Universal Journal of Computer Science and Engineering Technology
1 (1), 55-63, Oct. 2010.
© 2010 UniCSE, ISSN: 2219-2158

Effective Method for Extracting Rules from Fuzzy Decision Trees based on Ambiguity and Classifiability

Hesham A. Hefny
Dept. of Computer and Information Sciences, Institute of Statistics, Cairo University, Egypt.
hehefny@hotmail.com

Ahmed S. Ghiduk
Dept. of Mathematics, Faculty of Science, Beni-Suef University, Beni-Suef, Egypt.
asaghiduk@yahoo.com

Ashraf Abdel Wahab
Dept. of Computer and Systems, Electronics Research Institute, Cairo, Egypt.
awahab@mcit.gov.eg

Mohammed Elashiry
Dept. of Management, Academy of Specialized Studies, Beni-Suef, Egypt.
ashiry@aun.edu.eg

Abstract—Crisp Decision trees (*CDT*) algorithms have been the most widely employed methodologies for symbolic knowledge acquisition. There are many methodologies have been presented to address the problems of the continuous data, multi-valued data, missing data, uncertainty data and noisy features. Recently, due to the widespread use of the fuzzy representation, a lot of researchers have utilized the fuzzy representation in decision trees to overcome the preceding problems. Fuzzy decision trees (*FDT*) are generalization for the *CDT*. *FDTs* are built by using fuzzy or crisp attributes and classes which often need pruning to reduce their size. *FDTs* have been successfully used to extract knowledge in uncertain classification problems. In this paper, we present a technique to build FDT by employing the ambiguity of attributes and classifiability of instance. Our technique builds a reduced *FDT* which does not need for applying the pruning algorithms to reduce the size. The paper also presents the results of a set of empirical studies conducted on a dataset of *UCI* Repository of Machine Learning Database that evaluate the effectiveness of our technique compared to Fussy Iterative Dichotomiser 3 (*FID3*), ambiguity, and *FID3* with classifiability techniques. The studies show the effective of our technique in reducing the number of the extracted rules without loosing of the rules accuracy.

Keywords-Fuzzy decision tree; Fuzzy entropy; Fuzzy Ambiguity; Fuzzy rules; Classifiability of Instances.

I. INTRODUCTION

FDT is an extension to the crisp decision tree which is based on fuzzy valued attributes and classes. *FDT* has the ability for dealing with ambiguity and vagueness attributes associated with human thinking as well as crisp attributes.

A lot of algorithms have been proposed for producing small size decision trees [1,4,17,18]. Some of these algorithms use different criteria such as: Fuzzy Entropy (Information Gain), Fuzzy Gain Ratio and Fuzzy Ambiguity to split the instances at each node [4,17,18] while the other algorithms such as classifiability of instances are based on pruning the decision tree [1,17].

Both of Fuzzy Information Gain (*FIG*) and Fuzzy Gain Ratio (*FGR*) algorithms use the concept of entropy to reduce the depth of the fuzzy decision tree by determining the quantitative value of the uncertainty in each attribute of a data set. FIG and FGR are suitable for originally categorical data. The drawbacks nodes within the hidden level are generated until the fuzzy entropy is reduced to zero [3,4,5,6, 7,9,10,11,12,13,16,17].

Fuzzy ambiguity algorithm is based on the possibility to find the ambiguity in each attribute. Fuzzy ambiguity algorithm is suitable for numerical and categorical data. Fuzzy ambiguity algorithm is more accurate than fuzzy entropy algorithm in finding the uncertainty attribute, directly measure the quality of classification rule in decision node, significant level which will affect the generation of *FDT* [4, 11, 12, 16, 17, 18].

Fuzzy Information Gain, Fuzzy Gain Ratio, and Fuzzy ambiguity select and isolate a single attribute for building the decision tree without considering the other related attributes. Thus, the decision tree needs to burn for reducing the number of the extracting rules[3,4,5,7,9,10,11,12,13,14,16,18, 22].

Classifiability algorithm selects an attribute with considering the instances in the other attributes. Classifiability algorithm is considered a decision tree pruning technique. Therefore, it improves the performance of decision making, accurate prediction, efficiency and comprehensibility of the generated useful rules to electric power companies [1, 6].

Finally there are two stages to extracting rules from dataset, firstly create Fuzzy Decision Tree, secondly pruning a tree, our proposed will be reduced the steps also, reduce tree size and enhanced accuracy with using an algorithm.

Corresponding Author: Hesham A. Hefny, Dept. of Computer and Information Sciences, Cairo University, Egypt.

All the previous techniques contain two stages for extracting rules from a dataset. The first stage creates fuzzy decision tree. The second stage prunes the fuzzy decision tree to reduce the size of the tree.

In this paper, we present a technique to build FDT by employing the ambiguity of attributes and classifiability of instance. Our technique builds a reduced FDT which does not need for applying pruning algorithms to reduce the size. The paper also presents the results of a set of empirical studies conducted on a dataset of UCI Repository of Machine Learning Database that evaluate the effectiveness of our technique compared to Fussy Iterative Dichotomiser 3 (FID3), ambiguity, and FID3 with classifiability techniques. The studies show the effective of our technique in reducing the number of the extracted rules without loosing of the rules accuracy.

The rest of the paper is organized as follow. Section II gives some basic concepts and definitions. Section III introduces a number of the problems of the control dependencies based fitness function and presents two schemes and the key ingredients to overcome these problems. Section IV provides the related work. Section V gives conclusions and future work.

II. BASIC CONCEPTS

This section gives some basic concepts. There are some of heuristic proposed to generate roles next subsection displays some of them as the following:

A. Information Gain

The earlier version of *ID3* which is based on minimum information entropy to select expand attributes. This heuristic is an improved heuristic of famous *ID3*, information gain used the machine learning using decision tree in calculating significance of attributes [11, 8, 7, 9, 15].

Consider a non-leaf node *S* consisting of *n* attributes $A^{(0)}$,...,$A^{(n)}$ to be selected. For each $k\ (1 \le K \le n)$, the attribute $A^{(K)}$ takes m_k values of fuzzy subsets, $A_1^{(K)}$,...... $A_m^{(k)}$ and the fuzzy classification is $A^{(C)}$.

The averaged fuzzy classification entropy of the k-th attribute is defined as:

$$E_k = \sum_{i=1}^{m_k} \left(\frac{M(A_i^{(K)})}{\sum_{j=1}^{m_k} M(A_j^k)} \right) Entr_i^k \qquad (1)$$

where, $Entr_i^{(k)} = -\Sigma_{j=1,mc} P_{ij}^{(k)} \log_2 P_{ij}^{(k)};\ 1 \le k \le n;\ 1 \le i \le m_k;\ 1 \le j \le m_C$.

Fuzzy *ID3* heuristic aims to search for an attribute such that its averaged fuzzy classification entropy attains minimum, i.e. selecting such an integer k_0 (the k_0-th attribute) that $E_{ko} = \mathrm{Min}_{i<k<n}(E_k)$.

B. Minimum Classification ambiguity

Instead of using minimum fuzzy entropy [18, 11, 4, 12, 2] this heuristic uses the minimum classification ambiguity to select expanded attributes. The classification ambiguity with fuzzy attribute $A^{(k)}$ is

$$G(A^k) = \sum_{i=1}^{mk} \left(\frac{\mu(A_i^k)}{\sum_{j=1}^{mk} \mu(A_j^k)} \right) G(A_i^k) \qquad (2)$$

$G(A_m^{(k)}) = g\ (\ \Pi\ (C \mid A_m^{(k)}))$, which is measured based on the Possibility $\Pi\ (C|A_m^{(k)})$.

C. Compare between Fuzzy -ID3 and Fuzzy Ambiguity

1) Properties Fuzzy ID3

- Entropy determines the quantitative value of the uncertainty carried by each attribute in a data set but not the whole data set [6, 19].
- The nodes within the hidden level are generated until the fuzzy entropy is reduced to zero [18].
- That heuristic attempts to reduce the average depth of tree [5,3].
- $E(A^k_i) = -\Sigma_{j=1,mc} P^k_{ij} \log_2 P^k_{ij}$ and minimum fuzzy entropy is selected [3].
- If $P^k_{ij} = 0$ or 1 then Entropy(u) = 0 [18].
- Method is more suitable for originally categorical data.

2) Properties Fuzzy Ambiguity

- Directly measure the quality of classification rule in decision node [18].
- An option of significant level which will effect the generation of FDT [3].
- That heuristic attempts to reduce the average depth of tree [3].

$$G(A_i^k) = \sum_{s=1}^{n_c} \left(\overset{*}{\underset{s}{\Pi}} - \overset{*}{\underset{s+1}{\Pi}} \right) \ln s$$

- Minimum Fuzzy ambiguity implies to minimum fuzzy entropy [3, 4, 12].
- $\Pi(x) = 1$ mean x is fully possible and $\Pi(x) = 0$ mean x is impossible

  $\Pi^*_2 = 0 \Rightarrow$ Ambig=0
  $\Pi^*_n = 1 \Rightarrow$ Ambig=Ln(n)

  Minimum Fuzzy ambiguity Implies to Minimum Fuzzy Entropy
- Method is more suitable for originally numerical data.

D. Classifiability of attribute by instances

Decision tree pruning is useful in improving the generalization performance of decision trees. The criteria is based on Classifiability measure , that considers the number of pattern instances of different classes at node and spatial distribution or these instances to estimate the effect of further splitting the node [6] .

L(k) is measure of Classifiability of attribute K within instances given by:

$$L^{(k)} = \sum_{i=1}^{C} w_{ii}^{(k)} - \sum_{i=1}^{C} \sum_{\substack{i=1 \\ j \ne i}}^{C} w_{ii}^{(k)} \qquad (3)$$

Since W is Local co-occurrence matrix after attribute k is selected calculated as:

$$W^{(K)} = \sum_{i=1}^{m_k} \sum_{x} P(x) \qquad (4)$$

where x is any instance in the i-th child node of current node.

P(X) is Local co-occurrence matrix for instance x. It capture the distribution of instances around a specific instance, the size of Local co-occurrence matrix is C x C since C is number of classes, as in eq.(5)

$$P(x) = \sum_{y, D_{xy} \le r} \mu(x)^T \mu(y). \qquad (5)$$

$p_{ij} = \Sigma_{y, D_{xy} \le r}\, \mu_i(x)\, \mu_j(y)$ of matrix P show the number of class j instance that are within the neighborhood r of instance x when instance x belongs to class I within membership $\mu_i(x)$

Distance between instances defined by

$$D_{xy} = \sum_{k=1}^{n} \sum_{i=1}^{m_k} \left| \mu_i^{(k)}(x) - \mu_i^{(k)}(y) \right| \qquad (6)$$

For any instance in the data set, we can find those instances that are within a circular neighborhood r, Membership values of instance x for classes $\mu(x) = [\mu 1(x), \ , \mu(x)]$.

For example P(X=1) for A12 (sub attribute 2 of attribute 1), mean Local co-occurrence matrix for instance X=1, remove all instances have zero value in A12, find all distance between instance X=1 and X={1,…, N}, choose instances with neighborhood less than or equal r of instance X=1 and sum results of multiply vectors which represent membership of classes of instance X=1 and neighborhood instance. Repeat that for all instances and attributes.

J (k) = any measure of classification – L (k) (7)

III. THE PROPOSED FUZZY DECISION TREES CLASSIFIABILITY OF INSTANCES AND AMBIGUITY OF ATTRIBUTES

Aim to build tree with both of the ambiguity of attributes and Classifiability of instances.

A. Proposed of heuristic fuzzy ambiguity with classifiability "AMCL"

Through following levels fuzzy decision tree was creating as:

Level (-1):

1- Consider a non-leaf node S consisting of n attributes $A^{(0)}, \dots, A^{(n)}$ to be selected. For each k ($1 \le K \le n$), the attribute $A^{(K)}$ takes m_k values of fuzzy subsets, $A_1^{(K)}, \dots\dots A_m^{(k)}$

Level (0):

1-For all Attribute $A^{(K)}$ calculate $G(A^{(K)})$ and $L(A^{(K)})$,$1 \le K \le n$ them compute objective function $J(A^{(K)}) = G(A^{(K)}) - L(A^{(K)})$

2- The root of decision is $\min\{J(A^{(K)})\}$ root is $A_{root} = \mathrm{MIN}\{ J(A^{(K)})\}$

Level (1): Delete all empty branches of the decision node. For each nonempty branch of the decision node calculate the truth level of classifying all objects within the branch into each class as a leaf, according to the following:

1- Attribute $A_{root} = (A_{root1}, \dots\dots, A_{root\,m})$ and classes $C = (C_1, C_2, \dots, C_w)$

2- Calculate the classification truth level $S(A_{root1}, C_j)$classes C_j , $1 \le j \le w$

3- **IF** $S(A_{root\,i}, C_j)$,$1 \le i \le m, 1 \le j \le w$, greater than truth level β **THEN** $A_{root\,i}$ become a leaf with label C_j

Else Partitioning the branch $A_{root\ i}$ with different attributes according to the following:

I-Compute $J(A_{root\,i})$, $J(A^K \mid A_{root\,i})$, where $1 \le k \le n$ and $K \ne$ root for each K

II-**IF** $J(A^K | A_{root\,i}) \le J(A_{root\,i})$

THEN A^K is branch from $A_{root\,i}$

Else $A_{root\,i}$ leaf with label that have greater value of S (Aroot i , Cj)

Level (2) : Repeat steps of level (1) to create level (2) of tree and so .

B. Flow Chart of Proposed Fuzzy Decision Trees

The algorithm aims to generate fuzzy decision trees based on minimum objective function, see equation (7), of each attribute in dataset. The objective function J measures pure ambiguity of attributes by eliminating classifiability of attributes from ambiguity of attributes. Figure 3.1 shows the flow chart of the AMCL proposed algorithm.

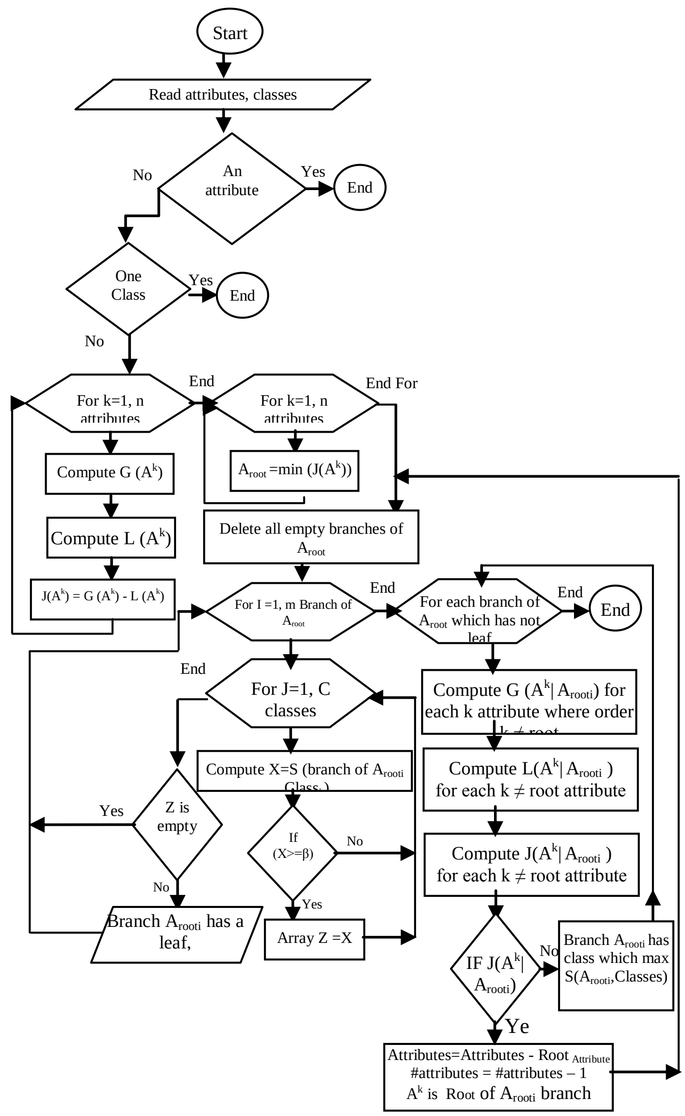

Figure 3.1: Flow chart of the algorithm of the proposed fuzzy decision tree

IV. EXPERIMENTS

The datasets which have been used for the experiments are obtained from [8] their features are briefly described in table 4.1. The next subsections present data sets.

TABLE 4.1: THE FEATURES OF DATA SETS

Datasets	*Samples*	*Attributes*	*Classes*
Weather	16	4	3
Wisconsin Cancer	699	9	2
Monk	432	7	2
Iris	150	4	3
Pima diabetes	768	8	2

Tree Quality Measures; in all the following data sets, calculate tree quality using two measures:

- Size of tree is number of leaf nodes (number of rules).
- Classification accuracy of rules.

A. *Weather Dataset of Quinlan's*

The following sections compare between algorithms fuzzy ID3 "*IG*", fuzzy ID3 with classifiability "*IGCL*", fuzzy ambiguity denote as "*AM*" and a new heuristic fuzzy ambiguity with classifibility "*AMCL*". Set of rules and the accuracy are summarizes in figure 4.1 with different value of fuzzy filter α.

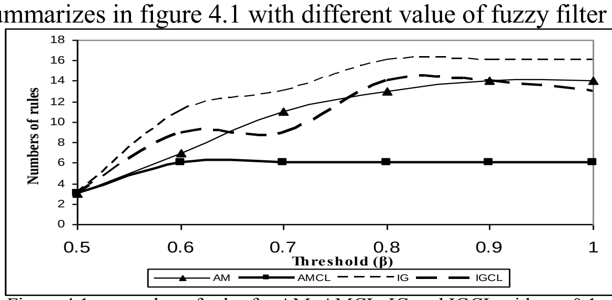

Figure 4.1.a: number of rules for AM, AMCL, IG and IGCL with α = 0.1

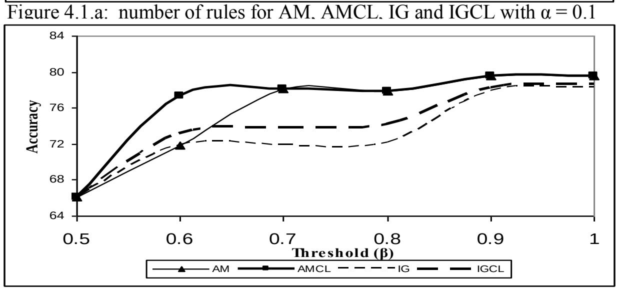

Figure 4.1.b: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.1

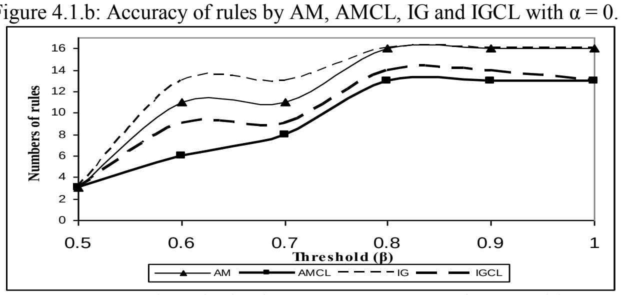

Figure 4.1.c:Number of rules for AM , AMCL, IG and IGCL with α = 0.2

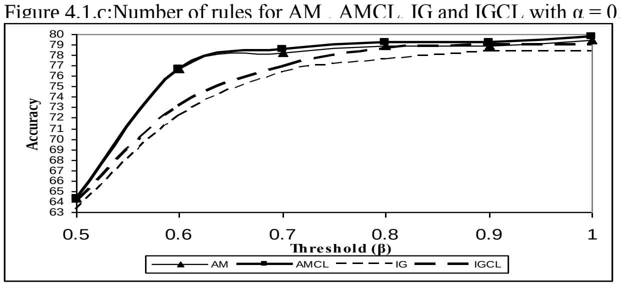

Figure 4.1.d: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.2

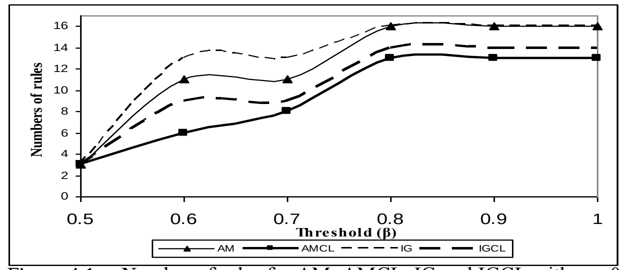

Figure 4.1.e: Number of rules for AM, AMCL, IG and IGCL with α = 0.3

In addition, comparison between IG, IGCL, AM and AMCL from figure 4.2, which summarized in table 4.2, the size of tree AMCL is smaller size than IG, IGCL and AM.

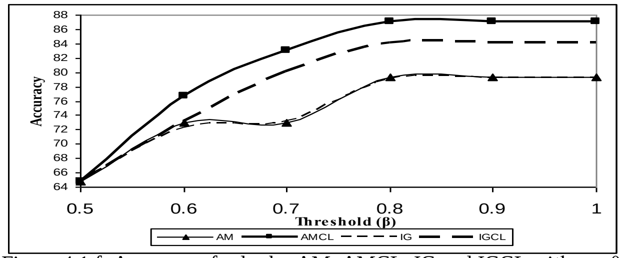

Figure 4.1.f: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.3

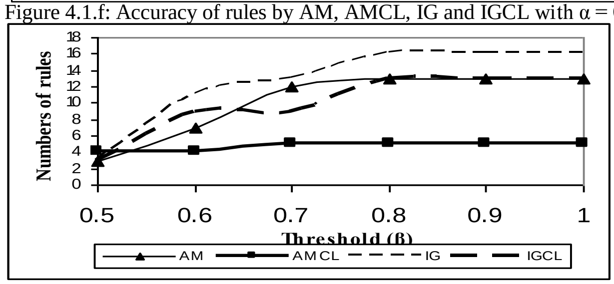

Figure 4.1.g: Number of rules for AM, AMCL, IG and IGCL with α = 0.4

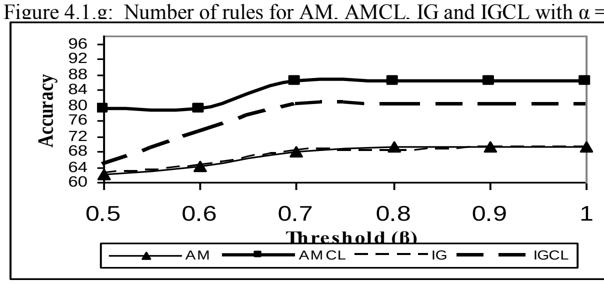

Figure 4.1.h: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.4

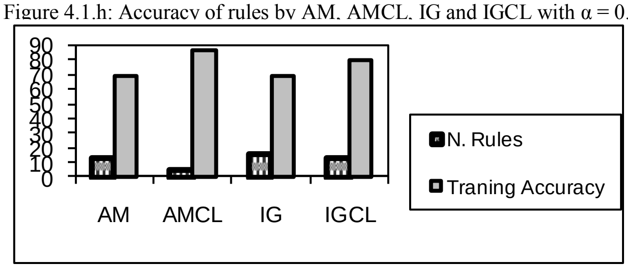

Figure 4.2: Size of tree AMCL is smaller size than IG, IGCL and AM.

TABLE 4.2: NUMBERS OF RULES AND TRAINING ACCURACY

	AM	*AMCL*	*IG*	*IGCL*
N. Rules	13	5	16	13
Training Accuracy	69.2	86.1	69	80.4

B. Wisconsin Breast Cancer (Real World Datasets)

Mangasarian and Bennett (1996) have compiled data on the problem of diagnosing breast cancer to test several new classification methods [Merz (1996)].

Each pattern in the data set has nine inputs and an associated with class label malignant and benign. Total numbers of instances are 699 patients with breast cancer, 458 benign patients and 241 malignant patients, 60% of data for training and 40% for testing and average of the results obtained are shown in figure 4.3.

Moreover, comparison between IG, IGCL, AM and AMCL is shown in figure 4.4 which summarized in table 4.3, illustrate that size of tree by AMCL is smaller szie than tree from IG, IGCL and AM algorithms

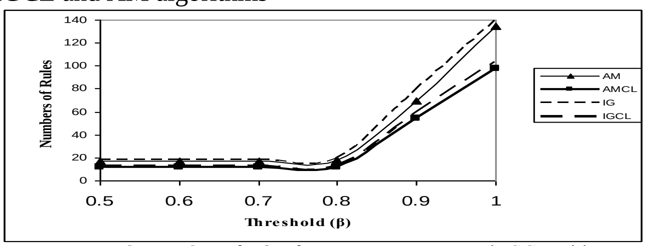

Figure4.3.a: The number of rules for AM,AMCL, IG and IGCL with α=0.1

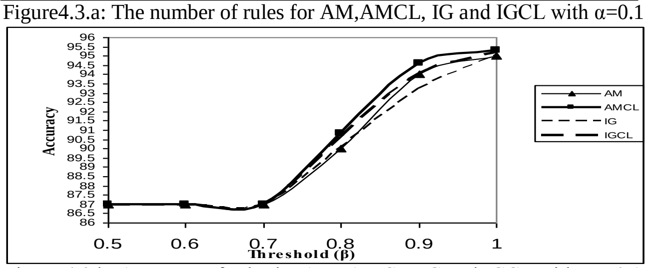

Figure 4.3.b: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.1

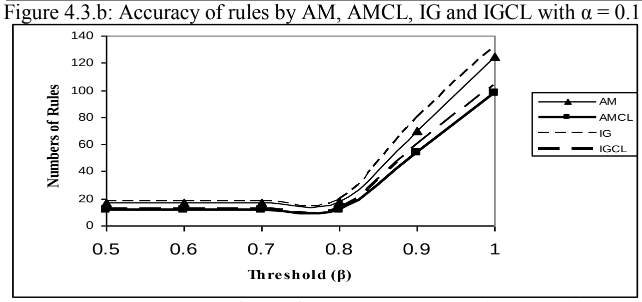

Figure 4.3.c: The number of rules for AM,AMCL,IG and IGCL with α=0.2

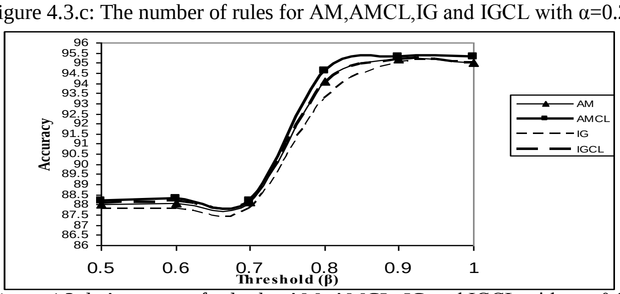

Figure 4.3.d: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.2

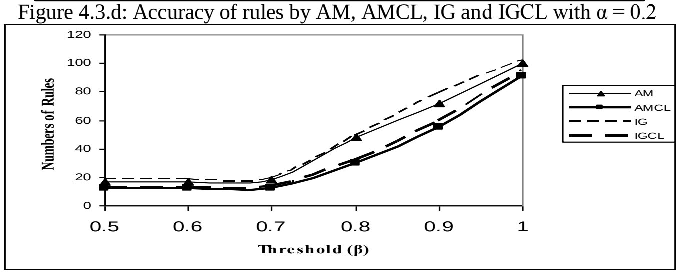

Figure 4.3.e: The number of rules for AM,AMCL,IG and IGCL with α=0.3

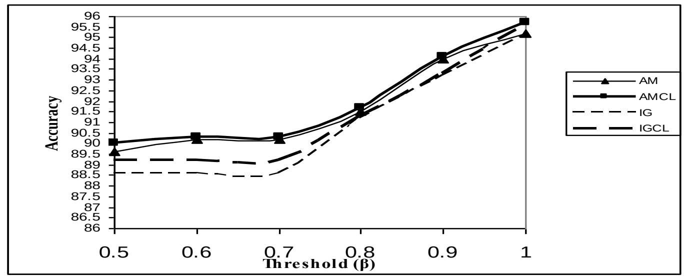

Figure 4.3.f:Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.3

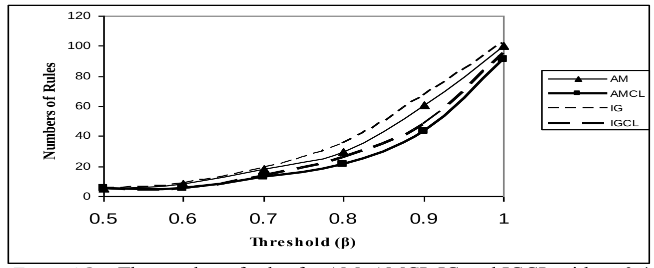

Figure4.3.g:The number of rules for AM ,AMCL,IG and IGCL with α=0.4

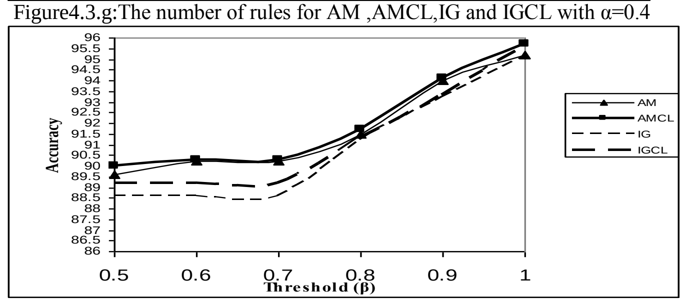

Figure 4.3.h: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.4.

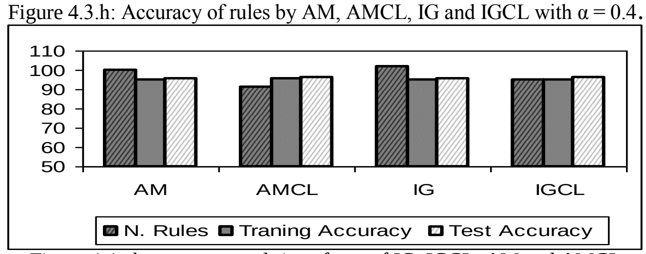

Figure 4.4: the accuracy and size of tree of IG, IGCL, AM and AMCL

Table 4.3: Numbers of rules, training and testing accuracy

	AM	***AMCL***	***IG***	***IGCL***
N. Rules	100	91	102	95
Training Accuracy	95.2	95.7	95.1	95.3
Test Accuracy	95.6	96.1	95.6	96

C. Dataset of Monk_1 Problem

Number of instances is 432, number of attributes is 8 including class attribute. Information of attributes is [8]:

1. Id unique symbol for each instance 2. A1 1, 2, 3

3. A2 1, 2, 3 4. A3 1, 2

5. A4 1, 2, 3 6. A5 1, 2, 3, 4

7.A6 1, 2 8.Class 0, 1

Missing Attribute Values: None

Target Concepts associated to the MONK-1 problem (a1 = a2) or (a5 = 1). 60% of data for training and 40% for testing and average of the results obtained are shown in figure 4.5. Set of rules, the accuracy are summarizes in figure 4.5 with different value of fuzzy filter α.

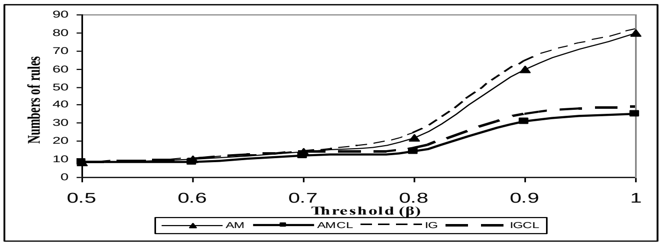

Figure 4.5.a: The number of rules for AM,AMCL,IG and IGCL with α = 0.1

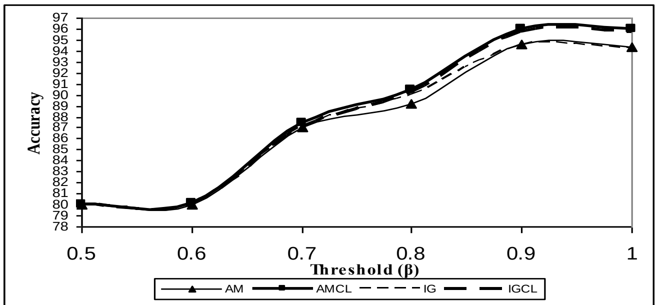

Figure 4.5.b: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.1

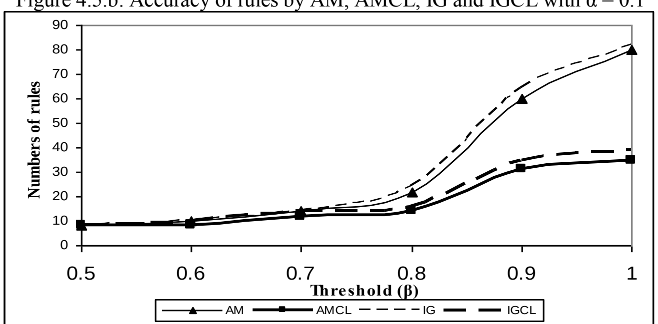

Figure 4.5.c: The number of rules for AM,AMCL,IG and IGCL with α = 0.2

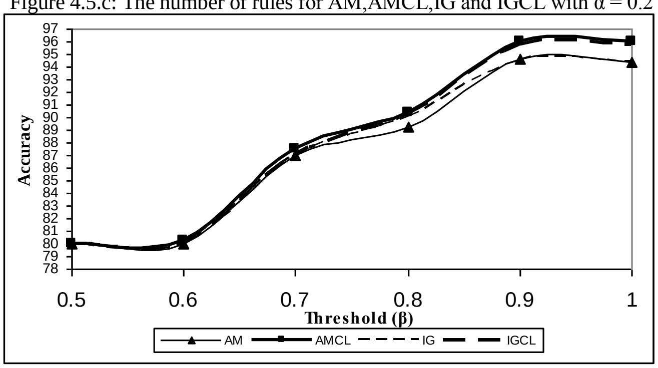

Figure 4.5.d: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.2

Moreover, the comparison between IG, IGCL, AM and AMCL is shown in figure 4.6, which summarized in table 4.4, the size of tree by AMCL is smaller that the other.

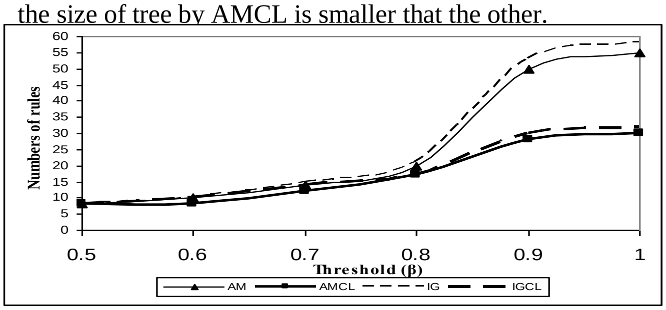

Figure 4.5.e: The number of rules for AM, AMCL,IG and IGCL with α=0.3

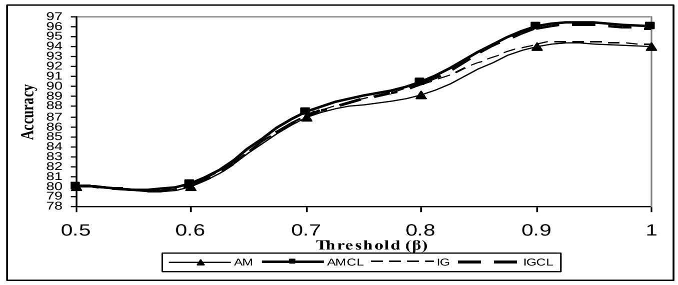

Figure 4.5.f: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.3

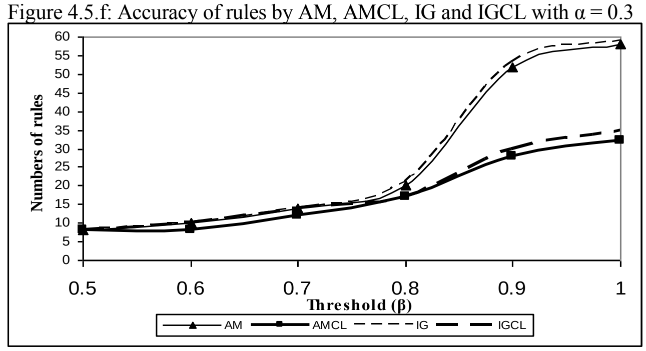

Figure 4.5.g: The number of rules for AM,AMCL,IG and IGCL with α = 0.4

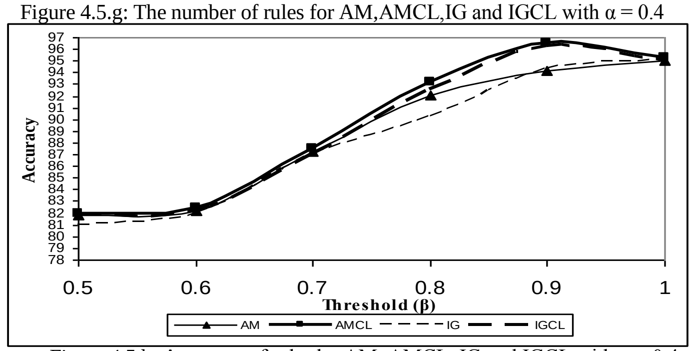

Figure 4.5.h: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.4

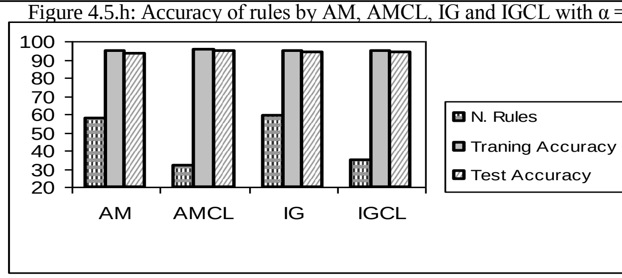

Figure 4.6: Accuracy of AMCL is higher and tree is smaller than other algorithms.

TABLE 4.4: NUMBERS OF RULES, TRAINING AND TESTING ACCURACY

	AM	*AMCL*	*IG*	*IGCL*
N. Rules	58	32	59	35
Training Accuracy	95	95.3	95	95
Test Accuracy	93.02	94.61	94	94.1

D. Dataset of Iris Problem

This is Fisher's famous Iris data, which has been extensively studied in the statistics and machine learning. Number of instances is 150; numbers of attributes are four attribute Information [8] :

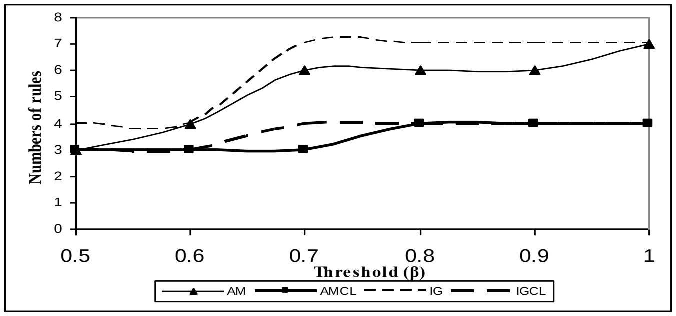

Figure 4.8.a: The number of rules for AM,AMCL,IG and IGCL with α = 0.1

1. Sepal length in cm
2. Sepal width in cm
3. Petal length in cm
4. Petal width in cm
5. Class {Iris Setosa, Iris Versicolour, Iris Virginica}

70% of data for training and 30% for testing and average of the results obtained are shown in figure 4.8. Set of rules, the accuracy are summarizes in figure 4.8 with different value of fuzzy filter α.

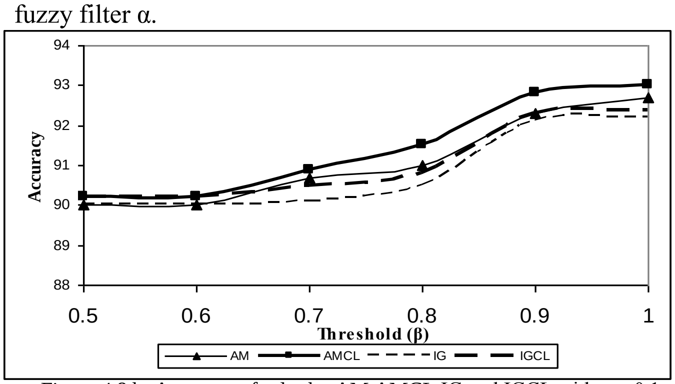

Figure 4.8.b: Accuracy of rules by AM,AMCL,IG and IGCL with α = 0.1

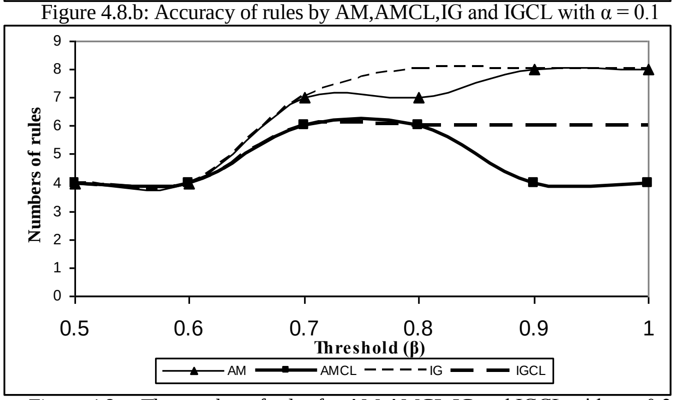

Figure 4.8.c: The number of rules for AM,AMCL,IG and IGCL with α = 0.2

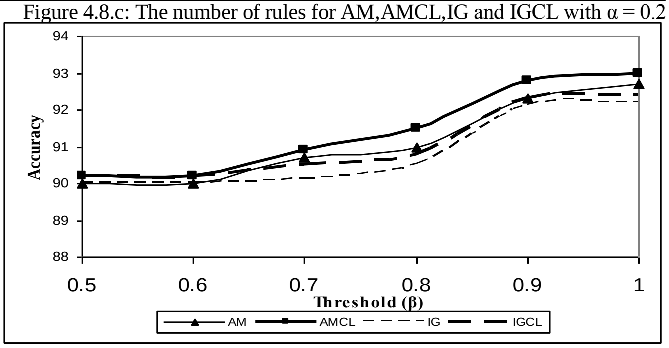

Figure 4.8.d: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.2

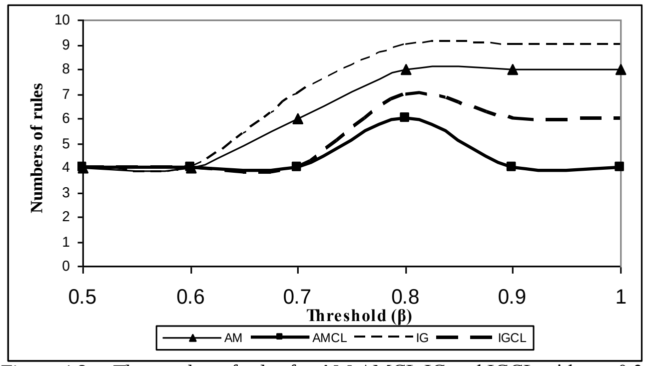

Figure 4.8.e: The number of rules for AM,AMCL,IG and IGCL with α = 0.3

Moreover, the comparison between IG, IGCL, AM and AMCL is shown in figure 4.9, which summarized in table 4.6 the size of tree by AMCL is smaller than others.

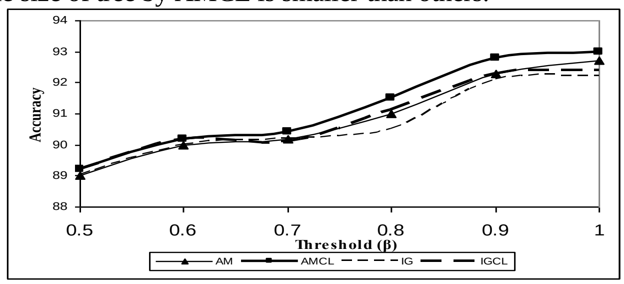

Figure 4.8.f: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.3

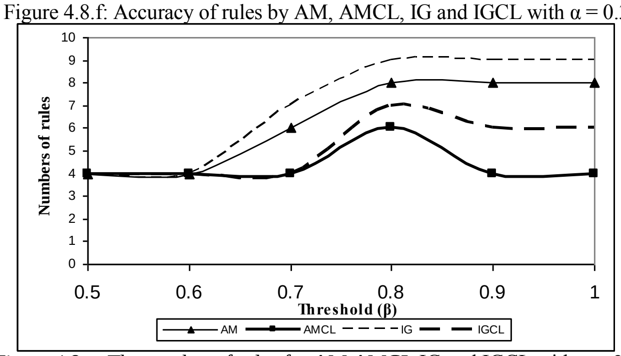

Figure 4.8.g: The number of rules for AM,AMCL,IG and IGCL with α = 0.4

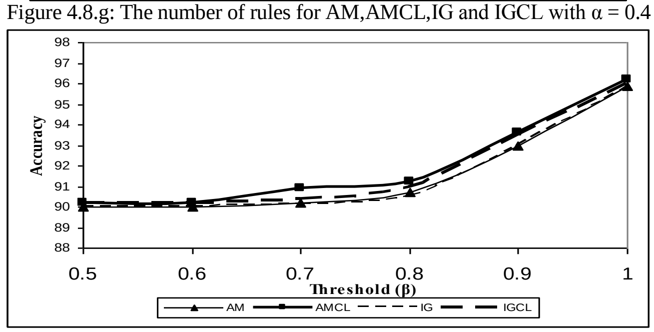

Figure 4.8.h: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.4

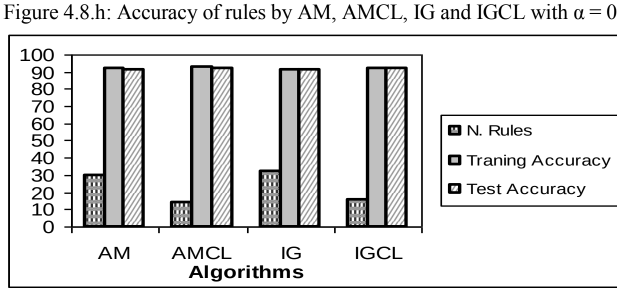

Figure 4.9: Accuracy of AMCL is higher and tree is smaller than other algorithms.

Table 4.6: Numbers of rules, training and testing accuracy

	AM	AMCL	IG	IGCL
N. Rules	8	4	9	6
Training Accuracy	91.9	92.6	91.2	92.4
Test Accuracy	95.2	96	96	96.4

E. Pima Indians Diabetes Database

This data catalogs the presence or absence of diabetes among Pima Indian females, the original source of the data is the National Institute of Diabetes, Digestive, and Kidney Disease (Indian), and it is available in UCI repository The training data has number of instances is 768 and number of Attributes is 8 plus class for each attribute (all numeric-value) [8]:

1. Number of times pregnant
2. Plasma glucose concentration a2 hours in an oral glucose tolerance test
3. Diastolic blood pressure (mm Hg)
4. Triceps skin fold thickness (mm)
5. 2-hour serum insulin (mu .U/ml)
6. Body mass index (weight in kg/(height in m)^2)
7. Diabetes pedigree function
8. Age (years)
9. Class variable (0 or 1) class 0 has 500 instances and class 1 has 268 instances, 50% of data for training and 50% for testing and average of the results obtained are shown in figure 4.11. Set of rules, the accuracy are summarizes in figure 4.11 with different value of fuzzy filter α.

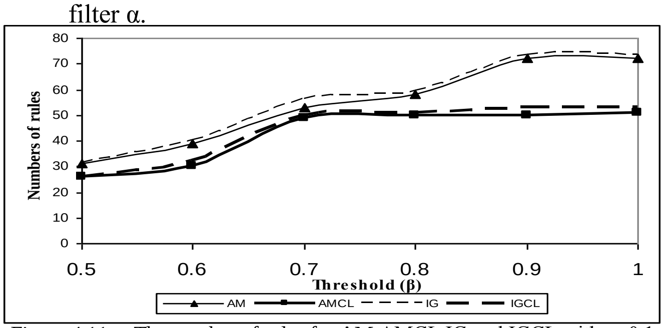

Figure 4.11.a: The number of rules for AM,AMCL,IG and IGCL with α=0.1

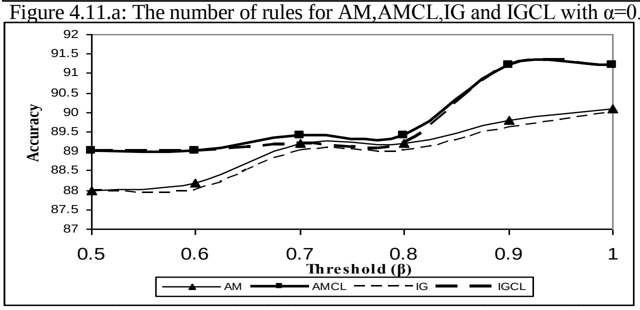

Figure 4.11.b: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.1

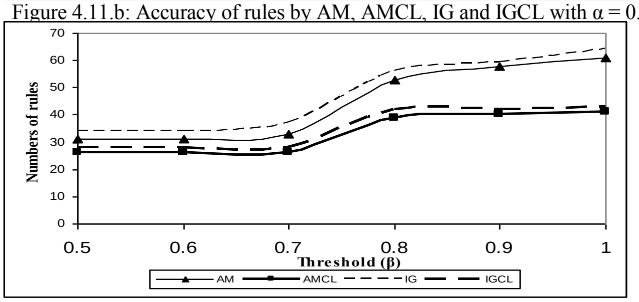

Figure 4.11.c:The number of rules for AM ,AMCL,IG and IGCL with α=0.2

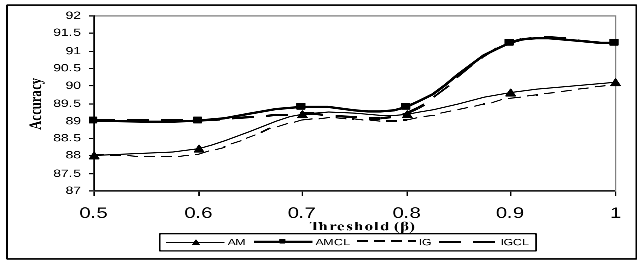

Figure 4.11.d: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.2

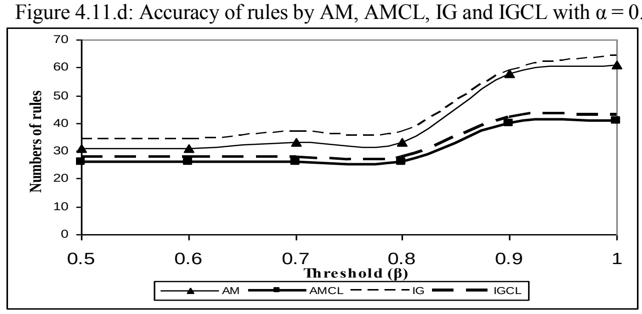

Figure 4.11.e: The number of rules for AM,AMCL,IG and IGCL with α=0.3

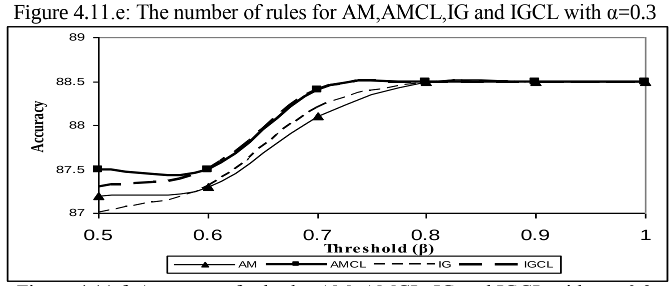

Figure 4.11.f: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.3

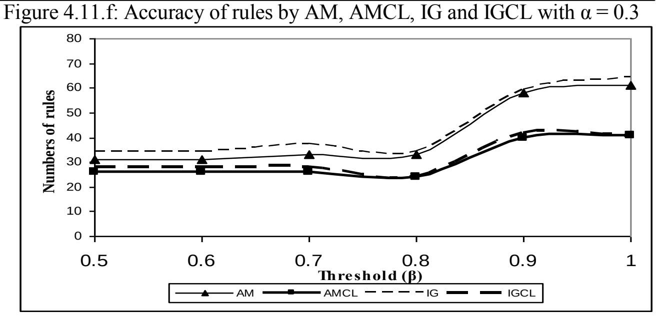

Figure 4.11.g: The number of rules for AM,AMCL,IG and IGCL with α=0.4

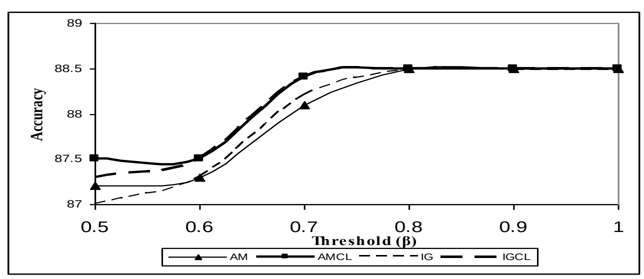

Figure 4.11.h: Accuracy of rules by AM, AMCL, IG and IGCL with α = 0.4

Comparison between IG, IGCL, AM and AMCL is shown in figure 4.12, which summarized in table 4.8 size of tree by AMCL is smaller than others.

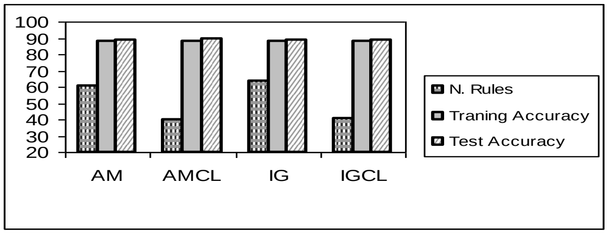

Figure 4.12: Tree size of AMCL is smaller than other algorithms.

Table 4.8: Numbers of rules, training and testing accuracy

	AM	*AMCL*	*IG*	*IGCL*
N. Rules	61	40	64	41
Training Accuracy	88.5	88.5	88.48	88.5
Test Accuracy	89.2	89.4	88.9	89.2

V. CONCLUSION

Tree induction has become an important technique for machine learning, expert system and prediction analysis and so on. Most existing methods are crisp and fuzzy decision tree induction. When choosing a decision tree induction method to classify unseen instance, we mainly consider the generalization capability of tree induction. This paper analyzes and compares the generalization capability of decision tree between fuzzy and crisp tree algorithms. The initial conclusion is that, for the classification problem of numerical attributes; the fuzzy decision tree has the stronger generalization capability than crisp one.

In this paper we presented a novel approach for evaluation the classifiability of instance based on evaluating the texture of class label surface. Based on that we also proposed an algorithm for fuzzy decision tree induction using the classifiability of instance, the proposed method AMCL when combined Ambiguity heuristic with the CLassifiability of instance. More specifically, it results in smaller decision tree and as a consequence better generalization (test) performance.

REFERENCES

[1] Dong, M. and R. Kothari, "Classifiability based pruning of decision trees" IEEE Transactions on Fuzzy Systems, vol. 9, no.2, pp. 1739-1743, 2001b.

[2] Huang, Z. , T. D. Gedeon and M. Nikravesh, "Pattern Trees Induction: A New Machine Learning Method", IEEE Transactions on Fuzzy Systems, vol. 16 no. 4, August 2008.

[3] Isao, H., "A formulation of learning type fuzzy ID3 ", available at: "http://www.cs.toronto.edu/~delue", 2002.

[4] Juan, S., D.S. Yeung and X.Z. Wang, "An initial comparison of generalization capability between crisp and fuzzy decision trees", Proceedings of 1st International Conference on Machine Learning and Cybernetics, Bijing, 2004.

[5] Kim, M.W., L.G. Joong and M. Changwoo, "Efficient rule generation based on fuzzy decision tree for data mining ", IEEE International fuzzy systems Conference proceedings, vol. 1 p.p. 1223-1228, August 1999.

[6] Kothari, R. and M. Dong, "Lookahead based fuzzy decision tree induction", IEEE Transactions on Fuzzy Systems vol. 9, No.3, pp. 461-468, 2001a.

[7] Liu, C.S., "Learning fuzzy rules from examples by fuzzy decision tree", available at: "http://neuron.et.ntust.edu.tw /m8802126", 2002.

[8] Merz, J.C. and P.M. Murphy UCI Repository of Machine Learning Database, available at: "ftp://ftp.ics.uci.edu/pub/machine-learning-database", 1996.

[9] Myung, K.L., M.L. Kyung, J.H. Lee and H.L. Hwang, "A fuzzy decision trees induction method for fuzzy data ", IEEE International fuzzy systems Conference proceedings, vol. 1 p.p. 16-21, August 1999.

[10] Quilnlan, J.R., "Induction of decision trees", Machine learning Vol. 1 pp. 81-106, 1986.

[11] Tsang, E.C.C and D.S. Yeung, X.Z. Wang, "A comparative study on heuristic algorithms for generating fuzzy decision trees". IEEE SMC ' 99 conference proceedings.IEEE International Conference on Systems, Man. And Cybernetics, vol. 3. pp343 -348, 1999.

[12] Wang, T., Z. Li, Y. Yan, and H. Chen," A Survey of Fuzzy Decision Tree Classifier Methodology", Fuzzy information and Engineering (ICFIE), ASC 40, pp. 959–968, 2007.

[13] Wang, X.Z. and H.M. Guang, "A comparison between fuzzy-ID3 and OFFSS-Based fuzzy-ID3", Proceedings of 3rd International Conference on Machine Learning and Cybernetics, Shanghai, 2004.

[14] Wang, X.Z., D.S. Yeung and E.C.C. Tsang, "Fuzzy rule mining by fuzzy decision tree induction based on fuzzy feature subset", International Conference on Systems, Man, and Cybernetics, Tunisia, vol. 4, pp. 599 - 604, 2002.

[15] Xue, L., H.Z. Xiao and D.Z.Dong "Four Matching Operators of Fuzzy Decision Tree Induction", Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007) vol.4, pages 674-678, 2007.

[16] Yeung, D.S., X.Z. Wang and E.C.C. Tsang, "A comparative study on heuristic algorithms for generating fuzzy decision trees", IEEE Transaction on systems, Man, and Cybernetics, Part B: Cybernetics, vol. 31, no.2, pp. 215-226, April 2001.

[17] Yeung, D.S., X.Z. Wang and E.C.C. Tsang, "Leaning weighted fuzzy rules from examples with mixed attributes by fuzzy decision trees", IEEE SMC'99 Conference Proceedings, IEEE International Conference on Systems, Man. And Cybernetics, vol. 3, pp343 -348, 1999.

[18] Yuan, Y. and M. J. Shaw, " Induction of fuzzy decision trees", Fuzzy Sets and systems, Vol. 69, pp.125-139, 1995.

[19] Zhao, X.W., F. Chao and J. Sun, "Analysis on fuzzy filter in fuzzy decision trees", Proceedings of 2nd International Conference on Machine Learning and Cybernetics.2-5 November, 2003.

AUTHORS PROFILE

Hesham Ahmed Hefny is an assistant professor and the head of Computer & Information Sciences Department at the Institute of Statistical Studies and research, Cairo University. His research interests include Artificial Neural Networks, Fuzzy Systems, Genetic Algorithms, Swarm Intelligence, Pattern Recognition, and Data Mining. Dr. Hesham has published over 35 peer refereed papers in academic journals and conferences on topics within Artificial Intelligence and related areas.

Ahmed S. Ghiduk is an assistant professor at Beni-Suef University, Egypt. He received the BSc degree from Cairo University, Egypt, in 1994, the MSc degree from Minia University, Egypt, in 2001, and a Ph.D. from Beni-Suef University, Egypt in joint with College of Computing, Georgia Institute of Technology, USA, in 2007. His research interests include software engineering especially search-based software testing, genetic algorithms, and ant colony. Currently, Ahmed S. Ghiduk is an assistant professor at College of Computers and Information Systems, Taif University, Saudi Arabia.

Ashraf Abdelwahab is a professor of Computer Engineering, Electronics Research Institute, Cairo, Egypt. He received his M.Sc. in 1988, Faculty of Engineering, Cairo University in the area of Artificial Intelligence. In 1992 he received his Ph.D. degree in Machine Learning and Evolutionary Algorithms. He has published over 150 technical papers in national and international journals and conferences in the areas of Evolutionary Algorithms, Machine Learning and Data Mining. Currently, Dr. Abdelwahab is Senior Advisor to the Minister of State for Administrative Development. He is responsible of developing, monitoring and coordinating E-government initiatives, supervising the coordination with other ministries and international relations.

Mohamed Aboubaker is a Ph.D. student in Computer Sciences Department at the Institute of Statistical Studies and Research, Cairo University. His Ph.D. in the filed of decision tree. His research interests include rough and fuzzy set theory, pattern recognition, data mining, and knowledge discovery.