Masters Research Project

Time Series Analysis

Supervisor: 清野健

João Gabriel Segato Kruse
Student ID 29B23068
Division Bioengineering

Graduate School of Egineering Sciences

May, 2023

Contents

Abstract

This project has as a main goal of studying and analysing time series data from biomedical sources to obtain a better understanding of their underlying properties.

1 Introduction

A time series is a collection of measurements made sequentially through time, and they are present in a variety of fields of study [1, Chapter 5]. Stock markets, electrocardiograms, weather forecasts, seismic activities, etc, are just a few examples of time series like data that appears in different areas. With such a presence, robust methods to analyse and extract information of time series are necessary, and have been active topic in high level research for a long time. Tasks such as forecasting, anomaly detection, classification and comparison methods for time series are of vital importance to many industries and people on a daily basis.

1.1 Time Series Properties

Now that the importance of understanding and analysing time series was shown, we will take a deeper look into the series themselves and some of their properties.

Continuous x Discrete Deterministic x Stochastic Description x Explanation x Prediction x Control Seasonality x Trend x Noise Stationary Time Series

1.2 Time Series in Biomedical Data

Heart Rate Electroencephalogram

2 Research Project

2.1 Proposal

2.2 Motivation

2.3 Previous Works

3 Time Series Forecasting

3.1 Problem Outline

3.2 Traditional Approaches

3.3 Deep Learning Approaches

4 Anomaly Detection

4.1 Problem Outline

Anomaly detection consists on the task of identifying patterns in data that are not consistent with its expected behaviour [2]. As such, it is applied in many different areas, from fraud detection in bank transactions all the way to health monitoring systems, as will be shown later in more details. It might be important first to distinguish between anomalies and novelties in the data, with the difference being that the later has patterns that are incorporated into the data after their first appearance. Although anomalies can intuitively be fairly easy to understand (such as the example shown in Figure 1) there are many challenges related to their actual detection, with some of the most relevant ones being:

- Defining a normal region that actually encompasses every possible normal behaviour;
- Anomalous patterns can change with time;
- Small quantities of labeled data for training;
- Noise in the samples may have similar properties to anomalies;
- The actual description of the anomaly is strictly related to the application at hand.

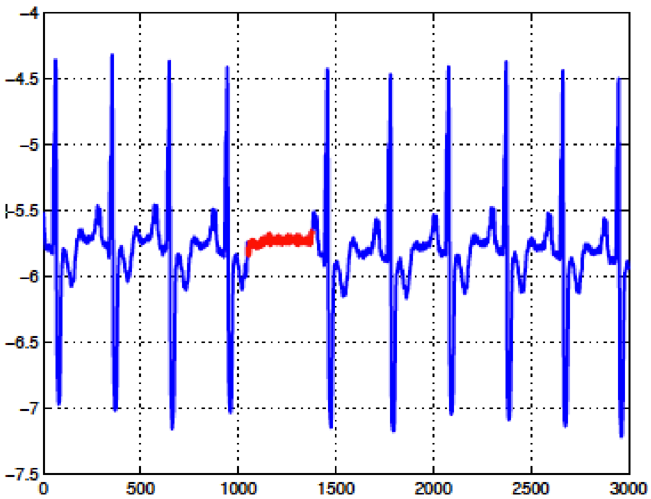

Figure 1: ECG signal showing the data deemed normal in blue and the anomaly marked in red. In this case, the anomaly correspond to an Atrial Premature Contraction(Image adapted from [2])

Anomalies can come in a few different forms. When there is a single instance of the data that has an anomalous behaviour when compared to the rest, it is called a **point anomaly**. When the anomalous behaviour is only classified as so in a specific context, being considered normal in others, it is called **contextual** or **conditional anomaly**. When a collection of related instances is anomalous with respect to the rest of the data (even if those individual instances are not anomalies by themselves) is called a **collective anomaly**.

Another important distinction to make is in the nature of the available data, that can be either labeled or unlabeled, depending on whether the dataset contains the information on the anomalies locations or not. When choosing an anomaly detection algorithm, this characteristic of the data determines whether a supervised, semi-supervised or unsupervised approach will be used. In supervised learning techniques, we have the data labeled between the normal and anomalous classes and usually train predictive models that classify the new data between the labels. One problem with this approach is that, by the nature of anomalies, datasets are highly biased with normal instances, which might affect the final predictions if not taken into account. In semi-supervised techniques, the dataset contains only one of the classes (usually the normal one), and then a model for this classes behaviour and apply it to see whether the instance follows it or not. The last type possible is unsupervised learning, in which the data available has no labels indicating what samples are normal and what are anomalous, and the model itself has to learn that in its training. The unsupervised models are the most widely applicable precisely because of this property of not requiring training data.

4.2 Traditional Approaches

4.3 Chaotic Time Series Anomalies

4.4 Deep Learning Approaches

4.5 Applications in Biosignal Analysis

When dealing with anomaly detection in medical and health monitoring applications, most of the times we are dealing with patient records. In this context, the anomalies present in the data can be due to several reasons, from indications of abnormal patient condition to instrumentation errors. As this project is related to time series analysis, anomaly detection in bio signals within this context will be the focus of analysis, but readers interested in research on other types of data can refer to [3] for a review on deep learning approaches for general medical data and [4] for a review on anomaly detection in medical images.

As was shown already, many datasets, especially when dealing with biosignals in smart wearables and such, are time series in nature.

5 Data Augmentation

5.1 Problem Outline

5.2 Traditional Approaches

5.3 Deep Learning Approaches

6 Methods

6.1 Datasets

6.2 Algorithm Description

6.3 Implementation

7 Discussion

7.1 Results

7.2 Comparison with previous methods

7.3 Future Works

8 Conclusion

References

1. Chatfield, C. & Xing, H. *The Analysis of Time Series: An Introduction with R* ISBN: 9781138066137. https://books.google.co.jp/books?id=9tP0wQEACAAJ (CRC Press, Taylor & Francis Group, 2019).

2. Chandola, V., Banerjee, A. & Kumar, V. Anomaly Detection: A Survey. *ACM Comput. Surv.* **41.** ISSN: 0360-0300. https://doi.org/10.1145/1541880.1541882 (2009).

3. Fernando, T., Gammulle, H., Denman, S., Sridharan, S. & Fookes, C. *Deep Learning for Medical Anomaly Detection – A Survey* 2021. arXiv: 2012.02364 [cs.LG].

4. Taboada-Crispi, A., Sahli, H., Hernandez-Pacheco, D. & Falcon-Ruiz, A. in *Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications* 426–446 (IGI Global, 2009).