National Conference on Advances in Computer Science and Applications with International Journal of Computer Applications (NCACSA 2012)
Proceedings published in International Journal of Computer Applications® (IJCA)

Temporal Cluster graphs for visualizing Trends

B.Ratnamala, P.M.Kiran
Assistant Prof.- Gayatri Vidya Parishad College of Engineering, Department of Computer Applications, Visakhapatnam, India.

ABSTRACT

Organizations and firms are capturing increasingly more data about their customers, suppliers, competitors and business environment. Most of this data is multidimensional and temporal in nature. Data mining and business intelligence technique are often used to discover in such data. We propose a new data analysis and visualization technique for representing trends and temporal data using K-means clustering based approach. And we introduce a system that implements the temporal clustered graph construct which maps temporal data to a two dimensional directed graph that identifies trends in dominant data types over time. In this paper, we present our temporal clustered based technique and its implementation and performance.

Keywords

Clustering, Data knowledge and visualization, Data mining, Temporal data mining, Trend analysis

1. INTRODUCTION

Business intelligence applications represent to help firms gather and analyze information about their performance, customers,information about their performance, customers, competitors, and business environment. Knowledge representation and data visualization tools constitute one form of business intelligence techniques that present information to users in a manner that supports business decision-making processes. In this paper, we develop a new data analysis and visualization technique that presents complex multi-attribute temporal data in a cohesive graphical manner by building on well-established data mining methods. Business intelligence tools gain their strength by supporting decision-makers, and our technique helps the users leverage their domain expertise to generate knowledge visualization diagrams from complex data and further customize them. Organizations and firms are capturing increasingly more data, and this data is often transactional in nature, containing multiple attributes and some measure of time. For example, through their websites, e-commerce firms capture the click stream and purchasing behavior of their customers, and manufacturing companies capture logistics data (e.g., on the status of orders in production or shipping information). One of the common analysis tasks for firms is to determine whether trends exist in their transactional data. For example, a retailer may wish to know if the types of its regular customers are changing over time, a financial institution may wish to determine if the major types of credit card fraud transactions change over time, and a website administrator may wish to model changes in website visitors' behavior over time. Visualizing and analyzing this type of data can be extremely difficult because it can have numerous attributes (dimensions). Additionally, it is often desired to aggregate over the temporal dimension (e.g., by day, month, quarter, year, etc.) to match corporate reporting standards. The approach that we take in the paper for addressing these types of issues is to mine the data according to specific time periods and then compare the data mining results across time periods to discover similarities.

2. RELATED WORK

The field of data mining has developed a number of methods for identifying patterns in data to provide insights and decision support to users. Data mining and business intelligence approaches are often used for class identification and data visualization in knowledge management systems. Increasingly, knowledge discovery in data (KDD) techniques are providing new analytical structures that complement and sometimes replace existing human expert-based techniques to provide improved support for decision making. Identifying and visualizing temporal relationships (e.g., trends) in data constitutes an important problem that is relevant in many business, scientific, and academic settings. In this section, we provide a brief review of related research in the temporal data mining and visualization streams.

2.1 Temporal Data Mining

Temporal data mining is a growing stream in the knowledge discovery research field, and the technique we propose relates well to this stream. The goal of temporal data mining is to discover relationships among events and their sequences that have some form of temporal dependency. It also has the capability of mining activity rather than just states, which can lead to inferences about relationships and cause-effect association

2.2 Data Visualization

Data mining requires the inclusion of the human in the data exploration process in order to be effective. Visual data exploration is the process of presenting data in some visual form and allowing the human to interact with the data to create insightful representations. It typically follows the "visual information seeking mantra" overview, zoom and filter, and details on demand. Most formal models of information visualization are concerned with presentation graphics and scientific visualization provide taxonomies of visualization-based data exploration approaches and note that these approaches can be classified by 1) the type of data, 2) the visualization technique, and 3) the interaction techniques. With the dramatic increase in the amount of data being captured by organizations, multidimensional visualization techniques have become an important area of data mining research. Representing multidimensional data in a two- or three-dimensional visual graphic cannot be achieved through simple mapping, and many data visualization techniques have been developed

2.3 k-Means Method

This algorithm takes the input parameter k and partitions are a set of n objects into k clusters. It randomly selects k of the objects each of which initially represents a cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is the more similar based on the distance between the object and cluster mean.

The basic algorithm is

1: Select k points as the initial centroids

2**: repeat**

3: From k clusters by assigning all points to the closest centroid

4: Recompute the centroid of each cluster

5: **Until** the centroids don't change.

The details of K-means clustering algorithm is

a) Initial centroids are often chosen randomly

b) The centroid is mean of the points in the cluster

c) distance is measured by Euclidean distance

d) K-mean centroids move at each iteration

3. ANALYSIS AND RESULTS

The temporal cluster graph technique is applied on share values of different companies. The share values of all different companies basically which are in un-clustered format are grouped using k-means clustering algorithms so that the data of all different companies share values are compared at once and the comparison is shown by using some visualization techniques so that it is very easy to study the trends in share values. So the ability to identify trends in general temporal data can provide significant benefits, such as competitive advantages to a firm performing forecasts or making decisions on future investments and strategies.

The implementation process involves

- Creation of dataset
- Clustering of relevant data
- Extracting values according to user input N - (N - no of elements)
- Performing Visualization on clustered data for single company and multiple companies.

For creation of data set the data is collected from internet and stored in the excel sheet. Data is stored into dataset through import and export data wizard present in the SQL-Server and irrelevant data which is present in the dataset is separated and grouped using the clustering algorithm and Values are extracted from the clustered dataset details according to the value given by the user and display it in same page through which the comparison of current trend analysis of each company with the other companies takes place. Extracted values from the above process are given as the input to this module and the results are shown in the form of graph. The figures shows sample temporal cluster graph process for different company share values.

The proposed method is developed using JAVA programming language and a SQL-Server database. The Dataset is handled with the support of SQL-Server. For our work in the dataset, data is taken for five years with respect to every month for some companies. So this Un-clustered data is clustered using the K-Means technique. This Clustered data is visualized in the form of graphs with the support of JFreechart which consists of set of classes for representing graphs.

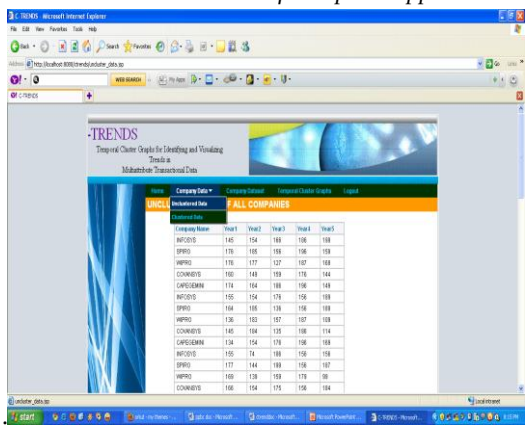


Fig 1: Un-clustered dataset details

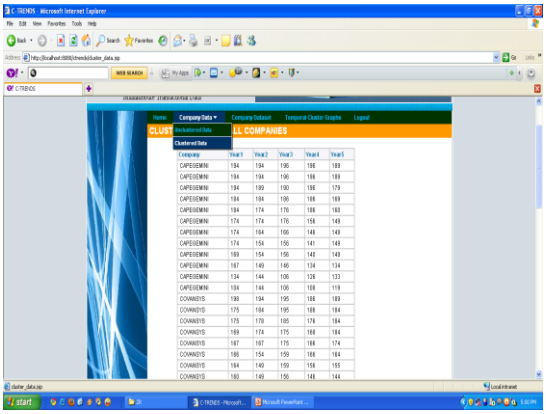


Fig 2: clustered dataset by company details

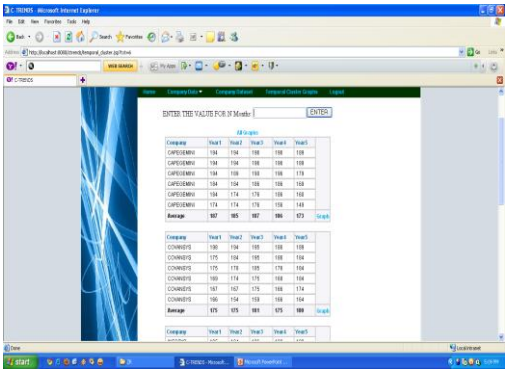
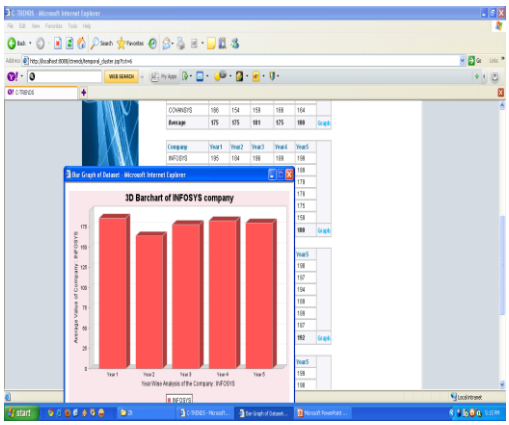


Fig 4: clustered data depends on Value N.



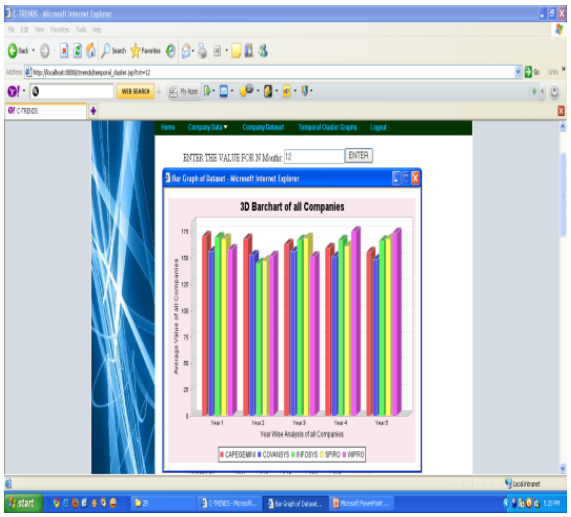


Fig 6: Comparison Graph for Multi-attribute representation

5. CONCLUSION

Temporal cluster graphs for visualizing trends presents temporal data in a unique and intuitive manner that emphasizes trends between dominant transaction types over time, and its output graphs resemble evolutionary diagrams and naturally portray the changes in data characteristics over time. It is a meta-analysis tool for data mining results and, therefore, it is designed to provide the domain expert with substantial control over the data presentation. By harnessing computational techniques of data mining, a new temporal clustering technique is developed for discovering, analyzing, and visualizing trends in multi attribute temporal data.

6. REFERENCES

[1] J. Abello and J. Korn, "MGV: A System of Visualizing Massive Multi-Digraphs," IEEE Trans. Visualization and Computer Graphics, vol. 8, no. 1, pp. 21-38, Jan.-Mar. 2001.

[2] R. Agrawal, K.I. Lin, H.S. Sawhney, and K. Shim, "Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time- Series Databases," Proc. 21st Int'l Conf. Very Large Data Bases (VLDB '95), pp. 490-501, 1995.

[3] M.S. Aldenderfer and R.K. Blashfield, Cluster Analysis.Sage Publications, 1984.

[4] C.M. Antunes and A.L. Oliveira, "Temporal Data Mining: An Overview," Proc. ACM SIGKDD Workshop Data Mining,pp. 1-13, Aug. 2001.

[5] C. Apte, B. Liu, E. Pednault, and P. Smyth, "Business Applications of Data Mining," Comm. ACM, vol. 45, no. 8, pp. 49-53, 2002.

[6] G.C. Battista, P. Eades, R. Tamassia, and I.G. Tollis, Graph Drawing. Prentice Hall, 1999.

[7]. Data Mining Concepts and Techniques, Jiawei Han and MichelineKamber

[8] .Data warehousing, Data mining and OLAP, Alex Berson, smith.j. Stephen

[9]Data Mining Introductory and Advanced topics, Margaret H Dunham PEA

[10] The Data Warehouse lifecycle toolkit, Ralph Kimball Wiley student Edition