Fuzzy Decision Tree Based Approach to Predict the Type of Pavement Repair

DEVINDER KAUR [a], HARICHARAN PULUGURTA [b]
[a]Electrical and Computer Science Engineering
[b]Civil Engineering
University of Toledo
2801 W Bancroft St
USA

Abstract: - Data mining is the process of extraction of hidden predictive information from large databases and expressing them in a simple and meaningful manner. This paper explains the use of Fuzzy logic as a data mining process to generate decision trees from a pavement (road) database containing historical pavement information. Generally there are many attributes in the pavement database and often it is a complicated process to develop any mathematical model to classify the data. This paper demonstrates the use of fuzzy logic to generate decision tree to classify the pavement data. The fuzzy decision tree is then converted to fuzzy rules. These fuzzy rules assist decision-making process for selecting a particular type of repair on a pavement based on its current condition. The fuzzy decision tree induction method used is based on minimizing the measure of classification ambiguity for different attributes. The model was developed and tested using the ODOT (Ohio Department of Transportation) data set.

Key-Words: - Classification, Pavement Management, Classification Ambiguity, fuzzy ID3

1 Introduction

Data mining is the process of extracting hidden information from large databases. Data mining models search databases for hidden patterns, finding classification and predictive information that experts miss because it lies outside their expectations. The classification and prediction problems, where the target attribute is respectively discrete (nominal) and continuous (numerical), are two main issues in data mining and machine learning fields. General methods for these two problems discover rules and models from a database of examples. IF … THEN … rules, neural nets, Bayesian nets, and decision trees are examples of such models. [2]. There are different ways of representing the patterns discovered by machine learning. Each one of the ways dictates the kind of technique to be used to generate that output. General methods of decision trees and classification rules are basic knowledge representation styles that machine learning methods use. Quinlan [2] with ID3 that stands for Interactive Dichotomizer 3 popularized the concept of decision trees. Systems based on this approach use an information theoretic measure of entropy for assessing the discriminatory power of each attribute [6].

The fusion of fuzzy sets with decision trees enables one to combine the uncertainty handling and approximate reasoning capabilities of the former with the comprehensibility and ease of application of the latter [1]. These models overcome the sharp boundary problems, providing soft controller surface and good accuracy dealing with continuous attributes and prediction problems. The information measure described by ID3 use to split a node is modified to introduce the Fuzzy concept. Sushmita et al. [1] discretize continuous attributes based on the distribution of pattern points in the feature space in linguistic terms using quantiles and use of fuzzy entropy and tree evaluation concept, in terms of compactness and performance. M. J. Kim et al. [3] describe hybrid knowledge integration mechanism using fuzzy genetic algorithm for the optimized integration of knowledge from several sources such as machine knowledge, expert knowledge and user knowledge. Baldwin and Xie [4] describe use of expected entropy and renormalized branch probability in modified fuzzy ID3 algorithm. Olaru and Wehenkel [7] introduce a new method of fuzzy decision trees called soft decision trees (SDT). This method combines tree growing and pruning, to determine the structure of the soft decision tree, with refitting and back fitting, to improve its generalization capabilities. Yuan and Shaw [2] induce a fuzzy decision tree by reducing classification ambiguity with fuzzy evidence. The input data is fuzzified using triangular membership functions around cluster centers obtained using Kohonen's feature map [2].

This study follows the approach proposed by Yuan and Shaw [2] and incorporation of fuzziness at the input by Sushmita et al. [1]. The system is then applied on the pavement management database. A pavement management database stores historical data about pavements (roads) of a network such as its present condition, past condition, geographical location, length, environmental conditions, etc and the attributes are both continuous and discrete. The main purpose of maintaining a pavement management database is to make informed decisions such as type of repairs to be performed on the pavements based on their condition. Given the number of attributes that are present in the pavement management database, it generally requires complex statistical models to model the data. In the current study, fuzzy logic is used to simplify the entire process of decision-making process. Simple rules are generated using fuzzy logic to assist in the decision making process.

2 Methodology

The Fuzzy logic used in the current study works by measuring the cognitive uncertainty. Cognitive uncertainty is the uncertainty that deals with phenomena arising from human thinking, reasoning from human thinking, reasoning, cognition and perception process, or cognitive information in general [2]. The cognitive uncertainty can be further classified into two subcategories: vagueness and ambiguity. Once the fuzzy sets are introduced, the cognitive uncertainties represented by fuzzy can therefore be measured.

2.1 The Measure of Vagueness

The vagueness or fuzziness of a fuzzy set can be measured by fuzzy entropy [2]. Let A denote a fuzzy set on the universe U with member ship function $\mu_A(u)$ for all $u \in U$. If U is a discrete set $U = \{u_1, u_2, ..., u_n\}$ and $\mu_i = \mu_A(u_i)$, the vagueness or the fuzziness of fuzzy set A is defined by

$$E_V(A) = -\frac{1}{m}\sum_{i=1}^{m}(\mu_i \ln\mu_i + (1-\mu_i)\ln(1-\mu_i)) \qquad (1)$$

where $E_V(A)$ measures the fuzziness or vagueness of a fuzzy set A. The degree of fuzziness expresses the average amount of ambiguity in taking a decision as to whether an element belongs to the set.

2.2 The Measure of Ambiguity

A fuzzy membership function $\mu(x)$ of a variable Y defined on X can also be interpreted as the possibility of taking value x for Y among all elements in X [2]. In this case $\pi(x) = \mu(x)$ for all $x \in X$, can be viewed as a possibility distribution of Y on X. The possibilistic measure of ambiguity or non-specificity is defined as

$$E_a(Y) = g(\pi) = \sum_{i=1}^{m}(\pi_i^* - \pi_{i+1}^*)\ln(i), \qquad (2)$$

where $\pi^* = \{\pi_1^*, \pi_2^*, ..., \pi_n^*\}$ is the permutation of the possibility distribution $\pi = \{\pi(x_1), \pi(x_2), ..., \pi(x_n)\}$, sorted so that $\pi_i^* \geq \pi_{i+1}^*$ for all $i = 1, ..., n$, and $\pi_{n+1}^* = 0$. To measure the ambiguity (overlapping) of an attribute A among its linguistic terms $T(A) = \{T_1, T_2, ..., T_n\}$, [2] interpret the membership functions $\{\mu_{T_1}(u_i), \mu_{T_2}(u_i), ..., \mu_{T_n}(u_i)\}$ as a possibility distribution for object u_i to take linguistic term on term label space $T(A) = \{T_1, T_2, ..., T_n\}$. To normalize the possibility distribution, let

$$\pi_{T_S}(u_i) = \mu_{T_S}(u_i) / Max_{1\leq j \leq S}\{\mu_{T_j}(u_i)\}, s = 1, ..., S. \qquad (3)$$

The ambiguity of the attribute A for object u_i therefore can be measured by [2]

$$E_a(A(u_i)) = g(\pi_T(u_i)). \qquad (4)$$

The ambiguity of attribute A then is

$$E_a(A) = \sum_{i=1}^{m} w(E_a(A(u_i))E_a(A(u_i)) \qquad (5)$$

Where $w(E_a(A(u_i))$ is the weight which represents the relative size. The ambiguity of classes can be measured in same way as attributes.

2.3 Classification Ambiguity

Knowing single evidence, such as a particular value of an attribute, the classification ambiguity can be defined [2] as follows:

$$G(E) = g(\pi(C/E)), \qquad (6)$$

which is measured on the possibility distribution of $\pi(C/E)$ which is defined as [2]

$$\pi(C_i/E) = S(E, C_i) / Max_j S(E, Cj), \qquad (7)$$

where $S(E, C_i)$ represents the *degree of truth* for the classification rule "IF E THEN C_i ", and $\pi(C/E) = \{\pi(C_i/E), i = 1,...,L\}$ is a normalized possibility distribution on the no fuzzy label space $C = \{C_1, C_2,..., C_L\}$. Given a fuzzy evidence F and a set of fuzzy evidences $P = \{E_1, E_2,..., E_K\}$ defined on object space U, the fuzzy partition P on F is defined [2] as $P/F = \{E_1 \cap F,..., E_K \cap F\}$, where each object defined in F is partitioned to E_i with membership $\mu_{E_i \cap F}$. The classification ambiguity of fuzzy partition can be defined as follows [2]:

$$G(P/F) = \sum_{i=1}^{K} w(E_i/F) G(E_i \cap F) \tag{8}$$

where $G(E_i \cap F)$ is the classification ambiguity with fuzzy evidence $E_i \cap F$, $w(E_i/F)$ is the weight which represents the relative size of subset $E_i \cap F$ in F

$$w(E_i/F) = M(E_i \cap F) / \sum_{j=1}^{K} M(E_j \cap F) \tag{9}$$

Significant level [2] α for a fuzzy evidence E with membership $\mu_E(u)$, is defined as

$$\mu_{E\alpha}(u) = \begin{cases} \mu_E(u) & \text{if } \mu_E(u) \geq \alpha \\ 0 & \text{if } \mu_E(u) \leq \alpha \end{cases} \tag{10}$$

i.e., if the membership value of an attribute is less than α, it is not considered for the analysis.

3 Induction of Fuzzy Decision Tree

Yuan and Shaw [2] construct fuzzy decision trees by reducing classification ambiguity with accumulated fuzzy evidences where fuzzy evidence is the knowledge about a particular attribute. The selection of fuzzy evidence is based on its contribution in reducing the classification ambiguity. The method is similar to the non-fuzzy decision tree induction method such as ID3. The fuzzy decision tree induction process suggested in [2] consists of following steps:

(1) Fuzzifying the training data
(2) Inducing the fuzzy decision tree
(3) Converting the decision tree into a set of rules
(4) Applying fuzzy rules for classification

3.1 Fuzzifying the Training Data

Any input feature value is described in terms of some combination of overlapping membership values in the linguistic property sets *low* (L), *medium* (M) and *high* (H). When input feature is numerical [1] divide it into three partitions (with range [0,1]) using only two parameters P_{j1} and P_{j2}. Let F_{jMax} and F_{jMin} denote maximum and minimum values encountered along feature F_j. The value of P_{j1} is the value of F_j that exceeds one-third of the measurements and less than two-thirds. The value of second quantile P_{j2} is the value of F_j that exceeds two-third of the measurements and less than remaining one-thirds.

3.2 Inducing Fuzzy Decision Tree

With given evidence significant level and truth level induction process consists of following steps [2]:
1) Select the attribute with the smallest classification ambiguity as root node
2) Delete all empty branches of decision node. For each non-empty branch of decision node, terminate the branch as leaf if the truth level of classifying into one class is above a given threshold β. Otherwise, investigate if an additional attribute will further partition the branch and further reduce the classification ambiguity. If yes, select the attribute with smallest classification ambiguity as a new decision node from the branch. If not, terminate this branch as leaf.
3) Repeat step 2 for all newly generated decision nodes until no further growth is possible, the decision tree is complete.

The input attributes considered for generating a fuzzy decision tree in this study are shown in Table 1. The meanings of each attribute are described following the table.

Table 1. Fuzzy Attributes

Attribute	Type	Representation
PCR	Numerical	0 – 100
HCS	Numerical	>0
AvgADT, AvgTADT	Numerical	> 0
Functional Class	Categorical	1,2,6,7,8,9,11,12,14, 16,17,19
Code 1-15	Categorical	LO,LF,LE,MO,MF, ME,HO,HF,HE, NO(NULL)
Activity Code-1	Categorical	10,20,30,35,40,45,50, 55,60,70,77,90,95, 100,110,120
Activity Code	Categorical	10,20,30,35,40,45,50, 55,60,70,77,90,95, 100,110,120

The overall pavement condition is represented by *PCR* (Pavement Condition Rating) 0 represents the worst pavement condition while 100 the perfect condition. PCR is calculated from 15 different variables (*Code 1- Code 15*) called distresses or observable faults on a pavement, which are in turn defined by different categories (Table 1). For example, a code category of "LO" indicates that a particular severity of a particular distress is **L**ow on the pavement and it is seen only **O**ccasionally. Half Car Simulation (*HCS*) represents the pavement ride condition (rough to smooth). Traffic data is defined by two different attributes that are *ADT* and *ATDT* (average daily traffic and average truck daily traffic). Functional class represents the type of road (for example, 1=freeway/interstate, 9=local roads etc). *Activity Code-1* is the previous treatment that was performed while *Activity code* represents the repairs performed on the pavements. Activity code and Activity code-1 are categorical where 10 to 45 represent maintenance activities (small repairs to the pavement), 50 to 60 represent minor activities (minor repairs on the pavements) and 70 to 120 represent major activities (reconstruction of the pavement). Table 2 shows a sample of the data set used in this research. The data are divided into training set and test set (ensuring that the training set contains all the attribute ranges found in the data set).

The significance level and degree of truth are considered as 0.5 and 0.6. Significance level of 0.5 indicates the membership values less than 0.5 are not considered. If the truth level for an attribute at a branch exceeds 0.6, it becomes leaf [2]. A small training set of 329 cases is selected to generate the decision tree. The decision tree is then applied to test data. The process of generating decision tree is done through a user as shown in Fig. 1.

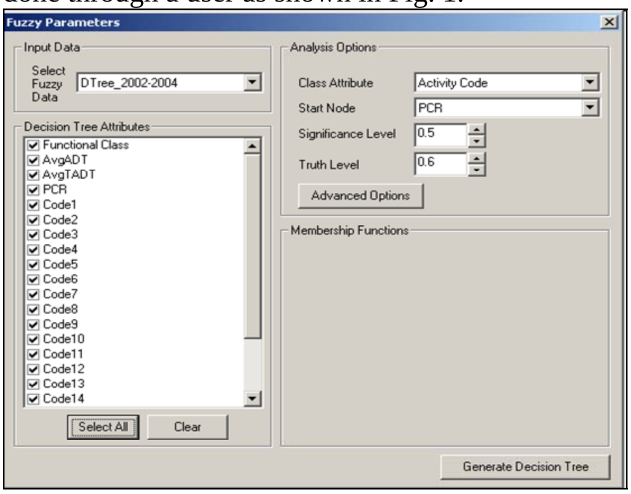

Fig. 1. User Interface to generate decision trees

The user interface provides an option to force the decision tree with a selected root node. In the present study, PCR represents the overall pavement condition and hence PCR is forced as root node incase the Fuzzy ID3 algorithm does not identify it as a root node. Fig. 2 shows the decision tree generated with significance level of 0.5 and truth level of 0.6. Each path of the branches from root to leaf can be converted into a rule with condition part represents the attributes on the passing branches from root to the leaf and the conclusion part represents the class at the leaf with the highest classification truth level [2]. Fig. 3 shows the rules generated from the decision tree.

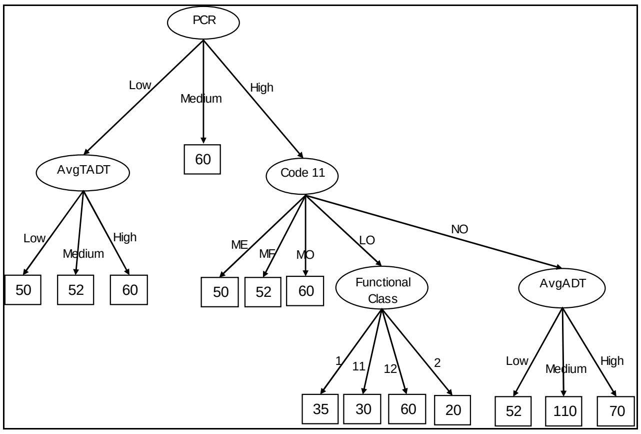

Fig. 2. Fuzzy Decision Tree

Each path of the branches from root to leaf can be converted into a rule with condition part represents the attributes on the passing branches from root to the leaf and the conclusion part represents the class at the leaf with the highest classification truth level [2]. Fig. 3 shows the 14 rules from the decision tree.

R1 IF (PCR=Low AND AvgTADT=Low) THEN Class=50
R2 IF (PCR=Low AND AvgTADT=Med) THEN Class=52
R3 IF (PCR=Low AND AvgTADT=High) THEN Class=60

R4 IF (PCR=Med) THEN Class=60

R5 IF (PCR=High AND Code11=ME) THEN Class=60
R6 IF (PCR=High AND Code11=MF) THEN Class=60
R7 IF (PCR=High AND Code11=MO) THEN Class=30
R8 IF (PCR=High AND Code11=LO AND Functional Class=1) THEN Class=35
R9 IF (PCR=High AND Code11=LO AND Functional Class=11) THEN Class=60
R10 IF (PCR=High AND Code11=LO AND Functional Class=12) THEN Class=60
R11 IF (PCR=High AND Code11=LO AND Functional Class=2) THEN Class=20
R12 IF (PCR=High AND Code11=NO AND AvgADT=Low) THEN Class=52
R13 IF (PCR=High AND Code11=NO AND AvgADT=Med) THEN Class=110
R14 IF (PCR=High AND Code11=NO AND AvgADT=High) THEN Class=70

Fig. 3. Fuzzy Rules

With the classification rules generated form the decision tree, classification results when applied on a small test set. For each rule, the membership of the condition is calculated for the object based on its attributes.

Table 2. Data Sample

Pave Section	Functional Class	Avg ADT	Avg TADT	PCR	Code							HCS	STRD	Activity Code-1	Class
					1	2	3	...	13	14	15				
1	1	42310	14380	62	LE	NO	MF		MO	E	NO	86	21.07	0	100
2	12	31250	1660	76	ME	NO	NO		LO	E	NO	113	9.04	0	100
3	12	31250	1660	67	ME	NO	HF		NO	E	NO	153	14.59	0	100
4	2	29150	3180	84	LF	NO	LO		NO	O	NO	93	10.11	0	20
5	2	29150	3180	85	LF	NO	NO		NO	F	NO	0	8.24	0	20
6	11	84590	12710	80	LE	NO	LO		LO	O	NO	73	9.12	60	30
7	11	84590	12710	80	LE	NO	LO		LO	O	NO	73	9.12	60	30
8	1	35050	13640	84	LF	NO	NO		NO	F	NO	47	6.96	0	35
.															
.															
.															
322	11	32870	5010	70	MF	NO	HO		MO	E	NO	62	11.6	0	52
323	11	32870	5010	70	MF	NO	HO		MO	E	NO	62	11.6	0	52
324	12	68540	3860	72	ME	NO	MO		NO	E	NO	110	10.2	0	52
325	12	68540	3860	65	ME	NO	MO		NO	E	NO	80	17.92	0	52
326	12	19890	1370	62	MF	NO	HO		LO	E	NO	176	19.84	0	52
327	12	19890	1370	62	MF	NO	HO		LO	E	NO	176	19.84	77	52
328	12	19890	1370	62	MF	NO	HO		LO	E	NO	120	19.84	77	52
329	11	134200	14130	69	MO	NO	HO		NO	E	NO	112	16	0	70

The membership of conclusion (classification of each class) is set equal to the membership of condition [2]. The rule with maximum membership value is considered as correct rule and the object is assigned to the class of the outcome of the rule. Table 3 shows the membership for each rule and also the actual class (Activity Code) for each object.

Table 3. Rule Strengths

Object	Rules membership when applied on test data														Actual Classification of Test Data
	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	
1	0.0	0.0	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**0.6**	70
2	0.0	0.0	0.2	**0.7**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	70
3	0.3	0.0	0.0	**1.0**	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	60
4	0.0	0.0	0.4	**0.7**	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	60
5	0.1	0.0	0.0	0.2	0.0	0.0	**0.6**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**0.8**	0.0	0.0	52
7	**1.0**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50
8	**0.7**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**1.0**	0.0	0.0	0.0	20
10	0.0	0.0	0.0	0.0	0.0	0.0	**0.9**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	**0.9**	0.4	110
12	0.0	0.0	0.0	0.0	0.0	0.0	**1.0**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	110

Consider object 6 from Table 3, according to maximum rule strength, Rule 12 is chosen and Rule 12 predicts the class as 52 same as the actual class. Finally, out of the 12 pavements that are shown in Table 3, 9 cases are predicted correctly.

4 Conclusion

Pavement management database consists of many different attributes that are both continuous and categorical in nature. It is often required in pavement management to determine the type of repair needed for a pavement. This decision is based on the condition of the pavement whether it is in good condition or fair condition and also with respect to different attributes such as traffic, weather conditions etc. It is a complicated process to develop a statistical model based on all these attributes. In this study a more straightforward approach is used and is demonstrated using actual data. A fuzzy decision tree is generated which is then converted to simple rules. The rules are then tested on test data set of 12 pavements. These 12 pavements were not used to develop the decision tree. The results showed that the fuzzy rules accurately predicted treatment types for 9 pavements out of 12 pavements.

5 Future Work

Varying the values of significance level and truth level will generate different rules. For example, Figure 4 shows the fuzzy rules when the truth level used is 0.9. Sushmita *et al.* [1] use a performance measure called T-Measure to select right sized tree. Breiman *et al.* [9] suggest a two step approach to select right sized tree. Grow a tree and prune it upward and calculate misclassification cost to select the right sized tree. Yuan and Shaw [2] simplify the rules by removing one attribute term at a time from the IF Part. If the truth level of the new rule is not lower then the truth level threshold or the truth level of the original rule, the simplification is successful.

1. IF (PCR=Low AND AvgTADT=Low) THEN Class=50
2. IF (PCR=Low AND AvgTADT=Med) THEN Class=52
3. IF (PCR=Low AND AvgTADT=High) THEN Class=60

4. IF (PCR=Med AND Code11=HO) THEN Class=60
5. IF (PCR=Med AND Code11=MO) THEN Class=60
6. IF (PCR=Med AND Code11=LO AND Activity Code-1=0) THEN Class=100
7. IF (PCR=Med AND Code11=LO AND Activity Code-1=73) THEN Class=60
8. IF (PCR=Med AND Code11=NO AND Code1=HO) THEN Class=52
9. IF (PCR=Med AND Code11=NO AND Code1=LE) THEN Class=70
10. IF (PCR=Med AND Code11=NO AND Code1=ME) THEN Class=52
11. IF (PCR=Med AND Code11=NO AND Code1=MF) THEN Class=52
12. IF (PCR=Med AND Code11=NO AND Code1=MO) THEN Class=70

13. IF (PCR=High AND Code11=ME) THEN Class=60
14. IF (PCR=High AND Code11=MF) THEN Class=60
15. IF (PCR=High AND Code11=LO AND Functional Class=11) THEN Class=30
16. IF (PCR=High AND Code11=LO AND Functional Class=2) THEN Class=20
17. IF (PCR=High AND Code11=MO AND Code3=LO) THEN Class=30
18. IF (PCR=High AND Code11=MO AND Code3=MO) THEN Class=30
19. IF (PCR=High AND Code11=MO AND Code3=NO) THEN Class=50
20. IF (PCR=High AND Code11=NO AND AvgADT=High) THEN Class=70
21. IF (PCR=High AND Code11=NO AND AvgADT=Low AND Code10=HF) THEN Class=52
22. IF (PCR=High AND Code11=NO AND AvgADT=Low AND Code10=MO) THEN Class=50
23. IF (PCR=High AND Code11=NO AND AvgADT=Low AND Code10=NO) THEN Class=52
24. IF (PCR=High AND Code11=NO AND AvgADT=Med AND Activity Code-1=0) THEN Class=52
25. IF (PCR=High AND Code11=NO AND AvgADT=Med AND Activity Code-1=50) THEN Class=110
26. IF (PCR=High AND Code11=NO AND AvgADT=Low AND Code10=ME AND Code13=LO) THEN Class=52
27. IF (PCR=High AND Code11=NO AND AvgADT=Low AND Code10=ME AND Code13=NO) THEN Class=31
28. IF (PCR=High AND Code11=NO AND AvgADT=Low AND Code10=MF AND Code14=O) THEN Class=20

Fig. 4. Fuzzy Rules (Truth Level =0.7)

The next part of the paper involves in optimizing the values of significance level and truth level for an optimized sized tree and clear rules.

References:

[1] S. Mitra , K. M. Konwar, and S. K. Pal, “Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-Based Network: Generation and Evaluation,” in *IEEE Trans. Systems Man Cybernet* , vol. 32, Nov. 2002, pp. 328-339.

[2] Y. Yuan and M. J. Shaw, “Induction of fuzzy decision trees,” *Fuzzy Sets Syst.*, vol. 69, pp. 125–139, 1995.

[3] H. J. Kim, I. Han, and K. C. Lee, “Hybrid Knowledge Integration Using the Fuzzy Genetic Algorithm: Prediction of the Korea Stock Price Index,” in *Intell. Sys. Acc. Fin. Mgmt.* 12: 43-60 (2004).

[4] J. F. Baldwin and D. W. Xie, “Simple Fuzzy Logic Rules Based on Fuzzy Decision Tree for Classification and Prediction Problem,” in Intelligent Information Processing (IIP) 2004 Available: http://eis.bris.ac.uk/~enxdx/files/ICIIPBaldwinandXie.pdf

[5] S. K. Pal and S. Mitra, *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*. New York: Wiley, 1999, ch. 2.

[6] J. R. Quinlan, “Induction on decision trees,” *Machine Learning*, vol. 1, pp. 81–106, 1986.

[7] C. Olaru and L. Wehenkel, “A Complete Fuzzy Decision Tree Technique,” *Fuzzy Sets Syst.*, vol. 138, pp. 221–254, 2003.

[8] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.

[9] L. Breiman and J. M. Friedman and R. A. Olshen and C. J. Stone, *Classification and Regression Trees*. Belmont, California: Wadsworth International Group, 1984, ch. 3.