International Journal of Engineering Research and General Science Volume 2, Issue 4, June-July, 2014
ISSN 2091-2730

SVM Based Spatial Data Mining for Traffic Risk Analysis

Roopesh Kumar[1],Diljeet Singh Chundawat[2], Prabhat Kumar Singh[3]

[1,2,3] Lecturer Deptt.of Computer Science & Engineering,

MIT,Mandsaur M.P.(India)

roopesh.kumar@mitmandsaur.info

Abstract: - Extracting knowledge from spatial data like GIS data is important to reduce the data and extract information. GIS data also contains information about accidents at certain place and road conditions. Such data contains useful information for the traffic risk analysis. But such information are not directly present in the dataset. Hence spatial data mining technique is needed to extract knowledge from these databases. The previous work shows unpractical approach to multi- layer geo-data mining. That means the information from various sources are combined based on the data relation and data mining is performed on the relation. The efficiency of risk factor evaluation requires

Key Word:- GIS, SVM,Clustering,Geo-Datamining, Ant Colony Optimization

INTRODUCTION

Automatic filtering of spatial relationships, on the whole the outcome of decision tree has dependence on the initial data which are incomplete, incorrect or non- relevant which inevitably cannot deliver error free results. The suggested model develops a SVM based technique to achieve the same by first training support vector machine with risk pattern and further classifying the data based on the training model. Therefore not only the result is based on relational model but also based on complex kernel techniques, compared to other existing approaches using non-intelligent decision tree heuristics.

LITERATURE REVIEW

Spatial data mining fulfills real needs of many geomantic applications. It allows taking advantage of the growing availability of geographically referenced data and their potential richness. This includes the spatial analysis of risk such as epidemic risk or traffic accident risk in the road network. This work deals with the method of decision tree for spatial data classification.

This method differs from conventional decision trees by taking account implicit spatial relationships in addition to other object attributes. Ref [2, 3] aims at taking account of the spatial feature of the accidents and their interaction with the geographical environment. It involves a new field of data mining technology that is spatial data mining. In the previous work, the system has implemented some spatial data mining methods such as generalization and characterization. This work [3] presents the approach to spatial classification and its application to extend TOPASE.

Clustering in spatial data mining is to group similar objects based on their distance, connectivity, or their relative density in space. In the real world, there exist many physical obstacles such as rivers, lakes and highways, and their presence may affect the result of clustering substantially. In this project, the system studies the problem of clustering in the presence of obstacles and defines it as a COD (Clustering with Obstructed Distance) problem. As a solution to this problem, the system proposes a scalable clustering algorithm, called COD- CLARANS [5,6].

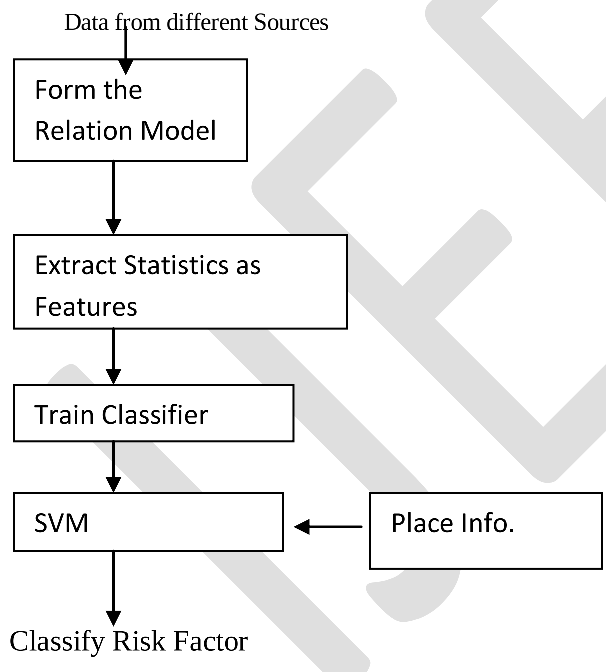
Spatial Clustering with Obstacles constraints (SCOC) has been a new topic in Spatial Data Mining (SDM). In [8] the author proposes an Improved Ant Colony Optimization (IACO) and Hybrid Particle Swarm Optimization (HPSO) method for SCOC. In the process of doing so, the system first use IACO to obtain the shortest obstructed distance, which is an effective method for arbitrary shape obstacles, and then the system develop a novel HPKSCOC based on HPSO and K-Medoids to cluster spatial data with obstacles, which can not only give attention to higher local constringency speed and stronger global optimum search, but also get down to the obstacles constraints.
Spatial clustering is an important research topic in Spatial Data Mining (SDM). Many methods have been proposed in the literature, but few of them have taken into account constraints that may be present in the data or constraints on the clustering. These constraints have significant influence on the results of the clustering process of large spatial data. In this project, the system discuss the problem of spatial clustering with obstacles constraints and propose a novel spatial clustering method based on Genetic Algorithms (GAs) and KMedoids, called GKSCOC, which aims to cluster spatial data with obstacles constraints.[9] Spatial data mining method is used to enrich Customer Intelligence analytical function in this project. The system first proposes a spatial data classification method which can handle the uncertainty property of customer data. On the basis of spatial classification rules, the system then proposes a detection method of potential customers by map overlapping. Deep spatial analytical function is

realized in customer intelligence system which cannot be done by traditional data mining method. With the coming of E-business, the enterprises are now faced harder competition than before. So they now focus their attention on customers instead of their production only. In order to win the competition, the enterprises have to provide their customers more individualized and more efficient service. Customer intelligence (C1) system appears in recent years to meet the need of above emergence. From the analytical function of the system, customer intelligence is a decision analytical methodwhich includes customer identification, customer selection, customer acquirement, customer improvement and customer maintenance [12]. The spatial co-location rule problem is different from the association rule problem, since there is no natural notion of transactions in spatial data sets which are embedded in continuous geographic space. In this project, the system provides a transaction-free approach to mine co-location patterns by using the concept of proximity neighborhood. A new interest measure, a participation index, is also proposed for spatial co-location patterns. The participation index is used as the measure of prevalence of a co-location for two reasons.

Modeling spatial context (e.g., autocorrelation) is a key challenge in classification problems that arise in geospatial domains. In [13] Markov random fields (MRF) are a popular model for incorporating Spatialcontext into image segmentation and land-use classification problems. The spatial auto regression (SAR) model [14], which is an extension of the classical regression model for incorporating spatial dependence, is popular for prediction and classification of spatial data in regional economics, natural resources, and ecological studies.

GENETIC AND ACO BASED SPATIAL DATA MINING MODEL

The proposed spatial data mining model uses ACO integrated with GA for risk pattern storage. The proposed ant colony based spatial data mining algorithm applies the emergent intelligent behavior of ant colonies. The proposed system handle the huge search space encountered in the discovery of spatial data knowledge. It applies an effective greedy heuristic combined with the trail intensity being laid by ants using a spatial path. GA uses searching population (set) to produce a new generation population. It evolves into the optimum state progressively by exerting a series of genetic operators such as selection, crossover and mutation etc on traffic risk patterns. The proposed system develops an ant colony algorithm for the discovery of spatial trends in a GIS traffic risk analysis database. Intelligent ant agents are used to evaluate valuable and comprehensive spatial patterns.



CONCLUSION

It is our survey paper and we have gone through for different aspects of data mining and find out problem in previous approaches and find out the way of solution using SVM so that we have proposed some method .We will do implementation and will take as future work

REFERENCES:

[1] [Agrawal & Srikant1994] Agrawal, R., and Srikant, R. 1994. Fast algorithms for Mining Association Rules. In Proc. of Very Large Database.

[2] [Anselin1988] Anselin, L. 1988. Spatial Econometrics: Methods and Models. Dordrecht, Netherlands: Kluwer.

[3] [Anselin1994] Anselin, L. 1994. Exploratory Spatial Data Analysis and Geographic Information Systems. In Painho, M., ed., New Tools for Spatial Analysis, 45-54.

[4] [Anselin1995] Anselin, L. 1995. Local Indicators of Spatial Association: LISA. Geographical Analysis 27(2):93{115.

[5] [Barnett & Lewis1994] Barnett, V., and Lewis, T. 1994. Outliers in statistical Data. John Wiley, 3rd Edition.

[6] [Besag1974] Besag, J. 1974. Spatial Interaction and Statistical Analysis of Lattice Systems. Journal of Royal Statistical Society: Series B 36:192-236.

[7] [Bolstad2002] Bolstad, P. 2002. GIS Foundamentals: A Fisrt Text on GIS.Eider Press.

[8] [Cressie1993] Cressie, N. 1993. Statistics for Spatial Data (Revised Edition). New York: Wiley.

[9] [Han, Kamber, & Tung2001] Han, J.; Kamber, M.; and Tung, A. 2001. Spatial Clustering Methods in Data Mining: A Survey. In Miller, H., and Han, J., eds., Geographic Data Mining and Knowledge Discovery. Taylor and Francis.

[10] [Hawkins1980] Hawkins, D. 1980. Identification of Outliers. Chapman and Hall. [Jain & Dubes1988] Jain, A., and Dubes, R. 1988. Algorithms for Clustering Data. Prentice Hall.

[11] [Jhung & Swain1996] Jhung, Y., and Swain, P. H. 1996. Bayesian Contextual Classification Based on Modified M-Estimates and Markov Random Fields. IEEE Transaction on Pattern Analysis and Machine Intelligence 34(1):67.

[12] [Koperski & Han1995] Koperski, K., and Han, J. 1995. Discovery of Spatial Association Rules in Geographic Information Databases. In Proc. Fourth International Symposium on Large Spatial Databases, Maine. 47-66.

[13] [Shekhar et al.2002] Shekhar, S.; Schrater, P. R.; Vatsavai, R. R.; Wu, W.; and Chawla, S. 2002. Spatial Contextual Classification and Prediction Models for Mining Geospatial Data. IEEE Transaction on Multimedia 4(2).

[14] [Zhang et al.2003] Zhang, P.; Huang, Y.; Shekhar, S.; and Kumar, V. 2003. Exploiting Spatial Autocorrelation to Efficiently Process Correlation-Based Similarity Queries. In Proc. of the 8th Intl. Symp. on Spatial and Temporal Databases