

Predicting Septic Shock Outcomes in a Database with Missing Data using Fuzzy Modeling

Influence of Pre-processing Techniques on Real-World Data-based Classification

Rúben D.M.A. Pereira,
André S. Fialho, Federico Cismondi
MIT-Portugal Program, Bio-Engineering Systems
MIT / HMS / IST
Cambridge, MA, USA / Lisbon, Portugal

Rui J. Almeida, Uzay Kaymak
Erasmus School of Economics, Econometric Institute
Erasmus University Rotterdam
Rotterdam, The Netherlands

Susana M. Vieira, João M.C. Sousa
Dept. of Mechanical Engineering, CIS/IDMEC-LAETA
Instituto Superior Técnico (IST)
Lisbon, Portugal

Shane R. Reti, Michael D. Howell
Beth Israel Deaconess Medical Center
Harvard Medical School (HMS)
Boston, MA, USA

Stan N. Finkelstein
Engineering Systems Division and Harvard-MIT Division of Health Sciences & Technology
Massachusetts Institute of Technology (MIT)
Cambridge, MA, USA

Abstract – Real-world databases often contain missing data and existing correction algorithms deliver varying performance. Also, most modeling techniques are not suitable to deal with them automatically. In this study we examine different approaches to predicting septic shock in the presence of missing data. Some pre-processing techniques for managing missing data include disregarding data, or replacing it with information that by design introduces bias. In this study, we show that predictive performance improves by employing a minimum pre-processing technique, the Zero-Order-Hold (ZOH) method, by applying a Fuzzy C-Means clustering technique based on the partial distance calculation strategy (FCM-PDS) and by computing the final classification regarding the samples from each patient. Performance improvements continue to occur where up to approximately 60% of the data is missing, though for higher percentage the classification performance still is statistically improved. We further validate this approach by making comparisons with previous studies.

Missing Data; Septic Shock; Fuzzy C-Means; Partial Distance Strategy; Zero-Order-Hold

I. INTRODUCTION

Severe sepsis remains both an important clinical challenge and an economic burden in intensive care. This is especially so for septic shock, which is characterized by the persistence of a

hypotensive state, despite adequate fluid resuscitation [1]. The mortality rate among septic shock patients is very high (about 50%), as are the comparative costs of treatment. Cost-of-illness studies in the US focusing on direct costs per sepsis patient support the premise of, and attention to, increasing global health care costs [2]. Prevention and quality are therefore embraced as cost efficient contributors to a solution, and it so the ability to predict and prevent expensive septic shock hospitalizations.

A review of the literature highlights previous studies in this area that have applied knowledge-based soft computing techniques to various scenarios associated with septic shock [3, 4, 5, 6]. In [5], however, feature selection is applied in order to reduce data dimensionality by removing inputs that do not improve the prediction performance of the model.

The aim of this study is explore the influence of pre-processing and new modeling techniques for the prediction of septic shock. By using two sets of 12 and 28 selected features from the pre-processed MEDAN database [6] described in [4] and [5], respectively, results were compared to these studies.

Real-world databases often contain many missing values and artifacts, either due to lack of sampling or post-acquisition processing techniques (e.g. re-sampling). Most learning rule-based algorithms deal poorly with this. However, different pre-processing techniques can be applied to minimize this issue. In some cases, if the amount of missing data is small, it is possible to delete all incomplete data. In other cases, it can be more useful to attempt to find suitable replacement values.

This work is supported by the Portuguese Government under the programs: Project PTDC/SEM-ENR/100063/2008, Fundação para a Ciência e Tecnologia (FCT), and by the MIT-Portugal Program and FCT grants SFRH/BD/51028/2010, SFRH/43043/2008 and SFRH/43081/2008.

There are two types of missing data: missing not at random (MNAR) and missing at random (MAR) [7]. The second group can be further divided into missing completely at random (MCAR) and general MAR. MCAR values can be interpreted as a random reduction of the dataset, because their probability does not depend on the observed data or the unobserved data [8, 9]. Thus, they provide no further information for assigning incomplete features vectors to clusters. Alternatively, MAR values are not randomly distributed across all observations, but are randomly distributed within one or more subsamples instead. This is the scenario for the database used in this study.

Having discussed the unsatisfactoriness of disregarding missing data, replacing missing values creates problems of bias with the introduction of new data [10]. An alternative then to these two methods is to extend the data analysis method so that the method can deal directly with incomplete data. In this work we use a modified fuzzy c-means (FCM) clustering method that allows the identification of Takagi-Sugeno fuzzy models from incomplete data by means of product-space clustering [11]. This method uses the extension of the Fuzzy C-Means algorithm with the Partial Distance Strategy (FCM-PDS), discussed in [12, 13], and does not require any kind of imputation or iterative guess-estimation of the missing values. The algorithm possesses all the standard convergence properties of FCM. Additionally, both the global convergence theory and the local alternating optimization convergence theory are applicable [13]. Previous works [11] showed that the performance of these models is comparable to the one obtained when using a complete data set [11] making it a suitable tool for the present study.

The modeling technique used will be briefly discussed in Section II. In Section III the methodology is described followed by results and discussion in Section IV. Section V concludes the paper.

II. MODELING

Takagi-Sugeno (TS) fuzzy models can be used as universal approximators in non-linear studies as this.

A. Takagi-Sugeno Fuzzy Modeling

A TS fuzzy model [14] is a fuzzy rule-based model where the rule consequents are crisp functions of the model input:

$$R^k : \text{If } \mathbf{x} \text{ is } A^k \text{ then } y^k = f^k(\mathbf{x}), k=1, 2, \dots, K, \quad (1)$$

where R^k denotes the k -th rule, K is the number of rules, $\mathbf{x}$ is the antecedent variable, y is the one dimensional consequent variable and A^k are the (multidimensional) antecedent fuzzy sets of the k -th rule. Each rule k has a different function f^k yielding a different value y^k for the output. The most simple consequent function f^k in (1) is the linear affine form:

$$R^k : \text{If } \mathbf{x} \text{ is } A^k \text{ then } y^k = (\mathbf{a}^k)^T \mathbf{x} + b^k \quad (2)$$

where $\mathbf{a}^k$ is a vector of parameter and b^k is a scalar offset. This model is called an affine TS model.

B. Fuzzy Modeling with Missing Values

Several techniques can be used in fuzzy identification. One possibility is to use identification by product-space clustering [15]. This principle involves approximating a nonlinear problem by decomposing it into several sub-problems. In general, the identification of fuzzy models is solved in two steps: structure identification and parameter estimation. Structure identification can be obtained by fuzzy clustering. Parameter identification consists of determining the parameters for the antecedent functions and the parameters of the rule consequents.

With missing values, product-space clustering can also be used as long as the distance metric is modified in order to deal with those values. It is possible to calculate partial distances using all non-missing feature values, and scaling afterwards this quantity by the reciprocal of the proportion of components used [16]. An example of the D_{ik}^2 calculation of the Euclidean distance is [13]:

$$\begin{aligned} D_{ik}^2 &= \|z_k - v_i\|_2^2 \\ &= \left\| \begin{pmatrix} ? & 2 & ? & ? & 5 \end{pmatrix}^T - \begin{pmatrix} 6 & 7 & 8 & 9 & 10 \end{pmatrix} \right\|_2^2 \\ &= \frac{5}{5-3} \left((2-7)^2 - (5-10)^2 \right) \end{aligned} \quad (3)$$

The general formula for the partial distance calculation of D_{ik}^2 is given by:

$$D_{ik}^2 = \frac{n}{I_k} \sum_{j=1}^n (x_{kj} - v_{ij})^2 I_{kj} \quad (4)$$

where

$$I_{kj} = \begin{cases} 0, & \text{if } x_{kj} \in X_M \\ 1, & \text{if } x_{kj} \in X_P \end{cases}, \text{ for } 1 \leq j \leq n \text{ and } 1 \leq k \leq N \quad (5)$$

$$\text{and } I_k = \sum_{j=1}^n I_{kj}$$

X_M and X_P are the subsets of X containing the missing and the present values from the total datum, respectively. This is the partial distance strategy (PDS) mentioned before and the operational basis of modified FCM the algorithm used in this study. The FCM-PDS is the regular FCM algorithm where the calculation of the distance D_{ik}^2 is substituted by (4) and also the cluster prototype centers, v_{ij} , are calculated according to:

$$v_{ij} = \frac{\sum_{k=1}^N (\mu_{ik})^m I_{kj} x_{kj}}{\sum_{k=1}^N (\mu_{ik})^m I_{kj}} \quad (6)$$

where μ_{ik} is the membership degree of x_k to the cluster center v_i .

The antecedent membership functions can be obtained by projecting the multidimensional fuzzy sets defined point wise in the rows of the partition matrix of the model onto the individual antecedent variables of the rules. However, by computing the membership degrees directly in the space of the antecedent variables, each cluster represents one fuzzy rule, as in (2). The multidimensional membership functions A^k are given analytically by computing the distance from the projection of the cluster center v_i onto X , and then computing the membership degree, $\beta_i(\mathbf{x})$, in an inverse proportion to the distance, with the form:

$$\beta_i(\mathbf{x}) = \left[\sum_{j=1}^c \left(D_{ik}^2(\mathbf{x}, v_i^x) / D_{jk}^2(\mathbf{x}, v_j^x) \right)^{\frac{2}{m-1}} \right]^{-1} \quad (7)$$

The consequent parameters for each rule are estimated by the least-squares method. The set of optimal parameters with respect to the model output is estimated by minimizing the global prediction error, taking into account the aggregation of the rules [11].

III. METHODOLOGY

A. MEDAN Database Pre-processing

The pre-processed MEDAN database [6] originates from 71 German intensive care units (ICUs) using several data quality techniques [3] (e.g. data re-sampling, errors filtration, standardization of measuring units). However, to use the database for our objectives, several pre-processing methodologies are required such as identification and removal of outliers and management of missing data.

In [4], data from a suitable normal distribution (noise) was inserted randomly so that the algorithms could not learn from the added noise data. In particular, the aim was to prevent situations where data with missing value replacements might be alternatively classified [17]. Contrastingly, a linear regression strategy followed by the Zero-Order-Hold (ZOH) method was used in [5]. The ZOH method can be applied causally by holding in time a sample value until a new one is acquired. This was done for each feature along the time. In the case where no value is present in the initial time stamp of a patient, the dataset will contain missing values until one value is acquired for that feature. This method constitutes a suitable pre-processing technique for medical data records as we are dealing with time-dependent signals series, many of them with low changing frequencies [5]. Schematically, the dataset preprocessing will happen as follows:

(Without ZOH pre-processing)

$$original\ dataset = \begin{bmatrix} ? & ? & ? & 13 & 17 & 21 & ? & 29 \\ 2 & ? & 10 & ? & ? & 22 & ? & ? \\ 3 & 7 & ? & ? & 19 & ? & ? & ? \\ 4 & 8 & 12 & 16 & ? & ? & 28 & ? \end{bmatrix} \quad (8)$$

(With ZOH pre-processing)

$$ZOH\ dataset = \begin{bmatrix} ? & ? & ? & 13 & 17 & 21 & ? & 29 \\ 2 & ? & 10 & 13 & 17 & 22 & ? & 29 \\ 3 & 7 & 10 & 13 & 19 & 22 & ? & 29 \\ 4 & 8 & 12 & 16 & 19 & 22 & 28 & 29 \end{bmatrix} \quad (9)$$

Approximately 24h spaced samples were used for comparison with [4]. Though in [4] the daily mean values (after input of missing data) were used, comparison with this work still is valid. Two studies were performed here: Predicting septic shock by applying the FCM-PDS method without any pre-processing of the dataset (FCM-PDS); and predicting septic shock by applying the FCM-PDS method with the ZOH pre-processing method (ZOH-FCM-PDS). Thus, for the first study the datasets contain all the missing data from the original MEDAN database. For the second, the datasets will only contain missing data until at least one value is acquired in every feature, in the original database.

From these studies we can compare the use of the FCM-PDS algorithm with the results obtained in [4] and evaluate the differences in performance when using a suitable pre-processing technique.

B. Experimental Setup

This study makes use of a publicly available database [6], containing data from a total of 382 patients with septic shock associated with abdominal causes. For each patient, the database holds personal records, physiological parameters, diagnoses/therapies, medication, and respective outcomes (survived or deceased).

As in previous works we focused exclusively on the physiological parameters (a total of 103 features). However, only 139 patients were found to meet initial inclusion criteria and not all of them have entries for the 103 features [3, 4]. From the total features available then, two different datasets of 12 features [4] and 28 features [5] were selected to perform the studies, resulting in a total of 121 patients and 81 patients in each set, respectively.

The models were iteratively built (100 times for statistical purposes) for the two sets of features and the performance of the models was evaluated in terms of accuracy (correct classification rate), sensitivity (true positive classification rate) and specificity (true negative classification rate), and area under the ROC (Receiver Operating Characteristic) curve (AUC):

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (11)$$

$$Specificity = \frac{TN}{TN + FP} \quad (12)$$

Where TP is the number of true positive classifications obtained, FP is the number of false positive classifications, TN is the number of true negative classifications and FN is the number of false negative classifications. The ROC curve is obtained by plotting the true positive classifications rate (TPR, or sensitivity) against the false positive classifications rate (FPR, or 1-specificity) obtained for a binary classifier system as its discrimination threshold is varied. The area under that curve is calculated as:

$$AUC = \int_0^1 ROC = \int_0^1 TPR.dFPR \quad (13)$$

The train and test sets in each set of features were obtained by randomly dividing the available data (50% train and 50% test). These subsets were maintained the same throughout the simulation.

The results were then compared with [4], where a neural-network-based modeling approach is used to classify abdominal septic shock patients using the same subset of 12 features that was used in this work.

C. Computing Classification

When performing the first tests with the FCM-PDS algorithm the accuracy obtained was very similar to that obtained in [4] (in contrast, sensitivity and specificity were balanced). This is not unexpected as neither of the pre-processing techniques adds any additional information to the missing dataset. Therefore, after testing the model created with the test set, the results were aggregated so that the predicted classification for each patient results from the mean of all its predicted classifications for each time samples. Schematically, consider the dataset configuration used:

$$\begin{bmatrix} id & time & feature1 & \dots & featureN & Classification \\ id_{p1} & t_{p1,1} & x_{1,p1,1} & \dots & x_{N,p1,1} & class_{p1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ id_{p1} & t_{p1,T1} & x_{1,p1,T1} & \dots & x_{N,p1,T1} & class_{p1} \\ id_{p2} & t_{p2,1} & x_{1,p2,1} & \dots & x_{N,p2,1} & class_{p2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ id_{p2} & t_{p2,T2} & x_{1,p2,T2} & \dots & x_{N,p2,T2} & class_{p2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ id_{pk} & t_{pk,1} & x_{1,pk,1} & \dots & x_{N,pk,1} & class_{pk} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ id_{pk} & t_{pk,Tk} & x_{1,pk,Tk} & \dots & x_{N,pk,Tk} & class_{pk} \end{bmatrix} \quad (14)$$

Where id_{pk} is the identifier of patient pk , x_N is the value of feature N , Tk is the number of time stamps for patient pk , and $class_{pk}$ is the observed classification. The clustering method

treats each time sample individually. Consequently, we can compute the predicted classification of each patient using the predicted classification for each of its time stamps:

$$class_{pk,final}^{\wedge} = \frac{1}{Tk} \sum_{i=1}^{Tk} class_{pk}^{\wedge}(t_{pk,i}) \quad (15)$$

This procedure maintains some of the time series information by not treating each time sample as an individual sample. This method was used in both approaches (with and without pre-processing) presented in this work.

IV. RESULTS AND DISCUSSION

The 12 feature and 28 feature datasets created without pre-processing contained approximately 54% and 67% missing data, respectively. In contrast, same datasets obtained using the ZOH method only contained 2% and 4% missing data, respectively (the remaining missing values are placed in the initial positions of each feature (9)).

In Fig.1 one can observe the performance test of the ZOH-FCM-PDS method for the 12 feature dataset. The threshold value selected for comparison was one for which accuracy is near the maximum, and sensitivity and specificity are relatively balanced. In this case we selected 0.35.

Table I presents the percentage of correct classifications achieved by each of the approaches. Although the ZOH-FCM-PDS approach performs better than the other methods for the 12 feature dataset, the differences between this method and the FCM-PDS method become faded for the 28 feature dataset. In Table II, sensitivity and specificity are presented for the 12 features dataset studies and, again, the best results for these criteria are achieved by the ZOH-FCM-PDS method. This method performs well overall as the missing data percentage decreases. In Table III, where the results for sensitivity and specificity for the studies using the 28 feature dataset are presented, the clearest differences between the two approaches are the standard deviations. The mean values are very similar which means that the FCM-PDS approach is not as robust as it is when preceded by the pre-processing. This is also observed in Table I and is due to the fact that we are not dealing with MCAR values, and therefore the configuration of the dataset influences the results drastically. This effect is best observed when dealing with the 28 feature dataset (67% missing values downgraded to 4%) where, by replacing the great majority of the missing values using the ZOH method, the model becomes more robust maintaining its prediction power. This means that the information contained in each time sample is related to the previous ones, as expected by physicians, otherwise the ZOH-FCM-PDS method would perform worse overall. Additionally, this reinforces the observations that for percentages of missing data up to approximately 70% the FCM-PDS model can perform well though the mean error increases as one approaches upper limits [11].

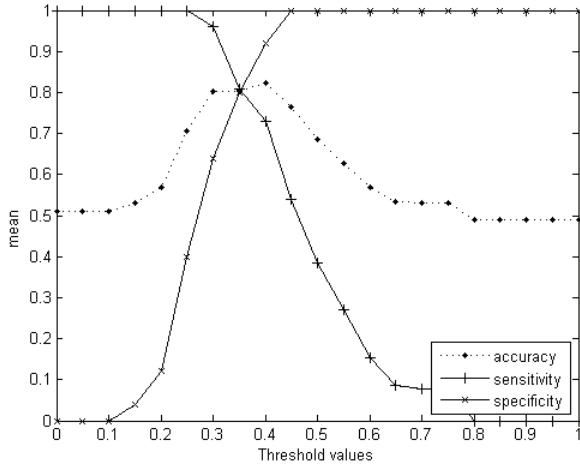


Figure 1. Performance of the ZOH-FCM-PDS method using the 12 feature dataset. The threshold value chosen for presenting the results here was 0.35.

TABLE I. ACCURACY (PERCENTAGE OF CORRECT CLASSIFICATIONS)

	12 feature		28 feature	
	Mean	Std	Mean	Std
NN [4]	69.00	4.37	-	-
ZOH-FCM-PDS	80.39	0.00	70.59	0.00
FCM-PDS	69.39	0.00	64.76	7.73

TABLE II. SENSITIVITY AND SPECIFICITY USING 12 FEATURES

	12 feature			
	Sensitivity		Specificity	
	Mean	Std	Mean	Std
NN [4]	15.01	-	92.26	-
ZOH-FCM-PDS	80.77	0.00	80.00	0.00
FCM-PDS	76.00	0.00	62.50	0.00

TABLE III. SENSITIVITY AND SPECIFICITY USING 28 FEATURES

	28 feature			
	Sensitivity		Specificity	
	Mean	Std	Mean	Std
ZOH-FCM-PDS	66.67	0.00	73.68	0.00
FCM-PDS	65.67	12.98	63.75	24.03

The method used in [4] performs similarly in terms of accuracy to FCM-PDS method. However, for sensitivity and specificity the approach used in this study clearly performs better.

In Table IV the AUC values obtained for the two datasets of features when using the two approaches discussed are presented. Again, the use of the ZOH pre-processing method constitutes a significant difference for the 12 feature dataset, resulting in better AUC values. The variability of the results due to the high percentage of missing data is noticeable for the 28 feature dataset. Fig.2 permits graphical comparison of the ROC curves obtained for all the studies. Both methods perform similarly for the 28 feature dataset.

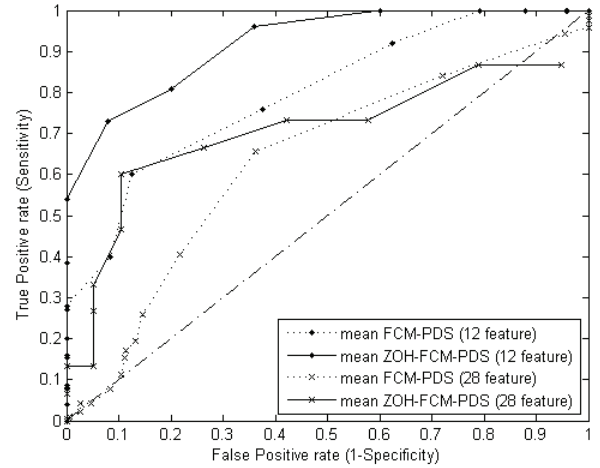


Figure 2. ROC curves obtained for the 12 and 28 feature datasets studies.

The technique used in this study to compute the predicted classification of each patient clearly improves the classification when compared to results in [4]. Although the accuracy of both is similar, the differences can be seen in Table II as the balance between sensitivity and specificity. As mentioned before, both [4] and the FCM-PDS approach do not use pre-processing techniques that could bias the information contained in the dataset.

TABLE IV. AUC

	12 feature		28 feature	
	Mean	Std	Mean	Std
ZOH-FCM-PDS	89.85	0.00	64.91	0.00
FCM-PDS	78.58	0.00	63.14	11.73

In [4], a potential improvement to the classification performance was to use only the three last 24h time samples from each patient. However, this work demonstrates a viable alternative method, and notes that other techniques could also be used in order to favor the positive or negative classifications, for example, as in the “balanced accuracy”. Computing the classification of each patient after using the FCM-PDS method in datasets containing limited time windows (e.g. last three days) could constitute a better improvement to the septic shock prediction study.

In reference [17] missing values are replaced by centers of gravity associated with fuzzy rules. This procedure unjustly improves the statistical properties of the original data by replacing the missing values in the best possible manner for the system [3, 4]. However, such a replacement strategy, like performing a linear regression in the existing values [5], offers some advantages that warrant further examination and comparison in the development of more robust models (e.g. applications in data-based knowledge extraction).

V. CONCLUSIONS

Real-world databases usually contain missing values. Using the MEDAN database we studied the application of a modified fuzzy c-means (FCM) clustering method that allows the identification of Takagi-Sugeno fuzzy models from incomplete data, without using complex pre-processing methods. This paper demonstrates a real-time modeling strategy for medical data that is reliable, with high fidelity and non-complex interpretability. Applying the models developed in this study with time limited physiological datasets and suitable pre-processing techniques demonstrates the opportunities to improve classification performance and risk prediction in clinical settings.

REFERENCES

- [1] American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference, "Definitions for sepsis and organ failures guidelines for the use of innovative therapies in sepsis," *Critical Care Medicine*, vol. 20, pp. 864–874, 1992.
- [2] H. Burchardi and H. Schneider, "Economic aspects of severe sepsis: a review of intensive care unit costs, cost of illness and cost effectiveness of therapy," *Pharmacoeconomics*, vol. 22, no. 12, pp. 793–813, 2004.
- [3] J. Paetz, B. Arlt, K. Erz, K. Holzer, R. Brause and E. Hanisch, "Data quality aspects of a database for abdominal septic shock patients," *Computer Methods and Programs in Biomedicine*, vol. 75, pp. 23–30, 2004.
- [4] J. Paetz, "Knowledge-based approach to septic shock patient data using a neural network with trapezoidal activation functions," *Artificial Intelligence in Medicine*, vol. 28, pp. 207–230, 2003.
- [5] A. S. Fialho and F. Cismonti, et al., "Predicting outcomes of septic shock patients using feature selection based on soft computing techniques," *Communications in Computer and Information Science*, Part II, vol. 81, pp. 65–74, 2010.
- [6] E. Hanisch, R. Brause, B. Arlt, J. Paetz, K. Holzer, "The MEDAN Database," 2003 (<http://www.medan.de>).
- [7] H. Timm, C. Döring, and R. Kruse, "Differentiated treatment of missing values in fuzzy clustering," in *Proceedings of the 10th International Fuzzy Systems Association World Congress, Lecture Notes in Artificial Intelligence*, T. Bilgic, B. D. Baets, and O. Kaynak, Eds. Istanbul, Turkey: Springer-Verlag, vol. 2715, June 2003, pp. 354–361.
- [8] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed. New York: Wiley and Sons, 2002.
- [9] J. L. Schafer, *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall, 1997.
- [10] A. Farhangfar, L. Kurgan and J. Dy, "Impact of imputation of missing values on classification error for discrete data," *Pattern Recognition*, vol. 41, no. 12, pp. 3692–3705, 2008.
- [11] R. J. Almeida, U. Kaymak and J. M. C. Sousa, "A new approach to dealing with missing values in data-driven fuzzy modeling," *IEEE International Conference on Fuzzy Systems*, pp. 1–7, July 2010.
- [12] H. Timm and R. Kruse, "Fuzzy cluster analysis with missing values," *Conference of the North American Fuzzy Information Processing Society*, pp. 242–246, Aug 1998.
- [13] R. J. Hathaway and J. C. Bezdek, "Fuzzy c-means clustering of incomplete data," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 31, no. 5, pp. 735–744, October 2001.
- [14] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.
- [15] R. Babuska and H. B. Verbruggen, "Constructing fuzzy models by product space clustering," in *Fuzzy Model Identification: Selected Approaches*, H. Hellendoorn and D. Driankov, Eds. London, UK: Springer-Verlag, 1997, pp. 53–90.
- [16] J. Dixon, "Pattern recognition with partly missing data," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 617–621, 1979.
- [17] M. R. Berthold, "Fuzzy-models and potential outliers," in *Proceedings of the 18th International Conference of the North America Fuzzy Information Processing Society*, R. N. Dave, T. A. Sudkamp, Eds. New York, USA, Los Alamitos: IEEE Press, 1999, pp. 532–535.