

SD-miner System to Retrieve Probabilistic Neighborhood Points in Spatial Data Mining

¹Asst. Prof. S. G. Kulkarni, ²Asst. Prof. Padma D, ³Mr. Manjunath R. H
^{1,2,3}Department of Computer Science, Gogte Institute of Technology, Belgaum.

Abstract:

In GIS or Geographic Information system technology, a vast volume of spatial data has been accumulated, thereby incurring the necessity of spatial data mining techniques. Displaying and visualizing such data items are important aspects. But no RDBMS software is loaded with displaying the spatial result over a MAP overlay or answer spatial queries like “all the points within” certain Neighborhood. In this project, we propose a new spatial data mining system named SD-Miner. SD-Miner consists of three parts: A Graphical User Interface for inputs and outputs, a Data Mining Module that processes spatial data mining functionalities, a Data Storage Model that stores and manages spatial as well as non-spatial data by using a DBMS. In particular, the data mining module provides major spatial data mining functionalities such as spatial clustering, spatial classification, spatial characterization, and spatio-temporal association rule mining. SD-Miner has its own characteristics:

- (1) It supports users to perform non-spatial data mining functionalities as well as spatial data mining functionalities intuitively and effectively.
- (2) It provides users with spatial data mining functions as a form of libraries, thereby making applications conveniently use those functions.
- (3) It inputs parameters for mining as a form of database tables to increase flexibility. Result shows that significantly reduced and precise data items are displayed through the result of this technique.

I. Introduction

Due to the development of information technology, a vast volume of data is accumulated on many fields. Since automated methods for filtering/analyzing the data and also explaining the results are required, a variety of data mining techniques finding new knowledge by discovering hidden rules from vast amount of data are developed. In the field of geography, due to the development of technology for remote sensing, monitoring, geographical information systems, and global positioning systems, a vast volume of spatial data is accumulated. Also, there have been many studies

of discovering meaningful knowledge from the spatial data. Since the spatial data has its own characteristics different from the non-spatial data, direct using of general data mining techniques incurs many difficulties. So there have been many studies of spatial data mining techniques considering the characteristics of the spatial data. However, commercial tools for spatial data mining have not been provided. Currently, many commercial data mining tools are available, but these tools not support the spatial data mining functionalities. Also, while some academic spatial data mining tools such as Geo-Miner are available, there are almost no commercial spatial data mining tools. So, for easy using of spatial data mining for real spatial data applications, developments of spatial data mining tools are needed. In this paper, author proposes a new spatial data mining system named SD-Miner. SD-Miner supports four important spatial data mining functionalities: spatial clustering, spatial classification, spatial characterization, and spatio-temporal association rule mining. We first analyze characteristics of previous spatial data mining techniques and suggest techniques to improve their efficiency in developing SD-Miner [1].

II. Motivation

Spatial data mining that is, discovery of interesting, implicit knowledge in spatial databases is a highly demanding field because very large amount of spatial data have been collected in various applications, ranging from remote sensing, to Geographical Information Systems (GIS), computer cartography, environmental assessment and planning etc. Spatial data has its own characteristic different from non spatial data. Direct use of general data mining techniques incur many difficulties, so we need a spatial data mining tool for easy using of spatial data, for spatial data application, which can be easily done by SD-miner. The SD-miner supports four important spatial data mining functionalities as spatial clustering, spatial classification, spatial characterization and spatial temporal association rule mining [1].

III. Proposed Approach

This work attempts to develop a spatial data mining system, the SD-miner which consists of three parts: A graphical user interface for inputs and outputs, a data mining module that processes spatial data mining functionalities, a data storage model that stores and manages spatial as well as non spatial data by using a DBMS.

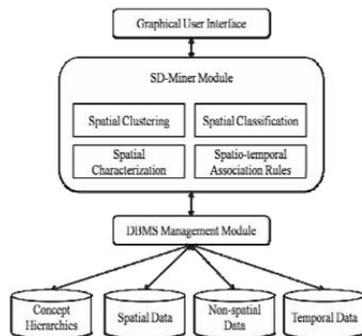
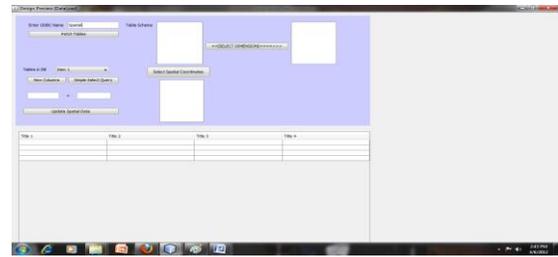


Fig 1: Architecture of SD-Miner

SD-Miner is composed of three main parts (Figure 1): Graphical User Interface (GUI), SD-Miner Module, and Data Storage Module (also, called DBMS management module) [1].

Graphical User Interface

A GUI is a type of computer human interface on a computer. It solves the blank screen problem that confronted early computer users. These early users sat down in front of a computer and faced a blank screen, with only a prompt. The computer gave the user no indication what the user was to do next. GUI is an attempt to solve this blank screen problem. Graphical user interface (GUI), a computer program that enables a person to communicate with a computer through the use of symbols, visual metaphors, and pointing devices. Best known for its implementation in Apple Inc.'s Macintosh and Microsoft Corporation's Windows operating system, the GUI has replaced the arcane and difficult textual interfaces of earlier computing with a relatively intuitive system that has made computer operation not only easier to learn but more pleasant and natural. The GUI is now the standard computer interface, and its components have themselves become unmistakable cultural artifacts.



Snap1- GUI

At a conceptual level, a computer human interface is a means by which "people and computers communicate with each other". In computer science terms, the GUI is a visual operating display that the monitor presents on the monitor to the computer operator. More specifically, a GUI is a specification for the look and feel of the computer system. GUI usually have common characteristic such as windows, icons, menus, and push-buttons (WIMP). Collectively, WIMP (Windows, Icons, Mouse, Pointer) are pictures that bring forth a certain action or an action space. The user issues commands via the GUI to computer applications. GUI usually has three major components. These three components are: a windowing system, an imaging model, and an application program interface (API). The windowing system builds the windows, menus, and dialog boxes that appear on the screen. The imaging model defines the fonts and graphics that appear on the screen. WIMP is products of both the windowing system and imaging model. Finally, the API is the means in which the user specifies how and what windows and graphics appear on the screen. The GUI contains some of the buttons and combo boxes which performs actions like:

1. Fetch tables- Which fetches the tables from the database.
2. .View column-Which displays columns according to the query written.
3. Simple select query-which specifies the condition in term of query.
4. Select spatial co-ordinates- this button helps us to select the required longitude and latitude.
5. Select dimensions.

SD-miner Module

The SD-Miner module processes respective data mining functions and transfers the results to the data storage module. This module provides four data mining functions as shown in fig-1.

Four important characteristics of SD-miner system:

Spatial Clustering

Spatial clustering classifies spatial objects as multiple groups according to its positional or geographical characteristics. This proposed work uses K means clustering. K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters). The main idea is to define k centroids, one for each cluster [2].

Spatial Classification

Usually, in the spatial classification, the objects are classified with considering of spatial and non-spatial attributes. The spatial classification also uses the decision tree. A big difference between the spatial classification and the classification is that the aggregation value of the spatial objects in a near region is used in the spatial classification. For making of a decision tree, the technique additionally uses predicates on relationship among spatial objects as decision criteria. For this, as the first step, the spatial attributes are represented as spatial predicates and then the possible useful predicates are extracted. For the second step, the decision tree is constructed with the predicates. The benefits of this technique are as follows. Since the decision tree is constructed after discarding ineffective predicates, the tree construction cost is greatly reduced. Speedy and correct classification can be made [3].

Spatial Characterization

Spatial characterization extracts a global outline of data classes for a given spatial region by using the spatial objects of the region. It gives simple and clear abstract information of the region. Spatial characterization evaluates whether the characteristics of given spatial objects are expanded to near region. To do this, the objects are defined as a neighbor of each other with considering of their distance or direction. The neighbor information is managed by using the neighbor table. The region handled by spatial characterization can be expanded with a spatial expansion algorithm using the neighbor table [4].

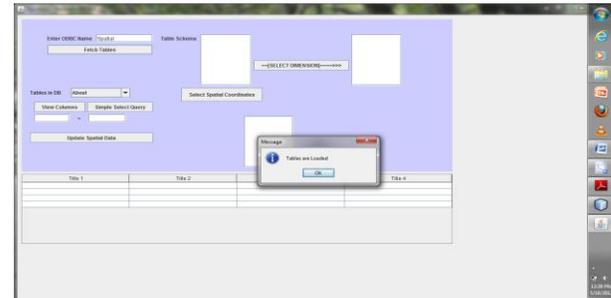
Spatio-Temporal Association Rule Mining

By using spatial association rule mining, proposed system can represent topological relationship and distance relationship of spatial objects via analyzing the relationship among spatial data and between spatial and non-spatial data. Also, by adding of temporal data analysis, proposed system can use spatio-temporal association rule mining. SD-Miner uses spatio-temporal

association rule mining. In order to use spatial association rule mining, the spatial relationships between spatial objects must be defined. The relationships are represented as spatial predicates. The predicates defined by a user are stored as concept hierarchy data in the database of SD-Miner. If the spatial relationship is defined as the predicates, the predicates can be regarded as non-spatial attributes. So spatio-temporal association rules can be extracted by using the well-known Apriori algorithm. Spatial association rule mining can be applied whether the predicates are spatial or non-spatial. So association rules can be extracted among spatial predicates or among spatial and non-spatial predicates. This technique can be applied hierarchically by using level-based predicates if exist. So, proposed system can extract detailed as well as abstracted association rules in this case [5].

RESULTS

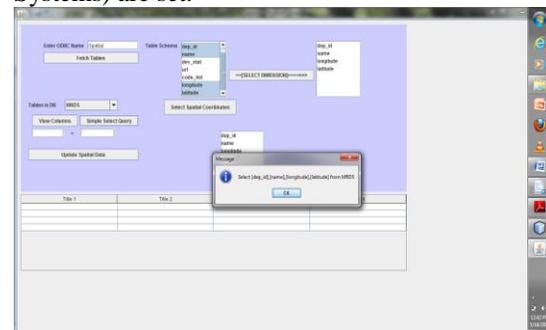
STEP 1: Load Data Base



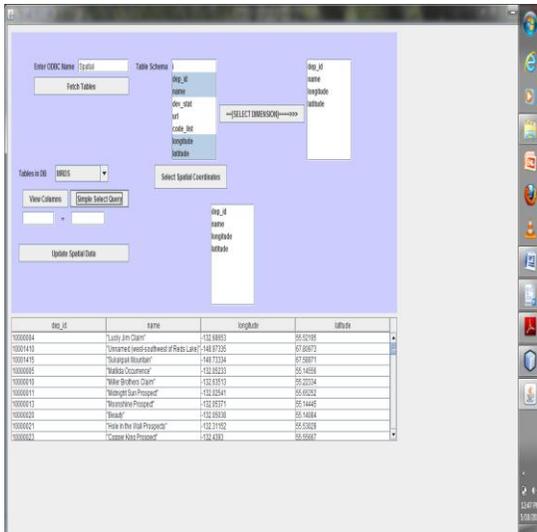
Snap 2: Illustrates Loading of Tables.

STEP 2: In Spatial Characterization phase, target objects will be set with respect to the database containing them as a description of the spatial properties.

In this proposed approach spatial properties like longitude, latitude and MRDS (Mineral Resources Data Systems) are set.



Snap 3: Illustrates Selecting of Required Characteristics.

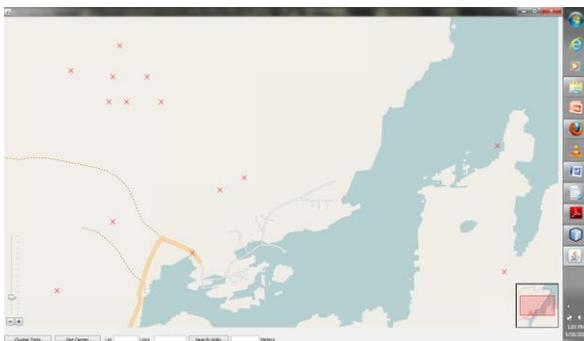


Snap 4: Illustrates list of selected characteristics in the form of columns and rows

Number of clusters	Range in meters
1	1000
4	2000
8	3000
25	4000
30	5000
38	6000
45	7000
52	8000
65	9000
71	10000

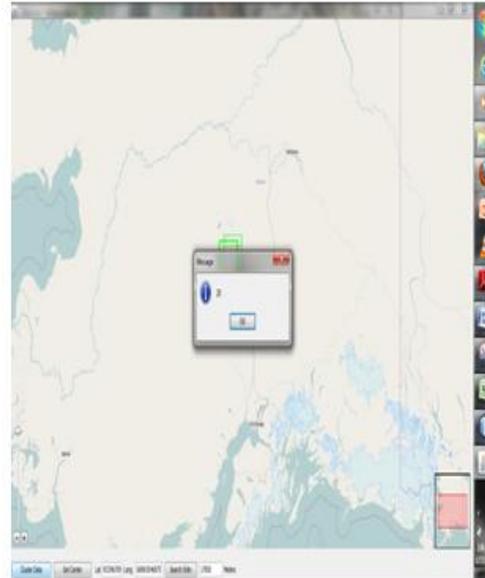
given set of classes based on the attribute values like longitude and latitude of the object.

STEP 3: In Classification phase, the objects (MRDS) are assigned to a respective classes from a



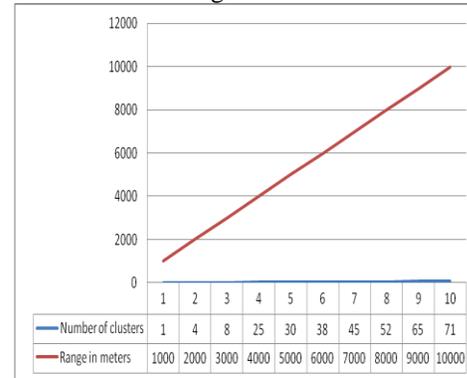
Snap 5: Illustrates The Classification Processes.

STEP 4: clustering classifies spatial objects (MRDS) as multiple groups according to specified longitude and latitude.



Snap 6: Illustrates clustering process.

Table 1: Illustrating number of clusters vs. Range



Graph: Number clusters vs. Range

The above table and graph shows us the variations of clusters according to the given range. As and when we change the parameters longitude and latitude the number of clusters will change, because it searches for the specified characteristic in the given range.

CONCLUSION

Spatial data has positional and topological data that do not exist in general data, and its structure is different according to the kinds of spatial data. Also, the objects on space affect each other and the relationship of objects is also different according to the kinds of objects. There have been many researches on spatial data mining considering these characteristics of spatial data. In this work, author explains the concept of spatial clustering, spatial classification, spatial characterization, and spatio-temporal association rule mining. We present our experiences in developing a spatial data mining system called SD-Miner that provides proper spatial data mining techniques with improved effectiveness and efficiency for real applications. SD-Miner adopts following techniques. For spatial clustering, proposed system adapt NN Superseeding algorithm. For spatial characterization, proposed system uses a binary decision tree and the RELIEF algorithm for efficiency. For spatial characterization, we use the neighbor table for spatial extension of the current characterization method. For spatio-temporal association rule mining, we use temporal data with spatial association rule mining using the spatial concept layer. SD-Miner uses spatial data mining functions extended from general mining functions. So, it can be applied to both of spatial and non-spatial data. Without special intervention of a user, it automatically recognizes which type of data is used. All functions are developed in library style, and the functions can be reused in another system.

IV. REFERENCES

- [1] SD-Miner: A SPATIAL DATA MINING SYSTEM Duck-Ho Bae, Ji-Haeng Baek, Hyun-Kyo Oh, Ju-Won Song, Sang-Wook Kim
- [2] J. Sander et al., "Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications," *Data Mining and Knowledge Discovery*, Vol. 2, No. 2, pp. 169-194, 1998
- [3] M. Ester et al., "Algorithms for Characterization and Trend Detection in Spatial Databases," In *Proc. Int'l. Conf. on Knowledge Discovery and Data Mining*, KDD, pp. 44-50, 1998.
- [4] K. Koperski, J. Han, and N. Stefanovic, "An Efficient Two-Step Method for Classification of Spatial Data," In *Proc. Int'l. Symp. On Spatial Data Handling*, SDH, pp. 45-54, 1998.
- [5] Spatio-Temporal Data Mining On Moving Objects In DBMS Yahaya Bin Abd Rahim GIS Section OTB Research Institute for Housing, Urban and Mobility Studies Delft University of Technology November 15 2007.
- [6] M. Ester et al., "Spatial Data Mining: Database Primitives, Algorithms and Efficient DBMS Support," *Data Mining and Knowledge Discovery*, Vol. 4, pp. 193-216, 2000.
- [7] M. Ester, H. Kriegel, and J. Sander, "Algorithms and Applications for Spatial Data Mining," *Geographic Data Mining and Knowledge discovery*, 2001.
- [8] J. Han, K. Koperski, and N. Stefanovic, "GeoMiner: A System Prototype for Spatial Data Mining," In *Proc. ACM Int'l. Conf. on Management of Data*, ACM SIGMOD, pp. 553-556, 1997.
- [9] W. Lu, J. Han, and B. Ooi, "Discovery of General Knowledge in Large Spatial Databases," In *Proc. Far East Workshop on Geographic Information Systems*, pp. 275-289, 1993.
- [10] E. Knorr and R. Ng, "Finding Aggregate Proximity Relationships and Commonalities in Spatial Data Mining," *IEEE Trans. On Knowledge and Data Engineering*, IEEE TKDE, Vol. 8, pp. 884-897, 1996.