Using Game Theory to Handle Missing Data at Prediction Time of ID3 and C4.5 Algorithms

Halima Elaidi[1], Zahra Benabbou[2], Hassan Abbar[3]
Laboratory of Information and Decision Support Systems
Hassan 1st University Settat, Morocco

Abstract—**The raw material of our paper is a well-known and commonly used type of supervised algorithms: decision trees. Using a training data, they provide some useful rules to classify new data sets. But a data set with missing values is always the bane of a data scientist. Even though decision tree algorithms such as ID3 and C4.5 (the two algorithms with which we are working in this paper) represent some of the simplest pattern classification algorithms that can be applied in many domains, but with the drawback of missing data the task becomes harder because they may have to deal with unknown values in two major steps: at training step and at prediction step. This paper is involved in the processing step of databases using trees already constructed to classify the objects of these data sets. It comes with the idea to overcome the disturbance of missing values using the most famous and the central concept of the game theory approach which is the Nash equilibrium.**

Keywords—*Decision tree; ID3; C4.5; missing data; game theory; Nash equilibrium*

I. INTRODUCTION

Machine learning is a discipline where knowledge is created automatically from raw data. Several algorithms have been developed for this purpose. This knowledge is then exploited to make decisions. Naturally, good decisions are made when data is of a good quality. Even though decision trees have proved that they are efficient classification tools, they remain, just like any other machine learning technique, helpless in front of missing data. This paper proposes to employ the concept of Nash equilibrium which is a fundamental concept of the theory of non-cooperative games with perfect information, to put an end to the disturbance caused by missing data. We only consider trees constructed by the use of algorithms ID3 or C4.5, and we suppose that these trees are perfectly constructed, the reason why our proposed method intervenes in the step of utilizing the resulting decision rules (trees) to classify new data sets containing observations with missing values.

When data is missing, it does not mean that we are allowed to ignore the corresponding records or observations. Because if we do ignore them, we are immediately causing a partial loss of information about the population we are studying through this data set. On contrary, we should treat them very accurately and try to find some useful techniques we can use to deal with missing values in a given data set. As a result researchers have developed several methods to handle this problem [1], such as:

- Deleting the records with missing values.

- Allocating the missing value of an attribute by its amount if it is quantitative or by the most frequent value if it is qualitative.

- Looking for the maximum likelihood between the records, etc.

The imputation technique this paper proposes is based on a mathematical approach which can be considered as one of the most fundamental and important discoveries of the last century: game theory.

Furthermore, the technique we are proposing can be considered as an improvement of both algorithms ID3 and C4.5 at the same time. Because, the calculations based on the Nash equilibrium that we will present later on our paper might be added as instructions or steps to the algorithm structure. As a result, for a given training data, the algorithms ID3 and C4.5 with their new structures permit to produce trees that are able to deal with data sets containing attributes with some missing values, thing which makes it possible to classify their records without any problem and with no need to look for a method among those that already exist to handle or impute the missing data.

Our document will be organized as follows: we will start by presenting the theory of decision trees and its algorithms we are interested in for this research. Then, we will introduce in details the problem to which we are proposing a solution in this paper. Next, Section 4 will be about the game theory and Nash equilibrium concepts. Section 5 will present in details the proposed imputation method, which is at the same time a way to improve the performance of the algorithms discussed in Section 2. And finally we will conclude with a brief and concise discussion.

II. THEORY OF DECISION TREES

Decision trees constitute simple tools for decision making [2]. Actually, they are used in various fields. Their form of graphical tree representation makes of them a very simple tool, but also a very powerful one. Decision trees are the result of a set of algorithms which identify different ways of dividing a database into branches called segments, these segments form a tree characterized by a root node at the top of it.

In the same paper, Quinlan claimed "Decision-makers need to make predictions...One sound basis for such predictions is an extrapolation of past, known cases" this population of known cases is called, in the field of machine learning, the training data. It represents the principal raw material of a

decision tree. In fact, a decision tree describes how to divide a population into homogeneous groups depending on the discriminant variables, since each node is just a choice on an attribute. The method used to separate the training data differs from one algorithm to another. However they all aim at making the best separation possible at each node of the tree by testing the "goodness of split" of each attribute [3].

Decision tree learning is a powerful tool and one of the most widely used and practical methods in the domain of machine learning. It is one of the supervised methods whose idea consists of classifying objects according to their characteristics or attributes and then the way these classes are formed should be used so that the resulting decision tree learn how to classify the elements of every treated data-set [4]. Several decision tree learning algorithms have been developed, but in this paper we are going to be interested in the two famous algorithms of Quinlan ID3 and C4.5. Their process of construction is based on the concept of gain (Profit or benefit): ID3 uses "Information gain" as its attribute selection measure, while the C4.5 algorithm which is the successor of ID3 uses the "gain ratio" as its attribute selection measure. Let's briefly present an overview of each of the two algorithms.

A. The ID3 Algorithm

ID3 is the well-known decision tree algorithm [5]. It is based on a recursive top-down approach; Giving a training data in which each observation is described in terms of a set of attributes, the ID3 algorithm uses the information gain as an attribute selection measure in order to separate recursively that set of examples. The information gain is calculated using the entropy:

$$E(S) = -\sum_{i=1}^{k} p_i log_2(p_i) \tag{1}$$

Where;

- **E** is the entropy function.

- **S** the set of examples that can be divided into classes $C_1, C_2, \ldots, C_k$

- **p_i** is the probability that a set of objects from S belongs to class C_i

The resulting entropy value for a treated attribute gives an idea about its randomness or uncertainty, in a way that the attribute with the smallest entropy value is the best to use in data separation. Contrariwise, the more information gain value is important, the more the tested attribute is gainful for the separation. This property of information gain to vary in the opposite direction of variation of entropy is explained by its formula:

$$Gain(A, S) = E(S) - \sum_{j=1}^{n} \frac{|S_j|}{|S|} E(S_j) \tag{2}$$

- **A** is the treated attribute.

- **n** the number of possible values of attribute A.

- **S_j** are the subsets of S containing objects with the same value of attribute A.

Even though the ID3 algorithm works well in some cases, it remains powerless with attributes having a significant number of values, continuous data and missing values. These limitations of ID3 were the reason why J.Ross Quinlan developed C4.5.

B. The C4.5 Algorithm

The C4.5 decision tree algorithm [6] was developed in order to overcome the limitations of ID3 mentioned previously. Just like the ID3 algorithm, C4.5 has as a starting point a given training data, but this time the measure used to split the data is the Gain Ratio which is none other than a normalized information gain. Its formula is written as follows:

$$GainRatio(A, S) = \frac{Gain(A,S)}{SplitInfo(A,S)} \tag{3}$$

where;

$$SplitInfo(A, S) = -\sum_{j=1}^{n} \frac{|S_j|}{|S|} \log_2 \left(\frac{|S_j|}{|S|} \right) \tag{4}$$

The present gain formula (Gain Ratio) intervenes to put an end to the weakness of information gain in front of attributes with a large number of values, because the information gain used as a splitting measure by the ID3 algorithm favors attributes with a significant number of values. Furthermore, C4.5 is said to be more efficient than ID3 in view of the fact that it is able to overcome the problem of features with continuous values as well as missing data, which is not the case for the ID3 algorithm. Another advantage of C4.5 over ID3 is that it can produce pruned decision trees. Pruning technique aims at reducing the size of a tree that over-fits the training data, which allows decreasing the prediction error rate [7].

III. PROBLEM TO BE SOLVED

Our paper assumes that a tree is perfectly constructed using a given training set and one of the algorithms we discussed previously (ID3 and C4.5). Since decision trees are developed for the purpose of making decisions and classifying data, the produced tree can be used to classify the elements of any data set structured in the same way of the training data i.e. a data set where the objects are described using the same attributes of the training data. The set of rules established after constructing this tree may be useless if the object we are trying to classify has one or several attributes with missing values.

This work comes to remedy that problem of unknown data by using the game theory approach (more precisely, the Nash equilibrium technique). In order to fully understand the way game theory works for helping an ID3 or C4.5 algorithm to overcome the missing data problem, we propose a simple example. Thus we will consider the same example (playing tennis) already introduced by the paper "Induction of decision trees" [4]. It is an ID3 algorithm example.

The training set of our example is presented by table1

By using this data set and applying the ID3 algorithm steps on it, we obtain the decision tree of figure 1:

"N" to indicate the decision "Not play" and "P" to indicate the decision "Play".

TABLE I.   A TRAINING SET EXAMPLE

| Day	Outlook	Temperature	Humidity	Windy	Class
1	sunny	hot	high	false	N
2	sunny	hot	high	true	N
3	overcast	hot	high	false	P
4	rain	mild	high	false	P
5	rain	cool	normal	false	P
6	rain	cool	normal	true	N
7	overcast	cool	normal	true	P
8	sunny	mild	high	false	N
9	sunny	cool	normal	false	P
10	rain	mild	normal	false	P
11	sunny	mild	normal	true	P
12	overcast	mild	high	true	P
13	overcast	hot	normal	false	P
14	rain	mild	high	true	N
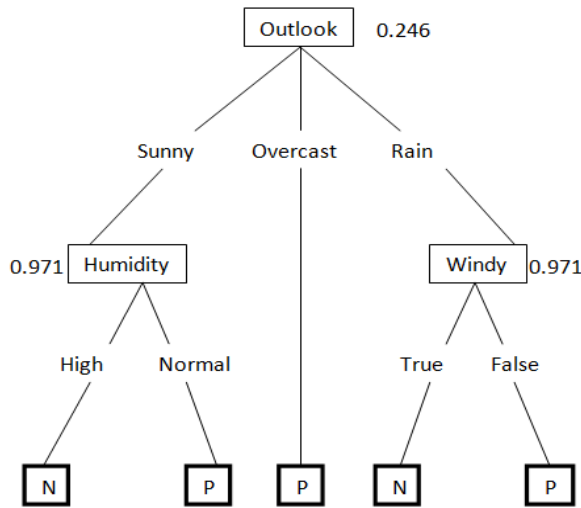
Fig. 1.   The Resulting Tree using the ID3 Algorithm.

The classification rules of our example appear clearly on the established tree, we can then easily classify new observations described and defined by the use of the same set of attributes. But, suppose that while processing a data set, our classifier has encountered an observation like that of the 25th day appearing on table 2

It is absolutely clear that the decision rules of our classifier are helpless in such a case. That is why in the following parts of our paper, we will present the discipline (game theory) that will help us adjust this limitation.

TABLE II.   EXAMPLE OF A RECORD WITH MISSING VALUES

Day	Outlook	Temperature	Humidity	Windy
..	..	..	..	..
25	?	mild	?	false
..	..	..	..	..

IV. GAME THEORY AND NASH EQUILIBRIUM APPROACHES

Despite of the fact that the fundamentals of game theory began to emerge earlier, this mathematical approach became more famous as a discipline only after publishing the book "Theory of Games and Economic Behavior" by J.V.Neumann and O.Morgenstern in 1944 [8]. And in spite of the presence of the word "Game" in the theory appellation, game theory remains very useful and helpful in plenty of domains which are crucial and of a major importance such as biology, economics and business, political science, engineering, computer science...and many others. The purpose of this paper is to find a solution for a problem often encountered in one of machine learning branches (decision trees).

In the field of game theory, the player is the principle element. His definition is extremely large: he can be an individual, a firm, a political party...In general, he is the decision-maker, conscious of his choices and their results, looking forward to ensure a gainful position in the game, and aware of the fact that his decisions and actions depend on those of other players i.e. he is supposed to be rational.

Thus, the main idea of game theory is to model the behavior of a set of players by observing and analyzing their strategic interactions. Usually, a game is defined by three elements: the set of players, the set of strategies (a strategy of each player is the set of his decisions) and utilities (the preference indicators of each decision) [9].

A. Mathematical Representation of a Game

First of all, note that our proposed game is a non-cooperative one. Because as we will discover later on this paper, the players of our game do not make any agreements that can bind them.

In the field of game theory, a normal form game is defined as follows:

$$G = (N, (S_i)_{i \in N}, (U_i)_{i \in N}) \qquad (5)$$

where;

- N is the set of players (Card $(N) = n$).

- S_i is the strategies set of player i, namely the set of decisions the player i can make.

- U_i: $S_1 \times S_2 \times ... \times S_n \rightarrow \mathbb{R}$ is the utility function of each player i (i=1,2, ..., n) called also the payoff function.

Concerning the game of this paper: The set of players is defined as the attributes with missing values of a given observation (or object). The strategy of each of the players is given by the set of values of each attribute. We will suppose that the utility or payoff of a player when making a decision, corresponds to the value of information gain (or Gain Ratio; depending on the algorithm constructing the decision tree) realized by the node that comes immediately after the branch representing the taken decision. But in this case, the utility value remains the same whatever the decisions made by the rest of players, which makes of the player a non-rational one. Because, the payoff of a rational player participating in a non-cooperative game as presented with (5) should depend not only

on his own strategy, but also on the decisions of other players. Thus, for a player (attribute) making a decision, his utility is equal to the amount of information gain (Gain Ratio) provided by this decision (which is always the value of information gain mentioned on the node that comes immediately after the branch representing the decision) multiplied by a proportion that we determine with the help of the training data. Indeed, for all observations on training data that have the same decisive value for the attribute (player) in question, we must look for the proportion of records that respect each of the decisions made by the rest of the players. For instance, considering the same example presented by the tree of figure 1, assume that "outlook" and "humidity" are the attributes players of the game as shown on table 2: the utility of the attribute "outlook" when "sunny" is its decision and "normal" is the decision of "humidity" is equal to 0.971*(2/5) = 0.388. We can then conclude that the utility of a player i is calculated using the following formula:

$$u_i(s_i, s_{-i}) = G_{s_i} \frac{K_{s_i, s_{-i}}}{K_{s_i}} \tag{6}$$

Where;

- s_i is the decision of player i.

- s_{-i} is representing the strategic profile of the rest of players.

- $K_{s_i, s_{-i}}$ is the number of records from the training data whose attributes players have the profile of strategies (s_i, s_{-i}).

- K_{s_i} is the number of records from the training data where player i plays his strategy si.

Intuitively, G_{s_i} is equal to 1 when the successor node of the branch representing the decision si of the player i is a leaf, because the information is fully provided. On the other hand, if an attribute does not appear at least in one of the classification rules of the established tree, G_{s_i} will be equal to 0 for all the possible values of this attribute, i.e. the attribute does not provide any information.

B. Nash Equilibrium

The notion of Nash equilibrium can be considered as the most brilliant as well as influential game theoretical concept that was invented by the "beautiful mind" John Nash. It is defined as a stable situation, where each player (from the set of players in interaction) is not ready to deviate of his decision. Because if he does, while the rest of players are keeping their strategies, his utility will immediately decrease, thing which is not gainful for a rational decision-maker (or simply player) [10].

1) Pure nash equilibrium: The normal form of a game as it is previously presented (5) is considered in the field of game theory as a game with pure strategies. The Nash equilibrium (the pure strategy Nash equilibrium) for a set of players participating in such a game is given by a profile of strategies $(s_1^*, s_2^*, \dots, s_n^*)$ where each s_i^* is representing the best decision made as a response to other players' strategies [9, 10]. Mathematically, this can be written as follows:

For all players i = 1, 2, ..., n

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*) ; \forall s_i \in S_i \tag{7}$$

where

$$s_{-i} = (s_1, s_2, \dots, s_{i-1}, s_{i+1}, \dots, s_n)$$

2) Mixed Nash equilibrium: The concept of mixed strategy can be adjudged to be the generalization of pure strategy, which comes to provide a much clearer vision about the real behavior of a rational player. A mixed strategy is simply a pure strategy associated to a distribution of probability with which each player is making his choice.

So, for a set of pure strategies Si, let $\Delta(S_i)$ denote the set of probability distributions over it, so that:

$$P_i \in \Delta(S_i) \Longrightarrow \begin{cases} P_i : S_i \rightarrow [0, 1] \\ s_i \rightarrow P_i(s_i) \\ \sum_{s_i \in S_i} P_i(s_i) = 1 \end{cases} \tag{8}$$

Thus Pi which is a probability distribution over Si represents the mixed strategy of player i.

Therefore, in a case where the decisions of players are in the form of mixed strategies, the mixed Nash equilibrium is defined as a profile of mixed strategies$(P_1^*, P_2^*, \dots, P_n^*)$, such as:

For all players i=1, 2,..., n

$$u_i^{mixed}(P_i^*, P_{-i}^*) \geq u_i^{mixed}(P_i, P_{-i}^*); \forall P_i \in \Delta(S_i) \tag{9}$$

Where; u_i^{mixed} is the expected utility function defined as follows:

$$u_i^{mixed}(P_i) = \sum_{s_i \in S_i} \left(\prod_{j \in N} P_j(s_j) \right) u_i(s_i) \tag{10}$$

V. PROPOSED METHOD

As it is already mentioned at the beginning of the paper, our work assumes that a decision tree is constructed using one of the famous algorithms of Quinlan (ID3 or C4.5). It also supposes that the stage of construction did not face any obstacle. As a matter of fact, the present work comes to fix a problem which appears during the use of the established tree for classifying new data, more precisely data with missing values. Our method assumes that, for a given observation, the imputation of its missing values is a strategic game of the form (5). According to the theorem of existence of the Nash equilibrium [10], the proposed game accepts at least a Nash equilibrium in mixed strategies, because the number of players is finite as well as the number of strategies for each player i.

Since the construction of decision trees using ID3 and C4.5 algorithms is a task which is mainly based on the theory of information, we then had the idea of using the same theory to impute the missing values of a data set at prediction time. In fact, Quinlan's splitting measures were based on the information theory of shannon in order to find how well each attribute by its own classifies the records of the training set, then the attribute with the highest value of information gain or gain Ratio is the one that can generate the best partition i.e. that attribute is providing the best quantity of information. Hence the inspiration of making from this concept of information gains a useful tool to handle missingness at prediction process.

The idea is to impute missing values of each observation by values maximizing the quantity of information, respecting of course the structure and the characteristics of the training data; to that end, we are suggesting the use of the game theory approach every time when encountering a data entry where at least two attributes' values are missing: each of those attributes is represented as a player and the strategy set of each one consists of the possible values in the range of the attribute. The payoffs correspond to that quantity of information we cited previously. As a matter of fact, the Nash equilibria would yield "balanced" ways of substituting the missingness: all values used for imputation had the same objective which is "maximization of information gain".

Using the example presented in section 3 (table 1), let's assume that the established decision rules (tree) (figure 1) are used to classify the elements of a given database containing an observation where the values of attributes "outlook" and "humidity" are missing (table 2). We can look for the Nash equilibrium in pure strategies for this game with two players: outlook and humidity. Notwithstanding this equilibrium point does not exist all the time, but if it does exist, the values of the corresponding attributes represent the ones that should be used to impute the missing values. If a Nash equilibrium in pure strategies does not exist, then we proceed to a Nash equilibrium in mixed strategies which always exists, depending on the theorem of Nah we discussed above.

It is to highlight that the job becomes difficult when the problem of missingness concerns continuous attributes (This difficulty is encountered only while working with the C4.5 algorithm). But as it is widely known, C4.5 converts continuous values to nominal ones by proposing to perform binary splits based on a threshold value. As a matter of fact two intervals should be obtained: [minimum value, threshold] and]threshold, maximum value]. Then we propose using the centers of those intervals to impute missing values.

A. Imputation by the use of Pure Nash Equilibrium

Always in the case of our explanatory example, the payoff matrix of the game is given by table 3:

According to the rule of determining the pure Nash equilibrium (7), we can deduce that this game admits two equilibrium profiles which are (high, sunny) and (normal, rain) ("humidity" is the first player and "outlook" is the second one). As a result, for the observation of table 2, attributes outlook and humidity can take respectively the values (high and sunny) or (normal and rain).

TABLE III.   THE PAYOFF MATRIX OF THE GAME

		Outlook		
		sunny	overcast	rain
Humidity	normal	(0.286, 0.388)	(0.286, 0.5)	(0.429, 0.583)
	high	(0.429, 0.583)	(0.286, 0.5)	(0.286, 0.388)

TABLE IV.   THE PAYOFF MATRIX OF THE NEW GAME

		Player 2 (Attribute 2)	
		C	D
Player 1 (Attribute 1)	A	(0.7 , 0.97)	(0.88 , 0.1)
	B	(0.81 , 0.5)	(0.6 , 0.92)

B. Imputation by the use of Mixed Nash equilibrium

In this subsection, we will treat another case in which we are confronting a situation where the game does not accept any Nash equilibrium in pure strategies, the reason why we are forced to look for Nash equilibrium in mixed strategies. Giving an example remains the best way to explain a technique. Thus, for a given training data (different from the one we worked with previously), assume that by using once again the ID3 (or C4.5) algorithm, we got a new tree and certainly different values of the players' utilities. Then suppose that while working with the obtained tree for the purpose of classifying new data, we have encountered an observation with two attributes whose values are missing. Therefore, the first thing to do is to construct the payoff matrix, which is given by table 4:

It is quite clear that the game in question does not admit a Nash equilibrium in pure strategies, thereby we will look for the mixed equilibrium. In fact, note α as the probability with which player 1 plays the strategy "A" and (1 – α) the probability with which he plays strategy "B". Similarly, player 2 plays strategy "C" with probability β and strategy "D" with probability (1 – β).

According to these probability values, the expected utility of player 1 is written as follows:

- $0.7\,\beta + 0.88\,(1 - \beta) = 0.88 - 0.18\,\beta$; if player 1 chooses to play strategy A.

- $0.81\,\beta + 0.6\,(1 – \beta) = 0.6 + 0.21\,\beta$; if player 1 chooses to play strategy B.

Accordingly:

$$u_1(P_1, P_2) = \alpha\,(0.88 – 0.18\,\beta) + (1 – \alpha)\,(0.6 + 0.21\,\beta)$$

$$= (0.6 + 0.21\,\beta) + \alpha\,(0.28 – 0.39\,\beta)$$

Which is a function increasing in α if $(0.28 – 0.39\,\beta) > 0$ and decreasing if $(0.28 – 0.39\,\beta) < 0$.

Consequently, the strategy A constitute the best response of player 1 in mixed strategies if and only if β < 0.72, while B is the best response of player 1 in mixed strategies if and only if β > 0.72, but when β=0.72 he still indifferent between the two strategies.

Similarly, the expected utility of player 2 is written as follows:

- $0.97\,\alpha + 0.5\,(1 - \alpha) = 0.5 + 0.47\,\alpha$; if player 2 chooses to play strategy C.

- $0.1\,\alpha + 0.92\,(1 - \alpha) = 0.92 - 0.82\alpha$; if player 2 chooses to play strategy D.

Thus;

$$u_2(P_1, P) = \beta\,(0.5 + 0.47\,\alpha) + (1 – \beta)\,(0.92 – 0.82\,\alpha)$$

$$= (0.92 – 0.82\,\alpha) + \beta\,(1.29\,\alpha – 0.42)$$

Which is a function increasing in β if $(1.29\,\alpha - 0.42) > 0$ and decreasing if $(1.29\,\alpha - 0.42) < 0$. Therefore, C is the best response of player 2 in mixed strategies if and only if α > 0.33 and D is the best response of player 2 in mixed strategies if and

only if α < 0.33. But in the case where α = 0.33 the player 2 is indifferent between the two strategies C and D.

The equilibrium of a game in mixed strategies is established when the players are indifferent in their choices of strategies. Concerning our example, the equilibrium is presented as a profile of probabilities: [(0.33 , 0.67) ; (0.72 , 0.28)] where player 1 chooses strategy A with a probability of 0.33 and strategy B with a probability of 0.67, and similarly player 2 chooses 72% of the time the strategy C and 28% of the time the strategy D.

VI. Discussion and Conclusion

Just like any other machine learning algorithm, techniques used for classification tasks are all the time facing the problem of missing data. In fact, in real data applications, the presence of missing data is a general and challenging problem [11]. A decision tree classifier may encounter this problem in two contexts: values may be missing in the training data (at induction time) or while predicting the classes to which new records should belong (at prediction time) [12]. Concerning the method we are proposing, it aims at handling missing values at prediction time and it only concerns two types of algorithms constructing decision trees: ID3 and C4.5.

Specialists in the field of data missingness insist on the necessity of making assumptions about what caused the data to be unknown. Thereby, they identify three categories or types of missing data:

- **MCAR**: (Missing Completely At Random) refers to data that were collected randomly which means that for an observation where a feature's value is missing, that missingness does not depend on any variable of the data set.

- **MAR**: (Missing At Random) requires that the cause of missingness is not related to the unknown feature while it could be conditional on some of the rest of variables in the data set.

- **NMAR**: (Not Missing At Random) this case takes place when the missingness is not random and depending on the actual value of the missing data.

"When data are MCAR or MAR, the missing data mechanism is termed ignorable" [13], the approach we are proposing does not require any prior knowledge about the reasons of data missingness i.e. we are assuming data to be MCAR or MAR. Classification approaches and methods have proved their usefulness in many problem domains. However, they have to deal with the problem of missing data which is a common drawback when solving a real life classification task. A wide range of techniques were elaborated to handle the limitations caused by unknown values such as:

- Case deletion

- Mean imputation

- Multiple imputation

- Hot and cold deck imputation

- Maximum likelihood

As their being classifiers, ID3 and C4.5 can use such approaches to face unknown values while processing a data set for prediction. Of course each one of these methods has its advantages and disadvantages. But logically, each classification technique has some specific characteristics related to the way the classifier should be constructed, from this point we thought that the approach adopted by every classifier to handle the limitations of missing data should respect the characteristics and essential elements of the classifier. This work concerns two well-known classification algorithms (ID3 and C4.5) that are based on the notion of "quantity of information gained", for that reason a method respecting this notion should be used when handling data missingness. Thereby, our proposed method consists of maximizing the gain of information while imputing unknown values in treated observation: our method covers all the features of only one record at the same time. Thus, working on a data set for classification purpose, we propose to handle the unknown values as a first step by the use of our method based on game theory approach (as seen on section 5), then the classifier can be applied to determine the class of each record.

Note that we are proposing a method under two forms: imputation using pure Nash equilibrium and imputation using mixed Nash equilibrium. The first one seems to be easier, but the second one is the most relevant as it always gives results and it is more realistic. In fact, records having exactly the same features whose values are unknown will use the same form of game to impute missing data. And as it was already mentioned, the solution in mixed strategies comes in the form of probabilities. For instance, assume that an attribute is part of a game and that attribute can take two possible values A and B, at the Nash equilibrium in mixed strategies, A can be the best decision with a probability of P% and B can be the best decision with a probability of (1 - P)%. We can then conclude that for all records utilizing the same game, A will be assigned to P% of these observations and B will be assigned to (1 - P)% of that observations (N.B for an observation, when only one variable is missing, we assign to it the value with the maximum payoff).

In general, features are considered the fundamental elements for constructing classifiers such as ID3 and C4.5 and they still unchanged while processing data sets. Assume that we are working on a data set with N variables, the number of possible games that can be used to deal with missing data is $\sum_{k=1}^{N-1} C_N^k$; it is a finite number of cases. Consequently, it is a step which can be added to both of algorithms as an improvement of algorithms ID3 and C4.5.

References

[1] J.L. Schafer, J.W. Graham, "Missing Data: Our View of the State of the Art" Psychological methods 2002, Vol. 7, N° 2, American Psychological Association, PP. 147-177.

[2] J.R. Quinlan, "Decision trees and decision-making" IEEE Trans. Syst. Man Cybern. 1990 Vol. 20, N° 2, PP 339-346, DOI: 10.1109/21.52545.

[3] L. Breiman, J. Friedman, R. Olshen, C. Stone, "Classification and regression trees" Pacific Grove, CA: Wadsworth 1984.

[4] J.R. Quinlan, "Induction of Decision Trees" Machine Learning 1986, 1, 81-106.

[5] J.R. Quinlan, "Discovering rules by induction from large collections of examples". In D. Michie (Ed.), Expert systems in the micro electronic age 1979, Edinburgh University Press.

[6] J.R. Quinlan, "Comparing connectionist and symbolic learning methods". In S. Hanson, G. Drastal, And R. Rivest (Eds.), Computational Learning Theory and Natural Learning Systems: Constraints and Prospects 1993, Cambridge, MA: MIT Press.

[7] H. Elaidi, Z. Benabbou, H. Abbar, "A comparative study of algorithms constructing decision trees: ID3 and C4.5" LOPAL'18 Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications 2018, DOI: 10.1145/3230905.3230916.

[8] J. von Neumann, O. Morgenstern, "Theory of Games and Economic Behavior" Princeton University Press, Princeton, New Jersey, 1944.

[9] M.J. Osborne, "An Introduction to Game Theory" Oxford University Press, Oxford, 2002.

[10] J. Nash, "Non-cooperative games" The Annals of Mathematics, Second Series 1951, Vol. 54, N° 2, PP. 286-295.

[11] Q. Song, M. Shepperd, "Missing Data Imputation Techniques" Int. J. Business Intelligence and Data Mining 2007, Vol. 2, N° 3, 261-291.

[12] M. SAAR-TSECHANSKY, F. PROVOST, "Handling Missing Values when Applying Classification Models" Journal of Machine Learning Research 2007, 8, 1625-1657.

[13] P.J. Garcia-Laencina, J.L. Sancho-Gomez, A.R. Figueiras-Vidal, "Pattern classification with missing data: a review" Neural Comput and Applic 2010, 19, 263–282, DOI: 10.1007/s00521-009-0295- 6.