An Attribute Reduction Algorithm Based on Rough Set Theory and an Improved Genetic Algorithm

Lian Chen
Computer Center, Nanchang University, Nanchang 330031, China
Email: nc_chenlian@163.com

Hongling Liu and Zilong Wan
Computer Center, Nanchang University, Nanchang 330031, China
Email: {nc_liuhongling@163.com, ncuwzl@126.com}

Abstract—**Because the existing attribute reduction algorithms based on rough set theory and genetic algorithm have the main problems: the complexity in calculating fitness function and slow speed in convergence. An attribute reduction algorithm based on rough set theory and an improved genetic algorithm is proposed in this paper. In order to simplify the calculation of fitness function under the condition of keeping the algorithm correct, the relative importance of chromosome is used to define the fitness function. Beyond that, by introducing the core attributes into the initial population and using an improved mutation operator, the algorithm can not only maintain the feature of whole optimization, but also have a quicker convergence speed. The experimental results show that the algorithm can obtain the optimal attribute reduction set of information system quickly and efficiently.**

Index Terms—**rough set theory, relative attribute importance, genetic algorithm, relative importance of chromosome**

I. INTRODUCTION

Rough set theory was proposed by Professor Pawlak in 1982, who worked in Warsaw University of Poland, it is a mathematical tool to deal with imprecise and incomplete knowledge [1]. Its main idea is to use the knowledge that has been mastered to infer the fuzzy or uncertain knowledge with a certain algorithm. It has the following characteristics [2,3]:

(1) When carry out data analysis with rough set theory, we just need the information provided by the data itself, without needing any extra knowledge. Although fuzzy set theory, evidence theory and probability statistics method can also be used to deal with uncertain and incomplete information, but they all need some additional knowledge, while these extra information is often uneasy to get. Compared to these theories, the biggest characteristic of rough set theory is that it does not need to provide additional information but the data themselves when solve a problem. Thus, the analysis of the data based on the rough set theory is more objective, simple and practical.
(2) It is a powerful tool for data analysis and can be used to carry out analysis on the past data to help make a decision in the future.
(3) It can be used to deal with all kinds of irregular data, including the incomplete data and uncertain data.
(4) It can be used for data preprocessing, including fill up missing data, discretize continuous data, attribute reduction and extract rules from decision table, etc.
(5) It can be used to identify and evaluate dependencies among the data and reveal simple patterns of some concepts. Rough set theory becomes an indispensable tool in the field of data mining.

It develops quickly in a short time after being founded, at present, it has been used widely in the fields, such as: expert system, artificial intelligence, decision support system, knowledge discovery (data mining), pattern recognition and medical diagnostics [2], etc. It is complementary to the probability theory, the fuzzy theory, the evidence theory and other theories of processing uncertain and imprecise problems. Especially in the field of knowledge discovery, it can discover the dependencies among data, evaluate the importance of a particular classification, eliminate the redundant attributes, reduce dimension of data set, and acquire the decision rules, etc. Attribute reduction is a very important research field of rough set theory. The so-called attribute reduction is to delete the irrelevant or unimportant attributes under the condition of keeping the classification and the decision-making capacity of the knowledge unchanged [4]. Due to the combinatorial explosion problem of attributes, obtaining a minimum attribute reduction set is an NP-hard problem. The general method to solve this problem

---

The work is supported by the National Natural Science Foundation of China (Grant No.61070139).

Corresponding author.

Email address: nc_liuhongling@163.com (Hongling Liu).

doi:10.4304/jsw.9.9.2276-2282

is up to heuristic algorithms, namely researchers make use of the importance of attributes as heuristic information to get the optimal attribute reduction set. The traditional attribute reduction algorithms of rough set theory are mainly based on the concept of information, the discernibility matrix and the positive region, and so on[4,5,6]. However, many of them are not complete, and even cannot be guaranteed to get the right attribute reduction result. Genetic algorithm has the advantages: global optimization, implicit parallelism, and it does not depend on specific areas, etc, thus it can be used to solve the NP-complete problem [7]. During the whole process, the genetic algorithm basically only uses the fitness function as the evolutionary basis. First, the fitness function of individual in the population is calculated. Then, the fitness values are regarded as the heuristic information to go forward. Therefore, a good design of fitness function is very essential to the genetic algorithm. The major disadvantage of the traditional genetic algorithms is its lower efficiency in their complicated calculation of the fitness function [8]. Especially for a large-scale information system, the time complexity mainly concentrates on calculating the fitness function. The organic combination of rough set theory and genetic algorithm is appropriate for attribute reduction and knowledge discovery.

In order to reduce the search space in the process of attribute reduction and simplify the calculation of fitness function and improve the performance of algorithm, we improve the algorithm from three aspects in this paper: first, the relative importance of attribute [9] is regarded as the heuristic information, second, the fitness function is defined with the relative importance of chromosomes, third, the revising operator is used to ensure the minimum reduction. Experimental results show that the improved algorithm is superior to the existing algorithms in several aspects.

II. ROUGH SET THEORY

Rough set theory is gaining more and more attention from the researchers with its unique advantages, however, it is still a very young discipline and there are many drawbacks. The current research hotspots of rough set theory mainly include the following aspects [5, 6, 10]:

(1) Fill missing data: An information system that have missing data is commonly called an incomplete information system. Due to the classical rough set theory is based on a complete information system, in order to make it suitable for processing the incomplete information system, some measures should be taken to deal with the missing data or extend the rough set model.

(2) Discretize continuous data: As we all know, rough set theory can only be used to deal with discrete data, but in the real world, continuous data exist widely. Therefore, the discretization of continuous data becomes extremely important. At present, there have been some related research in this aspect, but all of these methods more or less have some defects, and there is no an axiomatic method. So the study in this field is still a research hotspot in the future.

(3) The problem of large data set: In the real world, databases are becoming larger and larger, therefore, how to reduce the execution efficiency and complexity of the algorithm, and to find the most useful data from numerous data are challenges of the rough set theory. Although some research achievements have already achieved in this field, but not perfect, so further study still needed.

(4) Efficient attribute reduction algorithm: Attribute reduction is a NP difficult problem, the main reason is the combination explosion among attributes.The attribute reduction methods based on the heuristic knowledge are usually adopted to get an acceptable reduction result. Efficient reduction algorithm is the basis of rough set theory being applied in the field of knowledge discovery. Domestic and foreign scholars have done a lot of research in this aspect, but there is still no a very effective method, so seeking rapid and effective reduction algorithms is still one of the research hotspots in the rough set theory.

(5) Multiple method fusion: Due to the rough set has some shortcomings when processing the data, so it is necessary to combine the rough set theory with other methods. For example, the combination of rough set with neural network, fuzzy set and genetic algorithm, etc. Although some achievements have been made in this aspect, but there are still a lot of problems having not been solved, so further research is still needed.

The main feature of rough set theory is that it is not necessary to provide any priori information, except for the data set needs processing itself. So it is objective. In rough set theory, knowledge is considered to be the ability to classify objects, namely, knowledge means the division of the data [10]. The attribute reduction and the rule extraction are the core of rough set theory [11]. With other analytical methods, such as genetic algorithm, artificial neural networks and data mining, etc, it can enhance the ability to deal with the uncertainty problem further.

Definition 1: Information System S can be expressed as $S = (U, A, V, f)$, where U is a non-empty finite set of objects, called the domain, A is a non-empty finite set of attributes, $V = \bigcup_{a \in A} V_a$, V_a represents the value of attribute A. $f : U \times A \to V$ is an information function, it assigns a value for each attribute of each object, namely, $\forall a \in A, x \in U$, $f(x,a) \in V_a$. If the attribute set A can be divided into condition attribute set C and decision attribute set D, and $A = C \cup D$, $C \cap D = \Phi$. Then the information system can be referred to as the decision system or decision table, where D is generally contains only an attribute.

Definition 2: In information system S, a set $P \subseteq A$, objects $x, y \subseteq U$, binary equivalence relation $\mathrm{IND}(P)=\{(x,y)\in U\times U\mid \forall a\in P, f(x,a)=f(y,a)\}$, called indiscriminative relation of S [12]. Indiscriminative relation is an equivalence relation, we can obtain a division of a decision table or decision system by an indiscriminative relation.

Definition 3: Let R be a cluster of equivalence relation, $p\in R$, if $\mathrm{IND}(R)=\mathrm{IND}(R-\{p\})$, then we called P is unnecessary to R, otherwise P is necessary to R. For each $p\in R$ is necessary, then R is called independent, otherwise R is dependent [12].

Set $P\subseteq R$, If P is independent, and $\mathrm{IND}(P)=\mathrm{IND}(R)$, then P is a reduction of R.

Definition 4: Given a decision table $S=(U,A,V,f)$, where $A=C\cup D$, C, D are the sets of condition attribute and decision attribute respectively, the discernibility matrix $M_D=\{m_{ij}\}$ is defined as[13]:

$$m_{ij}=\begin{cases}\{a\in C\mid f(x_i,a)\neq f(x_j,a)\}, f(x_i,D)=f(x_j,D);\\ \Phi, f(x_i,D)\neq f(x_j,D).\end{cases} \qquad (1)$$

From the definition, we know that the discernibility matrix is symmetric. So we only need to compute the upper triangular or lower triangular [13].

Definition 5: $\mathrm{Core}(R)=\cap\,\mathrm{Re}\,d(R)$, where $\mathrm{Re}\,d(R)$ represents all reduction sets of R.

It should be noted that the result of attribute reduction is not unique, but the core is unique. The set, which includes the least number of attributes is called the optimal attribute reduction set.

Definition 6: Assume R is a tuft of equivalence relation, $p\subseteq R$, $a\in R-P$, the importance of attribute a relative to set P is defined as:

$$\mathrm{SIG}(p\mid a)=\frac{\left|U/_{P}\cup\{a\}\right|-\left|U/_{P}\right|}{U/_{R}}, \qquad (2)$$

where $U/_{P}$ denotes the number of equivalence class of P relative to U.

Property: Each attribute $a\in R$ is unnecessary, which is equal to $\mathrm{SIG}(R-\{a\}\mid a)=0$.

Proof: Necessity—suppose $a\in R$ is unnecessary, according to the definition of the necessity, we can know:

$$\mathrm{IND}(R-\{a\})=\mathrm{IND}(R)$$
$$\Rightarrow U/(R-\{a\})=U/R$$
$$\Rightarrow \mathrm{SIG}(R-\{a\}\mid a)=\frac{|U/R|-|U/(R-\{a\})|}{|U/R|}=0$$

Sufficiency— suppose $\mathrm{SIG}(R-\{a\}\mid a)=0$,

Then, $|U/R|=|U/(R-\{a\})|$

Because of $R-\{a\}\subseteq R$, for $\forall X\in U/R$, existing $X\subseteq Y$,making $U/(R-\{a\})=U/R$, so $U/(R-\{a\})=U/R$, namely, $\mathrm{IND}(R-\{a\})=\mathrm{IND}(R)$. So the attribute $a\in R$ is not unnecessary.

With the definition of relative importance of attribute and its property, we can know that the greater the importance of the attribute a is relative to the set P, the more number of increased equivalence class and the more close to the division of attribute set is relative to the domain when add the attribute a into the set P.

Definition 7: The relative importance of chromosome—the relative importance values of attributes corresponding to the gene bit 1 in the chromosome x are cumulative.

$$\mathrm{SIG}(x)=\sum_{i=j_1}^{i=j_k}\mathrm{SIG}(x\mid a_i),\ i\in\{j_1,j_2,\dots,j_k\} \qquad (3)$$

III. GENETIC ALGORITHM

In fact, attribute reduction is a kind of combinatorial optimization problem. As genetic algorithm is a global search method based on biological evolution, so it is a good idea to combine genetic algorithm and rough set theory to carry out attribute reduction. Attribute reduction algorithms based on rough set theory and genetic algorithm usually can get the optimal or the relative optimal attribute reduction result.

Genetic algorithm (GA) is put forward for the first time in the late 1960s by John Holland, who was a professor in the University of Michigan, its basic idea is to form a kind of algorithm for searching optimal solution by simulating the natural genetic mechanism and the biological evolutionism. It can solve the optimization problems of complexity, nonlinear and even involving space. At present, it has been used widely in the aspects, such as: computer science, engineering design, automatic control, pattern recognition, fault diagnosis and data mining [14], and so on. It has the following features [15]:

(1) It is a kind of intelligent algorithm that has self-learning, self-organization and adaptive ability. In the design process of traditional algorithms, all the features and attributes of the problems need to be described in advance, according to the different characteristics of the problems, the algorithms should take different measures. It is because that genetic algorithm does not need to do this, so it can be used to solve complex and unstructured problems.

(2) It can deal with the parameters of code set directly, rather than the parameters of the problem itself.

(3) It uses fitness function to evaluate the intermediate individuals and guide the search direction in the process of search. The search process is neither bound by the continuity of the optimizing function, nor by the derivability of the optimizing function.

(4) It is a kind of parallel algorithm. For each iteration, it is based on the population rather than an individual to complete the search process in the solution space.

(5) Its expression is easy to understand, its basic idea is simple, and its operation mode and implementation steps are standard.

Genetic algorithm includes five factors [16]: encoding, generate initial population, compute the fitness function, determine the genetic operator and set the relative parameters. The flowchart of genetic algorithm is showed as follows:

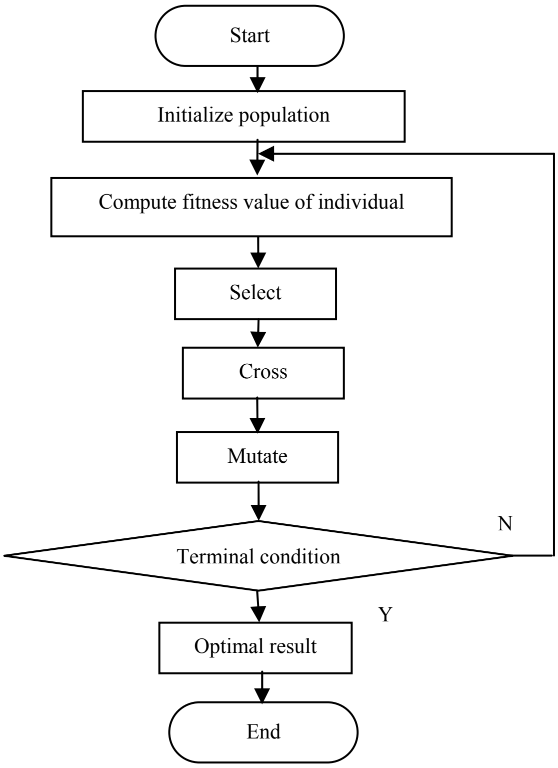

Figure I Flow chart of genetic algorithm

A. Encoding Method

The optimization process of genetic algorithm cannot act on the data in the solution space directly, but can act on the corresponding code, which is expressed as a genotype string in the genetic space through encoding [7]. Fixed-length binary array of one-dimensional is used here, the length of chromosome equals to the number of condition attributes. Each gene bit corresponds to a condition attribute, 1 denotes that the attribute is selected, and 0 denotes that the attribute is not selected.

B. Generate Initialize Population

The initial population is generated as the way: the gene bits corresponding to the attributes in the core are set as 1, the others are set as 0 or 1 randomly.

C. Design Fitness Function

According to the definition of attribute reduction, the fitness of individuals mainly depends on two aspects [8]: the number of attributes that they contained (it should be as little as possible) and their ability to distinguish between individuals (it should be more powerful). Thus, the fitness function is defined as follows (the fitness function of the individual x is defined as):

$$f(x)=\frac{1}{\text{count}(x)}+\text{SIG}(x) \tag{4}$$

D. Selection Operation

We use the optimal preservation strategy [14] which is the most commonly used method of selection to select the individuals. Set the number of population as n, and the fitness function of the individual i is f_i, then the probability of individual i selected is p_{si} [12]:

$$p_{si}=\frac{f_i}{\sum_{i}^{n} f_i} \tag{5}$$

Probability p_{si} reflects the proportion of the fitness value of the individual i in the sum of fitness values of the all individuals.

After getting a new generation of individuals using the optimal preservation strategy, in order to make the algorithm have a fast speed of convergence, if the fitness value of the worst individual in the current generation is less than the fitness value of the best individual in the previous generation, then we use the best individual in the previous generation to replace the worst individual in the current generation. This method can ensure that the individuals getting from the genetic process are not damaged by the crossover and mutation operators, it is an important guarantee for the convergence of genetic algorithm. It is also easy to make the local optimal individuals not be eliminated, so it can strengthen the global search ability of the algorithm.

E. Cross Operation

Cross operation is the main method to generate new chromosomes. We make use of the single-point crossover strategy here. An individual is chosen to take part in crossover at certain probability p_c. For each pair of individuals, a cross-point is selected randomly, and we exchange part chromosomes of the individuals on the cross-point, thus, next generation of individuals will be generated through the process.

F. Mutation Operator

General mutation operator reversals bits randomly according to a given mutation probability[13]. The heuristic information that is expressed as the relative importance of attribute is used to describe the mutation operator. So we can keep the classification ability of the selected subset of attributes unchanged and the gene bit corresponding to the attributes in the core immutable. For a given bit string of chromosome $s=a_1a_2...a_n$, specific operation is as follows:

$$a_i^{'}=\begin{cases}1-a_i, \max(\text{SIG}(s\mid a_i))\\ a_i, \min(\text{SIG}(s\mid a_i))\end{cases} \tag{6}$$

Then, a new individual $s^{'}=a_1^{'}a_2^{'}...a_n^{'}$ is generated, where $i\in\{j_1, j_2,..., j_k\}$, $j_l(l=1,2,...,k)$ denotes that the gene bits corresponding to the attributes without being included in the core.

IV. ALGORITHM DESCRIPTION

The main ideas of the improved algorithm in this paper are: (1) In order to simplify the calculation of fitness function under the condition of keeping the correctness of the attribute reduction result, we make use of relative importance of chromosome to define the fitness function. (2) In order to speed up the convergence of the algorithm, we introduce the attributes in the core to initialize the population. (3) To realize the local search capability, we introduce into the heuristic information by constructing a new mutation operator. Thus, we make the algorithm not only maintain the overall optimization features and has faster speed of convergence. The specific description of the improved algorithm is as follows:

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

Output: An optimal attribute reduction result of the attribute set A.

Step 1: Obtain the core of the information system, the first generation population of the binary string are initialized randomly. For the attributes in the core, the corresponding bits are set to 1, the other bits are set to 0 or 1 randomly.

Step 2: First, compute the relative importance of all chromosomes according to the definition, then, compute the fitness values.

Step 3: According to the probability of selection, using the roulette wheel selection method and the optimal preservation strategy to act selection operator on the population.

Step 4: Using the single-point crossover strategy and the probability of crossover to act crossover operator on the population.

Step 5: Using the proposed heuristic mutation operator to act mutation operator on the population.

Step 6: In order to ensure the algorithm running in the feasible solution space, we restore the population according to the revising operator. If the fitness value of the best individual is keeping constant after continuous breeding for many generations, then the algorithm terminates. Then we can reduce the number of attributes in the attribute reduction set according to the relative algorithm below. Otherwise, return to Step 3.

Modified algorithm—for each chromosome:

(1) Start from the beginning, set the gene bit of 0 as 1 until the positive domain of the attribute set equals to the positive domain of the original attribute set.

(2) From the first to the last one, set the gene bit of 1 as 0, if the positive domain of the attribute set equals to the positive domain of the original attribute set, go forward the above process. Otherwise, reset the gene bit as 1.

Reduce the number of attributes in the attribute reduction set:

(1) If the attribute reduction set R contains attributes that their relative importance have not been computed, then transfer to (2), otherwise terminates.

(2) Select an attribute that has not been dealt with in R, then compute its relative importance of attribute $\mathrm{SIG}(R-\{a\}\mid a)$, if $\mathrm{SIG}(R-\{a\}\mid a)=0$, then set the corresponding bit from 1 to 0.

TABLE I
DISEASE DECISION TABLE

object	a	b	c	d	e	f	g	h	i	D
U1	−1	−1	−1	0	−1	0	0	0	0	1
U2	−1	0	0	0	−1	−1	0	0	0	2
U3	0	−1	−1	0	−1	0	0	0	0	1
U4	0	0	−1	0	−1	0	1	0	0	4
U5	−1	0	0	0	−1	0	0	−1	0	5
U6	0	0	0	0	−1	0	0	0	0	3
U7	0	0	0	1	−1	0	0	0	0	6
U8	0	−1	0	0	−1	0	0	−1	0	1

V. EXAMPLE ANALYSIS

TABLE I is a decision table which is based on the agricultural data of crop disaster [17, 18].

Where C is the set of condition attributes, decision attribute is D means the disease-level.

C = {a = description of the symptom,
b = the shape of the disease,
c = disease - location,
d = disease - rate,
e = time of onset,
f = the color of the diseased parts,
g = relative humidity,
h = the number of rainy days,
i = precipitation }

Through this example, the improved attribute reduction algorithm is analyzed here. To the data in TABLE I, we also experiment on the attribute reduction algorithm based on rough set theory in Ref.[13], each parameter in the experiment is set as follows:

the size of the population is $n=40$, the probability of crossover is $P_c=0.65$. Through running 10 times, we get the average time: 1923ms and the attribute reduction result is $\{a,c,d,h\}$. Under the same conditions, running the improved algorithm 10 times, we get the average time: 1086ms, at this point, not only the individuals in the population tend to be the same for 101100010 (the corresponding condition attributes are: description of

TABLE II
COMPARISON OF REDUCTION ALGORITHMS

algorithm	attribute reduction set	average running time
algorithm in Ref.[10]	$\{a,c,d,h\}$	1923ms
the proposed algorithm	$\{a,c,d,h\}$	1086ms

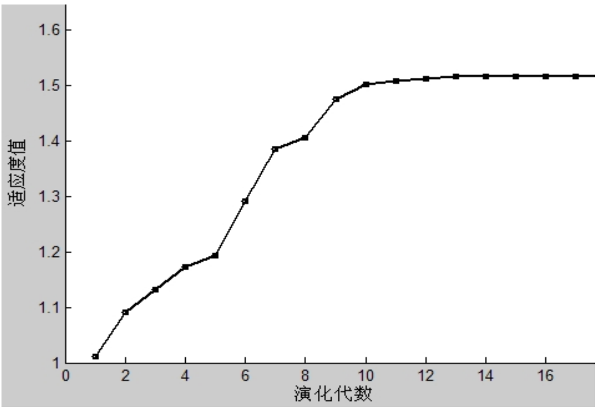

FIG I. Experimental result

TABLE III
EXPERIMENTAL RESULT

The algebra of population	The optimal individual	The fitness values of the optimal individuals
1	101101001	1.011531
4	101100101	1.172526
7	101101010	1.385790
10	101101010	1.502213
13	101100010	1.517229
15	101100010	1.517229
18	101100010	1.517229

symptom, disease-location, disease-rate and the number of runny days, namely, a , c , d and h), but also the fitness values of the individuals tend to be the same for 1.5 . So the final attribute reduction result is $\{a,c,d,h\}$. As is shown in TABLE II, TABLE III and FIG I.

From the experimental results, we know that the improved algorithm not only simplifies the calculation of the fitness function, ensures the correctness of the attribute reduction result and raises the efficiency of the process, but also reduces the time complexity of attribute reduction greatly and keeps the classification ability of the selected attribute subset unchanged. Observing the computing process, we know that the algorithm in Ref.[13] cannot really keep the gene bits corresponding to the attributes in the core unchanged, while the improved algorithm can guarantee the gene bits corresponding to the attributes in the core unchanged.

VI. CONCLUSION

By redefining the fitness function, the improved algorithm can simplify the calculation of the fitness function. We introduce the heuristic information through constructing a new mutation operator to keep the overall optimization feature of genetic algorithm while maintaining its fast rate of convergence. The experimental results show that the improved algorithm can quickly and efficiently obtain the optimal attribute reduction set of the decision table. The disadvantage of this algorithm is that it can only be applied for attribute reduction of complete decision table, the next step we will improve it further to deal with attribute reduction of incomplete decision table by revising the heuristic standards appropriately.

ACKNOWLEDGMENT

The work is supported by the National Natural Science Foundation of China (Grant No.61070139). The authors wish to thank the anonymous reviewers for their valuable comments and suggestions to improve the presentation of this paper.

REFERENCES

[1] Z. PAWLAK. “Rough sets,” *Information and Computer Science*, vol. 11, no. 5, pp. 41-356, 1982.

[2] Min Li. “Quick Attribute Reduction Based on Approximation Dependency Degree,” *Journal of Computers*, vol. 8, no. 4, pp.920-927, 2013.

[3] Yandong Zhai, Chunguang Zhou and Yingjuan Sun, etc. “Approach of Rule Extracting Based on Attribute Significance and Decision Classification,” *Journal of Computers,* vol. 7, no. 2, pp. 514-520, 2012.

[4] Y. L. Ye, and Y. San. “The mixed data attribute reduction of rough set genetic algorithm-based,” *Journal of Harbin Institute of Technology*, vol. 40, no. 5, pp. 683-687, 2008.

[5] Z. J. Yang, H. Ge, and Z. S. Wang. “Research overview on rough set-based attribute reduction methods,” *Journal of Computer Application*, vol. 29, no. 1, pp. 16-20, 2012.

[6] Taorong Qiu, Yuyuan Lin and Xiaoming Bai. “The difference degree of condition attributes and its application in the reduction of attributes,” *Journal of Computers*, vol. 7, no. 5, pp. 1087-1091, 2012.

[7] J. Q. Cao, and J. P. Hong. “An attribute reduction algorithm of rough set based on genetic algorithm,” *Mechanical Engineering and Automation*, no. 1, pp.28-30, 2011.

[8] B. Yang, Z. Y. Xu, and W. H. Shu. “A fast attribute reduction of rough set based on genetic algorithm,” *Mini-Micro Systems*, vol. 33, no. 1, pp. 140-144, 2012.

[9] P. Wang, X. F. Wang, and G. F. Wu. “Genetic algorithm-based attribute reduction algorithm of rough set,” *Computer Software and Theory*, vol. 25, no. 5, pp. 42-44, 2008.

[10] H. M. Li, G. H. Zhou, and K. J. Wang. “Method of knowledge discovery based on rough set and genetic algorithm,” *Journal of Computer Application*, no. 8, pp. 76-78, 2007.

[11] Y. Hu, Y. J. Zhang, and D. M. Yang. “The application in data mining of rough set and genetic algorithm,” *Journal of Computer Application*, 2006, 26(6): 98-99,107.

[12] W. Pan, Y. F. Wang, and Y. San. “Adaptive genetic algorithm-based knowledge reduction algorithm of

rough set," *Computer Engineering and Application*, vol. 44, no. 15, pp. 1-3, 11, 2008.

[13] X. Chen, J. Lei, and M. Fu. "Improved genetic algorithm-based attribute reduction algorithm of rough set," *Computer Engineering and Design*, vol. 31, no. 3, pp. 602-604, 2010.

[14] Zhengjiang Wu, Jingmin Zhang and Yan Gao. "An Attribute Reduction Algorithm Based on Genetic Algorithm and Discernibility Matrix," *Journal of Software*, vol. 7, no. 11, pp.2640-2648, 2012.

[15] Xia Bian, Liang Mi. "Genetic Algorithm (GA) Theory and Its Progress on Application and Research," *Computer Application Research*, vol. 27, no. 7, pp. 2425-2429, 2010.

[16] Shifei Ding, Xinzheng Xu, Hong Zhu. "Studies on Optimization Algorithms for Some Artificial Neural Networks Based on Genetic Algorithm (GA)," *Journal of Computers*, vol. 6, no. 5, pp. 939-946, 2011.

[17] Lian Chen, Hongling Liu, Zilong Wan. "An Improved Attribute Reduction Algorithm and Its Application in Agriculture," *Journal of Information and Computational Science*, vol. 11, no. 5, 2014.

[18] C. Y. Xia, X. B. Feng, M. S. Luo, and W. Sun. "The application of rough set theory in agriculture," *Journal of Anhui Agriculture*, vol. 39, no. 30, pp. 18391-18392, 2011.

Lian Chen received the B.E. degree in computer science from Chengdu Telecom Engineering college (University of Electronic Science and Technology), Sichuan, China in July 1984. He is a Professor at the Computer Center, Nanchang University, Jiangxi, China. His research interests include database theory and technology, software engineering. He has published more than 40 academic papers, including decision table reduction of rough set based on the binary, attribute reduction method based on matrix method clearly, intelligent data analysis of air quality based on rough set theory and so on. Most of them were published in the Chinese core journals.

Hongling Liu received the B.E. degree incomputer science and technology from Jianghuai College of Anhui University in June 2011, Anhui, China. She is currently a graduate student and working towards her Master of IT at Nanchang University, Jiangxi, China in June 2014. Her current research interest includes computer software and theory.

Zilong Wan received the B.E. degree in computer software from Software College of Nanchang University, Jiangxi, China in June 2011. He is currently a graduate student and working towards his Master of IT at Nanchang University, Jiangxi, China in Jun 2014. His current research interest includes computer software and theory.