Contents

[Introduction 2](#_Toc487201810)

[Research Problem 2](#_Toc487201811)

[Aim and Objectives 2](#_Toc487201812)

[Research Aims 2](#_Toc487201813)

[Research Objectives 2](#_Toc487201814)

[Research Questions 2](#_Toc487201815)

[Main Question 2](#_Toc487201816)

[Sub Question 2](#_Toc487201817)

[Limitations of the study 2](#_Toc487201818)

[2 Literature Review and Related Work 3](#_Toc487201819)

[2.1 Time Series 3](#_Toc487201820)

[2.2 Types and Components of Time Series Models 4](#_Toc487201821)

[2.3 Time series data mining tasks 4](#_Toc487201822)

[2.4 What is an Anomaly? 4](#_Toc487201823)

[2.5 Different Types of Anomaly Detection Algorithms 4](#_Toc487201824)

[Research Methodology 5](#_Toc487201825)

[Results 5](#_Toc487201826)

[Discussion 5](#_Toc487201827)

[Result Discussion 5](#_Toc487201828)

[Relation with previous Researches 5](#_Toc487201829)

[Limitations 5](#_Toc487201830)

[Future Work 5](#_Toc487201831)

[Conclusion 5](#_Toc487201832)

[References 5](#_Toc487201833)

[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 [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] [4]. Here are two examples of Time Series data.

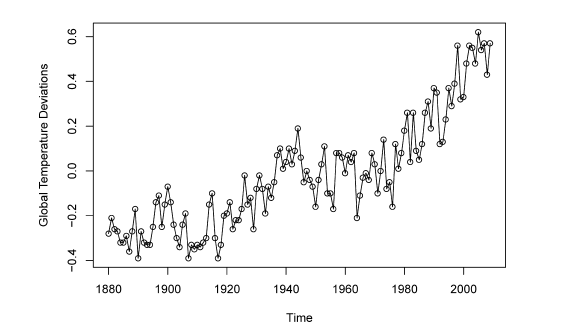


Figure 1 : Global temperature deviations (1880 – 2000) in degree centigrade (Discrete time series)

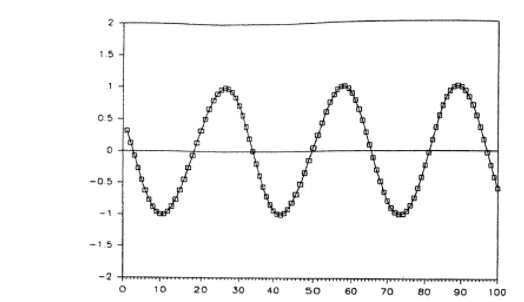


Figure 2: Current through a resistor (100 observations of the series x(t)=cos (.2t + π/3)) (Continuous time series)

## 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.

* **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.

* **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 a outlier in time series.

According to EGADS

* Forecasting
* Filtering
* Anomaly Detection

## 2.4 What is an Anomaly? Classes of Time Series Anomalies

Outlier

Change point detection

Anomalous series

This thesis deals with the first type of anomaly. But can be extended.

## 2.5 Different Types of Anomaly Detection Algorithms

You can call R functions from python using [rpy2](https://pypi.python.org/pypi/rpy2) Install rpy2 using pip with: pip install rpy2 Then use this wrapper: <https://gist.github.com/andreas-h/7808564> to call the STL functionality provided by R

# 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] | R. H. Shumway and D. S. Stoffer, Time series analysis and its applications, Springer Science & Business Media, 2010. |
| [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] | O. Vallis, J. Hochenbaum and A. Kejariwal, "A Novel Technique for Long-Term Anomaly Detection in the Cloud.," in *HotCloud*, Twitter Inc., 2014. |
| [7] | 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. |
| [8] | 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. |
| [9] | 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. |
| [10] | 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. |
| [11] | D. Goldberg and Y. Shan, "The Importance of Features for Statistical Anomaly Detection.," in *HotCloud*, 2015. |
| [12] | 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. |
| [13] | 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. |
| [14] | S. Salvador, P. Chan and J. Brodie, "Learning States and Rules for Time Series Anomaly Detection.," in *FLAIRS Conference*, 2004. |
| [15] | 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. |
| [16] | P. J. Brockwell and R. A. Davis, Time series: theory and methods, Springer Science & Business Media, 2013. |