"Data Mining and Data Warehousing: Applications Measures of Multi-Dimensional Moving Data Object through Trajectory Data Warehouse"

Priya Vij

Research Scholar, Sai Nath University, Ranchi, Jharkhand

Abstract – Trajectory means a path subsequently affecting object. Spatio temporal clustering is a development of alliance objects based on their spatial. It is also known as Trajectory. Moving object is fully based on clustering algorithms.

Keywords: Trajectory, Data Mining, Knowledge, Moving Object

----------♦----------

INTRODUCTION

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Although data mining is a relatively new term, the technology is not. Companies have used powerful computers to sift through volumes of supermarket scanner data and analyze market research reports for years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost.

Data:

Data are any facts, numbers, or text that can be processed by a computer. Today, organizations are accumulating vast and growing amounts of data in different formats and different databases. This includes:

operational or transactional data such as, sales, cost, inventory, payroll, and accounting nonoperational data, such as industry sales, forecast data, and macro-economic data meta data - data about the data itself, such as logical database design or data dictionary definitions

Information:

The patterns, associations, or relationships among all this *data* can provide *information*. For example, analysis of retail point of sale transaction data can yield information on which products are selling and when.

Knowledge:

Information can be converted into *knowledge* about historical patterns and future trends. For example, summary information on retail supermarket sales can be analyzed in light of promotional efforts to provide knowledge of consumer buying behavior. Thus, a manufacturer or retailer could determine which items are most susceptible to promotional efforts.

REVIEW OF LITERATURE:

A spatio-temporal data warehouse like any data warehouse has three main segments: extract, transform, and load usually known as "ETL". In [1] the researcher put in with the proposal of the Extract-Transform-Load process for reconstructing moving objects, also they addressed the distinct presence measure. In [1, 2] the researcher assume a discrete model for representing moving objects trajectories, they present a technique for counting the distinct number of objects existing in a cell. The technique uses a star schema that involves both spatial and temporal dimensions. The authors also classified the aggregate functions into three collections: algebraic, distributive and holistic. Finally, in [3] the researcher presented an approach to compute the algebraic presence determines

using an SQL computation approach.

Warehousing spatial and mobility data:

The pioneering work by Han et al. [5] introduces the concept of spatial data warehousing (SDW). The authors extend the idea of cube dimensions so as to include spatial and non-spatial ones, and of cube measures so as to represent space regions and/or calculate numerical data. In [6], spatial OLAP operators are studied. One step further from modeling a SDW is modeling a TDW.

Data Warehouses:

Dramatic advances in data capture, processing power, data transmission, and storage capabilities are enabling organizations to integrate their various databases into *data warehouses*. Data warehousing is defined as a process of centralized data management and retrieval. Data warehousing, like data mining, is a relatively new term although the concept itself has been around for years. Data warehousing represents an ideal vision of maintaining a central repository of all organizational data. Centralization of data is needed to maximize user access and analysis. Dramatic technological advances are making this vision a reality for many companies. And, equally dramatic advances in data analysis software are allowing users to access this data freely. The data analysis software is what supports data mining. How does data mining work?

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought:

Classes: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.

Clusters: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.

Associations: Data can be mined to identify associations. The beer-diaper example is an example of associative mining.

Sequential patterns: Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

Data mining consists of five major elements:

Extract, transform, and load transaction data onto the data warehouse system.

Store and manage the data in a multidimensional database system.

- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table.

Different levels of analysis are available:

- **Artificial neural networks:** Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- **Genetic algorithms:** Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.
- **Decision trees:** Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID) . CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.
- **Nearest neighbor method:** A technique that classifies each record in a dataset based on a combination of the classes of the *k* record(s) most similar to it in a historical dataset (where *k* 1). Sometimes called the *k*-nearest neighbor technique.
- **Rule induction:** The extraction of useful if-then

rules from data based on statistical significance.

- **Data visualization:** The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

CONCLUSION:

Data mining (DM) techniques can be employed in order to convert the enormous amount of raw data into useful knowledge. The high volume of generated data arises the challenge of applying analytical techniques on such data. In order to achieve this aim, we have to take into consideration the complex nature of spatiotemporal data and thus to extend appropriately the two aforementioned techniques to handle them in an efficient way.

REFERENCES:

[1] E. F. G. Marketos, I Ntoutsi, N. Pelekis, A. Raffaeta, and Y. Theodoridis, " Building real world trajectory warehouses," in Proc. 7th International ACM SIGMOD Workshop on Data Engineering for Wireless and Mobile Access (MobiDE'08), Canada, 2008, pp. 8-15.

[2] S. Orlando, Orsini, R., Raffaetà, A., Roncato, A., and Silvestri, C, "Trajectory Data Warehouses: Design and Implementation Issues." Journal of Computing Science and Engineering, vol. 1, pp. 211–232, 2007.

[3] G. M. H. M. O. Mokhtar "Querying trajectory data warehouses," in DBKDA: The 1st International Conference on Advances in Databases, Knowledge, and Data Applications, March 2009, pp. 101–107.

[4] Leila Esheiba, Hoda M.O.Mokhtar, and Mohamed El-Sharkawi "Spatio-temporal queries for moving objects data warehousing" ,International Journal of Database Management Systems (IJDMS) Vol.5, No.3, June 2013

[5] Han, J., Stefanovic, N., and Koperski, K. Selective Materialization: An Efficient Method for Spatial Data Cube Construction. Proc. PAKDD, 1998.

[6] Papadias, D., Kalnis, P., Zhang, J., and Tao, Y. Efficient OLAP Operations in Spatial Data Warehouses. Proc. SSTD, 2001.