JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EXPLORATORY VISUAL SEQUENCE MINING BASED ON PATTERN-GROWTH

DR.P. YOGANANDHINI
ASSISTANT PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE
VELLALAR COLLEGE FOR WOMEN, ERODE, TAMILNADU, INDIA

ABSTRACT

Exploring event sequences in big data is challenging. Sequential pattern mining finds applications in numerous diverging fields. Due to the problem's combinatorial nature, two main challenges arise. First, existing algorithms output large numbers of patterns many of which are uninteresting from a user's perspective. Second, as datasets grow, mining large numbers of patterns gets computationally expensive. Though many mining algorithms have been developed to derive the most frequently occurring and the most meaningful sequential patterns, it is yet difficult to make sense of the results. This work tackles this problem by combining interactive visualization with sequential pattern mining in order to create a "transparent box" execution model. Our proposed approach describes the design of Peek quence, which aims to increase the interpretability of machine learning-based sequence mining algorithms.

Keyword definition : sequential pattern mining, data mining, machine learning.

Introduction

As the size of data increases and focus shifts to the identification of meaningful and interesting sets of patterns, new approaches that can further and flexibly reduce the search space need to be used. Therefore, research aiming at integrating data mining with visualization has been gaining increasing interest, which is also the focus of our work. In doing so, the value of going away from "black box" model analysis approaches towards more transparent approaches has been lifted and the notion of progressive visual analytics was introduced promoting the importance of the interaction of the analyst with the mining algorithm.

To our knowledge, the possibility of changing and refining constraints while a particular sequence pattern is being built has not yet been considered, and it is an approach that addresses both challenges described above. To this end, we aim to investigate the possibility of breaking down existing algorithms into incremental steps making it possible to check point the mining process, display the current status and allow a user to intervene by imposing constraints that steer the algorithm in the direction of interesting patterns.

Problem definition

Sequential pattern mining addresses the problem of detecting sequences of events as patterns in data. Identification and analysis of sequential patterns are of increasing importance in a range of top priority application domains such as electronic health record analysis, process control, cyber security and safety, autonomous systems and software, and aid in the understanding and debugging of machine learning systems. There are, however, two main challenges that need to be addressed before sequential pattern mining can be fully utilized. The first challenge is based on the vast number of possible patterns.

PROPOSED SYSTEM

It has proposed a clustering algorithm to find the group relationships for query and data aggregation efficiency. The differences of and this work are as follows: First, since the clustering algorithm itself is a centralized algorithm, in this work, we further consider systematically combining multiple local clustering results into a consensus to improve the clustering quality and for use in the update-based tracking network.

Second, when a delay is tolerant in the tracking application, a new data management approach is required to offer transmission efficiency, which also motivates this study. We first introduce to distributed mining algorithm to approach the object clustering problem and discover group patterns.

ADVANTAGES OF PROPOSED SYSTEM

- The proposed system is efficiently mine much better the frequent patterns from database.
- It discovers the frequent item sets with item set generation
- It can help users understand algorithmic uncertainties, as well as trust the results of algorithms.

DATABASE DESIGN

Database design is the process of producing a detailed data model of a database. This logical data model contains all the needed logical and physical design choices and physical storage parameters needed to generate a design in a Data Definition Language, which can then be used to create a database. A fully attributed data model contains detailed attributes for each entity.

The term database design can be used to describe many different parts of the design of an overall database system. Principally, and most correctly, it can be thought of as the logical design of the base data structures used to store the data.

In the relational model these are the tables and views. In an object database the entities and relationships map directly to object classes and named relationships. However, the term database design could also be used to apply to the overall process of designing, not just the base data structures, but also the forms and queries used as part of the overall database application within the database management system (DBMS). The process of doing database design generally consists of a number of steps which will be carried out by the database designer. Not all of these steps will be necessary in all cases. Usually, the designer must:

- Determine the data to be stored in the database
- Determine the relationships between the different data elements
- Super impose a logical structure upon the data on the basis of these relationships.

Within the relational model the final step can generally be broken down into two further steps that of determining the grouping of information within the system, generally determining what are the basic objects about which information is being stored, and then determining the relationships between these groups of information, or objects. This step is not necessary with an Object database.

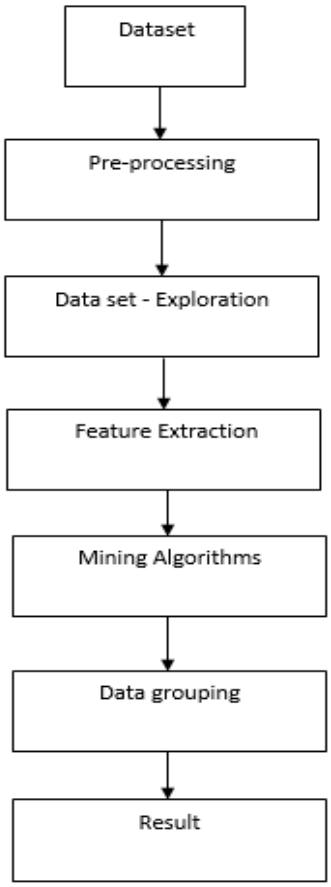
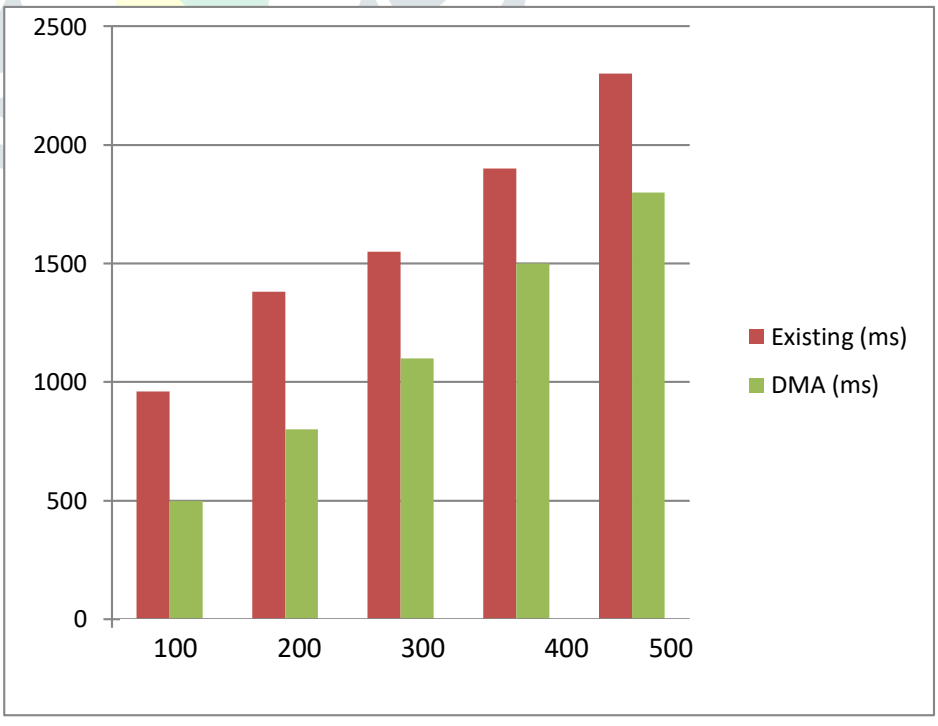


Fig 1. System design and development

Result analysis

The presented work is that the system has not been evaluated with high dimensional data. Sequences and events can be associated with surrounding information and it should be possible to incorporate this information in the mining process.Animal data are used as input in our system. Then we cluster the animal data present by using the distributed Mining algorithm.

Data	Existing (ms)	DMA (ms)
100	960	500
200	1380	800
300	1550	1100
400	1900	1500
500	2300	1800



Conclusion

In conclusion, this study explored the effectiveness of distributed Mining algorithm-based clustering process. The main contribution of the proposed work is an interactive sequence mining approach that allows a user to progressively refine constraints while pattern sequences are being built, enhancing in this way user exploration and

control over the search for interesting patterns. In this way, the user can guide the execution of the underlying mining algorithm at suitable points. The combination of the views provides additional context to the mining process by revealing how the patterns appear in the data and can therefore provide guidance to the user.

FUTURE ENHANCEMENTS

In future systems other patterns may be used for recalling purpose like touch of smells, study shows that these patterns are very useful in recalling the associated objects like images or text. It is concluded that the application works well and satisfy the end users. The application is tested very well and errors are properly debugged. The application is simultaneously accessed from more than one system

REFERENCE

- J. Han, J. Pei, and X. Yan, “Sequential pattern mining by pattern growth: principle and extensions,” Studies in Fuzziness and Soft Computing, vol. 180, pp. 183–220, 2005.
- B. C. Kwon and A. Perer, “Peekquence: Visual Analytics for Event Sequence Data,” in KDD 2016 Workshop on Interactive Data Exploration and Analytics (IDEA’16), San Francisco, CA, USA, 2016, pp. 72–75.
- C. H. Mooney and J. F. Roddick, “Sequential Pattern Mining - Approaches and Algorithms,” ACM Computing Surveys, vol. 45, no. 2, 2013.
- M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman, “Temporal event sequence simplification,” IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2227–36, 2013.
- Perer and F. Wang, “Frequence : Interactive Mining and Visualization of Temporal Frequent Event Sequences,” in Int’l Conf on Intelligent User Interfaces. Haifa, Israel: ACM, 2014, pp. 153–162.
- C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman, “LifeLines: visualizing personal histories,” in CHI ’96: Proc. of the SIGCHI conference on Human factors in computing systems. New York, NY, USA: ACM, 1996, pp. 221–227.
- P. J. Polack, S.-T. Chen, M. Kahng, K. de Barbaro, M. Sharmin, R. Basole, and D. H. Chau, “Chronodes: Interactive Multi-focus Exploration of Event Sequences,” CoRR, vol. abs/1609.0, 2016.
- K. Vrotsou, K. Elleg°ard, and M. Cooper, “Exploring Time Diaries Using Semi-Automated Activity Pattern Extraction,” electronic International Journal of Time Use Research, vol. 6, no. 1, pp. 1–25, 2009.
- T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, and S. Murphy, “Aligning Temporal Data by Sentinel Events : Discovering Patterns in Electronic Health Records,” CHI 2008 Proceedings Health and Wellness, pp. 457–466, 2008.
- T. D. Wang, C. Plaisant, B. Shneiderman, N. Spring, D. Roseman, G. Marchand, V. Mukherjee, and M. Smith, “Temporal Summaries: Supporting Temporal Categorical Searching, Aggregation and Comparison,” IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 1049–1056, 2009.