

# Mining Of Spatial Co-location Pattern from Spatial Datasets

G.Kiran Kumar  
Associate Professor  
Department of CSE  
MLR IT, Hyderabad  
A.P.

P.Premchand  
Professor  
Department of CSE  
O.U.College of Engineering  
Hyderabad, A.P.

T.Venu Gopal  
Associate Professor  
Department of CSE  
JNTUHCE Karimnagar  
A.P. India

## ABSTRACT

Spatial data mining, or knowledge discovery in spatial database, refers to the extraction of implicit knowledge, spatial relations, or other patterns not explicitly stored in spatial databases. Spatial data mining is the process of discovering interesting characteristics and patterns that may implicitly exist in spatial database. A huge amount of spatial data and newly emerging concept of Spatial Data Mining which includes the spatial distance made it an arduous task. Knowledge discovery in spatial databases is the extraction of implicit knowledge, spatial relations and discovery of interesting characteristics and patterns that are not explicitly represented in the databases. Co-location pattern discovery is the process of finding the subsets of features that are frequently located together in the same region. Spatial co-location patterns associate the co-existence of non-spatial features in a spatial neighborhood. The Previous methods of mining co-location patterns, converts neighborhoods of feature instances to item sets and applies mining techniques for transactional data to discover the patterns, combines the discovery of spatial neighborhoods with the mining process. It is an extension of a spatial join algorithm that operates on multiple inputs and counts long pattern instances. Previous works on discovering co-location patterns is based on participation index and participation ratio. In this paper we address the problem of mining co-location patterns with a novel method called Mediod participation index Our technique is an extension of maximal participation ratio and deploys the idea of K-mediods from clustering algorithms.. As demonstrated by experimentation, our method yields significant performance improvements compared to previous approaches.

## Keywords

Spatial Data Mining, Association Rules, Co-location rules, Participation Index, Apriori algorithm, Participation ratio.

## 1. INTRODUCTION

### 1.1 Data Mining

Extraction of knowledge from a data set in a human understandable form is the goal of data mining. The primary source to perform data mining [1,2] is the place where the data resides i.e database. It is a huge collection of facts related to some domain in the outside world in an organized manner. One of the primary aims or objectives of data mining is to analyze regularities that enable us for a detailed study of the domain explained by the database. Data mining is also known as Knowledge Discovery. It leaves ample scope to discover the process and patterns from a large database. It identifies the valid, potentially useful, understandable patterns & relationships in data. It discovers the new patterns from the large database[3,4]. It is evolved as a new interdisciplinary field of computer science.

### 1.2 Spatial Database and its Applications

Earlier in the research of data mining, only non-spatial data was concentrated. This resulted in the great achievement in finance, GIS, business, insurance, etc. Later the focus of research has been shifted to spatial data from non-spatial data. The important difference between non-spatial data and spatial data is that spatial data includes the constraints like scales, coordinates and have pictorial information.

Owing to some constraints, the covert knowledge embedded in the spatial database cannot be retrieved using conventional database management systems. It needs advanced management systems to tackle with. Use of spatial data enhances the productivity of data mining. This was recognized by scientists, environmental researchers, administrators and business professionals. Spatial data cannot be treated in the same way as it was with the other forms of data. A spatial database management system is used for this purpose. The primary technology for SDBMS is Geographic Information System. The difference between the GIS & SDBMS is that the GIS has a rich set of operations over few objects and layers, whereas SDBMS provides simpler operations on a set of objects and set of layers. Internet provides a spatially sensitive search engine. Queries such as "Find all Udipi Hotels in Hyderabad" can be answered by such search engines. Spatial technologies can be used to determine the location of the mobile phones which helps in providing location-based services. The applications of spatial data are those which involve decisions. For example, In Geo-Marketing, a store can establish its trade area, i.e the spatial boundary of its customers. Then the profile of those customers is analyzed on the basis of both their properties and the properties related to the area where they live.

In Geography, the main basis of data analysis is conventional statistics and multi-dimensional analysis which doesn't consider spatial data [5]. However the geographic data implies that the observations that are near-by in the space tend to share similar (or correlated) attribute values. This comprises the fundamental of distinct scientific area that focuses on inter-dependence of near-by observations, unlike conventional statistics. This is known as "spatial statistics". The extension of spatial statistics is Multi-dimensional analytical methods[6]. The study proves that spatial statistics is an integral part of spatial data mining as it provides data driven analysis. Some of these methods are now implemented in operational GIS or analysis tool. This paper is organized as follows. Firstly, spatial data mining is discussed in section 2, section 3 discusses co-location mining in which a case study is presented in-addition to the discussion related to existing and proposed system. Experimental setup is discussed in Section 4 in which the algorithm present in section 3 is programmed and results are analyzed, conclusion and scope for future work is given in section 5.

## 2. SPATIAL DATA MINING

The extra features that distinguish spatial data from other forms of data are spatial co-ordinates, topological distance, and direction information. By inclusion of many features, query language has become too complicated. The process to discover knowledge is not so easy as earlier now. The reason is “spatial statistics”. In contrast to the mining in relational databases, spatial data mining algorithms need to consider the objects that are near-by in order to extract useful knowledge as there is influence of one object on the neighboring object. Different approaches have been developed for knowledge discovery from spatial data such as Spatial Classification, Spatial association rule mining, Spatial Clustering [7].

### 2.1. Spatial classification

The basic objective of classification is to assign an object to a class from a given set of classes based on the attribute values of the object. In spatial classification the attribute values of neighboring objects may also be relevant for the membership of objects and therefore have to be considered as well. Grouping of data items according to their attribute values into categories is known as Classification [8]. This is identified as supervised classification; unsupervised classification is called as clustering. “Supervised” classification needs a training dataset to train or configure the classification model, a validation dataset to validate (or optimize) the configuration, and a test dataset to test or evaluate the performance of the trained model [9,10]. Classification methods include, for example, decision trees, linear discriminant function, support vector machines (SVM), artificial neural networks, maximum likelihood estimation, nearest neighbor methods and case-based reasoning .

### 2.2. Spatial association

The original aim of association rule mining is to identify the regularities between the items in the large transactional databases. Generally the strength is indicated by confidence and support denotes the frequency of the association rule. Rules that have reasonably large support needs more concentration [11]. Spatial properties and predicates are considered while formulating spatial association rules in spatial database which is similar to mining of the association rules in transactional or relational databases. Spatial co-location pattern mining differs with association rule mining [14] only in the technical manner, where as they are conceptually similar. A co-location pattern can be obtained by the input of a dataset of spatial features and their locations. These patterns represent subsets of features that are located together more frequently. For example, certain species of bird tend to habitat with certain type of trees. A location is not a transaction and two features rarely exist at exactly the same location. To check the features that co-locate in the same neighborhood, a user-specified neighborhood is needed. Many measures and algorithms have been proposed to mine spatial co-location patterns.

### 2.3. Spatial clustering

Clustering is interpreted as the task of collecting the objects of a database into meaningful detectable subclasses (that is clusters) so that the members of a cluster are as similar as possible whereas the members of different clusters differ as much as possible from each other. In Data analysis Cluster analysis is very frequently used, which organizes a set of data items into groups (or clusters) so that items in the same group are similar to each other and different from those in other groups [6,15]. Clustering methods can be broadly classified into Five groups they are Partitioning algorithms, Density based clustering, Hierarchical Algorithms, Grid-Based Methods and Model-Based Clustering Methods. Example algorithms of the above classification are K-Means, K-medoids, DBSCAN [16] GDBSCAN [17] Chameleon. To consider spatial information in

clustering, three types of clustering analysis are existing, they are spatial clustering, regionalization, and point pattern analysis.

This paper primarily focuses on spatial data mining, with a specific reference to spatial co-location pattern mining which is conceptually similar to association mining but technically very different.

## 3. CO-LOCATION PATTERNS

Spatial co-location pattern mining is similar to association mining. A spatial association rule is a rule of the form “ $A \rightarrow B$ ” where A and B are sets of predicates and some of which are spatial ones. In a large database many association relationships may exist but some may occur rarely or may not hold in most cases [18,19].

### 3.1 Application of Association Rules on Spatial Data

Spatial and non-spatial relationships at a concept level are discovered in spatial data mining[20]. Spatial objects are defined as merged spatial regions or clustered spatial points. Spatial or non-spatial relationships which contain spatial predicates such as adjacent\_to, occur\_together, near\_by, inside, close\_to etc cannot be identified by the above methods. Object/predicate relationships are represented by Spatial Association rules containing spatial predicates, For example, the following rules are spatial association rules “is\_a(k, gas station)  $\rightarrow$  close\_to(k, highway)”.

To facilitate the specification of the primitives for the spatial data mining, an SQL-like spatial data mining query interface, which is designed based on a spatial SQL proposed in [21], has been specified for an experimental spatial data mining system prototype, GeoMiner.

### 3.2 Modeling of Co-location rules

Co-location pattern discovery is the process of finding the subsets of features that are frequently located together in the same region. Spatial co-location patterns associate the co-existence of non-spatial features in a spatial neighborhood.

Owing to the following reasons Co-location pattern discovery is an ardent task [22]. With traditional association rule mining algorithms it is difficult to find co-location patterns since there is no concept of traditional “transaction” in most of spatial datasets. The instances of a spatial feature are getting distributed in a spatial framework, these instances share complex spatial neighborhood relationships with other spatial instances. Time complexity to generate the table instances of co-location pattern is very high. A model based approach for finding co-location rules is discussed [23]. The Previous methods of mining co-location patterns, converts neighborhoods of feature instances to item sets and applies mining techniques for transactional data to discover the patterns [15]. Some earlier methods combined mining process with the discovery of spatial neighborhoods and they are extensions of a spatial join algorithm that operates on multiple inputs and counts long pattern instances[14].

A Co-location rule is of the form:  $L_1 \rightarrow L_2(p, cp)$  where  $L_1$  and  $L_2$  are colocations, p and cp represents prevalence measure and conditional probability[24]. The definition of neighbor relation R is an input and is based on the semantics of application domains. It may be defined using graph theory such as connected, adjacent or metric relationship using distance measures such as earth mover distance, Euclidean distance, city block distance etc or a combination. The table instance of a co-location is the collection of all its row instances. Previous works on discovering co-location patterns is based on participation index and maximal participation ratio.

**Participation Ratio** : The Participation ratio  $P_f$  for a feature type  $f_i$  of a co-location L is the fraction of instances of  $f_i$  which participate in any row instance of co-location L. It is defined as  $pr(L, f_i) = \pi_f(\text{table instance}(L)) / (\text{instances}(f_i))$

$L = \{f_1; f_2; \dots; f_k\}$

**Participation index (PI)**  $pi(L) = \min_{i=1}^k Pr(L, f_i)$

Participation ratio and participation index are monotonically decreasing with respect to co-location size.

### 3.3 Our Proposition

In this section, we review a framework of mining co-location patterns, since our proposed solution in this paper is based on this model. For any co-location mining algorithm major challenge is neighborhood enumeration, identifying row instances of co-location. Our approach follows apriori algorithm, it is an extension of the participation index [14] and maximal participation ratio[22]. The following algorithm generates all the co-location rules with high prevalence's and conditional probabilities.

#### Algorithm

Step 1: Initialize prevalent co location set of size 1 along with table instance

Step 2: Generates co location rules of size 2

Step 3: For co-location of size (2, 3, 4 ...K-1) do

Begin

Step 4: Generate candidate prevalent co-location

Step 5: Generate table instance

Step 6: Prune co-locations using prevalence threshold

Step 7: Generate co-locations rules

Step 8: Stop

Explanation Of Step 4: Candidate prevalent co-locations of size K+1 are generated using combinatorial approach from prevalent co-locations of size K along with their table instances.

#### Case Study

A Spatial Data Set with spatial feature set  $F = \{A, B, C, D, E\}$  is shown in figure1. Objects with various shapes represent different spatial features. Each instance is uniquely identified by its instance-id. We have 20 instances in the database.

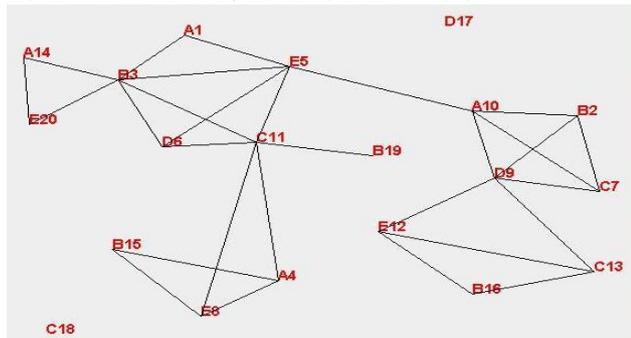


Figure 1: An Example data set

Object A has 4 instances with the instance ids {1,4,10,14}, Object B has 5 instances with the instance ids {2,3,15,16,19}, Object C has 4 instances with the instance ids {7,11,13,18} Object D has 3 instances with the instance ids {6,9,17}, Object E has 4 instances with the instance ids {5,8,12,20}. In figure 1, the neighborhood relation  $R$  is defined based on the Euclidean distance. Two instances are neighbors if their Euclidean distance is less than a user specified threshold. Neighboring instances are connected by edges. For instance {1,3,6,5}, {16,13,9,12} and {10,2,7,9, 12} are all neighbor-sets because each set forms a clique. Here, we use the instance-id to refer to an object in figure 1. Additional neighbor-sets include {14,20}, {4,15,8} and {10,2,7,9}.  $\{A,B,C,D,E\}$  is a co-location pattern. The neighborhood-set {10,2,7,9,12} is a row instance of pattern  $\{A,B,C,D,E\}$  but the neighborhood-set {10,2,7,9,12,13} is not a row instance of co-location  $\{A,B,C,D,E\}$  because it has a proper subset {10,2,7,9,12} which contains all the features in  $\{A,B,C,D,E\}$ . Finally  $rowset(\{A,B,C,D,E\})$  include only {10,2,7,9,12} but not {10,2,7,9,12,13}.

#### 3.3.1 Conditional Probability

For a co-location rule  $R: A \rightarrow B$ , the *conditional probability*  $cp(R)$  of  $R$  is defined as

$$\frac{|\{L \in rowset(A) \exists L' \text{ s.t. } (L \subseteq L') \wedge (L' \in rowset(A \cup B))\}|}{|rowset(A)|}$$

$rowset(\{A,B,C,D,E\})$  contains following eight patterns {{12,9,2, 10,13,} {12,9,16,10,13}{5,6,3,10,11}{5,6,3,14,11}{5,6,3,1,11} {12,9,2,10,7}{12,9,16,10,7}{5,6,3,10,7}} as shown in figure 2. and  $rowset(\{A,B,C\})$  contains following four patterns {{15,4,11,}{3,14,11,} {3,1,11} {2,10,7,}. Since  $|rowset(\{A,B,C\})| = 4$  only 3 rows of  $\{A,B,C\}$  satisfy the subset condition,  $\{A,B,C,D,E\}$  i.e., rows {10,2,7}, {1,3,11}, {14,3,11}. The conditional probability  $cp(\{A,B,C\} \{D,E\}) = 3/4 = 75\%$ .

Let us consider that the occurrence of objects A,B and C represent the occurrence of a traffic jam and the occurrence of the objects D and E represent the occurrence of the Violation of road rules. When all the objects occur simultaneously, it represents the occurrence of a road accident.

$cp(\{A,B,C\} \rightarrow \{D,E\})$  represent the conditional probability of the occurrence of the road accident whenever the traffic is jammed.

Traffic is jammed in four cases i.e {4,15,11}, {10,2,7}, {1,3,11} and {14,3,11}. Among these four, Accident occurs only when the traffic is jammed along with the violation of the road rules i.e – {10,2,7}, {1,3,11} and {14,3,11}.

Hence we conclude that, For every 4 traffic jams, there is a probability of 3 times occurring of an accident i.e. 75%.

#### 3.3.2 Participation ratio

Given a spatial database  $S$ , to measure how a spatial feature  $f$  is co-located with other features in co-location pattern  $C$ , a *participation ratio*  $pr(C, f)$  can be defined as

$$pr(C, f) = \frac{|\{r | (r \in S) \wedge (r.f = f) \wedge (r \text{ is in a row instance of } C)\}|}{|\{r | (r \in S) \wedge (r.f = f)\}|}$$

We first identify the row-sets of  $\{A,B,C,D,E\}$  that include 8 patterns. Among all the four instances of A i.e. (1,4,10,14) three of them, namely {1,10,14} has B, C, D and E in a neighbor-set. So the participation ratio  $pr(\{A,B,C,D,E\}, A) = 3/4$ . Similarly we can have  $pr(\{A,B,C,D,E\}, B) = 3/5$ ,  $pr(\{A,B,C,D,E\}, C) = 3/4$ ,  $pr(\{A,B,C,D,E\}, D) = 2/3$  and  $pr(\{A,B,C,D,E\}, E) = 1/2$ .

Participation ratio of each object in the given event represents the ratio of the distinct instances of the objects participating in the event to the total number of instances of the respective object.

$pr(\{A,B,C,D,E\}, A)$  represent the participation ratio of A in the event  $\{A,B,C,D,E\}$ , We know that A has four instances {1,4,10,14}. Among these four, only three instances {1,10,14} are participating in the event  $\{A,B,C,D,E\}$ , whereas instance {4} is not participating in the event. Hence participation ratio of A is  $pr(\{A,B,C,D,E\}, A) = 3/4$ .

#### 3.3.3 Participation Index

Participation index (PI) of the event is the probability of occurring of an event based on the participation ratios of the objects in the event. The event can occur only when all the objects in the event occur simultaneously. For example, an accident can occur when both the traffic jam and violation of road rules occur simultaneously. It depends on the participation ratios of the objects. When the least participated object occurs, surely all the remaining objects participate in the event. Hence the participation index is the minimal of the participation ratios



In the event {A,B,C,D,E}, by using the factor maxPI, we got the features A and C that has to be taken care. But these features are contextually most useful. Hence these cannot be avoided. Now by applying Mediod Participation index, we got the feature D to be taken care which is comparatively less contextually important than A and C.

#### 4. EXPERIMENTAL SETUP

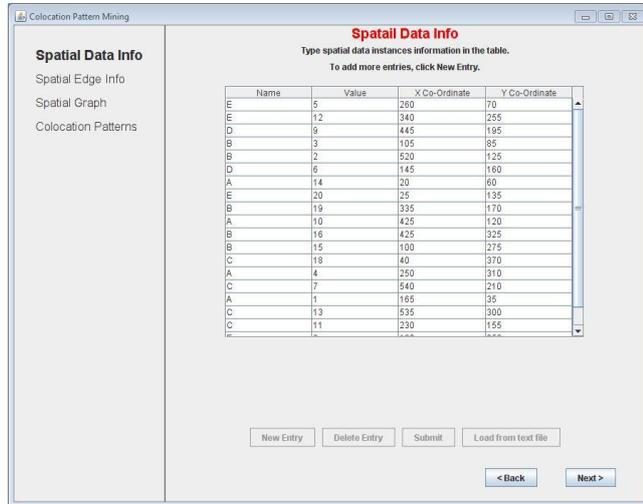


Figure 4: snapshot of spatial data

In this section we present experiments to evaluate the performance of three algorithms - a) Participation Index b) Maximum Participation Index c) Proposed co-location mining algorithm using Mediod Participation Index.

Figure 5: Snapshot of results calculating PI, MaxPI, Med PI

Co-loc	Rowset	PR	PI	Max. PI	Med. PR
{E}	{{5},{12},{20},{8},}	{1.0}	1	1	1
{D}	{{9},{6},{17},}	{1.0}	1	1	1
{B}	{{3},{2},{19},{16},{15},}	{1.0}	1	1	1
{A}	{{14},{10},{4},{1},}	{1.0}	1	1	1
{C}	{{18},{7},{13},{11},}	{1.0}	1	1	1
{ED}	{{5,6},{12,9},}	{0.5,0.67}	0.5	0.67	0.5
{EB}	{{8,15},{20,3},{12,16},{5,3},}	{1.0,0.6}	0.6	1	0.6
{EA}	{{5,10},{8,4},{20,14},{5,1},}	{0.75,1.0}	0.75	1	0.75
{EC}	{{8,11},{12,13},{5,11},}	{0.75,0.5}	0.5	0.75	0.5
{DB}	{{6,3},{9,2},}	{0.67,0.4}	0.4	0.67	0.4
{DA}	{{9,10},}	{0.34,0.25}	0.25	0.34	0.25
{DC}	{{6,11},{9,13},{9,7},}	{0.67,0.75}	0.67	0.75	0.67
{BA}	{{2,10},{15,4},{3,14},{3,1},}	{0.6,1.0}	0.6	1	0.6
{BC}	{{19,11},{16,13},{3,11},{2,7},}	{0.8,0.75}	0.75	0.8	0.75
{AC}	{{4,11},{10,7},}	{0.5,0.5}	0.5	0.5	0.5
{EDB}	{{12,9,16},{5,6,3},{12,9,2},}	{0.5,0.67,0.6}	0.5	0.67	0.6
{EDA}	{{5,6,10},{5,6,1},{12,9,10},}	{0.5,0.67,0.5}	0.5	0.67	0.5
{EDC}	{{12,9,13},{5,6,11},{12,9,7},}	{0.5,0.67,0.75}	0.5	0.75	0.67
{EBA}	{{5,3,10},{8,15,4},{20,3,14},{5,3,14},{5,3,1},{20,3,1},}	{0.75,0.4,1.0}	0.4	1	0.75
{EBC}	{{8,15,11},{12,16,13},{5,3,11},{20,3,11},}	{1.0,0.6,0.5}	0.5	1	0.6
{EAC}	{{8,4,11},{5,10,11},{5,1,11},{5,10,7},}	{0.5,0.75,0.5}	0.5	0.75	0.5
{DBA}	{{9,2,10},{6,3,14},{6,3,1},}	{0.67,0.4,0.75}	0.4	0.75	0.67
{DBC}	{{6,3,11},{9,2,13},{9,2,7},}	{0.67,0.4,0.75}	0.4	0.75	0.67
{DAC}	{{9,10,13},{9,10,7},}	{0.34,0.25,0.5}	0.25	0.5	0.34
{BAC}	{{15,4,11},{3,14,11},{3,1,11},{2,10,7},}	{0.6,1.0,0.5}	0.5	1	0.6
{EDBA}	{{12,9,2,10},{5,6,3,10},{5,6,3,14},{5,6,3,1},{12,9,16,10},}	{0.5,0.67,0.6,0.75}	0.5	0.75	0.6
{EDBC}	{{12,9,16,13},{12,9,2,13},{5,6,3,11},{12,9,16,7},{12,9,2,7},}	{0.5,0.67,0.6,0.75}	0.5	0.75	0.6
{EDAC}	{{12,9,10,13},{5,6,10,11},{5,6,1,11},{12,9,10,7},{5,6,10,7},}	{0.5,0.67,0.5,0.75}	0.5	0.75	0.5
{EBAC}	{{8,15,4,11},{5,3,10,11},{5,3,14,11},{5,3,1,11},{20,3,14,11},}	{0.75,0.4,1.0,0.5}	0.4	1	0.5
{DBAC}	{{6,3,14,11},{6,3,1,11},{9,2,10,13},{9,2,10,7},}	{0.67,0.4,0.75,0.75}	0.4	0.75	0.67
{EDBAC}	{{12,9,2,10,13},{12,9,16,10,13},{5,6,3,10,11},{5,6,3,14,11},}	{0.5,0.67,0.6,0.75,0.75}	0.5	0.75	0.67
	{5,6,3,1,11},{12,9,2,10,7},{12,9,16,10,7},{5,6,3,10,7},}				

Fourth column gives participation index which is the minimum of participation ratio, fifth column gives max participation index which is maximum of participation ratio and the last column will give medoid participation index which is median of participation ratio as shown in figure 5.

## 5. CONCLUSION AND FUTURE WORK

In this paper we have extracted co-location patterns having high prevalence using a novel medoid approach. In literature the co-locations are based on participation index (PI) and MaxPI. Our experiments are based on the case study given in figure 1. we have extracted co-locations for all the five features as shown in figure 5. MaxPI gives the feature which is most crucial in the event in our case occurring of accidents. In this case if we want to control road accidents we have to stop moving of cars on roads which is ridiculous. So we cannot stop moving of cars on roads to avoid accidents. Hence it is concluded that generally maxPI gives feature which is contextually cannot be removed. Semantic gap will be more when we consider maxPI for identifying the feature to be checked in co-location mining. When we apply medoidPI the feature it gave is traffic police among three traffic police accidents are occurring while two traffic police were there. Hence action can be taken on these two traffic policeman so that accidents are to be controlled. The medoid approach can be used which gives the optimal results with a least contextually used feature as a factor. By using MedoidPI semantic gap is also reduced. In Future we want to extend the work on very large spatial database and conclude the results and we want to address the issue of semantic gap in formulating co-location patterns.

## 6. REFERENCES

- [1] Dasu, T., Johnson, T. *Exploratory Data Mining and Data Cleaning*, Wiley, 2003.
- [2] Hand, D., Mannila H., Smyth, P., *Principles of Data Mining*, MIT Press, 2001.
- [3] Griffith D. "Statistical and Mathematical sources of regional Science theory. Map pattern analysis as an example." Regional science RSAI 1999.
- [4] . M. S. Chen, J. Han, P. S. Yu. "Data mining, an overview from database perspective", IEEE Transactions on Knowledge and data Engineering, 1997.
- [5] . K Zeitouni "A survey of spatial data mining methods databases and statistics point of views", Data warehousing and web engineering, 2002 - books.google.com.
- [6] Longley P. A., Goodchild M. F., Maguire D. J., Rhind D. W., *Geographical Information Systems - Principles and Technical Issues*, John Wiley & Sons, Inc., Second Edition, 1999
- [7]. Diansheng Guo, Jeremy Mennis ,2009 "*Spatial data mining and geographic knowledge discovery—An introduction*", Computers, Environment and Urban Systems, Elsevier.
- [8]. Gordon, A. D. 1996. "Hierarchical classification. Clustering and classification" (pp. 65–122). River Edge, NJ, USA: World Scientific Publisher.
- [9]. Deren LI and Shuliang WANG, "Concepts, principles and applications of spatial data mining and knowledge discovery", in ISSTM 2005, August, 27-29, 2005, Beijing, China.
- [10]. Manuel Alfredo PECHPALACIO, "*Spatial Data Modeling and Mining using a Graph-based Representation*", PhD Thesis.
- [11]. R. Agarwal, T. Imielinski, and A. Swami. "Mining association rules between sets of items in large databases," in Proc. of the ACM SIGMOD Conference on Management of Data, Washington, DC, pp. 207–216, 1993.
- [12]. Han, J., Kamber, M., 2001, *Data Mining: Concepts and Techniques* (San Francisco: Academic Press)
- [13]. S Shekhar, P Zhang. *Data Mining and Knowledge Discovery 2010 – Springer*.
- [14]. Shekhar, S., & Huang, Y. (2001). *Discovering spatial co-location patterns: A summary of results*. In C. Jensen, M. Schneider, B. Seeger, & V. Tsotras (Eds.), *Advances in spatial and temporal databases, proceedings, lecture notes in computer science* (pp. 236–256). Berlin: Springer-Verlag.
- [15]. X. Zhang, N. Mamoulis, D.W.L Cheung, and Y. Shou. "Fast mining of spatial collocations", in *Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, pp. 384–393, 2004.
- [16] Ester M., Kriegel H.-P., Sander J., and Xu X. 1996 "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise". Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining. Portland, Oregon, AAAI Press, Menlo Park, California, pp. 226-231.
- [17] Sander J., Ester M., Kriegel H.-P., and Xu X. 1998 "Density-Based Clustering in Spatial Databases: A New Algorithm and its Applications", *Data Mining and Knowledge Discovery, an International Journal*, Kluwer Academic Publishers, Vol.2, No. 2.
- [18]. Li D.R., Wang S.L., Li D.Y. and Wang X.Z., 2002, *Theories and technologies of spatial data knowledge discovery. Geomatics and Information Science of Wuhan University 27(3)*, 221-233.
- [19]. Spielman, S. E., & Thill, J. C. , 2008, "*Social area analysis, data mining and GIS*". *Computers Environment and Urban Systems*, 32(2), 110–122.
- [20]. Ester, M., Kriegel, H. P., & Sander, J. ,1997. "*Spatial data mining: A database approach*". In *Advances in spatial databases* (pp. 47–66). Berlin: Springer-Verlag Berlin
- [21]. Martin Ester, Hans-Peter Kriegel, Jörg Sander "*Algorithms and Applications for Spatial Data Mining*", *Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS*, Taylor and Francis, 2001.
- [22]. Yan Huang, Jian Pei, Hui Xiong, "Mining Co-Location Patterns with Rare Events from Spatial Data Sets *Geoinformatica* (2006) 10: 239–260.
- [23]. G. Kiran Kumar, T.Venu gopal and P.Premchand "A Novel method of modeling Spatial Co-location patterns on spatial Database", 2nd International conference ICFOCS 2011 held at IISc Bangalore, India Aug 7-9, 2011.
- [24]. S. Shekhar, and Y. Huang. "Co-location rules mining: A summary of results," in *Proc. 7th Intl.Symposium on Spatio-temporal Databases*, Springer, Berlin Heidelberg New York, p.236, 2001.