Measuring the Directional Distance Between Fuzzy Sets

Josie McCulloch
School of Computer Science
University of Nottingham
Nottingham, United Kingdom
psxjm5@nottingham.ac.uk

Christian Wagner
School of Computer Science
University of Nottingham
Nottingham, United Kingdom
christian.wagner@nottingham.ac.uk

Uwe Aickelin
School of Computer Science
University of Nottingham
Nottingham, United Kingdom
uwe.aickelin@nottingham.ac.uk

Abstract—The measure of distance between two fuzzy sets is a fundamental tool within fuzzy set theory. However, current distance measures within the literature do not account for the direction of change between fuzzy sets; a useful concept in a variety of applications, such as Computing With Words. In this paper, we highlight this utility and introduce a distance measure which takes the direction between sets into account. We provide details of its application for normal and non-normal, as well as convex and non-convex fuzzy sets. We demonstrate the new distance measure using real data from the MovieLens dataset and establish the benefits of measuring the direction between fuzzy sets.

Index Terms—distance measure, fuzzy sets, Hausdorff metric, directional distance

I. INTRODUCTION

Distance measures for fuzzy sets are an important tool and have been applied to many fields. For example, Bonissone [1] illustrates examples of applying distance measures in decision analysis and artificial intelligence and Wang and Xing [2] demonstrate distance measures applied to pattern recognition, particularly to the problem of classification. Turksen and Zhang [3] also demonstrate the applicability of similarity based on distance measures in fuzzy logic inference based on analogical reasoning.

A function $d(A, B) \rightarrow \mathbb{R}^+$, for which A and B are fuzzy sets in the universe of discourse X, is commonly called a distance measure if it satisfies the following properties [4]:

1) $d(A, B) = d(B, A)$
2) $d(A, A) = 0$
3) $d(D, D^c) = max_{A,B \in X} d(A, B)$
4) If $A \subset B \subset C$, then $d(A, B) \leq d(A, C)$ and $d(B, C) \leq d(A, C)$

Distance measures that are currently in the literature do not account for the "change in direction" between fuzzy sets. That is, they reveal the distance between two fuzzy sets, but they do not indicate if a fuzzy set is placed to the left or right of another fuzzy set; a concept which will prove useful within Computing with Words (CW) and the ranking of fuzzy numbers. This paper discusses the value of using a distance

measure which identifies the direction of change between fuzzy sets, and proposes an extension of the Hausdorff metric to implement it. Additional extensions of the distance measure are also presented to solve measuring the distance between non-normal and non-convex fuzzy sets.

In Section II we present some background information of fuzzy sets, followed by a discussion presenting the importance of using α-cuts in distance measures of fuzzy sets as well as introducing a number of alpha-cut based distance measures currently used in the literature. Section III introduces a new direction-based distance measure, followed by demonstrations of the new measure compared with current measures in Section IV. Demonstrations using synthetic fuzzy sets and real data are presented. Sections V and VI look at extensions of the newly proposed distance measure for non-normal and non-convex fuzzy sets, respectively. Finally, conclusions are presented in Section VII.

II. BACKGROUND

A. Fuzzy Sets

Fuzzy sets were first introduced by Zadeh [5] in 1965 and have since been applied to many fields, including data mining [6], time-series prediction [7] and CW [8]. A fuzzy set is a set in which the membership of each element is no longer a Boolean, i.e. not 0 or 1, but instead its membership lies in the interval [0,1]. A fuzzy set F may be viewed as a set of ordered pairs as follows [9]:

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

where $\mu_F(x)$ indicates the membership grade of the element x in the fuzzy set F. In a discrete universe of discourse, the fuzzy set F can also be written as [9]

$$F = \sum_x \mu_F(x) \, / \, x \qquad (2)$$

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

B. α-cuts vs Vertical Slices

In [10] it is noted that a distance measure for fuzzy sets should ideally focus on the ordering within the x-axis as it is

this axis that holds the important information regarding where a set's membership lies.

To show the importance of α-cuts in distance measures, consider the following example. A survey of three restaurants (A, B and C) is taken to find out how delicious the food is at each restaurant, on a scale of 0 to 10. Fuzzy sets are then constructed using the results of the survey; Fig. 1(a) shows an example of such fuzzy sets. A distance measure can be used on these fuzzy sets to determine how much more delicious the food of one restaurant is compared to another. Two common methods of measuring the distance (or similarity) between fuzzy sets are vertical slices (shown in Fig. 1(b)) and α-cuts (shown in Fig. 1(c)). The α-cut of the fuzzy set A is a non-fuzzy set comprised of all the elements whose membership grade within A is greater than or equal to α [11]; this is written as $A_\alpha = \{x \mid \mu_A(x) \geq \alpha\}$.

Using a vertical slice approach will give an indication of how much the two sets overlap, capturing the respective food quality. This is useful and most often used when measuring the similarity between the two sets e.g. their intersection, but it does not indicate how far apart they are along the x-axis. For example, if the intersection between two fuzzy sets is the empty set then both sets could be infinitely far apart or right next to each other.

By using α-cuts to measure the distance between fuzzy sets we get an approximation of how far apart the two sets are in terms of their universe of discourse. Thus, the result of the distance measure will be a value that is meaningful to the user. For example, if the distance between two of the restaurants is 5 then the user can understand that one of the restaurants was rated approximately 5 points higher than the other restaurant.

Two equations ((3) and (4)) have been created to demonstrate this idea. (3) uses vertical slices and is written as follows:

$$d(A, B) = \frac{1}{n} \sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)| \quad (3)$$

where n is the total number of discretisations on the x-axis. The next equation (4) uses α-cuts and is written as:

$$d(A, B) = \frac{1}{m} \sum_{i=0}^{m} max\{|A_{\alpha_{i_l}} - B_{\alpha_{i_l}}|, |A_{\alpha_{i_r}} - B_{\alpha_{i_r}}|\} \quad (4)$$

where $A_{\alpha_{i_l}}$ is the left point of the α-cut of A at α_i, $A_{\alpha_{i_r}}$ is the right point of the α-cut of A at α_i, and m is the total number of α-cuts. The fuzzy sets used in this demonstration are shown in Fig. 1(a). Both the x-axis and y-axis were discretised into points each at a distance of 0.1. The results of this demonstration are displayed in Table I. Examining the fuzzy sets in Fig. 1(a), it would be sensible to describe set C as being at a greater distance from set A than set B is from A. However, it can be clearly seen that when using vertical slices the result of the distance measure decreases as the sets are placed further apart along the x-axis. This is a result of the equation measuring the distance between the fuzzy sets according to their membership values on the y-axis, which can lead to unintuitive results when the fuzzy sets being measured
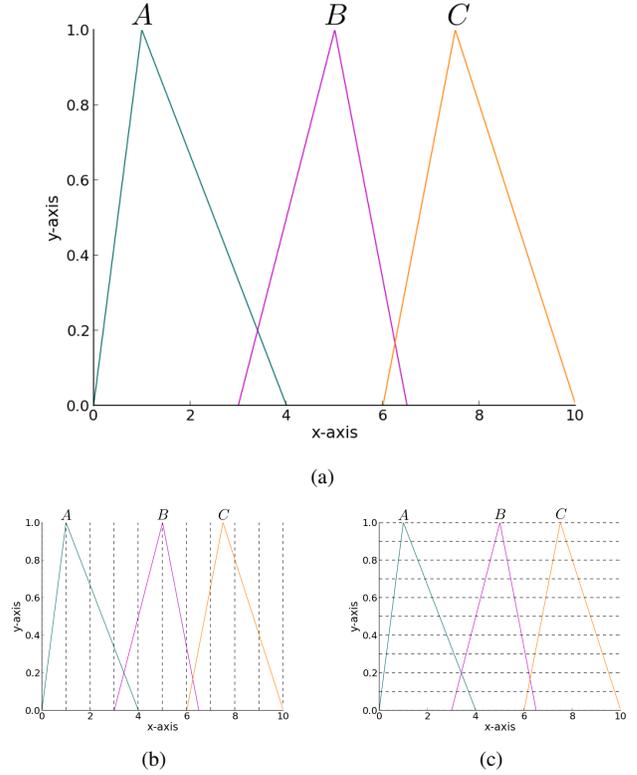

(a)

(b)

(c)

Fig. 1. (a) Three fuzzy sets to demonstrate vertical slices and α-cuts; A, B and C (b) Dashed lines representing vertical slices. (c) Dashed lines representing α-cuts.

TABLE I
A COMPARISON OF VERTICAL SLICE AND α-CUT APPROACHES ON MEASURING DISTANCE USING THE FUZZY SETS IN FIG. 1(A).

	(A, A)	(A, B)	(A, C)
Vertical slice based distance (3)	0.0	0.538	0.396
α-cut based distance (4)	0.0	3.495	6.262

are disjoint. For example, between $x = 4$ and $x = 6$ the distance between sets A and C is 0 because they both have the same membership value at these coordinates.

Considering the α-cut approach, the distance according to the measure increases as the sets move further apart along the x-axis. This seems logical as it is the x-axis that is of most importance when defining sets so it is generally this axis that a user is interested in when comparing different fuzzy sets.

C. Current Distance Measures

In this section, the Hausdorff metric for two intervals is reviewed, followed by two existing distance measures for fuzzy sets which use α-cuts.

The Hausdorff metric gives a generalisation of the distance between two non-empty crisp sets. The Hausdorff distance between two intervals is defined as [10]:

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

where $\bar{A} = [\bar{A}_l, \bar{A}_r]$ and $\bar{B} = [\bar{B}_l, \bar{B}_r]$. This is a common metric used to measure the distance between two α-cuts. Two common distance measures for fuzzy sets which use the Hausdorff metric are introduced next.

Ralescu and Ralescu [12] introduced the following generalisation of the Hausdorff metric to measure the distance between fuzzy sets:

$$d_{RR}(A, B) = \int_{\alpha=0}^{1} h(A_\alpha, B_\alpha) d\alpha \qquad (6)$$

where h is the conventional Hausdorff metric as shown in (5).

Chaudhuri and Rosenfeld also proposed a new metric to determine the distance between fuzzy sets based on the Hausdorff metric as follows [13].

$$d_{CR}(A, B) = \frac{\sum_{\alpha=1}^{m} y_\alpha \, h(A_\alpha, B_\alpha)}{\sum_{\alpha=1}^{m} y_\alpha} \qquad (7)$$

where the y-axis is discretised into m points $(y_1, y_2, ..., y_m)$, A_α is the non-fuzzy α-cut set of the fuzzy set A at y-coordinate y_α, and h is the conventional Hausdorff metric in (5).

Both Ralescu & Ralescu's and Chaudhuri & Rosenfeld's distance measures assume the fuzzy sets being measured are normal (i.e. $\exists x \in X \ \mu_A(x) = 1$ and $\exists x \in X \ \mu_B(x) = 1$). However, an extension of Chaudhuri & Rosenfeld's measure (7) for non-normal fuzzy sets (i.e. $\forall x \in X \ \mu_A(x) < 1$ or $\forall x \in X \ \mu_B(x) < 1$) is given in [13] and is discussed in Section V.

Having reviewed the core concepts of this paper, we proceed to develop direction-based distance measures for normal and convex, as well as non-normal and non-convex fuzzy sets.

III. DIRECTIONAL DISTANCE MEASURES

The following proposes the idea that the property of symmetry for distance measures (i.e., $d(A, B) = d(B, A)$) is not ideal for every problem. For example, consider two fuzzy sets that represent on a scale of 1 to 10 how fun two different roller coasters are according to a public survey (1 denoting not at all and 10 meaning very much). When comparing the distance between these fuzzy sets, the result will indicate how similar the roller coasters are in terms of how fun they are, but it will not indicate which is more fun. For example, if $d(\text{rollercoasterA}, \text{rollercoasterB}) = 6$, we will know that one of them was given, roughly speaking, "6 more points" than the other. However, there is no way to determine which is the more fun roller coaster without visually checking which set is actually on the left or right of the other. This can be time-consuming and tedious if many comparisons need to be made.

Instead, it would be ideal if the result were a signed value. For example, the result will be a positive value if rollercoasterB is more fun than rollercoasterA, and will be a negative value if the opposite case is true. For example $d(\text{rollercoasterA}, \text{rollercoasterB}) = 6$ indicates that rollercoasterB was rated approximately 6 points higher than rollercoasterA. For the same case, it will be true that

$d(\text{rollercoasterB}, \text{rollercoasterA}) = -6$, indicating that rollercoasterA was rated approximately 6 points less than rollercoasterB. Now, it is clear from the result of d which roller coaster is the most fun and by how much.

By taking this approach, the distance measure will no longer have the property of symmetry (i.e., $d(A, B) \neq d(B, A)$), however the absolute values of $d(A, B)$ and $d(B, A)$ will be equal (i.e. $|d(A, B)| = |d(B, A)|$).

The concept of a directional distance measure will prove useful in analysing survey data as shown in Section IV-B and will be a valuable tool in CW [8] and, more generally, in the evaluation of ratings and rankings of fuzzy numbers and sets.

As mentioned in (5), when comparing intervals, the Hausdorff metric is described by

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

where $\bar{A} = [\bar{A}_l, \bar{A}_r]$ and $\bar{B} = [\bar{B}_l, \bar{B}_r]$ [10]. Currently, this will never give a negative value for a negative distance. However, it can be modified as follows:

$$h(\bar{A}, \bar{B}) = \begin{cases} \bar{B}_l - \bar{A}_l, & \text{if } |\bar{B}_l - \bar{A}_l| > |\bar{B}_r - \bar{A}_r|. \\ \bar{B}_r - \bar{A}_r, & \text{otherwise.} \end{cases} \qquad (8)$$

This ensures that both the maximum distance and the sign are preserved. For example, in Fig. 2, set $\bar{A}$ is defined as $[1, 3]$ and set $\bar{B}$ is defined as $[5, 11]$. By using (8) for $h(\bar{A}, \bar{B})$ the result of h is 8. Alternatively, when computing $h(\bar{B}, \bar{A})$ the result of h is -8.

By testing $|\bar{B}_l - \bar{A}_l| > |\bar{B}_r - \bar{A}_r|$ within (8), the absolute values of $h(\bar{A}, \bar{B})$ and $h(\bar{B}, \bar{A})$ remain the same as when using the conventional Hausdorff metric (5). It also follows that the modified property $|d(A, B)| = |d(B, A)|$ is satisfied.
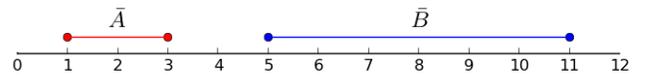

Fig. 2. Two interval sets, $\bar{A}$ and $\bar{B}$.

IV. DEMONSTRATIONS

This section presents demonstrations of the new direction-based distance measure. The distance measures by Ralescu and Ralescu given in (6) and Chaudhuri and Rosenfeld (7) are presented using the regular Hausdorff metric for intervals as shown in (5), and the new direction-based measure shown in (8). Two demonstrations are presented in this section. The first uses synthetic fuzzy sets, and the second uses fuzzy sets constructed from real data.

A. Synthetic Example

The fuzzy sets shown in Fig. 3 were constructed to demonstrate the new direction-based distance measure compared to conventional measures. For this demonstration, 51 α-cuts were taken for the distance measures. We next present and discuss the results of this demonstration shown in Table II.

It can be seen through the measurements based on sets A through C, that as the distance between the sets along the x-axis increases, the value of the distance measures also increases. Also, for the proposed direction-based measures based on (8), if the fuzzy set given as the second parameter to the distance function is placed to the right of the fuzzy set given as the first parameter then the distance according to the measure is a positive value, and if the fuzzy set of the second parameter is to the left of the fuzzy set of the first parameter then the distance measure gives a negative value. Also, it can be seen by the results between fuzzy sets A and D and fuzzy sets A and E that if the peak of the second fuzzy set is positioned to the right of the peak of the first fuzzy set then the result of the distance measure will be a positive value.

Note from Table II that when using (8), the modified property of distance measures has been obeyed. Only the sign of the value has changed, but their absolute values are the same. For example, $d(A, C)$ and $d(C, A)$ share the same absolute value, it is only their signs that differ. Additionally, it can be seen in Table II that the results from Ralescu & Ralescu's and Chaudhuri & Rosenfeld's equations share the same absolute value using (8) and (5).

B. Real World Example

MovieLens is a dataset of movie ratings developed by the GroupLens project at the University of Minnesota. Their datasets are available at *http://www.grouplens.org*. The dataset used for this demonstration is the 100k MovieLens dataset, which consists of 100,000 ratings from 943 users on 1682 movies, where each rating is between 1 and 5.

Fuzzy sets were constructed for each film by calculating the histogram of all ratings given to each film, and linear interpolation was used to find the values between known points. The distance measures introduced thus far have been designed only for normalised fuzzy sets, so it is necessary that the histograms of the films are normalised so that the distance measures can be used. The sets were normalised by dividing the y value at each x-coordinate by the peak y value; this ensures that the peak value of each fuzzy set is now 1. The histograms for the three films *Super Mario Bros.* (SMB), *Mars Attacks!* (MA) and *Star Wars* (SW) and their normalised fuzzy sets are shown in Fig. 4. The two previously shown distance measures by Ralescu and Ralescu (RR) (6) and by Chaudhuri and Rosenfeld (CR) (7) are demonstrated using the conventional Hausdorff metric in (5) and the proposed, modified, direction-based Hausdorff measure in (8). These measures were applied to the fuzzy sets in Fig. 4(b) to
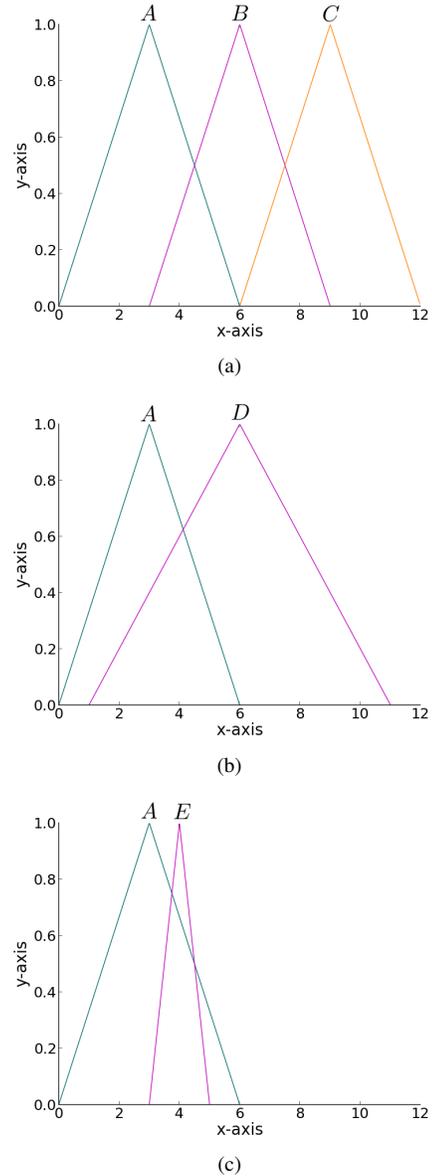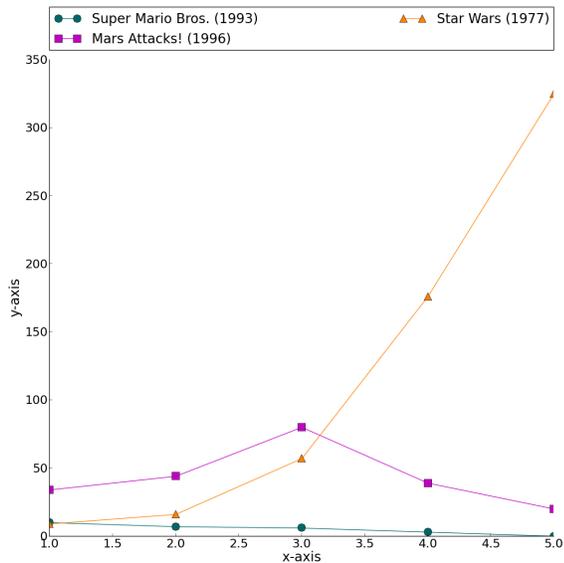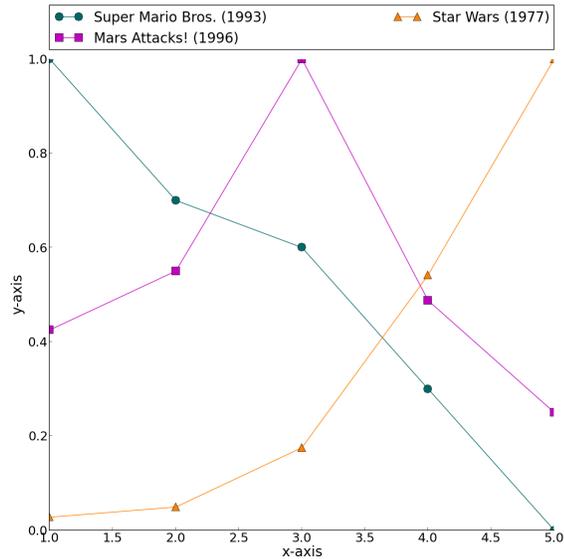

Fig. 3. Fuzzy sets used to demonstrate the distance measures. (a) Fuzzy sets A, B and C. (b) Fuzzy sets A and D. (c) Fuzzy sets A and E.

TABLE II
RESULTS OF DISTANCE MEASURE APPLIED TO THE FUZZY SETS IN FIG. 3.

	(A,A)	(A,B)	(A,C)	(C,A)	(B,A)	(A,D)	(D,A)	(A,E)	(E,A)
Ralescu & Ralescu & (5)	0.0	3.0	6.0	6.0	3.0	4.0	4.0	2.0	2.0
Ralescu & Ralescu & (8)	0.0	3.0	6.0	-6.0	-3.0	4.0	-4.0	2.0	-2.0
Chaudhuri & Rosenfeld & (5)	0.0	3.0	6.0	6.0	3.0	3.65	3.65	1.65	1.65
Chaudhuri & Rosenfeld & (8)	0.0	3.0	6.0	-6.0	-3.0	3.65	-3.65	1.65	-1.65

(a)

(b)

Fig. 4. (a) Histogram based fuzzy sets of three film ratings from the MovieLens dataset. (b) Normalised histogram based fuzzy sets of the same films.

	(SMB, MA)	(SMB, SW)	(MA, SW)	(MA, SMB)	(SW, SMB)	(SW, MA)
RR & (5)	1.194	2.775	1.974	1.194	2.775	1.974
RR & (8)	1.194	2.775	1.974	-1.194	-2.775	-1.974
CR & (5)	1.399	3.270	2.097	1.399	3.270	2.097
CR & (8)	1.399	3.270	2.097	-1.399	-3.270	-2.097

is between SMB and SW as their membership functions share an inverse relationship, and the peaks of their membership functions are the furthest apart than any other two films. In Table III it can be seen that SMB and SW do have the greatest distance by a considerable margin.

Observing the measures by Ralescu & Ralescu and Chaudhuri & Rosenfeld when using the conventional Hausdorff metric (5), the difference in ratings between each film is clear. For example, in both measures the distance between SMB and SW is clearly greater than the distance between SMB and MA, however, it is not clear which film is better. When using the direction-based Hausdorff measure in (8) the latter is not the case. Thus, the extended measure has shown additional information whilst also maintaining the same absolute values that are produced when using the conventional Hausdorff metric.

V. NON-NORMAL FUZZY SETS

Thus far, the distance measures introduced can only be applied to normalised fuzzy sets, where the peak of a set's membership is 1. However, it is common for sets to be non-normal. Referring to Fig. 4, it is clear from the original histograms in Fig. 4(a) that SW received a higher proportion/number of ratings of 5 than MA did of 3, however, this information is lost in the normalised histograms in Fig. 4(b). The following process was performed to retain this information. For every film, each value on the y-axis pertaining to the number of times a rating on the x-axis was given is divided by the total number of ratings given for that film; ergo the membership value $\mu_A(x)$ now indicates the percentage of people that gave the rating x for the film A. This was applied to all of the films in Fig. 4(a) resulting in the fuzzy sets shown in Fig. 5.

It can be seen that the current α-cut approach to measuring distance cannot be used on these sets. For example, comparing SMB with SW at $\alpha = 0.5$, the α-cut of SMB is $[null, null]$ because the set is not present at this α-level. However, SW has an α-cut of $[4.8, 5.0]$. This poses the question 'how can the distance between these two fuzzy sets be measured?' Existing solutions from the literature include the following approaches by Chaudhuri and Rosenfeld [13] and Fan [14].

Chaudhuri and Rosenfeld put forward an extension of their distance measure (7) in [13]. Firstly, each set is modified so that the maximum membership value of each fuzzy set is 1, and these modified fuzzy sets are applied to (7). Next, the original, non-modified fuzzy sets are applied to the following

determine how much better or worse each film was rated compared to each other film. The results of this experiment, for which 51 α-cuts were taken, are shown in Table III.

The results of the distance measures shown in Table III indicate that the difference in rating between MA and SW is slightly greater than the distance between SMB and MA Referring to the fuzzy sets in Fig. 4(b), these results are expected; the membership functions of SMB and MA are much closer to each other compared to the membership functions of MA and SW. Additionally, referring to Fig. 4(b), it is expected that the greatest distance between any two films

equation [13]

$$e(A, B) = \varepsilon \frac{\sum_{x \in X} |\mu_A(x) - \mu_B(x)|}{\sum_{x \in X} x} \qquad (9)$$

where ε is a small positive constant, and its value is determined by the importance of the equation. Finally (7) and (9) are joined together as follows [13]:

$$d_{CR}(A, B) = \frac{\sum_{\alpha=1}^{m} y_\alpha \; h(A_\alpha, B_\alpha)}{\sum_{\alpha=1}^{m} y_\alpha} + \varepsilon \frac{\sum_{x \in X} |\mu_A(x) - \mu_B(x)|}{\sum_{x \in X} x} \qquad (10)$$

Using Chaudhuri and Rosenfeld's method in (10) with the direction-based Hausdorff measure in (8), we have altered the extension for non-normal fuzzy sets to no longer take the absolute value $|\mu_A(x_i) - \mu_B(x_i)|$. This is in keeping with the desire to maintain information regarding the direction of change between the fuzzy sets, and ensures that the property $|d(A, B)| = |d(B, A)|$ is maintained. Thus, Chaudhuri and Rosenfeld's measure is altered to

$$d_{CR}(A, B) = \frac{\sum_{\alpha=1}^{m} y_\alpha \; h(A_\alpha, B_\alpha)}{\sum_{\alpha=1}^{m} y_\alpha} + \varepsilon \frac{\sum_{x \in X} \mu_B(x) - \mu_A(x)}{\sum_{x \in X} x} \qquad (11)$$

Fan [14] also put forward the following extension of the Hausdorff metric for non-normal fuzzy sets within the compact metric space S. Let $h(\emptyset, \emptyset) = 0$ and $h(\emptyset, U) = h(U, \emptyset) = w$ for all non-empty sets U, where $w = \sup h(U, V)$ for all non-empty compact subsets $U, V \subset S$. Based on Fan's measure, we propose the following extension.

Let $h(A_\alpha, \emptyset) = h(A_{\alpha_k}, B_{\alpha_k})$ where α_k is the α-level at $max\{|h(A_\alpha, B_\alpha)|\} \; \forall \alpha \; A_\alpha \neq \emptyset \wedge B_\alpha \neq \emptyset$. Likewise, let $h(\emptyset, A_\alpha) = h(B_{\alpha_k}, A_{\alpha_k})$ where α_k is the α-level at $max\{|h(B_\alpha, A_\alpha)|\} \; \forall \alpha \; A_\alpha \neq \emptyset \wedge B_\alpha \neq \emptyset$. By using this, the distance between an α-cut and the empty set is the maximum distance between the non-empty α-cuts of the sets being measured. Additionally, this approach also ensures that the sign of the distance is maintained.

We propose to disregard $h(\emptyset, \emptyset)$ for the following reason. Using (8), the distance between SMB and SW in Fig. 5 at $\alpha = 0.36$ is 2.52. Next, the distance at $\alpha = 0.46$ (at which the α-cut of SMB is the empty set) is 2.56. This was calculated using the newly introduced extension based on Fan's extension [14]. Considering this, it is unrealistic to state that the distance at $\alpha = 0.56$, at which the α-cut of each film is the empty set, is 0. Though neither fuzzy set is present at this α-cut, it is nonsensical to describe the fuzzy sets as having a distance of 0 when all other α-cuts denote otherwise. Taking this into account, we propose the following new distance measure for non-normal fuzzy sets:

$$d_{CRF}(A, B) = \frac{\sum_{\alpha \in [0, \lambda]]} y_\alpha \; h(A_\alpha, B_\alpha)}{\sum_{\alpha \in [0, \lambda]} y_\alpha} \qquad (12)$$

where $\lambda = \sup\{\alpha \in [0, 1] : A_\alpha \neq \emptyset \vee B_\alpha \neq \emptyset\}$ and h is described in (8). A numerical example of measuring the distance between non-normal fuzzy sets using (12) is presented in the appendix.
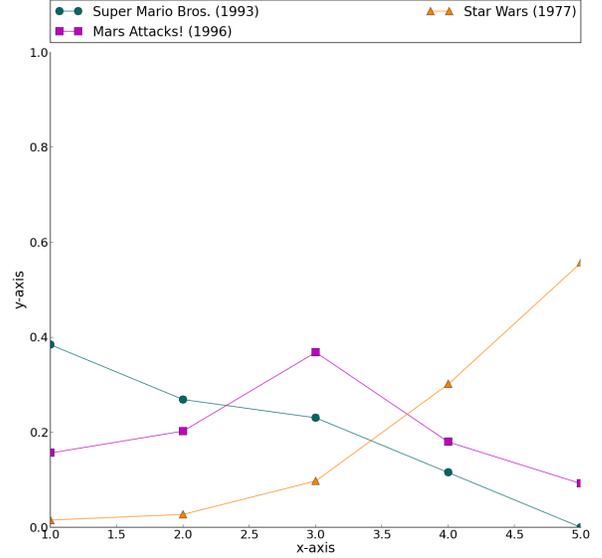

Fig. 5. Three non-normal film ratings from the MovieLens dataset.

TABLE IV
RESULTS OF DISTANCE MEASURES APPLIED TO THE FUZZY SETS IN FIG. 5.

	(SMB, MA)	(SMB, SW)	(MA, SW)	(MA, SMB)	(SW, SMB)	(SW, MA)
(11) & (8)	1.431	3.261	2.057	-1.431	-3.261	-2.057
(12) & (8)	0.904	2.374	2.348	-0.904	-2.374	-2.348

Demonstrations were performed on these non-normal fuzzy sets using Chaudhuri and Rosenfeld's altered method given in (11), and the proposed method shown in (12). Both equations use the modified direction-based Hausdorff measure in (8). Each axis was discretised into 51 equidistant points, and ε was set as 1. The results are shown in Table IV.

Comparing the results on the normalised fuzzy sets in Table III with the results on the non-normal fuzzy sets in Table IV, the distances of the non-normal sets have changed significantly. For example, in Table IV, according to (12), SMB and SW are now approximately as close to each other as MA and SW. Whereas in the previous experiment in Table III the films SMB and SW have the greatest distance by a considerable margin. In Fig. 5, the difference between the membership functions of SMB and MA is much smaller than in Fig. 4 because their membership functions have been compressed. It is due to their membership functions being more similar that the distance of these films from SW is also similar. The distances according to (11), however, are approximately the same in both experiments. Note that in Table IV both Chaudhuri & Rosenfeld's altered measure and the proposed measure based on Fan's extension still satisfy $|d(A, B)| = |d(B, A)|$.

VI. NON-CONVEX FUZZY SETS

The distance measures discussed so far can only be utilised on convex fuzzy sets. For example, consider the fuzzy set for
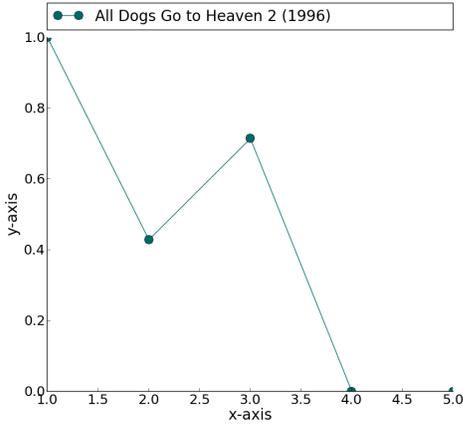
Fig. 6. The non-convex fuzzy set for the film *All Dogs Go to Heaven 2*.

the film *All Dogs Go to Heaven 2* shown in Fig. 6. The α-cut for this film at $y = 0.6$ yields four points. The Hausdorff measure in (5) and (8) can so far only be applied when there are two distinct end points, and therefore cannot be applied to this set without loss of information, for example discarding two data points so that only two remain.

Firstly, to solve this problem, it needs to be decided what the result of the distance measure should be. For example, in Fig. 7, should the distance increase as the fuzzy set A becomes more concave, or should the distance decrease? It could be argued that since the similarity between the shapes of the two sets decreases, the distance should increase. The following proposes a method to solve this problem.

Fig. 7(e) shows two fuzzy sets, A and B. At $\alpha = 0.8$, $A_\alpha = [1.8, 2.6, 3.5, 4.3]$, and $B_\alpha = [6.8, 9.2]$. One method to calculate the distance at this α-cut is to split A_α into two intervals, $A_{\alpha 1}$ and $A_{\alpha 2}$. Next, we calculate the distance between $A_{\alpha 1}$ and B_α and the distance between $A_{\alpha 2}$ and B_α. Then, to reduce the resulting two distances to one value, a fair approach is to take their average. Thus, the α-cut at $\alpha = 0.8$, is calculated as follows.

First, A_α is split into two intervals, $A_{\alpha 1} = [1.8, 2.6]$ and $A_{\alpha 2} = [3.5, 4.3]$. Next, using (8), $h(A_{\alpha 1}, B_\alpha) = 6.6$ and $h(A_{\alpha 2}, B_\alpha) = 4.9$. Finally, the average of these is 5.75 and is used as the result of $h(A_\alpha, B_\alpha)$ at $\alpha = 0.8$.

To test this method, an experiment was carried out on the sets in Fig. 7 using this extension with RR's (6) and CR's (7) measures. The non-convex set (A) was given as the first parameter of the distance measure, and the convex set (B) was given as the second parameter; thus, each measure should result in a positive distance. The results of the experiment are displayed in Table V. These results show that as the fuzzy set A becomes increasingly concave, the result of the distance measure increases.

Using this method, it is now possible, for example, to compare the ratings of the film *All Dogs Go to Heaven 2* in Fig. 6 with any of the films in 4(b).
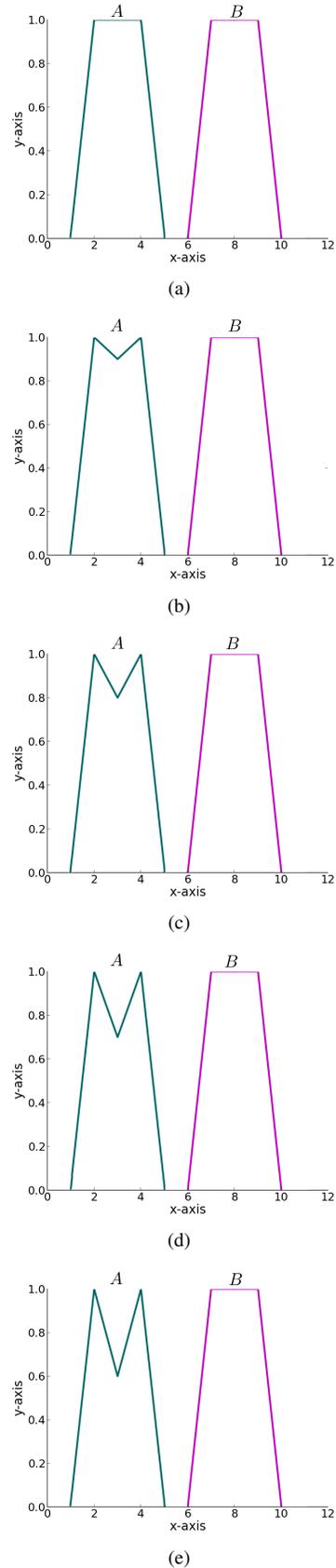

(a)

(b)

(c)

(d)

(e)

Fig. 7. Comparing distance between a non-convex fuzzy set and a convex fuzzy set.

TABLE V
RESULTS OF THE EXTENDED HAUSDORFF MEASURE APPLIED TO THE
FUZZY SETS IN FIG. 7; I.E. $d(A,B)$

	a	b	c	d	e
RR	4.99	5.061	5.141	5.227	5.317
CR	5.00	5.183	5.312	5.436	5.552

VII. CONCLUSIONS

In this paper we have introduced a distance measure for fuzzy sets which accounts for their direction of change, and have presented extensions of this distance measure for non-normal and non-convex fuzzy sets. We have demonstrated the applicability of the new distance measure using the MovieLens dataset and have asserted the advantage of using the new measure over distance measures which do not account for distance.

In the future, we plan to further develop and test the new distance measures, and implement extensions for interval and general type-2 fuzzy sets [15]. We will also apply the new distance measures within applications in CW, using fuzzy sets to construct word models of subjective information [16] and apply distance-based reasoning.

REFERENCES

[1] P. P. Bonissone, "A fuzzy sets based linguistic approach: Theory and applications," in *Proceedings of the 12th conference on Winter simulation*. IEEE Press, 1980, pp. 99–111.
[2] W. Wang and X. Xin, "Distance measure between intuitionistic fuzzy sets," *Pattern Recognition Letters*, vol. 26, no. 13, pp. 2063–2069, 2005.
[3] I. Turksen and Z. Zhong, "An approximate analogical reasoning approach based on similarity measures," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 6, pp. 1049–1056, 1988.
[4] L. Xuecheng, "Entropy, distance measure and similarity measure of fuzzy sets and their relations," *Fuzzy Sets and Systems*, vol. 52, no. 3, pp. 305–318, 1992.
[5] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338 – 353, 1965.
[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] S. Liao, T. Tang, and W.-Y. Liu, "Finding relevant sequences in time series containing crisp, interval, and fuzzy interval data," *IEEE Trans. Syst., Man, Cybern. B*, vol. 34, no. 5, pp. 2071 –2079, Oct. 2004.
[8] L. Zadeh, "Fuzzy logic = computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 103 –111, May 1996.
[9] J. Mendel, *Uncertain rule-based fuzzy logic systems: introduction and new directions*. Prentice Hall PTR, 2001.
[10] R. Zwick, E. Carlstein, and D. V. Budescu, "Measures of similarity among fuzzy concepts: A comparative analysis," *International Journal of Approximate Reasoning*, vol. 1, no. 2, pp. 221–242, 1987.
[11] 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.
[12] A. L. Ralescu and D. A. Ralescu, "Probability and fuzziness," *Information Sciences*, vol. 34, no. 2, pp. 85–92, 1984.
[13] B. Chaudhur and A. Rosenfeld, "On a metric distance between fuzzy sets," *Pattern Recognition Letters*, vol. 17, no. 11, pp. 1157–1160, 1996.
[14] J.-l. Fan, "Note on Hausdorff-like metrics for fuzzy sets," *Pattern Recognition Letters*, vol. 19, no. 9, pp. 793–796, 1998.
[15] J. McCulloch, C. Wagner, and U. Aickelin, "Extending similarity mesaures of interval type-2 fuzzy sets to general type-2 fuzzy sets," in *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, 2013.
[16] S. Miller, C. Wagner, J. Garibaldi, and S. Appleby, "Constructing General Type-2 fuzzy sets from interval-valued data," in *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, 2012, pp. 1–8.

APPENDIX

NUMERICAL EXAMPLE ON NON-NORMAL FUZZY SETS

The proposed direction-based distance measure using (12) and (8) is demonstrated using the non-normal fuzzy sets SMB and SW in Fig. 5. The fuzzy sets for SMB and SW are distributed as follows:

$$SMB = 0.385/1, 0.269/2, 0.231/3, 0.115/4, 0.0/5$$
$$SW \;= \;0.015/1, 0.027/2, 0.098/3, 0.302/4, 0.557/5$$

Using linear interpolation, the α-cuts of each film, given in the format $\alpha/[x_l, x_r]$, are:

$$SMB = 0.1/[1.0, 4.0], 0.2/[1.0, 4.0], 0.3/[1.0, 1.0]$$
$$SW \;= \;0.1/[2.44, 5.0], 0.2/[3.08, 5.0], 0.3/[3.36, 5.0],$$
$$0.4/[3.64, 5.0], 0.5/[3.92, 5.0]$$

The α-cut at 0.0 has been disregarded because it does not contribute to (12). The distance measure using (12) and (8) is calculated as follows where $\lambda = 0.5$:

$$\text{At } \alpha = 0.1, h(SMB_\alpha, SW_\alpha) = 1.44$$
$$\text{At } \alpha = 0.2, h(SMB_\alpha, SW_\alpha) = 2.08$$
$$\text{At } \alpha = 0.3, h(SMB_\alpha, SW_\alpha) = 2.36$$
$$\text{At } \alpha = 0.4, h(SMB_\alpha, SW_\alpha)$$
$$= h(A_{\alpha_k}, B_{\alpha_k})$$
$$= 2.36$$
$$\text{At } \alpha = 0.5, h(SMB_\alpha, SW_\alpha)$$
$$= h(A_{\alpha_k}, B_{\alpha_k})$$
$$= 2.36$$

Finally, combining these results in (12) gives

$$d(A,B)$$
$$= \frac{0.1 \times 1.44 + 0.2 \times 2.08 + 0.3 \times 2.36 + 0.4 \times 2.36 + 0.5 \times 2.36}{0.1 + 0.2 + 0.3 + 0.4 + 0.5}$$
$$= \frac{3.392}{1.5}$$
$$= 2.261$$