Incorporating fuzzy operators in the decision network to improve classification reliability

Jung-Hsien Chiang *, Yan-Cheng Chen

Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan, ROC

Received 28 April 2000; accepted 19 December 2000

Abstract

One type of hierarchical fuzzy-operator-based network implementation is investigated. In this approach, we generalized the Dombi operator as an effective component for decision analysis and making. This methodology provides several advantages due to the fact that the input to each node is the evidence supplied by the degree of satisfaction of sub-criteria and the output is the aggregated evidence. Thus, the decision making process is to aggregate and propagate the evidence information through such a hierarchical network. This trainable network is able to perceive and interpret complex decisions by using those transparent fuzzy models. This study examines the behavior of the fuzzy additive operator in more detail and the results show that the proposed framework exhibits reliable decision in the pattern classification domain. © 2001 Published by Elsevier Science Ltd.

Keywords: Fuzzy connectives; Aggregation operators; Neural networks; Pattern classification; Decision making

1. Introduction

The primary interest in a decision making framework may be in the relationship between inputs, which can be manipulated, and outputs, which determine whether or not system goals are met. This relationship can be thought of as a function approximation between them. Neural network paradigms are gaining popularity throughout diverse fields ranging from cognitive science to pattern recognition [8,11,23,28]. A typical engineering application in decision making involves a trainable neural network that can aggregate measurement information obtained from input sources. In this scenario, the neural network framework realizes unknown functions based...
on training examples, and is able to generalize beyond the particular examples learned. These useful abilities have been encouraged neural network research activities in the past few years. New paradigms and learning algorithms have been developed and applied to solve difficult problems in a variety of domains. But one aspect of current neural network research remains a bottleneck that could lead it to something of a black art. This bottleneck is the problem of network transparency.

From the decision making perspective, the most important and informative aspects of any neural network paradigm are the hierarchical framework that performs specific computational operations as well as inference process. Most existing neurocomputation methodologies employ learning algorithms to construct desired input/output relationships from examples. However, those paradigms are black boxes. Both hierarchy information and inference characteristics for network interpretation are unavailable. In this paper, we explore the properties of the fuzzy operators and their utilization in fuzzy neurocomputation. This paradigm attempts to use learning algorithm to find appropriate parameters of functionally distinct nodes in hierarchy. Moreover, the overall decision boundaries of networks are transparent, i.e., after training, the fuzzy-operator-based neuron can be analyzed as a “sub-decision”, and the network itself can be interpreted as a collection of hierarchical sub-decisions. Clearly this will lead us to implement a reliable decision network on classification problems.

Fuzzy models have been applied to neural network implementation in a number of ways [1,13,17,26]. Dubois [6] and Klir [15] summarized the classes of fuzzy operators based on their aggregation behaviors. Other interesting fuzzy neurocomputation paradigms that have been suggested include the fuzzy-set-based neurons [25], which realize aggregation of the input signals and carry out some referential processing, the ordered weighted averaging (OWA) operator [7,32], which is capable of performing aggregations according to linguistic quantifiers. More recently, Keller et al. have suggested that additive $\gamma$-models with Yager’s operators can be trained to learn the tasks of function approximation between input and output spaces [14]. Other fuzzy neural networks [21,31] are able to perform soft computing in various applications. The paper is organized as follows. In Section 2, the basic properties of the fuzzy operators are considered. Section 3 presents network implementation and learning of fuzzy operators. The experimental results are presented in Section 4, and both synthetic data set and practical classification problem are considered. The summary is presented in Section 5.

2. The properties of fuzzy operators

In this section we briefly describe basic fuzzy set operators and their generalizations. Several properties of the fuzzy additive operators are then discussed.

The union (or intersection) of two fuzzy sets is in general a function

$f : [0, 1] \times [0, 1] \rightarrow [0, 1]$ such that it must satisfy the commutativity, monotonicity, associativity, and boundary conditions. Certainly, there are a number of interesting families of union and intersection operators in terms of the underlying fuzzy set theory [12,18,29]. Two particular interest examples are:
the maximum operator
\[ u(x_1, x_2, \ldots, x_n) = \max(x_1, x_2, \ldots, x_n) \]

the minimum operator
\[ i(x_1, x_2, \ldots, x_n) = \min(x_1, x_2, \ldots, x_n) \]

In many decision-making applications one is likely to take a position between the two extremes of min and max. In particular, a certain amount of compensation is desirable in real situations. Several compensative operators have been proposed in the literature \[9,33\]. In this work, a fuzzy additive operator (FAO) is a mapping
\[
f : \mathbb{R}^n \rightarrow \mathbb{R}
\]
such that
\[
y = f(x_1, x_2, \ldots, x_n; \gamma) = (1 - \gamma)y_i + \gamma y_u
\]
where \( y_i \) and \( y_u \) are intersection and union operators, respectively. As can be seen, the fuzzy operator can act as a pure intersection or union at the extremes: \( \gamma = 0 \) and \( 1 \), respectively. It allows the intersection and union to compensate for each other when \( 0 < \gamma < 1 \). Thus \( \gamma \) can be regarded as the parameter that controls the degree of compensation.

Following are two types of FAO models that have been used in this work.

- The Dombi operators (refer to D-Model)
  \[
y_i = \frac{1}{1 + \left[ \sum_j \left( \frac{1-x_j}{x_j} \right)^p \right]^{1/p}} \\
y_u = \frac{1}{1 + \left[ \sum_j \left( \frac{1-x_j}{x_j} \right)^{-p} \right]^{-1/p}} \quad p \in (0, \infty)
\]

- The weighted Dombi operators (refer to W-Model)
  \[
y_i = \frac{1}{1 + \left\{ \sum_j \left[ w_j \left( \frac{1-x_j}{x_j} \right)^p \right] \right\}^{1/p}} \\
y_u = \frac{1}{1 + \left\{ \sum_j \left[ w_j \left( \frac{1-x_j}{x_j} \right)^{-p} \right] \right\}^{-1/p}} \quad \text{where } w_j > 1
\]

The definitions given above for \( y_i \) and \( y_u \) satisfy De Morgan’s theorem, i.e.,
\[
y_i(a_1, a_2, \ldots, a_n) = \neg y_u(\neg a_1, \neg a_2, \ldots, \neg a_n) \quad \text{and} \quad y_u(a_1, a_2, \ldots, a_n) = \neg y_i(\neg a_1, \neg a_2, \ldots, \neg a_n)
\]
where \( \neg a \) denotes negation.
From the above we can see that each model has different parameters associated with the input sources, which drive the activation of the model up toward (or back down toward) the maximum (or the minimum). We begin our study with a few basic properties of the FAO models. A number of important properties of the FAO model can be associated with the basic fuzzy set theoretic connectives [5]. We only point out some of these.

**Property 1.** The ranges of the two models are as follows:

\[
\text{D-model} \quad i_{\min}(x_1, x_2, \ldots, x_n) \leq y \leq u_{\max}(x_1, x_2, \ldots, x_n)
\]

where

\[
i_{\min}(x_1, x_2, \ldots, x_n) = \begin{cases} x_k & \text{when } x_i = 1 \text{ for all } i \neq k \\ 0 & \text{otherwise} \end{cases}
\]

\[
u_{\max}(x_1, x_2, \ldots, x_n) = \begin{cases} x_k & \text{when } x_i = 0 \text{ for all } i \neq k \\ 1 & \text{otherwise} \end{cases}
\]

\[
\text{W-model} \quad \tilde{i}_{\min}(x_1, x_2, \ldots, x_n) \leq y \leq \tilde{u}_{\max}(x_1, x_2, \ldots, x_n)
\]

where

\[
\tilde{i}_{\min}(x_1, x_2, \ldots, x_n) = \begin{cases} 1 & \text{when all } x_i = 1 \text{ for } i = 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}
\]

\[
\tilde{u}_{\max}(x_1, x_2, \ldots, x_n) = \begin{cases} 0 & \text{when all } x_i = 0 \text{ for } i = 1, 2, \ldots, n \\ 1 & \text{otherwise} \end{cases}
\]

**Property 2.** Following table summarizes function behaviors with respect to various parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>D-model</th>
<th>W-model</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>( c )</td>
<td>▲</td>
<td>▼</td>
</tr>
<tr>
<td>( x_j )</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>( y_i )</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>( y_u )</td>
<td>▼</td>
<td>▼</td>
</tr>
<tr>
<td>( P )</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>( w_j )</td>
<td>▼</td>
<td>△</td>
</tr>
</tbody>
</table>

(▲) monotonically increasing, (▼) monotonically decreasing, (△) monotonically nondecreasing and (▼) monotonically nonincreasing.

Fuzzy operator model clearly represents a parameterized integration operation. The full scope of fuzzy operator is shown in Fig. 1. The ranges of the fuzzy operators can be parameterized to extend to one (or zero). It can be seen that the fuzzy operator gets optimistic (or pessimistic) with increasing (or decreasing) \( \gamma \). The \( \gamma \) values of the fuzzy operators indicate where the actual output is located between the min and max. Thus, by varying the value of \( \gamma \) we can achieve the required degree of compromise.
Obviously, this formulation has the advantage of allowing the parameters to govern the desired behavior of the model instead of requiring knowledge to describe approximation process. It also has the characteristic that the final model tends to represent unknown approximated function transparently. These desirable properties provide the flexibility in network implementation. In Section 3, we will describe a multilayer network based on the fuzzy operator models in more detail.

3. Fuzzy-operator-based decision framework

The basic objective of decision process in engineering applications is to obtain a set of optimal discriminant functions that will allow assignment of an object (or an unknown pattern) into one of $k$ categories with minimum error. In this scenario, an unknown object is passed through each of the $k$ discriminant functions. However, the predefined discriminant functions may not be optimal and the concept of specifications can be very vague. The Bayes’ decision strategy is to obtain the estimation in such a way that the “expected error” is minimized. Fig. 2 depicts the fuzzy-operator-based decision framework. We refer to the fuzzy operator portion as the confidence estimation mechanism since the outputs of operators are overall confidence values of an unknown pattern belonging to certain category. The final stage is the rejection mechanism that relies on output level of the confidence value. It is concerned with the issue of classification reliability. Final decision can be obtained from output of the rejection mechanism.

Let us represent the confidence estimation mechanism by the decision function $f(v, \theta)$ where $v$ represents the input vector and $\theta$ represents the parameters that define the FAO. In this decision framework, we will consider to reject the decision result if the confidence value is small. In this case, the decision rule will be

$$\begin{align*}
\text{decide category } \omega_i & \quad \text{if } (f(v, \theta_i) = \max_i f(v, \theta_j)) \wedge (f(v, \theta_j) > \lambda_j) \quad \text{for } i = 1, 2, \ldots, k \\
\text{reject} & \quad \text{else}
\end{align*}$$

(8)
where \( f(v, \theta_j) \) represents \( j \)th set of decision function, \( \lambda_j \) represents certain threshold value, and \( \wedge \) stands for a logical and. Moreover, the error criterion is minimized by the confidence estimation mechanism as following:

\[
E^2 = \int_x [f(v, \theta) - g(v)]^2 p(v) \, dv \tag{9}
\]

where \( p(v) \) is the probability density function governing \( v \) and \( g(v) \) is the function of the posterior probability. Therefore, confidence estimation mechanism finds a minimum mean square error (MSE) approximation function by adjusting the parameters of \( f(v, \theta) \). In other words, to minimize \( E \) can be achieved by minimizing the MSE between the output of the confidence mechanism and the posterior probability. How good an approximation can be achieved depends on the training data and also on the implementation of the confidence estimation mechanism. We will describe this in experiments.

Applying the techniques that have proved successful in the learning algorithms used in neural networks [34], we now present a derivation example for parameters learning. The architectural graph shown in Fig. 3 illustrates the layout of a multilayer FAO network for the case of a single

---

**Fig. 2.** The fuzzy-operator-based decision framework.

**Fig. 3.** Architectural graph of the FAO network.
hidden layer. The input vector consists of \( n \) information sources, and an integrating function called fuzzy operator that performs the “transfer” function in each node.

Let us assume that there are \( n \) inputs to the node, and the training data for this node consists of \( M \) sets of inputs \( x^k_1, \ldots, x^k_n \) with \( M \) corresponding desired outputs \( d^k \) (for \( k = 1, 2, \ldots, M \)). The learning process is to determine the best set of \( \{p, \gamma\} \) values for the fuzzy operator model in such a way that the discrepancy between the desired and actual output behavior is minimized. One measure that is commonly used as discrepancy is the sum of squared error

\[
E = \frac{1}{2} \sum_k E_k = \frac{1}{2} \sum_k (y^k - d^k)^2
\]

where \( y^k \) denotes the \( k \)th set of final output value. The D-model is then optimized by minimizing \( E \) with respect to the parameters. Thus, we update \( \gamma \) and \( p \) value using the following equations based on gradient descent

\[
\gamma^{\text{new}} = \gamma^{\text{old}} - \eta_1 \frac{\partial E}{\partial \gamma} = \gamma^{\text{old}} - \eta_1 \sum_k (y^k - d^k) \frac{\partial y}{\partial \gamma}
\]

\[
p^{\text{new}} = p^{\text{old}} - \eta_2 \frac{\partial E}{\partial p} = p^{\text{old}} - \eta_2 \sum_k (y^k - d^k) \frac{\partial y}{\partial p}
\]

where \( \eta_1 \) and \( \eta_2 \) are suitable positive constants and

\[
\frac{\partial y}{\partial \gamma} = y_u - y_i
\]

\[
\frac{\partial y}{\partial p} = (1 - \gamma) \frac{\partial y_i}{\partial p} + \gamma \frac{\partial y_u}{\partial p}
\]

and

\[
\frac{\partial y_i}{\partial p} = \frac{[\sum_j^u (\frac{1}{x_j} - 1)^p]^{1/p} \{\sum_j^u [(\frac{1}{x_j} - 1)^p \ln (\sum_j^u (\frac{1}{x_j} - 1)^p)/(\frac{1}{x_j} - 1)^p]\}}{\lambda^2 \sum_j^u (\frac{1}{x_j} - 1)^p \{1 + [\sum_j^u (\frac{1}{x_j} - 1)^p]^{1/p}\}^2}
\]

\[
\frac{\partial y_u}{\partial p} = \frac{[\sum_j^u (\frac{1}{x_j} - 1)^p]^{-1/p} \{\sum_j^u [(\frac{1}{x_j} - 1)^{-p} \ln (\sum_j^u (\frac{1}{x_j} - 1)^{-p})/(\frac{1}{x_j} - 1)^{-p}]\}}{\lambda^2 \sum_j^u (\frac{1}{x_j} - 1)^{-p} \{1 + [\sum_j^u (\frac{1}{x_j} - 1)^{-p}]^{-1/p}\}^2}
\]

This learning process is repeated until there is no change in \( \gamma \) and \( p \). The choice of \( \eta_1 \) and \( \eta_2 \) is important and it determines the speed and reliability of convergence [34]. Fig. 4 depicts decision surfaces of both D-model and W-model on union-like and intersection-like data sets that were used by Keller et al. [14]. The final parameters were obtained from gradient-descent-based learning algorithm described above.
4. Experimental results

In this section, we give two numerical examples to illustrate the characteristics of the FAO networks. First we deal with the non-linearly separable exclusive-OR (XOR) problem and its fuzzy version. The second example describes a reliable framework based on the FAO networks for handwritten digits classification.

Example 1. Non-linearly separable examples (The fuzzy XOR problems): The first example is based on the XOR problem in which there are two non-linearly separable classes. In particular the diagonal pairs (0,1) and (1,0) constitute one class and the other diagonal patterns (0,0) and (1,1) constitute the other. In this example, we investigate a fuzzy version of the XOR problem. We have two classes, each being a Gaussian distribution centered on the respective diagonal position with identical variances. In this case, we have three different sizes of data sets which consist of 4, 16,
and 60 patterns, respectively. We used an FAO network with two D-models as hidden nodes and one D-model as output unit to classify those two classes. In order to compare the fuzzy operator network with standard back-propagation (BP) network, we retrain same configuration of the BP networks to perform the same problem. Fig. 5 depicts the decision boundaries for different sizes of data sets when the number of training patterns is varied from 4 to 60. Decision contours were obtained from output nodes of both the FAO and BP networks. The results were encouraging.

In this case the decision boundaries of the BP networks are varied significantly when the number of training patterns were increased (or decreased). In contrast to this, the fuzzy operator networks produce almost identical decision boundaries for different data sets in spite of changes in the number of the training patterns. The inconsistent decision boundaries of the BP network can be explained because of arriving to an available decision arbitrarily at the end of training, which is the result of random initialization. From the decision point of view, this could draw attention to the proposed fuzzy operator networks that produce consistent final decision reliably for analogous data sets. In other words, contrary to normal neural networks, the FAO networks can generate an effective confidence judgement for rejecting ambiguous patterns. We will further describe this in next example.

Fig. 5. Decision boundaries and scatter diagrams of three data sets for (a) BP and (b) FAO networks.
Example 2. In this example, we investigate the classification reliability of handwritten digits using fuzzy operator network involving rejection mechanism. Clearly, handwritten digits classification is a difficult problem [19,20,24,30]. General classifiers can perform classification of handwritten digits very well, if a perfect writing digit is used as input. In real applications, the inherent difficulty of personal written styles and image fields segmentation imply that the input to a classifier may not be a perfect digit. General classifiers can sometimes make an incorrect decision on imperfect input. The worst thing is that after time-consuming preprocessing the original image is still recognized incorrectly. Thus, there is a need for a reliable classification framework that can lead to better rejection of imperfect samples than the general classification methodology [2]. This is consistent with Marr's Principle of Least Commitment: “Don't do something that may later have to be undone” [22]. In other words, to improve the whole classification system reliability, the framework should be able to decide to read the digits with extremely high recognition rate (e.g. at least 99% correct rate) or to process them manually after rejection. In this example, we used the fuzzy operator network to implement the classification framework.

The data sets used for the experiments consist of handwritten numeric digits extracted from the Computer and Communication Research Lab Database [3] and partially from the CEDAR CDROM-1 Database [10]. In this experiment 1000 samples per class of numeric digits were selected for the training set. For the test set, another 1000 samples per class were selected. Samples of digit from the training data set used in our experiments are shown in Fig. 6.

We now describe the classification framework which consists of the following three stages:

1. Self-organization stage
2. Confidence estimation stage
3. Rejection stage

Fig. 7 illustrates the three-stage framework for digit classification in our research. In the self-organization stage, we use the unsupervised self-organizing feature map (SOM) algorithm [16,27] to create prototypes. Each prototype represents an allograph [4]. An allograph is defined as a certain shape of pattern that can be used to represent a digit (e.g., cursive or printed styles of the same digit). Example of one of an allograph “3” was shown as gray level image in the figure. We used at most nine allographs for each class, and each allograph is thought of as a prototype function. Given an unknown input pattern, it is desired to evaluate with each prototype function and obtain a confidence value. The confidence value here can be interpreted as the degree to which the input pattern matches the prototype function.

We now describe the confidence estimation and learning in the second stage. This stage has incorporated ranking, normalization, and fuzzy operator model learning into a confidence esti-
The idea is to evaluate the ordering confidence values from each individual class for the purpose of choosing best fuzzy operator model. In other words, we seek for an FAO model that best fit the prototype function with high confidence from the first stage. The parameters of 10 FAO models, one model for each category, were determined from the training data set using off-line learning. The output value obtained from the confidence estimation mechanism with the highest confidence here can be considered to be the classification result without rejection.

The third stage is performed for rejection purpose. For each pattern, the corresponding highest confidence category is considered as "legal decision" if the final confidence value obtained from the second stage is higher than a threshold value. Specifically, for a certain threshold (or decision boundary) we reject an input pattern if the final confidence value of the top ranked category is below the threshold. We also measure rejection rate and reliability with respect to the threshold level. The reliability is defined as the number of correct responses dividing by the number of responses where number of responses is the number of input patterns that are not rejected. In order to compare reliability of the fuzzy operator network with standard BP network, we retrain same configuration of the BP networks to replace the confidence estimation mechanism in the second stage. For training data set, all of the three models achieve approximately 97% correct rate without rejection. Fig. 8 shows the reliability curves obtained from each individual model on test data set. It can be seen that in the W-model we have a reliability of 99.3% on test data for the rejection rate 5%. In the same rejection level, the reliability obtained from the standard BP model is 98.4%. This result suggests that the W-model provides a higher reliable classification performance involving rejection than standard BP network.
5. Summary

The fuzzy-operator-based decision framework described in this paper illustrates that the fuzzy models can be easily integrated into neural networks. We examined the use of the fuzzy additive operators and their generalizations as neural nodes in network implementation for decision making. The proposed method is capable of aggregating information to arrive at appropriate overall decision functions. Transparent network interpretation through this methodology is quite possible. In the empirical experiments, we demonstrated that the proposed network can be used to obtain reliable decision in the cases of synthetic and practical classification problems. Our results show that the proposed model can generate an effective confidence judgement for rejecting ambiguous patterns.

Acknowledgements

This research work was supported in part by the National Science Council under Grant NSC-892213-E-006-119, Taiwan.

References

Jung-Hsien Chiang received the B.Sc. degree in electrical engineering from National Taiwan Institute of Technology, Taiwan and the M.S. and Ph.D. degrees in computer engineering from the University of Missouri, Columbia, in 1991 and 1995, respectively. He is an Associate Professor in the Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan. Previously, he worked as a researcher at the Computer and Communication Research Laboratory, Industrial Technology Research Institute, Taiwan. His current research interests include fuzzy modeling, fuzzy systems, neural networks, pattern recognition, and the modeling of uncertainty decision-making problem.

Yan-Cheng Chen was born in Yun-Lin, Taiwan, ROC, in 1975. He received the B.S. degree in information and computer engineering from Chung Yuan Christian University, Chung-Li, Taiwan in 1999. He is currently working towards the M.S. degree at the Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan. His research interests include information retrieval, fuzzy logic, and neural networks.