A Survey on Periodicity Detection Techniques in Time Series Databases

Vimal Vinod
PG Student, Department of Computer Science
Karunya University

S. Leela
Assistant Professor,
Department of Computer Science,
Karunya University

Abstract

In recent years, periodic patterns are gaining much importance, so various periodicity detection algorithms were developed. Time series database is a collection of data gathered at certain intervals to reflect certain behaviour of an entity. By analysing time series database we can find how frequent a particular pattern is present and the number of occurrences can be counted. Temporal regularity of a pattern can be found using periodic pattern mining technique. Periodic pattern mining can be used to find periodicity of many real life problems and can be used for prediction. Various periodicity determining algorithms are compared.

Keywords: **Data mining, Periodicity detection, Periodic pattern mining**

I. Introduction

Data mining is the process of extracting knowledge or patterns from very large databases. It is a computational process which can be used to discover patterns from large datasets and can be from any area. The mining technique can convert the extracted data to understandable format. Data mining task can be either automatic or semi-automatic. This technique helps in extracting previously unknown interesting patterns, unusual behaviour and other dependencies. Several data mining techniques are available and are emerging today.

In this paper we deal various periodicity detection in time series database. Time series database is a collection of data gathered at regular intervals of time. The data will be a sequence of data values which were measured and marked at proper intervals of time. Generally time series data will have a natural temporal ordering. Different types of periodicity like symbol, sequence and segment periodicity can be determined. Periodicity detection is useful in making various predictions. Based on the type of periodicity the behaviour of the patterns can be understood. Partial periodicity can be detected from a portion of time series database [8].

This paper is organized as follows. Section II includes key concepts in this paper; section III gives related works, section IV gives conclusion and section V gives references.

2. Key concepts

The periodicity detection algorithms are compared on the following metrics and are given as follows:

1. Complexity in computation.
2. Performance in presence of noise.
3. Patterns of user's interest.

Periodicity detection is used to find whether the given pattern is periodic in nature. It searches for temporal regularity. The various algorithms used for periodicity determination in time series database are given as follows.

3. Literature Review

A. Periodcity detection algorithms

1. Efficient periodicity mining in time series databases using suffix tree

Rasheed et al., [1] proposed efficient periodicity mining in time series databases using

suffix tree. Time series database is a collection of data values stored at uniform interval of time. Periodicity of several real life time series databases like weather conditions, stock growth, power consumption etc. can be found. Periodicity detection algorithm is used for detecting temporal regularities within the time series. This algorithm works well even in the presence of noise. The objective of time series database analysis is to find the periodic patterns present in the dataset. Three different types of periodicity can be detected in this analysis: they are symbol, sequence and segment periodicity. This paper works accurately on both synthetic and real data sets. In this paper, STNR as a suffix-tree based algorithm is used to detect periodicity in time series data. STNR is a noise-resilient algorithm. The worst case time complexity is O (kn2).

2. Efficient method for de-duplication and periodicity mining in time series databases.

Drishya et al., [2] proposed efficient method for de-duplication and periodicity mining in time series databases. In this paper, an algorithm which can detect symbol, sequence and segment periodicity which works well in presence of noise is explained. Noise may be due to replacement, insertion, deletion or a combination of these. Two phases are involved in periodicity detection algorithm: (i) build the suffix tree for time series database and (ii) periodicity of time series database is calculated. To find periodicity the difference between two characters are determined. In presence of noise, a time tolerant concept is introduced. Redundant periods are handled by pruning approach.

3. WARP: Time Warping for Periodicity Detection

Elfeky et al., [3] proposed a Time Warping for Periodicity Detection (WARP). In this paper, a time warping algorithm which can detect periodicity in the presence of noise is explained. To efficiently handle different types of noise, WARP can be extended or shrinked on the time axis at various locations and this helps in removing the noise efficiently. Another important achievement is that an online version of WARP that fits the data stream model is also developed by them. A study on this algorithm shows that there is a tradeoff between noise resiliency and time performance. Advantage of WARP is that it is more noise resilient, accurate even under low memory resources. But disadvantage of WARP is it requires more processing time, than other periodicity detection algorithms.

4. Mining partially periodic event patterns with unknown periods

Sheng Ma et al., [4] proposed mining partially periodic event patterns with unknown periods proposed an algorithm to find the period using Chi-Squared test to test in the presence of noise. Two other algorithms are used to discover p-patterns which are based on periods or associations and tradeoff between these two are studied. The effectiveness of this algorithm gets reduced immediately when the noise-to-signal ratio exceeds 1. The robustness can be increased by considering n-order inter-arrivals.

5. Periodicity detection method for small-sample time series data sets.

Daisuke Tominaga [5] proposed periodicity detection method for small-sample time series datasets. This paper explains the periodic behaviour of gene expression under the influence of signal pathways. It also represents a model which incorporates noise and harmonics. Two conventional methods are used in periodicity detection. One is Quantile method, in this DFT is applied and power of harmonic is determined. Quantile points and IQR are calculated. These are then compared with outlier bound. If the power is a large value when compared with the outlier bound then we can conclude that the given time series database is a periodic one. Dixon's Q test is used for outlier detection and it ignores redundant data. The critical is found out. The computational cost of these test increases as the data length increases.

6. Periodicity detection in time series databases.

Elfeky et al., [6] proposed an efficient algorithm for detecting various types of periodicity. In this paper two types of periodicities for time series databases has been defined. The periodicity of the symbols or single characters in the time series is called as symbol periodicity. If the entire

segment is periodic, then it is called as segment periodicity. The proposed algorithm is scalable, computationally efficient algorithm for detecting each type of periodicity. The computational time complexity is O (nlogn), for a time series database of length n. When compared with symbol periodicity, segment periodicity detection takes less time for execution but symbol periodicity can detect more periods. An important fact on this algorithm is that for symbol periodicity iterations are not possible. A study on this algorithm reveals that this is more effective for partial periodic patterns.

7. Effective periodic pattern mining in time series database.

Nishi et al., [7] proposed an effective periodic pattern mining in time series database. The goal of this paper is to find the periodicity of the patterns in time series database. The user can generate patterns of user's interest by skipping intermediate or don't care events. Algorithm can detect different types of periodicity i.e. symbol, sequence and segment periodicity. The processes involved in this are: discretization, mining process and joining of two patterns. Confidence function is calculated to filter the patterns according to the user's preference. The parameters controlled are: alphabet size, data size and period size. Maximum event skipping threshold plays an important role to skip the events. The experimental results explains that this algorithm can be used for real life time datasets. This algorithm can be used to find periodicity of real life data like road transport survey, weather prediction, stock market predictions etc. by skipping unwanted or don't care events which are of less interest or of no interest. This algorithm is best to find periodicity and is user-friendly and user-friendly and user can interactively mine patterns of user's interest which works well in presence of noise. Effective periodic pattern mining in time series database.

Table 1 Comparison of performance metrics

Algorithms	High computational complexity	Performance	User interested patterns
Efficient periodicity mining in time series databases using suffix tree		√	
Efficient method for de-duplication and periodicity mining in time series databases.	√		
Time Warping for Periodicity Detection (WARP).		√	
Mining partially periodic event patterns with			

unknown periods		√	
Periodicity detection method for small-sample time series data sets	√		
Periodicity detection in time series databases	√		
Effective periodic pattern mining in time series database.		√	√

4. Conclusion

Pattern mining is an important way of mining the interested pattern. Periodicity of the mined pattern can be can be determined by using various periodicity detection algorithms. Different periodicity detection algorithms are compared and it is concluded that effective periodic pattern mining is the best among them. As this algorithm is fast and can generate patterns in which the users' are interested.

5. References

[1] Faraz Rasheed, Mohammed Alshalalfa, and Reda Alhajj, "Efficient periodicity mining in time series databases using suffix tree", 2011.

[2] S. Drishya, I. Nancy Jeba Jingle, "Efficient method for de-duplication and periodicity mining in time series databases", 2012.

[3] M.G. Elfeky, W.G. Aref, and A.K. Elmagarmid, — "WARP: Time Warping for Periodicity Detection", 2005

[4] Sheng Ma, IBM T.J. Watson, Hawthorne, Hellerstein. J. L, "Mining partially periodic event patterns with unknown periods", 2001.

[5] Daisuke Tominaga "Periodicity detection method for small-sample time series data sets", 2010.

[6] Mohamed G. Elfeky, Walid G. Aref, Ahmed K. Elmagarmid, "Periodicity detection in time series databases", 2005.

[7] Manziba Akanda Nishi, Chowdhury Farhan Ahmed, Md. Samiullah, Byeong-Soo Jeong "Effective periodic pattern mining in time series database", 2012.

[8] Chang Sheng Wynne Hsu Mong Li Lee, "Mining Dense Periodic Patterns in Time Series Data", 2006.