CHEMICAL ENGINEERING TRANSACTIONS
VOL. 33, 2013
Guest Editors: Enrico Zio, Piero Baraldi

ISBN 978-88-95608-24-2; **ISSN** 1974-9791

A publication of
AIDIC
The Italian Association
of Chemical Engineering
Online at: www.aidic.it/cet

DOI: 10.3303/CET1333104

Machinery Prognostic Method Based on Multi-Class Support Vector Machines and Hybrid Differential Evolution – Particle Swarm Optimization

James K. Kimotho*, Christoph Sondermann-Woelke, Tobias Meyer, Walter Sextro

University of Paderborn, Mechatronics and Dynamics, Pohlweg 47-49, 33098 Paderborn, Germany
james.kuria.kimotho@uni-paderborn.de

Recently, focus on maintenance strategies has been shifted towards prognostic health management (PHM) and a number of state of the art algorithms based on data-driven prognostics have been developed to predict the health states of degrading components based on sensory data. Amongst these algorithms, Multiclass Support Vector Machines (MC-SVM) has gained popularity due to its relatively high classification accuracy, ability to classify multiple patterns and capability to handle noisy /incomplete data. However, its application is limited by the difficulty in determining the required kernel function and penalty parameters. To address this problem, this paper proposes a hybrid differential evolution – particle swarm optimization (DE-PSO) algorithm to optimize the MC-SVM kernel function and penalty parameters. The differential algorithm (DE) obtains the search limit for the SVM parameters, while the particle swarm optimization algorithm (PSO) determines the global optimum parameters for a given training data set. Since degrading machinery components display several degradation stages in their lifetime, the MC-SVM trained with optimum parameters are used to estimate the health states of a degrading machinery component, from which the remaining useful life (RUL) is predicted. This method improves the classification accuracy of MC-SVM in predicting the health states of a machinery component and consequently increases the accuracy of RUL predictions. The feasibility of the method is validated using bearing prognostic run-to-failure data obtained from NASA public data repository. A comparative study between MC-SVM with parameters obtained using simple grid search with n-fold cross validation and MC-SVM with DE-PSO based on prognostic performance metrics reveals that the proposed method has better performance, with all the cases considered falling within a 10 % error margin. The method also outperforms other soft computing methods proposed in literature.

1. Introduction

In order to increase machinery availability and safety, focus has been shifted towards prognostic health management (PHM) maintenance strategy. PHM aims at monitoring the current health of a system or component and predicting the condition of the component or system at a future time (ISO 13381, 2004). The predictions can then be used to schedule appropriate maintenance or to adaptively control a system to ensure that its mission is accomplished, for instance in critical mission systems in aircrafts, autonomous vehicles and automatic manufacturing machinery. Several algorithms for PHM have been developed and can be broadly categorized into three groups: reliability based (Zio et al., 2012), model based and data driven based (Tobon-Mejia et al., 2011). In recent years, focus is increasingly being directed towards data driven methods since they do not require expert knowledge of the system being monitored and can be adapted to different kinds of systems, as long as there is some historical condition monitoring data for training the prognostic models. Health indicatory features are extracted from sensory data to define various stages in the lifetime of a component. The trained models are used in on-line monitoring of the health of a component. Commonly used algorithms include: support vector machines (SVM), artificial neural networks (ANN), Neural Fuzzy Networks, Hidden Markov Models, Bayesian Networks.

One of the methods used in diagnosis and prognosis of machinery is to estimate the health states of a degrading component. This approach is motivated by the fact that a system or machinery component displays several degradation stages in its lifetime. Tobon-Mejia et al. (2011) employed Hidden Markov Models (HMM) to classify the health of rolling element bearings into three conditions: healthy, faulty and failed. Camci and Chinnam (2010) employed hierarchical HMM for online estimation of health states of drill bits in a milling machine as they deteriorate. Lee et al. (2006) proposed a model that employs quantitative health assessment to evaluate the overlap between normal behaviour, most recent behaviour and faulty behaviour in machinery. However, our preliminary investigation showed that a component may display more than one faulty condition, thereby giving rise to more than three health states. Kim et al. (2012) proposed the use of health state probability estimates using SVM to conduct diagnosis and prognosis of machinery. The approach uses the historical RUL of similar components at each trained health state, an approach which would not be suitable for components that display varying lifetimes under similar operating conditions.

This work aims at developing a prognostic model based on the health states of a component using MC-SVM. The optimum penalty and kernel parameters are obtained through hybrid differential evolution and particle swarm optimization. The proposed approach consists of two parts. The first part involves training the SVM with the optimum parameters from the hybrid DE-PSO algorithm to classify the health states of a degrading component. The second part involves estimating the RUL from the health state probabilities, the percentage RUL at each trained health state, the spent lifetime of the components and the duration of stay at each health state. The feasibility of the proposed approach is validated using run to failure rolling element bearing data obtained from the NASA Data Repository (Nectoux et al., 2012).

2. Multi-Class Support Vector Machines

Support Vector Machines (SVM) is a machine learning technique for binary data classification, which uses a kernel function to transform the input data into a higher dimensional feature space. The binary classification problem is then solved by separating the data using a hyperplane (Chang and Lin, 2011). The hyperplane is defined by support vectors, a subset of data that defines the boundary between the two classes.

Multi-class classification is achieved by constructing and combining several binary classifiers. Several methods of constructing multi-classification SVM have been proposed. These include: pairwise (one versus one), one versus all and directed acrylic graph SVM (DAGSVM) (Hsu and Lin, 2002). Pairwise method was employed in our work as it has been shown to be more suitable for practical applications (Hsu and Lin, 2002). In this method, $n(n-1)/2$ SVM classifiers where each one is trained on data from two classes are constructed. Based on the training data, SVM produces a model which predicts the target values of the test data given only the attributes of the test data (Hsu et al., 2010). Given l training data $(\boldsymbol{x_{t1}}, y_{t1}), \dots .(\boldsymbol{x_{tl}}, y_{tl})$, where $\boldsymbol{x_{ti}} \in R^n, i = 1,2,..l,$ is the health indicator (extracted feature) at time t and $y_{ti} \in \{1,2,..n\}$ is the health state of $\boldsymbol{x_{ti}}$, the SVM solves the following classification problem for training data from ith and jth health states as shown below (Chang and Lin, 2011):

Minimize: $\frac{1}{2}(\boldsymbol{w^{ij}})^T\boldsymbol{w^{ij}} + C\sum_t \xi_t^{ij}\,(\boldsymbol{w^{ij}})^T,$ (1)

Subject to: $(\boldsymbol{w^{ij}})^T\phi(\boldsymbol{x_i}, \boldsymbol{x_j}) + b^{ij} \geq 1 - \xi_t^{ij}, \quad \text{if } y_t = i,$

$$(\boldsymbol{w^{ij}})^T\phi(\boldsymbol{x_i}, \boldsymbol{x_j}) + b^{ij} \geq -1 + \xi_t^{ij}, \text{ if } y_t \neq i,$$

$$\xi_t^{ij} \geq 0,$$

where the training data $\boldsymbol{x_{ti}}$ are mapped to a higher dimensional space by the kernel function $\phi(\boldsymbol{x_i}, \boldsymbol{x_j})$. Also $\boldsymbol{w} \in R^n$ is the support vector, $b \in R$ is the hyperplane bias, C is the penalty parameter and ξ_t^{ij} is the slack variable which defines the margin of the support vectors from the hyperplane. The error term $C\sum_t \xi_t^{ij}(\boldsymbol{w^{ij}})^T$ reduces the number of training errors when data are not linearly separable (Hsu and Lin, 2002). The radial basis function (RBF) shown in Eq. (2) was employed as the kernel function due to its high accuracy in classifying non-linear data (Chang and Lin, 2011):

$$\phi(\boldsymbol{x_i}, \boldsymbol{x_j}) = \exp\left(-\gamma\|x_i - x_j\|^2\right), \ \gamma > 0, \quad (2)$$

where γ is the kernel parameter. In the SVM parameter optimization problem, a combination of C and γ that minimizes the cross validation error during training is obtained. To determine health state probabilities, the voting method proposed by Wu and Lin (2004) was employed.

2.1 Hybrid Particle Swarm Optimization and Differential Evolution

Differential evolution was employed to determine the search limit of the optimum parameters for each training data set, while PSO was employed to determine the global optimum parameters for each training data set. DE was selected because it is simple and straight forward to implement and converges fast compared to other optimization algorithms (Elsayed et al., 2012). On the other hand, PSO has the ability to accelerate towards the global best solution thereby avoiding convergence at the local optimum solution (Mirjalili et al., 2012).

2.2 Differential Evolution

Differential Evolution is a population based stochastic real parameter optimization algorithm that operates through similar computational steps as other evolutionary algorithms, that is, mutation, cross over and selection (Elsayed et al., 2012). In the mutation process, the parameters to be optimized (population) in a given problem are first initialized in vector form. DE then mutates randomly chosen vectors to produce an intermediary population of mutants. Each vector in the current population is combined with a mutant to produce a trial population in the crossover. Selection is then done to determine whether the target or trial vector survives to the next generation (Elsayed et al., 2012).

2.3 Particle Swarm Optimization (PSO)

PSO is inspired by the behaviour of social groups such as bird flocks and fish swarms. Each member of the swarm, called particle flies around the search space by following simple position and velocity update equations while evaluating the objective function at each new position (Mirjalili et al., 2012). The best position found by any particle within the search space is its own personal best towards which it accelerates in the next iteration. The best combination of parameters obtained from PSO is used to train the MC-SVM for health state classification.

3. Proposed PHM Method

The proposed method involves training MC-SVM to estimate the health states of a degrading component. Figure 1 shows the workflow of the proposed machinery prognostic method.

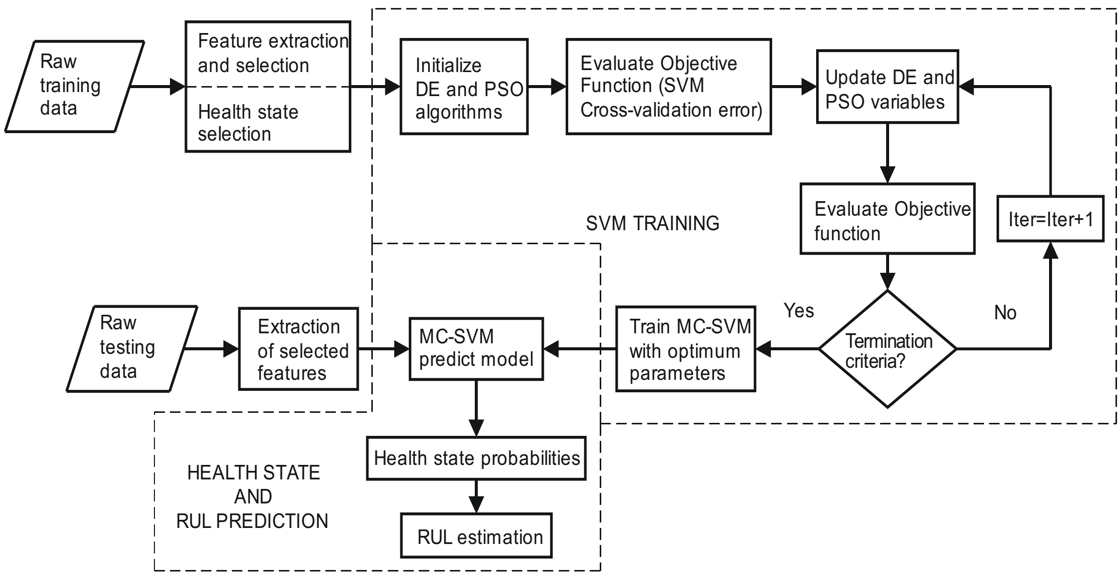

Figure 1: Workflow of the proposed machinery prognostics method

4. Data Analysis

Prognostic run to failure data of ball bearings obtained from NASA data repository (Nectoux et al., 2012) were used in the training and testing of the proposed model. The testing data sets are truncated run to failure data sets, and the prognostic algorithm is required to estimate the operation time from the truncation point to failure. Each training/ testing data set consists of vibration signals sampled at a frequency of 25.6 kHz with 2560 samples recorded at intervals of 10 seconds.

4.1 Feature Extraction and Selection

Signals from faulty bearings are considered non-stationary and therefore time-frequency feature extraction methods are recommended. In this study, wavelet packet decomposition (WPD) with 3 decomposition levels was considered. The raw time domain signals were decomposed into the detail and approximate signals, with the approximate signal being further decomposed to its detail and approximate signals at each level, yielding a total of six signals for each accelerometer reading. The wavelet energy of each sampled signal was computed and a correlation analysis performed. Features with a correlation coefficient greater than 0.9 and less than 0.5 were selected as SVM inputs. WPD was selected because it is capable of handling non-stationary signals.

4.2 Health States Selection

The training data was analysed to determine the number of possible health states in the lifetime of a given set of components, in this case ball bearings. Power spectral density (PSD) was performed to convert time domain data into frequency domain data. A plot of the frequencies at the peak amplitude of the PSD spectrum against time revealed existence of different health states, where a new set of frequencies appeared or disappeared in the lifetime of the bearing (Sutrisno et al., 2012), as shown in Figure 2 (a). The health states were confirmed using statistical features such as root mean square (rms) of the vibration signals as shown in Figure 2 (b).

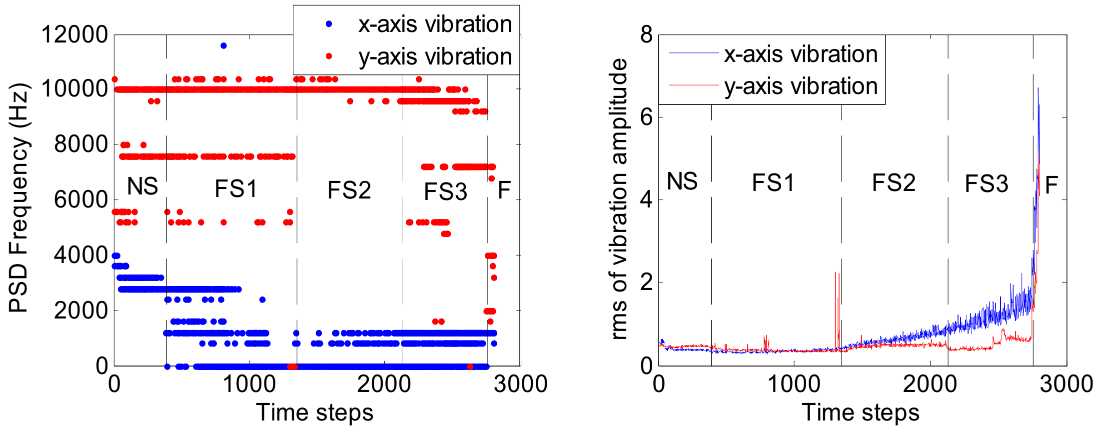

Figure 2: Observed health states in the lifetime of training bearing 1 from the first operation condition, (a) Frequency at peak amplitude, (b) Root mean square of the vibration signal, where NS is Normal state, FS1 is faulty state 1, FS2 is faulty state 2, FS3 is faulty state 3 and F is failed.

Five health states (ranging from normal condition to failure) were observed as shown in Figure 2. The bearings were assumed to have failed when the accelerometers recorded a value of 20 g, to avoid propagation of damage to other test rig components (Nectoux et al., 2012). The percentage remaining useful life at each health state was computed and a weighted average was obtained from the two training data sets provided. Table 1 shows the computed percentage remaining useful life at each health state. The target vector for SVM training consisted of the health state class for each training feature data point.

Table 1: Observed health states and their percentage remaining useful life

Health state class	Health state	Percentage RUL(R_i)
1	Normal state	89 – 100 %
2	Faulty state 1	55 – 89 %
3	Faulty state 2	25 – 55 %
4	Faulty state 3	2 – 25 %
5	Failure	0 – 2 %

4.3 RUL Estimation

The remaining useful life was estimated from the health state probability and the percentages of the historical RUL as shown in the following equation (Kim et al., 2012):

$RUL_t = \sum_{i=1}^{n} P_i R_i L_{it} - t_i,$ (3)

where P_i is the probability that a given data point belongs to class i, which is an output of the SVM algorithm; R_i is the percentage remaining useful life of health state i, obtained from Table 1; n is the number of health states or classes; L_{it} is a function of the spent lifetime and t_i is the time spent at health state i, a novel factor that has significant influence on the accuracy of the method.

5. Results and Discussions

The above procedures were implemented in MATLAB and the LIBSVM toolbox for SVM was employed (Chang and Lin, 2011). Test data provided for validating the PHM model was analysed and 12 features were extracted using wavelet packet decomposition (WPD) technique. The resulting vectors were fed into the SVM predict model as the input. Figure 3 shows the predicted health state probability at each given data point for the first test bearing.

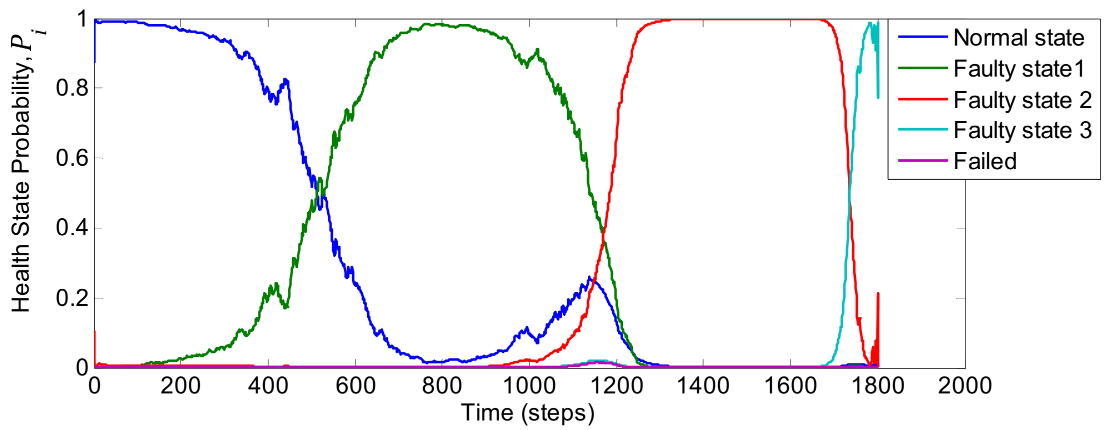

Figure 3: Predicted health states in the lifetime of test bearing 3 from the first operating condition

The test data provided were truncated run-to-failure data of the bearings and from Figure 3 it can be observed that the bearing had degraded up to faulty state 3. The remaining useful life was then calculated using Eq. (3). The performance of the algorithm was evaluated based on percentage error and percentage accuracy by:

$\%Error = \frac{ActRUL - RUL}{ActRUL} \times 100,$ (4)

$\%Accuracy = 100 - |\%Error|,$ (5)

where $ActRUL$ is the actual remaining useful life of the test data provided for validation of the prognostic methods and RUL is the predicted remaining useful life. The test data consists of truncated run-to-failure accelerometer readings of five bearings and similar analyses were conducted on all the other test bearings. The prognostic performance metrics are shown in Table 2. A comparison between the performances of the proposed DE-PSO MC-SVM algorithm, with MC-SVM trained with kernel parameters obtained from a simple grid search was conducted as shown in Table 2.

Table 2: Performance evaluation of the proposed PHM model (Operating condition 1 (Nectoux et al., 2012))

Test	% Error		% Accuracy	
	DE-PSO SVM	Simple grid search SVM	DE-PSO SVM	Simple grid search SVM
Bearing 1_3	3.19	7.26	96.81	92.74
Bearing 1_4	9.03	-23.50	90.97	76.50
Bearing 1_5	1.87	-5.81	98.12	94.19
Bearing 1_6	-4.81	-12.18	95.19	87.82
Bearing 1_7	-0.97	29.58	99.03	70.42

The negative error in Table 2 indicates a late prediction or overestimation. In machinery maintenance good prognostic performance relates to early RUL predictions (Nectoux et al., 2012). It can be observed that the proposed PHM model is fairly accurate and outperforms the SVM method with a simple grid search of the SVM parameters, with the majority of the predictions being early predictions. The method also performs better than other soft computing methods proposed in literature (Sutrisno et al., 2012).

6. Conclusion

This study presents a prognostic approach based on MC-SVM with kernel parameters optimized using hybrid DE-PSO. The method employs DE-PSO MC-SVM to estimate the health states of a degrading component, from which the remaining useful life is predicted. Results show that the proposed method can be used to accurately predict the health states of a degrading component and estimate the remaining useful life. The method outperforms MC-SVM with kernel parameters obtained from a simple grid search and also other data driven methods proposed in literature.

Acknowledgement

The authors wish to acknowledge the support of the Chair, Department of Mechatronics and Dynamics, University of Paderborn, German Academic Exchange Service (DAAD) and the Ministry of Higher Education Science and Technology (MoHEST), Kenya.

References

Camci F., Chinnam R. B., 2010, Health-State Estimation and Prognostics in Machining Processes, IEEE Transactions on Automation Science and Engineering, 7, 581-597.

Chang C.-C., Lin C.-J., 2011, LIBSVM: A Library for Support Vector Machines, ACM Transactions on Intelligent Systems and Technology, 2, 27:1-27:27.

Elsayed S. M., Sarker R. A., Ray T., 2012, Parameters Adaptation in Differential Evolution, IEEE World Congress on Computational Intelligence, Brisbane, Australia.

Hsu C.-W., Lin C.-J., 2002, A Comparison of Methods for Multiclass Support Vector Machines, IEEE Transactions on Neural Networks, 13, 415-425.

ISO 13381, 2004, Condition monitoring and diagnostics of machines - Prognostics - Part 1: General Guidelines.

Kim H.-E., Tan A. C., Mathew J., Choi B.-K., 2012, Bearing Fault Prognosis Based on Health State Probability Estimation, Expert Systems with Applications, 39, 5201-5213.

Lee J., Ni J., Djurdjanovic D., Qiu H., Liao H., 2006, Intelligent prognostics tools and e-maintenance, Computers in Industry, 57, 476-489.

Mirjalili S. A., Hashim S. Z. M., Sardroudi H. M., 2012, Training Feedforward Neural Networks using Hybrid Particle Swarm Optimization and Gravitational Search Algorithm, Applied Mathematics and Computation, 218, 11125-11137.

Nectoux P., Gouriveau R., Medjaher K., Ramaso E., Morello B., Zerhouni N., Varnier C., 2012, PRONOSTIA: An Experimental Platform for Bearing Accelerated Degradation Tests, IEEE Conference on Prognostics and Health Management, Denver, CO, USA.

Sutrisno E., Oh H., Vasan A. S. S., Pecht M., 2012, Estimation of Remaining Useful Life of Ball Bearings using Data Driven Methodologies, IEEE Conference on Prognostics and Health Management, Denver, CO, USA.

Tobon-Mejia D. A., Medjaher K., Zerhouni N., 2011, Estimation of the remaining useful life by using wavelet packet decomposition and HMMs, IEEE Aerospace Conference AIAA, Montana, USA.

Wu T.-F., Lin C.-J., 2004, Probability Estimates for Multi-class Classification by Pairwise Coupling, Journal of Machine Learning Research, 5, 975-1005.

Zio E., Broggi M., Golea L. R., Pedroni N., 2012, Failure and Reliability Predictions by Infinite Impulse Response Locally Recurrent Neural Networks, Chemical Engineering Transactions, 26, 117-122.