SIMPLE FUZZY LOGIC RULES BASED ON FUZZY DECISION TREE FOR CLASSIFICATION AND PREDICTION PROBLEM

J. F. Baldwin and Dong (Walter) Xie
Department of Engineering Mathematics, Faculty of Engineering, University of Bristol
Bristol, BS8 1TR, United Kingdom
Jim.Baldwin@bristol.ac.uk *D.Xie@bristol.ac.uk*

Abstract: In data mining for knowledge explanation purposes, we would like to build simple transparent fuzzy models. Compared to other fuzzy models, simple fuzzy logic rules (IF ... THEN... rules) based on triangular or trapezoidal shape fuzzy sets are much simpler and easier to understand. For fuzzy rule based learning algorithms, choosing the right combination of attributes and fuzzy sets which have the most information is the key point to obtain good accuracy. On the other hand, the fuzzy ID3 algorithm gives an efficient model to select the right combinations. We therefore discover the set of simple fuzzy logic rules from a fuzzy decision tree based on the same simple shaped fuzzy partition, after dropping those rules whose credibility is less than a reasonable threshold, only if the accuracy of the training set using these rules is reasonably close to the accuracy using fuzzy decision tree. The set of simple fuzzy logic rules satisfied with this condition is also able to be used to interpret the information of the tree. Furthermore, we use the fuzzy set operator "OR" to merge simple fuzzy logic rules to reduce the number of rules.

Key words: Simple shaped fuzzy partition, fuzzy ID3 decision tree, simple fuzzy logic rules (SFLRs), classification problem, prediction problem.

1. INRODUCTION

The classification and prediction problems, where the target attribute is respectively discrete (nominal) or 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.

To be able to handle imprecision and uncertainty of the representation of concepts and words in the real world, these models have been used with fuzzy logic [3] introduced by Zadeh in 1965. These fuzzy models overcome the sharp boundary problems [5], providing a soft controller surface and good accuracy in dealing with continuous attributes and prediction problem.

In classification and prediction problems, we would like the fuzzy model to be as simple as possible and provide an easy means of providing an explanation for the result. The fuzzy logic rules (IF ... THEN... rules) are good choice, because they are not only much simpler than the other models but also formulate human reasoning and decision-making into a set of easily understandable linguistic clauses. For explanation purpose, we have to use the simple triangular or trapezoidal shape fuzzy sets, so that simple fuzzy logic rule model based on these fuzzy sets are produced.

In order to use less number of simple fuzzy logic rules to provide reasonable accuracy, we firstly discover a fuzzy ID3 decision tree with post-pruning [2][4] based on the simple triangular fuzzy sets, and transfer the tree into a set of simple fuzzy logic rules after dropping those rules whose credibility is less than a reasonable threshold, only if the accuracy of the training set using simple fuzzy logic rules is reasonably close to the accuracy using fuzzy decision tree.

In Fril [1], a symbolic AI uncertainty logic programming system combining fuzzy reasoning, possibility and probability reasoning, we interpret the simple fuzzy logic rules as conditionalisations rather than as implications. Defuzzification in Fril takes a very simple form.

To reduce the complexity of our model, we merge simple fuzzy logic rules with neighbouring fuzzy sets to give trapezoidal fuzzy sets.

2. SIMPLY SHAPED FUZZY PARTITION AND FUZZY ID3 DECISION TREE

2.1 Simply triangular or trapezoidal shape fuzzy sets

When Zadeh proposed fuzzy set theory [3] in 1965, the use of simple linguistic words in place of numbers for computing and reasoning was one of the key ideas. This provides fuzzy logic with a simplified explanation power of being a suitable interface between human users and computing systems. This power does, though, depend on the form of fuzzy sets. If the fuzzy sets are simple triangular or trapezoidal in shape, then they can be given an easy interpretation. If they have a complicated shape, such as Figure 1, they do not provide a useful linguistic description.

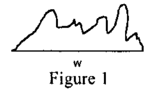

Figure 1

Optimised fuzzy sets, such as neuro-fuzzy sets [1], are used to obtain good accuracy but they have no explanation power because of their complicated shape. We investigate methods of deriving rules and models using a simple shaped fuzzy partition for each attribute, which is defined as a family of triangular or trapezoidal fuzzy sets in Definition 1 such that for any argument value the memberships add to1. [1][2]

Definition 1: A ***simply shaped fuzzy partition*** $\{f_i\}$ is a set of triangular or trapezoidal fuzzy sets such that

$\sum_i \chi_{f_i}(x) = 1$ for any data point x ∈ X where X is the universal set.

2.2 Fuzzy partition model and membership function

In this paper, we use equal data points fuzzy sets (EDP-FS) model in Definition 2 for continuous (numerical) attributes, which are normally asymmetric, and still use crisp sets as a special case of fuzzy sets for discrete (nominal) attributes.

Definition 2: Equal data points fuzzy sets (EDP-FS)

In this model, the number of data examples in each interval covered by a triangular fuzzy set in the universal set [a, b] is equal. For *n* fuzzy sets and *m* examples in database sorted in ascending order, if the value of example *x* is *val* (*x*) where $x \in [1, m]$, then the fuzzy partition is illustrated in Figure 2.

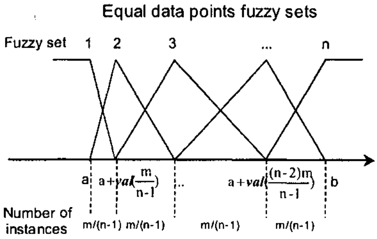

Figure 2

Mass Assignment Theory [1][2] proposed by Baldwin in 1991 integrated fuzzy logic and probability theory, points out that for simple shaped fuzzy partition $\{f_i\}$ of *kth* attribute, input such as x = g, where g can be point value, fuzzy set or probability distribution, is translated into distribution over fuzzy sets of words using membership function χ_f: X → [0, 1] [3]. The membership values $\chi_f(x)$ where x ∈ X is the conditional probability of each fuzzy set given input Pr (f_i | g). [1][2]

Definition 3: ***Membership value*** $\chi_{f_i}(x)$ = Pr (f_i | g), where if g is the point value, then Pr (f_i | g) $\chi_{f_i}(g)$ = , otherwise we will use point value semantic unification [1][2] to calculate it.

2.3 Fuzzy ID3 decision tree

Fuzzy ID3 algorithm [2][4] developed by Baldwin and co-workers and described below is an efficient algorithm to generate fuzzy decision tree.

```
Input 3 parameters of model: training set S, the number of fuzzy sets $f$ and the depth of decision tree $l$,
Start to form fuzzy decision tree from the top level,
Do loop until (1) the depth of the tree gets to $l$ or (2) there is no node to expand
    a) Determine expected entropy $\mathit{E}\mathrm{E}(A_k)$ for each attribute of S not already expanded in this branch,
    b) Expand the attribute x with the minimum expected entropy $\mathit{E}\mathrm{E}(A_x)$,
    c) Stop expansion of the leaf node $A_{kf}$ of attribute k if entropy $E(A_{kf}) = 0$ or nearly 0,
    d) Use post pruning to prune the tree and stop.
End do loop
```

During the process of learning fuzzy decision tree, the leaf nodes A_{kf} in each stage have the entropy

$$E(A_{kf}) = \sum_i \Pr(t_i \mid A_{kf}) \times \mathrm{Ln}(\Pr(t_i \mid A_{kf}))$$

where the node belongs to the *kth* attribute and *fth* fuzzy set, and t_i is *ith* class or fuzzy set of the target attribute. $\Pr(t_i \mid A_{kf})$ is the conditional probabilities associated with each class in the target attribute.

Definition 4: For the *kth* attribute, the ***expected entropy*** is

$$\mathit{E}\mathrm{E}(A_k) = \sum_f \Pr(A_{kf}) \times E(A_{kf})$$

where the renormalized ***branch probability*** passed in each branch is

$$\Pr(A_{kf}) = \mathit{ReNorm}(\sum_T \sum_{A_{1f}} \cdots \sum_{A_{(k-1)f}} \sum_{A_{(k+1)f}} \cdots \sum_{A_{nf}} \Pr(A_{1f\ldots,} A_{nf}, T))$$

where the subscript f could be the different number of fuzzy sets in each attribute and the set of nodes $\{A_{1f},\ldots,A_{kf},\ldots,A_{nf},T\}$ comprises the branch which is the path of the target T.

We modify Laplace's formula to prune the fuzzy decision tree. The error of the *fth* children node S_f of any node S in fuzzy decision tree is

$$\mathrm{Error}(S_f) = \frac{N \times \Pr'(S_f) - N \times \Pr'(S_f) \times \Pr(t_i \mid S_f) + k - 1}{N \times \Pr'(S_f) + k}$$

where N is the number of examples in the training set, and $\Pr'(S_f)$ is the probability passed in the branch before renormalization in Definition 4. Then, we calculate backup error of node S. If BackUpError(S) $\geq$ Error(S), the tree is pruned by halting at S and cutting all its children nodes. [2]

3. SIMPLE FUZZY LOGIC RULES BASED ON FUZZY DECISION TREE

For machine learning and data mining purpose, various types of fuzzy rules and models can be used, such as general Fril rules [1], fuzzy decision trees, fuzzy Bayesian nets and IF...THEN...fuzzy logic rules. Depending on the simply shaped fuzzy sets, fuzzy logic rules provide a simple transparent formulation of human reasoning and hence can be explained easily. Though those rules over optimised fuzzy sets would provide good accuracy, they lose their main advantage of fuzzy logic rules in original. We therefore only use those over simply triangular or trapezoidal shape fuzzy sets that are called as simple fuzzy logic rules in this paper.

3.1 Simple fuzzy logic rules (SFLRs)

Suppose there are k attributes and the *jth* attribute is the target attribute, the simple fuzzy logic rule (SFLR) based on the simple shaped fuzzy sets is of the form shown in (1):

$$\begin{array}{l}(A_j \text{ is large}) \text{ IF} \\ \quad (A_1 \text{ is small}) \text{ AND} \ldots \text{ AND } (A_{j-1} \text{ is small}) \text{ AND } (A_{j+1} \text{ is medium}) \text{ AND} \ldots \text{ AND } (A_k \text{ is large})\end{array} \tag{1}$$

where the term on the left side of IF is the head of this rule and the set of terms on the right is the body, and the clauses of the terms are words of attributes defined by fuzzy sets. Every SFLR has support and credibility defined as below.

Definition 5: The joint probability p_r = Pr(A_1 is small $\wedge \ldots \wedge$ A_{j-1} is small $\wedge$ A_j is large $\wedge$ A_{j+1} is medium $\wedge \ldots \wedge$ A_k is large) is the ***support*** of the simple fuzzy logic rule in (1). [2]

The support of a SFLR represents the frequency of occurrence of the particular combination of attribute values in the training set.

Let $\boldsymbol{p}$ = $\Pr(A_1 \text{ is small} \wedge \ldots \wedge A_{j-1} \text{ is small} \wedge A_{j+1} \text{ is medium} \wedge \ldots \wedge A_k \text{ is large})$

$= \sum_{A_j \text{ is large}} \Pr(A_1 \text{ is small} \wedge \ldots \wedge A_k \text{ is large})$, then

Definition 6: the value of $\frac{p_r}{p}$ is the ***credibility*** (confidence) of the simple fuzzy logic rule in (1). [2]

The credibility of a SFLR represents how often it is likely to be true.

Only the SFLRs whose credibility is greater than or equal to the credibility threshold ε are chosen. Those SFLRs are likely to be true, if ε is reasonably high.

3.2 Simple fuzzy logic rules from fuzzy ID3 decision tree

All kinds of decision trees can be changed into IF...THEN...rules. In our model, fuzzy ID3 decision tree is transferred into a set of SFLRs with one of the model parameters --- a credibility threshold ε. The head of a SFLR is the class or fuzzy set of a leaf node with maximum conditional probability, and this conditional probability is equivalent to the credibility of the SFLR transferred. The body is the path of this target in the tree. Any SFLR whose credibility is less than ε is dropped.

For example, in the Pima Indian Diabetes classification problem, we discovered a fuzzy decision tree [2][4] shown in Figure 3 using 3 fuzzy sets defined in Definition 2 in each attribute and assigning the depth of tree as 3, where each pair of integer numbers in a bracket on the left side represents a node in the tree, and the first number represents an attribute number and the second represents a selected fuzzy set of this attribute. Those float numbers on the right side are the conditional probabilities associated with each class in the target attribute. For instance, the first path of the tree shows the node with 2nd attribute 1st fuzzy set leads to the target node with probability equal to 0.862725 for the 1st class and 0.137275 for the 2nd class.

Fuzzy decision tree:

1.	(2 1)	(0.862725 0.137275)
2.	(2 2)	(0.668541 0.331459)
3.	(2 3)(8 2)	(0.308811 0.691189)
4.	(2 3)(8 3)	(0.275786 0.724214)
5.	(2 3)(8 1)(7 1)	(0.814304 0.185696)
6.	(2 3)(8 1)(7 2)	(0.525588 0.474412)
7.	(2 3)(8 1)(7 3)	(0.236044 0.763956)

Credibility >= 0.6

Simple fuzzy logic rules (SFLRs):

1. (attribute_9 is class_1) IF (attribute_2 is fuzzy_set_1)

2. (attribute_9 is class_1) IF (attribute_2 is fuzzy_set_2)

3. (attribute_9 is class_2) IF
(attribute_2 is fuzzy_set_3) AND (attribute_8 is fuzzy_set_2)

4. (attribute_9 is class_2) IF
(attribute_2 is fuzzy_set_3) AND (attribute_8 is fuzzy_set_3)

5. (attribute_9 is class_1) IF
(attribute_2 is fuzzy_set_3) AND (attribute_8 is fuzzy_set_1)
AND (attribute_7 is fuzzy_set_1)

7. (attribute_9 is class_2) IF
(attribute_2 is fuzzy_set_3) AND (attribute_8 is fuzzy_set_1)
AND (attribute_7 is fuzzy_set_3)

Figure 3

As we can see in Figure 3, these SFLRs with information of fuzzy ID3 decision tree have the simpler form and are easier to understand than the decision tree.

Our model has two uses: one is to efficiently discover a set of SFLRs with good accuracy, when their training set accuracy is reasonably close to the fuzzy decision tree; the other is to use the set of SFLRs transferred from fuzzy decision tree to interpret the information of the tree, if the training set accuracy of SFLRs is reasonably close to the accuracy of fuzzy decision tree.

3.3 Using simple fuzzy logic rules (SFLRs) to evaluate new case for classification or prediction problem

In Fril we interpret the IF...THEN... simple fuzzy logic rules as conditionalisations rather than as implications [1]. This makes more sense when uncertainties are involved since it is the conditional probabilities that are naturally apparent in data.

To evaluate a new example $\{x_1,...,x_k\}$ over k attributes using the SFLR in (1), we calculate Pr (body) = Pr (small | x_1) × ... × Pr (small | x_{j-1}) × Pr (medium | x_{j+1}) × ... × Pr (large | x_k) over k-1 attributes in the body. Let Pr (body) = ϕ, Fril inference of the SFLR is formulated as Pr (head) = Pr (head | body) × Pr (body) + Pr (head | $\neg$ body) × Pr ($\neg$ body) = $1 \times \phi + [0, 1] \times (1 - \phi) = [\phi, 1]$ [1][2]

Definition 7: For $\{t_i\}$ classes in the target attribute, the Fril inference of SFLRs will give $\{t_i : [\phi_i, 1]\}$ for each class. We therefore choose the class t_i of the target attribute as the predicted class that has maximum inference $\underset{t_i}{MAX}\ (\phi_i)$.

Definition 8: For $\{f_i\}$ fuzzy sets in the target attribute, the Fril inference of SFLRs will give $\{f_i : [\phi_i, 1]\}$ for each fuzzy set, where there is $\sum_i \phi_i \le 1$. Let $\chi_{f_i}(m_i) = 1$, the predicted value is:

$$x = (x_u + x_l)/2, \text{ where}$$

$$x_u = \underset{\{\theta_i\}}{MAX} \sum_i m_i \theta_i \ s.t\ \phi_i \le \theta_i \le 1\ (all\ i), \sum_i \theta_i = 1$$

$$x_l = \underset{\{\theta_i\}}{MIN} \sum_i m_i \theta_i \ s.t\ \phi_i \le \theta_i \le 1\ (all\ i), \sum_i \theta_i = 1$$

The predicted value x equals to the average of possible maximum value x_u of x and possible minimum value x_l. In the formula, we take the maximum value m_i of each fuzzy set i $(\chi_{f_i}(m_i) = 1)$ that is multiplied by the probability distribution θ_i associated with *ith* fuzzy set. To make x_u maximal, we keep probability distribution θ as much as possible in the fuzzy set whose maximum value m is maximal among all fuzzy sets, and then the rest of θ_i is assigned to the associated Fril inference ϕ_i. Vice versa for x_l.

For example, suppose we have Firl inferences $\{f_{small} : [0.2, 1], f_{medium} : [0.5, 1], f_{large} : [0.1, 1]\}$, where $\chi_{small}(1) = 1$, $\chi_{medium}(5) = 1$, and $\chi_{large}(9) = 1$.

Then $x_u = 1 \times 0.2 + 5 \times 0.5 + 9 \times (1 - 0.2 - 0.5) = 5.4$ and $x_l = 1 \times (1 - 0.5 - 0.1) + 5 \times 0.5 + 9 \times 0.1 = 3.8$. The predicted value $x = (5.4 + 3.8) / 2 = 4.6$. This provides the defuzzification.

3.4 Merging simple fuzzy logic rules (SFLRs)

The number of rules is the main measurement of the complexity of rule-based model. To reduce the number of SFLRs, we merge those rules where their heads are the same, and fuzzy sets separately in one term of their bodies are neighbouring but the other terms are same, by using fuzzy set operator "OR". The Fril inference of one merged rule equals the sum of inferences of those rules before mergence shown in Figure 4, because of Definition 1. The mergence therefore would not affect the accuracy of classification or prediction at all.

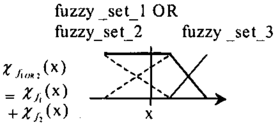

Figure 4

For example, the first 4 SFLRs in Figure 3 can be merged into two rules: (attribute_9 is class_1) IF (attribute_2 is [fuzzy_set_1 OR fuzzy_set_2]) and (attribute_9 is class_2) IF (attribute_2 is fuzzy_set_3) AND (attribute_8 is [fuzzy_set_2 OR fuzzy_set_3]). In result, the number of SFLRs in Figure 4 is reduced from 6 into 4.

4. EXPERIMENTS

To evaluate our models, we choose some typical databases in UCI Machine Learning Repository [6] to separate each of them into training set and test set by selecting database examples randomly, and then use the same training set and test set to get the accuracy each time.

In the following tables, "Number of Fuzzy Sets" represents the number of fuzzy sets we used for each attribute of database. "Depth of Tree" and "Credibility Threshold" are also model parameters mentioned in section 2.3 and 3.2. "Number of Leaf Nodes" and "Number of Rules" respectively show the complexity of fuzzy decision tree and SFLRs, where the number on the left of "→" is the number of SFLRs before merging and one on the right is after merging. The accuracy for "Training set" and "Test set" using fuzzy ID3 decision tree and SFLRs are in the percentage format and respectively calculated in (2) or (3) for classification or prediction problems:

$$\text{Accuracy} = \frac{\text{the number of successfully classified instances}}{\text{the number of instances in total in the dataset}} \qquad (2)$$

$$\text{Accuracy} = 1 - \frac{|T_{\text{predicted}} - T_{\text{original}}|}{\textit{range}\ (\text{T})} \qquad (3)$$

where $T_{\text{predicted}}$ is the predicted target value, T_{original} is the original target value in the dataset, and *range*(T) is the range of the target attribute T.

In the bottom of tables, we use models of Weka [7] to compare our model in classification problem using the same training set and test set.

Tree examples in classification are shown in Table 1, 2, 3 below.

Table 1. Pima Indians Diabetes Database

Number of Fuzzy Sets	Depth of Tree	Leaf Nodes	Training Set (ID3)	Test Set (ID3)	Credibility Threshold	Number of Rules	Training Set (SFLRs)	Test Set (SFLRs)
3	1	3	71.09 %	76.82 %	0.6	3 → 3	71.09 %	76.82 %
3	5	9	72.13 %	75.52 %	0.6	9 → 6	71.35 %	75.78 %
4	5	19	75.26 %	78.64 %	0.6	18 → 11	75.26 %	78.64 %
5	5	69	80.21 %	79.17 %	0.6	64 → 41	78.38 %	79.43 %
Weka J48 (C4.5 decision tree)					7 leaf nodes*		76.30 %	78.12 %
Weka Naïve Bayes							74.22 %	77.60 %
Weka Neural Network					7 nodes**		80.47 %	77.60 %

Table 2. Sonar Data

Number of Fuzzy Sets	Depth of Tree	Leaf Nodes	Training Set (ID3)	Test Set (ID3)	Credibility Threshold	Number of Rules	Training Set (SFLRs)	Test Set (SFLRs)
3	5	37	90.38 %	87.50 %	0.6	37 → 29	89.42 %	84.61 %
5	5	65	93.27 %	71.15 %	0.6	62 → 33	92.31 %	71.15 %
Weka J48 (C4.5 decision tree)					8 leaf nodes*		97.11 %	74.03 %
Weka Naïve Bayes							75.96 %	73.08 %
Weka Neural Network					33 nodes**		100 %	84.61 %

Table 3. Vision Data

Number of Fuzzy Sets	Depth of Tree	Leaf Nodes	Training Set (ID3)	Test Set (ID3)	Credibility Threshold	Number of Rules	Training Set (SFLRs)	Test Set (SFLRs)
4	5	28	65.75 %	65.78 %	0.4	27 → 15	67.18 %	68.25 %
5	5	37	66.98 %	67.15 %	0.4	33 → 14	65.42 %	65.97 %
6	5	71	68.41 %	69.07 %	0.4	56 → 23	68.04 %	68.86 %
Weka J48 (C4.5 decision tree)					625 leaf nodes*		91.45 %	73.18 %
Weka Naïve Bayes							49.36 %	50.47 %
Weka Neural Network					20 nodes**		76.47 %	75.69 %

* The tree is pruned by using Weka's default pruning. [7]

** Those nodes include all of nodes (internal and external nodes) in the neural network.

Table 4 is a example in prediction, where the 3-attribute and 529-data-point training set is created by function Z = Sin(X*Y) plotted in Figure 6, but the test set has 2209 data points. "5-6-6" in Table 4 represents using 5 fuzzy sets in the target attribute and 6 fuzzy sets in the other attributes.

Table 4. Function SinXY

Number of Fuzzy Sets	Depth of Tree	Leaf Nodes	Training Set (ID3)	Test Set (ID3)	Credibility Threshold	Number of Rules	Training Set (SFLRs)	Test Set (SFLRs)
5-6-6	2	31	90.66 %	90.68 %	0.3	29 → 17	92.35 %	92.40 %
5-8-8	2	57	94.54 %	94.46 %	0.3	57 → 34	96.25 %	96.42 %
5-10-10	2	82	95.36 %	95.33 %	0.3	82 → 44	95.69 %	95.90 %
5-12-12	2	122	97.16 %	97.03 %	0.3	122 → 55	96.02 %	96.44 %
13-13-13	2	145	97.89 %	97.66 %	0.2	145 → 107	98.02 %	98.04 %

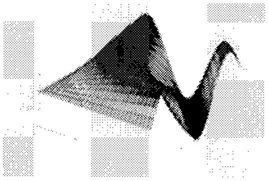

Figure 6: Original Function

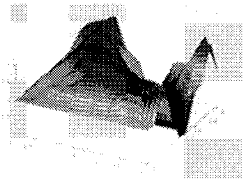

Figure 7: Test Set (SFLRs) 5-10-10

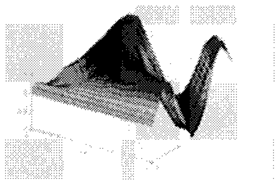

Figure 8: Test Set (SFLRs) 13-13-13

As we can see, if the training set accuracy using the set of SFLRs transferred from fuzzy decision tree is reasonably close to the training set accuracy using decision tree, the test set accuracy using SFLRs is reasonably close to the other or even better than it.

Furthermore, comparing other models, the SFLRs based on decision tree have a reasonable accuracy with less complexity (the number of rules). Considering their advantages of simplicity, transparency, and linguistic explanation power ability, it is one of most useful models in data mining and machine learning.

REFERENCES

1. J. F. Baldwin, T. P. Martin and B. W. Pilsworth, "Fril – Fuzzy and Evidential Reasoning in Artificial Intelligence", Research Studies Press Ltd. (John Wiley), 1995.
2. Jim Baldwin, http://www.enm.bris.ac.uk/teaching/enjfb/emat31600/
3. Lotfi A. Zadeh, George J. Klir, "Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh", World Scientific Publishing Company, May 1996.
4. Jim Baldwin, Sachin Karale, "New Concept of Fuzzy Partition, Defuzzification and Derivation of Probabilistic Fuzzy Decision Trees", Proceedings of the 2003 UK Workshop on Computational Intelligence.
5. Susan M. Bridges, Rayford B. Vaughn, "Fuzzy Data Mining and Genetic Algorithms Applied to Intrusion Detection", 23rd National Information Systems Security Conference, 2000.
6. UCI Machine Learning Repository, http://www1.ics.uci.edu/~mlearn/MLSummary.html
7. University of Waikato, Weka 3.4, http://www.cs.waikato.ac.nz/ml/weka/