

“© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Multiple Relevant Feature Ensemble Selection Based on Multilayer Co-Evolutionary Consensus MapReduce

Weiping Ding¹, Member, IEEE, Chin-Teng Lin,² Fellow, IEEE, and Witold Pedrycz³, Fellow, IEEE

Abstract—Although feature selection for large data has been intensively investigated in data mining, machine learning, and pattern recognition, the challenges are not just to invent new algorithms to handle noisy and uncertain large data in applications, but rather to link the multiple relevant feature sources, structured, or unstructured, to develop an effective feature reduction method. In this paper, we propose a multiple relevant feature ensemble selection (MRFES) algorithm based on multilayer co-evolutionary consensus MapReduce (MCCM). We construct an effective MCCM model to handle feature ensemble selection of large-scale datasets with multiple relevant feature sources, and explore the unified consistency aggregation between the local solutions and global dominance solutions achieved by the co-evolutionary memplexes, which participate in the cooperative feature ensemble selection process. This model attempts to reach a mutual decision agreement among co-evolutionary memplexes, which calls for the need for mechanisms to detect some noncooperative co-evolutionary behaviors and achieve better Nash equilibrium resolutions. Extensive experimental comparative studies substantiate the effectiveness of MRFES to solve large-scale dataset problems with the complex noise and multiple relevant feature sources on some well-known benchmark datasets. The algorithm can greatly facilitate the selection of relevant feature subsets coming from the original feature space with better accuracy, efficiency, and interpretability. Moreover, we apply MRFES to human cerebral cortex-based classification prediction. Such successful applications are expected to significantly scale up classification prediction for large-scale and complex brain data in terms of efficiency and feasibility.

Index Terms—Cerebral cortex classification, co-evolutionary consensus MapReduce, consistency aggregation, multiple relevant feature selection, Nash equilibrium.

I. INTRODUCTION

IN RECENT years, massive amounts of data have become available for all kinds of industrial applications, and big data has emerged as an important research topic and a visible application domain. Big data analytics can definitely reveal valuable knowledge. Reflecting the very nature of the data in big data, we often refer to so-called the “5V” aspect: 1) volume; 2) variety; 3) velocity; 4) veracity; and 5) value [1]. It is critical to extract knowledge and build models from big data. But the real challenge comes with the requisition of such knowledge, which is quantitative, defined across multiple space-time scales, and capable of prediction with sufficient accuracy [2]. It poses evident demands on conventional methods currently used in data mining and machine learning, including transmission, storage, processing, and optimization [3], [4]. Furthermore, many big datasets increase dynamically in size and contain various elements of noise. Many features are likely redundant or irrelevant. These useless features often diminish the learning process associated with classification algorithms, decreasing their overall performance. Various real-world big data applications can be formulated as feature selection problems. Hence, to analyze the high-dimensional datasets with huge numbers of features is a thought-provoking problem, that has led to new and important developments in feature selection techniques [5], [6].

Pawlak’s rough sets offer a theoretical framework supporting feature selection. Feature selection based on rough set theory (RST) has become an effective way to use information granules to build efficient models for complex applications, especially in the context of big data. The feature subset obtained by a feature selection algorithm is realized through attribute reduction, which selects a subset of features with the highest discriminatory power [7], [8]. The algorithm aims to select a relevant feature subset from the original feature space to improve the accuracy, efficiency, and interpretability of a predictive model. With feature relevance with respect to a class attribute, an input feature can be classified as strongly relevant, irrelevant, redundant, or nonredundant. Some feature selection algorithms based on RST have attracted more attention in different research areas [9]–[14].

Manuscript received February 17, 2018; revised June 2, 2018; accepted July 21, 2018. This work was supported in part by the Australian Research Council under Grant DP180100670 and Grant DP180100656, in part by the National Natural Science Foundation of China under Grant 61300167, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20151274, in part by the Jiangsu Provincial Government Scholarship Program under Grant JS-2016-065, in part by the Six Talent Peaks Project of Jiangsu Province under Grant XYDXXJS-048, and in part by the Applied Basic Research Program of Nantong under Grant GY12016014. This paper was recommended by Associate Editor K.-C. Tan. (Corresponding author: Weiping Ding.)

W. Ding is with School of Computer Science and Technology, Nantong University, Nantong 226019, China (e-mail: dwp9988@163.com).

C.-T. Lin is with the Centre for Artificial Intelligence, FEIT, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: chin-teng.lin@uts.edu.au).

W. Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada (e-mail: wpedrycz@ualberta.ca).

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/TCYB.2018.2859342

Models and applications of feature selection based on RST have been discussed extensively. Although they can select better features that preserve discernibility, they seldom attempt to maximize class separability. This can produce some undesirable feature subsets from multiple relevant feature sources with many inconsistent feature boundaries. When the number of features involving noise increases dynamically in large datasets, traditional methods often must repeatedly implement feature selection, resulting in huge computing overhead and memory requirements. Moreover, with the ever-increasing speed and volume of data, the performance of existing methods rapidly deteriorates. Hence, these algorithms cannot perform satisfactorily for large-scale complex features in real-world applications. In addition, more challenges emerge with large-scale and streaming data for processing, and practical applications with the increased complexity and dimensionality of search spaces usually presented by large datasets. Therefore, classification or prediction tasks involving high-dimensional heterogeneous data have become common in analytical engineering data and medical imaging, creating an urgent need for more efficient feature selection algorithms to acquire better-quality solutions and reduce computational complexity.

The co-evolutionary paradigm is inspired by the reciprocal evolutionary change driven by the cooperative or competitive interaction between species, and it can be broadly classified into the categories of cooperative and competitive co-evolution. In the case of cooperative co-evolutionary behaviors, individuals are rewarded when they work well with others and are punished when they perform poorly together [15]. But competitive co-evolutionary behaviors mean that various sub-populations always fight to gain an advantage over others to obtain common resources.

Cooperative co-evolution (CC) employs a divide-and-conquer strategy to solve large-scale optimization problems [16]. CC can decompose a large-scale problem into a number of subproblems, each initialized and optimized by the traditional evolutionary algorithm [17]. Essentially, each memplex can search a collection of variables, and the fitness of a memplex is an estimate of how well it cooperates with other memplexes to produce good solutions. These collaborators are updated at specific intervals. The final solution is a chain combination containing the representative subsolutions for all subproblems. The performance of CC relies heavily on decomposition strategies. For better decomposition, most decomposition strategies have been developed by obtaining the correct interdependency information and requiring many fitness identification evaluations [18]–[21]. MapReduce is a popular parallel model used to analyze large datasets. A number of traditional methods have been combined with MapReduce [22]–[24]. MapReduce-based feature selection for massive datasets is also gaining more attention across different research domains. But existing algorithms often assume that datasets are loaded into the main memory of a single machine, which is obviously infeasible for large-scale data, especially for multiple relevant feature sources.

To address these multiple challenges, this paper presents a multiple relevant feature ensemble selection (MRFES)

algorithm by combining the advantages of the MapReduce model and consensus CC. This approach aims not only to achieve a better multilayer co-evolutionary consensus MapReduce (MCCM) model for feature ensemble selection but also to guarantee consistency between local solutions and global dominance solutions with the better Nash equilibrium. MRFES, validated on both synthetic data and in vivo brain images, shows promising results in terms of accuracy, efficiency, and robustness, and accords better reality than some state-of-the-art algorithms. We discuss the main properties of MRFES and elaborate on its potential applications. The main contributions of this paper are as follows.

First, we propose an MCCM model for feature selection of large-scale datasets with multiple relevant feature sources. This model attempts to achieve mutual agreement among a group of co-evolutionary decisions, which implies the need for mechanisms to facilitate the detection of noncooperative co-evolutionary MapReduce behaviors and achieve the Nash equilibrium resolution. We believe this is the first model whose co-evolutionary consensus strategy accounts for different MapReduce structure organizations for nonseparable multiple relevant feature sources.

Second, we adopt a unified consistency dominance strategy to achieve the adaptive balance between the local solutions and global dominance solutions of co-evolutionary memplexes, which further guarantees the choice of the optimal feature subsets, including strongly multiple relevant and nonredundant features. Moreover, the ensemble feature set based on MapReduce can be accelerated so that the entire dominance solution of feature selection can be achieved accurately and efficiently.

Third, unlike some previous work, we apply MRFES to the human cerebral cortex-based classification, where can be better adapted to derive from the cortical folding surfaces and better achieve a consistency term adaptive to the temporal similarities of different cortical cortices. These results further confirm that MRFES can consistently provide highly accurate segmentations to find the complex cerebral cortex features with the highest likelihood and can provide satisfactory classification prediction.

This paper is organized as follows. Some related work is reviewed and existing challenges are analyzed in Section II. An MCCM model for feature ensemble selection is presented in detail in Section III. Section IV describes the primary steps and underlying processing of MRFES. Extensive experimental evaluations of MRFES are provided in Section V. The application performance of MRFES is evaluated in human cerebral cortex-based classification in Section VI. Some discussion on MRFES is presented in Section VII. The conclusions are drawn and further research is suggested in Section VIII.

II. RELATED STUDIES AND CHALLENGES

Recent years have seen significant advances of RST in science and engineering, and numerous feature selection algorithms have been proposed. In the classical RST model, classification quality, information entropy, positive regions,

and lower approximation bounds under each decision class vary consistently and monotonically during feature selection. For example, Yao and Zhao [25] proposed a discernibility-matrix-based feature selection method with a simplified matrix to reduce the computational cost. Chen *et al.* [26] presented a sample pair selection procedure to efficiently complete feature reduction. Li *et al.* [13] developed a decision-theoretic RST model based on neighborhoods to analyze positive regions-related feature selection. Chen *et al.* [27] explored an integrated classification mechanism for simultaneous extraction of fuzzy rules and selection of useful features. Zhang and Yang [28] proposed a feature selection and approximate reasoning framework by α -dominance-based quantitative RST for large-scale set-valued information tables. Particle swarm optimization (PSO) as an efficient optimization strategy has been applied to feature selection. Shen *et al.* [29] combined the basic concepts of attribute reduction in rough sets and PSO, and developed a reduction algorithm based on PSO. Zhang *et al.* [30] proposed a method to find optimal feature subsets through barebones PSO with a reinforced memory strategy. Zhang *et al.* [31] presented multiobjective PSO for cost-based feature selection problems to generate a Pareto front of nondominated solutions. Fong *et al.* [32] designed a novel lightweight feature selection particularly for mining streaming data by using accelerated PSO. Tran *et al.* [33] proposed a potential PSO for discretization-based feature selection by a new representation to reduce the search space of the problem. Shuffled frog leaping algorithm as a fast and robust algorithm with the efficient global co-search capability also has been applied to feature selection. Ding *et al.* [34] put forward an efficient and self-adaptive feature selection algorithm by combining quantum elitist frogs and cloud model operators.

Recent work to improve feature selection algorithms includes the following. Wang *et al.* [35] introduced a variable-precision fuzzy neighborhood rough set model for feature subset selection that can tolerate noise in data. This model can decrease the possibility that a sample is classified into the wrong category and can select out the relatively small numbers of features. So, it obtains higher classification performance. In this model, however, the optimal solutions of two important parameters for each dataset cannot be automatically set, and they must be set by users in advance. Jiang *et al.* [36] proposed a relative decision entropy-based feature selection approach. Unlike existing information entropy models in rough sets, relative decision entropy is defined using two basic concepts in rough sets: 1) roughness and 2) degree of dependency. The experimental results show that it is efficient for feature selection. In particular, it can achieve good scalability for large datasets. But this approach still cannot deal with continuous features without discretization. Zhang *et al.* [37] extended the idea of the firefly algorithm by introducing binary variables and proposed a return-cost-based binary firefly algorithm for feature selection, employing strategies such as return-cost attractiveness, Pareto dominance-based selection, and binary movement with the adaptive jump. The experimental results showed effectiveness at solving feature selection problems. Ding *et al.* [38] proposed a rough feature selection

algorithm by the layered co-evolutionary strategy with neighborhood radius hierarchy. The experimental results substantiate it can achieve better effectiveness and accuracy of feature selection.

Although these algorithms are dominant in feature selection, most do not perform well in large-scale datasets because of noise and complex dimensionality of big data. When massive new objects are generated, this wastes computing time and space by using existing feature selection algorithms. Moreover, the processing time to carry out the structuralized reasoning of feature selection will grow tremendously with increasing numbers of features.

To deal with dynamically increasing large datasets, there has been historically little research on finding features using MapReduce based on RST. Several researchers have proposed some feature selection algorithms based on the MapReduce technique to address this problem. Zhang *et al.* [39] adopted the parallel algorithm to compute equivalence classes and decision classes based on MapReduce and to update rough set approximations. Zhang *et al.* [40] also introduced parallel large-scale RST methods for feature selection and implemented them on some representative MapReduce systems. Qian *et al.* [41] exploited the $\langle key, value \rangle$ pair framework based on MapReduce to accelerate the traditional feature selection process. A common feature of these algorithms is the use of the framework structure of a $\langle key, value \rangle$ pair to accelerate computation of equivalence classes and attribute significance. However, many real-world data from applications with large-large, noisy, and uncertain datasets link the multiple relevant feature sources. When multiple relevant features are generated at the same time in a database, the previously noted algorithms may be inefficient because they must be executed repeatedly to handle ever-increasing numbers of features. By dynamically adjusting the framework structure of $\langle key, value \rangle$, the complexity of feature selection in big data is alleviated to some extent. So, the paradigm of MapReduce has great potential to facilitate the implementation process of dynamical large datasets with multiple relevant feature sources.

As discussed, the complex big data is widely seen as an essential problem in the design of complex systems. Feature selection plays a visible role in such technologies as data mining, machine learning, and knowledge reasoning with uncertainty. As a matter of fact, the divide-and-conquer strategy has long been used in large-scale retrieval and learning, just like the framework of CC and MapReduce. No general theoretical framework or computational model, however, has been designed. Although much effort has been dedicated to feature selection to solve large-scale problems in the face of uncertainties, the noted algorithms suffer from the following limitations and challenges.

- 1) Approximation and granulation are two important issues when applying feature selection algorithms to real-world large-scale datasets. The previous studies have focused almost exclusively on how to define some efficient rough approximation operators, but limited work has been devoted to the problem of extracting multiple relevant feature relations from large-scale datasets. Most feature selection models have not systematically discussed

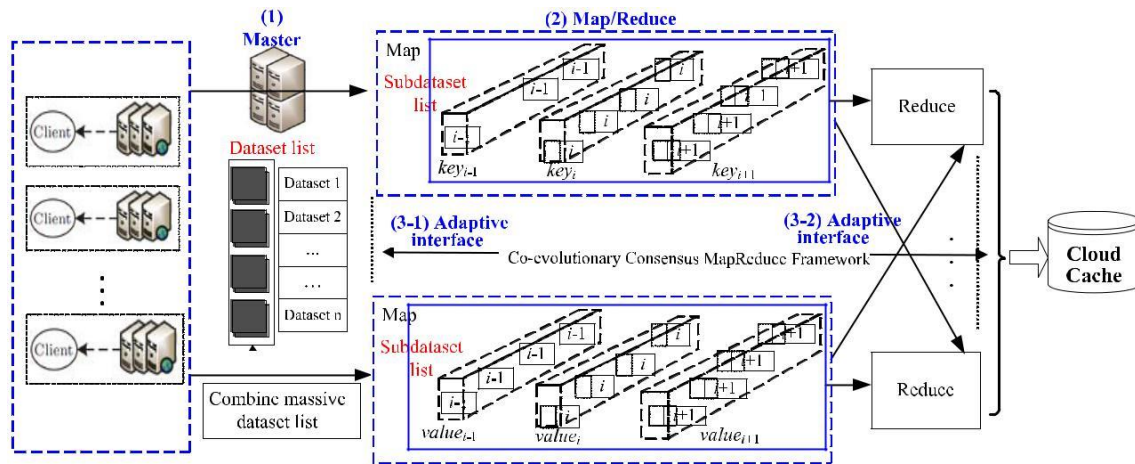


Fig. 1. Processing framework of MCCM model.

how to deal with similarity relations of multiple relevant features from large datasets.

- 2) A few efforts have been made to define, characterize, and assess feature selection with MapReduce, but this has not been studied in depth. For example, how does one deal with multiple relevant feature variables in different practical attribute sets? How do we achieve better feature selection for large-scale complex attributes in changing and interconnected real-world applications? How do we address the limitations of existing MapReduce structures and interactions through dynamic adaptation with a reorganization model?
- 3) In fact, it seems that similarity relations of multiple relevant features are not easily computed in real-world applications. Nevertheless, the way to extract reliable relations from multiple relevant features directly affects the performance of models, and no general approach has been designed or even has been discussed. Therefore, it is important to develop a systematic and effective approach combined with CC and MapReduce to determine similarity relations from multiple relevant feature variables in large datasets.

Although some progress has been made in feature selection, many challenges remain to be addressed in feature selection models and algorithms for multiple relevant feature variables in large datasets. We introduce an MCCM model, and propose an MRFES algorithm for MRFES. We demonstrate the comparative merits of MRFES using some benchmark datasets and real-world applications in the prediction of human cerebral cortex-based classification.

III. MULTILAYER CO-EVOLUTIONARY CONSENSUS MAPREDUCE MODEL

This section presents an MCCM model for feature ensemble selection is, which can deal with large-scale datasets coming from multiple relevant feature sources. This model is unique in its way of reaching mutual agreement among different co-evolutionary memplexes, as well as dealing with elitists with unified consistency aggregation for Nash dominance (UCAND) solutions. This implies the need for mechanisms to

facilitate the detection of some noncooperative co-evolutionary behaviors during co-evolutionary consensus MapReduce. In this selection, the main phases of MCCM description are composed of two parts. The MCCM framework is described in Section III-A. The UCAND solutions is implemented in Section III-B.

We implemented the co-evolutionary consensus scheduling of a large dataset processing architecture on MapReduce, which includes three main operations, as shown in Fig. 1.

A. MCCM Model

- 1) Master operation is responsible for the parallel dataset list allocation and scheduling of multiple relevant features from different clients. The module “combine massive dataset list” is run in Master operation to process multiple relevant features by one Map task.
- 2) Map/Reduce operation. After collecting subdataset lists from Master operation, the Map operation can deal with multiple relevant features by assigning the relevant results to Reduce by Master. The results display in the form of $\langle key_i, value_i \rangle$ pairs, which will be aggregated by the Reduce operation.
- 3) Adaptive interface operation provides the unified interface for subdataset lists from different kinds of multiple relevant features. The final results will be stored in Cloud Cache. Its identification and partition process for nonseparable multiple relevant features are described in Fig. 2.

The main steps of MCCM are shown in Algorithm 1.

B. Unified Consistency Aggregation for Nash Dominance Solutions

A theoretical framework and practical algorithm is needed for big data the analysis based on granular computing, and it is of interest to construct an analytical model based on evolutionary game theory. In fact, the Nash equilibrium theory has been used for the theoretical model in multigranular representation and learning, just like the MapReduce framework.

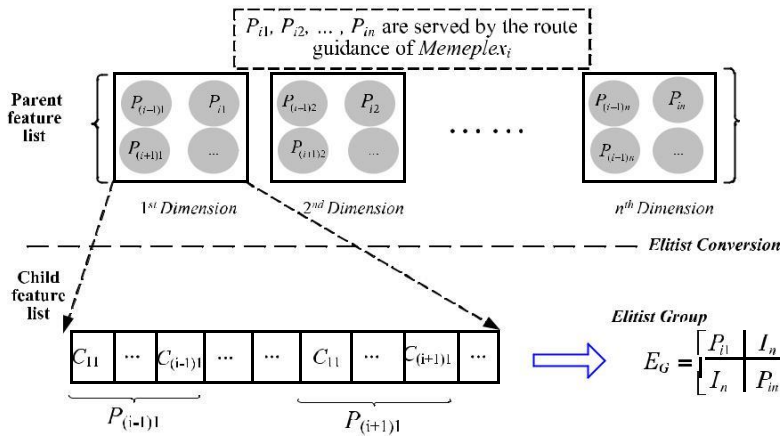


Fig. 2. Identification process for nonseparable multiple relevant features.

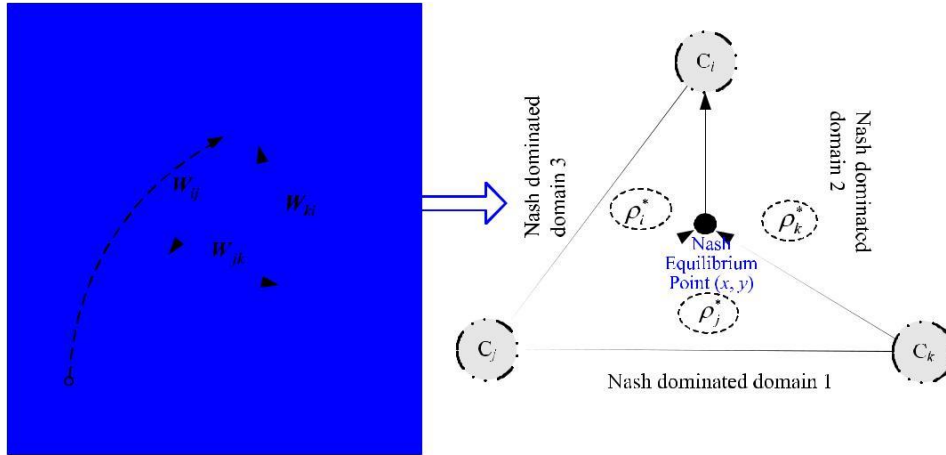


Fig. 3. Process to visualize UCAND solutions.

This section adopts a unified consistency dominance strategy to achieve the adaptive Nash equilibrium between the local solutions and global dominance solutions, as shown in Fig. 3, which will further guarantee the choice of the optimal feature subsets including strongly multiple relevant and nonredundant features. To focus our attention on feature selection dynamics, we design the unified consistency aggregation strategy for Nash dominance solutions to eliminate confounding factors, providing an elegant framework within which one can balance the behaviors of various co-evolutionary MapReduce operators. Here, the strength of the Nash equilibrium rests on such vital assumptions about the rational agents with the complete knowledge of their domains. For clarity, this polymorphic Nash equilibrium involves three Nash-dominated domains, in which the Nash equilibrium will be evaluated by the fitness equilibrium. So, the entire dominance solution of feature selection can be better achieved. The UCAND solutions are constructed through a sequence of steps.

This unified consistency aggregation can adjust Nash dominance solutions by a complementary diversity-preservation mechanism between local solutions and global dominance solutions of co-evolutionary memplexes. This approach is used to exploit the deep similarity between pairs of feature decision classes in the same neighborhood radius vector. In the global exploration, the evolution of isolated feature decision

classes tends to select more diverse features by using different ensemble weight vectors. In local exploitation, a diverse feature class is driven by the necessity to select different situations posed by the other feature decision classes. This will further guarantee the selection of the optimal feature subsets, including strongly multiple relevant and nonredundant features. Hence, the ensemble feature sets based on MapReduce can be accelerated to select out, and the entire dominance solution of feature selection can be achieved.

IV. PROPOSED MRFES ALGORITHM

To fully explore the property of an MCCM model for extracting multiple relevant feature relations from large datasets, we propose the MRFES algorithm based on MCCM and UCAND, to serve as a guide to conduct feature selection based on MapReduce. This algorithm can detect different elitists' noncooperative behaviors for feature selection and achieve Nash dominance solutions. The adaptive consensus scheme of MRFES allows for easy automation of human-moderator tasks, thus removing inherent subjective bias, and reaches a mutual decision agreement among co-evolutionary memplexes based on MapReduce. It can achieve Nash dominance between local solutions and global dominance solutions by the co-evolutionary MapReduce. The main steps of MRFES are shown in Algorithm 3.

Algorithm 1: MCCM

- 1) Construct the parallel operation $\langle key_i, value_i \rangle$ for Map/Reduce of memplex_{*i*} as

$$key_i = \frac{J_{ELITIST} - J_{Elitist_i}}{J_{ELITIST}}, \quad (1)$$

and

$$value_i = \frac{J_{Elitist_i}}{\sum_{j=1}^r J_j}, \quad (2)$$

where $J_{Elitist_i}$ is the local best fitness of elitists in memplex_{*i*}, $J_{ELITIST}$ is the global best fitness of all memplexes, $J_{ELITIST} = \text{Min}\{J_{Elitist_1}, J_{Elitist_2}, \dots, J_{Elitist_N}\}$, J_j is the fitness of the j^{th} elitist, and r is the number of co-evolutionary elitists in memplex_{*i*}.

- 2) Design the parallel operation matrix of memplexes as

$$\mathbf{KV} = \begin{bmatrix} kv_{11} & \dots & kv_{1n} \\ \vdots & \ddots & \vdots \\ kv_{n1} & \dots & kv_{nn} \end{bmatrix}_{n \times n}, \quad (3)$$

where $kv_{ij} = \frac{key_i}{value_j}$.

- 3) Compute the collective preference and proximity matrices for elitists, and list the adjacent matrix set of elitists as $\mathbf{P} = \{P_1, P_2, \dots, P_i, \dots, P_n\}$, where

$$P_i = \frac{m}{\sum_{j=1}^m w_j} \times \frac{value_i}{\sum_{i=1}^m w_j}, \quad (4)$$

where w_j is the j^{th} elitist's importance weight to be updated before

- 4) A proximity matrix \mathbf{PE} between each elitist's preference relation and the adjacent matrix \mathbf{P} is defined as

$$\mathbf{PE} = \begin{bmatrix} pe_{i1} & \dots & pe_{in} \\ \vdots & \ddots & \vdots \\ pe_{i1} & \dots & pe_{in} \end{bmatrix}_{n \times n}, \quad (5)$$

where the feature relation value pe_{lk} is obtained for each pair of elitists ($Elitist_l, Elitist_k$) as

$$pe_{lk} = \frac{p \times value_l - value_k}{i \times key_l - key_k}$$

Feature relation values are used to identify the furthest preferences from the collective opinion, which should be modified by some elitists.

- 5) Conduct the identification process for nonseparable multiple relevant features among co-evolutionary memplexes according to the following steps:

(i) Select the Parent feature list, and $P_{i1}, P_{i2}, \dots, P_{in}$ are served by

the route guidance of Memplex_{*i*}.

(ii) Group all multiple relevant features into different memplexes by the incremental Dimension, and construct Cartesian coordinates of the $\{P_{(i-1)1}, P_{i1}, P_{(i+1)1}\}$ in the form of three-dimensional subvectors in the Child feature list as

$$P_{(i-1)1} = [C_{11}, C_{21}, C_{31}, \dots, C_{(i-1)1}]^t$$

$$P_{i1} = [C_{11}, C_{21}, C_{31}, \dots, C_{i1}]^t,$$

$$P_{(i+1)1} = [C_{11}, C_{21}, C_{31}, \dots, C_{i1}, C_{(i+1)1}]^t, \quad (7)$$

- (iii) Normalize the adjacent matrix set of elitists \mathbf{P} as eigenvectors with the empirical covariance matrix Cov_i corresponding to the largest eigenvalues $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_n]^t$ (where $\lambda_j \geq \lambda_{j+1}$) as follows:

$$Cov_i = \frac{\sum_{i=1}^n (value_j - \mu) \times (key_j - \mu)}{n - 1} = \frac{\sum_{j=1}^n \lambda_j \times value_j \times key_j}{n - 1}, \quad (8)$$

where μ is the mean number of Child feature lists, $value_j$ and

key_j are two parallel operations of Map/Reduce of memplex_{*i*}, and λ_j is the j^{th} eigenvalue of eigenvectors.

- 6) Construct the values of the eigenvectors to follow the principle of the multivariate Gaussian distribution with zero mean, and the elitist conversion matrix is defined as

$$\mathbf{ECM} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \quad (9)$$

where $\text{diag}(\)$ represents a diagonal matrix with the independent diagonal elements $(\lambda_1, \lambda_2, \dots, \lambda_n)$.

- 7) Determine the number of components to keep \mathbf{ECM} on a basis of η , and the proportion of total variance can be preserved in this model

($\eta = 0.92$) as

$$covi = \min(covi), \text{ subject to: } \frac{\sum_{j=1}^p \lambda_j}{\sigma_{tot}^2} \geq \eta, \quad (10)$$

where σ_{tot}^2 is the total variance, defined as

$$\sigma_{tot}^2 = \sum_{j=1}^n \lambda_j. \quad (11)$$

- 8) Pairs of alternatives ($Elitist_i, Elitist_j$) whose consensus degrees cai and cpj are not sufficiently high are identified as

$$CC = \{(Elitist_i, Elitist_j) | cai < \sigma_{tot}^2 \wedge cpj < \sigma_{tot}^2\}. \quad (12)$$

- 9) The average proximity P_{in} of an elitist group is calculated as

$$P_{in} = \frac{\sum_{i=1}^n P_i}{n}. \quad (13)$$

- 10) Achieve the elitist group matrix of nonseparable multiple relevant features as

is a n -dimensional identity matrix.

$$EG = \frac{P_{ii}}{n} \Big| \frac{I_n}{P_{in}} , \quad (14)$$

w
h
e
r
e

 I
 n

V. EXPERIMENTAL STUDIES

We performed the following experiments to show the efficiency and robustness of MRFES. The experimental system was a four-node cluster connected with gigabit Ethernet. Each node was a dual-way x86 server equipped with two Intel Xeon E5-2630 V2 CPU (2.6 GHz, 6 cores, 12 threads, 15 MB cache), and one NVIDIA Tesla C2050. One server was used as both master and worker, and the other servers were used as workers. Linux (Redhat 6.4) and Hadoop (version 0.20.203.0) were used in the evaluation. The software used in all experiments was Microsoft Visual Studio 2015,

and the programming language was C# 5.0. To decrease the randomness in different methods, all values of evaluation measures were averages of five independent tenfold cross-validation experiments.

A. Evaluation of Feature Selection Comparison on Different Datasets

We conducted this experiment to evaluate the feature selection and classification efficiency of MRFES, compared with the feature selection algorithms as FCBF [42], PCFS [43], MRF [44], and NSPSOFS [45], on two groups

Algorithm 2: UCAND

- 1) Classify whole memplexes into different clusters, where the cluster centers C_i in the current round are computed from the adjacent matrix set of elitists \mathbf{P} .
- 2) Connect each elitist with two neighbors in one cluster group and establish the ring topology. Then randomly choose a cluster center C_i corresponding to other C_j , to construct the new topology relationship of long-distance links with probability η .
- 3) Given any pair of cluster points (d_i, d_j) in the same neighborhood radius vector and an arbitrary point d_k in their contrary group, the weight metric between C_i and C_j is

$$W_{ij} = \begin{cases} 1 & \text{if } (d_i \text{ is } d_k \text{ nearest neighbors of } d_j \\ & \text{or } d_j \text{ is } d_k \text{ nearest neighbors of } d_i), \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

$\rho_i = \sum_{j=1}^N W_{ij}$ is the weight of cluster point d_i , where N is the ensemble size of cluster weights.

- 4) Construct the i^{th} ensemble weight vector $EW_i (i = 1, 2, \dots, ND)$, and the i^{th} co-evolutionary decision class $DC_j (j = 1, 2, \dots, ND)$, where ND is the number of feature decision classes. So, the feature decision class can be viewed as a matrix of weight metric distributions $\{W_{ij}\}$, which indicates the classification prediction from the ensemble weight vector EW_i for co-evolutionary decision class DC_j .
- 5) Divide the Nash dominance domain into three equal-area triangles ³, in which the arrow indicates the direction for sorting cluster points in each triangle. Then perform the pairwise comparison of ensemble weight vectors $EW_i (i = 1, 2, \dots, ND)$ in the same Nash dominance triangle. So, the elimination and merging of feature decision classes can be carried out as follows:
 - (i) If the elimination criterion is met, feature decision classes are eliminated, one by one, and distributed between the other $ND - 1$ decision classes with a new starting cluster point C_i .
 - (ii) After ND is decremented by 1, the algorithm continues with steps 3 and 4 if $ND > 1$.
- 6) Let $U_i(\rho)$, $U_j(\rho)$ be the payoff to any cluster points C_i , C_j in the neighborhood radius vector. The co-evolutionary MapReduce strategy profile $\rho^* = (\rho_1^*, \rho_2^*, \dots, \rho_n^*)$ is a Nash equilibrium for the game as follows:

$$U_i(\rho_i^*, \rho_j^*) \geq U_i(\rho_i, \rho_j), \quad \forall \rho_i, \rho_j \in EW_i, \quad \forall i, j \in \{1, 2, \dots, ND\}. \quad (16)$$

- 7) At the Nash equilibrium point of Nash dominance solutions, no decision class can benefit by unilaterally deviating from the current Nash dominance triangle ³, as shown in Fig. 3. Regarding the basins of attraction by using different numbers of decision classes with very high proportions, three Nash dominance triangles ³ converge to the Nash equilibrium point (x, y) .
- 8) Construct the highest average prediction degree of co-evolutionary feature decision classes with Nash equilibrium point (x, y) as

$$PD = \frac{ND}{\sum_{i=1}^{ND} DC_i}. \quad (17)$$

of datasets [46]. We randomly divided the datasets into ten subsets (we used eight as training sets and the remaining as testing sets). We adopted the typical SVM and C4.5 as the classifiers, which followed their recommended parameter values in this experiment [47]. They have been identified as 2 of the 10 top algorithms of machine-learning and have been used widely in classification problems. The implementation parameters of MRFES were determined experimentally to better deal with the complex datasets linking the multiple relevant feature sources. We computed the average classification values using two classifiers based on a derived feature selection set. Meanwhile, we implemented 30 independent runs in each

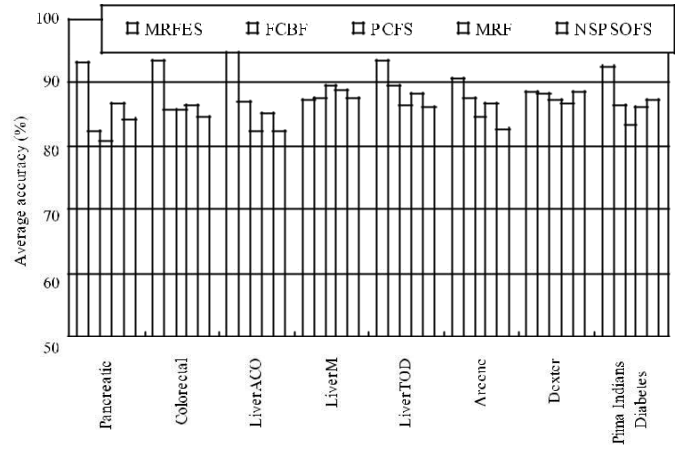


Fig. 4. Classification accuracy of MRFES against traditional FS algorithm on the second group of datasets with SVM classifier.

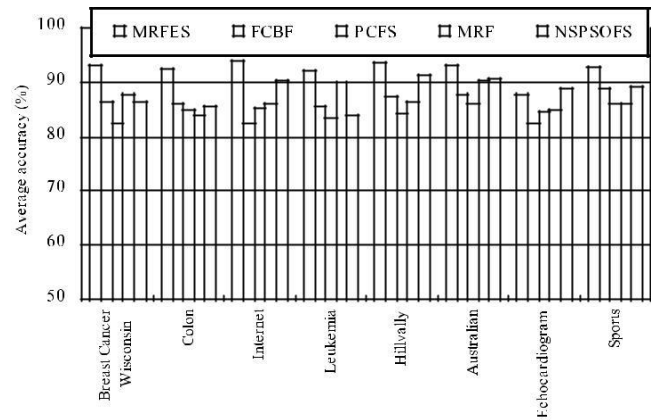


Fig. 5. Classification accuracy of MRFES against traditional FS algorithm on the second group of datasets with C4.5 classifier.

dataset and compared the noted algorithms with the best results across MRFES.

Figs. 4 and 5 show the classification accuracy with respect to MRFES and traditional FS algorithms with SVM and C4.5 classifiers, respectively. It can be observed that MRFES surpassed all four compared algorithms on two classifiers. For the SVM classifier, except for the LiverM and Dexter datasets, MRFES always obtained the best result. MRFES worked especially well with the SVM classifier for the Pancreatic, Colorectal, and LiverACO datasets, and FCBF and PCFS performed worst. Such datasets contain high uncertainty with multiple relevant features, and many objects are in boundary regions, which prevent the compared FS algorithms from finding the minimal feasible feature subset because they cannot consider the degree of overlap between the desired target feature sets. But MRFES can deal with the uncertainty with multiple relevant features more efficiently, as it uses the MCCM model. Thus, the most irrelevant and redundant features can be eliminated by MRFES to improve the classification accuracy compared with traditional FS algorithms. A similar situation can be also observed with the C4.5 classifier. Classification of Australian datasets suffered from a high error rate because of their high dimensionality and small sample size characteristics with multiple relevant feature sets,

Algorithm 3: MRFES

- 1) Initialize the search space, elitists of different memplexes for feature selection. Each memplex aims to optimize a respective feature subset.
- 2) To deal with noncooperating elitists based on elitists' importance weights, each elitist of memplex $Elitist_j \in E$ has an associated importance weight $w_j \in [0, 1]$, which is initially set to 1, $\forall i \in \{1, 2, \dots, n\}$, and the memplex matrix \mathbf{MM} are described as

$$\mathbf{MM} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_{ND} \\ 2 & 2 & \dots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ n & n & \dots & n \\ x_1 & x_2 & \dots & x_{ND} \end{pmatrix} \begin{matrix} fitness(Elitist_1) \\ fitness(Elitist_2) \\ \vdots \\ fitness(Elitist_n) \end{matrix} \quad (18)$$

- 3) Compute the similarity degree of each pair of elitists ($Elitist_i, Elitist_j$), ($i < j$) in the elitist group Ex_{g_i} according to their assessments (p^{lk}, p^{lk}) by computing the similarity function

$$sm_{ij} = 1 - \frac{|p_i^{lk} - p_j^{lk}|}{p_i^{lk} + p_j^{lk}} \quad (19)$$

- 4) Construct the similarity matrix \mathbf{SM} of each pair of elitists as follows:

$$\mathbf{SM} = \begin{pmatrix} 1 & sm_{1n} \\ \vdots & \vdots \\ sm_{n1} & \dots & sm_{nn} \end{pmatrix} \quad (20)$$

- 5) Given $P \subseteq C$ and $U/D = \{D_1, D_2, \dots, D_r\}$, where U is a finite nonempty set of objects, C is a condition attribute set, D is a decision

attribute set, and $C \setminus D \neq \emptyset$, the positive region of D with respect to the condition attribute subset P is defined as

$$POS_P(D) = \bigcup_{k=1}^r PD_k \quad (21)$$

- 6) Define the degree of dependency for feature selection $\gamma_P(D)$ as

$$\gamma_P(D) = \frac{|POS_P(D)|}{|U|} \quad (22)$$

- 7) Calculate the upper $\overline{\gamma}_{A_i}(D)$ and lower $\underline{\gamma}_{A_i}(D)$ related to each feature

$A_i \in C$, and then use **Algorithm 1** (MCCM) to select out the most

- 8) Conduct memplex-role to group according to the interdependencies of feature subsets. Many feature sets are decomposed into the same subsets because of complex interdependencies. By this regrouping strategy, the reasonable decomposition of nonseparable multiple relevant features will benefit the discovery of complex features' interdependencies.

(i) Compute the maximum value of the average weight cm^{lk}_{ij} in the i^{th} elitist memplex as $(cm^{lk})_{\max}$, and its minimum value as $(cm^{lk})_{\min}$.

(ii) Because the similarity degree sm^{lk}_{ij} is computed by aggregating similarity matrices \mathbf{SM} , taking into account the importance

weights associated with each pair of elitists ($Elitist_i, Elitist_j$), we set their weight parameter in the elitist group as $\lambda = sm^{lk}_{ij}$ which can be used to limit the number of rounds conducted in the cases that consensus cannot be achieved. So, the total average weight can be expressed as

$$cm^{lk}_{\text{Total}} = \lambda \times cm^{lk}_{\text{max}} + (1 - \lambda) \times cm^{lk}_{\text{min}} \quad (23)$$

- (iii) Define the consensus degree EM_i of the i^{th} elitist memplex as

$$EM_i = \frac{\sum_{k=1, k \neq i}^n cm^{lk}_{\text{Total}}}{|SM|} \quad (24)$$

- 9) Achieve the degree level of preference relationship coordination in the i^{th} elitist memplex as

$$Emp_i = \frac{EM_i - |SM|}{2^i} \quad (25)$$

- 10) Choose attribute subsets (Sub_attribute) $_i$, and use **Algorithm 2** (UCAND) to select out feature subsets. For the remaining features of $C \setminus C$ ($C = \emptyset$), repeat from steps 7 to 10 until the desired feature subset can be selected out.

- 11) Construct the vector of feature selection set $\{FS_1, FS_2, \dots, FS_i, \dots, FS_n\}$ in elitist memplexes ($Emp_1, Emp_2, \dots, Emp_i, \dots, Emp_n$) as follows:

$$\mathbf{FS} = \begin{pmatrix} FS_1 & FS_i & FS_n \\ FS_1 & FS_2 & FS_n \\ \vdots & \vdots & \vdots \\ FS_n & FS_n & FS_n \end{pmatrix} \quad (26)$$

- 12) Assign the corresponding weights i for the i^{th} vector of feature selection by

$$i = \frac{pe - p_i}{|PE_i|} \quad (27)$$

- 13) Optimize the vector of feature selection by the functions as

$$FS^1 = 1 \times FS_1, FS^2 = 2 \times FS_2, \dots, FS^i = i \times FS_i, \dots, FS^n = n \times FS_n \quad (28)$$

- 14) Achieve the overall vector of feature selection sets as

$$\mathbf{FS} = \begin{pmatrix} n \\ \vdots \\ i \times FS_i \end{pmatrix} \quad (29)$$

where \mathbf{FS}_i is the i^{th} vector of the feature selection set and i is the assigned corresponding weight of \mathbf{FS}_i .

whereas, the C4.5 classifier built on the obtained feature subsets by MRFES were combined into a stronger classifier that was applied to classify Australian datasets well. Hence, the superiority of MRFES was demonstrated on these large and complex datasets with multiple relevant feature sources.

In general, we can observe that the classification systems employing MRFES as a feature selection method have quite often led to superior classification accuracy. In the few cases in which MRFES was not the best, it still outperformed most algorithms, which is in agreement with our analysis. Specifically, MRF was almost always inferior to MRFES, and this is not surprising because that MRFES can accurately

capture uncertainties associated with the multiple relevant features, and can eliminate most irrelevant feature sets present in original datasets. MRFES excels in terms of classification efficiency and accuracy, along with much smaller feature subsets.

To conclude, as shown in Figs. 4 and 5, the proposed MRFES algorithm can achieve close to the best feature values. This finding shows that better classification performance can be reached by adopting the MCCM model for feature ensemble selection of large-scale datasets, and reaching a mutual agreement among the decision group of co-evolutionary memplexes. Hence, MRFES can achieve the

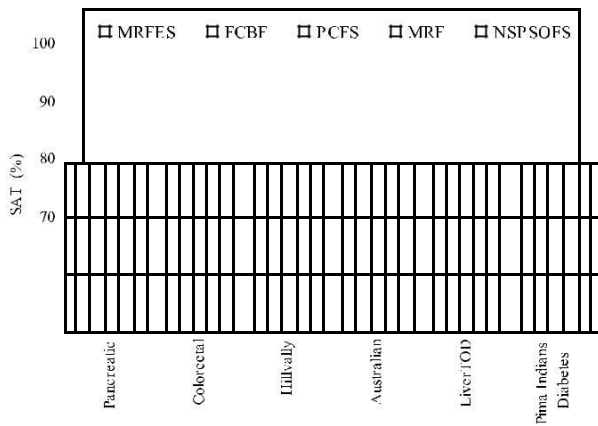


Fig. 6. Tradeoff comparison between stability and accuracy using SVM classifier.

highest classification accuracy in most the cases, irrespective of datasets and classifiers being used.

To further evaluate the tradeoff between classification accuracy and stability of the feature selection algorithm, the stability-accuracy tradeoff (SAT) is defined as

$$\text{SAT} = \frac{\#}{2 \cdot \text{stability} \cdot \text{accuracy}} \quad (30)$$

$$\text{stability} + \text{accuracy}$$

where the stability can be computed as in [48], while accuracy is evaluated using the classification results based on the selected features. SAT results for the SVM classifier are displayed in Fig. 6. It can be observed that MRFES can provide the better tradeoff between classification accuracy and stability, compared with the traditional FS algorithms.

B. Robustness Evaluation of MRFES on Two Classifiers

Despite its appealing performance in the accuracy-oriented classification system, varying noise levels will affect the performance of MRFES. Here, we employed two benchmark datasets coming from the UCI repository, such as Pima Indians Diabetes and Breast Cancer Wisconsin, to test the robustness performance of MRFES with the SVM classifier and C4.5 classifier, respectively. To obtain noisy classes, we randomly drew $k\%$ objects from two datasets and their decision labels were replaced by arbitrary candidates, where $k = 4, 8, 12, 16, 20,$ and 24 . These k objects were considered as the noisy ones and they were put back into the two datasets. With these k noise levels, the variation of classification accuracy is reported in Figs. 7 and 8 for the SVM and C4.5 classifiers, respectively. Note that the x -axis represents the noise level and the y -axis represents the classification accuracy variation by MRFES and some traditional FS algorithms. By looking at the curves in Figs. 7 and 8, it can be concluded that with increasing noise, the classification accuracy variation of MRFES became smaller than that of the traditional algorithms. It displayed a general downward trend in stability. We also see that the classification accuracy of the traditional algorithms decreased faster than that of MRFES. This indicates that the different levels of noise will degrade the classification performance. As can be seen in Figs. 7 and 8, however, the value

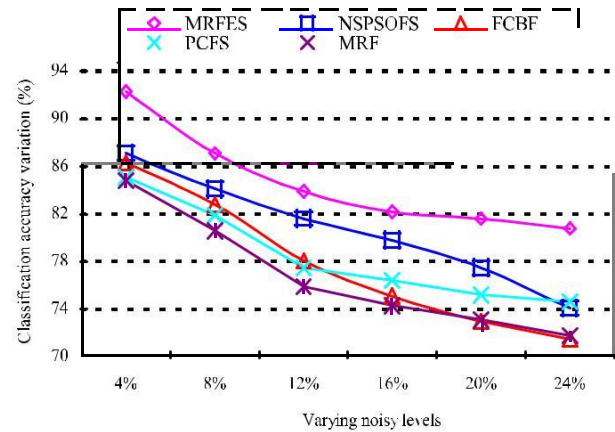


Fig. 7. Comparison of classification accuracy variation at varying noise levels on Pima Indians Diabetes with SVM classifier.

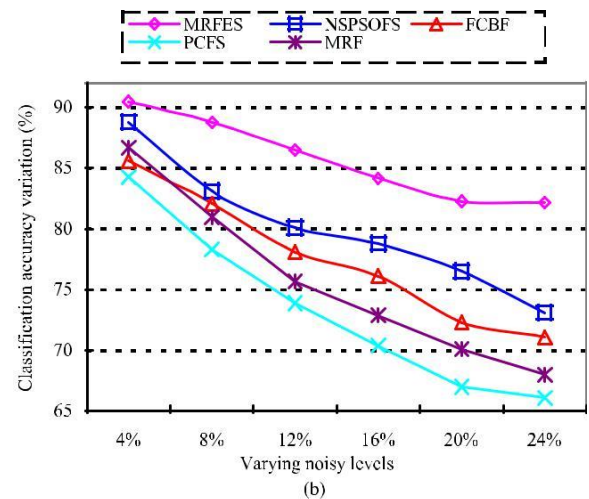


Fig. 8. Comparison of classification accuracy variation at varying noise levels on Breast Cancer Wisconsin with C4.5 classifier.

variation of MRFES is small in most cases. More important, its variation curves tend to achieve stability with increasing noise.

This is because that MRFES adopts UCAND solutions to achieve an adaptive balance between the local solutions and global dominance solutions, and it further guarantees the choice of an optimal feature subset, including strongly multiple relevant and nonredundant features. This allows accurate computation for the approximation space and reaches a mutual agreement among a group of co-evolutionary decisions for feature ensemble selection of large-scale datasets.

Hence, this test demonstrates that MRFES can select most compact feature subsets in a stable manner, with performance comparable to its supervised counterparts. This algorithm is more robust to noise than traditional algorithms. Hence, MRFES should be considered a powerful algorithm to deal with the uncertainty of large-scale datasets with varying noisy levels.

VI. APPLICATION IN HUMAN CEREBRAL CORTEX-BASED CLASSIFICATION

Health care is an important area with significant applications for big data, and successful applications of feature

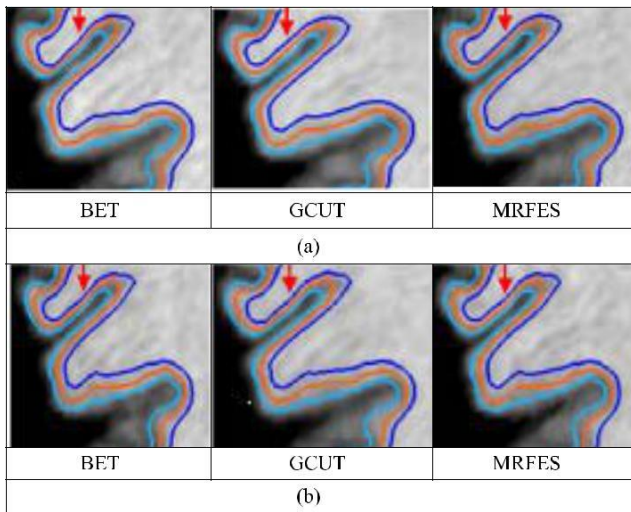


Fig. 9. Segmented longitudinal cortical surface by MRFES and compared algorithms. (a) Infant subject 1 with additional 8% Gaussian noise. (b) Infant subject 2 with additional 10% Gaussian noise.

selection algorithms are expected to dramatically scale up their efficiency and feasibility for large and complex big data. The human cerebral cortex is a very thin and highly folded sheet of gray matter, which is a typical kind of large-scale dataset with multiple relevant feature sources. Accurate and consistent classification of cortical surfaces from longitudinal human brain magnetic resonance images (MRIs) is highly important to the forecasting, diagnosis, and treatment of brain diseases [49]. Although several efforts have been made toward their classification, application of the existing methods to classify cortical surfaces may produce longitudinally inconsistent cortical surfaces because of the inconsistency of skull stripping and the evolution of surface tessellation [50]–[52]. Because longitudinal cortical changes in a short time usually are subtle, especially for infants at birth months, this requires much more accurate and consistent cortical surface classification and representation. This section investigates the proposed MRFES algorithm on consistent classification of deep buried sulci and gyri in some longitudinal images, which is important for the prediction of brain disease.

Fig. 9 shows the longitudinal surface-labeling boundaries of aligned longitudinal spherical surfaces for two typical infant subjects with longitudinal inner (blue curves), central (orange curves), and outer (light blue curves), where red arrows indicate selected cortical thinning regions. To demonstrate the consistency of longitudinal surface labeling results of cerebral cortex, we deliberately made the cortex thicker in two subjects to account for cortical evolution. As can be seen, MRFES achieved the more consistent labeling boundaries than two popular and well-established methods, namely BET [53] and GCUT [54]. We computed the average value of the symmetric boundary distance of labeling regions between each pair of aligned longitudinal surfaces. The average boundary distances are 0.60 ± 0.15 mm (MRFES), 0.71 ± 0.03 mm (BET), and 0.69 ± 0.25 mm (GCUT), respectively. MRFES significantly outperformed the other two methods and achieved a much lower boundary distance. It can be

seen that MRFES can be adapted to derive from the cortical folding surfaces, and can improve the reliability in cases in which the same cortical surfaces are used to train the classifier. Although the edges of different organizations of the infant cortical surface are fuzzy, and some nonbrain regions may be easily mistaken for value brain regions, MRFES can substantially improve the accuracy and robustness of cortical surface classification, with keeping the consistent details of complex human cerebral regions.

To further verify the effectiveness of MRFES, we evaluated the performance of MRFES against BET [53] and GCUT [54] for hemispheric vertex position asymmetries, selecting ten cerebral cortices of healthy infants, each with serial MRIs acquired at 4~8 birth months. Fig. 10(a) and (d) describes the average cortical surfaces of left hemispheres, and Fig. 10(b) and (e) describes the average cortical surfaces of mirror-flipped right hemispheres. Fig. 10(c) and (f) shows the significant clusters (nonwhite colors) of vertex position asymmetries. On the lateral surfaces in Fig. 10(a) and (b), MRFES can consistently identify the temporal lobe and inferior parietal cortex as regions with significant asymmetries. It has consistently larger sizes of significant clusters than those obtained BET and GCUT.

Prominent asymmetries also are consistently obtained around the supramarginal gyrus. On the medial surface in Fig. 10(d) and (e), by using MRFES, both the parieto-occipital sulcus and medial orbital frontal cortices are consistently identified with significant asymmetries. BET and GCUT, however, do not always find prominent asymmetries around the cerebral cortices.

From Fig. 10, it can be seen that MRFES significantly improves GCUT. BSE has about the same performance as GCUT, but it is not better than MRFES. MRFES can maintain good recognition performance of hemispheric vertex position asymmetries of cerebral cortex.

The performances of GCUT and BSE become increasingly unsatisfactory, however, when the magnitude of the deformation field increases from the left hemisphere surface to the mirror-flipped right hemisphere surface. This causes the evolved curves to be trapped between the ground-truth curves and constraint curves. This improvement by MRFES is mainly because that the better initial segmentation is conducive to more accurate alignment between left hemispheres and mirror-flipped right hemispheres, and this allows for more accurate segmentation guidance for hemispheric vertex position asymmetries.

To quantitatively evaluate the consistency of classification accuracy of MRFES on the publicly available datasets, we compared MRFES with more popular, publicly available methods—namely BET [53], GCUT [54], LPG-PCA [55], AFCM [56], and LongSeg [57], for brain classification accuracy within neonate, infant, and child groups. We compared the automated classification regions with manual classification regions by using the voxel-based Dice similarity coefficient (DSC), defined as $S(X, Y) = [2|X \cap Y|]/(|X| + |Y|)$. This is the most spread performance metric encountered in the segmentation study. Fig. 11 displays the average DSC values of three age-groups using MRFES. Their average values are

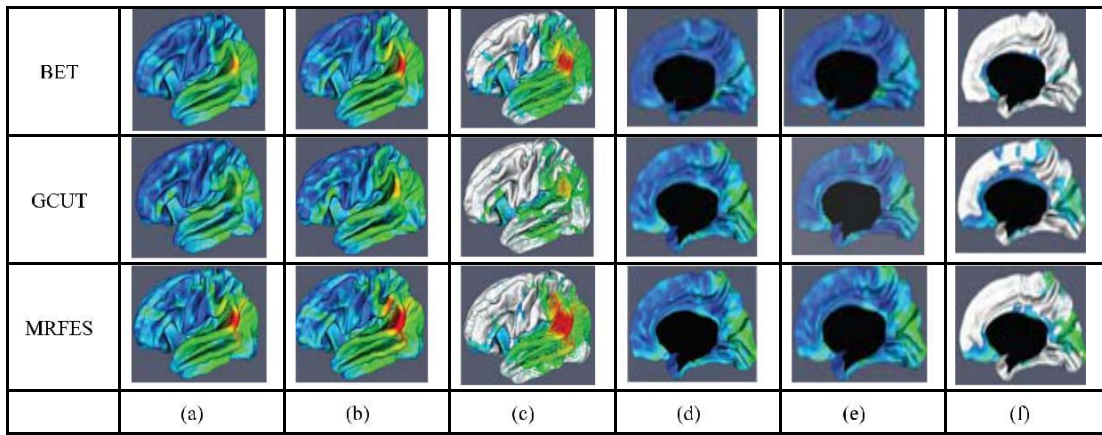


Fig. 10. Reconstructed hemispheric vertex position asymmetries of cerebral cortex. (a)–(c) Lateral surfaces. (d)–(f) Medial view surfaces.

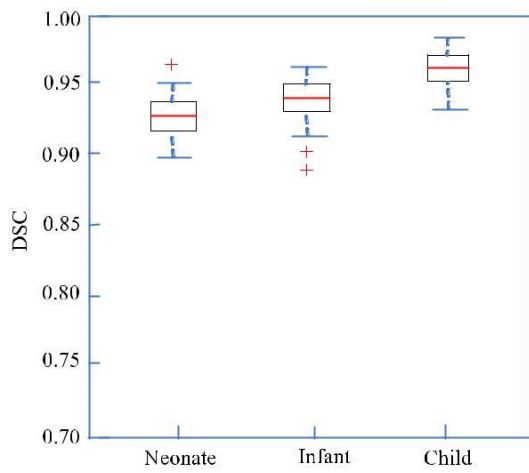


Fig. 11. Average DSC of MRFES on three age-groups.

0.925 for neonate-group, 0.939 for infant-group, and 0.960 for child-group, respectively. The DSC of the child group is higher than those of other two groups.

This finding may be a result of the fact that not only that can child brains release some useful signals to identify the nonbrain tissues in larger external brain spaces, but also that MRFES was originally developed for adults, and thus performed better on child-group brains. So, according to the MCCM model, MRFES can find the brain tissue’s contour with the highest likelihood

Fig. 12(a) and (b) shows the false-positive rate (FPR) and false-negative rate (FNR) for MRFES and five popular algorithms on 100 neonatal group subjects. The boundaries of neonatal MRIs may be quite fuzzy in some locations, especially central cortical regions, which show the low contrast and dynamic intensity changes in the early development of the neonatal brain. As indicated in Fig. 12, BET performs slightly better than LPG-PCA because it combines morpho-logical operation with edge detection to locate the brain boundaries. Both figures show well-removed nonbrain tissues, as indicated by the lower FPR in Fig. 12(a), and show rel-atively large incorrectly removed brain, as seen in the high FNR in Fig. 12(b). This result mainly is due to the fact that

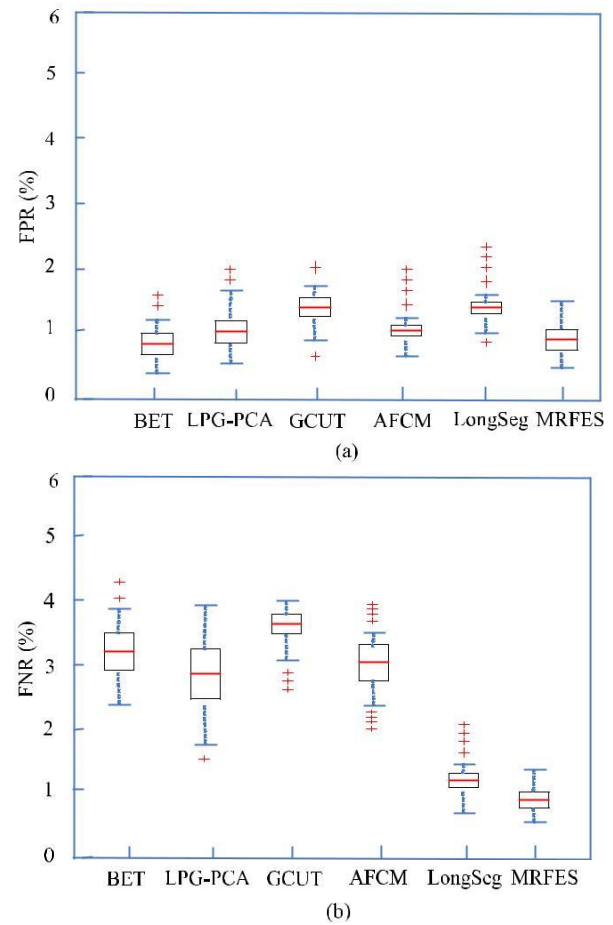


Fig. 12. Boxplot comparisons of (a) FPR and (b) FNR for MRFES and five popular algorithms.

they sometimes fail to burn some bridges so that chunks of skull cannot completely disconnect from brain boundaries. So, both tend to over-segment brain boundaries. Similar to BET, GCUT shows no significant difference in accuracy, but lower FPR and higher FNR, which causes the lowest average accuracy. AFCM shows better potential performance than GCUT by such measures as accuracy, FPR, and FNR, but it usu-ally retains a large amount of nonbrain tissue in final results, which leads to the lowest average accuracy. LongSeg indeed

TABLE I
QUANTITY COMPARISON OF DIFFERENT ALGORITHMS ABOUT DSC, FPR, AND FNR

Methods	Neonate			Infant			Child		
	DSC	FPR %	FNR %	DSC	FPR %	FNR %	DSC	FPR %	FNR %
BET	0.898	0.721	3.721	0.905	0.409	3.891	0.921	0.231	3.219
BSE	0.903	0.677	4.678	0.909	0.424	3.411	0.914	0.221	2.178
GCUT	0.918	0.691	2.237	0.895	0.587	2.708	0.929	0.412	0.668
AFCM	0.927	0.425	2.890	0.922	0.409	3.571	0.927	1.009	0.765
LongSeg	0.913	0.329	2.381	0.918	0.347	4.561	0.922	0.549	1.209
MRFES	0.936	0.258	1.921	0.945	0.256	3.101	0.953	0.205	0.649

eliminates most of false positives, but MRFES achieves the deformable surface of brain boundaries significantly and outperforms five popular methods in accuracy, generally with balanced FPR and FNR. In addition, the five popular methods also produce a number of outliers, especially AFCM and LongSeg, as indicated in Fig. 12, in which the cerebellum is over-segmented. But MRFES displays few outliers, and it produces robust results in the publicly available testing datasets.

In the following, we quantitatively calculated the average DSC, FPR, and FNR by MRFES and five popular methods, with results are summarized in Table I, and the best result in each column is marked in bold. According to experimental results, MRFES typically shows the highest quantity and outperforms five popular methods in all three age-groups by achieving an average DSC of 0.936 in neonatal group, 0.945 in infant group, and 0.953 in child group. For infant groups, MRFES fails to obtain the optimal FNR, but its performance is close to the best achieved by GCUT. However, GCUT has the lowest FNR in infant age groups at the cost of the high-est FPR. The results show that MRFES provides significantly improved performance measures in the neonate, infant, and child groups. Hence, MRFES can guarantee the classification of different tissues to be a plausible brain shape, for the better classification precision and robustness. These results provide encouraging evidence, indicating that MRFES can find more profitable longitudinal dynamic cortical surfaces for the neonate, infant, and child groups.

As shown previously, both qualitative and quantitative evaluation results demonstrate that MRFES can combine an MCCM model and UCAND solutions, to reach the temporal consistency term adapted to the temporal similarities of the cortical folding. It achieves significantly improved classification performance for complex human cerebral cortexes, and the detail preservation shows more robustness. Extensive experimental comparative studies confirm that MRFES outperforms five popular methods in terms of efficiency, accuracy, and stability.

VII. ANALYSIS AND DISCUSSION

In RST, the size of the boundary region affects the feature selection process because it reflects the characterizing ability

of available condition features. The larger the boundary region, the weaker the feature selection performance. Compared with tradition algorithms, MRFES can locate small boundary regions well, which indicates that MRFES can greatly improve the feature selection process. The results show that MRFES has potential for some application scenarios, especially cere-bral cortex-based classification. The strategy of the UCAND solutions ensures that the underlying similarity between any pair of feature points in the same feature decision class can be fully reflected, and results in better generalization abilities in the same neighborhood radius vector. Thus, MRFES is preferred to some other methods.

The applicability of MRFES has been shown on large datasets and has been evaluated by various statistical measures. Overall, its performance is comparable and satisfactory. We have demonstrated its suitability for classification and selection compared with four traditional feature selection algorithms, and it is efficient in terms of computational complexity. This paper has provided a fundamental way to explore an MCCM model for feature ensemble selection of large-scale datasets with the multiple relevant feature sources.

The proposed unified consistency dominance strategy to achieve the adaptive balance between local solutions and global dominance solutions by co-evolutionary memplexes is more informative compared with other methods, which will further guarantee that it selects out the optimal feature subsets including strongly multiple relevant and nonredundant features. We tested MRFES for various complex large datasets and real human cerebral cortex MRIs. The results were very encouraging and surpassed those of other state-of-the-art methods.

The experimental results also revealed that different feature selection algorithms display distinctive characteristics. FCBF can consistently deliver compact feature subsets. PCFS obtains high classification accuracy, but with larger sized feature subsets. NSPSOFS can achieve a proper balance between feature selection accuracy and dimensionality reduction, but with only a few large solutions by occasion. In addition, FCBF and PCFS do not always yield better feature subsets because remaining redundant subsets through feature selection may decrease overall performance. But MRFES significantly outperforms

its counterparts, making it an efficient and flexible model for feature selection and classification in large-scale datasets.

It is evident that the large number of large datasets requires substantial time feature selection, whose process relies on the efficacy of some operators of co-evolutionary consensus MapReduce. For extremely large datasets, it may be beneficial to choose the MCCM scheme, reducing time complexity compared with traditional algorithms.

Nevertheless, there are some limitations when using MRFES in large datasets with multiple relevant feature sources.

First, we ignore the missing data in original feature sets, because the integrity of data may reduce the efficiency of feature ensemble selection. Therefore, it is necessary to estimate missing values to facilitate the enough detection of much more noncooperative co-evolutionary behaviors to maintain the ensemble consensus co-evolution of memplexes. Therefore, it is expected that a more adaptive and robust consensus scheme must be designed for multiple-relevance-feature ensemble selection.

Second, we adopt the strategy of unified consistency aggregation for Nash dominance, which helps to achieve adaptive balance between the local solutions and global dominance solutions for feature ensemble selection by the co-evolutionary memplexes. This could provide more robust feature selection sets. This strategy will further guarantee the choice of optimal feature subsets, including strongly multiple-relevant and nonredundant features. This strategy can adaptively produce feature sets by avoiding partial overcrowding as well as by guiding the search toward different directions in the archive.

Third, some performance parameters of MRFES are empirically determined by preliminary repeated trials. A comparatively good solution is to attempt a multiobjective optimization algorithm to find the optimal settings of these parameters and then to generate a more profitable co-evolutionary consensus MapReduce strategy.

VIII. CONCLUSION

In this paper, we have proposed a novel MRFES algorithm. It not only constructed an MCCM model for feature selection of large scale datasets with multiple relevant feature sources and but also established a unified consistency framework between the local solutions and global dominance solutions achieved by the co-evolutionary memplexes that participate in the cooperative feature ensemble selection process. This MCCM model attempted to reach a mutual decision agreement among a group of co-evolutionary memplexes. A series of detailed experiments illustrated the thorough complexity analysis of MRFES, including aspects of the accuracy, efficiency, and robustness. This paper demonstrated the application of MRFES to human cerebral cortex-based classification in complex brain analysis, where it was shown to perform well and provide high classification accuracy to extract effectively good features.

Although be in promising, much can be done to further realize the potential of this approach. With the rapid development of the Internet of Intelligent Things and

advanced electroencephalography sensing technology, we plan to enhance MRFES by using personalized adaptive architecture to predict human cognitive states assessments from sensed brain activity and other physiological signals.

ACKNOWLEDGMENT

The authors would like to express their sincere appreciation to the anonymous reviewers for their insightful comments, which greatly improved the quality of this paper.

REFERENCES

- [1] A. Wright, "Big data meets big science," *Commun. ACM*, vol. 57, no. 7, pp. 13–15, Jul. 2014.
- [2] X. D. Wu *et al.*, "Knowledge engineering with big data," *IEEE Intell. Syst.*, vol. 30, no. 5, pp. 46–55, Sep./Oct. 2015.
- [3] D. Kumar *et al.*, "A hybrid approach to clustering in big data," *IEEE Trans. Cybern.*, vol. 46, no. 10, pp. 2372–2385, Oct. 2015.
- [4] T.-M. Choi, H. K. Chan, and X. H. Yue, "Recent development in big data analytics for business operations and risk management," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 81–92, Jan. 2017.
- [5] D. S. Wu and J. R. Birge, "Risk intelligence in big data era: A review and introduction to special issue," *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1718–1720, Aug. 2016.
- [6] Z.-H. Zhou, N. V. Chawla, Y. C. Jin, and G. J. Williams, "Big data opportunities and challenges: Discussions from data analytics perspectives," *IEEE Comput. Intell. Mag.*, vol. 9, no. 4, pp. 62–74, Nov. 2014.
- [7] J. Y. Liang, F. Wang, C. Y. Dang, and Y. H. Qian, "A group incremental approach to feature selection applying rough set technique," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 2, pp. 294–308, Feb. 2014.
- [8] A. Phophalia, A. Rajwade, and S. K. Mitra, "Rough set based image denoising for brain MR images," *Signal Process.*, vol. 103, no. 10, pp. 24–35, Oct. 2014.
- [9] W. Ding, J. Wang, Z. Guan, and Q. Shi, "Ensemble feature selection integrating elitist roles based quantum game algorithm," *J. Syst. Eng. Electron.*, vol. 26, no. 3, pp. 584–594, Jun. 2015.
- [10] A. Amin *et al.*, "Customer churn prediction in the telecommunication sector using a rough set approach," *Neurocomputing*, vol. 237, pp. 242–254, May 2017.
- [11] R. Jensen and Q. Shen, "Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 12, pp. 1457–1471, Dec. 2004.
- [12] S. Zhao, H. Chen, C. Li, X. Du, and H. Sun, "A novel approach to building a robust fuzzy rough classifier," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 4, pp. 769–786, Aug. 2015.
- [13] W. W. Li, Z. Q. Huang, X. Y. Jia, and X. Y. Cai, "Neighborhood based decision-theoretic rough set models," *Int. J. Approx. Reason.*, vol. 69, pp. 1–17, Feb. 2016.
- [14] Y. Yao and Y. She, "Rough set models in multigranulation spaces," *Inf. Sci.*, vol. 327, pp. 40–56, Jan. 2016.
- [15] M. A. Potter and K. A. De Jong, "Cooperative coevolution: An architecture for evolving coadapted subcomponents," *Evol. Comput.*, vol. 8, no. 1, pp. 1–29, Mar. 2000.
- [16] X. Li and X. Yao, "Cooperatively coevolving particle swarms for large scale optimization," *IEEE Trans. Evol. Comput.*, vol. 16, no. 2, pp. 210–224, Apr. 2012.
- [17] T. Jansen and R. P. Wiegand, "The cooperative coevolutionary (1+1) EA," *Evol. Comput.*, vol. 12, no. 4, pp. 405–434, 2004.
- [18] L. Sun, S. Yoshida, X. Cheng, and Y. Liang, "A cooperative particle swarm optimizer with statistical variable interdependence learning," *Inf. Sci.*, vol. 186, no. 1, pp. 20–39, Mar. 2012.
- [19] J. Fan, J. Wang, and M. Han, "Cooperative coevolution for large-scale optimization based on kernel fuzzy clustering and variable trust region methods," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 4, pp. 829–839, Aug. 2014.
- [20] N. R. Sabar, J. Abawajy, and J. Yearwood, "Heterogeneous cooperative co-evolution memetic differential evolution algorithm for big data optimization problems," *IEEE Trans. Evol. Comput.*, vol. 21, no. 2, pp. 315–327, Apr. 2017.
- [21] X.-M. Hu, F.-L. He, W.-N. Chen, and J. Zhang, "Cooperation coevolution with fast interdependency identification for large scale optimization," *Inf. Sci.*, vol. 381, pp. 142–160, Mar. 2017.

- [22] W. Fang, B. He, Q. Luo, and N. K. Govindaraju, "Mars: Accelerating MapReduce with graphics processors," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 4, pp. 608–620, Apr. 2011.
- [23] K. Guo, W. Guo, Y. Chen, Q. Qiu, and Q. Zhang, "Community discovery by propagating local and global information based on the MapReduce model," *Inf. Sci.*, vol. 323, pp. 73–93, Dec. 2015.
- [24] M. M. Rashid, I. Gondal, and J. Kamruzzaman, "Dependable large scale behavioral patterns mining from sensor data using Hadoop platform," *Inf. Sci.*, vol. 379, pp. 128–145, Feb. 2017.
- [25] Y. Yao and Y. Zhao, "Discernibility matrix simplification for constructing attribute reducts," *Inf. Sci.*, vol. 179, no. 7, pp. 867–882, Mar. 2009.
- [26] D. Chen, S. Zhao, L. Zhang, Y. Yang, and X. Zhang, "Sample pair selection for attribute reduction with rough set," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 11, pp. 2080–2093, Nov. 2012.
- [27] Y.-C. Chen, N. R. Pal, and I. F. Chung, "An integrated mechanism for feature selection and fuzzy rule extraction for classification," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 4, pp. 683–698, Aug. 2012.
- [28] H. Y. Zhang and S. Y. Yang, "Feature selection and approximate reasoning of large-scale set-valued decision tables based on α -dominance-based quantitative rough sets," *Inf. Sci.*, vol. 378, pp. 328–347, Feb. 2017.
- [29] H. Shen, S. Yang, and J. Liu, "An attribute reduction of rough set based on PSO," in *Proc. 5th Int. Conf. Rough Set Knowl. Technol. (RSKT)*, Beijing, China, 2010, pp. 695–702.
- [30] Y. Zhang, D. Gong, Y. Hu, and W. Zhang, "Feature selection algorithm based on bare bones particle swarm optimization," *Neurocomputing*, vol. 148, no. 1, pp. 150–157, Jan. 2015.
- [31] Y. Zhang, D.-W. Gong, and J. Cheng, "Multi-objective particle swarm optimization approach for cost-based feature selection in classification," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 14, no. 1, pp. 64–75, Jan./Feb. 2017.
- [32] S. Fong, R. Wong, and A. V. Vasilakos, "Accelerated PSO swarm search feature selection for data stream mining big data," *IEEE Trans. Services Comput.*, vol. 9, no. 1, pp. 33–45, Jan./Feb. 2016.
- [33] B. Tran, B. Xue, and M. Zhang, "A new representation in PSO for discretization-based feature selection," *IEEE Trans. Cybern.*, vol. 48, no. 6, pp. 1733–1746, Jun. 2018.
- [34] W. Ding, Z. Guan, Q. Shi, and J. Wang, "A more efficient attribute self-adaptive co-evolutionary reduction algorithm by combining quantum elitist frogs and cloud model operators," *Inf. Sci.*, vol. 293, pp. 214–234, Feb. 2015.
- [35] C. Wang, M. Shao, Q. He, Y. Qian, and Y. Qi, "Feature subset selection based on fuzzy neighborhood rough sets," *Knowl. Based Syst.*, vol. 111, pp. 173–179, Nov. 2016.
- [36] F. Jiang, Y. Sui, and L. Zhou, "A relative decision entropy-based feature selection approach," *Pattern Recognit.*, vol. 48, no. 7, pp. 2151–2163, Jul. 2015.
- [37] Y. Zhang, X.-F. Song, and D.-W. Gong, "A return-cost-based binary fire-fly algorithm for feature selection," *Inf. Sci.*, vols. 418–419, pp. 561–574, Dec. 2017.
- [38] W. Ding, Z. Guan, J. Wang, and D. Tian, "A layered co-evolution based rough feature selection using adaptive neighborhood radius hier-archy and its application in 3D-MRI," *Chin. J. Electron.*, vol. 26, no. 6, pp. 1168–1176, Dec. 2017.
- [39] J. B. Zhang, T. R. Li, D. Ruan, Z. Z. Gao, and C. B. Zhao, "A parallel method for computing rough set approximations," *Inf. Sci.*, vol. 194, pp. 209–223, Jul. 2012.
- [40] J. B. Zhang, J. S. Wong, T. R. Li, and Y. Pan, "A comparison of parallel large-scale knowledge acquisition using rough set theory on different MapReduce runtime systems," *Int. J. Approx. Reason.*, vol. 55, no. 3, pp. 896–907, Mar. 2014.
- [41] J. Qian, D. Q. Miao, Z. H. Zhang, and X. D. Yue, "Parallel attribute reduction algorithms using MapReduce," *Inf. Sci.*, vol. 279, pp. 671–690, Sep. 2014.
- [42] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, no. 12, pp. 1205–1224, Dec. 2004.
- [43] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artif. Intell.*, vol. 151, nos. 1–2, pp. 155–176, Dec. 2013.
- [44] Q. Cheng, H. Zhou, and J. Cheng, "The Fisher–Markov selector: Fast selecting maximally separable feature subset for multiclass classification with applications to high-dimensional data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1217–1233, Jun. 2011.
- [45] B. Xun, M. J. Zhang, and W. N. Browne, "Particle swarm optimization for feature selection in classification: A multi-objective approach," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1656–1671, Dec. 2013.
- [46] A. Frank and A. Asuncion. (2015). *UCI Machine Learning Repository*. [Online]. Available: <https://archive.ics.uci.edu/ml/index.html>
- [47] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, Jan. 2008.
- [48] Y. Li, J. N. Si, G. J. Zhou, S. Huang, and S. Chen, "FREL: A stable feature selection algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 7, pp. 1388–1401, Jul. 2015.
- [49] S. Abe, K. Takagi, T. Yamamoto, Y. Okuhata, and T. Kato, "Assessment of cortical gyrus and sulcus formation using MR images in normal fetuses," *Prenatal Diagn.*, vol. 23, no. 3, pp. 225–231, Mar. 2003.
- [50] N. K. Kasabov, M. G. Doborjeh, and Z. G. Doborjeh, "Mapping, learning, visualization, classification, and understanding of fMRI data in the NeuCube evolving spatiotemporal data machine of spiking neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 887–899, Apr. 2017.
- [51] J. E. Iglesias, C. Y. Liu, P. Thompson, and Z. W. Tu, "Robust brain extraction across datasets and comparison with publicly available methods," *IEEE Trans. Med. Imag.*, vol. 30, no. 9, pp. 1617–1634, Sep. 2011.
- [52] L. Wang *et al.*, "Segmentation of neonatal brain MR images using patch-driven level sets," *NeuroImage*, vol. 84, pp. 141–158, Aug. 2014.
- [53] S. Smith, "Fast robust automated brain extraction," *Human Brain Map.*, vol. 17, no. 3, pp. 143–155, Nov. 2002.
- [54] S. Sadanathan, W. Zheng, M. Chee, and V. Zagorodnov, "Skull stripping using graph cuts," *NeuroImage*, vol. 49, no. 1, pp. 225–239, Jan. 2010.
- [55] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognit.*, vol. 43, no. 4, pp. 1531–1549, Apr. 2010.
- [56] C. Xu, D. L. Pham, M. E. Rettmann, D. N. Yu, and J. L. Prince, "Reconstruction of the human cerebral cortex from magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 18, no. 6, pp. 467–480, Jun. 1999.
- [57] F. Shi *et al.*, "Neonatal brain image segmentation in longitudinal MRI studies," *NeuroImage*, vol. 49, no. 1, pp. 391–400, Aug. 2010.



Weiping Ding (M'16) received the Ph.D. degree in computation application from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013.

He was a Visiting Researcher with the University of Lethbridge, Lethbridge, AB, Canada, in 2011. From 2014 to 2015, he was a Post-Doctoral Researcher with the Brain Research Center, National Chiao Tung University, Hsinchu, Taiwan. In 2016, he was a Visiting Scholar with the National University of Singapore, Singapore. He was a Visiting Scholar with the University of Technology Sydney, Ultimo, NSW, Australia, from 2017 to 2018. He is currently an Associate Professor with Nantong University, Nantong, China. He has published over 50 papers in flagship journals and conference proceedings as the first author. He holds 10 approved invention patents in total over 18 issued patents. His current research interests include data mining, machine learning, and granular computing.

Dr. Ding was a recipient of Computer Education Excellent Paper Award (First-Prize) from the National Computer Education Committee of China, in 2009, the Excellent-Young Teacher (Qing Lan Project) of Jiangsu Province in 2014, the High-Level Talent (Six Talent Peak) of Jiangsu Province in 2016, the Best Paper of ICDMA'15, the Outstanding Teacher of Software Design and Entrepreneurship Competition by the Ministry of Industry and Information Technology, China, in 2017, and two Chinese Government Scholarships for Overseas Studies in 2011 and 2016. He served/serves as an Associate Editor for the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Information Sciences*, *Swarm and Evolutionary Computation*, and IEEE ACCESS. He is a member of IEEE, ACM and Senior CCF.



Chin-Teng Lin (F'05) received the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989 and 1992, respectively.

He is currently the Distinguished Professor with the Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW, Australia, a University Chair Professor of electrical and computer engineering with National Chiao Tung University, Hsinchu, Taiwan, the International Faculty of the University of

California at San Diego, San Diego, CA, USA, and an Honorary Professorship of the University of Nottingham, Nottingham, U.K. He has published over 200 journal papers (total citation: 19 166, *H*-index: 53, and *i10*-index: 332) in the areas of fuzzy systems, neural networks, and cognitive neuroengineering, including approximately 109 IEEE journal papers.

Dr. Lin was elected as the Editor-in-Chief of the IEEE TRANSACTIONS ON FUZZY SYSTEMS from 2011 to 2016. He also served on the Board of Governors at IEEE Circuits and Systems Society from 2005 to 2008, IEEE Systems, Man, Cybernetics Society from 2003 to 2005, and IEEE Computational Intelligence Society (CIS) from 2008 to 2010. He was the Distinguished Lecturer of IEEE CIS Society from 2015 to 2017. He served as the Deputy Editor-in-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: EXPRESS BRIEFS from 2006 to 2008. He was elevated to be an IEEE Fellow for his contributions to biologically inspired information systems in 2005, and was elevated to be an International Fuzzy Systems Association Fellow in 2012.



Witold Pedrycz (F'99) received the M.Sc., Ph.D., and D.Sci. degrees from the Silesian University of Technology, Gliwice, Poland.

He is a Professor and the Canada Research Chair in computational intelligence with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland. He holds an appointment of special professorship with the School of Computer Science, University of Nottingham,

Nottingham, U.K. He has also authored 15 research monographs covering various aspects of computational intelligence, data mining, and software engineering. His current research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering. He has published numerous papers in the above areas.

Dr. Pedrycz was a recipient of the prestigious Norbert Wiener Award from the IEEE Systems, Man, and Cybernetics Society in 2007, the IEEE Canada Computer Engineering Medal 2008, the Cajastur Prize for Soft Computing from the European Centre for Soft Computing for "pioneering and multifaceted contributions to granular computing" in 2009, the Killam Prize in 2013, and the Fuzzy Pioneer Award 2013 from the IEEE Computational Intelligence Society. He was elected as a Foreign Member of the Polish Academy of Sciences in 2009 and a fellow of the Royal Society of Canada in 2012. He has been a member of numerous program committees of the IEEE conferences in the area of fuzzy sets and neurocomputing. He is intensively involved in editorial activities. He is the Editor-in-Chief of *Information Sciences*, and *WIREs Data Mining and Knowledge Discovery* (Wiley). He serves on an Advisory Board of the IEEE TRANSACTIONS ON FUZZY SYSTEMS and is a member of a number of editorial boards of other international journals.