USC-SIPI REPORT #369

The Choquet and Sugeno Fuzzy Integrals: A Tutorial

by

Arjun Bharadwaj and Jerry M. Mendel

December 2005

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Viterbi School of Engineering
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Suite 400
Los Angeles, CA 90089-2564 U.S.A.

Contents

List of Tables

List of Figures

Abstract

In this report we present a tutorial on fuzzy measures, the Choquet Fuzzy integral (CFI) and the Sugeno Fuzzy Integral(SFI). We also provide applications of these integrals using heuristic and numerical examples. Finally, we provide practical applications in pattern classification and reservoir management.

Chapter 1

Introduction

The fuzzy integral is a non-linear approach to combine multiple sources of uncertain information (e.g., in pattern recognition applications, where results from multiple classifiers will be combined). The function being integrated provides a confidence value for each information source for a particular hypothesis, and the integral is evaluated over the set of information sources. The Choquet (and Sugeno) fuzzy integral is a specific type of fuzzy integral which combines information from multiple sources by taking into account subjective evaluation of the worth of each of the sources.

Chapter 2

Fuzzy Measures

The fuzzy integral relies on the concept of a fuzzy measure which in turn is a generalization of the concept of a probability measure. Consider a finite universal set $X = \{x_1, \ldots, x_n\}$ which can be interpreted in a number of ways, e.g.

- X is a set of expert judgments concerned with decision making.
- X is a set of attributes or features. Each element of X is used to calculate a degree of membership for an object $u \in U$ with respect to a class $w \in \Omega$.
- X is a set of classifier outputs. This is different from the previous interpretation in that the classifier outputs can be represented as confidence levels for associating an object with a particular class.

Let $P(X)$ be the power set of X. A *fuzzy measure* over the set X is the set function

$$g : P(X) \rightarrow [0, 1] \tag{2.1}$$

such that

1. $g(\emptyset) = 0$, $g(X) = 1$
2. If $A, B \subset P(X)$ and $A \subset B$, then $g(A) \leq g(B)$.

Usually $g(A)$ is viewed as the importance or power of an individual source or subset of sources (A) within the set X.

Another way of interpreting fuzzy measures is by considering the effect of the contribution of an element to a union or subset of elements. The contribution or added value of element X_i in union A is defined by $g(A \cup X_i) - g(A)$.

According to Sugeno [7], a fuzzy measure $g\left(A\cup B\right)$, which specifies the importance of the union of disjoint subsets A and B, cannot be completely ascertained from the component measures $g(A)$ and $g(B)$. Consequently, he introduced λ - fuzzy measures (also called "Sugeno measures") which satisfy the additional property, that:

$$g_\lambda(A\cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B)\ ,\ \lambda > -1 \tag{2.2}$$

for all $A, B \subset X$ with $A \cap B = \emptyset$. Sugeno fuzzy measures are typically denoted as g_λ; however, since they have found widespread use in applications (especially those involving fuzzy integrals), it has become common practice to denote them just as g. In this report, g and g_λ are used interchangeably.

The following are important properties of fuzzy measures:

1. If $\lambda = 0$, then the fuzzy measure g_λ becomes a probability measure in that $g\left(A\cup B\right) = g(A) + g(B)$. If $\lambda < 0$, then g_λ shows sub-additivity in that $g\left(A\cup B\right) < g(A) + g(B)$ and if $\lambda > 0$, then g_λ shows super-additivity in that $g\left(A\cup B\right) > g(A) + g(B)$.

2. Let X be a finite set of information sources $X = \{x_1, \ldots, x_n\}$ and let $g_\lambda^i = g_\lambda(\{x_i\})$. The values $g_\lambda^1, g_\lambda^2, \ldots, g_\lambda^n$ are called *fuzzy densities*[1] and represent the importance of the individual information sources.

3. Let

$$A_i = \{x_1,\ \ldots,\ x_i\} \subseteq X \tag{2.3}$$

We can then form a sequence of nested sets $A_1, \ldots, A_n$, starting from $A_1 = \{x_1\}$, and subsequently adding in elements $x_2, \ldots, x_n$, one at a time (note that $A_n = X$ and $A_0 = \emptyset$). The measure $g(A_i)$ is calculated from the following recursive formula which can be derived from (2.2) (see Section A.1):

$$g\left(A_i\right) = g\left(A_{i-1}\cup\{x_i\}\right) = g^i + g\left(A_{i-1}\right) + \lambda g^i g\left(A_{i-1}\right) \quad \text{for } 1 \le i \le n \tag{2.4}$$

where $g(A_1) = g^1$ and $g(A_n) = g(X)$.

4. Given the fuzzy densities for a set of sources X, it is important to determine the measures of the elements of the power set $P(X)$. This is essential in many applications and, as we explain next, can be done by using the λ - fuzzy measure. Let $A_i =$

[1] They are called "fuzzy" because $g_\lambda^1, g_\lambda^2, \ldots, g_\lambda^n$ are the values of the membership function of the fuzzy set g defined on X.

$\{x_1, \ldots, x_i\} \subseteq X$. According to (2.4) we can write (see Section A.2)

$$g(A_n) = \sum_{j=1}^{n} g^j + \lambda \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} g^j g^k + \cdots + \lambda^{n-1} g^1 \cdots g^n \tag{2.5}$$

$$g(A_n) = \left[\prod_{x_i \in A} \left(1 + \lambda g^i\right) - 1 \right] \left(\frac{1}{\lambda}\right), \quad \lambda \neq 0 \tag{2.6}$$

The value of λ can then be found by solving the equation

$$g(A_n) = g(X) = 1 \tag{2.7}$$

From (2.6) and (2.7), this is equivalent to solving the following equation for λ:

$$\lambda + 1 = \prod_{i=1}^{n} \left(1 + \lambda g^i\right) \tag{2.8}$$

Hence, if we know the fuzzy densities g^i, $i = 1, \cdots, n$, we can construct the λ - fuzzy measure. We first solve (2.8) for λ, and then compute the $g(A_i)$'s using (2.4).

5. For a fixed set of densities $\{g^i\}$, $0 < g^i < 1$, there exists a unique $\lambda \in (-1, \infty)$ where $\lambda \neq 0$ which satisfies (2.8).

6. Let A_i be defined according to (2.3). For densities $\{g^i\}$, $0 < g^i < 1$, we have (see Section A.3)

$$0 \leq g(A_i) \leq 1 \; \forall i \tag{2.9}$$

with equality when $i = 0$ and $i = n$, i.e., $g(A_0) = 0$ and $g(A_n) = 1$.

Estimating the individual fuzzy densities, $\{g^1, g^2, \ldots, g^n\}$, is an important problem in all applications. The behavior of fuzzy integrals (both the Choquet and the Sugeno) is heavily dependent on the choice of these fuzzy densities. In some applications it is possible to estimate these densities from training data [10]. For example, in a pattern recognition application where the output of different classifiers are fused, the densities could be the performance of the individual classifiers. Liang et al. [6] used a genetic algorithm to determine the fuzzy densities from training data.

Chapter 3

The Choquet Fuzzy Integral

Let h be a measurable function

$$h : X \rightarrow [0, 1] \tag{3.1}$$

The Choquet fuzzy integral (CFI) defined below is the integral of h with respect to a fuzzy measure g_λ. Note that in (3.1), X could be a set of classifier outputs and $h(x)$ could be the soft output of the classifier (the confidence or evidence grade of the classifier) denoting that an input sample is from a particular class. In general, $X = \{x_1, \ldots, x_n\}$ is a set of information sources and $h(x_i)$ is the confidence grade of source i that a particular hypothesis is true. A λ-fuzzy measure provides the importance of each subset of sources X for this hypothesis evaluation.

To begin, we provide a definition of a fuzzy integral. Given a class of functions $F \subseteq \{h : X \rightarrow R\}$ and a class of fuzzy measures $m \subseteq M$, a functional

$$I : F \times m \rightarrow R \tag{3.2}$$

is a fuzzy integral [2]. Consider a specific function h associated with fuzzy density g_λ. Then, we can define a fuzzy integral as

$$h, g \rightarrow I(h, g) \tag{3.3}$$

There are a number of families of fuzzy integrals in terms of the underlying fuzzy measures that have been described in the literature. We are particularly interested in the *Choquet Fuzzy Integral (CFI)* which is a nonlinear functional defined over measurable sets that combines multiple sources of uncertain information. It provides a computational scheme for aggregating information.

The CFI of h over X with respect to a fuzzy measure g_λ is defined as

$$E_g(h) = \int_X h \circ g = \sum_{i=1}^{n} g(A_i) \left[h\left(x_i\right) - h\left(x_{i+1}\right)\right] \tag{3.4}$$

where $h(x_1) \geq h(x_2) \geq \cdots \geq h(x_n)$ and $h(x_{n+1}) = 0$. Set A is as defined in (2.3); i.e. $A_i = \{x_1, ..., x_i\} \subseteq X$, $g(A_n) = g(X)$, and $g(A_0) = 0$. The CFI can also be expressed as (see Section A.4)

$$E_g(h) = \int_X h \circ g = \sum_{i=1}^{n} h(x_i) \left[g\left(A_i\right) - g\left(A_{i-1}\right)\right] \tag{3.5}$$

If the function h is reordered such that $h(x_1) \leq h(x_2) \leq \cdots \leq h(x_n)$, then the CFI has the following form (see Section A.5) :

$$E_g(h) = \int_X h \circ g = \sum_{i=1}^{n} g\left(X - \{A_{i-1}\}\right) \left[h\left(x_i\right) - h\left(x_{i-1}\right)\right] \tag{3.6}$$

$$= \sum_{i=1}^{n} g\left(x_i, x_{i+1}, ..., x_n\right) \left[h\left(x_i\right) - h\left(x_{i-1}\right)\right] \tag{3.7}$$

All three forms of the CFI are identical. See Sections A.4 and A.5 to see how one form leads to another.

In comparison with probability theory, the CFI corresponds to the concept of expectation, and it has found extensive use in combining feature and algorithm confidence values [4].

Important properties of the CFI are (See Appendix B for their proofs):

1. The CFI is a monotonically increasing function with respect to $h(x)$ [2].

2. For all $h, g \in [0, 1]$, the range of the CFI is

$$h_{\text{min}} \leq E_g(h) \leq h_{\text{max}} \tag{3.8}$$

   where $h_{\text{min}} = \min\left(h(x_1), h(x_2), \ldots, h(x_n)\right)$ and $h_{\text{max}} = \max\left(h(x_1), h(x_2), \ldots, h(x_n)\right)$. The CFI attains its lower bound when $g_\lambda^i = 0$ for all i and it attains its upper bound when $g_\lambda^i = 1$ for all i.

3. If $h(x_i) = c$ for all i, where $0 \leq c \leq 1$, then

$$\int_X h \circ g = c \tag{3.9}$$

4. If $h_1(x_i) \leq h_2(x_i)$ for all i, then

$$E_g(h_1) = \int_X h_1 \circ g \leq \int_X h_2 \circ g = E_g(h_2) \tag{3.10}$$

5. If $B \in X$, $C \in X$ and $B \subset C$, then

$$\int_B h \circ g \leq \int_C h \circ g \tag{3.11}$$

6. If the λ-fuzzy measure g_λ is a probabily measure, i.e. $\sum_i g_\lambda^i = 1$ and $\lambda = 0$, the CFI becomes a weighted average. In the special case where all the fuzzy density values are equal, the CFI is equivalent to the arithmetic mean. This corresponds to the case where $g_\lambda^i = 1/n$.

7. If $g_\lambda^j = 0$ for some j, then

$$E_g(h) = \int_X h \circ g = \sum_{i=1, i \neq j}^{n} h(x_i) \left[g\left(A_i\right) - g\left(A_{i-1}\right) \right] \tag{3.12}$$

This property shows that the CFI values are determined only by the input sources that have non-zero densities.

Chapter 4

Generic Applications of the Choquet Fuzzy Integral

The CFI aggregates the elements of the source information set X according to a specified criterion, while incorporating the relative importance of each of the elements. In this section, we present the CFI for some simple "textbook" applications so as to foster a better understanding of the integral and gain insight into why it works in aggregation.

A Worker Productivity

Consider the example of productivity[1] in a workshop. Let $X = \{x_1, \ldots, x_n\}$ be a set of workers. Suppose that each worker x_i works $h(x_i)$ hours a day from the opening hour. Without loss of generality, the function that defines the number of work hours for each worker is ordered such that $h(x_1) \leq h(x_2) \leq \cdots \leq h(x_n)$, where worker x_1 works the least amount of time and worker x_n works the most; thus, for $i \geq 2$, $h(x_i) - h(x_{i-1}) \geq 0$.

The fuzzy measure is defined as the number of products made by the workers in one hour, with the implicit assumption that the productivity of the workers remains constant throughout the day; hence, $g(x_i)$ denotes the number of products made by worker x_i in one hour, and g is a measure of productivity. A group of workers $A \subset X$ produces the amount $g(A)$ in one hour. A product can be made either by one worker or by a group of workers. Hence, the number of products produced by 2 or more workers working together is larger than the sum of the products produced by each individual worker, if he were working alone.

Next we show that the CFI can be used to find the total number of products produced by all the workers in one work day. The working hours of all the workers are aggregated in

[1]This example has been paraphrased from [7].

the following way. First, the whole group X with n workers works $h(x_1)$ hours. Next, the group $X - \{x_1\} = \{x_2, x_3, \ldots, x_n\}$ works $h(x_2) - h(x_1)$ hours as the worker x_1 is no longer at work. Then, the group $X - \{x_1, x_2\} = \{x_3, x_4, \ldots, x_n\}$ works $h(x_3) - h(x_2)$ hours, and so on. Lastly, one worker x_n works for $h(x_n) - h(x_{n-1})$ hours. Since group A produces the amount $g(A)$ in one hour, the total number of products produced by all workers in one day can be expressed as:

$$\begin{aligned} h(x_1)g(X) &+ [h(x_2) - h(x_1)]\, g(X - \{x_1\}) \\ &+ [h(x_3) - h(x_2)]\, g(X - \{x_1, x_2\}) + \cdots \\ &+ [h(x_n) - h(x_{n-1})]\, g(\{x_n\}) \\ &= \sum_{i=1}^{n} [h(x_i) - h(x_{i-1})]\, g(\{x_i, x_{i+1}, \ldots, x_n\}) \quad \text{where } h(x_0) = 0 \\ &= \sum_{i=1}^{n} [h(x_i) - h(x_{i-1})]\, g(X - A_{i-1}) \quad \text{where } A_i = \{x_1, \ldots, x_i\} \text{ and } A_0 = \{\emptyset\} \\ &\triangleq E_g(h) \end{aligned} \tag{4.1}$$

This example shows that (4.1) fits the definition of the CFI in (3.6), and demonstrates the aggregation logic behind the CFI.

B A Collection of Rare Books

Consider a particularly rare book[2] that comes in two volumes. The first and second volumes are denoted by x_1 and x_2, respectively. The fuzzy measure is defined as the price of the two volumes. The price of the first volume is given by $g(\{x_1\})$, the price of the second volume by $g(\{x_2\})$ and the price of the complete set by $g(\{x_1, x_2\})$. The complete set is considered to be more valuable than the combination of the two volumes; hence, this fuzzy measure becomes a $\lambda-$fuzzy measure since

$$g(\{x_1, x_2\}) > g(\{x_1\}) + g(\{x_2\}) \tag{4.2}$$

A certain person sells $h(x_1)$ copies of the first volume and $h(x_2)$ copies of the second volume. Without loss of generality we can assume that $h(x_1) < h(x_2)$. The number of complete book sets (both volumes) sold is $h(x_1)$. The number of copies of the second volume sold separately

[2] This example has been paraphrased from [7].

is $h(x_2) - h(x_1)$. The total amount of money the seller gets is

$$h(x_1)\, g(\{x_1, x_2\}) + [h(x_2) - h(x_1)]\, h(x_2) \tag{4.3}$$

This expression is also similar to (3.6) and is another example of combining measurable functions with respect to densities using the CFI.

C Multiple Judges of a Sporting Event

A numerical example that demonstrates the calculations of the CFI, is presented next; it is adapted from [5].

Table 4.1: Scores for the participant u from the five judges

Unordered			Ordered		
Judge (x_i)	Score ($h(x_i)$)	Expertise ($g(x_i)$)	Judge (x_i')	Score ($h(x_i')$)	Expertise ($g(x_i')$)
1	0.5	0.8	2	0.7	0.5
2	0.7	0.5	4	0.6	0.7
3	0.2	0.4	5	0.6	0.7
4	0.6	0.7	1	0.5	0.8
5	0.6	0.7	3	0.2	0.4

Let the set $X = \{x_1, \ldots, x_5\}$ represent five judges at a sporting event. Assume that the participant u has obtained the scores shown in Table 4.1 from the $n = 5$ judges. Experts rate the judges' expertise, and their ratings are also shown in the table. A rating of 1 indicates that the judge is an expert while a rating of zero indicates a totally unknowledgeable judge (naturally a judge with a rating of zero would not be considered for aggregation). The expertise of the judges can be considered to be the densities g_λ of the $\lambda-$fuzzy measure. The judges' scores need to be aggregated so as to determine the final score for the participant; hence, the scores become the values of the function h aggregated in the CFI.

Given the fuzzy densities g^i of the set X, λ has to be computed using (2.8). Solving this equation with the density set $[0.8, 0.5, 0.4, 0.7, 0.7]$, we get a unique root > -1 which is $\lambda = -0.9943$. Next, the fuzzy measures of the power set can be determined by using (2.4).

The aggregate score is computed using the CFI by first ordering the scores and the corresponding densities such that the scores are in decreasing order. Prior to ordering, the scores are assigned as $h(x_1) = 0.5$, $h(x_2) = 0.7$, ..., $h(x_5) = 0.6$. Let X' be the ordered set so that $h(x'_1) \geq h(x'_2) \geq \cdots \geq h(x'_n)$. Then,

$$[0.5, 0.7, 0.2, 0.6, 0.6] \rightarrow [0.7, 0.6, 0.6, 0.5, 0.2] \quad \text{(scores)} \tag{4.4}$$

$$[0.8, 0.5, 0.4, 0.7, 0.7] \rightarrow [0.5, 0.7, 0.7, 0.8, 0.4] \quad \text{(densities)} \tag{4.5}$$

After ordering, the scores are assigned (see Table 4.1) as $h(x'_1) = 0.7$, $h(x'_2) = 0.6$, $\cdots$, $h(x'_5) = 0.2$ and the corresponding densities are $g(x'_1) = 0.5$, $g(x'_2) = 0.7, \ldots, g(x'_5) = 0.4$. The new arrangement of the judges (set X') according to the re-ordering, is $X' = \{x'_1, x'_2, x'_3, x'_4, x'_5\} = \{x_2, x_4.x_5, x_1, x_3\}$. Let $A_i = \{x'_1, \ldots, x'_i\}$. The fuzzy measures for all A_i, $i = 1, \ldots, n$ are computed recursively according to (2.4), as:

$$\begin{aligned} g(A_1) &= 0.5 \\ g(A_2) &= 0.7 + 0.5 - 0.9943(0.7)0.5 = 0.8520 \\ g(A_3) &= 0.7 + 0.8520 - 0.9943(0.7)0.8520 = 0.9590 \\ g(A_4) &= 0.8 + 0.9590 - 0.9943(0.8)0.9590 = 0.9962 \\ g(A_5) &= 0.4 + 0.9962 - 0.9943(0.4)0.9962 = 1.0 \end{aligned} \tag{4.6}$$

Since $n = 5$, $g(A_5) = g(X) = 1.0$. The CFI can now be computed using (3.4) as:

$$\begin{aligned} E_g(h) &= (0.7 - 0.6)0.5 \\ &+ (0.6 - 0.6)0.8520 \\ &+ (0.6 - 0.5)0.9590 \\ &+ (0.5 - 0.2)0.9962 \\ &+ (0.2 - 0)1.0 \\ &= 0.64 \end{aligned} \tag{4.7}$$

According to the CFI, the aggregated score of participant u is 0.64.

In this example we cannot use the arithmetic average as a tool for aggregation of the judges' scores because the sum of the densities $\sum_{i=1}^{5} g(x_i) = 3.10 > 1$. In order to compute the weighted average, we have to normalize the densities as

$$w_i = \frac{g(x_i)}{\sum_{i=1}^{5} g(x_i)} \tag{4.8}$$

The scores and the normalized weights are given in Table 4.2. The weighted average of the judges' scores can now be computed as

Table 4.2: Scores for participant u from the five judges and their corresponding weights

Judge (x_i)	Score ($h(x_i)$)	Normalized Expertise (w_i)
1	0.5	0.26
2	0.7	0.16
3	0.2	0.13
4	0.6	0.225
5	0.6	0.225

$$W_g(h) = \sum_{i=1}^{5} h(x_i)\, w_i = 0.5(0.26) + 0.7(0.16) + 0.2(0.13) + 0.6(0.225) + 0.6(0.225) = 0.538 \tag{4.9}$$

If the normalized weights were to be considered as a λ-fuzzy measure, then $\lambda = 0$ and hence the CFI reduces to the weighted average (see Property 6 in Chapter 3).

The aggregated result of the CFI is higher than that of the weighted average because it takes into account the increased *value* or in this case the increased expertise of two or more judges who agree on a particular score. This has occurred because of the computation of the λ-fuzzy densities for *subsets* of the set of judges X.

Chapter 5

The Sugeno Fuzzy Integral

Let h be a measurable function $h: X \rightarrow [0,1]$, one that is ordered such that $h(x_1) \geq h(x_2) \geq \cdots \geq h(x_n)$. Let $A_i = \{x_1, \ldots, x_i\} \subseteq X$ where $X = \{x_1, \ldots, x_n\}$, a finite set. The A_i's form a sequence of nested sets $A_1, \ldots, A_n$, starting from $A_1 = \{x_1\}$, and then subsequently adding elements x_2 to x_n, one at a time to get $A_n = X$, as mentioned earlier in property 3 in Chapter 2. Let $g^i = g_\lambda(\{x_i\})$ be the fuzzy densities of the set X. The Sugeno fuzzy integral (SFI) with respect to a fuzzy measure g is given by [9]:

$$F_g(h) \quad = \quad \int_X h \circ g \quad = \quad \sup_\alpha \; \{t\,(\alpha, g\,(H_\alpha))\} \tag{5.1}$$

where H_α is the α-cut of h and t is a t-norm.

The α-cut of a fuzzy set A on U is the set $A_\alpha = \{u|\, u \in U,\, \mu_A(u) \geq \alpha\}$ where $\mu_A(u)$ is the membership function of A. In our case the α-cut of h is the set $H_\alpha = \{u|\, u \in X,\, h(u) \geq \alpha\}$. In Sugeno's original formula, the t-norm used was the minimum. In order to use (5.1), the t-norm should follow the distributive law under the supremum operator [4]. Examples of such t-norms are minimum, product, bounded difference, and drastic product.

Since X is finite, h has at most n different α-cuts ranging from $H_0 = X$ to $H_{height(H)}$. The latter only contains the element(s) that reach the maximum level of the function h. Since $h(x_1) \geq h(x_2) \geq \cdots \geq h(x_n)$ and $A_j = \{x_1, \ldots, x_j\} \subseteq X$, each $A_j \subseteq X$ is the $h(x_j)$-cut of h; thus, (5.1) can be expressed as

$$F_g(h) = \int_X h \circ g = \max_{j=1,\ldots,n} \{t\,(h(x_j), g\,(A_j))\} \tag{5.2}$$

which is computationally simpler than (5.1). In Fig 5.1 the $h(x_i)^{th}$-cut of h is shown. This is the set $\{x_1, \ldots, x_i\}$ which we have defined to be A_i. In applications of the SFI to pattern recognition the t-norm most commonly used is the minimum; hence, the most common form

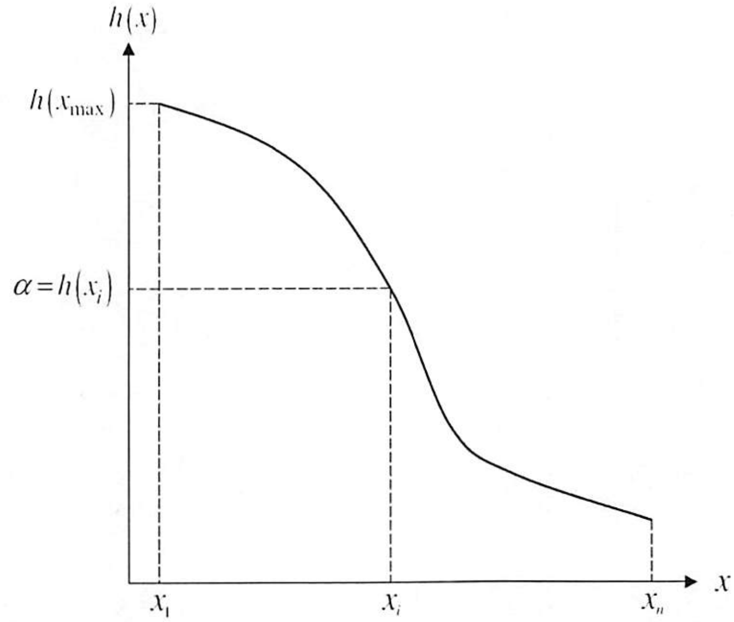

Figure 5.1: Plot of $h(x)$ versus x. Note that the α-cut shown in this figure is the set A_i. This is also the $h(x_i)^{th}$-cut. The continuous curve is for purposes of illustration only, i.e., it actually consists of lines connecting discrete values of $h(x_i)$ for $x = \{x_1, ..., x_n\}$.

of the SFI is given by

$$F_g(h) = \int_X h \circ g = \bigvee_j [h(x_j) \wedge g(A_j)] \tag{5.3}$$

where the values of $g(A_i)$ are determined recursively according to (2.4). Figure 5.2 provides a graphical interpretation of the SFI, which attempts to find the best *consensus* between the function value and the importance attribute. This is where the SFI fundamentally differs from the CFI. While the CFI quantitatively weights the function by the *jump* in the fuzzy measure, the SFI associates each function value with the corresponding importance measure.

Consider the worker-productivity problem in Section 4.A. The SFI finds the best consensus between the number of work hours and productivity whereas the CFI computes the improvement in productivity weighted by the number of hours. Clearly, the CFI makes more heuristic sense in this quantitative framework.

Next, consider the numerical example in Section 4.C where judges scored partipants at a sporting event. Recall that participant u has obtained the scores shown in Table 4.1 from the 5 judges; densities g_λ of the $\lambda-$fuzzy measure are given in (4.6), and the fuzzy measures denote the expertise of the judges. The scores obtained for participant u can now be aggregated as follows using the SFI so as to produce a final score. The scores are again reordered so that $h(x_1') \geq h(x_2') \geq \cdots \geq h(x_n')$ where $X' = \{x_1', x_2'.x_3', x_4', x_5'\} =$

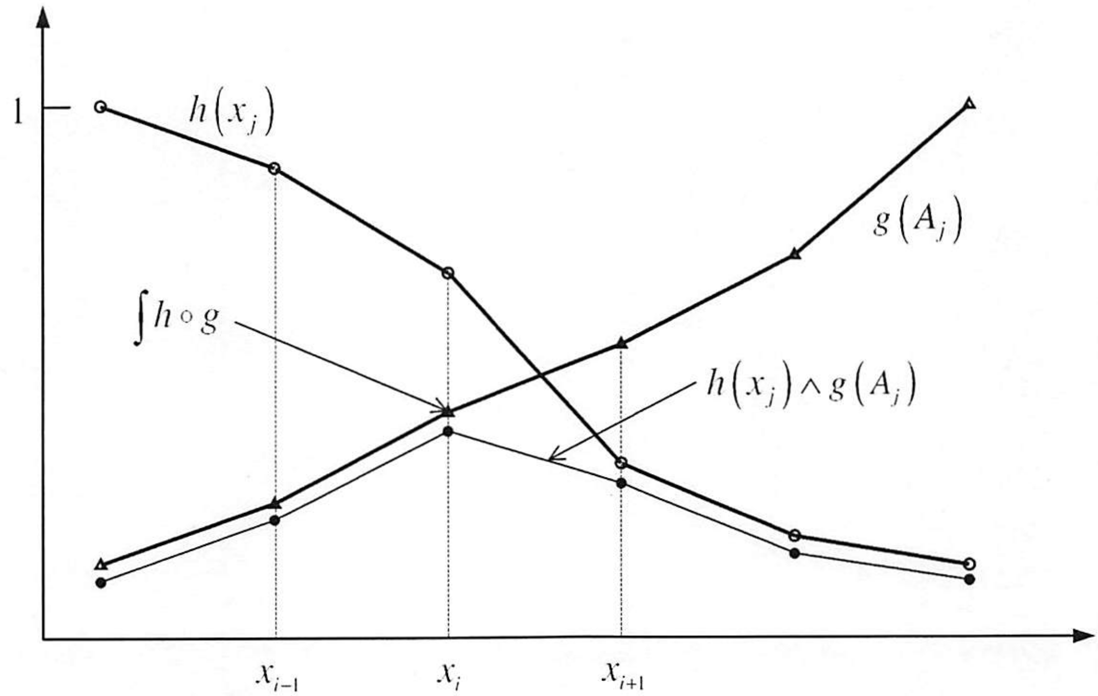

Figure 5.2: Graphical illustration of the calculation of the SFI

$\{x_2, x_4.x_5, x_1, x_3\}$ is the reordered set of judges as in (4.4), and the corresponding densities are also reordered as in (4.5). Using the defnision of the SFI in (5.3) and the computed values of the $g(A_i)$ in (4.6) we find:

$$\begin{aligned} F_g(h) &= \max\left[\min(0.7, 0.5), \min(0.6, 0.8520), \cdots \right. \\ &\qquad \left. \min(0.6, 0.9590), \min(0.5, 0.9962), \min(0.2, 1)\right] \\ &= \max\left[0.5, 0.6, 0.6, 0.5, 0.2\right] \\ &= 0.6 \end{aligned}$$

We see that the aggregated value of the scores $F_g(h)$, according to the stated fuzzy measure g using the SFI is 0.6. The SFI computes the highest score that most judges agreed upon corresponding to their expertise.

In general, the CFI is used when the problem framework involves quantitative measures and the SFI is used when qualitative measures (e.g., expertise, accuracy, talent etc) define the importance attribute. An application where quantitative measures are the densities, and hence the CFI is used, is described in Section 6.C.

The Generalized Sugeno Fuzzy Integral (GSFI) is an extended version of the SFI that is formed when each $h(x_i)$ is not a real value in $[0, 1]$ but is a fuzzy number within the universal set $[0, 1]$ and was presented by Auephanwiriyakul, et al. [1]. In the case of the numerical example just presented, if the judges had scored the participants qualitatively (e.g., good, excellent), then the scores would themselves be fuzzy. In this case the GSFI would be used to aggregate the scores. The output of the GSFI is a type-1 fuzzy set.

Chapter 6

Pattern Recognition Applications

In this section, we apply the CFI to pattern classification. Because of the versatility of the fuzzy integral, it can potentially be applied in multiple stages in the pattern classification problem. The CFI can first be used to combine classifier outputs from multiple feature sets, thereby improving the decision making ability of the classifier. Then, the CFI can be used to combine classifier outputs from different classifiers, (each having different performance levels), so as to improve overall probability of classification.

A Feature Aggregation by Decision Fusion

Let x_i denote a set of features, and let $X = \{x_1, x_2, \ldots, x_n\}$. Each x_i is characteristic of an object M that has to be classified into a set of predefined classes $C = \{c_1, \ldots, c_m\}$. Each x_i is a different set of features which could potentially be generated by different feature extraction algorithms, e.g. in face recognition, x_1 could be the quadrant energies of the Fourier transform of the image M and x_2 could be the set of spatial or temporal frequencies in the image M. Since there is inherent uncertainty in the pattern classification problem, there is normally no set of features that can precisely distinguish one object from another; thus, these feature sets are considered as evidence of characteristics which aid in identifying and classifying the object. Each of these feature sets would have a degree of importance in identification of object M, since usually all the feature sets do not carry equal importance in classifying the object. This notion of importance translates directly into a fuzzy measure.

There are many algorithms that have been described in the classification literature [3] (e.g. perceptrons, K-means, Bayesian classifiers, etc.) which can be used to classify the object M using one or more of the feature sets in X. Consider only one of these algorithms and let $h_k : X \to [0, 1]$ be the output of that classifier, i.e. $h_k(x_i)$ is the evaluation of the object M for class $c_k\,(k = 1, \ldots, m)$ using the feature set $x_i\,(i = 1, \ldots, n)$. A single

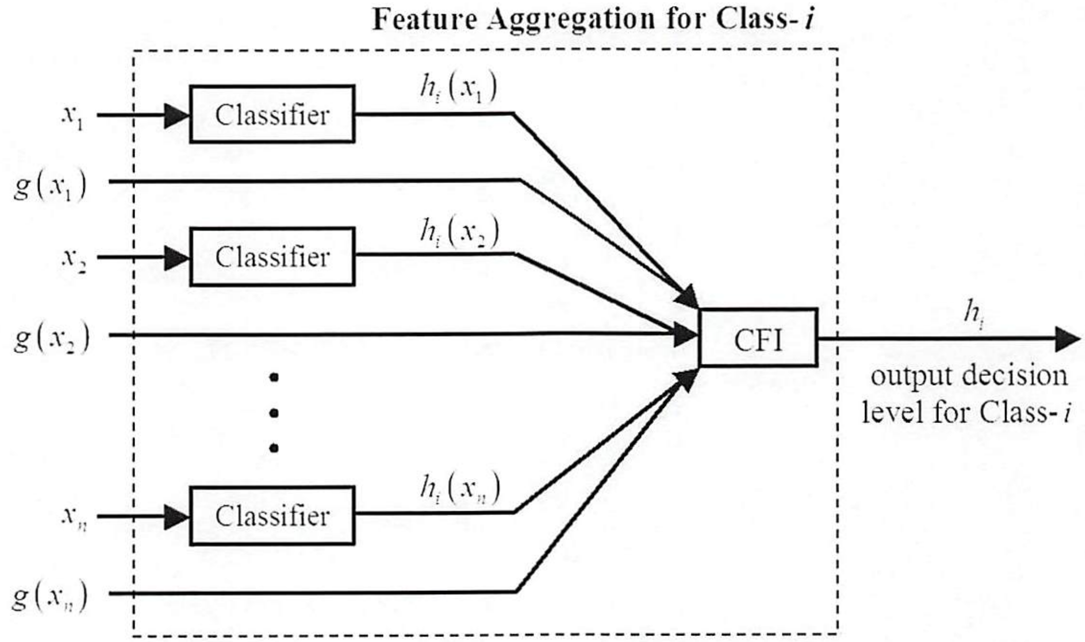

Figure 6.1: Block diagram representation of feature aggregation using decision fusion for class-i

classification algorithm can be trained using the different feature sets x_i $(i = 1, \ldots, n)$, so that $h_k(x_i)$ is the output of the classifier trained on the i^{th} feature set. $h_k(x_i)$ ranges from zero to one. An output of $h_k(x_i) = 1$ indicates absolute certainty that the object M is in class c_k, whereas an output of $h_k(x_i) = 0$ indicates absolute certainty that the object M is not in class c_k. The objective is to combine the classifier outputs $h_k(x_i)$ for all the feature sets x_i $(i = 1, \ldots, n)$ to obtain a decision level for the class c_k. Fig 6.1 shows a block diagram illustrating feature aggregation for a particular class. Note that the classifier algorithms for the n classifier are the same, but the effects of different input feature vectors are integrated using the CFI.

The degree of importance g^i associated with each feature set x_i denotes the importance of that feature set in the classification problem and must be ascertained prior to fusion. The g^i's are the fuzzy densities for the feature sets and can be specified by an expert or generated from training data. For a given classification algorithm the g^i's can be chosen to be the performance of the classifier for each of the x_i's for a particular class c_k. If the probability of classification error using training data for a feature set x_i is $P(x_i)$, then the fuzzy density $g^i = P(x_i)$. Since we are considering only λ-fuzzy densities in this report, λ can be calculated using (2.8) and then the complete fuzzy measure g_λ can be constructed. Using (3.4) (3.5) or (3.6), the CFI can then be computed.

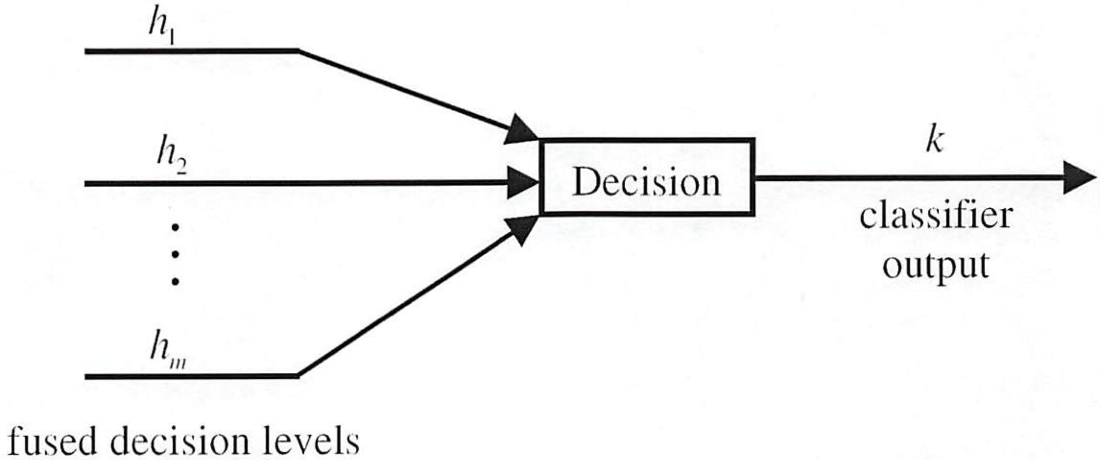

Figure 6.2: Block diagram of the decision strategy of the fused classifier

The CFI can be computed for each class c_k, $k = 1, ..., m$. If a decision needs to be made at this point, the class c_k with the largest integral value is chosen as the one most similar to the object M. See Fig 6.2 for an illustration on how a decision could be made using the classifier outputs. In the figure, h_k represents the outputs of the classifiers for class k. The decision box essentially computes the maximum of all the h_k's.

The set of CFI outputs can also be considered as partial evaluations and submitted as input to a higher level of recognition [10, 4]. This is explained next.

B Multisource Integration (or Multiclassifier Combining)

The CFI can be used to combine decisions from different information sources which could be decisions from different classifiers, and each of these classifiers could have a different degree of importance. Application of the CFI provides decisions supporting or rejecting the existence of the object under consideration in the class set C.

Under certain conditions, some classifiers may produce more reliable decisions than others, in which case we can integrate the decisions of the classifiers with respect to their relative importance to produce a more reliable decision, e.g. In [10], Tahani and Keller used the CFI to aggregate decisions from different image classifiers while maintaining a degree of flexibility so as to incorporate the specific attributes of the decision-maker.

Suppose there are p classifiers $Y = \{y_1, y_2, \ldots, y_p\}$ which make decisions about an object M which has to be classified into a set of predefined classes $C = \{c_1, \ldots, c_m\}$. Using the output of the decision box h_k in Section A, or a good feature extraction algorithm, these

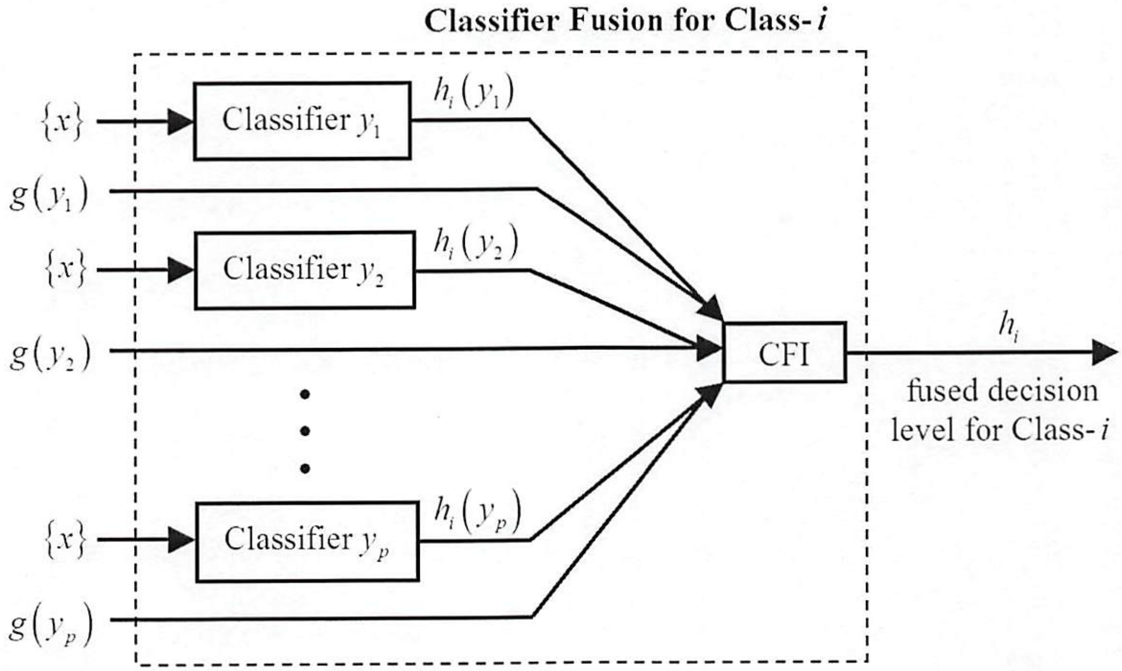

Figure 6.3: Block diagram representaion of multiclassifier fusion for Class-1

classifiers generate evaluations of the object M for each class c_k based on the degree of similarity between the object M and the class c_k. The outputs of these classifiers are then combined at a higher level to produce a final evaluation of the object M for each class. See Fig 6.3 for an illustration of Multiclassifier fusion. Note that the input feature set $\{X\}$ is the same for all the different classifier algorithms and the CFI is used to fuse classifier outputs based on classifier performance. Each of these classifiers has varying levels of reliability and performance; hence, each classifier has a different degree of importance in the recognition of the classes. The concept of "importance or reliability" associated with each classifier translates into the fuzzy density of that classifier and these densities are generated from training data. If the probability of classification error using training data for classifier y_i is $Q(y_i)$, then the fuzzy density for classifier y_i is $g^i = Q(y_i)$. For a given class c_k there are p fuzzy densities $g_k^i, i = 1, \dots, p$, one for each classifier, and p outputs $h_k(y_i)$. The decisions $h_k(y_i)$ are then integrated with respect to the fuzzy densities g_k^i over the classifier set Y. This is done for each class, resulting in a final confidence level for each class c_k. To make a final decision the class c_k with the highest confidence level is chosen (see Fig 6.2).

It has been shown [4] that the classification rates obtained by using the CFI are significantly higher than those obtained by using just the individual classifiers.

As explained above and in the previous section, the CFI can potentially be used at different stages in the pattern classification problem, e.g. combination of feature information, combination of sensor-based information, and fusion of classifier decisions. The core problem in using the CFI is in effectively estimating the densities. This is application specific and some methods have been proposed in the literature for applications like handwritten character classification, landmine detection, image segmentation, etc. A specific example of classifier fusion where the densities are generated from the training samples is described in the next section.

C CFI Applied to a Binary Classification Problem

Wu and Mendel [11] studied the binary classification problem of identifying tracked versus wheeled vehicles from acoustic data that was available for different runs which were segmented into data blocks from which features were extracted. The magnitudes of the second through 12th harmonics of each prototype (data block) were used as features. Because of the inherent uncertainties in the data, type-1 and type-2 fuzzy sets were used to model the uncertainties in the feature set.

Type-1 Fuzzy Logic - Rule Based Classifier (FL-RBC) and interval type-2 FL-RBC were designed and used for classification. In the type-1 FL-RBC, nine rules were used for the tracked versus wheeled classification problem, one for each of the nine kinds of vehicles for which measured data was available.

Let $X = \{x_1, x_2, \ldots, x_{11}\}$ be the feature set extracted from the data. The antecedents in each of the $l\,(l = 1, ..., 9)$ rules, $F_k^l\,(k = 1, \ldots, 11)$ (one for each feature) were modeled as type-1 fuzzy sets with membership functions (MF's) $\mu_{F_k^l}(x_k)$, and the consequents q^l were modeled as crisp numbers. Given an extracted feature vector $X' = [x'_1, ..., x'_{11}]$, the type-1 FL-RBC modeled each x'_k as a type-1 fuzzy set X_k and computed the firing degree $f^l(x')$ for each rule. The rules were then combined through defuzzification to obtain a crisp output $y(x')$. The decision about the feature vector X' depended on the sign of $y(x')$. If $y(x')$ was positive then x' was classified as a tracked vehicle, and if $y(x')$ was negative then x' was classified as a wheeled vehicle.

In the interval type-2 FL-RBC, the rule structure was the same as that of the type-1 FL-RBC, but the antecedents $\widetilde{F}_k^l\,(k = 1, \ldots, 11)$ and the feature sets were modeled as interval type-2 fuzzy sets. The consequent q^l was still modeled as a crisp number. Given an extracted feature vector $X' = [x'_1, ..., x'_{11}]$, the FLS modeled each x'_k as an interval type-2 fuzzy set $\widetilde{X}_k$ and computed the upper and lower firing degrees, $\overline{f}^l(x')$and $\underline{f}^l(x')$, for each rule. The rules were combined through type reduction to obtain a type-reduced set $[y_l(x'), y_r(x')]$. Finally, the type-2 FL-RBC defuzzified the type-reduced set to get a crisp output $y(x') =$

$[y_l(x') + y_r(x')]/2$. As in the case of the type-1 FL-RBC, the decision on the extracted feature vector X' depended on the sign of $y(x')$, where if $y(x')$ was positive then x' was classified as a tracked vehicle, and if $y(x')$ was negative then x' was classified as a wheeled vehicle.

A leave-one-out design was performed. It consisted of 88 designs (i.e., classifiers) for the type-1 and the type-2 classifiers. The parameters of all classifiers were optimized using the training data and a steepest descent algorithm. After training, the performance of the k^{th} classifier was characterized by its false alarm rate (FAR), p_k. Given a new input feature vector $X' = [x'_1, ..., x'_{11}]$, both the type-1 and type-2 FL-RBCs generated outputs $y(X')$ for each of their 88 classifier designs.

The CFI could be used to combine all 88 outputs of the type-1 and type-2 classifiers with respect to their corresponding FARs. The 88 classifier designs become the finite set of classifiers Y that are described in Sections A and B. The numerical outputs of the classifier designs are the functions $h_k(X)\,(k = 1, \ldots, 88)$ for a given set of input feature vectors X, and the fuzzy densities g^k correspond to the performance of each classifier design. If the probability of classification error using training data for for design y_k is $Q(y_k)$, then the fuzzy density for classifier y_k is $g^k = Q(y_k)$ The outputs of the individual classifier designs could be combined using the CFI as in (3.4), and a final output $y(X')$ could be obtained which could then be thresholded to make a final decision as to whether the input feature set X' corresponds to a wheeled or a tracked vehicle.

Chapter 7

Petroleum Reservoir Management

A petroleum reservoir is a dynamical system which produces observable output signals that are affected by external stimuli. The reservoir is a dynamical structure of many variables which can be observed (e.g., fluid rates) and that respond to the action of variable injection rates (and other inputs), and measurable and unmeasurable disturbances. Well fluid-rate-control has been related to optimization of fluid displacements in porous media [8]. There is a need to develop effective control schemes that will control the hardware especially for smart wells.

Here we examine a simple problem and demonstrate the application of the CFI to aid in decision fusion so as to provide better control inputs to the hardware. Consider the injector-producer model in a reservoir. The measurable outputs in a reservoir are the oil rate q_o, the water rate q_w and the gas rate q_g. It has been shown [8] that the reservoir pressure $\overline{p}$ and the well flowing pressure are linearly related to the oil, water and gas flow rates. The objective is to determine an optimal water injection-rate w_{inj} so as to meet target output fluid rates. This can be done in the following ways:

1. Since the reservoir pressure also depends on the water injection-rate, an adaptive mathematical reservoir model can be built to control the water injection-rate w_{inj} so as to achieve set-point target flow-rates $[q_{o,sp}, q_{w,sp}, q_{g,sp}]$; hence, one control output is obtained by mathematical *reservoir modeling.*

2. An empirical model can be built to control the water injection-rate w_{inj}. This method depends only on the sensor outputs from the reservoir and does not take into account the mathematical modeling of the reservoir; however, the model takes advantage of measured data to estimate w_{inj}; hence, another control output can be obtained by *empirical modeling.*

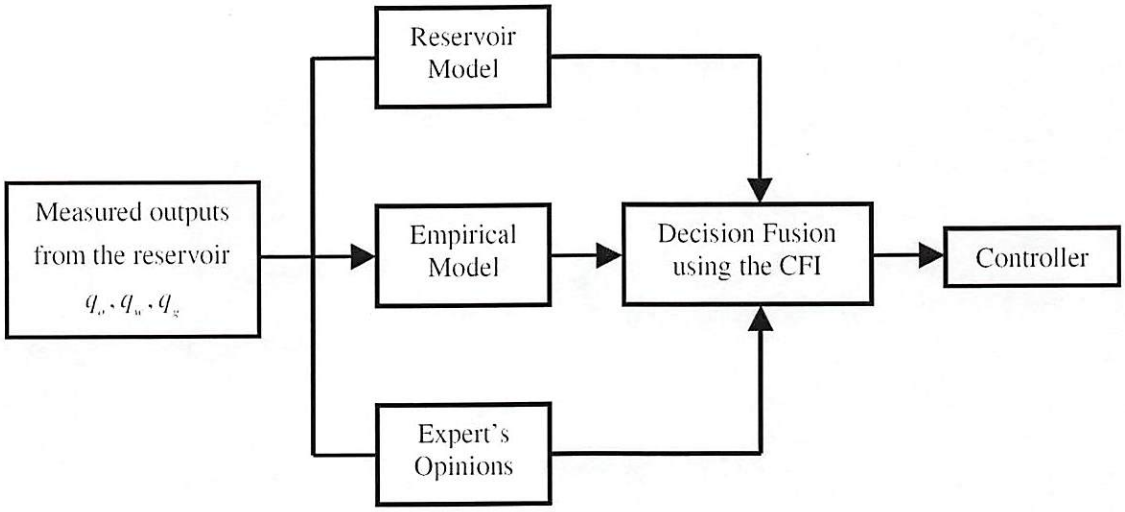

Figure 7.1: Representation of the injection-production problem

3. Experts in oilfield management can analyze the data and determine water injection-rates so as to attain set point targets for reservoir pressure and fluid rates; hence, a third control output comes from *expert opinion.*

The CFI could be used to combine these outputs so as to determine an aggregated optimal water injection rate w_{inj} (Fig. 7.1). The assignment of fuzzy measures (densities) or importance of each method, g^1, g^2, g^3, remains an open research problem. These densities could be obtained from field data or measured outputs of fluid flow fluctuations in response to changes in input for each of the methods. They need to be estimated once for a particular method, however, if the algorithms for computing control outputs change, then the densities need to be re-evaluated. Having determined the fuzzy densities g^i and the outputs $h^i = w^i_{inj}$, the λ - fuzzy measures could be determined by using (2.8), and the CFI could then be used according to (3.4) to obtain the final optimal water flow injection rate w_{inj}.

Chapter 8

Conclusions

In this report, we presented a tutorial on fuzzy measures and the CFI and the SFI. Generic applications and numerical examples were given to illustrate the use of the CFI and the SFI. Finally, applications of the CFI to pattern classification and reservoir management problems were proposed.

Appendix A

Derivation of Some of the Properties of Fuzzy Measures

A.1 Derivation of (2.4)

We know from (2.2) and (2.3) that for a λ-fuzzy measure

$$g\left(A_1\right) = g\left(\{x_1\}\right) = g^1 \tag{A.1}$$

$$g\left(A_2\right) = g\left(\{x_1, x_2\}\right) = g\left(\{x_1\} \cup \{x_2\}\right) = g^1 + g^2 + \lambda g^1 g^2 \tag{A.2}$$

$$g\left(A_3\right) = g\left(\{x_1, x_2, x_3\}\right) = g\left(\{x_1, x_2\} \cup \{x_3\}\right) = g\left(A_2\right) + g^3 + \lambda g^3 g\left(A_2\right) \tag{A.3}$$

From (A.1), (A.2), and (A.3) and extrapolating to n, we find

$$g\left(A_n\right) = g\left(A_{n-1} \cup \{x_n\}\right) = g^n + g\left(A_{n-1}\right) + \lambda g^n g\left(A_{n-1}\right) \tag{A.4}$$

Note that $g\left(A_n\right) = g\left(X\right) = 1$, and $g\left(A_0\right) = \emptyset = 0$ (by definition); therefore, we can state in general that

$$g\left(A_i\right) = g\left(A_{i-1} \cup \{x_i\}\right) = g^i + g\left(A_{i-1}\right) + \lambda g^i g\left(A_{i-1}\right) \quad \text{for } 1 \leq i \leq n \tag{A.5}$$

A.2 Derivation of (2.6)

Eqn (A.2) can be written as

$$g(A_2) = \sum_{j=1}^{2} g^j + \lambda g^1 g^2 \tag{A.6}$$

Multiplying and dividing by λ and then adding and subtracting 1 we get

$$g(A_2) = \left(\frac{1}{\lambda}\right)\left[1 + \lambda g^1 + \lambda g^2 + \lambda^2 g^1 g^2 - 1\right] \tag{A.7}$$

$$= \left(\frac{1}{\lambda}\right)\left[\left(1 + \lambda g^1\right)\left(1 + \lambda g^2\right) - 1\right] \tag{A.8}$$

$$= \left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{2}\left(1 + \lambda g^i\right) - 1\right] \tag{A.9}$$

Similarly, for $g(A_3)$, substituting (A.9) in (A.3), we see that

$$g(A_3) = \left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{2}\left(1 + \lambda g^i\right) - 1\right] + g^3 + \lambda g^3\left(\left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{2}\left(1 + \lambda g^i\right) - 1\right]\right) \tag{A.10}$$

$$= \left(\left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{2}\left(1 + \lambda g^i\right) - 1\right]\right)\left(1 + \lambda g^3\right) + g^3 \tag{A.11}$$

$$= \left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{3}\left(1 + \lambda g^i\right)\right] - \left(\frac{1}{\lambda}\right)\left(1 + \lambda g^3\right) + g^3 \tag{A.12}$$

$$= \left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{3}\left(1 + \lambda g^i\right) - 1\right] \tag{A.13}$$

For $n = k$, we assume that

$$g(A_k) = \left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{k}\left(1 + \lambda g^i\right) - 1\right] \tag{A.14}$$

Now, for $n = k + 1$, we have (according to (A.5))

$$g(A_{k+1}) = g^{k+1} + g(A_k) + \lambda g^{k+1} g(A_k) \tag{A.15}$$

$$= \left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{k}(1+\lambda g^i)-1\right] + g^{k+1} + \lambda g^{k+1}\left(\left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{k}(1+\lambda g^i)-1\right]\right) \tag{A.16}$$

$$= \left(\left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{k}(1+\lambda g^i)-1\right]\right)(1+\lambda g^{k+1}) + g^{k+1} \tag{A.17}$$

$$= \left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{k+1}(1+\lambda g^i)-1\right] - \left(\frac{1}{\lambda}\right)(1+\lambda g^{k+1}) + g^{k+1} \tag{A.18}$$

$$= \left(\frac{1}{\lambda}\right)\left[\prod_{i=1}^{k+1}(1+\lambda g^i)-1\right] \tag{A.19}$$

From (A.9) , (A.14), (A.19) and the induction hypothesis, we conclude that

$$g(A_n) = \left[\prod_{i=1}^{n}(1+\lambda g^i)-1\right]\left(\frac{1}{\lambda}\right), \qquad \lambda \neq 0 \tag{A.20}$$

A.3 Derivation of (2.9)

From the defnition of the fuzzy measure (see (2.1)) and (2.7), we have $0 < g^i < 1$ and $g(X) = g(A_n) = 1$. From (2.4), we get

$$g(A_{i-1}) = \frac{g(A_i) - g^i}{(1+\lambda g^i)} \tag{A.21}$$

When $i = n$, (A.21) becomes

$$g(A_{n-1}) = \frac{1 - g^n}{(1+\lambda g^n)} \tag{A.22}$$

Since $\lambda > -1$, $0 < g(A_{n-1}) < 1$ and consequently $0 < g(A_{n-2}) < 1$ [by substituting $i = n-1$ into (A.21)]. Similarly, $0 < g(A_i) < 1$ $\forall i$. Equality is obtained when $i = 1$, i.e., $g(A_n) = 1$ [by (2.7)].

A.4 Derivation of (3.5)

Expand (3.4) for $k-1$ and k as follows:

$$E_g(h) = \cdots + g\left(A_{k-1}\right)\left[h\left(x_{k-1}\right) - h\left(x_k\right)\right] + g\left(A_k\right)\left[h\left(x_k\right) - h\left(x_{k+1}\right)\right] + \cdots \tag{A.23}$$

Collect terms with $h(x_k)$ as the common factor, to obtain

$$E_g(h) = \cdots + h\left(x_{k-1}\right) g\left(A_{k-1}\right) + h\left(x_k\right)\left[g\left(A_k\right) - g\left(A_{k-1}\right)\right] + h\left(x_{k+1}\right)\left[-g\left(A_k\right)\right] \cdots \tag{A.24}$$

which when summed leads to (3.5).

A.5 Derivations of (3.6) and (3.7)

Let $Z = \{z_1, \ldots, z_n\}$ such that $z_1 = x_n, z_2 = x_{n-1}, \ldots, z_n = x_1$. Because $h(x_1) \leq h(x_2) \leq \cdots \leq h(x_n)$, it follows that $h(z_1) \geq h(z_2) \geq \cdots \geq h(z_n)$. This one-to-one correspondence between X and Z can be specified by the relation $x_i = z_{n-i+1}$. Replacing x by z in (3.7), we find

$$E_g(h) = \sum_{i=1}^{n} g\left(\{z_{n-i+1}, z_{n-i}, \ldots, z_1\}\right)\left[h\left(z_{n-i+1}\right) - h\left(z_{n-i+2}\right)\right] \tag{A.25}$$

Letting $j \equiv n - i + 1$, and $B_j \equiv \{z_1, \ldots, z_j\}$, (A.25) can be expressed as:

$$E_g(h) = \sum_{j=1}^{n} g\left(\{z_j, z_{j-1}, \ldots, z_1\}\right)\left[h\left(z_j\right) - h\left(z_{j+1}\right)\right] \tag{A.26}$$

$$= \sum_{j=1}^{n} g\left(B_j\right)\left[h\left(z_j\right) - h\left(z_{j+1}\right)\right] \tag{A.27}$$

which is in exactly the form of (3.4). This proves that (3.7) is a form that is equivalent to the CFI in (3.4). Finally (3.7) is equivalent to (3.6) because $X = \{x_1, x_2, \ldots, x_n\}$ and $A_i = \{x_1, x_2, \ldots, x_i\}$ and so, $g\left(X - \{A_{i-1}\}\right) = g\left(x_i, x_{i+1}, \ldots, x_n\right)$.

Appendix B

Proofs of the Properties of the CFI

B.1 Property 1

Let us take the partial derivative of CFI in (3.5) with respect to $h(x_i)$ for any i, i.e.

$$\frac{\partial E_g(h)}{\partial h(x_i)} = g(A_i) - g(A_{i-1}) \tag{B.1}$$

Using (2.4) to expand $g(A_i)$, (B.1) can be expressed as

$$\frac{\partial E_g(h)}{\partial h(x_i)} = g^i + \lambda g^i g(A_{i-1}) = g^i \left(1 + \lambda g\left(A_{i-1}\right)\right) \tag{B.2}$$

Because $g^i > 0$, $0 \leq g\left(A_i\right) \leq 1$ and $\lambda \geq -1$, the expression $g^i \left(1 + \lambda g\left(A_{i-1}\right)\right) \geq 0$, which proves that the CFI is a monotonically increasing function with respect to $h(x)$.

B.2 Property 2

When $g^i = 1$ for all $i = 1, 2, \ldots, n$, then $g(A_i) = 1$ for all $i = 1, 2, \ldots, n$ (from (2.7) and (2.4)); therefore, the CFI becomes (using (3.5))

$$E_g(h) = \sum_{i=1}^{n} h\left(x_i\right)\left[g\left(A_i\right) - g\left(A_{i-1}\right)\right] = h\left(x_1\right) \tag{B.3}$$

since $g(A_0) = 0$. Because $h(x_1) \geq h(x_2) \geq \cdots \geq h(x_n)$, it follows that

$$h(x_1) = \max\left(h(x_1), h(x_2), \ldots, h(x_n)\right) \tag{B.4}$$

When $g^i = 0$ for all i, then, $g(A_n) = 1$, $g(A_i) = 0$ for all $i \neq n$ (from (2.7) and (2.4)). The CFI then becomes (using (3.5))

$$E_g(h) = \sum_{i=1}^{n} h(x_i)\left[g(A_i) - g(A_{i-1})\right] = h(x_n) \tag{B.5}$$

Because $h(x_1) \geq h(x_2) \geq \cdots \geq h(x_n)$, it follows that

$$h(x_n) = \min\left(h(x_1), h(x_2), \ldots, h(x_n)\right) \tag{B.6}$$

When $0 \leq g^i \leq 1$ for all i, then $g(A_n) = 1$, and $0 \leq g(A_i) \leq 1$ for all i. Since the maximum value of $h(x_1)$ is obtained only when all the density values are one, in all other cases $E_g(h) \leq x_{\max}$. Similarly, $x_{\min} \leq E_g$ since the minimum value is obtained only when all the densities equal zero.

B.3 Property 3

Using (3.4), $g(A_n) = 1$, and $h(x_{n+1}) \equiv 0$, we have

$$E_g(h) = \sum_{i=1}^{n} g(A_i)\left[h(x_i) - h(x_{i+1})\right] = g(A_n)\, h(x_n) = h(x_n) = c \tag{B.7}$$

since all the other terms vanish because the h terms cancel as i ranges from 1 to n.

B.4 Property 4

In (3.5), let $a_i = g(A_i) - g(A_{i-1})$ for all i. In (3.10) g is the same for both h_1 and h_2; hence,

$$\int_X h_1 \circ g = E_g(h_1) = \sum_{i=1}^{n} h_1(x_i)\left[g(A_i) - g(A_{i-1})\right] = h_1(x_1)\, a_1 + h_1(x_2)\, a_2 + \cdots + h_1(x_n)\, a_n$$

$$\leq h_2(x_1)\, a_1 + h_2(x_2)\, a_2 + \cdots + h_2(x_n)\, a_n = \int_X h_2 \circ g \tag{B.8}$$

since $h_1(x_i) \leq h_2(x_i)$ for all i.

B.5 Property 5

Let $g^i (i = 1, \ldots n)$ be the fuzzy densities of the universal set X, $B = \{x_1, \ldots, x_l\} \subset X$, $C = \{x_1, \ldots, x_m\}$ and $l < m \leq n$. Then using (3.4), we see that

$$\int_C h \circ g = \sum_{i=1}^{m} g(A_i)[h(x_i) - h(x_{i+1})]$$

$$= \sum_{i=1}^{l} g(A_i)[h(x_i) - h(x_{i+1})] + \sum_{i=l+1}^{m} g(A_i)[h(x_i) - h(x_{i+1})]$$

$$= \int_B h \circ g + \sum_{i=l+1}^{m} g(A_i)[h(x_i) - h(x_{i+1})] \geq \int_B h \circ g \quad \text{(B.9)}$$

because $g(A_i) \geq 0$ and $h(x_i) - h(x_{i+1}) \geq 0$.

B.6 Property 6

Given that

$$g(A_n) = 1 = \sum_{j=1}^{n} g^j + f(\lambda) \quad \text{(B.10)}$$

where $f(\lambda)$ is a function of λ as in (2.5), and $\sum_{j=1}^{n} g^j = 1$, we see that $\lambda = 0$ for this equation to be satisfied. In this case, the $g(A_i)$'s become additive measures, i.e., $g(A_i) = \sum_{j=1}^{i} g^j$, and the CFI in (3.4) simplifies to (refer to (2.5)):

$$E_g(h) = \sum_{i=1}^{n} h(x_i)[g(A_i) - g(A_{i-1})] = \sum_{i=1}^{n} h(x_i)\left[\sum_{j=1}^{i} g^j - \sum_{j=1}^{i-1} g^j\right] \quad \text{(B.11)}$$

$$= \sum_{i=1}^{n} h(x_i) g^i \quad \text{(B.12)}$$

which is a weighted average. In the specific case when the densities all equal $1/n$, (B.12) simplifies further to:

$$E_g(h) = \frac{1}{n} \sum_{i=1}^{n} h(x_i) \quad \text{(B.13)}$$

which is the arithmetic mean.

B.7 Property 7

If $g^j = 0$, then according to (2.4)

$$g(A_j) = g^j + g(A_{j-1}) + \lambda g^j g(A_{j-1}) = g(A_{j-1}) \tag{B.14}$$

In this case, (3.5) becomes

$$E_g(h) = \sum_{i=1}^{n} h(x_i) \left[g\left(A_i\right) - g\left(A_{i-1}\right)\right] = \sum_{i=1, i \neq j}^{n} h(x_i) \left[g\left(A_i\right) - g\left(A_{i-1}\right)\right] \tag{B.15}$$

since the $g(A_i)$ terms cancel out when $i = j$.

Bibliography

[1] S. Auephanwiriyakul, J. M. Keller, and P. D. Gader, "Generalized Choquet fuzzy integral fusion," *Information Fusion*, vol. 3, no. 1, pp. 69–85, 2002.

[2] J.-H. Chiang, "Choquet fuzzy integral-based hierarchical networks for decision analysis," *IEEE Trans on Fuzzy Systems*, vol. 7, pp. 63–71, 1999.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: John Wiley & Sons, Inc., 2001.

[4] J. Keller, P. Gader, H. Tahani, J. H. Chiang, and M. Mohamed, "Advances in fuzzy integration for pattern recognition," *Fuzzy Sets and Systems*, vol. 65, pp. 273–283, 1994.

[5] L. I. Kuncheva, *Fuzzy Classifier Design*. New York: Physica-Verlag, 2000.

[6] J. Liang, W. Yang, and X. Cai, "Decision fusion using the fuzzy integral method," *Proc of SPIE Conf. Sensor Fusion: Architecture, Algorithms, and Applications IV*, pp. 468–475, 2000.

[7] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure," *Fuzzy Sets and Systems*, vol. 29, pp. 201–227, 1989.

[8] L. Saputelli, M. Nikolaou, and M. J. Economides, "Self-learning reservoir management," *Society of Petroleum Engineers*, 2003.

[9] M. Sugeno, "Fuzzy measures and fuzzy integrals - a survey," *Fuzzy Automata and Decision Processes*, pp. 89–102, 1977.

[10] H. Tahani and J. Keller, "Information fusion in computer vision using the fuzzy integral," *IEEE Trans on Systems, Man and Cybernetics*, vol. 20, pp. 733–741, 1990.

[11] H. Wu and J. M. Mendel, "Binary classification of ground vehicles based on the acoustic data using fuzzy logic rule-based classifiers," Signal and Image Processing Institute, University of Southern California, Tech. Rep. 356, 2002.