A Preliminary Study on Missing Data Imputation in Evolutionary Fuzzy Systems of Subgroup Discovery

C. J. Carmona
Department of Computer Science
University of Jaen
Jaen, Spain
ccarmona@ujaen.es

J. Luengo
Department of Civil Engineering
University of Burgos
Burgos, Spain
jluengo@ubu.es

P. González and M. J. del Jesus
Department of Computer Science
University of Jaen
Jaen, Spain
pglez@ujaen.es | mjjesus@ujaen.es

***Abstract*—In real-life data, a loss of information is frequent in data mining due to the presence of missing values in the attributes. Missing values can occur due to problems in the manual data entry procedures, equipment errors or incorrect measurements. The presence of missing values in attributes conditions the results obtained by any knowledge extraction approach. Specifically, this problem could lead in subgroup discovery to a loss of quality of results obtained by subgroups on measures such as sensitivity, confidence, significance or unusualness.**

This paper presents an experimental study to analyse the effect of different missing data imputation mechanisms within subgroup discovery algorithms based on evolutionary fuzzy systems presented throughout the literature. The analysis is carried out with a large number of data sets obtained from KEEL repository. Among all the imputation techniques, the imputation method *K-Nearest Neighbour* outstands as the best option. In summary, if experts need to analyse a problem with a high percentage of missing values they must use this imputation method in order to treat data in a correct way and also to obtain a meaningful descriptive knowledge. In addition, results also show that the evolutionary fuzzy system with the best results is the algorithm NMEEF-SD in the missing values scenario.

***Index Terms*—Subgroup Discovery, Evolutionary Fuzzy System, Missing Data Imputation.**

I. Introduction

Collecting data in real-world applications is not perfect. This leads to the appearance of imperfections in the datasets. One of the most common imperfections that can be usually found are missing values (MVs). The existence of MVs comes from different causes, such as manual data entry procedures, equipment errors and incorrect measurements. The presence of MVs in data mining produce several problems in the knowledge extraction process [1]:

- loss of efficiency,
- complications in to manage and analyse data, and
- bias resulting from differences between missing and complete data.

In order to avoid these negative effects in the analysis of data mining algorithms when MVs are present, different approaches are employed to prepare and clean data. This is critical as many existing industrial and research data sets contain MVs. In the specialised literature the presence of such imperfections requires a pre-processing stage in which data is prepared and cleaned [2], in order to be useful and sufficiently clear for the knowledge extraction process. This step improves the knowledge extraction process and, in consequence, the results obtained by any data mining algorithm. The simplest way of dealing with this problem, to discard the examples containing MVs, maybe inappropriate. This approach is practical only when data contains a relatively small number of examples with MVs and analysis of the complete examples does not lead to serious bias during the inference [3]. Thus, more sophisticated methods have been proposed in order to deal with MVs, where imputation methods are the most common ones due to their independence with respect to the data mining algorithm applied afterwards.

The aim of this contribution is to analyse the effectiveness of different MVs pre-processing methods in the subgroup discovery task. Subgroup discovery is a descriptive data mining technique but using supervised learning, i.e. subgroup discovery [4], [5] is a descriptive technique that attempts to extract knowledge respect to a variable of interest where unusual and interesting relationships between data with respect to a class are obtained. This technique is also affected by the presence of MVs in data, which could lead to a loss of precision of the model obtained.

An experimental study with subgroup discovery algorithms based on the evolutionary fuzzy systems presented throughout literature is carried out. Evolutionary fuzzy systems are fuzzy systems augmented by a learning process based on evolutionary computation [6]. Usually this kind of systems considers a model structure in the form of fuzzy rules. The resulting systems are called fuzzy rule based systems, which have demonstrated their ability with respect to different problems like control problems, modelling, classification or data mining in a large number of applications.

The results and conclusions obtained in this paper present to the experts the best MVs pre-processing method for evolutionary fuzzy systems in subgroup discovery task, where preprocessing method with the best results will be the optimal method in order to clean and pre-process data with MVs when experts need to apply subgroup discovery algorithm based on evolutionary fuzzy systems in a real-world application.

The rest of the contribution is organised as follows. Section II presents the definition of subgroup discovery, main properties, evolutionary fuzzy systems and quality measures employed. Section III shows a brief review about the use of MVs in the literature. Section IV presents the experimental study performed with evolutionary fuzzy systems for subgroup discovery task and different MVs pre-processing methods. Finally, some concluding remarks are outlined in Section V.

II. Subgroup discovery

The concept of subgroup discovery was initially introduced by Kloesgen [4] and Wrobel [5], and more formally defined by Siebes [7] but using the name Data Surveying for the discovery of interesting subgroups. It can be defined as [8]:

> *"In subgroup discovery, we assume we are given a so-called population of individuals (objects, customer, ...) and a property of those individuals we are interested in. The task of subgroup discovery is then to discover the subgroups of the population that are statistically "most interesting", i.e., are as large as possible and have the most unusual statistical (distributional) characteristics with respect to the property of interest."*

Considering this definition, the main property of this task is the search of partial relations where the majority of the examples for the property of interest (or target variable) will be covered. In addition, the relations must be interesting with an unusual behaviour respect to the full data set.

In order to represent the knowledge, subgroup discovery employs a rule (R) which consists of an induced subgroup description. It can be formally defined as:

$$R : Cond \ \rightarrow \ TargetVar$$

where $TargetVar$ is a value for the variable of interest (target variable) for the subgroup discovery task (which also appears as $Class$ in the literature), and $Cond$ is commonly a conjunction of features (attribute-value pairs) which is able to describe an unusual statistical distribution with respect to the $TargetVar$.

In Fig. 1 is represented a subgroup with two values for the target variable ($TargetVar = o$ and $TargetVar = x$). In this representation a subgroup for the first value of the target variable can be observed, where the rule attempts to cover a high number of objects with a single function, a circle in this case. As can be observed, the subgroup does not cover all the examples of the target value o even not all the examples covered belong to the target vaule, but the function is uniform and simple.

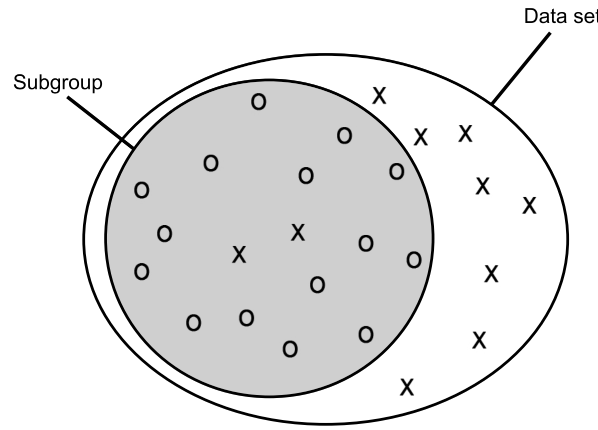

Fig. 1. Representation of a subgroup discovery rule with respect to a value (o) of the target variable

Throughout the literature, a wide number of algorithms have been presented for the subgroup discovery task [9], as for example proposals based on adaptations of classification algorithms, based on association rules algorithms or evolutionary fuzzy systems for subgroup discovery. This paper is focused in evolutionary fuzzy systems. Next, the evolutionary fuzzy systems for subgroup discovery and the quality measures employed by these algorithms are presented.

A. Evolutionary Fuzzy Systems in Subgroup Discovery

Subgroup discovery is a task which can be solved as optimisation and search problems. Due to the fact that evolutionary algorithms [10] in general imitate the principles of natural evolution in order to form searching processes, they can be used in this task. The heuristic used by this type of algorithm is defined by a fitness function, which determines which individuals (rules in this case) will be selected in the competition process. This makes genetic algorithms very useful for the subgroup discovery task.

On the other hand, fuzzy systems are one of most important areas for the application of the fuzzy set theory [11]. Usually these systems consider a model structure in the form of fuzzy rules. The use of these systems in the algorithms avoids the need to perform a previous crisp discretisation to analyse data, because this previous step could lead to a loss of information in the model obtained. The interpretability of the rules is improved because the experts can study the behaviour of different properties of the problem with linguistic labels, depending on the definition of the problem, instead of numbers or intervals. Specifically, uniform partitions with triangular membership functions for continuous variables are employed in this work, as shown in Fig. 2 for a variable with three linguistic labels: *Low*, *Medium* and *High*.

Fuzzy systems augmented by a learning process based on evolutionary computation are called evolutionary fuzzy systems [12], within evolutionary computation can be found

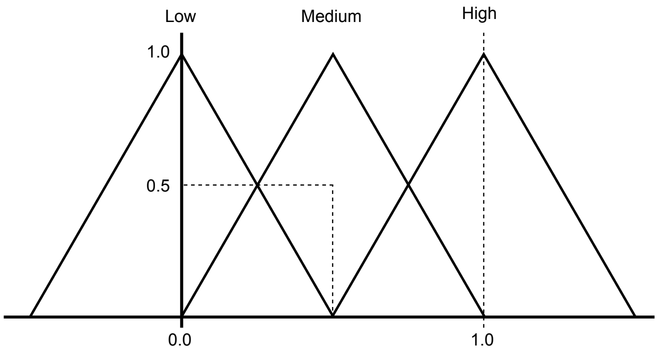

Fig. 2. Example of fuzzy partition for a continuous variable with three labels

genetic algorithms, genetic programming and evolutionary strategies, among others [13]. This type of systems for subgroup discovery task have demonstrated throughout the literature its utility in different real-world problems [14]–[18].

Different evolutionary fuzzy systems for subgroup discovery task can be found in the literature. These algorithms are briefly described below:

- SDIGA [14] is an evolutionary fuzzy rule induction system based on a mono-objective evolutionary algorithm where an aggregation function with different objectives is used like fitness.
- MESDIF [19] is a multi-objective evolutionary algorithm for the extraction of fuzzy rules which describe subgroups. The algorithm extracts a variable number of different rules expressing information on a single value of the target variable. The search is based on the multi-objective SPEA2 [20] approach.
- NMEEF-SD [21] is an evolutionary fuzzy system whose objective is to extract descriptive fuzzy and/or crisp rules for the subgroup discovery task, depending on the type of variables present in the problem. NMEEF-SD has a multi-objective approach whose search strategy is based on NSGA-II algorithm [22].

An important property of these algorithms is that they are able to obtain crisp and/or fuzzy rules depending on the nature of the data.

B. Quality Measures of Evolutionary Fuzzy Systems for Subgroup Discovery

One of the most important aspects in the development of subgroup discovery algorithms is the use of quality measures, both to guide the search process by the algorithms and to evaluate the quality of the subgroups obtained. In this paper are considered the following quality measures:

- *Significance*. This measure indicates the significance of a finding, it measured by the likelihood ratio of a rule [4].

$$Sign(R) = 2 \cdot \sum_{k=1}^{n_c} n(TargetVar_k \cdot Cond) \cdot \\ log \frac{n(TargetVar_k \cdot Cond)}{n(TargetVar_k) \cdot p(Cond)} \quad (1)$$

where $n(TargetVar \cdot Cond)$ is the number of examples which satisfy the conditions and also belong to the value for the target variable in the rule, $n(TargetVar)$ are all the examples of the target variable, $p(Cond)$ is used as a normalised factor, and n_c is the number of values of the target variable. It must be noted that although each rule is for a specific $TargetVar$, the significance measures the novelty in the distribution impartially, for all the values.

- *Unusualness*: This measure is defined as the weighted relative accuracy of a rule [23]. It can be computed as:

$$Unus(R) = \frac{n(Cond)}{n_s} \cdot \\ \left(\frac{n(TargetVar \cdot Cond)}{n(Cond)} - \frac{n(TargetVar)}{n_s} \right) \quad (2)$$

The unusualness of a rule can be described as the balance between the coverage of the rule $p(Cond_i)$ and its accuracy gain $p(TargetVar \cdot Cond) - p(TargetVar)$.

- *Sensitivity*: This measure is the proportion of actual matches that have been classified correctly [4]. It can be computed as:

$$Sens(R) = \frac{TP}{Pos} = \frac{n(TargetVar \cdot Cond)}{n(TargetVar)} \quad (3)$$

This quality measure was used in [14] as *Support based on the examples of the class*. Sensitivity combines precision and generality related to the target variable.

- *Fuzzy Confidence*: It measures the relative frequency of examples satisfying the complete rule among those satisfying only the antecedent in fuzzy rules [14]. This quality measure is specific for subgroup discovery algorithms based on fuzzy logic, but in the literature can be found crisp confidence, too. This can be computed as:

$$FCnf(R) = \frac{\sum_{E^k \in E / E^k \in TargetVar} APC(E^k, R)}{\sum_{E^k \in E} APC(E^k, R)}, \quad (4)$$

The following assumptions are important to understand this fuzzy quality measure:

  - An example E^k verifies the APC of a rule R_i if

$$APC(E^k, R_i) = T(\mu_{LL_1^1}(e_1^k), \ldots, \\ \mu_{LL_{n_v}^{l_{n_v}}}(e_{n_v}^k)) > 0 \quad (5)$$

  where APC (Antecedent Part Compatibility) is the degree of compatibility between an example and the antecedent part of a fuzzy rule, i.e., the degree of

membership for the example to the fuzzy subspace delimited by the antecedent part of the rule,
where:

* $\mu_{LL_{n_v}^{l_{n_v}}}(e_{n_v}^k)$ is the degree of membership for the value of the feature n_v for the example E^k to the fuzzy set corresponding to the linguistic label l_{n_v} for this variable (n_v);
* T is the $t-norm$ selected to represent the meaning of the AND operator (the fuzzy intersection) in our case the minimum $t-norm$.

– An example E^k is covered by a rule R_i if

$$APC(E^k, R_i) > 0 \quad AND \quad E^k \in TargetVar_j. \tag{6}$$

This means that an example is covered by a rule if the example has a degree of membership higher than 0 to the fuzzy input subspace delimited by the antecedent part of the fuzzy rule, and the value indicated in the consequent part of the rule agrees with the value of the target feature for the example. For the categorical variables, the degrees of membership are 0 or 1.

III. Missing data imputation

As mentioned above, the presence of MVS affects the results of data mining algorithms in general, and those of the subgroup discovery algorithms in particular.

First of all, it is important to categorise the mechanisms which lead to the introduction of MVs [3]. The assumptions we make about the mechanism producing MVs and the missing data pattern of MVs can affect which imputation method could be applied, if any. As [3] stated, there are three different mechanisms for missing data induction:

1) Missing completely at random (**MCAR**), when the distribution of an example having a MV for an attribute does not depend on either the observed data or the missing data.
2) Missing at random (**MAR**), when the distribution of an example having a MV for an attribute depends on the observed data, but does not depend on the missing data.
3) Not missing at random (**NMAR**), when the distribution of an example having a MV for an attribute depends on the MVs.

In the case of the MCAR mode, the assumption is that the underlying distributions of missing and complete data are the same, while for the MAR mode they are different, and the missing data can be predicted by using the complete data [3]. These two mechanisms are assumed by the imputation methods so far. Another approach is to convert the missing values to a new value (encode them into a new numerical value), but such a simplistic method was shown to lead to serious inference problems [24].

In this paper, single imputation methods are used due to the time complexity of the multiple imputation schemes, and the assumptions they make regarding data distribution and MV randomness; that is, that we should know the underlying distributions of the complete data and missing data prior to their application. Only in the cases of MCAR or MAR the imputation can be carried out [3], which are assumed to the the underlying ones in the real-world data sets [25].

We consider the following imputation methods as they are the most well-known approaches [26]. We also compare them with the lack of imputation:

- Case deletion or Ignore Missing (**IM**). Using this method, all instances with at least one MV are discarded from the data set.
- Concept Most Common Attribute Value for Symbolic Attributes, and Concept Average Value for Numerical Attributes (**CMC**) [27]. MV is replaced by the most repeated value (for nominal ones) or by the mean value (for numerical ones), but considering only the instances with the same target value as the reference instance.
- Imputation with K-Nearest Neighbour (**KNNI**) [26]. Using this instance-based algorithm, every time an MV is found in a current instance, KNNI computes the k nearest neighbours and a value from them is imputed. For nominal values, the most common value among all neighbors is taken, and for numerical values the average value is used. Therefore, a proximity measure between instances is needed for it to be defined. The Euclidean distance (it is a case of a L_p norm distance) is the most commonly used in the literature. Specifically, in this paper we use $k = 10$.

IV. Experimental study

This section presents an experimental study with the different evolutionary fuzzy systems for subgroup discovery presented throughout the literature and briefly described in Section II-A. In the study, several data sets with MVs are used. In order to analyse the quality of different imputation methods for MVs, the previously presented pre-processing approaches are applied in these data sets. So, the subgroup discovery algorithms will be applied to the data sets resulting from the imputation methods used.

The experimentation was undertaken with data sets from KEEL[1] repository [28], [29] containing MVs.

The properties of data sets are presented in Table I including: number of variables (n_v), number of discrete variables (n_{vD}), number of continuous variables (n_{vC}), number of classes (n_c), number of examples (n_s) and percentage of examples with MVs $(MV(\%))$ which indicates percentage of instances with at least one MV.

In order to perform the experimental study, subgroup discovery algorithms are executed using the ten-fold cross-validation (10-fcv) procedure. Therefore, the results shown for

[1]http://www.keel.es

TABLE I
PROPERTIES OF THE DATA SETS USED FROM THE KEEL REPOSITORY

$Name$	n_v	n_{vD}	n_{vC}	n_c	n_s	$MV(\%)$
Adult	14	8	6	2	48842	7.41
Bands	19	6	13	2	539	32.28
Breast	9	9	0	2	286	3.15
Cleveland	13	0	13	5	303	1.98
Crx	15	12	3	2	690	5.36
Dermatology	34	34	0	6	366	2.19
Hepatitis	19	13	6	2	155	48.39
Horse-colic	23	16	7	2	368	98.10
Housevotes	16	16	0	2	435	46.67
Mammographic	5	5	0	2	961	13.63
Mushroom	22	22	0	2	8124	30.53
Wisconsin	9	9	0	2	699	2.29

the experiments are the average of the results obtained for each data set for the different partitions. The results showed for each subgroup discovery algorithm are the average of 50 executions (5 executions per cross validation group).

The parameters used by the subgroup discovery algorithms are presented in Table II.

The average results obtained by the evolutionary fuzzy systems for subgroup discovery task are presented in Table III. The values of the quality measures presented correspond to the average for the rule sets in all the data sets studied. In this table, the subgroup discovery $Algorithm$, the preprocessing MVs method ($MVmethod$) employed, and the average results of the quality measures of fuzzy confidence ($FCNF$), sensitivity ($SENS$), significance ($SIGN$) and unusualness ($UNUS$) are shown. Complete results obtained by any algorithm for each data set are available in the Website http://simidat-web.ujaen.es/MVs/FUZZIEEE12.

Before analysing the results obtained by the algorithms, the main guidelines to be satisfied by any subgroup discovery algorithm are presented. Due to the analysis performed in [9], the use of different quality measures in the algorithms presented throughout the literature and the formal definition of subgroup discovery task, it could be considered that a subgroup discovery approach must satisfy different guidelines in order to measure its quality:

1) Interpretability. This guideline measures the number of rules and variables obtained by a subgroup discovery model. Optimal interpretability for a subgroup discovery approach is the obtaining of few rules that containing a low number of variables in order to help to the experts to understand and use the extracted knowledge because algorithms search for simple and interpretable subgroups through partial relations.
2) Relation between sensitivity and confidence. This objective quantifies a good compromise between both measures, i.e. the algorithm must achieve the best possible relation between sensibility and confidence. Both quality measures are essential to provide the experts subgroups which describe correctly as much examples as possible. This is very difficult to be achieved by the algorithms, as normally the improvement of one of the measures brings the worsening of the other.
3) Interest or Novelty. This final guideline is related with the search of interesting and unusual relations in data. A subgroup discovery model must contribute novel knowledge to the problem. This objective could be measured with a wide number of quality measures as novelty, interest or significance, among others. Nevertheless, it is important to highlight the use of the unusualness to measure this objective because it contributes with generality and confidence to the problem.

In order to measure each guideline, the following quality measures are considered: (i) number of rules and number of variables for interpretability, (ii) sensitivity and confidence for the relation between both measures and (iii) unusualness and significance with respect to the interest.

The interpretability obtained by the algorithms with the different missing data imputations is similar. Thus the analysis is focused on the remaining two guidelines: “relation sensitivity-confidence” and “interest”. In order to present a comprehensible analysis of this experimental study, a summary of the results obtained for each algorithm is presented:

- For the SDIGA algorithm the best relation between sensitivity and confidence is obtained by the missing data imputation of *KNNI* as it obtains the best values in both quality measures. With respect to the interest quality measure, although *IM* obtains the best value in unusualness, its value in significance is very poor. In this way, the best results in interest for this algorithm are obtained with *CMC* and *KNNI* method.
- MESDIF obtains the best results in both interest and relation sensitivity-confidence with the *KNNI* imputation method.
- For NMEEF-SD algorithm the best relation sensitivity-confidence is obtained by *KNNI* method because this one obtains the best results in both quality measures. In significance quality measures can be observed clearly that the best results are obtained by the same method, and for unusualness although *KNNI* does not obtain the best results is very similar.

From this analysis can be observed that the best results are obtained with the missing data imputation of *KNNI*. This algorithm allows all the subgroup discovery algorithms analysed to obtain the best relation between sensitivity and confidence. In addition to this, this algorithm gets excellent values in unusualness and significance. NMEEF-SD is the best approach, showing the best synergy with the imputation step, and specially with the *KNNI* algorithm.

V. CONCLUSIONS

A study about the influence of MVs in evolutionary fuzzy systems for subgroup discovery techniques is carried out in this paper, where different missing data imputations are

TABLE II
PARAMETERS OF ALGORITHMS EMPLOYED

Algorithm	Parameters
SDIGA	Population size=100, evalutions=10000, crossover probability=0.60, mutation probability=0.01, minimum confidence=0.6, representation of the rule=canonical and linguistic labels=3, objective1=sensitivity (weight=0.4), objective2=fuzzy confidence (weight=0.3) and objective3=unusualness (weight=0.3).
MESDIF	Population size=100, evalutions=10000, crossover probability=0.60, mutation probability=0.01, elite population size=3, representation of the rule=canonical, linguistic labels=3, objective1=sensitivity, objective2=fuzzy confidence and objective3=unusualness.
NMEEF-SD	Population size=50, evalutions=10000, crossover probability=0.60, mutation probability=0.1, minimum confidence=0.6, representation of the rule=canonical, linguistic labels=3, objective1=sensitivity and objective2=unusualness.

TABLE III
AVERAGE RESULTS OBTAINED FOR SUBGROUP DISCOVERY ALGORITHMS WITH DIFFERENT MVS APPROACHES

$Algorithm$	MV_{method}	$SIGN$	$UNUS$	$SENS$	$FCNF$
SDIGA	CMC	**21.718**	0.046	0.701	0.525
	IM	12.931	**0.050**	0.631	0.510
	KNNI	21.711	0.048	**0.741**	**0.549**
MESDIF	CMC	11.623	0.026	0.290	**0.477**
	IM	11.466	0.025	0.258	0.399
	KNNI	**13.257**	**0.034**	**0.330**	0.475
NMEEF-SD	CMC	17.226	**0.111**	0.906	0.783
	IM	17.688	**0.111**	0.849	0.779
	KNNI	**18.046**	0.110	**0.908**	**0.788**

analysed in order to obtain the best method for this type of algorithms. The paper is supported by a complete experimental study with data sets from KEEL containing MVs.

The analysis performed on the experimental study indicates that the best results are obtained using the pre-processing method based on *KNNI* algorithm. This pre-processing method obtains the best relation sensitivity-confidence in all evolutionary fuzzy systems analysed and good novelty results.

The best results are obtained by NMEEF-SD algorithm which obtains the best values in the majority of the quality measures: unusualness, sensitivity and fuzzy confidence. It is interesting to remark that NMEEF-SD obtains a sensitivity upper than 90% with a confidence close to 80% with pre-processing method of *KNNI*. As a consequence, NMEEF-SD is the best alternative to apply when MVs are found in the data.

In summary, this paper presents an alternative to improve the results obtained by any evolutionary fuzzy system for subgroup discovery in environments with MVs in the attributes. It arises an interesting line of work, where future efforts must involve the analysis of more complex and related imputation algorithms over controlled amounts of MVs. Knowledge from other related supervised learning fields can be useful in order to guide these efforts.

ACKNOWLEDGMENTS

This paper was supported by the Spanish Ministry of Education, Social Policy and Sports under project TIN-2008-06681-C06-02, FEDER Founds, and by the Andalusian Research Plan under project TIC-3928, FEDER Founds.

REFERENCES

[1] J. Barnard and X. Meng, "Applications of multiple imputation in medical studies: From AIDS to NHANES," *Statistical Methods in Medical Research*, vol. 8, no. 1, pp. 17–36, 1999.

[2] D. Pyle, *Data Preparation for Data Mining*. Morgan Kaufmann, 1999.

[3] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 1st ed., ser. Wiley Series in Probability and Statistics. New York: Wiley, 1987.

[4] W. Kloesgen, "Explora: A Multipattern and Multistrategy Discovery Assistant," in *Advances in Knowledge Discovery and Data Mining*. American Association for Artificial Intelligence, 1996, pp. 249–271.

[5] S. Wrobel, "An Algorithm for Multi-relational Discovery of Subgroups," in *Proceedings of the 1st European Symposium on Principles of Data Mining and Knowledge Discovery*, ser. LNAI, vol. 1263. Springer, 1997, pp. 78–87.

[6] O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*. World Scientific, 2001.

[7] A. Siebes, "Data Surveying: Foundations of an Inductive Query Language," in *Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining*. AAAI Press, 1995, pp. 269–274.

[8] S. Wrobel, *Inductive logic programming for knowledge discovery in databases*. Springer, 2001, ch. Relational Data Mining, pp. 74–101.

[9] F. Herrera, C. J. Carmona, P. González, and M. J. del Jesus, "An overview on Subgroup Discovery: Foundations and Applications," *Knowledge and Information Systems*, vol. 29, no. 3, pp. 495–525, 2011.

[10] T. Bäck, D. Fogel, and Z. Michalewicz, *Handbook of evolutionary computation*. Oxford University Press, 1997.

[11] L. A. Zadeh, "The concept of a linguistic variable and its applications to approximate reasoning. Parts I, II, III," *Information Science*, vol. 8-9, pp. 199–249,301–357,43–80, 1975.

[12] F. Herrera, "Genetic fuzzy systems: taxomony, current research trends and prospects," *Evolutionary Intelligence*, vol. 1, pp. 27–46, 2008.

[13] A. E. Eiben and J. E. Smith, *Introduction to evolutionary computation*. Springer, 2003.

[14] M. J. del Jesus, P. González, F. Herrera, and M. Mesonero, "Evolutionary Fuzzy Rule Induction Process for Subgroup Discovery: A case study in marketing," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 4, pp. 578–592, 2007.

[15] C. Romero, P. González, S. Ventura, M. J. del Jesus, and F. Herrera, "Evolutionary algorithm for subgroup discovery in e-learning: A prac-

tical application using Moodle data," *Expert Systems with Applications*, vol. 36, pp. 1632–1644, 2009.
[16] C. J. Carmona, P. González, M. J. del Jesus, C. Romero, and S. Ventura, "Evolutionary algorithms for subgroup discovery applied to e-learning data," in *Proceedings of the IEEE International Education Engineering*, 2010, pp. 983–990.
[17] C. J. Carmona, P. González, M. J. del Jesus, M. Navío, and L. Jiménez, "Evolutionary Fuzzy Rule Extraction for Subgroup Discovery in a Psychiatric Emergency Department," *Soft Computing*, vol. 15, no. 12, pp. 2435–2448, 2011.
[18] C. J. Carmona, P. González, M. J. del Jesus, and S. Ventura, "Subgroup discovery in an e-learning usage study based on Moodle," in *Proceedings of the International Conference of European Transnational Education*, 2011, pp. 446–451.
[19] M. J. del Jesus, P. González, and F. Herrera, "Multiobjective Genetic Algorithm for Extracting Subgroup Discovery Fuzzy Rules," in *Proceedings of the IEEE Symposium on Computational Intelligence in Multicriteria Decision Making*. IEEE Press, 2007, pp. 50–57.
[20] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization," in *International Congress on Evolutionary Methods for Design Optimization and Control with Applications to Industrial Problems*, 2002, pp. 95–100.
[21] C. J. Carmona, P. González, M. J. del Jesus, and F. Herrera, "NMEEF-SD: Non-dominated Multi-objective Evolutionary algorithm for Extracting Fuzzy rules in Subgroup Discovery," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 5, pp. 958–970, 2010.
[22] K. Deb, A. Pratap, S. Agrawal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
[23] N. Lavrac, P. A. Flach, and B. Zupan, "Rule Evaluation Measures: A Unifying View," in *Proceedings of the 9th International Workshop on Inductive Logic Programming*, ser. LNCS, vol. 1634. Springer, 1999, pp. 174–185.
[24] J. L. Schafer, *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall, 1997.
[25] 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.
[26] G. Batista and M. Monard, "An analysis of four missing data treatment methods for supervised learning," *Applied Artificial Intelligence*, vol. 17, no. 5, pp. 519–533, 2003.
[27] J. Grzymala-Busse, L. Goodwin, W. Grzymala-Busse, and X. Zheng, "Handling missing attribute values in preterm birth data sets," in *Proceedings of 10th international conference of rough sets and fuzzy sets and data mining and granular computing*, 2005, pp. 342–351.
[28] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, "KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework," *Journal of Multiple-Valued Logic and Soft Computing*, vol. 17, no. 2-3, pp. 255–287, 2011.
[29] J. Alcalá-Fdez, L. Sánchez, S. García, M. del Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fernández, and F. Herrera, "KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems," *Soft Computing*, vol. 13, no. 3, pp. 307–318, 2009.