

Havens, Timothy C. and Anderson, Derek T. and Wagner, Christian (2015) Data-informed fuzzy measures for fuzzy integration of intervals and fuzzy numbers. IEEE Transactions on Fuzzy Systems, 23 (5). pp. 1861-1875. ISSN 1941-0034

Access from the University of Nottingham repository:

http://eprints.nottingham.ac.uk/37872/1/Data-Informed%20Fuzzy%20Measures%20for %20Fuzzy%20Integration%20of%20Intervals%20and%20Fuzzy%20Numbers.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the University of Nottingham End User licence and may be reused according to the conditions of the licence. For more details see: http://eprints.nottingham.ac.uk/end_user_agreement.pdf

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk

Data-Informed Fuzzy Measures for Fuzzy Integration of Intervals and Fuzzy Numbers

Timothy C. Havens, Senior Member, IEEE, Derek T. Anderson, Senior Member, IEEE, and Christian Wagner, Senior Member, IEEE

Abstract-The fuzzy integral (FI) with respect to a fuzzy measure (FM) is a powerful means of aggregating information. The most popular FIs are the Choquet and Sugeno and most research focuses on these two variants. The arena of the FM is much more populated, including numerically-derived FMs such as the Sugeno λ -measure and decomposable measure, expert-defined FMs, and data-informed FMs. The drawback of numerically-derived and expert-defined FMs is that one must know something about the relative values of the input sources. However, there are many problems where this information is unavailable, such as crowd-sourcing. This paper focuses on datainformed FMs, or those FMs that are computed by an algorithm that analyzes some property of the input data itself, gleaning the importance of each input source by the data they provide. The original instantiation of a data-informed FM is the agreement FM, which assigns high confidence to combinations of sources that numerically agree with one another. This paper extends upon our previous work in data-informed FMs by proposing the uniqueness measure and additive measure of agreement for interval-valued evidence. We then extend data-informed FMs to fuzzy number (FN)-valued inputs. We demonstrate the proposed FMs by aggregating interval and FN evidence with the Choquet and Sugeno FIs for both synthetic and real-world data.

Index Terms—fuzzy measure; fuzzy integral; data fusion; sensor fusion; Sugeno integral; Choquet integral

I. INTRODUCTION

Aggregating multiple information sources is one basic approach used to answer questions or prove hypotheses. A way that this is done with fuzzy sets is with the *fuzzy integral* (FI). The discrete FI aggregates sources of information by a weighted aggregation, where the weights are computed by a *fuzzy measure* (FM) that models the (typically subjective) "worth" of subsets of the sources. In most applications, the measure is specified by an expert or learned (e.g., by a genetic algorithm [1, 2]).

FIs and FMs have been proposed for many applications and for many types of data, from simple numeric data to intervals

This work was funded in part by the U.S. Army (W909MY-13-C0013, W909MY-13-C0029) and Army Research Office (W911NF-14-1-0114) in support of the U.S. Army RDECOM CERDEC NVESD, the RCUK Horizon Digital Economy Research Hub Grant (EP/GO65802/1), the EPSRC Towards Data-Driven Environmental Policy Design (EP/K012479/1), and by the National Institute of Justice (2011-DN-BX-K838).

T.C. Havens is with the Departments of Electrical and Computer Engineering and Computer Science, Michigan Technological University, Houghton, MI USA (e-mail: thavens@mtu.edu).

D.T. Anderson is with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS USA (e-mail: anderson@ece.msstate.edu).

C. Wagner is with the School of Computer Science, University of Nottingham, Nottingham, UK (e-mail: christian.wagner@nottingham.ac.uk).

and type-2 fuzzy sets [1-16]. While manual specification of the FM works for small sets of sources (there are already 16 possible combinations of sources in the power set of 4 sources), manually specifying the values of the FM for large collections of sources is all but impossible. Thus, automatic methods have been proposed, such as the Sugeno λ -measure [17] and the decomposable measure [18], which build the measure from the densities (the worth of individual sources), and genetic algorithm [1, 2], Gibbs sampling [19] and other learning methods [20, 21], which build the measure by using training data. These methods work well for scenarios in which the worth of individual sources is known or training data are available, but one could easily imagine aggregation problems for which this information is unavailable (i.e., all problems where no solution-ground truth-is known, thus training or learning as such is not possible).

One area where this scenario is encountered is crowdsourcing. In this application, very little or no information is available about the overall solution or the individual contributors and, hence, each individual is usually considered as having an equally worthy contribution. However, this is clearly not the reality as the crowd is usually composed of people with varying levels of expertise on the topic. Furthermore, some individuals may not have altruistic aims and may knowingly (or unknowingly) contribute erroneous information. In [22, 23], methods were developed to filter contributors by ranking them; however, these methods are purpose-built strategies and are not appropriate for generalized approaches to aggregating crowdsourced information.

Another area where this scenario appears is that of battlefield situation awareness and intelligence gathering. While commanders are very good at applying relative worth to multiple sources of information by extracting qualitative features such as inflection, past experience, and intuition / instinct, it is still a grand challenge for computers to accurately interpret human communication in an efficient manner. The measures we propose can be used by analysts to examine information aggregation with respect to mathematical representations of human-based aggregation factors, such as agreement, accord, uniqueness, and specificity.

In our proposed approach—following on from our original work on this topic [24, 25]—we aim to use the overall data set to extract the worth of individual sources. In [24], we focused on the creation of meta-measures (viz., combinations of measures) with the purpose of aggregating based on multiple criteria such as agreement and specificity, and in [25] we further extended the notion of agreement to propose

TABLE I ACRONYMS AND NOTATION

FS	fuzzy set		FN	fuzzy number
FM	fuzzy measure	e	AG	agreement
GenA	generalized a	ecord	AA	additive agreement
SP	specificity		UQ	uniqueness
	$X = \{x_i\}$	set of sour	ces (sen	sors, inputs, etc.)
	$h(x_i), h_i$	real-valued	eviden	ce offered by x_i
\bar{h}_i interval-val			ued evi	dence offered by x_i
\hat{H}_i fuzzy num		ber-valu	ed evidence offered by x_i	
$\stackrel{ riange}{H}$ triangular f			uzzy nu	ımber
q^i density, nu		meric worth of x_i		
\hat{H}_{α} level-cut of		f fuzzy-	number \hat{H} at α	
π permutation		n functio	on of h	
	\bar{A}_i	permuted	set	of evidence, $\bar{A}_i =$
		$\{\bar{h}_{\pi(1)},\ldots,$	$., \bar{h}_{\pi(i)}$	}
$\bar{U}_K(\bar{A}_i), \hat{U}_K(\hat{A}_i)$ sets defined		d in equ	ations (8) and (29), respec-	
		tively		· · · •
	i = [m]	notation fo	r i = 1	$, 2, \ldots, m$
	i = [l:m]	notation fo	r i = l,	$l+1,\ldots,m$
	$\int h \circ g$	fuzzy integ	ral of h	a w.r.t. g
	$\int f(a)da$	Riemann in	ntegral o	of $f(a)$ w.r.t. a
$\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_{n_s})^T$ discrete fuz			zzy set s	sampled at $(y_1, \ldots, y_{n_s})^T$

the FM of *generalized accord* (GenA). In this paper, we do a complete investigation of the work initially proposed in [24, 25], propose an expanded set of data-derived FMs, and develop ways in which we can characterize the behavior of these FMs in relation to common aggregators such as min, max, and average.

Section II introduces the FM and FI, then Section III describes methods for characterizing the behavior of a given FM in terms of its similarity to max, min, and average aggregation operators. In Section IV, we first discuss the existing FMs of agreement and generalized accord [24, 25], exposing drawbacks to these FMs. To combat these drawbacks, we propose the additive FM of agreement, which, like GenA, is based on aggregation of set-similarity functions (i.e., Jaccard and Dice). We then propose a new FM of uniqueness, which is the antithesis (or complement) of the agreement-type measures. The last FM we discuss is the measure of specificity. The FMs proposed in Section IV only work with interval-valued evidence. Thus, we move to Section V, which describes the FN extensions of the FMs proposed in Section IV. Lastly, in Section VI we demonstrate the FMs by using the FI to aggregate information in both synthetic and real-world data. To assist the reader with the numerous acronyms and notation, we have compiled a selected list in Table I.

II. FUZZY MEASURES AND INTEGRALS

Let $X = \{x_1, \ldots, x_n\}$ be a non-empty finite set (typically of information sources or evidence) and $g: 2^X \to [0, \mathbb{R}^+]$ be a FM with the following properties [3]:

(P1) $g(\emptyset) = 0;$

(P2) If $A \subseteq B \subseteq X$ then $g(A) \leq g(B) \leq 1$ (g is monotonic and non-decreasing).

Note that there is a third property for continuous FMs which is not applicable in the case of discrete FMs such as those used in this paper. Also, in most cases, the worth of the universal set is constrained to 1, i.e., g(X) = 1, and we will assume this for the rest of this article. The measure g is the (possibly subjective) confidence or worth of each subset of X; hence, P1 tells us that the worth of no sources, viz., the empty set \emptyset , is 0. P2 follows intuition, in that if sources B contains sources A, then B is worth as much, if not more, than A.

A couple well-known FMs are the Sugeno λ -measure [17] and the decomposable measure [18]. Both are calculated by a computational method where one only has to provide the measure values of the densities, i.e., $g^i = g(x_i)$. The FM values for the non-singleton sets are built from the g^i values. However, this is the drawback of these FMs, and others like them. They do not consider any synergistic combination of sources, resulting in a boost of the FM values for combinations of complementary sources. This drawback is the main reason for our proposed research on this topic.

There are many forms of the FI; see [3] for detailed discussion. In practice, FIs are mostly used for evidence fusion [1, 4–7]. They combine sources of information by accounting for both the support of the question (the evidence) and the expected worth of each subset of sources (as supplied by the FM g). Here, we focus on the discrete fuzzy Sugeno and Choquet integrals, proposed by Murofushi and Sugeno [26, 27]. Let $h : X \to \mathbb{R}$ be a real-valued function that represents the evidence or support of a particular hypothesis.¹ The discrete fuzzy Sugeno and Choquet integral are defined, respectively, as

$$\int_{S} h \circ g = S_g(h) = \bigvee_{i=1}^{n} \left(h(x_{\pi(i)}) \wedge g(A_i) \right), \tag{1a}$$

$$\int_{C} h \circ g = C_g(h) = \sum_{i=1}^{n} h(x_{\pi(i)}) \left[g(A_i) - g(A_{i-1}) \right], \quad (1b)$$

where π is a permutation of X, such that $h(x_{\pi(1)}) \ge h(x_{\pi(2)}) \ge \ldots \ge h(x_{\pi(n)})$, $A_i = \{x_{\pi(1)}, \ldots, x_{\pi(i)}\}$, and $g(A_0) = 0$ [10, 28]. Detailed treatments of the properties of FIs can be found in [10, 28, 29].

In some cases, the evidence h cannot, or should not, be represented simply by numbers; h would be better represented as an interval-valued or FN-valued function. An example is the survey question, "How many bottles of wine should I purchase for the reception?" Many people would answer this question with an interval, e.g., "between 20 and 30," or a FN, e.g., "about 25." Furthermore, there is much work on how to extract intervals and fuzzy sets from data in real-world problems [30, 31]. Extensions of both the fuzzy Sugeno and fuzzy Choquet integral have been proposed for both interval-valued and also FN-valued integrands [8–12], which we now discuss.

Let $I(\mathbb{R}) = \{\bar{u} \subset \mathbb{R} | \bar{u} = [u^-, u^+], u^- \leq u^+\}$ be the set of all closed intervals over the real numbers. Dubois and Prade showed that if a function ϕ is continuous and non-decreasing, then, when defined on intervals, ϕ produces an interval the endpoints of which are equal to the function values on the lower and upper bound of the individual intervals, viz., $\phi(\bar{u}) = [\phi(u^-), \phi(u^+)]$. This approach benefits us in computing FIs

¹Generally, when dealing with information fusion problems it is convenient to have $h: X \to [0, 1]$, where each source is normalized to the unit-interval. This is also the appropriate space of h for use with the Sugeno FI when using the definition $g \in [0, 1]$.

for intervals as they are continuous and non-decreasing. Let $\bar{h}: X \to I(\mathbb{R})$, where $\bar{h}_i = \bar{h}(x_i) = [h_i^-, h_i^+]$ is intervalvalued integrand (evidence). The FI on \bar{h} is

$$\int \bar{h} \circ g = \left[\int h^{-} \circ g, \int h^{+} \circ g \right], \qquad (2)$$

where the output $\int \bar{h} \circ g$ is itself interval-valued.

Now, let $\hat{H}: X \to FN(\mathbb{R})$ be a FN-valued integrand. Using the Extension Principle, the FI of \hat{H} w.r.t. g is then defined as

$$\left(\int \hat{H} \circ g\right)(a) = \sup_{\alpha \in [0,1]} \left\{ a \in \int {}^{\alpha} \hat{H} \circ g \right\}, \qquad (3)$$

where ${}^{\alpha}\hat{H} = [{}^{\alpha}h^{-}, {}^{\alpha}h^{+}]$ are the closed intervals of the levelcuts of the members of \hat{H} at α . Note that ${}^{\alpha}\hat{H}$ is not necessarily one continuous interval, but the set of intervals (or discontinuous interval) formed by taking the level-cut of each $\hat{H}_i \in \hat{H}$. Hence ${}^{\alpha}\hat{H} = \{[({}^{\alpha}\hat{H}_1)^{-}, ({}^{\alpha}\hat{H}_1)^{+}], \dots, [({}^{\alpha}\hat{H}_n)^{-}, ({}^{\alpha}\hat{H}_n)^{+}]\}$ and ${}^{\alpha}h^{-} = \{({}^{\alpha}\hat{H}^{1})^{-}, \dots, ({}^{\alpha}\hat{H}^{n})^{-}\}$ (and similarly defined for ${}^{\alpha}h^{+}$). Equation (3) can alternatively be written as

$$\left(\int \hat{H} \circ g\right)(a) = \bigcup_{\alpha \in [0,1]} \alpha \left[\left(\int^{\alpha} \hat{H} \circ g\right)(a) \right], \quad (4a)$$

$$= \bigcup_{\alpha \in [0,1]} \alpha \left[\left(\int^{\alpha} h^{-} \circ g \right)(a), \left(\int^{\alpha} h^{+} \circ g \right)(a) \right], \quad (4b)$$

where a is the independent variable of the FI (i.e., $a \in [0, 1]$ for the Sugeno FI or $a \in \mathbb{R}$ for the Choquet FI).

Other extensions to FIs are proposed for sub-normal FSs [1, 12], sub-normal non-convex FSs [32], type-2 FNs [11, 33, 34], and for non-numeric FMs [2, 35–38].

III. INDICES OF FUZZY MEASURE BEHAVIOR

While there are several ways to compose a FM—e.g., the aforementioned Sugeno λ -measure and decomposable measure, or by experts—there are also measures that produce well-known aggregation operators directly when coupled with the Choquet FI. For example, with respect to the Choquet FI, the average, minimum, and maximum manifest as the following measures, respectively,

$$g^{AVG}(A_i) = i/n, \quad i = [0:n],$$
 (5a)

$$g^{MIN}(A_i) = \begin{cases} 1 & i = n, \\ 0 & i < n, \end{cases}$$
(5b)

$$g^{MAX}(A_i) = \begin{cases} 1 & i > 0, \\ 0 & i = 0. \end{cases}$$
(5c)

Hence, FMs can be characterized by comparing them with these well-known operators, e.g., FMs that have values all near to i/n mimic the average. We now propose an index of similarity by which any FM can be compared to the operators at (5).

Consider the distance between two measures, g^1 and g^2 , as

$$\Delta(g^1, g^2; A) = \frac{\|\mathbf{g}^1 - \mathbf{g}^2\|^2}{n-1},$$

where $\mathbf{g} = (g(A_1), \dots, g(A_{n-1}))^T$ is the vector of the (possibly unique) FM values computed on each of A_i , i = [n-1],



Fig. 1. Membership function with input of similarity value S and output of membership in the respective aggregation operator.

and $\|\cdot\|^2$ is squared Euclidean distance. The FM values, $g(A_1)$ to $g(A_{n-1})$, are only used because $g(A_0) = 0$ and $g(A_n) = 1$; thus, these two boundary conditions on the measure provide no information on the distance between two measures. It can be shown that $0 \leq \Delta(g^1, g^2; A) \leq 1$; hence, the similarity between two measures is computed as

$$\mathcal{S}(g^1, g^2; A) = 1 - \Delta(g^1, g^2; A).$$
(6)

We use $S(g^1, g^2; A)$ to compare any given measure to the average, minimum, and maximum measures (w.r.t. the Choquet integral) at (5). These prototype similarities are notated as $S_{AVG}(g) = S(g, g^{AVG})$, $S_{MAX}(g) = S(g, g^{MAX})$, and $S_{MIN}(g) = S(g, g^{MIN})$, respectively. We drop the notation of the set ordering A, assuming that g and, say, g^{AVG} are computed on the same set ordering. However, the next remark illustrates an intuitive problem with only using the similarities (and not some processed form of them as proposed next).

Remark 1. It is interesting to note that $S_{MIN}(g^{MAX}) = S_{MAX}(g^{MIN}) = 0$ (as they are duals) and $S_{MIN}(g^{AVG}) = S_{MAX}(g^{AVG}) = S_{AVG}(g^{MIN}) = S_{AVG}(g^{MAX}) = 1 - \sum_{i=1}^{n-1} (i/n)^2/(n-1)$. In the limit as $n \to \infty$, the similarity of g^{AVG} to both g^{MIN} and g^{MAX} goes to 2/3. Furthermore, it can be shown that, $\forall g$,

$$0 \le S_{MIN}(g) \le 1; \quad 0 \le S_{MAX}(g) \le 1;$$

$$2/3 \le 1 - \sum_{i=1}^{n-1} (i/n)^2 / (n-1) \le S_{AVG}(g) \le 1.$$

Because of this, we believe that this similarity index is much better at determining the behavior of a FM relative to min and max, but not as resolved when comparing to an average aggregator. Also, intuitively we wish to create a method to index the behavior of a FM such that it is definitively (in the fuzzy sense) one of average, min, or max. Hence, we propose the fuzzy membership function in Fig. 1, which takes as input $S_{AVG}(g)$, $S_{MIN}(g)$, or $S_{MAX}(g)$, and returns the membership of the FM in the respective aggregation operator. We will use the similarity function at (6) together with the membership function

$$\mu(\mathcal{S}) = \begin{cases} 0 & \mathcal{S} < 2/3, \\ 3\mathcal{S} - 2 & \mathcal{S} \ge 2/3, \end{cases}$$

as shown in Fig. 1 to characterize the data-derived FMs in our experiments.

IV. DATA-INFORMED FUZZY MEASURES ON INTERVALS

Consider a collection of inputs X, such that you have no prior knowledge about the relative worth of each input. Examples include crowd-sourced informations, ranges of sensor readings, and anonymous survey results.

Consider the interval-valued evidence $\bar{h} = {\bar{h}_1, ..., \bar{h}_n}$. The FM of agreement proposed in [24] is computed as

$$\tilde{g}^{AG}(\bar{A}_0) = \tilde{g}^{AG}(\bar{A}_1) = 0,$$
(7a)

$$\tilde{g}^{AG}(\bar{A}_{i}) = \left| \bigcup_{k_{1}=1} \bigcup_{k_{2}=k_{1}+1} \bar{h}_{\pi(k_{1})} \cap \bar{h}_{\pi(k_{2})} \right| z_{2} +$$
(7b)
$$\left| \bigcup_{k_{1}=1}^{i-2} \bigcup_{k_{2}=k_{1}+1}^{i-1} \bigcup_{k_{3}=k_{2}+1}^{i} \bar{h}_{\pi(k_{1})} \cap \bar{h}_{\pi(k_{2})} \cap \bar{h}_{\pi(k_{3})} \right| z_{3} + \dots + \left| \bar{h}_{\pi(1)} \cap \bar{h}_{\pi(2)} \cap \dots \cap \bar{h}_{\pi(i)} \right| z_{i}, \quad i = [2:n],$$

where $\bar{A}_0 = \emptyset$, $\bar{A}_i = \{\bar{h}_{\pi(1)}, \ldots, \bar{h}_{\pi(i)}\}$ is the permuted set of intervals, and z_i is the weight of each term.² Note that \tilde{g}^{AG} in (7b) is not, yet, a true FM as it is not normalized; hence, we include the tilde notation. In [24], $z_i = i/n$; however, one could use any values for z_i , such that $z_2 \leq z_3 \leq \ldots \leq z_n$. Let $\bar{U}_K(\bar{A}_i)$ be the interval defined by (8), then (7b) can be rewritten as

$$\tilde{g}^{AG}(\bar{A}_i) = \sum_{K=2}^{i} \left| \bar{U}_K(\bar{A}_i) \right| z_K, \quad i = [2:n].$$
(9)

Note that $\overline{U}_K(\overline{A}_i)$ is not necessarily a closed (i.e., continuous) interval; hence, the length of $\overline{U}_K(\overline{A}_i)$ is the summed length of all the continuous sub-intervals that comprise $\overline{U}_K(\overline{A}_i)$. This formulation shows that \tilde{g}^{AG} is the weighted summation of the cardinalities (lengths) of (i-1) (perhaps, discontinuous) intervals, where the Kth term, $\overline{U}_K(\overline{A}_i)$, is composed of the union of the intersections of the K-tuples in $\overline{A}_i \subseteq \overline{h}$. The measure of agreement formulation at (9) leads to the following theorem as well as a computationally efficient way of calculating \tilde{g}^{AG} for use in the Choquet integral.

In practice, g^{AG} is normalized by

$$g^{AG}(\bar{A}_i) = \frac{\tilde{g}^{AG}(\bar{A}_i)}{\tilde{g}^{AG}(\bar{h})}, \quad i = [2:n],$$
 (10)

so that $g^{AG}(\bar{h}) = 1$. We now show that g^{AG} is a FM and that it can be computed efficiently.

Theorem 1. The FM of agreement at (7) is monotonic and non-decreasing.

Before we prove this theorem, we require a couple corollaries.

Corollary 1. If a function f is monotonic, and non-decreasing then αf is also monotonic, non-decreasing for $\alpha > 0$.

This corollary shows that we only need to prove that \tilde{g}^{AG} is monotonic and non-decreasing as $g^{AG} = \alpha \tilde{g}^{AG}$, where $\alpha = (\tilde{g}^{AG}(\bar{h}))^{-1}$.

Corollary 2.
$$|\bar{U}_K(\bar{A}_{i-1})| \le |\bar{U}_K(\bar{A}_i)|, K = [i], i = [2:n].$$

Proof: First, by setting K = 1, we see that

$$\bar{U}_1(\bar{A}_i) = \bigcup_{k_1=1}^i \bar{h}_{\pi(k_1)}, \quad i = [n]$$

Thus, we see that $\overline{U}_1(\overline{A}_i) = \overline{U}_1(\overline{A}_{i-1}) \cup \overline{h}_{\pi(i)}, i = [2:n]$. It follows that

$$|U_1(A_{i-1})| \le |U_1(A_i)|,$$

which proves the corollary for K = 1, i = [2 : n].

By inspection, we can write (8) as (11), and by the distributive property, we can further reduce (11) to

$$\bar{U}_K(\bar{A}_i) = \bar{U}_K(\bar{A}_{i-1}) \cup \left(\bar{U}_{K-1}(\bar{A}_{i-1}) \cap \bar{h}_{\pi(i)}\right), \quad (12)$$

which holds true for K = [2:i], i = [2:n]. The proof is complete because (12) implies that

$$|\bar{U}_K(\bar{A}_{i-1})| \le |\bar{U}_K(\bar{A}_i)|, \quad K = [2:i], i = [2:n].$$

We can now prove Theorem 1.

Proof: For set $\bar{A}_{i-1} \subset \bar{A}_i$, Corollary 2 shows that

$$\tilde{g}^{AG}(\bar{A}_{i-1}) = \sum_{K=2}^{i-1} |\bar{U}_K(\bar{A}_{i-1})| z_K$$
$$\leq \sum_{k=2}^{i} |\bar{U}_K(\bar{A}_i)| z_K = \tilde{g}^{AG}(\bar{A}_i),$$

proving the theorem.

The formulation of $g^{AG}(\bar{A}_i)$, coupled with Theorem 1, shows that g^{AG} obeys the properties P1 and P2 of a FM.

Remark 2. Equation (12) gives us a computationally efficient way to compute $\tilde{g}^{AG}(\bar{A}_i)$, i = [2 : n]. The method is as follows:

- 1) $g^{AG}(\bar{A}_1) = 0;$
- 2) For i = [2:n], compute and store $\overline{U}_K(\overline{A}_i)$, K = [2:i], by the recursive formulation at (12);
- 3) Compute $\tilde{g}^{AG}(\bar{A}_i)$ at (9);
- 4) Normalize by (10).

A drawback of the AG measure becomes evident when it is used with the FI. Figure 2 shows the result of the Choquet FI on two sets of intervals, where the FM is calculated by the AG measure. As the figure shows, the only difference between the two sets of intervals is that the end-points are flipped (each end point is made negative). However, the resulting Choquet integrals of \bar{h} and $-\bar{h}$ show that the result is not symmetric; i.e., $C_g(\bar{h}) \neq -C_g(-\bar{h})$ —where we would prefer intuitively that $C_q(\bar{h}) = -C_q(-\bar{h})$. This is because the AG measure of the singletons is zero, i.e., $g^{AG}(\bar{A}_1) = 0$. Hence, if we examine the FI integral equations at (1), we see that source $h(x_{\pi(1)})$ is always ignored. The ignored source is only dependent on the ordering, not on any perceived agreementthere is a disconnect between our modeling of the concept of agreement as a FM and the what happens when we use it with the Choquet FI (or Sugeno FI, for that matter).³

²Note that (7) was written incorrectly in [24]. The last term in the summation was given a weight of z_n , which should be z_i as shown in (7). Furthermore, we add the tilde notation to (7b) because it is later normalized to ensure $g^{AG}(\bar{h}) = 1$.

³Unfortunately, the GenA measure also suffers from this drawback. Later, we will define an additive measure of agreement that solves this problem.

$$\bar{U}_{K}(\bar{A}_{i}) = \bigcup_{k_{1}=1}^{i-K+1} \bigcup_{k_{2}=k_{1}+1}^{i-K+2} \dots \bigcup_{k_{K}=k_{K-1}+1}^{i} \left(\bar{h}_{\pi(k_{1})} \cap \bar{h}_{\pi(k_{2})} \cap \dots \cap \bar{h}_{\pi(k_{K})} \right), \quad K = [i], i = [n], \tag{8}$$

$$\bar{U}_{K}(\bar{A}_{i}) = \bar{U}_{K}(\bar{A}_{i-1}) \cup \left\{ \bigcup_{k_{1}=1}^{((i-1)-(K-1)+1)} \bigcup_{k_{2}=k_{1}+1}^{(i-1)-(K-1)+2)} \cdots \bigcup_{k_{K-1}=k_{K-2}+1}^{i-1} \left(\bar{h}_{\pi(k_{1})} \cap \bar{h}_{\pi(k_{2})} \cap \dots \cap \bar{h}_{\pi(k_{K-1})} \cap \bar{h}_{\pi(i)}\right) \right\}$$
(11)



Fig. 2. Example that shows the non-symmetric result of the FI using the AG measure; i.e., $C_g(\bar{h}) \neq -C_g(-\bar{h})$.



Fig. 3. Two scenarios that each show the union of the intersections of the 2-tuples and 3-tuples for three sources of information.

A second potential drawback of the AG measure is that the cardinality of the interval(s) composed by (8) only captures the overall length and not necessarily the agreement of the intervals. Consider the example shown in Fig. 3. Scenarios A and B each show three sources of information as intervals, sources $\{(1), (2), (3)\}$. We believe that most human observers would say that the sources in Scenario B agree more than the sources in Scenario A. However, if we examine the lengths of the intervals $\overline{U}_2(\overline{A}_3)$ and $\overline{U}_3(\overline{A}_3)$ that would be summed to compute $g^{AG}(\bar{A}_3)$, we clearly see that Scenario A would give a much higher value of g, and thus more importance to the combination of the three sources (recall that $g^{(AG)}(\bar{A}_3) =$ $|\overline{U}_2(\overline{A}_3)|z_2 + |\overline{U}_3(\overline{A}_3)|z_3)$. Furthermore, it can be shown that every combination of sources in Scenario A are given a higher agreement by q^{AG} than those in Scenario B. This drawback leads us to define the measure of generalized accord next.

B. Measure of generalized accord

We begin our derivation of generalized accord with a definition of similarity indices on sets of intervals.

Definition 1. The function $s : \bar{A}_i \to [0, 1]$ is a similarity index on sets of intervals $\bar{A}_i = \{\bar{h}_{\pi(1)}, \bar{h}_{\pi(2)}, \dots, \bar{h}_{\pi(i)}\}$ such that 1) $s(\emptyset) = 0$; (boundary at 0)

- 2) $s(\bar{A}_i) = 1$ iff $\bar{h}_{\pi(1)} = \bar{h}_{\pi(2)} = \ldots = \bar{h}_{\pi(i)}$; (boundary at 1)
- 3) s is symmetric.

Two similarity indices that fit this definition are Jaccard's similarity index and Dice's coefficient, defined respectively as

$$s_J(\bar{h}) = \frac{|\bigcap_{i=1}^{|\bar{h}|} \bar{h}_i|}{|\bigcup_{i=1}^{|\bar{h}|} \bar{h}_i|};$$
(13a)

$$s_D(\bar{h}) = \frac{|\bar{h}||\bigcap_{i=1}^{|\bar{h}|} \bar{h}_i|}{\sum_{i=1}^{|\bar{h}|} |\bar{h}_i|},$$
(13b)

where \bar{h} is some set of interval evidence, $|\bar{h}|$ is the cardinality or number of intervals in \bar{h} , and $|\bar{h}_i|$ is the length of interval \bar{h}_i —e.g., in Eq.(14), the input to s is $\{\bar{h}_{\pi(k_1)}, \bar{h}_{\pi(k_2)}, \ldots, \bar{h}_{\pi(k_K)}\}$. The strength of Dice's coefficient is that the interval union does not need to be calculated, resulting in a less computationally complex similarity index. Furthermore, both Jaccard and Dice's similarity index are dissimilarity (distance) metrics when transformed by (1 - s).

Using s, we propose the GenA measure at (14), where the binomial coefficient term $\binom{n}{K}^{-1}$ accounts for the number of possible K-tuples in \bar{h} . Equation (14) can be written in a similar spirit as (9),

$$g^{GenA}(\bar{A}_i) = \alpha_{\bar{h}} \sum_{K=2}^{i} \mathbf{s}_K(\bar{A}_i), \quad i = [2:n], \quad (15)$$

where

$$\mathbf{s}_{K}(\bar{A}_{i}) = {\binom{n}{K}}^{-1} \sum_{k_{1}=1}^{i-K} \sum_{k_{2}=k_{1}+1}^{i-K+1} \dots$$

$$\sum_{k_{K}=k_{K-1}+1}^{i} s\left(\left\{\bar{h}_{\pi(k_{1})}, \bar{h}_{\pi(k_{2})}, \dots, \bar{h}_{\pi(k_{K})}\right\}\right).$$
(16)

The quantity $\mathbf{s}_K(\bar{A}_i)$ is the sum of the similarities of the *K*-tuples in \bar{A}_i , weighted by an overall value of $\binom{n}{K}^{-1}$. Hence, $\mathbf{s}_K(\bar{h})$ is the average value of the similarities of all *K*-tuples in \bar{h} .

Remark 3. The constant multiple $\alpha_{\bar{h}}$ is computed as

$$\alpha_{\bar{h}} = \left(\sum_{K=2}^{n} \mathbf{s}_{K}(\bar{A}_{n})\right)^{-1},\tag{17}$$

so that $g^{GenA}(\bar{h}) = 1$. If $\bar{h}_1 = \bar{h}_2 = \ldots = \bar{h}_n$ then $\alpha_{\bar{h}} = 1/(n-1)$. This results from the boundary condition of $s(\bar{B}) = 1$, where \bar{B} is any non-empty subset of \bar{h} , and that there are $\binom{n}{K}$ terms in each of the K inner summations in (17); hence, $\mathbf{s}_K(\bar{A}_n) = 1, K = [2:n]$.

$$g^{GenA}(\bar{A}_i) = \begin{cases} 0, & i = 0, 1 \\ \alpha_{\bar{h}} \sum_{K=2}^{i} {n \choose K}^{-1} \left[\sum_{k_1=1}^{i-K} \sum_{k_2=k_1+1}^{i-K+1} \cdots \sum_{k_K=k_{K-1}+1}^{i} s\left(\left\{\bar{h}_{\pi(k_1)}, \bar{h}_{\pi(k_2)}, \dots, \bar{h}_{\pi(k_K)}\right\}\right) \right], & i = [2:n], \end{cases}$$
(14)

Theorem 2. The GenA measure at (14) and (15) is monotonic and non-decreasing, and $g^{GenA}(\bar{h}) = 1$.

Proof: Proving that $g^{GenA}(\bar{h}) = 1$ is trivial. Simply substitute $\alpha_{\bar{h}}$ at (17) into (14) and notice that $\bar{A}_n = \bar{h}$ and, thus, i = n in (14).

To prove that (14) is monotonic, we use the formulation at (15) and rewrite it as (18). Since all quantities in (18) are \geq 0, this shows that $g^{GenA}(\bar{A}_i) \ge g^{GenA}(\bar{A}_{i-1})$. Furthermore, because $\bar{A}_{i-1} \subset \bar{A}_i$, this proves that the GenA measure at (14) is monotonic and non-decreasing.

Remark 4. Theorem 2 shows that the GenA measure is a FM that obeys properties P1 and P2. This is true for any similarity index $s: \overline{A}_i \to [0,1]$ that adheres to Definition 1.

Remark 5. The equation at (18) also illustrates how the GenA measure can be efficiently computed by incrementing through the lattice \bar{A}_i , i = [n]. In practice, one can simply start with $g^{GenA}(\bar{A}_1) = 0$ in the recursive formulation at (18). Also, we can distribute $\alpha_{\bar{h}}$ out of (18) and simply recursively calculate the value $g^{GenA}(\bar{A}_i)/\alpha_{\bar{h}}$ for the entire lattice, saving the calculation of $\alpha_{\bar{h}}$ for the end of the computation (i.e., the normalization step).

Remark 6. The formulation of the GenA measure is generalized such that Theorem 2 applies for similarity indices on other types of inputs, such as FNs (type-1, type-2, etc.), distributions, etc. The only requirement is that the similarity function s be designed for the given input type according to Definition 1. We will discuss this further in Section V.

Using the AG or GenA measure, we can calculate the FI for interval evidence \bar{h} by the following steps:

- 1) Determine the permutation functions π^- and π^+ , such that $h^-_{\pi^-(1)} \ge h^-_{\pi^-(2)} \ge \ldots \ge h^-_{\pi^-(n)}$ and $h^+_{\pi^+(1)} \ge$ $h_{\pi^+(2)}^+ \ge \dots \ge h_{\pi^+(n)}^+;$
- 2) Calculate the portions of the FM lattice, $g(\bar{A}_i^-)$ and $g(\bar{A}_{i}^{+}), \text{ using the AG or GenA measure, where } \bar{A}_{i}^{-} = \{\bar{h}_{\pi^{-}(1)}, \dots, \bar{h}_{\pi^{-}(i)}\}, \ \bar{A}_{i}^{+} = \{\bar{h}_{\pi^{+}(1)}, \dots, \bar{h}_{\pi^{+}(i)}\};$ 3) Compute $\int \bar{h} \circ g = [\int h^{-} \circ g, \int h^{+} \circ g].$

The AG and GenA measures are interesting in their formulation of worth by aggregating (agreements or) similarities of sets of K-tuples of intervals. However, they both suffer from asymmetry between positive- and negative-valued intervals. The next measure, the additive measure of agreement, keeps the same spirit of valuing intervals that agree, but it also alleviates the asymmetry issue.

C. Additive measure of agreement

Consider a similarity function s as in Definition 1. The additive measure of aggreement (AA) is defined as

$$\tilde{g}^{AA}(\bar{A}_i) = \tilde{g}^{AA}(\bar{A}_{i-1}) + \sum_{\substack{j=1\\j\neq i}}^n s^p(\bar{h}_j, \bar{h}_{\pi(i)}), \ i = [n], \ p \ge 0,$$
(19)

where p is a tuning parameter such that p > 1 weights high similarity more heavily and p < 1 weights similarity more evenly. Note that p = 0 produces a FM that weights each piece of evidence equally. Unless otherwise mentioned, we will use a value of p = 1 for this paper. The FM g^{AA} is formed by the normalization step $g^{AA}(\bar{A}_i) = \tilde{g}^{AA}(\bar{A}_i)/\tilde{g}^{AA}(\bar{A}_n)$. The AA measure is clearly an additive measure, i.e., $g^{AA}(\bar{A} \cup \bar{B}) =$ $g^{AA}(\bar{A}) + g^{AA}(\bar{B}), \ \bar{A} \cap \bar{B} = \emptyset$. Hence, the Choquet integral (operated on the left or right interval end-points) reduces to

$$C_{g^{AA}}(h) = \frac{\sum_{i=1}^{n} h_{\pi(i)} \sum_{\substack{j=1\\j\neq i}}^{n} s^{p}(\bar{h}_{i}, \bar{h}_{\pi(i)})}{\sum_{i,j=1}^{n} s^{p}(\bar{h}_{i}, \bar{h}_{j})},$$

where, clearly, this is just a weighted average; each source is weighted by its aggregated similarity with all other sources divided by the total similarity of all pairs of sources-or, in words, how much each source agrees with all other sources.

D. Measure of uniqueness

We propose that the combination of *unique* sources is the antithesis of the notion of agreement or accord. That is, a set of unique sources is that where each source is providing a unique (or different) answer to the question. Hence, sets of sources that do not agree are sets of sources that are unique.

Consider the dissimilarity function $d: \bar{h} \to [0, 1]$, which is related to the similarity function s at Definition 1 by

$$d(h) = 1 - s(h).$$
(20)

With this notion of dissimilarity, we can directly substitute $d(\bar{h})$ into the GenA or AA measures, directly producing uniqueness measures. We will denote these as $q^{UQ_{GenA}}$ and $q^{UQ_{AA}}$.

The FM of uniqueness based on the GenA measure for interval inputs \bar{h} is

$$g^{UQ_{GenA}}(\emptyset) = g^{UQ_{GenA}}(\bar{A}_1) = 0, \qquad (21a)$$

$$g^{UQ_{GenA}}(\bar{A}_i) = \beta_{\bar{h}} \sum_{K=2}^{\circ} \mathbf{d}_K(\bar{A}_i),$$
(21b)

$$\mathbf{d}_{K}(\bar{A}_{i}) = {\binom{n}{K}}^{-1} \sum_{k_{1}=1}^{i-K} \sum_{k_{2}=k_{1}+1}^{i-K+1} \dots$$
(21c)

$$\sum_{k_K=k_{K-1}+1}^{\circ} d\left(\left\{\bar{h}_{\pi(k_1)}, \bar{h}_{\pi(k_2)}, \dots, \bar{h}_{\pi(k_K)}\right\}\right).$$

$$g^{GenA}(\bar{A}_{i}) = \alpha_{\bar{h}} \left[\mathbf{s}_{i}(\bar{A}_{i}) + \sum_{K=2}^{i-1} {\binom{n}{K}}^{-1} \sum_{k_{1}=1}^{(i-1-K)} \sum_{k_{2}=k_{1}+1}^{(i-1-K)+1} \dots \sum_{k_{K}=k_{K-1}+1}^{i-1} s\left(\left\{ \bar{h}_{\pi(k_{1})}, \bar{h}_{\pi(k_{2})}, \dots, \bar{h}_{\pi(k_{K})}, \bar{h}_{\pi(i)} \right\} \right) \right] + g^{GenA}(\bar{A}_{i-1}), \quad i = [2:n]$$

$$(18)$$

where $\mathbf{d}_K(\bar{A}_i)$ at (21c) is formed analogous to $\mathbf{s}_K(\bar{A}_i)$ at (16) and $\beta_{\bar{h}}$ is a normalization factor. We now show that we do not need to calculate $g^{UQ_{GenA}}$ by directly summing the $\mathbf{d}_K(\bar{A}_i)$ terms. Instead, we formulate $g^{UQ_{GenA}}$ directly from g^{GenA} , as the following proposition and theorem show.

Proposition 1. The aggregated dissimilarities at (21c) can be computed as

$$\mathbf{d}_K(\bar{A}_i) = 1 - \mathbf{s}_K(\bar{A}_i). \tag{22}$$

Proof: Substituting (20) into (21c) and collecting terms proves the proposition.

Theorem 3. The FM of uniqueness can be calculated in terms of the FM of generalized accord as

$$g^{UQ_{GenA}}(\bar{A}_i) = \frac{(i-1)\alpha_{\bar{h}} - g^{GenA}(\bar{A}_i)}{(n-1)\alpha_{\bar{h}} - 1}, \quad i = [2:n],$$
(23)

where $\alpha_{\bar{h}}$ is the normalization factor given at (17).

Proof: First, we substitute (22) into (21b) giving

$$g^{UQ_{GenA}}(\bar{A}_i) = \beta_{\bar{h}} \sum_{K=2}^{i} (1 - \mathbf{s}_K(\bar{A}_i)),$$
 (24a)

$$=\beta_{\bar{h}}\left(i-1-\sum_{K=2}^{i}\mathbf{s}_{K}(\bar{A}_{i})\right).$$
 (24b)

Using the FM property that $g(\bar{A}_n) = 1$, the normalizing coefficient $\beta_{\bar{h}}$ is

$$\beta_{\bar{h}} = \left(n - 1 - \sum_{K=2}^{n} \mathbf{s}_{K}(\bar{A}_{n})\right)^{-1} = \frac{\alpha_{\bar{h}}}{(n-1)\alpha_{\bar{h}} - 1}, \quad (25)$$

wherein we have substituted the formulation for $\alpha_{\bar{h}}$ at (17). Finally, we substitute (25) into (24b), resulting in the formulation for $g^{UQ_{GenA}}$ at (23).

Remark 7. The formulation for $g^{UQ_{GenA}}$ at (23) is fast to compute as one can use the recursive calculation of the GenA measure and directly follow with the calculation of $g^{UQ_{GenA}}$. Also, this shows the inverse relationship of uniqueness and accord, which is intuitively pleasing.

Remark 8. If the intervals in inputs \bar{h} are all completely disjoint then $\alpha_{\bar{h}} = 0$ and $g^{GenA}(\bar{A}_i) = i/n$, i = [n]. Hence, (23) shows that $g^{UQ}(\bar{A}_i) = g^{GenA}(\bar{A}_i) = i/n$ for completely disjoint inputs. In the spirit of uniqueness this is intuitively pleasing; when all sources are unique, then all sources should be weighted equally in the FI. This is also the case when $\bar{h}_1 = \bar{h}_2 = \ldots = \bar{h}_n$, or when no source is unique.

The UQ measure based on the AA measure is created by

substituting the dissimilarity function d into (19) producing

$$\tilde{g}^{UQ_{AA}}(\bar{A}_i) = \tilde{g}^{UQ_{AA}}(\bar{A}_{i-1}) + n - 1 - \sum_{\substack{j=1\\j\neq i}}^n s(\bar{h}_j, \bar{h}_{\pi(i)}),$$

which can by shown to be equivalent to

$$\begin{split} \tilde{g}^{UQ_{AA}}(\bar{A}_i) = & g^{UQ_{AA}}(\bar{A}_{i-1}) + \sum_{\substack{j=1\\j\neq i}}^n d(\bar{h}_j, \bar{h}_{\pi(i)}), \\ = & g^{UQ_{AA}}(\bar{A}_{i-1}) + \sum_{\substack{j=1\\j\neq i}}^n \left(1 - s(\bar{h}_j, \bar{h}_{\pi(i)})\right), \\ = & \tilde{g}^{UQ_{AA}}(\bar{A}_{i-1}) + n - 1 \\ & - \left(\tilde{g}^{AA}(\bar{A}_i) - \tilde{g}^{AA}(\bar{A}_{i-1})\right). \end{split}$$

Like the GenA-based UQ measure, this shows that the AAbased UQ measure is also easy to compute once you have computed the AA measure. Furthermore, it is also an additive measure, and when coupled with the Choquet integral produces a weighted average where the *i*th source is weighted by $n - 1 - (\tilde{g}^{AA}(\bar{A}_i) - \tilde{g}^{AA}(\bar{A}_{i-1}))$.

Remark 9. Consider a set of sources $\bar{h} = \{\bar{h}_{11}, \ldots, \bar{h}_{1n_1}, \bar{h}_{21}, \ldots, \bar{h}_{2n_2}, \ldots\}$ where there are multiple copies of several sources, i.e., there are n_i sources that are equal, $\bar{h}_{i1} = \bar{h}_{i2} = \ldots = \bar{h}_{in_i}$, for i = [n], where n is some number of unique sources. It can be shown that in this scenario that the UQ_{AA} measure with the Choquet integral simplifies to

$$C_g(\bar{h}) = \left[\frac{1}{n}\sum_{i=1}^n h_{i1}^-, \frac{1}{n}\sum_{i=1}^n h_{i1}^+\right]$$

or to the average of the unique sources (in this equation, the first instance of each group), where each unique source \bar{h}_{i1} is only considered once in the average. In essence, the UQ_{AA} measure with the Choquet integral is only aggregating one instance from each matching group of sources. Hence, each unique group of evidence is considered equally, no matter how many sources fall into each group.

We now turn to our last proposed FM for sets of intervalbased evidence.

E. Measure of specificity

The FM of specificity, unlike the agreement and accord measures, is defined on the singletons or densities of the FM. First, the relative length as compared to the smallest intervalvalued source is calculated,

$$\delta(\bar{h}_i) = \frac{\min_{\bar{h}_j \in \bar{h}} \{h_j^+ - h_j^-\}}{h_i^+ - h_i^-}.$$
 (26a)



(a) Interval-valued sources with Choquet FI results for data-derived FMs



(b) Lattice of FM values

Fig. 4. Comparison of FMs and Choquet integral results for four intervalvalued inputs.

Then the densities are computed as the normalized values of these lengths,

$$g(\bar{h}_i) = \frac{\delta(h_i)}{\sum_{j=1}^n \delta(\bar{h}_j)}.$$
(26b)

Clearly, the sum of the densities are equal to 1; hence, the specificity measure is calculated as the probability measure

$$g^{SP}(\bar{A}_i) = \sum_{k=1}^{i} g(\bar{h}_{\pi(k)}).$$
 (26c)

The specificity measure weights each interval source inversely proportional to the source's length. Thus, relatively certain (or small length) intervals are weighted higher than uncertain (or long length) intervals.

Example 1. Figure 4 shows an example of using the proposed data-derived FMs for a set of 4 interval-valued sources. The values of these sources are $\bar{h}_1 = [0, 2]$, $\bar{h}_2 = [3, 4.1]$, $\bar{h}_3 = [4, 4.1]$, and $\bar{h}_4 = [4, 4.1]$. The Jaccard similarity was used for the GenA and AA measures. The result of the fuzzy Choquet integral using the proposed data-derived FMs is shown along with the illustration of the 4 sources in view (a), while the FM lattice is shown in view (b) of Fig. 4.

Comparing the AG, GenA, and AA results shows that each of these FMs places more weight on sets of evidence that include sources 2, 3, and 4; however, both GenA and AA clearly give higher weight to the combination of sources $\{3 \text{ and } 4\}$, which according to those measures agree more than $\{2 \text{ and } 3\}$ or $\{2 \text{ and } 4\}$; this is intuitively pleasing. The AG measure actually weights the combinations of $\{3 \text{ and } 4\}$, $\{2 \text{ and }$

and 3}, and {2 and 4} as all having equal agreement with a FM value of 0.4; this is because the length of the interval of the intersection of those three combinations is the same. Unlike the AG and GenA measures, the AA measure also weights the singletons and shows that sources 3 and 4 have the highest weight of 0.46, compared to 0.08 of source 2 (which agrees slightly with 3 and 4) and 0 of source 1. In essence, the AA measure represents the agreement of sources all throughout the lattice of the measure, including at the singleton level. Lastly, as the Choquet integral results show, the GenA and AA aggregation results are biased more towards sources 3 and 4 than the AG measure (which considers 2, 3, and 4 to have equal weight).

The UQ measures both show that the Choquet integral result is now including the unique evidence, i.e., source 1, in its aggregation, with the GenA-based UQ result being slightly more biased towards source 1 than the AA-based UQ result. Finally, the SP measure, which weights sources according to their specificity, is weighted very much toward sources 3 and 4, which are the most specific. This toy example clearly illustrates the behavior of these measures. We will investigate more results, including those with real data, in Section VI. We now extend the proposed data-informed FMs to FN-valued evidence.

V. DATA-INFORMED FUZZY MEASURES ON FUZZY NUMBERS

Consider a set of FN-valued inputs $\hat{H} = \{\hat{H}_1, \dots, \hat{H}_n\},$ $\hat{H}_i \in FN(\mathbb{R})$. We now extend the measures proposed in Section IV to FN-valued inputs \hat{H} .

A. Measure of agreement

The notion of agreement for interval-valued sources was that the length of the intersection of the sources is proportional to the level of agreement of the sources (more completely, agreement is the weighted sum of the lengths of unions of intersections of k-tuples in the set of intervals). For FNvalued sources, we propose that agreement is captured by a function r that represents some property about the union of the intersections of K-tuples in \hat{A}_i (namely, the FS $U_K(\hat{A}_i)$). The function $r: U \to \mathbb{R}$ takes a FS as an input and returns a real value that is proportional to the height or size of U.

Hence, we define the measure of agreement g^{AG} for the subset of FN-valued sources \hat{A}_i as

$$g^{AG}(\hat{A}_0) = g^{AG}(\hat{A}_1) = 0, \qquad (27a)$$

$$g^{AG}(\hat{A}_i) = \sum_{K=2}^{n} r(U_K(\hat{A}_i)) z_K, \quad i = [2:n],$$
 (27b)

where $U_K(\hat{A}_i)$ is defined at (28) and r is a function on the FS U_K . Note that \bar{U}_K for interval sources \bar{A}_i is a (perhaps, discontinuous) interval. However $U_K(\hat{A}_i)$ is a FS, but not necessarily a FN; hence, we do not use the $\hat{}$ notation on U_K in (28). The measure g^{AG} is then normalized to the FM \tilde{g}^{AG} , as in (10), such that $\tilde{g}^{AG}(\hat{H}) = 1$.

$$U_{K}(\hat{A}_{i}) = \bigcup_{k_{1}=1}^{i-K+1} \bigcup_{k_{2}=k_{1}+1}^{i-K+2} \dots \bigcup_{k_{K}=k_{K-1}+1}^{i} \left(\hat{H}_{\pi(k_{1})} \cap \hat{H}_{\pi(k_{2})} \cap \dots \cap \hat{H}_{\pi(k_{K})} \right), \quad K = [i], i = [n],$$
(28)

We propose two functions for r that accomplish the semantic definition of agreement. First, in the spirit of the interval length calculation in g_{AG} , we propose

$$r_{CARD}(U) = |U| = \int_{0}^{U} U(y) dy, \qquad (29a)$$

where U(y) is the membership of U at y and ⁰U is the support of U (or the level-cut at $\alpha = 0$). The drawback of r_{CARD} is that this calculation can be computationally expensive, both in coding and calculation (however, we present a computationally efficient solution in Section V-F). Hence, we also propose a simpler method based on the maximum of U,

$$r_{MAX}(U) = \max_{0} \{ U(y) \}.$$
 (29b)

In the spirit of FS similarity, the quantity $1 - r_{MAX}$ is analogous to the "degree of separation" proposed in [39]. The drawback of r_{MAX} is that it does not consider the overall shape of U. Figure 5 shows an example of this calculation for two different scenarios of the combination of three FNs. The example was designed to show that while r_{MAX} is simpler to compute, it may not capture the overall agreement between two FNs. In this example, r_{MAX} is equal in both scenarios. But, clearly, the sets in Scenario 1 "agree more," which is captured by r_{CARD} . This issue is analogous to the problem of height versus centroid defuzzifiers in fuzzy control problems r_{CARD} is accurate, while r_{MAX} is efficient. In Section V-F, we propose an efficient approximation to r_{CARD} ; hence, we recommend this function and use r_{CARD} exclusively in our results.

B. Measure of generalized accord

The strength of the GenA measure is its generalization as an aggregation of similarity functions; hence, the only part of GenA that needs to be adapted to accommodate any type of input is the similarity function s. There have been many similarity functions proposed for fuzzy sets. A good comprehensive review on this topic can be found in [40]. However, most of the existing similarity functions are for measuring the similarity of only two input sets and are thus inappropriate for use with the proposed FMs, which need to consider combinations of two or more. We propose the following adaptations of set-theoretic similarity measures, generalized for multiple FS inputs,

$$s_J(\hat{H}) = \frac{r\left(\bigcap_{i=1}^{|\hat{H}|} \hat{H}_i\right)}{r\left(\bigcup_{i=1}^{|\hat{H}|} \hat{H}_i\right)};$$
(30a)

$$s_D(\hat{H}) = \frac{|\hat{H}| \cdot r\left(\bigcap_{i=1}^{|\hat{H}|} \hat{H}_i\right)}{\sum_{i=1}^{|\hat{H}|} r\left(\hat{H}_i\right)};$$
(30b)

where r is, again, a function on a FS returning a measure of its size or height, $|\hat{H}|$ is the cardinality or number of FNs



Fig. 5. Two scenarios showing the resulting values of $r_{CARD}(U_2(A_3))$ and $r_{MAX}(U_2(A_3))$. Clearly, r_{MAX} is equal for each scenario, while r_{CARD} (Scenario 1) > r_{CARD} (Scenario 2).

in \hat{H} , and \hat{H} is some input set of FN evidence. Note that if r is r_{CARD} at (29a), then (30a) is analogous to Dubois and Prade's dissimilarity measure [18] and (30b) is equivalent to an extended form of the Z-similarity proposed in [41].

The GenA FM for FN-valued inputs can be directly calculated by substituting a FN similarity function, such as those at (30), into the formulation at (14), where the input to the GenA measure is now the set of FNs \hat{A}_i . Furthermore, the same recursive method proposed for the interval-valued GenA measure can be used for FN-valued inputs. As stated in Remark 6, Theorem 2 also applies to the GenA measure on FNs; viz., the GenA measure on FNs is monotonic, nondecreasing, and $g^{GenA}(\hat{H}) = 1$.

C. Additive measure of agreement

Much like the extension of the GenA measure to FN-valued inputs, the AA measure can be extended by using a FN-appropriate similarity function, such as those at (30), and substituting this directly into the AA measure definition at (19).

D. Measure of specificity

The SP measure for FN-valued inputs is calculated in the same spirit as the SP measure for intervals: the more specific the FN evidence, the higher the weight of that evidence. Hence, the SP measure for FN input \hat{H} starts with the

calculation of the inverse relative cardinality of each set in the evidence,

$$\delta(\hat{H}_i) = \frac{\min_{\hat{H}_j \in \hat{H}} \{ |H_j| \}}{|\hat{H}_i|},$$
(31a)

where $|\cdot|$ indicates cardinality as in (29a). Then, the densities and measure are calculated similarly as for the interval-valued inputs,

$$g(\hat{H}_i) = \frac{\delta(\hat{H}_i)}{\sum_{j=1}^n \delta(\hat{H}_j)},$$
(31b)

$$g(\hat{A}_i) = \sum_{k=1}^{i} g(\hat{H}_{\pi(k)}), \quad i = [n].$$
(31c)

Remark 10. If one is using only triangular FNs, then cardinality is directly proportional to the support of the FN; hence, (31a) can be written as

$$\delta(\stackrel{\triangle}{H}) = \frac{\min_{\hat{H}_j \in \hat{H}} \{|^0 \stackrel{\triangle}{H}_j|\}}{|^0 \stackrel{\triangle}{H}|},\tag{32}$$

where the cardinality is now just over the interval ${}^{0}\overset{\triangle}{H}$ at the support of \hat{H} , a distinct advantage of using triangular FNs. **Remark 11.** If any \hat{H}_i is a singleton FN, i.e., it has a

Remark II. If any H_i is a singleton FN, i.e., it has a membership function

$$\hat{H}_i(y) = \begin{cases} 1 & y = a, \\ 0 & \text{else,} \end{cases}$$
(33)

where a is a number, then $\delta(\hat{H}_i)$ is undefined. For this case, we set the value of $\delta(\hat{H}_i) = 1$ for all singleton evidence, and 0 else. Thus, the Choquet FI reduces to the average of the singleton evidence sets, ignoring all "uncertain" evidence (i.e., those that are not singletons). Examples of this will be shown in Section VI-B2.

E. Measure of uniqueness

Again, the beauty of the GenA and AA FMs is that they are defined over a generalized similarity function; hence, they are appropriate for any type of input for which a similarity function can be appropriately designed. Because the FM of uniqueness is simply defined by substituting the dissimilarity function (1 - s) into the formulations of GenA or AA, the FM of uniqueness for FN-valued inputs is simply calculated the same as with the interval-valued inputs. Hence, GenA and AA, and their UQ antitheses, can be easily generalized to any type of evidence, assuming an appropriate similarity-function can be proposed.

F. Computational solution

Computing the similarity function $r_{CARD}(U)$ can be computationally expensive, especially if it is exactly calculated as this involves a complicated geometrical solution. But an efficient computational solution can be proposed in the spirit of centroid defuzzification. However, by discretizing the support of the overlap area U, we can easily and efficiently calculate both $r_{CARD}(U)$ and $r_{MAX}(U)$. Consider the discretized support of U, i.e., $(y_1, \ldots, y_{n_s})^T$, $\mathbf{x}_i \in {}^0U$. Then if we assume that $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_{n_s})^T$ is the sampling of the fuzzy set U, where $\mathbf{u}_i = U(y_i)$, then the cardinality and max in (29a) and (29b), respectively, can be efficiently and accurately approximated as

$$|U| \approx \sum_{x_i \in {}^0 U} \mathbf{u}_i \Delta; \tag{34a}$$

$$\max_{x \in {}^{0}U} \{U(x)\} \approx \max_{x_i \in {}^{0}U} \mathbf{u}_i;$$
(34b)

where Δ is the sampling spacing, i.e., $\Delta = x_2 - x_1$. This can interpreted as the box-car integration of U at sampling spacing Δ . Note that in practice, one does not need to multiply by the sampling spacing Δ as it is unimportant in the calculations of the FMs.

First, we discretize the support of \hat{H} . For the examples shown in this paper, we discretized the support at $n_s = 100$ equally spaced points. Assume that $\hat{\mathbf{H}}_j = (\mathbf{h}_{j1}, \dots, \mathbf{h}_{jn_s})$ is the discretized version of the FN $\hat{H}_j \in \hat{H}$. Hence, the intersections and unions in (30) can be calculated as

$$\bigcap_{i=1}^{n} \hat{H}_{i} \approx \left(\min_{i=1}^{n} \mathbf{h}_{i1}, \dots, \min_{i=1}^{n} \mathbf{h}_{in_{s}}\right);$$
(35a)

$$\bigcup_{i=1}^{n} \hat{H}_{i} \approx \left(\max_{i=1}^{n} \mathbf{h}_{i1}, \dots, \max_{i=1}^{n} \mathbf{h}_{in_{s}} \right);$$
(35b)

where it is clear that we are using Zadeh's definition of intersection and union.

In practice, we accomplish the computations of (30) by first discretizing \hat{H} in n_s equally spaced points over the support of the union of all sets in \hat{H} . We then use (35) and (34) to efficiently calculate the appropriate contributions to the AG, GenA, and AA measures (and their UQ antitheses). The SP measure (for non-triangular FNs) is calculated in the same spirit by using the discretized cardinality at (34a) to compute $\delta(\hat{H}_i)$ at (31a).

VI. RESULTS

We now present several experiments that show the behavior of each measure when used with the FI. First, we will show results of our data-informed FMs for interval-valued data. Next, we demonstrate our FMs for FN-valued data on both synthetic and real data. We stress that in these examples, there are no "right" answers, each of the FMs produce different results by the way in which they weight different combinations of sources. What we aim to do instead with these demonstrations is provide a good look at how these FMs work and how they produce the weight values when used to aggregate interval and FN sources with the FI.

A. Interval data

In this experiment, we created four sets of synthetic interval data. These data are shown as the blue bars in Fig. 6. We then used each of the data-driven FMs, described in Section IV, to compute the Choquet integral aggregation. The results are shown as the black bars in Fig. 6. Scenario 1, in view (a), displays a case where there are three smaller (more certain) intervals-sources 4, 5, and 6-that agree very well and then three larger (less certain) intervals-sources 1, 2, and 3-that agree "somewhat." The results of the FI show that the AG measure clearly prefers the three larger intervals, sources 1-3, in the aggregation, while the AA and GenA measure put more weight on the smaller intervals because they agree more relative to their size. The UQ results show that the "unique" intervals, sources 1 and 2, have more weight in the result. The SP (specificity) measure places the most weight on the smaller, more certain, sources (4-6) and clearly shows its preference for these more certain sources. Finally, the equallyweighted average is displayed for comparison. The last row of the table is the result of a survey of 40 electrical and computer engineering college students who were asked to draw the interval that best represents the combination of all of the sources. The max and min values for the interval end-points are shown by the red bow-tie plots. The black bar indicates the mean of their results. While this is not a good comparison for theoretical work, such as this, it does show that the maximum end point is placed at the top end of sources 4-6, while the minimum end point is biased towards the middle of the scale. Hence, the humans also show some preference in moving the aggregated result toward the more certain, and "agreeing," sources 4-6.

Views (b–d) in Figure 6 display more synthetic examples of the data-derived FMs used with the Choquet integral. Interestingly, in Scenario 2 in view (b), the humans choose a wide interval result, while all the data-derived FMs produce very narrow results that look much like the inputs (at least in certainty level). However, in the scenarios shown in views (c,d), the data-derived measures agree well with the humandetermined aggregation.

B. Fuzzy number data

The Jaccard similarity measure at (30a) and r_{CARD} at (29a) were used in all these results. Triangular FNs are used throughout this section for computational ease and are denoted by (y_l, y_c, y_r) , i.e.,

$$\overset{\triangle}{H}_{i}(y) = \begin{cases} 0, & y \leq y_{l}, \ y > y_{r}, \\ (y - y_{l})/(y_{c} - y_{l}), & y_{l} < y \leq y_{c}, \\ (y_{r} - y)/(y_{r} - y_{c}), & y_{c} < y \leq y_{r}. \end{cases}$$

1) Synthetic data examples: Figure 7 shows four examples of using the Choquet integral to aggregate FNs with the FMs described in Section V. Table II contains the parameters of the triangular FNs used in these examples. The membership of each FM with respect to the AVG, MIN, and MAX aggregation operators is also shown in each view of Fig. 7.

Example 1 in view (a) has three narrow (more certain) inputs with support between 4 and 5 and two more uncertain inputs with support between 0 and 4. The AG-based Choquet integral shows that the result tends towards the more uncertain inputs that seemingly agree more, according to the AG measure. However, the GenA and AA results show that they put more weight on the 3 sources in the right side of the plot. These

	$ \overset{\bigtriangleup}{H} = (y_l, y_c, y_r) $
Example 1	$\{(0, 1.8, 3), (1, 2.2, 4), (4, 4.5, 5), (4.1, 4.4, 4.5), \}$
	(4.2, 4.6, 4.6)
Example 2	$\{(0,4,5), (0,1,2), (0,1.1,3), (4,4.1,5)\}$
Example 3	$\{(0, 0.5, 1), (1, 4, 5), (1, 4, 5)\}$
Example 4	$\{(0, 0.5, 1), (1, 4, 5), (1, 4, 5), (1, 4, 5), (1, 4, 5)\}$

sources agree more with respect to the Jaccard similarity measure. The UQ measures of GenA and AA show that they put more emphasis on the sources that are not as in agreement (with respect to the Jaccard). Finally, the SP measure clearly prefers the more certain sources at the right of the plot, which is expected.

Example 2 in view (b) presents a synthetic data experiment that further explores the notion of (relative) agreement between sources. There are two sources on the left side of the plot: the source with support between 0 and 2 (say Source 1) and the source between 0 and 3 (say Source 2). Then, there is a source with support between 0 and 5 (say Source 3). Finally, there is a source at the right of the plot with a support between 4 and 5. Clearly, Sources 1 and 2 agree the most with each and they also agree slightly with the uncertain source, Source 3. Source 4 has no overlap with Sources 1 and 2, but is completely contained in Source 5. The AG, GenA, and AA measures clearly put more weight on Sources 1 and 2. The SP measure prefers Source 4 as it is the most certain, but also puts some weight on Source 1 and 2. The comparison between the UQ measures based on GenA and AA is interesting. The UQ-GenA measure seems to put more emphasis on Source 4 (as it the output moves to the right). However, the UQ-AA measure moves much more to the right than the UQ-GenA measure. This is because it weights Source 4 heavily, as it is the source that only agrees with one other FN.

Examples 3 and 4 explore how many sources that completely agree are combined with one source that is completely disjoint from the others. View (c) shows the aggregation of one unique source and two exactly equal sources (those with support between 1 and 5). View (d) shows the aggregation of the same unique source with four exactly equal sources. Clearly, the AG, GenA, and AA measures recognize this situation and return an output that is exactly equal to the sources that agree for both Examples 3 and 4. However, the UQ measures show interesting behavior. The UQ-AA measure in both examples returns an equally weighted average between the unique source and one copy of the agreeing sources. But the UQ-GenA measure is confused by the many copies of the agreeing sources. View (d) shows that the UQ-GenA source tends toward the agreeing sources as the number of them increase. This is counterintuitive to the notion of uniqueness. Hence, we believe that the UQ-AA measure is much more useful as it basically ignores the presence of multiple copies of agreeing sources and instead considers them to be one unique source.

2) Real data: For this demonstration, we use environmental management data gathered by the Government of Western



Fig. 6. Results of fuzzy Choquet integral using proposed data-informed FMs. Human result is of a survey on 27 people-mean shown as thick black bar, max and min values shown in red bow tie plot.

Australia Departments of Parks and Wildlife. Eight stakeholder representatives responsible for making conservation decisions were asked to rank nine categories in terms of their value in making these decisions-the categories are *adequate* resources, aesthetics, health (physical environment), health (protection from other organisms), knowledge and education, meaningful occupation, philisophical/spiritual, recreation, and future options. Once they had ranked the categories, they were then asked to add a "minimum/maximum" bound in terms of how low/high they would go for the rank of each category (essentially adding an uncertainty to their original ranking). These were converted to triangular FNs as (min, rank, max). The eight responses for two categories-resources and meaningful occupation-are shown in Table III. As the table shows, some stakeholders provided very certain responses, indicated by singleton FNs. However, many of the stakeholders expressed uncertainty in their original rankings, giving us responses with which we could build triangular FNs.

The stakeholder responses are anonymized; hence, no information is known about the relative merit of each person's rankings. This is a perfect scenario for where our data-

TABLE III FN VALUE RANKINGS GIVEN BY CONSERVATION STAKEHOLDERS

Category	FN Value Rankings, $\stackrel{\triangle}{H} = (y_l, y_c, y_r)$
resources	$ \begin{array}{c} (1,1,1), (4,5,5), (6,6,8), (5,5,5), \\ (7,7,7), (7,9,9), (1,1,3), (1,3,3) \end{array} $
meaningful occupation	$\begin{array}{c} (8,8,8), (8,9,9), (7,7,8), (7,9,9), \\ (1,1,1), (4,5,5), (5,6,6), (6,7,8) \end{array}$

informed FMs can be used to assign weight to combinations of sources in order to aggregate the rankings into a composite rank for each category. Figure 8 shows the results of using the Choquet FI to aggregate the stakeholder responses with respect to the data-informed FMs. View (a) shows the aggregation for the *resources* category and view (b), the *meaningful occupation* category. The FN inputs in view (a) span the entire spectrum of the rankings; however, two of the responders agree fairly well that the category is highly ranked ('near 1'); hence, the AG, GenA, and AA FMs (all of which place importance on agreement of some type) give a composite result of a ranking 'around 3.' The UQ FMs cause an aggregation of the individual unique responses and results



Fig. 7. Synthetic data examples of aggregating FNs using the Choquet integral with data-derived FMs. Tables show the membership of each FM to each in the three base aggregation operators.

in ranking of 'around 4.' In this real-world example, the UQ aggregation could be thought of as a tool to remove the bias caused by cliques in the data, i.e., groups of people that are answering similarly because of surrounding peers. As Remark 11 states, if singleton inputs exist, then the SP FM prefers them exclusively over the uncertain inputs; hence, the SP FM for the *resources* category produces an average of the three singleton responses—i.e., 1, 5, and, 7—giving a composite rank of 4.33.

Figure 8(b) shows the Choquet FI aggregation of the *mean-ingful occupation* FN ranks. As we can see in the FN inputs, most stakeholders rank this category low (between 5 and 9) and one response is at 1. Hence, we would intuit that the composite aggregation would be a rank of 'around 7.' Indeed, the AG, GenA, and AA FMs produce results that matches our intuition. The UQ FMs reduce the effect of the "agreement" at the low ranks; thus, the lone responder at 1 has more of an effect on the composite ranking, as shown by the UQ results of 'around 6.' Last, the SP FM is again dominated by the presence of the singleton inputs, 1 and 8, producing a crisp composite output of 4.5.

VII. CONCLUSIONS

The focus of this paper was on the aggregation of intervaland fuzzy number-valued evidence using the fuzzy integral with respect to a fuzzy measure. The challenge we addressed is that of the formation of the fuzzy measure. Specifying a fuzzy measure is a difficult task for some data sets, especially those where this is little to no information about the individual sources, such as crowd-sourced data, battlefield situational awareness data, or anonymized survey data. Hence, we proposed six fuzzy measures that compute the lattice of the fuzzy measure according to functions on the input data itself. Table IV provides an overview of the proposed data-informed measures. The AG, GenA, and AA fuzzy measures place value on combinations of sources that agree with one another, while the UQ fuzzy measures provide an antithesis to the GenA and AA measures which allow one to only consider the uniquelyvalued sources (removing the synergy of redundant sources). Lastly, the SP measure places importance on inputs that are certain or specific in their response. Theoretical discussions on these data-informed measures presented evidence against the use of the AG and GenA measure, resulting in the proposal of the AA measure as a solution to these problems.

Next, we extended the notion of these measures to the case of fuzzy number-valued inputs and presented a computational approach for efficiently computing the fuzzy integral with respect to the data-informed measures. Two methods were presented for representing agreement in combinations of fuzzy number-valued sources—we recommend the cardinality-based method, i.e., r_{CARD} at (29a), which is grounded in the work of Dubois and Prade's dissimilarity measure [18] and Z-



Fig. 8. Aggregation of stakeholders' opinions on the value ranking of each category in making conservation decisions. FN rankings are aggregated using the Choquet integral with data-derived FMs. Tables show the membership of each FM to each in the three base aggregation operators.

TABLE IV OVERVIEW OF DATA-INFORMED FUZZY MEASURES

EM	D	C
FM	Description	Comments
AG	Provides higher weight on sets that have a larger intersection.	Does not normalize the intersection of the combination by the cardinality of the individuals; hence, more-uncertain sets could "agree" less than more-certain sets but have a larger overall weight in the aggregation. Suffers from asymmetry when negative-valued integrands are considered.
GenA	Provides higher weight on sets that have a larger similarity value; i.e., size of the intersection with respect to their individual cardi- nalities.	Suffers from asymmetry when negative-valued integrands are con- sidered.
AA	Provides higher weight on sets that have a larger similarity value.	Fixes the asymmetry issue of AG and GenA.
UQ	Provides equal weight to each unique set in the collection of sources; i.e., it removes the effect that redundant sources are represented more heavily in the aggregation.	There is a UQ measure for both the AA and GenA measures.
SP	Provides a weight inversely proportional to the cardinality of each	Sets that are more certain are represented more heavily in the
	input set.	aggregation.

similarity [41].

Demonstrations of the data-informed fuzzy measures provided an important view of how these measures cause different behaviors in fuzzy integral aggregation, with several synthetic examples shown as to reinforce the theoretical discussions earlier in the article. Finally, a real-world data example was shown which aggregated the value ranking of stakeholders' opinions on how different categories affected their decisions. Two specific categories were shown to showcase the datainformed fuzzy measures with real data.

In the future we will present a more comprehensive article on the results of the stakeholder opinion study, using our data-informed fuzzy measures to produce composite ranks of each of the nine available categories. We will also extend our analysis to larger data sets, such as internet crowdsourced responses and social networking, as well as consider other aggregation methods beyond the fuzzy integral that can leverage the proposed data-informed fuzzy measures. In summary, the proposed data-informed fuzzy measures pave the way for an expanded use of the fuzzy integral to aggregate sources of information for which no contextual knowledge is known about the contributing sources. This paper has laid the theoretical foundations and provided initial insights into the behavior and properties of data-informed fuzzy measures, which can be employed in combination with the fuzzy integral. Based on the these insights, we have proposed guidelines on

the applicability of the given fuzzy measures and in the future are looking forward to further explore their potential in real world applications.

ACKNOWLEDGEMENTS

We would like the thank the Government of Western Australia Department of Parks and Wildlife and, in particular, Ken Wallace and Michael Smith for providing us with the environmental management survey data.

REFERENCES

- M. Anderson, D. T. Anderson, and D. Wescott, "Estimation of adult skeletal age-at-death using the Sugeno fuzzy integral," *American Journal* of *Physical Anthropology*, vol. 142, no. 1, pp. 30–41, 2009.
- [2] D. T. Anderson, J. M. Keller, and T. C. Havens, "Learning fuzzy-valued fuzzy measures for the fuzzy-valued Sugeno fuzzy integral," in *Proc. IPMU*, Dortmund, Germany, 2010.
- [3] M. Grabisch, T. Murofushi, and M. Sugeno, Eds., Fuzzy Measures and Integrals: Theory and Applications. New York: Physica-Verlag, 2000.
- [4] M. Grabisch, Fuzzy Measures and Integrals: Theory and Applications. New York: Physica-Verlag, 2000, ch. Fuzzy integral for classification and feature extraction, pp. 415–434.
- [5] J. M. Keller, P. Gader, and A. K. Hocaoglu, *Fuzzy Measures and Integrals: Theory and Applications*. New York: Physica-Verlag, 2000, ch. Fuzzy integral in image processing and recognition, pp. 435–466.
- [6] S. Auephanwiriyakul, J. M. Keller, and P. Gader, "Generalized Choquet fuzzy integral fusion," *Information Fusion*, vol. 3, pp. 69–85, 2002.
- [7] H. Tahani and J. M. Keller, "Information fusion in computer vision using the fuzzy integral," *IEEE Trans. Systems Man Cybernet.*, vol. 20, no. 3, pp. 733–741, 1990.

- [8] D. Zhang and Z. Wang, "Fuzzy integrals of fuzzy valued functions," *Fuzzy Sets and Systems*, vol. 54, pp. 63–67, 1993.
- [9] R. Yang, Z. Wang, P. Heng, and K. Leung, "Fuzzified Choquet integral with a fuzzy-valued integrand and its application on temperature prediction," *IEEE Trans. SMC-B*, vol. 38, no. 2, pp. 367–380, Apr. 2008.
- [10] M. Grabisch, H. T. Nguyen, and E. A. Walker, Fundamentals of Uncertainty Calculi, With Applications to Fuzzy Inference. Dordrecht: Kluwer Academic, 1995.
- [11] T. C. Havens, D. T. Anderson, and J. M. Keller, "A fuzzy Choquet integral with an interval type-2 fuzzy number-valued integrand," in *Proc. IEEE Int. Conf. Fuzzy Systems*, Barcelona, Spain, 2010, pp. 1–8.
- [12] D. T. Anderson, T. C. Havens, C. Wagner, J. M. Keller, M. Anderson, and D. Wescott, "Sugeno fuzzy integral generalizations for sub-normal fuzzy set-valued inputs," in *Proc. IEEE Int. Conf. Fuzzy Systems*, Brisbane, Australia, 2012, pp. 1–8.
- [13] X. Wang, A. Chen, and H. Feng, "Upper integral network with extreme learning mechanism," *Nuerocomputing*, vol. 74, no. 16, pp. 2520–2525, 2011.
- [14] L. Wang, "An improved multiple fuzzy nnc system based on mutual information and fuzzy integral," *Int. J. Mach. Learn. and Cyber.*, vol. 2, no. 1, pp. 25–36, 2011.
- [15] J. Zhai, H. Xu, and Y. Li, "Fusion of extreme learning machine with fuzzy integral," *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 21, no. 2, pp. 23–34, 2013.
- [16] X. Liang, C. Wei, and Z. Chen, "An intuitionistic fuzzy weighted OWA operator and its applications," *Int. J. Mach. Learn. and Cyber.*, vol. 4, no. 6, pp. 713–719, 2013.
- [17] M. Grabisch, T. Murofushi, and M. Sugeno, Eds., Fuzzy Measures and Integrals. Theory and Applications. Berlin: Physica Verlag, 2000.
- [18] D. Dubois and H. Prade, Fuzzy Sets and Systems: Theory and Application. New York, NY: Academic Press, 1980.
- [19] A. Mendez-Vazquez and P. D. Gader, "Sparsity promotion models for the choquet integral," in *Proc. IEEE Symp. Foundations Comp. Int.*, 2007, pp. 454–459.
- [20] J. M. Keller and J. Osborn, "A reward/punishment scheme to learn fuzzy densities for the fuzzy integral," in *Proc. Int. Fuzzy Sys. Assoc. World Cong.*, 1995, pp. 97–100.
- [21] J. M. Keller and J. Osborn, "Training the fuzzy integral," Int. J. Approx. Reasoning, vol. 15, no. 1, pp. 1–24, 1996.
- [22] A. Doan, R. Ramakrishnan, and A. Y. Halevy, "Crowdsourcing systems on the world-wide web," *Communications of the ACM*, vol. 54, no. 4, pp. 86–96, 2011.
- [23] S. Saroiu and A. Wolman, "I am a sensor, and I approve of this message," in *Proc. 11th Workshop on Mobile Computing Systems and Applications*, San Diego, CA, 2010, pp. 37–42.
- [24] C. Wagner and D. T. Anderson, "Extracting meta-measures from data for fuzzy aggregation of crowd sourced information," in *Proc. IEEE Int. Conf. Fuzzy Systems*, Brisbane, Australia, 2012, pp. 1–8.
- [25] T. C. Havens, D. T. Anderson, C. Wagner, H. Deliamsalehy, and D. Wonnacott, "Fuzzy integration of intervals using a measure of generalized accord," in *Proc. IEEE Int. Conf. Fuzzy Systems*, Hyderabad, India, 2013, pp. 1–8.
- [26] G. Choquet, "Theory of capacities," Analles de l'Institit Fourier, vol. 5, pp. 131–295, 1953.
- [27] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measure and the Choquet integral as an integral with respect to a fuzzy measure," *Fuzzy Sets and Systems*, vol. 29, no. 2, pp. 201–227, 1989.
- [28] M. Sugeno, "Theory of fuzzy integral and its applications," Ph.D. dissertation, Tokyo Institute of Technology, 1974.
- [29] M. Grabisch, "Fuzzy integral in multicriteria decision making," Fuzzy Sets and Systems, vol. 69, pp. 279–298, 1995.
- [30] D. Wu, J. M. Mendel, and S. Coupland, "Enhanced interval approach for encoding words into interval type-2 fuzzy sets and its convergence analysis," *IEEE Trans. Fuzzy Systems*, vol. 20, no. 3, pp. 499–513, 2012.
- [31] C. Wagner, S. Miller, J. Garibaldi, D. T. Anderson, and T. C. Havens, "From interval-valued data to general type-2 fuzzy sets," 2014, doi:10.1109/TFUZZ.2014.2310734.
- [32] C. Wagner, D. T. Anderson, and T. C. Havens, "Generalization of the fuzzy integral for discontinuous interval- and non-convex interval fuzzy set-valued inputs," in *Proc. Int. Conf. Fuzzy Systems*, 2013, pp. 1–8.
- [33] O. Mendoza and P. Melin, "Extension of the Sugeno integral with interval type-2 fuzzy logic," in *Proc. NAFIPS*, 2008, pp. 1–8.

- [34] H.-C. Liu, "Type 2 generalized intuitionistic fuzzy Choquet integral operator for multi-criteria decision making," in *Proc. Int. Symp. Parallel* and Distributed Processing with Applications, 2010, pp. 605–611.
- [35] C. Guo, D. Zhang, and C. Wu, "Fuzzy-valued measures and generalized fuzzy integrals," *Fuzzy Sets and Systems*, vol. 97, pp. 255–260, 1998.
- [36] W. Congxin, W. Shuli, and M. Ming, "Generalized fuzzy integrals: part 1. fundamental concepts," *Fuzzy Sets and Systems*, vol. 57, pp. 219–226, 1993.
- [37] D. Zhanga and C. Guo, "Generalized fuzzy integrals of set-valued functions," *Fuzzy Sets and Systems*, vol. 76, no. 3, pp. 365–373, 1995.
- [38] G. Wang and X. Li, "Generalized Lebesgue integrals of fuzzy complex valued functions," *Fuzzy Sets and Systems*, vol. 127, no. 3, pp. 363–370, 2002.
- [39] Y. Enta, "Fuzzy decision theory," in Applied Systems and Cybernetics, Proc Int. Cong. Applied Systems Research and Cybernetics, G. E. Lasker, Ed., 1980, pp. 2980–2990.
- [40] R. Zwick, E. Carlstein, and D. V. Budescu, "Measures of similarity among fuzzy concepts: a comparitive analysis," *Int. J. Approx. Reasoning*, vol. 1, no. 221–242, 1987.
- [41] Z. Mitrovic and S. Rusov, "Z similarity measure among fuzzy sets," *FME Transactions*, vol. 34, pp. 115–119, 2006.



Tim Havens received the Ph.D. degree in electrical and computer engineering from the University of Missouri, Columbia, in 2010.

He is currently the William and Gloria Jackson Assistant Professor with the Departments of Electrical and Computer Engineering and Computer Science at Michigan Technological University.

Dr. Havens received the Best Paper Award at the 2012 IEEE International Conference on Fuzzy Systems, the IEEE Franklin V. Taylor Award for Best Paper at the 2011 IEEE Conference on Systems,

Man, and Cybernetics, and the Best Student Paper Award from the Western Journal of Nursing Research in 2009. He is an associate editor of the IEEE Transactions on Fuzzy Systems.



Derek Anderson received the Ph.D. degree in electrical and computer engineering from the University of Missouri, Columbia, in 2010.

He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at Mississippi State University.

Dr. Anderson received the Best Paper Award at the 2012 IEEE International Conference on Fuzzy Systems and the Best Student Paper Award at the 2008 IEEE International Conference on Fuzzy Systems. He is an associate editor of the IEEE Transactions

on Fuzzy Systems.



Christian Wagner received the Ph.D. degree in computer science from the University of Essex, UK, in 2009.

He is currently an Assistant Professor of Computer Science at the University of Nottingham, UK.

Dr. Wagner received the Outstanding Paper Award for the IEEE Transactions of Fuzzy Systems in 2013 (for paper published in 2010) and the Best Paper Award at the 2012 IEEE International Conference on Fuzzy Systems. He is chair of the IEEE CIS Task Force on Affective Computing and an associate

editor of the IEEE Transactions on Fuzzy Systems.