# Trend detection and prediction in spatial databases: A practical report

Arvind Sharma
RJIT Tekanpur
arvinddevansh@rediffmail.com

R K Gupta
MITS Gwalior
iiitmrkg@gmail.com

***Abstract:*** **Data mining is a process of extracting potential useful and interesting pattern hidden in large data set. Spatial data mining is emerged as new era in research community. Spatial data mining is mining of spatial data.A trend may be time based pattern that can change according to distance and other nonspatial attributes.Trends are set in many real world applications like rising or falling of interest rates ,rates of unemployment ,rent cost according to location, in weather forecasting, geo referenced data applications etc. A spatial trend may be defined as a pattern of systematic change of one or more non spatial attributes in multidimensional space.In this paper, a complete review of spatial data mining is demonstrated with application of real data sets. Specially, Kmeans++ and Spectral clustering method is demonstrated with results.**

***Keywords:*****Spatial Data Mining, Spatial Databases, Kmeans, Spectral clustering, Trend Detection, Trend Prediction etc.**

## I. INTRODUCTION

Data mining is a process of extracting unknown useful information from large dataset. In today's scenario the most active research area covers a correlation of mining information with spatial dataset leading to the evolution of Spatial data mining (SDM). Spatial data mining stores the information in spatial database which helps to discover the interesting and previously unknown useful pattern from large datasets and stores the space objects [1] such as roads, rivers, forests, deserts, buildings, cities etc. Various geometric type database are used to store spatial data such as point, line string, polygon, arc line string, arc polygon, compound polygon etc.

The higher organization including NASA, the National Geospatial-Intelligence Agency [2], The National Cancer Institute [3], and the US Department of Transportation [4]execute their major decisions by extracting information from geo spatial data which are further extracted from spatial database spreading across many application domains such as ecology and environmental management [5, 6, 7, 8], public safety, transportation [9], Earth science [10] Epidemiology [11], crime analysis [12] and climatology [13, 14]. Spatial database along with spatial data consists of geographical data where relationships of variables are depicted to avoid error between identically distributed data which are spread over the geographical area. A major application of SDM is implemented in social media which differentiate the data mining as well as spatial data mining in terms of extracting information from data. The data interpolated in data mining is plain data whereas spatial data mining interpolate both spatial as well as non-spatial data mining.

Thus one main important aspect of data mining is to identify patterns in given data. In case of SDM the spatial patterns may be categories as: 1. Discontinuities and Spatial outlier 2.Co-occurrence and Co-location 3.Spatial clusters and Hotspot 4. Location Prediction.

This research paper has been divided into various sections and each section led to make more understanding of SDM. Section II explains about spatial data and data relationship. Section III gives idea about location prediction problem and algorithms in SDM, Section IV about statistics and machine learning concept in SDM, Section V summarize concept of social media mining to spatial mining followed by different application of SDM in Section VI.

## II. SPATIALDATAANDDATARELATIONSHIP

Data used in mining can be categories in two categories: Simple data also called Non-Spatial Data and Geo-reference data also called Spatial data. Non-spatial data are generally denoted as number or text string whether as Spatial data are denoted by longitude, latitude, elevation (location attributes), neighborhood and extent. For example city name and its population is non-spatial data whether as its city location isspatialdata.

To represent spatial data we need to separate data types. These data types may be categories in following three categories:

1. **Raster data type:** Raster data is similar to our digital camera dataset or satellite imagery data, which we see in Google earth. Whole area is divided into grids (i.e. row and column and its intersection point also called pixel). One

purpose of this data type is to know land type whether given area is forest orriver. **Raster** data represents the landscape as a rectangular matrix of square cells.As above mentioned the raster data is represented as gridded space i.e. row,col,and pixels form . Examples of raster data type are Satellite images, captured pictures of large areas and many more. These images are used to identify division of land in lakes,buildings,empty land etc. A satellite image of a city is shown in figure below-


Figure 1 - Example of raster data in grid form as satellite image.Raster (Courtesy:UMN)

2. **Vector data type:**Example of vector data types are point, line and polygon. Vector maps are composition of these three datatypes.Polygon features are two dimensional and therefore can be used to measure the area and perimeter of a geographic feature. Boundaries of a city or country may be represented by polygon.
Line (or arc) data is used to represent linear features. Common examples would be rivers, trails, and streets.

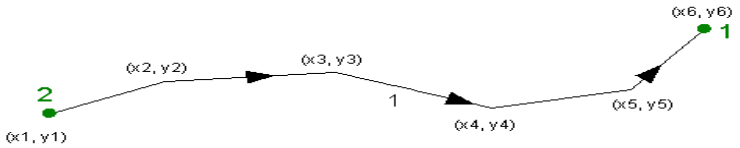


Figure 2 : Vector form of data representation as a line.

3. **Graph data type:**Here data is represented as node, edge and path. GPS devices use this for navigation and routing purpose. Here GPS devices may use road intersection as node, there segment as edge and this complete model asgraph.In figure 3, black dots represent nodes(major city locations) and linkages represent edges(links and routes b/w locations)

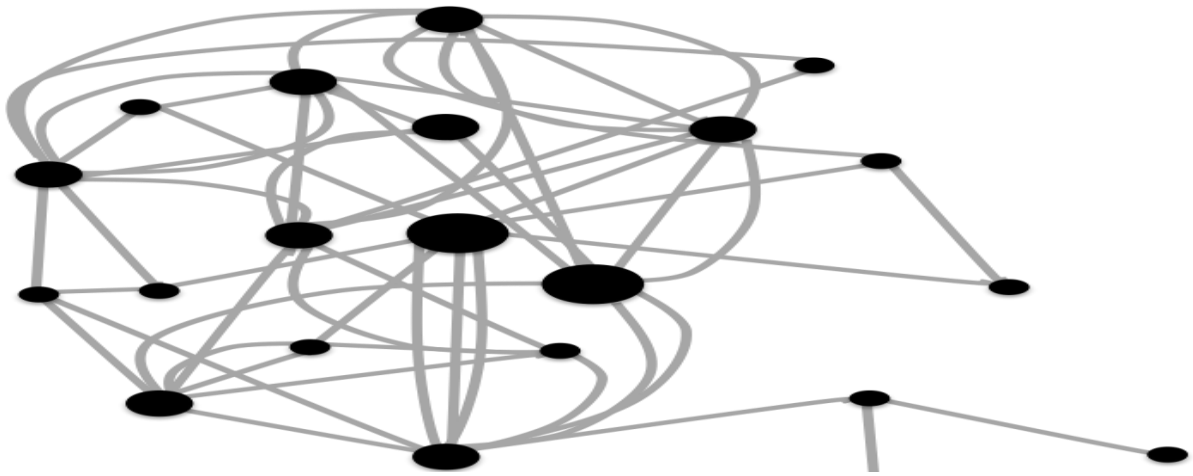

Figure 3 : A well connected transportation network of a city.

Data relationship is also an important factor in SDM as here data relationship is different from traditional data set. This relationship may be categories asfollowing manner:

1. Non Spatial Relationship:These are explicitly stored in database. Example- New Delhi is the capital of India.
2. Spatial Relationship: These are very complex and backbone of spatial data mining process.

**Topological:** Intersect, meet, overlap, with-in, boundary, etc.

**Metric:** Area, distance, perimeter etc.

**Directional:** North, South, east, west

**Raster relationship:**

i. Local

ii. Focal

iii. Zonal

These data types and relationships are formally defined as OGC(Open GIS Consortium) standard and supported by most of the software like Oracle, IBM DB2, PostGIS etc. Still there are scopes of improvement on direction prediction, three dimensional data mining, visibility, moving object etc.

In detail relationship of spatial data is shown below in table[2]-

| Data Type | Non spatial Data | Spatial data |
|---|---|---|
| Relationship details | Explicit | Many are found implicit |
| | Numerical,arithmetic,ranking,ordering etc. | Vector or topological relationships |
| | Object is instance of a class | Set oriented : membership union and intersection etc |
| | Class is a subclass of another class | Topological : Within ,Overlap |
| | Object is part of another object | Directional: North,NE,Left,Down,above,behind etc. |
| | Object is a member of a set | Metric : area<br>Euclidian :Distance, Perimeter<br>Dynamic : Modify,Create ,delete |

| | | |
|---|---|---|
| | Examples: Operational data such as ,sales,cost etc.<br>Non operational data such as Industry sales.<br>Meta data such as data dictionary definition.<br>A person's height, age, mass are non spatial data because they are independent of person's location.<br>It is clear that non spatial data is also called attribute or characteristics data as numbners,characters, logical etc. are very good example of non spatial data.<br><br>File types: .tiff ,.gif etc. | Shape based and visibility.<br><br>Examples : An inventory of assets, A person's weight and any other object which shows relation with space and location.<br><br>Spatial data=Spatial (Where) + Data (What)<br>File types : .csv(comma separated variable), .lyr(Layer file), . shp (Shape file), TIN(Triangular Irregular Network),.rgs ,.prj(Projection file), .dbf(data base file), .tab(Tabular file),.txt(Text File) , .mif ,.tgt ,.grd ,.e00 etc. are basic file types which are used for storage of spatial data.<br>Here it is noticeable that mass is an example of non spatial data while weight is spatial data. |

Table 2: Spatial vs Non Spatial relationships details.

Here a data set is given in the form of CSV format[23][24] and its cluster forming is shown with the help of kmeans++ method and spectral clustering methods.

Data set in CSV file format[23][24].

| Id | Location | Opening Hours | Dates | Seniors/children | Latitude | Longitude |
|---|---|---|---|---|---|---|
| 1 | Church of Nazarene, 6B Meadowlands Road, Carina | 9am - 11am | Tuesday weekly | children/seniors | -27.4913 | 153.1128 |
| 2 | Chermside Library- meeting room, 375 Hamilton Road Chermside | 1pm - 3pm | Tuesday weekly | children/seniors | -27.3858 | 153.0351 |
| 3 | Corinda Community Health Centre, 2 Clara St, Corinda | 1.30pm - 3pm | Thursday weekly | children/seniors | -27.5371 | 152.9849 |
| 4 | Forest Lake Community House, corner Forest Lake Boulevard and Alpine Place, Forest Lake | 9am - 10.30am | Monday fortnightly | children/seniors | -27.6112 | 152.9617 |
| 5 | Holy Family Church Hall (lower level), ward St, Indooroopilly | 9am - 11am | Wednesday weekly | children/seniors | -27.4999 | 152.9818 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 6 | North West Community Health Centre, 49 Corrigan Street, Keperra | 9am - 11am | Monday fortnightly | children/seniors | -27.4134 | 152.9588 |
| 7 | Mitchelton Library, 37 Heliopolis Parade, Mitchelton | 9am - 11am | Thursday weekly | children/seniors | -27.4167 | 152.9783 |
| 8 | New Farm Library- meeting room, 135 Sydney Street, New Farm | 1pm - 3pm | Monday fortnightly | children/seniors | -27.4663 | 153.0498 |
| 9 | Nundah Community Health Centre, 10 Nellie Street Nundah | 1pm - 3pm | Monday fortnightly | children/seniors | -27.3987 | 153.0647 |
| 10 | Paddington Community Health Centre, 202 Given Terrace Paddington | 1pm - 3pm | Wednesday weekly | children/seniors | -27.4615 | 153.0058 |
| 11 | British Australia Club (lower level) Hancock Street Sandgate | 9am - 10.45am | Wednesday weekly | children/seniors | -27.3208 | 153.0677 |
| 12 | Sunnybank Hills Library, Sunnybank Hills Shopping Centre, corner of Crompton and Calam Roads, Sunnybank Hills | 1pm - 3pm | Thursday weekly | children/seniors | -27.6109 | 153.0542 |
| 13 | Garden City Library, Corner Logan and Kessels Road, Upper Mount Gravatt | 1pm - 3pm | Wednesday weekly | children/seniors | -27.5595 | 153.0811 |
| 14 | Wynnum Hall, Corner Bay Terrace and Cedar Street, Wynnum | 9am - 10.45am | Thursday weekly | children/seniors | -27.4467 | 153.1769 |

Table 1: A sample of Data set for different public locations and offices in Australia.[23]

**k-Means:** The k-Means [HK00][27] algorithm takes the input parameter,k, and partitioned a set of n objects into k clusters so that the resulting intra clustering similarity is high but the intercluster similarity is low.Cluster similarity is measured in regard to the mean value of the objects in a cluster,which can be viewed as the cluster's center of gravity.

The k-Means algorithm works as follows –

First it randomly select k of the objects, each of which initially represent a cluster mean. For each of the remaining objects ,an object is assigned to the cluster to which it is the most similar,based on the distances between the objects and the clusters centers.Then it computes the new mean for each cluster.This process continues until the criterion function converges.The squared error criterion is used ,defined as

$E = \sum_{i=1}^{k} \sum_{p \in c_i} |p - m_i|^2$ ,

Where E is the sum for square error for all objects in the database ,p is the point in space representing a given object, and $m_i$ is the mean of the cluster $c_i$ .

A very good advantage of k-means algorithm is that it is very easy to implement.

A major drawback of this algorithm is that it is sensitive to outliers because an object with extremely large value may distort the distribution of the data.

**Spectral clustering** : Spectral clustering in spatial data mining plays a very important and innovative role due to its capacity of handling of large size of data ,effective application of linear algebra to solve graphical representation and problems,and application of very low cost of clustering algorithms like k-nearest or $\epsilon$ neighbourhood graph.Most of the research in this area is focused on efficient query processing for static or dynamic data. . Spectral clustering, as its name implies, makes use of the spectrum (or eigenvalues) of the similarity matrix of the data. It examines the *connectedness* of the data, whereas other clustering algorithms such as k-means use the *compactness* to assign clusters. Basically spectral clustering is a large family of combination of methods and it plays active role in research of machine learning, data mining due to its universality, efficiency and supporting result of practical output.

Spectral clustering[26] is especially applicable on moving objects and trajectories, spatially connected social network, spatial information in web based documents, geographic and graphical representation of multidimensional objects etc.

A sample output is shown in figure 1 which is based on spectral clustering (Laplacian methods).

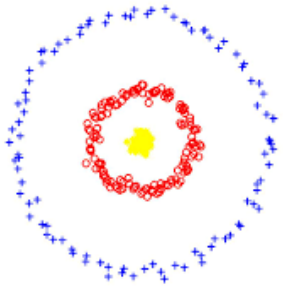

Figure 4: Result of spectral clustering algorithm

On the basis of spatial attributes it is shown in results that all those locations and offices may form clusters which are similar in spatial attributes. On the basis of two attributes as longitude and latitude , following results are generated.

**Result of k-Means++ method-**

---

Number of clusters: 5
Transformation: Standardize
Initialization method: k-Means++
Initialization re-runs: 50
Maximal iterations: 1000
Method: Arithmetic Mean
Distance function: Euclidean
Cluster centers:

| | Latitude | Longitude |
|---|---|---|
| C1 | 0.249561 | -0.664577 |
| C2 | 1.2236 | 0.304746 |
| C3 | -1.31702 | 0.49228 |
| C4 | 0.0451998 | 1.71988 |
| C5 | -1.18748 | -1.00784 |

The total sum of squares: 26
Within-cluster sum of squares:

| | Within cluster S.S. |
|---|---|
| C1 | 1.96608 |
| C2 | 0.643499 |
| C3 | 0.273073 |
| C4 | 0.656959 |
| C5 | 0.4449 |

The total within-cluster sum of squares: 3.98451
The between-cluster sum of squares: 22.0155
The ratio of between to total sum of squares: 0.846749

**Result of spectral clustering method-**

---

Number of clusters: 5
Gamm (Guassian Kernel): 1.0
Transformation: Standardize
Distance function: Euclidean
(K-Means) Initialization method: KMeans++
(K-Means) Initialization re-runs: 50
(K-Means) Maximal iterations: 300
(K-Means) Method: Arithmetic Mean
Cluster centers:

| | Latitude | Longitude |
|---|---|---|
| C1 | -0.286795 | -0.871242 |
| C2 | 0.457934 | -0.410504 |
| C3 | -0.210705 | 1.0484 |
| C4 | -0.91717 | 0.743978 |
| C5 | 1.7819 | 0.493815 |

The total sum of squares: 26
Within-cluster sum of squares:

| | Within cluster S.S. |
|---|---|
| C1 | 3.10164 |
| C2 | 3.74777 |
| C3 | 3.21664 |
| C4 | 1.41677 |
| C5 | 0 |

The total within-cluster sum of squares: 11.4828
The between-cluster sum of squares: 14.5172
The ratio of between to total sum of squares: 0.558353

According to application of two major clustering methods these results are generated . In above paragraph number of clusters and relationship between clusters are shown which are directly or indirectly applicable in real world problems as land cost estimation, density of residential area etc.

## III. Location Prediction

The first law of geography says that nearby things are similar. This indicates that spatial data samples are not independent. Some example of location prediction problems are: Where specific species of birds like to make their nests? Predicting spread of some infectious disease, predicting tornado or time and location of recent storm in India called HUDHUD.

If we talk about algorithms used in Data Mining some common algorithm for location prediction are: Linear regression model, Bayesian classifiers, neural networks, decision trees and SVM algorithms.

As nature of SDM is different these algorithms are not fully able to solve the problems. In SDM we use *Spatial autoregressive model (SAR)* and *Markov random field (MRF) based Bayesian Classifier* algorithms.

*Spatial Autoregressive Model (SAR):*

One of the commonly used auto regression models is spatial autoregressive regression Model (SAR) which uses the concept of contiguity matrix. Using the model [22] spatial autoregressive regression can be written as:

$$Y=\rho WY+X\beta+\varepsilon \quad (1)$$

Where

Y depicts the dependent variable
$\rho$ depicts the parameter of spatial autoregressive W depicts the contiguity matrix
$\beta$ depicts the regression coefficient
X depicts the error which is not observable
$\varepsilon$depicts independent variable

The SAR model is reduced to ordinary least square regression where $\rho$=0.The variability change with the distance basically plots the dissimilarity against the spatial separation [23].
Memory required: = O ($n^2$)
Disadvantage: Due to its I/O challenges several efficient techniques have been proposed to solve SAR. Many of these techniques have been studied and compared in [24]

## IV. STATISTICS IN SPATIALDATA

Input in spatial data mining consists of two distinct types of attributes: spatial and non-spatial where spatial is discrete representation of continuous phenomenon where spatial locations are defined and the spatial objects are extended[16, 17].There are basically two different spatial mining tasks to summarize the data in order to find the rules to characterize the data so that deviation can be detected easily. These are:

i. Statistics
ii. Machine learning

Different methods are carried out with these two approaches to describe data in global way which includes the factorial or variance analysis whereas machine learning applies the concept of generalization method to raise the abstract level of non-spatial attributes. To partition a given dataset depending on a similarity function the spatial data statistics uses the database approach where similarity function is classed as semantic distance. To reflect the data relationship the approach called local autocorrelation, assessing the degree of spatial dependencies and machine learning use association rule [18] with the help of technique "Kriging" technique to predict the location value outside the sample [19].there are basically two different types of Kriging technique:

*i.* *Ordinary kriging*: a technique which assumes random variable as a stationary moment
E(Z(xi))= E(Z(x0))= m, (2) where m is unknown.

*ii.* *Unbiased kriging*: Since the random function [20] is stationary, the following constraint is observed:

$$E(\epsilon(x_0)) = 0 \Leftrightarrow \sum_{i=1}^{N} w_i(x_0) \times E(Z(x_i)) - E(Z(x_0)) = 0 \Leftrightarrow$$

$$\Leftrightarrow m\sum_{i=1}^{N} w_i(x_0) - m = 0 \Leftrightarrow \sum_{i=1}^{N} w_i(x_0) = 1 \Leftrightarrow \mathbf{1}^T \cdot W = 1 \quad (3)$$

A spatial and non-spatial dependency is created when a strong bonding amongst an inclinational of an event is formed exhibiting the similar pattern which is calculated by spatial autocorrelations. There are basically two types of autocorrelation exhibited by spatial data:

**i.** **Positive auto correlation**: A positive auto correlation exhibits the dependencies of events creating the similar patterns.

**ii.** **Negative auto correlation**: A negative auto correlation represents the random pattern.

## V. SOCIAL MEDIA MINING TO SPATIALMINING

The technologies related to social network represents the geographic pattern of the communication network [23].Basically social media uses the geographical distribution of the users communicating with each other. The social media movement reflects the geospatial characteristics and information sharing practices of social media in which users are engaged in the occupied communication. There are so many social networking sites activated including Facebook, GooglePlus, Twtter, Instagram out of which taking into account twitter is popular social networking site in recent literature[25] where "Twitter" is considered as most active microblogging site where easily anyone can tweet his ideas and aspects where interstate and intrastate communication follows the different communication patterns[25].

Many challenges lies ahead of spatial data mining but still it needs to be more explored with the promising and fruitful

research results.

To correlate spatial data mining to social networking it can be relate as the events social networking users are highly influenced by the events of others which lead them to connect to each other forming the networks.

*Information extraction:*

The whole scene of social media is of extracting the more up-to-date information. The more up-to-date information means the more number of active users. The information extraction includes different pattern [25] where data can be analyzed effectively at various high levels. Different patterns and problems to effective area can be computed such as the interstate and intrastate communication pattern. It also concludes that how the infected disease affects and the precautionary aspects of affected areas. Recently one city of India named, Visakhapatnam battered by cyclone Hudhud, the damage assess caused the calamity. With the industrial sector suffering massive damages in the cyclone, the damage caused by Cyclone Hudhud not only changed the landscape of the port city, but also made it the first city inthe country to be directly hit by a cyclone since 1891. Visakhapatnam looks like Battered and bruised after Cyclone Hudhud struck the port town with wind speed of over 200 kmph. The eye of the storm passed at Kailashgiri where the

Indian Meteorological Department's weather station is located, damaging equipment, ironically not allowing measurement of wind speed to happen at the place where the eye of the storm had landfall. The streets littered with fallen trees, banners ripped out, broken statues of leaders, traffic lights lying broken on the road and many telephone and electricity poles down. Even the widespread of cyclone used the concept of Spatial data mining where the information was analyzed with the knowledge base management system and the database was maintained.We have tried to depict the internal workflow of SDM in the following figure 5.

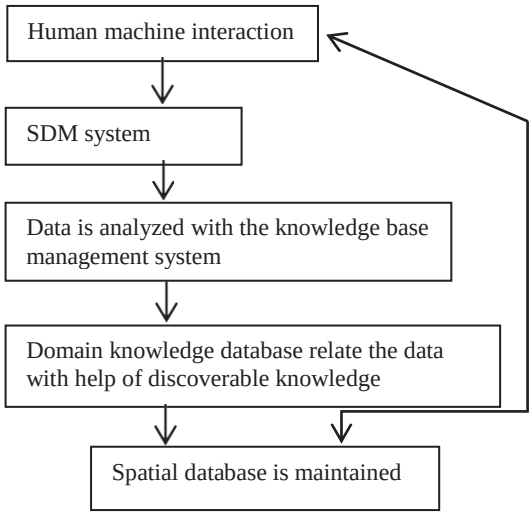


Figure.5: Work model of SDM

In the above figure the user face is the human machine interaction which forwards the data to the SDM system where data is analyzed with the knowledge base management system. The database related to domain knowledge relates the data with the help of discoverable knowledge maintaining the database of spatial data mining. The database maintained under the spatial data mining includes the following:

i. Set of variables describing theobjects

ii. Set of messages for communicating with different objects.

There[22] has been relatively little cross fertilization of data where different techniques investigate the pattern and process within geo referenced data to meet the ground level of priori and a posteriori points of views of the widespread disease such as “huddhudd” and many more where infected

disease follows a landscape pattern .Various researches are being conducted to provide a comprehensive set of

analytical tools in order to provide a link between the landscape pattern and disease influence which prevail with space and time. The contiguity of the relationship between the spatial data and the pattern is exclusively intra thematic where objects belongs to the same theme using join operator with various spatial criteria which are complex and time consuming and various authors have tried to optimize it using various techniques[21].

## VI. APPLICATIONS OF SDM

One of the major applications of spatial data mining is geographic information system in which data is represented in a grid of cells called pixels where each cell resolution depends on its pixel size [25] where more the pixel size more is theresolution.

### *i. Data structure in SDM*

Data structure in spatial data mining also correlate the communication networks in SDM the neighborhood graph and the path leads to the connection of objects. The relation neighbor depicts the topological and the metric relation in which distance of the objects can be measured to check the conductivity or disjunctively of the previous relation[15] Whereas non-spatial features of the objects including name, population and unemployment rate for city is characterized with the help of non-spatial data.

### *Spatio-temporal DM:*

The events and processes representing the change of some kind is represented using spatio-temporal DM where an ongoing phenomenal represent the activities without specifying the endpoints.

### *Criminology:*

Every individual has a specific mindset of their rational choice. Spatial process leads to two concepts either heterogeneity or dependency where at local level a change occurs across a defined space where conditions of racial composition vary evenly.In the case of criminals, they have a wide range of variability where crime chose targets within context. Even the criminals think rationally thus making calculative decisions and always commit a crime against the potential benefits while if constrained by time, cognitive ability and the availability of information results in a limited rationality[20]

## VII. CONCLUSION

The research work conducted in this paper is an attempt to give an insight to recent trends and techniques related to spatial data mining domain. Various trends and techniques have been explained in this paper which are being used in spatial data mining.In this paper, a review of spatial data mining is demonstrated with real time data sets and two most common and popular methods(Kmeans++,Spectral clustering) are applied. On the basis of spatial attributes it is shown in the results that all those locations and offices situated in an European country may form clusters which are similar in spatial attributes. Two basic attributes as longitude and latitude have been used.

## REFRENCES


[1] AakunuriManjula, Dr.G.Narsimha “A review on spatial data mining methods and applications” International Journal of Computer Engineering and Applications, Volume VII, Issue I, Part II, July14

[2] Krugman. “Development, geography, and economic theory”.MIT Press, Cambridge, MA, 1995

[3] P.Albert and L. McShane.” A generalized Estimating Equations Approach for Spatially Correlated Binary Data:Applications to the Analysis of Neuroimaging Data. Biometrics” (Publisher:Washington, Biometric Society, etc.), 51:627–638, 1995

[4] S. Shekhar, T. Yang, and P. Hancock” An Intelligent Vehicle Highway Information Management System.” Intl Jr. on Microcomputers in Civil Engg(Publisher:Blackwell Publishers, 8, 1993)

[5] Issaks, Edward, and M. Svivastava “Applied Geostatistics”.In Oxford University Press, Oxford, 1989.

[6] R.J.Haining” Spatial Data Analysis in the Social andEnvironmental Sciences”. Cambridge University Press, Cambridge, U.K., 1989.),

[7] J.-F. Roddick and M. Spiliopoulou.” A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research.” SIGKDD Explorations 1(1): 34-38 (1999),1999

[8] R. Scally” GIS for Environmental Management” ESRI Press,2006. ISBN: 978-589481429

[9] L. Lang. “Transportation GIS” ESRI Press, 1999. ISBN: 978- 1879102471

[10] M. Hohn and L. G. A.E. Liebhold.” A Geostatistical model for Forecasting the Spatial Dynamics of Defoliation caused by the Gypsy Moth,

Lymantriadispar (Lepidoptera:Lymantriidae). Environmental Entomology"(Publisher: Entomological Society of America), 22:1066–1075,1993.)

[11] P. Elliott, J. Wakefield, N. Best, and D. Briggs. "Spatial Epidemiology: Methods and Applications." Oxford University Press, 2000. ISBN:978-0192629418

[12] M. R. Leipnik and D. P. Albert." GISin Law Enforcement: Implementation Issues and Case Studies"CRC, 2002. ISBN: 978- 0415286107.

[13] P. Stolorz, H. Nakamura, E. Mesrobian, R. Muntz, E. Shek, J.Santos, J. Yi, K. Ng, S. Chien, R. Mechoso, and J. Farrara." Fast Spatio- Temporal Data Mining of Large Geophysical Datasets"In Proceedings of the First International Conference on Knowledge Discovery and Data Mining, AAAI Press, 300-305,1995

[14] Y. Yasui and S. Lele. "A Regression Method for Spatial Disease Rates: An Estimating Function Approach. "Journal of the American Statistical Association, 94:21–32,1997

[15] Google Earth, 2006,http://earth.google.com

[16] P.Bolstad "GIS fundamentals : A first text on GIS" eider press,2002

[17] A.R. ganguly and k.Steinhauser"data mining for climate change and impacts" in ICDM workshops, pages 385-394,2008

[18] karineZeitouni"survey of SDM methods database and statistics point of views" research of PSIG project of CASSINI network, Florence from THEMAlaboratory

[19] Isobel C., "practicalgeostatistics ", applied science publisher, reprinted1987.URL:http://curie.ej.jrc.it/faq/introduction.html

[20] URL:http://en.wikipedia.org/wiki/Kriging

[21] MatheusC.J:chanP.K;andPiatetsky Shapiro G 1993 "systems for Knowledge discovery in database IEEE transaction on knowledge and databaseengineering.

[22] Arvind Sharma and R K Gupta "i*ntelligent knowledge discovery in spatial datasets".* IJRECE Vol.4,issue 1,Jan-March. 2016. pp 67-73.

[23] http://diva-gis.org/data

[24] https://freegisdata.rtwilson.com

[25] Rana Majumdar,Arvind Kumar at el ." Spatial Data Mining: Recent Trends and Techniques". ICCCS 2015

[26] ] Arvind Sharma,R k gupta .Spatial Data Mining with the application of spectral clustering:A trend detection approach.IJCA ,volume 173,No.2,Sept.,2017.

[27] J Han and M kamber .Data Mining concepts and techniques.Morgan Kaufmann ,2000.