An Attribute Reduction Algorithm Based on Genetic Algorithm and Discernibility Matrix

Wu Zhengjiang, Zhang Jingmin and Gao Yan
School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: {wuzhengjiang, gaoyan}@hpu.edu.cn

***Abstract*—In order to effectively solve the problem between genetic algorithm convergence and a local optimal solution, this paper presents an attribute reduction algorithm based on genetic algorithm with improved selection operator and discernibility matrix. In the algorithm, from the point of view of granular computing, rough set decision tables based on partition and covering are researched by measuring granularity again. The practical results show that the average convergence generation of modified algorithm is obviously superior to not modified algorithm, which is generally applicable in rough set decision tables based on partition and covering**

***Index Terms*—rough set, genetic algorithm, discernibility matrix, selection operator, attribute reduction**

I. INTRODUCTION

Rough set theory proposed by Pawlak in 1982 is a mathematical theory that deals with imprecise and uncertain information [1]. It provides a systematic approach for classification of objects through an indiscernability relation. Many examples of applications of the rough set method to pattern recognition, expert system, medical diagnosis, environmental science, biochemistry, chemistry psychology, conflict analysis, economics, process control, and elsewhere can be found in [2-7]. The further investigation into rough set theory and its extension will find new applications and new theories [8].

The classical rough set theory is based on equivalence relations, but this requirement is too restrictive for many applications. Therefore, rough set theory has been extended to similarity relation [9], tolerance relation [10], arbitrary binary relation [11, 12-15], covering [16-17] and others [18-20].

The notion of a covering generalized rough set is regarded as a meaningful extension of the traditional rough set model to deal with more complex practical problems. The literature [16-17, 21-32,46] has already provided several models of covering based rough sets. Multiple fuzzy rough set models based on coverings have been established by some researchers [18-20].

In rough set theory, the attribute reduction is one of the most important research contents. Most of attribute reduction algorithm base on discernibility matrix developed by Skowron and Rauszer [33-35], Both the rows and columns of the matrix correspond to the objects. An element of the matrix is the set of all attributes that distinguish the corresponding object pairs, namely, the set consists of all attributes on which the corresponding two objects have distinct values. One can construct a Boolean discernibility function from a discernibility relation, with attributes as Boolean variables. Skowron and Rauszer [33] showed that the set of attribute reductions are in fact the set of prime implicants of the reduced disjunctive form of the discernibility function. This provides a logic foundation for the study of reducts. On this basis, many researchers studied reduction construction by using the discernibility information in the discernibility matrix [25], [36], [8] and [10]. The attribute reduction of covering rough set, one of rough set extension models, can also adopt discernibility matrix, such as Tsang[4] etc proposed a covering attribute reduction algorithm on basis of discernibility matrix.

The prerequisite for reduction that is called the weak reduction in this paper is to guarantee invariant classification capacity of knowledge base. In practice, we often hope to can ensure this prerequisite, K-nearest neighbor (KNN) is an effective and powerful lazy learning algorithm, notwithstanding its easy-to-implement. KNN is to find the nearest k neighbors for a query point from a given data set and is widely used in pattern classification applications [37, 38]. Therefore, we can evaluate the quality of weak reduction through the accuracy in KNN model.

In knowledge discovery, genetic algorithms have been used for classification, model selection, and other optimization tasks. Practically, genetic algorithms provides a general-purpose search methodology, which uses principles of the natural evolution [39]. It can effectively solve the problem of large scale, complicated structure, and low computational efficiency. Therefore, it is suitable to solve this kind of questions like NP-hard reduction. Genetic algorithm is firstly used for solving the rough set attribute reduction by Wroblewski in literature [40], after, many scholars have studied one after another[47]. So this paper proposes an attribute reduction algorithm based on genetic algorithm and discernibility matrix about rough set decision tables based on partition and covering.

However, sometimes the algorithm above cannot converge within 200 generations, so cannot get the global optimal solution. We know that behavior and performance of genetic algorithms are directly affected

doi:10.4304/jsw.7.11.2640-2648

by the interaction between their parameters [41], which have fixed values in the simple genetic algorithm [42]. Therefore poor parameter settings usually lead to several problems such as premature convergence. So we can consider the three basic genetic operator (selection operator, crossover operator and mutation operator) to solve this problem, How to choose these parameters will directly affect the algorithm performance and search speed. This paper used improved selection operator increases the probability of the algorithm convergence in the global optimal solution, and thus don't easily fall into the local optimum. Following the modified instruction the average convergence algebra is improved.

The rest of the paper is organized as follows. Section 2 reviews some basic concepts of the Pawlak's rough set and genetic algorithm. Section 3 describes the attribute reduction algorithm based on genetic algorithm and discernibility matrix. The improved algorithm is described in Section 4. Section 5 presents the examples and analyses of the algorithms, the relationship and performance analysis of the two algorithms are studied and tabulated in each section. Section 6 concludes the paper.

II. PRELIMINARY

In this section , we briefly introduce the basic ideas of rough sets, genetic algorithm and K-nearest neighbor algorithm.

A. Rough Set

In rough set theory, the data is organized in a table called decision table. Rows of the decision table correspond to objects, and columns correspond to attributes. In the data set, a class label indicates the class to which each row belongs. The class label is called as decision attribute, the rest of the attributes are the condition attributes. Let (U, A, F) is a knowledge expression system. Here, $A = C \bigcup D$, C is used to denote the condition attributes, D for decision attributes, where $C \bigcap D = \varnothing$. The knowledge expression system possessed the condition attributes and the decision attribute is called the decision table.

Pawlak defines rough set as following definition 2.1.

Definition 2.1.[43] Let U be a finite set and R be an equivalence relation on U. R generates a partition $U/R = \{Y_1, Y_2, \cdots, Y_m\}$ where $Y_1, Y_2, \cdots, Y_m$ are the equivalence classes generated by the equivalence relation R. In the rough set theory, these are also called elementary sets of R. Let $\varnothing$ denote the empty sets. For any $X \subseteq U$, we can describe X by the elementary sets of R and the two sets.

$$R_*(X) = \cup\{Y_i \in U/R \mid Y_i \subseteq X\}$$

$$R^*(X) = \cup\{Y_i \in U/R \mid Y_i \cap X \neq \varnothing\}$$

are called the lower and the upper approximation of X, respectively.

Definition 2.2.[43] An information system is a triplet (U, A, F) ,where $U = \{x_1, x_2, \cdots, x_n\}$ is a nonempty finite set of objects x_i , $i = 1, 2, ..., n$ called the universe of discourse, $A = \{a_1, a_2, \cdots, a_m\}$ is a nonempty finite set of attributes a_j , $j = 1, 2, ..., m$; $F = \{f_j \mid j = 1, 2, ..., m\}$ is a set of information functions such that $f_j(x_i) \in V_j$ for all $x_i \in U$, where V_j is the domain of attribute a_j .

Theorem 2.1. [44] Let (U, A, F) be an information system and $B \subseteq A$. The indiscernibility relation is defined as

$$R_B = \{(x_i, x_j) \mid x_i \in U, x_j \in U, f_l(x_i) = f_l(x_j), \forall a_l \in B\}.$$

Then, R_B is an equivalence relation, and produces a partition on $U : U/R_B = \{[x_i]_B \mid x_i \in U\}$, in which,

$$[x_i]_B = \{x_j \mid (x_i, x_j) \in R_B\}.$$
$$= \{x_j \mid \forall a_l \in B, f_l(x_i) = f_l(x_j)\}$$

The classical rough set analysis depends on the indiscernibility relation that describes indistinguishability of objects. Indiscernibility relations are equivalences that are interpreted so that two objects are equivalent if one cannot distinguish them by using existing information.

Definition 2.3.[44] Let (U, A, F) be an information system and $D \subseteq A$ .D is referred to as a consistent set of (U, A, F) if $R_D = R_A$. If D is a consistent set and $R_{D-\{d\}} \neq R_A$ for all $d \in D$, then D is referred to as a reduction of (U, A, F) . The set of all reductions of an information table *S* is denoted as *RED*(*S*).

Since $R_D = R_A \Leftrightarrow U/R_D = U/R_A$, the attribute reduction of an information system means it can preserve the partition (classification) of the object set *U*.

Definition 2.4. Let (U, A, F) be an information system and $D \subseteq A$. if $R_D = R_A$, the D is referred to as a weak reduction of (U, A, F) .

Compared with reduction, the weak reduction defined in this paper doesn't consider the minimum condition attribute but considers only two object that have the same classification ability.

Definition 2.5. [16] Let U be a finite universe of discourse, C a family of subsets of U. If none of the subsets in C is empty and $\cup C = U$, C is called a covering of U.

It is clear that a partition of U is certainly a covering of U. and the concept of a covering is an extension of the concept of a partition.

Definition 2.6.[33] Discernibility matrix. Two objects are discernible if their values are different in at least one attribute. Skowron and Rauszer suggested a matrix representation for storing the sets of attributes that discern pairs of objects, called a discernibility matrix.

Given an information table S, its discernibility matrix $M=(M(x,y))$ is a $|U|\times|U|$ matrix, in which the element $M(x,y)$ for an object pair (x,y) is defined by $M(x,y)=\{a\in A|\ f_a(x)\neq f_a(y)\}$

The physical meaning of the matrix element $M(x,y)$ is that objects x and y can be distinguished by any attribute in $M(x,y)$. The pair (x,y) can be discerned if $M(x,y)\neq\varnothing$. A discernibility matrix M is symmetric, i.e., $M(x,y)=M(y,x)$, and $M(x,y)\neq\varnothing$. Therefore, it is sufficient to consider only the lower triangle or the upper triangle of the matrix.

An attribute set $B\subseteq A$ can discern an object pair (x,y) if $B\cap M(x,y)\neq\varnothing$.

Definition 2.7. [33] (Discernibility function) The discernibility function of a discernibility matrix is defined by

$$f(M)=\wedge\{\vee(M(x,y))\,|\,\forall x,y\in U, M(x,y)\neq\varnothing\}$$

The expression $\vee(M(x,y))$ is the disjunction of all attributes in $M(x,y)$, indicating that the object pair (x,y) can be distinguished by any attribute in $M(x,y)$ The expression $\wedge\{\vee(M(x,y))\}$ is the conjunction of all $\vee(M(x,y))$, indicating that the family of discernible object pairs can be distinguished by a set of attributes satisfying $\wedge\{\vee(M(x,y))\}$.

The discernibility function can be used to state an important result regarding the set of reductions of an information table, as shown by the following theorem from Skowron and Rauszer.

Theorem 2.2.[33] The reduction set problem is equivalent to the problem of transforming the discernibility function to a reduced disjunctive form. Each conjunctor of the reduced disjunctive form is called a prime implicant. Given the discernibility matrix *M* of an information table *S*, an attribute set $R=\{a_1,a_2,\cdots a_p\}$ is a reduction if and only if the conjunction of all attributes in *R*, denoted as $a_1\wedge a_2\cdots\wedge a_p$, is a prime implicant of *f(M)*.

In order to derive the reduced disjunctive form, the discernibility function *f(M)* is transformed by using the absorption and distribution laws. Accordingly, finding the set of reductions can be modelled based on the manipulation of a Boolean function. The set *RED(S)* of reductions of an information table is equivalent to the set of prime implicants of the discernibility function.

Based on the results of Theorem 2.2, Skowron and Rauszer also suggested an alternative characterization of a reduction in terms of the discernibility matrix as shown by the next theorem.

Theorem 2.3 [33] Given the discernibility matrix M of an information table S, an attribute set R is a reduction if and only if

(i). For $\forall x,y\in U\times U$, such that

$$M(x,y)\neq\varnothing\Rightarrow R\cap M(x,y)\neq\varnothing.$$

(ii). For $\forall a\in R$, there exists $(x,y)\in U\times U$, such that

$$M(x,y)\neq\varnothing\wedge((R-\{a\})\cap M(x,y)=\varnothing.$$

Property (i) shows that R is jointly sufficient for distinguishing all discernible object pairs. In fact, the set of attributes formed by the union of all elements of the discernibility matrix satisfies property (i). Property (ii) shows that each attribute in R is individually necessary. The result of Theorem 2 provides a convenient way to test if a subset of attributes is a reduction. However, it does not directly offer a method to compute a reduction. Many authors have proposed and studied various algorithms to construct a reduction based on the discernibility matrix[45].

B. Genetic Algorithm

Genetic algorithm provides a general-purpose search methodology, which uses principles inspired by natural genetics to evolve solutions to problems [39] . The simple genetic algorithm starts off with a population of randomly generated chromosomes, each representing a candidate solution to the concrete problem being solved, and advances towards better chromosomes by applying genetic operators, which correspond to those occurring in nature. This population evolves over time through a successive iteration process of competition and controlled variation. Each state of population is called generation. Associated with each chromosome at every generation is a fitness value, which indicates the quality of the solution and the chromosome values lead to. Based upon these fitness values, the selection of the chromosomes, which form the new generation, takes place. Like in nature, the new chromosomes are created using genetic operators such as crossover and mutation.

The traditional genetic algorithm generally includes selection operator, crossover operator, mutation operator and optimal preservation strategy.

The one-point crossover operator is used in this paper. The concrete implementation process is as follows:

a) Randomly match by pairs for population.

b) Randomly choose an intersection point for the paired individuals.

c) Exchange two individuals part chromosomes at the set crossover probability in the intersection point for the paired individuals.

We take a basic bit mutation operator. Firstly, each gene locus of each individual is specified for an aberrance point at the set mutation probability, and then the value of each set aberrance point is taken inversion operation, so as to form one new individual. The elitist strategy is that after getting the new individual in contemporary generation, if the fitness value of the worst individual (the minimum fitness value) is less than the fitness value of

the best individual (the maximum fitness value) of last generation, then the best individual of last generation replaces the worst individual of contemporary generation.

So far, genetic algorithm have demonstrated impressive results in a wide range of domains. However, it has also been found that the genetic algorithm behavior is strongly determined by the interaction between their parameters, which have fixed values in the simple genetic algorithm. Therefore poor parameter settings usually lead to several problems such as premature convergence.

Genetic algorithm generally takes the roulette wheel scheme to select. Although it is widely performed, it still exist two problems. One is that in the early evolution, there may be a fitness high individual is selected at a high probability, thus reproduces many offspring, so it is easy to appear the local optimal situation because a single individual be unable to continue to evolve; The other is that in the later evolution, when the gap of each individual fitness is not big, this method is no longer has the ability to select, so doesn't reflect the superiority of individuals.

Hence, this paper improves selection operator. Firstly, select the individual directly into mating pool that the fitness value is 1 from the initial population, and then take the roulette wheel scheme. By taking this improved selection operator, the individual number of the better fitness value is increased, thus the above problems are solved to some extent.

C. K-nearest Neighbor Algorithm

The KNN method [37] is widely used in pattern recognition. This method is useful both for estimation of densities and for classification. As a classifier, this algorithm involves the following three steps: (1) calculating the distances between a query sample and all training samples, (2) choosing the k nearest training samples to the query sample, and (3) assigning a class label by applying the majority rule to the k nearest samples.

The process of KNN algorithm is as follows: given a test document x, find the K nearest neighbors of x among all the training documents, and score the category candidates based the category of K neighbors. The similarity of *x* and each neighbor document is the score of the category of the neighbor document. If several of the K nearest neighbor documents belong to the same category, then the sum of the score of that category is the similarity score of the category in regard to the test document *x*. By sorting the scores of the candidate categories, system assigns the candidate category with the highest score to the test document *x*. The decision rule of KNN can be written as:

$$f(x) = \arg\max_j Score(x, C_j) = \sum_{d_i \in KNN} sim(x, d_i) y(d_i, C_j)$$

where $f(x)$ is the label assigned to the test document x; $Score(x, C_j)$ is the score of the candidate category C_j with respect to x; $sim(x, d_i)$ is the similarity between x and the training document d_i; $y(d_i, C_j) \in \{0,1\}$ is the binary category value of the training document d_i with respect to C_j ($y = 1$ indicates document d_i is part of category C_j, or $y = 0$).

This approach is effective, non parametric and easy to implement. However, its classification time is long and the accuracy is severely degraded by the presence of noisy training document.

III. An Attribute Reduction Algorithm Based on Genetic Algorithm and Discernibility Matrix

A. The Coding of Chromosome

Considering the practical characteristic of attribute reduction, the binary encoding is introduced. And the coding strategy is as follows: let $C = \{c_1, c_2, \cdots c_n\}$ is the condition attributes set, the condition attributes space can easily map the chromosome of genetic algorithm. Chromosome is the binary string that the length is n, each bit of which corresponds a condition attribute. If the bit is 1, which means select the corresponding condition attribute, and otherwise means not select the corresponding condition attribute. For example, the code 0001100100001 shows the condition attributes of this reduction have 13 bits and select the corresponding the fourth, the fifth, the eighth and the thirteenth condition attributes. Since binary encoding has the feature of simple and convenient, in the whole, genetic algorithm in easy to operate.

B. The Fitness Function

In the genetic algorithm, the fine level of the optimal solution can be achieved by making use of the fitness function which is to measure each individual' s optimization calculation in the groups. In this paper, the fitness function is defined as follows:

$$F(x) = C_n / | f | \quad (1)$$

Where $| f |$ indicates the number of terms of discernibility function. C_n indicates the number of terms of discernibility function that is covered by the chromosome *n*.

C. Selection Operator

Take the roulette wheel scheme to select, we copy the excellent individual selected from the contemporary population to the next generation population. The concrete implementation process is as follows:

a) Calculating the sum of all individuals fitness;

b) Calculating relative fitness of all individuals by type (2), that is to say, the probability of each individual inherits in the next generation

$$C(x_i) = F(x_i) / \sum_{i=1}^{n} F(x_i), i = 1, 2, \cdots n \quad (2)$$

c) Taking the roulette wheel scheme to ascertain the times of each individuals selected.

D. Algorithm Description

According to the analyses above, an attribute reduction algorithm based on genetic algorithm and discernibility matrix is described as follows:

The weak reduction is called advantage weak reduction which more than the original accuracy of k-nearest neighbor. The weak reduction is called reliable weak reduction which greater than or equal the original accuracy of k-nearest neighbor.

Algorithm 1(also shown in figure 1):

Input: a decision table $S = (U, A, V, f), A = C \bigcup D$, where C is the condition attribute, D is the decision attribute

Output: one or more advantage weak reduction

Step 1: Calculating the number of terms of discernibility function $| f |$.

Step 2: Initializing the population: randomly generate m numbers initial population that length is n

Step 3: According to the type (1) to calculate the fitness value of each individual, if the fitness values of all individuals are 1, turn to step 9

Step 4: By type (2) to calculate the probability of each individual selected, and then taking the roulette wheel scheme.

Step 5: By the crossover probability p_c to proceed the one-point crossover operator

Step 6: By the mutation probability p_m to proceed the basic bit mutation operator

Step 7: By type (1) to calculate the fitness values of new individuals; and then taking the optimal preservation strategy

Step 8: If all individuals are 1, turn to step 9, otherwise, turn to step 3

Step 9: Judging whether the optimal individual's accuracy of k-nearest neighbor is more than the original accuracy of k-nearest neighbor, if more than, then input , otherwise ,turn to step 2

IV. An Attribute Reduction Algorithm Based on Genetic Algorithm with Improved Selection Operator and Discernibility Matrix

A. Improved Selection Operator

The process of improved selection operator is as follows:

a) Selecting the individual directly into the mating pool that the fitness value is 1 from initial population.

b) Taking the roulette wheel scheme for all individuals

B. Algorithm Description

According to the analyses above, an attribute reduction algorithm based on genetic algorithm with improved selection operator and discernibility matrix is described as follows:

Algorithm 2:

Input: a decision table $S = (U, A, V, f), A = C \bigcup D$, where C is the condition attribute, D is the decision attribute

Output: one or more advantage weak reduction

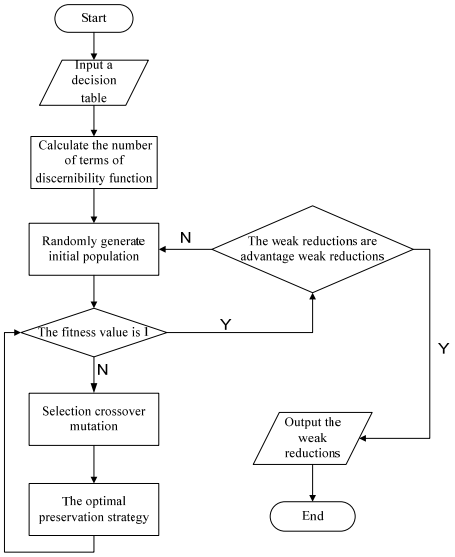

Figure 1. Algorithm Flowchart

Step 1: Calculating the number of terms of discernibility function $| f |$.

Step 2: Initializing the population: randomly generate m numbers initial population that length is n

Step 3: According to the type (1) to calculate the fitness value of each individual, if the fitness values of all individuals are 1, turn to step 9

Step 4: Taking the improved selection operator.

Step 5: By the crossover probability p_c to proceed the one-point crossover operator

Step 6: By the mutation probability p_m to proceed the basic bit mutation operator

Step 7: By type (1) to calculate the fitness values of new individuals; and then taking the optimal preservation strategy

Step 8: If all individuals are 1, turn to step 9, otherwise, turn to step 3

Step 9: Judging whether the optimal individual's accuracy of k-nearest neighbor is more than the original accuracy of k-nearest neighbor, if more than, then input ,otherwise ,turn to step 2.

C. The Analysis of Algorithm Complexity

Time complexity of the algorithm consists of three parts. The first part is to calculate the time complexity of discernibility function by discernibility matrix $O(|C||U^2|)$; The second part is to calculate the time complexity of the genetic algorithm $O(mn|f|+|C|)$, where the m is the population numbers, the n is the termination generations of genetic algorithm; The third part is to calculate the time complexity of more than the original accuracy of k-nearest neighbor $O(m|C||U^2|)$. So the total time complexity is:

$$O(|C||U^2|+mn|f|+|C|+m|C||U^2|)$$
$$= max(O(mn|f|), O(m|C||U^2|))$$

In algorithm 2, step 4 is the only difference from algorithm 1. So, in theory, algorithm complexities of algorithm 1 and algorithm 2 are the same, but in fact, the average running time of algorithm 2 is about 48% faster than algorithm 1. It is the effect of the improved selection in section 4.1.

V. Algorithm Examples and Analysis

In this section, we present the examples and analyses of the algorithm 1 and algorithm 2, the experimental result of the two algorithms are studied and tabulated.

We use the wine data set (13 condition attributes and a decision attribute, 178 records) in UCI machine learning database to test. In order to reduce the influence on outcomes for the inconformity of each attribute dimension, we standardize the value of all continuous attributes to [0, 1] interval with the biggest-minimum method. The rough set based on partition is measured with the condition attributes divided by particle size, the particle size selected is 0.21, if two condition attributes divided by 0.21 after rounding take the same, and then we think them belong to the same category, otherwise not belong to the same category. The rough set based on covering is measured through the particle size to change its neighborhood size of covering element, coverings of these neighborhoods meet a properties compared with generally covering: for any neighborhood, there always exist some elements that cannot be covered by other neighborhoods in addition to itself. The intervals selected are [0,0.22),[0.18,0.42),[0.38,0.62),[0.58,0.82),[0.78,1]. If two condition attributes fall in the same interval, and then we think them belong to the same category, otherwise not belong to the same category.

All in the experiments ,we take the test set TE and training set TR which are completely disjoint sets, and randomly set the proportion of TR and TE is 2:1, the experimental parameters are as follows: initial population m is 10, the biggest algebra n sets 200, the crossover probability p_c is 0.7, the mutation probability p_m is 0.01.

A. The Experimental Result of the Algorithm 1

In this subsection, we in detail state the experimental result of the algorithm 1. In order to more easy to analysis , the experimental result is made line charts and table .

In order to see the effect more visualized, algorithm 1 is run 100 times repeatedly to make line charts, As figure 2 and figure 3 show.

The advantages of the advantage weak reduction can be saw clearly and visually from figure 2 and figure 3. There are two cases that can't get the advantage weak reduction in the experiments. Case 1: when the original accuracy of k-nearest neighbor is 100%, by the definition of the advantage weak reduction, we know that this kind of situation has no advantage weak reduction; Case 2: algorithm can't converge within 200 generations, so we can't find the global optimal solution. It is because of these two cases that figure 2 and figure 3 have no 100 times records.

In order to more specific expression result of algorithm 1, we draw once experimental result from figure 2 and figure 3, such as table 1 and table 2 shows:

The effectiveness of algorithm 1 can be shown from table 1, table 2, figure 2and figure 3.

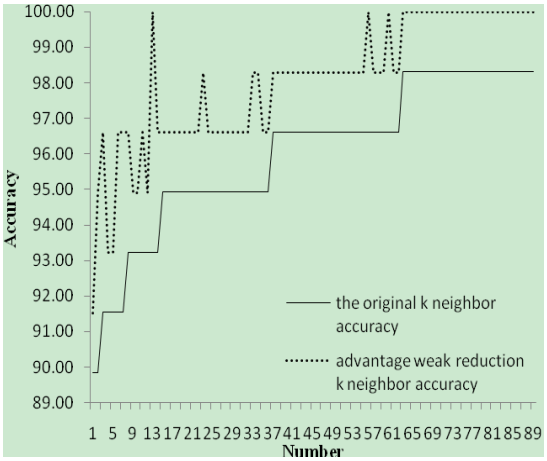

Figure 2. Algorithm flowchart. Accuracy Contrast Diagram of the Rough Set based on Partition

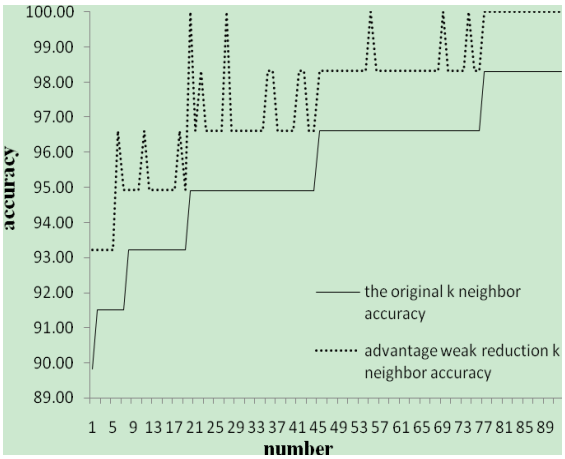

Figure 3. Algorithm flowchart. Accuracy Contrast Diagram of the Rough Set based on covering

TABLE I.
THE ROUGH SET BASED PARTITION

the original accuracy of k-nearest neighbor	weak reduction	the advantage weak reduction and its accuracy	convergence algebra
96.61%	1111011101010 1011111100010 1111011101000 1010010011100 1011101111111 1011001111111	1011001111111 1011101111111 98.31%	7

TABLE II.
THE ROUGH SET BASED COVERING

the original accuracy of k-nearest neighbor	weak reduction	the advantage weak reduction and its accuracy	convergence algebra
94.92%	0011111111100 0111111111100 1101001101100 0010101101101 1101001110111 1101111010100 1101001101101	1101001101101 98.31% 1101001110111 100%	4

B. The Experimental Results Contrast of Algorithm 1 and 2

In this subsection, we in detail state the experimental result contrast of algorithm 1 and 2. In order to more easy to analysis and contrast, the experimental result is made line charts and table .

In order to contrast the convergence of algorithm 1 and algorithm 2 based on partition and covering, we make 100 times contrast experiments between algorithm 1 and 2 with the same initial individual, As figure 4 and figure 5 show.

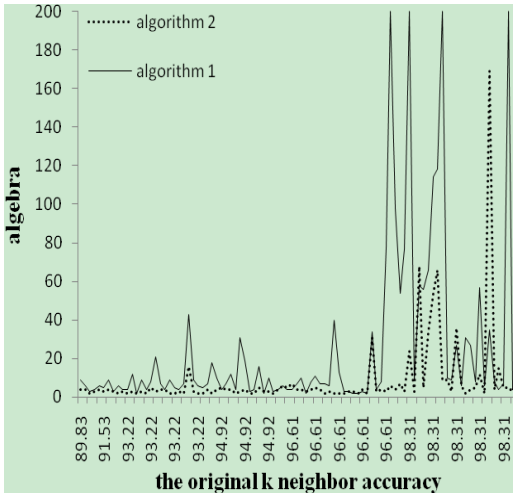

Figure 4. Convergence Algebra Contrast Diagram of the Rough Set Based on Partition

The cause of figure 4 and 5 no 100 times records is the same as figure 2 and 3. In figure4, the average convergence algebra of algorithm 1 is 25, and the average convergence algebra of algorithm 2 is 9, algorithm 2 is faster 16 generations than algorithm 1 on an average, that is to say, the generation of algorithm 1 is 2.8 times faster than algorithm 2. Similarly, in figure5, the average convergence algebra of algorithm 1 is 25, and the average convergence algebra of algorithm 2 is 10, algorithm 2 is faster 15 generations than algorithm 1 on an average, that is to say the generation of algorithm 1 is 2.5 times faster than algorithm 2. So we can see the convergence of the algorithm 2 is better than algorithm 1. In order to more specific reflect the advantage of algorithm 2 in the convergence, we draw once experiment result from the figure 4 and figure 5, such as table 3 and table 4 shows

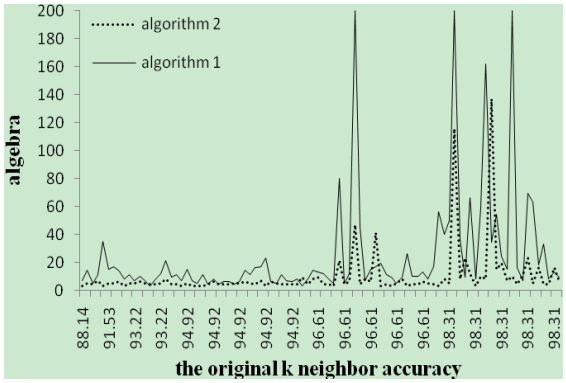

Figure 5. Convergence Algebra Contrast Diagram of the Rough Set Based on Covering

TABLE III.
ROUGH SET BASED ON PARTITION

	the original accuracy of k-nearest neighbor	weak reduction	convergence algebra	the advantage weak reduction and its accuracy
Alg. 1	98.31%	1111011001100 1111011011100 1110011011100	200	null
Alg. 2	98.31%	1110010011101 0011111111101 1011111111101 0011001011111 0011110110101 0011011110100 1010101111101	24	100% 1010101111101

TABLE IV.
ROUGH SET BASED ON COVERING

	the original accuracy of k-nearest neighbor	weak reduction	convergence algebra	the advantage weak reduction and its accuracy
Alg. 1	98.31%	1101010110011 1101010110010 1111110111111 1111000110011 1101000110011 1101110110011 1101110111011	200	null
Alg. 2	98.31%	1100010111101 1011001011101 1100100111101 0111000111101 1100011111101 1100001111001 0111000110101	10	100.0% 1011001011101

From the table 3 and table 4, we can see algorithm 1 can't converge within 200 generations in the same initial population, however, algorithm 2 can quickly converge in

optimal solution, at this time, the advantage of algorithm 2 is more apparent in the convergence.

The discussion above is the advantage weak reduction, the following simply discuss about the reliable weak reduction according to the analysis method of advantage weak reduction. For the rough set based on partition, the average convergence algebra of algorithm 1 is 8, and the average convergence algebra of algorithm 2 is 4, algorithm 2 is faster 4 generations than algorithm 1 on an average, that is to say, the generation of algorithm 1 is 2 times faster than algorithm 2. For the rough set based on covering, the average convergence algebra of algorithm 1 is 5, and the average convergence algebra of algorithm 2 is 12, algorithm 2 is faster 7 generations than algorithm 1 on an average, that is to say, the generation of algorithm 1 is 2.4 times faster than algorithm 2. The time of getting the advantage of weak reduction is 5 or 6 times faster than the reliable of weak reduction, in practice, we can choose according to the decision table's size and needs.

VI. Conclusion

In this paper, we propose an attribute reduction algorithm based on genetic algorithm and discernibility matrix and improve it. The convergence speed of algorithm improved before is slow, and it is easy to fall into the local optimal solution. So we uses improved selection operator which increases the probability of converged on the global optimal solution to solve this problem. The experimental results show that the algorithm can faster converge on the global optimal solution compared with the algorithm improved before. We also contrast the proportion of advantage (reliable) weak reduction in weak reduction about rough set decision tables based on partition and covering, the experimental results show that the proportion don't improve compared with the algorithm improved before. Hence, in future articles, we will study how to improve the proportion.

Acknowledgment

This work was supported by grants from The National Science Foundation of China under Grant No. 61202286, 61175047; The Soft Science Research Program of Henan Province under Grant No. 122400450212 and The Innovation Foundation of Henan Polytechnic for the Master Thesis under Grant No: 2001-M-36.

References

[1] Pawlak Z, Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11 (5): 341-356.

[2] F. Angiulli, C. Pizzuti, Outlier mining in large high-dimensional data sets, IEEE Transactions on Knowledge and Data Engineering 2005, 17 (2) 203–215.

[3] F. Hu, G.Y. Wang, H. Huang, a.l. et Incremental Attribute Reduction Based on Elementary Sets, RSFDGrC 2005, 3641, 2005, pp. 185–193.

[4] C.T. Su, J.H. Hsu, An extended chi2 algorithm for discretization of real value attributes, IEEE Transactions on Knowledge and Data Engineering 2005, 17 (3) 437–441.

[5] L. Polkowski, A. Skowron (Eds.), Rough Sets and Current Trends in Computing, vol. 1424, Springer, Berlin, 1998.

[6] L. Polkowski, A. Skowron (Eds.), Rough Sets in Knowledge Discovery, vol. 1, Physica-Verlag, Berlin, 1998.

[7] L. Polkowski, A. Skowron (Eds.), Rough Sets in Knowledge Discovery, vol. 2, Physica-Verlag, Berlin, 1998.

[8] Z. Pawlak, A. Skowron, Rough sets: some extensions, Inform. Sci. 2007, 177 (1) 28–40.

[9] R. Slowinski, D. Vanderpooten, A generalized definition of rough approximations based on similarity, IEEE Trans. Knowledge Data Eng. 2000, 12 (2) 331–336.

[10] A. Skowron, J. Stepaniuk, Tolerance approximation spaces, Fundamenta Informaticae. 1996, 27: 245–253.

[11] G.L. Liu, W. Zhu, The algebraic structures of generalized rough set theory, Inform. Sci. 2008, 178 (21) : 4105–4113.

[12] W. Zhu, Generalized rough sets based on relations, Inform. Sci. 2007. 177 (22): 4997–5011.

[13] Y.Y. Yao, On generalizing pawlak approximation operators, in: LNAI, vol. 1424, 1998, pp. 298–307.

[14] Y.Y. Yao, Relational interpretations of neighborhood operators and rough set approximation operators, Inform. Sci. 1998: 111 (1–4) 239–259.

[15] Y.Y. Yao, Constructive and algebraic methods of theory of rough sets, Inform. Sci. 1998: 109: 21–47.

[16] Z. Bonikowski, E. Bryniarski, U.W. Skardowska, Extensions and intentions in the rough set theory, Information Sciences 1998. 107: 149–167.

[17] E. Bryniarski, A calculus of rough sets of the first order, Bull. Pol. Acad. Sci. 1989, 36 (16) 71–77.

[18] T.J. Li, Y. Leung, W.X. Zhang, Generalized fuzzy rough approximation operators based on fuzzy coverings, International Journal of Approximation Reasoning 2008. 48: 836–856.

[19] T.J. Li, W.X. Zhang, Fuzzy rough approximations on two universes of discourse, Information Science 2008. 178:829–906.

[20] T. Deng, Y. Chen, W. Xu, Q. Dai, A noval approach to fuzzy rough sets based on a fuzzy covering, Information Science 2007. 177: 2308–2326.

[21] W. Zhu, Topological approaches to covering rough sets, Information Sciences 2007. 177: 1499–1508.

[22] W. Zhu, Relationship between generalized rough sets based on binary relation and coverings, Information Sciences 2008.

[23] Eric C.C. Tsang, D. Chen, D.S. Yeung, Approximations and reducts with covering generalized rough sets, Computers and Mathematics with Applications 2008. 56: 279–289.

[24] W.H. Xu, W.X. Zhang, Measuring roughness of generalized rough sets induced by a covering, Fuzzy Sets and Systems 2007. 158: 2443–2455.

[25] D. Chen, C. Wang, Q. Hu, A new approach to attributes reduction of consistent and inconsistent covering decision systems with covering rough sets, Information Sciences 2007. 177: 3500–3518.

[26] Z. Xu, Q. Wang, On the properties of covering rough sets model, Journal of Henan Normal University (Natural Science) 2005. 33 (1): 130–132.

[27] E. Tsang, D. Chen, J. Lee, D.S. Yeung, On the upper approximations of covering generalized rough sets, in: Proceedings of the 3rd International Conference on Machine Learning and Cybernetics, 2004, pp. 4200–4203.

[28] W. Zhu, F.Y. Wang, A new type of covering rough sets, in: IEEE IS 2006, London, 4–6 September, 2006, pp. 444–449.

[29] W. Zhu, F.Y. Wang, On three types of covering rough sets, IEEE Transactions on Knowledge and Data Engineering . 2007. 19 (8): 1131–1144.
[30] Q.H. Hu, D.R. Yu, J.F. Liu, C.X. Wu, Neighborhood rough set based heterogeneous feature subset selection, Information Sciences. 2008. 178: 3577–3594.
[31] Q.H. Hu, D.R. Yu, Z.X. Xie, Neighborhood classifiers, Expert Systems with Applications 2008. 34:866–876.
[32] Q.H. Hu, Z.X. Xie, D.R. Yu, Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, Pattern Recognition. 2007. 40: 3509–3521.
[33] A. Skowron, C. Rauszer. The discernibility matrices and functions in information systems, in: R. Slowiński (Ed.), Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory, Kluwer, 1992. Dordrecht .
[34] Yao Y Y, Zhao Y. Discernibility Matrix Simplification for Constructing Attribute Reducts. Information Sciences, 2009, 179: 867-882.
[35] Zhao Y, Yao Y Y, F.Luo, Data Analysis Based on Discernibility and Indiscernibility. Information Sciences, 2007, 177: 4959-4976.
[36] Y. Leung, W.Z. Wu, W.X. Zhang Knowledge acquisition in incomplete information systems: a rough set approach European Journal of Operational Research, 2006, 168, pp. 164–180
[37] T.M. Cover, P.E. Hart Nearest neighbor pattern classification IEEE Trans. Inform. Theory, 1967, 13: pp. 21–27
[38] K. Hattori, M. Takahashi A new nearest-neighbor rule in the pattern classification problem Pattern Recognition, 1999. 32 (3): 425–432
[39] J.H. Holland, Genetic algorithms, Sci. Am. 1992. 267 (1) 44–150.
[40] Wroblewski, J. Finding minimal reducts using genetic algorithms Proceedings of the Second Annual Joint Conference on Information Sciences. Wrightsville Beach, USA, 1995: 186-189.
[41] K. Deb, S. Agrawal, Understanding interactions among genetic algorithm parameters, in: W. Banzhaf, C. Reeves (Eds.) Foundations of Genetic Algorithm 5, Morgan Kaufmann, San Francisco, CA, 1998, pp. 265–286.
[42] Holland J H. Adaptation in Natural and Artificial Systems.Ann Arbor, Michigan: University of Michigan Press,1975,1-17.
[43] Z. Pawlak. Rough sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Dordrecht, Boston 1991
[44] W.X. Zhang, Y. Leung, W.Z. Wu. Information Systems and Knowledge Discovery. Science Publishing Company, Beijing . 2003. (in Chinese.)
[45] Z. Pawlak, A. Skowron Rough Sets and Boolean reasoning Information Sciences, 2007. 177: 41–73
[46] Tian Yang. Qingguo Li Reduction about approximation spaces of covering generalized rough sets, International Journal of Approximate Reasoning ,2010,335-345.
[47] Jianyong Chen, Changsheng Zhang. Efficient Clustering Method Based on Rough Set and Genetic Algorithm, Procedia Engineering 2011. 15 :1498 – 1503.

Wu Zhengjiang was born in 1981. He received the Ph.D. degree in 2009 from Southwest Jiaotong university. He is a lecturer at Henan Polytechnic University. His research interests include rough set theory, data mining, etc.

Zhang jingmin was born in 1987. She is a master in Henan Polytechnic University. Her interesting includes rough set theory and genetic algorithm.

Gao Yan. Profession He received the Ph.D. degree in 2012 from Southwest Jiaotong university. He is a lecturer at Henan Polytechnic University. His research interests include rough set theory, mining safety etc.