Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 38 (2012) 2832 – 2845

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference on Modeling Optimization and Computing-ICMOC-2012

Discovering Co-location Patterns from Spatial Domain using a Delaunay Approach

Venkatesan Meenakshi Sundaram[a],Arunkumar thnagavelu[b], Prabhavathy Paneer[c]

[a] *Assistant Professor,SCSE, Vellore institute of Technology ,Vellore-14,Inida*
[b] *Professor,SCSE,Vellore institute of Technology,Vellore-14,India,*
[c] *Assistant Professor,SITE,VIT University,Vellore-14,Inida*

Abstract

Spatial co-location patterns represent the subsets of boolean spatial features whose instances are often located in close geographic proximity. Spatial statistics and data mining approaches are used to identify co-location patterns from spatial data sets. Spatial proximity is the important concept to determine the co-location patterns from massive data sets. A Delaunay diagram based co-location mining approach is developed to mine co-location patterns from spatial data by using the concept of spatial proximity. Delaunay diagram is used to model the spatial proximity between the objects. This approach eliminates the parameters from the user to define neighborhood of objects and avoids multiple test and trail repetitions in the process of mining. An algorithm to discover co-location patterns are designed which generates candidate locations and their table instances. Finally the co-location rules are generated to identify the patterns. The results of the experiments have been discussed.

Keywords: Co-location,Proximity,Delaunay,Association,Membership Ratio,Spatial

1. Introduction

Very large amount of Geo-spatial data leads to definition of complex relationship, which creates challenges in today data mining research. Geo-spatial data can be represented in raster format and vector format. Raster data are represented in n-dimensional bit maps or pixel maps and vector data information can be represented as unions or overlays of basic geometric constructs, such as points, lines and polygons. Spatial data mining refers to the extraction of knowledge, spatial relationships, or other interesting patterns not explicitly stored in spatial data sets. As family of spatial data mining, spatial Co-location pattern detection aim to discover the objects whose spatial features/events that are frequently co-located in

1877-7058 © 2012 Published by Elsevier Ltd.
doi:10.1016/j.proeng.2012.06.332

the same region. It may reveal important phenomena in a number of applications including location based services, geographic information systems, geo-marketing, remote sensing, image database exploration, medical imaging, navigation, traffic control and environmental studies. Some types of services may be requested in proximate geographic area, such as finding the agricultural land which is nearest to river bed. Location based service providers are very interested in finding what services are requested frequently together and located in spatial proximity. The co-location pattern and the rule discovery are part of spatial data mining process. The differences between spatial data mining and classical data mining are mainly related to data input, statistical foundation, output patterns, and computational process

Co-location rules[12] are models to infer the presence of boolean spatial features in the neighborhood of instances of other boolean spatial features. Co-location rule discovery is a process to identify co-location patterns from large spatial datasets with a large number of boolean features. The spatial co-location rule discovery problem looks similar to, but, in fact, is very different from the association rule mining problem [2] because of the lack of transactions. Association rule mining is transaction based and transactions are defined around instance of special spatial feature. It uses spatial predicates as item types. But decomposing spatial data into transactions may alter patterns. So usage of co-location mining increases the efficiency of finding the interesting patterns from the very large spatial data. Co-location patterns are discovered by using neighborhood definition and spatial joins. This paper discusses the detection of co-location pattern from the complex Geo-Spatial data by using event centric model approach

Geographical data is represented by three basic topological concepts- the point, the line, and the area. Despite of the fact that the point is the most primitive one, it is not easy to define point proximity as a discrete relation. The spatial analysis of discrete non-connected objects (points) has been approached using distance concepts in the traditional vector and raster models. The Voronoi diagram[16] has been proposed as an alternative way to model proximity amongst points, overcoming the limitations of conventional data models First, it represents topology (spatial adjacency) explicitly. If two points share a Voronoi edge, then they are treated as neighbors. This adjacency information is so fundamental that constitutes the dual known as the Delaunay triangulation [17, 18, 19]. Delaunay triangulation possesses a number of unique properties, which can be successfully used for spatial proximity. Delaunay diagram, constructed by removing all ambiguous diagonals in the Delaunay triangulation, is the base of edge proximity analysis. Delaunay diagram is a structure representing the neighborhood of objects in a succinct and unambiguous manner. In this structure there is no doubt which objects are neighbors. It is a significant advantage in comparison with means of determining neighborhood in spatial data mining methods described in the literature so far. Definitions of neighborhood used there are ambiguous.

This paper is structured as follows: Section 2 discuses existing methods available to discover neighborhoods and co-location pattern approaches. Section 3 describes the use of Delaunay diagram to model the spatial proximity. Section 4 discusses the methodology, the proposed system design and co-location algorithm to discover co-location patterns. Section 5 deals Computing of spatial Co-location rules from the spatial patterns . Section 6 summarizes the performance analysis and comparison our approach with the existing methods and Section 7 discusses the conclusions and future enhancements of the proposed system

2. Literature Review

Approaches to discovering co-location rules in the literature can be categorized into three classes, namely spatial statistics, data mining, and the event centric approach. Spatial statistics-based approaches use measures of spatial correlation to characterize the relationship between different types of spatial features using the cross-K function with Monte Carlo simulation and quadrant count analysis. Computing spatial correlation measures for all possible co-location patterns can be computationally expensive due to the exponential number of candidate subsets given a large collection of spatial boolean features. Data mining approaches can be further divided into a clustering-based map overlay approach and association rule-based approaches. Association rule-based approaches can be divided into transaction-based approaches and distance-based approaches Association rule-based approaches focus on the creation of transactions over space so that an apriori like algorithm [2] can be used. Transactions over space can use a reference-feature centric [3] approach or a data-partition approach [4]. The reference feature centric model is based on the choice of a reference spatial feature and is relevant to application domains focusing on a specific boolean spatial feature, e.g., incidence of cancer. Domain scientists are interested in finding the co-locations of other task relevant features to the reference feature [3]. Transactions are created around instances of one user specified reference spatial feature. The association rules are derived using the apriori[2] algorithm. The rules found are all related to the reference feature.

Defining transactions by a data-partition approach [4] defines transactions by dividing spatial datasets into disjoint partitions. There may be many distinct ways of partitioning the data, each yielding a distinct set of transactions, which in turn yields different values of support of a given co-location. The conditional probability for the co-location rule is: spatial feature A at location l → spatial feature type B in neighborhood is 100%.This yields a well-defined prevalence measure (i.e., support) without the need for transactions. A clustering-based map overlay approach [7], [6] treats every spatial attribute as a map layer and considers spatial clusters (regions) of point-data in each layer as candidates for mining associations. Given X and Y as sets of layers, a clustered spatial association rule is defined as X=>Y (CS,CC%), for $X \cap Y = \emptyset$, where CS is the clustered support, defined as the ratio of the area of the cluster (region) that satisfies both X and Y to the total area of the study region S, and CC% is the clustered confidence, which can be interpreted as CC% of areas of clusters (regions) of X intersect with areas of clusters (regions) of Y.

A distance-based approach [4],[5] was proposed called k-neighboring class sets. In this the number of instances for each pattern is used as the prevalence measure, which does not possess an anti-monotone property by nature. However a non-overlapping instance constraint can be used to get the anti-monotone property for this measure. In contrast an event centric model was developed which does away with the non-overlapping instance constraint. It also defined a new prevalence measure called the participation index. This measure possesses the desirable anti-monotone property. Prevalence measures and conditional probability measures, called interest measures, are defined differently in different models of co-location mining [14]. The reference feature centric and data partitioning models materialize transactions and thus can use traditional support and confidence measures. The event centric approach defined new transaction free measures, e.g., the participation index [5].Co-location pattern mining general approach [8] formalized the co-location problem and showed the similarities and differences between the co-location rules problem and the classic association rules problem as well as the difficulties in using traditional measures (e.g., support, confidence) created by implicit, overlapping and potentially infinite transactions in spatial data sets. [8] proposed the notion of user-specified proximity neighborhoods [13][15] in place of transactions to specify groups of items and defined interest measures that are robust in the face of potentially infinite overlapping proximity neighborhoods[24] .A novel Joinless approach[9] for efficient co-location pattern mining uses an instance-lookup scheme instead of an expensive spatial or instance join operation for identifying co-location instances. A Partial join approach [9]for spatial data which are clustered in neighborhood area.

Mining co-location patterns with rare spatial features [10] proposes a new measure called the maximal participation ratio (maxPR) and shown that a co-location pattern with a relatively high maxPR value corresponds to a co-location pattern containing rare spatial events. Furthermore, it also identifies a weak monotonicity property of the maxPR measure. This property can help to develop an efficient algorithm to mine patterns with high maxPR values. A novel order-clique-based approach [11] is used to mine maximal co-locations. The efficiency of the approach is achieved by two techniques: (1) the spatial neighbor relationships and the size-2 prevalence co-locations are compressed into extended prefix-tree structures, which allows the order-clique-based approach to mine candidate maximal co-locations and co-location instances; and (2) the co-location instances[23,24] do not need to be stored after computing some characteristics of the corresponding co-location, which significantly reduces the execution time and space required for mining maximal co-locations. In this paper distance based approach is used to find the co-location patterns from the spatial data. The participation index is used to prune the data to accept only the interesting patterns. Two algorithms DF-NMColoc and BF-NMColoc[27]were used for finding *N*-most prevalent colocation patterns[25,26].Where *N* is the desired number of colocated event sets with the highest interest measure values per each pattern size.

AMOEBA [16] was a clustering method especially designed for spatial data sets. It uses the Delaunay diagram to incorporate spatial proximity. It does not require any prior knowledge about the data set, nor does it require parameters from the user. It incorporates global first-order effects and local second-order effects. Hence, it is less sensitive to noise, outliers and the type of distribution. AUTOCLUS[17]T was the Effective and efficient method for discovering cluster boundaries in point data sets. The approach automatically extracts boundaries based on Voronoi modeling and Delaunay diagrams [19]. Parameters are not specified by user in AUTIOCLUST. Spatial clustering algorithm TRICLUST [20] based on Delaunay triangulation treats clustering task by analyzing statistical features of data. For each data point, its values of statistical features are extracted from its neighborhood which effectively models the data proximity. TRICLUST is able to effectively handle data set with clusters of complex shapes and non-uniform densities, and with large amount of noises. An adaptive spatial clustering algorithm [21] employs both statistical features of the edges of Delaunay triangulation and a novel spatial proximity definition based upon Delaunay triangulation to detect spatial clusters. FARICS[22] A novel algorithm for finding spatial rules and collocations using Delaunay diagram has been presented. The approach allows eliminating the parameters defining neighborhood of objects, thus avoiding multiple “test and trial” repetitions of the process of mining for various parameter values.

3. Modeling Spatial Proximity with Delaunay Diagram

The raster and vector representations are widely used in modeling geographical databases; they do not consistently capture spatial proximity for point data sets. The Voronoi diagram offers an alternative to overcome the problems of conventional data models. The Voronoi diagram uniquely captures spatial proximity and represents the topology explicitly with the dual graph know as the Delaunay Triangulation. Each Delaunay edge represents neighborhood relation between points. Delaunay diagram is used, because Delaunay Triangulation is not unique when co-circularity occurs. Even in the presence of co-circularlty, Delaunay diagram guarantee a unique topology. Modeling spatial proximity for a discrete point data set $P=\{p_1,\ldots p_n\}$, indicates the neighbors of point p_i and how far are the neighbors relative to the context of the entire data set P.

The main idea of using Delaunay diagram is to avoid the need of defining distance threshold for determining neighborhoods ,do not have to reiterate the process of finding neighborhoods for various user defined parameters and finds neighborhood dynamically. A Delaunay diagram is a sub graph of every

Delaunay triangulation, it is a planar graph whose bounded faces are convex polygons all of whose vertices are co-circular and if no four points of P are co-circular then all bounded faces are triangles and the Delaunay diagram is a triangulation.

Let us denote the Euclidean distance between two points p and q by dist(p,q). In the plane we have:

$$dist(p,q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}$$

Let $P = \{p_1, p_{2,} .....p_n\}$ be a set of *n* distinct points in the plane; these points are called the *sites*.

Definition:1 We define the *Voronoi diagram* of *P* as the subdivision of the plane into *n* cells, one for each site in *P*, with the property that a point *q* lies in the cell corresponding to a site p_i if and only if $dist(q, p_i) < dist(q, p_j)$ for each $p_j \in P$, with $j \neq i$ .We denote the Voronoi diagram of *P* by *Vor*(*P)*. An example of a Voronoi diagram is shown in Fig.1

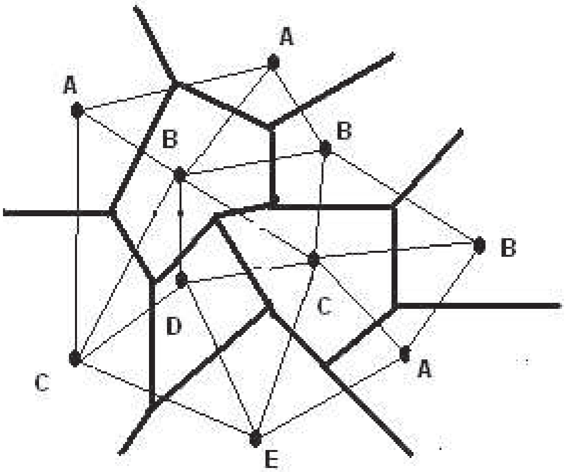


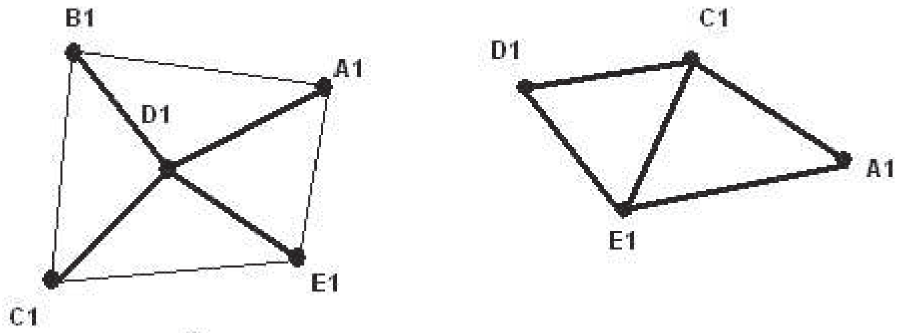
Fig.1. Representation of neighborhood of objects with Delaunay Diagram and the dual of voronoi diagram (darked line)

Definition 2 : We define the Delaunay diagram, D(P), to be the straight-line dual of the Voronoi diagram, Vor(P). By definition, we place a Delaunay node (vertex) at each site p_i and we join two sites p_i and p_j with a straight line segment iff the Voronoi cells corresponding to p_i and p_j share a common boundary segment (i.e., iff the Voronoi edge Vor({p_i, p_j}) exists).

Definition 3: Neighbors of a given object are those objects, for which there is a direct line connection in the Delaunay diagram. A concept of k-neighborhood of two points is defined as the shortest path between those points in the diagram that has the length of k.

The set of all types of objects represented in the Delaunay diagram *D(P)* is denoted by $T = \{t_1, t_2, ..., t_k\}$, where t_i represents one type of objects. Two types of neighborhoods are considered in our approach. There are namely neighborhoods with a central point, and neighborhoods used for computing co-locations.

A neighborhood with a central point is depicted in Fig. 2a. D is a central object here and its neighbors are objects connected to it with a solid bold line. The thin line indicates triangles existing in the Delaunay diagram used to create this neighborhood.



(a) (b)

Fig.2. Neighborhoods: a) neighborhood with a central object, b) neighborhood for co-location computation

The needed neighborhoods are the triangles from the Delaunay diagram which are interested in finding spatial co-locations.

4. Co-location Pattern Mining

4.1. Co-location Concepts

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. Ecologists are interested in finding subsets of spatial features likely to occur in a neighborhood around instance of given subsets of event types. Co-location patterns are discovered by using any one of the model such as reference feature centric model, window centric model and event centric model. The prevalence measure and the conditional probability measure are called interesting measures used to determine useful co-location patterns from the spatial data. The interesting measures are defined differently in different models. Our approach is to find the co-location pattern from the spatial by using Delaunay diagram where the interesting measures are membership index and membership ratio. The objective of co-location pattern mining is to find frequently co-located subsets of spatial features. For example, a co-location {traffic jam, police, car accident} means that a traffic jam, police, and a car accident frequently occur in a nearby region. To capture the concept of "nearby," the concept of Delaunay diagram was introduced. .

Figure.1 shows a spatial data set with a spatial feature set F = {A, B, C, D,E}, which will be used as the running example in this paper.

Definition 4: A co-location group is defined as a set of objects being neighbors to each other.

Sample co-location groups from Fig. 1 include: {C2,D1,B1},{D1,B1},{C2,B1},but {C2,D1,B2} is not a co-location, as the objects D1 and B2are not neighbors.

Definition 5: A co-location instance of a group of types $g = \{t_1, t_2, ..., t_k\}$, $t_i \in T$, is a set $C_g = \{i_1, i_2, ..., i_k\}$, where for every $t_j \in g$, there is one (and only one) instance i_s of the type t_j in C_g. Sample co-location instances for g = {C, D, E} from Fig.1 are co-locations :{ C1, D1, E1}, {C2, D1,E1}.

Having Delaunay diagram(P) and any group of types g, we denote by G_g the set of all co-location instances of group of types g occurring in the diagram.Fig.1 shows, there are no more instances of the group of types{C,D,D} than the two listed above, they constitute the set of all co-location instances of this group.

The set of all occurrences of objects of type t_i in the diagram, where $t_i \in T$ is denoted by G_{t_i} The set of all occurrences of objects of type C from fig.1 is

$G_C = \{\{C_1\}, \{C_2\}\}$

Definition 6: A membership ratio $MR_g(g, t_i)$ for the group of types $g = \{t_1, t_2, ..., t_k\}, g \subset T, t_i \in g$, can be described by the following formula

$$MR_g(g, t_i) = \frac{\Pr esence(t_i, G_{g)}}{\left|G_{t_i}\right|}$$

Where $\Pr esence(t_i, G_{g)}$ denotes number of unique instances of the objects of the type t_i that present in the set of all co-location instances of g, $\left|G_{t_i}\right|$ denotes the cardinality of the set of all occurrences of the objects of type t_i. In fig.1 the co-location instances of the group types g_1 ={A,C} are co-locations {A1,C1} and {A3,C2} .Only two objects out of three of type A participate in the co-location instances of g_1, thus MR_{g1}({A, C, A} =2/3.

All objects of type C participate in the co-location instances of g_1, MR_{g1}({A,C},C)=2/2. Let us consider another group of types g2 = {A, E}. The only co-location instance for g_2 is {A3, E1}. The membership ratios for g_2

MR_{g2}({A, E}, A)=1/3, MR_{g2}({A,E},E)=1/1=1.

Definition 7: A membership index MI_g for any group of types $g = \{t_1, t_2, ..., t_k\}$, is defined as

$\min_{i=1}^{k}\{MR_g(g, t_i)\}$ For the case as in fig.1 the membership ratios are , MR_g({A,C},A)=2/3,

MR_g({A,C},C)=2/2=1. The membership index for g= {A, C} is equal to min (2/3,1)=2/3.

Definition 8: A co-location rule is the rule of the form g1→g2 (p%, c%)where g1 and g2 are groups of types and do not intersect, p% is the prevalence of the rule, c% denotes confidence of the rule. Membership index is used as the prevalence measure. Confidence of a co-locartion rule Conf(g1→g2) is calculated using the following formula

$$Conf(g_1 \rightarrow g_2) = \frac{C_g(g_1 \cup g_2)}{\left|G_{g1}\right|}$$

Where $C_g(g_1 \cup g_2)$ denotes the number of all unique co-location instances of a group of types $g_1 \cup g_2$. $\left|G_{g1}\right|$ denotes the cardinality of the set of all co-location instances of a group of types g_1. The confidence of a co-location rule approximates the probability of finding co-location instances of objects belonging to group of types $g_2$in the neighbourhood of co-location instances of objects belonging to the group of types g_1. Having defined the prevalence measures of MR and MI for Co-locations, we are describing the algorithm of finding the co-location patterns in the next section.

4.2. Co-location mining Architecture and Algorithm

Co-location architecture input mainly consists of a spatial data which is processed to derive the co-ordinates item instances. The co-location algorithm is used to generate item sets from those co-ordinates based on the Delaunay diagram. When the algorithm is applied the co-ordinates are mapped in a Delaunay diagram..

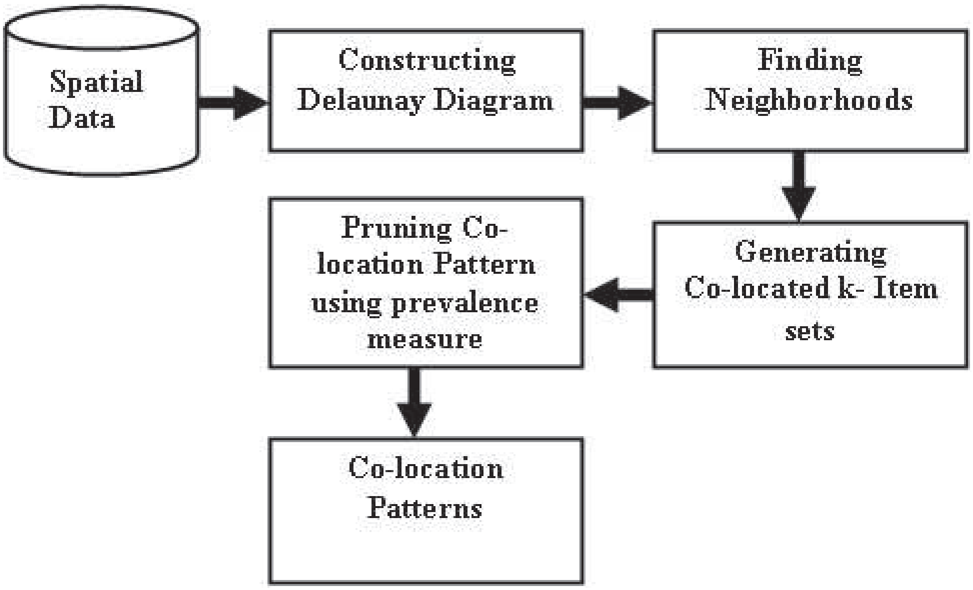


Fig.3. General Architecture of Co-location Pattern Analysis

The co-located item sets are generated based on the neighborhood between the object instances. The co-located item sets are pruned if patterns don't have minimum membership index. The non-pruned item sets are used to calculate the 3-item co-located sets. The interesting patterns are identified after pruning depending on the membership index. One instance of an item is compared with all the instances of other item and checked for neighbourhood and membership index is found out and according to membership index the co-location pattern is predicted.

Co-location Algorithm

1 .Read the Spatial dataset
2. Construct the Delaunay diagram
3. Find the neighbourhood between the object instances
4. Calculate Membership ratio of co-location item sets
5. Calculate Membership Index of Co-location
6. Initialize pruning index
7. Compare Membership index value and pruning index and consider only the item set that are above the pruning index.
8. The items that are pruned out in n-item set calculation are ruled out in n+1-item set Calculation
9. Generate the co-location pattern
10. Calculate the confidence of the co-location item sets
11. Generate the co-location rules based on the confidence measure

5. Computing spatial Co-location rules

In this section we will illustrate the algorithm of calculating co-locations, based on the example from Fig.1 For each group created during the clustering phase we consider the consecutive triangles belonging to that group that build up the Delaunay diagram. Each triangle is a co-location instance. The diagram is composed of triangles and polygons, obtained by removing internal edges of incident triangles in a triangulation process. The co-location groups from Fig.1 needed for the calculations are presented in Fig. 6

A1 A2 B1
B1 A2 B2
B1 D1 C2
B1 C2 B2
B2 C2 B3
C1 E1 D1
D1 E1 C2
C2 E1 A1
C2 A3 B3
A1 C1

Fig.6.Co-location groups for calculating spatial co-location patterns

Table 1 Candidate one element itemsets

A	B	C	D	E
A1	B1	C1	D1	E1
A2	B2	C2		
A3	B3			

The first step in the process of finding spatial co-locations among objects of various types is the creation of one-element itemsets. According to the adopted definition, all of them are frequent and there is no need to either calculate the prevalence or filter those sets out, based on the value of prevalence. One-element itemsets from Fig. 1 and their instances are shown in Table 1. The next step is to create two- and three-element candidate itemsets representing neighboring objects. For this purpose we use the structure of the Delaunay diagram. The fig.7a shows two and three element co-location itemsets which are generated by applying joins between the itemsets. Fig.7b shows the co-location rule which satisfies high confidence value.

{B,A}
{B,C}
{B,D}
{B,E}
{A,C}
{A,E}
{A,D}
{C,E}
{C,D}
{E,D}
{B,A,C}
{B,A,D}
{B,A,E}
{B,C,D}
{B,C,E}
{B,D,E}
{A,C,E}
{A,C,D}
{A,E,D}
{C,E,D}

{B} → {D}
{D} → {B}
{A} → {C}
{C} → {A}
{A} → {E}
{E} → {A}
{C} → {E}
{D} → {E}
{A,C} → {B}
{B,C} → {A}
{C,D} → {B}
{B,D} → {C}
{C,E} → {A}
{A,E} → {C}
{A,C} → {E}
{E,D} → {C}
{C,E} → {D}

Fig.7. (a) Two and three element co-location itemesets from the co-location groups (b) Spatial co-location rules

6. Performance Analysis

The proposed approach is a data-driven method in the sense that it does not require control parameters from the user. The rules discovered with this method apply to the immediate neighborhood of respective objects. We have compared the efficiency of our approach with two most popular techniques for the neighborhood generation such as distance and window based approaches. The approach proposed by Shekhar and Huang (2001), treating as neighbors those objects that are located within the given distance being the radius of a circle with the considered main object as the centre .The window method, discussed in Estivill-Castro and Lee (2000a), where the regarded data space is divided into square shaped windows with a user-defined length to find the neighborhoods. The following figure shows the computing time of neighbourhoods generation for Delaunay and window approaches.

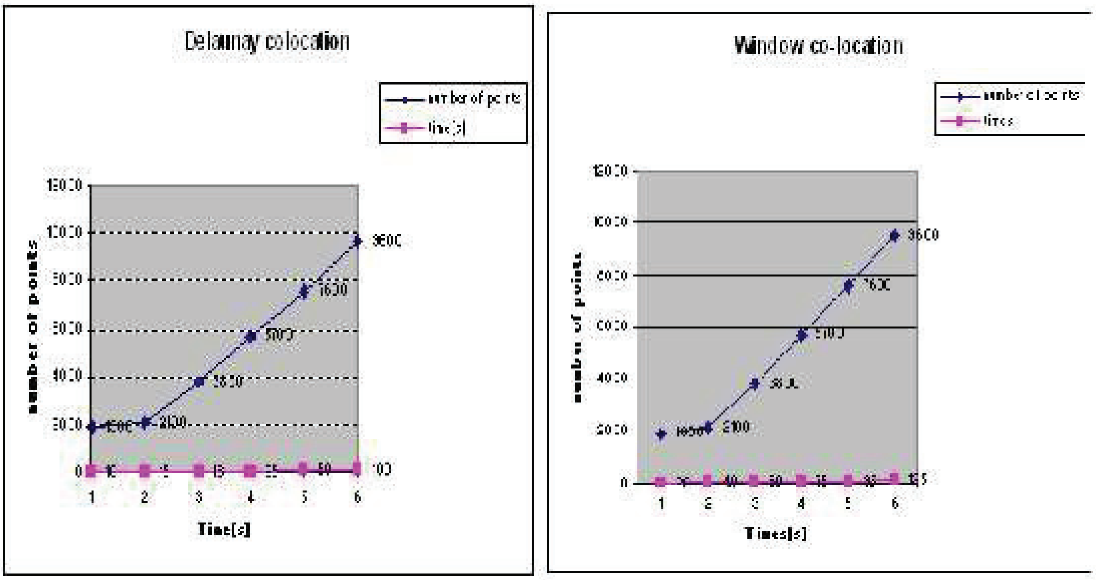


Fig.8. Computing time of neighborhoods generation for (a) Delaunay based co-location (b) Window based co-location

Comparing our approach with the *window* and *dis* methods, it finds properly all the neighborhoods independently of the density for both association rules and co-locations, whereas for the other two algorithms it is necessary to choose the window/distance parameter, which is often problematic. Especially, if the variations are higher, the problem may result in the need of multiple runs of *dis* and *windows* in order to find out proper distance threshold parameters, which may influence efficiency of the data mining process. Clearly, it will also result in the quality of the retrieved rules. The number of co-locations obtained in the Delaunay method is considerably limited and quality patterns than other methods as shown in fig.9

The execution of the proposed method is built up of two phases that are significant from the computational viewpoint: (1) the construction of the Delaunay diagram and the finding neighborhoods and (2) generating spatial co-locations. For the first phase it requires the time complexity of $O(N \log N)$,where N represents the number of points. The complexity $O(N \log N)$ actually reflects the time that is necessary to construct the Delaunay diagram, as the remaining part of phase (1) is performed in the linear time. For performing phase (2) we use the co-location algorithm.

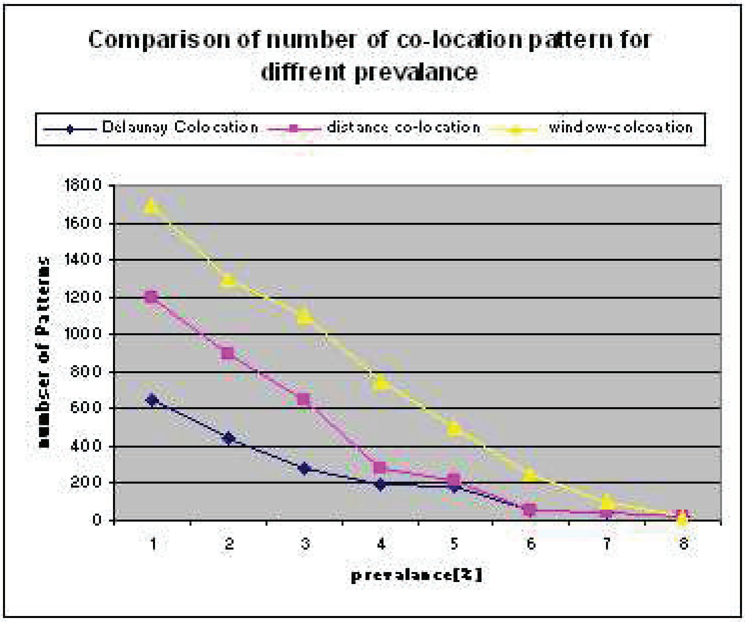


Fig.9.Comparisons of number of co-location patterns for different prevalence values in various approaches

Our Co-location algorithm takes minimum time to generate more number of instances in co-location pattern analysis. Our approach does not need the constraint of "any point object must belong to only one instance" since we do not use the number of instances for a pattern as its prevalence measure. We used Membership index as the prevalence measure, which possesses a desirable antimonotone property for effectively reducing the search space.

7. Conclusions and Future Enhancements

In this paper we presented a co-location mining algorithm for discovering spatial co-location patterns. In our approach we used the Delaunay diagram to find neighborhood of objects Delaunay diagram is a structure representing the neighborhood of objects in a succinct and unambiguous manner. Spatial data mining without unequivocal neighborhood definition returns different results depending on the assumed window size or the radius in which other objects are deemed neighbors and forces multiple execution of the mining process for different values of the appropriate parameters. An interest measure, a Membership index, is used for spatial co-location patterns as it possesses an anti-monotone property. The Co-location algorithm to mine co-location patterns from the spatial data was presented and analyzed. In future, the co-location mining problem should be investigated to account categorical and continuous data and also extended for spatial data types, such as line segments and polygons.

REFERENCES

[1].Huang. Y, Shekhar. S, and Xiong. H, "Discovering Colocation Patterns from Spatial Data Sets: A General Approach" IEEE Transactions on Knowledge and Data Engineering, 16, (12) pp. 1472-1485, 2004.

[2]. R.Agarwal and R.Srikant, "Fast Algorithms for Mining Association Rules," Proc.20th Int'l Conf. Very Large Data Bases, 1994.

[3]. K.Koperski and J.Han, "Discovery of Spatial Association Rules in Geographic Information Databases," Proceeding .Fourth Int'l Symp. Spatial Databases, 1995.

[4]. Morimoto .Y, "Mining Frequent Neighboring Class Sets in Spatial Databases". In Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2001

[5].Shekhar.S and Huang. Y, "Co-location Rules Mining: A Summary of Results". In Proc. of the 7th Int'l Symp. on Spatial and Temporal Databases,2001

[6].V. Estivill-Castro and I. Lee, "Data Mining Techniques for Autonomous Exploration of Large Volumes of Geo-Referenced Crime Data," Proc. Sixth Int'l Conf. Geo Computation, 2001.

[7].V. Estivill-Castro and A. Murray, "Discovering Associations in Spatial Data—An Efficient Medoid Based Approach," Proc. Second Pacific-Asia Conf. Knowledge Discovery and Data Mining, 1998.

[8].Yan Huang, Shashi Shekhar and Hui Xiong, "Discovering Co-location Patterns from Spatial Data Sets: A General Approach", IEEE Transactions On Knowledge And Data Engineering, Vol. 16, No. 12, December 2004

[9].Jin Soung Yoo and Shashi Shekhar, "A Joinless Approach for Mining Spatial Co-location patterns", IEEE Transactions On Knowledge And Data Engineering, Vol. 18, No. 10, October 2006

[10].Yan Huang · Jian Pei · Hui Xiong, "Mining Co-Location Patterns with Rare Events from Spatial Data Sets", Geoinformatica (2006) 10: 239–260

[11].Lizhen Wang, Lihua Zhou , Joan Lu and Jim Yip. "An order-clique-based approach for mining maximal co- locations", Information Sciences 179 (2009) 3370–3382

[12].Zhan-Quan Wang, Hai-Bo Chen, Hui-Qun Yu, "Spatial Co-Location Rule Mining Research In Continuous Data" Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 August 2006

[13].Feng Qian, Qinming He, and Jiangfeng He, "Mining Spatial Co-location Patterns with Dynamic Neighborhood Constraint", ECMLPKDD 2009, , LNAI 5782, pp. 238–253, 2009

[14].Marko Salmenkivi, "Efficient Mining of Correlation Patterns in Spatial Point Data" PKDD 2006, LNAI 4213, pp. 359–370, 2006.

[15].Mete Celik, Shashi Shekhar, James P. Rogers, and James A. Shine, "Mixed-Drove Spatiotemporal Co-Occurrence Pattern Mining" IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 10, October 2008.

[16] Estivill-Castro, V., Lee, I., 2000. AMOEBA: Hierarchical clustering based on spatial proximity using Delaunay diagram. In Proc. 9th Internat Symposium on Spatial Data Handling, pp. 7a.26–7a.41.

[17] Estivill-Castro, V., Lee, I., 2000. AUTOCLUST: Automatic clustering via boundary extraction for mining massive point-data sets. In: Proc. 5th Internat. Conf. on Geo computation.

[18] Okabe, Atsuyuki, Boots, Barry, Sugihara, Kokichi, Chiu, Sung Nok, 2000. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, second ed. John Wiley & Sons.

[19] Estivill-Castro, V., & Lee, I. (2002b). Argument free clustering for large spatial point data sets. Computers, Environment and Urban Systems, 26, 315–334.

[20] Dongquan Liu, Gleb V. Nosovskiy , Olga Sourina, Effective clustering and boundary detection algorithm based on Delaunay triangulation, Pattern Recognition Letters 29 (2008) 1261–1273

[21] Robert Bembenik · Henryk Rybinski, FARICS: a method of mining spatial association rules and collocations using clustering and Delaunay diagrams, J Intell Inf Syst (2009) 33:41–64

[22] Min Deng , Qiliang Liu , Tao Cheng , Yan Shi, An adaptive spatial clustering algorithm based on delaunay triangulation Computers, Environment and Urban Systems 35 (2011) 320–332

[23] Shekhar S, Chawla S (2003) Spatial databases: a tour. Prentice Hall, Upper Saddle River

[24] Zou S, Zhao Y, Guan J, Huang J (2005) A neighborhood-based clustering algorithm. In: Proceedings of the Pacific-Asia conference on knowledge discovery and data mining

[25] Wan Y, Zhou J (2008) Knfcom-t: a k-nearest features-based co-location pattern mining algorithm for large spatial data sets by using t-trees. Int J Bus Intell Data Min 3(4):375–389

[26] Yoo JS, Bow M (2009) Finding N-most prevalent colocated event sets. In: Proceedings of the international conference on data warehousing and knowledge discovery (DaWak)

[27] Jin Soung Yoo · Mark Bow, Mining spatial colocation patterns: A different Framework, Data Min Knowl Disc,Springer,2011