Pattern Detection in Time-Series Sequences by using pattern mining

[1]MEKALA HARISH, [2]M.SANTHI SUDHA, [3]D.Krishna
M.Tech(CSE) Pursuing, [2]Assistant Professor,[3]Associate Professor
[3]HOD of CSE Dept
kingharish123@gmail.com
Jawaharlal Nehru Institute of Technology

ABSTRACT: ***Periodic pattern detection in time-ordered sequences is an important data mining task, which discovers in the time series all patterns that exhibit temporal regularities. Periodic pattern mining has a large number of applications in real life; it helps understanding the regular trend of the data along time, and enables the forecast and prediction of future events. An interesting related and vital problem that has not received enough attention is to discover outlier periodic patterns in a time series. Outlier patterns are defined as those which are different from the rest of the patterns; outliers are not noise. While noise does not belong to the data and it is mostly eliminated by pre-processing, outliers are actual instances in the data but have exceptional characteristics compared with the majority of the other instances. Outliers are unusual patterns that rarely occur, and, thus, have lesser support (frequency of appearance) in the data. Outlier patterns may hint toward discrepancy in the data such as fraudulent transactions, network intrusion, change in customer behaviour, recession in the economy, epidemic and disease biomarkers, severe weather conditions like tornados, etc. We argue that detecting the periodicity of outlier patterns might be more important in many sequences than the periodicity of regular, more frequent patterns. In this paper, we present a robust and time efficient suffix tree-based algorithm capable of detecting the periodicity of outlier patterns in a time series by giving more significance to less frequent yet periodic patterns. Several experiments have been conducted using both real and synthetic data; all aspects of the proposed approach are compared with the existing algorithm Info Miner; the reported results demonstrate the effectiveness and applicability of the proposed approach.***

KEYWORDS: series, pattern, prediction, noise.

INTRODUCTION:

ATIME series records options captured often at uniform interval of your time. Reality has many samples of time series like weather record, stock worth movement, road or network traffic density pattern, sensory information, transactions record, etc. Time-series info records such time-ordered options, and data processing aims to get within the info hidden information, that can't be found by classical info question languages like SQL. Usually the data is within the kind of patterns, which may lead toward some type of rules. For example, association rules mining discovers patterns during which varied data things seem along so try and formulate some rules based on the discovered association. Periodicity detection in time-series databases may be a data mining problem wherever sporadically continuance patterns area unit discovered. Periodic patterns area unit found in weather information, transactions history, stock worth movement, road and network traffic density, organic phenomenon, etc. In time-series analysis, periodicity detection identifies the periodic functions to capture seasonality, e.g., victimization Fourier transforms [3]. Periodic pattern mining is a very important task, because it facilitates information analysis leading to prediction or forecast of future events and patterns. Periodic pattern mining algorithms sometimes provide additional significance to patterns that seem additional oft or have higher support within the analyzed sequence .For instance, a large number of power consumption patterns of a housing community are perennial weekly. A stimulating extension to the current problem is discovering the cyclicity of outlier or shocking patterns. For instance, variety of master card transactions completed over time and involving little amounts might hint toward fraudulent activity. Further, by considering the recent economy crises, it ought to be doable to appreciate that similar economic things have been determined within the past shall the historical knowledge be obtainable and well analyzed. it might be attention-grabbing to mine out whether or not the flip down in economy includes a periodic pattern. Discovering however sporadically continuance alternative options (e.g., oil price) square measure aligned with the flip down in economy would possibly facilitate in identifying the link between totally different options. Although the problem has been handled earlier in time-series literature, it is clear that decline within the economy is associate degree outlier, unusual, or shocking phenomenon, that doesn't repeat often; additionally, the repetitions don't seem to be strictly periodic (say specifically once 5 years), and therefore the amount price could be considerably larger than the regular frequent patterns (compare 5 years with regular weekly or monthly periods). Discovering the latter patterns may be a challenge as a result of cyclicity detection algorithms sometimes report large number of patterns and manual discovery of bizarre patterns from such sizable amount of frequent and regular patterns is very troublesome, time intense, and error prone.

LITERATURE SURVEY:

Efficiently finding the most unusual time series sub-sequence

In this work, we introduce the new problem of finding time series discords. Time series discords are subsequences of a longer time series that are maximally different to all the rest of the time series subsequences. They thus capture the sense of the most unusual subsequence within a time series. Time series discords have many uses for data mining, including improving the quality of clustering, data cleaning, summarization, and anomaly detection. As we will show, discords are particularly attractive as anomaly detectors because they only require one intuitive parameter (the length of the subsequence) unlike most anomaly detection algorithms that typically require many parameters. We evaluate our work with a comprehensive set of experiments. In particular, we demonstrate the utility of discords with objective experiments on domains as diverse as Space Shuttle telemetry monitoring, medicine, surveillance, and industry, and we demonstrate the effectiveness of our discord discovery algorithm with more than one million experiments, on 82 different datasets from diverse domains.

A practical visualization tool for working with large time series databases

The increasing interest in time series data mining in the last decade has resulted in the introduction of a variety of similarity measures, representations and algorithms. Surprisingly, this massive research effort has had little impact on real world applications. Real world practitioners who work with time series on a daily basis rarely take advantage of the wealth of tools that the data mining community has made available. In this work we attempt to address this problem by introducing a simple parameter-light tool that allows users to efficiently navigate through large collections of time series. Our system has the unique advantage that it can be embedded directly into the any standard graphical user interface, such as Microsoft Windows, thus making deployment easier. Our approach extracts features from a time series of arbitrary length, and uses information about the relative frequency of its features to color a bitmap in a principled way. By visualizing the similarities and differences within a collection of bitmaps, a user can quickly discover clusters, anomalies, and other regularities within their data collection. We demonstrate the utility of our approach with a set of comprehensive experiments on real datasets from a variety of domains.

Periodicity detection in time series database

Periodicity mining is used for predicting trends in time series data. Discovering the rate at which the time series is periodic has always been an obstacle for fully automated periodicity mining. Existing periodicity mining algorithms assume that the periodicity rate (or simply the period) is user-specified. This assumption is a considerable limitation, especially in time series data where the period is not known a priori. In this paper, we address the problem of detecting the periodicity rate of a time series database. Two types of periodicities are defined, and a scalable, computationally efficient algorithm is proposed for each type. The algorithms perform in time for a time series of length n. Moreover, the proposed algorithms are extended in order to discover the periodic patterns of unknown periods at the same time without affecting the time complexity. Experimental results show that the proposed algorithms are highly accurate with respect to the discovered periodicity rates and periodic patterns. Real-data experiments demonstrate the practicality of the discovered periodic patterns.

Mining partially periodic event patterns with unknown periods

Periodic behavior is common in real-world applications. However in many cases, periodicities are partial in that they are present only intermittently. The authors study such intermittent patterns, which they refer to as p-patterns. The formulation of p-patterns takes into account imprecise time information (e.g., due to unsynchronized clocks in distributed environments), noisy data (e.g., due to extraneous events), and shifts in phase and/or periods. We structure mining for p-patterns as two sub-tasks: (1) finding the periods of p-patterns and (2) mining temporal associations. For (2), a level-wise algorithm is used. For (1), we develop a novel approach based on a chi-squared test, and study its performance in the presence of noise. Further we develop two algorithms for mining p-patterns based on the order in which the aforementioned sub-tasks are performed: the period-first algorithm and the association-first algorithm. Our results show that the association-first algorithm has a higher tolerance to noise; the period-first algorithm is more computationally efficient and provides flexibility as to the specification of support levels. In addition, we apply the period-first algorithm to mining data collected from two production computer networks, a process that led to several actionable insights.

PROPOSED SYSTEM

The proposed algorithm consistently outperforms the existing approach Info Miner. Additionally, our outlier detection algorithm, being an extension of the STNR periodicity detection framework can achieve things like detection of periodic patterns in subsections, drifted or shifted periodic occurrence detection using time tolerance window, and works with noisy series containing any of insertion, deletion, and replacement noise. We have also shown that our STNR is not only time efficient but also space efficient .Finally; we are currently working on the following aspect. The definition of surprising patterns can be further improved; some possibilities include the exclusion of user-specified minimum surprise value and integration of standard deviation in the definition of candidate outlier patterns.

ADVANTAGES

- The proposed algorithm is not only time efficient but also space efficient.
- Higher performances compare to existing approaches.
- The algorithm, being simple in nature, is flexible enough to work with DNA and protein sequences.

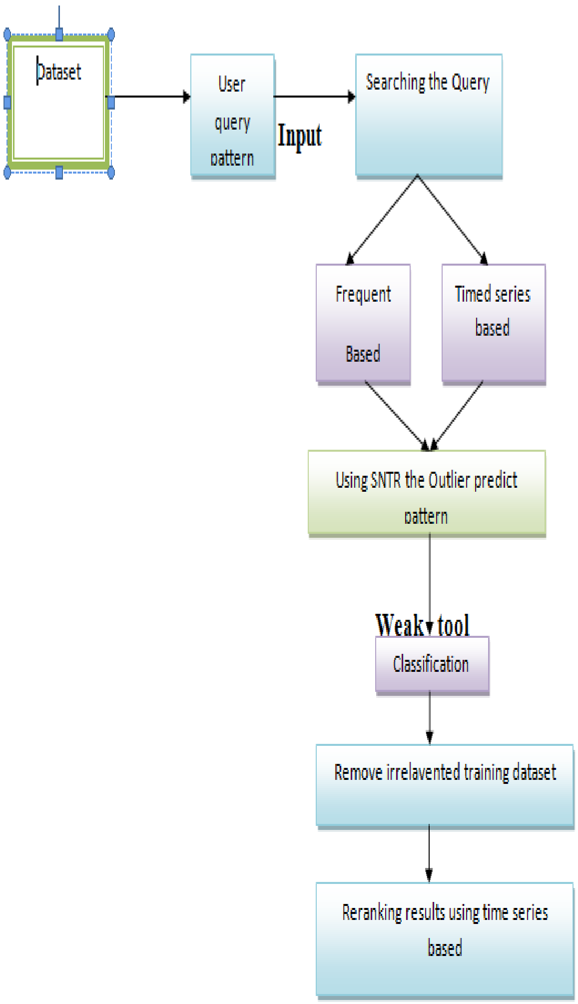

Fig:1 Architecture Diagram

CONCLUSION:

In this paper, we've got conferred a unique algorithmic program for the periodicity detection of outlier, surprising, or uncommon patterns. Our notion of a shocking or uncommon pattern takes into consideration the frequency of a pattern with patterns of comparable length. The algorithmic program conjointly takes into consideration the coverage space of the pattern and therefore the chance of pattern incidence to classify it as associate outlier pattern. This definition isn't restricted to the assumption that patterns involving less frequent events are uncommon patterns. The training/testing section, as delineated in [7]. With our definition, we can conjointly establish outlier patterns that will involve some (or all) frequent events, as we have a tendency to check the repetitions of combination of events and not simply the individual events. The experimental

results show that the projected algorithmic program systematically outperforms the prevailing approach Info Miner. We have conjointly shown that our STNR is not solely time economical however conjointly area economical. Finally, we are presently performing on the subsequent side. The definition of surprising patterns is additional improved; some potentialities include the exclusion of user-specified minimum surprise price and integration of normal deviation within the definition of candidate outlier patterns.

REFERENCES

[1] E. Keogh, J. Lin, and A. Fu, "HOT SAX: Efficiently finding the most unusual time series sub-sequence," in *Proc. IEEE Int. Conf. Data Mining*, Houston, TX, USA, Nov. 2005, pp. 226–233.

[2] N. Kumar, N. Lolla, E. Keogh, S. Lonardi, C. A. Ratanamahatana, and L. Wei, "Time-series bitmaps: A practical visualization tool for working with large time series databases," in *Proc. SIAM Int. Conf. Data Mining*,

Newport Beach, CA, USA, 2005, pp. 531–535.

[3] M. G. Elfeky, W. G. Aref, and A. K. Elmagarmid, "Periodicity detection in time

series databases," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 7, pp. 875–887, Jul. 2005.

[4] M. G. Elfeky,W. G. Aref, and A. K. Elmagarmid, "WARP: Time warping for periodicity detection," in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2005, pp. 8–15.

[5] J. Han, W. Gong, and Y. Yin, "Mining segment-wise periodic patterns in time related databases," in *Proc. ACM Int. Conf. Knowl. Discov. Data Mining*, vol. 8, no. 1, pp. 53–87, Aug. 1998.

[6] P. Indyk, N. Koudas, and S. Muthukrishnan, "Identifying representative trends in massive time series data sets using sketches," in *Proc. Int. Conf. Very Large Data*, Sep. 2000, pp. 363–372.

[7] S. Ma and J. Hellerstein, "Mining partially periodic event patterns with unknown periods," in *Proc. IEEE Int. Conf.Data Eng.*,Apr. 2001, pp. 205– 214.

[8] J.Yang,W.Wang, and P.Yu, "InfoMiner+: Mining partial periodic patterns with gap penalties," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2002, pp. 725–728.

[9] C. Berberidis, W. Aref, M. Atallah, I. Vlahavas, and A Elmagarmid, "Multiple and partial periodicity mining in time series databases," in *Proc. Eur. Conf. Artif. Intell.*, Jul. 2002, pp. 79–94.

[10] C. Sheng,W. Hsu, andM.-L. Lee, "Mining dense periodic patterns in time series data," in *Proc. IEEE Int. Conf. Data Eng.*, 2005, p. 115.

[11] C. Sheng, W. Hsu, and M.-L. Lee, "Efficient mining of dense periodic patterns in time series," Nat. Univ. Singapore, Singapore, Tech. Rep. 1, 2005.

AUTHOR 1

MEKALA HARISH
M.Tech(CSE) Pursuing
kingharish123@gmail.com
Jawaharlal Nehru Institute of Technology

AUTHOR 2

Santhi Sudha Mathi, B.Tech (CSE), M.Tech (CSE) is having 7+ years of relevant work experience in Academics, Teaching. At present, she is working as an Assistant Professor, Jawaharlal Nehru Institute of Technology, Ibrahimpatnam, Hyderabad, A.P, India. Her areas of interest are Computer Networks, Database management system, Network security.

AUTHOR 3

D.Krishna, B.Tech (CSE) M.Tech (CSE) is having 12+ years of relevant work experience in Academics, Teaching, and Lifetime Member of

ISTE. At present, he is working as an Associate Professor, HOD of CSE Dept, Jawaharlal Nehru Institute of Technology, Ibrahimpatnam, Hyderabad, Telangana State, India and utilizing his teaching skills, knowledge, experience and talent to achieve the goals and objectives of the Engineering College in the fullest perspective. He has attended seminars and workshops. He has published more than fifteen research papers in International journals. He has also guided ten postgraduate students. His areas of interest Data Mining, Data Warehousing, Cloud computing, Network security, Automata theory & Compiler Design.