Volume 4, Issue 10, October 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey of Efficient Periodicity Mining Techniques for Time-Series Data

Mr. Manish S. Sapkal
Computer Engineering & University of Pune
Pune, India

Mrs. V. S. Nandedkar
Information Technology & University of Pune
Pune, India

Abstract— Research on periodic pattern mining has attained a great focus on nowadays. Periodicity mining or discovering the periodic patterns from the time series, has always been a problem for fully automated periodicity mining. A time series is a collection of data values gathered generally at uniform interval of time to reflect certain behaviour of an entity. Real life has several examples of time series such as weather conditions of a particular location, spending patterns, stock growth, transactions in a superstore, network delays, power consumption, computer network fault analysis and security breach detection, earthquake prediction, gene expression data analysis, etc. A time series is mostly characterized by being composed of repeating cycles. This paper deals with the rigorous survey for implementation of techniques for Efficient Periodicity Mining for Time-Series Data. Here we compare different type of algorithms on the basis of technique they have employed.

Keywords—Periodicity mining, Time Series Data

I. INTRODUCTION

A time series is a collection of data values gathered generally at uniform intervals of time to reflect certain behavior of an attribute or entity. Examples of time series data are meteorological data containing several measurements, e.g., temperature and humidity; stock prices depicted in financial market; power consumption data reported in energy companies; and event logs monitored in computer networks. Periodicity analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of data. Periodic pattern mining or periodicity detection has a number of applications such as prediction, forecasting and detection of unusual activities. Discovering the periodicity rate of time series data has drawn the attention of the data mining research community.

Data mining (DM), also called Knowledge-Discovery and Data Mining, is the process of automatically searching large volumes of data for patterns using association rules. It is a fairly recent topic in computer science but utilizes many older computational techniques from statistics, information retrieval, machine learning and pattern recognition. The Periodicity are of three types 1) Symbol periodicity 2) Partial periodicity 3) Segment periodicity[1].

Symbol periodicity means only one symbol is periodic. Ex T= *acdabcadb*.Here only a is periodic with p = 5. If more than one symbol is periodic and occur partially it is called partial periodicity .Eg: T = *bbaaabbdabcaabbcabcd* . Here sequence ab is periodic with periodicity p=4 and periodicity starts from the position 4 .Segment periodicity means the whole time series is represented as a periodic pattern. In time series periodicity can be full or partial[4]. Also several noise can be present in the time series. Methods used for finding full periodicity cannot be used for detecting partial periodicity .Partial periodicity is a looser kind of periodicity than full periodicity and it exist ubiquitously in the real world[4] An example partial periodic pattern may state that Jim reads the Vancouver Sun newspaper from 7:00 to 7:30 every weekday morning but his activities at other times do not have much regularity. So partial periodic detection is a expensive mining process because it is the mixture of periodic and non periodic events. Another problem occur in periodic detection is the presence of noise .Most of the algorithm have poor resilience to noise.

Another problem in periodicity detection is perfect periodicity[1]. All the periodicity in time series database is not perfect. The degree of perfection of time series can be represented in terms of confidence .Confidence is defined as the ratio of actual frequency in the series over its expected perfect frequency in the time series.

Consider an imperfect periodicity in the example T = abefd abcde acbfe abedf, here the confidence is 4/5. The periodicity mining algorithm requires user to specify a periodic length that determines the rate at which the time series is periodic. This cannot be done in trial and error method. The solution of this problem is to devise a technique for discovering the potential periods in the time series data followed by the application of any existing pattern mining technique to determine the interesting pattern.[4] To sum up, time series exist frequently in our daily life and their analysis could lead to valuable discoveries. So there is need for noise resilience algorithm that can tackle the problem of i) Identifying three different type of periodic pattern ii)handling asynchronous periodicity. Here a comparative study of four algorithm are done such as Time wrapping for Periodicity Detection usually called WARP, Periodicity Detection Algorithm using Convolution usually called CONV ,Partial Periodicity Detection usually called Parper and Periodicity Detection using suffix tree usually called STNR.

II. LITERATURE SURVEY

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 (kn2) 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 2i-1 at most. Thus, a length-k pattern can be detected in [log2 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 super pattern 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.

There is a vast amount of literature on mining databases for frequent pattern [6], [27], [34]. The problem of mining for subsequence was introduced in [1]. Subsequence mining has several applications, and many algorithms like [33], [36], and [38] have been proposed to find patterns in the presence of noise. However, they primarily focus on subsequence mining, while we focus on contiguous patterns. A host of techniques have been developed have been developed to find sequence in a time series database that are similar to a given query sequence [4], [11], [32], [39]. The existing algorithm [2], [10], [13], [20], [37] requires the user to specify the period and patterns occurring with that period, otherwise which look for all possible periods in the time series.

Some algorithms are classified based on the detection type of periodicity for symbol, sequence or segment. Another algorithm that finds frequent trends in time series data was proposed in [31]. However, this algorithm is also limited to a simple mismatch based noise model. In addition, this is a probabilistic algorithm, and is not always guaranteed to find all existing patterns. The algorithms specified in [8], [9], [17], [26], looks for all possible periods by considering the range. COVN [8] fails to perform well when the time series contains insertion and deletion noise. *WARP* [9] can detect segment periodicity; it cannot find symbol or sequence periodicity. Sheng et al. [29], [30] developed algorithm based on [15] *ParPer* to detect periodic patterns in a section of the time series; their algorithm requires the user to provide the expected period value. COVN, *WARP* and *ParPer* are augmented to look for all possible periods, and which last till the very end of the time series. Cheung [5] used suffix tree similar to *STNR* [26] which is not beneficial in terms of growth of tree. Huang and Chang [16] and *STNR* [26] presented their algorithm for finding periodic patterns, with allowable range along the time axis. Both finds all type of periodicity by utilizing the time tolerance window and could function when noise is present. *STNR* [26] can detect patterns which are periodic only in a subsection of the time series. Periodic check in *STNR* last for all the positions of a particular pattern, which in our algorithm is been reduced.

Several approaches described in the literature handle both structured motif extraction problem [22], [23] and periodicity among subsection of the time series. However, our approach described in this paper is capable of handling both motif extraction and reporting all type of periodicity. In this paper, we present a flexible algorithm that handles general extended structured motif extraction problem and uses *CBPM* to build Consensus tree. *CBPM* is capable of reporting all types of periods with or without the presence of noise in the data up to a certain level. We believe that this is an interesting problem since it allows mining for useful motif patterns with all type of period, without requiring specific knowledge about the characteristics of the resulting motif.

Many researchers have tried to use data mining technologies in areas related to meteorology and weather prediction. Kotsiantis et al. [12] predict daily average, maximum and minimum temperature for Patras city in Greek by using six different data mining methods: Feed-Forward Back Propagation (BP), k-Nearest Neighbor (KNN), M5rules algorithm, linear least-squares regression (LR), Decision tree and instance based learning (IB3). They use four years period data [2002-2005] of temperature, relative humidity and rainfall. The results they obtained in this study were accurate in terms of Correlation Coefficient and Root Mean Square. The emphasis in [4] is on using DBSCAN (Density Based Spatial Clustering of Applications with Noise) clustering algorithm to categorize Turkey into regions according to climatic characteristics. They use the daily maximum and minimum temperature records between 1930 and 1996 from 258 stations. They draw that this type of data mining application can help meteorological to create faster forecast and decisions and provide more performance and reliability than any other methods.

Data mining have been employed successfully to build a very important applications in the field of meteorology like predicting abnormal events like hurricanes, storms and river flood prediction [2][15]. These applications can maintain public safety and welfare. In this context, Zhang and Huang [22] propose a new framework to discover dynamic inter-dimension association rules for local-scale weather prediction of Dallas City. The usefulness of applying association mining is to find a strong relation between severe conditions and the change tendencies of the measurements of the weather. The authors conclude with some predicates extracted from the obtained rules. Another contribution to detect severe events using data mining is by [14] and [18]. Peters at el. [18] used the volumetric radar data to detect storm events and classify them into four types: hail, heavy rain, tornadoes, and wind.

Using data mining in meteorological application is not limited to prediction, but it also extend to participate in many important fields like water resource management [11] and air pollution management [13].

Mining techniques also can be applied to various types of data like weather images and radar maps extend to characteristic features extracted from this weather images can be used to represent various weather patterns [21].

Prediction of the future values by analyzing Temperature and humidity data is one of the important parts which can be helpful to the society as well as to the economy. Work has been

done in this constrain since years. Different techniques have been applied to predict the temperature and humidity and other parameters of weather. Some of the work in this area is as follows : In data mining, the unsupervised learning technique of clustering is a useful method for ascertaining trends and patterns in data. Most general clustering techniques do not take into consideration the time-order of data. *Tasha R. Inniss* used a mathematical programming and statistical techniques and methodologies to develop a seasonal clustering technique for determining clusters of time series data, and applied this technique to weather and aviation data to determine probabilistic distributions of arrival capacity scenarios, which

can be used for efficient traffic flow management. The seasonal clustering technique is modeled as a set partitioning integer programming problem and resulting clustering's are evaluated using the mean square ratio criterion [2]. The resulting seasonal distributions, which have satisfied the mean square ratio criterion, can be used for the required inputs (distributions of airport arrival capacity scenarios) into stochastic ground holding models.

In combination, the results would give the optimal number of flights to ground in a ground delay program to aid more efficient traffic flow management [2].

S. Kotsiantis, A. Kostoulas, S. Lykoudis, A. Argiriou, K. Menagias investigate the efficiency of data mining techniques in estimating minimum, maximum and mean temperature values. Using temperature data from the city of Patras in Greece, a Regression algorithm is applied for the number of results. The performance of these algorithms has been evaluated using standard statistical indicators, such as Correlation Coefficient, Root Mean Squared Error, etc.

[1] ***Godfrey C. Onwubolu1, Petr Buryan, Sitaram Garimella,Visagaperuman Ramachandran,Viti Buadromo and Ajith Abraham,*** presented the data mining activity that was employed in weather data prediction or forecasting. The approach employed is the enhanced Group Method of Data Handling (e-GMDH). The weather data used for the research include daily temperature, daily pressure and monthly rainfall [3]. The results of e-GMDH were compared with those of PNN and its variant, e-PNN. E-GMDH outperformed PNN an its variant in most modeling and prediction problem. They showed that end users of data mining should endeavor to follow the methodologies of data mining since suspicious data points or outliers in a vast amount of data could give unrealistic results which may affect knowledge inference.

S. Kotsiantis, A. Kostoulas, S. Lykoudis, A. Argiriou, K.Menagias proposed a hybrid data mining technique that can be used to predict more accurately the mean daily temperature values [4], it was found that the regression algorithms could enable experts to predict temperature values with satisfying accuracy using as input the temperatures of the previous years. The hybrid data mining technique produce the most accurate results.

Simple temperature prediction methods mining in the past weather data records produced accurate prediction for development of intelligent control solutions. The problem was closely related to the prediction of the actual weather conditions within the immediate environment of the greenhouse, an intelligent greenhouse collects its own climate data, with time weather records from weather station localized strictly by the greenhouse were mined to the algorithm, increasing the prediction accuracy. *Peter Eredics* demonstrates the limited performance of uninformed, simple methods for temperature forecasts, and introduces more accurate solutions using information from the problem domain [5].

Discovering the periodicity rate of time series data has drawn the attention of the data mining research community very recently. Indyk et al. [15] have addressed this problem under the name periodic trends and have developed an O(n $\log 2^n$) time algorithm, where n is the length of the time series. Their notion of a periodic trend is the relaxed period of the entire time series, which is similar to our notion of segment periodicity. However, our proposed algorithm for segment periodicity detection performs in O(n log n) time. We conduct a thorough performance study to compare our proposed segment periodicity detection algorithm to the periodic trends algorithm of [15]. In addition to the saving in time performance, our proposed segment periodicity detection algorithm is more resilient to noise and produces more accurate periods. The proposed segment periodicity detection algorithm favors the shorter periods rather than the longer ones that are favoured by the periodic trends algorithm of [15]. The shorter periods are more accurate than the longer ones since they are more informative. For example, if the daily power consumption of a specific customer has a weekly pattern, it is more informative to report a period, say, of length 7, than to report the periods 14, 21, or other multiples of 7. Specific to partial periodic patterns, Ma and Hellerstein [19] have developed a linear distance-based algorithm for discovering the potential periods regarding the symbols of the time series. In [24], a similar algorithm has been proposed with some pruning techniques. However, both algorithms miss some valid periods since they only consider the adjacent inter arrivals. For example, consider a symbol that occurs in a time series in positions 0, 4, 5, 7, and 10. Since that symbol occurs in positions 0, 5, and 10, one of the underlying periods for that symbol should be 5. However, a distance-based algorithm only considers the adjacent inter arrival times 4, 1, 2, and 3 as candidate periods, which clearly do not include the value 5. Should it be extended to include all possible interarrivals, the complexity of a distance-based algorithm [24], [19] would increase to O(n^2). Although Berberidis et al. [7] have proposed an algorithm that considers all possible potential periods, their algorithm is inefficient as it considers one symbol at a time. Moreover, the algorithms of [24], [19], [7] require additional phase over the time series in order to output the periodic patterns. Not only does our proposed symbol periodicity detection algorithm perform in O(n log n), it also discovers all possible potential periods as well as their corresponding periodic patterns simultaneously.

Elfekey et al. [5] propose algorithms to mine two pre-defined types of periodicities in time-series data. Berberidis et al. [1] propose an algorithm that mines a set of candidate periods featured in a time-series that satisfy a minimum confidence threshold. Elfekey et al. [3] propose an algorithm for mining periodic patterns in time-series databases with unknown or obscure periods.Yang et al. [18] and Huang and Chang [6] propose algorithms for mining asynchronous periodic patterns in time-series data. Lai et al. [11] address the problem of mining periodicity of patterns that occur across artificial boundaries. Karli and Saygin [9] propose two techniques for mining periodic spatio-temporal patterns at different time granularities. Zhang et al. [19] present practical algorithms to solve the problem of mining frequently occurring periodic patterns with a gap requirement from sequences. Ma and Hellerstein [13] study partial periodic patterns taking into account imprecise time information, noisy data and shifts in phase and/or periods. Algorithms for incremental mining of partial periodic patterns in time-series archives are proposed and analyzed empirically by Elfekey et al. [4]. Lee et al. [12] address the problem of mining multiple partial periodic patterns in a parallel computing environment. Mahanta et al. [14] and Dutta and Mahanta [2] propose algorithms to extract calendar-based periodic temporal patterns across discrete domains. In this paper, a generalized method is proposed to detect calendar-based periodicities of temporal patterns occurring across a sequence of time-intervals not only in a discrete domain but across a continuous domain also. A theorem establishing an interesting relationship between periodicities of patterns at different levels of a time-hierarchy is also formulated.

Over the last decade, many interesting techniques of periodicity mining were proposed to detect various types of periodicities namely symbol, sequence and segment periodicity. The existing Periodicity Mining techniques are Text-Based which are more appropriate for numerical and alpha numerical data for deriving periodic patterns. Of the various mining techniques, not all techniques detect all the different types of periodicities. The ability to handle the imperfectly occurring periodicities is limited to certain techniques at the cost of poor memory management and restricted types of periodicity detection. Only few techniques are resilient to noise, but those techniques possess greater response time. The traditional Apriori Mining Technique [1] needs multiple scans of the database for generating the candidate keys. Its application is also limited to Sequence Periodicity. The technique that incorporates Fast Fourier Transform namely the Convolution [9] and Filter Refine Paradigm [3] for reducing the time complexity achieves performance at the cost of increased computational complexity. The application of Fast Fourier Transform (FFT) also complicates the process of finding the Partial Periodicity, since it considers the time series as whole, not separate entities. The LSI (Longest Subsequence Identification) algorithm [7] on the other hand handles both perfect and imperfect periodicities, but can report only the longest subsequence, not every pattern. The only algorithm that produces more accurate results in the presence of insertion, deletion and replacement noise is the Dynamic Time Warpimg algorithm [6] which is restricted to address only the segment periodicity. It also possesses higher time complexity, when compared to other mining techniques. The Suffix Tree based Noise Resilient (STNR) algorithm [10] is the only algorithm that addresses all types of periodicities that have perfect and imperfect occurrences. It also possesses high resilience towards noise. This STNR algorithm uses a Suffix tree data structure [11], [12], [13] that has been proven to be very useful in string processing. It can be efficiently used to find a substring in the original string and to find the frequent substrings. But this STNR algorithm is not the most appropriate for multimedia data due to its limited application to strings. Also in the worst case the time complexity of this algorithm may go up to O(n3), which is alleged to be very poor. It also poses some difficulty in interpreting the patterns from the Suffix tree when it grows larger for huge data volumes. Among the various periodicity mining techniques, the only technique that can be applied to multimedia data represented in the form of digital signals is Convolution [9] [14]. But this technique also has its own limitations. As FFT is used, it is restricted to determine only symbol and segment periodicities. It does not perform well, when the time series contains imperfect patterns. Its resilience is also limited to replacement noise and not insertion or deletion noise. The above discussions clearly show that none of the existing periodicity detection techniques can be deployed for mining multimedia data, due to the various limitations and restrictions involved. There is always a trade-off between the performance of the approaches and their ability to deal with the various types and occurrences of the periodicities.

There are various kinds of time series data related research, for example, finding similar time series(Agrawal etal.,1993a;Berndt and Clifford,1996;ChanandFu,1999), sub-sequence searching in time series (Faloutsos etal.,1994), dimensionality reduction (Keogh, 1997b;Keoghetal.,2000) and segmentation(Abonyi et al.,2005). Those research been studied in considerable detail by both database and pattern recognition communities for different domains of time series data(Keogh andKasetty,2002). In the context of time series data mining ,the fundamental problem is how to represent the time series data. One of the common approaches is transforming the time series to another domain for dimensionality reduction followed by an indexing mechanism. Moreover, similarity measure between time series or time series sub-sequences and segmentation are two core tasks for various time series mining tasks. Based on the time series representation, different mining tasks can be found in the literature and they can be roughly classified into four fields: pattern discovery and clustering, classification, rule discovery and summarization. Some of the research concentrates on one of these fields, while the others may focus on more than one of the above processes.

III. CONCLUSION

In this paper, we have discussed different algorithms that can detect periodicity in time series database. Some authors proposed separate algorithm for different periodicity. The periodicity detection in time series play an important role in many application. Periodic detection algorithm should detect all type of periodicity and also partial periodicity. The single algorithm can find symbol, sequence (partial periodic), and segment (full cycle) periodicity in the time series. It can also find the periodicity within a subsection of the time series. We performed several experiments to show the time behaviour, accuracy, and noise resilience characteristics of the data.

REFERENCES

[1] Faraz Rasheed, Mohammed Alshalfa, Reda Alhaji Associate Member,IEEE," Efficient periodicity Mining In Time Series Databases Using Suffix Trees" IEEE Trans. Knowledge and Data Eng., vol.23, no.1, pp.79-94, January2011.

[2] M.G. Elfeky, W.G. Aref, and A.K. Elmagarmid, "PeriodicityDetection in Time Series Databases," IEEE Trans. Knowledge andData Eng., vol. 17, no. 7, pp. 875-887, July 2005.

[3] M.G. Elfeky, W.G. Aref, and A.K. Elmagarmid, "WARP: TimeWarping for Periodicity Detection," Proc. Fifth IEEE Int'l Conf.Data Mining, Nov. 2005.

[4] J. Han, Y. Yin, and G. Dong, "Efficient Mining of Partial PeriodicPatterns in Time Series Database," Proc. 15th IEEE Int'l Conf. DataEng., p. 106, 1999.

[5] N. Kumar, N. Lolla, E. Keogh, S. Lonardi, C.A. Ratanamahatana,and L. Wei, "Time-Series Bitmaps: A Practical Visualization Toolfor Working with Large Time Series Databases," Proc. SIAM Int'lConf. Data Mining, pp. 531-535, 2005.

[6] F. Rasheed and R. Alhajj, “STNR: A Suffix Tree Based NoiseResilient Algorithm for Periodicity Detection in Time SeriesDatabases,” Applied Intelligence, vol. 32, no. 3, pp. 267-278, 2010.

[7] Mukesh Kumar,Arvind Kalia ,”Preprocessing and Symbolic Representation of Stock Data ”,2012 Second International Conference on Advanced Computing & Communication Technologies

[8] Hiroshi Sugimura,Kazunori Matsumoto,”Classification System for Time Series Data Based onFeature Pattern Extraction”,2011 IEEE

[9] R. Grossi and G.F. Italiano, “Suffix Trees and Their Applicationsin String Algorithms,” Proc. South Am. Workshop String Processing,pp. 57-76, Sept. 1993.

[10] E.F. Glynn, J. Chen, and A.R. Mushegian, “Detecting PeriodicPatterns in Unevenly Spaced Gene Expression Time Series UsingLomb-Scargle Periodograms,”Bioinformatics, vol. 22, no. 3 pp. 310316,Feb. 2006.

[11] C. Berberidis, W. Aref, M. Atallah, I. Vlahavas, and A.Elmagarmid, “Multiple and Partial Periodicity Mining in TimeSeries Databases,” Proc. European Conf. Artificial Intelligence, July2002.

[12] E. Keogh, J. Lin, and A. Fu, “HOT SAX: Efficiently Finding theMost Unusual Time Series Subsequence,” Proc. Fifth IEEE Int’lConf. Data Mining, pp. 226-233, 2005.

[13] Elfeky MG, Aref WG, Elmagarmid AK (2004a) Using convolution to mine obscure periodic patterns in one pass. In: EDBT 2004: Proceedings of the ninth international conference on extending database technology. LNCS 2992, Berlin, Springer-Verlag, pp 605-620

[14] Yang J, Wang W, Yu PS (2003) Mining asynchronous periodic patterns in time series data. IEEE Trans Knowl Data Eng 15(3):613-628

[15] Huang K, Chang C (2005) SMCA: A general model for mining asynchronous periodic patterns in temporal databases. IEEE Trans Knowl Data Eng 17(6):774-785

