Approaches to Attributes Reduction Based on Ant Colony Optimization

Honghua Chen Zheng Pei Li Zhang

School of Mathematics & Computer Engineering, Xihua University, Chengdu 610039, China

Abstract

As we known, ants are blind and single ant's capacity is limited. But researchers have studied that ant colonies have egregious ability to search shorter path which loads between their nests and foods. Inspired from this and based on ant colony optimization(ACO), we propose a new approach about attributes reduction of rough sets. Using this approach, we can reduce attributes as maximum as possible, obtain many different results synchronously and reduce the scope of core attributes. Experiment shows that by this new approach, better results can be obtained.

Keywords: Rough sets, Attributes reduction, Ant colony optimization, Binary discernibility matrices

1. Introduction

The need for discovering knowledge from information increases with the forward rapid development of the recent civilization [1]-[3]. One of the most useful methods for knowledge discovering is the rough set theory(RST) which proposed by Z. Pawlak[4]. And attributes reduction is one of the most important subjects of RST. Due to attributes reduction is a NP-hard problem, most of approaches about attributes reduction are heuristic[6], [7]. As an approach of attributes reduction, binary discernibility matrices have many interesting properties[8]. In this paper, we propose a new approach about attributes reduction which based on binary discernibility matrices and ant colony optimization(ACO).

As we known, a wonderful self-organization behavior will usually be produced from the collective behavior of social animals. Take a colony of ants for example, biologists had studied the phenomenon carefully and found that ants cooperate to find shorter routing path by means of indirect communications using a kind of substance called "pheromone"[10]. Inspired from this, ant system(AS) was proposed by Italian researchers M. Dorigo, V. Maniezzo and A. Colorni[13]. This algorithm was proposed to solve the travelling salesman problem(TSP) firstly. These years, M. Dorigo *etc.* have expanded AS algorithm, and proposed a new generales optimization technique: ant colony optimization(ACO)[10]. All the algorithms which accord with the frame of ACO can be called as ACO algorithm[14]. Many scholars are attracted to study ACO and in the past ten years the algorithm has been widely applied to the fields of combinatorial optimization, network routing, data mining *etc.*, [15]-[18].

In the next Section, some preliminaries knowledge about attributes reduction and binary discernibility matrix is introduced. Section 3 is devoted to propose a new approach based on ACO and binary discernibility matrix. Experiment about the new approach is given in Section 4. Conclusion is Section 5.

2. Preliminaries knowledge

Let an information system be $S = \langle U, A, V, f\rangle$, where U is a non-empty finite set of objects, A is a non-empty finite set of attributes, $V = \bigcup_{a\in A} V_a$ and V_a is the domain of a, $f : U \times A \to V$ is information function. Let R be an equivalence relation on U. And a set of equivalence classes with respect to R is as follows:

$$U/R = \{[x_i]_R | x_i \in U\},$$

where $[x_i]_R = \{x_j | (x_i, x_j) \in R\}$. Each non-empty subset $B \subseteq A$ determines an indiscernibility relation as follows:

$$R_B = \{(x, y) \in U^2 | f(x, a_l) = f(y, a_l), \forall a_l \in B\},$$

then R_B is an equivalence relation on U and let $[x_i]_B = \{x_j | (x_i, x_j) \in R_B\}$.

Definition 1 *[5]* $\langle U, A, V, f\rangle$ *is an information system, for* $B \subseteq A$*, if* $R_A = R_B$*, then* B *is a partition consistent set. If* $R_A = R_B$*, and* $\nexists B' \subset B$*, s.t.,* $R_{B'} = R_A$*, then* B *is a partition reduct.*

Table 1: A general information system

$U \setminus A$	humidity (a_1)	rainy (a_2)	frosty (a_3)	sunny (a_4)	windy (a_5)
Monday(x_1)	Low	No	Yes	Shady	Low
Tuesday(x_2)	Average	Yes	No	Commonly	Average
Wednesday(x_3)	Average	No	No	Sunshine	Average
Thursday(x_4)	High	Yes	Yes	Shady	Average
Friday(x_5)	High	Yes	No	Shady	High

Definition 2 *$S = \langle U, A, V, f\rangle$ is an information system, $\overrightarrow{U} = \{(x_i, x_j) \in U^2 | \exists a \in A$ s.t., $f(x_i, a) \neq f(x_j, a)\}$. Let $\psi : \overrightarrow{U} \times A \to \{0, 1\}$. Define*

$$\psi((x_i, x_j), a_l) = \begin{cases} 1 & \text{if } f(x_i, a_l) \neq f(x_j, a_l), \\ 0 & \text{otherwise.} \end{cases}$$

Then $\langle \overrightarrow{U}, A, \{0, 1\}, \psi\rangle$ is called binary discernibility matrix with respect to S.

Definition 3 *[9] $S = \langle U, A, V, f\rangle$ is an information system, $B_k(k \leq r)$ is a partition reduct, let*

$$C = \bigcap_{k \leq r} B_k \tag{1}$$

denote the core attributes set.

For an information system, attributes reduction is a NP-hard problem. But it can be solved by heuristic approaches. In heuristic approaches, weights of attributes are the most important. And for a binary discernibility matrix $M = \langle \overrightarrow{U}, A, \{0, 1\}, \psi\rangle$, weights of attributes are obtained as follows[11]:

$$W(a) = \sum_{(x_i, x_j) \in \overrightarrow{U}} w(x_i, x_j) \times \psi((x_i, x_j), a), \tag{2}$$

where $a \in A$, $w(x_i, x_j) = \frac{1}{\sum_{b \in A} \psi((x_i, x_j), b)}$.

Example 1 *Considering a general information system (See Table 1.), it is about the weather of five workdays. Where U = {Monday, Tuesday, Wednesday, Thursday, Friday}, A = {humidity, rainy, frosty, sunny, windy}. The matrix M of $\langle \overrightarrow{U}, A, \{0, 1\}, \psi\rangle$ can be calculated as follows according to Definition 2.:*

$$M = \left(\begin{array}{c|ccccc} \overrightarrow{U} \setminus A & a_1 & a_2 & a_3 & a_4 & a_5 \\ \hline (x_1, x_2) & 1 & 1 & 1 & 1 & 1 \\ (x_1, x_3) & 1 & 0 & 1 & 1 & 1 \\ (x_1, x_4) & 1 & 1 & 0 & 0 & 1 \\ (x_1, x_5) & 1 & 1 & 1 & 0 & 1 \\ (x_2, x_3) & 0 & 1 & 0 & 1 & 0 \\ (x_2, x_4) & 1 & 0 & 1 & 1 & 0 \\ (x_2, x_5) & 1 & 0 & 0 & 1 & 1 \\ (x_3, x_4) & 1 & 1 & 1 & 1 & 0 \\ (x_3, x_5) & 1 & 1 & 0 & 1 & 1 \\ (x_4, x_5) & 0 & 0 & 1 & 0 & 1 \end{array} \right).$$

According to Equation (2), the weight of attribute a_1 can be calculated as follows:

$$\begin{aligned} W(a_1) &= \frac{1}{5} + \frac{1}{4} + \frac{1}{3} + \frac{1}{4} + \frac{1}{3} + \frac{1}{3} + \frac{1}{4} + \frac{1}{4} \\ &= \frac{132}{60}. \end{aligned}$$

3. Approach based on ant colony optimization

Let $b_i(t_l)(i = 0, 1, \cdots, |A|)$ denote the number of agents(artificial ants) in node i at the time t_l, then $m = \sum_{i=0}^{n} b_i(t_l)$ denotes the total amount of agents. Let $\tau_{ij}(t_l)$ denote the amount of pheromone on the edge: (i, j), and $tabu_k$ denote the nodes which agent k has gone.

Definition 4 *For a general information system $\langle U, A, V, f\rangle$ and its binary discernibility matrix $\langle \overrightarrow{U}, A, \{0, 1\}, \psi\rangle$. $\forall a_i \in A$, let a_i denote a node, and add a virtual node noted as Θ. $\forall (a_i, a_j) \in (A \cup \{\Theta\})^2$, there is an edge connects between node a_i and node a_j, and $\forall (x_i, x_j) \in \overrightarrow{U}$, let $\psi((x_i, x_j), \Theta) = 0$. Then, let all the agents move from Θ till making circles. And the anticipant degree of moving from a_i to a_j is defined as follows:*

$$\eta_{a_i a_j} = \sum_{(x_i, x_j) \in \overrightarrow{U}} (\psi((x_i, x_j), a_j) - \vee_{a \in tabu_k} (\psi((x_i, x_j), a_j) \wedge \psi((x_i, x_j), a))). \tag{3}$$

Where $a_j \in J_k = (A - tabu_k) \cup \{\Theta\}$, and J_k denotes the nodes which agent k can go. When $tabu_k = \{\Theta\}$, $\eta_{\Theta a_j} = \sum_{(x_i,x_j)\in\overrightarrow{U}} \psi((x_i, x_j), a_j)$.

In the iterative process: $t_l \to t_{l+1}$, the probability of agent k select node a_j when it is in node a_i is defined as follows:

$$p^k_{a_ia_j}(t_l) = \begin{cases} \frac{[\tau_{a_ia_j}(t_l)]^\alpha \cdot [\eta_{a_ia_j}]^\beta}{\sum_{a\in J_k}[\tau_{a_ia}(t_l)]^\alpha \cdot [\eta_{a_ia}]^\beta}, & \text{if } a_j \in J_k; \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Where, $\eta_{a_ia_j}$ is a heuristic parameter which denotes the anticipant degree of moving from node a_i to node a_j, and α, β are both parameters which denote the significance of accumulated information and heuristic factor respectively. In many cases, $\eta_{a_ia_j} = \frac{1}{d_{a_ia_j}}$, where $d_{a_ia_j}$ denotes the distance between node a_i and a_j. But in this approach, $\eta_{a_ia_j}$ is obtained according to Equation (3). After agents making circles, update the amount of pheromone on these paths according to the following equation:

$$\tau_{a_ia_j}(t_{l+1}) = (1-\rho)\cdot\tau_{a_ia_j}(t_l) + \Delta\tau_{a_ia_j}(t_l), \quad (5)$$

where $\rho (0 < \rho < 1)$ denotes the vaporing parameter of pheromone, $1-\rho$ denotes the supportable parameter of pheromone, and $\Delta\tau_{a_ia_j}(t_l)$ denotes the increaser amount of pheromone on the path: $a_i \to a_j$ in the iterative process: $t_l \to t_{l+1}$. And let

$$\Delta\tau_{a_ia_j}(t_l) = \sum_{k=1}^{m} \Delta^k\tau_{a_ia_j}(t_l). \quad (6)$$

Where, $\Delta^k\tau_{a_ia_j}(t_l)$ denotes the amount of pheromone which agent k leaves on the path: $a_i \to a_j$ in the iterative process: $t_l \to t_{l+1}$, and m denotes the number of agents. Let

$$\Delta^k\tau_{a_ia_j}(t_l) = \begin{cases} \frac{Q}{N_k(t_l)}, & \text{if agent } k \text{ moves through path: } a_i \to a_j; \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

Where Q is a constant and $N_k(t_l)$ denotes the number of nodes that agent k has passed(except the node Θ) in the iterative process: $t_l \to t_{l+1}$.

Remark 1 *Based on ACO, we can construct an ant system with respect to an information system. Then let agents simulate the real ants to search shorter paths. As we known, ants select paths according to the pheromone left on the paths. Then we define some parameters in order to guide these agents to make circles. And these circles are all the partition consistent sets of the information system.*

Theorem 1 *For a general information system $\langle U, A, V, f\rangle$ and an agent is in node a_i, $\forall a_j \in (A - tabu_k)$, $\eta_{a_ia_j} = 0$ if and only if $tabu_k - \{\Theta\}$ is a partition consistent set.*

Proof 1 $\forall a_j \in (A - tabu_k)$, $\eta_{a_ia_j} = \sum_{(x_i,x_j)\in\overrightarrow{U}}(\psi((x_i, x_j), a_j) - \vee_{a\in tabu_k}(\psi((x_i, x_j), a_j) \wedge \psi((x_i, x_j), a))) = \sum_{(x_i,x_j)\in\overrightarrow{U}}(\psi((x_i, x_j), a_j) - \psi((x_i, x_j), a_j) \wedge (\vee_{a\in tabu_k}\psi((x_i, x_j), a)))$. *Due to $\forall(x_i, x_j) \in \overrightarrow{U}$, $\psi((x_i, x_j), a_j) \geq \psi((x_i, x_j), a_j) \wedge (\vee_{a\in tabu_k}\psi((x_i, x_j), a))$, then $\eta_{a_ia_j} = 0 \Rightarrow \forall(x_i, x_j) \in \overrightarrow{U}$, $\psi((x_i, x_j), a_j) = \psi((x_i, x_j), a_j) \wedge (\vee_{a\in tabu_k}\psi((x_i, x_j), a)) \Rightarrow \forall(x_i, x_j) \in \overrightarrow{U}$, $\psi((x_i, x_j), a_j) \leq \vee_{a\in tabu_k}\psi((x_i, x_j), a) \Rightarrow \forall(x_i, x_j) \in \overrightarrow{U}$, if $\psi((x_i, x_j), a_j) = 1$, then $\exists a \in tabu_k$, s.t., $\psi((x_i, x_j), a) = 1 \Rightarrow (x_i, x_j) \in \overrightarrow{U}$, if $\psi((x_i, x_j), a_j) = 1$, then $\exists a \in (tabu_k - \{\Theta\})$, s.t., $\psi((x_i, x_j), a) = 1 \Rightarrow \forall(x_i, x_j) \in \overrightarrow{U}$, if $f(x_i, a_j) \neq f(x_j, a_j)$, then $\exists a \in (tabu_k - \{\Theta\})$, s.t., $f(x_i, a) \neq f(x_j, a)$. Then $\forall(x_i, x_j) \notin R_A \Rightarrow \exists b \in A$, s.t., $f(x_i, b) \neq f(x_j, b)$, if $b \in (A - tabu_k)$, then $\exists a \in (tabu_k - \{\Theta\})$, s.t., $f(x_i, a) \neq f(x_j, a)$, i.e., $(x_i, x_j) \notin R_{tabu_k - \{\Theta\}}$; if $b \in (tabu_k - \{\Theta\})$, and $f(x_i, b) \neq f(x_j, b)$ then $(x_i, x_j) \notin R_{(tabu_k - \{\Theta\})}$. Therefore, $R_{(tabu_k - \{\Theta\})} \subseteq R_A$, and due to $R_A \subseteq R_{(tabu_k - \{\Theta\})}$ holds all the time, so, $R_{(tabu_k - \{\Theta\})} = R_A$, i.e., $(tabu_k - \{\Theta\})$ is a partition consistent set.*

On the other hand, when $(tabu_k - \{\Theta\})$ is a partition consistent set, if $\exists a_j \in (A - tabu_k)$, s.t., $\eta_{a_ia_j} \neq 0$. And according to above discussion, $\forall(x_i,x_j) \in \overrightarrow{U}$, $\psi((x_i, x_j), a_j) \geq \psi((x_i, x_j), a_j) \wedge (\vee_{a\in tabu_k}\psi((x_i, x_j), a))$, so, $\exists(x_i, x_j) \in \overrightarrow{U}$, s.t., $\psi((x_i, x_j), a_j) = 1$, and $\vee_{a\in tabu_k}\psi((x_i, x_j), a)) = 0$, i.e., $\forall a \in (tabu_k - \{\Theta\})$, $\psi((x_i, x_j), a)) = 0$, namely, $\forall a \in (tabu_k - \{\Theta\})$, $f(x_i, a) = f(x_j, a)$, so, $(x_i, x_j) \in R_{(tabu_k - \{\Theta\})}$, but $(x_i, x_j) \in \overrightarrow{U} \Rightarrow \exists b \in A$, s.t., $f(x_i, b) \neq f(x_j, b) \Rightarrow (x_i, x_j) \notin R_A$. Therefore, $R_{(tabu_k - \{\Theta\})} \subseteq R_A$ does not hold. It is inconsistent with that $(tabu_k - \{\Theta\})$ is a partition consistent set. So, $\forall a_j \in (A - tabu_k)$, $\eta_{a_ia_j} = 0$.

Note 1 *For agent k is in node a_i, if $\forall a \in J_k$, $\eta_{a_ia} = 0$, then for any $a_j \in J_k$, let*

$$p^k_{a_ia_j}(t_l) = \frac{[\tau_{a_ia_j}(t_l)]^\alpha \cdot [\eta_{a_ia_j}]^\beta}{\sum_{a\in J_k}[\tau_{a_ia}(t_l)]^\alpha \cdot [\eta_{a_ia}]^\beta} = \frac{[\tau_{a_ia_j}(t_l)]^\alpha}{\sum_{a\in J_k}[\tau_{a_ia}(t_l)]^\alpha}. \quad (8)$$

Table 2: Status of crops in China

$U \setminus A$	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}
1983	10,350	29,854	16,695	1,402	1,316	32	55	0	332	0	546
1984	10,839	30,679	15,840	1,690	1,445	37	59	149	432	0	619
1985	11,382	33,140	15,588	1,927	1,655	47	59	160	535	0	705
1986	11,896	33,719	16,623	2,112	1,796	59	62	188	555	333	824
1987	12,191	32,773	18,034	2,216	1,835	79	72	219	590	379	955
1988	12,538	34,222	20,153	2,480	2,018	96	80	274	696	419	1,061
1989	12,805	35,281	21,164	2,629	2,123	107	96	282	720	436	1,152
1990	13,021	36,241	21,002	2,857	2,281	126	107	323	795	475	1,237
1991	13,193	36,965	20,621	3,144	2,452	154	118	395	922	524	1,351
1992	13,485	38,421	20,733	3,431	2,635	180	125	454	1,020	564	1,557
1993	13,988	39,300	21,731	3,842	2,854	234	137	274	1,180	564	1,823
1994	14,919	41,462	24,053	4,499	3,205	327	101	755	1,479	609	2,143
1995	15,862	44,169	27,865	5,260	3,648	415	202	935	1,677	673	2,517
1996	13,361	36,284	23,728	4,584	3,158	356	181	833	1,897	736	2,813
1997	14,542	40,035	25,576	5,269	3,596	441	213	979	1,897	681	3,602
1998	14,803	42,256	26,904	5,724	3,884	480	235	1,056	2,021	745	3,907
1999	15,025	43,020	27,926	5,949	3,891	505	251	1,116	2,135	807	4,122
2000	15,152	44,682	29,032	6,125	4,031	533	274	1,208	2,243	919	4,279
2001	14,996	45,743	29,826	6,334	4,184	549	293	1,210	2,337	1,123	4,374
2002	15,189	46,292	31,655	6,587	4,327	585	317	1,250	2,462	1,400	4,566
2003	15,500	46,602	34,054	6,933	4,519	631	357	1,312	2,607	1,849	4,705
2004	15,738	48,189	36,639	7,245	4,702	676	399	1,351	2,724	2,368	4,902

Theorem 2 *Agents make circles(coming back to node Θ) only when $tabu_k - \{\Theta\}$ is a partition consistent set in general information system.*

Proof 2 $\forall (x_i, x_j) \in \overrightarrow{U}$, $\psi((x_i,\ x_j),\ \Theta) = 0$, *so, when an agent is in node a_i, $\eta_{a_i\Theta} = 0$, then $p^k_{a_i\Theta}(t_l) \neq 0$ if and only if $\forall a \in J_k$, $\eta_{a_i a} = 0$(according to Note 1). So agent k make circles, i.e., coming back to node Θ only when $\forall a_j \in J_k = (A - tabu_k) \cup \{\Theta\}$, $\eta_{a_i a_j} = 0$ i.e., only when, $\forall a_j \in (A - tabu_k)$, $\eta_{a_i a_j} = 0$. Then according to Theorem 1, $tabu_k - \{\Theta\}$ is a partition consistent set.*

Remark 2 *According to Theorem 2, agents makes circles are all partition consistent sets in general information system. Then using ant colony optimization, we can obtain many different partition consistent sets synchronously, and we can obtain the partition reduct possibly. Furthered, the minimal reduct can be obtained possibly.*

4. Experiment

Table 2. is a general information about the status of crops in China(http://www.agri.gov.cn/sjzl/baipsh/WB2005.htm#12). There are 22 objects which represent the years and 11 attributes which represent the outputs of various crops, and denoted as a_1, $a_2 \cdots$, a_{11}.

4.1. Experiment about reducing attributes as maximum as possible

In this experiment, let $\alpha = 1$, $\beta = 1$, $Q = 10$, $\rho = 0.5$, and according to attributes' value orderly, we disperse their domain into average 4 classes which denoted as 1, 2, 3, 4. Then the continuous value information system can be transformed into a discrete value information system. According to Theorem 2, for each agent in every cycle, there is a partition consistent set(partition reduct) correspondingly. Our experiment has been performed 4 times according to 4 different cases: 10 agents and 5 cycles, 10 agents and 10 cycles, 100 agents and 10 cycles, 10 agents and 100 cycles (See Fig 1.).

From Fig 1., we can find that: for different cases, the number of partition consistent sets or partition reducts which contain 5 attributes is the maximum. And we have affirmed that the number of attributes in the minimal reducts of this information system is 5 when it is dispersed averagely

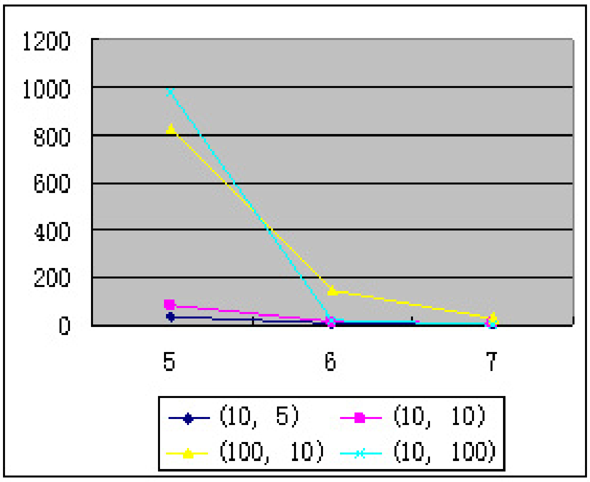

Fig. 1: Numbers of partition consistent sets in four cases

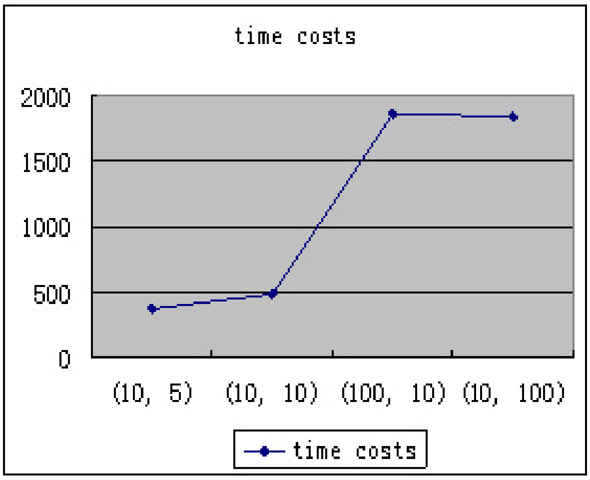

Fig. 2: Time costs in four cases

into a discrete 4-value information system.

According to the theory of ACO, we may get in the partial optimization sometimes. In this paper, that is partition consistent set or partition reduct with more attributes than other partition reduct. But we can still use this new approach to obtain many results synchronously.

Remark 3 *The time cost: T_c for this approach can be estimated as: $T_c \le O(m \cdot n \cdot |U|^2 \cdot |A|^2)$. Where, m denotes the number of agents and n denotes cycle count.*

For the binary discernibility matrix $\langle \overrightarrow{U}, A, \{0,1\}, \psi \rangle$ with respect to a given information system $\langle U, A, V, f \rangle$, $|\overrightarrow{U}| \le \frac{|U| \cdot (|U|-1)}{2}$. And for agent k, when it is in the original node, *i.e.*, Θ, the time cost: t_1^k for selecting next node is: $t_1^k \le O(\frac{|U| \cdot (|U|-1)}{2} \cdot |A|)$. Then, when this agent has select one node, there are $|A| - 1$ attributes left, so, the time cost: t_2^k for selecting next node is: $t_2^k \le O(\frac{|U| \cdot (|U|-1)}{2} \cdot (|A| - 1))$. And so on, the time cost for agent k coming back to the node Θ is: $t^k \le \sum_{i=0}^{|tabu_k|-1} O(\frac{|U| \cdot (|U|-1)}{2} \cdot (|A| - i)) \le \sum_{i=0}^{|A|} O(\frac{|U| \cdot (|U|-1)}{2} \cdot (|A| - i)) = O(\frac{|U| \cdot (|U|-1)}{2} \cdot \frac{|A| \cdot (|A|+1)}{2})$. So, for m agents and n cycle count, the total time cost: $T_c \le n \cdot \sum_{k=1}^{m} t^k \le O(m \cdot n \cdot |U|^2 \cdot |A|^2)$. From Fig 2., we can obtain affirmance of the time cost in this experiment, and we can reach a conclusion that the time cost is polynomial with numbers of attributes, objects, agents and cycle count, not exponential.

4.2. Remark about obtaining many different results synchronously and reducing the scope of core attributes

When using the approach which based on weights of attributes to reduce information system, we can obtain only one result always in this experiment. Based on the discussion above, for kinds of cases(different numbers of agents or cycles), we can obtain at most $a \times b$ partition consistent sets. Where a denotes the number of agents and b denotes the number of cycles. In despite of that there are much of the same results, we can still obtain many different results synchronously through this new approach. Then these different results can be selected to make different decisions.

And according to Definition 3., ordinarily, the more different partition consistent sets(partition reducts) we have, the smaller scope of core attributes is. Because we can obtain many different results(partition consistent sets or partition reducts) using this new approach, then we can reduce the scope of core attributes. Later, we can use the most important attributes in restricted cases.

5. Conclusion

In this paper, we present an approach about attributes reduction based on ant colony optimization(ACO). By this approach, we can obtain the results as simply as possible and reduce the scope of core attributes in order to assist managers. In the future, we can apply this approach into much larger information systems and knowledge reduction of decision information systems.

Acknowledgement

The authors would like to thank the anonymous referees for their very constructive comments. This work is supported by the Young Foundation of Sichuan Province (Grant No.06ZQ026-037) and the Education Department Foundation of Sichuan Province (Grant No.2005A121, 2006A084).

References

[1] B. Malcolm, Reducts within variable precision rough sets model: A further investigation. *European Journal of Operational Research*, 134:592-605, 2001.

[2] E. F. Lashin, T. Medhat, Topological reduction of information systems. *Solitons and Fractals*, 25:277-286, 2005.

[3] Z. Pei, R. Germano, J. V. D. W. Ariën, K. Y. Qin, Y. Xu, Interpreting and extracting fuzzy decision rules from fuzzy information systems and their inference. *Information Sciences*, 176:1869-1897, 2006.

[4] Z. Pawlak, Rough Sets. *International Journal of Computer and Information Science*, 11(5):341-356, 1982.

[5] Z. Pawlak, *Rough Sets: Theoretical aspects of reasoning about data*, Kluwer Academic Publishers, Boston, 1991.

[6] A. Skowron,C. Rauszer, The discernibility matrices and functions in information systems. In: Słowiński, R (ed.): Intelligent Decision Support-Handbook of Applications and Advances of the Rough Sets Theory, System Theory, *Knowledge Engineering and Problem Solving*, Kluwer, Dordrecht, 11:331-362, 1992.

[7] H. S. Nguyen, Approximate boolean reasoning. *Foundations and Applications in Data Mining, Transactions on Rough Sets V*, 4100:334-506, Springer, LNCS, 2006.

[8] R. Felix, T. Ushio, Rough sets-based machine learning using a binary discernibility matrix. *Proceeding of the second International Conference on Intelligent Proceeding and Manufacturing of Materials*, pp. 299-305, 1999.

[9] W. X. Zhang,Q. F. Qiu, *Uncertain Decision Making Based on Rough Sets*, Tsinghua University Publisher, Beijing(in Chinese), 2005.

[10] X. B. Hu, *The theory and applications of ant colony optimization*, College of Automation. Chongqing University. Ph. D. Thesis, 2004.

[11] Q. Y. Lin, *Study of Attribute Reduction Algorithms Based on Binary Discernibility Matrix and of Rule Extraction*, Mr. Thesis(in Chinese), 2004.

[12] T. Y. Zhi,D. Q. Miao, The transform of binary discernibility matrix and algorithm's construction about attributes reduction efficiently. *Computer Sciences*, 29(2):140-142, 2002.

[13] M. Dorigo,V. Maniezzo,A. Colorni, Positive feedback as a search strategy. *Technical Report. Dipartimento di Elettronica, Politecnico di Milano, IT,* pp.91-016, 1991.

[14] M. Dorigo, G. C. Di, The Ant Colony Optimization meta-heuristic, In D. Corne, M. Dorigo, and F. Glover, editors, *New Ideas in Optimization*, McGraw Hill. London. UK, 1999.

[15] M. Dorigo, G. D. Caro, L. M. Gambardella, Ant algorithms for discrete optimization. *Artificial Life*, 5(2):137-172, 1999.

[16] M. Dorigo, V. Maniezzo, A. Colorni, The Ant System: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems*, Man, and Cybernetics-Part B, 26(1):29-41, 1996.

[17] D. G. Caro, M. Dorigo, AntNet: Distributed Stigmergetic Control for Communications Networks. *Journal of Artificial Intelligence Research*, 9:317-365, 1998.

[18] S. P. Rafael, S. L. Heitor, A. F. Alex, Data Mining With an Ant Colony Optimization Algorithm. *IEEE Transactions on Evolutionary Computing*, 6(4):321-332, 2002.