무선 센서 네트워크에서 노이즈 감지를 위한 DWT-PCA 조합

당 띠엔 빈*, 김문성**, 추현승*
*성균관대학교 전자전기컴퓨터공학과
**서울신학대학교 교양학부
e-mail: *{binh01t2, choo}@skku.edu, **moonseong@stu.ac.kr

DWT-PCA Combination for Noise Detection in Wireless Sensor Networks

Thien-Binh Dang*, Duc-Tai Le*, Moonseong Kim**, Hyunseung Choo*
* Dept. of Electrical and Computer Engineering, Sungkyunkwan University
**Dept. of Liberal Arts, Seoul Theological University

Abstract

Discrete Wavelet Transform (DWT) is an effective technique that is commonly used for detecting noise in collected data of an individual sensor. In addition, the detection accuracy can be significant improved by exploiting the correlation in the data of neighboring sensors of Wireless Sensor Networks (WSNs). Principal component analysis is the powerful technique to analyze the correlation in the multivariate data. In this paper, we propose a DWT-PCA combination scheme for noise detection (DWT-PCA-ND). Experimental results on a real dataset show a remarkably higher performance of DWT-PCA-ND comparing to conventional PCA scheme in detection of noise that is a popular anomaly in collected data of WSN.

1. Introduction

In Wireless Sensor Networks (WSN), the limitation on size and cost of a sensor make it a weak device, such as low computational speed, small memory, limited energy and restricted communication bandwidth [1]. Thus, the sensory data of WSN often contains noise. The noisy data collected from sensors can leads data analysts to improper decisions. To ensure accuracy and reliability of the sensory data, an efficient noise detection algorithm needs to be developed. Discrete Wavelet Transform is a powerful technique to detect noise. However, it works on the data of an individual sensor which is not suitable to analyze the multivariate data of WNSs. Principal Component Analysis (PCA) is an effective technique to exploit the correlation in multivariate data obtained from WSNs. It was used in many existing works to detect noisy data such as [1], [2], [3], [4], [5]. However, the conventional PCA is not sensitive enough to recognize small noise whose data has small differences to the normal data. In this paper, we propose a DWT-PCA combination scheme for noise detection (DWT-PCA-ND) that focuses on improving the sensitiveness of conventional PCA.

2. DWT-PCA Combination Scheme for Noise Detection

The key idea is that the data of an individual sensor is first analyzed by DWT. PCA is then applied to the outputs of DWT of sensors for noise detection. The proposed scheme has two phases: training phase and testing phase. In training phase, considering 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 its normal operation; M is the number of sensors. DWT is applied on $\boldsymbol{x}^i$ to get L detail coefficients vectors $\{\boldsymbol{D}_1^i, \boldsymbol{D}_2^i, \dots, \boldsymbol{D}_L^i\}$ where L is the decomposition level. The decomposition vectors of level i of M sensors are collected into a matrix $\boldsymbol{D}_i$ to finally obtain L matrixes $\boldsymbol{D}_1, \boldsymbol{D}_2, \dots, \boldsymbol{D}_L$. Then, PCA is applied on these matrices independently to compute Square Prediction Error (SPE) control limit. In testing phase, the SPE statistics of incoming data is first calculated. The data is normal if its SPE value is less or equal to the

SPE control limit that is calculated form the training phase, otherwise the data is detected as noisy data.

The advantage of PCA is that it captures the correlation of sensory data by projecting sensors' data into a lower dimension space that still preserves maximum variance of the original data in a minimum number of dimensions. In order to apply PCA, the data matrix $\boldsymbol{D}_i$ is first normalized to zero-mean and scaled to unit variance. Then, the co-variance matrix $\boldsymbol{C}$ of matrix $\boldsymbol{D}_i$ is constructed by $\boldsymbol{C} = \boldsymbol{D}_i^T \boldsymbol{D}_i$ where $\boldsymbol{D}_i^T$ is the transpose matrix of matrix $\boldsymbol{D}_i$. In the next step, Singular Value Decomposition (SVD) is performed on $\boldsymbol{C}$ as $\boldsymbol{C} = \boldsymbol{V \Lambda V}^T$ where $\boldsymbol{\Lambda}$ is the diagonal matrix containing M eigenvalues of matrix $\boldsymbol{C}$ in descending order $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M \geq 0)$ and matrix $\boldsymbol{V}$ is the collection of M eigenvectors of $\boldsymbol{C}$. SPEs of the data are calculated for noise detection based on the loading matrix $\widehat{\boldsymbol{P}}$ which is formed by l smallest eigenvectors. SPE statistics can be calculated by the following equation: $SPE = \left\|(\boldsymbol{I} - \widehat{\boldsymbol{P}}\widehat{\boldsymbol{P}}^T)\boldsymbol{D}_i\right\|$. The data is considered normal if its $SPE \leq \delta^2$, where δ^2 is expressed as follows:

$$\delta^2 = \theta_1 \left[\frac{C_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}}$$

$$h_0 = \frac{2\theta_1 \theta_3}{\theta_2^2},$$

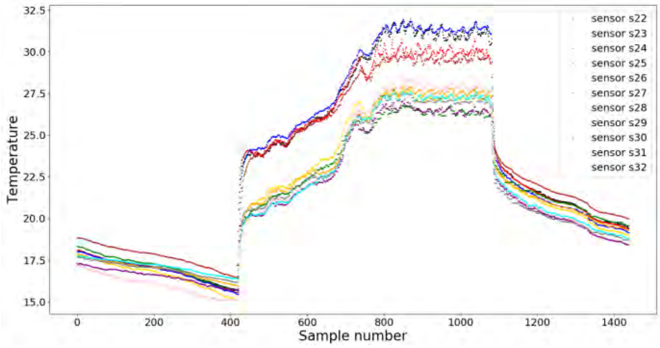
$$\theta_i = \sum_{j=l+1}^{L} \lambda_j^i,$$

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 Gaussian distribution.

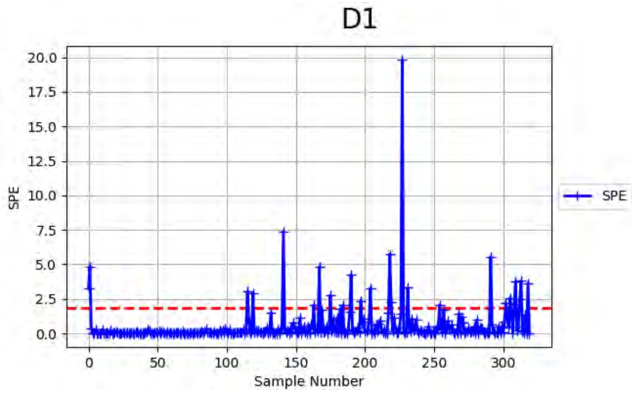
3. Performance Evaluation

In this research, a real WSN from Intel Berkeley Research lab (IBRL) is used for evaluating the efficiency of DWT-PCA-ND. Without loss of generality, we only choose temperature measurements from eleven sensors whose IDs are 22, 23, …, 32 for experiment. In our experiment, temperature readings from these eleven 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 (Figure 1), the decomposition level L is set to 2 and the training data is considered as normal data. Because $L = 2$ so we have $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ for noise analysis. The SPE control limit of $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ in the training data is shown in Figure 2 and Figure 3.

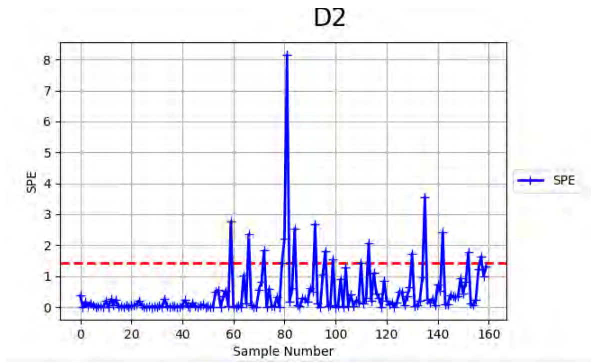
For testing, we generate and inject noise into this normal data (i.e. an example of artificial injected noise is shown in Figure 4) and comparing the performance of the conventional PCA and DWT-PCA-ND.



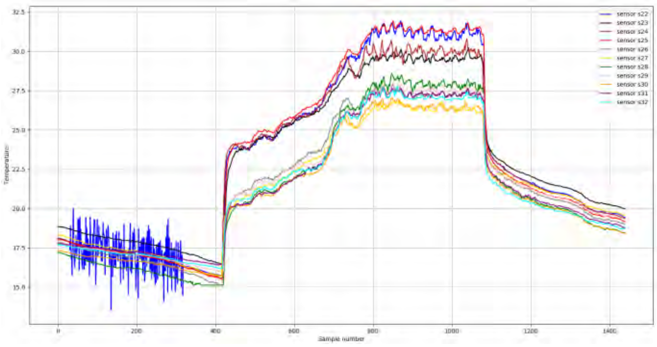
(Figure 1) Training data.



(Figure 2) SPE control limit of $\boldsymbol{D}_1$.

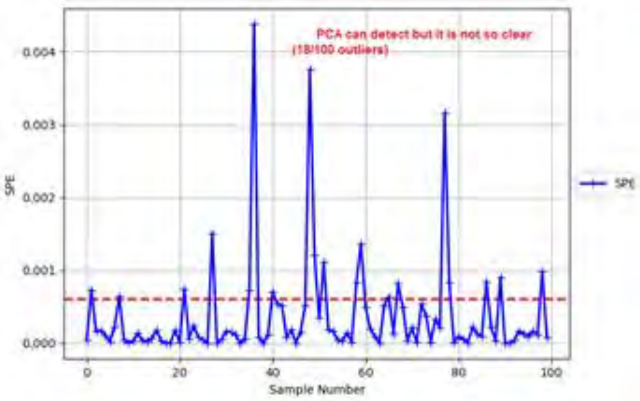


(Figure 3) SPE control limit of $\boldsymbol{D}_2$.

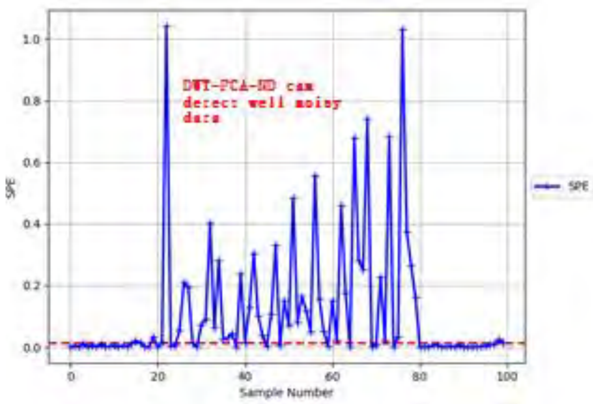


(Figure 4). Noise in sensor 22.

The experiment result in Figure 5 showed that the conventional PCA can detect a few noises represented by a small number of SPE points that is higher than the SPE control limit (the red dot line). In the case DWT-PCA-ND, there are many SPE points that is higher than the SPE control limit (the red dot line) depicted in Figure 6. This stated that DWT-PCA-ND can detect noisy data with high accuracy.



(Figure 5). Detection result of conventional PCA with noise.



(Figure 6). Detection result of DWT-PCA-ND with noise.

4. Conclusion

In this work, we propose a DWT-PCA combination scheme for detecting noise in the sensory data of WSNs. This combination adopts the advantages of both DWT and PCA in analyzing multivariate data. Consequently, the experiment results stated that DWT-PCA-ND is efficient in detecting noise and outperforms the conventional PCA.

ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 보통신기획평가원의 ICT명품인재양성 사업(IITP-2020-2051-001), Grand ICT연구센터지원사업(IITP-2020-2015-0-00742)과 2020년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2020R1A2C2008447).

References

[1] D. W Carman, P. S Kruus, and B. Matt, “Constraints and approaches for distributed sensor network security (final),” 10 2000.
[2] X. Miao, Y. Liu, H. Zhao, and C. Li, “Distributed online one-class support vector machine for anomaly detection over networks,” IEEE transactions on cybernetics, vol. 49, no. 4, pp. 1475–1488, 2018
[3] Y. Chen and S. Li, “A lightweight anomaly detection method based on svdd for wireless sensor networks,” Wireless Personal Communications, vol. 105, no. 4, pp. 1235–1256, 2019.
[4] Z. Chen, C. K. Yeo, B. S. Lee, and C. T. Lau, “Detection of network anomalies using improved-mspca with sketches,” Computers & Security, vol. 65, pp. 314–328, 2017.
[5] T. Yu, X. Wang, and A. Shami, “A novel r-pca based multivariate faulttolerant data aggregation algorithm in wsns,” in 2016 IEEE International Conference on Communications (ICC). IEEE, 2016, pp. 1–5.