ROBUST SENSOR ESTIMATION USING TEMPORAL INFORMATION

Chao Yuan and Claus Neubauer

Siemens Corporate Research, 755 College Road East, Princeton NJ, 08540

ABSTRACT

We propose a dynamic Bayesian framework for sensor estimation, a critical step of many machine condition monitoring systems. The temporal behavior of normal sensor data is described by a stationary switching autoregressive (SSAR) model that possesses two advantages over traditional switching autoregressive (SAR) models. First, the SSAR model removes time dependency of signals during mode switching and fits sensor data better. Secondly, the SSAR model is stationary in that at each time, sensor data have the same distribution which represents the normal operating range of a system; this ensures that estimates are accurate and are not distracted by deviations. During monitoring the deviation covariance is estimated adaptively, which effectively handles variable levels of deviations. Tests on gas turbine data are presented.

Index Terms— Machine condition monitoring, autoregressive, Kalman filter, Gaussian mixture model

1. INTRODUCTION

The objective of machine condition monitoring is to predict failures of complex systems to avoid costly damages. When a machine works properly, its sensor data should be distributed in a *normal operating range*. Deviations from this range may indicate a fault. We address sensor estimation, a critical step in many condition monitoring systems [1]. Specifically, given the observation of sensor values $\mathbf{y}$, we estimate the hidden $\mathbf{x}$, the fault-free values that sensors should have if the system operates normally. Based on the locations and magnitudes of deviations $(\mathbf{y} - \mathbf{x})$, various rules can be designed for fault diagnosis.

Most widely used sensor estimation algorithms such as Auto-Associative Neural Networks (AANN)[1], the multivariate state estimation technique (MSET) [2] and support vector regression (SVR) [1], attempted to establish a deterministic mapping network from $\mathbf{y}$ to $\mathbf{x}$. These methods desire a very large training pair set $\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1:N}$ representative of all possible deviations, which is seldom available. Our previous work [3] achieved much better performances via a Bayesian framework, in which we mapped $\mathbf{y}$ to $\mathbf{x}$ probabilistically and adaptively without the need for a large training pair set. However, all the above methods ignored the temporal information richly present in sensor signals and only used $\mathbf{y}_t$ at time t to estimate $\mathbf{x}_t$.

In this paper, we extend our previous work by using all available observations $\mathbf{y}_{1:t}$ to estimate $\mathbf{x}_t$. In particular, we present a stationary switching autoregressive (SSAR) model to model the joint distribution of $\mathbf{x}_{1:t}$. Our SSAR model differs from the conventional switching autoregressive (SAR) model [4] mainly in the following aspects.

First, the SAR model assumed that dependency between $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ always exists. However, for a machine such dependency becomes very weak during mode switching and it is not appropriate to use an autoregressive (AR) model at such time. In contrast, our SSAR model assumes that $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ are independent during mode switching. Such independence assumption was also found to be useful in modeling music signals [5].

Secondly and more importantly, the SAR model did not offer a way to describe the normal operating range of a machine. This does not prevent its success because in many applications, the signal $\mathbf{x}_t$ (e.g., the velocity and position of a target in a visual tracking system) is not confined to a certain range. However, in machine condition monitoring the normal signal $\mathbf{x}_t$ must be within the normal operating range. Our SSAR model is *strictly stationary* in that for all time $t = 1, 2, ..., \infty$, $\mathbf{x}_t$ has the same distribution $P(\mathbf{x}_t)$, which represents the normal operating range. This is essential to ensure that estimates are not distracted by any deviations.

This paper is organized as follows. In Sect.2, we briefly review our prior work. The new model is described in Sect.3. Test results are presented in Sect.4. Sect.5 summarizes this paper.

2. OUR PREVIOUS ALGORITHM

In our previous study, we presented a Bayesian sensor estimation algorithm without using temporal information. During training, we used normal operation data to learn the distribution $P(\mathbf{x}_t)$ of normal sensor signals via a Gaussian mixture model (GMM). During monitoring, the observation $\mathbf{y}_t$ was modeled as a Gaussian distribution conditioned on $\mathbf{x}_t$:

$$P(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t|\mathbf{C}\mathbf{x}_t, \mathbf{\Theta}_t), \tag{1}$$

where $\mathbf{C}$ is the observation matrix and is needed when sensor data at consecutive data points are concatenated into one column vector $\mathbf{x}_t$. $\mathbf{\Theta}_t$, the covariance of deviation $(\mathbf{y}_t - \mathbf{C}\mathbf{x}_t)$ is

an unknown diagonal matrix. The sensor estimation problem was formulated as computing the conditional expectation of $\mathbf{x}_t$ given $\mathbf{y}_t$:

$$E(\mathbf{x}_t|\mathbf{y}_t, \mathbf{\Theta}_t). \quad (2)$$

We employed the Expectation-Maximization (EM) algorithm [6] to estimate (2) and $\mathbf{\Theta}_t$, simultaneously. The EM algorithm is a powerful tool for maximum likelihood estimation of parameters of $P(\mathbf{y}_t)$ in the presence of hidden variables. In our case, $\mathbf{x}_t$ is viewed as the hidden variable and $\mathbf{\Theta}_t$ is the parameter. We alternate the computation of (2) in the E-step and the estimation of $\mathbf{\Theta}_t$ in the M-step. We output the result of (2) as the estimate of $\mathbf{x}_t$ when the algorithm converges. This model is able to handle different levels of deviations through $\mathbf{\Theta}_t$ which is adaptively estimated for each input $\mathbf{y}_t$. If the ith sensor value y_i is normal, the corresponding variance θ_i will be small such that $x_i \approx y_i$; if a sensor value y_i is faulty, θ_i will be large to allow x_i to be different from y_i.

3. THE PROPOSED ALGORITHM

We give an overview about how our new method works. Similarly to [3], the normal operating range of a machine is presented by a Gaussian mixture model (GMM). However, in [3], $\mathbf{x}_t$ was independently drawn from this GMM at time t. In contrast, in our new model the normal signal $\mathbf{x}_t$ can evolve in each Gaussian distribution (mode) following an autoregressive model or choose to jump to another mode. We first describe how normal signal $\mathbf{x}_t$ evolves in each mode in Sect.3.1. In Sect.3.2 we elucidate how $\mathbf{x}_t$ progresses in the full normal operating range. During monitoring, given a sequence of observations $\mathbf{y}_{1:t}$, we estimate $\mathbf{x}_t$, as shown in Sect.3.3.

3.1. The stationary autoregressive model

We first present a stationary autoregressive (STAR) model. The dynamics between $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ is described by a conditional Gaussian distribution:

$$P(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t|\mathbf{A}(\mathbf{x}_{t-1} - \mathbf{m}) + \mathbf{m}, \mathbf{Q}), \quad (3)$$

where $\mathbf{A}$ is the state transition matrix and covariance $\mathbf{Q}$ specifies the uncertainty of such transition. Eq.(3) depicts an AR model with mean $\mathbf{m}$. We extend it to the STAR model by additionally requiring:

$$P(\mathbf{x}_1) = \mathcal{N}(\mathbf{x}_1|\mathbf{m}, \mathbf{V}), \mathbf{V} = \mathbf{AVA}^T + \mathbf{Q}. \quad (4)$$

We now sketch the proof of this model's stationarity. Based on (3) and (4), $P(\mathbf{x}_2) = \mathcal{N}(\mathbf{x}_2|\mathbf{m}, \mathbf{V})$. Using induction, we have $\mathbf{x}_t = \mathcal{N}(\mathbf{x}_t|\mathbf{m}, \mathbf{V})$. Thus, $P(\mathbf{x}_t)$ is the same function for any t, which completes the proof.

The parameters of the STAR model include $\mathbf{m}$, $\mathbf{V}$, $\mathbf{A}$ and $\mathbf{Q}$. Standard methods such as maximum likelihood estimation (MLE) or Yule-Walker estimation for AR models [7] is not directly applicable due to the new constraints in (3) and (4). Standard ways to solve this problem are usually gradient-based methods (e.g., constrained quasi-Newton method), which are both time consuming and easily trapped in local extrema. Using the stationarity of $\mathbf{x}_t$, we can derive the following simple but very efficient learning method. First, $\mathbf{m}$ and $\mathbf{V}$ are learned from sample mean and sample covariance, respectively. We then learn $\mathbf{A}$ from (3) using MLE. Finally, $\mathbf{Q}$ is set to $\mathbf{V} - \mathbf{AVA}^T$. $\mathbf{Q}$ obtained in this way is not necessarily positive definite. Thus, we search for the largest $r \leq 1$ such that $\mathbf{Q} = \mathbf{V} - r^2\mathbf{AVA}^T$ is positive definite. The $\mathbf{A}$ accordingly is replaced by $r\mathbf{A}$. In our experience, r ranges from 0.95 to 1.

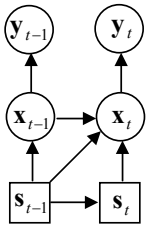

Fig. 1. The proposed model for sensor estimation. During training, we learn the SSAR model (consisting of $\mathbf{x}$ and s nodes). During monitoring, we estimate $E(\mathbf{x}_t|\mathbf{y}_{1:t})$.

3.2. The stationary switching autoregressive model

We now describe all the dependences and constraints in the SSAR model:

$$P(s_t = j|s_{t-1} = i) = Z_{ij},\ P(s_1) = \pi_j, \quad (5)$$

$$P(\mathbf{x}_t|\mathbf{x}_{t-1}, s_t = j, s_{t-1} = i) = \begin{cases} \mathcal{N}(\mathbf{x}_t|\mathbf{A}_j(\mathbf{x}_{t-1} - \mathbf{m}_j) + \mathbf{m}_j, \mathbf{Q}_j) & \text{if } i = j \\ \mathcal{N}(\mathbf{x}_t|\mathbf{m}_j, \mathbf{V}_j) & \text{if } i \neq j \end{cases}, \quad (6)$$

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{Z},\ \mathbf{V}_j = \mathbf{A}_j\mathbf{V}_j\mathbf{A}_j^T + \mathbf{Q}_j. \quad (7)$$

In the above equations, Eq.(5) describes a Markov chain with M possible modes. s_t is the discrete mode variable taking values between 1 and M. Eq.(6) describes possible transitions from $\mathbf{x}_{t-1}$ to $\mathbf{x}_t$. If mode s_t is the same as the previous mode s_{t-1}, the previous STAR model continues (see Sect.3.1); during mode change, the previous $\mathbf{x}_{t-1}$ is forgotten and $\mathbf{x}_t$ is randomly drawn from the Gaussian distribution of the new mode. Fig.1 (excluding $\mathbf{y}$ nodes) visualizes all these dependencies. The stationarity of the SSAR model is enforced by (7). We omit the proof due to space limit.

The SSAR model includes the following parameters: $\boldsymbol{\pi}$, $\mathbf{Z}$, $\mathbf{m}_j$, $\mathbf{V}_j$, $\mathbf{A}_j$, $\mathbf{Q}_j$, where $j = 1, 2, ..., M$. By using the stationarity of $\mathbf{x}_t$, we learn these parameters sequentially. In general, we proceed by first learning $\mathbf{m}_j$, $\mathbf{V}_j$ and $\mathbf{Z}$, similarly to the learning algorithm of the hidden Markov model (HMM). Then, $\mathbf{A}_j$ is estimated only using (6) via MLE. Finally, two constraints are enforced by setting $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{Z}$ and $\mathbf{Q}_j = \mathbf{V}_j - \mathbf{A}_j\mathbf{V}_j\mathbf{A}_j^T$.

A simpler algorithm can be derived, if training data can be partitioned into different modes. With the partitioned data, we can easily obtain $\mathbf{Z}$ and we set $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{Z}$. A STAR model

described by $\mathbf{m}_j$, $\mathbf{V}_j$, $\mathbf{A}_j$, $\mathbf{Q}_j$ for mode j is then learned using the training data under this mode (see Sect.3.1).

3.3. Monitoring

During monitoring, we inherit the observation model from our previous work; i.e., $\mathbf{y}_t$ is conditioned on $\mathbf{x}_t$ via (1). Fig.1 shows the dependencies of all random variables. The sensor estimation problem is now formulated as computing the conditional expectation of $\mathbf{x}_t$ given $\mathbf{y}_{1:t}$:

$$E(\mathbf{x}_t|\mathbf{y}_{1:t}, \mathbf{\Theta}_t). \tag{8}$$

We again employ the EM algorithm to compute (8) by maximizing $P(\mathbf{y}_{1:t})$. This EM procedure resembles the learning of a switching Kalman filter (SKF) [4] but is much simpler, because the only unknown parameter is $\mathbf{\Theta}_t$ ($\mathbf{\Theta}_{1:t-1}$ were estimated in the past) and all SSAR-related parameters were already learned during training. Learning SKF is a standard procedure (see [4] for details). We only sketch the major steps here.

In the E-step, the goal is to compute $P(\mathbf{x}_t|\mathbf{y}_{1:t})$ using the current estimate of $\mathbf{\Theta}_t$. For the standard SKF, this is a mixture of M^2 Gaussians. Each Gaussian $P(\mathbf{x}_t|\mathbf{y}_{1:t}, s_t = j, s_{t-1} = i)$ with a probability $P(s_t = j, s_{t-1} = i|\mathbf{y}_{1:t})$ is obtained by Kalman filtering. To avoid exponential mode growth, collapsing was used to merge M Gaussians $P(\mathbf{x}_t|\mathbf{y}_{1:t}, s_t = j, s_{t-1} = i)$ (where $i = 1, 2, ..., M$) into one Gaussian $P(\mathbf{x}_t|\mathbf{y}_{1:t}, s_t = j)$. In contrast, our Fig.1 model only involves $2M$ Gaussians, since all $M - 1$ Gaussians $P(\mathbf{x}_t|\mathbf{y}_{1:t}, s_t = j, s_{t-1} = i)$ (where $i \neq j$) are equal. Thus, in our collapsing step, only two Gaussians are merged into one. In the M-step, $\mathbf{\Theta}_t$ is estimated. We output the result of (8) after the EM algorithm converges.

4. TEST RESULTS

We applied the proposed algorithm to monitor gas turbines. In its steady state, a gas turbine typically operates either in a full load mode or a part load mode. We refer to the switching between these load modes as a transition mode. There are thus a total of three modes ($M = 3$). We consider artificial faults, because by knowing the ground truth (original) value of each sensor we can evaluate sensor estimation accuracy quantitatively.

The error measures we use are now introduced. *Estimation error* of a sensor is the absolute difference between the estimate and the ground-truth value. We distinguish three types of estimation errors. E_n is the average estimation error of all sensors in the normal time range. E_{nf} is the average estimation error of all normal sensors in the faulty time range. E_{ff} is the average estimation error of all faulty sensors in the faulty time range. Small values are preferred for all these errors.

We selected 22 sensors useful for fault diagnosis to build our models. These sensors are gas flow, power, inlet guide vane (IGV) actuator position, inlet temperature, shell temperature, shell pressure, eight blade path temperature sensors and eight exhaust temperature sensors. Two types of deviations were considered. In a step deviation, the deviation stays at value v during the faulty time; in a drift deviation, the deviation linearly varies from 0 to v during the faulty time. We compared our SSAR model with SVR, GMM [3], hidden Markov model (HMM) and the SAR model [4]. The HMM was adapted from the SSAR model by removing dependency $\mathbf{x}_{t-1} \rightarrow \mathbf{x}_t$ and $s_{t-1} \rightarrow \mathbf{x}_t$ in Fig.1. For SVR, GMM and HMM, both $\mathbf{y}_t$ and $\mathbf{x}_t$ are 22-dimensional vectors. For our SSAR model and SAR model, each $\mathbf{x}_t$ is a 44-dimensional vector containing data at t and $t-1$; thus, the observation matrix $\mathbf{C}$ is a 22×44 matrix. The HMM and SAR model were placed similarly in Fig.1 whose deviation covariance $\mathbf{\Theta}_t$ was adaptively estimated; this improved the performances of both of them.

During training, data for each load mode were extracted by applying an 1D edge detection algorithm to the IGV actuator position sensor data. The remaining data were assigned to the transition mode. These partitioned training data were used to train GMM, HMM, SAR and SSAR models. A transition mode cannot be well represented by a STAR model, since its data contain very high frequency. Therefore, we removed time dependency from its STAR model by setting $\mathbf{A}_j = 0$ and $\mathbf{Q}_j = \mathbf{V}_j$. During monitoring, because any sensor can experience deviations, the 1D edge detection used in training is not reliable and is not used.

The tests are now detailed. We selected eight normal data sets, each from a different power plant. The first 70 percent of each four-month data set was used for training and the rest was for testing (monitoring). Data for each sensor were normalized using the mean and standard deviation computed from the training data. The test data were evenly split into two parts; faults were added to the second part and the first part was left intact. For each randomly selected faulty sensor, we randomly picked a deviation type and randomly generated v within $\pm 10\sigma$ (where σ is this sensor's standard deviation) and then added this deviation to the sensor in the faulty time range. Such a trial was repeated four times. The number of faulty sensors was also varied from one to four. There are thus a total of 8 (data sets) $\times 4$ (trials) $\times 4$ (faulty sensors) = 128 tests.

Fig.2 shows the comparison results of the SSAR model vs. other methods in terms of different estimation errors. The scores of GMM are very similar to those of HMM and are thus omitted. There are 128 (tests) $\times$ 3 (other methods) = 384 points in the E_{nf} and E_{ff} plots, but there are only 8 (data sets) $\times$ 3 (other methods) = 24 points in the E_n plot, because E_n does not vary with different faults. Note that certain points representing the SVR and SAR models are located outside the upper border of some plots due to very large errors. The

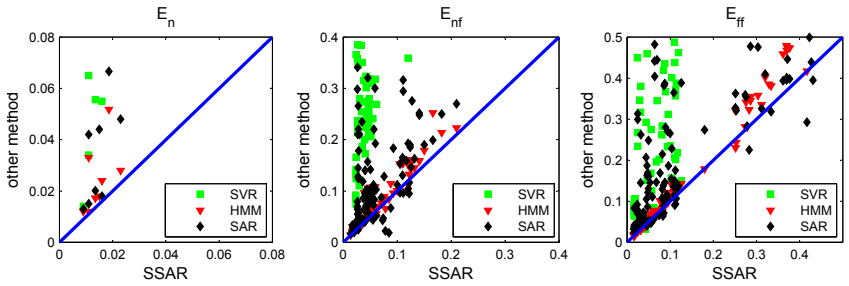

Fig. 2. Estimation errors of the SSAR model vs. the SVR, HMM and SAR models

SSAR model is shown to outperform all other methods, because the majority of points lie above the diagonal line (equal error line) in each error plot.

We adopt the Wilcoxon signed-ranks test to check whether our conclusion is statistically significant as suggested by [8]. In our case, the null hypothesis is that the SSAR model does not perform better than the other method. The p-values for all 3 (other methods) $\times$ 3 (error measures) = 9 hypothesis tests are below 0.0001. Therefore, we reject the null hypothesis and can claim that the SSAR model significantly outperforms the other method in terms of all three error measures.

5. DISCUSSION AND SUMMARY

We use two examples in Fig.3 to show how the SSAR model outperformed the SAR model. Fig.3 (top) shows the power sensor data of a normal data set, for which estimates should be close to observations. Both models produced nice estimates in the full load (power ≈ 225) and the part load mode (power ≈ 150). However, during mode switching (e.g., at data points 530, 595), the SAR model continued fitting the data with an AR model, leading to poor estimation results. Fig.3 (bottom) shows a real-fault case where blade path temperature sensors such as BPTC12A started drifting down at data point 170. The SAR model continued fitting faulty data. As a stationary model, our SSAR model stopped following and produced correct estimates within the normal operating range.

In a summary, we present a dynamic Bayesian model for sensor estimation in machine condition monitoring. The normal behavior of sensor signals is modeled by a stationary switching autoregressive model. Our SSAR model not only describes the dynamics of normal sensor signals but also represents the normal operating range. During monitoring, different levels of deviations are handled by the deviation covariance which is adaptively estimated via the EM algorithm.

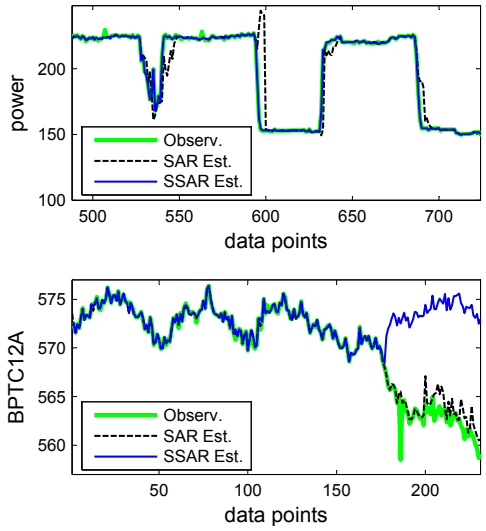

Fig. 3. Sensor estimation examples for the SAR and SSAR models.

6. REFERENCES

[1] J. W. Hines, A. Gribok, and B. Rasmussen, "On-line sensor calibration verification: A survey," in *Proceeding of International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management*, 2001.

[2] A. V. Gribok, J. W. Hines, and R. E. Uhrig, "Use of kernel based techniques for sensor validation in nuclear power plants," in *Proceeding of International Topical Meeting on Nuclear Plant Instrumentation, Controls and Human-Machine Interface Technologies*, 2000.

[3] C. Yuan and C. Neubauer, "Bayesian sensor estimation for machine condition monitoring," in *IEEE Conf. on Acoustics, Speech and Signal Processing*, 2007.

[4] K. P. Murphy, "Switching Kalman filters," Tech. Rep. 98-10, Compaq Cambridge Research Lab, 1998.

[5] A. T. Cemgil, H. J. Kappen, and D. Barber, "A generative model for music transcription," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 14, pp. 679–694, March 2006.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, pp. 1–38, 1977.

[7] P. J. Brockwell and R. A. Davis, Eds., *Introduction to Time Series and Forecasting*, Springer-Verlag, 2002.

[8] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.