Time Series Anomaly Detection Using Enhanced Spectral Residual Approach

Manjunath Adinarayana, Yixuan Wang
manjunath.adinarayan@rakuten.com,yixuan.wang@rakuten.com
Rakuten Mobile, Inc.
Tokyo, Japan

ABSTRACT

Time series anomaly detection is crucial for large enterprises in monitoring of real time services and applications. In this paper, we propose enhancement to the novel algorithm based on Spectral Residual(SR) and we call it as Enhanced Spectral Residual(ESR). In our work, we have borrowed the SR concept from computer vision's saliency detection methods. We have combined innovatively the RNN, filtering of SR to obtain ESR. Our approach achieves superior experimental results compared to state-of-the-art baseline solutions on public datasets.

CCS CONCEPTS

• **Computer methodologies** → **Machine learning**; **Unsupervised learning**; **Anomaly detection**; **Signal Processing**; • **Mathematics of computing** → *Time series analysis.*

KEYWORDS

time series; anomaly detection; spectral residual; saliency map

ACM Reference Format:
Manjunath Adinarayana, Yixuan Wang. 2020. Time Series Anomaly Detection Using Enhanced Spectral Residual Approach. In *Proceedings of KDD'20,*. ACM, New York, NY, USA, 4 pages.

1 INTRODUCTION

The question of anomaly detection has interested researchers in different domains for more than a century. The question has taken many names such as "anomaly detection", "outlier detection", "novelty detection" and "peak detection." There is no universally accepted definition of anomaly or outlier. Nevertheless, many definitions depending on the context have been proposed. In general, an anomaly is a data sample that is considerably different from the whole data set.

Data associated with time is common to see at various industries. Normally, these time series data are related to tracking or monitoring services or applications at real time. It is important to explore the solutions for anomaly detection in time series because unexpected changes in data including drops, spikes, and shifts might stand for problems occur in services or applications that can cause severe revenue loss, especially in business and technology fields. In order to troubleshoot potential problems in systems and platforms on time and prevent enterprises from losing revenue, enterprises had made many researches and developed different strategies in anomaly detection. For example, recently Microsoft had launched a new anomaly detection algorithm [2] based on Spectral Residual(SR) and Convolutional Neural Network(CNN), which proved to be accurate, efficient and general. Anomaly detection service should be used to monitor different types of time series data coming from various sources, which will help to develop more efficient troubleshooting process. In this paper, we will introduce new algorithm based on existing methods and show significant improvements in existing methods.

For real-world data in industries, there are several challenges we will face and these challenges lead to specific demands for anomaly detection including:

- Generalization
  - Time series data from different sources can have diverse characteristics. For example, common patterns of time series data in industry include trend, seasonal and cyclic. In order to deal with all kinds of time series data in an efficient way, the algorithm we develop needs to have the ability of generalization.
- No need of labels
  - The definition of anomaly do not need to be defined. In other words, the anomaly detector will learn and define itself. In real world scenario, it is hard or even impossible to label all kinds of time series, and due to the changing nature of time series data, brand new patterns might appear constantly. Thus, it is important that an anomaly detection algorithm can distinguish anomaly itself.

To address these challenges, we came up with a solution that generalizes for various types of anomalies and is unsupervised in nature. At the same time the solution is robust, efficient and scalable.

In this paper, we borrowed the Spectral Residual approach [3] from the computer vision saliency detection technique. We are inspired by Microsoft's work [2] of time series anomaly detection using SR. Our work is enhancing the SR approach of Microsoft [2] and we obtained superior results on public datasets when compared to all other state-of-the-art baseline solutions. Our intuition behind considering the time-series anomaly detection same as computer vision saliency detection is that the saliency(important regions) that dominates in images catch the focus quite easily would be same in the time-series with anomalies catching the focus easily when observed visually. Anomalies in time-series can be considered

KDD'20,, August 22 -27, 2020, San Diego, CA, USA,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00

equivalent to saliencies in images. We propose enhancement to the novel algorithm based on SR and we term it as Enhanced Spectral Residual(ESR).

Our Contributions in this paper for ESR are mentioned below:

- For the first time we came up with idea of Average-filtering multiple times the Saliency Map of the time series to get superior results on public datasets.
- We introduced the formula based threshold rather than fixed threshold for anomaly detection in Saliency Map which is inspired by saliency detection in computer vision[3].
- We came up with innovative idea for extrapolation of the data points by forecasting using RNN-Seq2Seq.

2 METHODOLOGY

2.1 Enhanced Spectral Residual (ESR)

Inspired by the algorithm in computer vision, Microsoft built an anomaly detection method called Spectral Residual [3] based on Fast Fourier Transform (FFT) [4]. In computer vision, the idea of spectral residual focus on simulating the behavior of pre-attentive visual search. More specifically, log spectrum of each image is analyzed and spectral residual is obtained and transformed into the saliency map, which shows the positions of proto-objects [3]. By exploring the similar nature of detecting objects from backgrounds and detecting anomaly in time series, Microsoft applied Spectral Residual method on anomaly detection. On top of that, we improved Microsoft's SR algorithm by applying various enhancements like formula based threshold and filtering of the Saliency Map multiple times on evaluation of saliency map. We came up with 3 different versions of Enhanced Spectral Residual(ESR) namely ESR-1, ESR-2 and ESR-3.

In the below equations, after Fourier transform of input time series, $A(f)$ and $P(f)$ stand for amplitude and phase of the spectrum; $L(f)$ and $AL(f)$ stand for log amplitude specturm and averaged log amplitude specturm. At last, $R(f)$ and $S(\mathbf{x})$ stand for spectral residual and saliency map. $AS(\mathbf{x})$ is the averaged saliency map using moving average method. τ is the threshold used in evaluation of saliency map result $S(\mathbf{x})$, $FS(\mathbf{x})$ or $MFS(\mathbf{x})$. Different types of threshold are applied in different ESR models.

2.1.1 ESR-1.

In ESR-1 we use formula based threshold which is inspired from saliency detection of computer vision[3] and we directly consider the Saliency Map to check for anomalies without average-filtering the Saliency Map. Below is the mathematical explanation of ESR-1:

$$A(f) = Amplitude(\mathfrak{F}(\mathfrak{x})) \tag{1}$$
$$P(f) = Phrase(\mathfrak{F}(\mathfrak{x})) \tag{2}$$
$$L(f) = log(A(f)) \tag{3}$$
$$AL(f) = h_q(f) \cdot L(f) \tag{4}$$
$$R(f) = L(f) - AL(f) \tag{5}$$
$$S(\mathbf{x}) = ||\mathfrak{F}^{-1}(exp(R(f) + iP(f))|| \tag{6}$$
$$\tau = 7 \times E(S(\mathbf{x})) \tag{7}$$

We mark anomalies if $S(\mathbf{x}) > \tau$.

2.1.2 ESR-2.

In ESR-2, we consider averaging the Saliency Map once and there after check for anomalies using fixed threshold. Below is the mathematical explanation of ESR-2:

$$A(f) = Amplitude(\mathfrak{F}(\mathfrak{x})) \tag{8}$$
$$P(f) = Phrase(\mathfrak{F}(\mathfrak{x})) \tag{9}$$
$$L(f) = log(A(f)) \tag{10}$$
$$AL(f) = h_q(f) \cdot L(f) \tag{11}$$
$$R(f) = L(f) - AL(f) \tag{12}$$
$$S(\mathbf{x}) = ||\mathfrak{F}^{-1}(exp(R(f) + iP(f))|| \tag{13}$$
$$FS(\mathbf{x}) = \frac{S(\mathbf{x}) - AS(\mathbf{x})}{AS(\mathbf{x})} \tag{14}$$
$$\tau = 2.8 \tag{15}$$

We mark anomalies if $FS(\mathbf{x}) > \tau$.

2.1.3 ESR-3.

In ESR-3, we consider averaging the Saliency Map multiple times with different window lengths and there after check for anomalies using fixed threshold. Below is the mathematical explanation of ESR-3:

$$A(f) = Amplitude(\mathfrak{F}(\mathfrak{x})) \tag{16}$$
$$P(f) = Phrase(\mathfrak{F}(\mathfrak{x})) \tag{17}$$
$$L(f) = log(A(f)) \tag{18}$$
$$AL(f) = h_q(f) \cdot L(f) \tag{19}$$
$$R(f) = L(f) - AL(f) \tag{20}$$
$$S(\mathbf{x}) = ||\mathfrak{F}^{-1}(exp(R(f) + iP(f))|| \tag{21}$$

$A_4S(\mathbf{x})$ is 4-time filtered Saliency Map using different window lengths all the four times

$$MFS(\mathbf{x}) = \frac{S(\mathbf{x}) - A_4S(\mathbf{x})}{A_4S(\mathbf{x})} \tag{22}$$

$$\tau = 4.8 \tag{23}$$

We mark anomalies if $MFS(\mathbf{x}) > \tau$.

2.2 Extrapolation using Seq2seq RNN

The SR approach works well for those points that fall within certain window and for predicting whether the most recent data points are anomalous or not, we expect the most latest points in data to fall in centre of the window by extrapolating the needed extra points of the window. To extrapolate these needed data points we used seq2seq RNN-LSTM with feedback at the decoder for multi-step forecasting.

We even experimented with the extrapolation method by Microsoft [2]. We didn't achieve any significant improvement by using seq2seq RNN-LSTM extrapolation when compared to simple linear method of extrapolation given by Microsoft. The seq2seq RNN-LSTM extrapolation technique comes with complexity when considered for large scale anomaly detection compared to simple extrapolation method.

Figure 1: Example of a signal with anomalies found

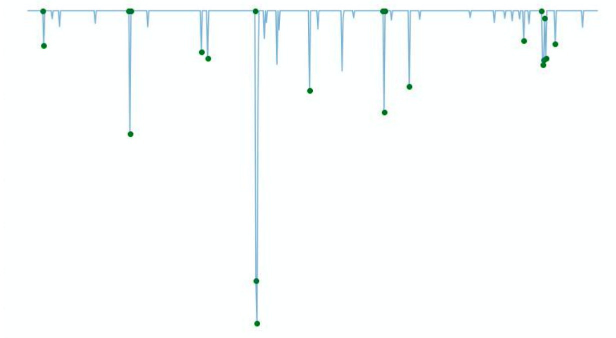
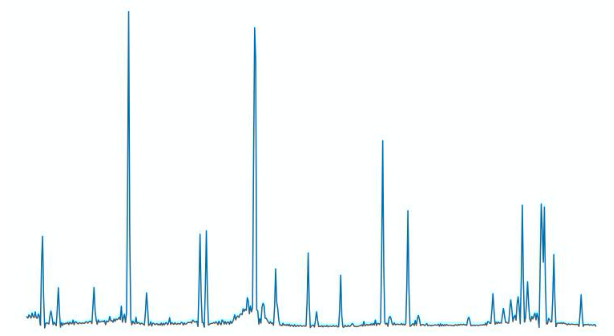


Figure 2: Saliency map of the signal



3 EXPERIMENT

3.1 Dataset

For our experiment, we used a anomaly detection dataset published by yahoo lab [1]. In the dataset, there are four groups of datasets, some of which are real traffic data of Yahoo services and others are simulated data. This dataset contains relatively typical time series data including anomaly points, trends, change points and seasonality.

3.2 Metric

The models were evaluated in two criteria including accuracy and generality. We measured accuracy by calculating test statistics of precision, recall and f1-score. Generality is also one important aspect we wanted to focus on, so we ran test on all datasets in the Yahoo benchmark dataset, which contains several typical scenarios of time series data.

3.3 Results

In ESR, shape of $h_q(f)$ q equals 3, the number of extrapolated points are 3 for the latest data points in the data, and 32 is the windowing length of the averaging filter of Saliency Map in ESR-2. 32, 12, 8, 6 are the different windowing length of the averaging filters used to filter the Saliency Map in ESR-3.

Yahoo Benchmark Dataset			
Model	Precision	Recall	f-1 score
FFT	0.202	0.517	0.291
Twitter-AD	0.166	0.462	0.245
Luminol	0.254	0.818	0.388
SR	0.404	0.765	0.529
SR-CNN	0.786	0.561	0.655
ESR-1	0.651	0.778	0.650
ESR-2	0.596	0.869	**0.659**
ESR-3	0.697	0.806	**0.693**

Table 1: Result comparison of full data(cold-start)

Yahoo Benchmark Dataset			
Model	Precision	Recall	f-1 score
SPOT	0.269	0.454	0.338
DSPOT	0.241	0.458	0.316
DONUT	0.013	0.825	0.026
SR	0.451	0.747	0.563
SR-CNN	0.816	0.542	0.652
ESR-1	0.675	0.756	**0.657**
ESR-2	0.658	0.831	**0.694**
ESR-3	0.713	0.790	**0.695**

Table 2: Result comparison of test data

3.4 Explanation

We report Precision, Recall, F1-score for the Yahoo dataset. We can see that ESR-3 significantly outperforms current state-of-the-art methods. Table-1 shows the comparison results of FFT, Twitter-AD, Luminol, SR and SR-CNN on complete data(cold-start). Our ESR-3 solution achieves 3.8% improvement in F1-score when compared to the best result achieved by baseline solutions. Table-2 shows the comparison results of FFT, Twitter-AD, Luminol, SR and SR-CNN on test data(second half of the data). Our ESR-3 solution achieves 4.3% improvement in F1-score when compared to the best result achieved by baseline solutions.

4 CONCLUSION AND FUTURE WORK

In this paper, we proposed an enhanced version of Spectral Residual approach for time series anomaly detection. In addition, for the first time we combined SR with RNN model to achieve an outperformed detection system. In the future, we are planning to ensemble the state-of-art methods together in order to provide a more robust anomaly detection system. We also want to implement reinforcement learning for anomaly detection if we have chance.

REFERENCE

[1] Yahoo. 2015. Yahoo benchmark dataset for time series anomaly. https://yahooresearch.tumblr.com/post/114590420346/a-benchmark-dataset-for-time-series-anomaly

[2]Hansheng Ren et al. 2019. Time-Series Anomaly Detection Service at Microsoft. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* - KDD 19 (2019). DOI:http://dx.doi.org/10.1145/3292500.3330680

[3]Xiaodi Hou and Liqing Zhang. 2007. Saliency Detection: A Spectral Residual Approach. *2007 IEEE Conference on Computer Vision and Pattern Recognition* (2007). DOI:http://dx.doi.org/10.1109/cvpr.2007.383267

[4]Charles Van Loan. 1992. Computational Frameworks for the Fast Fourier Transform. (1992). DOI:http://dx.doi.org/10.1137/1.9781611970999