Indices for Introspection on the Choquet Integral

Stanton R. Price[1], Derek T. Anderson[1], Christian Wagner[2], Timothy C. Havens[3], and James M. Keller[4]

[1]Electrical and Computer Engineering Department
Mississippi State University, Mississippi State, MS, USA
sp438@msstate.edu , anderson@ece.msstate.edu
[2]School of Computer Science, Horizon Digital Economy Research Institute
University of Nottingham, Nottingham, UK
christian.wagner@nottingham.ac.uk
[3]Electrical and Computer Engineering, Computer Science
Michigan Technological University, Houghton, MI, USA
thavens@mtu.edu
[4]Electrical and Computer Engineering
University of Missouri, Columbia, MO, USA
kellerj@missouri.edu

Abstract. *Fuzzy measures* (FMs) encode the *worth* (or importance) of different subsets of information sources in the *fuzzy integral* (FI). It is well-known that the *Choquet FI* (CFI) often reduces to an elementary aggregation operator for different selections of the FM. However, FMs are often learned from training data or they are derived from the densities (worth of just the singletons). In these situations an important question arises; what is the resultant CFI really doing? Is it aggregating data relative to an additive measure, a possibility measure, or something more complex and unique? Herein, we introduce new indices (distance formulas) and fuzzy sets that capture the degree to which the CFI is behaving like a set of known aggregation operators. This has practical application in terms of gaining a deeper understanding into a given problem, guiding new learning methods and evaluating the CFI's benefit.

DRAFT

1 Introduction

Fuzzy measures (FMs) are often used to encode the (possibly subjective) *worth* of different subsets of information sources. The *Choquet fuzzy integral* (CFI), an aggregation operator, is a way to combine the information encoded in a FM with the (objective) evidence or support of some hypothesis, e.g., sensor data, algorithm outputs, expert opinions, etc. FMs can be obtained in a number of ways: *quadratic program* (QP) [1], learning algorithm (e.g., genetic algorithm [2], punishment-reward [3], gradient descent [4]), S-Decomposable measure (e.g., Sugeno λ-fuzzy measure [5], belief measure), etc. However, the vast majority of FM learning and density deriving techniques (i.e., a FM acquired from just the worth of the singletons) do not provide us with any knowledge about how the resultant CFI is truly aggregating the data. Characterization methods are needed to help us discover what the CFI is really doing relative to a given (learned or derived) FM. The applicability of such a tool is wide-ranging: from an increased low-level understanding of how the sources are being fused to possibly helping us learn which sources are needed and which are providing negligible information to the problem at hand. Such a tool can also help us justify when the CFI is indeed doing something unique by acting outside the bounds of a simple known aggregation operator (e.g., weighted sum). It also has the potential to aid us in predicting the benefit of adding (or aiding the design of) a new sensor in multi-sensor fusion [6, 7].

Herein, we put forth new indices (distance formulas) and corresponding fuzzy sets that help assess the degree of similarity a learned or density-derived FM, which drives the behavior of the CFI, has to a subset of known aggregation operators. As stated above, there are many potential applications for indices that characterize the CFI. The work put forth herein is a first step in terms of identifying such indices. Later, focused work will improve upon these building blocks, indices, and explore their role in fusion.

The remainder of this article is structured as follows. Section 2 is a quick review of the FM, followed by learning and density-based derivation methods used herein to acquire it. Equations for characterizing the FM's behavior is introduced in Section 3. Last, preliminary experimental results are reported in Section 4.

2 Background

The FM does not possess the restrictive property of additivity. Instead, it has the weaker property of monotonicity (and typically normality). With respect to a set of N information sources, $X = \{x_1, ..., x_N\}$, the FM is often used to encode the (possibly subjective) worth of each subset in the power set 2^X. In the CFI, the FM is the mechanism that ultimately dictates how information is aggregated.

Definition 1. ***(Fuzzy Measure)*** *For a finite set of N information sources, X, the FM is a set-valued function $g : 2^X \rightarrow [0, 1]$ with the following conditions:*

1. *(Boundary Conditions) $g(\phi) = 0$ and $g(X) = 1$,*
2. *(Monotonicity) If A, $B \subseteq X$ with $A \subseteq B$, then $g(A) \leq g(B)$.*

Note, if X is an infinite set, there is a third condition guaranteeing continuity.

Definition 2. ***(Choquet Fuzzy Integral)*** *For a finte set of N information sources, X, FM g, and partial support function $h : X \rightarrow [0, 1]$, the CFI is*

$$\int h \circ g = \sum_{i=1}^{N} \omega_i h(x_{\pi(i)}), \tag{1}$$

DRAFT

where $\omega_i = (G_{\pi(i)} - G_{\pi(i-1)})$, $G_{(i)} = g(\{x_{\pi(1)}, ..., x_{\pi(i)}\})$, $G_{\pi(0)} = 0$, $h(x_i)$ *is the strength in the hypothesis from source* x_i, *and* $\pi(i)$ *is a sorting on* X *such that* $h(x_{\pi(1)}) \geq \ldots \geq h(x_{\pi(N)})$.

In contrast to relatively simple and classical additive aggregation techniques, e.g., weighted sum, the CFI is a more powerful aggregation operator that can account for important *interactions* (when/if present) between different subsets of information sources. Interactions, as referred to herein, are the combining of hypotheses between the subsets of X (e.g., combining $h(x_1)$ with $h(x_2)$ results in an increase or decrease to the overall confidence of a decision). The way in which these interactions impact the aggregated result is guided by the selection of FM. An example of the CFI is the following. Consider the case of multi-sensor (e.g., electro optical infrared (EO/IR), ground penetrating radar (GPR), and visual spectrum (VS)) explosive hazard detection. Furthermore, assume these sensors are co-registered, meaning, we know the mapping from one pixel, (i,j), in sensor (image) k to its corresponding pixel in sensor (image) m, or each source is individually mapped into some global coordinate system. Therefore, $N = 3$ and x_1 is EO/IR, x_2 is GPR, and x_3 is VS. For a specific world location, such as (33.4526 , −88.7874), a hypotheses might be, "there is an explosive hazard at this world location." For simplicity, assume the information provided by each sensor is its individual strength, in [0,1], in the above hypothesis. Thus, $g(x_1, x_3)$ is the importance of EO/IR and VS relative to answering this hypothesis and the CFI is the combined belief that the location is a hazard and dangerous.

Next, we review a few FM learning and density deriving methods since they guide the work and are a large part of the proposed experiments. Specifically, we review learning a FM from data via QP, the class of S-Decomposable measures, possibility and necessity measures, and the Sugeno λ-FM.

2.1 Quadratic Program for Learning the FM from Data

In [8], Grabisch shows how to compute the FM using QP. Training data is used and the FM that makes the FI fit the data with minimum *sum of squared error* (SSE) is used. Let $T = \{(o_j, \alpha_j) : j = 1, ..., m\}$ denote the set of training data, where o_j is the the j^{th} object/instance and $\alpha_j \in [0, 1]$ is the label of o_j. Referring to the multi-sensor EHD system example from Section 2, instance o_j would be a $[1 \times N]$ vector with confidence values for each of the N information sources and α_j would have a label of target or not target. With respect to instance j, the CFI can be compactly represented in linear algebra form as $\boldsymbol{A}_{o_j}^t \boldsymbol{g} + h(o_j; x_{\pi(N)})$, where $\boldsymbol{g}$ is a vector (of size $2^N - 1 \times 1$) of lexicographic ordered FM terms (e.g., $\boldsymbol{g}(1) = g(\{x_1\})$, $\boldsymbol{g}(2) = g(\{x_2\})$, $\boldsymbol{g}(N+1) = g(\{x_1, x_2\})$, $\boldsymbol{g}(2^N - 1) = g(\{x_1, x_2, \ldots, x_N\})$, etc.) and $\boldsymbol{A}_{o_j}$ is a vector (of size $2^N - 2 \times 1$) of differences in h values. For example, if the largest h value index is k, then $\boldsymbol{A}_{o_j}(k) = h(o_j; x_{\pi(1)}) - h(o_j; x_{\pi(2)})$. The FM is computed by minimizing

DRAFT

$$\frac{1}{2}\boldsymbol{g}^t(2\sum_{j=1}^{m}\boldsymbol{A}_{o_j}\boldsymbol{A}_{o_j}^t)\boldsymbol{g} + \sum_{j=1}^{m}(2(h(o_j;x_{\pi(N)}) - \alpha_j)\boldsymbol{A}_{o_j}^t)\boldsymbol{g}, \tag{2}$$

subject to $(-C)\boldsymbol{g} - \boldsymbol{b} \geq \boldsymbol{0}$ and $\boldsymbol{0} \leq \boldsymbol{g} \leq \boldsymbol{1}$, where $\boldsymbol{b}$ is a vector (of size $N(2^{N-1} - 1) \times 1$) of all 0's except for the last N entries, which are of value -1, and C is a matrix (of size $N(2^{N-1} - 1) \times 2^N - 2$) of monotonic fuzzy constraints (see [8] for details).

2.2 S-Decomposable Measure

The Sugeno λ-FM is one of the most widely used methods for deriving a FM from just the densities. However, the Sugeno λ-FM is a member of a much larger class of FMs, S-decomposable measures.

Definition 3. ***(S-Decomposable Measure)*** *Let S be a t-conorm. A FM g is called an S-decomposable measure if $g(\phi) = 0$, $g(X) = 1$, and for all A, B such that $A \cap B = \phi$,*

$$g(A \cup B) = S(g(A), g(B)).$$

2.3 Possibility and Necessity Measures

Popular examples of S-decomposable measures include the following.

Definition 4. ***(Possibility Measure)*** *A FM Π is called a possibility measure if $\Pi(\phi) = 0$, $\Pi(X) = 1$, if $A \subseteq B$, $\Pi(A) \leq \Pi(B)$, and*

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)). \tag{3}$$

Definition 5. ***(Necessity Measure)*** *A FM Nec is called a necessity measure if $Nec(\phi) = 0$, $Nec(X) = 1$, if $A \subseteq B$, $Nec(A) \leq Nec(B)$, and*

$$Nec(A \cap B) = \min(Nec(A), Nec(B)). \tag{4}$$

2.4 Sugeno λ-Fuzzy Measure

Definition 6. ***(Sugeno λ-Fuzzy Measure)*** *A measure g_λ is called a Sugeno λ-FM if it is a FM, [Def 1], and if for $A, B \subseteq X$ and $A \cap B = \phi$,*

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B), \lambda > -1.$$

Sugeno showed that λ can be found by solving

$$\lambda + 1 = \prod_{i=1}^{N}(1 + \lambda g(\{x_i\})), \tag{5}$$

where Sugeno showed that there is exactly one real solution such that $\lambda > -1$.

DRAFT

3 Indices for Analysis of the Choquet Integral

In this section, we put forth a new set of indices (distance formulas) and fuzzy sets that capture the degree to which the CFI is behaving like a known aggregation operator for a given FM. Specifically, we put forth definitions for max, min, average, the generic case of an *ordered weighted average* (OWA), and the degree to which a FM is in accordance with a possibility FM. These FMs are based on the theoretical investigation of the CFI for different FMs by Keller et al. [9].

Before moving on, a few terms used herein need to be established. First, a *lattice* is induced according to the set of monotonicity constraints of the FM. It is simple to think about and talk about a FM in terms of its lattice. Second, a *node* in the lattice refers to the value of the FM for a given *non-empty* subset of X. *Layer* k in the lattice, denoted $L(k)$, is the set of all FM terms for subsets of 2^X that have cardinality equal to k. For example, if $N = 3$, $L(1) = \{g(x_1), g(x_2), g(x_3)\}$, $L(2) = \{g(x_1, x_2), g(x_1, x_3), g(x_2, x_3)\}$, and $L(3) = g(x_1, x_2, x_3) = g(X)$.

DRAFT

3.1 Maximum Operator

It is well-known that the CFI is a max operator when the value at each node (used herein when referring to non-empty subsets of X) in the lattice is 1. A naive approach for measuring the distance in which a given FM behaves like the max operator is

$$\frac{1}{2^N - 2} \sum_{I \in 2^X \setminus \{\phi, X\}} (1 - g(I)). \tag{6}$$

Equation 6 is a distance, which has a value of 0 when we have a max FM and a value 1 in the case of minimum (dual of max). It can be easily converted into a similarity by subtracting it from 1. However, Eqn. 6 is lacking as it assigns equal importance to each node in the lattice and it does little-to-nothing to guarantee the specific properties that we expect a max operator to possess. It is also very hard to interpret what the distance is really measuring (outside comparing it to the extreme case of min). It appears, to us, that whatever the measure is, it should calculate its distance relative to a FM of all 1s (max FM). We assert that a quality distance measure should enforce the idea that a given FM needs to indeed behave like an OWA and differences at different layers in the lattice are not equal (as we expect max to move to soft max and then a different operator from there). We begin by assigning a weight (penalty or cost) for each layer in the FM,

$$\boldsymbol{W} = \frac{\left[\frac{1}{N}, \ldots, 1\right]}{\sum_{i=1}^{N} \frac{i}{N}}. \tag{7}$$

Now, we put forth a formula that adheres to the concepts discussed above,

$$D_{\max} = \sum_{k=1}^{1} \frac{\boldsymbol{W}(k)}{2}(T_1 + T_4) + \left[\sum_{k=2}^{N} \frac{\boldsymbol{W}(k)}{3} (T_1 + T_2 + T_4) \right], \tag{8}$$

$$T_1 = 1 - \left(\frac{\sum_{I \in L(k)} g(I)}{|L(k)|} \right), \tag{9}$$

$$T_2 = \left(\frac{\sum_{I \in L(k)} g(I)}{|L(k)|} \right) - \left(\frac{\sum_{J \in L(k-1)} g(J)}{|L(k-1)|} \right), \tag{10}$$

$$T_3 = \frac{\sum_{I \in L(k)} g(I)}{|L(k)|}, \tag{11}$$

$$T_4 = \frac{\sum_{I \in L(k)} (g(I) - T_3)^2}{|L(k)| - 1}. \tag{12}$$

Specifically, T_1 measures the difference from the expected value of 1, T_2 is the extraction of the OWA weights (which should be $[1, 0, \ldots, 0]$ for max), and T_4 is an unbiased estimator of the variance of the values at layer k in the lattice. While $D_{\max}$ measures the distance of a given FM to max, what we really want is a membership degree in which 0 means not max and 1 means it is max. However, how should this membership function be defined on the domain $D_{\max}$? We propose that one specifies a definition for soft max and measures its corresponding $D_{\max}$ value. For example, in our experiments section we used $N = 3$ and obtained $D_{\max} \approx 0.0637$. Next, a membership function is selected; herein, we used a Gaussian shaped membership function whose peak is centered at 0 and σ is derived so that the 0.5 membership degree is at the corresponding location ($D_{\max}$) for soft max. In this respect, one calibrates the concept of compatibility relative to when we expect a FM to become increasingly dissimilar to max.

3.2 Minimum Operator

With respect to the min operator, each node in the lattice, except for $g(X)$ which is 1 by definition, should be 0. We also expect that the FM is an OWA and we calibrate it relative to soft min. As min and max are duals, their resulting calculations are largely similar,

$$D_{\min} = \sum_{k=1}^{1} \frac{\boldsymbol{W_2}(k)}{2} (T_3 + T_4) + \left[\sum_{k=2}^{N-1} \frac{\boldsymbol{W_2}(k)}{3} (T_3 + T_2 + T_4) \right], \tag{13}$$

where the weights are

$$\boldsymbol{W_2} = \frac{\left[1, \ldots, \frac{1}{N-1}\right]}{\sum_{i=1}^{N-1} \frac{i}{N-1}}. \tag{14}$$

DRAFT

Note, T_3 measures the difference of a layer to the expected value of 0. Furthermore, we follow a similar calibration process to that outlined above (Sec. 3.1) to calibrate min, with the only difference being we specify a definition for soft min instead of soft max.

3.3 Average Operator

A unique property of the average operator, with respect to the FM, is it has value $\frac{k}{N}$ at layer k in the lattice. This is captured in the following equation,

$$D_{avg} = \frac{1}{2^N - 2} \sum_{k=1}^{N-1} \sum_{I \in L(k)} \left| g\left(I\right) - \frac{k}{N} \right|. \tag{15}$$

Unlike max and min, it is not as clear what the membership function and specific calibration (point where the membership drops off to value 0.5 and shape of the function) should be. It could however, be calibrated or fit to a user's specification or application. Herein we fit a z-membership function which was calibrated to have membership value 0.5 at $\left(1/N\right)^2$. Note that the value used for calibration was experimentally determined based on our preliminary experiments.

3.4 OWA Operator

Based on the definitions above, measuring the degree to which the CFI, for a given FM, is compatible with an OWA operator is relatively simple. The criteria is that sets of equal size cardinality in the lattice have equal measure value,

$$D_{OWA} = \frac{1}{N-1} \sum_{k=1}^{N-1} \sqrt{T_4}. \tag{16}$$

Again, it is not obvious how to take this distance and construct a membership function that everyone would agree with that captures the exact rate at which a given FM is moving away from being an OWA operator. The membership function can be specified by a user or specific application. Herein, we fit a triangular membership function across the domain $[0, 1]$ such that the left and center points are at 0 and the right point is at 1. This membership function was chosen intuitively and supported experimentally through our preliminary experiments.

3.5 Possibility Measure

The final index put forth here captures the degree to which the CFI is aggregating data under the guidance of a possibility measure. A possibility measure has the unique distinction that for $A \subset X$, $g(A)$ is the maximum over the densities corresponding to A. We therefore put forth the following distance measure,

$$D_{Poss} = \frac{\sum_{A \in 2^X \setminus \{\phi, \{x_1\}, \ldots, \{x_N\}, X\}} |g(A) - \max_{i \in A} g(\{x_i\})|}{2^N - 2 - N}. \tag{17}$$

DRAFT

As in the sections above, we empirically calibrated the membership function. Herein, we fit a triangular membership function to D_{Poss} such that the left and center points are at 0 and the right point is at 1.

4 Preliminary Results

In this section, we report preliminary results that demonstrate the behavior of our proposed indices. Once again, these indices are put forth in order to characterize the behavior of the CFI. The goal of this work is to explore a few such indices. The focus of this work is not to vigorously analyze the results of different learning or data driven experiments for different application domains. That is the subject of future work. Two sets of experiments were conducted: using known FMs (Sec. 4.1), and using learned and density derived FMs (Sec. 4.2). For simplicity of reporting and describing the following experiments, the following cases are performed for $N = 3$.

4.1 Experiment 1: Known Aggregation Operators

The experiments presented in Table 1 are put forth to demonstrate the behavior of our indices with respect to known user specified FMs.

Table 1. Results of proposed indices for the case of known FMs and $N = 3$.

	User Specified FM		**Max**	**Min**	**Avg**	**OWA**	**Poss**
(a)	Max	dist.	0	0.44	0.5	0	0
		memb.	1	0	0	1	1
(b)	Soft Max OWA weights of (0.7, 0.2, 0.1)	dist.	0.07	0.35	0.3	0	0.2
		memb.	0.5	0	0	1	0.8
(c)	Min	dist.	0.36	0	0.5	0	0
		memb.	0	1	0	1	1
(d)	Average	dist.	0.18	0.22	0	0	0.33
		memb.	0.01	0.01	1	1	0.66
(e)	OWA weights of (0.52, 0.08, 0.40)	dist.	0.16	0.24	0.12	0	0.08
		memb.	0.04	0	0.35	1	0.92
(f)	Possibility Measure densities of (0.4, 0.6, 0.8)	dist.	0.12	0.31	0.18	0.15	0
		memb.	0.13	0	0.02	0.84	1
(g)	Sugeno λ-FM densities of (0.1, 0.1, 0.1)	dist.	0.28	0.1	0.27	0	0.25
		memb.	0	0.41	0	1	0.74
(h)	Sugeno λ-FM densities of (0.5, 0.2, 0.9)	dist.	0.13	0.35	0.25	0.27	0.07
		memb.	0.11	0	0	0.72	0.92

DRAFT

As expected, each of the simple aggregation operators have full membership to their respective index. Also, note that a given FM can be many types of FMs at once (i.e., for the max FM, it is also an OWA and possibility FM). Further, soft max, 1.(b), returns a membership of 0.5 to the max index, and the possibility measure, 1.(f), has full membership for its respective index. Interestingly, both of the Sugeno λ-FMs, one being subadditive, 1.(g), and the other superadditive, 1.(h), have relatively high possibility memberships, which indicate it is acting in accordance to a possibility FM despite the disparity between the two densities.

4.2 Experiment 2: Learned and Density Derived FMs

The baseline experiments presented in Table 2 are put forth to demonstrate the behavior of our indices with respect to learned and density derived FMs. For the case of the QP, 100 random samples were produced, labels were generated using a known FM (which varies per experiment), and the QP was then used to approximate the FM from that data (input plus known labels).

Table 2. Results of proposed indices for learned and density derived FMs for $N = 3$.

	Method used to obtain FM		Max	Min	Avg	OWA	Poss
(a)	QP fit to data from soft max FM OWA weights of (0.7, 0.2, 0.1)	dist. memb.	0.07 0.5	0.35 0	0.3 0	0 1	0.2 0.8
(b)	QP fit to 0.9 of OWA and 0.1 noise[1] OWA used was $[0.52, 0.08, 0.4]$	dist. memb.	0.14 0.07	0.26 0	0.1 0.51	0 0.99	0.11 0.88
(c)	Sugeno λ-FM using densities learned in Table 2.b	dist. memb.	0.11 0.19	0.3 0	0.19 0.01	0 0.99	0.29 0.7
(d)	QP fit to random (but valid) FM FM of $(0.5, 0.1, 0.11, 0.51, 0.79, 0.89, 1)$	dist. memb.	0.2 0	0.23 0	0.18 0.02	0.21 0.78	0.36 0.64
(e)	QP fit to valid FM FM of $(0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 1)$	dist. memb.	0.24 0	0.14 0.14	0.16 0.1	0 1	0 1

The first experiment, shown in Table 2.(a), shows that the indices are still performing as expected when the FMs are learned using a QP. More interestingly, however, is comparing experiments 2.(b)-(c) with 1.(e). Memberships for all indices in 2.(b), are relatively close to the "calibrated" (expected) results computed in 1.(e) even with a slight amount of noise added to the OWA. However, 2.(c), the Sugeno λ-FM derived using the densities found in 2.(b), does not return memberships consistent with that found in 2.(b), which is undesired if we recognize

[1] In the case of noise, 90% of the label is from the OWA and 10% is random noise from the interval $[0, 1]$.

DRAFT

2.(b) as being the optimal FM for that data. In fact, the Sugeno λ-FM appears to make the CFI act a bit more like a max operator. Additionally, as seen in 1.(g)-(h), the Sugeno λ-FM continues to have a high membership to behaving in accordance to possibility FMs. The last two experiments, 2.(d)-(e), show that the indices are able to characterize the behavior of FMs picked at random (but checked to ensure monotonicity) in a manner that is expected.

These preliminary results provide a proof of concept for the indices put forth herein. From these experiments it has been shown that the indices do return characterizations of the FM that reflect what one may expect to see for a given FM. It is important to note that a thorough investigation with respect to the experimental application of these indices is needed and planned as future work to further develop this FM characterization tool. This set of preliminary results gives a glimpse at the utility and need for such a tool; however, further analysis needs to be done. For example, a more intuitive and defined method for creating the fuzzy sets needs to be introduced, as these dictate the membership degree. Though there are areas of these indices that can likely be refined, this initial set of CFI characterization tools begins a movement towards filling a void in the understanding of what the CFI is really doing and how it is behaving.

5 Summary

In this paper, we put forth a set of new indices (distance formulas) and fuzzy sets to help address the question of what the CFI is doing for the case of learned or density derived FMs. The proposed tools have practical application in terms of gaining a deeper understanding into a given problem, guiding new learning methods (i.e., learning a FM that is a possibility measure, additive measure, etc.) and evaluating the benefit of using the CFI. Herein, we defined a few initial indices and we presented preliminary results for the behavior of these indices for the cases of a number of known FMs (and thus aggregation operators), simple learned (using the QP), and density derived FMs. While the results are preliminary, we believe they show the promise of such an approach to formally characterize the FM and CFI. In future work we will improve upon the definitions and further refine (calibrate) the fuzzy sets. In addition, we will then look to (experimentally) use these indices to help discover interesting and unknown behaviors of FMs and CFIs for the bigger purpose of guiding new theorems to help us better understand the aggregation behavior of mutli-sensor fusion systems.

DRAFT

References

1. Grabisch, M., Nguyen, H.T., Walker, E.A.: Fundamentals of uncertainty calculi with applications to fuzzy inference. Kluwer Academic Publishers (1994)
2. Anderson, D., Keller, J., Havens, T.: Learning fuzzy-valued fuzzy measures for the fuzzy-valued sugeno fuzzy integral. In: Computational Intelligence for Knowledge-Based Systems Design. Volume 6178 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2010) 502–511

3. Keller, J.M., Osborn, J.: Training the fuzzy integral. International Journal of Approximate Reasoning **15**(1) (1996) 1 – 24
4. Mendez-Vazquez, A., Gader, P., Keller, J., Chamberlin, K.: Minimum classification error training for choquet integrals with applications to landmine detection. Fuzzy Systems, IEEE Transactions on **16**(1) (2008) 225–238
5. Sugeno, M.: Theory of Fuzzy Integrals and Its Applications. Ph.D. thesis. Tokyo Institute of Technology (1974)
6. Keller, J.M., Auephanwiriyakul, S., Gader, P.D.: New fuzzy set tools to aid in predictive sensor fusion. Proc. SPIE **4038** (2000) 1497–1507
7. Keller, J.M., Auephanwiriyakul, S., Gader, P.D.: Experiments in predictive sensor fusion. Proc. SPIE **4394** (2001) 1047–1058
8. Grabisch, M., Roubens, M.: Application of the choquet integral in multicriteria decision making. In: Fuzzy measures and integrals, Physica Verlag (2000) 348–374
9. Keller, J.M., Gader, P., Tahani, H., Chiang, J.H., Mohamed, M.: Advances in fuzzy integration for pattern recognition. Fuzzy Sets and Systems **65**(23) (1994) 273 – 283

DRAFT