Article

Anomaly Detection on Gas Turbine Fuel System Using a Sequential Symbolic Method

Fei Li [1], Hongzhi wang [1,*], Guowen Zhou [2], Daren Yu [2], Jiangzhong Li [1] and Hong Gao [1]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China; feili_hit@yahoo.com (F.L.); lijzh@hit.edu.cn (J.L.); honggao@hit.edu.cn (H.G.)

[2] School of Energy Science and Engineering, Harbin Institute of Technology, Harbin 150001, China; 16S102117@stu.hit.edu.cn (G.Z.); yudaren@hit.edu.cn (D.Y.)

* Correspondence: wangzh@hit.edu.cn;

Abstract: Anomaly detection plays a significant role in helping gas turbines run reliably and economically. Considering collective anomalous data and both sensitivity and robustness of the anomaly detection model, a sequential symbolic anomaly detection method is proposed and applied to the gas turbine fuel system. A structural Finite State Machine is to evaluate posterior probabilities of observing symbolic sequences and most probable state sequences they may locate. Hence an estimating based model and a decoding based model are used to identify anomalies in two different ways. Experimental results indicates that these two models have both ideal performance overall, and estimating based model has a strong ability in robustness, while decoding based model has a strong ability in accuracy, particularly in a certain range of length of sequence. Therefore, the proposed method can well facilitate existing symbolic dynamic analysis based anomaly detection methods especially in gas turbine domain.

Keywords: gas turbine fuel system; anomaly detection; symbolic dynamic analysis; time series

1. Introduction

Gas turbine engines, one of the most sophisticated devices, performs an essential role in industry. However, gas turbine maintenance is a great challenge for detecting anomalies and eliminating faults since they always run in variable operating conditions that make anomalies seem like normal. So Engine Health Management (EHM) policy is implemented to help gas turbines run in reliable, safe and efficient states and facilitates operating economy benefits and security levels. [1, 2]. In the framework of EHM, many works have been devoted into anomaly detection and fault diagnosis on gas turbine engines. Since Urban [3] first got involved in EHM research, many techniques and method are subsequently proposed. Previous anomaly detection works mainly contains two categories, model based method and data driven based methods. Model based methods are typically includes linear gas path analysis [4–6], nonlinear gas path analysis [7, 8], Kalman filters [9, 10] and expert systems [11]. Data driven based methods typically include artificial neural networks [12, 13], support vector machine [14, 15], Bayesian approaches [16, 17], genetic algorithms [18] and fuzzy reasoning [19–21].

All the mentioned previous works are mainly based on simulation data or continuous observation data. Simulation data are sometimes too simple to reflect actual operating conditions while real data usually contain many interferences that make anomalous observation appear normal. So this is a consensus of challenge in anomaly detection, especially in gas turbines which are operating in sophisticated and severe conditions. There are two possible routes on improving anomaly detection performance for gas turbines, the first one is on perspective of anomaly data and the second one is on perspective of detecting model.

On the one hand, anomalies occurred in gas turbine operating usually involve collective anomaly. Collective anomaly is defined that if a collection of related data instances is anomalous with respect to the entire data set, it is termed a collective anomaly [22]. Each single data doesn't

seem as anomaly, but their occurrence together may be considered as anomaly. For example, when a combustion nozzle is damaged, one of the exhaust temperature sensors records a continuously lower temperature than other sensors. Collective anomalies have been explored for sequential data [23, 24]. The collective anomaly detection has been widely applied in several domains such as intrusion detection, commercial fraud detection, medical and health detection, etc. [22] In industrial anomaly detection, some of structural damage detection are applied by using Statistical [25], parametric statistical modeling [26], mixture of models [27], rule based models [28] and neural based models [29], which can sensitively detect highly complicated anomalies. However, in gas turbine anomaly detection, these methods have some common demerits. Data preprocessing such as feature selection and dimension reduction are highly complicated when confronting such devices' observation data and extremely affect the whole performance. Besides this, many interferences comprised in data such as ambient conditions, normal patterns changes, even sensors observing deviation offer extraneous information for detecting anomalies and usually conceal essential factors that may critically helpful for detection, e.g. small degeneration in gas turbine fuel system may be covered by normal flow changes, thus making obstructs in detecting anomalies in device's early faults.

On the other hand, detection model for gas turbine requires both sensitivity and robustness capabilities. The sensitivity ensures higher detection rate and the robustness ensures fewer misjudging. The symbolic dynamic filtering (SDF) [30] was proposed and yield good performance in anomaly detection for robustness in comparison to other method such as principal component analysis (PCA), ANN and Bayesian approach [31] as well as proper accuracies. Gupta *et al.* [32] and Sarkar *et al.* [33] presented a SFD- based model for detecting fault of gas turbine subsystem and used it to estimate multiple component faults. Sarkar *et al.* [34] proposed an optimized feature extraction method under the SDF framework [34]. And then they applied symbolic dynamic analysis (SDA) based method in fault detection in gas turbine. They proposed Markov based analysis in transient data during takeoff other than quasi-stationary steady-states data and validated the method by simulation on the NASA Commercial Modular Aero Propulsion System Simulation (C-MAPSS) transient test-case generator. However, current SDF-based models usually use simulating data or generated data in laboratories, especially in gas turbine domain. Performance of these methods are still unconfirmed in real data, for instance, from a long time operated gas turbine devices which contains lots of flaws and defects and the sensors may not always be available for data acquisition.

Considering the two solutions on improving the effect of anomaly detection for gas turbine with their advantages and disadvantages, in this paper, we combine these two strategies by building a SDA-based anomaly detection model and processing collective anomalous sequential data in order to establish more sensitive and robust models and to eliminate their intrinsic demerits, and then apply this method in anomaly detection for gas turbine fuel system.

In this paper, the observing data from offshore platform gas turbine engine are first partitioned into symbolic sequential data to construct SDA based model, finite state machine, which reflects texture of a system operating tendency. Then two methods, estimating based model and detecting based model, are proposed on basis of the sequential symbolic reasoning model, one is with high robustness performance and the other is more sensitive, and they can be well integrated in different practice scenes and meanwhile eliminates irrelevant interferences for detecting anomalies. The comparison among collective anomaly detection, symbolic anomaly detection and our method is concluded in Table 1.

This paper is organized in 6 sections. In section 2, preliminary mathematical theories on symbolic dynamic analysis are given. Then in section 3, data used in this study and symbol partition are introduced. The finite state machine training and the two anomaly detection models are proposed in section 4. Hence experimental results and comparison between two models from several perspectives are expatiated in section 5 and the conclusion and discussion are briefly presented in section 6.

Table 1. Comparison between different current methods and ours

	Collective anomaly detection	Symbolic anomaly detection	Our method
Strategy	Sequential based analysis;	Symbolic and semantic based analysis;	Sequential and symbolic integrated analysis;
Advantages	• Detecting complicated industrial anomalies; • High sensitivity;	• High robustness; • An easy operating vector representation in anomaly detection;	• Confront real operating data, guaranteeing both sensitivity and robustness; • High dimension data converted to simple individual symbol sequences; • Eliminating interferences.
Disadvantages	• High dimension data to be preprocessed; • Interferences influence detection remarkably;	• Most based on simulating data or experimental data, lack complexity and credibility in actual scenes;	

2. Preliminary mathematical theories

2.1. Discrete Markov Model

Consider a time series of states $\omega(t)$ in length T, marked with $\boldsymbol{\omega}^T = \{\omega(1), \omega(2), ..., \omega(T)\}$. A system could be in a same state in different time and it doesn't need to achieve all possible states at one time.

Stochastic series of state is generated through equation (1) which is named transfer probability:

$$a_{ij} = P\left(\omega_j(t+1) \middle| \omega_i(t)\right), \tag{1}$$

it represents a conditional probability that a system will transfer to state ω_j in the next time point when it is in state ω_i currently. a_{ij} is not relevant to the time. In Markov model diagram, each discrete state ω_i is denoted as a node and the line which links two nodes is denoted as transfer probability. A typical Markov model diagram is represented in figure 1. The system is in state $\omega(t)$ at moment t currently, while the state in moment (t+1) is a random function which is both related to the current state and the transfer probability. Therefore, a specific time series of states generated in probability is actually a consecutive multiply operation of each transfer probability in this series.

2.2. Finite State Machine

Assume that in moment t, a system is in state $\omega(t)$, and meanwhile the system activates a visible symbol $\upsilon(t)$. So a specific time series would activate a specific series of visible symbols: $\boldsymbol{\upsilon}^T = \{\upsilon(1), \upsilon(2), ..., \upsilon(T)\}$. In this Markov model, state $\omega(t)$ is invisible but activated visible symbol $\upsilon(t)$ can be observed. We defined this model is finite state machine (FSM) shown in figure 2.

The activation probability of a FSM is defined by equation (2),

$$b_{jk} = P\left(\upsilon_k(t) \middle| \omega_j(t)\right). \tag{2}$$

In this equation, we can only observe the symbol $\upsilon(t)$. In figure 2 it can be seen that the FSM has 4 invisible states linked with transfer probabilities and each of them can activate 3 different visible

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2017 doi:10.20944/preprints201704.0071.v1

Peer-reviewed version available at Energies 2017, 10, , 724; doi:10.3390/en10050724

symbols. The FSM is strictly subject to causality, which means the probability in the future depends on current probability conclusively.

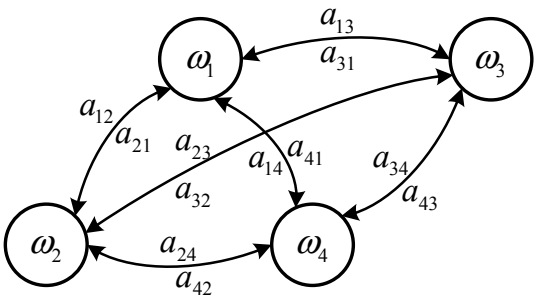


Figure 1. Discrete Markov Model

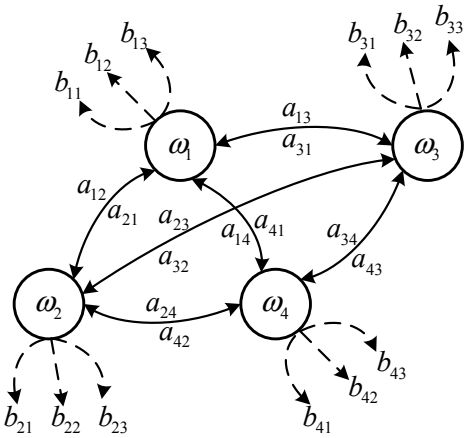


Figure 2. Finite state machine

3. Data and symbolization

In gas turbine monitoring, all operation data observed by sensors are continuous and there is no such an observation that can acquire several discrete operating conditions automatically. So in this section we focus on symbol extraction which can represents different load condition patterns discretely and after symbolization many irrelevant interferences are dismissed. In this paper, data recourse is from a SOLAR, Titan 130 Gas Turbine where the parameters to be used are listed in table 2 below and the structure of the turbine fuel system overview is shown in figure 3.

Many strategies for data symbolization or discretization have been proposed and are well discussed in papers [36-38]. In summary, two kind of approaches, splitting and merging, can be used in data symbolization. They discretizes a feature by either splitting the interval of continuous values or by merging the adjacent intervals [36]. It is very hard for splitting applied in our study since the interval we can't properly preset. Therefore, in this study, a simple merging-based strategy is used in data symbolization. We apply a cluster method to symbol extractions—K means cluster method. The main reason that in a FSM, number of the hidden states and visible symbols are both finitude and KM has a specific cluster boundary and expected cluster numbers initially. Samples in a cluster are regarded as the same symbol. There are 7 clusters that corresponds to 7 different load conditions, showed in table 3. So in this study K is equal to 7 for the KM model.

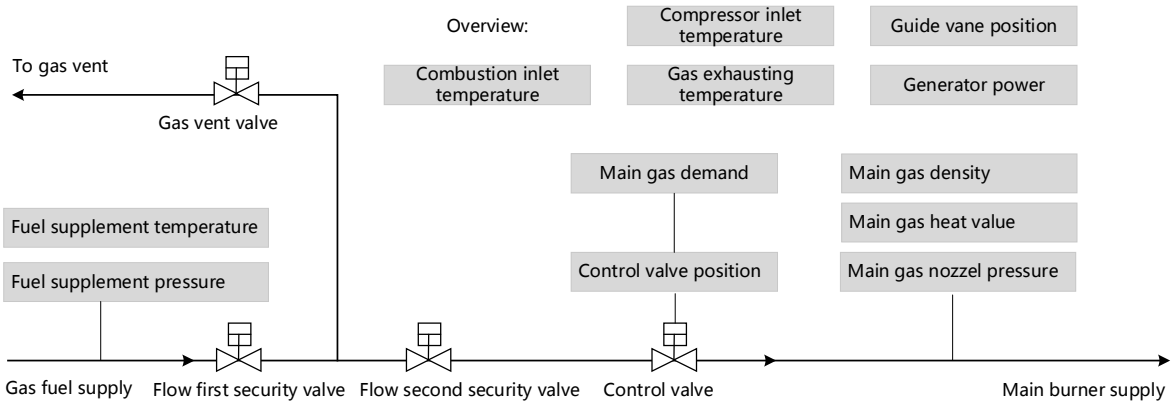


Figure 3. Overview on SOLAR Gas Turbine fuel system

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2017 doi:10.20944/preprints201704.0071.v1

Peer-reviewed version available at Energies 2017, 10, , 724; doi:10.3390/en10050724

Table 2. Monitoring sensors used in symbol extraction

Sensor No.	Parameters	Unit
92	Average gas exhaust temperature	℃
84	Average fuel supplement temperature	℃
81	Main gas supplement pressure	kPa
85	Main gas valve demand	%
98	Average combustion inlet temperature	℃
148	Generator power	KW
88	guide vane position	%
82	Main gas nozzle pressure	kPa
95	Main gas heat value	kJ/kg
96	Main gas density	kg/m^3
112	Average compressor inlet temperature	℃

Table 3. Corresponding list between clusters and labels

Cluster No.	Label	Symbol
#0	Flow rise in low load	FRLL
#1	Steady in high load	SHL
#2	Steady in low load	SLL
#3	Fast changing condition	FCC
#4	Flow rise in high load	FRHL
#5	Flow drop in high load	FDHL
#6	Flow drop in low load	FDLL

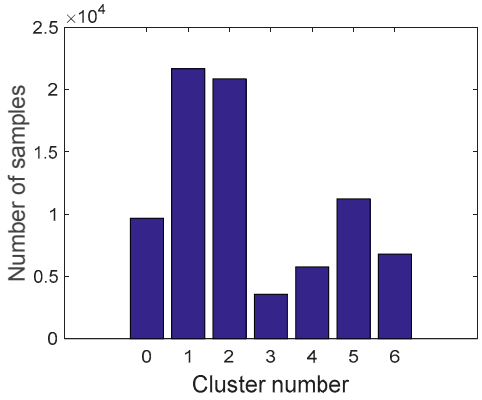


Figure 4. Distribution on total symbol data

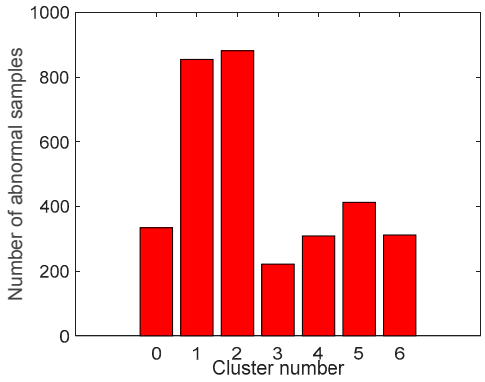


Figure 5. Distribution on anomaly data

The purpose of KM method is to divide original samples into k clusters, which ensure the high similarity of the samples in each cluster and low similarity among different clusters. The main procedures of this method are as follows:

First, choose k samples in dataset as the initial center for each cluster, and then evaluate distance between the rest samples and every cluster centers. Each sample will be assigned to a cluster to which the sample is closest. Second, renew the clusters center through those samples nearby and then reevaluating distance. Repeat this process until cluster centers converge. Generally, distance evaluation is used in Euclidean distance and the convergence judgement is square error criterion, as shown in equation (3):

$$\mathbf{E} = \sum_{i=1}^{k} \sum_{p \in C_i} \left| \mathbf{p} - \mathbf{m}_i \right|^2 , \tag{3}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2017 doi:10.20944/preprints201704.0071.v1

Peer-reviewed version available at Energies 2017, 10, , 724; doi:10.3390/en10050724

where E is sum error of total samples, P is the position of the samples and $\mathbf{m}_i$ is the center of cluster C_i. Iteration terminates when E becomes stable, which is less than 1e-6. Therefore, the final clusters and samples are converted from continuous variable to discrete visible symbols. In this study, altogether 79580 samples that are divided in to 7 symbols. The sampling interval is 10min and symbol extraction results are shown in figure 4. In figure 4, the most frequent symbols are Steady in high load (SHL) and Steady in low load (SLL). The intermediate ones are Flow rise in low load (FRLL), Flow rise in high load (FRHL), Flow drop in low load (FDLL) and Flow drop in high load (FDHL). The least frequent symbol is Fast changing condition (FCC). There are 3327 samples labeled in anomaly and the distribution of the abnormal samples is shown in figure 5. All of the abnormal labels are derived from the device operation logs and maintenance records manually. To simplify our modeling, we classify all kinds of anomalies or defects in records into one class – anomaly.

After symbolization, we use these symbols generated by KM cluster method to construct a FSM to measure a series of symbols whether they are on normal load conditions or on an abnormal one.

4. FSM modeling and anomaly detection methods

Finite state machine, widely applied in symbolic dynamic analysis, has shown great superiority in comparison to other techniques [30, 31]. Therefore, the main tasks of establishing a sequential symbolic model in gas turbine fuel system anomaly detection are to build a FSM to estimate posterior probabilities of sequences and build detection models to define whether a sequence is abnormal. But before this, the discrete symbols extracted in section 3 need to be sequenced into series, time series segments, in length T.

For FSM modeling, one aspect of this work is to determine several parameters. There are 5 states defined in this model, normal state (NS), anomaly state (AS), turbine startup (ST+), turbine shutdown (ST-) and halt state (HS). Actually, anomaly is a hidden state that usually invisible in operating compared to the other 3 states. Till now, we have defined the model structure of the FSM which contains 5 states and 7 visible symbols. In each state on a moment, there are 7 possible symbols could be observed in probability b_{jk} and a_{ij} is the probability that state transfers from one to another.

Another concerning about FSM modeling is how long the time series are, which can be efficiently involved in transfer and activation probabilities estimation. Figure 6 shows that different length T of segments may lead to different results in classification labels. A segment is defined in anomaly only if it contains at least one abnormal sample so that the time series in T=5 and T=10 are significantly different.

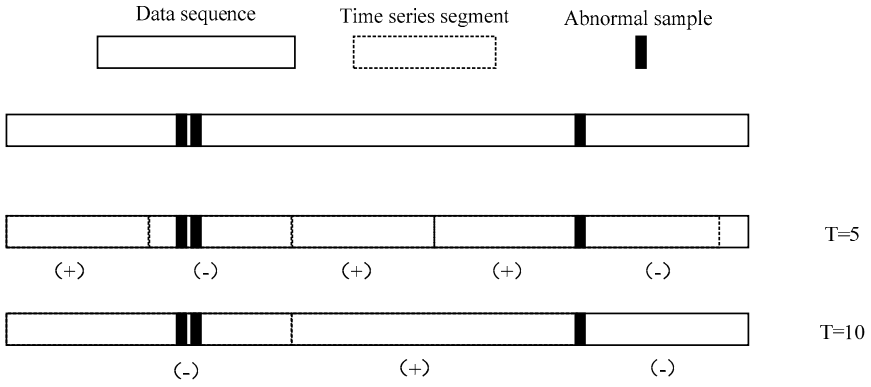


Figure 6. Time series generation with different length. Mark (+) denotes the normal series and mark (-) denotes the abnormal series.

It is hard to primordially define a proper T that can be well applied in parameter estimation and anomaly detection, so another way we take is to optimize length T recursively until it has the best performance in anomaly detection section. We declare that in actual data preprocessing, time series are generated by sliding window with length T in order to ensure that the size of time series can matches the original dataset, which means the total number of time series is 79581-T.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2017 doi:10.20944/preprints201704.0071.v1

Peer-reviewed version available at Energies 2017, 10, , 724; doi:10.3390/en10050724

For anomaly detection, two strategies are used, estimating based model and decoding based model. Estimating based model is that we use FSM to calculate probability of a symbol sequence. In this case, FSM is built by the training data precluding abnormal samples, so the abnormal sequences contained in testing data will be in very low probabilities through calculating. Decoding based model is that we use a FSM to decode a most probable state sequence which generates an observed symbol sequence. If the estimated state sequence contains anomaly state (AS), the symbol sequence is judged to anomaly.

According to aforementioned contents, the schematic of a FSM modeling is expatiated in figure 7, as illustrated below. First, data are sequenced into a pool of time series by initial sliding window with length T. Then data are divided into two parts, training data and testing data. Training data are used to construct a FSM, to estimate transfer and activation probabilities of unknown states and visible symbols. Testing data are used to evaluate performance of FSM. After modeling the FSM, performance will be tested through different anomaly detection strategies, decoding based model and estimating based model. Hence updating length T until the models achieve the best performance as well as FSM.

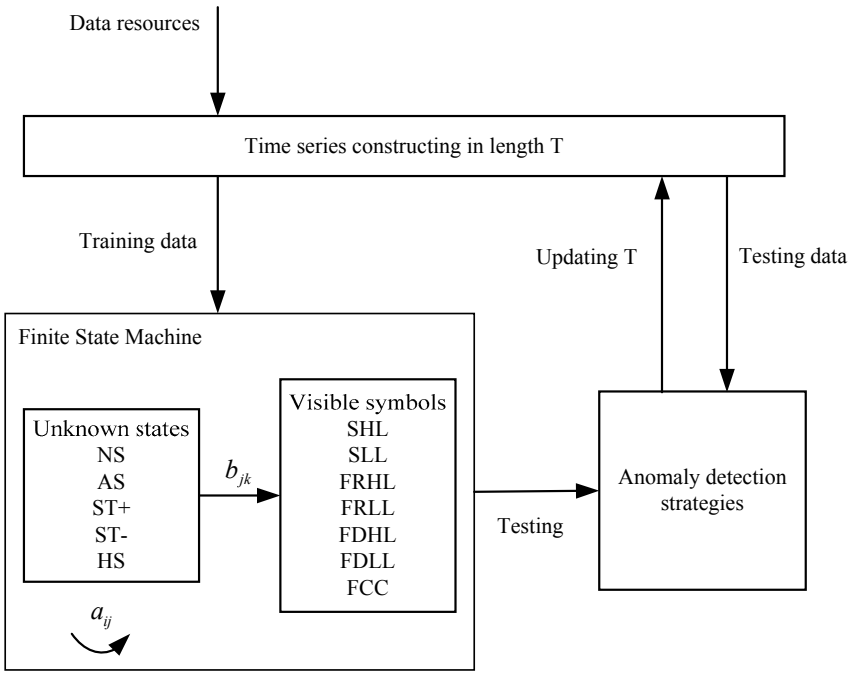


Figure 7. Schematic of FSM modeling procedures

4.1. Training a FSM

The main task of training a FSM is to estimate a group of transfer probabilities a_{ij} and activation probabilities b_{jk} over a pool of training samples. The normalized premise of a_{ij} and b_{jk} is that:

$$\begin{aligned} \sum_j a_{ij} &= 1, \text{for all } i \\ \sum_k b_{jk} &= 1, \text{for all } j \end{aligned} . \tag{4}$$

In this study, Baum-Welch algorithm [39] is used in estimation. Define a forward recursive equivalence in equation (5), where $\alpha_i(t)$ denotes the probability in state ω_i on moment t which has generated t symbols of the sequence $\mathbf{v}^T$ before. a_{ij} is transfer probability from state $\omega_i(t-1)$ to state $\omega_j(t)$ and b_{jk} is the probability of activated symbol v_k on state $\omega_j(t)$.

$$\alpha_j(t)=\begin{cases}1 & t=0\\ \sum_i \alpha_i(t-1)a_{ij}b_{jk} & \end{cases}. \tag{5}$$

Define a backward recursive equivalence in equation (6) either, where $\beta_i(t)$ denotes the probability in state ω_i on moment t which generates symbols of the sequence $\mathbf{v}^T$ from moment t+1 to T. a_{ij} is transfer probability from state $\omega_i(t)$ to state $\omega_j(t+1)$ and b_{jk} is the probability of activated symbol v_k on state $\omega_j(t+1)$.

$$\beta_i(t)=\begin{cases}1 & t=T\\ \sum_j \beta_j(t+1)a_{ij}b_{jk} & \end{cases}, \tag{6}$$

the parameters a_{ij},b_{jk} in the above 2 equations remain unknown so an expectation maximized strategy can be used in estimation. According to equation (5) and (6), define the transfer probability $\gamma_{ij}(t)$ that is from state $\omega_i(t-1)$ to state $\omega_j(t)$ under the condition of observed symbol v_k:

$$\gamma_{ij}(t)=\frac{\alpha_i(t-1)\alpha_{ij}b_{jk}\beta_j(t)}{P(\mathbf{v}^T|\boldsymbol{\theta})}, \tag{7}$$

where $P(\mathbf{v}^T|\boldsymbol{\theta})$ is the probability of the symbol sequence $\mathbf{v}^T$ generated through any possible state sequence. Expectation transfer probability through a sequence from state $\omega_i(t-1)$ to state $\omega_j(t)$ is $\sum_{t=1}^{T}\gamma_{ij}(t)$ and the total expectation transfer probability from state $\omega_i(t-1)$ to any state is $\sum_{t=1}^{T}\sum_k \gamma_{ik}(t)$. Therefore, an estimated transfer probability $\hat{\alpha}_{ij}$ is described as:

$$\hat{\alpha}_{ij}=\frac{\sum_{t=1}^{T}\gamma_{ij}(t)}{\sum_{t=1}^{T}\sum_k \gamma_{ik}(t)}. \tag{8}$$

Similarly, estimation of activation probability $\hat{b}_{jk}$ is described as:

$$\hat{b}_{jk}=\frac{\sum_{t=1}^{T}\gamma_{jk}(t)}{\sum_{t=1}^{T}\sum_l \gamma_{jl}(t)}, \tag{9}$$

where expectation activation probability of $v_k(t)$ on state $\omega_j(t)$ is $\sum_{t=1}^{T}\gamma_{jk}(t)$ and the total expectation activation probability on state $\omega_j(t)$ is $\sum_{t=1}^{T}\sum_l \gamma_{jl}(t)$, l is number of symbols.

According to the analysis above a_{ij},b_{jk} can be gradually approximated by $\hat{\alpha}_{ij},\hat{b}_{jk}$ through equations (8) and (9) until converged. Pseudocode of the estimation algorithm is shown in algorithm 1 below. The first a_{ij},b_{jk} are generated randomly at beginning. Hence evaluate $\hat{\alpha}_{ij},\hat{b}_{jk}$ using a_{ij},b_{jk} estimated in the former generation. Repeat this process until residual between $\hat{\alpha}_{ij},\hat{b}_{jk}$ and a_{ij},b_{jk} is less than a threshold ε and then the optimized a_{ij},b_{jk} are used in the FSM.

4.2. Anomaly detection based on estimating strategy

An estimating strategy is used in anomaly detection over FSM, in this part, which is inspired from anomaly detection approaches. Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior [22]. FSM estimating strategy is to calculate posterior probabilities of the symbol sequences generated by FSM. It can detect anomalies efficiently if it has been built by normal data. Therefore, this strategy conforms to the basic idea of what anomaly detection does. In this strategy thereby, FSM is utilized to establish an expected pattern and estimating process is to find out the nonconforming sequences, in other word, anomalies. Figure 8 illustrates the schematic of anomaly detection using estimating strategy. In order to establish an expected normal pattern, training data used in FSM modeling are utterly normal sequences.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2017 doi:10.20944/preprints201704.0071.v1

Peer-reviewed version available at Energies 2017, 10, , 724; doi:10.3390/en10050724

Parameters a_{ij}, b_{jk} are estimated by FSM training, which indicates the intrinsic normal pattern recognition capability contained in these parameters. Hence, estimate probabilities of testing symbol sequences in which abnormal sequences are included. Probabilities in normal sequences will be much higher than that in abnormal ones. So detecting indicator is actually a preset threshold that judges the pattern of test sequences.

Algorithm 1. Procedures of a_{ij}, b_{jk} estimation

Input:
Initial parameters $a_{ij}(0), b_{jk}(0)$, Training set $\mathbf{v}^T$, convergence threshold ε, $z \leftarrow 0$

Output:
Finial estimated FSM parameters $\hat{\alpha}_{ij}, \hat{b}_{jk}$

1 **Loop** $z \leftarrow z+1$
2 Estimating $\hat{a}_{ij}(z)$ by $a(z-1), b(z-1)$ in equation (8)
3 Estimating $\hat{b}_{jk}(z)$ by $a(z-1), b(z-1)$ in equation (9)
4 $a_{ij}(z) \leftarrow \hat{a}_{ij}(z)$
5 $b_{jk}(z) \leftarrow \hat{b}_{jk}(z)$
6 **Until** $\max_{i,j,k}\left[a_{ij}(z)-a_{ij}(z-1); b_{jk}(z)-b_{jk}(z-1)\right] < \varepsilon$
7 **Return** $a_{ij} \leftarrow a_{ij}(z)$, $b_{jk} \leftarrow b_{jk}(z)$

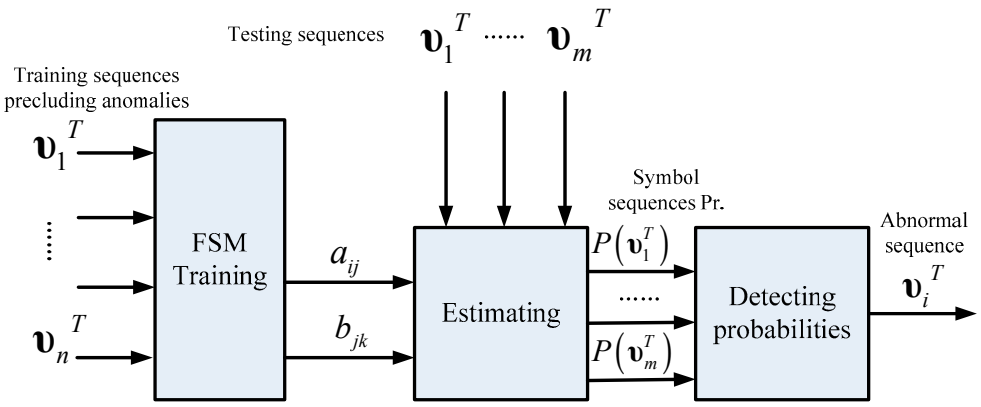


Figure 8. Schematic of anomaly detection using estimating strategy

The probability of a symbol sequence generated by a FSM can be described as:

$$P(\mathbf{v}^T) = \sum_{r=1}^{r_{\max}} P(\mathbf{v}^T | \boldsymbol{\omega}_r^T) P(\boldsymbol{\omega}_r^T), \tag{10}$$

where r is the index of state sequence in length T, $\boldsymbol{\omega}_r^T = \{\omega(1), \omega(2), ..., \omega(T)\}$. If there are, for instance, c different states in this model, the total number of possible states sequence is $r_{\max} = c^T$. Enormous amount of possible state sequences need to be considered to calculate probability of a generated symbol sequence $\mathbf{v}^T$, as shown in equation (10). The second part of the equation can be descried as:

$$P(\boldsymbol{\omega}_r^T) = \prod_{i=1}^{T} P(\omega(t) | \omega(t-1)). \tag{11}$$

That probability $P(\boldsymbol{\omega}_r^T)$ is continuous and chronological multiplication. Assume that activation probability of a symbol critically depends on current state, so the probability $P(\mathbf{v}^T | \boldsymbol{\omega}_r^T)$ can be described in equation (12) and then equation (13) is the other description of equation (10):

$$P\left(\mathbf{v}^T \left| \boldsymbol{\omega}_r{}^T \right.\right) = \prod_{t=1}^{T} P\left(v(t)\left|\omega(t)\right.\right), \tag{12}$$

$$P\left(\mathbf{v}^T\right) = \sum_{r=1}^{r_{\max}} \prod_{t=1}^{T} P\left(v(t)\left|\omega(t)\right.\right) P\left(\omega(t)\left|\omega(t-1)\right.\right). \tag{13}$$

However, computing cost of the above equation is $O\left(c^T T\right)$ which is too high to finish this process. There are two alternative methods that can dramatically simplify computing process, forward algorithm and backward algorithm, which are depicted through equation (5) and (6), respectively. Computing cost of these algorithms are both $O\left(c^2 T\right)$, c^{T-2} times faster than original strategy. Algorithm 2 shows detection indicator based on forward algorithm. Initialize a_{ij}, b_{jk}, training sequences $\mathbf{v}(t)$ precluding anomalies, $\alpha_j(0) = 1$, t=0. Hence update $\alpha_j(t)$ until t=T and probability of $\mathbf{v}(t)$ is $\alpha_j(T)$. If the probability is higher than the preset threshold, then the symbol sequence is classified into positive, normal class. Otherwise the symbol sequence is classified into negative, anomalous class.

Algorithm 2. Anomaly detection based on estimating strategy

Input:
$t \leftarrow 0$, a_{ij}, b_{jk}, sequence $\mathbf{v}(t)$, $\alpha_j(0) = 1$, threshold θ
Output:
Classification result $Class\left(\mathbf{v}(t)\right)$

For $t \leftarrow t+1$
  $\alpha_j(t) \leftarrow b_{jk} v(t) \sum_{i=1}^{c} \alpha_i(t-1) a_{ij}$
Until $t = T$
Return $P\left(\mathbf{v}(t)\right) \leftarrow \alpha_j(T)$
If $\alpha_j(T) > \theta$
  $Class\left(\mathbf{v}(t)\right) \leftarrow Positive$
Else $Class\left(\mathbf{v}(t)\right) \leftarrow Negative$
Return $Class\left(\mathbf{v}(t)\right)$

There are two points to be mentioned. In the first place, decision judgement is very simple for use of anomaly detection by threshold θ. This convenience is ascribed to normal pattern constructed by FSM and the probabilities estimated from FSM intuitively simulate the possibilities of the symbol sequences emerging in real system. Furthermore, performance of this strategy is not only dependent to efficiency of the trained FSM, but also depends on the proper threshold. Therefore, in modeling process, we optimized the threshold by traversing different values in order to get the highest overall accuracy of normal and abnormal sequences.

4.3. Anomaly detection based on decoding strategy

Comparing to estimating strategy, a FSM can detect anomalies either if it can recognize whether there are hidden anomalous states in sequences. Fortunately decoding a state sequence is available in FSM so another way to detect anomalies is to decode a symbol sequence to state sequence and then to find whether the state sequence contains abnormal states. Unlike previous estimating strategy, decoding strategy is a sort of, somehow, optimization algorithm in which searching function is used. Figure 9 illustrates the schematic of anomaly detection using decoding strategy. In this strategy, FSM is trained by all sequences that contain both normal and abnormal sequences so the FSM reflects the system that may run on normal pattern or occur anomalies. After that, decoding process is to search the most probable state sequence that the symbol sequence corresponds to. A

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2017 doi:10.20944/preprints201704.0071.v1

Peer-reviewed version available at Energies 2017, 10, , 724; doi:10.3390/en10050724

greedy based method is applied in searching that in each step the most possible state is chosen and added to path and the final path is the decoded state sequence. At last, judge a symbol sequence by its state sequence on whether it contains anomaly states.

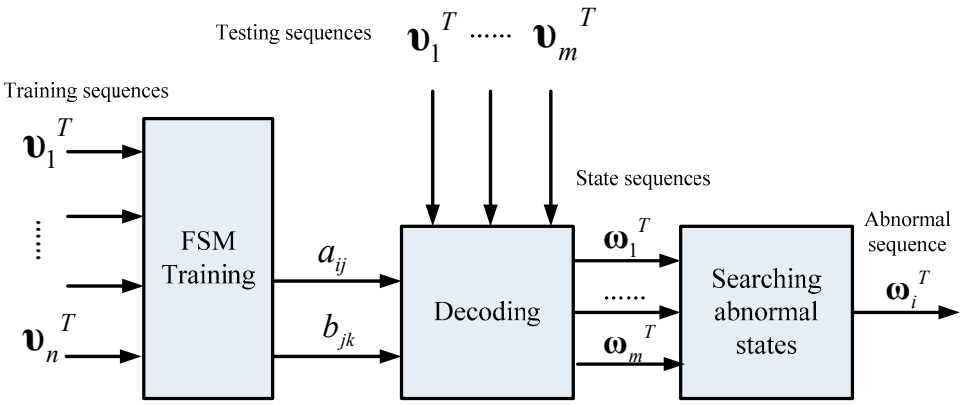


Figure 9. Schematic of anomaly detection using decoding strategy

Algorithm 3 shows the procedures of the decoding based anomaly detection strategy. Initialize parameters a_{ij}, b_{jk}, testing sequence $\upsilon(t)$, $\alpha_j(0)=1$, and path. Hence, update $\alpha_j(t)$ and in each moment t, traverse all state candidates and the most possible state in this moment is the one who makes $\alpha_j(t)$ biggest and then add the state into the path until the end of sequence. After that, scan the decoded state sequence if there is at least one anomaly state (AS), the observed sequence is classified into negative, anomalous class, otherwise classified into positive, normal class.

Algorithm 3. Anomaly detection based on decoding strategy

Input:

$t \leftarrow 0$, a_{ij}, b_{jk}, $\upsilon(t)$, $\alpha_j(0)=1$, $path \leftarrow \{\ \}$

Output:

Classification result $Class(\upsilon(t))$

For $t \leftarrow t+1$

$j \leftarrow 1$

For $j \leftarrow j+1$

$\alpha_j(t) \leftarrow b_{jk}\upsilon(t)\sum_{i=1}^{c}\alpha_i(t-1)a_{ij}$

Until $j=c$

$j' \leftarrow \arg\max_j \alpha_j(t)$

Add $\omega_{j'}$ to $path$

Until $t=T$

If $path$ contains state (AS)

$Class(\upsilon(t)) \leftarrow Negative$

Else $Class(\upsilon(t)) \leftarrow positive$

Return $Class(\upsilon(t))$

Compared to estimating strategy, there are some advantages and disadvantages on decoding strategy. First of all, detection indicator based on decoding strategy may help the anomaly detection system improve sensitivity, which means it can precisely warn an anomaly once it occurs. And besides, it helps system locate anomaly emerging time in a high resolution. For example, a sequence in T=10 and the sample interval is 10 min so the length of the sequence is 100 min. If the sequence is abnormal and the two system, estimating based model and decoding based model, both alert, the later can provide a more precise anomalies occurring time, for instance, in 20 min and 60 min possibly and the former can only provide a probability that an anomaly may have occurred.

However, this point may be, somehow in another aspect, a disadvantage because of lacking robustness of decoding strategy. This detection strategy is a local optimization algorithm that may reach a local minimum point that actually not global optimized solution. It may have searched an utterly different state sequence than the real one. Furthermore, error rate accumulates with the growth of searching path, particular in longer length T, and the false positive rate may be very high, which means many normal sequences are classified into anomalies. So the length of sequence is a critical impact on detection performance. This issue will be analyzed in the next section.

5. Experiments design and results analysis

5.1. Performance evaluation

Data used in this section is from gas turbine power generator on offshore oil platform, SOLAR Titan 130. Data list and symbol distribution are shown in table 2 and figure 4 earlier. Overall (79581-T) sequences are used in training and testing by using cross validation. The dataset was divided equally and sequentially into ten folders. Each folder was used in turn as testing data, with the other nine as training data, until every folder had been tested by the others. Hence the final result consists of a performance average together with a standard deviation.

The labels of the samples are grouped into positive (normal) and negative (anomaly). The detection performance is measured by the true positive rate (TP rate) and true negative rate (TN rate) [40]. Table 4 shows the definition of confusion matrix which measures 4 possible outcomes.

The TP rate is defined as the ratio of the number of samples correctly classified as positive to the number of samples that are actually positive:

$$\mathrm{TPrate} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}. \tag{14}$$

The TN rate represents the ratio of the number of samples correctly classified as negative to the number of actual negative samples:

$$\mathrm{TNrate} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}. \tag{15}$$

Table 4. Definition on confusion matrix

	Detecting Positive	**Detecting Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

In this paper, anomaly detection on gas turbine fuel system is a class imbalanced problem that normal class is much bigger than abnormal class. So performance of imbalanced data can be measured by AUC that is area under ROC curve, as shown in figure 10. ROC curve is a plot of the FP rate on the X axis versus the TP rate on the Y axis. It shows the differences between the FP rate and the TP rate based on different rules.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2017 doi:10.20944/preprints201704.0071.v1

Peer-reviewed version available at Energies 2017, 10, , 724; doi:10.3390/en10050724

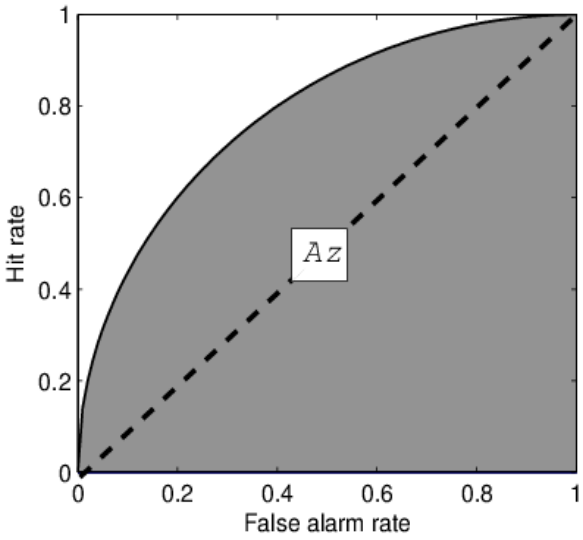


Figure 10. Schematic diagram of AUC. Surface of the area is Receiver Operating Characteristic curve and the area of the shadow is AUC.

5.2. Exemplary cases and experimental results

To illustrate how the detecting system works, two exemplary cases are used to be expatiated. One is a normal sequence and another is abnormal sequence that is in fuel nozzle valve anomaly, which may cause fuel flow fluctuation or drop and in turn cause load drop. The selected sequences, after normalization, are in length T=10, 10 min sampling interval, total 100 min and the anomalous sequence is shown in figure 11, the red sequence, original parameters 'Main gas valve demand' and 'Power'. The two sequences, partitioned into symbols, are described in figure 12. The routine map is consisted by 7 different symbols and 10 time points and the black row is normal sequence and the red dash row is anomalous sequence. Besides, the labeled state sequence of normal and anomalous cases are {NS, NS, NS, ST-, ST-, NS, ST-, ST-, NS, ST-}.and {NS, NS, NS , NS , NS ,AS, AS, AS, NS, NS}, respectively. The results are subsequently made by the trained FSM and the two detecting models, estimating and decoding based models. Threshold θ in this model is 0.00743 and the posterior probabilities of normal and anomalous sequences made by estimating based model are 0.01563 and 0.0009235. The decoded most probable state sequences made by decoding based model are {NS, NS, NS, NS, ST-, ST-, ST-, NS, ST-, NS}.and {NS, NS, NS, NS, NS, AS, AS, AS, AS, NS} and the anomalous sequence contains AS, yet the normal doesn't. As a result, these two models have both made correct decisions.

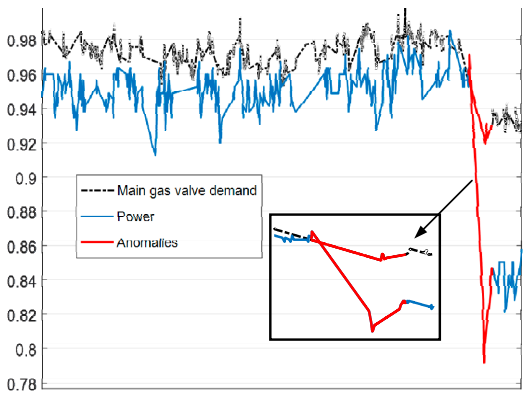


Figure 11. An exemplary anomaly sequence

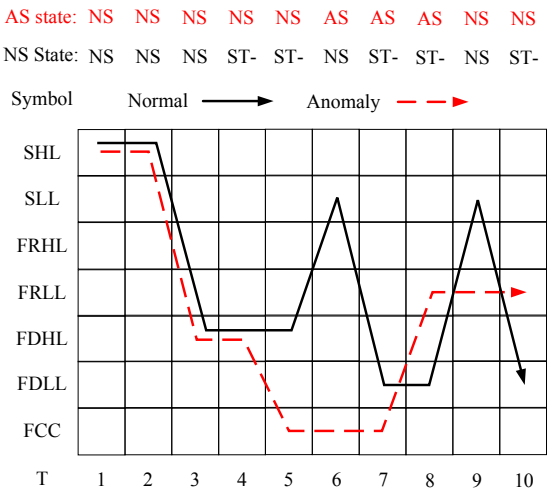


Figure 12. Comparison between a normal sequence and an anomaly sequence in symbolic description

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2017 doi:10.20944/preprints201704.0071.v1

Peer-reviewed version available at Energies 2017, 10, , 724; doi:10.3390/en10050724

Performance of the two models in each data group within cross-validation are given in table 5. It can be seen from the table that the overall performance of estimating based model is better than that of decoding based model. Specifically, the estimating based model's evaluation in 8 groups for TP rate, 6 groups for TN rate and 8 groups for AUC are better than another model. Similarly, table 6 and table 7 show the evaluation of confusion matrixes for the two models. Results illustrate that the estimating based model outperform decoding based model in overall accuracy as well as deviation. Therefore, it can be concluded that in terms of T=10, θ=0.00743, the estimating based model can resolve anomaly detection in gas turbine fuel system more efficiently. However, as is analyzed earlier, threshold θ and length of sequence can drastically influence detecting performance. As a result, several impacts need to be proposed into further discussions.

Table 5. Performance of the two model in each data group

Models	**Estimating based model**			**Decoding based model**		
Measurement	**TP rate**	**TN rate**	**AUC**	**TP rate**	**TN rate**	**AUC**
1	0.9577	0.8977	0.9277	0.9478	0.9009	0.9244
2	0.9692	0.9052	0.9372	0.954	0.8938	0.9239
3	0.9549	0.906	0.9305	0.9393	0.9032	0.9213
4	0.9617	0.9111	0.9364	0.9336	0.8931	0.9134
5	0.9679	0.8986	0.9333	0.9334	0.9052	0.9193
6	0.9549	0.9043	0.9296	0.9567	0.9089	0.9328
7	0.9574	0.9039	0.9307	0.9445	0.9188	0.9317
8	0.9488	0.9101	0.9295	0.9444	0.8976	0.921
9	0.9699	0.908	0.939	0.9587	0.8861	0.9224
10	0.9442	0.9164	0.9303	0.9487	0.8905	0.9196
Mean value	0.9587	0.9061	0.9324	0.9461	0.8998	0.9229

Table 6. Performance of estimating based model

	Detecting Normal	**Detecting Anomaly**
Actual Normal	0.9587±0.0086	0.0413±0.0086
Actual Anomaly	0.0939±0.0056	0.9061±0.0056

Table 7. Performance of decoding based model

	Detecting Normal	**Detecting Anomaly**
Actual Normal	0.9461±0.0089	0.0539±0.0089
Actual Anomaly	0.1011±0.0097	0.8989±0.0097

5.3. Threshold determination strategy

Despite of training a FSM, a core problem of building a estimating based detection model is to determine a proper threshold θ. In a particular sequence length, we can find a most suitable value that can classify testing sequence well. Actually, with different thresholds, we will gain different classification results, as is shown in figure 13, which is a ROC curve of estimating based model with different thresholds. The optimized threshold is one of them on the curve.

A sequence is judged into anomaly when its posterior probability is less than θ. When θ=0, the TN rate, correct detected anomalous sequence among all anomalies, is 0 initially. And then TN rate increases with the growth of θ to 1 when θ reach to a certain point. On the contrary, the TP rate is 1 initially when θ=0 since all the posterior probabilities is over 0. And then TP rate starts to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2017 doi:10.20944/preprints201704.0071.v1

Peer-reviewed version available at Energies 2017, 10, , 724; doi:10.3390/en10050724

decrease with the growth of θ. This regular is illustrated in figure 14. One concern is that both TP rate and TN rate are important to the model performance, so the suitable threshold should be found in a synthesized highest place, which is measured by average accuracy that is the mean value of TP rate and TN rate. In this experiment, we searched the θ in step of 0.0001 until it reached the peak average accuracy when θ=0.0074.

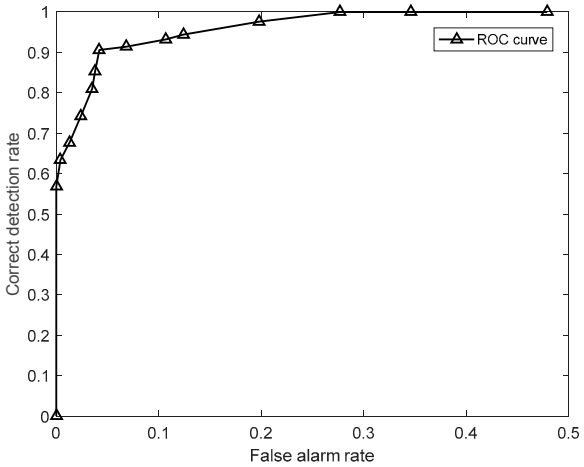


Figure 13. ROC curve of estimating based model with different thresholds

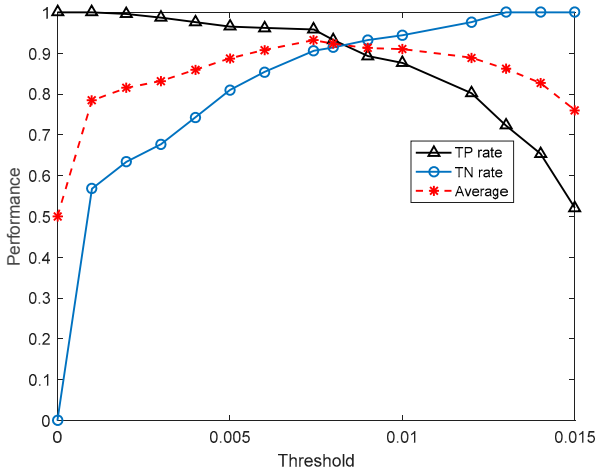
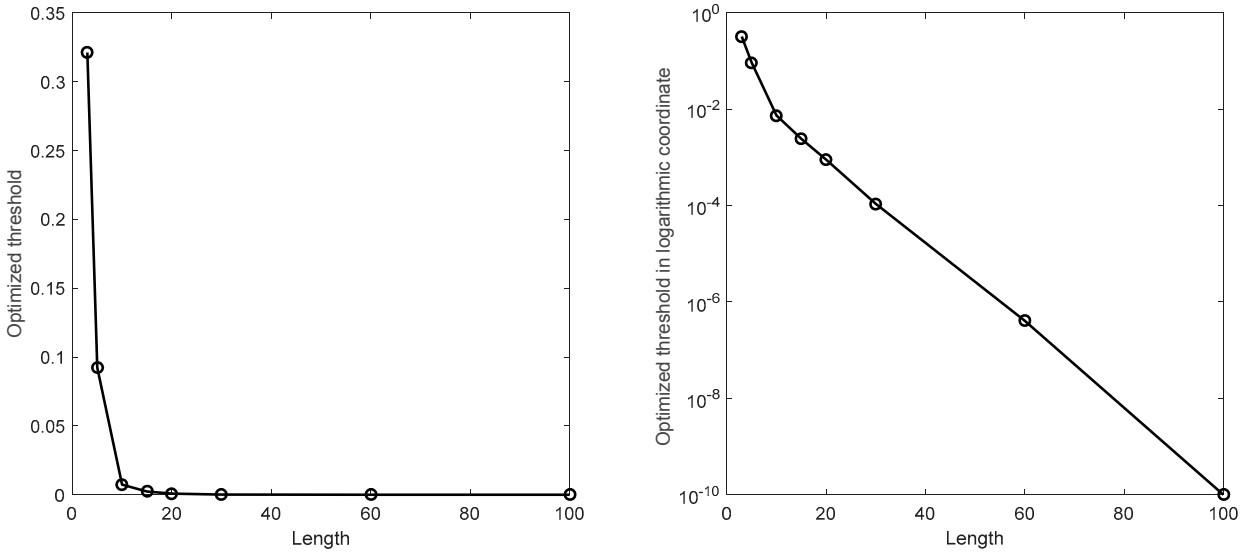


Figure 14. Performances changing with growth of threshold

The optimized threshold will drop along with the change of length T because the posterior probabilities are continuous multiplying so the probabilities drop exponentially when T increases. Figure 15 shows optimized thresholds in different length of the sequence. The sub-figure (1) is depicted in Cartesian coordinates and the sub-figure (2) is depicted in Logarithmic coordinate in Y axis. It can be clearly seen that the θ is in exponential tendency.



(1) Y axis is Cartesian coordinate (2) Y axis in Logarithmic coordinate

Figure 15. Optimized thresholds in different length of the sequence

5.4. Comparison between the two models on different length of sequence

Another main factor that influence performance of the detecting models is the length of sequence. Figure 16 and Figure 17 illustrate the comparison between estimating based model and decoding based model in TN rate and TP rate in different length of sequence, respectively. With the growth of length, TN rate and TP rate of estimating based model gradually decrease and then become stable after about T>60, steadying in 0.82 in TN rate and 0.89 in TP rate. However, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2017 doi:10.20944/preprints201704.0071.v1

Peer-reviewed version available at Energies 2017, 10, , 724; doi:10.3390/en10050724

performance of decoding based model seems unlikely to that of estimating based model. TN rate of decoding based model rises to nearly 1 when length T>30 and TP rate decreases drastically when T>20. The tremendous difference between the two models is ascribed in several reasons. First, detecting mechanism of the two models are not conformed. Estimating based model is built upon the premise of normal pattern which is constructed by a normal pattern based FSM while decoding based model is built upon the trained FSM containing normal and anomalous sequences. It points out that estimating based model concerns those data without anomalies while decoding based model concerns all the data whatever class they belong to. Second, detecting indicator used in estimating based model is posterior probabilities while detecting indicator used in decoding based model is states. With the growth of length T, the sequences become longer and more complicated for classification using threshold because there are more single symbols in a sequence. So the performance of estimating based model decreases and eventually can ensure steady because of a suitable threshold and classification capability of FSM. Regarding decoding based model, it is more complicated than estimating based model. With the growth of length T, the sequences become longer but it is easier to find a possible abnormal state. A sequence is judged into anomaly only if there can be found at least one abnormal state. The longer the sequence is, the more probable abnormal states hide, so the TN rate rises when length is growing. On the contrary, it can be easy to understand that long, complicated sequences would make the model's incorrect judgement because the more states to be decoded, the higher possible to misjudge. So the TP rate decrease along with the growth of T.

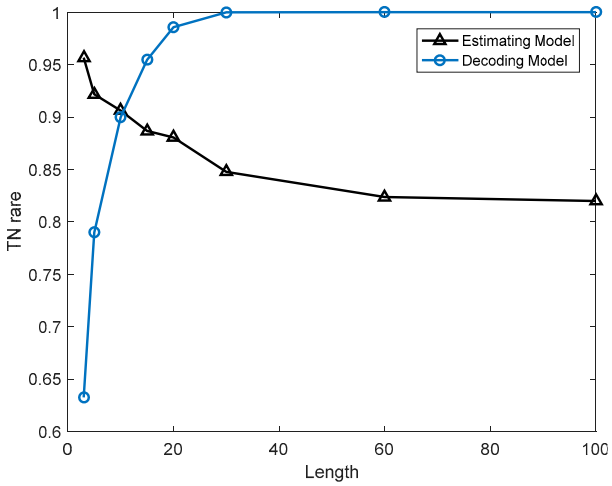


Figure 16. Comparison between estimating based and decoding based models on TN rate

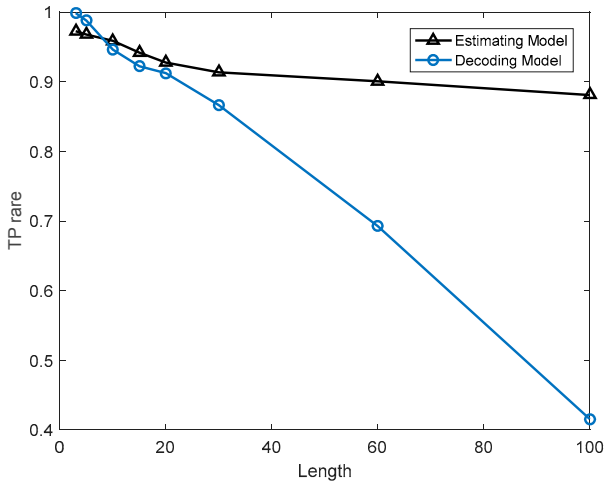


Figure 17. Comparison between estimating based and decoding based models on TP rate

Based on these points presented, we can in turn elicit a conclusion: the estimating based model has a better robust performance than decoding based model while the decoding based model may yield better performance in particular intervals. Figure 18 shows the comparison between estimating based model and decoding based model on AUC, which is an evaluation measurement that reflects overall classification performance of a model for class imbalanced problems. AUC of estimating based model gradually decreased as well as TP rate and TN rate along with the growing of length T. However, AUC of decoding based model rises to a peak value when T=20 and then start to drop. In this figure, we can see that before T=12, the performance of estimating based model is better than decoding based model. Between T=12 and 53, decoding based model outperform estimating based model and when T>53, estimating based model is much more efficient than decoding based model. In a conclusion, if we need a short term anomaly detection system, e.g. less than 1 hours observing window, the estimating based model will be satisfied. If a long term anomaly detection system is needed, e.g. 1 hours to 8 hours, that decoding based model will be efficient. And for an ultra-long term anomaly detection system, e.g. over 8 hours, there should be an estimating based model.

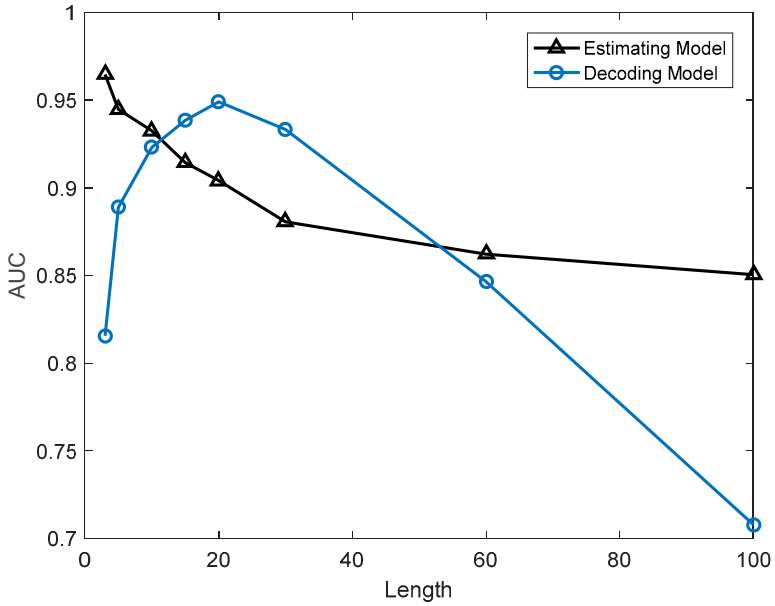


Figure 18. Comparison between estimating based model and decoding based model on AUC

6. Conclusion and discussion

The essential issue of anomaly detection is to detect when and how an anomaly would happen sensitively and effectively. Many conventional anomaly detection methods for points or collective anomalies are mostly established on continuous real-time sensor observations. Noise, operating fluctuation or ambient conditions may be all contained in raw data, which make anomalous observation appear normal. The anomalous features often hide in the structural data which reflect various patterns the device operating. Thus we first partitioned original data into classes by using K-means clustering and symbolized each class to construct a sequence based feature-structure which intrinsically on behalf of operating patterns of a device. When operating pattern changes in structural level, the device might occur anomaly that is easier to observe. So the proposed method in this paper satisfied the basic idea of anomaly detection, finding patterns in data that do not conform to expected behavior.

Second, we built the core computing unit for anomaly detection, finite state machine, on a large quantity of training sequential data and then used it to estimate posterior probabilities or find most probable states. And in turn two detection models were generated, estimating based model and decoding based model. They were quite different to each other. In estimating based model, FSM was used to calculate posterior probabilities for normal sequences, so it was trained by utterly normal data. When FSM was calculating probabilities for testing sequences that are anomalous data, the results would be extremely lower than that of normal data. Therefore, optimizing a proper threshold classifying two classes would be of great use for an efficient model performance. In decoding based model, FSM was used to find most probable state sequence of the observed sequence. The main task of FSM was to detecting anomalous state efficiently so it was trained by both normal and abnormal data. Once an anomalous state was detected, the sequence was judged to anomaly.

Third, these two models have their own advantages and weakness. Estimating based model has strong robustness since the performance is stable high when length of sequence is growing. However in decoding based model, performance varies on different length of sequence. The anomalous sequences are easier to be detected in longer length than in shorter one but there will be a high false alarm rate in longer length, which means many normal sequences are misclassified into anomalies. Thus the overall performance of decoding based model has a trend of rise first then fall. Besides, there is another merit of the decoding based model that it can help to locate precise anomaly occurring point in a high time resolution, which means it can give us what symbol points are in anomaly state precisely other than only probabilities of sequences made by estimating based model.

Experimental results indicated that estimating based model is more suitable for anomaly detection for gas turbine fuel system when observation window is less than 1 hour or over 8 hours. And decoding based model is more suitable when observation window is between 1 and 8 hours. So it may help people choose the most efficient detecting model for different demands.

Further work may be concentrated on algorithm optimizing and application extension. As is described above, decoding based model used local searching algorithm which may cause high deviation in some circumstances though it could be operated in a high computing speed. So the algorithms need to be optimized. We need also apply this method into other domains of gas turbine anomaly detection, such as gas path, combustion components, etc. All these need further attentions.

Acknowledgments: This paper was partially supported by National Natural Science Foundation of China (NSFC) grant U1509216, 61472099, National Sci-Tech Support Plan 2015BAH10F01, the Scientific Research Foundation for the Returned Overseas Chinese Scholars of Heilongjiang Province LC2016026 and MOE–Microsoft Key Laboratory of Natural Language Processing and Speech, Harbin Institute of Technology.

Author Contributions: Fei li developed the idea, performed the experiments and wrote the draft paper. Hongzhi Wang reviewed and supervised the whole paper. Guowen Zhou got involved in data processing and experimental analysis, Daren Yu and Jiangzhong Li significant improved the paper in scientific ideas and helped modify the paper. Hong Gao provided the raw data and involved in data preprocessing and improved the paper from algorithmic perspective.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Volponi, A. J. In *Gas Turbine Engine Health Management: Past, Present and Future Trends*, ASME Turbo Expo 2013: Turbine Technical Conference and Exposition, 2013; 2013; pp 433-455.
2. Marinai, L.; Probert, D.; Singh, R., Prospects for aero gas-turbine diagnostics: a review. *Applied Energy* **2004,** 79, (1), 109-126.
3. Urban, L. A., Gas Path Analysis Applied to Turbine Engine Condition Monitoring. *Journal of Aircraft* **1973,** 10, (7), 400-406.
4. Pu, X.; Liu, S.; Jiang, H.; Yu, D., Sparse Bayesian Learning for Gas Path Diagnostics. *Journal of Engineering for Gas Turbines & Power* **2013**, 135, (7), 071601.
5. Doel, D. L., TEMPER—A Gas-Path Analysis Tool for Commercial Jet Engines. *Journal of Engineering for Gas Turbines & Power* **1992**, 116, (1), V005T15A013.
6. Gulati, A.; Zedda, M.; Singh, R. In *Gas turbine engine and sensor multiple operating point analysis using optimization techniques*, Aiaa/asme/sae/asee Joint Propulsion Conference and Exhibit, 2000; 2000; pp 323-331.
7. Mathioudakis, K., Comparison of Linear and Nonlinear Gas Turbine Performance Diagnostics. *Journal of Engineering for Gas Turbines & Power* **2003**, 127, (1), 451-459.
8. Stamatis, A.; Mathioudakis, K.; Papailiou, K. D., Adaptive simulation of gas turbine performance. *Journal of Engineering for Gas Turbines & Power* **1989**, 238, (112), 168-175.
9. Meskin, N.; Naderi, E.; Khorasani, K., Nonlinear Fault Diagnosis of Jet Engines by Using a Multiple Model-Based Approach. *Journal of Engineering for Gas Turbines & Power* **2011**, 13, (1), 63-75.
10. Kobayashi, T.; Simon, D. L., Evaluation of an Enhanced Bank of Kalman Filters for In-Flight Aircraft Engine Sensor Fault Diagnostics. *Journal of Engineering for Gas Turbines & Power* **2004**, 127, (3), 635-645.
11. Doel, D. In *The Role for Expert Systems in Commercial Gas Turbine Engine Monitoring*, 1990; 1990.
12. Loboda, I.; Feldshteyn, Y.; Ponomaryov, V. In *Neural Networks for Gas Turbine Fault Identification: Multilayer Perceptron or Radial Basis Network?*, ASME 2011 Turbo Expo: Turbine Technical Conference and Exposition, 2011; 2011.
13. Bettocchi, R.; Pinelli, M.; Spina, P. R.; Venturini, M., Artificial Intelligence for the Diagnostics of Gas Turbines—Part I: Neural Network Approach. *Journal of Engineering for Gas Turbines & Power* **2007**, 129, (3), 19-29.
14. Lee, S. M.; Choi, W. J.; Roh, T. S.; Choi, D. W., A study on separate learning algorithm using support vector machine for defect diagnostics of gas turbine engine. *Journal of Mechanical Science and Technology* **2008**, 22, (12), 2489-2497.

15. Lee, S. M.; Roh, T. S.; Choi, D. W., Defect diagnostics of SUAV gas turbine engine using hybrid SVM-artificial neural network method. *Journal of Mechanical Science and Technology* **2009**, 23, (2), 559-568.
16. Romessis, C.; Mathioudakis, K., Bayesian Network Approach for Gas Path Fault Diagnosis. *Journal of Engineering for Gas Turbines & Power* **2004**, 128, (1), 691-699.
17. Lee, Y. K.; Mavris, D. N.; Volovoi, V. V.; Yuan, M.; Fisher, T., A fault diagnosis method for industrial gas turbines using Bayesian data analysis. *Journal of Engineering for Gas Turbines & Power* **2010**, 132, (4), 041602.
18. Li, Y. G.; Ghafir, M. F. A.; Wang, L.; Singh, R.; Huang, K.; Feng, X., Non-Linear Multiple Points Gas Turbine Off-Design Performance Adaptation Using a Genetic Algorithm. *Journal of Engineering for Gas Turbines & Power* **2011,** 133, (7), 521-532.
19. Ganguli, R., Application of Fuzzy Logic for Fault Isolation of Jet Engines. *Journal of Engineering for Gas Turbines & Power* **2003,** 125, (3), V004T04A006.
20. Shabanian, M.; Montazeri, M., A neuro-fuzzy online fault detection and diagnosis algorithm for nonlinear and dynamic systems. *International Journal of Control, Automation and Systems* **2011,** 9, (4), 665.
21. Dan, M., Fuzzy Logic Estimation Applied to Newton Methods for Gas Turbines. *Journal of Engineering for Gas Turbines & Power* **2007,** 129, (1), 787-797.
22. Chandola, V.; Banerjee, A.; Kumar, V., Anomaly detection:A survey. *Acm Computing Surveys* **2009,** 41, (3), 1-58.
23. Lei, W.; Wang, S.; Zhang, J.; Liu, J.; Yan, Z. In *Detecting Intrusions Using System Calls: Alternative Data Models*, Security and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on, 1999; 1999; pp 133-145.
24. Sun, P.; Chawla, S.; Arunasalam, B. In *Mining for Outliers in Sequential Databases*, Siam International Conference on Data Mining, April 20-22, 2006, Bethesda, Md, Usa, 2006; 2006; pp 94--106.
25. Manson, G.; Pierce, G.; Worden, K., On the Long-Term Stability of Normal Condition for Damage Detection in a Composite Panel. *Key Engineering Materials* **2001,** 204-205, (204), 359-370.
26. Ruotolo, R.; Surace, C., A statistical approach to damage detection through vibration monitoring. *Bmc Proceedings* **1997,** 3 Suppl 7, (Suppl 7), : S25.
27. Hollier, G.; Austin, J. In *Novelty detection for strain-gauge degradation using maximally correlated components*, Esann 2002, Eurorean Symposium on Artificial Neural Networks, Bruges, Belgium, April 24-26, 2002, Proceedings, 2002; 2002; pp 257-262.
28. Yairi, T.; Kato, Y.; Hori, K., Fault Detection by Mining Association Rules from Housekeeping Data. *Proc of International Symposium on Artificial Intelligence Robotics & Automation in Space* **2001**.
29. Brotherton, T.; Johnson, T. In *Anomaly detection for advanced military aircraft using neural networks*, Aerospace Conference, 2001, IEEE Proceedings, 2001; 2001; pp 3113-3123 vol.6.
30. Ray, A., Symbolic dynamic analysis of complex systems for anomaly detection. *Signal Processing* **2004,** 84, (7), 1115-1130.
31. Rao, C.; Ray, A.; Sarkar, S.; Yasar, M., Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns. *Signal, Image and Video Processing* **2009,** 3, (2), 101-114.
32. Gupta, S.; Ray, A.; Sarkar, S.; Yasar, M., Fault detection and isolation in aircraft gas turbine engines. Part 1: Underlying concept. *Proceedings of the Institution of Mechanical Engineers Part G Journal of Aerospace Engineering* **2008,** 222, (G3), 307-318.
33. Yamaguchi, T.; Mori, Y.; Idota, H., Fault detection and isolation in aircraft gas turbine engines. Part 2: Validation on a simulation test bed. *Proceedings of the Institution of Mechanical Engineers Part G Journal of Aerospace Engineering* **2008,** 222, (3), 319-330.
34. Soumik, S.; Jin, X.; Asok, R., Data-Driven Fault Detection in Aircraft Engines With Noisy Sensor Measurements. *Journal of Engineering for Gas Turbines & Power* **2011,** 133, (8), 783-789.
35. Sarkar, S.; Mukherjee, K.; Sarkar, S.; Ray, A., Symbolic Dynamic Analysis of Transient Time Series for Fault Detection in Gas Turbine Engines. *Journal of Dynamic Systems Measurement & Control* **2012,** 135, (1), 014506.
36. Liu, H.; Hussain, F.; Tan, C. L.; Dash, M., Discretization: An Enabling Technique. *Data Mining and Knowledge Discovery* **2002,** 6, (4), 393-423.
37. Tsai, C. J.; Lee, C. I.; Yang, W. P., A discretization algorithm based on Class-Attribute Contingency Coefficient. *Information Sciences* **2008,** 178, (3), 714-731.
38. Gupta, A.; Mehrotra, K. G.; Mohan, C., A clustering-based discretization for supervised learning. *Statistics & Probability Letters* **2010,** 80, (9–10), 816-824.
39. Baum, L. E.; Petrie, T., Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Annals of Mathematical Statistics* **1966**, 37, (6), 1554-1563.

40. Witten, I. H.; Frank, E.; Hall, M. A., Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems). *Acm Sigmod Record* **2011,** 31, (1), 76-77.