An Adaptable System to Construct Fuzzy Decision Trees

Christophe Marsala

LIP6, mailbox 169

Université P. et M. Curie

4 place Jussieu

75252 Paris cedex 05, France

Christophe.Marsala@lip6.fr

Bernadette Bouchon-Meunier

LIP6, mailbox 169

Université P. et M. Curie

4 place Jussieu

75252 Paris cedex 05, France

Bernadette.Bouchon-Meunier@lip6.fr

Abstract

Nowadays, data mining is an active domain that is linked to data management and machine learning techniques. However, even if inductive learning methods work well when handling symbolic attributes, problems arise when considering numerical or numerical-symbolic (num/symb) attributes. This problem can be solved by introducing tools from the fuzzy set theory to handle such kind of data. In this paper, we present our adaptable system to construct and to use fuzzy decision trees by means of several kinds of operators.

1. Introduction

Data mining is a new domain highly active but it is a difficult domain to define. This domain was introduced at the beginning of the 1990s. The first definition that appeared concerned knowledge discovery: "Knowledge Discovery in Databases (KDD) is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns from data." [6].

At this time, KDD was highly connected to databases, as mentionned: "... database mining as the confluence of machine learning techniques and the performances emphasis of database technology (...), a number of database mining problems can be uniformly viewed as requiring discovery of rules embedded in massive data." [1].

New researchers from the machine learning domain begun to be interested in this topic. Important works were presented by Fayyad related to the use of a learning algorithm (construction of decision trees) to mine

a big amount of data. The definition of KDD evolved slightly: "..., KDD refers to the overall process of discovering useful knowledge from data, and Data Mining (DM) refers to a particular step in this process. DM is the application of specific algorithms for extracting patterns from data." [5].

Finally, the DM process was linked to the kind of data to be mined, in the overall process of KDD and always: "DM is defined from the kind of knowledge to mine." [4].

Moreover, DM and KDD are highly connected to commercial applications, that leads to a new kind of definition of this topic: "DM is the exploration and analysis of a big amount of data in order to find forms and significant rules with automatic or semi-automatic means... The *aim* of DM is to enable a plant to enhance its functions through a better comprehension of its clients." [2].

So, we can summarize these definitions, saying that the aim of knowledge discovery is to discover knowledge from a set of data to enrich its self knowledge. Data mining is the learning step of this process.

Inductive learning raises the particular to the general. This process is well-adapted to associate a computer with learning abilities and it can be used as a data mining process.

More particularly, a set of classes C is considered, representing a physical or a conceptual phenomenon. This phenomenon is described by means of a set of attributes $\mathcal{A} = \{A_1, ..., A_N\}$. Each attribute A_j can take a value v_{jl} in a set $\{v_{j1}, ..., v_{jm_j}\}$ of possible values. A *description* is a N-tuple of attribute-value pairs (A_j, v_{jl}). Each description is associated with a particular class c_k from a set $C = \{c_1, ..., c_K\}$ to make up an *instance* (or *example*, or *case*) e_i of the phenomenon.

Inductive learning is a process to generalize from a *training set* $\mathcal{E} = \{e_1, ..., e_n\}$ of examples to a general law to bring out relations between descriptions and
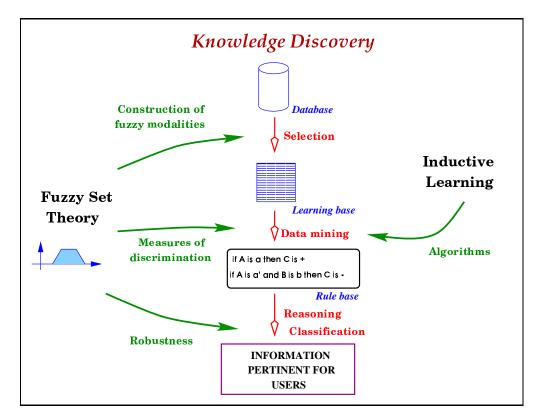
Figure 1. Processus of Knowledge Discovery

classes in C.

Inductive learning methods work well when handling *symbolic* attributes. However, problems arise when considering *numerical* or *numerical-symbolic* (num/-symb) attributes. Values of a num/symb attribute can be either numerical or symbolic. Intrinsically, a num/-symb modality is imprecise by nature. Such a modality must not be excluded from the learning process because it is already more general than a precise value. Thus, it represents an information to take advantage of and to take into account.

Fuzzy set theory (FST) defines tools useful to handle num/symb data. It is an efficient way to take into account imprecision in the description of a case, produced by its measurement for instance, and to set up an interface between numerical values and symbolic values. Moreover, the FST enables an inductive algorithm to handle missing data that can exist in a training set.

To take into account num/symb modalities, classical decision trees have been enhanced by the use of the FST into *fuzzy decision trees* [9]. In a fuzzy decision tree, nodes are associated with num/symb attributes, edges are labeled by fuzzy modalities and leaves can be associated with a fuzzy subset of the class. As studied in [11], methods to construct fuzzy decision trees differ

in their choice of the measure of discrimination.

This paper is composed as follows. In Section 2, we recall algorithms to construct fuzzy decision trees. In Section 3, we present our model to study discrimination measures. In Section 4, we introduce the adaptable system *Salammbô* we have constructed [3, 9] that enables us to construct a fuzzy decision tree by means of several kinds of measures of discrimination.

2. Fuzzy Decision Trees

Methods to construct decision trees in presence of num/symb attributes are less common and have been developed for few years. The FST has been introduced to handle numerical attributes when constructing decision trees. Such methods define a generalization of decision trees: the *fuzzy decision trees* (FDT) [8, 3, 16, 18, 19].

Two kinds of methods to construct fuzzy decision trees can be brought out [11]. The first one deals with methods based on a generalized Shannon entropy measure, the *star-entropy measure*, as a measure of discrimination. The second one deals with methods based on another family of fuzzy measures.

In presence of num/symb data, the commonly used

measure of discrimination is a measure of *entropy of fuzzy events*, also called *star-entropy* [16]. It corresponds to the Shannon entropy extended to fuzzy events by substituting probabilities of fuzzy events to classical probabilities.

$$H_S^*(C|A_j) =$$
$$-\sum_{l=1}^{m_j} P^*(V_{jl}) \sum_{k=1}^{K} P^*(C_k|V_{jl}) \log(P^*(C_k|V_{jl}))$$

This measure is obtained from the classical Shannon measure of entropy by substituting the Zadeh's probability measure P^* of fuzzy events by the classical probability [20].

The SAFI system [16] was one of the first methods of construction of FDT in the domain of inductive learning in presence of num/symb data. In this software, the star-entropy is used as a measure of discrimination. Fuzzy modalities on the universe of values of numerical attributes are considered as provided by experts of the given domain.

In the same way, several systems have been introduced to construct FDT by means of the star-entropy, for instance Weber's [18] or Janikow's [8] systems. All these systems construct FDT in presence of given fuzzy modalities defined on the universe of values of numerical attributes. Moreover, in his system, Janikow proposed to optimize the fuzzy partitions by means of genetic algorithms [8].

Likewise, the measure of discrimination used in the system described by [17] is a particular case of the star-entropy. Here, it is an example of system where the fuzzy modalities are determined by means of an automatic algorithm. The Umano's method was also improved by [7] which proposed a method to build fuzzy modalities from training data. Their method to compute fuzzy partitions is based on the use of the theory of evidence.

The star-entropy is one of the measures of discrimination implemented in the system *Salammbô* we have constructed [3, 9].

All the methods we have mentioned in this section are equivalent in the sense that they use the star-entropy as a measure of discrimination although they sometimes use a different method to build or to optimize fuzzy partitions. Moreover, they are different with respect to the method they implement to use a FDT when classifying new cases. This last point is not developed in this paper but it is detailed in [11]. Measures of discrimination must be defined differently in order to enable the construction of fuzzy decision trees from a set of examples described by means of numeric-symbolic attributes.

Adapted measures have to take into account the knowledge of membership degrees to a fuzzy value for each example of the training set.

3. Measures of Discrimination

In order to find an attribute to split the training set[1], each attribute of $\mathcal{A}$ is tested and its discriminating power with regard to the classes is valued. The attribute with the highest discriminating power is selected to construct a node in the tree.

We need to measure the suitability of a set of examples induced by a modality v_{jl} of an attribute A_j with a set of examples induced by a class c_k. In [12], a hierarchical $\mathcal{FGH}$-model is proposed to validate and to construct discrimination measures (Figure 2). This model enables us to evaluate the pertinence of a function used to measure the discriminating power of an attribute in an inductive process. This pertinence is evaluated with regard to the properties required for such a fonction that enable it to measure the adequation of the set of examples related to the class.
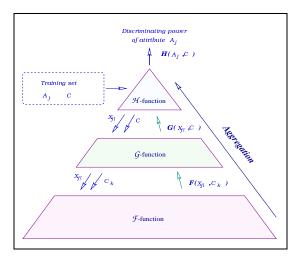

Figure 2. Hierarchical model of functions

Other classical measures based on a measure of probability can be extended in the same way, to handle numeric-symbolic values. The difficulty in this case lies in the validation of such a measure to sort the attributes on their discriminating power. We have presented in [9, 12] a framework based on properties required from measures of discrimination in the context of decision tree construction. This framework admits the already presented measures of discrimination as particular cases and enables us to define new measures of

---

[1]For more details on an algorithm to construct a decision tree, see [3, 9].

discrimination well-suited with the process of construction.

4. The Salammbô system

The adaptable Salammbô system we have constructed is a computer software to build fuzzy decision trees [3, 9] and it enables us to test several kinds of parameters (Figure 3). Several measures of discrimination are implemented (Shannon measure of entropy, Gini test of impurity, Entropy star measure) in order to test the corresponding results for each parameters.

This software enables us to construct a FDT by means of measures chosen in a general family we have introduced on the basis of interesting properties required for the discrimination process. Moreover, an automatic method to build a fuzzy partition on the set of values of a numerical attribute is introduced [10] and it enables us to avoid the prior definition of fuzzy modalities of attributes by an expert.

Various parameters (t-norms, t-conorms, implications) can be used in the *Salammbô* software and we test them in the process of classification on different kind of databases. A t-norm is chosen as the AND operator for the premises, an implication enables us to deduce a conclusion associated with a set of premises, and a t-conorm is chosen as the OR operator for the aggregation of several conclusions.

5. Validation

Tests have been conducted that highlighted the interest of fuzzy decision trees for such learning, their expressiveness and their qualities when handling continuous attributes.

For instance, in chemistry, associations between the structures of chemical compounds and the quality of their odors have been brought out by means of the construction of FDT with the software *Salammbô* [14]. Applications were conducted to extract knowledge from other kinds of domains: from a geographical oriented-object spatial database [13], and from an electrical domain database [9].

Here, we present some results of two construction methods, obtained on two common training databases with a cross validation test. These databases are available on the ftp site of the University of Irvine, California (ftp://ftp.ics.uci.edu/pub/machine-learning-databases).

The first comparison concerns the iris database. In this database, examples are described by means of 4 numeric attributes and there are 3 classes to recognize. The second comparison concerns the database of

Base	Method	Size	Classification rate
Iris	Classic	8.5	95.2%
	Salammbô	4.0	96.0%
Waveform	Classic	44.6	72.7%
	Salammbô	66.9	78.2%

Table 1. Results of execution

Breiman's waveforms. In this database, examples are described by means of 21 numeric attributes and there are 4 classes to recognize.

In Table 1, results with the *Classic* method concern the classical ID3 method and are those given by [15] for the C4.5 algorithm, adapted to numeric attributes. In this case, the decision trees are also pruned. Results with the *Salammbô* method concern the construction of fuzzy decision trees with the entropy star measure as measure of discrimination. The *Size* is the average number of paths of the built trees, and the *Classification rate* is the number of test examples that are well classified by means of the built tree.

It can be observed that the fuzzy decision trees provide better classification rate than the classical decision trees. In the case of the iris database, the size of fuzzy decision trees, even with no pruning, is highly smaller than the size of the classical decision tree. In the case of the waveform database, the size of the tree should be minimized by means of a pruning phase.

6. Conclusion

In this paper, a presentation of fuzzy decision trees is done. After a review of some algorithms to construct such trees, we introduce a model to study measures of discrimination used in such methods. Afterwards, we present the system Salammbô to construct fuzzy decision trees by means of several kind of discrimination measures.

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Database mining: a performance perspective. *IEEE Transactions on Knowledge and Data Engineering*, 5(6), 1993.

[2] M. Berry and G. Linof. *Data Mining: Techniques appliquées au marketing, à la vente et aux services clients*. Interéditions, 1997.

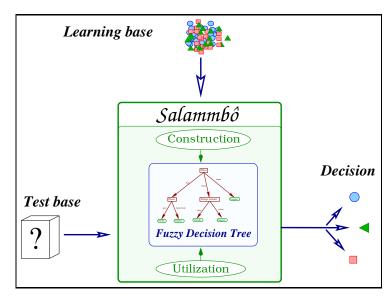[3] B. Bouchon-Meunier, C. Marsala, and M. Ramdani. Learning from imperfect data. In D. Dubois,

Figure 3. The Salammbô system

H. Prade, and R. R. Yager, editors, *Fuzzy Information Engineering: a Guided Tour of Applications*, pages 139–148. John Wileys and Sons, 1997.

[4] M. Chen, J. Han, and P. Yu. Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6):866–883, 1996.

[5] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17(3):37–54, Fall 1996.

[6] W. Frawley, G. Piatetsky-Shapiro, and C. Matheus. Knowledge discovery in databases: An overview. In *In Knowledge Discovery in Databases*, pages 1–27. AAAI/MIT Press, Cambridge, 1991.

[7] H. Ichihashi, T. Shirai, K. Nagasaka, and T. Miyoshi. Neuro-fuzzy ID3: a method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method for incremental learning. *Fuzzy Sets and Systems*, 81(1):157–167, July 1996.

[8] C. Z. Janikow. Fuzzy decision trees: Issues and methods. *IEEE Transactions on Systems, Man and Cybernetics*, 28(1):1–14, 1998.

[9] C. Marsala. *Apprentissage inductif en présence de données imprécises : construction et utilisation d'arbres de décision flous.* Thèse de doctorat,

Université Pierre et Marie Curie, Paris, France, Janvier 1998. Rapport LIP6 nº 1998/014.

[10] C. Marsala and B. Bouchon-Meunier. Fuzzy partioning using mathematical morphology in a learning scheme. In *Proceedings of the 5th IEEE Int. Conf. on Fuzzy Systems*, volume 2, pages 1512–1517, New Orleans, USA, September 1996.

[11] C. Marsala and B. Bouchon-Meunier. Construction methods of fuzzy decision trees. In *Proceedings of the JCIS'98 Conference*, volume 4, pages 17–20, October 1998.

[12] C. Marsala, B. Bouchon-Meunier, and A. Ramer. Hierarchical model for discrimination measures. In *Proc. of the IFSA'99 World Congress*, pages 339–343, Taiwan, 1999.

[13] C. Marsala and N. Martini Bigolin. Spatial data mining with fuzzy decision trees. In N. F. F. Ebecken, editor, *Data Mining*, pages 235–248. WIT Press, 1998. Proceedings of the International Conference on Data Mining, Rio de Janeiro, Sept. 1998.

[14] C. Marsala, M. Ramdani, M. Toullabi, and D. Zakaria. Fuzzy decision trees applied to the recognition of odors. In *Proceedings of the IPMU'98 Conference*, volume 1, pages 532–539, Paris, July 1998. Editions EDK.

[15] J. R. Quinlan. Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*, 4:77–90, 3 1996.

[16] M. Ramdani. Une approche floue pour traiter les valeurs numériques en apprentissage. In *Journées Francophones d'apprentissage et d'explication des connaissances*, 1992.

[17] M. Umano, H. Okamoto, I. Hatono, H. Tamura, F. Kawachi, S. Umedzu, and J. Kinoshita. Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems. In *Proceedings of the 3rd IEEE Conference on Fuzzy Systems*, volume 3, pages 2113–2118, Orlando, june 1994.

[18] R. Weber. Fuzzy-ID3: A class of methods for automatic knowledge acquisition. In *IIZUKA'92 Proceedings of the 2nd International Conference on Fuzzy Logic*, pages 265–268, 1992.

[19] Y. Yuan and M. Shaw. Induction of fuzzy decision trees. *Fuzzy Sets and systems*, 69:125–139, 1995.

[20] L. Zadeh. Probability measures of fuzzy events. *Journal Math. Anal. Applic.*, 23, 1968. reprinted in "*Fuzzy Sets and Applications: selected papers by L. A. Zadeh*", R. R. Yager, S. Ovchinnikov, R. M. Tong and H. T. Nguyen eds, pp. 45–51.