Searching for Feature Sets for Misalignment Classification Using Experimental Data and Data Mining

Sebastian Bold and Sven Urschel

Working Group for Electrotechnical Systems of Mechatronics − Kaiserslautern University of Applied Sciences, Kaiserslautern, Germany
{sebastian.bold,sven.urschel}@hs-kl.de

Abstract. Misalignment causes heat, which leads to increased wear, stator-rotor friction, and, in the worst case, fatigue cracks. Misalignment can occur in any motor-driven assembly unless the motor is integrated into the enclosure, such as in pumps, fans, or gearboxes. In order to avoid short operation times or even unexpected down-times, companies spend a lot of effort on avoiding misalignment. But even with high precision devices such as optical alignment systems, it is not always possible to avoid the occurrence of misalignment, especially during processes resulting from heat expansion or vibration.

These problems can be overcome with a monitoring approach. Besides optical systems based on high-speed cameras, approaches based on the analysis of the motor current are also promising. By comparison, currentbased approaches are cheaper due to the simpler equipment. Approaches like motor current signature analysis (MCSA) are well described in the literature. To execute this approach, the spectra of the motor current and the fault-related frequencies are calculated. By comparing the magnitude at the specific frequency with a limit, an alarm or warning can be sent. In [1] and [2], it is shown that misalignment has an influence on the motor current. The main disadvantage of MCSA is the load dependence, which was investigated in [3]. This means that it is not possible to distinguish between load variation and an increase of misalignment.

More sophisticated classifiers such as k-nearest neighbors or artificial neural networks may solve these problems by using more sources of information. To apply such a classifier, a new feature set needs to be identified. In order to find the feature set, this work focuses on data mining, which includes two steps. Since no public data for misalignment was available, an experiment was needed to create such data. The first step in data mining is to conduct an experiment to create a database that includes variation of the misalignment and distortion such as load variation and motor size. In the second step, the data is processed by data extraction, followed by the application of a feature selection algorithm.

The experiment must be adapted to the following data mining. To extract valuable features from the data, measurement of 3-phase motor current and one line-to-line voltage was performed in 42 states. The states result from the variation of the target (misalignment) and two distortions (load and motor size). Regarding misalignment four levels were chosen from zero to 1.5 times the alarm level defined by an optical alignment system producer. Regarding load, 100%, 75%, and 50% of the rated load were selected. Regarding size, a 1.1kW and a 7.5kW motor were investigated. All magnitudes were sampled with 10kHz to assure that all natural distortions are included. The time of the measurement cycle was five times the fundamental period. The number of cycles, which led to the number of feature samples, is 2000 for each state and is needed to guarantee the stability of the feature selection as examined in [4].

For the data mining, the feature extraction is followed by a feature selection. The approaches used for data mining are described in [5]. In the feature extraction step, the raw data from the experiment were processed to create physically interpretable values. This is necessary to allow discussing the results of the data mining. Since the extraction process compresses the data, as many features from different domains as possible must be calculated to avoid the loss of important information. In practice, the signals from the experiment were used to calculate the output of additional virtual sensors, in this case the spectra, the space vector represented by its length and angle components, and the spectra of the space vector. All signals were used to calculate values such as the *root mean square*, the *maximum value*, or the *signal-to-noise ratio*. Besides generally used values, values from MCSA theory were also calculated, for example the magnitude of the frequency that correlates with air gap eccentricity (ECC). All the calculated values, based on different signals, create the feature space for the feature selection. A wrapper type approach as described in [6] was chosen for the selection. The wrapper approach uses a learning model to find information rich features, which in this case was a k-nearest neighbor classifier. To ensure generalization of the classifier, it was embedded into a 10-fold cross-validation with random selection of the samples. For the data mining, the measured states were combined into groups, which differ in terms of the target being either parallel or angular misalignment. In addition, all groups contain data with a variation in load and size. A problem with feature selection is stability, which is the sensitivity of the algorithm to perturbation in the training data. In [4], the dependence of stability is examined, including data distribution. Since it is possible to find redundant features because of the input data, the data processing was repeated several times. In the first round, the full feature space was used for data mining. In the second round, previously found features were excluded, and in the last round, only MCSA features were used. The results from all turns lead to a better understanding of the affects which misalignment has on the electrical signals and avoid confusion due to redundant feature sets.

The results show that in the first round, which used the full feature space for the search, five features are needed to reach an error rate below 0.1%. In the second round, nine features are needed to reach the threshold, and in the last round, the threshold could not be reached.

Keywords: data mining; misalignment; feature extraction; feature selection.

References

1. Verma, A.K., Sarangi, S., Kolekar, M.H.: Shaft misalignment detection using stator current monitoring
2. Popaleny, P., Antonino-Daviu, J.: Electric motors condition monitoring using currents and vibrations analyses. In: 2018 XIII International Conference on Electrical Machines (ICEM), IEEE (9 2018)
3. Obaid, R., Habetler, T.: Effect of load on detecting mechanical faults in small induction motors. In: 4th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, 2003. SDEMPED 2003., IEEE (2003)
4. Alelyani, S., Liu, H., Wang, L.: The effect of the characteristics of the dataset on the selection stability. In: 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence, IEEE
5. U.R., A., Paul, S.: Feature selection and extraction in data mining. In: 2016 Online International Conference on Green Engineering and Technologies (IC-GET), IEEE
6. Kaur, A., Guleria, K., Trivedi, N.K.: Feature selection in machine learning: Methods and comparison. In: 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), IEEE