Supervised Transfer Learning Framework for Fault Diagnosis in Wind Turbines

Kenan Weber and Christine Preisach

Hochschule Karlsruhe - University of Applied Sciences
`kenan.weber@h-ka.de`
`christine.preisach@h-ka.de`

Abstract. Common challenges in fault diagnosis include the lack of labeled data and the need to build models for each domain, resulting in many models that require supervision. Transfer learning can help tackle these challenges by learning cross-domain knowledge. Many approaches still require at least some labeled data in the target domain, and often provide unexplainable results. To this end, we propose a supervised transfer learning framework for fault diagnosis in wind turbines that operates in an *Anomaly-Space*. This space was created using SCADA data and vibration data and was built and provided to us by our research partner. Data within the Anomaly-Space can be interpreted as anomaly scores for each component in the wind turbine, making each value intuitive to understand. We conducted cross-domain evaluation on the train set using popular supervised classifiers like Random Forest, Light-Gradient-Boosting-Machines and Multilayer Perceptron as metamodels for the diagnosis of bearing and sensor faults. The Multilayer Perceptron achieved the highest classification performance. This model was then used for a final evaluation in our test set. The results show, that the proposed framework is able to detect cross-domain faults in the test set with a high degree of accuracy by using one single classifier, which is a significant asset to the diagnostic team.

Keywords: Condition Monitoring, Wind Turbines, Anomaly Detection, Fault Detection, Fault Diagnosis, Transfer Learning

1 Introduction

In Germany, electricity generated from wind turbines (WTs) makes up a large portion of the total generated energy from renewable energy sources [1]. In order to increase the total energy yield, it is very important to reduce total downtimes by monitoring critical aspects of WTs. With condition monitoring, faults can be detected early and maintenance times and measures can be planned accordingly. This further reduces the risk of total failure due to the propagation of the faults to other areas of the machine. The total amount of WTs is increasing, with each of them equipped with more and more sensors over the years. Hence, this raises the amount of components, that can be measured which leads to higher costs and more signals, that need to be monitored. Furthermore, when a fault has been detected, we still need to infer the type of fault and localize it, e.g. which component(s) is/are affected. With an increasing amount of signals, there is need for more and more highly specialized personnel in order to monitor these machines manually. Therefore, we propose a solution that automates this whole process from end-to-end.

Supervised learning is a method for performing intelligent fault diagnosis on WTs. In literature, numerous solutions based on supervised learning are available [2, 3, 4]. However, these solutions typically focus solely on diagnosing faults in one particular machine, resulting in the development of separate models for each machine. Transfer learning is a promising approach to use knowledge extracted from a subset of WTs to multiple other ones. This can decrease the total model count needed for reliable fault diagnosis. Additionally, transfer learning based fault diagnosis is not limited to detect only those fault types, that have been occurred in the past on that particular WT, since fault data from other WTs can also be used for the diagnosis. Many transfer learning solutions for WTs exist. Zhang et al. [5] built a fully connected neural network which is able to detect, whether ice are on the WT blades by only using SCADA data. A small data set from another WT was used to fine-tune the model. Yang et al. [6] are able to detect blade defects by segmenting blade images with the otsu threshold segmentation algorithm and then using a pre-trained Alexnet classifier for the feature extraction. A Random Forest was used for the fault diagnosis in the last step. Li et al. [7] pre-trained a convolutional autoencoder on SCADA data from 14 WTs and fine-tuned the model on data from the 15th WT in order to detect fault types like high temperature in gearbox or generator or low pressure of the hydraulic system. A stacked autoencoder was employed by Deng et al. [8] by pretraining the model on source data and then utilizing and fine-tuning a fully connected layer for the diagnosis in the target domain.

The aforementioned solutions have the limitation, that some labeled data in the source domain have to be available. Furthermore, features extracted with the help of neural networks are mostly abstract and not interpretable for diagnosticians. If the fault diagnosis system indicates that a fault is present, then it has to be clear to the diagnostician, *how* the decision was made. To this end, we propose a fault diagnosis framework, which operates in an *Anomaly-Space*. This new feature space provides several normalized anomaly scores for each WT component, for every available WT. Values above 1.0 are considered anomalous. It was built and provided to us by our research partner EnBW Energie Baden-Württemberg AG. Both SCADA data and vibration data were used for the creation of the Anomaly-Space. A supervised classifier takes data from the Anomaly-Space as input and provides fault diagnosis results. These results can easily be interpreted by diagnosticians, since features in the Anomaly-Space represent deviations from the normal behavior of the WT. This can be seen as a feature-based transfer learning approach, where the Anomaly-Space represents the domain-shared feature space.

In summary, the contributions of this paper are the following:

1. Fault diagnosis based on derived signals from SCADA data and vibration data, that are easily interpretable, in contrast to many other transfer learning approaches.

2. Extensive model training and evaluation with stratified cross-validation from real data across 5 WTs from 4 wind parks and comparing classification performance of popular supervised classifiers, such as Random Forest (RF), Light-Gradient-Boosting-Machines (LightGBM) and Multilayer Perceptron (MLP).

3. Showing transfer learning capabilities by applying the best performing classifier from the aforementioned analysis on a new test set, which consists of 2 WTs from 2 wind parks, one of which is a completely different wind park compared to the train set.

This paper is organized as follows: In section 2, we give an overview about the dataset. This includes a brief description and explanation of the Anomaly-Space. In section 3,

some background information about transfer learning in fault diagnosis is given. Our supervised fault diagnosis framework is introduced in section 4. Results are shown in section 5 and conclusions are made in section 6.

2 Dataset Description

The dataset we used contains fault-types and anomaly scores, which are deduced from SCADA data and vibration data. Two common faults can be found within the data: bearing fault and sensor fault. Bearing faults are common and severe faults in WTs. Ignoring these can result in a total failure of the machine and leads to substantial downtimes and repair costs. Therefore, there is a huge interest in detection of bearing faults as early as possible.

Faulty sensors are also very common in WTs, but are very cheap to replace in terms of raw material cost and are not damaging the WT itself. At a first glance, this might not be as important, but these type of faults can lead to several problems. Unnecessary maintenance works could be performed if sensor faults are confused with a more serious disturbance. Consequently, diagnosticians and technicians might increasingly distrust the fault diagnosis application and/or mistake a serious fault with a sensor fault.

SCADA data are typically used for condition monitoring in WTs. In general, the term SCADA stands for "Supervisory Control and Data Acquisition" and refers to the monitoring and control of technical processes using data that originates from sensors, actuators and other field devices and is sent to a control system. Among other things, process variables such as temperature, pressure and similar values are recorded. Each recorded 10-minute window is aggregated into four scalar values: minimum, maximum, standard deviation and average. Vibration sensors are able to capture data in a much higher sample rate than SCADA and are commonly used to identify early signs of wear, imbalance, or misalignment in rotating machinery.

The data originates from a total of 7 WTs, across 5 different wind parks. All fault cases with further information are listed in Table 1. Our train set (case 1 to 6) consists of data from 5 WTs, which are from 4 different wind parks. The test set (case 7 and 8) has data from 2 WTs, one of which is from a completely different wind park.

Table 1. More information about the data. The train data is considered the source data, validation and test data are considered target data. P = park, U = unit, (N)DE = (Non-)Drive End.

Park/Unit	Fault-Type	Fault-Location	Dataset	Case-No.
P1/U1	sensor fault	temperature generator phase 3	train/validation	1
P2/U1	sensor fault	temperature transformator phase 3	train/validation	2
P2/U1	sensor fault	temperature generator phase 1	train/validation	3
P3/U1	bearing fault	Fast Shaft Bearing DE	train/validation	4
P3/U2	bearing fault	Fast Shaft Bearing NDE	train/validation	5
P4/U1	bearing fault	Fast Shaft Bearing DE	train/validation	6
P1/U2	sensor fault	temperature generator phase 2	test	7
P5/U1	bearing fault	Fast Shaft Bearing DE	test	8

2.1 Anomaly-Space

The Anomaly-Space refers to a feature space, that has been constructed by using multiple proprietary algorithms, denoted as detectors. The input of these detectors are SCADA data and vibration data.

Detectors are monitoring critical components of the WT and provide anomaly scores, which represent a deviation from the normal behavior. These values are normalized such that measurements with values above 1.0 are considered anomalous.

Broad-Band-Characteristic-Value (bbcv) is one of the detectors. This detector first captures multiple windows of vibration data when pre-defined conditions have been met (e.g. approximately constant wind speeds). Several features are then extracted from the raw vibration data and from the frequency domain, after applying the Fast-Fourier-Transformation (FFT), such as skewness, kurtosis and average values. In the last step, the trendiness of the aforementioned features with the help of hypothesis testing is being calculated. Bearing faults will usually result in an increased trendiness in multiple features.

Another detector is the tuplet detector. This detector is designed to detect SCADA data deviations from groups of semantically similar components, for example the generator temperature of all three voltage supply phases. This is being achieved by monitoring the variance of the measurements of these components. A sensor fault in any of these components significantly increases the variance. A statistical test quantifies the difference to the expected null hypothesis, null hypothesis being a variance value of 0.

The procedure of both detectors are depicted in Figure 1 and Figure 2.

The Anomaly-Space is provided to us by our research partner EnBW Energie Baden-Württemberg AG and is used as input data to our preprocessing steps and metamodels.
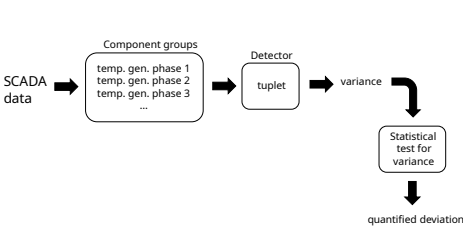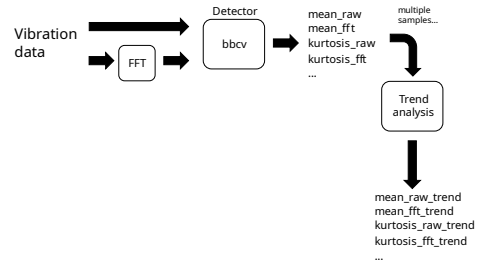

Fig. 1. tuplet detector procedure

Fig. 2. bbcv detector procedure

3 Transfer Learning in Fault Diagnosis

Fault diagnosis can be decomposed into fault detection and fault classification. Fault detection (e.g. using anomaly/outlier detection) is the detection of deviations from normal behavior in the data. These deviations come from either a fault within the monitored system, or from a faulty measuring device. The transformation of the input data can help in detecting faults, that are otherwise not visible.

To be able to diagnose the type of fault, the transformed data can then be classified into pre-defined fault-classes (labels) using supervised classification algorithms.

One of the challenges of fault diagnosis is the lack of labels, since faults occur rarely. In the context of WTs, there might be many fault-labels available across all available WTs. However, there are many reasons why it is still not eligible to train a ML model on one WT and apply it on another without some modifications. For example, WTs may come from different manufacturers and therefore have different components, which results in non-similar signal patterns. Consequently, at least one model for each WT needs to be created.

Transfer learning aims to capture domain-invariant knowledge from just few available domains and apply this knowledge onto other domains. In the context of WTs, for instance, the objective is to leverage insights gained from a few WTs and extend them to others.

There are 4 types of transfer learning methods according to Lei et al. [9]: **feature-based** approaches, **GAN-based** approaches, **instance-based** approaches and **parameter-based** approaches. Feature-based approaches map cross-domain data into a common feature space and decrease the distribution discrepancy before applying a classifier. GAN-based approaches utilize the GAN framework in order to learn the distribution of the target data and generate new ones to improve the classifier. Instance-based approaches reweight misclassified instances from source and target domain, increasing/decreasing the influence of those instances on the fault diagnosis classifier. Parameter-based approaches train models (e.g. neural networks) on the source data and fine-tune the learned model parameters on the target data. Instance- and parameter-based approaches assume, that few labeled samples of the target domain are available. Our solution can be regarded as a feature-based TL approach without the step of further decreasing the distribution discrepancy. This space is designed to represent deviations from the individual normal behavior, making each value intuitive to understand.

4   Supervised Fault Diagnosis Framework

A general overview of the proposed fault diagnosis framework is depicted in Figure 3.

The dataset labels are generated using fault time frames (appearance dates and repair dates) provided by diagnosticians. Data from within the fault time frames are labeled with the corresponding fault type. Data outside these time frames are labeled as "Normal". Data that are not within the normal operating-mode of the WT (e.g. stillstand or wind speeds below a specified threshold) have been omitted. A time-based forward fill was applied, filling missing values for up to 3 hours after the first occurrence. Remaining data gaps are filled with the value 0.0.

The resulting data frame has data from each available WT component as observations (rows) with each detector output being a feature (column).

The amount of features provided by the bbcv detector is reduced to one single feature, by only keeping the feature with the largest variance. Consequently, both detectors provide only a single feature each, resulting so far in 2 total features.

Sliding-window based feature extraction has been employed in order to capture the relationship between neighbored data samples, with a window-size of 144 and a stride of 1. The extracted features are trend-certainty (tc) and variance (var). The Mann-Kendall
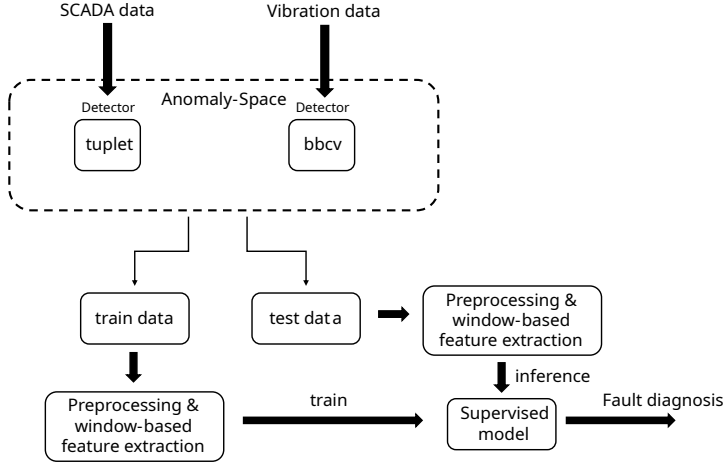
Fig. 3. Proposed fault diagnosis framework

test is a statistical test for determining trends in data. We used Equation 1 to determine the value of our trend-certainty feature, with p_{mk} being the p-value for having a positive trend within a window. The values of both features have been manually set to 0.0, if the window contains only values below 1.0. The final feature count has been increased to 6 after extracting those features.

For MLP, the base features, e.g. the outputs of both detectors, have been normalized with a min-max scaler.

$$tc = \begin{cases} 1 & \text{if } p_{mk} < 0.001, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

5 Results & Discussion

Stratified 3-fold cross-validation was used for evaluation on the train data. We chose the f_β-score with $\beta = 0.5$ as our evaluation metric. This way, precision has a larger impact than recall. This choice was made because trusting the fault diagnosis is crucial, which can be achieved by minimizing false positive predictions; therefore, precision should be weighted more heavily. Additionally, measurements from faulty sensors could occasionally resemble healthy ones, if the root-cause is loose contact. There could be multiple days of data within the fault time frame, where the sensor delivers non-faulty measurements. Weighting precision and recall equally (e.g. with $\beta = 1$) would provide overly pessimistic evaluation values in these cases.

5.1 Comparison between different classifiers on train data

Several popular classifiers were used for our evaluation on the train data: RF, LightGBM and MLP. We created a baseline model which classifies each instance with bbcv values above 1.0 as bearing fault and tuplet values above 1.0 as sensor fault, which we simply termed *Above-One*. The results are depicted in Figure 4. The best performing model

is MLP, achieving a $F_{0.5}$ score of **0.874** with the following hyperparameters: ReLU activation function, Adam optimizer, learning rate of 0.001 and 1 hidden layer with 5 neurons.
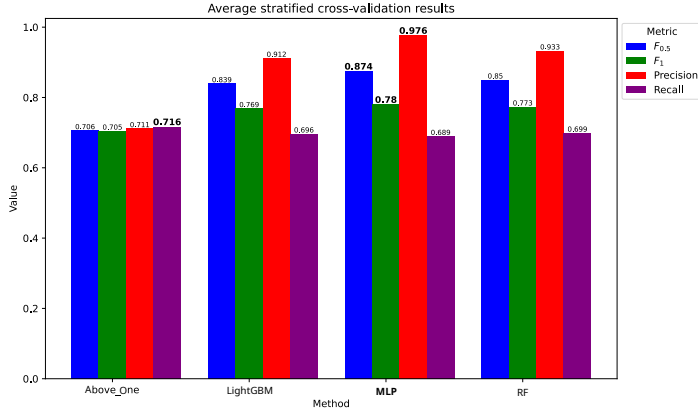

Fig. 4. Average stratified cross-validation results

5.2 Evaluating the best classifier on test data

The MLP was trained on the whole train data with the best parameters and then applied on the test data. Table 1 shows the results. The model achieved a $F_{0.5}$ score of **0.937**.

Table 2. Evaluation results on test data.

Method	$F_{0.5}$ score	F_1 score	Precision	Recall
mlp	**0.937**	0.871	0.992	0.789

5.3 Discussion

It can be seen from Figure 4 and Table 2, that higher evaluation scores have been achieved in the test data (0.874 vs. 0.937). This is due to data quality issues, that are present in the train data. More precisely, there is a loose contact sensor fault case, in which multiple signal segments appear normal, within the fault time frame. These segments can be present multiple days. This complicates the fault diagnosis evaluation, since our window size for our feature extraction methods is approximately one day. Possible solutions would be to increase the window size or to split the fault time frame into multiple smaller ones to increase the coverage of the visible fault pattern.

6 Conclusion & Future Work

In this paper, a fault diagnosis framework based on an Anomaly-Space is proposed. The Anomaly-Space is a feature space, in which deviations from normal behavior (anomaly

scores) for each WT component are encoded. Window-based features are then extracted from the Anomaly-Space, such as trend values extracted with the Mann-Kendall test. This new feature space provides intuitive values which can help explain fault diagnosis results, since these values represent deviations from the normal behavior in contrast to many other approaches. This framework can be regarded as a feature-based transfer learning method without further decreasing the distribution discrepancy. Supervised classifiers such as Random Forest, Light-Gradient-Boosting-Machines and Multilayer Perceptron are compared on the train data with stratified cross-validation. The Multilayer Perceptron achieved the highest classification performance in diagnosing bearing and sensor faults and was tested on 2 new WTs, one of which stems from a different wind park, compared to the train data. This final evaluation also showed good results, making this a promising fault diagnosis approach for cross-domain fault diagnosis. Future work could include Out-Of-Distribution (OOD) detection to the framework, in order to detect previously unseen fault types.

Acknowledgement

References

[1] Geschäftsstelle der Arbeitsgruppe Erneuerbare Energien-Statistik (AGEE-Stat) am Umweltbundesamt. *Erneuerbare Energien in Deutschland Daten zur Entwicklung im Jahr 2023*. Accessed: 7-19-2024. URL: https://www.umweltbundesamt.de/publikationen/erneuerbare-energien-in-deutschland-2023.

[2] Guoqian Jiang et al. "Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox". In: *IEEE Transactions on Industrial Electronics* 66.4 (2019), pp. 3196–3207. DOI: 10.1109/TIE.2018.2844805.

[3] Yanting Li, Shujun Liu, and Lianjie Shu. "Wind turbine fault diagnosis based on Gaussian process classifiers applied to operational data". In: *Renewable Energy* 134 (2019), pp. 357–366. ISSN: 0960-1481. DOI: 10.1016/j.renene.2018.10.088.

[4] Zhi-Xin Yang, Xian-Bo Wang, and Jian-Hua Zhong. "Representational Learning for Fault Diagnosis of Wind Turbine Equipment: A Multi-Layered Extreme Learning Machines Approach". In: *Energies* 9.6 (2016). ISSN: 1996-1073. DOI: 10.3390/en9060379.

[5] Chengkai Zhang, Junchi Bin, and Zheng Liu. "Wind turbine ice assessment through inductive transfer learning". In: *2018 ieee international instrumentation and measurement technology conference (i2mtc)*. IEEE. 2018, pp. 1–6. DOI: 10.1109/I2MTC.2018.8409794.

[6] Xiyun Yang et al. "Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier". In: *Renewable Energy* 163 (2021), pp. 386–397. DOI: 10.1016/j.renene.2020.08.125.

[7] Yanting Li et al. "Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data". In: *Renewable Energy* 171 (2021), pp. 103–115. DOI: 10.1016/j.renene.2021.01.143.

[8] Ziwei Deng et al. "A deep transfer learning method based on stacked autoencoder for cross-domain fault diagnosis". In: *Applied Mathematics and Computation* 408 (2021), p. 126318. DOI: `10.1016/j.amc.2021.126318`.

[9] Yaguo Lei et al. "Applications of machine learning to machine fault diagnosis: A review and roadmap". In: *Mechanical systems and signal processing* 138 (2020), p. 106587. DOI: `10.1016/j.ymssp.2019.106587`.