Outlier Detection for Distributed Services using Multi-Frequency Patterns

Lu Xu
Sino-German Joint Software Institute
Beihang University
Beijing, China
xulu_416@buaa.edu.cn

Zhongzhi Luan ✉
Sino-German Joint Software Institute
Beihang University
Beijing, China
luan.zhongzhi@buaa.edu.cn

Carol J Fung
Computer Science Department
Virginia Commonwealth University
Richmond, USA
cfung@vcu.edu

Guang Wei
Sino-German Joint Software Institute
Beihang University
Beijing, China
guang.wei@buaa.edu.cn

Depei Quian
Beijing municipal key laboratory of network technology
Beihang University
Beijing, China
depeiq@buaa.edu.cn

Abstract—Outlier detection has been commonly used in large scale distributed systems to detect abnormalities and service failures. Traditional abnormality detection methods are mostly based on statistical analysis or simple thresholds to detect outliers. However, those solutions do not consider the periodic patterns of data which is a common feature in many systems due to users' repetitive activities on a daily basis. In our work, we investigate a method that can significantly improve the effectiveness of outliers detection through utilizing periodic patterns in the system data. To this purpose we use State-Frequency-Memory (SFM) Recurrent Neural Networks (RNN) to analyze the cyclical patterns of monitoring data and use cumulative sum (CUMSUM) to identify outliers. Our real data evaluation shows that our detection method can achieve 99% accuracy. Finally, our method also applies to many other fields, such as network traffic security monitoring, bank fraud detection, etc.

Keywords—Outlier Detection, Data Mining, Machine Learning, Service Management, Neural Network, Deep Learning

I. Introduction

Nowadays, many companies, such as Google and Yahoo, are using distributed systems to provide users with stable services. However, once an abnormality occurs in the system, users' experience may be compromised. Therefore, it is highly important to be able to detect abnormalities in a timely manner and inform the occurrences to the administrator.

Traditional anomaly detection methods are based on anomaly detection through statistical methods [16] [17] [18]. Once the observations are outside the normal range, it is considered an outlier. However, we found that data set may follow some patterns. Therefore, many outliers can be detected based on whether they conform to the expected patterns, our approach is to identify these patterns and apply them to anomaly detection.

In this work, we propose the use of the SFM-RNN to predict a sequence of data and then calculate the difference between the predicted values and the actual data. We use the Cumulative Sum method to identify anomalies. SFM-RNN was first proposed by Hao Hu and Guo-Jun Qi in 2017 [1], which is a novel recurrent architecture that allows modeling the dynamic patterns across different frequency components. Evaluations on several temporal modeling tasks demonstrate the SFM can yield competitive performances [20].

To the best of our knowledge, our work is the first to apply SFM-RNN to the field of anomaly detection for Distributed Systems. Our prediction model is based on dataset provided by Yahoo! Web-scope program which was approved to be used in non-commercial research. The goal of this work is to provide a novel approach of outlier detection in distribute systems.

The key contributions of this paper can be summarized as follows:

1) Modeling distributed system monitoring data using the SFM-RNN Model.
2) Identifying sensors that exhibit abnormal behavior using CUSUM method.
3) Validation of the proposed method on a Distributed System Benchmark testbed.

The remaining of our paper is organized as follows: Section II reviews relevant literature on different outlier detection techniques. In section III we describe our proposed model. In section IV, we briefly describe the data and the detect method. We present our results on the Yahoo! Web-scope program dataset [2] in section IV. Section V concludes this paper.

Corresponding author: Zhongzhi Luan

978-3-903176-14-0 © 2018 IFIP

II. Related works

Outlier detection lies in the category of data mining. The definition of outlier detection is stated in Angiulli and Pizzuti (2005) [3], given a dataset, the definition of exceptional data should be defined before the exceptional data could be detected.

During the past many years, many anomaly detection methods have been created. Most popular ones can be categorized into distance-based approach and model-based approach.

In the distance-based approach, given a specific distance measure, the data are categorized as outliers if the distance between the data to its nearest neighbors is higher than expected [3], [4], [5], [6].

The second approach is termed as the model-based approach. In this approach, a model is applied to describe the data. Any data that cannot be described will be deemed as outliers [7].

In the model-based approach, the primary way to detect outliers is based on regression methods, including autoregressive-based methods and network-based methods. The autoregressive model mainly includes AR model [8], MA model [9], ARIMA model [10] [14], and SARIMA model [11]. The auto-regressive model not only considers the values of the observed data, but also the correlation among data. The neural network based models mainly includes RNN [12] and LSTM [13]. In this paper, we use the SFM-RNN model which is a recurrent neural network combined with different frequency components as our prediction model. The SFM model was first proposed by Hao Hu and Guo-Jun Qi in 2017 [1]. The work of L.Zhang, Charu Aggarwal [20] verified the validity of the SFM model in stock data prediction. We will compare our model with the existing model based on a set of performance metrics.

III. Outlier Detection Model

In this section we describe the SFM model and the CUMSUM method that we use for outlier detection. We will first introduce the architecture of the State Frequency Memory (SFM) recurrent neural networks (RNN). Then we apply the SFM to data prediction with historical data. Finally we use Custom Sum (CUSUM) to calculate the difference between the predicted outputs and the actual sensor data to detect outlier data.

A. Formulation of Outlier Detection Problem using SFM

Inspired by Discrete Fourier Transform (DFT), the SFM decomposes the hidden states of memory cells into multiple frequency components. Each component models a particular frequency of latent data generation pattern underlying the fluctuation of monitoring data. Then the future monitor data are predicted as a nonlinear mapping of the combination of these components in an Inverse Fourier Transform (IFT) fashion. The SFM model can simulate the pattern of monitoring data better by adding the different frequency unit. The structure of SFM is shown in Figure 1.

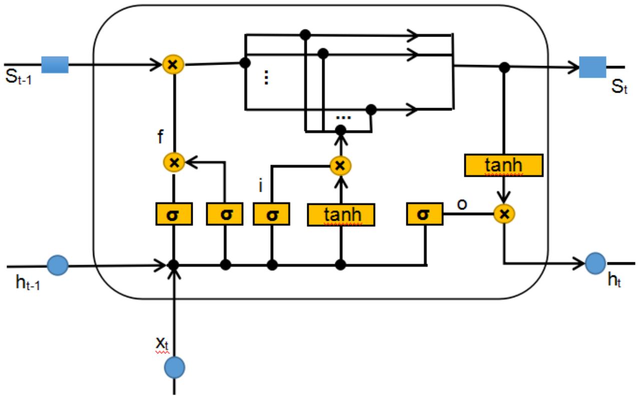


Fig. 1: The structure of SFM

1) Updating State-Frequency Memory: The input of SFM model is a sequence $X_{1:T} = [x_1, x_2, ... , x_T]$ which means T observations of N-dimensional space, i.e., $x_t \in R^N$ for t = 1, ... , T. SFM decompose the memory hidden states into a set of frequency components, denoted by $\{w_k = \frac{2\pi k}{K} | k = 1, 2, ..., K\}$. For this purpose, we define the hidden state of SFM model as a matrix $S_t \in C^{D*K}$ at each time t. The rows of the matrix mean D dimensional states and the columns mean K frequencies. This forms a joint state-frequency decomposition to model the temporal context of the input sequence across different states and frequencies.

The SFM matrix of a memory cell is updated by combining the past memory and the new input. the updating rule for the state-frequency matrix is formulated below:

$$S_t = f_t \cdot S_{t-1} + (i_t \cdot c_t^{\sim}) \cdot \begin{bmatrix} e^{j\omega_1 t} \\ e^{j\omega_2 t} \\ ... \\ e^{j\omega_k t} \end{bmatrix}^T \quad (1)$$

where the operator · is an element-wise multiplication and j =$\sqrt{-1}$ and $[e^{j\omega_1 t}, e^{j\omega_2 t}, ..., e^{j\omega_k t}]$ are the Fourier basis of K frequency components of the state sequence. $f_t \in R^{D*K}$ is a joint state frequency forget gate matrix to control how much information on various states and frequencies should be kept in the memory cell. $i_t \in R^D$ is the input gate similar to the traditional LSTM model. The input modulation $c_t^{\sim}$ aggregates the current inputs fed into the memory cell at time t.

Based on the Euler formula and Complex Fourier Transform, the updating rule can be separated into the real and virtual parts of the state-frequency matrix S_t:

$$ReS_t = f_t \cdot ReS_{t-1} + (i_t \cdot c_t^{\sim})[cos_1 t, ..., cos_K t] \quad (2)$$

$$ImS_t = f_t \cdot ImS_{t-1} + (i_t \cdot c_t^{\sim})[sin_1 t, ..., sin_K t] \quad (3)$$

The polar coordinate representation of a complex number is represented by its magnitude and phase angle which can be seen below:

$$A_t = |S_t| = \sqrt{(ReS_t)^2 + (ImS_t)^2} \in C^{D*K} \tag{4}$$

$$\angle S_t = arctan(\frac{ImS_t}{ReS_t}) \in [-\frac{\pi}{2}, \frac{\pi}{2}]^{D*K} \tag{5}$$

where arctan(·) is an element-wise inverse tangent function; A_t is the amplitude of S_t; and $\angle S_t$ is the phase angle of S_t. The amplitude A_t is used to obtain the output hidden state $h_t \in R^D$. We ignore the phase $\angle S_t$ as we found it has no significant impact on the results in our experiments but incurs extra computational and memory overheads.

2) The Joint State-Frequency Forget Gate: The hidden state S_t is a matrix containing state information and frequency information. Therefore we define a joint state frequency forget gate matrix $f_t \in R^{D*K}$ which can be decomposed into two matrices : f_t^{std} and f_t^{ste}. They can be formulated as follows:

$$f_t^{ste} = \sigma(W_{std}x_t + U_{ste}h_{t-1} + b_{ste}) \in R^D \tag{6}$$

$$f_t^{fre} = \sigma(W_{fre}x_t + U_{fre}h_{t-1} + b_{fre}) \in R^K \tag{7}$$

$$f_t = f_t^{ste} \times f_t^{fre} \in R^{D*K} \tag{8}$$

where × is an outer product; f_t^{ste} is a state forget gate and f_t^{fre} is a frequency forget gate vector. Through the outer product, f_t can be seen as a composition gate over different states and frequencies to control the information flowing into the memory cell.

3) Gates and Modulations: The input gate i_t and the input modulation $c_t^{\sim}$ can be defined as below:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{9}$$

$$c_t^{\sim} = \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{10}$$

The input gate i_t controls how much information of the input modulation $c_t^{\sim}$ should be kept in the memory cell. $c_t^{\sim}$ is the tanh activation of linear transformation combinations of the current observation x_t and the output hidden state h_{t-1}. $i_t \cdot c_t^{\sim}$ represents the value of the Fourier base.

4) Multi-Frequency Outputs and Modulations: As we discussed earlier, the amplitude A_t of hidden state S_t is given, we should also get an output state h_t. An output vector $h_t \in R^M$ can be calculated as follows:

$$c_t = \tanh(A_t u_a + b_a) \tag{11}$$

$$o_t = \sigma(U_o c_t + W_o h_{t-1} + V_o x_t + b_o) \tag{12}$$

$$h_t = o_t \cdot c_t \tag{13}$$

where $A_t \in R^{D*K}$ is the state amplitude; c_t is the hidden unit of SFM, o_t is controlling how much information could be allowed to output from the memory cell; h_t is the output vector of the SFM cell.

After training, the unique complex-value memory states of SFM model can learn multiple patterns with different frequencies.

B. Cumulative Sum (CUSUM)

Based on the prediction of the SFM, we then calculate the difference between the predicted outputs and the observed output. In this paper, we use CUSUM [15] to calculates all cumulative sums of the positive and negative changes to detect small deviations over time. The formula of calculating the CUSUMs are shown below :

$$x_0 = 0 \tag{14}$$

$$SH_i = MAX(0, x_i - Target - b) \tag{15}$$

$$SL_i = MIN(0, x_i - Target + b) \tag{16}$$

where SH means the high cumulative sum, SL is the low cumulative sum and x_i is the difference between the predicted value and the actual value at time i. Target and b are pre-defined safety limit with the allowable slack. If x_i is higher than $Target + b$ or x_i is lower than $Target - b$, then we classify the data to be an outlier.

IV. Evaluation results and analysis

We have implemented SFM algorithm and compare the performance with three other methods. The first method is LSTM model which is combined with CUSUM to detect anomies. The other two methods are LOF model and Isolation model(both are Distance-based anomaly detection method). We use the data from the A4Benchmark of the Yahoo Webscope anomaly detection dataset [2] in our experiment.

A. Experimental setup

1) Models: In this subsection, we briefly overview the three other methods we will use as a baseline to compare with in our experiments.

- LSTM model: The LSTM model has made a remarkable achievement in the fields of time series prediction. In this experiment, we use the LSTM model to perform data predictions, and then use the difference between the actual data and the predicted data to determine whether the actual data is an outlier or not.
- Isolation forest: The strategy of Isolation forest to perform outlier detection is to use a random hyperplane to split a data space. It is straight forward to see that the clusters with high density can be cut many more times before the loop stops, while anomaly data take fewer rounds to be isolated.
- LOF model: The Local Outlier Factor (LOF) algorithm reflects the abnormality of a sample by calculating a numeric score. The meaning of this score can be explained as a ratio that represents the average density of around a sample point compared to the density on the sample point.

2) Metrics: We will use a few metrics to compare the performance of all methods. The first metric is the accuracy rate. Accuracy is one of the most important criteria in a typical classification problem. The second

metric is the ROC value. We will plot the ROC curves for all methods. After comparing the above evaluation metrics, we will summarize the performance of all three methods.

3) Dataset: In this experiment, We will use the data from the A4Benchmark of the Yahoo Webscope anomaly detection dataset [2]. We pick two groups of correlated series to do our experiment, after correlation analysis of the data in the collection(Group 1: TS6, TS16, TS34, Group2: TS4, TS13, TS90).

4) Experimental Design: In this experiment, we use the first 90% of the dataset as the training set and the last 10% of the dataset as the validation set. Our experiments were run on top of a Windows server with the configuration of Core I7-870 CPU, 4G RAM, and WINDOWS8.

B. Result value and analysis

In this section, we present the experimental results based on the metrics we have proposed in Section IV-A.

1) Effectiveness of prediction: Since our idea of anomaly detection is based on regression to make predictions, we first look at the prediction effect of the LSTM model and the SFM model. In this experiment, the parameters of the LSTM model were set to be num_{units}=128, and the optimization algorithm used was the Adam algorithm [24]. The parameters of the SFM model are: $hidden_{dim} = 50$, $fre_{dim} = 4$, and the parameter optimization algorithm is the SGD algorithm [25]. The MSE comparison of these two model are shown in Table1. We can see that the SFM model predicts better than the LSTM model.

TABLE I: the MSE of LSTM and SFM model

	LSTM	SFM
MSE(Group 1)	193687.72	22307.32
MSE(Group 2)	663536.44	32312.13

2) Detection accuracy: After obtaining the corresponding predicted value, the difference between the actual value and the predicted value is calculated. After that the CUSUM algorithm is used to classify the outliers. In this experiment, we consider there is an abnormality at the moment when SHi or SLi is not 0. Fig.2 shows the detection accuracy of different methods. From Fig.2, we can see that the SFM-CUSUM model can achieve the highest accuracy.

3) ROC comparison: The experimental results of ROC curves are shown in Figure 3. We can see that the SFM achieves better results with lower false positive rates and higher true positive rates. Although Isolation forest method can achieve higher true positive rates, it is still worse than SFM model.

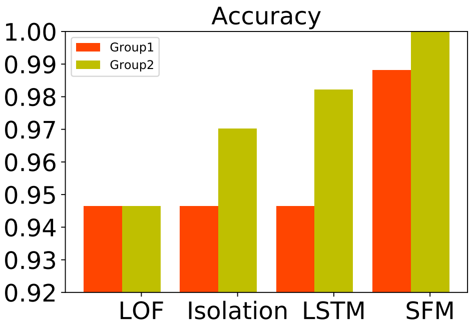
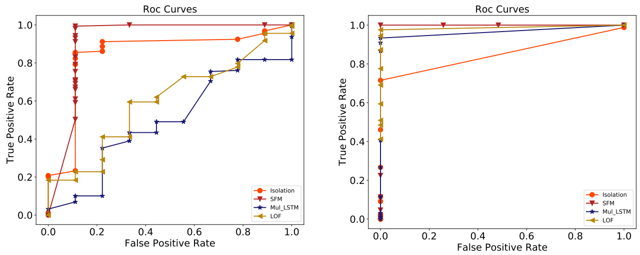


Fig. 2: the accuracy of different method



(a) group1

(b) group2

Fig. 3: Roc curves of different mothods

Through all the experiments above, we can see that the SFM-CUSUM model outperforms the other two methods in the anomaly detection field. The accuracy of SFM-RNN algorithm is higher than the traditional methods. Through our experiments, we verified the feasibility of anomaly detection based on regression, and introduced the SFM-RNN model into the anomaly detection field.

V. CONCLUSION

In this paper, we present SFM-CUSUM, a novel unsupervised neural network model combined with Fourier transform for the purpose of detecting outliners in distributed systems. More specifically, we introduce the SFM model into the anomaly detection field and use CUSUM to calculate the difference between the predicted value and the observed value. If the difference is too large, the data point is classified to be an abnormal value. In this way, we can better detect local anomalies. Our results show that the SFM-RNN achieves high detection accuracy and outperforms several other algorithms including LSTM, Isolation, and LOF methods.

Acknowledgment

This research is supported by the National Key R&D Program (Grant No. 2018YFB0204003)

References

[1] Hao Hu ,Guo-Jun Qi . 2017. State-Frequency Memory Recurrent Neural Networks. Proceedings of the 34 th International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017

[2] Y.Webscope,"S5 - a labeled anomaly detection dataset" https://webscope.sandbox.yahoo.com/catalog.php?datatype=s.

[3] Angiulli, F., Pizzuti, C., 2005. Outlier mining in large high-dimensional data sets. IEEE Trans. Knowl. Data Eng. 17 (2), 203- 215.

[4] Angiulli, F., Basta, S., Pizzuti, C., 2006. Distance-based detection and prediction of outliers. IEEE Trans. Knowl. Data Eng. 18 (2), 145-160.

[5] Angiulli, F., Basta, S., Lodi, S., Sartori, C., 2013. Distributed strategies for mining outliers in large data sets. IEEE Trans. Knowl. Data Eng. 25 (7), 1520- 1532

[6] Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation forest."Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on. IEEE, 2008.

[7] Albanese, A., Pal, S.K., Petrosino, A., 2014. Rough sets, kernel set, and spatiotemporal outlier detection. IEEE Trans. Knowl. Data Eng. 26 (1), 194–207.

[8] M. C. Hau,H. Tong,A practical method for outlier detection in auto-regressive time series modeling.Stochastic Hydrol. Hydraul. 3 (1989) 241-260.

[9] Enders, Walter (2004). "Stationary Time-Series Models". Applied Econometric Time Series (Second ed.). New York: Wiley. pp. 48-107. ISBN 0-471-45173-8.

[10] Munim, Ziaul Haque; Schramm, Hans-Joachim (2017)." Forecasting container shipping freight rates for the Far East – Northern Europe trade lane". Maritime Economics and Logistics. 19 (1): 106–125. doi:10.1057/s41278-016-0051-7

[11] Hyndman, Rob J; Athanasopoulos, George. "8.9 Seasonal ARIMA models". Forecasting: principles and practice. oTexts. Retrieved 19 May 2015.

[12] TAYLOR, R. N., AND OSTERWEIL, L. J. 1980. Anomaly detection in concurrent software by static data flow analysis. IEEE

[13] J. Goh, S. Adepu, M. Tan, and Z. S. Lee, Anomaly detection in cyber physical systems using recurrent neural networks, in High Assurance Systems Engineering (HASE), 2017 IEEE 18th International Symposium on. IEEE, 2017, pp. 140- 145.

[14] Qin Yu,Lyu Jibin,and Lirui Jiang, An Improved ARIMA-Based Traffic Anomaly Detection Algorithm for Wireless Sensor Networks Hindawi Publishing Corporation International Journal of Distributed Sensor Networks ,Volume 2016, Article ID 9653230

[15] Mishra, S., Vanli, O. A., and Park, C (2015). "A Multivariate Cumulative Sum Method for Continuous Damage Monitoring with Lamb-wave Sensors", International Journal of Prognostics and Health Management, ISSN 2153-2648

[16] T. Johnson, I. Kwok, and R. Ng. Fast Computation of 2-dimensional Depth Contours. ACM KDD Conference, 1998.

[17] J. Laurikkala, M. Juholal, and E. Kentala. Informal Identification of Outliers in Medical Data. Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology, pp. 20- 24, 2000.

[18] J. Laurikkala, M. Juholal, and E. Kentala. Informal Identification of Outliers in Medical Data. Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology, pp. 20- 24, 2000.

[19] P. Rousseeuw and A. Leroy. Robust Regression and Outlier Detection. Wiley, 2003.

[20] L.Zhang ,Charu Aggarwal and Guo-Jun Qi.Stock Price Prediction via Discovering Multi-Frequency Trading Patterns .KDD '17, August 13-17, 2017.

[21] S.Zhang, C.Feng and L.Zhong. PSOM: Periodic Self-Organizing Maps for Unsupervised Anomaly Detection in Periodic Time Series. ISQoS, June, 2017.

[22] A time series library in TensorFlow. https://www.tensorflow.org/api_docs/python/tf/contrib/timeseries. July, 2018.

[23] Machine Learning Models in Python with scikit-learn. http://scikit-learn.org/stable/. July, 2018.

[24] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 2014.

[25] Zhang, Sixin, Anna E. Choromanska, and Yann LeCun. "Deep learning with elastic averaging SGD." Advances in Neural Information Processing Systems. 2015.