Network Architecture to Identify Spatial Knowledge for Dengue

N K KAMESWARA RAO
Assoc.Professor, It Dept.
Srkr Engineering College

Dr. G P SARADHI VARMA
Professor & Hod, It Dept.
Srkr Engineering College

Dr.M.NAGABHUSHANA RAO
Professor, Cse Dept.
Swarnandra Engg. College

Abstract: **Recent developments in information technology have enabled collection and processing of vast amounts of personal data, business data and spatial data. 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 study is carried out on the way to provide the mission-goal strategy (requirements) to predict the disaster. The co-location rules of spatial data mining are proved to be appropriate to design nuggets for disaster identification and the state-of-the-art and emerging scientific applications require fast access of large quantities of data. Here both resources and data are often distributed in a wide area networks with components administrated locally and independently, a framework has been suggested for the above. Our contribution in this paper is to design network architecture for disaster identification.**

Keywords: Spatial data mining, Algorithms, Network architecture, Web based architecture, Components.

I. INTRODUCTION:

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][2][3].

A formula implemented as Hazard science to Risk Science, towards understanding the hazards and their consequences (risks), following a probabilistic approach using spatial data mining [1].Due to larger heterogeneity of spatial data, the providers of geographic data specify different models for same spatial objects. Context specific semantics is one of the best approach suggested which deals with provision of feature space derivations. Unknown and unexpected patterns, trends or relationships can hide deep in a huge feature space.

A hypothesis space is formed by all possible configurations of the tools used to detect patterns in a feature space. Characteristically, however, the hypothesis space for a large and high dimensional geographic dataset has an extreme degree of complexity. and make it very hard for analytical methods or visual approaches to find. *(Miller and Han 2000).*

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 [4].

2.1. 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[5][6][7]. Classes of features are organized into concept hierarchies [8].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.

2.2. 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 [9]. 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** $\rightarrow$ **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 $\subseteq$ 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* (A$\cup$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*.

2.3. FINDING/ESTIMATING SYMPTOMS TO BUILD COLLOCATIONS

In the imaginary figure the landscape describes two important spatial marks, sea and lake. The epidemic spread is noted in the **Figure-1.** The water in the lake is afflicted by the lichens and mosses at the western zone of the lake as most of the water is stagnant and covered by marsh. The people utilizing the water resources at this zone may be affected by so many kinds of fecal contamination in water and food. The water in the sea is contaminated with the high salts and the crude oil and base products, as the people cannot take the water for the domestic purposes. As it is not easy to detect the spread of virulent features from the spatial data, the features are tested in the affected people who are native of the zone.

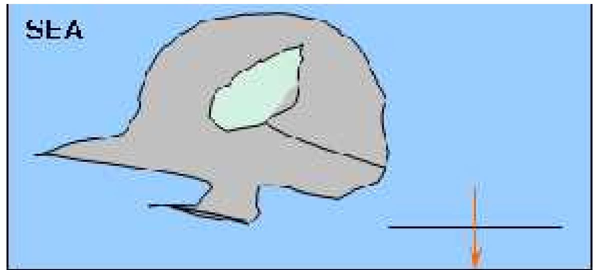

Figure-1: Lake affected by lichens & mosses

III. RELATED WORK

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	*RNA flavivirus or DENV virus*	*infected mosquitos or monkeys*	*Humans*

Table-1. The table clearly explains about the causative agent, Sources and reservoirs of the disease

3.1. COURSE OF THE DISEASE

Dengue (also called Break bone *fever*) is an tropical disease caused by the dengue virus. Dengue is transmitted by several species of mosquito within the genus *Aedes*, principally *A. aegypti*. Symptoms include Sudden, high fever,Severe headaches,Pain behind the eyes,Severe joint and muscle pain,Nausea,Vomiting,Skin rash, which appears three to four days after the onset of fever,Mild bleeding (such a nose bleed, bleeding gums, or easy bruising). Other things that are found are loss of appetite and bladder problems.If left untreated it can cause death.

These following are the stages of Dengue Disease

Stage1: Acute fever and upper respiratory symptoms

Stage2: Fever subside

Stage3: Circulatory failure, neurological problems, and hemorrhaging

Stage 4: Shock and death (10% of all cases reach this stage)

IV. THE LAW OF TOTAL PROBABILITY

Although there are many solutions to prevent diseases, finding the right area to apply the prevention measure with right inputs becomes the criterion. The Bayes' theorem evaluates the reverse of conditionality of events; where the symptoms and the causative-agents are analyzed and found with a reciprocal equivalence. The **Table-1** describes the most probable symptoms that cause the epidemics. The fact that the person had a positive reaction to the test may be considered as our data to build the collocation pattern.

The conditional probability of the collocation is the probability that a neighbor-set explaining the features of *existence of causative agent*, *infection sources*, is a part of the global neighbor-set in the *spatial domain* for this epidemic application Given a *spatial domain* in a database view ***S***, to measure the implication strength of a spatial feature in a collocation pattern, *a participation ratio Pr(C,f)* has to be defined. A feature *f* has a participation ratio *Pr(C,f)* in pattern *C* means whenever the feature *f* is observed, with probability *Pr(C,f)*, all other features in *C* are also observed in a neighbor-set. In spatial application domain, as there are no natural transactions, for a continuous space, a *participation index* is proposed to measure the implication strength of a pattern from spatial features in the pattern. For a collocation pattern *C*, the participation index *PI(C)* = min $_{f€C}$ {*Pr(C, f)*}. In other words, wherever a feature in C is observed, with a probability of at least *PI(C)*, all other features in C can be observed in a neighbor-set. A high participation index value indicates that the spatial features in a collocation pattern likely show up together [10].

Let us consider a collocation pattern:

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

V. PROBLEM

<u>*Detection of the Epidemic*</u>

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** Ć **b** Ć **c**} → **d** representation**.**

Similarly, considering the collocation pattern for the problem:

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

The collocation pattern is considered with practically proved parameters for dengue as follows …

C1: { *Fever, headache, Myalgia or bone pains, skin rash, nausea, Vomiting* } ⇒ { *Bleeding, low-levels of blood platelets, Low BP (dengue shock)* }

<u>Probability</u> *<excreted along with innumerable Vibrios with high probability>*

C2: { ***C1,*** *Bleeding, low-levels of blood platelets, Low BP* } ⇒ { *Fluid accumulation in chest & abdominal cavity leakage* }

<u>Probability</u> *<loss of fluid, electrolyte imbalance with high probability>*

C3: { ***C1, C2,*** *fluid accumilation in chest & abdominal cavity leakage* } ⇒ { *Depletion of fluid from circulation & decreased blood supply to vital organs* }

As the disease reaches "Collapse" stage, the circulation is almost completely arrested, accelerated respiration, weak pulse, decreased systolic blood pressure, diminished or no urine output.

C4: { ***C1, C2, C3,*** *decreased blood supply to vital* } ⇒ { ***Dengue Shock Syndrome*** }

<u>Probability:</u> *<the <u>loss</u> of fluid, electrolyte <u>imbalance</u> with high probability>*

C5: { **X, infected mosquitos or monkeys, DENV virus** } ⇒ { ***Dengue Shock Syndrome*** }

Assuming ***X*** as defined representation of collocated sequence of patterns i.e., C_1, C_2, C_3, C_4 the resultant collocation C_5 is determined.. If the lead feature of the collocation contains higher probability then collocation is considered as highly important. The probabilities mentioned in the problem are <*excreted along with innumerable Vibrios*>, <*loss of fluid, electrolyte imbalance*>. If one of them or some of them exhibit high probability, then there is a high significance of occurring the disease severely, for low exhibition of probability, the existing of the disease will be indicative.

VI. ALGORITHM

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

1. Data collection from the patients.
2. Attributes related to cholera are selected.
3. Collocation rule is applied.
4. Spatial predicate is applied.
5. Source (Area) of disaster identified.

Synthesized data is generated using the rand () function. Parameters of cholera are chosen to work out from database.

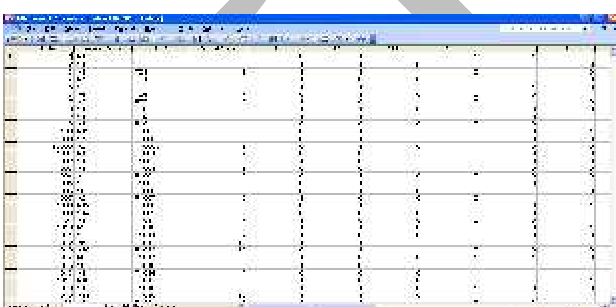

Table-2: patient's information

6.1 RESULTS

With threshold Value of 40%

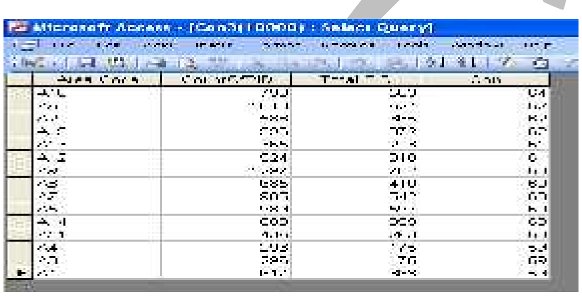

Table-3: Dengue effected areas

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

VII. NETWORK ARCHITECTURE

The built of framework explains the elements of the spatial knowledge support system in a *work flow strategy* and *component architecture strategy* [11][12].

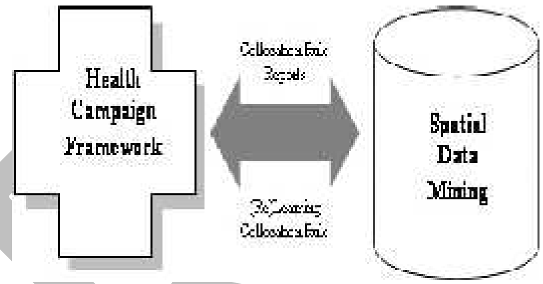

Figure-2: Components of Frame work

The above conceptual design contains two important components, the spatial data mining infrastructure and the health campaign framework. The figure shows the detail process design framework of work flow strategy. The collocation pattern formed by this sample region acts as a cautious measure or the forecast for the bio-medical researchers, analysts and other health-care-takers of the spatial zone which will be useful for them to take suitable remedial campaigns.

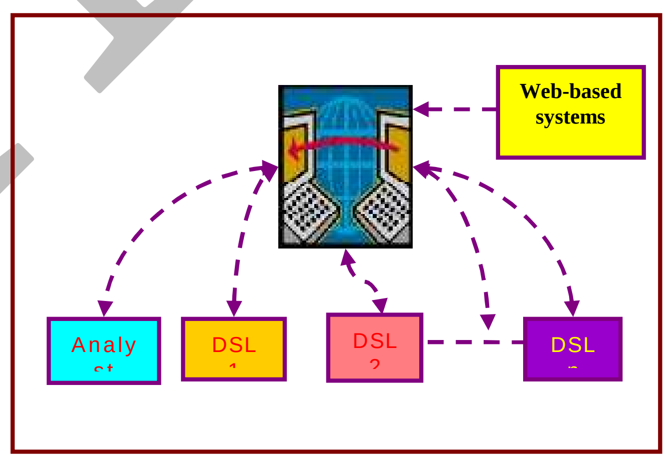

Figure-3: Web-based Architecture for disaster Management

In Figure-3 web–based architecture was proposed for processing epidemic. DSL's are nothing but stations across different locations, connected to central server through web.

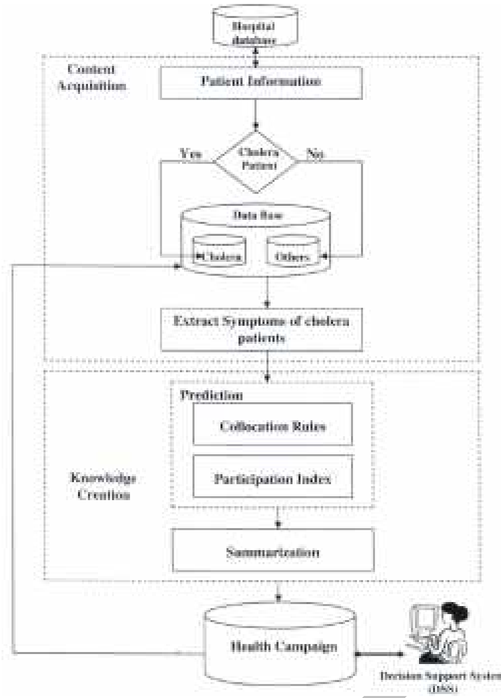

Figure-4: **Architecture of Epidemic System**

Figure - 4 explains data flow of the system. In figure-5 proposed Client Server Architecture was given for the epidemic system [10].

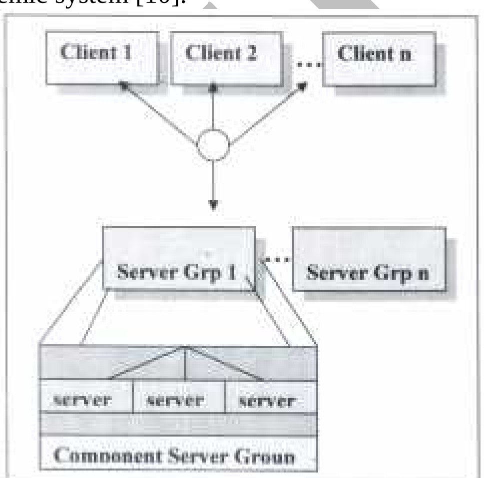

Figure-5: Client-Server Software Architecture and its interacting components [Courtesy from Rainbow Architecture]

Socio-statistical methods related to health-science can be implemented to regulate the input variables that play a parametric role of collocation rule formation, in order to prevent the epidemic in the spatial zone, if not permanently, at least suitable preventive measures can be undertaken for the affect of such candidate epidemic in the interested spatial zone.

VIII. 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. A Collocation rule is defined as a syntactic representation of the parameters in the form of antecedent and consequent. Using the Collocation rule, the effected area of cholera is found and results are obtained. Framework is described for the application of collocation rules i.e., spatial knowledge for the health campaign. **Network architecture** for spatial knowledge i.e. to identify cholera was proposed.

IX. ACKNOWLEDGEMENTS

Our sincere thanks to management, Principal and faculty members of our institute (Sagi Ramakrishnam Raju Engineering College).

REFERENCES

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

[2] Hardy Pundt, Evaluating tyhe 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.

[3] 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/

[4] 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.

[5] 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.

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

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

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

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

[10] Ojha Ruby,Naresh Raheja, Sunil R Mallik, Report on “Role of internet-based GIS in effective natural disaster management”, R. M. Software India Pvt. Ltd. (RMSI), 2004. www.gisdevelopment.net/technology/gis/techgi0030.htm

[11] 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.

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

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

AUTHORS

N K Kameswara Rao did M Tech in Software Engineering from JNT University-Hyderabad. Presently working as Associate Professor in Information Technology Department, SRKREC – Bhimavaram.

Dr. Saradhi Varma did Ph.D in computer Science & systems Engineering from Andhra University-visakhapatnam. Presently working as Professor & HOD in Information Technology Department, SRKREC – Bhimavaram.

Dr. M. Nagabhushana Rao did Ph.D in Spatial Data Mining from Sri Venkatateswara University-Tirupathi. Presently working as Professor in CSE Department, SCET – Narsapuram.

