A SURVEY ON PERIODICITY DETECTION IN TIME SERIES DATABASE

B.Sujatha[*1], Dr.S.Chenthur Pandian[2]

[*1]Asst.Prof, Department of CSE, Sengunthar Engineering College, India

[2]Principal, Dr.Mahalingam College of Engineering and Technology, India

sujab15@yahoo.co.in

Abstract: Research on periodic pattern mining has attained a great focus on nowadays. It is the problem that regards temporal regularity. There are many emerging applications in periodic pattern mining, including weather predictions, computer networks and biological data. The discovery of patterns with periodicity is of great importance and has been rapidly developed in recent years. The problem of discovering periods for time series databases, referred as periodicity detection. These types of periodicities are available such as symbol periodicity, sequence periodicity and segment periodicity and they are identified even in the presence of noise in the time series database. Using pruning strategy some of these patterns are identified and extracted from the given time series database. There are different techniques already exists for periodic pattern mining. Those existing techniques have their own merits and demerits. This paper presents a survey on some of the existing periodic pattern mining techniques.

Keywords: periodic pattern mining, temporal regularity, symbol periodicity, sequence periodicity, segment periodicity.

INTRODUCTION

A time series database is one that supplied with data evolving over time. The several examples of time series data are meteorological data, shack market data, power consumption data and computer network data. Data mining is the process of discovering patterns and trends from the large amounts of data using techniques that uses mathematical and statistical concepts.

Research in time series data mining has concentrated on discovering different types of patterns. The periodicity mining technique requires the user to specify a periodic length that determines the rate at which the time series is periodic. These techniques assume that users either knows the length of the period previously or that that they are willing to try different periodic length values till the required periodic pattern emerge. The mining process must be executed repeatedly to get good results.

The solution to these problems is to find the method of discovering potential periods in the time series data, followed by the application of any existing pattern mining technique to extract the interesting pattern.

The problem of discovering periods for time series databases, referred as periodicity detection. These types of periodicities are available such as symbol periodicity, sequence periodicity and segment periodicity and they are identified even in the presence of noise in the time series database. There are different techniques available for periodicity detection. They have their own merits and demerits. This paper provides some discussion about some of the techniques available for periodic mining.

LITERATURE SURVEY

This Section discusses some of the periodic pattern mining techniques available today.

Efficient Periodicity mining in time series databases using suffix tree is proposed by Faraz Rasheed et al., [1].Time series database is a collection of data values stored at uniform interval of time to show the behavior of an entity. Periodicity detection is a method for detecting temporal regularities within the time series and the goal of analyzing this database is to find whether and how frequent a periodic pattern is repeated within the series. Here, the data to be analyzed are mostly noisy and there of different periodicity types. The author used STNR as a suffix-tree based algorithm for periodicity detection in time series data. This algorithm is noise-resilient and run in $O(kn^2)$ in the worst case. This method also found symbol, sequence and segment periodicity in the time series.

Jinlin Chen [2] presented an updown directed acyclic graph approach for sequential pattern mining. Sequential pattern mining is an important data mining problem that detects frequent subsequences in a sequence database. The author proposed an UDDAG for fast pattern growth. It is a new novel data structure, which supports bidirectional pattern growth from both ends of detected patterns. With UDDAG, at level i recursion, we may grow the length of patterns by 2^{i-1} at most. Thus, a length-k pattern can be detected in $[\log_2 k+1]$ levels of recursion at best and that will give result in fewer levels of recursion and faster pattern growth.

Jae-Gil Lee et al. [3] proposed a technique for mining discriminative patterns for classifying trajectories on road networks. Feature-based classification is used in the field of data mining. Using this method, features are extracted from the data points and that points are transformed into feature vector. Each vector represents the existence of features in its corresponding data point. For effective classification, we require the discovery of discriminative features. This method uses frequent pattern for classification. To know the usefulness of frequent pattern, in the classification first analyze the behavior of trajectory data on road networks. By analyzing it, what they have observed means, in addition to the location where vehicles have visited, the order of these

locations is important one for improving classification accuracy. Based on the author's analysis, he assured that frequent sequential patterns are compressed with previous method that uses only individual good feature candidates since they maintain this order information. This pattern also improves classification accuracy by 10-15%.

Avrilia Floratou et al. [4] give a technique for efficient and accurate discovery of patterns in sequence datasets. The main aim of sequential data mining applications is to discover frequently occurring patterns. The challenge behind this frequent pattern is allowing some noise in the matching process. The main thing is the definition of a pattern and the definition of similarity between two patterns. This definition of similarity can vary from one application to another. The Author presents a new algorithm called FLAME (Flexible and Accurate Motif Detector) is a flexible suffix tree based algorithm that can be used to find frequent patterns with a variety of definition of motif (pattern) models. FLAME is accurate, fast and scalable one.

David Lo et al. [5] provides mining iterative generators and representative rules for the specification of software. It is best if the software is developed with clear, precise and documented specifications. But the software products are often come with poor, incomplete and even without any documented specifications. These factors are contributed to high software maintenance cost. This is mainly due to the effort put in comprehending or understanding the software code base. So, to improve program understanding, author introduces iterative pattern mining that outputs pattern that are occurred frequently within a program trace. Frequent program behaviors that in turn represents software specifications. So, author introduces mining closed iterative patterns (ie) maximal patterns without any superpattern having the same support. These generators can be joined with the closed patterns to produce a set of rules called representative rules for forward, backward in-between temporal conditions among events in one general representation.

Obules u et al., [6] suggests a pruning strategy to remove redundant data in spatiotemporal database. The spatiotemporal data movements obey periodic patterns. (ie) the objects follow the same route over regular time intervals. Author presented the pattern matching technique to find the patterns that were repeated in the time-series database. Three kinds of patterns such as symbols, sequence and segment periodicity are also discovered. Using pruning strategy redundant data are deduced in order to reduce the memory usage and complexities.

A suffix tree based noise resilent algorithm for periodicity detection in time series database is proposed by Faraz Rasheed et al., [7]. They present a noise resilent algorithm using suffix tree as an underlying data structure. This algorithm not only calculates symbol and segment periodicity, but also detects the partial periodicity in time series. It also efficiently detects periodicity in the presence of noise compared with existing algorithm. It detects periodicity in the presence of replacement, insertion, deletion or a mixture of any of this type of noise. The authors improve their previous algorithm by incorporating

the time tolerance window so as to make it more silent to insertion and deletion noise.

David Lo et al., [8] put forth a novel method, frame work, and tool for mining inter-object scenario-based specifications in the form of a UML2-compliant variant of Damm and Harel's live sequence charts (LSC). LSC as a specification language extends the partial order semantics of sequence diagram with temporal liveness and symbolic class level lifeliness to generate compact specifications. The output of this algorithm is satisfying the given thresholds of support and confidence, mined from an input program execution race. The author uses search pruning strategy, specifically adapted to LSCs, which provides efficient mining of scenarios of arbitrary size.

Live sequence charts (LSC), a visual model, scenario-based, inter-object language is proposed by David Lo et al., [9] to investigate the problem of mining scenario-based triggers and effects from program execution tracers. The author uses data mining methods to provide significant and complete results of modulo user-defined thresholds. The input trigger and effect scenarios and the resulting candidate modal scenarios are represented and visualized using a UML2-complaint variant of LSC.

CONCLUSION

Periodicity detection is a process of finding temporal regularities within the time-series and the goal of analyzing a time series database is to find how frequent a periodic pattern is repeated within time intervals. Three types of periodic patterns can be detected in a time series database using different techniques. Using pruning strategy, we can reduce the redundant data there by reducing memory space and complexities. These sequences can be used anywhere in any field according to the needs of the user, especially in earthquake prediction, weather forecasting, power consumption and fraud detection applications. It is also applicable in the areas of biological and DNA sequences.

REFERRENCES

[1]. Faraz Rasheed et. Al., "Efficient Periodicity Mining in Time Series Databases Using Suffix Trees", IEEE Transactions On Knowledge And Data Engineering, Vol. 23, No. 1, January 2011

[2]. Jinlin Chen et. Al., "An UpDown Directed Acyclic Graph Approach for Sequential Pattern Mining", IEEE Transactions On Knowledge And Data Engineering, Vol. 22, No. 7, July 2010

[3]. Jae-Gil Lee et. Al., "Mining Discriminative Patterns for Classifying Trajectories on Road Networks", IEEE Transactions On Knowledge And Data Engineering, Vol. 23, No. 5, May 2011

[4]. Avrilia Floratou et. al., "Efficient and Accurate Discovery of Patterns in Sequence Datasets", IEEE Transactions on Knowledge and Data Engineering, 2011

[5]. David Lo et. Al., "Mining Iterative Generators and Representative Rules for Software Specification Discovery", IEEE Transactions On Knowledge And Data Engineering, Vol. 23, No. 2, Febrauary 2011

[6]. O.Obulesu et. Al., " Finding Maximal Periodic Patterns and Pruning Strategy in Spatiotemporal Databases", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 4, April 2012 ISSN: 2277 128X

[7]. F. Rasheed and R. Alhajj, "STNR: A Suffix Tree Based Noise Resilient Algorithm for Periodicity Detection in Time Series Databases," Applied Intelligence, vol. 32, no. 3, pp. 267-278, 2010.

[8]. D. Lo, S. Maoz, and S.-C. Khoo, "Mining Modal Scenario-Based Specifications from Execution Traces of Reactive Systems," Proc. ACM/IEEE Int'l Conf. Automated Software Eng., 2007.

[9]. D. Lo and S. Maoz, "Mining Scenario-Based Triggers and Effects," Proc. ACM/IEEE Int'l Conf. Automated Software Eng., 2008.

Short Bio Data Authors Profile

B.Sujatha received her B.E. degree in Computer Science& Engineering from the Madurai Kamarajar University, Tamil Nadu, and M.E. degree in Computer Science & Engineering from VMRFDU, Tamil Nadu. She is currently pursuing PhD in Anna University, Chennai under the supervision of Dr.S.Chenthur Pandian. Her research interests include time series data mining, periodicity detection in time series databases.

Dr. S. Chenthur Pandian Completed his Ph.D in Periyar University, Salem. He had served in Indian Air Force for 15 years from 1979 to 1994. He had worked in KSR College of Technology, Tiruchengode as Professor and Head of Electrical & Electronics Engineering, Vice-Principal and Director till 2006. From 2006 to 1st June, 2010, he had served in Selvam College of Technology, Namakkal as Founder Principal. He has assumed the office as Principal in Dr. Mahalingam College of Engineering and Technology with effect from 2nd June, 2010 FN onwards. He has totally 31 years of rich experience in teaching, industry, research and development.