Data-driven Fault Diagnosis of Oil Rig Motor Pumps Using a Novel Multiple Classifier System

Estefhan D. Wandekoken, Eduardo Mendel,
Fábio Fabris, Marcelo Valentim,
Flávio M. Varejão, Thomas W. Rauber
Department of Computer Science
Federal University of Espírito Santo
Vitória, ES, Brazil
thomas@inf.ufes.br

Rodrigo J. Batista
Espírito Santo Exploration and Production Business Unit. Automation, Inspection and Maintenance Engineering.
Petróleo Brasileiro S.A. PETROBRAS
Vitória, ES, Brazil

Abstract—We report about fault diagnosis experiments to improve the maintenance quality of motor pumps installed on oil rigs. Our work is motivated by the diversity of the studied defects and the availability of real data from operational oil rigs. In this work we present a fault diagnosis system that is better suited to overcome the difficulties that arise from real-world fault diagnosis, for instance the occurrence of multiple coexistent defects. Each fault is predicted by a distinct ensemble of Support Vector Machine (SVM) classifiers which differ among themselves on the feature set they use as well as on their intrinsic parameters. In order to build the ensemble we apply a novel approach based on the outputs of several stepwise wrapper feature selection methods. Our method requires a minimum of a priori knowledge about the plant because the faults predictor is automatically defined based on training data, allowing the method to be easily extended to many equipments, sensors, and failures types.

Keywords- feature selection, ensemble classifier system, motor pumps, SVM, diagnosis, classification.

I. INTRODUCTION

The detection and diagnosis of faults in complex machinery is advantageous for economical and security reasons [1]. The main objective is to detect defects still at an early stage, while the machine is under normal operating conditions. Thereby damaged components can be repaired during planned maintenance, which minimizes machinery standstill and increases security. The fault diagnosis module of an intelligent monitoring system aims to aid the human users in the monitoring tasks, by outputting the system predictions. However, a single reliable diagnosis procedure for any type of fault based on noninvasive signals is not established yet [2]. To contribute to this field we present a novel, artificial-intelligence-based fault predictor, which is able to analyze data from diverse sources like electrical, chemical, thermal and mechanical vibration sensors. One advantage of our method is that it requires a minimum of a priori knowledge about the plant because the faults predictor is automatically defined based on training data, allowing the method to be easily extended to many equipments, sensors, and failures types.

In this work the equipment we focus on are horizontal motor pumps with extended coupling between the electric motor and the pump, and the faults are detected by analyzing vibrational signals. Accelerometers are placed along the main directions to capture specific vibrations of the main shaft which provides a multichannel time domain raw signal. We further apply signal processing techniques like Fourier Transform or Envelope analyses in order to extract the features that describe the vibrational signals. Several faults can simultaneously occur on the motor pump equipment, and the faults predictor has to be robust enough to overcome this difficult. For instance, many faults cause vibrations in similar frequency bands, like the first, the second, and the third harmonics of the shaft rotation frequency, in such a way that the faults can not be detected by just seeking for their well-known characteristic signatures.

We use the supervised learning [3] classification paradigm as the primal mechanism to automatically generate the faults predictor. A strong motivation of our work is the availability of statistically significant amounts of real data from operational oil rigs. The total amount of 2000 examples of vibrational signals obtained from operational motor pumps was regularly acquired from 25 different oil platforms off the Brazilian coast during the period of five years. Each example presents at least one of the six types of considered faults (and also, rarely, other faults). The examples labels were assigned by human experts in maintenance engineering. We consider the fault diagnosis problem as a multi-label [5] classification task in which several distinct labels (fault classes) can be simultaneously assigned to an example. To deal with that we build an independent binary predictor for each considered fault type which outputs whether an input pattern presents or not that considered fault. Then, the decision about which faults are present in an input signal is taken simply by assigning to it the result of the predictor of each fault type.

A central issue regarding real-world fault diagnosis is the definition of which aspects of the input signals, i.e features, the predictor of each defect should analyze. The traditional approach is the manual definition of the features used by the predictor of each fault type. Nevertheless, in [7] we showed that the manual setting of the used features demands specialized knowledge and can result in low accurate classifiers due to the occurrence of multiple coexistent faults. An approach for building a fault predictor avoiding the manual definition of the important features relies on the initial extraction of a large, comprehensive feature set, and then the usage of a method for automatically defining an optimized way of

employing those features. For instance, a popular approach is to perform *feature selection* [6] in order to build a single classifier which uses just a reduced, optimized feature set. However, recent research in *ensemble of classifiers* [14] indicates that much better generalization power can be achieved when the final decision is not taken only by a single accurate classifier, but taken as a combination of decisions given by many distinct accurate classifiers that collectively disagree on their predictions as much as possible.

In this work we propose a novel ensemble creation method to better deal with the real-world fault diagnosis domain. Each considered fault is predicted by a distinct ensemble of classifiers. In order to build it, first, different feature selection methods are used to produce a set of classifiers that are the candidates to compose the ensemble. Each of those classifiers is induced using a feature set defined by taking a distinct number of features selected by a feature selection algorithm, and we also automatically tune the intrinsic parameters of each classifier aiming to increase the individual accuracies and therefore the ensemble accuracy. We perform *wrapper* [6] feature selection, using stepwise greedy search strategies, and evaluate the merit of a candidate feature set as being the value of the Area Under the ROC Curve (AUC) [4] of a Support Vector Machine (SVM) [13] classifier which uses that feature set. After building those candidate classifiers, a predefined number of SVM classifiers are chosen to finally compose the ensemble by employing a forward stepwise search, selecting classifiers to compose the ensemble (instead of selecting features to compose a single classifier).

II. FEATURE SELECTION

Feature selection is the process of identifying and selecting an optimized subset of features from a larger set that may contain irrelevant or redundant features. That feature subset is further used to describe the input space of an induced classifier. In this work, we apply the wrapper approach to feature selection, in which the learning algorithm itself is used to estimate the usefulness of features by evaluating classifiers that use those candidate feature sets. Thus, the selection criterion J used to estimate the performance of a candidate feature set F_k is the cross-validation Area Under the ROC Curve (AUC) [4] on training data achieved by a SVM classifier that uses the feature set F_k. We focus on stepwise greedy search algorithms due to their good compromise between computational cost and covering of the search space, and employ the following feature selection techniques [6]: Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS). The Sequential Forward Selection (SFS) search starts with an empty set of currently selected features. At each step, one feature is definitely included in the set of selected features. Consider that k < d features have already been selected and included in the feature set X_k. If G is the total set of all $|G|$ features, then $G \setminus X_k$ is the set of $|G| - k$ candidates f_j. To include one more feature in X_k, each non-selected feature f_j must be tested individually together with the already selected features (being individually included in X_k) and ranked according to the criterion J. The Sequential Backward Selection (SBS) method operates in a similar way of SFS, but the latter includes features, while the former removes features. Thus SBS starts with every feature of G in X_k and at each step one feature is removed from X_k, namely the one that provides the highest criterion J with its individual exclusion from X_k.

III. ENSEMBLE OF CLASSIFIERS

Combining decisions of multiple predictors into an ensemble decision is becoming one of the most important techniques for improving classification accuracy. An efficient ensemble is formed by predictors that are very accurate and also diverse, in the sense that the prediction errors occur in different regions of the feature space. Three approaches have become popular for achieving diversity between the classifiers within an ensemble: using a different subset of training data for each classifier; setting different parameters for each classifier; and employing a distinct feature subset for each classifier. In this work we employ both the second and third approaches (we do not employ the first approach as we desire to use every available data for training the classifiers).

In order to combine the decisions of the individual classifiers of the ensemble we perform aggregation by averaging due to its simplicity and good results. Thus the estimated a posteriori probability value $P(\mathbf{x})_{pos}$ of a pattern $\mathbf{x}$ belonging to the positive class ω_{pos} is calculated as the average of the posterior probabilities values $P(\mathbf{x})^{c_l}_{pos}$ that each classifier c_l of the ensemble outputs for $\mathbf{x}$. Thereby the ensemble predicts $\mathbf{x}$ as belonging to ω_{pos} if $P(\mathbf{x})_{pos} > 0.5$ or to ω_{neg} otherwise.

A. Previous Works in Ensemble of Classifiers

The first proposed ensemble methods relied on using a different training subset for each classifier, for instance *Bagging* [14] and *Boosting* [15]. Both methods have been employed for fault diagnosis (see for instance [11], [12], [10]). The Random Subspaces Method (RSM) was the first proposed algorithm for creating an ensemble by varying the feature subsets of the individual classifiers. In RSM, the features of each classifier are randomly taken from the initial feature set (the number of classifiers and the number of features per classifier are predefined parameters). Some works have further proposed ensemble feature selection methods in order to improve the accuracy of the RSM ensemble [16], [17]. See for instance Hill Climbing (HC) [17] in which, for each feature of the feature set of each classifier, an evaluation is done with that feature being selected (if not previously selected) or removed (if it is selected); that change is kept for the next step if the ensemble accuracy increases.

The major drawback in the RSM-based methods is the possibility of producing an ensemble that lacks in accuracy due to severe over-fitting to the training data, as the features of the classifiers within the ensemble are individually selected aiming to increase the whole ensemble accuracy. On the other hand, the ensemble creation method proposed in this work relies on the

usage of different feature selection methods aiming to generate the highest accurate individual classifiers as possible, which tends to generate an accurate ensemble.

B. The Proposed Ensemble Classifier System

After extracting the global pool of features G composed of many relevant features and probably also some irrelevant or noisy ones, composed by $|G|$ features, we employ m distinct stepwise greedy feature selection methods, $\{S_1, \ldots, S_i, \ldots, S_m\}$, which can operate in a forward or backward manner. We want each S_i to determine feature sets with every possible distinct cardinality, thus we require Si to select $|G|-1$ features if S_i is a forward search or we require S_i to select 1 feature if S_i is a backward search, so the experiment outputs the feature set $X_k^{S_i}$ associated to each number k of selected features. We also consider the global pool of feature G itself. Thereby Ξ is composed by every produced $X_k^{S_i}$ and also G. The next step consists in producing a set of classifiers C composed by candidate classifiers to compose the ensemble. For that, for each distinct $X_k^{S_i}$ in Ξ, we build a classifier c_j that uses it, and in order to increase the accuracy of j and thereby the ensemble accuracy, we automatically tune its intrinsic parameters. The last step is the selection of an optimized, reduced set of produced classifiers to compose the ensemble. For that, we employ a forward search, but now selecting classifiers to compose the ensemble, instead of selecting features to compose a single classifier. The selection criterion J_e is the AUC achieved by the candidate ensemble of classifiers. Figure 1 presents a diagram of the training process of the proposed ensemble.

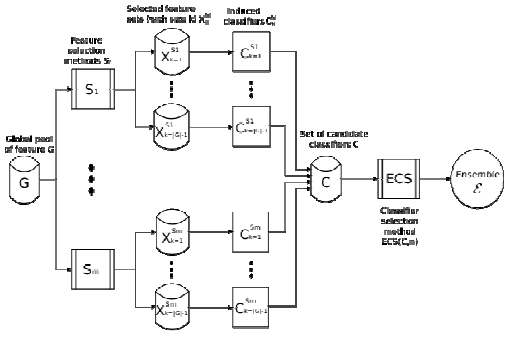

Figure 1. A diagram of the training process of the predictor of a specific kind of fault built as the proposed ensemble of classifiers. First, several SVM are produced, with tuned parameters and with different feature sets. Then, a reduced, optimized set of SVMs is selected to compose the ensemble.

IV. Experimental Results

In order to perform the experiments, we randomly divided the complete database of 2000 examples into a pair of approximated 1000 training examples and the remaining 1000 test examples, in a stratified manner in order to have approximately the same number of examples of each fault class in each database. Then one of the databases is used to train the classifier (training database) and the other is used to test it (test database), resulting in the accuracy achieved in that test database. Following that, we use the databases in a reversed manner, with the one used for training now used for test and vice-versa. This procedure was repeated five times, obtaining in this way ten different train-test pairs. The final accuracy on test data is therefore estimated averaging the ten different test accuracies. This estimation process is called 5x2 cross-validation.

In order to train the SVM classifier we use the *libsvm* library, employing the C-SVM [3] SVM architecture with a radial basis function kernel; thus two parameters must be defined, C and γ. The *libsvm* library provides us with a parameter search technique to identify good parameters that maximizes cross-validation accuracy estimated on training data, taking the pair of values (from a set of predefined pairs) that provided the highest cross-validation accuracy.

A. Initially Extracted Global Pool of Features

Every fault classifier uses the global pool of features G which is composed of 67 features extracted from the Fourier spectrum of the vibrational signals and also 28 features from the Envelope spectrum. Each feature is extracted from a given position (1, 2, 3 or 4) and from a given direction of measurement (vertical, horizontal or axial). For a given motor pump, the features of G are extracted from each of the three directions, selecting from a subset of positions depending on the fault under consideration (for instance, taking from positions 2 or 3 for misalignment fault).

The Fourier features correspond mostly to the RMS values of some important bands in the frequency spectrum of the velocity signal, for instance the bands 0.9x-1.1x, 1.4x-1.6x, 1.9x-2.1x, ..., 5.4x-5.6x; the 10% large narrow band around the pump blade pass frequency (BPF) and its harmonics; and the RMS value of the noise in important bands (such as 0x-3.0x), calculated with the median filter. The features extracted from the Envelope spectrum that compose the global pool G are defined as narrow bands around the first five harmonics of the bearing characteristic frequencies (BPFI, BPFO, FTF and BSF).

B. Faults Predictor Creation and Performance

In order to build each fault predictor, after extracting the global pool of features G we ran two distinct feature selection experiments, namely SFS and SBS. The selection criterion J was the AUC achieved by a SVM, estimated by 10-folds cross-validation on the training data. The SVM parameters were set as $C=8.0$ and $\lambda=0.5$ for feature selection. Following, for each selected feature set, we built a SVM classifier with tuned parameters C and λ. Finally, we employed a SFS search in order to select a total of 10 SVM classifiers in order to compose the ensemble, with the selection criterion J_e being the AUC achieved by the candidate ensemble of SVM classifiers.

Table I presents the accuracy estimated on test data by 5x2 cross-validation, individually achieved by the classifier of each considered fault. Our objective is to show the superiority of the proposed ensemble classifier. Thus, we compare it to the accuracies obtained by the

following models: a single SVM classifier that uses the complete global feature set *G*, i.e. that uses a non-optimized high-dimensional feature space; and a single SVM classifier with an optimized feature set, namely the one that achieved the highest training cross-validation AUC considering both feature selection experiments. Both SVMs are induced using tuned parameters.

The results of table I indicate that the usage of a single SVM as fault predictor performed worst than the usage of the ensemble of SVM classifiers built by BSFS. It is interesting to note that the usage of feature selection to retain an optimal feature set, a method widely used for fault diagnosis, achieved an accuracy that is just slightly higher than the one achieved by SVM with the complete global pool of features. That fact suggests that the SVM classification architecture is well adapted to deal with unnecessary features. Satisfactorily, we see the superiority of the proposed SVM ensemble when compared to a single SVM, as for every considered fault it achieved the higher accuracy, thanks to the averaging of the decisions of accurate, divergent classifiers.

TABLE I. TEST DATA CLASS DISTRIBUTION AND ACCURACY

Fault Classifier	Percen-tage of Negative Class Data	Single SVM with Global Feature Set	Single SVM with Optimized Feature Set	Proposed Ensemble classifier
Misalign-ment	57.4%	75.9%	75.6%	**79.0%**
Bearing	64.3%	83.5%	86.1%	**87.9%**
Unbalance	75.1%	81.4%	81.3%	**82.9%**
Hydrody-namical	57.6%	84.0%	85.0%	**86.7%**
Structural Looseness	78.8%	80.5%	81.2%	**81.7%**
Mechanical Looseness	88.0%	90.0%	90.0%	**90.5%**

V. CONCLUSIONS AND FUTURE WORK

We presented supervised learning classification experiments applied to data-driven fault diagnosis of operational motor pumps installed at oil rigs. We evaluate effectiveness of the proposed ensemble architecture in real-world operating industrial machines instead of using well-behaved data from a controlled laboratory environment which is almost always found in literature. We propose a novel ensemble architecture based on the automatic variation of the feature sets of the classifiers within an ensemble, each one using a distinct number of features. Our method initially extracts a large and comprehensive feature set, and then automatically defines how and which of those various features are used by the diagnosis module. Future developments will analyze the benefits of robustly introducing diversity within the ensemble in order to build it, developing a metric that explicitly expresses both individual accuracy and collective divergence among the classifiers of an ensemble.

ACKNOWLEDGMENT

We would like to thank CNPq (Grant n. 620165/2006-5) and COPES-Petrobrás for the financial support given to the research from which this work originated.

REFERENCES

[1] P. J. Tavner and J. Penman, Conditiong Monitoring of Electrical Machines. Wiley, 1987.

[2] A. Bellini, F. Filippetti, C. Tassoni, and G.-A. Capolino, “Advances in diagnostic techniques for induction machines,” *IEEE Transactions on Industrial Electronics*, vol. 55, Dec. 2008.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin: Springer, 2007.

[4] T. Fawcett, “An introduction to ROC analysis”, *Pattern Recognition Letters*, vol. 27, 2006.

[5] Letters, vol. 27, pp. 861–874, 2006.A. F. D. Souza, F. Pedroni, E. Oliveira, P. M. Ciarelli, W. F. Henrique, L. Veronese, and C. Badue, “Automated multi-label text categorization with VG-RAM weightless neural networks,” *Neurocomputing*, vol. 72, pp. 2209 – 2217, 2009.

[6] Mineichi Kudo and Jack Sklansky,"Comparison of algorithms that select features for pattern classifiers", Pattern Recognition, vol. 33, pp. 25-41, 2000

[7] E. D. Wandekoken, F. T. A. Franzosi, T. W. Rauber, F. M. Varejão, and R. J. Batista, “Data-driven fault diagnosis of oil rig motor pumps applying automatic definition and selection of features,” in *Proceedings of the 7th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, Carg`ese, France, August 2009.

[8] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.

[9] B. S. Yang, T. Han, and J. L. An, “Art-kohonen neural network for fault diagnosis of rotating machinery,” *Mechanical Systems and Signal Processing*, vol. 18, pp. 645–657, 2004.

[10] J. Zhang, “Improved on-line process fault diagnosis through information fusion in multiple neural networks,” *Computers and Chemical Engineering*, vol. 30, pp. 558–571, 2006.

[11] J. bo Yu and L. feng Xi, “A neural-network ensemble-based model for on-line monitoring and diagnosis of out-of-control signals in multivariate manufacturing processes,” *Expert Systems with Applications*, vol. 36, pp. 909–921, 2009.

[12] Q. Hu, Z. He, Z. Zhang, and Y. Zi, “Fault diagnosis of rotating machinery based on improved wavelet package transform and svms ensemble,” *Mechanical Systems and Signal Processing*, vol. 21, pp. 688– 705, 2007.

[13] C. Cortes and V. Vapnik, “Support-vector network,” *Machine Learning*, vol. 20, pp. 273–297, 1995.

[14] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24(2), pp. 123–140, 1996.

[15] Y. Freund and R. Schapire, “Experiments with a new boosting algorithm,” in *Proceedings of the 13th international conference on machinelearning.*, Bari, Italy, 1996.

[16] Opitz, David W., "Feature selection for ensembles, AAAI '99/IAAI '99: Proceedings of the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence, pp. 379-384, Florida: American Association for Artificial Intelligence, 1999

[17] A. Tsymbal, P. Cunningham, M. Pechenizkiy, S. Puuronen, "Search Strategies for Ensemble Feature Selection in Medical Diagnostics," Computer-Based Medical Systems, IEEE Symposium on, pp. 124, 16th IEEE Symposium on Computer-Based Medical Systems (CBMS'03), 2003.