SPATIAL DATA MINING FEATURES BETWEEN GENERAL DATA MINING

Mohit Rastogi*

*Assistant Professor, Department of Finance,
Teerthanker Mahaveer Institute of Management and Technology,
Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, INDIA
Email id: mohit.management@tmu.ac.in

DOI: 10.5958/2249-877X.2021.00116.8

ABSTRACT

Data mining is often described as the process of finding, analyzing, and sifting through vast quantities of data in order to discover connections, patterns, or statistical correlations. The practice of finding interesting, valuable, non-trivial patterns of information or knowledge from big geographical datasets is known as spatial data mining (SDM). Due to the complexity of geographical data types, spatial connections, and spatial auto-correlation, finding interesting and meaningful patterns from spatial datasets must be more challenging than extracting the equivalent patterns from conventional numeric or categorical data. Emphasis was placed on the distinctive characteristics that differentiate geographic data mining from traditional data mining, as well as the significant achievements of spatial data mining research. In precision agriculture, community planning, resource finding, and other fields, extracting intriguing patterns and rules from spatial information, such as remotely sensed images and related ground data, may be useful.

KEYWORDS: *Connection, Data Mining, KDD, Software, SDM.*

1. INTRODUCTION

In general, data mining (also known as data or knowledge discovery) is the act of evaluating data from many angles and synthesizing it into valuable information - information that may be utilized to generate revenue, reduce expenses, or both. Data mining software is one of several analytical techniques available for data analysis. It enables users to examine data from a variety of perspectives, classify it, and describe the connections discovered. Data mining is the process of discovering patterns or correlations in hundreds of variables in big relational databases[1]–[5].

The automated, exploratory examination and modelling of massive data repositories is known as Knowledge Discovery in Databases (KDD). The systematic process of finding legitimate, new, valuable, and comprehensible patterns from vast and complicated data sets is known as knowledge discovery.The heart of the KDD process is Data Mining (DM), which entails inferring algorithms that examine the data, build the model, and uncover previously undiscovered patterns. The model is used to explain phenomena based on a variety of facts, as well as to analyze and forecast outcomes.

Knowledge discovery and Data Mining have become more important and necessary because of the accessibility and quantity of data available today. For all situations, no one technique is better than the others are. The purpose of this article is to provide performance assessment methodologies and techniques, as well as to illustrate the usage of various methods in the spatial data-mining field using examples and software tools. The process of finding valid, innovative, valuable, and comprehensible patterns from big geographical datasets is known as spatial knowledge discovery in databases (SKDD).

The heart of the SKDD process is Spatial Data Mining (SDM), which entails inferring algorithms that investigate geo-data, build models, and find important patterns - the essence of usable information. The rapid increase of geographical data and extensive usage of spatial databases highlights the need for automated spatial knowledge discovery. The process of finding intriguing and previously unknown, but possibly valuable patterns from a set of geographical datasets is referred to as spatial data mining. Traditional Data Mining methods for extracting geographical patterns are limited by the complexity of spatial data and inherent spatial connections.

Statistical Data Miner and SAS Enterprise Miner are general-purpose data mining tools for analyzing big business datasets. Due to the complexity of geographical data types, spatial connections, and spatial auto correlation, finding interesting and meaningful patterns from spatial data sets is more challenging than extracting comparable patterns from conventional numeric and categorical data[6].

Our fuzzy classifiers' classification performance was similar, but in most instances inferior, to that of the support vector machine. This trend was particularly noticeable when the number of class data points was minimal. My research focus has moved away from fuzzy classifiers with strong generalization potential and toward support vector machine–based classifiers. To identify regularities and relationships between data points, data mining employs a huge quantity of computer power applied to a big collection of data. To search big datasets automatically, algorithms that utilize statistics, machine learning, and pattern recognition methods are employed. Knowledge-Discovery in Databases is another name for data mining (KDD)[7]–[9].

1.1 Additional Information on Spatial Data Mining:

What distinguishes geographical data mining from other types of data mining?

To begin, a spatial pattern includes the following features: Location prediction models, Spatial clusters (such as hotspots of global safety, Location as attribute (Location as attribute in spatial data mining), and so on are examples of spatial outliners.

Second, spatial data mining is utilized in a variety of applications, including Earth scientific data (such as the Earth Observing System), crime mapping, census data, transportation, social health, and public safety. For instance, we may respond to the following questions. How is the global Earth system evolving utilizing geographical data mining technology?

What is the Earth system's main forcing? What are the effects of natural and human-caused changes on the Earth system? What are the ramifications for human civilisation of changes in the earth system?

How effectively can we forecast changes in the Earth system in the future?

Finally, since data is distinct in the two main types of data mining, the related words, such as data, information, and knowledge, are also different. In the mind, many spatial connections between geographic things are not formally recorded. Different degrees of granularity allow for the construction of a mental representation appropriate for addressing a particular spatial issue at the granularity level.

1.2 Spatial Data Mining System Components:

There are five main components to data mining:

- Load transaction data into the data warehouse system by extracting, transforming, and loading transaction data.
- Use a multidimensional database system to store and manage the data.
- Give business analysts and information technology expert’s access to data.
- Use application software to analyze the data.
- Present the information in a logical manner, such as a graph or table.

Spatial data mining systems may be split into input, output, and computing process components, much like any other general-purpose data mining system. When data is entered into a data-mining program, it may be used to create a table with numerous columns representing each characteristic of a carton geographical feature. Figure 1 shows the raster and vector data for a campus.

Figure 1: The above figure shows the raster and vector data for a campus.

For example, tid: a feature's identifier; fi: spatial characteristics of the feature that are spatially linked; and non-spatial attributes such as conventional relational database.

The data in this database table may be divided into two main types: non-spatial and spatial.

Non-spatial data includes:

- data in conventional data mining
- Numerical, category, ordinal, Boolean, and other types of information. For example, the name of the city and the population of the city.

Spatial attribute:

- Spatially referenced, scope, and neighbourhood, such as feature location, e.g., longitude, latitude, and elevation.
- Methods for spatial data representation in the Raster data structure model, grid space, and in Vector, point, line, polygon, or Graph, node, edge, path, and so on.

Measurement may be used to determine a distance: Extended objects (buffer-based), Graph (shortest route), and Point (Euclidean) Due to the various data structures, transaction types have Circles centered at reference features, Gritted cells, Min-cut partitions, and Voronoi diagram space partitions.

1.3 Spatial Data Mining Methods

Because of the huge quantity of apprehension data utilized in this research, the data reduction technique of categorization may be able to uncover patterns that would otherwise be obscured by noise. The decision tree algorithm is a common classification method. The trained modal can find the smallest variables that may be used to explain certain occurrences. It also helps in focusing on smaller regions and identifying variable relationships. A part of a trained classifier mode is shown in the diagram below[10]–[12].

Clustering analysis may identify groups of points with similar magnitude values as well as outlier points, which are points that are not found in clusters. For example, the variable "Apprehension age" shows that the many low right points have comparatively low index values when it comes to apprehension age. This implies that these individuals' apprehension ages are vastly different. If the index value is positive, on the other hand, that subject's values should be comparable to those of its neighbours.

1.4 Spatial Data Mining Visualization

Visualization is crucial in multi-dimensional inference since it aids in the construction of hypotheses. Revisualization provides data pattern hints, which are subsequently used to drive data mining queries. The data mining findings are then displayed, resulting in more hints and more concrete study ideas. In our applications, this dynamic and interactive data pattern-searching technique has proved to be successful.

In our research, visualization is very important. Several tools have been put to the test to see whether they are suitable for the data type. A well-known collection of indices, such as mean, variance, and range, may be used to describe tables. It is critical to be able to characterize the potential connections between the values and ultimately identify clusters of data when working with big and multivariate datasets, as we did in our research.

It is possible to map multivariable GIS datasets into 3D views with Common GIS, a free Java program for geographic and geostatistical analysis, and overlay the findings with additional layers and information.

This has shown, for example, that when illegal immigrants are between the ages of 18 and 27, they prefer to enter the US from the interior. The majority of elderly and younger people are stopped closer to the border.

The depiction of the geographical distribution of the density of apprehension events over time is another intriguing use of GIS technologies and statistical analysis. Furthermore, by combining various periods, it is possible to construct animations showing density change over a longer period, utilizing weekly or monthly data as frame units.

Some popular graph techniques have been used to show the change and variability in time of the overall number of apprehensions or of repeated people, such as bar charts, line graphs, pie charts, and so on. The following findings may be allocated to a spatial sector of the area of research using GIS tools. For instance, a series of bar charts may be shown over a map of Mexican states to show how migration patterns have changed over time in each state.

1.5 Exploration of spatial data

The application of data mining techniques to spatial data is known as spatial data mining. The goal of spatial data mining is to discover patterns in data that are related to geography. Until now, data mining and Geographic Information Systems (GIS) were two distinct technologies with their own methodologies, traditions, and approaches to data presentation and analysis. Most modern GIS, in particular, only offer extremely limited spatial analysis capabilities. The massive influx of geographically referenced data brought on by technological advancements, digital mapping, remote sensing, and the worldwide adoption of GIS highlights the need of creating data-driven inductive methods to geographic analysis and modelling.

For GIS-based applied decision-making, data mining has a lot of promise. The challenge of merging these two technologies has recently become essential, particularly as many public and commercial sector organizations with large databases of thematic and spatially linked data recognize the enormous potential of the data contained within. These are some of the organizations involved:

- Offices that need geo-referenced statistics data analyzed or disseminated.
- Public health officials are looking for reasons for illness clustering.
- Environmental authorities are assessing the effect of shifting land-use patterns on climate change.
- Customers are segmented depending on their geographic location by remarketing firms.

Problems with Spatial Mining: Large geospatial data stores are common. Furthermore, current GIS datasets are often fragmented into feature and attribute components, which are then preserved in hybrid data management systems.

2. DISCUSSION

The author discussed about the data mining, Database systems for the administration of spatial data are known as spatial database systems. Geographical data mining techniques are critical for detecting implicit regularities, rules, or patterns buried in huge spatial datasets, such as for traffic management or environmental research. In general, predictive analytics (also known as data or feature learning) is the process of analyzing data from a variety of perspectives and synthesizing it into useful information, which may be used to create income, cut costs, or do both. For data analysis, data mining software is one of many analytical methods accessible. It allows users to look at data from different angles, categorize it, and explain the relationships they find. The technique of finding patterns or correlations in millions of variables in large relational databases is known as data mining.

3. CONCLUSION

The author has concluded about the data mining, Database systems for the administration of spatial data are known as spatial database systems. Relational (attribute) data management and topological (feature) data management have quite different algorithmic needs. The variety and diversity of geographic data formats, which offer distinct difficulties, is related to this. Beyond the conventional "vector" and "raster" forms, the digital spatial data revolution is spawning new data formats. Irregularly organized data, such as photography and geo-referenced multi-media, is increasingly being included in geographic data repositories. Geographical data mining techniques are critical for detecting implicit regularities, rules, or patterns buried in huge spatial datasets, such as for traffic management or environmental research.

REFERENCES

1. R. Tamilselvi and S. Kalaiselvi, "An Overview of Data Mining Techniques and Applications Keywords: Data mining Techniques; Data mining algorithms; Data mining applications 1. Overview of Data Mining," *Int. J. Sci. Res.*, 2013.

2. P. VIKRAMA, P and Radha Krishna, "Data Mining Data mining," *Min. Massive Datasets*, 2005.

3. F. A. Hermawati, "Data Mining Data mining," *Min. Massive Datasets*, 2005.

4. A. Twin, "Data Mining Data mining," *Min. Massive Datasets*, 2005.

5. B. A. B. Ii, "Data Mining Data mining," *Min. Massive Datasets*, 2005.

6. Y. Chen, D. Hu, and G. Zhang, "Data mining and critical success factors in data mining projects," in *IFIP International Federation for Information Processing*, 2006, doi: 10.1007/0-387-34403-9_39.

7. K. M. Raval, "Data Mining Techniques | Data Mining Articles," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, 2012.

8. T. L. Yang, P. Bai, and Y. S. Gong, "Spatial data mining features between general data mining," in *2008 International Workshop on Education Technology and Training and 2008 International Workshop on Geoscience and Remote Sensing, ETT and GRS 2008*, 2008, doi: 10.1109/ETTandGRS.2008.167.

9. E. T. L. Kusrini, "Data Mining Data mining," *Min. Massive Datasets*, 2005.

10. D. J. H. N. M.Adams, "Data Mining Data mining," *Min. Massive Datasets*, 2015.

11. G. Wang, J. H. Si, and C. F. Yang, "Research of data mining system," *Beijing Gongye Daxue Xuebao / J. Beijing Univ. Technol.*, 2005, doi: 10.2991/ameii-15.2015.78.

12. H. Petersohn, "Data-Mining-Anwendungsarchitelctur," *Wirtschaftsinformatik*, 2004, doi: 10.1007/BF03250992.