A Study of Feature Selection Algorithms for the Detection of Gear Damage in Vibration Data

Andreas Beering, Jakob Döring, Karl-Ludwig Krieger

Chair for Automotive Electronic Systems (ITEM.ae), 28359 Bremen, Germany

email: {beering, doering, krieger}@item.uni-bremen.de

Abstract

Nowadays, classification algorithms are used in a variety of areas, for example in the field of condition monitoring to detect machine damage. In this context, features are often extracted from measurement data in order to use them for classification. Due to increasing amounts of data and large feature sets, the resulting computational cost is tremendous and usually not processable by low-cost embedded hardware. To reduce this, algorithms for feature selection can be used. These algorithms reduce the number of features and thus the computing effort significantly. In addition, due to the curse of dimensionality, it is possible to achieve better classification performance by reducing the number of features. In this paper, a study of feature selection algorithms for vibration data of damaged gears is carried out. For this purpose, labelled data for the feature selection is produced in a laboratory setup using a piezoelectric vibration transducer. Based on the data set, sequential forward selection (SFS), sequential forward floating selection (SFFS), sequential backward selection (SBS) and sequential backward floating selection (SBFS) are compared. The studied feature selection algorithms reduce the number of features with at least the same classifier performance.

Introduction

During harvest time, agricultural machines require a very high level of availability. Due to the harsh environmental conditions, the gearboxes of these machines are exposed to a numerous number of loads. In particular, dust and soil, picked up by the agricultural machine, can lead to various types of damage and in the long term even to total failure of the moving and rotating machine parts. Despite regular maintenance of the machines, heavy-duty machine components, such as gearboxes, can fail due to such damage during harvesting. In this context, classification approaches can be used to detect damage at an early stage. For local detection on the vehicle usually low-cost embedded hardware is used, which often reach its limits due to the huge amount of data and feature sets. Feature selection algorithms can be used to create a subset of features in this area in order to obtain reliable classification results with lower computational cost.

The remainder of this paper is organized as follows. In the second section, we outline several feature selection algorithms that were compared within the investigations. Afterwards the used data set is described, as well as its acquisition and the split into training and test data. Subsequently, the feature selection algorithms are compared using different criterion functions and classifiers and the results are presented. Finally, a conclusion is drawn.

Feature Selection Algorithms

Feature selection (FS) algorithms are divided into three categories - filter, wrapper or embedded methods - depending on their interaction with learning models. Filter methods are based on a statistical measure between the features and are therefore independent from learning models. The mutual information, relief or pearson's correlation are commonly used filter approaches. In contrast, the performance of wrapper methods depends directly on a classifier. Due to repeated learning and validation steps during the feature selection, the computation cost is higher. Commonly used approaches are for example sequential forward selection (SFS) or sequential backward selection (SBS). Embedded methods are a special form of wrapper methods, in which the feature search and the classification are not separated. This results in a lower computational effort. Examples for embedded feature selection are LASSO or Ridge regression. In this paper the feature selection is used to optimize two existing classifications via k-nearest neighbors (KNN) and feedforward neural network (NN). Since for these classifiers especially wrapper methods are useful, we will discuss sequential forward and backward selection (SFS and SBS) in more detail in the following. In the sequential approaches, we start with an empty or the complete feature set and add or eliminate features iteratively. The disadvantage of this method is that once eliminated/added features can not be added/eliminated again. This so-called nesting effect can reduce the performance [1]. To avoid this nesting effect, extensions of the SFS and SBS (the so called sequential forward floating selection (SFFS) and sequential backward floating selection (SBFS)) are used [2]. These offer the advantage that features already eliminated/added can be added/eliminated again during the floating inclusion/exclusion. The sequence of the SFFS is shown in Algorithm 1 as pseudo code. Since the SFFS is an extension of the SFS, the pseudo code in the first two steps shows the actual SFS.

In the first step of the SFFS, the feature subset $\mathcal{X}$ is initialized as an empty set $\varnothing$ and the number of features k as 0. For all feature selection algorithms considered, the subset $\mathcal{X}_k$ is a true subset of the entire feature set $\mathcal{Y}$ for the number of features k. In the second step of the feature selection, the feature for the inclusion x^+

Algorithm 1 Sequentiel forward floating selection (SFFS) pseudo code.

1. $\mathcal{X}_0 = \varnothing, k = 0$
2. Select the best feature (Inclusion):
   $x^+ = \underset{x \in \mathcal{Y} \setminus \mathcal{X}_\mathrm{k}}{\arg\max}\, J(\mathcal{X}_k \cup \{x\})$
   if $J(\mathcal{X}_k) < J(\mathcal{X}_k \cup \{x^+\})$ **then**
   $\quad\mathcal{X}_{k+1} = \mathcal{X}_k \cup \{x^+\}$
   $\quad k = k + 1$
   else
   $\quad$End feature selection
   end if
3. Floating exclusion:
   $x^- = \underset{x \in \mathcal{X}_\mathrm{k}}{\arg\max}\, J(\mathcal{X}_k \setminus \{x\})$
   while $J(\mathcal{X}_k) < J(\mathcal{X}_k \setminus \{x^-\})$ **do**
   $\quad x^- = \underset{x \in \mathcal{X}_\mathrm{k}}{\arg\max}\, J(\mathcal{X}_k \setminus \{x\})$
   $\quad\mathcal{X}_{k-1} = \mathcal{X}_k \setminus \{x^-\}$
   $\quad k = k - 1$
   end while
   go to 2.

Algorithm 2 Sequentiel backward floating selection (SBFS) pseudo code.

1. $\mathcal{X}_0 = \mathcal{Y} = \{x_1, x_2, ..., x_d\}, k = d$
2. Select worst feature (Exclusion):
   $x^- = \underset{x \in \mathcal{X}_\mathrm{k}}{\arg\max}\, J(\mathcal{X}_k \setminus \{x\})$
   if $J(\mathcal{X}_k) < J(\mathcal{X}_k \setminus \{x^-\})$ **then**
   $\quad\mathcal{X}_{k-1} = \mathcal{X}_k \setminus \{x^-\}$
   $\quad k = k - 1$
   else
   $\quad$End feature selection
   end if
3. Floating inclusion:
   $x^+ = \underset{x \in \mathcal{X}_\mathrm{k}}{\arg\max}\, J(\mathcal{X}_k \cup \{x\})$
   while $J(\mathcal{X}_k) < J(\mathcal{X}_k \cup \{x^+\})$ **do**
   $\quad x^+ = \underset{x \in \mathcal{Y} \setminus \mathcal{X}_\mathrm{k}}{\arg\max}\, J(\mathcal{X}_k \cup \{x\})$
   $\quad\mathcal{X}_{k+1} = \mathcal{X}_k \cup \{x^+\}$
   $\quad k = k + 1$
   end while
   go to 2.

from the set $\mathcal{Y} \setminus \mathcal{X}_\mathrm{k}$ is searched for, which in combination with the current subset $\mathcal{X}_k$ gives the best performance. For this a criterion function $J(\cdot)$ is used. On the one hand, the presented approach examined the classification accuracy as a criterion function $J(\cdot)$. On the other hand, the area under the receiver operating characteristic curve (ROC), abbreviated AUC, was examined. Both are known performance criteria for classifications in the literature [3]. Following the new criterion value $J(\mathcal{X}_\mathrm{k} \cup \{x^+\})$ is compared to the old one. If no performance increase is achieved, the feature selection ends. If there is an increase in performance, the floating exclusion is carried out. During this step previously excluded features can be included again if they result in an increase of the criterion. If no performance increase can be achieved during this exclusion, the SFFS continues in including features.

The analogue sequence of the SBFS is shown in Algorithm 2. Here again, the first two steps lead to the actual SBS. Compared to the SFFS, the SBFS assumes in the first step the entire feature set $\mathcal{X}_0 = \mathcal{Y}$ with $k = d$ features. Afterwards, in the second step, the algorithm searches for the feature x^-, which increases the performance the most when excluded. Here, the search is also terminated, if no improvement is achieved. During the floating inclusion in the third step, features, which have already been eliminated can be included again if they increase the criterion value $J(\cdot)$.

Data Sets

The data set used for the investigation of the feature selection algorithms consists of vibration signals from measured gearboxes. In total, one intact gearbox, one with synthetic tooth damage on the first shaft and one with synthetic tooth damage on the second shaft were examined. Figure 1 shows a schematic structure of the used gearbox type with the positions of the synthetic damages.

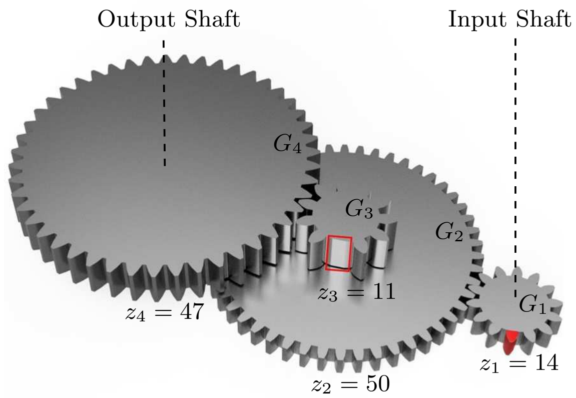

Figure 1: Schematic structure of the two-stage gearbox with synthetic tooth damage on gear G_1 and G_3.

The vibration signal was sampled via a piezoelectric vibration transducer with a resolution of 24 bit and a sampling rate of 51.2 kHz.

Figure 2 shows an example of the characteristic signal course of the vibration signal in case of a broken tooth. For this purpose, 50 recorded measurement series were aligned by correcting the temporal offset via cross-correlation. The individual signals are shown in grey and the mean value in red. The influence of the tooth damage on the vibration signal is largely determined by the speed and torque. In total, the training data set consists of measurements on a laboratory set-up with different speeds at the input shaft from 100 to 500 min^{-1} and torques up to 250 Nm. For the test data set, in contrast, measurements from the real machine, in which the gearboxes were integrated, are used. The entire data set used, as well as its split, is shown in Table 1. For the feature selection, the two measurements of gear 2 and 3 are combined, resulting in a balanced data set of two classes with 494 elements each. The same grouping is selected for the test data set, resulting in a balanced data

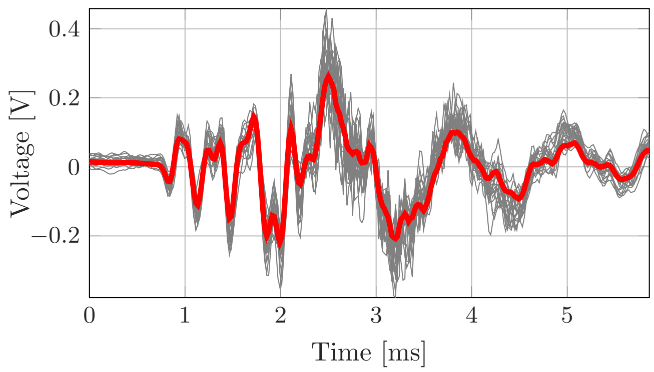

Figure 2: Characteristic vibration signal during tooth engagement with a tooth fracture.

Table 1: Listing of the dataset used for the training and testing.

	Description	$\overline{N}_{\text{train}}$	$\overline{N}_{\text{test}}$
Gear 1	Intact	494	129
Gear 2	Damage wheel G_1	247	123
Gear 3	Damage wheel G_3	247	6

set of 129 elements each. Each element has a size of 2195 samples.

Experimental Study

Each feature selection algorithm was examined 100 times with a randomly varying data set of half the training and test data set. Therefore, for each iteration a training data set of 494 elements and a test data set of 129 elements was considered. First, the feature selection for the k-nearest neighbor (KNN) classification and the AUC criterion is considered. The histogram of the selected features during the feature selection is shown in Figure 3.

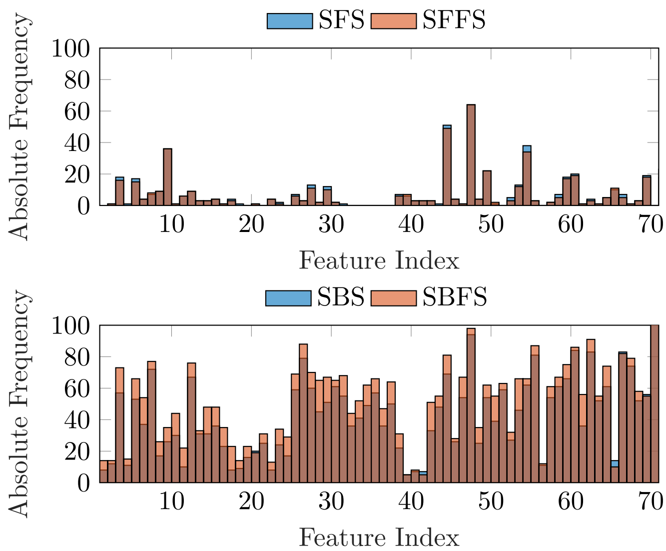

Figure 3: Histogram of the chosen features for KNN classification and AUC as criterion for 100 iterations.

The figure shows which features were selected most frequently for the respective selection algorithms. It can be seen that the number of features for all methods can be reduced. Table 2 summarizes the number of features, the AUC and the accuracy for the KNN classification with AUC as selection criterion.

Table 2: Comparison of FS algorithms for KNN classification with AUC as criterion.

FS Type	$\overline{N}_{\text{Feats}}$	AUC	Accuracy [%]
none	72.0	0.951	92.92
SFS	5.2	0.998	97.48
SFFS	4.9	0.998	97.53
SBS	30.3	0.985	95.29
SBFS	36.2	0.983	95.19

It can be seen that for all selection algorithms the initial AUC of 0.951 can be improved and the number of features can be reduced from the initial 72. The forward methods achieve better results, since the remaining number of features is considerably smaller and the AUC is higher compared to the backward methods. Furthermore, accuracy also increases for all cases. Table 3 summarizes the number of features, the accuracy and the AUC for the KNN classification with accuracy as selection criterion.

Table 3: Comparison of FS algorithms for KNN classification with accuracy as criterion.

FS Type	$\overline{N}_{\text{Feats}}$	Accuracy [%]	AUC
none	72.0	92.92	0.951
SFS	3.5	98.39	0.988
SFFS	3.4	98.38	0.988
SBS	10.7	97.29	0.971
SBFS	12.6	97.79	0.977

With feature selection, using accuracy as a criterion, comparable results to the previous observation can be seen. The accuracy of initially 92.92 % increases and the number of features decreases for all cases. Overall, the number of features is reduced more in comparison to feature selection with AUC as criterion (especially for the backward methods). For this case the forward methods also achieve slightly better results than the backward methods. Besides KNN the feature selection for a feed-forward neural network (NN) with 10 hidden layers and 2 output layers was investigated. The histogram of the selected features during the selection with NN classification and AUC criterion is shown in Figure 4. It can be seen that the forward methods show a similar trend to the consideration with KNN. With the backward methods, however, very few features are excluded and in total, considerably more features remain. It is assumed that this is due to the functioning of neural networks, where poorly suited characteristics have less influence due to the low weighting. This weighting could lead to the fact that barely any improvement is achieved by removing a feature and thus very few features are removed. A tabular comparison of the feature selection is given in Table 4 for the selection with AUC as criterion and in Table 5 for the selection via accuracy. The tables show that for the classification using the NN a better performance (with AUC as well as accuracy) and a reduced number of features

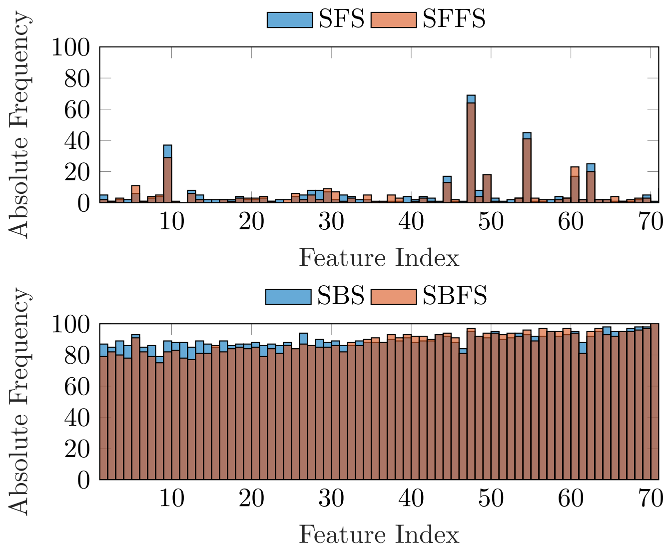

Figure 4: Histogram of the chosen features for NN classification and AUC as criterion for 100 iterations.

Table 4: Comparison of FS algorithms for NN classification with AUC as criterion.

FS Type	$\overline{N}_{\text{Feats}}$	AUC	Accuracy [%]
none	72.0	0.985	96.34
SFS	4.0	0.998	95.57
SFFS	3.6	0.999	94.94
SBS	63.8	0.999	95.32
SBFS	62.7	0.999	95.92

Table 5: Comparison of FS algorithms for NN classification with accuracy as criterion.

FS Type	$\overline{N}_{\text{Feats}}$	Accuracy [%]	AUC
none	72.0	96.34	0.986
SFS	3.4	98.74	0.984
SFFS	3.4	98.82	0.983
SBS	60.7	99.02	0.970
SBFS	59.9	99.02	0.969

can be achieved. As with the KNN, a slightly smaller number of features remain when selecting via accuracy. The forward methods again show better results regarding the number of features. Based on the calculated results it can be argued that the backward approaches for the NN can barely achieve a reduction of the feature space.

In order to achieve a further reduction of the feature subset in the backward methods, the existing approach is extended by a rounding. For this purpose, the comparison of the criteria within the backward methods is rounded to 1% according to Equation 1.

$$\lfloor 100 \cdot J(\mathcal{X}_k) \rfloor < \lfloor 100 \cdot J(\mathcal{X}_k \setminus \{x^-\}) \rfloor \qquad (1)$$

Thus, features which gain no or only slight improvement are excluded. However, this rounding can also reduce performance during an iteration of the selection algorithm. The results of the feature selection with these approximations ($\text{SBS}_{\text{round}}$ and $\text{SBFS}_{\text{round}}$) for 100 iterations compared to the normal methods (SBS and SBFS) for the KNN classification is listed in Table 6 and for NN in Table 7.

Table 6: Comparison of FS algorithms for KNN with approximation in backward methods.

	$J_{\text{AUC}}(\cdot)$		$J_{\text{Accuracy}}(\cdot)$	
FS Type	$\overline{N}_{\text{Feats}}$	AUC	$\overline{N}_{\text{Feats}}$	Accurcy [%]
SBS	30.3	0.985	10.7	97.29
$\text{SBS}_{\text{round}}$	5.9	0.996	9.5	97.26
SBFS	36.2	0.983	12.6	97.79
$\text{SBFS}_{\text{round}}$	6.6	0.996	11.4	97.71

Table 7: Comparison of FS algorithms for NN with approximation in backward methods.

	$J_{\text{AUC}}(\cdot)$		$J_{\text{Accuracy}}(\cdot)$	
FS Type	$\overline{N}_{\text{Feats}}$	AUC	$\overline{N}_{\text{Feats}}$	Accurcy [%]
SBS	63.8	0.999	60.7	99.02
$\text{SBS}_{\text{round}}$	8.0	0.996	52.8	99.08
SBFS	62.7	0.999	59.9	99.02
$\text{SBFS}_{\text{round}}$	9.4	0.998	57.6	99.12

For both classifiers a further reduction of the feature subset can be achieved by this approach. Especially when selecting via the AUC criterion, a significant reduction of features can be achieved with almost the same performance. For the selection using accuracy as a criterion, a majority of the features are still contained in the subset. However, the forward methods still result in a smaller feature subset.

Conclusion

In this paper a study of feature selection algorithms for the detection of gear damage in vibration data is presented. For this purpose, vibration data of different gearboxes with fractured gears has been recorded. Based on this data, SFS, SFFS, SBS and SBFS, as well as modified forms of the two backward methods with rounding were examined. Overall, the two forward methods reduce the feature space more effectively. For all FS algorithms it was shown that a reduction of the feature space can lead to better performance with a reduced feature set. In selecting the criterion, accuracy achieves a slightly higher reduction of the feature subset. In terms of classification, the neural network achieves a better performance than k-nearest neighbors.

References

[1] B. Venkatesh and J. Anuradha: A Review of Feature Selection and Its Methods. Cybernetics and Information Technologies 19, pp. 3–26, 2019.

[2] P. Pudil, J. Novovičová and J. Kittler: Floating search methods in feature selection. Pattern Recognition Letters 15,pp. 1119-1125, 1994.

[3] J. Huang and C. X. Ling: Using AUC and Accuracy in Evaluating Learning Algorithms. IEEE Transactions on Knowledge and Data Engineering 17, pp. 299-310, 2005.