An Overproduce-and-Choose Strategy to Create Classifier Ensembles with Tuned SVM Parameters Applied to Real-World Fault Diagnosis

Estefhan Dazzi Wandekokem, Flávio M. Varejão, and Thomas W. Rauber

Department of Computer Science, Federal University of Espírito Santo. Vitória, ES, Brazil

Abstract. We present a supervised learning classification method for model-free fault detection and diagnosis, aiming to improve the maintenance quality of motor pumps installed on oil rigs. We investigate our generic fault diagnosis method on 2000 examples of real-world vibrational signals obtained from operational faulty industrial machines. The diagnostic system detects each considered fault in an input pattern using an ensemble of classifiers, which is composed of accurate classifiers that differ on their predictions as much as possible. The ensemble is built by first using complementary feature selection techniques to produce a set of candidate classifiers, and finally selecting an optimized subset of them to compose the ensemble. We propose a novel ensemble creation method based on feature selection. We work with Support Vector Machine (SVM) classifiers. As the performance of a SVM strictly depends on its hyperparameters, we also study whether and how varying the SVM hyperparameters might increase the ensemble accuracy. Our experiments show the usefulness of appropriately tuning the SVM hyperparameters in order to increase the ensemble diversity and accuracy.

Keywords: Fault diagnosis, feature selection, feature extraction, classifier ensemble, Support Vector Machine, multi-label classification.

1 Introduction

The detection and diagnosis of faults in industrial machines is advantageous for economical and security reasons. The objective is to repair damaged components during planned maintenance, which minimizes machinery standstill besides providing more secure operations.

Two principal approaches to create a fault predictor exist: model-based techniques and model-free techniques. The model-based line of research relies on an analytical model of the studied process, involving time dependent differential equations [1]. However in real-world processes the availability of an analytical model is often unrealistic or inaccurate due to the complexity of the process. In this case model-free techniques are an alternative method [2].

We present a model-free method based on the *supervised learning* [3] classification paradigm as the primal mechanism to automatically generate the fault classifiers. This presents as advantage the requirement of a minimum of a priori knowledge about the plant, as the fault predictor is automatically defined based on training data. We work with 2000 examples of vibrational signals obtained from operational faulty motor pumps, acquired from 25 oil platforms off the Brazilian coast during five years. Human experts provided a label for every fault present in each acquired example.

We focus on the horizontal motor pump with extended coupling between the electric motor and the pump. Accelerometers are placed at strategic positions along the main directions to capture specific vibrations of the main shaft. To extract features, we apply well known signal processing techniques like Fourier transform, envelope analysis based on the Hilbert transform [9] and median filtering. So the features correspond to the vibrational energy in a predetermined frequency band. The cardinality of an extracted feature vector G is 95. Several faults can simultaneously occur in a motor pump, which increases the diagnosis complexity. We build a predictor for detecting six fault categories in an input pattern: rolling element bearing failures; pump blade unbalance; hydrodynamic fault; shaft misalignment; mechanical looseness; and structural looseness.

The novelty of this paper is two-fold. First, we present a generic model-free diagnosis procedure for diagnosing faults using features extracted from the machine signals. Each fault is predicted by a distinct ensemble [7] of support vector machine (SVM) [4] classifiers. We propose a novel ensemble creation method well suited to fault diagnosis as suggested by our experiments. Second, we work with data from real-world operating industrial machines instead of using data from a controlled laboratory environment which is almost always found in the literature. That is highly desirable, as laboratory hardware cannot realistically represent intricate real-world fault occurrences.

The remainder of this paper is organized as follows. Our model-free approach to fault diagnosis based on feature extraction, feature selection and ensemble classification is explained in section 2. Section 3 outlines the proposed ensemble creation method for dealing with real-world fault diagnosis. In section 4 we show the experimental results achieved by the studied faults predictors using the acquired database. Finally, section 5 draws conclusions and points out to future research.

2 Model-Free Approach to Fault Diagnosis

We formulate the fault diagnosis problem as a multi-label classification task in which several labels (fault classes) may be simultaneously assigned to an example. Each fault category is represented by a binary predictor, diagnosing the presence or absence of that individual fault in an input pattern. Therefore the problem at hand is the one of creating accurate binary predictors.

We work with the *support vector machine* [4] (SVM) classifier which is currently considered one of the most powerful machine learning approaches for solving binary classification problems. We use a widely adopted SVM model, namely

a Radial Basis Function kernel and the C-SVM classification architecture [3]. So we work with two hyperparameters, namely the regularization parameter C which controls the model complexity and the kernel parameter γ which controls the nonlinear mapping of the features. As the performance of a SVM strictly depends on its hyperparameters, we use an effective method to tune them, namely grid search combined with cross-validation on each candidate parameter vector.

To increase the accuracy achieved by an individual classifier, the research in *classifier ensembles* [7] indicates that a better generalization power is achieved when the class of an input pattern is predicted by a set of very accurate classifiers that collectively disagree on their predictions as much as possible. The traditional approach to create an ensemble is to vary the training data set used by the classifiers (see [5] for a reference on SVM ensembles). However such an approach usually is not well suited to SVMs, as a small variation on the training data set tends to cause a small variation on the SVM decision function. On the other hand, varying the SVM parameters does decisively change the SVM decision function [10]. This is useful for ensemble creation because the divergence among the SVMs in an ensemble increases, and so does its accuracy [8].

Another useful approach for ensemble creation is to vary the feature set of the classifiers (see [11] for a reference on fault diagnosis). But at the present time the role of SVM parameters in feature-based ensembles of SVM classifiers has not been investigated yet. In this work we propose a novel ensemble creation method based on feature selection, and study whether and how varying also the SVM hyperparameters might increase the ensemble accuracy.

3 Feature-Based Classifier Ensembles

In this work we propose a novel method for creating an ensemble. This method presents as advantages simplicity and a relatively low computational cost. Besides, it is well suited to fault diagnosis as it is based on feature selection, an approach that allows the crucial features to be detected and prioritized.

3.1 Feature Selection

Feature selection [6] is the process of choosing an optimized subset of features for classification from a larger set. In this work we employ the wrapper approach in which the learning algorithm itself is used to access the saliency of the candidate feature sets. So the selection criterion J used to estimate the performance of a candidate feature set X_k is the Area Under the ROC Curve (AUC) achieved by a classifier which uses X_k (estimated by cross-validation).

As an exhaustive search is not feasible in general we work with suboptimal search strategies. We work with two complementary hill-climbing searches. The *Sequential Forward Selection* (SFS) search method starts with an empty set of currently selected features, and at each step one feature is definitely included in it. Consider that k features have already been selected and included in the feature set X_k . To include one more feature in X_k , each non-selected feature

ξ_j must be tested individually together with the already selected features and ranked according to the criterion J , so that the feature ξ_h which provided the highest criterion J is selected and included in X_k . The *Sequential Backward Selection* (SBS) search method operates in a similar way as SFS, but SFS includes features, while SBS removes features. SBS starts with every feature already included within the set of selected features X_k . At each step, one feature is definitely removed from the set, namely the one that provided the highest criterion J with its individual exclusion from X_k .

3.2 Best Selected Feature Subsets Ensemble Creation

The proposed ensemble creation method, which we call *Best Selected Feature Subsets* (BSFS), is based on a two-stage *overproduce-and-choose* [7] strategy. It operates by initially using complementary hill-climbing feature selection methods to build a large set $\mathcal{L}$ of classifiers that are candidates to constitute the ensemble, and further using a hill-climbing search to select just a reduced, optimized subset of classifiers $\mathcal{E}$ from $\mathcal{L}$.

To combine the individual predictions of the classifiers in an ensemble into a single decision we use an effective, simple method, namely averaging the classification scores given to an input pattern by the classifiers in that ensemble.

The classifier overproduction stage. To create the set $\mathcal{L}$ of candidate classifiers to compose the ensemble, we first build a set of feature sets Ξ composed of several promising feature sets. Each of these feature sets uses features from the global pool G of available features. We perform m distinct sequential feature selection searches, $\{\mathcal{S}_1, \dots, \mathcal{S}_i, \dots, \mathcal{S}_m\}$, so that the feature sets in Ξ are determined by taking each produced feature set $X_k^{\mathcal{S}_i}$ which uses a number k of features selected by a search $\mathcal{S}_i$, for every combination of k and i with $k = 1, 2, \dots, |G|$ and $i = 1, 2, \dots, m$. To obtain the feature sets, if the search $\mathcal{S}_i$ operates in a forward way (for instance SFS), we require $\mathcal{S}_i$ to select a total of $|G| - 1$ features. On the other hand if the search $\mathcal{S}_i$ operates in a backward way (for instance SBS), we require $\mathcal{S}_i$ to select a total of 1 feature.

Then the set of candidate classifiers $\mathcal{L}$ is defined by building, for each feature set $X_k^{\mathcal{S}_i}$ in the set Ξ , a classifier c_j that uses this feature set, and we also automatically tune the hyperparameters of this classifier c_j aiming to increase its accuracy. So $\mathcal{L}$ is composed of every produced c_j .

The classifier selection stage. After building the set of candidate classifiers $\mathcal{L}$, we use a method to select an optimized set of n_c classifiers to compose the final ensemble $\mathcal{E}$, selecting from $\mathcal{L}$. For performing this ensemble classifier selection $\text{ECS}(\mathcal{L}, n_c)$ we employ the Sequential Forward Selection (SFS) search. We define the criterion J of a particular candidate subset of classifiers to compose the ensemble (a subset of $\mathcal{L}$) as the AUC on training data achieved by this candidate ensemble. The score that a candidate ensemble gives to a training pattern $\mathbf{x}$ can be obtained by averaging the scores given to $\mathbf{x}$ by the classifiers in that ensemble (the scores of the training data are estimated by cross-validation).

As we use the SFS search to select the classifiers in the ensemble, the first selected classifier c_l to compose the ensemble is the one with the highest individual cross-validation AUC. Following, each next selected classifier is the non-selected one which enables the highest criterion J with its individual inclusion in the current ensemble. When n_c classifiers are selected, the inclusion process stops, so the ensemble $\mathcal{E}$ is finally built.

4 Experimental Results

To access the effectiveness of the studied classification approaches, we performed a stratified 5×2 cross-validation [7]. So in the experiments we performed five replications of a 2-fold cross-validation. In each replication, the complete database of 2000 examples was randomly partitioned, in a stratified manner, into two sets each one with approximately 1000 examples (the stratification process preserves the distribution of the six fault categories between both sets). Then in each replication each considered classification model for creating the predictor of a fault was trained on a set and tested on the remaining one; so after the five replications we averaged the ten distinct test accuracies it achieved.

4.1 Studied Classification Approaches

For each of the six considered faults, we studied four different classification models for creating the predictor of that fault. Our objective is to evaluate the efficiency of feature-based SVM ensembles besides studying an approach for varying the hyperparameters of the SVM classifiers in an ensemble.

For the studied classification models based on the proposed BSFS ensemble creation method, to build the set $\mathcal{L}$ of candidate classifiers we ran four distinct feature selection experiments $\{\mathcal{S}_1, \dots, \mathcal{S}_4\}$, which were: the SFS search using the SVM hyperparameters values ($C = 8.0, \gamma = 0.5$) for building SVMs to estimate the selection criterion (that was the AUC estimated by 10-fold cross-validation); SFS using ($C = 2.0, \gamma = 8.0$); SBS using ($C = 8.0, \gamma = 0.5$); and SBS using ($C = 2.0, \gamma = 8.0$). Then for each produced feature subset $X_k^{\mathcal{S}_i}$ (obtained by using, for each $\mathcal{S}_i$, each number of selected features from $k = 1$ to $k = 95$) we built a SVM classifier to compose $\mathcal{L}$. In order to select a subset $\mathcal{E}$ of classifier from $\mathcal{L}$ to compose the ensemble, we set the desired ensemble size as $n_c = 10$ as we observed a tendency of an AUC decrease using a larger set.

The SVM classification model. We studied the effectiveness of a single SVM which used as feature set the complete global pool of features G . We used the grid-search parameter optimization method to tune its hyperparameters. We refer to this classification model as SVM.

The BSFS- n classification model. In this experiment we aim to get an insight into the effectiveness of the proposed ensemble creation method to generate a feature-based ensemble, without varying the SVM hyperparameters of the classifiers. So after performing the four feature selection experiments, we built

every SVM in the set of candidate classifiers $\mathcal{L}$ using the hyperparameters values $C = 8.0$, $\gamma = 0.5$ (which usually provided accurate SVMs). So the selected SVMs in the final ensemble $\mathcal{E}$ also used $C = 8.0$, $\gamma = 0.5$. We refer to this classification model as **BSFS-n**.

The BSFS-t classification model. We studied the usefulness of directly tuning the SVM hyperparameters of the classifiers in an SVM ensemble, aiming to increase the individual accuracy of each of those SVMs. So we used the grid search method to tune the hyperparameters of each SVM in an ensemble initially defined by **BSFS-n**. We refer to this classification model as **BSFS-t**.

The BSFS-o classification model. Finally we evaluated an ensemble in which we used the grid search method to tune the hyperparameters of every SVM in the set of candidate classifiers $\mathcal{L}$. It is expected that the subsequent selection process actively searches for a set of diverse SVMs which differ on their feature set and also on their SVM hyperparameters, therefore composing a more diverse and accurate ensemble. We refer to this classification model as **BSFS-o**.

4.2 Misalignment Predictor Overproduced SVMs

We present some graphs related to the construction of the misalignment predictor, for the first pair of train-test data of the 5×2 cross-validation process. Figures 1 and 2 present, respectively to the **BSFS-o** and **BSFS-n** classification models, the test data AUC achieved by each produced SVM in the set $\mathcal{L}$. It can be seen that the AUC value individually achieved by the SVMs presented much more variation among them for the **BSFS-o** model (in figure 1) than for the **BSFS-n** model (in figure 2), as even for similar feature sets (with a similar

Fig. 1. AUC on test data achieved by the SVMs in the set $\mathcal{L}$ of candidate classifiers, for the **BSFS-o** classification model. For comparison we also show as a horizontal line the test data AUC achieved by the final ensemble.

Fig. 2. AUC on test data achieved by the SVMs in the set $\mathcal{L}$ of candidate classifiers, for the BSFS- n classification model

number of features) the AUC value considerably changed. This suggests that BSFS- o is able to produce more diverse SVMs than BSFS- n due to the different hyperparameters values that are used.

4.3 5×2 Cross-Validation Estimation Results

Table 1 presents, for each considered fault, the percentage of negative class (non-faulty) data, and also the accuracy and AUC estimated on test data by the

Table 1. Class distribution and 5x2 test data accuracy, AUC

Considered fault classifier	Percentage of negative class data	The SVM model	The BSFS- n model	The BSFS- t model	The BSFS- o model
Misalignment	57.4%	75.9% 0.829	77.8% 0.851	77.5% 0.855	79.3% 0.872
Bearing	64.3%	83.4% 0.909	86.5% 0.932	86.7% 0.933	88.0% 0.945
Unbalance	75.1%	81.5% 0.836	82.4% 0.861	82.6% 0.867	83.1% 0.882
Hydrodynamic	57.6%	84.0% 0.912	85.9% 0.928	85.8% 0.929	86.5% 0.936
Structural looseness	78.8%	86.7% 0.873	86.9% 0.877	87.2% 0.881	87.8% 0.919
Mechanical looseness	89.0%	92.9% 0.878	93.0% 0.880	93.1% 0.889	93.5% 0.901

5×2 cross-validation estimation process, individually achieved by each considered classification model for creating the predictor of that fault.

5 Conclusions and Future Work

In this paper we presented a data-driven supervised learning classification method for performing real-world fault diagnosis. We proposed a novel approach for creating ensembles of SVM classifiers based on the variation of the feature sets and also the SVM hyperparameters.

We focused on the role of SVM parameters in feature-based SVM ensembles. Our experiments show that tuning the hyperparameters of each candidate SVM to compose the ensemble, aiming to increase its individual accuracy, enables the creation of a more accurate ensemble, as an optimized subset of SVMs might be selected to finally constitute the ensemble. In this case the ensemble is composed of accurate SVMs which are as divergent as possible due different feature sets and also SVM hyperparameters. In opposition to that, our experiments show that the ensemble accuracy gain is lower if that process of hyperparameters tuning is only performed in an already defined ensemble. This suggests that, to create a more accurate ensemble, the divergence among the produced classifiers should be taken into account during the process of varying the SVM hyperparameters.

We intend to acquire more real-world data, from different machines and also from more sources than just vibrational signals, which increases the classification accuracy as the features are extracted from complementary information sources. Thus we plan to develop a multiparametric diagnostic system, which uses vibration signals complemented with electrical signals such as current and power.

Acknowledgments

This work was supported by COPES-Petrobrás.

References

1. Angelo, C.H.D., Bossio, G.R., Giaccone, S.J., Valla, M.I., Solsona, J.A., Garcia, G.O.: Online model-based stator-fault detection and identification in induction motors. *IEEE Transactions on Industrial Electronics* 56 (November 2009)
2. Bellini, A., Filippetti, F., Tassoni, C., Capolino, G.A.: Advances in diagnostic techniques for induction machines. *IEEE Transactions on Industrial Electronics* 55 (December 2008)
3. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, Berlin (2007)
4. Cortes, C., Vapnik, V.: Support-vector network. *Machine Learning* 20, 273–297 (1995)
5. Evgeniou, T., Pontil, M., Elisseeff, A.: Leave-one-out error, stability, and generalization of voting combinations of classifiers. *Machine Learning* (2002)
6. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182 (2003)

7. Kuncheva, L.I.: *Combining Pattern Classifiers: Methods and Algorithms*. Springer, Heidelberg (2004)
8. Li, X., Wang, L., Sung, E.: Adaboost with svm-based component classifiers. *Engineering Applications of Artificial Intelligence* 21 (2008)
9. Mendel, E., Mariano, L.Z., Drago, I., Loureiro, S., Rauber, T.W., Varejao, F.M.: Automatic bearing fault pattern recognition using vibration signal analysis. In: *Proc. of the IEEE International Symposium on Industrial Electronics (ISIE 2008)* (2008)
10. Valentini, G., Dietterich, T.G.: Bias-variance analysis of support vector machines for the development of svm-based ensemble methods. *The Journal of Machine Learning Research* 5 (January 2004)
11. Zio, E., Baraldi, P., Gola, G.: Feature-based classifier ensembles for diagnosing multiple faults in rotating machinery. *Applied Soft Computing* 8, 1365–1380 (2008)