A Novel Negative-Transfer-Resistant Fuzzy Clustering Model With a Shared Cross-Domain Transfer Latent Space and its Application to Brain CT Image Segmentation

Yizhang Jiang, Xiaqing Gu, Dongrui Wu, Wenlong Hang, Jing Xue, Shi Qiu, and Chin-Teng Lin

Abstract—Traditional clustering algorithms for medical image segmentation can only achieve satisfactory clustering performance under relatively ideal conditions, in which there is adequate data from the same distribution, and the data is rarely disturbed by noise or outliers. However, a sufficient amount of medical images with representative manual labels are often not available, because medical images are frequently acquired with different scanners (or different scan protocols) or polluted by various noises. Transfer learning improves learning in the target domain by leveraging knowledge from related domains. Given some target data, the performance of transfer learning is determined by the degree of relevance between the source and target domains. To achieve positive transfer and avoid negative transfer, a negative-transfer-resistant mechanism is proposed by computing the weight of transferred knowledge. Extracting a negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space (called NTR-FC-SCT) is proposed by integrating negative-transfer-resistant and maximum mean discrepancy (MMD) into the framework of fuzzy c-means clustering. Experimental results show that the proposed NTR-FC-SCT model outperformed several traditional non-transfer and related transfer clustering algorithms.

Index Terms—Medical image segmentation, fuzzy clustering, transfer learning, negative transfer

1 INTRODUCTION

With the development of electronic information and computer technology, medical imaging and image processing technology have developed rapidly. Medical imaging equipments collect images for a short time, and are less affected by external factors. Today, medical imaging technology has become a powerful tool and core technology for modern clinical diagnosis and treatment. Commonly used medical imaging techniques include Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Computed Radiography (CR), Ultrasound, and so on. CT scans produce clearer images than conventional x-ray for internal organs, bone and soft tissues. MRI scans furnish greater clearness and higher resolution than CT scans with lower resolution [1]. Medical image segmentation is a basic and important step in medical image processing and analysis. It is also the basis of medical image registration, medical image information fusion and 3D visualization. In the current clinical practice, manual segmentation based on visual recognition and empirical judgment by doctors is still the most typical and common segmentation method. However, manual segmentation is tedious, time consuming and subjective. For example, in the Isointense Infant Brain Segmentation Challenge (ISEG2017), manual segmentation of each brain MRI scan took an average of one week for neuroradiologists [2]. Furthermore, the differences in physician experience and uncertain factors such as visual fatigue will affect the correct analysis of segmentation results.

With the rapid growth of the image processing technology, many automatic image processing techniques have appeared in recent years [3], [4], [5]. Image segmentation methods can be broadly classified into four categories: graph based methods, classification methods, deep learning methods and clustering methods [6]. A medical image in graph based image segmentation is presented as a weighted undirected graph [7]. Each pixel or region in the image is treated...
as a vertex of a graph, and the set of edges can be connected by adjacent pixels or two adjacent regions. Then the image is divided into several parts according to the relationship of the adjacent pixels. The second category, also called supervised methods, use labeled segmented images to extract features and train a segmentation model, such as k-nearest neighbor (KNN) [8], neural network [9], support vector machine (SVM) [10], and so on. A major drawback of classification methods is that they require sufficient labeled training images. But in the areas of medical imaging, it is relatively easy and inexpensive to obtain a large amount of unlabeled data [11]. Deep learning learns the feature representation of tissue contour based on deep convolutional neural networks [12], [13]. Deep learning methods have successfully applied for medical image segmentation in recent years. However, deep learning methods usually need a large number of training dataset and special hardware devices. It is known that the medical image segmentation problem can be considered as classifying the pixels of images into homogeneous regions. This process can be viewed as clustering problem. Clustering method, as an unsupervised machine learning approach, is the process of grouping a set of data points into subsets so that data points in the same subset are similar (according to some criteria). Widely used clustering methods include expectation-maximization, spectral clustering and fuzzy clustering, and so on. In the last decade, clustering-based methods have attracted great interest in the segmentation of medical images [14]. Portela et al. [15] proposed clustering based semi-supervised classification for brain image segmentation, which used K-means clustering as initial processing to select brain slices. Ortiz et al. [16] improved brain image segmentation by using self-organizing maps (SOMs) based voxel clustering to extract features, and then used entropy-gradient clustering to segment brain images. Based on the idea of multiobjective optimization, Saha et al. [17] proposed semi-supervised clustering in the intensity space for medical image segmentation. Abdel-Maksouda et al. [18] combined K means and FCM clustering to propose a hybrid clustering method for tumour segmentation from brain image.

To make cluster-based segmentation methods perform better, the training medical image data needs to be representative of the target data. However, medical images are often collected with different scanners and scanning parameters, and medical images may have large differences in image quality due to machine performance or scanning technology, such as varying degrees of rotation, noise, etc. The requirement of training and target data under the same distribution prevents the use of clustering methods in larger research and clinical practice. Since the above scenarios exist in a large number of real-world environments, this leads to unsatisfactory segmentation results and the risk of algorithm failure.

To solve this problem, researchers have introduced the idea of transfer learning into clustering methods [19], [20]. With the help of some knowledge of auxiliary domain (called source domain), transfer learning handles the cases where the distribution, feature space or tasks are different between source domain and testing domain (called target domain). In transfer learning, the auxiliary knowledge from the source domain involves the data sample, feature representations, parameters and relationships [21], [22], [23]. The knowledge is usually obtained from certain precise procedures and reliable theory through some specific perspectives. Jiang et al. [24] proposed a transfer spectral clustering method, which used both data manifold and feature manifold between related clustering tasks. Deng et al. [25] proposed a transfer prototype-based fuzzy clustering method, which incorporated prototype knowledge induced from source domain to implement the clustering in the target domain. Qian et al. [20] proposed a cross-domain maximum entropy clustering method, which utilized the auxiliary knowledge from cluster centers and fuzzy memberships belonging to source data. However, these methods have a common assumption that source domain and target domain must have the same number of clusters. Moreover, most existing transfer clustering methods are not developed for noisy scenarios. Thus, these methods may not be suitable for noisy medical image segmentation.

Since the medical images of different domains may have variations due to changes caused by noise, field offset and bias field, in this paper, we study the problem of medical image segmentation in a noisy scenario by transferring medical images collected from related scenarios. We consider the new noisy scenario as the target domain and the existing medical image dataset from related scenario elsewhere as the source domain, and then use the learning on clean images of source data to improve the clustering in target data. To improve the transfer learning performance, we consider learning the negative-transfer-resistant mechanism, so that the influence of positive transfer knowledge is reinforced and the influence of negative transfer knowledge is reduced or even eliminated. Meanwhile, we think medical images in different scenarios may share certain common representations such as bone and soft tissue, and the shared representations could be preserved in a shared space. Inspired of maximum mean discrepancy (MMD) [26], we learn the shared latent space for source and target domains such that the distributions in different domains are close to each other. We investigate transferring ability of each cluster belonging to source domain in the shared latent space for medical image segmentation modeling. We use the clustering centers in the source domain as the transfer knowledge, regardless of whether the number of clusters in the source domain is the same as that in the target domain. With the above ideas, we propose a negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space (called NTR-FC-5CT). The motivation of NTR-FC-SCT is shown in Fig. 1. Two cluster centers presented as black triangle and circle have positive transfer

![Image](https://via.placeholder.com/150)
influence to the clustering in the target domain, while the cluster center presented as Black Square has negative influence to the clustering in the target domain. NTR-FC-SCT will automatically resist black square participating in the clustering in the target domain by using the negative-transfer-resistance mechanism. We evaluate the proposed model on real-world datasets, and compare with several non-transfer and related transfer clustering methods. The results on real-world brain CT dataset demonstrate that NTR-FC-SCT is more robust than the comparison methods.

The novelty of this study is as follows. 1) We formulate the problem of insufficient and noisy medical image segmentation as a model of transfer clustering task. To the best of our knowledge, our study is the first attempt to address this issue. 2) The negative-transfer-resistance mechanism is proposed to identify and resist negative source transfer knowledge. 3) The MMD is introduced into NTR-FC-SCT to unify the representation of image data of different domains in the shared transfer latent space, which helps transferring knowledge across different domains. 4) Clustering centers based transfer matching scheme is used to deal with the inconsistency problem of clustering numbers between source and target domains, so that more robust cluster performance can be promoted.

The rest of this paper is organized as follows. Concepts related to FCM, transfer learning and MMD are reviewed in Section 2. In Section 3, the negative-transfer-resistant mechanism and new proposed algorithm is introduced. Its parameter learning based on iteratively optimization strategy is then presented accordingly. The experimental results in real-world brain CT image datasets are reported in Section 4. Conclusions are given in the last section.

2 RELATED WORK

2.1 Conventional FCM for Image Segmentation

Fuzzy C-means (FCM) clustering [27] is one of most commonly used fuzzy clustering methods. FCM allows data points to belong to more than one cluster defined by a membership matrix. Let \( X = \{x_1, x_2, \ldots, x_N\} \) be a given dataset where \( d \) and \( N \) are data dimension and capacity, respectively. Suppose \( C \) clusters exist in \( X \), FCM derives the following objection function:

\[
\min_{U,V} J = \sum_{i=1}^{N} \sum_{j=1}^{C} \mu_{ij}^m \|x_i - v_j\|^2
\]

\[
s.t. 0 \leq \mu_{ij} \leq 1, \sum_{j=1}^{C} \mu_{ij} = 1,
\]

where \( U = [\mu_{ij}]_{N \times C} \) denotes the fuzzy membership matrix, and \( V = [v_1, v_2, \ldots, v_C]^T \) denotes the clustering center matrix. \( m (m > 1) \) denotes the fuzzy index.

FCM finds an optimal group of sets to explain the data samples into \( C \) clusters via matrixes \( Y \) and \( U \). FCM minimizes the total membership weighted distance of each sample \( x_i \) to the clustering center \( v_j \). FCM can easily optimize the objective functions by an iterative technique. Among center-based clustering methods, FCM was simple, efficient and high popularity. It is widely used in transfer clustering methods. For example, a FCM-based transfer learning was proposed in [28], which combined with Gini-Simpson diversity index and quadratic weights on membership. A knowledge-leveraged transfer FCM (KL-TCM) was proposed in [29], which used three-interlinked framework of knowledge extraction, knowledge matching, and knowledge utilization to leverage source information to help clustering in the target domain. However, the above FCM-based transfer learning clustering methods are completed in the original space, and while they do not consider resisting negative transfer.

2.2 Transfer Learning and Maximum Mean Discrepancy

Currently, when the training data is not enough to represent the data in the current domain, transfer learning, multi-task learning and co-clustering are three effective techniques that can enhance the clustering performance in the current domain. Multi-task learning performs multiple learning tasks together through by sharing certain knowledge among all tasks [30], [31]. Co-clustering performs clustering on both objects and features to exploit the clear duality between rows and columns of a contingency table [32]. Transfer learning clustering enhances the clustering performance in the target domain by leveraging useful knowledge from different but related domains. Many researches show that the transfer clustering methods have better learning ability to obtain an effective model with the idea of transfer learning [28], [29], [33]. In real applications, due to the existences of noise and field offset etc., the insufficient medical images are inadequate to complete image segmentation. Therefore, we think transfer learning clustering can be used as an effect technology to promote the segmentation of insufficient and noisy medical images in the new domain.

In transfer learning, a fundamental problem is to evaluate the distribution difference between source domain and target domain. Many criteria, like Kullback-Leibler (KL) divergence can be used for distribution estimation. But some criteria need density estimation, some are parametric and not suitable for high-dimensional data [34], [35]. In these cases, maximum mean discrepancy (MMD) as a nonparametric estimate criterion receives is widely used for comparing distributions. Let \( X_s = \{x_{1s}, x_{2s}, \ldots, x_{Ns}\} \) and \( X_t = \{x_{1t}, x_{2t}, \ldots, x_{Nt}\} \) denote the samples from distributions \( P_{source}(X_s) \) and \( P_{target}(X_t) \) belonging to source and target domains, respectively. The MMD for comparing distributions between \( P_{source}(X_s) \) and \( P_{target}(X_t) \) is defined as

\[
Dist(P_{source}(X_s), P_{target}(X_t)) = \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} f(x_{is}) - \frac{1}{N_t} \sum_{i=1}^{N_t} f(x_{it}) \right\|^2.
\]

MMD is based on reproducing kernel Hilbert space (RKHS). Suppose \( f : X \rightarrow H \), \( H \) is a universal RKHS. By inducing nonlinear mapping \( \phi \), function evaluation can be represented \( f(x) = \langle \phi(x), f \rangle \), then equation (2) can be rewritten as

\[
Dist(P_{source}(X_s), P_{target}(X_t)) = \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(x_{is}) - \frac{1}{N_t} \sum_{i=1}^{N_t} \phi(x_{it}) \right\|^2.
\]

When the difference between source and target domains is small, the relationship between two domains is strong and the transfer knowledge can be fully utilized. However, when the
data in the source domain is not sufficiently related, the clustering performance in the target domain may not only fail to promote, it may even actually decrease. Thus, transfer learning would resist negative transfer when source and target domains are not a good match. One strategy of resisting negative transfer is to identify and reject unhelpful knowledge from source domain. Some data selection and source selection methods have been proposed. The former implements some rules to select data samples to reconstruct the training set of source domain. For example, Rosenstein et al. [36] proposed a detecting negative transfer algorithm based on naive Bayes classification model. Croonenborghs et al. [37] proposed an option-based transfer in reinforcement learning algorithm to achieve a balance between positive and negative transfer. The source selection methods are applicable for multiple source scenarios, which select the best source domain (task) for transfer learning. An example of this strategy is Talvitie and Singh [38] proposed a Markov decision process to select the proper source task.

3 Negative-Transfer-Resistant Fuzzy Clustering Model With a Shared Cross-Domain Transfer Latent Space

The schematic diagram of NTR-FC-SCT is shown in Fig. 2. In the noisy transfer learning scenario, compared with data sample, feature representations, parameters and relationships are usually considered as being more insightful and more resistant to noise. In this study, we use the cluster centers in the source domain as auxiliary knowledge. The reason is that cluster centers are computed by certain reliable theories and rigorous procedures; such that the obtained cluster centers can well represent a cluster and all affiliated samples in a cluster.

3.1 Negative-Transfer-Resistant Mechanism

To resist the negative transfer, our objective is to discard bad cluster centers in the source domain and select helpful cluster centers that can help the clustering task satisfactorily. Let we have a total of $N_{SD}$ training images in the source domain $x_{i,s}(x_{i,s} \in \mathbb{R}^{d \times 1}, i = 1, 2, \ldots, N_{SD})$ and $N_{TD}$ training images in the target domain $x_{i,t}(x_{i,t} \in \mathbb{R}^{d \times 1}, i = 1, 2, \ldots, N_{TD})$, where $N_{TD} \ll N_{SD}$. We consider there exists a shared latent space, spanned by a projection matrix $\Theta \in \mathbb{R}^{r \times d}$, where $r$ is the dimensions of the shared latent space. In this way, the known cluster centers $\mathbf{v}_{h}^{SD}(h = 1, 2, \ldots, C_{SD})$ in the source domain obtained by a certain clustering method could be represented as $\Theta \mathbf{v}_{h}^{SD}$. The projection of a target domain sample $x_{i,t}$ could be represented as $\Theta x_{i,t}$. Suppose $\tilde{\mathbf{v}}_{j}^{TD}$ ($j = 1, 2, \ldots, C_{TD}$) is the unsolved cluster centers in the target domain in the shared latent space. We consider the following optimization term for resisting negative transfer:

$$
\min_{\tilde{\mathbf{v}}} J(\tilde{\mathbf{v}}) = \sum_{j=1}^{C_{TD}} \sum_{h=1}^{C_{SD}} \| \tilde{\mathbf{v}}_{j}^{TD} - S_{jh} \Theta \mathbf{v}_{h}^{SD} \|^{2},
$$

where the parameter $S_{jh}$ is called the weight of transfer knowledge, and its value is in the range $[0, 1]$. $S_{jh}$ denotes the matching degree between the $j$th cluster center of the target domain and the $h$th cluster center of the source domain. To find useful transfer knowledge from the cluster centers in the source domain, it is needed to devise a strategy to set the values of $S_{jh}$ to high values for positive transfer and low values for negative transfer. In Eq. (4), we set $S_{jh}$ as follows:

$$
S_{jh} = 1 - \frac{\sum_{h'=1}^{C_{SD}} \| \tilde{\mathbf{v}}_{j}^{TD} - \Theta \mathbf{v}_{h}^{TD} \|^{2}}{\| \tilde{\mathbf{v}}_{j}^{TD} - \Theta \mathbf{v}_{h}^{TD} \|^{2}}.
$$

When $S_{jh}$ tends to 1, $\tilde{\mathbf{v}}_{j}^{TD}$ exactly matches $\Theta \mathbf{v}_{h}^{SD}$. In this case, the clustering results will be coherent and these two centers clusters of different domains are much closer to each other. When $S_{jh}$ tends to 0, $\tilde{\mathbf{v}}_{j}^{TD}$ does not match $\Theta \mathbf{v}_{h}^{SD}$. In this case, we make $\mathbf{v}_{h}^{SD}$ participate in transfer learning as little as possible. That is to say, the transfer performance will be at least no worse than performing the target clustering without transfer. In other words, if the source clustering transfer ability of $\Theta \mathbf{v}_{h}^{SD}$ decreases cluster performance of target domain data, it means the partial source clustering may be not related or the relationship is not sufficiently leveraged, then the negative transfer has occurred. By adjusting the value of $\Theta \mathbf{v}_{h}^{SD}$, Eq. (4) can help the transfer model make positive transfer when two domains are appropriately matched and resist negative transfer when two domains are not matched. At one extreme, $S_{jh}$ is set to be 1, the transferred knowledge from the source domain are completely helpful, such that the cluster centers in the source and target domains are coincided with each other, and all transferred knowledge are completely adopted. At the other extreme, $S_{jh}$ tends to 0, it means the transferred knowledge from the source domain are unhelpful. The clustering on the target domain will disregard the transferred knowledge. But in most cases part transferred knowledge are selectively keep and the other parts are disregard.

3.2 The Proposed NTR-FC-SCT

To find a proper projection matrix $\Theta$, we think the difference between source and target domains in the shared latent space should be as small as possible, such that the relationship between domains is strengthened, and the transfer knowledge of data in the source domain will be more helpful to complete medical segmentation in the target domain. Based on the
The solution of objective function in Eq. (9) relates to the definition of MMD, the difference between two domains in the shared latent space. Let

\begin{equation}
\Omega = \frac{1}{(N^T)^2} \sum_{i=1}^{N^T} \sum_{j=1}^{N^T} x_i x_j^T + \frac{1}{(N^S)^2} \sum_{i=1}^{N^S} \sum_{j=1}^{N^S} x_i x_j^T - \frac{2}{N^TN^S} \sum_{i=1}^{N^T} \sum_{j=1}^{N^S} x_i x_j^T.
\end{equation}

The optimization of \(d(P_{\text{source}}, P_{\text{target}})\) can be simplified as

\begin{equation}
\begin{aligned}
\min_{\Theta} d(P_{\text{source}}, P_{\text{target}}) &= \min_{\Theta} \Theta \Omega \Theta^T, \\
s.t. \quad &\Theta \Theta^T = I_{r \times r}.
\end{aligned}
\end{equation}

where \(I_{r \times r}\) is a \(r \times r\) identity matrix, such that the projection matrix \(H\) is orthogonal.

Coming to the transfer learning tasks, we incorporate Eqs. (4) and (8) into the FCM framework. We obtain the objective function of NTR-FC-SCT as follows:

\begin{equation}
J = \sum_{i=1}^{N^T} \sum_{j=1}^{C^{TD}} (\mu_{ij}^{TD})^m \left( \|\Theta x_{i,t} - \tilde{v}_j^{TD}\|^2 \right) + \lambda_1 \sum_{j=1}^{C^{TD}} \sum_{h=1}^{C^{SD}} \left( \tilde{v}_j^{TD} - S_{jh} \Theta v_h^{SD} \right)^2 + \lambda_2 \Theta \Theta^T, \\
s.t. \quad \Theta \Theta^T = I_{r \times r}, \sum_{j=1}^{C^{TD}} \mu_{ij}^{TD} = 1,
\end{equation}

where parameters \(\lambda_1 > 0\) and \(\lambda_2 > 0\) are the coefficients of transfer optimization term and MMD term, respectively. In NTR-FC-SCT, the parameter \(\lambda_1\) is used to control the influence of transfer optimization term \(\sum_{j=1}^{C^{TD}} \sum_{h=1}^{C^{SD}} \|\tilde{v}_j^{TD} - S_{jh} \Theta v_h^{SD}\|^2\) to the entire objective function. The greater the \(\lambda_1\), the greater the contribution of the transfer term will be. In this case, the unsolved cluster centers \(\tilde{v}_j^{TD}\) in the target domain should be close to \(S_{jh} \Theta v_h^{SD}\) in the shared latent space. Conversely, when \(\lambda_1\) tends to 0, the contribution of the transfer term is weakened, the difference between the unsolved cluster centers and known cluster centers in two different domains can be relaxed.

### 3.3 Optimization of NTR-FC-SCT

The solution of objective function in Eq. (9) relates to the matrixes \(\Theta, U\) and \(\tilde{V}\). In the following, we optimize them one by one using the iteratively optimization strategy. In terms of the Lagrange optimization, the minimization of \(J\) in Eq. (9) by introducing the Lagrangian multiplier \(\alpha\) in Eq. (9) can be converted to the following unconstrained minimization problem:

\begin{equation}
L = \sum_{i=1}^{N^T} \sum_{j=1}^{C^{TD}} \left(\mu_{ij}^{TD}\right)^m \left( \|\Theta x_{i,t} - \tilde{v}_j^{TD}\|^2 \right) + \lambda_1 \sum_{j=1}^{C^{TD}} \sum_{h=1}^{C^{SD}} \left( \tilde{v}_j^{TD} - S_{jh} \Theta v_h^{SD} \right)^2 + \lambda_2 \Theta \Theta^T + \sum_{i=1}^{N^T} \alpha_i \left(1 - \sum_{j=1}^{C^{TD}} \mu_{ij}^{TD}\right).
\end{equation}

In the first step, we fix parameters \(\Theta\) and \(U\), and only consider \(\tilde{V}\). To minimize this objective on parameter \(\tilde{V}\), we set the derivative with regard to \(\tilde{V}\) to zero:

\begin{equation}
\frac{\partial L}{\partial \tilde{v}_j^{TD}} = - \sum_{i=1}^{N^T} \left(\mu_{ij}^{TD}\right)^m \left(\Theta x_{i,t} - \tilde{v}_j^{TD}\right) + \lambda_1 \sum_{h=1}^{C^{SD}} \left( \tilde{v}_j^{TD} - S_{jh} \Theta v_h^{SD} \right) = 0
\end{equation}

\begin{equation}
\Leftrightarrow \tilde{v}_j^{TD} = \frac{\sum_{i=1}^{N^T} \left(\mu_{ij}^{TD}\right)^m \Theta x_{i,t} + \sum_{h=1}^{C^{SD}} \lambda_1 S_{jh} \Theta v_h^{SD}}{\sum_{i=1}^{N^T} \left(\mu_{ij}^{TD}\right)^m + \lambda_1 C^{SD}}.
\end{equation}

We can get \(\tilde{V}\) in a closed form as follows

\begin{equation}
\tilde{v}_j^{TD} = \frac{\sum_{i=1}^{N^T} \left(\mu_{ij}^{TD}\right)^m \Theta x_{i,t} + \sum_{h=1}^{C^{SD}} \lambda_1 S_{jh} \Theta v_h^{SD}}{\sum_{i=1}^{N^T} \left(\mu_{ij}^{TD}\right)^m + \lambda_1 C^{SD}}.
\end{equation}

Likewise, in the next step, we fix parameters \(\Theta\) and \(\tilde{V}\), and only consider \(U\). The minimization problem of Eq. (10) with respect to \(U\) can be equivalent to the following problem,

\begin{equation}
\frac{\partial L_{\mu_{ij}^{TD}}}{\partial \mu_{ij}^{TD}} = m(\mu_{ij}^{TD})^{m-1} \left(\|\Theta x_{i,t} - \tilde{v}_j^{TD}\|^2 - \alpha_i\right) = 0
\end{equation}

\begin{equation}
\Leftrightarrow \mu_{ij}^{TD} = \left(\frac{\alpha_i}{m}\right)^{1\over m-1} \left(\|\Theta x_{i,t} - \tilde{v}_j^{TD}\|^2\right)^{1\over m-1}.
\end{equation}

In light of \(\sum_{j=1}^{C^{TD}} \mu_{ij}^{TD} = 1\), we can obtain

\begin{equation}
\mu_{ij}^{TD} = \left(\frac{1}{\|\Theta x_{i,t} - \tilde{v}_j^{TD}\|^2}\right)^{1\over m-1} \left(\sum_{k=1}^{C^{TD}} \left(\frac{1}{\|\Theta x_{i,t} - \tilde{v}_k^{TD}\|^2}\right)^{1\over m-1}\right)^{-1\over m-1}.
\end{equation}

In the next step, we update matrix \(\Theta\) and fix parameters \(\tilde{V}\) and \(U\). Let

\begin{equation}
\tilde{U} = [\tilde{u}_{11}, \ldots, \tilde{u}_{1i}, \ldots, \tilde{u}_{NN_T}] \in \mathbb{R}^{1 \times N^T}.
\end{equation}
\[ \mathbf{\hat{U}} = [\hat{U}_1, \ldots, \hat{U}_{C_{TD}}] \in \mathbb{R}^{1 \times C_{TD \times TD}}, \mathbf{U} = \text{diag}(\mathbf{U}) \in \mathbb{R}^{C_{TD \times TD}}. \]

Let
\[ \omega = \left( \mathbf{E}, \ldots, \mathbf{E} \right) \in \mathbb{R}^{C_{TD \times TD}}, \]

where \( \mathbf{E} \in \mathbb{R}^{C_{TD \times TD}} \). Let \( \mathbf{Q} = [\mathbf{q}_i, \ldots, \mathbf{q}_{C_{TD}}] \in \mathbb{R}^{C_{TD \times TD}}, \) where \( \mathbf{q}_i = [\mathbf{v}_i^{TD}, \ldots, \mathbf{v}_i^{TD}] \in \mathbb{R}^{C_{TD}}. \) Substituting Eqs. (16), (17) and (18) back to Eq. (9), the minimization problem of Eq. (10) with respect to \( \Theta \) can be equivalent to the following problem,
\[ \ell(\Theta) = tr\left( (\Theta \mathbf{x}_\omega - \mathbf{Q}) (\Theta \mathbf{x}_\omega - \mathbf{Q})^T \right) + \lambda_1 tr\left( (\mathbf{\Theta} \mathbf{V}^{SD})^T (\mathbf{\Theta} \mathbf{V}^{SD}) - (\mathbf{\hat{V}}^{TD})^T (\mathbf{\hat{V}}^{TD})^T \right) + \lambda_2 tr(\Theta \Theta^T), \]

where
\[ \Omega = \frac{1}{(N_{TD})} \mathbf{x}_i^{T} \mathbf{x}_i^{TD} - \frac{1}{(N_{SD})} \mathbf{x}_i^{T} \mathbf{x}_i^{SD}. \]

Likewise, the minimization problem of Eq. (10) with respect to \( \Theta \) can be equivalent to the following problem,
\[ \frac{\partial \ell}{\partial \Theta} = (\Theta \mathbf{x}_\omega - \mathbf{U}(\omega)) (\hat{\mathbf{U}}(\omega))^T (\mathbf{x}_i)^T - \mathbf{Q} (\hat{\mathbf{U}}(\omega))^T (\hat{\mathbf{x}}_i)^T \]
\[ + \lambda_1 \left( \mathbf{\Theta} \mathbf{V}^{SD} \mathbf{S} \left( \mathbf{\hat{V}}^{SD} \right)^T - \mathbf{\hat{V}}^{TD} \mathbf{S} \left( \mathbf{\hat{V}}^{TD} \right)^T \right) + \lambda_2 \Theta. \]

In this study, the widely gradient descent method is adopted to compute the optimal \( \Theta \). By setting the initial value \( \Theta = \Theta^0 \), the gradient descent method successively optimal \( \Theta \) as follows
\[ \Theta^t = \Theta^{t-1} - \eta \frac{\partial \ell}{\partial \Theta} |_{\Theta=\Theta^{t-1}}, \]

where \( \eta \) is the learning rate and \( t \) is the iteration number. Considering the constraint of \( \Theta \Theta^T = 1 \). After each updating step of \( \Theta^t \), let \( \Theta^t = \Theta^t \mathbf{R} \) be the QR decomposition of \( \Theta^t \), where \( \Theta^t \) has orthogonal columns and \( \mathbf{R} \) is an upper triangle. Then we replace \( \Theta^t \) with \( \Theta^t \) for the next iteration. Eq. (19) will be iteratively solved until the convergence condition is satisfied.

Based on the above analysis, the proposed NTR-FC-SCT model is presented in Algorithm 1.

### 4 Experiments

#### 4.1 Data Sets and Settings

We use ultrashort echo time (UTE) and modified Dixon brain image datasets [39], [40]. It consists of 256 brain CT image slices of 10 patients, with each image of a resolution of 256 x 256 pixels. All CT images with corresponding manual segmentation are segmented into three classes: bone, water and soft issues. These class labels are assigned by physicians or technicians. We randomly select 20 brain CT images as the original target domain data, and the rest 236 brain CT images as source domain data. We consider the application of NTR-FC-SCT in the scenario of target images polluted by noise. To this aim, all target images were corrupted by 5, 10, 15, 20, 25 and 30 percent Gaussian noise. The example images in the source and target domains are shown in Fig. 3. Following the training protocol established in [41], we construct a total training data set combining 236 source brain images and random 8 target brain images, while the remaining 12 target brain images are used as testing brain images. We repeat the experiment for 10 runs and record the experimental results.

#### Algorithm 1. NTR-FC-SCT Model

- Initialize Set the maximum number of iterations \( t_{max} \), the fuzzy index \( m \), the regularization parameters \( \lambda_1 \) and \( \lambda_2 \), and the learning rate \( \eta \).

- Repeat:
  - Extracting transfer knowledge form the source domain;
  - Perform soft-partition clustering methods in the source domain, such as FCM, and obtain the cluster centers of data in the source domain;
  - t = t + 1;

- Initialize the clustering centers of data in the target domain;

- Compute the weight of transfer knowledge \( S_{\beta} \) using Eq. (10);

  - Fix \( U(t) \) and \( \Theta(t) \), obtain \( \hat{V}^{TD}(t) \) using Eq. (12);

  - Fix \( \hat{V}^{TD}(t) \) and \( \Theta(t) \), obtain \( U(t) \) using Eq. (14);

  - Fix \( U(t) \) and \( \hat{V}^{TD}(t) \), obtain \( \Theta(t) \) using Eqs. (18) and (19);

  - Compute \( J(t) \) using Eq. (9);

  - Until ||J(t) - J(t - 1)|| ≤ \delta or t ≥ t_{max};

To compare the segmentation performance of NTR-FC-SCT with that of existing methods, complete image segmentation is obtained and compared with segmentations obtained by FCM [27], transfer spectral clustering (TSC) [24], and type-I knowledge-transfer-oriented c-means (T1-KT-FCM) [28]. As introduced in Section 2, FCM is the baseline algorithm of NTR-FC-SCT. TSC performs transfer learning based on bipartite graph co-clustering, which adopts both the data manifold and sample manifold shared among different domains. T1-KT-FCM makes the cluster centers in the source domain as the transfer information to control the knowledge transfer in the test images. T1-KT-FCM incorporates this idea into FCM to
achieve automatic image segmentation. To obtain the optimal parameters in all four methods, the common used grid search is conducted. Fuzzy index $m$ in all fuzzy clusters is set within the grid $[1, 1.5, 2, 2.5]$. The $K$-nearest parameters in TSC are set within the grid $\{0, 0.005, 0.1, 0.5, 0.7, 1, 1.5, 10, 50, 100\}$. The $\lambda, \gamma$ parameters in T1-KT-FCM are set within the grid $[10^{-4}, 10^{-2}, \ldots, 10]$. The parameters $\lambda_1, \lambda_2$ in NTR-FC-SCT are set within the grid $[10^{-4}, 10^{-2}, \ldots, 10]$, learning rate $\eta$ in NTR-FC-SCT are set within the grid $[10^{-4}, 10^{-2}, \ldots, 10]$, and the maximum number of iterations is 10E5.

In this study, the performance of image segmentation by clustering methods is evaluated in terms of two validity indicators: normalized mutual information (NMI) [42] and adjusted rand index (ARI) [43]. NMI and ARI can efficiently evaluate the agreement degree between the known clusters and the estimated data structure. Both NMI and ARI take values from 0 to 1, and larger value means better cluster performance. Experimental environment is Intel Core i3-4170 3.7 GHz CPU and 12 GM RAM, Windows 10, and MATLAB R2016a in this study.

### 4.2 Performance Comparison

The clustering performance of four methods is reported in the following. The mean and standard deviation of NMI and ARI for all compared clustering methods are displayed in Tables 1, 2, 3, 4, 5, 6, respectively. The experimental results show that three transfer learning methods are superior to FCM. The introduction of transfer knowledge from source domain has indeed promoted the cluster performance of data in the target domain. FCM is not a transfer learning cluster method which simply combines the source domain and target domain data as the training data. Due to underlying noise or outliers in the target domain, the distribution difference between source and target domains are significant different. Thus, FCM cannot obtain good clustering performance in terms of NMI and ARI. Our model achieves the best performance in all datasets. TSC may be not suitable for the transfer scenario in noisy medical image segmentation, since the character of medical image are usually different in noisy scenario, while the manifold and sample manifold shared among different domains can not resist negative transfer for TSC. T1-KT-FCM exploits the transfer knowledge across domains in the original data space; however, the limited transfer knowledge can not be fully exploited in such original space. NTR-FC-SCT has shown better performance than the other comparison methods in terms of NMI and ARI. Both the reliable knowledge obtained in the source domain and the ability of resisting negative transfer has the important influence on the segmentation performance of NTR-FC-SCT. To better observe the behavior of all four compared clustering methods are displayed in Tables 1, 2, 3, 4, 5, 6.
algorithms, Figs. 4, 5, 6, 7, 8, 9 graphically shows the segmentation results of all comparison methods obtained on subject1 with different noise. Similar to the results in the Tables 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, NTR-FC-SCT obtains the best segmentation results for distinguishing the bone, water and soft issues. The boundaries between different organizations are smooth, and obvious are relatively clearer than the other three methods.

4.3 Flexibility Evaluation of NTR-FC-SCT

To validate the effect of two regularization terms on the performance of NTR-FC-SCT, we present two comparison methods NTR-FC-SCT \((\lambda_1 = 0)\) and NTR-FC-SCT \((\lambda_2 = 0)\), obtained with the parameter \(\lambda_1 = 0\) and \(\lambda_2 = 0\) in NTR-FC-SCT, respectively. We compare them with FCM and NTR-FC-SCT on Subjects 1-8 and show their mean and standard deviation of NMI and ARI in Tables 13, 14, respectively. The experimental results show that the performances of both NTR-FC-SCT \((\lambda_1 = 0)\) and NTR-FC-SCT \((\lambda_2 = 0)\) are better than baseline method FCM. The regularization term in NTR-FC-SCT \((\lambda_2 = 0)\) is \(\sum_{j=1}^{C^{TD}} \sum_{h=1}^{C^{SD}} \| \bar{V}_j^{TD} - S_h \theta_k \|_2^2 \), which can effectively resist negative transform by using the transfer optimization strategy and improve the segmentation performance in

<table>
<thead>
<tr>
<th>Dataset</th>
<th>FCM Means Std</th>
<th>TSC Means Std</th>
<th>T1-KT-FCM Means Std</th>
<th>NTR-FC-SCT Means Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject1</td>
<td>0.0107 0.0058</td>
<td>0.5078 0.2154</td>
<td>0.4969 0.0043</td>
<td>0.5469 0.0104</td>
</tr>
<tr>
<td>Subject2</td>
<td>0.1708 0.1001</td>
<td>0.5792 0.1741</td>
<td>0.5842 0.0076</td>
<td>0.6572 0.0182</td>
</tr>
<tr>
<td>Subject3</td>
<td>0.2460 0.0981</td>
<td>0.5803 0.1685</td>
<td>0.5232 0.0145</td>
<td>0.5940 0.0093</td>
</tr>
<tr>
<td>Subject4</td>
<td>0.0121 0.0081</td>
<td>0.5177 0.2390</td>
<td>0.5045 0.0112</td>
<td>0.5717 0.0197</td>
</tr>
<tr>
<td>Subject5</td>
<td>0.1503 0.1169</td>
<td>0.5621 0.1599</td>
<td>0.5580 0.0122</td>
<td>0.6090 0.0083</td>
</tr>
<tr>
<td>Subject6</td>
<td>0.2199 0.0660</td>
<td>0.5662 0.1533</td>
<td>0.5434 0.0311</td>
<td>0.5935 0.0142</td>
</tr>
<tr>
<td>Subject7</td>
<td>0.0284 0.0285</td>
<td>0.5331 0.2528</td>
<td>0.5156 0.0137</td>
<td>0.5677 0.0095</td>
</tr>
<tr>
<td>Subject8</td>
<td>0.1320 0.1375</td>
<td>0.5705 0.1577</td>
<td>0.5547 0.0156</td>
<td>0.6272 0.0136</td>
</tr>
<tr>
<td>Subject9</td>
<td>0.2594 0.1054</td>
<td>0.5859 0.1542</td>
<td>0.5682 0.0124</td>
<td>0.6437 0.0127</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>FCM Means Std</th>
<th>TSC Means Std</th>
<th>T1-KT-FCM Means Std</th>
<th>NTR-FC-SCT Means Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject1</td>
<td>0.0120 0.0093</td>
<td>0.5086 0.2028</td>
<td>0.5569 0.0124</td>
<td>0.6567 0.0084</td>
</tr>
<tr>
<td>Subject2</td>
<td>0.2553 0.1770</td>
<td>0.5833 0.1497</td>
<td>0.6333 0.0113</td>
<td>0.7439 0.0109</td>
</tr>
<tr>
<td>Subject3</td>
<td>0.2010 0.0804</td>
<td>0.5526 0.1537</td>
<td>0.5872 0.0206</td>
<td>0.6602 0.0110</td>
</tr>
<tr>
<td>Subject4</td>
<td>0.0891 0.1752</td>
<td>0.5143 0.2144</td>
<td>0.5664 0.0050</td>
<td>0.6720 0.0114</td>
</tr>
<tr>
<td>Subject5</td>
<td>0.1820 0.1376</td>
<td>0.5663 0.1463</td>
<td>0.6172 0.0193</td>
<td>0.7080 0.0028</td>
</tr>
<tr>
<td>Subject6</td>
<td>0.2316 0.1094</td>
<td>0.5695 0.1928</td>
<td>0.6062 0.0120</td>
<td>0.6872 0.0086</td>
</tr>
<tr>
<td>Subject7</td>
<td>0.1080 0.1686</td>
<td>0.5316 0.2091</td>
<td>0.5769 0.0172</td>
<td>0.6966 0.0054</td>
</tr>
<tr>
<td>Subject8</td>
<td>0.1755 0.1045</td>
<td>0.5696 0.1777</td>
<td>0.6231 0.0097</td>
<td>0.7037 0.0103</td>
</tr>
<tr>
<td>Subject9</td>
<td>0.2584 0.0893</td>
<td>0.5913 0.1901</td>
<td>0.6295 0.0216</td>
<td>0.7356 0.0200</td>
</tr>
<tr>
<td>Subject10</td>
<td>0.0113 0.0135</td>
<td>0.5751 0.2901</td>
<td>0.6247 0.0198</td>
<td>0.7441 0.0115</td>
</tr>
<tr>
<td>Subject11</td>
<td>0.2121 0.2388</td>
<td>0.5827 0.2652</td>
<td>0.6346 0.0212</td>
<td>0.7491 0.0042</td>
</tr>
<tr>
<td>Subject12</td>
<td>0.1547 0.1170</td>
<td>0.5702 0.1941</td>
<td>0.6226 0.0045</td>
<td>0.7256 0.0095</td>
</tr>
</tbody>
</table>
FCM, (b)TSC, (c) T1-KT-FCM, (d)NTR-FC-SCT. 

TABLE 7
ARI Performance of all Comparison Methods on 5 percent Noisy CT Image Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>FCM</th>
<th>TSC</th>
<th>T1-KT-FCM</th>
<th>NTR-FC-SCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject1</td>
<td>0.0021</td>
<td>0.0043</td>
<td>0.3399</td>
<td>0.1756</td>
</tr>
<tr>
<td>Subject2</td>
<td>0.0854</td>
<td>0.0814</td>
<td>0.3931</td>
<td>0.1561</td>
</tr>
<tr>
<td>Subject3</td>
<td>0.0861</td>
<td>0.0521</td>
<td>0.4144</td>
<td>0.1478</td>
</tr>
<tr>
<td>Subject4</td>
<td>0.0013</td>
<td>0.0041</td>
<td>0.3466</td>
<td>0.1701</td>
</tr>
<tr>
<td>Subject5</td>
<td>0.1186</td>
<td>0.0877</td>
<td>0.3884</td>
<td>0.1761</td>
</tr>
<tr>
<td>Subject6</td>
<td>0.0839</td>
<td>0.0560</td>
<td>0.4153</td>
<td>0.1393</td>
</tr>
<tr>
<td>Subject7</td>
<td>0.0012</td>
<td>0.0026</td>
<td>0.3993</td>
<td>0.1994</td>
</tr>
<tr>
<td>Subject8</td>
<td>0.0651</td>
<td>0.0732</td>
<td>0.3902</td>
<td>0.1753</td>
</tr>
<tr>
<td>Subject9</td>
<td>0.0658</td>
<td>0.0682</td>
<td>0.4153</td>
<td>0.1524</td>
</tr>
<tr>
<td>Subject10</td>
<td>0.0701</td>
<td>0.0815</td>
<td>0.3806</td>
<td>0.2090</td>
</tr>
<tr>
<td>Subject11</td>
<td>0.0771</td>
<td>0.1614</td>
<td>0.3858</td>
<td>0.1903</td>
</tr>
<tr>
<td>Subject12</td>
<td>0.0094</td>
<td>0.0109</td>
<td>0.3744</td>
<td>0.1365</td>
</tr>
</tbody>
</table>

TABLE 8
ARI Performance of all Comparison Methods on 10 percent Noisy CT Image Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>FCM</th>
<th>TSC</th>
<th>T1-KT-FCM</th>
<th>NTR-FC-SCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject1</td>
<td>0.0026</td>
<td>0.0113</td>
<td>0.3403</td>
<td>0.1791</td>
</tr>
<tr>
<td>Subject2</td>
<td>0.1798</td>
<td>0.1182</td>
<td>0.3955</td>
<td>0.1613</td>
</tr>
<tr>
<td>Subject3</td>
<td>0.1158</td>
<td>0.0542</td>
<td>0.4121</td>
<td>0.1696</td>
</tr>
<tr>
<td>Subject4</td>
<td>0.0486</td>
<td>0.0063</td>
<td>0.3433</td>
<td>0.1729</td>
</tr>
<tr>
<td>Subject5</td>
<td>0.1051</td>
<td>0.0953</td>
<td>0.3848</td>
<td>0.1535</td>
</tr>
<tr>
<td>Subject6</td>
<td>0.1523</td>
<td>0.0812</td>
<td>0.4168</td>
<td>0.1373</td>
</tr>
<tr>
<td>Subject7</td>
<td>0.0008</td>
<td>0.0012</td>
<td>0.3607</td>
<td>0.1920</td>
</tr>
<tr>
<td>Subject8</td>
<td>0.1140</td>
<td>0.0988</td>
<td>0.3915</td>
<td>0.1311</td>
</tr>
<tr>
<td>Subject9</td>
<td>0.1202</td>
<td>0.1143</td>
<td>0.4210</td>
<td>0.1781</td>
</tr>
<tr>
<td>Subject10</td>
<td>0.0011</td>
<td>0.0089</td>
<td>0.3800</td>
<td>0.1963</td>
</tr>
<tr>
<td>Subject11</td>
<td>0.0493</td>
<td>0.1062</td>
<td>0.3848</td>
<td>0.1608</td>
</tr>
<tr>
<td>Subject12</td>
<td>0.0484</td>
<td>0.0401</td>
<td>0.3737</td>
<td>0.1681</td>
</tr>
</tbody>
</table>

noisy scenario. The regularization term in NTR-FC-SCT (λ₁ = 0) is 0.97, which finds a shared latent space for data cross domains, so that the projection data distributions of the source and target domains are close to each other. NTR-FC-SCT has the advantages of both NTR-FC-SCT (λ₁ = 0) and NTR-FC-SCT (λ₂ = 0). It can exploit more transfer knowledge; meanwhile, and achieve a good balance between making use of positive transfer and resisting negative transfer.

Next, we discuss the influence of the number of samples in the source domain on the performance of NTR-FC-SCT. We randomly select 10, 30, 50, 70, 90 and 100 percent proportion of training samples in the source domain as the source training dataset. To make the results fair, we repeat the above sampling 10 times for each sample size. NMI and ARI performances of NTR-FC-SCT on Subject1 and Subject2 with 5 percent noisy are shown in Figs. 10 and 11, respectively. The experimental results show that the values of NMI and ARI increase with the increase of the number of samples in the source domain. The reason is that NTR-FC-SCT can
In the experiments, the parameters $\lambda_1$ and $\lambda_2$ are determined in a given search grid. In the following, we discuss the performance of NTR-FC-SCT using different parameters. Tables 15, 16 show the means of NMI and ARI on the subject using different $\lambda_1$ and $\lambda_2$, while fixing the parameter $m = 2$.

1) NTR-FC-SCT is sensitive to parameters $\lambda_1$ and $\lambda_2$. Different $\lambda_1$ and $\lambda_2$ lead to different cluster performance of NTR-FC-SCT in terms of NMI and ARI. It can be found that in most situations when the value of NMI is better, the value of ARI is also better. Thus, it is feasible to use NMI and ARI as performance criterions to determine the suitable parameters.

2) Fixed the value of $m$, NTR-FC-SCT obtains the worst NMI and ARI when $\lambda_1 = 0$ and $\lambda_2 = 0$. The clustering performance of NTR-FC-SCT is improved when $\lambda_1$ and $\lambda_2$ are not equal to 0. Since when $\lambda_1 = 0$ and $\lambda_2 = 0$ NTR-FC-SCT is degenerated to the classical FCM clustering.

3) We can find that when the value of $\lambda_1$ is large, NTR-FC-SCT obtains the satisfactory performance in terms of NMI and ARI. This further demonstrates that the proposed negative-transfer-resistance mechanism has played an effective role. Thus, in the subsequent experiments, we can reduce the search grid of $\lambda_1$ in the range $\{0, 10e-4, \ldots, 10e6\}$. We can’t find the rule to select parameter $\lambda_2$. We think it is reasonable to select optimal $\lambda_2$ within the search grid. The range $\{0, 10e-4, \ldots, 10e6\}$ is appropriate.

### 5 Conclusion

In this study, we have addressed the problem of medical image segmentation with insufficient and noisy samples,
and proposed NTR-FC-SCT model for leveraging source knowledge to improve the segmentation performance of target domain. We explore the negative-transfer-resistant mechanism to reinforce the influence of positive transfer and reduce, or even eliminate, the negative transfer. In particular, we find a shared latent space based on the idea of MMD, in which the mapped data distributions of source domain and target domain are close to each other. The experiments focus on noisy brain CT images. The experimental results show that with insufficient and noisy medical images, it is possible to build an efficient segmentation model with the help of medical images from the related scenarios. Future work will extend our algorithm to other medical image segmentation applications. We will extend the framework so as to apply various clustering algorithms in order to obtain more satisfactory medical image segmentation results. We will also study how many images in the source domain can be considered sufficient, and how to select the important images to further improve the transfer. In addition, how to speed up NTR-FC-SCT is worthy to be studied in the future.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grants 61806026, 61702225, 61772241, and 61711540041, in part by the Natural Science Foundation of Jiangsu Province under Grants BK20180956, BK20161268, and BK20160187, in part by the Fundamental Re-search Funds for the Central Universities under Grants JUSRP51614A and JUSRP11737, in part by 2016 Qinglan Project of Jiangsu Province, and in part by Six Talent Peaks Project of Jiangsu Province under Grant XYDXX-127.

REFERENCES


Xiaqing Gu received the PhD degree from Jiangnan University, in 2017. She is currently an associate professor with the School of Information Science and Engineering, Changzhou University, Changzhou, China. She has published more than 30 papers in international/national authoritative journals. Her current research interests include pattern recognition and machine learning.

Dongrui Wu (S’05–M’09–SM’14) received the BE degree in automatic control from the University of Science and Technology of China, in 2003, the ME degree in electrical engineering from the National University of Singapore, in 2005, and the PhD degree in electrical engineering from the University of Southern California, in 2009. He is currently a professor with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China, and the deputy director of the Key Laboratory of Image Processing and Intelligent Control, Ministry of Education. His research interests include affective computing, brain computer interfaces, computational intelligence, and machine learning. He has more than 120 publications, including a book entitled Perceptual Computing (Wiley-IEEE Press, 2010). He received the IEEE Computational Intelligence Society Outstanding PhD Dissertation Award, in 2012, IEEE TRANSACTIONS ON FUZZY SYSTEMS Outstanding Paper Award, in 2014, NAIFIPS Early Career Award, in 2014, IEEE Systems, Man and Cybernetics (SMC) Society Early Career Award, in 2017, and the IEEE SMC Society Best Associate Editor Award, in 2018. He was also a finalist of another three Best Paper Awards. He was/is an associate editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS (2011-2018), IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS (2014-), IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE (2017-), and IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING (2019). He is a Senior member of the IEEE.

Wenlong Hang received the PhD degree from Jiangnan University, in 2017. He is currently a lecturer with the School of Computer Science and Technology, Nanjing Tech University, Nanjing, China. He has published more than 10 papers in international/national authoritative journals. His current research interests include pattern recognition and machine learning.

Chin-Teng Lin (S’88–M’91–SM’99–F’05) received the BS degree in control engineering from the National Taiwan Institute of Technology (Taipei), Taiwan, in 1986, and the master’s and PhD degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989 and 1992, respectively. He is currently a distinguished professor of the Faculty of Engineering and Information Technology, University of Technology Sydney. He also holds an Honorary chair professorship with Electrical and Computer Engineering, NCTU, International Faculty of University of California at San-Diego, La Jolla, CA, USA, and an honorary professorship with the University of Nottingham, Nottingham, UK. He is the co author of Neural Fuzzy Systems (Prentice-Hall, 1996), and the author of Neural Fuzzy Control Systems with Structure and Parameter Learning (World Scientific, 1994). He has published more than 220 journal papers (Total Citation: 18,236, H-index: 57, 110-index: 295) and holds 97 patents in the areas of computational intelligence, fuzzy neural networks, natural cognition, brain–computer interface, intelligent system, multimedia information processing, machine learning, robotics, and intelligent sensing and control, including approximately 108 IEEE journal papers. He was elevated to a fellow of the IEEE for his contributions to biologically inspired information systems, in 2005, and was elevated to an International Fuzzy Systems association fellow, in 2012. He received the Merit National Science Council Research Fellow Award, Taiwan, in 2009, Outstanding Electrical and Computer Engineer, Purdue University, in 2011, Outstanding Achievement Award by Asia Pacific Neural Network Assembly, in 2013, and IEEE Fuzzy Systems Pioneer Award, in 2016. He served as the deputy editor-in-chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II in 2006–2008 and as editor-in-chief of the IEEE TRANSACTIONS ON FUZZY SYSTEMS from 2011 to 2016. He also served on the Board of Governors of the IEEE Systems, Man, Cybernetics Society during 2003–2005, IEEE Circuits and Systems Society during 2005–2008, and IEEE Computational Intelligence Society (CIS) during 2008–2010. He was the chair of the IEEE Taipei Section during 2009–2010. He was a distinguished lecturer of the IEEE CAS Society from 2003 to 2005 and the CIS Society from 2015 to 2017. He was the program chair of the IEEE International Conference on Systems, Man, and Cybernetics, in 2005 and the general chair of the 2011 IEEE International Conference on Fuzzy Systems.

Shi Qiu received the MS degree from Xidian University, Xi’an, P. R. China, in 2012, the PhD degree from the University of Chinese Academy of Sciences, Xi’an, P. R. China, in 2016. He is currently an associate research fellow with the Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an, P. R. China. His research interests include image restoration, sparse representation, and object detection.

Jing Xue received the MM degree from the Department of Nephrology, Nanjing Medical University, China. She received the MD (doctor of medicine) degree from the Department of Nephrology, Nanjing Medical University, China, in 2019. Her research interests include machine learning and its applications on medicine. She is a doctor with the Affiliated Wuxi People’s Hospital of Nanjing Medical University, China.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.