Conference Paper

"Machine Learning and Neural Network for Maintenance Management"

- Lecture Notes on Multidisciplinary Industrial Engineering -

Accepted: 29 June 2017

Alfredo arcos Jiménez
Ingenium Research Group, Universidad de Castilla-La Mancha
Alfredo.Arcos@alu.uclm.es

Carlos Quiterio Gómez Muñoz
Ingenium Research Group, Universidad de Castilla-La Mancha
CarlosQuiterio.Gomez@uclm.es

Fausto Pedro García Márquez
Ingenium Research Group, Universidad de Castilla-La Mancha
FaustoPedro.Garcia@uclm.es

Cite as: Jiménez, A. A., Muñoz, C. Q. G., & Márquez, F. P. G. (2017, July). Machine Learning and Neural Network for Maintenance Management. In International Conference on Management Science and Engineering Management (pp. 1377-1388). Springer, Cham.

DOI: https://doi.org/10.1007/978-3-319-59280-0_96

Contents

Chapter 1
Machine Learning and Neural Network for Maintenance Management

Alfredo Arcos Jiménez, Carlos Quiterio Gómez Munoz and Fausto Pedro Garcła Marquez

Abstract The paper presents a novel approach that allows to optimize the ultrasonic wave sensors for a condition monitoring system employing. It can detect and diagnosis different faults with a signal, such as delamination, mud or ice on blades of wind turbines. This methodology allows to avoid the redundancy of sensors, since a specific number of ultrasonic transducers can determine the structural condition using guided waves. The signal is pre-processed with the aim of removing the noise, then extracted and selected features to be later classified by Machine Learning and Neural Networks. Finally, for each damage or anomaly, the best classifier will be evaluated. The best classifier of each damage will act on a parallel network that will process the signal sent by the sensor.

Keywords Non-destructive testing · Fault detection and diagnosis · Condition monitoring system · Wavelet transforms · Machine learning · Neuronal network

1.1 Introduction

The objective in the next decades is to obtain a competitive and sustainable energy to all countries, allowing to reduce dependence on fuels to households, industries and transportation. Wind power and Concentrated Solar Power (CSP) are being two of the main renewable energy sources. Its importance in the energy market is being essential, and the forecasts show it will continue in the near future.

Renewable energy industry requires significant improvements in reliability, lifetime or availability that it is done by an efficient maintenance based on Condition Monitoring Systems (CMS). CMS is the process of determining the condition of system [37, 38, 49], where one of the main proposes is to identify changes between

---

A. Alfredo (✉) · Q. Carlos · P. Fausto
Alfredo Arcos Jiménez, Ingenium Research Group, Castilla-La Mancha University, Spain
e-mail: Alfredo.Arcos@alu.uclm.es

J. Xu et al. (eds.), *International Conference on Management Science and Engineering Management 2017*,
Advances in Intelligent Systems and Computing XXX, DOI: XXXXXXXXXX

two main states of the structure, damaged and undamaged [36, 39, 41]. They will provide different patterns that will be used in order to analyze the condition of the system [19].

A complex CMS contains numerous sensors that generates different type of signals with information and sampling frequencies [41]. It needs to be processed, preferable on-line, with a minimum computational cost and with high accuracy to reduce false alarms [39].

This paper present a novel approach to reduce the number of sensors, maximizing the signal processing analysis to detect different damages. The guided wave ultrasonic signals are employed for fault detection and classification.

Structural Health Monitoring (SHM) is a technology that combines advanced CMS, together with signal processing, to determine the condition of the structures on line or not [19, 36, 41]. SHM leads to increase the Reliability, Availability, Maintainability and Safety (RAMS) of the system [16, 41]. SHM allows also to know the different levels of the defect severity. It will be useful for an optimal maintenance management to reduce costs and increase the profitability.

Farrar and Sohn [16] considered pattern recognition in SHM. This methodology consists in four stages shown in Fig. 1.1

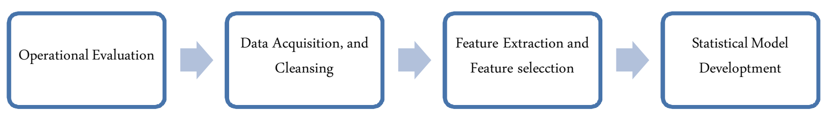

Fig. 1.1 Methodology SHM.

The problems caused by the implementation of damage identification system is assessed in operational evaluation phase. Data acquisition and normalization involves the choice of sensors and their location, the normalization procedure of the data collected and the selection of the data to be used in the feature selection process. Dimensionality reduction techniques have been developed to remove irrelevant data (noise) and redundant features, and categorized mainly into Feature Extraction (FE) and Feature Selection (FS). FE contains information of the signal that distinguishes between a damaged or non-damaged structures [57]. FS selects the data that are better for the detection of damage and executes a condensation of the data [16, 40]. Finally, classifiers, based on statistical models, calculate and cluster the dataset depending on the damage. These classifiers are grouped as: Supervised Learning (SL) and Unsupervised Learning [35].

This paper proposed patterns recognition of each damage by Machine Learning and Neural Network.Then approach minimize the number of sensors, where through an ultrasonic signal emitted by a sensor is able to classify different states of different damages or anomalies.

This process is complex because each damage must be analyzed individually. On the other hand, there is no generic classifier that provides the best results. Even the same classifier for the same signal depends on the methodology of FE and its subsequent FS.

1.2 Approach

Fig. 1.2 shows the schematic approach for determining the level of damage or anomaly based on SHM employing ultrasonic guided waves.

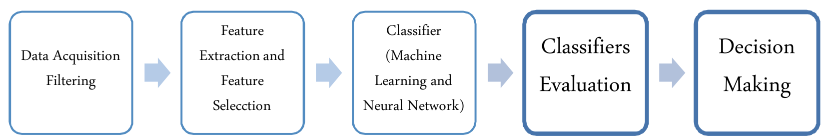

Fig. 1.2 Methodology schedule for determining the level of damage or anomaly.

1.2.1 Signal Pre-processing

Sensors generate noise from a variety of sources and therefore a signal pre-processing is necessary to eliminate/reduce the dataset that does not provide useful informatio [45]. Standard statistical techniques have been employed in references [10, 53]. Yu et al. [64] were able to reduce global noise using averaging techniques and Daubechies Wavelet (DW) to eliminate local high-frequency perturbations. Denoising and compression signal in Guided Waves (GW) based on the Discrete Wavelet Transform (DWT) was employed by Rizzo and di Scalea [55]. There are a large number of research and reviews articles on filtering in the treatment of ultrasonic guided waves [46]. Hamming [21] performs a review about low-pass filters available for data smoothing. This paper will consider Wavelet transforms for filtering the signals.

The denoising of the signal is performed employing a multilevel 1-D wavelet analysis using Daubechies family. The wavelet decomposition structure of the signal is extracted. The threshold for the de-noising is obtained by a wavelet coefficients selection rule using a penalization method provided by Birg-Massart [47]. An overly aggressive filtering could eliminate data that should show, for example, small echoes that come from defects. Fig. 1.3 shows the original signal and the de-noised signal when it is applied the wavelet de-noised filter. The Wavelet de-noising filter does not produce an unwanted signal delay in contrast to other digital filters.

It is observed that the filter removes noise significantly, and does not eliminate information that is related to different structural features.

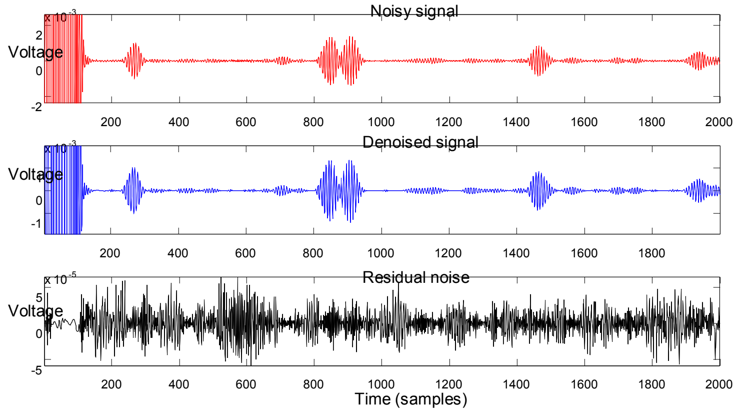

Fig. 1.3 Decomposition detail five (D5), De-noised D5 and extracted residual noise.

1.2.2 FE Methods

FE affects the learning and classification process. FE and FS are capable of improving learning performance, presenting less computational complexity and historical dataset [58].

Classical dimensional reduction techniques such as Principal Component Analysis (PCA) [26] and Multidimensional Scale (MDS) [29] are extensively employed. However, the Linear and Nonlinear Regression systems are used when the signal type is complex and the first order interactions are not enough to derive good results [20].

Two autoregressive models have been applied in this paper for FE: Linear (AR) and nonlinear (NARX).AR is a lineal mathematical model commonly employed for FE because of its high computational performance. AR method is popular because estimation of AR parameters (features) is achieved easily by solving linear Yule-Walker equations [13]. NARX presents a better behavior in the most of the nonlinear systems. NARX was introduced by Leontaritis and Billings [30] and it has been widely used in the identification of nonlinear dynamic models and feature selection [6]. NARX presents a good generalization capability, as effective training algorithm, a fast convergence rate, and can represent a wide class on nonlinear systems [3, 18, 25].

Within the NARX model, the lineal-in-the-parameters structure is a method widely used in nonlinear dynamics models. It is a linear combination of model terms, or basis functions, that are some nonlinear-usually polynomial- functions of the system variables [32]. In this case, the monomials of the linear combination are the features. The linear-in-the-parameter models are well structured to adaptive learning, with learning conditions and demonstrable convergence and capacity parallel processing, in addition to numbers applications in signal processing and pattern recognition [24, 36].

1.2.3 FS Methods

FS techniques can be categorized into three approaches: embedded, wrapper or filter [33]. The main methods for FS are Information Gain, Relief, Fisher Score and Lasso [27]. In this paper, the main objective to reduce dimensionality and it is not intended to select the best features.

SHM literature propose several methods to select the AR model order: AIC [2]; Akaike's Final Prediction Error [1](FPE); Partial Autocorrelation Function (PAF) [60], and; Root Mean Square (RMS) [31]. AIC statistical technique has been used to reduce the dimensionality of the feature extraction. The AIC is one of the most efficient techniques for order optimization [35]. The AIC is a measure of the goodness-of-fit of an estimated statistical model, based on the trade-off between fitting accuracy and the number of estimated parameters.

1.2.4 Classifiers

(1) Machine Learning Approach
① Decision Trees

Decision Tree (DT)is a classifier used in many fields to study if the data contains different classes of objects that can be interpreted significantly in the context of a substantive theory [34, 48, 51, 52].

DT generates a split of space from a labeled training set. The objective is to separate the elements of each class into different labeled regions (leaves) minimizing the local error. Each internal node in the tree is a question (decision) that determines which branch of the tree must be taken to reach a leaf. DT is determined for; 1) how to split the space (Splitting Rules); 2) stop condition of splitting; 3) labeling function of a region, and; 4) measurement of error.

The purpose of the Splitting Rule is to minimize the impurity of the node. The recursive splitting algorithm stops when it finds any of the following conditions: the node is pre-set maximum deep; all elements of the node are same class; there is no empty sub node or; SR does not reach a pre-set value.

To label a leaf or region once it is already considered as a terminal is considered to develop a DT. Eq. (1.1) establishes the labeling function.

$$l' = \arg\min l' \left\{ \sum_{l=1}^{k} N_I \times c_{l,l'} \right\}, \tag{1.1}$$

where N_I is the number of elements of class l in the region, l' is the class to label and $c_(l,l')$ is the labeling cost.

$c_(l,l')$ with all classes is calculated and l' is selected, which minimizes the error. The label that minimizes the error of a region is the most populated class. In case of a tie, a random one is chosen. Eq. (1.2) gives the classification average error

$$\varepsilon = \frac{1}{N}\sum_{l=1}^{N} \mathrm{err}(l, l_i'), \tag{1.2}$$

where $\mathrm{err}(l, l_i')$ is the error of labeling a class l as l'. This error is solved by splitting the space and assigned to each split a label.

② Discriminant Analysis (DA)

The ultrasonic signals considered are homogeneous for the same frequency and, therefore, the classifier that provides better results is LDA [15]. This classifier is valid for ultrasonic signals obtained at different frequencies and used for speech recognition. This type of classifier is based on a geometrical approach. A linear function representing a decision limit divides a feature space into regions that have common properties. The aim of this method is to find the expression of linear discriminant functions that enable the objects classification in the considered classes.

③ Support Vector Machine (SVM)

SVM [9] is a supervised multivariate classification method. Supervised refers to a training step where the algorithm learns the differences between pre-specified groups to be classified [59]. SVM treats each feature as a point in a high-dimensional space, being the number of dimensions the same to the number of rating levels. Each feature is assigned to a group and the linear classification function (1.3) learns the characteristics to discriminate among five groups. A limit of decision, or hyperplane (a generalization of a plane of n-1 dimensions which divides an n-dimensional space), must be defined to separate the data based on class membership and to classify linearly the dataset. However, for a linearly separable problem, there are an infinite number hyperplanes correctly classified data. SVM algorithm finds the optimal one characterized by the largest margin between classes. The margin is defined as the distance of the closest training data points of the hyperplane. These points are the most difficult to classify and they are called support vectors. The hyperplane is defined by a weight vector, which is a linear combination of the support vectors, and specifies both a direction and a displacement which together define the maximum margin classifier. The decision function $D(x)$ is given by Eq. (1.3),

$$D(x) = w\emptyset(x) + b, \tag{1.3}$$

where w and b are the SVM parameters, and $\emptyset(x)$ is a kernel function

The hyperplane is defined by Eq. (1.3), and the distance between the hyperplane and pattern x can be written by Eq. (1.4).

$$\frac{D(x)}{\|w\|} \tag{1.4}$$

A training classifier is designed to find the value of w that maximizes the margin between the class boundary and the training patterns. The objective function of the training algorithm is given by Eq. (1.5).

$$J = \frac{\|w\|^2}{2}. \tag{1.5}$$

In this case, the linear function Kernel and one vs one method have been employed [43].

④ k-Nearest Neighbors

k-Nearest Neighbors(k-NN) [11, 12] is a high-performance classifier widely used in Machine Learning [42]. k-NN rule classification is an extension of the Nearest-Neighbor (NN) rule. Given a set x of the k samples, the rule assign to each sample is to label most frequently represented among the k nearest samples [15]. k-NN search technique and k-NN based algorithms are widely used as benchmark learning rules. The relative simplicity of the k-NN search technique makes it easy to compare the results from other classification techniques to k-NN results.

The accuracy of k-NN classification significantly depends on the metric used to compute distances between different samples [61]. In most cases, the best performing classifier is Fine k-NN [62], using metric Euclidean distance in Eq. (1.6).

$$d_{st}^2 = (x_s - y_t) \cdot (x_s - y_t)' \tag{1.6}$$

⑤ Ensemble Subspace Discriminant (ESD)

Ensemble methods are learning algorithms that construct a set of classifiers whose individual decisions are combined in some way (generally by weighted or unweighted voting) to classify the set of features of each pattern [14]. Bagging, boosting, and random subspaces are general techniques that can be used with any type of base classifier. Ensemble Bagged Tree (EBT) and Random Subspace Method (RSM) have been the most successful methods. EBT uses the Breiman's 'random forest' algorithm [7]. RSM is a parallel learning algorithm proposed by Ho [23].

(2) Artificial Neuronal Network (ANN)

ANN unidirectional supervised through a MLP with training by backpropagation algorithm [56] has been applied. Backpropagation with algorithm scaled conjugate gradient and performance Cross Entropy [28, 44] with 'Early Stopping' to avoid overfitting [50] has been training mode. ANN is given by Eq. (1.7),

$$z_k = \sum_j w'_{kj} y_j - \theta'_i = \sum_j w'_{kj} f\left(\sum_i w_{ji} x_i - \theta_j\right) - \theta'_j, \tag{1.7}$$

where x_i is ANN input, y_i is hidden layer output, z_i is final layer output, t_k is targets output, $w_j i$ is hidden layer weight, w'_{kj} is final layer weight, θ_j is hidden layer bias, θ'_k is final layer bias and $f(.)$ is the activation function of sigmoid type, employed as the activation function of the ANN.

The MLP process tests initially with an ANN architecture and is trained with 70% of the total of the experiments and 30% of them to test the network. Then the ANN architecture is chosen according to the accuracy and performance. Finally, ANN is tested with different cases (30%) to know if the learning is right and to the accuracy.

Backpropagation (BP) is one of the simplest and most general methods for supervised training of multi-layer ANNs. Scaled, standardization, normalization that perform pre-processing inputs techniques have been employed to accelerate the BP training. Scaled conjugate gradient and performance Cross Entropy are employed to increase the computational learning cost of ANN.

1.2.5 Evaluation Classifier

Analysis Receiver Operating Characteristic (ROC), as Confusion Matrix (CM), is employed to evaluate of the classification. CM determines the accuracy of a classifier and measures its performance. The main parameters in CM are:

True Positive (TP) : corresponds to real successes of the classifier.
False Positives(FP) : is the sum of the values of a class in the corresponding CM column, excluding the TP.
False Negative(FN) : is the sum of the values of a class in the corresponding CM row, excluding the TP.
True Negatives(TN) : is the sum of all columns and rows, excluding that class's column and row.

Recall function, R, given by Eq. (1.8), provides the probability to be correctly classified, called true positive rate or hit rate:

$$R = \frac{TP}{TP + FN}. \tag{1.8}$$

Specificity, S, also called the true negative rate, measures the proportion of negatives that are correctly identified as negatives.

$$S = \frac{TN}{FP + FN}. \tag{1.9}$$

Precision, P:

$$P = \frac{TP}{FP + TP}. \tag{1.10}$$

F-score, F [54]

$$F = \frac{2 \times P \times R}{P + R}. \tag{1.11}$$

There are two conventional methods to establish the average performance in all categories: macro-averaging and micro-averaging [63]:

Macro-average: P^M, R^M and F^M is obtained are the mean of all P_i^M, R_i^M and F_i^M where M denotes Macro-average, and i is the scenario, then for each category they are calculated, i.e. they are valued are evaluated locally P_i^M, R_i^M and F_i^M, and then globally P^M, R^M and F^M.

Micro-average: P^M, R^M and F^M value is obtained as: i) TP_i, FP_i, FN_i values is calculated for each of the scenario; ii) the value of TP, FP and FN are calculated as the sum of TP_i, FP_i and FN_i, and; iii) to applying the equation of the measure that correspond.

There are several indices extracted from the ROC curve to evaluate the efficiency of a classifier. Area Under Curve (AUC) represents the area between the ROC curve and the negative diagonal [5, 22], being between 0.5 and 1. $Values \leq 0.5$ indicates that the classifier is invalid and a value of 1 indicates a perfect rating because there is a region in which, for any point cut, the values of R and Pare equal to unit. The statistical property of AUC is equivalent to the Wilcoxon test of ranks [22]. The AUC is also related to the Gini coefficient [8], which is twice the area between the diagonal and the ROC curve.

The recommendations by Demsar [4], and the extensions by Garcia and Herrera [17], have been employed to perform the comparative analysis of classifiers. Friedman Test will be used to test the null hypothesis that all classifiers achieve the same average. Bonferroni-Dunn Test is applied to determine significant differences between the top-ranked classifier and the following. Holm Test is used to contrast the results.

1.3 Approach Scheme

The methodological process is represented schematically in Fig. 1.4. Firstly, for each damage or anomaly, the best classifier is selected.

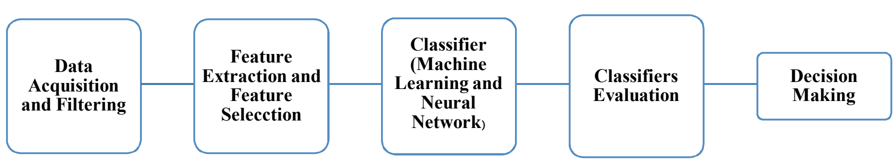

Fig. 1.4 Methodology flow for determining the classifier.

The methodological process will follow the scheme given in Fig. 1.5 when the best classifier for each damage is set.

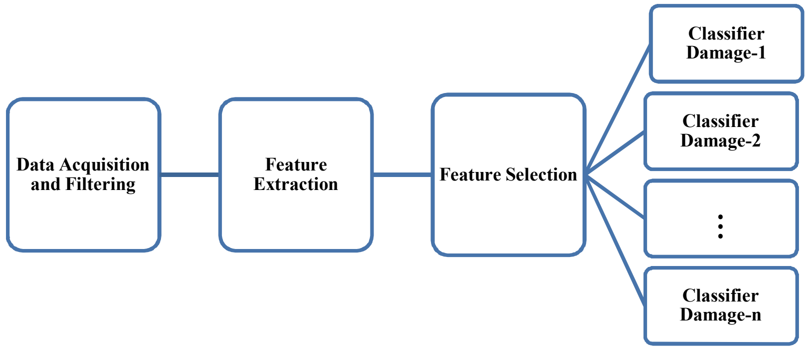

Fig. 1.5 Methodology flow for determining the damage.

1.4 Conclusions

The paper presents a novel approach to optimize the sensors in a condition monitoring system employing ultrasonic waves. The approach can detect different potential faults with a single signal emitted by a sensor, such as delamination, mud or ice on blades of wind turbines. This methodology allows to avoid the redundancy of sensors, since a specific number of ultrasonic transducers can determine the structural condition using guided waves. The signal is pre-processed with the aim of removing the noise, then extracted and selected features to be then classified by Machine Learning and Neural Networks. Finally, for each damage or anomaly, the best classifier will be evaluated. The best classifier of each damage will work in a parallel network that will process the signal sent by the sensor.

Acknowledgements The work reported herewith has been financially supported by the Spanish Ministerio de Economía y Competitividad, under Research Grant DPI2015-67264-P and RTC-2016-5694-3.

References

1. Akaike H (1969) Fitting autoregressive models for prediction. Annals of the Institute of Statistical Mathematics 21(1):243–247
2. Akaike H (1974) A new look at the statistical model identification. Automatic Control IEEE Transactions on 19(6):716–723
3. Alazrai R, Lee CSG (2012) An narx-based approach for human emotion identification. In: Ieee/rsj International Conference on Intelligent Robots and Systems, pp 4571–4576
4. Ar J (2006) Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7(1):1–30
5. Bradley AP (1997) The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognition 30(7):1145–1159

6. Brankovic A, Falsone A, et al (2016) Randomised algorithm for feature selection and classification. Xiv preprint
7. Breiman L (2001) Random forests. Machine Learning 45(1):5–32
8. Breiman L, Friedman J, et al (1984) Classification and regression trees. CRC press
9. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2(2):121–167
10. Chopra I (2002) Review of state of art of smart structures and integrated systems. AIAA Journal 40(11):2145–2187
11. Cover T, Hart P (1967) Nearest neighbor pattern classification. IEEE Transactions on Information Theory 13(1):21–27
12. Dasarathy BV (1990) Nearest neighbor (nn) norms: nn pattern classification techniques. Los Alamitos IEEE Computer Society Press 13(100):21–27
13. De Lautour OR, Omenzetter P (2010) Damage classification and estimation in experimental structures using time series analysis and pattern recognition. Mechanical Systems and Signal Processing 24(5):1556–1569
14. Dietterich TG (2000) Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Systems, pp 1–15
15. Duda RO, Hart PE, Stork DG (2012) Pattern classification. John Wiley and Sons
16. Farrar CR, Doebling SW, Nix DA (2001) Vibrationcbased structural damage identification. Philosophical Transactions of the Royal Society B Biological Sciences 359(359):131–149
17. Garc S, Herrera F (2008) An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. Journal of Machine Learning Research 9(12):2677–2694
18. Ghosh S, Maka S (2009) A narx modeling-based approach for evaluation of insulin sensitivity. Biomedical Signal Processing and Control 4(1):49–56
19. González-Carrato RRDLH, Márquez FPG, Dimlaye V (2015) Maintenance management of wind turbines structures via mfcs and wavelet transforms. Renewable and Sustainable Energy Reviews 48:472–482
20. Guyon I, Elisseeff A (2006) An introduction to feature extraction. Springer Berlin Heidelberg
21. Hamming RW (1989) Digital filters (3rd ed.). Prentice Hall
22. Hanley JA, Mcneil BJ (1982) The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology 143(1):29–36
23. Ho TK (1998) The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(8):832–844
24. Hong X, Mitchell R, et al (2008) Model selection approaches for non-linear system identification: a review. International Journal of Systems Science 39:925–946
25. Jiang C, Song F (2010) Forecasting chaotic time series of exchange rate based on nonlinear autoregressive model. In: Advanced Computer Control (ICACC), 2010 2nd International Conference on, pp 238–241
26. Jolliffe I (2002) Principal component analysis. Wiley Online Library
27. Jovic A, Brkic K, Bogunovic N (2015) A review of feature selection methods with applications. In: International Convention on Information and Communication Technology, Electronics and Microelectronics, pp 1200–1205
28. Kroese DP, Rubinstein RY, et al (2013) Cross-entropy method. In: Encyclopedia of Operations Research and Management Science, Springer, pp 326–333
29. Kruskal JB, Wish M (1978) Multidimensional scaling. Book on Demand Pod
30. Leontaritis IJ, Billings SA (1985) Input-output parametric models for non-linear systems part ii: stochastic non-linear systems. International Journal of Control 41(2):303–328
31. Levinson N (1949) The wiener (root mean square) error criterion in filter design and prediction. Studies in Applied Mathematics 25(1-4):261C278
32. Li K, Peng JX, Bai EW (2006) A two-stage algorithm for identification of nonlinear dynamic systems. Automatica 42(7):1189–1197
33. Liu H, Motoda H (2008) Less is more. Computational Methods of Feature Selection, H Liu and e Motoda, H(Eds) Chapman and Hall/CRC pp 3–17

34. Loh WY, Shih YS (1997) Split selection methods for classification trees. Statistica Sinica 7(4):815–840
35. Lordo RA (2001) Learning from data: Concepts, theory, and methods. Technometrics 43(1):105–106
36. Márquez FPG, Munoz JMC (2012) A pattern recognition and data analysis method for maintenance management. International Journal of Systems Science 43(6):1014–1028
37. Márquez FPG, Pedregal DJ, Roberts C (2013) New methods for the condition monitoring of level crossings. In: International Journal of Systems Science, pp 878–884
38. Márquez FPG, Pardo IPG, Nieto MRM (2015) Competitiveness based on logistic management: a real case study. Annals of Operations Research 233(1):157–169
39. Márquez FPG, Pérez JMP, et al (2016) Identification of critical components of wind turbines using fta over the time. Renewable Energy 87:869–883
40. Martinez-Luengo M, Kolios A, Wang L (2016) Structural health monitoring of offshore wind turbines: A review through the statistical pattern recognition paradigm. Renewable and Sustainable Energy Reviews 64:91–105
41. Marugán AP, Márquez FPG, Pérez JMP (2016) Optimal maintenance management of offshore wind farms. Energies 9(1):46
42. Michalski RS, Carbonell JG, Mitchell TM (2013) Machine learning: An artificial intelligence approach. Springer Science and Business Media
43. Milgram J, Cheriet M, Sabourin R (2006) "one against one" or "one against all": Which one is better for handwriting recognition with svms? Proceedings of International Workshop on Frontiers in Handwriting Recognition
44. Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 6(4):525–533
45. Muñoz JMC, Márquez FPG, Papaelias M (2013) Railroad inspection based on acfm employing a non-uniform b-spline approach. Mechanical Systems and Signal Processing 40(2):605–617
46. Munoz CG, Márquez FG (2016) A new fault location approach for acoustic emission techniques in wind turbines. Energies 9(1):40
47. Munoz CQG, Márquez FPG, Tomás JMS (2016) Ice detection using thermal infrared radiometry on wind turbine blades. Measurement 93:157–163
48. Pal M, Mather PM (2003) An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment 86(4):554–565
49. Papaelias M, Cheng L, et al (2016) Inspection and structural health monitoring techniques for concentrated solar power plants. Renewable Energy 85:1178–1191
50. Prechelt L (1998) Automatic early stopping using cross validation: Quantifying the criteria. Neural Networks the Official Journal of the International Neural Network Society 11(4):761
51. Quinlan JR (1986) Induction of decision trees. Machine Learning 1(1):81–106
52. Quinlan JR (2014) C4. 5: programs for machine learning. Elsevier
53. Raghavan AC, Cesnik CES (2007) Review of guided-wave structural health monitoring. Shock and Vibration Digest 39(2):91–114
54. Rijsbergen CJV (1979) Information retrieval. Butterworth-Heinemann
55. Rizzo P, Scalea FLD (2004) Discrete wavelet transform to improve guided-wave-based health monitoring of tendons and cables. Proceedings of SPIE - The International Society for Optical Engineering 5391:523–532
56. Rumelhart DE, McClelland JL, et al (1988) Parallel distributed processing, vol 1. IEEE
57. Staszewski WJ, Robertson AN (2007) Time-frequency and time-scale analyses for structural health monitoring. Philosophical Transactions of the Royal Society A Mathematical Physical and Engineering Sciences 365(1851):449
58. Tang J, Alelyani S, Liu H (2014) Feature selection for classification: A review. Documentación Administrativa pp 313–334
59. Vapnik VN (1999) An overview of statistical learning theory. IEEE Transactions on Neural Networks 10(10):988–999
60. Wei WWS (1994) Time series analysis. Addison-Wesley publ Reading

61. Weinberger KQ, Saul LK (2009) Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research 10(1):207–244
62. Xu Y, Zhu Q, et al (2013) Coarse to fine k nearest neighbor classifier. Pattern Recognition Letters 34(9):980–986
63. Yang Y (1999) An evaluation of statistical approaches to text categorization. Information Retrieval Journal 1(1):69–90
64. Yu L, Bao J, Giurgiutiu V (2004) Signal processing techniques for damage detection with piezoelectric wafer active sensors and embedded ultrasonic structural radar. Proceedings of SPIE - The International Society for Optical Engineering 5391:492–503

Author Index

J. Xu et al. (eds.), *International Conference on Management Science and Engineering Management 2017*,

Advances in Intelligent Systems and Computing XXX, DOI: XXXXXXXXXX