On anomalies detection in time series and transfer learning

Arturo L. Zamorategui

November 12, 2017

Abstract

In this report I discuss the main issues regarding anomalies detection in time series. Also, I mention some of the main techniques used in the literature to analyze times series depending on their properties: is the time series stationary? is it correlated or is there some structure in the time series? Are we interested in detecting anomalous time series from a database or rather an anomalous segment of it? The answers to these questions might tell us whether or not we can classify time series describing a given production process from the information learned from other time series associated to a similar process.

1 Background

In order to detect anomalous time series from a database of time-series we need to know first what the normal behaviour is. Defining such normal or standard behaviour is not always a simple task. Also, it is important to know the nature of the data and the anomaly we are looking for: either a point anomaly, a contextual anomaly (occurying at a certain time in the series), or non-contextual anomaly [1]. From a probabilistic point of view, the normal behaviour of the time series is defined both by the expected value and the small fluctuations around the mean value (confidence interval). Then, an anomaly or a large deviation from the mean value will be given by non-Gaussian fluctuations. The most common anomalies present in a time series are a spike, an abrupt drop and trend changes. Some common statistical techniques used to analyze time series are a seasonal-trend decomposition and the auto-regressive moving averages techniques (ARMA, ARIMA) which consist of sliding a window along the time series [4, 2].

In order to extract the normal behaviour we can also use supervised and unsupervised learning methods. In the former, we assume that we know *a priori* what time series behave normally and those that behave anomalously. The drawback with these models is that usually the number of time series behaving normally outnumbers anomalous time series since the latter are rare. This fact might bias the model during the training process. For unsupervised learning, we need to build a model that describes the normal behaviour of the time series. In this case we migth take the whole database as training set.

Another issue is to decide whether we analyze the raw data in a given time series or whether we extract some features from the time series before classifying them. Extracting some feature from a time series is strongly related with dimensionality reduction and can be essential in large time-series or when a large number of samples is available, for instance.

One of the first methods to extract some features from a time series is the Fast-Fourier transform (FFT). In this case, we obtain the principal frequencies present in the signal. Other technique is the discrete Wavelet transform (DWT) that gives information both about the frequencies in the signal and the time they are present. Using dynamical systems concepts we can compute lyapunov exponents associated to the time series as well as the entropy of a signal as a measure of its randomness [5]. Further, if we decide to extract features of a time series not only by analyzing segments of the time series, we could gain insight into the structure of the time series. Some methods proposed recently regarding the segmentation analysis are the bag-of-patterns and the bag-of-words. The latter is inspired in text analysis where sequences of word present in the text can help classify the text in a specific group [5].

The final step to detect anomalous time series from the feature vector is to use a classification method. From the features extracted from the time series, the next step is to build a classifier in order to detect outliers. The most common classifiers used in biomedical signal classification are Artificial Neural Networks (ANN), Support Vector Machines (SVM), decision trees, k-means clustering and k-nearest neighbors (kNN).

Another issue that is not discussed in this report is the fact that time series can be irregular.

2 Future directions

The bag-of-words described in [5] seems to be a promising technique to detect anomalous time series. The main idea is to decompose the time series into segments as the words in a text and to extract some features of the segments via the DWT. Providing we have a histogram of codewords associated to any time series, thus a codebook associated to the database of these time series, the final step is to cluster such histograms via a classifier method. In [5], the authors use 1-Nearest Neighbor classifier and show that the bag-of-words method is effective to characterize biomedical signals such as EEGs and ECGs.

Regarding "transfer learning", in order to extend previous knowledge extracted from a database of time series to another database describing a similar process, I might propose the following methodology:

1. One might expect that two time series describing similar processes might have a similar structure.

2. Using the idea of principal component analysis, we want to extract the main components of a given series describing a particular process. One might expect that a database of time series describing the same process have a common structure as having a common fingerprint (plus noise). Therefore, another database with time series describing a similar production process might (or not) have a similar structure.

3. In order to characterize such "fingerprint" associated to a time-series database, I would use the codebook idea to build a histogram of codewords associated to that production process.

4. Next, to compare the histograms of two different databases, i.e. two different codebooks, one can define similarity measures such as Euclidean distance, Chi-squared distance, or Jensen-Shannon distance [3] to name a few.

References

[1] Chandola, V., Banerjee, A., and Kumar, V. Anomaly detection: A survey. *ACM computing surveys (CSUR) 41*, 3 (2009), 15.

[2] datascience.com. Data Science introduction to anomaly detection, 2017.

[3] Endres, D. M., and Schindelin, J. E. A new metric for probability distributions. *IEEE Transactions on Information theory 49*, 7 (2003), 1858–1860.

[4] Gupta, M., Gao, J., Aggarwal, C., and Han, J. Outlier detection for temporal data. *Synthesis Lectures on Data Mining and Knowledge Discovery 5*, 1 (2014), 1–129.

[5] Wang, J., Liu, P., She, M. F., Nahavandi, S., and Kouzani, A. Bag-of-words representation for biomedical time series classification. *Biomedical Signal Processing and Control 8*, 6 (2013), 634–644.