

“© 2016 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.”

Attribute Equilibrium Dominance Reduction Accelerator (DCCAEDR) Based on Distributed Co-evolutionary Cloud and Its Application in Medical Records

Wei-Ping Ding, Chin-Teng Lin, *Fellow, IEEE*, Mukesh Prasad, *Member, IEEE*, Sen-Bo Chen, and Zhi-Jin Guan

Abstract—Aimed at the tremendous challenge of attribute reduction for big data mining and knowledge discovery, we propose a new attribute equilibrium dominance reduction accelerator (DCCAEDR) based on the distributed co-evolutionary cloud model. First, the framework of N -populations distributed co-evolutionary MapReduce model is designed to divide the entire population into N subpopulations, sharing the rewards of different subpopulations' solutions under a MapReduce cloud mechanism. Because the adaptive balancing between exploration and exploitation can be achieved in a better way, the reduction performance is guaranteed to be the same as those using the whole independent dataset. Second, a novel Nash equilibrium dominance strategy of elitists under the N bounded rationality regions are adopted to assist the subpopulations necessary to attain the stable status of Nash equilibrium dominance. This further enhances the accelerator's robustness against complex noise on big data. Third, the approximation parallelism mechanism based on MapReduce is constructed to implement rule reduction by accelerating the computation of attribute equivalence classes. Consequently, the entire attribute reduction set with the equilibrium dominance solution can be achieved. Extensive simulation results have been used to illustrate the effectiveness and robustness of the proposed DCCAEDR accelerator for attribute reduction on big data. Furthermore, the DCCAEDR is applied to solve attribute reduction for traditional Chinese medical records and to segment cortical surfaces of the neonatal brain 3D-MRI records, and the DCCAEDR

shows the superior competitive results, when compared with the representative algorithms.

Index Terms—Attribute reduction accelerator, bounded rationality region, distributed co-evolutionary cloud, equilibrium dominance strategy, MapReduce framework.

I. INTRODUCTION

THE increasingly complex and dynamic change in big data processing is currently an issue of widespread concern. One of primary topics for study is how to use the existing information to achieve the rapid updating knowledge. Granular computing, a new computing paradigm of information, has become an emerging research field [1]. Granular computing can provide solutions based on the granularity and the relationships between these granularities. As an important part of granular computing, rough set theory characterizes the boundaries of approximation, which is a valid mathematical tool to describe and handle imprecision, uncertainty and vagueness [2],[3]. Rough set theory has been successfully applied into several data analysis tasks [4]-[6]. The effective computation of approximations is vital to improve the performance of feature selection. Attribute reduction offers a systematic theoretic framework for consistency-based feature selection, which does not attempt to maximize the class separability, but rather aims to retain the discriminatory power of the original features. As such attribute reduction can be considered a kind of specific feature selection and can select useful features from a given data table. Thus, it allows us to acquire different features that can be used to induce different concise sets of knowledge rules.

Over recent years, attribute reduction has played an important role in the fields of machine learning, pattern recognition and data mining. However, the rapid increases and updates in big data bring a new challenge for traditional attribute reduction algorithms [7]. Despite many successful applications, the computation of most algorithms is still very expensive, and the distribution of solutions will not be uniform enough. Most algorithms do not work well when dealing with large-scale data sets with high dimensions and

Manuscript received November 23, 2014; revised May 11, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61300167, 61171132 and 61139002, the International Center of Excellence in Advanced Bio-engineering sponsored by the Taiwan National Science Council International Research-Intensive Centers of Excellence Program under Grant NSC-100-2911-I-009-101, Open Project Program of State Key Lab. for Novel Software Technology, Nanjing University under Grant KFKT2015B17, 2015 Natural Science Foundation of Jiangsu Province, Sponsored by Qing Lan Project of Jiangsu Province, and Starting Function for Doctoral Scientific Research, Nantong University under Grant 14B20.

W. P. 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 Institute of Electrical Control Engineering and Brain Research Center, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: ctlin@mail.nctu.edu.tw).

M. Prasad is with the Department of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: mukeshnctu.cs99g@nctu.edu.tw).

S. B. Chen is with College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: sam_c@163.com).

Z. J. Guan is with School of Computer Science and Technology, Nantong University, Nantong 226019, China (e-mail: guan.zj@ntu.edu.cn).

complex noise; consequently, their performance deteriorates sharply as the dimensionality of the space rapidly increases [8]. Currently, the primary obstacle of attribute reduction for big data is that traditional algorithms are often computationally time-consuming, which greatly restricts the development and practical application of the rough set theory and method. Therefore, it is necessary to thoroughly investigate some novel and effective attribute reduction algorithms to accelerate the knowledge discovery process in the real world applications of big data.

To overcome the bottleneck of performance in big data, Dean and Ghemawat [9],[10] from Google firstly presented a parallel programming MapReduce model, which was a framework for processing big data on different kinds of distributable problems. Following Google's work, Ekanayake *et al.* [11] presented the CGL-MapReduce model as a streaming-based MapReduce implementation. Berlińska *et al.* [12] also proposed a divisible load computation model and two load-partitioning algorithms; the mapping and reducing operations were adopted as two divisible applications with precedence constraints on the cloud-computing platform. The introduced MapReduce technique has received a great deal of attention from both the scientific community and industry for its applicability in big data analysis.

In this paper, by considering the outstanding superiority of the MapReduce technique on the cloud platform when dealing with the large-scale datasets, we designed the N -populations distributed co-evolutionary MapReduce model as an instantiation to solve the attribute reduction problem associated with big data. Meanwhile, a novel Nash equilibrium strategy using evolutionary game theory (EGT) was adopted to improve the optimal dominance performance of dynamic elitists. Therefore, we propose a new attribute equilibrium dominance reduction accelerator (DCCAEDR) based on the distributed co-evolutionary cloud model, and its innovations and advantages cover as follows:

- First, the framework of the N -populations distributed co-evolutionary MapReduce model is designed to decompose the entire population into N subpopulations and share their rewards by using the MapReduce mechanism. Because the adaptive balancing between exploration and exploitation can be better achieved, the reduction performance is surely guaranteed to be the same as those using the whole independent big data set.
- Second, we adopt a novel equilibrium dominance strategy of elitists under the N bounded rationality regions to assist the co-adaptive subpopulations in order to reach a stable status with the equilibrium dominance solution. This will further enhance the DCCAEDR's equilibrium robustness and stability when dealing with big data.
- Third, the MapReduce operations are constructed and implemented. The approximation parallelism rule set based on the MapReduce mechanism is accelerated in order to compute; the Pareto dominance solution sets are then attained by reducing the space cost. Thus the entire equilibrium dominance solution of attribute reduction can be achieved.

The remainder of this paper is organized as follows. First, we review some related work in Section II. Section III presents some definitions of attribute reduction and the MapReduce model. In Section IV, the distributed co-evolutionary MapReduce model is introduced and its advantage is analyzed. Subsequently, a novel Nash equilibrium dominance strategy of co-evolutionary elitists is presented in Section V. Section VI details the primary steps and flowchart of DCCAEDR. The extensive experimental evaluation of DCCAEDR is provided in Section VII, which includes a number of special applications in the traditional Chinese medical records and neonatal brain 3D-MRI records. Finally, discussions are presented and conclusions are drawn, and further research is also directed in Section VIII.

II. LITERATURE REVIEW

Attribute reduction has been widely used in data mining and machine learning. As it normally needs to deal with very large-scale data sets, a fast and efficient attribute reduction algorithm is especially important for practical applications. In recent years, a number of various attribute reduction algorithms have been proposed. These algorithms can be classified into three main categories as follows: the first is the traditional attribute reduction algorithms with the discernibility matrix and positive region; the second is attribute reduction algorithms based on evolutionary optimization; and the third is the large-scale attribute reduction algorithms based on the MapReduce technique.

Over the past few decades, traditional attribute reduction algorithms have been widely studied according to the discernibility matrix and positive region. Many algorithms have been developed using the rough set theory. Hu *et al.* [13] gave an extension of the positive region reduction for hybrid attribute reduction in the framework of the fuzzy rough set. Meng *et al.* [14] provided a relatively systematic and comprehensive study of attribute reduction in inconsistent incomplete decision systems based on three kinds of representative entropy. Wang *et al.* [15] developed a novel attribute reduction algorithm for data sets with varying data values. Qian *et al.* [16] presented a theoretic framework of positive approximations to accelerate the performance of heuristic attribute reduction algorithms. Zhao *et al.* [17] set up a fuzzy variable precision rough set model by combining the fuzzy rough set and the variable precision rough set, with the goal of establishing the fuzzy rough set as a special case. Liang *et al.* [18] proposed an efficient attribute reduction algorithm for large-scale decision tables in which a sub-table of the large-scale data set could be considered as a single, small granularity. Wang *et al.* [19] presented some basic structural properties of attribute reduction by covering rough sets, and developed a heuristic algorithm to find the attribute subset that approximates to be a minimal reduction. Ye *et al.* [20] analyzed the major relationships between attribute reduction and attribute-generalization, and put forward a heuristic algorithm for attribute-generalization reduction.

To summarize, these traditional attribute reduction

algorithms cannot be implemented very well in large-scale data sets due to the huge spatial requirements. With the dramatically increasing dimensions and numbers of attributes, the processing time for attribute reduction will grow tremendously.

Evolutionary algorithms have shown to be powerful when solving these global optimization problems. The high complexity of attribute reduction has motivated a few investigators to propose some new attribute reduction algorithms based on evolutionary optimization to find near optimal reductions. Slezak *et al.* [21] presented the attribute reduction algorithm of approximate entropy using an order-based genetic algorithm. Ke *et al.* [22] introduced the ant colony optimization approach to attribute reduction, reporting some interesting results afterward. Ye *et al.* [23] applied particle swarm optimization to attribute reduction and provided a number of competitive solutions. Hedar *et al.* [24] proposed a memory-based heuristic of Tabu search to solve the attribute reduction problem. Ding *et al.* [25] presented an enhanced minimum attribute reduction approach based on the quantum-inspired shuffled frog leaping algorithm.

However, the above algorithms often slow the premature convergence speed in the approximate attribute space, and suffer from the difficulties in balancing between exploration and exploitation for iterative optimization. Moreover, these algorithms are still too computationally time-consuming to deal with large-scale data sets. Consequently, these attribute reduction algorithms based on evolutionary optimization are unable to achieve superior high-performance for large-scale complex attributes in many real-world applications.

The ubiquity of mobile computing technology has produced an abundance of large data sets. More and more research and application areas focus on dealing with big data problems. Hence, there is a great deal of need to study the attribute reduction algorithm for large-scale data sets. The recently introduced MapReduce technique has received much attention for its applicability in big data analysis. Several researchers have proposed some attribute reduction algorithms based on the MapReduce technique. Zhang *et al.* [26] presented a parallel algorithm for computing the equivalence classes, decision classes based on the MapReduce model, that can be used to deal with the massive amount of data by updating rough set approximations. Yang *et al.* [27] designed a parallel attribute reduction algorithm based on the MapReduce for big data, which was based on the MapReduce's parallel program mode. Qian *et al.* [28] adopted the parallel computations of equivalence classes and the attribute significance for attribute reduction, and proposed a hierarchical attribute reduction algorithm by using the MapReduce model. Zhang *et al.* [29] proposed parallel large-scale rough set methods for knowledge acquisition, and implemented them on such representative MapReduce runtime systems as Hadoop, Phoenix and Twister in order to mine knowledge from big data. Qian *et al.* [30] designed a novel structure of the $\langle \text{key}, \text{value} \rangle$ pair to speed up the computation of equivalence classes and attribute significance, parallelizing the traditional attribute reduction process based on the

MapReduce mechanism.

Nevertheless, the above-mentioned attribute reduction algorithms based on MapReduce are not usually guaranteed to produce the same results as those achieved when implemented on the whole, and non-separated, large-scale datasets. These subsystems on separated datasets that have used MapReduce are independent, and they do not share any respective performance information with neighboring subsystems. Obviously, this can be detrimental to the attribute reduction process. Therefore, in most cases these algorithms are unable to acquire the exact reduction set from big data, which also limits their practical applications.

In recent years, large-scale optimization issues have been solved through cooperative co-evolution, which is the parallel implementation for multi-populations architectures [31]-[36]. In such cooperative co-evolutionary architecture, the whole population is divided into several subpopulations, each of which is in charge of optimizing a subset of global solutions. A number of the studies described so far have introduced the cooperative co-evolutionary architecture as a significant improvement when solving large-scale optimization problems through a divide-and-conquer paradigm, which are described as follows:

Yang *et al.* [37] proposed a random grouping scheme and adaptive weighting for large-scale evolutionary optimization through the use of cooperative co-evolution. Omidvar *et al.* [38] employed the delta grouping strategy to divide a greater number of interacting variables into one group for cooperative co-evolution in the large-scale optimization. Most recently, Fan *et al.* [39] introduced a cooperative co-evolutionary strategy incorporated with a kernel fuzzy C-means clustering, which was used to divide high-dimensional large-scale problems into subproblems while exploring their search spaces. Mei *et al.* [40] proposed a novel cooperative co-evolutionary approach with the route distance grouping for large-scale capacitated arc routing problems.

Although cooperative co-evolution has been proposed as a promising framework, it is ineffective when the number of interacting variables grows beyond a certain number of variables or other non-separable subcomponents in the objective function. In addition, most populations in the cooperative co-evolutionary process focus on optimizing their respective objectives without an information-sharing strategy. This may lead the individuals in each subpopulation to bring forward the profit of the corresponding sub-objectives, resulting in an inefficient approximation of the whole populations. The performance of cooperative co-evolution is also limited due to the lack of elitists and the localized perception of the Pareto optimality. Furthermore, it may also be impractical to deal with a large-scale dataset only through the existing cooperative co-evolutionary framework, which would be a particularly challenging task. Therefore, we should find a novel and efficient cooperative co-evolutionary algorithm to exploit the hidden structure of the large-scale optimization problem, and then motivate different subpopulations to share their search strategies with each other. The whole solution with the equilibrium dominance could thus

be acquired efficiently.

Because cooperative co-evolution is composed of some distributed subpopulations through the evolution of independent objects [41], we consider using the MapReduce framework to further improve the parallel performance of cooperative co-evolution. To the best of our knowledge, most of the traditional attribute reduction algorithms are only run on a single computer to deal with small data sets. So the question of how best to combine the MapReduce framework and cooperative co-evolution in order to greatly enhance the performance of attribute reduction has emerged as an urgent problem that also motivates us to investigate more deeply.

At the same time, another problem we must solve is the issue of traditional attribute reduction algorithms that are not robust against complex noise. In this paper, we introduced EGT [42] to help analyze the dynamic behavior of cooperative co-evolution for the purpose of attribute reduction. This strategy enhanced the stability and the degree of noise resistance to attribute reduction in big data. Recently, some promising theoretical research in the EGT field has been explored. For example, Ficici *et al.* [43] introduced EGT in a simple co-evolutionary algorithm to investigate the selection dynamics, and focused on symmetric variable-sum games that have the polymorphic Nash equilibrium. Liu *et al.* [44] introduced EGT in traditional particle swarm optimization to overcome the premature convergence and improve its performance. Koh [45] proposed a novel evolutionary algorithm based on Nash dominance for equilibrium problems with equilibrium constraints. Additionally Razi *et al.* [46] discussed how best to determine the Nash equilibrium point of nonlinear non-cooperative games, by using co-evolutionary strategies.

EGT can provide a better understanding of dynamic performance mechanisms in cooperative co-evolution. In most cases, the optimality of cooperative co-evolution is best expressed as a self-adaptive Nash equilibrium achieved in the co-evolutionary populations [47], which would further enhance its high efficiency to solve some difficult NP-hard optimizations.

It is worth mentioning that despite the increasing interest in cooperative co-evolution over the last decade, few extensive studies have been devoted to discussing EGT in the attribute co-evolutionary reduction. Consequently, the question of how to apply EGT into a cooperative co-evolutionary analysis of attribute reduction to achieve better Nash equilibrium has become an interesting problem. This has been extended to the better performance in both solution quality and competitive complexity for attribute reduction in complex data sets.

Enlightened by these studies, our work will focus on extending the attribute reduction algorithm to deal with big data tasks. Since EGT can be seen as the Nash equilibrium of self-adaptive balance, the extension of cooperative co-evolution incorporated into EGT can be potentially helpful for attribute reduction. If the improved MapReduce framework with cooperative co-evolution can be assured to be valid, it can be incorporated into the proposed algorithm. In this paper, we propose a new attribute equilibrium dominance

reduction accelerator (DCCAEDR) based on the distributed co-evolutionary cloud model. This work will provide some new insights into the attribute reduction problems for big data mining and knowledge discovery.

III. PRELIMINARY KNOWLEDGE

The section will provide the relevant definitions of attribute reduction, and a summary of the MapReduce model.

A. Relevant Definitions of Attribute Reduction

An approximation space $K = (U, R)$ is characterized by an information system $S = (U, A, V, f)$, where U is a non-empty finite set of objects, called a universe, A is a non-empty finite set of attributes, V equals to $\bigcup_{a \in A} V_a$, V_a is a domain of the attribute a , and f is an information function $U \times A \rightarrow V$ such that $f(x, a) \in V_a$ for each $x \in U, a \in A$.

Specifically, $S = (U, A, V, f)$ is called a decision table if $A = C \cup D$, where C is a set of condition attributes and D is a set of decision, and $C \cap D = \emptyset$.

Definition 1 (Positive Region) [1] The C -positive region of D is the set of all objects from U , and it can be classified with certainty into classes of U/D employing attributes from C . It is defined as

$$POS_C(D) = \bigcup_{X \in U / ind(D)} \underline{C}X. \quad (1)$$

Definition 2 (Degree of Dependency) [2] The degree of dependency $\gamma_C(D)$ is used as a criterion for attribute selection as well as the stop condition, defined as

$$\gamma_C(D) = \frac{|POS_C(D)|}{|U|}. \quad (2)$$

Definition 3 (Attribute Reduction) [3] For the attribute reduction of rough set, a given decision table may have many attribute reductions, and the set of all reductions is defined as

$$RED = \{R \subseteq C \mid \gamma_R(D) = \gamma_C(D), \forall B \subset R, \gamma_B(D) \neq \gamma_C(D)\}. \quad (3)$$

Definition 4 (Minimal Reduction) [5] A reduction with minimal cardinality aims to locate a single element of the minimal reduction set $RED_{\min} \subseteq RED$ as

$$RED_{\min} = \{R \in RED \mid \forall R' \in RED, |R| \leq |R'|\}. \quad (4)$$

Definition 5 (Adaptive Fitness Function) The usage of an adaptive fitness function is very important in addressing the

attribute reduction problem. A key question in the fitness function evaluation is how stable progress can be guaranteed. In the proposed DCCAEDR accelerator, the adaptive fitness function of attribute reduction with the pairwise constraints is constructed as follows:

$$Fit(x) = \min \left(\bar{\rho} \times \frac{|C(x)| - |R(x)|}{|C(x)|} + \bar{\delta} \times \frac{|Core(\xi(x))|}{\gamma_{\xi(x)}(D)} \right) + \Phi(x) \quad (5)$$

where $|C(x)|$ is the total number of attribute features, $|R(x)|$ is the length of selected attribute subsets, $\xi(x)$ is the attribute subsets, $Core(\xi(x))$ is the reduction core of attribute subsets, and $\gamma_{\xi(x)}(D)$ is the reduction quality of $\xi(x)$ relative to D .

The coefficients in (5) $\bar{\rho}$ and $\bar{\delta}$ are defined as follows:

$$\bar{\rho} = \frac{1}{N} \left(\sum_{i=1}^N f_{Elitist_i} \right) \quad (6a)$$

$$\bar{\delta} = \frac{1}{N} \left(\sum_{i=1}^N f_{Subpopulation_i} \right) \quad (6b)$$

where $\bar{\rho}$ denotes the average fitness of elitists in N subpopulations in which $f_{Elitist_i}$ is the fitness of $Elitist_i$ in $Subpopulation_i$, $\bar{\delta}$ denotes the total average fitness of N subpopulations and $\overline{f_{Subpopulation_i}} = \frac{1}{r} \sum_{j=1}^r f_j$ where f_j indicates the best fitness of the j^{th} co-evolutionary individual and r is the number of co-evolutionary individuals in $Subpopulation_i$. Two coefficients $\bar{\rho}$ and $\bar{\delta}$ are introduced to reduce the influence because the fitness value may become too large or too small during the process of solving the optimal set of attribute reduction.

The $\Phi(x)$ in (5) is the adaptive penalty function, which is formulated as follows:

$$\Phi(x) = \Gamma_i \times \left(\frac{\gamma_{\xi(x)}(D)}{\gamma_c(D)} \right) \quad (7a)$$

$$\text{where } \left(\Gamma_i = \frac{1}{N} \sum_{j=1}^N \left| \frac{f_{Elitist_i} - \overline{f_{Subpopulation_j}}}{f_{Elitist_i}} \right| \right) \quad (7b)$$

The penalty function $\Phi(x)$ can be adapt to adjust the fitness value according to the proximity degree for which the target solution of attribute reduction approximates the optimal solution. This will greatly improve the convergence of DCCAEDR in the attribute reduction of big data.

The adaptive fitness function $Fit(x)$ considers the different relationships of attribute interactions to maintain diverse

solutions, and it uses population elitists to improve its efficiency so as to choose between global exploration and local exploitation. So this fitness function has the extremely strong performance to improve the reduction diversity of an optimal solution in the attribute approximation space.

B. Summary of MapReduce Model

The MapReduce model provides a parallel strategy to the distributed computation without burdening the programmer with detail computing [9]. Furthermore, MapReduce implementations usually supply their own distributed file systems that can provide a scalable mechanism for storing large amounts of datasets. The user of the MapReduce library usually expresses two such computation functions as *Map* and *Reduce* [10].

The summary of the MapReduce model can be represented as follows:

Map : $\langle in_key, in_value \rangle$
 $\rightarrow \{ \langle key_i, value_i \rangle \mid i = 1, 2, \dots, k \};$
Reduce : $(key, [value_1, value_2, \dots, value_k])$
 $\rightarrow \langle final_key, final_value \rangle.$

The MapReduce framework is described in Fig.1.

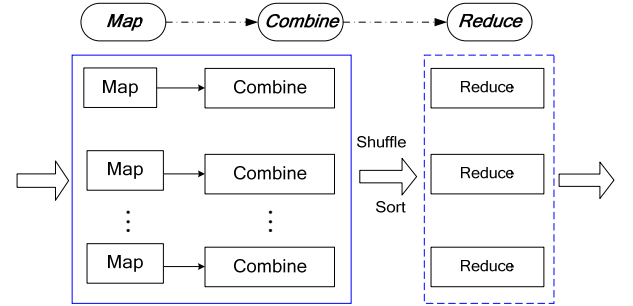


Fig. 1. Framework of MapReduce.

Its paradigm can be summarized as follows:

- *Map* takes an input pair $\langle in_key, in_value \rangle$ and produces a set of intermediate *key/value* pairs $\langle key_i, value_i \rangle$. The MapReduce library collects all intermediate values associated with the same intermediate *key I* and transforms them using the *Combine/Reduce* function.
- *Combine* is a local *Reduce*, which can perform local computations to lessen the burden. It accepts a *key I* and a set of values for that key from the local *Map*. Next, it merges those values together to form a possibly smaller set of values. All results are shuffled, sorted, and sent to the *Reduce* function.
- *Reduce* accepts an intermediate *key I* and a set of values for that key. It merges those values to form a possibly smaller set of values $\langle final_key, final_value \rangle$. Lastly, it can produce more outputs.

Since the MapReduce framework can split the whole dataset into many data subsets, we focus on designing the proper pairs $\langle key_i, value_i \rangle$ to implement the *Map* and *Reduce* functions for attribute cooperation co-evolutionary reduction.

IV. DISTRIBUTED CO-EVOLUTIONARY MAPREDUCE MODEL (DCMM)

Based on the MapReduce model in cloud computing, this paper aims to improve the cooperative co-evolutionary performance of attribute reduction in big data, and to design a novel N -Populations distributed co-evolutionary MapReduce model (DCMM) to accelerate the implementation of attribute reductions simultaneously and cooperatively. DCMM can distribute the workload among multiple independent subpopulations. Its architecture is shown in Fig. 2.

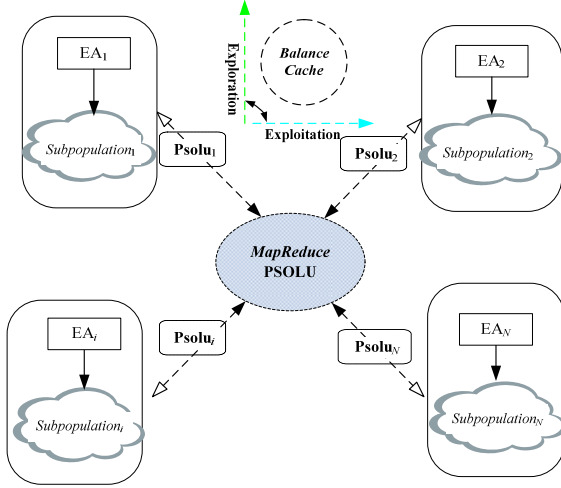


Fig. 2. Framework of N -Populations DCMM.

In order to exploit and explore the inherent parallelism of co-evolutionary populations for attribute reduction, the proposed DCMM model is constructed using the following steps:

Algorithm 1. Distributed co-evolutionary MapReduce model (DCMM)

- 1) Using the MapReduce framework, we divide the entire population into N subpopulations: $Subpopulation_1, Subpopulation_2, \dots, Subpopulation_i, \dots, Subpopulation_N$.
- 2) Each $Subpopulation_i$ shares the respective best solution with its neighbor's distributed co-evolutionary subpopulations, in order to expedite the parallel computational speed.
- 3) Each selected evolutionary algorithm EA_i will be in charge of optimizing its respective assigned $Subpopulation_i$. At the end of each iteration, each $Subpopulation_i$ will focus on producing its best solution with the self-adaptive probability p_i , which represents the probability of $\langle key_i, value_i \rangle$ on MapReduce. This probability p_i is defined as follows:

$$p_i = \left| \frac{f_{ELITIST} - \omega_i f_{Elitist_i}}{f_{ELITIST}} \right| \quad (8a)$$

$$\omega_i = \frac{\sum_{j=1}^N |f_{Elitist_i} - f_j|}{\sum_{i=1}^N \sum_{j=1}^N |f_{Elitist_i} - f_j|} \quad (8b)$$

where $f_{ELITIST} = \text{Min} \{f_{Elitist_1}, f_{Elitist_2}, \dots, f_{Elitist_N}\}$, which is the global best fitness for the best elitist in the entire population, and f_j and $f_{Elitist_i}$ are the same as those in (6a) and (6b).

- 4) Construct the parallel operation $\langle key_i, value_i \rangle$ for each $Subpopulation_i$ under the MapReduce formwork as follows:

$$key_i = p_i + \left| \frac{f_{ELITIST} - f_{Elitist_i}}{f_{ELITIST}} \right| \quad (9a)$$

$$value_i = w_i \times key_i. \quad (9b)$$

- 5) The archiving strategy records the best solution attained by each $Subpopulation_i$. Then all solution sets are combined to form the whole central solution set as follows:

$$\mathbf{PSOLU} = \bigcup_{i=1}^N \mathbf{Psolui} \quad (10a)$$

where

$$\mathbf{Psolui} = \begin{bmatrix} Psolu_{i1} \\ Psolu_{i2} \\ \vdots \\ Psolu_{iN} \end{bmatrix} \times \left[\frac{key_1}{value_1}, \frac{key_2}{value_2}, \dots, \frac{key_i}{value_i} \right]. \quad (10b)$$

As it is shown in Fig. 2, the dashed line indicates the execution sequence order: the black solid arrow contributes to the completed solution construction of the distributed co-evolutionary $Subpopulation_i$, and the white arrow is the corresponding feedback state for $Subpopulation_i$. These subpopulations share their rewards equally from the whole central solution set \mathbf{PSOLU} using the MapReduce framework. As a result, the running processes of the DCMM model demonstrate the superior performance when updating the co-adaptation of all subpopulations.

It is known that the existing cooperative co-evolutionary algorithms usually suffer from difficulties in balancing between exploration and exploitation. Exploitation leads to locally optimal policies, possibly differing from a globally optimal one. However, this DCMM model can contribute to better balance between exploration in breadth and exploitation in depth. An adaptive method would be to start with exploration, and then gradually change to exploitation. Finally, the globally optimal balancing can be reached.

In order to achieve better performance of attribute reduction for big data, the DCMM model excels at exploring the search space and locating the region of a global minimum. Meanwhile, the model also performs well in its exploitation-searching tendency for global optimization, which results in the efficient approximation of all subpopulations with adequate information sharing. Because the co-adaptation of different subpopulations can be reached, their reductions will be surely guaranteed to be the same as those using the whole independent dataset. This determines the adaptive dynamics stability of distributed co-evolutionary subpopulations for attribute reduction in a large-scale complex dataset. So DCMM can be used as the more effective optimization model

to improve the performance of attribute reduction for some complex big data problems.

V. EQUILIBRIUM DOMINANCE STRATEGY OF CO-EVOLUTIONARY ELITISTS (EDSCE)

Due to its parallel and distributed nature, the N -populations DCMM model is more efficient and flexible than traditional cooperative co-evolutionary algorithms. However, another central issue of attribute co-evolutionary reduction is how the cooperation mechanism can be best established and how its stability of fixed points can be realized. Clearly, EGT is the perfect strategy to address this perspective [43].

Currently, most attribute reduction algorithms are sensitive to big data with complex noise, which results in some inaccurate or unexpected reduction results. In order to enhance the robustness of attribute reduction, in this section we propose a novel equilibrium dominance strategy of co-evolutionary elitists (EDSCE). This strategy ensures that the N co-evolutionary subpopulations in the DCMM model can converge to the Pareto-optimal set, and can maintain the diversity of reduction solutions.

This equilibrium dominance strategy of co-evolutionary elitists will be carried out under the N bounded rationality regions, as described in Fig. 3. Here we suppose that all strategies conducted in the polymorphic Nash equilibrium will be similar to the fitness equilibrium among the subpopulations. The strength of the polymorphic Nash equilibrium depends largely upon such vital assumptions that all bounded rationality regions have strategies available to all subpopulations, and thus they can get their payoffs under the complete knowledge of co-evolutionary elitists.

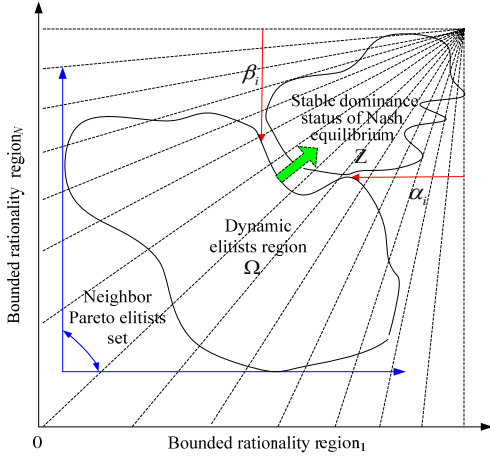


Fig. 3. Equilibrium dominance strategy of co-evolutionary elitists under N bounded rationality regions.

During the co-evolutionary process of elitists, we can regard N bounded rationality regions as the N multiple co-evolutionary elitists' game. The proposed EDSCE strategy meets the searching optimization of the neighbor Pareto elitists, which results in the stable adaptive dominance of an ideal Nash equilibrium set. This Nash equilibrium will be formed more spontaneously without adscititious assumptions. Therefore, the co-evolutionary elitists' game will have a

polymorphic Nash equilibrium that involves more strategies. The elitists under the N bounded rationality regions can greatly improve the optimal performance of dynamic elitists. The steps of EDSCE are described as follows:

Algorithm 2. Equilibrium dominance strategy of co-evolutionary elitists (EDSCE)

- 1) Let all bounded rationality regions be in Ω , bounded rationality region₁ as α_i , bounded rationality region _{N} as β_i , and $f_i^{\min} \leq \alpha_i < \beta_i \leq f_i^{\max}$, so

$$\Omega = \bigcup_{i=1}^N [\alpha_i, \beta_i]. \quad (11)$$

- 2) When the p^{th} neighbor Pareto elitist is a preference node, its bounded rationality region is described as $[f_p^{\min}, f_p^{\max}]$. If there are m neighbor Pareto elitists, the Ω can be refined as follows:

$$\Omega = \bigcup_{i=1, i \neq p}^{N-m} [\alpha_i, \beta_i] \times \bigcup_p^m [f_p^{\min}, f_p^{\max}]. \quad (12)$$

- 3) Judge whether the elitist's solution $f_j(x_i)$ in Ω can enter into the stable solution sets of the Nash equilibrium or not, and then record the result as follows:

$$a_{ij} = \begin{cases} 0, & \alpha_i \leq f_j(x_i) \leq \beta_i \\ 1, & \text{otherwise} \end{cases} \quad (13)$$

- 4) Update the dynamical elitists region Ω as follows:
 - i) Perform the removal of the non-dominance solutions on the region Ω to determine better solutions for the Pareto elitists.
 - ii) Adopt the dynamic neighbor-based Pareto elitists by means of cooperating with neighbor elitists. This will expedite the exploration and exploitation process of co-evolutionary elitists in the dynamic elitists region Ω .
- 5) Define the radius boundaries of a stable Nash equilibrium as follows:

$$r_i = \sqrt{\sum_{j=1}^m \left(\frac{a_{ij} \cdot \min(|f_j(x_i) - \alpha_i|, |f_j(x_i) - \beta_i|)}{f_p^{\max} - f_p^{\min}} \right)^2}. \quad (14)$$

- 6) Construct a desirable judging function of elitists by using the formed stable dominance status of Nash equilibrium as follows:
 - i) Assume P_{ni} as the probability of elitists in *Subpopulation_i* which is used to find the convergence extent of the elitist region $(r_i - \delta_i)$ with the threshold $\delta_i < |\alpha_i - \beta_i|$, and P_{pi} as the probability of finding $(r_i + \delta_i)$ with $\delta_i > |\alpha_i - \beta_i|$.
 - ii) Compute the average weighted probability score WAPS as follows:

$$WAPS = \sum_{i=1}^N \left(\frac{c_i r_i}{w_i} \cdot \log_2 \left(\frac{r_i}{w_i} \right) \right) \quad (15a)$$

$$\text{where } w_i = \begin{cases} 1 - P_{ni}, & 0 \leq c_i \leq (r_i - \delta_i) \\ 1 - P_{pi}, & (r_i - \delta_i) < c_i \leq (r_i + \delta_i) \end{cases} \quad (15b)$$

and c_i is the radius region value of elitists in *Subpopulation_i*.

- 7) Define Z as the radius region set of stable Nash equilibrium dominance. The average radius region set under the respective weighted probability is formulated as follows:

$$Z = \frac{1}{WAPS} \bigcup_{i=1}^N (r_i \cdot w_i). \quad (16)$$

By using this EDSCE strategy, most solutions will be kept within the stable elitists region. The dynamic stability of elitists' behavior adapts to improve the optimal performance of the dynamic elitists region sets. As such, it will guide the co-evolutionary elitists to get closer to the favorite region of optimum equilibrium, and enable the attribute reduction performance to converge quickly to the Nash equilibrium instead of locating the Pareto front.

When the number of participating elitists increases, it is difficult to collaborate for higher profits but the EDSCE strategy can present a stability strategy to enhance the dynamics of co-evolutionary elitists, and the Nash equilibrium yields a maximized, but not optimal, profit result. As expected, the final situation converges to the optimal dominance status of Nash equilibrium. This result would be in accordance with the intended purpose.

VI. PROPOSED DCCAEDR ACCELERATOR

Based on the above-proposed DCMM model and EDSCE strategy, we propose a new attribute equilibrium dominance reduction accelerator (DCCAEDR) based on the distributed co-evolutionary cloud model. Its aim is to retain the discriminatory power of original attribute sets and improve the approximation computations of attribute co-evolutionary reduction in big data. Its kernel ideas are designed as follows:

First, divide a large-scale attribute set into different rationality regions in order to collaborate and form dynamic elitists region. For the DCMM model with MapReduce, we employ the same QSFLA [25] as each evolutionary algorithm, namely EA_i as shown in Fig. 2, to optimize each subpopulation. The goal is to improve the computational parallel performance of the DCCAEDR accelerator, while retaining the same selecting performance.

Second, construct neighbor Pareto elitists using the EDSCE strategy until the elitist region Ω can reach the stable region Z with Nash equilibrium dominance.

Third, set up the proper pairs $\langle key_i, value_i \rangle$ and implement parallel accelerating operations of *Map* and *Reduce*.

Fourth, compute the reduction rule set based on the MapReduce approximation parallelism, and then perform above process repeatedly until the reduction results can achieve entire reduction sets with the equilibrium dominance solution.

The DCCAEDR accelerator can construct a parallel framework to compute the attribute reduction subsets using MapReduce. This avoids the problem that most existing algorithms only run on a single computer to deal with small data sets. The flowchart of DCCAEDR accelerator is illustrated in Fig. 4, and is explained step by step below:

Algorithm 3. Attribute equilibrium dominance reduction accelerator (DCCAEDR) based on distributed co-evolutionary cloud

Input: $S = (U, A, V, f)$

Output: Entire the attribute reduction with the equilibrium dominance solution Red_{Eed}

- 1) Set up a searching space for the attribute reduction, including m co-evolutionary subpopulations. Decomposed the attribute set into m attribute subsets $Sub_attribute_i$. Each subpopulation and attribute subset must remain in the same rationality region. Thus, each subpopulation optimizes its respective attribute subset.
- 2) Calculate the weight of each attribute subset $Sub_attribute_i$ using the following formula:
$$Weight_i = \gamma_{core(c)-Sub_attribute_i}(D) - \gamma_{core(c)}(D). \quad (17)$$
- 3) Project each subpopulation into the condition attribute subset, and limit it into the defined reduction as
$$Weight_{i_p} = \frac{|Weight_i - Weight_{min}|}{|Weight_{max} - Weight_i|}. \quad (18)$$
- 4) Perform the Map and Reduce procedures under the N bounded rationality regions using the proposed DCMM model in **Algorithm 1** as follows:
 - i) For each attribute subset do
 - { Set c be flag as the equivalence class $\bar{x}_{C-\{c\}}$;
 - $key \leftarrow c + \bar{x}_{C-\{c\}}$;
 - ii) Conduct two operations as
 - $value \leftarrow I_d(x)$ and $c_AttriSig \leftarrow 0$.
 - iii) Denote the list of decision attribute as V .
 - iv) For each $j \in V$ do
 - { $N_p^j \leftarrow$ the frequency of decision attribute;
 - $c_AttriSig \leftarrow \sum_{j=1}^N N_p^j$;
 - v) Select the flag c from the equivalence class key.
 - vi) Implement two pairs of operations as
 - $key_i \leftarrow c$ and $value_i \leftarrow c_AttriSig$.
 - vii) Merge these values together to form the proper pairs set $\langle key, value \rangle$.
- 5) Construct neighbor Pareto elitists, obtain \mathbf{Psolu}_i for the rationality region _{i} , and then collaborate them to form a distributed co-evolutionary dynamic elitists region.
- 6) Carry out the EDSCE strategy in **Algorithm 2** until the elitists regions Ω can reach the stable Nash equilibrium region Z .
- 7) Perform the parallel accelerating reduction of the rule set

as follows:

- { i) Select attribute subset $Sub_attribute_i \subset Z$.
- ii) Execute the parallel attribute subsets by using the DCMM model in **Algorithm 1**.
- iii) Reduce rule R_i with Map_i by using \mathbf{Psolu}_i with the self-adaptive probability P_i .
- iv) Calculate the degree of dependency $\gamma_C(D)$ as the criterion for attribute selection.
- v) Compare fitness $Fit(x)$, and select out the best attribute reduction subset R_i^{best} for each $Subpopulation_i$.

- 8) Achieve the entire attribute reduction set with the equilibrium dominance as follows:

$$Red_{Eed} = \bigcup_{i=1}^Z \{R_i^{best}\}. \quad (19)$$

- 9) Stop if convergence is attained or the number of iterations exceed the criterion. Otherwise, go to Step 4).
- 10) Output the entire attribute reduction with equilibrium dominance solution as Red_{Eed} .

VII. EXPERIMENTAL STUDIES

In this section, we conduct a series of studies of the proposed DCCAEDR accelerator, and compare experimental results with several representative algorithms. Our experiments run on the Apache Hadoop platform [48], which is an open source software framework with supporting data-intensive distributed applications. Hadoop version 1.0.1 and Java 1.6.0.12 are employed as the MapReduce system. For distributed experiments, we run algorithms on fifteen nodes, in which one is set as a master node and the remaining nodes are configured as slave nodes. Each node has 64 GB main memory and an AMD Opteron Processor 2376 with 2 Quad-Core CPUs. The operating system in these machines is Linux CentOS 5.2 Kernel 2.6.18. For the following results, we present the average values of 30 runs, and the best average value achieved for testing is shown in **bold type**.

This experimental section includes five subsections. The comparative study between DCCAEDR and various representative algorithms will be performed on new large-scale global optimization functions in subsection A, and on ten UCI machine learning datasets (<http://www.ics.uci.edu/mllearn/MLRepository.html>) in subsection B, respectively.

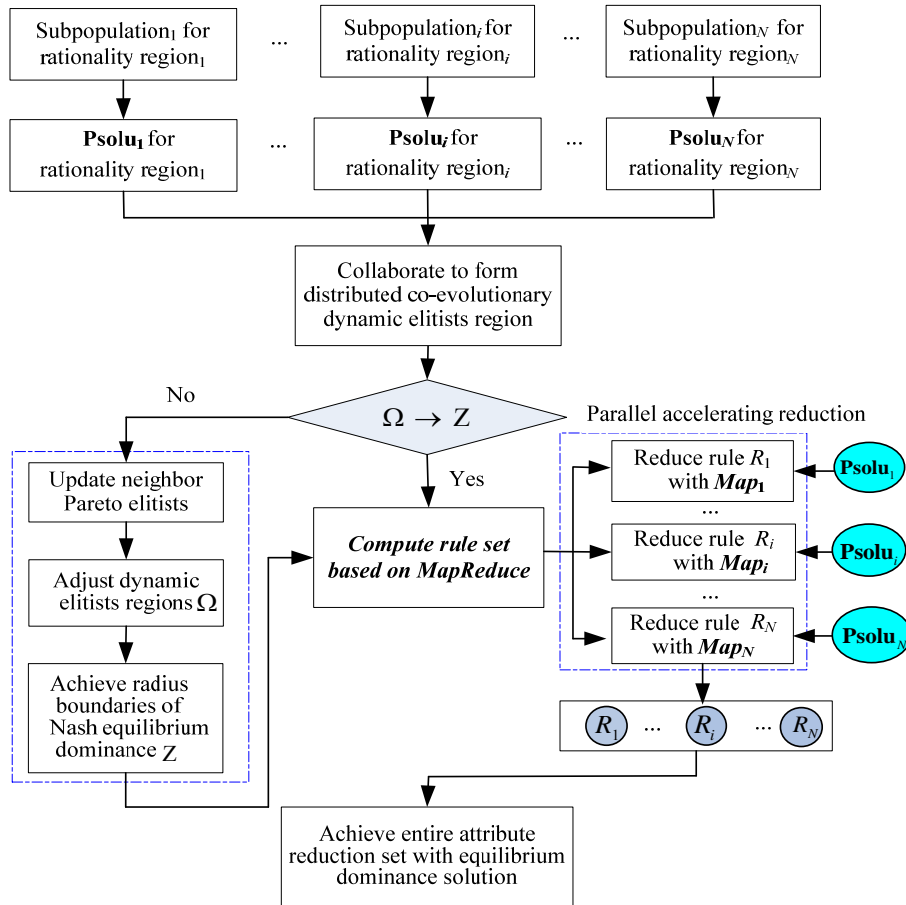


Fig. 4. Flowchart of DCCAEDR accelerator.

Subsequently, DCCAEDR is further determined to test the classification performance on UCI datasets with different attribute-noise levels in subsection C, and to evaluate its stability with the increasing number of datasets in subsection D. Finally, DCCAEDR is applied to solve attribute reduction for traditional Chinese medical records and to segment the cortical surface of neonatal brain 3D-MRI records in subsection E.

A. Optimization on Benchmark Functions for CEC'2013

We evaluate the optimization performance of DCCAEDR with “Benchmark Functions for the CEC'2013 Special Session and Competition on Large Scale Global Optimization”[49], in which 15 large-scale benchmark problems are proposed as the extension to the CEC'2010 large-scale global optimization benchmark functions. It is known that following factors make this type of large-scale global optimization exceedingly difficult. First, the search space of the problem grows exponentially as the number of decision variables increases. Second, the properties of the search space may change as the number of dimensions increases. Third, this factor contributes to the difficulty of large-scale optimization with interacting variables. Its goal is to better represent a wider range of real-world large-scale optimization problems and to provide convenience and flexibility when comparing various algorithms for large-scale global optimization. Therefore, we selected 5 different benchmark functions to carry out this experiment on large-scale global optimization. Here, f_2 is a fully separable function, f_5 is a separable subcomponent function, f_9 is a nonseparable subcomponent function, f_{13} is an overlapping function, and f_{15} is a nonseparable function. Details of these functions can be found in Reference [49]. Experiments are conducted on 1000- D for these functions. Here we compare DCCAEDR with three representative co-evolutionary algorithms: CoBRA [36], DECC-G [37], and CCNSGAI [50]. CoBRA is a co-evolutionary bi-level algorithm using repeated methods, and is able to face general bi-level optimization problems involving the complex large-size searching spaces. DECC-G is a differential evolutionary algorithm based on the cooperative co-evolution with the grouping structure and it can decompose a nonseparable problem, and optimize a group of tightly interdependent variables together. CCNSGAI is a cooperative co-evolutionary non-dominated sorting genetic algorithm which can use different representative selections of cooperative co-evolution by choosing a random individual from the best non-domination level. The running number is set as 30 per function and the maximum value of fitness evaluations is set as $FE = 3 \times 10^6$. The experimental results (Best fitness, Worst fitness, Average fitness, Standard Deviation and Running time) for each algorithm on the five functions are listed in Table I.

From the results, we can see that DCCAEDR performs quite well for 4 out of 5 large-scale global optimization functions, including the fully separable function f_2 , the separable subcomponent function f_5 , the nonseparable subcomponents

function f_9 , and nonseparable function f_{15} . For the overlapping function f_{13} , DCCAEDR fails to obtain a few of the optima values, but it is close the best performance achieved by CCNSGAI. The reason for this phenomenon is that not only there are a large number of local optimal solutions in the overlapping function f_{13} , but also some subpopulations might lapse into the degeneracy so that the successive range may be greatly restricted.

Function f_{15} is a completely non-separable function in which any two variables will interact so they cannot be optimized independently in order to determine the global optimum of the objective function. However, DCCAEDR can pursue a better near-optimum value. The reason why it performs better when discovering and exploiting problem structures is because it can incorporate with various features like archiving and adequate information-sharing, and maintain the uniform solutions distribution using MapReduce. Meanwhile, DCCAEDR has an obvious advantage in running time. DCCAEDR can construct a completed solution with the speed-up factor of the Nash equilibrium optimum.

The experimental results verify DCCAEDR's higher effectiveness in exploring and exploiting the inherent structures of large-scale global optimization functions, although no prior knowledge is available. DCCAEDR has achieved significant results, which further confirms its obvious superiority for optimization on large-scale global functions.

B. Attribute Reduction Evaluation on UCI Datasets

Ten UCI machine learning datasets, the numbers of which have been magnified to be 10^5 times that of the original datasets, are selected to verify the performance of DCCAEDR, compared with such representative parallel attribute reduction algorithms as PACCA [30] and E-FSA [51]. PACCA is the parallel algorithm for computing the core attributes, and it can compute equivalence classes and reduce the search space. E-FSA is the efficient rough feature selection algorithm, and it can select sub-tables from the large-scale data table and fuse the final reductions on all selected sub-tables. In Table II, that $|Core|$ denotes the number of *Core* attributes, $|U'|$ denotes the number of elements in the reduced decision table, namely $|U/C|$, RL represents the number of condition attributes in the minimum reduction, and R_{E-FSA} , R_{PACCA} , $R_{DCCAEDR}$ all represent the reduction results of E-FSA, PACCA and DCCAEDR, respectively. The numbers in the attribute reduction results represent the serial number of the condition attributes, while the underline represents *Core* attributes. In Table III, *Time* is the running time, *Space* is the space consumption and *FSn* is the number of selected features.

As it is shown in Tables II and III, it is clear that the solutions obtained by DCCAEDR are better than those found by E-FSA and PACCA. Each algorithm usually comes with a substantially reduced computing time and space load. The performance usually improves in the presence of a big dataset: the bigger the dataset, the more profound the computing savings. It is remarkable that in our implementation the big space consumption also leads to the memory overflow for

E-FSA in Hayes-roth, Airfoil and Forest fire datasets, and the same memory overflow is for PACCA in the 9Gauss and Hayes-roth datasets. But DCCAEDR could deal with these situations. It is noted that DCCAEDR clearly outperforms E-FSA and PACCA in most datasets. As for 9Gauss, Hayes-roth, and Forest fire datasets, the significant improvements are brought by DCCAEDR.

For example, DCCAEDR spends the space consumption (*Space*) 65.32 % of PACCS in the Airfoil dataset and 60.04 % in the Forest fire dataset. The reason for this is the total main population is decomposed into the corresponding number of subpopulations, and those distributed co-evolutionary subpopulations can evolve in parallel using the DCMM model fast.

TABLE I
AVERAGE RESULTS OF FOUR ALGORITHMS ON FIVE LARGE SCALE GLOBAL OPTIMIZATION FUNCTIONS

Function	Algorithm	Best	Worst	Avg.	StDev.	Running time (s)
f_2	DECC-G	3.2143e-024	2.1335e-018	3.1534e-020	1.3107e-020	0.341
	CCNSGAI	4.2276e-026	3.5160e-017	4.4613e-023	1.6690e-023	0.401
	CoBRA	4.2467e-025	5.2579e-019	4.3729e-023	2.5578e-023	0.360
	DCCAEDR	5.6149e-036	9.9445e-031	2.8827e-034	1.2140e-034	0.207
f_5	DECC-G	3.2349e-017	9.9069e-014	6.5299e-016	3.1976e-016	0.583
	CCNSGAI	1.6946e-015	4.2108e-013	2.1996e-014	1.3598e-014	0.629
	CoBRA	5.3348e-014	3.0917e-010	3.1525e-012	1.3840e-012	0.439
	DCCAEDR	1.2040e-021	3.0187e-017	8.1587e-020	3.1398e-020	0.395
f_9	DECC-G	1.3490e-003	1.5690e-001	1.3400e-002	2.3411e-003	0.934
	CCNSGAI	6.2309e-005	3.4578e-003	9.0876e-003	4.5090e-003	1.394
	CoBRA	1.8956e-004	2.2456e-002	3.8934e-003	3.4572e-003	1.132
	DCCAEDR	3.2908e-006	4.8903e-004	1.3423e-004	3.0009e-004	0.897
f_{13}	DECC-G	2.9809e-005	3.4560e+000	1.0098e-003	3.5690e-003	1.325
	CCNSGAI	4.9098e-007	4.4589e-005	3.4590e-005	2.8934e-005	1.213
	CoBRA	1.4589e-005	1.5546e-002	3.7245e-003	9.0867e-003	1.519
	DCCAEDR	2.3400e-006	5.8909e-005	1.5590e-006	8.5656e-006	1.250
f_{15}	DECC-G	5.9812e-005	3.0564e-003	4.8909e-004	4.5678e-004	2.531
	CCNSGAI	6.9866e-004	1.5637e-002	6.7854e-003	1.8734e-002	2.209
	CoBRA	4.3490e-005	4.3490e-003	4.7823e-003	3.4590e-002	1.998
	DCCAEDR	3.9878e-006	1.5434e-004	1.7645e-004	3.2390e-004	1.531

TABLE II
RESULTS OF ATTRIBUTE REDUCTION FOR DCCAEDR WITH E-FSA AND PACCA
(“—” INDICATING NO TRIAL CAN REACHED ACCEPTABLE SOLUTION)

Dataset	Core	U'	RL	Result of attribute reduction		
				R_{E-FSA}	R_{PACCA}	$R_{DCCAEDR}$
Pendigits	6	187	4	{2, <u>5</u> , 12, <u>13</u> }	{ <u>5</u> , 6, 8, <u>13</u> }	{3, <u>5</u> , 11, <u>13</u> }
Hepatitis	7	325	4	{ <u>1</u> , <u>2</u> , <u>5</u> , 9}	{ <u>1</u> , <u>2</u> , <u>5</u> , 7}	{ <u>1</u> , <u>2</u> , <u>5</u> , 8}
Vertebral	10	782	7	{1, 2, 4, 5, 7, 8, 11}	{ <u>2</u> , <u>3</u> , <u>4</u> , <u>5</u> , <u>6</u> , <u>8</u> , 10}	{ <u>2</u> , <u>3</u> , <u>4</u> , <u>5</u> , <u>6</u> , <u>8</u> , 9}
Ecoli	5	798	8	{ <u>2</u> , <u>4</u> , <u>6</u> , 8, <u>14</u> , <u>16</u> , 17, <u>21</u> }	{ <u>2</u> , <u>4</u> , 5, <u>6</u> , <u>14</u> , <u>16</u> , <u>21</u> , 31}	{ <u>2</u> , <u>4</u> , <u>6</u> , 8, <u>14</u> , <u>16</u> , <u>21</u> , 23}
Parkinson	27	3196	6	{ <u>2</u> , <u>19</u> , 21, <u>22</u> , 27, <u>32</u> }	{2, 8, 10, 14, 17, 28}	{ <u>2</u> , 8, <u>19</u> , <u>22</u> , 29, <u>32</u> }
9Gauss	7	8124	4	{8, 11, 16, 24}	—	{ <u>3</u> , <u>5</u> , <u>8</u> , 19}
Hayes-roth	8	4002	7	—	—	{ <u>1</u> , <u>2</u> , <u>4</u> , <u>5</u> , <u>6</u> , <u>7</u> , <u>11</u> }
Airfoil	5	7020	6	—	{5, 11, 13, 19, 21, 22}	{ <u>2</u> , <u>3</u> , <u>5</u> , <u>6</u> , <u>7</u> , <u>10</u> }
Forest fire	7	5320	6	—	{3, 6, 8, 11, 14, 15, 19}	{ <u>2</u> , <u>6</u> , <u>8</u> , <u>10</u> , <u>13</u> , <u>15</u> }
Arrhythmia	8	6780	8	{ <u>5</u> , 6, <u>7</u> , 9, <u>12</u> , <u>15</u> , <u>17</u> , <u>20</u> }	{ <u>5</u> , <u>7</u> , <u>12</u> , <u>15</u> , <u>17</u> , 19, <u>20</u> , 29}	{ <u>5</u> , 6, <u>7</u> , <u>12</u> , 14, <u>15</u> , <u>17</u> , <u>20</u> }

TABLE III
COMPARISON OF ATTRIBUTE REDUCTION PERFORMANCE FOR DCCAEDR WITH E-FSA AND PACCA

Dataset	E-FSA			PACCA			DCCAEDR		
	Time (s)	Space (M)	FSn	Time (s)	Space (M)	FSn	Time (s)	Space (M)	FSn
Pendigits	0.335	1.221	7	1.204	0.742	9	0.132	0.531	5
Hepatitis	3.103	1.987	8	1.293	1.802	11	0.861	1.065	6
Vertebral	2.698	7.231	11	2.067	3.910	8	2.219	4.112	8
Ecoli	0.680	2.945	6	0.792	3.172	5	0.522	1.837	4
Parkinson	15.728	19.192	23	21.028	23.490	31	10.543	10.515	17
9Gauss	23.109	24.786	27	Memory Overflow			17.324	16.743	15
Hayes-roth	Memory Overflow			Memory Overflow			19.764	20.278	12
Airfoil	Memory Overflow			16.213	28.765	16	11.171	18.786	10
Forest fire	Memory Overflow			29.361	27.278	16	17.239	16.378	9
Arrhythmia	27.450	32.109	21	32.098	40.125	28	20.561	25.870	19

The Map and Reduce working in Hadoop is launched, and have the best speed for above big datasets. The intermediate data produced by Map is written to the local file system which will greatly improve the efficiency of DCCAEDR for attribute reduction. Therefore, when attributes in larger datasets are reduced, the Map and Reduce can perform faster, allowing DCCAEDR to significantly reduce the computational time and space load as well as maintaining the excellent performance.

According to comparisons of the above listed quality criteria, DCCAEDR achieves satisfying attribute reduction results. Therefore, it is feasible and efficient to perform attribute reduction on the big datasets.

C. Classification Performance on UCI Datasets with Different Attribute-noise Levels

To further test the application feasibility of DCCAEDR, we quantitatively compare its classification accuracy with representative algorithms. In five selected UCI datasets, we use 55% as the training set, 15% as the validation set and 30% as the testing set. The comparisons of classification accuracy are performed when the classifiers C4.5 and Naive Bayes are employed to test the classified datasets using DCCAEDR and PACCA [30], E-FSA [51] and FVPRS [52], respectively. FVPRS is the better model of fuzzy variable precision rough sets, and it's less sensitive to both misclassification and perturbation in a fuzzy information system. Tables IV and V depict the classification accuracy of different algorithms when dealing with five datasets with different attribute-noise levels. The 5%, 10%, and 20% attribute-noise levels are added into the raw datasets, and then we apply four algorithms on these datasets to compute their classification accuracies. The "Avg." row records the average values of accuracies.

TABLE IV
CLASSIFICATION ACCURACY OF FOUR ALGORITHMS WITH C4.5 CLASSIFIER IN DIFFERENT ATTRIBUTE-NOISE LEVELS

Dataset	Attribute-noise level (%)	FVPRS (%)	E-FSA (%)	PACCA (%)	DCCAEDR (%)
Thyroid	5	89.23	84.43	87.12	90.23
	10	82.58	76.32	83.24	87.90
	20	75.12	69.15	80.32	85.89
Musk Version2	5	86.12	90.34	88.79	89.16
	10	82.32	85.13	82.31	86.23
	20	80.09	80.15	75.92	80.15
Arrhythmia	5	79.16	84.19	80.32	87.25
	10	75.68	80.43	74.52	85.21
	20	69.75	75.35	69.98	79.88
Audiology	5	87.46	90.49	89.71	90.13
	10	81.19	84.25	86.18	89.21
	20	76.08	72.68	81.65	83.40
Weka-3.2G	5	87.19	89.29	92.17	87.45
	10	82.31	84.14	85.21	84.13
	20	76.22	78.17	83.47	81.29
Avg.		80.70	82.25	82.11	85.83

TABLE V
CLASSIFICATION ACCURACY OF FOUR ALGORITHMS WITH NAIVE
BAYES CLASSIFIER IN DIFFERENT ATTRIBUTE-NOISE LEVELS

Dataset	Attribute-noise level (%)	FVPRS (%)	E-FSA (%)	PACCA (%)	DCCAEDR (%)
Thyroid	5	89.23	84.43	87.12	90.23
	10	82.58	76.32	83.24	87.90
	20	75.12	69.15	80.32	85.89
Musk Version2	5	86.12	90.34	88.79	89.16
	10	82.32	85.13	82.31	86.23
	20	80.09	80.15	75.92	80.15
Arrhythmia	5	79.16	84.19	80.32	87.25
	10	75.68	80.43	74.52	85.21
	20	69.75	75.35	69.98	79.88
Audiology	5	87.46	90.49	89.71	90.13
	10	81.19	84.25	86.18	89.21
	20	76.08	72.68	81.65	83.40
Weka-3.2 G	5	87.19	89.29	92.17	87.45
	10	82.31	84.14	85.21	84.13
	20	76.22	78.17	83.47	81.29
Avg.		80.70	82.25	82.11	85.83

The results in Table IV indicate that four algorithms can improve the classification accuracy by eliminating certain irrelevant attributes in the five datasets. It can be observed that in most cases DCCAEDR achieves higher classification accuracy than the three compared representative algorithms. For example, the number of cases in which DCCAEDR achieves higher classification accuracy over FVPRS, E-FSA and PACCA are five, five and four out of five datasets, respectively, with the C4.5 classifier. In particular, on the Musk Version2 dataset the classification accuracy of DCCAEDR can be enhanced substantially from 80.15% to 86.23%, and 89.16%, respectively, when the attribute-noise level reduces from 20 to 10, and 5.

From Table V, the experimental results with the Naive Bayes classifier also show similar results. It can be seen that in most datasets the accuracy of the Naive Bayes classifier based on DCCAEDR is significantly better than the classifier based on the compared algorithms. Meanwhile, from the view of the ‘‘Avg.’’ values, the average accuracy of DCCAEDR surpasses the compared algorithms, irrespective of different classifiers. Therefore, the experimental results indicate that DCCAEDR provides an effective way to attain better classification results on UCI datasets with different attribute-noise levels.

In addition, these results show better variations in the classification performance when the attribute-noise level increases. It is apparent that the accuracies of two classifiers based on FVPRS, E-FSA and PACCA drop sharply, having performing very poorly in most datasets. However the

classifier based on DCCAEDR is more robust and its classification accuracy does not change especially much when the attribute-noise level increases.

According to the aforementioned experimental analysis, the following conclusions can be drawn: First, DCCAEDR can decompose a large-scale dataset into many data subsets using MapReduce, acquiring a consistent decision table and attribute reduction. It can significantly reduce the computational time necessary for selecting candidate attributes with different attribute-noise levels. So DCCAEDR is feasible and efficient. Second, from the perspective of classification performance, in most cases DCCAEDR experienced no losses in the classification performance when compared with representative algorithms, and it rapidly converges to the best solutions with smaller error rates. Therefore, it is much more robust in reducing the noise-related influence. The results further demonstrate the high efficiency and accuracy of DCCAEDR, although big data sets have different attribute-noise levels.

D. Stability Evaluation of DCCAEDR

Stability is another metric used to characterize the robustness of different algorithms. In this subsection, we compare the stability of DCCAEDR, PACCA [30], E-FSA [51] and FVPRS [52], and further discuss their properties. Here we test four algorithms on the selected specific Weka-3.2G dataset, and the same conclusions can be extended to other datasets. In this experiment, the selected Weka-3.2 G dataset is decomposed into 20 subsets of the same size, and we evaluate the stability of attribute reduction with 20-fold cross validations. The variation tendency of accuracy curves in Fig. 5 is employed to evaluate the stability of attribute reduction.

In Fig. 5, we can see the accuracy curve of DCCAEDR has improved only slightly and it can remain stable with an increasing number in the Weka-3.2G dataset. However, the accuracy curves of the compared algorithms increase much more obviously. The reason for this is that the equilibrium dominance strategy of elitists under the N bounded rationality regions used in DCCAEDR exhibits better stability performance. Thus, the Pareto-optimal solutions can effectively be found using the EDSCE strategy. This strategy can meet the Nash equilibrium of the neighbor Pareto elitists spontaneously, and ensure the stable equilibrium dominance result while minimizing the total cost. This stability will remain consistent with the theoretical prediction.

The above comparisons demonstrate the strong superior performance of DCCAEDR, in which different subpopulations can deeply communicate and share their search strategies with neighbor subpopulations in order to acquire equilibrium dominance more efficiently. This can improve the execution time and boost accuracy. Therefore, DCCAEDR can choose the same results of attribute reduction as the original version in big datasets, which attains the same or significantly higher performance. From the perspective of big data, DCCAEDR is very suitable for the attribute reduction problem.

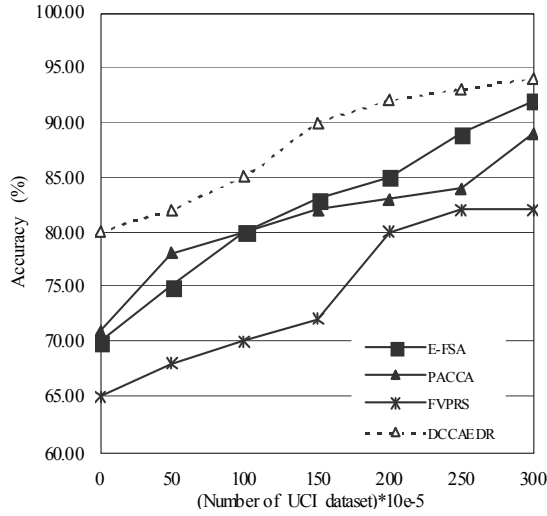


Fig. 5. Accuracy stability comparison of DCCAEDR with E-FSA, PACCA and FVPRS.

E. Application Performance Assessment of DCCAEDR

In recent years, different attribute reduction algorithms in Rough set theory have devoted great effort from laboratory demonstrations to daily-life applications. In the following experiments, we will assess the performance of DCCAEDR on attribute reduction for traditional Chinese medical (TCM) records and its segmentation of cortical surfaces in neonatal brain 3-D MRI records, which is an important implication for the forecasting, diagnosis and treatment evaluation of certain medical diseases. The experimental results will further exhibit DCCAEDR's application performance to deal with real large-scale datasets with complex noise.

Experiment 1: Application on the TCM System

The principle of TCM diagnosis is mainly based on the information obtained from four diagnostic processes: inspection, listening and smelling, inquiry, and palpation. As a complete knowledge system, the TCM system focuses on researching human health conditions via different approaches compared to orthodox medicine. However, most TCM rules and knowledge could not be retained scientifically and quantitatively from the TCM system with complex medical noise. Therefore, it is necessary to build an intelligent diagnosis system to enhance the accuracy of the computer-assisted TCM platform [53-55]. Moreover, no extensive work has yet been devoted to discussing the influence of noise on attribute reduction in rough set theory for TCM system so far. Surely, the noise has a different influence on dependence estimation when the attribute reduction is considered in TCM. In this subsection, a developed TCM system using DCCAEDR can detect diseases and assist patients to take precautions and take advantage of available therapies. This disease diagnosis database includes approximately 8954 kinds of diseases, more than 10034 kinds of Pharmaceuticals, more than 50000 kinds of Chinese herbal formulas, and more than 28900 kinds of pharmaceutical

application information. The medical data in this TCM database has such complicated and unique features that the traditional medical intelligent hypothesis does not work well. However, our proposed TCM system can not only extract some useful empirical regularities, but it can also assist medical experts in judging whether or not the discovered regularities are effective for related diseases.

In this experiment, we select the data source of Cold Syndrome attributes as a TCM record, and load these medical records into the TCM system. We remove the redundant information from a series of symptoms, extract valuable attributes, and finally classify the related symptoms.

First, we extract the related condition attributes from the Cold Syndrome database, and each Cold Syndrome corresponds to its condition attributes (A1-A8): A1 as Fever, A2 as Thirst, A3 as Limbs, A4 as Languid, A5 as Cough, A6 as Pulse Condition late/tight, A7 as Dizziness and Tinnitus, and A8 as Chest Distress. In addition, A9 is the Cold Syndrome decision attribute. As it is shown in Fig. 6(a), the number "1" represents a case showing this corresponding symptom, but "0" represents a case that does not show this corresponding symptom. Analyzing these symptoms is an important step in the quantitative conditions of patient diagnosis using the Cold Syndrome database.

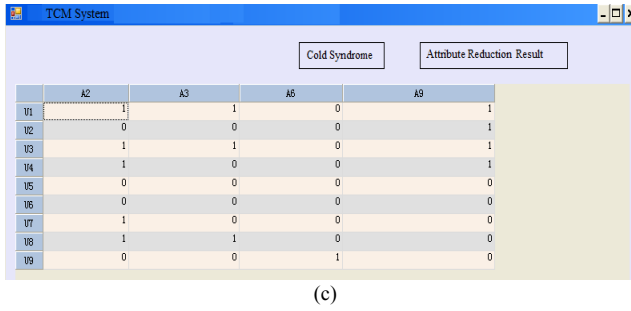
Second, we divide the properties of A9 into: {U1, U2, U3, U4} and {U5, U6, U7, U8, U9}. The set {U1, U2, U3, U4} indicates that the patient is suffering from Cold Syndrome; the set {U5, U6, U7, U8, U9} represents that the patient is not suffering from Cold Syndrome. The attribute reduction procedure of Cold Syndrome using DCCAEDR is described in Fig. 6 (b) and (c).

TCM System										
Cold Syndrome										
	A1	A2	A3	A4	A5	A6	A7	A8	A9	
U1	1	1	1	1	0	0	0	1	1	
U2	0	1	0	0	1	1	0	0	1	
U3	1	1	1	1	1	0	0	1	1	
U4	1	1	1	0	1	0	0	1	1	
U5	0	0	0	0	0	0	0	1	0	
U6	0	0	0	0	1	1	0	0	0	
U7	0	1	1	0	1	1	0	0	0	
U8	0	0	1	1	0	1	0	0	0	
U9	0	1	0	0	1	0	1	1	0	

(a)

TCM System										
Cold Syndrome Attribute Reduction Procedure										
	U5	U6	U7	U8	U9					
U1	A1A2A3A4	A1A2A3A4A5A6A8	A1A4A5A6A8	A1A2A5A8	A1A3A4A5A7					
U2	A2A5A6A8	A2	A3	A2A3A4A5	A6A7A8					
U3	A1A2A3A4A5	A1A2A3A4A6A8	A1A4A6A8	A1A2A5A6A8	A1A3A4A7					
U4	A1A2A3A5	A1A2A3A6A8	A1A6A8	A1A2A4A5A6A8	A1A3A7					

(b)



	A2	A3	A6	A9
U1	1	1	0	1
U2	0	0	0	1
U3	1	1	0	1
U4	1	0	0	1
U5	0	0	0	0
U6	0	0	0	0
U7	1	0	0	0
U8	1	1	0	0
U9	0	0	1	0

(c)

Fig. 6. Attribute reduction for Cold Syndrome record in TCM system: (a) Original decision information table. (b) Procedure of attribute reduction. (c) Optimal attribute reduction table.

As it is seen from Fig. 6, three attributes “A2 as Thirsty,” “A3 as Limbs,” and “A6 as Pulse Condition late/tight,” are the inevitable symptoms for “A9 as Cold Syndrome,” and the remaining five symptom attributes for “Cold Syndrome” are almost redundant, which is a the better guiding value for the practical “Cold Syndrome” diagnosis. According to the experimental results, the overall correctness rate for “Cold Syndrome” diagnosis can exceed 91.50%. It will reach a consistent conclusion which the clinical doctor obtains using conventional judgment. This TCM system can immensely reduce a lot of vague TCM values, extract useful rules, and enhance the effectiveness of TCM diagnoses. Further, combining the domain knowledge of clinical doctors with the proposed TCM system has shown that the majority of valuable TCM attribute rules will produce more and more accurate results and knowledge, which will be worthy of a diagnosis using a TCM record.

Experiment 2: Application on Segmentation of Cortical Surface of Neonatal Brain 3D-MRI Records

The segmentation of a neonatal brain 3D-MRI into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF), is challenging due to low spatial resolution, high image noise, and the dynamic myelination process. These

inconsistent segmentation results usually generate some longitudinally-inconsistent cortical surfaces because of more severe inconsistencies in tissue segmentation, surface tessellation, and surface evolution. Moreover, all longitudinal surfaces of the same subject also need to be normalized onto a common space in order to measure longitudinal development. This is usually done using the feature-based registration, but it also leads to temporally incorrect or bumpy correspondences especially in the cortical surface regions. This may affect the eventual accurate measurement of cortical development [56-59].

In order to accurately and consistently measure the development of the cortical surface and related cortex attributes for neonatal brain 3D-MRI records, DCCAEDR is firstly applied to the segmentation of the neonatal brain 3D-MRI slice, and each longitudinal image is acquired at 5, 10, 15 and 18 months. Fig. 7 shows the consistently reconstructed longitudinal inner and outer cortical surfaces embedded in their respective brain image spaces, in which the red and blue curves indicate inner and outer surfaces, respectively.

Fig. 8 shows the surface distance at 5, 10, 15 and 18 months with adding 10% Gaussian noise into the neonatal brain 3D-MRI record, color-coded by surfaces. The edges along different organizations of raw brain regions are fuzzy and non-uniform. The misclassification rate is always too high, and the non-brain region may be easily mistaken for the brain region. As seen in Fig. 8, the DCCAEDR is used to segment them, and the outline of the watershed is taken as the initial curves of the level set in order to realize the automatic segmentation of brain tissues (WM, GM and CSF). Moreover, the noise of the neonatal brain could be reduced significantly as well as preserving the details of the brain image. The cortical thickness develops dynamically, especially from 10 to 15 months. The results will reach the better balance between removing the noise and preserving the true brain regions, while seeking higher segmentation accuracy.

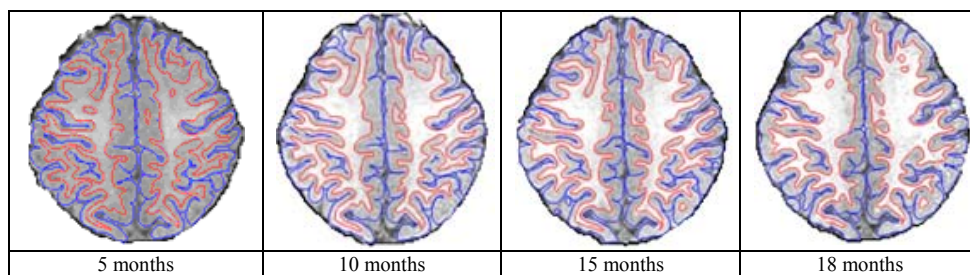


Fig. 7 Segmentation of cortical surfaces of neonatal brain MRI slice by DCCAEDR.

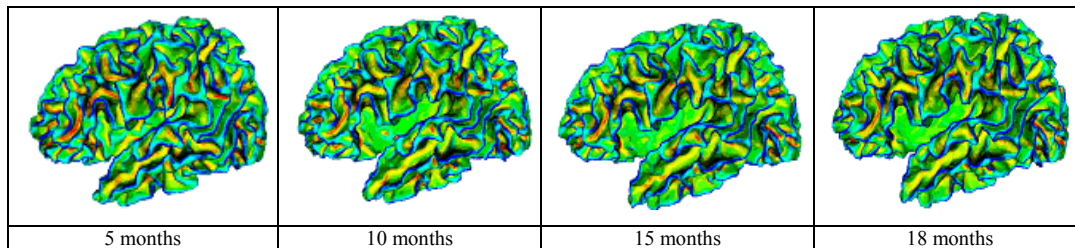


Fig. 8 Segmentation of neonatal brain 3D-MRI record with 10% Gaussian noise by DCCAEDR.

These results demonstrate the feasibility and practicality of DCCAEDR applied to the segmentation of neonatal brain 3D-MRI records.

In order to quantitatively characterize the segmentation accuracy for neonatal brain 3D-MRI records, the typical Dice Ratio (DR) [60] is employed to evaluate the performance of three kinds of surfaces: the modeled inner surface, the reconstructed inner surface and the reconstructed outer surface. The results are depicted in Fig. 9. The box plots have lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to the most extreme data value. The outliers are data with values beyond the ends of whiskers, displayed by some additional noise. The segmentation accuracy of DCCAEDR on the neonatal brain 3D-MRI record is significantly higher on each surface because the larger external CSF spaces in neonatal brains can ease the problem of identifying non-brain tissue.

The experimental results show that even when 10% Gaussian noise is added to the neonatal brain 3D-MRI record, DCCAEDR still has the better DR with adequate denoising capability. Since two abilities of exploration and exploitation will be well balanced, DCCAEDR can combine the advantages of both local and global segmentation. When the contour is close to different surface boundaries, the local intensity fitting exploitation becomes dominant, which attracts the contour toward object boundaries and, finally, stops the contour. Consequently, DCCAEDR can recover the boundaries of WM, GM, and CSF accurately.

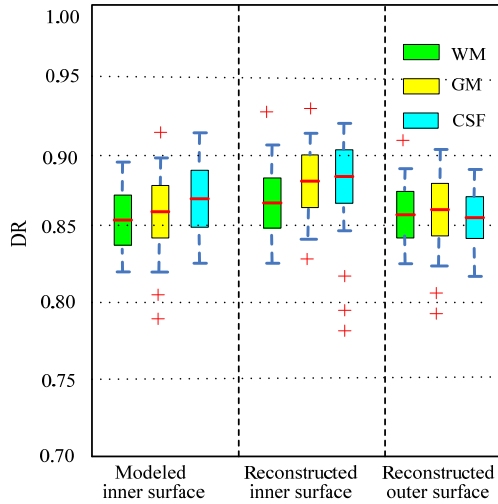


Fig. 9. Quantitative comparisons of three kinds of cortical surfaces on neonatal brain 3D-MRI record with 10% Gaussian noise.

Furthermore, in order to obtain the quantitative comparison of DCCAEDR's advantage for the segmentation of neonatal brain 3D-MRI records with 10% Gaussian noise, Table VI gives three kinds of different quantitative segmentation accuracy values of cortical surfaces using DCCAEDR and three compared methods. The results further confirm that DCCAEDR provided high-performing segmentation of complex neonatal brain 3D-MRI records as

well as 3D-MRI detail preservation, and it exhibits the segmentation improvement of neonatal brain 3D-MRI records with 10% Gaussian noise.

TABLE VI
SEGMENTATION ACCURACY COMPARISONS OF NEONATAL BRAIN 3D-MRI RECORD WITH 10% GAUSSIAN NOISE

	References			
	Smith et al. [56] (%)	Battaglini et al. [57] (%)	Wang et al. [59] (%)	DCCAEDR (%)
Modeled inner surfaces	89.43	87.15	90.32	92.29
Reconstructed inner surfaces	87.39	81.23	87.90	89.87
Reconstructed outer surfaces	80.45	78.23	82.21	84.32
Avg.	85.76	82.20	86.48	88.83

Based on the conducted experiments, the proposed DCCAEDR accelerator has higher efficiency and robustness when it can be applied to dealing with the attribute reduction for traditional Chinese medical records and the segmentation of large-scale neonatal brain 3D-MRI records with complex noise in real-world big datasets. It can achieve the most desirable performance and satisfactory accuracy. The results can assist medical experts to realize the diagnosis using the Chinese medicine and brain 3D-MRI diseases, and to design best course of medical treatment.

VIII. DISCUSSION AND CONCLUSIONS

With an increasing amount of scientific and industrial datasets, extracting useful information from big data is a growing requirement. Enlarged data make the algorithms that have been based on rough set theory a challenging task. As this paper shows, there are few parallel algorithms to handle attribute reduction for big data. But noise is one of the primary sources of uncertainty in today's applications, and most of them are sensitive to noise in big datasets. Even though few of them can improve their accuracy, such improvement is not so obvious. Consequently, accuracy and stability are the focus points in need of improvement. In this paper, we put forward a new attribute equilibrium dominance reduction accelerator (DCCAEDR) based on distributed co-evolutionary cloud model to solve the optimization problem of attribute reduction with complex noise in big datasets.

Why is DCCAEDR superior to the compared related methods? There are two reasons: the framework of N -populations distributed co-evolutionary MapReduce model (DCMM) and the novel equilibrium dominance strategy (EDSCE) of elitists under the N bounded rationality regions. These are adopted to enhance the attribute reduction performance in big datasets with complex noise. DCCAEDR can produce a superior global optimization performance with a diverse and uniformly distributed solution set using the DCMM model, particularly when compared to three representative co-evolutionary algorithms (CoBRA, DECC-G and CCNSGAI). DCCAEDR is capable of grouping

interacting variables with great accuracy for the majority of nonseparable functions in order to keep interdependencies to be a minimum, and thus it will achieve a dynamic balance between exploring and exploiting inherent structures of large-scale global optimization functions. Meanwhile, on ten number-magnified UCI datasets, DCCAEDR outperforms the representative attribute reduction algorithms (PACCA, E-FSA and FVPRS) through extensive simulation studies, including an efficiency study and an accuracy study. The elitists using the EDSCE strategy are driven towards neighbor Pareto dominance. They construct an entire solution with a speedup factor for the Nash equilibrium optimum, which will enhance DCCAEDR's robustness against complex noise and accelerate its stability with the increasing numbers of datasets, while PACCA, E-FSA and FVPRS performed poorly in attaining the classification results. Our research is expected to shed light on how cooperative co-evolution can be incorporated into MapReduce and EGT works during the attribute cooperative co-evolutionary reduction.

In addition, DCCAEDR has been applied to the attribute reduction for TCM records and segmentation of the inner and outer cortical surface of neonatal brain 3-D MRI records. These promising applications further demonstrate the significant application performance of DCCAEDR. Hence, we strongly recommend the DCCAEDR as a good choice for dealing with large-scale and complicated datasets in real-world applications.

In future work, we will continue to develop parallel MapReduce speed-up by considering a fair selection of partners among distributed co-evolutionary subpopulations. Another target will focus on DCCAEDR's relation to different applying problem characteristics and extending the competitive complementary strategy to minimize the influence of complex noise in big datasets.

ACKNOWLEDGEMENTS

The authors would like to thank Professor Nikhil R. Pal in the Electronics and Communication Science Unit of the Indian Statistical Institute, for providing very helpful suggestions, and also would like to thank Mr. Paul D. Found from United Kingdom for his time to proof reading this paper. The authors express the sincere appreciation to six anonymous reviewers and Associate Editor for their insightful comments which greatly improve the quality of this paper.

REFERENCES

- [1] W. Pedrycz, *Granular Computing: analysis and design of intelligent systems*. CRC Press/Francis Taylor, Boca Raton, 2013.
- [2] Z. Pawlak, "Rough sets," *Int. J. Comput. Inform. Sci.*, vol. 11, no. 5, pp. 341-356, 1982.
- [3] Z. Pawlak, and A. Skowron, "Rough sets: Some extensions," *Inform. Sci.*, vol. 177, pp. 28-40, Jan. 2007.
- [4] Z. Pawlak, "Rough sets and intelligent data analysis," *Inform. Sci.*, vol. 147, pp. 1-12, Nov. 2002.
- [5] Z. Pawlak, and A. Skowron, "Rudiments of rough sets," *Inform. Sci.*, vol. 177, pp. 3-27, Jan. 2007.
- [6] F. Fernandez-Riverola, F. Diaz, and J. M. Corchado, "Reducing the Memory Size of a Fuzzy Case-Based Reasoning System Applying Rough Set Techniques," *IEEE Trans. Syst., Man, Cybern. C, Appl., Rev.* vol. 37, no. 1, pp. 138-146, Jan. 2007.
- [7] M. Q. Ye, X. D. Wu, X. G. Hu, and D. H. Hu, "Knowledge reduction for decision tables with attribute value taxonomies," *Knowl. Based Syst.*, vol. 56, pp. 68-78, Jan. 2014.
- [8] C. L. Philip Chen, and C.Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Inform. Sci.*, vol. 275, pp. 314-347, Aug. 2014.
- [9] J. Dean, and S. Ghemawat, "MapReduce: simplified data processing on large clusters," in *Proc. of Oper. Syst. Design and Implementation*, 2004, pp. 137-150.
- [10] J. Dean, and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107-113, Jan. 2008.
- [11] J. Ekanayake, S. Pallickara, and G. Fox, "MapReduce for data intensive scientific analyses," in *Proc. 4th IEEE Int. Conf. eScience*, 2008, pp. 277-284.
- [12] J. Berlińska, and M. Drozdowski, "Scheduling divisible MapReduce computations," *J. Parallel Distr. Com.*, vol. 71, no. 3, pp. 450-459, Mar. 2011.
- [13] Q. H. Hu, Z. X. Xie, and D. R. Yu, "Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation," *Pattern Recognit.*, vol. 40, pp. 3509-3521, Dec. 2007.
- [14] Z. Q. Meng, and Z. Z. Shi, "Extended rough set-based attribute reduction in inconsistent incomplete decision systems," *Inform. Sci.*, vol. 204, pp. 44-69, Oct. 2012.
- [15] F. Wang, J. Y. Liang, and C. Y. Dang, "Attribute reduction for dynamic data sets," *Appl. Soft Comput.*, vol. 13, pp. 676-689, Jan. 2013.
- [16] Y. H. Qian, J. Y. Liang, W. Pedrycz, and C.Y. Dang, "Positive approximation: an accelerator for attribute reduction in rough set theory," *Artif. Intell.*, vol. 174, pp. 597-618, Jun. 2010.
- [17] S. Y. Zhao, E. C. C. Tsang, and D. G. Chen, "The model of fuzzy variable precision rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 2, pp. 451-467, Apr. 2009.
- [18] J. Y. Liang, F. Wang, C. Y. Dang, and Y. H. Qian, "An efficient rough feature selection algorithm with a multi-granulation view," *Int. J. Approx. Reason.*, vol. 53, no. 6, pp. 912-926, Sep. 2012.
- [19] C. Z. Wang, Q. He, D. G. Chen, and Q. H. Hu, "A novel method for attribute reduction of covering decision systems," *Inform. Sci.*, vol. 254, pp. 181-196, Jan. 2014.
- [20] M. Q. Ye, X. D. Wu, X. G. Hu, and D. H. Hu, "Knowledge reduction for decision tables with attribute value taxonomies," *Knowl. Based Syst.*, vol. 56, pp. 68-78, Jan. 2014.
- [21] D. Slezak, J. Wroblewski, "Order based genetic algorithms for the search of approximate entropy reducts," in *Proc. 9th Int. Conf. Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, 2003, pp. 608-621.
- [22] B. B. Zhou, Z. R. Feng, and Z. G. Ren, "An efficient ant colony optimization approach to attribute reduction in rough set theory," *Pattern Recognit. Lett.*, vol. 29, pp. 1351-1357, Jul. 2008.
- [23] D. Y. Ye, Z. J. Chen, and J. K. Liao, "A new algorithm for minimum attribute reduction based on binary particle swarm optimization with vaccination," in *Proc. 11th Pacific-Asia Conf. Knowledge Discovery and Data Mining*, 2007, pp. 1029-1036.
- [24] A. R. Hedar, J. Wang, and M. Fukushima, "Tabu search for attribute reduction in rough set theory," *Soft Comput.*, vol. 12, no. 9, pp. 909-918, Jul. 2008.
- [25] W. P. Ding, J. D. Wang, Z. J. Guan, and Q. Shi, "Enhanced minimum attribute reduction based on quantum-inspired shuffled frog leaping algorithm," *J. Syst. Eng. Electron.*, vol. 24, no. 3, pp. 426-434, Jun. 2013.
- [26] J. B. Zhang, T. R. Li, D. Ruan, Z. Z. Gao, and C. B. Zhao, "A parallel method for computing rough set approximations," *Inform. Sci.*, vol. 194, pp. 209-223, Jul. 2012.
- [27] Y. Yang, Z. R. Chen, Z. Liang, and G.Y. Wang, "Attribute reduction for massive data based on rough set theory and MapReduce," in *Proc. 5th Int. Conf. Rough Sets and Knowledge Technology*, 2010, pp. 672-678.
- [28] J. Qian, P. Lv, X. D. Yue, C. H. Liu, and Z. J. Jing, "Hierarchical attribute reduction algorithms for big data using MapReduce," *Knowl. Based Syst.*, vol. 73, pp. 18-31, Jan. 2015.
- [29] 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.
- [30] J. Qian, D. Q. Miao, Z. H. Zhang, and X. D.Y. ue, "Parallel attribute reduction algorithms using MapReduce," *Inform. Sci.*, vol. 279, pp. 671-690, Sep. 2011.
- [31] R. Chandra, M. Frean, M. J. Zhang, and C. W. Omlin, "Encoding subcomponents in cooperative coevolutionary recurrent neural networks," *Neurocomputing*, vol. 74, no. 17, pp. 3223-3234, Oct. 2011.

- [32] R. A. Krohling, and L. dos Santos Coelho, "Coevolutionary particle swarm optimization using Gaussian distribution for solving constrained optimization problems," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 6, pp. 1407-1416, Dec. 2006.
- [33] X. D. 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.
- [34] C. J. Lin, C. H. Chen, and C. T. Lin, "A hybrid of cooperative particle swarm optimization and cultural algorithm for neural fuzzy networks and its prediction applications," *IEEE Trans. Syst., Man, Cybern. C, Appl., Rev.*, vol. 39, no. 1, pp. 55-68, Jan. 2009.
- [35] J. Derrac, I. Triguero, S. Garcia, and F. Herrera, "Integrating instance selection, instance weighting, and feature weighting for nearest neighbor classifiers by coevolutionary algorithms," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 5, pp. 1383-1397, Oct. 2012.
- [36] F. Legillon, A. Liefoghe, and E. G. Talbi, "CoBRA: A cooperative coevolutionary algorithm for bi-level optimization," in *Proc. IEEE World Congr. Comput. Intell.*, 2012, pp. 1-8.
- [37] Z. Y. Yang, K. Tang, and X. Yao, "Large scale evolutionary optimization using cooperative coevolution," *Inform. Sci.*, vol. 178, no. 15, pp. 2985-2999, Aug. 2008.
- [38] M. N. Omidvar, X. D. Li, and X. Yao, "Cooperative co-evolution with delta grouping for large scale non-separable function optimization," in *Proc. IEEE Congr. Evol. Comput.*, 2010, pp. 1762-1769.
- [39] J. C. 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.
- [40] Y. Mei, X. D. Li, and X. Yao, "Cooperative coevolution with route distance grouping for large-scale capacitated arc routing problems," *IEEE Trans. Evol. Comput.*, vol. 83, no. 3, pp. 435-449, Jun. 2014.
- [41] R. Subbu, and A. C. Sanderson, "Modeling and convergence analysis of distributed coevolutionary algorithms," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 2, pp. 806-822, Apr. 2004.
- [42] R. P. Wiegand, W. C. Liles, and K. A. De Jong, "Analyzing cooperative coevolution with evolutionary game theory," in *Proc. of IEEE Congr. on Evol. Comput.*, 2002, pp. 1600-1605.
- [43] S. G. Ficici, O. Melnik, and J. B. Pollack, "A game-theoretic and dynamical systems analysis of selection methods in coevolution," *IEEE Trans. Evol. Comput.*, vol. 9, no. 6, pp. 580-601, Dec. 2005.
- [44] W. B. Liu, and X. J. Wang, "An evolutionary game based particle swarm optimization algorithm," *J. Comp. Appl. Math.*, vol. 214, no. 1, pp. 30-35, Apr. 2008.
- [45] A. Koh, "An evolutionary algorithm based on Nash dominance for equilibrium problems with equilibrium constraints," *Appl. Soft Comput.*, vol. 12, no. 1, pp. 161-173, Jan. 2012.
- [46] K. Razi, S. H. Shahri, and A. R. Kian, "Finding Nash equilibrium point of nonlinear non-cooperative games using coevolutionary strategies," in *Proc. of the Seventh Int. Conf. on Intell. Syst. Design and Appl.*, 2007, pp. 875-880.
- [47] M. Staudigl, "Co-evolutionary dynamics and Bayesian interaction Games," *Int. J. Game Theory*, vol. 42, no. 1, pp. 179-210, Feb. 2013.
- [48] T. White, *Hadoop: The Definitive Guide*, 2nd edition. O'Reilly Media /Yahoo Press, 2010.
- [49] X. D. Li, K. Tang, M. Omidvar, Z. Yang, and K. Qin, "Benchmark Functions for the CEC'2013 Special Session and Competition on Large Scale Global Optimization," Technical Report, Evol. Comput. and Machine Learning Group, RMIT Univ., Australia, Jun. 21, 2013.
- [50] B. Dorransoro, G. Danoy, A. J. Nebro, and P. Bouvry, "Achieving super-linear performance in parallel multi-objective evolutionary algorithms by means of cooperative coevolution," *Comput. Oper. Res.*, vol. 40, no. 6, pp. 1552-1563, Jun. 2013.
- [51] J. Y. Liang, F. Wang, C. Y. Dang, and Y. H. Qian, "An efficient rough feature selection algorithm with a multi-granulation view," *Int. J. Approx. Reason.*, vol. 53, no. 6, pp. 912-926, Sep. 2012.
- [52] S. Y. Zhao, E. C. C. Tsang, and D. G. Chen, "The model of fuzzy variable precision rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 2, pp. 451-467, Apr. 2009.
- [53] S. Lukman, Y. He, and S. C. Hui, "Computational methods for traditional Chinese medicine: a survey," *Comput. Meth. Prog. Bio.*, vol. 88, no. 3, pp. 283-294, Dec. 2007.
- [54] X. Z. Zhou, S. B. Chen, B. Y. Liu, R. S. Zhang, Y. H. Wang, P. Li, Y. F. Guo, H. Zhang, Z. Y. Gao, and X. F. Yand, "Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support," *Artif. Intell. Med.*, vol. 48, no. 2-3, pp. 139-152, Feb.-Mar. 2010.
- [55] Y. Wang, L. Z. Ma, and P. Liu, "Feature selection and syndrome prediction for liver cirrhosis in traditional Chinese medicine," *Comput. Meth. Prog. Bio.*, vol. 95, no. 3, pp. 249-257, Sep. 2009.
- [56] S. M. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapp.*, vol. 17, no. 3, pp. 143-155, Nov. 2002.
- [57] M. Battaglini, S. M. Smith, S. Brogi, and N. D. Stefano, "Enhanced brain extraction improves the accuracy of brain atrophy estimation," *Neuroimage*, vol. 40, no. 2, pp. 583-589, Apr. 2008.
- [58] L. P. Zhou, Y. P. Wang, Y. Li, P. T. Yap, and D. G. Shen, "Hierarchical anatomical brain networks for MCI prediction: Revisiting volumetric measures," *PLoS ONE*, vol. 6, no. 7, pp. 1-14, Jul. 2011.
- [59] L. Wang, F. Shi, G. Li, Y. Gao, W. Lin, J. H. Gilmore, and D. Shen, "Segmentation of neonatal brain MR images using patch-driven level sets," *NeuroImage*, vol. 84, pp. 141-158, Jan. 2014.
- [60] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297-302, Jul. 1945.



Wei-Ping Ding received the B.S. degree in Compute Science and Technology, Nantong University, Nantong, China, in 2002, the M.S. degree in Software Engineer from Soochow University, SuZhou, China, in 2005, and the Ph.D. degree in Computer Application, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013. He was a visiting researcher at Department of Mathematics & Computer Science, University of Lethbridge, Lethbridge, Alberta, Canada. In 2014, he was a Postdoctoral Researcher at the Brain Research Center, National Chiao Tung University (NCTU) with Professor Chin-Teng Lin, Hsinchu, Taiwan. Now, he is an Associate Professor in School of Computer Science and Technology, Nantong University, Nantong, China. He is a member of Association of Computing Machinery (ACM), IEEE Computer Society (IEEE-CS), and China Computer Federation (CCF). He has authored or co-authored more than 60 papers in journals and conference proceedings. His current research interests included machine learning, data mining, granular computing, and their applications in big data.

Dr. Ding was a recipient of National Natural Science Young Foundation of China in 2013. He was awarded an excellent-young teacher of Jiangsu Province sponsored by Qing Lan Project, Jiangsu Province, China, in 2014.



Chin-Teng Lin (S'88-M'91-SM'99-F'05) received the B.S. degree from National Chiao-Tung University (NCTU), Hsinchu, Taiwan, in 1986, and 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 Chair Professor of Electrical and Computer Engineering, and the Director of the Brain Research Center, NCTU. He has published more than 120 journal papers in the areas of neural networks, fuzzy systems, multimedia hardware/software, and cognitive neuro-engineering, including approximately 74 IEEE

journal papers. He is the co-author of Neural Fuzzy Systems (Prentice-Hall) and the author of Neural Fuzzy Control Systems with Structure and Parameter Learning.

Dr. Lin is an Editor-in-Chief of the IEEE TRANSACTIONS ON FUZZY SYSTEMS. He served on the Board of Governors at the IEEE Circuits and Systems Society from 2005 to 2008, the IEEE Systems, Man, Cybernetics Society from 2003 to 2005, the IEEE Computational Intelligence Society from 2008 to 2010, and was a Chair of the IEEE Taipei Section from 2009 to 2010. He served as the Deputy Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II from 2006 to 2008.

Dr. Lin has own the Outstanding Research Award granted by National Science Council (NSC), Taiwan, since 1997 to present, the Outstanding Professor Award granted by the Chinese Institute of Engineering (CIE) in 2000, and the 2002 Taiwan Outstanding Information-Technology Expert Award. He was also elected to be one of 38th Ten Outstanding Rising Stars in Taiwan (2000). He was honored with Outstanding Electrical and Computer Engineer (OECE), Purdue University in 2011.

Dr. Lin was elevated to IEEE Fellow in 2005 for contributions to biologically inspired information systems.

Zhi-Jin Guan received the B.S. Degree in Mathematics, University of Harbin Normal University, Harbin, China, in 1986, and Ph.D. degree in Computer Application, Nanjing



University of Aeronautics and Astronautics, Nanjing, Jiangsu, China, in 2008. Now he is a professor in School of Computer Science and Technology, Nantong University, Nantong, China. From 2003 to 2004, he was a visiting researcher at Colorado School of Mines, Colorado, USA. In 2011, he was a visiting

researcher at Department of Mathematics & Computer Science, University of Lethbridge, Lethbridge, Alberta, Canada. His research interests include reversible computation, intelligence system and information security. He has authored or co-authored more than 80 papers in journals and conference proceedings.



Mukesh Prasad received his Master's degree in Computer Application (MCA) from Jawaharlal Nehru University (JNU), New Delhi, India, in 2009 and currently pursuing Ph.D degree in Department of Computer Science at National Chiao-Tung University, Hsinchu, Taiwan. He has published

several journal and international conference papers. His current research interest includes machine learning, big data, pattern recognition, fuzzy systems and neural networks.



Sen-Bo Chen received the B.S. degree in Computer Science and Technology, Nantong University, Nantong, China, in 2003, the M.S. degree in Computer Application from Yangzhou University, Yangzhou, China, in 2009, He is currently pursuing the Ph. D. degree in Computer Application, Nanjing University of Aeronautics and Astronautics, Nangjing,

China. He is a member of ACM, IEEE-CS and CCF. His current research interests include service computing, social network, data mining and machine learning.