Operations Research Perspectives 5 (2018) 232–244

Contents lists available at ScienceDirect

Operations Research Perspectives

journal homepage: www.elsevier.com/locate/orp

Detecting operation regimes using unsupervised clustering with infected group labelling to improve machine diagnostics and prognostics

Juhamatti Saari*,[a,b], Johan Odelius[b]

[a] SKF-LTU University Technology Centre, Luleå University of Technology, Luleå, SE-97187, Sweden
[b] Division of Operation, Maintenance and Acoustics, Luleå University of Technology, Luleå, SE-97187, Sweden

ARTICLE INFO

Keywords:
Maintenance
Operation regime
Clustering
Data mining
LHD

ABSTRACT

Estimating the stress level of components while operation modes are varying is a key issue for many prognostic models in condition monitoring. The identification of operation profiles during production is therefore important. Clustering condition monitoring data with regard to operation regimes will provide more detailed information about the variation of stress levels during production. The distribution of the operation regimes can then support prognostics by revealing the cause-and-effect relationship between the operation regimes and the wear level of components.

In this study unsupervised clustering technique was used for detecting operation regimes for an underground LHD (load-haul-dump machine) by using features extracted from vibration signals measured on the front axle and the speed of the Cardan axle. The clusters were also infected with a small portion of the data to obtain the corresponding labels for each cluster. Promising results were obtained where each sought-for operation regime was detected in a sensible manner using vibration RMS values together with speed.

1. Introduction

Prognostic and health management (PHM) of a system is a discipline that link studies of failure mechanism to system lifecycle management [25]. One of the challenges of PHM is to estimate the stress level of the components of a system when operation modes vary. For many systems, it is either impossible or impractical to measure component stress accurately, so the next best thing may be to detect operation profiles during production. However, for complex systems, even the operation profile can be unknown and may change on daily basis. There is a need for methods which can use pre-existing data (condition monitoring or process data), often collected for other purposes, to detect operation regimes. Results can be used to predict different life scenarios in case of incipient faults or to determine the correct time and place to apply diagnostic techniques.

Machine learning and pattern recognition techniques for data mining have been improving dramatically recently, with many more areas of application, including PHM. They have been adapted for and are used in the PHM of machines in the automotive industry [6], defense and space programs [30] and heavy industries [33]. Machine learning techniques used in PHM can be divided into three rough categories: classification, regression and clustering techniques.

Classification algorithms are used to classify two or more categories by using data to distinguish, for example, a faulty system from a healthy one [19,27]. Regression models are mainly used for prognosis where the time to failure is estimated using existing historical data (see for instance [28]). Regression analysis involves the use of such techniques as neural networks, fuzzy logic systems and simpler univariate regression models; these techniques are not strictly reserved for regression analysis and can also be used for data mining. Recently Hanafizadeh et al. [16] used supervised neural networks to identify flow regimes in a pipe to determine when the flow type was changing during operation. This technique aims to improve the control of the process by determining when it is not optimal. However, it is not the most practical one for identifying operation regimes of complex machines; the data need to be labelled while training the model, and this is seldom done in a varying operating environment, as, for instance, with mobile machines. Suarez et al. [26] tracked real-time onboard damage accumulation using a model called PHM/ALPS. The goal was to evaluate the current mission profile (operating conditions) using past mission profiles (historical data) to demonstrate independent life prediction capability. It is difficult to adapt this type of technique for operation regime detection, however, unless several mission profiles are pre-recorded or simulated. Unsupervised clustering techniques, may be more practical than supervised ones in some cases since they do not require historical data from several different operating conditions. The benefit of

* Corresponding author at: SKF-LTU University Technology Centre, Luleå University of Technology, Luleå, SE-97187, Sweden
E-mail address: juhamatti.saari@ltu.se (J. Saari).

https://doi.org/10.1016/j.orp.2018.08.002
Received 7 February 2018; Received in revised form 6 July 2018; Accepted 3 August 2018
Available online 04 August 2018

unsupervised techniques is the possibility of finding natural groups and patterns in the data by optimizing the boundaries and the clusters in the data. Mostly these techniques are used for anomaly detection, where several clusters are formed to characterize typical system behaviour and alarm is send when data vector is outside of clusters [17]. Perhaps the most common unsupervised clustering method is the k-means algorithm [18]. This algorithm is initialized by picking k initial cluster points and allocating all data points to the closest one. Another popular cluster algorithm proven to be successful in many situations is the expectation-maximization [9,11]. When detecting operation regimes, there are limitations to using these algorithms (see for instance [13,31]). Perhaps the biggest problem when trying to implement these techniques is the need to set the number of clusters in advance, as this is rarely known for complex machines operating under unknown conditions or in a changing environment. For instance, load change in one time and position during production might create three separate clusters which cannot be treated as one mode.

To overcome this problem, Corduneanu and Bishop [10] have developed a variational Bayesian Gaussian mixture (VBGM) model. With this algorithm, it is not necessary to know the exact number of clusters (k) in the beginning, since only the maximum number of clusters needs to be set. Similar techniques have been applied to defining operation regimes in the process industry using control parameters, such as valve openings or temperature [34]. but in these techniques, the k value is defined using another algorithm [12].

Although unsupervised techniques have advantages when compared to supervised ones, there are some practical limitations. One is the validation of cluster labels, i.e., what each cluster actually represents. To overcome this problem, we propose a method where the VBGM algorithm is first used to separate a large set of condition monitoring data into groups (clusters) which are later infected with a smaller set of data with labels. We apply the method to the analysis of vibration data collected from a complex machine operating under harsh conditions (underground mining loader, LHD). Aim is to see how the unsupervised algorithm, together with infection data, can be applied for separating operation modes using only condition monitoring data. We use vibration measurement data collected for diagnosis purposes and consisting of noise from many natural sources. Work is novel in that it applies the VGBM clustering algorithm to real data and explains how it can be used generically with infection data to predict labelled clusters.

2. Background and labelling operation regimes

Clustering technique (VBGM) used in this study for separating data for different clusters is based on the work by Corduneanu and Bishop [10], which can be also found in the book written by Bishop [7]. When the VBGM algorithm is used for mining condition monitoring data, more specifically, to separate data into meaningful operation regimes, it is not necessary to know the exact number of clusters, since components whose expected mixing coefficients are numerically indistinguishable from zero are not plotted [7]. The method is also more practical (generalizable) since it can rely on data when the training set is large and on the prior distribution assumption when the data set is small.

In Gaussian mixture model for each observation x_n we have a corresponding latent variable z_n comprising a 1-of-K binary vector with elements z_{nk} for k = 1,... ,K. Denotation for observed data set is $\mathbf{X} = x_1,...,x_N$, similarly latent variables are denoted as $\mathbf{Z} = z_1,...,z_N$.

Conditional distribution of $\mathbf{Z}$, given the mixing coefficients \boldmath π, is defined as follows [7]

$$p(Z \mid \pi) = \prod_{n=1}^{N} \prod_{k=1}^{K} {\pi_k}^{z_{nk}}. \tag{1}$$

For the observed data, the conditional distribution, given the latent variables and the component parameters, is as follows [7]

$$p(X \mid Z, \mu, \Lambda) = \prod_{n=1}^{N} \prod_{k=1}^{K} \mathcal{N}(x_n \mid \mu_k, \Lambda_k^{-1})^{z_{nk}}, \tag{2}$$

where $\mu = \{\mu_k\}$ is mean and $\Lambda = \{\Lambda_k\}$ is precision.

Using conjugate prior distribution and choosing a Dirichlet distribution over the mixing coefficient \boldmath π, which is defined as [7]

$$p(\pi) = Dir(\pi \mid \alpha_0) = C(\alpha_0) \prod_{k=1}^{K} \pi_k^{\alpha_0 - 1}, \tag{3}$$

where $C(\alpha_0)$ is the normalization constant for the Dirichlet distribution. Hyperparameter α_0 can be interpreted as the effective number of observations associated with each component of a mixture. If α_0 is small, the posterior distribution will be influenced primarily by the data rather than the prior.

By introducing independent Gaussian-Wishart prior governing the mean and precision of each Gaussian component, the distribution can be written as [7]

$$p(\mu, \Lambda) = p(\mu \mid \Lambda) p(\Lambda) = \prod_{k=1}^{K} \mathcal{N}(\mu_k \mid m_0, (\beta_0 \Lambda_k)^{-1}) \mathcal{W}(\Lambda_k \mid W_0, \upsilon_0). \tag{4}$$

Joint distribution of all of the random variables, is given by the equation [7]

$$p(X, Z, \pi, \mu, \Lambda) = p(X \mid Z, \mu, \Lambda) p(Z \mid \mu, \Lambda) p(\pi) p(\pi \mid \Lambda) p(\Lambda). \tag{5}$$

In the Eq. 3 only the variables $\mathbf{X}$ are observed.

Considering a variational distribution which factorizes between the latent variables and the parameters, so that [7]

$$q(Z, \pi, \mu, \Lambda) = q(Z) q(\pi, \mu, \Lambda). \tag{6}$$

With this assumption it is possible to obtain a traceable practical solution to the Bayesian mixture model. The optimal solution is found by seeking a distribution for which the lower bound is largest.

A toolbox for the algorithm is publicly available at Mathworks [22]. In this study, we kept the parameter settings at default each time the algorithm was run.These parameters, α_0, was 1 and β_0, which affects to the initial precision value (Λ), was 1.

To overcome the problem of not knowing what each cluster represent, we propose method to collect another set of data which is much smaller than the training set (See Fig. 1). This smaller set of data can be used to infect some or all of the found clusters in order to know what they represent by predicting their clusters using already trained models. Benefit of the technique is that the training can be carried out for a much larger data set and rare patterns which may occur during production in some situations, will be included in the model. However disadvantage may be the difficulties of interpreting cluster labels, if data is distributed evenly among clusters. In these cases, parameters needs to re-selected or use different initial parameter values to achieve better results. Infection data should be collected in such manner that one complete cycle of the operation is present.

With this technique, once he computationally demanding training phase is over (although it is the same as compared to traditional maximum likelihood ones), real time or near real time cluster prediction for new data set is achievable for several system/components by using on-site feature extraction and wireless communication together with centralized computing.

The ideal way to collect infection data set would be to let the operator determine when to acquire data during operation (first-hand knowledge) or to automatize data collection and use RFID tags or other similar techniques. These are used in many industries to keep track of mobile machines (for instance, in mining industry). Time period for data collection should cover the whole operation mode in the beginning and only later, if the operation mode is distributed evenly into many clusters, a deeper analysis and better selection should be done.