HAL open science

Forests of Fuzzy Decision Trees

Christophe Marsala, Bernadette Bouchon-Meunier

▶ To cite this version:

Christophe Marsala, Bernadette Bouchon-Meunier. Forests of Fuzzy Decision Trees. IFSA'97 - Seventh International Fuzzy Systems Association World Congress, Jun 1997, Prague, Czech Republic. hal-02182889

HAL Id: hal-02182889
https://hal.sorbonne-universite.fr/hal-02182889v1

Submitted on 13 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FORESTS OF FUZZY DECISION TREES

Christophe Marsala, Bernadette Bouchon-Meunier

LIP6, *Université Pierre et Marie Curie, Case 169, 4 place Jussieu, 75252 Paris cedex 05, France.*
e-mail: {marsala, bouchon}@laforia.ibp.fr

Abstract. In inductive learning, to build decision trees is often arduous when there exists more than two classes to learn. In this paper, a method of decomposition of problems with more than two classes into several problems with only two classes is proposed. This decomposition enables the construction of a *forest of fuzzy decision trees* where each fuzzy decision tree is dedicated to the recognition of a single class against a combination of all the other classes. The construction of fuzzy decision trees is based on an extension of the ID3 algorithm which handles imprecision in data. A method to use such a forest of fuzzy decision trees to classify new cases is also proposed.

Keywords: fuzzy decision tree, inductive learning, data fusion, evidence theory

1 Introduction

Inductive learning of a phenomenon from a given domain is based on a set of *examples*. Each example is a case already solved on the given phenomenon. It is associated with a pair [description, class] where the *description* is a set of pairs [attribute, value] which is the available knowledge on the given phenomenon. That is, the knowledge which could be obtained by the observation of the phenomenon or by measures. The *class* of the example is the decision (or category, or solution...) associated with the given description. Such a set of examples is called a *training set*. The aim of inductive learning, or induction, is to find general rules enabling us to classify any description of the phenomenon, *i. e.* to generalize the knowledge obtained by the observation of some occurrences of the phenomenon to any future observations.

The inductive method used in this paper is based on the construction of a decision tree from a training set. An advantage of such kind of method lies in the easiness of its implementation. Moreover, a decision tree is a tool easily understandable and easily explainable: a decision associated to a description by a tree is easy to justify to an expert of the domain (which could not also be an expert in computer science).

The simplest decision trees correspond to symbolic trees, built from training sets where all the attributes take their values (or modalities) in a finite set. However, other kinds of attributes can occur as numeric attributes which take their values in a continuous set, or numeric-symbolic attributes, the values of which are either numeric precise (*6'2"*) or numeric imprecise (*about 6 feet*) or symbolic (*tall*). To handle this particular kind of attributes, new methods must be introduced, either to construct decision trees or to use them in a generalization process.

Usual methods to take into account numeric attributes discretize the universe of values of the attributes. Some thresholds are computed and used as test values in nodes of the decision trees [5]. Tools from fuzzy set theory are useful to handle the imprecision upon such thresholds: they enable to smooth the boundaries induced by these thresholds. For instance, decision trees are built with these thresholds considered as fuzzy values during the generalization phase [2].

When considering that the symbolic values of a numeric-symbolic attribute are fuzzy modalities on the numeric universe of its values, particular methods from fuzzy set theory to treat these values enable to take into account this kind of attributes. These methods enable to build fuzzy decision trees [17], [20], [18], [10], [22], [2], [13], [1], [3]. During the generalization phase, the use of fuzzy decision trees to classify new examples enables to associate more than one class to them, each class weighted by a membership degree.

Other problems to build decision trees occur when the examples are associated with more than two classes. In this paper, an association of classes is proposed, before the learning process. The training set is split into several subsets, each one induced by a class to recognize. Here, the process of learning from a training set is decomposed into several learning processes on training subsets where a class has to be learnt against all the other classes. In such a system, called a *forest of decision trees*, each further classified example will be associated with a membership degree to classes, given by each fuzzy decision tree of the forest. Then, these degrees have to be aggregated in order to determine a single membership degree to each class.

In the first section of this paper, a method to build and to use fuzzy decision trees is recalled. In the second section, the proposed method to build a forest of fuzzy decision trees, when more than two classes must be learnt, is presented. In the third section, results with such a method are reported.

Finally, this paper ends with a conclusion on this method and some perspectives are presented.

2 Fuzzy decision trees

A fuzzy decision tree is a particular case of questionnaire [14]. It is composed by three kinds of elements: nodes, edges and leaves. A path is composed of edges and ends in a leaf. A node is associated with a question on the values of an attribute and each edge going out of a node is associated with a particular value (or modality) of this attribute. A leaf, which is a terminal node, is labelled with the modality of the class associated with the path from the root to this leaf.

To build a questionnaire is equivalent to choose an order on the questions to ask on the values of the attributes in order to determine a class. Usually, a question related to an attribute is selected by means of a measure of discrimination from the set of all possible questions [14]. An example of such a measure is Shannon's measure of entropy. In Artificial Intelligence, this algorithm is called the ID3 algorithm [15].

The ID3 algorithm is a top down algorithm, the training set is split by means of a question on an attribute. The question is chosen with the Shannon entropy and labels a node of the tree. Each edge from this node is labeled by a value of the chosen attribute. The modalities of the chosen attribute split the training set into subsets on which the process is done again, until all the examples of the training set pertain to a single class.

Methods to build decision trees from imprecise knowledge can be compared with methods to take into account numeric attributes [5], [16]. But specific techniques must be added to take into account imprecision and fuzziness of the knowledge. Such techniques enable the construction of a particular kind of decision trees: the fuzzy decision trees.

Several methods exist to build fuzzy decision trees, most of them are based on the ID3 algorithm and use particular techniques to take into account the imprecision in the knowledge [17], [20], [18], [10], [22], [1]. Differences between these methods lie essentially in the choice of a new measure of discrimination to use during the construction of a fuzzy decision tree and in the discretization method to construct the modalities associated with edges. This measure should take into account the discriminating power of an attribute and, also, fuzzy modalities for the numeric-symbolic attributes. Moreover, the use of such fuzzy decision trees to classify examples is based on an extension of the classic method of utilization of decision tree. With a fuzzy decision tree, the attribute values in the description to classify do not necessarily label an edge from a node associated with a question related to the corresponding attribute. Thus, this new kind of data has to be taken into account.

The application *Salammbô* enables the construction of a fuzzy decision tree by means of an algorithm based on SAFI algorithm [18]. Moreover, from a training set composed of attributes with numeric values, Salammbô generates a fuzzy partition on the universe of values of such attributes [12]. A fuzzy decision tree which handles numeric values and fuzzy modalities is built by means of the ID3 algorithm where the Shannon entropy is replaced by a fuzzy entropy taking into account the new kind of information upon the data and extends Shannon entropy to fuzzy subsets. The fuzzy entropy of the decision C, with modalities c_j, related to the attribute A, with modalities v_i, is called the *entropy star* and is defined as [18]:

$$E_A^* = -\sum_i P^*(v_i) \sum_j P^*(c_j|v_i)\ \log P^*(c_j|v_i)$$

where $P^*(v_i) = \sum_{1\leq l\leq n} f_{v_i}(x_l)P(x_l)$ is the probability of the fuzzy modality v_i [23]. This probability is computed from the n examples x_l belonging to the training set.

After the construction of a fuzzy decision tree, Salammbô is able to classify new examples. Each classified example is associated with a membership degree to each class. A fuzzy decision tree is considered as a set of fuzzy rules on which usual inference methods can be applied. We need to evaluate the adequation of the values associated with the corresponding modalities labeling the edges of the tree. We use a measure of satisfiability for this evaluation. Each leaf is labelled by classes with membership degrees obtained by aggregating the satisfiability degrees attached to each edge of the path from the root to the leaf. All the membership degrees related to a class are again aggregated to obtain one single degree related to the class. Thus, an example can fired several paths of the tree and can be associated with several classes, each class associated with a membership degree (for more details, see [13], [3]).

3 A method when there exists more than two classes

3.1 Learning

Learning with decision trees is difficult when there exists more than two classes. Difficulties lie in the chosen measure of discrimination. Measures of discrimination enable to find the attribute with the best discriminating power related to the whole set of classes. However, the chosen attribute is not necessarily the attribute with the best discriminating power for a single class [7]. To handle this problem,

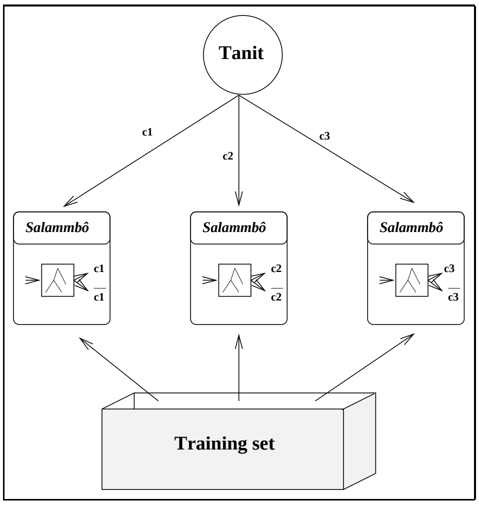

Figure 1 - *Learning*

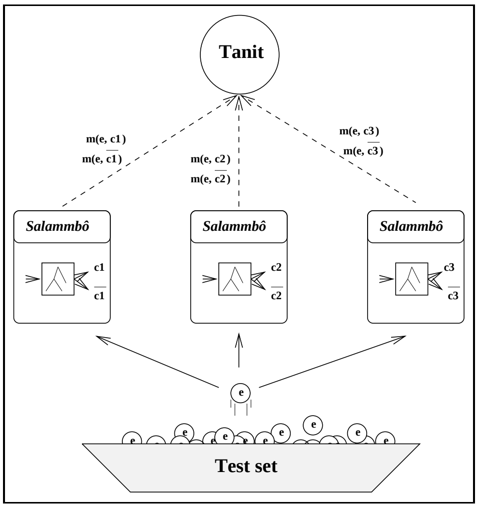

Figure 2 - *Classifying*

either a new kind of measures can be introduced [7] or the training set with more than two classes can be transformed into a training set with only two classes. For instance, a solution proposed by [6] is to combine classes at each step of the construction of the tree. But this solution is time consuming to find the best combination during the learning process.

Another solution is to combine classes before the building of the tree. Here, the training set is duplicated into several sets. In each such set, a class is kept and all the other classes are combined to create a single class. This process is called the binarization of the set of classes. There exist n different ways to binarize a set of classes $\{c_1, c_2 \ldots c_n\}$: $\{c_1, \overline{c_1}\}$, $\{c_2, \overline{c_2}\} \ldots \{c_n, \overline{c_n}\}$ where $\overline{c_i}$ represents the set of all the classes different from c_i. This is equivalent to learn a class against all the other classes together. *Figure 1* shows this process of learning.

The application, called *Tanit*, supervises the whole process. During the learning phase, Tanit binarizes the set of classes and runs as many Salammbô applications as classes in the training set. Each Salammbô has to build a fuzzy decision tree from a special training set where a class is outlined against all the other classes together.

For instance, if there exists three classes c_1, c_2 and c_3 in the training set, Tanit creates three Salammbô, one to recognize the class c_1, another to recognized the class c_2 and the last one to recognize the class c_3.

3.2 Classifying

Now, it is important to be able to utilize each so built fuzzy decision tree to classify examples from a test set (that is a set of examples whose the only description is known). In this phase, each Salammbô must classify examples from the test set. This classification is related to the class that the corresponding Salammbô is able to recognize.

Thus, for each example, Tanit aggregates the degrees of membership to classes from all the Salammbô. It is a problem of data fusion. Lots of works study this kind of aggregation [21], [9], [11]. For instance, [8] studies the aggregation of a forest of decision trees. During the learning phase, the system builds several decision trees by means of several measures of discrimination. If several attributes are chosen by these measures, each attribute will enable the construction of a new tree. Thus, a high number of trees could be generated. On the contrary, in the Tanit system, the number of trees is only determined by the number of classes.

During the generalization phase, degrees of membership to classes are provided by each Salammbô for each example. For each example, the Salammbô adapted to recognize class c_i returns a membership degree $m_{c_i}(\{c_i\})$ to the class c_i and a membership degree $m_{c_i}(\{\overline{c_i}\})$ to $\overline{c_i}$ (*Figure 2*). These two degrees are returned to Tanit which centralizes all the membership degrees provided by all the Salammbô. All the membership degrees related to an example are aggregated to obtain a single membership degree $\mu(c_i)$ for each class c_i. Various methods exist to realize such an aggregation. For instance, the voting method or the method based on Dempster combination rule in Dempster-Shafer theory of evidence [19].

3.2.1 The voting method

The membership degrees are considered as votes. Each vote for each class is added to elect the most popular class. The membership degrees given by the Salammbô are added:

$$\mu(c_i) = m_{c_i}(\{c_i\}) + \frac{1}{|n|} \sum_{c_j \neq c_i} m_{c_j}(\{\overline{c_j}\})$$

The membership degrees associated with a combination of classes ($\overline{c_j}$) are weighted by the number of classes n in order to limit the influence of trees not specialized in the recognition of the class c_i. Moreover, the degree $m_{c_i}(\{\overline{c_i}\})$ has to be split for the computation of the votes for all the classes which belong to $\overline{c_j}$.

Finally, the elected class c_i is the one with the higher ranking vote $\mu(c_i)$.

3.2.2 Dempster's rule method

In evidence theory, each Salammbô is considered as an expert which expresses its opinion on a given situation. The universe of discourse is then $\Omega = \{c_1, \ldots c_n\}$ and the two values $m_{c_i}(\{c_i\})$ and $m_{c_i}(\{\overline{c_i}\})$ returned by Salammbô are normalized to be considered as basic probability numbers.

The aggregation of degrees is done by means of Dempster combination rule. For a given example, the membership degree $\mu(c_i)$ to each class c_i is computed as:

$$\mu(c_i) = K\ m_{c_i}(\{c_i\}) \prod_{c_j \neq c_i} m_{c_j}(\{\overline{c_j}\})$$

(Where K is a normalization coefficient)

For instance, when there exists three classes, the following computation is done by Tanit:

$$\begin{aligned}
\mu(c_1) &= K\ m_{c_1}(\{c_1\})\ m_{c_2}(\{\overline{c_2}\})\ m_{c_3}(\{\overline{c_3}\}) \\
\mu(c_2) &= K\ m_{c_2}(\{c_2\})\ m_{c_1}(\{\overline{c_1}\})\ m_{c_3}(\{\overline{c_3}\}) \\
\mu(c_3) &= K\ m_{c_3}(\{c_3\})\ m_{c_1}(\{\overline{c_1}\})\ m_{c_2}(\{\overline{c_2}\})
\end{aligned}$$

Finally, the class associated with the example is the class with the higher membership degree.

4 Results

The two applications Salammbô and Tanit have been implemented in Language C on Sun-Sparc station. The application Tanit runs as many Salammbô applications as classes in the training set. Each Salammbô is associated with a class to recognize. The whole system runs on a single computer but a distributed version is being implemented. In this version, each Salammbô will be run on a machine apart, the communication between all the Salammbô and Tanit will be done by means of the RPC protocol. Thus, the time of building a forest of fuzzy decision trees will be greatly minimized.

The current Tanit application has been tested on various kinds of 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 databases which have been used are those with a high proportion of numeric data in order to highlight the utility of fuzzy decision trees.

The results with the iris database are shown in the first table. In this database, examples are described by means of 4 numeric attributes and there are 3 classes to recognize.

Method	*Error rate*	
	Classic tree	*Fuzzy tree*
Single tree	4,8%	4,7%
Vote	4,7%	**4,0%**
Dempster's rule	5,3%	5,3%

Result with the iris database

The results with the database of Breiman's waveforms [4] are shown in the second table. In this database, examples are described by means of 21 numeric attributes and there are 4 classes to recognize.

Method	*Error rate*	
	Classic tree	*Fuzzy tree*
Single tree	27,3%	23,7%
Vote	26,7%	**20,2%**
Dempster's rule	28,5%	28,4%

Result with the waveform database

Results with *Classic tree*, that is a tree provided by the classical ID3 method which discretizes the universe of numeric values and does not consider the threshold as fuzzy, are given in the first column of the table (*Classic tree*). For the Classic *Single tree* results, the results are those given by [16] for the C4.5 algorithm, adapted to numeric attribute. Results with a forest of (classic or fuzzy) decision trees are shown in the two other lines.

5 Conclusion

In this paper, we present a method of inductive learning by means of a forest of fuzzy decision trees. In this method, several fuzzy decision trees are constructed, each one is dedicated to the recognition of a single class against a combination of all the other classes.

The use of such combination of classes enables the use of discriminating measures adapted to the ordering of attributes when only two classes are present in the training set. Moreover, the roots of all the fuzzy decision trees of the forest are rarely labeled by the same attribute. Thus, an attribute does not appear necessarily in each rule of the rule base induced by a forest of decision trees.

The aggregation step of this system could be enhanced and some works are conducted at the moment to find a good method of combination of membership degrees in order to enhance the predicting power of the forest of fuzzy decision trees.

Finally, this method enables us to study the best discriminating measure to build a fuzzy decision tree. The advantages of using a forest of fuzzy decision trees lie in the fact that the discriminating measure to study should reflects the discriminating power of an attribute related to only two classes.

References

1. S. Bothorel. *Analyse d'image par arbre de décision flou. Application à la classification sémiologique des amas de microcalcifications.* PhD thesis, Université P. et M. Curie, Paris, France, Décembre 1996.

2. B. Bouchon-Meunier, C. Marsala, and M. Ramdani. Arbres de décision et théorie des sous-ensembles flous. In *Actes des 5èmes journées du PRC-GDR d'Intelligence Artificielle*, pages 50–53, 1995.

3. B. Bouchon-Meunier, C. Marsala, and M. Ramdani. Learning from imperfect data. In H. Prade D. Dubois and R. R. Yager, editors, *Fuzzy Information Engineering: a Guided Tour of Applications*, pages 139–148. John Wileys and Sons, 1997.

4. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification And Regression Trees.* Chapman and Hall, New York, 1984.

5. J. Catlett. On changing continuous attributes into ordered discrete attributes. In Y. Kodratoff, editor, *Machine Learning – Proceedings of EWSL-91*, number 482 in Lecture notes in Artificial Intelligence, pages 164–178. Springer-Verlag, 1991.

6. F. D'Alché-Buc. *Modèles neuronaux et algorithmes constructifs pour l'apprentissage de règles de décision.* PhD thesis, Université Paris-Sud, Orsay, France, décembre 1993.

7. U. M. Fayyad and K. B. Irani. The attribute selection problem in decision tree generation. In *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 104–110. AAAI, 7 1992.

8. N. Feray and D. Brucq. Fusion d'information provenant d'une famille d'arbres de décision. application à la reconnaissance de chiffres manuscrits. In *Information Signal ImageS (ISIS), Journées thématiques en fusion d'informations, GDR-PRC*, pages 4–13, Paris, France, octobre 1996.

9. T. K. Ho, J. J. Hull, and S. N. Srihari. On multiple classifier systems for pattern recognition. In *Proceedings of the 11th IAPR International Conference on Pattern Recognition*, volume II, pages 84–87, The Hague, Netherlands, September 1992.

10. C. Z. Janikow. Fuzzy decision trees: FIDMV. In *JCIS'94 Proceedings of the Joint Conference on Information Sciences*, pages 232–235, Pinehurst, 1994.

11. P. Loonis. *Contribution à la minimisation de l'a priori en Reconnaissance des Formes.* PhD thesis, Université de La Rochelle, France, Janvier 1996.

12. C. Marsala and B. Bouchon-Meunier. Fuzzy partioning using mathematical morphology in a learning scheme. In *FUZZ'IEEE'96*, New Orleans, USA, September 1996.

13. C. Marsala and M. Ramdani. Connaissances expertes floues et système d'apprentissage descendant. In *Rencontres Francophones sur la Logique Floue et ses Applications, LFA'95*, pages 163–168, Paris, Novembre 1995. Cépaduès éditions.

14. C. F. Picard. *Graphes et Questionnaires.* Gauthier-Villars, 1972.

15. J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):86–106, 1986.

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

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

18. M. Ramdani. *Système d'Induction Formelle à Base de Connaissances Imprécises.* PhD thesis, Université P. et M. Curie, Paris, France, Février 1994. Aussi publiée en rapport du LAFORIA-IBP n^o TH94/1.

19. G. A. Shafer. *Mathematical Theory of Evidence.* Princeton University Press, New Jersey, 1976.

20. R. Weber. Fuzzy-ID3: A class of methods for automatic knowledge acquisition. In *IIZUKA'92 Proceedings of the 2nd International Conference on Fuzzy Logic*, volume 1, pages 265–268, july 1992.

21. L. Xu, A. Krzyzak, and C. Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3):418–435, May/June 1992.

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

23. L. A. 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 and S. Ovchinnikov and R. M. Tong and H. T. Nguyen eds, pp. 45–51.