Large-Scale Unusual Time Series Detection

Rob J Hyndman
Monash University
Victoria, Australia
Rob.Hyndman@monash.edu

Earo Wang
Monash University
Victoria, Australia
yiru.wang@monash.edu

Nikolay Laptev
Yahoo Research
California, USA
nlaptev@yahoo-inc.com

ABSTRACT

It is becoming increasingly common for organizations to collect very large amounts of data over time, and to need to detect unusual or anomalous time series. For example, a large internet company has banks of mail servers that are monitored over time. Many measurements on server performance are collected every hour for each of thousands of servers. We wish to identify servers that are behaving unusually.

We compute a vector of features on each time series, measuring characteristics of the series. The features may include lag correlation, strength of seasonality, spectral entropy, etc. Then we use a principal component decomposition on the features, and use various bivariate outlier detection methods applied to the first two principal components. This enables the most unusual series, based on their feature vectors, to be identified. The bivariate outlier detection methods used are based on highest density regions and α-hulls.

Keywords

Feature Space, Multivariate Anomaly Detection, Outliers, Time Series Characteristics

1. INTRODUCTION

In the past decade a lot of work has been done on finding the *most* similar time series efficiently [21, 13]. In this paper we focus on finding the *least* similar time series in a large set. We shall refer to such time series as *unusual* or *anomalous*. Figure 1 gives a visual motivation for our approach. Each graph in the left column shows a collection of 100 time series, two of which are outliers having an abnormal trend or seasonality. The second column shows the first two principal components which we use to identify unusual time series. Some unusual time series are not easy to identify (e.g., seasonality anomalies in Figure 1); for this reason, a robust, accurate and automated solution is critical.

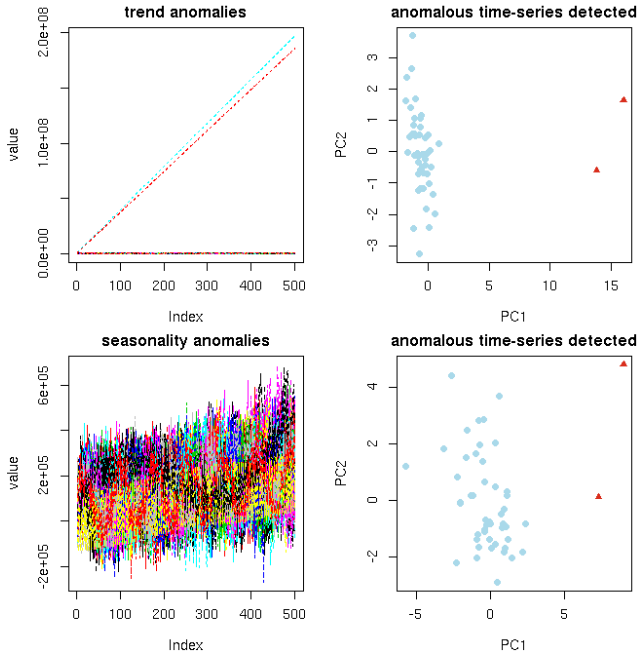

Figure 1: Different types of anomalies and corresponding first two principal components which our method uses for unusual time series detection. These types of anomalous time series may be due to an abnormal server or a malicious user.

An important motivation for efficiently finding anomalous time series comes from large internet companies. At such companies, thousands of servers power user services providing an uninterrupted and secure user experience. It is therefore critical to monitor the server metrics (e.g., latency, cpu), represented by time series, for any unusual behavior.

We are interested in the time series that are anomalous *relative* to the other time series in the same cluster, or more generally, in the same set. This type of anomaly detection is different from univariate anomaly detection or even from a multivariate point anomaly detection [6] because we are interested in identifying *entire* time series that are behaving unusually in the context of other metrics. Early detection of these anomalous time series is critical for taking preemptive action to protect users and provide a better user-experience. The solution presented in this paper has been deployed at scale within a large internet company and the open-source

version of the proposed method is being released as an R package [16]. As shown in Section 4, the proposed method has impressive performance for a wide variety of anomalies present in the time series, making it applicable to other use-cases such as identifying anomalous users, data-base transactions, retail sales and many others.

We make three fundamental contributions. First, we introduce a novel and accurate method of using PCA with α-convex hulls for finding anomalous time series. Second we perform a study of possible features that are useful for the types of time series dynamics seen in web-traffic time series. Lastly we perform experiments on both synthetic and real world data and demonstrate the usefulness and wide applicability of our method to finding *interesting* time series in a collection of other time series.

In Section 2 we present our approach that uses PCA and α-convex hulls. In Section 3 we look at the features used for explaining the variance in different scenarios. Experiments of the method are described in Section 4. Related work and conclusions are presented in Sections 5 and 6 respectively.

2. APPROACH

We first extract n features (see Section 3) from m time series. We then use Principal Component Analysis (PCA) (similar to [21]) to identify the patterns (i.e., principal components). The first two principal components (PCs) are then selected and a two dimensional outlier detection algorithm is used to find the top $k \in m$ outliers.

PCA is a tool for dimension reduction in high dimensional data. A principal component is a combination of the original variables after a linear transformation. For example the first principal component captures the maximum variation in the rows of the $m \times n$ matrix. More formally, the first principal component c_1 is given by $c_1 = \arg\max_{||c||=1||y_c||}$. Therefore, loosely speaking the first k principal components capture the k most prevalent patterns in the data.

Figure 2 shows the fraction of the variance captured by the first k principal components from real time series. We found that using the first two principal components was sufficient for our use-cases. To find anomalies in the first two PCs we use a multi-dimensional outlier detection algorithm. We have implemented a density-based and an α-hull based multi-dimensional outlier detection algorithms.

The density based multi-dimensional anomaly detection algorithm [7] finds points in the first two principal components with lowest density. The α-hull method [15] is a generalization of the convex hull [6] which is a bounding region of a point set. The α parameter in the α-hull method defines a generalized disk of radius α. When α is sufficiently large, the α-hull method is equivalent to the convex hull. Given α, an edge of the α-shape is drawn between two members of the finite point set if there exists a generalized disk of radius α containing the entire point set and the two points lie on its boundary.

3. FEATURES

We now describe the time series features we use in the PCA. While we focus on our use-case of identifying anomalous servers in a large internet company, we attempt to make our approach general and applicable to other use-cases where finding anomalous time series is critical.

The features identified should capture the global information of the time series. The features identified in our research add to an already existing set of well established features that describe time series [4] including measures of trend, seasonality, and serial correlation [20] and spectral entropy [5]. Some of features have been specifically selected to address our use-case. For example we divide a series into blocks of 24 observations to remove any daily seasonality. Then the variances of each block are computed and the variance of the variances across blocks measures the "lumpiness" of the series. Some of our features rely on a robust STL decomposition [3]. For example, the size and location of the peaks and troughs in the seasonal component are used, and the *spikiness* feature is the variance of the leave-one-out variances of the remainder component. Other features measure structural changes over time. The "level shift" is defined as the maximum difference in mean between consecutive blocks of 24 observations, "variance change" is computed similarly using variances, and the Kullback-Leibler (KL) score is the maximum difference in KL divergence (measured using kernel density estimation) between consecutive blocks of 48 observations. "Flat spots" are computed by dividing the sample space of a time series into ten equal-sized intervals, and computing the maximum run length within any single interval. Finally, "crossing points" are defined as the number of times a time series crosses the mean line.

A more detailed look at the features will be presented in the longer version of our paper.

Feature	Description
Mean	Mean.
Var	Variance.
ACF1	First order of autocorrelation.
Trend	Strength of trend.
Linearity	Strength of linearity.
Curvature	Strength of curvature
Season	Strength of seasonality.
Peak	Strength of peaks.
Trough	Strength of trough.
Entropy	Spectral entropy.
Lumpiness	Changing variance in remainder.
Spikiness	Strength of spikiness
Lshift	Level shift using rolling window.
Vchange	Variance change.
Fspots	Flat spots using disretization.
Cpoints	The number of crossing points.
KLscore	Kullback-Leibler score.
Change.idx	Index of the maximum KL score.

Table 1: Summary of features used for detecting unusual time series.

4. EXPERIMENTS

We now evaluate the effectiveness of our anomaly detection method using real-world and synthetic data comprising normal and anomalous time series. Our goal is to detect anomalous time series accurately.

The real dataset comes from a large internet company and represents the various server metrics (e.g., memory usage,

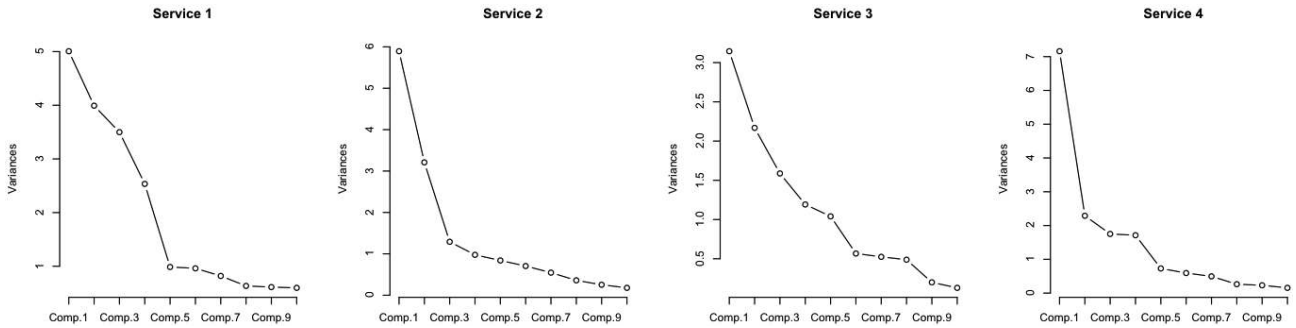

Figure 2: Scree plots showing that on our real dataset, a significant proportion of the variation can be captured using the first three to five components. For unusual time series detection we found that the first 2 components are sufficient.

Baseline Method	Description
Baseline 1	Computes Mean Absolute Difference between time series.
Baseline 2	Computes similarity between time series using discrete wavelet transform (DWT) [9].
Baseline 3	Uses PCA to extract raw time series features and uses K-Means for clustering. The time series in the smallest cluster are labeled as outliers [18].

Table 2: Summary of the baseline method.

latency, cpu). The unusual time series in the real dataset are based on a malicious activity, new feature deployment or a traffic shift. The synthetic dataset was generated by varying various time series parameters such as the trend, seasonality and noise. Both the synthetic and real datasets contain approximately 1500 time series with labeled anomalies.

4.1 Overall Detection Accuracy

Here we evaluate the average performance of our method relative to the baseline methods. Recall that our approach first extracts the two most significant principal components (PC)s from all time series and then determines the outliers in the new 2D "feature space". For PC extraction, we have tested the regular PCA and Robust PCA (RPCA). For multidimensional outlier detection on the PC space we show results for the density-based method (HDR) and for the α-hull method.

The baselines are described in Table 2. Because our method has no direct competitor, we use time series similarity and clustering techniques as baselines to detect unusual time series. We label a time series as unusual if it has a low average similarity score or it belongs to the smallest cluster.

For this experiment both real and synthetic datasets were used. For the synthetic dataset, 10 sets of time series were created. Each set consists of 1500 time series, 5 of which were creating with unusual features (e.g., unusually high seasonality). All methods were evaluated in terms of the average $accuracy = \frac{\#correct}{\#total}$ across both real and synthetic datasets.

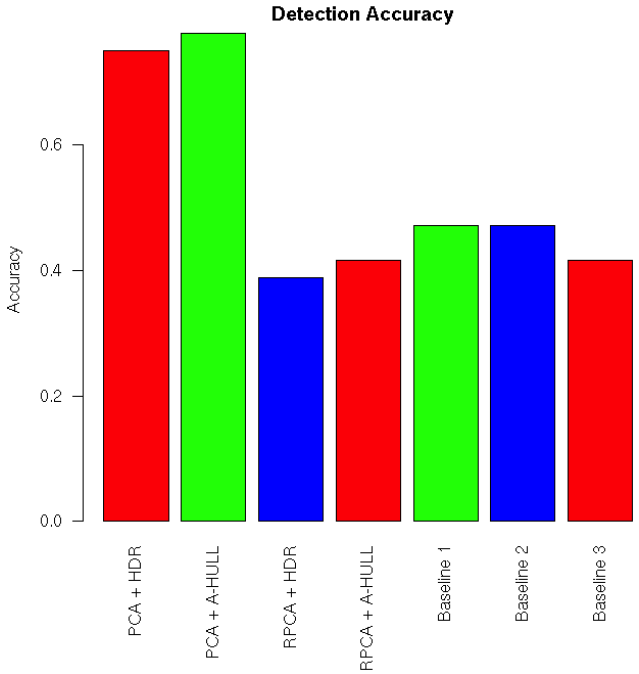

Figure 3: Average accuracy of our method compared to baseline approaches.

Figure 3 shows that our *PCA + α-hull* approach performs the best. While it is not surprising that our technique outperformed the baselines because we use a well-researched feature-space, it is surprising that the *Robust PCA* method did not perform well. This, however, can be explained by looking at the optimization equation of Robust PCA [2] which ignores outliers thereby potentially missing the principal component that explains the variance better.

4.2 Performance

Here we evaluate the performance of our algorithms compared to the baseline methods. The performance is measured in seconds as the number of total time series increases. Note that the number of unusual time series also increases proportionally to constitute roughly 2% of all time series. We can observe from Figure 4 that our approach performs favorably compared to others. Note that we were not able to run Baseline method #2 due to extremely slow performance above

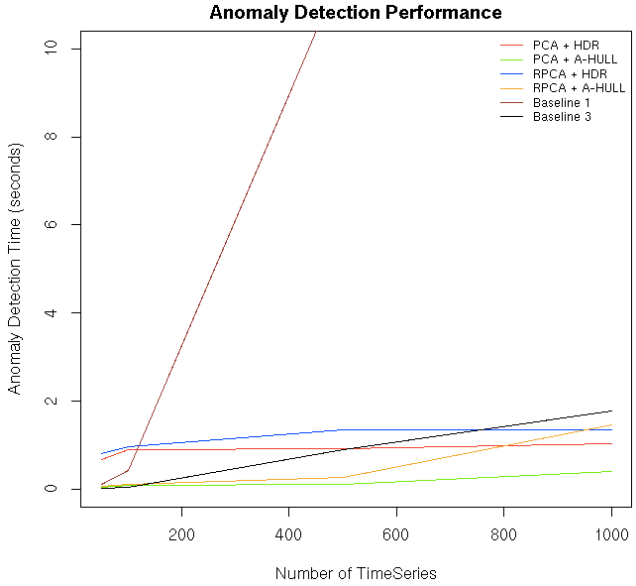

Figure 4: Scalability performance

100 time series therefore we do not include it in the comparison. Also note that the feature extraction and the anomaly detection of the PCA + α-hull increases only slightly as the number of time series is increased by an order of magnitude.

5. RELATED WORK

While our approach of identifying *entire* anomalous time series is novel, there are some parallels with existing work. For example authors in [10, 11, 12, 1] look at unusual subsequences within a single time series. PCA has also been used for detecting anomalous functions in a sample of functions [8], and for detecting univariate anomalies by [17, 19]. In addition to anomaly detection, PCA has been employed as a similarity measure used in clustering [21, 13]. Authors in [14] use PCA for a multi-dimensional visualization of a large collection of time series. None of the above methods, however, address our problem of finding unusual time series in a large collection of time series.

6. CONCLUSION

We propose using Principal Component Analysis (PCA) together with multi-dimensional anomaly detection to identify unusual time series in a large collections of time series. Our method is robust and accurate as demonstrated by the experiments over synthetic and real data from a large internet company. Our approach achieves a detection accuracy of over 80% (compared to 42% for baseline methods) and requires less than 0.5 seconds to process 1000 time series which is at least 3x faster than baseline algorithms. More experiments such as the effect on performance as the number of principle components used by the outlier detection method increases are to be presented in our full paper. Our method requires no *a priori* labeling or tuning of parameters other than the user-acceptable sensitivity threshold. Our method incorporates thoughtful selection of features that measure the types of anomalous behavior likely to occur in the time series collection. The presented approach is to be open-sourced and is already deployed at scale within a large internet company.

7. REFERENCES

[1] J. D. Brutlag. Aberrant behavior detection in time series for network monitoring. In *LISA '00 Proc. 14th USENIX Conf. System Admin.*, pages 139–146, Berkeley, CA, USA, 2000.

[2] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, 58(3):11:1–11:37, June 2011.

[3] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. STL: a seasonal-trend decomposition procedure based on loess. *J Official Stat.*, 6(1):3–73, 1990.

[4] B. D. Fulcher and N. S. Jones. Highly comparative feature-based time-series classification. *IEEE Trans. Knowl. Data Eng.*, 26(12):3026–3037, Dec. 2014.

[5] G. Goerg. Forecastable component analysis. In *Proc. 30th Int. Conf. Machine Learning*, pages 64–72, 2013.

[6] V. Hodge and J. Austin. A survey of outlier detection methodologies. *Artif. Intell. Rev.*, 22(2):85–126, 2004.

[7] R. J. Hyndman. Computing and graphing highest density regions. *Amer. Statist.*, 50(2):120–126, 1996.

[8] R. J. Hyndman and H. L. Shang. Rainbow plots, bagplots, and boxplots for functional data. *J Comp. Graph. Stat.*, 19(1):29–45, 2010.

[9] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems*, 3(3):263–286, 2001.

[10] E. Keogh, J. Lin, and A. W. Fu. HOT SAX: Efficiently finding the most unusual time series subsequence. In *5th IEEE Int. Conf. Data Mining*, pages 226–233, 2005.

[11] E. Keogh, J. Lin, A. W. Fu, and H. Van Herle. Finding unusual medical time-series subsequences: Algorithms and applications. *IEEE Trans. Inf. Tech. Biomedicine*, 10(3):429–439, 2006.

[12] E. Keogh, S. Lonardi, and B. Y. Chiu. Finding surprising patterns in a time series database in linear time and space. In *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, pages 550–556. ACM Press, 2002.

[13] E. J. Keogh and M. J. Pazzani. A simple dimensionality reduction technique for fast similarity search in large time series databases. In *Knowledge Discovery & Data Mining: Current Issues & New Applications*, pages 122–133. Springer, 2000.

[14] D. T. Nhon and L. Wilkinson. Timeseer: detecting interesting distributions in multiple time series data. In *Proc 5th Int. Symposium on Visual Information Communication and Interaction*, pages 43–51. ACM, 2012.

[15] B. Pateiro-López and A. Rodrıguez-Casal. Generalizing the convex hull of a sample: The R package alphahull. *J. Stat. Soft.*, 34(5):1–28, 4 2010.

[16] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

[17] M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn, and L. Chang. Principal component-based anomaly detection scheme. In *Foundations & Novel Approaches in Data Mining*, pages 311–329. Springer, 2006.

[18] A. Singhal and D. E. Seborg. Clustering multivariate time-series data. *J. Chemometrics*, 19(8):427–438, 2005.

[19] B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Towards detecting anomalous user behavior in online social networks. In *USENIX*, pages 223–238, Aug. 2014.

[20] X. Wang, K. Smith, and R. J. Hyndman. Characteristic-based clustering for time series data. *Data Min Knowl Discov*, 13(3):335–364, 2006.

[21] K. Yang and C. Shahabi. A PCA-based similarity measure for multivariate time series. In *Proc. 2nd ACM Int. Workshop Multimedia Databases*, pages 65–74, 2004.