Research Combination Method of Parameter m based on Ant Colony Clustering

Qun Xiao1
Information technology department
QiongTai Teachers College
Haikou,China
xjq@qttc.deu.cn

Jieqing Xing 2
Information technology department
QiongTai Teachers College
Haikou,China
58628511@qq.com

Xiangyi Li3
Information technology department
QiongTai Teachers College
Haikou,China
xiaoqun@qttc.deu.cn

Abstract—With the development of data mining, a large number of new clustering algorithms, such as ant colony clustering algorithm, are emerging.The ant-based clustering parameter values in different circumstances often will has a significant impact on the performance and efficiency of the algorithm. In this paper, based on the research of ant colony clustering and K- method combination algorithm (KMAOC), we discuss and analyze the m value of the parameters of the ant colony. Respectively experimental values by several groups of experimental verification are provided to show the better proposal that a KMAOC ant algorithm parameters are better to configure the number of m.

Keywords-Clustering; Ant colony algorithm; K-means; Clustering combination;

I. INTRODUCTION

Clustering plays a very important role in scientific data detection, image processing, pattern recognition, medical diagnosis, computational biology, document retrieval, Web analysis and so on. It has become a very active research topic in data mining research. Classic clustering method includes hierarchical clustering algorithm and the method of dividing, such as K means algorithm, fuzzy C means algorithm, neural network method, as well as statistical methods. Recently, with the development of data mining, a large number of new clustering algorithms, such as ant colony clustering algorithm, are emerging. Ant colony algorithm is a kind of pioneering biological simulation algorithm, which has been widely used because of its excellent properties such as parallelism, robustness and so on. As the research history of ant colony algorithm is still very short, the application of ant colony algorithm is still less, because there are many problems to be further studied, such as the need to set many parameters, the difficult to set the parameters i.,the cost of algorithm convergence and computing. The ant colony clustering method is designed with the use of the information according to the ant foraging principle. If the value of its information begins from the 0 or the equivalent value, it often takes a long time for each path of the pheromone to be distinguished. Research shows that the combinatorial method of ant colony clustering algorithm and K-means algorithm can make up for these defectsii. At present, the research on the combination of ant colony algorithm is also carried out, such as Yang Yan and so on. In the literature they pointed out that in order to improve the quality of the clustering analysis, a combination of threshold and ant colony algorithm is carried outiii. According to this method, first the clustering algorithm is based on the threshold to generate the clustering center, and the number of clusters is determined. Then the transfer probability of the ant colony algorithm is introduced into the K- algorithm, and the results are optimized by two times. The two algorithms can be used to complement each other and avoid the limitations of a single algorithm. While in research literatureiv of Gao Shang and so on, some methods are proposed to overcome the difficulty of finding the common ground in the output results of different clustering algorithms and improve the clustering quality. The problem of cluster analysis is established. The advantages and disadvantages of K- mean algorithm is analyzed. The annealing algorithm and basic ant colony algorithm are simulated. While the results of the K- method are used as the initial value, the ant colony algorithm is improved. After comparing and testing, the effect of two kinds of hybrid ant colony algorithm is better, and the second hybrid method is the best of two. In this paper, based on the research of ant colony clustering and K- method combination algorithm (KMAOC), we discuss and analyze the m value of the parameters of the ant colony.

II. RESEARCH ON THE COMBINATION OF ANT COLONY CLUSTERING ALGORITHM

Because of the characteristics and problems of ant colony algorithm, the ant colony algorithm and other algorithms are used together to solve the existing problems in the practical application. One of the more common is the combination of genetic algorithm, immune algorithm, fuzzy neural network algorithm and so on. In the process of the combination of the algorithm, it shows that the combination algorithm is better

than the ant colony algorithm and other algorithms in the data clustering of one aspect. There are several aspects of the combination of algorithms.

A. Combined with genetic algorithm

From above we know the classical ant colony algorithm has very strong positive feedback ability. In the latter, it can accelerate the speed of evolution of the algorithm, which makes the algorithm converge quickly, but in the early stage of the algorithm, the information element is short, and the evolution speed is slow. The genetic algorithm has the ability to search the global search ability, but cannot use the feedback information of the system. Using the good ability of coupling of the ant colony algorithm, genetic algorithm to generate the information element initial distribution and ant colony algorithm to give the precision of the solution, the combination of ant colony algorithm and genetic algorithm can realize the complementary advantages of two algorithms.

Gong Daoxiong et al. proposed genetic algorithm based on ant colony algorithm. In this algorithm, the population of genetic algorithm corresponds to the ant colony, and the chromosome of the genetic algorithm is the ant colony algorithm. In this paper, the validity of the algorithm is verified by the experimental research on the problem of boundary and TSP. Take advantage of genetic algorithm to search the fast global search ability and the positive feedback of ant colony algorithm, Zou Yuanqiang et al. give out a new combination algorithm. From the simulation examples, the proposed algorithm has certain advantages over the standard ant colony clustering algorithm and fuzzy ant colony clustering algorithm in optimizing performance and time performance. And Xiao Hongfeng et al. also uses the genetic algorithm to solve the continuous function of the ant colony algorithm in the binary encoding form. According to the advantages and disadvantages of genetic algorithm and ant colony algorithm, two hybrid ant colony algorithm is constructed. The new algorithm of genetic algorithm can produce a new path of the ant colony algorithm, so as to achieve the fusion of genetic algorithm and ant colony algorithm. In terms of convergence rate, hit rate and accuracy, the correctness of the hybrid algorithm is also verified by using Rosenbrock and Shubert.

B. Combined with neural network algorithm

The extensive mapping ability of neural network and rapid global convergence of ant system can be obtained by combining ant system and neural network. Cao Xiaoxin et al. proposed a fuzzy neural network algorithm based on ant colony clustering. Neural network uses RBF network node structure. Clustering is used as a clustering algorithm of the two level structures and fuzzy C- means clustering (FCM) is used as a first level clustering algorithm for clustering two. The clustering method is used in the construction of fuzzy neural network. The simulation results show that it has the characteristics of parallel real time and clustering ability. Hong Bingrong et al. propose that the ant colony algorithm and neural network are combined to realize the control and identification of nonlinear model of inverted pendulum. The experimental results show that the algorithm can train the neural network with ant colony algorithm, which can be used in the performance of the fast global convergence of the neural network and ant colony algorithm..

C. Combined with immune algorithm

The effectiveness of the enhanced ant colony algorithm of the combination of ant colony algorithm and immune algorithm is using immune algorithm to obtain the key parameters of the ant colony algorithm.

To solve the contradiction between convergence and premature convergence and stagnation in the traditional ant colony algorithm, Chen Xu et al. proposed that the immune process is accomplished by two steps of vaccination and immune selection. Then, a novel immune ant algorithm is designed and applied to solving complex TSP problem. Numerical simulations show that the proposed algorithm can overcome the defect of the basic ant colony algorithm, and has better global search capability and stability. Base on the research status for Ad hoc network Qos routing, Lu Wei proposes a Qos routing algorithm based on immune ant colony algorithm. The algorithm uses artificial immune algorithm to quickly search for feasible solutions. In the process of the algorithm, ant colony algorithm is used to obtain the optimal feasible solution, and further improve the solving efficiency. The algorithm combines the advantages of artificial immune algorithm and ant colony algorithm for the two.

III. K-MEANS CLUSTERING ALGORITHM AND THE CLASSICAL ANT COLONY ALGORITHM

A. K-means algorithm is based on clustering algorithm, which is also the most common clustering algorithmv.

The algorithm calculates the center of each cluster. It is the average value of the object in the cluster. And it also is the new seed of clustering algorithm. The K-means algorithm is trying to find the minimum of the K of the square error function. While the difference between the cluster and the cluster is obvious, the K-means algorithm is effective. While processing large data sets, it has good stability and high efficiency. While the difference is not obvious, the K-means algorithm's effect is poor. The disadvantage of this algorithm is that the number of clusters must be generated in advance

B. Classical ant colony clustering algorithm.

Characteristics of ant colony algorithmvi:

The ant colony algorithm has a strong ability to find better solutions. Because the algorithm uses the principle of positive feedback, it accelerates the process of evolution, and is not easy to fall into local optimal solution;

The ant colony algorithm has a strong parallelism. It is beneficial to the better solution that the information exchanges and transfer between the individual. The single individual is easy to converge to local optimum, and many individuals can converge to a subset of the solution space quickly. It is beneficial to further explore the solution space, and thus find a better solution.

Problems in ant colony clustering algorithmvii:

The efficiency of the algorithm: the ant colony clustering algorithm to process is relatively slow. Especially in the initial stage, because the system parameters change very

slowly, the whole calculation process will take a long time. In the clustering analysis based on the principle of ant colony foraging, the initialization of pheromone on each path is usually assigned to the same constant C (usually 0) in order to simplify the operation. Therefore, the value of the pheromone is from the same constant C, the information on each path to clearly distinguish generally takes a long time. The KMAOC algorithm can solve this problem well.

The choice of initial value: the choice of initial value has a great influence on the final result of the clustering. In the classical ant colony algorithm, the initial parameters are generally lack of the known experience and the initial parameters are determined with a great blindness. In the clustering method, the selection of (α，β，m) has a great influence on the efficiency of the algorithm and the results of clustering. Choosing improper will affect the efficiency and effectiveness of the algorithm. We can try different methods to avoid the local optimal algorithm. In this paper, we will focus on the effect of m parameters on the algorithm.

IV. STUDY ON PARAMETERS OF BASIC ANT COLONY ALGORITHM SUCH AS M

The parameters in the ant colony algorithm can have a significant impact on the performance and efficiency of the algorithm. Ant colony algorithm is an adaptive, positive feedback and distributed simulation optimization algorithms, which have some advantages in solving complex combinatorial optimization problems. The good combination of alpha and beta, rho, m can result a good quality and good stability. But if the parameters of the ant colony algorithm are not properly selected, the ant colony algorithm converges quickly to the local optimum or the slow convergence, which has a great influence on the performance of the algorithm.

In the literature of Zhang Jiehui, in order to verify this idea that the number of ants is not the more the better, they chose the wpbc data set as the experimental data. Fig .1isthe experimental results of the choice of the number of different ants, and it denied the assumption that the number of ants is the more the better. In the experiment, the parameters of the experimental data m=5.

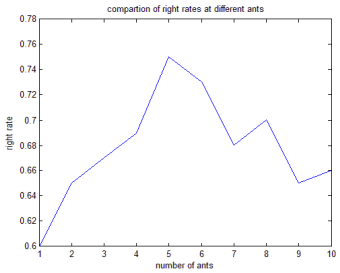

Figure 1. the classification performance of different antsviii

By the research, the selection of the parameters m is not the bigger the better, but it cannot be taken too small. In the TSP experiment, the number of ants is more conducive to the solution, but the number of iterations calculation will also change. According to the experiment, the number of ants is generally set to the number of cities in the number of 1/2 to 2/3 which is more suitable .

V. ANT COLONY CLUSTERING ALGORITHM BASED ON K-MEANS

The basic ant colony algorithm (AOC) is introduced into K-Means as a per computation algorithm for solving clustering problem. The steps of KMAOC algorithm are as follow:

Step1 Optional K initial clustering center: C1, C3, C2... CK;

Step2 One by one, each data of the set {X} is assigned to one of the K cluster centers Cj according to the principle of minimum distance;

Step3 Calculate the new clustering center Cj'(j=1,2, …,k), $C_j' = \frac{1}{N}\sum_{X \in Sj} X$, among Nj is the number of J clusters Sj contains;

Step4 If $C_j' \neq C_j$ (j=1,2, …,k) and not classify to the setting clustering effect threshold γ fast or reach to the specified number of times then go to Step2;

Step5 $nc \leftarrow 0$ (nc is the Cycle times),the clustering center calculated by K-means algorithm is Cj(j=1,2, …,k), work out τij(0)(i,j=1,2, …,N) according to the initial value of Xi correspondent each sample data Cj. The values of Q, P (pheromone persistence) and n (number of ants) are given. Distribution plan is given randomly.

Step6 the next node is selected by the transfer probability PIJ (T) of every ant.

Step7 Calculate the new clustering center, calculate the distance between each sample data to the new cluster center, modify the pheromone strength ij (T) according to the ant colony clustering formula.

Step8 if nc< the predetermined number of iterations and non-degenerate behavior (find the same solution), then output best solution, or go to Step 6.

VI. ALGORITHM TEST

In this paper, we use external evaluation method F-measure ix and the total running time to evaluate and compare the proposed clustering algorithm with the K means algorithm and the standard ant colony algorithm. The value of F-measure is in [0，1]. The value is better and closer to 1. Experimental data are from wine data set from the UCI machine learning database. The data sets have their own classification, which can be used for the evaluation of clustering performance. The performance evaluation of clustering algorithms usually uses two kinds of external and internal, which depending on the prior knowledge of the data set.

Table 1 the description of data base

Fig .2shows the values of Runtime①、F –measure in 8 different values of KMAOC algorithm parameters m.

database name	Data size	Attribute Number	Classification number
Wine	178	13	3

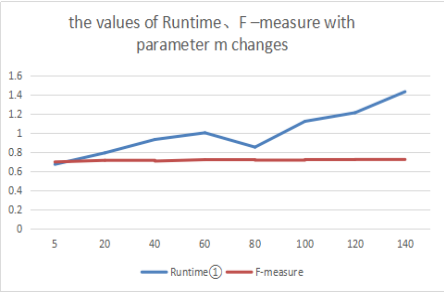

Figure 2. the values of Runtime①、F –measure with parameter m changes

Note① running on the same computer with KMAOC algorithm parameter values: α=1，β=5，ρ=0.99，Q=80，the count of ants m=60. Iteration number NC is 200 times. To its standard time, the value is 1. The value 0f Runtime is equal to the actual running time of the algorithm parameters m divided by the actual running time when KMAOC algorithm parameter m is 60.

Experimental results show: the number of iterations and other α，β，ρ parameters under the condition of the same, Runtime and F-measure values are different when KMAOC algorithm parameter m's value is different. But by comparison, The F-measure and Runtime values are best when the m value of the data set is 60. If the m value is reduced, the convergence time is reduced, but the F-measure is also smaller. If the m value is large, the F-measure value is increased, but the convergence time is increased. By experiment, we know the m value is determined by the practical application background, and the Runtime and F-measure values should be chosen to be the ideal situation.

VII. CONCLUSION

In this paper, the values of the parameters of the M are discussed when K-Means is used as the ant colony algorithm in the preprocessing process. When Parameter values are proposed according to different data objects, the parameter m will get a good range. Because the KMAOC algorithm can avoid the shortcomings of the classical ant colony algorithm, whose initial stage of the learning is slow, so compared with the classical ant colony algorithm parameter m, KMAOC algorithm parameter m can be obtained relatively large. It is suggested that the number of M should be 1.6 to 2 times of the number of clusters.

REFERENCES

[1] Xia Yamei, Bo Cheng, Chen Junliang, Meng Xiangwu, Beijing 100022). Based on improved ant colony algorithm of service composition optimization [J]. Chinese Journal of computers, 2012, 02:2270-2281.

[2] Wu Huafeng, Hsin Chiang Chen, Mao Qi Huang, Qian Nan Zhang, Zhang Shou Chun. Based on natural selection strategy of ant colony algorithm to solve TSP problem. Journal of communication, 2013,04:165-170.

[3] J Handl, J Knowles, M Dorrigo. On the performance of ant-based clustering [C].In: Proc of the 3rd Int. Conf. on Hybrid Intelligent Systems, IOS Press, Australia, 2003-12.

[4] Han Jiawei, KAMBER M. Data mining: concepts and techniques [M]. Beijing: Machinery Industry Press, 2001 : 223-251.

[5] Zeng Zhou et al. Uncertainty analysis of ant colony algorithm [J], Computer application, 2004; 10:136-138.

[6] Chen Yingxian, Improvement of the method of adjacent objects in the ant colony clustering algorithm[J], Computer engineering and Application, 2009;45(18):144-145.

[7] Xing Jieqing, Research on combination method of ant colony clustering [J], Computer engineering and Application, 2009;45(18):146-148.

[8] Yang Yan, Clustering method based on the combination of threshold and ant colony algorithm [J], Journal of Southwest Jiao Tong University, 2006; 41(6):719-723.

[9] Wang Jun, ant colony algorithm to solve the parameters of [J] set up the research TSP, science and technology and engineering, 2007; 7 (17): 4501-4505.

[10] Huang Han, etc. The convergence rate of ant colony algorithm for the analysis of [J].. 2007 (08).

[11] Gao Shang, a new clustering method based on hybrid ant colony algorithm [J], Microelectronics and computer, 2006;23 (12): 38-42.

[12] C.Rijsbergen . Information Retrieval, 2nd edition [M], Butterworths, London, UK, 1979,99.

[13] Zhang Jiehui, etc. Combined feature selection algorithm based on adaptive ant colony algorithm [J], system simulation, 2009, 3 (6): 1605-1610.

[14] Xu Ning, etc., several modern optimization algorithms of [J], system engineering and electronic technology, 2002; 24 (12): 101-105.

[15] Jiang Yan, etc. Parameter analysis of the ant colony algorithm [J], computer engineering and application, 2007, 43 (20): 31-37.

..