An Analysis of Algorithms for Periodic Pattern Mining in Time Series

Yeping Peng

Software College of Jishou University, Zhangjiajie, Hunan Province, China

Keywords: Periodic pattern mining; Periodic length; Mining Algorithm

Abstract: Periodic pattern mining in time series is of great practical significance to scientific research and practical application as it is beneficial to improve prediction and trend analysis. Current algorithms for periodic pattern mining in time series can be divided into three categories according to whether the parameters of periodic length are known or not. This paper discusses the specific classification of periodic patterns and summarizes advantages and disadvantages of these algorithms, with a purpose to provide reference for the practical application of these algorithms.

1. Introduction

Researchers have been carrying out many studies of periodic pattern mining in time series and have put forwards different periodic pattern mining algorithms in terms of data features and application situations. Current algorithms for periodic pattern mining in time series can be divided into three categories according to whether the parameters of periodic length are known or not, i.e., periodic pattern mining with known periodic length, with given periodic length threshold and with unknown periodic length. This paper will provide an analysis of the specific classification of the periodic patterns and summarize advantages and disadvantages of the mining algorithms, with a purpose to provide reference for the practical application of these algorithms.

2. Classification of Periodic Pattern Mining in Time Series

2.1 Periodic Pattern Mining with Known Periodic Length

Periodic length is an important parameter in periodic pattern mining. Most of the algorithms are based on the premise that the parameters of periodic length are known. The classification of different periodic pattern mining algorithms is as follows:

2.2 Full Periodic Pattern Mining and Partial Periodic Pattern Mining

Full periodic pattern mining was first proposed by Ozden in 1998, of which the input data is a series of transaction data sets. Each transaction record is randomly combined by a series of item sets. In addition, each transaction record is marked the occurrence time. The goal of the algorithm is to mine association rules with repeatability. Full periodic pattern means that a sequence is divided into equal-length subsequences at equal intervals, and each subsequence is strictly the same periodic pattern. [1]

In fact, Partial Periodic Patterns exist more widely in practical application. Partial periodic pattern was first proposed by Jiawei Han in 1999. Most of the algorithms for partial periodic pattern mining are based on Apriori property. [2] Han et al. further proposed several new algorithms to mine partial periodic patterns by exploring the unique properties of Partial Periodic Patterns, including Apriori heuristic properties and Max-subpattern hit set attributes. The attributes of the largest sub-pattern hit set include the subset of time series patterns in the maximum hit sub-pattern tree, so mining part of periodic patterns only needs to scan the time series database twice to separate all frequent patterns in time series. At the same time, the algorithm can also mine partial periodic patterns of multi-period time series. The incremental periodic pattern mining algorithm proposed by Rasheed et al. greatly improves the online analytical processing ability of periodic patterns while posing higher requirements for the mining of periodic patterns.[3]

  DOI: 10.25236/isrme.2019.067

Full periodic pattern emphasizes that every element in the sequence strictly participates in the formation of periodic pattern while not every element in partial periodic pattern contributes to periodicity. In fact, partial periodic pattern is a looser than full periodic pattern and has received wider popularity in practical application. Therefore, the study of partial periodic pattern is more meaningful and it is also one the current hotspots within the research field.

2.3 Synchronized Periodic Pattern Mining and Asynchronous Periodic Pattern Mining

In a Synchronous Periodic Pattern, a periodic subsequence is strictly aligned with the original sequence. For example, in the character sequence{a,f,c,a,d, c, a, g, c, d, c, a},the subsequences{a, *, c},{a,*,*},{*,*, c}are all synchronous periodic patterns of the original sequence. If the original sequence mismatches the original sequence due to noise interference or element loss, then a new periodic pattern called Asynchronous Periodic Pattern will be generated. In the above-mentioned character sequence, the subsequences {a,*},{c,*},{c, a} are all asynchronous periodic patterns of the original sequence.

In practical application, if synchronous periodic pattern is blindly pursued, some rules will be missed in one way or another. As long as the periodic pattern exists objectively, the dislocation range is within the maximum deviation, and the duration of interference is within the detecTable range. Hence, the periodic pattern of the event concerned can still be discovered. Latent Semantic Analysis, as proposed in view of the above situation, belongs to asynchronous periodic pattern mining algorithm, which defines two parameters: (1)min_rep-the minimum repetition number of periodic patterns; and (2)max_dis-the maximum allowable interval between two consecutive and effective segments. Based on the above two parameters, a two-stage mining algorithm is thus proposed. Firstly, candidate periodic patterns are generated by pruning the parameters of “maximum allowable interval distance”. Secondly, the validity of candidate periodic patterns is verified iteratively, and the longest periodic subsequences are generated.

2.4 Surprising Periodic Pattern Mining

As is expected, SMCA algorithm is further proposed on the basis of LSI algorithm. Its improvement is that besides the two parameters of LSI algorithm, SMCA algorithm adds a new parameter-the minimum number of repetitions of valid sequences, namelyglobal_rep, and SMCA algorithm allows multiple events to occur at the same time point, which enables the processing of more complex data. There is still asynchronous periodic mining algorithm that has been newly worked out to focus both the frequency of patterns and the periodic mining of elements in the same pattern by setting multiple minimum itemset support thresholds.[4]

Obviously, previous periodic pattern mining algorithms focus on frequent periodic pattern mining. Scholars put forwards another algorithm InfoMiner that can be applied to the mining of rare but statistically important periodic patterns, i.e., Surprising Periodic Patterns.[5]Such patterns embody a new information gain pattern to replace the previous support pattern based on the concepts of Information and Information Gainas specified in information theories. They focus more on the mining of periodic sequential patterns when the probability of occurrence of events is very small. In this algorithm, the Information Gain index is used to measure the Surprise of each pattern in the sequence, and the periodic pattern whose Surprise is greater than the user’s predefined threshold is selected. Because the location information of periodic patterns is not taken into account in the above algorithm, an improved InfoMiner plus algorithm is proposed by scholars, which measures the Surprise of each pattern with generalized Information Gain index, so as to find out all part of periodic subsequences with statistical significance.[6]

2.5 Periodic Pattern Mining with Given Periodic Length Threshold

In fact, it is rare to predict the period length in advance in the periodic pattern mining algorithm, but users may provide some useful information about the period length. For example, when the rainy season comes, there will be a heavy rain every three to five days, which indicates that the period length threshold is 3 to 5 days. This is also of great significance to the mining of periodic patterns.

Approximate periodic pattern mining not only allows the occurrence of events in sequential patterns to be approximately periodic, but also allows the structure of sequential patterns to be approximately periodic. The famous recent periodic pattern mining algorithm was proposed by Dehne et al. [7]. The algorithm defines a parameter$\alpha \in [0,1]$, making the maximum periodic ratio, i.e., ration of maximum to minimum absolute difference between adjacent time intervals of a single event-period ratio is 1+a.The disadvantage of the algorithm is that the time complexity of the algorithm is O (n3). When the amount of input data is too large, the time complexity of the algorithm will too high. Therefore, by defining a new parameter ε, a subsequence whose period ratio is at most (1+α) (1+ε) can be found to reduce the time complexity to O (n1+γ), whereinγ is an arbitrarily small normal number. The approximate periodic pattern mining algorithm proposed by Amir et al. allows the occurrence of events in sequential patterns to be approximately periodic. The algorithm pre-sets a parameter $\varepsilon \in [0,1]$ to find the longest periodic pattern whose absolute difference between the two adjacent occurrences of the same event is between P and p(1+ε). Compared with previous approximate periodic pattern mining algorithms, the time complexity of the algorithm is sub-cubic and in most cases O (n2), which greatly improves the time complexity of the algorithm.[8]

2.6 Periodic Pattern Mining with Unknown Periodic Length

All the above algorithms require users to input the period length parameters subjectively, that is, users need to know the threshold range of the period length or period length in advance, which makes these algorithms more suiTable for natural period length data, such as the data with the period length of hours, days, weeks, months and years. However, most data sets may not repeat with natural period length, so a more applicable algorithm is needed to extract the implicit period length and period pattern.

In view of this, an algorithm is proposed for mining periodic patterns with unknown period length. The mining of periodic pattern association rules can be divided into two sub-tasks: discovery of periodic length of sequential patterns and mining of sequential association rules. Two algorithms are proposed in this paper. The first one is to discover the periodic length of sequential patterns, and the second one is to mine the association rules of periodic patterns. Among them, the former uses chi-square test to find the periodic length. The latter first mines association rules between events, and then discovers the cycle length of sequential patterns. Experiments show that the algorithm of inter-event association rule priority mining is less affected by noise, but the algorithm of cycle length priority mining is more efficient. [9] Besides, by defining the tolerance rol of the occurrence time point of periodic events based on ant colony algorithm, another periodic pattern mining algorithm with weak constraints can be more practical than the strict periodic pattern mining algorithm.[10]

In summary, the algorithms for mining periodic patterns in time series are compared in terms of problems solved, number of parameters, and advantages, as shown in Table 1.

Table 1 Comparison between Different Algorithms

Typical Algorithm	Problems Solved	Number of Parameters	Advantages
Max-Subpattern Hit Set Method	Partial periodic pattern mining	2	It only needs to scan the database twice.
LSI (Latent Semantic Analysis)	Asynchronous periodic pattern mining	2	Time complexity is in direct proportion to the length of the input sequence.
InfoMiner	Surprising periodic pattern mining	2	It raises new information gain measurement.
Longest Approximate Periodic Pattern	Approximate periodic pattern mining	2	It enables the longest approximate periodic pattern to be found with the minimal time complexity while solving 3SUM problem.
Finding Unknown Periods	Unknown periodic length mining	3	It enables the unknown periodic length to be found through Chi-square test while being insensitive to noise.

3. Summary

Periodic pattern mining in time series is of great practical significance to scientific research and practical application, and it has contributed much to prediction and trend analysis. This paper has reviewed the specific classification of three popular periodic patterns and summarized their advantages and disadvantages. It is aimed at providing reference for the practical application of periodic pattern mining algorithms in time series.

Acknowledgement

This research was financially supported by the General Project of Hunan Provincial Department of Education (No. 16C1312).

References

[1] B.Ozden, S. Ramaswamy, A. Silberschatz. Cyclic association rules, in proceedings of 14th International Conference on Data Engineering (1998) 322-326.

[2] J. W. Han, W. Gong and Y. Yin. Efficient mining partial periodic patterns in times series database, in proceedings of International Conference on Data Engineering (1999) 106-115.

[3] K. G. Cheng, X. L. Dong. Efficient mining of local frequent periodic patterns in time series Database, IEEE 2009 International Conference on Machine Learning and Cybernetics (2009)183 - 186

[4] K. Y. Huang, C. H. Chang. SMCA: a general model for mining asynchronous periodic patterns in temporal databases, IEEE Transactions on Knowledge & Data Engineering, 17, 6 (2005)774-785.

[5] J. Yang, W. Wei, P. S. Yu. Infominer: mining surprising periodic patterns. Proc AcmSigkdd International Conference on Knowledge Discovery & Data Mining (2001) 395-400.

[6] M. A. Nishi, C. F. Ahmed, M. Samiullah, et al.. Effective periodic pattern mining in time series databases, Expert Systems with Applications, 40, 8 (2013)3015-3027.

[7] F. Dehne, A. Ferreira, E. Cáceres, S. W. Song, A. Roncato, Efficient parallel graph algorithms for coarse-grained multicomputers and BSP, Algorithmica, 33, 2 (2002)183-200.

[8] K. J. Yang, T. P. Hong, Y. M. Chen, et al.. Projection-based partial periodic pattern mining for event sequences, Expert Systems with Applications, 40,10 (2013)4232-4240.

[9] S. Ma, J. L. Hellerstein. Mining partially periodic event patterns with unknown periods, in proceedings of International Conference on Data Engineering,(2000) 205-214.

[10] R. S. Parpinelli, H. S. Lopes, A.A. Freitas, Data mining with an ant colony optimization algorithm, Evolutionary Computation IEEE Transactions on Evolutionary Computation, 6, 4 (2002)321-332.