Data Mining Based Induction Motor Inter-Turn Short-Circuit Fault Diagnosis

Maria Valdivia [1], Antonio Zamarón[1], Miguel A. Casillas[1], Héctor Fraire[2]

[1] Instituto Tecnológico de Léon Av. Tecnológico S/N Frac. Ind. Julián de Obregón. C.P. 37290 León Guanajuato, México.
[2]Instituto Tecnológico de Ciudad Madero Av. 1o. de Mayo esq. Sor Juana Inés de la Cruz s/n Col. Los Mangos C.P.89440 Cd. Madero Tamaulipas, México.

Abstract. A fault diagnosis system of inter-turns short-circuits in three-phase squirrel cage induction motor based on Data Mining is presented in this paper. The system uses the stator current time domain to obtain the inter-turns short-circuit information. Average and variance are features calculated from stator current data. These features are analyzed combining Decision Stump classifier and learning model LWL (Local Weighted Learning) in order to classify the faults. Experimental results indicate that Data Mining is an effective technique in the diagnosis of inter-turns short-circuits in induction machines.

Keywords: Data mining, Decision Stump, Fault diagnosis, Induction motor, LWL.

1 Introduction

The intelligent fault diagnosis technology is the combination of artificial intelligence and fault diagnosis. It is an artificial approach of computer simulating human experts to diagnose faults in complex systems [1]. In general, the task of fault diagnosis is to determine whether there is any fault, and to identify the fault in a system by reasoning machine. By other hand Data Mining is a piece of the Artificial Intelligent that can be defined as the nontrivial extraction process of implicit, previously unknown, potentially useful and ultimately user-understandable knowledge from large-scale database [2]. Data Mining is used to analyse enormous historical data, disclose the implicit regular contents, and the further develop a modelling analysis method. By data mining different types of models of the general system and sub-systems can be set up. These models can not only describe the current situation and regularity, but also can be used to forecast the possible situation under changing conditions. In this work a data mining model based on the Decision Stump classifier and LWL is used to diagnosis turn-to-turn short circuits in induction motors. In recent years there has been an increasing interest in induction motor intelligent fault diagnosis, an online stator winding turn fault detection based on a fuzzy model is presented in [3], in [4] is shown a neural network method to diagnose broken rotor bars and [5] uses a genetic algorithm and artificial neural networks to diagnose different faults. It is well-known that induction motor is a rotating machine with wider application in the industrial field. Thanks to its robustness, low cost, easy maintenance and versatility have become popular with applications ranging from household devices to more sophisticated equipment for industrial processes. In spite of preventive maintenance routines that can be applied, the motor may present diverse incipient or imminent faults. For these reasons, much researchers have been motivated to develop induction motor diagnosis. The anticipated detection of a possible cause of failure allows to scheduling maintenance aims. Therefore it is desirable that the fault diagnosis be realized by means of noninvasive techniques allowing the machine to continue in operation. The most common faults that could occur in the motor are: bearing fault, rotor bar, stator windings and shaft / coupling. Stators windings failures account for 21% of faults in the machine. Two main classes of stator winding failures can be considered: asymmetry in the stator windings such as an open-phase failure and short-circuit of a few turns in a phase winding. A short circuit is recognized as one of the most difficult failures to detect [6]. Inter-turn short circuits faults that occur on the stator, start with a few inter-turns until reach a more severe failure.

This paper presents a new method of fault diagnosis of inter-turns short circuits based on the analysis of the stator phase currents using Data Mining. The technique used for the fault diagnostic is based on Decision Stump classifier and LWL learning model.

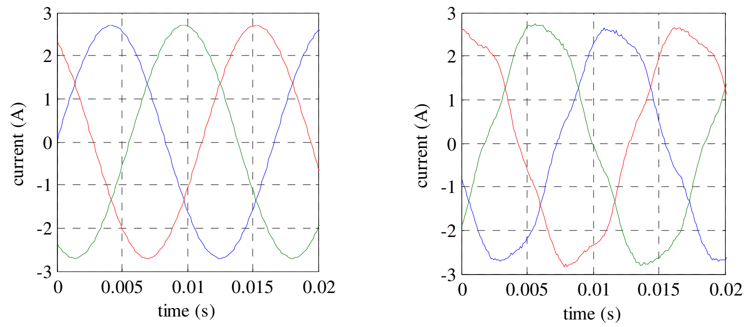
2 Preparing the Data for Classification

The preparation of the data is an important step in the process of classification or prediction, because the data in its original form are rarely suitable for classification; this step is applied to improve the accuracy, efficiency, and scalability of the classification or prediction process. Table 1 shows the results obtained without pre-processing data using some classifiers in WEKA software. All classifiers parameters used in the calculations are set in WEKA by default (version 3.5.8). As can be seen none of these classifiers show an efficiency estimated over 50%. Data are for a fixed motor speed (1800 rpm).

Table 1. Classifiers Efficiency tested without data pre-processing

Classifier	Efficiency Estimated (%)	Estimated Error (%)	No. of Faults	Training (%)	Motor Speed (rpm)
NaiveBayes	10.19	89.81	10	75	1800
LWL	13.07	86.93	10	75	1800
DecisionStump	11.36	88.64	10	75	1800
NBTree	22.11	77.81	10	75	1800
RandomTree	38.11	61.89	10	75	1800
DecisionTable	25.42	74.58	10	75	1800

Because of these results, it is necessary to use a pre-processing data approach to try to improve them. To apply a preprocessing approach is necessary to take into account some considerations. The 60 Hz cycle stator current waves obtained from the machine are not perfect sine waves and the presence of a turn-to-turn short circuit fault is reflected in a deformation of the sine waves. If the sine waves were ideal like shown in Figure 1a formed by points uniformly distributed in the full wave, its average would be zero because both the negative and the positive half-wave are equal. Figure 1b shows an experimental stator currents waves for a health machine. In this case the average of data is not exactly zero, this is why it may make sense for a classification algorithm.



(a) (b)

Figure 1 Stator current evolution. a) Simulation data for a health machine. b) Experimental data for a health machine.

According to experimental stator currents waves obtained for both healthy and faulty machine, the signals are not perfect sine waves, therefore the average is not exactly zero. This feature can be used in part to distinguish one fault from each other. Considering the average and variance as extracted features of electrical current cycle, these can be used as a pattern cycle in the data mining.

Table 2 Results of classification using preprocessed data

Classifier	**Efficiency Estimated (%)**	**Estimated Error (%)**	**No. of Faults**	**Training (%)**	**Motor Speed (rpm)**
NaiveBayes	100	0	10	75	1800
LWL	76.67	23.33	10	75	1800
DecisionStump	14.67	85.33	10	75	1800
NBTree	99.33	0.67	10	75	1800
RandomTree	85.33	14.67	10	75	1800
DecisionTable	100	0	10	75	1800

Table 2 shows the results of classifiers using the average and variance in the preprocessing. As can be seen the efficiency estimated was improved compared to Table 1. In Table 2 also can be seen that decision stump classifier presents a very low efficiency (14.66%) by itself but hybridized with LWL learning model it increase significantly (76.66%). Table 1 and 2 are for fixed motor speed (1800 rpm).

3 Local Weighted Learning (LWL)

LWL is a learning model that belongs to the category of lazy classifiers; Lazy type classifiers are called in this way because they learn from their neighbors and always use the original data. When it is making a classification or prediction, lazy learners can be computationally expensive. This classifier types are known like instance based and they require the handling of large amounts of information which can be a disadvantage. WEKA work by default with LWL model and Decision Stump in combination as classifier. Decision Stump usually is used in conjunction with a boosting algorithm.

Boosting is one of the most important recent developments in classification methodology. Boosting works by sequentially applying a classification algorithm to reweighted versions of the training data, and then taking a weighted majority vote of the sequence of classifiers thus produced. For many classification algorithms, this simple strategy results in dramatic improvements in performance. This seemingly mysterious phenomenon can be understood in terms of well known statistical principles, namely additive modeling and maximum likelihood. For the two-class problem, boosting can be viewed as an approximation to additive modeling on the logistic scale using maximum Bernoulli likelihood as a criterion.

We are trying to find the best estimate for the outputs y_i, using a local model that is an hiper-plane. Distance weighting the data training points corresponds to requiring the local model to fit nearby points well, with less concern for distant points:

$$C - \sum_i (x_i{}^T \beta - y_i)^2 \tag{1}$$

This process has a physical interpretation. The strength of the springs are equal in the unweighted case, and the position of the hiper-plane minimizes the sum of the stored energy in the springs (Equation 2). We will ignore a factor of 1/2 in all our energy calculations to simplify notation. The stored energy in the springs in this case is C of Equation 1, which is minimized by the physical process.

$$w = \int dFx = \int Kdx.x = K \int xdx = \frac{Kx^2}{2} \tag{2}$$

The linear model in the parameters β can be expressed as:

$$x_i^{\mathrm{T}}\beta = y_i \tag{3}$$

In what follows we will assume that the constant **1** has been appended to all the input vectors $\boldsymbol{x_i}$ to include a constant term in the regression. The data training points can be collected in a matrix equation:

$$X\beta = y \tag{4}$$

where ***X*** is a matrix whose ***i***th row is $\boldsymbol{x_i}^{\mathbf{T}}$ and ***y*** is a vector whose ***i***th element is $\boldsymbol{y_i}$***.*** Thus, the dimensionality of ***X*** is $\boldsymbol{n\ x\ \ d}$ where ***n*** is the number of data training points and ***d*** is the dimensionality of ***x.*** Estimating the parameters β using an unweighted regression minimizes the criterion given in equation 1 [7]. By solving the normal equations

$$(\mathrm{X}^{\mathrm{T}}\mathrm{X})\beta = \mathrm{X}^{\mathrm{T}}\mathrm{y} \tag{5}$$

for $\boldsymbol{\beta}$:

$$\beta = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{y} \tag{6}$$

Inverting the matrix $\mathbf{X^TX}$ is not the numerically best way to solve the normal equations from the point of view of efficiency or accuracy, and usually other matrix techniques are used to solve Equation 5.

Additive Logistic Regression

An explanation of logistic regression begins with an explanation of the logistic function:

$$f(z) = \frac{1}{1+e^{-z}} \tag{7}$$

The logistic function is useful because it can take as an input any value, from negative infinity to positive infinity, whereas the output is confined to values between 0 and 1. The variable z represents factors like average and variance for each line of features, while □(□) represents the probability of a particular outcome, given that set of factors. The variable z is a measure of the total contribution of all the average and variance for each line of features used in the model and is known as the *logit* model [8].

The variable z is usually defined as

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_6 x_6 \tag{8}$$

Where β_o is called the "intercept" and β_1, β_2, β_3 and so on, are called the regression coefficients of x_1, x_2, x_3 respectively. The intercept is the value of z when the value of all features is zero. Each of the regression coefficients describes the size of the contribution of that factor. A positive regression coefficient means that the feature increase the probability of that outcome, while a negative regression coefficient means that the feature decreases the probability of that outcome; a large regression coefficient means that the feature considered strongly influences the probability of that outcome; while a near-zero regression coefficient means that the feature has little influence on the probability of that outcome.

Logistic regression is a useful way of describing the relationship between one or more electrical current data (e.g., average phase A, variance phase A, average phase B, variance phase B, etc.) and an outcome, expressed as a probability, that has only two possible values, such belong or not belong to a certain class of fault.

4 Experimental setup.

Experiments in an induction motor have been carried out in order to obtain turn-to-turn short circuit faults data. The machine was modified to have the ability to short adjacent turns. In Figure 2 the experimental setup is shown. It includes a Foucault currents based break dynamic used to control the speed and a tachometer to measure it. Three Hall effect based sensors are used to measure the stator currents and a Data Acquisition Card (National Instruments DAQ-Card USB-9162) to send them to the PC. The software used to acquire the data is implemented in LabVIEW version 8.2. Data are sampled at 9840 samples per second in order to have 164 samples for a 60 Hz cycle.

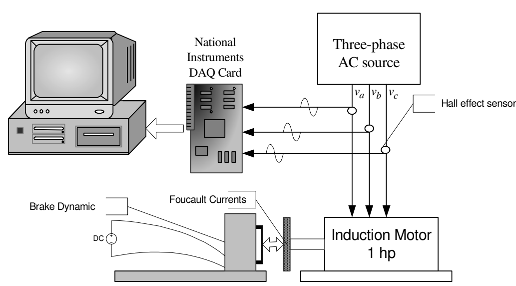


Figure 2 Experimental setup.

5 Results and Conclusions

In this section the results and conclusions obtained are presented. In Table 4 is illustrated de efficiency estimated for a set of data at 1800 rpm and other set of data that includes seven different speeds (1800, 1790, 1780, 1770, 1760, 1750 and 1745 rpm). It can be seen that naive Bayes classifier give a 100% of efficiency estimated for one speed but this efficiency is reduced to 11.98% when the set of data include several speeds. The same behavior is presented in the others classifiers tested. In some of them the reduction of the efficiency estimated is big enough. LWL classifier was stopped after five hours of computations because for same quantity of data the other classifiers take just several minutes (this case is marked by "*" in Table 4). This may be because LWL learning model use for default the maximum neighbors in the neighborhood.

Table 4 Comparison of classification results at 1800 rpm speed and seven different speeds.

Classifier	Efficiency Estimated (%) (1800 rpm)	Estimated Error (%)	Efficiency Estimated (%) for seven speed set	Estimated Error (%)	Training (%)
NaiveBayes	100.00	0.00	11.98	88.02	75
LWL	76.66	23.34	*	*	75
DecisionStump	14.66	85.34	12.67	87.33	75
NBTree	99.33	0.67	97.63	2.37	75
RandomTree	85.33	14.67	83.68	16.32	75
DecisionTable	100.00	0.00	92.64	7.36	75

Until this point all computations has been carried out using default parameter set by WEKA. In order to improve the efficiency estimated on LWL classifier the neighborhood size is adjusted the parameter KNN = 9. With this modification the efficiency estimated at fixed speed is increased from 76.66% to 96.66%. In the case of efficiency estimated for several speeds the result is 93.25%, it also was reached in several minutes, equivalent to the time required for the other classifiers.

The classifier model was tested with a new data set carried out at fixed speed getting an efficiency estimated of 90.70% compared to 96.66% obtained from the training data set. In order to make data mining several features can be used to recognize pattern in our case two statistical metrics were use average and variance for this purpose. As was shown is this document this metrics improved significantly the efficiency estimated. Experimental results indicate that classification techniques included in the data mining are effective techniques in the diagnosis of inter-turns short-circuit in induction machines, similar at used in [10].

Acknowledgments

The autors wish to acknowledge the financial support and motivation provided by CONACYT, Instituto Tecnológico de León and Instituto Tecnológico de Cd. Madero.

References

1. Chunjie Bai and Yexin Song, "Structure Design of Intelligent Fault Diagnosis System Based on Data Mining", Proceedings of the 6th World Congress on Intelligent Control and Automation, June 21-23, 2006, Dailan, China.
2. Xianghui Wang, Xiaoyu and Weida Liu, "The Technology od data mining and its application on decision supporting system", Computing Technology and Automation, vol. 23, no. 4, pp 99-102, 2004.
3. Wang Hu-hong and He Yi-gang, "Fuzzy Model based On-line Stator Winding Turn Fault Detection for Induction Motors", Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06), 2006.
4. Qing He and Dong-mei Du, "Fault Diagnosis of Induction Motor using Neural Networks", Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007
5. Tian Han, Bo-Suk Yang and Jong Moon Lee, "A New Condition Monitoring and Fault Diagnosis System of Induction Motors using Artificial Intelligence Algorithms", IEEE International Conference on Electric Machines and Drives, May 2005, pp. 1967 – 1974.
6. Bellini Alberto, Filippetti Fiorenzo, Tassoni Carla and Capolino Gérard-André, "Advances in Diagnostic Tecniques for Induction Machines", IEEE Transactions on Industrial Electronics, December 2008, pp. 4109-4125.
7. Christopher G. Atkeson, Andrew W. Moore and Stefan Schaal, "Locally Weighted Learning" College of Computing. Georgia Institute of Technology, October 1996.
8. Hilbe, Joseph M., "Logistic Regression Models" Chapman & Hall/CRC Press, 2009.
9. Picard, Richard; Cook, Dennis, "Cross-Validation of Regression Models". Journal of the American Statistical Association, vol.**79**, 1984, pp. 575-583.
10. Pablo Serrano, Antonio Zamarrón, Arturo Hernández and Alberto Ochoa, "Artificial Neural Networks for Diagnosing Stator Induction Motor Faults", Research in Computer Sciencies Vol. 39, pp. 251-262. October, 2008.