Induction Motor Fault Classification with Topological Data Analysis

Wang, Bingnan

TR2024-145 October 22, 2024

Abstract

In this paper, we apply topological data analysis (TDA) method for the processing of timedomain stator current signals of an induction motor under various fault conditions, and show that it can effectively reveal data features related to the fault condition. We show that classification accuracy of machine learning models for motor faults can be greatly improved when trained with TDA processed data, in comparison with models trained with time-domain stator current data. As a mathematical tool, TDA is effective in the development of datadriven fault detection and classification for motor applications.

IEEE Energy Conversion Congress and Exposition (ECCE) 2024

© 2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Mitsubishi Electric Research Laboratories, Inc. 201 Broadway, Cambridge, Massachusetts 02139

Induction Motor Fault Classification with Topological Data Analysis

Bingnan Wang, Senior Member, IEEE Mitsubishi Electric Research Laboratories (MERL) 201 Broadway, Cambridge, MA 02139, USA bwang@merl.com

Abstract—In this paper, we apply topological data analysis (TDA) method for the processing of time-domain stator current signals of an induction motor under various fault conditions, and show that it can effectively reveal data features related to the fault condition. We show that classification accuracy of machine learning models for motor faults can be greatly improved when trained with TDA processed data, in comparison with models trained with time-domain stator current data. As a mathematical tool, TDA is effective in the development of data-driven fault detection and classification for motor applications.

Index Terms—Electric motors, fault diagnosis, fault classification, topological data analysis

I. INTRODUCTION

Electric motors are essential components in various industries, and their healthy operation is uttermost important. Condition monitoring and condition-based maintenance of electric machines are highly desirable with the latest development of advanced sensing, data analysis and machine learning technologies. In particular, induction motors are widely used in numerous industrial applications and factories. Various faults can occur during the operation of induction motors, and the effective detection and diagnosis of these faults are essential to reduce downtime and prevent asset losses. A lot of research efforts have been devoted to understand the mechanism and corresponding consequences of each type of motor faults, and a number of fault detection mechanisms have been proposed and investigated, using signals such as noise and acoustic emission [1], vibration [2], airgap and stray magnetic flux [3], and phase current [4].

Motor current signature analysis (MCSA) method aims to detect various motor faults using stator current signal only, therefore has the advantages of simple implementation and low-cost, due to the fact that no additional sensor installations are required for this method. When fault occurs in an electric motor, the permeance function and air gap magnetic flux are often modulated periodically; some of these harmonics show up in the induced voltage in the stator windings, and are eventually reflected in the stator current spectrum. Both physics-based and data-driven approaches have been proposed for MCSA based motor fault detection. However, challenges still remain in the adoption of the method. For physics-based approaches, detailed models have been established to describe motor under fault conditions, and understand the signatures in stator current signals for each type of motor fault [5]. It has been shown that fault signatures in the stator current signals are typically much smaller than the dominating fundamental component at supply frequency. Involved frequency analysis and signal processing are typically required to identify these small fault signatures. In addition, the fault signatures in the stator current can vary for motors with different power ratings, and at different speed and loads conditions for the same machine, making it challenging to establish threshold values using those fault indicators identified through physics models. For data-driven approaches, due to the challenge of small fault signatures, many machine learning models often fail to distinguish current signals measured at healthy and faulty conditions due to the subtlety. Therefore, feature engineering and signal processing based on domain knowledge and physical models is typically needed to effectively extract fault related features in the stator current data and enable the development of effective machine learning approaches.

In recent years, topological data analysis (TDA) has been gaining popularity in a wide range of applications owing to the development of methods such as persistent homology [6]. TDA extracts the shape information for a given data space, such as connected components, holes, and other higher dimensional features [6]. Compared with other geometrical data analysis methods, TDA is coordinate-free and robust, therefore attractive for data analysis tasks ranging from material science, biology, to image and time-series data analysis [7]-[10]. In particular, persistent homology based TDA has recently been applied to MCSA based motor fault diagnosis, and shown to be very effective in extracting features in stator current signals associated with eccentricity fault and distinguishing data measured at different eccentricity levels [11], [12]. The extracted topological features are then used to develop regression models for eccentricity fault level prediction, with much improved accuracy as compared with models trained with time-domain stator current data.

In this paper, we apply TDA method to MCSA for fault classification problem, considering various induction motor faults including bearing faults, eccentricity faults, and broken bar faults. A winding function based physical model [11] is utilized for the dynamic simulation of an induction motor under faulty conditions, and the simulated time-domain stator current data is processed with the TDA method. We show that the TDA processed data can significantly improve the distinction of data between different fault conditions. We then establish fault classification models, and demonstrate that models trained on the TDA processed data can achieve much higher accuracy, as compared with models trained directly on time-domain data. Compared with conventional signal processing methods, the TDA approach does not require physical model or domain knowledge of the motor system, and needs only a very short data sequence to extract fault features and make reasonable predictions.

The rest of the paper is organized as follows: In Section II, we introduce the method for simulating faulty conditions and generating the dataset, and analyze the data; in Section III, we present the TDA process and discuss the obtained Betti sequence data; in Section IV, we present the classification models and discuss the results with TDA; in Section V, we conclude the paper and discuss future work.

II. DATA GENERATION METHOD

In this work, to evaluate the effectiveness of TDA for motor fault classification, we consider a total of seven conditions for an induction motor: healthy, static eccentricity (SE), dynamic eccentricity (DE), mixed eccentricity (ME), inner race bearing fault (IR), outer race bearing fault (OR), and broken bar fault (BB).

A dynamic model based on coupled circuits with winding function method for inductance calculation is established to simulate these faulty conditions, and the simulated stator current data are used for the subsequent fault classification study.

A. Simulation Model

The motor dynamics is described with multiple coupled circuits. Consider a three-phase squirrel-cage induction motor with rotor bar number N_b with no fault, the stator voltage $V_s = [v_{s1}, v_{s2}, v_{s3}]^{\mathsf{T}}$ and flux linkage $\Lambda_s = [\lambda_{s1}, \lambda_{s2}, \lambda_{s3}]^{\mathsf{T}}$ are modeled as

$$V_s = R_s I_s + \frac{\mathrm{d}}{\mathrm{d}t} \Lambda_s,\tag{1}$$

$$\Lambda_s = L_{ss}I_s + L_{sr}I_r,\tag{2}$$

where $I_s = [i_{s1}, i_{s2}, i_{s3}]^{\mathsf{T}}$ is the stator current, $I_r =$ $[i_{r1}, i_{r2}, ..., i_{rN_b}, i_e]^{\mathsf{T}}$ is the rotor current with dimension R+1to describe rotor loop currents and the end ring current respectively. R_s is stator resistance matrix, L_{ss} is stator inductance matrix, and L_{sr} is the mutual inductance matrix between stator phase windings and rotor loops with dimension $3 \times (N_b + 1)$.

The squirrel cage of the rotor is modeled as multiple coupled-circuits, with each bar and end ring modeled as separate elements. The rotor voltage and flux linkage are described by

$$V_r = R_r I_r + \frac{\mathrm{d}}{\mathrm{d}t} \Lambda_r,\tag{3}$$

$$\Lambda_r = L_{rs}I_s + L_{rr}I_r,\tag{4}$$

where $L_{rs} = L_{sr}^{T}$, and L_{rr} is the $(N_b + 1) \times (N_b + 1)$ self- where e_0 is the eccentricity level caused by the bearing fault, inductance matrix for rotor loops.

The torque equation of the induction motor is

$$T_e = \frac{1}{2} I_s^{\mathsf{T}} \frac{\partial L_{ss}}{\partial \theta_r} I_s + I_s^{\mathsf{T}} \frac{\partial L_{sr}}{\partial \theta_r} I_r + \frac{1}{2} I_r^{\mathsf{T}} \frac{\partial L_{rr}}{\partial \theta_r} I_r, \qquad (5)$$

where θ_r is the rotor mechanical angle.

In addition, the mechanical dynamics of the motor is described by

$$\dot{\omega}_r = \frac{1}{J} (T_e - T_L), \tag{6}$$

$$\dot{\theta}_r = \omega_r,\tag{7}$$

where ω_r is the rotation speed, T_L is load torque, and J is the rotor inertia.

The inductances between windings are calculated by a modified winding function model [13]. In general, the inductance between a pair of windings labeled i and j within a motor is calculated by

$$L_{ij}(t) = \mu_0 lr \int_0^{2\pi} n_i(\phi, t) M_j(\phi, t) g^{-1}(\phi, t) \mathrm{d}\phi, \quad (8)$$

where μ_0 is the free-space permeability, r is air gap radius, l is the stack length of the motor, $n_i(\phi, t)$ is the winding turns function for winding i, and $M_i(\phi, t)$ is the modified winding function for winding j, which is derived from

$$M(\phi,t) = n(\phi,t) - \left(M(t)\right),\tag{9}$$

where

$$\left\langle M(t)\right\rangle = \frac{1}{2\pi\langle g^{-1}(\phi,t)\rangle} \int_0^{2\pi} n(\phi,t)g^{-1}(\phi,t)\mathrm{d}\phi, \quad (10)$$
 and

$$g^{-1}(\phi,t) = \frac{1}{2\pi} \int_0^{2\pi} g^{-1}(\phi,t) \mathrm{d}\phi.$$
 (11)

Mechanical faults including eccentricity and bearing faults are described by the periodic modulation of the air gap function $q(\phi, t)$.

Under SE and DE conditions, the air gap function can be written as:

$$g(\phi, t) = g_0 K_c - \delta_{SE} g_0 \cos(\phi) - \delta_{DE} g_0 \cos(\phi - \omega_r t).$$
(12)

where g_0 is nominal air gap length, K_c is Carter's coefficient to quantify the slotting effect, δ_{SE} and δ_{DE} are the SE and DE amplitude respectively [14].

Bearing faults can also be described by periodic modulation of the air gap function. When a point defect of the bearing comes into contact with another bearing element, a radial displacement of the rotor center, or eccentricity is created. The periodic modification of the air gap length due to bearing fault induced eccentricity can be modeled by a series of Dirac delta functions. In presence of bearing fault, the air gap function of a motor at stator angle ϕ and time t is described by

$$g(\phi,t) = g_0 \left[K_c - e_0 \cos(\phi + \psi(t)) \sum_{k=-\infty}^{+\infty} \delta\left(t - \frac{k}{f_c}\right) \right], \quad (13)$$

and $\psi(t)$ is the defect position angle at time t. In practical

implementation, rectangular shaped pulses with finite width are used instead of the pulses [15], [16]. The time-dependence of the fault position depends on the fault type. For outer race fault, the outer race stays in place at all time, and the fault position is fixed, $\psi(t) = 0$ without loss of generality. For inner race fault, the fault rotates along with the inner race and rotor, so that $\psi(t) = 2\pi f_r t$.

Denote the resistance of each rotor bar as r_b , and the resistance of the end winding piece connecting two neighboring rotor bars as r_e , the rotor resistance matrix can be written as equation (14).

In case of a broken bar fault, the corresponding bar is removed from the circuit system, and neighboring loops are combined as one new loop. The coupled-circuit equations update subsequently [17]. The dimension of rotor resistance and inductance matrices is reduced from $(N_b + 1) \times (N_b + 1)$ to $N_b \times N_b$, and the updated rotor resistance matrix with one broken bar is written in equation (15). The inductance terms also need to be updated accordingly. Broken bar fault is reflected in the stator current spectrum at side bands around the supply frequency:

$$f_{BB} = (1 \pm 2ks)f_s, \tag{16}$$

where f_s is the supply frequency, s is the slip of the induction motor, and k is an integer.

B. Simulated Data

A 3-phase 4-pole squirrel cage induction machine is used in the study. The stator has 36 slots, with 37 winding turns in each slot. The air gap length $g_0 = 0.28$ mm, and the air gap radius r = 41.6 mm. The squirrel cage has 28 bars, and the stack length is l = 80 mm. The Carter's coefficient $K_c = 1.40$. Supply frequency is 60 Hz.

Simulation is performed with the method described in the previous subsection for each faulty case under no-load condition, and the stator current data is recorded at 10 kHz sampling frequency for a length of 30 seconds.

To simulate bearing fault conditions, we assume that two 6022-ZZ bearings are mounted on the load side and the opposite side respectively, with each bearing has 8 balls. The inner race defect characteristic frequency for these bearings is approximately $f_i \approx 0.6nf_r = 4.8f_r$, while an outer race defect gives characteristic frequency of $f_o \approx 0.4nf_r = 3.2f_r$.

Fig. 1 shows an example of simulated three-phase stator current data for the motor without fault. Fig. 2 compares the simulated phase A stator current data under different fault conditions. While the signal without fault is generally smooth sinusoidal, signals under faulty conditions show more small features. However, it is not straightforward to tell these timedomain signals apart.

We examine the frequency spectrum of the obtained stator current to make sure the simulation model generates reasonable signals. In Fig. 3 we compare the frequency spectrum of phase A current under inner race and outer race bearing faults. Different characteristic frequencies corresponding to the respective bearing faults are seen in the two conditions. Fig. 4 plots the frequency spectrum around the supply frequency under broken bar condition, which clearly shows the side bands corresponding to the the broken bar fault frequencies listed in equation (16). These results show that the simulation model is reasonable in generating stator current signals under these fault conditions.

III. TDA PROCESS

Typically, MCSA based motor fault detection requires detailed analysis of the current spectrum with domain knowledge to identify frequency signatures of each fault. In this work, we utilize TDA method to process the obtained stator current data and reveal the features in the data space that are related to motor faults, without the need of domain knowledge or physical models.

For each investigated faulty condition, we take a segment of 1024 consecutive data points from the simulated time-domain stator current data, which corresponds to roughly 0.1 s. The three-phase data segment forms a point cloud in 3D Euclidean space, with the dominating shape a large circle corresponds to the sinusoidal waveform at supply frequency. We perform TDA following the process established in Ref. [12] to calculate the persistent homology of each data segment, and then obtain the H_0 and H_1 Betti sequences, which corresponds to connected components and holes in the data space respectively.

The obtained Betti sequences corresponding to data samples from each fault condition are plotted in Fig. 5. Note that the x-axis of the curves indicates the filtering radius. H_0 Betti sequence describes how the connected components in the point cloud formed by the three-phase current data evolves with the filtering radius. We can generally understand that motor faults introduces additional components in the stator current data, causing the neighboring points in the point cloud formed by the three-phase current data to be further apart, which alters the line shape of the corresponding Betti curve. H_1 Betti sequence describes how the holes formed in the point cloud evolves with the filtering radius. Without fault, the point cloud is almost an ideal circle in three dimensional space, and there are very few H_1 features. When fault occurs, localized holes form and merge in the point cloud at various filtering radii, introducing additional features in the H_1 Betti sequence. While it is not straightforward to interpret the shapes of each curve in details accordingly to the motor fault, we do notice that they are more separable than the original time-domain data as shown in Fig. 2.

The separability of the Betti sequences under different fault conditions is more obvious when we process all the obtained data samples with TDA, and visualize them in the t-SNE plot. As shown in Fig. 6(a), the time-domain data are essentially indistinguishable, as they are all mixed together. On the other hand, As shown in Fig. 6(b), the combined H_0 and H_1 Betti sequence data largely clusters according to fault condition, indicating that the data are much more linearly separable than the original time-domain data.





Fig. 1. Simulated three-phase stator current for the inductor motor without fault.

With the Betti sequences obtained from TDA, we can design and train fault classification models to effectively detect and classify motor faults for given stator current data.

 TABLE I

 MOTOR FAULT CONDITIONS AND CORRESPONDING LABEL

Label	0		1	2	3	4	5	6
Condition	Healthy		SE	DE	ME	IR	OR	BB

IV. FAULT CLASSIFICATION RESULTS

With the model established for each fault condition, dynamic simulations are performed to obtain the corresponding time-domain (TD) stator current signals. For machine learning purposes, the obtained stator current signal is segmented into samples of length 1,024 and labelled with corresponding fault condition as indicated in Table I. For comparison, the timedomain data samples are processed with the TDA method, and a new dataset is assembled with the H_0 and H_1 Betti sequences. For convenience, all H_0 and H_1 Betti sequences



Fig. 2. Comparison of phase A stator current data under different fault conditions. Each curve is shifted vertically by 2 A from previous one for better visibility.

are kept the same length of 1,024 as their time-domain data counterparts.

The dataset is then shuffled and partitioned into training and test set with a split ratio of 80:20. Commonly used machine learning classification models, including decision tree, multi-layer perceptron (MLP), Naive Bayes, nearest neighbors, random forest, and linear support vector machine (SVM), are trained on the training dataset and the performance is evaluated on the test dataset according to the classification accuracy. Hyperparameters for each classification model are kept the same for all input data types.

The results for models trained on TD, H_0 , H_1 and combined $H_0 + H_1$ data respectively are compared side-by-side in Table II. For models trained on TD data, they all fail to predict the correct fault condition, and the performances are



Fig. 3. Phase A current spectrum under bearing fault conditions.



Fig. 4. Phase A current spectrum under broken bar fault condition.



Fig. 5. H_0 and H_1 Betti curves for induction motor phase current signals under different fault conditions.

no better than random guess. When trained with H_0 data, the classification accuracy for all models is improved to around 90%. Models trained on H_1 data also yield prediction accuracy above 80%. When using both H_0 and H_1 Betti sequence data for training, the overall accuracy of models trained (last column of the table) is further improved to over 90%, with the highest accuracy of 96.9% achieved by linear SVM classifier.

We further evaluate the model prediction performance by examining the confusion matrix between true label and predicted label on the test dataset for the best performing classification models for TD data (random forest model, with



Fig. 6. t-SNE plot in 2d for (a) time-domain phase current data, and (b) H_0 and H_1 sequences combined for different fault conditions.

TABLE II CLASSIFICATION ACCURACY OF MODELS TRAINED ON TIME-DOMAIN DATA AND TDA PROCESSED DATA FOR MOTOR FAULT CONDITIONS CORRESPONDING TO TABLE I

	Input Data Type								
Classifier Model	TD	$ $ H_0	$ $ H_1	$H_0 + H_1$					
Decison Tree	17.2%	85.8%	80.2%	91.2%					
Multi-layer Perceptron	13.5%	86.9%	88.3%	95.3%					
Naive Bayes	12.1%	89.9%	86.4%	95.6%					
Nearest Neighbors	11.9%	84.4%	86.8%	94.1%					
Random Forest	18.6%	85.3%	81.7%	92.8%					
Support Vector Machine	11.5%	89.9%	89.6%	96.9%					

prediction accuracy of 18.6%) and Betti sequence data (linear SVM, with prediction accuracy 96.6%), which are shown in Fig. 7. For the random forest model trained on TD data, the prediction is pretty much random. For the linear SVM model trained on H_0 and H_1 Betti sequence data, the model gives correct prediction on fault condition for most test data samples.

These results show the effectiveness of TDA process to extract fault related features in stator current signals, without



Fig. 7. Confusion matrix of the best performing classifiers for (a) time-domain data, and (b) Betti sequence data.

the need of spectral analysis and physical model. Simple classification models that are easy to train and deploy can be build to make accurate predictions to motor fault conditions with the TDA processed data.

V. CONCLUSIONS

In this paper, we evaluated topological data analysis method for stator current data based induction motor fault classification. With coupled circuit model and winding function approach, we simulated the motor dynamics, and obtain the stator current data for each faulty condition. We then processed the segmented time-domain signals with TDA method, and trained machine learning models for fault classification. We showed that the data features related to the fault condition can be effectively extracted with the method, and the classification accuracy of trained machine learning models can be greatly improved, as compared with models trained with time-domain stator current data. The method requires neither physical knowledge in fault signatures for each case, nor parameters for motors or bearings. In the next step, we will conduct experiments to obtain data under various faulty conditions, and apply the method to the experiment data to evaluate its effectiveness on real data.

ACKNOLEDGEMENT

The authors thank Chungwei Lin, Lei Zhou, and AKM Khaled Ahsan Talukder for helpful discussions.

REFERENCES

- [1] O. AlShorman, F. Alkahatni, M. Masadeh, M. Irfan, A. Glowacz, F. Althobiani, J. Kozik, and W. Glowacz, "Sounds and acoustic emission-based early fault diagnosis of induction motor: A review study," *Advances in Mechanical Engineering*, vol. 13, no. 2, p. 1687814021996915, 2021. [Online]. Available: https://doi.org/10.1177/1687814021996915
- [2] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics – a comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020.
- [3] E. Mazaheri-Tehrani and J. Faiz, "Airgap and stray magnetic flux monitoring techniques for fault diagnosis of electrical machines: An overview," *IET Electric Power Applications*, vol. 16, no. 3, pp. 277–299, 2022. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/elp2.12157
- [4] M. E. H. Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE transactions on industrial electronics*, vol. 47, no. 5, pp. 984–993, 2000.
- [5] W. T. Thomson and I. Culbert, Current signature analysis for condition monitoring of cage induction motors: Industrial application and case histories. John Wiley & Sons, 2017.
- [6] C. Gunnar, "Topology and data," Bulletin of the American Mathematical Society, vol. 46, no. 2, pp. 255–308, 2009.
- [7] Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escolar, K. Matsue, and Y. Nishiura, "Hierarchical structures of amorphous solids characterized by persistent homology," *Proceedings of the National Academy of Sciences*, vol. 113, no. 26, pp. 7035–7040, 2016. [Online]. Available: https://www.pnas.org/content/113/26/7035
- [8] Y. Lee, S. D. Barthel, P. Dłotko, S. M. Moosavi, K. Hess, and B. Smit, "Quantifying similarity of pore-geometry in nanoporous materials," *Nature communications*, vol. 8, no. 1, pp. 1–8, 2017.
- [9] T. Qaiser, Y.-W. Tsang, D. Taniyama, N. Sakamoto, K. Nakane, D. Epstein, and N. Rajpoot, "Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features," *Medical image analysis*, vol. 55, pp. 1–14, 2019.
- [10] F. A. Khasawneh and E. Munch, "Chatter detection in turning using persistent homology," *Mechanical Systems and Signal Processing*, vol. 70, pp. 527–541, 2016.
- [11] B. Wang, H. Inoue, and M. Kanemaru, "Motor eccentricity fault detection: Physics-based and data-driven approaches," in 2023 IEEE 14th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED), 2023, pp. 42–48.
- [12] B. Wang, C. Lin, H. Inoue, and M. Kanemaru, "Induction motor eccentricity fault detection and quantification using topological data analysis," *IEEE Access*, vol. 12, pp. 37 891–37 902, 2024.
- [13] N. Al-Nuaim and H. Toliyat, "A novel method for modeling dynamic air-gap eccentricity in synchronous machines based on modified winding function theory," *IEEE Transactions on Energy Conversion*, vol. 13, no. 2, pp. 156–162, 1998.
- [14] B. Wang, M. W. Albader, H. Inoue, and M. Kanemaru, "Induction motor eccentricity fault analysis and quantification with modified winding function based model," in 2022 25th International Conference on Electrical Machines and Systems (ICEMS), 2022, pp. 1–6.
- [15] P. McFadden and J. Smith, "Model for the vibration produced by a single point defect in a rolling element bearing," *Journal of Sound and Vibration*, vol. 96, no. 1, pp. 69–82, 1984.
- [16] N. Tandon and A. Choudhury, "An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect," *Journal of Sound and Vibration*, vol. 205, no. 3, pp. 275–292, 1997.
- [17] F. Karami, J. Poshtan, and M. Poshtan, "Broken bar fault detection in induction motors based on modified winding function," in 2010 IEEE International Conference on Control Applications, 2010, pp. 1951–1956.