iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** IV **Month of publication:** April 2014

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Multiple Periodicity Detectionandpattern Mining In Time Series Database

Vimal Vinod

PG student,

Karunya University, Coimbatore

Abstract: ***The time series database is analysed to retrieve the relevant patterns and to examine whether it is periodic in nature. Existing algorithms failed to generate flexible user interested patterns. Another drawback was the user defined period were required to mine the patterns. This paper introduces an algorithm which can mine the user interested pattern from the database in less time. The algorithm helps to provide user interested patterns quicklywithout wasting time in making decisions.***

Keywords: Time series database, periodicity, pattern mining, episode mining, temporal regularities.

I. INTRODUCTION

Data mining is the process of extracting or mining knowledge from large amounts of data. It is a computational process. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable form so that user can use it accordingly. It helps in analysing the data from different perspectives and summarizing it into useful information. The data mining task can be automatic or semi-automatic analysis of very large databases.

Pattern mining is an important technique in data mining. It is used to retrieve a particular pattern from the database. Different types of pattern mining are frequent, sequential, periodic, inter transaction, and episode mining. Periodic pattern mining can be carried on a time series database to determine the periodicity of the patterns. Periodicity detection is a process for finding temporal regularities within the time series database. Periodicity detection or periodic pattern mining can be used for applications such as prediction, weather forecasting, unusual activity detection etc.

The existing pattern mining techniques were used to determine the periodicity and have some limitations. For instance the suffix tree based algorithm failed to retrieve the user-interested patterns. So in order to overcome the drawbacks of STNR algorithm (Rasheed et al., 2011)[2], a new algorithm (Effective periodic pattern mining)[1] (Rasheed et al., 2012) has been developed. To determine the periodicity of database.

II. RELATED WORKS

Periodic pattern mining is used in several areas. It also helps to make several critical decisions. Some techniques failed to generate user-interested patterns (Rasheed &Alshalalfa, 2011)Some special techniques for specific periodicity mining problems are in (Kolpakov& Kucherov,1999 [14]; Rasheed &Alhajj, 2010) because it requires multiple phases to retrieve the output. Periodicity detection is useful to predict the behavior and the future trends of the time series database (Weigend&Gershenfeld, 1994)[12]. Such as, periodic pattern mining is proved to be useful in predicting the computer network fault analysis, stock price movement and earth-quake prediction, and gene expression analysis (Glynn, Chen, &Mushegian, 2006) [7]. Some of the existing algorithms deal with unknown periods (Ma &Hellerstein, 2001) [4] and periodic patterns (Tanbeer et al., 2009) [9]. The input data is a set of transactions (Ozden et

al., 1998) [13], this consists of a set of items. Periodicity mining allows an energy company to analyse power consumption patterns and predict periods of high and low usage so that proper planning may take place[8].

A.Definitions and Terminologies

Definition 1(*Time series database*): A time series database is a set of observations taken at specified times, usually at 'equal intervals'[1].

Definition 2(*Periodic pattern*): An ordered list of events repeats itself in the event sequence is termed as periodic pattern.

Definition 3(*Periodic pattern mining*): Periodic pattern mining refers to the mining of patterns which are periodic within the time series database.

Definition 4 (*Symbol, partial and full cycle periodicity*): A time series database is said to have symbol periodicity if at least one symbol is repeated periodically.

Definition 5 (*Occurrence vector*): Occurrence vector is a list of the index positions at which any substring exists in the original string.

Definition 6 (*Confidence*): The confidence of a periodic pattern X, occurring in time series T, is the ratio of its actual periodicity to its expected perfect periodicity.

$$conf(p, stPos, X) = \frac{Actual\_Periodicity(p, stPos, X)}{Perfect\_Periodicity(p, stPos, X)}$$

$$Perfect\_Periodicity(p, stPos, X) = \frac{|T| - stPos + 1}{p}$$

B.*Existing Algorithm*

There several algorithms available for periodicity detection and pattern mining. These include STNR algorithm (Rasheed et al., 2011)[1], it fails to generate user interested pattern. Indyk et al., have addressed periodic trends problem. Elfeky et al., 2005b [6] developed two separate algorithms to detect segment and symbol periodicity. Elfeky et al., 2005a [5] presented an algorithm which uses time warping technique to deal with the noisy data, it can detect only segment periodicity but failed to generate symbol and sequence periodicity.

Effective periodic pattern mining in time series databases

The recent algorithm which performs periodic pattern mining in time series database (Nishi et al., 2012) [1]. There are three phases in this algorithm. In the first phase, discretization process is done. In this phase, the each event will be represented by a character. After applying discretization process the time series database can be converted into a string.

Eg: T = {accxacxbabcxbacx}

In the second phase mining process is done. Using apriori based level-by-level sequential pattern mining to get a specific pattern. Smaller length patterns are joined gradually to get a larger pattern.

In the third phase using event identifiers the patterns are joined. That is the Event Id of the items are maintained as Event Id (X_1)< Event Id (X_2) <Event Id (X_3)<Event Id (X_{i-1})<Event Id (X_i). The eventidentifier for each of the event in the pair is joined only if the event identifier of the first event in pair is lower than the event identifier of the second event otherwise the pair is neglected. This process is continued until the user defined number of pass is finished.

Algorithm

1. Perform discretization operation and construct a string.
2. For event obtained assign occurrence vector.
3. Eliminate don't care events.
4. Generate pattern.
5. Determine occurrence vector for each pattern.
6. Join occurrence vectors by combining event-Id.
7. Determine periodicity by calculating difference vector.

C.*Proposed algorithm*

In the proposed algorithm, a new technique is introduced for periodic pattern mining of web log data. This technique is a combination of two algorithms PUCC algorithm [3] and Effective periodic pattern mining algorithm. Using this technique the periodic patterns in a web log data can be mined easily.

[i] *Discretization*: Each event will be represented by a character. After applying discretization process the time series database can be converted into a string.

[ii] *Mining*: Using apriori based level-by-level sequential pattern mining to get a specific pattern. Smaller length patterns are joined gradually to get a larger pattern.

Algorithm

1. Perform discretization operation and construct a string.
2. For event obtained assign occurrence vector.
3. Eliminate don't care events.
4. Retrieve the pattern using pattern matching technique.
5. Determine periodicity by calculating difference of position vectors.

III. ARCHITECTURE OF PROPOSED SYSTEM

The architecture of proposed system is given as follows.

(i)Applying Discretization

The discretization is a mapping among the range of values of an entityto represent a specific event. As an example, consider the time series database. If we apply ''Discretization Technique'', where the discretization function can be expressed as, if any employee login in any slot withinany period/day, then a character ''a'' will represent the event and the event of being logout of that particular employee in any timeslot can be expressed by ''x''. The time slot before login can be represented by ''b'', before logout can be denoted by ''c'' and after logout can be expressed by ''d''. Hence, the time series database after applying discretization can be represented by the string, T = {accxacxdaxddbacx}.

(ii) Mining Process

Since our proposed algorithm provides a flexible pattern mining facility to the users, we have mined the string in a special fashion.Initially mined single length patterns and gradually generated larger length patterns in each pass.

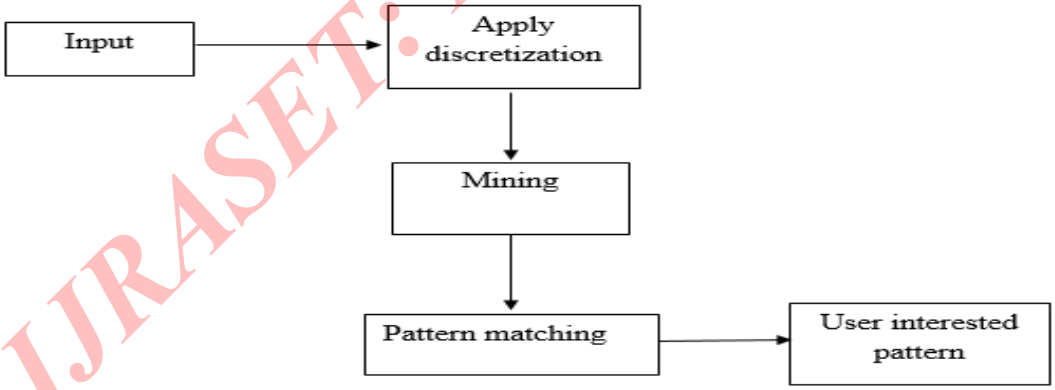


Fig 1: Architecture of proposed system

In the next step, the algorithm generates all possible exclusive interesting patterns by allowing event skipping among intermediate events by using the technique of pattern mining. The number of allowed event skipping within any two events can be determined by the difference vector. Then the

generated patterns are tested for periodicity using periodicity detection algorithm.

Experimental environment settings

The proposed algorithm is implemented in Java SDK programming language using eclipse android SDK and tested extensively in Windows 8 operating system. The experimental results are shown using a 2 GB RAM and core i3 processor.

A. Experimental Results

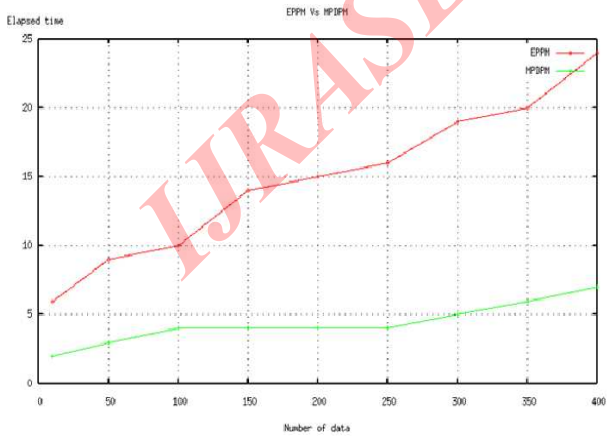
Performance analysis

The performance of the proposed algorithm depends on alphabet size, size of data and period size that means the number of unique symbols in the data, the length of the time series database which is the number of symbols in the data and the period size respectively.

The performance of this paper is compared with the paper "Effective periodic pattern mining in time series database"[2] and the resulting graph is as given below. The performance of this proposed work is improved by decreasing the time elapsed to return the pattern.

B. Graph

The graph is against elapsed time and number of data retrieved.

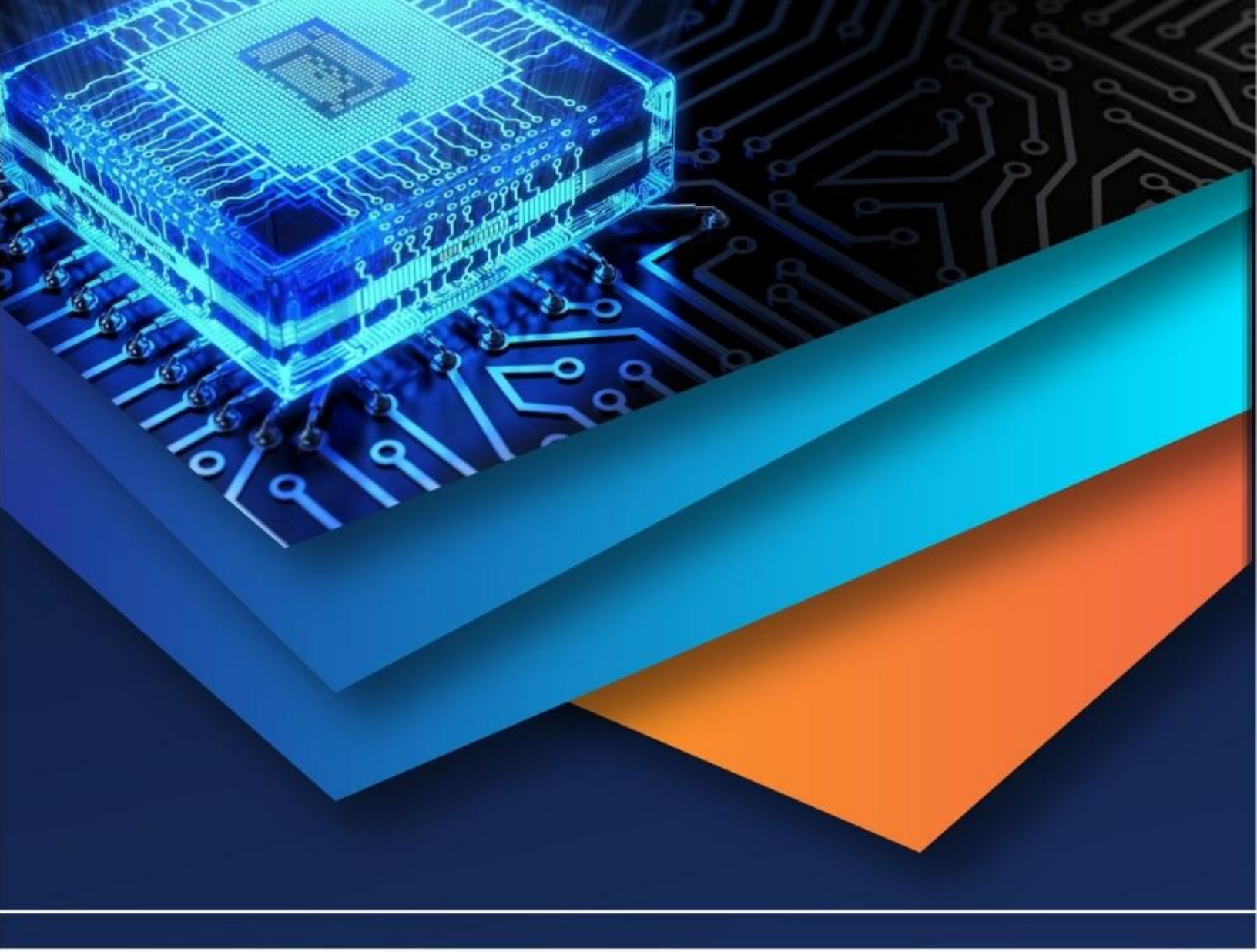


IV. CONCLUSION

This paper explains a technology that can be used to improve the web browsing technology. Using this algorithm the browsing recommendations can be improved with respect to time. It can also be developed in a mobile device for a quick reference.

REFERENCES

[1]ManzibaAkanda Nishi, Chowdhury Farhan Ahmed, Md. Samiullah, Byeong-SooJeong (2012), "Effective periodic pattern mining in time series databases".
[2] Rasheed, F., Al-Shalalfa, M., &Alhajj, R. (2011). "Efficient periodicity mining in time series databases using suffix trees".
[3] V. Sujatha&Punithavalli (2011), "Improved use navigation pattern prediction technique from web log data".
[4] Ma, S., &Hellerstein, J. (2001). Mining partially periodic event patterns with unknown periods.
[5] Elfeky, M., Aref, W., &Elmagarmid, A. (2005a). "Warp: time warping for periodicity detection".
[6] Elfeky, M., Aref, W., &Elmagarmid, A. (2005b). "Periodicity detection in time series databases".
[7] Glynn, E., Chen, J., &Mushegian, A. (2006). "Detecting periodic patterns in unevenly spaced gene expression time series using lomb–scargleperiodograms".
[8] Han, J., Dong, G., & Yin, Y. (1999)." Efficient mining of partial periodic patterns in time series database".
[9] Tanbeer, S. K., Ahmed, C. F., Jeong, B.-S., & Lee, Y.-K. (2009). "Discovering periodic frequent patterns in transactional databases".
[10] Cooley, R., Srivastava, J. and Mobasher, B, "Web mining: Information and pattern discovery on the world wide web, Tools with Artificial Intelligence" 1999.
[11] Huysmans, J., Baesens, B. and Vanthienen, J.), "Web Usage Mining: A Practical Study", Katholieke Universities Leuven, Dept. of Applied Economic Sciences (2003).
[12] Weigend, A., &Gershenfeld, N. (1994). "Time series prediction: Forecasting the future and understanding the past".
[13] Ozden, B., Ramaswamy, S., &Silberschatz, A. (1998). "Cyclic association rules. In Data Engineering".
[14] Kolpakov, R., &Kucherov, G. (1999). Finding maximal repetitions in a word in linear time.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY