Linguistic Summarization Using IF–THEN Rules and Interval Type-2 Fuzzy Sets

Dongrui Wu, *Member, IEEE*, and Jerry M. Mendel, *Life Fellow, IEEE*

***Abstract*—Linguistic summarization (LS) is a data mining or knowledge discovery approach to extract patterns from databases. Many authors have used this technique to generate summaries like "Most senior workers have high salary," which can be used to better understand and communicate about data; however, few of them have used it to generate IF–THEN rules like "IF X is large and Y is medium, THEN Z is small," which not only facilitate understanding and communication of data but can also be used in decision-making. In this paper, an LS approach to generate IF–THEN rules for causal databases is proposed. Both type-1 and interval type-2 fuzzy sets are considered. Five quality measures—the degrees of truth, sufficient coverage, reliability, outlier, and simplicity—are defined. Among them, the *degree of reliability* is especially valuable for finding the most reliable and representative rules, and the *degree of outlier* can be used to identify outlier rules and data for close-up investigation. An improved parallel coordinates approach for visualizing the IF–THEN rules is also proposed. Experiments on two datasets demonstrate our LS and rule visualization approaches. Finally, the relationships between our LS approach and the Wang–Mendel (WM) method, perceptual reasoning, and granular computing are pointed out.**

***Index Terms*—Data mining, fuzzy set (FS), granular computing, IF–THEN rules, interval type-2 (IT2) FS, knowledge discovery, linguistic summarization (LS), parallel coordinates, perceptual reasoning, rule visualization, Wang–Mendel (WM) method.**

I. Introduction

THE RAPID progress of information technology has made huge amounts of data accessible to people. 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 [35], "*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 paper means to [48] "*grasp and briefly describe trends and characteristics appearing in a dataset, without doing (explicit) manual 'record-by-record' analysis.*"

There can be two approaches to summarize a dataset: numerical summarization and linguistic summarization (LS). Statistical characteristics, such as mean, median, variance, etc., are examples of numerical summarization; however, as pointed out by Yager [75], "*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 LS of databases, which outputs summaries like "*Most senior workers are well-paid*" or "*IF X is large and Y is medium, THEN Z is small*," 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 LS of databases [9], [10], [53], [55], and time series [7], [25]. The fuzzy set (FS) based approach, introduced by Yager [75]–[78] and advanced by many others [13], [25], [28], [48], [53], [56], is the most popular one. It has been used in

1) LS of sales data [26]–[28], e.g., *about one half of sales in autumn is of accessories, much sales on saturday is about noon*, etc.;
2) LS of worker information [47], [50], e.g., *about half of workers are about 30, many of workers, who are about 30 earn about 4000*, etc.;
3) LS of the performance of intelligent algorithms [47], e.g., *about half of scores given by Algorithm 2 are equal to scores by Expert 3, many scores given by Algorithm 1 are equal or almost equal to the median*, etc.;
4) LS of time series [25], e.g., *among all trends of a low variability most are short, among all medium trends, at least around a half is of medium variability*, etc.

Most of the previous works focus on type-1 (T1) FSs [38], [80]. Niewiadomski *et al.* [47]–[52] are to date the only ones working on LS using interval and general type-2 FSs (see [38], [40]–[42], [46], [64], [81]; see also Section III-A).

In this paper, we focus on generating IF–THEN rules from *causal*[1] databases, e.g., "*IF X is large and Y is medium, THEN Z is small*," because our primary goal is to use LS to generate a rulebase for decision-making [43], [44], [46], [64], [66], and IF–THEN rules are used in almost all fuzzy logic systems rather than Yager *et al.*'s summaries. There have been only a few publications [20]–[22] in this direction, e.g., Ishibuchi *et al.* [21]

Manuscript received February 2, 2010; revised June 18, 2010 and October 3, 2010; accepted October 9, 2010. Date of publication October 18, 2010; date of current version February 7, 2011.

D. Wu was with the Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564 USA. He is now with the Industrial Artificial Intelligence Laboratory, General Electric Global Research, Niskayuna, NY 12309 USA (e-mail: drwu09@gmail.com).

J. M. Mendel is with the Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: mendel@sipi.usc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2010.2088128

[1] According to Wikipedia [2], "*causality is the relationship between an event (the cause) and a second event (the effect), where the second event is a consequence of the first.*" In this paper, we consider more general cases, where there can be multiple causes.

TABLE I
EXPLANATIONS OF THE SYMBOLS USED IN THIS PAPER

Symbol	Meaning	Example
$\mathbb{D}$	The complete database	Auto MPG dataset [1]
$\mathbb{Y}$	The set of all objects in the database	All auto models in the database
M	Number of objects in $\mathbb{Y}$	The total number of models (392)
y_m	The m^{th} object in the database	The m^{th} model in the dataset
v_n	Name of the n^{th} attribute	#cylinder (1^{st} attribute)
$\mathbb{X}_n$	The domain of v_n	[3, 8] for #cylinder
$\mathbb{V}$	A set of all attribute names	<#cylinder, Displacement, ..., MPG>
v_n^m	Value of the n^{th} attribute for y_m	8 (#cylinder of the 1^{st} model)
$\mathbf{d}_m$	A complete record related to y_m with values assigned to all attributes in $\mathbb{V}$	<8, 307, ..., 18> for the 1^{st} model
S_n	Summarizer	Small horsepower, Around 1971, etc.
Q	Quantifier	Most, about half, more than 100, etc.
w_g	Qualifier	Small horsepower, Around 1971, etc.
T	Degree of truth	Any value in $[0, 1]$
C	Degree of sufficient coverage	Any value in $[0, 1]$
R	Degree of reliability	Any value in $[0, 1]$
O	Degree of outlier	Any value in $[0, 1]$
S	Degree of simplicity	Any value in $[0, 1]$
Q_M	Quality measure, any of T, C, R, O, and S	Any value in $[0, 1]$

$n = 1, 2, \ldots, N$ and $m = 1, 2, \ldots, M$.

and Ishibuchi and Yamamoto [22] generated weighted rules like "*IF x_1 is small and x_2 is large, THEN Class1 with w*" for pattern classification. Our work can be viewed as an extension of theirs. Our contributions are as follows.

1) We use interval type-2 (IT2) FSs instead of T1 FSs in the IF–THEN rules. As argued in Section III-B, IT2 FSs enables us to model both intrapersonal and interpersonal uncertainties about linguistic terms, whereas T1 FSs can only model intrapersonal uncertainties.
2) We introduce five quality measures (QMs) (the degrees of truth, sufficient coverage, reliability, outlier, and simplicity) to quantify different properties of the IF–THEN rules. Degrees of reliability and outliers, which are the most important QMs in this paper, have not been used by others.
3) We propose a parallel coordinates approach for rule visualization. It is the first time that such a visualization approach is introduced to the fuzzy logic community.

The rest of this paper is organized as follows: Section II introduces our LS approach to generate IF–THEN rules using T1 FSs and its associated QMs. Section III extends the results in Section II to IT2 FSs. Section IV illustrates our LS approach for two datasets. Section V discusses the relationships between our LS approach and the Wang–Mendel (WM) method, perceptual reasoning, and granular computing. Section VI draws conclusions.

II. Linguistic Summarization Using T1 Fuzzy Sets

The main purpose of this paper is to propose an LS approach using IT2 FSs. For ease in understanding, we start with LS using T1 FSs; however, this does not mean we advocate that T1 FSs should be used in LS. In fact, we always argue that IT2 FSs should be used in LS, because they can model both intrapersonal and interpersonal uncertainties, as explained in the next section.

A. Data Description

Let us define[2] a set of M objects $\mathbb{Y} = \{y_1, y_2, \ldots, y_M\}$ and a set of N attributes $\mathbb{V} = \{v_1, v_2, \ldots, v_N\}$. 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 nth attribute for the mth object $(m = 1, 2, \ldots, M)$. Hence, the database $\mathbb{D}$, which collects information about elements from $\mathbb{Y}$, is in the form of

$$\begin{aligned}\mathbb{D} &= \{\langle v_1^1, v_2^1, \ldots, v_N^1\rangle, \langle v_1^2, v_2^2, \ldots, v_N^2\rangle, \ldots \\ &\quad \langle v_1^M, v_2^M, \ldots, v_N^M\rangle\} \\ &\equiv \{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_M\} \qquad (1)\end{aligned}$$

where $\mathbf{d}_m = \langle v_1^m, v_2^m, \ldots, v_N^m\rangle$ is a complete record about object y_m.

For example, for the auto miles per gallon (MPG) dataset [1] used in Section IV-A, there are 392 auto models ($M = 392$), and hence, $\mathbb{Y} = \{$Model1, Model2, ..., Model392$\}$. Each model has eight attributes ($N = 8$), and $\mathbb{V} = \langle$#cylinder, displacement, horsepower, weight, acceleration, year, origin, MPG$\rangle$. For #cylinder, its value ranges from 3 to 8; therefore, its domain $\mathbb{X}_1 = [3, 8]$. Model1, which was a U.S. car made in 1970, has eight cylinders, 307 displacement, 130 hp, weighs 3504 lb, 12 s acceleration, and 18 mi/gal. Therefore, the complete record for Model1 is $\mathbf{d}_1 = \langle$ 8, 307, 130, 3504, 12, 1970, U.S., 18$\rangle$.

B. Linguistic Summarization Using IF–THEN Rules and Type-1 Fuzzy Sets

Only single-antecedent single-consequent rules are considered in this section. Multiantecedent multiconsequent rules are considered in Sections II-J and III-E.

Because we are interested in generating IF–THEN rules from a causal dataset, our *canonical form for LS using T1 FSs* is as

[2]For easy reference, our most frequently used symbols are summarized in Table I.

follows:

$$\text{IF } v_1 \text{ is/has } S_1, \text{ THEN } v_2 \text{ is/has } S_2 \quad [Q_M] \tag{2}$$

where S_1 and S_2 are words modeled by T1 FSs,[3] and $Q_M \in [0, 1]$ is a *QM*, which indicates how good the rule is. One example of such a rule is as follows:

$$\text{IF } \underbrace{\text{horsepower}}_{v_1} \text{ is } \underbrace{\text{large}}_{S_1}, \text{ THEN } \underbrace{\text{MPG}}_{v_2} \text{ is } \underbrace{\text{very low}}_{S_2} [Q_M]. \tag{3}$$

Once a dataset is given, the antecedents and consequents of the rules are determined. A user needs to specify the words used for each antecedent and consequent, as well as their corresponding FS models. Then, all possible combinations of the rules can be constructed. The challenge is to compute Q_M, which can have different definitions.

C. Quality Measures of LS Using Type-1 Fuzzy Sets

According to Hirota and Pedrycz [19], the following five features[4] 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 [57] "*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.
5) *Simplicity*: This measure concerns the syntactic complexity of the summaries. Generally, simpler summaries are easier to understand and, hence, are preferred.

Next, we propose five QMs for T1 FS LS, corresponding to *validity*, *generality*, *usefulness*, *novelty*, and *simplicity*, respectively.

D. Degree of Truth T

Validity is represented by the *degree of truth* T, which is computed as follows:

$$T = \frac{\sum_{m=1}^{M} \min(\mu_{S_1}(v_1^m), \mu_{S_2}(v_2^m))}{\sum_{m=1}^{M} \mu_{S_1}(v_1^m)}. \tag{4}$$

T is the same as Kosko's subsethood measure [30] for T1 FSs. This kind of formula has also been used in Zadeh's calculus of linguistically quantified proposition to assess the truth value of a linguistic proposition [83], computing the conditional probability for fuzzy events [58], the confidence of (fuzzy) association rules [11], [20]–[22], the fuzzy matching degree of the SaintEtiQ approach [54], and the certainty factor of a decision rule [14]. Roughly speaking, T increases as more data satisfying the antecedent also satisfy the consequent.

A different representation of the degree of truth T defined in (4) is introduced next, because it will lead easily to the computation of T for LS using IT2 FSs, as will be shown in Section III-C; first, two related definitions are introduced.

Definition 1: The *cardinality* of a T1 FS S_1 on database $\mathbb{D}$ is defined as follows:

$$c_{\mathbb{D}}(S_1) = \sum_{m=1}^{M} \mu_{S_1}(v_1^m) \tag{5}$$

where v_1^m is the value of the mth datum in the universe of discourse of S_1.

Definition 2: The *joint cardinality* of T1 FSs $\{S_1, \ldots, S_N\}$ on database $\mathbb{D}$ is defined as follows:

$$c_{\mathbb{D}}(S_1, \ldots, S_N) = \sum_{m=1}^{M} \min\{\mu_{S_1}(v_1^m), \ldots\ldots, \mu_{S_N}(v_N^m)\}. \tag{6}$$

Using the cardinality $c_{\mathbb{D}}(S_1)$ and joint cardinality $c_{\mathbb{D}}(S_1, S_2)$, (4) can be reexpressed as follows:

$$T = \frac{c_{\mathbb{D}}(S_1, S_2)}{c_{\mathbb{D}}(S_1)}. \tag{7}$$

It is worthwhile to mention the analogy between (7) and conditional probability in probability theory. Consider S_1 and S_2 in (2) as two events. Then, the conditional probability of S_2 given S_1, $P(S_2|S_1)$ is computed as follows:

$$P(S_2|S_1) = \frac{P(S_1, S_2)}{P(S_1)} \tag{8}$$

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 (7), the numerator can be viewed as the total degree that S_1 and S_2 are satisfied simultaneously [which is analogous to $P(S_1, S_2)$], and the denominator can be viewed as the total degree that only the prerequisite S_1 is satisfied [which is analogous to $P(S_1)$].

E. Degree of Sufficient Coverage C

Generality is represented by the *degree of sufficient coverage* C, which describes whether a rule is supported by enough data. It is independent of the degree of truth because a rule with high C may have low T, i.e., there are many data supporting this rule, but also many data not supporting this rule. To compute C, we first compute the *coverage ratio*, which is as follows:

$$r_c = \frac{\sum_{m=1}^{M} t_m}{M} \tag{9}$$

where

$$t_m = \begin{cases} 1, & \mu_{S_1}(v_1^m) > 0 \quad \text{and} \quad \mu_{S_2}(v_2^m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

[3]These T1 FS word models are predefined before LS is carried out. They can be easily constructed by users who are familiar with FSs.

[4]There are many other QMs for association rules in the literature, e.g., confirmation measure [14], interestingness measure [11], [12], etc. We use Hirota and Pedrycz's five measures, since they adequately quantify the properties of a summary from different aspects. Other QMs will be considered in our future research.

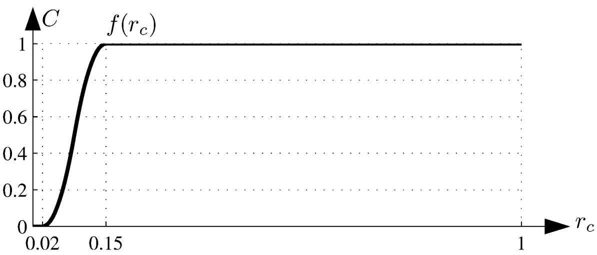

Fig. 1. S-shape function $f(r_c)$ used in this paper.

i.e., r_c is the percentage of data, which fit both the antecedent and the consequent of the rule at nonzero degrees. Because each rule only covers a small region of the high-dimensional input–output space, r_c is usually very small (e.g., mostly smaller than 0.1). Therefore, $r_c = 0.15$ 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 about sufficient coverage:

$$C = f(r_c) \tag{11}$$

where f is a function that maps r_c into C. The S-shape function $f(r_c)$ used in this paper is shown in Fig. 1. It is determined by two parameters r_1 and r_2 ($0 \le r_1 < r_2$), i.e.,

$$f(r_c) = \begin{cases} 0, & r_c \le r_1 \\ 2\left(\dfrac{r_c - r_1}{r_2 - r_1}\right)^2, & r_1 < r_c < \dfrac{r_1 + r_2}{2} \\ 1 - 2\left(\dfrac{r_2 - r_c}{r_2 - r_1}\right)^2, & \dfrac{r_1 + r_2}{2} \le r_c < r_2 \\ 1, & r_c \ge r_2 \end{cases} \tag{12}$$

and $r_1 = 0.02$ and $r_2 = 0.15$ are used in this paper. $f(r_c)$ can be modified according to the user's requirement about sufficient coverage.

F. Degree of Reliability R

The *degree of reliability* R, as its name suggests, describes how reliable a summary is. A rule is reliable if and only if we have the following.

1) It has high degree of truth, i.e., most of the data satisfying the rule's antecedents also have the behavior described by its consequent.
2) It has sufficient coverage, i.e., enough data are described by it.

Hence, R is computed as follows:

$$R = \min(T, C). \tag{13}$$

G. Degree of Outlier O

Novelty means *unexpectedness*. There are different understandings of unexpectedness, e.g., the *degree of appropriateness* defined by Kacprzyk and Strykowski [24] considers the independency of the summarizers. In this paper, unexpectedness is related to the *degree of outlier O*, which indicates the possibility that a rule describes only outliers instead of a useful pattern. Clearly, the degree of sufficient coverage for an outlier rule must be very small, i.e., it only describes very few data; however, small C alone is not enough to identify outliers rules, and the degree of truth should also be considered. When C is small, T can be small (close to 0), medium (around 0.5), or large (close to 1), as shown in Fig. 2, where the rule "IF v_1 is Low, THEN v_2 is High" is illustrated for three different cases.

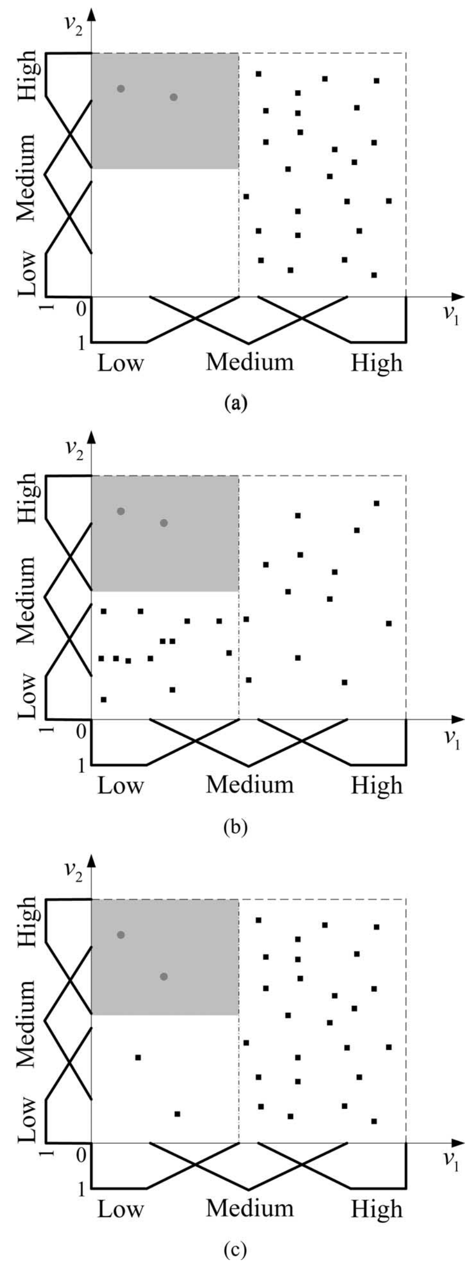

Fig. 2. Three cases for the rule "IF v_1 is Low, THEN v_2 is High," whose C is small. (a) T is large. (b) T is small. (c) T is medium.

1) For the rule illustrated by the shaded region in Fig. 2(a), T is large because all data satisfying the antecedent (v_1 is Low) also satisfy the consequent (v_2 is High), i.e., $\sum_{m=1}^{M} \min(\mu_{\text{Low}}(v_1^m), \mu_{\text{High}}(v_2^m))$ is close to $\sum_{m=1}^{M} \mu_{\text{Low}}(v_1^m)$. Visual inspection suggests that this

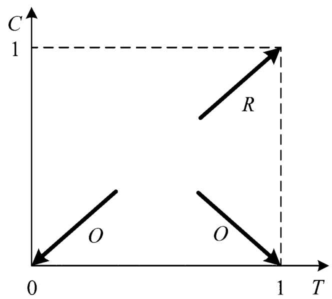

Fig. 3. Useful rules and outlier rules determined by T and C.

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. 2(b), T is small because most data satisfying the antecedent (v_1 is Low) do not satisfy the consequent (v_2 is High), i.e., $\sum_{m=1}^{M} \min(\mu_{\text{Low}}(v_1^m), \mu_{\text{High}}(v_2^m))$ much smaller than $\sum_{m=1}^{M} \mu_{\text{Low}}(v_1^m)$. Visual inspection suggests that this rule should also 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. 2(c), T is medium because the data satisfying the antecedent (v_1 is Low) are distributed somewhat uniformly in the v_2 domain, i.e., $\sum_{m=1}^{M} \min(\mu_{\text{Low}}(v_1^m), \mu_{\text{High}}(v_2^m))$ is about half of $\sum_{m=1}^{M} \mu_{\text{Low}}(v_1^m)$. By visual inspection, this rule should not be considered as an outlier (although it is not a good rule as R would be small) because its data are not so isolated from the rest.

In summary, an outlier rule must satisfy the following.

1) The degree of truth T must be very small or very large.
2) The degree of sufficient coverage C must be very small.

Finally, note that the purpose of finding an outlier rule is to help people identify possible outlier data, and then, to further investigate them. Therefore, we need to exclude a rule with $T = 0$ from being identified as an outlier because in this case the rule does not describe any data. The following formula is used in this paper to compute the degree of outlier is as follows:

$$O = \begin{cases} \min(\max(T, 1-T), 1-C), & T > 0 \\ 0, & T = 0. \end{cases} \tag{14}$$

The term $\max(T, 1-T)$ converts a small T (close to 0) or a large T (close to 1) to a large number in [0, 1], which is required by the first criterion of an outlier rule, and $\min(\max(T, 1-T), 1-C)$ further imposes the constraint that C must be small, which is the second criterion for an outlier rule. Note that the closer O is to 1, the more a rule is judged to be an outlier.

A graph illustrating the dependence of R in (13) and O in (14) on T and C is shown in Fig. 3. R or O increases as (T, C) moves in the directions indicated by the arrows, e.g., R moves toward 1 as both T and C increase.

H. Degree of Simplicity S

The *simplicity* of a summary can be measured by its length, i.e., how many antecedents and consequents the rule has. We define the *degree of simplicity* S of a rule by

$$S = 2^{2-l} \tag{15}$$

where l is the total number of antecedents and consequents of the rule. Clearly, $S \in (0, 1]$, and the simplest rule ($S = 1$) has only one antecedent and one consequent. As the number of antecedents and/or consequents increases, S decreases, and a rule becomes more difficult to understand and communicate.

TABLE II
CORRESPONDENCES BETWEEN THE CONCEPTS PROPOSED BY HIROTA AND PEDRYCZ [19] AND OUR QMS

Hirota and Pedrycz's Concepts	Our Quality Measures
Validity	Degree of truth (T)
Generality	Degree of sufficient coverage (C)
Usefulness	Degree of reliability (R)
Novelty	Degree of outlier (O)
Simplicity	Degree of simplicity (S)

I. Summary of the Quality Measures

A summary of the correspondences between the concepts proposed by Hirota and Pedrycz [19] and our QMs is given in Table II. Note that Hirota and Pedrycz only proposed the concepts but did not define these measures.

J. Multiantecedent Multiconsequent Rules

The generalization of the results for single-antecedent single-consequent rules to multiantecedent multiconsequent rules is straightforward. Consider the following multiantecedent multiconsequent rule:

$$\begin{aligned} &\text{IF } v_1 \text{ is/has } S_1 \text{ and} \ldots \text{and } v_K \text{ is/has } S_K \\ &\text{THEN } v_{K+1} \text{ is/has } S_{K+1} \text{ and} \ldots \text{and } v_N \text{ is/has } S_N \quad [Q_M]. \end{aligned} \tag{16}$$

The degree of truth T is computed as follows:

$$T = \frac{c_{\mathbb{D}}(S_1, \ldots, S_N)}{c_{\mathbb{D}}(S_1, \ldots, S_K)}. \tag{17}$$

The coverage ratio r_c is computed by redefining t_m as follows:

$$t_m = \begin{cases} 1, & \mu_{S_n}(v_n^m) > 0 \quad \forall n = 1, \ldots, N \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

Once r_c is obtained, C is computed by (11). Because both T and C are crisp numbers, (13) and (14) can again be used to compute R and O. The degree of simplicity S is still computed by (15).

Comment: Lee [31] considers multiantecedent multiconsequent rules in fuzzy logic control. By assuming, the consequents are independent control actions, he proposes to decompose such a rule into q multiantecedent single-consequent rules (see [31, p. 426]), where q is the number of consequents in the original multiantecedent multiconsequent 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 degree of truth, 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 a multiantecedent multiconsequent rule is decomposed into multiantecedent single-consequent rules. For example, the rule in (16) is not equivalent to the combination of the following $N-K$ multiantecedent single-consequent rules:

IF v_1 is/has S_1 and ... and v_K is/has S_K

THEN v_{K+1} is/has S_{K+1} $[T_1]$

IF v_1 is/has S_1 and ... and v_K is/has S_K

THEN v_{K+2} is/has S_{K+2} $[T_2]$

$\vdots$

IF v_1 is/has S_1 and ... and v_K is/has S_K

THEN v_N is/has S_N $[T_{N-K}]$.

III. Linguistic Summarization Using Interval Type-2 Fuzzy Sets

The canonical form of LS using IT2 FSs and its associated QMs are proposed in this section. All are extensions of the previous section's results on LS using T1 FSs.

A. Interval Type-2 Fuzzy Sets

A T1 FS has membership grades that are crisp, whereas an IT2 FS [38], [40]–[42], [46], [64], [81] has membership grades that are intervals. Such a set is particularly useful in circumstances, where it is difficult to determine the exact membership function (MF) for an FS, e.g., approximate reasoning [15], [66], [69], recognition and classification [36], [74], [86], system modeling and control [5], [6], [16], [17], [23], [32], [33], [38], [61], [70]–[73], word modeling [34], [45], [46], [67], etc.

Definition 3 [38], [41]: An IT2 FS $\tilde{A}$ is characterized by the MF $\mu_{\tilde{A}}(x,u)$, where $x \in X$ and $u \in J_x \subseteq [0,1]$, i.e.,

$$\tilde{A} = \{((x,u), \mu_{\tilde{A}}(x,u) = 1) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\} \tag{19}$$

where x, which is called the *primary variable*, has domain X; $u \in [0,1]$, which is called the *secondary variable*, has domain $J_x \subseteq [0,1]$ at each $x \in X$; J_x is also called the *primary membership* of x, and is defined in (21), and $\mu_{\tilde{A}}(x,u)$, which is called a *secondary grade* of x, equals 1 for $\forall x \in X$ and $\forall u \in J_x \subseteq [0,1]$.

An example of an IT2 FS is shown in Fig. 4. It can be viewed as a blurred T1 FS, and all elements in the blurred area have the same secondary membership grade, which is 1.

Definition 4: Uncertainty about $\tilde{A}$ is conveyed by the union of all its primary memberships, which is called the *footprint of uncertainty* (FOU) of $\tilde{A}$ (see Fig. 4), i.e.,

$$FOU(\tilde{A}) = \bigcup_{\forall x \in X} J_x. \tag{20}$$

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

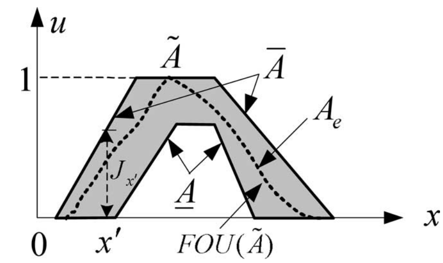

Fig. 4. IT2 FS and its associated quantities.

Definition 5: The *upper MF* and *lower MF* of $\tilde{A}$ are two T1 FSs $\overline{A}$ and $\underline{A}$ that bound the FOU (see Fig. 4).

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

$$J_x = \left[\mu_{\underline{A}}(x), \mu_{\overline{A}}(x)\right]. \tag{21}$$

Using (21), FOU($\tilde{A}$) can also be expressed as follows:

$$\text{FOU}(\tilde{A}) = \bigcup_{\forall x \in X} \left[\mu_{\underline{A}}(x), \mu_{\overline{A}}(x)\right]. \tag{22}$$

A very compact way to describe an IT2 FS is as follows:

$$\tilde{A} = 1/\text{FOU}(\tilde{A}) \tag{23}$$

where this notation means that the secondary grade equals 1 for all elements of FOU($\tilde{A}$). 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 6: An *embedded T1 FS* A_e of $\tilde{A}$ is as follows:

$$A_e = \int_{x \in X} u/x, \quad u \in J_x \tag{24}$$

where $\int$ means union instead of integral.

The upper and lower MFs represent two embedded T1 FSs.

Finally, note that there are more general T2 FSs [38] for which the secondary grades are different over the FOU and that an IT2 FS is a special case of those T2 FSs.

B. Which Type of Fuzzy Sets Should Be Used to Model Words in Linguistic Summarization

Both T1 and IT2 FSs have been used in modeling words [38], [84]. In this paper, we suggest that IT2 FSs should be used in LS for the following reasons.

1) There are at least two types of uncertainties associated with a word [39], [60]: *intrapersonal uncertainty* and *interpersonal uncertainty*. Intrapersonal uncertainty describes [39] "*the uncertainty a person has about the word*." It is also explicitly pointed out by Wallsten and Budescu [60] 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 T1 FS. Interpersonal uncertainty describes [39] "*the uncertainty that a group of people have about the word*." It is pointed out by Mendel [38] as *"words mean different things to different people"* and

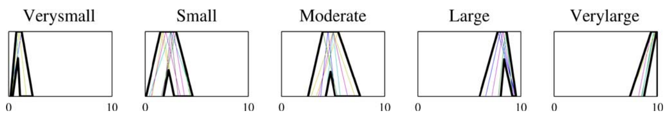

Fig. 5. Five examples of word FOUs obtained from the interval approach [34]. The areas between the thick curves are FOUs, and the curves within the FOUs are embedded T1 FSs mapped from individuals' endpoint data.

Wallsten and Budescu [60] as *"different individuals use diverse expressions to describe identical situations and understand the same phrases differently when hearing or reading them."* Because an IT2 FS has an FOU, which can be viewed as a group of T1 FSs (see Fig. 5), it can model both types of uncertainty [39]; hence, we suggest IT2 FSs be used in modeling words [37]–[39], [46], [64].

2) IT2 FS word models in LS can be constructed from the interval approach [34] or the enhanced interval approach [8] (see also Section III-C).By default, both approaches output an IT2 FS model for each word; however, if there is only one user, or all users give the same boundary for a word, then that word is modeled as a T1 FS. Therefore, starting from IT2 FS word models does not eliminate the possibility of T1 FS word models but not *vice versa*.

C. Linguistic Summarization Using IF–THEN Rules and Interval Type-2 Fuzzy Sets

When IT2 FSs are used in LS to generate IF–THEN rules, our canonical form in (2) becomes

$$\text{IF } v_1 \text{ is/has } \tilde{S}_1, \text{ THEN } v_2 \text{ is/has } \tilde{S}_2 \quad [Q_M] \tag{25}$$

where $\tilde{S}_1$ and $\tilde{S}_2$ are words modeled by IT2 FSs, and $Q_M \in [0,1]$ is a QM.

The IT2 FS word models should be constructed before LS is carried out. This can be done with the *interval approach* [34] or *enhanced interval approach* [8]. First, for each word in an *application-dependent* encoding vocabulary, a group of subjects are asked the following question:

On a scale of $x_{\min}$ to $x_{\max}$, what are the end-points of an interval that you associate with the word ____?

After some preprocessing, 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*. A simple codebook is shown in Fig. 5. Software for the interval approach and enhanced interval approach can be downloaded from J. M. Mendel's website at http://sipi.usc.edu/∼mendel.

Next, we explain how to compute the five different QMs.

D. Quality Measures for Linguistic Summarization Using Interval Type-2 Fuzzy Sets

Recall from (7) that the degree of truth for LS 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 7: The *cardinality* of an IT2 FS $\tilde{S}_1$ on dataset $\mathbb{D}$ is defined as follows:

$$C_{\mathbb{D}}(\tilde{S}_1) \equiv [c_{\mathbb{D}}(\underline{S}_1), c_{\mathbb{D}}(\overline{S}_1)] = \left[\sum_{m=1}^{M} \mu_{\underline{S}_1}(v_1^m), \sum_{m=1}^{M} \mu_{\overline{S}_1}(v_1^m)\right] \tag{26}$$

and the *average cardinality* is as follows:

$$c_{\mathbb{D}}(\tilde{S}_1) = \frac{c_{\mathbb{D}}(\underline{S}_1) + c_{\mathbb{D}}(\overline{S}_1)}{2}. \tag{27}$$

Definition 8: The *joint cardinality* of IT2 FSs $\{\tilde{S}_1, \ldots, \tilde{S}_N\}$ on database $\mathbb{D}$ is defined as follows:

$$\begin{aligned} C_{\mathbb{D}}(\tilde{S}_1, \ldots, \tilde{S}_N) &\equiv \left[c_{\mathbb{D}}(\underline{S}_1, \ldots, \underline{S}_N), c_{\mathbb{D}}(\overline{S}_1, \ldots, \overline{S}_N)\right] \\ &= \left[\sum_{m=1}^{M} \min\{\mu_{\underline{S}_1}(v_1^m), \ldots, \mu_{\underline{S}_N}(v_N^m)\},\right. \\ &\qquad \left.\sum_{m=1}^{M} \min\{\mu_{\overline{S}_1}(v_1^m), \ldots, \mu_{\overline{S}_N}(v_N^m)\}\right] \end{aligned} \tag{28}$$

and the *average joint cardinality* is as follows:

$$c_{\mathbb{D}}(\tilde{S}_1, \ldots, \tilde{S}_N) = \frac{c_{\mathbb{D}}(\underline{S}_1, \ldots, \underline{S}_N) + c_{\mathbb{D}}(\overline{S}_1, \ldots, \overline{S}_N)}{2}. \tag{29}$$

A straightforward extension of (7) to LS using IT2 FSs is to define a truth quantity

$$\tilde{T} = \frac{C_{\mathbb{D}}(\tilde{S}_1, \tilde{S}_2)}{C_{\mathbb{D}}(\tilde{S}_1)}. \tag{30}$$

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. However, as it is difficult and unnecessary[5] to compute an interval truth quantity, a crisp degree of truth is defined in this paper based on average cardinalities instead of cardinalities.

By substituting the cardinalities in (7) by their respective average cardinalities, T in (25) is computed as follows:

$$T = \frac{c_{\mathbb{D}}(\tilde{S}_1, \tilde{S}_2)}{c_{\mathbb{D}}(\tilde{S}_1)} \tag{31}$$

[5] $\tilde{T}$ cannot be computed using simple interval arithmetic, i.e.,

$$\tilde{T} \neq \left[\frac{\sum_{m=1}^{M} \min\{\mu_{\underline{S}_1}(v_1^m), \mu_{\underline{S}_2}(v_2^m)\}}{\sum_{m=1}^{M} \mu_{\overline{S}_1}(v_1^m)}, \frac{\sum_{m=1}^{M} \min\{\mu_{\overline{S}_1}(v_1^m), \mu_{\overline{S}_2}(v_2^m)\}}{\sum_{m=1}^{M} \mu_{\underline{S}_1}(v_1^m)}\right]$$

because $\tilde{S}_1$ appears in both the numerator and the denominator of (30), 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 endpoints in the aforementioned equation, different embedded T1 FSs of $\tilde{S}_1$ are used in the numerator and the denominator (e.g., $\underline{S}_1$ is used in the numerator of the first term in the aforementioned equation, whereas $\overline{S}_1$ is used in the denominator). Although it is possible to derive an interval $\tilde{T}$ based on the representation theorem for IT2 FSs [42], the computation is complicated, and as explained at the end of this section, it is also unnecessary.

which is essentially Vlachos and Sergiadis's subsethood measure [59], [64], [68] for interval-valued FSs.

Like its T1 counterpart (see Section II-B), (31) is also analogous to the conditional probability $P(\tilde{S}_2|\tilde{S}_1)$, which is computed as follows:

$$P(\tilde{S}_2|\tilde{S}_1) = \frac{P(\tilde{S}_1, \tilde{S}_2)}{P(\tilde{S}_1)} \tag{32}$$

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 prerequisite $\tilde{S}_1$ is satisfied [analogous to $P(\tilde{S}_1)$].

For LS using IT2 FSs, the coverage ratio is still computed by (9), but t_m is defined differently

$$t_m = \begin{cases} 1, & \mu_{\overline{S}_1}(v_1^m) > 0 \quad \text{and} \quad \mu_{\overline{S}_2}(v_2^m) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{33}$$

i.e., we count all objects with nonzero membership (i.e., J_x in (21) does not equal [0, 0]) on both antecedent and consequent. Once the coverage ratio r is obtained, the degree of sufficient coverage is computed by (11). Because both T and C are crisp numbers, (13) and (14) can again be used to compute the degree of reliability and the degree of outlier. The degree of simplicity S is still computed by (15).

Comment: A reader may argue that information is lost when the QM of an IT2 FS linguistic summary is described 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 quality 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, and hence, it is necessary to model the terms used in the content of a rule by IT2 FSs. The QM is used to rank the rules and, hence, to find the best rules; however, how it should be used in decision-making is still an open problem. A single-number QM is easier to compute and more convenient in ranking rules than an interval measure; therefore, the former is used in this paper.

E. Multiantecedent Multiconsequent Rules

The generalization of the results for single-antecedent single-consequent rules to multiantecedent multiconsequent rules is straightforward. Consider the following multiantecedent multiconsequent rule:

IF v_1 is/has $\tilde{S}_1$ and ... and v_K is/has $\tilde{S}_K$

THEN v_{K+1} is/has $\tilde{S}_{K+1}$ and ... and v_N is/has $\tilde{S}_N$ $[T]$. (34)

The degree of truth T is computed as follows:

$$T = \frac{c_{\mathbb{D}}(\tilde{S}_1, \ldots, \tilde{S}_N)}{c_{\mathbb{D}}(\tilde{S}_1, \ldots, \tilde{S}_K)} \tag{35}$$

and the coverage ratio r_c is computed by redefining t_m as follows:

$$t_m = \begin{cases} 1, & \mu_{\overline{S}_n}(v_n^m) > 0 \quad \forall n = 1, \ldots, N \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

Once r_c is obtained, C is computed by (11). Because both T and C are crisp numbers, (13) and (14) can again be used to compute R and O. The degree of simplicity S is still computed by (15).

IV. Applications

A MATLAB-based GUI was created to demonstrate the IT2 FS LS approach. Two functions were implemented.

1) *Global top rules:* Given the number of antecedents, the program finds top rules that give the maximum T, C, R, or O.
2) *Local top rules:* Given the number of antecedents and a desired attribute, the program finds top rules that contain that attribute.

Two datasets were used, and their results are presented in this section.

A. Auto Miles Per Gallon Dataset

The auto MPG dataset was obtained from the University of California at Irvine (UCI) machine-learning repository [1]. It contains 392 entries (after removing incomplete entries) about the configurations of auto models and their MPGs. LS was used to find the relationship between the following inputs and *MPG*, which is a continuous value in [9, 46.6]:

1) *#cylinder*: Discrete values in {3, 4, 5, 6, 8};
2) *Displacement*: Continuous values in [68, 455];
3) *Horsepower*: Continuous values in [46, 230];
4) *Weight*: Continuous values in [1613, 5140];
5) *Acceleration*: Continuous values in [8, 24.8];
6) *Model year*: Integer values in [1970, 1982];
7) *Origin*: Categorical values in {U.S., Germany, Japan}.

This dataset was chosen because the attributes consist of both continuous and discrete, and both numerical and categorical, values. Therefore, the ability of LS to handle diverse attributes can be demonstrated.

The "global top rules" function is used to automatically find global top rules according to the ranking criterion a user chooses. Figs. 6–9 show global top rules when[6] T, C, R, and O are used as the ranking criterion, respectively. A user first specifies the number of antecedents. The program then computes T, C, R, and O for all possible combinations of words with such number of antecedents. By default, top rules are selected according to R and displayed at the top-left corner of the GUI; however, a user can change the ranking criterion by clicking on the four push buttons on the top-right corner of the GUI. The rules are then updated accordingly.

A user can also click on a certain radio button to select a specific rule. All cases that support and violate that rule are displayed by an improved parallel coordinates approach [4] in the middle of the GUI, where each coordinate represents an attribute, and the two numbers labeled at the two ends of each coordinate represent the range of that attribute, e.g., observe

[6]The degree of simplicity O was not considered because in the GUI, all rules have the same number of antecedents and consequents, i.e., O for all rules are equal.

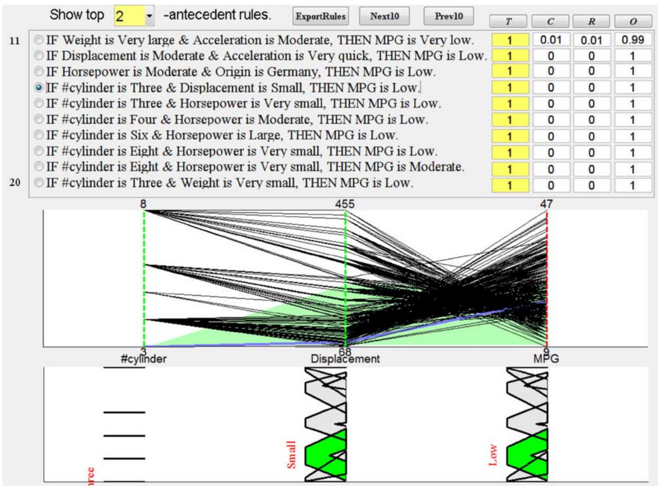

Fig. 6. Auto MPG dataset. Global top 11–20 rules according to T: the degree of truth. The middle and bottom parts illustrate the $14\mathrm{th}$ rule.

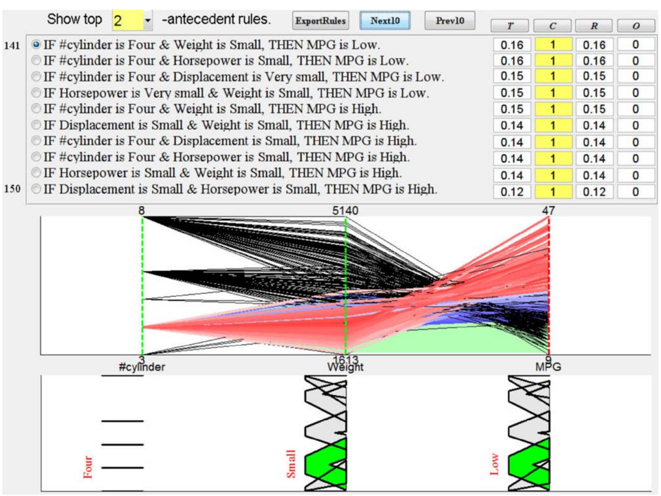

Fig. 7. Auto MPG dataset. Global top 141–150 rules according to C: the degree of sufficient coverage. The middle and bottom parts illustrate the $141\mathrm{st}$ rule.

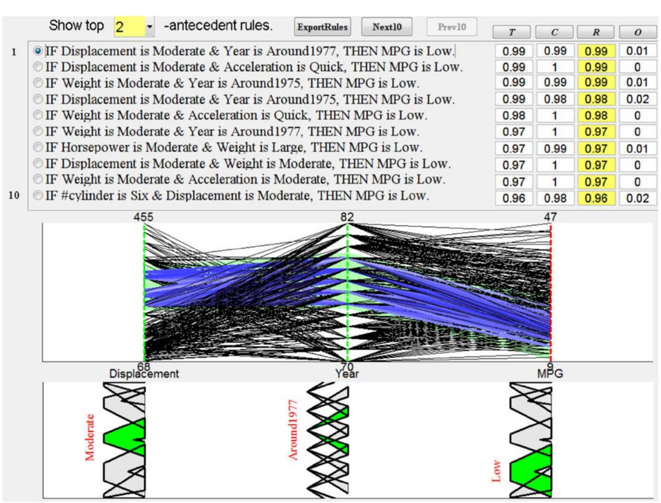

Fig. 8. Auto MPG dataset. Global top 1–10 rules according to R: the degree of reliability. The middle and bottom parts illustrate the first rule.

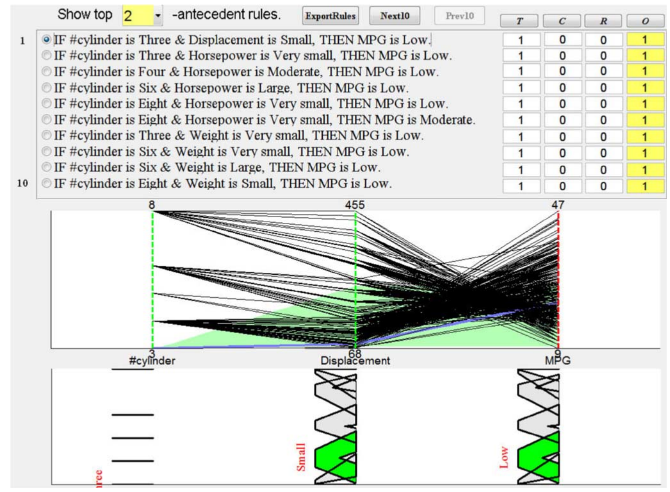

Fig. 9. Auto MPG dataset. Global top 1–10 rules according to O: the degree of outlier. The middle and bottom parts illustrate the first rule.

from Fig. 6 that *#cylinder* has range [3, 8]. Each case is represented in the middle of Fig. 6 as a piecewise linear curve. The blue curves represent those cases, which support the current rule under consideration at degrees larger than 0 (i.e., those cases satisfying *both* the antecedents and the consequent of the rule at degrees larger than 0), and the strength of supporting is proportional to the depth of the blue color. The red curves represent those cases violating the current rule (i.e., those cases satisfying *only* the antecedents of the rule), and the strength of violating is proportional to the depth of the red color. The black curves are cases irrelevant to the current rule (i.e., those cases *not* satisfying the antecedents of the rule). The light green region indicates the area covered by the current rule.

The bottom axes in Fig. 6 show the IT2 FSs used for each attribute. They were constructed by the authors for illustration purpose. The IT2 FSs that are used in the current rule are highlighted in green, and their names are also displayed.

Observe the following.

1) From Fig. 6, when T is used as the ranking criterion, a rule with high T may describe very few cases; therefore, it is very possible that this rule describes only outliers and, hence, cannot be trusted, e.g., the $14\mathrm{th}$ rule "IF #cylinder is Three and Displacement is Small, THEN MPG is Small" has $T = 1$, but from the middle part of Fig. 6, we see that only one case falls into the region described by it. Indeed, this rule seems counterintuitive. This suggests that T alone is not a reliable QM for LS.
2) From Fig. 7, when C is used as the ranking criterion, a rule with high C may have a low degree of truth, e.g., the $141\mathrm{st}$ rule "IF #cylinder is Four and Weight is Small, THEN MPG is Small" has $C = 1$, which means many cases support this rule, but from the bottom part of Fig. 7, we see that many cases violate it too (that is why its $T = 0.16$, which is a very small number). Indeed, this rule seems counterintuitive. Therefore, C alone is not a good QM either.

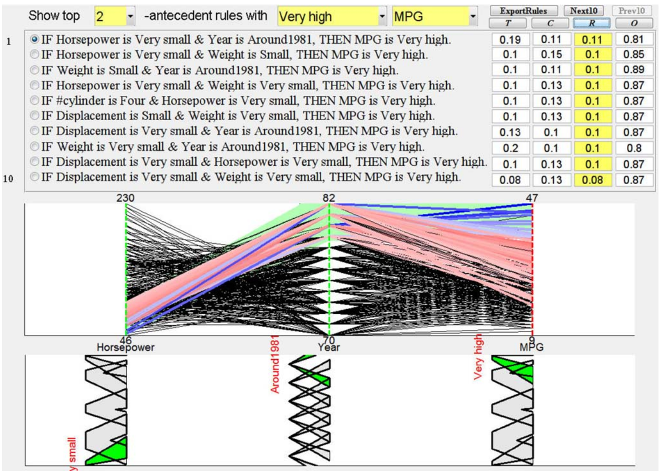

Fig. 10. Auto MPG dataset. Local top 1–10 rules according to R: the degree of reliability. The middle and bottom parts illustrate the first rule.

3) From Fig. 8, when R is used as the ranking criterion, a rule with high R has both high degree of truth and sufficient coverage (e.g., the first rule "IF Displacement is Moderate and Year is Around1977, THEN MPG is Small" has $R = 0.99$, and from the middle part of Fig. 8, we see most cases that fit its antecedents support the rule at different degrees), and hence, it describes a useful rule. Therefore, R is a comprehensive and reliable QM for LS.
4) From Fig. 9, when O is used as the ranking criterion, a rule with high O usually describes a very small number of cases (e.g., the first rule "IF #cylinder is Three and Displacement is Small, THEN MPG is Small" has $O = 1$, and from the middle part of Fig. 9, we see that only one case fits this rule), which should be considered as outliers. Therefore, O is useful in finding unexpected data and rules.

In summary, it appears that R and O proposed in this paper are better QMs for LS than T, which is dominant in previous LS literature: A high R identifies a useful rule with both high degree of truth and sufficient coverage, whereas a high O identifies outliers in the dataset that are worthy of further investigation.

The "local top rules" function is very similar to the 'global top rules" function, except that an attribute of the rules is specified by the user, e.g., a user may only want to know what combinations of attributes would lead to very large MPG. Fig. 10 shows the local top rules when R is used as the ranking criterion. Observe that the maximum R for two-antecedent rules, which lead to very large MPG, is 0.11 (a very small number), which means that it may be impossible to predict very large MPG using only two antecedents. Although no reliable rules can be found in this situation, LS also provides us with valuable information about the dataset.

B. Pima Indians Diabetes Dataset

The Pima Indians diabetes dataset was also obtained from the UCI machine-learning repository [3]. It contains 768 cases from females at least 21 years old of Pima Indian heritage. LS

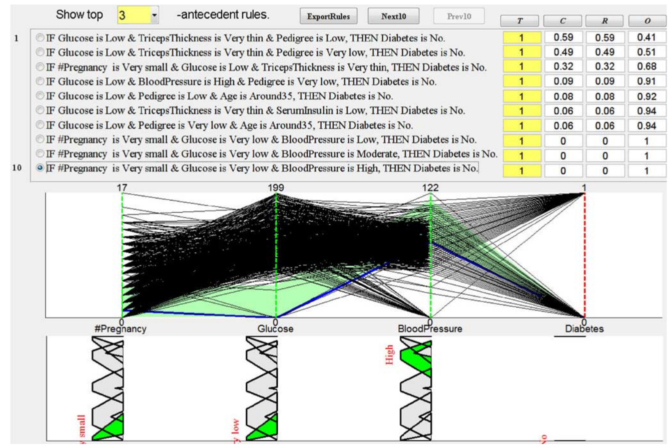

Fig. 11. Pima Indians diabetes dataset. Global top 1–10 rules according to T: the degree of truth. The middle and bottom parts illustrate the 10th rule.

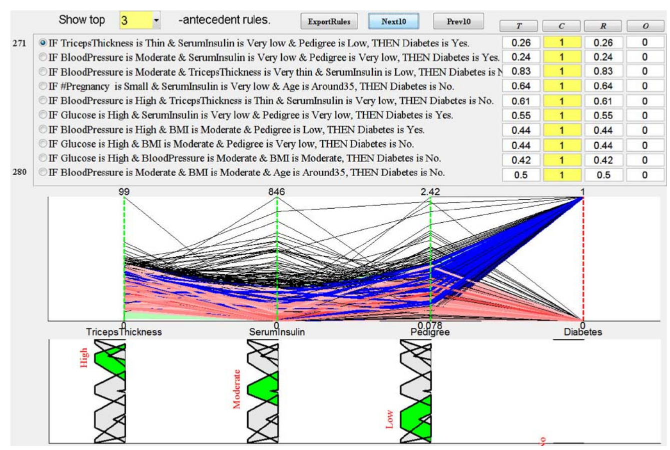

Fig. 12. Pima Indians diabetes dataset. Global top 271–280 rules according to C: the degree of sufficient coverage. The middle and bottom parts illustrate the 271st rule.

was used to find the relationship between the following inputs and *whether or not a person has diabetes*.

1) *#Pregnant*, which is the number of times pregnant;
2) *Glucose*, which is the plasma glucose concentration in an oral glucose tolerance test;
3) *BloodPression*, which is the diastolic blood pressure (in mm Hg);
4) *TricepsThickness*, which is the triceps skin fold thickness (in mm);
5) *SerumInsulin*, which is the 2-h serum insulin (in mu U/ml);
6) *BMI*, which is the body mass index;
7) *Pedigree*, which is the diabetes pedigree function;
8) *Age*, which is the age of the person.

Figs. 11–14 show global top rules when T, C, R, and O are used as the ranking criterion, respectively. The same conclusions about the roles of T, C, R, and O can be drawn here.

V. Discussions

In this section, the relationships between LS and the WM method [38], [63], perceptual reasoning [46], [64], [66], and

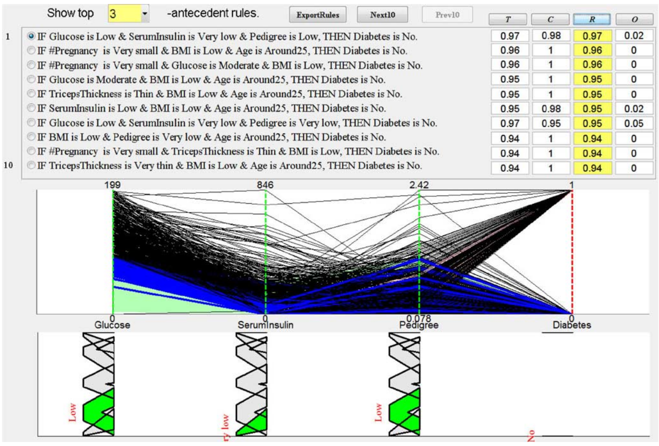

Fig. 13. Pima Indians diabetes dataset. Global top 1–10 rules according to R: the degree of reliability. The middle and bottom parts illustrate the first rule.

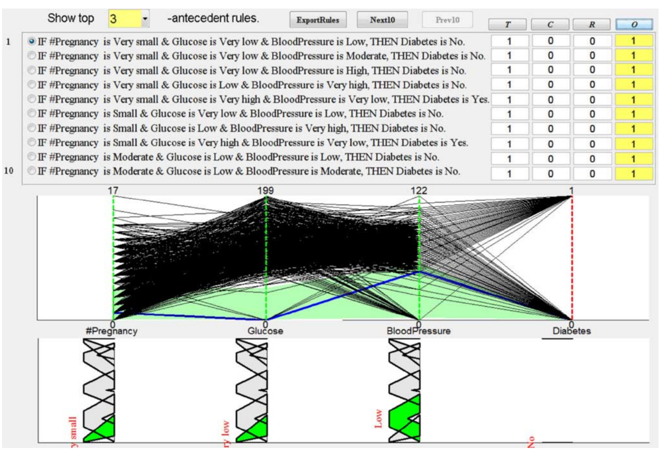

Fig. 14. Pima Indians diabetes dataset. Global top 1–10 rules according to O: the degree of outlier. The middle and bottom parts illustrate the first rule.

granular computing [19], [29], [79], [82], [85] are discussed. Because currently the WM method and granular computing mainly focus on T1 FSs, only T1 FSs are used in the discussion; however, our results can be extended to IT2 FSs without problems.

A. Linguistic Summarization and the Wang–Mendel Method

The WM method [38], [63] is a simple yet effective method to generate fuzzy rules from training examples. We use Fig. 15, where the 18 training data points are represented by squares,[7] 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. 15, each of the x and y domains is partitioned into three overlapping intervals by the FSs low, medium, and high. An interval in the x domain and an interval in the y domain 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. 15.
2) Because of overlapping MFs, it frequently happens that a datum is in more than one region, e.g., the diamond in Fig. 15 belongs to the region determined by high x and low y, as well as to the region determined by High x and Medium y. For each (x, y), one evaluates its degrees of belonging in regions, where it occurs, assigns it to the region with maximum degree, and generates a rule from it. For example, the degree of belonging of the diamond in Fig. 15 to the region determined by High x and Low y (the shaded region in the lower right corner) is $\mu_{\text{High}}(x)\mu_{\text{Low}}(y) = 1 \times 0.1 = 0.1$, and its degree of belonging to the region determined by High x and Medium y is $\mu_{\text{High}}(x)\mu_{\text{Medium}}(y) = 1 \times 0.8 = 0.8$; therefore, the diamond should be assigned to the region determined by High x and Medium y. Consequently, the corresponding rule generated from this diamond is as follows:

$$\text{IF } x \text{ is High, THEN } y \text{ is Medium} \qquad (37)$$

and it is also assigned a degree of 0.8. Similarly, a rule generated from the cross in Fig. 15 is as follows:

$$\text{IF } x \text{ is High, THEN } y \text{ is Low} \qquad (38)$$

and it has a degree of $\mu_{\text{High}}(x)\mu_{\text{Low}}(y) = 1 \times 1 = 1$.
3) To resolve conflicting rules, i.e., rules with the same antecedent MFs and different consequent MFs, one chooses the rule with the highest degree and discards all other rules, e.g., Rules (37) and (38) are conflicting, and Rule (38) is chosen because it has a higher degree.

Finally, the three rules generated by the WM method for the Fig. 15 data are as follows:

$$\text{IF } x \text{ is Low, THEN } y \text{ is High}$$

$$\text{IF } x \text{ is Medium, THEN } y \text{ is Medium}$$

$$\text{IF } x \text{ is High, THEN } y \text{ is Low.}$$

The first rule seems counter-intuitive, but it is a true output of the WM method. It is generated by the circle in Fig. 15 with a degree $\mu_{\text{Low}}(x)\mu_{\text{High}}(y) = 1 \times 1 = 1$, i.e., its degree is higher than two other possible rules, IF x is Low, THEN y is Low and

[7]Three points are represented by different shapes only for easy reference purpose.

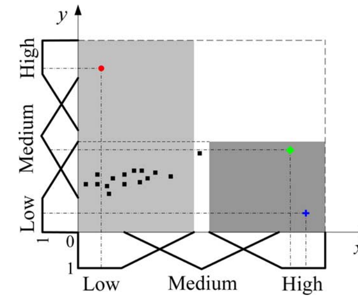

Fig. 15. Example to illustrate the difference between the WM method and LS. When x is Low, the WM method generates a rule "IF x is Low, THEN y is High," whereas LS generates a rule "IF x is Low, THEN y is Low."

IF x is Low, THEN y is Medium, although these two rules have more data to support them and, hence, look more reasonable. Note, however, that this example considers an extreme case. In practice, the WM method usually generates very reasonable rules, which is why it is popular.

Once the rules are generated, the degrees associated with them are discarded as they are no longer useful.

Example 1: Fig. 15 can also be used to illustrate the difference between the WM method and LS. Consider the shaded region, where x is Low. There are three candidates for a rule in this region:

$$\text{IF } x \text{ is Low, THEN } y \text{ is High} \tag{39}$$

$$\text{IF } x \text{ is Low, THEN } y \text{ is Medium} \tag{40}$$

$$\text{IF } x \text{ is Low, THEN } y \text{ is Low.} \tag{41}$$

For Rule (39)

$$c_{\mathbb{D}}(\text{Low}_x, \text{High}_y) = \sum_{m=1}^{18} \min(\mu_{\text{Low}_x}(x_m), \mu_{\text{High}_y}(y_m)) = 1 \tag{42}$$

$$c_{\mathbb{D}}(\text{Low}_x) = \sum_{m=1}^{18} \mu_{\text{Low}_x}(x_m) = 12.8 \tag{43}$$

$$T = \frac{c_{\mathbb{D}}(\text{Low}_x,\ \text{High}_y)}{c_{\mathbb{D}}(\text{Low}_x)} = 0.08. \tag{44}$$

Because the dataset consists of 18 points and there is only one datum that falls in the region determined by Low x and High y, the coverage ratio [see (9)] and degree of sufficient coverage [see (11)] are as follows:

$$r_c = \frac{1}{18} \tag{45}$$

$$C = f(r_c) = 0.15 \tag{46}$$

and hence, $R = \min(T, C) = 0.08$ and $O = \min(\max(T, 1 - T), 1 - C) = \min(\max(0.08, 0.92), 1 - 0.15) = 0.85$.

Similarly, for Rule (40), LS gives the following:

$$T = 0.31, \quad C = 1, \quad R = 0.31, \quad O = 0 \tag{47}$$

and for Rule (41), LS gives the following:

$$T = 0.71, \quad C = 1, \quad R = 0.71, \quad O = 0. \tag{48}$$

By ranking R and O, LS would select Rule (41) as the most useful rule with $R = 0.71$ and Rule (39) as an outlier with $O = 0.85$. These results are more reasonable than the rules generated by the WM method.

Repeating the aforementioned procedure for the other two regions, the following three rules are generated when R is used as the ranking criterion:

IF x is Low, THEN y is Low

$$T = 0.71, \quad C = 1, \quad R = 0.71, \quad O = 0.$$

IF x is Medium, THEN y is Medium

$$T = 0.82, \quad C = 1, \quad R = 0.82, \quad O = 0.$$

IF x is High, THEN y is Low

$$T = 0.57, \quad C = 0.82, \quad R = 0.57, \quad O = 0.18.$$

In summary, the differences between the WM method and LS are as follows.

1) The WM method tries to construct a predictive model,[8] whereas LS primarily constructs a descriptive model,[9] although the rules in this descriptive model may also be used for classification and prediction. According to [18], "*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 dataset. 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.*" As pointed out by Duch *et al.* [10], "*formulation of understandable rules derived from analysis of data is not the same as creating predictive models of data.*"
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 there are, whereas LS does not, e.g., if a region has very few data in it, then these data may be considered as outliers and no reliable rule is generated for this region.
3) The rules obtained from LS have several QMs associated with them; therefore, the rules can be sorted according to different criteria, whereas the rules obtained from the WM method are considered equally important.[10]

B. Linguistic Summarization and Perceptual Reasoning

Perceptual reasoning has been introduced by Mendel and Wu in [44], [46], [64], and [66]. It is different from most other approximate reasoning methods in that it requires the inference result to resemble the FS word models in the codebook, i.e., the inference result should be a normal FS[11] so that it can be mapped into a word in that codebook.

Perceptual reasoning considers the following problem: *Given a rulebase with K rules, each of the form:*

$$R^k\text{: IF } x_1 \text{ is } \tilde{F}_1^k \text{ and } \ldots \text{ and } x_p \text{ is } \tilde{F}_p^k, \text{ THEN } y \text{ is } \tilde{G}^k \quad k = 1, \ldots, K \tag{49}$$

[8]Predictive models 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.

[9]Descriptive models include clustering (finding natural groupings in the data), association models (discovering cooccurrence relationships among the data), feature extraction (creating new attributes as a combination of the original attributes), etc.

[10]There is an improved version of the WM method [62] that assigns a degree of truth to each rule; however, the degree of truth is computed differently from T in this paper, and the rule consequents are numbers instead of words modeled by FSs; therefore, it is not considered in this paper.

[11]A normal FS must have at least one point in its universe of discourse, whose membership grade is 1.

where $\tilde{F}_j^k$ and $\tilde{G}^k$ are words modeled by IT2 FSs, and a new input $\tilde{\mathbf{X}}' = (\tilde{X}_1, \ldots, \tilde{X}_p)$, where $\tilde{X}_j$ $(j = 1, \ldots, p)$ are also words modeled by IT2 FSs, then what is the output IT2 FS $\tilde{Y}_{PR}$?

Usually the scenario described by $\tilde{\mathbf{X}}'$ does not exist in the rulebase, which is why inference is needed. In similarity-based perceptual reasoning [46], [64], [66] one computes

$$\tilde{Y}_{\mathrm{PR}} = \frac{\sum_{k=1}^{K} f^k(\tilde{\mathbf{X}}')\tilde{G}^k}{\sum_{k=1}^{K} f^k(\tilde{\mathbf{X}}')} \tag{50}$$

where $f^k(\tilde{\mathbf{X}}')$ is the firing level of R^k, i.e.,

$$f^k(\tilde{\mathbf{X}}') = \prod_{j=1}^{p} s_J(\tilde{X}_j, \tilde{F}_j^k) \tag{51}$$

in which $s_J(\tilde{X}_j, \tilde{F}_j^k)$ is the Jaccard similarity for IT2 FSs [65] defined in

$$s_J(\tilde{X}_j, \tilde{F}_j^k) = \frac{\int_X \min(\overline{X}_j(x), \overline{F}_j^k(x))dx + \int_X \min(\underline{X}_j(x), \underline{F}_j^k(x))dx}{\int_X \max(\overline{X}_j(x), \overline{F}_j^k(x))dx + \int_X \max(\underline{X}_j(x), \underline{F}_j^k(x))dx}. \tag{52}$$

It has been mathematically proven [46], [64], [66] that $\tilde{Y}_{\mathrm{PR}}$ resembles the FOUs of the words in a codebook when the words are modeled using the interval approach. Another approach that uses firing intervals instead of firing levels is described in [44].

A rulebase is needed before perceptual reasoning 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, as mentioned in Section I, data are usually readily available in this information explosion age. However, note that rules extracted from LS construct a descriptive model instead of a predictive model; therefore, optimizations may be needed before these rules are used for classification and prediction. In addition, the rules have QMs associated with them, which have not been considered in perceptual reasoning. How to make use of the QMs is an open problem. One idea is to use them as weights of the rules, as in Ishibuchi and Yamamoto's approach [21], [22].

Additionally, the LS 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 to discover inconsistencies between the data and experience, e.g., 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; therefore, this inconsistency would suggest that it is necessary to update the operator's understanding about the process.

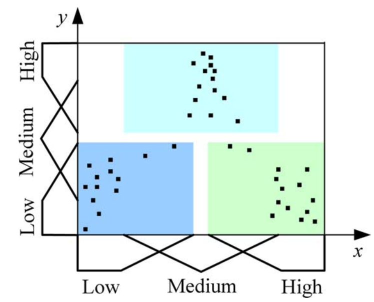

Fig. 16. Example to illustrate the idea of granular computing.

C. Linguistic Summarization and Granular Computing

Granular computing [19], [29], [79], [82], [85] 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. Although the name was first invented by Zadeh [85], according to Hirota and Pedrycz [19], "*the idea of information granulation has existed for a long time... For instance, an effect of temporal granulation occurs in 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.*"

LS can be viewed as a granular computing approach, as demonstrated by the following example.

Example 2: Consider the example shown in Fig. 16, 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*. LS 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 High

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.

VI. Conclusions

LS is a data mining or knowledge discovery approach to extract patterns from databases. Many authors have used this technique to generate summaries like "Most senior workers have high salary," which can be used to better understand and communicate about data; however, few of them have used it to generate IF–THEN rules like "IF X is large and Y is medium, THEN Z is small," which not only facilitate understanding and communication of data but can also be used in decision-making. In this paper, an LS approach to generate IF–THEN rules from causal databases has been proposed. Both T1 and IT2 FSs are considered. Five QMs for such summaries have been proposed:

1) the degree of truth, which quantifies the validity (confidence) of a rule;
2) the degree of sufficient coverage, which describes how many data support a rule and is related to the generality of the rule;
3) the degree of reliability, which finds rules with both high validity and sufficient coverage;
4) the degree of outlier, which describes the novelty of rules, i.e., the degree to which the summaries deviate from our expectations;
5) the degree of simplicity, which quantifies the syntactic complexity of the summaries.

Among them, the degree of reliability is especially useful to find the most reliable and representative rules, and the degree of outlier can be used to identify outlier rules and data for close-up investigation. These five QMs also correspond to the concepts of validity, generality, usefulness, novelty, and simplicity, which are five essential measures of a summary proposed by Hirota and Pedrycz [19].

Experiments on two datasets demonstrated our LS approach and a parallel coordinates rule visualization approach. The relationships between LS and the WM method, perceptual reasoning, and granular computing were also pointed out.

1) LS is similar to the WM method; however, LS is mainly used to discover patterns in data, whereas the WM method is used to construct a predictive model from the data.
2) The rules generated by LS can be used to initialize the rulebase in perceptual reasoning for decision-making.
3) LS can be viewed as a granular computing approach.

Our future work includes

1) to further study the applications of LS, e.g., how to use LS to rank the importance of inputs and, hence, to select the most important ones;
2) to design more efficient algorithms for LS. Currently, we use an exhaustive search method, where all possible combinations of rules are evaluated and then ranked according to a certain QM to find the top rules. This approach is feasible for small datasets, e.g., for the Pima Indians Diabetes Dataset in Section IV-B, which consists of 768 cases, eight inputs, five MFs for each input, and two MFs for the output, to compute T, C, R, and O together for all three-antecedent rules (the total number of rules is $\binom{8}{3} \times 5^3 \times 2 = 14\,000$) takes about 5 s on an IBM T43 notebook. However, the computational cost of this approach increases rapidly when the size of the database increases, and/or the number of antecedents increases, and/or the number of FSs associated with each attribute increases. More efficient algorithms are necessary to facilitate the applications of LS. One idea is to use some heuristics to eliminate some less promising rules from evaluation. Additionally, the algorithm should be incremental, i.e., the QMs should be updated incrementally as new data comes in.

References

[1] "Auto MPG data set." (2010, Nov. 16). [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Auto+MPG

[2] "Causality." (2010, Oct. 1). [Online]. Available: http://en.wikipedia.org/wiki/Causality.

[3] "Pima Indians diabetes data set." (2010, Nov. 16). [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Diabetes

[4] "Xmdv tool home page." (2010, Nov. 16). [Online]. Available: http://davis.wpi.edu/∼xmdv/

[5] L. Astudillo, O. Castillo, and L. T. Aguilar, "Intelligent control for a perturbed autonomous wheeled mobile robot: A type-2 fuzzy logic approach," *J. Nonlinear Stud.*, vol. 14, no. 3, pp. 37–48, 2007.

[6] P. Baguley, T. Page, V. Koliza, and P. Maropoulos, "Time to market prediction using type-2 fuzzy sets," *J. Manuf. Technol. Manag.*, vol. 17, no. 4, pp. 513–520, 2006.

[7] D. A. Chiang, L. R. Chow, and Y. F. Wang, "Mining time series data by a fuzzy linguistic summary system," *Fuzzy Sets Syst.*, vol. 112, pp. 419–432, 2000.

[8] S. Coupland, J. M. Mendel, and D. Wu, "Enhanced interval approach for encoding words into interval type-2 fuzzy sets and convergence of the word fous," in *Proc. IEEE World Congr. Comput. Intell.*, Jul. 2010, pp. 1–8.

[9] D. Dubois and H. Prade, "Gradual rules in approximate reasoning," *Inf. Sci.*, vol. 61, pp. 103–122, 1992.

[10] W. Duch, R. Setiono, and J. Zurada, "Computational intelligence methods for rule-based data understanding," *Proc. IEEE*, vol. 92, no. 5, pp. 771–805, May 2004.

[11] A. A. Freitas, "On rule interestingness measures," *Knowl.-Based Syst.*, vol. 12, no. 5–6, pp. 309–315, 1999.

[12] L. Geng and H. J. Hamilton, "Interestingness measures for data mining: A survey," *ACM Comput. Surv.*, vol. 38, no. 3, pp. 1–32, 2006.

[13] 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.

[14] S. Greco, B. Matarazzo, and R. Slowinski, "Rough membership and bayesian confirmation measures for parameterized rough sets," *Lecture Notes Comput. Sci.*, vol. 3641, pp. 314–324, 2005.

[15] L. Gu and Y. Q. Zhang, "Web shopping expert using new interval type-2 fuzzy reasoning," *Soft Comput.*, vol. 11, no. 8, pp. 741–751, 2007.

[16] H. Hagras, "A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 4, pp. 524–539, Aug. 2004.

[17] H. Hagras, "Type-2 FLCs: A new generation of fuzzy controllers," *IEEE Comput. Intell. Mag.*, vol. 2, no. 1, pp. 30–43, Feb. 2007.

[18] D. J. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*. Boston, MA: MIT Press, 2001.

[19] K. Hirota and W. Pedrycz, "Fuzzy computing for data mining," *Proc. IEEE*, vol. 87, no. 9, pp. 1575–1600, Sep. 1999.

[20] T. P. Hong, C. S. Kuo, and S. C. Chi, "Trade-off between computation time and number of rules for fuzzy mining from quantitative data," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 9, no. 5, pp. 587–604, 2001.

[21] H. Ishibuchi, T. Nakashima, and T. Murata, "Three-objective genetics-based machine learning for linguistic rule extraction," *Inf. Sci.*, vol. 136, no. 1–4, pp. 109–133, 2001.

[22] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 428–435, Aug. 2005.

[23] E. A. Jammeh, M. Fleury, C. Wagner, H. Hagras, and M. Ghanbari, "Interval type-2 fuzzy logic congestion control for video streaming across IP networks," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 5, pp. 1123–1142, Oct. 2009.

[24] J. Kacprzyk and P. Strykowski, "Linguistic summaries of sales data at a computer retailer via fuzzy logic and a genetic algorithm," in *Proc. Congr. Evol. Comput.*, vol. 2, Washington, DC, Jul. 1999, pp. 937–943.

[25] J. Kacprzyk, A. Wilbik, and S. Zadrozny, "Linguistic summarization of time series using a fuzzy quantifier driven aggregation," *Fuzzy Sets Syst.*, vol. 159, pp. 1485–1499, 2008.

[26] J. Kacprzyk and R. Yager, "Linguistic summaries of data using fuzzy logic," *Int. J. General Syst.*, vol. 30, pp. 133–154, 2001.

[27] J. Kacprzyk, R. Yager, and S. Zadrożny, "A fuzzy logic based approach to linguistic summaries of databases," *Int. J. Appl. Math. Comput. Sci.*, vol. 10, pp. 813–834, 2000.

[28] J. Kacprzyk and S. Zadrożny, "Linguistic database summaries and their protoforms: Towards natural language based knowledge discovery tools," *Inf. Sci.*, vol. 173, pp. 281–304, 2005.

[29] J. Kacprzyk, "Linguistic summaries of static and dynamic data: Computing with words and granularity," in *Proc. IEEE Conf. Granular Comput.*, Silicon Valley, CA, Nov. 2007, pp. 4–5.

[30] B. Kosko, "Fuzziness vs. probability," *Int. J. General Syst.*, vol. 17, pp. 211–240, 1990.

[31] C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller — Part II," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, no. 2, pp. 419–435, Mar./Apr. 1990.

[32] Q. Liang, N. N. Karnik, and J. M. Mendel, "Connection admission control in ATM networks using survey-based type-2 fuzzy logic systems," *IEEE Trans. Syst., Man, Cybern.*, vol. 30, no. 3, pp. 329–339, Aug. 2000.

[33] Q. Liang and J. M. Mendel, "MPEG VBR video traffic modeling and classification using fuzzy technique," *IEEE Trans. Fuzzy Systems*, vol. 9, no. 1, pp. 183–193, Feb. 2001.

[34] F. Liu and J. M. Mendel, "Encoding words into interval type-2 fuzzy sets using an interval approach," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1503–1521, Dec. 2008.

[35] I. Mani and M. Maybury, *Advances in Automatic Text Summarization*. Cambridge, MA: MIT Press, 1989.

[36] P. Melin, J. Urias, D. Solano, M. Soto, M. Lopez, and O. Castillo, "Voice recognition with neural networks, type-2 fuzzy logic and genetic algorithms," *J. Eng. Lett.*, vol. 13, no. 2, pp. 108–116, 2006.

[37] J. M. Mendel, "Computing with words, when words can mean different things to different people," in *Proc. 3rd Int. ICSC Symp. Fuzzy Logic Appl.*, Rochester, NY, Jun. 1999, pp. 158–164.

[38] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[39] J. M. Mendel, "Computing with words and its relationships with fuzzistics," *Inf. Sci.*, vol. 177, pp. 988–1006, 2007.

[40] J. M. Mendel, H. Hagras, and R. I. John, "Standard background material about interval type-2 fuzzy logic systems that can be used by all authors," (2010, Nov. 16). [Online]. Available: http://ieee-cis.org/_files/standards.t2.win.pdf

[41] J. M. Mendel and R. I. John, "Type-2 fuzzy sets made simple," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 117–127, Apr. 2002.

[42] J. M. Mendel, R. I. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 6, pp. 808–821, Dec. 2006.

[43] J. M. Mendel and D. Wu, "Perceptual reasoning: A new computing with words engine," in *Proc. IEEE Conf. Granular Comput.*, Silicon Valley, CA, Nov. 2007, pp. 446–451.

[44] J. M. Mendel and D. Wu, "Perceptual reasoning for perceptual computing," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1550–1564, Dec. 2008.

[45] J. M. Mendel and D. Wu, "Computing with words for hierarchical distributed decision making," in *Computational Intelligence in Complex Decision Systems*, D. Ruan, Ed. Paris, France: Atlantis, 2010.

[46] J. M. Mendel and D. Wu, *Perceptual Computing: Aiding People in Making Subjective Judgments*. Hoboken, NJ: Wiley-IEEE Press, 2010.

[47] A. Niewiadomski, *Methods for the Linguistic Summarization of Data: Applications of Fuzzy Sets and Their Extensions*. Portland, OR: Warszawa, 2008.

[48] J. M. Mendel and D. Wu, "A type-2 fuzzy approach to linguistic summarization of data," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 1, pp. 198–212, Feb. 2008.

[49] A. Niewiadomski and M. Bartyzel, "Elements of type-2 semantics in summarizing databases," *Lecture Notes Artif. Intell.*, vol. 4029, pp. 278–287, 2006.

[50] A. Niewiadomski and P. Szczepaniak, "News generating based on type-2 linguistic summaries of databases," in *Proc. Int. Conf. Inf. Process. Manag. Uncertainty Knowl.-Based Syst.*, Paris, France, Jul. 2006, pp. 1324–1331.

[51] A. Niewiadomski, "On two possible roles of type-2 fuzzy sets in linguistic summaries," *Lecture Notes Comput. Sci.*, vol. 3528, pp. 341–347, 2005.

[52] A. Niewiadomski, "Type-2 fuzzy summarization of data: An improved news generating," *Lecture Notes Comput. Sci.*, vol. 4585, pp. 241–250, 2007.

[53] G. Raschia and N. Mouaddib, "Using fuzzy labels as background knowledge for linguistic summarization of databases," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Melbourne, Australia, Dec. 2001, pp. 1372–1375.

[54] G. Raschia and N. Mouaddib, "SAINTETIQ: A fuzzy set-based approach to database summarization," *Fuzzy Sets Syst.*, vol. 129, pp. 137–162, 2002.

[55] D. Rasmussen and R. Yager, "Finding fuzzy and gradual functional dependencies with SummarySQL," *Fuzzy Sets Syst.*, vol. 106, pp. 131–142, 1999.

[56] R. Saint-Paul, G. Raschia, and N. Mouaddib, "Database summarization: The SaintEtiQ system," in *Proc. IEEE Int. Conf. Data Eng.*, Istanbul, Turkey, Apr. 2007, pp. 1475–1476.

[57] A. Silberschatz and A. Tuzhilin, "On subjective measures of interestingness in knowledge discovery," in *Proc. 1st Int. Conf. Knowl. Discov. Data Mining*, Montreal, QC, Canada, Aug. 1995, pp. 275–281.

[58] J. van den Berg, U. Kaymak, and W.-M. van den Bergh, "Fuzzy classification using probability-based rule weighting," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Honolulu, HI, May 2002, pp. 991–996.

[59] I. Vlachos and G. Sergiadis, "Subsethood, entropy, and cardinality for interval-valued fuzzy sets—An algebraic derivation," *Fuzzy Sets Syst.*, vol. 158, pp. 1384–1396, 2007.

[60] T. S. Wallsten and D. V. Budescu, "A review of human linguistic probability processing: General principles and empirical evidence," *Knowl. Eng. Rev.*, vol. 10, no. 1, pp. 43–62, 1995.

[61] C. H. Wang, C. S. Cheng, and T. T. Lee, "Dynamical optimal training for interval type-2 fuzzy neural network (T2FNN)," *IEEE Trans. Syst., Man, Cybern.*, vol. 34, no. 3, pp. 1462–1477, Jun. 2004.

[62] H. Wang and D. Qiu, "Computing with words via Turing machines: A formal approach," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 6, pp. 742–753, Dec. 2003.

[63] L.-X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 2, pp. 1414–1427, Nov./Dec. 1992.

[64] D. Wu, "Intelligent systems for decision support," Ph.D. dissertation, Univ. Southern California, Los Angeles, CA, May 2009.

[65] D. Wu and J. M. Mendel, "A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets," *Inf. Sci.*, vol. 179, no. 8, pp. 1169–1192, 2009.

[66] D. Wu and J. M. Mendel, "Perceptual reasoning for perceptual computing: A similarity-based approach," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 6, pp. 1397–1411, Dec. 2009.

[67] D. Wu and J. M. Mendel, "Computing with words for hierarchical decision making applied to evaluating a weapon system," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 3, pp. 441–460, Jun. 2010.

[68] D. Wu and J. M. Mendel, "Interval type-2 fuzzy set subsethood measures as a decoder for perceptual reasoning," Signal Image Process. Inst., Univ. Southern California, Los Angeles, CA, Tech. Rep. USC-SIPI Rep. 398, 2010.

[69] D. Wu and J. M. Mendel, "Social judgment advisor: An application of the perceptual computer," in *Proc. IEEE World Congr. Comput. Intell.*, Barcelona, Spain, Jul. 2010, pp. 10–8.

[70] D. Wu and W. W. Tan, "A type-2 fuzzy logic controller for the liquid-level process," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, vol. 2, Budapest, Hungary, Jul. 2004, pp. 953–958.

[71] D. Wu and W. W. Tan, "Type-2 FLS modeling capability analysis," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Reno, NV, May 2005, pp. 242–247.

[72] D. Wu and W. W. Tan, "Genetic learning and performance evaluation of type-2 fuzzy logic controllers," *Eng. Appl. Artif. Intell.*, vol. 19, no. 8, pp. 829–841, 2006.

[73] D. Wu and W. W. Tan, "A simplified type-2 fuzzy controller for real-time control," *ISA Trans.*, vol. 15, no. 4, pp. 503–516, 2006.

[74] H. Wu and J. M. Mendel, "Classification of battlefield ground vehicles using acoustic features and fuzzy logic rule-based classifiers," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 1, pp. 56–72, Feb. 2007.

[75] R. Yager, "A new approach to the summarization of data," *Inf. Sci.*, vol. 28, pp. 69–86, 1982.

[76] R. Yager, "On linguistic summaries of data," in *Knowledge Discovery in Databases*. G. Piatetsky-Shapiro and B. Frawley, Eds. Cambridge, MA: MIT Press, 1991, pp. 347–363.

[77] R. Yager, "Linguistic summaries as a tool for database discovery," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Yokohama, Japan, 1995, pp. 79–82.

[78] R. Yager, "Database discovery using fuzzy sets," *Int. J. Intell. Syst.*, vol. 11, pp. 691–712, 1996.

[79] Y. Y. Yao, "Granular computing for data mining," in *Proc. SPIE Conf. Data Mining, Intrusion Detection, Inf. Assur., Data Netw. Security*, Orlando, FL, Apr. 2006, pp. 1–12.

[80] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, pp. 338–353, 1965.

[81] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-1," *Inf. Sci.*, vol. 8, pp. 199–249, 1975.

[82] L. A. Zadeh, "Fuzzy sets and information granularity," in *Advances in Fuzzy Set Theory and Applications*. M. Gupta, R. Ragade, and R. Yager, Eds. Amsterdam, The Netherlands: North Holland, 1979, pp. 3–18.

[83] L. A. Zadeh, "A computational approach to fuzzy quantifiers in natural languages," *Comput. Math. Appl.*, vol. 9, pp. 149–184, 1983.

[84] L. A. Zadeh, "Fuzzy logic: Computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 103–111, May 1996.
[85] L. A. Zadeh, "Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems," *Soft Comput.*, vol. 2, pp. 23–25, 1998.
[86] J. Zeng and Z.-Q. Liu, "Type-2 fuzzy hidden Markov models and their applications to speech recognition," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 3, pp. 454–467, Jun. 2006.

Dongrui Wu (S'05–M'09) received the B.E. degree in automatic control from the University of Science and Technology of China, Hefei, China, in 2003, the M.E. degree in electrical engineering from the National University of Singapore, Singapore, in 2005, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 2009.

He is currently with the Industrial Artificial Intelligence Laboratory, General Electric Global Research, Niskayuna, NY. He has authored or coauthored more than 30 publications, including a book entitled *Perceptual Computing: Aiding People in Making Subjective Judgments* (Hoboken, NJ: Wiley-IEEE, 2010). He has reviewed more than 160 papers from 20 journals and various international conferences. His research interests include affective computing, computational intelligence, human–machine interaction, intelligent control, machine learning, and speech and physiological signal processing.

Dr. Wu was the recipient of the Best Student Paper Award from the IEEE International Conference on Fuzzy Systems, Reno, NV, in 2005.

Jerry M. Mendel (S'59–M'61–SM'72–F'78–LF'04) received the Ph.D. degree in electrical engineering from the Polytechnic Institute of Brooklyn, Brooklyn, NY.

Since 1974, he has been with the University of Southern California, Los Angeles, where he is currently a Professor of electrical engineering. He has authored or coauthored more than 470 technical papers. He is author and/or editor of eight books, including *Uncertain Rule-based Fuzzy Logic Systems: Introduction and New Directions* (Upper Saddle River, NJ: Prentice-Hall, 2001). His research interests include type-2 fuzzy logic systems and their applications to a wide range of problems, including smart oil field technology and computing with words.

Dr. Mendel is a Fellow of International Fuzzy Systems Association and a Distinguished Member of the IEEE Control Systems Society. He was the President of the IEEE Control Systems Society in 1986. He is currently the Chairman of the Fuzzy Systems Technical Committee and an elected member of the Administrative Committee of the IEEE Computational Intelligence Society. He was the recipient of the 1983 Best Transactions Paper Award of the IEEE Geoscience and Remote Sensing Society; the 1992 Signal Processing Society Paper Award; the 2002 Transactions on Fuzzy Systems Outstanding Paper Award; the 1984 IEEE Centennial Medal; an IEEE Third Millenium Medal; a Pioneer Award from the IEEE Granular Computing Conference, May 2006, for Outstanding Contributions in Type-2 Fuzzy Systems; and the 2008 Fuzzy Systems Pioneer Award from the IEEE Computational Intelligence Society.