Anomaly Detection from Multivariate Time-Series with Sparse Representation

Naoya Takeishi
Department of Aeronautics and Astronautics
The University of Tokyo
Tokyo, Japan
Email: takeishi@space.rcast.u-tokyo.ac.jp

Takehisa Yairi
Research Center for Advanced Science and Technology
The University of Tokyo
Tokyo, Japan
Email: yairi@space.rcast.u-tokyo.ac.jp

***Abstract*—Anomaly detection from sensor data is an important data mining application for efficient and secure operation of complicated systems. In this study, we propose a novel anomaly detection method for multivariate time-series to capture relationships of variables and time-domain correlations simultaneously, without assuming any generative models of signals. The supposed framework in this study is a semi-supervised anomaly detection where we seek unusual parts of test data compared with reference data. The proposed method is based on feature extraction with sparse representation and relationship learning with dimensionality reduction. Our idea comes from the similarity between a sparse feature matrix extracted from multivariate time-series and a term-document matrix. We conducted experiments with synthetic and simulated data, and confirmed that the proposed method successfully detected anomalies in multivariate time-series signals. Especially, it demonstrated superior performance with anomalies in which only relationships of time-series patterns are changed from reference data (multivariate anomalies).**

I. Introduction

For efficient and secure operation of complicated modern systems such as plants, cars and artificial satellites, detecting faults or anomalies from their sensor data is a very important data mining application. Since most modern systems are equipped with many sensors and generate signals sequentially at some time intervals, we have to detect anomalous behaviors from those multivariate time-series. However, anomaly detection from multivariate time-series remains controversial despite the large amount of literature. There are mainly two types of anomalies in multivariate time-series; one type is abnormal observation values or unusual subsequences within individual variable, and the other type is unexpected changes of relationships among multiple variables. In the following, we refer to the former anomalies as univariate anomalies, and the latter as multivariate anomalies. We provide an example of anomalies in Fig. 1 for clarity. The univariate anomaly can be found investigating the individual variable, but the multivariate anomaly cannot be found without watching the multiple variables at once.

A number of anomaly detection methods for multivariate time-series data have been proposed until now [1]. One of them is to regard time-series as a set of independent data samples that are distributed in high-dimensional space. Since the data samples can be embedded in lower-dimensional subspace owing to constraints of the system behavior, we can find anomalous observation by watching deviation from the subspace. This method can capture interdependency among multiple variables, but ignores time-domain correlations of signals. The most fundamental method to identify the subspace is the principal component analysis (PCA), which is a linear dimensionality reduction technique. As PCA is not suitable for nonlinear systems, several extensions to capture nonlinearity of data have been developed. For example, a recent study [2] showed that mixtures of probabilistic PCA [3] are advantageous for anomaly detection in spacecraft telemetry data, because spacecraft systems usually have several distinct operational modes.

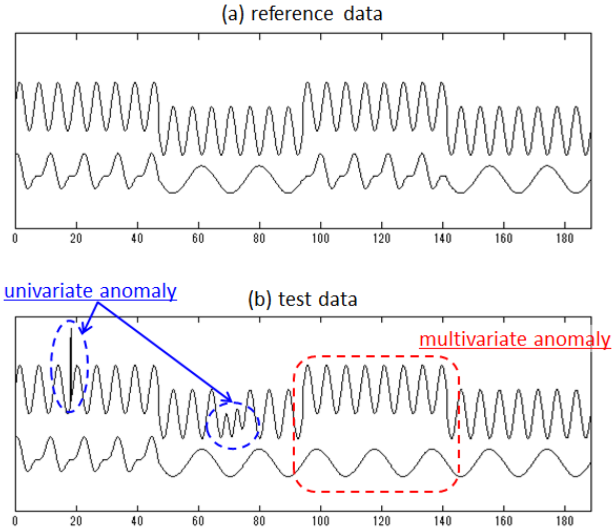


Fig. 1. Example of two types of anomalies; univariate anomaly (highlighted with blue circles) and multivariate anomaly (highlighted with red square). All the anomalies are defined by their unusuality, i.e., signals in test data are regarded as anomaly if they are not contained in reference data.

Another approach to detecting anomalies in multivariate time-series is to estimate their generative models, such as vector autoregressive (VAR) models and state space models (SSMs). Generative models can explain spatial relationships and time-domain correlations of variables if constructed appropriately, but it is difficult to estimate generative models without knowledge on characteristics of the system. Moreover, in anomaly detection, we are often interested in apparent patterns of time-series, rather than in assumed models behind the signals.

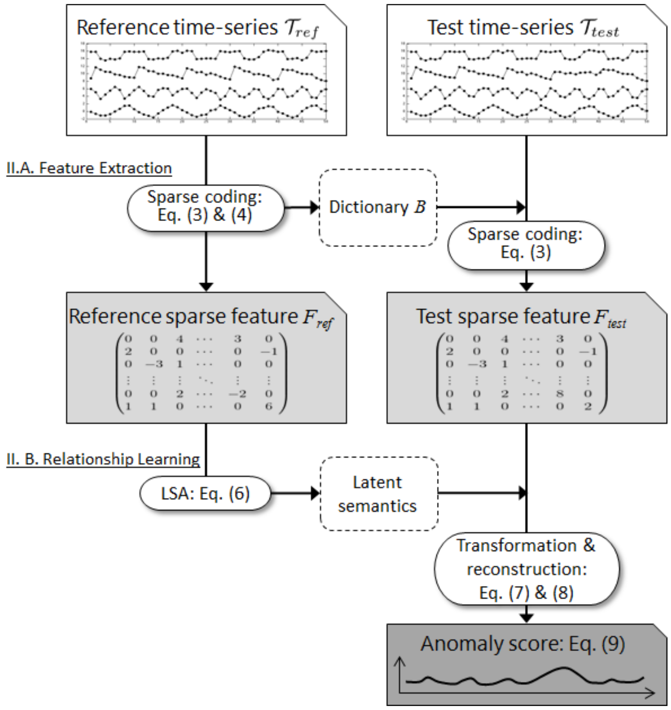

Reference time-series $\mathcal{T}_{ref}$
Test time-series $\mathcal{T}_{test}$
II.A. Feature Extraction
Sparse coding: Eq. (3) & (4)
Dictionary B
Sparse coding: Eq. (3)
Reference sparse feature F_{ref}
Test sparse feature F_{test}
II. B. Relationship Learning
LSA: Eq. (6)
Latent semantics
Transformation & reconstruction: Eq. (7) & (8)
Anomaly score: Eq. (9)

Fig. 2. Framework of the proposed method. The first stage is feature extraction with sparse coding. After extracting sparse features of local patterns of time-series, their co-occurrence relationships are learned using the latent semantic analysis (LSA). Then, anomaly scores are computed as deviations from the learned co-occurrence relationships in reference data.

Some other studies focused on the graph structures constructed from the nature of time-series signals for anomaly detection. Ide et al. [4] defined graphs based on the correlation between sensor signals and watched the graph with an assumption that the neighborhood structures were preserved under normal system operation. Qiu et al. [5] proposed to investigate graphical models of Granger causality of multivariate time-series. Cheng et al. [6] proposed to build graphs with kernel matrix and its alignment for unsupervised anomaly detection. Some of the graph-based techniques can investigate spatial relationships and time-domain correlations simultaneously. However, the number of edges of the graphs becomes too large if there are many variables, and techniques to reduce the computational cost are necessary.

In this study, we propose a novel anomaly detection technique for multivariate time-series to capture relationships of multiple variables and time-domain correlations simultaneously, without assuming any generative models or graph structures of signals. The proposed method is based on a natural idea that the relationships among variables can be captured by investigating co-occurrences of local patterns of the time-series.

II. Proposed Method

We focus on a semi-supervised anomaly detection, which assumes that only the labels for the normal class are given in training [1]. In other words, we seek unusual patterns or relations that do not appear in a reference time-series signal. We depict the framework of the proposed method in Fig. 2. The proposed method consists of two stages; the first is feature extraction with sparse coding [7] and the second is relationship learning with dimensionality reduction. We describe them in detail in the following subsections.

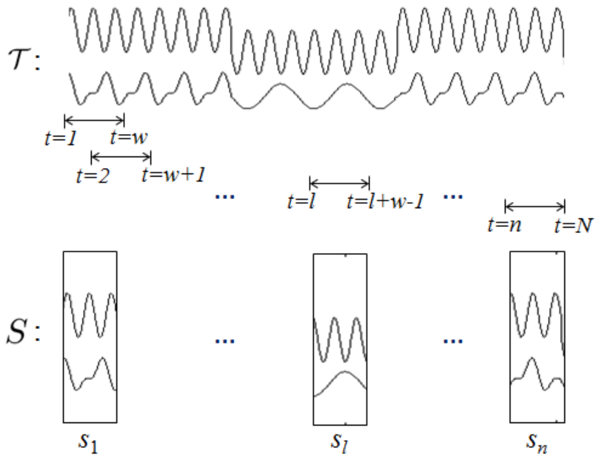


Fig. 3. Conceptual diagram of taking a sliding window from time-series. We run a overlapping sliding window that starts at $t = l$ for $l = 1, 2, \dots, n$, and obtain subsequences $s_1, \dots, s_n$, where $n = N + w - 1$.

In the following, $T^{(k)} = \{t_1^{(k)}, \dots, t_N^{(k)}\}$ denotes the k-th univariate time-series with the length of N and $\mathcal{T} = \{T^{(1)}, \dots, T^{(d)}\}$ denotes multivariate time-series with d variables. The subscript 'ref' means reference data or the training phase, and 'test' means test data or the test phase.

A. Feature Extraction with Sparse Representation

We first introduce the general framework of sparse representation [7]. The notations below are similar to those in [8]. With regard to a signal $\mathbf{y} \in \mathcal{R}^w$, a sparse representation $\mathbf{x} \in \mathcal{R}^m$, which is a sparse vector with a small number of nonzero components, is obtained by optimization as follows:

$$\underset{\mathbf{x}}{\text{minimize}} \quad ||\mathbf{y} - B\mathbf{x}||_2^2 + \gamma||\mathbf{x}||_1, \tag{1}$$

where $B = (\mathbf{b}_1 \cdots \mathbf{b}_m)$ is a basis dictionary that consists of a set of bases $\mathbf{b}_j$ $(j = 1, \dots, m)$ of the signal. A dictionary B is obtained by the following optimization:

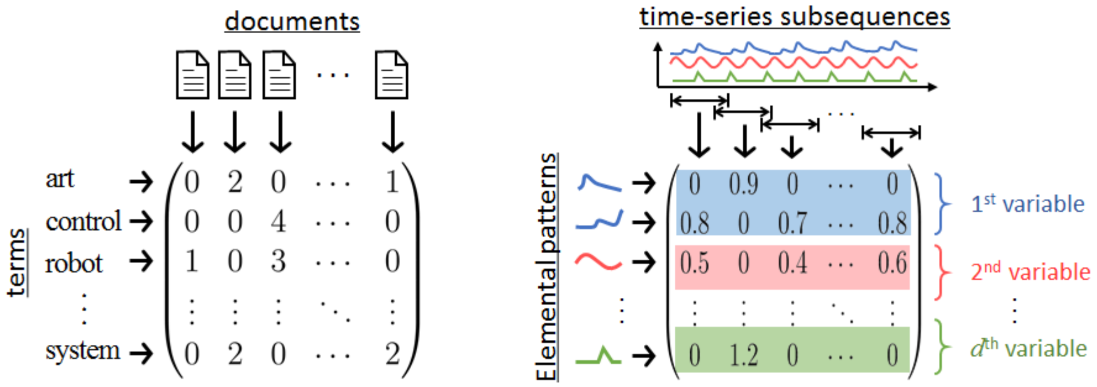
$$\underset{B}{\text{minimize}} \quad ||\mathbf{y} - B\mathbf{x}||_2^2 + \lambda\sum_{j=1}^{m}||\mathbf{b}_j||_2^2. \tag{2}$$

For n samples of the signal $\mathbf{y}_i(i = 1, \dots, n)$, (1) and (2) can be written in a matrix form respectively,

$$\underset{X}{\text{minimize}} \quad ||Y - BX||_2^2 + \gamma\sum_{i=1}^{n}||\mathbf{x}_i||_1, \tag{3}$$

$$\underset{B}{\text{minimize}} \quad ||Y - BX||_2^2 + \lambda\sum_{j=1}^{m}||\mathbf{b}_j||_2^2, \tag{4}$$

where $Y \in \mathcal{R}^{w \times n}$ is a signal matrix, and $X \in \mathcal{R}^{m \times n}$ is a matrix of the sparse representations. It is known that (3) and (4) are convex respectively, but not convex jointly. Thus, we solve (3) and (4) iteratively, optimizing X with B fixed and vice versa. Several efficient algorithms with global convergence have been proposed for this L1-regularized problem. We adopt



$$\begin{array}{l|ccccc} \text{art} \rightarrow & 0 & 2 & 0 & \cdots & 1 \\ \text{control} \rightarrow & 0 & 0 & 4 & \cdots & 0 \\ \text{robot} \rightarrow & 1 & 0 & 3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{system} \rightarrow & 0 & 2 & 0 & \cdots & 2 \end{array}$$

$$\begin{pmatrix} 0 & 0.9 & 0 & \cdots & 0 \\ 0.8 & 0 & 0.7 & \cdots & 0.8 \\ 0.5 & 0 & 0.4 & \cdots & 0.6 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1.2 & 0 & \cdots & 0 \end{pmatrix}$$

(a) term-document matrix A

(b) sparse feature matrix F

Fig. 4. Conceptual diagram of the analogy between a term-document matrix A and a sparse feature matrix F. (a) In a term-document matrix, the columns correspond to documents and the rows correspond to terms, and each element denotes the frequency of appearance. (b) In a sparse-feature matrix, the columns correspond to time-series subsequences and the rows correspond to elemental patterns, and each element denotes the weight of the elemental patterns.

the algorithm proposed by Lee et al. [8], which is based on estimation of feature signs of $\mathbf{x}_i$ and the Lagrange dual of the problem.

We utilize the sparse coding technique to extract elemental patterns of time-series. To capture local patterns, we first run a sliding window on time-series and obtain sets of overlapping subsequences,

$$S^{(k)} = (s_1^{(k)} \cdots s_n^{(k)}),$$

where $s_l^{(k)}$ is a subsequence of the time-series $T^{(k)}$ that begins at $t = l$, i.e., $s_l^{(k)} = (t_l^{(k)}, \dots, t_{l+w-1}^{(k)})^T$, $n = N - w + 1$, and w is the size of the sliding window. We show the concept of the sliding window in Fig. 3.

In the training phase, we run optimization (3) and (4) iteratively for each time-series variable $Y = S_{\text{ref}}^{(1)}, \dots, S_{\text{ref}}^{(d)}$, and obtain sparse representations $X_{\text{ref}}^{(k)}$ and dictionaries $B_{\text{ref}}^{(k)}$ respectively for d variables of reference time-series. Then in the test phase, only sparse representations $X_{\text{test}}^{(k)}$ are optimized by (3) with the fixed dictionary $B_{\text{ref}}^{(k)}$ for each variable of test time-series. The learned dictionary $B_{\text{ref}}^{(k)}$ consists of elemental patterns that compose the local patterns of the time-series $T_{\text{ref}}^{(k)}$ within a time scale according to the size of the sliding window w. The sparse representation $S^{(k)}$ is a set of coefficients of the bases and used as a feature of the time-series signals in the latter half of the proposed method.

To treat the d-variable time-series at once, we stack all the sparse features $X^{(k)}$ for d variables:

$$F = \begin{pmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(d)} \end{pmatrix} \in \mathcal{R}^{dm \times n},$$

and define a transformation of the original multivariate time-series $\mathcal{T} = \{T^{(1)}, \dots, T^{(d)}\}$ into the sparse features F as follows:

$$\mathcal{T} \rightarrow F. \tag{5}$$

B. Relationship Learning with LSA

After the feature extraction (5) of the reference time-series $\mathcal{T}_{\text{ref}}$, we learn co-occurrence relations of the local patterns. Our idea is based on an analogy between a sparse feature matrix F and a term-document matrix (see Fig. 4). A term-document matrix is a sparse matrix whose (i, j) element denotes the frequency of the i-th term in the j-th document. In other words, each column of A corresponds to each document and denotes the frequency of appearance about all the terms.

In the area of natural language processing, co-occurrence analysis from term-document matrices has been done with a dimensionality reduction technique called latent semantic analysis (LSA) [9]. LSA begins with a matrix decomposition of term-document matrix A by singular value decomposition (SVD):

$$A = U \Sigma V^T,$$

where U and V are sets of singular vectors, and Σ is a diagonal matrix of the singular values. In this notation, the left singular vectors $\mathbf{u}_1, \dots, \mathbf{u}_r$ correspond to the terms, and the right singular vectors $\mathbf{v}_1, \dots, \mathbf{v}_r$ correspond to the documents in semantic space. To preserve only the essential semantic relationship of terms, the term-document matrix is approximated as follows:

$$A = U_k \Sigma_k V_k^T, \tag{6}$$

where U_k, Σ_k and V_k are the matrices corresponding to the k-th largest singular values. A new document vector $\mathbf{a}$ can be transformed into a lower-dimensional semantic space by linear transformation:

$$\hat{\mathbf{a}} = \Sigma_k^{-1} U_k^T \mathbf{a}, \tag{7}$$

and it can be reconstructed by inverse transformation:

$$\tilde{\mathbf{a}} = U_k \Sigma_k \hat{\mathbf{a}}. \tag{8}$$

Based on the analogy between the term-document matrix and the sparse feature matrix depicted in Fig. 4, we utilize the same technique to extract co-occurrence relations of the local patterns of multivariate time-series. Both in training and test phases, the multivariate time-series $\mathcal{T}_{\text{ref}}$ and $\mathcal{T}_{\text{test}}$ are transformed into the sparse feature matrix F_{ref} and F_{test}, whose

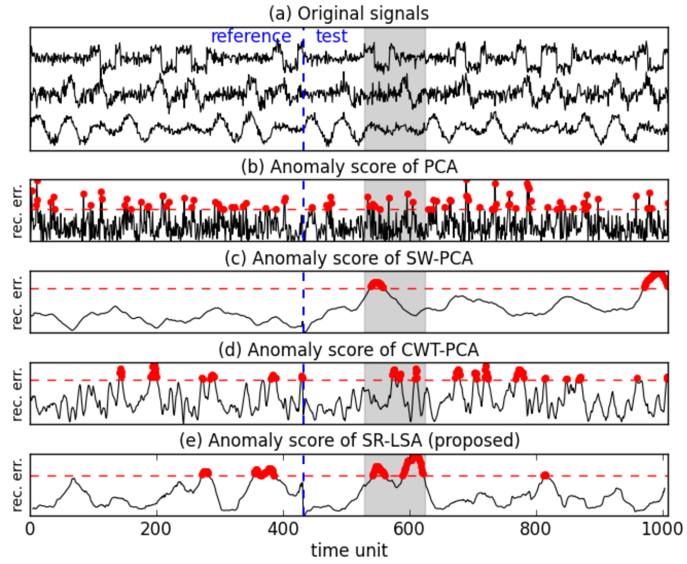


Fig. 5. (a) Original time-series used in Experiment 1; (b) anomaly scores of PCA; (c) anomaly scores of SW-PCA; (d) anomaly scores of CWT-PCA; and (e) anomaly scores of the proposed method. (b)-(e) plot the sum of squared reconstruction errors as anomaly scores, and large anomaly scores may indicate the presence of anomalies. The horizontal red line indicates a threshold of anomaly scores, which is decided as the 90th percentile of the scores in this figure.

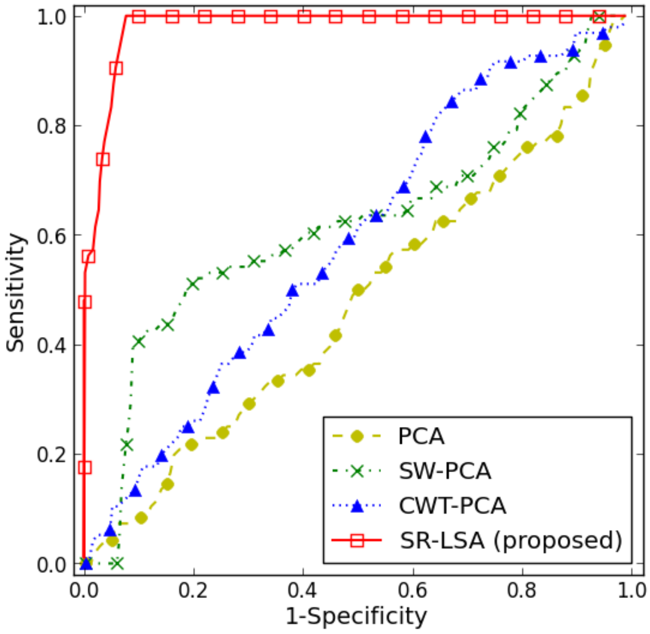


Fig. 6. ROC curves plotted changing the thresholds of anomaly scores in Experiment 1. Areas under the curve (AUCs) are; 0.465 (PCA); 0.608 (SW-PCA); 0.580 (CWT-PCA); and 0.968 (SR-LSA).

rows correspond to the local patterns and columns correspond to respective subsequences in the time-series. In the training phase, we apply (6) to the feature matrix F_{ref} and obtain $U_{k\text{ref}}$ and $\Sigma_{k\text{ref}}$. In the test phase, the feature matrix F_{test} is transformed into a semantic space by (7) and reconstructed into the original space by (8) using the learned matrices $U_{k\text{ref}}$ and $\Sigma_{k\text{ref}}$.

Because the rank-reduced matrices $U_{k\text{ref}}$ and $\Sigma_{k\text{ref}}$ preserve only the essential latent semantics, (8) cannot reconstruct the original feature perfectly and produces reconstruction errors. If the latent semantics (co-occurrence relations) in the test time-series data are not different from those in the reference data, the reconstruction errors will be small. However, if the test data contain anomalies, the reconstruction errors will be large on the anomalous elements of the feature matrix. Hence, we regard the squared reconstruction errors as anomaly scores;

$$(\text{Anomaly score}) = (F - \tilde{F}) \circ (F - \tilde{F}), \tag{9}$$

where $\tilde{F} = U_k \Sigma_k \hat{F}$, $\hat{F} = \Sigma_k^{-1} U_k^T F$, and $\circ$ is the entrywise product. Note that these anomaly scores are computed with regard to each sliding window, and an anomaly score of each timestamp is given as the average of scores of the corresponding windows.

The idea introduced above is intuitive, but the following points should be noted. First, components of a reconstructed matrix $\tilde{F}$ follow normal distribution [10], while components of an original matrix F follow Laplace distribution. Secondly, we assume exchangeability of sliding windows (documents) of time-series, that is, we focus only on local time-correlations in each sliding window, and ignore wider correlations among windows. Thirdly, the amount of training data for sparse coding and dictionary learning should be as large as possible to make a dictionary B contain every normal elemental pattern.

III. Experimental Results

This section describes the experimental results of the proposed method. We conducted experiments applying three types of anomaly detection based on principal component analysis (PCA) as well as the proposed method. One of the PCA-based methods is very naive one, which treats time-series signals as a set of independent observation (denoted by 'PCA' in the following). Other two types of PCA-based methods extract elementary features from the original time-series before applying PCA; one just takes sliding windows of size w and stacks for all variables (denoted by 'SW-PCA') [1], and the other computes continuous wavelet transforms (CWTs) with Daubechies wavelets and stacks for all variables (denoted by 'CWT-PCA'). The proposed method is denoted by 'SR-LSA' in the following.

The optimization parameters in (3) and (4) were set as following; $\gamma = 3$ in Experiment 1, $\gamma = 2$ in Experiment 2 and 3, and $\lambda = 1$ for all the experiments. The size of the sliding window w was set to 30 in Experiment 1 and 2, and $w = 60$ in Experiment 3. For dimensionality reduction with PCA or LSA, the reduced dimension k was chosen so that the cumulative contribution ratio of eigenvalues (singular values) just exceeded 0.9. The number of bases m was decided as; $m = 2w$ in Experiment 1 and 2, and $m = 6w$ in Experiment 3. Scale factors of CWTs were set to $1, \ldots, 2w$ in every experiment.

A. Experiment 1: Pattern combinations

The first experiment was conducted on the multivariate time-series shown in Fig. 5 (a), which were generated by

[1] SW-PCA corresponds to the singular spectrum analysis (SSA) [11] or the dynamic PCA (DPCA) [12].

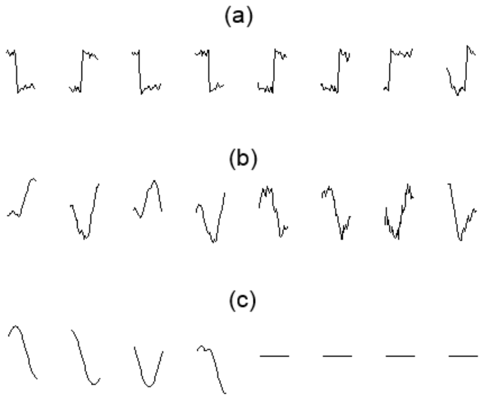


Fig. 7. A small part of the learned bases in Experiment 1; (a) bases learned with the time-series plotted in top of Fig. 5 (a); (b) bases learned with the middle time-series in Fig. 5 (a); and (c) bases learned with the undermost time-series in Fig. 5 (a).

combining two patterns for each variable. The total length of the generated time-series was about 1,000, and the first 430 signals were used as reference time-series, and the latter 570 were used as test time-series. Multivariate anomalies were put on the signals from 530 to 620 (as shaded in Fig. 5) as unusual relations of the patterns.

Figs. 5 (b), (c), (d) and (e) show the sum of anomaly scores with thresholds defined as the 90th percentile of scores, and Fig. 6 shows ROC curves plotted changing the thresholds. We can confirm that the proposed method successfully detected the inserted anomaly, while other methods cannot find the anomaly well. Especially, the naive PCA can hardly detect the anomaly because it ignores time-domain correlations.

For clarity, we show a small part of learned bases for each variable in Experiment 1. Fig. 7 (a) corresponds to the time-series plotted in top of Fig. 5 (a), Fig. 7 (b) corresponds to the middle time-series, and Fig. 7 (c) corresponds to the undermost time-series. These three sets of bases reflect the nature of each time-series variable, for example, stepwise bases shown in Fig. 7 (a) are similar to the local patterns of square waves in the time-series plotted in top of Fig. 5 (a).

B. Experiment 2: VARMA process

We conducted an experiment on the time-series shown in Fig. 8 (a), which were generated with a vector autoregressive moving average (VARMA) model. The first half of the signals was used as reference time-series, and the latter half was used as test time-series. The AR coefficient begins to change gradually at the time 2800, so anomalies should be found after 2800 as shaded in Fig. 8.

Figs. 8 (b), (c), (d) and (e) show the sum of anomaly scores with thresholds defined as the 90th percentile of scores, and Fig. 9 shows ROC curves plotted changing the thresholds. Every method succeed in detecting the anomalous behavior of the signal, but the PCA-based methods contain many false alarms. The proposed method is more robust than other methods based on dimensionality reduction, although the detectability of this kind of anomalies is not dictated so much by time-domain information.

C. Experiment 3: Wind turbine faults

The third experiment was conducted on the more realistic data shown in Fig. 10 (a). These data were generated with the wind turbine FDI (fault detection and isolation) benchmark model [13]. This benchmark model simulates a generic variable-speed wind turbine, causing faults of sensors, actuators and systems. We subsampled the generated time-series to 1/100, and used only the signals in the constant power production phase. The reference time-series contain no faults, whereas the test time-series contain six faults on the sensors, actuators or systems.

Figs. 10 (b), (c), (d) and (e) show that both of the PCA-based methods and the proposed method successfully detected the second fault, which is a scaling error on a rotor and generator speed sensor. On the other hand, only the proposed method was able to detect the fifth fault, which represents an anomalous offset in the converter torque control. Note that the first, third, fourth and sixth faults were not captured with any applied methods, and that the difference among the ROC curves is slight.

IV. Conclusion

In this study, we proposed a novel method for anomaly detection from multivariate time-series. The proposed method is based on sparse coding and co-occurrence learning using dimensionality reduction. The experimental results show that the proposed method can detect various types of anomalies well. Moreover, we confirmed that the proposed method was especially good at the multivariate anomalies in which only the relationships among time-series variables changed (Experiment 1).

As further study, validity of sparse coding, which is a kind of soft clustering, on time-series subsequence must be verified. We are also planning to apply more sophisticated co-occurrence analysis such as hierarchical latent topic models to overcome shortcomings of the SVD-based model.

References

[1] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” *ACM Computing Surveys*, vol. 41, no. 3, 2009.

[2] T. Yairi, M. Inui, A. Yoshiki, Y. Kawahara, and N. Tanaka, “Spacecraft telemetry data monitoring by dimensionality reduction techniques,” in *SICE Annual Conference*, 2010.

[3] M. E. Tipping and C. M. Bishop, “Mixtures of probabilistic principal component analysers,” *Neural Computation*, vol. 11, no. 2, 1999.

[4] T. Ide, S. Papadimitriou, and M. Vlachos, “Computing correlation anomaly scores using stochastic nearest neighbors,” in *IEEE International Conference on Data Mining*, 2007.

[5] H. Qiu, Y. Liu, N. A. Subrahmanya, and W. Li, “Granger causality for time-series anomaly detection,” in *IEEE International Conference on Data Mining*, 2012.

[6] H. Cheng, P. N. Tan, C. Potter, and S. A. Klooste, “Detection and characterization of anomalies in multivariate time series,” in *SIAM International Conference on Data Mining*, 2009.

[7] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[8] H. Lee, A. Battle, R. Raina, and A. Y. Ng, “Efficient sparse coding algorithms,” in *Neural Information Processing Systems (NIPS)*, 2007.

[9] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, “Indexing by latent semantic analysis,” *Journal of the American Society for Information Science*, vol. 41, no. 6, 1990.

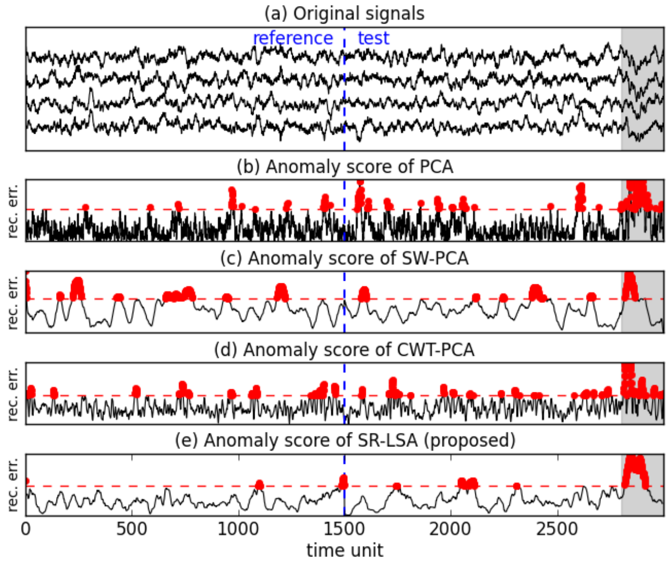


Fig. 8. (a) Original time-series used in Experiment 2; (b) anomaly scores of PCA; (c) anomaly scores of SW-PCA; (d) anomaly scores of CWT-PCA; and (e) anomaly scores of the proposed method. (b)-(e) plot the sum of squared reconstruction errors as anomaly scores, and large anomaly scores may indicate the presence of anomalies. The horizontal red line indicates a threshold of anomaly scores, which is decided as the 90th percentile of the scores in this figure.

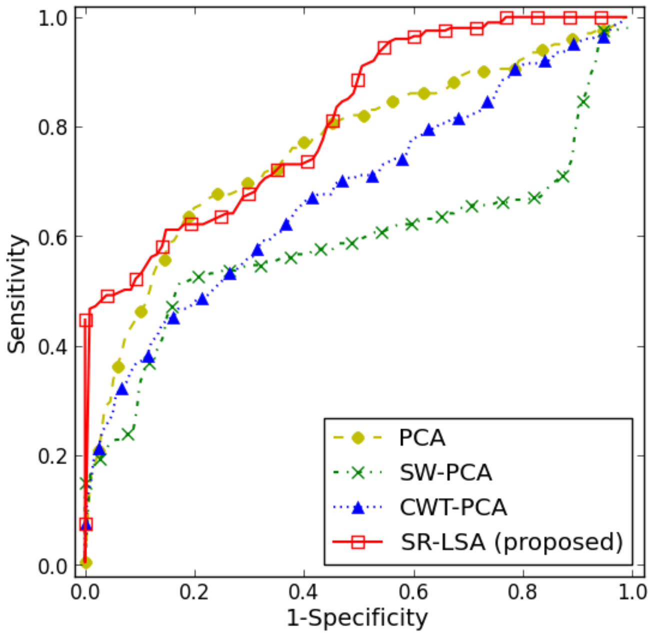


Fig. 9. ROC curves plotted changing the thresholds of anomaly scores in Experiment 2. Areas under the curve (AUCs) are; 0.751 (PCA); 0.579 (SW-PCA); 0.671 (CWT-PCA); and 0.802 (SR-LSA).

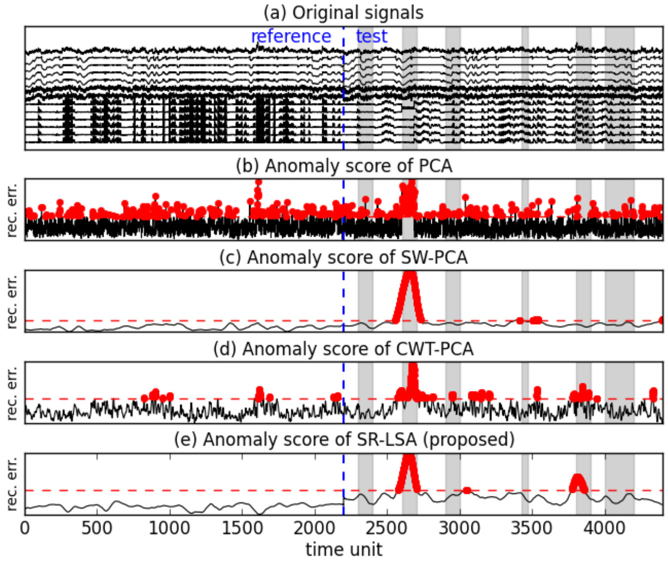


Fig. 10. (a) Original time-series used in Experiment 3; (b) anomaly scores of PCA; (c) anomaly scores of SW-PCA; (d) anomaly scores of CWT-PCA; and (e) anomaly scores of the proposed method. (b)-(e) plot the sum of squared reconstruction errors as anomaly scores, and large anomaly scores may indicate the presence of anomalies. The horizontal red line indicates a threshold of anomaly scores, which is decided as the 90th percentile of the scores in this figure.

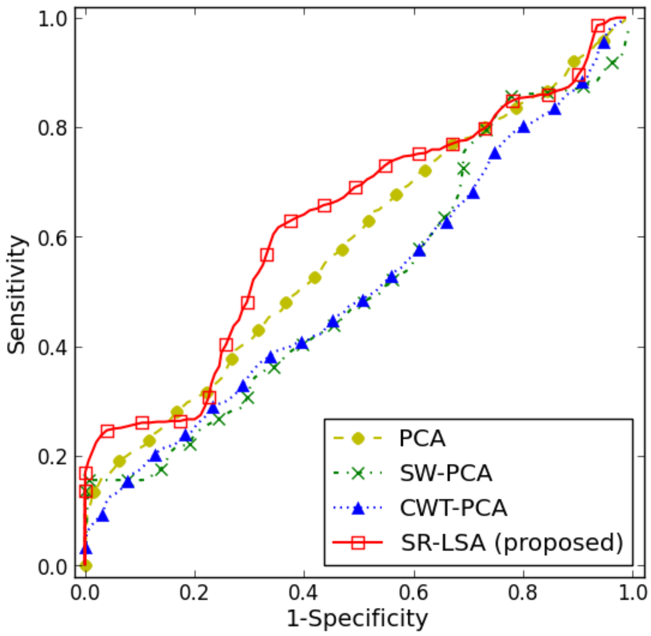


Fig. 11. ROC curves plotted changing the thresholds of anomaly scores in Experiment 3. Areas under the curve (AUCs) are; 0.574 (PCA); 0.510 (SW-PCA); 0.500 (CWT-PCA); and 0.614 (SR-LSA).

[10] C. D. Manning and H. Schtze, *Foundations of Statistical Natural Language Processing*. The MIT Press, 1999.

[11] N. Golyandina, V. Nekrutkin, and A. A. Zhigljavsky, *Analysis of time series structure: SSA and related techniques*. CRC Press, 2001.

[12] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 30, 1995.

[13] P. F. Odgaard, J. Stoustrup, and M. Kinnaert, "Fault tolerant control of wind turbines – a benchmark model," in *IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, 2009.