

A brief review of the main approaches for treatment of missing data

Luciana O. Silva and Luis E. Zárate*

Applied Computational Intelligence Laboratory, Department of Computer Science, Pontifical Catholic University of Minas Gerais, Minas Gerais, Brazil

Abstract. Missing data is a significant problem found in data mining projects and data analysis. Despite being a common problem, the missing data is dealt in a simplistic way and may lead to inconsistent knowledge discovery. Through literature review, it was possible to observe that the missing data mechanisms are not always considered when methods of treatment or imputation are chosen. This work presents a review about the main treatment methods of missing data that can be considered in a process of knowledge discovery in database. We emphasize that attention should be given to the identification of the absence mechanism, for the choice the most appropriate treatment method.

Keywords: Missing data, missing-data mechanism, data mining

1. Introduction

The knowledge Discovery in databases has the objective of extracting knowledge, from potentially useful unknown implicit data patterns [22]. In the preparation stage, a considerable effort is necessary to design and apply procedures such as cleaning, outliers analysis, data transformation, reduction of dimensionality and treatment of missing data, etc. In real domain, one of the great problems faced in data mining is the occurrence of missing data, which may happen in one attribute, in many attributes, in various instances or in a random fashion in database.

A missing value can also be called non-observed value or unknown value. This lack of information may result from uncontrolled circumstances, in which its value has not been inserted into the database, although there is a value in the domain to which it belongs, that is, there is a value for the attribute in the real world [83]. In some situations, the attributes have not been informed due to a matter of privacy or refusal. Sometimes it is just a consequence of certain measures not being available at the moment of collection of data. In other situations, the values are lost due to the failures that occur in automated or manual systems of measurements. Inside the KDD process the occurrence of missing data is extremely harmful and it can lead to wrong patterns and conclusions or decision making mistaken.

The first works related to missing data appeared in the 1930s [6]. The first literature review concerning data analysis with missing values was presented in 1966 [4]. Other studies identified in the literature include [18,32,54,55,57,76], which are purely statistical contributions.

*Corresponding author: Luis E. Zárate, Department of Computer Science, Pontifical Catholic University of Minas Gerais, Minas Gerais, Av. Dom José Gaspar 500, Coração Eucarístico, Belo Horizonte, 30535-610, MG, Brazil. E-mail: zarate@pucminas.br.

Table 1
Examples of missing mechanisms

| X | Y | MCAR absent 30% | MAR $x \geq 0.12$ | NMAR $y \geq 1755$ | X | Y | MCAR absent 30% | MAR $x \geq 0.12$ | NMAR $y \geq 1755$ |
|-------|---------|--------------------|----------------------|-----------------------|-------|---------|--------------------|----------------------|-----------------------|
| 0.144 | 1613.62 | 1613.62 | | 1613.62 | 0.096 | 1716.93 | 1716.93 | 1716.93 | 1716.93 |
| 0.112 | 1915.24 | 1915.24 | 1915.24 | | 0.144 | 1329.95 | 1329.95 | | 1329.95 |
| 0.123 | 1557.62 | 1557.62 | | 1557.62 | 0.117 | 1634.15 | | 1634.15 | 1634.15 |
| 0.117 | 2090.81 | 2090.81 | 2090.81 | | 0.101 | 1775.00 | 1775.00 | 1775.00 | |
| 0.107 | 1673.59 | | 1673.59 | 1673.59 | 0.123 | 1463.49 | 1463.49 | | 1463.49 |
| 0.144 | 1568.50 | 1568.50 | | 1568.50 | 0.096 | 1827.25 | 1827.25 | 1827.25 | |
| 0.096 | 1669.28 | | 1669.28 | 1669.28 | 0.096 | 1860.24 | | 1860.24 | |
| 0.096 | 1331.87 | | 1331.87 | 1331.87 | 0.107 | 1565.86 | 1565.86 | 1565.86 | 1565.86 |
| 0.107 | 1604.21 | 1604.21 | 1604.21 | 1604.21 | 0.112 | 2055.79 | 2055.79 | 2055.79 | |
| 0.101 | 1260.30 | 1260.30 | 1260.30 | 1260.30 | 0.112 | 1379.22 | 1379.22 | 1379.22 | 1379.22 |
| 0.107 | 1944.06 | 1944.06 | 1944.06 | | 0.144 | 1617.71 | 1617.71 | | 1617.71 |
| 0.096 | 1482.93 | 1482.93 | 1482.93 | 1482.93 | 0.101 | 1882.04 | 1882.04 | 1882.04 | 1882.04 |
| 0.096 | 1246.05 | | 1246.05 | 1246.05 | 0.144 | 1563.51 | | | 1563.51 |
| 0.112 | 1207.06 | | 1207.06 | 1207.06 | 0.117 | 2079.32 | 2079.32 | 2079.32 | |
| 0.107 | 1507.78 | 1507.78 | 1507.78 | 1507.78 | 0.101 | 1787.43 | 1787.43 | 1787.43 | |
| 0.133 | 1739.79 | | | 1739.79 | 0.139 | 1746.10 | | | 1746.10 |
| 0.107 | 1819.95 | 1819.95 | 1819.95 | | 0.101 | 1905.32 | | 1905.32 | |
| 0.128 | 1627.87 | 1627.87 | | 1627.87 | 0.128 | 1743.91 | 1743.91 | | 1743.91 |
| 0.101 | 2133.93 | | 2133.93 | | 0.112 | 1688.91 | 1688.91 | 1688.91 | 1688.91 |
| 0.101 | 1635.50 | 1635.50 | 1635.50 | 1635.50 | 0.112 | 1543.62 | 1543.62 | 1543.62 | 1543.62 |
| 0.096 | 1498.84 | | 1498.84 | 1498.84 | 0.128 | 1580.29 | 1580.29 | | 1580.29 |
| 0.107 | 1682.44 | 1682.44 | 1682.44 | | 0.112 | 2023.90 | | 2023.90 | |
| 0.128 | 1648.09 | 1648.09 | | 1648.09 | 0.101 | 1223.90 | 1223.90 | 1223.90 | 1223.90 |
| 0.107 | 1405.13 | 1405.13 | 1405.13 | 1405.13 | 0.112 | 1898.66 | | 1898.66 | |
| 0.128 | 1648.08 | 1648.08 | | 1648.08 | 0.112 | 2108.30 | 2108.30 | 2108.30 | |
| 0.128 | 1592.11 | | | 1592.11 | 0.107 | 1397.29 | 1397.29 | 1397.29 | 1397.29 |
| 0.101 | 1556.22 | 1556.22 | 1556.22 | 1556.22 | 0.139 | 1700.49 | 1700.49 | | 1700.49 |
| 0.101 | 1420.39 | | 1420.39 | 1420.39 | 0.096 | 2026.11 | 2026.11 | 2026.11 | |
| 0.117 | 2251.83 | 2251.83 | 2251.83 | | 0.096 | 1756.42 | 1756.42 | 1756.42 | |
| 0.096 | 1689.46 | 1689.46 | 1689.46 | 1689.46 | | | | | |

A great milestone of the missing data literature was the formalization of the mechanisms of absence proposed in [89]. The author concludes that to deal with missing data, the mechanism that causes the absence must be explicitly considered, and for that to happen, it is necessary to identify the models that represent such processes. Rubin proposed a taxonomy to the mechanisms that generate missing data: MCAR (missing completely at Random), MAR (missing at random) and NMAR (not missing at random). Table 1 shows the effect of the missing data mechanisms on the variables X and Y . The data are a sample for visualization the impact of the mechanisms on the original probability distribution function (Fig. 1). The MCAR mechanism occurs when the probability of absence for a variable Y is not related to the value of Y itself as well as to any other variable in the data base. For this mechanism the absence of data in the variable Y is totally random. In this case the distribution of the variable Y containing missing data retains their original characteristics of the distribution containing complete data. Since the observed cases are a random sub-sample of all the case. Figures 2(a)–(c) show the effect of MCAR mechanism to 25%, 50% and 75% of absence. Note that the distribution functions retain their original characteristics.

The MAR mechanism occurs when the probability of absence for a variable Y depends on other variables, but does not depend on the variable Y itself. As example, suppose that systolic blood pressure of voluntary participants is recorded during two months. But only participates in the second month those who had their measurements above 140, a level used for diagnosing hypertension. Notice that on the

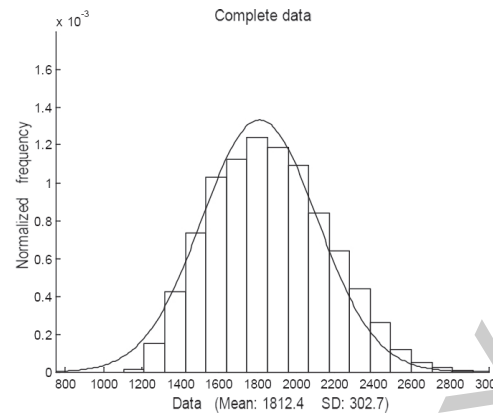


Fig. 1. Normalize histogram for complete data.

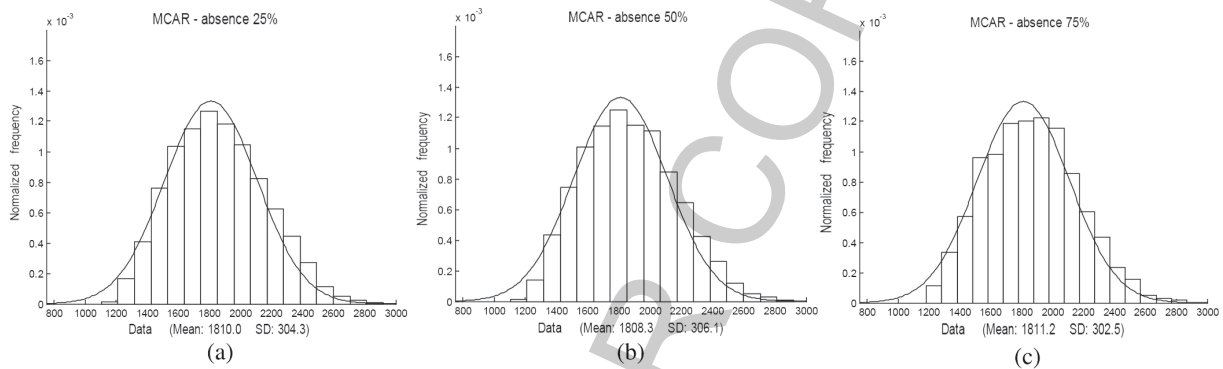
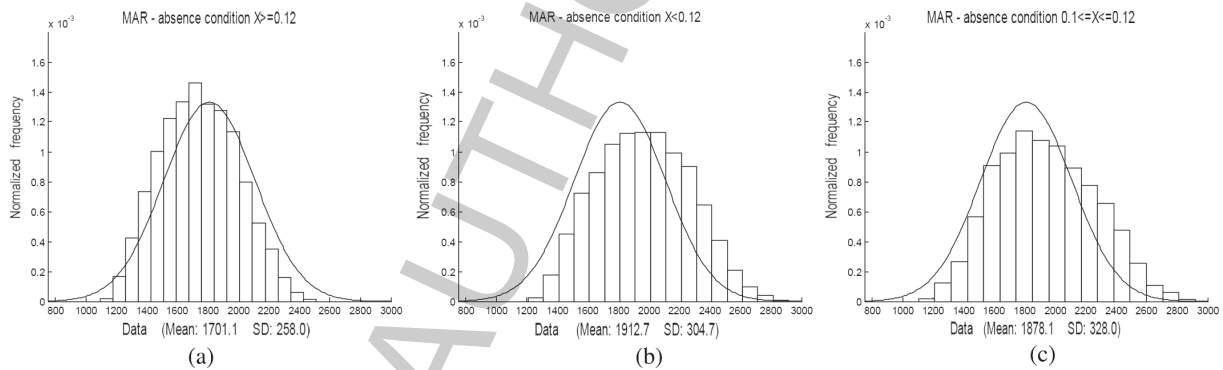


Fig. 2. (a) Normalize histogram for MCAR, absence 25%; (b) Normalize histogram for MCAR, absence 50%; (c) Normalize histogram for MCAR, absence 75%.

Fig. 3. (a) Normalize histogram for MAR, absence condition $X \geq 0.12$; (b) Normalize histogram for MAR, absence condition $X < 0.12$; (c) Normalize histogram for MAR, absence condition $0.1 < X \leq 0.12$.

second month there may be records of participants with absent information. Figures 3(a)–(c) show the effect of MAR mechanism for three conditions of absence in the X variable. Notice that the distribution functions are different of the original distribution, Fig. 1.

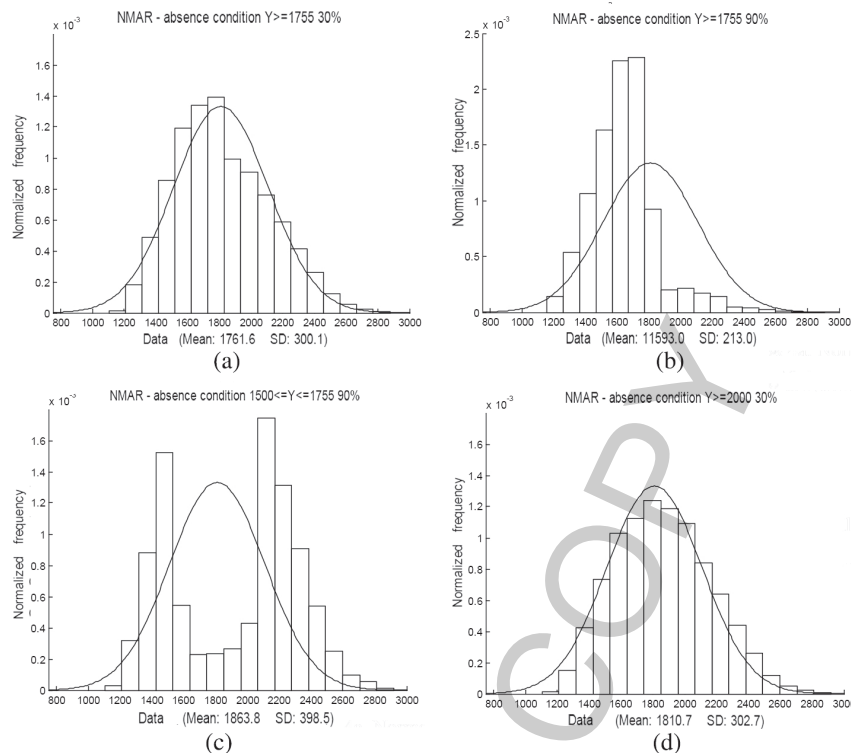


Fig. 4. (a) Normalize histogram for NMAR, absence condition $Y \geq 1755$, absence 30%; (b) Normalize histogram for NMAR, absence condition $Y \geq 1755$, absence 90%; (c) Normalize histogram for NMAR, absence condition $1500 \leq Y \leq 1755$, absence 90%; (d) Normalize histogram for NMAR, absence condition $Y \geq 2000$, absence 30%.

Finally, the NMAR mechanism, also known as ‘not ignorable’, occurs when the probability of absence for the variable Y is related to the own value of Y . For example, people who earn a higher salary, prefer not to inform or lying about their salary. Figures 4(a)–(d) show the effect of NMAR mechanism for four absence conditions in the Y variable. Notice that the distribution functions have variations evident in relation to the original characteristics. Experimentally it was possible to observe that for absences above 30%, the original characteristics are lost, see Fig. 4(d). Notice that, a datum containing false information should be also considered a missing value. Hence the difficulty and challenges of dealing with databases containing missing data. On the other hand, considering the example that, after a certain age, women avoid informing their age. It is possible to observe that there is a hybrid mechanism, situation where the mechanism can be MAR and NMAR. In this case, the mechanism of absence depends of the gender and of advanced age of the people. Summarizing, for the MCAR mechanism, the cause of absence is completely random. As for the MAR, the cause obeys a conditional randomness, while the NMAR mechanism shows systematic reasons for missing data. In [12], the authors simplify the taxonomy of [89] by pointing out the causes of missing data. In a general way, one can not affirm beforehand that the absence mechanism is MCAR, MAR or NMAR. There are not specific techniques to determine what is the mechanism of absence for a determined set of data. If the mechanism is identified, more efficient methods for recovery or imputation of missing data could be proposed.

The treatment of missing values must be carefully analyzed, because if it is performed in an inadequate way, many problems may occur in the discovery of patterns in the data mining processes. Perhaps the most serious problem is the elimination of missing data with the NMAR pattern. This pattern can not be

ignored, since it causes severe distortions on the results. Another problem comes up when the volume of missing data is massive. In this case, the elimination may occur on the records or on the attributes that contain them. Anyway, the elimination of such instances can be a factor that affects the final quality of the extracted knowledge.

The first authors to affirm that the best way of dealing with missing data is by not having them were Orchard and Woodbury [76]. The authors compare missing data with the accidents, which are not planned, but one can avoid them by taking cautionary measures. Thus, precaution must be taken in capture of data, in order to avoid the occurrence of missing data. Later on, [7,12] reiterated the statement of Orchard and Woodbury [76]. That is an unrealistic recommendation, since missing data is a real problem in data bases of various fields of knowledge, which has been the target of many researches in the past few years. Lynch [63] mentions the difficulty to publish empirical studies on Sociology without the discussion of how the missing data were treated. In his turn, Shadish [97], in the field of Psychology, admits the rejection of scientific papers that do not make the treatment of missing data explicit. In many areas of knowledge, review papers and tutorials are periodically published with the objective of supporting researchers in the treatment of missing data. Schafer and Graham [94] present a broad review of the existing methods for dealing with missing data. Other review studies are found in [69,107]. Additionally, [2,12,95] compared the techniques available in statistical packages.

The missing data analysis can begin by defining a matrix R of absences (Little and Rubin [56]), that describes which values are observed and which ones are missing ($r_{i,j} = 0$ observed value, and $r_{i,j} = 1$ missing value). In matrix R , rows correspond to observational units (instances) and columns correspond to variables (attributes). The distribution of the absence depends on the complexity of the pattern. According to [56], basically the patterns can be general (random) or specific. In the general pattern the missing data are found in any records of the data set. The specific pattern can be univariate, in other words, is restricted to one variable only of the data base, or can be monotone, when the amount of missing values, from one sample to the other, is always growing. As an example of this last case, Hruschka [37] presents a research with data collect performed in three stages. In the first stage, there is no missing data; in the second stage, performed a few years after the first one, with the same people, not all participants were found and consequently some data are missing. Finally, in the last stage, only the participants of the second stage were looked for and, again, not all of them were found, which increased the number of missing values.

In modern approaches of missing-data the absence is considered as a probabilistic phenomenon and the matrix R contain a set of random variables having a joint probability distribution.

Another important contribution in research area of missing data was the formalization of the Expectation-Maximization (EM) algorithm [18], a computational method for efficient estimation of incomplete data. The EM algorithm represented a milestone in the way as statisticians have treated the missing data. From that study, statisticians began to observe missing data as a source of variability to be taken into consideration in statistics analysis.

In this study, a brief review of the main contributions about the techniques and methods for dealing with missing data in data set, which should be considered in data mining processes, is presented. In order to organize the techniques and methods, we prefer consider three kinds of procedures adopted in the literature: based on removal and imputation, based on models and based on direct manipulation of missing data.

2. Technique and methods for missing data treatment

The specialized literature classifies the methods of dealing with missing data in different ways. Def-

initely, there is no consensus among the authors about the classification of the existing missing data techniques. The literature proposes classifications with a higher or lower level of generalization. For example, Batista and Monard [10] organized the techniques and methods as: a) Ignoring and disposing missing data (disposing records with attributes that present too many missing data); b) Parameters estimation; and c) Imputation. From the statistical viewpoint, according to Soares [101] the methods for analysis of data set containing incomplete data are grouped in a) imputation procedures; and b) procedures based on models. In the first group, procedures have as aim to complete the missing data, while in the second group, probability models, as Bayesian inferences models, are proposed.

Hruschka et al. [38,39] summarize methods for treatment of missing values as: a) ignoring the records with missing values; b) filling them in manually; c) replacing the missing value by a constant, by the average or mode; and d) assigning the most probable value. The techniques b) e c) correspond to imputation procedures. These methods can distort the characteristics of the records, since that, will be changes the scenario in which the data were generated.

Myrtveit et al. [70] divide the techniques into: a) ignoring the records with missing values; b) imputation-based techniques; c) weight-assignment-based techniques; and d) models-based techniques.

In their turn, the classification proposed by [56] is restricted to statistics methods. According to them, the procedures are grouped in four categories from the viewpoint of data analysis: a) procedures based on complete cases; b) procedures based on available cases; c) procedures based on imputation; and d) procedures based on weight-assignment.

Magnani [65] proposes a classification that is more didactical and more refined, which tries to take into consideration all the possible techniques for dealing with missing data. They are: a) conventional methods as imputation and parameters estimation; and b) direct manipulation of missing data.

As mentioned, in this study we prefer to group procedures adopted in literature based on the way as these treat the database containing missing data: 1) procedures based on removal; 2) procedures based on imputation; 3) procedures based on models; and 4) procedure based on direct manipulation of missing data. The techniques and methods were organized as following:

- 1) Procedures based on removal of records.
 - a. Removal of incomplete records.
 - b. Removal in pairs.
 - c. Removal of attributes with missing values.
 - d. Removal with weights.
- 2) Procedures based on imputation.
 - a. Global imputation based on the missing attributes.
 - b. Global imputation based on non-missing attributes.
 - c. Local imputation.
 - d. Multiple imputation.
 - e. Composite imputation.
- 3) Procedures based on models.
 - a. Maximum Likelihood methods.
 - b. Bayesian models.
 - c. Models distribution for the missing values.
 - d. Models based on Neural Networks.
- 4) Procedures based on direct manipulation of missing data.

2.1. Procedures based on complete cases

The procedures based on removal of records containing missing value are simple and are usually a default option of the statistical packages.

The Removal of incomplete records is called Listwise deletion, which is the simplest method of obtaining a complete data base [74]. This method is satisfactory when the quantity of missing data is small. On the other hand, when used in situations with many missing data it implies loss of data and information. Besides, this method must be used only if the mechanism of absence obeys the MCAR pattern. Since that the observed cases are a random sub-sample of all the case. In data mining context, the removal of records reduces the ability of finding consistent patterns, implying worse results, since fewer records are used in the training process or in the adjustments of the prediction model [23].

The method of Removal in pairs (pairwise deletion) is a variation of the listwise deletion method, in which incomplete records are kept only when the attribute under analysis does not have missing value. So, like the listwise deletion method, the pairwise deletion can cause biased results if the missing data is not MCAR [56]. The pairwise deletion method has the advantage of using all the variables available and, consequently, of increasing the statistical power for analyzing the data set. However, it requires elaborate procedures to deal with all the absences in the database.

Another conventional method is the Removal of all the variables with missing values, which, in its turn, is not an interesting option, due to evident reasons.

Finally, Removal with weights is a more refined version of the records exclusion method. After the exclusion of incomplete records, the remaining data are re-weighted in a way that their distribution looks like the complete sample. The weighting procedures are often used in the application of polls questionnaire, for the adjustment of non-answered units [56]. In [75], a comparative study about of weighting method approach was presented.

2.2. Procedures base on imputation

Imputation is the procedure of replacing the missing values. According to Rubin [91], imputation has two objectives. The basic objective is to allow final data users to use their analysis tools on any database as if there were no missing data. The second objective is to obtain, statistically valid inferences. On the other hand, Junninen et al. [48] points out that imputation methods cannot be used as a kind of “statistical alchemy” to help remedy missing data situations. In specific situations, the imputation is not applicable. For example, in asynchronous distributed data collection, in which the nature of the data is temporal, and/or spatial [28]. For example, data originated from an intruder detection system, in a computer network, have these characteristics. In this case, the absence of data may be even apparent, because the missing data can be empty information, if many sources that generate intrusion are considered. Pyle [83] defines as empty datum, the datum that does not exist in the real world, and for this situation, the author recommends the use of techniques that can deal with missing data. Donders et al. [19] presented an introduction to imputation methods and their implications in real problems. Hu et al. [40] briefly describe about thirty distinct imputation methods.

Twala et al. [108] point out that imputation hides the inaccuracy of the data, leading to invalid test intervals and confidence, since the imputed values are derived from the existing data. The imputation methods normally predict values that are more “well behaved” than real values. This fact can increase the risk of simplifying excessively the problem under study, considering that the most important aspect of substitution of missing data is the non-distortion of the original characteristics of the sample [37].

The global imputation method, based on missing attribute uses the existing values to fill in the missing values. These methods use a central tendency measure as the average, median or mode. In this methods, the standard deviation is distorted, because the extreme values are brought into the distribution, thus reducing the variance, even on the MCAR pattern. It is advisable that the imputation does not alter the variance of the sample. An alternative is the introduction of a random disturbance into the value of the average. Although such disturbance tries to avoid the distortion of the results it is not a satisfactory method [65], especially by the distortion of instance or record.

The global imputation based on non-missing attributes uses the existing correlations between variables with missing values and available variables. One strategy is the imputation through regression. The attributes containing missing values are treated like independent variables, and the regression is executed to impute the missing values. One disadvantage is that regression is based on a chosen model, which not always represents the data in a satisfactory way [52].

Another global imputation technique is based on association rule. The association rules seek inter-connections of records it attempting to expose characteristics and their dependencies. The objective is to find rules in the forms of antecedent and consequent, which are frequent and valid, and that meet minimum criteria of support and confidence. The support is related to how frequently a rule occurs in a table and the confidence reflects the validity and expresses the quality of a rule, indicating how much the occurrence of their antecedent can assure the occurrence of its consequent.

It is important to observe that the missing data bring problems to the algorithms based on association rules, because they distort the measures of support and confidence, which imply in the loss of the good rules. An innovation for dealing with missing data was the creation of extensibility and representativeness and measures, which complement the measures of support and confidence. Representativeness indicates the quantity of item sets that do not have missing values in the attributes. An itemset is extensible if it has a frequent representative superset. An implementation of these new measures is found in [13], through the Xminer algorithm. In their turn, Nayak and Cook [71] present the approximate association rule algorithm that considers the existence of noisy and missing data.

Finally, Mampaey [66] points out that the generated rules can contain items with absences and that this characteristic can be explored in order to mine interesting rules about the mechanism of absence that affects the database. An item is an attribute value written as $(A = a_i)$. The absence in the item is written as $(A = ?)$, indicating a missing value for the attribute A . In this way, rules of format $(A = a_i) \Rightarrow (B = ?)$ can be found.

In [122] a method based on Formal Concept Analysis to find errors in objects (examples or instances) not inserted in the formal context was discussed. This method proposes to use a implication basis (generated from formal context) which represents the dependencies between attributes, to find errors in the objects. This method can be extended as a new mechanism to impute value for missing data.

The MVC (Missing Value Completion) method, proposed by Ragel and Crémilleux [85], uses robust association rules for dealing with missing data through the RAR-Robust Association Rules algorithm. In Ragel and Crémilleux [86] the application of the MVC method in a real database is demonstrated. The RAR algorithm finds out association rules in sets of data which do not contain missing data in their records. These rules are obtained from a subset of data which do not have any missing value, called valid database. The MVC algorithm decides which value to impute based on the consequents of the rules. As a heuristic criterion, the user intervention in the process of data complementation can occur. The user decides which rule or rules to use, based on parameters like support and confidence of each rule, and the consequent of the rule is assumed as the value to be imputed. A practical utilization of association rules to complete missing data is found in [117]. The work is based on the idea that rules of association

describe the dependencies of relationships between attributes in a database, where every itemset that has attribute with missing value, must also have similar relationships.

The procedures based on local imputation presents three methods: hot-deck, cold-deck and KNN. The hot-deck method is divided into two stages: first the records are subdivided into classes and then, for each incomplete record, the imputed values are chosen based on records of the same class. Magnani [65] points out some advantages of the hot-deck: reduction of the standard error without imposing a rigid model; production of a data set without missing data, and preservation of population distribution. Moreover, this method allows distinct imputation techniques to be used for each group generated. Soares [102] observes that the division of the data into groups can bring great benefits to the imputation process.

In its turn, the cold-deck method replaces the missing values by a constant. This external constant can be obtained from historical data or from knowledge of the problem domain. Little and Rubin [56] state that the theoretical base for this method is not satisfactory.

The K-Nearest-Neighbor (KNN) method is a variation of the One Nearest Neighbor (1NN) method. KNN is often used in the records classification task. KNN, in turn, to estimate the class of a new pattern X , calculates the k nearest neighbors to X and classifies X as belonging to the class that appears the most frequently among its k -neighbors. The parameter " k " indicates the number of neighbors used by the algorithm during the test phase, and allows a more refined classification [47]. KNN is also used as a method of imputation and, in this case, the k records near the record with missing values are combined in the estimation. Batista and Monard [9] analyze KNN as an imputation method and they conclude that the results obtained with KNN are superior to those obtained by the learning algorithms CN2, C4.5 and by the mode or average imputation. Besides, they point out that KNN, even facing a great amount of missing data, maintains the quality of imputation. Cover and Hart [16] proved that, for any number n of samples, the 1NN rule has strictly lower probability of error than any other KNN rule against certain classes of distributions, and hence is admissible among the KNN rules.

Multiple imputation (MI), proposed in [89], came as a flexible alternative to the similarity methods. Multiple imputation was used for the first time in public data from census and polls, due to the need to make complete data available for final users. Multiple imputation aims to reduce the uncertainty inherent to the imputed data and it is made up of three stages: imputation, analysis, and results combination. In the imputation stage, the sets of plausible values for missing observations are created by using some proper imputation method. A proper imputation method can be based either on an explicit model (parametric, as linear regression) or on an explicit method (as, for example, hot-deck) which incorporates the appropriate variability between repetitions. Each set of plausible values is used to fill in the missing data and to create a complete database. In the analysis stage, each database is analyzed by using complete data methods. Finally, in the results combination stage, the results are combined, aiming to reduce the uncertainty.

Rubin [90] recommend that the imputations are created through a Bayesian process, specifying a parametric model for complete data, apply a previous distribution for the unknown parameters of the model, and simulate several executions from the conditional distribution of the missing data, according to the data observed, by using the Bayes Theorem. According to Shafer and Graham [94], multiple imputation, from the Bayesian perspective, is a natural motivation. Other than the Bayesian perspective it is possible to use other imputation methods; among them the likelihood method, discriminant analysis, and logistic regression. To apply the Multiple imputation method, the mechanism must obey the MAR pattern [90].

Multiple imputation is an approximation of the Maximum Likelihood method (ML). Multiple imputation tries a few plausible values for the missing data, while the ML method integrates all possible values, giving more weight to those values that are more plausible. The result of the ML estimation is the same

as if infinite imputations were executed in multiple imputation. The disadvantages of the ML method are the assumptions about the distribution of the missing data. So, when it is possible to live with those assumptions, ML becomes the ideal option; otherwise, MI (Multiple Imputation) is more flexible.

Composite imputation was proposed in [102]. In this method, the imputation process of a missing data is preceded by the application of other tasks as, for example, the data clustering and the attribute selection, with the objective to improve the quality of the imputed data. The method proposes the application of data complementation committees for the imputation process. The committees are formed by a group of trained classifiers iteratively. The simplest approach is a simple majority vote, where the classifiers are combined into a voting strategy. As a final result, the response with the most votes is considered the committee's response. The classification committees use many suggestions for a decision-making process or for a combination of results that can produce a new classification for a database record. In Farhangfar et al. [21] a study about the effect of imputation on classification error for discrete data, considering imputation methods as: mean, hot-deck, Naïve-Bayes and others are presented. The analysis showed that all imputation methods except the mean imputation, improve classification error.

Recently, in Luengo et al. [61,62] several imputation methods for the classification task were evaluated. In [61] the aim was to improve the performance of Radial Basis Function Network classifier in data sets containing missing data. The results showed that Event-Covering imputation method [117] presents a good performance with this type of classifier. The Event-Covering approach detects statistically significant event associations and can deduce a certain structure of inherent data relationships. The approach allows the selection of useful information inherent in the data when the causal relationship is uncertain or unknown and discover and disregard uncertain events which are erroneous or simply irrelevant.

2.3. Procedures based on models

Data modeling includes statistical, probabilities and learning techniques for obtaining a model that is able to represent, in a generic way, the characteristics of the data. Those models can be used as inference mechanism in the imputation of the appropriated values. In this category, the likelihood algorithms, the Bayesian methods, the selection models, the pattern-mixture method, and the neural networks stand out.

Maximum Likelihood methods (ML estimate the parameters of the statistical distribution function for an incomplete sample. The objective is to find a model that represents the data set and, by doing so, make the regression of any existing missing value. The Expectation Maximization algorithm (EM), proposed by Dempster et al. [18], consists of an iterative process of likelihood which estimates the parameters of the probability density function of a sample. This method is based on the relationships observed among all variables and injects a degree of random error in order to reflect the uncertainty of the imputation. In case there is an idea of the missing values, the parameters estimation of the model becomes simple. Similarly, once the parameters of the model are known, then it is possible to obtain non-distorted predictions for the missing data. This interdependence between the parameters of the model and the missing values suggests an iterative method, in which first the missing values are predicted, based on the values assumed for the parameters, then these predictions are used to update the estimated parameters and so successively, until the convergence of the algorithm. A way of monitoring the EM convergence is by checking the loglikelihood function, which must build up at each iteration, until it becomes stable [93].

The iterative algorithm can take long time before convergence, mainly in presence of many missing values and many variables. The convergence rate of EM is proportional to the amount of information missing in the database [56]. However, as any other non-linear optimization procedure, it suffers from the

local minimum problem, being sensitive to the initial values of the parameter. To use it, it is necessary to specify the distribution of the sample, which is not always possible in the KDD process [65]. Variations of this method aim to improve the convergence rate as, for example, the ECM, AECM, and PX-EM algorithms [92].

Williams et al. [116] presents a logistic regression algorithm for classification of incomplete data. The conditional density function is estimated by using a Gaussian Mixed Model (GMM) and the estimation of the parameter uses the EM and the Variational Bayes EM (VB-EM) algorithms. In this study, the problem of missing data is solved by avoiding the heuristics of imputation.

Bayesian models are broadly used for treating missing data, both in the statistical area and in the data mining area.

According to Magalhães [64], Bayesian networks are structures that combine the distribution of probabilities and graphs with a set of variables of interest. A G graph is made up of a set of V vertices, connected by a set of A arcs that represent the connections between vertices. A Bayesian network needs a directed graph (or digraph) that establishes a relation of direct dependence (cause/effect) between the vertices. In the directed graph, the edges are called arcs, and the relation defined by the set of arcs A is not symmetric, and there is an orientation in the relation between the vertices. An acyclic directed graph (ADG) is a graph composed by vertices and arcs in which cycles do not occur. A Bayesian network is a pair $B = (G, \theta)$, defined over a set of random variables $X = \{X_1, X_2, \dots, X_n\}$, where each X_i corresponds to a vertex, G is an acyclic directed graph, called structure, and θ is a set of parameters that specify the distribution of conditional probabilities that satisfy the Markov's condition. Markov's condition for a Bayesian network says that any vertex of the network is conditionally independent of its non-descendants conditioned to their parents. Once the network structure and its set of parameters θ are defined, it is possible to calculate a function that attributes a value for each ADG, based on the data, called score function, whose calculation depends on the distribution of probabilities associated to the random variables of the problem in question. A Bayesian model can be used in the prediction of a parameter θ or any other value that is not observed in any of the other variables of the model. This characteristic of working with unknown values is a motivation to use this theory in the treatment of missing data.

García and Hruschka [25] use the Bayesian classifier as imputation tool for classification problems, considering the Rubin's taxonomy ([89]) for the absence mechanism. The work illustrates how the allocation process influences the classification task. A similar work is proposed in [59], which proposes a Bayesian imputation method, using the imputation attribute as class attribute to build the Bayesian classifier.

Friedman [24] showed a method for learning the Bayesian network structure and estimation of parameters from incomplete data, based on an extension of the EM algorithm.

A new imputation method based on the representation of the knowledge, through Bayesian networks, is presented in [37]. In that study, the Bayesian networks are used as a inference mechanism in the prediction of the adequate values for the substitution of the missing values.

Riggelsen and Feelders [87] propose a Bayesian method for learning Bayesian networks with incomplete data. The objective is to obtain the posterior distribution of the models from the data observed.

It is important to point out that the Bayesian inference process involves the calculations of the posterior distributions from a previous distribution. Such calculation is restricted to the data observed; so, the bigger the sample used for obtaining the posterior distribution the smaller is the distribution influence on the previous prediction [37]. However, in presence of many missing data, the calculations of posterior distributions become compromised.

On the other hand, considering the Models distribution for the missing values mechanism. For example, for the NMAR mechanism, one must explicitly specify a distribution for the absence that will be added to the model complete data. There are two methods of implementation of this process: Models selection, and Pattern-mixture.

The Models selection method was firstly used in Econometrics to describe how the problem of an answer to a questionnaire item depends on the item itself [8,33]. In Models selection, distribution for the complete data is specified first, after that; the assumption of the distribution for the missing data is required. It is necessary that distribution assumptions of the missing data be done for identifying the model. However, such assumptions are not always enough to identify the model. Thus, Models selection presents convergence problems due to the estimation of many parameters. These models are considered unstable for scientific applications and generally generate more questions than answers [49].

As an alternative for models selection, Little [53] describes a new class of methods based on the formulation of Pattern-mixture. This class of models does not require precise specification of the absence mechanism. The Pattern-mixture model categorizes the different absence patterns into a predictor variable, incorporated into the statistical model. So, it is possible to determine if the absence pattern has a predictive power in the model, by itself (main effect), or together with other predictors (interaction effect). The primary disadvantage of this method concerns to the convergence. If the number of patterns and the number of variables with absence are too high, in relation to the number of cases or records of the sample, the model may not converge, due to the insufficiency of data to support the use of many main effects and many interaction effects.

The patterns-mixture is a multiple imputation procedure that is executed over a variety of assumptions about the mechanism of absence. As an example, consider a database of people with the attribute weight. In multiple imputation, it is assumed that the individuals who informed their weights are similar to those who did not. The Patterns mixture model, for example, assumes that the individuals who did not inform their weight are, in average, heavier than the ones who did. Clearly, it is an arbitrary assumption: the idea of patterns mixing is to try a variety of plausible assumptions and to check them against the result. Thijs et al. [105] describe adjustment strategies for the patterns mixture model, like restrictions identification and adjustment through pattern.

Models based on Neural Networks are mentioned in many studies as a strategy for dealing with estimating or imputation missing data. Artificial neural networks are non-linear mathematical functions that detect and represent sophisticated patterns and are used for predictive tasks (classification) and descriptive tasks (grouping). There are also some studies that bet on the capacity of networks to deal with missing data, due to its robust structure and generalization capacity. Among the studies that use neural networks as estimators, some stand out: [1,26,81,100,114,120]. Zárate et al. [120] demonstrated the utilization of an ANN that estimates missing values, in an imbalanced database with massive missing data, from a Minimum Set of Consistent Records (MSCR). The problem of database was conducted seeking a subset of data sufficiently consistent for learning of the neural network. Moreover, they point out that the efficiency of the neural representation depends on the quality more than the number of examples considered to represent the problem domain. The studies of [46,73,77,110] show evidence of the capacity of these networks for dealing with incomplete data situations.

Wie and Tang [114] bet on the data preparation stage and propose a framework for the imputation of missing data, by using a Self Organization Map (SOM) neural network, through which the data are grouped in subsets. This way, the imputation accuracy can be increased. Yu et al. [119] proposed an integrated scheme of data preparation for data analyses that use neural networks. The scheme proposed helps the learning process, reduces the complexity of neural model and increases the performance of the data analysis task.

In [1], the method presented combines genetic algorithms and neural networks to help in missing value estimation. The genetic algorithm is used to minimize the error function derived from an auto-associative neural network. In turn, [81] present the model ISOM-DH, based on principal component analysis and Self Organization Map (SOM) networks to deal with missing data. Still using ANN, García-Laencina et al. [26] proposed a multitask learning method that parallelly performs the imputation task and the classification to estimate missing data. The imputation is guided by the classification task, and the imputed values are those that contribute to improve learning.

Another proposal presented in [72] uses sets of Fuzzy-ARTMAPs for the classification task, and multi-layer perceptrons for the regression task, as an estimator. The method proposed is adequate for online estimation operations using previously trained neural networks.

Jiang et al. [46] propose a model of artificial neural networks for the classification of incomplete data. The incomplete database is divided into groups of complete sub-databases, which are used as training subsets. The method proposed uses all the information provided by the data with missing value, keeps maximum consistence of the incomplete data and avoids dependence on distribution assumptions.

Following another line of investigation based on modification and adaptation of the neural network structure, [110] bet on artificial neural networks capable of dealing with presence of missing data. The main idea is to find an appropriate configuration for the neural representation. The proposed algorithm adapts the network structure according to presence and absent of individual values of the input vector. When a value is absent the algorithm selects the protection status of the network neurons, according to missing values of the data vectors. The protection of the neuron is performed through a flag that indicates if the neuron is protected or not. If the neuron is protected, it means that this will not be considered in the weights calculation of the neural network. Following these line of investigation, [77] using the recurrent Bayesian Confidence propagation Neural Network (BCPNN), proposed in [35], to deal with missing data in the training stage. BCPNN uses weights which are less sensitive to co-inactivity of nodes and thus has the potential for efficient pattern recognition.

Ingrassia and Domma [43] relate the statistical models and the neural models. The training of an ANN is compared to a maximum likelihood estimation problem. A study comparing the techniques of auto-associative neural networks combined with genetic algorithms to those of the Expectation Maximization (EM) is presented in [73]. The conclusion emphasizes that the EM techniques present superior performance when there is little or no interdependence between the input variables, while the ANN techniques and the genetic algorithms show better performance when there are inherent nonlinear relationships between some input variables, which actually happens in real problems of data mining.

2.4. Procedures based on direct manipulation of missing data

The methods of direct manipulation of missing data can treat missing values without the need for imputation. In this category, the algorithms based on decision tree stand out. The methods based on trees do not make any assumption about the way data are distributed and do not demand a structured specification of the model, that is, they are nonparametric methods. Although the low correlation of the attributes can negatively influence the performance of these algorithms. The concern of making the algorithms robust and efficient with partial absence of the data is made evident in [84].

Weiss and Indurkha [115] discuss methods for learning and application of decision rules for data classification in datasets containing large quantity of missing values. The authors present a method to obtain induction rules, from records with missing data, in which the format of the rules is no different than rules for data without missing values, and no special features are specified to prepare the original data or to apply the induced rules. This method generates rules in the Disjunctive Normal Form (DNF).

Fortes et al. [23] and Kewett [34] present a method for dealing with missing data, based on the Top-down Induction Decision Tree algorithm. Three aspects were observed: the splitting criterion, value allocation for attributes with missing values, and predictions of new observations. In relation to the second aspect, the assignation of values to missing values by means of a decision theory approach, taking into account the information of observations and classes with associated parameters (confidence and error). The error parameter measures how near or how far the value is from the original value of the attribute. After applying a splitting criterion, a decision tree is obtained. This decision tree can be used to predict the class of an observation. The authors assume that the unknown values are missing completely at random (MCAR), which restricts the generalization of the method.

Gorodetsky et al. [28] present a method that generates two sets of rules used as inferior and superior limits for other sets of rules that correspond to the associations of missing data. So, based on these sets, the subset of rules is selected to be used in the classification. This method is recommended for applications in which the imputation is not theoretically justified.

Schoner [96], in his turn – motivated by the fact that imputation introduces dependences or regularities not present in the real data – explores new directions through the utilization of *SVM (Support Vector Machine)* and *maximum entropy*, and concludes that these methods can improve the prediction for an incomplete database. For example, SVM is a supervised learning method, used for classification and regression. One property of this method is that it simultaneously minimizes the classification error, and maximizes the geometrical margin. The SVM training always finds a global minimum and, its geometrical interpretation creates fertile ground for further investigations. On the other hand, entropy is a measure of the information of a probability distribution [99]. The maximum entropy principle allows selecting, from a family from consistent distributions, the one that maximizes the entropy [45]. This theory is able to determine probabilities distribution with small data samples, which makes it efficient in studies where available data are scarce.

Other contributions explore the rough sets theory, introduced in [78]. This theory has properties that allow eliminating irrelevant attributes, through the process of reduction of the information system, based on the definition of reducts, which are the subsets of attributes capable to keep the same properties of the extracted knowledge, when the latter is made by considering all attributes of database. The objects (instances) contained in a system, are grouped in classes. The objects of a class are indiscernible from each other. This theory can administrate imprecision, noisy and incomplete information. The objective of the rough set theory is to reduce the information system and, consequently, the generation of rules. However, the classical formalization of the rough sets theory does not address the problem of the missing value. The first approach considering rough sets to lead with missing values were the LEM1 and the LEM2 modified algorithms [29,31].

Li and Cercone [58] present the RSFit method, that predicts missing values, based on the rough sets theory and on the frequent itemsets.

Another alternative for dealing with missing data, inside the rough sets theory, is the decomposition method, in which neither the process nor the indiscernibility relation are modified. In the decomposition, the database is decomposed in subsets without missing values. Then, methods of classification through induction are applied on these subsets. Finally, conflict resolution methods are used for obtaining the final classification, from the partial classifications. The decomposition method introduces difficulties for the interpretation of the last step, which is conflicts resolution. Normally, a vote strategy is used to solve them. The improvement of the decomposition compared to other studies is to avoid the need to combine many classifiers [50]. In Hong et al. [36] a learning algorithm, based on rough sets, which can derive fuzzy rules from incomplete quantitative data sets and estimate the missing values in the learning process was proposed.

Partial reconstruction, proposed in [5], creates conceptual representations with reduced dimensions. This method uses a covariance matrix and principal components to obtain the concepts that hold the minimum possible information contained in the original variables. The covariance matrix can not be computed in a database containing missing values; so, the estimation of missing values, is based on the EM algorithm.

The issue of missing data must be carefully addressed, since that ignoring this problem can introduce bias into the models being evaluated and lead to inaccurate conclusions in knowledge discovery processes.

Besides the statistical approach, one notices, in the data mining field, a concern about missing data. Pearson [79,80] alerts that besides the missing data problem, there is the disguised data problem. According to the author, disguised data are the data that is coded as valid but it are not. The non valid data are explicitly represented but the value compromises the interpretation of the knowledge extracted. The main effect is the introduction of significant distortions in the analysis results. Hua and Pei [41] developed a method that is capable of identifying frequent missing data disguised. The method is based on distribution model of the missing data disguised and on the heuristics that uses the non-distorted samples. In the literature of data mining, there is concern about treatment of missing data in the data preparation phase, and about the performance of the algorithm. In [11] the impact of missing data on the main algorithms, of prediction, estimation, classification, and association rules, were evaluated.

Some studies present frameworks for the automatic choice of the method for missing data treatment. Zou [121] presents a framework based on the number of records, number of attributes, number of symbolic attributes, database entropy, number of classes, and missing data rate. Shalabi et al. [98] propose a framework that implements four techniques for treating missing data, and the choice of the technique to be used is based on the rough sets theory.

Generally speaking, in the literature of missing data, in the data mining context, few studies take the absence mechanism into consideration. The studies are basically divided into two categories: comparison of techniques and proposal of methods. Among them, [10,25,42,103] stand out. Among the comparative studies on techniques for dealing with missing data, [10,30,42,60,103,106,109] can be cited.

Acuña and Rodriguez [3] evaluated the classification error rate of the classifiers Linear Discriminant Analysis (LDA) and KNN against the methods for treatment of missing data: case deletion, average imputation, median imputation, and KNN.

Farhanglar et al. [20] compare the methods of average imputation, hot-deck, probabilistic algorithms, decision tree, and decision rules, considering the following characteristics of the database: number of records, number of attributes, proportion of Boolean attributes, and number of classes.

Delavallade and Dang [17] propose a new technique, based on the entropy measure, that finds a distribution value with more discrimination power for each missing value. Besides, they propose a new taxonomy for the methods, dividing them into: observation space or variable space, iterative or non-iterative, local information or global information, stochastic or deterministic, prediction model or class information.

A modification of the K-means algorithm, proposed in [11] incorporates information about the missing data. This information, called KSC (K-means with soft constraint) uses complete data and do not dispose the data with missing values. The complete data is used for clustering and the missing data for generating a set of restrictions for the clustering algorithm. A practical demonstration of the KSC algorithm is found in [112].

2.5. Considerations about the mechanisms of absence

The identification of absence mechanisms is an important task, because the properties of the statistical methods that deal with missing data strongly depend on the nature of such mechanisms [56]. No method can be considered optimum for all situations of missing data and, almost always, statistical methods make assumptions on the absence mechanism, which is not entirely correct.

The treatment of data with MAR or NMAR absence patterns considerably increases the complexity of the problem, since in such cases, the absence occurs because of a conjunction of factors, as far as one knows, unknown. So, unless you know the conditions under which the data became MAR or NMAR or treat them as MCAR is presented as a good first option [102].

Collins et al. [15] investigated ways to discover absence mechanism, through the incorporation of auxiliary variables. The strategies were called restrictive and inclusive. First, the variables are inserted to improve the performance of the imputation methods. The restrictive strategy incorporates few auxiliary variables, while the inclusive strategy uses all or almost all auxiliary variables available. Thus, it is possible to identify subcategories of the MAR mechanism, which are called MAR-linear, MAR-convex, and MAR-sinister. For example, suppose that X and Y are the observed variables, Y a variable with missing values, and Z an auxiliary variable that contains the possible cause of absence in Y . For the MCAR mechanism, the missing values are imposed to Y , independently of X , Y and Z . For the MAR-linear mechanism, the probability of absence is linearly related to Z . For the MAR-convex, the probability of absence is highest at the extremes of Z and lower at the middle. But the MAR-sinister, the probability of absence is a correlation function between Z and X . The authors conclude that the adoption of inclusive strategies help clear up the causes of absence. Moreover, they point out other plausible varieties of MAR.

Another study that investigates varieties of the MAR mechanism is presented in [82], which points out the related mechanisms MAR-Strong and MAR-Weak, respectively, as high and low correlation of variables.

In the real domain, a purely MAR or a purely NMAR absence is hard to be found, that is, MAR and NMAR can coexist in a database with missing values [27]. This observation leads to a new investigation of a fourth mechanism of absence denominated HMAR (hybrid missing at random). Suppose a database containing the sex and age attributes, where the missing values are associated to the age attribute. When the absence of age is related to sex attribute, then there is a uniquely MAR mechanism. When the absence of age is related to value of age itself, then there is a uniquely NMAR mechanism. However, in real situations, it is acceptable that the absence is related to a set of factors as, for example: female individuals over 30 years old do not inform their age; in this case, the HMAR hybrid mechanism occurs.

The NMAR mechanism is the most complex absence pattern. The NMAR mechanism is also called non-ignorable mechanism, because the conventional models of data analysis can not efficiently deal with this kind of absence. The existence of the NMAR mechanism is very difficult to be identified from the observed data. So, the literature recommends the exploration of the results, by using sensitivity analysis [44]. The sensitivity analysis is used for evaluating the degree of confidence of the results in situations of uncertain decisions or assumptions about the data. In general, applying sensitivity analysis is a good practice when employing different techniques of missing data treatment. Thus, one can make sure that the conclusions of each method are robust. Molenberghs et al. [67] observes that the sensitivity analysis for the specification of a model is a must, and subjective knowledge can play a critical role.

Due to the uncertainty that originates from the non-ignorable missing data (NMAR), great precaution must be taken when making any inference about a model or mechanism. In [14,88] demonstrated the distortion of the result when the NMAR data are treated as MAR data.

In turn, [68] demonstrated that empirical distinctions are not possible between MAR and NMAR. Through adjustments, an NMAR model can be exactly reproduced by a MAR mechanism.

Finally, another line of investigation corresponds for absence visualization as a means of helping identify the absence mechanism, as shown in [104,113].

3. Conclusion

In summary, there is no definitive conclusion about the better methods for missing data treatment. The method more adequate depends of the considered database, taking into account: the quantity of missing data and the absence mechanism or mechanisms that working on it. Opting for the removal of incomplete records (listwise deletion) can be acceptable when the quantity of missing data is small. In situations where the missing data is massive this can be disastrous. On the other hand, some authors consider that the procedures based on imputation are a “statistical alchemy” although it preserves statistical characteristics for inference, can not be useful to characterize groups of objects, due to modification of their “instances”. For a task of clustering of data, this can be a problem, because the object can be converted in an outlier.

The statistical literature recommends the multiple imputation (MI) methods and the maximum likelihood (ML). The great problem with these techniques is that they need strong assumptions of the model, which are difficult to justify, in the KDD context. On the other hand, the machine learning literature presents a variety of methods, among them the KNN, ANN, association rules, rough sets, decision trees and Bayesian classifiers. In the data mining context, there is not a coherent guide of methods to be used in databases containing missing values [50].

Considering that the identification of the absence mechanism is not a trivial task, and that methods of missing data treatment should be related to the absence mechanism, new methods that help identify the absence mechanism, for a correct application of the treatment methods available, are necessary. The biggest challenges are when the absence mechanism is NMAR, the most difficult one to be dealt with and also the most likely to occur.

This study presented a brief bibliographic review – in the statistics area and in the data mining area – of the main procedures, methods and techniques, used for the treatment of missing data. The aim of this work is alerting those responsible for projects in data mining of the importance of treatment of missing data and problems of the data imputation without a knowledge of the mechanisms of absence. It was possible to observe that many methods have been proposed for the treatment of missing data; but notices that the methods apply to certain absence mechanisms and, most times, they are applied from assumptions about the mechanism. The identification of the absence mechanism is not a trivial task. In literature there is not methods capable of identifying, precisely, all the mechanisms of absence. Methods of identification could be found only for the MCAR mechanism [51]. The lack of methods for the identification of absence mechanism serves as a challenge for the proposal of new research. Therefore, the proposal of methods for the identification of absence mechanisms corresponds to a latent contribution in the missing data area. The estimation or the imputation of the missing value and the identification of the absence mechanism can bring satisfactory results to the process of missing data recovery, because the identification of the mechanism can feed the estimation and the imputation processes, making it possible the utilization of the most appropriate technique for each mechanism. Thus, the process of recovery of data base containing missing data can be compared to a learning process, in which the identification mechanism is a procedure that perfects the estimation or the imputation process. For some authors, only

when the additional information of the missing data generation process is available it is possible to identify which method is the most adequate.

Another suggestion for a future study corresponds to the definition of the hybrid mechanism of absence. Graham [27] states that there can be many causes for the absence, and purely MAR cases or purely NMAR cases are only theoretical simplifications; the author considers the hypothesis of the existence of partially MAR and NMAR cases, defined as a hybrid mechanism.

Finally, a complementary study that deserves a biggest investigation is the visualization of missing data. Templ and Filzmoser [104,123] presented a visualization tool (R package) that allows the exploration of the missing data. According to the authors, the visualization of incomplete data allows to simultaneously explore the data and the structure of missing values. This is helpful for learning about the distribution of the incomplete information and identifies possible absence mechanisms and their relation to the available information (see Figs 1–4). For the authors, specialists in statistical which work with only one kind of database, they know the most relevant relationship and behavior in their data. For specialists in Data Mining, the database and applications domains are always different. Visualization tools may help to gain this knowledge in less time. In turn, [113] use SOM networks to visualize the absence patterns in the classification tasks. The visualization of absences can help identify the absence mechanism and, somehow, improve the performance of the identification methods of the mechanism.

Acknowledgments

The authors acknowledge the financial support received from the Foundation for Research Support of Minas Gerais State, FAPEMIG, through Project CEX PPM 107/12, and the National Council for Scientific and Technological Development, CNPq, Brazil.

References

- [1] M. Abdella and T. Marwala, Treatment of missing data using neural networks and genetic algorithms, in: *Proceedings of the IEEE International Joint Conference on Neural Networks* **1** (2005), 598–603.
- [2] A. Acocck, Working with missing values, *Journal of Marriage and Family* **67** (2005), 1012–1028.
- [3] E. Acuña and C. Rodriguez, The treatment of missing values and its effect in the classifier accuracy, in: *Classification, Clustering and Data Mining Application, Studies in Classification, Data Analysis, and Knowledge Organization*, D. Banks, L. House, F. McMorris, P. Arabic and W. Gauls, eds, Springer-Verlag **0**(7) 639–648.
- [4] A. Afifi and R. Elashoff, Missing observations in multivariate statistics I: Review of the literature, *Journal of the American Statistical Association* **61** (1966), 595–604.
- [5] C. Aggrwal and S. Parthasarathy, Mining massively incomplete data sets by conceptual reconstruction, in: *Proceedings of the Seventh ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (2001), 227–232.
- [6] F.G. Allan and J. Wishart, A method of estimating the yield of missing plot in field experiments, *J Agric Sci* **20** (1930), 399–406.
- [7] P.D. Allison, Multiple imputation for missing data: A cautionary tale, *Sociological Methods & Research* **28** (2000), 301–309.
- [8] T. Amemiya, Tobit models: A survey, *Journal of Econometrics* **24** (1984), 3–61.
- [9] G.E.A.P.A. Batista and M.C. Monard, A study of k-nearest neighbour as a model-based method to treat missing data, in: *Proceedings of the Argentine Symposium on Artificial Intelligence (ASAI'01)*, Buenos Aires, Argentina, (2001), 1–9.
- [10] G.E.A.P.A. Batista and M.C. Monard, An analysis of four missing data treatment methods for supervised learning, *Applied Artificial Intelligence* **17**(5) (2003), 519–533.
- [11] M.L. Brown and J.F. Kros, data mining and the impact of missing data, *Industrial Management & Data Systems* **103**(8) (2003), 611–621.
- [12] E.R. Buhi, P. Goodson and T.B. Neilands, Out of sight, not out of mind: Strategies for handling missing data, *American Journal Health Behaviour* **32** (2008), 83–92.

- [13] T. Calders, B. Goethals and M. Mampaey, Mining itemsets in the presence of missing values, in: *Proceedings of the 2007 ACM Symposium on Applied Computing* (2007), 404–408.
- [14] M.Y. Chang, Adjusting for nonignorable missing data with nonignorable sampling design, *ASA Section on Survey Research Methods* (2005), 2810–2814.
- [15] L.M. Collinns, J. Schafer and C. Kam, A comparison of inclusive and restrictive strategies in modern missing data procedures, *Psychological Methods* **6**(4) (2001), 330–351.
- [16] T.M. Cover and P.E. Hart, Nearest neighbor pattern classification, *IEEE Transactions on Information Theory* **13**(1) (1967), 21–27.
- [17] T. Delavallade and T.H. Dang, Using entropy to impute missing data in a classification task, *IEEE International Fuzzy Systems Conference* (2007), 1–6.
- [18] A.P. Dempster, N.M. Laird and D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B (Methodological)* **39**(1) (1967), 1–38.
- [19] A.R.T. Donders, G.J.M.G. Heijden and T. Stijnen, Review: A gentle introduction to imputation of missing values, *Journal of Clinical Epidemiology* **59** (2006), 1087–1091.
- [20] A. Farhangfar, L. Kurgan and W. Pedrycz, Experimental analysis of methods for imputation values in databases, in: *Intelligent Computing: Theory and Applications II Conference, in Conjunction with the SPIE Defense and Security Symposium*, Orlando, FL, (2004), 172–182.
- [21] A. Farhangfar, L. Kurgan and J. DY, Impact of imputation of missing values on classification error for discrete data, *Pattern Recognition* **41** (2008), 3692–3705.
- [22] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, From data mining to knowledge discovery in databases, *AI Magazine, American Association for Artificial Intelligence* (1996), 37–54.
- [23] I. Fortes, L. Mora-López and F. Triguero, Inductive learning models with missing values, *Mathematical and Computer Modelling* **44** (2006), 790–806.
- [24] N. Friedman, Learning belief networks in the presence of missing values and hidden variables, in: *Fourteenth International Conference on Machine Learning (ICML97)* (1997), 125–133.
- [25] A.J.T. Garcia and E.R. Hruschka, Naive bayes as an imputation tool for classification problems, in: *Proceedings of the Fifth International Conference on Hybrid Intelligent Systems (HIS'05)* (2005), 497–499.
- [26] P.J. Garcia-Laencina, J.L. Sancho-Gómez and A.R. Figueira-Vidal, Pattern classification with missing values using multitask learning, in: *Proceedings of the IEEE International Joint Conference on Neural Networks* (2006), 3594–3601.
- [27] J.W. Graham, Course session #14 Missing data analysis and design, <http://ies.ed.gov/ncer/whatsnew/conferences/rct-traininginstitute/presentations.asp>, 2007, Accessed January 30, 2009.
- [28] V. Gorodetsky, O. Karsaev and V. Samoilov, Direct mining of rules from data with missing values, in: *Studies in Computational Intelligence*, T.Y. Lin, S. Ohsuga, C.J. Liao, X.T. Hu and S. Tsumoto, eds, *Foundation of Data Mining and Knowledge Discovery* (Vol. 6), Springer, 1996, pp. 233–264.
- [29] J.W. Grzymala-Busse, Rough set strategies to data with missing attribute values, in: *Proceedings of the Workshop on Foundations and New Directions in Data Mining, associated with the third IEEE International Conference on Data Mining*, Melbourne, FL, USA, (19–22 November 2003), 56–63.
- [30] J.W. Grzymala-Busse and M. Hu, A comparison of several approaches to missing attribute values in data mining, in: *Proceedings 2nd International Conference on Rough Sets and Current Trends in Computing RSTC*, (2000), 340–347.
- [31] J.W. Grzymala-Busse and A.Y. Wang, Modified algorithms LEM1 and LEM2 for rule induction from data with missing attribute values, in: *Proceedings 5th International Workshop on Rough Sets and Soft Computing RSSC'97*, at *The Third Joint Conference on Information Sciences (JCIS'97)*, 1997, 69–72.
- [32] H.O. Hartley and R. Hocking, The analysis of incomplete data, *Biometrics* **27** (1971), 783–808.
- [33] J. Heckman, The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models, *Annals of Economic and Social Measurements* **5** (1976), 475–492.
- [34] R. Hewett, Decision making using incomplete data, in: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (2004), 182–187.
- [35] A. Holst, The use of a bayesian neural network model for classification tasks, PhD thesis, *Dept of Numerical Analysis and Computing Science, Royal Institute of Technology*, Stockholm, Sweden, Sept. 2004, TRITA-NA-P9708.
- [36] T.P. Hong, L.H. Tseng and B.C. Chien, Mining from incomplete quantitative data by fuzzy rough sets, *Expert Systems With Applications* **37**(3) (2010), 2644–2653.
- [37] E.R. Hruschka, Imputação bayesiana no contexto da mineração de dados, Doctoral dissertation, COPPE/FEDERAL UNIVERSITY OF RIO DE JANEIRO, UFRJ, Rio de Janeiro, RJ, Brasil, 2003.
- [38] E.R. Hruschka, E.R. Hruschka, Jr. and N.F.F. Ebecken, A nearest-neighbor method as a data preparation tool for a clustering genetic algorithm, in: *Proceedings of the 18^o Simpósio Brasileiro de Banco de Dados (SBBBD)* Manaus, AM, (2003), 319–327.
- [39] E.R. Hruschka, E.R. Hruschka, Jr. and N.F.F. Ebecken, Evaluating a nearest-neighbor method to substitute continuous missing values, in: *The 16th Australian Joint Conference on Artificial Intelligence – AI 03, Perth Lecture Notes in*

- Artificial Intelligence (LNAI 2903)* Heidelberg, Springer-Verlag, **2903** (2003), 723–734.
- [40] M. Hu, S.M. Salvucci and M.P. Cohen, Evaluation of some popular imputation algorithms, in: *ASA Section on Survey Research Methods* (1998), 308–313.
 - [41] M. Hua and J. Pei, Cleaning disguised missing data: A heuristic approach, in: *ACM KDD 07 Industrial and Government Track Paper*, ACM, (2007), 950–958.
 - [42] J. Hulse, J.T.M. Khoshgoftaar and C. Seiffert, A comparison of software fault imputation procedures, in: *Proceedings of the 5th International Conference on Machine Learning and Applications (ICMLA'06)* (2006), 135–142.
 - [43] S. Ingrassia and F. Domma, Statistical methods and perspectives in neural network training from incomplete data, in: *Proceedings Atti XL Riunione Scientifica della Società Italiana di Statistica* (2000), 353–364.
 - [44] I. Jansen, N. Hens, G. Molenberghs, M. Aerts, G. Verbeke and M.G. Kenward, The nature of sensitivity in monotone missing not at random models, *Computational Statistics & Data Analysis* **50** (2006), 830–858.
 - [45] E.T. Jaynes, Information theory and statistical mechanics, *Physical Review* **106**(4) (1957).
 - [46] K. Jiang, H. Chen and S. Yuan, Classification for incomplete data using classifier ensembles, in: *Proceedings of the IEEE International Conference on Neural Networks and Brain* **1** (2005), 559–563.
 - [47] P. Jonsson and C. Wohlin, An evaluation of k-nearest neighbour imputation using likert data, in: *Proceedings of the 10th IEEE International Symposium on Software Metrics (METRICS'04)* Chicago, USA, (2004), 108–118.
 - [48] H. Junninen, H. Niska, K. Tuppurainen, J. Ruuskanen and M. Kolehmainen, Methods for imputation of missing values in air quality datasets, *Journal of Atmospheric Environment* **38** (2004), 2895–2907.
 - [49] N.M. Laird, Discussion of informative drop-out in longitudinal data analysis by P.J. Diggle and M.G. Kenward, *Applied Statistics* **43**(84) (1994).
 - [50] R. Latkowski and M. Mikolajczyk, Data decomposition and decision rule joining for classification of data with missing values, in: *Rough Sets and Current Trends in Computing, RSCTC'2004* Springer-Verlag Berlin Heidelberg, (2004), 299–320.
 - [51] R.J.A. Little, A test of missing completely at random for multivariate data with missing values, *Journal of the American Statistical Association* **83** (1998), 1198–1202.
 - [52] R.J.A. Little, Regression with missing X's: A review, *Journal of the American Statistical Association* **87** (1992), 1227–1237.
 - [53] R.J.A. Little, Pattern-mixture for multivariate incomplete data, *Journal of the American Statistical Association* **88**(421) (1993), 125–134.
 - [54] R.J.A. Little, Biostatistical analysis with missing data, *Encyclopedia of Biostatistics*, P. Armitage and T. Colton, eds, Wiley: London, 1997.
 - [55] R.J.A. Little and D.B. Rubin, Incomplete data, *Encyclopedia of the Statistical Sciences* **4** (1983), 46–53.
 - [56] R.J.A. Little and D.B. Rubin, Statistical analysis with missing data, *Series in Probability and Statistics*, Wiley, 2nd Edition, 2002.
 - [57] R.J.A. Little and N. Schenker, Missing data, in: *Handbook for Statistical Modeling in the Social and Behavioral Sciences*, G. Arminger, C.C. Clogg and M.E. Sobel, eds, Plenum, New York, 1994, pp. 39–75.
 - [58] J. Li and N. Cercone, Predicting missing attribute values based on frequent itemset and RSFit, Technical Report CS-2006-13, School of Computer Science, University of Waterloo, 2006.
 - [59] P. Liu and L. Lei, Missing data treatment methods and NBI model, in: *Proceedings on the Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06)* (2006), 633–638.
 - [60] P. Liu, L. Lei and N. Wu, A quantitative study of the effect of missing data in classifiers, in: *Proceedings of the 2005 Fifth International Conference on Computer and Information Technology, CIT'05*, (2005), 28–33.
 - [61] J. Luengo, S. García and F. Herrera, A Study on the use of imputation methods for experimentation with radial basis function network classifiers handling missing attribute values: The good synergy between RBFs and EventCovering method, *Neural Networks* **23** (2010), 406–418.
 - [62] 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* (2011), To appear.
 - [63] S.M. Lynch, Missing Data, http://www.princeton.edu/~slynch/SOC_504/missingdata.pdf, 2003, Accessed July 30, 2007.
 - [64] I.B. Magalhães, Avaliação de redes Bayesianas para imputação em variáveis qualitativas e quantitativas, doctoral dissertation, São Paulo University-USP, São Paulo, SP, Brasil, 2007.
 - [65] M. Magnani, Techniques for dealing with missing data in knowledge discovery tasks, <http://magnanim.web.cs.unibo.it/index.html>, 2004, Accessed July 30, 2007.
 - [66] M. Mampaey, Association rule mining met missing values, Master Thesis, University of Antwerp, Belgium, 2006.
 - [67] G. Molenberghs, E. Goetghebeur, S.R. Lipsitz and M.G. Kenward, Non-random missingness in Categorical data: strengths and limitations, *The American Statistician* **53**(2) (May 1999), 110–118.
 - [68] G. Molenberghs, C. Beunckens, C. Sotito and M.G. Kenward, Every missing not at random model has got a missing at random bodyguard, Technical Report 06103, IAP Statistics Network, Interuniversity Attraction Pole, Institut de

- statistique, Université catholique de Louvain, Belgium, 2002.
- [69] W.R. Myers, Handling missing data in clinical trials: An overview, *Drug Information Journal* **34** (2000), 525–533.
 - [70] I. Myrtveit, E. Stensrud and U.H. Olsson, Analyzing data sets with missing data: An empirical evaluation of imputation methods and likelihood-based methods, *IEEE Transactions on Software Engineering* **27**(11) (Nov 2001), 999–1013.
 - [71] J.R. Nayak and D.J. Cook, Approximate association rule mining, in: *Proceedings of the Florida Artificial Intelligence Research Symposium* (2001), 259–263.
 - [72] F.V. Nelwamondo and T. Marwala, Fuzzy artmap and neural network approach to online processing of inputs with missing values, *SAIEE Africa Research Journal* **98**(2) (June 2007), 1–7.
 - [73] F.V. Nelwamondo, S. Mohamed and T. Marwala, Missing data: A comparison of neural network and expectation maximisation techniques, *Current Science* **93**(11) (2007), 1514–1521.
 - [74] N. Nie, C. Hull, J. Jenkins, K. Steinbrenner and B. Bent, SPSS, 2nd Edition McGraw-Hill, New York, 1975.
 - [75] J. Ning and P.E. Cheng, A comparison study of nonparametric imputation methods, *Statistics and Computing*, (2011), 1–13.
 - [76] T. Orchard and M.A. Woodbury, A missing information principle: Theory and applications, in: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* **1** (1972), 697–715.
 - [77] R. Orre, A. Bate, G.N. Norén, E. Swahn, S. Arnborg and I.R. Edwards, A bayesian recurrent neural network for unsupervised pattern recognition in large incomplete data sets, *International Journal of Neural Systems* **15**(3) (June 2005), 207–222.
 - [78] Z. Pawlak, Rough sets, *Intelligent Journal Computer and Information Science* **11** (1982), 341–356.
 - [79] R. Pearson, The problem of disguised missing data, *SIGKDD Explorations* **8**(1) (1982), 83–92.
 - [80] R. Pearson, *Mining Imperfect Data: Dealing with Contamination and Incomplete Records*, SIAM Press, Philadelphia, Pennsylvania, 2005.
 - [81] H. Peng and S. Zhu, Handling of incomplete data sets using ICA and SOM in data mining, *Neural Comput & Applic* **16** (2007), 167–172.
 - [82] J.S. Preisser, K.K. Lohman and P.J. Rathouz, Performed of weighted estimating equations for longitudinal binary data with drop-outs missing at random, *Statistics in Medicine* **21** (2002), 3035–3054.
 - [83] D. Pyle, *Data Preparation for Data Mining*, Morgan Kaufmann Publishers, São Francisco, CA, 1999.
 - [84] J.R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Series in Machine Learning, São Francisco, CA, 1993.
 - [85] A. Ragel and B. Crémilleux, Treatment of missing values for association rules, in: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Melbourne, Australia, (1998), 258–270.
 - [86] A. Ragel and B. Crémilleux, MVC-A preprocessing method to deal with missing values, *Knowledge-Based Systems* **12**(5–6) (1999), 285–291.
 - [87] C. Riggelsen and A. Feelders, Learning bayesian network models from incomplete data using importance sampling, in: *Proceedings Tenth International Workshop on Artificial Intelligence and Statistics*, (2005), 301–308.
 - [88] P.R. Rosenbaum and D.B. Rubin, The central role of the propensity score in observational studies for causal effects, *Biometrika* **70** (1983), 41–56.
 - [89] D.B. Rubin, Inference and missing data, *Biometrika* **63**(3) (1976), 581–592.
 - [90] D.B. Rubin, An overview of multiple imputation, *ASA, Section on Survey Research Methods* (1988), 79–84.
 - [91] D.B. Rubin, Multiple imputation after 18+ years, *Journal American Statistical Association* **91** (1996), 473–489.
 - [92] D.B. Rubin, Conceptual, computational and inferential benefits of the missing data perspective in applied and theoretical statistical problem, *Allgemeines Statistisches Archiv* **90**(4) (2006), 501–513.
 - [93] J.L. Schafer and M.K. Olsen, Multiple imputation for multivariate missing-data problems: A data analyst's perspective, *Multivariate Behavioral Research* **33** (1998), 545–571.
 - [94] J.L. Schafer and J.W. Graham, Missing data: Our view of the state of the art, *Psychological Methods* **7**(2) (2002), 147–177.
 - [95] J. Scheffer, Dealing with missing data, *Research Letters in the Information and Mathematical Sciences* **3** (2002), 153–160.
 - [96] H. Schoner, Working with real-world datasets: Preprocessing and prediction with large incomplete and heterogeneous datasets, Doctoral dissertation, Berlin University of Technology, Berlin, 2004.
 - [97] W. Shadish, Finding a solution for missing Data, <http://www.apa.org/monitor/julaug02/missingdata.html>, 2002, Accessed July 30, 2008.
 - [98] L.A. Shalabi, M. Najjar and A. Kayed, A framework to deal with missing data in data sets, *Journal of Computer Science* **2** (2006), 740–745.
 - [99] C.E. Shannon, A mathematical theory of communication, *Bell Syst Tech* **27** (1948), 379–423.
 - [100] E.L. Silva-Ramírez, R. Pino-Mejías, M. López-Coello and M.D. Cubiles-De-La-Vega, Missing value imputation on missing completely at random data using multilayer perceptrons, *Neural Networks* **24**(1) (2011), 121–129.
 - [101] P.J.J. Soares, Análise bayesiana de dados deficientemente categorizados, Doctoral dissertation, Universidade Técnica

- de Lisboa, Lisboa, Portugal, 2004.
- [102] J.A. Soares, Pré-processamento em mineração de dados: Um estudo comparativo em complementação, Doctoral dissertation, Federal University of Rio de Janeiro, COPPE, Rio de Janeiro, RJ, Brasil, 2007.
 - [103] Q. Song and M. Shepperd, A new imputation method for small software project data sets, *Journal of System and Software* **80** (2007), 51–62.
 - [104] M. Templ and P. Filzmoser, Visualization of missing values using the R-package VIM, Forschungsberit CS-2008-1. Institut f. Statistik u. Wahrscheinlichkeitstheorie, Wien, Austria, <http://www.statistik.tuwien.ac.at>, 2008.
 - [105] H. Thijs, G. Molenberghs, B. Michiels, G. Verbeke and D. Curran, Strategies to fit pattern-mixture models, *Biostatistics* **3** (2002), 245–265.
 - [106] M.S. Tsechansky and F. Provost, Handling missing values when applying classification models, *Journal of Machine Learning Research* **8** (2007), 1625–1657.
 - [107] N. Tsikriktsis, A review of techniques for treating missing data in OM survey research, *Journal of Operations Management* **24** (2005), 53–62.
 - [108] B. Twala, M. Cartwright and M. Shepperd, Comparison of various methods for handling incomplete data in software engineering databases, in: *Proceedings of the 2005 International Symposium on Empirical Software Engineering* (2005), 105–114.
 - [109] B. Twala, M. Cartwright and M. Shepperd, Ensemble of missing data techniques to improve software prediction accuracy, in: *Proceedings of the 28th International Conference on Software Engineering ACM*, Shanghai, China, (2006), 909–912.
 - [110] Zs.J. Viharos, L. Monostori and T. Vincze, Training and application of artificial neural networks with incomplete data, *Lecture Notes of Artificial Intelligence, LNAI 2358, Proceedings of the Fifteenth International Conference on Industrial & Engineering Application of Artificial Intelligence & Expert Systems* Springer Computer Science, (2002), 649–659.
 - [111] K.L. Wagstaff, Clustering with missing values: No imputation required, in: *Proceedings of the Meeting of the International Federation of Classification Societies* Springer, (2004), 649–658.
 - [112] K.L. Wagstaff and V.G. Laidler, Making the most of missing values: Object clustering with partial data in astronomy, *Astronomical Data Analysis Software and System, XIV ASP Conference Series*, (2005), 172.
 - [113] H. Wang and S. Wang, Visualization of the critical patterns of missing values in classification tasks, *Lectures Notes in Computer Science* **4781** (2007), 267–274.
 - [114] W. Wei and Y. Tang, A generic neural network approach for filling missing data in data mining, in: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* **1** (2003), 862–867.
 - [115] S. Weiss and N. Indurkha, Decision-rule solutions for data mining with missing values, Tech Report RC-21783, IBM, 1999.
 - [116] D. Williams, X. Liao, Y. Xue and L. Carin, On classification with incomplete data, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(3) (2007), 427–436.
 - [117] C. Wu, C. Wun and H. Chou, Using association rules for completing missing data, in: *Proceedings of the Fourth IEEE International Conference on Hybrid Intelligent System, (HIS'04)* (2004), 236–241.
 - [118] A.K. Wong and D.K.Y. Chiu, Synthesizing statistical knowledge from incomplete mixed-mode data, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **9** (1987), 796–805.
 - [119] L. Yu, S. Wang and K. Lai, An integrated data preparation scheme for neural network data analysis, *IEEE Transactions on Knowledge and Data Engineering* **18**(2) (2006), 217–230.
 - [120] L.E. Zárate, B.M. Nogueira and T.R.A. Santos, Comparison of classifiers efficiency on missing values recovering: Application in a marketing database with massive missing data, in: *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining, IEEE CIDM Hawaii*, (2007), 66–72.
 - [121] Y. Zou, A. An and X. Huang, Evaluation and automatic selection of methods for handling missing data, in: *Proceedings of the IEEE International Conference in Granular Computing* (2005), 728–733.
 - [122] A. Revenko, O.S. Kuznetsov and B. Ganter, Finding errors in new object intents, in: *Proc CLA 2012*, L. Szathmary and U. Priss, eds, Universidad de Málaga Spain, 2012, pp. 151–162.
 - [123] M. Templ, A. Alfons and P. Filzmoser, Exploring incomplete data using visualization techniques, *Advances in Data Analysis and Classification* **6**(1) (2012), 29–47.