

# The research of fuzzy decision trees building based on entropy and the theory of fuzzy sets

S B Begenova<sup>1</sup> and T V Avdeenko<sup>1</sup>

<sup>1</sup>Novosibirsk State Technical University, Karla Marks ave 20, Novosibirsk, Russia, 630073

**Abstract.** Decision trees are widely used in the field of machine learning and artificial intelligence. Such popularity is due to the fact that with the help of decision trees graphic models, text rules can be built and they are easily understood by the final user. Because of the inaccuracy of observations, uncertainties, the data, collected in the environment, often take an unclear form. Therefore, fuzzy decision trees are becoming popular in the field of machine learning. This article presents a method that includes the features of the two above-mentioned approaches: a graphical representation of the rules system in the form of a tree and a fuzzy representation of the data. The approach uses such advantages as high comprehensibility of decision trees and the ability to cope with inaccurate and uncertain information in fuzzy representation. The received learning method is suitable for classifying problems with both numerical and symbolic features. In the article, solution illustrations and numerical results are given. Also the comparison of fuzzy logic approaches for building fuzzy rules and classification trees are given.

## 1. Introduction

Nowadays, in the era of big data, the extraction of knowledge is a bottleneck in the field of knowledge engineering. Computer programs that extract knowledge from data successfully try to solve this problem. Among these programs, systems for building decision trees for decision-making and classification tasks are very popular. The knowledge acquired in the form of decision trees and inference procedures is highly valued for the clarity and visibility of the data. Such assessment, at one time, aroused interest of scientists, which led to a number of methodological and empirical achievements. However, initially decision trees were popularized by Quinlan and his ID3 algorithm [1].

One of the extensions of the classical construction of decision trees is an approach based on fuzzy logic. Fuzzy approach is becoming increasingly popular in solving problems of uncertainty, noise and inaccurate data. It is successfully applied to problems in many industrial spheres. Most studies on the application of this representative framework to existing methodologies are focused mainly on new areas, such as neural networks and genetic algorithms. Nowadays, the fuzzy approach that integrates the concepts of fuzzy sets and entropy is becoming popular.

This article presents a method that includes the features of the two above-mentioned approaches: a graphical representation of the rules system in the form of a tree and the fuzzy representation of the data. Section 2 describes the principle of the decision trees, their advantages and disadvantages, algorithms for their construction. Section 3 shows the principle of constructing fuzzy decision trees, introduces the concepts of fuzzy logic. Section 4 describes the results of the study and the last section gives a conclusion.

## 2. Decision trees

A decision tree (DT) is a common formalization for mapping the transitions of attribute values to classes in the form of a map, which consists of attribute nodes or so-called tests that can have two or more subtrees, leaves, or decision nodes that are labeled with a class indicating the solution. The main advantage of this approach is the visualization of the solution. One of the most commonly used algorithms for constructing decision trees is the ID3 method, formalized by Quinlan in 1986 [1].

Decision trees create efficient models for machine learning [11, 12]. Let us give the following characteristics of decision trees:

- they are easily interpretable and visible;
- the model can be expressed both graphically and with text rules;
- they are competitive in comparison with more expensive approaches;
- decision trees are scalable;
- they can process discrete and continuous data;
- decision trees can be applied to different sizes of data sets, including large sample sets.

In the process of tree constructing, the pattern is represented by a set of features that are expressed in some descriptive language. Samples whose characteristics are known are called examples. The purpose of constructing a tree is to solve the problem of classification or regression.

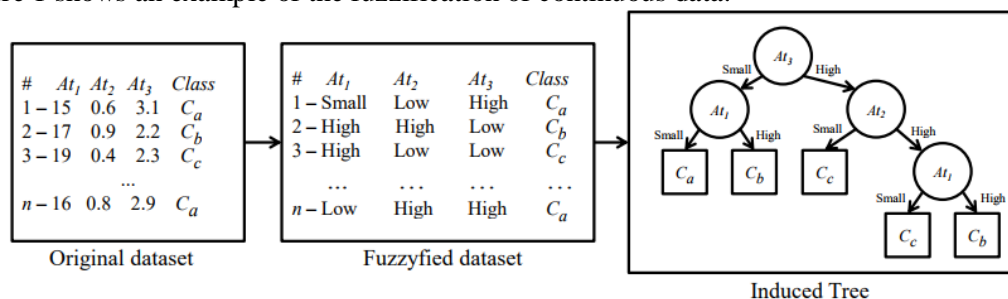
ID3 and CART are the two most important discriminating learning algorithms that work by recursive partitioning. Their basic ideas are approximately the same: splitting the incoming sample into subsets and representing the partitions as a tree. An important property of these algorithms is that they simultaneously try to minimize the size of the tree with the optimization of some quality measure. Subsequently, they use the same logical inference.

## 3. Fuzzy decision trees

To construct a fuzzy decision tree, the following procedure is proposed [4]:

1. Define the fuzzy data base, i.e., the fuzzy granulation for the domains of the continuous features.
2. Replace the continuous attributes of the training set using the linguistic labels of the fuzzy sets with highest compatibility with the input values [5, 6].
3. Calculate the entropy and information gain of each feature to split the training set and define the test nodes of the tree until all features are used or all training examples are classified.

Figure 1 shows an example of the fuzzification of continuous data.



**Figure 1.** Algorithm for constructing a fuzzy decision tree.

The first block of Figure 1 illustrates a dataset with  $n$  examples, three attributes ( $At_1, At_2, At_3$ ) and a class attribute. The fuzzified version of this dataset is presented in the second block. This fuzzified set of examples is used to induce the final DT, illustrated in the last block of Figure 1.

The entropy and information gain formulas remain the same for the classical version of the ID3 algorithm [10]. Let us introduce the following notation:

$U = \{u_1, u_2, \dots, u_s\}$  – set of data samples;  $C = \{c_1, c_2, \dots, c_n\}$  – set of attributes;  $D = \{d\}$  – a singleton set with a solution attribute or class attribute. Let this attribute have  $m$  different values, then,  $n_i$  is the number of samples of set  $U$  in class  $d_i$ .

Information gain  $I$  relative to subset  $S_j$  is equal to:

$$I(s_{1j}, s_{2j}, \dots, s_{mj}) = -\sum_{i=1}^m p_{ij} \log_2 p_{ij},$$

$$p_{ij} = \frac{s_{ij}}{|S_j|},$$

where  $|S_j|$  – the number of samples in a subset of S.

Entropy  $E(c_i)$  is:

$$E(c_i) = \sum_{j=1}^v \frac{s_{1j} + s_{2j} + \dots + s_{mj}}{s} I(s_{1j}, s_{2j}, \dots, s_{mj});$$

accordingly, the criterion for selecting an attribute is the increase in information:

$$Gain(c_i) = I(s_{1j}, s_{2j}, \dots, s_{mj}) - E(c_i).$$

The difference between common algorithm ID3 and the fuzzy version of algorithm ID3 is that the attributes of objects have degrees of belonging to a particular node, and it is quite possible that an attribute with certain probabilities belongs to several nodes.

Figure 2 shows two decision trees that were built using the above-mentioned algorithms.

As an example, a classic data set was taken, Fisher's iris [9], which has 4 attributes: the length and width of the cup, the length and width of the petal and the three resultant classes - setosa, versicolor and virginica.

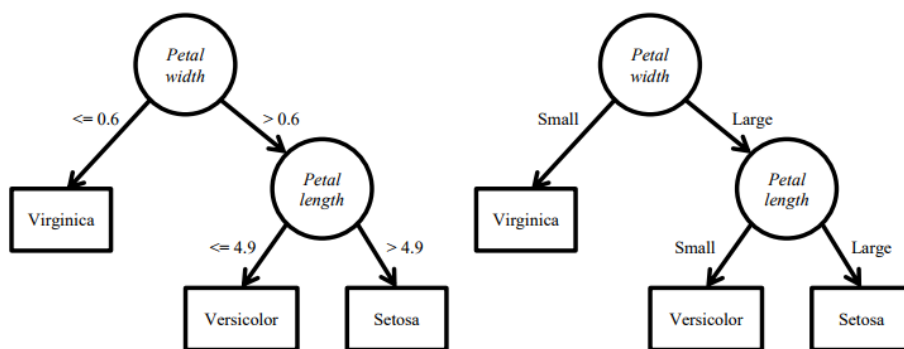


Figure 2. Classical (left) and fuzzy (left) decision trees.

#### 4. Theresearch results

As a research object, just as in the previous example, a Fisher's iris data set was used. To construct a fuzzy decision tree at the first stage, it is necessary to perform a fuzzification procedure.

While performing the fuzzification procedure, the definition set of fuzzy attributes is divided into fuzzy subsets. The value of the fuzzy attribute is put in correspondence with the term, and this correspondence is found using the membership function. The division of the definition set into fuzzy subsets can be made evenly, that is, the definition set is divided into equal intervals. However, in most real data sets obtained from the environment, it is preferable to perform the partitioning taking into account the features of the original sample. For example, it may happen that most of the sampling objects lie in the first third of the definition set and, in this case, uniform partitioning will not give the desired effect. The results of fuzzification for attributes SepalLength, SepalWidth, PetalLength and PetalWidth are shown in Figures 3, 4, 5 and 6 respectively.

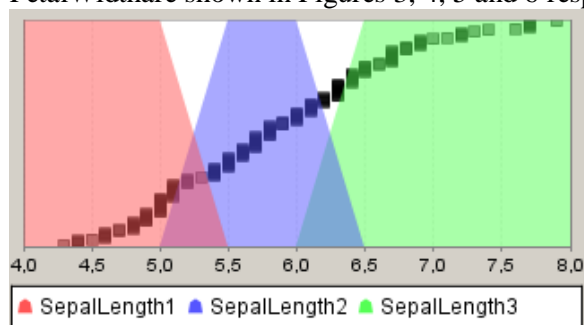


Figure 3. Sepal Length Attribute fuzzification.

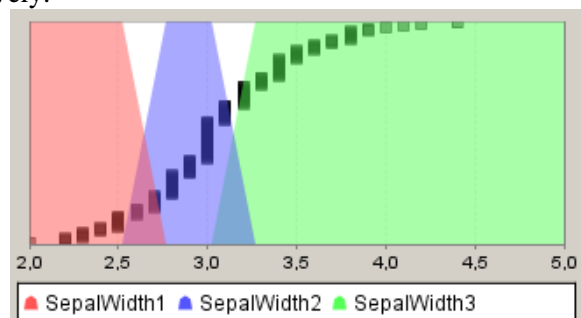


Figure 4. Sepal Width Attribute fuzzification.

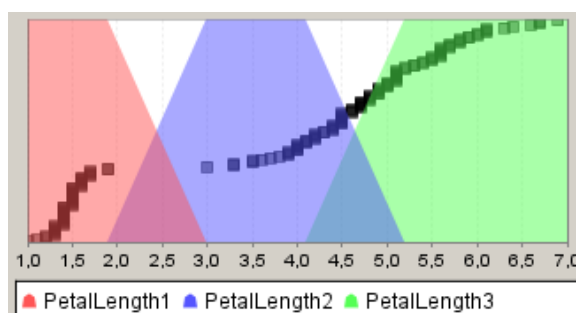


Figure 5. Petal Length Attribute fuzzification.

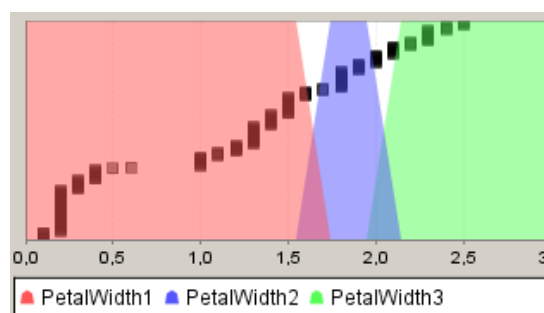


Figure 6. Petal Width Attribute fuzzification.

Numeric attributes have been assigned 3 terms (low, medium, high). Let us present the results of the classification obtained with the help of fuzzy decision trees. In the context of this studies, let us introduce the following notations:

Correct - the number of correctly classed sample objects.

Incorrect - the number of incorrectly classified sample objects.

WithoutClass - the number of objects without a class.

PercentCorrect is the percentage of correctly-categorized sample objects that is calculated as follows:

$$PercentCorrect = 100 - \frac{(Incorrect + WithoutDecision) * 100}{Correct}$$

To study the hypothesis that with a decrease of the sample size, the accuracy of the classification of fuzzy decision trees is better than classifying with classical ones, the dependence of the classification accuracy on the number of instances in the data set was constructed.

In this study, the trees were constructed for 3 randomly selected N samples and the table shows the averaged values obtained (the sum of the values of the attributes / 3).

According to the data presented in Table 1, it can be seen that when the sample is reduced from 150 to 90, the accuracy of classification using fuzzy decision trees is three percent higher than results of classifying with classical decision trees, and when the sample is reduced to 60, the accuracy is higher by 0.82 percent.

Table 1. Comparison of classification results obtained using fuzzy decision trees and classical decision trees.

The number of instances in the data set (N)	Fuzzy decision trees	Classical decision trees
120	Correct = 115 Incorrect = 5 WithoutClass = 0 PercentCorrect= 95.65	Correct = 117 Incorrect = 3 WithoutClass = 0 PercentCorrect= 97.43
90	Correct= 88 Incorrect = 2 WithoutClass = 0 PercentCorrect= 97.72	Correct= 85 Incorrect = 5 WithoutClass = 0 PercentCorrect= 94.11
60	Correct= 58 Incorrect = 2 WithoutClass = 0 PercentCorrect= 95.55	Correct= 57 Incorrect = 3 WithoutClass = 0 PercentCorrect= 94.73

Table 2 shows the dependence of the accuracy of data classification on the number of terms.

According to the data in the table, it is clear that the optimal number of terms for the test set of data is 5. Such quantity gave a higher percentage of correctly classified data compared to 3 terms.

Table 3 shows the dependence of the accuracy of the data classification on the value of the information gain. In this method, the increment of information will be the breakpoint of the algorithm, that is, when the specified value is reached, further building of the tree stops. According to the data it is clear that the lower the information gain, the more accurate and "deeper" the tree will be built.

**Table 2.** Dependence of the accuracy of classification of data on the number of terms.

The number of terms	Classification results
3	Correct = 142 Incorrect = 8 WithoutClass = 0 Percent Correct= 94.36
5	Correct = 143 Incorrect = 7 WithoutClass = 0 Percent Correct = 95.33
7	Correct = 143 Incorrect = 7 WithoutClass = 0 Percent Correct= 95.33

**Table 3.** Dependence of the accuracy of classification of data on the information gain.

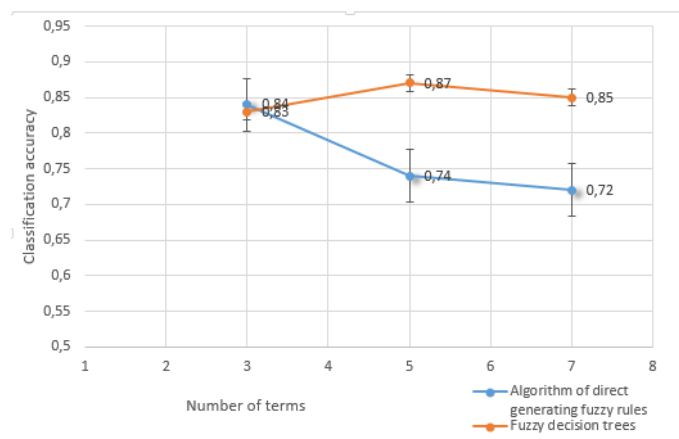
Information gain	Classification results
0.02	14 leaves Correct = 142 Incorrect = 8 WithoutClass = 0 PercentCorrect= 94.67
0.2	5 leaves Correct = 139 Incorrect = 11 WithoutClass = 0 PercentCorrect= 92.67
0.4	3 leaves Correct = 119 Incorrect = 31 WithoutClass = 0 PercentCorrect= 79.33

As a part of fuzzy decision trees research, we compare two methods of classification based on fuzzy logic. The first one is algorithm of direct generating fuzzy linguistic rules, proposed in [3]. The second method of fuzzy decision trees which is proposed in this article was used. Fig.7 gives visual illustration of comparison between the two methods for sequentially growing number of terms.

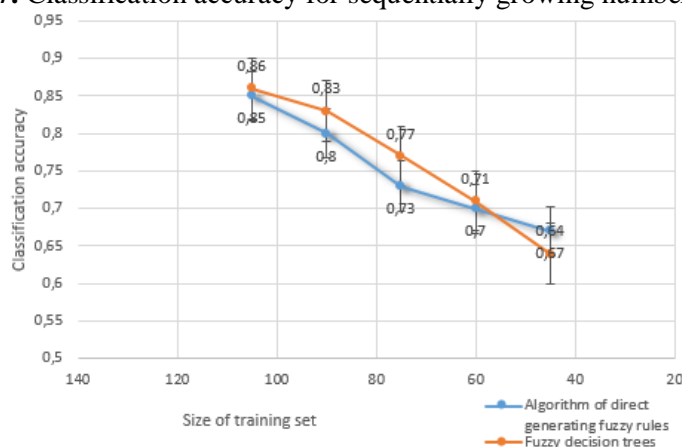
Here we can observe that the classification accuracy for sequentially growing number of terms (from 3 cases to 7) remains quite high. Method of fuzzy decision trees is better for medium and high sizes of terms while the method of direct generating fuzzy rules is better for small size of training sample. Further research will be in development of algorithm based on combination of both approaches.

On fig. 8 we can observe that the classification accuracy for sequentially reducing size of training sample (from 105 cases to 45) remains quite high. Method of fuzzy decision trees is better for medium size of training sample while the method of direct generating fuzzy rules is better for small size of training sample. Further research will be in development of algorithm based on combination of both approaches.

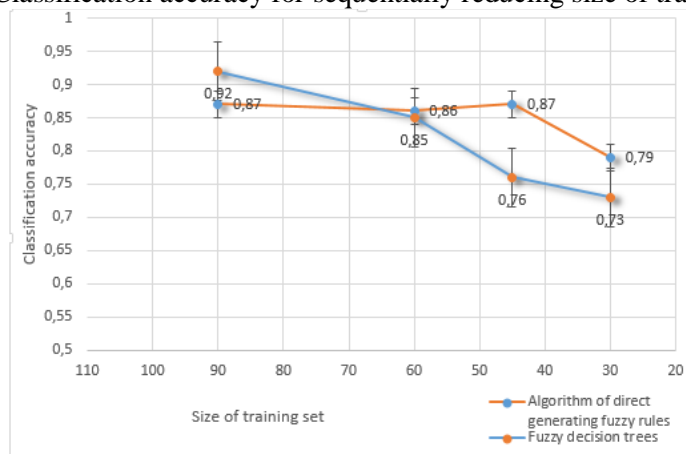
Fig. 9 and Fig. 10 illustrate the comparison between the two methods with sequentially reducing size of training sample using T-class and S- class membership functions respectively.



**Figure 7.** Classification accuracy for sequentially growing number of terms.



**Figure 8.** Classification accuracy for sequentially reducing size of training sample.



**Figure 9.** Classification accuracy for sequentially reducing size of Iris dataset training sample (T – class).

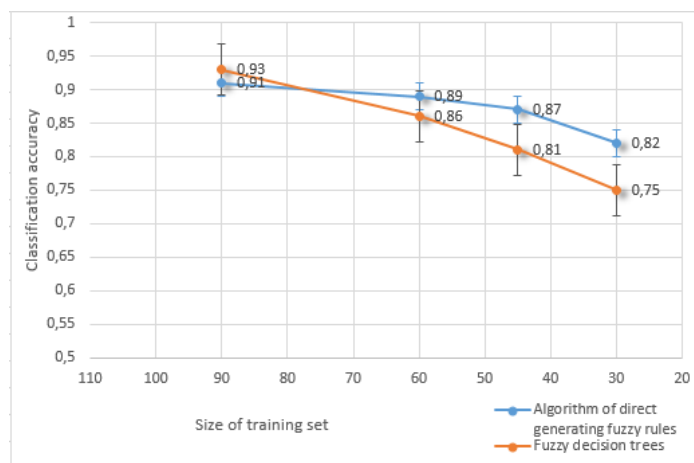
T – class membership function also known as triangular is specified by three parameters {a, b, c} as follows:

$$t(x, a, b, c) = \begin{cases} 0, & x \leq a \\ \frac{x - a}{b - a}, & a \leq x \leq b \\ \frac{c - x}{c - b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases}$$

S – class membership function is specified by two parameters {a, b} as follows:

$$s(x, a, b) = \begin{cases} 0, & x \leq a \\ 1 & \\ 1 + e^{-a(x-b)} & \\ 0, & x \geq b \end{cases}$$

For T-class membership functions, we observe that fuzzy decision trees method is better for large size of training sample while the method of direct generating fuzzy rules is better for small and medium size of training sample. In case of S-class membership functions, we observe the same situation.



**Figure 10.** Classification accuracy for sequentially reducing size of Iris dataset training sample (S – class).

## 5. Conclusion

Decision trees are successfully used to solve regression and classification problems. They are popular in the field of machine learning, because decision trees build graphic models, along with text rules that are easily interpreted by the users. On the other hand, fuzzy systems can solve classification problems with input inaccurate and noisy data.

The combination of fuzzy trees and fuzzy logic makes it possible to construct intuitive graphic models for qualitative and quantitative data [2, 7, 8]. Usage of this type of decision tree gives us several solutions with different probabilities of belonging to a particular class.

In addition, in the course of the conducted studies, the advantage of classification using fuzzy decision trees with respect to classical ones was revealed, by comparing the percentage of correctly classed objects. Also, a direct correlation between the accuracy of the classification and the value of the information gain was revealed (the increment is a criterion for stopping the further construction of the tree).

The comparison between algorithm of direct generating fuzzy linguistic rules and method of fuzzy decision trees didn't reveal the one and only right one. Both methods show high classification accuracy under certain conditions. The proposed approach can be applied to built fuzzy neural networks [13].

## 6. References

- [1] Quinlan J R 1986 Induction of decision trees *Machine learning* **1** 81-106
- [2] Cintra M E, Meira C A A, Monard M C, Camargo H A and Rodrigues L H 2011 The use of fuzzy decision trees for coee rust warning in Brazilian *Int. Conf. Int. Sys. Design & Applications* **1** 1347- 1352
- [3] Avdeenko T V and Makarova E S 2017 Acquisition of knowledge in the form of fuzzy rules for cases classification *Lecture Notes in Computer Science* **10387** 536-544
- [4] Cintra M E, Monard M C and Camargo H A 2012 Fuzzy DT- a fuzzy decision tree algorithm based on C4.5 *CBSF -Brazilian Congress on Fuzzy Systems* 199-211

- [5] Janikow C Z 1998 Fuzzy Decision Trees: Issues and Methods *IEEE Transactions of Man, Systems, Cybernetics* **28(1)**
- [6] Faifer M and Janikow C Z 2000 Bottom-up Partitioning in Fuzzy Decision Trees *Proceedings of the 19th International Conference of the North American Fuzzy Information Society* 326-330
- [7] Tokumaru M and Muranaka N 2010 Impression analysis using fuzzy c4.5 decision tree *Int. con. on Kansei engineering and emotion research*
- [8] Janikow C Z 2004 Fid 4.1: an overview *Proc. of the North American Fuzzy Information Processing Society* 877-881
- [9] *Machine Learning Repository* (Access mode: <http://archive.ics.uci.edu/ml/datasets/Iris>) (30.05.2018)
- [10] Begenova S B and Avdeenko T V 2018 Building of fuzzy decision trees using ID3 algorithm *Journal of Physics: Conference Series* **1015**
- [11] Cintra M E, Meira C A A, Monard M C, Camargo H A and Rodrigues L H 2011 The use of fuzzy decision trees for coffee rust warning in Brazilian crop *Int. Conf. Int. Sys. Design & Applications* **1** 1347-1352
- [12] Olaru C and Wehenkel L 2003 Fuzzy Sets and Systems **138** 221-254
- [13] Soldatova O P, Lezin I A, Lezina I V, Kupriyanov A V and Kirsh D V 2015 Application of fuzzy neural networks to determine the type of crystal lattices observed on nanoscale images *Computer Optics* **39(5)** 787-795 DOI: 10.18287/0134-2452-2015-39-5-787-794

### **Acknowledgments**

The work is supported by a grant from the Ministry of Education and Science of the Russian Federation within the framework of the project part of the state task, project No. 2.2327.2017 / 4.6 “Integration of knowledge representation models based on intellectual analysis of large data to support decision making in the field of software engineering.”