International Journal of Database Theory and Application
Vol.10, No.8 (2017), pp.11-20
http://dx.doi.org/10.14257/ijdta.2017.10.8.02

An Intelligible Illustration of an Epidemic Spatio – Temporal Statistics on GIS

B. Renuka Devi[1], Venkata Naresh Mandhala[2], Debnath Bhattacharyya[3] and Hye-jin Kim[4]

*[1]Department of Computer Science and Engineering,
Vignan's Nirula Institute of Technology & Science for Women,
Peda Palakaluru, Guntur, AP, India
[2]Department of Computer Science and Engineering
KL University, Vaddeswaram, Guntur, AP, India
[3]Department of Computer Science and Engineering,
Vignan's Institute of Information Technology,
Duvvada, Visakhapatnam, AP, India
[4]Sungshin Women's University, 2, Bomun-ro 34da-gil,
Seongbuk-gu, Seoul, Korea
[1]brdcse@gmail.com, [2]mvnaresh.mca@gmail.com, [3]debnathb@gmail.com,
[4]hyejinaa@daum.net
(Corresponding Author)*

Abstract

Innovations in the sector of information technology have enabled the collection and processing of enormous amounts of spatial data. The goal of data mining is to determine nuggets. Spatial data mining identifies the collocation rules. Spatial data are considered from the spatial objects. The considered spatial data is preprocessed by using the data mining tool. To the preprocessed data, collocation rule is applied for detecting the frequent item sets. Disaster impacted areas were predicted by applying the collocation rule. In particular to spatial data mining, when spatial data are comparatively represented in time series, a spatio-temporal significance is concluded. In this perspective, the collocation rule that is an epitome for the spatial data acquires changes with temporal impact. Therefore, the changes that arise to the spatial knowledge are the spatio-temporal transactions. Extracting the spatio-temporal transactions and finding the various behavioral aspects of collocation is one of the considerable activities of GIS. By implementing the collocation rule with "nearby" as the predicate, disaster affected areas are identified follows the representation of the spatial data on Geographical Information Systems (GIS) by various colored pinpoints for all the quarters of a year. From that, the regions at risk zone of disaster were predicted, then the analyzed spatial data will be redirected to the health organizations for supervising campaigns. Our focus is to forecast the disaster, design the spatio-temporal trees for all the quarters of a year and to represent the spatial nuggets on GIS. Therefore, a spatio-temporal disaster management system is designed and implemented. A novel data structure for the spatio-temporal data is proposed.

Keywords: *Collocation rule, Dengue Fever, Geographical Information Systems, Pre-processing, Spatial Data Mining, Spatio-temporal trees*

1. Introduction

Spatial data mining is a significant extension that discovers the non-trivial and potentially valuable datasets from huge data sets. Mining useful patterns from huge spatial

ISSN: 2005-4270 IJDTA

databases, which cover various technical overheads like, spatial data infrastructures, spatial relationships, spatial autocorrelation and some others related to spatial-geometry [1]. Temporal data mining is the mining of data based on time-series[2]. Temporal data mining can be defined as the action of looking for attention-grabbing correlations or patterns in large sets of temporal data acquired for other purposes. Integrating the semantic support from spatial data mining and temporal data mining, with a meager correlation, the idea of spatio-temporal mining helps us to give, the temporal meaning to the evolving and ever changing collocation rules[3]. However, the characteristic fuzzy or Boolean, the problem of collocation is whether a static or a dynamic is complicated for several GIS application users. However, the rules are characterized as out-of-date which does not please the decisive factors of GIS apprehension, such are disposed. And with the historical rules that illustrate the incident, which are mandatory, in addition to which are used to identify the state transformations in terms of size and shape of the elementary parts of the rules are of very significant concern when related to spatial, several algorithms for mining the spatio-temporal conclusion encompasses much importance [4]. In general the potentiality of transformations occurred to the collocation, is to realize the time components associated to the transformations and extracting the temporal sequences that identify the sequence of transformations.

2. History

Spatial data mining extracts potentially helpful patterns from large databases [5]. The difficulty of discovering the collocation rules of spatial data is commenced by Shashi Shekhar et. Al. It was followed by consecutive refinement and improvement and given a discrete data model demonstration [6]. In the conceptual representation for the collocation that is planned for non-spatial features of the spatial objects have been discussed, supporting to that, a transformation has been deduced [7]. To support additional experimental work the semantic representation of the data structure to accumulate the collocation has designed [8].

According to the spatio-temporal mining will be the subjective principle that will set up the correlation between the time components and the spatial aspects [9]. But the experiment approved over in this work, is related to the time components that are defined as even with a broader intervals and the spatial knowledge as the spatial objects.

Examples of information stored in the GIS are: patients' particulars, locations of Aedes breeding, larval densities, species of vectors, habitat types, premises types, and ovitrap locations [10]. The GIS enables us to imagine at a glance "hot spots" where cases or breeding are concentrated so that early control operations can be executed. We can also perform spatial and temporal analysis of the data for future planning, such as the review of dengue and cholera sensitive areas and for day-to-day process planning such as the boundary of control operations in outbreak areas, the development of an outbreak etc.

Dengue (pronounced den' gee) the most prevalent Arthropod-borne viral (Arbor virus) belonging to the family Flaviviridae. The major dengue vector in urban areas is Aedes aegypti but Aedes albopticus is also in attendance. It breeds in pond of water [11]. Symptoms include rigorous and continuous pain in the abdomen, the flow of blood from the nose, mouth, high fever, relentless headache, retro-orbital pain, severe joint pain, muscle pains, general weakness, vomiting and cold skin. There is no particular treatment for dengue, but closely medical attention and clinical management saves the lives of several patients [12]. At present, the solitary method of controlling dengue is to conflict with the vector mosquito through chemical control and environmental management. Invasion, collapse and reaction are the three stages of mosquito.

3. Proposed System

Patient database is collected. The disaster that is dengue is recognized using spatial mining techniques. Then the analyzed data are represented on a map based on the time series for all the quarters of a year for the predicted spatial data [13]. The entire work is a threefold procedure. In the first step is to structure the conclusion i.e. collocation with transformations. The second step is to predict the disease affected areas using collocation rule. In the third step is to represent the knowledge of Geographical Information Systems [14].

4. Collocation Identification

Detection of the Epidemic

Risk areas are predicted by implementing collocation rules to the symptoms [15]. The pattern of collocation for the difficulty in the nearby region with high probability is as follows:

C: {cause of epidemic} → { causative agent, infection sources}

Where 'C' is the collocation rule. By implementing collocation rules to the symptoms, the rigorousness of the disease is identified [15]. The collocation rule is applied to the symptoms. High Fever → Retro orbital , Severe Joint Pains → General Weakness, nausea → Rashes, Damage of blood vessels→ Bleeding from the nose, gums or under the skin Hemorrhagic fever → Dengue Death.

5. Spatial Knowledge Representation

The collocation is a fundamental pattern and the collocation rule is the spatial information [3]. As the collocation of our preceding experiments has an agreement with a fuzzy set combination of values for the characteristics, the intensity of a characteristic and its participation grows tremendously. So the collocation cannot be symbolized as simple antecedent and consequent, moderatively a group of features takes part as the antecedent set and consequent group. The antecedent set of the features is granted to have additional weight and the consequent set is having reasonable weight to rely on the organization of the collection. Therefore, the representation has been accomplished in our previous works as a superior-ranking structure [6]. Where the weighted features are coagulated as superior and others are linked to the superior as a group of sets called as the rank of the superior. The feature grouped as in the rank describes the predictability and the well-built relationship of the features in the superior.

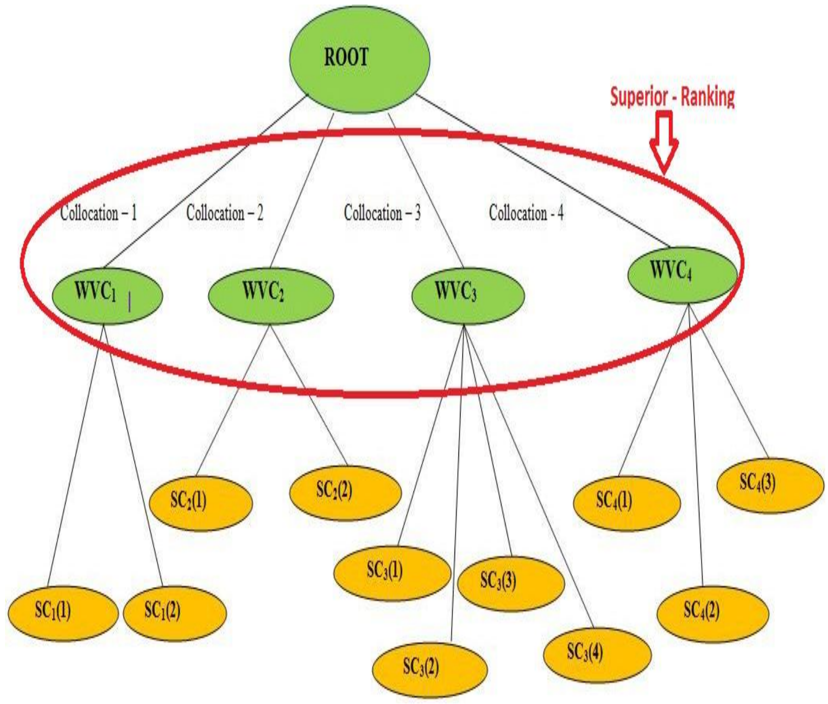


Figure 1. Illustrating the Superior-ranking Structure for Spatial Information Representation

6. Temporal Model

In the area of temporal data mining, the kinds of temporal data being examined is of primary importance for the temporal knowledge discovery procedure.

From the large quantity of literature, it is nonetheless important to remember that a fully temporal database is not necessary for temporal knowledge discovery and that temporal rules can also be resulted from the series of static data sets. Temporal reasoning can be applied to several motionless snapshots of the collocations in order to store numerous rule sets that are later correlated to illustrate conclusions concerning the change in data over time [10].

Though, the data are not reserved in temporal structures; rules describing the change in the data over time can only be resulting indirectly from changes in the stored rule set. The survival of some temporal knowledge can be used to make extraction easier. The reality of calendars is used to segment a pattern, thus making the problem well-mannered. Temporality can be generally classified within the data as static, sequences, time-stamped, fully-temporal.

Various memory optimistic trees and connection optimistic trees have been available which give more financial importance to pile up the data sets. In order to save the disk space, common paths and links between the features are maintained only once, since they are distributed among the structures. The group of structures can be represented as an acyclic graph rather than a collection of independent tree structures [10]. The temporality of the data sets is observed with various important conditions according to the preceding works. The conditions address with respect to the data type and data sets as: (i) maintenance of the data types - assume the collocation storage structures, (ii) support of the time dimension(s) – valid transaction and bi-temporal time components, (iii) mobility of the data sets - with respect to the altering of collocation and the cardinality of the data set through time, namely evolving, growing and full-dynamic.

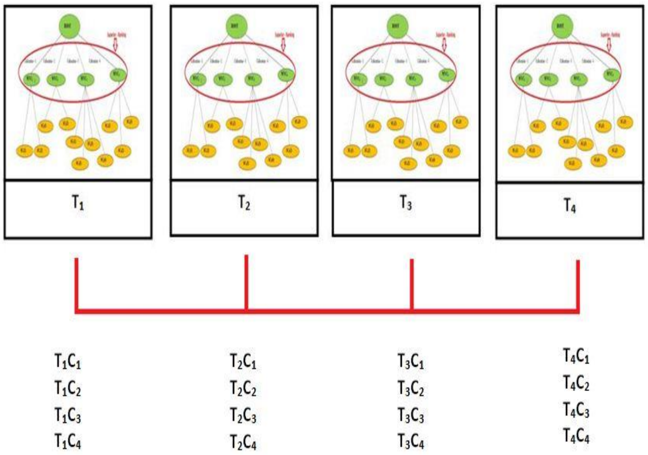


Figure 2. Temporality of Collection of Collocation Instances

7. Algorithm

Using the raw data, the disaster is analyzed and represented in GIS [16]. The algorithm applied is as follows:

1. From the spatial objects, raw data are considered.
2. Preprocessing of noisy data using WEKA tool.
3. Implementation of collocation rules.
4. The predicate "nearby" is applied.
5. The area with disaster is predicted.
6. Representation of knowledge in GIS.

Data Flow Diagram

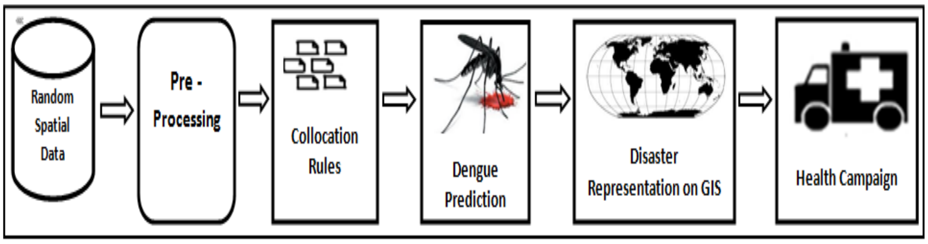


Figure 3. Flow Chart of Spatio-Temporal Representation of Disaster System

8. Identification of Disaster Affected Areas

8.1. Raw Data

Spatial data associated with the spatial objects *i.e.*, patients were collected. Dengue disease comprises of different symptoms like high fever, muscle pain, prostration, swollen lymph nodes, bleeding from nose, nausea *etc*. For the existence of each symptom a binary value "1" is assigned and for nonexistence of each symptom a binary value "0" is assigned. Some noisy data (missing values) were found in the raw data. The collected raw data is as shown in Figure 4.

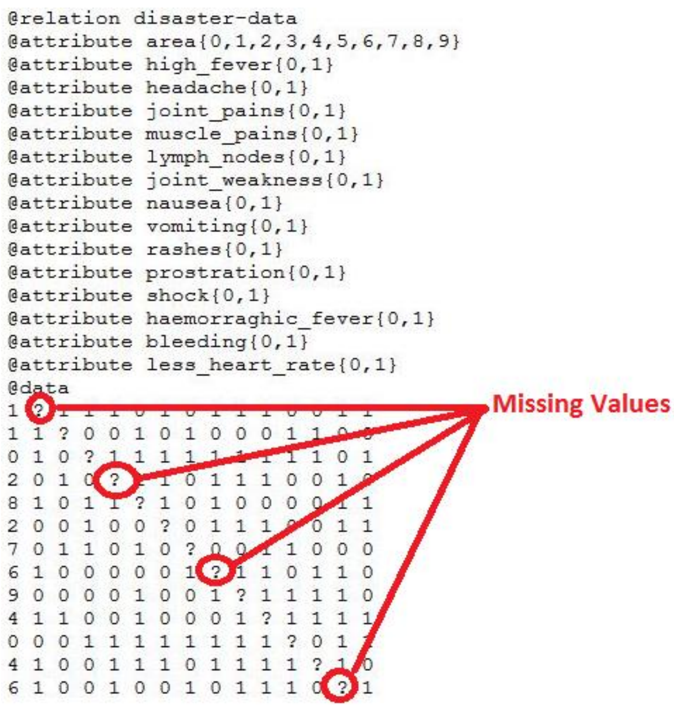
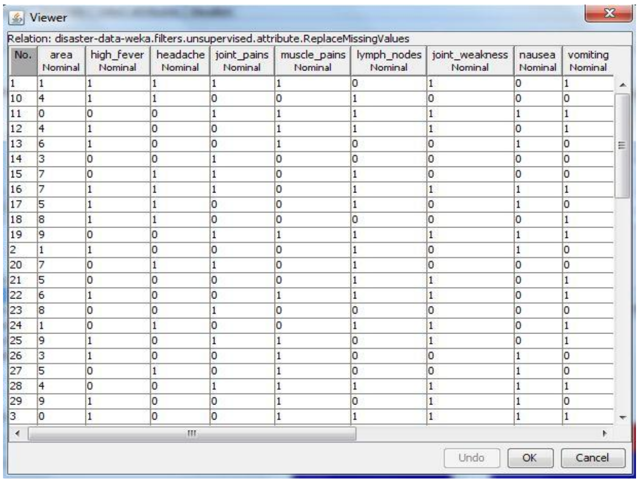


Figure 4. Each Record Refers to One Patient. By the Occurrence of Each Symptom a Binary Value “1” is Read and in the Dearth of Each Symptom a Binary Value “0” Reads in all the Records

8.2. Preprocessed Data

The raw data is preprocessed using the WEKA data mining tool. Missing values are substituted by the mean values or by the highest occurred value. Preprocessed data is shown in table 1 as follows:

Table 1. Preprocessed Spatial Data Using the WEKA Data Mining Tool



Viewer

Relation: disaster-data-weka.filters.unsupervised.attribute.ReplaceMissingValues

No.	area Nominal	high_fever Nominal	headache Nominal	joint_pains Nominal	muscle_pains Nominal	lymph_nodes Nominal	joint_weakness Nominal	nausea Nominal	vomiting Nominal
1	1	1	1	1	1	0	1	0	1
10	4	1	1	0	0	1	0	0	0
11	0	0	0	1	1	1	1	1	1
12	4	1	0	0	1	1	1	0	1
13	6	1	0	0	1	0	0	1	0
14	3	0	0	1	0	0	0	0	0
15	7	0	1	1	0	1	0	0	0
16	7	1	1	1	0	1	1	1	1
17	5	1	1	0	0	1	0	1	0
18	8	1	1	0	0	0	0	0	1
19	9	0	0	1	1	1	1	1	1
2	1	1	0	0	0	1	0	1	0
20	7	0	1	1	0	1	0	0	0
21	5	0	0	0	0	1	1	0	1
22	6	1	0	0	1	1	1	0	1
23	8	0	0	1	0	0	0	0	0
24	1	0	1	0	0	1	1	0	1
25	9	1	0	1	1	0	1	0	1
26	3	1	0	0	1	0	0	1	0
27	5	0	1	0	1	0	0	1	0
28	4	0	0	1	1	1	1	1	1
29	9	1	0	0	1	0	1	1	0
3	0	1	0	0	1	1	1	1	1

Undo OK Cancel

Noisy data (in the raw data) is preprocessed using the WEKA data mining tool.

8.3. Frequent Item Set Generation

Collocation rule is applied to the considered input data. By implementing the apriori algorithm 1-item sets, 2-item sets and 3-item sets were generated.

Spatial Collocation:

Collocation is implemented for point data in space. It does not need any transactions but works directly with continuous space. It uses a neighborhood approach and spatial joins. Participation index is used as a measure for collocation. A group of spatial features that are frequently co-located is described as a spatial collocation. Correctness, completeness and efficiency are the objectives of a spatial collocation.

8.4. Report Generation

The report containing dengue affected areas is as shown below in the Figure 5.

This Report is Forwarded to HEALTH CAMPAIGN..

AREA CODE	AREA	TOTAL	SUPPORT	%AFFECTED
A0	Narsapuram	5	3	60
A1	Bhimavaram	6	3	50
A3	Palakollu	7	1	14
A4	Krishna	7	0	0
A5	Guntur	8	7	87
A6	Prakasham	5	2	40

Figure 5. Disaster Report Containing Percentage of People Affected

9. Visualization of Spatial Data on GIS

Spatio-Temporal Mining

Spatio-temporal mining is a forthcoming research field which comprises of the growth and practice of innovative techniques for the determination of huge spatio-temporal databases [17]. Spatial mining is the mining of previously unknown, veiled spatial data from large amounts of spatial and non-spatial data [18]. Temporal mining is the mining of useful information from temporal data. Both temporal dimensions and spatial dimensions add difficulty to data mining tasks [19]. Our focus is on spatio-temporal tree which describes the spatial data at different time series [20]. Spatial data is a collection of spatial data. In this paper, temporal data include information on spatial data for every quarterly time series. The spatio - temporal tree is drawn for every quarter of a year. Therefore we can visualize the knowledge in GIS for every quarter of a year [20].

10. Results

First Quarter

All the regions in India are visualized in GIS using colored pin points *i.e.*, green, orange, pink, yellow. Green colored pin points indicate that the area is in safe zone, yellow colored pinpoints indicate that the area is in the least dangerous zone, pink colored pinpoints indicate that area is more dangerous zone and red colored pinpoints indicate that area is in the risk zone.

Second Quarter

The changes are observed in the knowledge from Q1 to Q2. Andra Pradesh is changed from most dangerous zone to least dangerous zone. The pinpoint color of Narsapuram is changed from red to yellow in GIS. Palakollu, Bhimavaram and Krishna in GIS changed from most dangerous zone to safe zone. The pinpoint color of Palakollu, Bhimavaram and Krishna are changed from red to green in GIS. The changes from Q1 to Q2 are as shown below in Figure 6 & 7 with red colored arrowed oval symbols.

Third Quarter

The changes are observed in the knowledge from Q2 to Q3. Krishna is changed from safe zone to least dangerous zone. The pinpoint color of Krishna is changed from green to yellow in GIS. Guntur is changed from most dangerous zone to least dangerous zone. The pinpoint color of Guntur is changed from red to yellow in GIS.

Fourth Quarter

Kakinada (which is labelled) is changed from most dangerous zone to least dangerous zone. The pinpoint color of Kakinada in GIS is changed from red to yellow.

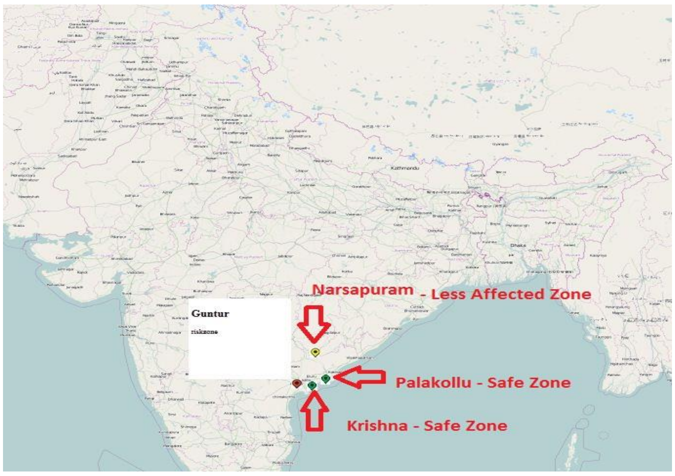


Figure 6. Visualization of Knowledge for the II Quarter of a Year in GIS

Figure 7. Representation of Nuggets for the II Quarter of a Year in GIS

11. Conclusion

In this paper we have acknowledged the need of the data structure to represent the collocation and the spatial knowledge. An epidemic is the multiplying of a disease because of variation in ecology. A collocation rule is applied to the disease symptoms to recognize the epidemics. Identified epidemics are represented in Geographical Information Systems (GIS) for all the quarters of a year. A spatio-temporal tree is proposed. The temporal work is made on the group of such structures and a spatio-temporal coincidence has been improved with the definitions given before scheduling into the experiment. A data structure is designed for the spatial and temporal data. Hence, the spatio-temporal trees along with the visualized nuggets on GIS will be forwarded to the hospitals for conducting health campaigns in the regions which are in the danger zone.

References

[1] H. Pundt, "Evaluating the relevance of spatial data in time critical situations", University of Applied Sciences and Research, Faculty of Automatisation and Computer Science, Wernigerode (DE), **(2005)**, pp. 779–788.

[2] M. N. Rao, M. M. Rao and P. Vedavathi, "Disaster Prediction System using IBM SPSS Data Mining Tool for Cholera", IJCST, vol. 4, no. 2, **(2013)**, pp. 136-140.

[3] M. Yoshioka, Y. Shamoto and T. Tomiyama, "An Application of the Knowledge Intensive Engineering Framework to Building Foundation Design", The International Federation for Information Processing, vol. 33, **(2000)**, pp. 197-212.

[4] R. Naresh, R. Ojha and S. R. Mallik, "Role of internet-based GIS in effective natural disaster management".

[5] A. T. Murray, I. McGuffog, J. S. Western and P. Mullins, "Exploratory Spatial Data Analysis for Examining Urban Crime", **(2001)**.

[6] A. T. Murray, I. McGuffog, J. S. Western and P. Mullins, "Exploratory spatial data analysis techniques for examining urban crime", Br J Criminol, vol. 41, no. 2, pp. 309–329.

[7] S. C. Feng and Y. Zhang, "Conceptual Process Planning - A Definition and Functional Decomposition", Manufacturing Engineering Laboratory, National Institute of Standards and Technology, **(1997)**.

[8] Y. Huang, H. Xiong, S. Shekhar and J. Pei, "Mining Confident Collocation Rules without A support Threshold", Symposium on applied computing, Proceeding of the ACM Symposium on Applied Computing, **(2003)**; Florida.

[9] A. K. Tripathy, J. Adinarayana and D. Sudharsan, "Geospatial data mining for Agriculture pest management - a framework", 17th International Conference on Geoinformatics, **(2009)**; Fairfax, VA.

[10] A. Tung, J. Hou and J. Han, "Spatial clustering in the presence of obstacles", Proceedings of 17th International Conference on Data Engineering, **(2001)**; Heidelberg.

[11] B. He, F. Tao and Z. Guo, "Uncertainty in spatial data mining", Proceedings of International Conference on Machine Learning and Cybernetics, **(2004)**; Xuzhou, China.

[12] J. Zhang, J. Liu and B. Wang, "Spatial cluster based on map algebra", International Conference on E-Business and E-Government (ICEE), **(2010)**; Guangzhou.

[13] K. Polous, A. Farshad, M. Zarinkafsh and M. H. Roozitalab, "Evaluation of the spatio-temporal of soil salinity variation using data mining approach", IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), **(2011)**; Fuzhou.

[14] L. Van Quoc Anh and M. Gertz, "Mining Spatio-temporal Patterns in the Presence of Concept Hierarchies", IEEE 12th International Conference on Data Mining Workshops (ICDMW), **(2012)**; Brussels.

[15] N. Rao and S. R. Babu, "Spatial Knowledge Algorithm For epidemics Using Data Mining Techniques", International conference-ICSCI, **(2008)**; Hyderabad.

[16] B. R. Devi, K. N. Rao, S. P. Settya nd M. N. Rao, "Disaster Prediction System Using IBM SPSS Data Mining Tool", International Journal of Engineering Trends and Technology (IJETT), vol. 4, no. 8, **(2013)**, pp. 3352-3357.

[17] P.-R. Lei, T.-J. Shen, W.-C. Peng and I.-J. Su, "Exploring Spatial-Temporal Trajectory Model for Location Prediction", 12th IEEE International Conference on Mobile Data Management (MDM), **(2011)**; Lulea.

[18] Y. Jun, "Spatial data storage model based on map calculus for online map generalization", IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), **(2011)**; Fuzhou.

[19] Y. Ye, Y. Zheng, Y. Chen and J. Feng, "Xing Xie, Mining Individual Life Pattern Based on Location History", Proceedings of the Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, **(2009)**.

[20] H.-Y. Yi and S.-Z. Chen, “Design and implementation of roadway geological, sketch picture drawing system based on MapGIS”, 2nd International Conference on Information Science and Engineering (ICISE), **(2010)**; Hangzhou, China.

