Applied Mechanics and Materials *Online: 2013-08-30*
ISSN: 1662-7482, Vols. 380-384, pp 753-756
doi:10.4028/www.scientific.net/AMM.380-384.753

Entropy-based City Tunnel real-time traffic incident detection Clustering Algorithm

Xiaofeng Li[1, a], Weiwei Gao [2,b] , Xuemei Wang [3,c]

[1,2]Department of Information Science, Heilongjiang International University, Harbin 150025, China

[3]Department of Computer Science and Technology, ChengDong College Of Northeast Agricultural University, Harbin 150025, China

[a]email: mberse@123.com, [b]email: 48848805@qq.com, [c]email:Master2@163.com

Keywords: Information entropy; City Tunnel real-time; Spatial data mining; Clustering

Abstract. The use of spatial clustering technology has important practical significance to obtain useful information. According to the characteristics of city tunnel real-time traffic ,then, put forward ECRT (Entropy-based City Tunnel Real-time), the object associated with the city tunnel as real-time traffic properties to calculate the entropy of information between the city tunnel, based on information entropy change to achieve real-time traffic urban tunnel clustering. Algorithm used in the actual data set ECRT test. The results showed that the algorithm ECRT is effective.

Introduction

With the continuous development of informatization of human society, the quantity and complexity of spatial data have been increasing, which makes it impossible to provide conclusive and useful information for the customers through mere inquiry and search of the database. As a result, a large amount of information in the spatial database has not been fully utilized. Against this backdrop, the spatial data mining (SDM) and knowledge discovery (KD) have emerged as an automatic and intelligent method to assist people in appreciating and applying the spatial database in a better way[1-2].

Spatial clustering is an important method to mine spatial data. In general, it refers to divide spatial data object sets into clusters of similar items; the items in the same cluster share great similarities while different clusters have great difference. The combination of different clusters results in various congested areas, which refer to some spaces in which the items are distributed in relatively high density, and when the density is higher than some certain thresholds, the area is "congested". As the major application in the field of knowledge discovery, clustering technology has been studied intensively by the scholars both at home and abroad, who have come up with many practical clustering. By mining data of the characteristics of traffic accidents as well as the factors of people, transports, and roads that lead to traffic accidents in accordance with the association rules, and by simulating the process of traffic accidents with virtual reality technology, document gives an analysis to the traffic accident data. Document [3-4] describes the appliances of Floating Car Data (FCD) technology in the simulation of real-time traffic. With the multiple linear regression patterns simulating the tunnel traffic speed, the appliances can reasonably and effectively simulate the real-time traffic in the tunnels so as to truly reflect the actual speed of cars in the tunnel, To provide strategies for the prevention of traffic accidents[5].

While elaborating the data mining, that is, the Association Rules Mining Algorithm of common traffic accidents in urban tunnels, in this article, an improved algorithm - the Detection of Urban Tunnel Real Time Traffic Accidents Based on Information Entropy is mainly proposed and put into experiments, which turns out to be able to improve the automation level of spatial data analysis and appliance. This algorithm takes into consideration the sophistication of spatial data and the links of data, such as the tunnel emergencies, particularly the links between the implicit and extract knowledge.

The Clustering Designed for the Detection of Real-time Traffic Accidents in Urban Tunnels Based on Information Entropy

Because the contemporary clustering doesn't well take into account the complexity of spatial data and the links between the data, plus the low degree of precision of clustering, the article provides a new spatial clustering - ECRT of urban tunnels based on Information Entropy. Next are detailed elaborations on the algorithm's design philosophy and procedures[6].

The Design Philosophy

Spatial data sets include many types of spatial objects, which possess the others' properties, quantities, directions, distances or topological relations [7]. These features of the spatial objects make it impossible to directly use the existing spatial clustering to the clustering analysis of spatial data sets. However, information entropy can well solve the problem. To a certain space, when the objects are similar, the information entropy is smaller than that when the objects are different. Based on this theory, we introduce information entropy to the spatial clustering. This will help solve the problems of clustering of the spatial objects with topological relations in the complicated spatial data sets.

$S=\{S_1,S_2,...,S_m\}$ is an aggregate of spatial object group composing spatial data set D. CL refers to one of the clustering of S, $CL=\{C_1,C_2,...,C_k\}$, C_l is the l th cluster; $C_l \subseteq S$, $l \in [1,k]$, k refers to the number of clustering, CL minimizes the value of the target function F; the target function F follows the formula (1), in which E_l is the total value of information entropy of all the spatial object groups in the l cluster, and the total value is reached by the formula (2).

$$F=\sum_{l=1}^{k} E_l \tag{1}$$

$$E_l = E(\{S_i \mid S_i \in C_l\}) = -\sum_{h=1}^{n}\sum_{t=1}^{T} p_{ht} * \lg p_{ht} \tag{2}$$

The Procedures

The major procedures of ECRT are: based on the corresponding topological relations, through the changes of information entropy in the spatial data objects' neighborhood, to instruct Space Object to pick up or drop spatial data, thus clustering spatial data sets. The detailed procedures are as follows:

① For $iter=1$ to $N_{iteration}$
② For $q=1$ to N_{ant}
③ if (Space Object$_q$ isload=false), then Space Object$_q$ pick up Space Object S_i, $i \in [1,m]$;
④ else calculate the information entropy value E_1 when Space Object$_q$ in the neighborhood didn't drop S_i;

Calculate the information entropy value E_2 when Space Object$_q$ in the neighborhood dropped S_i;

⑤ if($E_1 >= E2 \,\|\, fail > N_{fail}$) then Space Object$_q$ put down S_i at the current position;
⑥ else Space Object$_q$ loads S_i and choose a direction rationally to move with the distance d to next position, $fail++$;
⑦ End if
⑧ End if
⑨ End For
⑩ End For

In the process of iteration, the value of Space Object's searching radius r and step length d can be adjusted, which not only enhances efficiency, but also reduces the likelihood of the method to turn to local optimum. After finite times' iteration, the method put out the results of clustering.

The algorithm optimization of ECRT

The results of Ant Colony Clustering Algorithm based on information entropy are crude; in the process of clustering of Ant Colony Algorithm, Space Objects only take into account piling up similar spatial targets and isolating different space, rather than separating different clusters, which makes it vague to define different clusters. To make the algorithm more practical, some improved procedures are listed below:

① For $i = 1$ to m m refers to the total number of spatial data sets

② p stands for the current position of S_i; Space Object picks up S_i; the failure times of Space Object $fail = 0$;

③ Calculate the Space Object's value of information entropy in position p and that in the neighborhood of S_i; S_i refers to the position after the Space Object moves one step in six directions. Take the minimum value and record the corresponding position p'.

④ if($fail < N_{fail}$ && $p \neq p'$), then Space Object carries S_i to position p', and if $fail$++, $p = p'$, repeat ③;

⑤ when Space Object is in position p', put down S_i;

⑥ End if

⑦ End For

Experiment and Result Analysis

Verification of ECRT

The real-time speeds in the table above are the average value of 30 days' observation in a row. The experiment analysis above and ECRT bring out the results below. Figure 1 shows the distribution of the originally input data. Figure shows the clustering results from ECRT.

Figure 2 implies that if the data are similar, they will get together to the same cluster; if something abnormal occurs, it would be isolated points. This clearly shows the features after a large quantity of data piles up. Chart 2 also shows that nothing abnormal happened in the tunnel during the experiment period.

Performance Analysis of ECRT

In order to verify the performances of ECRT, the author made a comparison of the experiment data listed above in the same circumstances with the Association Rules and ECRT separately. Figure 3 shows the error accuracy. From the Figure we can find that ECRT has lower error rate, particularly when with abundant data, the rate of accuracy is higher.

Conclusion

In response to the need of real-time detection in tunnels, and based on the existing Association Rules Analysis, a new method is put forward and the experiments are adopted to verify the effectiveness of this method in this article. which lays a solid foundation for the widespread use of this method. Through comparison with Association Rules, it can be concluded that this method possesses higher performances.

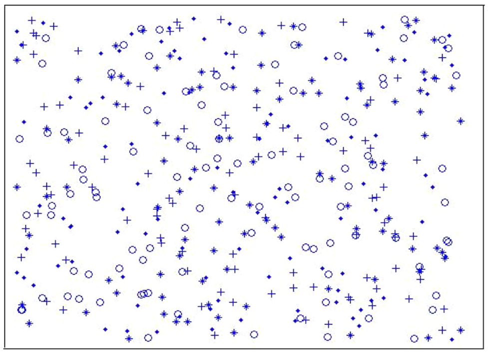

Figure 1 Output of the Analysis of Original Data

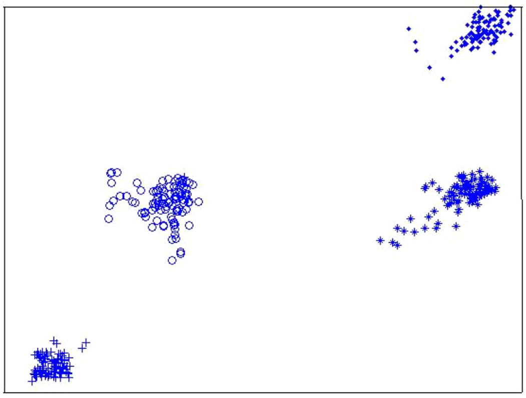

Figure 2 The output of ECRT improved method

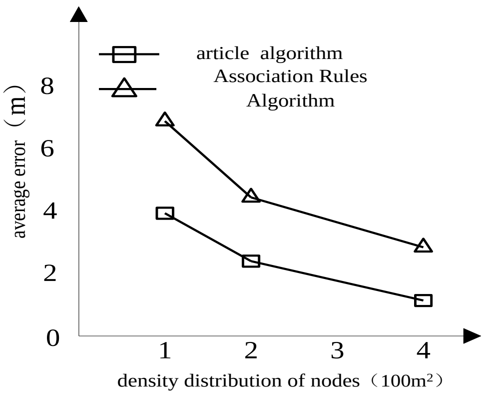

Figure 3 The degrees of Accuracy of the two methods

References

[1] Liu chun , Huang meixian. Based on the data of the floating car city road tunnel real-time simulation DuoYuan linear regression [J]. Traffic and computer, 2008:26-38

[2]Lin Chengru, Liu Kenhao, Chen Ming-syan.Dual Clustering:Integrating Data Clustering over Optimization and Constraint Domains[C].Proceedings of the IEEE Transactions on Knowledge and Data Engineering[C].Piscataway:IEEE Educational Activities Department,2005:628-637

[3] Zhang Xueping,Wang Jiayao,Wu Fang,et al.A Novel Spatial Clustering with Obstacles Constraints Based on Genetic Algorithms and K-Medoids[C].Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA' 06),2006,1:605-610

[4]Huang Ming,Bian Fuling.A Grid and Density Based Fast Spatial Clustering Algorithm.2009 International Conference on Artificial Intelligence and Computational Intelligence,2009,4:260-263

[5]Hu Caiping,Qin Xiaolin.A novel spatial cluster algorithm with sampling[C].Berlin Heidelberg:Springer Verlag,2007:216-225

[6]M ao demei ,Ding ruiguo In the data mining association rules algorithm was studied[J]. Wanxi college Journal 2006(10):27-30

[7] Xiang Laisheng,Guo Yajun,Lan Tian.Topological cluster:a generalized view for density-based spatial clustering[C].International Conference on Management Science& Engineering(14th).Harbin,China:IEEE Service Center,2007:422-428