ISSN : 0974 - 7435 Volume 10 Issue 5

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 10(5), 2014 [1327-1336]

ACO prototype system optimization –based k means clustering algorithm research

Xin Wang[1*], Zhi Xu[1], Wei Yuan[2]

[1]Institute of electrical and automation engineering, Sanjiang University, Yuhuatai District, Nanjing, 210012, Jiangsu, (CHINA)

[2]College of elect ronic information, Jiangsu University of science and technology, Zhenjiang, 212003, Jiangsu, (CHINA)

ABSTRACT

Cluster analysis is an important method in image identification, information retrieval, data mining and spatial database research, from which K means algorithm is a kind of clustering algorithm based on classification method, the algorithm thought is providing K pieces of classification on N pieces of objects, and every classification of them represents a cluster, by comparing every cluster calculated mean and all patterns samples mean, it gets a most similar cluster, constantly repeat such process till objects in cluster all are similar and different clusters' objects are different, while objective function convergence lets square error function value to be the minimum one. ACO(Ant Colony Optimization)is a kind of simulating ant colony foraging behaviors' bio-inspired optimization calculation, due to the algorithm reflects prominent applicability in complex optimization problems' solution aspect, let it to get well applied in robot system, picture processing, manufacturing system, vehicle route system and communication system. Therefore, the paper analyzes K means clustering algorithm, it gets the algorithm shortcomings, and uses ACO prototype system to optimize K means clustering algorithm, and states the algorithm feasibility and superiority.

KEYWORDS

ACO prototype system;
K means clustering algorithm;
Pheromone;
Ant colony.

INTRODUCTION

In daily life, it often sees ant colony, these ant colony always can find relative shortest route between its nest and food source, on the basis of above phenomenon, scholars make research on it and finds that every ant is uncertain about which route is the shortest beforehand, they only focus on small range pheromone concentration, and then use some simple rules to make decision, these pheromones are odors in colony cooperative foraging route, the whole ant colony is using the pheromone to mutual cooperate and communicate, during the process these pheromones form into positive feedback that let multiple routes ants gradually gather in the shortest route. Simulated ant colony foraging behavior bio-inspired optimization algorithm is called ACO(Ant Colony Optimization), the algorithm has rapidity, distributed and global optimization features in complex optimization problems calculation aspect, from which look for optimal solution rapidity is up to pheromone positive feed-

FULL PAPER

back mechanism, and the algorithm distributed calculation reasonably avoids algorithm premature convergence. Cluster analysis makes classification on pattern samples according to different objects differences and specialized criterion, is a kind of mathematical analysis method, and also a kind of machine learning method without supervisor monitoring that belongs of NP difficult problems, but the analysis method has been widely applied in image identification, information retrieval, data mining, statistics, machine learning, spatial database, biology and marketing as well as other fields, in order to better realize cluster analysis, the paper provides a kind of ACO prototype system optimization-based K means clustering algorithm that provides theoretical basis for cluster analysis development.

Regarding ACO development, it should bring up from M.Dorigo and others' proposed improvement in 1996, in the beginning, ACO was applied into solving TSP problems and got better efficiency, but it was not fit for solving large scale complex optimization problems, based on the shortcoming, M.Dorigo and others based on ant system, they made three aspects improvements as providing a better state transferring selection strategy, updating everywhere that only applied in current optimal ant route, and applying partial information into updating pheromone during constructing solution process, and then applied ant colony system after improving into TSP problems solution that could get better application. For ant colony algorithm, lots of people have made efforts, used researches to let the algorithm to have stronger operability, such as Yan'an university computing center's Du Li-Feng and Niu Yong-Jie in "Ant colony algorithm implementation in Matlab", they described ant colony algorithm operating principle, provided the algorithm detailed steps in implementation in Matlab, finally respectively took 17, 21, 24, 48, 51, 70 as cities scale to verify algorithm, provided algorithm operation optimal result, worst effects, average result and running time as well as result table, which provided basis for algorithm applying in other fields and further improving[1]; Besides, Wu Qin-Hong, Zhang Ji-Hui and Xu Xin-He in "Ant colony algorithm with variation features", they detailed introduced the algorithm, and then stated ant colony algorithm with variation features, which provided theoretical basis for ACO development[2]; Wei Ping and Xiong Weil-Qing in "A kind of solving functions optimized ant colony algorithm", they proposed a kind of improved ant colony algorithm that made contributions to quicken computation speed and improve accuracy[3]; Yan Chen-Yang, Zhang You-Peng and Xiong Wei-Qing in "A kind of new ant colony optimization algorithm pheromone updated strategy", they made research on information updating method, and got a kind of new optimization algorithm, which provided well improvement directions for ant colony algorithm computing speed and global solution[4]. The paper on the basis of previous research, it puts forward a kind of ACO prototype system optimization-based K means clustering algorithm that provides theoretical basis for ACO and cluster analysis combined development.

ACO PROTOTYPE SYSTEM AND OPTIMIZATION AND IMPROVEMENT

ACO has achieved better effects in a series of combinatorial optimization solution, and apply ACO into practical problems, researchers put forward artificial ant colony concept, artificial ant colony and real ant colony have lots of similarities, but it also has unique skills that real ant colony doesn't possess, in the section, on the basis of introducing ACO thought, it designs the algorithm prototype system and provides optimization and improvement opinions for the system.

ACO algorithm thought

Most of artificial ant colony behaviors are from real ant colony, is ant colony searching route shortest foraging process, during the process, pheromone plays an important role, route with more pheromones, and number of ants that crossing the route will be more, after

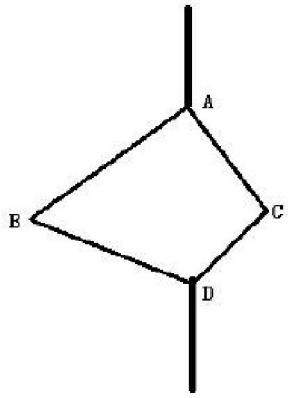

Figure 1: Ant colony foraging schematic diagram

positive feedback for a while, it will let more ants gather in the relative shortest route. As Figure 1 shows the nest, foot source and route schematic diagram.

In Figure 1, point A is ant nest, point D is food source, in the figure, it lists two routes, one of them is A-B-D, and the other is A-C-D, and it has AB=BD=2AC=2CD.

Assume that in every Δt time interval, it has n ants move from point A to point D, these ants climb with a unit speed, one ant can leave a unit pheromone in a unit length route, and assume when $t=0$, route A-B-D and route A-C-D pheromones are 0, so the n ants move from point A to point D route selection probabilities are the same, then now it has $\frac{n}{2}$ ants walk in the route A-B-D, another $\frac{n}{2}$ ants walk in the route A-C-D.

Based on above assumption, start from moment $t=0$ and after Δt time interval, depart from point A and through route A-C-D the $\frac{n}{2}$ ants arrive at point D in advance, and in the route A-C-D, it leaves $\frac{n}{2}$ pheromone, when ants find it then retrace the paths, after Δt time interval, they return to point A, while in the route A-C-D, they also leave $\frac{n}{2}$ pheromone, so after $2\Delta t$ time interval, route A-C-D pheromone is n, however depart from point A and through route A-B-D the $\frac{n}{2}$ ants after $2\Delta t$ time interval, they just walk to point D, so in the time $2\Delta t$, route A-B-D pheromone is $\frac{n}{2}$, when start from moment $t=2\Delta t$, it has n ants will go to point D to forage by pheromone different concentrations, and it will have $\frac{2n}{3}$ ants foraging in route A-B-D, repeat by times in this way, it will have most ants to walk in the relative shortest route.

According to above principle, it can deduce that in case ant colony has numerous foraging routes, it is also according to pheromone concentration changes that lets ant colony to more gather in relative shortest route till all ants pass through the shortest route.

ACO prototype system

Ant colony algorithm whole process includes ants' allocation and move, partial pheromone updating and volatilization, global pheromone updating such three parts. So the constructed ACO prototype system is up to above three parts.

Assume there are M ants randomly placed in N pieces of points, lie in i point k ant selects j point climbing probability is as formula (1)show

$$p^k(i,j)=\begin{cases}\dfrac{[\tau(i,j)]^{\alpha}\bullet[\eta(i,j)]^{\beta}}{\sum_{s\notin tabu_k}[\tau(i,s)]^{\alpha}\bullet[\eta(i,s)]^{\beta}} & \text{,if } j\notin tabu_k\\ 0 & \text{otherwise}\end{cases} \quad (1)$$

In Formula (1), $\tau(i,j)$ expresses edge (i,j) pheromone concentration, $\eta(i,j)=\frac{1}{d(i,j)}$ represents heuristic information, from which $d(i,j)$ represents edge (i,j) length, α,β respectively reflect pheromone and heuristic information relative importance degree, $tabu_k$ shows ant k crawl crossed points list.

When all ants get through all points and back to origin, pheromone updating is as formula (2) show

$$\tau_{ij}(t+n)=\rho\cdot\tau_{ij}(t)+\Delta\tau_{ij} \quad (2)$$

In Formula (2), ρ shows constant that less than 1, it shows information durability, $\Delta\tau_{ij}=\sum_{k=1}^{M}\Delta\tau_{ij}^{k}$ and $\Delta\tau_{ij}^{k}$ are as formula (3)show

$$\Delta\tau_{ij}^{k}=\begin{cases}\dfrac{Q}{L_k} & ij\in l_k\\ 0 & \text{otherwise}\end{cases} \quad (3)$$

In Formula (3), Q represents a constant, l_k shows the k ant walked route in this iteration, L_k is route length.

ACO algorithm prototype system designing flow is as Figure 2 show.

According to Figure 2, designing iteration process algorithm steps are as following.

Initialization, random place ants and establish $tabu$

FULL PAPER

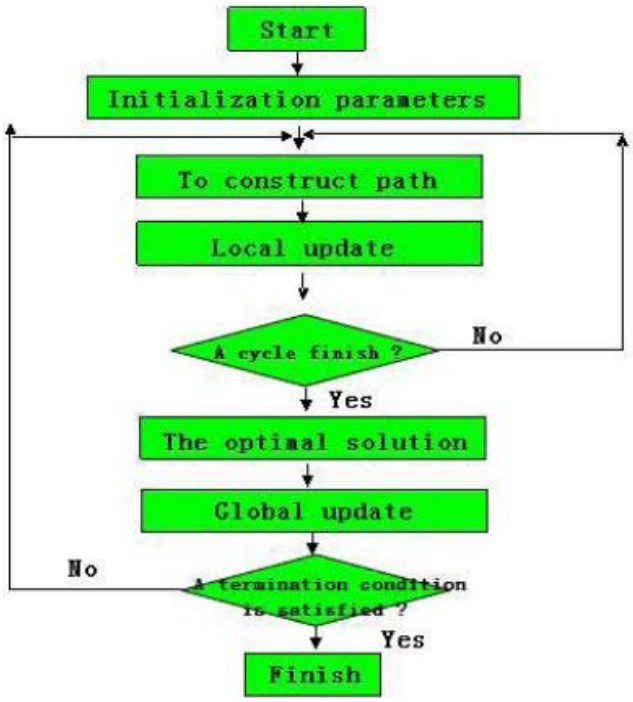

Figure 2 : Ant colony algorithm prototype system designing flow chart

for every ant, place initial node into tabu;

Iteration process:

$k = 1$

while $k \le It \quad Count \quad do$ (Implement iteration)

for $i = 1 \quad to \quad M \quad do$ (cycle M ants)

for $j = 1 \quad to \quad N - 1 \quad do$ (cycle N points)

Place node j into $tabu$, transfer ant into node j as initial node, repeat above steps

end for

end for

Calculate every ant route length, according to above steps, update all ant routes information;

Output result, end algorithm.

By above algorithm, it is clear that for large scale complex optimization problems, ACO needs longer searching time, due to individual ant movement randomness decides that though pheromone indirect coordinate can let ant colony to crawl towards optimal route, it is difficult to find a good route from lots of disordered routes in a shorter time, and ACO is easy to occur stagnation phenomenon and appear excessive early convergence, it cannot let individual ant to search optimal solution, based on above algorithm disadvantages, in the following, it provides improvement and optimization methods.

ACO prototype system optimization and improvement

In 2.2, before ant every step transferring node selection, it should consider all possible node set, its time complexity is $O(N^2)$, so most of system running time is applying and calculating optional nodes efficiency, if constructed node subset is not suitable, it will reduce algorithm searching effects, besides system is prone to appear premature convergence, these problems are mainly up to pheromone updated rules. For ACO prototype system optimization and improvement, it can start from following three aspects:

Start from system overall structure

Factors that need to consider except for ant colony system overall structure and organization pattern, it also should consider system each sub part information connection pattern, such as when it needs to expand the system to multi-population ant colony system and mixed ant colony system, it needs some ant colonies to joint solve problems, for information exchanging among ant colonies layers are optimized from system overall structure;

Start from system specific parameters setting and adjustment strategy

In ACO prototype system implementation process, it needs to set system parameters options and change patterns, in maximum and minimum ant colony algorithm, it is an improved method that restricts pheromone concentration, introduces variation mechanism into basic ACO, then makes variation improvement on parameters, such improvement can let algorithm to have faster solving speed and higher solving precise. Self-adaption ant colony algorithm and dynamical ant colony algorithm are one kind of balanced distributed ant colony algorithms, it can get balance between speeding up convergence and preventing prematurely, stagnation behaviors, it can also introduce random features into algorithm parameters setting, so that can form into random perturbation ant colony algorithm, and can adopt objective functions values-based pheromone allocation strategy, according to objective functions values, it can self-adapted adjust ant colony search directions, by numerous researchers' experiments, they show that these

improvement schemes can better optimize system functions;

Start from combining with other intelligent algorithms

ACO has advantages that is easily combining with other intelligent algorithms, so by combination, sometimes it can also get efficiency of complementing algorithms' advantages.

K MEANS CLUSTERING ALGORITHM

K means clustering algorithm is proposed by J.B.MacQueen in 1967, is a kind of typical classified cluster algorithm, is by far the most widely, mature cluster analysis method, due to the algorithm has advantages as simple fast and is fit for processing with big data set, by far it has already been widely applied in scientific research and industrial applications.

K means clustering algorithm thought

K means clustering algorithm is a kind of typical cluster-based hard cluster algorithm, the algorithm usually adopts error sum of squares function as optimized objective function, error sum of squares function is as formula (4) show

$$\mathbf{E}=\sum_{\mathbf{j}=1}^{\mathbf{K}}\sum_{\mathbf{x}\in\mathbf{C_j}}\left\|\mathbf{x}-\mathbf{m_j}\right\|^2 \quad (4)$$

In Formula (4) K shows amount of clustering, $C_j\left(j=1,2,\cdots,K\right)$ shows clustering the j cluster, x shows cluster C_j any data object, m_j shows cluster C_j mean value, E shows data samples and cluster gravity center difference degree squares sum and its size is up to K pieces of clustering central points, when E value gets smaller, its clustering result quality will be better, so the purpose of K means clustering algorithm is trying to find out clustering result when clustering criterion function E value arrives at minimum.

K means algorithm basic thought is firstly during data set that contains n pieces of data objects, it random selects K pieces of data objects as initial center, and then calculates every data object distance from each center, according to nearest neighborhood principle, it classifies all data objects into their nearest center represented cluster, subsequently respectively calculates new generated each cluster data object mean value as each cluster new center, compares new center and last time obtained center, if new center doesn't change, then it is algorithm convergence, outputs the result; If new center changes by comparing with last time center, and then it should make classification on all data objects according to new center till it meets algorithm convergence condition.

K means algorithm mainly flow description is as following:

Input: K value and contained n pieces of data objects data set X;

Output: K pieces of kinds that let error sum of squares arrive at minimum.

① From n pieces of data objects contained data set X, random select K pieces of data objects as initial clustering center;

② Respectively calculate data set every data object to each clustering center distance, according to nearest neighborhood principle, classify data one by one into its nearest clustering center represented cluster;

③ Update cluster center, it is respectively calculating each cluster all data objects means as each cluster new center, use new clustering center to calculate error sum of squares criterion function value;

④ Compare step ③ calculated E value with previous E value, if the two difference absolute value is not bigger than preestablished threshold value that judge clustering criterion function is convergent, move to step⑤, otherwise move to step②;

⑤ Output K pieces of clustering.

Above algorithm flow chart is as Figure 3 show.

K means algorithm analysis

1) K means clustering algorithm, thought it has advantages as simple process, fast and effective, fit for processing with lots of data set, itself still has some shortcoming and drawbacks, which restrict its application and development to a certain degree, it mainly reflects in following four aspects: K means clustering algorithm is very sensitive in initial central point selection, is prone to let algorithm to get caught in partial optimal solution. K means clustering algorithm K each initial clustering center is totally se-

FULL PAPER

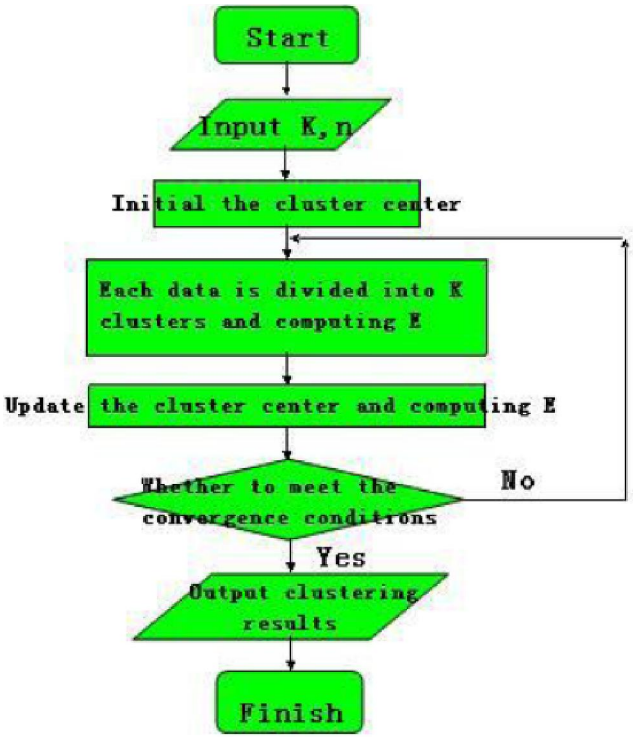

Figure 3 : K means clustering algorithm flow chart

lected according to random principles, which causes result fluctuation range to be bigger, stability to be worse, generally, K means clustering algorithm adopted clustering objective function is error sum of squares function, the function is a kind of non-convex function, usually the function simultaneously has many partial optimal values, and only has one global optimal value, K means clustering algorithm is gradient descent algorithm that based on error sum of squares function, one kind of initial center selection way represents one kind of searching way, the algorithm initial center is random selected, then there will have many initial center selection ways, and also have lots of algorithm searching ways, when the searching to error sum of squares criterion function value that doesn't reduce any more, the algorithm ends, and then along searching directions to search, when algorithm ends, obtained clustering result tends to be partial optimal saluting rather than global optimal solution, so make improvement on K means clustering algorithm initial central value selection sensitivity is one kind of important ways in the algorithm development;

2) K means clustering algorithm is relative sensitive to noisy data and outliers, form algorithm clustering process, it is clear that cluster center every time updating is obtained by calculating cluster all data objects mean values, while noisy data and outliers are usually far away from data samples space dense region, if add noisy data and outliers into cluster central updated calculation, it will surely generate important impacts on cluster center calculating, even may let newly calculated cluster center seriously to deviate from sample space dense region, use such cluster center to cluster that will generate extremely bad impacts on clustering result that is why K means clustering algorithm is sensitive to noisy data and outliers, reduce K means clustering algorithm sensitivity to noisy data and outliers is one kind of important ways in the algorithm development;

3) K means clustering algorithm generally is difficult to find out other shapes cluster except for sphere cluster, due to the algorithm is a kind of hard clustering algorithm, it generally adopts Euclidean distance as data samples similarities measuring methods, algorithm objective function usually adopts error sum of squares function, the objective function have better effects on handling with data set that data samples distribution is relative concentrate and each kind of data sample quantity has no big differences, but when handle with data set that each kind of data sample that quantity has great disparity, each boundary distinguishing is fuzzy and each kind of shapes differences are big, to minimize objective function, it may have a bigger class to be cut into several small classes and leads to clustering result becomes undesirable, so capacity of finding any shapes clustering plays certain roles in the algorithm development;

4) Cluster amount has an important impacts on K means clustering algorithm clustering result, if provided cluster amount K value is improper, then it will let clustering result to be undesirable, and in practical application, to a given data set, users cannot know how many clusters to be divided for a data can get most ideal clustering result in advance, so define cluster amount Kin advance adds burdens to users, and reasonable defining cluster amount K is also a direction that the algorithm should be improved.

ACO PROTOTYPE SYSTEM OPTIMIZATION-BASED CLUSTER ALGORITHM

Design thought and scheme

For K means clustering algorithm shortcomings, it can use method that combines with ACO to optimize K means clustering algorithm, due to K means clustering algorithm is prone to suffer initial information and noisy data as well as outliers' impacts that cause clustering result deviations, therefore the paper introduces ACO prototype system to make improvement on K means algorithm.

Firstly according to clustering problems definition variables, set N to be total of pattern samples, M is total amount of ants, V represents pattern samples number, K represent clustering centers number, S represents stored ants constructing solution (length is using N to express), NC represents iteration times, τ represents stored pheromone $N \times K$ matrix, w represents $N \times K$ matrix, its element w_{ij} is 0-1 variable and when $x_i \in Cluster_j$, $w_{ij} = 1$ others are 0, x represents pattern samples data information, x_{iv} represents sample i the v attribute value, m represents $K \times V$ clustering center matrix, m_{jv} represents class j all samples attributes v average value, p represents $N \times K$ matrix that is used to store ants selected clustering center probability, d represents $N \times K$ matrix used to store each sample to clustering center distance, $n_{ij} = \frac{1}{d_{ij}}$ represents sample to clustering center visibility.

Clustering objective function is as Formula (5)show

$$\mathbf{F(w,m)} = \sum_{j=1}^{K}\sum_{i=1}^{N}\sum_{v=1}^{V} \mathbf{w_{ij}} \left\| \mathbf{x_{iv}} - \mathbf{m_{jv}} \right\|^2 \quad \textbf{(5)}$$

In Formula (5), w_{ij} should meet Formula (6)

$$\begin{cases} \sum_{j=1}^{K} \mathbf{w_{ij}} = 1 \quad , \mathbf{i = 1,2,\cdots, N} \\ \sum_{j=1}^{K} \mathbf{w_{ij}} \geq 1 \quad , \mathbf{j = 1,2,\cdots, K} \end{cases} \quad \textbf{(6)}$$

And m_{jv} expression is as Formula (7)show

$$\mathbf{m_{jv}} = \frac{\sum_{i=1}^{N} \mathbf{w_{ij} x_{iv}}}{\sum_{i=1}^{N} \mathbf{w_{ij}}}, \mathbf{j = 1,2,\cdots, K, v = 1,2,\cdots, V} \quad \textbf{(7)}$$

After getting above definition, it can research on ants' constructing solution strategy, according to ant colony prototype system, ants moving selection strategy is proceeding according to formula (1), therefore it can change clustering rule into formula (8).

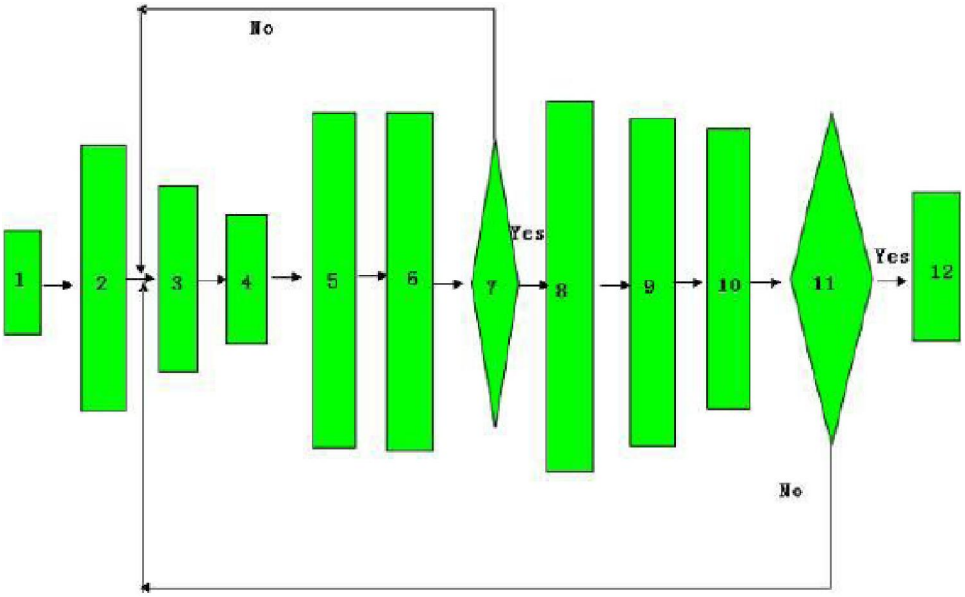

Figure 4 : ACO prototype system optimized K means clustering algorithm flow chart

FULL PAPER

TABLE 1 : Flow chart serial numbers' definitions table

No.	Definition
1	Start
2	Random generate K pieces of clustering centers
3	Set parameters, initialize ant colony
4	Ants' constructing solution S
5	Calculate solution S corresponding objective function E
6	Apply refuse to update rule
7	Successive cycle is over or not?
8	Calculate optimal solution corresponding K each clustering center
9	Calculate optimal solution objective function
10	Apply global updating rule
11	Meet ending conditions or not?
12	End

$$s = \begin{cases} \underset{S}{\text{select max}} \left\{ [\tau_{ij}]^{\alpha} [\eta_{ij}]^{\beta} \right\} & \text{, if } \quad q \le q_0 \\ \text{Otherwise} \end{cases} \tag{8}$$

Therefore, it can get ACO prototype system optimization-based K means clustering algorithm designing flow chart, as Figure 4 show.

In Figure 4, 1, 2, ……, 12 definition is as TABLE 1 show.

Apply random constructing solution method to make initialization that all pattern samples are randomly enclosed to affiliated clusters, every ant uses random constructing solution to generate initial cluster, apply formula (7) into calculating each ant cluster center, and according to objective function (5), calculate its objective function value, find out minimum objective function value's ant constructed solution, and take ant cluster center as initial cluster center, and then is parameter setting, main parameters in the algorithm is as TABLE

TABLE 2 : Symbol description

Symbol	Definition
α	Pheromone concentration importance parameter
β	Heuristic information importance parameter
Ψ	Partial information updating parameter
ρ	One parameter in control pheromone attenuation process
Q_0	Control threshold value that shift according to probability
M	Ants amount

TABLE 3 : Program annotation table

Line No.	Annotation
4	Each sample point to its clustering distance sum
5	Store ants constructed solution
6	Clustering center
7	Stack that stores K pieces of clusters
8	Each cluster samples number
9	Initialize ants
10	Pattern samples total amount
11	Clustering centers number
12	Store N pieces of samples data information
13	Random distribute solution
14	Update clustering center
15	Add all samples into corresponding clusters
16	Calculate objective function value
17	Output each cluster
18	Release ants

```
1  class CAnt
2  {
3  public:
4       double F;
5       int     *S;
6       CDoublePoint*Center;
7       CPoint**Cluster;
8       int     *Top;
9       void Init Ant(int i,int j,CPoint*);
11      int     K;
12      CPoint   *AllData;
13      void Random_solution();
14      void Update_center();
15      void Add_all_to Cluster();
16      void Fun();
17      void Out_cluster();
18      virtual~CAnt();
19 };
```

Figure 5 : Cant definition

```
1  class CACOKMeans
2  {
3  public:
4        double ALPHA;
5        double BETA;
6        double RHO;
7        double Q0;
8        int     M;
9        int     N;
10       int     K;
11       long double MinF;
12       CPoint*AllData;
13       double **TAU;
14       CDoublePoint*OutCenter;
15       CPoint    **OutCluster;
16       int     *OutTop;
17       CACOKMeans(double rho,double alpha,double beta,double q0,int m);
18       void InitData();
19       void Init_tau(CAnt*);
20       int Init_Center(CAnt*);
21       double ETA(int i,int j);
22       void BuildSolution(CAnt*,int i);
23       int    Min_F(CAnt*);
24       void Global_update_rule(CAnt*,int);
25       double Transition(int i,int j);
26       double Sum_Transition(int i);
27       int    Choose(int);
28       int    MaxP(int);
29       void Add_all_togood Cluster(CAnt*,int m);
30       void Add_all_togoodout Cluster();
31       void Out_good_Cluster();
32       void Update_good_center();
33       virtual~CACOKMeans();
34 };
```

Figure 6 : CACOKMeans definition

2 show, after fulfilling parameters setting, ant can construct solution according to our designed rules, after all ants completing successive cycling, it can find out optimal solution, and take optimal solution corresponding clustering center as current optimal clustering center, and calculate current optimal objective function value, with global pheromone updating, above process are always repeating and ends when it meets ending conditions.

According to ant colony clustering algorithm, it designs a CAnt, the class defines ants basic functions in K means clustering problems, as Figure 5 show.

In Figure 5, annotation from line four to line eighteen is as TABLE 3 show.

According to K means clustering algorithm features, combine with ant colony algorithm prototype system designing pattern, design a relative mean values ant colony clustering algorithm system class ACOKMeans, the design mainly includes each parameters setting, objective functions calculation, clustering center calculation, ants solution construction, global pheromone updating, ants transfer strategy, optimal solution selection and result output eight contents. K means ant colony clustering system class CACOKMeans definition is as Figure 6 show.

In Figure 6, annotation from line four to line thirty-three is as TABLE 4 show.

CONCLUSION

Introduce ACO prototype system into K means clustering algorithm can let K means clustering algorithm to get optimized;ACO prototype system has its own advantages and disadvantages, the paper points out improvement orientations in the way of statement;K means clustering algorithm has some shortcomings, it gets the algorithm improvement orientation by researching ; K means clustering algorithm has good ports with other algorithms, it can apply other algorithms to optimize itself, except for the paper explored cases, it can try to combine with other algorithms ;ACO is a kind of newest developed bio-inspired optimization algorithm, it can be well applied into data mining multiple fields; ACO is a kind of algorithm with a good development prospect.

REFERENCES

[1] Du Li-Feng, Niu Yong-Jie; Implementation of ant colony algorithm in MATLAB. Information Technology, **6**, 115-118 **(2011)**.

[2] Wu Qing-Hong, Zhang Ji-Hui, Xu Xin-He; An ant colony algorithm with mutation features. Journal of Computer Research and Development, **36(10)**, 1240-1245 **(1999)**.

[3] Ma Xiao-Li, Cai Xiu-Feng, Zhao Zhan-Fang, Wano Nan; Comparison and Parameter Optimization of Ant Colony System. Digital Community & Smart Home, **3**, 1324-1325 **(2007)**.

[4] Yan Chen Yang, Zhang You Peng, Xiong Wei Qing; Analysis on Performance of Novel Pheromone Trails Update Strategy in Ant Colony Optimization. Application Research of Computers, **24(7)**, 86-88 **(2007)**.

[5] Ji Yaofeng, Dang Pei, Guo Xiaobo; Study of Job Shop Scheduling Problem Based on an Improved Ant Colony Algorithm. Computer and Digital Engi-

neering, **39(1)**, 4-6, 52 **(2011)**.

[6] Zhu Kaiyan, Cheng Xuxiao; Dual Population Ant Colony Optimization Based on Pheromone Diffusion. Microcomputer Applications, **31(7)**, **(2010)**.

[7] Luo Jia-Wei, Chen Fei, Peng Dong-Hai; Hybrid behavior based ant colony pairwise alignment method. Computer Engineering and Applications, **45(11)**, 150-153 **(2009)**.

[8] Wan Bo, Lu Yu, Chen Liyun, He Ruibo; Optimization for Multiple Constraints Transportation Path Based on Mutated Ant Colony Algorithm. Microcomputer Applications, **32(7)**, 6-12 **(2011)**.