

# Information Attribute Reduction Based on the Rough Set Theory

YU Bo-wen, Du Chao, Wang Zhi-guo

**Abstract**—The genetic algorithm is used to optimize the algorithm of attribute reduction in data preprocessing, and the rough approximation precision in the rough set theory is utilized to determine the importance of information attribute. From which the decision table is constituted by selecting the attributes which have higher degree of attribute importance, and the attribute core of decision information is obtained by using the identification matrix. The initial population is constructed on the standard of the attribute core, the search area of genetic algorithm is reduced. Finally, the correction operator based on the rough approximation precision is introduced, and the algorithm is made to conduct in the correct solution space, thus the speed of attribute reduction is improved, furthermore, the optimal results of attribute reduction are obtained.

**Index Terms**—Genetic algorithm, The rough set, The rough approximation precision, Attribute reduction.

## I. INTRODUCTION

Attribute reduction is one of the main research direction of rough set theory. Attribute reduction is to keep the original data classification ability under the premise of get rid of those who are not related to characterize the properties. It has been proved the computation of minimal reduction and full reduction both is NP-hard problem, information of permutation and combination is the important factor to the NP-hard. For example, Guo-yin Wang, using the subset was proposed to calculate the minimalist attribute reduction, the method of time complexity is exponential, exponentially with the increase of attributes its complexity growth, so this method is not suitable for practical application. Duo-Qian Miao proposed a reduction method based on information entropy and information entropy to define attribute importance often cannot get the smallest reduction, is also likely to get reduction results.

For information attribute reduction, optimization algorithm is usually adopted, which adds and sets standard information according to the characteristics of attribute information to reduce the search attribute range to obtain the desired final result. According to researches, the attribute reduction in the rough set theory can be regarded as a combination optimization process, so the genetic algorithm can be introduced into attribute reduction. Genetic algorithm refers to a global search algorithm, and is featured by good stability and parallel execution ability.

In the process of past research, a new heuristic genetic algorithm for production scheduling is proposed by Jian-hua

dai, select individual state to establish the initial group by means of random, this method requires a great deal of calculation and evolution algebra to search all state, this algorithm under a large amount of information efficiency to be improved. describe the importance of attribute rough approximation accuracy is proposed by Dong-yi ye, according to the concept of rough approximation precision a greedy algorithm is designed, it's Characteristics is to realize the process is relatively simple, can quickly find attribute reduction in a large amount of condition attributes.

In this paper, the concept of attribute significance is introduced on the basis of the concept of rough approximation accuracy, and attribute information of high attribute significance among the original attribute information is selected according to this standard to form the initial population of genetic algorithm, to narrow the search range of solution space and raise the speed of searching the optimal reduction results. At last, it is proved through experiments that the genetic algorithm based on the rough set theory can greatly raise the accuracy of reduction results and reduction efficiency.

## II. BASIC CONCEPTS

### A. Rough Set

The rough set theory is a mathematic tool that allows various interferences such as inaccurate analysis, disaffinity and incompleteness put forward by a Polish mathematician named Pawlak.Z in 1982. Through scholars' unremitting efforts in studying the properties and laws of rough set in the past years, a substantive leap has been made in the rough set theory. Since rough set is superior in data preprocessing, rough set has a good application prospect in the field of data mining. Rough set is useful for standardizing and denoising data, processing missing data, data reducing, and identifying correlation. Rough set has been successfully applied in other related fields. Thus, the rough set theory is of great significance to the field of data mining.

### B. Genetic Algorithm

In the 1960s, an American professor Holland put forward a new theory. He had built an artificial intelligence model by imitating the biological evolutionism base on Darwin's theory of "survival of the fittest", and applied the idea of genetic variation of organisms for adapting themselves to changes in nature in the process of optimization. It is called genetic algorithm (GA). With research and development in recent years, great achievements of the application of the genetic algorithm in other fields have been made. Genetic algorithm is a bionic algorithm, namely carrying out space optimization search by simulating the process of organisms changing with the environment, performing genetic operation over individuals via genetic operators, and forming a new

YU Bo-wen, Automatization Engineering College, Nanjing University of Science and Technology, Nanjing, China, +86 18545055186.

DU Chao, Second Team of Special Process, Underground Operation Branch of Daqing Oilfield Company, +86 13936955212.

Wang Zhi-guo, Planning and Design Institute, Petrochina Daqing Oilfield No.9 Production Plant, +86 18645913307.

evolutionary population with evolved individuals.

C. Attribute Significance Based on Rough Approximation Accuracy

1) Rough Approximation Accuracy

Definition of rough approximation accuracy: Suppose  $P \subseteq C$ ,  $\{X_1, X_2, X_3, \dots, X_k\}$  is expressed by the decision attribute of  $U$ , and  $X \subseteq U$ , then the approximation accuracy of the attribute set  $P$  is

$$a_p = \frac{\text{card}(P_-(X))}{\text{card}(P^-(X))} . \text{card}(X) \text{ refers to the number of}$$

$X$ , based on which the degree of attribute reduction of the set can be determined. The smaller its value is, the higher the reduction degree is. The rough approximation progress of the

attribute set  $P$  by  $L$  is:  $\gamma_p(L) = \sum_{i=1}^k \frac{\text{card}(P_-(X_i))}{\text{card}(U)}$ .

Definition 1: The rough approximation accuracy of the

attribute set  $P$  by  $L$  is:  $\alpha_p(L) = \sum_{i=1}^k \frac{\text{card}(P_-(X_i))}{\text{card}(P^-(X_i))}$ .

2) Attribute Significance

According to the definition of rough approximation accuracy, it is put forward in this paper to calculate attribute significance based on rough approximation accuracy, to describe the significance level of attribute information. Suppose  $C$  refers to basic attribute,  $D$  to decision attribute,  $C$  contains  $n$  basic attributes  $C_1, C_2, \dots, C_n$ ,  $L$  by the decision attribute is expressed as  $\{x_1, x_2, \dots, x_n\}$ , the rough approximation accuracy of each attribute can be calculated out, and the expectation  $\alpha_{C_i}$  and variance  $\sigma_{C_i}$  of the  $k+2$  values can be calculated out based on the  $L$  rough approximation accuracy.

Definition 2:  $\mathcal{C}_i$  significance attribute function:

$$S_{C_i} = \frac{\alpha_{C_i}}{\beta + \sigma_{C_i}} , \text{wherein } \beta \text{ refers to an auxiliary parameter.}$$

D. Deduction and Proof

Deductions 1 and 2 are obtained according to the definition of rough approximation accuracy:

Deduction 1: Suppose a decision information system  $S = (U, C \cup D, V, f)$ , then its  $L$  is divided according to  $U$  based on the decision attribute  $D$ , so  $\gamma_C(L) = 1$ .

Proof: Suppose  $U / IND(C) = \{X_1, X_2, X_3, \dots, X_n\}$ , in which  $n$  refers to the number of classification of conditions of characterization by the domain of discourse  $U$ . According to

the definitions above,  $\gamma_C(L) = \sum_{i=1}^k \frac{\text{card}(P_-(X_i))}{\text{card}(U)}$ , in

which  $k$  refers to the statistic of the decision attribute of the decision information. In the decision information table, the division of the domain of discourse  $U$  by the conditions of

characterization  $C$  depends on the division of the domain of discourse  $U$  by the decision attribute  $D$ , thus  $IND(C) \subseteq IND(D)$ ; and the conditions of characterization  $C$  divides the domain of discourse  $U$  into several classifications  $X_i$ , and clearly defines it in  $L$ , so

$$\text{card}(U) = \sum_{i=1}^n \text{card}(X_i) . \text{According to the definition of}$$

$$\gamma_C(L) = \sum_{i=1}^k \frac{\text{card}(P_-(X_i))}{\text{card}(U)} , \gamma_C(L) = 1 .$$

Deduction 2: The components modified with modified operators all comply with  $POS_{C'}(D) = POS_C(D)$ .

Proof: The stop condition of the modification process is  $\gamma_{C'}(L) = \gamma_C(L)$ . According to Deduction 1,

$\gamma_{C'}(L) = \gamma_C(L) = 1$ . If  $C' = \{C'_1, C'_2, \dots, C'_r\}$  and

$L = \{L_1, L_2, \dots, L_k\}$  are equivalence relation sets of  $C$  and  $D$ , it can be obtained that

$$\gamma_{C'}(L) = \sum_{i=1}^k \frac{\text{card}(P_-(L_i))}{\text{card}(U)}$$
 according to the definition of

rough approximation accuracy as above.

Suppose  $POS_{C'}(D) \neq POS_C(D)$ , then  $\exists x \in U$  and

$$x \notin POS_C(D) , \exists x \notin \bigcup_i^d [x]_{C'} \subseteq L_i ,$$

$$\text{and } [x]_{C'} \not\subseteq \bigcup_{i=1}^d L_i = U \Rightarrow \exists j [x]_{C'} \cap L_j \neq \emptyset \Rightarrow \gamma_{C'} < 1 .$$

The result contradicts the proposition, then  $POS_{C'}(D) = POS_C(D)$ .

III. GENETIC ALGORITHM BASED ON ROUGH SET THEORY

The genetic algorithm based on the rough set theory has preserved the basic characteristics of the original genetic algorithm; and is to select attributes at high significance level to form an information decision system according to the attribute significance of Definition 2, calculate the core of attribute reduction with a discernibility matrix, and determine the initial population of the algorithm according to the core attribute, to develop a strong search capability of the algorithm within the local search space. Besides, modified operators based on rough approximation accuracy have introduced into the algorithm to reduce the population, determine that every chromosome corresponds to a candidate reduction, and guarantee the algorithm is operated in a correct solution space by the constraint of rough approximation accuracy. The algorithm in this paper also has the overall optimization characteristics of the original algorithm, and has improved the speed of attribute reduction and reduction accuracy. The details of the algorithm are as follows:

Input: Information decision system  $S = (U, C \cup D, V, f)$ , crossover probability  $P_r$ ,

selective probability  $P_m$ , attribute significance parameter  $\beta$ ;

Output: Attribute reduction of  $(C \cup D)$ ;

1) Calculate the rough approximation accuracy by the

$$\text{condition attribute set } C: \gamma_C(L) = \sum_{i=1}^k \frac{\text{card}(P_-(X_i))}{\text{card}(U)};$$

2) Calculate the significance  $S_{C_i} = \frac{\alpha_{C_i}}{\beta + \sigma_{C_i}}$  of each

characterization  $C_i$ , sequence the calculations, identify the attributes of which the function value of significance attribute is large, and form  $S_C$  therewith. Calculate the attribute core of the decision system with the discernibility matrix, generate an initial population, and determine the generation number of the population,  $k=1$ ;

3) Calculate the fitness information of every elements of the population according to the fitness function preset in the system;

4) Judge whether the algorithm is terminated:

If the result complies with the termination end, the algorithm is terminated;

If not, the selective probability of corresponding individuals should be calculated, so as to generate a new population,  $k=k+1$ ;

5) Perform crossover operation over individuals;

6) Perform mutation operation over individuals;

7) Modify the result, determine that the attributes of which the attribute significance level is high are the search range of attributes, and return to the third step of the algorithm;

The algorithm flow chart is as shown in Fig. 1:

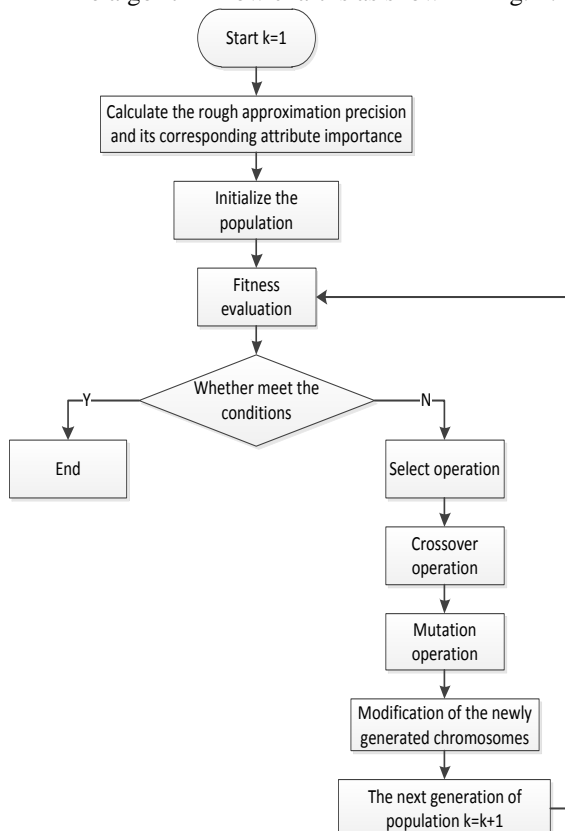


Fig. 1 Flow Chart of Improved Algorithm

#### IV. ALGORITHM ANALYSIS

##### A. Modified Operator

Reduce the population with modified operators, determine that every chromosome corresponds to a candidate reduction, guarantee the algorithm is operated in a correct solution space by the constraint of rough approximation accuracy, and calculate the local optimal solution based on the chromosomes in the  $k+1$  optimization result. According to Deduction 2, in modification and verification, the solution space can be planned on the basis of rough approximation accuracy, which should be specifically determined by the correlation between  $\gamma_R(L)$  and  $\gamma_C(L)$ . Select a characterization of which the rough approximation accuracy value is comparatively larger from the characterization set  $C$  not included in the  $k$ th generation of optimization result, and add it into the planned search space, to make preparations for getting satisfactory attribute reduction. The specific flow chart is as shown in Fig. 2.

Steps:

1) Calculate the rough approximation accuracy of the existing characterization set:

$$\gamma_C(L) = \sum_{i=1}^k \frac{\text{card}(P_-(X_i))}{\text{card}(U)}; \text{ If } \gamma_C(L) < \gamma_R(L), \text{ repeat}$$

steps 2 and 3; If  $\gamma_C(L) = \gamma_R(L)$ , turn to step 4;

2) Select the maximum  $S_{C_i} = \frac{\alpha_{C_i}}{\beta + \sigma_{C_i}}$  from the characterization set;

3) If the code position in correspondence with  $C_j$  is "0", change it into "1", and turn to Step 1;

4) Modification ends.

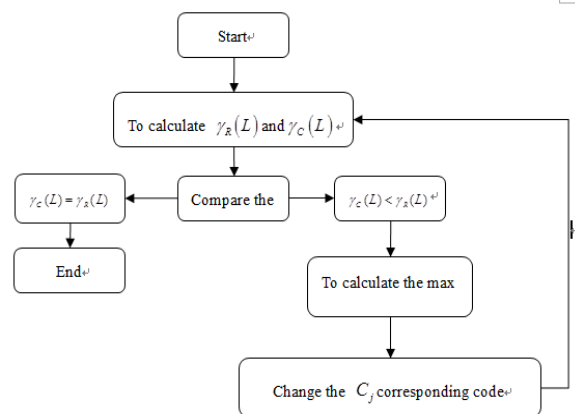


Fig. 2 Modification Flow

##### B. Setting of Initial Population Size

In the reduction process based on genetic algorithm in Literature [9][10], the initial population is not set. In practice, if the initial population is approximate to the problem solution, the time taken to solve algorithm can be saved, and it is easier to calculate the optimal solution with the algorithm. Therefore, the setting of the initial population is of great importance to the execution efficiency of genetic algorithm. If the population size is set to be large, the population can be restrained from premature convergence, which complicates the execution of algorithm; if the population size is set to be

excessively large, the optimal performance (reduction of system attributes) of the algorithm may decline. Thus, an ideal population size should be:

$$m = 2^{\delta_s/2}, (\delta_s = (n-1)(1-p_s)/p_c \geq \delta(H))$$

Wherein,  $\delta(H)$  refers to the length of template,  $n$  to the number of coded message positions of individual chromosome,  $\delta(H)$  to crossover probability, and to selective  $p_s$  probability.

Suppose an information system has its own core, then in the discernibility matrix, if there is a line consisting of only one "1" (the other elements all are "0"), it indicates that the attribute cannot be distinguished, and is the attribute core. Every core corresponds to only one decision information table, and is in the reduction of its corresponding decision information table. For purpose of this paper, attributes of which the significance function value is high will be selected to form the initial population of the algorithm, to further improve the selection quality of the initial population. The core attribute codes in correspondence with all chromosomes of the initial population all are "1", and every initial population contains the required core attributes.

### C. Coding

Binary sequence is adopted to represent chromosome information. Each code position has its own specific attributes. "0" and "1" are used to distinguish whether the selected information set contains an attribute value or not. The initial population should be set to contain these core attributes. For example, suppose the attribute core of a decision table  $\{a_1, a_2, a_3, \dots, a_{10}\}$  subject to 10 condition attributes is  $\{a_1, a_2, a_6\}$ , then it is required to select chromosomes of which the codes are 11\*\*\*1\*\*\*\* to form the initial population.

### D. Fitness Function

The target value approximation, selection of proper attribute set, and global control ability in algorithm all are determined by a fitness function. The selection of fitness function determines the convergence direction of algorithm. Attribute reduction is to obtain a minimalist attribute set and own the original information processing ability. A reduction set should meet: minimum number of attributes, and attribute classification ability.

Since the rough approximation accuracy of attribute is set as the criterion for determining the initial population in this paper, it is not necessary to consider attribute classification ability, and it only needs to consider the number of attributes. Hence, fitness function is defined as:

$$f(x) = \text{card}(c) - \text{count}(x)$$

$\text{card}(c)$  refers to the number of condition attributes;  $\text{count}(x)$  to the number of condition attributes contained in chromosome.

The fitness function of the improved algorithm in this paper is  $f(x) = \text{card}(c) - \text{count}(x)$ . According to this formula, the fitness function equals to the difference between the number of condition attributes and that of condition attributes contained in chromosome. Since rough approximation accuracy is introduced, it only needs to control the difference between the said numbers, which simplifies the

calculation process of the algorithm. Since the solution space is normalized by rough approximation accuracy, and relatively important condition attributes are determined, with algorithm execution, the smaller the number of condition attributes contained in chromosome is, the closer the result is to the required minimalist reduction. Hence, the larger the difference value between the number of condition attributes and that of condition attributes contained in chromosome is, the closer the result is to the minimalist reduction.

### E. Feasibility and Performance Analysis

#### 1) Feasibility of Fitness Function

The fitness function of the improved algorithm in this paper is  $f(x) = \text{card}(c) - \text{count}(x)$ . According to this formula, the fitness function equals to the difference between the number of condition attributes and that of condition attributes contained in chromosome. Since rough approximation accuracy is introduced, it only needs to control the difference between the said numbers, which simplifies the calculation process of the algorithm. Since the solution space is normalized by rough approximation accuracy, and relatively important condition attributes are determined, with algorithm execution, the smaller the number of condition attributes contained in chromosome is, the closer the result is to the required minimalist reduction. Hence, the larger the difference value between the number of condition attributes and that of condition attributes contained in chromosome is, the closer the result is to the minimalist reduction.

#### 2) Algorithm Complexity Analysis

For the algorithm, the evolutionary direction of the population are effectively controlled with rough approximation accuracy and fitness function, so that the result gets closer and closer to the minimalist reduction, and finally reach the minimalist reduction of the decisive system. If these conditions are not set, the solution space range of the algorithm is  $2^m$ ; suppose the number of attribute cores of the decisive system is  $n$ , after rough approximation accuracy is taken as the inspiration information, the solution space is narrowed to  $2^{m-h}$ . It is thus clear that the introduction of inspiration information into the algorithm narrows the search space.

#### 3) Algorithm Convergence Analysis

Usually, the convergence of genetic algorithm is determined by the design, crossover probability and mutation probability of fitness function. Definition 3: If  $X_e'(s)$  is the off-line performance of implementation strategy  $\mathcal{S}$  in the environment  $e$ , then:

$$X_e'(s) = \frac{1}{T} \sum_{t=1}^T f_e'(t)$$

In the off-line performance,  $f_e'(t) = \text{best}\{f_e(1), f_e(2), \dots, f_e(t)\}$ ,  $f_e(t)$  refers to the  $t$ th generation of fitness function relative to the environment  $e$ . For purpose of this paper,  $f(x) = \text{card}(c) - \text{count}(x)$ . Off-line performance is used to describe the average value of the predefined fitness function of the algorithm, with which the convergence of the algorithm can be measured. For building the fitness function, the chromosome individuals are taken into account in detail.

Since attribute core information is introduced at the time of building the initial population, the fitness function is designed to equal to the difference value between the total number of attribute characterizations and the number of reduced attribute characterizations. Besides, a modification strategy based on rough approximation accuracy is introduced in follow-up execution. Comparing with algorithm for which no attribute core information is introduced, this algorithm is helpful to approach the output result faster, to prevent the expectation of each generation of population from greatly changing, and then make this algorithm have a better convergence.

F. Termination Conditions

Since there are no definite termination conditions or model for attribute reduction, there is no definite termination function. According to the actual attribute reduction flow, if the fitness function values of consecutive k generations of population don't change, it can be regarded that the desired optimization result has been achieved, and then the algorithm execution can be terminated.

V. INFORMATION REDUCTION

A. Figures and Tables

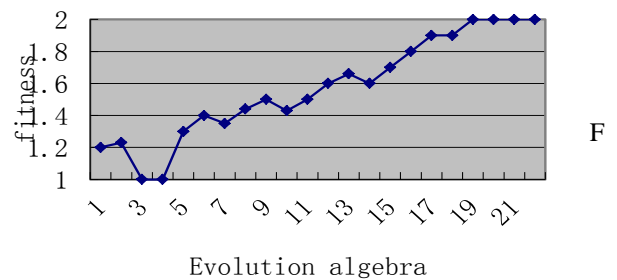
B According to the decisive information table given in Literature [8], as shown in Tab. 1, the reduction result is  $\{a, c, d\}$  and  $\{b, c, d\}$ . According to the reduction algorithm in the Literature[10], the attribute reduction result of Tab. 1 is  $\{a, b, c\}$ ,  $\{a, b, d\}$ ,  $\{a, c, d\}$ , and  $\{b, c, d\}$ , while the real minimalist reduction of the decisive information table is  $\{c, d\}$ . Thus,  $\{a, b, c\}$  and  $\{a, b, d\}$  are not correct reduction results,  $\{a, c, d\}$  and  $\{b, c, d\}$  are not the minimalist reduction.

Tab. 1 Literature [10] Decision information table

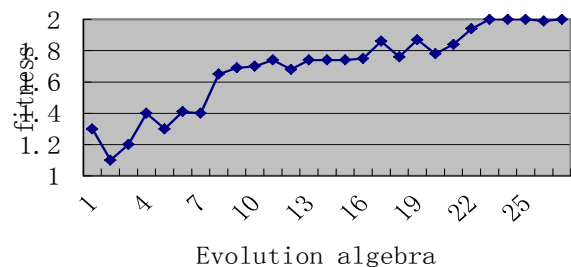
U	a	b	c	d	class
1	1	1	1		1
2	2	2	1	2	2
3	3	3	3	3	3
4	4	4	3	4	4
5	5	5	5	5	5
6	6	5	5	6	6
7	7	7	7	7	7
8	7	7	6	7	8
9	8	8	8	8	9
10	9	9	10	10	10
11	11	11	11	11	11
12	12	12	11	12	12
13	13	13	13	13	13
14	14	14	13	14	14
15	15	15	15	15	15
16	16	15	15	16	16
17	17	17	17	17	17
18	17	18	18	17	18

Execute attribute reduction of the decisive information as shown in Tab.1 with the algorithm put forward herein, the

final chromosome codes of the algorithm population all are 0011. According to the fitness function  $f(x) = card(c) - count(x)$ , the corresponding fitness value is 2. The attributes in correspondence with the code contents are  $C$  and  $d$ , and their corresponding attribute reduction is  $\{c, d\}$ . Figs. 3 and 4 show the comparison between the reduction process of the contents of Tab. 1 with the algorithm in this paper and that with the algorithm in Literature [8]. For the reduction process, it is predefined that the number of the initial population is 30, the crossover probability is  $p_r = 0.7$ , the mutation probability is  $p_m = 0.05$ , the attribute significance parameter  $\beta$  is 0.1, and the termination condition is that the fitness function values of several consecutive generations of population don't change.



igs. 3 Genetic algorithm based on rough set



Figs. 4 Genetic algorithm of Literature [8]

Through comparison with the algorithm of Literature [8], correct fault attribute reduction can be obtained with the algorithm of this paper, which verifies the perfrmability of the genetic algorithm of this paper. For a same fault data decisive table, the fitness function of the genetic algorithm based on rough set put forward herein roughly remains unchanged after the 19<sup>th</sup> generation, while the fitness function of the genetic algorithm of Literature [10] tends to be stable at least after the 25<sup>th</sup> generation. It is also verified that the algorithm of this paper can greatly reduce the iterations of the algorithm itself, enhance the convergence of genetic algorithm, and quicken the reduction speed.

Since the concept of attribute significance is introduced into the algorithm of this paper as the standard for selecting the initial population, and the core attributes of which the significance level is high are selected from it to form a specific individual population, then several data sets of which

the core attributes largely differ are selected from the UCI data set to verify that the efficiency of mass data processing of the algorithm of the paper is relatively high. It is predefined that the number of the initial population is 30, the crossover probability is  $P_r = 0.9$ , the mutation probability is  $p_m = 0.05$ , and the termination condition is that the fitness function values of several consecutive generations of population don't change. The comparison between the output result of the algorithm of this paper and that of the algorithm of Literature [11] and Literature [12] is as shown in Tab. 2.

Tab. 2 The experimental results compared

The number of instances		101	335	1484
Number of attributes		18	17	9
The core attributes number		1	6	4
Literature [11] algorithm	The number of iterations	37	49	93
	Time (s)	21.581	214.857	401.73
Literature [12] algorithm	The number of iterations	25	31	36
	Time (s)	5.356	16.465	36.802
The algorithm of this paper	The number of iterations	16	20	24
	Time (s)	2.342	12.073	25.607

According to the comparison experiment as above, the larger the amount of data information is, the longer the time taken to execute the traditional genetic algorithm is. Both the execution time and iterations of the algorithm of this paper are reduced. It is thus clear that the improved algorithm can greatly raise the efficiency of attribute reduction without reduction in the global search ability of the original algorithm. For the genetic algorithm based on rough set, core attribute is introduced, so that the initial population is set to be close to the final reduction result, the search range is reduced, the local search ability is enhanced, and the convergence speed is quickened while the attribute reduction accuracy is guaranteed.

VI. CONCLUSION

In this paper, the genetic algorithm based on rough set is detailed. The concept of attribute significance is put forward based on the concept of rough approximation accuracy as the standard for selecting attribute information, and introduced into the algorithm put forward in this paper, to determine the core attribute of attribute reduction. It is proposed to determine the initial population of genetic algorithm according to characterization attributes information, so as to raise the execution efficiency of the algorithm, and achieve a good convergence. A simple fitness function is designed. The fitness function equals to the difference between the total number of attributes and the number of the attributes obtained via reduction. In this way, the calculation is simplified. An attribute signification modification strategy based on rough approximation accuracy is adopted, so that satisfactory solutions can be obtained within the local space with the algorithm, the reduction result contains fewer attributes and remains the classification ability as same as that of the original data, search within the specified feasible solution space is guaranteed, and the convergence is guaranteed. At last, example analyses are made to verify that the genetic algorithm based on rough set is effective in solving attribute reduction.

REFERENCES

[1] GUO Qing-lin, ZHENG Ling. A Novel Approach for Fault Diagnosis of Steam Turbine Unit Based on Fuzzy Rough Set Data Mining Theory[J]. Proceedings of the CSEE, 2007,27(8):81-87.

[2] WANG Ying-ying, LUO Yi, TU Guang-yu. Fault Diagnosis Method for Distribution Networks Based on the Rough Sets and Decision Tree Theory[J]. High Voltage Engineering, 2008,34(4):794-798.

[3] Li Quan, Zhou Xing she. Spacecraft Fault Diagnosis Technology Based on Measurement and Control Data Mining[J]. Computer Measurement & Control, 2011,19(3):500-502.

[4] ZHENG Rao, WANG KuiSheng, ZHANG BeiKe, WANG Yun. A fault diagnosis method for complex processes based on Regflow graph and data mining[J]. Journal of Beijing University of Chemical Technology (Natural Science), 2013,40(5):104-109.

[5] WANG Guo-Yin, YU Hong, YANG Da-Chun. Decision Table Reduction based on Conditional Information Entropy[J]. CHINESE JOURNAL OF COMPUTERS, 2002,25(7):759-766.

[6] MIAO Duo-Qian, HU Gui-Rong. A HEURISTIC ALGORITHM FOR REDUCTION OF KNOWLEDGE[J]. JOURNAL OF COMPUTER RESEARCH & DEVELOPMENT, 1999,36(6):681-684.

[7] YIN Lin-Zi, YANG Chun-Hua, WANG Xiao-Li, GUI Wei-Hua. An Incremental Algorithm for Attribute Reduction Based on Labeled Discernibility Matrix[J]. ACTA AUTOMATICA SINICA, 2014,40(3):397-403.

[8] Dai Jianhua, Li Yuanxiang. Heuristic Genetic Algorithm for Reduction of Attributes in Rough Set Theory[J]. JOURNAL OF XI'AN JIAOTONG UNIVERSITY, 2002,36(12):186-190.

[9] Ye Dongyi, Huang Cui-wei, Zhao Bin. A Greedy Algorithm for Attribute Reduction in Rough Set[J]. Systems Engineering and Electronics, 2000,22(9):63-65.

[10] YE Dong-yi, CHEN Zhao-jiong. An improved reduction algorithm of attribute in rough set[J]. Journal of Fuzhou University(Natural Science), 2000,28(5):9-12.

[11] REN Yong-gong, WANG Yang, YAN De-qin. Rough Set Attribute Reduction Algorithm Based on GA. MICRO SYSTEMS, 2006,27(5):862-865.

[12] LI Wei-sheng, YI Zhe. Rough Set Attribute Reduction Algorithm Based on GA[J]. MICROELECTRONICS & COMPUTER, 2010,27:71-74.



First Author Name: YU Bo-wen

Publications :

Communication Module of FC-AE-1553 Interface  
Improved Apriori Algorithm Based on Matrix Reduction