Recent Fuzzy Decision tree Algorithms

Dr. Jaibhagwan, Dr. Parvesh Kumar, Dr. Pushpander Kadian,

[1]Associate Professor, Pt. N.R.S. Govt.College, Rohtak (Haryana), India

[2]Associate Professor, Pt. N.R.S. Govt.College, Rohtak (Haryana), India

[3]Associate Professor, Govt.PG Nehru College, Jhajjar (Haryana), India

Abstract:

The digital revolution has made digitized information easy to capture and fairly inexpensive to store. The rate at which such data is stored is growing at a phenomenal rate. As a result, traditional ad hoc mixtures of statistical techniques and data management tools are no longer adequate for analysing this large collection of data. Soft computing methodologies (involving fuzzy sets, neural networks, genetic algorithms, and decision tree) are emerged as an alternative to traditional techniques. In present paper, we have studied recent developments in the field of fuzzy logic-based decision tree algorithms and the possibilities of future improvements.

Keywords: Algorithm, Fuzzy Set, Rough Set, Decision Tree, Leaf Node.

Introduction:

With the new innovations in the field of computer hardware & software and the rapid computerization of business, large amount of information has been collected and stored in databases. As a result, traditional statistical techniques and data management algorithms are no longer relevant for analysing this large collection of data. Soft computing techniques are the alternate to traditional techniques and provide new information processing capabilities for handling real world problems.

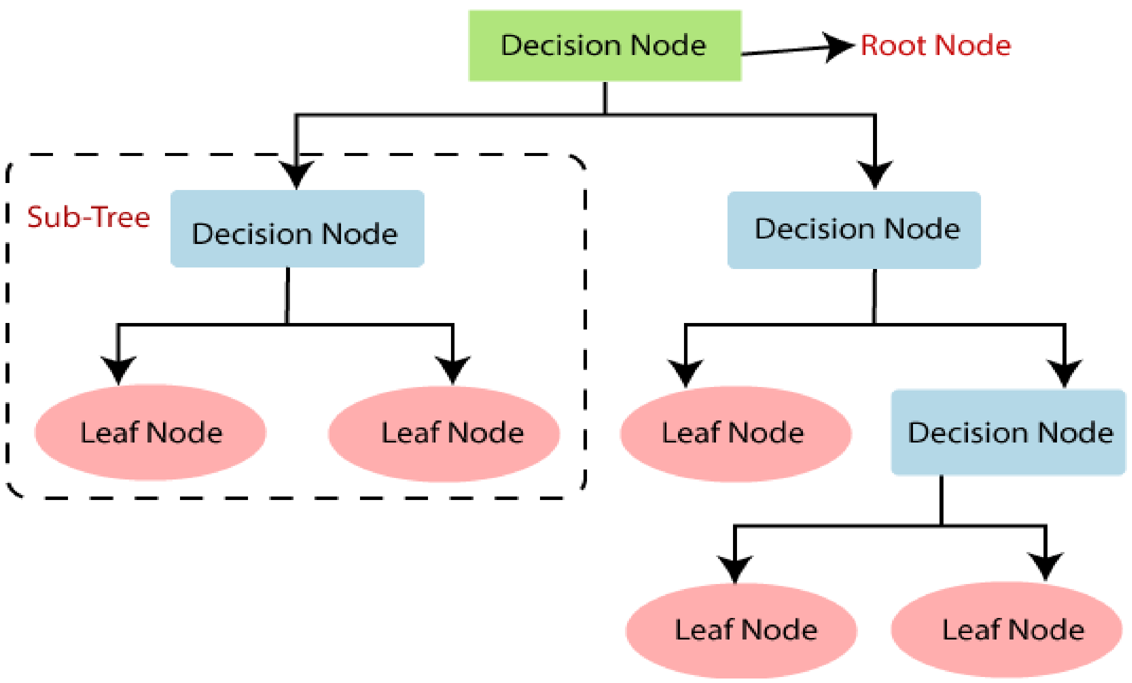
Classification technique is a 2-step technique which involves learning phase and testing phase in machine learning. In the learning phase, the model is prepared using the given training dataset while in the testing phase; the model is prepared to predict the possible outcome for given data. Decision Tree is one of the popular classification techniques to understand and interpret the data and information. Decision tree algorithms, their basic structure and the developments in the field of decision tree algorithms are briefed in this article. Fuzzy sets are sets whose elements have degrees of membership. Fuzzy classification is the procedure of dividing elements into fuzzy sets whose membership functions are defined by the truth value of a fuzzy propositional function. The brief introduction to fuzzy logic and various developments in decision tree algorithms with fuzzy logic has also been discussed herewith.

Decision Tree is a supervised learning technique that can be used for both classification and regression problems, but mostly it is preferred for solving classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, the Decision

Node and the Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas the Leaf nodes are the output of those decisions and do not contain any further branches. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. Basic steps involved in a decision tree algorithm are: -

- Step-1: Start the tree with the root node, says T, which includes the whole dataset.
- Step-2: Find the best attribute in the complete dataset using Attribute Selection Measure (ASM).
- Step-3: Divide the dataset T into subsets that contains possible values for the best attributes selected in step-2.
- Step-4: Generate the decision tree node, which contains the best attribute.
- Step-5: In this step we recursively make new decision trees using the subsets of the dataset created in step -3. This process continues till a stage is arrived where we cannot further classify the nodes and this final node is known as a Leaf node.

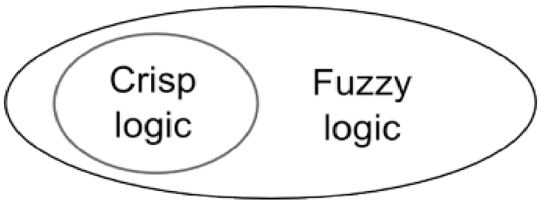
ID3 (Iterative Dichotomiser 3), a basic and first decision tree algorithm, was introduced in 1986 by Quinlan Ross. It is based on Hunt's algorithm and uses information theory invented by Shannon in 1948. ID3 algorithm does not guarantee an optimised solution. In most of the cases, this algorithm converges upon local optimum solution. ID3 algorithm can over fit the training dataset. It is difficult to use for the continuous data. If the values of any given attribute in the given dataset are continuous, then there are many more possibilities to split the dataset on this attribute and it make the process of splitting more time-consuming for the best value.



CART (Classification and regression trees) is another variation of decision tree algorithms and given by Leo Breimanin in 1984, which can handle both classification and regression tasks. CART algorithm uses Gini Impurity to split the dataset into a decision tree.It does that by searching for the best homogeneity for the sub nodes, with the help of the Gini index criterion.

SLIQ (Supervised Learning In Quest) (Mehta et al, 1996), SPRINT (Scalable Parallelizable Induction of decision Tree algorithm) (Shafer et al, 1996), QUEST(Quick, Unbiased, Efficient, Statistical Tree) (Loh and Shih, 1997) algorithm and BOAT(Bootstrap Optimistic Algorithm for Tree Construction) (Gehrke et al, 1999) are another improvements over the years in the field of decision tree algorithms.

The term fuzzy refers to things that are not clear or are vague. In the real world many times we encounter a situation when we can't determine whether the state is true or false, their fuzzy logic provides very valuable flexibility for reasoning. Fuzzy logic is an approach to variable processing that allows for multiple possible truth values to be processed through the same variable. Fuzzy logic attempts to solve problems with an open, imprecise spectrum of data and heuristics that makes it possible to obtain an array of accurate conclusions. Fuzzy logic is an extension of Boolean logic by Zadeh in 1965 based on the mathematical theory of fuzzy sets, which is a generalization of the classical set theory. By introducing the notion of degree in the verification of a condition which enables a condition to be in a state other than true or false, fuzzy logic provides a very valuable flexibility for reasoning, which makes it possible to take into account inaccuracies and uncertainties.



The major task of fuzzy-based pattern classification is the extraction of knowledge from numerical data to build a rule base, which will permit the classification of new data members.

Recent developments in fuzzy logic-based decision tree Algorithms:

Fuzzy logic allows decision trees to handle with continuous data and noise in a better way. FID decision tree differs from traditional decision trees in two ways: first, splitting criteria used in developing decision tree is based on fuzzy restrictions and secondly, inference procedures are also different. The FID algorithm was developed by Janikow in 1998 which can handle dataset containing various types of attributes like discrete, nominal, continuous and many more. Its construction is similar to the traditional decision tree, with a recursive depth-first method and different information content formula representing partial membership. Leaf node of fuzzy decision tree generally has samples of different classes with different degrees of membership. FID2.0, FID3.2, FID4, FID4.1 are some improvements of FID algorithm given by Janikow.

Fuzzy Decision Forest (FDF) is another development by Janikow in 2000 which is an extension of FID algorithm, builds the tree in the same way as in FID algorithm. But FDF allows multiple choices of alternative tests in some or all nodes of the decision tree which will help to improve accuracy of the algorithm and better choice selection at node level. Genetically Optimized Fuzzy Decision Tree (G-DT) is a decision tree algorithm given by Pedrycz & Sosnowski in the year 2005 which combines the concept of fuzzy logic and genetic optimization in decision trees to improve the performance of decision tree.

C-Fuzzy decision tree introduces a concept and design of decision trees based on information granules (Fuzzy clusters). The decision tree for C-Fuzzy DT grows around the fuzzy clusters created by FCM (Fuzzy c-means). These clusters are treated as generic building blocks for the tree. Initially the dataset is divided into clusters so that similar data samples are put together and these clusters are the top nodes of the tree. After that cluster (node) with highest value of heterogeneity criterion is divided further into clusters(nodes). For the C-decision trees, the number of nodes is equal to the number of clusters multiplied by the number of iterations. Here each node is associated by three components: heterogeneity criterion, no. of samples with it and list of these samples with degree of belongingness. This algorithm returns with compact decision trees and also uses all the attributes at a time. An improved version of this algorithm is proposed by Chiu et al. in 2006 by giving reasonable definition of the distance function and constructing local linear model for each leaf node.

A fuzzy supervised learning in Quest (SLIQ) decision tree (FS-DT) algorithm (Chandra & Paul, 2008) modifies the SLIQ decision tree algorithm to construct a fuzzy binary decision tree without converting the quantitative values into fuzzy linguistic terms and produces a tree of significantly reduced size. The entire decision tree is traversed to make an inference for a test sample. The classification accuracy is significantly better in the case of FS-DT compared to that of SLIQ.

FuzzyDT, proposed by M.E. Cintra & H.A. Camargo, uses the same measures of the classic C4.5 algorithm (entropy and information gain) to decide on the importance of the features. It also uses the same induction strategy to recursively partition the feature space creating branches until a class is assigned to each branch. However, for FuzzyDT, continuous features are defined in terms of fuzzy sets before the induction of the tree. This way, the process of inducing a tree using FuzzyDT takes a set of discretized features, since the continuous features are defined in terms of fuzzy sets and the training set is fuzzified before the decision tree induction takes place.

Fuzzy Entropy Based Fuzzy Partitioning (FEBFP) is new two-step algorithm developed by M. Zeinalkhani & M. Eftekhariin in 2014 which uses discretization algorithms for initial partitions and in second step,a membership function is defined for each partition generated in first step.

Fuzzy Logic Based Decision Tree Algorithms since 1998			
S.No.	**Decision Tree Algorithm based Fuzzy Logic**	**Author**	**Year**
1.	FID(Fuzzy Decision Tree)	Janikow	1998
2.	FDF(Fuzzy Decision Forest)	Janikow	2000
3.	IFN(Info-Fuzzy Network)	Last et al.	2002
4.	FRID(Fuzzy-Rough Interactive Dichotomizers)	Bhatt et al.	2004
5.	G-DT(Genetic-Decision Tree)	Pedrycz&Sosnowski	2005
6.	C-Fuzzy DT(c-Fuzzy DecisionTree)	Pedrycz&Sosnowski	2005
7.	Improved C-Fuzzy DT	Chiu et al.	2006
8.	Fuzzy SLIQ decision tree algorithm	B.Chandra&P.Paul Varghese	2008
9.	FuzzyDT	M.E. Cintra & H.A. Camargo	2010
10.	Fuzzy Entropy Based Fuzzy Partitioning (FEBFP)	M. Zeinalkhani&M. Eftekhari	2014
11.	FDT-Boost	Marco Barsacchi et al.	2017
12.	A three-way classification with FDT	Xiaoyu Han et al.	2023

FDT-Boost (Marco Barsacchi et al., 2017) is a boosting approach shaped according to the multi-class SAMME-AdaBoost scheme that employs size-constrained fuzzy binary decision trees as weak classifiers.

A three-way classification with FDT is developed by Xiaoyu Han recently in the year 2023 based on three-way classification mechanism realized through fuzzy decision trees and uncertainty measure methods to filter out dataset with large uncertainty. The developed mechanism creates a fuzzy decision tree by fuzzifying Boolean boundaries of already constructed decision tree algorithm and gives three techniques to calculate the level of ambiguity of the classification belongingness for an object. Here fuzzy sets are used to describe the boundaries of decision trees to produce fuzzy decision trees and three-way classification is introduced into fuzzy decision trees, which allows data with high uncertainty to be filtered out and further considered by the users. Secondly, three methods are used in this algorithm to quantify the level of uncertainty of the classification results and analyse the relationship between classification rates and rejected data. Users have options to choose an appropriate threshold to filter the dataset with high uncertainty according to the relationship and the cost of misclassification in real-world applications. Thirdly, conventional algorithms for making three-way decisions require a pair of thresholds while the developed mechanism in this algorithm could construct three regions just by using single threshold. There were some fuzzy decision trees in the existing literatures while they did not have these capabilities.

Conclusions:

As it appears, Decision tree algorithms have great promise for efficient classification of large and heterogeneous datasets. These techniques are very helpful in the solution of real-life problems and can suggest better solutions. However, concept of rough sets can also be combined with fuzzy decision trees to improve these algorithms. Fuzzy granulation, through rough-fuzzy computing, can be used to perform operations which will provide both information compression and gain in computation time; thereby making it suitable for data mining applications which will revolutionise classification and analysis techniques in future.

References:

1. B. Chandra and P. P. Varghese, On improving efficiency of SLIQ decision tree algorithm, in Proc. IEEE IJCNN, pp. 66–71, 2007.
2. B. Chandra and P. Pallath, Fuzzy SLIQ Decision Tree Algorithm, IEEE Trans. On systems, man and cybernetics, vol. 38, no. 5, 2008.
3. C. Z. Janikow, Fuzzy decision trees: Issues and methods, IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 28, no. 1, pp. 1–14, Feb. 1998.
4. C. Z. Janikow, Fuzzy processing in decision trees, in Proc. 6th Int. Symp. Artif. Intell., pp. 360–367, 1993.
5. D.F. Madau - Influence value defuzzification method. Fuzzy Systems, Proceedings of the Fifth IEEE International Conference, 3:1819 – 1824. [cited at p. 15],1996
6. Freund, Y., Mason, L.: The alternating decision tree learning algorithm. In: Proceeding of the Sixteenth International Conference on Machine Learning, Bled, Slovenia, 124-133, 1999.
7. J. C. Bezdek: "Pattern Recognition with Fuzzy Objective Function Algoritms", Plenum Press, New York, 1981.
8. J. R. Quinlan, C4.5: Programs for Machine Learning. Morgan Kaufman.
9. L. A. Zadeh : Fuzzy sets. Information and Control 8 (3) 338–353, 1965.
10. Leo Breiman, Jerome H. Friedman, Richard A. Olshen, Charles J. Stone. Classification and Regression Trees. Wadsworth International Group, Belmont, California, 1984.
11. M. E. Cintra and H. A. Camargo- Feature subset selection for fuzzy classification methods. In International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems - IPMU, volume 80, pages 318-327. Springer, 2010.
12. M. Mehta, R. Agrawal, and J. Rissanen, SLIQ: A fast scalable classifier for data mining, in Proc. Extending Database Technol., Avignon, France, pp. 18–32, 1996.
13. M. Barsacchi et al. Multi-class boosting with fuzzy decision trees Proc. of 2017 IEEE int'l conf. on fuzzy systems (FUZZ-IEEE) (2017)
14. Marco Barsacchi, Alessio Bechini, Francesco Marcelloni - An analysis of boosted ensembles of binary fuzzy decision trees, Expert Systems with ApplicationsVolume 154, 15 September 2020
15. Mohsen Zeinalkhani, Mahdi EftekhariFuzzy partitioning of continuous attributes through discretization methods to construct fuzzy decision tree classifiers Information Sciences, Volume 278, 2014, pp. 715-735
16. Niels Landwehr, Mark Hall, Eibe Frank. Logistic Model Trees. Machine Learning. 95(1-2):161-205, 2005.

17. R.B. Bhatt and M.Gopal, ―FRID: Fuzzy-Rough Interactive Dichotomizers,‖ IEEE International Conference on Fuzzy Systems (FUZZIEEE' 04), pp. 1337–1342, 2004.
18. Rajen B. Bhatt and M. Gopal, ―FRCT: Fuzzy-Rough Classification Trees‖, Pattern Analysis and Applications, 11(1), pp. 73-88, 2008.
19. Ross Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA, 1993.
20. W. V. Leekwijck and E.E. Kerre- Defuzzification: criteria and classification. Fuzzy Sets and Systems, 108(2):159 – 178, 1999
21. Xiaoyu Han et al. - A three-way classification with fuzzy decision trees, Applied Soft ComputingVolume 132, January 2023