

Novelty Detection based on Extensions of GMMs for Industrial Gas Turbines

Yu Zhang, Chris Bingham, Michael Gallimore

School of Engineering
University of Lincoln
Lincoln, U.K.

{yzhang; cbingham; mgallimore}@lincoln.ac.uk

Darren Cox

Siemens Industrial Turbomachinery LTD
Lincoln, U.K.

darren.cox@siemens.com

Abstract— The paper applies the application of Gaussian mixture models (GMMs) for operational pattern discrimination and novelty/fault detection for an industrial gas turbine (IGT). Variational Bayesian GMM (VBGMM) is used to automatically cluster operational data into steady-state and transient responses, where extraction of steady-state data is an important pre-processing scenario for fault detection. Important features are extracted from steady-state data, which are then fingerprinted to show any anomalies of patterns which may be due to machine faults. Field data measurements from vibration sensors are used to show that the extensions of GMMs provide a useful tool for machine condition monitoring, fault detection and diagnostics in the field. Through the use of experimental trials on IGTs, it is shown that GMM is particularly useful for the detection of emerging faults especially where there is a lack of knowledge of machine fault patterns.

Keywords— Gaussian mixture model; operational pattern discrimination; feature extraction; novelty detection; industrial gas turbine.

I. INTRODUCTION

The use of industrial gas turbines (IGTs) is widespread on a global scale. IGTs have design features that are specific for industrial applications. They can be categorised into packaged (less than ~20MW) or heavy classifications dependant on their design. IGT units use a direct-coupled power turbine that extracts the energy required to drive the compressor and transmit the power required to drive any equipment such as pumps or generators. An essential part of these complex systems are the control and monitoring units, which are required by legislation, and play an important role in ensuring system safety and performance.

Vibration, especially in fluid film bearings in IGTs can have an extremely detrimental effect and cause hundreds of thousands in repair bills and lost production. The impact this paper results in can mean that machinery faults, damage and catastrophic machine failures could be minimised or removed completely. Recurring faults or one-off problems will also be easily detected and then a decision can be made as to whether the problem needs looking into further or the turbine can be allowed to continue running. Faults that become gradually worse will also be easily identified by the shift of patterns in data measurements.

Most existing condition monitoring techniques focus on detecting abnormal trends or out-of-range amplitudes of measured signals [1,2], however, such methodologies are generally only suitable for steady-state unit operation, and often fail in situations where measurement transients caused by changes of power and/or loading occur, by indicating numerous false alarms. To address this issue, [3] has proposed an automatic on-line system to distinguish the input daily batch data to two categories—steady-state operation and operation with transients—in order to accomplish the pre-processing scenario of operational pattern discrimination, through the use of Gaussian mixture models (GMMs), especially variational Bayesian GMM (VBGMM) in this case.

The use of GMM techniques has also attracted significant attention for fault detection. Specially, in [4], GMMs have been applied to the detection of ‘novel’ vibration signatures in gearboxes, with experimental results showing good fault detection properties with known classifications. In this paper, outlier components are considered as background ‘noise’ for the GMM in the final mixture model [5,6], i.e. the use of a GMM with an outlier component (GMM_OC). In this way, measurements are still characterised as a GMM, but novel behaviour i.e. data that has a very low probability of being clustered in any existing distributions in the GMM, are clustered as outliers. It is the identification of such outliers that is shown to provide a mechanism for fault detection in the IGT applications.

The advantages of extensions of GMMs, such as ‘automatically calculating’ the number of clusters of VBGMM and the novelty detection capability of the GMM_OC, therefore provides particularly good properties for the operational pattern discrimination and detection of emerging faults on the sub-15MW IGTs considered here. Measurements from experimental trials are used to demonstrate the efficacy of the proposed methodology.

II. METHODOLOGY

The proposed methodology is depicted in the flowchart and table in Table 1, with the overview principles and characteristics of the underpinning techniques described in the following sub-sections.

TABLE I. OUTLINE METHODOLOGY

Steps	Inputs	Methods	Results	Purposes
Data collection	Running turbine	Machine probe sensor	Data to analyse	Check for faults or issues
Operational pattern separation	Vibration data	VBGMM	Steady state data	Pre-process data to remove transient data
Feature extraction	Daily steady state data	Statistical methods	Mean, variance etc.	Extract feature information from data
Novelty detection	Extracted features	GMM_OC	Showing anomalies in data	Early identification warning of a bearing fault

A. GMM

A GMM approximates the probability density function of sample data through a linear combination of Gaussian distributions $\mathcal{N}(x|\mu, \sigma)$ [7]. In this way, the probabilistic distribution of sensor measurements can be expressed as a sum of K Gaussian distributions with their corresponding means μ_k and standard deviations σ_k . A GMM consisting of K mixture components is therefore denoted as

$$p(x|\theta, \mu, \sigma) = \sum_{k=1}^K \theta_k \mathcal{N}(x|\mu_k, \sigma_k), \quad (1)$$

where $\mathcal{N}(x|\mu, \sigma)$ represents a Gaussian distribution density function of a multidimensional variable x , with mean $\mu \equiv \{\mu_k\}$ and standard deviation $\sigma \equiv \{\sigma_k\}$, as follows:

$$\mathcal{N}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad (2)$$

and θ_k are the mixing coefficients that are required to be chosen.

Let $X \equiv \{x_n\}$ denote a set of N data samples, where each sample consists of the (possibly) multidimensional variable x , and $n=1, \dots, N$. Assuming that x_n are statistically independent, the probability function of the data set can be expressed as the multiplication of the probabilities of the N data points:

$$p(X|\theta, \mu, \sigma) = \prod_{n=1}^N \left(\sum_{k=1}^K \theta_k \mathcal{N}(x_n|\mu_k, \sigma_k) \right). \quad (3)$$

which is the likelihood function, and through taking the natural log, takes the simplified form:

$$\mathcal{L}(X|\theta) = \ln p(X|\theta, \mu, \sigma) = \sum_{n=1}^N \ln \left(\sum_{k=1}^K \theta_k \mathcal{N}(x_n|\mu_k, \sigma_k) \right) \quad (4)$$

The resulting mixture density is denoted:

$$\rho_{nk} = \theta_k \mathcal{N}(x_n|\mu_k, \sigma_k), \quad (5)$$

which indicates the probability density of the k th mixture component on the n th data sample, and the conditional probability of a selected component k on the given data sample x_n is calculated as a proportion of ρ_{nk} , expressed as:

$$p_{nk} = \frac{\rho_{nk}}{\sum_{j=1}^K \rho_{nj}}, \quad (6)$$

To evaluate (4), an expectation-maximization (E-M) optimization procedure is commonly adopted for maximising the log-likelihood function [7].

B. VBGMM

To select the necessary number of mixture components, a variational Bayesian (VB) technique is used. For each data sample x_n , there exists a set of binary latent variables $z_{nk} \in \{0,1\}$ that specify the relationship between the N data points and the K mixture components, effectively indicating which data sample belongs to which mixture component. In the raw GMM technique, the number K is pre-determined, but in VBGMM, K is solved as an unknown through the solutions of $\{z_{nk}\}$.

The joint probability of all variables in the GMM is then given by

$$p(X, Z, \theta, \mu, \sigma) = p(X|Z, \mu, \sigma) p(Z|\theta) p(\theta) p(\mu|\sigma) p(\sigma) \quad (7)$$

where $p(\theta)$, $p(\mu|\sigma)$, and $p(\sigma)$ are Dirichlet, Gaussian and Wishart distributions, respectively [8]. Marginalizing (7) with respect to Z , μ and σ is intractable. However, through use of a VB framework, a lower bound on $p(X|\theta) = p(X|Z, \mu, \sigma) p(Z|\theta)$ can be found [9].

Denoting $\Psi \equiv \{Z, \mu, \sigma\}$, the marginal likelihood can be written as:

$$p(X|\theta) = \int p(X, \Psi|\theta) d\Psi. \quad (8)$$

$p(\Psi|\theta)$ can be approximated by a variational distribution, i.e. $q(\Psi) \approx p(\Psi|\theta)$. The Kullback-Leibler (KL) divergence of p from q is commonly used to make the evaluation (8) tractable, and it holds that:

$$\ln p(X|\theta) = D_{KL}(q\|p) + \mathcal{L}(q), \quad (9)$$

where $\mathcal{L}(q)$ is the lower bound, and $D_{KL}(q\|p)$ is the KL divergence. Maximizing the lower bound $\mathcal{L}(q)$ is equivalent to minimizing the KL divergence, which can be accomplished by appropriate choice of the q distributions.

The variational posterior distribution $q(\Psi)$ can be considered to factorize over the subsets $\{\Psi_i\} = \{Z, \mu, \sigma\}$, i.e.

$$q(Z, \mu, \sigma) = q_Z(Z)q_\mu(\mu)q_\sigma(\sigma), \quad (10)$$

The best distribution for each factor can be found by a free-form minimization over q_i :

$$q_i^*(\Psi_i) = \frac{\exp(E_{j \neq i}[\ln p(X, \Psi)])}{\int \exp(E_{j \neq i}[\ln p(X, \Psi)]) d\Psi_i}, \quad (11)$$

where $E_{j \neq i}[\cdot]$ represents the expectation of the distributions $q_j(\Psi_j)$ for all $j \neq i$.

Since the variational factors are mutually coupled, they can be solved using the expectation step (E-step). In the interests of brevity, for a more in-depth discussion of the properties and application of the VB framework, the reader is directed to [9].

C. GMM_OC

GMM_OC is a mixture model methodology now used for novelty detection, and is an extension of the traditional GMM [5,6]. It is established by adding an outlier component to traditional GMM, where the extra noise components (the outliers) are modeled by a uniform distribution. The resulting mixture model is then expressed as:

$$p(x) = \sum_{k=1}^K \theta_k \mathcal{N}(x|\mu_k, \sigma_k) + \left(1 - \sum_{k=1}^K \theta_k\right) \mathcal{U}(x). \quad (12)$$

Since the $\mathcal{U}(x)$ term is uniformly distributed, it will not unduly affect the underlying E-M procedure used for the GMM method. In reality, the noise term is normally assume small for initialization purposes. The total data are clustered into $(K+1)$ clusters, including K clusters of Gaussian distributions, and an outlier cluster of uniform distribution. Using the E-M

procedure, μ_k , σ_k and θ_k are estimated in the normal manner, and if the probability of an object x_i clustering into any of the K Gaussian distributions is less than a pre-set threshold, the data sample is considered to be clustered as an outlier, and hence provides an indication of novelty (emerging fault).

III. CASE STUDY

A. VBGMM for Operational Pattern Separation

VBGMM is now applied to cluster one-dimensional measurements (power readings in this case) from a sub-15MW IGT in order to classify its operational behaviour. Fig. 1 shows an example power signal, along with the calculated numbers of clusters. One month's data are used (31 days), and data for each day contains 1440 data points (i.e. 1440 minutes).

It can be seen that when the unit is considered to be operating in steady-state, the classification is 1, and otherwise is classified > 1 , by separating the results into example classifications over individual days. Cluster number index 0 indicates constant readings of 0s throughout the day, which are precluded from pattern analysis, and considered as a continuous shut down 'operation'. The corresponding vibration signals (from an inlet bearing sensor for instance) for the month are shown in Fig. 2.

Through VBGMM, vibration data for only steady state running operation can be extracted and kept for subsequently analysis. In this case, days 1-12 and days 29-31 are deemed to provide steady state characteristics, with the vibration data saved, as shown in Fig. 3.

This shows that the VBGMM approach provides a convenient pre-processing tool for subsequent fault detection, diagnostic and prognostic algorithms, by readily discriminating between transient and steady-state operation. In this way, it also provides an effective filter for more traditionally used diagnostic techniques that are only applicable to steady-state operation.

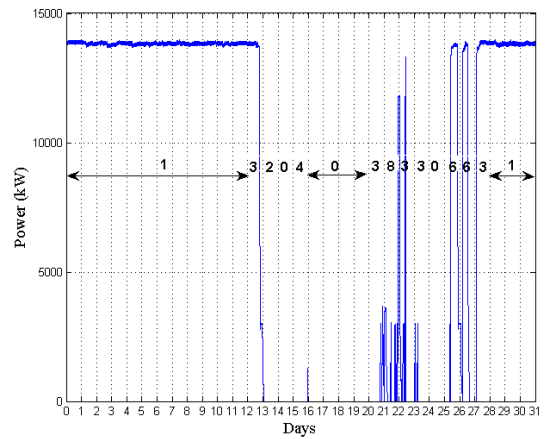


Fig. 1. Automatic clustering of power readings

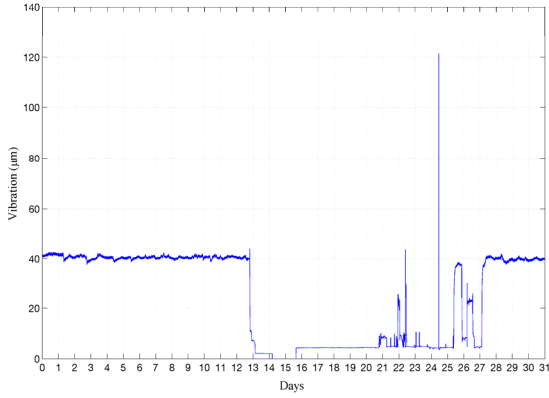


Fig. 2. Corresponding vibration measurement

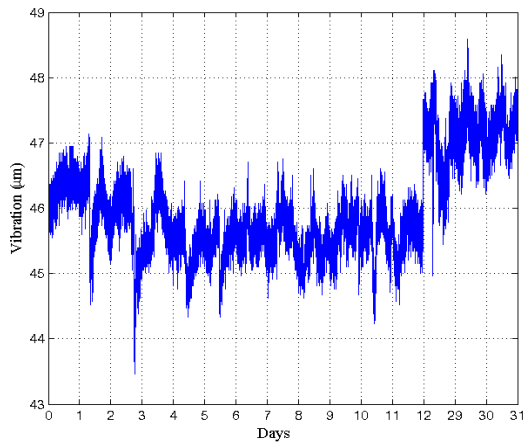


Fig. 3. Identified steady state data

B. Feature Extraction

Using the steady state data, a feature extraction method is used to extract important information from the data. Many algorithms have been previously reported to effect feature extraction, based on statistical methods, Fourier transform, principal component analysis, wavelets, empirical mode decomposition, data filters, etc. [9]

In this paper, classical statistics -- mean (that carries the information about the equilibrium of the signal) and the standard deviation (that carries information about the signal power)) are employed as a proof of principle. For the data in Fig. 3, the mean and the standard deviation of each day's data are calculated and plotted in Fig. 4, with indication number of the corresponding days.

From these results it is clear that, between days 29-31 and days 1-12's features diverge. However, for statistical study (if two years' data are studied for instance), the characteristics from Fig. 4 become condensed. For example, with two more months' data, the extracted features are shown in Fig. 5.

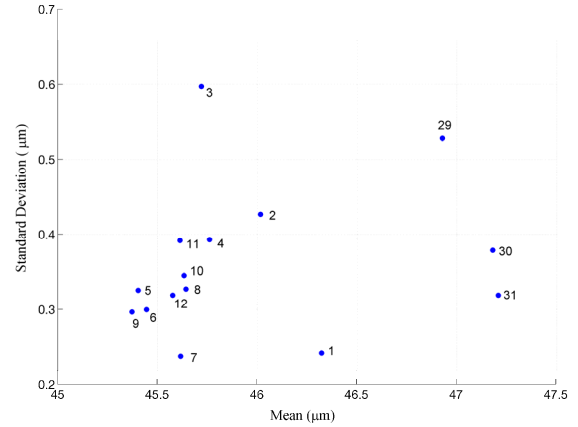


Fig. 4. Mean and standard deviation of steady-state data

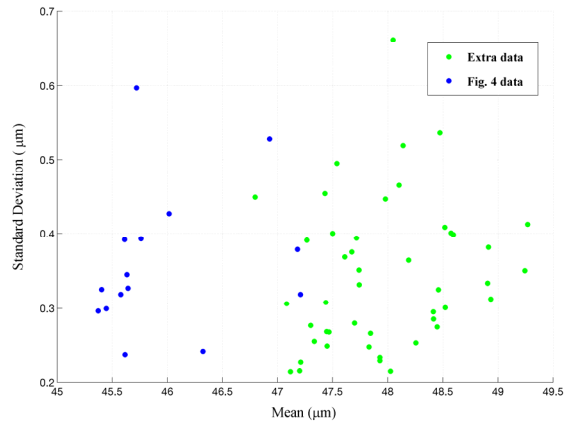


Fig. 5. Extracted features for 3 months' data

C. GMM_OC for Novelty Detection

Through use of 1-cluster GMM, a fingerprint can be built for the extracted features (data from Fig. 5 in this case are considered as normal data), and GMM_OC is utilized for novelty detection. Daily data from a subsequent month of measurements can therefore be investigated, leading to the vibration measurements shown in Fig. 6.

Faulted bearing measurements are often preceded by emerging increases in vibration amplitude levels, which eventually lead to poor performance and breakage. As shown in Fig. 7, an ellipse boundary line is drawn according to 2-D 1-cluster GMM model. For a 99.99% confidence level, all normal data are included within the boundary line, whereas outlier data from the original data can be removed as a result of lowering the confidence level, for example, a 95% confidence level boundary line is also shown. With a subsequent month's of data being studied, where a step change of vibration amplitudes has occurred, the measurements have been correctly been identified as outliers (day 9 and day 10 in this case), whereas data from days 1-4 are correctly considered to correspond to normal behaviour.

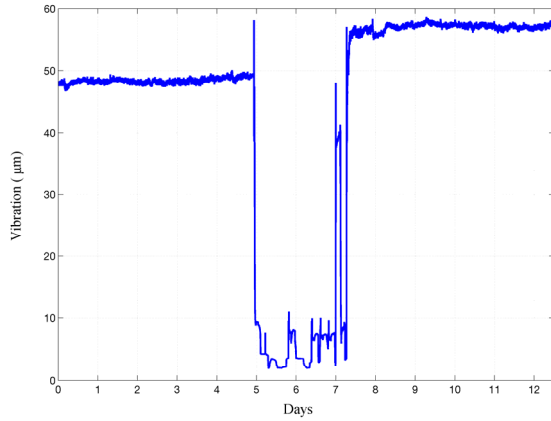


Fig. 6. Test data

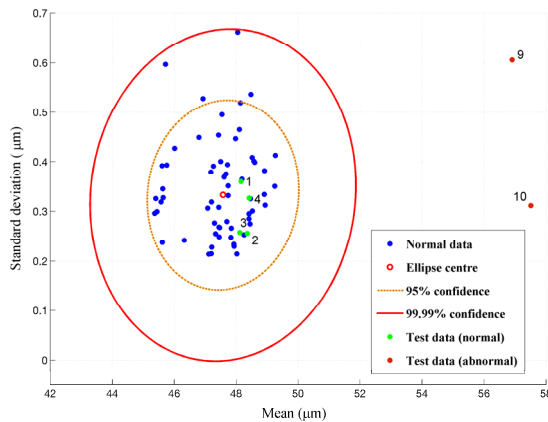


Fig. 7. GMM_OC novelty detection results

Noticeably, after several vibration amplitude step changes, eventually the bearing failed after ~ 3 months. After the shut down and inspected the inlet bearings were removed and the damage was noted (Fig. 8). The resulting history therefore forms a further fingerprint to identify subsequent related failures.

IV. CONCLUSION

This paper has proposed a novelty/fault detection strategy for IGT systems through the use of extensions of GMMs. VBGMM is used to discriminate between steady state operation from normal transient operational conditions that can be a result of simple changes of power demand or load for instance. Specially, in this paper, experimental trials of bearing vibration measurements are studied to demonstrate the efficacy of the proposed methodology, and show that correct identification of normal and transient behaviour due to the emergence of faults, can be identified.

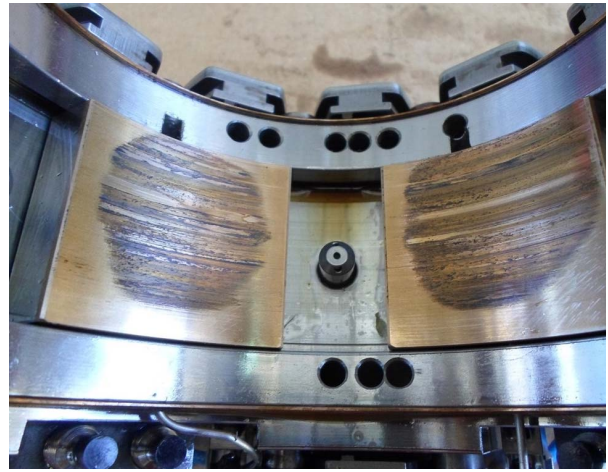


Fig. 8. Damaged bearing inspected

ACKNOWLEDGMENT

The authors would like to thank Siemens Industrial Turbomachinery, Lincoln, U.K., for providing research support and access to real-time data to support the research outcomes.

REFERENCES

- [1] Y. Zhang, C.M. Bingham, M. Gallimore, Z. Yang, J. Chen. "Sensor Fault Detection and Diagnosis Based on SOMNNs for Steady-State and Transient Operation." *Proc. of IEEE Int. Conf. on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, pp. 110-115. Milan, Italy, 15-17th July, 2013.
- [2] Y. Zhang, C.M. Bingham, M. Gallimore. "Applied Sensor Fault Detection, Identification and Data Reconstruction." *Advances in Military Technology*, vol. 8, no. 2, pp.13-26, 2013.
- [3] Y. Zhang, C.M. Bingham, M. Gallimore, J. Chen. "Steady-state and Transient Operation Discrimination by Variational Bayesian Gaussian Mixture Models." *Proc. of IEEE Int. Workshop on Machine Learning for Signal Processing*, pp. 1-5. Southampton, UK, 22-25th Sept, 2013.
- [4] T. Heynsa, P.S. Heyns, J.P. DeVilliers. "Combining Synchronous Averaging with a Gaussian Mixture Model Novelty Detection Scheme for Vibration-based Condition Monitoring of a Gearbox." *Mechanical Systems and Signal Processing*, vol. 32, pp. 200-215, 2012.
- [5] M.K. Titsias. "Unsupervised Learning of Multiple Objects in Images." PhD thesis, School of Informatics, University of Edinburgh, U.K. 2005.
- [6] P. Coretto. "The Noise Component in Model-based Clustering." PhD thesis, Department of Statistical Science, University of London, U.K. 2008.
- [7] C.M. Bishop. "Chapter 9: Mixture Models and EM." *Pattern Recognition and Machine Learning. Series: Information Science and Statistics*. Springer, Heidelberg. 2006.
- [8] K. Watanabe, S. Watanabe. "Stochastic Complexities of Gaussian Mixtures in Variational Bayesian Approximation." *Journal of Machine Learning Research*, vol. 7, pp. 625-644. 2006.
- [9] A. Corduneanu, C.M. Bishop. "Variational Bayesian Model Selection for Mixture Distributions." *Artificial Intelligent Statistics*, pp. 27-34. 2001.
- [10] K. Worden, W.J. Staszewski, J.J. Hensman. "Natural Computing for Mechanical Systems Research: A Tutorial Overview." *Mechanical Systems and Signal Processing*, vol. 25, pp. 4-111, 2011.