Strategies for Fuzzy Inference within Classifier Systems

Keeley A Crockett, Zuhair Bandar, David Mclean

The Intelligent Systems Group, The Manchester Metropolitan University,
Chester Street, Manchester, UK, M15 GD.
Email: K.Crockett@mmu.ac.uk

Abstract

When designing any type of fuzzy rule based system, considerable effort is placed in identifying the correct number of fuzzy sets and the fine tuning of the corresponding membership functions. Once a rule base has been formulated a fuzzy inference strategy must be applied in order to combine grades of membership. Considerable time and effort is spent trying to determine the number of fuzzy sets for a given system while substantially less time is invested in obtaining the most suitable inference strategy. This paper investigates a number of theoretical proven fuzzy inference strategies in order to assess the impact of these strategies on the performance of a fuzzy rule based classifier system. A fuzzy inference framework is proposed, which allows the investigation of five pure theoretical fuzzy inference operators in two real world applications. An additional two novel fuzzy-neural strategies are proposed and a comparative study is undertaken. The results show that the selection of the most suitable inference strategy for a given domain can lead to a significant improvement in performance.

1. Introduction

Fuzzy inference has been applied to numerous control and classification problems in order to provide the mechanism for aggregating the rule strengths in any type of rule-based system. In particular, the application of fuzzy theory to classification problems has generally proved successful by allowing overlapping class definitions which consequently results in less uncertainty. It is often recognized that the key to a successful fuzzy classifier is the correct determination of the membership functions through careful partitioning of the input-output space. Early fuzzy systems used human experts in the specified domain to define sets of membership functions and rules. However, there was no way to assess whether they were a correct representation of the sample training set. Recently the task of generating fuzzy rules without the use of domain experts has been investigated. Methods include the use of clustering techniques, genetic algorithms and neural networks (Bouchon-Meunier 1997; Castellano and Fanelli 2000; Lee and Narayanan 2003; Zadeh 1992; Zadeh 1994).

Once a fuzzy rule-base has been determined, a way of conducting inference is required. Inference is a way in which a conclusion is drawn from a sample set of data and a collection of rules. Zadeh first formulated a set of inference operators, which could be applied to fuzzy sets (Zadeh 1965). These were union (represented by the Maximum operator) and intersection (Minimum operator). In most fuzzy systems, a variant of Zadeh's original operators are often used, i.e. *product* composition for intersection and *min* for union (Zadeh 1992; Klir and Folder 1988; Lee and Narayanan 2003). Whereas time is often invested in tuning membership grades to give the system optimal performance, little thought is given to the selection of the most suitable set of inference operators for a specific application domain.

Significant literature (Dubois and Prade 1982; Dombi 1982; Klir and Folder 1988; Novak and Pedrycz 1990; Yager 1980; Zadeh 1965) has been published concerning the appropriate definitions for intersection and union of fuzzy sets. The criterion for choosing a particular set of these inference operators over another includes examining the domain of the chosen application and the ability to capture formal properties. Other issues to consider are the accuracy of the fuzzy model, the degree of simplicity and the type of hardware implementation.

This paper investigates a number of theoretical proven fuzzy inference strategies within the context of a data driven fuzzy rule based classifier system. In addition two novel fuzzy-neural inference strategies are proposed and compared with the pure fuzzy inference strategies.

2. Components of a Fuzzy Rule-based Classifier

Let F be a single output fuzzy classifier system in an n-dimensional input space $[0,1]^n$. The training set for the system consists of a series of input-output pairs, *i*. Then

$$F= \{(x_i;y_i) \mid i = 1,2,..n\} \tag{1}$$

Where $x_i = (x_{i1}, x_{i2} \dots x_{in})$ is the input vector of the ${}_ith$ input-output pair and y_i is the corresponding output.

2.1 Fuzzification Interface

Fuzzification is the process of converting crisp values into their fuzzy representations. The construction of a fuzzy rule base requires the identification of all attributes and outcome variables within the system which need to be represented as fuzzy sets. Each attribute and variable is then decomposed into a series of fuzzy regions. Each region is represented by a fuzzy set with a linguistic identifier. They collectively are referred to as the term set for a particular variable. The proposed fuzzy rule base classifier system can incorporate both linear and non-linear membership functions in order to define each fuzzy set. For the purpose of this work linear membership functions were used to enable the focus to be on establishing the gains that could be made by applying different inference strategies. Within this generalised framework, a Genetic Algorithm (GA) was used to search for these boundaries for each domain (Goldberg 1998). Figure 1 illustrates how it is possible for a GA to tune the positions and size of a fuzzy set young, by directing points (a,b) with the fundamental constraint (a < b) to produce (a',b') for a linear increasing membership function.

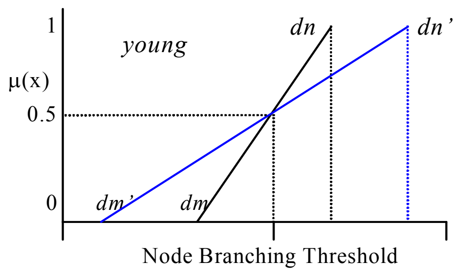

Figure 1.

The membership functions of each rule were encoded onto a chromosome. Each gene represented a real value used in the determination of one membership function domain delimiter (dm_i, dn_i). Hence one chromosome will represent one possible set of membership functions.

2.2 Rule Base Generation

The rule base consists of a series of fuzzy IF-THEN rules and a coterie of fuzzy sets. A typical simplified rule has the form:

$R_{j1..jn}$: IF x_1 is $A1_{j1}$ ..... AND X_n is An_{jn} THEN y is $b_{j1.jn}$,
Where

$j_1 = 1,2,...K1;\quad j_n = 1,2,...,K_n$

Where $R\, j_1..j_n$ is the label of each fuzzy if-then rule,
$b\, j_1..j_n$, is the consequent real number and
K is the number of fuzzy sub-spaces. (2)

These simplified fuzzy rules produce a singleton as the output however the majority of fuzzy models usually produce a fuzzy region. In this proposed generalized framework the fuzzy rule base is constructed by producing a C4.5-type crisp binary decision tree (Quinlan 1993; Quinlan 2002) and transforming each tree into a set of rules using a one-to-one mapping. Initial membership functions are selected to be equivalent to crisp sets with the initial fuzzy classifier system becoming an alternative representation of the decision tree.

For example, Figure 2 shows a C4.5-type highly optimised crisp decision tree, which was generated, from the binary outcome Diabetes in Prima Indians data set (Sigillito 2002). Each leaf node has an associated leaf probability which represents the probability that an example reaching a leaf node will have the same outcome as the leaf. The probability of the dominant outcome is defined as

$$P = \frac{C_d.W_d.NF_d}{\sum_{i=1}^{n} C_i.W_i.NF_i} \qquad (3)$$

where C_i, W_i and NF_i are the frequency, weight and normalisation factor respectively of outcome i.
C_d, W_d and Nf_d are the frequency, weight and normalisation factor respectively of the displayed outcome d,
and n is the number of outcome values.

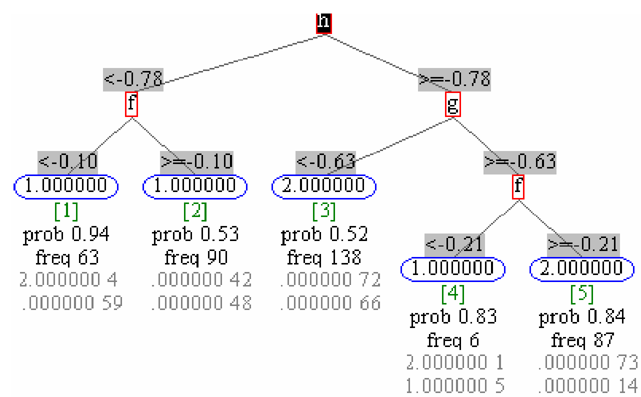

Figure 2.

In the one-to-one mapping the leaf probability becomes the associated rule probability for each given rule and after fuzzification the attribute boundaries identified in the tree rules cease to exist. The tree in figure 2 thus becomes as a one-to-one mapping for the following set of rules (Figure 3) where h, f and g are attributes from the domain.

R1: IF h is LOW and f is LOW THEN outcome is 1 with probability 0.94
R2: IF h is LOW and f is HIGH THEN outcome is 1 with probability 0.53
R3: IF h is HIGH and g is LOW THEN outcome is 2 with probability 0.52
R4: IF h is HIGH and g is HIGH and f is LOW THEN outcome is 1 with probability 0.83
R5: IF h is HIGH and g is HIGH and f is HIGH THEN outcome is 2 with probability 0.84

Figure 3.

2.3 Inference Mechanism

The inference mechanism deduces new conclusions from the given information in the form of fuzzy IF-THEN rules. One form of fuzzy inference is called generalised Modus-Ponens (Pedrycz, 1996) where

Implication : IF x is A THEN y is B
Premise : x is A'
Conclusion : y is B'

Where x and y are linguistic variables and A, A', B, B' are fuzzy sets representing linguistic labels over the universe of discourse, such as small, medium and large.

2.4 Defuzzification Interface

The defuzzification interface performs a mapping from fuzzy output of a Fuzzy Rule Based Classifier System to a non-fuzzy output. If the output is not a singleton, the process may involve performing an operation on an output fuzzy region in order to establish the expected value of the solution variable.

3. Fuzzy Inference Framework

The application of a selected inference strategy is formally defined as follows:
A set of data S will consist of i-attributes $\{A_1,A_2....A_i\}$ of domain D which are used to describe a single object. The process of learning from S involves a transformation function F which accepts as input S and produces a defuzzified outcome O, which is a mapping

$$F(S) \rightarrow O \quad (4)$$

Applying an inference technique to an existing rule base consisting of x rules involves the combination of V membership function values $\{\mu_1, \mu_2,...\mu_v\}$ of all root antecedents. Let T be a set of all possible outcomes $\{t_1,t_2,..t_y\}$ where y is the total number of outcomes.

Fuzzy inference of S will involve an inference mechanism, IM which consists of an intersection function $f_\cap$, which takes in V and produces a set of minimum outcomes Min $\{Min_1,Min_2.......Min_j\}$ where j is the number of rules, and a union function $f_\cup$, which combines output from $f_\cap$ to produce a maximum membership grade O.

Let $f_\cap, f_\cup$, O $\in$ {0.0,1} consisting of real numbers, $\Re$.

- Applying the fuzzy intersection function, $f_\cap$

This involves combining membership grades of the antecedent parts of all rules.

$$f_\cap (\{\mu_1, \mu_2,...\mu_v\}) \rightarrow \text{Min } \{Min_1,Min_2...Min_j\} \quad (5)$$

- Applying rule probabilities

Let P be a set of rule probabilities $\{p_1,p_2...p_y\}$ where y is the total number of outcomes then

$$f_\cap(\{\mu_1, \mu_2,..\mu_v \}) \rightarrow \text{Min}\{(Min_1*p_1),(Min_2*p_2)\ldots(Min_j*p_y)\} \quad (6)$$

Each rule probability is applied to the corresponding membership grade after the intersection operation.

- Applying fuzzy union function $f_\cup$

The fuzzy union operator is applied in-order to combine the membership grades from all rules in order to produce a representative final grade of membership.

$$f_\cup (\{(Min_1*p_1),(Min_2*p_2)......(Min_j*p_y)\}) \rightarrow O \quad (7)$$

O is the fuzzy singleton used to determine the success of a correct classification having taken place for S.

3.1 Pure Fuzzy Inference Operators

For the purposes of this paper, five contrasting fuzzy inference techniques (Table 1) have been selected to combine grades of membership generated by linear membership functions for each attribute featured within the rule base. Zadeh's min-max technique is often used as the standard benchmark inference technique in many fuzzy systems. However, it is sometimes criticized for not allowing interaction of membership grades (Zadeh 1965, Klir and Folder 1988). If the fuzzy subsets are restricted to crisp sets then the operators become the conjunction and disjunction of classical set theory. This is known as the correspondence principle. As the set tends towards fuzziness, different values are obtained for the min and max.

Yager uses parameters: w_u, $w_i \in [0,\infty]$ in order to soften the union and intersection operators and in order to make them more adaptable (Yager 1980). For each variation of the parameters w_u and w_i a different fuzzy intersection or union is obtained. In essence w determines the strength of the operation carried out. Dubois and Prade also offer a general class of fuzzy connectives and make the assumption that the grade of uncertainty of a union of mutually exclusive events can be obtained by combination of grades of uncertainty of each of the events (Dubois and Prade 1982). The strengths / weakness are again determined by two operators α_i, α_u where α_i, $\alpha_u \in [0,1]$.

Two other parameterised inference operators that also offer contrasting weighting schemes are those proposed by Hamacher (Klir and Folder 1988) and Dombi (Dombi 1982). Table 1 defines each set of fuzzy inference operators which are investigated.

Ref	Operators	Range
(Zadeh 1965)	$\mu_{(A\cup B)} = \max[\ \mu(A), \mu(B)\ ]$ $\mu_{(A\cap B)} = \min[\ \mu(A), \mu(B)\ ]$	n/a
(Yager 1980)	$\mu_{(a\cup b)} = \min[1,\ (a^{wu} + b^{wu})^{1/wu}]$ $\mu_{(a\cap b)} = 1-\min[1, ((1-a)^{wi} + (1-b)^{wi})^{1/wi}]$	$w_i, w_u \in (0,\infty)$
(Dubois and Prade 1982)	$\mu_{(a\cup b)} = \dfrac{a+b-ab-\min(a,b,1-\alpha u)}{\max(1-a,1-b,\alpha u)}$ $\mu_{(a\cap b)} = \dfrac{ab}{\max(a,b,\alpha i)}$	$\alpha_i, \alpha_u \in (0,1)$
Hamacher (Klir and Folder 1988)	$\mu_{(a\cup b)} = \dfrac{a+b-(2-qu)ab}{1-(1-qu)ab}$ $\mu_{(a\cap b)} = \dfrac{ab}{qi+(1-qi)(a+b-ab)}$	$q_i, q_u \in (0,\infty)$
(Dombi 1982)	$\mu_{(a\cup b)} = \dfrac{1}{1+\left[\left(\frac{1}{a}-1\right)^{-w}+\left(\frac{1}{b}-1\right)^{-w}\right]^{-1/w}}$ $\mu_{(a\cap b)} = \dfrac{1}{1+\left[\left(\frac{1}{a}-1\right)^{w}+\left(\frac{1}{b}-1\right)^{w}\right]^{1/w}}$	$w_u, w_i \in (0,\infty)$

Table 1. Fuzzy Inference Techniques

3.2 Back-Propagation

Back-Propagation is probably the most commonly used and well-documented FFNN (Feed-Forward Neural Network) training algorithm (Plaut, Nowlan and Hinton 1986; Sethi 1990). The problem of training a BP-FFNN involves finding a mapping that approximately transforms all the input vectors in the training set into their associated class. This is broken down into a set of transformation sub-problems that must be solved at each layer of neurons.

Firstly a FFNN's ability to approximate a function mapping of arbitrary complexity through consecutive spatial transformations is defined. These transformations are implemented using decision regions constructed from the hyperplanes of neuronal layers. A set of data T consists of a number p of tuples of arity 2, t(v,c) where v is a set of n attributes {I1,I2,..,In} from the domain, from which T is taken, that describe a single object or occurrence and c is the associated class or set of outcomes {ζ1, ζ2,.., ζm}. The task of learning from this data set can be considered as finding an approximation F', of a mapping function F which transforms any set of attributes from the domain v, to its corresponding set of outcomes c. The function F may be linearly inseparable and thus require multiple layers of neurons to approximate it. In this case the transformation is broken down into simpler mapping functions where each layer of neurons provides a further abstraction until the overall mapping is accomplished.

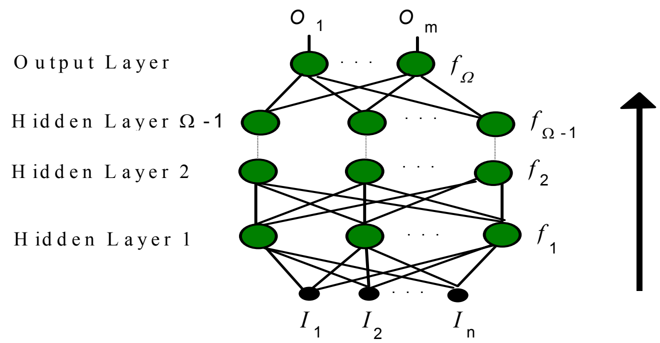

Figure 4. Mapping Functions in a Layered BPNN

Each layer of the network above implements a mapping function fω which performs a transformation on the vectors fed to its neurons. The composition of these functions F', approximately maps all the subsets of input vectors $V_\alpha = \{I_1^\alpha, I_2^\alpha, .., I_n^\alpha\}$ in the training set, to the corresponding output vectors $O_\alpha = \{O_1^\alpha, O_2^\alpha, .., O_m^\alpha\}$ where each $O_i^\alpha \approx \varsigma_i^\alpha$.

At each iteration BP attempts to improve on each of the transformations fΩ to f1 (Figure 4) individually and in that order. Each error-correcting step is based on the error at the outputs of the current layer ω, of neurons and the outputs of the previous layer ω-1. The weights are adjusted so as to improve the transformation fω of the previous layer's outputs (the domain D) to the desired outputs at the current layer (the range R) and thus reduce the error term for each training pattern.

3.2.1 Fuzzy-neural Inference Operators

In this section, a novel hybrid strategy is proposed to combining membership grades uses a fuzzy-neural inference process. This strategy is based around the selection of a fuzzy intersection operator from a predefined fuzzy inference technique (such as those defined in Table 1) and then applying the outputs generated at each leaf node to a Back Propagation Feed Forward Neural Network (BP-FFNN). To summarise, the Fuzzy-Neural inference approach requires:

1. For each input record that is applied to the fuzzy rule base, determine the membership grades of each rule antecedent.
2. Apply a pre-selected fuzzy intersection operator to combine values to produce resultant membership grades.

3. Use the resultant grades as input into a BP-FFNN to provide the classification for any given case.

3.3 Optimisation of Inference parameters

A GA was used to optimize the weights of the union and intersection operators for each of the parameterised inference techniques. In order to accomplish this, the chromosome structure used to encode the membership functions (section 2.1) was extended to allow simultaneous optimization of fuzziness and the strengths of the inference operators.

4. Experimental Results

4.1 Domain

Two real world binary outcome data sets known as 'Mortgage' (Attar, 2002) and 'Diabetes in Pima Indians' (Sigillito, 2002) will be used for the purpose of this investigation. These two data sets were chosen due to their varying size and complexity. The Mortgage data set investigates the possibility of a person acquiring a mortgage and comprises 8611 records featuring 11 discrete and 14 continuous attributes. (4306 representing a Good Risk and 4305 depicting a Bad Risk). The second set, known as 'Diabetes in Pima Indians', investigates whether Pima Indian patients show signs of diabetes, and comprises of 768 records featuring 9 continuous attributes (500 Class 1, indicating that a person has diabetes, 268 Class 2, which represents a person who shows no signs of the disease).

4.2 Methodology

Each technique comprises of two operators, fuzzy union and fuzzy intersection whose strengths of application are determined by two parameters, $p_{\cup}$ and $p_{\cap}$ respectively. For example, every value of $p_{\cup is}$ will generate a different fuzzy union for each inference strategy, *is*. The aim of these experiments was to explore the effect of $p_{\cup is}$ and $p_{\cap is}$ for each inference technique when applied within a real world application. For each data set, ten complete 10-fold cross validations were carried out. For each data set, the training cases were partitioned into 10 equal-sized blocks with similar class distributions. Highly optimized binary C4.5-type trees were first created using the statistical chi-square pruning technique with a significance level of 0.1%. Using a one-to-one mapping each tree was transformed into a fuzzy rule base. Each dataset in turn was then used as test data for the crisp and the five pure fuzzy inference and the two fuzzy-neural strategies generated from the remaining nine blocks. Membership grades would then be combined firstly using pure fuzzy inference and then using the Fuzzy-Neural inference approach. For the purpose of experimentation, Zadeh's and Yager's intersection operators were used in the fuzzy-neural inference strategy. These pure inference operators were chosen to investigate the effect of both non-parameterised and parameterised operators when combined using a BPNN.

4.3 Results

Each table shows firstly, the classification results obtained for the crisp classifier i.e. when initially the membership functions are selected to be equivalent to crisp sets and thus no fuzzification is applied. Secondly the results obtained when using each of the five pure fuzzy inference strategies defined in Table 1, and finally the results of using two different combinations of Fuzzy-Neural inference (FNIA).

INFERENCE	%AVG TEST	% CLASS 1	%CLASS 2
Crisp Classifier	70.0	89.0	52.0
Yager	74.3	73.0	75.5
Zadeh	74.1	81.6	66.7
Dubois/Prade	75.2	72.9	77.4
Hamacher	74.1	72.1	76.1
Dombi	74.5	71.7	77.7
FNIA – Zadeh	76.0	81.0	71.0
FNIA – Yager	75.1	77.2	73.1

Table 2: Diabetes

INFERENCE	%AVG TEST	% CLASS 1	%CLASS 2
Crisp Classifier	67.0	70.0	64.0
Yager	71.4	78.4	64.4
Hamacher	71.2	72.3	69.9
Dombi	71.2	77.7	67.0
Dubois/Prade	70.7	74.2	67.3
Zadeh	70.2	71.2	69.2
FNIA – Zadeh	70.0	78.2	61.8
FNIA – Yager	70.8	72.4	69.2

Table 3: Mortgage

4.4 Evaluation

The results in tables 2 and 3 show that the choice of inference technique and the strength to which it is applied has an effect on the performance of the fuzzy rule-based classifier. Generally, increasing $p_{\cup is}$ resulted in a stronger fuzzy unions whilst increasing $p_{\cap is}$ gave weaker fuzzy intersections. For each technique it is generally found that weak unions coupled with strong intersections gave the best overall performance. Deviations from this pattern caused a decline in the accuracy. The selection of parameter values for each inference operator by the GA was generally found to be domain dependent, however both Hamacher's and Dubois and Prade's techniques gave identical parameter values on both data sets. It is not however, possible to state whether these values will be the same for any other data domains. The results in Tables 2 and 3 show significant improvement in the classification accuracy has been achieved by using each inference strategy within the fuzzy rule-based classifier system. Fuzzification of rules alone provides a reasonable improvement that is reflected in the results obtained using Zadeh's operators with an average 4.1% (Diabetes) and 3.2% (Mortgage) increase respectively when compared with a crisp classifier. The results show that by focusing on optimizing the strength of application of both the union and intersection of the parameterised operators the performance of the classifier system can improve further. The two new strategies of fuzzy-neural inference also show improvements, but the issue of transparency in the decision-making problem remains an issue.

5. Conclusion

This paper has presented a number of strategies for fuzzy inference within rule based classifier systems. A generalized framework for fuzzy inference was proposed which enabled the testing of seven strategies on two real world data sets. A GA was used to determine a number of high performance membership functions and simultaneously optimise the strength controlling union / intersection parameters for each specific inference technique. The results clearly show that the choice of strategy can improve the overall performance of the classifier and justifies the need to investigate the selection of inference parameters when designing fuzzy rule based systems.

References

Attar Software Limited, Newlands House, Newlands Road, Leigh, Lancashire, WN7 4HN, UK.

Bouchon-Meunier, B. Mamdani, R, 1987. Learning From Imperfect Data, *Fuzzy Information Engineering A Guided Tour of Applications*, Eds. Dubois, D. Prade, H. Yager, R, R, John Wiley and Sons : 139-148

Castellano, G. Fanelli, A, M. 2000. A Staged Approach for Generation and Compression of Fuzzy Classification Rules, *The 9th IEEE International Conference on Fuzzy Systems,* Vol 1, 42-47,

Ching-Chang, W. Hsuan-Ming F. 2003. Design of Fuzzy Systems with Fewer Rules, *Cybernetics and Systems* 34(2), 139-160.

Dubois, D. Prade, H. 1982. A Class Of Fuzzy Measures Based On Triangular Norms. *General Systems* 8, 43-61

Dombi, J. 1982. A General Class of Fuzzy Operators, the Demorgan Class of Fuzzy Operators and Fuzziness Measures Induced by Fuzzy Operators, *Fuzzy Sets and Systems* 8.

Goldberg, D. 1998. A Meditation on the Application of Genetic Algorithms. IlliGal Technical Report No. 98003. Solution: http://gal4.ge.uiuc.edu/illigal.home.html.

Janikow, C. 1998 Fuzzy Decision Trees: Issues and Methods, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 28, Issue 1, 1-14.

Kim, C, J. 1997. An algorithmic approach for fuzzy inference, *IEEE Transactions on Fuzzy Systems* vol. 5, no. 4, 585-98.

Klir, G. Folder, T. 1988. Fuzzy Sets, Uncertainty and Information. Prentice Hall.

Lee, C. Narayanan, S. 2003. Emotion Recognition Using a Data-Driven Fuzzy Inference System, *8th European Conference Speech Communication Technology*, Switzerland.

Novak, V. Pedrycz, W. 1990. Fuzzy Sets and T-norms in the Light of Fuzzy Logic. *Machine Learning, Uncertain Reasoning and Knowledge Based Systems* (3), Academic Press.

Pedrycz, W. 1996. Fuzzy Modeling Paradigms and Practice, International Series in Intelligent Technologies.

Plaut D., Nowlan S. 1996. and Hinton G. Experiments on Learning by Back Propagation, *Technical report CMU-CS-86-126,* Dept. of Computer Science, Carnegie Mellon University, Pittsburgh.

Quinlan, J, R. 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.

Quinlan, J.R. 2002. Is C5.0 Better Than C4.5. Available: http://www.rulequest.com/see5-comparison.html

Sethi I. K. 1990. Entropy Nets: From Decision Trees to Neural Networks. *Proceedings Of The IEEE*, 78(10), 1605-1613.

Shyi-Ming Chen, Cheng-Hsuan Kao, Cheng-Hao Yu. 2002. Generating Fuzzy Rules From Training Data Containing Noise For Handling Classification Problems. *Cybernetics and Systems 33*(7), 723-748.

Sigillito, V. 2002. Original Owners: National Institute of Diabetes, Digestive and Kidney Diseases. Donor of Database: Vincent Sigillito (vgs @aplcen.apl.jhu.edu), Research Centre, RMI Group Leader. Applied Physics Laboratory. The Johns Hopkins University, Johns Hopkins Road, Laurel, MD 20707. (301) pp. 953-6231. Actual data source obtained from Statlog Project at ics.uci.edu in directory pub/machine-learning-databases.

Yager,R. 1980. On a General Class Of Fuzzy Connectives. *Fuzzy Sets and Systems* 4.

Zadeh, 1965. L. Fuzzy Sets. *Information and Control 8*, 228-353.

Zadeh, L. 1992. Knowledge Representation In Fuzzy Logic. In *An Introduction To Fuzzy Logic Applications In Intelligent systems*, Eds. Yager, R and Zadeh, L, Kluwer Academic Publishers.