

Interval-Valued Fuzzy Decision Trees

Youthachai Lertworapachaya, Yingjie Yang, and Robert John

Abstract—This research proposes a new model for constructing decision trees using interval-valued fuzzy membership values employing on look-ahead based fuzzy decision tree induction and interval-valued fuzzy sets. Most existing fuzzy decision trees do not consider the uncertainty associated with their membership values. However, precise values of fuzzy membership values are not always possible. In this paper, we represent fuzzy membership values as intervals to model uncertainty and employ the look-ahead based fuzzy decision tree induction method and Hamming distance of interval-valued fuzzy sets to construct decision trees. An example is given to demonstrate the effectiveness of the approach.

I. INTRODUCTION

Fuzzy decision trees (FDTs) [1], [7], [8], [9], [10] can deal with fuzzy data. FDT have been constructed using type-1 fuzzy sets which are precise in nature. Determining type-1 fuzzy sets is well known to be difficult [16]. For example, experts will disagree about the shape and position of a type-1 fuzzy set representing a linguistic term. Existing FDT models can not process uncertain fuzzy membership values. There have been applications of type-2 fuzzy sets in decision tree construction [17]. However, these methods are designed to work with the same data as type-1 models. In real world data, the uncertainty of fuzzy membership values may be represented as interval values. Interval-valued fuzzy sets [2], [3], [4] allow for uncertain membership functions. Here, we apply interval-valued fuzzy sets to construct an interval-valued fuzzy decision tree. This new model can construct the tree from data with uncertain fuzzy membership values. In this way, a FDT is extended to an Interval-Valued Fuzzy Decision Tree (IVFDT). Ming Dong and Ravi Kothari (2001) proposed a look-ahead based fuzzy decision tree (LAFDT) induction method for data represented with type-1 fuzzy sets [8]. The LAFDT method can evaluate the classifiability of instances, that are split along branches of a given node based on evaluating the texture of the class label surface. Moreover, this particular method consists in finding the instance that are within a distance r from a given instance. In this paper, we propose a new model to construct a decision tree using interval-valued fuzzy membership values based on LAFDT induction and interval-valued fuzzy sets. It is called Look-Ahead Based Interval-Valued Fuzzy Decision Tree (LAIVFDT).

Section II provides some necessary material on LAFDT, and interval-valued fuzzy sets (IVFS). Section III proposes an extension of LAFDT induction to LAIVFDT. Section IV demonstrates a simple example in the application of

LAIVFDT to data with uncertain fuzzy membership values. Section V summaries the results of the study.

II. PRELIMINARIES

A. The Look-Ahead Based Fuzzy Decision Tree

There are many different models of FDT [1], [10]. Look-ahead based fuzzy decision tree (LAFDT) is one of the latest models [1], [7], [8]. In a FDT, the key is to find the appropriate attribute to split samples into different branches along the tree [1], [12], [13]. The LAFDT has a particular method of evaluating the classifiability of attributes along the branches of a node to split and produce a smaller decision tree. A nonparametric method in LAFDT is to characterise the classifiability of attributes using an occurrence matrix. The usual approach for of LAFDT is, for any instance x , to measure a distance r between two instances that are within a circular neighbourhood of radius r based on the distance in equation (1). The r distance assists in filtering the instances which exceed the radius r , and to reduce computing time.

Considering the universe of objects described by n attributes, an attribute has values of fuzzy subsets $A_1^k, A_2^k, \dots, A_{m_k}^k$. The distance between two objects (or instance x and y) can be measured using their fuzzy memberships.

Definition 1: [8] Let $\mu_i^{(k)}(x)$ ($1 \leq k \leq n$, $1 \leq i \leq m_k$, $1 \leq x \leq N$) denote the membership value of instance x for the i^{th} value of the k^{th} attribute. The distance between instance x and y is defined by

$$D_{xy} = \sum_{k=1}^n \sum_{i=1}^{m_k} |\mu_i^{(k)}(x) - \mu_i^{(k)}(y)|. \quad (1)$$

For any object x in the universe, we can restrict its circular neighbourhood to those objects within a radius r of x . Then local occurrence matrix P for object x is defined as follows.

Definition 2: [8], [9] Let $\mu_j(x)$, $1 \leq j \leq C$ denote the membership value of instance x for class j and let $\mu_j(x) = [\mu_1(x), \dots, \mu_C(x)]$. The local co-occurrence matrix of instance x is define by

$$P(x) = \sum_{y, D_{xy} \leq r} \mu(x)^T \times \mu(y). \quad (2)$$

where, $\mu(x)^T$ is a transpose matrix and r is the neighbourhood radius of x .

The authors are with the Centre for Computational Intelligence, Faculty of Technology, De Montfort University, The Gateway, and Leicester, LE1 9BH, UK.(corresponding author to provide phone: (0116)207-8408; e-mail: youd@dmu.ac.uk, yyang@dmu.ac.uk and rij@dmu.ac.uk).

With a local occurrence matrix, we can derive the co-occurrence matrix for each attribute.

Definition 3: [8], [9] The local co-occurrence matrix after attribute k is selected.

$$W^{(k)} = \sum_{i=1}^{m_k} \sum_x P(x). \quad (3)$$

Then, the classifiability of attribute k is

$$L^{(k)} = \sum_{i=1}^C W_{ij}^{(k)} - \sum_{i=1}^C \sum_{j=1, i \neq j}^C W_{ij}^{(k)}. \quad (4)$$

According to the values of $L^{(k)}$, we can identify the attribute with the highest classifiability in order to build a decision tree.

B. Interval-Valued Fuzzy Sets

Definition 4: [2], [4], [15] Let X denote a universe of discourse. An interval-valued fuzzy set is an expression A denoted by

$$A = \{(x_i, \mu_A(x_i)) | x_i \in X; i = 1, 2, \dots, n\}. \quad (5)$$

where $\mu_A(x_i) : X \rightarrow D([0, 1])$, and $x_i \rightarrow \mu_A(x_i) = [\underline{\mu}_A(x_i), \bar{\mu}_A(x_i)] \in D([0, 1])$.

Compared with type-1 fuzzy sets, the membership is represented as an interval within $[0, 1]$.

If we represent the interval relationship with $\underline{\mu}_A(x_i)$ and $\nu_A(x_i) = 1 - \bar{\mu}_A(x_i)$ then we get intuitionistic fuzzy sets [2], [3], [4]. The interval of intuitionistic fuzzy sets is denoted by $[\mu_A(x_i), 1 - \nu_A(x_i)]$. In this paper, we transform the intuitionistic fuzzy sets into interval-valued fuzzy sets as follows.

Given $\mu_A(x_i) = [\underline{\mu}_A(x_i), \bar{\mu}_A(x_i)] = [\mu_A(x_i), 1 - \nu_A(x_i)]$. Then we can repeat the distance in the form of interval-valued fuzzy sets:

$$d' = \frac{1}{2} \sum_{i=1}^n [|\underline{\mu}_A(x_i) - \underline{\mu}_B(x_i)| + |\bar{\mu}_A(x_i) - \bar{\mu}_B(x_i)|]. \quad (6)$$

III. THE PROPOSED LOOK-AHEAD BASED INTERVAL-VALUED FUZZY DECISION TREE

In section II, it is clear that the key step in LAFDT is the calculation of the distance between two samples or instances. In LAIVFDT, the two samples or instances involved are as combined using two interval-valued fuzzy sets with elements of attributes. The distance between two instances is calculated as the distance between two fuzzy

sets, as shown in equation (1). Obviously, equation (1) is not applicable here, and the distance between two interval-valued fuzzy sets in equation (6) should be applied.

Considering the same universe for attributes and instances in section II, we have the following definition for the distance between two instances.

Definition 5: Let $\mu_i^{(k)}(x) = [\underline{\mu}_i^{(k)}(x), \bar{\mu}_i^{(k)}(x)]$ and $\mu_i^{(k)}(y) = [\underline{\mu}_i^{(k)}(y), \bar{\mu}_i^{(k)}(y)]$; ($1 \leq k \leq n, 1 \leq i \leq m_k$) denote the interval-valued fuzzy membership value of instance x and y for the i^{th} value of k^{th} attribute. The distance between the two instances is

$$\hat{D}_{xy} = \frac{1}{2} \sum_{k=1}^n \sum_{i=1}^{m_k} [|\underline{\mu}_i^{(k)}(x) - \underline{\mu}_i^{(k)}(y)| + |\bar{\mu}_i^{(k)}(x) - \bar{\mu}_i^{(k)}(y)|]. \quad (7)$$

where x is x_i in $\mu_A^{(k)}(x_i)$ and y is x_i in $\mu_B^{(k)}(x_i)$ for \hat{D}_{xy} . For any instance in the universe, we can restrict its circular neighbourhood to those objects within a radius r of x . Then a local co-occurrence matrix P for object x is defined.

Definition 6: Let $\mu_j(x)$; $1 \leq j \leq C$ denote the membership value of instance x of class j and let $\hat{\mu}(x) = [[\underline{\mu}_1(x), \bar{\mu}_1(x)], \dots, [\underline{\mu}_C(x), \bar{\mu}_C(x)]]$. The local co-occurrence matrix of instance x is

$$\hat{P}(x) = \sum_{y, D_{xy} \leq r} \hat{\mu}(x)^T \times \hat{\mu}(y). \quad (8)$$

where, $\hat{\mu}(x)^T$ is a transpose matrix and r is neighbourhood radius of x .

According to the $\hat{P}(x)$ matrix, each matrix element is represented by an interval value. Schneider et al.(1996) described an interval X as a closed, bounded set of real numbers, in which $\{x | \underline{X} \leq x \leq \bar{X}, x \in X\}$ can be denoted as $X = [\underline{X}, \bar{X}]$ [6]. For all real number $\underline{X}, \bar{X}, \underline{Y}$ and \bar{Y} . Such that $0 \leq \underline{X} \leq \bar{X} \leq 1$ and $0 \leq \underline{Y} \leq \bar{Y} \leq 1$ [5], [6], [7]. The rules of interval arithmetic are as follows

- Addition: $[\underline{X}, \bar{X}] + [\underline{Y}, \bar{Y}] = [\underline{X} + \underline{Y}, \bar{X} + \bar{Y}]$.
- Subtraction: $[\underline{X}, \bar{X}] - [\underline{Y}, \bar{Y}] = [\underline{X} - \bar{Y}, \bar{X} - \underline{Y}]$.
- Multiplication: $[\underline{X}, \bar{X}] * [\underline{Y}, \bar{Y}] = [\underline{X} * \underline{Y}, \bar{X} * \bar{Y}]$.
- Division: $[\underline{X}, \bar{X}] / [\underline{Y}, \bar{Y}] = [\underline{X} / \bar{Y}, \bar{X} / \underline{Y}]$.
assuming $0 \leq \underline{Y}$
- Distribution law:
 $\min([\underline{X}, \bar{X}], [\underline{Y}, \bar{Y}]) = [\min(\underline{X}, \underline{Y}), \min(\bar{X}, \bar{Y})]$ and
 $\max([\underline{X}, \bar{X}], [\underline{Y}, \bar{Y}]) = [\max(\underline{X}, \underline{Y}), \max(\bar{X}, \bar{Y})]$.

Note that the operation $[\underline{X}, \bar{X}] / [\underline{Y}, \bar{Y}]$ is undefined, if $\underline{Y} = 0$, $\bar{Y} = 0$, or if both $\underline{Y} = 0$ and $\bar{Y} = 0$ [6].

Therefore, the rules of interval arithmetic above are employed for calculating the $\hat{P}(x)$ matrix in equation (8), the $\hat{W}^{(k)}$ matrix in equation (9) and the $\hat{L}^{(k)}$ matrix in equation (10). With the local occurrence matrix, we can derive the co-occurrence matrix for each attribute.

TABLE I
INTERVAL-VALUED WEATHER DATA SET

Outlook			Temperature			Humidity		Wind		Plan		
Sunny	Cloudy	Rain	Hot	Mild	Cool	Humid	Dry	Windy	Calm	Plan A	Plan B	Plan C
0.6-0.8	0.0-0.2	0.0-0.1	0.9-1.0	0.0-0.1	0.0-0.1	0.7-0.9	0.1-0.3	0.3-0.5	0.5-0.7	0.0-0.1	0.3-0.5	0.5-0.7
0.2-0.4	0.7-0.9	0.0-0.1	0.5-0.7	0.3-0.5	0.0-0.1	0.0-0.1	0.9-1.0	0.0-0.1	0.9-1.0	0.9-1.0	0.2-0.4	0.0-0.1
0.0-0.1	0.6-0.8	0.2-0.4	0.7-0.9	0.1-0.3	0.0-0.1	0.0-0.2	0.8-1.0	0.1-0.3	0.7-0.9	0.2-0.4	0.5-0.7	0.0-0.2
0.0-0.3	0.6-0.8	0.0-0.2	0.2-0.4	0.6-0.8	0.0-0.1	0.1-0.3	0.7-0.9	0.2-0.4	0.6-0.8	0.8-1.0	0.0-0.2	0.0-0.1
0.0-0.1	0.0-0.2	0.8-1.0	0.6-0.8	0.2-0.4	0.0-0.1	0.4-0.6	0.4-0.6	0.4-0.6	0.4-0.6	0.0-0.1	0.0-0.1	0.9-1.0
0.0-0.1	0.6-0.8	0.2-0.4	0.0-0.1	0.2-0.4	0.6-0.8	0.6-0.8	0.2-0.4	0.3-0.5	0.5-0.7	0.1-0.3	0.0-0.1	0.7-0.9
0.0-0.1	0.2-0.4	0.6-0.8	0.0-0.1	0.0-0.1	0.9-1.0	0.0-0.1	0.9-1.0	0.0-0.2	0.8-1.0	0.0-0.1	0.0-0.1	0.9-1.0
0.0-0.1	0.9-1.0	0.0-0.1	0.0-0.1	0.1-0.3	0.7-0.9	0.1-0.3	0.7-0.9	0.0-0.1	0.9-1.0	0.6-0.8	0.0-0.1	0.2-0.4
0.4-0.6	0.6-0.8	0.0-0.1	0.9-1.0	0.0-0.1	0.0-0.1	0.5-0.7	0.3-0.5	0.6-0.8	0.2-0.4	0.1-0.3	0.7-0.9	0.0-0.1
0.4-0.6	0.5-0.7	0.0-0.1	0.0-0.1	0.2-0.4	0.6-0.8	0.0-0.1	0.9-1.0	0.8-1.0	0.0-0.2	0.0-0.1	0.2-0.4	0.6-0.8
0.6-0.8	0.2-0.4	0.0-0.1	0.9-1.0	0.0-0.1	0.0-0.1	0.9-1.0	0.0-0.1	0.1-0.3	0.7-0.9	0.3-0.5	0.8-1.0	0.0-0.1
0.1-0.3	0.5-0.7	0.1-0.3	0.0-0.1	0.9-1.0	0.0-0.1	0.2-0.4	0.6-0.8	0.2-0.4	0.6-0.8	0.6-0.8	0.1-0.3	0.0-0.2
0.8-1.0	0.0-0.2	0.0-0.1	0.1-0.3	0.7-0.9	0.0-0.1	0.2-0.4	0.8-1.0	0.9-1.0	0.0-0.1	0.0-0.1	0.0-0.1	0.9-1.0
0.0-0.1	0.8-1.0	0.0-0.2	0.0-0.1	0.8-1.0	0.0-0.2	0.0-0.2	0.8-1.0	0.6-0.8	0.2-0.4	0.0-0.1	0.0-0.1	0.9-1.0
0.0-0.1	0.0-0.1	0.9-1.0	0.0-0.1	0.0-0.1	0.9-1.0	0.9-1.0	0.0-0.1	0.7-0.9	0.1-0.3	0.0-0.1	0.0-0.1	0.9-1.0
0.9-1.0	0.0-0.1	0.0-0.1	0.7-0.9	0.4-0.6	0.0-0.1	0.0-0.1	0.9-1.0	0.0-0.1	0.9-1.0	0.3-0.5	0.4-0.6	0.0-0.1

Definition 7: The local co-occurrence matrix for attribute k is selected as follows:

$$\hat{W}^{(k)} = \sum_{i=1}^{m_k} \sum_x \hat{P}(x). \quad (9)$$

then, the classifiability of attribute k is

$$\hat{L}^{(k)} = \sum_{i=1}^C w'_{ii} - \sum_{i=1}^C \sum_{j=1, i \neq j}^C w'_{ij}. \quad (10)$$

where, w'_{ii} and w'_{ij} are elements of $\hat{W}(x)$. C is the number of fuzzy subsets of the classification attribute.

For classification of $L^{(k)}$ values, we can identify the attribute k with the highest classifiability to build an interval-valued fuzzy decision tree. For LAFDT, it is worked out comparing between two or more $L^{(k)}$ values and the highest single value is chosen. When the membership values are represented by intervals, the $\hat{L}^{(k)}$ value is an interval. The comparison of $\hat{L}^{(k)}$ values is not as simple as other values. In this case, the values of $\hat{L}^{(k)}$ are compared through their probability [6]. The probability is used to consider the chance of the occurrence of a sample x and y in the intervals.

For example, let $X=[\underline{x}, \bar{x}]$ and $Y=[\underline{y}, \bar{y}]$ denote the interval of X and Y , respectively. Suppose that the sample x is in the interval X and y is in the interval Y . The relationship between x and y can be either $x < y$, $x = y$ or $x \geq y$. Only the probability of $x < y$ is needed in this context. The probability of $x < y$ is split into three types: $P(x \leq y)_I$, $P(x \leq y)_P$ and $P(x \leq y)_F$ [6].

- 1) $P(x < y)_I$: x and y are in $X_I = X \cap Y$ and $x < y$ is within the intersection.
- 2) $P(x < y)_P$: x is in the intersecting interval and precedes the intersecting intervals.
- 3) $P(x < y)_F$: x is in the intersecting interval and y follows.

According to [6], there are six possible cases of the probability $P(x < y)$. The six possible situations could be represented using probability $P(x < y)$ as follows [6].

- 1) If x precedes y entirely, there is no overlapping. ($X \cap Y = \Phi$ and X precedes Y then $P(x < y) = 1$ (Figure 1).
- 2) If $(X \cap Y) \neq \Phi$ and $(x < y)$ then $P(x < y) = P(x \leq y)_I + P(x \leq y)_P + P(x \leq y)_F$ (Figure 2).
- 3) If $(x < y)$ and x is in $(X \cap Y)$ then $P(x < y) = P(x \leq y)_I + P(x \leq y)_P$ (Figure 3).
- 4) If $(x < y)$ and x is in $(x < y)$ and y follows then $P(x < y) = P(x \leq y)_I + P(x \leq y)_F$ (Figure 4).
- 5) If $(x < y)$ is in $(X \cap Y)$ then $P(x < y) = P(x \leq y)_I$ (Figure 5).
- 6) If x follows y entirely, there is no overlapping. ($X \cap Y = \Phi$ and X follows Y . Then $P(x < y) = 0$ (Figure 6).

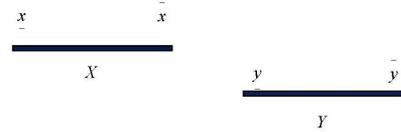


Fig. 1. The probability of $P(x < y) = 1$

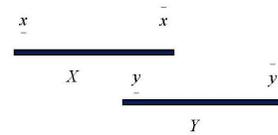


Fig. 2. The probability of $P(x < y) = P(x \leq y)_I + P(x \leq y)_P + P(x \leq y)_F$

In general, the equation of $P(x < y)$ is $P(x < y) = P(x \leq y)_I + P(x \leq y)_P + P(x \leq y)_F$. Obviously, if $P(x < y) > 0.5$, it means that there is greater opportunity

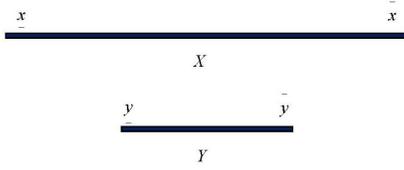


Fig. 3. The probability of $P(x < y) = P(x \leq y)_I + P(x \leq y)_P$

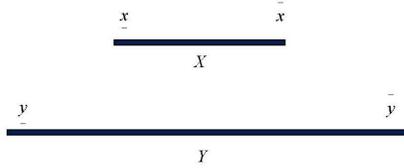


Fig. 4. The probability of $P(x < y) = P(x \leq y)_I + P(x \leq y)_F$

for $x < y$ then $x > y$. In this case, we can consider $x < y$ more possible than $x > y$. Then the attribute represented by y should have priority over the attribute. The probability of $P(x < y)$ is employed to find the maximum $\hat{L}^{(k)}$ value. It is discussed in section IV.

IV. APPLICATIONS

A. Data Set

A data set in [8] is adapted into interval membership values to verify the proposed methods. Each precise membership value in the data set is transformed into an interval. There are four attributes: “Outlook”, “Temperature”, “Humidity”, “Wind” and one classification attribute: “Plan” [8]. Outlook can be sunny, cloudy or rain; temperature can be hot, mild or cool; humidity can be humid or dry; wind can be windy or calm and plan can be A, B or C. Suppose a precise membership value is denoted by $\tilde{\mu}_A(x)$ in [8]. Mendel et al. (2006) suggested that a membership degree $\mu_A(x)$ can be provided by an expert with an appropriate degree $\tilde{\mu}_A(x)$ and a bound Δx describing his uncertainty [14]. For example, an interval of possible values of uncertainty can be expressed as $[\underline{\mu}_A(x), \overline{\mu}_A(x)] = [\tilde{\mu}_A(x) - \Delta x, \tilde{\mu}_A(x) + \Delta x]$. Therefore, an interval-valued membership value is assigned as $[\underline{\mu}_A(x), \overline{\mu}_A(x), 0 \leq \underline{\mu}_A(x), \overline{\mu}_A(x) \leq 1]$.

B. Interval-Valued Fuzzy Decision Tree Algorithm

For a given data set, a fuzzy decision tree can be constructed by the following algorithm.

- Step 1: Fuzzify the training data and testing data into interval-valued fuzzy sets.
- Step 2: Compute the distance \hat{D}_{xy} all between instance x and y by equation (7).
- Step 3: Calculate the local co-occurrence matrix $\hat{P}(x)$ by comparison with the r value which is defined by the user. Compute the local co-occurrence matrix $\hat{P}(x)$ by equation (8). $\hat{P}(x)$ is subject to the restriction of r .

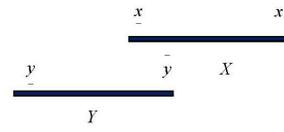


Fig. 5. The probability of $P(x < y) = P(x \leq y)_I$

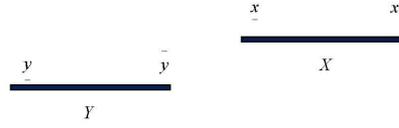


Fig. 6. The probability of $P(x < y) = 0$

Step 4: Select an attribute and sum the local co-occurrence matrix ($\hat{W}(x)$) along each branch using equation (9).

step 5: Normalise the matrix and calculate $\hat{W}(x)$ using equation (10).

Step 6: Repeating step 4 to step 5 for all attributes.

Step 7: The attribute with the maximum probability for having greater $\hat{L}^{(k)}$ than others is selected for the corresponding node to split the sample set into next layer branches.

- 7.1 For the root node, select the attribute with the highest possibility for having a greater value of the look-ahead term $\hat{L}^{(k)}$ than others.
- 7.2 For each child node, the attribute with the highest possibility to have a greater $\hat{L}^{(k)}$ value than that of the left attribute is selected to further split branches of the decision tree.
- 7.3 The node is a leaf node if enough of the instances corresponds to a same class of classification.

C. Results

We used 16 instances of the interval valued data set and tested them with $r = 0.5, 1, 2, 3, 4, 5$ and 6. Figure 7 to 9 illustrate the decision trees with different r values. The trees in figures 7 and 9 have 19 nodes and figure 8 has 21 nodes; the number of nodes in figure 7 and 9 is less than in figure 8. Thus, the decision tree in figures 7 and 9 are better than the tree in figure 8. The interval-valued fuzzy decision tree with $r = 2$ in figure 7 and $r = 0.5$ in figure 9 can be selected for a root node of the tree.

D. Analysis

For the given data set in table I, we constructed decision trees with $r = 0.5, 1, 2, 3, 4, 5$ and 6. When $r = 0.5$ and 2 we obtain the smallest tree with 19 nodes. when $r = 1$ the trees have 21 nodes. When $r = 3, 4, 5$ and 6 the trees could not be constructed, because there was not a dominant attribute for a root node. This is discussed later in the same section. For example, if we select $r = 0.5$, we get the results of $\hat{L}^{(k)}$ for each attribute at root node as follows (see table II):

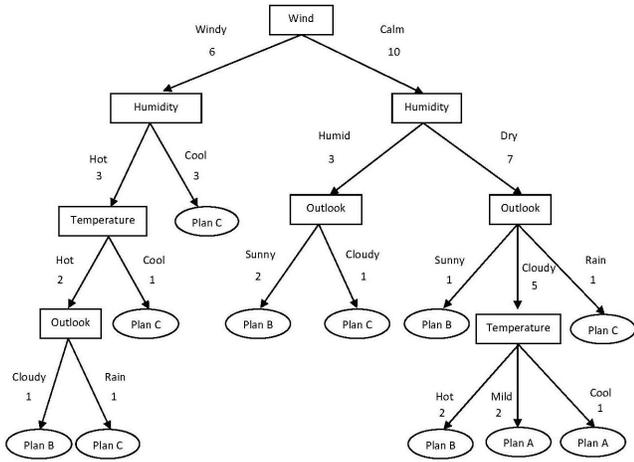


Fig. 7. Interval-valued fuzzy decision tree of weather data set with $r = 2$

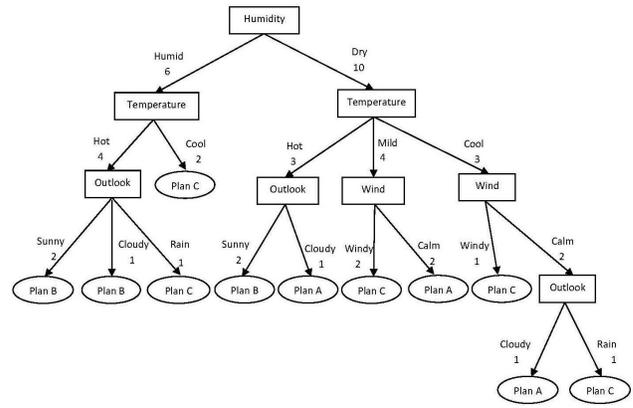


Fig. 9. Interval-valued fuzzy decision tree of weather data set with $r = 0.5$

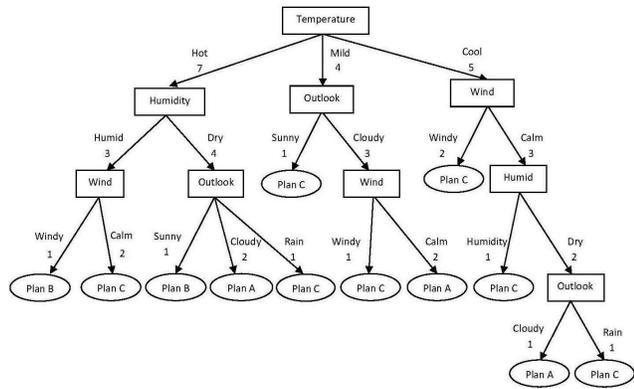


Fig. 8. Interval-valued fuzzy decision tree of weather data set with $r = 1$

$$\begin{aligned} \hat{L}(\text{Outlook}) &= [-0.89, 0.75] \\ \hat{L}(\text{Temperature}) &= [-1.54, 1.40] \\ \hat{L}(\text{Humidity}) &= [-0.97, 0.83] \\ \hat{L}(\text{Wind}) &= [-1.22, 1.01] \end{aligned}$$

Table II illustrates the effect of r value on the values of $\hat{L}(x)$.

TABLE II
THE RESULTS OF $\hat{L}(x)$

r	Outlook	Temperature	Humidity	Wind
0.5	[-0.89, 0.75]	[-1.54, 1.40]	[-0.97, 0.83]	[-1.22, 1.01]
1	[-0.94, 0.76]	[-1.48, 1.34]	[-1.04, 0.89]	[-1.33, 1.13]
2	[-1.15, 0.66]	[-1.66, 1.04]	[-1.27, 1.04]	[-1.62, 1.13]
3	[-1.51, 0.74]	[-1.96, 1.22]	[-1.34, 0.61]	[-1.57, 0.82]
4	[-1.61, 0.67]	[-2.04, 1.10]	[-1.39, 0.40]	[-1.75, 0.75]
5	[-1.61, 0.66]	[-2.08, 1.14]	[-1.35, 0.38]	[-1.71, 0.70]
6	[-1.56, 0.62]	[-2.06, 1.11]	[-1.36, 0.39]	[-1.70, 0.70]

Table III illustrates the probability of $x < y$ for each pair of $\hat{L}^{(k)}$ with $r = 0.5, 1, 2, 3, 4, 5$ and 6 . Using the algorithm in section IV, each pair of $\hat{L}^{(k)}$ is compared, e.g. the probability for $\hat{L}(\text{Wind}) \leq \hat{L}(\text{Humidity})$ with $r = 0.5$ is 0.516 . As we can see, it has a confidence or about 51.6% . From table III,

we can see all values in Humidity column with $r = 0.5$ are greater or equal to 0.5 . It indicates that the probability for any other attribute to have a lower $\hat{L}(x)$ value than Humidity is greater than 0.5 . Therefore, we can draw our conclusion that Humidity should be selected as the root attribute to split the tree into branches.

The probability for $\hat{L}(\text{Wind}) \leq \hat{L}(\text{Temperature})$ is 0.511 . As we can see, it has a confidence or about 51.1% . From table III, we can see all values in the Temperature column with $r = 0.5$ are greater or equal to 0.5 . It indicates that the probability for any other attribute to have a lower $\hat{L}(x)$ value than Temperature is greater than 0.5 . Therefore, we can draw our conclusion that Temperature should be selected as the root attribute to split the tree into branches.

However, the probability for Temperature to have a larger $L^{(k)}$ value than Wind is 0.512 , but the probability between Humidity and Wind is 0.516 and the probability of $L^{(k)}$ for Humidity is greater than Wind. Thus, we should select Humidity.

The probability for $\hat{L}(\text{Temperature}) \leq \hat{L}(\text{Wind})$ is 0.524 . From table III, we can see all values in the Wind column with $r = 2$ are greater or equal to 0.5 . It indicates that the probability for any other attribute to have a lower $\hat{L}(x)$ value than Wind is greater than 0.5 . Therefore, we can draw our conclusion that Wind should be selected as the root attribute to split the tree into branches.

With $r = 3$, the probability for $\hat{L}(\text{Outlook}) \leq \hat{L}(\text{Wind})$ is 0.188 , $\hat{L}(\text{Temperature}) \leq \hat{L}(\text{Wind})$ is 0.498 and $\hat{L}(\text{Humidity}) \leq \hat{L}(\text{Wind})$ is 0.496 . We can see the values in the Wind column are not greater than 0.5 . The probability for $\hat{L}(\text{Outlook}) \leq \hat{L}(\text{Humidity})$ is 0.129 , $\hat{L}(\text{Temperature}) \leq \hat{L}(\text{Humidity})$ is 0.502 , and $\hat{L}(\text{Wind}) \leq \hat{L}(\text{Humidity})$ is 0.504 . We can see some values in the Humidity column are not greater than 0.5 . The probability for $\hat{L}(\text{Outlook}) \leq \hat{L}(\text{Temperature})$ is 0.267 , $\hat{L}(\text{Humidity}) \leq \hat{L}(\text{Temperature})$ is 0.498 and $\hat{L}(\text{Wind}) \leq \hat{L}(\text{Temperature})$ is 0.502 . We can see some values in the Temperature column are not greater than 0.5 . The probability for $\hat{L}(\text{Temperature}) \leq \hat{L}(\text{Outlook})$ is 0.116 , $\hat{L}(\text{Humidity}) \leq \hat{L}(\text{Outlook})$ is 0.871 and $\hat{L}(\text{Wind}) \leq \hat{L}(\text{Outlook})$ is 0.812 . We can see some values in the Outlook column are not greater

than 0.5. In conclusion, there is not a dominating attribute with $r = 3$. Thus, we cannot select the attribute for a root node of decision tree and cannot construct the tree. For $r = 4, 5$ and 6 , we still cannot construct the tree with the same reason of $r = 3$.

TABLE III
THE PROBABILITY OF $P(x < y)$ OF $\hat{L}^{(k)}$

$x \setminus y$	r	Outlook	Temperature	Humidity	Wind
Outlook	0.5	—	0.500	0.500	0.484
Temperature		0.279	—	0.500	0.488
Humidity		0.500	0.500	—	0.484
Wind		0.516	0.512	0.516	—
Outlook	1	—	0.507	0.508	0.496
Temperature		0.301	—	0.498	0.489
Humidity		0.492	0.502	—	0.490
Wind		0.504	0.511	0.510	—
Outlook	2	—	0.476	0.485	0.500
Temperature		0.335	—	0.513	0.524
Humidity		0.515	0.487	—	0.511
Wind		0.500	0.476	0.489	—
Outlook	3	—	0.267	0.129	0.188
Temperature		0.116	—	0.502	0.498
Humidity		0.871	0.498	—	0.496
Wind		0.812	0.502	0.504	—
Outlook	4	—	0.244	0.067	0.166
Temperature		0.107	—	0.492	0.490
Humidity		0.933	0.508	—	0.498
Wind		0.834	0.510	0.502	—
Outlook	5	—	0.252	0.063	0.154
Temperature		0.102	—	0.495	0.489
Humidity		0.937	0.505	—	0.492
Wind		0.846	0.511	0.508	—
Outlook	6	—	0.252	0.070	0.162
Temperature		0.098	—	0.497	0.492
Humidity		0.930	0.503	—	0.494
Wind		0.838	0.508	0.506	—

Table IV, V and VI demonstrate the component of $P(x < y)$: $P(x < y)_I$, $P(x < y)_P$ and $P(x < y)_F$. The probability values in these tables show that we cannot neglect any component in the determination of the dominant attribute. For example, table V indicates that Temperature should be selected in any situation. It is clearly not acceptable according to table II.

Obviously, with the proposed LAIVFDT, data with uncertain fuzzy membership values could be adopted to construct a fuzzy decision tree. Therefore, a precise fuzzy membership is not a precondition to construct a decision tree anymore. Such relaxation can significantly benefit data mining where precise fuzzy membership values are difficult to get. In this paper, we tested with different r such as $r = 0.5, 1, 2, 3, 4, 5$ and 6 , respectively. The difference of fuzzy decision trees using LAIVFDT and LAFDT are listed as follows:

- 1) We obtain the the smaller decision tree with $r = 0.5$ and 2 in LAIVFDT and type-1 FDT is 3 , which the r value in LAIVFDT is less than r value in LAFDT method.
- 2) If the distance r changes then the dominant attribute is changed. Thus, r is significant is constructing the tree.

TABLE IV
THE PROBABILITY OF $P(x < y)_I$ OF $\hat{L}^{(k)}$

$x \setminus y$	r	Outlook	Temperature	Humidity	Wind
Outlook	0.5	—	0.279	0.456	0.368
Temperature		0.279	—	0.306	0.379
Humidity		0.456	0.306	—	0.404
Wind		0.368	0.379	0.404	—
Outlook	1	—	0.301	0.440	0.346
Temperature		0.301	—	0.342	0.436
Humidity		0.440	0.342	—	0.392
Wind		0.346	0.436	0.392	—
Outlook	2	—	0.335	0.455	0.329
Temperature		0.335	—	0.369	0.476
Humidity		0.455	0.369	—	0.362
Wind		0.329	0.476	0.362	—
Outlook	3	—	0.116	0.129	0.155
Temperature		0.116	—	0.307	0.376
Humidity		0.129	0.307	—	0.408
Wind		0.155	0.376	0.408	—
Outlook	4	—	0.107	0.067	0.134
Temperature		0.107	—	0.285	0.398
Humidity		0.067	0.285	—	0.358
Wind		0.134	0.398	0.358	—
Outlook	5	—	0.102	0.063	0.137
Temperature		0.102	—	0.269	0.374
Humidity		0.063	0.269	—	0.359
Wind		0.137	0.374	0.359	—
Outlook	6	—	0.098	0.070	0.129
Temperature		0.098	—	0.276	0.379
Humidity		0.070	0.276	—	0.365
Wind		0.129	0.379	0.365	—

TABLE V
THE PROBABILITY OF $P(x < y)_P$ OF $\hat{L}^{(k)}$

$x \setminus y$	r	Outlook	Temperature	Humidity	Wind
Outlook	0.5	—	0.000	0.000	0.000
Temperature		0.000	—	0.194	0.109
Humidity		0.044	0.000	—	0.000
Wind		0.148	0.000	0.112	—
Outlook	1	—	0.000	0.000	0.000
Temperature		0.000	—	0.156	0.053
Humidity		0.052	0.000	—	0.000
Wind		0.159	0.000	0.118	—
Outlook	2	—	0.000	0.000	0.000
Temperature		0.000	—	0.144	0.015
Humidity		0.060	0.000	—	0.000
Wind		0.171	0.000	0.127	—
Outlook	3	—	0.000	0.000	0.000
Temperature		0.000	—	0.195	0.123
Humidity		0.687	0.000	—	0.000
Wind		0.657	0.000	0.096	—
Outlook	4	—	0.000	0.000	0.000
Temperature		0.000	—	0.207	0.092
Humidity		0.777	0.000	—	0.000
Wind		0.700	0.000	0.144	—
Outlook	5	—	0.000	0.000	0.000
Temperature		0.000	—	0.227	0.115
Humidity		0.780	0.000	—	0.000
Wind		0.710	0.000	0.149	—
Outlook	6	—	0.000	0.000	0.000
Temperature		0.000	—	0.221	0.114
Humidity		0.777	0.000	—	0.000
Wind		0.708	0.000	0.142	—

- 3) LAIVFDT can construct a decision tree using interval-valued fuzzy membership values

TABLE VI
THE PROBABILITY OF $P(x < y)_F$ OF $\hat{L}^{(k)}$

x \ y	r	Outlook	Temperature	Humidity	Wind
Outlook	0.5	—	0.221	0.044	0.117
Temperature		0.000	—	0.000	0.000
Humidity		0.000	0.194	—	0.081
Wind		0.000	0.133	0.000	—
Outlook	1	—	0.206	0.067	0.150
Temperature		0.000	—	0.000	0.000
Humidity		0.000	0.160	—	0.098
Wind		0.000	0.074	0.000	—
Outlook	2	—	0.141	0.030	0.171
Temperature		0.000	—	0.000	0.032
Humidity		0.000	0.119	—	0.149
Wind		0.000	0.000	0.000	—
Outlook	3	—	0.510	0.000	0.033
Temperature		0.000	—	0.000	0.000
Humidity		0.055	0.192	—	0.088
Wind		0.000	0.126	0.000	—
Outlook	4	—	0.137	0.000	0.032
Temperature		0.000	—	0.000	0.000
Humidity		0.090	0.223	—	0.140
Wind		0.000	0.111	0.000	—
Outlook	5	—	0.149	0.000	0.017
Temperature		0.000	—	0.000	0.000
Humidity		0.093	0.236	—	0.133
Wind		0.000	0.137	0.000	—
Outlook	6	—	0.155	0.000	0.033
Temperature		0.000	—	0.000	0.000
Humidity		0.083	0.227	—	0.129
Wind		0.000	0.129	0.000	—

- 4) The decision tree from uncertain membership values is different from that of precise membership values. We cannot take the average value of an interval-valued to construct the tree.

V. CONCLUSION

The original LAFDT method requires precise membership values. Precise membership values are not always available in real world applications. In this paper, we propose LAIVFDT method to apply interval-valued fuzzy sets to construct an interval-valued fuzzy decision tree. In the proposed model, Hamming distance between two interval-valued fuzzy sets is applied to measure the distance between the two instances. A probability model is employed to compare intervals to determine the classifiability of each attribute. A systematic algorithm is established to construct a decision tree from data with uncertain membership values. Our example demonstrates that the proposed method does construct an acceptable decision tree when interval-valued fuzzy membership values are involved in the data set. There are still further work to do in the proposed algorithm, such as the determination of the neighbourhood radius r . We will discuss it in future publications.

REFERENCES

- [1] K.M. Lee, K.M. Lee, J.H. Lee and H.L. Kwang, "A fuzzy decision tree induction method for fuzzy data", *IEEE International Fuzzy Systems Conference Proceedings*, pp. 16-21, 1999.
- [2] P. Grzeorzewski, "Distance between intuitionistic fuzzy sets and/or interval-valued fuzzy set based in the Hausdorff metric", *Fuzzy Sets and Systems*, Vol. 148, pp.319-328, 2004.
- [3] Y. Yang and F. Chiclana, "Intuitionistic fuzzy sets: spherical representation and distance", *International Journal of Intelligent Systems*, Vol. 24, pp.399-420, 2009.
- [4] H. Zhang, W. Zhang and C. Mei, "Entropy of interval-valued fuzzy sets based on distance and its relationship with similarity measure", *Knowledge-Based Systems*, Vol. 22, pp.449-454, 2009.
- [5] S.Ferson, J. Hajagos, D. Berleant, J. Zhang, W. T. Tucker, L. Ginzburg and W. Oberkamf, "Dependence in Dempster-Shafer theory and probability bounds analysis", *Applied Biomathematics*, U.S.A., pp.1-138, 2004.
- [6] M. Schneider, A. Kandel, G. Langholz and G. Chew, "Fuzzy Expert System Tools", John Wiley & Sons Ltd., England, 1996.
- [7] Y. Yuan and M. J. Shaw, "Induction of fuzzy decision tree", *Fuzzy Sets and Systems*, Vol. 69, pp.125-139, 1995.
- [8] M. Dong and R. Kothari, "Look-ahead based fuzzy decision tree induction", *IEEE Transactions on Fuzzy Systems*, Vol 9, No 3, pp. 461-468, 2001.
- [9] M. Dong, R. Kothari, and M. Visscher and S. B. Hoath, "Evaluating skin using a new decision tree induction algorithm", *Proceedings of the 2001 International Joint Conference on Neural Networks*, Vol. 4, pp. 2456-2460, 2001.
- [10] X. Wang, B. Chen, G. Qian and F. Ye, "On the optimization fuzzy decision tree", *Fuzzy Sets and Systems*, Vol. 112, pp. 117-125, 2000.
- [11] K. T. Atanassov and G. Gargov, "Interval valued intuitionistic fuzzy sets", *Fuzzy Sets and Systems*, Vol. 31, No. 3, pp.343-349, 1989.
- [12] C. Y. C. ChiangLin, "Comparison between Crisp and Fuzzy Stock Screening Models", [Online]. Available: http://www.atlantipress.com/php/download_paper.php?id=33.
- [13] J. Zeidler and M. Schlosser, "Continuous-valued attributes in fuzzy decision trees", *Proceeding of the 6th International Conference on Information Proceeding and Management of Uncertainty in Knowledge-Based Systems*, pp. 395-400, 1996.
- [14] J. M. Mendel, H. Wu, V. Kreinovich and G. Xiang, "Fast Computation of centroids for constant-width interval-valued fuzzy sets", *Fuzzy Information Processing Society, NAFIPS 2006, Annual Meeting of the North American*, pp. 621-626, 2006.
- [15] J. M. Mendel and R. I. John and F. Liu, "Interval type-2 fuzzy logic systems made simple", *IEEE Transactions on Fuzzy Systems*, Vol.14, No. 6, pp. 808-821, 2006.
- [16] D. Wu, "A vector similarity measure for type-1 fuzzy sets", *Lecture Note In Artificial Intelligence, Proceeding of the 12th International Fuzzy Systems Association World Congress on Foundations of Fuzzy Logic and Soft Computing*, Vol. 4529, pp.575-583, 2007.
- [17] K. McKarty and M. Manic, "Contextual fuzzy type-2 hierarchical for decision trees (CoFuH-DT) - An accelerated data mining technique", *Conference on Human Systems Interactions*, pp. 699-704, 2008.