

Analysing Fuzzy Sets Through Combining Measures of Similarity and Distance

Josie McCulloch *Student Member, IEEE*, Christian Wagner *Senior Member, IEEE* and Uwe Aickelin

Abstract—Reasoning with fuzzy sets can be achieved through measures such as similarity and distance. However, these measures can often give misleading results when considered independently, for example giving the same value for two different pairs of fuzzy sets. This is particularly a problem where many fuzzy sets are generated from real data, and while two different measures may be used to automatically compare such fuzzy sets, it is difficult to interpret two different results. This is especially true where a large number of fuzzy sets are being compared as part of a reasoning system. This paper introduces a method for combining the results of multiple measures into a single measure for the purpose of analysing and comparing fuzzy sets. The combined measure alleviates ambiguous results and aids in the automatic comparison of fuzzy sets. The properties of the combined measure are given, and demonstrations are presented with discussions on the advantages over using a single measure.

I. INTRODUCTION

TO compare two fuzzy sets (FSs) one may consider their similarity or distance. To assess their similarity, we measure the similarity of the membership values for each element in each set. The result is given within the interval $[0, 1]$, where 0 indicates that there are no elements shared between both FSs and 1 indicates that the sets are identical. Alternatively, to assess the distance between FSs, given as a value in \mathbb{R} , we measure the distance between the elements which belong to each set; typically the distance between elements is also weighted by their membership values.

Measures of similarity and distance have frequently been applied to a variety of different applications. For example, similarity has often been used to measure the similarity between different word models [1], [2], or to find similar patterns in classification [3] and clustering [3]. Distance Measures (DMs), though less commonly researched, have been used to compare FSs, for example, in the ranking of fuzzy numbers [4].

Josie McCulloch, Christian Wagner and Uwe Aickelin are with the School of Computer Science, University of Nottingham, Nottingham, United Kingdom (email: {psxjm5, christian.wagner, uwe.aickelin}@nottingham.ac.uk)

This work was partially funded by the EPSRCs Towards Data-Driven Environmental Policy Design grant, EP/K012479/1 and the RCUKs Horizon Digital Economy Research Hub grant, EP/G065802/1.

Measures of similarity and distance evaluate two fundamentally different aspects of FSs, and it is due to the unique properties of these measures, or more directly, through the nature of *what* the measures actually measure that their applicability to a given problem setting is determined. For example, there are cases in which a similarity measure (SM) may not be useful, such as when the FSs are disjoint. In this case, the result of the SM is always zero. This does not tell us how far apart the FSs are placed in the universe of discourse (UoD); they may be far apart or right next to each other. Where this is of concern, a DM may be beneficial. However, likewise, a DM is also not always a useful measure, for example when one FS is a subset of another. In this case the results become ambiguous as DMs are not ideal for detecting overlap between FSs.

Current research within the literature has generally made a choice between using either measures of similarity or distance, however in many cases, it is not trivial to make this choice, in particular when FSs are dynamically created from data such as for approaches like [2] and [5]. This paper proposes the fusion of both measures into a single measure which can be applied in the comparison of FSs and produces meaningful results regardless of the exact nature of the FSs to be measured. The fusion is achieved by an ordered weighted average (OWA) operator, and is applied to data-driven FSs to demonstrate the benefits of the measure.

Section II introduces FSs, SMs, DMs, and OWA operators, followed by an examination of what exactly the measures measure in Section III. In Section IV, a combined measure is presented which utilises the unique properties of both similarity and distance, and demonstrations of the combined measure are shown in Section V. Finally, conclusions are given in Section VI.

II. BACKGROUND

A. Fuzzy Sets

Fuzzy sets have been applied to many applications in which uncertainty is present; some examples of which include data mining [6] and Computing with Words [7]. Unlike traditional logic, for which the membership of each element to the set is a Boolean value, the elements of a FS have a membership value that lies anywhere in

the interval $[0, 1]$. A FS F may be represented as a set of ordered pairs as follows [8]:

$$F = (x, \mu_F(x)) \mid x \in X \quad (1)$$

where $\mu_F(x)$ indicates the membership value of the element x in the FS F . For a discrete UoD, the FS F may be written as [8]

$$F = \sum_x \mu_F(x) / x \quad (2)$$

where \sum denotes the collection of all points $x \in X$ with associated membership value $\mu_F(x)$.

B. Similarity Measures

SMs are a common tool used within fuzzy logic. A SM $s(A, B) \rightarrow [0, 1]$ calculates how similar two FSs are to each other through a comparison of the degrees of membership within each set. Common properties of a SM s for FSs A , B and C are as follows:

Reflexivity: $s(A, B) = 1 \iff A = B$

Symmetry: $s(A, B) = s(B, A)$

Transitivity: If $A \subseteq B \subseteq C$, then $s(A, B) \geq s(A, C)$

Overlapping: If $A \cap B \neq \emptyset$, then $s(A, B) > 0$; otherwise, $s(A, B) = 0$

Note that it is not necessary for a SM to have all of these properties as the application for which the measure is used may not depend on all of them. However, it is typical that a SM always follows the property of reflexivity.

Throughout this paper, similarity is measured using the Jaccard SM, which supports all of the four properties listed above [1]. The Jaccard measure s for FSs A and B is given as:

$$s(A, B) = \frac{\sum_{i=1}^n \min(\mu_A(x_i), \mu_B(x_i))}{\sum_{i=1}^n \max(\mu_A(x_i), \mu_B(x_i))} \quad (3)$$

where n is the total number of discretisations along the x -axis.

C. Distance Measures

A DM $d(A, B) \rightarrow \mathbb{R}^+$ is used to assess the distance between FSs by calculating the distances between the elements in each set.

A DM d on FSs A , B and C holds the following properties:

Self-identity: $d(A, A) = 0$

Separability: $d(A, B) \geq 0$

Symmetry: $d(A, B) = d(B, A)$

Transitivity: If $A \subseteq B \subseteq C$, then $d(A, B) \leq d(A, C)$

Triangle inequality: $d(A, C) \leq d(A, B) + d(B, C)$

The distance between two FSs is most commonly measured by taking α -cuts of FSs and measuring the distance between the α -cuts. The α -cut of the FS

A is a non-FS comprised of all the elements whose membership grade within A is greater than or equal to α [9]; this is written formally as $A_\alpha = \{x \mid \mu_A(x) \geq \alpha\}$.

Chaudhuri and Rosenfeld [10] proposed the following metric to measure the distance between two convex, normal FSs A and B :

$$d(A, B) = \frac{\sum_{i=1}^m y_{\alpha_i} h(A_{\alpha_i}, B_{\alpha_i})}{\sum_{i=1}^m y_{\alpha_i}} \quad (4)$$

where the y -axis is discretised into m points (y_1, y_2, \dots, y_m), A_{α_i} is the non-fuzzy α -cut (given as an interval) of the FS A at y -coordinate y_{α_i} , and h is the conventional Hausdorff metric for two continuous intervals \bar{A} and \bar{B} as follows [11]:

$$h(\bar{A}, \bar{B}) = \max\{|\bar{A}_l - \bar{B}_l|, |\bar{A}_r - \bar{B}_r|\} \quad (5)$$

where $\bar{A} = [\bar{A}_l, \bar{A}_r]$ and $\bar{B} = [\bar{B}_l, \bar{B}_r]$.

In addition to the Hausdorff distance given above, a directional DM (DDM) is given as follows [12]:

$$h(\bar{A}, \bar{B}) = \begin{cases} \bar{B}_l - \bar{A}_l, & \text{if } |\bar{B}_l - \bar{A}_l| > |\bar{B}_r - \bar{A}_r|. \\ \bar{B}_r - \bar{A}_r, & \text{otherwise.} \end{cases} \quad (6)$$

for which a positive distance is given where $A < B$, and a negative value of distance is given where $B > A$. The DDM, however, does not hold the property of symmetry and instead follows partial symmetry, defined as $|d(A, B)| = |d(B, A)|$ and $d(A, B) \neq d(B, A)$ where $A \neq B$. Throughout this paper, (4) is used in conjunction with (6).

Having reviewed SMs and DMs, a brief overview of OWA operators is given next, which will be used to aggregate similarity and distance.

D. Ordered Weighted Average

OWA operators [13] are used to aggregate sub-components of a problem. An OWA involves assigning objects to an ordered set of weights $w = \{w_1, w_2, \dots, w_n\}$, for which $w_i \in [0, 1]$ and $\sum_{i=1}^n w_i = 1$. The objects which are to be aggregated are sorted into descending order, and each object is multiplied by the corresponding weight. Thus, for a given list of objects a_1, a_2, \dots, a_n and weights w_1, w_2, \dots, w_n , the OWA is calculated as follows [13]:

$$F(a_1, a_2, \dots, a_n) = w_1 b_1 + w_2 b_2 + \dots + w_n b_n \quad (7)$$

where b_i is the i_{th} largest element in the collection a_1, a_2, \dots, a_n .

OWAs have been commonly used in the literature to solve a variety of problems. For example, [14] uses an OWA in decision making applied to the personnel selection problem. In [15], an OWA is used to aggregate

different performance indicators to assess the performance of small drinking water utilities, and [15] uses and OWA to aid in the selection of financial products.

III. COMPARISON OF MEASURES ON FUZZY SETS

In this Section, SMs and DMs are compared on a series of real-data driven FSs. This is in order to clarify their respective outputs in an applied context, and to demonstrate the proposition that it can be more beneficial to use a combination of both measures.

As previously discussed, SMs and DMs have unique properties which lead to them measuring fundamentally different concepts. To demonstrate the nature of the measures, and the strengths of using both similarity and distance together to analyse FSs, consider the FSs shown in Fig. 1. These FSs have been constructed from the Movie Lens data set [16], in which films are rated between 1 (poor) and 5 (great). Histograms were created to represent the distribution of ratings and each histogram was normalised by dividing the membership value at each x -coordinate by the peak membership value of the histogram. Linear-interpolation was used to determine membership values between known points.

The SM and DDM introduced in Sections II-B and II-C were applied to each pair of movies, respectively. Their results are shown in Table I. The results of the combined measure are also shown in Table I for comparison purposes, and will be introduced in the next section. For each pair, the FS A was given as the first parameter for the measure, and the FS B was given as the second parameter.

TABLE I
VALUES GIVEN BY SMS AND DMs ON THE FSs IN FIG. 1

Fig. 1 - part:	a	b	c	d	e	f
Similarity (3)	0.050	0.067	0.170	0.242	0.0	0.892
Distance (4)	-3.628	2.936	2.723	-1.999	3.258	0.169
Comparative (8)	-0.915	0.864	0.857	-0.806	0.938	0.072

The following is a discussion of the results for similarity (3) and distance (4) in Table I for the FSs in Fig. 1. For each case, a brief discussion highlights where both measures contribute information that is particularly helpful when considered together.

Sets a & b For the sets in Fig. 1(a) and 1(b), the SM indicates that the FSs are almost disjoint, but there is a small degree of similarity between them. However, there is no indication of where this similarity lies and how much the FSs differ. One can, however, see that the sign of the DM may be helpful to indicate the actual region of similarity. In this case, the direction of the DM tells us that the similarity is likely towards the lower end of

the UoD of the FS A for Fig. 1(a) and the higher end of A for Fig. 1(b).

Sets c & d The SM indicates a small difference in similarity between the sets of c and the sets of d , but it tells us very little else; in both cases, there is a small amount of overlap but we do not know where. However, the DM reveals that this overlap is to the right of the FS A for c , and to the left A for d .

Sets e In this case, the SM indicates that there is no similarity between the FSs, i.e. they are disjoint, and the DM indicates that there is a large amount of distance between the FSs.

Sets f Both the SM and DM are able to identify when two FSs are identical or, in this case, almost identical. For the results of Fig. 1(f), each measure indicates that the membership functions of both FSs are very close to each other.

Given the results above, it is clear that SMs and DMs are each unique functions with distinct properties. This results in the common necessity to choose between both types of measure or indeed to apply both individually. While the application of both measures individually as conducted here can provide some insight, it can be challenging to interpret the two distinct outputs simultaneously for given FSs.

In the next section, the measures are combined into a single measure resulting in a single value, which can be used to determine the similarity, distance and direction between FSs.

IV. COMBINING MEASURES

The comparative measure removes the need to choose between measures of similarity or distance, and by combining both measures it creates a more detailed comparison of FSs. Both of these aspects are particularly important in cases where a potentially large number of FSs are generated from data. In such cases, an appropriate decision between the individual measures (and/or joint result interpretation) cannot be conducted by a human expert but has to be done automatically. Thus, a single measure is proposed to provide a detailed comparison of FSs.

A. A Single Comparative Measure

Note that both measures commonly yield results in different domains; SMs within $[0, 1]$ and DMs within \mathbb{R} (or \mathbb{R}^+ if it is non-directional). A decision must therefore be made as to which domain will be used for the results of the combined measure. The following presents a measure which yields results in $[0, 1]$, for which the value 0 indicates minimum distance/maximum similarity, and the value 1 indicates maximum distance/minimum similarity.

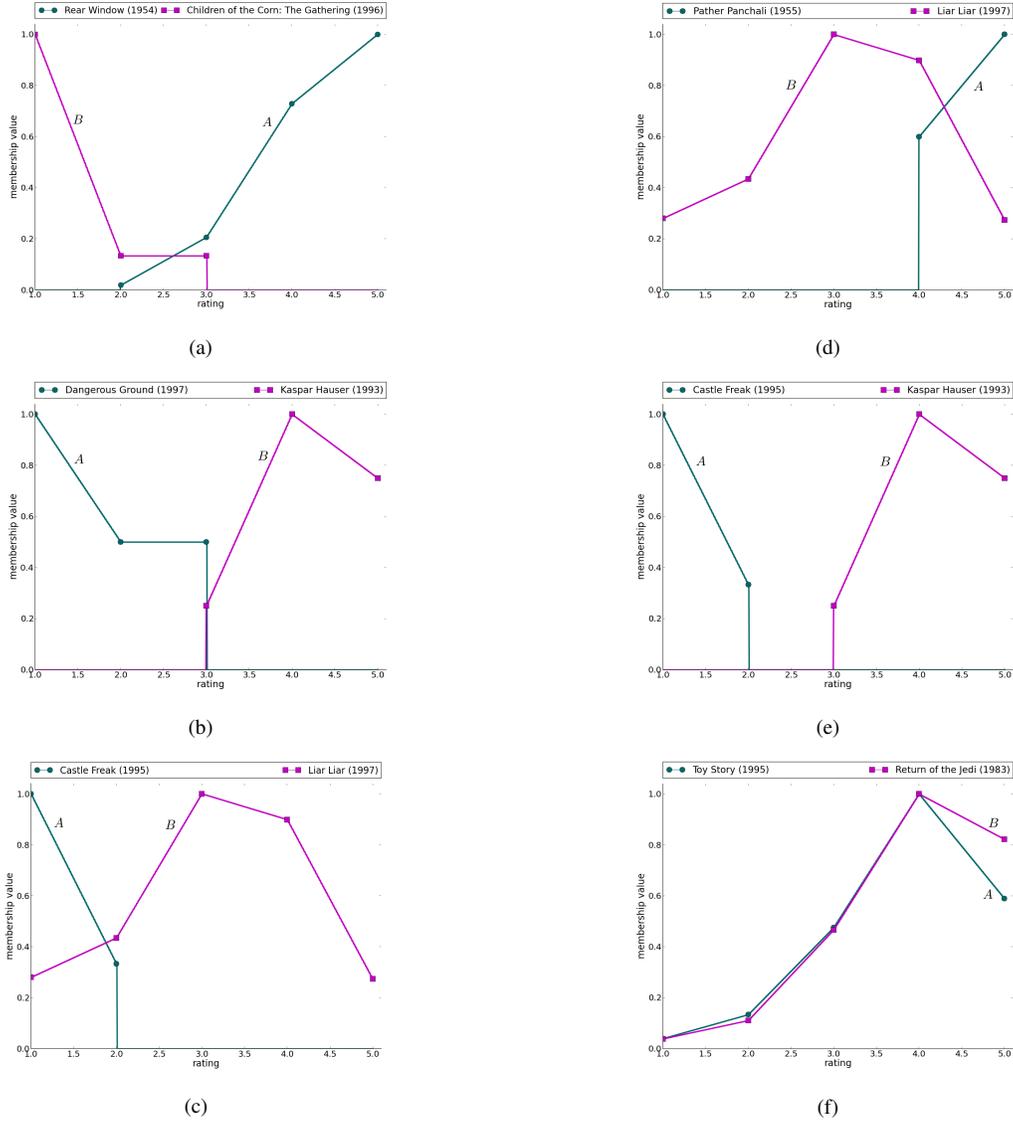


Fig. 1. Fuzzy sets used to demonstrate the attributes of SMs and DMs in Table I.

To fuse the measures, it is important to consider that similarity and distance represent two fundamentally different comparisons of FSs; i.e. both measures measure “opposite” concepts. The SM indicates how *similar* or how *close* two FSs are placed, and the DM indicates how *far apart* they are positioned. To fuse these measures they must both represent the same concept, either both *similarity/closeness* or both *dissimilarity/distance*. The following considers the latter case.

As similarity is within the domain $[0, 1]$, to achieve a measure of dissimilarity for the combined measure the complement of the SM may be used (i.e. $1 - s(A, B)$) [17]. This can then be used in conjunction with the DM.

Considering the DM is within \mathbb{R} , it should be changed

such that the result falls within $[0, 1]$ to enable a meaningful fusion of both measures. To alter the result, it is necessary to take into account the UoD in which the measure has been applied. For example, if the UoD is in $\{1, 2, 3, 4, 5\}$ then the maximum distance that can be achieved is 4. The result of the DM may be given as a ratio to the maximum possible distance. Taking the above into account, $\frac{d(A,B)}{\lambda}$ is used to obtain a ratio of closeness from the DM, where λ is the largest possible distance within the UoD. For a finite UoD X described as $\{x_1, x_2, \dots, x_n\}$, λ will be $x_n - x_1$.

Given the above, the OWA provides a reasonable approach to fusing both measures ((3) and (4)). (8)

presents the comparative measure as an OWA based aggregation of the measures of similarity and distance for FSs A and B :

$$c(A, B) = \begin{cases} F((1-s(A, B)), (\frac{d(A, B)}{\lambda})), & d(A, B) \geq 0 \\ F(-(1-s(A, B)), (\frac{d(A, B)}{\lambda})), & \text{otherwise} \end{cases} \quad (8)$$

where F is an OWA as shown in (7) with weights $w = \{0.7, 0.3\}$ and d is the DDM (4) with (6).

The weights $w = \{0.7, 0.3\}$ are chosen such that the largest of the dissimilarity measure and normalised DM within (8) is assigned the weight 0.7, and the smallest is assigned the weight 0.3. Note that the absolute values of the measures are used when assigning the weights, thus a measure of -0.45 is considered larger than a measure of 0.3. These weights have been determined heuristically as outlined in Section IV-B, and in the future other ways of determining such weights may be investigated.

Note that if the result from the DDM gives a negative value then the result of (8) will also be a negative value. Likewise, if the DDM gives a positive value then the result of the combined measure will also be positive.

In (8), a value of 0 represents identical FSs, as proven in theorem 1 below, and a value of 1 (or -1) represents the maximum distance possible of two disjoint FSs. If one wishes to have the value 1 to represent identical FSs, the complement of (8) may be used as

$$c'(A, B) = \begin{cases} 1 - c(A, B), & \text{if } c(A, B) \geq 0 \\ -1 - c(A, B), & \text{otherwise} \end{cases} \quad (9)$$

Note that (9) maintains the direction according to the comparative measure (8).

Within (8), the measure of similarity is altered such that it reflects dissimilarity or distance. Note that this is just one method of combining the measures proposed because of both its simplicity and its ability to represent both similarity and distance as demonstrated in the examples within the next section. Another method, for example, is the special case where the weights are both 0.5, resulting in the standard average of both measures. It is also possible that the result may be altered to yield results in the domain \mathbb{R} by multiplying the SM by the value λ and fusing the result with the unaltered DM.

B. Choosing the OWA Weights

The following discusses how the weights of the OWA operator (7) may be chosen for the comparative measure (8). Referring to Fig. 2, the FS pairs (A, B) and (A, C) are compared. According to the DM (4) the distance for both pairs is 0.331. However, according to the SM (3)

the similarity of (A, B) is 0.182, and the similarity of (A, C) is 0.0. Due to the DM giving the same result for (A, B) and (A, C) one would assume that the B and C are the same FS. It is only by also referring to the SM (or by viewing the FSs) that it becomes clear that the FSs are different. By using the comparative measure, however, it is possible to distinguish between different pairs of FSs which give equal values of similarity or distance. It is also important to note that, by using the comparative measure, this can be confirmed by using a single measure; a user does not have to check the results of both the SM and DM to ensure pairs of FSs are different.

The weights of the comparative measure play an important role in determining the difference between different pairs of FSs which result in equal values from a measure. Table II shows the difference between pairs (A, B) and (A, C) using a variety of weights. As the first weight increase in value the difference between the two pairs also increases; the results begin to signify that B is closer to A than C is to A . This is because the dissimilarity measure gives 1 for disjoint sets (such as pairs (A, C)) and thus, in such cases, will always be given the first weight. As the first weight increases in value the overall measure is, in effect, placing more importance on the fact that the sets are disjoint.

However, it is unhelpful to have too large of a value for the first weight. If the first weight equals 1 then the output of the combined measure will always equal 1 for disjoint sets. Considering this, the first weight should be low enough such that it is possible to distinguish between different pairs of disjoint sets. However, it must also be large enough such that it is possible to make a distinction between FSs which give equivalent values of similarity or distance, such as those in Fig. 2.

Ideally, when the FSs are known beforehand, the weights should be tuned such that the widest range of values are given by the measure. This decreases possible confusions over pairs of FSs which would give close or identical values from a single measure. However, if the weights cannot be tuned, the weights $\{0.7, 0.3\}$ are ideal and are used throughout this paper. This is because tests showed that these weights are useful for preventing disjoint FSs from resulting in a lower distance/dissimilarity than non-disjoint FSs.

C. Properties of the combined measure

This Section introduces and proves the properties of the combined measure (8).

Theorem 1 (Self-identity). *The comparative measure (8) follows the property of self-identity. That is, $c(A, B) = 0 \iff A = B$.*

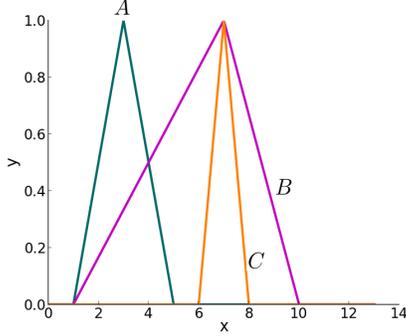


Fig. 2. Three FSs, A, B and C.

TABLE II
COMPARATIVE MEASURE ON THE FSS WITHIN FIG. 2 USING
DIFFERENT WEIGHTS (* INDICATES THE CHOSEN WEIGHTS
WITHIN THIS PAPER).

Weight 0	Weight 1	c(A,B)	c(A,C)
0.0	1.0	0.331	0.331
0.1	0.9	0.380	0.398
0.2	0.8	0.428	0.465
0.3	0.7	0.477	0.532
0.4	0.6	0.526	0.598
0.5	0.5	0.574	0.665
0.6	0.4	0.623	0.732
*0.7	0.3	0.672	0.799
0.8	0.2	0.721	0.866
0.9	0.1	0.769	0.933
1.0	0.0	0.818	1.0

Proof: If $A = B$ then $s(A, B) = 1$ according to the property of reflexivity, and so $w(1 - s(A, B)) = 0$. Also, if $A = B$ then $d(A, B) = 0$ according to the property of self-identity, and so $w\left(\frac{d(A, B)}{\lambda}\right) = 0$ for any w . Thus $c(A, B) = 0$ if $A = B$. Alternatively, if $A \neq B$ then $s(A, B) \neq 1$ and $d(A, B) \neq 0$, thus $c(A, B) \neq 0$. \square

Theorem 2 (Symmetry). *The comparative measure (8) follows the property of symmetry. That is, $c(A, B) = c(B, A)$.*

Proof: If the SM and DM that are aggregated are both symmetrical, then the same values will be given to the comparative measure for both $c(A, B)$ and $c(B, A)$, thus the comparative measure is also symmetrical. \square

Theorem 3 (Partial Symmetry). *Where the DDM (6) is used with the comparative measure (8), the property of partial symmetry holds. That is $|c(A, B)| = |c(B, A)|$, and $c(A, B) \neq c(B, A)$ where $A \neq B$*

Proof: When the distance is a positive value, the

values of both distance and dissimilarity given to the OWA are in the positive domain. Likewise, where the distance is negative, both inputs given to the OWA are in the negative domain. In each case, the absolute values of the positive and negative inputs are the same, thus the absolute values of the outputs are also the same, and the sign of the final value is in same domain as the input values. \square

Theorem 4 (Separability). *The result of the comparative measure is always greater than or equal to zero, i.e. $c(A, B) \geq 0$ when aggregating with the non-DDM.*

Proof: Given that $1 - s(A, B) \in [0, 1]$, and $\frac{d(A, B)}{\lambda} \in [0, 1]$, as λ never exceeds the maximum value of $d(A, B)$, it follows that $c(A, B) \in [0, 1]$ thus $c(A, B) \geq 0$. \square

Note, however, if the DDM is used to construct the comparative measure, then $c(A, B) \in [-1, 1]$. Thus, separability does not apply where the DDM is used.

Theorem 5 (Transitivity). *The comparative measure (8) follows the property of transitivity. That is, if $A \subseteq B \subseteq C$, then $d(A, B) \leq d(A, C)$*

Proof: Given that both the dissimilarity measure and the DM follow transitivity, when they are aggregated the resulting comparative measure also follows transitivity. \square

Theorem 6 (Triangle Inequality). *The comparative measure (8) follows the property of triangle inequality. That is, $c(A, C) \leq c(A, B) + c(B, C)$*

Proof: Given that both the dissimilarity measure [17] and the DM follow triangle inequality, when they are aggregated the resulting comparative measure also follows triangle inequality. \square

The complement of the comparative measure (9), likewise, follows the properties of symmetry, separately, transitivity and triangle inequality. However, the complement does not satisfy self-identity and instead follows reflexivity ($c'(A, B) = 1 \Leftrightarrow A = B$). This is because the complement uses 1 to indicate identical FSs, where as the comparative measure uses 0 instead. It is trivial to see from theorem 1 that the complement of the comparative measure satisfies reflexivity.

Note that the comparative measure does not follow the property of overlapping (i.e. if $A \cap B \neq \emptyset$, then $c(A, B) > 0$, otherwise $c(A, B) = 0$) unless the weights $w = [1.0, 0]$ are given, such that the maximum weight is given to the dissimilarity measure when the FSs are identified as disjoint.

V. DEMONSTRATIONS

Examples of the comparative measure (8) are given in Table I in which the measure is applied to the FSs in Fig. 1. A demonstration and discussion of the comparative measure in an applied context are presented next, and compared against using a single measure of similarity or distance.

A. Demonstration - MovieLens

The following is a discussion of comparisons between the FSs in Fig. 1 according to the comparative measure (8), the results of which are shown in Table I.

Sets a & b The results of a and b in Table I have a high degree of dissimilarity/distance. Additionally, the sign of the comparative measure shows in which direction the dissimilar regions of the FSs reside. To the left of A in the case of the FSs in a and to the right of A in the case of the sets in b . The comparative measure also shows that the FSs within b are closer than those in a .

Sets c & d The FSs within c and d both have a slightly increased degree of dissimilarity/distance than indicated by the original SM and DM. This is due to a large difference between the range of elements contained within the sets, which increases the dissimilarity according to the measure. In both cases, the FS B covers the range $[1, 5]$ whereas A only covers $[1, 2]$ in c , and $[4, 5]$ in d . It can also now be observed, by using the comparative measure alone, that B is to the right of A in c , and is to the left of A in d . The ordering and direction of the distance between c and d are the same using both the DM and the comparative measure.

Sets e The FSs within e are disjoint and were thus given the value 0 by the Jaccard SM. However, the comparative measure gives a non-zero value for e . Note that this value is still the largest dissimilarity/distance compared to the other pairs within Table I. This value also helps to identify the direction of the FSs indicating that the FS B is to the right of A .

Sets f The comparative measure indicates with a high degree of certainty that the FSs of f are nearly identical.

A possible application of the comparative measure is to the problem of ranking. Comparing the comparative measure against the DM, which may also be used for ranking [4], the ordering of the FSs differs. By observing the absolute values of the measures, according to the DM the most distant pair is a and the second most distant is e , however, it is the other way round according to the comparative measure. This is because the SM indicates there is some similarity between the FSs within Fig. 1(a), which causes the comparative measure to decrease in dissimilarity/distance. However, the FSs

TABLE III
SIMILARITY AND DISTANCE BETWEEN THE RESTAURANT AND WORD MODELS IN FIG. 3

	Poor	OK	Great
Similarity (3)	0.081	0.493	0.469
Distance (4)	5.573	1.064	-3.360
Comparative (9)	0.171	0.609	-0.516

in Fig. 1(e) are disjoint, so the dissimilarity/distance remains high. This leaves the FSs with no similarity as the most distant. One could argue that this is an expected result of the comparative measure because the FSs within e are disjoint where as the FSs within a are not disjoint. Thus the measure may be considered more intuitive as it is natural to consider the sets in e as being more distant than the sets in a .

As stated earlier, the unique properties of similarity and distance enable the measures to be applied to a wide variety of fields, and the same can be said for the comparative measure, which, as demonstrated, can be used in terms of a measure of similarity and a measure of distance. For example, with the FSs in Fig. 1 the comparative measure may be used to find similarly rated films by choosing FSs with a low value of dissimilarity/distance, or it may be used to rank the film ratings by ordering the results of the measure.

B. Demonstration - Classification

This section presents a synthetic example of the comparative measure applied to the problem of classification. In this example, three initial FSs are given which represent three different descriptions. In this case, they each represent different levels of ambience within an establishment on a scale from 1 to 10. These levels, as shown in Fig. 3, are labelled as *Poor*, *OK* and *Great*. Given a FS representing the ambience of a restaurant, as shown in Fig. 3, the aim is to classify which description best fits the restaurant.

In Table III, comparisons are given between the ambience of the restaurant and the descriptions. The measures of similarity (3) and distance (4) are shown, as well as the complement of the comparative measure (9). For each measure, the word model is given as the first parameter, and the restaurant is given as the second parameter. The complement of the comparative measure is given to match the SM, such that both measures give the value 1 for identical FSs.

According to the SM (3), the restaurant's ambience is similar to the descriptions *poor*, *OK* and *great* to the degree 0.081, 0.493 and 0.469, respectively. It is clear from these results that the restaurant's ambience cannot

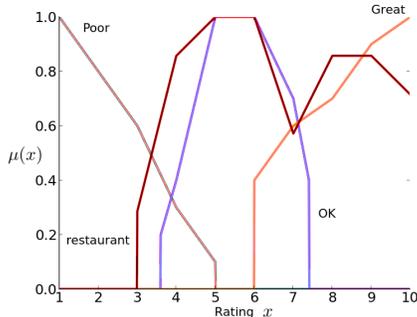


Fig. 3. Three FSs modelling degrees of ambience, with a FS representing the ambience of a given restaurant.

be described as *poor*, however, it is almost equally valid that it may be described as *OK* or *Great*.

The DM (4), however, gives a clearer view of which FS the restaurant most closely matches; the restaurant has a smaller distance to *OK* than to *Great*. Thus, by fusing the distance and similarity as in (8) and (9), a more distinct match is achieved. Now, it is clear from the results of the comparative measure in Table III that the restaurant most closely matches *OK* ambience to the degree of 0.609, whereas it only matches *Great* by 0.516 and *Poor* by 0.171. It can now be determined with greater certainty that *OK* is the correct classification. It should be noted that this can be determined by observing a single measure (9), rather than viewing the SM and DM separately.

VI. CONCLUSIONS

This paper has introduced a novel measure, referred to as a comparative measure, which analyses and compares FSs by combining a SM and DM. When these measures are viewed separately the results may be difficult and time-consuming to interpret as similarity and distance each measure fundamentally different concepts. By joining the measures together, the comparison of FSs is simplified by reducing any ambiguity in the results. Additionally, compared to a single measure, the combined measure provides a richer comparison as it may be swayed towards a preference in representing similarity or distance. This is especially useful for the automatic comparison of a large number of FSs which have been constructed from data. Additionally, through using an OWA operator, it is possible to refine the weights to further alleviate ambiguous values resulting from the original measures.

Demonstrations using data-driven FSs have shown that the comparative measure may be applied in terms

of both similarity and distance, and as such may be applied to applications of these measures. Though the demonstrations have been applied to type-1 FSs only, as the comparative measure uses the outputs of the SM and DM, it may also be applied to type-2 FSs, where the original measures are a type-2 SM and DM.

Future work will look at measures which indicate similarity and distance as a FS, which better reflects the uncertainty inherent in FSs.

REFERENCES

- [1] D. Wu and J. M. Mendel, "A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets," *Information Sciences*, vol. 179, no. 8, pp. 1169–1192, 2009.
- [2] C. Wagner, S. Miller, and J. Garibaldi, "Similarity based applications for data-driven concept and word models based on type-1 and type-2 fuzzy sets," in *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, 2013, pp. 1–9.
- [3] W. Wang and X. Xin, "Distance measure between intuitionistic fuzzy sets," *Pattern Recognition Letters*, vol. 26, no. 13, pp. 2063–2069, 2005.
- [4] C.-H. Cheng, "A new approach for ranking fuzzy numbers by distance method," *Fuzzy Sets and Systems*, vol. 95, no. 3, pp. 307–317, 1998.
- [5] S. Coupland, J. Mendel, and D. Wu, "Enhanced interval approach for encoding words into interval type-2 fuzzy sets and convergence of the word focus," in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, 2010, pp. 1–8.
- [6] K. Hirota and W. Pedrycz, "Fuzzy computing for data mining," *Proceedings of the IEEE*, vol. 87, no. 9, pp. 1575–1600, Sep. 1999.
- [7] L. Zadeh, "Fuzzy logic = computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 103–111, May 1996.
- [8] J. Mendel, *Uncertain rule-based fuzzy logic systems: introduction and new directions*. Prentice Hall PTR, 2001.
- [9] L. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.
- [10] B. Chaudhuri and A. Rosenfeld, "On a metric distance between fuzzy sets," *Pattern Recognition Letters*, vol. 17, no. 11, pp. 1157–1160, 1996.
- [11] R. Zwick, E. Carlstein, and D. V. Budeanu, "Measures of similarity among fuzzy concepts: A comparative analysis," *International Journal of Approximate Reasoning*, vol. 1, no. 2, pp. 221–242, 1987.
- [12] J. McCulloch, C. Wagner, and U. Aickelin, "Measuring the directional distance between fuzzy sets," in *Computational Intelligence (UKCI), 2013 13th UK Workshop on*. IEEE, 2013, pp. 38–45.
- [13] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 1, pp. 183–190, 1988.
- [14] L. Cans and V. Liern, "Soft computing-based aggregation methods for human resource management," *European Journal of Operational Research*, vol. 189, no. 3, pp. 669–681, 2008.
- [15] R. Sadiq, M. J. Rodriguez, and S. Tesfamariam, "Integrating indicators for performance assessment of small water utilities using ordered weighted averaging (owa) operators," *Expert Systems with Applications*, vol. 37, no. 7, pp. 4881–4891, 2010.
- [16] "Movielens dataset," Last Access: 12/2013. [Online]. Available: <http://grouplens.org/datasets/movielens/>
- [17] A. Lipkus, "A proof of the triangle inequality for the tanimoto distance," *Journal of Mathematical Chemistry*, vol. 26, no. 1-3, pp. 263–265, 1999.



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

McCulloch, J; Wagner, C; Aickelin, U

Title:

Analysing Fuzzy Sets Through Combining Measures of Similarity and Distance

Date:

2014-01-01

Citation:

McCulloch, J., Wagner, C. & Aickelin, U. (2014). Analysing Fuzzy Sets Through Combining Measures of Similarity and Distance. 2014 IEEE INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS (FUZZ-IEEE), pp.155-162. IEEE. <https://doi.org/10.1109/FUZZ-IEEE.2014.6891672>.

Persistent Link:

<http://hdl.handle.net/11343/241191>

File Description:

Accepted version