International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.9, (2016), pp.429-438
http://dx.doi.org/10.14257/ijsip.2016.9.9.39

Research on Application of Ant Colony Clustering Algorithm to Xinjiang Ethnic Groups' Evaluation System of Chinese Teaching

Tao Liu[1] and Xia Peng[2]

[1]*College of Literature and Art, Shihezi University, Shihezi, Xinjiang, China, 832003*
[2]*Mechanical and Electric Engineering College, Shihezi University, Shihezi, Xinjiang, China, 832003*
ltpx123@163.com, Penxia521@163.com

Abstract

With the continuous development of science and technology, computer technology and clustering technology of DM have become a part of the process of data analysis in this new era. How to use data mining technology to have better service for teaching evaluation field has been a research focus in the related research institutions at home and abroad.

This paper has a research on application of ant colony clustering algorithm to Xinjiang ethnic groups' evaluation system of Chinese teaching based on above. After briefly introducing the related theory and technology, this paper uses exploratory research method and survey method and designs evaluation system of Chinese teaching based on ant colony algorithm. Through the intensive study of a variety of data mining algorithm, this paper designs evaluation model of Chinese teaching using AHP. Then this paper gets data index through related survey and measures results after introducing ant colony clustering algorithm in evaluation system of Chinese teaching. The results show that application of ant colony clustering algorithm to Xinjiang ethnic groups' evaluation system of Chinese teaching is more effective for teaching evaluation and improves the stability of the evaluation system and makes the system professional and provides effective reference experience to introduce ant colony clustering algorithm to other areas.

Keywords: *ant colony clustering algorithm; Chinese teaching; teaching evaluation system*

1. Introduction

Computer technology has become an important technology in the development of human society in the new century, information technology which the core is computer technology and data mining technology not only has played a large role in people's life and work and has caused wide public concern in the field of education. Research on application of ant colony clustering algorithm to Xinjiang ethnic groups' evaluation system of Chinese teaching and collection and statistical analysis for the data of evaluation system of Chinese teaching are built on a foundation with a large amount of accurate user data

Since 1993, Harvard University will be carried out summer training program for Chinese learners and Chinese language teachers every summer to evaluate the Chinese language teaching activities [1]. There are several facets of the main evaluation content including Chinese pronunciation, vocabulary of Chinese language, the form of Chinese and Chinese articles and research and so on. There are also many researches on evaluation system of Chinese teaching in China. Many research institutions and related scholars have done a lot of fruitful researches. Cheng Guanglei in Fudan University put forward a new evaluation index in 2002 which is the most comprehensive and specific ethnic Chinese

ISSN: 2005-4254 IJSIP

teaching evaluation system by far and the teaching evaluation system can cover the different teaching object oriented Chinese teacher [2]. Zhang Baoling put forward that with the increasingly competition of education market and society raised more and more quality requirements for Chinese education teaching, the only way is setting up new education teaching evaluation system which can adapt to the demand of the modern society and promote the development of the ethnic Chinese teaching and establish a preliminary model of reform [3]. Yang Huiyuan of Beijing Language and Culture University detailed that how to apply the evaluation system of Chinese teaching and the positive role of the system for Chinese language teachers and putting forward the relevant evaluation principles [4].

At this stage, Chinese teaching evaluation system of ethnic minorities mainly has the disadvantages that scientific nature and credibility are low and it lacks of cross-cultural sensitivity. This paper mainly adopts the exploratory research and uses ant colony clustering algorithm and AHP and designs evaluation system of Chinese teaching based on ant colony algorithm and builds the relevant evaluation model.

The rest of this article is organized as follows. Chapter 2, introduces ant colony clustering algorithm and AHP. Chapter 3, studies the ant colony clustering algorithm and designs ethnic Chinese teaching evaluation system based on ant colony clustering algorithm. Chapter 4 analyzes and evaluates teaching evaluation system through the investigation. Chapter 5 is conclusion.

2. State of the Art

2.1. Ant Colony Clustering Algorithm

In recent years, the further study of ant colony algorithm can be used to analysis of clustering problem and people can get better clustering results through the use of ant colony algorithm to solve clustering problems [5]. Scientists found many colonies clustering algorithm model based on ant behavior of ant by simulating various behavior of ant colony in nature. These models can be divided into two kinds: ant clustering model based on the ant larvae classification and body disposal and ant clustering model based on ants foraging [6].

Ant clustering model based on the ant larvae classification is to simulate the behavior that worker-priests move the dead ants to the ant heap. At first, the ant heap is small and can constantly attract worker-priests to carry dead ants. With the increasing of concentration of moving information in the process of moving to the ant heap, it can attract more worker-priests to participate in moving. This is a positive feedback process and it can realize clustering through the process of positive feedback that the ant heap is bigger and bigger [7].

Ant clustering model that is on the basis of the principle of ants foraging simulates ants foraging behavior. It makes the data objects as different ants and the food source is the center of the cluster. The ant group can always find the nearest way to the path of food in the absence of any hint. Even if the surrounding environment changes in the process, the ants also can give rapid responses and find the shortest path again [8].

Clustering algorithm based on ant colony algorithm has a very big difference with the classical clustering algorithm. Ant colony algorithm is based on bionics and this algorithm has a unique advantage of bionics. It don't need to tip information algorithm and can be applied in different areas [9].

2.2 AHP

Analytic Hierarchy Process has a short name, AHP. It was put forward by operational research expert, Thomas L.Saaty of University of Pittsburgh [10]. AHP can quantify policymakers experienced judgments and provide the basis of decision-making for

decision-makers. In general, AHP can give complex problems clear hierarchy models and give quantitative descriptions combined with subjective judgments from mathematical algorithm and provide objective basis to solve problems [11]. AHP method is introduced to our country in the 1990 s. With the combination of qualitative and quantitative characteristics to deal with problems, it has a great development in all areas of our country. Using AHP method to calculate index's weight coefficient has some basic methods. They are as follows. According to the essence of the problem and the purpose of the problems and the relationships between these factors, we let the problem hierarchical and be decomposed into different form factors in order to construct a multi-level analysis structure model; We determine the judgment standard for level factors; By opinions of experts and combining with all kinds of comprehensive analysis, we construct judgment matrix; We have a weight translation and conclude the index at the bottom; We calculate the best solutions [12]

3. Methodology

3.1. Ant Colony Clustering Algorithm Based on Relative Canberra Distance

Canberra distance is put forward by Lance and Williams in 1967. It is a new distance definition formula and computes the sum of the subtle differences between the two vector objects. Weight difference between the two vector objects is between 0 and 1. The distance is very sensitive about tiny changes between two vectors [13]. Canberra distance is similar to the Manhattan distance. We can regard it as a kind of weighted Manhattan distance. Now Canberra distance is a measurement standard to compare and test score list [14].

Concrete definition is following:

$$d(X,Y)=\sum_{i=1}^{n}\frac{|x_i-y_i|}{|x_i|+|y_i|} \tag{1}$$

Of which $X=(x_1,x_2,...,x_n)$ and $Y=(y_1,y_2,...,y_n)$ is a n d vector.

And this paper uses the relative Canberra distance which combines the ideas of Euclidean distance and forms a new distance formula of similarity. It magnifies the function of longest difference property in distance measure between two vectors. It have the both nature of the Euclidean distance and Canberra distance [15].

Concrete definition of relative Canberra distance is following:

$$d_{canberra}(X,Y)=\sqrt{\sum_{i=1}^{n}(\frac{x_i-y_i}{x_i})^2} \tag{2}$$

Of which $X=(x_1,x_2,...,x_n)$ and $Y=(y_1,y_2,...,y_n)$ is a n d vector.

The formula of relative Canberra distance has the following properties: The greater function value is, the greater the distance between two data objects is and the less the similarity is. On the other hand, the less the function value is, the shorter distance between two data objects is and the more the similarity is.

3.2. Comparison of Measure of Similarity of Vectors

Generally there are two ways of measure of similarity of vectors: measurement method based on distance and measurement based on the similarity function. Measurement method based on distance has a lot, the typical methods are Minkowski distance, the

Manhattan distance and Euclidean distance. Measurement based on the similarity function has angle cosine and correlation coefficient method.

We establish $X=(x_1,x_2,...,x_n)$ and $Y=(y_1,y_2,...,y_n)$ as P degree vector. The similarity between them can be measured by using a variety of forms, concrete definition is as following.

Minkowski distance:

$$d(x,y)=\sqrt[q]{(|x_1-y_1|^q+|x_2-y_2|^q+\cdots+|x_p-y_p|^q)} \tag{3}$$

When q=1 in Minkowski distance, it becomes Manhattan distance:

$$d(x,y)=|x_1-y_1|+|x_2-y_2|+\cdots+|x_p-y_p| \tag{4}$$

When q=2 in Minkowski distance, it becomes Euclidean distance:

$$d(x,y)=\sqrt{(|x_1-y_1|^2+|x_2-y_2|^2+\cdots+|x_p-y_p|^2)} \tag{5}$$

Also we can use weighted distance, according to the degree of importance of attributes in actual application, we give the weights. This can better describe the data object and the calculation results are more accurate, such as weighted Euclidean distance:

$$d(x,y)=\sqrt{(\omega_1|x_1-y_1|^2+\omega_2|x_2-y_2|^2+\cdots+\omega_p|x_p-y_p|^2)} \tag{6}$$

A similarity measure is the basis of cluster analysis which affects the quality of the clustering results. The data set having numerical attributes compared to Canberra distance has two advantages. One is that it won't be affected by the measuring unit compared to Canberra distance; the other is that it has good robustness for loss data sets.

3.3. The Establishment of the Database Model

Information system modeling method can be divided into process oriented modeling, data oriented modeling, information oriented modeling, decision oriented modeling and object oriented modeling.

Data oriented modeling make input and output the most important. Therefore, we should first define the data structure and process modules are derived from the data structure that is function with data.

Model database is not in the real world, but the user model modeling about its business environment. Appropriate criteria to judge a data model is to see if it is in accordance with the user model. Arguing which one is suitable for the realistic world most is meaningless.

College teachers' data model of teaching evaluation can establish entity relationship diagram from three perspectives that are selection, evaluation and data processing. The first one is the student course selection. Student course selection is a process of operation and every student can choose courses according to their own plans. Obviously the relationship between students in course selection and course is a many-to-many relationship. The second one is the teaching evaluation. Teaching evaluation of teachers in terms of teachers is an individual and it is an integral to the whole school. In the establishment of teaching evaluation' entity relationship diagrams, it is usually a one-to-many relationship. Only the relationship between questionnaire and teachers is usually a one-to-one relationship. The last is the data statistical analysis, this is a one-to-one relationship. Entity relationship' data statistical analysis model is in figure 1.

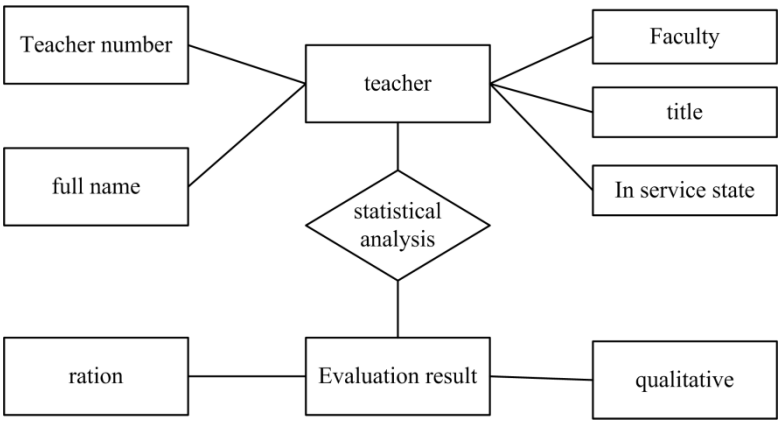

Figure 1. E-R Model of Data Analysis

After completion of the above three local model diagram design, we should precede a few local model diagram for integration, the method is based on the principles of sorting all the points. We give the front points the first integration and integrate the similar entities and attributes to merge. As shown in Figure 2, it is the evaluation system of Chinese.

Figure 2. Chinese Teaching Evaluation System

4. Result Analysis and Discussion

4.1. Results of Comparative Tests of Clustering Algorithm

The experimental data is Iris data set downloaded from UCI machine learning repository. The data set will be described in detail in the fifth chapter; here we only make a contrast experiment on classification effect of the Iris data set.

Through the compared experiment results in Iris data sets, it can be seen that the probability of this class has improved after using relative Canberra distance and the specific experimental results are shown in Table 1:

Table 1. Two Distance Region Division Results

	Class 1 distinguish the same kind of probability	Class 2 distinguish the same kind of probability	Class 3 distinguish the same kind of probability
Euclidean distance quality	0.9996	0.7473	0.7136
Relative Canberra distance quality	1.0000	0.8588	0.7808

We use figure 3, shows distinguishing the same kind of probability between the Euclidean distance and relative Canberra distance better. It can be seen from the figure that when we in distinguish the difference between the second and third class data using the mixed measure, its discrimination improves more than the traditional Euclidean distance's.

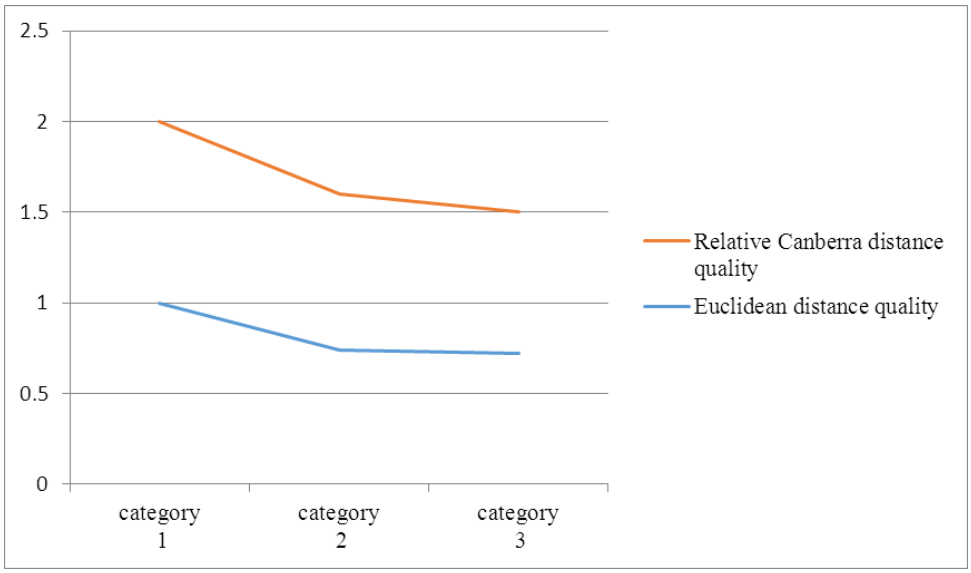

Figure 3. Comparison of Two Kinds of Probability of Distance

4.2. Experimental Result of the Optimal Parameters

First we have experiment in view of the optimal values of the input parameters in the traditional ant colony clustering algorithm. After repeated experiments and analysis, we take different values for each input parameters and calculate the corresponding F-measure value evaluated by clustering results. The greater the F-measure value is, the better clustering results are.

Because randomness of ant colony clustering algorithm is very big, so each occurrence of each input parameters has 10 independent experiment and finally we will average F-measure value and get relation schema between r1 and f shown in Figure 5 and relation schema between r2 and f shown in Figure 6.

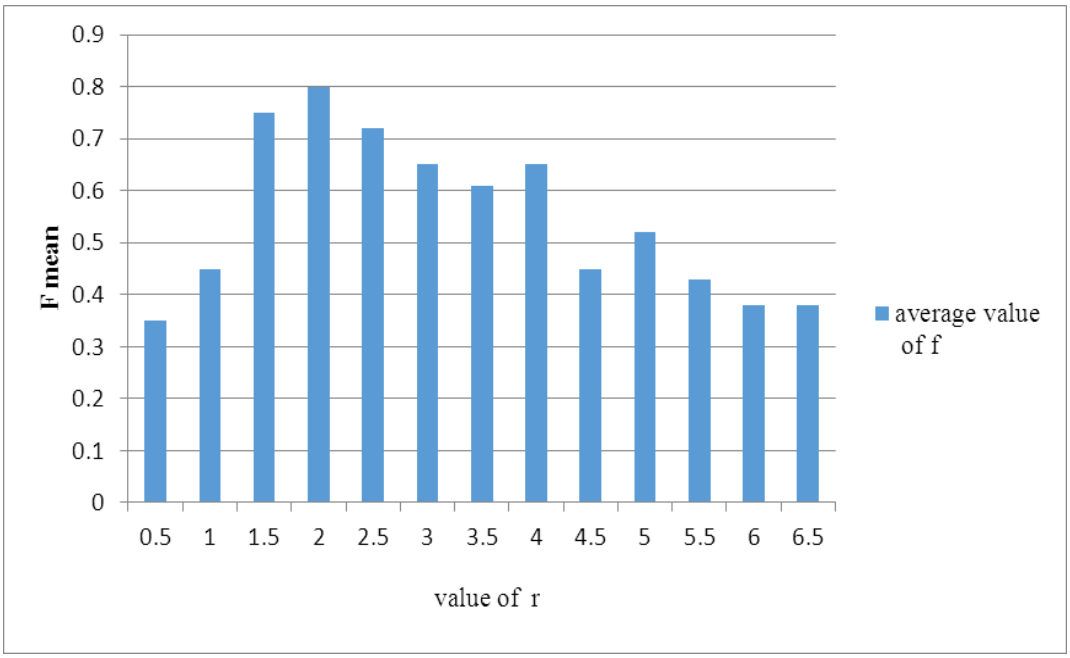

Figure 4. Relationship Diagram of Parameter r and f

From figure 4, we can see that the parameter r's value has a great influence with the result of clustering. When r' s value is equaled to 2, f value is maximum. That is to say, this point is the best clustering result so that the optimal value of r is 2.0.

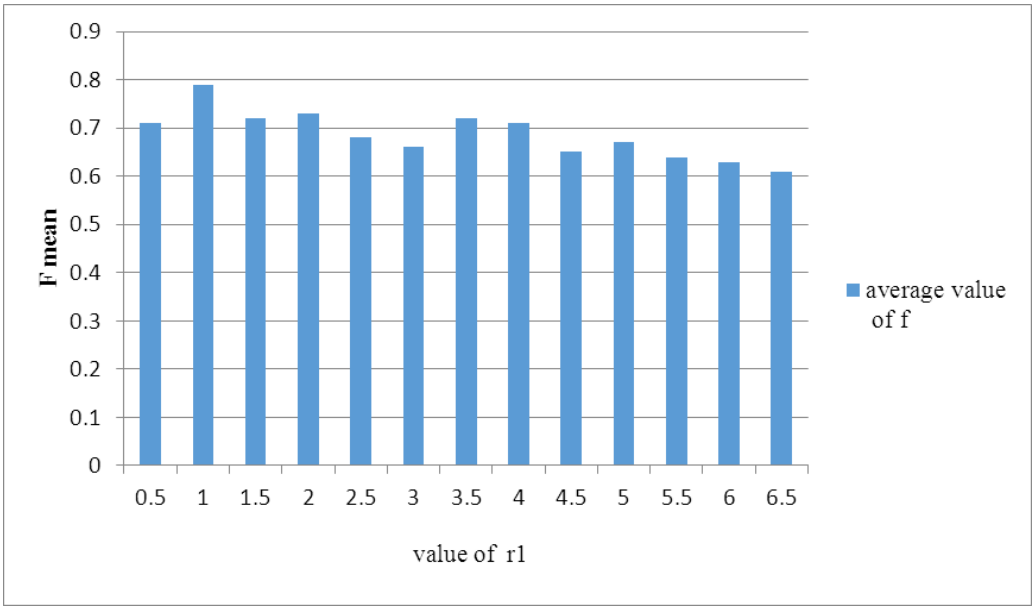

Figure 5. Relationship Diagram of Parameter r1 and f

From figure 5, we can see that the parameter r1' s value don't have a great influence with the result of clustering. When r1' s value is equaled to 0.2, f value is maximum.

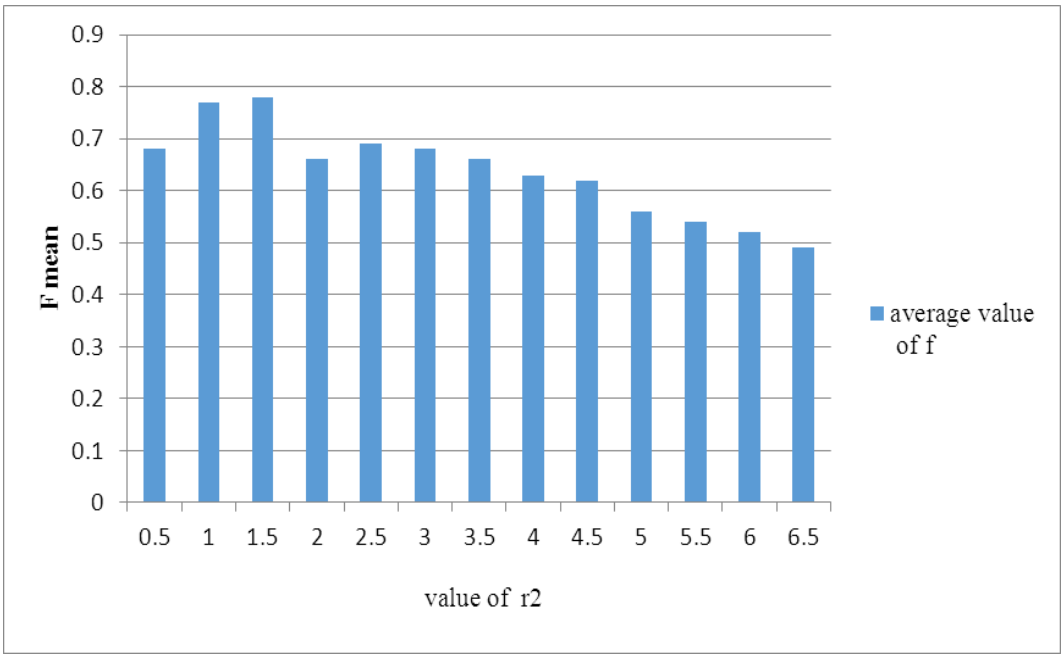

Figure 6. Relationship Diagram of Parameter r2 and f

From Figure 6, we can see that the greater the r2 value is, the worse clustering results are. From the Figure, the optimal value of r2 is 0.3.

4.3. Analysis and Assess of Xinjiang Ethnic Groups' Chinese Teaching

The teachers and students on the importance of the evaluation standard of Chinese language teachers have significant differences. for example, teachers focus on an evaluation of the project, but students don't think it is very important. Questionnaire is based on this assumption. By investigating the teachers and students' thinking about sorting the importance of the evaluation standard for Chinese language teachers and testing whether the hypothesis is correct, we can show that different evaluation objects' evaluation standards need to be adjusted.

We make the existing investigation to teaching evaluation of the Chinese language teachers of Xinjiang Normal University as the foundation and sum up Chinese teaching evaluation system from eight perspectives. We set up the evaluation items and sort options by 12 investigators. Importance degree is decreased by the set 1, 2, 3... And finally we determine the order according to how many the number of each choice. We choose 50 foreign language teachers from Xinjiang Normal University and Xinjiang University and 50 overseas students from School of Intercultural Studies of Xinjiang University and distribute 100 questionnaires and recycle effective 90 questionnaires 90 including 50 teachers' questionnaires and students' 40 questionnaires.

Table 2. Teachers and Students Input Questionnaire Statistics

	Teaching objectives	Teaching method	Content of courses	Languag e expression	Teaching attitude
teacher	21	11	10	8	7
student	19.1	12	10.7	8.2	7.1

By above tables, teachers and students sort the the importance of the evaluation project with basic coincidence and no too much difference on the whole. There have a few differences only in some detail aspects such as teachers' teaching manners, teaching blackboard writing, mandarin, volume, speed, and homework assignments. The remaining several projects are basically identical.

5. Conclusion

Under the era with rapid development of science and technology, in order to better use of data mining and technology to serve for the education teaching and promote the computer data mining research, this paper has detailed analysis the defects of the traditional evaluation system of Chinese ethnic minority and deep research on application of ant colony clustering algorithm to Xinjiang ethnic groups' evaluation system of Chinese teaching at the same time. On the basis of thorough understanding about the technology, we aim at putting forward a new intelligent optimization algorithm——ant colony clustering algorithm, based on how to reflect the teacher's teaching quality standard , ant colony clustering algorithm authentic and we systematively solve the problems in the process of teaching evaluation in Chinese ethnic. Through the investigation and experiment and analysis of the experimental data, we can prove that the algorithm has the obvious advantages in efficiency and accuracy compared to other data mining algorithm. Based on the design of ant colony clustering algorithm to Xinjiang ethnic groups' evaluation system of Chinese teaching, it has a great help to enhance the effectiveness of the teaching evaluation and the Xinjiang ethnic groups' Chinese language teaching. The development of technology, of course, will have more and more high the requirements for data analysis, so the research still has a lot of need to improve. It needs to have a further study.

Acknowledgments

Social science fund of Xinjiang production and Construction Corps "A survey study on the popularization and use of the national common language in Xinjiang" (Item Number: 13QN04); Xinjiang Education Science Program "Xinjiang Uygur Autonomous Region family language education and the present situation of the school of Mandarin promotion survey" (Item Number: 145030).

References

[1] G. X. Zhang and L. M. Li, "Chinese language teaching in the UK: Present and future", Language Learning Journal, vol. 38, no. 1, **(2010)**, pp. 87-97.
[2] C. U. I. Xiliang, "On the" Three Concerns" of Teaching Chinese as a Second Language", Chinese Teaching in the World, vol. 1, **(2010)**, pp. 013.
[3] Y. G. Butler, "The implementation of communicative and task-based language teaching in the Asia-Pacific region", Annual Review of Applied Linguistics, vol. 31, **(2011)**, pp. 36-57.
[4] M. Cargill, P. O'Connor and Y. Li, "Educating Chinese scientists to write for international journals: Addressing the divide between science and technology education and English language teaching", English for Specific Purposes, vol. 31, no. 1, **(2012)**, pp. 60-69.
[5] F. Gao, "Teacher identity, teaching vision, and Chinese language education for South Asian students in Hong Kong", Teachers and Teaching, **(2012)**, vol. 18, no. 1, pp. 89-99.
[6] W. Shouren, "Some Thoughts on College English Teaching in China", Foreign Language Learning Theory and Practice, vol. 1, **(2011)**, pp. 003.
[7] D. Karaboga and C. Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm", Applied soft computing, vol. 11, no. 1, **(2011)**, pp. 652-657.
[8] C. Zhang, D. Ouyang and J. Ning, "An artificial bee colony approach for clustering", Expert Systems with Applications, vol. 37, no. 7, **(2010)**, pp. 4761-4767.
[9] V. Selvi and D. R. Umarani, "Comparative analysis of ant colony and particle swarm optimization techniques", International Journal of Computer Applications , vol. 5, no. 4, **(2010)**, pp. 0975–8887.

[10] B. C. Mohan and R. Baskaran, "A survey: Ant Colony Optimization based recent research and implementation on several engineering domain", Expert Systems with Applications, vol. 39, no. 4, **(2012)**, pp. 4618-4627.
[11] J. Yang, M. Xu and W. Zhao, "A multipath routing protocol based on clustering and ant colony optimization for wireless sensor networks", Sensors, vol. 10, no. 5, **(2010)**, pp. 4521-4540.
[12] X Liu and H. Fu, "An effective clustering algorithm with ant colony", Journal of Computers, vol. 5, no. 4, **(2010)**, pp. 598-605.
[13] C. Lin, G. Wu and F. Xia, "Energy efficient ant colony algorithms for data aggregation in wireless sensor networks", Journal of Computer and System Sciences, vol. 78, no. 6, **(2012)**, pp. 1686-1702.
[14] D. Karaboga and C. Ozturk, "Fuzzy clustering with artificial bee colony algorithm", Scientific research and Essays, vol. 5, no. 14, **(2010)**, pp. 1899-1902.
[15] O. A. M. Jafar and R. Sivakumar, "Ant-based clustering algorithms: A brief survey", International Journal of Computer Theory and Engineering, vol. 2, no. 5, **(2010)**, pp. 787.

Authors

Tao Liu, he was born in 1978 in Hunan Province. Now he has got the Master Degree of Art, and now he is a lecturer of College of Literature and Art in Shihezi University.

Xia Peng, she was born in 1980 in Sichuan Province. Now she has got the PhD of Engineering. Now, she is a lecturer of Mechanical and Electric Engineering College in Shihezi University.

