Key Issues on Computational Intelligence Techniques for Missing Data Imputation – A Review

Fulufhelo V. Nelwamondo and Tshilidzi Marwala

***Abstract*— The need for the use of computational intelligence techniques for missing data imputation has been overlooked in the literature. This is due to the complexity of computational intelligence techniques and as a result, researchers devote to easier and perhaps, inefficient techniques. In this paper, the theory and practice of using computational intelligence techniques for missing data imputation is explored. Commonly used techniques are reviewed and the latest computational intelligence techniques and their applications are presented. The paper also addresses the trade-offs between traditional heuristics and computational intelligence techniques. Research implications and major challenges in the subject area are discussed whereas new research directions are pointed by highlighting the key issues in the field.**

I. Introduction

Datasets are often characterized by incompleteness. Some of the reasons for missing data are sensor failures, omitted entries in databases and non-response to questions in questionnaires. In most cases, data collectors put in place, strict measures to avoid incompleteness during data collection. It is unfortunate that for some reasons, data are not obtainable for every observation of every variable. More often, the reason for incompleteness in not known and as a result, techniques for the preventing missing data run short of the success. The unavailability of the data hinders the decision making processes due to the dependencies of decisions on information. Most scientific, business and economic decisions are somehow related to the information available at the time of making such decisions. As an example, most business evaluations and decisions are highly dependent on the availability of sales and other information, whereas advances in research are based on discovery of knowledge from various experiments and measured parameters. There are also many situations in fault detection and identification where the data vector is partially corrupted, or otherwise incomplete.

Many decision making processes use predictive models that take observed data as inputs. Such models breakdown when one or more inputs are missing. In many applications, simply ignoring the incomplete record is not an option. This is mainly due to the fact that ignorance can lead to biased results in statistical modeling or even damages in machine control. For this reason, it is often essential to make the decision based on available data. Most decision making tools such as the commonly used neural networks, support vector machines and many other computational intelligence techniques can not be used for decision making if data are not complete. In such cases, the optimal decision output should still be maintained despite the missing data. In cases of incomplete data vectors, the first step toward decision making is to estimate the missing values. Once missing values have been estimated, pattern recognition tools for decision making can then be used. The challenge missing data pose to the decision making process is more evident in on-line applications where data has to be used almost instantly after being obtained. Computational intelligence techniques such as neural networks and other pattern recognition techniques have recently become very common tools in decision making processes. In a case where some variables are not measured, it becomes difficult to continue with the decision making process. The biggest challenge is that the standard computational intelligence techniques are not able to process input data with missing values and hence, cannot perform classification or regression.

II. Historical Evolution of Missing Data Handling Techniques

Prior to 1970s, missing data were solved by editing [1], whereby a missing item could be logically inferred from other data that have been observed. A framework of inference from incomplete data was only developed in 1976. Shortly afterward, Dempster et al. [2] formulated the Expectation Maximisation (EM) algorithm that led to the full use of Maximum Likelihood (ML) techniques in the missing data problem. Only after a decade, Little and Rubin [3] and Rubin [4] documented the shortcomings of case deletion and single imputations and introduced Multiple Imputations (MI). Multiple Imputations would not have been possible without the advancements in computational resources [5] as they are computationally expensive. The years 1995 till today have discussed many techniques of solving the missing data problem in different applications. Latest research is now beginning to analyze the sensitivity of the results to the distribution of missing data [6].

A great deal of research has recently been done to discover new ways of approximating the missing values. Among others, [7], [8] and [9] used neural networks and some evolutionary techniques for missing data imputation. Dhlamini et al. [10] have used evolutionary computing in condition monitoring of high voltage (HV) bushings in the presence of missing data. In their study, auto-associative neural networks were used together with Genetic Algorithms

This work was supported by the National Research Foundation (NRF) of South African

All the authors are with the Faculty of Engineering and the Built Environment, School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, Republic of South Africa. Correspondence email address: fulufhelo.nelwamondo@wits.ac.za

(GA) or Particle Swarm Optimization (PSO) to predict the missing values and also to optimize the prediction. On the other hand, [11] used semi hidden Markov models in prediction of missing data in mobility tracking whereas [12] used the pseudo-nearest-neighbour approach for missing data recovery on random Gaussian data sets. Various techniques and approaches have since been developed and discussed in detail in the current literature, including but limmited to, multi-task leaning, bootstrap, fuzzy sets and rough sets.

III. Patterns of Missing Data

Data can be missing in various patterns as shown in Figure 1. In the figure, rows correspond to observational units whereas columns are variables [1]. Univariate pattern occurs when data is missing from one variable as shown by Y in Figure 1 (a). Monotone pattern occurs when data is missing from a number of variables, but, missing data follows a particular pattern. Lastly, an arbitrary pattern occurs when data is missing following some random pattern as shown in Figure 1 (c). The pattern that the data will follow depends on the application. Sensor failure is more likely to follow the pattern in Figure 1 (a) or (b) whereas for databases where information is recorded by different individuals, as in medical databases, the pattern as in (c) is most likely to be observed. The next section discusses the mechanisms of missing data.

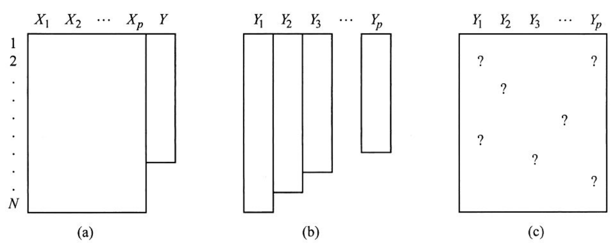

Fig. 1. Patterns of missing data in rectangular databases: (a) univariate pattern, (b) monotone pattern, and (c) arbitrary pattern [1]

There is no doubt that the pattern of missing data can help in dictating which technique needs to be used. This, however, calls for the appropriate pre-processing techniques for a particular pattern. The next section highlights the mechanisms of missingness.

IV. The Effect of Missing Data Mechanism on Imputation

For the problem of missing data, it is very important to know the reason why data are missing. When the reason is known, an appropriate technique for imputation may be selected or derived, leading to a higher efficiency and prediction accuracy. In most cases, data collectors may know such reasons whereas statisticians and data users may not have that information available to them. In such cases, data users may have to use other means that can help in the data analysis to see how the missing data relate to the collected data and as a result, possible reasons may be derived.

A variable or a feature in the dataset is viewed as a mechanism, if it helps in explaining why other variables are missing or not missing, which ever is the case. In datasets collected through surveys, variables that are mechanisms are often associated with information that people are ashamed to disclose and such details can be deduced from the data that has been given. As an example, people who earn very little may be embarrassed to disclose their salary, but may disclose their highest level of education. Data users may exploit the provided information to get an insight to the missing data. Variables that are commonly used as mechanisms include, among others, education, race, age, gender, salary scale and health status.

Considering the notation used in Figure 1, let Y be the variable of interest, $Y_{complete}$ be the complete dataset and X_{obs} be the observed values, then,

$$Y_{complete} = f(Y, X_{obs}) \tag{1}$$

Little and Rubin [3] and Rubin [4] distinguish between three missing data mechanisms. These types are referred to as ‘Missing at Random’, (MAR), ‘Missing Completely at Random’, (MCAR) and ‘Missing Not at Random’, (MNAR) and are described below.

A. Missing at Random

Missing at Random requires that the cause of missing data be unrelated to the missing values themselves. However, the cause may be related to other observed variables. MAR is also known as the ignorable case [13] and occurs when cases with missing data are different from the observed cases but the pattern of missing data is predictable from other observed variables. Differently said, the cause of the missing data is due to external influence and not to the variable itself. Suppose there are two sensors namely S and *T*. For MAR to hold, the probability of datum d from a sensor *S* to be missing at random should be dependent on other measured variables in the database.

B. Missing Completely at Random

Missing Completely At Random refers to a condition where the probability of data missing is unrelated to the values of any other variables, whether missing or observed. In this mechanism, cases with complete data are indistinguishable from cases with incomplete data. Neither, Y_{miss} nor Y_{obs} can help in predicting the missing values.

C. Missing Not at Random

Missing not at random (MNAR) implies that the missing data mechanism is related to the missing values. A good example of data missing not at random can result from a situation where two databases from different cities where merged. Suppose one database lacks some features that have been measured on the other database. In this condition, why some data are missing can be explained. However, this explanation is only dependent on the same variables that are missing and can not be explained in terms of any other variables in the database. Data are classified in this category if they are neither MAR nor MCAR. There has been some different views by researcher, when data are not missing, but are simply answered as '*not applicable*' or as '*don't know*'. Acork [14] defines this as missing by definition of the subpopulation. More clarity is still required in this matter.

V. THE AD HOC TECHNIQUES

There are many ad hoc techniques that many researchers have resorted to, due to their readily availability or their easiness in computing. Some of these techniques include: list deletion, pairwise deletion, simple rule prediction, mean substitution, hot deck imputation, cold deck imputation, regression imputations, regression-based nearest neighbor hot decking, tree based imputation and stochastic imputation. This paper will not detail how these techniques work, but will only generalize the major drawbacks of such techniques.

There is an overwhelming evidence in the literature that shows how poor these methods perform for missing data imputation. Most of these methods either reduce or exaggerate statistical power and as a result, lead to biased estimations [14]. Undoubtedly, listwise deletion is the most commonly technique used. In cases if missing data, the entire observation is removed from the database [3]. The biggest drawback of this method is the amount of information lost in the process. If there is a data set with 50 variables and each variable has a 2% probability of being missing, then, there will be a less than 40% probability that there is an instance with a complete observation of 50 variables. To be precise, if these data were to be observed according to Bernoulli process, only

$$0.98^{50} = 0.364$$

is the expected propotion of data to be complete. This will essentially imply that if there are 100 instances of recorded data and listwise deletion has to be applied, only 37 instances are likely to be used. The mean imputaion on the other hand attenuates the variance, leading to underestimating the correlation of one variable with the other [14] [3].

A. What are the Conclusions?

None of these ad hoc methods gives an optimal solution to the problem of missing data. However, the relevance of a method is a function of how accurate one wants the predictions. They, however, lead to biases except in a few exceptional and specialized conditions. Listwise deletion has been argued to work well for MCAR and when the dataset is very large. However, proving that the dataset follows the MCAR mechanism is difficult and MCAR is often unreasonable in rectangular survey data. The use of mean substitution cannot be justified at any situation as it distorts the variance. This calls for an urgent need to use imputation based techniques as discussed in the next section.

VI. MODERN TECHNIQUES: WHICH ONE IS THE STATE-OF-THE-ART?

There are many techniques that have emerged lately to overcome the traditional techniques for missing data estimation. These techniques can be classified into two broad categories. One category models the probability of missing variables and uses some form of probability weighting to estimate the missing values. The second category models the distribution of missing data and uses explicit imputation techniques. Maximum likelihood methods on the other hand use implicit imputation but also fall within the second category. Some cases, derived from hot-deck imputations, find complete cases with similar values and observed values for the unobserved values. The key here is that all models used, must preserve the relationship in the data. Some of the techniques that meet this goal are discussed shortly. For a very long time, the expectation maximization (EM) method was considered the state-of-the-art. The question is, *is it still the state-of-the-art?* The EM algorithm is presented next, followed by other newly developed methods.

A. Expectation Maximization for Missing Data Imputation

The EM algorithm is a general technique capable of fitting models to incomplete data and capitalizes on the relationship between missing data and the known parameters of a data model. If the missing values were known, then estimating the model parameters would be straightforward. Similarly, if the parameters of the data model were known, then it would be possible to obtain unbiased predictions for the missing values. This interdependence between model parameters and missing values suggests an iterative method where, firstly, the missing values are predicted based on assumed values for the parameters and these predictions are used to update the parameter estimates, and the process is repeated until convergence. The sequence of parameters converges to maximum-likelihood estimates that implicitly average over the distribution of the missing values. In simple terms, EM operates by using an iterative procedure that can be explained as follows [3]:

1) Replace missing data with estimates;
2) Estimate the model parameters of interest;
3) Repeat steps (1) and (2) until convergence.

The key idea that differentiates EM algorithms form any other iterative algorithms is that, missing values themselves are not necessarily estimated by the EM. Instead, the EM only finds the conditional expectations of the missing data using the observed and the estimated parameters [3].

B. Neural Networks Combined with Evolutionary Computing for Missing Data

The method used here combines the use of auto-associative neural networks with an evolutionary computing method. An auto-associative neural network or simply an autoencoder is a neural network that is trained to recall its inputs and is characterised by the number of hiddne nodes which is smaller than the number of inputs. Evolutionary computing techniques such as genetic algorithms and swarm intelligence optimisation are used to approximate missing data as shown in Figure 2. This method has been developed to approximate missing data in a database by Abdella and Marwala [7]. Other evolutionary computing methods have been investigated for this problem by [10]. For illustration purposes, genetic algorithm is used to *estimate* the missing values by optimizing an objective function. The complete vector combining the estimated and the observed values is fed into the autoencoder as input, as shown in Figure 2. Symbols X_k and X_u represent the known variables and the unknown or missing variables, respectively. The combination of X_k and X_u represent the full input space

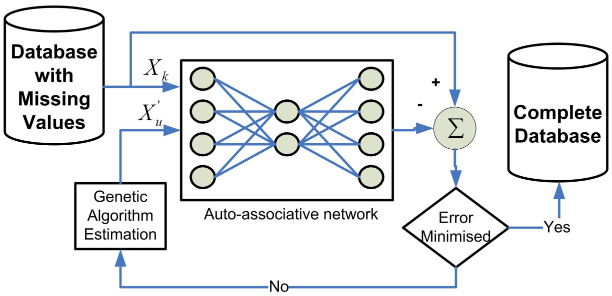

Fig. 2. Autoencoder and GA based missing data estimator structure

Considering that the method uses an autoencoder, one will expect the input to be very similar to the output for a well chosen architecture of the autoencoder. This is, however, only expected on a dataset similar to the problem space from which the inter-correlations have been captured. The difference between the target and the actual output is used as the error and this error is defined as follows:

$$\varepsilon = \vec{x} - f(\vec{W}, \vec{x}) \tag{2}$$

where $\vec{x}$ and $\vec{W}$ are input and weight vectors, respectively. To make sure the error function is always positive, the square of the equation is used. This leads to the following equation:

$$\varepsilon = (\vec{x} - f(\vec{W}, \vec{x}))^2 \tag{3}$$

Since the input vector consist of both the known, X_k and unknown, X_u entries, the error function can be written as follows:

$$\varepsilon = \left(\left\{ \begin{array}{c} X_k \\ X_u \end{array} \right\} - f(\vec{W}, \left\{ \begin{array}{c} X_k \\ X_u \end{array} \right\}) \right)^2 \tag{4}$$

and this equation is used as the objective function that is minimized using evolutionary computing techniques.

C. Ensemble Approaches

Ensemble based approaches are based the assumption that 'two are better than one' and there are many methods in the literature, many of which are based on bootstrapping techniques. These techniques use an ensemble networks to estimate the missing value and use a technique of rewarding or penalizing a particular network using the known performance of the network. A similar approach has been presented by Perrone and Cooper [15] where they showed that the committee of networks can optimize the decision. The final output is then a weighted combination of individual networks that have been used in the prediction. The researcher here still has to find a sensible way of combining the outputs from the ensemble. In discrete values, majority or the weighted majority techniques may be used. There are, however, a few challenges at continuous variables. In some cases, each network is assigned weight based on its performance on the validation set and the weight can be determined in a number of ways. As an example, the weight for some network can be assigned such that:

$$\alpha_i = \frac{1 - E_i}{\sum_{j=1}^{N} (1 - E_i)} \tag{5}$$

where E_i is the estimate of model i's error on the validation set. This kind of weight assignment intuitively ensures that model i gets more weight if its performance is higher than the performance of other models. The objective here is to have a set of models which are likely to have uncorrelated errors [16]. It is also very important to choose an appropriate size of an ensemble as it may increase the complexity of the technique and as a result, might take longer to impute the missing data.

D. Rule-based Imputation via Rough Sets

There is a substantial amount of work in the literature that demonstrates how certain and possible decisions rules may be computed from incomplete decision tables. A well known *Learning from Examples Module* (LEM2) rule induction algorithm [17] has been explored for rule extraction. LEM2 is a component of the *Learning from Examples based on Rough Sets* (LERS) data mining system. The work of [18] and [19] also uses neuro-fuzzy techniques in the presence of missing data.

E. Techniques that Avoid Imputing for Missing Data

There are several methods that do not necessarily impute for missing data. They, rather, classify and regress for decision making in the presence of missing data. There are some other methods that are based on multi-tasked learning, where the decision variable is predicted first and if necessary, the missing data is imputed. Various neural network architectures are have been used in various applications and Nelwamondo and Marwala [20] have developed a Fuzzy ARTMAP for classification tasks with incomplete data. Their approach was also extended to regression using normal multilayer perceptron neural networks. Another approach replaces incomplete patterns with fuzzy patterns. The patterns without missing values are, along with fuzzy patterns, used to train the neural network, which learns to classify without actually predicting the missing data[21].

A thorough review of existing methods for coping with missing data in decision trees is given by Twala [22]. In his research, [22] found the implementation of multiple imputation using EM algorithms to be consistently the best of the existing methods investigated. Nelwamondo et al. [23] compared multiple imputation using EM algorithm to the combination of neural networks and GA and their findings showed that EM algorithm is not any better.

VII. Incompleteness in Time Series Data with Concept Drift

There are many nonstationary quantities in nature that fluctuate with time and their measurements can only be sampled after a certain time period, thereby forming a time series. Common examples are the stock market, weather, heartbeats, seismic waves and animal populations. Engineering examples of such systems include velocity component at a point in a turbulent flow and various measurement systems such as those measuring the heart beat rate [24]. Another example is the electrical transmission system; when this system is pushed to its capacity limit it can exhibit chaotic behavior and failure [24]. There are some engineering and measurement systems that are dedicated to measuring nonstationary quantities. Such instruments are not immune to failures. When instrument or sensor failure occurs, it becomes difficult to estimate the missing values. This difficulty is increased by the chaotic and unpredictable nature of the data. The 2003 Nobel Prize Laureates in Economics, Granger [25] and Engle [26] had an excellent contribution to non-linear data. Although he was not addressing the problem of missing data, Granger showed that the traditional statistical methods could be misleading if applied to variables that wander over time without returning to some long-run resting point [25]. Engle on the other hand had a ground-breaking discovery of Autoregressive Conditional Heteroskedasticity (ARCH), a method to analyse unpredictable movements in financial market prices and also applicable in risk assessment [26].

Some work has been done in dealing with missing data in nonstationary time series [27]. In most cases, attempts are made to make the data stationary using deferencing procedures [27]. In cases of missing data, applying the differencing techniques proposed in literature usually widens the gap or even introduces additional gaps, thereby, not solving the problem of missing data [28]. An interesting method was proposed by [27], which operates by completing missing values at the level of uncorrelated residuals after removing any systematic trends such as periodic components. A sample of the results obtained using their method is presented in Figure 3.

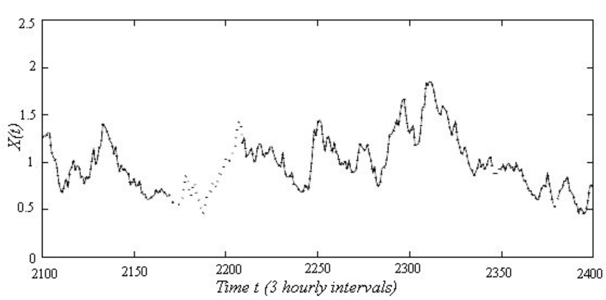

Fig. 3. Incomplete time series (solid line) and the completed one (dotted line) [27]

The major challenge in this regard is that the time series data are usually very unpredictable, most particularly in measurement systems such as those in engineering applications. Furthermore, the frequency at which the data are collected affects the decisions on how to impute for missing data. As an example, an online condition monitoring can be based on the data for monitoring if some plant parameter is healthy and the sampling for this is performed every five seconds. If the sensor fails, depending on the application, it will be essential to have the missing data imputed within the five seconds period to avoid having a long vector of holes in the data. This calls for quick and perhaps inefficient techniques for imputation. However, in some application, there is no need to have quick imputations. Thus, the speed of imputation will vary from one application to the other. Missing data in fast moving data such as the stock market need to be imputed withing the shortest possible time, whereas for political science data such as the infant mortality rate can be dealt with as it suits the users of the data.

VIII. Key Challenges on Missing Data Imputation Through Computational Intelligence Techniques

One of the biggest challenges that have remained with no single answer, depending on who was asked, relates to the size of the dataset. One observation is that smaller datasets are often complete whereas larger datasets are often incomplete or somehow, characterized by outliers and other inaccuracies. Research questions therefore should be aimed at finding if the choice of an imputation method relates to the size of the dataset. Detecting missing data raises another challenge: it is easy to note missing variables, but what happens when the entire observation is missing? Are there techniques that can be used to recover such data?

Lately, there has been an exponential growth in terms of information and machine learning techniques. It seem that having so many tools at our disposal is also becoming a problem. The challenge in this regard is knowing what machine learning tools to use. Moreover, most of these tools can also be combined with other techniques, forming hybrid tools. How then, do we make sure that we have selected the right tools? Is better to use Bayesian statistics compared to maximum likelihood methods?

What optimization techniques should be used? Researchers need to look at the speed of convergence as well as where the convergence occurs (local versus global optimal outputs). Can something be done about the "black-box" nature of such techniques? There are quite a few issues about bounds in evolutionary computing. There are several questions on how appropriate bounds can be found. Furthermore, there are some evolutionary computing techniques that mathematicians find to be mathematically unsound. In this applications, researchers are concerned about the usability of such techniques.

Another challenge comes to the ways outputs are combined from committee or ensemble networks. Majority voting can work in classification tasks, but what should we do in cases of regression? Do we average them and ignore the effect of biasing the answer? In time-series data, one can rest on the assumption that data from some areas such as the political sciences changes very slowly over the years and as a result, the value for this year may not differ much from the value of last year. The onus is therefore left upon the researcher to see which assumptions are valid in their applications. The relationship between the one variable can also correlate to another variable and as a result, it might be necessary to extract such relationships. The challenge here remains to detect the speed of the drift and knowing if the drift is cyclic or gradual.

A great deal of methods have been compared with each other with varying results. There are a few issues of concern on how the comparisons are done. Firstly, there is no standard way of measuring the error between the predictions and the actual values. Errors used by many researchers include, standard errors, sum of the square errors, $|(predicted - actual)|/actual$ often referred as the percentage error, the correlation between the predicted vector and the actual vector, etc. Most of these techniques have less meaning in some applications and do not indicate meaningful results. For example, correlation does not indicate accuracy, but how the predicted vector changes with respect to the actual data. The percentage error may be very strict and harsh when it comes to smaller numbers and allows bigger errors for bigger numbers. While some techniques are more suitable for some applications, for the interest of research, it would be better to have a method that quantifies the performance appropriately. Although much comparison has been done in this regard, not so many researchers have done a test of significance that outlines how significant one methods is better than the other. A mere report on the improvement, eg, 13% better, does not reflect significance at all. It is therefore very important to have such test so that readers can have a deeper insinght into the methods.

IX. Concluding Remarks

This paper has raised several conceptual challenges to the problem of missing data, hence, has broadened the research direction. It is evident that the research direction can still be broadened as there are several questions that need to be answered. Answering some of these questions can provide a rich guideline on how to grow the use of computational intelligence for missing data imputation. It is in the view of research that imputation techniques should not compromise the statistical accuracy of the data and should be conducive to time efficiency.

References

[1] J. L. Schafer and J. W. Graham, "Missing data: Our view of the state of the art," *Psychological Methods*, vol. 7, no. 2, pp. 147–177, 2002.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood for incomplete data via the EM algorithm," *Journal of Royal Statistic Society*, vol. B39, pp. 1–38, 1977.

[3] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*. New York: Wiley, 1987.

[4] D. B. Rubin, *Multiple Imputation for Nonresponse in Surveys*. New York:Wiley, 1987.

[5] J. L. Schafer and M. K. Olsen, "Multiple imputation for multivariate missing-data problems: A data analysts perspective," *Multivariate Behavioural Research*, vol. 33, no. 4, pp. 545–571, 1998.

[6] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag, 2000.

[7] M. Abdella and T. Marwala, "The use of genetic algorithms and neural networks to approximate missing data in database," *Computing and Informatics*, vol. 24, pp. 1001–1013, 2006.

[8] S. Mohamed and T. Marwala, "Neural network based techniques for estimating missing data in databases," in *The 16th Annual Symposium of the Pattern Recognition Association of South Africa*, (Langebaan, South Africa), pp. 27–32, 2005.

[9] W. Qiao, Z. Gao, and R. G. Harley, "Continuous online identification of nonlinear plants in power systems with missing sensor measurements," in *IEEE International Joint Conference on Neural Networks*, (Montreal), pp. 1729–1734, 2005.

[10] S. M. Dhlamini, F. V. Nelwamondo, and T. Marwala, "Condition monitoring of HV bushings in the presence of missing data using evolutionary computing," *WSEAS Transactions on Power Systems*, vol. 1, no. 2, pp. 280–287, 2006.

[11] S. Yu and H. Kobayashi, "A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking," *Signal Processing*, vol. 83, no. 2, pp. 235–250, 2003.

[12] X. Huang and Q. Zhu, "A pseudo-nearest-neighbour approach for missing data recovery on Gaussian random data sets," *Pattern Recognition Letters*, vol. 23, pp. 1613–1622, 2002.

[13] J. Schafer, *Analysis of Incomplete Multivariate Data*. Chapman & Hall, 1997.

[14] A. C. Acork, "Working with missing values," *Journal of Marriage and Family*, vol. 67, pp. 1012–1028, 2005.

[15] P. Perrone and L. N. Cooper, "When networks disagree: Ensemble methods for hybrid neural networks," *Neural Networks for Speech and Image Processing (in R. J. Mammone (Ed))*, pp. 126– 142, 1993.

[16] C. J. Merz, "Using correspondence analysis to combine classifiers," *Machine Learning*, pp. 1–26, 1997.

[17] J. W. Grzymala-Busse, *LERS A system for learning from examples based on rough sets*. Handbook of Applications and Advances of the Rough Sets Theory: Kluwer Academic Publishers, 1992.

[18] D. Nauck and R. Kruse, "Learning in neuro-fuzzy systems with symbolic attributes and missing values," in *Proceedings of the IEEE International Conference on Neural Information Processing*, (Perth), pp. 142–147, 1999.

[19] B. Gabrys, "Neuro-fuzzy approach to processing inputs with missing values in pattern recognition problems," *International Journal of Approximate Reasoning*, vol. 30, pp. 149–179, 2002.

[20] F. V. Nelwamondo and T. Marwala, "Fuzzy artmap and neural network approach to online processing of inputs with missing values," *SAIEE Africa Research Journal, (appearing in a special issue)*, 2007.

[21] S. Wang, "Classification with incomplete survey data: a Hopfield neural network approach," *Computers & Operations Research*, vol. 24, pp. 53–62, 2005.

[22] B. E. T. H. Twala, *Effective Techniques for Handling Incomplete Data Using Decision Trees*. PhD thesis, The Open University, UK, 2005.

[23] F. V. Nelwamondo, S. Mohamed, and T. Marwala, "Missing data: A comparison of neural networks and expectation maximization techniques," *Current Science*, vol. 93, no. 11, pp. 1514–1521, 2007.

[24] D. L. Turcotte and J. B. Rundle, "Self-organized complexity in the physical, biological and social sciences," in *Proceedings of the National Academy of Sciences*, (USA), pp. 2463–2465, PNAS, 2002.

[25] C. W. J. Granger, "Time series analysis, cointegration and applications," *Nobel Price Lecture*, pp. 360–366, 2003.

[26] R. Engle, "Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation," *Econometrica*, vol. 50, pp. 987–1008, 1982.

[27] C. Stefanakos and G. A. Athanassoulis, "A unified methodology for analysis, completion and simulation of nonstationary time series with missing values, with application to wave data," *Applied Ocean Research*, vol. 23, pp. 207–220, 2001.

[28] G. M. Ljung, "A note on the estimation of missing values in time series," *Commun Statist*, vol. 18, no. 2, pp. 459–465, 1989.