International Journal of Computer Applications Technology and Research
Volume 3– Issue 4, 228 - 238, 2014, ISSN: 2319–8656

Analysis of Various Periodicity Detection Algorithms in Time Series Data with Design of New Algorithm

Shital P. Hatkar
BAMU University
Aditya Engineering
College Beed, India

S.H.Kadam
BAMU University
Aditya Engineering
College, Beed, India

Syed A.H
BAMU University
Aditya Engineering
College, Beed, India

Abstract:Time series datasets consist of sequence of numeric values obtained over repeated measurements of time. They are Popular in many applications such as stock market analysis, power consumption, economic and sells forecasting, temperature etc. Periodic pattern mining or periodicity detection is process of finding periodic patterns in time series database. It has a number of applications, such as prediction, forecasting, detection of unusual activities, etc. Periodicity mining needs to give more attention as its increased need in real life applications. The types of periodicities are symbol periodicity, sequence periodicity and segment periodicity and they should be identified even in the presence of noise in the time series database. There are number of algorithms exists for periodic pattern mining. Those algorithms have some advantages and disadvantages. In this paper, we have compared different periodicity mining algorithms and given plan for developing efficient periodicity mining algorithm which detect symbol periodicity, sequence periodicity and segment periodicity and noise-resilient .

Keywords: periodic pattern mining, periodicity detection, time series data, symbol periodicity, sequence periodicity, segment periodicity.

1. INTRODUCTION

Time series data captures the evolution of a data value over time. Life includes several examples of time series data. Examples 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 A time series data is a collection of values represented at uniform intervals of time.A pattern is said to be periodic if, it appears again and again in time series at uniform intervals of time. The length of interval defines period of the frequent pattern. Periodicity mining deals with the methods for analyzing time series data in order to extract meaningful statistics and other characteristics of data. Periodicity mining is a tool that helps in predicting the behavior of time series data. For example, periodicity mining allows an energy company to analyze power consumption patterns and predict periods of high and low usage so that proper planning may take place. Data mining, which is also called Knowledge-Discovery from data. It is the process of searching enormous volumes of data for patterns using association rules with the help of computer programs. It is one of the most important area of research. A time series is mostly characterized by being composed of repeating cycles. For instance, there is a traffic jam twice a day when the schools are open; number of transactions in a superstore is high at certain periods during the day, certain days during the week, and so on. Identifying repeating (periodic) patterns could reveal important observations about the behavior and future trends of the case represented by the time series and hence would lead to more effective decision making. In other words, periodicity detection is a process for finding temporal regularities within the time series, and the goal of analyzing a time series is to find whether and how frequent a periodic pattern (full or partial) is repeated within the series. In this paper our focus is on analysing various periodicity detection algorithms in time series data.

1.1 Types Of Periodic Patterns

In general, three types of periodic patterns can be detected in a time series:

1.1.1 Symbol Periodicity

A time series is said to have symbol periodicity if at least one symbol is repeated periodically For example, in time series T =abd acb aba abc, symbol a is periodic with periodicity p = 3, starting at position zero (stPos= 0).

1.1.2 Sequence Periodicity

A pattern consisting of more than one symbol may be periodic in a time series; and this leads to partial periodic patterns. For instance, in time series T = bbaa abbd abca abbc abcd, the sequence ab is periodic with periodicity p= 4, starting at position 4 (stPos= 4);and the partial periodic pattern ab _ _ exists in T, where _denotes any symbol or don’t care mark.

1.1.3 Segment Periodicity.

If the whole time series can be mostly represented as a repetition of a pattern or segment, then this type of periodicity is called segment or full-cycle periodicity.For instance, the time series T = abcab abcab abcab has segment periodicity of 5 (p = 5) starting at the first position (stPos 0), i.e., T consists of only three occurrences of the segment abcab. It is not necessary to always have perfect periodicity in a time series as in the above three examples.Usually the degree of perfection is represented by confidence, which is 100 percent in the above three examples. On the other hand, real-life examples are mostly characterized by the presence of noise in the data and this negatively affects the confidence level.

1.2Periodicity Detection In Time Series Data In Presence Of Noise

Three types of noise usually measured in time series data are substitution, addition, and deletion noise. In substitution noise, various symbols in the dishonored time series are restored at arbitrary with further symbols. In case of addition and deletion noise, several symbols are interleaved or deleted, correspondingly arbitrarily at diverse locations (or time values). Noise is also a combination of these three kinds, for example, substitution type noise resources the consistent combination of replacement (R) and addition (I) noise. When the time series is moreover completely cyclic or includes only substitution noise and achieves inadequately in the occurrence of addition or deletion noise. This is since insertion and deletion noise develops or deals the time axis important to move of the imaginative time series values.This paper provides some discussion about some of the available algorithms for periodicity detection. It also includes design of efficient technique that has optimum performance parameter.

2. LITERATURE SURVEY

B.Sujatha and Dr.S.Chenthur Pandian have Presented a survey on periodicity detection in time series database[1]. Their paper presents a survey on some of the existing periodic pattern mining techniques.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.

Jisha Krishnan and Chitharanjan K have studied different periodicity algorithm and done comparison among four algorithms[2]. They state that Periodicity mining is used for predicting different applications such as prediction, forecasting etc. It has several application in Time series databases. Several algorithms are present for detecting the periodicity. But most of the algorithm do not take into account the presence of noise or partial periodicity. Based on time wrapping, the first algorithm wraps the time axis to optimally remove the noise at various locations. The Partial Periodic Detection Algorithm algorithm can be viewed as a variation of the approximate string matching algorithm. Periodic Detection using convolution method algorithm is used for partial periodicity detection and in the Periodic Detection using SuffixTree the periodic detection is done using suffix tree.This algorithms detects periodicity in noise and also detects partial periodicity.

Christos Berberidis, Walid G. Aref, Mikhail Atallah, Ioannis Vlahavas, Ahmed K. Elmagarmid[3] have put limelight on Multiple and Partial Periodicity Mining in Time Series Databases. Periodicity search in time series is a problem that has been investigated by mathematicians in various areas, such as statistics, economics, and digital signal processing. For large databases of time series data, scalability becomes an issue that traditional techniques fail to address. In existing time series mining algorithms for detecting periodic patterns, the period length is user specified. This is a drawback especially for datasets where no period length is known in advance. They propose an algorithm that extracts a set of candidate periods featured in a time series that satisfy a minimum confidence threshold, by utilizing the autocorrelation function and FFT as a filter.

G.N.V.G. Sirisha , M. Shashi & G.V. Padma Raju have studied the periodic pattern mining[4]. The merits and demerits of different algorithms was given in their paper. Owing to a large number of applications periodic pattern mining has been extensively studied for over a decade. Periodic pattern is a pattern that repeats itself with a specific period in a given sequence. Periodic patterns can be mined from datasets like biological sequences, continuous and discrete time series data, spatiotemporal data and social networks. Periodic patterns are classified based on different criteria. Periodic patterns are categorized as frequent periodic patterns and statistically significant patterns based on the frequency of occurrence. Frequent periodic patterns are classified as perfect and imperfect periodic patterns, full and partial periodic patterns, synchronous and asynchronous periodic patterns, dense periodic patterns, approximate periodic patterns. Their paper presents a survey of research on periodic pattern mining algorithms and their application areas. They also presents a brief overview of algorithms that can be applied for specific types of datasets like spatiotemporal data and social networks.

M.G. Elfeky, W.G. Aref, and A.K. Elmagarmid have developed an efficient algorithm for detecting each type of periodicity in O(nlogn)[5]. In their paper, they defined two types of periodicities for time series databases. Whereas symbol periodicity addresses the periodicity of the symbols in the time series, segment periodicity addresses the periodicity of the entire time series regarding its segments. They have proposed a scalable, computationally efficient algorithm for detecting each type of periodicity in O(nlogn) time, for a time series of length n. An empirical study of the algorithms using real- world and synthetic data sets proves the practicality of the problem, validates the accuracy of the algorithms, and validates the usefulness of their outputs. Moreover, segment periodicity detection takes less execution time whereas symbol periodicity detects more periods. They conclude that in practice, segment periodicity detection could be applied first and if the results are not sufficient, or not appealing, symbol periodicity detection can be applied. They also have studied the integration of their proposed periodicity detection algorithms in the entire process of time series mining, and have proved its effectiveness in the case of partial periodic patterns mining.

M.G. Elfeky, W.G. Aref, and A.K. Elmagarmid, mentioned the concept of time warping for periodicity algorithm[6]. In that paper, they have proposed a time warping algorithm, named WARP, for periodicity detection in the presence of noise. To handle efficiently all types of noise, WARP extends or shrinks the time axis at various locations to optimally remove the noise. Furthermore, he have proposed an online version of WARP that fits the data stream model. An empirical study using synthetic data shows that there is a tradeoff between noise resiliency and time performance. WARP is more noise resilient, yet requires more processing time, than the previous periodicity detection algorithms. Moreover, Online WARP is shown empirically to be reasonably accurate, even under low memory resources.

David Lo et al has 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)[7]. 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., 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.

Jinlin Chen has presented an updown directed acyclic graph approach for sequential pattern mining[8]. 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.

A suffix tree based noise resilent algorithm for periodicity detection in time series database is proposed by Faraz Rasheed et al [9]. 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.

Dr.Ramachandra, V.Pujeri, G.M.Karthik[10] have proposed a novel algorithm that uses FP tree as underlying structure. The algorithm can detect symbol, sequence and segment periodicity as well as present the patterns that are periodic.The search for the periodicity in time-series database has a number of application, is an interesting data mining problem. In real world dataset are mostly noisy and rarely a perfect periodicity, this problem is not trivial. Periodicity is very common practice in time series mining algorithms, since it is more likely trying to discover periodicity signal with no time limit. They propose an algorithm uses FP-tree for finding symbol, partial and full periodicity in time series. They designed the algorithm complexity as $O(kN)$, where N is the length of input sequence and k is length of periodic pattern. They have shown their algorithm is fixed parameter tractable with respect to fixed symbol set size and fixed length of input sequences. Experiment results on both synthetic and real data from different domains have shown their algorithms has time efficient and noise-resilient feature. A comparison with some current algorithms demonstrates the applicability and effectiveness of the proposed algorithm.

Efficient periodicity mining in time series databases using suffix tree is proposed by Faraz Rasheed et al.[11] 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.

David Lo et al [12] 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 (i.e) 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.

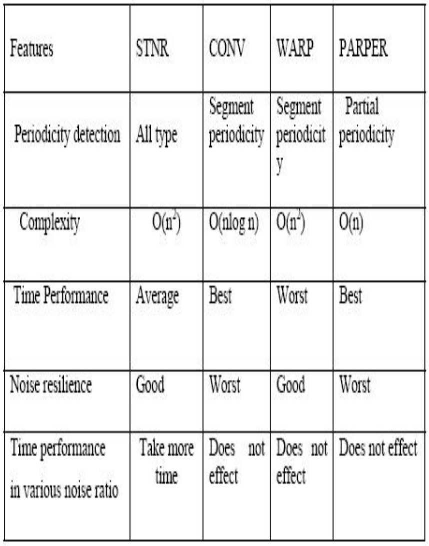
Avrilia Floratou et al [13] give a technique for efficient and accurate discovery of patterns in sequence datasets. The main aim of sequential data mining applications is to discover frequently occurring patterns. The challenge behind this frequent pattern is allowing some noise in the matching process. 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.

Obules u et al., suggests a pruning strategy to remove redundant data in spatiotemporal database[14]. 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.

S. Drishya, and I. Nancy Jeba Jingle have implemented the concept of periodicity mining from the generated time series database[15]. This will make the prediction process in the time series database to be more efficient. This database contains information about the data which can be duplicated for a particular time period. Periodicity detection is a process for finding temporal regularities within the time series, and the goal of analyzing a time series is to find whether and how frequent a periodic pattern (full or partial) is repeated within the series. Suffix tree formation is useful to find the repeated patterns. Existing work mainly focused on predicting future data based pattern only which is not adopted in all types of data. They have implemented the concept of periodicity mining from the generated time series database. This will make the prediction process in the time series database to be more efficient. This database contains information about the data which can be duplicated for a particular time period. Periodicity detection is a process for finding temporal regularities within the time series, and the goal of analyzing a time series is to find whether and how frequent a periodic pattern (full or partial) is repeated within the series. Suffix tree formation is useful to find the repeated patterns. Existing work mainly focused on predicting future data based pattern only which is not adopted in all types of data. To avoid it they have implemented Suffix tree formation for predicting future data in Time Series Databases.

Eammon Keogh, Jessica Lin and Ada fu have put watch on Finding the Most Unusual Time Series Subsequence[16].In their work, they 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 they have shown, 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. They evaluate their work with a comprehensive set of experiments. In particular, they demonstrate the utility of discords with objective experiments on domains as diverse as Space Shuttle telemetry monitoring, medicine, surveillance, and industry, and demonstrate the effectiveness of our discord discovery algorithm with more than one million experiments, on 82 different datasets from diverse domains.

3. COMPARISON BETWEEN VARIOUS ALGORITHMS



Features	STNR	CONV	WARP	PARPER
Periodicity detection	All type	Segment periodicity	Segment periodicit y	Partial periodicity
Complexity	$O(n^2)$	O(nlog n)	$O(n^2)$	O(n)
Time Performance	Average	Best	Worst	Best
Noise resilience	Good	Worst	Good	Worst
Time performance in various noise ratio	Take more time	Does not effect	Does not effect	Does not effect

For comparing different types of algorithms such as WRAP,CONV, Parper, STNR etc different features are considered. The type of periodicity detected, noise ratio, time performance and algorithmic complexies are take into account. By using STNR all type of periodicity such as symbol ,segment and partial periodicity can be detected .By using CONV only symbol and segment periodicity is detected. By using WRAP only segment periodicity is detected and Parper detects only partial periodicity Next feature considered is the time performance. By comparing STNR against Parper, Parper performs well since it specify only partial periodicity. while STNR is general and finds the periodicity for all patterns which are periodic for any periodic value starting and ending anywhere in the time series.So time performance of Parper is better than Stnr but it only detects partial periodicity. By comparing CONV with STNR the time performance of CONV is better than STNR. By comparing the complexity of the four algorithm CONV perform best running time complexity of O(nlogn).The time complexity of STNR is O(n2) but it perform better because Stnr applies various optimization strategies like redundant period pruning technique.The complexity of WRAP is worse than remaining algorithm since its running time is O(n2). Parper has best running algorithm of O(n) but it is not best because it only detect partial periodicity When the noise ratio increases the efficiency of STNR increases. WRAP and STNR have good effect on noise resilience.

3.1Disadvantages Of Existing Algorithms

Algorithms based on convolution Technique has complexity of O(nlogn) and it fails to perform well when the time series contain insertion and deletion noise. CONV and Parper does not have any effect on noise ratio .When the noise increases then the efficiency of WRAP decreases. Algorithm based on time warping technique reporting O(n^2) WRAP has the worse time performance among the four type of algorithms WRAP detect only segment periodicity. The efficiency of WRAP in various noise ratio is worst. When comparing the time performance of algorithms in different noise ratio, the STNR takes more time if the noise ratio is small. PARPER needs the user to provide the expected period values and runs in linear O(n) time. It detect only partial periodic pattern. The existing algorithm has some advantages and disadvantages so we have decided to develop new algorithm.

4.PROPOSED PLAN FOR WORK

In our paper, we will study different periodicity mining algorithm and compare those algorithm and find out solution to design an efficient periodicity mining algorithm to overcome the potential demerits. For this purpose we are going to use temperature, weather and power consumption data as a input and results will be represented.

4.1Basic Plan To Design New Algorithm

It consist of some phases. In the first phase, we will study and analyse the existing algorithm so that we should aware of their merits , dmerits and complexities. In the second phase, we will present the design of the proposed algorithm. The basic structure of algorithm will be given in this phase. In next phase, we will give implementation details of proposed algorithm. In this phase the input dataset,output dataset ,flowchart, hardware and software requirements will be discussed. In the last phase, testing will be done on designed algorithm with temperature, weather and power consumption databases which are examples of time series databases.

4.2 Methodology

we will use MATLAB 7.10 to implement and evaluate the performance of the algorithm. Datasets can be used for evaluation are temperature and weather data The parameters required for evaluation are

4.2.1 Accuracy

The accuracy measure is the ability of the algorithm to detect the different periodicities that are embedded in the time series. The parameters that determine accuracy are data distribution, alphabet size (number of unique symbols in the data), size of the data (number of symbols in the data), period size, and the type and amount of noise in the data.

4.2.2 Time Complexity

The algorithm is examined to be efficient as it has the capability to detect all different kinds of periodicities (Symbol, Sequence & Segment) at their different occurrences (perfect & imperfect) within this time complexity. The time complexity is approximately the same for all different kinds of data namely the text, images and audio. The time complexity of an algorithm quantifies the amount of time taken by an algorithm to run as a function of the length of the string representing the input.

In our work, the following input parameters along with time series data will be Considered. They are Input Pattern (String), Confidence Measure , Minimum Support and Period Of Interest. The result of experiment on time series data set will be presented and evaluated on the basis of key factors accuracy and time performance and the input data will be processed using proposed algorithm in order to mine periodicity efficiently. Results will be displayed in tabular and graphical form.

5.IMPLEMENTATION PLAN

5.1 Input Dataset

Input dataset is time series databases that can be weather,temperature or power consumption data.The time series data can be represented in different forms

5.1.1 Data Adaptive

It includes Sorted Coefficients ,Piecewise Polynomial, Singular value Decomposition ,symbolic ,Trees.The Piecewise Polynomial can be subdivided in to piecewise Linear Approximation and Adaptive piecewise constant Approximation. Symbolic can be classified as Natural Language and Strings.

5.1.2 Non Data Adaptive

It is subdivided as Wavelets, Random mappings, Spectral and Piecewise Aggregate approximation. Wavelets can be represented in Orthonormal and Biorthonormal. Spectral has subtypes as discrete fourier transform and discrete cosine transform.

5.2 Pre-Processing

In this, the system is concerned with proper categorization of data stream provided as input. In addition to that the data will be normalized in proper range of values.

5.3 Symbolization

The time series database is a large volume of data, non-finite, noise interference forms. It is infeasible to analyze large data manually. So automatic or semi automatic tools are used for data analysis. Symbolization technique can be used to reduce the number of values for a given continuous attribute, by dividing the range of attribute into interval. The interval labels (input symbols) can then be used to replace actual data values. The time series database should be symbolized in order to improve analysis that is complex.The symbolization methods are evaluate in terms of information loss and compression factor.

5.4 Binarization

Binarizes a set of real-valued time series using k-means clustering edge detection, or scan statistics.

5.4.1 K-Means Clustering

For each gene, k-means clusterings are performed to determine a good separation of groups. The values belonging to the cluster with the smaller centroid are set to 0, and the values belonging to the greater centroid are set to 1.

5.4.2 Edge Detector

This approach first sorts the measurements for each gene.In the sorted measurements, the algorithm searches for differences of two successive values that satisfy a predefined condition:If the "firstEdge" method was chosen, the pair of values whose difference exceeds the scaled average gradient of all values is chosen and used as maximum and minimum value of the two groups.If the "maxEdge" method was chosen,the largest difference between two successive values is taken.

5.4.3 Scan Statistic

The scan statistic assumes that the measurements for each gene are uniformly and independently distributed independently over a certain range. The scan statistic shifts a scanning window across the data and decides for each window position whether there is an unusual accumulation of data points based on an approximated test statistic. The window with the smallest p-value is remembered. The boundaries of this window form two thresholds, from which the value that results in more balanced groups is taken for binarization.Depending on the supplied significance level, gene binarizations are rated according to the p-value of the chosen window.

5.5 Fast Fourier Transform

FFT=Fast Fourier Transform. The FFT is a faster Version of the Discrete Fourier Transform. The FFT utilizes some clever algorithms to do the same thing as the DTF, but in much less Time. FFT represent periodic time series data as a sum of sinusoidal components(sine and cosine).It also represent time series in the frequency domain We can Filter data/Extract pattern with Fourier Transform as it is one of the application of FFT. If we want to extract a pattern from time series data it is better to see it in the graph.so for that reason we are using a tool like Matlab.

5.6 Autocorrelation

We can look directly at yhe time series and ask how much information there is in the previous value that help to predict the current value acf function looks at the correlation between now and various points in the past. Autocorrelation refers to the correlation of a time series with its own past and future values. Autocorrelation is also sometimes called “lagged correlation” or “serial correlation”, which refers to the correlation between members of a series of numbers arranged in time. Positive autocorrelation might be considered a specific form of “persistence”, a tendency for a system to remain in the same state from one observation to the next. For example, the likelihood of tomorrow being rainy is greater if today is rainy than if today is dry. Geophysical time series are frequently autocorrelated because of inertia or carry over processes in the physical system. For example, the slowly evolving and moving low pressure systems in the atmosphere might impart persistence to daily rainfall. Or the slow drainage of groundwater reserves might impart correlation to successive annual flows of a river. Or stored photosynthates might impart correlation to successive annual values of tree-ring indices.Autocorrelation complicates the application of statistical tests by reducing the number of independent observations .Autocorrelation can also complicate the identification of significant covariance or correlation between time series. Autocorrelation can be exploited for predictions. An autocorrelated time series is predictable because future values depend on current and past values.Three tools for assessing the autocorrelation of a time series are the time series plot, the lagged scatterplot, and the autocorrelation function.

5.7 Output

Results will be displayed in tabular and graphical form. In output we have to consider Periodic Patterns, Period and Confidence Measure.

5.8 Platform Required

5.8.1 Software Specification

Operating System: Windows XP/7/8
Technology: MATLAB 7.10
Database : MySQL

5.8.2 Hardware Specification

Processor: Pentium Processor
Ram: 64 MB
Hard Disk: 2 GB

6. CONCLUSION

In this paper we have discussed various periodicity detection algorithms and done comparision among those algorithms. After understanding the drawbacks of existing algorithm we proposed the design of efficient algorithm. For testing the efficiency of our algorithm we have selected input as weather, temperature and power consumption data.

7. ACKNOWLEDGMENTS

I would like to express my deepest appreciation to all authors whose research papers provided me the possibility to complete this paper. A special gratitude I give to my guide whose contribution in stimulating suggestions and encouragement, helped me in writing this paper.

8. REFERENCES

[1] B. Sujatha, Dr.S. Chenthur Pandian "A Survey On Periodicity Detection In Time Series Database".Journal of Global Research in Computer Science, Volume 4, No. 5, May 2013.

[2] JishaKrishnan and Chitharanjan K "Periodicity Detection Algorithms In Time Series Databases-A Survey", International Journal of Computer Science & Engineering Technology, Vol. 4 No. 01 Jan 2013.

[3] Christos Berberidis, Walid G. Aref, Mikhail Atallah, Ioannis Vlahavas, Ahmed K. Elmagarmid "Multiple and Partial Periodicity Mining in Time Series Databases"2002.

[4] G.N.V.G. Sirisha, M. Shashi & G.V. Padma Raju, "Periodic Pattern Mining Algorithms and Applications", Global Journal of Computer Science and Technology Software & Data Engg.,Volume 13 Issue 13 Version 1.0 Year 2013.

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

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

[7] 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.

[8] 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.

[9] 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.

[10] Dr.Ramachandra.V.Pujeri,G.M.Karthik,"Constraint Based Periodicity Mining in Time Serie databases,"I.J. Computer Network and Information Security, 2012, 10, 37-46.

[11] 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

[12] 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.

[13] Avrilia Floratou et. al., "Efficient and Accurate Discovery of Patterns inSequenceDatasets",IEEE Trans-

actions on Knowledge and Data Engineering, 2011.

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

[15] S. Drishya, I. Nancy Jeba Jingle “Efficient Method for De-Duplication and Periodicity Mining In Time Series Databases”, International Journal of Engineering and Advanced Technology, Volume-1, Issue-5, June 2012.

[16] Eamonn Keogh, Jessica Lin, Ada Fu “HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence”, Proceedings of the Fifth IEEE International Conference on Data Mining, 2005.