ISSN: 0975-766X
CODEN: IJPTFI
Available Online through
www.ijptonline.com
Research Article

AN EMPIRICAL STUDY OF FUZZY DECISION TREES FOR PREDICTING THE PATIENTS MEDICAL BEHAVIOUR

Swathi J.Narayanan*[a], Priya. V[b], Boominathan Perumal[a], Ramani.S[a],
[a]School of Computing Science and Engineering, VIT University, Vellore-14.
[b]School of Information Technology and Engineering, VIT University, Vellore-14.
Email: swathi.jns@gmail.com

Received on 10-05-2016 *Accepted on 09-06-2016*

Abstract

In health sciences, prescribing drugs for the patients depends on the diagnosis conducted by the doctors or experts. As the world is moving toward the automation of diagnosis using computers, the prediction of diseases in the human can be automated by using machine learning algorithms which perform the task of classification/ prediction when given the patient details. To predict the patient diseases, a set of historical data is passed to the machine learning algorithms and the algorithm is trained to learn for prediction. Among several machine algorithms, decision tree is the prominent technique to understand how the decision is taken as it represents the classification knowledge in hierarchical form. As decision trees take crisp decision, it is not possible to handle fuzziness which is common in real world data, so fuzzy decision trees emerged. In this paper, we present an empirical study of fuzzy decision trees (FDTs) towards the classification of several medical datasets so as to facilitate an intelligent system to predict the disease and further it can lead to an option of automatic prescription of drugs for the patients as per the predictions made by the machine learning algorithm which can later be verified by the experts if desired.

Keywords: Fuzzy ID3, fuzzy c- means, pattern classification, fuzzy decision tree.

1. Introduction

Several machine learning algorithms are widely used for classification problem. But if the experts need to understand how a decision has been taken, then most of the algorithms suffer from lack of readability. Decision trees[1] are the ones which manage to give satisfactory results by providing understandable decision making rules. Decision tree namely ID3 (Interactive Dichotomiser3) was first proposed by Quinlan [1]. This initial algorithm takes only discrete data as input where as in real world several continuous attributes exists in most of the datasets. Further, several improvements emerged to deal with continuous attributes as well.

Still decision trees had issues in dealing with noisy data. i.e., if there is a small noise in data, which slightly change attribute values, classification of pattern starts taking completely different path in the tree. To overcome this issue, various researchers introduced Fuzzy Decision Tree (FDT) induction algorithms [3, 4] where cut points are defined by fuzzy functions and depending on overlap among fuzzy membership functions of a given variable, pattern travels on multiple paths. Using fuzzy decision trees.

We obtain understandable classification rules in the form of *if-then* kind of rules which represents classification knowledge extracted from data more naturally in the form of human thinking.

The FDT induction process consists of three modules namely fuzzy clustering, induction of fuzzy decision tree and fuzzy rule inference for classification. The fuzzy clustering can be performed using several clustering techniques available in literature. The most commonly used are Fuzzy c-means (FCM), Grid Partitioning and Subtractive Clustering.

The clustered data are later approximated to fuzzy membership functions like Triangular, Trapezoidal or Gaussian Membership functions. Swathi et.al [5] has come up with heuristic approaches to approximate the clustered data to triangular and trapezoidal membership function. The points ([a, b, c]) for triangular approximation and [a, b, c, d] for trapezoidal approximation) are obtained automatically from the FCM clustered raw data. Approximation to Gaussian membership functions is quite a straight forward and discussed in detail in Bhatt and Gopal [6].

The induction process of FDT follows a top-down recursive divide and conquer approach that makes locally optimal decisions at each node and the training set is recursively partitioned into smaller subsets as the tree is being built. The output of the induction process would be set of fuzzy rules which are further used to predict the class/ category of an unseen pattern by applying fuzzy reasoning mechanism on the FDT. The aim of any classification study is to find a mapping function between the input $x_i \mid i = 1,..p$ and output $y_i \in \{1,....,q\}$ based on training patterns $D = \{(x_i, y_i \mid i = 1,..n)\}$. The function *F*(*X, D*) predicts the class of unseen pattern by applying suitable fuzzy inference mechanism on the FDT.

In this paper, we study the variants of FDT for predicting the patient's medical behaviour in order to classes for the medical datasets so as to facilitate the doctors for automatic prescription of drugs and later experts can authenticate the same.

The rest of the paper has been organized as follows. In Section 2, the induction of FDTs using fuzzy ID3 is discussed. Section 3 provides the computational experimental results. Section 4 presents the concluding remarks.

2. Induction of Fuzzy Decision Tree

A. Pattern Classification

Classification problems assign patterns to classes. Class is considered to be a group of similar patterns. Each pattern is in the form of a set of attribute values, which are used as the basis for classification. For example,Table 1 provides the sample patterns from Thyroid Gland dataset [9].

Table 1. Sample Patterns from Thyroid Gland dataset.

T3-resin uptake test	**Total Serum thyroxin**	**Total serum triiodothyronine**	**basal thyroid-stimulating hormone (TSH)**	**Maximal absolute difference of TSH**	**Class label**
107	10.1	2.2	0.9	2.7	1
89	14.3	4.1	0.5	0.2	2
120	6.8	2.1	10.4	38.6	3

The row vector represents the pattern. The column vector is the attribute. The values of the attribute T3-resin uptake test are 107,89 and 120. There are five conditional attributes plus one decision attribute the "Class label". The conditional attributes can be numerical or categorical. The decision attribute is always considered as discrete in classification problem. Each of the patterns has to be classified into either of the three classes: 1 for normal, 2 for hyper, 3 for hypo.

B. Fuzzy Classification Problem

Given a set of n input-output training patterns $D=\left\{\left(x^i, y^i\right) \mid i=1,....,n\right\}$, where each training pattern x^i has been described by a set of p conditional (or input) attributes $\left(x_1,.....,x_p\right)$ and one corresponding discrete class label y^i where $y^{(i)} \in (1,.....,q)$ and q is the number of classes ,each attribute x_j is fuzzified into c_j fuzzy sets $\left\{F_{jk}; k=1,....,c_j\right\}$. $\mu_{F_{jk}}(x_j^{\,i})$ is the membership degree of the i^{th} value of attribute x_j on the fuzzy set F_{jk}. The decision attribute y^i represents a posteriori knowledge regarding class of each pattern. An arbitrary class has been indexed by l $(1 \le l \le q)$ and each class l has been modeled as a crisp set. The membership degree of the i^{th} value of the decision attribute (y^i) concerning the l^{th} class is defined as follows:

$$\mu_l(y^i) = \begin{Bmatrix} 1; & \text{if } y^i \text{ belongs to } l^{th} \text{ class} \\ 0; & \text{otherwise} \end{Bmatrix} \tag{1}$$

C. **Fuzzy ID3 [7]**

Fuzzy ID3 uses fuzzy classification entropy to determine the best root/internal node during FDT induction. For each fuzzy set $\{F_{jk} \mid j=1,...,p; k=1,...,c_j\}$ of the attribute x_j, certainty factor concerning the l^{th} class is defined as [7]

$$\beta_{jk}^{l} = \frac{\sum_{i=1}^{n} \min\{\mu_{F_{jk}}(x_j^i), \mu_l(y^i)\}}{\sum_{i=1}^{n} \mu_{F_{jk}}(x_j^i)}; 0 \le \beta_{jk}^{l} \le 1 \qquad (2)$$

The equation (2) is a subsethood measure which determines the degree to which all possible attribute values of fuzzy set F_{jk} belong to the class l.

The fuzzy classification entropy of F_{jk} is defined as[7]

$$Entr_{jk} = -\sum_{l=1}^{q} \beta_{jk}^{l} \times \log_2(\beta_{jk}^{l}) \qquad (3)$$

The averaged fuzzy classification entropy of x_j is defined as[7]

$$E_j = \sum_{k=1}^{c_j} w_{jk} \times Entr_{jk} \qquad (4)$$

Where w_{jk} denotes the weight of the k^{th} fuzzy set of the j^{th} attribute and is defined as[7]

$$w_{jk} = \frac{\sum_{i=1}^{n} \mu_{F_{jk}}(x_j^i)}{\sum_{k=1}^{c_j}(\sum_{i=1}^{n} \mu_{F_{jk}}(x_j^i))} \qquad (5)$$

The general procedure for generating fuzzy decision trees using Fuzzy ID3(Wang et al.,2001) is outlined as follows:

Prerequisites:

A Fuzzy Partition Space, leaf selection threshold β_{th}, best node selection criterion.

Procedure:

While there exist candidate nodes ***DO***

Select one of them using a search strategy,

Generate its child-nodes,

Child-nodes meeting the leaf threshold areleveled as leaf-nodes, otherwise the remaining child-nodes are regarded as new candidate nodes and the procedure is repeated until the stopping criterion is met.

end

Before training the initial data, the α -cut is usually used for the initial data to reduce the fuzziness.

The α -cut of a fuzzy set A is defined as:

$$\mu_{A_\alpha}(a) = \begin{cases} \mu_A(a); \mu_A(a) \geq \alpha \\ 0; \mu_A(a) \leq \alpha \end{cases} \tag{6}$$

Using the Fuzzy ID3 procedure described above a fuzzy decision tree is generated as shown in Figure.1 for Thyroid gland dataset available in UCI Machine learning repository [35].

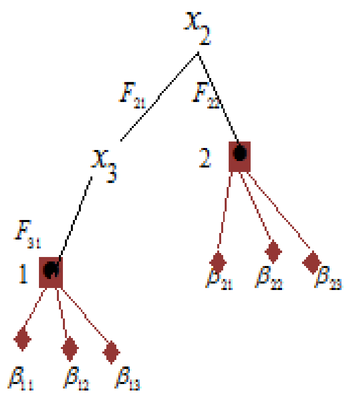

Figure 1. Fuzzy Decision Tree for Thyroid Gland.

The problem is to determine whether or not a patient has one of the three types of thyroid gland. There are three classes, five numerical attributes, and 215 patterns. The class label y_1 is for first type of thyroid gland(normal), y_2 is for second type(hyper), y_3 is for third type of thyroid gland(hypo).

The FCM generated fuzzy partitions are used for the induction of FDT. In the process of expanding a non-leaf node on the fuzzy decision tree, the Fuzzy ID3 algorithm selects the expanded attribute using Average Fuzzy Classification Entropy given in equation (4) and uses the certainty factor of a fuzzy cluster at a node as the leaf criterion. That is, the node will be regarded as a leaf node; if the certainty factor exceeds a given threshold (β_{th}). We have chosen threshold value of 0.75. The certainty factor of a Fuzzy Classification rule can be measured by fuzzy subset hood $\beta\ (F_V\ , l)$, which measures the degree to which F_V isa subset of class *l*.

$$\beta(F_V, l) = \frac{\sum_{i=1}^{n} \min\left\{\mu_{F_V}(x^i), \mu_l(y^i)\right\}}{\sum_{i=1}^{n} \mu_{F_V}(x^i)} \tag{7}$$

For $F_V = F_{1k} \cap .....\cap F_{pk}$, $\mu_{F_V}(x^i) = \min(\mu_{F_{1k}}(x_1^i), \mu_{F_{2k}}(x_2^i), ...\mu_{F_{pk}}(x_p^i))$

In Figure.1, Patterns are classified by starting with the root node and then by reaching to one or more than one leaf nodes by following the path of degree of memberships which are greater than zero.

A traversing path from root node x_2 to first leaf node is represented by:

$$path_1 = x_2 \text{ is } F_{21} \wedge x_3 \text{ is } F_{31}.$$

$$leaf_1 : y_1 = \text{normal}(\beta_{11}),\ y_2 = \text{hyper } 2(\beta_{12}),\ y_3 = \text{hyper } 2(\beta_{12}).$$

$\wedge$ is a fuzzy T-norm operator and it is taken as a *product* here.

To fuzzify input attributes, the Gaussian membership function is chosen due to its differentiable property. For i^{th} pattern, membership degree of $path_m$ can be calculated by

$$\mu^i path_m = \prod_j \mu F_{jm}(x_j^i) = \prod_j e^{\left(-\frac{(x_j^i - c_j^m)^2}{2\sigma_{jm}^2}\right)} \qquad (8)$$

Where c_{jm} and σ_{jm} are center and standard deviation (width) of Gaussian membership function of j^{th} input attribute on m^{th} path.In Figure.1, the membership degree of arbitrary pattern to *path₁* can be calculated by $\mu^i_{path_1} = \mu_{F_{21}}(x_2^i) \times \mu_{F_{31}}(x_3^i)$.$leaf_m$is the degree of certainty β_{ml} ($0 \leq \beta_{ml} \leq 1$; $l = 1, . . . , q$) with which $path_m$ can classify the class l. β_{11} is the degree of certainty with which all possible attribute values of $path_1$can belong to class 1 thyroid gland. β_{12} is the degree of certainty with which all possible attribute values of $path_1$can belong to class 2 thyroid gland. β_{13} is the degree of certainty with which all possible attribute values of $path_1$can belong to class 3 thyroid gland. Putting $path_m$ and $leaf_m$ together, it gives a weighted fuzzy classification rule of the form " $\textit{If } path_m \textit{ then } leaf_m$".

The fuzzy classification rules obtained from Figure 1 are:

$$\textit{if } (x_2 \textit{ is } F_{21}) \textit{ and } (x_3 \textit{ is } F_{31}) \textit{ then } y_1 = 0.097(\beta_{11}), y_2 = 0.9861(\beta_{12}) \textit{ and } y_3 = 0.0042(\beta_{13})$$
$$\textit{if } (x_2 \textit{ is } F_{22}) \textit{ then } y_1 = 0.8108(\beta_{21}), y_2 = 0.0269(\beta_{22}) \textit{ and } y_3 = 0.1623(\beta_{23})$$

To carry out the inference for novel test patterns, the certainty factors from all the leaf nodes corresponding to class 1 are summed up to obtain y_1. In same way, certainty factors corresponding to class 2 are summed up to obtain y_2 and class 3 are summed up to obtain y_3 as shown in Figure 2. The same procedure can be followed for q classes.

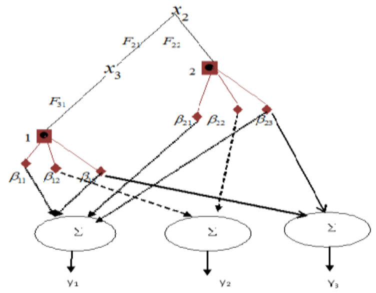

Figure 2. Fuzzy Decision Tree with prediction certainty

For an arbitrary pattern, $\mu^i path_m \times \beta_{ml}$ gives the firing strength of m^{th} class at m^{th} leaf node.

$$y_l^i = \sum_{m=1}^{M} \mu^i{}_{path_m} \times \beta_{ml} \qquad (9)$$

Equation (9) predict the degree of certainty for every class l. To determine the unique class for a pattern p, the winner-take all rule is applied where the class with maximum membership degree is selected, i.e., classify given pattern to class l^i, where

$$l^i = \max_{l=1,\ldots,q} \{y_l{}^i\} \qquad (10)$$

3. Computational Experiments

In this paper, we have presented a study on variants of FDT for predicting the class labels of ten medical datasets using fuzzy ID3 algorithm. The fuzzy partitioning of inputs are obtained using FCM algorithm.FCM generates cluster centers and degree of memberships of each pattern to each fuzzy cluster **[**8].The cluster centers are chosen to be the prototypes. FCM with the fuzzifier exponent value of 2 is used in our experiment. The datasets considered for our experiments from available in UCI repository [9] are given in Table 2.

Table-2. Dataset and its Characteristics.

Dataset	**Number of attributes**	**Number of patterns**	**No.of Labels**	**Type**
Cleveland Heart Disease	13	270	2	Continuous, Categorical, Discrete
Liver disorder	6	345	2	Continuous
Wisconsin BC	9	683	2	Continuous
Veteran lung cancer	7	137	2	Continuous, Categorical, Discrete
Stat log Heart Disease [13]	13	270	2	Continuous, Categorical, Discrete
Thoracic surgery	16	470	2	Continuous, Categorical
Mammographic Mass	5	830	2	Continuous, Categorical
Thyroid Gland	5	215	3	Continuous
Post-Operative Patient	8	87	3	Continuous
Pima Indian Diabetes	8	768	2	Continuous

During FDT induction, the continuous attributes are fuzzy partitioned into three sets and later approximated to Gaussian membership functions [6]. The categorical attributes are fuzzified into fuzzy singletons. The Alpha cutis chosen to be 0.05. Terminating a rule as a leaf is based on the leaf selection threshold (β_{th}) which is compared against the certainty factor of the rule obtained using equation 7.

If the obtained certainty factor is higher than leaf selection threshold then the path is considered for leaf node otherwise the tree is grown further. The leaf selection threshold (β_{th}) values of range between0.6 to 0.8 are chosen for our experiments. The different variants of FDT considered for our study are as given below. The first one is fuzzy version of ID3 which uses the measure called average fuzzy classification entropy of attribute(s). The second FDT variant is proposed by Yuan and Shaw [4] which uses the average fuzzy classification ambiguity of attribute(s) for FDT induction. The third variant we have given is by Yeung et al. [10] which uses the average degree of importance of attributes.

The comparison of these three attributes selection measures are given by Wang et al. [7]. The last two algorithms for the induction of fuzzy decision trees have been proposed by Rajen and Gopal [11-12].For each of the experiments conducted, we have used tenfold cross validation i.e., the whole dataset is divided in to ten parts with uniform class distribution. Nine out of ten parts are used for training and the remaining one part is used for testing. This process is repeated for 10 times in such way every part is used for testing and training. The final accuracies obtained are average of the tenfold cross validation. Table 3 and Table 4 show the computational experimental results for the test data.

Table-3. Classification Accuracy (%).

Sl. No	Dataset	Fuzzy ID3	Yuan and Shaw	Yeung, Wang and Tsang	FRCT	FRID ver1.1
1	Cleveland Heart Disease	80.26	69.73	57.23	76.09	77.63
2	Liver disorder	71.92	67.32	56.54	69.72	79.83
3	Wisconsin BC	90.03	79.82	88.01	90.93	89.03
4	Veteran lung cancer	92.34	91.23	90.12	91.56	91.32
5	Stat log Heart	89.16	84.43	87.61	88.76	84.32
6	Thoracic surgery	78.08	76.34	77.21	78.11	76.34
7	Mammographic Mass	70.65	67.52	66.34	72.54	71.56
8	Thyroid Gland	89.47	88.55	86.55	88.76	83.55
9	Post-Operative Patient	72.52	71.54	68.23	71.54	69.23
10.	Pima Indian Diabetes	72.39	65.12	72.39	71.48	72.39

Table 4. Number of Rules.

Sl. No	Dataset	Fuzzy ID3	Yuan and Shaw	Yeung, Wang and Tsang	FRCT	FRID ver1.1
1	Cleveland Heart Disease	30	1	3	15	6
2	Liver disorder	3	4	4	8	6

3	Wisconsin BC	3	2	2	5	2
4	Veteran lung cancer	3	1	1	3	2
5	Stat log Heart	5	6	6	5	7
6	Thoracic surgery	7	8	6	9	9
7	Mammographic Mass	3	2	2	4	6
8	Thyroid Gland	2	3	3	2	3
9	Post-Operative Patient	4	5	5	8	5
10.	Pima Indian Diabetes	3	3	3	20	3

4.

5.

6.

7.

8.

9. Conclusion

In this paper, we present a study on FDTs for classifying medical datasets. From the results obtained, we observe that Fuzzy ID3 algorithm followed by Fuzzy rough classification trees outperforms other techniques in obtaining better classification rates for most of the medical datasets. Intelligent systems with appropriate machine learning algorithms helps the doctors in medical field to automate the decision making process of diagnosing what disease the patient has and further it can be extended for choosing the appropriate drugs in automated way based on the predictions made. Such classifications schemes combined with doctor's expertise leads to monitored intelligent system that automates diagnosis and prescription of drugs for the patients. Further optimization techniques can be used for continuous attributes to further improve the algorithm results by tuning few parameters of FDT.

References

1. J. R. Quinlan, Induction of decision trees, Machine Learning (1) 81–106, (1986).
2. Quinlan JR. C4. 5: programs for machine learning. Morgan Kaufmann Publishers; 1993.
3. 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 Proc. IEEE Int. Conf. FuzzySystems, pp. 2113–2118,(1994).
4. Y. Yuan and M. J. Shaw, Induction of fuzzy decision trees, Fuzzy Sets and Systems 69 , 125–139(1995).
5. Rajen B. Bhatt, Swathi J. Narayanan, Ilango Paramasivam, Khalid M, Approximating Fuzzy Membership Functions from Clustered Raw Data , Proc. IEEE International Conference on Innovations in Social and Humanitarian Engineering(IEEE-INDICON 2012),pages 487-492,Dec 7-9,Kochi,Kerala (2012)
6. R. B. Bhatt and M. Gopal, Neuro-Fuzzy Decision Trees, Int. J. Neural Systems 16(1) 63-78, (2006).
7. X.-Z.Wang, D. S. Yeung and E. C. C. Tsang, A comparative study on heuristic algorithms for generating fuzzy decision trees, IEEE Transactions on SMC-B 31(2), 215–226, (2001).

8. L. Black and C. J. Merz, UCI Repository of Machine Learning Databases [http: //www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, Department of Information and Computer Science.
9. Yeung DS, Wang XZ, Tsang ECC, Learning weighted fuzzy rules from examples with mixed attributes by fuzzy decision trees. In: Proceedings of the IEEE international conference on SMC, Tokyo, Japan, October 12–15, 349–354, (1999).
10. Rajen B. Bhatt, Gopal M, FRID: Fuzzy-Rough Interactive Dichotomizers. In: Proceedings of the IEEE international conference on Fuzzy Systems, IEEE-FUZZ'04, Budapest, Hungary, July 26–29, 1337–1342, (2004).
11. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", NY: Plenum Press, (1982).
12. Bhatt, Rajen B., and M. Gopal, FRCT: fuzzy-rough classification trees, Pattern analysis and applications 11(1), 73-88, (2008).
13. Nozaki K, Ishibuchi H, Tanaka H, A simple but powerful heuristic method for generating fuzzy rules from numerical data. Fuzzy Sets Syst (86), 251–270, (1997).
14. http://lib.stat.cmu.edu/datasets/veteran

Corresponding author:
Dr. Swathi J Narayanan*,
Email: *swathi.jns@gmail.com*