Historical Seismic Intensity Determination Method Based on Ant K-means clustering Algorithm

Wenming Zhu

School of Computer & Communication Engineering
University of Science and Technology Beijing
Beijing, 100083China
E-mail: zf10053@126.com

Abstract—**Aiming at the subjectivity and ambiguity problems of historical seismic intensity determination, a k-means clustering determination method combined with ant algorithm is proposed in this paper. Intensity features are extracted based on the latest historical seismic intensity table and the quantization process of historical earthquake records is dynamically adjusted by using exponential fitting method. The sensitivity problem of initial cluster centers in k-means clustering is resolved by using ant algorithm. Simulation results prove that the proposed method has high accuracy and practicality.**

Keywords- Historical seismic intensity; K-means clustering; Ant colony optimization; Exponential fitting

I. INTRODUCTION

After an earthquake, because the diffusivity and attenuation of seismic energy, different regions are affected differently by the earthquake,and the seismic intensityis a term used to describe the scale of earthquake influence strength degree [1]. Thanks to a long history in China, the long time series of seismic records can provide good data foundation for studying the earthquake, and the earthquake intensity as the key parameters of the epicenter and magnitude, is very meaningful to realize accurate judgement. But on the other hand, the historical earthquake records are predominantly descriptive language, and lack of unified format, so it's very difficult to inductive characteristic of earthquake intensity and quantitative, and many problems lack of characteristic information records, this makes historyrecords hard to be used inthe classification and intensity determination methods which is based on modern detailed earthquake records of data [2-5].

"Earthquake intensity scale seismic intensity scale" is the basic reference standard of historical earthquake intensity determine.However, both historical earthquake records, or the table of the historical earthquake intensity is qualitative description and quantitative degree is low, which makes determining work of the current historical earthquake intensity mainly depend on artificial judgment of the history experts [6]. Despite the strong speciality, personal subjective opinions are inevitably involved. Given the large number ofhistorical earthquake record, artificial evaluation method is difficult to realize the bulk processing.Therefore, it is very necessary to design an accurate, objective and efficient historical earthquake intensity method.

According to the above problem, this paper proposes a k-means clustering method combining of ANT algorithm thought (referred to as the ANT-KM algorithm) to realize the intensity determine of historical earthquake records.Through the contrast experiment,we proved the effectiveness of the proposed method

II. ANT K-MEANS CLUSTERING METHOD

A. Idea of ANT-KM algorithm

K-means algorithm can be simply expressed as: given k classes and randomly select k objects as the initial cluster centers. Then calculate the distance between other objects and each cluster center, and then assign these objects to its nearest cluster center. Every cluster center and its assigned object represent a cluster. Once all the objects are finally assigned,the cluster center of every cluster will be recalculated, according to the existing object in every cluster. This process will repeat until producing stable cluster results.

K-means algorithmis simple in principle, which has faster calculation speed for large data sets, and thus it is more suitable for broad historical earthquake records. But a big disadvantage of the algorithm is that it's very sensitive to the selection of the initial cluster centers, randomly selected different initial cluster centers may lead to clustering results differ significantly, thereby affecting the effectiveness of their clustering. To solve the problem, combined with ant algorithm, this paper allocate objects of k clusteringby using ants, so that the algorithm can adaptively adjust the cluster center, thereby reducing the impact on the quality of the initial cluster centers selected   to algorithm performance.

Specifically, the central value of each cluster in the algorithm of this paper is gradually fine-tuned based on 'one-entered' approach, after the first clustering assignment, the algorithm isn't directly calculate new cluster center to recalculate, instead of studying current k clustering "chaos degree", and select one of the cluster (set as out) a "arrangement", then examine the chaos degree contribution of every objects in cluster out, and select one of the object(set as obj) to leave the cluster, finally, inspect the "fit" between each cluster and obj, and put objinto a cluster. Repeat the process until convergence. It should be noted that the above behavior of ants, though closely related to the cluster center, but in line with the thinking of ant algorithm pheromone selection, that these selecting behaviors are carried out according to the probability, for example, when ants select a clustering to organize, the more chaotic

clustering evaluation has the greater probability of finishing, but the ants will not necessarily choose the most confusing clustering to organize, so it can well avoid the local best excellent, while ensuring the algorithm to converge on the direction we want, which means each initial confusing cluster is organized into orderly.

In addition, due to the fuzziness of the historical earthquake record itself, it makes the intensity dividend of historical intensity table is also based on qualitative description. And the intensity table uses plenty of vague words, making artificial intensity assessment became largely an empirical behavior, assessment results necessarily include rater experience because of the amount of false positive caused by subjectivity. However, the mere use of the machine determinesthe intensity table is rote and lacking of oversight, and due to advances in thehistory of disaster narrative civilization, historical disaster records increase sharply, Qin dynasty multiple records major natural disasters and gives a brief description, while the state records and county records is widely used in Ming and Qing dynasty, disaster statistical sources greatly enriched and detailed narration. If there was considered less Qin disasters and lightly, obviously is not objective, and therefore the application of intensity scale must be dynamically corrected, which is also a question needed to be pay special attention to while assessing the historical intensity by machine.

B. Description of ANT-KM algorithmic process

According to the latest "2010 historical earthquake intensity scale " (referred to 2010 Table), due to the intensity I and II only reflects people's slight feeling, and the historical earthquake records were not documented, so history intensity is divided from the beginning of scale degree III. This article is based on the table in 2010 on historical earthquake records to quantify intensity characteristics, and determine intensity III-XII.

Let the set of historical earthquake data is S, with each data p (where p = 4) attributes , the first data set S randomized to k (where k = 10) classes, all types of seismic data set as number tn (n∈[1, k]). Sequentially treated as follows:

1) Initial quantization

According to historical earthquake intensity scale (2010), the seismic intensity characteristics is divided into "human and social repercussions ", " house building ", " structure" and " surface phenomenon" four categories(ie, p=4), and according to the feature section of table 2010 to quantify the historical earthquake records, using codes "0-n" corresponding to the intensity characterization of severity, where " 0 " indicates that the recording feature is missing , "1-n" , respectively, the severity of each feature.When the severity of an historical earthquake records appears numbers of features descriptions produce a cross, whichever the greater value is to be encoded.

2) Dynamic correction

Due to the increasing popularity of the history of civilization, as well as paper , distant earthquakes recorded age is very simple, records close to the times is relatively detailed, showing a more recent era ago, the number of

seismic records and the more content-rich features, so disasters recorded earthquakes of various historical periods should have treated differently.If only unified treatment according to the 2010 statement of these historical earthquake records, then the ancient times' earthquake which describe simply because of the characteristics of the times may mix with the recent simple earthquake disaster records which means the disaster is gently. Thus it's far from the actual historical situation.

Figure 1 is the exponential fit to historical earthquake frequency using SPSS, which shows its upward trend is in line with the characteristics of the index. It's also consistent with the conclusions of the history experts, so we use the exponential function to fit the historical earthquake records, and to derive the exponential function to simulate the dynamic characteristics of the era of history, and correct the historical earthquake intensity characteristic value, as follows :

Let the exponential function fitted by SPSS is F(x), as shown in Equation (1) below, where x represents the historical year, a, b for the adjustment parameter.

$$F(x) = a * \exp(bx) \tag{1}$$

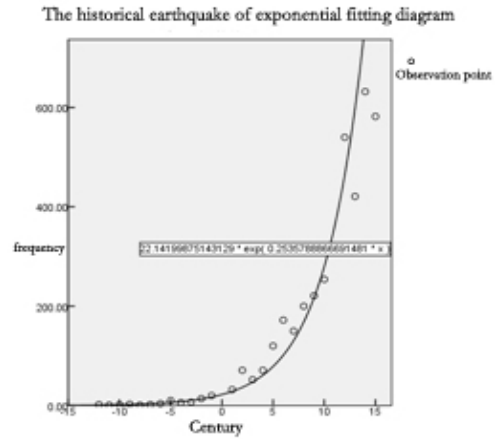The historical earthquake of exponential fitting diagram

Figure 1 Exponential fitting diagram

Followed by various dynasties to divide the interval, introducing analysis of gray correlation, setting the correlation of recent Qing dynasty is 1 (named Qing dynasty as the weighted base), getting the weights wepochof every dynasties based on Equation (2), where xbeginand xendmeans the beginning and ending year of the dynasty, xbegini and xendimeans the the beginning and ending year of the dynasty which wepochlies on.

$$w_{epoch} = 2 - \cfrac{1}{2 - \cfrac{\ln F(x_{endi}) + \ln F(x_{begini})}{\ln F(x_{end}) + \ln F(x_{begin})}} \tag{2}$$

The characteristic values of the various historical earthquake records multiplied by their respective weights

dynasties, we can get quantitative results after correction. Since the divided n values each of characteristic intensity are not the same, it also need to be normalized, as shown in Equation (3), the correction value of the jth feature of the ith seismic data is set to be xij, after normalization, changes into yij, where xmaxj, xminj means the maximum and minimum values respectively in the ith eigenvalue of the set of seismic data.

$$y_{ij} = (x_{ij} - x_{\min j})/(x_{\max j} - x_{\min j})$$ (3)

3) Selecting arranged classes

Firstly, ants need to examine the chaos degree of each current class, and select a category to finishing. Here, we indicate the chaos degree of each class with chaosn(n ∈ [1,k]), whose larger values indicates a more chaotic class(That is the whole difference between data in class and its cluster center is larger), then it's more likely that ant choose this class to sort. The chaos degree of nth class is shown in Equation (4) below:

$$chaos_n = \sqrt{\frac{\sum_{i=1}^{t_n} (x_i - avg_n)^2}{t_n}}$$ (4)

Where $x_i = \sum_{j=1}^{p} y_j^2$, represents the square of each attribute in the ith seismic data of nth class; $avg_n = \left(\sum_{i=1}^{t_n} x_i\right)/t_n$ represents the average of everyxi of the nth class; that is chaosn represents the standard deviation of each sum of square of the seismic data attributes of the nth class.

Then based on chaos value of each class, ants use the roulette wheel selection method to elect a class to sort.

4) Selecting arranged data

After finishing the class to be elected, then the ant need to assess the differences between seismic data and the cluster center of the class, that is the attribute to the chaos, represented by contribution. The larger the values of contribution is, the larger the difference between the seismic data and its corresponding class is, then the probability of that ants removed from the seismic data is larger. Contribution of the ith seismic data in class is shown in Equation (5) below:

$$contribution_i = (x_i - avg_n)^2$$ (5)

Then based on contribution value of each seismic data, ants use the roulette wheel selection method to arrive at a data to be processed.

5) Selecting data vesting class

Finally, ants need to decide put theselected data into which class, which have to assess thematched-degree of this data and each current class, represented by similarity n(n ∈

[1,k]), a lager similarityshows that thematched-degree of this data and each current class is higher, then probability of putting it into corresponding class is larger. The fit with the nth class as shown in Equation (6) as follows:

$$similarity_n = \frac{1}{(x_i - avg_n)^2}$$ (6)

Then based on contribution value of each seismic data, ants use the roulette wheel selection method to get a data attribution category, and put the selected seismic data into this class.

6) Selecting sorting scheme

To make the algorithm converges, we use a number of ants, respectively, let them step 3-5, then according to equation (7) to assess the quality of each ant's finishing solution, represented by fitness, the larger the value, the higher the quality of ants' finishing work, using the maximum fitness value corresponding finishing solution to change the original cluster. The sorting quality of the nth ant as shown below:

$$fitness_n = \frac{1}{t_n * chaos_n}$$ (7)

Where tn represents the size of nth class, chaosn is the corresponding chaos degree. Those values aresolved based on current ants' sorting solution.

7) Convergence

Circulating above steps 3-6 until the algorithm converges, then the seismic data selected from one class by ants will be putting back to original class, thus members of each class will not change again. Finally, the different types of data in turn divided into the corresponding intensity according to the order of the various types of cluster centers.

III.   SIMULATION RESULTS

In order to verify the validity of the proposed method in this paper, we implemented the programming algorithm in MATLAB 7.0, and randomly selected 1000 historical earthquake records which isfrom the pre-qin to qing dynasty from the historical disaster database.At the same time, two historical disaster experts assessed the intensity of the 1000 records according to the 2010 table.Because of fully reference the history of the seismic background factors in the process of expert assessment, the accuracy of the results is pretty high, which can be used as algorithm performance standards. We also choose the classic k means clustering algorithm, and the k - means clustering algorithm combined simulated annealing optimization [11] (hereinafter referred to as the SA - KM algorithm) as contrast algorithm. All the three algorithms usedthe fixed quantitative data by the proposed method.

On the experimental parameters setting, for SA - KM algorithm, the initial temperature t0 = 10, temperature coefficient of alpha = 0.90, the number of used ant in ANT - KM algorithm is 10.In addition, to enhance the efficiency,

some conditions are not set such asterminated annealing temperature, whilethe convergence condition is set as clustering results does not change within 10 loop.

TABLE 1 ALGORITHM MATCHING RATE

Algorithm	matching rate With expert 1	matching rate With expert 2	matching rate With expert 3
ANT-KM	86.3%	79.5%	82.9%
SA-KM	68.5%	73.9%	71.2%
K-MEANS	48.4%	57.4%	52.9%

The matching ratebetween algorithm of clustering results and the results of experts is shown in table 1. We use the matching rate with expert evaluation result as a reference of the algorithm accuracy. The table shows that matching rate of between the classic k means clustering algorithm and the experts was obviously lower, while the ANT - KM algorithm is better than that of SA - KM algorithm. Figure 2 is the algorithm convergence times under different seismic record numbers, we can see that the convergence speed of classic k - means clustering algorithm is faster than other two kinds of algorithm, the ANT - KM algorithm is faster than SA - KM algorithm. Inaddition, with the increase of data amount the needed convergence number of the ANT - KM and SA - KM algorithm also shows a rapid growth trend (SA - KM algorithm performances more obviously), it also reflects the ANT - KM algorithm requires further optimization.

It is important to note that due to the lack of unified standard of quantization, historical earthquake intensity experts give the result only as the reference of determination, and the opinions of the experts are not completely consistent. However, if the clustering result can achieve a high matching rate with a number of experts, we should consider that the clustering result itself has a certain guidance and practical. In large-scale intensity determination of historical earthquake records, we think it is very appropriate to use this algorithm to do the work of intensity initial dividing.
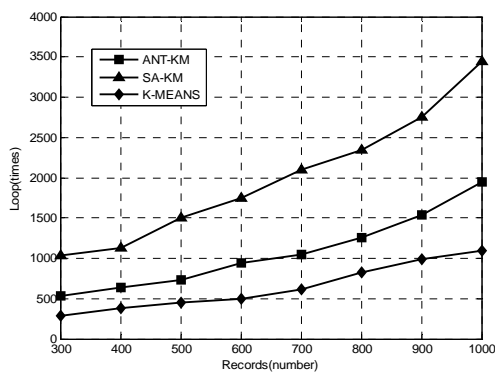

Figure 2.comparison of algorithm convergence time

IV.   CONCLUSION

In this paper,contrapose the long time series of historical earthquake records, use index fitting method to realize the dynamic adjustment of quantifying intensity characteristics,

which makesthe quantitative result conform to the historical truth well. And combined with ant algorithm, k means clustering algorithm was improved, in order to overcome the problem of the sensitivity to initial clustering center. Finally, comparison result proved the feasibility of this method through an example. As a kind of machine recognition method, compared with the artificial identification, it can be used more efficiently to large quantities of historical earthquake intensity determination, and has certain promotion value.

REFERENCES

[1]   Zhang Minzheng. Seismic intensity and Intensity Assessment [J].Journal of Institute of Disaster-Prevention Science and Technology ,2010,12(1):1-6

[2]   HUANG Han-ming, Zhou Hai-jun, BIAN Yin-ju. The Development of Decision Support System for Identifying Explosion Events by Seismic Waves[J]. JDCTA, 2011,5(4) : 107-112

[3]   MaoGuomin , GuJianhua , WuXinyan. Method Study Of Classification and Gradation of Earthquake Disasters[J].ActaSeismologicaSinica, 2007,29(4):426-436.

[4]   OUYANG Cheng-xin, XI Xiao-shuang. Seismic Evaluation of Building with Additional Factor[J]. JCIT, 2011,6(9) : 59-65

[5]   T. K. Nian, R. Q. Huang, K. L. Zhang, G. Q. Chen. Seismic stability analysis of three-dimensional slopes using Elasto-plastic FE method[J]. IJACT, 2011,3(9) : 367-376

[6]   Yan Jiaquan, Zhang ZhangZhizhong.A Study on seismic intensity scale of Chinese historical earthquakes[J].ActaSeismologicaSinica,2011,33(4):515-531

[7]   Huang Yonglin,GaoJunsuo.Historical data compilation of the present situation and the development of the earthquake [J].Journal of Catastrophology,2009,24(3):103-108

[8]   Chen Surong,ZhuXiaohui.K-means algorithm based on fuzzy logic is studied[J].Computer Engineering and Science,2012,34 (12): 155-159

[9]   XiongZhongyuan,ChenRuotian, Zhang Yufang. Effective method for cluster centers initialization in K-means clustering[J].Application Research of Computers ,2011,28(11):4188-4190

[10]   Hu Mingsheng,JiaZhijuan ,Liu Si.Approach for historical disaster correlation analysis based on ant colony optimisation[J].Computer Applications and Software[J],2012,29(10):62-64

[11]   Hu Yanwei ,Qin Zheng, Zhang Zhizhong.Intrusion detection algorithm based on simulated annealing and K-mean clustering [J].Computer Science,2010,37(6):122-124.