UDC 519.651
DOI: 10.15587/1729-4061.2024.312610

IDENTIFICATION OF THE ELECTRIC MOTOR MATHEMATICAL MODEL BASED ON A DATA SAMPLE WITH FEATURE ENGINEERING

Anton Korotynskyi
Corresponding author
PhD, Senior Lecturer*
E-mail: ihfantkor@gmail.com
Liudmyla Zhuchenko
PhD, Assistant*
Vitalii Tsapar
PhD, Associate Professor*
Andrii Savula
PhD Student*
*Department of Technical and Software Automation
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
Beresteiskyi ave., 37, Kyiv, Ukraine, 03056

The object of this study is a mathematical model of a synchronous electric motor, obtained on the basis of experimental data, which takes into account the temperature mode and uses artificial features to increase the accuracy of its operation. A characteristic feature of this work is that the model takes into account the temperature mode as a component of the technical-operational state of the object. The resulting mathematical model could make it possible to synthesize an optimal automatic control system in terms of the operational state of the object.

The problem addressed was to increase the accuracy of the identified mathematical models by applying the approach of feature engineering.

The results showed that the identification of mathematical models by the initial data leads to a low level of accuracy of the obtained models, namely 65–70 % for the first output channel, 80–85 % for the second, and 75–80 % for the third, fourth, and fifth output channels.

Accordingly, building models with a higher threshold of accuracy requires the use of other, more significant data for identification. This paper reports a method for reformatting the original data into artificial features and provides results of their effectiveness in relation to the original channels.

The resulting artificial features and the original features were used for further identification; the resulting mathematical model has on average higher accuracy thresholds, namely 82 %, 93 %, 88 %, 85 % for the corresponding output channels. The results prove the effectiveness of applying the principle of feature engineering since the accuracy of the resulting model is 5–10 % higher compared to the baseline.

The scope of practical application of the results includes the synthesis of automatic control systems based on mathematical models of control objects obtained as a result of identification

Keywords: mathematical model of a synchronous electric motor, mathematical model identification, mutual information, correlation analysis of electric motor operating parameters, artificial feature engineering

Received date 24.06.2024
Accepted date 27.09.2024
Published date 25.10.2024

How to Cite: *Korotynskyi, A., Zhuchenko, L., Tsapar, V., Savula, A. (2024). Identification of the electric motor mathematical model based on a data sample with feature engineering. Eastern-European Journal of Enterprise Technologies, 5 (1 (131)), 91–98. https://doi.org/10.15587/1729-4061.2024.312610*

1. Introduction

The operation of an electric motor is determined by the electricity required for its functioning, so the task of finding effective control over the electric motor to achieve its optimal performance is obvious. However, when designing an optimal electric motor control system, it is necessary to take into account its current technical and operational condition. Due to the peculiarity of the structure of the electric motor, one of these characteristics is the temperature of the rotor, which cannot be directly measured during its operation with the help of thermal sensors.

To solve this problem, first of all, it is necessary to build a mathematical model of electric motor operation taking into account its technical and operational condition, which can be carried out by identifying the mathematical model with the current temperature of the rotor based on the experimental data from laboratory studies. However, the resulting mathematical model according to technical and operational indicators should meet the specified accuracy, which would allow it to be used in the development of optimal automated control systems.

Conducting scientific research in the field of identification of the technical and operational condition of electric motors is of practical importance, as such research allows for the development of more accurate models that improve the operation of automated control systems. Such research contributes to increasing the energy efficiency, reliability, and durability of electric motors, which has a significant impact on the economy and safety of industrial processes. Therefore, studies aimed at the identification of the technical and operational condition of electric motors are relevant.

2. Literature review and problem statement

Some works are aimed at the correct understanding of data reflecting the operating states of engines and can be used for early identification of potential failures [1]. But it is not always possible to measure these data in real time, due to the complex structure of the object. An option to overcome the difficulties may involve the construction of a model for calculating non-measurable data from measurable data – forecasting based on mathematical models.

Paper [2] describes methods for identifying mathematical models of asynchronous electric motors, presents an overview and analysis of the main methods for determining the parameters of an asynchronous machine. It is important to add that despite the review of the existing methods for identifying the parameters of asynchronous machines, the main unsolved problem in the work is insufficient accuracy of the models under difficult operating conditions, which requires further research and optimization of algorithms capable of taking into account the variable modes of operation of the engine.

The authors of work [3] reported a separate identification of electrical and mechanical parameters. A new algorithm for identifying the electrical parameters of an asynchronous motor is proposed, obtained on the basis of the analysis of the mathematical model of the machine at a constant rotation frequency. But the issue related to determining the technical and operational condition of the engine during its operation remained unresolved.

Artificial intelligence-based fault diagnosis methods have become widespread in recent years and have been successful in many applications for electric machines and drives [4, 5]. Objective difficulties associated with the development of these methods are the need to have a sufficiently large dataset for training artificial neural networks. Works [4, 5] indicate the difficulty of extracting informative features from available data, which is a critical problem for the accuracy of models. The insufficient amount of high-quality data limits the effectiveness of both neural networks and traditional signal analysis methods.

An option to solve this problem can be the synthesis of artificial data; however, this approach could lead to a change in the characteristics of the input data, which would both positively and negatively affect the work results.

For example, a two-stage learning method involving sparse filtering and a neural network was proposed to form an intelligent fault diagnosis method for learning features from raw signals [6]. It is important to note that although the two-step method described in the paper has demonstrated effectiveness in learning features from raw signals, the main problem is the difficulty of adapting this method for different operating conditions and the need for a large amount of well-labeled data.

A neural network using the Levenberg-Marquardt algorithm demonstrated a new way of detecting and diagnosing faults in asynchronous machines [7], in which the results were not influenced by load conditions and types of faults. Although the work demonstrated the effectiveness of neural networks for fault diagnosis, one of the main problems remains the sensitivity of the models to changes in the input data. Networks can show good results only under certain operating conditions.

Although all of these studies have demonstrated the benefits of above approaches for determining engine performance, most of these approaches are based either on training, which requires high-quality training data, or on the analysis of existing data. However, simply getting enough data would not suffice. Many tasks of determining the technical and operational condition depend on the extraction of features from the measured signals.

In the current literature, many feature extraction methods are suitable for fault diagnosis tasks, such as statistical analysis in the time domain, spectral analysis in the frequency domain [8, 9]. They are a powerful tool specifically for feature extraction, while the problem is the limited number of features that can be used. Available studies use feature extraction methods such as spectral analysis and time domain statistical analysis. However, as indicated in the works, these methods have limitations in the number of features that can be extracted from signals, especially under conditions of incomplete or insufficiently informative data. This limitation significantly affects the overall accuracy of fault diagnosis under real operating conditions.

As a result of the analysis, it was established that most of the existing methods for diagnosing faults or determining the technical and operational condition of electric motors have limitations. These limitations are related to the insufficient amount or quality of data, as well as the difficulty of extracting relevant features. Therefore, an unsolved scientific problem is the insufficient amount of high-quality informative data that could be used both in the training of artificial neural networks and in the construction of models for the calculation of immeasurable data. This allows us to state that it is appropriate to conduct a study aimed at the development of methods for artificial feature engineering to increase the accuracy of the informativeness of data. That, in turn, will allow us to build more accurate models based on experimental data and improve their application under actual operating conditions.

3. The aim and objectives of the study

The purpose of our work is to determine the possibility of applying the approach of artificial feature engineering to build a more accurate mathematical model based on artificially generated data. This will make it possible to identify the mathematical model of the electric motor based on the results of experimental studies, respectively, with high accuracy and the possibility of simulating the temperature of the stator and rotor in real time.

To achieve the goal, the following tasks must be solved:

– to conduct a preliminary analysis of the collected data to identify the main regularities and dependences between input and output parameters;

– to use identification methods to build a basic mathematical model of an electric motor based on the collected experimental data, evaluate the accuracy of the resulting model;

– to devise a method of artificial feature engineering to expand the sample of input data, to conduct a preliminary analysis of artificial features to identify the main regularities and dependences between input and output parameters;

– to use artificially formed features for re-identification, evaluate the degree of accuracy of the resulting model compared to the base model, determine the effectiveness of the applied approach.

4. The study materials and methods

4. 1. The object and hypothesis of the study

The object of our study is an electric motor, namely its mathematical model, which was built by identification based on experimental data.

The main hypothesis of the study assumes that the use of the method of artificial feature engineering could make it possible to significantly expand the data sample with new features. This, in turn, after filtering the generated features, would lead to an increase in the accuracy of the identified mathematical model.

Assumptions adopted in the study temperature indicators of the stator and rotor can be used as indicators of the technical and operational condition of the engine.

Simplifications accepted in the study:

– the influence of noise in the data is considered to be minimal and has little effect on the identification results;

– all measured parameters are considered to be linearly dependent, which allows the use of correlation analysis, or non-linearly dependent, which in turn allows the use of mutual information analysis.

4. 2. Description of initial experimental data

As a result of conducting a field study, experimental data on the operation of a synchronous motor with permanent magnets, installed on a test bench, were obtained. The data set consists of several measurement sessions [10, 11].

Description of basic data:

– ambient – ambient temperature measured by a thermal sensor located near the stator;

– coolant – the temperature of the coolant (the engine is cooled by water; the measurement is performed at the outlet);

– u_d – D-component of voltage;

– u_q – Q-component of voltage;

– motor_speed – motor speed;

– torque – rotating moment;

– i_d – d-component of the current;

– i_q – q-component of the current;

– the surface temperature of the permanent magnet, which represents the temperature of the rotor, was measured using an infrared temperature sensor;

– stator_yoke – stator yoke temperature;

– stator_tooth – stator tooth temperature;

– stator_winding – stator winding temperature.

A fragment of the data obtained as a result of the experiments is given in Table 1.

As input parameters of the mathematical model of the electric motor, the following are considered: the temperature of the coolant and the environment, D and Q voltage components; d and q components of the current. As the initial parameters of the model, the following are considered: engine speed and torque, temperature of the yoke, tooth, and stator winding. Therefore, the structure of the proposed model will have 6 inputs and 5 outputs.

Table 1

A fragment of data obtained as a result of field studies [10]

u_q	coolant	stator_winding	u_d	stator_tooth	motor_speed
–0.4507	18.8052	19.0867	–0.3501	18.2932	0.0029
–0.3257	18.8186	19.0924	–0.3058	18.2948	0.0003
–0.4409	18.8288	19.0894	–0.3725	18.2941	0.0024
–0.327	18.8356	19.083	–0.3162	18.2925	0.0061
–0.4712	18.857	19.0825	–0.3323	18.2914	0.0031
–0.539	18.9015	19.0771	0.0091	18.2906	0.0096
i_d	i_q	pm	stator_yoke	ambient	torque
0.0044	0.0003	24.5542	18.3165	19.8507	0.1871
0.0006	–0.0008	24.5381	18.315	19.8507	0.2454
0.0013	0.0004	24.5447	18.3263	19.8507	0.1766
0	0.002	24.554	18.3308	19.8506	0.2383
–0.0643	0.0372	24.5654	18.3267	19.8506	0.2082
–0.6136	0.3367	24.5736	18.3239	19.8506	0.4762

4. 3. Data analysis procedures

Processing of initial experimental data was carried out by constructing a correlation matrix and a matrix of mutual information. That made it possible to determine the nature of the relationship between input and output parameters. Both linear (correlation) and non-linear (mutual information) methods were used for the analysis, which provided a deeper understanding of interactions between parameters.

4. 4. Identification of mathematical models

Identification of the basic mathematical model of the electric motor was carried out on the basis of collected experimental data. To this end, MIMO (multiple-input multiple-output) identification methods of objects in the Simulink environment, such as Subspace, Prediction Error methods, were used. The model included 6 input and 5 output parameters.

4. 5. Artificial feature engineering

To increase the accuracy of mathematical models, the method of artificial feature engineering was used. The input parameters of the model were expanded through artificially created features, such as the difference between parameters, their product, quotient, square, logarithm, and square root. That made it possible to increase the informativeness of the sample of input data and improve the accuracy of the models.

5. Results of research on the use of artificially formed features in the identification of mathematical models

5. 1. Results of preliminary analysis of experimental data

A preliminary analysis of the output data was carried out to determine the nature of the relationship between the input and output values. To that end, a correlation matrix of model parameters and a matrix of mutual information were constructed.

In this study, the use of both specified matrices makes it possible to get a more complete picture of the relationships between the parameters of the electric motor:

1. The correlation matrix helps identify the main linear relationships between input and output parameters.

2. The matrix of mutual information makes it possible to identify both linear and non-linear relationships.

Therefore, the combination of these two methods of analysis provides a deeper understanding of the interactions between parameters, which contributes to the construction of a more accurate and reliable mathematical model of the electric motor.

Fig. 1 shows the correlation matrix of the initial experimental data.

The correlation matrix demonstrates a significant linear dependence of D and Q voltage components and d and q current components on engine speed and torque. The temperature of the coolant and the environment has a significant influence on the temperature of the yoke, tooth, and stator winding.

Fig. 2 shows the mutual information matrix of the initial experimental data.

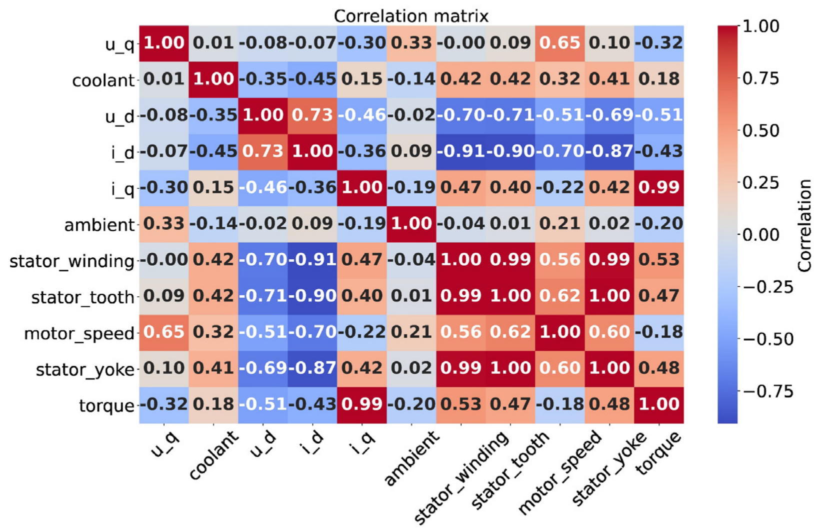

Fig. 1. Correlation matrix of input data

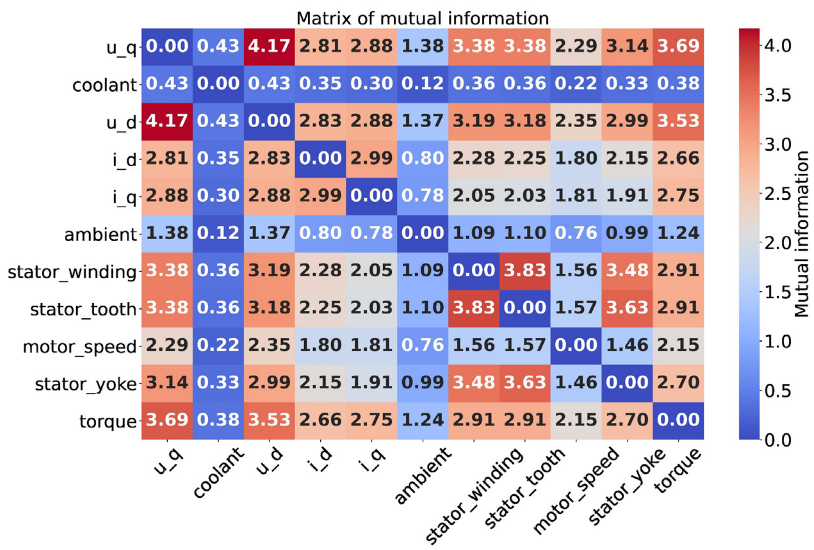

Fig. 2. Matrix of mutual information

The matrix of mutual information also demonstrates a significant dependence of D and Q voltage components and d and q current components, the temperature of the coolant and the environment on the engine speed and its torque. The temperature of the yoke, tooth, and stator winding is significantly influenced by the environment, D and Q components of the voltage.

5. 2. Identification results based on the base sample

Mathematical models of different structure and type obtained as a result of identification show on average 65–70 % accuracy on the first output, 80–85 % on the second output, and 75–80 % accuracy on the third, fourth, fifth output, Fig. 3.

The general structure of the identified model based on the base sample:

$$x(t+Ts)=Ax(t)+Bu(t)+Ke(t),$$

$$y(t)=Cx(t)+Du(t)+e(t),$$

where

$$A=\begin{array}{cccccc} & x1 & x2 & x3 & x4 & x5 \\ x1 & 0.9998 & -8.976e\text{-}05 & -6.038e\text{-}05 & 0.0001387 & 0.000113 \\ x2 & -0.0001159 & 0.9979 & -0.000394 & 7.185e-05 & -0.0002946 \\ x3 & -0.0001743 & 0.0006817 & 0.9986 & -0.002534 & 0.0007535 \\ x4 & -0.001283 & 0.001122 & -0.001419 & 0.9946 & -0.0004948 \\ x5 & -0.001474 & 0.0004323 & -0.000869 & 0.001267 & 0.9937 \end{array},$$

$$B=\begin{array}{ccccccc} & u1 & u2 & u3 & u4 & u5 & u6 \\ x1 & 7.082e\text{-}08 & -2.268e\text{-}07 & 8.959e\text{-}08 & -7.127e\text{-}08 & -7.58e\text{-}08 & -7.006e\text{-}08 \\ x2 & -2.577e\text{-}08 & -3.682e\text{-}06 & 1.599e\text{-}08 & -1.1176e\text{-}07 & -3.355e\text{-}07 & 7.224e\text{-}07 \\ x3 & 4.806e\text{-}07 & -2.692e\text{-}06 & -7.351e\text{-}08 & 4.266e\text{-}07 & 1.899e\text{-}06 & -2.181e\text{-}06 \\ x4 & 1.504e\text{-}06 & 7.46e\text{-}06 & -9.402e\text{-}08 & -1.588e\text{-}06 & 4.764e\text{-}06 & -1.138e\text{-}05 \\ x5 & -9.7e\text{-}07 & -3.652e\text{-}06 & 1.982e\text{-}07 & 9.264e\text{-}07 & -3.081e\text{-}06 & 2.374e\text{-}05 \end{array},$$

$$C=\begin{array}{cccccc} & x1 & x2 & x3 & x4 & x5 \\ y1 & -118.8 & 220.8 & -805.9 & -246.7 & 142.9 \\ y2 & -209.7 & 88.29 & -255.7 & -0.3889 & 274.5 \\ y3 & -1.682e+05 & 1.151e+04 & -7210 & 4175 & -2541 \\ y4 & -180.8 & -592.7 & -85.55 & -32.55 & 20.51 \\ y5 & 1129 & -103.6 & -630.2 & 428.7 & -79.36 \end{array},$$

$$D=\begin{array}{c|cccccc} & u1 & u2 & u3 & u4 & u5 & u6 \\ \hline y1 & 0 & 0 & 0 & 0 & 0 & 0 \\ y2 & 0 & 0 & 0 & 0 & 0 & 0 \\ y3 & 0 & 0 & 0 & 0 & 0 & 0 \\ y4 & 0 & 0 & 0 & 0 & 0 & 0 \\ y5 & 0 & 0 & 0 & 0 & 0 & 0 \end{array},$$

$$K=\begin{array}{c|ccccc} & y1 & y2 & y3 & y4 & y5 \\ \hline x1 & 3.007e\text{-}05 & -7.36e\text{-}05 & -6.842e\text{-}06 & -9.636e\text{-}05 & 6.939e\text{-}05 \\ x2 & 0.0002437 & -7.324e\text{-}05 & -1.639e\text{-}06 & -0.001664 & -3.828e\text{-}06 \\ x3 & -0.007599 & -0.0001877 & -1.04e\text{-}05 & -0.0002127 & -0.0002149 \\ x4 & -0.000926 & -6.495e\text{-}05 & -3.181e\text{-}05 & -0.0002982 & 0.001439 \\ x5 & 0.0006994 & 0.001514 & -6.664e\text{-}05 & 0.0003792 & 0.0005905 \end{array}.$$

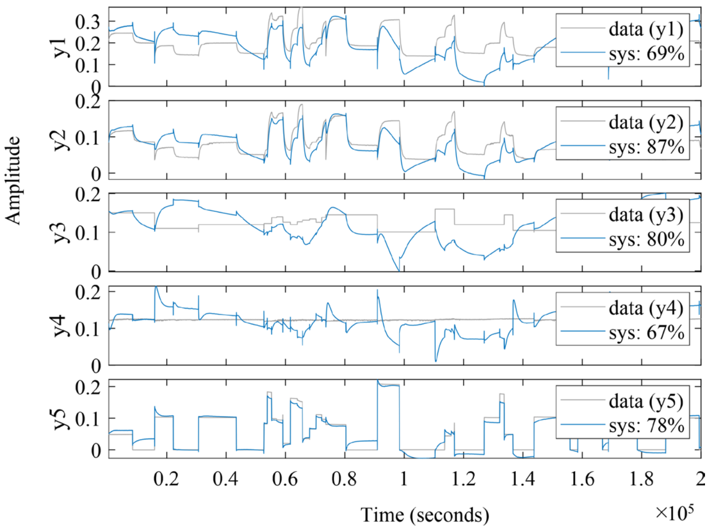

Fig. 3. Results of identification of the mathematical model of the electric motor

The resulting mathematical models do not meet the accuracy conditions proposed for models of this type. For example, when developing an automated control system for the process of baking carbon articles, the accuracy of the mathematical model, and therefore the correctness of the control action, directly affects the quality indicators of the process, namely the number of defective products [11].

5. 3. Artificial feature engineering

That is why there is a need to artificially construct the input parameters of the model by distorting the initial experimental data to obtain a greater relationship between the input characteristics and the output of the model.

To increase the informativeness of the initial data and improve the accuracy of the mathematical model, artificial feature engineering was applied. The essence of this method is that, based on existing input data, new features were obtained that reflect various mathematical relationships between the input parameters:

– diff (difference) – the difference between two input parameters. This feature reflects the change between two parameters, which may indicate important dependences between them;

– mult (product) – the product of two parameters. The product makes it possible to see the relationships between the parameters, especially when their simultaneous change may have a cumulative effect;

– div (part) – ratio of one parameter to another. The ratio makes it possible to measure the influence of one parameter on another and display their proportional dependence;

– exp (exponential) – exponential function of one of the parameters. The use of an exponential function can represent non-linear dependences between parameters that are important for dynamic systems;

– square – raising the parameter to a square. The square of the parameter helps reveal the effect of changing the parameter on the nonlinear behavior of the system;

– log (logarithm) – the natural logarithm of the parameter. The logarithmic function is useful for detecting dependences that have properties of diminishing returns or effects;

– sqrt (square root) – the square root of the parameter. The square root allows a better understanding of the interrelationships of parameters that have weak non-linear effects.

The paper proposed the following mathematical formulae (Table 2), according to which the original value was distorted, resulting in the formation of artificial features.

Table 2

Artificial features

Feature ID	Feature formula
diff	x_1-x_2
mult	x_1*x_2
div	x_1/x_2
epx	exp(x_1)
square	x_1^2
log	log(x_1)
sqrt	$x_1^{1/2}$

Note: x_1, x_2 – parameters of the input experimental sample

The best examples of artificially formed features and their correlation with the original initial features sorted by the correlation threshold value of 0.6 are shown in Fig. 4.

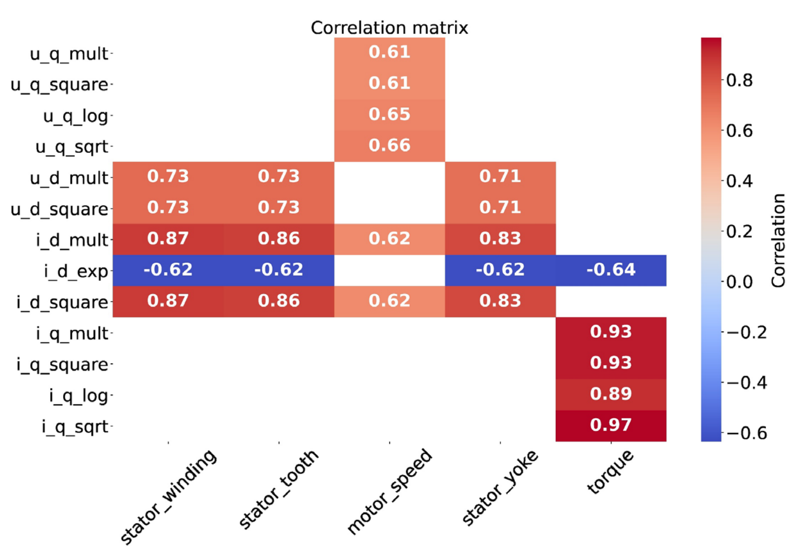

Fig. 4. Correlation matrix of artificially formed values

Applying additional features to the initial sample will increase its informativeness in accordance with the original features.

5. 4. Identification results on a sample with artificially formed features

The identification of the mathematical model was carried out based on the most correlated initial and artificially formed features. To this end, the method of identifying MIMO objects in the Simulink environment was used. As a result of model identification, accuracy within 80–95 % was obtained for different channels; the obtained identification results are shown in Fig. 5.

The general structure of the identified model based on the extended sample:

$$x(t+Ts)=Ax(t)+Bu(t)+Ke(t),$$

$$y(t)=Cx(t)+Du(t)+e(t),$$

where

$$A=\begin{array}{cccccc} & x1 & x2 & x3 & x4 & x5 \\ x1 & 0.9997 & -2.754e\text{-}05 & -0.0001484 & -1.19e\text{-}05 & 0.0001989 \\ x2 & -0.0003198 & 0.9965 & 0.0002705 & -0.000548 & -0.0004585 \\ x3 & -0.0001976 & 0.0003447 & 0.9973 & -0.003781 & 0.0008617 \\ x4 & -0.001238 & -0.0007357 & -0.00105 & 0.9933 & -0.0005935 \\ x5 & -0.004378 & -0.001583 & -0.0004495 & 0.001402 & 0.9938 \end{array},$$

$$B=\begin{array}{cccccc} & x1 & x2 & x3 & x4 & x5 \\ u1 & 0.004262 & 0.01108 & -0.002396 & 0.03761 & 0.02634 \\ u2 & -0.003392 & -0.009198 & 0.002096 & -0.03104 & -0.02152 \\ u3 & -0.001497 & -0.004254 & 0.001001 & -0.01487 & -0.009267 \\ u4 & 1.158e\text{-}06 & 4.632e\text{-}06 & -1.163e\text{-}06 & 1.669e\text{-}05 & 8.924e\text{-}06 \\ u5 & 1.725e\text{-}06 & 7.014e\text{-}06 & -2.122e\text{-}06 & 2.478e\text{-}05 & 1.356e\text{-}05 \\ u6 & -9.109e\text{-}06 & -3.751e\text{-}05 & 1.324e\text{-}05 & -0.0001313 & -7.173e\text{-}05 \\ u7 & 1.244e\text{-}06 & 5.882e\text{-}06 & 1.401e\text{-}06 & 2.155e\text{-}05 & 1.082e\text{-}05 \\ u8 & 1.992e\text{-}06 & 8.781e\text{-}06 & -2.463e\text{-}06 & 3.233e\text{-}05 & 1.603e\text{-}05 \\ u9 & 3.149e\text{-}06 & 1.341e\text{-}05 & -3.626e\text{-}06 & 4.902e\text{-}05 & 2.495e\text{-}05 \\ u10 & -1.424e\text{-}06 & -6.399e\text{-}06 & 1.505e\text{-}06 & -2.424e\text{-}05 & -1.139e\text{-}05 \\ u11 & -1.226e\text{-}05 & -5.092e\text{-}05 & 1.686e\text{-}06 & -0.0001803 & -9.669e\text{-}05 \\ u12 & -1.906e\text{-}06 & -7.531e\text{-}06 & 2.226e\text{-}06 & -2.746e\text{-}05 & -1.413e\text{-}05 \\ u13 & -1.158e\text{-}06 & -4.632e\text{-}06 & 1.163e\text{-}06 & -1.669e\text{-}05 & -8.924e\text{-}06 \\ u14 & -0.01444 & -0.03985 & 0.001629 & -0.1272 & -0.1015 \\ u15 & 1.226e\text{-}05 & 5.092e\text{-}05 & -1.686e\text{-}05 & 0.0001803 & 9.669e\text{-}05 \\ u16 & -1.226e\text{-}05 & -5.092e\text{-}05 & 1.686e\text{-}05 & -0.0001803 & -9.669e\text{-}05 \\ u17 & 0.001716 & 0.004841 & -0.0003512 & 0.01557 & 0.01211 \\ u18 & -4.037e\text{-}05 & -0.0001131 & 5.406e\text{-}06 & -0.0003611 & -0.0002863 \end{array},$$

$$C=\begin{array}{cccccc} & x1 & x2 & x3 & x4 & x5 \\ y1 & -186{,}4 & 112 & -690.5 & -207.6 & 109.5 \\ y2 & -211 & 21.83 & -218.9 & 4.379 & 246.9 \\ y3 & -1.42e{+}05 & 6429 & -24.7 & 4534 & -3389 \\ y4 & -130.1 & -788.2 & -29.21 & -29.46 & 4.569 \\ y5 & 927.9 & -131.5 & -485 & 424 & -73.25 \end{array},$$

$$D=\begin{array}{ccccccccccccccccccc} & u1 & u2 & u3 & u4 & u5 & u6 & u7 & u8 & u9 & u10 & u11 & u12 & u13 & u14 & u15 & u16 & u17 & u18 \\ y1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ y2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ y3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ y4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ y5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array},$$

$$K=\begin{array}{cccccc} & y1 & y2 & y3 & y4 & y5 \\ x1 & -3.522e\text{-}06 & -0.000131 & -8.208e\text{-}06 & -7.765e\text{-}05 & 0.0001087 \\ x2 & 0.0003453 & -0.0005231 & -6.992e\text{-}06 & -0.00175 & 0.0003072 \\ x3 & -0.001026 & -0.000259 & -7.285e\text{-}06 & -7.355e\text{-}05 & -7.987e\text{-}05 \\ x4 & -0.0008064 & -0.0005067 & -2.332e\text{-}05 & -0.0003362 & 0.002297 \\ x5 & 0.000549 & 0.0008484 & -5.857e\text{-}05 & 0.000376 & 0.001943 \end{array}.$$

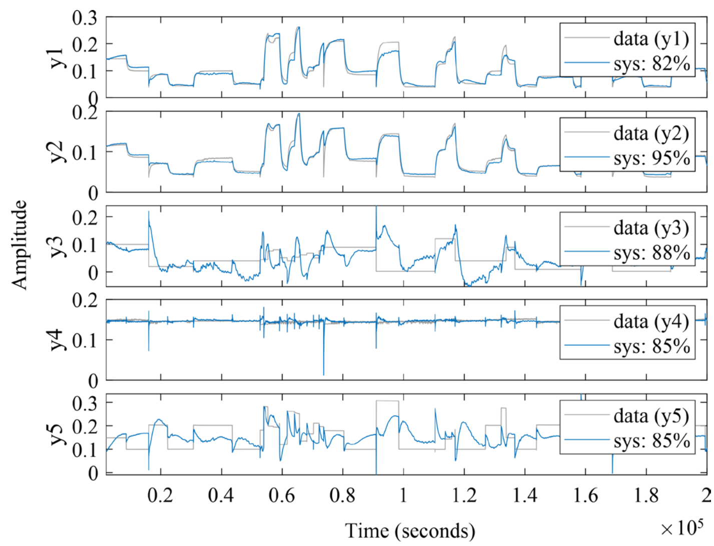

Fig. 5. Results of identification of a mathematical model of an electric motor with artificially formed features

With the help of methods of artificial feature engineering, new features diff, mult, div, epx, square, log, sqrt were formed, which were used to build models. Fig. 5 demonstrates that the artificially formed features have a high correlation with the original parameters, which confirms their informativeness. From Fig. 6, it can be seen that the accuracy of models with artificially generated features has improved significantly, reaching 80–95 % for different channels. The results of a comparison of the accuracy of the basic and extended models for all output channels are shown in Fig. 6.

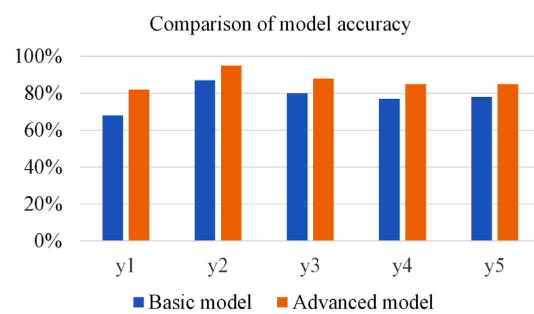

Fig. 6. Results of the comparison of the accuracy of the basic and advanced models

The results in Fig. 6 show that the use of artificial feature engineering has made it possible to increase the accuracy of mathematical models, which confirms the feasibility of using this approach.

6. Discussion of results of investigating the use of artificially formed features in the identification of mathematical models

The identification of the mathematical model of the electric motor showed that the base models have accuracy in the range of 65–85 % for different model outputs. This can be explained by the high level of linear dependence between input and output parameters, which can be seen from the correlation matrix (Fig. 2).

The main advantage of the proposed solutions is the use of the method of artificial feature engineering, which made it possible to increase the accuracy of the models (Fig. 7).

Fig. 5 shows that the artificially formed features have a high correlation with the original parameters, which confirms their informativeness. Owing to this, the accuracy of models with artificially formed features has improved significantly (Fig. 6). This result allows us to use the generated features, in the approaches given in work [4, 5], as data for more effective training of artificial neural networks.

Unlike conventional methods that use only raw data [1, 2], this approach makes it possible to create additional features that significantly increase the accuracy of models. For example, unlike data augmentation methods, where artificial data are added to increase the sample size, artificial feature engineering forms new, informative parameters, the application of which increases the accuracy of the model by 10–20 % (Fig. 6). This becomes possible due to a deeper relationship between new input features and output parameters.

One of the main problems in building mathematical models is that the initial set of input features can be limited. The selection of existing features may not always be effective since these features may not fully reflect the complex relationships between the input and output parameters of the system. In this case, it is important to expand the feature space beforehand by creating new, artificial features that may include relationships or non-linear relationships between the initial parameters. Unlike [8, 9], in our case, only redundant, artificial features are selected, which solves the problem of the limited number of features. This becomes possible due to the fact that as a result of the presence of a large number of artificial, more informative features, the basic features, if necessary, can be left unchanged.

The proposed solution solves the problem of an insufficient amount of high-quality informative data that can be used to build a model of technical and operational characteristics in the development of automatic control systems. And therefore, it makes it possible to obtain a mathematical model of a high-precision engine not only with the main, power characteristics, but also with auxiliary technical and operational ones.

However, this study has limitations. First, the models were identified based on data collected under specific experimental conditions, which may limit their applicability in other settings. Second, the use of artificial features could lead to a significant increase in model structure if the new data are not sufficiently informative.

The disadvantage of the study is that all experiments were conducted under laboratory conditions. Under the actual operating conditions of the electric motor, additional factors may arise that were not taken into account during the simulation. This can affect the accuracy and reliability of the resulting models.

The development of this research may involve the following:

1. Expansion of the database with experimental data collected under different operating conditions of the electric motor.

2. Using machine learning techniques, such as neural networks, to improve the accuracy of the original models.

3. Development of adaptive models capable of changing their parameters in real time depending on operating conditions.

4. Researching the possibilities of integrating models with automated electric motor control systems to increase the efficiency and reliability of their operation.

Our results and conducted analyses confirm the expediency of further development and application of the proposed approaches for controlling electric motors under various operating conditions.

7. Conclusions

1. Preliminary data analysis made it possible to determine key dependences that should be taken into account during further identification of the model, such as *d*-component of voltage and *d*-component of current (correlation values are close to –1), as well as to highlight parameters, for example, *q*-component of voltage (the correlation is close to 0), which have no significant influence.

2. The structure of the proposed model has 6 inputs and 5 outputs. Accuracy in the range of 65–85 % was obtained for various model outputs. The models obtained as a result of identification on the basis of experimental data showed a low threshold of their accuracy, and therefore, the impracticality of their use in the synthesis of control systems.

3. More than 100 artificial features were synthesized. Based on the results of the correlation analysis of artificially formed features, the best representatives with correlations above 0.6 were selected, which made it possible to generally increase the informativeness of the sample for further identification.

4. The structure of the proposed model has 18 inputs and 5 outputs. The accuracy of the models was obtained in the range of 80–95 % for different channels, which confirms the effectiveness of the applied approach. The resulting model proves the effectiveness of applying the principle of artificial feature engineering since the accuracy of the resulting model is 5–10 % higher than the basic model.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Funding

The study was conducted without financial support.

Data availability

The manuscript has associated data in the data warehouse.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. Drif, M., Cardoso, A. J. M. (2014). Stator Fault Diagnostics in Squirrel Cage Three-Phase Induction Motor Drives Using the Instantaneous Active and Reactive Power Signature Analyses. IEEE Transactions on Industrial Informatics, 10 (2), 1348–1360. https://doi.org/10.1109/tii.2014.2307013
2. Trzynadlowski, A. M. (2001). Construction and steady-state operation of induction motors. Control of Induction Motors, 15–41. https://doi.org/10.1016/b978-012701510-1/50002-7
3. Lindegger, M. (2009). Economic viability, applications and limits of efficient permanent magnet motors. Switzerland: Swiss Federal Office of Energy.
4. Zhang, M., Tang, J., Zhang, X., Zhang, J. (2016). Intelligent diagnosis of short hydraulic signal based on improved EEMD and SVM with few low-dimensional training samples. Chinese Journal of Mechanical Engineering, 29 (2), 396–405. https://doi.org/10.3901/cjme.2015.1214.147
5. Matić, D., Kulić, F., Pineda-Sánchez, M., Kamenko, I. (2012). Support vector machine classifier for diagnosis in electrical machines: Application to broken bar. Expert Systems with Applications, 39 (10), 8681–8689. https://doi.org/10.1016/j.eswa.2012.01.214
6. Lei, Y., Jia, F., Lin, J., Xing, S., Ding, S. X. (2016). An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data. IEEE Transactions on Industrial Electronics, 63 (5), 3137–3147. https://doi.org/10.1109/tie.2016.2519325
7. Boukra, T., Lebaroud, A., Clerc, G. (2013). Statistical and Neural-Network Approaches for the Classification of Induction Machine Faults Using the Ambiguity Plane Representation. IEEE Transactions on Industrial Electronics, 60 (9), 4034–4042. https://doi.org/10.1109/tie.2012.2216242
8. Wang, J., Gao, R. X., Yan, R. (2014). Multi-scale enveloping order spectrogram for rotating machine health diagnosis. Mechanical Systems and Signal Processing, 46 (1), 28–44. https://doi.org/10.1016/j.ymssp.2013.06.001
9. Boashash, B. (2015). Time-frequency signal analysis and processing: A comprehensive reference. Academic Press.
10. Electric Motor Temperature. Available at: https://www.kaggle.com/wkirgsn/electric-motor-temperature
11. Korotynskyi, A., Zhuchenko, O. (2020). A system of automated control for the baking process that minimizes the probability of defects. Eastern-European Journal of Enterprise Technologies, 1(2 (103)), 58–67. https://doi.org/10.15587/1729-4061.2020.195785