A Health Monitoring Method for Wind Power Generators with Hidden Markov and Probabilistic Principal Components Analysis Models

Riku Sasaki[1], Naoya Takeishi[2], Takehisa Yairi[3], Koichi Hori[4], Kazunari Ide[5], and Hiroyoshi Kubo[6]

[1,2,3,4]*The University of Tokyo, Bunkyo-ku, Tokyo, 113-0033, Japan*
sasaki@g.ecc.u-tokyo.ac.jp
takeishi@ailab.t.u-tokyo.ac.jp
yairi@ailab.t.u-tokyo.ac.jp
hori@g.ecc.u-tokyo.ac.jp

[5,6]*Mitsubishi Heavy Industries, Ltd., Nagasaki, Nagasaki, 851-0392, Japan*
kazunari_ide@mhi.co.jp
hiroyoshi_kubo@mhi.co.jp

ABSTRACT

In this work, we propose a data-driven health monitoring method for wind power generators, which learns an empirical model from the time-series sensor data and detects irregularities or faults in the turbines and blades. Our main objective is to predict any symptoms of faults as early as possible before the generators fall into malfunction. The data obtained from the wind power generators are strongly correlated multidimensional time-series with multiple states. In this study, we take such features into account and develop the probabilistic model for them, namely, *hidden Markov and probabilistic principal component analysis*. Once the model is learned with the data that contain no faulty events, it can be used to detect faults in new data by comparing the original sensor values and reconstructed ones. In this research, we apply this method to synthetic data and real-world wind turbine data and show the results of experiments to confirm the availability of the proposed method.

1. INTRODUCTION

In operation of large-scale engineering systems like power plants and grids, it is important to detect irregularities or faults before the systems get into malfunction. Since such systems are composed of many equipment and sensors that are intricately correlated, searching for the irregularities by inspecting system's status manually is often prohibitive. To overcome the limitation of the manual inspection, data-driven methods based on machine learning techniques are attracting attention for the health monitoring of engineering systems. Although such method is versatile and can be utilized for various systems, one must consider characteristics of the data in hand carefully and adapt the method to them appropriately.

In this work, we propose a data-driven health monitoring method especially for wind power generators. We focus on the following three characteristics of the data obtained from wind power generators. Firstly, those data comprise time-series that are non-stationary due to state transitions such as switching from ceasing to working. Secondly, they comprise many types of variables (sensor readings) with strong correlations among them. Thirdly, most faulty behaviors cannot be known precisely beforehand because of the changes of the environmental properties. Considering the abovementioned three points, we propose a probabilistic model to model "normal" data of the generators, and apply it to data of turbine blades of a wind power generator.

The remainder of this paper is organized as follows. After reviewing the background in Section 2, we introduce the proposed method in Section 3. The experimental settings and results are shown in Section 4, and this paper ends with conclusions in Section 5.

2. BACKGROUND

While there have been proposed various types of algorithms for fault detection (see, e.g., Chandola et al., 2009), we adopt an "unsupervised" approach based on probabilistic generative models for modeling the data without any faulty events because we have little information on the faulty behaviors beforehand. In this approach, we learn the parameters of the models in a training phase and search for deviants from the learned model in a test (detection) phase.

According to the characteristics of the data of wind power generators, we focus on the following two popular probabilistic models: the mixtures of probabilistic principal component analysis (MPPCA) (Tipping & Bishop, 1999) and the hidden Markov models (HMMs) (see e.g., Bishop, 2006). The former can model the correlations among multiple variables (sensors) of the data, as well as their cluster structures. The latter takes the temporal correlation of the data into account and estimates discrete hidden states, and thus it is suitable for the time-series wherein the state transitions are present. In this work, we combine these two probabilistic models; we model the latent variables that indicate assignments of the mixture components of MPPCA by a Markovian model. The detail of the proposed probabilistic model is described in the next section.

3. PROPOSED METHOD

3.1. Overview

In the proposed method, the sensor data that do not contain faulty behaviors (*training data*) are preprocessed and modeled with the *hidden Markov and probabilistic principal component analysis* (HM-PPCA), which is a combination of MPPCA and HMM. After the "normal" model being estimated, we detect faults in another set of data that may contain faults (*test data*) by monitoring the reconstruction errors by the estimated model. The proposed method is schematically shown in Figure 1. In the subsequent sections, we introduce the method for estimating the parameters of HM-PPCA with the training data and the way to compute the reconstruction errors on the test data.

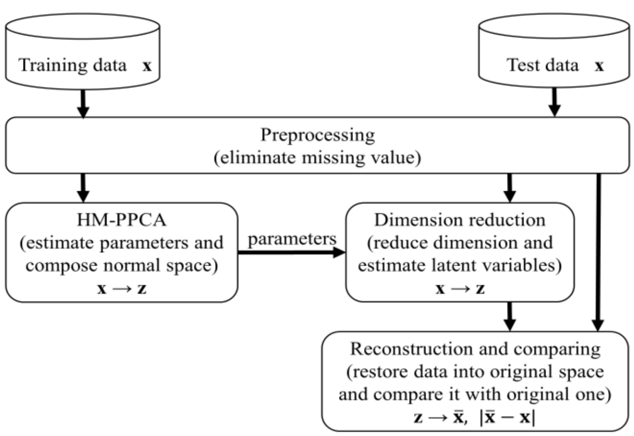

Figure 1. Framework of the proposed method.

3.2. Parameter Estimation of HM-PPCA

The probabilistic model of HM-PPCA is as follows.

$$p(u_t = k_2 | u_{t-1} = k_1) = A_{k_1,k_2}, \quad (1)$$

$$p(z_{t,k}) = \mathcal{N}(0, I), \quad (2)$$

$$p(x_t | u_t = k, z_{t,k}) = \mathcal{N}(W_k z_{t,k} + \mu_k, \sigma_k^2 I). \quad (3)$$

In the above equations, $x_t \in \mathbb{R}^D$ represents the data vector at the t-th timestamp, $z_{t,k} \in \mathbb{R}^d$ is the (continuous) latent variable of the k-th mixture component of MPPCA, and the (discrete) latent variable $u_t \in \{1, \ldots, K\}$ indicates the mixture component of MPPCA at the t-th timestamp and works as the hidden state of HMM. The matrix $A \in \mathbb{R}^{K \times K}$ is the state transition matrix and A_{k_1,k_2} denotes the (k_1, k_2)-th element of A. Matrix W_k is the factor loading matrix of the k-th component of MPPCA, μ_k is the corresponding bias term, and σ_k^2 is the noise variance.

We infer the latent variables and estimate the parameters of HM-PPCA, $\theta = (A, W_{1:K}, \mu_{1:K}, \sigma_{1:K}^2)$, by an EM algorithm (known as Baum-Welch algorithm for HMM; see, e.g., Bishop, 2006), where inference of the latent variables (E-step) and the maximization of the expectation of log likelihood (M-step) are conducted iteratively. In the M-step, the update of the parameters is conducted following the computation of the maximum likelihood estimator of MPPCA (Tipping & Bishop, 1999).

3.3. Computation of Reconstruction Errors

After estimating the parameters of HM-PPCA with the training data, we run the Viterbi algorithm (Forney, 1973) on the test data with the parameters being fixed to decide the assignment of the mixture components on each timestamp of test data. Once the mixture component is determined, the reconstruction error e_t is defined as follows:

$$e_t = x_t - \bar{x}_t, \quad (4)$$

$$E(z_t | x_t) = M^{-1} W^{\mathrm{T}} (x_t - \mu), \quad (5)$$

$$\bar{x}_t = W (W^{\mathrm{T}} W)^{-1} M E(z_t | x_t) + \mu, \quad (6)$$

$$M = W^{\mathrm{T}} W + \sigma^2 I, \quad (7)$$

where we dropped the subscript k for simplicity.

When the magnitude of e_t is large, it means that the data at that time is not being described well by the learned probabilistic model. Hence, we investigate the magnitude of the reconstruction errors calculated as above to detect the faults of the system.

3.4. Implementation

The estimations shown above are implemented according to the following procedures.

Algorithm 1. (Fault detection by HM-PPCA)

1 Initialize parameters A, W, μ, σ

2 Estimate parameters with Baum-Welch algorithm, i.e., iterate 2.1 and 2.2 until it converges.

- *2.1* Compute ξ, γ using current estimation of parameters A, W, μ, σ.
- 2.2 Find parameters which maximize the likelihood, i.e, update A using ξ, γ and update W, μ, σ using γ, d.

3 Estimate the assignment of latent discrete variable δ using Viterbi algorithm with θ.

4 Reconstruct data in subspace and calculate errors, i.e., calculate $\bar{x}$ using W, μ, σ.

Note that we define ξ, γ and δ as follows.

$$\xi_t(i,j) = p(u_t = k_i, u_{t+1} = k_j | x, \theta) \tag{8}$$

$$\gamma_t(i) = p(u_t = k_i | x, \theta) \tag{9}$$

$$\delta_t = \max_i p(u_1, \dots, u_{t-1}, u_t = k_i, x_1, \dots, x_t | \theta) \tag{10}$$

4. EXPERIMENT

We conducted experiments of fault detection of a wind power generator with synthetic and real-world datasets. In the experiments introduced below, we set the number of mixture components by $K = 3$, since there are three possible states of the wind power generator: *running*, *stopped* and *transition*. The *transition* state is considered for the transient phase between the *running* and *stopped* states, and thus it would be rare compared to the other two.

4.1. Synthetic Data

We conducted experiment on synthetic data that imitate the real-world data introduced in the next subsection. The synthetic data comprise 12 variables, each of four variables corresponds to one of three components of a wind turbine. These variables indicate zero when the state is *stopped* and 15 when *running*. The transitions of these two states were smoothed to generate regions corresponding to *transition* state. Moreover, we added noise on those data, as shown in Figure 2. For simplicity, we assumed all of the 12 variables follow the same state changes, so that data are one dimension essentially. Hence, we defined the dimensionality of latent variable z as $d = 1$. To prepare the test data that contain "faulty" behaviors, we added the gradual increase of the values to simulate unexpected loads on one of the three components. In Figure 3, we show the example of such artificial faults.

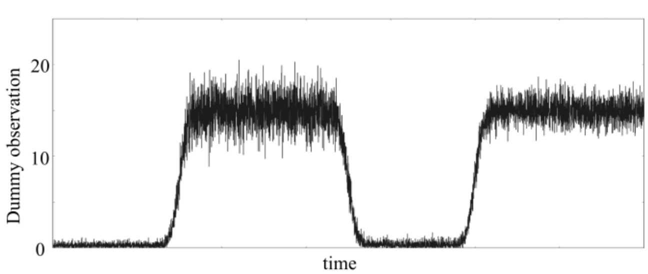

Figure 2. A part of synthetic data without any faulty events.

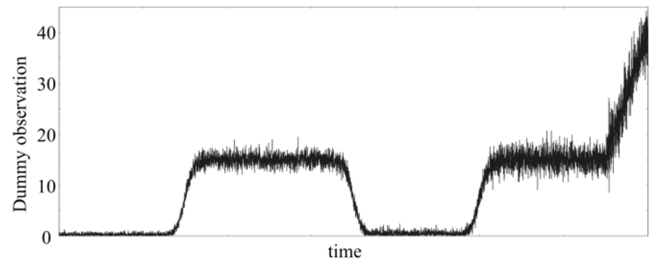

Figure 3. A part of synthetic data with a faulty event.

We show the reconstruction errors on the test data by HM-PPCA in Figure 4, wherein the errors of the 12 variables are shown in the three plots and each plot shows summed four variables in each component. When the artificial fault occurs in the last part of the test data, the reconstruction errors in the corresponding part of the data become obviously larger than the ones in the other parts.

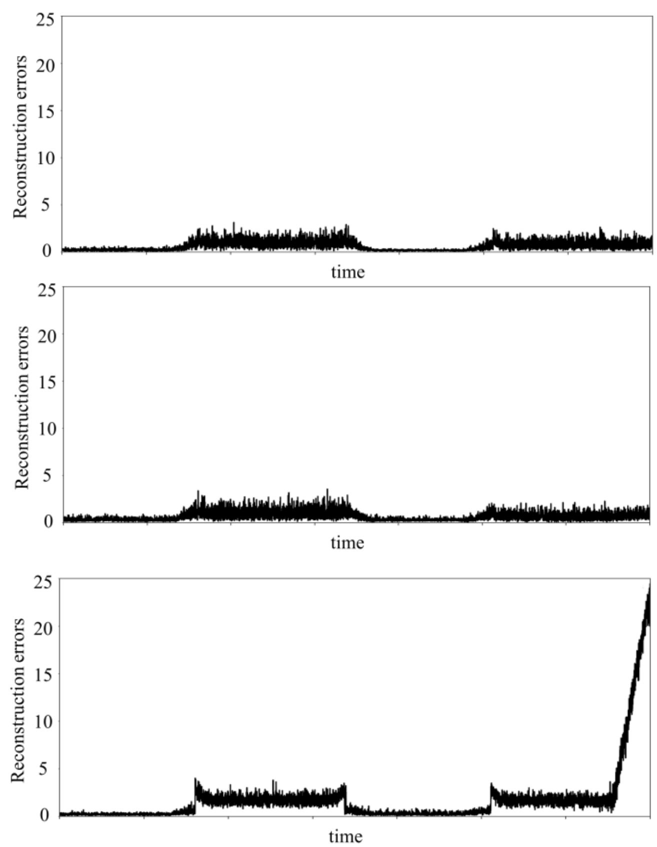

Figure 4. Reconstruction errors on synthetic data.

4.2. Wind Turbine Data

We conducted another experiment using real-world data obtained from strain sensors attached to turbine blades of a wind power generator. The wind power generator has three turbine blades that have four strain sensors each, and thus we obtain 12 dimensional time-series data. The manner to attach the four sensors on each blade is similar among the three blades. We prepared both a set without any faulty events as training data and a set with a fault as test data. In the last part of the test data, a fault occurred on the turbine blade, and our task is to detect this fault from data. In this experiment, we defined the dimensionality of latent variable z as $d = 6$.

We show the reconstruction errors in Figure 5, wherein the errors of the 12 variables are shown in the three plots and each plot shows summed four variables in each component. The fault in the test data is successfully detected as the large magnitude of the reconstruction errors.

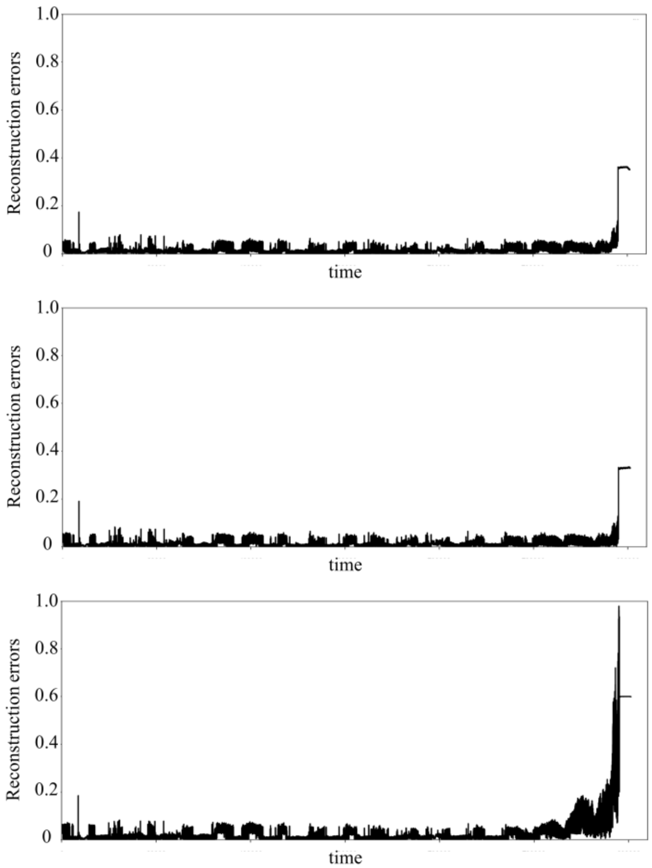

Figure 5. Reconstruction errors on a wind turbine system.

Next, let us look at the values of state transition matrix A estimated with the real-world data. We show the schematic diagram of the transitions following A in Figure 6, and the values of the elements are shown in Table 1. One can see that the transitions between the same states are overwhelming and the transition between *running* and *stopped* hardly occurs.

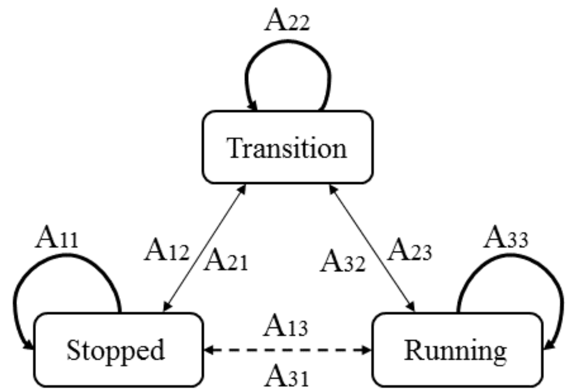

Figure 6. A schematic diagram of state transitions.

Table 1. Values of the state transition probability.

		Destination state		
		stopped	transition	running
Source state	stopped	A_{11}=0.999	A_{12}=0.001	A_{13}=0.000
	transition	A_{21}=0.001	A_{22}=0.988	A_{23}=0.011
	running	A_{31}=0.000	A_{32}=0.024	A_{33}=0.976

5. CONCLUSIONS

In this work, we proposed a data-driven health monitoring method for wind power generators, focusing on the characteristics of the data. The proposed method relies on the probabilistic model termed hidden Markov and probabilistic principal component analysis (HM-PPCA), which is a combination of HMM and MPPCA. The experimental results show the availability of the proposed method for fault detection. Online updates of parameters to deal with gradual degradation of the system would be one of the future researches, and applying the proposed method to the system with more sensors is also an important challenge.

REFERENCES

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York, NY: Springer-Verlag New York, inc.

Chandola, V., Banerjee, A. and Kumar, V. (2009). Anomaly Detection: A Survey. *ACM Computing Surveys*. vol. 41(3). doi:10.1145/1541880.1541882

Forney, G. D. (1973). The Viterbi Algorithm. *Proceedings of the IEEE*. Vol. 61(3), pp. 268–278. doi:10.1109/PROC.1973.9030

Tipping, M. E. and Bishop, C. M. (1999). Mixtures of Probabilistic Principal Component Analysers. *Neural Computation*. vol. 11(2), pp. 443–482. doi:10.1162/089976699300016728