An Evolutionary Missing Data Imputation Method for Pattern Classification

Fabio M. F. Lobato
Technological Institute
Federal University of Pará
Belém, PA, Brazil
lobato.fabio@ufpa.br

Vincent W. Tadaiesky
Technological Institute
Federal University of Pará
Belém, PA, Brazil
vincent@ufpa.br

Igor M. de Araújo
Technological Institute
Federal University of Pará
Belém, PA, Brazil
igoraraujo@ufpa.br

Adamo L. Santana
Technological Institute
Federal University of Pará
Belém, PA, Brazil
adamo@ufpa.br

ABSTRACT

Data analysis plays an important role in our Information Era; however, most of statistical and machine learning algorithms were not developed to tackle the ubiquitous issue of missing values. In pattern classification, several strategies have been proposed to handle this problem, where missing data imputation is the most used one, which can be viewed as an optimization problem where the goal is to reduce the bias imposed by the absence of information. Although most imputation methods are restricted to one type of variable only (categorical or numerical), they usually ignore information within incomplete instances. To fill these gaps, we propose an evolutionary missing data imputation method for pattern classification, based on a genetic algorithm, which is suitable for mixed-attribute datasets and takes into account information from incomplete instances and model building – more specifically, the classification accuracy. To assess the performance of our method, we used three algorithms in order to represent the three groups of classification methods: 1) rule induction learning, 2) approximate models and 3) lazy learning. Experiments have shown that the proposed method outperforms some well-established missing value treatment methods.

Categories and Subject Descriptors

H.2.8 [**Database management**]: Database Applications—*Data mining*; I.5.2 [**Pattern Recognition**]: Design methodology—*Pattern analysis*

GECCO '15, July 11 - 15, 2015, Madrid, Spain

ACM 978-1-4503-3488-4/15/07$15.00
DOI: http://dx.doi.org/10.1145/2739482.2768451

General Terms

Incomplete data; Data mining; Supervised learning by classification

Keywords

Missing data; data imputation; evolutionary computing; genetic algorithms

1. INTRODUCTION

The important role of data analysis is unquestionable in our Information Era. Even so, most statistical and machine learning methods are not robust enough to be unaffected by a ubiquitous problem in data analysis: Missing Data (MD). Missing data affects both hard and soft sciences [13] and is a recurring problem in pattern classification. By way of illustration, about 45% of datasets in the UCI repository have missing values [11, 17].

In order to mitigate the harmful consequences of Missing Data, several studies have been conducted that aim to tackle this issue. The most accepted way to handle this problem is by means of Missing Data Imputation (MDI), which denotes the estimation of plausible values in order to substitute the missing ones [28, 22]. Based on the definition, MDI can be viewed as optimization problems, where the goal is to find the best values to impute which will reduce the bias imposed by the absence of information. In this context, metaheuristics – such as Evolutionary Algorithms (EA) – have been successfully applied to solve optimization problems.

Despite the paradigm in which the imputation method is based, some restrictions that are present in these state-of-the-art techniques should be pointed out. For instance, [33] makes known that most of the available imputation methods are restricted to one type of variable only (categorical or numerical). In other words, these methods handle variables of different types separately, losing possible relationships between them. It is critical to remember that this kind of correlation is usually explored by classification algorithms, thus it is important to treat MD in mixed-attributes datasets properly. Two other important restrictions are: 1) imputation methods evaluation cannot be properly evaluated apart from the modelling task [29] and 2) we should avoid complete-case analysis, where information of instances or attributes with missing values are removed.

Aiming to fill these gaps, we present in this paper an evolutionary missing data imputation method for pattern classification called GAI, which is based on a genetic algorithm. GAI aims to treat mixed-attributes datasets properly, considering incomplete instances and information of the generated model, more specifically, the classification accuracy obtained from the three algorithms which represent the three groups of classification methods: rule induction learning, approximate models and lazy learning.

We compare the performance of GAI with some well- accepted methods for handling missing values in pattern classification, using benchmarking datasets obtained from the UCI repository [17] that already have missing values. Our results show that GAI reaches a very high performance, moreover, our method proved to be suitable for tackling missing values in mixed-attribute datasets.

The rest of the paper is organized as follows. In Section 2 we present a brief theoretical background about missing values, followed by an overview of related work on Missing Data treatment in Section 3. We then describe our evolutionary imputation method in Section 4. Section 5 discusses our experimental methodology and results. Finally, conclusions and some open issues are presented in Section 6.

2. THEORETICAL BACKGROUND

As said previously, missing values can be defined as the absence of information in instances, which brings harmful consequences to the validity of the subsequent analyzes. Table 1 shows an example of dataset with missing values. The instances 1–5 are commonly called "complete cases" because they have no missing values, while instances 6–8 are called "incomplete cases" because they have missing values, usually represented by "?".

Table 1: Example of a dataset with missing values.

ID	Color	Weight	Broken	Class
1	Black	80	Yes	1
2	Yellow	100	No	2
3	Yellow	120	Yes	2
4	Blue	90	No	2
5	Blue	85	No	2
6	**?**	60	No	1
7	Yellow	100	**?**	2
8	**?**	40	**?**	1

The causes of this problem are diverse and are related to the application domain. For example, a common cause of MD occurrence in behavioral sciences is the refusal of respondents to answer certain questions; while the inability to perform a specific exam most affects health research; and, finally, equipment failure is frequent in areas dependent on sensor networks, such as: traffic monitoring, industrial processes and satellite information processing [4]. In the classification context, by way of illustration, approximately 45% of datasets in the UCI repository have missing values [11, 17].

By means of these examples is possible to attest to the ubiquity of this problem, which makes clear the importance of developing treatment methods to handle it. The choice of a suitable way to tackle this issue depends on its causes and can be regarded as a probabilistic phenomenon [26, 22]. A proposed mathematical device called "missingness mechanisms" aims to describe the missingness characteristics and to identify possible relationships between the observed and missing items [19]. It defines three different patterns [28]:

- Missing completely at random (MCAR): missing data that does not depend on the observed or missing data itself;
- Missing at random (MAR): inferences can be made from the observed data, but are independent of the missing variables;
- Missing not at random (MNAR): the absence of data is not stochastic because it depends on the missing data itself.

It is important to note that "missingness mechanism" denotes a statistical association between the observed and missing data, but not a causal relationship. In this sense, MCAR and MAR are considered to be non-informative patterns because the unavailability of data does not deliver important information about missingness. Due to this fact, these mechanisms are termed as "ignorable". The importance of such observations is that the data analyst can ignore the reasons for the missing data, making the subsequent analysis less laborious. On the other hand, MNAR are considered to be informative. For example, in a classification dataset, for a label k, the attribute j is always missing, thus it gives important clues to identify instances that belong to label k.

Due to the simplicity and frequency of occurrences, the majority of research covers cases (or assumes that) where missing data belongs to non-informative patterns (MAR and MCAR). Until the 1970s, missing values were handled primarily by manual edition and complete-case analysis [28], nowadays we have a large number of treatment methods [13] for MD. According to [11], pattern classification with missing data shows two different problems: 1) handle MD and 2) perform the classification itself. In views of this it, the authors categorized methods for pattern classification with missing data into four main groups:

- Case deletion;
- Missing data imputation;
- Model-based procedures;
- Machine learning methods for handling missing data.

The first method, case deletion, consists in removing examples or attributes with missing values, this is also known as complete case analysis [27, 20]. In real cases this method should be avoided because useful information can be lost and consequently the method increases the data acquisition costs in order to deliver more complete-cases. However, as stated by [23], it is the most commonly used approach for dealing with missing data. In the pattern classification context, this approach has an even greater impact when associated with unbalanced datasets, making the use of the most sophisticated treatment methods necessary

In this sense, methods based on the second method, missing data imputation, should be highlighted. MDI means to estimate plausible values in order to substitute the missing ones. There are several ways to estimate the value to be imputed, from naïve approaches (such as: mean and mode substitution), to machine learning and statistical based methods (such as multiple imputation [20], Bayesian imputation [15], k-nearest neighbors imputation [3], autoenconders neural networks imputation [24]).

The third method, proposed by [11], is the Model-based approach, where the data analysts have to make assumptions about the joint distributions of all studied variables. One of the most accepted model-based methods is the mixture models trained with expectation-maximization algorithm [28, 5]. The fourth method is the machine learning method for handling missing data, which aims to develop machine learning techniques that are more robust to missing data incidence. The most prominent examples of this category are ensemble classifiers [25], fuzzy procedures [10] and hybrid approaches [18].

3. RELATED WORK

This section briefly reviews some recent work related to our proposed method and it covers three main work groups: reviews and comparison studies, research on data imputation's impact on different data types and, finally, evolutionary approaches to dealing with missing values.

3.1 Reviews and comparisons studies

A recent literature review about pattern classification with missing data is given by [11], where the missing value problem and its impact are discussed. The authors also provide an outline about well-known methods used to tackle this problem, from naïve approaches (*e.g.* mean and hot-deck imputation) to more robust classification methods. As a conclusion, [11] gives some important research directions:

- Data imputation is widely used because data analysis softwares cannot tackle MD;
- In pattern classification with missing values, the main goal of a treatment method is to improve the reliability of classification, in other words, to enhance the classification accuracy;
- There is no unique-best solution that provides optimum results for each classification domain;
- The choice of treatment method for missing data is a complex task.

In this sense, many studies have focused on the comparison of existing imputation methods in the classification context. [31] provides a review and comparison of the possible strategies for handling missing data in a separate-and-conquer rule learning. The imputation methods compared are based on the most common value imputation, except the "Predicted value strategy" which is based on k-nn imputation. In this study, the authors conclude that the strategies analyzed has its particular strengths and weaknesses, making them perform well on some datasets and poorly on others, making hard to detect clear-cut differences.

A more detailed study about nearest neighbor-based imputation is given by [29], the authors studied the influences of five nearest neighbor based imputation algorithms, taking as baseline mean and majority imputation methods. The behavior of these imputation methods was evaluated using synthetic datasets, the authors concluded that the best prediction results, regarding the distance between the original and imputed values, do not necessarily yield smaller classification biases.

It is important to point out that the traditional approaches used to evaluate imputation methods, which take into account the distance between the original and imputed values, are not suitable due to low correlations between the distance and the classification bias. This suggests that the best predictive accuracy results do not necessarily lead to the lowest classification bias [14, 11, 29]. Thus, it is clear that the imputation methods cannot be properly evaluated apart from the modelling task, in our case, the pattern classification.

Another important comparative study is presented by [22], which consists of a meta-learning study about the imputation method choice. It considers three groups of classification methods: 1) rule induction learning, 2) approximate models and 3) lazy learning. The contributions of this study are: the testbed used to perform the comparisons (which is comprised of 14 imputation methods, 21 benchmarking datasets and 23 classification methods) and the impact evaluation of each imputation method in relation to classification group.

It is important to point out that some of the imputation methods studied by [22] are specifically designed for discrete or continuous data. In view of this fact, some studies are dedicated to assessing the impact of the data types in the treatment of missing values.

3.2 Data types

Missing data imputation offers a good solution to many application domains, however most of available imputation methods are restricted to one type of variable only: continuous or categorical [30]. So, for mixed-attribute datasets, these sort of methods handle different data types separately. As a result, this strategy ignores the possible relations between the variable types. This has a negative impact in the classification context, because such relationships are usually explored by machine learning methods.

As seen in Subsection 3.1, nearest neighbor imputation approaches are well-accepted; but, as pointed by [32], they are usually based on Minkowsky distance or its variants, which are generally efficient for numerical variables and do not perform well for categorical ones. For this reason, the authors propose a new kNN imputation method, based on gray distance, which works better with mixed-datasets. Other studies propose different strategies under the mixed-attribute perspective, such as [33, 30]. Through the result analysis of these studies, it is possible to attest that mixed-attributes approaches were proven to be more robust, since they take into account the relationships between variables.

3.3 Evolutionary approaches

As seen previously, several data imputation methods have been proposed, some of them utilize evolutionary approaches. At this point, these studies can be divided into two groups: 1) methods that apply evolutionary algorithms to improve the convergence of other imputation methods [24, 1] and 2) the ones that use EA to perform the imputation itself. The latter will be discussed in this subsection.

In this perspective, in [6, 7] the authors describe a missing data estimation method in time series data based on an evolutionary algorithm, more specifically, using genetic algorithms. As heuristics to guide the evolution process, the authors use an autocorrelation function, mean and variance, because these statistics are useful to build some well-known linear time series models.

Despite the innovation regarding the use of genetic algorithm to perform data imputation itself, this approach shows some weaknesses considering scalability and accuracy, since the individuals are codified into a m size vector, instead of using subsets/partitions [2, 12]. A modified version of these approaches was proposed for multivariate data [8] and it is important to highlight that this method falls into complete-case analysis, since the statistic information used in fitness function are extracted from examples without missing values, thus, important information is lost. In addition, as stated by [21], the covariance criterion is not related to the classification criterion.

4. EVOLUTIONARY DATA IMPUTATION

In mixed-attribute datasets with missing data, the imputation process can be viewed as a Mixed-Variable Optimization Problem (MVOP), which consists of a model $R = (\mathbf{S}, \Omega, f)$, where $\mathbf{S}$ is the search space defined over a finite set of both discrete and continuous decision variables; Ω is a set of constraints among the variables; and $f : \mathbf{S} \to \mathbb{R}_0^+$ to be minimized [16].

Many techniques have been proposed to solve optimization problems, in which metaheuristics are well-accepted. An example of a metaheuristic that is widely used is the genetic algorithms, which belongs to the EA category and draws attention due to its exploitation and exploration abilities, easy adaptation to specific problems and computation efficiency. For these reasons, this evolutionary method has been applied successfully in many areas, including data mining and machine learning [9]. In this section, we present a genetic algorithm for missing data imputation, which considers information within incomplete instances, takes into account the model construction performance and, furthermore, is suitable for mixed-attribute datasets.

4.1 Workflow

The proposed genetic algorithm workflow is shown in Figure 1. Firstly, the dataset is divided into $k * C$ subdatasets, where k is a user-defined constant and C is the number of classes. Those subsets with missing values will undergo the evolutionary imputation process, which follows the basic steps of a genetic algorithm: *i)* population of individuals initialization; *ii)* fitness evaluation of each individual; *iii)* evaluation of stop criterion; and if the stop criterion is not satisfied, then *iv)* apply selection, crossover and mutation operators in order to produce a new generation of individuals; back to Step *ii*, if it is satisfied, then stop. The Steps *ii* to *iv* are repeated for many generations until the stop criterion is satisfied.

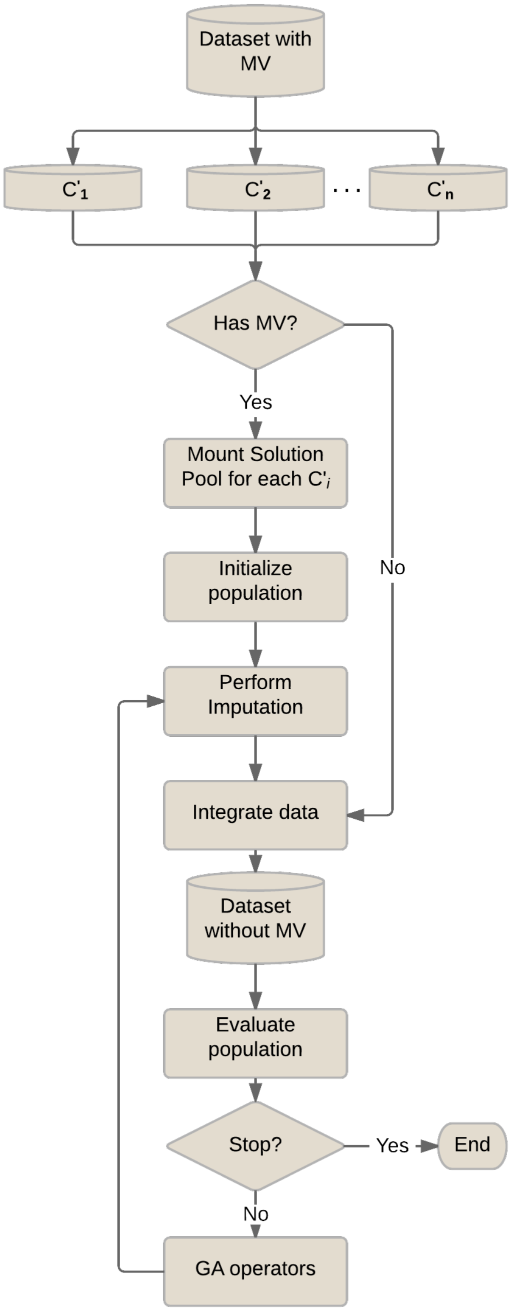


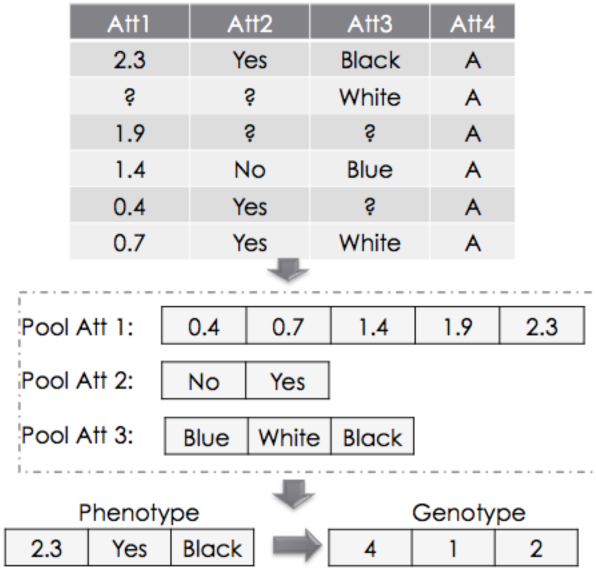
Figure 1: Workflow of the proposed genetic algorithm.

4.2 Individual encoding and genetic operators

Each subset with missing values shown in Figure 1 represents a gene that will compose the chromosome. A gene consists of alleles which contain the value to substitute the missing ones for each attribute (Figure 2 illustrates this description).

Two important concepts of our approach are summarized in Figure 2, the solution pool and the gene composition. The first one, the solution pool, consists of ordered sets of all possible values for each attribute with missing observations at each subdataset; it is based on these sets that the gene is formed. The chromosome size is equal to the number of attributes with missing values of that subset, since each attribute is represented by one allele.

The index of the solution pool is used as the genotype and the phenotype is the value referred by the index itself. So, using the illustrative example of Figure 2, after imputation process the Att3 will be = *Black, White, Black, Blue, Black, White*, the rest follows the same logic. In brief, the chromosome is the assembly of all genes, combining the solutions of



Att1	Att2	Att3	Att4
2.3	Yes	Black	A
?	?	White	A
1.9	?	?	A
1.4	No	Blue	A
0.4	Yes	?	A
0.7	Yes	White	A

Pool Att 1:	0.4	0.7	1.4	1.9	2.3
Pool Att 2:	No	Yes			
Pool Att 3:	Blue	White	Black		

Figure 2: Gene codification scheme.

each subdataset and, consequently, resulting in a complete solution.

The crossover operation is a variant of n-point crossover, where n is the number of genes, so the genetic information is exchanged by genes, generating new individuals. For numerical attributes, the mutation is based on Gaussian distribution, where the mean is the actual index (genotype), so a random (and feasible) number is generated accordingly to Gaussian distribution, thus, as the values are sorted in the solution pool, the mutation does not often cause abrupt changes. For categorical attributes, the mutation is a random change in value, which is chosen from the solution pool.

4.3 Fitness function and selection method

As seen in Subsection 4.2, the proposed model was designed to consider information within incomplete instances and treat mixed-attributes properly. The last feature to cover is to take into account the model construction feedback, which is done by the fitness function. Therefore, to measure the quality of a candidate imputation, the fitness function is calculated according to Eq. 1, which represents the average of classification accuracies.

$$Fitness = \frac{1}{n}\sum_{i=1}^{n} acc_i \qquad (1)$$

In Eq. 1, n is the number of classifiers used and the acc_i is the accuracy of the i-th classifier. The selection of individuals to cross-over is done by tournament and the elitism operator is also applied to generate a new population.

5. DISCUSSION

This section presents the experimental setup adopted to assess the performance of the proposed method, followed by the results and further discussion.

5.1 Experimental methodology

The tests were conducted using three classification methods: C4.5, Naïve-Bayes and kNN. These methods were selected in order to represent three categories of classifiers: rule induction learning, approximate models and lazy learning, respectively.

The datasets used in the experiments were obtained from the UCI repository [17]and were selected because they already have missing values; additionally, the data represent numeric, nominal and mixed-attributes. Table 2 2 summarizes some information about the datasets, where %IMD denotes the ratio of instances with missing values.

Table 2: Datasets description.

Dataset	Acron.	%IMD	Att. Type
audiology	AUD	98.23%	Nominal
autos	AUT	22.44%	Mixed
cleveland	CLE	1.98%	Numeric
hepatitis	HEP	48.39%	Mixed
lung-cancer	LUN	15.15%	Nominal
mammographic	MAM	13.63%	Numeric

The performance of the GAI method was compared to four well-known imputation methods: 1) Concept most common attribute value for symbolic attribute and concept average value for numeric attribute (CMC); 2) k-nearest neighbor imputation (kNNI); 3) ignore missing (IM); and 4) event covering (EC). As an evaluation criterion of the imputation methods, the classifier accuracies were adopted. To assess the statistical significance of obtained comparisons between the imputation methods, the Wilcoxon signed-rank test with a confidence level of 95% was employed.

The genetic algorithm parameters are shown in Table 3. Due to the similarity in dataset complexity, the same parameters were adopted for all datasets, with the number of generations as a stop criterion.

Table 3: Genetic algorithm parameters.

Parameter	Value
Population Size	50
Mutation rate	10%
Cross-over rate	90%
Elitist individuals	3
Individuals per tournament	4
Number of generations	30

5.2 Results

Table 4 presents the performance of each imputation method with respect to classifiers' accuracy. The best results for the combination between classification algorithm and dataset are highlighted in bold type.

According to Table 4, the proposed method outperforms the baseline methods in most scenarios, except Naïve-Bayes when GAI loses 50% of datasets; for the other two classification algorithms GAI wins in 4 out of 5 datasets. For this reason, it is possible to state that GAI overcomes the others' imputation methods, as evidenced in the boxplot of overall accuracies shown in Figure 3.

The results presented in Table 4 are summarized in Figure 3. The results show the robustness of GAI, since it has the highest average and smaller variance in relation to the

Table 4: Performance of each imputation method in terms of accuracy.

	Datasets	Missing Values Treatment Methods				
		IM	EC	KNNI	CMC	GAI
C4.5	AUD	0	78.42	78.42	72.47	**79.83**
	CLE	53.52	53.75	53.46	53.14	**55.38**
	LUN	**83.33**	43.33	43.33	43.33	50
	MAM	82.32	82.84	81.17	83.04	**84.76**
	AUT	80.93	73.52	82.19	78.91	**84.29**
	HEP	82.98	83.88	76.79	86.25	**91.42**
Naïve-Bayes	AUD	20	72.15	72.61	72.63	**74.60**
	CLE	54.91	55.46	55.46	55.46	**55.71**
	LUN	**66.67**	52.5	49.17	52.5	53.12
	MAM	82.9	82.52	81.69	**83.56**	83.07
	AUT	**72.24**	68.86	69.35	69.87	58.68
	HEP	82.65	85.08	81.25	86.46	**87.61**
3-NN	AUD	0	68.58	73.04	69.94	**74.69**
	CLE	54.8	54.12	55.05	55.05	**57.09**
	LUN	31.67	31.67	31.67	31.67	**50**
	MAM	80.25	64.21	80.44	**81.8**	81.63
	AUT	65.34	1.46	65.15	64.28	**72.97**
	HEP	82.32	79.37	81.42	83.96	**87.80**

accuracy of the classifiers. In addition, there is no correlation between the losses and attribute type, thus the GAI is considered to be suitable for mixed-attribute datasets.

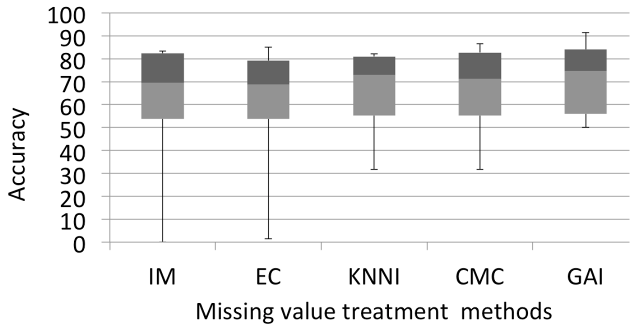


Figure 3: Boxplot of overall accuracies for each imputation method.

Finally, the results shown in Figure 4 represent the number of wins, draws and losses of each imputation method according to the Wilcoxon signed-rank test, with a confidence level of 95%. For instance, CMC has one win, two draws and one loss; while GAI has four wins and no losses or draws. In summary, the rank presented in Figure 4 shows that GAI is the first, followed by CMC, KNNI and IM are tied, and EC appears in last position.

6. CONCLUSIONS AND FUTURE WORK

Missing data imputation estimates plausible values to substitute the missing ones, aiming to reduce the bias imposed by this issue. In pattern classification, the missing data imputation process can be viewed as an optimization problem, where the goal is to find the best values to impute which will increase the accuracy of the classifier.

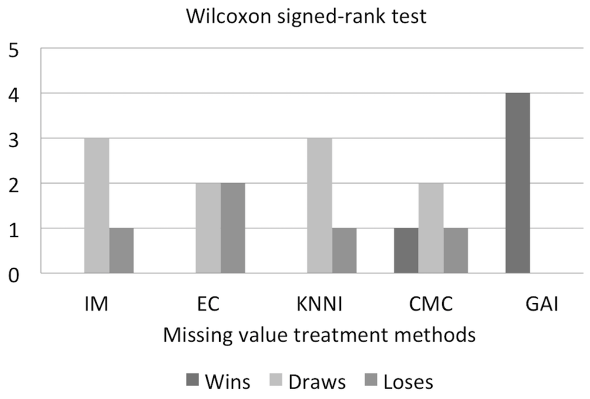


Figure 4: Wilcoxon signed-rank test scoreboard.

In this paper we proposed the GAI method, a novel data imputation method based on a genetic algorithm. GAI takes into account information from incomplete instances and classifier building, more specifically, the classification accuracy. The experiments were performed using six bench-marking datasets and three classification algorithms, in order to represent the three groups of classification methods: rule induction learning, approximate models and lazy learning. The experimental results indicate that GAI obtains results that are significantly superior to the other missing data treatment methods analyzed. Moreover, GAI has proved to be suitable for tackling mixed-attribute datasets with missing data.

For future research, we intend to expand the analysis to include more datasets and classification algorithms, and then to perform sensitive analysis of the results in order to understand the impact of GAI's parameters in the final analysis. We also plan to incorporate some statistical information into fitness functions and, finally, adapt it to regression tasks.

7. REFERENCES

[1] M. Abdella and T. Marwala. The use of genetic algorithms and Neural Networks to approximate missing data in database. *Computing and Informatics*, 24:577–589, 2005.

[2] A. Aussem and S. Rodrigues de Morais. A conservative feature subset selection algorithm with missing data. *Neurocomputing*, 73(4-6):585–590, Jan. 2010.

[3] G. E. A. P. A. Batista and M. C. Monard. An analysis of four missing data treatmente methods for supervised learning. *Applied Artificial Intelligence*, (Dm):519–533, 2003.

[4] M. L. Brown and J. F. Kros. The Impact of Missing Data on Data Mining. In J. Wang, editor, *Data Mining: Opportunities and Challenges*, chapter The impact, pages 174–198. IGI Publishing, Hershey, PA, USA, 2003.

[5] E. Eirola, A. Lendasse, V. Vandewalle, and C. Biernacki. Mixture of Gaussians for distance estimation with missing data. *Neurocomputing*, 131:32–42, May 2014.

[6] J. Figueroa García, D. Kalenatic, and C. Lopez Bello. Missing Data Imputation in Time Series by Evolutionary Algorithms. In D.-S. Huang, D. Wunsch

II, D. Levine, and K.-H. Jo, editors, *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence SE - 34*, volume 5227 of *Lecture Notes in Computer Science*, pages 275–283. Springer Berlin Heidelberg, 2008.

[7] J. C. Figueroa García, D. Kalenatic, and C. A. López Bello. An Evolutionary Approach for Imputing Missing Data in Time Series. *Journal of Circuits, Systems and Computers*, 19(01):107–121, Feb. 2010.

[8] J. C. Figueroa García, D. Kalenatic, and C. A. Lopez Bello. Missing data imputation in multivariate data by evolutionary algorithms. *Computers in Human Behavior*, 27(5):1468–1474, Sept. 2011.

[9] A. A. Freitas. *Data Mining and Knowledge Discovery with Evolutionary Algorithms.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002.

[10] B. Gabrys. Neuro-fuzzy approach to processing inputs with missing values in pattern recognition problems. *International Journal of Approximate Reasoning*, 30(3):149–179, Sept. 2002.

[11] P. J. García-Laencina, J.-L. Sancho-Gómez, and A. R. Figueiras-Vidal. Pattern classification with missing data: a review. *Neural Computing and Applications*, 19(2):263–282, Sept. 2009.

[12] M. Ghannad-Rezaie, H. Soltanian-Zadeh, H. Ying, and M. Dong. Selection-Fusion Approach for Classification of Datasets with Missing Values. *Pattern recognition*, 43(6):2340–2350, June 2010.

[13] J. W. Graham. Missing data analysis: making it work in the real world. *Annual review of psychology*, 60:549–76, Jan. 2009.

[14] E. R. Hruschka, A. J. T. Garcia, E. R. Hruschka Jr., and N. F. F. Ebecken. On the influence of imputation in classification: practical issues. *Journal of Experimental & Theoretical Artificial Intelligence*, 21(1):43–58, Mar. 2009.

[15] E. R. Hruschka, E. R. Hruschka, and N. F. F. Ebecken. A Bayesian imputation method for a clustering genetic algorithm. *Journal of Computational Methods in Sciences and Engineering*, 11:173–183, 2011.

[16] T. Liao, K. Socha, M. A. Montes de Oca, T. Stutzle, and M. Dorigo. Ant Colony Optimization for Mixed-Variable Optimization Problems. *Evolutionary Computation, IEEE Transactions on*, 18(4):503–518, 2014.

[17] M. Lichman. UCI machine learning repository, 2013.

[18] C.-P. Lim, J.-H. Leong, and M.-M. Kuan. A hybrid neural network system for pattern classification tasks with missing features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):648–653, 2005.

[19] R. J. A. Little and D. B. Rubin. *Statistical analysis with missing data.* Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, 1 edition, 1987.

[20] R. J. A. Little and D. B. Rubin. *Statistical Analysis with missing data.* Wiley, New York, 2 edition, 2002.

[21] Y. Liu and S. D. Brown. Comparison of five iterative imputation methods for multivariate classification. *Chemometrics and Intelligent Laboratory Systems*, 120:106–115, Jan. 2013.

[22] J. Luengo, S. García, and F. Herrera. On the choice of the best imputation methods for missing values considering three groups of classification methods. *Knowledge and Information Systems*, 32(1):77–108, June 2011.

[23] P. Mcknight, K. Mcknight, S. Sidani, and A. Figueredo. *Missing Data: A Gentle Introduction (Methodology In The Social Sciences).* The Guilford Press, Apr. 2007.

[24] V. Miranda, J. Krstulovic, H. Keko, C. Moreira, and J. Pereira. Reconstructing missing data in state estimation with autoenconders. *IEEE Transactions on Power Systems*, 27, 2012.

[25] L. Nanni, A. Lumini, and S. Brahnam. A classifier ensemble approach for the missing feature problem. *Artificial Intelligence in Medicine*, 55(1):37–50, May 2012.

[26] D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[27] J. L. Schafer. *Analysis of Incomplete Multivariate Data*, volume 11 of *C&H/CRC Monographs on Statistics & Applied Probability.* Chapman & Hall, 1997.

[28] J. L. Schafer and J. W. Graham. Missing data: our view of the state of the art. *Psychological Methods*, 7(2):147–177, 2002.

[29] J. D. A. Silva and E. R. Hruschka. An experimental study on the use of nearest neighbor-based imputation algorithms for classification tasks. *Data & Knowledge Engineering*, 84:47–58, Jan. 2013.

[30] D. J. Stekhoven and P. Bühlmann. MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics (Oxford, England)*, 28(1):112–8, Jan. 2012.

[31] L. Wohlrab and J. Fürnkranz. A review and comparison of strategies for handling missing values in separate-and-conquer rule learning. *Journal of Intelligent Information Systems*, 36(1):73–98, Apr. 2010.

[32] S. Zhang. Nearest neighbor selection for iteratively kNN imputation. *Journal of Systems and Software*, 85(11):2541–2552, Nov. 2012.

[33] S. Zhang, Z. Jin, and X. Zhu. Missing value estimation for mixed-attribute data sets. *IEEE Transactions on Knowledge and Data Engineering*, 23(3):110–121, Mar. 2011.