# INTELLIGENT SYSTEMS FOR DECISION SUPPORT

by

Dongrui Wu

---

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(ELECTRICAL ENGINEERING)

May 2009

## Dedication

To my parents and Ying.

# Acknowledgments

I would like to express my first and most earnest gratitude to my advisor and my Committee Chair, Prof. Jerry M. Mendel, for his invaluable guidance, supervision, encouragement and constant support during my PhD study at the University of Southern California. His mentorship was paramount in shaping various aspects of my professional and personal life.

I am also sincerely thankful to Prof. C.-C. Jay Kuo and Prof. Iraj Ershaghi for their advices and serving in my Qualifying and Dissertation Committees, Prof. Viktor K. Prasanna and Prof. B. Keith Jenkins for serving in my Qualifying Committee, and Prof. Woei Wan Tan with the National University of Singapore for helping me discover such a fruitful research field.

I must also thank the Center for Interactive Smart Oilfield Technologies (CiSoft), a joint USC-Chevron initiative, for providing me scholarships and wonderful research opportunities.

Special thanks are conveyed to my teammates in Prof. Mendel's group, Feilong Liu, Joaquín Rapela, Daoyuan Zhai and Jhiin Joo, and my peers in CiSoft. Discussions with them gave me lots of inspirations. They also made my life in USC more enjoyable and memorable.

My final and most heartfelt acknowledgment must go to my parents and my wife, Ying, for their encouragement and love. Without them this work would never have come into existence.

# Table of Contents

# List of Tables

# List of Figures

# Abstract

This research is focused on multi-criteria decision-making (MCDM) under uncertainties, especially linguistic uncertainties. This problem is very important because many times linguistic information, in addition to numerical information, is an essential input of decision-making. Linguistic information is usually uncertain, and it is necessary to incorporate and propagate this uncertainty during the decision-making process because uncertainty means risk.

MCDM problems can be classified into two categories: 1) multi-attribute decision-making (MADM), which selects the best alternative(s) from a group of candidates using multiple criteria, and 2) multi-objective decision-making (MODM), which optimizes conflicting objective functions under constraints. Perceptual Computer, an architecture for computing with words, is implemented in this dissertation for both categories. For MADM, we consider the most general case that the weights for and the inputs to the criteria are a mixture of numbers, intervals, type-1 fuzzy sets and/or words modeled by interval type-2 fuzzy sets. Novel weighted averages are proposed to aggregate this diverse and uncertain information so that the overall performance of each alternative can be computed and ranked. For MODM, we consider how to represent the dynamics of a process (objective function) by IF-THEN rules and then how to perform reasoning based on these rules, i.e., to compute the objective function for new linguistic inputs. Two approaches

for extracting IF-THEN rules are proposed: 1) linguistic summarization to extract rules from data, and 2) knowledge mining to extract rules through survey. Applications are shown for all techniques proposed in this dissertation.

# Chapter 1

# Introduction

## 1.1 Multi-Criteria Decision-Making (MCDM)

Decision-making is an essential part of everyday life, e.g., we make decisions on which restaurant to have dinner, which car to buy, how to design an optimal investment strategy to balance profit and risk, etc. Multi-criteria decision-making (MCDM) refers to making decision in the presence of multiple and often conflicting criteria, where *criteria* means the standards of judgment or rules to test acceptability [72]. All MCDM problems share the following common characteristics [46]:

- *Multiple criteria*: Each problem has multiple criteria.
- *Conflict among criteria*: Multiple criteria usually conflict with each other.
- *Incommensurable units*: Multiple criteria may have different units of measurement.

- *Design/selection*: Solutions to an MCDM problem are either to design the best alternative(s) or to select the best one(s) among a pre-specified finite set of alternatives.

In this dissertation we focus on MCDM under uncertainties, especially linguistic uncertainties. Uncertainties are emphasized here because according to Harvard Business Essentials ( [23], pp. 59), "*in business, uncertainty of outcome is synonymous with risk, and you must factor it into your evaluation.*"

MCDM problems can be broadly classified into two categories — *multi-attribute decision-making* (MADM) and *multi-objective decision-making* (MODM). The main difference between them is that MADM focuses on discrete decision spaces whereas MODM focuses on continuous decision spaces [192].

## 1.2 Multi-Attribute Decision-Making (MADM)

A typical MADM problem is formulated as [72]:

$$\text{(MADM)} \begin{cases} \text{select } A_i \text{ from } \ A_1, \ldots, A_n \\ \text{using } \ C_1, \ldots, C_m \end{cases} \tag{1.1}$$

where $\{A_1, \ldots, A_n\}$ denotes $n$ alternatives, and $\{C_1, \ldots, C_m\}$ represents $m$ criteria. The selection is usually based on maximizing a multi-attribute utility function.

There are four steps in MCDM:

1. Define the problem.

2. Identify alternatives.

3. Evaluate the alternatives.

4. Identify the best alternative(s).

The last two steps are particularly difficult because the alternatives may have diverse inputs and uncertainties, as illustrated by the following [15, 16]:

**Example 1** *A contractor has to decide which of the three companies (A, B or C) is going to get the final mass production contract for a missile system. The contractor uses five criteria to base his/her final decision (see the first column in Table 1.1), namely:* tactics, technology, maintenance, economy *and* advancement*. Each of these criteria has some associated sub-criteria, e.g., for* tactics *there are seven sub-criteria, namely* effective range, flight height, flight velocity, reliability, firing accuracy, destruction rate, *and* kill radius, *whereas for* economy *there are three sub-criteria, namely* system cost, system life *and* material limitation*. Each criterion and sub-criterion also has its weight, as shown in the second column of Table 1.1, where $\tilde{n}$ means a type-1 fuzzy set "about n." The performances of the three companies for each sub-criterion are also given in Table 1.1. Observe that some of them are* words *instead of numbers.*

*To select the best missile system, the decision-maker needs to determine:*

1. *How to model linguistic uncertainties expressed by the words.*

2. *How to aggregate the diverse inputs and weights consisting of numbers, type-1 fuzzy sets, and words. Presently no method in the literature can do this.*

*3. How to rank the final aggregated results to find the best missile system.* ■

These questions will be answered in Chapters 2-4, and we will return to this example in Chapter 5.

Table 1.1: Criteria with their weights, sub-criteria with their weights and sub-criteria data for the three companies [15, 16].

| Item | Weighting | Company A | Company B | Company C |
|---|---|---|---|---|
| **Criterion 1: Tactics** | $\tilde{\mathbf{9}}$ | | | |
| 1. Effective range (km) | $\tilde{7}$ | 43 | 36 | 38 |
| 2. Flight height (m) | $\tilde{1}$ | 25 | 20 | 23 |
| 3. Flight velocity (M. No) | $\tilde{9}$ | 0.72 | 0.80 | 0.75 |
| 4. Reliability (%) | $\tilde{9}$ | 80 | 83 | 76 |
| 5. Firing accuracy (%) | $\tilde{9}$ | 67 | 70 | 63 |
| 6. Destruction rate (%) | $\tilde{7}$ | 84 | 88 | 86 |
| 7. Kill radius (m) | $\tilde{6}$ | 15 | 12 | 18 |
| **Criterion 2: Technology** | $\tilde{\mathbf{3}}$ | | | |
| 8. Missile scale (cm) (l×d–span) | $\tilde{4}$ | 521×35–135 | 381×34–105 | 445×35–120 |
| 9. Reaction time (min) | $\tilde{9}$ | 1.2 | 1.5 | 1.3 |
| 10. Fire rate (round/min) | $\tilde{9}$ | 0.6 | 0.6 | 0.7 |
| 11. Anti-jam (%) | $\tilde{8}$ | 68 | 75 | 70 |
| 12. Combat capability | $\tilde{9}$ | Very Good | Good | Good |
| **Criterion 3: Maintenance** | $\tilde{\mathbf{1}}$ | | | |
| 13. Operation condition requirement | $\tilde{5}$ | High | Low | Low |
| 14. Safety | $\tilde{6}$ | Very Good | Good | Good |
| 15. Defilade[a] | $\tilde{2}$ | Good | Very Good | Good |
| 16. Simplicity | $\tilde{3}$ | Good | Good | Good |
| 17. Assembly | $\tilde{3}$ | Good | Good | Poor |
| **Criterion 4: Economy** | $\tilde{\mathbf{5}}$ | | | |
| 18. System cost (10,000) | $\tilde{8}$ | 800 | 755 | 785 |
| 19. System life (years) | $\tilde{8}$ | 7 | 7 | 5 |
| 20. Material limitation | $\tilde{5}$ | High | Low | Low |
| **Criterion 5: Advancement** | $\tilde{\mathbf{7}}$ | | | |
| 21. Modularization | $\tilde{5}$ | Average[b] | Good | Average[b] |
| 22. Mobility | $\tilde{7}$ | Poor | Very Good | Good |
| 23. Standardization | $\tilde{3}$ | Good | Good | Very Good |

[a] *Defilade* means to surround by defensive works so as to protect the interior when in danger of being commanded by an enemy's guns.

[b] The word *general* used in [16] has been replaced by the word *average*, because it was not clear to us what *general* meant.

## 1.3 Multi-Objective Decision-Making (MODM)

Mathematically, an MODM problem can be formulated as [72]:

$$\text{(MODM)} \begin{cases} \max & \mathbf{f}(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in R^m, \mathbf{g}(\mathbf{x}) \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0} \end{cases} \tag{1.2}$$

where $\mathbf{f}(\mathbf{x})$ represents $n$ conflicting objective functions, $\mathbf{g}(\mathbf{x}) \leq \mathbf{b}$ represents $m$ constraints in continuous decision spaces, and $\mathbf{x}$ is an $m$-vector of decision variables.

To solve an MODM problem, first the objective functions $\mathbf{f}(\mathbf{x})$ must be defined. Sometimes this is trivial, e.g., in [2, 133] a subset of transportation projects are selected for implementation subject to budget constraints, and the objective function is a weighted average of these projects' impacts on traffic flow, traveler's safety, economic growth, environment, etc. However, sometimes the objective functions are difficult to calculate, as illustrated by the following:

**Example 2** *Fracture stimulation in an oilfield is to inject specially engineered fluids under high pressure into the channels of a low permeability reservoir to crack the reservoir and hence improve the flow of oil. It is a complex process involving many parameters [24, 42, 52, 92], e.g., the porosity and permeability of the reservoir, the number of stages, the number of holes and the length of perforations during well completion, the injected sand, pad and slurry volumes during fracture stimulation, etc. The last six parameters are adjustable; so, an interesting problem is to optimize fracture simulation design by maximizing post-fracturing oil production under a cost constraint. Unfortunately,* *post-fracturing oil production is difficult to compute because a model is needed to predict it from well parameters whereas it is very difficult to find such a model. Presently all successful approaches [52, 92] are black-box models, i.e., they are not very useful in helping people understand the relationship between fracture parameters and the post-fracturing oil production. An approach that can describe the dynamics of fracture stimulation and also is easily understandable, e.g., in terms of IF-THEN rules, would be highly desirable.* ■

Such an approach, called linguistic summarization, will be introduced in Chapter 6.

Linguistic summarization extracts rule from data; however, sometimes we do not have such training data, as illustrated by the following:

**Example 3** *A Social Judgment Advisor (SJA) is developed in [153] to describe the relationship between behavioral indicators [65] (e.g., touching, eye contact, acting witty, primping, etc) and flirtation level. Because numerical values are not appropriate for such an application and there is no training data, words are used in a survey to obtain IF-THEN rules. However, different people may give different responses for the same scenario; so, the survey result for each question is usually a histogram instead of a single word. How should a rulebase be constructed from such word histograms?* ■

This problem will be solved by a knowledge mining approach and we will return to this example in Chapter 7.

Once a rulebase is constructed, either from linguistic summarization or knowledge mining, it can be used to compute a linguistic objective function for new inputs. How this can be done will be shown in Chapter 8.

## 1.4 Perceptual Computer (Per-C) for MCDM

The above three examples show that *computing with words* (CWW) is very important for MCDM under linguistic uncertainties. According to Zadeh [181, 183], the father of fuzzy logic, CWW is *"a methodology in which the objects of computation are words and propositions drawn from a natural language."* It is *"inspired by the remarkable human capability to perform a wide variety of physical and mental tasks without any measurements and any computations."* Nikravesh [101] further pointed out that CWW is *"fundamentally different from the traditional expert systems which are simply tools to 'realize' an intelligent system, but are not able to process natural language which is imprecise, uncertain and partially true."*

There are at least two types of uncertainties associated with a word [135]: *intra-personal uncertainty* and *inter-personal uncertainty*. The former is explicitly pointed out by Wallsten and Budescu [135] as *"except in very special cases, all representations are vague to some degree in the minds of the originators and in the minds of the receivers,"* and they suggest to model it by a type-1 fuzzy set (T1 FS, Section 2.1.1). The latter is pointed out by Mendel [83] as *"words mean different things to different people"* and Wallsten and Budescu [135] as *"different individuals use diverse expressions to describe identical situations and understand the same phrases differently when hearing or reading them."* Because an interval type-2 FS (IT2 FS, Section 2.2.1) can be viewed as a group of T1 FSs, it can model both types of uncertainty; hence, we suggest IT2 FSs be used in

CWW [77,79,83]. Additionally, Mendel [84] has explained why it is scientifically incorrect to model a word using a T1 FS.

A specific architecture is proposed in [78] for making subjective judgments by CWW, as shown in Fig. 1.2. It is called a *perceptual computer*—Per-C for short. In Fig. 1.2, the *encoder*[1] transforms linguistic perceptions into IT2 FSs that activate a *CWW engine*. The CWW engine performs operations on the IT2 FSs. The *decoder*[2] maps the output of the CWW engine into a recommendation, which can be a word, rank, or class.

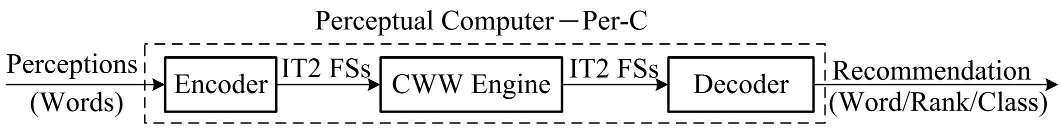



Fig. 1.1: Conceptual structure of Per-C.

When specified to MCDM, the Per-C can be described by the diagram shown in Fig. 1.2. The encoder has been studied by Liu and Mendel [70]. This dissertation proposes methods to construct the CWW engines and the decoder so that the Per-C can be completed.

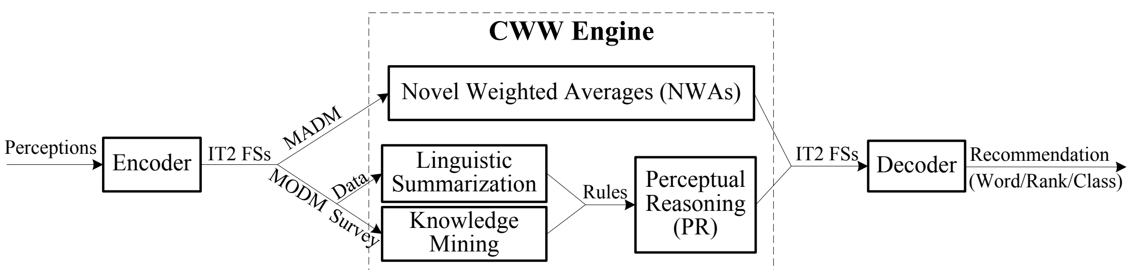



Fig. 1.2: Per-C for MCDM.

[1] Zadeh calls this *constraint explicitation* in [181, 183]. In [184, 185] and some of his recent talks, he calls this *precisiation*.

[2] Zadeh calls this *linguistic approximation* in [181, 183].

## 1.5 Dissertation Outline

The rest of this dissertation is organized as follows. Chapter 2 introduce the basic concepts of type-1 and interval type-2 fuzzy sets and systems, and Liu and Mendel's Interval Approach for word modeling. Chapter 3 is about decoding. Because the output of Per-C is a recommendation in the form of word, rank or class, three different decoders corresponding to these three outputs are proposed. Chapter 4 proposes novel weighted averages as a CWW engine for MADM, which can aggregate mixed signals, e.g., numbers, intervals, T1 FSs and words modeled by IT2 FSs. Chapter 5 applies NWA to the missile evaluation application introduced in Example 1. Chapter 6 introduces linguistic summarization, a data mining approach to extract rules from data. Chapter 7 introduces a knowledge mining approach to construct rules through survey. Chapter 8 proposes perceptual reasoning as a CWW engine for MODM, which performs approximate reasoning based on rules. Finally, Chapter 9 draws conclusions and proposes future works.

# Chapter 2

# Background Knowledge

Fuzzy set theory was first introduced by Zadeh [177] in 1965 and has been successfully used in many areas, including modeling and control [6, 9, 51, 129, 141, 157–160, 171], data mining [3, 44, 104, 132, 169, 173, 182], time-series prediction [58, 68, 134], decision making [83, 87, 88, 152, 153], etc. Background knowledge on type-1 and interval type-2 fuzzy sets and systems, and Liu and Mendel's Interval Approach for word modeling [70] are briefly introduced in this chapter.

## 2.1 Type-1 Fuzzy Logic

### 2.1.1 Type-1 Fuzzy Sets (T1 FSs)

**Definition 1** *A* type-1 fuzzy set *(T1 FS) $X$ is comprised of a* domain $D_X$ *of real numbers (also called the* universe of discourse *of $X$) together with a* membership function *(MF)* $\mu_X : D_X \rightarrow [0,1]$, *i.e.,*

$$X = \int_{D_X} \mu_X(x)/x \tag{2.1}$$

*Here* $\int$ *denotes the collection of all points* $x \in D_X$ *with associated* membership grade $\mu_X(x)$. ■

Two examples of T1 FSs are shown in Fig. 2.1. A T1 FS $X$ and its MF $\mu_X(x)$ are synonyms and are therefore used interchangeably, i.e., $X \Leftrightarrow \mu_X(x)$. Additionally, the terms *membership*, *membership function* and *membership grade* are also used interchangeably.

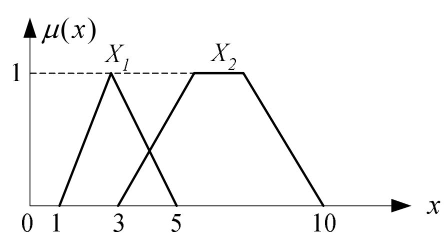



Fig. 2.1: Examples of T1 FSs. The universe of discourse is [0, 10].

In general, MFs can either be chosen arbitrarily, based on the experience of an individual (hence, the MFs for two individuals could be quite different depending upon their experiences, perspectives, cultures, etc.), or, they can be designed using optimization procedures [45, 50, 138, 139].

The centroid of a T1 FS is equivalent to the mean of a random variable in probability theorem, and hence it is useful in ranking T1 FSs [144, 164].

**Definition 2** *The* centroid *of a T1 FS X is defined as*

$$c(X) = \frac{\int_{D_X} x\mu_X(x)dx}{\int_{D_X} \mu_X dx} \tag{2.2}$$

*if $D_X$ is continuous, or*

$$c(X) = \frac{\sum_{i=1}^{N} x_i\mu_X(x_i)}{\sum_{i=1}^{N} \mu_X(x_i)} \tag{2.3}$$

*if $D_X$ is discrete.* ■

Cardinality of a crisp set is a count of the number of elements in that set. Cardinality of a T1 FS is more complicated because the elements of the FS are not equally weighted as they are in a crisp set. Definitions of the cardinality of a T1 FS have been proposed by several authors [7, 22, 38, 60, 63, 161, 180]. Basically there have been two kinds of proposals [26]: (1) those that assume that the cardinality of a T1 FS can be a crisp number, and (2) those that claim that it should be a fuzzy number. De Luca and Termini's [22] definition of the cardinality of a T1 FS is used in this dissertation.

**Definition 3** *The* cardinality *of a T1 FS $X$ is defined as*

$$\text{card}(X) = \int_{D_X} \mu_X(x)dx \tag{2.4}$$

*when $D_X$ is continuous, or*

$$\text{card}(X) = \sum_{i=1}^{N} \mu_X(x_i) \tag{2.5}$$

*when $D_X$ is discrete.* ■

### 2.1.2 Set Theoretic Operations for T1 FSs

Just as crisp sets can be combined using the union and intersection operations, so can FSs.

**Definition 4** *Let T1 FSs $X_1$ and $X_2$ be two T1 FS in $D_X$ that are described by their MFs $\mu_{X_1}(x)$ and $\mu_{X_2}(x)$. The* union *of $X_1$ and $X_2$, $X_1 \cup X_2$, is described by its MF $\mu_{X_1 \cup X_2}(x)$, where*

$$\mu_{X_1 \cup X_2}(x) = \max[\mu_{X_1}(x), \mu_{X_2}(x)] \quad \forall x \in X. \tag{2.6}$$

*The* intersection *of $X_1$ and $X_2$, $X_1 \cap X_2$, is described by its MF $\mu_{X_1 \cap X_2}(x)$, where*

$$\mu_{X_1 \cap X_2}(x) = \min[\mu_{X_1}(x), \mu_{X_2}(x)] \quad \forall x \in X. \quad ■ \tag{2.7}$$

Although $\mu_{X_1 \cup X_2}(x)$ and $\mu_{X_1 \cap X_2}(x)$ can be described using different $t$-conorms and $t$-norms [64], in this paper only the maximum $t$-conorm and the minimum $t$-norm are used in (2.6) and (2.7), respectively.

**Example 4** *The union and intersection of the two TI FSs that are depicted in Fig. 2.1 are shown in Figs. 2.2(a) and 2.2(b), respectively.* ■

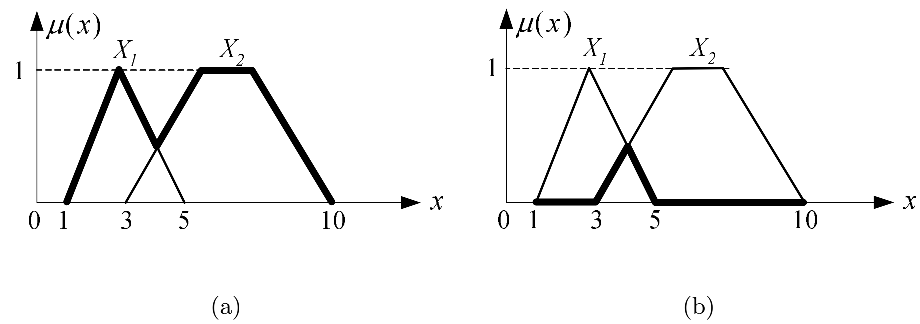



Fig. 2.2: Set theoretic operations for the two T1 FSs $X_1$ and $X_2$ depicted in Fig. 2.1. (a) Union and (b) intersection.

### 2.1.3 $\alpha$-cuts and a Decomposition Theorem for T1 FSs

**Definition 5** *The $\alpha$-cut of T1 FS $X$, denoted $X(\alpha)$, is an interval of real numbers, defined as:*

$$X(\alpha) = \{x | \mu_X(x) \geq \alpha\} = [a(\alpha), b(\alpha)] \tag{2.8}$$

*where* $0 \leq \alpha \leq 1$. ■

An example of an $\alpha$-cut is depicted in Fig. 2.3, and in this example, $X(\alpha) = [2.8, 5.2]$.

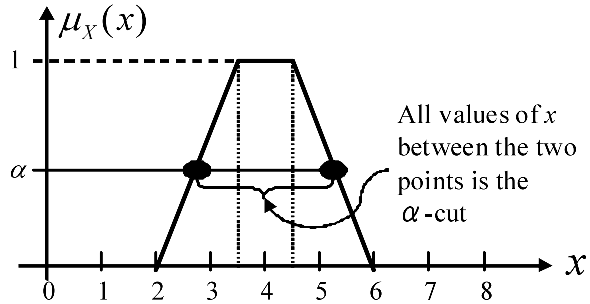



Fig. 2.3: A trapezoidal T1 FS and an $\alpha$-cut.

One of the major roles of $\alpha$-cuts is their capability to represent a T1 FS. In order to do this, first the following *indicator function* is introduced:

$$I_{X(\alpha)}(x) = \begin{cases} 1, & \forall\ x \in X(\alpha) \\ 0, & \forall\ x \notin X(\alpha) \end{cases} \tag{2.9}$$

Associated with $I_{X(\alpha)}(x)$ is the following *square-well function*:

$$\mu_X(x|\alpha) = \alpha I_{X(\alpha)}(x) \tag{2.10}$$

This function, an example of which is depicted in Fig. 2.4, raises the $\alpha$-cut $X(\alpha)$ off of the $x$-axis to height $\alpha$.

**Theorem 1** ***(T1 FS Decomposition Theorem)*** *[64] A T1 FS $X$ can be represented as:*

$$\mu_X(x) = \bigcup_{\alpha\in[0,1]} \mu_X(x|\alpha) \tag{2.11}$$

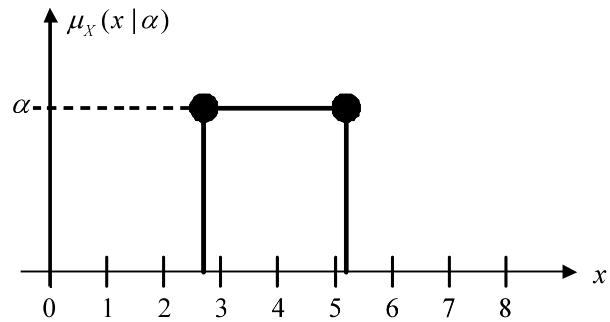



Fig. 2.4: Square-well function $\mu_X(x|\alpha)$.

*where* $\mu_X(x|\alpha)$ *is defined in (2.10) and* $\cup$ *(which is over all values of* $\alpha$*) denotes the standard union operator, i.e. the* supremum *(often the* maximum*) operator.* ■

This theorem is called a "Decomposition Theorem" [64] because $X$ is decomposed into a collection of square well functions that are then aggregated using the union operation.

An example of (2.11) is depicted in Fig. 2.5. When the dark circles at each $\alpha$-level (e.g., $\alpha_3$) are connected, $\mu_X(x|\alpha)$ is obtained. Note that greater resolution is obtained by including more $\alpha$-cuts, and the calculation of new $\alpha$-cuts does not affect previously calculated $\alpha$-cuts.


Fig. 2.5: Illustration of the T1 FS Decomposition Theorem when $n$ $\alpha$-cuts are used.

### 2.1.4 Type-1 Fuzzy Logic System (T1 FLS)

A *T1 fuzzy logic system* (FLS) uses only T1 FSs. It consists of four components — *rulebase, fuzzifier, inference engine* and *defuzzifier*, as shown in Fig. 2.6. The fuzzifier maps the crisp inputs into T1 FSs. The inference engine operates on these T1 FSs according to the rules in the rulebase, and the results are also T1 FSs, which will be mapped into a crisp output by the defuzzifier. By a *rule*, we mean an IF-THEN statement, such as:

$$R^i: \text{ IF } x_1 \text{ is } F_1^i \text{ and } \cdots \text{ and } x_p \text{ is } F_p^i, \text{ THEN } y \text{ is } G^i \qquad i = 1, ..., N \tag{2.12}$$

where $F_j^i$ and $G^i$ are T1 FSs.

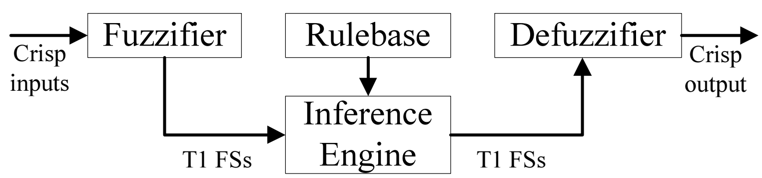



Fig. 2.6: A type-1 fuzzy logic system.

## 2.2 Interval Type-2 Fuzzy Logic

### 2.2.1 Interval Type-2 Fuzzy Sets (IT2 FSs)

Despite having a name which carries the connotation of uncertainty, researches have shown that there are limitations in the ability of T1 FSs to model and minimize the effect of uncertainties [39, 83, 159]. This is because a T1 FS is certain in the sense that its membership grades are crisp values. Recently, type-2 FSs [179], characterized by MFs that are themselves fuzzy, have been attracting interests. Interval type-2 (IT2) FSs [83], a special case of type-2 FSs, are considered in this dissertation for their reduced computational cost.

**Definition 6** *[80, 83] An IT2 FS $\tilde{X}$ is characterized by its MF $\mu_{\tilde{X}}(x,u)$, i.e.,*

$$\begin{aligned}
\widetilde{X} &= \int_{x\in D_{\tilde{X}}} \int_{u\in J_x\subseteq[0,1]} \mu_{\tilde{X}}(x,u)/(x,u) \\
&= \int_{x\in D_{\tilde{X}}} \int_{u\in J_x\subseteq[0,1]} 1/(x,u) \\
&= \int_{x\in D_{\tilde{X}}} \left[\int_{u\in J_x\subseteq[0,1]} 1/u\right] \Big/ x
\end{aligned} \tag{2.13}$$

*where $x$, called the* primary variable, *has domain $D_{\tilde{X}}$; $u \in [0,1]$, called the* secondary variable, *has domain $J_x \subseteq [0,1]$ at each $x \in D_{\tilde{X}}$; $J_x$ is also called the* primary membership *of $x$, and is defined below in (2.15); and, the amplitude of $\mu_{\tilde{X}}(x,u)$, called a* secondary grade *of $\tilde{X}$, equals 1 for $\forall x \in D_{\tilde{X}}$ and $\forall u \in J_x \subseteq [0,1]$.* ■

An example of an IT2 FS is shown in Fig. 2.7.

**Definition 7** *Uncertainty about $\tilde{X}$ is conveyed by the union of all its primary memberships, which is called the* footprint of uncertainty *(FOU) of $\tilde{X}$ (see Fig. 2.7), i.e.,*

$$FOU(\tilde{X}) = \bigcup_{\forall x\in D_{\tilde{X}}} J_x = \{(x,u) : u \in J_x \subseteq [0,1]\}. \quad \blacksquare \tag{2.14}$$

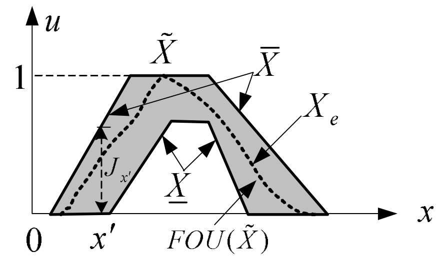



Fig. 2.7: An interval type-2 fuzzy set.

The size of an FOU is directly related to the uncertainty that is conveyed by an IT2 FS. So, an FOU with more area is more uncertain than one with less area.

**Definition 8** *The* upper membership function *(UMF) and* lower membership function *(LMF) of $\tilde{X}$ are two T1 MFs $\overline{X}$ and $\underline{X}$ that bound the FOU (see Fig. 2.7).* ■

Note that the primary membership $J_x$ is an *interval*, i.e.,

$$J_x = \left[\mu_{\underline{X}}(x), \mu_{\overline{X}}(x)\right] \tag{2.15}$$

Using (2.15), $FOU(\tilde{X})$ can also be expressed as

$$FOU(\tilde{X}) = \bigcup_{\forall x \in D_{\tilde{X}}} \left[\mu_{\underline{X}}(x), \mu_{\overline{X}}(x)\right] \tag{2.16}$$

A very compact way to describe an IT2 FS is:

$$\tilde{X} = 1/FOU(\tilde{X}) \tag{2.17}$$

where this notation means that the secondary grade equals 1 for all elements of $FOU(\tilde{X})$. Because all of the secondary grades of an IT2 FS equal 1, these secondary grades convey no useful information; hence, *an IT2 FS is completely described by its FOU.*

**Definition 9** *For continuous universes of discourse $D_{\tilde{X}}$ and $U$, an* embedded T1 FS $X^e$ *is*

$$X^e = \int_{x \in D_{\tilde{X}}} u/x, \quad u \in J_x. \quad \blacksquare \tag{2.18}$$

The set $X^e$ is embedded in $FOU(\tilde{X})$. An example of $X^e$ is given in Fig. 2.7. Other examples are $\overline{X}$ and $\underline{X}$.

**Definition 10** *The* centroid *of an IT2 FS is an interval determined by the centroids of all its embedded T1 FSs, i.e.,*

$$C(\tilde{X}) = [c_l(\tilde{X}), c_r(\tilde{X})] \tag{2.19}$$

*where*

$$c_l(\tilde{X}) = \min_{\forall \mu_X(x_i) \in [\mu_{\underline{X}}(x_i), \mu_{\bar{X}}(x_i)]} \frac{\sum_{i=1}^N x_i \mu_X(x_i)}{\sum_{i=1}^N \mu_X(x_i)} \tag{2.20}$$

$$c_r(\tilde{X}) = \max_{\forall \mu_X(x_i) \in [\mu_{\underline{X}}(x_i), \mu_{\bar{X}}(x_i)]} \frac{\sum_{i=1}^N x_i \mu_X(x_i)}{\sum_{i=1}^N \mu_X(x_i)} \tag{2.21}$$

*in which $N$ is the number of discretizations in $D_{\tilde{X}}$.* $\blacksquare$

It has been shown [57, 83] that $c_l(\tilde{X})$ and $c_r(\tilde{X})$ can be re-expressed as:

$$c_l(\tilde{X}) = \frac{\sum_{i=1}^{L} x_i \mu_{\overline{X}}(x_i) + \sum_{i=L+1}^{N} x_i \mu_{\underline{X}}(x_i)}{\sum_{i=1}^{L} \mu_{\overline{X}}(x_i) + \sum_{i=L+1}^{N} \mu_{\underline{X}}(x_i)} \tag{2.22}$$

$$c_r(\tilde{X}) = \frac{\sum_{i=1}^{R} x_i \mu_{\underline{X}}(x_i) + \sum_{i=R+1}^{N} x_i \mu_{\overline{X}}(x_i)}{\sum_{i=1}^{R} \mu_{\underline{X}}(x_i) + \sum_{i=R+1}^{N} \mu_{\overline{X}}(x_i)} \tag{2.23}$$

where $L$ and $R$ are called *switch points*. There are no closed-form solutions for $L$ and $R$; however, they can be computed iteratively by the Karnik-Mendel (KM) [57, 83] or Enhanced KM (EKM) Algorithms presented in Appendix A.

The centroid of an IT2 FS provides a legitimate measure of its uncertainty [147]. The average centroid, or the center of centroid, of an IT2 FS is very useful in ranking IT2 FSs ( [144]; see also Section 3.3).

**Definition 11** *The* average centroid, *or* center of centroid, *of an IT2 FS $\tilde{X}$ is*

$$c(\tilde{X}) = \frac{c_l(\tilde{X}) + c_r(\tilde{X})}{2}. \quad \blacksquare \tag{2.24}$$

The average cardinality [147] of an IT2 FS is very useful in computing the average subsethood of an IT2 FS in another and hence in decoding ( [151]; see also Section 3.4.2).

**Definition 12** *The* cardinality *of an IT2 FS $\tilde{X}$ is the union of the cardinalities of all its embedded T1 FSs $X^e$, i.e.,*

$$\text{card}(\tilde{X}) = \bigcup_{\forall X^e} \text{card}(X^e) = [\text{card}(\underline{X}), \text{card}(\overline{X})] \tag{2.25}$$

*The* average cardinality *of an IT2 FS $\tilde{X}$, $AC(\tilde{X})$, is the center of its cardinality, i.e.,*

$$\mathrm{AC}(\tilde{X}) = \frac{\mathrm{card}(\underline{X}) + \mathrm{card}(\overline{X})}{2}. \quad \blacksquare \tag{2.26}$$

### 2.2.2 Representation Theorems for IT2 FSs

So far the *vertical-slice representation (decomposition)* of an IT2 FS, given in (2.16), has been emphasized. In this section a different representation is provided for such an IT2 FS, one that is in terms of so-called *wavy slices* [80]. It is stated here for a discrete IT2 FS.

**Theorem 2** ***(Wavy Slice Representation Theorem for an IT2 FS)*** *Assume that primary variable $x$ of an IT2 FS $\tilde{X}$ is sampled at $N$ values, $x_1, x_2, \ldots, x_N$, and at each of these values its primary memberships $u_i$ are sampled at $M_i$ values, $u_{i1}, u_{i2}, \ldots, u_{iM_i}$. Let $X^{e,j}$ denote the $j^{\mathrm{th}}$ embedded T1 FS for $\tilde{X}$. Then $FOU(\tilde{X})$ in (2.17) can be represented as*

$$FOU(\tilde{X}) = \bigcup_{j=1}^{n_X} X^{e,j} \equiv [\underline{X}, \overline{X}] \tag{2.27}$$

*and $n_X = \prod_{i=1}^{N} M_i$.* $\blacksquare$

This theorem expresses $FOU(\tilde{X})$ as a union of simple T1 FSs. Note that both the union of the vertical slices and the union of embedded T1 FSs can be interpreted as *covering representations*, because they both cover the entire FOU.

In the sequel it will be seen that one does not need to know the explicit natures of any of the wavy slices in $FOU(\tilde{X})$ other than $\mu_{\underline{X}}(x)$ and $\mu_{\overline{X}}(x)$. In fact, for an IT2 FS, everything can be determined just by knowing its lower and upper MFs.

### 2.2.3 Set Theoretic Operations for IT2 FSs

The Wavy-Slice Representation Theorem and the formulas for the union and intersection of two T1 FSs can be used to derive the union and intersection of two IT2 FSs [83].

**Definition 13** *For continuous universes of discourse, (a) the* union *of two IT2 FSs,* $\tilde{X}_1$ *and* $\tilde{X}_2$*,* $\tilde{X}_1 \cup \tilde{X}_2$*, is another IT2 FS, i.e.,*

$$\tilde{X}_1 \cup \tilde{X}_2 = 1/FOU(\tilde{X}_1 \cup \tilde{X}_2) = 1/[\mu_{\underline{X}_1}(x) \vee \mu_{\underline{X}_2}(x), \mu_{\overline{X}_1}(x) \vee \mu_{\overline{X}_2}(x)] \tag{2.28}$$

*where* $\vee$ *denotes the* disjunction *operator (e.g.,* maximum*); (b) the* intersection *of two IT2 FSs,* $\tilde{X}_1$ *and* $\tilde{X}_2$*,* $\tilde{X}_1 \cap \tilde{X}_2$*, is also another IT2 FS, i.e.,*

$$\tilde{X}_1 \cap \tilde{X}_2 = 1/FOU(\tilde{X}_1 \cap \tilde{X}_2) = 1/[\mu_{\underline{X}_1}(x) \wedge \mu_{\underline{X}_2}(x), \mu_{\overline{X}_1}(x) \wedge \mu_{\overline{X}_2}(x)] \tag{2.29}$$

*where* $\wedge$ *denotes the* conjunction *operator (e.g.,* minimum*).* ■

It is very important to observe, from (2.28) and (2.29), that all of their calculations only involve calculations between T1 FSs.

**Example 5** *Two IT2 FSs,* $\tilde{X}_1$ *and* $\tilde{X}_2$ *are depicted in Fig. 2.8(a). Their union and intersection are depicted in Figs. 2.8(b) and 2.8(c), respectively.* ■

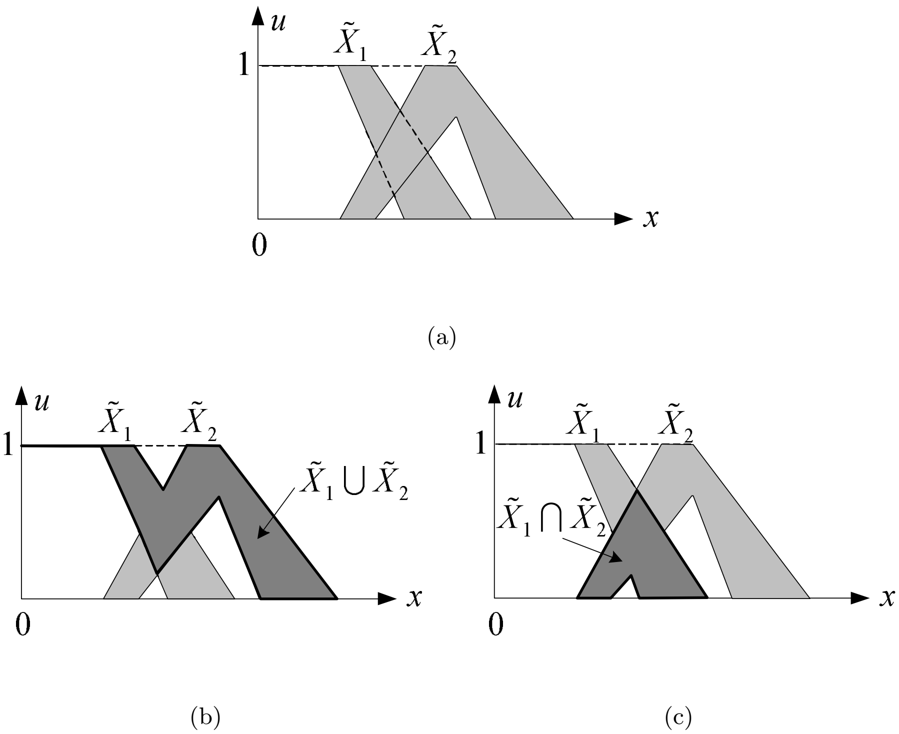



Fig. 2.8: Set theoretic operations for IT2 FSs. (a) Two IT2 FSs $\tilde{X}_1$ and $\tilde{X}_2$, (b) $\tilde{X}_1 \cup \tilde{X}_2$, and (c) $\tilde{X}_1 \cap \tilde{X}_2$.

### 2.2.4 Interval Type-2 Fuzzy Logic System (IT2 FLS)

An IT2 FLS is depicted in Fig. 2.9. Each input is fuzzified into an IT2 FS, after which these FSs activate a subset of rules in the form of

$$R^i: \text{ IF } x_1 \text{ is } \tilde{F}_1^i \text{ and } \cdots \text{ and } x_p \text{ is } \tilde{F}_p^i, \text{ THEN } y \text{ is } \tilde{G}^i \qquad i = 1, ..., N \tag{2.30}$$

where $\tilde{F}_j^i$ and $\tilde{G}^i$ are IT2 FSs. The output of each activated rule is obtained by using an extended sup-star composition [83]. Then all of the fired rule outputs are blended in some way and reduced from IT2 FSs to a number.

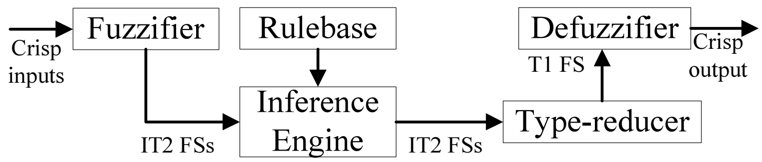



Fig. 2.9: An interval type-2 fuzzy logic system.

The first step in this chain of computations is to *compute a firing interval*. This can be a very complicated calculation, especially when the inputs are fuzzified into IT2 FSs, as they would be when the inputs are words. For the minimum $t$-norm, this calculation requires computing the sup-min operation between the lower (upper) MFs of the FOUs of each input and its corresponding antecedent [83]. The firing interval propagates the uncertainties from all of the inputs through their respective antecedents. An example of computing the firing interval is depicted in the left-hand part of Fig. 2.10 for a rule that has two antecedents.

For Mamdani Inference, the next computation after the firing interval computation is the *meet* operation between the firing interval and its consequent FOU, the result being a *fired-rule output FOU*. Then all fired rule output FOUs are aggregated using the *join* operator, the result being yet another FOU.

An example of this computing is depicted in the right-hand part of Fig. 2.10, and an example of aggregating two fired-rule output FOUs is depicted in Fig. 2.11. Fig. 2.11(a)

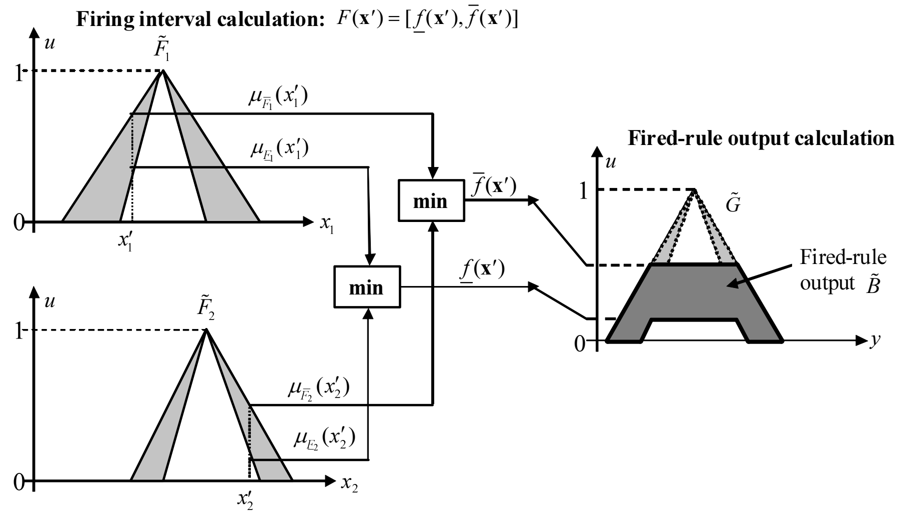



Fig. 2.10: Mamdani Inference for IT2 FLSs: from firing interval to fired-rule output FOU.

shows the fired-rule output sets for two fired rules, and Fig. 2.11(b) shows the union of those two IT2 FSs. Observe that the union tends to spread out the domain over which non-zero values of the output occur, and that $\tilde{B}$ does not have the appearance of either $\tilde{B}^1$ or $\tilde{B}^2$.

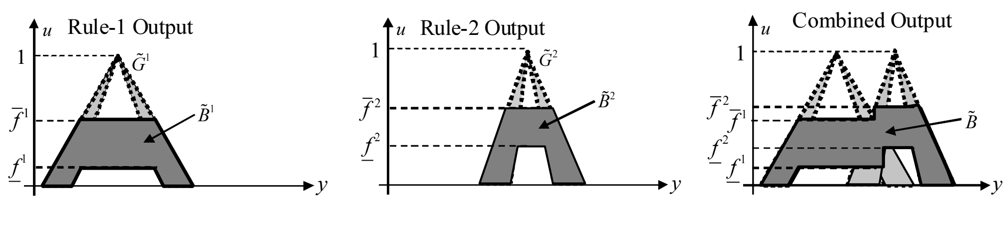



Fig. 2.11: Pictorial descriptions of (a) Fired-rule output FOUs for two fired rules, and (b) combined fired output FOU for the two fired rules in (a) using Mamdani Inference.

Referring to Fig. 2.9, this aggregated FOU is then type-reduced, the result being an interval-valued set, after which that interval is defuzzified by taking the average of the interval's two end-points.

## 2.3 Encoding: The Interval Approach

Liu and Mendel proposed an Interval Approach to for word modeling [70], i.e., to construct the decoder in Fig. 1.2. First, for each word in an application-dependent encoding vocabulary, a group of subjects are asked the following question:

> *On a scale of 0-10, what are the end-points of an interval that you associate with the word ____?*

After some pre-processing, during which some intervals (e.g., outliers) are eliminated, each of the remaining intervals is classified as either an interior, left-shoulder or right-shoulder IT2 FS. Then, each of the word's data intervals is individually mapped into its respective T1 interior, left-shoulder or right-shoulder MF, after which the union of all of these T1 MFs is taken. The result is an FOU for an IT2 FS model of the word. The words and their FOUs constitute a *codebook*.

The dataset used in this dissertation was collected from 28 subjects at the Jet Propulsion Laboratory[1] (JPL). 32 words were randomly ordered and presented to the subjects. Each subject was asked to provide the end points of an interval for each word on the

[1]This was done in 2002 when J. M. Mendel gave an in-house short course on fuzzy sets and systems at JPL.

scale 0-10. The 32 words can be grouped into three classes: small-sounding words (*little, low amount, somewhat small, a smidgen, none to very little, very small, very little, teeny-weeny, small amount* and *tiny*), medium-sounding words (*fair amount, modest amount, moderate amount, medium, good amount, a bit, some to moderate* and *some*), and large-sounding words (*sizeable, large, quite a bit, humongous amount, very large, extreme amount, considerable amount, a lot, very sizeable, high amount, maximum amount, very high amount* and *substantial amount*). The 32 word FOUs obtained from the interval approach are depicted in Fig. 2.12. Observe that only three kinds of FOUs emerge, namely, left-shoulder (the first six FOUs), right-shoulder (the last six FOUs) and interior FOUs. The parameters of the FOUs, their centroids and centers of centroid are given in Table 2.1. Note that each FOU can be represented by nine parameters shown in Fig. 2.13, where $\tilde{X}$ is a left-shoulder when $a = b = e = f = 0$ and $h = 1$, and $\tilde{X}$ is a right-shoulder when $c = d = g = i = 10$ and $h = 1$.

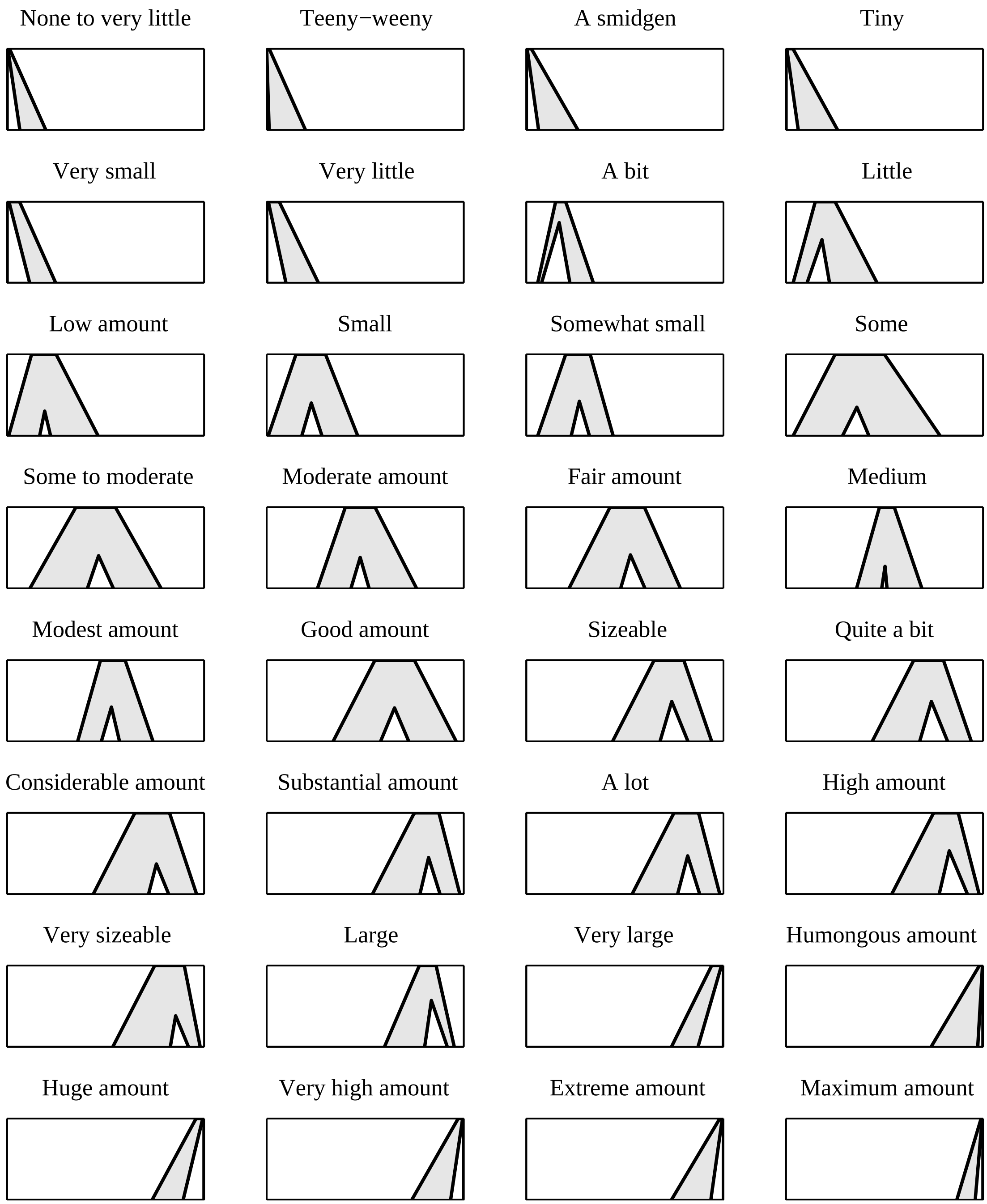



Fig. 2.12: The 32 word FOUs ranked by their centers of centroid. To read this figure, scan from left to right starting at the top of the page.

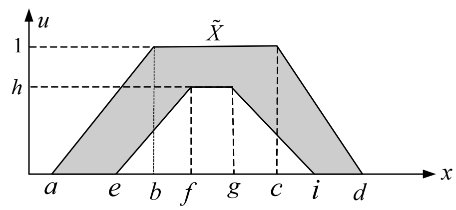



Fig. 2.13: The nine parameters to represent an IT2 FS.

Table 2.1: Parameters of the 32 word FOUs. Each UMF is represented by $[a, b, c, d]$ in Fig. 2.13 and each LMF is represented by $[e, f, g, i, h]$ in Fig. 2.13.

| Word | UMF | LMF | $C(\tilde{X})$ | $c(\tilde{X})$ |
|---|---|---|---|---|
| 1. None to very little | [0, 0, 0.14, 1.97] | [0, 0, 0.05, 0.66, 1] | [0.22,0.73] | 0.47 |
| 2. Teeny-weeny | [0, 0, 0.14, 1.97] | [0, 0, 0.01, 0.13, 1] | [0.05,1.07] | 0.56 |
| 3. A smidgen | [0, 0, 0.26, 2.63] | [0, 0, 0.05, 0.63, 1] | [0.21,1.05] | 0.63 |
| 4. Tiny | [0, 0, 0.36, 2.63] | [0, 0, 0.05, 0.63, 1] | [0.21,1.06] | 0.64 |
| 5. Very small | [0, 0, 0.64, 2.47] | [0, 0, 0.10, 1.16, 1] | [0.39,0.93] | 0.66 |
| 6. Very little | [0, 0, 0.64, 2.63] | [0, 0, 0.09, 0.99, 1] | [0.33,1.01] | 0.67 |
| 7. A bit | [0.59, 1.50, 2.00, 3.41] | [0.79, 1.68, 1.68, 2.21, 0.74] | [1.42,2.08] | 1.75 |
| 8. Little | [0.38, 1.50, 2.50, 4.62] | [1.09, 1.83, 1.83, 2.21, 0.53] | [1.31,2.95] | 2.13 |
| 9. Low amount | [0.09, 1.25, 2.50, 4.62] | [1.67, 1.92, 1.92, 2.21, 0.30] | [0.92,3.46] | 2.19 |
| 10. Small | [0.09, 1.50, 3.00, 4.62] | [1.79, 2.28, 2.28, 2.81, 0.40] | [1.29,3.34] | 2.32 |
| 11. Somewhat small | [0.59, 2.00, 3.25, 4.41] | [2.29, 2.70, 2.70, 3.21, 0.42] | [1.76,3.43] | 2.59 |
| 12. Some | [0.38, 2.50, 5.00, 7.83] | [2.88, 3.61, 3.61, 4.21, 0.35] | [2.04,5.77] | 3.90 |
| 13. Some to moderate | [1.17, 3.50, 5.50, 7.83] | [4.09, 4.65, 4.65, 5.41, 0.40] | [3.02,6.11] | 4.56 |
| 14. Moderate amount | [2.59, 4.00, 5.50, 7.62] | [4.29, 4.75, 4.75, 5.21, 0.38] | [3.74,6.16] | 4.95 |
| 15. Fair amount | [2.17, 4.25, 6.00, 7.83] | [4.79, 5.29, 5.29, 6.02, 0.41] | [3.85,6.41] | 5.13 |
| 16. Medium | [3.59, 4.75, 5.50, 6.91] | [4.86, 5.03, 5.03, 5.14, 0.27] | [4.19,6.19] | 5.19 |
| 17. Modest amount | [3.59, 4.75, 6.00, 7.41] | [4.79, 5.30, 5.30, 5.71, 0.42] | [4.57,6.24] | 5.41 |
| 18. Good amount | [3.38, 5.50, 7.50, 9.62] | [5.79, 6.50, 6.50, 7.21, 0.41] | [5.11,7.89] | 6.50 |
| 19. Sizeable | [4.38, 6.50, 8.00, 9.41] | [6.79, 7.38, 7.38, 8.21, 0.49] | [6.17,8.15] | 7.16 |
| 20. Quite a bit | [4.38, 6.50, 8.00, 9.41] | [6.79, 7.38, 7.38, 8.21, 0.49] | [6.17,8.15] | 7.16 |
| 21. Considerable amount | [4.38, 6.50, 8.25, 9.62] | [7.19, 7.58, 7.58, 8.21, 0.37] | [5.97,8.52] | 7.25 |
| 22. Substantial amount | [5.38, 7.50, 8.75, 9.81] | [7.79, 8.22, 8.22, 8.81, 0.45] | [6.95,8.86] | 7.90 |
| 23. A lot | [5.38, 7.50, 8.75, 9.83] | [7.69, 8.19, 8.19, 8.81, 0.47] | [6.99,8.83] | 7.91 |
| 24. High amount | [5.38, 7.50, 8.75, 9.81] | [7.79, 8.30, 8.30, 9.21, 0.53] | [7.19,8.82] | 8.01 |
| 25. Very sizeable | [5.38, 7.50, 9.00, 9.81] | [8.29, 8.56, 8.56, 9.21, 0.38] | [6.95,9.10] | 8.03 |
| 26. Large | [5.98, 7.75, 8.60, 9.52] | [8.03, 8.36, 8.36, 9.17, 0.57] | [7.50,8.75] | 8.12 |
| 27. Very large | [7.37, 9.41, 10, 10] | [8.72, 9.91, 10, 10, 1] | [9.03,9.57] | 9.30 |
| 28. Humongous amount | [7.37, 9.82, 10, 10] | [9.74, 9.98, 10, 10, 1] | [8.70,9.91] | 9.31 |
| 29. Huge amount | [7.37, 9.59, 10, 10] | [8.95, 9.93, 10, 10, 1] | [9.03,9.65] | 9.34 |
| 30. Very high amount | [7.37, 9.73, 10, 10] | [9.34, 9.95, 10, 10, 1] | [8.96,9.78] | 9.37 |
| 31. Extreme amount | [7.37, 9.82, 10, 10] | [9.37, 9.95, 10, 10, 1] | [8.96,9.79] | 9.38 |
| 32. Maximum amount | [8.68, 9.91, 10, 10] | [9.61, 9.97, 10, 10, 1] | [9.50,9.87] | 9.69 |

# Chapter 3

# Decoding: From FOUs to a Recommendation

## 3.1 Introduction

Recall that a Per-C (Fig. 1.2) consists of three components: Encoder, which maps words into IT2 FS models; CWW engine, which operates on the inputs words and whose outputs are FOU(s); and decoder, which maps these FOU(s) into a recommendation. The decoder is discussed in this chapter.

The recommendation from the decoder can have several different forms:

1. *Word*: This is the most typical case, e.g., for the social judgment advisor developed in [88, 153], Perceptual reasoning (Chapter 8) is used to compute an output FOU from a set of rules that are activated by words. This FOU is then mapped into a

codebook word so that it can be understood. The mapping that does this imposes two requirements, one each on the CWW engine and the decoder.

First, the output of the CWW engine must resemble the word FOU in the codebook. Recall that in Section 2.3 it has been shown that there are only three kinds of FOUs in the codebook — left-shoulder, right-shoulder and interior FOUs — all of which are *normal*; consequently, the output of the CWW engine must also be a normal IT2 FS having one of these three shapes. Perceptual reasoning introduced in Chapter 8 lets us satisfy this requirement.

Second, the decoder must compare the similarity between two IT2 FSs so that the output of the CWW engine can be mapped into its most similar word in the codebook. Several similarity measures [12, 37, 90, 144, 150, 188] for IT2 FSs are discussed in Section 3.2.

2. *Rank*: In some decision-making situations several alternatives are compared so that the best one(s) can be chosen, e.g., in the procurement judgment advisor developed in [88, 152], three missile systems are compared to find the one with the best overall performance. In these applications, the outputs of the CWW Engines are always IT2 FSs; hence, the decoder must rank them to find the best alternative(s). Ranking methods [91, 144] for IT2 FSs are discussed in Section 3.3.

3. *Class*: In some decision-making applications the output of the CWW engine must be mapped into a class. In the journal publication judgment advisor developed in [87, 88], the outputs of the CWW engine are IT2 FSs representing the overall

quality of a journal article obtained from reviewers. These IT2 FSs must be mapped into one of three decision classes: accept, rewrite, and reject. How to do this is discussed in Section 3.4.

It is important to propagate linguistic uncertainties all the way through the Per-C, from its encoder, through its CWW engine, and also through its decoder; hence, our guideline for developing decoders is to *preserve and propagate the uncertainties through the decoder as far as possible.* More will be said about this later in this chapter.

## 3.2 Similarity Measure Used As a Decoder

In this section, six similarity measures for IT2 FSs are briefly introduced, and their performances as a decoder are compared. The best of these measures is suggested for use as a decoder in CWW, and is the one used in later chapters.

### 3.2.1 Definitions

*Similarity*, *proximity* and *compatibility* have all been used in the literature to assess agreement between FSs [20]. There are many different definitions for the meanings of them [20, 30, 59, 76, 124, 170, 178].

According to Yager [170], a *proximity relationship* between two T1 FSs $X_1$ and $X_2$ on a domain $D_X$ is a mapping $p$: $D_X \times D_X \rightarrow T$ (often $T$ is the unit interval) having the properties:

1. *Reflexivity:* $p(X_1, X_1) = 1$;

2. *Symmetry:* $p(X_1, X_2) = p(X_2, X_1)$.

According to Zadeh [178] and Yager [170], a *similarity relationship* between two FSs $X_1$ and $X_2$ on a domain $D_X$ is a mapping $s$: $X \times X \to T$ having the properties:

1. *Reflexivity:* $s(X_1, X_1) = 1$;

2. *Symmetry:* $s(X_1, X_2) = s(X_2, X_1)$;

3. *Transitivity:* $s(X_1, X_2) \geq s(X_1, X_3) \wedge s(X_3, X_2)$, where $X_3$ is an arbitrary FS on domain $D_X$.

Observe that a similarity relationship adds the additional requirement of transitivity to proximity, though whether or not the above definition of transitivity is correct is still under debate [18, 62].

There are other definitions of transitivity used in the literature [12, 33, 128], e.g., the one used by Bustince [12] is:

*Transitivity′:* If $X_1 \leq X_2 \leq X_3$, i.e., $\mu_{X_1}(x) \leq \mu_{X_2}(x) \leq \mu_{X_3}(x)$ $\forall x \in D_X$ (see Fig. 3.1), then $s(X_1, X_2) \geq s(X_1, X_3)$.

Bustince's transitivity is used in this dissertation.

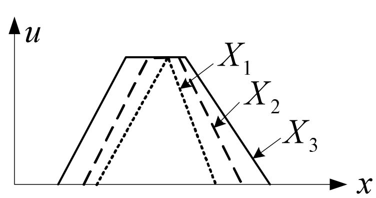



Fig. 3.1: An illustration of $X_1 \leq X_2 \leq X_3$.

*Compatibility* is a broader concept. According to Cross and Sudkamp [20], "*the term compatibility is used to encompass various types of comparisons frequently made between objects or concepts. These relationships include similarity, inclusion, proximity, and the degree of matching.*"

In summary, similarity is included in proximity, and both similarity and proximity are included in compatibility. This chapter focuses on similarity measures.

### 3.2.2 Desirable Properties for an IT2 FS Similarity Measure

Let $s(\tilde{X}_1, \tilde{X}_2)$ be the similarity measure between two IT2 FSs $\tilde{X}_1$ and $\tilde{X}_2$ in $D_{\tilde{X}}$, and $c(\tilde{X}_1)$ be the center of the centroid of $\tilde{X}_1$ (see Definition 11).

**Definition 14** *$\tilde{X}_1$ and $\tilde{X}_2$ have the same shape if $\mu_{\overline{X}_1}(x) = \mu_{\overline{X}_2}(x+\lambda)$ and $\mu_{\underline{X}_1}(x) = \mu_{\underline{X}_2}(x+\lambda)$ for $\forall x \in D_{\tilde{X}}$, where $\lambda$ is a constant.* ■

**Definition 15** *$\tilde{X}_1 \le \tilde{X}_2$ if $\mu_{\overline{X}_1}(x) \le \mu_{\overline{X}_2}(x)$ and $\mu_{\underline{X}_1}(x) \le \mu_{\underline{X}_2}(x)$ for $\forall x \in D_{\tilde{X}}$.* ■

Two examples of $\tilde{X}_1 \le \tilde{X}_2$ are shown in Fig. 3.2.

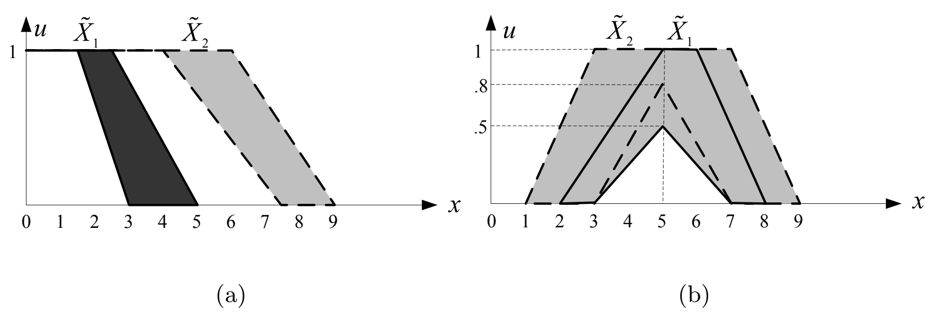



Fig. 3.2: Two examples of $\tilde{X}_1 \le \tilde{X}_2$. In both figures, $\tilde{X}_1$ is represented by the solid curves and $\tilde{X}_2$ is represented by the dashed curves.

**Definition 16** *$\tilde{X}_1$ and $\tilde{X}_2$ overlap, i.e., $\tilde{X}_1 \cap \tilde{X}_2 \neq \emptyset$, if and only if $\exists x \in D_{\tilde{X}}$ such that* $\min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x)) > 0$. ■

Two examples of overlapping $\tilde{X}_1$ and $\tilde{X}_2$ are shown in Fig. 3.2.

**Definition 17** *$\tilde{X}_1$ and $\tilde{X}_2$ do not overlap, i.e., $\tilde{X}_1 \cap \tilde{X}_2 = \emptyset$, if and only if* $\min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x)) = 0$ *for $\forall x \in D_{\tilde{X}}$.* ■

Non-overlapping $\tilde{X}_1$ and $\tilde{X}_2$ have no parts of their FOUs that overlap. In Fig. 3.7, $\tilde{X}_1$ and *Rewrite* and $\tilde{X}_1$ and *Reject* do not overlap.

Let $X_1^e$ be an embedded T1 FS of $\tilde{X}_1$. Because $\mu_{X_1^e}(x_i) \leq \mu_{\overline{X}_1}(x_i)$ for all embedded T1 FSs $X_1^e$ and $\mu_{\underline{X}_2}(x) \leq \mu_{\overline{X}_2}(x)$, $\min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x)) = 0$ means $\min(\mu_{X_1^e}(x_i), \mu_{\overline{X}_2}(x)) = 0$ and $\min(\mu_{X_1^e}(x_i), \mu_{\underline{X}_2}(x)) = 0$ for $\forall x \in D_{\tilde{X}}$ and $\forall\ X_1^e$, i.e., the following lemma follows from Definition 17:

**Lemma 3** *If $\tilde{X}_1$ and $\tilde{X}_2$ do not overlap, then* $\min(\mu_{X_1^e}(x_i), \mu_{\overline{X}_2}(x)) = 0$ *and* $\min(\mu_{X_1^e}(x_i), \mu_{\underline{X}_2}(x)) = 0$ *for $\forall x \in D_{\tilde{X}}$ and $\forall\ X_1^e$.* ■

The following four properties [150] are considered desirable for an IT2 FS similarity measure:

1. *Reflexivity:* $s(\tilde{X}_1, \tilde{X}_2) = 1 \Leftrightarrow \tilde{X}_1 = \tilde{X}_2$.

2. *Symmetry:* $s(\tilde{X}_1, \tilde{X}_2) = s(\tilde{X}_2, \tilde{X}_1)$.

3. *Transitivity:* If $\tilde{X}_1 \leq \tilde{X}_2 \leq \tilde{X}_3$, then $s(\tilde{X}_1, \tilde{X}_2) \geq s(\tilde{X}_1, \tilde{X}_3)$.

4. *Overlapping:* If $\tilde{X}_1 \cap \tilde{X}_2 \neq \emptyset$, then $s(\tilde{X}_1, \tilde{X}_2) > 0$; otherwise, $s(\tilde{X}_1, \tilde{X}_2) = 0$.

Observe that the first three properties are the IT2 FS counterparts of those used in Zadeh and Yager's definition of T1 FS similarity measures, except that a different definition of transitivity is used. The fourth property of overlapping is intuitive and is used in many T1 FS similarity measures [20], so, it is included here as a desirable property for IT2 FS similarity measures.

### 3.2.3 Problems with Existing IT2 FS Similarity Measures

Though "*there are approximately 50 expressions for determining how similar two (T1) fuzzy sets are*" [13], to the best knowledge of the authors, there are only six similarity (compatibility) measures for IT2 FSs [12, 37, 90, 144, 150, 188]. The drawbacks of five of them are pointed out in this subsection (an example that demonstrates each of the drawbacks can be found in [150]), and the sixth similarity measure (Jaccard similarity measure [144]) is introduced in the next subsection.

1. Gorzalczany [37] defined an interval compatibility measure for IT2 FSs; however, it is not a good similarity measure for our purpose because [150] as long as $\max_{x \in D_{\tilde{X}}} \mu_{\underline{X}_1}(x) = \max_{x \in D_{\tilde{X}}} \mu_{\underline{X}_2}(x)$ and $\max_{x \in D_{\tilde{X}}} \mu_{\overline{X}_1}(x) = \max_{x \in D_{\tilde{X}}} \mu_{\overline{X}_2}(x)$ (both of which can be easily satisfied by $\tilde{X}_1$ and $\tilde{X}_2$, even when $\tilde{X}_1 \neq \tilde{X}_2$), no matter how different the shapes of $\tilde{X}_1$ and $\tilde{X}_2$ are, it always gives $s_G(\tilde{X}_1, \tilde{X}_2) = s_G(\tilde{X}_2, \tilde{X}_1) = [1, 1]$, i.e., it does not satisfy Reflexivity.

2. Bustince [12] defined an interval similarity measure for IT2 FSs $\tilde{X}_1$ and $\tilde{X}_2$ based on the inclusion of $\tilde{X}_1$ in $\tilde{X}_2$. A problem with this approach is that [150] when $\tilde{X}_1$ and $\tilde{X}_2$ are disjoint, no matter how far away they are from each other, $s_{X_2}(\tilde{X}_1, \tilde{X}_2)$ will always be a nonzero constant, i.e., it does not satisfy Overlapping.

3. Mitchell [90] defined the similarity between two IT2 FSs as the average of the similarities between their embedded T1 FSs, when the embedded T1 FSs are generated randomly. Consequently, this similarity measure does not satisfy Reflexivity, i.e., $s_M(\tilde{X}_1, \tilde{X}_2) \neq 1$ when $\tilde{X}_1 = \tilde{X}_2$ because the randomly generated embedded T1 FSs from $\tilde{X}_1$ and $\tilde{X}_2$ vary from experiment to experiment [150].

4. Zeng and Li [188] defined the similarity between $\tilde{X}_1$ and $\tilde{X}_2$ based on the difference between them. A problem with this approach is that when $\tilde{X}_1$ and $\tilde{X}_2$ are disjoint, the similarity is a nonzero constant, or increases as the distance increases, i.e., it does not satisfy Overlapping.

5. Wu and Mendel [150] proposed a vector similarity measure, which considers the similarity between the shape and proximity of two IT2 FSs separately. It does not satisfy Overlapping [144].

### 3.2.4 Jaccard Similarity Measure for IT2 FSs

The Jaccard similarity measure for T1 FSs [48] is defined as

$$s_J(X_1, X_2) = \frac{f(X_1 \cap X_2)}{f(X_1 \cup X_2)} \tag{3.1}$$

where $f$ is a function satisfying $f(X_1 \cup X_2) = f(X_1) + f(X_2)$ for disjoint $X_1$ and $X_2$. Usually the function $f$ is chosen as the cardinality [see (2.25)], i.e., when $\cap \equiv \min$ and $\cup \equiv \max$,

$$s_J(X_1, X_2) \equiv \frac{\text{card}(X_1 \cap X_2)}{\text{card}(X_1 \cup X_2)} = \frac{\int_X \min(\mu_{X_1}(x), \mu_{X_2}(x))dx}{\int_X \max(\mu_{X_1}(x), \mu_{X_2}(x))dx}. \tag{3.2}$$

whose discrete version is

$$s_J(X_1, X_2) = \frac{\sum_{i=1}^N \min(\mu_{X_1}(x_i), \mu_{X_2}(x_i))}{\sum_{i=1}^N \max(\mu_{X_1}(x_i), \mu_{X_2}(x_i))} \tag{3.3}$$

where $x_i$ $(i = 1, \ldots, N)$ are equally spaced in $D_{\tilde{X}}$.

[144] proposes a new similarity measure for IT2 FSs, which is an extension of (3.2), and uses average cardinality, $AC$, as defined in (2.26), applied to both $\tilde{X}_1 \cap \tilde{X}_2$ and $\tilde{X}_1 \cup \tilde{X}_2$, where $\tilde{X}_1 \cap \tilde{X}_2$ and $\tilde{X}_1 \cup \tilde{X}_2$ are computed by (2.29) and (2.28), respectively, i.e.,

$$s_J(\tilde{X}_1, \tilde{X}_2) \equiv \frac{AC(\tilde{X}_1 \cap \tilde{X}_2)}{AC(\tilde{X}_1 \cup \tilde{X}_2)} = \frac{\int_X \min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))dx + \int_X \min(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))dx}{\int_X \max(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))dx + \int_X \max(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))dx}. \tag{3.4}$$

Note that each integral in (3.4) is an area, e.g., $\int_X \min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))dx$ is the area under the minimum of $\mu_{\overline{X}_1}(x)$ and $\mu_{\overline{X}_2}(x)$. Closed-form solutions cannot always be found for these integrals, so, the following discrete version of (3.4) is used in calculations:

$$s_J(\tilde{X}_1, \tilde{X}_2) = \frac{\sum_{i=1}^N \min(\mu_{\overline{X}_1}(x_i), \mu_{\overline{X}_2}(x_i)) + \sum_{i=1}^N \min(\mu_{\underline{X}_1}(x_i), \mu_{\underline{X}_2}(x_i))}{\sum_{i=1}^N \max(\mu_{\overline{X}_1}(x_i), \mu_{\overline{X}_2}(x_i)) + \sum_{i=1}^N \max(\mu_{\underline{X}_1}(x_i), \mu_{\underline{X}_2}(x_i))}. \tag{3.5}$$

**Theorem 4** *The Jaccard similarity measure, $s_J(\tilde{X}_1, \tilde{X}_2)$, satisfies reflexivity, symmetry, transitivity and overlapping.* ■

*Proof:* Our proof of Theorem 4 is for the continuous case (3.4). The proof for the discrete case (3.5) is very similar, and is left to the reader.

1. *Reflexivity:* Consider first the necessity, i.e., $s_J(\tilde{X}_1, \tilde{X}_2) = 1 \Rightarrow \tilde{X}_1 = \tilde{X}_2$. When the areas of the FOUs are not zero, $\min(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x)) < \max(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))$; hence, the only way that $s_J(\tilde{X}_1, \tilde{X}_2) = 1$ [see (3.4)] is when $\min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x)) = \max(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))$ and $\min(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x)) = \max(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))$, in which case $\mu_{\overline{X}_1}(x) = \mu_{\overline{X}_2}(x)$ and $\mu_{\underline{X}_1}(x) = \mu_{\underline{X}_2}(x)$, i.e., $\tilde{X}_1 = \tilde{X}_2$.

   Consider next the sufficiency, i.e., $\tilde{X}_1 = \tilde{X}_2 \Rightarrow s_J(\tilde{X}_1, \tilde{X}_2) = 1$. When $\tilde{X}_1 = \tilde{X}_2$, i.e., $\mu_{\overline{X}_1}(x) = \mu_{\overline{X}_2}(x)$ and $\mu_{\underline{X}_1}(x) = \mu_{\underline{X}_2}(x)$, it follows that $\min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x)) = \max(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))$ and $\min(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x)) = \max(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))$. Consequently, it follows from (3.4) that $s_J(\tilde{X}_1, \tilde{X}_2) = 1$.

2. *Symmetry:* Observe from (3.4) that $s_J(\tilde{X}_1, \tilde{X}_2)$ does not depend on the order of $\tilde{X}_1$ and $\tilde{X}_2$; so, $s_J(\tilde{X}_1, \tilde{X}_2) = s_J(\tilde{X}_2, \tilde{X}_1)$.

3. *Transitivity:* If $\tilde{X}_1 \leq \tilde{X}_2 \leq \tilde{X}_3$ (see Definition 15), then

$$\begin{aligned} s_J(\tilde{X}_1, \tilde{X}_2) &= \frac{\int_X \min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))dx + \int_X \min(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))dx}{\int_X \max(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))dx + \int_X \max(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))dx} \\ &= \frac{\int_X \mu_{\overline{X}_1}(x)dx + \int_X \mu_{\underline{X}_1}(x)dx}{\int_X \mu_{\overline{X}_2}(x)dx + \int_X \mu_{\underline{X}_2}(x)dx} \\ s_J(\tilde{X}_1, \tilde{X}_3) &= \frac{\int_X \min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_3}(x))dx + \int_X \min(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_3}(x))dx}{\int_X \max(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_3}(x))dx + \int_X \max(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_3}(x))dx} \end{aligned} \quad (3.6)$$

$$= \frac{\int_X \mu_{\overline{X}_1}(x)dx + \int_X \mu_{\underline{X}_1}(x)dx}{\int_X \mu_{\overline{X}_3}(x)dx + \int_X \mu_{\underline{X}_3}(x)dx} \tag{3.7}$$

Because $\tilde{X}_2 \le \tilde{X}_3$, it follows that $\int_X \mu_{\overline{X}_2}(x)dx + \int_X \mu_{\underline{X}_2}(x)dx \le \int_X \mu_{\overline{X}_3}(x)dx + \int_X \mu_{\underline{X}_3}(x)dx$, and hence $s_J(\tilde{X}_1, \tilde{X}_2) \ge s_J(\tilde{X}_1, \tilde{X}_3)$.

4. *Overlapping:* If $\tilde{X}_1 \cap \tilde{X}_2 \ne \emptyset$ (see Definition 16), $\exists x$ such that $\min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x)) > 0$, then, in the numerator of (3.4),

$$\int_X \min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))dx + \int_X \min(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))dx > 0 \tag{3.8}$$

In the denominator of (3.4),

$$\begin{aligned} &\int_X \max(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))dx + \int_X \max(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))dx \\ \ge &\int_X \min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))dx + \int_X \min(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))dx > 0 \end{aligned} \tag{3.9}$$

Consequently, $s_J(\tilde{X}_1, \tilde{X}_2) > 0$. On the other hand, when $\tilde{X}_1 \cap \tilde{X}_2 = \emptyset$, i.e., $\min(\mu_{\overline{X}_1}(x),$
$\mu_{\overline{X}_2}(x)) = \min(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x)) = 0$ for $\forall x$, then, in the numerator of (3.4),

$$\int_X \min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))dx + \int_X \min(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))dx = 0 \tag{3.10}$$

Consequently, $s_J(\tilde{X}_1, \tilde{X}_2) = 0$. ■

### 3.2.5 Simulation Results

The 32 word FOUs shown in Fig. 2.12 are used in this section. The similarities among all 32 words, computed using the Jaccard similarity measure in (3.4), are summarized in Table 3.1. The numbers across the top of this table refer to the numbered words that are in the first column of the table. Observe that the Jaccard similarity measure gives very reasonable results, i.e., generally the similarity decreases monotonically as two words get further away from each other[1]. The Jaccard similarity measure was also compared with five other similarity measures in [150], and the results showed that to-date it is the best one to use in CWW, because it is the only IT2 FS similarity measure that satisfies the four desirable properties of a similarity measure.

The fact that so many of the 32 words are similar to many other words suggest that it is possible to create many sub-vocabularies that cover the interval $[0, 10]$. Some examples of five word vocabularies are given in [70].

[1] There are some cases where the similarity does not decrease monotonically, e.g., Words 8 and 9 in the first row. This is because the distances among the words are determined by a ranking method (see Section 3.3) which considers only the centroids but not the shapes of the IT2 FSs.

Table 3.1: Similarity matrix for the 32 words when the Jaccard similarity measure is used.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. None to very little | 1 | .54 | .51 | .49 | .48 | .47 | .09 | .08 | .08 | .07 | .04 | .04 | .02 | .01 | .01 | 0 | 0 | .01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2. Teeny-weeny | .54 | 1 | .57 | .54 | .44 | .44 | .08 | .08 | .08 | .07 | .04 | .03 | .02 | .01 | .01 | 0 | 0 | .01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3. A smidgen | .51 | .57 | 1 | .96 | .76 | .78 | .15 | .13 | .12 | .10 | .07 | .05 | .03 | .01 | .01 | 0 | 0 | .01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4. Tiny | .49 | .54 | .96 | 1 | .79 | .81 | .15 | .14 | .12 | .10 | .07 | .05 | .03 | .01 | .01 | 0 | 0 | .01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5. Very small | .48 | .44 | .76 | .79 | 1 | .91 | .17 | .14 | .12 | .11 | .07 | .05 | .03 | .01 | .02 | 0 | 0 | .01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6. Very little | .47 | .44 | .78 | .81 | .91 | 1 | .18 | .15 | .13 | .12 | .08 | .06 | .03 | .02 | .02 | 0 | 0 | .01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7. A bit | .09 | .08 | .15 | .15 | .17 | .18 | 1 | .43 | .35 | .32 | .25 | .11 | .07 | .04 | .04 | .01 | .01 | .02 | .01 | .01 | .01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8. Little | .08 | .08 | .13 | .14 | .14 | .15 | .43 | 1 | .77 | .66 | .50 | .21 | .13 | .08 | .08 | .04 | .04 | .04 | .01 | .01 | .01 | .01 | .01 | .01 | .01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9. Low amount | .08 | .08 | .12 | .12 | .12 | .13 | .35 | .77 | 1 | .80 | .55 | .23 | .15 | .10 | .09 | .05 | .05 | .04 | .02 | .02 | .02 | .01 | .01 | .01 | .01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10. Small | .07 | .07 | .10 | .10 | .11 | .12 | .32 | .66 | .80 | 1 | .64 | .25 | .18 | .11 | .11 | .05 | .05 | .05 | .02 | .02 | .02 | .01 | .01 | .01 | .01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11. Somewhat small | .04 | .04 | .07 | .07 | .07 | .08 | .25 | .50 | .55 | .64 | 1 | .24 | .18 | .11 | .11 | .05 | .05 | .05 | .02 | .02 | .02 | .01 | .01 | .01 | .01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12. Some | .04 | .03 | .05 | .05 | .05 | .06 | .11 | .21 | .23 | .25 | .24 | 1 | .58 | .37 | .36 | .20 | .23 | .20 | .11 | .11 | .11 | .06 | .06 | .06 | .06 | .04 | .02 | .01 | .02 | .01 | .01 | .01 |
| 13. Some to moderate | .02 | .02 | .03 | .03 | .03 | .03 | .07 | .13 | .15 | .18 | .18 | .58 | 1 | .57 | .60 | .31 | .34 | .29 | .16 | .16 | .16 | .09 | .09 | .08 | .08 | .06 | .02 | .02 | .02 | .02 | .02 | .01 |
| 14. Moderate amount | .01 | .01 | .01 | .01 | .01 | .02 | .04 | .08 | .10 | .11 | .11 | .37 | .57 | 1 | .72 | .50 | .54 | .29 | .16 | .16 | .15 | .08 | .08 | .07 | .07 | .05 | .01 | .01 | .01 | .01 | .01 | 0 |
| 15. Fair amount | .01 | .01 | .01 | .01 | .02 | .02 | .04 | .08 | .09 | .11 | .11 | .36 | .60 | .72 | 1 | .50 | .53 | .36 | .21 | .21 | .20 | .11 | .11 | .10 | .10 | .07 | .02 | .02 | .02 | .02 | .02 | .01 |
| 16. Medium | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .04 | .05 | .05 | .05 | .20 | .31 | .50 | .50 | 1 | .61 | .20 | .12 | .12 | .11 | .06 | .06 | .05 | .05 | .03 | .01 | .01 | .01 | .01 | .01 | 0 |
| 17. Modest amount | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .04 | .05 | .05 | .05 | .23 | .34 | .54 | .53 | .61 | 1 | .30 | .18 | .18 | .16 | .09 | .09 | .08 | .08 | .05 | .01 | .01 | .01 | .01 | .01 | 0 |
| 18. Good amount | .01 | .01 | .01 | .01 | .01 | .01 | .02 | .04 | .04 | .05 | .05 | .20 | .29 | .29 | .36 | .20 | .30 | 1 | .50 | .50 | .50 | .27 | .27 | .25 | .25 | .18 | .07 | .05 | .06 | .05 | .05 | .02 |
| 19. Sizeable | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .01 | .02 | .02 | .02 | .11 | .16 | .16 | .21 | .12 | .18 | .50 | 1 | 1 | .84 | .47 | .47 | .43 | .42 | .32 | .09 | .07 | .08 | .08 | .07 | .03 |
| 20. Quite a bit | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .01 | .02 | .02 | .02 | .11 | .16 | .16 | .21 | .12 | .18 | .50 | 1 | 1 | .84 | .47 | .47 | .43 | .42 | .32 | .09 | .07 | .08 | .08 | .07 | .03 |
| 21. Considerable amount | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .01 | .02 | .02 | .02 | .11 | .16 | .15 | .20 | .11 | .16 | .50 | .84 | .84 | 1 | .49 | .49 | .44 | .45 | .32 | .09 | .08 | .08 | .08 | .08 | .03 |
| 22. Substantial amount | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .01 | .01 | .01 | .06 | .09 | .08 | .11 | .06 | .09 | .27 | .47 | .47 | .49 | 1 | .98 | .82 | .79 | .63 | .15 | .13 | .14 | .14 | .13 | .05 |
| 23. A lot | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .01 | .01 | .01 | .06 | .09 | .08 | .11 | .06 | .09 | .27 | .47 | .47 | .49 | .98 | 1 | .83 | .79 | .63 | .15 | .13 | .14 | .13 | .13 | .05 |
| 24. High amount | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .01 | .01 | .01 | .06 | .08 | .07 | .10 | .05 | .08 | .25 | .43 | .43 | .44 | .82 | .83 | 1 | .89 | .70 | .17 | .14 | .16 | .15 | .14 | .06 |
| 25. Very sizeable | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .01 | .01 | .01 | .06 | .08 | .07 | .10 | .05 | .08 | .25 | .42 | .42 | .45 | .79 | .79 | .89 | 1 | .64 | .15 | .14 | .14 | .13 | .13 | .05 |
| 26. Large | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .04 | .06 | .05 | .07 | .03 | .05 | .18 | .32 | .32 | .32 | .63 | .63 | .70 | .64 | 1 | .17 | .15 | .16 | .15 | .15 | .05 |
| 27. Very large | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .02 | .02 | .01 | .02 | .01 | .01 | .07 | .09 | .09 | .09 | .15 | .15 | .17 | .15 | .17 | 1 | .67 | .86 | .70 | .68 | .21 |
| 28. Humongous amount | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .02 | .01 | .02 | .01 | .01 | .05 | .07 | .07 | .08 | .13 | .13 | .14 | .14 | .15 | .67 | 1 | .66 | .68 | .68 | .22 |
| 29. Huge amount | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .02 | .02 | .01 | .02 | .01 | .01 | .06 | .08 | .08 | .08 | .14 | .14 | .16 | .14 | .16 | .86 | .66 | 1 | .83 | .80 | .25 |
| 30. Very high amount | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .02 | .01 | .02 | .01 | .01 | .05 | .08 | .08 | .08 | .14 | .13 | .15 | .13 | .15 | .70 | .68 | .83 | 1 | .96 | .25 |
| 31. Extreme amount | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .02 | .01 | .02 | .01 | .01 | .05 | .07 | .07 | .08 | .13 | .13 | .14 | .13 | .15 | .68 | .68 | .80 | .96 | 1 | .26 |
| 32. Maximum amount | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | .01 | .01 | 0 | .01 | 0 | 0 | .02 | .03 | .03 | .03 | .05 | .05 | .06 | .05 | .05 | .21 | .22 | .25 | .25 | .26 | 1 |

## 3.3 Ranking Method Used As a Decoder

Though there are more than 35 reported different methods for ranking T1 FSs [142,143], to the best knowledge of the authors, only one method on ranking IT2 FSs has been published, namely Mitchell's method in [91]. We will first introduce some reasonable ordering properties for IT2 FSs, and then compare Mitchell's method against them. A new ranking method for IT2 FSs is proposed at the end of this section.

### 3.3.1 Reasonable Ordering Properties for IT2 FSs

Wang and Kerre [142,143] performed a comprehensive study of T1 FSs ranking methods based on seven reasonable ordering properties for T1 FSs. When extended to IT2 FSs, these properties are[2]:

P1. If $\tilde{X}_1 \succeq \tilde{X}_2$ and $\tilde{X}_2 \succeq \tilde{X}_1$, then $\tilde{X}_1 \sim \tilde{X}_2$.

P2. If $\tilde{X}_1 \succeq \tilde{X}_2$ and $\tilde{X}_2 \succeq \tilde{X}_3$, then $\tilde{X}_1 \succeq \tilde{X}_3$.

P3. If $\tilde{X}_1 \cap \tilde{X}_2 = \emptyset$ and $\tilde{X}_1$ is on the right of $\tilde{X}_2$, then $\tilde{X}_1 \succeq \tilde{X}_2$.

P4. The order of $\tilde{X}_1$ and $\tilde{X}_2$ is not affected by the other IT2 FSs under comparison.

P5. If $\tilde{X}_1 \succeq \tilde{X}_2$, then[3] $\tilde{X}_1 + \tilde{X}_3 \succeq \tilde{X}_2 + \tilde{X}_3$.

[2]There is another property saying that "for any IT2 FS $\tilde{X}_1$, $\tilde{X}_1 \succeq \tilde{X}_1$;" however, it is not included here since it sounds weird, though our centroid-based ranking method satisfies it.

[3]$\tilde{X}_1 + \tilde{X}_3$ is computed using $\alpha$-cuts [64] and Extension Principle [179], i.e., let $\tilde{X}_1^\alpha$, $\tilde{X}_3^\alpha$ and $(\tilde{X}_1 + \tilde{X}_3)^\alpha$ be $\alpha$-cuts on $\tilde{X}_1$, $\tilde{X}_3$ and $\tilde{X}_1 + \tilde{X}_3$, respectively; then, $(\tilde{X}_1 + \tilde{X}_3)^\alpha = \tilde{X}_1^\alpha + \tilde{X}_3^\alpha$ for $\forall \alpha \in [0,1]$.

P6. If $\tilde{X}_1 \succeq \tilde{X}_2$, then[4] $\tilde{X}_1\tilde{X}_3 \succeq \tilde{X}_2\tilde{X}_3$.

where $\succeq$ means "*larger than or equal to in the sense of ranking*" and $\sim$ means "*the same rank.*"

All the six properties are intuitive. P4 may look trivial, but it is worth emphasizing because some ranking methods [142, 143] first set up reference set(s) and then all FSs are compared with the reference set(s). The reference set(s) may depend on the FSs under consideration, so it is possible (but not desirable) that $\tilde{X}_1 \succeq \tilde{X}_2$ when $\{\tilde{X}_1, \tilde{X}_2, \tilde{X}_3\}$ are ranked whereas $\tilde{X}_1 \prec \tilde{X}_2$ when $\{\tilde{X}_1, \tilde{X}_2, \tilde{X}_4\}$ are ranked.

### 3.3.2 Mitchell's Method for Ranking IT2 FSs

Mitchell [91] proposed a ranking method for general type-2 FSs. When specialized to $M$ IT2 FSs $\tilde{X}_m$ ($m = 1, \ldots, M$), the procedure is:

1. Discretize the primary variable's universe of discourse, $D_{\tilde{X}}$, into $N$ points, that are used by all $\tilde{X}_m$, $m = 1, \ldots, M$.

2. Find $H$ random embedded T1 FSs[5], $X_e^{mh}$, $h = 1, \ldots, H$, for each of the $M$ IT2 FSs $\tilde{X}_m$, as:

$$\mu_{X_e^{mh}}(x_n) = r_{mh}(x_n) \times [\mu_{\overline{X}_m}(x_n) - \mu_{\underline{X}_m}(x_n)] + \mu_{\underline{X}_m}(x_n) \qquad n = 1, 2, \ldots, N \quad (3.11)$$

[4] $\tilde{X}_1\tilde{X}_3$ is computed using $\alpha$-cuts [64] and Extension Principle [179], i.e., let $\tilde{X}_1^\alpha$, $\tilde{X}_3^\alpha$ and $(\tilde{X}_1\tilde{X}_3)^\alpha$ be $\alpha$-cuts on $\tilde{X}_1$, $\tilde{X}_3$ and $\tilde{X}_1\tilde{X}_3$, respectively; then, $(\tilde{X}_1\tilde{X}_3)^\alpha = \tilde{X}_1^\alpha\tilde{X}_3^\alpha$ for $\forall\alpha \in [0, 1]$.

[5] Visually, an embedded T1 FS of an IT2 FS is a T1 FS whose membership function lies within the FOU of the IT2 FS. A more precise mathematical definition can be found in [83].

where $r_{mh}(x_n)$ is a random number chosen uniformly in $[0,1]$, and $\mu_{\underline{X}_m}(x_n)$ and $\mu_{\overline{X}_m}(x_n)$ are the lower and upper memberships of $\tilde{X}_m$ at $x_n$.

3. Form the $H^M$ different combinations of $\{X_e^{1h}, X_e^{2h}, \ldots, X_e^{Mh}\}_i$, $i = 1, \ldots, H^M$.

4. Use a T1 FS ranking method to rank each of the $M^H$ $\{X_e^{1h}, X_e^{2h}, \ldots, X_e^{Mh}\}_i$. Denote the rank of $X_{mh}^e$ in $\{X_e^{1h}, X_e^{2h}, \ldots, X_e^{Mh}\}_i$ as $r_{mi}$.

5. Compute the final rank of $\tilde{X}_m$ as

$$r_m = \frac{1}{H^M} \sum_{i=1}^{H^M} r_{mi}, \quad m = 1, \ldots, M \tag{3.12}$$

Observe from the above procedure that:

1. The output ranking, $r_m$, is a crisp number; however, usually it is not an integer. These $r_m$ $(m = 1, \ldots, M)$ need to be sorted in order to find the correct ranking.

2. A total of $H^M$ T1 FS rankings must be evaluated before $r_m$ can be computed. For our problem, where 32 IT2 FSs have to be ranked, even if $H$ is chosen as a small number, say 2, $2^{32} \approx 4.295 \times 10^9$ T1 FS rankings have to be evaluated, and each evaluation involves 32 T1 FSs. This is highly impractical. Although two fast algorithms are proposed in [91], because our FOUs have lots of overlap, the computational cost cannot be reduced significantly. Note also that choosing the number of realizations $H$ as 2 is not meaningful; it should be much larger, and for larger $H$, the number of rankings becomes astronomical.

3. Because there are random numbers involved, $r_m$ is random and will change from experiment to experiment. When $H$ is large, some kind of stochastic convergence can be expected to occur for $r_m$ (e.g., convergence in probability); however, as mentioned above, the computational cost is prohibitive.

4. Because of the random nature of Mitchell's ranking method, it only satisfies P3 of the six reasonable properties proposed in Section 3.3.1.

### 3.3.3 A New Centroid-Based Ranking Method

A simple ranking method [144] based on the centroids of IT2 FSs is proposed in this subsection.

**Centroid-based ranking method**: [144] First compute the average centroid for each IT2 FS using (2.26) and then sort $c(\tilde{X}_i)$ to obtain the rank of $\tilde{X}_i$. ■

This ranking method can be viewed as a generalization of Yager's first ranking method for T1 FSs [164], which first computes the centroid of T1 FSs $X_i$ and then ranks them.

**Theorem 5** *The centroid-based ranking method satisfies the first four reasonable properties.* ■

*Proof:* P1-P4 in Section 3.3.1 are proved in order.

P1. $\tilde{X}_1 \succeq \tilde{X}_2$ means $c(\tilde{X}_1) \geq c(\tilde{X}_2)$ and $\tilde{X}_2 \succeq \tilde{X}_1$ means $c(\tilde{X}_2) \geq c(\tilde{X}_1)$, and hence $c(\tilde{X}_1) = c(\tilde{X}_2)$, i.e., $\tilde{X}_1 \sim \tilde{X}_2$.

P2. For the centroid-based ranking method, $\tilde{X}_1 \succeq \tilde{X}_2$ means $c(\tilde{X}_1) \geq c(\tilde{X}_2)$ and $\tilde{X}_2 \succeq \tilde{X}_3$ means $c(\tilde{X}_2) \geq c(\tilde{X}_3)$, and hence $c(\tilde{X}_1) \geq c(\tilde{X}_3)$, i.e., $\tilde{X}_1 \succeq \tilde{X}_3$.

P3. If $\tilde{X}_1 \cap \tilde{X}_2 = \emptyset$ and $\tilde{X}_1$ is on the right of $\tilde{X}_2$, then $c(\tilde{X}_1) > c(\tilde{X}_2)$, i.e., $\tilde{X}_1 \succeq \tilde{X}_2$.

P4. Because the order of $\tilde{X}_1$ and $\tilde{X}_2$ is completely determined by $c(\tilde{X}_1)$ and $c(\tilde{X}_2)$, which have nothing to do with the other IT2 FSs under comparison, the order of $\tilde{X}_1$ and $\tilde{X}_2$ is not affected by the other IT2 FSs. ■

The centroid-based ranking method does not always satisfy P5 and P6. A counter-example of P5 for $\tilde{X}_1$ and $\tilde{X}_2$ in Fig. 3.3(a) and $\tilde{X}_3'$ in Fig. 3.3(b) is shown in Fig. 3.3(c). In Fig. 3.3(a), $\overline{X}_1 = [0.05, 0.55, 2.55, 3.05]$, $\underline{X}_1 = [1.05, 1.55, 1.55, 2.05, 0.6]$, $\overline{X}_2 = [0, 1, 2, 3]$ and $\underline{X}_2 = [0.5, 1, 2, 2.5, 0.6]$. Because $c(\tilde{X}_1) = 1.55$ and $c(\tilde{X}_2) = 1.50$, $\tilde{X}_1 \succeq \tilde{X}_2$. In Fig. 3.3(b), $\overline{X}_3' = [0, 5.5, 6.5, 7]$, $\underline{X}_3' = [6, 6.5, 6.5, 7, 0.6]$, $\overline{X}_3'' = [0, 1.5, 2, 3]$ and $\underline{X}_3'' = [0.5, 1.5, 2, 2.5, 0.6]$. In Fig. 3.3(c), $c(\tilde{X}_1') = 6.53$ and $c(\tilde{X}_2') = 6.72$ and hence $\tilde{X}_1' \preceq \tilde{X}_2'$. A counter-example of P6 for $\tilde{X}_1$ and $\tilde{X}_2$ in Fig. 3.3(a) and $\tilde{X}_3''$ in Fig. 3.3(b) is shown in Fig. 3.3(d), where $c(\tilde{X}_1'') = 3.44$ and $c(\tilde{X}_2'') = 3.47$ and hence $\tilde{X}_1'' \preceq \tilde{X}_2''$. However, note that these counter examples happen only when $c(\tilde{X}_1)$ and $c(\tilde{X}_2)$ are very close to each other. For most cases, P5 and P6 are still satisfied.

In summary, the centroid-based ranking method satisfies three more of the reasonable ordering properties than Mitchell's method.

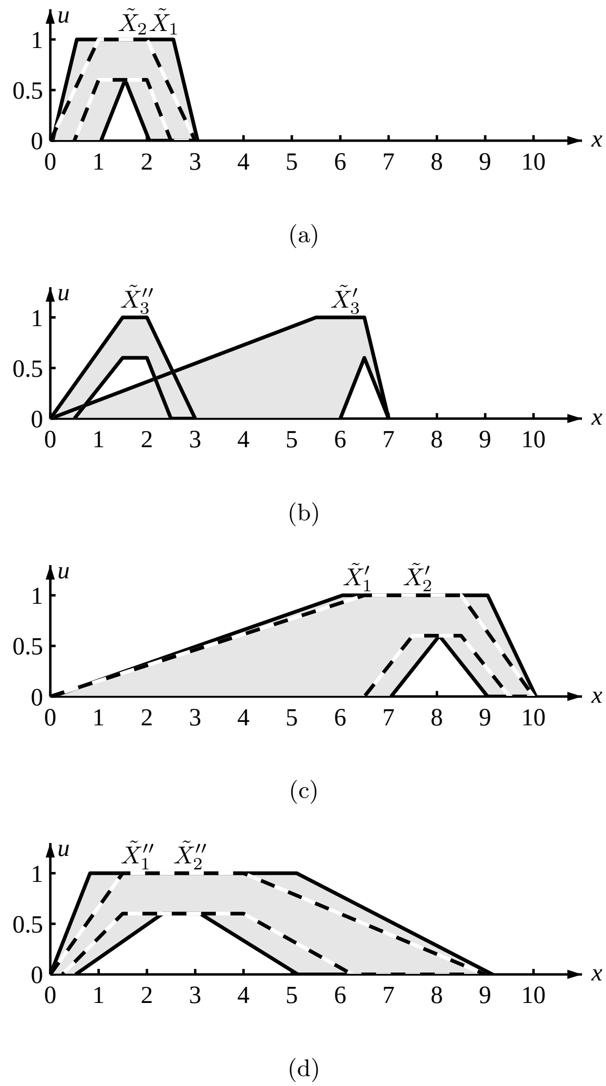



Fig. 3.3: Counter examples for P5 and P6. (a) $\tilde{X}_1$ (the solid curve) $\succeq \tilde{X}_2$ (the dashed curve). (b) $\tilde{X}_3'$ used in demonstrating P5 and $\tilde{X}_3''$ used in demonstrating P6. (c) $\tilde{X}_1' \preceq \tilde{X}_2'$, where $\tilde{X}_1' = \tilde{X}_1 + \tilde{X}_3'$ is the solid curve and $\tilde{X}_2' = \tilde{X}_2 + \tilde{X}_3'$ is the dashed curve. (d) $\tilde{X}_1'' \preceq \tilde{X}_2''$, where $\tilde{X}_1'' = \tilde{X}_1\tilde{X}_3''$ is the solid curve and $\tilde{X}_2'' = \tilde{X}_2\tilde{X}_3''$ is the dashed curve.

### 3.3.4 Comparative Study

In this section, the performances of the two IT2 FS ranking methods are compared using the 32 word FOUs.

The ranking of the 32 word FOUs using this centroid-based method has already been presented in Fig. 2.12. Observe that:

1. The six smallest terms are left shoulders, the six largest terms are right shoulders, and the terms in-between have interior FOUs.

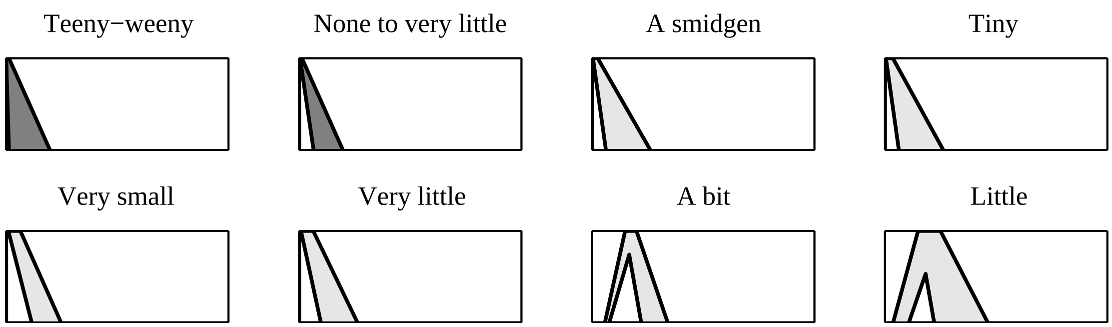
2. Visual examination shows that the ranking is reasonable; it also coincides with the meanings of the words.

Because it is computationally prohibitive to rank all 32 words in one pass using Mitchell's method, only the first eight words in Fig. 2.12 were used to evaluate Mitchell's method. To be consistent, the T1 FS ranking method used in Mitchell's method is a special case of the centroid-based ranking method for IT2 FSs, i.e., the centroids of the T1 FSs were computed and then were used to rank the corresponding T1 FSs. Ranking results with $H = 2$ and $H = 3$ are shown in Fig. 3.4(a) and Fig. 3.4(b), respectively. Words which have a different rank than that in Fig. 2.12 are shaded more darkly. Observe that:

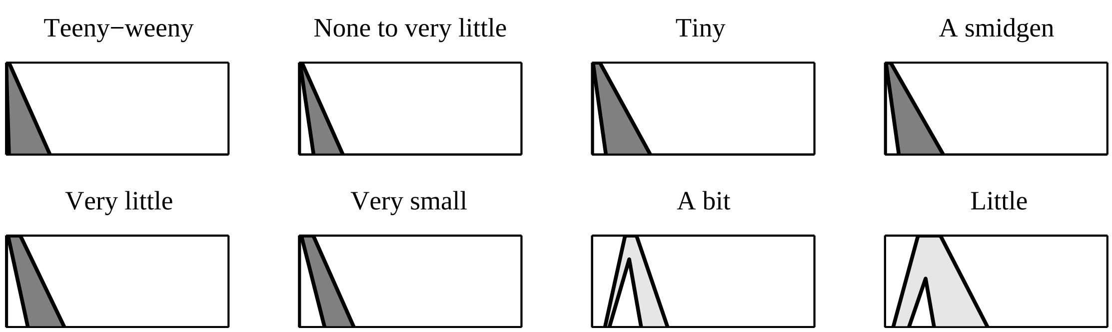
1. The ranking is different from that obtained from the centroid-based ranking method.

2. The rankings from $H = 2$ and $H = 3$ do not agree.

In summary, the centroid-based ranking method for IT2 FSs seems to be a good choice for the decoder in CWW decoder; however, note that it violates the guideline proposed at the end of Section 3.1, i.e., it first converts each FOU to a crisp number and then ranks them. To-date, an IT2 FS ranking method that can propagate FOU uncertainties does not exist; hence, the centroid-based ranking method is used in this dissertation.




(a)




(b)

Fig. 3.4: Ranking of the first eight word FOUs using Mitchell's method. (a) $H = 2$; (b) $H = 3$.

## 3.4 Classifier Used As a Decoder

Let $\tilde{X}_1$ be the output of the CWW engine. An average subsethood based classifier can be used to map $\tilde{X}_1$ into a class, according to the following procedure:

1. Construct class-FOUs, i.e., find an IT2 FS to represent each class.

2. Compute the average subsethood of $\tilde{X}_1$ in each class.

3. Map $\tilde{X}_1$ into the class with the maximum average subsethood.

How to construct class-FOUs and how to compute average subsethood are explained next.

### 3.4.1 Construct Class-FOUs

To construct class-FOUs, a decoding vocabulary must first be established, one that consists of the class names. Then, there are two ways to obtain FOUs for this vocabulary:

1. *Construct class-FOUs from a survey*: The Interval Approach introduced in Section 2.3 can be used to map the interval survey data into IT2 FSs.

2. *Construct class-FOUs from training*: A training pair is {*CWW engine output* $\tilde{X}_i$, *corresponding class* $C_i$}. Assume that $N_T$ such training pairs are available, and for an arbitrary set of class-FOUs, the outputs of the average subsethood based classifier are $C_i'$, $i = 1, ..., N_T$. Genetic algorithms [35] can be used to optimize the parameters of the class-FOUs so that the number of mismatches between $C_i$ and $C_i'$ is minimized. Assume there are $M$ classes. Because each IT2 FS word model

is defined by nine parameters (see Fig. 2.13), a total of $9M$ parameters need to be found during training.

**Example 6** *Here the journal publication judgment advisor developed in [87, 88] is used as an example to illustrate the two methods.* $\tilde{X}_1$ *is the overall quality of a paper, and it must be mapped into one of the three recommendation classes:* accept, rewrite *and* reject.

1. Construct class-FOUs from a survey*: Associate Editors and reviewers can be surveyed to provide data intervals (on a 0–10 scale) for the three classes, after which FOUs can be obtained from them.*

2. Construct class-FOUs from training*: Each paper in the training dataset would have the following pair associated with it:* {overall quality $\tilde{X}_i$, publication recommendation}*. The class-FOUs for* accept, rewrite *and* reject *can be found from the training examples so that the number of misclassified* $\tilde{X}_i$ *by the journal publication judgment advisor is minimized.* ■

### 3.4.2 Average Subsethood of IT2 FSs

Subsethood of FSs was first introduced by Zadeh [177] and then extended by Kosko [66], who defined the subsethood of a T1 FS $X_1$ in another T1 FS $X_2$ as

$$ss(X_1, X_2) = \frac{\sum_{i=1}^{N} \min(\mu_{X_1}(x_i), \mu_{X_2}(x_i))}{\sum_{i=1}^{N} \mu_{X_1}(x_i)} \tag{3.13}$$

Observe that $ss(X_1, X_2) \neq ss(X_2, X_1)$, and $ss(X_1, X_2) = 1$ if and only if $\mu_{X_1}(x_i) \leq \mu_{X_2}(x_i)$ for $\forall x_i$. Note also that $ss(X_1, X_2)$ in (3.13) and $s_J(X_1, X_2)$ in (3.3) have the same numerator but different denominators.

Rickard et al. [110] extended Kosko's definition of subsethood to IT2 FSs based on the Representation Theorem in Section 2.2.2.

**Definition 18** *Let $\tilde{X}_1$ and $\tilde{X}_2$ be two IT2 FSs, and $X_e^1$ and $X_e^2$ be their embedded T1 FSs. Then, the subsethood of $\tilde{X}_1$ in $\tilde{X}_2$, $SS(\tilde{X}_1, \tilde{X}_2)$, is defined as*

$$\begin{aligned} SS(\tilde{X}_1, \tilde{X}_2) &= \bigcup_{\forall X_e^1, X_e^2} ss(X_e^1, X_e^2) \\ &= \bigcup_{\forall X_e^1, X_e^2} \frac{\sum_{i=1}^N \min(\mu_{X_e^1}(x_i), \mu_{X_e^2}(x_i))}{\sum_{i=1}^N \mu_{X_e^1}(x_i)} \\ &\equiv [ss_l(\tilde{X}_1, \tilde{X}_2), ss_r(\tilde{X}_1, \tilde{X}_2)] \end{aligned} \tag{3.14}$$

*where*

$$ss_l(\tilde{X}_1, \tilde{X}_2) = \min_{\forall X_e^1, X_e^2} \frac{\sum_{i=1}^N \min(\mu_{X_e^1}(x_i), \mu_{X_e^2}(x_i))}{\sum_{i=1}^N \mu_{X_e^1}(x_i)} = \min_{\forall X_e^1} \frac{\sum_{i=1}^N \min(\mu_{X_e^1}(x_i), \mu_{\underline{X}_2}(x_i))}{\sum_{i=1}^N \mu_{X_e^1}(x_i)} \tag{3.15}$$

$$ss_r(\tilde{X}_1, \tilde{X}_2) = \max_{\forall X_e^1, X_e^2} \frac{\sum_{i=1}^N \min(\mu_{X_e^1}(x_i), \mu_{X_e^2}(x_i))}{\sum_{i=1}^N \mu_{X_e^1}(x_i)} = \max_{\forall X_e^1} \frac{\sum_{i=1}^N \min(\mu_{X_e^1}(x_i), \mu_{\overline{X}_2}(x_i))}{\sum_{i=1}^N \mu_{X_e^1}(x_i)} \quad \blacksquare \tag{3.16}$$

The second parts of (3.15) and (3.16) are obvious because $\mu_{X_e^2}(x_i)$ only appear in the numerators.

**Definition 19** *The* average subsethood *of* $\tilde{X}_1$ *in* $\tilde{X}_2$, $ss(\tilde{X}_1, \tilde{X}_2)$, *is the center of* $SS(\tilde{X}_1, \tilde{X}_2)$, *i.e.*,

$$ss(\tilde{X}_1, \tilde{X}_2) = \frac{ss_l(\tilde{X}_1, \tilde{X}_2) + ss_r(\tilde{X}_1, \tilde{X}_2)}{2}. \quad \blacksquare \tag{3.17}$$

To compute $ss(\tilde{X}_1, \tilde{X}_2)$, $ss_l(\tilde{X}_1, \tilde{X}_2)$ and $ss_r(\tilde{X}_1, \tilde{X}_2)$ must first be obtained. Define

$$\mu_{X_l}(x_i) = \begin{cases} \mu_{\overline{X}_1}(x_i), & \mu_{\underline{X}_2}(x_i) \le \mu_{\underline{X}_1}(x_i) \\ \mu_{\underline{X}_1}(x_i), & \mu_{\underline{X}_2}(x_i) \ge \mu_{\overline{X}_1}(x_i) \\ \{\mu_{\underline{X}_1}(x_i), \mu_{\overline{X}_1}(x_i)\}, & \mu_{\underline{X}_1}(x_i) < \mu_{\underline{X}_2}(x_i) < \mu_{\overline{X}_1}(x_i) \end{cases} \tag{3.18}$$

$$\mu_{X_r}(x_i) = \begin{cases} \mu_{\underline{X}_1}(x_i), & \mu_{\overline{X}_2}(x_i) \le \mu_{\underline{X}_1}(x_i) \\ \mu_{\overline{X}_1}(x_i), & \mu_{\overline{X}_2}(x_i) \ge \mu_{\overline{X}_1}(x_i) \\ \mu_{\overline{X}_2}(x_i), & \mu_{\underline{X}_1}(x_i) < \mu_{\overline{X}_2}(x_i) < \mu_{\overline{X}_1}(x_i) \end{cases} \tag{3.19}$$

Then, (3.15) and (3.16) can be computed as [110, 151]:

$$ss_l(\tilde{X}_1, \tilde{X}_2) = \min_{\mu_{X_l}(x_i) \text{ in } (3.18)} \left[ \frac{\sum_{i=1}^N \min\left(\mu_{X_l}(x_i), \mu_{\underline{X}_2}(x_i)\right)}{\sum_{i=1}^N \mu_{X_l}(x_i)} \right] \tag{3.20}$$

$$ss_r(\tilde{X}_1, \tilde{X}_2) = \frac{\sum_{i=1}^N \min\left(\mu_{X_r}(x_i), \mu_{\overline{X}_2}(x_i)\right)}{\sum_{i=1}^N \mu_{X_r}(x_i)} \tag{3.21}$$

The derivations of (3.20) and (3.21) are given in Appendix B.

Note that $ss_r(\tilde{X}_1, \tilde{X}_2)$ has a closed-form solution; however, because for each $x_i \in I_l = \{x_i | \mu_{\underline{X}_1}(x_i) < \mu_{\underline{X}_2}(x_i) < \mu_{\overline{X}_1}(x_i)\}$, $\mu_{X_l}(x_i)$ can have two possible values, to compute

$ss_l(\tilde{X}_1, \tilde{X}_2)$, $2^L$ evaluations of the bracketed terms in (3.20) have to be performed, where $L$ is the number of elements in $I_l$, and this can be a rather large number depending upon $L$. Note, also, that even though only the third line of $\mu_{X_l}(x_i)$ in (3.18) is used to established $I_l$, all three lines are used to compute $ss_l(\tilde{X}_1, \tilde{X}_2)$ because the summations in both the numerator and denominator of the bracketed function use all $N$ values of $\mu_{X_l}(x_i)$.

**Example 7** *Consider $\tilde{X}_1$ and $\tilde{X}_2$ shown in Fig. 3.5, where $\mu_{\underline{X}_1}(x_i)$, $\mu_{\overline{X}_1}(x_i)$, $\mu_{\underline{X}_2}(x_i)$, $\mu_{\overline{X}_2}(x_i)$, $\mu_{X_l}(x_i)$ and $\mu_{X_r}(x_i)$ are summarized in Table 3.2. Observe (see the $i = 5$ and 6 columns in Table 3.2) that $I_l = \{x_5, x_6\}$, $\mu_{X_l}(x_5) = \{\mu_{\underline{X}_1}(x_5), \mu_{\overline{X}_1}(x_5)\} = \{0, 1\}$ and $\mu_{X_l}(x_6) = \{\mu_{\underline{X}_1}(x_6), \mu_{\overline{X}_1}(x_6)\} = \{0, 1\}$. Because $L = 2$, $2^L = 4$ evaluations of the bracketed terms in (3.20) have to be performed before $ss_l(\tilde{X}_1, \tilde{X}_2)$ can be obtained:*

- *When $\mu_{X_l}(x_5) = 0$ and $\mu_{X_l}(x_6) = 0$,*

$$\frac{\sum_{i=1}^{10} \min\left(\mu_{X_l}(x_i), \mu_{\underline{X}_2}(x_i)\right)}{\sum_{i=1}^{10} \mu_{X_l}(x_i)} = \frac{0+0+0+0+0+0+0+0+0+0}{0+0.5+1+1+0+0+0+0+0+0} = 0$$

- *When $\mu_{X_l}(x_5) = 0$ and $\mu_{X_l}(x_6) = 1$,*

$$\frac{\sum_{i=1}^{10} \min\left(\mu_{X_l}(x_i), \mu_{\underline{X}_2}(x_i)\right)}{\sum_{i=1}^{10} \mu_{X_l}(x_i)} = \frac{0+0+0+0+0+0.4+0+0+0+0}{0+0.5+1+1+0+1+0+0+0+0} = 0.11$$

- *When* $\mu_{X_l}(x_5) = 1$ *and* $\mu_{X_l}(x_6) = 0$,

$$\frac{\sum_{i=1}^{10} \min\left(\mu_{X_l}(x_i), \mu_{\underline{X}_2}(x_i)\right)}{\sum_{i=1}^{10} \mu_{X_l}(x_i)} = \frac{0+0+0+0+0.2+0+0+0+0+0}{0+0.5+1+1+1+0+0+0+0+0} = 0.06$$

- *When* $\mu_{X_l}(x_5) = 1$ *and* $\mu_{X_l}(x_6) = 1$,

$$\frac{\sum_{i=1}^{10} \min\left(\mu_{X_l}(x_i), \mu_{\underline{X}_2}(x_i)\right)}{\sum_{i=1}^{10} \mu_{X_l}(x_i)} = \frac{0+0+0+0+0.2+0.4+0+0+0+0}{0+0.5+1+1+1+1+0+0+0+0} = 0.13$$

*It follows that* $ss_l(\tilde{X}_1, \tilde{X}_2) = \min\{0, 0.11, 0.06, 0.13\} = 0$. $ss_r(\tilde{X}_1, \tilde{X}_2)$ *has a closed-form solution, i.e.,*

$$ss_r(\tilde{X}_1, \tilde{X}_2) = \frac{0+0+0+0.2+0.4+0.6+0.5+0+0+0}{0+0+0.3+0.6+0.4+0.6+0.5+0+0+0} = 0.71$$

*and hence* $ss(\tilde{X}_1, \tilde{X}_2) = (0 + 0.71)/2 = 0.36$. ■

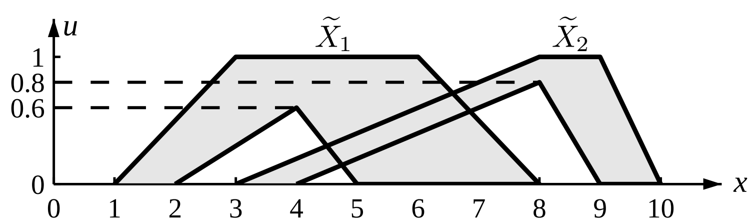



Fig. 3.5: $\tilde{X}_1$ and $\tilde{X}_2$ used to compute $ss(\tilde{X}_1, \tilde{X}_2)$.

Table 3.2: $\mu_{\underline{X}_1}(x_i)$, $\mu_{\overline{X}_1}(x_i)$, $\mu_{\underline{X}_2}(x_i)$, $\mu_{\overline{X}_2}(x_i)$, $\mu_{X_l}(x_i)$ and $\mu_{X_r}(x_i)$ for $\tilde{X}_1$ and $\tilde{X}_2$ shown in Fig. 3.5.

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $\mu_{\underline{X}_1}(x_i)$ | 0 | 0 | 0.3 | 0.6 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\mu_{\overline{X}_1}(x_i)$ | 0 | 0.5 | 1 | 1 | 1 | 1 | 0.5 | 0 | 0 | 0 |
| $\mu_{\underline{X}_2}(x_i)$ | 0 | 0 | 0 | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 0 | 0 |
| $\mu_{\overline{X}_2}(x_i)$ | 0 | 0 | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1 | 1 | 0 |
| $\mu_{X_l}(x_i)$ | 0 | 0.5 | 1 | 1 | {0, 1} | {0, 1} | 0 | 0 | 0 | 0 |
| $\mu_{X_r}(x_i)$ | 0 | 0 | 0.3 | 0.6 | 0.4 | 0.6 | 0.5 | 0 | 0 | 0 |

### 3.4.3 An Efficient Algorithm for Computing $ss_l(\tilde{X}_1, \tilde{X}_2)$

An efficient algorithm for computing $ss_l(\tilde{X}_1, \tilde{X}_2)$ is proposed in [151] and given on the next page. Its idea is similar to [28, 75]. The efficient algorithm can reduce the computational cost significantly, especially when $L$ is large [151].

Extensive simulations have been performed to compare the performances of the exhaustive computation approach in [110] [i.e., compute all possible $2^L$ combinations of the bracketed term in (3.20) and then choose the minimum] and the efficient algorithm for computing $ss_l(\tilde{X}_1, \tilde{X}_2)$. The platform was an IBM T43 notebook computer running Windows XP X32 Edition and Matlab 7.4.0 with Intel Pentium M 2GHz processors and 1GB RAM.

In the simulations, $N$, the number of samples in the $x$ domain, were chosen to be $\{10, 20, 50, 100, 1000, 10000\}$. For each $N$, 1,000 Monte Carlo simulations were used to compute $ss_l(\tilde{X}_1, \tilde{X}_2)$, i.e., for each $N$, 1,000 $\mu_{\underline{X}_2}(x_i)$ were generated using Matlab function *rand(1000,1)*, and 1,000 pairs of $\{\mu_{\underline{X}_1}(x_i), \mu_{\overline{X}_1}(x_i)\}$ were generated by using Matlab function *rand(1000,2)*. All $\mu_{\underline{X}_2}(x_i)$, $\mu_{\underline{X}_1}(x_i)$ and $\mu_{\overline{X}_1}(x_i)$ were constrained in $[0, 1]$, and

**Algorithm 1** Efficient Algorithm for Computing $ss_l(\tilde{X}_1, \tilde{X}_2)$

**Initialization:**

Find $I_l = \{x_i | \mu_{\underline{X}_1}(x_i) < \mu_{\underline{X}_2}(x_i) < \mu_{\overline{X}_1}(x_i)\}$, $\mu_{X_l}(x_i)$

$$\mu_{X_l}(x_i) = \begin{cases} \mu_{\overline{X}_1}(x_i), & \mu_{\underline{X}_2}(x_i) \le \mu_{\underline{X}_1}(x_i) \\ \mu_{\underline{X}_1}(x_i), & \mu_{\underline{X}_2}(x_i) \ge \mu_{\overline{X}_1}(x_i) \\ \mu_{\underline{X}_1}(x_i), & x_i \in I_l \end{cases}$$

$num = \sum_{i=1}^{N} \min(\mu_{X_l}(x_i), \mu_{\underline{X}_2}(x_i))$

$den = \sum_{i=1}^{N} \mu_{X_l}(x_i)$

$ss_l(\tilde{X}_1, \tilde{X}_2) = num/den$

$ss_{l0} = 1$

**Iteration:**

***while*** $ss_{l0} > ss_l(\tilde{X}_1, \tilde{X}_2)$

  set $ss_{l0} = ss_l(\tilde{X}_1, \tilde{X}_2)$

  ***for*** each $x_j \in I_l$

    ***if*** $\mu_{\overline{X}_1}(x_j)$ is used in the current $ss_l(\tilde{X}_1, \tilde{X}_2)$

      replace $\mu_{\overline{X}_1}(x_j)$ by $\mu_{\underline{X}_1}(x_j)$ and keep all other items in $ss_l(\tilde{X}_1, \tilde{X}_2)$ the same, i.e.,

$$num' = num - \min(\mu_{\overline{X}_1}(x_j), \mu_{\underline{X}_2}(x_j)) + \min(\mu_{\underline{X}_1}(x_j), \mu_{\underline{X}_2}(x_j))$$
$$den' = den - \mu_{\overline{X}_1}(x_j) + \mu_{\underline{X}_1}(x_j)$$

    ***else***

      replace $\mu_{\underline{X}_1}(x_j)$ by $\mu_{\overline{X}_1}(x_j)$ and keep all other items in $ss_l(\tilde{X}_1, \tilde{X}_2)$ the same, i.e.,

$$num' = num - \min(\mu_{\underline{X}_1}(x_j), \mu_{\underline{X}_2}(x_j)) + \min(\mu_{\overline{X}_1}(x_j), \mu_{\underline{X}_2}(x_j))$$
$$den' = den - \mu_{\underline{X}_1}(x_j) + \mu_{\overline{X}_1}(x_j)$$

    ***end if***

    $ss'_l(\tilde{X}_1, \tilde{X}_2) = num'/den'$

    ***if*** $ss'_l(\tilde{X}_1, \tilde{X}_2) < ss_l(\tilde{X}_1, \tilde{X}_2)$

      $ss_l(\tilde{X}_1, \tilde{X}_2) = ss'_l(\tilde{X}_1, \tilde{X}_2)$

      $num = num'$

      $den = den'$

    ***end if***

  ***end for***

***end while***

**return** $ss_l(\tilde{X}_1, \tilde{X}_2)$

$\mu_{\underline{X}_2}(x_i)$ were independent of $\mu_{\underline{X}_1}(x_i)$ and $\mu_{\overline{X}_1}(x_i)$. To make sure $\mu_{\underline{X}_1}(x_i) \leq \mu_{\overline{X}_1}(x_i)$, each pair of $\{\mu_{\underline{X}_1}(x_i), \mu_{\overline{X}_1}(x_i)\}$ was checked and the smaller value was assigned to $\mu_{\underline{X}_1}(x_i)$ and the larger value was assigned to $\mu_{\overline{X}_1}(x_i)$.

The computation time and average number of iterations[6] for the two algorithms are shown in Table 3.3 for different $N$. Observe that the efficient algorithm outperforms the exhaustive computation approach significantly, and it needs only a few iterations to converge.

Table 3.3: Computation time and average number of iterations for the two algorithms used to compute $ss_l(\tilde{X}_1, \tilde{X}_2)$. The results for $N \geq 100$ in the exhaustive computation approach are not shown because $2^L$ was too large for the computations to be performed.

| $N$ | Exhaustive Computation Approach | | Efficient Algorithm | |
|---|---|---|---|---|
| | Avg. Time (sec) | Avg. $2^L$ | Avg. Time (sec) | Avg. No. of Iterations |
| 10 | 0.0016 | 10 | 0.0001 | 1.8322 |
| 20 | 0.0331 | 97 | 0.0001 | 2.0451 |
| 50 | 17.8280 | 53,232 | 0.0001 | 2.2173 |
| 100 | — | — | 0.0001 | 2.3937 |
| 1,000 | — | — | 0.0004 | 2.9998 |
| 10,000 | — | — | 0.0040 | 3.0460 |

### 3.4.4 Properties of the Average Subsethood

The Jaccard similarity measure and the average subsethood bare a strong resemblance, hence it is interesting to check $ss(\tilde{X}_1, \tilde{X}_2)$ against the four properties of $s_J(\tilde{X}_1, \tilde{X}_2)$ (Section 3.2.2) — reflexivity, symmetry, transitivity and overlapping. First, several definitions that are used in the average subsethood properties are introduced.

[6]The number of iterations is defined as the number of times the *while* loop in the Efficient Algorithm is executed before $ss_l(\tilde{X}_1, \tilde{X}_2)$ is found.

The following theorem describes four properties for average subsethood, and the proof is given in the Appendix.

**Theorem 6** *$ss(\tilde{X}_1, \tilde{X}_2)$ defined in (3.17) has the following properties:*

1. *Reflexivity:* $ss(\tilde{X}_1, \tilde{X}_2) = 1 \Rightarrow \tilde{X}_1 \le \tilde{X}_2$.

2. *Asymmetry: Generally* $ss(\tilde{X}_1, \tilde{X}_2) \ne ss(\tilde{X}_2, \tilde{X}_1)$.

3. *Transitivity: If* $\tilde{X}_1 \le \tilde{X}_2$, *then* $ss(\tilde{X}_3, \tilde{X}_1) \le ss(\tilde{X}_3, \tilde{X}_2)$ *for any* $\tilde{X}_3$.

4. *Overlapping: If* $\tilde{X}_1 \cap \tilde{X}_2 \ne \emptyset$, *then* $ss(\tilde{X}_1, \tilde{X}_2) > 0$; *otherwise,* $ss(\tilde{X}_1, \tilde{X}_2) = 0$. ■

*Proof:* The four properties in Theorem 6 are considered separately.

1. *Reflexivity:* $ss(\tilde{X}_1, \tilde{X}_2) = 1$ means $ss(X_1^e, X_2^e) = 1$ for *every* pair of embedded T1 FSs $X_1^e$ and $X_2^e$, because otherwise $ss_l(\tilde{X}_1, \tilde{X}_2) = \min\limits_{\forall X_1^e, X_2^e} ss(X_1^e, X_2^e) < 1$ and hence $ss(\tilde{X}_1, \tilde{X}_2) < 1$. Choose $X_1^e = \overline{X}_1$ and $X_2^e = \underline{X}_2$; hence, it follows, from $ss(X_1^e, X_2^e) = 1$, that $ss(\overline{X}_1, \underline{X}_2) = 1$, i.e., $\mu_{\overline{X}_1}(x_i) \le \mu_{\underline{X}_2}(x_i)$ for $\forall x_i$ [see (3.13) and the comments under it]. This means that $\tilde{X}_1$ is completely below or touching $\tilde{X}_2$ [see Fig. 3.2(a)]. Consequently, $\tilde{X}_1 \le \tilde{X}_2$.

   Note that $\tilde{X}_1 \le \tilde{X}_2$ does not necessarily mean $ss(\tilde{X}_1, \tilde{X}_2) = 1$, as illustrated by Example 8.

2. *Asymmetry:* From (3.14), it is true that

$$SS(\tilde{X}_1, \tilde{X}_2) = \bigcup_{\forall X_1^e, X_2^e} \frac{\sum_{i=1}^N \min(\mu_{X_1^e}(x_i), \mu_{X_2^e}(x_i))}{\sum_{i=1}^N \mu_{X_1^e}(x_i)} \tag{3.22}$$

and

$$SS(\tilde{X}_2, \tilde{X}_1) = \bigcup_{\forall X_1^e, X_2^e} \frac{\sum_{i=1}^N \min(\mu_{X_1^e}(x_i), \mu_{X_2^e}(x_i))}{\sum_{i=1}^N \mu_{X_2^e}(x_i)} \tag{3.23}$$

i.e., $ss(\tilde{X}_1, \tilde{X}_2)$ and $ss(\tilde{X}_2, \tilde{X}_1)$ have the same numerators but different denominators. Generally, $SS(\tilde{X}_1, \tilde{X}_2) \neq SS(\tilde{X}_2, \tilde{X}_1)$; hence, $ss(\tilde{X}_1, \tilde{X}_2) \neq ss(\tilde{X}_2, \tilde{X}_1)$, as illustrated by Example 9.

3. *Transitivity:* According to Definition 15, $\tilde{X}_1 \leq \tilde{X}_2$ means $\mu_{\underline{X}_1}(x_i) \leq \mu_{\underline{X}_2}(x_i)$ and $\mu_{\overline{X}_1}(x_i) \leq \mu_{\overline{X}_2}(x_i)$ for $\forall x_i$; hence, for an arbitrary IT2 FS $\tilde{X}_3$, it follows from (3.15) and (3.16), that:

$$\begin{aligned} ss_l(\tilde{X}_3, \tilde{X}_1) &= \min_{\forall X_3^e} \frac{\sum_{i=1}^N \min(\mu_{X_3^e}(x_i), \mu_{\underline{X}_1}(x_i))}{\sum_{i=1}^N \mu_{X_3^e}(x_i)} \\ &\leq \min_{\forall X_3^e} \frac{\sum_{i=1}^N \min(\mu_{X_3^e}(x_i), \mu_{\underline{X}_2}(x_i))}{\sum_{i=1}^N \mu_{X_3^e}(x_i)} \\ &= ss_l(\tilde{X}_3, \tilde{X}_2) \end{aligned} \tag{3.24}$$

$$\begin{aligned} ss_r(\tilde{X}_3, \tilde{X}_1) &= \max_{\forall X_3^e} \frac{\sum_{i=1}^N \min(\mu_{X_3^e}(x_i), \mu_{\overline{X}_1}(x_i))}{\sum_{i=1}^N \mu_{X_3^e}(x_i)} \\ &\leq \max_{\forall X_3^e} \frac{\sum_{i=1}^N \min(\mu_{X_3^e}(x_i), \mu_{\overline{X}_2}(x_i))}{\sum_{i=1}^N \mu_{X_3^e}(x_i)} \\ &= ss_r(\tilde{X}_3, \tilde{X}_2) \end{aligned} \tag{3.25}$$

Consequently, $ss(\tilde{X}_3, \tilde{X}_1) \leq ss(\tilde{X}_3, \tilde{X}_2)$. The equality holds when $\tilde{X}_1$ and $\tilde{X}_2$ are the same, or $\tilde{X}_3 \leq \tilde{X}_1$ (and hence $\tilde{X}_3 \leq \tilde{X}_2$), as illustrated by Example 10.

4. *Overlapping:* According to Definition 16, $\tilde{X}_1$ and $\tilde{X}_2$ overlap if $\min(\mu_{\overline{X}_1}(x_i), \mu_{\overline{X}_2}(x_i))$ $> 0$ for at least one $x_i$; hence, if $\tilde{X}_1 \cap \tilde{X}_2 \neq \emptyset$, then

$$\begin{aligned} ss_r(\tilde{X}_1, \tilde{X}_2) &= \max_{\forall X_1^e} \frac{\sum_{i=1}^N \min(\mu_{X_1^e}(x_i), \mu_{\overline{X}_2}(x_i))}{\sum_{i=1}^N \mu_{X_1^e}(x_i)} \\ &= \max \left\{ \frac{\sum_{i=1}^N \min(\mu_{\overline{X}_1}(x_i), \mu_{\overline{X}_2}(x_i))}{\sum_{i=1}^N \mu_{\overline{X}_1}(x_i)}, \right. \\ &\qquad\qquad \left. \max_{\forall X_1^e \neq \overline{X}_1} \frac{\sum_{i=1}^N \min(\mu_{X_1^e}(x_i), \mu_{\overline{X}_2}(x_i))}{\sum_{i=1}^N \mu_{X_1^e}(x_i)} \right\} \\ &\geq \frac{\sum_{i=1}^N \min(\mu_{\overline{X}_1}(x_i), \mu_{\overline{X}_2}(x_i))}{\sum_{i=1}^N \mu_{\overline{X}_1}(x_i)} \\ &> 0 \end{aligned} \tag{3.26}$$

Consequently, $ss(\tilde{X}_1, \tilde{X}_2) > 0$.

On the other hand, according to Lemma 3, when $\tilde{X}_1 \cap \tilde{X}_2 = \emptyset$, $\sum_{i=1}^N \min(\mu_{X_1^e}(x_i), \mu_{\underline{X}_2}(x_i)) = 0$ for all embedded T1 FSs $X_1^e$; hence, from (3.15) it follows that $ss_l(\tilde{X}_1, \tilde{X}_2) = 0$. Similarly, when $\tilde{X}_1 \cap \tilde{X}_2 = \emptyset$, $\sum_{i=1}^N \min(\mu_{X_1^e}(x_i), \mu_{\overline{X}_2}(x_i)) = 0$ for all embedded T1 FSs $X_1^e$; hence, from (3.16) it follows that $ss_r(\tilde{X}_1, \tilde{X}_2) = 0$. As a result, $ss(\tilde{X}_1, \tilde{X}_2) = 0$. ■

**Example 8** *For $\tilde{X}_1$ and $\tilde{X}_2$ in Fig. 3.2(a), because $\mu_{\underline{X}_2}(x_i) \geq \mu_{\overline{X}_1}(x_i)$ for $\forall x_i$, it follows that $\mu_{X_l}(x_i) = \mu_{\underline{X}_1}(x_i)$ [see (3.18)], and hence (3.20) becomes $ss_l(\tilde{X}_1, \tilde{X}_2) = \sum_{i=1}^N \mu_{\underline{X}_1}(x_i)$ $/ \sum_{i=1}^N \mu_{\underline{X}_1}(x_i) = 1$. Similarly, because $\mu_{\overline{X}_2}(x_i) \geq \mu_{\overline{X}_1}(x_i)$ for $\forall x_i$, it follows that $\mu_{X_r}(x_i) = \mu_{\overline{X}_1}(x_i)$ [see (3.19)], and hence (3.21) becomes $ss_r(\tilde{X}_1, \tilde{X}_2) = \sum_{i=1}^N \mu_{\overline{X}_1}(x_i)$ $/ \sum_{i=1}^N \mu_{\overline{X}_1}(x_i) = 1$. Consequently, the average subsethood is $ss(\tilde{X}_1, \tilde{X}_2) = 1$.*

*For* $\tilde{X}_1$ *and* $\tilde{X}_2$ *in Fig. 3.2(b), because* $\mu_{\underline{X}_1}(x_i) \leq \mu_{\underline{X}_2}(x_i) \leq \mu_{\overline{X}_1}(x_i)$ *for* $\forall x_i$*, the efficient algorithm is needed to compute* $ss_l(\tilde{X}_1, \tilde{X}_2)$ *and the result is* $ss_l(\tilde{X}_1, \tilde{X}_2) = 0.72$*. Because* $\mu_{\overline{X}_2}(x_i) \geq \mu_{\overline{X}_1}(x_i)$ *for* $\forall x_i$*, it follows that* $\mu_{X_r}(x_i) = \mu_{\overline{X}_1}(x_i)$*, and hence (3.21) becomes* $ss_r(\tilde{X}_1, \tilde{X}_2) = \sum_{i=1}^N \mu_{\overline{X}_1}(x_i) / \sum_{i=1}^N \mu_{\overline{X}_1}(x_i) = 1$*. Consequently, the average subsethood is* $ss(\tilde{X}_1, \tilde{X}_2) = 0.86$*, i.e.,* $\tilde{X}_1 \leq \tilde{X}_2$ *does not necessarily mean* $ss(\tilde{X}_1, \tilde{X}_2) = 1$*.*

■

**Example 9** *Consider again* $\tilde{X}_1$ *and* $\tilde{X}_2$ *in Fig. 3.2(a). Example 8 has shown that* $ss(\tilde{X}_1, \tilde{X}_2) = 1$*. This example shows* $ss(\tilde{X}_2, \tilde{X}_1) < 1$*, and hence* $ss(\tilde{X}_1, \tilde{X}_2) \neq ss(\tilde{X}_2, \tilde{X}_1)$*.*

*Let* $X'_l$ *be the embedded T1 FS of* $\tilde{X}_2$ *from which* $ss_l(\tilde{X}_2, \tilde{X}_1)$ *is computed, and its MF be* $\mu_{X'_l}(x_i)$*. Then, by analogy to* $X_l$ *in (3.18),*

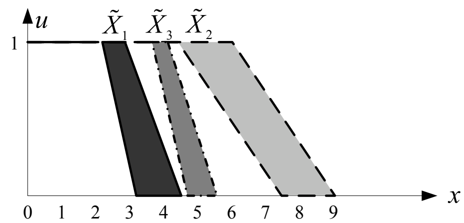
$$\mu_{X'_l}(x_i) = \begin{cases} \mu_{\overline{X}_2}(x_i), & \mu_{\underline{X}_1}(x_i) \leq \mu_{\underline{X}_2}(x_i) \\ \mu_{\underline{X}_2}(x_i), & \mu_{\underline{X}_1}(x_i) \geq \mu_{\overline{X}_2}(x_i) \\ \mu_{\underline{X}_2}(x_i) \text{ or } \mu_{\overline{X}_2}(x_i), & x_i \in I'_l \end{cases} \tag{3.27}$$

*where* $I'_l \equiv \{x_i | \mu_{\underline{X}_2}(x_i) < \mu_{\underline{X}_1}(x_i) < \mu_{\overline{X}_2}(x_i)\}$*. Because in Fig. 3.2(a)* $\mu_{\underline{X}_1}(x_i) \leq \mu_{\underline{X}_2}(x_i)$ *for* $\forall x_i$*, it follows that* $\mu_{X'_l}(x_i) = \mu_{\overline{X}_2}(x_i)$ *for* $\forall x_i$*. Consequently,*

$$\begin{aligned} ss_l(\tilde{X}_2, \tilde{X}_1) &= \min_{X'_l \text{ in } (3.27)} \frac{\sum_{i=1}^N \min(\mu_{X'_l}(x_i), \mu_{\underline{X}_1}(x_i))}{\sum_{i=1}^N \mu_{X'_l}(x_i)} \\ &= \frac{\sum_{i=1}^N \min(\mu_{\overline{X}_2}(x_i), \mu_{\underline{X}_1}(x_i))}{\sum_{i=1}^N \mu_{\overline{X}_2}(x_i)} \\ &= \frac{\sum_{i=1}^N \mu_{\underline{X}_1}(x_i)}{\sum_{i=1}^N \mu_{\overline{X}_2}(x_i)} < 1 \end{aligned} \tag{3.28}$$

*Similarly, it can be shown that* $ss_r(\tilde{X}_2, \tilde{X}_1) < 1$*; hence,* $ss(\tilde{X}_2, \tilde{X}_1) < 1$. ■

**Example 10** *Consider again* $\tilde{X}_1$ *and* $\tilde{X}_2$ *in Fig. 3.2(a), which are also depicted in Fig. 3.6. This example shows that when* $\tilde{X}_1 \leq \tilde{X}_2$, $ss(\tilde{X}_3, \tilde{X}_1) \leq ss(\tilde{X}_3, \tilde{X}_2)$ *for an arbitrary* $\tilde{X}_3$.



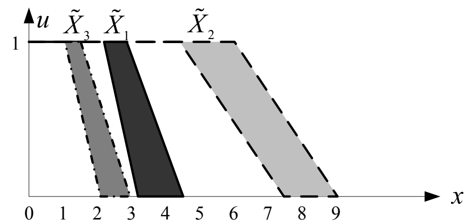




Fig. 3.6: (a) $\tilde{X}_1 \leq \tilde{X}_2$ and $ss(\tilde{X}_3, \tilde{X}_1) < ss(\tilde{X}_3, \tilde{X}_2)$; (b) $\tilde{X}_1 \leq \tilde{X}_2$ and $ss(\tilde{X}_3, \tilde{X}_1) = ss(\tilde{X}_3, \tilde{X}_2) = 1$. In both figures, $\tilde{X}_1$ is represented by the solid curves, $\tilde{X}_2$ is represented by the dashed curves, and $\tilde{X}_3$ is represented by the dash-dotted curves.

*For* $\tilde{X}_3$ *shown in Fig. 3.6(a),*

$$ss_l(\tilde{X}_3, \tilde{X}_1) = \min_{\forall X_3^e} \frac{\sum_{i=1}^N \min(\mu_{X_3^e}(x_i), \mu_{\underline{X}_1}(x_i))}{\sum_{i=1}^N \mu_{X_3^e}(x_i)} = \frac{\sum_{i=1}^N \mu_{\underline{X}_1}(x_i)}{\sum_{i=1}^N \mu_{\overline{X}_3}(x_i)} < 1 \tag{3.29}$$

$$ss_r(\tilde{X}_3, \tilde{X}_1) = \max_{\forall X_3^e} \frac{\sum_{i=1}^N \min(\mu_{X_3^e}(x_i), \mu_{\overline{X}_1}(x_i))}{\sum_{i=1}^N \mu_{X_3^e}(x_i)} = \frac{\sum_{i=1}^N \mu_{\overline{X}_1}(x_i)}{\sum_{i=1}^N \mu_{\underline{X}_3}(x_i)} < 1 \tag{3.30}$$

*and hence* $ss(\tilde{X}_3, \tilde{X}_1) < 1$*. On the other hand,*

$$ss_l(\tilde{X}_3, \tilde{X}_2) = \min_{\forall X_3^e} \frac{\sum_{i=1}^N \min(\mu_{X_3^e}(x_i), \mu_{\underline{X}_2}(x_i))}{\sum_{i=1}^N \mu_{X_3^e}(x_i)} = \min_{\forall X_3^e} \frac{\sum_{i=1}^N \mu_{X_3^e}(x_i)}{\sum_{i=1}^N \mu_{X_3^e}(x_i)} = 1 \tag{3.31}$$

$$ss_r(\tilde{X}_3, \tilde{X}_2) = \max_{\forall X_3^e} \frac{\sum_{i=1}^N \min(\mu_{X_3^e}(x_i), \mu_{\overline{X}_2}(x_i))}{\sum_{i=1}^N \mu_{X_3^e}(x_i)} = \max_{\forall X_3^e} \frac{\sum_{i=1}^N \mu_{X_3^e}(x_i)}{\sum_{i=1}^N \mu_{X_3^e}(x_i)} = 1 \tag{3.32}$$

*and hence* $ss(\tilde{X}_3, \tilde{X}_2) = 1$*, i.e.,* $ss(\tilde{X}_3, \tilde{X}_1) < ss(\tilde{X}_3, \tilde{X}_2)$.

*Similarly, it is easy to show that for* $\tilde{X}_1$*,* $\tilde{X}_2$ *and* $\tilde{X}_3$ *in Fig. 3.6(b),* $ss(\tilde{X}_3, \tilde{X}_1) = ss(\tilde{X}_3, \tilde{X}_2) = 1$*. In summary,* $ss(\tilde{X}_3, \tilde{X}_1) \leq ss(\tilde{X}_3, \tilde{X}_2)$ *when* $\tilde{X}_1 \leq \tilde{X}_2$*.* ■

### 3.4.5 Why Average Subsethood Instead of Similarity

Subsethood and similarity are closely related. A natural question then, is: Why should subsethood be used instead of similarity in classification?

Firstly, subsethood is conceptually more appropriate for a classifier, because it defines the degree that $\tilde{X}_1$ is contained in a class. It is not reasonable to compare the similarity between $\tilde{X}_1$ and the class-FOUs because they belong to different domains, e.g., $\tilde{X}_1$ represents the overall quality of a paper, whereas a class-FOU represents a recommendation.

Secondly, subsethood as a classifier gives more reasonable results, as illustrated by the following:

**Example 11** *Consider again the journal publication judgment advisor developed in [153]. The overall paper quality,* $\tilde{X}_1$*, and the three class-FOUs are shown in Fig. 3.7. It is visually clear that* $\tilde{X}_1$ *should be mapped to "Accept."*

*In this example,* $\tilde{X}_2 \equiv$ *Accept. Instead of computing* $ss_l(\tilde{X}_1, \tilde{X}_2)$ *and* $ss_r(\tilde{X}_1, \tilde{X}_2)$ *from (3.20) and (3.21), they can be easily computed directly from (3.15) and (3.16), because, for* $\forall X_e^1$ *(see Fig. 3.7)*

$$\min(\mu_{X_e^1}(x_i), \mu_{\underline{X}_2}(x_i)) = \mu_{X_e^1}(x_i) \tag{3.33}$$

$$\min(\mu_{X_e^1}(x_i), \mu_{\overline{X}_2}(x_i)) = \mu_{X_e^1}(x_i) \tag{3.34}$$

*hence,*

$$ss_l(\tilde{X}_1, \tilde{X}_2) = \min_{\forall X_e^1} \frac{\sum_{i=1}^N \mu_{X_e^1}(x_i)}{\sum_{i=1}^N \mu_{X_e^1}(x_i)} = 1 \tag{3.35}$$

$$ss_r(\tilde{X}_1, \tilde{X}_2) = \max_{\forall X_e^1} \frac{\sum_{i=1}^N \mu_{X_e^1}(x_i)}{\sum_{i=1}^N \mu_{X_e^1}(x_i)} = 1 \tag{3.36}$$

*so that the average subsethood of $\tilde{X}_1$ in "Accept" is 1. This indicates that this paper should be mapped unequivocally into class "Accept", which is consistent with our visual recommendation.*

*Next, let us compute $s_J(\tilde{X}_1, \tilde{X}_2)$ by using (3.4). Because (see Fig. 3.7)*

$$\begin{aligned} &\int_X \min(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))dx + \int_X \min(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))dx \\ = &\int_X \mu_{\overline{X}_1}(x)dx + \int_X \mu_{\underline{X}_1}(x)dx \end{aligned} \tag{3.37}$$

*and*

$$\begin{aligned} &\int_X \max(\mu_{\overline{X}_1}(x), \mu_{\overline{X}_2}(x))dx + \int_X \max(\mu_{\underline{X}_1}(x), \mu_{\underline{X}_2}(x))dx \\ = &\int_X \mu_{\overline{X}_2}(x))dx + \int_X \mu_{\underline{X}_2}(x))dx \end{aligned} \tag{3.38}$$

*it follows that*

$$s_J(\tilde{X}_1, \tilde{X}_2) = \frac{\int_X \mu_{\overline{X}_1}(x)dx + \int_X \mu_{\underline{X}_1}(x)dx}{\int_X \mu_{\overline{X}_2}(x))dx + \int_X \mu_{\underline{X}_2}(x))dx}. \tag{3.39}$$

*For $\tilde{X}_1$ in Fig. 3.7, one can compute the similarity between $\tilde{X}_1$ and "Accept" as 0.43. Moreover, as $\tilde{X}_1$ moves towards the right end of the domain, i.e., the overall quality of the paper gets better, the numerator of $s_J(\tilde{X}_1, \tilde{X}_2)$ decreases whereas the denominator remains the same; hence, the similarity between $\tilde{X}_1$ and "Accept" decreases, which is counter-intuitive.*

*Consequently, subsethood as a classifier gives much more reasonable results than similarity for this example.* ■

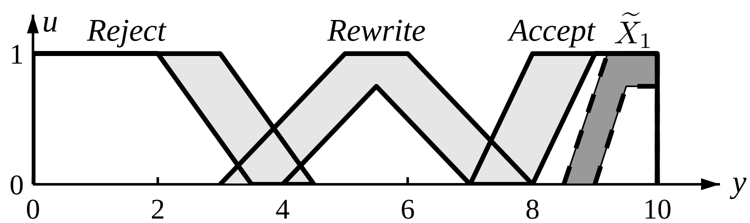



Fig. 3.7: Classifier as a decoder for the journal publication judgment advisor. $\tilde{X}_1$ is the overall quality of a paper.

# Chapter 4

# Novel Weighted Averages As a CWW Engine for MADM

Recall the Per-C depicted in Fig. 1.2, which consists of three components: encoder, decoder and CWW engine. The encoder transforms words into IT2 FSs that activate a CWW engine, as has been discussed in Section 2.3. The decoder maps the output of the CWW engine into a word and some accompanying data, as has been discussed in Chapter 3. The CWW engine maps IT2 FSs to IT2 FSs. There can be different kinds of CWW engines, e.g., novel weighted averages (NWAs) and perceptual reasoning. The NWAs are introduced in this chapter.

## 4.1 Novel Weighted Averages (NWAs)

The *weighted average* (WA) is arguably the earliest and still most widely used form of aggregation or fusion. We remind the reader of the well-known formula for the WA, i.e.,

$$y = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}, \tag{4.1}$$

in which $w_i$ are the weights (real numbers) that act upon the sub-criteria $x_i$ (real numbers). In many situations, however, providing crisp numbers for either the sub-criteria or the weights is problematic (there could be uncertainties about them), and it is more meaningful to provide intervals, T1 FSs, IT2 FSs, or a mixture of all of these, for the sub-criteria and weights.

**Definition 20** *An NWA is a WA in which at least one sub-criterion or weight is not a single real number, but is instead an interval, T1 FS or an IT2 FS, in which case such sub-criteria, weights, and the WA are called* novel models. ■

How to compute (4.1) for these novel models is the main subject of this chapter. What makes the computations challenging is the appearance of novel weights in both the numerator and denominator of (4.1). So, returning to the issue about normalized versus un-normalized weights, while everyone knows how to normalize a set of $n$ numerical weights (just divide each weight by the sum of all of the weights) it is not known how to normalize a set of $n$ novel weights.

Because there can be four possible models for sub-criteria or weights, there can be 16 different WAs, as summarized in Fig. 4.1.

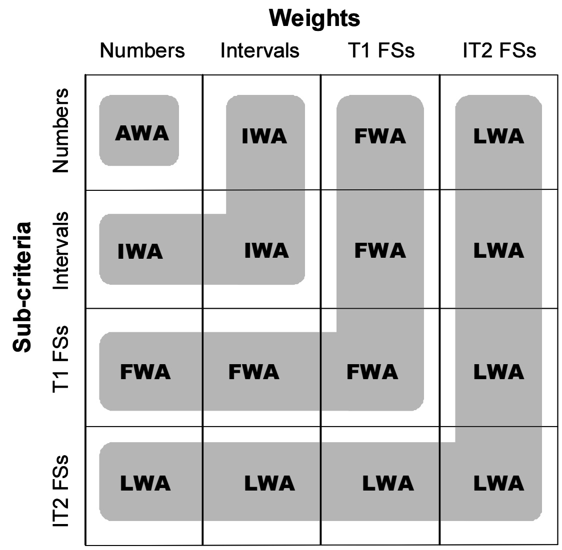



Fig. 4.1: Matrix of possibilities for a WA.

**Definition 21** *When at least one sub-criterion or weight is modeled as an interval, and all other sub-criteria or weights are modeled by no more than such a model, the resulting WA is called an* Interval WA (IWA). ■

**Definition 22** *When at least one sub-criterion or weight is modeled as a T1 FS, and all other sub-criteria or weights are modeled by no more than such a model, the resulting WA is called a* Fuzzy WA (FWA). ■

**Definition 23** *When at least one sub-criterion or weight is modeled as an IT2 FS, the resulting WA is called a* Linguistic WA (LWA). ■

**Definition 20** *(Continued): By a NWA is meant an IWA, FWA or LWA.* ■

From Fig. 4.1 it should be obvious that contained within the LWA are all of the other NWAs, suggesting that one should focus on the LWA and then view the other NWAs as its special cases (a top-down approach). Although this is possible, our approach will be to study NWAs from the bottom up, i.e. from the IWA to the FWA to the LWA, because (this is proved in Sections 4.3 and 4.4) the computation of a FWA uses a collection of IWAs, and the computation of a LWA uses two FWAs.

In order to reduce the number of possible derivations from 15 (the AWA is excluded) to three, it is assumed that: for the IWA *all* sub-criteria and weights are modeled as intervals, for the FWA *all* sub-criteria and weights are modeled as T1 FSs, and for the LWA *all* sub-criteria and weights are modeled as IT2 FSs.

## 4.2 Interval Weighted Average (IWA)

In (4.1) let

$$x_i \in [a_i, b_i] \qquad i = 1, ..., n \tag{4.2}$$

$$w_i \in [c_i, d_i] \qquad i = 1, ..., n \tag{4.3}$$

We associate interval sets $X_i$ and $W_i$ with (4.2) and (4.3), respectively, and refer to them as *intervals*.

The WA in (4.1) is now evaluated over the Cartesian product space

$$D_{X_1} \times D_{X_2} \times \cdots \times D_{X_n} \times D_{W_1} \times D_{W_2} \times \cdots \times D_{W_n}.$$

Regardless of the fact that this requires an uncountable number of evaluations[1], the resulting IWA, $Y_{IWA}$, will be a closed interval of non-negative real numbers, and is completely defined by its two end-points, $y_L$ and $y_R$, i.e.,

$$Y_{IWA} = [y_L, y_R] \tag{4.4}$$

Because $x_i$ $(i = 1, ..., n)$ appear only in the numerator of (4.1), the smallest (largest) value of each $x_i$ is used to find $y_L$ $(y_R)$, i.e.,

$$y_L = \min_{\forall w_i \in [c_i, d_i]} \frac{\sum_{i=1}^{n} a_i w_i}{\sum_{i=1}^{n} w_i} \tag{4.5}$$

$$y_R = \max_{\forall w_i \in [c_i, d_i]} \frac{\sum_{i=1}^{n} b_i w_i}{\sum_{i=1}^{n} w_i} \tag{4.6}$$

where the notations under min and max in (4.5) and (4.6) mean that $i$ ranges from 1 to $n$, and each $w_i$ ranges from $c_i$ to $d_i$.

It has been shown [83] that that $y_L$ and $y_R$ can be represented as

$$y_L = \frac{\sum_{i=1}^{L^*(\alpha)} a_i d_i + \sum_{i=L^*(\alpha)+1}^{n} a_i c_i}{\sum_{i=1}^{L^*(\alpha)} d_i + \sum_{i=L^*(\alpha)+1}^{n} c_i} \tag{4.7}$$

$$y_R = \frac{\sum_{i=1}^{R^*(\alpha)} b_i c_i + \sum_{i=R^*(\alpha)+1}^{n} b_i d_i}{\sum_{i=1}^{R^*(\alpha)} c_i + \sum_{i=R^*(\alpha)+1}^{n} d_i} \tag{4.8}$$

[1]Unless all of the $D_{X_i}$ and $D_{W_i}$ are first discretized, in which case there could still be an astronomically large but countable number of evaluations of (4.1), depending upon the number of terms in (4.1) (i.e., $n$) and the discretization size.

in which $L^*(\alpha)$ and $R^*(\alpha)$ are *switch points* that are found by using either KM or EKM Algorithms. In order to use these algorithms, $\{a_1, ..., a_n\}$ and $\{b_1, ..., b_n\}$ must be sorted in increasing order, respectively; hence, in the sequel, it is always assumed that

$$a_1 \leq a_2 \leq \cdots \leq a_n \tag{4.9}$$

$$b_1 \leq b_2 \leq \cdots \leq b_n \tag{4.10}$$

**Example 12** *Suppose for* $n = 5$, $\{x_i\}|_{i=1,\ldots,5} = \{8, 7, 5, 4, 1\}$ *and* $\{w_i\}|_{i=1,\ldots,5} = \{2, 1, 8, 4, 6\}$*, so that the arithmetic WA* $y_{AWA} = 4.14$*. Let* $\lambda$ *denote any of these crisp numbers. In this example, for the IWA,* $\lambda \to [\lambda - \delta, \lambda + \delta]$*, where* $\delta$ *may be different for different* $\lambda$*, i.e.,*

$$\{x_i\}|_{i=1,\ldots,5} \to \{[8.2, 9.8], [5.8, 8.2], [2.0, 8.0], [3.0, 5.0], [0.5, 1.5]\}$$

$$\{w_i\}|_{i=1,\ldots,5} \to \{[1.0, 3.0], [0.6, 1.4], [7.1, 8.9], [2.4, 5.6], [5.0, 7.0]\}$$

*It follows that* $Y_{IWA} = [2.02, 6.36]$*. Note that the average of* $Y_{IWA}$ *is 4.19, which is very close to the value of* $y_{AWA}$*. The important difference between* $y_{AWA}$ *and* $Y_{IWA}$ *is that the uncertainties about the sub-criteria and weights have led to an uncertainty band for the IWA, and such a band may play a useful role in subsequent decision-making.* ■

Finally, the following is a useful *expressive* way to summarize the IWA:

$$Y_{IWA} \equiv \frac{\sum_{i=1}^{n} X_i W_i}{\sum_{i=1}^{n} W_i} \tag{4.11}$$

where $X_i$ and $W_i$ are intervals whose elements are defined in (4.2) and (4.3), respectively, and $Y_{IWA}$ is also an interval. Of course, in order to explain the right-hand side of this expressive equation, one needs (4.4)-(4.6) and their accompanying discussions.

## 4.3 Fuzzy Weighted Average (FWA)

As for the IWA, let

$$x_i \in [a_i, b_i] \qquad i = 1, ..., n \tag{4.12}$$

$$w_i \in [c_i, d_i] \qquad i = 1, ..., n \tag{4.13}$$

but, unlike the IWA, where the membership grade for each $x_i$ and $w_i$ is 1, now the membership grade for each $x_i = x_i'$ and $w_i = w_i'$ is $\mu_{X_i}(x_i')$ and $\mu_{W_i}(w_i')$, respectively. So, now, T1 FSs $X_i$ and $W_i$ and their MFs $\mu_{X_i}(x_i)$ and $\mu_{W_i}(w_i)$ are associated with (4.12) and (4.13), respectively.

Again, the WA in (4.1) is evaluated over the Cartesian product space

$$D_{X_1} \times D_{X_2} \times \cdots \times D_{X_n} \times D_{W_1} \times D_{W_2} \times \cdots \times D_{W_n},$$

making use of $\mu_{X_1}(x_1), \mu_{X_2}(x_2), ..., \mu_{X_n}(x_n)$ and $\mu_{W_1}(w_1), \mu_{W_2}(w_2), ..., \mu_{W_n}(w_n)$, the result being a specific numerical value, $y$, as well as a degree of membership, $\mu_{Y_{FWA}}(y)$. How to compute the latter will be explained in Section 4.3.3 below. The result of each pair of computations is the pair $(y, \mu_{Y_{FWA}}(y))$, i.e.,

$$\Big\{(x_1, \mu_{X_1}(x_1)), ..., (x_n, \mu_{X_n}(x_n)), (w_1, \mu_{W_1}(w_1)), ..., (w_n, \mu_{W_n}(w_n))\Big\} \rightarrow \left(y = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}, \mu_{Y_{FWA}}(y)\right) \quad (4.14)$$

When this is done for all elements in the Cartesian product space, the FWA, $Y_{FWA}$, is obtained. By this explanation, observe that $Y_{FWA}$ is itself a T1 FS that is characterized by its MF $\mu_{Y_{FWA}}(y)$.

The FWA is a function of T1 FSs. The theory for computing any function of T1 FSs is introduced next. It uses Zadeh's Extension Principle [179].

### 4.3.1 Extension Principle

The *Extension Principle* was introduced by Zadeh in 1975 [179] and is an important tool in FS theory. It lets one extend mathematical relationships between non-fuzzy variables to fuzzy variables. Suppose, for example, one is given MFs for the FSs *small* and *light* and wants to determine the MF for the FS obtained by multiplying these FSs, i.e., *small* × *light*. The Extension Principle tells us how to determine the MF for *small* × *light* by making use of the non-fuzzy mathematical relationship $y = x_1 x_2$ in which the FS *small* plays the role of $x_1$ and the FS *light* plays the role of $x_2$.

Consider first a function of a single variable, $y = f(x)$, where $x \in D_X$ and $y \in D_Y$. A T1 FS $X$ is given, whose universe of discourse is also $D_X$, and whose MF is $\mu_X(x)$,

$\forall x \in D_X$. The Extension Principle [49, 141] states the image of $X$ under the mapping $f(x)$ can be expressed as another T1 FS $Y$, where

$$\mu_Y(y) = \begin{cases} \max\limits_{\forall x | y = f(x)} \mu_X(x) \\ 0 \quad \text{otherwise} \end{cases} \tag{4.15}$$

The condition in (4.15) that "$\mu_Y(y) = 0$ otherwise" means that if there are no values of $x$ for which a specific value of $y$ can be reached then the MF for that specific value of $y$ is set equal to zero. Only those values of $y$ that satisfy $y = f(x)$ can be reached.

So far, the Extension Principle has been stated just for a mapping of a single variable. Next, consider a function of more than one variable. Suppose that $y = f(x_1, x_2, ..., x_r)$, where $x_i \in D_{X_i}$ $(i = 1, ..., r)$. Let $X_1, X_2, ..., X_r$ be T1 FSs in $D_{X_1}, D_{X_2}, ..., D_{X_r}$. Then, the Extension Principle lets us induce from these $r$ T1 FSs a T1 FS $Y$ on $D_Y$, through $f$, i.e. $Y = f(X_1, X_2, ..., X_r)$, such that

$$\mu_Y(y) = \begin{cases} \sup\limits_{\forall (x_1, x_2, ..., x_r) | y = f(x_1, x_2, ..., x_r)} \min\{\mu_{X_1}(x_1), \mu_{X_2}(x_2), ..., \mu_{X_r}(x_r)\} \\ 0 \qquad \text{otherwise} \end{cases} \tag{4.16}$$

In order to implement (4.16), one must first find the values of $x_1, x_2, ..., x_r$ for which $y = f(x_1, x_2, ..., x_r)$, after which $\mu_{X_1}(x_1), \ldots, \mu_{X_r}(x_r)$ are computed at those values, and then $\min\{\mu_{X_1}(x_1), \mu_{X_2}(x_2), ..., \mu_{X_r}(x_r)\}$ is computed. If more than one set of $x_1, x_2, ..., x_r$ satisfy $y = f(x_1, x_2, ..., x_r)$, then this is repeated for all of them and the largest of the minima is chosen as $\mu_Y(y)$. Usually, the evaluation of (4.16) is very difficult, and the challenge is to find easier ways to do it than just described. The Function Decomposition Theorem that is given in Theorem 7 below is one such way.

Note, finally, that when it is necessary to extend an operation of the form $f(X_1, X_2, ..., X_r)$, where $X_i$ are T1 FSs, the individual operations like addition, multiplication, division, etc. that are involved in $f$ are not extended. Instead, the following is used [as derived from (4.16)]:

$$f(X_1, X_2, ..., X_r) = \int_{x_1 \in D_{X_1}} \int_{x_2 \in D_{X_2}} \cdots \int_{x_r \in D_{X_r}} \mu_Y(y)/f(x_1, x_2, ..., x_r) \tag{4.17}$$

where $\mu_Y(y)$ is defined in (4.16).

For example, if $y = f(x_1, x_2) \equiv (c_1 x_1 + c_2 x_2)/(x_1 + x_2)$, we write the extension of $f$ to T1 FSs $X_1$ and $X_2$ as

$$Y = f(X_1, X_2) = \int_{x_1 \in D_{X_1}} \int_{x_2 \in D_{X_2}} \mu_Y(y) \Bigg/ \frac{c_1 x_1 + c_2 x_2}{x_1 + x_2} \tag{4.18}$$

where

$$\mu_Y(y) = \begin{cases} \sup\limits_{\forall (x_1, x_2) | y = f(x_1, x_2)} & \min\{\mu_{X_1}(x_1), \mu_{X_2}(x_2)\} \\ 0 & \text{otherwise} \end{cases} \tag{4.19}$$

If we write it as $Y = f(X_1, X_2) \equiv (c_1 X_1 + c_2 X_2)/(X_1 + X_2)$, this does not mean that $f(X_1, X_2)$ is computed by adding and dividing the T1 FSs. It is merely an *expressive equation* computed by (4.18).

### 4.3.2 Computing a Function of T1 FSs Using $\alpha$-cuts

The ultimate objective of this section is to show that a function of T1 FSs can be expressed as the union (over all values of $\alpha$) of that function applied to the $\alpha$-cuts of the T1 FSs. The original idea, stated as the $\alpha$-cut Decomposition Theorem, is explained in [64]. Though that theorem does not require the T1 FSs to be normal, it does not point out explicitly how sub-normal T1 FSs should be handled. Because this theorem is so important, it is proved here for the convenience of the readers. Although the proof is very similar to that in [64], it emphasizes sub-normal cases as it is useful in Section 4.4.2.

We have just seen that the Extension Principle states that when the function $y = f(x_1, \ldots, x_r)$ is applied to T1 FSs $X_i$ $(i = 1, \ldots, r)$, the result is another T1 FS, $Y$, whose membership function is given by (4.16). Because $\mu_Y(y)$ is a T1 FS, it can therefore be expressed in terms of its $\alpha$-cuts as follows:

$$Y(\alpha) = \{y \,|\mu_Y(y) \geq \alpha\} \tag{4.20}$$

$$I_{Y(\alpha)}(y) = \begin{cases} 1, & \forall\, y \in Y(\alpha) \\ 0, & \forall\, y \notin Y(\alpha) \end{cases} \tag{4.21}$$

$$\mu_Y(y|\alpha) = \alpha I_{Y(\alpha)}(y) \tag{4.22}$$

$$\mu_Y(y) = \bigcup_{\alpha\in[0,1]} \mu_Y(y|\alpha) \tag{4.23}$$

In order to implement (4.21)-(4.23), a method is needed to compute $Y(\alpha)$, and this is provided in the following:

**Theorem 7** *(**Function Decomposition Theorem** [64]) Let $Y = f(X_1, \ldots, X_r)$ be an arbitrary (crisp) function, where $X_i$ ($i = 1, \ldots, r$) is a T1 FS whose domain is $D_{X_i}$ and $\alpha$-cut is $X_i(\alpha)$. Then under the Extension Principle:*

$$Y(\alpha) = f(X_1(\alpha), \ldots, X_r(\alpha)) \tag{4.24}$$

*and the height of $Y$ equals the minimum height of all $X_i$.* ■

*Proof*: For all $y \in D_Y$, from (4.20) it follows that[2]

$$y \in Y(\alpha) \Leftrightarrow \mu_Y(y) \geq \alpha \tag{4.25}$$

Under the Extension Principle in (4.16),

$$\mu_Y(y) \geq \alpha \Leftrightarrow \sup_{(x_1,\ldots,x_r)|y=f(x_1,\ldots,x_r)} \min\{\mu_{X_1}(x_1), \ldots, \mu_{X_r}(x_r)\} \geq \alpha \tag{4.26}$$

[2]The results in Theorem 7 are adapted from [64], Theorem 2.9, where they are stated and proved only for a function of a single variable. Even so, our proof of Theorem 7 follows the proof of their Theorem 2.9 very closely.

It follows that:

$$\sup_{(x_1,\ldots,x_r)|y=f(x_1,\ldots,x_r)} \min\{\mu_{X_1}(x_1),\ldots,\mu_{X_r}(x_r)\} \geq \alpha$$

$$\Leftrightarrow (\exists\ x_{10} \in D_{X_1} \text{ and } \cdots \text{ and } x_{r0} \in D_{X_r})$$

$$\ni (y = f(x_{10},\ldots,x_{r0}) \text{ and } \min\{\mu_{X_1}(x_{10}),\ldots,\mu_{X_r}(x_{r0})\} \geq \alpha)$$

$$\Leftrightarrow (\exists\ x_{10} \in D_{X_1} \text{ and } \cdots \text{ and } x_{r0} \in D_{X_r})$$

$$\ni (y = f(x_{10},\ldots,x_{r0}) \text{ and } [\mu_{X_1}(x_{10}) \geq \alpha \text{ and } \cdots \text{ and } \mu_{X_r}(x_{r0}) \geq \alpha])$$

$$\Leftrightarrow (\exists\ x_{10} \in D_{X_1} \text{ and } \cdots \text{ and } x_{r0} \in D_{X_r})$$

$$\ni (y = f(x_{10},\ldots,x_{r0}) \text{ and } [x_{10} \in X_1(\alpha) \text{ and } \cdots \text{ and } x_{r0} \in X_r(\alpha)])$$

$$\Leftrightarrow y \in f\left(X_1(\alpha),\ldots,X_r(\alpha)\right) \tag{4.27}$$

Hence, from the last line of (4.27) and (4.26),

$$\mu_Y(y) \geq \alpha \Leftrightarrow y \in f(X_1(\alpha),\ldots,X_r(\alpha)) \tag{4.28}$$

which means that

$$Y(\alpha) = f(X_1(\alpha),\ldots,X_r(\alpha)). \tag{4.29}$$

Because the right-hand-side of (4.26) (read from right to the left) indicates that $\alpha$ cannot exceed the minimum height of all $\mu_{X_i}(x_i)$ (otherwise there is no $\alpha$-cut on one or more $X_i$), the height of $Y$ must equal the minimum height of all $X_i$. ■

In summary, the Function Decomposition Theorem states that

*The MF for a function of T1 FSs equals the union (over all values of $\alpha$) of the MFs for the same function applied to the $\alpha$-cuts of the T1 FSs.*

The importance of this decomposition is that it reduces all computations to interval computations because all $\alpha$-cuts are intervals.

### 4.3.3 FWA Algorithms

The FWA is computed by using the Function Decomposition Theorem. There are three steps:

1. For each $\alpha \in [0, 1]$, the corresponding $\alpha$-cuts of the T1 FSs $X_i$ and $W_i$ must first be computed, i.e. compute

$$X_i(\alpha) = [a_i(\alpha), b_i(\alpha)] \quad i = 1, ..., n \tag{4.30}$$

$$W_i(\alpha) = [c_i(\alpha), d_i(\alpha)] \quad i = 1, ..., n \tag{4.31}$$

2. For each $\alpha \in [0, 1]$, compute the $\alpha$-cut of the FWA by recognizing that it is an IWA, i.e. $Y_{FWA}(\alpha) = Y_{IWA}(\alpha)$, where

$$Y_{IWA}(\alpha) = [y_L(\alpha), y_R(\alpha)] \tag{4.32}$$

in which [see (4.5) and (4.6)]

$$y_L(\alpha) = \min_{\forall w_i(\alpha)\in[c_i(\alpha),d_i(\alpha)]} \frac{\sum_{i=1}^{n} a_i(\alpha)w_i(\alpha)}{\sum_{i=1}^{n} w_i(\alpha)} \tag{4.33}$$

$$y_R(\alpha) = \max_{\forall w_i(\alpha)\in[c_i(\alpha),d_i(\alpha)]} \frac{\sum_{i=1}^{n} b_i(\alpha)w_i(\alpha)}{\sum_{i=1}^{n} w_i(\alpha)} \tag{4.34}$$

where the notations under min and max in (4.33) and (4.34) mean $i$ ranges from 1 to $n$, and each $w_i(\alpha)$ ranges from $c_i(\alpha)$ to $d_i(\alpha)$. From (4.7)-(4.10):

$$y_L(\alpha) = \frac{\sum_{i=1}^{L^*(\alpha)} a_i(\alpha)d_i(\alpha) + \sum_{i=L^*(\alpha)+1}^{n} a_i(\alpha)c_i(\alpha)}{\sum_{i=1}^{L^*(\alpha)} d_i(\alpha) + \sum_{i=L^*(\alpha)+1}^{n} c_i(\alpha)} \tag{4.35}$$

$$y_R(\alpha) = \frac{\sum_{i=1}^{R^*(\alpha)} b_i(\alpha)c_i(\alpha) + \sum_{i=R^*(\alpha)+1}^{n} b_i(\alpha)d_i(\alpha)}{\sum_{i=1}^{R^*(\alpha)} c_i(\alpha) + \sum_{i=R^*(\alpha)+1}^{n} d_i(\alpha)} \tag{4.36}$$

$$a_1(\alpha) \le a_2(\alpha) \le \cdots \le a_n(\alpha) \tag{4.37}$$

$$b_1(\alpha) \le b_2(\alpha) \le \cdots \le b_n(\alpha) \tag{4.38}$$

The KM or EKM Algorithms can be used to compute switch points $L^*(\alpha)$ and $R^*(\alpha)$. In practice, a finite number of $\alpha$-cuts are used, so that $\alpha \in [0,1] \to \{\alpha_1, \alpha_2, ..., \alpha_m\}$. If parallel processors are available, then all computations of this step can be done in parallel using $2m$ processors.

3. Connect all left-coordinates $(y_L(\alpha), \alpha)$ and all right-coordinates $(y_R(\alpha), \alpha)$ to form the T1 FS $Y_{FWA}$.

**Example 13** *This is a continuation of Example 12 in which each interval is assigned a symmetric triangular distribution that is centered at the mid-point* ($\lambda$) *of the interval, has distribution value equal to one at that point, and is zero at the interval end-points* ($\lambda - \delta$ *and* $\lambda + \delta$) *(see Fig. 4.2). The FWA is depicted in Fig. 4.3(c). Although* $Y_{FWA}$ *appears to be triangular, its sides are actually slightly curved.*

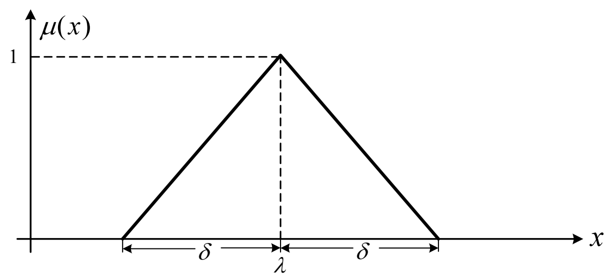


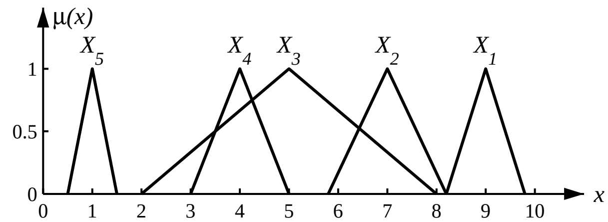

Fig. 4.2: Illustration of a T1 FS used in Example 13.

*The support of* $Y_{FWA}$ *is [2.02, 6.36], which is the same as* $Y_{IWA}$ *(see Example 12). This will always occur because the support of* $Y_{FWA}$ *is the* $\alpha = 0$ $\alpha$*-cut, and this is* $Y_{IWA}$.

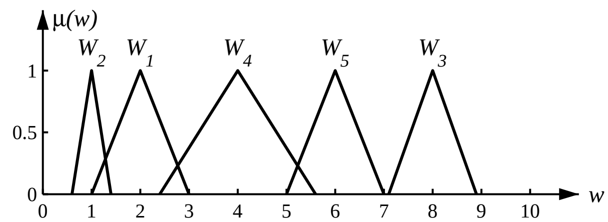
*The center of gravities of* $Y_{FWA}$ *and* $Y_{IWA}$ *are 4.15 and 4.19, respectively, and while close are not the same. The almost triangular distribution for* $Y_{FWA}$ *indicates that more emphasis should be given to values of variable y that are closer to 4.15, whereas the uniform distribution for* $Y_{IWA}$ *indicates that equal emphasis should be given to all values of variable y in its interval. The former reflects the propagation of the non-uniform uncertainties through the FWA, and can be used in future decisions.* ■

Finally, the following is a very useful *expressive* way to summarize the FWA:

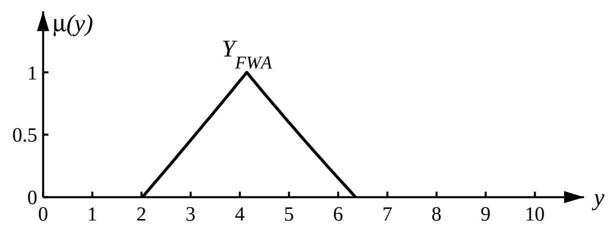
$$Y_{FWA} \equiv \frac{\sum_{i=1}^{n} X_i W_i}{\sum_{i=1}^{n} W_i} \tag{4.39}$$




(a)




(b)




(c)

Fig. 4.3: Example 13: (a) sub-criteria, (b) weights, and, (c) $Y_{FWA}$.

where $X_i$ and $W_i$ are T1 FSs that are characterized by $\mu_{X_i}(x_i)$ and $\mu_{W_i}(w_i)$, respectively, and $Y_{FWA}$ is also a T1 FS. Of course, in order to explain the right-hand side of this expressive equation, (4.14), (4.30)-(4.38), and their accompanying discussions are needed. Although the right-hand sides of (4.39) and (4.11) look the same, it is the accompanying models for $X_i$ and $W_i$ that distinguish one from the other.

## 4.4 Linguistic Weighted Average (LWA)

In the FWA, sub-criteria and weights are modeled as T1 FSs. In some situations it may be more appropriate to model sub-criteria and weights as IT2 FSs. When (4.1) is computed using IT2 FSs for sub-criteria and weights, then the result is the LWA, $\tilde{Y}_{LWA}$ [145, 148].

### 4.4.1 Introduction

As for the FWA, let

$$x_i \in [a_i, b_i] \quad i = 1, ..., n \tag{4.40}$$

$$w_i \in [c_i, d_i] \quad i = 1, ..., n \tag{4.41}$$

but, unlike the FWA, where the degree of membership for each $x_i = x_i'$ and $w_i = w_i'$ is $\mu_{X_i}(x_i')$ and $\mu_{W_i}(w_i')$, now the primary membership for each $x_i = x_i'$ and $w_i = w_i'$ is an interval $J_{x_i'}$ and $J_{w_i'}$, respectively. So, now, IT2 FSs $\tilde{X}_i$ and $\tilde{W}_i$ and their primary memberships $J_{x_i}$ and $J_{w_i}$ are associated with (4.40) and (4.41), respectively.

Now the WA in (4.1) is evaluated, but over the Cartesian product space

$$D_{\tilde{X}_1} \times D_{\tilde{X}_2} \times \cdots \times D_{\tilde{X}_n} \times D_{\tilde{W}_1} \times D_{\tilde{W}_2} \times \cdots \times D_{\tilde{W}_n},$$

making use of $J_{x_1}, J_{x_2}, ..., J_{x_n}$ and $J_{w_1}, J_{w_2}, ..., J_{w_n}$, the result being a specific numerical value, $y$, as well as the primary membership, $J_y$. Recall, from (2.14), that $J_{x_i} = [\underline{\mu}_{\tilde{X}_i}(x_i), \overline{\mu}_{\tilde{X}_i}(x_i)]$ and $J_{w_i} = [\mu_{\underline{W}_i}(w_i), \mu_{\overline{W}_i}(w_i)]$; consequently, $J_y = [\mu_{\underline{Y}_{LWA}}(y), \mu_{\overline{Y}_{LWA}}(y)]$. How to compute the latter interval of non-negative real numbers will be explained below[3]. The result of each pair of computations is the pair $(y, J_y)$, i.e.

$$\begin{aligned} &\{(x_1, J_{x_1}), ..., (x_n, J_{x_n}), (w_1, J_{w_1}), ..., (w_n, J_{w_n})\} \\ &\rightarrow \left(y = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}, J_y = \left[\mu_{\underline{Y}_{LWA}}(y), \mu_{\overline{Y}_{LWA}}(y)\right]\right) \end{aligned} \tag{4.42}$$

When this is done for all elements in the Cartesian product space $\tilde{Y}_{LWA}$ is obtained. By this explanation, observe that $\tilde{Y}_{LWA}$ is itself an IT2 FS that is characterized by its primary MF $J_y$, or equivalently by its FOU, $FOU(\tilde{Y}_{LWA})$, i.e.,

$$FOU(\tilde{Y}_{LWA}) = \bigcup_{\forall y \in D_{\tilde{Y}_{LWA}}} J_y = \left[\underline{Y}_{LWA}, \overline{Y}_{LWA}\right] \tag{4.43}$$

where $D_{\tilde{Y}_{LWA}}$ is the domain of the primary variable, and $\underline{Y}_{LWA}$ and $\overline{Y}_{LWA}$ are the LMF and UMF of $\tilde{Y}_{LWA}$, respectively, as shown in Fig. 4.4.

[3] A different derivation, which uses the Wavy Slice Representation Theorem (Section 2.2.2) for an IT2 FS, is given in [145, 148]; however, the results are the same as those presented in this chapter.

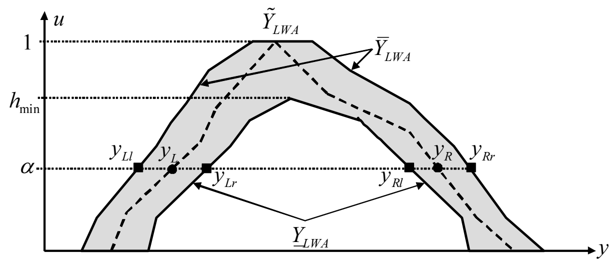



Fig. 4.4: $\tilde{Y}_{LWA}$ and associated quantities. The dashed curve is an embedded T1 FS of $\tilde{Y}_{LWA}$.

Similar to (4.39), the following is a very useful *expressive* way to summarize the LWA:

$$\tilde{Y}_{LWA} \equiv \frac{\sum_{i=1}^{n} \tilde{X}_i \tilde{W}_i}{\sum_{i=1}^{n} \tilde{W}_i} \tag{4.44}$$

where $\tilde{X}_i$ and $\tilde{W}_i$ are IT2 FSs that are characterized by their FOUs, and, $\tilde{Y}_{LWA}$ is also an IT2 FS.

Recall from the Wavy Slice Representation Theorem [(2.17) and (2.27)] that

$$\tilde{X}_i = 1/FOU(\tilde{X}_i) = 1/[\underline{X}_i, \overline{X}_i] \tag{4.45}$$

$$\tilde{W}_i = 1/FOU(\tilde{W}_i) = 1/[\underline{W}_i, \overline{W}_i] \tag{4.46}$$

as shown in Figs. 4.5 and 4.6. Because in (4.44) $\tilde{X}_i$ only appears in the numerator of $\tilde{Y}_{LWA}$, it follows that

$$\underline{Y}_{LWA} = \min_{\forall W_i \in [\underline{W}_i, \overline{W}_i]} \frac{\sum_{i=1}^{n} \underline{X}_i W_i}{\sum_{i=1}^{n} W_i} \tag{4.47}$$

$$\overline{Y}_{LWA} = \max_{\forall W_i \in [\underline{W}_i, \overline{W}_i]} \frac{\sum_{i=1}^{n} \overline{X}_i W_i}{\sum_{i=1}^{n} W_i} \tag{4.48}$$

By this preliminary approach to computing the LWA, it has been shown that it is only necessary to compute $\underline{Y}_{LWA}$ and $\overline{Y}_{LWA}$, as depicted in Fig. 4.4. One method is to compute the totality of all FWAs that can be formed from all of the embedded T1 FSs $W_i$; however, this is impractical because there can be infinite many $W_i$. An $\alpha$-cut based approach is proposed next. It eliminates the need to enumerate and evaluate all embedded T1 FSs.

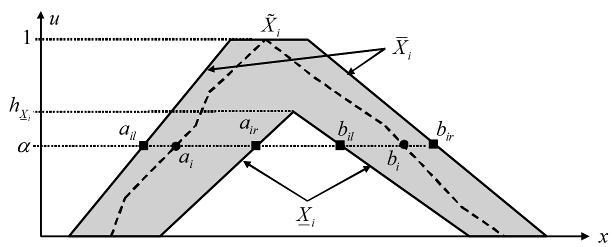



Fig. 4.5: $\tilde{X}_i$ and an $\alpha$-cut. The dashed curve is an embedded T1 FS of $\tilde{X}_i$.

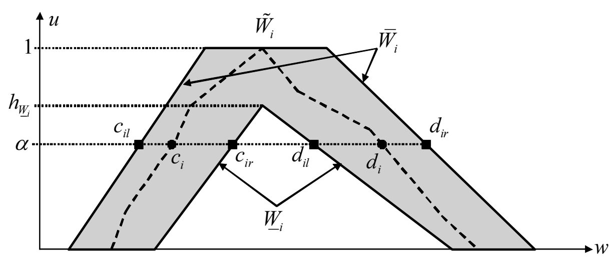



Fig. 4.6: $\tilde{W}_i$ and an $\alpha$-cut. The dashed curve is an embedded T1 FS of $\tilde{W}_i$.

### 4.4.2 Computing the LWA

Before $\underline{Y}_{LWA}$ and $\overline{Y}_{LWA}$ can be computed, their heights need to be determined. Because all UMFs are normal T1 FSs, $h_{\overline{Y}_{LWA}} = 1$. Denote the height of $\underline{X}_i$ as $h_{\underline{X}_i}$ and the height of $\underline{W}_i$ as $h_{\underline{W}_i}$. Let

$$h_{\min} = \min\{\min_{\forall i} h_{\underline{X}_i}, \min_{\forall i} h_{\underline{W}_i}\} \tag{4.49}$$

$h_{\min}$ is the smallest height of all FWAs computed from embedded T1 FSs of $\tilde{X}_i$ and $\tilde{W}_i$. Because $FOU(\tilde{Y}_{LWA})$ is the combination of all such FWAs, and $\underline{Y}_{LWA}$ is the lower bound of $FOU(\tilde{Y}_{LWA})$, it must hold that $h_{\underline{Y}_{LWA}} = h_{\min}$.

Let $[a_i(\alpha), b_i(\alpha)]$ be an $\alpha$-cut on an embedded T1 FS of $\tilde{X}_i$, and $[c_i(\alpha), d_i(\alpha)]$ be an $\alpha$-cut on an embedded T1 FS of $\tilde{W}_i$. Observe in Fig. 4.5, if the $\alpha$-cut on $\underline{X}_i$ exists, then the interval $[a_{il}(\alpha), b_{ir}(\alpha)]$ is divided into three sub-intervals: $[a_{il}(\alpha), a_{ir}(\alpha)]$, $(a_{ir}(\alpha), b_{il}(\alpha))$ and $[b_{il}(\alpha), b_{ir}(\alpha)]$. In this case, $a_i(\alpha) \in [a_{il}(\alpha), a_{ir}(\alpha)]$ and $a_i(\alpha)$ cannot assume a value larger than $a_{ir}(\alpha)$. Similarly, $b_i(\alpha) \in [b_{il}(\alpha), b_{ir}(\alpha)]$ and $b_i(\alpha)$ cannot assume a value smaller than $b_{il}(\alpha)$. However, if the $\alpha$-cut on $\underline{X}_i$ does not exist (e.g., $\alpha > h_{\underline{X}_i}$), then both $a_i(\alpha)$ and $b_i(\alpha)$ can assume values freely in the entire interval $[a_{il}(\alpha), b_{ir}(\alpha)]$, i.e.,

$$a_i(\alpha) \in \begin{cases} [a_{il}(\alpha), a_{ir}(\alpha)], & \alpha \in [0, h_{\underline{X}_i}] \\ [a_{il}(\alpha), b_{ir}(\alpha)], & \alpha \in (h_{\underline{X}_i}, 1] \end{cases} \tag{4.50}$$

$$b_i(\alpha) \in \begin{cases} [b_{il}(\alpha), b_{ir}(\alpha)], & \alpha \in [0, h_{\underline{X}_i}] \\ [a_{il}(\alpha), b_{ir}(\alpha)], & \alpha \in (h_{\underline{X}_i}, 1] \end{cases} \tag{4.51}$$

Similarly, observe from Fig. 4.6 that

$$c_i(\alpha) \in \begin{cases} [c_{il}(\alpha), c_{ir}(\alpha)], & \alpha \in [0, h_{\underline{W}_i}] \\ [c_{il}(\alpha), d_{ir}(\alpha)], & \alpha \in (h_{\underline{W}_i}, 1] \end{cases} \tag{4.52}$$

$$d_i(\alpha) \in \begin{cases} [d_{il}(\alpha), d_{ir}(\alpha)], & \alpha \in [0, h_{\underline{W}_i}] \\ [c_{il}(\alpha), d_{ir}(\alpha)], & \alpha \in (h_{\underline{W}_i}, 1] \end{cases} \tag{4.53}$$

In (4.50)-(4.53) subscript $i$ is the sub-criterion or weight index, $l$ means *left* and $r$ means *right*.

Using (4.50)-(4.53), let:

$$a_{ir}(\alpha) \triangleq \begin{cases} a_{ir}(\alpha), & \alpha \leq h_{\underline{X}_i} \\ b_{ir}(\alpha), & \alpha > h_{\underline{X}_i} \end{cases} \tag{4.54}$$

$$b_{il}(\alpha) \triangleq \begin{cases} b_{il}(\alpha), & \alpha \leq h_{\underline{X}_i} \\ a_{il}(\alpha), & \alpha > h_{\underline{X}_i} \end{cases} \tag{4.55}$$

$$c_{ir}(\alpha) \triangleq \begin{cases} c_{ir}(\alpha), & \alpha \leq h_{\underline{W}_i} \\ d_{ir}(\alpha), & \alpha > h_{\underline{W}_i} \end{cases} \tag{4.56}$$

$$d_{il}(\alpha) \triangleq \begin{cases} d_{il}(\alpha), & \alpha \leq h_{\underline{W}_i} \\ c_{il}(\alpha), & \alpha > h_{\underline{W}_i} \end{cases} \tag{4.57}$$

Then

$$a_i(\alpha) \in [a_{il}(\alpha), a_{ir}(\alpha)], \quad \forall \alpha \in [0, 1] \tag{4.58}$$

$$b_i(\alpha) \in [b_{il}(\alpha), b_{ir}(\alpha)], \quad \forall \alpha \in [0, 1] \tag{4.59}$$

$$c_i(\alpha) \in [c_{il}(\alpha), c_{ir}(\alpha)], \quad \forall \alpha \in [0, 1] \tag{4.60}$$

$$d_i(\alpha) \in [d_{il}(\alpha), d_{ir}(\alpha)], \quad \forall \alpha \in [0, 1] \tag{4.61}$$

Note that in (4.33) and (4.34) for the FWA, $a_i(\alpha)$, $b_i(\alpha)$, $c_i(\alpha)$ and $d_i(\alpha)$ are crisp numbers; consequently, $y_L(\alpha)$ and $y_R(\alpha)$ computed from them are also crisp numbers; however, in the LWA, $a_i(\alpha)$, $b_i(\alpha)$, $c_i(\alpha)$ and $d_i(\alpha)$ can assume values continuously in their corresponding $\alpha$-cut intervals. Numerous different combinations of $a_i(\alpha)$, $b_i(\alpha)$, $c_i(\alpha)$ and $d_i(\alpha)$ can be formed. $y_L(\alpha)$ and $y_R(\alpha)$ need to be computed for all the combinations. By collecting all $y_L(\alpha)$ a continuous interval $[y_{Ll}(\alpha), y_{Lr}(\alpha)]$ is obtained, and, by collecting all $y_R(\alpha)$ a continuous interval $[y_{Rl}(\alpha), y_{Rr}(\alpha)]$ is also obtained (see Fig. 4.4), i.e.

$$\underline{Y}_{LWA}(\alpha) = [y_{Lr}(\alpha), y_{Rl}(\alpha)], \quad \alpha \in [0, h_{\min}] \tag{4.62}$$

and

$$\overline{Y}_{LWA}(\alpha) = [y_{Ll}(\alpha), y_{Rr}(\alpha)], \quad \alpha \in [0, 1] \tag{4.63}$$

where $y_{Lr}(\alpha), y_{Rl}(\alpha), y_{Ll}(\alpha)$ and $y_{Rr}(\alpha)$ are illustrated in Fig. 4.4. Clearly, to find $\underline{Y}_{LWA}(\alpha)$ and $\overline{Y}_{LWA}(\alpha)$, $y_{Ll}(\alpha), y_{Lr}(\alpha), y_{Rl}(\alpha)$ and $y_{Rr}(\alpha)$ need to be found.

Consider $y_{Ll}(\alpha)$ first. Note that it lies on $\overline{Y}_{LWA}$, and is the minimum of $y_L(\alpha)$ but now $a_i(\alpha) \in [a_{il}(\alpha),\, a_{ir}(\alpha)]$, $c_i(\alpha) \in [c_{il}(\alpha),\, c_{ir}(\alpha)]$, and $d_i(\alpha) \in [d_{il}(\alpha),\, d_{ir}(\alpha)]$, i.e.

$$y_{Ll}(\alpha) = \min_{\substack{\forall a_i(\alpha) \in [a_{il}(\alpha),\, a_{ir}(\alpha)] \\ \forall c_i(\alpha) \in [c_{il}(\alpha),\, c_{ir}(\alpha)], \forall d_i(\alpha) \in [d_{il}(\alpha),\, d_{ir}(\alpha)]}} y_L(\alpha) \tag{4.64}$$

Substituting $y_L(\alpha)$ from (4.35) into (4.64), it follows that

$$y_{Ll}(\alpha) \equiv \min_{\substack{\forall a_i(\alpha)\in[a_{il}(\alpha),\, a_{ir}(\alpha)] \\ \forall c_i(\alpha)\in[c_{il}(\alpha),\, c_{ir}(\alpha)], \forall d_i(\alpha)\in[d_{il}(\alpha),\, d_{ir}(\alpha)]}} \frac{\sum_{i=1}^{L_1(\alpha)} a_i(\alpha)d_i(\alpha) + \sum_{i=L_1(\alpha)+1}^{n} a_i(\alpha)c_i(\alpha)}{\sum_{i=1}^{L_1(\alpha)} d_i(\alpha) + \sum_{i=L_1(\alpha)+1}^{n} c_i(\alpha)} \tag{4.65}$$

Observe that $a_i(\alpha)$ only appears in the numerator of (4.65); thus, $a_{il}(\alpha)$ should be used to calculate $y_{Ll}(\alpha)$, i.e.

$$y_{Ll}(\alpha) = \min_{\substack{\forall c_i(\alpha)\in[c_{il}(\alpha),c_{ir}(\alpha)] \\ \forall d_i(\alpha)\in[d_{il}(\alpha),d_{ir}(\alpha)]}} \frac{\sum_{i=1}^{L_1(\alpha)} a_{il}(\alpha)d_i(\alpha) + \sum_{i=L_1(\alpha)+1}^{n} a_{il}(\alpha)c_i(\alpha)}{\sum_{i=1}^{L_1(\alpha)} d_i(\alpha) + \sum_{i=L_1(\alpha)+1}^{n} c_i(\alpha)} \tag{4.66}$$

Following a similar line of reasoning, $y_{Lr}(\alpha)$, $y_{Rl}(\alpha)$ and $y_{Rr}(\alpha)$ can also be expressed as:

$$y_{Lr}(\alpha) = \max_{\substack{\forall c_i(\alpha)\in[c_{il}(\alpha),c_{ir}(\alpha)] \\ \forall d_i(\alpha)\in[d_{il}(\alpha),d_{ir}(\alpha)]}} \frac{\sum_{i=1}^{L_2(\alpha)} a_{ir}(\alpha)d_i(\alpha) + \sum_{i=L_2(\alpha)+1}^{n} a_{ir}(\alpha)c_i(\alpha)}{\sum_{i=1}^{L_2(\alpha)} d_i(\alpha) + \sum_{i=L_2(\alpha)+1}^{n} c_i(\alpha)} \tag{4.67}$$

$$y_{Rl}(\alpha) = \min_{\substack{\forall c_i(\alpha)\in[c_{il}(\alpha),c_{ir}(\alpha)] \\ \forall d_i(\alpha)\in[d_{il}(\alpha),d_{ir}(\alpha)]}} \frac{\sum_{i=1}^{R_1(\alpha)} b_{il}(\alpha)c_i(\alpha) + \sum_{i=R_1(\alpha)+1}^{n} b'_{il}(\alpha)d_i(\alpha)}{\sum_{i=1}^{R_1(\alpha)} c_i(\alpha) + \sum_{i=R_1(\alpha)+1}^{n} d_i(\alpha)} \tag{4.68}$$

$$y_{Rr}(\alpha) = \max_{\substack{\forall c_i(\alpha)\in[c_{il}(\alpha),c_{ir}(\alpha)] \\ \forall d_i(\alpha)\in[d_{il}(\alpha),d_{ir}(\alpha)]}} \frac{\sum_{i=1}^{R_2(\alpha)} b_{ir}(\alpha)c_i(\alpha) + \sum_{i=R_2(\alpha)+1}^{n} b_{ir}(\alpha)d_i(\alpha)}{\sum_{i=1}^{R_2(\alpha)} c_i(\alpha) + \sum_{i=R_2(\alpha)+1}^{n} d_i(\alpha)} \tag{4.69}$$

So far, only $a_i(\alpha)$ are fixed for $y_{Ll}(\alpha)$ and $y_{Lr}(\alpha)$, and $b_i(\alpha)$ are fixed for $y_{Rl}(\alpha)$ and $y_{Rr}(\alpha)$. As will be shown, it is also possible to fix $c_i(\alpha)$ and $d_i(\alpha)$ for $y_{Ll}(\alpha)$, $y_{Lr}(\alpha)$, $y_{Rl}(\alpha)$ and $y_{Rr}(\alpha)$; thus, there will be no need to enumerate and evaluate all of $\tilde{W}_i$'s embedded T1 FSs to find $\underline{Y}_{LWA}$ and $\overline{Y}_{LWA}$.

**Theorem 8** *It is true that:*

*(a) $y_{Ll}(\alpha)$ in (4.66) can be specified as*

$$y_{Ll}(\alpha) = \frac{\sum_{i=1}^{L_l^*(\alpha)} a_{il}(\alpha)d_{ir}(\alpha) + \sum_{i=L_l^*(\alpha)+1}^{n} a_{il}(\alpha)c_{il}(\alpha)}{\sum_{i=1}^{L_l^*(\alpha)} d_{ir}(\alpha) + \sum_{i=L_l^*(\alpha)+1}^{n} c_{il}(\alpha)}, \quad \alpha \in [0,1] \tag{4.70}$$

*(b) $y_{Lr}(\alpha)$ in (4.67) can be specified as*

$$y_{Lr}(\alpha) = \frac{\sum_{i=1}^{L_r^*(\alpha)} a_{ir}(\alpha)d_{il}(\alpha) + \sum_{i=L_r^*(\alpha)+1}^{n} a_{ir}(\alpha)c_{ir}(\alpha)}{\sum_{i=1}^{L_r^*(\alpha)} d_{il}(\alpha) + \sum_{i=L_r^*(\alpha)+1}^{n} c_{ir}(\alpha)}, \quad \alpha \in [0,h_{\min}] \tag{4.71}$$

*(c) $y_{Rl}(\alpha)$ in (4.68) can be specified as*

$$y_{Rl}(\alpha) = \frac{\sum_{i=1}^{R_l^*(\alpha)} b_{il}(\alpha)c_{ir}(\alpha) + \sum_{i=R_l^*(\alpha)+1}^{n} b_{il}(\alpha)d_{il}(\alpha)}{\sum_{i=1}^{R_l^*(\alpha)} c_{ir}(\alpha) + \sum_{i=R_l^*(\alpha)+1}^{n} d_{il}(\alpha)}, \quad \alpha \in [0,h_{\min}] \tag{4.72}$$

*(d) $y_{Rr}(\alpha)$ in (4.69) can be specified as*

$$y_{Rr}(\alpha) = \frac{\sum_{i=1}^{R_r^*(\alpha)} b_{ir}(\alpha)c_{il}(\alpha) + \sum_{i=R_r^*(\alpha)+1}^{n} b_{ir}(\alpha)d_{ir}(\alpha)}{\sum_{i=1}^{R_r^*(\alpha)} c_{il}(\alpha) + \sum_{i=R_r^*(\alpha)+1}^{n} d_{ir}(\alpha)}, \quad \alpha \in [0,1] \tag{4.73}$$

*In these equation $L_l^*(\alpha)$, $L_r^*(\alpha)$, $R_l^*(\alpha)$ and $R_r^*(\alpha)$ are switch points that are computed using KM or EKM Algorithms.* ■

*Proof:* Because the proofs of Parts (b)-(d) of Theorem 8 are quite similar to the proof of Part (a), only the proof of Part (a) is given here.

Let

$$g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j)) \equiv \frac{\sum_{i=1}^{L_1(\alpha_j)} a_{il}(\alpha_j) d_i(\alpha_j) + \sum_{i=L_1(\alpha_j)+1}^{n} a_{il}(\alpha_j) c_i(\alpha_j)}{\sum_{i=1}^{L_1(\alpha_j)} d_i(\alpha_j) + \sum_{i=L_1(\alpha_j)+1}^{n} c_i(\alpha_j)} \tag{4.74}$$

where $\mathbf{c}(\alpha_j) \equiv [c_{L_1(\alpha_j)+1}(\alpha_j), c_{L_1(\alpha_j)+2}(\alpha_j), \ldots, c_n(\alpha_j)]^T$, $\mathbf{d}(\alpha_j) \equiv [d_1(\alpha_j), d_2(\alpha_j), \ldots, d_{L_1(\alpha_j)}(\alpha_j)]^T$, $c_i(\alpha_j) \in [c_{il}(\alpha_j), c_{ir}(\alpha_j)]$ and $d_i(\alpha_j) \in [d_{il}(\alpha_j), d_{ir}(\alpha_j)]$. Then $y_{Ll}(\alpha_j)$ in (4.89) can be found by:

(1) Enumerating all possible combinations of $(c_{L_1(\alpha_j)+1}(\alpha_j), \ldots, c_n(\alpha_j), d_1(\alpha_j), \ldots, d_{L_1(\alpha_j)}(\alpha_j))$ for $c_i(\alpha_j) \in [c_{il}(\alpha_j), c_{ir}(\alpha_j)]$ and $d_i(\alpha_j) \in [d_{il}(\alpha_j), d_{ir}(\alpha_j)]$;

(2) Computing $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ in (4.74) for each combination; and,

(3) Setting $y_{Ll}(\alpha_j)$ to the smallest $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$.

Note that $L_1(\alpha_j)$, corresponding to the smallest $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ in Step (3), is $L_l^*(\alpha_j)$ in Theorem 8. In the following proof, the fact that there always exists such a $L_l^*(\alpha_j)$ is used.

(4.89) can be expressed as

$$y_{Ll}(\alpha_j) = \min_{\substack{\forall c_i(\alpha_j) \in [c_{il}(\alpha_j), c_{ir}(\alpha_j)] \\ \forall d_i(\alpha_j) \in [d_{il}(\alpha_j), d_{ir}(\alpha_j)]}} g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j)) \tag{4.75}$$

In [69] it is proved that $y_{Ll}(\alpha_j)$ has a value in the interval $[a_{L_l^*(\alpha_j),\,l}(\alpha_j),\ a_{L_l^*(\alpha_j)+1,\,l}(\alpha_j)]$; hence, at least one $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ must assume a value in this interval. In general there can be numerous $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ satisfying

$$a_{L_l^*(\alpha_j),\,l}(\alpha_j) \leq g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j)) \leq a_{L_l^*(\alpha_j)+1,\,l}(\alpha_j) \tag{4.76}$$

The remaining $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ must be larger than $a_{L_l^*(\alpha_j)+1,\,l}(\alpha_j)$, i.e. they must assume values in one of the intervals $(a_{L_l^*(\alpha_j)+1,\,l}(\alpha_j), a_{L_l^*(\alpha_j)+2,\,l}(\alpha_j)]$, $(a_{L_l^*(\alpha_j)+2,\,l}(\alpha_j), a_{L_l^*(\alpha_j)+3,\,l}(\alpha_j)]$, etc. Because the minimum of $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ is of interest, only those $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ satisfying (4.76) will be considered in this proof.

Next it is shown that when $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ achieves its minimum, (i) $d_i(\alpha_j) = d_{ir}(\alpha_j)$ for $i \leq L_l^*(\alpha_j)$, and (ii) $c_i(\alpha_j) = c_{il}(\alpha_j)$ for $i \geq L_l^*(\alpha_j) + 1$.

**i**. When $i \leq L_l^*(\alpha_j)$, it is straightforward to show that the derivative of $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ with respect to $d_i(\alpha_j)$, computed from (4.74), is

$$\frac{\partial g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))}{\partial d_i(\alpha_j)} = \frac{a_{il}(\alpha_j) - g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))}{\sum_{i=1}^{L_l^*(\alpha_j)} d_i(\alpha_j) + \sum_{i=L_l^*(\alpha_j)+1}^{n} c_i(\alpha_j)} \tag{4.77}$$

Using the left-hand side of (4.76), it follows that

$$-g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j)) \leq -a_{L_l^*(\alpha_j),\,l}(\alpha_j); \tag{4.78}$$

hence, in the numerator of (4.77),

$$a_{il}(\alpha_j) - g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j)) \leq a_{il}(\alpha_j) - a_{L_l^*(\alpha_j),\, l}(\alpha_j) \leq 0 \tag{4.79}$$

In obtaining the last inequality in (4.79) the fact that $a_{il}(\alpha_j) \leq a_{L_l^*(\alpha_j),\, l}(\alpha_j)$ when $i \leq L_l^*(\alpha_j)$ [due to the a priori increased-ordering of the $a_{il}(\alpha_j)$] was used. Consequently, using (4.79) in (4.77), it follows that

$$\frac{\partial g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))}{\partial d_i(\alpha_j)} \leq \frac{a_{il}(\alpha_j) - a_{L_l^*(\alpha_j),\, l}(\alpha_j)}{\sum_{i=1}^{L_l^*(\alpha_j)} d_i(\alpha_j) + \sum_{i=L_l^*(\alpha_j)+1}^{n} c_i(\alpha_j)} \leq 0 \tag{4.80}$$

(4.80) indicates that the first derivative of $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ with respect to $d_i(\alpha_j)$ ($i \leq L_l^*(\alpha_j)$) is negative; thus, $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ decreases when $d_i(\alpha_j)$ ($i \leq L_l^*(\alpha_j)$) increases. Consequently, the minimum of $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ must use the maximum possible $d_i(\alpha_j)$ for $i \leq L_l^*(\alpha_j)$, i.e. $d_i(\alpha_j) = d_{ir}(\alpha_j)$ for $i \leq L_l^*(\alpha_j)$, as stated in (4.70).

**ii**. When $i \geq L_l^*(\alpha_j) + 1$, it is straightforward to show that the derivative of $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ with respect to $c_i(\alpha_j)$, computed from (4.74), is

$$\frac{\partial g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))}{\partial c_i(\alpha_j)} = \frac{a_{il}(\alpha_j) - g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))}{\sum_{i=1}^{L_l^*(\alpha_j)} d_i(\alpha_j) + \sum_{i=L_l^*(\alpha_j)+1}^{n} c_i(\alpha_j)} \tag{4.81}$$

Using the right-hand side of (4.76), it follows that

$$-g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j)) \geq -a_{L_l^*(\alpha_j)+1,\, l}(\alpha_j) \tag{4.82}$$

Hence, in the numerator of (4.81),

$$a_{il}(\alpha_j) - g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j)) \geq a_{il}(\alpha_j) - a_{L_l^*(\alpha_j)+1,\,l}(\alpha_j) \geq 0 \tag{4.83}$$

In obtaining the last inequality in (4.83) the fact that $a_{il}(\alpha_j) \geq a_{L_l^*(\alpha_j)+1,\,l}(\alpha_j)$ when $i \geq L_l^*(\alpha_j)+1$ [due to the a priori increased-ordering of the $a_{il}(\alpha_j)$] was used. Consequently, using (4.83) in (4.81), it follows that

$$\frac{\partial g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))}{\partial c_i(\alpha_j)} \geq \frac{a_{il}(\alpha_j) - a_{L_l^*(\alpha_j)+1,\,l}(\alpha_j)}{\sum_{i=1}^{L_l^*(\alpha_j)} d_i(\alpha_j) + \sum_{i=L_l^*(\alpha_j)+1}^{n} c_i(\alpha_j)} \geq 0 \tag{4.84}$$

(4.84) indicates that the first derivative of $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ with respect to $c_i(\alpha_j)$ ($i \geq L_l^*(\alpha_j)+1$) is positive; thus, $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ decreases when $c_i(\alpha_j)$ ($i \geq L_l^*(\alpha_j)+1$) decreases. Consequently, the minimum of $g_{Ll}(\mathbf{c}(\alpha_j), \mathbf{d}(\alpha_j))$ must use the minimum possible $c_i(\alpha_j)$ for $i \geq L_l^*(\alpha_j)+1$, i.e. $c_i(\alpha_j) = c_{il}(\alpha_j)$ for $i \geq L_l^*(\alpha_j)+1$, as stated in (4.70). ■

Observe from (4.70), (4.73), and Figs. 4.5 and 4.6 that $y_{Ll}(\alpha)$ and $y_{Rr}(\alpha)$ only depend on the UMFs of $\tilde{X}_i$ and $\tilde{W}_i$, i.e., they are only computed from the corresponding $\alpha$-cuts on the UMFs of $\tilde{X}_i$ and $\tilde{W}_i$; so (this is an expressive equation),

$$\overline{Y}_{LWA} = \frac{\sum_{i=1}^{n} \overline{X}_i \overline{W}_i}{\sum_{i=1}^{n} \overline{W}_i}. \tag{4.85}$$

Because all $\overline{X}_i$ and $\overline{W}_i$ are normal T1 FSs, according to Theorem 5.3, $\overline{Y}_{LWA}$ is also normal.

Similarly, observe from (4.71), (4.72), and Figs. 4.5 and 4.6 that $y_{Lr}(\alpha)$ and $y_{Rl}(\alpha)$ only depend on the LMFs of $\tilde{X}_i$ and $\tilde{W}_i$; hence (this is an expressive equation),

$$\underline{Y}_{LWA} = \frac{\sum_{i=1}^{n} \underline{X}_i \underline{W}_i}{\sum_{i=1}^{n} \underline{W}_i}. \tag{4.86}$$

Unlike $\overline{Y}_{LWA}$, which is a normal T1 FS, the height of $\underline{Y}_{LWA}$ is $h_{\min}$, the minimum height of all $\underline{X}_i$ and $\underline{W}_i$.

### 4.4.3 LWA Algorithms

It has been shown in the previous subsection that computing $\tilde{Y}_{LWA}$ is equivalent to computing two FWAs, $\overline{Y}_{LWA}$ and $\underline{Y}_{LWA}$. To compute $\overline{Y}_{LWA}$:

1. Select appropriate $m$ $\alpha$-cuts for $\overline{Y}_{LWA}$ (e.g., divide [0, 1] into $m-1$ intervals and set $\alpha_j = (j-1)/(m-1)$, $j = 1, 2, ..., m$).
2. For each $\alpha_j$, find the corresponding $\alpha$-cuts $[a_{il}(\alpha_j), b_{ir}(\alpha_j)]$ and $[c_{il}(\alpha_j), d_{ir}(\alpha_j)]$ on $\overline{X}_i$ and $\overline{W}_i$ ($i = 1, ..., n$). Use a KM or EKM algorithm to find $y_{Ll}(\alpha_j)$ in (4.70) and $y_{Rr}(\alpha_j)$ in (4.73).
3. Connect all left-coordinates $(y_{Ll}(\alpha_j), \alpha_j)$ and all right-coordinates $(y_{Rr}(\alpha_j), \alpha_j)$ to form the T1 FS $\overline{Y}_{LWA}$.

To compute $\underline{Y}_{LWA}$:

1. Determine $h_{\underline{X}_i}$ and $h_{\underline{W}_i}$, $i = 1, \ldots, n$, and $h_{\min}$ in (4.49).

2. Select appropriate $p$ $\alpha$-cuts for $\underline{Y}_{LWA}$ (e.g., divide $[0, h_{\min}]$ into $p-1$ intervals and set $\alpha_j = h_{\min}(j-1)/(p-1)$, $j = 1, 2, ..., p$).

3. For each $\alpha_j$, find the corresponding $\alpha$-cuts $[a_{ir}(\alpha_j), b_{il}(\alpha_j)]$ and $[c_{ir}(\alpha_j), d_{il}(\alpha_j)]$ on $\underline{X}_i$ and $\underline{W}_i$. Use a KM or EKM algorithm to find $y_{Lr}(\alpha_j)$ in (4.71) and $y_{Rl}(\alpha_j)$ in (4.72).

4. Connect all left-coordinates $(y_{Lr}(\alpha_j), \alpha_j)$ and all right-coordinates $(y_{Rl}(\alpha_j), \alpha_j)$ to form the T1 FS $\underline{Y}_{LWA}$.

A flowchart for computing $\underline{Y}_{LWA}$ and $\overline{Y}_{LWA}$ is given in Fig. 4.7. For triangular or trapezoidal IT2 FSs, it is possible to reduce the number of $\alpha$-cuts for both $\underline{Y}_{LWA}$ and $\overline{Y}_{LWA}$ by choosing them only at *turning points*, i.e., points on the LMFs and UMFs of $X_i$ and $W_i$ $(i = 1, 2, ..., n)$ at which the slope of these functions changes.

**Example 14** *This is a continuation of Example 13 where each sub-criterion and weight is now assigned an FOU that is a 50% blurring of the T1 MF depicted in Fig. 4.2. The left half of each FOU (Fig. 4.8) has support on the $x$ ($w$)-axis given by the interval of real numbers* $[(\lambda-\delta)-.5\delta, (\lambda-\delta)+.5\delta]$ *and the right-half FOU (Fig. 4.8) has support on the $x$-axis given by the interval of real numbers* $[(\lambda+\delta)-.5\delta, (\lambda+\delta)+.5\delta]$. *The UMF is a triangle defined by the three points* $(\lambda-\delta-.5\delta, 0), (\lambda, 1), (\lambda+\delta+.5\delta, 0)$, *and the LMF is a triangle defined by the three points* $(\lambda-\delta+.5\delta, 0), (\lambda, 1), (\lambda+\delta-.5\delta, 0)$. *The resulting sub-criterion and weight FOUs are depicted in Figs. 4.9(a) and 4.9(b), respectively, and* $\tilde{Y}_{LWA}$ *is depicted in Fig. 4.9(c). Although* $\tilde{Y}_{LWA}$ *appears to be symmetrical, it is not.*

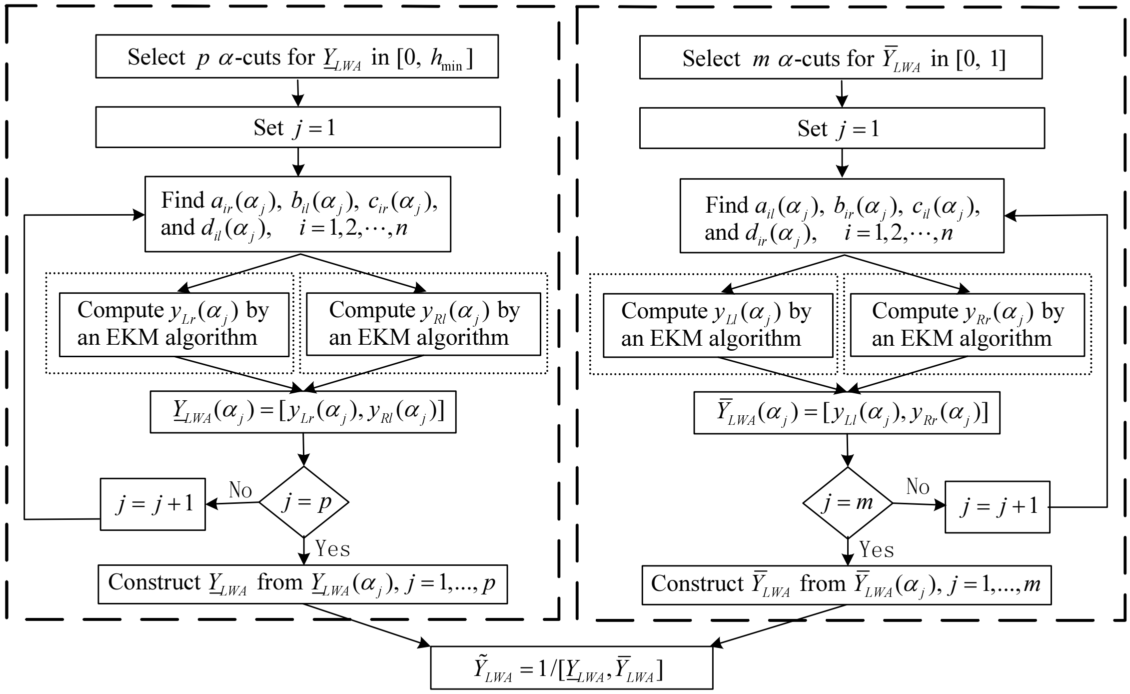



Fig. 4.7: A flowchart for computing the LWA [148].

*The support of the left-hand side of* $\tilde{Y}_{LWA}$ *is [0.85, 3.10] and the support of the right-hand side of* $\tilde{Y}_{LWA}$ *is [5.22, 7.56]; hence, the length of the support of the left-hand side of* $\tilde{Y}_{LWA}$ *is 2.25, whereas the length of the support of the right-hand side of* $\tilde{Y}_{LWA}$ *is 2.34. In addition, the centroid of* $\tilde{Y}_{LWA}$ *is computed using the EKM algorithms in Appendix A, and is* $C(\tilde{Y}_{LWA}) = [3.38, 4.96]$*, so that* $c(\tilde{Y}_{LWA}) = 4.17$.

*Comparing Figs. 4.9(c) and 4.3(c), observe that* $\tilde{Y}_{LWA}$ *is spread out over a larger range of values than is* $Y_{FWA}$*, reflecting the additional uncertainties in the LWA due to the blurring of sub-criteria and weights. This information can be used in future decisions.*

*Another way to interpret* $\tilde{Y}_{LWA}$ *is to associate values of* $y$ *that have the largest vertical intervals (i.e., primary memberships) with values of greatest uncertainty; hence, there is*

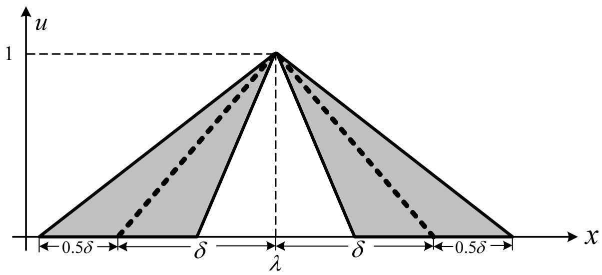
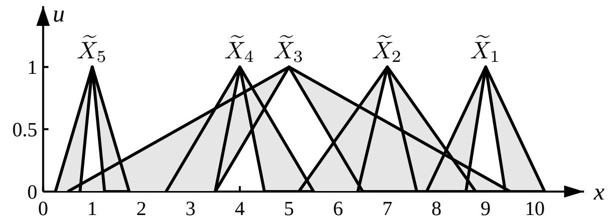



Fig. 4.8: Illustration of an IT2 FS used in Example 14. The dashed lines indicate corresponding T1 FS used in Example 13.

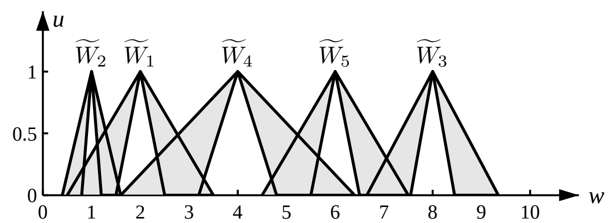
*no uncertainty at the three vertices of the UMF, and, e.g. for the right-half of* $\tilde{Y}_{LWA}$ *uncertainty increases from the apex of the UMF reaching its largest value at the right vertex of the LMF and then decreases to zero at the right vertex of the UMF.* ■

## 4.5 A Special Case of the LWA

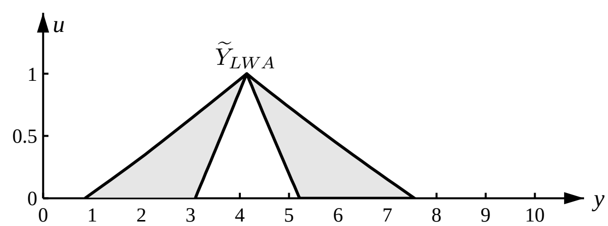
As shown in Fig. 4.1, there are many special cases of the general LWA introduced in the previous section, e.g., the weights and/or sub-criteria can be mixtures of numbers, intervals, T1 FSs, and IT2 FSs. The special case, where all weights are numbers and all sub-criteria are IT2 FSs, is of particular interest in this section because it is used in Chapter 8 for perceptual reasoning. Great simplifications of the LWA computations occur in this special case.




(a)




(b)




(c)

Fig. 4.9: Example 14: (a) $\tilde{X}_i$, (b) $\tilde{W}_i$, and, (c) $\tilde{Y}_{LWA}$.

Denote the crisp weights as $w_i$, $i = 1, \ldots, n$. Each $w_i$ can still be interpreted as an IT2 FS $\tilde{W}_i$, where

$$\mu_{\underline{W}_i}(w) = \mu_{\overline{W}_i}(w) = \begin{cases} 1, & w = w_i \\ 0, & w \neq w_i \end{cases} \tag{4.87}$$

i.e.,

$$c_{il}(\alpha) = c_{ir}(\alpha) = d_{il}(\alpha) = d_{ir}(\alpha) = w_i, \quad \alpha \in [0, 1] \tag{4.88}$$

Substituting (4.88) into Theorem 8, (4.70)-(4.73) are simplified to

$$y_{Ll}(\alpha) = \frac{\sum_{i=1}^n a_{il}(\alpha) w_i}{\sum_{i=1}^n w_i}, \quad \alpha \in [0, 1] \tag{4.89}$$

$$y_{Rr}(\alpha) = \frac{\sum_{i=1}^n b_{ir}(\alpha) w_i}{\sum_{i=1}^n w_i}, \quad \alpha \in [0, 1] \tag{4.90}$$

$$y_{Lr}(\alpha) = \frac{\sum_{i=1}^n a_{ir}(\alpha) w_i}{\sum_{i=1}^n w_i}, \quad \alpha \in [0, h_{\min}] \tag{4.91}$$

$$y_{Rl}(\alpha) = \frac{\sum_{i=1}^n b_{il}(\alpha) w_i}{\sum_{i=1}^n w_i}, \quad \alpha \in [0, h_{\min}] \tag{4.92}$$

where

$$h_{\min} = \min_{\forall i} h_{\underline{X}_i} \tag{4.93}$$

Note that (4.89)-(4.92) are arithmetic weighted averages, so they are computed directly without using KM or EKM algorithms.

**Example 15** *This is a continuation of Example 14, where the sub-criteria are the same as those shown in Fig. 4.9(a) and weights are crisp numbers* $\{w_i\}|_{i=1,\ldots,5} = \{2, 1, 8, 4, 6\}$, *i.e., they are the values of* $w$ *that occur at the apexes of* $\tilde{W}_i$ *shown in Fig. 4.9(b). The resulting* $\tilde{Y}_{LWA}$ *is depicted in Fig. 4.10. Observe that it is more compact than* $\tilde{Y}_{LWA}$ *in Fig. 4.9(c), which is intuitive, because in this example the weights have less uncertainties than those in Example 14. In addition, unlike the unsymmetrical* $\tilde{Y}_{LWA}$ *in Fig. 4.9(c),* $\tilde{Y}_{LWA}$ *in Fig. 4.10 is symmetrical*[4]. $C(\tilde{Y}_{LWA}) = [3.59, 4.69]$, *which is inside the centroid of* $\tilde{Y}_{LWA}$ *in Fig. 4.9(c).* ■

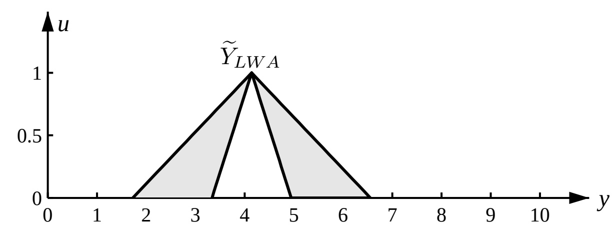



Fig. 4.10: $\tilde{Y}_{LWA}$ for Example 15.

## 4.6 Fuzzy Extensions of Ordered Weighted Averages (OWAs)

The ordered weighted average (OWA) operator [32, 71, 73, 131, 166, 172, 189–191] was proposed by Yager to aggregate experts' opinions in decision making.

[4]It can be shown that when all weights are crisp numbers, the resulting LWA from symmetrical $\tilde{X}_i$ is always symmetrical.

**Definition 24** *An OWA operator of dimension $n$ is a mapping $y_{OWA} : R^n \to R$, which has an associated set of weights $\mathbf{w} = \{w_1, \ldots, w_n\}$ for which $w_i \in [0, 1]$ and $\sum_{i=1}^{n} w_i = 1$, i.e.,*

$$y_{OWA} = \sum_{i=1}^{n} w_i x_{\sigma(i)} \tag{4.94}$$

*where $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is a permutation function such that $\{x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}\}$ are in* descending *order.* ■

Note that $y_{OWA}$ is a nonlinear operator due to the permutation of $x_i$. The most attractive feature of the OWA operator is that it can implement different aggregation operators by choosing the weights differently [32], e.g., by choosing $w_i = 1/n$ it implements the *mean* operator, by choosing $w_1 = 1$ and $w_i = 0$ $(i = 2, \ldots, n)$ it implements the *maximum* operator, and by choosing $w_i = 0$ $(i = 1, \ldots, n-1)$ and $w_n = 1$ it implements the *minimum* operator.

Yager's original OWA operator [166] considers only crisp numbers; however, experts may prefer to express their opinions in linguistic terms, which are modeled by FSs. Fuzzy extensions of OWAs [155, 189–191] are considered in this section.

### 4.6.1 Ordered Fuzzy Weighted Averages (OFWAs)

The ordered fuzzy weighted average (OFWA) was introduced by the authors in [155].

**Definition 25** *An OFWA is defined as*

$$Y_{OFWA} = \frac{\sum_{i=1}^{n} W_i X_{\sigma(i)}}{\sum_{i=1}^{n} W_i} \tag{4.95}$$

*where* $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ *is a permutation function such that* $\{X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(n)}\}$ *are in* descending *order.* ■

**Definition 26** *A group of T1 FSs* $\{X_i\}_{i=1}^{n}$ *are in* descending *order if* $X_i \succeq X_j$ *for* $\forall i < j$ *by a ranking method.* ■

Any T1 FS ranking method can be used to find $\sigma$. In this book, Yager's first method (see Section 3.3), which is a special case of the centroid-based ranking method, is used. Once $X_i$ are rank-ordered, $Y_{OFWA}$ is computed by an FWA.

### 4.6.2 Fuzzy Ordered Weighted Averages (FOWAs)

Zhou, et al. [190,191] introduced a fuzzy ordered weighed average (FOWA) operator which is different from the OFWA:

**Definition 27** *Given T1 FSs* $\{W_i\}_{i=1}^{n}$ *and* $\{X_i\}_{i=1}^{n}$*, an associated FOWA operator of dimension* $n$ *is a mapping:*

$$Y_{FOWA} : D_{X_1} \times \cdots \times D_{X_n} \to D_Y \tag{4.96}$$

*where*

$$\mu_{Y_{FOWA}}(y) = \sup_{\sum_{i=1}^{n} w_i' x_{\sigma(i)} = y} \min(\mu_{W_1}(w_1), \cdots, \mu_{W_n}(w_n), \mu_{X_1}(x_1), \cdots, \mu_{X_n}(x_n)) \quad (4.97)$$

*in which* $w_i' = \frac{w_i}{\sum_{j=1}^{n} w_j}$, *and* $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ *is a permutation function such that* $\{x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}\}$ *are in* descending *order.* ■

$\mu_{Y_{FOWA}}(y)$ can be understood from the Extension Principle (Section 4.3.1), i.e., first all combinations of $w_i$ and $x_i$ whose OWA is $y$ are found, and for the $j^{\text{th}}$ combination, the resulting $y_j$ has a membership grade $\mu(y_j)$ which is the minimum of the corresponding $\mu_{X_i}(x_i)$ and $\mu_{W_i}(w_i)$. Then, $\mu_{Y_{FOWA}}(y)$ is the maximum of all these $\mu(y_j)$.

$Y_{FOWA}$ can be computed efficiently using $\alpha$-cuts [189], similar to the way they are used in computing the FWA. Denote $Y_{FOWA}(\alpha) = [y_L'(\alpha), y_R'(\alpha)]$ and use the same notations for $\alpha$-cuts on $X_i$ and $W_i$ as those in (4.30) and (4.31). Then,

$$y_L'(\alpha) = \min_{\forall w_i(\alpha) \in [c_i(\alpha), d_i(\alpha)]} \frac{\sum_{i=1}^{n} a_{\sigma(i)}(\alpha) w_i(\alpha)}{\sum_{i=1}^{n} w_i(\alpha)} \quad (4.98)$$

$$y_R'(\alpha) = \max_{\forall w_i(\alpha) \in [c_i(\alpha), d_i(\alpha)]} \frac{\sum_{i=1}^{n} b_{\sigma(i)}(\alpha) w_i(\alpha)}{\sum_{i=1}^{n} w_i(\alpha)} \quad (4.99)$$

$y_L'(\alpha)$ and $y_R'(\alpha)$ can be computed using KM or EKM algorithms. Generally $\sigma$ is different for different $\alpha$ in (4.98) and (4.99), because for each $\alpha$ the $a_i(\alpha)$ or $b_i(\alpha)$ are ranked separately.

### 4.6.3 Comparison of OFWA and FOWA

Because $Y_{OFWA}$ uses the same $\sigma$ for all $\alpha \in [0,1]$ whereas $Y_{FOWA}$ computes the permutation function $\sigma$ for each $\alpha$ separately, generally the two approaches give different results, as illustrated in the following example.

**Example 16** *$X_i$ and $W_i$ shown in Figs. 4.11(a) and 4.11(b) are used in this example to illustrate the difference between $Y_{FOWA}$ and $Y_{OFWA}$. The former is shown as the dashed curve in Fig. 4.11(c). Because by the centroid-based ranking method, $X_1 \succ X_2 \succ X_3 \succ X_4 \succ X_5$, no reordering of $X_i$ is needed, and hence $Y_{OFWA}$ is computed as*

$$Y_{OFWA} = \frac{\sum_{i=1}^{5} W_i X_i}{\sum_{i=1}^{5} W_i} \tag{4.100}$$

*$Y_{OFWA}$ is shown as the solid curve in Fig. 4.11(c). Note that it is quite different from $Y_{FOWA}$. The difference is caused by the fact that the legs of $X_2$ cross the legs of $X_1$, $X_3$ and $X_4$, which causes the permutation function $\sigma$ to change as $\alpha$ increases. There will be no differences between $Y_{OFWA}$ and $Y_{FOWA}$ if $X_i$ do not have such kinds of intersections.*

■

### 4.6.4 Ordered Linguistic Weighted Averages (OLWAs)

An ordered linguistic weighed average (OLWA) was also proposed by the authors in [155].

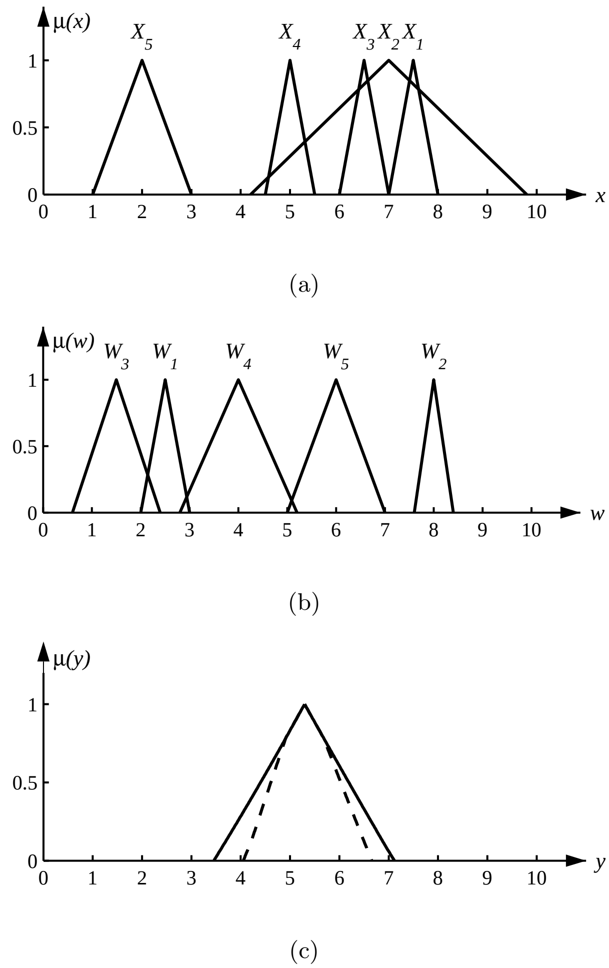



Fig. 4.11: Illustration of the difference between FOWA and OFWA for Example 16. (a) $X_i$, (b) $W_i$, and (c) $Y_{FOWA}$ (dashed curve) and $Y_{OFWA}$ (solid curve).

**Definition 28** *An OLWA is defined as*

$$\tilde{Y}_{OLWA} = \frac{\sum_{i=1}^{n} \tilde{W}_i \tilde{X}_{\sigma(i)}}{\sum_{i=1}^{n} \tilde{W}_i} \tag{4.101}$$

*where* $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ *is a permutation function such that* $\{\tilde{X}_{\sigma(1)}, \tilde{X}_{\sigma(2)}, \ldots, \tilde{X}_{\sigma(n)}\}$ *are in* descending *order.* ■

**Definition 29** *A group of IT2 FSs* $\{\tilde{X}_i\}_{i=1}^{n}$ *are in* descending *order if* $\tilde{X}_i \succeq \tilde{X}_j$ *for* $\forall i < j$ *by a ranking method.* ■

The LWA algorithm can also be used to compute the OLWA, except that the centroid-based ranking method must first be used to sort $\tilde{X}_i$ in descending order.

### 4.6.5 Linguistic Ordered Weighted Averages (LOWAs)

Zhou et al. [191] defined the IT2 fuzzy OWA, which is called a linguistic ordered weighted average (LOWA) in this dissertation, as:

**Definition 30** *Given IT2 FSs* $\{\tilde{W}_i\}_{i=1}^{n}$ *and* $\{\tilde{X}_i\}_{i=1}^{n}$*, an associated LOWA operator of dimension* $n$ *is a mapping:*

$$\tilde{Y}_{LOWA} : D_{\tilde{X}_1} \times \cdots \times D_{\tilde{X}_n} \to D_{\tilde{Y}} \tag{4.102}$$

*where*

$$\mu_{\tilde{Y}_{LOWA}}(y) = \bigcup_{\forall W_i^e, X_i^e} \left[ \sup_{\sum_{i=1}^{n} w_i' x_{\sigma(i)} = y} \min(\mu_{W_1^e}(w_1), \cdots, \mu_{W_n^e}(w_n), \mu_{X_1^e}(x_1), \cdots, \mu_{X_n^e}(x_n)) \right] \tag{4.103}$$

*in which* $W_i^e$ *and* $X_i^e$ *are embedded T1 FSs of* $\tilde{W}_i$ *and* $\tilde{X}_i$*, respectively,* $w_i' = \frac{w_i}{\sum_{j=1}^{n} w_j}$*, and* $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ *is a permutation function such that* $\{x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}\}$ *are in* descending *order.* ■

Comparing (4.103) with (4.97), observe that the bracketed term in (4.103) is an FOWA, and the LOWA is the union of all possible FOWAs computed from the embedded T1 FSs of $\tilde{X}_i$ and $\tilde{W}_i$. The Wavy Slice Representation Theorem is used implicitly in this definition.

$\tilde{Y}_{LOWA}$ can be computed efficiently using $\alpha$-cuts, similar to the way they were used in computing the LWA. Denote the $\alpha$-cut on the UMF of $\tilde{Y}_{LOWA}$ as $\overline{Y}_{LOWA}(\alpha) = [y'_{Ll}(\alpha), y'_{Rr}(\alpha)]$ for $\forall\alpha \in [0, 1]$, the $\alpha$-cut on the LMF of $\tilde{Y}_{LOWA}$ as $\underline{Y}_{LOWA}(\alpha) = [y'_{Lr}(\alpha), y'_{Rl}(\alpha)]$ for $\forall\alpha \in [0, h_{\min}]$, where $h_{\min}$ is defined in (4.49). Using the same notations for $\alpha$-cuts on $\tilde{X}_i$ and $\tilde{W}_i$ as in Section 4.4, it is easy to show that

$$y'_{Ll}(\alpha) = \min_{\forall w_i(\alpha) \in [c_{il}(\alpha), d_{ir}(\alpha)]} \frac{\sum_{i=1}^{n} a_{\sigma(i),l}(\alpha) w_i(\alpha)}{\sum_{i=1}^{n} w_i(\alpha)}, \quad \forall\alpha \in [0, 1] \tag{4.104}$$

$$y'_{Rr}(\alpha) = \max_{\forall w_i(\alpha) \in [c_{il}(\alpha), d_{ir}(\alpha)]} \frac{\sum_{i=1}^{n} b_{\sigma(i),r}(\alpha) w_i(\alpha)}{\sum_{i=1}^{n} w_i(\alpha)}, \quad \forall\alpha \in [0, 1] \tag{4.105}$$

$$y'_{Lr}(\alpha) = \min_{\forall w_i(\alpha)\in[c_{ir}(\alpha),d_{il}(\alpha)]} \frac{\sum_{i=1}^{n} a_{\sigma(i),r}(\alpha)w_i(\alpha)}{\sum_{i=1}^{n} w_i(\alpha)}, \quad \forall\alpha \in [0, h_{\min}] \tag{4.106}$$

$$y'_{Rl}(\alpha) = \max_{\forall w_i(\alpha)\in[c_{ir}(\alpha),d_{il}(\alpha)]} \frac{\sum_{i=1}^{n} b_{\sigma(i),l}(\alpha)w_i(\alpha)}{\sum_{i=1}^{n} w_i(\alpha)}, \quad \forall\alpha \in [0, h_{\min}] \tag{4.107}$$

$y'_{Ll}(\alpha)$, $y'_{Rr}(\alpha)$, $y'_{Lr}(\alpha)$ and $y'_{Rl}(\alpha)$ can be computed using KM or EKM algorithms. Because $\tilde{Y}_{LOWA}$ computes the permutation function $\sigma$ for each $\alpha$ separately, generally $\sigma$ is different for different $\alpha$.

### 4.6.6 Comparison of OLWA and LOWA

Again, because $\tilde{Y}_{OLWA}$ uses the same $\sigma$ for all $\alpha \in [0, 1]$ whereas $\tilde{Y}_{LOWA}$ computes the permutation function $\sigma$ for each $\alpha$ separately, generally the two approaches give different results, as illustrated in the following example.

**Example 17** *$\tilde{X}_i$ and $\tilde{W}_i$ shown in Figs. 4.12(a) and 4.12(b) are used in this example to illustrate the difference between $\tilde{Y}_{LOWA}$ and $\tilde{Y}_{OLWA}$. The former is shown as the dashed curve in Fig. 4.12(c). Because by the centroid-based ranking method, $\tilde{X}_1 \succ \tilde{X}_2 \succ \tilde{X}_3 \succ \tilde{X}_4 \succ \tilde{X}_5$, no reordering of $\tilde{X}_i$ is needed, and hence $\tilde{Y}_{OLWA}$ is computed as*

$$\tilde{Y}_{OLWA} = \frac{\sum_{i=1}^{5} \tilde{W}_i \tilde{X}_i}{\sum_{i=1}^{5} \tilde{W}_i} \tag{4.108}$$

*$\tilde{Y}_{OLWA}$ is shown as the solid curve in Fig. 4.12(c). Note that it is quite different from $\tilde{Y}_{LOWA}$. The difference is caused by the fact that the legs of $\tilde{X}_2$ cross the legs of $\tilde{X}_1$,*

*$\tilde{X}_3$ and $\tilde{X}_4$, since the permutation function $\sigma$ changes as $\alpha$ increases. There will be no differences between $\tilde{Y}_{OLWA}$ and $\tilde{Y}_{LOWA}$ if $\tilde{X}_i$ do not have such kinds of intersections.* ■

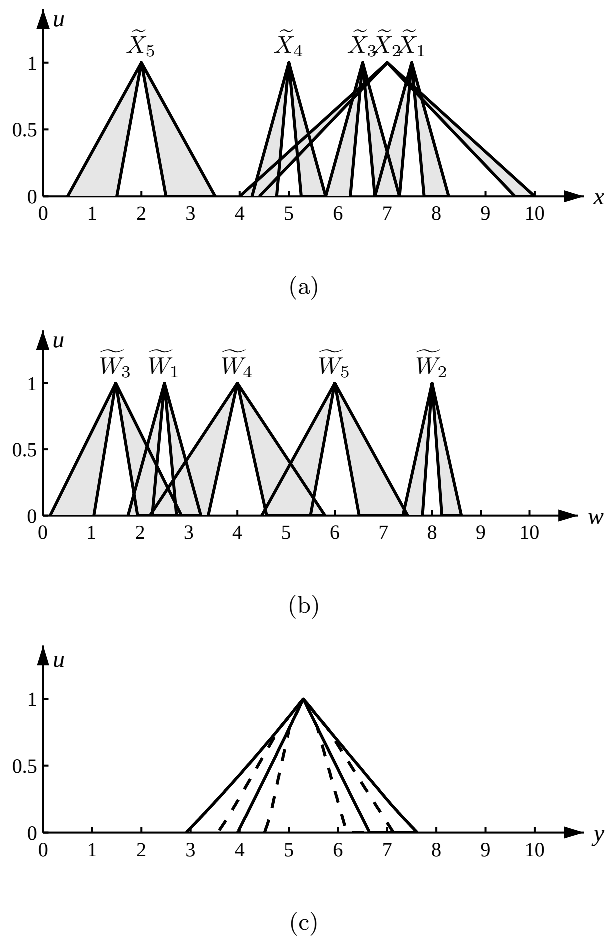



Fig. 4.12: Illustration of the difference between LOWA and OLWA for Example 17. (a) $\tilde{X}_i$, (b) $\tilde{W}_i$, and (c) $\tilde{Y}_{LOWA}$ (dashed curve) and $\tilde{Y}_{OLWA}$ (solid curve).

### 4.6.7 Comments

The FOWA and LOWA have been derived by considering each $\alpha$-cut separately, whereas the OFWA and OLWA have been derived by considering each sub-criterion as a whole.

Sometimes the two approaches give different results. Then, a natural question is: which approach should be used in practice?

We believe that it is more intuitive to consider an FS in its entirety during ranking of FSs. To the best of our knowledge, all ranking methods based on $\alpha$-cuts deduce a single number to represent each FS and then sort these numbers to obtain the ranks of the FSs. Each of these numbers is computed based only on the FS under consideration, i.e., no $\alpha$-cuts on other FSs to be ranked are considered. Because in OFWA and OLWA the FSs are first ranked and then the WAs are computed, they coincide with our "FS in its entirety" intuition, and hence they are preferred in this dissertation. Interestingly, this "FS in its entirety" intuition was also used implicitly in developing the *linguistic ordered weighted averaging* [43] and the *uncertain linguistic ordered weighted averaging* [163].

Finally, note that the OFWA can be viewed as a special case of the FWA, and the OLWA can be viewed as a special case of the LWA.

# Chapter 5

# Perceptual Computer for MADM: The Missile Evaluation Application

## 5.1 Introduction

The missile evaluation application, an MADM problem introduced in Example 1, is completed in this chapter. It was used because it has already appeared in several publications [14–16, 93] and also because the evaluations range from numbers to words.

As has been introduced in Example 1, a contractor has to decide which of three companies ($A$, $B$ or $C$) is going to get the final mass production contract for a missile system based on the criteria, sub-criteria, weights and inputs given in Table 1.1. Observe that:

1. The major criteria are not equally weighted, but instead are weighted using fuzzy numbers[1] (T1 FSs[2], as depicted in Fig. 5.1 and Table 5.1) in the following order of importance: *tactics, advancement, economy, technology* and *maintenance*. These weightings were established ahead of time by the contractor and not by the companies.

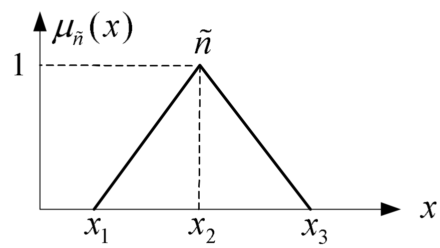



Fig. 5.1: Membership function for a fuzzy number $\tilde{n}$ (see Table 5.1).

2. *Tactics* has seven sub-criteria, *technology* and *maintenance* each have five sub-criteria, and, *economy* and *advancement* each have three sub-criteria; hence, there are 23 sub-criteria all of which were established ahead of time by the contractor and not by the companies.

[1]It is common practice to use a tilde over-mark to denote a fuzzy number that is modeled using a T1 FS. Even though it is also common practice to use such a tilde over-mark to denote an IT2 FS, we shall not change this common practice for a fuzzy number in this chapter. Instead, we shall indicate in the text when the fuzzy number $\tilde{n}$ is modeled either as a T1 or as an IT2 FS.

[2]Even though, in Table 5.1, these fuzzy numbers are called *triangular fuzzy numbers*, observe that $\tilde{1}$ is a left-shoulder MF and $\tilde{9}$ is a right-shoulder MF.

Table 5.1: Triangular fuzzy numbers and their corresponding MFs [14].

| **Triangular fuzzy numbers** | $(x_1, x_2, x_3)$ |
|---|---|
| $\tilde{1}$ | (1, 1, 2) |
| $\tilde{2}$ | (1, 2, 3) |
| $\tilde{3}$ | (2, 3, 4) |
| $\tilde{4}$ | (3, 4, 5) |
| $\tilde{5}$ | (4, 5, 6) |
| $\tilde{6}$ | (5, 6, 7) |
| $\tilde{7}$ | (6, 7, 8) |
| $\tilde{8}$ | (7, 8, 9) |
| $\tilde{9}$ | (8, 9, 9) |

3. All of the sub-criteria are weighted using fuzzy numbers. These weightings have also been established ahead of time by the contractor and not by the companies, and have been established separately within each of the five criteria and not simultaneously across all of the 23 sub-criteria.

4. The performance evaluations for all 23 sub-criteria are shown for the three companies, and are either numbers or words. It is assumed that each company designed, built and tested a small number of its missiles after which they were able to fill in the numerical performance scores. It is not clear how the linguistic scores were obtained, so it is speculated that the contractor provided them based on other evidence and perhaps on some subjective rules.

5. How to aggregate all of this data seems like a daunting task, especially since it involves numbers, fuzzy numbers for the weights, and words.

6. Finally, we believe there should be an uncertainty band for each numerical score because the numbers correspond to measurements of physical properties obtained

from an ensemble of test missiles. Those bands have not been provided, but will be assumed in this chapter to inject some additional realism into this application.

The missile evaluation problem can also be summarized by Fig. 5.2. It is very clear from this figure that this is a multi-criteria and two-level decision making problem. At the first level each of the three companies[3] is evaluated for its performance on five criteria: *tactics*, *technology*, *maintenance*, *economy* and *advancement*. The lines emanating from each of the companies to these criteria indicate these evaluations, each of which involves a number of important (but not shown) sub-criteria and their weighted aggregations that are described below. The second level in this hierarchical decision making problem involves a weighted aggregation of the five criteria for each of the three companies.

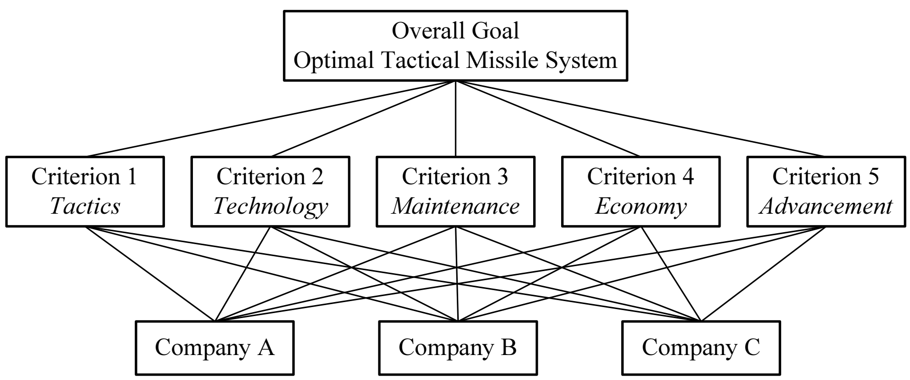



Fig. 5.2: Structure of evaluating competing tactical missile systems from three companies [93].

[3]The terms company and system are used interchangeably in this chapter.

## 5.2 A Per-C Approach for Missile Evaluation

Recall that the Per-C has three components: encoder, CWW engine and decoder. When Per-C is used for the missile evaluation problem each of these components must be considered.

### 5.2.1 Encoder

In this application, mixed data are used — crisp numbers, T1 fuzzy numbers and words. The codebook contains the crisp numbers, the T1 fuzzy numbers with their associated T1 FS models (Fig. 5.1 and Table 5.1), and the words and their IT2 FS models.

To ensure that LWAs would not be unduly-influenced by large numbers, all of the Table 1.1 numbers were mapped into [0, 10]. Let $x_1$, $x_2$ and $x_3$ denote the raw numbers for Companies A, B and C, respectively. For the 13 sub-criteria whose inputs are numbers, those raw numbers were transformed into:

$$x_i \to x_i' = \frac{10x_i}{\max(x_1, x_2, x_3)}. \tag{5.1}$$

Examining Table 1.1, observe that the words used for the remaining 10 sub-criteria are: *poor, low, average, good, very good* and *high.* Because this application is being used merely to illustrate how a Per-C can be used for missile system evaluation, and we do not have access to domain experts, interval-point data were not collected for these words in the context of this application. Instead, the codebook shown in Fig. 2.12 is used. Unfortunately, none of the six words that are actually used in this application appear in that codebook. So, each word was mapped into a word that was felt to be a synonym for it. The mappings are:

$$\left.\begin{array}{lcl} Poor & \rightarrow & Small \\ Low & \rightarrow & Low\ Amount \\ Average & \rightarrow & Medium \\ Good & \rightarrow & Large \\ Very\ Good & \rightarrow & Very\ Large \\ High & \rightarrow & High\ Amount \end{array}\right\} \tag{5.2}$$

The IT2 FS models of the six words are shown in Fig. 5.3.

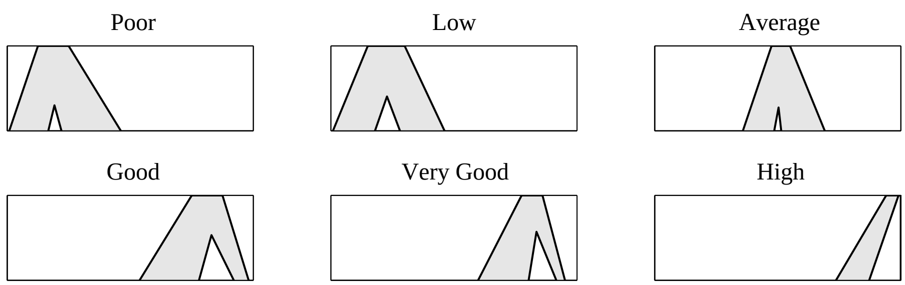



Fig. 5.3: IT2 FS models for the six words used in missile evaluation.

Observe from Table 1.1 that some sub-criteria may have a positive connotation and others may have a negative connotation. The following six sub-criteria have a negative connotation: *flight height*[4], *missile scale*[5], *reaction time, operation condition requirement, system cost* and *material limitation*. The first three sub-criteria have numbers as their

[4]The lower the flight height the better, because it is then more difficult for a missile to be detected by radar.

[5]A smaller missile is also harder to detect by radar.

inputs. For them, in addition to (5.1), a further step is needed to covert a large $x'$ into a low score and a small $x'$ into a high score:

$$x'_i \rightarrow x''_i = 10 - x'. \tag{5.3}$$

**Example 18** *Suppose that* $x_1 = 3$*,* $x_2 = 4$ *and* $x_3 = 5$*. Then when these numbers are mapped into* $[0, 10]$ *using (5.1), they become:* $x'_1 = 10(3/5) = 6$*,* $x'_2 = 10(4/5) = 8$ *and* $x'_3 = 10(5/5) = 10$*. On the other hand, for a sub-criterion with negative connotation, these numbers become:* $x''_1 = 10 - x'_1 = 4$*,* $x''_2 = 10 - x'_2 = 2$ *and* $x''_3 = 10 - x'_3 = 0$*.* ■

For the other three sub-criteria with a negative connotation (operation condition requirement, system cost, and material limitation), *antonyms* [61,102,126,184] are used for the words in (5.2), i.e.,

$$\mu_{10-A}(x) = \mu_A(10 - x), \quad \forall x \tag{5.4}$$

where $10 - A$ is the antonym of a T1 FS $A$, and 10 is the right end of the domain of all FSs used in this chapter. The definition in (5.4) can easily be extended to IT2 FSs, i.e.,

$$\mu_{10-\tilde{A}}(x) = \mu_{\tilde{A}}(10 - x), \quad \forall x \tag{5.5}$$

where $10 - \tilde{A}$ is the antonym of an IT2 FS $\tilde{A}$. Because an IT2 FS is completely characterized by its LMF and UMF, each of which is a T1 FS, $\mu_{10-\tilde{A}}(x)$ in (5.5) is obtained by applying (5.4) to both $\underline{A}$ and $\overline{A}$.

*Comment:* Using these mappings, the highest score for the 17 sub-criteria that have a positive connotation is always assigned the value 10, and the lowest score for the six sub-criteria that have a negative connotation is also always assigned the value 10. What if such scores are not actually "good" scores? Assigning it our highest value does not then seem to be correct.

In this type of procurement competition the contractor often sets specifications on numerical performance sub-criteria. Unfortunately, such specifications do not appear in any of the published articles about this application, so we have had to do the best we can without them. If, for example, the contractor had set a specification for *reliability* as 85%, then no company should get a 10. A different kind of normalization would then have to be used. ■

### 5.2.2 CWW Engine

NWAs are used as our CWW engine. Each of the major criteria had an NWA computed for it. Examining Table 1.1, observe that the NWA for *Tactics* ($\tilde{Y}_1$) is a FWA (because the weights are T1 FSs and the sub-criteria evaluations are numbers), whereas the NWAs for *Technology* ($\tilde{Y}_2$), *Maintenance* ($\tilde{Y}_3$), *Economy* ($\tilde{Y}_4$) and *Advancement* ($\tilde{Y}_5$) are LWAs (because at least one sub-criterion evaluation is a word modeled by an IT2 FS). More specifically:

$$\tilde{Y}_1 = \frac{\sum_{i=1}^{7} X_i W_i}{\sum_{i=1}^{7} W_i} \quad (5.6)$$

$$\tilde{Y}_2 = \frac{\sum_{i=8}^{12} \tilde{X}_i \tilde{W}_i}{\sum_{i=8}^{12} \tilde{W}_i} \quad (5.7)$$

$$\tilde{Y}_3 = \frac{\sum_{i=13}^{17} \tilde{X}_i \tilde{W}_i}{\sum_{i=13}^{17} \tilde{W}_i} \quad (5.8)$$

$$\tilde{Y}_4 = \frac{\sum_{i=18}^{20} \tilde{X}_i \tilde{W}_i}{\sum_{i=18}^{20} \tilde{W}_i} \quad (5.9)$$

$$\tilde{Y}_5 = \frac{\sum_{i=21}^{23} \tilde{X}_i \tilde{W}_i}{\sum_{i=21}^{23} \tilde{W}_i} \quad (5.10)$$

These six NWAs are then aggregated by another NWA to obtain the overall performance, $\tilde{Y}$, as follows:

$$\tilde{Y} = \frac{\tilde{9}\tilde{Y}_1 + \tilde{3}\tilde{Y}_2 + \tilde{1}\tilde{Y}_3 + \tilde{5}\tilde{Y}_4 + \tilde{7}\tilde{Y}_5}{\tilde{9} + \tilde{3} + \tilde{1} + \tilde{5} + \tilde{7}} \quad (5.11)$$

As a reminder to the reader, when $i = 2, 8, 9, 13, 18, 20$, (5.3) or the antonyms of the corresponding IT2 FSs must be used.

### 5.2.3 Decoder

The decoder computes ranking, similarity and centroid. Rankings of the three companies are obtained for the six LWA FOUs in (5.6)-(5.11) using the centroid-based ranking method [144]. The average centroids for Companies A, B and C are represented in all figures in Section 5.3 by $*$, $\diamond$ and $\circ$, respectively.

Similarity is computed only for the three companies' overall performances $\tilde{Y}$ so that one can observe how similar the overall performances are for them.

Centroids are also computed for the three companies' $\tilde{Y}$, and provide a measure of uncertainty for each company's overall ranking, since $\tilde{Y}$ has propagated both numerical and linguistic uncertainties through their calculations.

## 5.3 Examples

This section contains examples that illustrate the missile evaluation results for different scenarios. Example 19 uses the data that are in Table 1.1 as is. Examples 21-23 use intervals for all numerical values, i.e., in Example 21 each numerical value $x$ is changed to the interval $[x - 10\%x, x + 10\%x]$ for all three companies, in Example 22 each numerical value $x$ is changed to the interval $[x - 20\%x, x + 20\%x]$ for all three companies, and in Example 23 $x$ is changed to $[x - 30\%x, x + 30\%x]$ for Company B but is only changed to $[x - 5\%x, x + 5\%x]$ for Companies A and C. Using more realistic data intervals instead of numbers is something that was mentioned earlier at the end of Section 5.1 in Item 6.

**Example 19** *As just mentioned, this example uses the data that are in Table 1.1 as is. In all figures, System A is represented by the solid curve, System B is represented by the dashed curve, and System C is represented by the dotted curve. In order to simplify the notation in the figures, the notations $\tilde{Y}_{Aj}$, $\tilde{Y}_{Bj}$ and $\tilde{Y}_{Cj}$ are used for aggregated results* *for Criterion j and for Companies A, B and C, respectively. The caption of each figure indicates the name of Criterion j ($j = 1, 2, \ldots, 5$) and the numbering of the criteria corresponds to their numbering in Table 1.1.*

*FOUs for Tactics, Technology, Maintenance, Economy and Advancement are depicted in Figs. 5.4(a)-5.4(e), respectively. FOUs for Overall Performance are depicted in Fig. 5.4(f). Observe from Fig. 5.4(f) that System B is the best. This is because System B ranks first in Maintenance, Economy and Advancement and by significant amounts. Although it ranks last for Tactics and Technology its FOUs for these two criteria are very close to those of System's A and C. Not only is $FOU(\tilde{Y}_B)$ visually well to the right of the other two FOUs in Fig. 5.4(f), but its center of centroid (which is on the horizontal axis) is also well to the right of those for Companies A and C. So, based on ranking alone, Company B would be declared the winner.*

*Table 5.2 summarizes the similarities between $\tilde{Y}_A$, $\tilde{Y}_B$ and $\tilde{Y}_C$. Observe that $\tilde{Y}_B$ is not very similar to either $\tilde{Y}_A$ or $\tilde{Y}_C$, so choosing Company B as the winner is further reinforced, i.e. it is not a close call.*

Table 5.2: Similarities of $\tilde{Y}$ in Example 19 for the three companies.

| Company | $\tilde{Y}_A$ | $\tilde{Y}_B$ | $\tilde{Y}_C$ |
|---|---|---|---|
| $\tilde{Y}_A$ | 1 | 0.072 | 0.209 |
| $\tilde{Y}_B$ | 0.072 | 1 | 0.407 |
| $\tilde{Y}_C$ | 0.209 | 0.407 | 1 |

*Finally, the centroids of $\tilde{Y}_A$, $\tilde{Y}_B$ and $\tilde{Y}_C$ (Table 5.3) are $C_A = [5.889, 6.648]$, $C_B = [7.586, 8.148]$ and $C_C = [6.966, 7.651]$; the numerical rankings (computed from these centroids) are $c_A = 6.268$, $c_B = 7.867$ and $c_C = 7.308$; and, the half-lengths of each centroid*

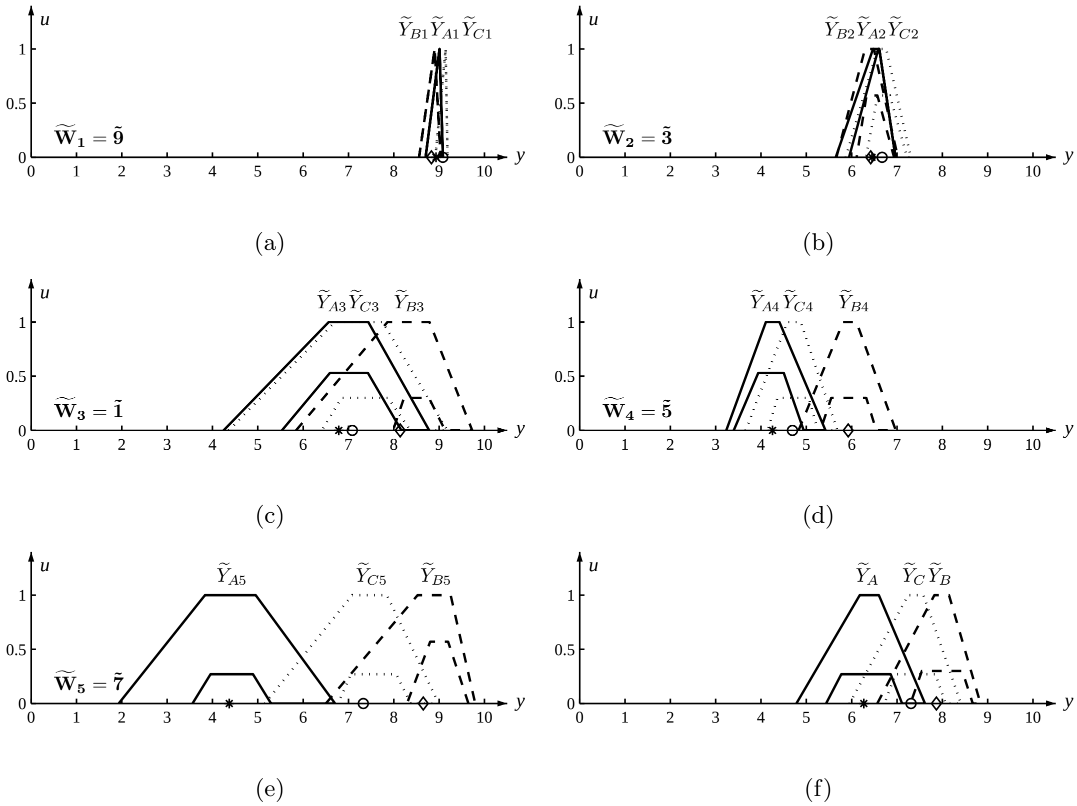



Fig. 5.4: Example 19: Aggregation results for (a) Criterion 1: *Tactics*; (b) Criterion 2: *Technology*; (c) Criterion 3: *Maintenance*; (d) Criterion 4: *Economy*; (e) Criterion 5: *Advancement*; and, (f) Overall performances of the three systems. The average centroids for Companies A, B and C are shown in all figures by $*$, $\diamond$ and $\circ$, respectively. The FOUs in (b)-(f) are not filled in so that the three IT2 FSs can be distinguished more easily.

*are* $l_A/2 = 0.380$*,* $l_B/2 = 0.281$ *and* $l_C/2 = 0.342$*. One way to use these half-lengths is to summarize the rankings as:* $r_A = 6.268 \pm 0.380$*,* $r_B = 7.867 \pm 0.281$ *and* $r_C = 7.308 \pm 0.342$*. Note that the centroids can also be interpreted as ranking-bands and that there is very little overlap of these bands in this example, and it is only between Systems B and C. All these results are summarized in Table 5.3.*

Table 5.3: Centroids, centers of centroid and ranking bands of $\tilde{Y}$ for various uncertainties.

| Company | | 0% for all three companies Example 19 | ±10% for all three companies Example 21 | ±20% for all three companies Example 22 | ±30% for Company B and ±5% for Companies A and C Example 23 |
|---|---|---|---|---|---|
| A | $C_A$ | [5.889, 6.648] | [5.763, 6.698] | [5.553, 6.651] | [5.843, 6.697] |
| | $c_A$ | 6.268 | 6.230 | 6.102 | 6.270 |
| | $r_A$ | 6.268±0.380 | 6.230±0.467 | 6.102±0.549 | 6.270±0.427 |
| B | $C_B$ | [7.586, 8.148] | [7.356, 8.067] | [7.141, 8.014] | [6.898, 7.928] |
| | $c_B$ | 7.867 | 7.712 | 7.578 | 7.413 |
| | $r_B$ | 7.867±0.281 | 7.712±0.356 | 7.578±0.437 | 7.413±0.515 |
| C | $C_C$ | [6.966, 7.651] | [6.828, 7.708] | [6.621, 7.659] | [6.902, 7.687] |
| | $c_C$ | 7.308 | 7.268 | 7.140 | 7.294 |
| | $r_C$ | 7.308±0.342 | 7.268±0.440 | 7.140±0.519 | 7.294±0.393 |

*Not only does Company B have the largest ranking but it also has the smallest uncertainty band about that ranking and* $\tilde{Y}_B$ *is not very similar to either* $\tilde{Y}_A$ *or* $\tilde{Y}_C$*. Choosing Company B as the winner seems the right thing to do.* ■

In reality though there are uncertainties about each of the numbers in Table 1.1, as noted in Harvard Business Essentials ( [23], pp. 80): "*... point estimates are almost always wrong. Worse, point estimates give the impression of certainty when there is none. What the decision maker needs is a range of possible outcomes for each uncertainty, as determined by experienced and knowledgeable informants.*"

In the remaining examples uncertainty intervals are assigned to each of these numbers, i.e.,

$$x_i \to [x_i - v\%x_i, \min(x_i + v\%x_i, \max(x_1, x_2, x_3))], \quad i = 1, 2, 3 \tag{5.12}$$

so that the effects of such uncertainties on the overall performances of the three companies can be studied. Note that $\max(x_1, x_2, x_3)$ is used as an upper limit so that the converted number is not larger than 10 [see (5.14)]. The specific choice(s) made for $v$ are explained in the examples.

For the 10 sub-criteria that have a positive connotation, (5.1) is used for the two end-points in (5.12), i.e.,

$$x_i - v\%x_i \to \frac{10(x_i - v\%x_i)}{\max(x_1, x_2, x_3)} \tag{5.13}$$

$$\min(x_i + v\%x_i, \max(x_1, x_2, x_3)) \to \frac{10\min(x_i + v\%x_i, \max(x_1, x_2, x_3))}{\max(x_1, x_2, x_3)} \tag{5.14}$$

and for the three sub-criteria that have a negative connotation, the mappings are

$$\begin{aligned} &[x_i - v\%x_i, \min(x_i + v\%x_i, \max(x_1, x_2, x_3))] \\ &\to \left[10 - \frac{10\min(x_i + v\%x_i, \max(x_1, x_2, x_3))}{\max(x_1, x_2, x_3)}, 10 - \frac{10(x_i - v\%x_i)}{\max(x_1, x_2, x_3)}\right] \end{aligned} \tag{5.15}$$

**Example 20** *As in Example 18, suppose that* $x_1 = 3$*,* $x_2 = 4$ *and* $x_3 = 5$*. Let* $v = 10$*, so that* $x_1 \rightarrow [2.7, 3.3]$*,* $x_2 \rightarrow [3.6, 4.4]$ *and* $x_3 \rightarrow [4.5, 5]$*. For a sub-criterion with positive connotation, using (5.12)-(5.14), one finds that*

$$[2.7, 3.3] \rightarrow [10(2.7/5), 10(3.3/5)] = [5.4, 6.6]$$

$$[3.6, 4.4] \rightarrow [10(3.6/5), 10(4.4/5)] = [7.2, 8.8]$$

$$[4.5, 5] \rightarrow [10(4.5/5), 10(5/5)] = [9, 10];$$

*and, for a sub-criteria with negative connotation, using (5.12) and (5.15), one finds that*

$$[2.7, 3.3] \rightarrow [10 - 6.6, 10 - 5.4] = [3.4, 4.6]$$

$$[3.6, 4.4] \rightarrow [10 - 8.8, 10 - 7.2] = [1.2, 2.8]$$

$$[4.5, 5] \rightarrow [10 - 10, 10 - 9] = [0, 1]. \quad \blacksquare$$

**Example 21** *In this example each numerical value* $x$ *in Table 1.1 is changed by the same percentage amount to the interval* $[x - 10\%x, x + 10\%x]$*. We are interested to learn if such uncertainty intervals change the rankings of the three companies. FOUs for Tactics, Technology, Maintenance, Economy and Advancement are depicted in Figs. 5.5(a)-5.5(e), respectively. The overall performances of the three systems are depicted in Fig. 5.5(f). System B still appears to be the winning system.*

*Comparing the results in Fig. 5.5 with their counterparts in Fig. 5.4, observe that generally the FOUs have larger support. Particularly, the T1 FSs in Fig. 5.4(a) are*

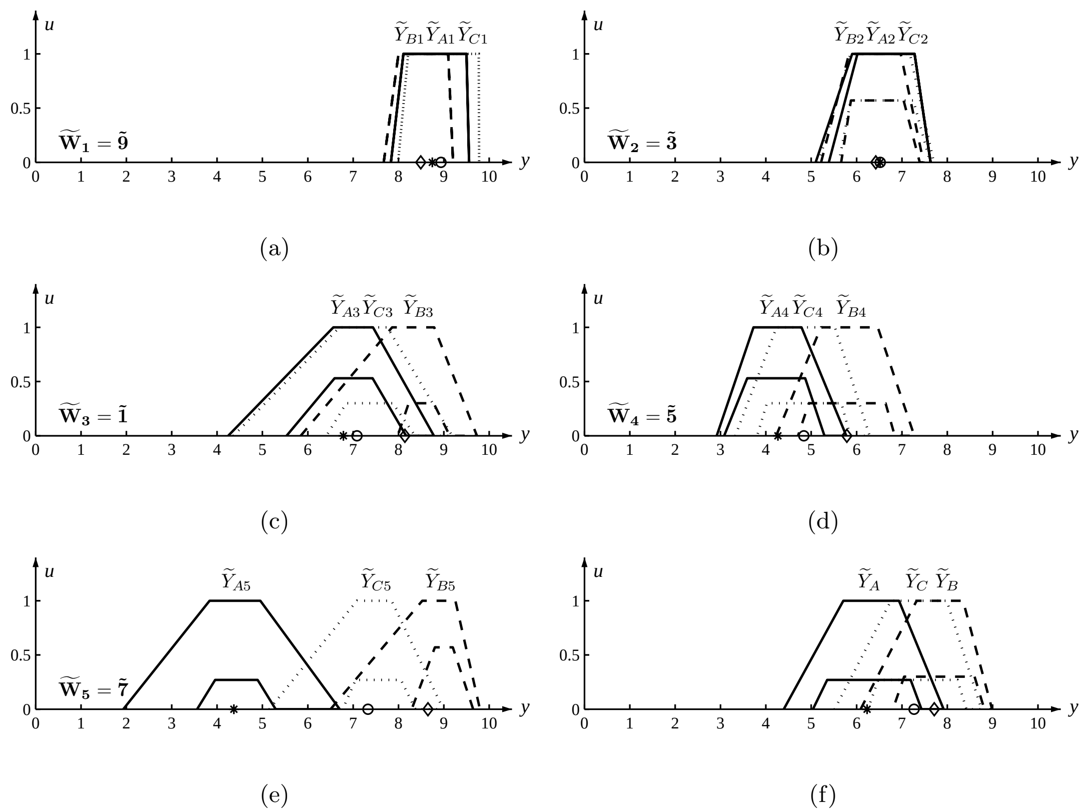



Fig. 5.5: Example 21: Aggregation results for (a) Criterion 1: *Tactics*; (b) Criterion 2: *Technology*; (c) Criterion 3: *Maintenance*; (d) Criterion 4: *Economy*; (e) Criterion 5: *Advancement*; and, (f) Overall performances of the three systems. The average centroids for Companies A, B and C are shown in all figures by $*$, $\diamond$ and $\circ$, respectively.

*triangular whereas the T1 FSs in Fig. 5.5(a) are trapezoidal. This is because in Fig. 5.4(a) the inputs to the sub-criteria are numbers and the weights are triangular T1 FSs, and hence the* $\alpha = 1$ $\alpha$*-cut on* $\tilde{Y}_{A1}$ *(*$\tilde{Y}_{B1}$*, or* $\tilde{Y}_{C1}$*) is an AWA, whereas in Fig. 5.5(a) the inputs to the sub-criteria are intervals and the weights are triangular T1 FSs, and hence the* $\alpha = 1$ $\alpha$*-cut on* $\tilde{Y}_{A1}$ *(*$\tilde{Y}_{B1}$*, or* $\tilde{Y}_{C1}$*) is an IWA.*

*Table 5.4 summarizes the similarities between* $\tilde{Y}_A$*,* $\tilde{Y}_B$ *and* $\tilde{Y}_C$*. Observe that* $\tilde{Y}_C$ *is much more similar to* $\tilde{Y}_B$ *in this example than it was in Example 19. Consequently, one may be less certain about choosing Company B as the winner when there is* $\pm 10\%$ *uncertainty on all of the numbers in Table 1.1 than when there is no uncertainty on those numbers.*

Table 5.4: Similarities of $\tilde{Y}$ in Example 21 for the three companies.

| Company | $\tilde{Y}_A$ | $\tilde{Y}_B$ | $\tilde{Y}_C$ |
|---|---|---|---|
| $\tilde{Y}_A$ | 1 | 0.189 | 0.371 |
| $\tilde{Y}_B$ | 0.189 | 1 | 0.637 |
| $\tilde{Y}_C$ | 0.371 | 0.637 | 1 |

*The centroids, centers of centroids and the ranking bands of* $\tilde{Y}_A$*,* $\tilde{Y}_B$ *and* $\tilde{Y}_C$ *are shown in Table 5.3. Observe that not only does Company B still have the largest ranking but it still has the smallest uncertainty band about that ranking. However, when there is* $\pm 10\%$ *uncertainty on all of the numbers in Table 1.1, not only do the numerical rankings for the three companies shift to the left (to lower values) but the uncertainty bands about those rankings increase. The overlap between the ranking bands of Systems B and C also increases.*

*In short, even though Company B could still be declared the winner, one is less certain about doing this when there is ±10% uncertainty on all of the numbers in Table 1.1.* ■

**Example 22** *In this example each numerical value $x$ in Table 1.1 is changed by the same percentage amount to the interval $[x-20\%x, x+20\%x]$. This is twice as much uncertainty as in Example 21. We are again interested to learn if such uncertainty intervals change the rankings of the three companies.*

*FOUs for Tactics, Technology, Maintenance, Economy and Advancement are depicted in Figs. 5.6(a)-5.6(e), respectively. The overall performances of the three systems are depicted in Fig. 5.6(f). System B still appears to be the winning system, but declaring Company B the winner is now more problematic as is demonstrated next.*

*Table 5.5 summarizes the similarities between $\tilde{Y}_A$, $\tilde{Y}_B$ and $\tilde{Y}_C$. Observe that $\tilde{Y}_C$ is even more similar to $\tilde{Y}_B$ in this example than in Example 21, so one may be even less certain about choosing Company B as the winner when there is ±20% uncertainty on all of the numbers in Table 1.1 than when there is no uncertainty on those numbers.*

Table 5.5: Similarities of $\tilde{Y}$ in Example 22 for the three companies.

| Company | $\tilde{Y}_A$ | $\tilde{Y}_B$ | $\tilde{Y}_C$ |
|---|---|---|---|
| $\tilde{Y}_A$ | 1 | 0.295 | 0.466 |
| $\tilde{Y}_B$ | 0.295 | 1 | 0.707 |
| $\tilde{Y}_C$ | 0.466 | 0.707 | 1 |

*The centroids, centers of centroids and the ranking bands of $\tilde{Y}_A$, $\tilde{Y}_B$ and $\tilde{Y}_C$ are shown in Table 5.3. Observe that not only does Company B still have the largest ranking, but it still has the smallest uncertainty band about that ranking. Notice however that when*

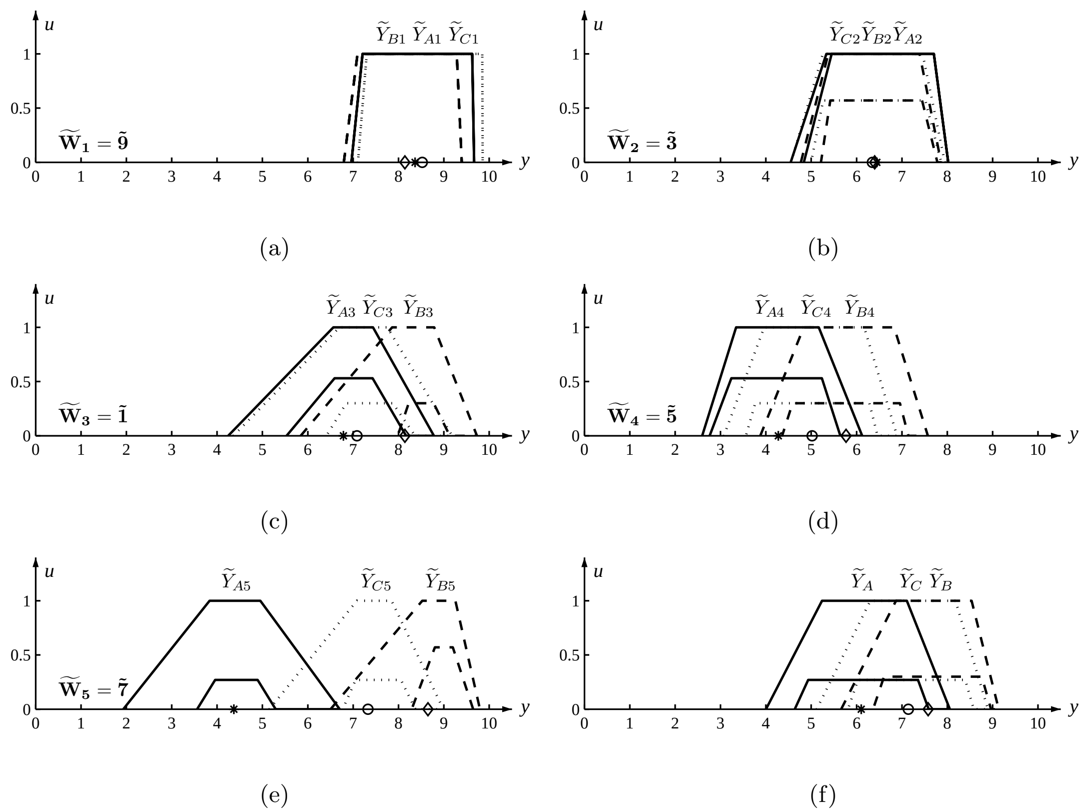



Fig. 5.6: Example 22: Aggregation results for (a) Criterion 1: *Tactics*; (b) Criterion 2: *Technology*; (c) Criterion 3: *Maintenance*; (d) Criterion 4: *Economy*; (e) Criterion 5: *Advancement*; and, (f) Overall performances of the three systems. The average centroids for Companies A, B and C are shown in all figures by $*$, $\diamond$ and $\circ$, respectively.

*there is ±20% uncertainty on all of the numbers in Table 1.1, not only do the numerical rankings for the three companies shift further to the left (to even lower values) than in the ±10% case, but the uncertainty bands about those rankings also increase, and in addition to Systems B and C, the ranking-bands of Systems A and C also have some overlap.*

*In short, even though Company B still could be declared the winner, one is even less certain about this when there is ±20% uncertainty on all of the numbers in Table 1.1. The rankings are getting uncomfortably close to each other for the three companies, making a declaration of a clear winner problematic.* ■

**Example 23** *In our previous examples, Company B always seems to be ahead of Companies A and C. In this example each numerical value $x$ in Table 1.1 is changed by the same percentage amount to the interval $[x - 30\%x, x + 30\%x]$ for Company B, but is only changed by $[x - 5\%x, x + 5\%x]$ for Companies A and C. Perhaps the tighter uncertainty bands for Companies A and C will change the results.*

*FOUs for Tactics, Technology, Maintenance, Economy and Advancement are depicted in Figs. 5.7(a)-5.7(e), respectively. The overall performances of the three systems are depicted in Fig. 5.7(f). Observe that the FOU of $\tilde{Y}_C$ is completely inside the FOU of $\tilde{Y}_B$; so, it is difficult to declare System B the winner.*

*Table 5.6 summarizes the similarities between $\tilde{Y}_A$, $\tilde{Y}_B$ and $\tilde{Y}_C$. Observe that $\tilde{Y}_C$ is again more similar to $\tilde{Y}_B$ in this example than in Example 19, so one may be less certain about choosing Company B as the winner when there is $\pm 30\%$ uncertainty on all of the* *numbers about Company B whereas there is only ±5% uncertainty on all of the numbers about Companies A and C.*

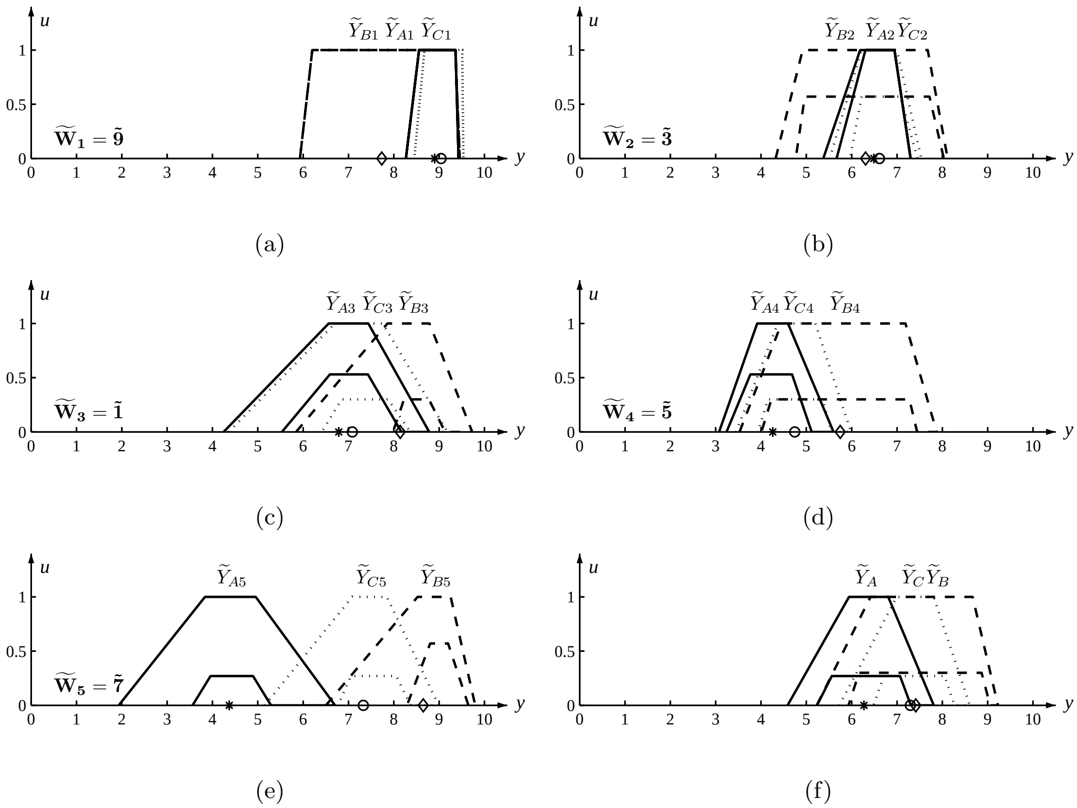



Fig. 5.7: Example 23: Aggregation results for (a) Criterion 1: *Tactics*; (b) Criterion 2: *Technology*; (c) Criterion 3: *Maintenance*; (d) Criterion 4: *Economy*; (e) Criterion 5: *Advancement*; and, (f) Overall performances of the three systems. The average centroids for Companies A, B and C are shown in all figures by $*$, $\diamond$ and $\circ$, respectively.

*The centroids, centers of centroids and the ranking bands of* $\tilde{Y}_A$, $\tilde{Y}_B$ *and* $\tilde{Y}_C$ *are shown in Table 5.3. Now the ranking bands for Systems B and C overlap a lot, and even the ranking bands for Systems A and B overlap significantly, which is why it is difficult to declare System B the winner.*

Table 5.6: Similarities of $\tilde{Y}$ in Example 23 for the three companies.

| Company | $\tilde{Y}_A$ | $\tilde{Y}_B$ | $\tilde{Y}_C$ |
|---|---|---|---|
| $\tilde{Y}_A$ | 1 | 0.380 | 0.303 |
| $\tilde{Y}_B$ | 0.380 | 1 | 0.567 |
| $\tilde{Y}_C$ | 0.303 | 0.567 | 1 |

*This example clearly demonstrates that providing only average values for the sub-criteria in Table 1.1 can lead to misleading conclusions. Uncertainty bands about those average values can change conclusions dramatically.* ■

## 5.4 Comparisons with Previous Approaches

In this section the results from our Per-C approach are compared with results from four previous approaches on the missile evaluation problem.

### 5.4.1 Comparison with Mon et al.'s Approach

Mon et al. [93] appear to be the first to work on "*performance evaluation and optimal design of weapon systems [as] multiple criteria decision making problems*" using FSs. They use fuzzy numbers to indicate the relative strength of the elements in the hierarchy, and build a *fuzzy judgment matrix* through comparison of performance scores. They begin with numerical scores for all of the sub-criteria (no words are used by them), after which they:

1. Aggregate the sub-criteria scores by first converting each number into either 1 if some (contractor's) sub-criterion is satisfied or 0.5 if the sub-criterion is not satisfied,

after which all these crisp numbers are added (implying that they are given the same weight), and the sum is then treated as a fuzzy number. This is done for each of the five criteria and for each of the three companies. These fuzzy numbers are put into a $3 \times 5$ *fuzzy judgment matrix*.

2. Assign fuzzy importance weights to each of the five criteria.

3. Compute a *total fuzzy judgment matrix* by multiplying each fuzzy number in the fuzzy judgment matrix by its respective fuzzy importance weight using $\alpha$-cuts. For each value of $\alpha \in [0, 1]$ the result is an interval of numbers $[a_l(\alpha), a_r(\alpha)]$. The result for each value of $\alpha$ is a $3 \times 5$ $\alpha$-cut *judgment matrix* $A_\alpha$.

4. Estimate a *degree of satisfaction*, $\hat{a}(\alpha)$, for each element of $A_\alpha$ by taking a linear combination of $a_l(\alpha)$ and $a_r(\alpha)$, i. e. $\hat{a}(\alpha) = \lambda a_l(\alpha) + (1 - \lambda) a_r(\alpha)$ in which $\lambda \in [0, 1]$ is an *index of optimism*. The resulting matrix $\hat{A}_\alpha$ is called a *crisp judgment matrix*.

5. Normalize each row of $\hat{A}_\alpha$ by dividing all of the row's elements by its largest element.

6. Compute an *entropy number* for each row (company).

7. Normalize the three entropies by dividing each entropy number by the sum of the three entropy numbers, leading to three entropy weights, one for each company. This is done for sampled values of $\alpha \in [0, 1]$ and specified values of $\lambda$.

8. Choose the winning company as the one whose entropy weight is the largest.

When $\lambda = 0$ or $1/2$ or $1$ the decision maker is called *pessimistic*, *moderate* or *optimistic*, respectively. The pessimistic decision maker uses worst values for $\hat{a}(\alpha)$, namely $a_l(\alpha)$; the optimistic decision maker uses best values for $\hat{a}(\alpha)$, namely $a_r(\alpha)$; and, the moderate decision maker uses the arithmetic average $[a_l(\alpha) + a_r(\alpha)]/2$.

The shortcomings of Mon et al.'s approach are: 1) when a sub-criterion is not satisfied, a company is assigned a score 0.5 no matter how far away it is from that sub-criterion, i.e., useful information is lost; 2) each sub-criterion is weighted the same in order to compute a sum for each criterion, which is counter-intuitive; and, 3) the crisp sum for each criterion is fuzzified to a fuzzy number to incorporate uncertainty, whereas the uncertainties should be considered at the beginning of the aggregation and be propagated. All these three short-comings are overcome in our Per-C approach, i.e., the numerical score for each sub-criterion is computed based on how far away a system's performance is from the best performance, the FOUs for the words are modeled a priori from a survey, and the sub-criteria are weighted.

### 5.4.2 Comparison with Chen's Approaches

Chen [14] uses a different approach in which he begins with tables of numerical or linguistic scores for the sub-criteria, after which he:

1. Ranks the three companies (1, 2 or 3) for each sub-criterion,

2. Adds the rankings for all of a criterion's sub-criteria[6].

[6]Note that it is possible to rank linguistic scores (which may be why ranking is used by Chen [14]), e.g., *higher* cost is worse than *average* cost, and *good* range is better than *average* range.

3. Treats the aggregated crisp ranking as a fuzzy number, leading to a $3 \times 5$ *fuzzy rank score matrix* (FRSM).

4. Assigns fuzzy importance weights to each of the five criteria.

5. Multiplies each element of the FRSM by its associated criterion's fuzzy importance weight leading to another fuzzy number.

6. Adds each company's five fuzzy numbers, the result being a *fuzzy ranking number.*

7. Defuzzifies each of the company's fuzzy ranking number by computing its centroid.

8. The winning company is the one with the smallest defuzzified value[7].

In a different chapter, Chen [15] further modifies Mon et al's method, i.e. he:

1. Assigns a *fuzzy importance number* to each of the sub-criteria, thereby not only overcoming the objection to Mon et al.'s method that every sub-criterion is weighted the same, but also introducing some uncertainty into the importance of each sub-criterion.

2. Ranks the sub-criteria using fuzzy ranking ($\tilde{1}$, $\tilde{2}$ or $\tilde{3}$), where again there can be ties in which two or all of the companies receive the same ranking, but this can only occur when either numerical or linguistic scores are the same.

3. Computes a *fuzzy score* for each company by multiplying each sub-criterion's fuzzy importance weight by its fuzzy ranking, using $\alpha$-cuts.

[7]The smallest value is the winner because 1 is of higher rank than is 2 or 3.

4. Uses Mon et al.'s [93] index of optimism idea to reach a final decision. The result (for each $\alpha$-cut) is an index of optimism for each of the three companies that is normalized by dividing it by the sum of the three indexes of optimism.

5. The winning company is the one with the largest normalized index of optimism.

Stepping back from the details of [14, 15], it is observed that Chen is also losing information by first ranking the sub-criteria and by then processing the ranked sub-criteria by using each criterion's fuzzy importance weight. Additionally, the fuzzy scores are not normalized, so that each score may be unduly influenced by one large fuzzy-number, something that was first pointed out in [16]. Both short-comings are overcome in our Per-C approach, e.g., the numerical score for each sub-criterion is computed based on how far away a system's performance is from the best performance, and a novel weighted average is used in the aggregation.

### 5.4.3 Comparison with Cheng's Approach

Cheng [16] proposes to overcome the normalization deficiency in [15] by using *fuzzy ratio scales* to indicate the relative importance of the five criteria and the three missile systems' scores for them (the scales are $\tilde{1}$, $\tilde{3}$, $\tilde{5}$, $\tilde{7}$, $\tilde{9}$, where $\tilde{1}$ denotes almost equal importance, $\tilde{3}$ denotes moderate importance of one over another, $\tilde{5}$ denotes strong importance, $\tilde{7}$ denotes very strong importance, and $\tilde{9}$ denotes extreme importance); however, he does not explain how each missile system's scores for the criteria are obtained. And, because he is still performing ranking before the other processing, he is also losing information.

In summary, the short-comings of the four previous approaches are: loss of information by pre-processing, inability to process a broad range of mixed data from numbers to words, and, inability to provide uncertainty information about the final results. All of these shortcomings are overcome in our use of a Per-C for the missile evaluation problem, as demonstrated in Section 5.2.

## 5.5 Conclusions

In this chapter it has been shown how the Per-C can be applied to a missile evaluation problem, which is a hierarchical MADM problem, and is representative of procurement judgment applications. Distinguishing features of our approach are:

1. No pre-processing of the sub-criteria scores (e.g., by ranking) is done and therefore no information is lost.

2. A wide range of mixed data can be used, from numbers to words. By not having to convert words into a pre-processed rank, information is again not lost.

3. Uncertainties about the sub-criteria scores as well as their weights flow through all NWA calculations, so that our final company performance FOUs not only contain ranking and similarity information but also uncertainty information. No other existing method contains such uncertainty information.

Although we have explained how the Per-C can be applied to a hierarchical MADM problem in the context of a specific application, the methodology of this Per-C is quite general and it can be applied to similar procurement applications.

# Chapter 6

# Extract Rules from Data: Linguistic Summarization

## 6.1 Introduction

The rapid progress of information technology has made huge amounts of data accessible to people, e.g., a single seismic survey of the BP Valhall field generates 7 TB of data [29], and the 2nd Palomar Observatory Sky Survey (POSS-II) conducted by the California Institute of Technology resulted in about 3,000 digital images of $23,040 \times 23,040$ 16-bit pixels each, totalling over 3 terabytes of data [31]. Unfortunately, the raw data alone are often hardly understandable and do not provide knowledge, i.e., frequently people face the "data rich, information poor" dilemma. Data mining approaches to automatically summarize the data and output human-friendly information are highly desirable. According to Mani and

Maybury [74], "*summarization is the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks).*" Particularly, data summarization in this dissertation means to [95] "*grasp and briefly describe trends and characteristics appearing in a dataset, without doing (explicit) manual 'record-by-record' analysis.*"

Statistics can be used to compute the mean, median, variance, etc, of a dataset and hence can be viewed as a simple form of summarization; however, as pointed out by Yager [165], "*summarization would be especially practicable if it could provide us with summaries that are not as terse as the mean, as well as treating the summarization of nonnumeric data.*" This suggests that linguistic summarization of databases, which outputs rules/patterns like "*most wells with high oil production also have high water production*" or "*IF oil production in a well is high, THEN water production of the well is also high*," is more favorable, because it can provide richer and more easily understandable information, and it also copes well with nonnumeric data.

There are many approaches for linguistic summarization of databases [25, 27, 107, 108] and time series [17, 54]. In this chapter we will follow the FS based approach introduced by Yager [165, 167–169] and advanced by many others [34, 54, 55, 95, 107, 123]. Most of these authors focus on T1 FSs. Niewiadomski et al. [95, 96, 98–100] are to date the only ones working on linguistic summarization using IT2 FSs; however, their results have limitations, and some of them are incorrect, as shown in Sections 6.2 and 6.3. So, a new IT2 FSs based linguistic summarization approach is proposed in this chapter.

## 6.2 Linguistic Summarization Using T1 FSs: Traditional Approach

Yager [165, 167–169] was the first to study linguistic summarization of databases using T1 FSs and proposed two canonical forms; however, he only considers the case where one summarizer is used. George and Srikanth [34] extended Yager's approach to more than one summarizer, and their approach is briefly reviewed in this section. For easy reference, our most frequently used symbols are collected in Table 6.1.

Table 6.1: Explanations of the symbols used in this chapter. $n = 1, 2, \ldots, N$ and $m = 1, 2, \ldots, M$.

| Symbol | Meaning | Example |
|---|---|---|
| $\mathbb{D}$ | The complete database | The fracture optimization dataset |
| $\mathbb{Y}$ | The set of all objects in the database | All wells in the fracture optimization dataset |
| $M$ | Number of objects in $\mathbb{Y}$ | 85 (wells) for the fracture optimization dataset |
| $y_m$ | The $m^{\text{th}}$ object in the database | The $m^{\text{th}}$ well in the fracture optimization dataset |
| $v_n$ | Name of the $n^{\text{th}}$ attribute | #Stages |
| $\mathbb{X}_n$ | The domain of $v_n$ | [3, 17] for #Stages |
| $\mathbb{V}$ | A set of all attribute names | <#Stages, lengthPerf, #Holes, Sand, Slurry, Pad, Oil> |
| $v_n^m$ | Value of the $n^{\text{th}}$ attribute for $y_m$ | #Stages of the $m^{\text{th}}$ well |
| $\mathbf{d}_m$ | A complete record related to $y_m$ with values assigned to all attributes in $\mathbb{V}$ | <4, 232, 290, 161000, 5715, 1183, 5278> for the first well in the dataset |
| $S_n$ | Summarizer | High oil production, low water production, etc |
| $Q$ | Quantifier | Most, about half, more than 100, etc |
| $w_g$ | Qualifier | High oil production, low water production, etc |
| $T$ | Truth level | Any value in $[0, 1]$ |
| $T_3$ | Degree of covering | Any value in $[0, 1]$ |
| $T_4$ | Degree of appropriateness | Any value in $[0, 1]$ |
| $T_c$ | Degree of sufficient coverage | Any value in $[0, 1]$ |
| $T_u$ | Degree of usefulness | Any value in $[0, 1]$ |
| $T_o$ | Degree of outlier | Any value in $[0, 1]$ |

### 6.2.1 Two Canonical Forms

Define a set of $M$ objects $\mathbb{Y} = \{y_1, y_2, \ldots, y_M\}$ and a set of $N$ attribute names $\mathbb{V} = \{v_1, v_2, \ldots, v_N\}$. Let $\mathbb{X}_n$ $(n = 1, 2, \ldots, N)$ be the domain of $v_n$. Then, $v_n(y_m) \equiv v_n^m \in \mathbb{X}_n$ is the value of the $n^{\text{th}}$ attribute for the $m^{\text{th}}$ object $(m = 1, 2, \ldots, M)$. Hence, the database $\mathbb{D}$, which collects information about elements from $\mathbb{Y}$, is in the form of

$$\begin{aligned} \mathbb{D} &= \{< v_1^1, v_2^1, \ldots, v_N^1 >, < v_1^2, v_2^2, \ldots, v_N^2 >, \cdots, < v_1^M, v_2^M, \ldots, v_N^M >\} \\ &\equiv \{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_M\} \end{aligned} \tag{6.1}$$

where $\mathbf{d}_m =< v_1^m, v_2^m, \ldots, v_N^m >$ is a complete record about object $y_m$.

For example, for the fracture dataset used in Section 6.6, there are 85 wells ($M = 85$), and hence $\mathbb{Y}$={Well1, Well2, ...,Well85}. Each well has seven attributes ($N = 7$), and $\mathbb{V}$=<#Stages, lengthPerf, #Holes, Sand, Slurry, Pad, Oil>. For #Stages, its value ranges from three to 17; so, its domain $\mathbb{X}_1 = [3, 17]$. Well1 has four stages, 232 feet of perforation and 290 holes in completion, and 161,000 barrels of sand, 5,715 barrels of slurry and 1,183 barrels of pad were injected in fracturing, and it produced 5,278 barrels of oil during the first 180 days after fracturing; so, the complete record for Well1 is $\mathbf{d}_1$=<4, 232, 290, 161,000, 5,715, 1,183, 5,278>.

Let $S_n$ be a label associated with a T1 FS in $\mathbb{X}_n$. Define

$$\mathbf{S} = \{S_1, S_2, \ldots, S_N\} \tag{6.2}$$

where

$$\mu_{\mathbf{S}}(\mathbf{d}_m) \equiv \min\{\mu_{S_1}(v_1^m), \cdots, \mu_{S_N}(v_N^m)\}, \quad m = 1, 2, \ldots, M \tag{6.3}$$

The two canonical forms of T1 FS linguistic summarization considered by George and Srikanth [34] are:

1. *$Q$ objects from $\mathbb{Y}$ are/have $\{S_1, S_2, \ldots, S_N\}$ [$T$]*, where $Q$ is a *linguistic quantifier* modeled by a T1 FS, e.g., about half, most, etc; $S_n$ ($n = 1, 2, \ldots, N$) is a *summarizer*, e.g., high oil production, low water production, etc; $T \in [0, 1]$ is a quality measure for the summary called the *truth level*. It describes how well the dataset fits the summary. Generally, $T$ increases as more data support the summary. $T$ is computed as

$$T = \mu_Q\left(\frac{r}{R}\right) \tag{6.4}$$

in which

$$r = \sum_{m=1}^{M} \mu_{\mathbf{S}}(\mathbf{d}_m) \tag{6.5}$$

$$R = \begin{cases} M, & Q \text{ is a relative quantifier (e.g., most)} \\ 1, & Q \text{ is an absolute quantifier (e.g., more than 100)} \end{cases} \tag{6.6}$$

i.e., we first compute $r/R$ as the portion (when a relative $Q$ is used) or the number (when an absolute $Q$ is used) of $\mathbb{Y}$ that have properties $\{S_1, S_2, \ldots, S_N\}$, and then map it to a truth level in $[0,1]$ according to the quantifier $Q$. Note that FS operations are used in the calculation, i.e., a well can partially fit the summary. This is different from the crisp case, where a well must either fit the summary completely (i.e., with degree 1) or not fit the summary at all (i.e., with degree 0).

An example of such a summary is:

$$\underbrace{\text{Most}}_{Q}\ \underbrace{\text{wells}}_{\mathbb{Y}}\ \text{have}\ \underbrace{\text{high oil production}}_{S_1}\ \text{and}\ \underbrace{\text{high water production}}_{S_2}\ [\underbrace{0.6}_{T}]$$

In this example $Q$ is a relative quantifier.

2. *$Q$ objects from $\mathbb{Y}$ being/with $w_g$ are/have* $\{S_1, \ldots, S_{g-1}, S_{g+1}, \ldots, S_N\}$ *[$T$]*, where $w_g = S_g$ is a pre-selected *qualifier* from $\{S_1, S_2, \ldots, S_N\}$. The truth level $T$ is computed as

$$T = \mu_Q\left(\frac{r}{\sum_{m=1}^{M} \mu_{w_g}(v_g^m)}\right) \tag{6.7}$$

Note that for this canonical form only relative $Q$ can be used[1]. An example of such a summary is:

$$\underbrace{\text{Most}}_{Q}\ \underbrace{\text{wells}}_{\mathbb{Y}}\ \text{with}\ \underbrace{\text{high water production}}_{w_g}\ \text{have}\ \underbrace{\text{high oil production}}_{S_1}\ [\underbrace{0.7}_{T}] \tag{6.8}$$

### 6.2.2 Additional Quality Measures

According to Hirota and Pedrycz [44], the following five features are essential to measure the quality of a summary:

1. *Validity*: The summaries must be derived from data with high confidence.
2. *Generality*: This describes how many data support a summary.
3. *Usefulness*: This relates the summaries to the goals of the user, especially in terms of the impact that these summaries may have on decision-making. Usefulness is strongly related to the concept of *interestingness*, which is [125] "*one of the central problems in the field of knowledge discovery.*"
4. *Novelty*: This describes the degree to which the summaries deviate from our expectations, i.e., how unexpected the summaries are.

[1]This means that summaries with absolute $Q$ like "*more than 100 wells with high water production have high oil production*" cannot be generated; however, because this kind of summary can be converted to the first canonical form, e.g., "*more than 100 wells have high water production and high oil production,*" only relative $Q$ being used in the second canonical form does not limit the application of linguistic summarization. Also note that when relative $Q$ is used, a summary in the second canonical form cannot be converted to the first canonical form, e.g., "*most wells with high water production have high oil production*" is different from "*most wells have high water production and high oil production.*"

5. *Simplicity*: This measure concerns the syntactic complexity of the summaries. Generally simpler summaries are easier to understand and hence are preferred.

The *validity* can be understood as the truth level $T$ introduced in the previous subsection. Several other quality measures for linguistic summaries have also been introduced [53,55]. For example, the degree of covering ($T_3$) is related to *generality*, the degree of appropriateness ($T_4$) is related to *novelty*, the length of summary ($T_5$) is related to *simplicity*, etc[2]. $T_3$ and $T_4$ are introduced next, and some problems with them are also pointed out.

The *degree of covering*, $T_3 \in [0,1]$, describes how many objects (in terms of portion) in the dataset satisfying $w_g$ are "covered" by a summary, and is defined in [53] as

$$T_3 = \frac{\sum_{m=1}^{M} t_m}{\sum_{m=1}^{M} h_m} \tag{6.9}$$

where

$$t_m = \begin{cases} 1, & \mu_{\mathbf{S}}(\mathbf{d}_m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.10}$$

$$h_m = \begin{cases} 1, & \mu_{w_g}(v_g^m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.11}$$

According to our experiments, $T_3$ defined in (6.9) does not provide much extra information beyond $T$, because the correlation between $T$ and $T_3$ is usually very large, e.g.,

[2] There is no quality measure related to usefulness.

larger than 0.9. So, a different quality measure that can indicate whether a rule has sufficient coverage and is more independent of $T$ is desired. Such a measure, called the degree of sufficient coverage, has been proposed in Section 6.4.3.

Next consider the *degree of appropriateness* defined in [53]. Suppose a summary containing summarizers $\mathbf{S}$ is partitioned into $N$ partial summaries, each of which consists of only one summarizer, e.g., "*$Q$ objects from $\mathbb{Y}$ being/with $w_g$ are/have $S_1$ and $S_2$*" can be partitioned into two partial summaries, "*$Q$ objects from $\mathbb{Y}$ being/with $w_g$ are/have $S_1$*" and "*$Q$ objects from $\mathbb{Y}$ being/with $w_g$ are/have $S_2$.*" Denote

$$r_n = \frac{\sum_{m=1}^{M} t_{n,m}}{M}, \quad n = 1, \ldots, N \tag{6.12}$$

where

$$t_{n,m} = \begin{cases} 1, & \mu_{S_n}(v_n^m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.13}$$

Then, the degree of appropriateness, $T_4$, is defined as [53]

$$T_4 = \left| \prod_{n=1}^{N} r_n - T_3 \right|. \tag{6.14}$$

Note the mismatch between the two components of $T_4$ in (6.14): $r_n$ is actually the degree of covering of the summary "*$Q$ objects from $\mathbb{Y}$ are/have $S_n$,*" whereas $T_3$ computes the degree of covering of the summary "*$Q$ objects from $\mathbb{Y}$* ***being/with*** $\mathbf{w_g}$ *are/have*

$\{S_1, \ldots, S_{g-1}, S_{g+1}, \ldots, S_N\}$." Since the first term does not consider the qualifier $w_g$ whereas the second does, the meaning of $T_4$ is unclear.

The following example, quoted from [55], was used to illustrate the meaning of the degree of appropriateness:

> *For a database concerning the employees, if 50% of them are less than 25 years old and 50% are highly qualified, then we may expect that 25% of the employees would be less than 25 years old and highly qualified; this would correspond to a typical, fully expected situation. However, if the degree of appropriateness is, e.g., 0.39 (i.e. 39% are less than 25 years old and highly qualified), then the summary found reflects an interesting, not fully expected relation in our data. This degree describes therefore how characteristic for the particular database the summary found is. $T_4$ is very important because, for instance, a trivial summary like "100% of sale is of any articles" has full validity (truth) if we use the traditional degree of truth but its degree of appropriateness is equal 0 which is correct.*

The above example is interesting; however, because it does not use a qualifier $w_g$ whereas $T_4$ is supposed to be applied only to summaries involving $w_g$, it cannot be used to illustrate $T_4$.

According to our understanding, the idea of $T_4$ is to test the independency of the summarizers. We first assume all summarizers are independent and decompose them into several sub-summaries each with only one summarizer. Then we compute the "expected" degree of covering by assuming the sub-summarizers are independent. Finally, the difference between the "expected" degree of covering and true degree of covering is computed. If the difference is significant, then the summary reflects something unexpected and hence interesting (note that an unexpected summary does not necessary have high $T$). So, a more reasonable way to compute $T_4$ is to re-define $r_n$ as

$$r_n = \frac{\sum_{m=1}^{M} t'_{n,m}}{\sum_{m=1}^{M} h_m}, \quad n = 1, \ldots, N \tag{6.15}$$

where

$$t'_{n,m} = \begin{cases} 1, & \mu_{S_n}(v_n^m) > 0 \text{ and } \mu_{w_g}(v_g^m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.16}$$

and then to compute $T_4$ by (6.14).

$T_4$ is not used in this chapter because we are more interested in finding outlier rules and data instead of studying the independency of summarizers. A degree of outlier will be introduced in Section 6.4.3.

## 6.3 Linguistic Summarization Using IT2 FSs: Niewiadomski's Approach

Though linguistic summarization of a database by T1 FSs has been studied extensively, as pointed out by Niewiadomski [95], *Type-1 membership functions are frequently constructed based on preferences of one expert. However, it may look arbitrary, since it seems more natural when two or more opinions are given to illustrate, e.g., a linguistic term, to model it as objectively as possible. Traditional fuzzy sets dispose no methods of handling these, usually different, opinions. The average or median of several membership degrees keep no information about those natural differences. For instance, the question* What is the compatibility level of the 36.5°C with "temperature of a healthy human body" *can be answered 0.5, 1.0, and 1.0 by three doctors, respectively, but the average 0.866 does not show that one of them remains unconvinced.*

So, Niewiadomski proposed to use interval or general T2 FSs in linguistic summarization [95, 96, 98–100]. The IT2 FS approach is considered in this chapter because we want to obtain the word models from a survey instead of arbitrarily, whereas presently there is no method to obtain general T2 FS word models.

According to Niewiadomski, to qualify as an IT2 FS linguistic summarization, at least one of $Q$ and $S_n$ must be modeled by an IT2 FS, and $\tilde{T} \subseteq [0,1]$ becomes a *truth interval*. Presently, his approach cannot handle the case when both the quantifier $Q$ and the summarizers $S_n$ are IT2 FSs; so, he considers the following two cases separately:

1. $\tilde{Q}$ is an IT2 FS and $S_n$ are T1 FSs.

2. $Q$ is a T1 FS and $\tilde{S}_n$ are IT2 FSs.

### 6.3.1 Summaries with IT2 FS Quantifier $\tilde{Q}$ and T1 FS Summarizers $S_n$

When $\tilde{Q}$ is an IT2 FS and $S_n$ are T1 FSs, the two canonical forms introduced in Section 6.2.1 become:

1. $\tilde{Q}$ *objects from* $\mathbb{Y}$ *are/have* $\{S_1, S_2, \ldots, S_N\}$ *[*$\tilde{T}$*].* The truth interval $\tilde{T}$ is computed as [recall from (2.15) that the membership of a number on an IT2 FS $\tilde{Q}$ is an interval determined by its LMF $\underline{Q}$ and UMF $\overline{Q}$]:

$$\tilde{T} = \left[\mu_{\underline{Q}}\left(\frac{r}{R}\right), \mu_{\overline{Q}}\left(\frac{r}{R}\right)\right] \tag{6.17}$$

   where $r$ and $R$ have been defined in (6.5) and (6.6), respectively.

2. $\tilde{Q}$ *objects from* $\mathbb{Y}$ *being/with* $w_g$ *are/have* $\{S_1, \ldots, S_{g-1}, S_{g+1}, \ldots, S_N\}$ *[*$\tilde{T}$*]*, where $w_g = S_g$ is a pre-selected T1 FS qualifier. The truth interval $\tilde{T}$ is computed as

$$\tilde{T} = \left[\mu_{\underline{Q}}\left(\frac{r}{\sum_{m=1}^{M} \mu_{w_g}(v_g^m)}\right), \mu_{\overline{Q}}\left(\frac{r}{\sum_{m=1}^{M} \mu_{w_g}(v_g^m)}\right)\right] \tag{6.18}$$

   Again, only relative $\tilde{Q}$ can be used in the second canonical form.

### 6.3.2 Summaries with T1 FS Quantifier $Q$ and IT2 FS Summarizers $\tilde{S}_n$

When $Q$ is a T1 FS and $\tilde{S}_n$ are IT2 FSs, the two canonical forms introduced in Section 6.2.1 become:

1. $Q$ *objects from* $\mathbb{Y}$ *are/have* $\{\tilde{S}_1, \tilde{S}_2, \ldots, \tilde{S}_N\}$ *[*$\tilde{T}$*].* The truth interval $\tilde{T}$ is computed as

$$\tilde{T} = \left[\inf_{r \in [\underline{r}, \overline{r}]} \mu_Q\left(\frac{r}{R}\right), \sup_{r \in [\underline{r}, \overline{r}]} \mu_Q\left(\frac{r}{R}\right)\right] \tag{6.19}$$

where

$$\underline{r} = \sum_{m=1}^{M} \min\{\mu_{\underline{S}_1}(v_1^m), \cdots, \mu_{\underline{S}_N}(v_N^m)\} \tag{6.20}$$

$$\overline{r} = \sum_{m=1}^{M} \min\{\mu_{\overline{S}_1}(v_1^m), \cdots, \mu_{\overline{S}_N}(v_N^m)\} \tag{6.21}$$

2. *Q objects from* $\mathbb{Y}$ *being/with* $w_g$ *are/have* $\{\tilde{S}_1, \ldots, \tilde{S}_{g-1}, \tilde{S}_{g+1}, \ldots, \tilde{S}_N\}$ *[*$\tilde{T}$*]*, where $w_g = S_g$ is a T1 FS qualifier. The truth interval $\tilde{T}$ is computed as

$$\tilde{T} = \left[ \inf_{r\in[\underline{r}',\overline{r}']} \mu_Q\left(\frac{r}{\sum_{m=1}^{M}\mu_{w_g}(v_g^m)}\right), \sup_{r\in[\underline{r}',\overline{r}']} \mu_Q\left(\frac{r}{\sum_{m=1}^{M}\mu_{w_g}(v_g^m)}\right)\right] \tag{6.22}$$

where

$$\underline{r}' = \sum_{m=1}^{M} \min\{\mu_{\underline{S}_1}(v_1^m), \cdots, \mu_{\underline{S}_N}(v_N^m)\} \tag{6.23}$$

$$\overline{r}' = \sum_{m=1}^{M} \min\{\mu_{\overline{S}_1}(v_1^m), \cdots, \mu_{\overline{S}_N}(v_N^m)\} \tag{6.24}$$

Note that (6.22) cannot handle the case when $w_g$ is modeled by an IT2 FS, which may be favorable in practice, as we will see in Section 6.5.

### 6.3.3 Additionally Quality Measures

Niewiadomski [95, 97] extended all other quality measures for T1 FS linguistic summarization to IT2 FS linguistic summarization, e.g., degree of covering ($\tilde{T}_3$), degree of appropriateness ($\tilde{T}_4$), etc. However, since they are based on Kacprzyk's definition of $T_3$ and $T_4$, they are problematic, as explained below.

The problem with $\tilde{T}_3$ in Niewiadomski's approach is similar to that with $T_3$ in the T1 FS case: it has high correlation with the truth level $\tilde{T}$.

When $\tilde{Q}$ is an IT2 FS and $w_g$ and $S_n$ are T1 FSs, the degree of appropriateness is a crisp number computed by (6.14). $T_4$ becomes an interval when $Q$ and $w_g$ are T1 FSs and $\tilde{S}_n$ are IT2 FSs, and it is computed as [97]:

$$\tilde{T}_4 = \left[ \left| \prod_{n=1}^{N} \underline{r}_n - \frac{\sum_{m=1}^{M} \overline{T}_m}{\sum_{m=1}^{M} h_m} \right|, \left| \prod_{n=1}^{N} \overline{r}_n - \frac{\sum_{m=1}^{M} \underline{T}_m}{\sum_{m=1}^{M} h_m} \right| \right] \tag{6.25}$$

where

$$\underline{T}_m = \begin{cases} 1, & \mu_{\underline{S}_n}(v_n^m) > 0, \forall n \neq g, \text{ and } \mu_{w_g}(v_g^m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.26}$$

$$\overline{T}_m = \begin{cases} 1, & \mu_{\overline{S}_n}(v_n^m) > 0, \forall n \neq g, \text{ and } \mu_{w_g}(v_g^m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.27}$$

$$\underline{r}_n = \frac{\sum_{m=1}^{M} \underline{t}_{n,m}}{M} \tag{6.28}$$

$$\overline{r}_n = \frac{\sum_{m=1}^{M} \overline{t}_{n,m}}{M} \tag{6.29}$$

in which

$$\underline{t}_{n,m} = \begin{cases} 1, & \text{if } \mu_{\underline{S}_n}(v_n^m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.30}$$

$$\overline{t}_{n,m} = \begin{cases} 1, & \text{if } \mu_{\overline{S}_n}(v_n^m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.31}$$

and $h_m$ is defined in (6.11).

Niewiadomski's definitions of the degree of appropriateness are based on Kacprzyk and Strykowski's [53] definition of $T_4$, so we question their rationale (see Appendix 6.2.2). Additionally, in (6.25) Niewiadomski does not consider the dependence between the internal variables, e.g., for $|\prod_{n=1}^{N} \underline{r}_n - \frac{\sum_{m=1}^{M} \overline{T}_m}{\sum_{m=1}^{M} h_m}|$, $\underline{S}_n$ are used to compute $\prod_{n=1}^{N} \underline{r}_n$ whereas $\overline{S}_n$ are used to compute $\frac{\sum_{m=1}^{M} \overline{T}_m}{\sum_{m=1}^{M} h_m}$.

## 6.4 Linguistic Summarization Using T1 FSs: Our Approach

The main purpose of this chapter is to propose a linguistic summarization approach using IT2 FSs. For ease in understanding, we start with linguistic summarization using T1 FSs; however, this does not mean we advocate that T1 FSs should be used in linguistic summarization. In fact, we always argue that IT2 FSs should be used in linguistic summarization.

### 6.4.1 The Canonical Form

Because we are interested in generating IF-THEN rules from a dataset, our canonical form for linguistic summarization using T1 FSs is:

$$\text{IF } \mathbb{X}_1 \text{ are/have } S_1, \text{ THEN } \mathbb{X}_2 \text{ are/have } S_2 \quad [T] \tag{6.32}$$

where $T$ is a *truth level*. One example of such a rule is:

$$\begin{aligned}&\text{IF } \underbrace{\text{the total length of perforations in a well}}_{\mathbb{X}_1} \text{ is } \underbrace{\text{small}}_{S_1},\\ &\text{THEN } \underbrace{\text{the 180-day cumulative oil production of the well}}_{\mathbb{X}_2} \text{ is } \underbrace{\text{tiny}}_{S_2} [\underbrace{0.9}_{T}]\end{aligned} \tag{6.33}$$

Only single-antecedent[3] single-consequent[4] rules are considered in this section. Multi-antecedent multi-consequent rules are considered in Section 6.5.3.

Our canonical form in (6.32) can be re-expressed as:

$$\textbf{All } \mathbb{Y} \text{ with } S_1 \text{ are/have } S_2 \quad [T] \tag{6.34}$$

It is analogous to Yager's second canonical form (see Appendix 6.2.1), which is

$$Q \text{ objects from } \mathbb{Y} \text{ with } w_g \text{ are/have } S \quad [T] \tag{6.35}$$

[3]Antecedents are the attributes in the IF part of a rule.

[4]Consequents are the attributes in the THEN part of a rule.

i.e., our IF-THEN rule is equivalent to Yager's second canonical form by viewing the word *All* as the *quantifier* $Q$ and $\mathbb{X}_1$ *are/have* $S_1$ as the *qualifier* $w_g$. For example, (6.33) can be understood as:

$$\begin{aligned} &\underbrace{\text{All}}_{Q}\ \underbrace{\text{wells}}_{\mathbb{Y}}\ \text{with}\ \underbrace{\text{small total feet of perforations}}_{w_g} \\ &\text{have}\ \underbrace{\text{tiny 180-day cumulative oil production}}_{S_2}\ [\underbrace{0.9}_{T}] \end{aligned} \tag{6.36}$$

The truth level $T$ for (6.32) is hence computed by using (6.7):

$$T = \mu_{All}\left(\frac{\sum_{m=1}^{M}\min(\mu_{S_1}(v_1^m), \mu_{S_2}(v_2^m))}{\sum_{m=1}^{M}\mu_{S_1}(v_i^m)}\right) \tag{6.37}$$

There can be different models for the quantifier *All*, as shown in Fig. 6.1. When *All* is modeled as a T1 FS, $T$ is a crisp number. When *All* is modeled as an IT2 FS, $T$ becomes an interval. When we model the quantifier *All* as the proportional function shown in Fig. 6.1(a), $\mu_{All}(x) = x$, so that (6.37) becomes

$$T = \frac{\sum_{m=1}^{M}\min(\mu_{S_1}(v_1^m), \mu_{S_2}(v_2^m))}{\sum_{m=1}^{M}\mu_{S_1}(v_1^m)} \tag{6.38}$$

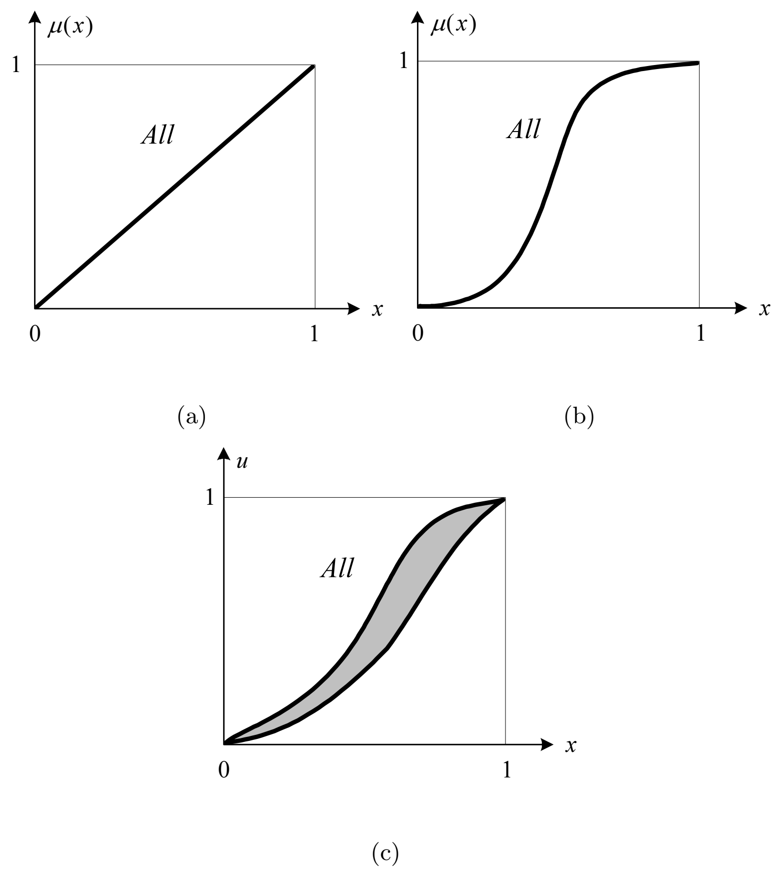



Fig. 6.1: Three possible models for the quantifier *All*. (a) and (b) are T1 FS models, and (c) is an IT2 FS model.

### 6.4.2 Another Representation of $T$

A different representation of the truth level $T$ defined in (6.38) is introduced in this subsection. It will lead easily to the computation of $T$ for linguistic summarization using IT2 FSs, as will be shown in Section 6.5.1. But first, two related definitions are introduced.

**Definition 31** *The* cardinality *of a T1 FS $S_1$ on database $\mathbb{D}$ is defined as*

$$c_{\mathbb{D}}(S_1) = \sum_{m=1}^{M} \mu_{S_1}(v_1^m). \quad \blacksquare \tag{6.39}$$

**Definition 32** *The* joint cardinality *of T1 FSs $\{S_1, ..., S_N\}$ on database $\mathbb{D}$ is defined as*

$$c_{\mathbb{D}}(S_1, ..., S_N) = \sum_{m=1}^{M} \min\{\mu_{S_1}(v_1^m), ..., \mu_{S_N}(v_N^m)\}. \quad \blacksquare \tag{6.40}$$

Using the cardinality $c_{\mathbb{D}}(S_1)$ and joint cardinality $c_{\mathbb{D}}(S_1, S_2)$, (6.38) can be re-expressed as:

$$T = \frac{c_{\mathbb{D}}(S_1, S_2)}{c_{\mathbb{D}}(S_1)}. \tag{6.41}$$

It is worthwhile to mention the analogy between (6.41) and conditional probability in probability theory. Consider $S_1$ and $S_2$ in (6.32) as two events. Then, the conditional probability of $S_2$ given $S_1$, $P(S_2|S_1)$, is computed as:

$$P(S_2|S_1) = \frac{P(S_1, S_2)}{P(S_1)} \tag{6.42}$$

where $P(S_1, S_2)$ is the joint probability of $S_1$ and $S_2$, and $P(S_1)$ is the probability of $S_1$. In (6.41) the numerator can be viewed as the total degree that $S_1$ and $S_2$ are satisfied simultaneously [analogous to $P(S_1, S_2)$], and the denominator can be viewed as the total degree that only the pre-requisite $S_1$ is satisfied [analogous to $P(S_1)$].

### 6.4.3 Additional Quality Measures

As has been mentioned in Section 6.2.2, the truth level $T$ is related to the *validity* of a summary. Three additional quality measures for T1 FS linguistic summarization, corresponding to *generality*, *usefulness* and *novelty*, are proposed in this section. The fifth measure, *simplicity*, is not used in our approach because we require a user to specify the length of the summaries, e.g., how many antecedents and consequents he or she wants to see.

*Generality* is related to the *degree of sufficient coverage*, $T_c$, which describes whether a rule is supported by enough data. It is independent of the truth level $T$ because a rule with high $T_c$ may have low $T$, i.e., there are many data supporting this rule, but also many data do not support this rule. To compute $T_c$, we first compute the *coverage ratio*, which is

$$r = \frac{\sum_{m=1}^{M} t_m}{M} \tag{6.43}$$

where

$$t_m = \begin{cases} 1, & \mu_{S_1}(v_1^m) > 0 \text{ and } \mu_{S_2}(v_2^m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.44}$$

i.e., $r$ is the percentage of data which fit both the antecedent and the consequent of the rule. The coverage ratio cannot be used directly because usually its value is very small (e.g., mostly smaller than 0.1), so $r = 0.15$ may be considered sufficient coverage with degree 1. The following mapping converts the coverage ratio into the appropriate degree of sufficient coverage, and agrees with our feeling:

$$T_c = f(r) \tag{6.45}$$

where $f$ is a function that maps $r$ into $T_c$. The S-shape function $f(r)$ used in this chapter is shown in Fig. 6.2. It is determined by two parameters $r_1$ and $r_2$ ($0 \le r_1 < r_2$), i.e.,

$$f(r) = \begin{cases} 0, & r \le r_1 \\ \frac{2(r-r_1)}{(r_2-r_1)^2}, & r_1 < r < \frac{r_1+r_2}{2} \\ 1 - \frac{2(r_2-r)}{(r_2-r_1)^2}, & \frac{r_1+r_2}{2} \le r < r_2 \\ 1, & r \ge r_2 \end{cases} \tag{6.46}$$

and $r_1 = 0.02$ and $r_2 = 0.15$ are used in this chapter. $f(r)$ can be modified according to the user's requirement on sufficient coverage.

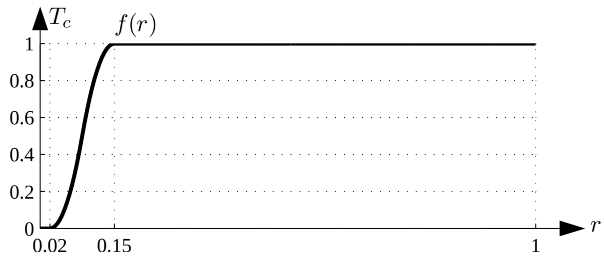



Fig. 6.2: The S-shape function $f(r)$ used in this chapter.

The *degree of usefulness* $T_u$, as its name suggests, describes how useful a summary is. A rule is useful if and only if:

1. It has high truth level, i.e., most of the data satisfying the rule's antecedents also have the behavior described by its consequent; and,

2. It has sufficient coverage, i.e., enough data are described by it.

Hence, $T_u$ is computed as

$$T_u = \min(T, T_c) \tag{6.47}$$

*Novelty* means *unexpectedness*. There are different understandings of unexpectedness, e.g., the *degree of appropriateness* defined by Kacprzyk and Strykowski [53] considers the independency of the summarizers (see Appendix 6.2.2). In this chapter unexpectedness is related to the *degree of outlier*, $T_o$, which indicates the possibility that a rule describes only outliers instead of a useful pattern. Clearly, the degree of sufficient coverage $T_c$ for an outlier rule must be very small, i.e., it only describes very few data; however, small $T_c$ alone is not enough to identify outliers rules, and the truth level $T$ should also be considered. When $T_c$ is small, $T$ can be small (close to 0), medium (around 0.5) or large (close to 1), as shown in Fig. 6.3, where the rule "IF $x$ is Low, THEN $y$ is High" is illustrated for different cases:

1. For the rule illustrated by the shaded region in Fig. 6.3(a), the truth level $T$ is large because all data satisfying the antecedent ($x$ is Low) also satisfy the consequent ($y$ is High). Visual inspection suggests that this rule should be considered as an outlier because the data described by it are isolated from the rest.

2. For the rule illustrated by the shaded region in Fig. 6.3(b), the truth level $T$ is small because most data satisfying the antecedent ($x$ is Low) do not satisfy the consequent ($y$ is High). Visual inspection suggests that this rule should be considered as an outlier because the data described by it are isolated from the rest.

3. For the rule illustrated by the shaded region in Fig. 6.3(c), the truth level $T$ is medium because the data satisfying the antecedent ($x$ is Low) are distributed somewhat uniformly in the $y$ domain. By visual inspection, this rule should not be considered as an outlier (although it is not a good rule as $T_u$ would be small) because its data are not so isolated from the rest.

In summary, an outlier rule must satisfy: 1) The degree of sufficient coverage, $T_c$, is very small; and 2) The truth level, $T$, must be very small or very large. Finally, note that the purpose of finding an outlier rule is to help people identify possible outlier data and then to further investigate them. So, we need to exclude a rule with $T = 0$ from

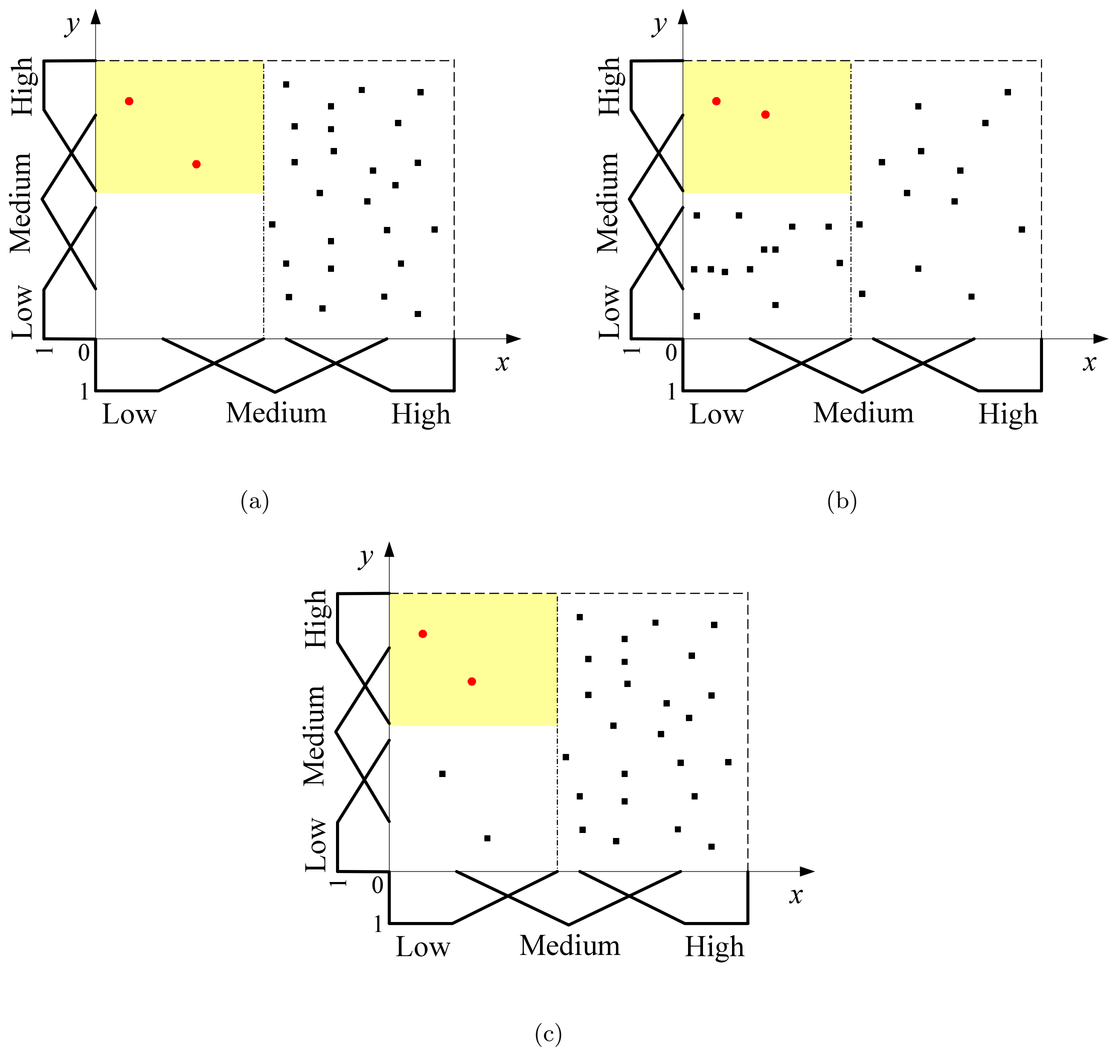



Fig. 6.3: Three cases for the rule "IF $x$ is Low, THEN $y$ is High," whose $T_c$ is small. (a) $T$ is large, (b) $T$ is small, and (c) $T$ is medium.

being identified as an outlier because in this case the rule does not describe any data. The following formulas are used in this chapter to compute the degree of outlier $T_o$:

$$T_o = \begin{cases} \min(\max(T, 1-T), 1-T_c), & T > 0 \\ 0, & T = 0 \end{cases} \tag{6.48}$$

The term $\max(T, 1-T)$ convert a small $T$ (close to 0) or a large $T$ (close to 1) to a large number in [0, 1], and $\min(\max(T, 1-T), 1-T_c)$ further imposes the constraint that $T_c$ must be small for an outlier rule.

A graph illustrating the location of useful rules (high $T_u$) and outlier rules (high $T_o$) in the domain formed by $T$ and $T_c$ is shown in Fig. 6.4.

A summary of the correspondences between the quality measures proposed by Hirota and Pedrycz [44] and us is given in Table 6.2.

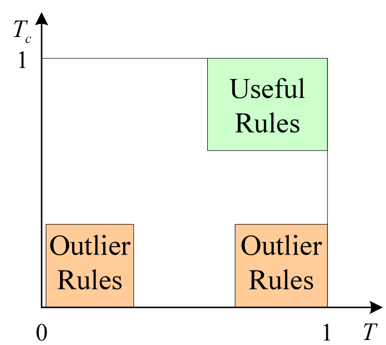



Fig. 6.4: Illustration of useful rules and outlier rules determined by $T$ and $T_c$. The small gap at $T = 0$ means that rules with $T = 0$ are excluded from being considered as outliers.

Table 6.2: Correspondences between the quality measures proposed by Hirota and Pedrycz [44] and us.

| Hirota and Pedrycz's Quality Measure | Our Quality Measure |
|---|---|
| Validity | Truth level ($T$) |
| Generality | Degree of sufficient coverage ($T_c$) |
| Usefulness | Degree of usefulness ($T_u$) |
| Novelty | Degree of outlier ($T_o$) |
| Simplicity | Not used in our method |

## 6.5 Linguistic Summarization Using IT2 FSs: Our Approach

The canonical form of linguistic summarization using IT2 FSs and the associated quality measures are proposed in this section. They are extended from the previous section's results on linguistic summarization using T1 FSs.

### 6.5.1 The Canonical Form

When IT2 FSs are used in linguistic summarization to generate IF-THEN rules, our canonical form becomes:

$$\text{IF } \mathbb{X}_1 \text{ are/have } \tilde{S}_1, \text{ THEN } \mathbb{X}_2 \text{ are/have } \tilde{S}_2 \quad [T] \tag{6.49}$$

(6.49) can be re-expressed as:

$$\textbf{All } \mathbb{Y} \text{ with } \tilde{S}_1 \text{ are/have } \tilde{S}_2 \quad [T] \tag{6.50}$$

It is analogous to Niewiadomski's second canonical form (see Section 6.3.1), which is

$$Q \text{ objects from } \mathbb{Y} \text{ with } w_g \text{ are/have } \tilde{S} \quad [T] \tag{6.51}$$

i.e., our IF-THEN rule is equivalent to Niewiadomski's second canonical form by viewing the word *All* as the *quantifier* $Q$ and $\mathbb{X}_1$ *are/have* $\tilde{S}_1$ as the *qualifier* $w_g$, except a small difference that in Niewiadomski's approach $w_g$ must be a T1 FS whereas we use an IT2 FS.

Recall from (6.41) that the truth level for linguistic summarization using T1 FSs is computed based on the cardinalities of T1 FSs on a database $\mathbb{D}$. To extend that result to IT2 FSs, the following definitions are needed.

**Definition 33** *The* cardinality *of an IT2 FS $\tilde{S}_1$ on dataset $\mathbb{D}$ is defined as*

$$C_{\mathbb{D}}(\tilde{S}_1) \equiv [c_{\mathbb{D}}(\underline{S}_1), c_{\mathbb{D}}(\overline{S}_1)] = \left[\sum_{m=1}^{M} \mu_{\underline{S}_1}(v_1^m), \sum_{m=1}^{M} \mu_{\overline{S}_1}(v_1^m)\right] \tag{6.52}$$

*and the* average cardinality *is*

$$c_{\mathbb{D}}(\tilde{S}_1) = \frac{c_{\mathbb{D}}(\underline{S}_1) + c_{\mathbb{D}}(\overline{S}_1)}{2}. \quad \blacksquare \tag{6.53}$$

**Definition 34** *The* joint cardinality *of IT2 FSs $\{\tilde{S}_1, ..., \tilde{S}_N\}$ on database $\mathbb{D}$ is defined as*

$$C_{\mathbb{D}}(\tilde{S}_1, ..., \tilde{S}_N) \equiv \left[c_{\mathbb{D}}(\underline{S}_1, ..., \underline{S}_N), c_{\mathbb{D}}(\overline{S}_1, ..., \overline{S}_N)\right]$$

$$= \left[ \sum_{m=1}^{M} \min\{\mu_{\underline{S}_1}(v_1^m), ..., \mu_{\underline{S}_N}(v_N^m)\}, \sum_{m=1}^{M} \min\{\mu_{\overline{S}_1}(v_1^m), ..., \mu_{\overline{S}_N}(v_N^m)\} \right] \tag{6.54}$$

*and the* average joint cardinality *is*

$$c_{\mathbb{D}}(\tilde{S}_1, ..., \tilde{S}_N) = \frac{c_{\mathbb{D}}(\underline{S}_1, ..., \underline{S}_N) + c_{\mathbb{D}}(\overline{S}_1, ..., \overline{S}_N)}{2}. \quad \blacksquare \tag{6.55}$$

A straight-forward extension of (6.41) to linguistic summarization using IT2 FSs is to define

$$\tilde{T} = \frac{C_{\mathbb{D}}(\tilde{S}_1, \tilde{S}_2)}{C_{\mathbb{D}}(\tilde{S}_1)}. \tag{6.56}$$

Because both $C_{\mathbb{D}}(\tilde{S}_1, \tilde{S}_2)$ and $C_{\mathbb{D}}(\tilde{S}_1)$ are intervals, $\tilde{T}$ is also an interval. $\tilde{T}$ cannot be computed using simple interval arithmetics, i.e.,

$$\tilde{T} = \left[ \frac{\sum_{m=1}^{M} \min\{\mu_{\underline{S}_1}(v_1^m), \mu_{\underline{S}_2}(v_2^m)\}}{\sum_{m=1}^{M} \mu_{\overline{S}_1}(v_1^m)}, \frac{\sum_{m=1}^{M} \min\{\mu_{\overline{S}_1}(v_1^m), \mu_{\overline{S}_2}(v_2^m)\}}{\sum_{m=1}^{M} \mu_{\underline{S}_1}(v_1^m)} \right] \tag{6.57}$$

because $\tilde{S}_1$ appears in both the numerator and the denominator of (6.56), which means the same embedded T1 FS of $\tilde{S}_1$ must be used in both places in computation, whereas in each of the two end-points in (6.57), different embedded T1 FSs of $\tilde{S}_1$ are used in the numerator and the denominator. Though it is possible to derive an interval $\tilde{T}$ based on the Representation Theorem for IT2 FSs [81], the computation is complicated, and as explained at the end of this subsection, it is also unnecessary. So, the truth level $T$ is defined as a number in this chapter based on average cardinalities instead of cardinalities.

By substituting the cardinalities in (6.41) by their respective average cardinalities, the truth level $T$ of (6.49) is thus computed as

$$T = \frac{c_{\mathbb{D}}(\tilde{S}_1, \tilde{S}_2)}{c_{\mathbb{D}}(\tilde{S}_1)}. \tag{6.58}$$

Like its T1 counterpart (see Section 6.4.1), (6.58) is also analogous to the conditional probability $P(\tilde{S}_2|\tilde{S}_1)$, which is computed as

$$P(\tilde{S}_2|\tilde{S}_1) = \frac{P(\tilde{S}_1, \tilde{S}_2)}{P(\tilde{S}_1)} \tag{6.59}$$

i.e., $c_{\mathbb{D}}(\tilde{S}_1, \tilde{S}_2)$ is the total degree that both $\tilde{S}_1$ and $\tilde{S}_2$ are satisfied [analogous to $P(\tilde{S}_1, \tilde{S}_2)$], and $c_{\mathbb{D}}(\tilde{S}_1)$ is the total degree that only the pre-requisite $\tilde{S}_1$ is satisfied [analogous to $P(\tilde{S}_1)$].

A reader may argue that information is lost as we describe the truth level of an IT2 FS linguistic summary using a number instead of an interval. Note that two categories of uncertainties need to be distinguished here: 1) uncertainties about the content of an IF-THEN rule, which are represented by IT2 FSs $\tilde{S}_1$ and $\tilde{S}_2$; and, 2) uncertainties about the validity of the rule, which may be described by an interval instead of a number. We think the first category of uncertainty is more important because it is the content of a rule that provides knowledge. The validity is used to rank the rules and hence to find the best; however, how should it be used in decision-making is still an open problem. A truth level is easier to compute and more convenient in ranking rules than a truth interval; so, it is used in this chapter.

### 6.5.2 Additional Quality Measures

For linguistic summarization using IT2 FSs, the coverage ratio is still computed by (6.43), but $t_m$ is defined differently:

$$t_m = \begin{cases} 1, & \mu_{\overline{S}_1}(v_1^m) > 0 \text{ and } \mu_{\overline{S}_2}(v_2^m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6.60}$$

i.e., we count all objects with non-zero membership ($J_x$ in (2.15) does not equal [0, 0]) on both antecedent and consequent. Once the coverage ratio $r$ is obtained, the degree of sufficient coverage $T_c$ is computed by (6.45). Because both $T$ and $T_c$ are crisp numbers, (6.47) and (6.48) can again be used to compute the degree of usefulness and the degree of outliers.

### 6.5.3 Multi-Antecedent Multi-Consequent Rules

The generalization of the results for single-antecedent single-consequent rules to multi-antecedent multi-consequent (MAMC) rules is straightforward. Consider an MAMC rule:

$$\begin{aligned} &\text{IF } \mathbb{X}_1 \text{ are/have } \tilde{S}_1 \text{ and ... and } \mathbb{X}_K \text{ are/have } \tilde{S}_K, \\ &\text{THEN } \mathbb{X}_{K+1} \text{ are/have } \tilde{S}_{K+1} \text{ and ... and } \mathbb{X}_N \text{ are/have } \tilde{S}_N \quad [T] \end{aligned} \tag{6.61}$$

The truth level $T$ is computed as

$$T = \frac{c_{\mathbb{D}}(\tilde{S}_1, ..., \tilde{S}_N)}{c_{\mathbb{D}}(\tilde{S}_1, ..., \tilde{S}_K)} \tag{6.62}$$

and the degree of sufficient coverage $T_c$ is computed by redefining $t_m$ as

$$t_m = \begin{cases} 1, & \mu_{\overline{S}_n}(v_n^m) > 0, \quad \forall n = 1, ..., N \\ 0, & \text{otherwise} \end{cases} \tag{6.63}$$

Once the coverage ratio $r$ is obtained, $T_c$ is computed by (6.45). Because both $T$ and $T_c$ are crisp numbers, (6.47) and (6.48) can again be used to compute $T_u$ and $T_o$.

*Comment:* In [67] Lee considers MAMC rules in fuzzy logic control. By assuming the consequents are independent control actions, he proposes to decompose such a rule into $q$ multi-antecedent single-consequent (MASC) rules (see Page 426 of [67]), where $q$ is the number of consequents in the original MAMC rule. Although his approach is appropriate for fuzzy logic control, it may not be applied to knowledge extraction because by using "*and*" to connect a group of consequents and computing a single truth level $T$ we consider explicitly the correlations among the consequents (i.e., Lee's assumption that the consequents are independent does not hold here), whereas the correlations are lost when an MAMC rule is decomposed into MASC rules. For example, the rule in (6.61) is not equivalent to the combination of the following $N - K$ MASC rules:

IF $\mathbb{X}_1$ are/have $\tilde{S}_1$ and ... and $\mathbb{X}_K$ are/have $\tilde{S}_K$, THEN $\mathbb{X}_{K+1}$ are/have $\tilde{S}_{K+1}$ $[T_1]$

IF $\mathbb{X}_1$ are/have $\tilde{S}_1$ and ... and $\mathbb{X}_K$ are/have $\tilde{S}_K$, THEN $\mathbb{X}_{K+2}$ are/have $\tilde{S}_{K+2}$ $[T_2]$

$\vdots$

IF $\mathbb{X}_1$ are/have $\tilde{S}_1$ and ... and $\mathbb{X}_K$ are/have $\tilde{S}_K$, THEN $\mathbb{X}_N$ are/have $\tilde{S}_N$ $[T_{N-K}]$. ■

## 6.6 Example 2 Completed

The fracture stimulation optimization problem has been introduced in Example 2. A Matlab-based Demo was created to demonstrate how our linguistic summarization techniques can be used to extract rules for the fracture process. Our dataset consists of 85 wells after pre-processing to remove outliers [52]. Linguistic summarization was used to find the relationship between the following inputs and 180-day cumulative oil production (*Oil* for short in all screenshots in this chapter):

- Number of stages (*#Stage* for short)
- Total number of holes (*#Holes* for short)
- Total length of perforations (*lengthPerf* for short)
- Total sand volume (*Sand* for short)
- Total pad volume (*Pad* for short)
- Total slurry volume (*Slurry* for short)

Four functions were implemented in the Demo, as shown in Fig. 6.5:

- *Simple Query*: Show wells with certain properties.
- *Rule Validation*: Given a conjecture in the form of an IF-THEN rule, compute $T$, $T_c$, $T_u$ and $T_o$ based on the dataset.
- *Global Top 10 Rules*: Given the number of antecedents, find the top 10 rules with the maximum $T$, $T_c$, $T_u$ or $T_o$.
- *Local Top 10 Rules*: Given the number of antecedents and the desired consequent, find the top 10 combinations of the antecedents that have the maximum $T$, $T_c$, $T_u$ or $T_o$.

These four functions are described in more details next.

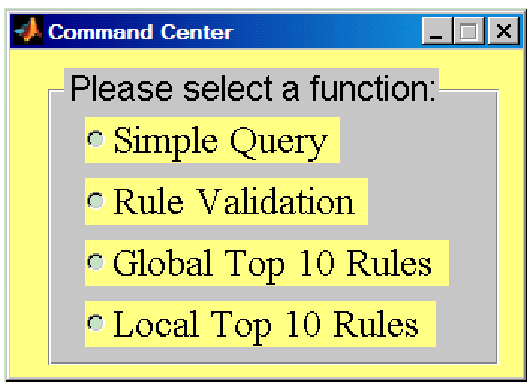



Fig. 6.5: The Command Center where the four functions can be launched.

### 6.6.1 Simple Query

The Simple Query function is simply a data visualization tool. It does not use any linguistic summarization techniques. A screenshot is shown in Fig. 6.6. A user can select the name and values of the properties he or she wants to query from the popup menus. Each property has five linguistic terms associated with it: *Tiny, Small, Medium, High* and *Huge*. They are collected from Fig. 2.12. The user can also click on the red *and* pushbutton to remove a property, or the blue *and* button to add a property.

The query results are displayed by a Parallel Coordinates approach [162], where each coordinate represents an input or output, and the two numbers labeled at the two ends of each coordinate represent the range of that variable, e.g., observe from Fig. 6.6 that *#Stage* has range [3, 17]. Each well is represented in Fig. 6.6 as a curve. The blue curves represent those wells satisfying the user's query. The light green region indicates the area covered by the query.

### 6.6.2 Rule Validation

A screenshot of the Rule Validation graphical user interface (GUI) is shown in Fig. 6.7. A user can select the number of antecedents, their names and values, and also the value for 180-day oil production. Once the rule is specified, linguistic summarization is used to compute $T$, $T_c$, $T_u$ and $T_o$ for it. The results are displayed similar to the way they are displayed in the Simple Query GUI, except that now more colors are used. The blue curves in the bottom axes represent those wells supporting the current rule under consideration (i.e., those wells satisfying *both* the antecedents and the consequents of the rule), and the strength of supporting is indicated by the depth of the blue color. The red curves represent those wells violating the current rule (i.e., those wells satisfying *only* the antecedent part of the rule), and the strength of violating is indicated by the depth of the

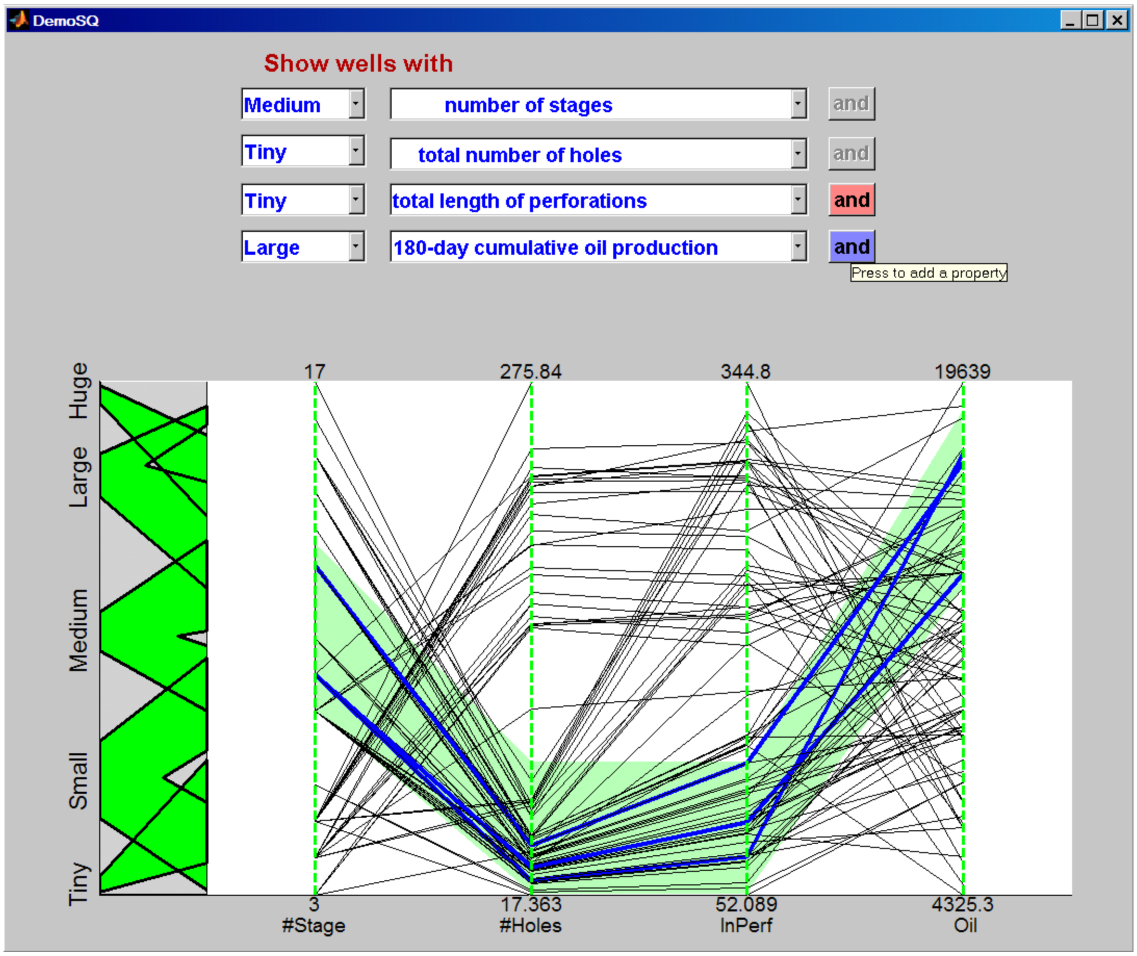



Fig. 6.6: A screenshot of the Simple Query GUI.

red color. The black curves are wells irrelevant to the current rule (i.e., those wells *not* satisfying the antecedent part of the rule). In addition to 180-day oil production, 180-day water and gas productions are also included in each figure for reference; however, they are not considered as the consequents of the rules, i.e., they are not used in computing the quality measures of the rule.

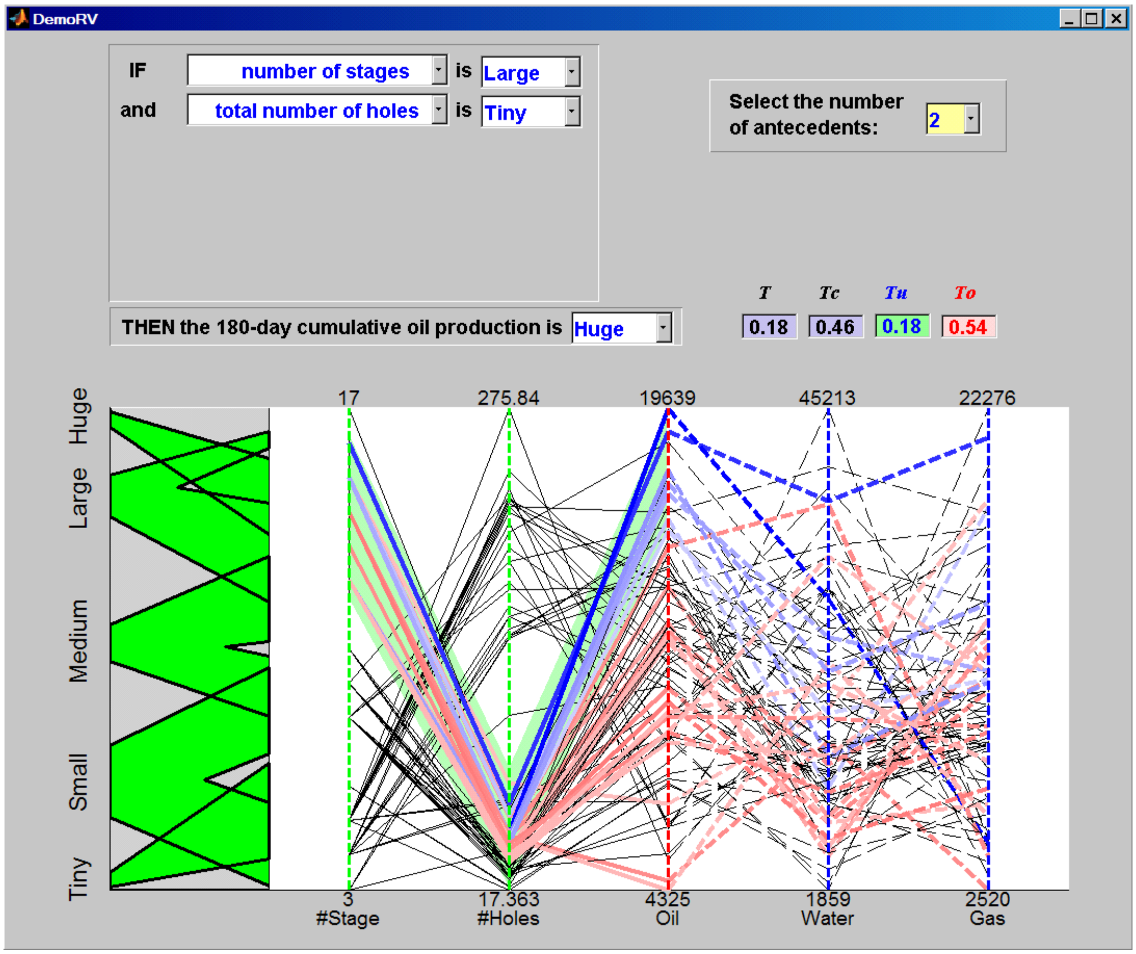



Fig. 6.7: A screenshot of the Rule Validation GUI.

### 6.6.3 Global Top 10 Rules

This function is used to automatically find the top 10 rules according to the ranking criterion a user chooses. Figs. 6.8-6.11 show the top 10 rules when $T$, $T_c$, $T_u$ and $T_o$ are used as the ranking criterion, respectively. A user first specifies the number of antecedents. The program then computes $T$, $T_c$, $T_u$ and $T_o$ for all possible combinations of rules with such number of antecedents. Since there are a total of six antecedents, and each antecedent and consequent domain consists of five MFs, the total number of all possible $k$-antecedent rules is[5]

$$\binom{k}{6} \times 5^{(k+1)}, \qquad k = 1, ..., 6 \tag{6.64}$$

By default, the top 10 rules are selected according to $T_u$; however, a user can change the ranking criterion by clicking on the four pushbuttons on the top right corner of the GUI. The rules are then updated accordingly. A user can also click on a certain radiobutton to select a specific rule. All wells that support and violate that rule are then highlighted in the bottom axes.

Observe:

1. from Fig. 6.8 that when $T$ is used as the ranking criterion, a rule with high $T$ may describe only one well, so it is very possible that this rule only describes an outlier

[5]This is usually a large number and it increases rapidly as the numbers of antecedents and MFs in each input and output domain increase; so, an efficient algorithm that can eliminate bad rules from the beginning is favorable.

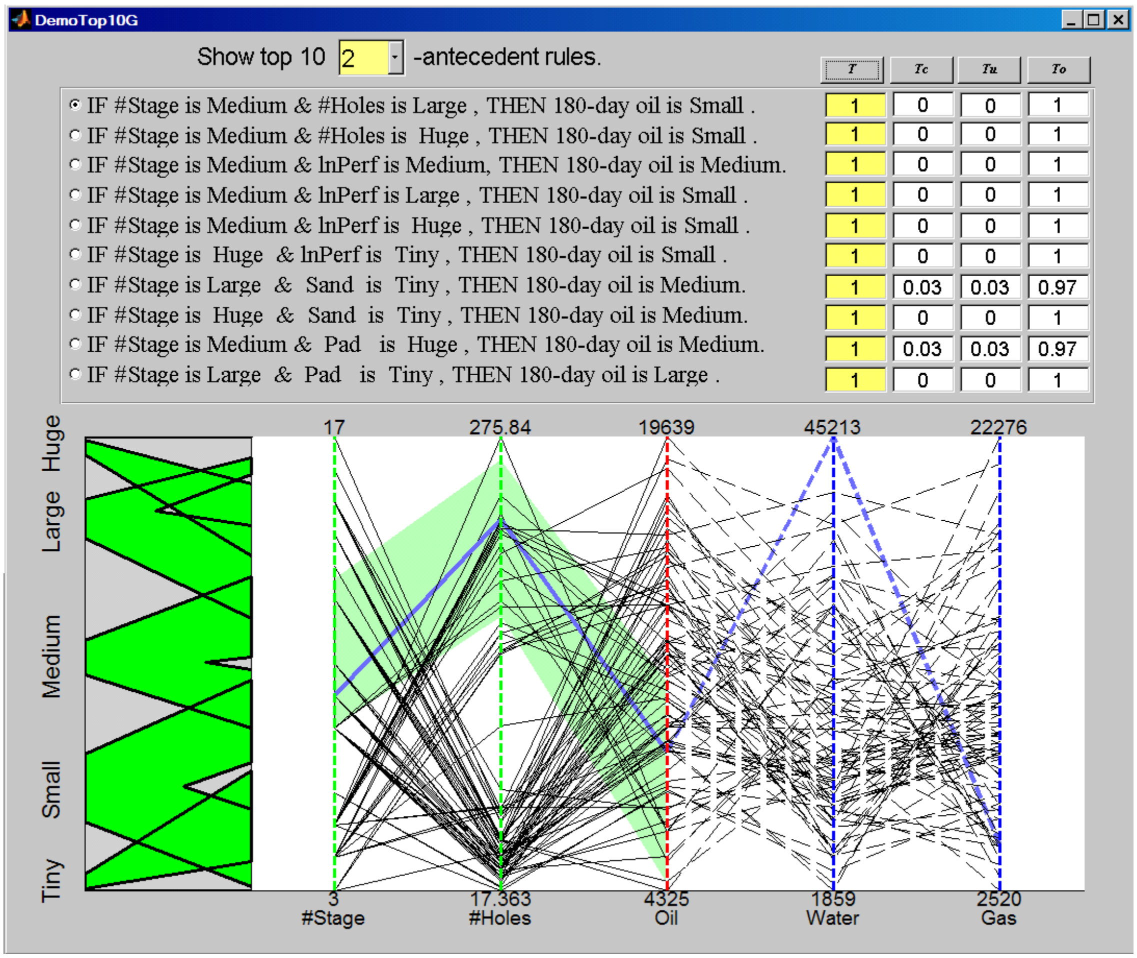



Fig. 6.8: The global top 10 rules according to $T$, the truth level.

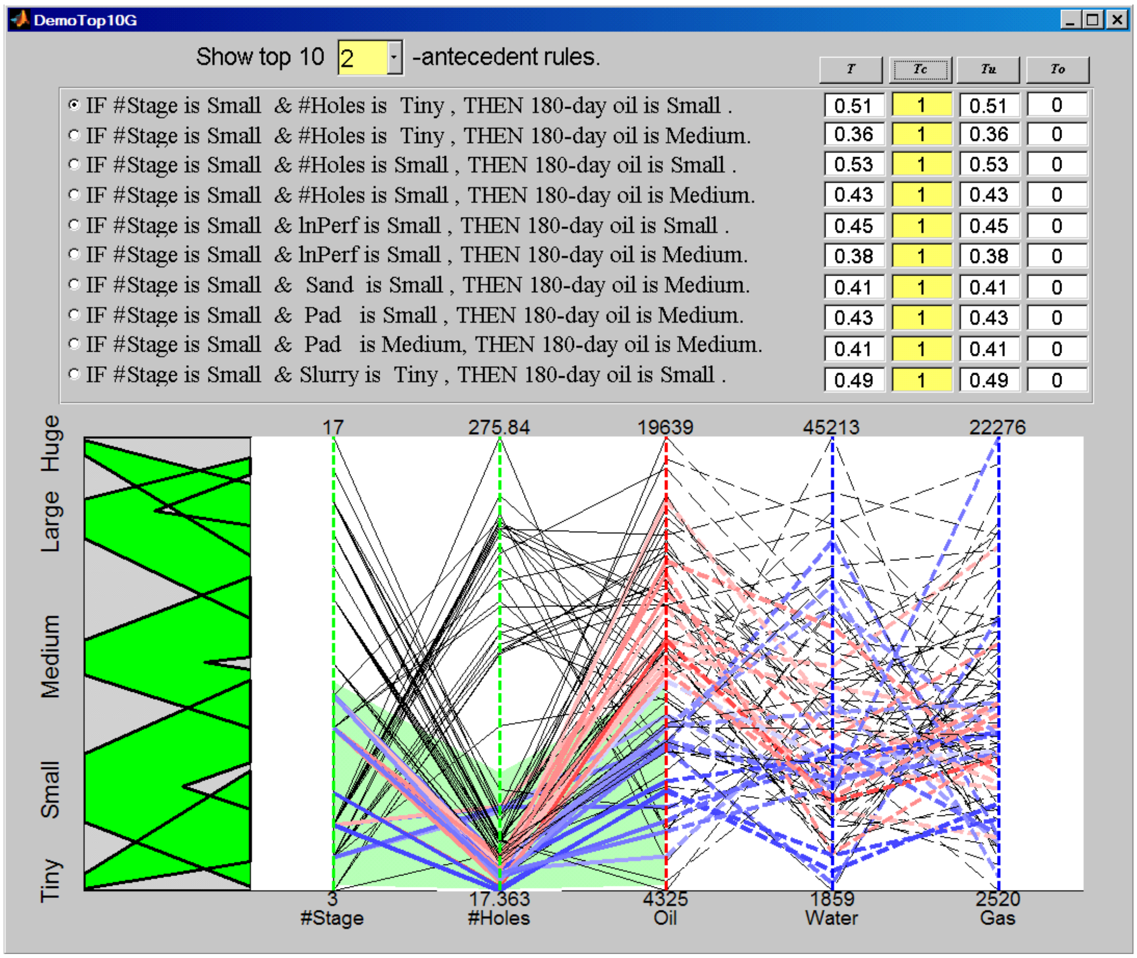



Fig. 6.9: The global top 10 rules according to $T_c$, the degree of sufficient coverage.

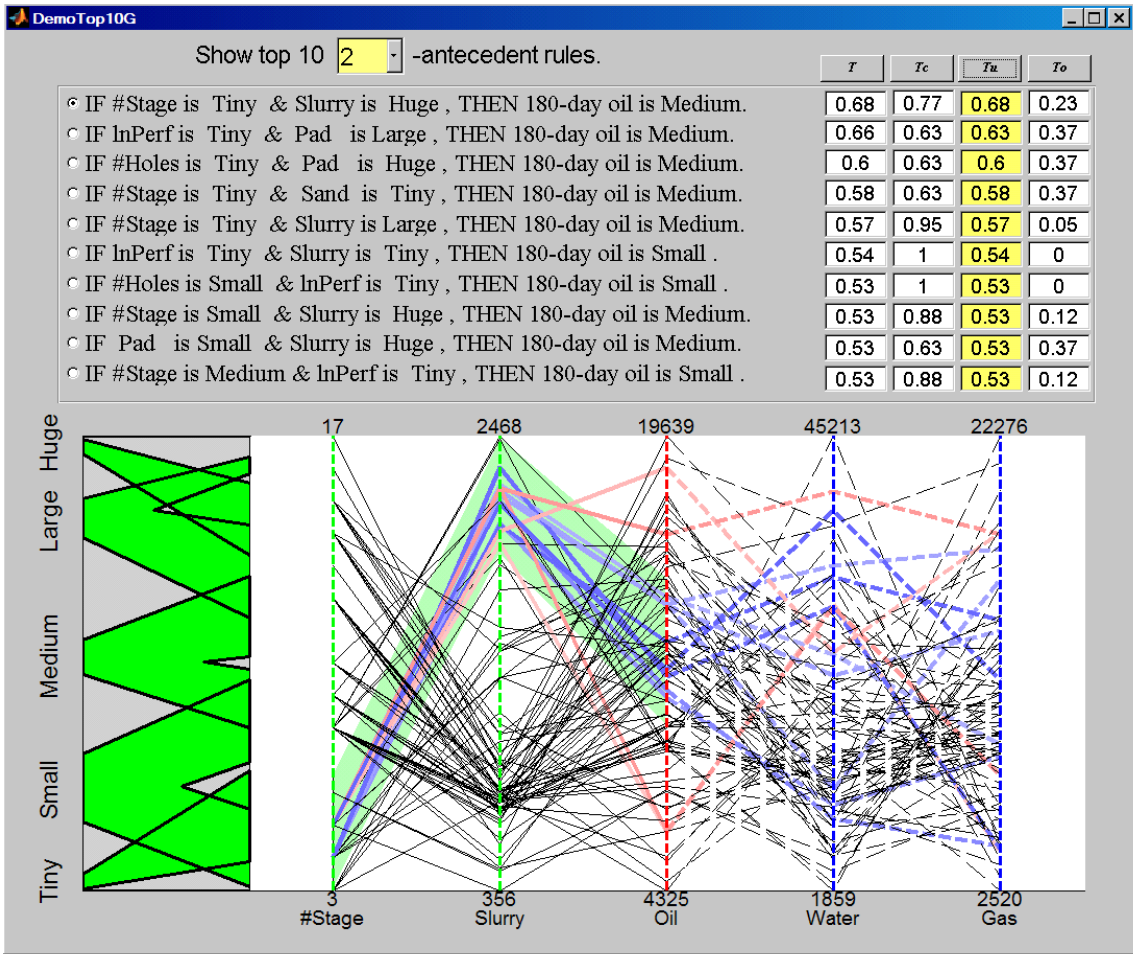



Fig. 6.10: The global top 10 rules according to $T_u$, the degree of usefulness.

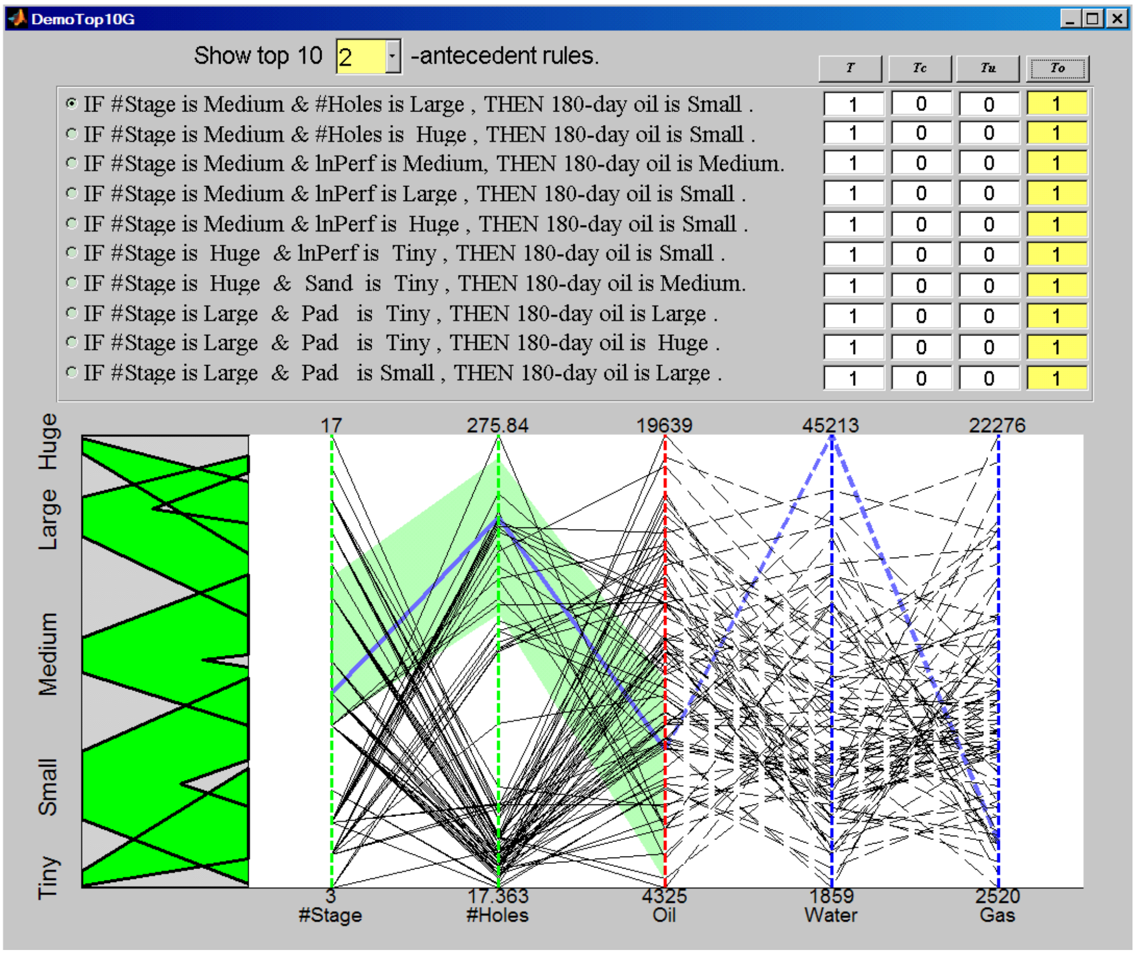



Fig. 6.11: The global top 10 rules according to $T_o$, the degree of outlier.

and hence cannot be trusted. This suggests that $T$ alone is not a reliable quality measure for linguistic summarization.

2. from Fig. 6.9 that when $T_c$ is used as the ranking criterion, a rule with high $T_c$ may have a low truth level, i.e., many wells support the rule but more violate it. So, $T_c$ alone is not a good quality measure.

3. from Fig. 6.10 that when $T_u$ is used as the ranking criterion, a rule with high $T_u$ has both high truth level and sufficient coverage, and hence it describes a useful rule. So, $T_u$ is a comprehensive and reliable quality measure for linguistic summarization.

4. from Fig. 6.11 that when $T_o$ is used as the ranking criterion, a rule with high $T_o$ usually describe only one well, which should be considered as an outlier. So, $T_o$ is useful in finding unexpected data/rules.

In summary, $T_u$ and $T_o$ proposed in this chapter are better quality measures for linguistic summarization than $T$ used in almost all other linguistic summarization literature: a high $T_u$ identifies a useful rule with both high truth level and sufficient coverage, whereas a high $T_o$ identifies outliers in the dataset that are worthy of further investigation.

### 6.6.4 Local Top 10 Rules

This function is very similar to the function to find global top 10 rules, except that the consequent of the rules is specified by the user, e.g., a user may only want to know what combinations of inputs would give huge oil production. Fig. 6.12 shows the top 10 rules when $T_u$ is used as the ranking criterion. A user first specifies the number of antecedents.

The program then computes $T$, $T_c$, $T_u$ and $T_o$ for all possible combinations of rules with such number of antecedents. The number of evaluations in this function is only 1/5 of that in finding global top 10 rules because all rules can have only one instead of five consequents.

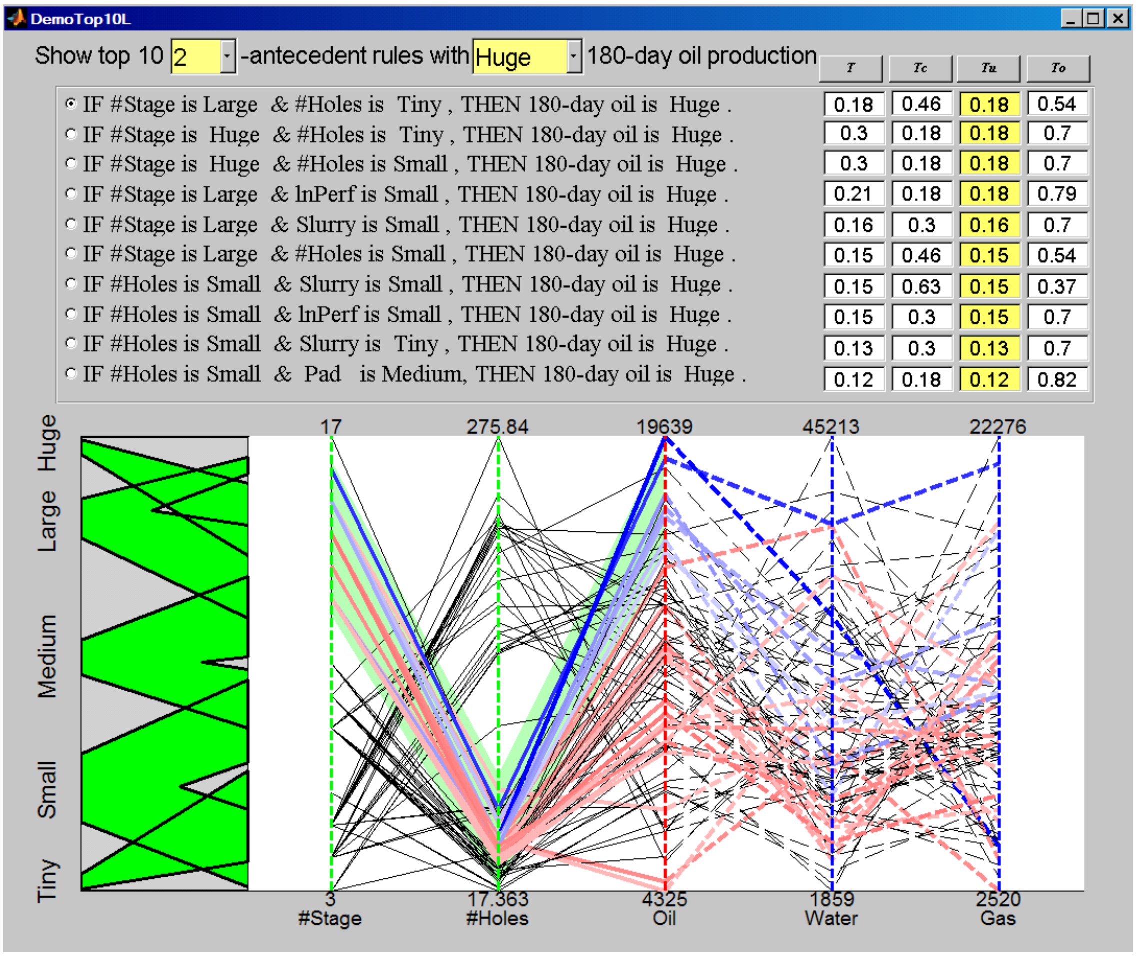



Fig. 6.12: The local top 10 rules according to $T_u$. Observe that there is no satisfactory combination of two properties that lead to huge 180-day oil production.

## 6.7 Discussions

In this section the relationships between linguistic summarization, perceptual reasoning, granular computing, and the Wang-Mendel (WM) method are discussed. Because currently granular computing and the WM method mainly focus on T1 FSs, only T1 FSs are used in the discussion; however, our results can be extended to IT2 FSs without problems.

### 6.7.1 Linguistic Summarization and Perceptual Reasoning

Perceptual reasoning (PR) will be introduced in Chapter 8. A rulebase is needed before PR can be carried out. There are two approaches to construct the rules: 1) from experience, e.g., survey the experts; and, 2) from data, e.g., summarize a database linguistically. The latter has become very convenient because data is usually readily available in this information explosion age.

Additionally, the linguistic summarization approach can serve as a preliminary step for the survey approach, i.e., potential rules can first be extracted from data and then presented to the experts for validation. This would save the time of the experts, and may also help us discover inconsistencies between the data and experience. For example, if from the input-output data of a process we extract a rule which says "*IF $x$ is large, THEN $y$ is medium*" whereas the operator thinks $y$ should be small when $x$ is large, then it is worthwhile to study why the data are not consistent with the operator's experience. It is possible that the dynamics of the process has been changing as time elapses; so, this inconsistency would suggest that it is necessary to update the operator's understanding about the process.

### 6.7.2 Linguistic Summarization and Granular Computing

Granular Computing (GrC) [44, 56, 173, 176, 186] is a general computation theory for effectively using granules such as classes, clusters, subsets, groups and intervals to build an efficient computational model for complex applications with huge amounts of data, information and knowledge. Though the name was first invented by Zadeh in 1998 [186], according to Hirota and Pedrycz [44], "*the idea of information granulation has existed for a long time... For instance, an effect of temporal granulation occurs in analog-to-digital (A/D) conversion equipped with an averaging window: one uniformly granulates an incoming signal over uniform time series. An effect of spatial granulation occurs quite evidently in image processing, especially when we are concerned with image compression.*"

Linguistic summarization can be viewed as a GrC approach, as demonstrated by the following example.

**Example 24** *Consider the example shown in Fig. 6.13, where the training data ($x$ is the input and $y$ is the output) are shown as squares. There is no simple correlation between $x$ and $y$; however, observe that generally as $x$ increases, $y$ first increases and then decreases. Assume each input and output domain is partitioned by three overlapping T1 FSs* Low, Medium *and* High. *Linguistic summarization considers these three intervals in the* $x$ *domain independently and outputs the following three rules for them:*

*IF* $x$ *is Low, THEN* $y$ *is Low*

*IF* $x$ *is Medium, THEN* $y$ *is Medium*

*IF* $x$ *is High, THEN* $y$ *is Low*

*which describe the trend correctly. The resolution of the summarization can be improved by using more MFs in each input/output domain.* ■

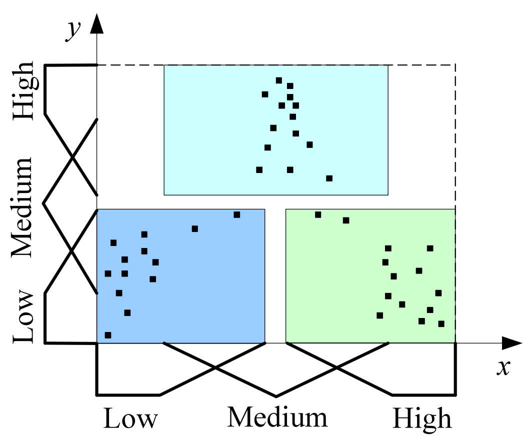



Fig. 6.13: An example to illustrate the idea of granular computing.

### 6.7.3 Linguistic Summarization and the WM Method

The Wang-Mendel (WM) method [83, 140] is a simple yet effective method to generate fuzzy rules from training examples (according to Google Scholar, it has been cited 1,119 times). We use Fig. 6.14, where the 18 training data points are represented by squares[6], to introduce its idea:

1. Each input ($x$) and output ($y$) domain is partitioned into $2L + 1$ (an odd number) overlapping intervals, where $L$ can be different for each variable. Then, MFs and labels are assigned to these intervals. In Fig. 6.14, each of the $x$ and $y$ domain is partitioned into three overlapping intervals by FSs Low, Medium and High. An interval in the $x$ domain and an interval in the $y$ domains together determine a region in the input-output space, e.g., the region determined by High $x$ and Low $y$ is shown as the shaded region in the lower right corner of Fig. 6.14.

2. Because of overlapping MFs, it frequently happens that a datum is in more than one region, e.g., the diamond in Fig. 6.14 belongs to the region determined by High $x$ and Low $y$, and also the region determined by High $x$ and Medium $y$. For each $(x, y)$, we evaluate its degrees of belonging in regions where it occurs, assign it to the region with maximum degree, and generate a rule from it. For example, the degree of belonging of the diamond in Fig. 6.14 to the region determined by High $x$ and Low $y$ (the shaded region in the lower right corner) is $\mu_{High}(x)\mu_{Low}(y) = 1 \times 0.1 = 0.1$, and its degree of belonging to the region determined by High $x$ and Medium $y$ is $\mu_{High}(x)\mu_{Medium}(y) = 1 \times 0.8 = 0.8$; so, the diamond should be assigned to the

[6]Three points are represented by different shapes only for easy reference purpose.

region determined by High $x$ and Medium $y$. The corresponding rule generated from this diamond is hence

$$\text{IF } x \text{ is High, THEN } y \text{ is Medium} \tag{6.65}$$

and it is also assigned a degree of 0.8. Similarly, a rule generated from the cross in Fig. 6.14 is

$$\text{IF } x \text{ is High, THEN } y \text{ is Low} \tag{6.66}$$

and it has a degree of $\mu_{High}(x)\mu_{Low}(y) = 1 \times 1 = 1$.

3. To resolve conflicting rules, i.e., rules with the same antecedent MFs and different consequent MFs, we choose the one with the highest degree and discard all others. For example, Rules (6.65) and (6.66) are conflicting, and Rule (6.66) is chosen because it has a higher degree.

Finally, the three rules generated by the WM method for the Fig. 6.14 data are:

IF $x$ is Low, THEN $y$ is High

IF $x$ is Medium, THEN $y$ is Medium

IF $x$ is High, THEN $y$ is Low

The first rule seems counter-intuitive, but it is a true output of the WM method. It is generated by the circle in Fig. 6.14 with a degree $\mu_{Low}(x)\mu_{High}(y) = 1 \times 1 = 1$, i.e., its degree is higher than two other possible rules, IF $x$ is Low, THEN $y$ is Low and IF $x$ is Low, THEN $y$ is Medium, through these two rules have more data to support them and hence look more reasonable. However, note that this example considers an extreme case. In practice the WM method usually generates very reasonable rules, that's why it is popular.

Once the rules are generated, the degrees associated with them are discard as they are no longer useful.

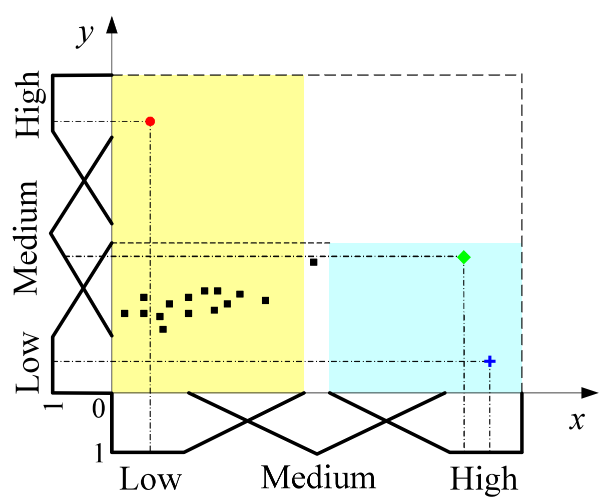



Fig. 6.14: An example to illustrate the difference between the WM method and linguistic summarization. When $x$ is Low, the WM method generates a rule "IF $x$ is Low, THEN $y$ is High" whereas linguistic summarization generates a rule "IF $x$ is Low, THEN $y$ is Low."

**Example 25** *Fig. 6.14 can also be used to illustrate the difference between the WM method and linguistic summarization. Consider the shaded region where $x$ is Low. There are three candidates for a rule in this region:*

$$IF\ x\ is\ Low,\ THEN\ y\ is\ High \tag{6.67}$$

$$IF\ x\ is\ Low,\ THEN\ y\ is\ Medium \tag{6.68}$$

$$IF\ x\ is\ Low,\ THEN\ y\ is\ Low \tag{6.69}$$

*For Rule (6.67),*

$$c_{\mathbb{D}}(Low_x,\ High_y) = \sum_{m=1}^{18} \min(\mu_{Low_x}(x_m), \mu_{High_y}(y_m)) = 1 \tag{6.70}$$

$$c_{\mathbb{D}}(Low_x) = \sum_{m=1}^{18} \mu_{Low_x}(x_m) = 12.8 \tag{6.71}$$

$$T = \frac{c_{\mathbb{D}}(Low_x,\ High_y)}{c_{\mathbb{D}}(Low_x)} = 0.08 \tag{6.72}$$

*Because the dataset consists of 18 points and there is only one datum falls in the region determined by Low $x$ and High $y$, the coverage ratio [see (6.43)] and degree of sufficient coverage [see (6.45)] are*

$$r = 1/18 \tag{6.73}$$

$$T_c = f(r) = 0.15 \tag{6.74}$$

*and hence* $T_u = \min(T, T_c) = 0.08$ *and* $T_o = \min(\max(T, 1-T), 1-Tc) = \min(\max(0.08, 0.92), 1-0.15) = 0.85$.

*Similarly, for Rule (6.68) linguistic summarization gives:*

$$T = 0.31, \quad T_c = 1, \quad T_u = 0.31, \quad T_o = 0 \tag{6.75}$$

*and for Rule (6.69), linguistic summarization gives:*

$$T = 0.71, \quad T_c = 1, \quad T_u = 0.71, \quad T_o = 0 \tag{6.76}$$

*By ranking* $T_u$ *and* $T_o$*, linguistic summarization would select Rule (6.69) as the most useful rule with* $T_u = 0.71$ *and Rule (6.67) as an outlier with* $T_o = 0.85$*. These results are more reasonable than the rule generated by the WM method.*

*Repeating the above procedure for the other two regions, the following three rules are generated when* $T_u$ *is used as the ranking criterion:*

| | |
|---|---|
| *IF x is Low, THEN y is Low* | $T = 0.71,\ T_c = 1,\ T_u = 0.71,\ T_o = 0$ |
| *IF x is Medium, THEN y is Medium* | $T = 0.82,\ T_c = 1,\ T_u = 0.82,\ T_o = 0$ |
| *IF x is High, THEN y is Low* | $T = 0.57,\ T_c = 0.82,\ T_u = 0.57,\ T_o = 0.18$ ■ |

In summary, the differences between the WM method and linguistic summarization are:

1. The WM method tries to construct a predictive model[7] whereas linguistic summarization tries to construct a descriptive model[8]. According to [41], "*a* descriptive model *presents, in convenient form, the main features of the data. It is essentially a summary of the data, permitting us to study the most important aspects of the data without their being obscured by the sheer size of the data set. In contrast, a* predictive model *has the specific objective of allowing us to predict the value of some target characteristic of an object on the basis of observed values of other characteristics of the object.*"

2. Both methods partition the problem domain into several smaller regions and try to generate a rule for each region; however, the WM method generates a rule for a region as long as there are data in it, no matter how many data are there, whereas linguistic summarization does not, e.g., if a region has very few data in it, then these data may be considered as outliers and no useful rule is generated for this region.

3. The rules obtained from linguistic summarization have several quality measures associated with them, so the rules can be sorted according to different criteria, whereas the rules obtained from the WM method are considered equally important[9].

---

[7]Predictive models [47] include classification (grouping items into classes and predicting which class an item belongs to), regression (function approximation and forecast), attribute importance determination (identifying the attributes that are most important in predicting results), etc.

[8]Descriptive models [47] include clustering (finding natural groupings in the data), association models (discovering co-occurrence relationships among the data), feature extraction (creating new attributes as a combination of the original attributes), etc.

[9]There is an improved version of the WM method [137] that assigns a degree of truth to each rule; however, the degree of truth is computed differently from $T$ in this chapter, and the rule consequents are numbers instead of words modeled by FSs; so, it is not considered in this chapter.

# Chapter 7

# Extract Rules through Survey: Knowledge Mining

As has been mentioned in Chapter 1, sometimes the rules describing the dynamics of a process can only be extracted though a survey. In this chapter the SJA introduced in Example 3 is revisited to illustrate this knowledge mining approach. Particularly, we focus on a fuzzy logic flirtation advisor.

Flirtation judgments offer a fertile starting place for developing an SJA for a variety of reasons. First, many behavioral indicators associated with flirtation have been well established [65]. Second, the indicators (e.g., smiling, touching, eye contact) are often ambiguous by themselves and along with a changing level of the behavior (along with other cues) the meaning of the behavior is apt to shift from one inference (e.g., friendly) to another (e.g., flirtation, seductive, or harassing). Third, participants are apt to have had a great deal of experience with flirtation judgments, and be therefore apt to easily make them. Finally, inferences made about the meaning of these behaviors are often sensitive to both the gender of the perceiver and the gender of the interactants [65].

Although our focus is on flirtation judgment, the methodology can also be applied to engineering judgments such as global warming, environmental impact, water quality, audio quality, toxicity, etc.

## 7.1 Survey Design

An SJA uses a rulebase, which is obtained from surveys. The following methodology can be used to conduct the surveys [82, 83]:

1. *Identify the behavior of interest.* This step, although obvious, is highly application dependent. As mentioned above, our focus is on the behavior of flirtation.

2. *Determine the indicators of the behavior of interest.* This requires:

   (a) Establishing a list of candidate indicators (e.g., for flirtation [82], six candidate indicators are *touching*, *eye contact*, *acting witty*, *primping*, *smiling*, and *complementing*).

   (b) Conducting a survey in which a representative population is asked to rank-order in importance the indicators on the list of candidate indicators. In some applications it may already be known what the relative importance of the indicators is, in which case a survey is not necessary.

(c) Choosing a meaningful subset of the indicators, because not all of them may be important. In Step 6, where people are asked to provide consequents for a collection of IF–THEN rules by means of a survey, the survey must be kept manageable, because most people do not like to answer lots of questions; hence, it is very important to focus on the truly significant indicators. The analytic hierarchy process [114] and factor analysis [36] from statistics can be used to help establish the relative significance of indicators.

3. *Establish scales for each indicator and the behavior of interest.* If an indicator is a physically measurable quantity (e.g., temperature, pressure), then the scale is associated with the expected range between the minimum and maximum values for that quantity. On the other hand, many social judgment indicators as well as the behavior of interest are not measurable by means of instrumentation (e.g., touching, eye contact, flirtation, etc.). Such indicators and behaviors need to have a scale associated with them, or else it will not be possible to design or activate an SJA. Commonly used scales are 1 through 5, 0 through 5, 0 through 10, etc. We shall use the scale 0 through 10.

4. *Establish names and collect interval data for each of the indicator's FSs and behavior of interest's FSs.* The issues here are:

   (a) What vocabulary should be used and what should its size be so that the FOUs for the vocabulary completely cover the 0-10 scale and provide the user of the SJA with a user-friendly interface?

(b) What is the smallest number of FSs that should be used for each indicator and behavior of interest for establishing rules?

This is the encoding problem and the IA [70] can be used to find the FOU word models once a satisfactory vocabulary has been established, and word data have been collected from a group of subjects using surveys.

5. *Establish the rules.* Rules are the heart of the SJA; they link the indicators of a behavior of interest to that behavior. The following issues need to be addressed:

   (a) How many antecedents will the rules have? As mentioned earlier, people generally do not like to answer complicated questions; so, we advocate using rules that have either one or two antecedents. An interesting (non-engineering) interpretation for a two-antecedent rule is that it provides the *correlation* effect that exists in the mind of the survey respondent between the two antecedents. Psychologists have told us that it is just about impossible for humans to correlate more than two antecedents (indicators) at a time, and that even correlating two antecedents at a time is difficult. Using only one or two antecedents does not mean that a person does not use more than this number of indicators to make a judgment; it means that a person uses the indicators one or two at a time (this should be viewed as a *conjecture*). This suggests the overall architecture for the SJA should be parallel or hierarchical (see Section 8.4.6).

(b) How many rulebases need to be established? Each rulebase has its own SJA. When there is more than one rulebase, each of the advisors is a social judgment sub-advisor, and the outputs of these sub-advisors can be combined to create the structure of the overall SJA. If, e.g., it has been established that four indicators are equally important for the judgment of flirtation, then there would be up to four single-antecedent rulebases as well as six two-antecedent rulebases. These rulebases can be rank-ordered in importance by means of another survey in which the respondents are asked to do this. Later, when the outputs of the different rulebases are combined, they can be weighted using the results of this step.

There is a very important reason for using sub-advisors for an SJA. Even though the number of important indicators has been established for the social judgment, it is very unlikely that they will all occur at the same time in a social judgment situation. If, for example, *touching, eye contact, acting witty* and *primping* have been established as the four most important indicators for flirtation, it is very unlikely that in a new flirtation scenario, all four occur simultaneously. From your own experiences in flirting, can you recall a situation when someone was simultaneously touching you, made eye contact with you, was acting witty and was also primping? Not very likely! Note that a missing observation is not the same as an observation of zero value; hence, even if it was possible to create four antecedent rules, none of those rules could be activated if one or more of the indicators had a missing observation. It is therefore very important to have sub-advisors that will be activated when one or two of these indicators are occurring.

More discussions about this are in Section 8.4.6.

6. *Survey people (experts) to provide consequents for the rules.* If, e.g., a single antecedent has five FSs associated with it, then respondents would be asked five questions. For two-antecedent rules, where each antecedent is again described by five FSs, there would be 25 questions. The order of the questions should be randomized so that respondents don't correlate their answers from one question to the next. In Step 4 earlier, the names of the consequent FSs were established. Each single-antecedent rule is associated with a question of the form:

    IF the antecedent is (state one of the antecedent's FSs),

    THEN there is (state one of the consequent's FSs) of flirtation.

    Each two-antecedent rule is associated with a question of the form:

    IF antecedent 1 is (state one of antecedent 1's FSs) and antecedent 2 is (state one of antecedent 2's FSs),

    THEN there is (state one of the consequent's FSs) of flirtation.

    The respondent is asked to choose one of the given names for the consequent's FSs. The rulebase surveys will lead to rule consequent histograms, because everyone will not answer a question the same way.

The following nine terms, shown in Fig. 7.1, are taken from the 32-word vocabulary[1] in Fig. 2.12, and are used as the codebook for the SJA: *none to very little* (NVL), *a bit* (AB), *somewhat small* (SS), *some* (S), *moderate amount* (MOA), *good amount* (GA), *considerable amount* (CA), *large amount* (LA), and *maximum amount* (MAA). Their FOUs and centroids have been given in Table 2.1. These FOUs are being used only to illustrate our SJA methodology. In actual practice, word survey data would have to be collected from a group of subjects, using the words in the context of flirtation.

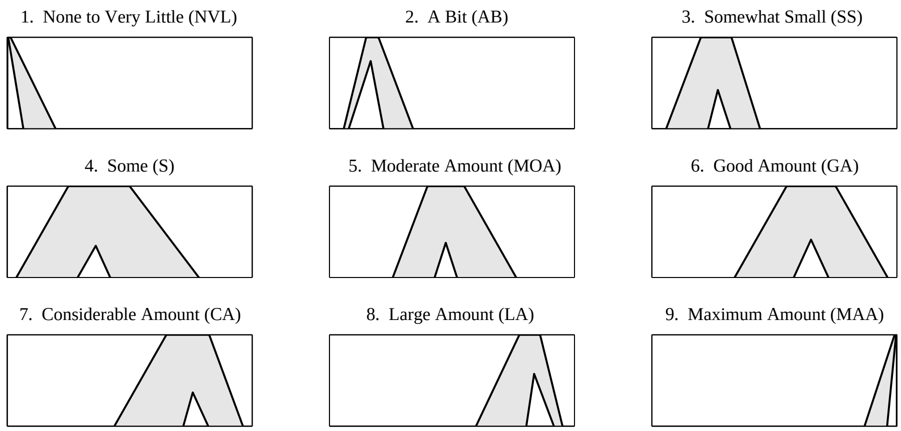



Fig. 7.1: Nine word FOUs ranked by their centers of centroid. Words 1, 4, 5, 8 and 9 were used in the Step 6 survey.

Our SJA was limited to rulebases for one- and two-antecedent rules, in which $x_1$ and $x_2$ denote touching and eye contact, respectively, and $y$ denotes flirtation level. Section 8.4.6 explains how to deduce the output for multiple antecedents using rulebases consisting of only one or two antecedents. For all of the rules, the following five-word subset of the codebook was used for both their antecedents and consequents: *none to very little*, *some*,

[1]They are selected in such a way that they are distributed somewhat uniformly in $[0, 10]$.

*moderate amount*, *large amount*, and *maximum amount*. It is easy to see from Fig. 7.1 that these words cover the interval [0, 10]. Tables 7.1-7.3, which are taken from [82] and Chapter 4 of [83], provide the data collected from 47 respondents to the Step 6 surveys.

Table 7.1: Histogram of survey responses for single-antecedent rules between touching level and flirtation level. Entries denote the number of respondents out of 47 that chose the consequent.

| **Touching** | **Flirtation** | | | | |
|---|---|---|---|---|---|
| | **NVL** | **S** | **MOA** | **LA** | **MAA** |
| 1. NVL | 42 | 3 | 2 | 0 | 0 |
| 2. S | 33 | 12 | 0 | 2 | 0 |
| 3. MOA | 12 | 16 | 15 | 3 | 1 |
| 4. LA | 3 | 6 | 11 | 25 | 2 |
| 5. MAA | 3 | 6 | 8 | 22 | 8 |

Table 7.2: Histogram of survey responses for single-antecedent rules between eye contact level and flirtation level. Entries denote the number of respondents out of 47 that chose the consequent.

| **Eye Contact** | **Flirtation** | | | | |
|---|---|---|---|---|---|
| | **NVL** | **S** | **MOA** | **LA** | **MAA** |
| 1. NVL | 36 | 7 | 4 | 0 | 0 |
| 2. S | 26 | 17 | 4 | 0 | 0 |
| 3. MOA | 2 | 16 | 27 | 2 | 0 |
| 4. LA | 1 | 3 | 11 | 22 | 10 |
| 5. MAA | 0 | 3 | 7 | 17 | 20 |

## 7.2 Data Pre-Processing

Inevitably, there are bad responses and outliers in the survey histograms. These bad data need to be removed before the histograms are used.

Data pre-processing consists of three steps: 1) bad data processing, 2) outlier processing, and, 3) tolerance limit processing, which are quite similar to the pre-processing steps

Table 7.3: Histogram of survey responses for two-antecedent rules between touching/eye contact levels and flirtation level. Entries denote the number of respondents out of 47 that chose the consequent.

| Touching/Eye Contact | Flirtation | | | | |
|---|---|---|---|---|---|
| | **NVL** | **S** | **MOA** | **LA** | **MAA** |
| 1. NVL/NVL | 38 | 7 | 2 | 0 | 0 |
| 2. NVL/S | 33 | 11 | 3 | 0 | 0 |
| 3. NVL/MOA | 6 | 21 | 16 | 4 | 0 |
| 4. NVL/LA | 0 | 12 | 26 | 8 | 1 |
| 5. NVL/MAA | 0 | 9 | 16 | 19 | 3 |
| 6. S/NVL | 31 | 11 | 4 | 1 | 0 |
| 7. S/S | 17 | 23 | 7 | 0 | 0 |
| 8. S/MOA | 0 | 19 | 19 | 8 | 1 |
| 9. S/LA | 1 | 8 | 23 | 13 | 2 |
| 10. S/MAA | 0 | 7 | 17 | 21 | 2 |
| 11. MOA/NVL | 7 | 23 | 16 | 1 | 0 |
| 12. MOA/S | 5 | 22 | 20 | 0 | 0 |
| 13. MOA/MOA | 2 | 7 | 22 | 15 | 1 |
| 14. MOA/LA | 1 | 4 | 13 | 17 | 12 |
| 15. MOA/MAA | 0 | 4 | 12 | 24 | 7 |
| 16. LA/NVL | 7 | 13 | 21 | 6 | 0 |
| 17. LA/S | 3 | 11 | 23 | 10 | 0 |
| 18. LA/MOA | 0 | 3 | 18 | 18 | 8 |
| 19. LA/LA | 0 | 1 | 9 | 17 | 20 |
| 20. LA/MAA | 1 | 2 | 6 | 11 | 27 |
| 21. MAA/NVL | 2 | 16 | 18 | 11 | 0 |
| 22. MAA/S | 2 | 9 | 22 | 13 | 1 |
| 23. MAA/MOA | 0 | 3 | 15 | 18 | 11 |
| 24. MAA/LA | 0 | 1 | 7 | 17 | 22 |
| 25. MAA/MAA | 0 | 2 | 3 | 12 | 30 |

used in [70]. Rule 1 in Table 7.1 is used below as an example to illustrate the details of these three steps. The number of responses before pre-processing are shown in the first row of Table 7.4.

*1) Bad Data Processing:* This removes gaps (a zero between two non-zero values) in a group of subject's responses. For Rule 1 in Table 7.1, the number of responses to the five consequents are $\{42, 3, 2, 0, 0\}$. Because there is no gap among these numbers, no response is removed, as shown in the second row of Table 7.4. On the other hand, for Rule 2 in Table 7.1, the numbers of responses to the five consequents are $\{33, 12, 0, 2, 0\}$. Observe that no respondent selected the word *MOA* between *S* and *LA*; hence, a gap exists between *S* and *LA*. Let $G_1 = \{NVL,\ S\}$ and $G_2 = \{LA\}$. Because $G_1$ has more responses than $G_2$, it is passed to the next step of data pre-processing and $G_2$ is discarded.

*2) Outlier processing:* Outlier processing uses a Box and Whisker test [136]. As explained in [70], outliers are points that are unusually too large or too small. A Box and Whisker test is usually stated in terms of first and third quartiles and an interquartile range. The first and third quartiles, $Q(0.25)$ and $Q(0.75)$, contain 25% and 75% of the data, respectively. The inter-quartile range, $IQR$, is the difference between the third and first quartiles; hence, $IQR$ contains 50% of the data between the first and third quartiles. Any datum that is more than $1.5\,IQR$ above the third quartile or more than $1.5\,IQR$ below the first quartile is considered an outlier [136].

Rule consequents are words modeled by IT2 FSs; hence, the Box and Whisker test cannot be directly applied to them. In our approach, the Box and Whisker test is applied to the *set of centers of centroids* formed by the centers of centroids of the rule consequents. Focusing again on Rule 1 in Table 7.1, the centers of centroids of the consequent IT2 FSs *NVL, S, MOA, LA* and *MAA* are first obtained Table 2.1, and are 0.48, 3.91, 4.95, 8.13 and 9.69, respectively. Then the set of centers of centroids is

$$\{\underbrace{0.48, \cdots, 0.48}_{42}, \underbrace{3.91, 3.91, 3.91}_{3}, \underbrace{4.95, 4.95}_{2}\} \tag{7.1}$$

where each center of centroid is repeated a certain number of times according to the number of respondents after bad data processing. The Box and Whisker test is then applied to this crisp set, where $Q(0.25) = 0.48$, $Q(0.75) = 0.48$, and $1.5\,IQR = 0$. For Rule 1, the three responses to $S$ and the two responses to $MOA$ are removed, as shown in the third row of Table 7.4. The new set of centers of centroids becomes

$$\{\underbrace{0.48, \cdots, 0.48}_{42}\} \tag{7.2}$$

*3) Tolerance limit processing:* Let $m$ and $\sigma$ be the mean and standard deviation of the remaining histogram data after outlier processing. If a datum lies in the tolerance interval $[m - k\sigma, m + k\sigma]$, then it is accepted; otherwise, it is rejected [136]. $k$ is determined such that one is 95% confident that the given limits contain at least 95% of the available data, and it can be obtained from a table look-up [94].

For Rule 1 in Table 7.1, tolerance limit processing is performed on the set of 42 centers of centroids in (7.2), for which $m = 0.48$, $\sigma = 0$ and $k = 2.43$. No word is removed for this particular example; so, only one consequent, *NVL*, is accepted for this rule, as shown in the last row of Table 7.4.

The final pre-processed responses for the histograms in Tables 7.1, 7.2 and 7.3 are given in Tables 7.5, 7.6 and 7.7, respectively. Comparing each pair of tables, observe that most responses have been preserved.

Table 7.4: Data pre-processing results for the 47 responses to the question *"IF there is* NVL *touching, THEN there is _____ flirtation."*

| Number of responses | NVL | S | MOA | LA | MAA |
|---|---|---|---|---|---|
| Before pre-processing | 42 | 3 | 2 | 0 | 0 |
| After bad data processing | 42 | 3 | 2 | 0 | 0 |
| After outlier processing | 42 | 0 | 0 | 0 | 0 |
| After tolerance limit processing | 42 | 0 | 0 | 0 | 0 |

Table 7.5: Pre-processed histograms of Table 7.1.

| **Touching** | **Flirtation** | | | | |
|---|---|---|---|---|---|
| | **NVL** | **S** | **MOA** | **LA** | **MAA** |
| 1. NVL | 42 | 0 | 0 | 0 | 0 |
| 2. S | 33 | 12 | 0 | 0 | 0 |
| 3. MOA | 12 | 16 | 15 | 3 | 0 |
| 4. LA | 0 | 6 | 11 | 25 | 2 |
| 5. MAA | 0 | 6 | 8 | 22 | 8 |

Table 7.6: Pre-processed histograms of Table 7.2.

| **Eye Contact** | **Flirtation** | | | | |
|---|---|---|---|---|---|
| | **NVL** | **S** | **MOA** | **LA** | **MAA** |
| 1. NVL | 36 | 0 | 0 | 0 | 0 |
| 2. S | 26 | 17 | 4 | 0 | 0 |
| 3. MOA | 0 | 16 | 27 | 0 | 0 |
| 4. LA | 0 | 3 | 11 | 22 | 10 |
| 5. MAA | 0 | 0 | 0 | 17 | 20 |

Table 7.7: Pre-processed histograms of Table 7.3.

| Touching/Eye Contact | Flirtation | | | | |
|---|---|---|---|---|---|
| | **NVL** | **S** | **MOA** | **LA** | **MAA** |
| 1. NVL/NVL | 38 | 0 | 0 | 0 | 0 |
| 2. NVL/S | 33 | 11 | 3 | 0 | 0 |
| 3. NVL/MOA | 0 | 21 | 16 | 0 | 0 |
| 4. NVL/LA | 0 | 12 | 28 | 0 | 0 |
| 5. NVL/MAA | 0 | 9 | 16 | 19 | 3 |
| 6. S/NVL | 31 | 11 | 4 | 0 | 0 |
| 7. S/S | 17 | 23 | 7 | 0 | 0 |
| 8. S/MOA | 0 | 19 | 19 | 0 | 0 |
| 9. S/LA | 0 | 8 | 23 | 13 | 2 |
| 10. S/MAA | 0 | 7 | 17 | 21 | 2 |
| 11. MOA/NVL | 0 | 23 | 16 | 0 | 0 |
| 12. MOA/S | 0 | 22 | 20 | 0 | 0 |
| 13. MOA/MOA | 0 | 7 | 22 | 15 | 1 |
| 14. MOA/LA | 0 | 4 | 13 | 17 | 12 |
| 15. MOA/MAA | 0 | 4 | 12 | 24 | 7 |
| 16. LA/NVL | 0 | 13 | 21 | 0 | 0 |
| 17. LA/S | 0 | 11 | 23 | 0 | 0 |
| 18. LA/MOA | 0 | 3 | 18 | 18 | 8 |
| 19. LA/LA | 0 | 0 | 0 | 17 | 20 |
| 20. LA/MAA | 0 | 0 | 0 | 11 | 27 |
| 21. MAA/NVL | 0 | 16 | 18 | 11 | 0 |
| 22. MAA/S | 0 | 9 | 22 | 13 | 1 |
| 23. MAA/MOA | 0 | 3 | 15 | 18 | 11 |
| 24. MAA/LA | 0 | 0 | 0 | 17 | 22 |
| 25. MAA/MAA | 0 | 0 | 0 | 12 | 30 |

## 7.3 Rulebase Generation

Observe from Tables 7.5, 7.6 and 7.7 that the survey and data pre-processing lead to rule consequent histograms, but how the histograms should be used is an open question. In [83] three possibilities were proposed:

1. Keep the response chosen by the largest number of respondents.
2. Find a weighted average of the rule consequents for each rule.
3. Preserve the distributions of the expert-responses for each rule.

Clearly, the disadvantage of keeping the response chosen by the largest number of respondents is that this ignores all the other responses.

The second method was studied in detail in [83]. Using that method, when T1 FSs were used (see Chapter 5 of [83]), the consequent for each rule was a crisp number, $c$, where

$$c = \frac{\sum_{m=1}^{5} c_m w_m}{\sum_{m=1}^{5} w_m} \tag{7.3}$$

in which $c_m$ is the centroid [83] of the $m^{\text{th}}$ T1 consequent FS, and $w_m$ is the number of respondents for the $m^{\text{th}}$ consequent. When IT2 FSs were used (see Chapter 10 of [83]), the consequent for each rule was an interval, $C$, where

$$C = \frac{\sum_{m=1}^{5} C_m w_m}{\sum_{m=1}^{5} w_m} \tag{7.4}$$

in which $C_m$ is the centroid [57, 83] of the $m^{\text{th}}$ IT2 consequent FS.

The disadvantages of using (7.3) or (7.4) are: (1) there is information lost when converting the T1 or IT2 consequent FSs into their centroids, and (2) it is difficult to describe the aggregated rule consequents ($c$ or $C$) linguistically.

Our approach is to preserve the distributions of the expert-responses for each rule by using a different weighted average to obtain the rule consequents, as illustrated by the following:

**Example 26** *Observe from Table 7.5 that when the antecedent is* some *(S) there are two valid consequents, so that the following two rules will be fired:*

$R_1^2$: *IF touching is* some*, THEN flirtation is* none to very little*.*

$R_2^2$: *IF touching is* some*, THEN flirtation is* some*.*

*These two rules should not be considered of equal importance because they have been selected by different numbers of respondents. An intuitive way to handle this is to assign weights to the two rules, where the weights are proportional to the number of responses, e.g., the weight for* $R_1^2$ *is* $33/45 = 0.73$*, and the weight for* $R_2^2$ *is* $12/45 = 0.27$*. The aggregated consequent* $\tilde{Y}^2$ *for* $R_1^2$ *and* $R_2^2$ *is*

$$\tilde{Y}^2 = \frac{33NVL + 12S}{33 + 12} \tag{7.5}$$

*The result is shown in Fig. 7.2.* ■

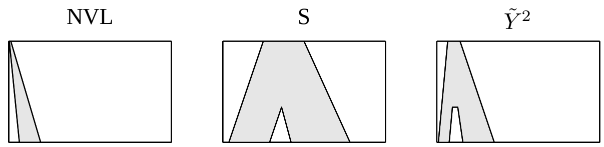



Fig. 7.2: $\tilde{Y}^2$ obtained by aggregating the consequents of $R_1^2$ (NVL) and $R_2^2$ (S).

Without loss of generality, assume there are $N$ different combinations of antecedents (e.g., $N = 5$ for the single-antecedent rules in Tables 7.5 and 7.6, and $N = 25$ for the two-antecedent rules in Table 7.7), and each combination has $M$ possible different consequents (e.g., $M = 5$ for the rules in Tables 7.5-7.7); hence, there can be as many as $MN$ rules. Denote the $m^{\text{th}}$ consequent of the $i^{\text{th}}$ combination of the antecedents as $\tilde{Y}_m^i$ ($m = 1, 2, \ldots, M, i = 1, 2, \ldots, N$), and the number of responses to $\tilde{Y}_m^i$ as $w_m^i$. For each $i$, all $M$ $\tilde{Y}_m^i$ can be combined first into a single IT2 FS by using the algorithm given in the Appendix:

$$\tilde{Y}^i = \frac{\sum_{m=1}^{M} w_m^i \tilde{Y}_m^i}{\sum_{m=1}^{M} w_m^i} \tag{7.6}$$

$\tilde{Y}^i$ then acts as the (new) consequent for the $i^{\text{th}}$ rule. By doing this, the distribution of the expert responses has been preserved for each rule. Examples of $\tilde{Y}^i$ for single-antecedent and two-antecedent rules are depicted in Figs. 7.3(a), 7.4(a) and 7.5, and are described in detail next.

## 7.4 Single-Antecedent Rules: Touching and Flirtation

In this section, the rulebase for a single-antecedent SJA, which describes the relationship between touching and flirtation, is constructed. The resulting SJA is denoted $SJA_1$.

When (7.6) is used to combine the different responses for each antecedent into a single consequent for the rule data in Table 7.5, one obtains the rule consequents depicted in Fig. 7.3(a). As a comparison, the rule consequents obtained from the original rule data in Table 7.1 are depicted in Fig. 7.3(b). Observe that:

1. The consequent for *none to very little* (NVL) touching is a left-shoulder in Fig. 7.3(a), whereas it is an interior FOU in Fig. 7.3(b). The former seems more reasonable to us.

2. The consequent for *some* (S) touching in Fig. 7.3(a) is similar to that in Fig. 7.3(b), except that it is shifted a little to the left. This is because the two largest responses [*large amount* (LA)] in Table 7.1 are removed in pre-processing.

3. The consequent for *moderate amount* (MOA) touching in Fig. 7.3(a) is similar to that in Fig. 7.3(b), except that it is shifted a little to the left. This is because the largest response [*maximum amount* (MAA)] in Table 7.1 is removed in pre-processing.

4. The consequent for *large amount* (LA) is similar to that in Fig. 7.3(b), except that it is shifted a little to the right. This is because the three smallest responses [*none to very little* (NVL)] in Table 7.1 are removed in pre-processing.

5. The consequent for *maximum amount* (MAA) is similar to those in Fig. 7.3(b), except that it is shifted a little to the right. This is because the three smallest responses [*none to very little* (NVL)] in Table 7.1 are removed in pre-processing.

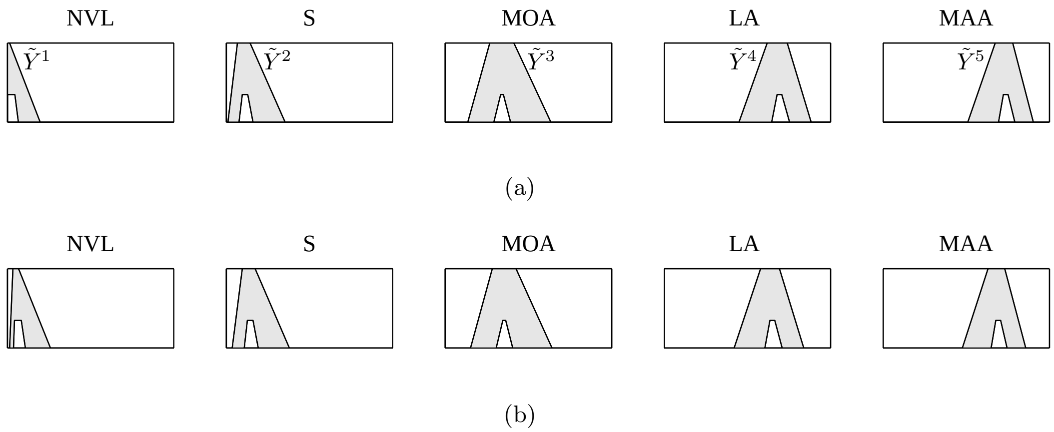



Fig. 7.3: Flirtation-level consequents of the five rules for the single-antecedent *touching* $SJA_1$: (a) with data pre-processing and (b) without data pre-processing. The level of touching is indicated at the top of each figure.

The consequents $\tilde{Y}^1$–$\tilde{Y}^5$ shown in Fig. 7.3(a) are used in the rest of this section for the consensus $SJA_1$. Its five-rule rulebase is

$R^1$: IF touching is *NVL*, THEN flirtation is $\tilde{Y}^1$.

$R^2$: IF touching is *S*, THEN flirtation is $\tilde{Y}^2$.

$R^3$: IF touching is *MOA*, THEN flirtation is $\tilde{Y}^3$.

$R^4$: IF touching is *LA*, THEN flirtation is $\tilde{Y}^4$.

$R^5$: IF touching is *MAA*, THEN flirtation is $\tilde{Y}^5$.

## 7.5 Single-Antecedent Rules: Eye Contact and Flirtation

In this section, the rulebase for another single-antecedent SJA, which describes the relationship between eye contact and flirtation, is constructed. The resulting SJA is denoted $SJA_2$.

When (7.6) is used to combine the different responses for each antecedent into a single consequent for the rule data in Table 7.6, one obtains the rule consequents depicted in Fig. 7.4(a). As a comparison, the rule consequents obtained from the original rule data in Table 7.2 are depicted in Fig. 7.4(b). The rule consequents for *NVL, MOA, LA* and *MAA* are different in these two figures. The consequents in Fig. 7.4(a) are used by $SJA_2$.

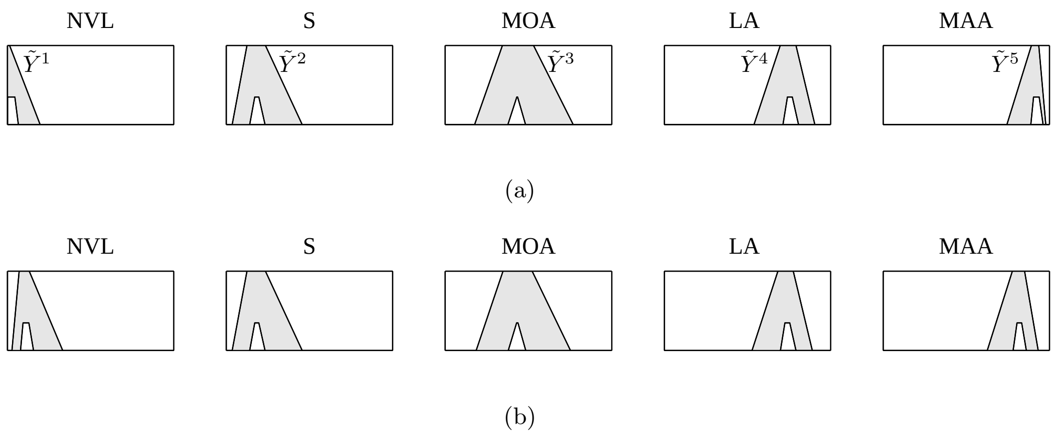



Fig. 7.4: Flirtation-level consequents of the five rules for the single-antecedent *eye contact* $SJA_2$: (a) with data pre-processing and (b) without data pre-processing. The level of eye contact is indicated at the top of each figure.

## 7.6 Two-Antecedent Rules: Touching/Eye Contact and Flirtation

The previous two sections have considered single antecedent rules. This section considers two-antecedent (touching and eye contact) rules, whose corresponding SJA is denoted $SJA_3$.

When (7.6) is used to combine the different responses for each pair of antecedents into a single consequent for the rule data in Table 7.7, one obtains the 25 rule consequents $\tilde{Y}^{1,1}$–$\tilde{Y}^{5,5}$ depicted in Fig. 7.5. Observe that the rule consequent becomes larger (i.e., moves towards right in the [0,10] interval) as either input increases, which is intuitive.

The 25-rule rulebase of $SJA_3$ is:

$R^{1,1}$: IF touching is $NVL$ and eye contact is $NVL$, THEN flirtation is $\tilde{Y}^{1,1}$.

$\vdots$

$R^{1,5}$: IF touching is $NVL$ and eye contact is $MAA$, THEN flirtation is $\tilde{Y}^{1,5}$.

$\vdots$

$R^{5,1}$: IF touching is $MAA$ and eye contact is $NVL$, THEN flirtation is $\tilde{Y}^{5,1}$.

$\vdots$

$R^{5,5}$: IF touching is $MAA$ and eye contact is $MAA$, THEN flirtation is $\tilde{Y}^{5,5}$.

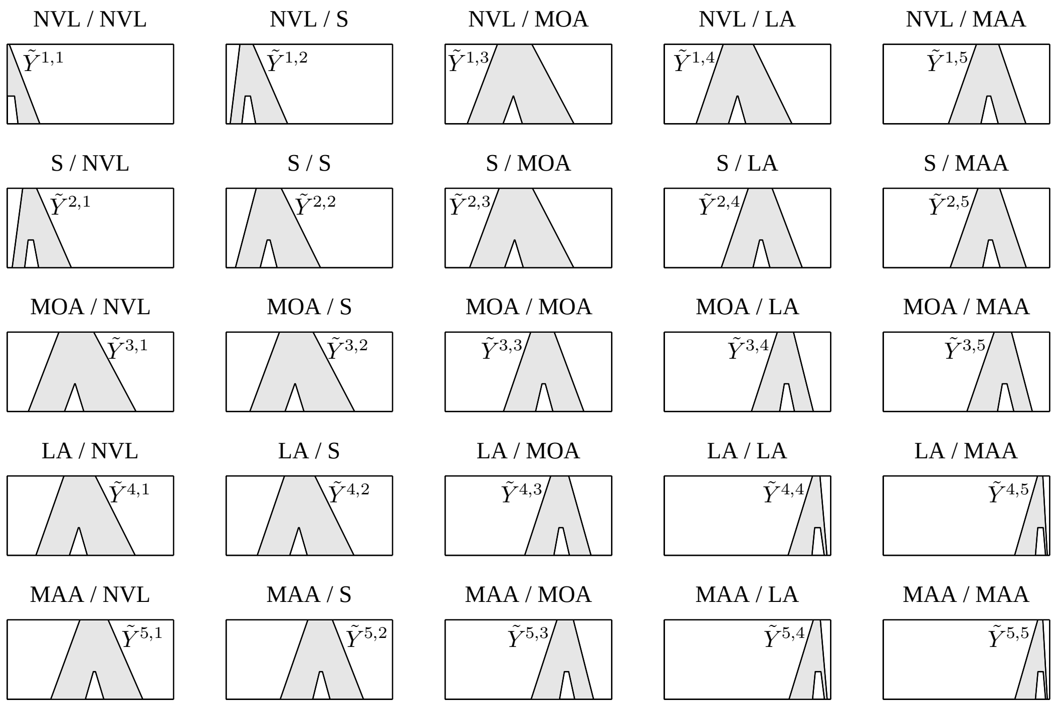



Fig. 7.5: Flirtation-level consequents of the 25 rules for the two-antecedent consensus $SJA_3$ with data pre-processing. The levels of touching and eye contact are indicated at the top of each figure.

## 7.7 Comparisons with Other Approaches

A prevailing paradigm for examining social judgments would be to examine the influence of various factors on the variable of interest using linear approaches, e.g., linear regression. Unfortunately, perceptions regarding the variable of interest may not be linear, but rather step-like. A linear model is unable to capture such non-linear changes, whereas the Per-C is able to do this because of its non-linear nature. In summary, the main differences between linear approaches and an SJA are [88]:

1. The former are determined only from numerical data (e.g., regression coefficients are fitted to numerical data) whereas the SJA is determined from linguistic information, i.e. a collection of IF-THEN rules that are provided by people.

2. The rules, when properly collected, convey the details of a *nonlinear* relationship between the antecedents of the rule and the consequent of the rule.

3. An SJA can directly quantify a linguistic rule and can provide a linguistic output; a regression model cannot do this.

4. Regression models can, however, include nonlinear regressors (e.g., interaction terms), which make them also nonlinear functions of their inputs; however, the structure of the nonlinearities in the SJA is not pre-specified, as it must be for a regression model; it is a direct result of the mathematics of the SJA.

5. An SJA is a *variable structure model*, in that it simultaneously provides excellent local and global approximations to social judgments, whereas a regression model

can only provide global approximations to social judgments. By “variable structure model” is meant that only a (usually) small subset of the rules are fired for a given set of inputs, and when the inputs change so does the subset of fired rules, and this happens automatically because of the mathematics of the SJA.

6. The way in which uncertainty is dealt with. Typically, in a linear regression model individuals are forced to translate their assessment into absolute numbers, e.g., 1, 2, 3, 4. In contrast a person can interact with the SJA using normal linguistic phrases, e.g., about eye contact (one of the indicators of flirtation), such as “eye contact is moderate.”

7. Finally, if determining the level of flirtation were easy, we would all be experts; but it is not, and we are not. In fact, many times we get “mixed signals.” Fuzzy logic leads to an explanation and potential resolution of “mixed signals,” though the simultaneous firing of more than one rule, each of which may have a different consequent. So the SJA also provides us with *insight* into why determining whether or not we are being flirted with is often difficult. The same should also be true for other social judgments. We do not believe that this is possible using a regression model.

The flirtation advisor was also studied in [83], and it is called a fuzzy logic advisor (FLA). Two kinds of FLAs were designed: a T1 FLA (see Chapter 5 of [83]; it is similar to the T1 FLA designed in [82]) which uses only T1 FSs and an IT2 FLA (see Chapter 10 of [83]) which uses IT2 FSs. The differences from these two FLAs and our SJA are:

1. Surveys were used to obtain the interval end-points of the *words* used in the rules for all three approaches; however, no pre-processing was used to removed bad data and outliers for the two FLAs. Additionally, for the T1 FLA, only the means and the standard deviations (stds) of the end-points were used to construct T1 FS models for the words, and for the IT2 FLA, an ad hoc *fraction of uncertainty* was used to blur the T1 FS word models into IT2 FSs. Both modeling approaches are not as intuitive as the Interval Approach used in the SJA, where interval end-points data are pre-processed, each interval is mapped into a T1 FS, and all T1 FSs are then combined using the Representation Theorem [81] to obtain the FOU for each word.

2. Weighted averages were used to combine multiple responses of a *rule* into a single consequent in all three approaches; however, no pre-processing was used to remove bad responses and outliers for the two FLAs. Additionally, for the T1 FLA, (7.3) was used to compute the consequent, the result being a crisp number, and, for the IT2 FLA, (7.4) was used to compute the consequent, the result being an interval. Both kinds of consequents can no longer be viewed as words. On the other hand, the SJA used a three-step pre-processing procedure to remove bad responses and outliers. It then used an LWA to combine the responses, the result being an FOU that resembles the three kinds of word FOUs in the codebook so that it can be easily mapped into a word in that codebook.

3. For both FLAs, their inputs can only be crisp numbers in [0, 10] instead of words, and their outputs are crisp numbers (for the T1 FLA) or intervals (for the IT2 FLA).

There is no procedure to map these numbers back into words; so, the two FLAs are not performing CWW, which we view as a mapping of words into a recommendation. On the other hand, as we will shown in Section 8.4, both the inputs and outputs of the SJA are words.

In summary, our approach for designing the SJA is significantly different from a linear regression model and the FLAs. Presently, it is the only approach that enable us to map words into a recommendation, as will be shown in Section 8.4.

# Chapter 8

# Perceptual Reasoning as a CWW Engine for MODM

One of the most popular CWW engines uses IF-THEN rules [see (2.12)]. This chapter is about such rules and how they are processed within a CWW engine so that their outputs can be mapped into a word-recommendation by the decoder. This use of IF-THEN rules is quite different from their use in most engineering applications of rule-based systems — FLSs — because in a FLS the output almost always is a number, whereas the output of the Per-C is a recommendation. This distinction imposes the following important requirement on the output of a CWW engine using IF-THEN rules:

*Requirement: The result of combining fired rules should lead to an FOU that resembles the three kinds of FOUs in a CWW codebook.*

## 8.1 Traditional Inference Engines

Approximate reasoning using the Mamdani inference has been described in Section 2.2.4. Two points about it are worth emphasizing:

1. *Each fired-rule output FOU does not resemble the FOU of a word in the Per-C codebook* (Fig. 2.12). This is because the meet operation between the firing interval and its consequent FOU results in an FOU whose lower and upper MFs are clipped versions of the respective lower and upper MFs of a consequent FOU.

2. *The aggregated fired rule output FOU also does not resemble the FOU of a word in the Per-C codebook.* This is because when the union operator is applied to all of the fired rule output FOUs it further distorts those already-distorted FOUs.

Mamdani inference does not let us satisfy this requirement; hence, we turn to an alternative that is widely used by practitioners of FLSs, one that blends *attributes* about the fired rule consequent IT2 FSs with the firing quantities.

Attributes of a fired rule consequent IT2 FS include its centroid and the point of symmetry of its FOU (if the FOU is symmetrical). The *blending* is accomplished directly by the kind of type-reduction that is chosen, e.g., center-of-sets type-reduction makes use of the centroids of the consequents, whereas height type-reduction makes use of the point of symmetry of each consequent FOU. Regardless of the details of this kind of type-reduction-blending[1], the type-reduced result is an interval-valued set after which that interval is defuzzified as before by taking the average of the interval's two end-points.

[1] More details about type-reduction can be found in [83].

It is worth noting that by taking this alternative approach there is no associated FOU for either each fired rule or all of the fired rules; hence, there is no FOU obtained from this approach that can be compared with the FOUs in the codebook. Consequently, using this alternative to Mamdani inference also does not let us satisfy the Requirement.

By these lines of reasoning we have ruled out the two usual ways in which rules are fired and combined for use by the Per-C.

## 8.2 Perceptual Reasoning: Computation

Perceptual reasoning[2] (PR) [85,86,149,156], which satisfies the Requirement, is introduced in this section.

Let $\tilde{\mathbf{X}}'$ denote an $N \times 1$ vector of IT2 FSs that are the inputs to a collection of $N$ rules, as would be the case when such inputs are words. $f^i(\tilde{\mathbf{X}}')$ denotes the firing level for the $i^{\text{th}}$ rule, and it is computed only for the $n \leq N$ number of fired rules, i.e., the rules whose firing levels do not equal zero. In PR, the fired rules are combined using a LWA. Denote

[2]Perceptual Reasoning is a term coined in [86] because it is used by the Per-C when the CWW Engine consists of IF-THEN rules. In [86] firing intervals are used to combine the rules; however, firing levels are used in this dissertation because, as shown in [156], they give an output FOU which more closely resembles the word FOUs in a codebook.

the output IT2 FS of PR as $\tilde{Y}_{PR}$. Then, $\tilde{Y}_{PR}$ can be written in the following expressive[3] way:

$$\tilde{Y}_{PR} = \frac{\sum_{i=1}^{n} f^i(\tilde{\mathbf{X}}')\tilde{G}^i}{\sum_{i=1}^{n} f^i(\tilde{\mathbf{X}}')} \tag{8.1}$$

This LWA is a special case of the more general LWA (Section 4.4) in which both $\tilde{G}^i$ and $f^i(\tilde{\mathbf{X}}')$ were IT2 FSs.

Observe that PR consists of two steps:

1. A *firing level* is computed for each rule, and
2. The IT2 FS consequents of the fired rules are combined using an LWA in which the "weights" are the firing levels and the "sub-criteria" are the IT2 FS consequents.

### 8.2.1 Computing Firing Levels

Similarity is frequently used in Approximate Reasoning to compute the firing levels [12, 106, 174], and it can also be used in PR to do this.

Let the $p$ inputs that activate a collection of $N$ rules be denoted $\tilde{\mathbf{X}}'$. The result of the input and antecedent operations for the $i^{\text{th}}$ fired rule is the firing level $f^i(\tilde{\mathbf{X}}')$, where

$$f^i(\tilde{\mathbf{X}}') = s_J(\tilde{X}_1, \tilde{F}_1^i) \star \cdots \star s_J(\tilde{X}_p, \tilde{F}_p^i) \equiv f^i \tag{8.2}$$

[3] As in Section 4.4.1, (8.1) is referred to as "expressive" because it is not computed using multiplications, additions and divisions, as expressed by it. Instead, $\underline{Y}_{PR}$ and $\overline{Y}_{PR}$ are computed separately using $\alpha$-cuts, as explained in Section 8.2.2.

where $s_J(\tilde{X}_j, \tilde{F}^i_j)$ is the Jaccard's similarity measure for IT2 FSs [see (3.4)], and $\star$ denotes a $t$-norm. The minimum $t$-norm is used in (8.2).

*Comment:* To use PR, we need a *codebook*[4] consisting of words and their associated FOUs so that a user can choose inputs from it. Once this codebook is obtained, the similarities between the input words and the antecedent words of the rules (i.e., $s_J(\tilde{X}_j, \tilde{F}^i_j)$, $j = 1, \ldots, p$, $i = 1, \ldots, N$) can be pre-computed and stored in a table (e.g., the similarity matrix shown in Table 3.1), so that $s_J(\tilde{X}_j, \tilde{F}^i_j)$ can be retrieved online to save computational cost.

### 8.2.2 Computing $\tilde{Y}_{PR}$

$\tilde{Y}_{PR}$ in (8.1) is a special case of the more general LWA introduced in Section 4.4. The formulas for this special case have been presented in Section 4.5, except that different notations were used. Because PR is widely used in the rest of this dissertation, these formulas are repeated here using the notations in this chapter.

An interior FOU for rule consequent $\tilde{G}^i$ is depicted in Fig. 8.1(a), in which the height of $\underline{G}^i$ is denoted $h_{\underline{G}^i}$, the $\alpha$-cut on $\underline{G}^i$ is denoted $[a_{ir}(\alpha), b_{il}(\alpha)]$, $\alpha \in [0, h_{\underline{G}^i}]$, and the $\alpha$-cut on $\overline{G}^i$ is denoted $[a_{il}(\alpha), b_{ir}(\alpha)]$, $\alpha \in [0, 1]$. For the left shoulder $\tilde{G}^i$ depicted in Fig. 8.1(b), $h_{\underline{G}^i} = 1$ and $a_{il}(\alpha) = a_{ir}(\alpha) = 0$ for $\forall \alpha \in [0, 1]$. For the right-shoulder $\tilde{G}^i$ depicted in Fig. 8.1(c), $h_{\underline{G}^i} = 1$ and $b_{il}(\alpha) = b_{ir}(\alpha) = M$ for $\forall \alpha \in [0, 1]$.

[4] The words used in the antecedents of the rules, as will the words that excite the rules, are always included in this codebook.

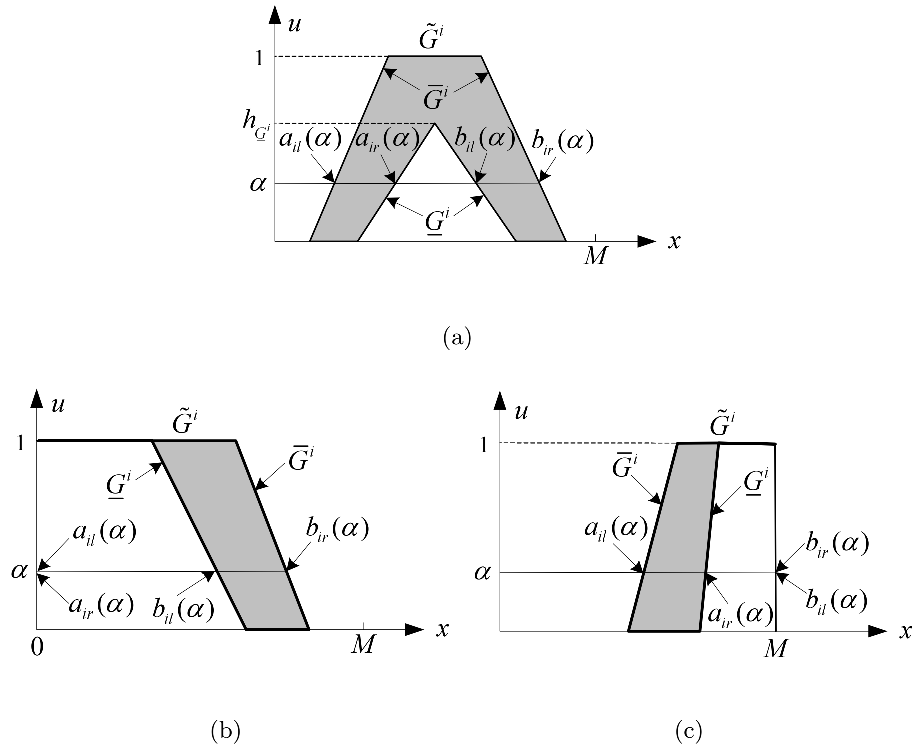



Fig. 8.1: Typical word FOUs and an $\alpha$-cut. (a) Interior, (b) left-shoulder, and (c) right-shoulder FOUs.

Because the output of PR must resemble the three kinds of FOUs in a codebook, $\tilde{Y}_{PR}$ can also be an interior, left shoulder or right shoulder FOU, as shown in Fig. 8.2 (this is actually proved in Section 8.3.2). The $\alpha$-cut on $\overline{Y}_{PR}$ is $[y_{Ll}(\alpha), y_{Rr}(\alpha)]$ and the $\alpha$-cut on $\underline{Y}_{PR}$ is $[y_{Lr}(\alpha), y_{Rl}(\alpha)]$, where, as explained in Section 4.5, the end-points of these $\alpha$-cuts are computed for (8.1) as:

$$y_{Ll}(\alpha) = \frac{\sum_{i=1}^{n} a_{il}(\alpha) f^i}{\sum_{i=1}^{n} f^i}, \quad \alpha \in [0,1] \tag{8.3}$$

$$y_{Rr}(\alpha) = \frac{\sum_{i=1}^{n} b_{ir}(\alpha) f^i}{\sum_{i=1}^{n} f^i}, \quad \alpha \in [0,1] \tag{8.4}$$

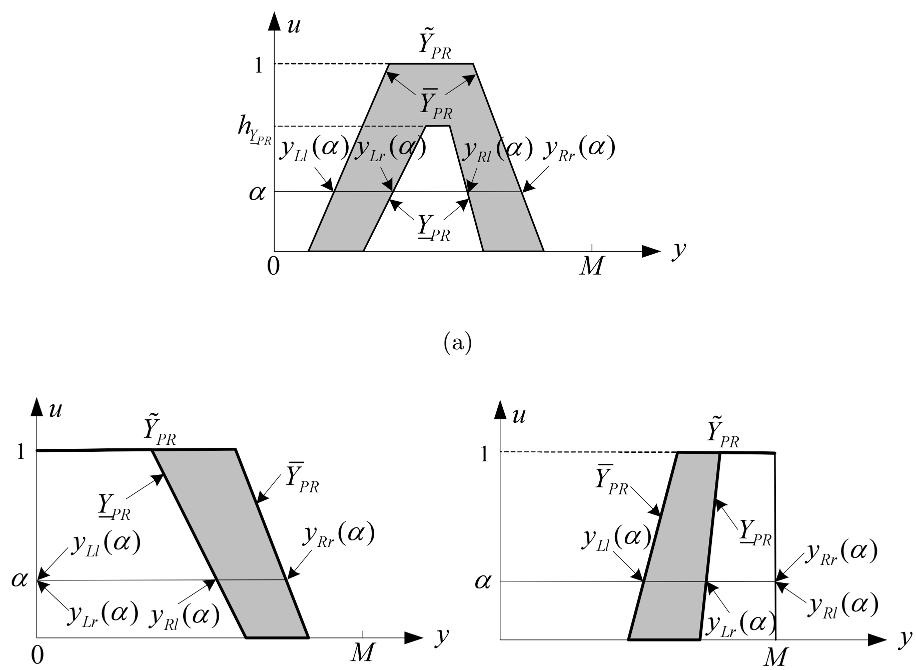
$$y_{Lr}(\alpha) = \frac{\sum_{i=1}^{n} a_{ir}(\alpha) f^i}{\sum_{i=1}^{n} f^i}, \quad \alpha \in [0, h_{\underline{Y}_{PR}}] \tag{8.5}$$

$$y_{Rl}(\alpha) = \frac{\sum_{i=1}^{n} b_{il}(\alpha) f^i}{\sum_{i=1}^{n} f^i}, \quad \alpha \in [0, h_{\underline{Y}_{PR}}] \tag{8.6}$$

where

$$h_{\underline{Y}_{PR}} = \min_i h_{\underline{G}^i} \tag{8.7}$$

Note that (8.3)-(8.6) are arithmetic weighted averages, so they are computed directly without using KM or EKM algorithms.




(a)

(b)

(c)

Fig. 8.2: PR FOUs and $\alpha$-cuts on (a) interior, (b) left-shoulder, and (c) right-shoulder FOUs.

Observe from (8.3) and (8.4) that $\tilde{Y}_{PR}$ is always normal, i.e., its $\alpha = 1$ $\alpha$-cut can always be computed. This is different from many other Approximate Reasoning methods, whose aggregated fired-rule output sets are not normal, e.g., the Mamdani-inference based method. For the latter, even if only one rule is fired (see Fig. 2.10), unless the firing level is one, the output is a clipped or scaled version[5] of the consequent IT2 FS instead of a normal IT2 FS. This may cause problems when the output is mapped to a word in the codebook.

In summary, knowing the firing levels $f^i$, $i = 1, ..., n$, $\overline{Y}_{PR}$ is computed in the following way:

1. Select $m$ appropriate $\alpha$-cuts for $\overline{Y}_{PR}$ (e.g., divide [0, 1] into $m-1$ intervals and set $\alpha_j = (j-1)/(m-1)$, $j = 1, 2, ..., m$).
2. For each $\alpha_j$, find the $\alpha$-cut $[a_{il}(\alpha_j), b_{ir}(\alpha_j)]$ on $\overline{G}^i$ ($i = 1, ..., n$) and compute $y_{Ll}(\alpha_j)$ in (8.3) and $y_{Rr}(\alpha_j)$ in (8.4).
3. Connect all left-coordinates $(y_{Ll}(\alpha_j), \alpha_j)$ and all right-coordinates $(y_{Rr}(\alpha_j), \alpha_j)$ to form the T1 FS $\overline{Y}_{PR}$.

Similarly, to compute $\underline{Y}_{PR}$:

1. Determine $h_{\underline{X}_i}$, $i = 1, \ldots, n$, and $h_{\min}$ in (8.7).
2. Select appropriate $p$ $\alpha$-cuts for $\underline{Y}_{PR}$ (e.g., divide $[0, h_{\min}]$ into $p-1$ intervals and set $\alpha_j = h_{\min}(j-1)/(p-1)$, $j = 1, 2, ..., p$).

[5] A scaled version of the consequent IT2 FS occurs when the product $t$-norm is used to combine the firing level and the consequent IT2 FS.

3. For each $\alpha_j$, find the $\alpha$-cut $[a_{ir}(\alpha_j), b_{il}(\alpha_j)]$ on $\underline{G}^i$ $(i = 1, ..., n)$ and compute $y_{Lr}(\alpha_j)$ in (8.5) and $y_{Rl}(\alpha_j)$ in (8.6).

4. Connect all left-coordinates $(y_{Lr}(\alpha_j), \alpha_j)$ and all right-coordinates $(y_{Rl}(\alpha_j), \alpha_j)$ to form the T1 FS $\underline{Y}_{PR}$.

## 8.3 Perceptual Reasoning: Properties

Properties of PR are presented in this section. All of them help demonstrate the Requirement for PR, namely, *the result of combining fired rules using PR leads to an IT2 FS that resembles the three kinds of FOUs in a CWW codebook.*

### 8.3.1 General Properties About the Shape of $\tilde{Y}_{PR}$

In this section, some general properties are provided that are about the shape of $\tilde{Y}_{PR}$. These general properties are used in Section 8.3.2.

**Theorem 9** *When all fired rules have the same consequent $\tilde{G}$, $\tilde{Y}_{PR}$ defined in (8.1) is the same as $\tilde{G}$.* ■

*Proof:* When all fired rules have the same consequent $\tilde{G}$, (8.1) simplifies to

$$\tilde{Y}_{PR} = \frac{\sum_{i=1}^{n} f^i \tilde{G}}{\sum_{i=1}^{n} f^i} = \tilde{G}\left(\frac{\sum_{i=1}^{n} f^i}{\sum_{i=1}^{n} f^i}\right) = \tilde{G}. \quad \blacksquare \tag{8.8}$$

Although Theorem 9 is true regardless of how many rules are fired, its most interesting application occurs when only one rule is fired, in which case the output from PR is the consequent FS, $\tilde{G}$, and $\tilde{G}$ resides in the codebook. On the other hand, when one rule fires, the output from Mamdani inferencing is a clipped version of $\tilde{G}$, $\tilde{B}$, as depicted in Fig. 2.10, and $\tilde{B}$ does not reside in the codebook.

**Theorem 10** *$\tilde{Y}_{PR}$ is constrained by the consequents of the fired rules, i.e.,*

$$\min_i a_{il}(\alpha) \leq y_{Ll}(\alpha) \leq \max_i a_{il}(\alpha) \tag{8.9}$$

$$\min_i a_{ir}(\alpha) \leq y_{Lr}(\alpha) \leq \max_i a_{ir}(\alpha) \tag{8.10}$$

$$\min_i b_{il}(\alpha) \leq y_{Rl}(\alpha) \leq \max_i b_{il}(\alpha) \tag{8.11}$$

$$\min_i b_{ir}(\alpha) \leq y_{Rr}(\alpha) \leq \max_i b_{ir}(\alpha) \tag{8.12}$$

*where $a_{il}(\alpha)$, $a_{ir}(\alpha)$, $b_{il}(\alpha)$ and $b_{ir}(\alpha)$ are defined for three kinds of consequent FOUs in Fig. 8.1.* ■

*Proof:* Theorem 10 is obvious because each of $y_{Ll}(\alpha)$, $y_{Lr}(\alpha)$, $y_{Rl}(\alpha)$ and $y_{Rr}(\alpha)$ is an arithmetic weighted average of the corresponding quantities on $\tilde{G}^i$. So, e.g., from (8.3), observe that

$$y_{Ll}(\alpha) = \frac{\sum_{i=1}^n a_{il}(\alpha) f^i}{\sum_{i=1}^n f^i} \geq \frac{\min_i a_{il}(\alpha) \cdot \sum_{i=1}^n f^i}{\sum_{i=1}^n f^i} = \min_i a_{il}(\alpha) \tag{8.13}$$

$$y_{Ll}(\alpha) = \frac{\sum_{i=1}^n a_{il}(\alpha) f^i}{\sum_{i=1}^n f^i} \leq \frac{\max_i a_{il}(\alpha) \cdot \sum_{i=1}^n f^i}{\sum_{i=1}^n f^i} = \max_i a_{il}(\alpha) \quad \blacksquare \tag{8.14}$$

The equalities in (8.9)-(8.12) hold simultaneously if and only if all $n$ fired rules have the same consequent. A graphical illustration of Theorem 10 is shown in Fig. 8.3. Assume only two rules are fired and $\tilde{G}^1$ lies to the left of $\tilde{G}^2$; then, $\tilde{Y}_{PR}$ lies between $\tilde{G}^1$ and $\tilde{G}^2$.

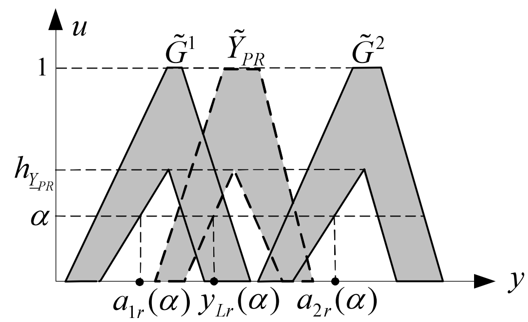



Fig. 8.3: A graphical illustration of Theorem 10, when only two rules fire.

Theorem 10 is about the location of $\tilde{Y}_{PR}$. Theorem 11 below is about the span of $\tilde{Y}_{PR}$; but first, the span of an IT2 FS is defined.

**Definition 35** *The* span *of the IT2 FS $\tilde{G}^i$ is $b_{ir}(0) - a_{il}(0)$, where $a_{il}(0)$ and $b_{ir}(0)$ are the left and right end-points of the $\alpha = 0$ $\alpha$-cut on $\bar{G}^i$, respectively.* ■

It is well-known from interval arithmetic that operations (e.g., $+$, $-$ and $\times$) on intervals usually spread out the resulting interval; however, this is not true for PR, as indicated by the following:

**Theorem 11** *The span of $\tilde{Y}_{PR}$, $y_{Rr}(0) - y_{Ll}(0)$, is constrained by the spans of the consequents of the fired rules, i.e.,*

$$\min_i (b_{ir}(0) - a_{il}(0)) \leq y_{Rr}(0) - y_{Ll}(0) \leq \max_i (b_{ir}(0) - a_{il}(0)). \quad \blacksquare \tag{8.15}$$

*Proof:* It follows from (8.3) and (8.4) that

$$\begin{aligned}
y_{Rr}(0) - y_{Ll}(0) &= \frac{\sum_{i=1}^{n}(b_{ir}(0) - a_{il}(0))f^i}{\sum_{i=1}^{n} f^i} \\
&\geq \frac{\min\limits_i(b_{ir}(0) - a_{il}(0)) \cdot \sum_{i=1}^{n} f^i}{\sum_{i=1}^{n} f^i} \\
&= \min_i(b_{ir}(0) - a_{il}(0)) \qquad (8.16) \\
y_{Rr}(0) - y_{Ll}(0) &= \frac{\sum_{i=1}^{n}(b_{ir}(0) - a_{il}(0))f^i}{\sum_{i=1}^{n} f^i} \\
&\leq \frac{\max\limits_i(b_{ir}(0) - a_{il}(0)) \cdot \sum_{i=1}^{n} f^i}{\sum_{i=1}^{n} f^i} \\
&= \max_i(b_{ir}(0) - a_{il}(0)). \quad \blacksquare \qquad (8.17)
\end{aligned}$$

Both equalities in (8.15) hold simultaneously if and only if all $n$ fired rules have the same span.

The following two definitions are about the shape of a T1 FS, and they are used in proving properties about the shape of $\tilde{Y}_{PR}$.

**Definition 36** *Let $A$ be a T1 FS and $h_A$ be its height. Then, $A$ is* trapezoid-looking *if its $\alpha = h_A$ $\alpha$-cut is an interval instead of a single point.* $\blacksquare$

$\underline{Y}_{PR}$ and $\bar{Y}_{PR}$ in Fig. 8.2(a) are trapezoid-looking.

**Definition 37** *Let $A$ be a T1 FS and $h_A$ be its height. Then, $A$ is* triangle-looking *if its $\alpha = h_A$ $\alpha$-cut consists of a single point.* $\blacksquare$

$\underline{Y}_{PR}$ in Fig. 8.3 is triangle-looking.

**Theorem 12** *Generally, $\underline{Y}_{PR}$ is trapezoid-looking; however, $\underline{Y}_{PR}$ is triangle-looking if and only if all $\underline{G}_i$ are triangles with the same height.* ■

*Proof:* Because $a_{ir}(\alpha) \leq b_{il}(\alpha)$ [see Fig. 8.1(a)], it follows from (8.5) and (8.6) that, for $\forall\alpha \in [0, h_{\underline{Y}_{PR}}]$,

$$y_{Lr}(h_{\underline{Y}_{PR}}) = \frac{\sum_{i=1}^{n} a_{ir}(h_{\underline{Y}_{PR}}) f^i}{\sum_{i=1}^{n} f^i} \leq \frac{\sum_{i=1}^{n} b_{il}(h_{\underline{Y}_{PR}}) f^i}{\sum_{i=1}^{n} f^i} = y_{Rl}(h_{\underline{Y}_{PR}}) \tag{8.18}$$

i.e., $y_{Lr}(h_{\underline{Y}_{PR}}) \leq y_{Rl}(h_{\underline{Y}_{PR}})$. The equality holds if and only if $a_{ir}(h_{\underline{Y}_{PR}}) = b_{il}(h_{\underline{Y}_{PR}})$ for $\forall i = 1, \ldots, n$, i.e., when all $\underline{G}_i$ are triangles with the same height $h_{\underline{Y}_{PR}}$. In this case, according to Definition 37, $\underline{Y}_{PR}$ is triangle-looking. Otherwise, $y_{Lr}(h_{\underline{Y}_{PR}}) < y_{Rl}(h_{\underline{Y}_{PR}})$, and according to Definition 36, $\underline{Y}_{PR}$ is trapezoid-looking. ■

**Theorem 13** *Generally, $\overline{Y}_{PR}$ is trapezoid-looking; however, $\overline{Y}_{PR}$ is triangle-looking when all $\overline{G}^i$ are triangles.* ■

*Proof:* Because $a_{il}(\alpha) \leq b_{ir}(\alpha)$ [see Fig. 8.1(a)], it follows from (8.3) and (8.4) that, for $\forall\alpha \in [0, 1]$,

$$y_{Ll}(1) = \frac{\sum_{i=1}^{n} a_{il}(1) f^i}{\sum_{i=1}^{n} f^i} \leq \frac{\sum_{i=1}^{n} b_{ir}(1) f^i}{\sum_{i=1}^{n} f^i} = y_{Rr}(1) \tag{8.19}$$

i.e., $y_{Ll}(1) \leq y_{Rr}(1)$. The equality holds if and only if $a_{il}(1) = b_{ir}(1)$ for $\forall i = 1, \ldots, n$, i.e., when all $\overline{G}_i$ are triangles. In this case, $\overline{Y}_{PR}$ is triangle-looking according to Definition 37. Otherwise, $y_{Ll}(1) < y_{Rr}(1)$, and hence $\overline{Y}_{PR}$ is trapezoid-looking according to Definition 36. ■

### 8.3.2 The Geometry of $\tilde{Y}_{PR}$ FOUs

The following three definitions are about the geometry of $\tilde{Y}_{PR}$ FOUs:

**Definition 38** *An IT2 FS $\tilde{Y}_{PR}$ is a* left shoulder FOU *[see Fig. 8.2(b)] if and only if $h_{\underline{Y}_{PR}} = 1$, and $y_{Ll}(\alpha) = 0$ and $y_{Lr}(\alpha) = 0$ for $\forall \alpha \in [0, 1]$.* ■

**Definition 39** *An IT2 FS $\tilde{Y}_{PR}$ is a* right shoulder FOU *[see Fig. 8.2(c)] if and only if $h_{\underline{Y}_{PR}} = 1$, and $y_{Rl}(\alpha) = M$ and $y_{Rr}(\alpha) = M$ for $\forall \alpha \in [0, 1]$.* ■

**Definition 40** *An IT2 FS $\tilde{Y}_{PR}$ is an* interior FOU *[see Fig. 8.2(a)] if and only if it is neither a left shoulder FOU nor a right shoulder FOU.* ■

Three lemmas derived from the above three definitions are used in the proofs of Theorems 17-19 in Section 8.3.3:

**Lemma 14** *An IT2 FS $\tilde{Y}_{PR}$ is a* left shoulder FOU *if and only if $h_{\underline{Y}_{PR}} = 1$ and $y_{Lr}(1) = 0$.* ■

*Proof:* According to Definition 38, one only needs to show that "$y_{Lr}(1) = 0$" and "$y_{Ll}(\alpha) = 0$ and $y_{Lr}(\alpha) = 0$ for $\forall \alpha \in [0, 1]$" are equivalent. When $h_{\underline{Y}_{PR}} = 1$, $y_{Ll}(\alpha) \leq$

$y_{Lr}(\alpha)$ holds for $\forall\alpha \in [0,1]$ for an arbitrary FOU [e.g., see Fig. 8.4]; hence, one only needs to show that "$y_{Lr}(1) = 0$" and "$y_{Lr}(\alpha) = 0$ for $\forall\alpha \in [0,1]$" are equivalent. Because only convex IT2 FSs are used in PR, $y_{Lr}(\alpha) \leq y_{Lr}(1)$ for $\forall\alpha \in [0,1]$ [e.g., see again Fig. 8.4]; hence, $y_{Lr}(1) = 0$ is equivalent to $y_{Lr}(\alpha) = 0$ for $\forall\alpha \in [0,1]$. ■

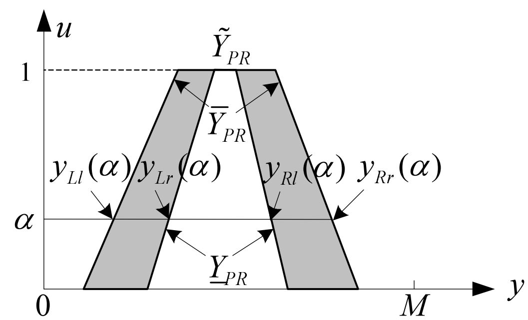



Fig. 8.4: An IT2 FS with $h_{\underline{Y}_{PR}} = 1$.

**Lemma 15** *An IT2 FS* $\tilde{Y}_{PR}$ *is a* right shoulder FOU *if and only if* $h_{\underline{Y}_{PR}} = 1$ *and* $y_{Rl}(1) = M$*.* ■

*Proof:* According to Definition 39, one only needs to show that "$y_{Rl}(1) = M$" and "$y_{Rl}(\alpha) = M$ and $y_{Rr}(\alpha) = M$ for $\forall\alpha \in [0,1]$" are equivalent. When $h_{\underline{Y}_{PR}} = 1$, $y_{Rr}(\alpha) \geq y_{Rl}(\alpha)$ holds for $\forall\alpha \in [0,1]$ [e.g., see Fig. 8.4]; hence, one only needs to show that "$y_{Rl}(1) = M$" and "$y_{Rl}(\alpha) = M$ for $\forall\alpha \in [0,1]$" are equivalent. Because only convex IT2 FSs are used in PR, $y_{Rl}(\alpha) \geq y_{Rl}(1)$ for $\forall\alpha \in [0,1]$ [e.g., see again Fig. 8.4]; hence, $y_{Rl}(1) = M$ is equivalent to $y_{Rl}(\alpha) = M$ for $\forall\alpha \in [0,1]$. ■

**Lemma 16** *An IT2 FS* $\tilde{Y}_{PR}$ *is an* interior FOU *if and only if:*

*(1)* $h_{\underline{Y}_{PR}} < 1$*; or*

*(2)* $h_{\underline{Y}_{PR}} = 1$, $y_{Lr}(1) > 0$ *and* $y_{Rl}(1) < M$. ■

*Proof:* (1) Because both left shoulder and right shoulder require $h_{\underline{Y}_{PR}} = 1$ (see Lemmas 14 and 15), $\tilde{Y}_{PR}$ must be an interior FOU when $h_{\underline{Y}_{PR}} < 1$.

(2) When $h_{\underline{Y}_{PR}} = 1$ and $y_{Lr}(1) > 0$, $\tilde{Y}_{PR}$ is not a left shoulder by Lemma 14. When $h_{\underline{Y}_{PR}} = 1$ and $y_{Rl}(1) < M$, $\tilde{Y}_{PR}$ is not a right shoulder by Lemma 15. Consequently, $\tilde{Y}_{PR}$ must be an interior FOU. ■

### 8.3.3 Properties of $\tilde{Y}_{PR}$ FOUs

In this subsection it is shown that $\tilde{Y}_{PR}$ computed from (8.1), that uses firing levels, resembles the three kinds of FOUs in a CWW codebook.

**Theorem 17** *Let $\tilde{Y}_{PR}$ be expressed as in (8.1). Then, $\tilde{Y}_{PR}$ is a* left shoulder FOU *if and only if all $\tilde{G}^i$ are left shoulder FOUs.* ■

*Proof:* From Lemma 14, $\tilde{Y}_{PR}$ is a left shoulder FOU if and only if $h_{\underline{Y}_{PR}} = 1$ and $y_{Lr}(1) = 0$, and similarly all $\tilde{G}^i$ are left shoulder FOUs if and only if $h_{\underline{G}^i} = 1$ and $a_{ir}(1) = 0$ for $\forall i$. To prove Theorem 17, one needs to show 1) "$h_{\underline{Y}_{PR}} = 1$" and "$h_{\underline{G}^i} = 1$ for $\forall i$" are equivalent; and 2) "$y_{Lr}(1) = 0$" and "$a_{ir}(1) = 0$ for $\forall i$" are equivalent.

The first requirement is obvious from (8.7). For the second requirement, it follows from (8.5) that

$$y_{Lr}(1) = \frac{\sum_{i=1}^{n} a_{ir}(1) f^i}{\sum_{i=1}^{n} f^i} \tag{8.20}$$

Because all $f^i > 0$, $y_{Lr}(1) = 0$ if and only if all $a_{ir}(1) = 0$. ■

**Theorem 18** *Let $\tilde{Y}_{PR}$ be expressed as in (8.1). Then, $\tilde{Y}_{PR}$ is a* right shoulder FOU *if and only if all $\tilde{G}^i$ are right shoulder FOUs.* ■

*Proof:* From Lemma 15, $\tilde{Y}_{PR}$ is a right shoulder if and only if $h_{\underline{Y}_{PR}} = 1$ and $y_{Rl}(1) = M$, and similarly all $\tilde{G}^i$ are right shoulders if and only if $h_{\underline{G}^i} = 1$ and $b_{il}(1) = M$ for $\forall i$. To prove Theorem 18, one only needs to show that 1) "$h_{\underline{Y}_{PR}} = 1$" and "$h_{\underline{G}^i} = 1$ for $\forall i$" are equivalent; and, 2) "$y_{Rl}(1) = M$" and "$b_{il}(1) = M$ for $\forall i$" are equivalent.

The first requirement is obvious from (8.7). For the second requirement, it follows from (8.6) that

$$y_{Rl}(1) = \frac{\sum_{i=1}^{n} b_{il}(1) f^i}{\sum_{i=1}^{n} f^i} \tag{8.21}$$

Because all $f^i > 0$, $y_{Rl}(1) = M$ if and only if all $b_{il}(1) = M$. ■

**Theorem 19** *Let $\tilde{Y}_{PR}$ be expressed as in (8.1). Then, $\tilde{Y}_{PR}$ is an* interior FOU *if and only if one of the following conditions is satisfied:*

1. *$\{\tilde{G}^i | i = 1, 2, \ldots, n\}$ is a mixture of both left and right shoulders.*

2. *At least one $\tilde{G}^i$ is an interior FOU.* ■

*Proof:* The sufficiency is proved first. Consider first Condition (1). Without loss of generality, assume $\{\tilde{G}^i | i = 1, \ldots, n_1\}$ are left shoulders and $\{\tilde{G}^i | i = n_1 + 1, \ldots, n\}$ are right shoulders, where $1 \le n_1 \le n - 1$. For each left shoulder $\tilde{G}^i$, it is true that $a_{ir}(1) = 0$

and[6] $b_{il}(1) < M$. For each right shoulder $\tilde{G}^i$, it is true that[7] $a_{ir}(1) > 0$ and $b_{il}(1) = M$. In summary,

$$a_{ir}(1) \begin{cases} = 0, & i = 1, \ldots, n_1 \\ > 0, & i = n_1 + 1, \ldots, n \end{cases} \tag{8.22}$$

$$b_{il}(1) \begin{cases} < M, & i = 1, \ldots, n_1 \\ = M, & i = n_1 + 1, \ldots, n \end{cases} \tag{8.23}$$

It follows that

$$y_{Lr}(1) = \frac{\sum_{i=1}^{n} a_{ir}(1) f^i}{\sum_{i=1}^{n} f^i} > \frac{\sum_{i=1}^{n_1} a_{ir}(1) f^i}{\sum_{i=1}^{n_1} f^i} = 0 \tag{8.24}$$

$$y_{Rl}(1) = \frac{\sum_{i=1}^{n} b_{il}(1) f^i}{\sum_{i=1}^{n} f^i} < \frac{\sum_{i=n_1+1}^{n} b_{il}(1) f^i}{\sum_{i=n_1+1}^{n} f^i} = M; \tag{8.25}$$

hence, $\tilde{Y}_{PR}$ is an interior FOU according to Part (2) of Lemma 16.

Next consider Condition (2). Without loss of generality, assume only $\tilde{G}^1$ is an interior FOU, $\{\tilde{G}^i | i = 2, \ldots, n_2\}$ are left shoulders, and $\{\tilde{G}^i | i = n_2 + 1, \ldots, n\}$ are right shoulders, where $2 \leq n_2 \leq n - 1$. Two sub-cases are considered:

i) When $h_{\underline{G}^1} < 1$, according to (8.7), $h_{\underline{Y}_{PR}} = h_{\underline{G}^1} < 1$, and hence $\tilde{Y}_{PR}$ is an interior FOU according to Part (1) of Lemma 16.

[6] $b_{il}(1)$ for a left shoulder cannot be $M$, because otherwise according to Lemma 15, $\tilde{G}^i$ would be a right shoulder.

[7] $a_{ir}(1)$ for a right shoulder cannot be 0, because otherwise according to Lemma 14, $\tilde{G}^i$ would be a left shoulder.

ii) When $h_{\underline{G}^1} = 1$, it follows from (8.7) that $h_{\underline{Y}_{PR}} = 1$, and from Lemma 16 applied to $\tilde{G}^1$ that $a_{1r}(1) > 0$ and $b_{1l}(1) < M$, i.e.,

$$a_{ir}(1) \begin{cases} = 0, & i = 2, \ldots, n_2 \\ > 0, & i = 1, n_2+1, \ldots, n \end{cases} \tag{8.26}$$

$$b_{il}(1) \begin{cases} < M, & i = 1, 2, \ldots, n_2 \\ = M, & i = n_2+1, \ldots, n \end{cases} \tag{8.27}$$

Consequently,

$$y_{Lr}(1) = \frac{\sum_{i=1}^{n} a_{ir}(1) f^i}{\sum_{i=1}^{n} f^i} > \frac{\sum_{i=2}^{n_2} a_{ir}(1) f^i}{\sum_{i=2}^{n_2} f^i} = 0 \tag{8.28}$$

$$y_{Rl}(1) = \frac{\sum_{i=1}^{n} b_{il}(1) f^i}{\sum_{i=1}^{n} f^i} < \frac{\sum_{i=n_2+1}^{n} b_{il}(1) f^i}{\sum_{i=n_2+1}^{n} f^i} = M \tag{8.29}$$

Again, $\tilde{Y}_{PR}$ is an interior FOU according to Part (2) of Lemma 16.

Next consider the necessity. $\{\tilde{G}^i | i = 1, 2, \ldots, n\}$ can only take the following four forms:

i) All $\tilde{G}^i$ are left shoulders.

ii) All $\tilde{G}^i$ are right shoulders.

iii) $\{\tilde{G}^i | i = 1, 2, \ldots, n\}$ is a mixture of both left and right shoulders.

iv) At least one $\tilde{G}^i$ is an interior FOU.

Assume $\tilde{Y}_{PR}$ is an interior FOU whereas $\{\tilde{G}^i | i = 1, 2, \ldots, n\}$ is not in Forms (iii) and (iv). Then, $\{\tilde{G}^i | i = 1, 2, \ldots, n\}$ must be in Form (i) or (ii). When $\{\tilde{G}^i | i = 1, 2, \ldots, n\}$

is in Form (i) (i.e., all $\tilde{G}^i$ are left shoulders), according to Theorem 17, $\tilde{Y}_{PR}$ must also be a left shoulder, which violates the assumption that $\tilde{Y}_{PR}$ is an interior FOU. Similarly, when $\{\tilde{G}^i | i = 1, 2, \ldots, n\}$ is in Form (ii) (i.e., all $\tilde{G}^i$ are right shoulders), according to Theorem 18, $\tilde{Y}_{PR}$ must be a right shoulder, which also violates the assumption. Hence, when $\tilde{Y}_{PR}$ is an interior FOU, $\{\tilde{G}^i | i = 1, 2, \ldots, n\}$ must be a mixture of both left and right shoulders, or at least one $\tilde{G}^i$ is an interior FOU. ■

Theorems 17-19 are important because they show that the output of PR is a normal IT2 FS and is similar to the word FOUs in a codebook[8] (see Fig. 3.17). So, the Jaccard similarity measure can be used to map $\tilde{Y}_{PR}$ to a word in the codebook. On the other hand, it is less intuitive to map a clipped FOU (see $\tilde{B}$ in Fig. 2.10), as obtained from a Mamdani inference mechanism, to a normal IT2 FS word FOU in the codebook.

## 8.4 Example 3 Completed

In Chapter 7 we have introduced how simplified rulebases can be generated from for SJAs. In this section, we explain how PR can be used in these SJAs.

### 8.4.1 Compute the Output of the SJA

First consider single-antecedent rules of the form

$$R^i: \text{ If } x \text{ is } \tilde{F}^i, \text{ Then } y \text{ is } \tilde{Y}^i \qquad i = 1, \ldots, N \tag{8.30}$$

[8]A small difference is that the LMFs of interior codebook word FOUs are always triangular, whereas the LMFs of interior $\tilde{Y}_{PR}$ are usually trapezoidal.

where $\tilde{Y}^i$ are computed by (7.6). In PR, the Jaccard similarity measure (3.4) is used to compute the firing levels of the rules, i.e., $f^i = s_J(\tilde{X}, \tilde{F}^i)$, $i = 1, \ldots, N$. Once $f^i$ are computed, the output FOU of the SJA is computed as [see (8.1)]

$$\tilde{Y}_C = \frac{\sum_{i=1}^{N} f^i \tilde{Y}^i}{\sum_{i=1}^{N} f^i} \tag{8.31}$$

The subscript $C$ in $\tilde{Y}_C$ stands for *consensus* because $\tilde{Y}_C$ is obtained by aggregating the survey results from a population of people, and the resulting SJA is called a *consensus SJA*. Because only the nine words in Fig. 7.1 are used in the SJAs, the similarities among them can be pre-computed, and $f^i$ in (8.31) can be retrieved from Table 8.1. Finally, $\tilde{Y}_C$ is mapped into a word in the Fig. 7.1 vocabulary also using the Jaccard similarity measure.

Table 8.1: Similarities among the nine words used in the SJAs.

| | NVL | AB | SS | S | MOA | GA | CA | LA | MAA |
|---|---|---|---|---|---|---|---|---|---|
| None to very little (NVL) | 1 | .11 | .08 | .05 | 0 | 0 | 0 | 0 | 0 |
| A bit (AB) | .11 | 1 | .40 | .21 | .02 | 0 | 0 | 0 | 0 |
| Somewhat small (SS) | .08 | .40 | 1 | .43 | .12 | .02 | 0 | 0 | 0 |
| Some (S) | .05 | .21 | .43 | 1 | .56 | .26 | .16 | .05 | 0 |
| Moderate amount (MOA) | 0 | .02 | .12 | .56 | 1 | .37 | .21 | .06 | 0 |
| Good amount (GA) | 0 | 0 | .02 | .26 | .37 | 1 | .63 | .32 | .03 |
| Considerable amount (CA) | 0 | 0 | 0 | .16 | .21 | .63 | 1 | .50 | .04 |
| Large amount (LA) | 0 | 0 | 0 | .05 | .06 | .32 | .50 | 1 | .05 |
| Maximum amount (MAA) | 0 | 0 | 0 | 0 | 0 | .03 | .04 | .05 | 1 |

Next consider two-antecedent rules of the form

$$R^i: \text{ If } x_1 \text{ is } \tilde{F}_1^i \text{ and } x_2 \text{ is } \tilde{F}_2^i, \text{ Then } y \text{ is } \tilde{Y}^i \qquad i = 1, \ldots, N \tag{8.32}$$

The firing levels are computed as

$$f^i = s_J(\tilde{X}_1, \tilde{F}_1^i) \star s_J(\tilde{X}_2, \tilde{F}_2^i) \qquad i = 1, \ldots, N \tag{8.33}$$

where in this paper $\star$ is the minimum $t$-norm. $s_J(\tilde{X}_1, \tilde{F}_1^i)$ and $s_J(\tilde{X}_2, \tilde{F}_2^i)$ can be obtained from the pre-computed similarities in Table 8.1. When all $f^i$ are obtained, the output FOU is computed again using (8.31) and then $\tilde{Y}_C$ is mapped back into a word in the Fig. 7.1 vocabulary using the Jaccard similarity measure.

### 8.4.2 Use SJA

As mentioned below (8.31), each SJA that is designed from survey is referred to as a *consensus SJA*, because it is obtained by using survey results from a group of people. There are at least two ways to make use of the consensus SJA:

1. Use it to infer outputs for new scenarios that are not considered in survey.

2. Use it to advise (counsel) an individual about a social judgment, as shown in Fig. 8.5. An individual is given a questionnaire similar to the one used in Step 6 of the knowledge mining process, and his/her responses are obtained for all the words in the vocabulary. These responses can then be compared with the outputs of the consensus SJA. If some or all of the individual's responses are "far" from those of the consensus SJA, then some action could be taken to sensitize the individual about these differences.

More details about both approaches are give in this section.

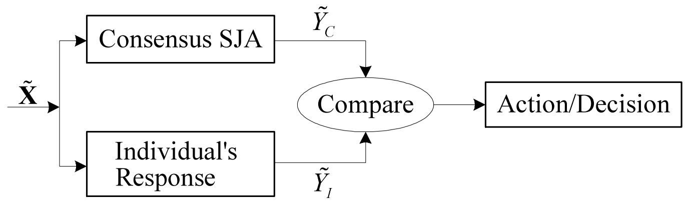



Fig. 8.5: One way to use the SJA for a social judgment.

### 8.4.3 Single-Antecedent Rules: Touching and Flirtation

This subsection shows how the consensus $\text{SJA}_1$ developed in Section 7.4 can be used.

For an input touching level, the output of $\text{SJA}_1$ can easily be computed by PR, as illustrated by the following:

**Example 27** *Let observed touching be somewhat small (SS). From the third row of Table 8.1 the following firing levels of the five rules are obtained:*

$$f^1 = s_J(SS, NVL) = 0.08$$

$$f^2 = s_J(SS, S) = 0.43$$

$$f^3 = s_J(SS, MOA) = 0.12$$

$$f^4 = s_J(SS, LA) = 0$$

$$f^5 = s_J(SS, MAA) = 0$$

*The resulting $\tilde{Y}_C$ computed from (8.31) is depicted in Fig. 8.6 as the dashed curve. The similarities between $\tilde{Y}_C$ and the nine words in the Fig. 7.1 vocabulary are computed to be:*

$$s_J(\tilde{Y}_C, NVL) = 0.17 \qquad s_J(\tilde{Y}_C, AB) = \boldsymbol{0.67} \qquad s_J(\tilde{Y}_C, SS) = 0.43$$

$$s_J(\tilde{Y}_C, S) = 0.24 \qquad s_J(\tilde{Y}_C, MOA) = 0.04 \qquad s_J(\tilde{Y}_C, GA) = 0$$

$$s_J(\tilde{Y}_C, CA) = 0 \qquad s_J(\tilde{Y}_C, LA) = 0 \qquad s_J(\tilde{Y}_C, MAA) = 0$$

*Because $\tilde{Y}_C$ and AB have the largest similarity, $\tilde{Y}_C$ is mapped into the word AB.* ■

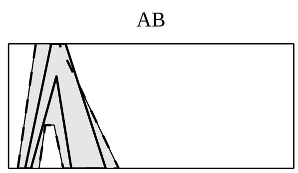



Fig. 8.6: $\tilde{Y}_C$ (dashed curve) and the mapped word (AB, solid curve) when touching is *somewhat small*.

When PR is used to combine the rules and any of the nine words in Fig. 7.1 are used as inputs, the outputs of the consensus $SJA_1$ are mapped to words shown in the second column of Table 8.2. Each of these words was determined by using the same kind of calculations that were just described in Example 27. Observe that generally the flirtation level increases as touching increases, as one would expect.

Next, assume for the nine codebook words, an individual gives the responses[9] shown in the third column of Table 8.2. Observe that this individual's responses are generally the same as or lower than $\tilde{Y}_C$. This means that this individual may under-react to touching.

[9]The individual is asked the following question for each of the nine codebook words: *"If there is* (one of the nine codebook words) *touching, then what is the level of flirtation?"* and the answer must also be a word from the nine-word codebook.

The similarities between the consensus outputs $\tilde{Y}_C$ and the individual's responses $\tilde{Y}_I$, computed by using (3.4), are shown in the fourth column of Table 8.2. $\tilde{Y}_I$ and $\tilde{Y}_C$ are said to be "significantly different" if $s_J(\tilde{Y}_C, \tilde{Y}_I)$ is smaller than a threshold $\theta$. Let $\theta = 0.6$. Then, for the last four inputs, $\tilde{Y}_I$ and $\tilde{Y}_C$ are significantly different. Some action could be taken to sensitize the individual about these differences.

Table 8.2: A comparison between the consensus $SJA_1$ outputs and an individual's responses.

| Touching | Flirtation level | | Similarity $s_J(\tilde{Y}_C, \tilde{Y}_I)$ |
|---|---|---|---|
| | Consensus ($\tilde{Y}_C$) | Individual ($\tilde{Y}_I$) | |
| None to very little (NVL) | NVL | NVL | 1 |
| A bit (AB) | AB | AB | 1 |
| Somewhat small (SS) | AB | AB | 1 |
| Some (S) | SS | SS | 1 |
| Moderate amount (MOA) | SS | SS | 1 |
| Good amount (GA) | S | SS | 0.12 |
| Considerable amount (CA) | MOA | SS | 0.56 |
| Large amount (LA) | GA | SS | 0.26 |
| Maximum amount (MAA) | CA | MOA | 0.21 |

### 8.4.4 Single-Antecedent Rules: Eye Contact and Flirtation

This subsection shows how the consensus $SJA_2$ developed in Section 7.5 can be used.

When PR is used to combine the rules and any of the nine words in Fig. 7.1 are used as inputs, the outputs of the consensus $SJA_2$ are mapped to words shown in the second column of Table 8.3. Observe that generally the flirtation level increases as eye contact increases, as one would expect.

Assume for the nine codebook words, an individual gives the responses shown in the third column of Table 8.3. Observe that this individual's responses are generally the same as or higher than those from the consensus $SJA_2$. This means that this individual may over-react to eye contact.

The similarities between the consensus outputs $\tilde{Y}_C$ and the individual's responses $\tilde{Y}_I$ are shown in the fourth column of Table 8.3. Again, let the threshold be $\theta = 0.6$. Then, for the last six inputs, $\tilde{Y}_I$ and $\tilde{Y}_C$ are significantly different. Some action could be taken to sensitize the individual about these differences.

Table 8.3: A comparison between the consensus $SJA_2$ outputs and an individual's responses.

| Eye contact | Flirtation level | | Similarity $s_J(\tilde{Y}_I, \tilde{Y}_C)$ |
|---|---|---|---|
| | Consensus ($\tilde{Y}_C$) | Individual ($\tilde{Y}_I$) | |
| None to very little (NVL) | NVL | NVL | 1 |
| A bit (AB) | AB | AB | 1 |
| Somewhat small (SS) | SS | SS | 1 |
| Some (S) | SS | S | 0.43 |
| Moderate amount (MOA) | S | MOA | 0.56 |
| Good amount (GA) | MOA | CA | 0.21 |
| Considerable amount (CA) | GA | LA | 0.32 |
| Large amount (LA) | CA | LA | 0.50 |
| Maximum amount (MAA) | LA | MAA | 0.05 |

### 8.4.5 Two-Antecedent Rules: Touching/Eye Contact and Flirtation

This subsection shows how the consensus $SJA_3$ developed in Section 7.6 can be used.

For input touching and eye contact levels, the output of $SJA_3$ can easily be computed by PR, as illustrated by the following:

**Example 28** *Let observed touching be a bit (AB) and observed eye contact be considerable amount (CA). Only 12 of the possible 25 firing levels are non-zero, and they are obtained from the second and the seventh rows of Table 8.1, as:*

$$f^{1,2} = \min\{s_J(AB, NVL), s_J(CA, S)\} = \min(0.11, 0.16) = 0.11$$

$$f^{1,3} = \min\{s_J(AB, NVL), s_J(CA, MOA)\} = \min(0.11, 0.21) = 0.11$$

$$f^{1,4} = \min\{s_J(AB, NVL), s_J(CA, LA)\} = \min(0.11, 0.50) = 0.11$$

$$f^{1,5} = \min\{s_J(AB, NVL), s_J(CA, MAA)\} = \min(0.11, 0.04) = 0.04$$

$$f^{2,2} = \min\{s_J(AB, S), s_J(CA, S)\} = \min(0.21, 0.16) = 0.16$$

$$f^{2,3} = \min\{s_J(AB, S), s_J(CA, MOA)\} = \min(0.21, 0.21) = 0.21$$

$$f^{2,4} = \min\{s_J(AB, S), s_J(CA, LA)\} = \min(0.21, 0.50) = 0.21$$

$$f^{2,5} = \min\{s_J(AB, S), s_J(CA, MAA)\} = \min(0.21, 0.04) = 0.04$$

$$f^{3,2} = \min\{s_J(AB, MOA), s_J(CA, S)\} = \min(0.02, 0.16) = 0.02$$

$$f^{3,3} = \min\{s_J(AB, MOA), s_J(CA, MOA)\} = \min(0.02, 0.21) = 0.02$$

$$f^{3,4} = \min\{s_J(AB, MOA), s_J(CA, LA)\} = \min(0.02, 0.50) = 0.02$$

$$f^{3,5} = \min\{s_J(AB, MOA), s_J(CA, MAA)\} = \min(0.02, 0.04) = 0.02$$

*The resulting* $\tilde{Y}_C$ *computed from (8.31) is depicted in Fig. 8.7 as the dashed curve. The similarities between* $\tilde{Y}_C$ *and the nine words in the Fig. 7.1 vocabulary are computed to be:*

| | | |
|---|---|---|
| $s_J(\tilde{Y}_C, NVL) = 0$ | $s_J(\tilde{Y}_C, AB) = 0.06$ | $s_J(\tilde{Y}_C, SS) = 0.21$ |
| $s_J(\tilde{Y}_C, S) = 0.64$ | $s_J(\tilde{Y}_C, MOA) = \boldsymbol{0.71}$ | $s_J(\tilde{Y}_C, GA) = 0.28$ |
| $s_J(\tilde{Y}_C, CA) = 0.15$ | $s_J(\tilde{Y}_C, LA) = 0.03$ | $s_J(\tilde{Y}_C, MAA) = 0$ |

*Because* $\tilde{Y}_C$ *and* MOA *have the largest similarity,* $\tilde{Y}_C$ *is mapped into the word* MOA. ■

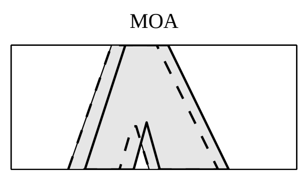



Fig. 8.7: $\tilde{Y}_C$ (dashed curve) and the mapped word (MOA, solid curve) when touching is $AB$ and eye contact is $CA$.

When PR is used to combine the rules, and any pair of the nine words in Fig. 7.1 are used as observed inputs for touching and eye contact, there are a total of 81 combinations of these two inputs. The 81 SJA outputs and the words that are most similar to them are shown in Fig. 8.8. Scan this figure horizontally from left-to-right to see the effect of varying touching on flirtation. Scan it vertically from top-to-bottom to see the effect of varying eye contact on flirtation. Scan it diagonally from top-left to bottom-right to see the simultaneous effects of varying touching and eye contact on flirtation. Observe that generally the flirtation level increases as either one or both inputs increase, as one would expect.

Once the consensus $SJA_3$ is constructed, one can again check an individual's responses against it, as he or she did for $SJA_1$ and $SJA_2$. The procedures are quite similar, so they are not repeated here.

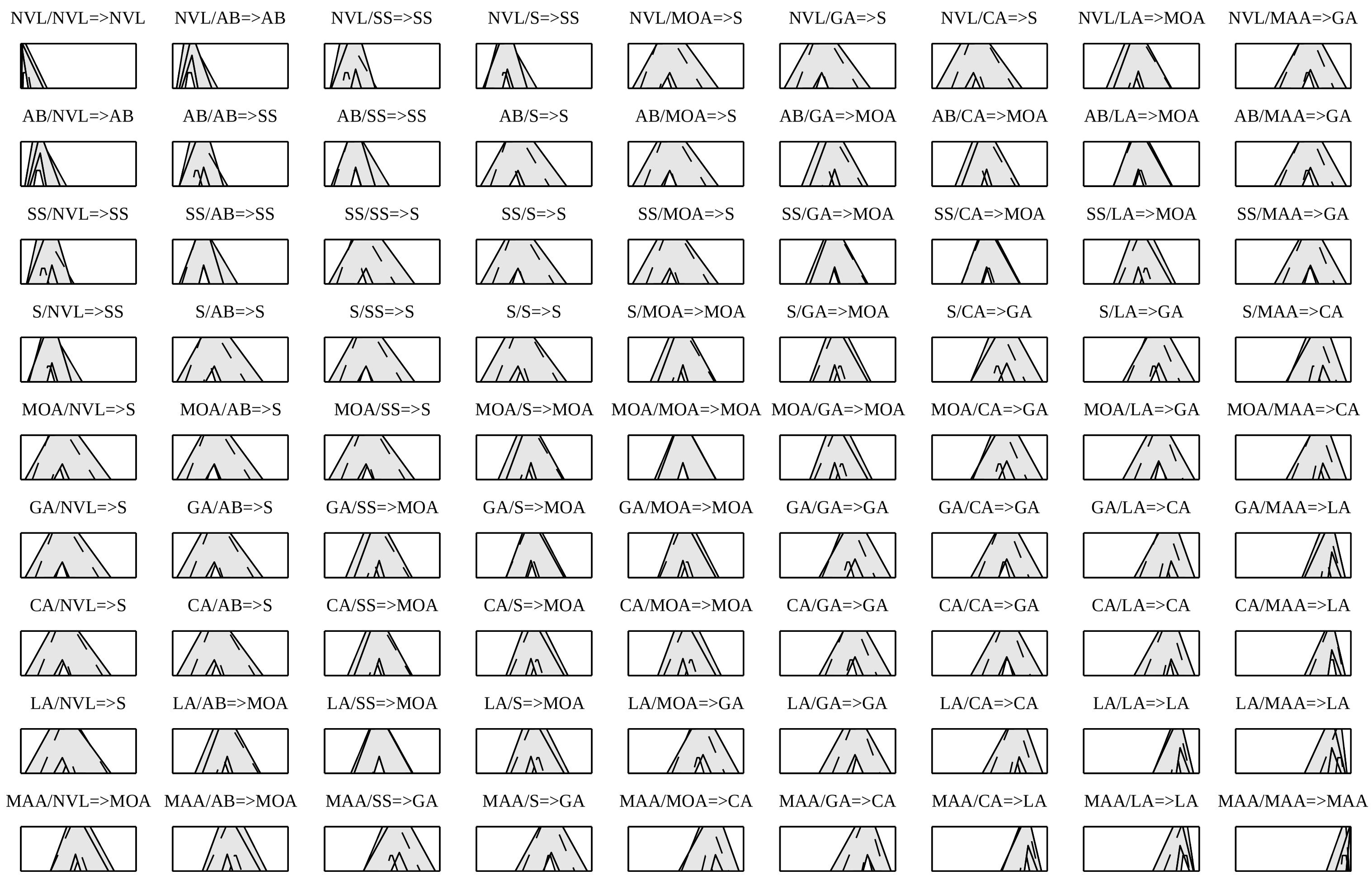



Fig. 8.8: $\tilde{Y}_C$ (dashed curve) and the mapped word (solid curve) for different combinations of touching/eye contact. The title of each sub-figure, $X_1/X_2 \Rightarrow Y$, means that "*when touching is $X_1$ and eye contact is $X_2$, the flirtation level is $Y$.*"

### 8.4.6 On Multiple Indicators

As has been mentioned in Example 3, people have difficulties in answering questions with more than two antecedents. So, in the survey each rule consists of only one or two antecedents; however, in practice an individual may observe one indicator or more than one indicators. An interesting problem is how to deduce the output for multiple antecedents using rulebases consisting of only one or two antecedents.

For the sake of this discussion, assume there are four indicators of flirtation, *touching, eye contact, acting witty* and *primping*, and that the following ten SJAs have been created:

$SJA_1$: IF *touching* is _____, THEN flirtation is _____.

$SJA_2$: IF *eye contact* is _____, THEN flirtation is _____.

$SJA_3$: IF *acting witty* is _____, THEN flirtation is _____.

$SJA_4$: IF *primping* is _____, THEN flirtation is _____.

$SJA_5$: IF *touching* is _____ and *eye contact* is _____, THEN flirtation is _____.

$SJA_6$: IF *touching* is _____ and *acting witty* is _____, THEN flirtation is _____.

$SJA_7$: IF *touching* is _____ and *primping* is _____, THEN flirtation is _____.

$SJA_8$: IF *eye contact* is _____ and *acting witty* is _____, THEN flirtation is _____.

$SJA_9$: IF *eye contact* is _____ and *primping* is _____, THEN flirtation is _____.

$SJA_{10}$: IF *acting witty* is _____ and *primping* is _____, THEN flirtation is _____.

These ten SJAs can be used as follows:

1. When only one indicator is observed, only one single-antecedent SJA from $SJA_1$–$SJA_4$ is activated.

2. When only two indicators are observed, only one two-antecedent SJA from $SJA_5$–$SJA_{10}$ is activated.

3. When more than two indicators are observed, the output is computed by aggregating the outputs of the activated two-antecedent SJAs[10]. For example, when the observed indicators are *touching*, *eye contact* and *primping*, three two-antecedent SJAs — $SJA_5$, $SJA_7$ and $SJA_9$ — are activated, and each one gives a flirtation level. The final output is some kind of aggregation of the results from these three SJAs. There are different aggregation operators, e.g., mean, linguistic weighted average, maximum, etc. An intuitive approach is to survey the subjects about the relative importance of the four indicators and hence to determine the linguistic relative importance of $SJA_5$–$SJA_{10}$. These relative importance words can then be used as the weights for $SJA_5$–$SJA_{10}$, and the final flirtation level can then be computed by a linguistic weighted average.

A diagram of the proposed SJA architecture for different numbers of indicators is shown in Fig. 8.9.

[10]Some of the four single-antecedent SJAs, $SJA_1$–$SJA_4$, are also fired; however, they are not used because they do not fit the inputs as well as two-antecedent SJAs, since the latter account for the correlation between two antecedents, whereas the former do not.

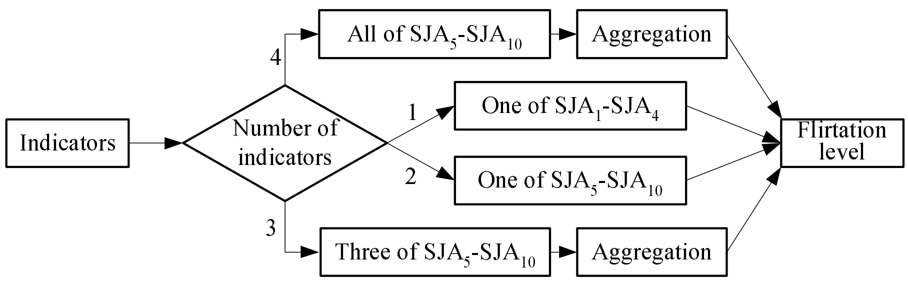



Fig. 8.9: An SJA architecture for one-to-four indicators.

# Chapter 9

# Conclusions and Future Works

## 9.1 Conclusions

In this dissertation, we have introduced the Per-C, a CWW architecture for MCDM. It consists of three components: encoder, which transforms inputs words into IT2 FS models; CWW engine, which performs operations on the IT2 FS word models; and decoder, which maps the output of the CWW engine into a recommendation (word, rank or class). The CWW engine and the decoder are the main focus of our work. Two CWW engines have been proposed: 1) novel weighted averages for MADM, which for the first time enable us to aggregate mixed signals consisting of numbers, intervals, T1 FSs and/or words modeled by IT2 FSs; and, 2) perceptual reasoning for MODM, which is an approximate reasoning method to infer an output for an input from rules. Two methods for rulebase construction — linguistic summarization to extract rules from data and knowledge mining to construct rules through survey — have also been introduced. Particularly, linguistic summarization can be used alone as a data mining approach for database understanding.

## 9.2 Future Works

Some future research works are proposed in this section.

### 9.2.1 Incorporate Uncertainties in the Analytic Hierarchy Process (AHP)

The analytic hierarchy process (AHP) is an MADM approach that uses multiple pairwise comparisons to rank order alternatives. It was first developed by Prof. Thomas L. Saaty [111], has been extensively studied and refined since then [89, 112, 113, 120], and has been used worldwide in a variety of decision-making situations in fields [113–116, 118] such as economics, finance, politics, social sciences, games, sports, etc. Its details are given in Appendix C.

There are different types of uncertainties in the AHP, e.g., the inconsistency[1] in the pairwise comparison matrices (PCMs), the uncertainties in expressing the preferences using crisp numbers, the change of judgments over time or in different scenarios, etc. Some approaches to incorporate uncertainties in the AHP are introduced next.

Poyhonen et al. [105] expressed concern about the numerical interpretation of the phrases that are used in the AHP, and chose to analyze the relationship between words and numbers. They created some experiments to find "*representative numerical counterparts for the verbal expressions used in the AHP*," to see if the results from the AHP were

[1]A positive $m \times m$ matrix $W = [c_{ij}]$ is *consistent* if $c_{ij}c_{jk} = c_{ik}$, $\forall i, j, k = 1, \ldots, m$.

sensitive to the numerical scale and "*to study whether the numerical counterparts of the verbal expressions vary from one decision problem to another.*" Their experiments: 1) "*Do not support the 1-9 scale as the default to represent numerical counterparts for the verbal expressions in the AHP*"; 2) "*Do not suggest any fixed numerical scale as a standard tool for the AHP, because the interpretation of verbal expressions varies from one person to another*"; and, 3) demonstrate that "*numerical counterparts of the verbal expressions vary according to the set of elements involved in the comparisons,*" i.e., they are application dependent.

Regarding these three conclusions and the material in this dissertation: 1) Although we began with the 0-10 scale, we did not pre-assign numbers to words; instead, our collected word-interval data and the Interval Approach, that mapped it into an FOU, located the word FOU on the 0-10 scale; 2) Variability from one person to another was not ignored by us, but instead was directly mapped into a word FOU by the Interval Approach; and, 3) From the very beginning we have advocated that the relationships between words and numbers (FOUs) are application dependent.

Beyth-Marom [5], Hamm [40] and Timmermanns [130] have observed that verbal expressions seem to be best modeled by ranges of values rather than by point estimates. Poyhonen et al. [105] state: "*... provided these results can be generalized to ratio comparisons of relative importance, it is possible that the exact numbers in the AHP should be replaced by intervals of numbers.*"

The data that we collect about a word are indeed ranges of values, one per subject; however, the Interval Approach provides us with an FOU for the word and not just an interval. If one wants to just use an interval of numbers for the AHP verbal expressions, then one way to obtain such an interval is: 1) Choose an appropriate scale for the application; 2) Collect interval end-point data for the words of that application as we have explained; 3) Map the collection of subject intervals into an FOU using the Interval Approach (modified to the scale if it is not the 0-10 scale); and 4) Compute the centroids of the word FOUs. The centroid is an interval of numbers on the given scale and provides a measure of the uncertainty of the entire FOU.

Paulson and Zahir [103] considered the uncertainty in alternative rankings [114] and the probability of rank reversals[2] as functions of the number of alternatives and of the layer of the hierarchy, and found that ranking uncertainty decreases as the number of alternatives or the layer of the hierarchy increases. The sole source of uncertainty was assumed to be the entries of the judgment matrices. Zahir [187] later showed how to compute uncertainties in the relative priorities of a decision.

Reuven and Wan [109] considered two types of uncertainties: 1) the future characteristics of the decision-making environment described by a set of scenarios, and 2) the decision-making judgments regarding each pairwise comparison. They proposed a simulation approach for handling both types of uncertainties in the AHP.

[2]Rank reversal denotes the phenomena that the rank of the alternatives may be changed by adding or deleting (irrelevant) alternatives.

Sugihara and Tanaka [127] proposed a linear programming approach to obtain interval weight vector and priority vectors from the crisp judgment matrices, i.e., the crisp PCMs are still used, but interval weight vector and priority vectors are computed from them by making use of the inconsistencies of the PCMs.

Beynon [4] proposed a DS/AHP method which combines the Dempster-Shafer (DS) theory of evidence with the AHP. This method allows judgments on groups of alternatives to be made, instead of pairwise comparisons used in the original AHP. It also provides a measure of uncertainty in the final results by evaluating the range of uncertainty expressed by the decision maker, and hence allows an understanding of the appropriateness of the rating scale values.

Fuzzy AHP [10, 11, 21, 121, 175] is a a popular way to incorporate uncertainties into the AHP. In this approach, T1 FSs instead of crisp numbers are used in the judgment matrices. Usually $\alpha$-cuts are used to decompose these T1 FSs into intervals, and for each $\alpha$, an eigenvalue interval can be computed. The problem is then to find the "best" (maximum) eigenvalue with small consistency ratio[3] (e.g., $<0.1$) and high $\alpha$. There are several different ways to set up such kinds of multi-objective optimization problems, e.g. [121],

1. Minimize the maximum eigenvalue (i.e., minimize the consistency ratio) while setting constraints on $\alpha$ (e.g., $\alpha > 0.5$).

[3]Consistency ratio [122] is a measure of inconsistency, and a smaller consistency ratio means better consistency.

2. Maximize $\alpha$ while setting a constraint on the consistency ratio (e.g., consistency ratio $< 0.1$).

3. Run simulations to determine the eigenvalue intervals for all $\alpha$ and then ask the user to make a decision.

Saaty and Tran [121] oppose the fuzzy AHP approach for the following reasons:

1. Improving consistency in the PCMs does not necessarily improve the validity of the output, whereas in many fuzzy AHP approaches people try to improve the consistency regardless of the consequences.

2. In many fuzzy AHP approaches people obtain certain and crisp judgments first and then fuzzify them to be fuzzy judgments, whereas it is more reasonable to obtain these uncertain judgments directly from the decision-maker.

We are in general agreement with Saaty and Tran about these objections; however, we think the Interval Approach could be used to extend the AHP to a *linguistic AHP* in which all pair-wise comparisons are expressed as linguistic terms that are modeled by FOUs. Although the linguistic AHP would use FSs, it would not use T1 FSs, and it does not simply fuzzify crisp numbers. The details for how to do this remain to be worked out.

### 9.2.2 Efficient Algorithm for Linguistic Summarization

Currently an exhaustive search method is used in linguistic summarization, i.e., to find the top $N$ rules with the maximum usefulness from a database, we need to compute the usefulness for all possible combinations of rules and then rank them. This is very time-consuming when the database is large, and/or each rule has multiple antecedents/consequents, and/or each antecedent/consequent has many MFs. An efficient algorithm that can eliminate non-interesting rules from the beginning and hence speed up the search is highly desirable.

### 9.2.3 Make Use of the Rule Quality Measures in Perceptual Reasoning

In linguistic summarization, we generate not only rules, but also quality measures for them, e.g., truth level, degree of sufficient coverage, degree of usefulness, and degree of outlier. Unfortunately, presently we do not know how to make use of them in perceptual reasoning. It is counter-intuitive to simply ignore them, because they indicate different degrees of importance for different rules.

# Bibliography


[1] "Analytic hierarchy process." [Online]. Available: http://en.wikipedia.org/wiki/Analytic_hierarchy_process.

[2] E. Avineri, J. Prashker, and A. Ceder, "Transportation projects selection process using fuzzy sets theory," *Fuzzy Sets and Systems*, vol. 116, pp. 35–47, 2000.

[3] J. F. Baldwin, "Knowledge from data using fuzzy methods," *Pattern Recongnition Lett.*, vol. 17, pp. 593–600, 1996.

[4] M. Beynon, "DS/AHP method: A mathematical analysis, including an understanding of uncertainty," *European Journal of Operational Research*, vol. 104, no. 1, pp. 148–164, 2002.

[5] R. Beyth-Marom, "How probable is probable? a numerical translation of verbal probability expressions," *J. Forecasting*, vol. 1, pp. 257–269, 1982.

[6] J. Bezdek, "Fuzzy models–what are they, and why?" *IEEE Trans. on Fuzzy Systems*, vol. 1, no. 1, pp. 1–5, 1993.

[7] N. Blanchard, "Cardinal and ordinal theories about fuzzy sets," in *Fuzzy Information and Decision Processes*, M. M. Gupta and E. Sanchez, Eds. Amsterdam: North-Holland, 1982, pp. 149–157.

[8] A. Blumenthal, *The Process of Cognition*. Englewood Cliffs, NJ: Prentice Hall, 1977.

[9] J. J. Buckley and H. Ying, "Expert fuzzy controller," *Fuzzy Sets and Systems*, vol. 43, pp. 127–137, 1991.

[10] J. J. Buckley, T. Feuring, and Y. Hayashi, "Fuzzy hierarchical analysis revisited," *European Journal of Operational Research*, vol. 129, no. 1, pp. 48–64, 2001.

[11] J. Buckley, "Fuzzy hierarchical analysis," *Fuzzy sets and systems*, vol. 17, no. 3, pp. 233–247, 1985.

[12] H. Bustince, "Indicator of inclusion grade for interval-valued fuzzy sets. Application to approximate reasoning based on interval-valued fuzzy sets," *Int'l. Journal of Approximate Reasoning*, vol. 23, no. 3, pp. 137–209, March 2000.

[13] H. Bustince, M. Pagola, and E. Barrenechea, "Construction of fuzzy indices from fuzzy DI-subsethood measures: Application to the global comparison of images," *Information Sciences*, vol. 177, pp. 906–929, 2007.

[14] S.-M. Chen, "Evaluating weapon systems using fuzzy arithmetic operations," *Fuzzy Sets and Systems*, vol. 77, pp. 265–276, 1996.

[15] ——, "A new method for evaluating weapon systems using fuzzy set theory," *IEEE Trans. on Systems, Man, and Cybernetics-A*, vol. 26, pp. 493–497, 1996.

[16] C.-H. Cheng, "Evaluating weapon systems using ranking fuzzy numbers," *Fuzzy Sets and Systems*, vol. 107, pp. 25–35, 1999.

[17] D. A. Chiang, L. R. Chow, and Y. F. Wang, "Mining time series data by a fuzzy linguistic summary system," *Fuzzy Sets and Systems*, vol. 112, pp. 419–432, 2000.

[18] M. D. Cock and E. Kerre, "On (un)suitable fuzzy relations to model approximate equality," *Fuzzy Sets and Systems*, vol. 133, no. 2, pp. 137–153, 2003.

[19] G. B. Crawford, "The geometric mean procedure for estimating the scale of a judgment matrix," *Mathematical modeling*, vol. 9, pp. 3–5, 1987.

[20] V. V. Cross and T. A. Sudkamp, *Similarity and Compatibility in Fuzzy Set Theory: Assessment and Applications*. Heidelberg, NY: Physica-Verlag, 2002.

[21] R. Csutora and J. J. Buckley, "Fuzzy hierarchical analysis: the lambda-max method," *Fuzzy Sets and Systems*, vol. 120, no. 2, pp. 181–195, 2001.

[22] A. De Luca and S. Termini, "A definition of nonprobabilistic entropy in the setting of fuzzy sets theory," *Information and Computation*, vol. 20, pp. 301–312, 1972.

[23] *Decision Making*, ser. Harvard Business Essentials. Boston, MA: Harvard Business School Press, 2005.

[24] F. F. Dengen Zhou, Jairam Kamath and M. Morea, "Identifying key recovery mechanisms in a diatomite waterflood," in *SPE/DOE 13th Symposium on Improved Oil Recovery*, Tulsa, OK, April 2002.

[25] D. Dubois and H. Prade, "Gradual rules in approximate reasoning," *Information Sciences*, vol. 61, pp. 103–122, 1992.

[26] ——, "Fuzzy cardinality and the modeling of imprecise quantification," *Fuzzy Sets and Systems*, vol. 16, pp. 199–230, 1985.

[27] W. Duch, R. Setiono, and J. Zurada, "Computational intelligence methods for rule-based data understanding," *Proc. IEEE*, vol. 92, no. 5, pp. 771–805, 2004.

[28] K. Duran, H. Bernal, and M. Melgarejo, "Improved iterative algorithm for computing the generalized centroid of an interval type-2 fuzzy set," in *Proc. Annual Meeting of the North American Fuzzy Information Processing Society*, New York, May 2008, pp. 1–5.

[29] "Fact sheet – Valhall life of field seismic (LOFS)." [Online]. Available: http://www.oyogeospace.com/pdfs/engineering_lofs_factsheet.pdf.

[30] J. Fan and W. Xie, "Some notes on similarity measure and proximity measure," *Fuzzy Sets and Systems*, vol. 101, pp. 403–412, 1999.

[31] U. M. Fayyad, "SKICAT: Sky image cataloging and analysis tool," in *Proc. Int'l Joint Conf. on Artificial Intelligence*, vol. 2, Montreal, Quebec, Canada, August 1995, pp. 2067–2068.

[32] D. Filev and R. Yager, "On the issue of obtaining OWA operator weights," *Fuzzy Sets and Systems*, vol. 94, pp. 157–169, 1998.

[33] J. L. García-Lapresta and L. C. Meneses, "An empirical analysis of transitivity with four scaled preferential judgment modalities," *Review of Economic Design*, vol. 8, pp. 335–346, 2003.

[34] R. George and R. Srikanth, "Data summarization using genetic algorithms and fuzzy logic," in *Genetic Algorithms Soft Comput.*, F. Herrera and J. Verdegay, Eds. Heidelberg, Germany: Springer-Verlag, 1996, pp. 599–611.

[35] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, 1989.

[36] R. L. Gorsuch, *Factor Analysis.* Hillsdale, NJ: Lawrence Erlbaum, 1983.

[37] M. B. Gorzalczany, "A method of inference in approximate reasoning based on interval-valued fuzzy sets," *Fuzzy Sets and Systems*, vol. 21, pp. 1–17, 1987.

[38] S. Gottwald, "A note on fuzzy cardinals," *Kybernetika*, vol. 16, pp. 156–158, 1980.

[39] H. A. Hagras, "A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots," *IEEE Trans. on Fuzzy Systems*, vol. 12, no. 4, pp. 524–539, Aug. 2004.

[40] R. M. Hamm, "Selection of verbal probabilities: a solution for some problems of verbal probability expressions," *Organization, Behavior, Human Decision Process*, vol. 48, pp. 193–223, 1991.

[41] D. J. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining.* Boston, MA: MIT Press, 2001.

[42] K. Hejl, A. Madding, M. Morea, C. Glatz, J. Luna, W. Minner, T. Singh, and G. Stanley, "Extreme multistage fracturing improves vertical coverage and well performance in the lost hills field," in *SPE Annual Technical Conference and Exhibition*, San Antonio, TX, September 2006.

[43] F. Herrera, "A sequential selection process in group decision making with linguistic assessment," *Information Sciences*, vol. 85, p. 223239, 1995.

[44] K. Hirota and W. Pedrycz, "Fuzzy computing for data mining," *Proc. IEEE*, vol. 87, no. 9, pp. 1575–1600, 1999.

[45] S. Horikawa, T. Furahashi, and Y. Uchikawa, "On fuzzy modeling using fuzzy neural networks with back-propagation algorithm," *IEEE Trans. on Neural Network*, vol. 3, pp. 801–806, 1992.

[46] C. Hwang and A. Masud, *Multiple Attribute decision making: methods and applications – A state-of-the-art survey*. Berlin: Springer-Verlag, 1981.

[47] "Introduction to Oracle Data Mining." [Online]. Available: http://download.oracle.com/docs/cd/B12037_01/datamine.101/b10698/1intro.htm

[48] P. Jaccard, "Nouvelles recherches sur la distribution florale," *Bulletin de la Societe de Vaud des Sciences Naturelles*, vol. 44, p. 223, 1908.

[49] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft-Computing*. Upper Saddle River, NJ: Prentice-Hall, 1997.

[50] J. R. Jang, "Self-learning fuzzy controllers based on temporal back-propagation," *IEEE Trans. on Neural Networks*, vol. 3, pp. 714–723, 1992.

[51] T. A. Johansen, "Fuzzy model based control: Stability, robustness, and performance issues," *IEEE Trans. on Fuzzy Systems*, vol. 2, no. 3, pp. 221–234, 1994.

[52] J. Joo, D. Wu, J. M. Mendel, and A. Bugacov, "Forecasting the post fracturing response of oil wells in a tight reservoir," in *SPE Western Regional Meeting*, San Jose, CA, March 2009.

[53] J. Kacprzyk and P. Strykowski, "Linguistic summaries of sales data at a computer retailer via fuzzy logic and a genetic algorithm," in *Proc. of Congress on Evolutionary Computation*, vol. 2, Washington DC, 1999, pp. 937–943.

[54] J. Kacprzyk, A. Wilbik, and S. Zadrozny, "Linguistic summarization of time series using a fuzzy quantifier driven aggregation," *Fuzzy Sets and Systems*, vol. 159, pp. 1485–1499, 2008.

[55] J. Kacprzyk and S. Zadrożny, "Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools," *Information Sciences*, vol. 173, pp. 281–304, 2005.

[56] J. Kacprzyk, “Linguistic summaries of static and dynamic data: Computing with words and granularity,” in *Proc. of IEEE Int’l Conf. on Granular Computing*, silicon valley, CA, November 2007, pp. 4–5.

[57] N. N. Karnik and J. M. Mendel, “Centroid of a type-2 fuzzy set,” *Information Sciences*, vol. 132, pp. 195–220, 2001.

[58] N. K. Kasabov and Q. Song, “DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction,” *IEEE Trans. on Fuzzy Systems*, vol. 10, no. 2, pp. 144–154, 2002.

[59] A. Kaufmann, *Introduction to the Theory of Fuzzy Sets*. NY: Academic Press, 1975.

[60] ——, “Introduction a la theorie des sous-ensembles flous,” in *Complement et Nouvelles Applications*, Masson, Paris, 1977, vol. 4.

[61] C. S. Kim, D. S. Kim, and J. S. Park, “A new fuzzy resolution principle based on the antonym,” *Fuzzy Sets and Systems*, vol. 113, pp. 299–307, 2000.

[62] F. Klawonn, “Should fuzzy equality and similarity satisfy transitivity? comments on the paper by M. De Cock and E. Kerre,” *Fuzzy Sets and Systems*, vol. 133, pp. 175–180, 2003.

[63] E. P. Klement, “On the cardinality of fuzzy sets,” in *Proc. 6th European Meeting on Cybernetics and Systems Research*, Vienna, 1982, pp. 701–704.

[64] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ: Prentice-Hall, 1995.

[65] L. B. Koeppel, Y. Montagne-Miller, D. O’Hair, and M. J. Cody, “Friendly? flirting? wrong?” in *Interpersonal communication: Evolving interpersonal relationship*, P. J. Kalbfleisch, Ed. Hillsdale, NJ: Erlbaum, 1993, pp. 13–32.

[66] B. Kosko, “Fuzziness vs. probability,” *Int’l J. General Systems*, vol. 17, pp. 211–240, 1990.

[67] C. Lee, “Fuzzy logic in control systems: Fuzzy logic controller — part II,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 20, no. 2, pp. 419–435, 1990.

[68] S. S. Liao, T. H. Tang, and W.-Y. Liu, “Finding relevant sequences in time series containing crisp, interval, and fuzzy interval data,” *IEEE Trans. on Systems, Man, Cybernetics, B*, vol. 34, no. 5, pp. 2071–2079, 2004.

[69] F. Liu and J. M. Mendel, “Aggregation using the fuzzy weighted average, as computed using the Karnik-Mendel algorithms,” *IEEE Trans. on Fuzzy Systems*, vol. 12, no. 1, pp. 1–12, 2008.

[70] ——, "Encoding words into interval type-2 fuzzy sets using an interval approach," *IEEE Trans. on Fuzzy Systems*, vol. 16, no. 6, pp. 1503–1521, 2008.

[71] X. Liu, "The solution equivalence of minimax disparity and minimum variance problems for OWA operators," *International Journal of Approximate Reasoning*, vol. 45, pp. 68–81, 2007.

[72] J. Lu, G. Zhang, D. Ruan, and F. Wu, *Multi-Objective Group Decision Making.* London: Imperial College Press, 2007.

[73] P. Majlender, "OWA operators with maximal Renya entropy," *Fuzzy Sets and Systems*, vol. 155, pp. 340–360, 2005.

[74] I. Mani and M. Maybury, *Advances in automatic text summarization.* Cambridge, MA: MIT Press, 1989.

[75] M. Melgarejo, "A fast recursive method to compute the generalized centroid of an interval type-2 fuzzy set," in *Proc. Annual Meeting of the North American Fuzzy information processing society*, San Diego, CA, June 2007, pp. 190–194.

[76] C. Mencar, G. Castellano, and A. M. Fanelli, "Distinguishability quantification of fuzzy sets," *Information Sciences*, vol. 177, pp. 130–149, 2007.

[77] J. M. Mendel, "Computing with words, when words can mean different things to different people," in *Proc. 3rd Int'l ICSC Symposium on Fuzzy Logic and Applications*, Rochester, NY, June 1999, pp. 158–164.

[78] ——, "An architecture for making judgments using computing with words," *Int'l Journal of Applied Mathematics and Computer Science*, vol. 12, no. 3, pp. 325–335, 2002.

[79] ——, "Computing with words and its relationships with fuzzistics," *Information Sciences*, vol. 177, pp. 988–1006, 2007.

[80] J. M. Mendel and R. I. John, "Type-2 fuzzy sets made simple," *IEEE Trans. on Fuzzy Systems*, vol. 10, no. 2, pp. 117–127, April 2002.

[81] J. M. Mendel, R. I. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," *IEEE Trans. on Fuzzy Systems*, vol. 14, no. 6, pp. 808–821, 2006.

[82] J. M. Mendel, S. Murphy, L. C. Miller, M. Martin, and N. Karnik, "The fuzzy logic advisor for social judgments," in *Computing with words in information/intelligent systems*, L. A. Zadeh and J. Kacprzyk, Eds. Physica-Verlag, 1999, pp. 459–483.

[83] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions.* Upper Saddle River, NJ: Prentice-Hall, 2001.

[84] ——, "Computing with words: Zadeh, Turing, Popper and Occam," *IEEE Computational Intelligence Magazine*, vol. 2, pp. 10–17, 2007.

[85] J. M. Mendel and D. Wu, "Perceptual reasoning: A new computing with words engine," in *Proc. IEEE Int'l Conf. on Granular Computing*, Silicon Valley, CA, November 2007, pp. 446–451.

[86] ——, "Perceptual reasoning for perceptual computing," *IEEE Trans. on Fuzzy Systems*, vol. 16, no. 6, pp. 1550–1564, 2008.

[87] ——, "Computing with words for hierarchical and distributed decision making," in *Computational Intelligence in Complex Decision Systems*, D. Ruan, Ed. Paris, France: Atlantis Press, to appear in 2009.

[88] ——, *Perceptual Computing: Aiding People in Making Subjective Judgments*. Wiley-IEEE Press, to appear in 2010.

[89] I. Millet and T. L. Saaty, "On the relativity of relative measures – accommodating both rank preservation and rank reversals in the AHP," *European Journal of Operational Research*, vol. 121, no. 1, pp. 205–212, 2000.

[90] H. B. Mitchell, "Pattern recognition using type-II fuzzy sets," *Information Sciences*, vol. 170, no. 2-4, pp. 409–418, 2005.

[91] ——, "Ranking type-2 fuzzy numbers," *IEEE Trans. on Fuzzy Systems*, vol. 14, no. 2, pp. 287–294, 2006.

[92] S. Mohaghegh, S. Reeves, and D. Hill, "Development of an intelligent systems approach for restimulation candidate selection," in *SPE/CERI Gas Technology Symposium*, Calgary, Alberta, Canada, April 2000.

[93] D.-L. Mon, C.-H. Cheng, and J.-L. Lin, "Evaluating weapon system using fuzzy analytic hierarchy process based on entropy weight," *Fuzzy Sets and Systems*, vol. 62, pp. 127–134, 1994.

[94] M. G. Natrella, *Experimental Statistics*. Washington, DC: National Bureau of Standards, 1963.

[95] A. Niewiadomski, "A type-2 fuzzy approach to linguistic summarization of data," *IEEE Trans. on Fuzzy Systems*, vol. 16, no. 1, pp. 198–212, 2008.

[96] A. Niewiadomski and M. Bartyzel, "Elements of type-2 semantics in summarizing databases," *Lecture Notes in Artificial Intelligence*, vol. 4029, pp. 278–287, 2006.

[97] A. Niewiadomski, J. Ochelska, and P. Szczepaniak, "Interval-valued linguistic summaries of databases," *Control and Cybernetics*, vol. 35, no. 2, pp. 415–443, 2006.

[98] A. Niewiadomski and P. Szczepaniak, "News generating based on type-2 linguistic summaries of databases," in *Proc. IPMU*, Paris, France, July 2006, pp. 1324–1331.

[99] A. Niewiadomski, "On two possible roles of type-2 fuzzy sets in linguistic summaries," *Lecture Notes in Computer Science*, vol. 3528, pp. 341–347, 2005.

[100] ——, "Type-2 fuzzy summarization of data: An improved news generating," *Lecture Notes in Computer Science*, vol. 4585, pp. 241–250, 2007.

[101] M. Nikravesh, "Soft computing for reservoir characterization and management," in *Proc. IEEE Int'l Conf. Granular Computing*, vol. 2, Beijing, China, July 2005, pp. 593–598.

[102] V. Novaka, "Antonyms and linguistic quantifiers in fuzzy logic," *Fuzzy Sets and Systems*, vol. 124, pp. 335–351, 2001.

[103] D. Paulson and S. Zahir, "Consequences of uncertainty in the analytic hierarchy process: A simulation approach," *European Journal of Operational Research*, vol. 87, no. 1, p. 4556, 1995.

[104] W. Pedrycz, "Fuzzy set technology in knowledge discovery," *Fuzzy Sets and Systems*, vol. 98, pp. 279–290, 1998.

[105] M. A. Poyhonen, R. P. Hamalainen, and A. A. Salo, "An experiment on the numerical modeling of verbal ratio statements," *J. of Multi-criteria Decision Analysis*, vol. 6, no. 1-10, 1997.

[106] S. Raha, N. Pal, and K. Ray, "Similarity-based approximate reasoning: methodology and application," *IEEE Transactions on Systems, Man and Cybernetics–A*, vol. 32, no. 4, pp. 541– 547, 2002.

[107] G. Raschia and N. Mouaddib, "Using fuzzy labels as background knowledge for linguistic summarization of databases," in *Proc. FUZZ-IEEE*, Melbourne, Australia, December 2001, pp. 1372–1375.

[108] D. Rasmussen and R. Yager, "Finding fuzzy and gradual functional dependencies with SummarySQL," *Fuzzy Sets and Systems*, vol. 106, pp. 131–142, 1999.

[109] R. Reuven and K. Wan, "A simulation approach for handling uncertainty in the AHP," *European Journal of Operational Research*, vol. 106, no. 1, pp. 1116–1122, 1998.

[110] J. T. Rickard, J. Aisbett, and G. Gibbon, "Fuzzy subsethood for fuzzy sets of type-2 and generalized type-n," *IEEE Trans. on Fuzzy Systems*, vol. 17, no. 1, pp. 50–60, 2009.

[111] T. L. Saaty, "A scaling method for priorities in hierarchical structures," *Journal of Math. Psychology*, vol. 15, pp. 234–281, 1977.

[112] ——, *Decision making with the analytic network process: economic, political, social and technological applications with benefits, opportunities, costs and risks*. Springer, 2006.

[113] ——, "Rank from comparisons and from ratings in the analytic hierarchy/network processes," *European Journal of Operational Research*, vol. 168, no. 2, pp. 557–570, 2006.

[114] ——, *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill, 1980.

[115] ——, *Decision making for leaders: the analytical hierarchy process for decisions in a complex world*. Lifetime Learning Publications, 1982.

[116] ——, *Conflict resolution: the analytic hierarchy approach*. Praeger, 1989.

[117] ——, "How to make a decision: The analytic hierarchy process," *European Journal of Operational Research*, vol. 48, no. 1, pp. 9–26, 1990.

[118] ——, *Prediction, projection, and forecasting: applications of the analytic hierarchy process in economics, finance, politics, games, and sports*. Kluwer Academic Publishers, 1990.

[119] ——, *The Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*. RWS Publications, 2000.

[120] ——, "Decision-making with the AHP: Why is the principal eigenvector necessary," *European Journal of Operational Research*, vol. 145, no. 1, pp. 85–91, 2003.

[121] T. L. Saaty and L. T. Tran, "On the invalidity of fuzzifying numerical judgments in the analytic hierarchy process," *Mathematical and Computer Modeling*, vol. 46, pp. 962–975, 2007.

[122] T. Saaty and M. Ozdemir, "Why the magic number seven plus or minus two," *Mathematical and Computer Modeling*, vol. 38, no. 3, pp. 233–244, 2003.

[123] R. Saint-Paul, G. Raschia, and N. Mouaddib, "Database summarization: The SaintEtiQ system," in *Proc. of 2007 IEEE Int'l Conf. on Data Engineering*, Istanbul, Turkey, April 2007, pp. 1475–1476.

[124] M. Setnes, R. Babuska, U. Kaymak, H. R. Van, and N. Lemke, "Similarity measures in fuzzy rule base simplification," *IEEE Transactions on Systems, Man and Cybernetics–B*, vol. 28, no. 3, pp. 376–386, 1998.

[125] A. Silberschatz and A. Tuzhilin, "On subjective measures of interestingness in knowledge discovery," in *Proc. 1st Int'l. Conf. on Knowledge Discovery and Data Mining*, Menlo Park, CA, 1995, pp. 275–281.

[126] A. D. Soto and E. Trillas, "On antonym and negate in fuzzy logic," *International Journal of Intelligent Systems*, vol. 14, pp. 295–303, 1999.

[127] K. Sugihara and H. Tanaka, "Interval evaluations in the analytic hierarchy process by possibility analysis," *Computational Intelligence*, vol. 17, no. 3, pp. 567–579, 2001.

[128] Z. Świtalski, "General transitivity conditions for fuzzy reciprocal preference matrices," *Fuzzy Sets and Systems*, vol. 137, pp. 85–100, 2003.

[129] W. W. Tan and D. Wu, "Design of type-reduction strategies for type-2 fuzzy logic systems using genetic algorithms," in *Advances in Evolutionary Computing for System Design*, L. Jain, V. Palade, and D. Srinivasan, Eds. Springer, 2007, pp. 169–188.

[130] D. Timmermanns, "The roles of experience and domain of expertise in using verbal probability terms in medical decisions," *Medical Decision Making*, vol. 14, pp. 146–156, 1994.

[131] V. Torra and Y. Narukawa, *Modeling Decisions: Information Fusion and Aggregation Operators*. Berlin: Springer, 2007.

[132] I. B. Turksen, "Fuzzy data mining and expert system development," in *Proc. IEEE Int'l. Conf. Systems, Man and Cybernetics*, Oct. 1998, pp. 2057–2061.

[133] G.-H. Tzeng and J.-Y. Teng, "Transportation investment project selection with fuzzy multiobjectives," *Transportation Planning Technology*, vol. 17, pp. 91–112, 1993.

[134] M. Versaci and F. C. Morabito, "Fuzzy time series approach for disruption prediction in tokamak reactors," *IEEE Trans. on Magnetics*, vol. 39, no. 3, pp. 1503–1506, 2003.

[135] T. S. Wallsten and D. V. Budescu, "A review of human linguistic probability processing: General principles and empirical evidence," *The Knowledge Engineering Review*, vol. 10, no. 1, pp. 43–62, 1995.

[136] R. W. Walpole, R. H. Myers, A. Myers, and K. Ye, *Probability & Statistics for Engineers and Scientists*, 8th ed. Upper Saddle River, NJ: Prentice-Hall, 2007.

[137] H. Wang and D. Qiu, "Computing with words via turing machines: A formal approach," *IEEE Trans. on Fuzzy Systems*, vol. 11, no. 6, pp. 742–753, 2003.

[138] L.-X. Wang and J. M. Mendel, "Back-propagation of fuzzy systems as nonlinear dynamic system identifiers," in *Proc. FUZZ-IEEE*, San Diego, CA, 1992, pp. 1409–1418.

[139] ——, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning," *IEEE Trans. on Neural Networks*, vol. 3, pp. 807–813, 1992.

[140] ——, “Generating fuzzy rules by learning from examples,” *IEEE Trans. on Systems, Man and Cybernetics*, vol. 22, no. 2, pp. 1414–1427, 1992.

[141] L.-X. Wang, *A Course in Fuzzy Systems and Control*. Upper Saddle River, NJ: Prentice Hall, 1997.

[142] X. Wang and E. E. Kerre, “Reasonable properties for the ordering of fuzzy quantities (I),” *Fuzzy Sets and Systems*, vol. 118, pp. 375–387, 2001.

[143] ——, “Reasonable properties for the ordering of fuzzy quantities (II),” *Fuzzy Sets and Systems*, vol. 118, pp. 387–405, 2001.

[144] D. Wu and J. M. Mendel, “A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets,” *Information Sciences*, vol. 179, no. 8, pp. 1169–1192, 2009.

[145] ——, “Aggregation using the linguistic weighted average and interval type-2 fuzzy sets,” *IEEE Trans. on Fuzzy Systems*, vol. 15, no. 6, pp. 1145–1161, 2007.

[146] ——, “Enhanced Karnik-Mendel algorithms for interval type-2 fuzzy sets and systems,” in *Proc. NAFIPS*, San Diego, CA, June 2007, pp. 184–189.

[147] ——, “Uncertainty measures for interval type-2 fuzzy sets,” *Information Sciences*, vol. 177, no. 23, pp. 5378–5393, 2007.

[148] ——, “Corrections to “Aggregation using the linguistic weighted average and interval type-2 fuzzy sets”,” *IEEE Trans. on Fuzzy Systems*, vol. 16, no. 6, pp. 1664–1666, 2008.

[149] ——, “Perceptual reasoning using interval type-2 fuzzy sets: Properties,” in *IEEE Int’l Conf. on Fuzzy Systems*, Hong Kong, June 2008, pp. 1219–1226.

[150] ——, “A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets,” *Information Sciences*, vol. 178, no. 2, pp. 381–402, 2008.

[151] ——, “Average subsethood as a decoder for perceptual reasoning,” submitted for publication in *IEEE Trans. on Fuzzy Systems*, 2009.

[152] ——, “Computing with words for hierarchical decision making applied to evaluating a weapon system,” submitted for publication in *IEEE Trans. on Fuzzy Systems*, 2009.

[153] ——, “Computing with words for making social judgments,” submitted for publication in *IEEE Trans. on Systems, Man and Cybernetics-A*, 2009.

[154] ——, “Enhanced Karnik-Mendel Algorithms,” *IEEE Trans. on Fuzzy Systems*, 2009, in press.

[155] ——, "Ordered fuzzy weighted averages and ordered linguistic weighted averages," in *Proc. IFSA/EUSFLAT*, Lisbon, Portugal, July 2009.

[156] ——, "Perceptual reasoning for perceptual computing: A similarity-based approach," submitted for publication in *IEEE Trans. on Fuzzy Systems*, 2009.

[157] D. Wu and W. W. Tan, "A type-2 fuzzy logic controller for the liquid-level process," in *Proc. FUZZ-IEEE*, vol. 2, Budapest, Hungary, July 2004, pp. 953–958.

[158] ——, "Type-2 FLS modeling capability analysis," in *Proc. FUZZ-IEEE*, Reno, NV, May 2005, pp. 242–247.

[159] ——, "Genetic learning and performance evaluation of type-2 fuzzy logic controllers," *Int'l. J. Engineering Applications of Artificial Intelligence*, vol. 19, no. 8, pp. 829–841, 2006.

[160] ——, "A simplified type-2 fuzzy controller for real-time control," *ISA Transactions*, vol. 15, no. 4, pp. 503–516, 2006.

[161] M. Wygralak, "A new approach to the fuzzy cardinality of finite fuzzy sets," *Busefal*, vol. 15, pp. 72–75, 1983.

[162] "Xmdv tool home page." [Online]. Available: http://davis.wpi.edu/~xmdv/.

[163] Z. Xu, "Uncertain linguistic aggregation operators based approach to multiple attribute group decision making under uncertain linguistic environment," 2004, vol. 168, p. 171184.

[164] R. Yager, "Ranking fuzzy subsets over the unit interval," in *Proc. IEEE Conf. on Decision and Control*, vol. 17, 1978, pp. 1435–1437.

[165] ——, "A new approach to the summarization of data," *Information Sciences*, vol. 28, pp. 69–86, 1982.

[166] ——, "On ordered weighted averaging aggregation operators in multi-criteria decision making," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 18, pp. 183–190, 1988.

[167] ——, "On linguistic summaries of data," in *Knowledge Discovery in Databases*, G. Piatetsky-Shapiro and B. Frawley, Eds. MIT Press, 1991, pp. 347–363.

[168] ——, "Linguistic summaries as a tool for database discovery," in *Proc. FUZZ-IEEE*, Yokohama, Japan, 1995, pp. 79–82.

[169] ——, "Database discovery using fuzzy sets," *Int'l. J. Intelligent Systems*, vol. 11, pp. 691–712, 1996.

[170] ——, "A framework for multi-source data fusion," *Information Sciences*, vol. 163, pp. 175–200, 2004.

[171] R. Yager and D. Filev, *Essentials of Fuzzy Modeling and Control*. John Wiley & Son, 1994.

[172] R. Yager and J. Kacprzyk, *The Ordered Weighted Averaging Operators: Theory and Applications*. Norwell, MA: Kluwer, 1997.

[173] Y. Yao, "Granular computing for data mining," in *Proc. SPIE*, vol. 6241, Orlando, FL, 2006, p. 624105.

[174] D. S. Yeung and E. C. C. Tsang, "A comparative study on similarity-based fuzzy reasoning methods," *IEEE Trans. on Systems, Man and Cybernetics–B*, vol. 27, pp. 216–227, 1997.

[175] C. Yu, "A GPAHP method for solving group decision making fuzzy AHP problems," *Computers and Operations Research*, vol. 29, no. 14, pp. 1969–2001, 2002.

[176] L. Zadeh, "Fuzzy sets and information granularity," in *Advances in Fuzzy Set Theory and Applications*, M. Gupta, R. Ragade, and R. Yager, Eds. Amsterdam: North-Holland Publishing Co., 1979, pp. 3–18.

[177] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[178] ——, "Similarity relations and fuzzy orderings," *Information Sciences*, vol. 3, pp. 177–200, 1971.

[179] ——, "The concept of a linguistic variable and its application to approximate reasoning-1," *Information Sciences*, vol. 8, pp. 199–249, 1975.

[180] ——, "Possibility theory and soft data analysis," in *Mathematical Frontiers of the Social and Policy Sciences*, L. Cobb and R. M. Thrall, Eds. Boulder, CO: Westview Press, 1981, pp. 69–129.

[181] ——, "Fuzzy logic = Computing with words," *IEEE Trans. on Fuzzy Systems*, vol. 4, pp. 103–111, 1996.

[182] ——, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets and Systems*, vol. 19, pp. 111–127, 1997.

[183] ——, "From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions," *IEEE Trans. on Circuits and Systems–I: Fundamental Theory and Applications*, vol. 4, pp. 105–119, 1999.

[184] ——, "Toward a generalized theory of uncertainty (GTU)–an outline," *Information Sciences*, vol. 172, pp. 1–40, 2005.

[185] ——, "Toward human level machine intelligence – Is it achievable? The need for a paradigm shift," *IEEE Computational Intelligence Magazine*, vol. 3, no. 3, pp. 11–22, 2008.

[186] L. Zadeh, "Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems," *Soft Computing*, vol. 2, pp. 23–25, 1998.

[187] M. Zahir, "Incorporating the uncertainty of decision judgments in the analytic hierarchy process," *European Journal of Operational Research*, vol. 53, no. 2, pp. 206–216, 1996.

[188] W. Zeng and H. Li, "Relationship between similarity measure and entropy of interval valued fuzzy sets," *Fuzzy Sets and Systems*, vol. 157, pp. 1477–1484, 2006.

[189] S.-M. Zhou, F. Chiclana, R. I. John, and J. M. Garibaldi, "A practical approach to type-1 OWA operation for soft decision making," in *Proc. of the 8th Int'l FLINS Conf. on Computational Intelligence in Decision and Control*, Madrid, Spain, 2008, pp. 507–512.

[190] ——, "Type-1 OWA operators for aggregating uncertain information with uncertain weights induced by type-2 linguistic quantifiers," *Fuzzy Sets and Systems*, vol. 159, no. 24, pp. 3281–3296, 2008.

[191] ——, "Type-2 OWA operators – aggregating type-2 fuzzy sets in soft decision making," in *Proc. FUZZ-IEEE*, Hong Kong, June 2008, pp. 625–630.

[192] H.-J. Zimmermann, *Fuzzy Set Theory and Its Applications*, 4th ed. Boston/Dordrecht/London: Kluwer Academic Publishers, 2001.

# Appendix A

# The Enhanced Karnik-Mendel (EKM) Algorithms

The **EKM Algorithm for computing** $c_l(\tilde{X})$ in (2.22) is [146, 154]:

1. Sort $\underline{x}_i$ $(i = 1, 2, \ldots, N)$ in increasing order and call the sorted $\underline{x}_i$ by the same name, but now $\underline{x}_1 \le \underline{x}_2 \le \cdots \le \underline{x}_N$. Match the weights $w_i$ with their respective $\underline{x}_i$ and renumber them so that their index corresponds to the renumbered $\underline{x}_i$.

2. Set $k = [N/2.4]$ (the nearest integer to $N/2.4$), and compute

$$a = \sum_{i=1}^{k} \underline{x}_i \overline{w}_i + \sum_{i=k+1}^{N} \underline{x}_i \underline{w}_i \tag{A.1}$$

$$b = \sum_{i=1}^{k} \overline{w}_i + \sum_{i=k+1}^{N} \underline{w}_i \tag{A.2}$$

and

$$y = a/b \tag{A.3}$$

3. Find $k' \in [1, N-1]$ such that

$$\underline{x}_{k'} \le y \le \underline{x}_{k'+1} \tag{A.4}$$

4. Check if $k' = k$. If yes, stop, set $y_l = y$ and call $k$ $L$. If no, continue.

5. Compute $s = \text{sign}(k' - k)$, and[1]

---

[1] When $k' > k$, it is true that

$$a' = a + s \sum_{i=k+1}^{k'} \underline{x}_i(\overline{w}_i - \underline{w}_i), \qquad b' = b + s \sum_{i=k+1}^{k'} (\overline{w}_i - \underline{w}_i)$$

and when $k > k'$, it is true that

$$a' = a + s \sum_{i=k'+1}^{k} \underline{x}_i(\overline{w}_i - \underline{w}_i), \qquad b' = b + s \sum_{i=k'+1}^{k} (\overline{w}_i - \underline{w}_i).$$

(A.5) and (A.6) express the above two cases in a more concise form.

$$a' = a + s \sum_{i=\min(k,k')+1}^{\max(k,k')} \underline{x}_i(\overline{w}_i - \underline{w}_i) \tag{A.5}$$

$$b' = b + s \sum_{i=\min(k,k')+1}^{\max(k,k')} (\overline{w}_i - \underline{w}_i) \tag{A.6}$$

$$y' = a'/b' \tag{A.7}$$

6. Set $y = y'$, $a = a'$, $b = b'$ and $k = k'$. Go to Step 3.

The **EKM Algorithm for computing** $c_r(\tilde{X})$ in (2.23) is [146, 154]:

1. Sort $\overline{x}_i$ $(i = 1, 2, \dots, N)$ in increasing order and call the sorted $\overline{x}_i$ by the same name, but now $\overline{x}_1 \leq \overline{x}_2 \leq \cdots \leq \overline{x}_N$. Match the weights $w_i$ with their respective $\overline{x}_i$ and renumber them so that their index corresponds to the renumbered $\overline{x}_i$.

2. Set $k = [N/1.7]$ (the nearest integer to $N/1.7$), and compute

$$a = \sum_{i=1}^{k} \overline{x}_i \underline{w}_i + \sum_{i=k+1}^{N} \overline{x}_i \overline{w}_i \tag{A.8}$$

$$b = \sum_{i=1}^{k} \underline{w}_i + \sum_{i=k+1}^{N} \overline{w}_i \tag{A.9}$$

and

$$y = a/b \tag{A.10}$$

3. Find $k' \in [1, N-1]$ such that

$$\overline{x}_{k'} \leq y \leq \overline{x}_{k'+1} \tag{A.11}$$

4. Check if $k' = k$. If yes, stop, set $y_r = y$ and call $k$ $R$. If no, continue.

5. Compute $s = \text{sign}(k' - k)$, and

$$a' = a - s \sum_{i=\min(k,k')+1}^{\max(k,k')} \overline{x}_i(\overline{w}_i - \underline{w}_i) \tag{A.12}$$

$$b' = b - s \sum_{i=\min(k,k')+1}^{\max(k,k')} (\overline{w}_i - \underline{w}_i) \tag{A.13}$$

$$y' = a'/b' \tag{A.14}$$

6. Set $y = y'$, $a = a'$, $b = b'$ and $k = k'$. Go to Step 3.

# Appendix B

# Derivations of (3.20) and (3.21)

Consider $ss_l(\tilde{X}_1, \tilde{X}_2)$ first. Define

$$f_l(\mu_{X_1^e}(\mathbf{x})) = \frac{\sum_{i=1}^{N} \min\left(\mu_{X_1^e}(x_i), \mu_{\underline{X}_2}(x_i)\right)}{\sum_{i=1}^{N} \mu_{X_1^e}(x_i)} \tag{B.1}$$

where $\mu_{X_1^e}(x_i) \in [\mu_{\underline{X}_1}(x_i), \mu_{\overline{X}_1}(x_i)]$. Then,

$$ss_l(\tilde{X}_1, \tilde{X}_2) = \min_{\mu_{X_1^e}(x_i) \in \left[\mu_{\underline{X}_1}(x_i), \mu_{\overline{X}_1}(x_i)\right]} f_l(\mu_{X_1^e}(\mathbf{x})) \tag{B.2}$$

Let $X_l$ be the embedded T1 FS from which $ss_l(\tilde{X}_1, \tilde{X}_2)$ is computed. For a particular $x_j$, there are three possible relationships between $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_1}(x_j)]$ and $\mu_{\underline{X}_2}(x_j)$:

1. When $\mu_{\underline{X}_1}(x_j) \geq \mu_{\underline{X}_2}(x_j)$, i.e., the entire interval $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_1}(x_j)]$ is larger than or equal to $\mu_{\underline{X}_2}(x_j)$, it follows that $\min(\mu_{X_1^e}(x_j), \mu_{\underline{X}_2}(x_j)) = \mu_{\underline{X}_2}(x_j)$, and hence

$$\frac{d(\min(\mu_{X_1^e}(x_j), \mu_{\underline{X}_2}(x_j)))}{d\mu_{X_1^e}(x_j)} = 0 \tag{B.3}$$

$$\frac{\partial f_l(\mu_{X_1^e}(\mathbf{x}))}{\partial \mu_{X_1^e}(x_j)} = -\frac{\sum_{i=1}^N \min(\mu_{X_1^e}(x_i), \mu_{\underline{X}_2}(x_i))}{\left(\sum_{i=1}^N \mu_{X_1^e}(x_i)\right)^2} \leq 0 \tag{B.4}$$

i.e., $f_l(\mu_{X_1^e}(\mathbf{x}))$ decreases as $\mu_{X_1^e}(x_j)$ increases; so, $ss_l(\tilde{X}_1, \tilde{X}_2)$, the minimum of $f_l(\mu_{X_1^e}(\mathbf{x}))$, is obtained when $\mu_{X_1^e}(x_j) = \mu_{\overline{X}_1}(x_j)$, i.e., $\mu_{X_l}(x_j) = \mu_{\overline{X}_1}(x_j)$ when $\mu_{\underline{X}_2}(x_j) \leq \mu_{\underline{X}_1}(x_j)$, which is the first line of (3.18).

2. When $\mu_{\overline{X}_1}(x_j) \leq \mu_{\underline{X}_2}(x_j)$, i.e., the entire interval $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_1}(x_j)]$ is smaller than or equal to $\mu_{\underline{X}_2}(x_j)$, it follows that $\min(\mu_{X_1^e}(x_j), \mu_{\underline{X}_2}(x_j)) = \mu_{X_1^e}(x_j)$, and hence

$$\frac{d(\min(\mu_{X_1^e}(x_j), \mu_{\underline{X}_2}(x_j)))}{d\mu_{X_1^e}(x_j)} = 1 \tag{B.5}$$

$$\frac{\partial f_l(\mu_{X_1^e}(\mathbf{x}))}{\partial \mu_{X_1^e}(x_j)} = \frac{\sum_{i=1}^N \mu_{X_1^e}(x_i) - \sum_{i=1}^N \min(\mu_{X_1^e}(x_i), \mu_{\underline{X}_2}(x_i))}{\left(\sum_{i=1}^N \mu_{X_1^e}(x_i)\right)^2} \geq 0 \tag{B.6}$$

The second part of (B.6) is true because $\mu_{X_1^e}(x_i) \geq \min(\mu_{X_1^e}(x_i), \mu_{\underline{X}_2}(x_i))$ for $\forall i$. (B.6) indicates that $f_l(\mu_{X_1^e}(\mathbf{x}))$ decreases as $\mu_{X_1^e}(x_j)$ decreases; so, $ss_l(\tilde{X}_1, \tilde{X}_2)$, the minimum of $f_l(\mu_{X_1^e}(\mathbf{x}))$, is obtained when $\mu_{X_1^e}(x_j) = \mu_{\underline{X}_1}(x_j)$, i.e., $\mu_{X_l}(x_j) = \mu_{\underline{X}_1}(x_j)$ when $\mu_{\underline{X}_2}(x_j) \geq \mu_{\overline{X}_1}(x_j)$, which is the second line of (3.18).

3. When $\mu_{\underline{X}_1}(x_j) < \mu_{\underline{X}_2}(x_j) < \mu_{\overline{X}_1}(x_j)$, i.e., $\mu_{\underline{X}_2}(x_j)$ is within the interval $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_1}(x_j)]$, $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_1}(x_j)]$ can be partitioned into two sub-intervals, $[\mu_{\underline{X}_1}(x_j),$

$\mu_{\underline{X}_2}(x_j)]$ and $[\mu_{\underline{X}_2}(x_j), \mu_{\overline{X}_1}(x_j)]$, and then the minimum for each sub-interval can be computed. The minimum over the entire interval is the smaller one of the minimums of the two sub-intervals. Note that for sub-interval $[\mu_{\underline{X}_1}(x_j), \mu_{\underline{X}_2}(x_j)]$, which is smaller than or equal to $\mu_{\underline{X}_2}(x_j)$, the result in Case 2 can be used, and the minimum is obtained when $\mu_{X_1^e}(x_j) = \mu_{\underline{X}_1}(x_j)$; and for sub-interval $[\mu_{\underline{X}_2}(x_j), \mu_{\overline{X}_1}(x_j)]$, which is larger than or equal to $\mu_{\underline{X}_2}(x_j)$, the result in Case 1 can be used, and the minimum is obtained when $\mu_{X_1^e}(x_j) = \mu_{\overline{X}_1}(x_j)$. $ss_l(\tilde{X}_1, \tilde{X}_2)$ is obtained by computing $f_l(\mu_{X_1^e}(\mathbf{x}))$ for both $\mu_{\underline{X}_1}(x_j)$ and $\mu_{\overline{X}_1}(x_j)$ and choosing the smaller value. This means, that when $\mu_{\underline{X}_1}(x_j) < \mu_{\underline{X}_2}(x_j) < \mu_{\overline{X}_1}(x_j)$, $\mu_{X_l}(x_j) = \{\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_1}(x_j)\}$, which is the third line of (3.18).

(3.20) is a summarization of the above results.

Consider $ss_r(\tilde{X}_1, \tilde{X}_2)$ next. Define

$$f_r(\mu_{X_1^e}(\mathbf{x})) = \frac{\sum_{i=1}^{N} \min\left(\mu_{X_1^e}(x_i), \mu_{\overline{X}_2}(x_i)\right)}{\sum_{i=1}^{N} \mu_{X_1^e}(x_i)} \tag{B.7}$$

where $\mu_{X_1^e}(x_i) \in [\mu_{\underline{X}_1}(x_i), \mu_{\overline{X}_1}(x_i)]$. Then,

$$ss_r(\tilde{X}_1, \tilde{X}_2) = \max_{\mu_{X_1^e}(x_i) \in \left[\mu_{\underline{X}_1}(x_i), \mu_{\overline{X}_1}(x_i)\right]} f_r(\mu_{X_1^e}(\mathbf{x})) \tag{B.8}$$

Let $X_1^e$ be the embedded T1 FS from which $ss_r(\tilde{X}_1, \tilde{X}_2)$ is computed. Again, for a particular $x_j$, there are three possible relationships between $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_1}(x_j)]$ and $\mu_{\overline{X}_2}(x_j)$:

1. When $\mu_{\underline{X}_1}(x_j) \geq \mu_{\overline{X}_2}(x_j)$, i.e., the entire interval $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_1}(x_j)]$ is larger than or equal to $\mu_{\overline{X}_2}(x_j)$, it follows that $\min(\mu_{X_1^e}(x_j), \mu_{\overline{X}_2}(x_j)) = \mu_{\overline{X}_2}(x_j)$, and hence

$$\frac{d(\min(\mu_{X_1^e}(x_j), \mu_{\overline{X}_2}(x_j)))}{d\mu_{X_1^e}(x_j)} = 0 \tag{B.9}$$

$$\frac{\partial f_r(\mu_{X_1^e}(\mathbf{x}))}{\partial \mu_{X_1^e}(x_j)} = -\frac{\sum_{i=1}^N \min(\mu_{X_1^e}(x_i), \mu_{\overline{X}_2}(x_i))}{\left(\sum_{i=1}^N \mu_{X_1^e}(x_i)\right)^2} \leq 0 \tag{B.10}$$

So, $ss_r(\tilde{X}_1, \tilde{X}_2)$ is obtained when $\mu_{X_1^e}(x_j) = \mu_{\underline{X}_1}(x_j)$, i.e., $\mu_{X_r}(x_j) = \mu_{\underline{X}_1}(x_j)$ when $\mu_{\overline{X}_2}(x_j) \leq \mu_{\underline{X}_1}(x_j)$, which is the first line of (3.19).

2. When $\mu_{\overline{X}_1}(x_j) \leq \mu_{\overline{X}_2}(x_j)$, i.e., the entire interval $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_1}(x_j)]$ is smaller than or equal to $\mu_{\overline{X}_2}(x_j)$, it follows that $\min(\mu_{X_1^e}(x_j), \mu_{\overline{X}_2}(x_j)) = \mu_{X_1^e}(x_j)$, and hence

$$\frac{d(\min(\mu_{X_1^e}(x_j), \mu_{\overline{X}_2}(x_j)))}{d\mu_{X_1^e}(x_j)} = 1 \tag{B.11}$$

$$\frac{\partial f_r(\mu_{X_1^e}(\mathbf{x}))}{\partial \mu_{X_1^e}(x_j)} = \frac{\sum_{i=1}^N \mu_{X_1^e}(x_i) - \sum_{i=1}^N \min(\mu_{X_1^e}(x_i), \mu_{\overline{X}_2}(x_i))}{\left(\sum_{i=1}^N \mu_{X_1^e}(x_i)\right)^2} \geq 0 \tag{B.12}$$

The second part of (B.12) is true because $\mu_{X_1^e}(x_i) \geq \min(\mu_{X_1^e}(x_i), \mu_{\overline{X}_2}(x_i))$ for $\forall i$. So, $ss_r(\tilde{X}_1, \tilde{X}_2)$ is obtained when $\mu_{X_1^e}(x_j) = \mu_{\overline{X}_1}(x_j)$, i.e., $\mu_{X_r}(x_j) = \mu_{\overline{X}_1}(x_j)$ when $\mu_{\overline{X}_2}(x_j) \geq \mu_{\overline{X}_1}(x_j)$, which is the second line of (3.19).

3. When $\mu_{\underline{X}_1}(x_j) < \mu_{\overline{X}_2}(x_j) < \mu_{\overline{X}_1}(x_j)$, i.e., $\mu_{\overline{X}_2}(x_j)$ is within the interval $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_1}(x_j)]$, $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_1}(x_j)]$ can be partitioned into two sub-intervals, $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_2}(x_j)]$ and $[\mu_{\overline{X}_2}(x_j), \mu_{\overline{X}_1}(x_j)]$, and then the maximum for each sub-interval can be computed. The maximum over the entire interval is the larger one of the maximums

of the two sub-intervals. Note that for sub-interval $[\mu_{\underline{X}_1}(x_j), \mu_{\overline{X}_2}(x_j)]$, which is smaller than or equal to $\mu_{\overline{X}_2}(x_j)$, the result in Case 2 can be used [where $\mu_{\overline{X}_2}(x_j)$ plays the role of $\mu_{\overline{X}_1}(x_j)$], and the maximum is obtained when $\mu_{X_1^e}(x_j) = \mu_{\overline{X}_2}(x_j)$; and for sub-interval $[\mu_{\overline{X}_2}(x_j), \mu_{\overline{X}_1}(x_j)]$, which is larger than or equal to $\mu_{\overline{X}_2}(x_j)$, the result in Case 1 can be used [where $\mu_{\overline{X}_2}(x_j)$ plays the role of $\mu_{\overline{X}_1}(x_j)$], and the maximum is also obtained when $\mu_{X_1^e}(x_j) = \mu_{\overline{X}_2}(x_j)$. So, $ss_r(\tilde{X}_1, \tilde{X}_2)$ is obtained when $\mu_{X_1^e}(x_j) = \mu_{\overline{X}_2}(x_j)$, i.e., $\mu_{X_r}(x_j) = \mu_{\overline{X}_2}(x_j)$ when $\mu_{\underline{X}_1}(x_j) < \mu_{\overline{X}_2}(x_j) < \mu_{\overline{X}_1}(x_j)$, which is the third line of (3.19).

(3.21) is a summarization of the above results.

# Appendix C

# The Analytic Hierarchy Process (AHP)

The Harvard psychologist Arthur Blumenthal [8] pointed out that there are two types of judgments: comparative judgment which *is the identification of some relation between two stimuli both present to the observer*, and absolute judgment which *is the identification of the magnitude of some simple stimulus. ... Absolute judgment involves the relation between a single stimulus and some information held in short-term memory — information about some former comparison stimuli or about some previously experienced measurement scale. On that basis, an observer identifies or rates a single stimulus.* In the AHP the first type of judgment is called *relative measurement* and the second is called *absolute measurement* [113]. In relative measurement each alternative is compared with many other alternatives, and in absolute measurement each alternative is compared with one ideal alternative the decision-maker knows of or can imagine, a process called "rating alternatives." Novel weighted averages introduced in Chapter 4 use absolute measurements (i.e., each alternative is evaluated independently). The *distributive mode AHP* introduced in this appendix uses relative measurements.

## C.1 The Distributive Mode AHP

In the *distributive mode AHP*, pair-wise comparisons are used to obtain the weights for the criteria and for the scores of the alternatives for each criterion, and then a weighted average is used to compute the overall performance of each alternative. It consists of the following four steps: 1) Identify the alternatives and criteria, 2) Compute the weights for the criteria, 3) Compute the priorities of the alternatives for each criterion, and, 4) Compute the overall priorities of the alternatives. These four steps are explained in more detail next[1].

### C.1.1 Identify the Alternatives and Criteria

In this step, first the alternatives that will be compared in a specific MCDM problem are identified. Denote them as $A_i$, $i = 1, ..., n$. Then, the major criteria for comparing the alternatives are identified[2]. Denote these criteria as $C_j$, $j = 1, ..., m$.

[1]There are many variants of the distributive model AHP, e.g., ideal mode AHP [119] and logarithmic least-squares method [11, 19].

[2]Each major criterion can have several sub-criteria; however, for simplicity no sub-criteria are considered in this section.

### C.1.2 Compute the Weights for the Criteria

Once the criteria are identified, their weights are computed through pair-wise comparisons[3]. A *pair-wise comparison matrix* (PCM) $W$ is constructed, whose $ij^{\text{th}}$ element, $c_{ij}$, is the ratio of the importance of $C_i$ to the importance of $C_j$. The comparisons are performed linguistically using the terms shown in the second column of Table C.1, and then the corresponding numerical intensities are used to fill the appropriate positions in matrix $W$.

Table C.1: The fundamental scale [117] for AHP. A scale of absolute numbers is used to assign numerical values to judgments made by comparing two elements, with the less important one used as the unit and the more important one assigned a value from this scale as a multiple of that unit.

| Intensity[a] | Definition | Explanation |
|---|---|---|
| 1 | Equal importance[b] | Two elements contribute equally to the objective |
| 3 | Moderate importance | Experience and judgment slightly favor one element over the other |
| 5 | Strong importance | Experience and judgment strongly favor one element over the other |
| 7 | Very strong or demonstrated importance | One element is favored very strongly over the other; its dominance is demonstrated in practice |
| 9 | Extreme importance | The evidence favoring one element over the other is of the highest possible order of affirmation |

[a] Intensities of 2, 4, 6 and 8 can be used for compromise between the above values.
[b] Intensities $\{1.1, \ldots, 1.9\}$ can also be used when elements are close and nearly indistinguishable.

Because it always holds that $c_{ii} = 1$ and $c_{ji} = 1/c_{ij}$, a total of $m(m-1)/2$ (instead of $m^2$) pair-wise comparisons need to be made, as illustrated in (C.1).

$$W = [c_{ij}]_{i,j=1,\ldots,m} \text{ where } c_{ii} = 1 \text{ and } c_{ji} = 1/c_{ij} \tag{C.1}$$

[3]This is very different from the way in which the weights are chosen when using a NWA.

This matrix cannot be used directly in the final aggregation. A *weight vector* $\mathbf{w} = (w_1, ..., w_m)^T$ corresponding to the weights of the criteria is needed. This $\mathbf{w}$ must be deduced from $W$. It has been shown [114] that $\mathbf{w}$ should be the principle eigenvector of $W$. Usually, $\mathbf{w}$ is normalized so that $\sum_{j=1}^{m} w_j = 1$.

### C.1.3 Compute the Priorities of the Alternatives for Each Criterion

The priorities of the $n$ alternatives to the $m$ criteria need to be determined so that they can be aggregated to obtain the overall priority. For criterion $C_k$ $(k = 1, ..., m)$, using a similar approach as used above to construct $W$, a PCM $X_k$ is constructed, as:

$$X_k = [a_{ij}]_{i,j=1,\ldots,n} \text{ where } a_{ii} = 1 \text{ and } a_{ji} = 1/a_{ij} \tag{C.2}$$

in which $a_{ij}$ is the relative importance of alternative $A_i$ over alternative $A_j$. Then, the normalized principal eigenvector of $X_k$, $\mathbf{x}_k$, is computed to represent the priorities of the alternatives for criterion $C_k$. Once this is done for all $m$ criteria, one ends up with $m$ *priority vectors* $\mathbf{x}_k$, $k = 1, ..., m$.

### C.1.4 Compute the Overall Priorities of the Alternatives

In the final step of the AHP, a vector $\mathbf{p} = (p_1, ..., p_n)^T$, representing the overall priorities of the $n$ alternatives, is derived from $\mathbf{x}_k$ $(k = 1, ..., m)$ and $\mathbf{w}$, as:

$$\mathbf{p} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_m]\mathbf{w} \tag{C.3}$$

Usually $\mathbf{p}$ is normalized so that $\sum_{i=1}^{n} p_i = 1$, though the normalization does not change the overall priorities of the alternatives.

## C.2 Example

The following example [1] is used to illustrate the procedures of the AHP.

**Example 29** *Suppose a family wants to buy a new car, and they consider four criteria: $C_1$ = Cost, $C_2$ = Safety, $C_3$ = Style and $C_4$ = Capacity. There are three candidates for selection: $A_1$ = Accord Sedan, $A_2$ = Pilot SUV and $A_3$ = Odyssey Minivan. The complete AHP hierarchy is shown in Fig. C.1.*

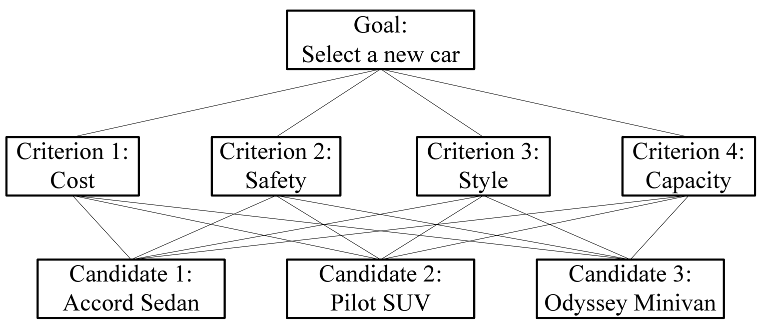



Fig. C.1: The AHP hierarchy for car selection.

*The relative importance of the four criteria is determined first by pair-wise comparisons. Assume the family thinks Cost is equally important as Safety ($c_{12} = 1$), very strongly more important than Style ($c_{13} = 7$), and moderately more important than Capacity ($c_{14} = 3$); Safety is extremely more important than Style ($c_{23} = 9$) and moderately* *more important than Capacity ($c_{24} = 3$); and, Capacity is strongly more important than Style ($c_{43} = 5$). Then, the PCM is constructed as*

$$W = \begin{array}{c} \\ Cost \\ Safety \\ Style \\ Capacity \end{array} \begin{array}{c} \begin{array}{cccc} Cost & Safety & Style & Capacity \end{array} \\ \left[\begin{array}{cccc} 1 & 1 & 7 & 3 \\ 1 & 1 & 9 & 3 \\ 1/7 & 1/9 & 1 & 1/5 \\ 1/3 & 1/3 & 5 & 1 \end{array}\right] \end{array} \tag{C.4}$$

*The relative importance of the three criteria is computed as the principle eigenvector of $W$, which is*

$$\mathbf{w} = (0.39,\ 0.41,\ 0.04,\ 0.16)^T \tag{C.5}$$

*According to the typical prices of the three models and the family's budget, they think Accord is very strongly preferred to Pilot ($a_{12} = 7$) and strongly preferred to Odyssey ($a_{13} = 5$), and Odyssey is moderately preferred to Pilot ($a_{32} = 3$); so, they construct the PCM $X_1$, for cost, as*

$$X_1 \text{ (for Cost)} = \begin{array}{c} \\ Accord \\ Pilot \\ Odyssey \end{array} \begin{array}{c} \begin{array}{ccc} Accord & Pilot & Odyssey \end{array} \\ \begin{bmatrix} 1 & 7 & 5 \\ 1/7 & 1 & 1/3 \\ 1/5 & 3 & 1 \end{bmatrix} \end{array}$$

*Assume for the other three criteria, after some research the family gives the following PCMs:*

$$X_2 \text{ (for Safety)} = \begin{array}{c} \\ Accord \\ Pilot \\ Odyssey \end{array} \begin{array}{c} \begin{array}{ccc} Accord & Pilot & Odyssey \end{array} \\ \begin{bmatrix} 1 & 1/3 & 1/7 \\ 3 & 1 & 1/5 \\ 7 & 5 & 1 \end{bmatrix} \end{array}$$

$$X_3 \text{ (for Style)} = \begin{array}{c} \\ Accord \\ Pilot \\ Odyssey \end{array} \begin{array}{c} \begin{array}{ccc} Accord & Pilot & Odyssey \end{array} \\ \begin{bmatrix} 1 & 5 & 3 \\ 1/5 & 1 & 1/3 \\ 1/3 & 3 & 1 \end{bmatrix} \end{array}$$

$$X_4 \text{ (for Capacity)} = \begin{array}{c} \\ Accord \\ Pilot \\ Odyssey \end{array} \begin{array}{c} \begin{array}{ccc} Accord & Pilot & Odyssey \end{array} \\ \begin{bmatrix} 1 & 1/5 & 1/5 \\ 5 & 1 & 1 \\ 5 & 1 & 1 \end{bmatrix} \end{array}$$

*Then, the corresponding priority vectors are*

$$\mathbf{x}_1 = (0.73,\ 0.08,\ 0.19)^T \tag{C.6}$$

$$\mathbf{x}_2 = (0.08,\ 0.19,\ 0.73)^T \tag{C.7}$$

$$\mathbf{x}_3 = (0.64,\ 0.10,\ 0.26)^T \tag{C.8}$$

$$\mathbf{x}_4 = (0.10,\ 0.45,\ 0.45)^T \tag{C.9}$$

*Consequently, the overall priority of the three cars is*

$$\mathbf{p} = [\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3\ \mathbf{x}_4]\mathbf{w} = (0.36,\ 0.18,\ 0.46)^T \tag{C.10}$$

*So, the choice would be the Odyssey Minivan.* ■

## C.3 AHP versus NWA

Comparisons of the NWA and AHP are shown in Table C.2. Each method has four steps, but only Step 1 is common to both the NWA and AHP. Weights, scores of the alternatives, and final rank are computed differently by the NWA and AHP.

Table C.2: A comparison of the NWA and AHP.

| **Step** | **NWA** | **AHP** |
|---|---|---|
| 1. Identify criteria and alternatives | This step is common to both the NWA and AHP. | |
| 2. Find weights for the criteria | Decision-makers express the weights linguistically, and then IT2 FSs are used to represent them. | A PCM is constructed, and then a weight vector is computed from it. |
| 3. Find scores of the alternatives for each criterion | Decision-makers express the scores linguistically, and then IT2 FSs are used to represent them. | A PCM is constructed for each criterion, and then a priority vector is computed from it. |
| 4. Compute the final rank | An NWA is computed for each alternative to obtain its overall performance, and then the final IT2 FSs are ranked. Similarities among the ranked alternatives are also computed. | The priority vectors are weighted by the weight vector to obtain the overall priority vector. |