Identification of Epidemic Dengue through Spatial Mining and Visualization with Map Server

*G.Prasanthi, #J.Maha Lakshmi,# M.Raj Kumar,#N.Sridhar Babu
Swarnandhra Institute of Engineering & Technology,Narsapur,# Krishna chaitanya inst.of engg.& technology,markapur,#Pragati engineering college,#Surampalem,Spanco Private Limited,Chandigar

Abstract

This paper develops the Identification of Epidemic Dengue through Spatial Mining and Visualization with Map Server. to search the disease with the help of the open maps just like the Google maps. On this paper apply the mathematical method. We propose a novel algorithm for co-location pattern mining which materializes spatial neighbour relationships with no loss of co-location instances and reduces the computational cost with the aid of the Prim's Algorithm. The spatially co-location mining algorithm is proficient since it generates and filters the candidate instances.

1.0 INTRODUCTION:

The main difference between data mining in relational DBS and in spatial DBS is that attributes of the neighbours of some object of interest may have an influence on the object and therefore have to be considered as well. The explicit location and extension of spatial objects define implicit relations of spatial neighbourhood (such as topological, distance and direction relations) which are used by spatial data mining algorithms.

1.1 OBJECTIVE:

Therefore, new techniques are required for effective and efficient data mining. We propose a novel algorithm for co-location pattern mining which materializes spatial neighbour relationships with no loss of co-location instances and reduces the computational cost with the aid of the Prim's Algorithm. The spatially co-location mining algorithm is proficient since it generates and filters the candidate instances. Subsequently, the neighbourhood relationships are carried out by the designed neighbourhood and the node membership functions which satisfy the minimum conditional threshold.

2.0 RESEARCH METHODOLOGY

2.1 Types of research :

Quantitative vs qualitative:

Quantitative research is based on the measurement of quantity or amount. this is applicable to phenomena that can be expressed in terms of quality, amount, size, or weight etc. Quantitative research on the other hand is concerned with the qualitative phenomena, i.e ,something relating to or involving quality. For example , research designed to find out how people feel of what they think about a particular subject or institution is qualitative research.

2.2 MATHEMATICAL MODELLING:

2.2.1 CLASSIFICATION BASED ON MATHEMATICAL PROPERTY:

LINEAR VS NON LINEAR: Mathematical models are composed of variables, which are abstractions of quantities of interest in the system described and operators that act on those variables. Operators includes algebraic operators , functions ,differential operators, etc.if all the operators in the mathematical model present linearity ,the resulting mathematical model is defined as linear , otherwise it is considered as nonlinear. Example of linear model are

$$L=ax+by$$

Where ‘a’ and ‘b’ are constents . and x and y are variables.

$$L=a(d^2x/dt^2) +b(dx/dt) + cx$$

Where a, b, c are constants

The question of linearity and nonlinearity depends on the context, and linear models may have nonlinear expressions in them. For example , differential equation is said to be linear if it can be written with linear differential operators, but still it can have nonlinear expression in it.

A common approach to nonlinear problem is linearization, but this can be problematic if one is trying to study aspects such as irreversibility which are strongly tied to nonlinearity.

2.3DETERMINISTICVS PROBABILISTIC:

A DETERMINISTIC model is a one in which every set of variables state is uniquely decided by parameters in the model and by set of previous states of this variables. Therefore deterministic models performes the same way for a given set of initial conditions. Conversely ,in a probabilistic or stochastic models , randomness in present , and variables states are not described by in unique values , but rather by probabilistic distributions.

2.4 STATIC VS DYNAMIC:

A static model does not account for the elements of time , while a dynamic model

does. Dynamic models are typically represented with difference equations or differential equations.

2.5LUMPEDVSDISTRIBUTED PARAMETERS:

If a model is homogeneous (consistent state throughout the system), its parameters are distributed. On the other hand if a model is heterogeneous (varying state within the system), then the parameters are lumped. Distributed parameters are typically represented with the partial differential equations.

2.6 CLASSIFICATION ON THE BASIS OF TRANPARENCY

A PRIORI/SUBJECTIVE INFORMATION - mathematical modeling probles are often classified into black box and white box models . according to how much a priori information is available of the system. A black box model is a system about which there is no priori information available , where as a white box model(glass box) is a system where all necessary information is available practically all systems are somewhere between the black and white box models. It is always preferable to use as much a priori information as possible to make the model are accurate. Therefore , white box models are usually consider easier , because if you have used the information correctly then the model behave correctly.

In black box models one tries to estimate both the functional form of relations between variables and the numerical parameters in those functions. an often used approach for black-box models is neutral network which usually does not make assumptions about incoming data. Sometimes it is useful incorporate subjective information into a mathematical model. This can be done based on the intuition, experience or expect opinion or based on convenience of mathematical form. Bayesian statistics provides a theoretical framework for incorporating such subjectivity into a rigorous analysis : one specifies a priori probability distribution first(which can be subjective) and then update this distribution based on empirical data.

3.0 Survey

3.1 Epidemic

Epidemic is a derivation of two Greek word epic (upon / among) and demos (people). It is the 'unusual' occurrence in a community or region of a disease specific health related events "clearly in excess" of the "expected occurrence". Thus, any disease, which occurs in numbers more than the excepted occurrence, constitutes an epidemic. It includes heart diseases, or even psychosomatic disorders. An epidemic may be restricted to one location, however if it spreads to other countries or continents and affects a substantial number of people, it may be termed a pandemic. The declaration of an epidemic usually requires a good understanding of a baseline rate of incidence; epidemics for certain diseases, such as influenza, are defined as reaching some defined increase in incidence above this baseline A few cases of a very rare disease may be classified as an epidemic, while many cases of a common disease (such as the common cold) would not.

Epidemic diseases may afflict large numbers of persons living in a particular geographical area. Symptoms of an epidemic disease also spread rather rapidly among many, but not all, persons in the same community. Epidemic diseases are frightening because of the number afflicted and because of the rapid spread of the disease. Some diseases, such as dengue fever and measles, disable victims for some days, but usually do not cause death. Others, such as cholera and yellow fever, cause a varying number of deaths. When particular groups have not been exposed to a disease vector, the outcome may be devastating, as appears to have been the case among the Plains Indians of Texas when they were exposed to the communicable diseases carried by the white settlers.

The origins of the word "dengue" are not clear, but one theory is that it is derived from the Swahili phrase *Ka-dingapepo*, which describes the disease as being caused by an evil spirit. The Swahili word *dinga* may possibly have its origin in the Spanish word *dengue*, meaning fastidious or careful, which would describe the gait of a person suffering the bone pain of dengue fever.However, it is possible that the use of the Spanish word derived from the similar-sounding Swahili. Slaves in the West Indies having contracted dengue were said to have

the posture and gait of a dandy, and the disease was known as "dandy fever".

Spatial Data Mining

It's a Search for spatial patterns. Non-trivial search – as "automated" as possible. Large search space of plausible hypothesis Ex. Asiatic cholera : causes water, food, air, insects. Interesting, useful, and unexpected spatial patterns. Useful in certain application domain Ex. Shutting off identified water pump => saved human lives. May provide a new understanding of the world Ex. Water pump – Cholera connection lead to the "germ" theory. The main difference between data mining in relational DBS and in spatial DBS is that attributes of the neighbors of some object of interest may have an influence on the object and therefore have to be considered as well. The explicit location and extension of spatial objects define implicit relations of spatial neighborhood (such as topological, distance and direction relations) which are used by spatial data mining algorithms. Therefore, new techniques are required for effective and efficient data mining.

Spatial trends describe a regular change of non-spatial attributes when moving away from certain start objects. Global and local trends can be distinguished. To detect and explain such spatial trends, e.g. with respect to the economic power, is an important issue in economic geography. Another important task of economic geography is to characterize certain target regions such as areas with a high percentage of retirees. Spatial characterization does not only consider the attributes of the target regions. New algorithms for spatial characterization and spatial trend analysis were developed.

For spatial characterization it is important that class membership of a database object is not only determined by its non-spatial attributes but also by the attributes of objects in its neighborhood. In spatial trend analysis, patterns of change of some non-spatial attributes in the neighborhood of a database object are determined. In efficient algorithms were proposed to mine spatial co-location patterns from spatial databases. When the number of objects of different spatial features spans a wide range, the popular features (features with a large number of objects) tend to get a low ratio compared to rare features (features with a small number of instances).

In spatial co-location patterns were generalized and expressed by multi-way spatial joins. The space partitioning algorithms were proposed to solve the spatial co-location pattern mining problem. The proposed algorithm is not restricted to a particular interesting measure.

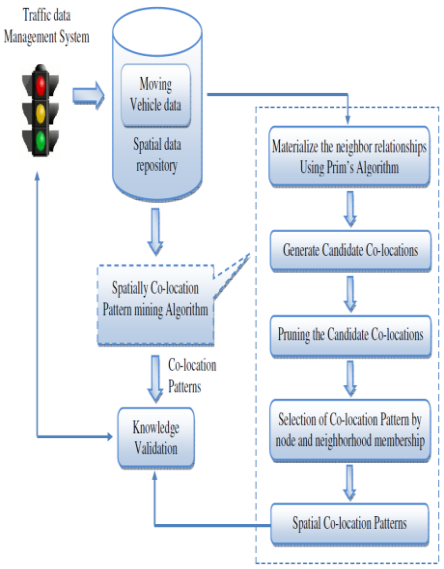
Co-Location

A *co-location pattern* is a group of spatial features/events that are frequently co-located in the same region. For example, human cases of West Nile Virus often occur in regions with poor mosquito control and the presence of birds. For collocation pattern mining, previous studies often emphasize the equal participation of every spatial feature. Advanced spatial data collecting systems, such as NASA Earth's Observing System (EOS) and Global Positioning System (GPS), have been accumulating increasingly large spatial data sets . For instance, since 1999, more than a terabyte of data has been produced by EOS every day. These spatial data sets with explosive growth rate are considered nuggets of valuable information.

The automatic discovery of interesting, potentially useful, and previously unknown patterns from large spatial datasets is being widely investigated via various spatial data mining techniques. Classical spatial pattern mining methods include spatial clustering, spatial characterization, spatial outlier detection, spatial prediction , and spatial boundary shape matching. Mining *spatial co-location patterns* is an important spatial data mining task. *A spatial co-location pattern is a set of spatial features that are frequently located together in spatial proximity*. In ecology, scientists are interested in finding frequent co-occurrences among spatial features, such as drought, EI Nino, substantial increase/drop in vegetation, and extremely high precipitation.

The overall block diagram of the proposed spatially co-located pattern mining algorithm is given as below in Figure. 1.

Figure 1: Block Diagram of the Proposed Algorithm



Geospatial Data Mining

The advent of remote sensing and survey technologies over the last decade has dramatically enhanced our capabilities to collect terabytes of geographic data on a daily basis. However, the wealth of geographic data cannot be fully realized when information implicit in data is difficult to discern. This confronts GIScientists with an urgent need for new methods and tools that can intelligently and automatically transform geographic data into information and, furthermore, synthesize geographic knowledge. Knowledge discovery (KD) technology empowers development of the next generation database management and information systems through its abilities to extract new, insightful information embedded within large heterogeneous databases and to formulate knowledge. A KD process includes "data warehousing, target data selection, cleaning, preprocessing, transformation and reduction, data mining, model selection (or combination), evaluation and interpretation, and finally consolidation and use of the extracted knowledge".

The term geostatistics is often associated with continuous geographic space, whereas the term spatial statistics is often associated with discrete space. In a statistical model that handles nonspatial data, one usually assumes statistical independence among different portions of data. However, different from traditional data sets, there is no such independence among spatially distributed data because in reality, spatial objects are often interrelated, or more exactly spatially co-located, in the sense that the closer the two objects are located, the more likely they share similar properties. For example, nature resource, climate, temperature, and economic situations are likely to be similar in geographically closely located regions.

There is currently a good deal of interest in geospatial data as a rich source of structure and pattern, making it ideal for data mining research (e.g. Koperski & Han,1995; Ester et al., 1996, 1998; Knorr & Ng, 1996; Koperski et al, 1999; Roddick & Spiliopoulou, 1999). Many of the very large consumer, medical and financial transaction databases currently being constructed contain spatial and temporal attributes and hence offer the possibility of discovering or confirming geographical knowledge (Miller & Han, 2000). For decision makers this knowledge represents improved decision power.

Three characteristics of geospatial data create special challenges to development of a robust data foundation. The characteristics that make geospatial data .special. as a computing problem have been acknowledged in many other writings, of course. Moreover, development of a data infrastructure needed to support GIScience in general forms a focus in another UCGIS initiative (spatial data infrastructure). Let us point out that the focus here is not on developing the spatial data infrastructure *per se* but on developing data mining within the emerging infrastructure. As argued below, the research problems solved by generating a solid data foundation can be shown to create the need for new developments in data mining and knowledge discovery.

Below Figure represents the geo spatial representation GIS map

Dengue fever

Dengue fever is a disease caused by a family of viruses that are transmitted by mosquitoes. It is an acute illness of sudden onset that usually follows a benign course with symptoms such as headache, fever, exhaustion, severe muscle and joint pain, swollen glands (lymphadenopathy), and rash. The presence (the "dengue triad") of fever, rash, and headache (and other pains) is particularly characteristic of dengue. Other signs of dengue fever include bleeding gums, severe pain behind the eyes, and red palms and soles.Dengue (pronounced DENG-gay) can affect anyone but tends to be more severe in people with compromised immune systems. Because it is caused by one of four serotypes of virus, it is possible to get dengue fever multiple times. However, an attack of dengue produces immunity for a lifetime to that particular serotype to which the patient was exposed.Dengue goes by other names, including "breakbone" or "dandy fever." Victims of dengue often have contortions due to the intense joint andmuscle pain, hence the name breakbone fever. Slaves in the West Indies who contracted dengue were said to have dandy fever because of their postures and gait.Dengue hemorrhagic fever is a more severe form of the viral illness. Symptoms include headache, fever, rash, and evidence of haemorrhage in the body. Petechiae (small red or purple splotches or blisters under the skin), bleeding in the nose or gums, black stools, or easy bruising are all possible signs of hemorrhage.

This form of dengue fever can be life-threatening and can progress to the most severe form of the illness, dengue shock syndrome.Dengue fever is a disease caused by a family of viruses that are transmitted by mosquitoes.Symptoms such as headache, fever, exhaustion, severe joint and muscle pain, swollen glands (lymphadenopathy), and rash. The presence (the "dengue triad") of fever, rash, and headache (and other pains) is particularly characteristic of dengue fever.The acute phase of the illness with fever and myalgias lasts about one to two weeks.Dengue hemorrhagic fever (DHF) is a specific syndrome that tends to affect children under 10 years of age. It causes abdominal pain, hemorrhage (bleeding), and circulatory collapse (shock).The prevention of dengue fever requires control or eradication of the mosquitoes carrying the virus that causes dengue.There is currently no vaccine available for dengue fever.***Dengue Points to remember***

Dengue haemorrhagic fever (DHF) is a **severe form of Dengue** fever caused by infection with more than one Dengue virus. The patient suffering from Dengue haemorrhagic fever develops **bleeding from nose, gums or skin.** Sometimes, the patient may have **coffee colored vomiting or black stools**. This indicates bleeding in gastro intestinal tracts which is serious. Rarely, the patient suffering from dengue may manifest rapid & weak pulse, low blood pressure, cold clammy skin and restlessness (manifestations of **shock**), then it is called Dengue shock syndrome (DSS).Dengue should be suspected when there sudden onset of high grade fever (103-105 degrees F or 39-40 degrees C).

The fever is accompanied with severe headache (mostly in the forehead), typically pain behind the eyes, body aches and joint pains, rash on the skin and nausea or vomiting. The fever usually lasts for 5-7 days. All the above symptoms and signs may not be present in the patient.If a patient suspected to be having dengue has **reduced platelet* count** (less than one lac/cubic mm) or an **increase in blood hematocrit*** (20% or more, than base line value), then the patient is considered to be a probable case of Dengue. Patients with dengue may not have a high hematocrit, if the person was anemic (low hemoglobin) to start with.*Platelets help to stop bleeding and Haematocrit indicates

hemoconcentration (i.e. thickness of blood).Aedes aegypti is can be easily recognized by its peculiar **white spotted body** and legs. It is a domestic mosquito which rests indoors, in closets and other dark places. Outside, it rests where it is cool and shaded. The female mosquito lays her eggs in water containers in and around the homes, and other dwellings. These eggs will develop, become larvae, and further develop into adults in about 10 days. It is to be noted that the mosquito is **day biter** and the bite is likely to go unnoticed as it does not cause deep skin irritation/ eruption like other mosquito bites.Dengue can only spread through the bite of an infected mosquito.

The symptoms of Dengue develop generally **after 4-6 days of infection**, There may be 4-5 persons with no symptoms (or with very mild symptoms), for every patient with signs and symptoms of Dengue. The signs and symptoms that are considered to be serious include **severe pain in the abdomen**, **persistent vomiting**, bleeding in the skin appearing as small **red or purplish spots**, bleed through nose, bleeding from gums, passage of black stools like coal tar. As soon as severe pain in the abdomen and persistent vomiting are detected, it is better to bring the patient to the hospital. Like most viral diseases there is no specific treatment for Dengue. Paracetamol is the drug of choice to bring down fever and relieve joint pain. Other medicines such as Aspirin and Ibuprofen should be avoided because of increased risk of bleeding. Antibiotics are usually not required. Most people who suffer from Dengue fever recover in 1-2 weeks time and usually there is no risk of death. Sometimes, the infection can become dangerous since it may cause damage to the blood vessels causing leakage of blood fluid/plasma into various organs (condition known as Dengue Hemorrhagic Fever or Dengue Shock Syndrome).

Good treatment provided in time can save most lives. Being affected by one strain offers no protection against the others. Therefore, it is possible to get dengue more than once. Due to increased tendency of viruses to mutate into more serious forms, it is imperative to have Dengue vaccines at the earliest. As yet, there are no vaccines available in the market. Strong interactions involving rare spatial features are often marked off in previous methods, since they require frequent co-occurrences of all features in the co-location patterns. Many measures are based on the measures of *frequency or minimum participation ratio* where rare events are unfavourable.

Bibliography

1. Spatial data mining : an emerging tool for policy makers- sanjay chawala shashi shekhar,Li wu,tan CURA Reporter, 2000 september.
2. Spatial databases Accomplishments and Research needs. Shashi shekhar,sanjay chawala,siva rawada IEEE,1999 January.
3. Theoretical framewors for data mining Heikki Mannila,ACM SIGKOD,2003 January.
4. A.Framework for Diagnosing changes in evdring data streams, Charu.C.Aggarwal ,,ACM SIGMOD,2003 june.
5. An Interitive framework for understanding changes in endruing data streams, Charu.C.Aggarwal ,ACH conference,1999.
6. Discovery spatial co-location patterns a summary of results, Shashi Shekhar,Yan Huary,7th Symposium,2000.
7. Fast mining of spatial collocation,Xin Zhang,Nikas,IEEE-KDD,2004 August.
8. Spatial Knowledge for rural development,Dr.P.Govindharajulu,Dr. M.N.Rao,proceedings of 41st annual convention,CSI-2006 Nov,Kolkata.