

# A Layered-Coevolution-Based Attribute-Boosted Reduction Using Adaptive Quantum Behavior PSO and Its Consistent Segmentation for Neonates Brain Tissue

Weiping Ding, *Member, IEEE*, Chin-Teng Lin, *Fellow, IEEE*, Mukesh Prasad, *Member, IEEE*, Zehong Cao, and Jiandong Wang

**Abstract**—The main challenge of attribute reduction in large data applications is to develop a new algorithm to deal with large, noisy, and uncertain large data linking multiple relevant data sources, structured or unstructured. This paper proposes a new and efficient layered-coevolution-based attribute-boosted reduction algorithm (LCQ-ABR\*) using adaptive quantum behavior particle swarm optimization (PSO). First, the quantum rotation angle of an evolutionary particle is updated by a dynamic change of self-adapting step size. Second, a self-adaptive partitioning strategy is employed to group particles into different memplexes, and the quantum-behavior mechanism with the particles' states depicted by the wave function cooperates to achieve superior performance in their respective memplexes. Third, a new layered co-evolutionary model with multi-agent interaction is constructed to decompose a complex attribute set, and it can self-adapt the attribute sizes among different layers and produce the reasonable decompositions by exploiting any interdependency among multiple relevant attribute subsets. Fourth, the decomposed attribute subsets are evolved to compute the positive region and discernibility matrix by using their best quantum particles, and the global optimal reduction set is induced successfully. Finally, extensive comparative experiments are provided to illustrate that LCQ-ABR\* has better feasibility and effectiveness of attribute reduction on large-scale and uncertain dataset problems with complex noise, compared with

representative algorithms. Moreover, LCQ-ABR\* can be successfully applied in the consistent segmentation for neonatal brain 3D-MRI, and the consistent segmentation results further demonstrate its stronger applicability.

**Index Terms**—Attribute-boosted reduction, adaptive quantum behavior PSO, layered-coevolution with multi-agent interaction, consistent segmentation for neonates brain tissue, sulci and gyrus estimate.

## I. INTRODUCTION

Attribute reduction is an important issue that has retained high interest in machine learning, data mining, and pattern recognition. Its aim is to discover a minimal feature subset from a problem domain while retaining high accuracy and efficiency in representing the original data [1]. Rough set theory (RST) is a very efficient and useful mathematical tool that can handle some information and knowledge with uncertainty, imprecision, and vagueness [2] [3]. Introducing RST makes attribute reduction much more popular in modeling and propagating uncertainty and vagueness. When we perform attribute reduction using RST, the main goal is to find the minimum attribute set and induce minimal length decision rules inherent in the information system with affordable algorithmic complexity and computational cost [4][5][6]. So, it usually refers to the preferred technique for data preprocessing in data mining and knowledge discovery [7]-[11].

In recent years, various attribute reduction algorithms and general frameworks for their unification have been discussed. Ke *et al.* [12] introduced ant colony optimization to the attribute reduction process to investigate a fast and effective approximation algorithm. This algorithm updated the pheromone trails of the edges connecting each two different attributes of the best-so-far solution, and used a rapid optimization procedure to construct candidate solutions. So, it had the ability to rapidly find solutions with very small cardinality. But this algorithm was unable to effectively deal with large-scale datasets. Yeung *et al.* [13] proposed the generalization of fuzzy rough sets, which defined the upper and lower approximation operators by using arbitrary fuzzy relations, and characterized different classes of generalized upper and lower approximation operators of fuzzy sets by different sets of axioms. This generalization was applied to a fuzzy reasoning system, and the results demonstrated it had a wider range of applications. But this model of fuzzy rough sets was sensitive to noisy real-valued attributes. Li *et al.* [14]

---

Manuscript received August 17, 2016; revised February 9, 2017; accepted June 8, 2017. This work was supported by the National Natural Science Foundation of China under Grant 61300167, Natural Science Foundation of Jiangsu Province under Grant BK20151274, Sponsored by Qing Lan Project of Jiangsu Province, Jiangsu Provincial Government Scholarship Program under Grant JS-2016-065, Six Talent Peaks Project of Jiangsu Province under Grant XYDXXJS-048, and Applied Basic Research Program of Nantong under Grant GY12016014. (*Corresponding author: Weiping Ding*)

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

C.-T. Lin is with the Computational Intelligence and Brain-Computer Interface (CI-BCI) Lab, CAI, University of Technology Sydney, Ultimo NSW 2007, Australia and also with the Institute of Electrical Control Engineering and Brain Research Center, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: chintenglin@gmail.com).

M. Prasad is with the School of Software, faculty of Engineering and Information Technology, University of Technology Sydney, Australia (e-mail: mukesh.prasad@uts.edu.au).

Z. Cao is with the Computational Intelligence and Brain-Computer Interface (CI-BCI) Lab, CAI, University of Technology Sydney, Ultimo NSW 2007, Australia (e-mail: zhcaonctu@gmail.com).

J. Wang is with College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: aics@nuaa.edu.cn).

proposed a neighborhood based decision-theoretic rough set model in which the positive region related attribute reduction and its minimum cost were analyzed. This model can overcome the shortcomings of the current decision-theoretic rough set model and obtain a short reduction set with competitive classification ability. Qian *et al.* [15] exploited a new framework structure to speed up the computation of equivalence classes and attribute significance by parallelizing the traditional attribute reduction process based on the MapReduce mechanism. This parallel attribute reduction algorithm performed efficiently on massive data. So, the paradigm with the MapReduce technique had good feasibility to facilitate big datasets. Maji *et al.* [16] put forward the IT2 fuzzy-rough set-based attribute selection method, where the lower and upper relevance and significance of attributes were defined for IT2 fuzzy approximation spaces, and then attributes were selected by maximizing the relevance and significance. This method had the good utility effectiveness for fuzzy attribute reduction. Real-world datasets are now available everywhere from the Web, sensor networks, social networks, and proprietary databases, which often link multiple relevant data sources, structured or unstructured. Moreover, the scale of datasets increases dynamically with time variation. So, an abundance of staggeringly complex large datasets has been produced [17] [18]. Most of the above-mentioned algorithms are inadequate and unreliable for attribute reduction for these datasets due to their ever-greater volume, complexity and diversity of structures. Besides, noise is one of the main sources of uncertainty in applications, and it has also been shown that traditional attribute reduction operators are not robust to noise. Hence, there is an urgent need for new and effective attribute reduction algorithms to remove irrelevant and redundant datasets while retaining the optimum salience in these complex large-scale datasets.

In the past few years, the quantum-inspired evolutionary algorithm (QEA) has attracted significant attention from researchers by using a Q-bit as the probabilistic representation, without numeric or symbolic representation. This appears to be a much better characterization of population diversity, and thus this representation has the strong advantage of denoting the linear superposition of evolutionary states [19]. It has been empirically and theoretically demonstrated that QEA always gains better performance than the traditional evolutionary algorithm (EA) [20][21][22]. We know that the implementation of attribute reduction algorithms on large and complex datasets is very time-consuming due to the dramatic increases in the number of attributes because of complex, fast-changing relationships between big data objects. Therefore, to propose an effective and efficient attribute reduction algorithm based on the superiority of QEA becomes a significant and urgent challenge. Some work is required to solve this problem. First, some effective quantum-inspired evolutionary operators and mechanisms are designed to enhance attribute reduction of datasets with high-dimension. Second, the interacting decision variables among various attribute sets can be handled well to achieve better performance in practical applications. In addition, some limitations of existing co-evolution structures can be addressed to decompose large-scale attribute sets by using the dynamic adaptation of a reorganization model.

This study proposes a new and efficient layered-coevolution-based attribute-boosted reduction algorithm (LCQ-ABR\*) using quantum behavior PSO. It aims to choose attribute subsets including strongly relevant and non-redundant attributes for large-scale, noisy, and uncertain datasets linking multiple relevant data sources. The main flowchart of LCQ-ABR\* is shown in Fig. 1. Its performance is extensively evaluated to solve large-scale and uncertain dataset problems with complex noise on some well-known benchmark datasets. The experimental results illustrate that LCQ-ABR\* has better feasibility and effectiveness than some representative algorithms to which it is compared. Moreover, LCQ-ABR\* is applied to the consistent segmentation for neonatal brain 3D-MRI, and it can achieve better feasibility and effectiveness in complex neonatal brain tissue.

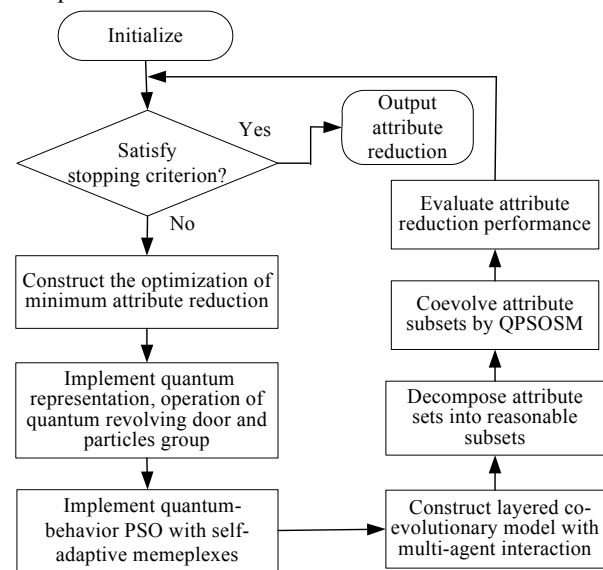


Fig. 1. Flowchart of LCQ-ABR\*

We briefly state key contributions of the work presented in this paper below.

- 1) The quantum rotation angle of evolutionary particles can be updated by the dynamic change of the self-adapting step size. A new self-adaptive partitioning strategy is employed to group particles into memplexes, and the quantum-behavior mechanism of particles cooperates better in their respective memplexes to achieve superior performance. These new quantum operators aim to strengthen the adaptive stability of particle memplexes for attribute reduction in large-scale datasets.
- 2) A new layered co-evolutionary model with multi-agent interaction is constructed to decompose attribute sets, and it can self-adapt the attribute sizes among different layers and produce reasonable decompositions by neighborhood vectors among multiple relevant attribute subsets. It adapts the dynamic stability of co-evolutionary particle behavior to achieve a complementary cumulative distribution. So, it boosts the optimal performance of attribute reduction.
- 3) LCQ-ABR\* is applied to the longitudinal cortical surface labeling of a neonatal brain 3D-MRI. It can be adaptive to derive from atlas surfaces and the cortical folding geometries, and to achieve the temporal consistency term adaptive to the similarities of cortical folding. This successful application is

expected to dramatically scale up attribute reduction algorithms for large-scale datasets in terms of efficiency and feasibility.

This paper is organized as follows. First, the optimization model of minimum attribute reduction is constructed in section II. Section III introduces the quantum-behavior PSO with self-adaptive memplexes (QPSOSM), which illustrates the quantum rotation angle, self-adaptive partitioning memplexes strategy and quantum-behavior mechanism of particles. A layered co-evolutionary model with multi-agent interaction (LCMMI) for attribute reduction is presented in section IV. In section V, the main steps of LCQ-ABR\* are stated in detail. Extensive experimental evaluation and discussion are provided in section VI. In section VII, the application performances of LCQ-ABR\* are assessed in the consistent segmentation for neonates brain tissue 3D-MRI. Section VIII presents some discussions of experimental results. Section IX provides the conclusion.

## II. OPTIMIZATION MODEL OF MINIMUM ATTRIBUTE REDUCTION

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

*Definition 1:* For any concept  $X \subseteq U$  and attribute subset  $R \subseteq A$ ,  $X$  can be approximated by the  $R$ -lower approximation  $\underline{R}X$  and  $R$ -upper approximation  $\overline{R}X$ , with

$$\underline{R}X = \{x \in U \mid [x]_R \subseteq X\} \quad (1)$$

$$\text{and } \overline{R}X = \{x \in U \mid [x]_R \cap X \neq \emptyset\}. \quad (2)$$

$\underline{R}X$  is the set of objects of  $U$  that will surely belong to  $X$ , whereas  $\overline{R}X$  is the set of objects of  $U$  that can possibly belong to  $X$ .

*Definition 2:* When  $C$  is a set of condition attributes,  $D$  is a set of decisions, and  $A = C \cup D$  ( $C \cap D = \emptyset$ ), the information system  $S = (U, A, V, f)$  is called a decision table. The  $C$ -positive region of  $D$  is the set of all objects from the universe  $U$  that can be classified with certainty into classes of  $(U/D)$  employing attributes from  $C$ , that is

$$POS_C(D) = \bigcup_{X \in U/D} \underline{C}X. \quad (3)$$

The criterion for attribute reduction is the degree of dependency  $\gamma_C(Q)$ , which is defined as

$$k = \gamma_C(Q) = \frac{|POS_C(Q)|}{|U|}. \quad (4)$$

*Definition 3:* For attribute reduction, a reduction is defined as a subset  $R$  of conditional attribute set  $C$  that is satisfied with  $\gamma_R(D) = \gamma_C(D)$ . The reduction set is given by

$$RED = \{R \subseteq C \mid \gamma_R(D) = \gamma_C(D), \forall B \subset R, \gamma_B(D) \neq \gamma_C(D)\}. \quad (5)$$

*Definition 4:* Suppose  $\{0, 1\}^m$  is the  $m$ -dimensional approximation space and  $\xi$  is the mapping from  $\{0, 1\}^m$  to the power set  $2^C$ , denoted as

$$x_i = 1 \Leftrightarrow a_i \in \xi(x) \quad i = 1, \dots, m, a_i \in C. \quad (6)$$

The optimization model of minimum attribute reduction is defined as

$$f(x) = \min(S(x))$$

$$s.t. \begin{cases} x \in \{0, 1\}^m \\ \gamma_{\xi(x)}(D) = \gamma_C(D) \\ \forall q \in \xi(x), \gamma_{\xi(x) \setminus \{q\}}(D) = \gamma_{\xi(x)}(D) \end{cases} \quad (7)$$

$$\text{subject to } 0 \leq S(x) = \sum_{i=1}^m x_i \leq m.$$

## III. QUANTUM-BEHAVIOR PSO WITH SELF-ADAPTIVE MEMPLEXES

In this section, we propose the quantum-behavior PSO with self-adaptive memplexes (QPSOSM) for attribute-boosted reduction. Assume that the PSO evolutionary system is a quantum system and each evolutionary individual is described by a particle in the quantum space. Then, the rationales for using this hybrid combination of PSO and quantum technology are as follows.

First, evolutionary particles are represented by multi-state quantum bits, which represent a linear superposition of states in the particle search space probabilistically, in order to increase the diversity of evolutionary particles.

Second, the adaptive updating strategy of the quantum rotation angle is adopted to update the operation of a quantum revolving door. This strategy speeds the search process for evolutionary particles in different memplexes, so that they can keep the balance between global search and local refinement during attribute reduction.

Third, a self-adaptive partitioning strategy is employed to group particles into different memplexes, and the state of a particle is represented as the wave function, instead of position and velocity in standard PSO. Each particle moves with the potential in each dimension through the establishment of a delta trap, which will greatly accelerate the evolution convergence. Two abilities of exploration and exploitation are well balanced to achieve better performance at attribute reduction.

### A. Adaptive Updating Strategy of Quantum Rotation Angle

QPSOSM uses a new representation of a quantum bit (Q-bit), where one Q-bit is denoted as a pair of complex numbers  $(\alpha, \beta)$  such as

$$|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad (8)$$

where  $\alpha$  and  $\beta$  are complex numbers that specify the probability amplitudes of corresponding states. The quantities

$|\alpha|^2$  and  $|\beta|^2$  give the probabilities that the Q-bit is found in the “0” or “1” state, respectively. The normalization condition of their states to unity guarantees that

$$|\alpha|^2 + |\beta|^2 = 1. \quad (9)$$

The quantum gate (Q-gate) is used for the rotation gate according to the relationship

$$U(\theta) = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{bmatrix}. \quad (10)$$

The  $i^{\text{th}}$  Q-bit is updated as shown such that

$$\begin{bmatrix} \alpha'_i \\ \beta'_i \end{bmatrix} = U(\theta) \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}, \quad (11)$$

where  $\theta_i$  is the rotation angle toward either the “0” or “1” state, depending on its objective sign.

For a decision variable  $x = (x_1, x_2, \dots, x_n)$ , where  $x_i \in [lower, upper], i = 1, 2, \dots, n$ , the quantum rotation angle is updated by the adaptive strategy as follows:

**Algorithm 1: Adaptive updating strategy of quantum rotation angle**

- 1) Each  $x_i$  is mapped into  $[0, 1]$  according to the following equation

$$x_i = \frac{x_i - lower}{upper - lower}, i = 1, 2, \dots, n. \quad (12)$$

- 2) Implement the quantum real coding to get  $x'$  as

$$x' = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \sqrt{1-(x_1)^2} & \sqrt{1-(x_2)^2} & \dots & \sqrt{1-(x_n)^2} \end{pmatrix}. \quad (13)$$

- 3) For some Q-bit  $i$  of  $x'$ , compute the rotation angle  $\varphi = \eta \times d(i, j)$ ,  $\eta \in rand(0.6, 0.9)$ , where  $d(i, j)$  is the random rotation direction.  $p'_j (j = 1, 2, \dots, n)$  is the state of the  $t^{\text{th}}$  iteration, defined as

$$p'_j = \begin{pmatrix} \alpha'_{11} & \alpha'_{12} & \dots & \alpha'_{1l} & \alpha'_{21} & \alpha'_{22} & \dots & \alpha'_{2l} & \dots & \alpha'_{m1} & \alpha'_{m2} & \dots & \alpha'_{ml} \\ \beta'_{11} & \beta'_{12} & \dots & \beta'_{1l} & \beta'_{21} & \beta'_{22} & \dots & \beta'_{2l} & \dots & \beta'_{m1} & \beta'_{m2} & \dots & \beta'_{ml} \end{pmatrix}. \quad (14)$$

- 4) Get  $x'_i$  as

$$x'_i = \sin(\varphi) \times x_i - \cos(\varphi) \times \sqrt{1-(x_i)^2}. \quad (15)$$

- 5) Update  $x'_i = \sqrt{\frac{(x'_i + x_i)}{2}}$  if the Q-bit is out of  $[0, 1]$ .
- 6) Repeat above operation until each Q-bit is within  $[0, 1]$ .

**B. New Strategy for Partitioning Particles into Self-adaptive Memeplexes**

The proposed kernel strategy for partitioning particles combines the geometric partitioning method and the memeplexes model that is used to represent “networks of particles.” Particles cooperate better in their respective memeplexes and achieve superior performance. The

partitioning method aims to group particles in one vicinity in the same memeplex. First, the center position of each memeplex is selected randomly, and then the particles are grouped according to their geometric distance from the best center particle of each memeplex, as illustrated in Fig. 2. To better represent the attribute reduction solutions, the particles in each partitioned memeplex are used to optimize their corresponding attribute variables. Fig. 3 shows the graphical procedure of the refined attribute reduction solution. Each memeplex is assigned  $n$  attribute variables representing the  $n$  candidate solutions. The main steps are as follows.

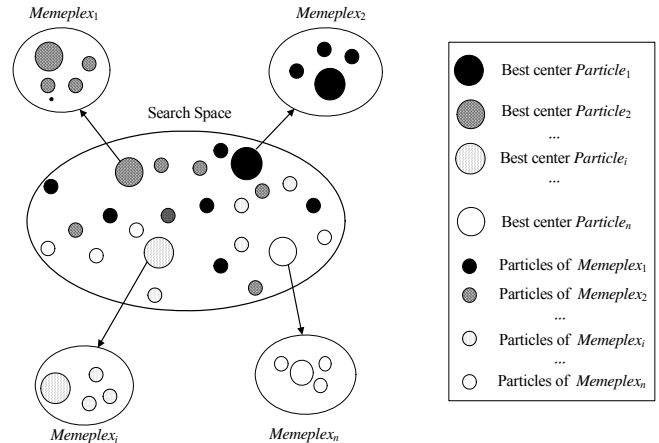


Fig. 2. Schematic diagram for partitioning particles into memeplexes

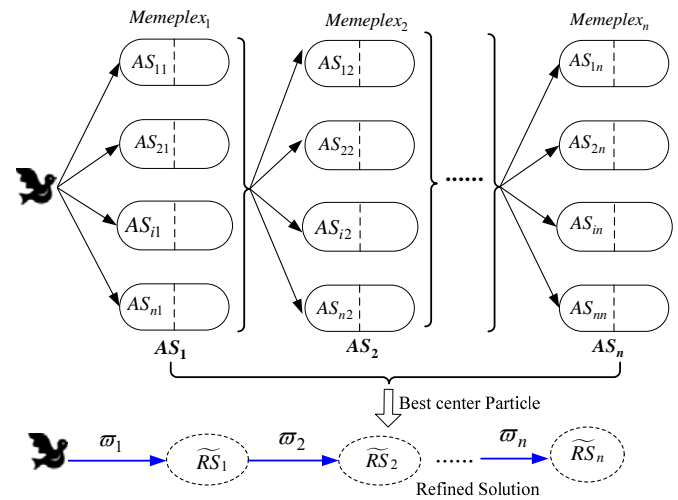


Fig. 3. Graphical representation of refined attribute reduction solution

**Algorithm 2: Self-adaptive memeplex of partitioning particles for attribute reduction**

- 1) Create evolutionary particles, initialize them in the quantum population space, and construct the memeplex of partitioning particles as follows. First, calculate the distance between center position 1 and all other particles, denoted as  $l_1$ . If the particles have less distance than  $l_1$ , they are assigned to  $Memeplex_1$ . Second, calculate the distance between center position 2 and all remaining particles, denoted as  $l_2$ . If the particles have less distance than  $l_2$ , they are assigned to the  $Memeplex_2$ . Continue this process until all particles are assigned to  $Memeplex_3, \dots, Memeplex_n$ , respectively.
- 2) Consider the  $n$  best particles with the best fitness values in

the respective  $n$  memplexes as the center position of their corresponding memplexes.

- 3) Along with the particles' partitioning, if the local best fitness of most memplexes reaches the same target area, the result is nearly the same even though the particles learn from any memplex. If we continue to partition particles, the consumption will be increased. So, we will merge the relevant memplexes into a new memplex according to the diversity value of the local best fitness of memplexes. Diversity is computed as

$$diversity(t) = \frac{1}{cn(t)}, \quad (16)$$

where  $cn(t)$  represents the sum of the difference numbers in each dimension that meet the threshold set  $\varepsilon$  between the local best fitness of the memplex in the  $t^{\text{th}}$  iteration and the average fitness from the 1<sup>st</sup> to the  $(t-1)^{\text{th}}$  iteration. It can be defined as

$$cn(t) = \sum_{k=1}^{n-1} f(lbest_k(t)) \quad (17a)$$

$$f(lbest_k(t)) = \begin{cases} 1, & \forall j \left| lbest_{kj}(t) - \overline{lbest_j(t)} \right| < \varepsilon \\ 0, & \text{otherwise.} \end{cases} \quad (17b)$$

$$\overline{lbest_j(t)} = \frac{\sum_{k=1}^{n-1} lbest_{kj}(t)}{n-1}, \quad (17c)$$

where  $t = 1, 2, \dots, n$ ,  $\overline{lbest_j(t)}$  represents the average fitness from the 1<sup>st</sup> to the  $(t-1)^{\text{th}}$  iteration.

If  $diversity(t) < 1/[c(n-1)]$ , where  $c$  is a uniform random number and  $c \in [0, 1]$ , the relevant memplexes will be merged.

- 4) A particle path starts from the nest (denoted as variable  $AS_{11}$ ), passes through nodes for variables  $AS_{12}, \dots, AS_{1(n-1)}$ , and stops at a node for variable  $AS_{1n}$ . A particle path includes  $n$  edges, and each edge will construct a solution component in the solution vector  $AS_i$ .
- 5) The credit assignment for partitioning particles is performed at the memplexes. It is computed by estimating how well the best center particle in the  $i^{\text{th}}$  memplex performs relative to its competitors in cooperating with other particles. In the  $i^{\text{th}}$  memplex, the best center particle is assigned credit  $\varpi_i$  based on the equation

$$\varpi_i = \left| \frac{f_{CP_i} - \overline{f_{Mem_i}}}{f_{CP_i}} \right|, \quad (18)$$

where  $f_{CP_i}$  is the fitness of the  $i^{\text{th}}$  best center particle and  $\overline{f_{Mem_i}}$  is the average local best fitness of the  $i^{\text{th}}$  memplex.

- 6) Set memplex  $i$  for representing attribute subset  $i$ , and best particle  $i$  for representing attribute  $i$ . For  $j=1$  to the  $i^{\text{th}}$  memplex  $|S_i|$  do

- (a) Assemble the complete solution with  $S_{ij}$  (the  $j^{\text{th}}$  particle of  $S_i$ ) and representative from the other memplexes;
- (b) Assign Pareto rank to  $AS_{ij}$ ;
- (c) Calculate the niche count of  $S_{ij}$ ;
- (d) Update the best center particle of  $AS_i$ .

- 7) Assigned a credit  $\Gamma_i$  to the best center particle of  $AS_i$ , and update the archive of the nondominated solution vectors as follows:

$$\widetilde{RS}_1 = \Gamma_1 \times AS_1; \quad \widetilde{RS}_2 = \Gamma_2 \times AS_2; \quad \dots$$

$$\widetilde{RS}_i = \Gamma_i \times AS_i; \quad \dots \quad \widetilde{RS}_n = \Gamma_n \times AS_n.$$

$$\text{So, } AS = \bigcup_{i=1}^n (\Gamma_i \times \widetilde{RS}_i).$$

- 8) Check elimination criterion if the maximum iteration number has reached the termination condition, otherwise continue with **Step 5**.

Using the proposed strategy of partitioning particles, each memplex has a vicinity filled by particles within a close distance of each other, with credit assigned for the best center particle. It can provide a way of complementing the diversity preservation, and can produce more diverse particles across the different memplexes in the quantum space.

### C. Quantum-Behavior Mechanism of QPSOSM

QPSOSM adopts a new quantum-behavior mechanism to guarantee the global convergence of PSO. The state of a particle in the quantum space is represented by the wave function  $\psi(x, t)$ , and each particle moves with  $\delta$  potential in each dimension through the establishment of a delta trap. By this wave function  $\psi(x, t)$ , the particle's position of a wave function and its probability density function are achieved. So, the quantum-behavior searching mechanism is deduced, and its steps are described as follows:

#### Algorithm 3: Quantum-behavior mechanism of QPSOSM

- 1) At the number of iterations  $t$ , the  $i^{\text{th}}$  particle in the  $j^{\text{th}}$  dimension quantum space has the attractor  $p_{ij}^t$ . The wave function at iteration  $t+1$  can be expressed as

$$\psi(x_{ij}^{t+1}) = \frac{1}{\sqrt{H_{ij}^t}} \exp(-|x_{ij}^t - p_{ij}^t|/H_{ij}^t), \quad (19)$$

where  $H_{ij}^t$  is the standard deviation of the double exponential distribution, and it will change with the number of iterations  $t$ .

- 2) Construct the Lévy flight model for quantum-behavior particles in the quantum space as

$$p_{\alpha, \beta}(k, \mu, \sigma) = \exp \left[ i\mu k - \sigma^\alpha |k|^\alpha (1 - i\beta \frac{k}{|k|} \delta(k, \alpha)) \right], \quad (20)$$

where  $k$  is the number of particles in each memplex,

$\mu$  is the displacement,  $\sigma$  is the measure,  $\alpha$  is the characteristic exponent,  $\beta$  is the skewness, and we set  $\alpha = 0.5, \beta = 1$ .

- 3) The probability density function with double exponential distribution  $Q$  is

$$Q(x_{ij}^{t+1}) = \text{Lévy}(\tilde{\lambda}) \cdot |\psi(x_{ij}^t)|^2 = \text{Lévy}(\tilde{\lambda}) \cdot \frac{1}{H_{ij}^t} \exp(-2|x_{ij}^t - p_{ij}^t|/H_{ij}^t) \quad (21)$$

where  $\text{Lévy}(\tilde{\lambda})$  denotes the random search vector of the Lévy distribution with parameters  $\tilde{\lambda}$  ( $1 < \tilde{\lambda} \leq 3$ ).

- 4) Compute the corresponding probability distribution function as

$$F(x_{ij}^{t+1}) = 1 - \exp(-2|x_{ij}^t - p_{ij}^t|/H_{ij}^t). \quad (22)$$

- 5) Implement the strategy of self-adapting step size  $\lambda$  as follows:

$$\lambda = U \left( k \times b^{\frac{g-1}{g_{\max}-1}}, b^{\frac{g-1}{g_{\max}-1}} \right), \quad (23)$$

where  $U(\cdot, \cdot)$  is a uniform random number,  $k$  is the number of particles in each memplex,  $g$  is the current iteration number, and  $g_{\max}$  is the total number of iterations. The parameter  $\lambda$  can adjust the step size dynamically near the better solutions and enhance the searching ability of particles as follows:

- (a) When  $g$  is quite small,  $\lambda$  is very large and the operation can achieve a comprehensive fast global search. The searching range of global search is  $[0, k]$ ;
- (b) When  $g$  gradually increases to  $g_{\max}$ ,  $\lambda$  can self-adaptively reduce to a small value and the improvement of local searching ability can be realized by reducing the step size. The range of the local search is  $[0, b]$ .

- 6) Obtain particle  $x_i$  in the  $j^{\text{th}}$  dimension quantum space and the  $(t+1)^{\text{th}}$  iteration, and update its position as follows:

$$x_{ij}^{t+1} = p_{ij}^t \pm \frac{1}{2} \lambda \cdot H_{ij}^t \cdot \ln(1/u_{ij}^t), \quad (24a)$$

$$H_{ij}^t = 2\alpha |c_j^t - x_{ij}^t|, \quad (24b)$$

where  $u_{ij}^t$  is a uniformly distributed random number in  $(0,1)$ , the parameter  $\alpha$  is set as  $\alpha < 1.782$  to guarantee convergence of the particle in each memplex, and  $\alpha$  can be controlled to decrease linearly from  $\alpha_0$  to  $\alpha_1$ , where  $\alpha_0, \alpha_1$  ( $\alpha_0 < \alpha_1$ ) are two uniformly distributed random numbers that can be selected from the range  $[0.7, 1.5]$ , and the values of  $\alpha_0, \alpha_1$  express the degree of relative importance of particle  $x_i$  and self-adapting step size  $\lambda$  in the updating process.

- 7) Design the global point of the particle population in the quantum space, denoted as  $C$ , which is defined as the average position of the best center particle among each

memplex. Its equation is

$$C = (c_1^t, c_1^t, \dots, c_j^t) = \left( \frac{1}{M} \sum_{i=1}^M x_{i1}^t, \frac{1}{M} \sum_{i=1}^M x_{i2}^t, \dots, \frac{1}{M} \sum_{i=1}^M x_{ij}^t \right), \quad (25)$$

where  $M$  is the number of updated memplexes and  $x_{ij}^t$

is the best position of particle  $i$  in the  $j^{\text{th}}$  dimension quantum space at the  $(t+1)^{\text{th}}$  iteration.

By the above-mentioned QPSOSM, the quantum rotation angle of evolutionary particle dimension in the quantum space is updated by the dynamic change of the self-adapting step size, and the quantum-behavior mechanism with the particles' states depicted by the wave function cooperates to achieve superior performance in their respective memplexes. Hence, QPSOSM will strengthen the adaptive stability of particle memplexes for attribute reduction in large-scale datasets.

#### IV. LAYERED CO-EVOLUTIONARY MODEL WITH MULTI-AGENT INTERACTION FOR ATTRIBUTE REDUCTION

To solve the optimization problem of large-scale attribute reduction, a layered co-evolutionary model with multi-agent interaction (LCMMI) is constructed to adaptively partition the multiple-relevance attribute sets into different subsets by avoiding dependency on prior domain knowledge. The model can keep the interdependencies of multiple relevance attribute sets to a minimum to achieve a dynamical balance between exploring and exploiting inherent structures of attribute sets.

Each agent corresponds to a neighborhood vector with different resolutions, and can capture the multiple-relevance attribute decision variables and group them together in one memplex. It allows a better approximation of the contribution of various neighborhood vectors, uncovers the underlying interaction structure of attribute decision variables, and optimizes attribute-boosted reduction.

This LCNRH constructs a hierarchical framework of neighborhood vectors for partitioning an attribute set as described in Fig. 4. The main steps are as follows:

First, according to factor  $h$  of the partition, each agent is divided into such parts as "Head ( $h$ )" and "Tail ( $t=(n-1)*h+1$ )". Thus, the attribute in each layer is divided into  $m$  agents.

Second, to map an attribute set into different layers, the computation of neighborhood vector interactions is initially conducted on Layer1, and then propagated to other layers effectively via the distribution neighborhood vector that indicates the relationship of neighborhood vectors from Layer2, Layer3, and Layer4.

Third, generate four interaction neighborhood vectors in  $L^4$  as  $\mathbf{R}_p^4$ ,  $\mathbf{R}_t^4$ ,  $\mathbf{R}_n^4$ , and  $\mathbf{R}_m^4$ , and decompose  $\mathbf{R}_p^4$ ,  $\mathbf{R}_t^4$ ,  $\mathbf{R}_n^4$ , and  $\mathbf{R}_m^4$  into four vector subsets as

$$\begin{aligned} \mathbf{R}_p^4 &= [R_p^1, R_p^2, R_p^3, R_p^4]^T, \mathbf{R}_t^4 = [R_t^1, R_t^2, R_t^3, R_t^4]^T, \\ \mathbf{R}_n^4 &= [R_n^1, R_n^2, R_n^3, R_n^4]^T, \mathbf{R}_m^4 = [R_m^1, R_m^2, R_m^3, R_m^4]^T. \end{aligned} \quad (26)$$

The layered co-evolutionary optimization strategy with multi-agent interaction is the most important strategy of the LCMMI model. Its main procedure is shown in Fig. 5. The pair-wise neighborhood vector iteration within layers and

cascade between layers are considered not only for attribute reduction with the same scale in the same layer, but for attribute reduction across different scales in different layers as well. So, this strategy ensures that the underlying similarity of

information between any pair of neighborhood vectors in the same layer can be fully reflected and results in better generalization ability. Its detailed investigation is included as follows.

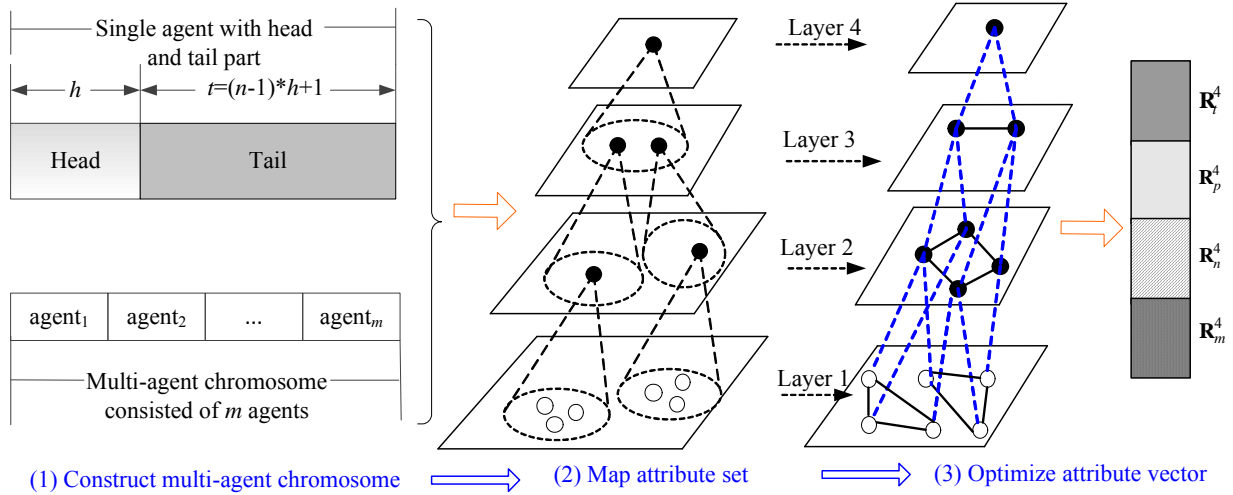


Fig. 4. Hierarchy framework of neighborhood vector for partitioning attribute set

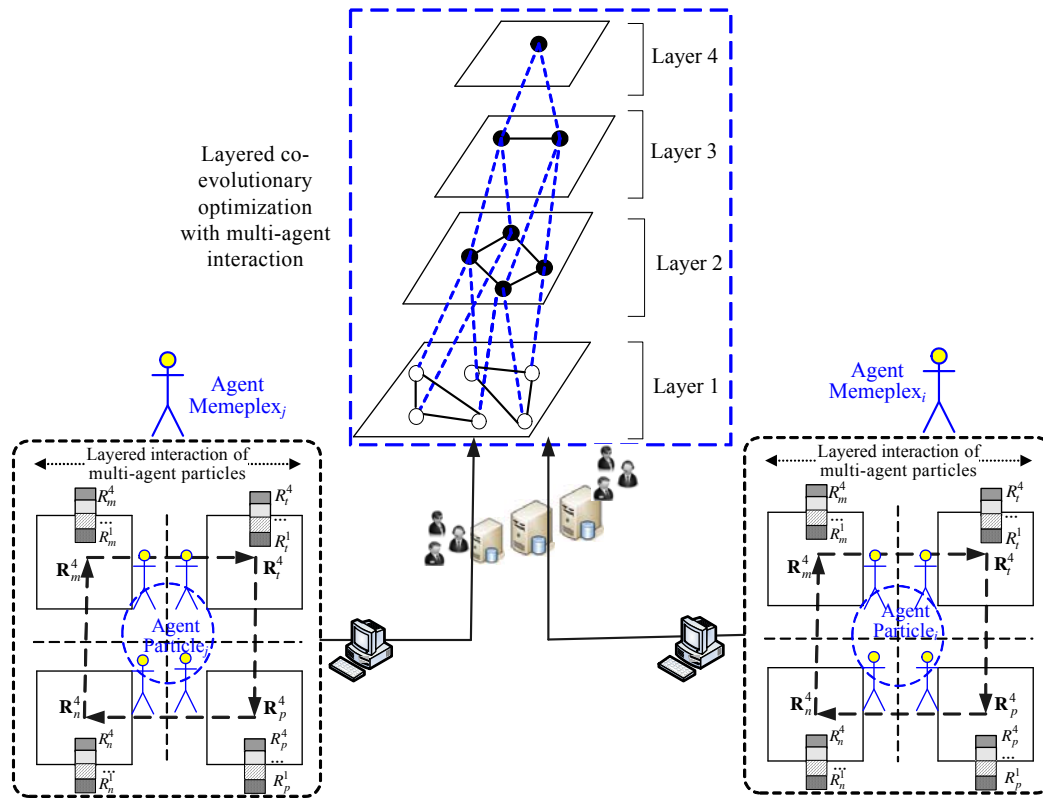


Fig. 5. Layered co-evolutionary optimization with multi-agent interaction

---

**Algorithm 4. Layered co-evolutionary optimization with multi-agent interaction**

---

- 1) Construct the four-layers structure based on the self-organizing neighborhood vector, in which the  $i^{\text{th}}$  layer corresponds to the efficient solution of the  $i^{\text{th}}$  neighborhood vector  $R^i$ .
- 2) Obtain the  $N^4 \times N^4$  matrix as  $C^4$ , where  $N$  is the number of neighborhood vectors in Layer4. The Pearson correlation between attribute vectors  $f_i$  and  $f_j$  is

$$C^4 = \sum_{i,j=1}^4 \text{corr}(f_i, f_j), \quad (27)$$

where  $f_i, f_j$  denote the volumetric ratios of attribute vectors of the  $i^{\text{th}}$  and  $j^{\text{th}}$  neighborhood vector in Layer4, respectively.

- 3) Construct the membership decomposition matrix  $M^i$ , which includes  $i$  rows and  $N^4$  columns. Each row corresponds to the simple neighborhood and each column corresponds to the single neighborhood in Layer1.
- 4) Calculate the interaction margin value of any two interaction neighborhood vectors in Layer4 as

$$\text{corr}(R_t^4, R_m^4) = \frac{1}{a \times b} \sum_{R_t^4 \in S_t^4} \sum_{R_m^4 \in S_m^4} \text{corr}(R_t^4, R_m^4), \quad (28)$$

where  $R_t^4$  and  $R_m^4$  represent the attribute neighborhood vector in Layer4, including  $a$  and  $b$  neighborhood vectors, respectively.  $S_t^4$  and  $S_m^4$  are two simple neighborhood sets that comprise  $R_t^4$  and  $R_m^4$ , respectively.

- 5) Compute the cross energy between the Agent\_Memplex $_i$  and Agent\_Particle $_j$  as

$$\begin{aligned} & \text{Energy}(\text{Agent\_Memplex}_i) \\ &= \frac{1}{2 * n} \left( 0.6 \times \sum_{j=1}^n \text{same}(x_{\text{Agent\_Memplex}_i}, x_{\text{Agent\_Particle}_j}) \right) \\ & \quad + 1.4 \times \sum_{k=1}^n \text{same}(f_{\text{Agent\_Memplex}_i}, f_{\text{Agent\_Particle}_k}) \end{aligned} \quad (29a)$$

where  $x_{\text{Agent\_Memplex}_i}$  is the best position of Agent\_Memplex $_i$ ,  $x_{\text{Agent\_Particle}_j}$  is the best position of Agent\_Particle $_j$ ,  $f_{\text{Agent\_Memplex}_i}$  is the best fitness of Agent\_Memplex $_i$ , and  $f_{\text{Agent\_Particle}_j}$  is the best fitness of Agent\_Particle $_j$ . The function same( ) is calculated as

$$\text{same}(x,y) = \begin{cases} 0, & x = y \\ 1, & \text{otherwise.} \end{cases} \quad (29b)$$

- 6) Construct the energy proximity matrix **EPM** of multi-agent interaction as

$$\mathbf{EPM} = \begin{bmatrix} En_1 & \pi_1^2 & \cdots & \pi_1^n \\ \pi_2^1 & En_2 & \cdots & \pi_2^n \\ \vdots & \vdots & \cdots & \vdots \\ \pi_n^1 & \pi_n^2 & \cdots & En_n \end{bmatrix}, \quad (30)$$

where  $En_i = \text{Energy}(\text{Agent\_Memplex}_i)$  and

---

$$\pi_i^j = \frac{f_{\text{Agent\_Memplex}_i}}{f_{\text{Agent\_Particle}_j}}.$$

- 7) Achieve the ensemble set of neighborhood vectors in the  $i^{\text{th}}$  layer as

$$\mathfrak{R} = \bigcup_{i=1}^N \left( R_i^i \otimes R_p^i \otimes R_n^i \otimes R_m^i \times \left| \frac{En_i}{\mathbf{EPM}} \right| \right), \quad (31)$$

where the symbol  $\otimes$  denotes the vector matrix product.

---

V. PROPOSED LCQ-ABR\* ALGORITHM

Based on the above-mentioned quantum-behavior PSO with self-adaptive memplexes (QPSOSM) and layered co-evolutionary model with multi-agent interaction (LCMMI), this paper proposes a new and efficient layered-coevolution-based attribute-boosted reduction algorithm (LCQ-ABR\*) using adaptive quantum behavior PSO. The decomposed attribute subsets are co-evolved by their best quantum particles with QPSOSM, respectively. The optimal reduction subset of each attribute subset can be easily obtained, and then the global optimal reduction set is also induced successfully. Its main steps are as follows:

---

**Algorithm 5: LCQ-ABR\***

---

- 1) Set up a searching space of  $m$  dimensions for the attribute reduction in the quantum space, and construct the co-evolutionary memplexes, in which each memplex represents its corresponding attribute subset.
- 2) Partition particles into different memplexes using **Algorithm 2**, and map each memplex into one condition attribute subset that is limited to the defined space of attribute reduction by

$$\text{Weight}_i = \frac{\text{Weight}_i - \text{Weight}_{\min}}{\text{Weight}_{\max} - \text{Weight}_i}. \quad (32)$$

- 3) Construct the optimization object of minimum attribute reduction as  $F(x) = \min(S(x))$ .
  - 4) Calculate the lower  $\underline{\gamma}_{A_i}(D)$  and upper  $\overline{\gamma}_{A_i}(D)$  relevance of each attribute  $A_i \in C$ , and then select out the most relevant attribute  $A_i$  with the highest lower relevance value  $\underline{\gamma}_{A_i}(D)$ .
  - 5) Generate a quantum particle chromosome by giving  $|\alpha|^2$  and  $|\beta|^2$  in each Q-bit individual. Encode the quantum evolutionary particle's position as the subset of condition attribute set  $C$ .
  - 6) Make quantum particle states  $Q(t)$ , where the observation individuals corresponding to the  $i^{\text{th}}$  Q-bit particle are represented by  $(X'_{i1}, X'_{i2}, \dots, X'_{im})$ .
  - 7) Update  $Q(t)$  to  $Q(t+1)$  by using the adaptive updating strategy of quantum rotation angle in **Algorithm 1**. Conduct the proposed quantum-behavior mechanism for best center particles in the memplexes by using **Algorithm 3**.
  - 8) Decompose attribute sets into reasonable subsets with the hierarchy framework of neighborhood vector, and carry out the layered co-evolution with multi-agent interaction
-



by using **Algorithm 4**. Both within- and between-layer neighborhood vector interactions are performed to achieve the ensemble set of neighborhood vectors.

- 9) **Do** {
- (a) Select attribute subsets  $(Sub\_attribute)_i$  with parallel model;
  - (b) Calculate the lower significance of  $A_j$  ( $A_j \in C$ ) with respect to each selected attribute  $A_i \in S$  by 
$$\sigma_{\{A_i, A_j\}}(D, A_j) = \gamma_{\{A_i, A_j\}}(D) - \gamma_{A_i}(D);$$
  - (c) Calculate the degree of dependency  $\gamma_R(D)$  as comparison criterion for attribute reduction;
  - (d) Remove  $A_j$  from  $C$  if  $\sigma_{\{A_i, A_j\}}(D, A_j) = 0$  for any attribute  $A_i \in S$ ;
  - (e) From the remaining attributes of  $C$ , repeat similar steps of the upper significance until the desired number of attributes is selected out or  $C = \emptyset$ ;
  - (f) Calculate the fitness  $Fit(x)$ , and select out the local best reduction subset in every memplex;
  - (g) Achieve the minimal solution of reduction subset  $R_i^{best}$ .
- } **While** (the stopping criterion is satisfied).
- 10) Evaluate whether attribute reduction accuracy is satisfied with the stopping criterion.
- If** satisfied, Output the entire minimum attribute reduction 
$$RED_{Emin} = \bigcup_{i=1}^n R_i^{best}.$$
- Otherwise**, Go to **Step 7** and continue to implement the attribute reduction procedures.

## VI. EXPERIMENTAL STUDIES

The objective of the following experiments is to show the effectiveness and efficiency of LCQ-ABR\* compared with traditional algorithms. To support data-intensive distributed applications, our experiments run on the open platform Apache Hadoop. We implement all algorithms on a cluster with ten nodes, and each node is provided with 128 GB main memory and an AMD Opteron Processor 2376 with 2 Quad-Core CPUs. For distributed experiments of large-scale datasets, one is configured as the master node and the rest are set as slave nodes. The operating system in these machines uses Linux CentOS 6.5 Kernel 3.18. We have chosen 16 benchmark datasets whose characteristics are summarized in Table 1. There are three protein datasets (Pancreatic, Colorectal and LiverACO), two metabolism datasets (LiverM and LiverTOD), four NIPS 2003 feature selection challenge datasets (Arcene, Dexter, Dorothea, and Madelon), four public microarray datasets (Colon, Prostate, Leukemia, and Lung-cancer), and three biomedical data-sets (Breast-cancer, Ovarian-cancer, and Hiva). These 16 benchmark datasets cover a wide range of real-world application domains, including protein, metabolism clinical image, gene expression, ecology, text categorization, and molecular biology; sample sizes (from 62 to 4,229); and features (from 115 to 100,000). Hence, they are significant

challenges to the construction and reduction of attribute sets by using some attribute reduction algorithms. For the four NIPS 2003 challenge datasets, we employ the originally provided training sets and validation sets; for four public microarray datasets, we randomly adopt the first 3/5 samples for training and the last 2/5 for testing; and for the rest we take ten-fold cross-validation. The software used is Microsoft Visual Studio 2015 and Visual C# 6.0. For the following results, we present the average testing values of 30 runs.

TABLE I  
CHARACTERISTICS OF 16 BENCHMARK DATASETS

Dataset	#Sample	#Feature	Dataset	#Sample	#Feature
1. Pancreatic	181	664	9. Madelon	2,000	500
2. Colorectal	112	779	10. Colon	62	2,000
3. LiverACO	975	129	11. Prostate	102	6,033
4. LiverM	126	115	12. Leukemia	72	7,129
5. LiverTOD	280	866	13. Lung -cancer	181	12,533
6. Arcene	100	10,000	14. Breast -cancer	286	17,816
7. Dexter	300	20,000	15. Ovarian -cancer	216	2,190
8. Dorothea	800	100,000	16. Hiva	4,229	1,617

In this experiment, the attribute reduction performance of the proposed LCQ-ABR\* algorithm is conducted to compare with such representative algorithms as FRFS [5], FRRFDM [9], and TDNEC [14]. Table 2 summarizes the comparative results of reduction running time (Time) and space consumption (Space) for FRFS, TDNEC, FRRFDM, and LCQ-ABR\*, respectively. The superscript symbol “<sup>†</sup>” means that the corresponding result is significantly best, “—” indicates no trial can reach an acceptable solution, and bold indicates that the mean value is greater than those of the other three algorithms.

As described in Table 2, LCQ-ABR\* typically shows the highest reduction speed, compared with FRFS, TDNEC, and FRRFDM. For the Arcene dataset, the reduction time by LCQ-ABR\* is 3.76 s, while the reduction times by FRFS, TDNEC, and FRRFDM are 7.78 s, 6.54 s, and 6.08 s, respectively. LCQ-ABR\* can significantly improve the reduction running time. In addition, LCQ-ABR\* needs lower space consumption. For example, LCQ-ABR\* spends 54.06% space consumption in the Dorothea dataset and 73.09% in the Breast-cancer dataset of FRRFDM. Similar results are evident in most datasets. Therefore, the experimental results indicate that, LCQ-ABR\* can find a feasible attribute reduction subset in much less time and space, compared with FRFS, TDNEC, and FRRFDM.

Moreover, we observe the variation of average misclassification cost on four kinds of datasets (Pancreatic, Dorothea, Leukemia, and Breast-cancer) with an added attribute noise rate. We assume  $n_{NI}$  is the number of objects classified incorrectly,  $n_{ND}$  is the number of objects with deferment decisions,  $\mu_{CP}$  is the cost for classifying an object into the negative region when it belongs to the positive region, and  $\mu_{CP}$  is the cost for classifying an object into a boundary region. The misclassification cost can be calculated as follows:

$$MCF\_cost = n_{ND} \cdot \mu_{CP} + n_{NI} \cdot \mu_{CP} \quad (33)$$

Fig. 6 presents the average comparison results of misclassification cost and its corresponding running time of attribute reduction. The  $x$ -axis contains different levels (2.0%~12.0%) of incremental noise rate, and the  $y$ -axis shows the variation of updating misclassification cost (MCF\_cost) and CPU time. We can observe the variation of MCF\_cost and CPU time of three algorithms as the level of noise rate increases. This indicates that the noise has a great effect on the MCF\_cost performance. But we can see from Fig. 6 that the value variation of LCQ-ABR\* is small in most cases. Taking the Dorothea dataset (one of the NIPS 2003 feature selection challenge datasets) as an example, when the level of noise rate is from 8.0% to 10.0%, the variation of MCF\_cost of LCQ-ABR\* is 1.98. And when the level of noise rate is from 10.0% to 12.0%, the variation of MCF\_cost is 1.63. But for TDNEC, the variation of MCF\_cost is larger. Despite its appealing performance, FRRFDM is also dominated by LCQ-ABR\* in most cases throughout our experiments. Furthermore, with the levels of noise rate dynamically increasing, the efficiency of LCQ-ABR\* becomes obvious, and it can achieve satisfactory results. Similar behaviors also hold for the other datasets.

The results clearly demonstrate that LCQ-ABR\* can lead to an appealing attribute reduction performance in these challenge datasets with real-world applications. This is not surprising, because LCQ-ABR\* can capture strong dependency and complex structures associated with attributes more accurately, and it can greatly eliminate irrelevant and redundant attributes without losing performance accuracy. There is a tradeoff between speed and quality. In effect, the reduction set of relevant and significant attributes can be much more stably obtained by using LCQ-ABR\* as the scale of these challenges datasets becomes larger.

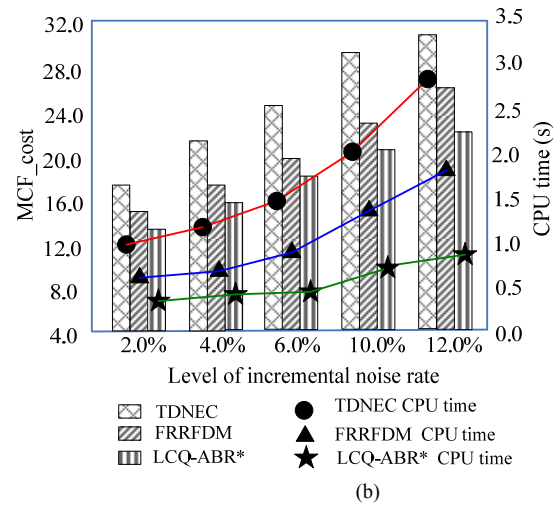
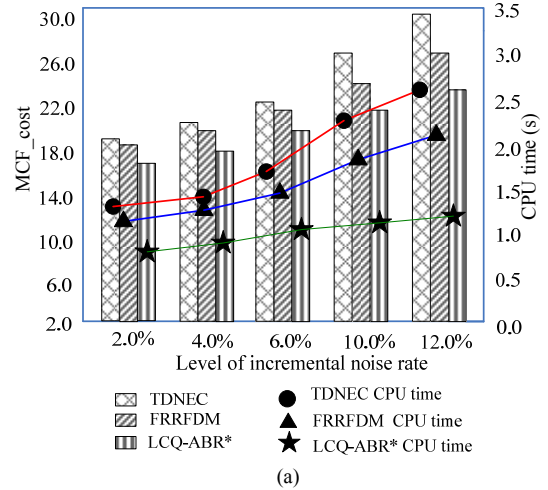


TABLE II  
TIME AND SPACE COMPARISONS OF FOUR ALGORITHMS ON 16 BENCHMARK DATASETS

Datasets	Time (s)				Space (M)			
	FRFS	TDNEC	FRRFDM	LCQ-ABR*	FRFS	TDNEC	FRRFDM	LCQ-ABR*
1. Pancreatic	0.89	0.98	0.76	<b>0.57</b>	1.12	1.65	0.87	<b>0.68</b> <sup>†</sup>
2. Colorectal	0.78	0.86	0.73	<b>0.65</b>	1.27	1.42	1.00	<b>0.88</b>
3. LiverACO	0.80	0.98	0.73	<b>0.69</b>	0.97	0.89	0.78	<b>0.51</b> <sup>†</sup>
4. LiverM	1.13	1.21	1.07	<b>0.72</b> <sup>†</sup>	1.39	1.09	1.05	<b>0.84</b> <sup>†</sup>
5. LiverTOD	3.32	2.98	<b>1.95</b>	2.17	1.58	1.50	<b>1.25</b>	1.38
6. Arcene	7.78	6.54	6.08	<b>3.76</b> <sup>†</sup>	4.84	4.14	3.96	<b>3.02</b>
7. Dexter	—	11.36	8.89	<b>6.58</b> <sup>†</sup>	—	8.08	7.17	<b>6.21</b>
8. Dorothea	41.16	36.34	34.43	<b>25.97</b> <sup>†</sup>	29.89	28.26	19.68	<b>10.64</b> <sup>†</sup>
9. Madelon	12.01	11.21	10.32	<b>9.87</b>	6.57	8.78	6.80	<b>5.79</b>
10. Colon	18.34	17.45	16.32	<b>14.50</b>	9.41	10.30	8.63	<b>7.16</b>
11. Prostate	—	19.56	17.32	<b>13.11</b> <sup>†</sup>	—	12.11	10.97	<b>7.39</b> <sup>†</sup>
12. Leukemia	21.23	20.09	<b>18.77</b>	19.11	12.57	10.12	<b>9.18</b>	9.34
13. Lung-cancer	29.78	—	21.89	<b>20.10</b>	23.89	—	18.84	<b>13.38</b>
14. Breast-cancer	31.90	28.56	26.35	<b>19.09</b> <sup>†</sup>	19.87	18.78	16.54	<b>12.09</b> <sup>†</sup>
15. Ovarian-cancer	10.45	9.32	8.27	<b>6.92</b>	8.90	7.09	6.80	<b>5.77</b>
16. Hiva	8.46	9.78	7.80	<b>5.28</b> <sup>†</sup>	6.75	5.89	5.48	<b>4.28</b> <sup>†</sup>

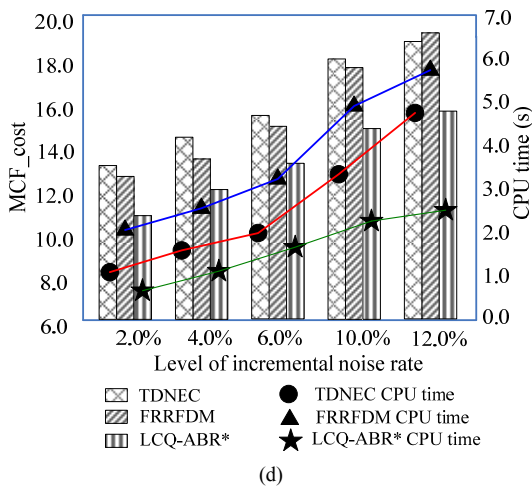
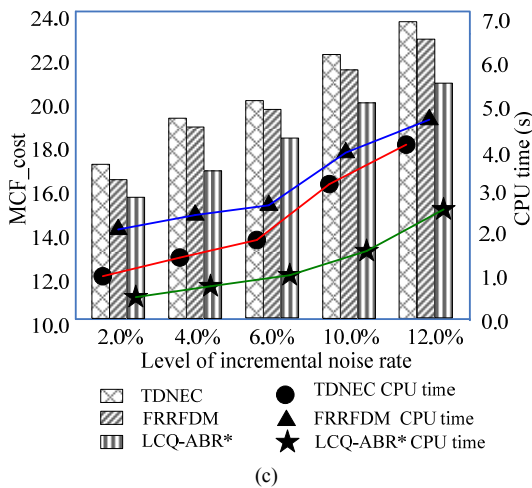


Fig. 6. Performance comparisons of three algorithms with different levels of incremental noise rate on (a) Pancreatic, (b) Dorothea, (c) Leukemia, and (d) Breast-cancer.

LCQ-ABR\* can overcome the nuanced challenges in most large-scale and complex datasets, and it can achieve the tradeoff between high efficiency and accuracy of attribute reduction. The main reasons for these advantages of LCQ-ABR\* are as follows: The layered co-evolutionary model with multi-agent interaction can decompose large-scale datasets quickly, and it can self-adapt the attribute sizes among

different layers and produce reasonable decompositions by neighborhood vectors among multiple relevant attribute subsets. It can be ensured that the underlying similarity interdependency among interacting decision variables can be fully determined, and it can keep the interdependencies of multiple relevance attribute sets to a minimum. The other main reason is that quantum-behavior PSO with self-adaptive memplexes has strong optimization performance for the decomposed attribute subsets. These employed quantum operators can strengthen the adaptive stability of particle memplexes and can achieve superior reduction performance in their respective attribute subsets. This will result in better generalization ability to remove the relative dispensable candidate attribute sets so that the reduction sets will be updated quickly. So, we consider that LCQ-ABR\* is an extremely promising attribute reduction algorithm in terms of efficiency and accuracy.

### VII. CONSISTENT SEGMENTATION APPLICATION IN NEONATAL BRAIN TISSUE

The human brain is a hierarchy of complex networks with different spatial and temporal scales. The studying of neonatal brain structure is currently booming [23][24]. This process of removing non-brain tissue is the first module of most brain structure studies. But there are lots of heterogeneous tissues with dynamic, changing characteristics in the neonatal brain structure. The mean tissue densities of gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) in the neonatal brain structure are usually used as features for the forecasting, diagnosis, and treatment of neonatal brain diseases. Hence, it is increasingly urgent to design methods to develop the related techniques of neonatal brain study.

In the following experiment, the proposed LCQ-ABR\* algorithm is applied to the multi-atlas-based simultaneous labeling of longitudinal dynamic cortical surfaces of neonatal brain 3D-MRI, to further evaluate its application performance. We select 10 neonate subjects with 2, 4, 6, and 8 birth months, respectively. Fig. 7 exhibits the close-up views of labeling results of longitudinal surfaces in one typical subject of neonatal brain 3D-MRI with 8% achieved by LCQ-ABR\* with algorithms of Li *et al.* [25], and Wang *et al.* [26]. We see that LCQ-ABR\* can make multi-atlas-based simultaneous labeling adaptive to derive segmentation from atlas surfaces

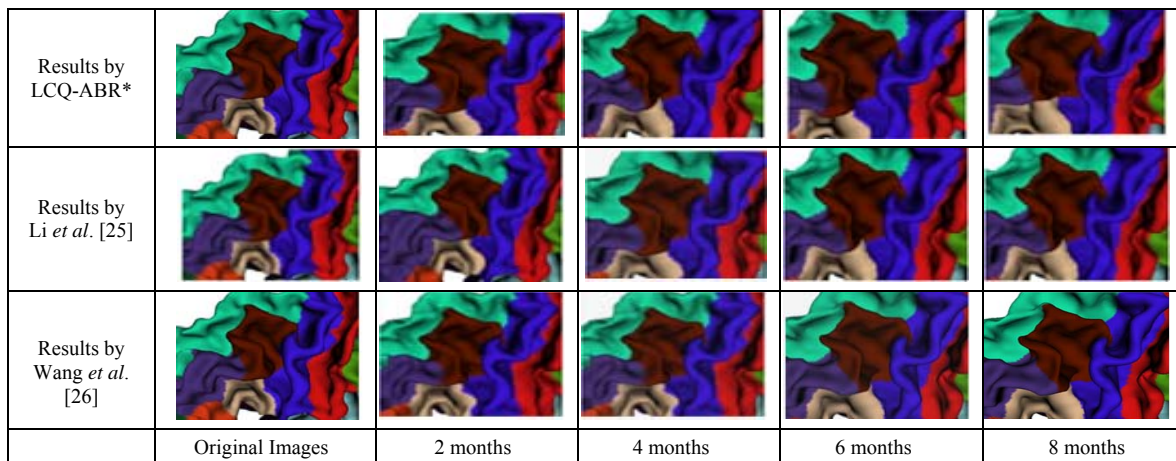


Fig. 7. Labeling results of longitudinal surfaces in the neonatal brain 3D-MRI with 8% Gaussian noise

and cortical folding geometries, and most interregional information is extracted. Because the edges of different organizations of the neonatal brain are very fuzzy at 2 months and 3 months and are easily taken into inappropriate regions, several regions with longitudinally-inconsistent labeling are caused with algorithms of Li *et al.* [25], and Wang *et al.* [26]. LCQ-ABR\* can substantially improve the accuracy of longitudinally-consistent labeling, while substantially maintaining the tissue details of the neonatal brain 3D-MRI.

In the following experiment, the mean classification accuracy (MCA) performance of LCQ-ABR\*, the Li *et al.* algorithm [25], and the Wang *et al.* algorithm [26] are evaluated using two feature types as cortical GM volume and cortical associated WM volume. The WM vs temporal scales and GM vs temporal scales are described in Fig.8. The MCA of different algorithms is improved with increasing temporal scales. MCA of the Wang *et al.* algorithm has a small speedup, just barely greater than that of the Li *et al.* algorithm. The MCA of LCQ-ABR\* shows greater accuracy than the others and finally reaches 94.6% for WM and 93.5% for GM at 8 months. It indicates that the superiority of LCQ-ABR\* can be used to better characterize the interior structure of neonatal brain tissue.

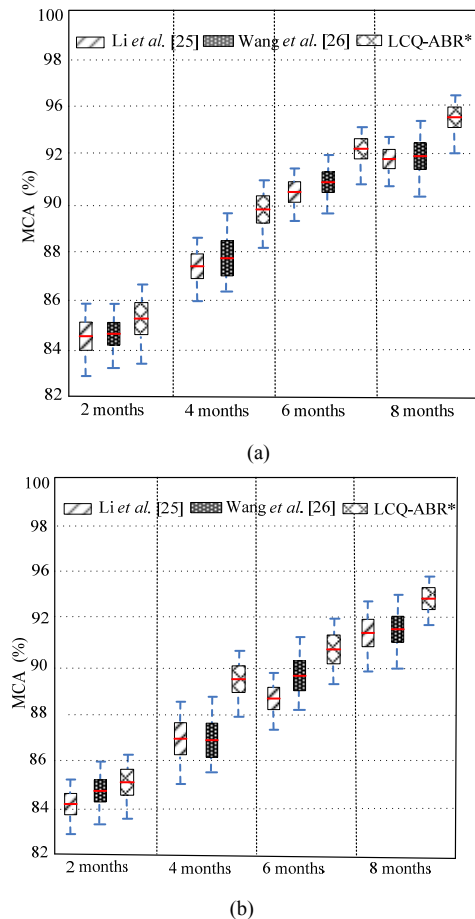


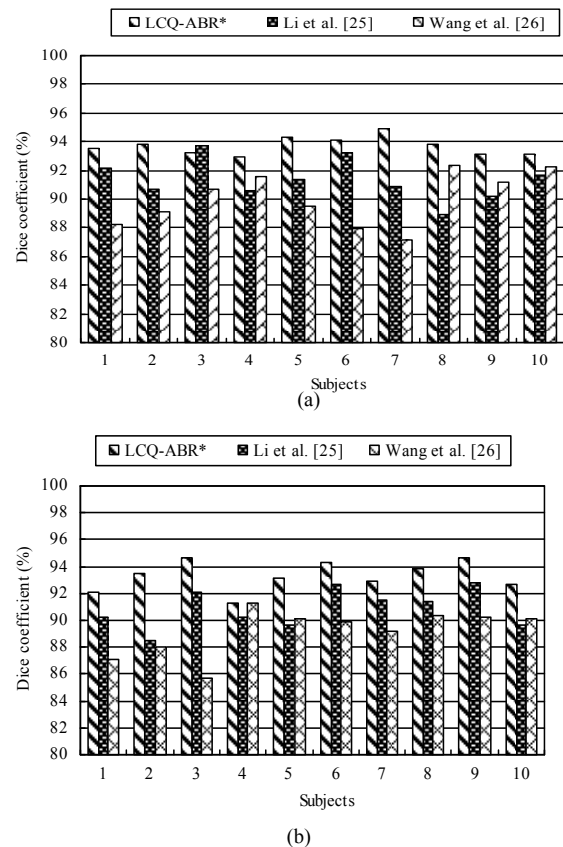
Fig. 8. Boxplot comparisons of MCA with increasing temporal scales for (a) WM, and (b) GM.

Finally, we calculate two kinds of labeling regions of longitudinal cortical surfaces. The gyri and sulci in the neonatal brain structure are two kinds of crucial organizing roles in the architectonic, connectional, and functional sense, but these organizations are not explicitly captured due to the dramatic

changes during early brain development. In this experiment, we invite six 3D-MRI experts to manually annotate the precentral gyrus (PreCG), superior temporal gyrus (STG), precentral cerebral sulci (PreCCS), and central lateral sulci (CLS) in the cortical surfaces of 10 subjects (2~10 months) according to the mean-curvature based cortical surfaces, and then we compute their manual average values of PreCG, STG, PreCCS and CLS, respectively. We further calculate the automatic labeling regions by LCQ-ABR\*. We adopt the Dice coefficient [27] to quantitatively evaluate LCQ-ABR\* in the segmentation performance of labeling regions of longitudinal cortical surfaces. The Dice coefficient can characterize how many pixels in the labeling region of the neonatal brain are correctly segmented and how many pixels outside the labeling region are correctly excluded, respectively. This is defined as

$$\text{Dice coefficient} = \frac{2|X_1 \cap X_2|}{|X_1| + |X_2|} \times 100\% \quad (34)$$

The quantitative comparisons of Dice coefficients of PreCG, STG, PreCCS, and CLS are illustrated in Fig. 9. The average Dice coefficients for PreCG/STG are 93.87%/93.89% (LCQ-ABR\*), 91.97%/90.12% (Li *et al.*), and 91.03%/88.79% (Wang *et al.*). As can be seen, LCQ-ABR\* achieves the highest Dice coefficient of gyri. Meanwhile, the average Dice coefficients for PreCCS/CLS are 92.19%/94.14 (LCQ-ABR\*), 89.19%/90.85% (Li *et al.*), and 88.66%/91.09% (Wang *et al.*). It is noticed that Dice coefficients of PreCG and STG by LCQ-ABR\* surpass those of PreCCS and CLS by Li *et al.* and Wang *et al.* This shows that LCQ-ABR\* also can automatically segment and label the sulci well, and these are complementary to the cortical internal surface.



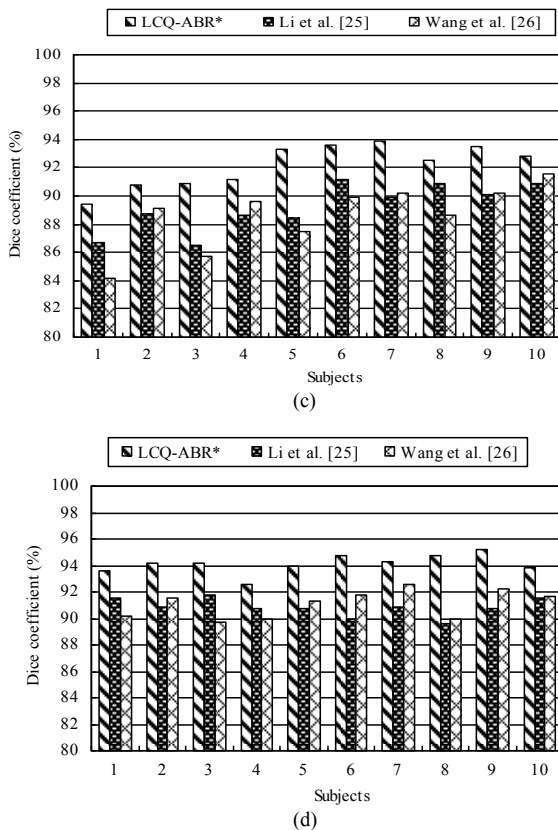


Fig. 9. Dice coefficient comparison on four quantitative evaluations as (a) PreCG, (b) STG, (c) PreCCS, and (d) CLS.

The above results and analysis further verify that LCQ-ABR\* has promising, consistent segmentation performance of neonatal brain 3D-MRI. It can be dynamically adaptive to derive from atlas surfaces and cortical folding geometries. Even though some Gaussian noise is involved in neonatal brain 3D-MRI, and LCQ-ABR\*, it does not have an obvious effect. Therefore, the above qualitative and quantitative results indicate that LCQ-ABR\* has more accurate and consistent performance for complex neonatal brain 3D-MRI with better detail preservation. This has a direct impact on fields where attribute reduction is critical for brain tissue analysis.

### VIII. DISCUSSION

From the above experimental results, it can be observed that LCQ-ABR\* can eliminate irrelevant attributes in original big datasets and efficiently select the useful candidate attributes from decision systems. In addition, LCQ-ABR\* is not sensitive to incremental noise influence, while FRRFDM and TDNEC show stronger sensitivity to noise. This will cause LCQ-ABR\* to be much more feasible for reduction of complex attributes in large-scale datasets. As traditional attribute reduction algorithms [5][9][14][28][29][30] mainly deal with categorical data, causing them to discretize numerical data, this preprocessing procedure will increase their processing time and the complexity of attribute reduction. This does not guarantee that the range of attribute subsets is proportionally represented. However, LCQ-ABR\* can deal with different datasets directly without preprocessing procedure and thus it has efficient

performance. The main reason is that the calculated starting center positions of corresponding neighborhood vectors using the LCMMI model can allow the discovery of better local minima, and multi-agents can self-adapt the attribute sizes among different layers to produce reasonable decompositions among multiple relevant attribute subsets. So, it can be ensured that the underlying similarity interdependencies among interacting decision variables in large-scale datasets can be fully determined by LCQ-ABR\*. This will achieve better generalization ability to remove the relative dispensable candidate attribute sets so that the reduction sets will be updated quickly

For the multiple features of large-scale benchmark datasets in Table 1, the accumulation will be a problem for representative algorithms such as FRFS [5], FRRFDM [9], and TDNEC [14]. From Table 2 and Fig. 6, it can be seen that LCQ-ABR\* is efficient for the NIPS 2003 feature selection challenge datasets (Arcene, Dexter, Dorothea, and Madelon), and is accurate for four public microarray datasets (Colon, Prostate, Leukemia, and Lung-cancer). But FRFS and TDNEC do not produce scalable solutions for these challenge datasets due to the non-smooth and highly complex solution cost terms. Although FRRFDM can make a part of scalable solutions for large-scale datasets by designing the robust fuzzy matrix of discerning certain object pairs, it suffers a degradation of quality with incremental noise rates in large-scale benchmark datasets. Hence, we recommend LCQ-ABR\* for extremely large datasets where FRFS, TDNEC, and FRRFDM are infeasible.

Another issue relates to the robustness of LCQ-ABR\*. Most traditional attribute reduction algorithms [5][9][14][15][31][32] are not robust to noise in massive data, resulting in some unconvincing results. Moreover, the variations of their attribute reduction performance will appear to be significant with increasing attribute noise levels. So, their stability is not consistent with theoretical predictions. But, as shown in Fig. 6, the MCF and CPU time of LCQ-ABR\* are improved only slightly and are more stable with increasing noise rates. The main reason is that the employed QPSOSM strategy with quantum adaptive updating strategy and quantum-behavior mechanism can help the particles jump out of local optima quickly. And it can ensure that the particles converge to the optimal set without the increasing noise influence. This property makes LCQ-ABR\* much more robust at reducing the influence of noise, and it is guaranteed to converge to a global optimal set of attribute reduction. So, it has both better speedup and quality than TDNEC and FRRFDM. With respect to robustness, LCQ-ABR\* is a better choice than most attribute reduction algorithms.

Besides these interesting advantages, there are some limitations that involve some future directions as follows.

First, the effectiveness of quantum-behavior PSO operators in LCQ-ABR\* usually relies on the location selection of different starting best center particles in the neighborhood vectors. However, most operators randomly determine the location selection of starting best center particles. This might lead to inefficiency when dealing with extremely large-scale datasets. Specifically, the particle's position cannot be represented well by the wave function. The exploitation-versus-exploration dilemma may have appeared. So, it is interesting to



extend LCQ-ABR\* to build the interrelationship among different memplexes for determining best center particles, which can help to balance the exploitation and exploration of the particle-search process in the quantum space.

Second, the disadvantage of the LCQ-ABR\* application in neonatal brain 3D-MRI is that it tends to over-smooth the contour of the brain tissue. In some cases, LCQ-ABR\* maybe leave out some gyrus and sulci matter, and fail to identify stable folds of sulci roots in the occipital lobe. This is caused by the high variability of CLS location, which disturbs the local curvature and depth of sulci. So, LCQ-ABR\* must be refined to detect some folding around sulcal roots with various orientations. For future work, we plan to add such dense information as curvature and depth to locally compensate for various orientations, which will provide better evidence of LCQ-ABR\* in a wide variety of neonatal brain 3D-MRI and improve our understanding of cortical variability. Meanwhile, we further design the complementary strategy to ameliorate the noise influence, such as 15~20%, and to conduct more testing work on conformity and sensibility, which can be further incorporated into more consistent and accurate segmentation applications in neonatal brain tissue.

## IX. CONCLUSION

At present, enlarging datasets make attribute reduction algorithms based on RST a challenging task. Moreover, the increasing noise of multiple relevant attribute sources is one main source of uncertainty in many practical applications. In this paper, we introduced two significant contributions to study the attribute reduction algorithm: 1) use of a quantum-behavior PSO with self-adaptive memplexes to strengthen the adaptive stability of particle memplexes for attribute reduction in large-scale datasets; and 2) a new layered co-evolutionary model with multi-agent interaction to adaptively decompose large-scale attribute sets. We presented a novel layered-coevolution-based attribute-boosted reduction algorithm using quantum behavior PSO. Extensive experimental comparative studies confirmed that LCQ-ABR\* can surpass traditional algorithms in terms of efficiency, accuracy, and robustness. Moreover, it was successfully applied to the consistent segmentation for neonatal brain tissue, demonstrating its validity. So, it can achieve significantly better attribute reduction performance, especially for large-scale, uncertain, and noisy datasets. This research offers a promising and flexible model of attribute reduction for the large, noisy, and uncertain datasets linking multiple relevant data sources in applications. The effective and robust segmentation results of neonatal brain 3D-MRI images further demonstrate its stronger applicability.

## ACKNOWLEDGEMENTS

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

## REFERENCES

- [1] R. Jensen, and Q. Shen, Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches. Wiley-IEEE Press, New York, NY, USA, 2008.

- [2] Z. Pawlak, "Rough sets and intelligent data analysis," *Inf. Sci.*, vol. 147, pp. 1-12, Nov. 2002.
- [3] Y.Y. Yao, and Y.H. She, "Rough set models in multigranulation spaces," *Inf. Sci.*, vol. 327, pp. 40-56, Jan. 2016.
- [4] D. Chen, S. Zhao, L. Zhang, Y. Yang, and X. Zhang, "Sample pair selection for attribute reduction with rough set," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 11, pp. 2080-2093, Nov. 2012.
- [5] R. Jensen, and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 824-838, Aug., 2009.
- [6] R. Susmaga, "Reducts and constructs in classic and dominance-based rough sets approach," *Inf. Sci.*, vol. 271, pp. 45-64, Jul., 2014.
- [7] P. Maji and S. K. Pal, "Fuzzy-rough sets for information measures and selection of relevant genes from microarray data," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 3, pp. 741-752, Jun. 2010.
- [8] W. Ding, Z. Guan, Q. Shi, and J. Wang, "A more efficient attribute self-adaptive co-evolutionary reduction algorithm by combining quantum elitist frogs and cloud model operators," *Inf. Sci.*, 2015, vol. 293, pp. 214-234, Feb. 2015.
- [9] S. Zhao, H. Chen, C. Li, M. Zhai, and X. Du, "RFRR: Robust fuzzy rough reduction," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 5, pp. 825 - 841, Oct. 2013.
- [10] W. P. Ding, C. T. Lin, M. Prasad, S.B. Chen, and Z. J. Guan, "Attribute equilibrium dominance reduction accelerator (DCCAEDR) based on distributed coevolutionary cloud and its application in medical records," *IEEE Trans. Syst. Man, Cybern. Syst.*, 2016, vol. 46, no. 3, pp. 384 - 400, Mar. 2016.
- [11] H. Chen, T. Li, D. Ruan, J. Lin, and C. Hu, "A rough-set-based incremental approach for updating approximations under dynamic maintenance environments," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 2, pp. 274-284, Feb. 2013.
- [12] L. J. Ke, 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.
- [13] D. S. Yeung, D. G. Chen, E. C. C. Tsang, J. W. T. Lee, and X. Z. Wang, "On the Generalization of Fuzzy Rough Sets," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 3, pp. 343-361, Mar. 2005.
- [14] W. W. Li, Z. Q. Huang, X. Y. Jia, and X. Y. Cai, "Neighborhood based decision-theoretic rough set models," *Int. J. Approx. Reason.*, vol. 69, pp. 1-17, Feb. 2016.
- [15] J. Qian, D. Q. Miao, Z. H. Zhang, and X.D. Yue, "Parallel attribute reduction algorithms using MapReduce," *Inf. Sci.*, vol. 279, pp. 671-690, Sep. 2014.
- [16] P. Maji and P. Garai, "IT2 fuzzy-rough sets and max relevance-max significance criterion for attribute selection," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1657-1668, Aug., 2015.
- [17] Y. C. Jin, and B. Hammer, "Computational intelligence in big data," *IEEE Comput. Intell.*, vol. 9, no. 3, pp. 12-13, Aug. 2014.
- [18] X. Wu, X. Zhu, G.Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 97-107, Jan. 2014.
- [19] K.H. Han, and J.H. Kim, "Quantum-inspired evolutionary algorithm for a class of combinatorial optimization," *IEEE Trans. Evol. Comput.*, vol. 6, no. 6, pp. 580-593, Dec. 2002.
- [20] J.W. Gu, M.Z. Gu, C.W. Cao, and X.S. Gu, "A novel competitive co-evolutionary quantum genetic algorithm for stochastic job shop scheduling problem," *Comput. Oper. Res.*, vol. 37, pp. 927-937, May. 2010.
- [21] W. Ding, and J. Wang, "A novel approach to minimum attribute reduction based on quantum-inspired self-adaptive cooperative co-evolution," *Knowl. Based Syst.*, vol. 50, pp. 1-13, Sep., 2013.
- [22] Z. A. E. M. Dahi, C. Mezioud, and A. Draa, "A quantum-inspired genetic algorithm for solving the antenna positioning problem," *Swarm Evol. Comput.*, vol. 31, pp. 24-63, Dec., 2016.
- [23] A. Afif, J. Trouillas, and P. Mertens, "Development of the sensorimotor cortex in the human fetus: A morphological description," *Surgical and Radiologic Anatomy*, vol. 37, no. 2, pp. 153-60, Jun., 2015.
- [24] G. Li, J. X. Nie, L. Wang, F. Shi, and J. H. Gilmore. W. L. Lin, and G. G. Shen "Measuring longitudinally dynamic cortex development in infants by deconstruction of consistent cortical surfaces," *2013 IEEE 10<sup>th</sup> International Symposium on Biomedical Imaging*, San Francisco, CA, USA, 2013, pp. 1368-1371.
- [25] G. Li, J. X. Nie, and G.R. Wu, et al, "Consistent reconstruction of cortical surfaces from longitudinal brain MR Images," *Neuroimage*, vol. 59, no. 4, pp. 3805-3820, Oct. 2012.

- [26] L. Wang, F. Shi, and G. Li, et al, "Segmentation of neonatal brain MR images using patch-driven level sets," *Neuroimage*, vol. 84, pp. 141-158, Jan. 2014.
- [27] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297-302, Nov., 1945.
- [28] W. Ziarko, "Variable precision rough set model," *J. Comput. Syst. Sci.*, vol. 46, no. 1, pp. 39-59, Feb., 1993.
- [29] Y. Yao, and Y. Zhao, "Discernibility matrix simplification for constructing attribute reducts," *Inf. Sci.*, vol.179,no.7, pp.867-882, Mar.,2009.
- [30] Y. Qian, J. Liang, W. Pedrycz, and C. Dang, "Positive approximation: an algorithm for attribute reduction in rough set theory," *Artif. Intell.*, vol. 174, no. 9, pp. 597-618, Jun. 2010.
- [31] 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, no.12, pp. 3509-3521, Dec. 2007.
- [32] 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.



**Weiping Ding** (M'16) received the Ph.D. degree in Computation Application, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013. He was a visiting Scholar at Department of Mathematics and Computer Science, University of Lethbridge, Canada, and Department of Electrical and Computer Engineer, National University of Singapore (NUS), Singapore. He was a Post-Doctoral Fellow at the Brain

Research Center, National Chiao Tung University, Hsinchu, Taiwan, in 2014. He is a member of IEEE, ACM and CCF. He has authored more than 60 papers in journals and conference proceedings. He holds 10 patents as inventor or co-inventor. His current research interests included machine learning, quantum computing, data mining and their applications in big data. Dr. Ding was awarded an excellent- young teacher of Jiangsu Province sponsored by Qing Lan Project in 2014, and high-level talent of Jiangsu Province "Six Talent Peak" in 2016. He was awarded the Best Paper of ICDMA'15, HongKong. Dr. Ding serves as an Associate Editor of *IEEE Transaction on Fuzzy Systems*, and *Information Sciences*.



**Chin-Teng Lin** (S'88-M'91-SM'99-F'05) received the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989 and 1992, respectively. He is currently the Distinguished Professor of Faculty of Engineering and Information Technology, University of Technology Sydney, University Chair Professor of Electrical and

Computer Engineering, NCTU, International Faculty of University of California at San-Diego (UCSD), and Honorary Professorship of University of Nottingham. Dr. Lin was elevated to be an IEEE Fellow for his contributions to biologically inspired information systems in 2005, and was elevated International Fuzzy Systems Association (IFSA) Fellow in 2012. He is elected as the Editor-in-chief of IEEE Transactions on Fuzzy Systems since 2011. He also served on

the Board of Governors at IEEE Circuits and Systems (CAS) Society in 2005-2008, IEEE Systems, Man, Cybernetics (SMC) Society in 2003-2005, and IEEE Computational Intelligence Society (CIS) in 2008-2010. Dr. Lin is the Distinguished Lecturer of IEEE CIS Society from 2015-2017. He served as the Deputy Editor-in-Chief of IEEE Transactions on Circuits and Systems-II in 2006-2008. Dr. Lin is the coauthor of *Neural Fuzzy Systems* (Prentice-Hall), and the author of *Neural Fuzzy Control Systems with Structure and Parameter Learning* (World Scientific). He has published more than 200 journal papers (Total Citation: 19,166, H-index: 53, i10-index: 332) in the areas of fuzzy systems, neural networks and neuro-engineering, including approximately 101 IEEE journal papers.



**Mukesh Prasad** (M'13) received his Master's degree in Computer Application from Jawaharlal Nehru University, New Delhi, India, in 2009 and Ph.D degree in Department of Computer Science at National Chiao-Tung University, Hsinchu, Taiwan, in 2015. Currently, He is a Lecturer the school of Software, Faculty of Engineering and information Technology, University of Technology Sydney, Australia. 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.



**Zehong Cao** received the B.E. degree in Electronic and Information Engineering, Northeastern University (NEU) in 2012, and the M.S. degree in Electronic Engineering from The Chinese University of Hong Kong (CUHK) in 2013. Currently, he is working toward the dual Ph.D. degree program at the Faculty of Engineering and Information Technology, University of Technology, Australia, and the Institute of Electrical and Control Engineering, National Chiao Tung University (NCTU), Taiwan, respectively. He is also a student member in Computational Intelligence and Brain-Computer Interface (CI-BCI) Lab, CAI, UTS. His research interests are in the area of biomedical signal processing, brain-computer interface, fuzzy entropy, machine learning, and their applications in clinical medicine.



**Jiandong Wang** received the B.S. degree in Radio Department of Shanghai Jiaotong University Shanghai, China, in 1967. He is a professor in College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China. His research interests are artificial intelligence and machine learning.