Multi-Resolution Pruning Based Co-Location Identification In Spatial Data

Sruthi K Surendran , S. Dhamodaran

Computer Science and Engineering Department ,
Sathyabama University,
Chennai 600119,Tamilnadu ,India.

Abstract*: A co-location spatial pattern is a pattern of multiple groups which co-relates spatial features or events that are frequently located in same zone. Co-location pattern mining emphasizes overall analysis by manipulating the proportion of spatial features and other relevant information's. In this paper we are trying to remove the problem of co-location mining from the large data set. Data is generated from wide range of data sources that are available in geographical space. One method to mine the widely spread co-location is to compute the participation index in measuring the prevalence of the co-location. First option from the above said the measure is closely inter-related to cross function, which often used for measuring the statistical among various pair of spatial features. Second option focused on the property of anti-monotones which can be included for computational perspective efficiency. In this paper, we are trying to incorporate a novel multi-resolution pruning technique to address the problem of mining co-location data pattern with rare spatial features.*
Keywords: *Spatial colocation,spatially uncertain data set, possible world, probabilistic prevalent colocation(PPCs)*

I.   INTRODUCTION

Spatial data mining is an application of data mining methods in spatial data. The main objective of spatial data mining is to found pattern in data with respect to geography. Spatial data mining is a process of finding interesting and useful patterns from spatial data that are generated from geographic space. It is very difficult to extract patterns from spatial data than transaction data because; it is due to the complexity of spatial data types, spatial relationship and autocorrelation.

Pattern mining is data mining method that find existing pattern in data .A spatial co-location pattern is a pattern which represents a subset of spatial features whose instance frequently located in close geographic proximity. A data set consist of instance of several Boolean spatial features, each represented by a distinct shape . Data input of spatial data mining is complex than input of transaction data mining. Spatial data mining include extended object such as point, lines and polygons. There are two different type of attributes used in spatial data mining as data input. They are non-spatial attributes and spatial attributes. In non-spatial attributes, they are used to characterize non spatial features of objects like name, population and unemployment rate of cities. In spatial attributes they are used to define the spatial location and spatial object. It includes longitude, latitude, elevation, slop etc. A careful reveals two co-location pattern. Real-world examples of co-location pattern include symbiotic species, e.g. the Nile, Crocodile and Egyptian plover in ecology.

Boolean spatial features describe the presence and absence of geographic object types at different location in a two dimensional or three dimensional metric space , such as surface of the earth. Example of Boolean spatial features include plant species, animal species, road type, cancer, crime, and business type.

Advanced spatial data collecting system, such as NASA Earth's Observing System (EOS) and Global Positioning System (GPS), has been accumulating increasingly large spatial data set. For insatance, since 1999, more than a terabyte of data has been produced by EOS every day. These spatial data set with explosive growth rate are considered nuggets of valuable information.

The automatic discovery of interesting, potentially useful, and previously unknown pattern from large spatial data set is being widely investigated via various spatial data mining technique. Classical spatial pattern mining method includes spatial clustering, spatial characterization spatial outlier detection, spatial prediction, and spatial boundary shape matching.

The problem of mining spatial co-location patterns can be related to various application domain. For example, in location based services, different services are requested by service subscribers from their mobile PDA's equipped with location devices such as GPS. Some type of services may be requested in proximate geographic area such as finding the nearest Italian restaurant and nearest parking place. Location based service providers are very interested in finding what services are requested frequently together and located in spatial proximity. . This information can help them improve the effectiveness of their location based recommendation system where users request a service in a nearby location. By knowing co-location pattern in location based

services may enable the use of pre-fetching to speed up service delivery. In ecology, scientists are interested in finding frequent co-occurrences among spatial features such as drought, substantial increase or drop in vegetation and extremely high precipitation.
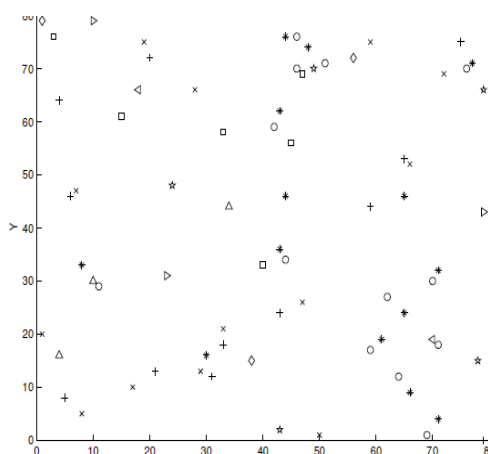

Fig.1 Colocation pattern illustration

       The previous studies on co-location pattern mining emphasize frequent co-occurrences of all the features involved. These marks off some valuable pattern involving rare spatial features We say a spatial feature is rare if its instances are substantially less than those of the other features in a co-location. This definition of "rareness" is relative with respect to other features in a co-location. A feature could be rare in one co-location but not rare in another. For example, if the spatial feature A has 10 instances, the spatial feature B has 20 instances, and the spatial feature

II.   Related Work
       Extraction of knowledge from huge amount of spatial data is a big challenge. Many technologies are available for knowledge discovery in large spatial database .Different type of data mining approaches are available, mainly extraction of implicit knowledge, spatial relation and other pattern which are not explicitly stored in spatial data base. Spatial data is categorized based on some rules that discovered in spatial data bases.

       A spatial characteristic rule is a description of set spatial-related data[9]. For example general description of whether pattern in a set of geographical region. A spatial discriminate rule is a description of contrasting features of class of spatial related data from other class. For example comparison of whether pattern in different geographic region. A spatial association rule describes implication of a set of features by another set in spatial data base. For example big city in Canada are close to Canada U.S border.

       Various kind of prediction constitutes a spatial association rule [8]. For example distance information like close-to or faraway. For finding spatial co-location a special case of mining spatial association is used. So association rule is one of data mining technique which helps to find frequent item. Association rule mining algorithms finite set of disjoint transaction are given as input to the algorithm. Algorithms like apriori is used to find from those frequent item set. Another approach can be based on the choice of reference spatial features to mine all association rules where at least one of the predicates is a spatial predicates. In fact association rule is derived using the appriori algorithm, But this approach does not find more general co-location pattern involving no reference spatial feature on the left hand side of association rule.

       A data mining form proposed by Agarval [2] for fault-zone recognition for flexible ac transmission (FACT) based transmission line together with thyristor-controlled series compensator (TCSC) and integrated power flow controller (UPFC) by mean of ensemble decision trees. The algorithm used here is experienced on replicated slipup data with extensive variation in service parameters of the supremacy system network; counting noisy environment provided a reliable measure with quicker response time. The performance accuracies that have been assessed in noisy environment is an critical issues in power system applications and also the improvement is required in individual trees that are noisy and unstable based on low bias.

       A multitask spectrum-sensing process is proposed by Hung [5] on the basis of spatiotemporal data mining ways. Within every cluster the intension of spectrum sensing is executed in a synchronized technique. They used DP-based MT-CS technique to cluster to cluster the remark from various clusters that may not allocate one common sparseness spectrum.But they need to maintain spectrum sensing performance and stable quality of service.

A data mining based method proposed by Aggarwal [1] for plastic bearing fault diagnostics by means of vibration sensors. The frequency domain CIs which used through a statistical classification model to recognize bearing outer race faults. The time domain CIs extracted using EMD are urbanized to build a k-nearest neighbor algorithm-based fault classifier to spot other type of bearing faults. But only first four IMF components were used because the first four components contained the majority of the energy of the inventive signals

III. Proposed Work

Here we introduce more accurate method to find the wide spread co-location which allow us to evaluate uncertain data in probabilistic way. In common, the problem of finding expected co-locations in spatially uncertain data is more difficult than finding frequent item set in transaction data. Uncertain data are essential in many applications and are usually described by specific probability. It is unproblematic to obtain precise probabilities from uncertain data in a application.

In this paper we study the problem of pulling out of co-location from spatially uncertain data with probability interval. It defines possible world model with prospect intervals and proves that probability intervals of all possible world are feasible. Based on applicable probability interval, it converts the probability interval of possible world in to point probabilities. Then it defines the interrelated concept of probabilistic wide spread co-location. Then it gives two lemmas for optimizing the computation of pre dominance point likelihood of a contender co-location.

Then it proves the closure possession of prevalence point probability. Finally, experiment on real and synthetic data set to prove that the algorithms are effective and significant. Here they provide a probabilistic data model which can be used from spatial co-location mining in spatially uncertain data. From this data model the expected prevalent co-location and probabilistic prevalent co-location can define. The uncertain data model used is based on the possible world semantic with the existential uncertain instances. From the possible world semantic with the existential uncertain

instances, the expected prevalent co-location can be introduced. After that it allows us to evaluate co-location in probabilistic way.
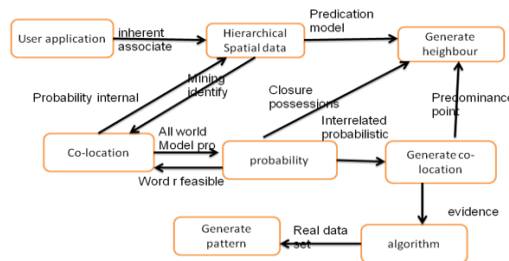

Fig 2:Data flow diagram

SPATIAL INFORMATION INPUT
Spatial means pertaining to or involving or having the nature of space. Spatial data is any data with a direct or indirect reference to specific location or geographical area.
There are two type of attributes present in input set of spatial data mining. They are spatial attributes and non-spatial attributes. Non-spatial attributes include whether condition, soil type and minerals available in particular location. Spatial attributes include latitude, longitude, slope, elevation of particular location, Spatial data is the modern means of digitally mapping features on the earth. Spatial data is used in geographic information systems (GIS) which merge cartography, statistical analysis and database techniques. In this module , the spatial information in the form of latitude and longitude which are gathered from the user as input. The data set generated contain set of spatial feature and their locations.

CO-LOCATION MINING IDENTIFICATION
Co-location are the neighborhood location of particular location under examination. In this, based on the location's latitude and longitude the neighborhood location are mined. The mining of the data takes place in term of distance relation in miles and kilometers. Miles and kilometers are the metric which is used to calculate the distance. Here if we are giving a particular location needed to know with there latitude and longitude value, it give the list of neighborhood locations. Thus identify the particular location. The co-location pattern discovery process finds subset of features that are frequently located together.

CO-LOCATION PREDICTION
A prediction is a statement about the way thing will happen in the future, often but not always thing will happen in the future, often but not always based on the experience or knowledge. According the

neighborhood location identification, the prediction about the location are formulated. Here the co-location is predicted with its best and mediate resources available. Here we implement not only the co-location mining but also there is a prediction of the available resource in the land within the neighborhood possibilities.

FILTERING AND INPUT VALIDATION

Filtering is the process here to segregate the co-location based on the available resources from the co-location prediction. Land resources can be taken to mean the resources available from the land. The agricultural land which contain natural fertilizer for the growth of the products, the underground water, the various minerals like coal, bauxite, gold , and other raw materials. Here the process of the filtration is carried out based on the type of input that defines the resources required in the particular location there by validating the input.

TOP-K PREFERABLE LIST GENERATION

Top-k products are widely applied for retrieving a ranked set of the k most interesting objects based on the preferences. Here reference means the selecting an item or land for decision or consideration.. Here a selection technique is used to define a list for the co-location with the available resources there by defining the severity of the resource available.

PERFORMANCEEVALUATION-CO-LOCATION VALIDATIONS

Performance is characterized by the amount of useful work accomplished by an application compared to the time and the resources used. Here the validation of the co-location with its available resources are checked for the accuracy. The validation of the co-location by the system are visualized graphically.
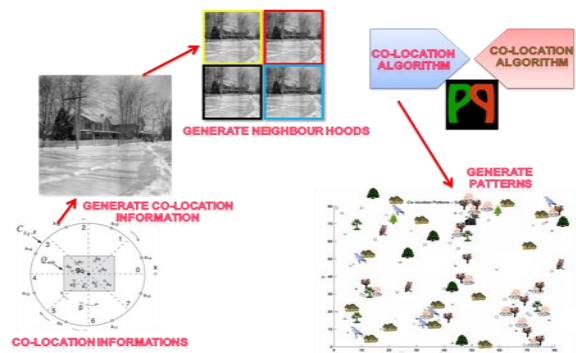

Fig:3 Architecture Diagram

SPATIAL CO-LOCATION MINING ALGORITHAM

Input: (a) E= {Event-ID, Event-Type, Location in Space } representing a set of events
                 (b) ET= {Set of K Boolean spatial event type }
     (c) Spatial relationship R
            (d)Minimum prevalence threshold θ and conditional probability threshold α
                Note that participation index is measure of prevalence

Output: A set of co-locations with prevalence and
                conditional probability values greater than
                user-specified minimum prevalence and
                conditional probability threshold .
Variables: k: co-location size
                     C_K: set of candidate size-k co-location
                   T_K: set of table instances of co-locations in C_K
                   P_K: set of prevalent size-k co-locations
                    R_K: set co-location rule of size-k
                   T-C_K:set of coarse-level table instances of
            size-k co-location in C_K
Steps:
  1. Co-location size k=1; C1=ET; P1=ET;
  2. T1= generate_table_instance(C1, E);
  3. If(fmul=TRUE) then
  4. T-C1= generate_table_insatance(C1,multi_event(E));

5. Initialize data structure $C_K, T_K, P_K.R_K, T-C_K$ to be empty for $1 < k \le K$;
6. While(not empty P_K and $k > K$) do {
7. C_K+1=generate_candidate_colocation(C_K, K);
8. If(fmul=TRUE) then
9. C_K+1=multi_resolution_pruning($\theta, C_K+1, T-C_K$, multi_rel(R));
10. T_K+1=generate_table_insatance(θ, C_K+1, T_K, R);
11. P_K+1=select_prevalent_colocation(θ, C_K+1, T_K+1);
12. R_K+1=generate_colocation_rule(α, P_K+1, T_K+1);
13. $K=k+1$;

IV. Conclusion And Future Enhancement

This paper studies the problem of pulling out co-location from spatially uncertain data with probability intervals. It has to defined the possible world model with prospect intervals, and proves that probability intervals, and proves that probability intervals of all possible world are feasible. It has also defined the interrelated concept of probabilistic widespread co-locations. Then it has proved the closure possessions of prevalence point probability.

Here the predictions of the available resource in the land take place with the neighborhood possibilities. In our future enhancement according to the land resource the life of the plant and other living things life which lis based on the wealth of the land is predicted.

REFERENCES

[1] C.C. Aggarwal et al., "Frequent Pattern Mining with Uncertain Data," Proc. 15th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 29-37, 2009.
[2] P. Agrawal, O. Benjelloun, A. Das Sarma, C. Hayworth, S. Nabar,T. Sugihara, and J. Widom, "Trio: A System for Data, Uncertainty, and Lineage," Proc. Int'l Conf. Very Large Data Bases (VLDB),pp. 1151-1154, 2006.
[3] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proc. Int'l Conf. Very Large Data Ba ses (VLDB),pp. 487-499, 1994.
[4] C.C. Aggarwal and P.S. Yu, "A Survey of Uncertain Data Algorithms and Applications," IEEE Trans. Knowledge and Data Eng. (TKDE), vol. 21, no. 5, pp. 609-623, May 2009.
[5] C.-K. Chui and B. Kao, "A Decremental Approach for Mining Frequent Itemsets from
[6] Uncertain Data," Proc. 12th Pacific-Asia Conf. Knowledge Discovery and Data Mining (PAKDD), pp. 64-75, 2008.
[7] T. Bernecker, H-P Kriegel, M. Renz, F. Verhein, and A. Zuefle, "Probabilistic Frequent Itemset Mining in Uncertain Databases," Proc. 15th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD '09), pp. 119-127, 2009.
[8] M. Ester, H.-P. Kriegel, and J. Sander, "Knowledge Discovery in Spatial Databases," Proc. 23rd German Conf. Artificial Intelligence (KI '99), (Invited Paper), vol. 1701, pp. 61-74, 1999.
[9] K. Koperski and J. Han, "Discovery of Spatial Association Rules in Geographic Information Databases," Proc. Int'l Symp. Large Spatial Databases (SSD '95), pp. 47-66, 1995.
[10] Y. Huang, S. Shekhar, and H. Xiong, "Discovering Co-Location Patterns from Spatial Data Sets: A General Approach," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 12, pp. 1472-1485, Dec. 2004.
[11] Huang, Y., Pei, J. and Xiong, H. "Mining Co-location Patterns with Rare Events from Spatial Data Sets", Geoinformatica 10:239-260.2006
[12] R. Ng and J. Han "Efficient and effective clustering method for spatial data mining." In Proc Int. Conf. VLDB. Pp-144-155, Santiago. Chile, Sept 1994.