2015 IEEE 14th International Conference on Machine Learning and Applications

Multiple Imputation of Missing Residuals for Fault Classification: a Wind Turbine Application

Eman M. Nejad, Roozbeh Razavi-Far, Q.M. Jonathan Wu, Mehrdad Saif
Department of Electrical and Computer Engineering
University of Windsor, 401 Sunset Avenue, Windsor, N9B 3P4, Canada
Email: (moham12b, roozbeh, jwu, msaif)@uwindsor.ca

***Abstract*—Handling the missing data is considered as a crucial requirement for the performance of diagnostic systems. In the proposed diagnostic system, the preprocessing module receives sets of residuals generated by a combined set of observers, and feeds the proceeded residuals to a fault classification module. It is necessary for the fault classification module to receive complete feature sets. Multiple missing data imputation techniques have been devised in the preprocessing module to guarantee feeding complete sets of features to the fault classification module. The proposed diagnostic scheme is validated using incomplete batch of residuals for sensor fault diagnosis in a doubly fed induction generator (DFIG) of a wind turbine.**

***Keywords*—multiple imputation, fault diagnosis, sensor faults, wind turbine**

I. Introduction

Electricity generation using the wind power requires a reliable operation of wind turbines. The faults within the wind turbine must be diagnosed as early as possible to prevent major component failures [1]. The doubly fed induction generator (DFIG) is considered as one of the most powerful and widely used electrical machines in wind turbines due to its reliable performance during the normal operation of wind turbines. However, their performance are sensitive to sensor faults [1]. In [1], an intelligent diagnostic system has been developed to diagnose the sensor faults of the DFIG. The developed diagnostic scheme has two major steps. In the first step, several observers employed to generate residual signals which reflects faults in the sensors. The generated residuals are related to stator voltage, current sensors, and rotor sensors [1]. In the second stage, the residuals are fed to a fault classification module for decision making [1-3].

A dynamic weighting ensembles (DWE) algorithm was used for diagnosing new classes of faults in [1]. The preprocessing module bridges the gap between the residual generation and fault classification modules (see Fig. 1), processes the residuals and then feeds the processed residuals to the DWE algorithm [1-4]. Although, the preprocessing module in [4] is able to handle the missing residuals, there is a need to study other strategies to deal with missing values to improve the fault classification performance.

It should be noted that contemporary fault classifiers demand complete feature data and, they are not capable to treat the incomplete data [5-7]. Discarding approaches, single and multiple imputations are considered as major techniques to handle missing data. This paper aims to study and compare the application of multiple imputation techniques as a part of the preprocessing module to impute the missing residual data and improve the fault classification performance using the complete feature sets.

The rest of the paper is organized as follows. The problem statement, preprocessing module, and diagnostic system are briefly described in Section 2. Section 3 presents the multiple imputation algorithms which are used in the preprocessing module. Section 4 presents the experimental results. Conclusions are drawn in Section 5.

II. System Description and Problem Statement

In this study, the considered faults are derived from the stator voltage, stator and rotor currents sensors. The diagnostic system includes three major steps. First, residual signals which are reflecting faults in the DFIG system are produced from sampled command inputs and sensor measurements [1-2]. Second, the preprocessing module transforms the residual signals to a feature space as required for an accurate fault classification to define the location and time of faults. Since each residual r contains two components $\{v_1, v_2\}$, preprocessing module receives a pattern of 18 features for each record $s_i = \{v_j\}_{j=1}^{18}$, and collects k patterns in every snapshot to form a feature set as $S = \{s_i\}_{i=1}^{k}$ [4]. Finally, the processed residuals are sequentially collected $\{S^1, S^2, ..., etc\}$ and gradually applied to the fault classification unit which aims to map each pattern of the feature space to a pre-assigned class of faults [1-4]. There are totally 10 classes including a fault free (ff) class and 9 classes of faults $(f_1, f_2, ..., f_9)$.

The produced residuals are collected in various batches and their signature trends vary in different time intervals [1-2]. The fault classification module, dynamic weighting ensemble algorithm, can dynamically learn the relation between residuals and faults and diagnose multiple classes of faults [1-3]. The DWE algorithm trains and appends a novel ensemble member when a new batch of residual data is emerged. Every ensemble contains a preassigned number of multilayer perceptron (MLP) classifiers. Each MLP network is trained on a diverse subset of the current training data which is derived according to an iteratively updated distribution [1-2]. Each MLP has three layers where the number of neurons in the input layer is equal to the number of residual components; and the number of output neurons is equal to the number of

978-1-5090-0287-0/15 $31.00 © 2015 IEEE
DOI 10.1109/ICMLA.2015.145

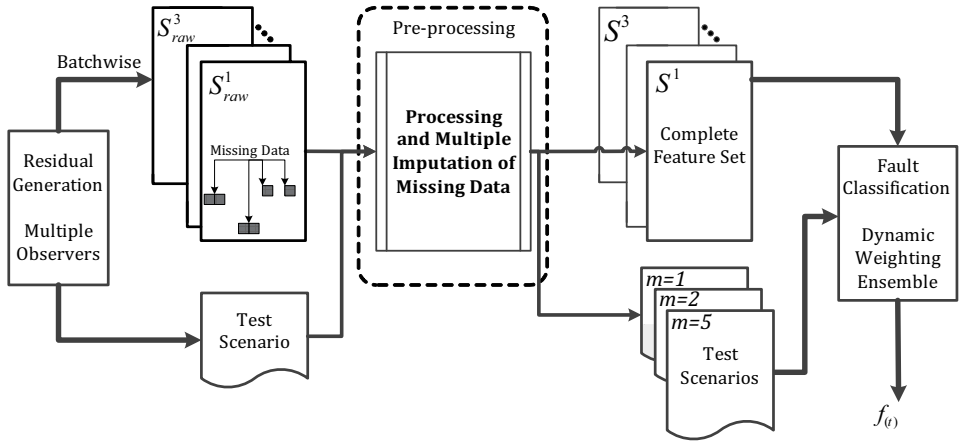

Fig. 1. General diagram of the diagnostic system: the missing residual data are multiply imputed in the preprocessing module to generate various complete feature sets for fault classification module

classes, $(ff, f_1, f_2, ..., f_9)$ [1-2]. The details and notations of the DWE algorithm are described in [1-2].

The individual classifiers fail to map the incomplete patterns [4]. Therefore, there is a need for the classifier to receive complete feature sets. It should be noted that discarding the incomplete residual patterns is not acceptable due to the loss of data for classification. Thus, the feature sets must be imputed in advance. Single imputation methods treat the imputed values as known in the analysis. This underestimates the variance of the estimates and overstates precision and optimistic results in confidence intervals. Multiple imputation takes into account the sampling variability due to the missing data and generates multiple imputed datasets [8]. Therefore, various multiple imputation algorithms are devised in the preprocessing module to generate m sets of complete features for the fault classification module, $m = 1, \ldots, M$. The imputed feature sets lead to make a proper decision and improve the classification accuracy when original batches of raw residual data include missing values. The developed diagnostic system including its preprocessing module and missing data imputation unit is shown in Fig.1.

III. Multiple Imputation Strategies For Missing Data Imputation

Multiple imputation is a three-step process [9]. First, multiple sets of reasonable values for missing features are generated forming M complete sets to reflect the uncertainty. Second, each of the M completed datasets are analyzed. The analysis model is considered as the completed-data model which is employed to obtain completed data estimates, $\widehat{Q}$, of parameters of interest, Q, and the estimate, m, of sampling variability associated with $\widehat{Q}$ [9]. Finally, the individual completed-data estimates $(\widehat{Q}, m)$ are pooled into $(\widehat{Q}_{MI}, T)$ to allow the uncertainty of the imputation counting into account, and form one repeated imputation inference [9]. As stated in [10], five number of imputations should be adequate to attain valid inference. In this paper, S shows the complete dataset including observed S_{obs} and unobserved values S_{mis}. At the following, this section briefly describes the seven multiple imputation algorithms used in this study.

Predictive Mean Matching (PMM): In order to impute the missing value, PMM selects a random value from a set of observed values whose predicted values are closest to the predicted value for the missing value from the simulated regression model [11]. In the regression technique, a regression model is fitted for a continuous variable with the covariates created using a set of effects. Based on the fitted regression model, a new regression model is simulated from the posterior predictive distribution of the parameters and is used to impute the missing values for each variable [10]. Multiply imputing for S_{mis} using the PMM is implemented as follows: Linear regression of S_{obs} given X_{obs} (observed values at the covariates) is used to estimate $\hat{\beta}$, $\hat{\alpha}$, and $\hat{\varepsilon}$ using the ordinary least squares, where β is the regression coefficient vector, and ε is a normally distributed random error with expectation zero and variance σ^2. Next, $\hat{S}_{obs}$ and $\hat{S}_{mis}$ are calculated by $\hat{S}_{obs} = X_{obs}\hat{\beta}$ and $S_{mis} = X_{mis}\hat{\beta}$. Then, for each $\hat{S}_{mis}$ the $\Delta = |\hat{S}_{obs} - \hat{S}_{mis}|$ is calculated. Next, a value from $(\Delta_1, \Delta_2, \Delta_3)$ is randomly selected, where $(\Delta_1, \Delta_2, \Delta_3)$ are the three smallest elements in Δ, and the corresponding $\hat{S}_{obs}$ is chosen as the imputation. For the multiple imputation, these steps are repeated M times, each time saving the completed dataset.

Mahalanobis Distance Method (MDM): For the patterns containing missing values, the distance between that pattern and all other patterns within the residual dataset is calculated. The distance is calculated using covariates specified where X is the vector of the covariates for the pattern with the missing value and S_{obs} is the vector for the $i-th$ fully observed pattern in the dataset. The expression for the mahalanobis distance is given as follows:

$$D_{(S_{obs},X)} = \sqrt{(S_{obs} - X)^T C^{-1} (S_{obs} - X)} \qquad (1)$$

where C is considered as the covariance matrix for the set of covariates which are employed in the calculation. Each of the missing values in the incomplete pattern is imputed using

the observed value of the pattern considering the minimum Mahalanobis distance to the missing data value that needs to be imputed [12]. For each missing value in the patterns that is to be imputed, a smaller sub-set is chosen on the basis of the association among the missing value and refinement variable. Then, the selected smaller sub-set is employed to generate the imputations. For each missing value from the residuals, the imputations are randomly performed according to the Approximate Bayesian Bootstrap technique using the selected sub-set of observed values from the residuals [10]. A random sample with replacement is randomly selected from the chosen sub-set of the observed values and the imputations are randomly done from this sample [10].

Markov Chain Monte Carlo (MCMC): The bayesian network is employed to impute missing data using MCMC [13]. The basic idea of imputation by means of Bayesian networks is to learn a model from the data, predict values for the missing data, update the model with the upgraded data and update imputes with the upgraded model. This progression is iterated several times until convergence criterion has been met. The general expression for the inference is described as follows:

$$p(S_{mis}|S_{obs}, G) = \frac{p(S_{mis}|G)p(S_{obs}|S_{mis}, G)}{\int p(S_{obs}|S_{mis}, G)p(S_{mis}|G)dS_{mis}} \quad (2)$$

where G signify the graphical model of a Bayesian network according to which the joint probability distribution is factored. Data augmentation is applied to Bayesian inference with missing data by repeating the imputation and posterior steps. In the imputation step, the missing values for each observation are estimated independently using the estimated mean vector and covariance matrix. If the features with missing values are signified by S_{mis} and the variables by observed values by S_{obs}, then the imputation step draws values for S_{mis} from a conditional distribution S_{mis} given S_{obs}. The second step estimates the posterior population mean vector and covariance matrix from the completed estimates. Then, these new estimates are used in the imputation step [13]. The two steps are repeated many times for the results to be reliable for a multiply imputed data set [13].

Predictive Model Based Method (PMB): The predictive information in a specified set of covariates is employed to impute the missing values as follows: First, the Predictive Model is estimated using the observed data. Using this estimated model, novel linear regression parameters are arbitrarily drawn from their Bayesian posterior distribution. The arbitrarily selected values are considered to impute the missing values, which include random deviations from the models predictions. Drawing the exact model from its posterior distribution guarantees that the extra uncertainty about the unknown true model is reflected. In the stated process, multiple regression estimates of parameters are attained by means of the method of least squares [14]. For the least square regression model, the S_{mis} and X are considered as the variable to be imputed and the set of covariates respectively. It should be noted that all the variables without any missing value are included in the set of covariates in this study.

Bootstrap EM algorithm Imputation (BEM): The mean and covariance matrix, (θ), are considered as the parameters of complete data for multiple imputation. In a model of residual data, the observed data is shown by S_{obs} and the missing values is presented by S_{mis}. Therefore, the likelihood of the observed data is $p(S_{obs}, S_{mis}|\theta)$. The likelihood of the observed data can be break up as follows:

$$p(S_{obs}, S_{mis}|\theta) = p(S_{mis}|S_{obs})p(S_{obs}|\theta) \quad (3)$$

Since only the inference on the complete data parameters is considered, the likelihood is described as follows:

$$L(\theta|S_{obs}) \ \propto p(S_{obs}|\theta) \quad (4)$$

The stated likelihood is rewritten by means of law of iterated expectations as follows:

$$p(S_{obs}|\theta) = \int p(S|\theta)dS_{mis} \quad (5)$$

The posterior is defined using the stated likelihood and a flat prior on θ, and described as follows:

$$p(\theta|S_{obs})\mu p(S_{obs}|\theta) = \int p(S|\theta)dS_{mis} \quad (6)$$

The main challenge in the analysis of incomplete sets of residuals is to draw from this posterior. The EM algorithm [15] is employed to find the mode of the posterior. The EM algorithm is combined with a bootstrap approach to take draws from the stated posterior. The data are bootstraped to simulate estimation uncertainity. Next, the EM algorithm is used to figure out the mode of the posterior for the bootstraped data. When the draws of the posterior of the complete data parameters are attained, the imputation is performed by drawing values of S_{mis} from its distribution conditional on S_{obs} and the draws of θ that is considered as a linear regression with parameters which are calculated directly from θ. The details of the BEM algorithm are completely described in [16].

Propensity Score Method (PSM): In the PSM, a propensity score is produced for each residual with missing values to define the probability of that observation being missing. Next, the observations are clustered based on the given propensity scores. Then, an approximate Bayesian bootstrap imputation is applied to each cluster to impute the missing values. The notation of the PSM algorithm is completely described in [17].

Multivariate Imputation by Chained Equations (MICE): A series of regression models are employed whereby each feature with missing data is modeled conditional upon the other residuals in the data. In other words, each residual can be modeled according to its distribution [18]. In order to impute missing values in residuals, the MICE algorithm is broken down into four major steps. First, the PMM algorithm is used to impute every missing value in the dataset. These PMM imputations are called place holders. Second, the place holder PMM imputations for one residual are set back to missing. Third, the observed values, S_{obs}, in the second step are regressed on the other residuals in the imputation model which consist of all residuals in the dataset. It should be noted

TABLE I
THE RANKING OF THE MI ALGORITHMS

Rank	MI Algorithm	ff	f_1	f_6	f_8	Total
1	MCMC	91%	85%	**73%**	75%	81%
2	MDM	91%	**91%**	50%	**79%**	77.75%
3	MICE	91%	80%	50%	63%	71%
4	PMM	91%	73%	50%	61%	68.75%
5	PSM	91%	69%	55%	59%	68.5%
6	BEM	91%	66%	50%	64%	67.75%
7	PMB	91%	53%	50%	70%	66%

that S_{obs} are considered as the dependent residuals. Finally, the S_{mis} are imputed using the regression model. At the end, steps 2 through 4 are repeated for a number of iterations to update the imputation at each cycle. The details of the MICE algorithm are presented in [18].

IV. EXPERIMENTAL RESULTS

The generated raw residuals are processed in the preprocessing module and then fed to the fault classification module. In the classification module, the DWE algorithm trains T_k individual fault classifier with a subset of S^k to form an ensemble $\mathcal{E}_k$. The training phase is fully described in [1]. The diagnostic system is validated using a realistic wind turbine sequence. The missing values are generated by failure in a sensor reading in the four related residual components because of the mutual relation among the sensor measurement and residual components [4]. The total number of failures in sensor reading generate 250 monotone and 500 non-monotone missing patterns. The pattern containing missing values are multiply imputed using the proposed techniques in section III, and then applied to the fault classification module.

In this study, multiple imputation involves imputing 5 values for each missing cell in the dataset and generating 5 completed data sets. In these completed data sets, the observed values are the same, but the missing values are filled in with a distribution of imputations that reveal the uncertainty regarding the missing data. The ranking of seven multiple imputation algorithms in terms of classification performance is stated in Table I. In the table, the bold entry in each column presents the best classification performance for each class. The patterns of the class ff are fully observed, and thus, the multiply-imputation methods are ineffective. Therefore, the classification performance with respect to the ff class is constant. The classification performance with respect to the f_6 using the MCMC is higher compare to other algorithms. The highest accuracy of fault classification for f_1 and f_8 is achieved by the Mahalanobis Distance method. The overall highest accuracy to impute missing data for fault classification is achieved by MCMC and the overall lowest accuracy is attained using the PMB algorithm. In general, the fault classification performances are significantly improved using the MCMC and MDM compare to other competitors.

V. CONCLUSION

This work studies the performance of multiple imputation algorithms in the preprocessing module of the diagnostic system in the controlled DFIG sensors. The diagnostic system has been validated on a three phase simulation. In the stated simulations, new classes of faults are generated at each phase and finally tested with respect to incomplete faulty scenarios. The major goal of this paper was to handle the missing residual data by resorting to multiple imputation algorithms. The missing residual data are multiply-imputed and the complete feature sets are fed to the fault classification module. This procedure leads to have an enhanced classification performance under missing data assumption. The highest classification accuracy for the simulated incomplete scenario was achieved using the MCMC algorithm.

REFERENCES

[1] R. Razavi-Far, M. Kinnaert, "A multiple observers and dynamic weighting ensembles scheme for diagnosing new class faults in wind turbines," Control Engineering Practice, vol. 21, no. 9, pp. 1165-1177, 2013.

[2] R. Razavi-Far, M. Kinnaert, "Incremental design of a decision system for residual evaluation: A wind turbine application" In 8th IFAC Conf. on fault detection, supervision and safety of technical processes, vol. 8(1), 2012, pp. 343-348.

[3] R. Razavi-Far, V. Palade, E. Zio, "Optimal detection of new classes of faults by an invasive weed optimization method," Proceedings of the International Joint Conf. on Neural Networks (IJCNN), pp. 91-98, 2014.

[4] R. Razavi-Far, E. Zio, V. Palade, "Efficient residuals pre-processing for diagnosing multi-class faults in a doubly fed induction generator, under missing data scenarios," Expert Systems with Applications, vol. 41, no. 14, pp. 6386-6399, 2014.

[5] R. Razavi-Far, H. Davilu, V. Palade, C. Lucas, "Neuro-fuzzy based fault diagnosis of a steam generator," In 7th IFAC Conf. on Fault Detection, Supervision and Safety of Technical Processes, Barcelona, Spain, 2009, pp. 1180-1185.

[6] P. Baraldi, R. Razavi-Far, E. Zio, "A method for estimating the confidence in the identification of nuclear transients by a bagged ensemble of FCM classifiers" In Proc. NPIC&HMIT, Las Vegas, NV, Nov. 7-11, 2010, pp. 283-293.

[7] R. Razavi-Far, P. Baraldi, E. Zio, "Dynamic weighting ensembles for incremental learning and diagnosing new concept class faults in nuclear power systems" IEEE Transactions on Nuclear Science, vol. 59, no. 5, pp. 2520-2530, 2012.

[8] D.B.R. Roderick J.A. Little, "Statistical analysis with missing data," 2nd Edition, Wiley, 2002.

[9] D.B. Rubin, "Multiple imputation after 18+ years," Journal of American Statistical Association, vol. 91, no. 434, pp. 473-489, 1996.

[10] D.B. Rubin, "Multiple imputation for nonresponse in surveys," Wiley, 1987.

[11] N. Schenker, J.M.G. Taylor, "Partially parametric techniques for multiple imputation," Computational Statistics and Data Analysis, vol. 22, no. 4, pp. 425-446, 1996.

[12] A. Grannell, "SOLAS for missing data analysis," White Paper on Mahalanobis Distance Imputation, December 2011.

[13] J. Schafer, "Analysis of incomplete multivariate data," London: Chapman and Hall, 1997.

[14] A. Grannell, "White paper on predictive model based imputation method," SOLAS For Missing Data Analysis, December 2011.

[15] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, no. 1, pp. 1-38, 1977.

[16] J. Honaker, G. King, M. Blackwell, "AMELIA II: a program for missing data," November 2014.

[17] P.R. Rosenbaum; D.B. Rubin, "The central role of the propensity score in observational studies for causal effects," Biometrika, vol. 70, no. 1, pp. 41-55, 1983.

[18] M.J. Azur, E.A. Stuart, C. Frangakis, P.J. Leaf, "Multiple imputation by chained equations: what is it and how does it work?," Int J Methods Psychiatr Res, vol. 20, no. 1, pp. 40-49, 2012.