A Parallel Minimum Attribute Co-reduction Accelerator based on Quantum-inspired SFLA and MapReduce Framework

Weiping Ding, Senbo Chen, and Xueyun Chen

***Abstract*—The fast increase and update of big data brings a new challenge to quickly acquire the useful information with classical attribute reduction methods. In this paper, a parallel minimum attribute co-reduction accelerator (QSMFAC) based on quantum-inspired SFLA and MapReduce framework is presented. First, a novel framework of *N*-populations distributed co-evolutionary cloud model is designed to divide the entire population into *N* subpopulations and share the rewards of different subpopulations' solutions under MapReduce mecha- nism. Second, the divided attribute subsets in subpopulations are coevolved by quantum-inspired SFLA in which evolutionary frogs are represented by quantum chromosome gene state, and the crossover co-evolutionary strategy between neighborhood subpopulations can adapt the consecutive sharing of better performance. Third, the MapReduce based approximation parallelism mechanism is adopted to conduct rules reduction to speed up the computation of attribute equivalence classes, so that it will be extended to high performance in both quality of solution and competitive computation complexity. Experimental results indicate the proposed accelerator has better on efficiency and accuracy of minimum attribute reduction than some representative algorithms. Moreover it is applied into MRI segmentation with intensity inhomogeneity, and the effective and robust segmentation results further indicate it has stronger superior for complex big data application.**

***Index Terms*—minimum attribute reduction, MapReduce framework, quantum-inspired SFLA, big data analysis, crossover co-evolutionary operator**

Manuscript received November 4, 2014; revised November 19, 2014. This work was supported by the National Natural Science Foundation of China under Grant 61300167, the Open Project Program of State Key Lab. for Novel Software Technology under Grant KFKE2012B28, Open Project Program of Jiangsu Provincial Key Laboratory of Computer Information Processing Technology, sponsored by Qing Lan Project of Jiangsu Province, Starting Foundation for Doctoral Scientific Research Nantong University, and Applying Study Project of Nantong under Grant BK2013043 and BK2012038.

Weiping Ding is with School of Computer Science and Technology, Nantong University, Nantong 226019, China (Tel.: +86-13921606118. Fax: +86-513-85012511, e-mail: dwp9988@163.com).

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

Chenyun Chen is with School of Computer Science and Technology, Nantong University, Nantong 226019, China (e-mail: chenxy@ntu.edu.cn).

I. Introduction

In tackling a large number of critical real-world problems, big data analysis has become increasingly popular. Effective and efficient solutions for many grand challenges in big data require better and faster computational methods [1]. As an important part of Granular computing (GrC) [2], Rough set theory (RST) is an effective mathematical tool to deal with the uncertainty and imprecise problems, and it has been successfully applied into many data analysis tasks in the field of artificial intelligence [3][4]. Attribute reduction in RST 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 original features for objects from the universe [5]. It plays an important role in the fields of machine learning, data mining and knowledge discovery [6-8]. But this new challenge of big data analysis has posed some new research question that can stimulate the rapid growth of attribute reduction algorithm.

In recent years, a novel parallel programming model (MapReduce) has received much attention from both scientific community and industry for its applicability in big data algorithm [9]. This also boosts a strong interesting and trend toward studying attribute reduction in big data analysis. Some last researches are described as follows: Zhang et al. [10-12] presented a parallel algorithm for computing the equivalence classes, decision classes based on the MapReduce model, and then designed two parallel incremental algorithms for updating rough set approximations in different incremental strategies to deal with big data. Qian et al. [13] designed the algorithm of computing equivalence classes for large-scale data set in cloud computing, and proposed the corresponding attribute reduction algorithms. Qian et al. [14] adopted the parallel computations of the equivalence classes and attribute significance for attribute reduction, and proposed the hierarchical attribute reduction algorithm for big data by using MapReduce. Zhang et al. [15] put up the parallel large-scale rough set based methods for knowledge acquisition, and then implemented them on

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

several representative MapReduce runtime systems: Hadoop, Phoenix and Twister to mine knowledge from big data. Qian et al. [16] designed a novel structure of <*key*, *value*> pair to speed up the computation of equivalence classes and attribute significance, and parallelized the traditional attribute reduction process based on MapReduce mechanism.

Nevertheless, the attribute reduction results of most above-mentioned algorithms are not usually guaranteed to be the same results as those achieved by implemented on the whole and non-separated big data set, because these subsystems on separated data sets by adopting MapReduce mechanism are independent, and they can not enough share the performance information with neighborhood and other subsystems. These will cause these methods to acquire the inexact results of attribute reduction from big data set in most cases. The current main obstacles are that most algorithms are often computationally time-consuming for big data sets, which will greatly restrict the development and practical application of the RST theory. Therefore it is necessary to investigate some novel and effective heuristic attribute reduction algorithms to speed up the knowledge discovery process, which can well adapt to the applications of real big data set.

As we all know, the quantum-inspired evolutionary algorithm (QEA) utilizes the concepts of quantum bit (Q-bit), superposition of states and collapse of states. A Q-bit individual has the special advantage that it can represent a linear superposition of states by using search space probabilistically. Thus, it has a better characteristic of population diversity than any other representation. Recently, some scholars have presented a few of efficient QEAs for global optimization problems [17-19]. We also presented quantum-inspired SFLA and successfully applied it into attribute reduction in RST [20]. Although our proposed algorithm has shown a nicer performance in several aspects compared with the existing algorithms, a critical confronting challenge is that it is still not effective enough for big data sets due to its high time consumption and space complexity.

In this paper, we propose a parallel minimum attribute co-reduction accelerator (QSMFAC) based on quantum-inspired SFLA and MapReduce framework. Our work is based on QSFLAR as in [20] and the aim is to develop its fast and improved version to deal with big data. The proposed QSMFAC can still reserve the merits of QSFLAR, and meanwhile it can effectively avoid the shortcomings of its high time consumption in attribute reduction of big data. First, we design a novel framework of *N*-populations distributed co-evolutionary cloud model (DCCM) to divide the entire population into *N* subpopulations and share all the rewards of different subpopulations' solutions under MapReduce mechanism. Then, the MapReduce based approximation parallelism mechanism is adopted to implement rules reduction by speeding up the computation of attribute equivalence classes. In addition, the crossover co-evolutionary strategy can adapt the consecutive sharing of better performance between the divided neighborhood quantum frog subpopulations so as to further improve better performance. Thus QSMFAC will become an integrated attribute co-reduction accelerator for big data analysis.

The rest of this paper is organized as follows. Section II describes minimum attribute reduction model in brief. In section III, by designing new operators of quantum-inspired SFLA and framework of *N*-populations DCCM Model, a new parallel minimum attribute co-reduction accelerator QSMFAC is proposed. Experimental studies are reported in Section IV, and some concluding remarks are given in the last section.

II. Minimum Attribute Reduction Model

In RST, an information system, which is also called a decision table, is defined as $S = (U, A, V, f)$, where U, called universe, is a nonempty set of finite objects; $A = C \bigcup D$ where C is the set of condition attributes and D is the set of decision attributes, V a set of values of attributes in A, and $f : A \to V$ a description function [3][4].

Definition 1 For any concept $X \subseteq U$ and attribute subset $R \subseteq C$, X can be approximated by the R-lower and R-upper approximation

The R-lower approximation of X is the set of objects of U that are surely in X, defined as

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

The R-upper approximation of X is the set of objects of U that are possibly in X, defined as

$$\overline{R}X = \{x \in U \mid [x]_R \bigcap X \neq \varnothing\} \tag{2}$$

Definition 2 Let P, $Q \subseteq A$, it is said that Q depends on P in a dependency degree $k(0 \le k \le 1)$, denoted $P \Rightarrow_k Q$, if

$$k = \gamma_P(Q) = \frac{\left|POS_P(Q)\right|}{|U|} \tag{3}$$

where $|\bullet|$ represents the cardinality of a set. $\left|POS_P(Q)\right|$, called positive region, is defined as

$$\left|POS_P(Q)\right| = \bigcup_{X \in U/Q} \underline{P}X \tag{4}$$

The positive region contains all the objects in U that can be uniquely classified to blocks of the partition U/Q by means of the knowledge in the attribute P.

Definition 3 Let $R \subseteq C$, R is said to be a reduction if

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

The intersection of all reductions is called as the attribute core of C, which is denoted as $Core(C)$.

Definition 4 Let $\{0,1\}^m$ be the m-dimensional Boolean space and ξ be a mapping from $\{0,1\}^m$ to the power set 2^C such as

$$x_i = 1 \Leftrightarrow a_i \in \xi(x) \qquad i=1,2,...,m,\ a_i \in C \tag{6}$$

Then minimum attribute reduction can be reformulated as the following constrained binary optimization model

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

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

where $0\le S(x)=\sum_{i=1}^{m}x_i\le m$.

Given a vector $x\in\{0,1\}^m$, if it is a feasible solution to (5), then its corresponding subset of attributes $\xi(x)$ is a reduction. Furthermore, if it is an optimal solution to (7), then $\xi(x)$ is a minimum attribute reduction.

Definition 5 Construct the fitness evaluation function of minimum attribute reduction as follows:

$$Fit(x)=\min\left(\overline{\rho}\times\frac{|C(x)|-|R(x)|}{|C(x)|}+\overline{\partial}\times\frac{|Core(\xi(x))|}{\gamma_{\xi(x)}(D)}\right)+\lambda\times\Phi(x) \quad (8)$$

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, and $Core(\xi(x))$ is the reduction core of attribute subsets. $\gamma_{\xi(x)}(D)$ is the reduction quality of attribute subsets $\xi(x)$ relative to decision attribute set D, $\overline{\rho}=\frac{1}{N}\left(\sum_{i=1}^{N}f_{Elitist_i}\right)$ represents the average fitness of elitists in N subpopulations where $f_{Elitist_i}$ is the fitness of the i^{th} elitist in subpopulation. Supposed that $\overline{f_{Subpopulation_i}}=\frac{1}{r}\sum_{j=1}^{r}f_j$, where r is the number of co-evolutionary individuals in the subpopulation, $\overline{\partial}=\frac{1}{N}\left(\sum_{i=1}^{N}\overline{f_{Subpopulation_i}}\right)$ represents the total average fitness of N subpopulations. The $\Phi(x)$ is the adaptive penalty function and it is formulated as follows:

$$\Phi(x)=\Gamma_i\times\left(\frac{\gamma_{\xi(x)}(D)}{\gamma_C(D)}\right) \quad (9)$$

$$\text{where}\quad \Gamma_i=\frac{1}{r}\sum_{j=1}^{r}\left|\frac{f_{Elitist_i}-\overline{f_j}}{f_{Elitist_i}}\right| \quad (10)$$

This penalty function $\Phi(x)$ can adaptively control the fitness complexity of minimum attribute reduction and induce its desired consumption.

III. Minimum Attribute Co-reduction Accelerator based on Quantum-inspired SFLA and MapReduce Framework (QSMFAC)

A. New operators of quantum-inspired SFLA

1) *Quantum chromosome gene operator*

The quantum-inspired SFLA (QSFLA) use a new quantum bit (Q-bit) representation. The initial quantum frogs are randomly generated according to the binary coding method as $P(t)=\{p_1^t,p_2^t,...,p_n^t\}$, where n is the number of quantum frogs. $p_j^t(j=1,2,...,n)$ is the quantum chromosome gene operator of the t^{th} iteration which is defined as follows:

$$p_j^t=\begin{pmatrix}\alpha_{11}^t|\alpha_{12}^t|...|\alpha_{1l}^t|\alpha_{21}^t|\alpha_{22}^t|...|\alpha_{2l}^t|...|\alpha_{m1}^t|\alpha_{m2}^t|...|\alpha_{ml}^t\\ \beta_{11}^t|\beta_{12}^t|...|\beta_{1l}^t|\beta_{21}^t|\beta_{22}^t|...|\beta_{2l}^t|...|\beta_{m1}^t|\beta_{m2}^t|...|\beta_{ml}^t\end{pmatrix} \quad (11)$$

where m is the number of the chromosome genes of quantum frogs, and l is the number of quantum bits of each gene. Both α and β are initialized as ($1/\sqrt{2}$), representing the same probability for the linear superposition. Such operations as quantum rotation gate and quantum mutation in [20] for the subpopulation of quantum frogs will effect on all possible information.

2) *Updated operator of worst quantum frog*

The main steps of quantum-inspired SFLA are in accordance with steps as described in [20]. But at the k^{th} iteration, the quantum frog with the worst fitness $P_w=(p_{w1},p_{w2},...,p_{wD})$ in each subpopulation is updated by new operator as following:

$$x_{wd}^{k+1}=x_{wd}^k+r_1\times\lambda\times\text{sgn}(p_{bd}^k-x_{wd}^k)+r_2\times(1-\lambda)\times\text{sgn}(p_{Bd}^k-x_{wd}^k) \quad (12)$$

$$\lambda=\left|\frac{f_{Elitist_i}-f_{ELITIST}}{f_{ELITIST}}\right| \quad (13)$$

where r_1,r_2 are uniform random numbers which can be selected from the range of parameters($r_1,r_2\in[0,1]$), and their values express the relative important degree of P_b and P_B in the leaping process for quantum-inspired SFLA. That $f_{ELITIST}$ represents the fitness of the global best quantum elitist frog. p_{bd}^k, p_{Bd}^k represent the each subpopulation's best solution and all global best solution, respectively, in the d^{th} dimension space and the k^{th} iteration. x_{wd}^{k+1} represents the position of the worst quantum frog in the d^{th} dimension space and the $(k+1)^{\text{th}}$ iteration, and the function sgn( ) is a sign function and its value is from $\{-1,0,1\}$.

2) *Crossover co-evolutionary operator*

A crossover co-evolutionary operator with competitive and cooperative mechanism is designed as a new boosted operator between neighborhood subpopulations as to implement consecutive sharing of better performance, described in Fig. 1. The individuals originally exist in the evolutionary tree, and elitist individuals in each subpopulation are dynamically created by the mechanism of self-adaptive competitive and cooperative co-evolution. When competitive co-evolution is performed, individuals on all levels of a sub-tree are mixed together, and two kinds of entities, the elitist and inferior, are selected by the fitness competition. When cooperative co-evolution is implemented, two special individuals in each subpopulation will share the successful and failure performance experiences with its corresponding elitist and inferior of neighborhood subpopulation. This will strengthen

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the sharing experience so as to further improve better performance among the divided quantum frog subpopulations.

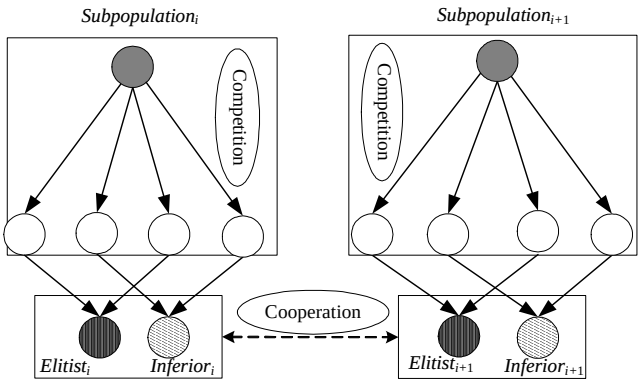

Fig. 1. A crossover co-evolutionary strategy between neighborhood subpopulations in the evolutionary trees

B. Framework of N-populations DCCM Model

Based on the MapReduce model in cloud computing [9], in order to improve the performance for attribute reduction in big data set, we design a *N*-Populations distributed co-evolutionary cloud model (DCCM) to accelerate reduction implementation, which architecture is shown in Fig. 2.

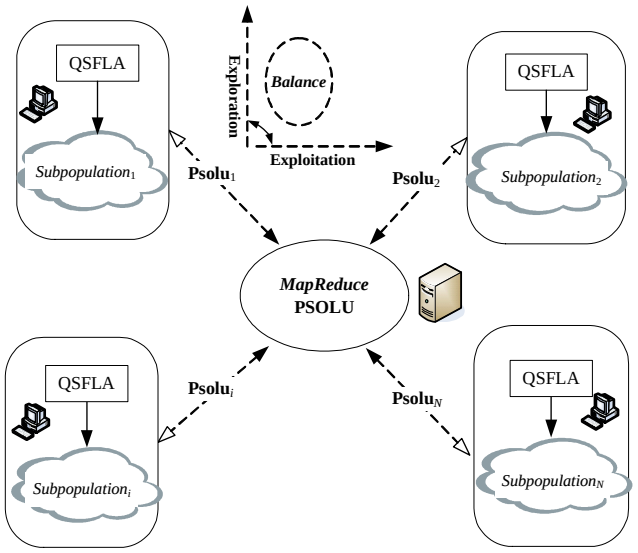

Fig. 2. Framework of parallel *N*-populations DCCM Model

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

1) Using MapReduce framework, we divide the entire population into *N* subpopulations as (*Subpopulation*$_1$,…,*Subpopulation*$_i$,…,*Subpopulation*$_N$);
2) Every subpopulation (*Subpopulation*$_i$) will share the respective current best solution with neighborhood distributed co-evolutionary subpopulations in order to expedite the entire parallel computational speed.
3) The proposed QSFLA with new operators will be in charge of optimizing the respective assigned subpopulation (*Subpopulation*$_i$). In the end of each iteration, each subpopulation focuses on resolution of *Subpopulation*$_i$ and will produce its best solution $\mathbf{Psolu}_i$ with the self-adaptive probability p_i, which stands for the probability $< key_i, value_i >$ on the MapReduce. And the probability p_i is defined as follows:

$$p_i = \left| \frac{f_{ELITIST} + \omega_i f_{Elitist_i}}{f_{ELITIST}} \right|, \quad \omega_i = \left| \frac{f_{Elitist_i}}{\frac{1}{r}\sum_{j=1}^{r} f_j} \right| \tag{14}$$

where f_j is the best fitness of the j^{th} evolutionary individual, r is the number of co-evolutionary individuals in *Subpopulation*$_i$, $f_{Elitist_i}$ is the local best fitness of elitist in *Subpopulation*$_i$, and $f_{ELITIST}$ is the global best fitness of best elitist in entire population.

4) The archiving strategy will record the best solution attained by each subpopulation. Then all solution sets are combined to form the complete solution, which is constructed as follows:

$$\mathbf{PSOLU} = \mathbf{Psolu_1} + \mathbf{Psolu_2} + ... + \mathbf{Psolu}_i + ... + \mathbf{Psolu}_N \tag{15}$$

$$\mathbf{Psolu}_i = \begin{bmatrix} Psolu_{i1} \\ Psolu_{i2} \\ \vdots \\ Psolu_{NN} \end{bmatrix} \tag{16}$$

As shown in Fig.1, the dash line indicates the execution sequence order, in which the black arrow contributes to the completed solution construction of distributed co-evolutionary *Subpopulation*$_i$, meanwhile white arrow is corresponding feedback states for *Subpopulation*$_i$. These subpopulations will share all the rewards equally from all solution set **PSOLU** by MapReduce architecture. As the result, the running processes of DCCM model will have superior performance in updating the co-adaptation of all subpopulations.

The proposed DCCM model can be good at exploring the search space and locating the global region, meanwhile two abilities of exploration and exploitation will be well balanced as to achieve better performance for attribute coevolutionary reduction, with the adequate information-sharing for all subpopulations. Moreover, as the coadaptation of different subpopulations can be attained, final reductions on divided data subsets are surely guaranteed to be the same result as those implemented on the whole independent big data set.

C. Proposed QSMFAC accelerator

Firstly, the QSMFAC can decompose the big data set into many data subsets by using DCCM model, and then set up the $< key, value >$ pairs on each data subset. Secondly, it implements Map/Reduce operations and computes reduction rule set based on the MapReduce parallelism until the global minimum attribute reduction can be achieved. The flowchart of QSMFAC accelerator is illustrated in Fig. 3.

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

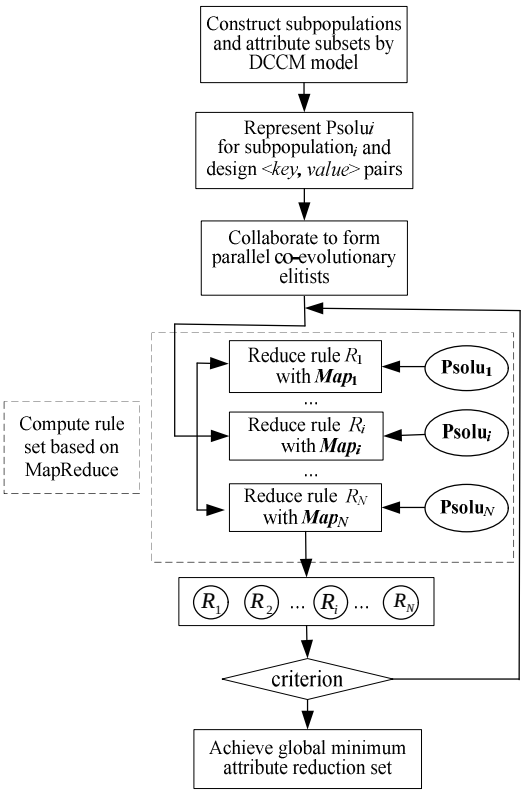

Fig. 3. Flowchart of QSMFAC accelerator

IV. Experimental Evaluations

In order to illustrate the efficiency and effectiveness of QSMFAC accelerator in big data analysis, we carry out some experiments on various big data sets, compared with selected methods. Some parameters of QSFLA are in accordance with those as in [20]. Hadoop version 1.0.1 and Java 1.6.0.12 are employed as MapReduce system. We run algorithms on a cluster of 10 nodes, in which one is set as a master node and the rest nine are configured as slave nodes. Each node has 16 GB main memory and AMD Opteron 2376 with Intel Core 2 Quad CPUs (8 cores in all, each has a clock frequency of 2.3 GHz), and connects via an Ethernet (100 Mbit/s). The operating system is Linux CentOS 4.7.

A. Classification comparison on UCI datasets

Five UCI machine learning datasets [21], which number have being magnified 10^5 times, are selected to verify the classification performance of QSMFAC accelerator. The accuracy comparisons are performed when the classifier Naive Bayes based on QSMFAC, ACOAR [5], PACCA [16] and QSFLAR [20] respectively, are employed to test the classified data sets. In selected five UCI datasets, we use 55% as training set, 15% for validation and 30% as testing set. We analyze the performance of different algorithms in dealing with datasets with different attribute-noises as described in Table I. These 5%, 8%, and 13% attribute-noises are incrementally added into the raw datasets, and then we apply the classifier Naive Bayes based on four algorithms on these datasets and compute the classification accuracies. The "Avg." row records the average accuracies.

TABLE I
Classification accuracy of Naïve Bayes classification based on four algorithms (%)

Dataset	Noise Level	ACOAR	PACCA	QSFLAR	QSMFAC
Thyroid	5	87.23	93.54	90.17	**94.80**
	8	84.56	90.23	89.02	**92.10**
	13	75.17	83.12	83.17	**87.17**
Musk Version2	5	84.32	89.96	87.32	**92.03**
	8	79.43	84.10	82.15	**88.38**
	13	68.16	75.19	74.32	**80.56**
Arrhythmia	5	84.82	**92.29**	89.14	91.35
	8	75.48	**88.98**	83.18	87.36
	13	67.57	**84.47**	78.07	83.52
Audiology	5	80.73	89.14	89.23	**91.47**
	8	76.14	80.27	83.28	**87.24**
	13	63.60	75.93	72.09	**82.37**
Weka-3.2 G	5	83.98	86.40	88.78	**90.23**
	8	77.25	82.39	83.08	**88.09**
	13	70.25	77.70	71.96	**84.92**
Avg.		77.25	84.91	83.01	**88.15**

As shown in Table I, the results indicate that four algorithms can all improve the classification accuracy by eliminating some irrelevant features in the original data set. Among four algorithms, one can observe that Naïve Bayes classifier based on QSMFAC achieves higher classification accuracy than three compared algorithms in most cases. It surpasses ACOAR, PACCA and QSFLAR by five, four and five out of five data sets respectively, and the average accuracy of QSMFAC surpasses all compared algorithms. Furthermore, on Musk Version2 dataset, the classification accuracy of the Naïve Bayes classifier based on QSMFAC is obviously enhanced from 80.56% to 88.38%, and 92.16% respectively, as noise level reduces from 13 to 8 and 5.

It can be seen in Table I the accuracy performances of the Naïve Bayes classifier based on ACOAR, PACCA and QSFLAR algorithms sharply drop, but the classifier based on QSMFAC is more robust and its classification accuracy does not change especially much.

B. Application comparison on the MRI

A major difficulty in the MRI segmentation is the complex intensity inhomogeneity due to the radio-frequency coils or acquisition sequences. In the following experiment, we will evaluate the segmentation performance on large-scale complex cerebrum MRI [22] segmentation of QSMFAC accelerator, compared with by ACOAR [5], PACCA [16] and QSFLAR [20].

Fig. 4(a) is the original cerebrum MRI corrupted by 8% intensity inhomogeneity, and Fig. 4 (b)-(e) show the segmentation results by using ACOAR, PACCA, QSFLAR, and QSMFAC, respectively. It can be noticeable that ACOAR and QSFLAR are much less fragmented than other two algorithms and have somewhat disadvantage of blurring of some details. Both PACCA and QSMFAC can achieve comparably satisfactory segmentation results for MRI with 8% intensity inhomogeneity, but QSMFAC produces more

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

accurate region boundaries of cerebrum MRI, as shown in Fig. 4(e).

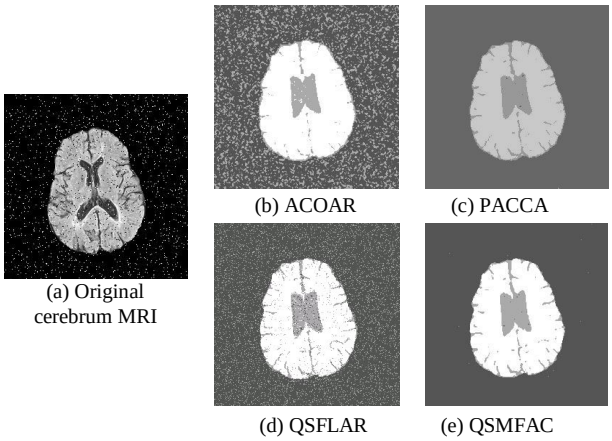

Fig. 4. Comparison of segmentation results on cerebrum MRI corrupted by 8% intensity inhomogeneity

To quantitatively demonstrate segmentation accuracy, Fig. 5 gives the distributions of segmentation accuracy metrics. It is apparent that the QSMFAC outperforms three compared algorithms for on cerebrum MRI with 8% intensity inhomogeneity, which demonstrates the stronger robustness of the QECMASCR.

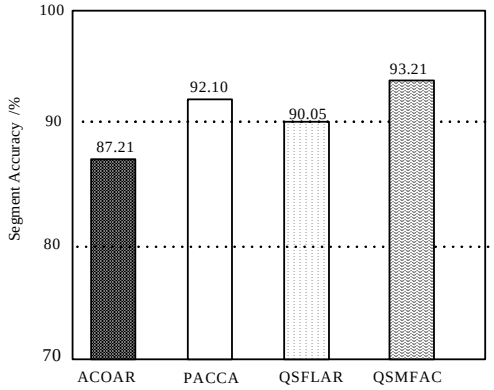

Fig. 5. Quantitative comparison of MRI segmentation accuracy

According to the above results and analysis, we strongly recommend QSMFAC as the preferred algorithm for finding minimum attribute reduction and the segmentation of complex medical MRI. So QSMFAC has more excellent feasibility and effectiveness for minimum attribute reduction in big data analysis.

V. Conclusions

More and more researches and applications need to deal with big data, which leads to a hot of studying novel attribute reduction algorithms as the important preprocessing step of knowledge discovery. But most existing attribute reduction algorithms in RST deem to fail to deal with big data. In this paper, we have proposed a parallel minimum attribute co-reduction accelerator based on quantum-inspired SFLA and MapReduce framework. The MapReduce parallelism mechanism and new operators of QSFLA are adopted to implement attribute reduction, which performs well and shows high accuracy for big data analysis. In the further work, we would extend this QSMFAC accelerator into electroencephalogram (EEG) data in order to achieve better performance in more complex big data with high attribute noises.

References

[1] C. L. P. Chen and C.Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Information Scien*ces, vol. 275, pp. 314–347, Aug. 2014.
[2] W. Pedrycz, "Granular Computing: analysis and design of intelligent systems," *CRC Press/Francis Taylor*, Boca Raton, 2013.
[3] Z. Pawlak, "Rough sets," International Journal of Computer and Information Sciences, vol. 11, no. 5, pp. 341–356, Oct. 1982.
[4] Z. Pawlak and A. Skowron, "Rough sets: Some extensions," Information Sciences, vol. 177, pp. 28–40, Jan. 2007.
[5] L. J. Ke, Z. R. Feng, and Z. G. Ren, "An efficient ant colony optimization approach to attribute reduction in rough set theory," *Pattern Recognition Letters*, vol. 29, no. 9, pp. 1351-1357, July, 2008.
[6] Y. H. Qian, J. Y. Liang, W. Pedrycz, and C.Y. Dang, "Positive approximation: an accelerator for attribute reduction in rough set theory," *Artificial Intelligence*, vol. 174, no. (9-10), pp. 597–618, June. 2010.
[7] W. P. Ding, J. D. Wang, and Z. J. Guan, "Cooperative extended rough attribute reduction algorithm based on improved PSO," *Journal of Systems Engineering and Electronics*, vol. 23, no. 1, pp. 160-166, Feb., 2012.
[8] J. H. Dai, and Q. Xu, "Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification," *Applied Soft Computing*, vol.13, no.1, pp. 211-221, Jan., 2013.
[9] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, Jan., 2008.
[10] J. B. Zhang, T. R. Li, Y. Pan, C. Luo, and F. Teng, "A parallel and incremental algorithm for updating knowledge based on rough sets in cloud computing platform," Journal of Software, 2014 (in Chinese). http://www.cnki.net/kcms/doi/10.13328/j.cnki.jos.004590.html.
[11] J. B. Zhang, T. R. Li, D. Ruan, Z. Z. Gao, and C. B. Zhao, "A parallel method for computing rough set approximations," *Information Sciences*, vol. 194, pp. 209–223, July, 2012.
[12] J. B. Zhang, T. R. Li, and Y. Pan, "Parallel rough set based knowledge acquisition using MapReduce from big data," in *Proc. the 1st Interational Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, BigMine'12, New York: ACM Press, pp. 20–27, 2012.
[13] J. Qian, D. Q. Miao, and Z. H. Zhang, "Knowledge reduction algorithms in cloud computing," *Chinese Journal of Computers*, vol. 34, no. 12, pp. 2332–2343, Dec. 2011.
[14] J. Qian, P. Lv, X. D. Yue, C. H. Liu, and Z. J. Jing, "Hierarchical attribute reduction algorithms for big data using MapReduce," *Knowledge-Based Systems*, Jan, 2014. http://www.sciencedirect.com/science/article/pii/S0950705114003311.
[15] 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," *International Journal of Approximate Reasoning*, vol. 55, no. 3, pp. 896–907, March, 2014.
[16] J. Qian, D. Q. Miao, Z. H. Zhang, and X.D. Yue, "Parallel attribute reduction algorithms using MapReduce," *Information Sciences*, vol. 279, pp. 671–690, Sep., 2014.
[17] L. C. Jiao, Y. Y. Li, M. G. Gong, and X. R. Zhang, "Quantum-inspired immune clonal algorithm for global numerical optimization", *IEEE Transactions on System, Man, and Cybernetics, Part B*, vol. 38, no. 5, pp. 1234-1253, Oct., 2008.
[18] 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", *Computers & Operations Research*, vol. 37, no. 5, pp.927-937, May, 2010.
[19] M. Ykhlef, "A quantum swarm evolutionary algorithm for mining association rules in large databases", *Journal of King Saud University Computer and Information Sciences*, vol. 23, no. 1, pp.1-6, Jan., 2011.
[20] W. P. Ding, J. D. Wang, Z. J. Guan, and Q. Shi, "Enhanced minimum attribute reduction based on quantum-inspired shuffled frog leaping algorithm," *Journal of Systems Engineering and Electronics*, vol. 24, no. 3, pp. 426–434, June, 2013.
[21] A. Asuncion and D. J. Newman. UCI Repository of Machine Learning Databases.Available: http://www.ics.uci.edu/~mlearn/mlrepository.htm.
[22] Montreal Neurological Institute, McGill University, McConnell Brain Imaging Centre (McBIC). Available: http://www.bic.mni.mcgill.ca/brainweb.

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)