Contents

[Introduction 2](#_Toc492568502)

[Research Problem 2](#_Toc492568503)

[Aim and Objectives 2](#_Toc492568504)

[Research Aims 2](#_Toc492568505)

[Research Objectives 2](#_Toc492568506)

[Research Questions 2](#_Toc492568507)

[Sub Question 2](#_Toc492568508)

[Limitations of the study 2](#_Toc492568509)

[2 Literature Review and Related Work 3](#_Toc492568510)

[2.1 Time Series 3](#_Toc492568511)

[2.2 Different Time Series Features 4](#_Toc492568512)

[2.3 Time series data mining tasks 4](#_Toc492568513)

[2.4 What is anomaly detection? Different Classes of Time Series Anomalies 5](#_Toc492568514)

[2.5 Related Work on Anomaly Detection Algorithm 6](#_Toc492568515)

[Research Methodology 6](#_Toc492568516)

[Results 6](#_Toc492568517)

[Discussion 6](#_Toc492568518)

[Result Discussion 6](#_Toc492568519)

[Relation with previous Researches 6](#_Toc492568520)

[Limitations 6](#_Toc492568521)

[Future Work 6](#_Toc492568522)

[Conclusion 6](#_Toc492568523)

[References 7](#_Toc492568524)

[Figure 1 : Global temperature deviations (1880 – 2000) in degree centigrade (Discrete time series) 3](#_Toc487201855)

[Figure 2: Current through a resistor (100 observations of the series x(t)=cos (.2t + π/3)) (Continuous time series) 4](#_Toc487201856)

# Introduction

## Research Problem

## Aim and Objectives

### Research Aims

### Research Objectives

## Research Questions

### Sub Question

## Limitations of the study

# 2 Literature Review and Related Work

*This section describes the concepts of time series data, Main data mining tasks on time series data objects and further discuss one of the main data mining task anomaly detection and previous studies on them.*

This thesis aims at finding out a general anomaly detection algorithm for random quasi periodic data such as telemetry data, sales data etc. There are handful of anomaly detection algorithms but most of them deals with strongly periodic data such as ECG data, gene sequence data etc. [1] [2] [3].

## 2.1 Time Series

A time series is an ordered sequence of observations xt, each recorded at a specified time t [4] [5] . A discrete-time series is a time series where these observations are made at discrete time, i.e. at fixed interval of time. On the other hand, a continuous-time series are obtained when the observations are recorded continuously over some interval of time [4]. In this thesis, we will mainly deal with the discrete time series. Time series is a temporal data object which can be obtained from different scientific, financial or software applications like Electrocardiogram (ECG), Speech Data, MRI imaging, Weather reports like daily temperature, yearly amount of rain, global warming deviation, earthquake and explosions; Business data like weekly sales totals, Quarterly earnings per share, Prices of mutual funds and stocks, Telemetry data like CPU usage, hourly users’ login, average memory usage etc [5] [6]. Here are two examples of Time Series data.

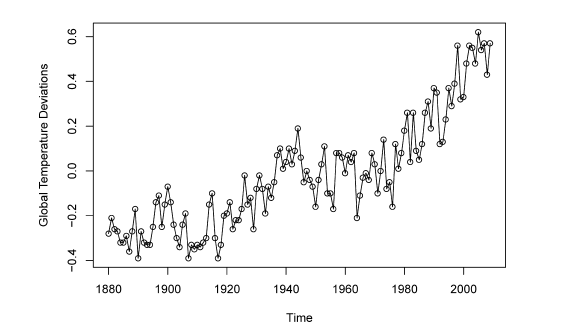


Figure 1 : Global temperature deviations (1880 – 2000) in degree centigrade (Discrete time series)

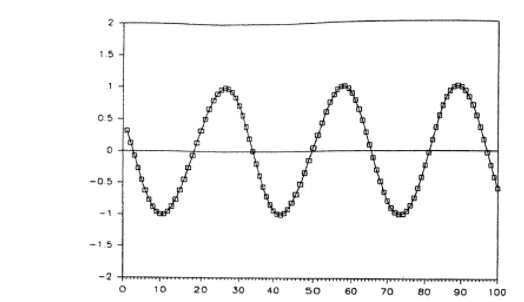


Figure 2: Current through a resistor (100 observations of the series x(t)=cos (.2t + π/3)) (Continuous time series)

## 2.2 Different Time Series Features

## 2.3 Time series data mining tasks

Time series is a popular and hugely researched data structure in data science. Several data mining tasks can be performed on a time series data. [7] [8] [4]

* **Indexing:**

Here the task is to find out the most similar time series C to a query time series Q from the time series database DB given some similarity or dissimilarity measure D (Q,C).

* **Clustering:**

Here the task is to group similar types of time series in the time series database using the similarity or dissimilarity measure D (Q,C).

* **Classification:**

Here the task is to classify or assign a query time series Q under one or more predefined class or group of time series.

* **Summarization:**

Time Series data often helps to summarize the whole story. Here the task is to describe the data graphically using some approximation process and maintaining the main features of the data but in more concise and summarized way.

* **Forecasting:**

Time series forecasting is one of the modern field of research where several classical statistical analyses are performed to predict the future. To obtain that, lot of historical data is analyzed and fit under a model to predict the future observation.

* **Anomaly detection**

This is the most recent research interest area in time series data. Here the task is to find out the data points or sections where the given query time series Q behaves unexpectedly and which can be considered as an outlier in time series.

## 2.4 What is anomaly detection? Different Classes of Time Series Anomalies

Any deviation from ‘normal’ behavior of a system can be considered as an anomaly. Generally, an anomaly is an event or observation that differs significantly from some standard referenced events, more than some threshold value. [9] [10] There are different approaches to determine the normal behavior and threshold value. “Novelty or anomaly detection refers to automatic identification of unforeseen or abnormal phenomena embedded in a large amount of normal data.” However, anomaly detection is a challenging topic due to the insufficient knowledge and different representations of the meaning of anomaly for a given system. [10] [11] [12] Often, it is use case and system specific.

For a time series data, there can be different types of anomalies. In this paper [7], the time series anomalies are broadly categorized in three different classes. [7]

|  |  |
| --- | --- |
| Outlier detection: | Given an input time series x, an outlier is a timestamp value pair (t, xt) where the observed value of xt is significantly different from the expected value at that time. |
| Change point detection: | Given an input time series x, a change point is a timestamp t where the behavior of the time series is significantly different before and after that. |
| Anomalous time series: | Given a set of time series X = {xi}, an anomalous time series xj is anomalous time series when its behavior is significantly different from the majority of time series. |

This thesis will address first two types of anomalies. But not the last one.

## 2.5 Related Work on Anomaly Detection Algorithm

There are handful of Anomaly detection algorithm for time series data. [11] [13] [14] [2]

# Research Methodology

# Results

# Discussion

## Result Discussion

## Relation with previous Researches

## Limitations

## Future Work

# Conclusion

# References

|  |  |
| --- | --- |
| [1] | N. Kumar, V. 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 *Proceedings of the 2005 SIAM international conference on data mining*, 2005. |
| [2] | D. Dasgupta and S. Forrest, "Novelty detection in time series data using ideas from immunology," 1996. |
| [3] | J. Lin, E. Keogh, S. Lonardi, J. P. Lankford and D. M. Nystrom, "Visually mining and monitoring massive time series," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004. |
| [4] | P. J. Brockwell and R. A. Davis, Time series: theory and methods, Springer Science & Business Media, 2013. |
| [5] | T.-c. Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence,* vol. 24, no. 1, pp. 164--181, 2011. |
| [6] | R. H. Shumway and D. S. Stoffer, Time series analysis and its applications, Springer Science & Business Media, 2010. |
| [7] | N. Laptev, S. Amizadeh and I. Flint, "Generic and scalable framework for automated time-series anomaly detection," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015. |
| [8] | J. Lin, E. Keogh, S. Lonardi and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, 2003. |
| [9] | M. Pradhan, S. K. Pradhan and S. K. d Sahu, "Anomaly detection using artificial neural network," *International Journal of Engineering Sciences \& Emerging Technologies,* vol. 2, no. 1, pp. 29--36, 2012. |
| [10] | J. Ma and S. Perkins, "Time-series novelty detection using one-class support vector machines," in *Neural Networks, 2003. Proceedings of the International Joint Conference on*, 2003. |
| [11] | A. Artemov and E. Burnaev, "Detecting Performance Degradation of Software-Intensive Systems in the Presence of Trends and Long-Range Dependence," in *Proceedings of the 16th IEEE International Conference on Data Mining Workshops (ICDMW), 2016*, 2016. |
| [12] | D. Goldberg and Y. Shan, "The Importance of Features for Statistical Anomaly Detection," in *HotCloud*, 2015. |
| [13] | P. K. Chan and M. V. Mahoney, "Modeling multiple time series for anomaly detection," in *Proceedings of the Fifth IEEE International Conference on Data Mining*, 2005. |
| [14] | C. Croux and P. J. Rousseeuw, Time-efficient algorithms for two highly robust estimators of scale, B-2610 Wilrijk: Universitaire Instelling Antwerpen. Department of Mathematics and Computer Science, 1992. |
| [15] | O. Vallis, J. Hochenbaum and A. Kejariwal, "A Novel Technique for Long-Term Anomaly Detection in the Cloud.," in *HotCloud*, Twitter Inc., 2014. |
| [16] | J. Ma and S. Perkins, "Time-series novelty detection using one-class support vector machines," in *Proceedings of the International Joint Conference on Neural Networks, 2003.*, 2003. |
| [17] | P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *Journal of the American Statistical association,* vol. 88, no. 424, pp. 1273--1283, 1993. |
| [18] | S. Papadimitriou, H. Kitagawa, P. B. Gibbons and C. Faloutsos, "Loci: Fast outlier detection using the local correlation integral," in *Proceedings of 19th IEEE International Conference on Data Engineering, 2003*, 2003. |
| [19] | S. Salvador, P. Chan and J. Brodie, "Learning States and Rules for Time Series Anomaly Detection.," in *FLAIRS Conference*, 2004. |
| [20] | C. Leys, C. Ley, O. Klein, P. Bernard and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *Journal of Experimental Social Psychology,* vol. 49, no. 4, pp. 764--766, 2013. |
| [21] | F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics,* vol. 11, no. 1, pp. 1--21, 1969. |
| [22] | B. Rosner, "Percentage points for a generalized ESD many-outlier procedure," *Technometrics,* vol. 25, no. 2, pp. 165--172, 1983. |