무선 센서 네트워크에서의 이상 징후 감지를 위한 공동 지수 평활법 및 추세 기반 주성분 분석

Thien-Binh Dang *, 양희규*, Manh-Hung Tran*, Duc-Tai Le*, 김문성**, 추현승*
*성균관대학교 소프트웨어대학
**서울신학대학교 교양학부
e-mail :{dtbinh, huigyu, hungtm, ldtai, choo} @skku.edu
moonseong@stu.ac.kr

Joint Exponential Smoothing and Trend-based Principal Component Analysis for Anomaly Detection in Wireless Sensor Networks

Thien-Binh Dang*, Manh-Hung Tran*, Duc-Tai Le*, Moonseong Kim**, Hyunseung Choo*
*College of Software, Sungkyunkwan University
**Dept. of Liberal Arts, Seoul Theological University

Abstract

Principal Component Analysis (PCA) is a powerful technique in data analysis and widely used to detect anomalies in Wireless Sensor Networks. However, the performance of conventional PCA is not high on time-series data collected by sensors. In this paper, we propose a Joint Exponential Smoothing and Trend-based Principal Component Analysis (JES-TBPCA) for Anomaly Detection which is based on conventional PCA. Experimental results on a real dataset show a remarkably higher performance of JES-TBPCA comparing to conventional PCA model in detection of stuck-at and offset anomalies.

1. Introduction

Wireless Sensor Networks (WSNs) have been being used in many critical applications, ranging from civilian fields (e.g. smart-homes) to military (e.g. battlefield surveillance systems) or industry (e.g. industrial control systems) [1]. The constraints on size and cost of a sensor make it an exiguous resource device, such as weak computational speed, small memory capacity, limited energy and restricted communication bandwidth [2]. Therefore, the WSNs are highly vulnerable to random faults and cyber-attacks which unavoidably cause anomalies. The anomalous data collected from sensors not only provides wrong information about phenomenon but also leads to improper decisions. In order to keep sensory data accurate and reliable, it is necessary to develop efficient anomaly detection algorithms. As stated by the literature [3], anomaly detection techniques have been generally recognized as effective methods against these anomalies. Principal Component Analysis (PCA) is a powerful tool to analyze multivariate data collected from WSN networks. There are many works using PCA for anomalous data detection [4], [5]. However, conventional PCA is not sensitive enough to capture anomalies whose data has small different comparing to normal data or contains significant amount of noises. In this paper, we propose a Joint Exponential Smoothing and Trend-based Principal Component Analysis (JES-TBPCA) for Anomaly Dectection focus on improving the sensitiveness of PCA by splitting data into approximately monotonic trends and using exponential smoothing technique to reduce noises.

2. Joint Exponential Smoothing and Trend-based PCA

The key idea is that noises of training data are reduced by using exponential smoothing technique and then the smoothed data is split into smaller sets whose overall trend is approximately monotonic increasing or approximately monotonic decreasing. The proposed scheme has two stages: training stage and testing stage. In training stage, considering the WSN has M sensors and the data matrix $\boldsymbol{X} = \{\boldsymbol{x}^1, \boldsymbol{x}^2, \dots, \boldsymbol{x}^M\}$ where column i contains N data samples $\boldsymbol{x}^i = \{x_0^i, x_1^i, \dots, x_{N-1}^i\}$ of sensor i collected from normal operation. The smoothed data matrix $\boldsymbol{Y} = \{\boldsymbol{y^1}, \boldsymbol{y^2}, \ldots, \boldsymbol{y^M}\}$ where:

$$y_0^i = x_0^i,$$
$$y_j^i = \alpha x_j^i + (1-\alpha).$$

Where α is the smoothing factor and $0 \le \alpha \le 1$. Then the smoothed data is manually split into trends $\mathbf{1 \ldots n}$ corresponding to time intervals ($t_0 \rightarrow t_1$), ($t_1 \rightarrow t_2$), etc. The minimal number of samples of each trend is predefined which ensures enough input data for training. PCA is then applied on these trends separately to compute Square Prediction Error (SPE) limits. In testing stage, first the data is split into different sets corresponding to time intervals which are established in training stage. PCA is then applied on these sets of data to calculate SPE values for each sample. The model checks whether a sample s is abnormal or normal by comparing its SPE values to corresponding SPE limits whose time interval contains sampling time of s. If the SPE values of s is less or equal to the SPE limits, the sample s is normal otherwise s is detected as anomalous data.

The advantage of PCA is that it can capture the correlation of by projecting sensors' data into a lower dimension space which still preserves maximum variance of the original data in minimum number of dimensions. In order to apply PCA, the data matrix is normalized to zero-mean and scaled to unit variance. Let $\boldsymbol{Y_s}$ is the normalized data, $\boldsymbol{Y_s}$ can be expressed as:

$$\boldsymbol{Y_s} = (\boldsymbol{Y} - \boldsymbol{\bar{Y}})D^{\frac{-1}{2}}$$

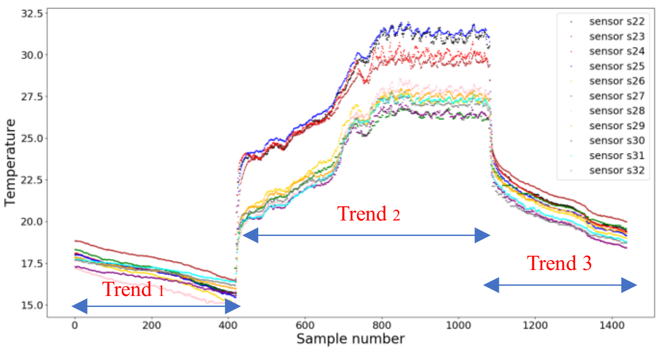
where $\boldsymbol{\bar{Y}} = 1(\frac{1}{N}1^T\boldsymbol{X})$ and $D = \frac{1}{N-1}[(\boldsymbol{X}-\boldsymbol{\bar{X}})^T(\boldsymbol{X}-\boldsymbol{\bar{X}})] \circ I_M$ where $\circ$ denoting the Hadamard multiplication and I_M is the identity matrix. Then, the co-variance matrix $\boldsymbol{R}$ of matrix $\boldsymbol{Y_s}$ is constructed by $\boldsymbol{R} = \boldsymbol{Y_s^T Y_s}$ where $\boldsymbol{Y_s^T}$ is the transpose matrix of matrix $\boldsymbol{Y_s}$. In next step, singular value decomposition (SVD) is performed on $\boldsymbol{R}$ as $\boldsymbol{R} = \boldsymbol{V \Lambda V^T}$ where $\boldsymbol{\Lambda}$ is the diagonal matrix containing M eigenvalues of matrix $\boldsymbol{R}$ in descending order ($\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_M \ge 0$) and matrix $\boldsymbol{V}$ is the collection of M eigenvectors of $\boldsymbol{R}$. The n, SPEs of data samples are calculated for anomalous detection. SPE measures how the testing data fits the model which is constructed in training phase. In more detail, SPE indicates the squared perpendicular distance from the sample and its projection in principal component space which is formed by l eigenvetors in the loading matrix $\boldsymbol{\hat{P}}$. SPE statistics can be calculated by the following equation: $SPE = \|(\boldsymbol{I} - \boldsymbol{\hat{P}\hat{P}^T})\boldsymbol{Y_s}\|$. The data is considered normal if its $SPE \le \delta^2$. The confident limit δ^2 is expressed as follows:

$$\delta^2 = \theta_1\left[\frac{C_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0-1)}{\theta_1^2}\right]^{\frac{1}{h_0}}$$
$$h_0 = \frac{2\theta_1\theta_3}{\theta_2^2},$$
$$\theta_i = \sum_{j=l+1}^{L} \lambda_j^i,$$

where λ_j is the eigenvalue associated with j_{th} the eigenvector, C_α is the standard normal deviation corresponding to the confident level of standard normal distribution.

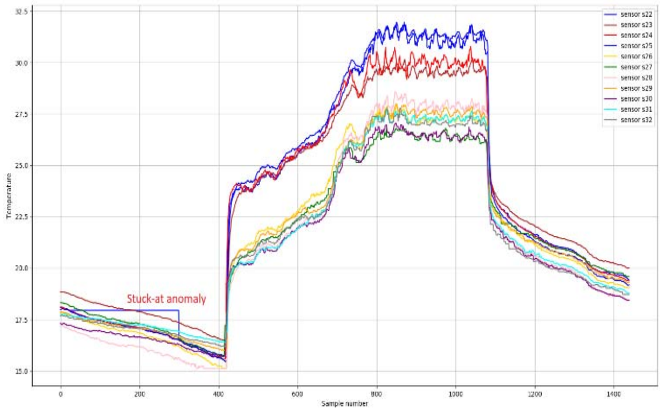
3. Numerical Experiments and Evaluation

In this research, we choose a real WSN from Intel Berkeley Research lab (IBRL) to evaluate the efficiency of the proposed model. We choose temperature measurements from eleven sensors whose IDs are 22, 23, …, 32 for experiment. In our experiment, temperature readings from these seven sensors are re-sampled every minute so a total of 1440 samples are taken in one day. We use 1400 samples of March 1st for training and we consider this training data is normal. For testing, we inject anomalous data into this normal data and evaluation the performance on conventional PCA and proposed scheme. In training phase, we split training data into three trends where trend 1 is from sample 1 to sample 416, trend 2 is from sample 417 to 1079 and trend 3 is from sample 1080 to 1440 as show in the Figure 1. Data of these three trends is processed separately to extract SPE limits for each trend.

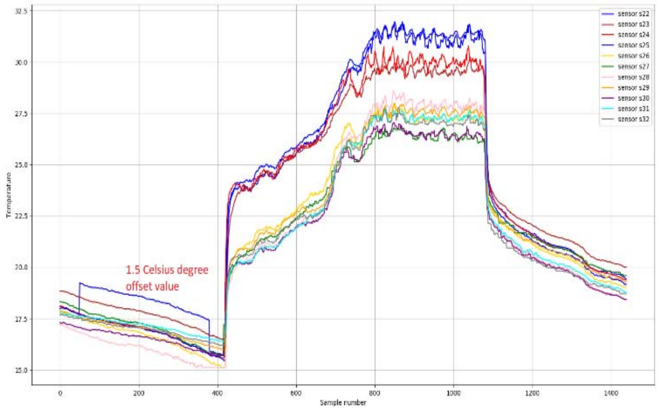


(Figure 1) Training data and extracted trends.

In testing phase, we do experiment on the stuck-at and offset anomalies. While stuck-at anomaly shows a series of data values with little or no variation for a period of time longer than expected, the offset anomaly is defined as a sudden deviation from the normal data with a constant amount. In the first experiment, we inject stuck-at to the normal data for testing as shown in the Figure 2.

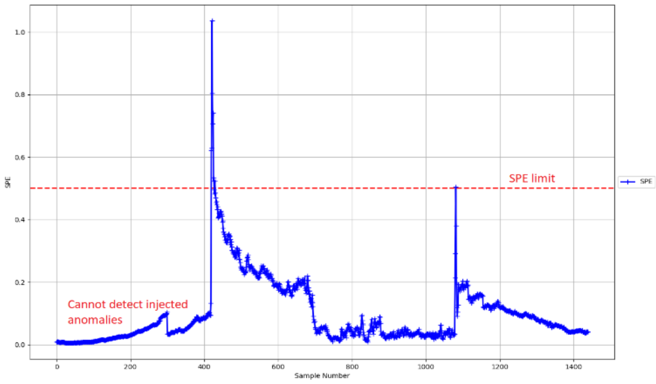


(Figure 2). Stuck-at anomaly in sensor 22.

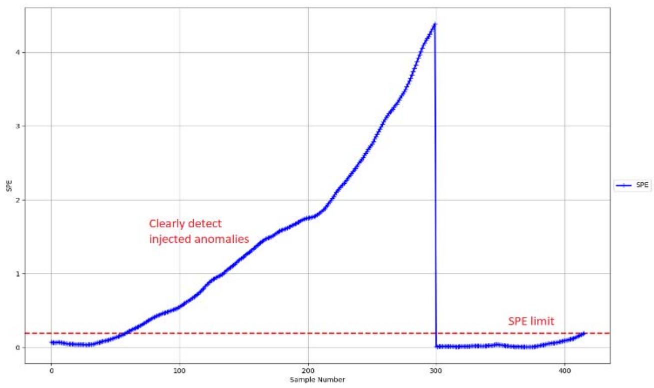


(Figure 5). Offset anomaly in sensor 22.

The experiment results in the Figure 3 and Figure 4 shown that the conventional PCA cannot detect the injected stuck-at anomaly but the JEM-TBPCA can detect clearly which is depicted by the number of SPE points higher than the SPE limit (the red dot line).



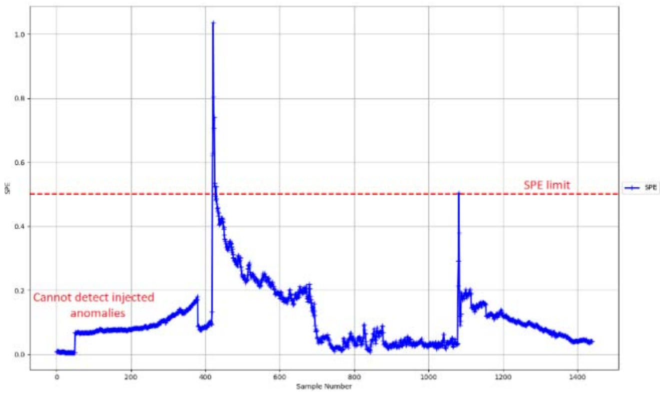
(Figure 3). Detection result of conventional PCA on stuck-at anomaly.



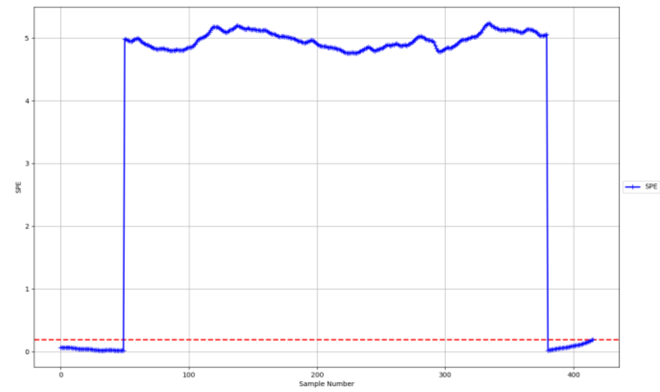
(Figure 4). Detection result of JES-TBPCA on stuck-at anomaly.

In the second experiment, we inject offset anomaly into the normal data for testing. The offset value which is the distance between the normal value and the abnormal value is set to 1.5 Celsius degree as shown in the Figure 5.

The experiment results in the Figure 6 and Figure 7 once again shown that the conventional PCA still cannot detect injected offset anomaly. On the contrary, JEM-TBPCA detects well in this experiment.



(Figure 6). Detection result of conventional PCA on offset anomaly.



(Figure 7). Detection result of JES-TBPCA on offset anomaly.

4. Conclusion

In this work, we proposed a joint exponential smoothing and trend-based PCA for anomaly detection in WSN. The experiment results show that the proposed scheme is more sensitive with stuck-at and offset anomalies and outperforms the conventional PCA.

Acknowledgement

본 논문은 과학기술정보통신부 및 정보통신기획평가원의 Grand ICT 연구센터지원사업 (IITP-2019-2015-0-00742), 과학기술정보통신부 및 정보통신기획평가원의 글로벌핵심인재양성지원사업(2019-0-01579)과 2019 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2019-0-00421, 인공지능대학원지원)의 연구결과로 수행되었음.

References

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," Computer Networks, vol. 38, no. 4, pp. 393 – 422, 2002.
[2] D. W Carman, P. S Kruus, and B. Matt, "Constraints and approaches for distributed sensor network security (final)," 10 2000.
[3] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Anomaly detection in wireless sensor networks," IEEE Wireless Communications, vol. 15, no. 4, pp. 34–40, Aug 2008.
[4] M. Livani and M. Abadi, "Distributed pca-based anomaly detection in wireless sensor networks," in Internet Technology and Secured Transactions (ICITST), 2010 International Conference.
[5] X. L. Zhang, F. Zhang, J. Yuan, J. lan Weng, and W. h. Zhang, "Sensor fault diagnosis and location for small and medium-scale wireless sensor networks," in 2010 Sixth International Conference on Natural Computation.