SPATIAL MINING SYSTEM FOR DISASTER MANAGEMENT

N K Kameswara Rao
Dept. of Information Technology
SRKR Engineering College
Bhimavaram

Dr. G P Saradhi Varma
Dept. of Information Technology
SRKR Engineering College
Bhimavaram

Dr. M. Nagabhushana Rao
Dept. of Computer Science & Engg.
Swarnandra Engineering College
Narasapuram

Abstract —***Information Systems enable us to capture up to date effects due to disaster. It has been widely recognized that spatial data analysis capabilities have not kept up with the need for analyzing the increasingly large volumes of geographic data of various themes that are currently being collected and archived. Our analysis is on disaster management through spatial Maps. Intelligent application algorithms are ideal for finding the rules and unknown information from the vast quantities of computer data. The Intelligence system is to obtain and process the data, to interpret the data, and to design the algorithms for decision makers (Health Companion) as a basis for action. Spatial Map for disaster identification is designed. The Intelligence in each of these algorithms are provided the point and multi-point decision making system to capacitive for evaluation of spreading the dengue. Our contribution in this paper is to design Spatial Maps for*** **Dengue.**

Keywords: Dengue Fever, spatial data mining, Map server, Spatial Rules, Spatial System

I. INTRODUCTION

Dengue virus and dengue hemorrhagic fever are amongst the most important challenges in tropical diseases due to their expanding geographical distribution, increasing outbreak frequency, hyperendemicity and evolution of virulence [Dengue Bulletin – Vol 28, 2004]. Artificial Intelligence (AI), with its various subfields, has a long history of knowledge extraction, representation, and inference in medicine. In dermatology, applications of AI methods, the focus has traditionally been on image analysis and understanding, aimed at providing decision support for physicians. The field of computer-assisted dermatology has thus benefited greatly from advances in knowledge representation techniques and machine learning algorithms.

Recently, increased connectivity and the ubiquitous availability of internet access have resulted in new opportunities for distributed and collaborative diagnosis. Clinical dermatology is mainly a visually dominated discipline. The recognition of signs and symptoms as well as their interpretation of patterns typical for specific diseases remains the core task for diagnosis.

During the last decade computer-assisted applications have proven to be of value for the diagnosis of various forms of skin cancer, especially cutaneous melanoma, Dengue, Malaria, Polio, etc...

The previous work Nagabhushana Rao etl; have proposed Co-location rules for epidemic like Cholera, Dengue[12].The work of Prasanthi. G etl; Have proposed the same with Prims Algorithm but not practical proved with data and Map server screens as output[13]. Here we have with Input data , Algorithm, Output Result and Proved with Map server Screens. Geography is an integrative discipline and geographic data under analysis often span across multiple domains. The complexity of spatial data and geographic problems, together with intrinsic spatial relationships, constitute an enormous challenge to conventional data mining methods and call for both theoretical research and development of new techniques to assist in deriving information from large and heterogeneous spatial datasets. ***(Han and Kamber 2001; Miller and Han 2001; Gahegan and Brodaric 2002).*** Health maps have become available as the use of geographical information systems in health related contexts increased. Many literary research works has been taken place such as [1][4][11].

A hypothesis space is formed by all possible configurations of the tools used to detect patterns in a feature space. This is caused by several factors. First, each pattern may involve a different subset of variables from the original data, and the number of such subsets (hereafter subspaces), i.e., possible combinations of attributes, is huge. Second, inside a subspace, potential patterns can be of various forms (e.g., clusters can be various shapes). Third, for a specific pattern form (e.g., cluster of a specific shape), its parameter space is still huge, i.e., there are many ways to configure its parameters. Fourth, patterns can vary over geographic space, i.e., patterns can be different from region to region.

II. APPLYING SPATIAL DATA MINING

Spatial data mining becomes more interesting and important as more spatial data have been accumulated in spatial databases [9].

A. Spatial Statistics

Using spatial statistics measures, dedicated techniques such as cross *k*-functions with *Monte Carlo simulations, lattice* method have been developed to test the collocation of two spatial features. At the outset the studies include, the spatial data mining problem of how to extract a special type of proximity relationship – namely that of distinguishing two *clusters* of points based on the types of their *neighboring features* is another study[2][6][8]. Classes of features are organized into concept hierarchies [3].A reasonable and rather popular approach to spatial data mining is the use of *clustering techniques* to analyze the spatial distribution of data. While such techniques are effective and efficient in identifying spatial clusters, they do not support further analysis and discovery of the properties of the clusters.

B. Mining Collocation Patterns

Mining collocation patterns gives the standard of observing the generic characteristics of a given spatial zone with more relevant boolean features with their *s%(support)* and *c(confidence)*[6]. The work of mining Collocation patterns into spatial statistics approaches and combinatorial approaches [7]. The spatial Collocation pattern mining framework presented in the erstwhile works has bias on popular events. It may miss some highly confident but "infrequent" Collocation rules by using only "*support*"-based pruning.

In a spatial database S, let F = $\{f_1,\ldots,f_k\}$ be a set of *boolean spatial features*. Let I = $\{i_1,\ldots,i_n\}$ be a set of n instances in the spatial database S, where each instance is a vector consisting of [instance-id, location, spatial features]. ~ *Neighborhood relation R over pair wise locations in S exists* ~ is assumed. The object of this collocation rule mining is to find rules in the form of **A** □ **B**, where **A** and **B** are subsets of spatial features. **A** determines the set of spatial features that form the antecedent part of the rule and **B** defines the action and its consequential parts the support and the confidence. The rule indicates the coincidence of the spatial collocation rule absorbs the action of the rule in the "*nearby*" regions of the spatial objects that comply with the collocation rule. A *collocation pattern* C is a set of spatial features, i.e., C□□□F. A neighbor-set L is said to be a *row* instance of collocation pattern C if every feature in C appears in an instance of L, and there exists no proper subset of L does so. We denote all row instances of a collocation pattern C as *rowset*(C). In other words, *rowset*(C) is the set of neighbor-sets where spatial features in C collocate. The conditional probability is the probability that a neighbor-set in *rowset*(A) is a part of a neighbor-set in *rowset*□(□□B). Intuitively, the conditional probability *p* indicates that, whenever we observe the occurrences of the spatial features in A, the probability to find the occurrence of B in a nearby region is *p*.

C. Finding/Estimating Symptoms To Build Collocations

Since 1998, we have developed and made use of the PC-based geographical information system (GIS) to manage the huge databases on cases and Aedes mosquitoes island-wide. Examples of information stored on the GIS are: patients' particulars, locations of Aedes breeding, larval densities, species of vectors, habitat types, premises types, and ovitrap locations[3]. The GIS enables us to visualize at a glance "hotspots" where cases or breeding are concentrated so that early control operations can be implemented (Figures-1is an example). We can also perform spatial and temporal analyses of the data for future planning, such as the review of dengue and cholera sensitive areas; and for day-to-day operation planning such as the boundary of control operations in outbreak areas, the progression of an outbreak, etc.

The majority of houses have a cement water container located in the bathroom to store water for bathing and a smaller container in the water closet (WC). Water containers made from clay or plastic barrels/jars are also kept in the kitchen for cooking or drinking purposes. Additional water containers may act as potential breeding sites both inside and outside houses. The people utilizing breed in pools of water.

A few epidemics that are spread due to common sources like contaminated water and contaminated food are shown below.

Common Source Epidemic Diseases			
Disease	**Causative Agent**	**Infection Sources**	**Reservoirs**
Bacillary	Shigella disenteriae (B)	Fecal contamination of food and water	Humans
Dengue	Arthropod-borne virus	Breed in pools of water	Humans

Table-1. The table clearly explains about the causative agent

Sources and reservoirs of the disease

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 present. It breeds in pools of water [13]. Only female can transmit the virus. Female mosquitoes can transmit the virus to the next generation of mosquitoes. Symptoms include severe and continuous pain in the abdomen, bleeding from the nose, mouth , skin bruising, frequent vomiting with or without blood, black stools like coal tar, excessive thirst, pale, cold skin. There is no specific treatment for dengue, but closely medical attention and clinical management saves the lives of many patients. At present, the only method of controlling dengue is to combat the vector mosquito through chemical control and environmental management. Remove tires, bottles, cans and other items that catch and retain water, eliminating potential breeding sites for vector mosquitoes.

The disease proceeds in possibly three stages:

(a) Invasion (b) Collapse (c) Reaction

III. PROBLEM

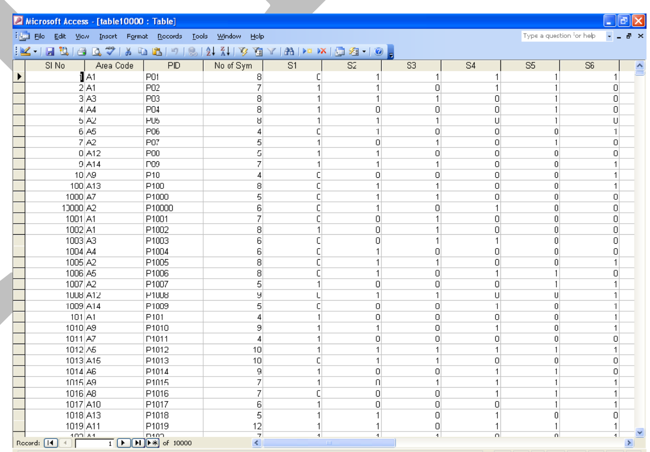
A. Detection of the Epidemic

The possibility that dengue outbreaks result from anomalous patterns of precipitation, we analyzed the relationships linking rainfall, the abundance of vector mosquitoes, the degree of source-reduction coverage, and the occurrence of dengue during the couple of year period of observation[16]. The collocation rules are very useful in detecting the affected areas by finding the symptoms of a disease and influence of symptoms in a disease by using sample identifiers, the collocation can be explained as follows: Assuming firstly, the '**b**' as the consequence of feature '**a**' is developed, forms a first level of collocation, which is identified by **a** → **b**, secondly, if the consequence '**c**' from the feature '**b**' is developed, forms a collocation, which is identified by **b** → **c**. As '**b**' already have an antecedent '**a**', the consolidated version of collocation, {**a, b**} → **c** can be formed. If '**c**' becomes another feature that can lead to the consequence of '**d**', then the notation wholly represents the cause of '**d**' as {**a, b, c**} → **d.** Also implies to {**a** ∪ **b** ∪ **c**} → **d** representation**.**

Similarly, considering the collocation pattern for the problem:

***C:** {cause of epidemic} {causative agent, infection sources}; in the <u>nearby</u> region with high probability.*

IV. ALGORITHM



Sl No	Area Code	PID	No of Sym	S1	S2	S3	S4	S5	S6
1	A1	P01	8	0	1	1	1	1	1
2	A1	P02	7	1	1	0	1	1	0
3	A3	P03	8	1	1	1	0	1	0
4	A4	P04	8	1	0	0	0	1	0
5	A2	P05	8	1	1	1	0	1	0
6	A5	P06	4	0	1	0	0	0	1
7	A2	P07	5	1	0	1	0	1	0
0	A12	P00	5	1	1	0	0	0	0
9	A14	P09	7	1	1	1	0	0	1
10	A9	P10	4	0	0	0	0	0	1
100	A13	P100	8	0	1	1	0	0	1
1000	A7	P1000	5	0	1	1	0	0	0
10000	A2	P10000	6	0	1	0	1	0	0
1001	A1	P1001	7	0	0	1	0	0	0
1002	A1	P1002	8	1	0	1	0	0	0
1003	A3	P1003	6	0	0	1	1	0	0
1004	A4	P1004	6	0	1	0	0	0	0
1005	A2	P1005	8	0	1	1	0	0	1
1006	A5	P1006	8	0	1	0	1	1	0
1007	A2	P1007	5	1	0	0	0	1	1
1008	A12	P1008	9	0	1	1	0	0	1
1009	A14	P1009	5	0	0	0	1	0	1
101	A1	P101	4	1	0	0	0	0	1
1010	A9	P1010	9	1	1	0	1	0	1
1011	A7	P1011	4	1	0	0	0	0	0
1012	A5	P1012	10	1	1	1	1	1	1
1013	A15	P1013	10	0	1	1	0	0	0
1014	A6	P1014	9	1	0	0	1	1	0
1015	A9	P1015	7	1	0	1	1	1	1
1016	A8	P1016	7	0	0	0	1	1	1
1017	A10	P1017	6	1	0	0	0	1	1
1018	A13	P1018	5	1	1	0	1	0	0
1019	A11	P1019	12	1	1	0	1	1	1

Table-2: patient's information

The following algorithm is to find the spatial knowledge i.e. dengue disaster from health demographic data.

1. Data collection from the patients.
2. Attributes are selected.
3. Collocation rule is applied.
4. Spatial predicate is applied.

Source (Area) of disaster identified

Results :

Count of PID: Total no of symptoms
Total PID: Confidence Value
Con: % of Effected dengue in each area

ISSN 2320 –5547

From the Output, we can say that near spatial object (Area) **"A10"** people are having more chances to get dengue (spreading of dengue).

Microsoft Access - [Con3(10000) : Select Query]

File Edit View Insert Format Records Tools Window Help

Area Code	CountOfPID	Total PID	Con
A10	790	503	64
A6	1013	624	62
A2	588	365	62
A15	596	372	62
A13	365	223	61
A12	524	318	61
A9	1297	782	60
A8	685	410	60
A7	905	542	60
A5	989	597	60
A14	603	359	60
A11	436	263	60
A4	296	175	59
A3	296	175	59
A1	617	363	59

Table-3: dengue effected areas

V. SPATIAL MAPS FOR DENGUE

Non Spatial features of the spatial object were collected from the hospitals in the respective places, i.e. cities / villages. Each place is given Identification (id). For each patient Non-Spatial feature of the spatial data i.e. symptoms were taken.

Our own proposed Algorithms for dengue was applied through that we have got the count of persons that were affected by Dengue were obtained in below Figure 1.

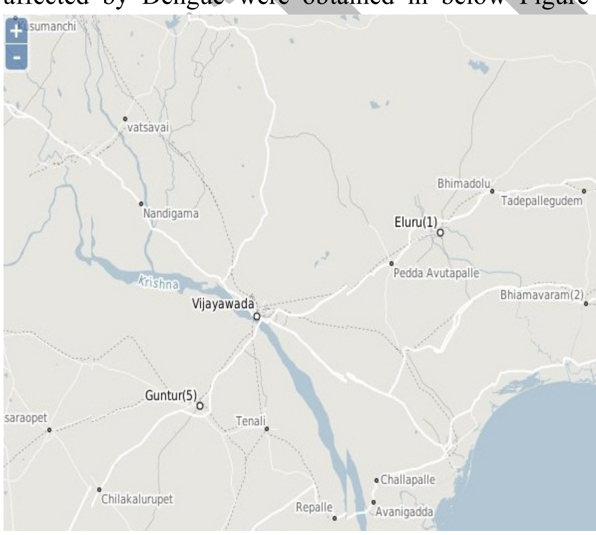


Fig.1Represents Dengue effected persons in different places

VI. CONCLUSION

Epidemics, chronic diseases which are the major social disasters follow strategic-virulent disasters that affect the ecosystem of a spatial zone probabilistic study is made on the health demographic data. Using the Collocation rule, the effected area of dengue is found and results are obtained. Using the Participation index the symptoms that influence spreading the dengue were found and results are obtained. i.e., spatial knowledge for the health campaign. Identifying dengue effected area using spatial maps as spatial knowledge i.e. to identify spreading of dengue was proposed and proved. Dengue effected Areas were shown on Map.

REFERENCES

[1] Alan T. Murray, Ingrid Mc Guffog, John S. Western and Patrick Mullins, Exploratory Spatial Data Analysis for Examining Urban Crime, 2001. www.gegraphy.hunter.cuny.edu

[2] Bavani Arunasalam, Sanjay Chawla, Pei Sun and Robert Munro, Mining Complex Relationships in the SDSS SkyServer Spatial Database, School of Information Technologies, University of Sydney, proceedings of 28th annual international computer software and application conference (CONPSAC-04), 2004, IEEE.

[3] Chawla, Shekkar, Spatial Databases: A Tour, 2003, Prentice all. (ISBN – 013 – 017 480 – 7)

[4] Hardy 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), pp 779– 788, 2005, www.springer.com/3-540-249988-5.

[5] Huang, Shekkar, Xiong, Discovery Collocation Patterns from Spatial Data Sets: A General Approach, IEEE-KDE,volume 16,No:12, dec 2004.

[6] Knorr and Ng, Extraction of Spatial Proximity Patterns by Concept Generalization, proceedings second international conference of KDD, pp. 347 – 350, aug 1996.

[7] Koperski and Han, Discovery of Spatial Association Rules in GI Databases, proceedings 4th international symposium large spatial databases, pp. 44-66, aug 1995.

[8] Munro, Chawla, Complex Spatial Relationships, proceedings of IEEE international conference on data mining, 2003.

[9] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Spatial Data Mining: A Database Approach, Institute for Computer Science, University of Munich, proceedings of 5th international symposium of large special databases, 2007.

[10] Masaharu Yoshioka, Yasuhiro Shamoto, Tetsuo Tomiyama, an Application of the Knowledge Intensive Engineering Framework to Architectural Design, a research project of GNOSIS, IMS Program, 1998.

[11] Naresh Raheja, Ruby Ojha, Sunil R Mallik, Role of internet-based GIS in effective natural disaster management,R.M.S.I. www.development.net/technology/gis/

[12] Nagabhushana Rao, Ramesh babu, Sangameswar, Spatial Knowledge Algorithm For epidemics Using Data Mining Techniques. International conference-ICSCI-2008, Hyderabad, Jan 2008.

[13] Prashanthi.G, Rajkumar.M,etl, Identification of Epidemic through Spatial Mining and Visulisation with Mapserver. IJERT, Issue 7, Vol:1, Sept-2012.

[14] Shaw C. Feng, Y. Zhang: Conceptual Process Planning - A Definition and Functional Decomposition, Manufacturing Engineering Laboratory, National Institute of Standards and Technology 1997.

[15] New Initiatives in Dengue Control in Singapore byTan Boon Teng. Dengue Bulletin – Vol 25, 2001

[16] Vector Densities That Potentiate Dengue Outbreaks In A Brazilian City, Ricardo J. S. Pontes, Jonathan Freeman, Jose´ Wellington Oliveira-Lima, J. Christina Hodgson, AndAndrew Spielman, Am. J. Trop. Med. Hyg., 62(3), 2000, pp. 378–383

