

# Privacy-Preserving Data Mining Algorithm Quantum Ant Colony Optimization

Wu Jue<sup>1</sup>, Yang Lei<sup>1</sup>, Peng Lingxi<sup>2,\*</sup> and Liu Feng<sup>3</sup>

<sup>1</sup>College of Computer science and technology, Southwest University of Science and Technology, MianYang, Sichuan, China

<sup>2</sup>College of Computer science and education software, Guangzhou University, Guangzhou, Guangdong, China

<sup>3</sup>Internataional School of Software, Wuhan University, China

Received: 19 Oct. 2012, Revised: 21 Dec. 2012, Accepted: 10 Jan. 2013

Published online: 1 May 2013

**Abstract:** Bayesian network has been used extensively in data mining. The Privacy-Preserving data mining algorithm based on quantum ant colony optimization is proposed in this paper. The algorithm is based on distributed database. The algorithm is divided into two steps. In the first step, the modified quantum ant colony optimization algorithm is used to get the local Bayesian network structure. The purpose of the second step is getting the global Bayesian network structure by using local ones. In order to protect the privacy, the secure sum is used in the algorithm. The algorithm is proved to be convergent on theory. Some experiments have been done on the algorithm, and the results proved that the algorithm is feasible.

**Keywords:** Quantum Ant colony optimization, Bayesian network, data mining.

## 1. Introduction

Bayesian network is a graphic model which presents the relationship of variables. It provides a method to present the causality of the information. This method can find the relationship in the data. Bayesian theory provides belief function calculation method on mathematics, so Bayesian network has firm mathematical basement. Bayesian network can process incomplete data and data with noisy in data mining. The probability measurement weight is used in Bayesian network to present the relativity among the data, and this method can solve the inconsistency even dependency problem of the data, even dependency problem. The graphic method is used in the method to present the relationship among the data. It is clear and comprehensible for the forecast analysis. Bayesian network has been used extensively in data mining [1-5].

The quantum ant colony optimization algorithm is based on the quantum state vector. The quantum bit probability amplitude is used to describe ant position in quantum ant colony optimization algorithm. The rotate quantum gate is used to update the information of the ant information. The not quantum gate is used to avoid getting local optimal solution. The ant scale is alterable, and it does not influent the performance of the algorithm.

The algorithm has a fast convergent rate. Privacy-preserving data mining algorithm[6,7] based on modified quantum ant colony optimization (PPDM-QANCO) is proposed in this paper. This algorithm is focused on the distributed database. The modified quantum ant colony optimization is used to solve the problem of Bayesian network structure learning on each data node. The privacy-preserving problem is focused on the communicate process on the distributed database. Only the local Bayesian network structure is need to be preserved in the algorithm. The secure sum scheme is used to preserve the local data in the algorithm. The Bayesian network learning based on modified quantum ant colony optimization algorithm is proved to be convergent on theory, and the experiment result is also show that the algorithm is convergent. As to the privacy-preserving, the secure sum scheme is proved to be feasible according to the experiment result.

\* Corresponding author e-mail: wujue826@yahoo.com.cn

## 2. Algorithm Model

### 2.1. Bayesian network

The method of the graph theory is used to express the joint distribution of the variables set according to the relationship of the variables. The Bayesian network can be defined as follows[8]. Definition 1: the Bayesian network can be described as a triple  $G = \langle V, \lambda, P \rangle$  Where  $V$  means a nodes set,  $V = \{X_1, X_2, \dots, X_n\}$ , every node represents an attribution.  $\lambda$  represents a set of edges with direction, and  $\lambda = \{< X_i, X_j > | X_i \neq X_j, X_i, X_j \in V\}$ .  $< X_i, X_j >$  represents that there is a dependency relationship between  $X_i$  and  $X_j$ .

The purpose of Bayesian network structure learning is to get an entire Bayesian network which includes the network topology and the conditional possibility table. The purpose of network learning is to find a Bayesian network which is matched well to the data samples. There is a fitness function which can describe the accuracy of the Bayesian network. The BIC[6] measure function is used as the fitness function in the algorithm. The problem can be described as formula (1).

$$f = BIC(\xi | D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \lg \frac{m_{ijk}}{m_{ij}} - \sum_{i=1}^n \frac{q_i(r_i - 1)}{2} \lg m \quad (1)$$

Where  $\xi$  the means the a Bayesian network structure which is composed by  $n$  variables  $X = \{x_1, x_2, \dots, x_n\}$ .  $q_i$  presents the number of  $x_i$  parent nodes value combination. If there is no parent node,  $q_i = 1$ .  $r_i$  presents the number of  $x_i$  value.  $m_{ijk}$  presents the value of sample when  $X_i$  parent node is  $j$  and  $x_i$  equals to  $j$ .  $m_{ij} = \sum_{k=1}^{r_i} m_{ijk}$ ,  $\theta_{ij} = \frac{m_{ijk}}{m_{ij}}$  presents the likelihood conditional probability. Where,  $0 \leq \theta_{ijk} \leq 1$ ,  $\sum_k \theta_{ijk} = 1$ .

### 2.2. Encoding Method

Quantum bit as the information storage unit is a two-state quantum system. It is a unit vector which is defined in a two-dimensional complex vector space. The space is composed of standard orthogonal basis  $\{|0\rangle, |1\rangle\}$ . Therefore, it can be in the superposition of quantum at the same time. The state can be represented as below[9, 10].

$$|\phi\rangle = \alpha|0\rangle + \beta|1\rangle \quad (2)$$

Where  $\alpha$  and  $\beta$  are complex numbers that specify the probability amplitudes of the corresponding states.  $|\alpha|^2$  gives the probability that the Q-bit will be found in the 0 state and  $|\beta|^2$  gives the probability that the Q-bit will be found in the 1 state. Normalization of the state to unity

guarantees the equation (3).

$$|\alpha|^2 + |\beta|^2 = 1 \quad (3)$$

If there is a system of  $m$  Q-bits, the system can represent  $2^m$  states at the same time. However, in the act of observing a quantum state, it collapses to a single state. It is encoded with the quantum bit of probability amplitude directly. The quantum  $q_i$  which presents the minimum distance equation can be described as follows.

$$q_i = \begin{bmatrix} \left| \cos(t_{i1}) \right| & \left| \cos(t_{i2}) \right| & \dots & \left| \cos(t_{in}) \right| \\ \left| \sin(t_{i1}) \right| & \left| \sin(t_{i2}) \right| & \dots & \left| \sin(t_{in}) \right| \end{bmatrix} \quad (4)$$

The position of an ant represents a solution of the problem in the continuous ant colony optimization algorithm. Supposed that there are  $m$  ants which are distributed randomly in the  $n$ -dimension space  $[-1, 1]^T$ . Each ant has  $n$  quantum bits. The probability amplitude represents the current position of the ant. In the adjacency matrix is used to describe the Bayesian network in the algorithm. The Bayesian network is directed acyclic graph, and so the data on the diagonal of the matrix are equal to zero, and in the encoding scheme the quantum on the matrix adjacency collapse to zero. The ant can be initialized as follow.

$$q = \begin{bmatrix} 1 & \sqrt{2}/2 & \dots & \sqrt{2}/2 \\ 0 & \sqrt{2}/2 & \dots & \sqrt{2}/2 \\ \sqrt{2}/2 & 1 & \dots & \sqrt{2}/2 \\ \sqrt{2}/2 & 0 & \dots & \sqrt{2}/2 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \sqrt{2}/2 & \sqrt{2}/2 & \dots & 1 \\ \sqrt{2}/2 & \sqrt{2}/2 & \dots & 0 \end{bmatrix} \quad (5)$$

### 2.3. Select Ants Moving Target Location

Set  $\tau(x_r)$  as the pheromone intensity of the  $k$  ant which is at the position of  $x_k$ . It is initialized as a constant. Set  $\eta(x_r)$  as the visibility of the position of  $x_k$ . The probability of the ant  $k$  moves from  $x_r$  to  $x_s$  is described as follow.

$$p_{xs} = \frac{[\tau(x_s)]^\alpha [\eta(x_s)]^\beta}{\sum_{x_s, x_u \in X} [\tau(x_u)]^\alpha [\eta(x_u)]^\beta} \quad (6)$$

### 2.4. Quantum Rotation Gate

The Rotation operation updates the Quantum bit (Q-bit) by using rotate gate. This operation makes the ant population develop to the best individual. The Q-gate is defined by formula (7).

$$U(\Delta\theta) = \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) \\ \sin(\Delta\theta) & \cos(\Delta\theta) \end{bmatrix} \quad (7)$$

Q-bit is updated as follow.

$$\begin{bmatrix} \cos(\Delta\theta) - \sin(\Delta\theta) \\ \sin(\Delta\theta) \cos(\Delta\theta) \end{bmatrix} \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix} = \begin{bmatrix} \cos(t + \Delta\theta) \\ \sin(t + \Delta\theta) \end{bmatrix} \quad (8)$$

We can see from the formula (8) that this update operation only changes phase of the Q-bit, but does not change the length of the Q-bit. Where  $\Delta\theta$  is a rotation angle of each Q-bit. The magnitude of  $\Delta\theta$  has an effect on the speed of convergence, but if it is too big the solution may diverge or converge prematurely to a local optimum. The sign of  $\Delta\theta$  determines the direction of convergence. This can be determined by the following method. Based that the  $\alpha_0\beta_0$  is the probability amplitude of the global optimal solution in the current search,  $\alpha_1\beta_1$  is the probability amplitude of Q-bit in current solution. Let us define A as follows.

$$A = \begin{vmatrix} \alpha_0 & \alpha_1 \\ \beta_0 & \beta_1 \end{vmatrix} \quad (9)$$

The direction of  $\Delta\theta$  can be determined as follows. If  $A \neq 0$  the direction is  $-\text{sgn}(A)$ , if  $A = 0$  the direction can be selected randomly. In order to avoid the premature convergence the size of  $\Delta\theta$  can be determined as formula (10) described. This method is a dynamic adjustment strategy and has nothing to do with the problem.  $pi$  is the circumferential ratio. The maxGen represents the maximal number of iterations times.

$$\Delta\theta = 0.5 * pi * \exp(-gen / \text{max Gen}) \quad (10)$$

## 2.5. The Update Rules of Pheromone Intensity and Visibility

The function of the optimal problem is responsible to evaluate the ant position, and the main idea of the pheromone intensity updating is that function value is integrated with the pheromone. The better the value of the function is, the higher the pheromone intensity is. The gradient information of the function is also integrated with the visibility, and this make that the position with greater gradient, the visibility is greater. The point found in this way not only has higher fitness also has higher rate of change. Each ant computes the function value after one step research, and updates the local pheromone intensity and the visibility according to the rules given as follow. Supposed that the previous position of the ant is  $x_q$ , the current position is  $x_r$ , the next position is  $x_s$ . The updating rule is described as follow.

$$\tau(x_s) = \tau(x_r) + \text{sgn}(f) * |f|^\alpha \quad (11)$$

$$f = f(x_s) - f(x_r) \quad (12)$$

$$\eta(x_s) = \eta(x_r) + \text{sgn}(\partial f) * |\partial f|^\beta \quad (13)$$

$$\partial f = \max_{1 \leq i \leq n} \left( \frac{\partial f}{\partial x_{si}} \right) - \max_{1 \leq i \leq n} \left( \frac{\partial f}{\partial x_{ri}} \right) \quad (14)$$

When the function f is a non-differentiable function, we can use the first difference. It is described as follow.

$$\partial f = \max_{1 \leq i \leq n} \left( \frac{f(x_s) - f(x_r)}{x_{si} - x_{ri}} \right) - \max_{1 \leq i \leq n} \left( \frac{f(x_r) - f(x_q)}{x_{ri} - x_{qi}} \right) \quad (15)$$

$$\tau(x_u) = \begin{cases} (1 - \rho)\tau(x_u) + \rho \text{fit}(x_u)x = \tilde{x} \\ (1 - \rho)\tau(x_u)x \neq \tilde{x} \end{cases} \quad (16)$$

## 2.6. Bayesian network classified problem based on modified ant colony optimization

The steps of the algorithm can be described as follows. Step 1: Initialization. Because that the Bayesian network is a directed acyclic graph, the data on the diagonal equal to zero as described in formula (5).

Step 2: Observe the quantum, and get the possible solutions. In the adjacency matrix, if there is a directed edge between point i and j,  $A[i, j] = 1$ . If there is no directed edge between i and j,  $A[i, j] = -1$ . Calculate the objective function value.

Step 3: The rotation gate and the mutation gate are used to update the ant colony. Observe the new quantum to get new possible solutions. The new ant colony is revised if it is not a directed acyclic graph.

Step 4: Calculate the objective function, supposed that the global optimal function  $f^*$ , the global optimal solution is  $x^*$ . If  $f^* > f_k$ ,  $f^* = f_k$ ,  $x^* = x_k$ .

Step 5: Update the phenomenon and the visibility according to the formula (11) and (13). Update the global phenomenon according to formula (16).

Step 6: Checking the end condition, if it is not satisfied turn to the step 2, otherwise the optimal solution can be output.

## 2.7. Secure Sum

The main idea of the algorithm is to generate a matrix which has the same size with the local data matrix, and add the random matrix on the local data matrix, sends the merged matrix to the following party. Each party receives the perturbed matrix, adds it with its local matrix, and passes it to the following party (the last party sends the matrix back to the first party). The first party subtracts the random matrix from the received matrix, which results in the matrix that adds the matrices of all parties, without disclosing their local matrices to each other. The method preserves the data privacy, sine only the original party gets to exactly see the data. The local model is directly

computed from the local data. But this algorithm is very fragile to the collusion attack. The parties preceding and following a party can collude to recover its local matrix [7,8]. Let  $x_i$  represents the data held by the party  $i$ ,  $k$  is the number of the data slice. All the data split into  $k$  slices. And the  $x_{ij}$  satisfies the equations (17). The  $x_{ij}$  and  $x_i$  are in the same interval.

$$x_i = \sum_{j=1}^k x_{ij} \quad (17)$$

1. Every party sends the  $x_{i2}, x_{im}$  to the other  $m - 1$  parties, and keep the  $x_{i1}$  as private.
2. Every party receives  $k - 1$  numbers which come from the other  $k - 1$  parties.
3. Every party adds the  $k - 1$  numbers to its own  $x_{i1}$ , and sends the sum to the data center.
4. What the data center gets is as follow.

$$\sum_{i=1}^n \sum_{j=1}^k x_{ij} = \sum_{i=1}^n x_i \quad (18)$$

From the formula (18), we can see what the data center gets is the global gram matrix of the distributed database. During the procedure of the algorithm, every party keeps the  $x_{i1}$  as private. And so it is very difficult to get the data of other parties. If the data center does not take part in the collusion attack, what the other  $m - 1$  parties can get is  $\lim_{k \rightarrow \infty} P\{f_k = f^*\} = 1$  and they cannot obtain the data  $x_{i1}$ . And so they cannot get the exact data of the party. This has nothing to do with other coefficients. If the data center takes part in the collusion attack, the collusion parties can get the exact data of one party only when they receive the  $k - 1$  data.

The data needed to be protected in the algorithm is the local Bayesian network structure. Supposed that the local Bayesian network is  $G = \langle V_i, E_i \rangle$ ,  $E_i$  represents the edge set,  $V_i$  represents the point set. The Bayesian network is described by matrix. The matrix element is -1 or 1. The secure sum scheme is used in the algorithm. Supposed the number of the node is  $S$ . If the matrix element is equal to  $S$ , the oriented edge saved in edges set  $E_a$ , otherwise the edges saved in edges set  $E_b$ .  $E_a$  is the edges in the global Bayesian network.  $E_b$  is the undetermined edges in the global Bayesian network. The mutual information of the edges in the set  $E_b$  is calculated in the next step. If the value of the mutual information is larger than threshold value, then add the corresponding edge to edges set  $E_a$ . The edges in set  $E_a$  is the global Bayesian network structure.

### 3. Convergence Analysis

Theorem 1: The PPDM-QANCO algorithm ant population sequence  $\{Q(t), t > 0\}$  is the finite homogeneous Markov chains. Proof: The Q-bits is used

in the algorithm. In the ant colony evolution algorithm, the value of ants is discrete. Assumed that the length of the ant is  $m$ , the ant population size is  $n$ , so the state space size is  $2^{mn}$ . Because of the continuity of the value of variable, the size of the state space size is infinite in theory, but the accuracy is finite during the calculating. Assumed that the dimension is  $V$ , then the state size of the population is  $V^{mn}$ , and so the population is finite. In the algorithm, the operation of rotate has nothing to do with the generation number, and the pheromone intensity updating is according to the formula (11), the visibility updating is according to the formula (13). All this have nothing to do with the generation number. Therefore the population sequence is a finite homogeneous Markov chain.

Definition 2:  $f_k = \max\{f(x_i), i = 1, 2, \dots, N\}$  is a random variant sequence. The variable represents the best solution in the  $k$  generation. If the condition meets the formula (19), then we can declare that the algorithm is convergent with probability 1 to the global optimal solution.

$$\lim_{k \rightarrow \infty} P\{f_k = f^*\} = 1 \quad (f^* = \max\{f(b) | b \in \Omega\}) \quad (19)$$

Theorem 2: PPDM-QANCO algorithm is convergent with probability 1 to the global optimal solution. Proof: According to the theorem 1, the ant colony is a finite homogeneous Markov chain. Supposed that the population of ants is  $m$ , the ants are points in research space  $\Omega$ ,  $X_i \in \Omega$  represents that the  $X_i$  is a point in  $\Omega$ .  $X_i = \{x_1, x_2, \dots, x_n\}$ ,  $X_k^i$  represents that in the  $k$  iteration, the ant  $X$  is at the point  $X_i$ . The ant moves to the point  $j$  from point  $i$  in the search space after a step of iteration in the randomize procedure. The transfer possibility is  $P\{X_{k+1}^j | X_k^i\}$ . During the transfer procedure, there are two special situations [14, 15].

(1)The first situation is that  $X_{k+1}^j \notin f^*$ ,  $X_k^i \in f^*$ . The optimal solution is preserved in the algorithm .This operation ensured that each iteration will not degenerate, and so the transfer possibility  $P\{X_{k+1}^j | X_k^i\} = 0$ .

(2)The second special is that  $X_{k+1}^j \in f^*$ ,  $X_k^i \notin f^*$ . According to the algorithm operation, we can come to a conclusion that the transfer possibility  $P\{X_{k+1}^j | X_k^i\} > 0$ .

After describing two special situations, we begin to proof the convergence. Supposed that  $X_k$  is at the position  $i$ , and the possibility is marked as  $p_k^i$ .

$$p_k = \sum_{X_i \notin f^*} p_k^i \quad (20)$$

$$p_{k+1} = \sum_{X_i \in f^*, X_j \notin f^*} P\{X_{k+1}^j | X_k^i\} + \sum_{X_i \notin f^*, X_j \notin f^*} P\{X_{k+1}^j | X_k^i\} \quad (21)$$

$$\sum_{X_i \notin f^*, X_j \in f^*} p\{X_{k+1}^j | X_k^i\} + \\ \sum_{X_i \notin f^*, X_j \notin f^*} p\{X_{k+1}^j | X_k^i\} = \sum_{X_i \notin f^*} p\{X_k^i\} = p_k \quad (22)$$

From the formula (22), we can get the formula (23).

$$\sum_{X_i \notin f^*, X_j \notin f^*} p\{X_{k+1}^j | X_k^i\} \\ = p_k - \sum_{X_i \notin f^*, X_j \in f^*} p\{X_{k+1}^j | X_k^i\} \quad (23)$$

The formula (23) is plugged into the formula (21), we get the formula(24)

$$0 \leq p_{k+1} < p_k - \sum_{X_i \notin f^*, X_j \in f^*} p\{X_{k+1}^j | X_k^i\} + \\ \sum_{X_i \in f^*, X_j \in f^*} P\{X_{k+1}^j | X_k^i\} \\ < p_k + \sum_{X_i \in f^*, X_j \in f^*} P\{X_{k+1}^j | X_k^i\} = p_k \quad (24)$$

According to the two special situation described above, we can get  $\lim_{k \rightarrow \infty} p_k = 0$ ,  $\lim_{k \rightarrow \infty} p(f_k = f^*) = 1 - \lim_{k \rightarrow \infty} p_k = 1$ . This means that the algorithm is global convergence. This means that the algorithm is global convergence.

#### 4. Simulation Experiment

The Asia network is used to be an example to do some experiment in the paper. It is a very small belief network for a fictitious medical example about whether a patient has tuberculosis, lung cancer or bronchitis, related to their X-ray, dyspnea, visit-to-Asia and smoking status. It is also called "Chest Clinic". There are eight variables in the Asia network. Each variable only has two possible values. The data in the experiment are created by the software Netica according to the probability of the Asia network. There are there data sets with different scale. In the experiment, there are three data set which scale is 10000, 30000, 50000 separately. Algorithm K2 is an important algorithm on Bayesian network structure learning. It is used to compare with the algorithm supposed in this paper. K2 algorithm and the modified ant colony optimization algorithm are used to get Bayesian network structure on three different data sets separately. From Table 1, we can draw a conclusion that the PPDM-QANCO algorithm can get more and more accurate Bayesian network structure than K2 algorithm with the development of number of data. Fig.1 is the best result of the ten times execute of PPDM-QANCO algorithm. Compared with the standard Asia Bayesian

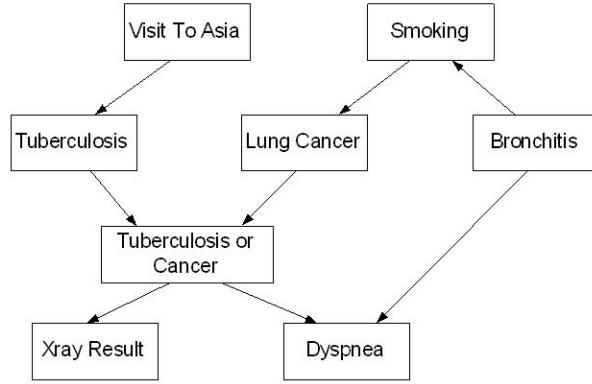


Figure 1: Asian Bayesian network structure learnt by PPDM-QANCO

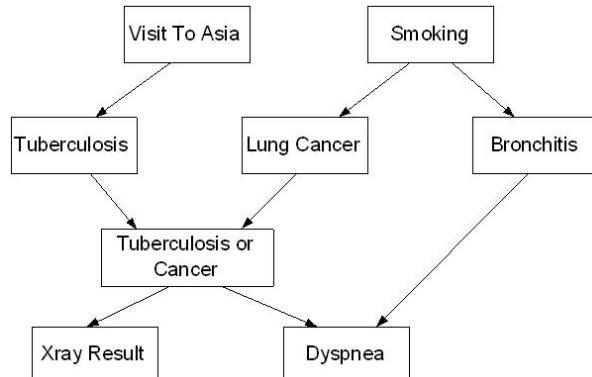


Figure 2: Standard Asia Bayesian network structure

network, we can see that there only one reversed edges. The PPDM-QANCO algorithm is more efficient than K2 algorithm. As to the privacy-preserving, the run time of the algorithm with the different party number is described on the Fig.4. From the Fig.5 we can draw a conclusion that the run time is not enhance too more with the enhancement of the party number. That is to say that the algorithm has the run time stabilization with the increasing party number.

Table 1: PPDM-QANCO compared with K2

Algorithm	10000			30000			50000		
PPDM-QANCO	1.0	1.0	0.8	1.0	1.0	0.8	0.6	0.8	0.0
K2	1.0	1.8	1.0	1.0	1.7	1.0	1.0	1.0	0.9

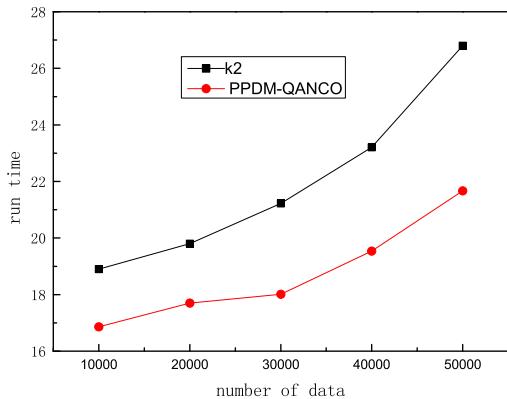


Figure 3: Run time compared with K2

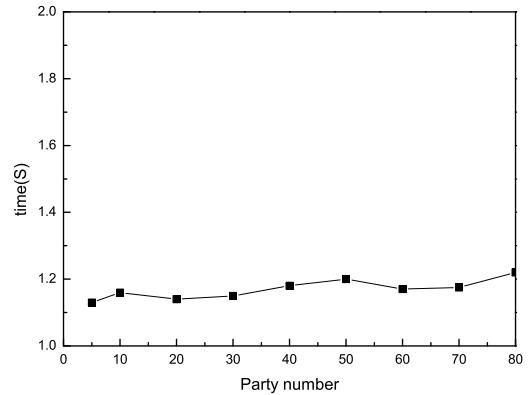


Figure 5: Run time with different distributed node number

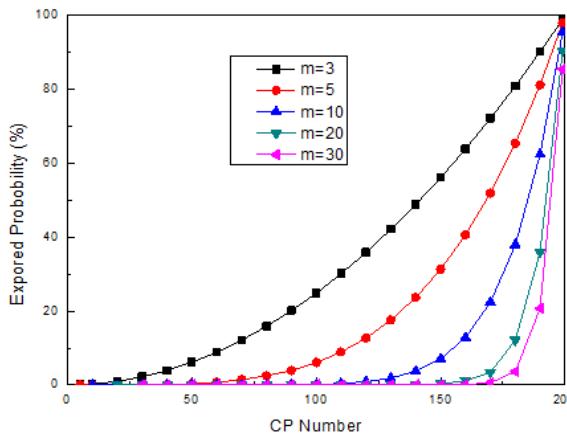


Figure 4: The PPDM-QANCO performance in collusion attack

## 5. Conclusion

The modified ant colony optimization algorithm is used to do data mining on distributed database. It is used on each data node to get the local Bayesian network structure. The privacy-preserving in distributed database is focused on the communication among the data nodes. The results of the local Bayesian network learning is needed to be protect during the process of private preserving data mining on distributed database. The secure sum is used to preserve the privacy in the algorithm to do the privacy-preserving. The algorithm is proved to be feasible on theory and experiment.

## Acknowledgement

The authors acknowledge the financial support of the National Natural Science Foundation of China under Grant No. 61100150, and the Natural Science Foundation of Guangdong Province of China under Grant No. S2011040004528 and No. S2011040003843.

## References

- [1] Xu LijiaHuang JianguoWang HoujunLong Bing, Computer-Aided Design & Computer Graphics **21**, 633 (2009).
- [2] Zhao JunshengLi YueguangZhang Yuanping, Computer Application and Software **27**, 133 (2010).
- [3] Xing Cheng Heng, Zhang Qin, Xian Hui Wang, information Technology Journal **3**, 540 (2006).
- [4] Wang WeiCai Lianhong Mini-micro System **23**, 435 (2002).
- [5] Wang Hongmei, Dissertation for Philosophy Degree in Tianjin University. **2**, (2006).
- [6] Dai, L (Dai, Lu); Ding, LX (Ding, Lixin); Lei, YW (Lei, Yunwen); Tian, YG, Applied Mathematics & Information Sciences. **6**, 705 (2012).
- [7] Lee, YC (Lee, Yung-Cheng), Applied Mathematics & Information Sciences. **6**, 361s (2012).
- [8] Wang HaishuLiu GangQi Zhaohui, Computer Engineering **34**, 229 (2008).
- [9] Li Panchi, song Kaoping, Systems Engineering-Theory and Practice **6**, 759 (2011).
- [10] Song Kaoping, Yang Erlong,Li Panchi, A Information and Control **39**, 681 (2010).
- [11] Wang Hongmei, Zeng Yuan, Zhao Zheng, Journal of Tianjin University **40**, 1025 (2007).
- [12] Ge Weiping, Wang Wei, Zhou Haofeng, Shi Baile, Journal of Computer Research and Development **43**, 39 (2006).
- [13] Jie Wang, Dissertation for Philosophy Degree in University of Kentucky, 485 (2010).
- [14] Xue Han, Jin Min, Ma Hong-xu, Y.Q. Shi, X.F. Zheng and H. Zhao, Journal of System Simulation **21**, 6462 (2009).

[15] Huang Jing-wen, Qin Chao-yong, Application Research of Computer **26**, 3660 (2009).



**Wu Jue** is presently working in Southwest University of science and technology. She obtained her PhD from Southwest Petroleum university (China) in 2012. Her recent research interests include privacy preserving in data mining, cloud computing.



**Peng Lingxi** obtained his PhD in computer science from Sichuan University (China) in 2008. He is currently an associate professor at the College of Computer Science and Educational Software, Guangzhou University. His research interests include

artificial immune, algorithm design and analysis, bioinformatics algorithms, database, network security, etc.