Proceedings of the 6th European Conference of the Prognostics and Health Management Society 2021 - ISBN – 978-1-936263-34-9

Feature Based Bearing Fault Detection With Phase Current Sensor Signals Under Different Operating Conditions

Tobias Wagner[1, 2], Sara Sommer[2]

[1] *Fulda University of Applied Sciences, Fulda, Hesse, 36037, Germany*
tobias.wagner@informatik.hs-fulda.de

[2] *Bosch Rexroth AG, Lohr, Bavaria, 97816, Germany*
tobias.wagner7@boschrexroth.de
sara.sommer@boschrexroth.de

Abstract

This study addresses the problem of bearing damage detection on Permanent Magnet Synchronous Motors (PMSMs) based on the internal phase current data sources. It focusses on tackling the accuracy degradations caused by variations of the motor parameters like rotational speed, load torque and radial forces. Therefore, we propose a feature based deep unsupervised domain adaptation method to improve the classification accuracies of two operating points by use of only one label set. Instead of analyzing the raw data directly, we perform a handcrafted feature extraction based on expert-driven features, related to the physical causes of bearing faults. Since this feature extraction stage can be performed on edge devices directly, it also reduces the necessary data traffic to the analytics server. Model training and domain adaptation is afterwards performed on server sided GPU devices. We evaluate the proposed approach on a benchmark dataset for bearing fault detection on PMSMs.

1. Introduction

To avoid sudden motor failures, which might lead to unexpected downtimes of entire production plants, predictive machinery maintenance is of great importance. Since more than 40 percent of all motor failures are reasoned by bearing defects (Raison, Rostaing, Butscher & Maroni (2002)), the detection of bearing defects is of particular importance in predictive maintenance applications. In the past, mainly external vibration sensors were used for fault detection (Carvalho, Soares, Vita, Francisco, Basto & Alcalá (2019)). However, in recent years, there has been increasing research into using phase currents for bearing damage detection, which has the advantage of saving additional sensors and thus costs ((Trajin, Regnier & Faucher, 2010), (Li, Xiong, Li, Su & Wu (2019)). Phase currents are sparely used compared to vibrational signals, which is reasoned by the fact that mechanical vibrations caused by the bearing faults can be perceived with less effort. Nevertheless, there are two major disadvantages of using such signals, which are on the one hand the required external hardware effort and associated costs and on the other hand the susceptibility to interfering signals. The latter is due to the installation situation and results in noise contamination of the measured signal, caused by the production process being transmitted via the flange-mounted motor to the vibration sensor attached to the motor housing. The practical use of phase currents is justified by the fact that bearing damage is accompanied by eccentricities between the rotor and stator, which affect the magnetic flux and thus also the phase currents (Rosero, Cusido, Garcia, Ortega & Romeral, 2006). For these reasons, we focus on bearing defect detection using phase currents.

By evaluating sensor data with machine and deep learning methods, models can be trained which can be used to determine the bearing state. The implementation of machine-learning based solutions for manufacturing poses several challenges, especially when deploying the models to the production environment (Höhr, Tasci & Verl (2019), Lade, Ghosh & Srinivasan (2017)). It becomes particularly difficult when the data distribution differs in model application compared to model training, which is often the case in manufacturing. For example, variations of the rotational shaft frequency, torque loads and radial forces applied to the bearings can have a strong influence on the applicability of a defect detection model trained with phase current data (Wagner & Sommer, 2020). This means, a defect detection model, which was trained with labeled data collected under a fixed operating condition is not necessarily able to make reliable classifications if the operating conditions differ from the training conditions. This problem is known as the domain shift problem. However, it is of utmost importance for practical applicability that a fault detection model is able to make reliable classifications under a variety of previously unknown operating conditions. To the best of our knowledge, no solution has been proposed to handle this domain shift problem for phase current based bearing fault detection yet.

To solve this problem, we present a deep learning approach for diminishing the influence of different operating conditions by utilizing unsupervised domain adaptation. Our approach has the following features /advantages:

- Computational feasibility thanks to domain related handcrafted features
- Bearing, resp. machine health states must only be known for the training data coming from the test rig, while no labels are needed for the inference data
- Accuracy degradations due to motor parameter variations (like rotational shaft speed, load torque and radial forces) are reduced in an unsupervised manner

The structure of the rest of the paper is as follows: In section 2, we give an introduction to existing work that handle the domain shift problem in bearing fault detection with vibration data. The proposed method is explained in detail in section 3. In section 4, we evaluate the method with a public available benchmark dataset. The paper ends with a conclusion and gives an outlook on upcoming studies.

2. Related Work

In current research of bearing fault detection, the use of external measured vibration signals is very established. To overcome the drawbacks associated with this, namely the need and costs of external sensors and hardware equipment, this paper considers phase current sensor signals, as they are directly measureable by the motor parameters. To substantiate the opportunity of detecting bearing damage based on phase currents, this chapter first introduces the considered motor family of PMSMs as well as some fundamentals on bearing damage and their physical effects on the motor system. Subsequently, the state of research is introduced based on selected and related works.

2.1. Influence of bearing faults on phase current signals in PMSMs

Mostly four types of bearing damages are considered. These are faults on the outer ring (OR) or inner ring (IR) as well as faults on the rolling elements (RE) or cage element (CE). However, it should be taken into account that a defect bearing must always be replaced completely and is not disassembled into its individual parts, which is why detecting the fault location is not mandatory. To compensate forces on the motor due to the application side, PMSMs often contain two different types of bearings. Since forces acting in radial orientation have the strongest effect on the shaft end facing outwards, to which the load is flanged, this bearing is loose coupled. The second, further inward placed bearing is not loose coupled - due to the strong impacts of axial acting forces, these bearings are more robustly dimensioned compared to the shaft-end bearings. In their bearing fault model, Schoen, Habetler, Kamran & Bartheld (1994) introduced the impact of bearing faults on the phase currents, caused by eccentricities between the rotational centers of rotor and stator. This causes irregularities in the air gap between inner- and outer ring, which in turn influences the magnetic flux density and thus the stator current frequency. These investigations were later supported by Blodt, Granjon, Raison & Rostaing (2008) under the assumption of neglected load zone effects in the bearing and with the consideration of the fault impact as series of dirac functions. With these assumptions, the authors introduced side band frequencies around the fundamental train frequency of the motor, as indicators for bearing faults. Since the fundamental and therefore side band frequency depend on both the bearing dimensions as well as on the rotational shaft frequency of the motor, the bearing fault defect frequencies can linked to the well-known ball pass frequencies for vibrational sensors. Table 1 gives the corresponding comparison.

Table 1. Comparison of ball pass frequencies for vibration (left) and phase current (right) sensor signals

Part	Vibration	Phase Current
OR	$f_{OR} = \frac{N_B}{2} * f_R * (1 - \frac{D_B * \cos(\theta)}{D_P})$	$f_S \pm k * f_{OR}$
IR	$f_{IR} = \frac{N_B}{2} * f_R * (1 + \frac{D_B * \cos(\theta)}{D_P})$	$f_S \pm k * f_{IR}$
RE	$f_{RE} = \frac{D_P}{2D_B} * f_R * (1 - \frac{{D_B}^2 * \cos^2(\theta)}{{D_P}^2})$	$f_S \pm k * f_{RE}$

2.2. Phase current based bearing fault analysis

Data analysis can be applied on the raw data directly in time domain, as well as in the frequency domain by also two-dimensional representations. Rosero, Cusido, Garcia Espinosa, Ortega & Romeral (2007) where the Gabor Spectrogram of the STFT was used in image representation, give an example for the latter. Image classification based on Convolutional Neural Networks (CNN) is performed for automatic bearing fault detection. As data, the authors used one phase current signal source.

As addition to this single feature approach, Hoang & Kang (2019) proposed a method to bundle the signals of multiple phases by information fusion and performed classification directly on the raw time series signals by use of recurrent- and convolutional layers. Since the feature spaces, especially for image inputs, becomes very large, feature reduction can be applied to reduce the model scope and thus increase performance. In their approach, Li, Xiong, Li, Su & Wu (2019) the fundamental architecture was based on an AutoEncoder structure to extract features from the time series raw data and a trailing Extreme Learning Machine (ELM) for the classification. To reduce the features of the AutoEncoders z-Space, the authors introduced the Sparse-Neighborhood-Representation (SNP) method, which creates a linear combination of the sampling points by use of k-nearest-neighbor algorithm to create a spare feature representation.

2.3. Approaches to handle the domain shift problem

Due to the fact that PSMS do not always operate at the same parameters for rotational speed, load torque and radial forces, it is necessary to take care of the applicability of the analyzation model to variations on the motor parameters. The difference between the motor parameter constellation the model was (supervised) trained on and the parameter constellations at application, is called domain shift and is a sub discipline of bearing fault detection on the way to analyzation methods for real application scenarios. To give a visual example, the following considerations shall serve, based on data of an industrial PMSM of Bosch Rexroth AG. Motors operate at different speeds. The rotational shaft speed is directly related to the period length of the phase currents: The higher the rotational shaft speed, the shorter the period duration (see Figure 1).

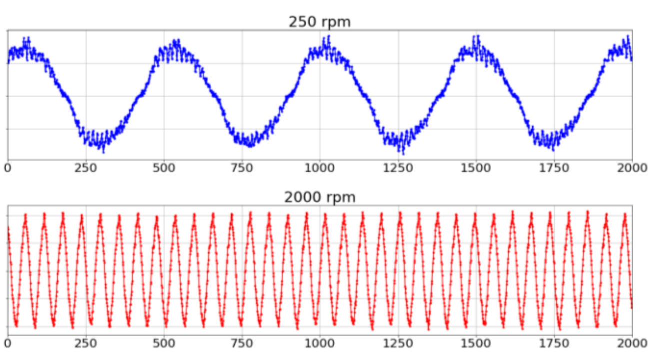

Figure 1: Phase current periods for different rotation shaft frequencies

It is obvious that models trained on, e.g. 250 rpm data will fail on data of 2000 rpm, since the input vectors will differ in a manner that basic architectures cannot generalize. Therefore, adaptations on the general model structures are necessary to adapt the models to data of previously unseen operating points.

Tong, Li & Zhang (2018) proposed a diagnosis approach using vibrational data for variable working conditions based on domain adaptation using feature transfer learning. Transferable features for training data and test data were obtained by reducing the discrepancy between two domains while strengthening the recognizable information via domain invariant clustering in an unsupervised manner.
To adapt not only to varying operating conditions instead of also varying device types, Zhou, Zheng, Wang, & Gogu (2020) proposed a model based on a CNN-Dense architecture with Softmax classification and a domain discrepancy reduction in the last dense layer by reducing the difference between features extracted from the different domains respectively operating points / device types. The authors evaluated their method on vibrational bearing fault data.

The idea of domain discrepancy reduction for the feature representations (extracted from vibrational data) of different operating points was taken up by Che, Wang, Fu & Ni (2019) and extended to several layers. For a detailed review of further research to tackle operating point shifts, we refer to the paper of Zheng, Wang & Yang (2019).

To the best of our knowledge, there does not exist any publications regarding the domain shift problem for phase current data in bearing fault detection. Hence, we propose a method for bearing fault detection on phase current data by also adapting to additional operating points in an unsupervised manner by using statistical features extracted from the time series signals of the motors phase currents.

3. PROPOSED APPROACH

The issue of bearing defect detection is considered in this paper as a multiclass classification problem. To achieve this, we have to collect phase current data from healthy motors and motors with bearing damages first. We then train a neural network as classification model with this data. What sounds relatively simple at first glance is actually much more complicated in practical application: Motors run at different speeds and are subjected to varying radial forces. The actual prevailing operating conditions are not necessarily known before a machine is put into operation and can change during its service life. Models intended to detect bearing damage on motors must therefore be able to make reliable classifications under a variety of previously unknown operating conditions. Since the operating conditions influence the phase currents, models trained with data from one operating condition will not work as well under a different operating point. This is caused by the fact that a classifier relies on the assumption that the features of the training set are drawn from the same distribution than the features at inference. Variations of the motor parameters like speed, load torque and radial forces, violate this assumption.

We propose the following approach for practical applications as depicted in Figure 2: First faulty and healthy motors are operated under one defined operating condition $\boldsymbol{A}$ on a test rig. During this process, phase current data $\mathcal{X}_A$ is collected and stored in a persistent database. The data emerging from this process are referred to as source data in the following. For the sake of simplicity, we suggest collecting data from only one operating point, although it is of course also possible to generate data from several operating points, which might lead to better defect detection accuracies.

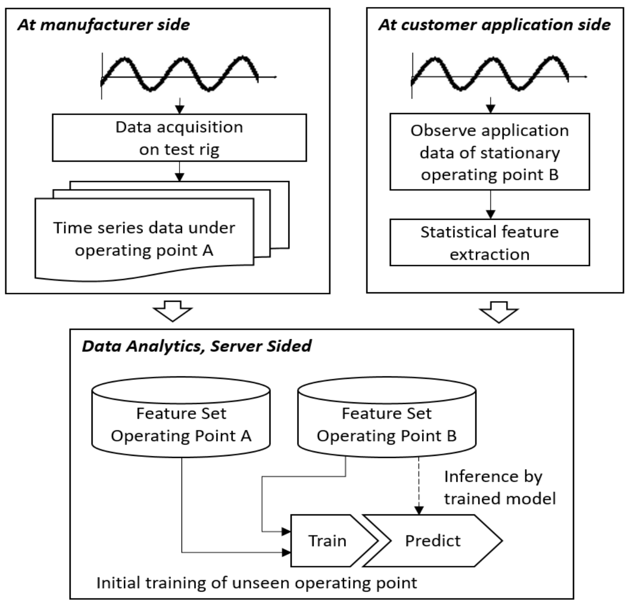

Figure 2: Proposed Data Mining workflow

When actually monitoring a motor in the field, the labeled source data $\mathcal{X}_A$ from operating point $\boldsymbol{A}$ and the unlabeled field data $\mathcal{X}_B$ generated under operating point $\boldsymbol{B}$ ($\boldsymbol{\mathcal{P}}(\boldsymbol{x_A}) \neq \boldsymbol{\mathcal{P}}(\boldsymbol{x_B})$) are then both used to train a defect detection model. This means that a separate model is trained for each motor to be monitored. This type of learning belongs to the field of unsupervised domain adaptation: To learn a model with partly labeled data from two different domains, the network architecture is built in such a way that the data from the different domains are handled separately. Due to the lower parameter- and training effort, we chose a multi-layer perceptron (MLP) architecture instead of convolutional or recurrent layers and adapted the MLP for domain adaptation functionalities by sharing some of the layers. The architecture as well as the layer hyperparameters were carried out by previous neural architecture searches (NAS) and are not described in further detail. Figure 3 shows the architecture of the proposed network with its two different processing pipelines: The upper branch processes the labeled source data (i.e. the test rig data); the lower branch processes the unlabeled target data (i.e. the data from the motor to be monitored). The only difference between the two branches is that the source-branch has a softmax function as output. The two branches are linked by sharing the same weights (dotted arrows in Fig. 3). Due to the shared weights concept, updates during backpropagation always affect the feature extraction in both branches simultaneously. As a side effect, by sharing the weights for the three dense layers, the number of parameters to be trained is halved.

In contrast to model training with only the classification metric, domain adaptation involves bringing together the feature distributions of two branches. Coupling the two branches through common weights is only the first step. The intuition is now the following: By having two layers with equal weights, the distributions of the features at the layer output can be compared to give a statement about the remaining discrepancy between the feature spaces of both branches respectively inputs (in our case the data of two different operating points). To evaluate the remaining discrepancy, a metric, which will be introduced in subchapter 3.3, is necessary.

3.1. Preprocessing of raw sensor data: feature extraction

In order to increase the customer's acceptance of the automated bearing fault solution in terms of data security and process data protection, the feature extraction of the field data directly takes place edge sided. This also reduces the data traffic to the analytics server. Only the anonymized statistical feature set is send to the server. For each new operating point, the initial model training, using both feature sets of test rig and field, is performed. From our previously studies in Wagner & Sommer (2020) on bearing fault detection on raw phase current data as input source, we found that using the time series data directly, causes a non-negligible extent of network parameters. As argued by Hoang & Kang (2019) this is due to the fact, that the phase current data needs deeper feature extraction and longer training time to converge compared to vibrational raw data. To counteract that we propose to extract features from the phase current signals, both from the time- and frequency domain. In total, we extracted 1187 features from the time series. Christ, Braun, Neuffer & Kempa-Liehr (2018) developed a python package for efficient feature extraction on industrial time series data, called tsfresh, from which we used the so called EfficientFCParameters featureset for feature extraction. This feature set contains common time- and frequency features without compelling reference to the bearing damage issue. To address the bearing damage domain, we additionally extracted the features suggested by Lessmeier, Kimotho, Zimmer & Sextro (2016) for bearing fault detection. These features are namely: Spectral energy, energy of Power Spectral Density (PSD), energy of the wavelet coefficients 1 to 3, fft peak value, skewness, shapefactor, clearance, rms, kurtosis, crestfactor and shannon entropy. Feature extraction is carried out for all available phase current sources.

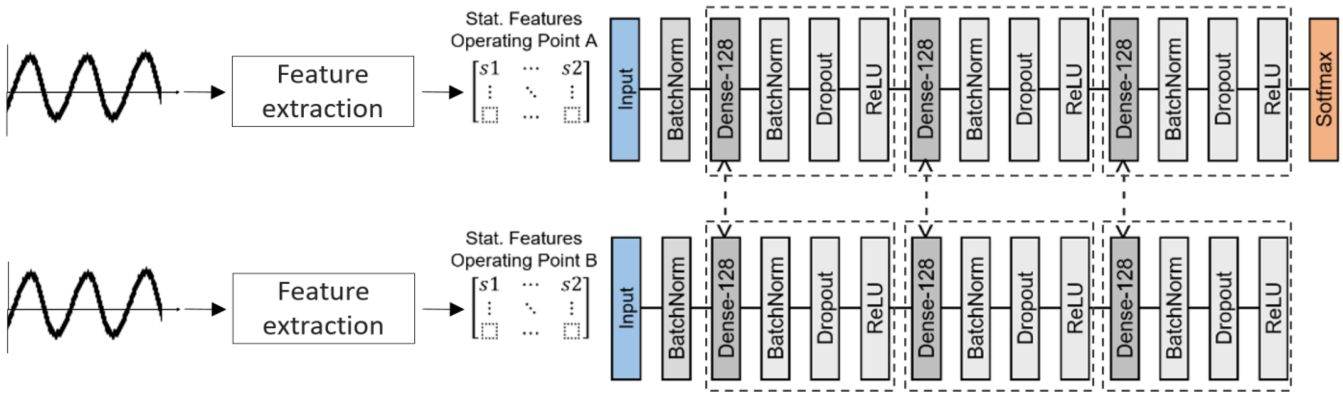

Figure 3: Proposed Domain Adaptation MLP

3.2. Model architecture

Since the input data is bound to take feature vectors from domain driven statistic feature extraction, the architecture of the deep neural network could be limited to only dense layers compared to the architecture of our previous paper where feature extraction was performed directly on the time series data by the use of recurrent and convolutional layers. With the statistical featureset, the classification and domain adaptation task was performed on a Multi-Layer Perceptron (MLP) architecture without the need of such computationally intensive recurrent or convolutional layers. The network architecture is composed by three blocks (outlined blocks in Fig. 3), each consists of the combination: Dense, Batch Normalization (BN), Dropout (DT) and ReLU activation Layer. All Dense layers consist of 128 neurons each. To prevent overfitting on the training data, dropout was applied with the rates: *0.8, 0.5, 0.5* (from the first to the third DT layer). To further archive better generalization, L1/L2 regularization was applied on the weights and bias terms of the Dense layers. The intention behind the two regularizers is that the L1 (lasso regression) term penalizes the sum of the weight values. Weights close to zero are petty since they influence the model predictions very low and thus become zero to reduce the parameter load. With the L2 (ridge regression or weight decay) term, the sum of the square of the weight values is penalized. The regularization terms and dropout layers reduce the network complexity, decrease the risk of overfitting while training and thus improve stability on predicting unseen samples at inference.

3.3. Loss function

Learning a model aims to find a model which minimizes a given loss function. In the task of unsupervised domain adaptation, the goal is to find a model that maximizes the classification accuracy on the source data and at the same time minimizes the difference between the source and the target feature embeddings. Thus, the loss function $\mathcal{L}$ of the model is the sum of the classification loss of the source branch and the domain adaptation loss $\mathcal{L}_{i,DA}$ in the *i-th* layer (L) (Che, Wang, Fu & Ni (2019)):

$$\mathcal{L} = \alpha * \mathcal{L}_{SM} + \sum_{i=1}^{L} \beta_i * \mathcal{L}_{i,DA} \tag{1}$$

The α and β terms in Eq. (1) are introduced to control the influence of the two sub losses on the total loss $\mathcal{L}$. The term $\mathcal{L}_{SM}$ refers to the softmax loss for the supervised classification task:

$$P(y = j|x) = \frac{e^{x^T w_j}}{\sum_{c=1}^{C} e^{x^T w_c}} \tag{2}$$

As the metric for measuring the domain discrepancy$\mathcal{L}_{i,DA}$, we decided to use the Maximum Mean Discrepancy (MMD), which is frequently used in domain adaptation related to bearing fault detection on vibrational data as Ma, Zhang, Fan, & Wang (2020) proposed. The MMD was introduced by Gretton, Borgwardt, Rasch, Schölkopf & Smola (2012) as a non-parametric distance metric to measure the distribution discrepancy between two domains, based on the differences of the feature means in a Reproducing Kernel Hilbert Space (RKHS):

$$MMD(\mathcal{P}(\mathcal{X}_s), \mathcal{P}(\mathcal{X}_t)) = \|\mu_S - \mu_T\|_{\mathcal{H}}^2 \tag{3}$$

For the problem setting faced in this paper, namely the adaptation of probability distributions of one domain to those of a different but related domain, the second order binomial of the empirical MMD is interpreted as the sum of the intrinsic- and extrinsic distribution similarities within and between the domains resp. operating points:

$$\begin{aligned} MMD(\mathcal{X}^S, \mathcal{X}^T) = \mathbb{E}_{\mathcal{X}^S}[k(\mathcal{X}^S, \mathcal{X}^{S\prime})] \\ -2 * \mathbb{E}_{\mathcal{X}^{S,T}}[k(\mathcal{X}^S, \mathcal{X}^T)] \\ +\mathbb{E}_{\mathcal{X}^T}[k(\mathcal{X}^T, \mathcal{X}^{T\prime})] \end{aligned} \tag{4}$$

In Eq. (4) the term $\mathbb{E}(P,Q)$ represents the expectation that a given sample of Q is drawn from the same distribution than P and k denotes the kernel function to map the original features to the higher dimensional RKHS. As the kernel we used a radial basis function (RBF). Recommendations of using multiple kernel functions were given by Li, Zhang, Ding & Sun (2019) to improve the domain discrepancy reduction and in general obtain a higher stability of the model performance for fault classification. Therefore, we applied three RBF kernels to each of the domain adaptation layers. The corresponding kernel sigma parameters (bandwidth) were set to *1.0*, *2.0*, *3.0*. The overall result is determined by averaging the single kernel results. The approach proposed by Long, Cao, Wang & Jordan (2015) applied domain adaptation loss calculation in multiple layers to overcome the deterioration of the transferability of the features. Therefore, we expanded the default MMD to use Multi-Kernels as well as applying it in a Multi-Layer setting, ML-MK-MMD for short.

4. Experimental Evaluation of the Proposed Approach

This section evaluates the proposed approach by experimentally studies on a multiclass bearing fault dataset, in particular the influence of the additional domain adaptation mechanism on the classification performance of both domain branches. In addition, further attention is paid on the aptitude of the handcrafted features to reach a more slender network compared to deep learning based feature extraction models.

4.1. Explanations on the used dataset and experiment settings

Regarding the given constraints on phase current based bearing fault detection, the number of datasets available is limited. Experiments were carried out by using a benchmark open source dataset for phase current bearing fault diagnosis, provided by Lessmeier, Kimotho, Zimmer & Sextro (2016) of Paderborn University. Figure 4 shows the test rig used for the data generation. The main parts of the test rig are (1) the examinee motor, (2) load motor and (3) the housing where the bearing under test is placed. The dataset contains data of 32 different bearing instances, all of the same bearing type 6203 from different manufacturers. The dataset contains four different working conditions, parametrized by the rotational shaft frequency of the examinee motor, the load torque generated by the load motor and the radial force directly applied to the bearing using a spring. Table 2 shows the available settings for the four operating conditions. The data of each bearing was generated by measuring each working condition for in total 80 seconds (20 repeats of 4 seconds each).

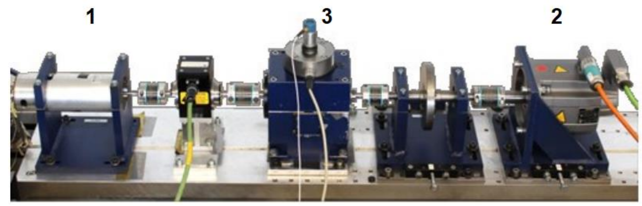

Figure 4: Paderborn Test rig by Lessmeier et al. (2016)

Table 2. Considered operating conditions

No.	Speed [rpm]	Torque [Nm]	Radial force [N]
1	1500	0.7	1000
2	900	0.7	1000
3	1500	0.1	1000
4	1500	0.7	400

The used bearings are either healthy (undamaged, "OK") or have outer ring ("OR") or inner ring ("IR") faults. To be precise, the following instances were used: K001-K006 ("OK"), KA01, KA03, KA05-KA09, KA15, KA16, KA22 & KA30 ("OR") and KI01, KI03, KI05, KI07 & KI08 ("IR"). We performed 12 transfer experiments between a source (S) and target (T) domain to validate the applicability of the proposed approach on different working conditions. Each transfer task was carried out by a five fold cross validation with a train/test ratio of 80/20. Table 3 gives an overview of the transfer scenarios and the averaged results. The overall loss function was optimized by the Adam optimizer with a learning rate of 1e-3. The weights for the classification and domain adaptation losses were set to: α = 4.0, $\beta_1 = 1.0$, $\beta_2 = 2.0$, $\beta_3 = 3.0$.

Table 3. Transfer scenarios and corresponding target results with and without (Baseline) domain adaptation

Task		**Acc. without DA (Baseline)**		**Acc. with DA**		**Acc. Improvements**	
S	**T**	**Src**	**Tar**	**Src**	**Tar**	**Src**	**Tar**
1	2	0.93	0.88	0.99	0.93	+ *0.06*	+ *0.06*
1	3	0.94	0.90	0.99	0.95	+ *0.05*	+ *0.05*
1	4	0.94	0.93	0.98	0.96	+ *0.04*	+ *0.03*
2	1	0.91	0.88	0.97	0.92	+ *0.06*	+ *0.05*
2	3	0.92	0.85	0.96	0.92	+ *0.04*	+ *0.07*
2	4	0.91	0.87	0.97	0.91	+ *0.06*	+ *0.04*
3	1	0.94	0.89	0.99	0.95	+ *0.05*	+ *0.06*
3	2	0.93	0.86	0.98	0.93	+ *0.05*	+ *0.07*
3	4	0.93	0.90	0.98	0.94	+ *0.05*	+ *0.05*
4	1	0.95	0.89	0.98	0.96	+ *0.03*	+ *0.07*
4	2	0.96	0.84	0.99	0.87	+ *0.03*	+ *0.03*
4	3	0.96	0.89	0.99	0.90	+ *0.03*	+ *0.01*

4.2. Results without Domain Adaptation (Baseline)

First, the proposed method was evaluated for the applicability on automated bearing fault detection using phase-current data. No attention to operating point transfer was spend. Therefore no domain adaptation was applied to obtain a baseline for each operating condition. Only the supervised branch of the network presented in Fig. 3 was deployed and train/test was performed on the same operating point.

Figure 5 shows the training history of the single (source-only) branch on operating point 4. The black dot highlights the inference accuracy of the trained model on data coming from operating point 1 and thus is interpreted as the baseline cross operating point accuracy. It stands out that the validation accuracy substantially exceeds the training accuracy, which is caused by the large regularization (Dropout & L1L2) which makes the training process artificially more complicated to converge. The abrupt collapse of the validation accuracy at initial epochs also indicates the regularization effect on the generalization performance of the model. Due to the muted units, some sample information is suppressed, which is why subsequent layers attempt predictions based on that incomplete embedding representations. At inference, no regularization is present and thus the entire computational strength of the model is unleashed. The accuracies on the target data is relatively high even if no domain adaptation is used. This advocates the suitability of the extracted features to address the bearing fault classification task. But nevertheless, a classifier trained on source domain data collapses in its accuracy by up to 12% at inference on target domain data, which also indicates a remaining domain discrepancy despite the suitable features. The further section discusses the proposed domain adaptation approach to improve the baseline accuracies for either the supervised source- as well as the unsupervised target domain.

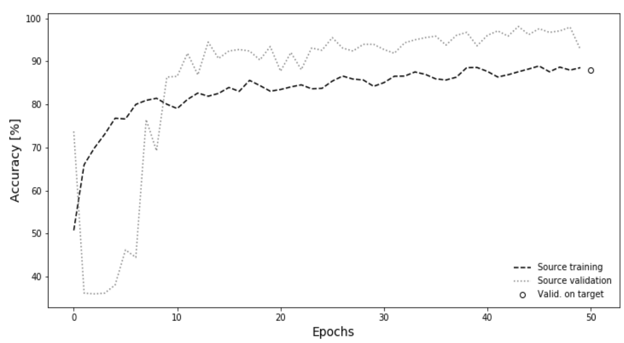

Figure 5: Training history on operating point 4

4.3. Influence of the domain adaptation approach

To improve the accuracies on both domains, the Domain Adaptation approach is evaluated by adding the DA terms to the model loss function. We first evaluate the impact of the proposed method on the source domain data. The baseline Source- as well as the DA-Source accuracy columns in table 3 give the corresponding comparison. The shared weights architecture with the applied domain discrepancy loss forces the network to extract domain invariant features while still increasing the performance of the source domain Softmax layer through the supervised training. An improvement of the source domain accuracy was found over all cross-validated transfer scenarios. Considering the cross-operating point accuracy (comp. Baseline), the proposed domain adaptation approach improves predictions on unseen operating points up to 7%. The applied weights for the classification- and domain adaptation loss circumvents the threats of negative transfer in a manner that improvements are found on limited extend but nevertheless on both domains. Figure 6 summarizes that circumstance by comparing the train/test history of the source only branch and the corresponding source- and target accuracies as well as the history of the proposed domain adaptation network with shared weights and applied MMD loss in all hidden layers.

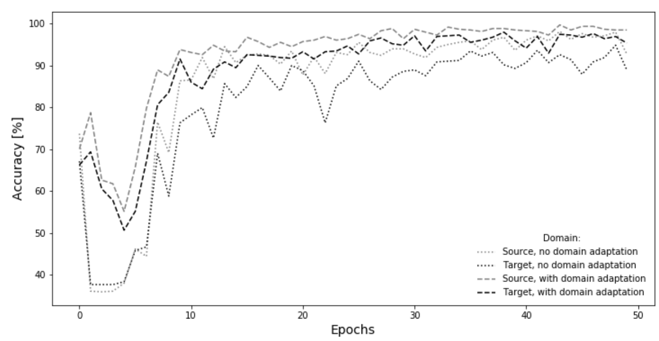

Figure 6: Comparison of baseline and domain adaptation performance on source and target domain

5. Contribution and Originality

The study mainly founds on the following three contributions:

- Bearing fault detection using internal motor signals instead of external sensors
- Domain knowledge driven feature engineering for bearing faults to downsize the necessary network parameters by enlarging the information quality of input variable space
- Coupling the feature spaces of data coming from different operating points to improve domain invariance and gain increased accuracies on both domains

As an alternative to vibration sensors, comparably good accuracies could be achieved based on phase current data. Compared to our previous studies, the overall model complexity could be reduced to a performant Dense-Only

(MLP) architecture by extracting domain-related features instead of using time series data directly. To overcome the drawback of the lower information content, extractable by the phase current raw signals, compared to vibration signals (see chap. 2), we resorted to features that specifically address the physical changes of the motor that are triggered by a bearing damage, namely torque fluctuations caused by air gap changes through eccentricities between rotor and stator. Like analyzation models trained on the raw data directly, also the domain referenced handcrafted features were found to be subjected to the operating point the sample was drawn at. The performance degradations triggered by this were mitigated by means of domain adaptation to couple different but related domains with a discrepancy loss to iteratively bring the feature distributions of the input embeddings closer together.

6. Future Work

As shown in Figure 6 the integration of a second branch for other operating point data has led to improvements on both domain accuracies even if there was labeled data available for only one branch. This suggests the idea that coupling even more branches of different operating points could bring further enhancements - but this requires a rethinking of the taken domain adaptation approach as the discrepancy metric in the presented approach is based on a two-sample test.

Apart from the methodology and model investigations, we believe it is unmitigated necessary to question the underlying dataset and test rig setup. Available data sets are limited to external placed bearings. However, for bearing damage analyses such as those required in the industrial environment for motor analysis, it is necessary to consider the bearing component in its installed state.

We also take up the critics of Smith & Randall (2015) regarding the selected operating point influences, especially that of the torque component, since this should defacto have no influence on the externally placed bearing.

References

Blodt, M., Granjon, P., Raison, B., Rostaing, G. (2008). Models for Bearing Damage Detection in Induction Motors Using Stator Current Monitoring. In: IEEE Trans. Ind. Electron. Vol.: 55, Pages: 1813-1822. DOI: 10.1109/TIE.2008.917108

Carvalho, T., Soares, F., Vita, R., Francisco, P., Basto, J. & Alcalá, S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. In: Computers & Industrial Engineering, Vol. 137

Che, C., Wang, H., Fu, Q. & Ni, X. (2019). Deep transfer learning for rolling bearing fault diagnosis under variable operating conditions. In: Advances in Mechanical Engineering. Vol.: 11, DOI: 10.1177/1687814019897212

Christ, M, Braun, N., Neuffer, J., Kempa-Liehr, A. (2018) Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package), Neurocomputing, Volume 307, 2018, Pages 72-77, ISSN 0925-2312

Gretton, A., Borgwardt K., Rasch, M., Schölkopf, B. and Smola, A. (2012). A Kernel Two-Sample Test. In *Journal of Machine Learning Research*, Vol. 13 (25), 723-773. http://jmlr.org/papers/v13/gretton12a.html

Hoang, D. T. & Kang, H. J. (2019). A Motor Current Signal Based Bearing Fault Diagnosis Using Deep Learning And Information Fusion. In: IEEE Transactions on Instrumentation and Measurement, 2019. DOI: 10.1109/TIM.2019.2933119

Höhr, S., Tasci, T. & Verl, A. (2019). Realization of Data Analytics Projects in Manufacturing Using a Microservice-Based Approach. In: 2019 IEEE International Conference on Mechatronics (ICM), Ilmenau, Germany, 2019, Pages: 321-326, doi: 10.1109/ICMECH.2019.8722926.

Lade, P, Ghosh, R. & Srinivasan, S. (2017). Manufacturing Analytics and Industrial Internet of Things. In: IEEE Intelligent Systems, vol. 32, no. 3, Pages: 74-79, May-June 2017, doi: 10.1109/MIS.2017.49.

Lessmeier, Chr., Kimotho, J. K., Zimmer D. & Sextro W. (2016). Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification. In *European Conference of the Prognostics and Health Management Society*, Volume 7, Pages 17. http://www.phmsociety.org/node/1834/

Li, K., Xiong, M., Li, F., Su, L. & Wu, J. (2019). A novel fault diagnosis algorithm for rotating machinery based on a sparsity and neighborhood preserving deep extreme learning machine. In: Neurocomputing (Vol. 350), Pages: 261-270. DOI: 10.1016/j.neucom.2019.03.084

Li, X.; Zhang, W.; Ding, Q. and Sun, J.-Q. (2019) Multi-Layer domain adaptation method for rolling bearing fault diagnosis. *Signal Processing*, Vol. 157 (Pages 180-197), April, DOI: 10.1016/j.sigpro.2018.12.005

Long, M., Cao, Y., Wang, J. & Jordan, M.I. (2015) Learning Transferable Features with Deep Adaptation Networks. *International Conference on Machine Learning*, Vol. 32 (97-105). France. http://proceedings.mlr.press/v37/long15.html

Ma, P.; Zhang, H.; Fan, W. & Wang, C. (2020). A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. Instrument Society of America (ISA) Transactions (Pages 465-478), April, https://doi.org/10.1016/j.isatra.2019.08.040

Raison, B. Rostaing, G., Butscher, O. & Maroni, C. (2002). Investigations of algorithms for bearing fault detection in induction drives. In: IEEE 2002 28th Annual Conference of the Industrial Electronics Society. IECON 02, vol. 2, Pages: 1696-1701

Schoen, R. R., Habetler, T.G., Kamran, F. & Bartheld R. (1994). Motor bearing damage detection using stator current monitoring. In: IEEE industry applications society: IAS ’94, 2-6 October 1994. Pages: 110-116. DOI: 10.1109/IAS.1994.345491

Rosero, J. Cusido, J., Garcia, A., Ortega, J. & Romeral, L. (2006). Broken bearings and eccentricity fault detection for a permanent magnet synchronous motor. *2006 1st International Conference on Nano-Networks and Workshops*, Pages: 964-969

Rosero, J., Cusido, J., Garcia Espinosa ,A., Ortega, J. A. & Romeral, L., (2007). Broken bearings fault detection for a permanent magnet synchronous motor under non-constant working conditions by means of a joint time frequency analysis. In: *2007 IEEE International Symposium on Industrial Electronics*, Pages: 3415–3419, Piscataway, N.J., 2007. DOI: 10.1109/ISIE.2007.4375165

Smith W.A. & Randall R.B. (2015). Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. In: *Mechanical Systems and Signal Processing*, Volumes: 64-65, 2015, DOI: https://doi.org/10.1016/j.ymssp.2015.04.021.

Tong, Z., Li, W. & Zhang, B. (2018). Bearing Fault Diagnosis Under Variable Working Conditions Based on Domain Adaptation Using Feature Transfer Learning.

Trajin, B., Regnier, J. & Faucher, J. (2010). Comparison between vibration and stator current analysis for the detection of bearing faults in asynchronous drives. In: Electric Power Applications IET, vol. 4, Pages: 90-100.

Wagner T. & Sommer S. (2020). Bearing fault detection using deep neural network and weighted ensemble learning for multiple motor phase current sources. In *International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 24-26 August, DOI: 10.1109/INISTA49547.2020.9194618.

Zheng, H., Wang, R., Yang, Y. et al. (2019): Cross-Domain Fault Diagnosis Using Knowledge Transfer Strategy: A Review.

Zhou, J., Zheng, L.Y., Wang, Y. & Gogu, C. (2020). A Multistage Deep Transfer Learning Method for Machinery Fault Diagnostics across Diverse Working Conditions and Devices. In: IEEE Access 8, Pages: 80879-80898. (2020). DOI: 10.1109/ACCESS.2020.2990739.