Research Article
Event Space-Correlation Analysis Algorithm Based on Ant Colony Optimization

[1,2]Mingsheng Hu, [1]Zhijuan Jia, [1]Si Liu and [2]Liu Hong
[1]Institute of Software, Zhengzhou Normal University, Zhengzhou, 450044, China
[2]Institute of Systems Engineering, Huazhong University of Science and Technology,
Wuhan, 430074, China

Abstract: Historical disaster events are taken as a case for space-correlation analysis, three-dimensional disasters space-time network are modeled and chain relationship of disaster nodes are mined by looking for similar space vector in network. Then transformed the vector discover problem into a path optimization problem and solved by using ant colony algorithm, where the pheromone parameter in the process of optimal-path finding is concerned as the algorithm result, in order to solve the problem of path competition which existed when only to solve the optimal path. Experimental results of MATLAB show that this method has high accuracy and practicality.

Keywords: Ant colony algorithm, disaster chain, path optimization, pheromone, space correlation

INTRODUCTION

More and more facts have shown that drought, floods, earthquake and other disaster events often do not exist in isolation, but have some connections. Such as, after 8.5-magnitude earthquake in Indonesia Sumatra in March 2005, major flood struck China's Pearl River Basin in July and after the 8.3-magnitude earthquake in Sumatra in September 2007, snow and ice storms struck the south of China in early 2008, because of this similar chain-like cognate phenomenon of disasters, the disaster chain and disaster evolution mechanism (Men, 2008) have been got more and more attention, many researchers of our country believe that the occurrence of the Sumatra earthquake is directly related to the above two disasters of our country, "(earthquake causes) the evaporation of sea water causes clouds and rain, they become flood in summer, become winter rain or snow in winter" (Yan *et al.*, 2011) and proposed the disaster's homology, chain nature and rhythmicity which reflected by the relationship between disasters, has a high value for the prediction of disasters. However, the relationship between disasters is not often instantaneous and self-evident, but is a large span of geographical and time, so the analysis and excavation of the relationship between disasters is a very complex problem.

Now the disaster correlation studies are mainly statistical analysis of disaster records, however, due to the space migration and time delayed etc., complex characteristics of the disaster correlation, it is difficult to effectively extract the association rules, thus the efficiency and credibility are low. Literature (Qiu *et al.*, 2011; Wu *et al.*, 2006) was based on time series similarity matching, achieved the correlation analysis

for the earthquake areas and achieved very good results, but the model can only obtain generally relevance degree, but cannot determine the causal relationship between the regions, namely, it shows the law of the region A cause the earthquake of region B, or shows the law of the region B cause the earthquake of region A, or both A and B are the areas of frequent earthquake, they shows a fake matching.

In view of this, this study attempts to describe the relationship between disasters with a historical three-dimensional space-time network model, this model reflects the causal relationship between the disasters by the vector space, thus transform the problem of disaster correlation analysis into the problem of excavating the similar vector of network and by the ant colony algorithm (Hu *et al.*, 2011) to improve velocity and precision of excavation, the experiments show that this method has a good application value.

SPACE-TIME NETWORK MODEL OF THE HISTORICAL DISASTERS

China has a long history, the account of natural disaster events in history books are characterized by huge number, rich type, long sequence, strong continuity etc and these records provide a valuable source of data for analyzing and summing the law of the disaster. So this study takes the historical disasters as the analysis object of the disaster correlation and maps various disasters in the history records to the map of China by GIS technology, then arranges the multi-layer disaster map according to the fixed time span (year or season) in order to build a three-dimensional network model of historical disasters as shown in Fig. 1.
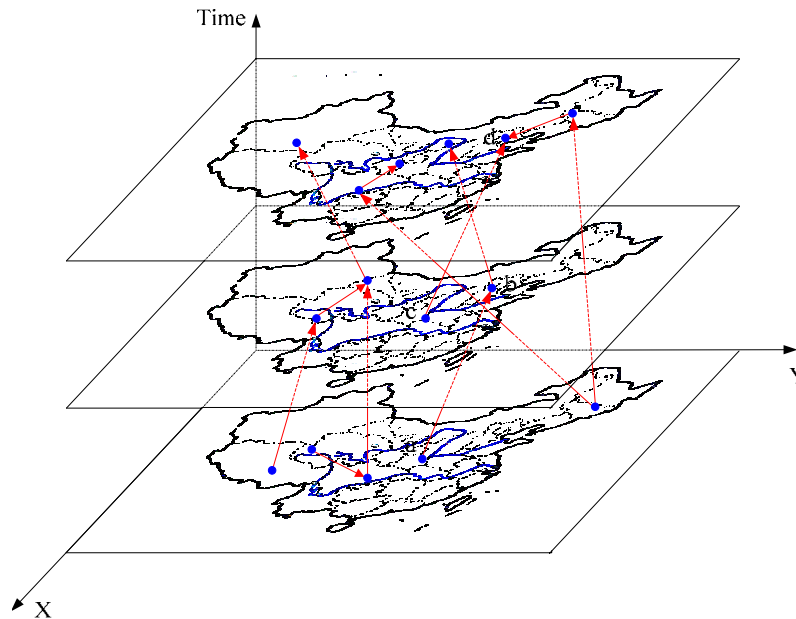
Fig. 1: Space-time network model of historical disasters

The blue spots represent disaster node, the red directed lines represent correlation vector (disaster chain, shows the previous disaster trigger the following disaster) in Fig. 1, the three-dimensional disaster vector can conveniently describe the evolving rule of disasters on time and space, so how to determine the disaster correlation vector needs to be addressed in this study. Multiple disaster nodes distributed each historical time layer, each node can choose a large number of nodes which can be linked with it and these nodes into a chain in the same layer or cross-layer, so only rely on analyzing the distribution rules of the entire three-dimensional network nodes to make sure the linked objects. Set the disaster node in the network is n, n can be defined as $n = (x, y, z, type)$, where x and y represent coordinates of disaster node on the map, z represents the layer of the disaster node, $type$ represents the category of disaster. For a node pair (n_i, n_j), if can find a number of node pairs which have the similar location and the same time span in the other layer of the network, then space vector which connected by these nodes is similar, the position of node a and node c are similar and they are separated by a layer in Fig. 1, the position of node b and node d are similar and they are separated by a layer, so the vector ab is similar with vector cd, if there are many vectors which are similar with the vector ab in the three-dimensional network, we can consider this type of vector space is disaster correlation vector, so can determine the chaining rules of the corresponding node location to the vector. Through the method of disaster correlation analysis which based on vector, not only can provide probability etc basis of being a chain for the further study of disaster network evolution model, but also can provide more specific advices for preventing disaster for the local governments.

In order to make the object of study more clearly, this study simplify the above model, only study the law of a single type and adjacent layers in single region of chaining disaster, such as Henan Province, study which areas of the earthquake in the underlying map will lead to flood in the adjacent upper map of Henan Province (earthquake-water disasters chain), then the problem becomes what vector can be found a number of similar vectors between the other adjacent layers in the all connected vectors that lower earthquake nodes point to the upper Henan Province floods nodes. If takes the interlamination vectors between these layers as a connection path between the connected layers and interlamination is tantamount to solving all possible paths from the bottom level to the top level, how to make the sub-paths in the total paths as much as possible similar, can be transformed into a path optimization problem.

Obviously, the above path optimization problem belongs to *np*-hard problem, the ant colony algorithm which as a classic met heuristic has better computational efficiency on the problem of path optimization. In addition, obtaining an optimal path is not a perfect scheme for solving the above problem, because optimal path can only select a sub-path between every two levels, but it is likely to have many types of similar sub-paths between the two layers, that is, there is the phenomenon of "competition". However, what we are looking for is not the path which has the most parallels in the similar sub-paths, but as long as has a certain amount parallels in the sub-path is valuable path, ant colony algorithm builds a new solution by releasing

pheromone on the path, this way provides a guideline for us to solve the "competition", in the algorithm, the more pheromone owned by a sub-path, the greater probability is selected by the ants, the algorithm converges can be considered if a path has accumulated a number of considerable pheromones, then this path is excellent, therefore, in the using of ant colony algorithm, is let the ant colony algorithm to find the optimal path as the media and the ultimate aim is to extract the sub-path which with high pheromones at the end of the algorithm, in a sense, it is a reverse using of the ant colony algorithm.

DESCRIPTION OF ALGORITHM PROCESS

The description of the algorithm is as follows; set the number of ants as *m*.

Data initialization:

Step 1: Import network nodes topology and generate the node path matrix, then initialize the pheromone matrix *pheromone* (*n*) (*n*), where the value of array element *pheromone* (*i*) (*j*) (*i* is an arbitrary node in lower layer, *j* is an arbitrary node in the selected region of adjacent upper layer) which represents the pheromone content of the path from node *i* to *j* is set a uniform initial value as shown in formula (1):

$$\text{Pheromone [i] [j]} = v_{\text{init Pheromone}} \qquad (1)$$

Step 2: Initialize the parameters of ants and algorithm, such as the maximum iterations *epoch-max*, the length of global optimal *path-best,* the lower limit of pheromone content *pheromone-min* and so on.

Iteration: When the iteration number is equal to *epoch-max*, the iteration end; otherwise circulating the following acts:

Step 1: Clear the way-finding memory of all ants on the previous iteration.
Step 2: All ants were randomly assigned to an initial node in the lowest layer.
Step 3: Every ant builds a full path: In each step of the path construction, ant *k* (*k* ∈ (1, *m*)) in accordance with the roulette wheel selection method (Zhijuan *et al.*, 2012) to decide the next node to go. The characteristic of this selection method is that the value of *pheromone* (*i*) (*j*) corresponding to a path is greater, the probability of the path chosen by ant *k* is greater. So on the one hand ants tend to choose the paths which is considered better currently, on the other hand the other paths are also have

the chance of being selected that can avoid algorithm falling into local convergence prematurely. The probability of an ant in node *i* will choose the node *j* as its next visit node is shown in formula (2), where *N* represents the set of optional nodes in next layer:

$$p_{ij} = \frac{pheromone[i][j]}{\sum\limits_{v \in N} pheromone[i][v]} , \quad j \in N \qquad (2)$$

Step 4: Calculate the fitness value of each path constructed by ants, then reduce the pheromone content of each sub-path has been visited (set the reduction as V_{sub}), so that the probability of other paths chosen by ants in next iteration is increased which could prevent the algorithm falling into a local optimum. If the pheromone content of a path is less than *pheromone-min*, then make it equal to *pheromone-min*, which can avoid the pheromone content of a path is too low so the path can hardly be selected. Algorithm need to find a full path contains the most similar sub-paths, so the fitness value is obtained by comparing the sub-path similarity, set the coordinate of starting node *i* of sub-path $p_{i,j}$ as (x_i, y_i, z_i) and the coordinate of tai node *j* as (x_j, y_j, z_j), the coordinate of starting node *u* of sub-path p_{uv} as (x_u, y_u, z_u) and the coordinate of tai node *v* as (x_v, y_v, z_v), then the similarity judging method for these two sub-paths is shown in formula (3):

$$sim = \sum\limits_{\substack{a=i,b=u; \\ a=j,b=v}} \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \leq threshold \qquad (3)$$

where, *threshold* represents the similarity threshold, two sub-paths are similar if the value of *sim* below the predefined threshold, then the similarity determination value (represent by *h*) of these two sub-paths is equal to 1. However, due to the error increases phenomenon in the comparing process, for example, there are three sub-paths set as *a*, *b* and *c*, *a* and *b* are similar, *a* and *c* are also similar, but *b* and *c* are not necessarily still similar, so in this study the fitness value is determined by comparing the sub-paths one by one, as shown in formula (4):

$$fitness = \alpha * \sum\limits_{sim_{ij} \leq threshold} h_{ij} \qquad (4)$$

where,
α : The adjustment factor
i, j : For any two sub-paths in constructed full path

Table 1: Path information matched with the actual history

Vector starting point	Vector endpoint	Average pheromone content	Number of similar records in history
Nanjing, Yangzhou	Northern Henan	30.7	94
Taiyuan	Jiaozuo, Zhengzhou	23.2	65
Baoding, Shijiazhua-ng	Kaifeng	17.4	50
Northern Shandong	Northern Henan	15.2	44
Linfen	Henan province	13.4	39
Wuhan	Kaifeng	8.1	33
Quzhou	Shangqiu	7.5	21

Step 5: Select the path which has the highest fitness value in this iteration and increase the pheromone content (set the increment as V_{add}) of this path, which is to reward the best path in this iteration so the probability of this path chosen by ants in next iteration is increased and V_{add} is significantly greater than V_{sub}. If the fitness value of this path is more than *path-best*, then use it to replace *path-best* and record the node order of this path.

Step 6: Go to the next iteration.

Iteration complete: Extract the sub-paths whose pheromone content is higher than the set threshold and combine the similar sub-paths with same kind, then return to network and conduct accurate data processing, so as to figure out the actual similar number of these sub-paths.

CASE STUDY

In order to verify the performance of the proposed algorithm, we programmed the algorithm under MATLAB 7.0, the experimental data are the nationwide drought records and locust plague records of Henan in Ming and Qing Dynasty (1368-1912), so as to study which region of China has a drought, will more likely to cause the occurrence of locusts in Henan, the data derived from historical disaster databases which constructed by relying Fund, where the record years of Henan locust are 216 years and in the 216 years with drought records at the same time in China there are 211 years, this study used the corresponding historical disaster records of the 211 years to establish the disaster network.

In specific experiment, the number of ants in the algorithm is set to 10, used time layer as year to conduct the network construction, the disaster geographical positions are sampling point marked by China map under the range of 1280 * 1024, the initial pheromone content of each path is 1, the pheromone reward for best path is 0.1, pheromone volatilization is 0.01, the similarity determine threshold is set to 15.0, the number of iterations is set to 2000. After observed repeated tests, set the sub-path whose pheromone content is greater than 7 is the similar sub-path after the end of iteration and finally sorted out a total of seven groups of similar
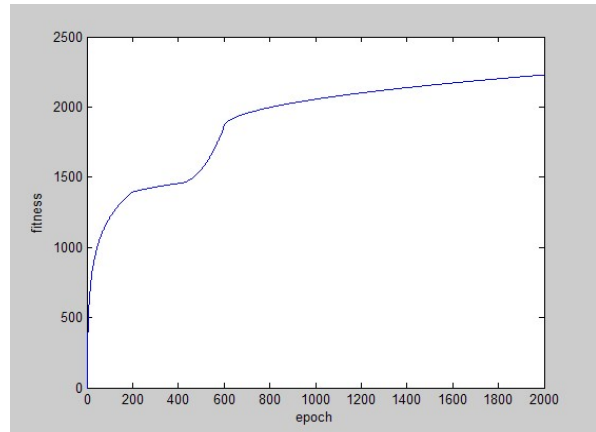

Fig. 2: Maximum fitness value in iteration

sub-paths and the vector information represented by these similar sub-paths is shown in Table 1, due to disasters node is determined according to the affected center which in fact has a certain affected range and when combine the similar vectors, the affected disaster range may further expand. The vector information in Table 1 has been translated into specific locations, from the table we can found the end points of vector are mostly distributed in the northern part of Henan, it has great relevance with the historical locust records in Henan are also concentrated in the northern part. According to the experimental results, it can be seen in history after the drought occurred in Nanjing, Taiyuan, Shijiazhuang, there are many locust records appeared in Henan (in particular, Nanjing-northern Henan vector has a very prominent number of similar records) and therefore these vectors have a high disaster relevance, worthwhile for the Government to pay early attention when the similar incident occurs.

Figure 2 is the parameter monitoring chart for the global highest value when the program is running, parameters, you can see the fitness value increased steeply in algorithm running prophase and improved and converged steadily in the later. In addition, because the algorithm is not purposed to obtain the optimal path, the set number of cycles can be further reduced, the distribution of sub-path pheromone content can be considered mature when the change of fitness value began to slow, so the computational efficiency of the proposed algorithm is actually higher than the normal ant colony algorithm.

CONCLUSION

This study through the establishment of three-dimensional disaster network to make the analysis of disaster correlation more clarity on the geographical and time scales and converted the purpose of using ant colony algorithm from solving the optimal path to extract high pheromone sub-paths, so that to solve the problem of path competition. However, the "single" disaster chain solved in this study is only a basic relational form of disasters evolution and for further disaster evolutionary modeling, it still need to do more in-depth discussion on the circumstances of composite hazard, cross-layer.

ACKNOWLEDGMENT

REFERENCES

Hu, M., J. Zhijuan, D. Xiangying and H. Liu, 2011. SA-ANT: A historical epidemic classification method based on simulated annealing and ACO. Int. J. Adv. Comput. Technol., 3(11): 47-54.

Men, K.P., 2008. Time-space order of severe earthquake disaster chain and its prediction. Progress Geophy., 22(2): 645-651.

Qiu, J.F., J. Xie, W. Li and J.W. Wang, 2011. Research on correlation and periodicity of moderate-strong earthquake. Comput. Eng., 37(10): 16-18.

Wu, S.C., G.F. Wu, W. Wang and Z.C. Yu, 2006. A time-sequence similarity matching algorithm for seismological relevant zones. J. Software., 17(2): 185-192.

Yan, J.P., J. Bai and K.H. Su, 2011. Research on symmetry and tendency of several major natural disasters. Geograph. Res., 30(7): 1159-1168.

Zhijuan, J., H. Mingsheng, D. Xiangying and H. Liu, 2012. Application of ant-clustering in classification of historical disaster. Adv. Infor. Sci. Serv. Sci., 4(2): 135-143.