Fuzzy Set-Theoretical Approach for Comparing Objects with Fuzzy Attributes

Y. Bashon, D. Neagu, M.J. Ridley
Department of Computing
University of Bradford
Bradford, BD7 1DP, UK
e-mail: {Y.Bashon, D.Neagu, M.J.Ridley}@bradford.ac.uk

Abstract—**In this paper we develop the similarity measure introduced by the *Tversky contrast model* and apply it on fuzzy sets using the cardinality of fuzzy sets and their operations. Based on this extended similarity definition we propose a new approach for comparing fuzzy objects and discuss some properties of the new similarity model. Some experimental examples are given to show the effectiveness of using this model against different cases. This work provides a method to compare objects with vague/fuzzy content and support further development of (fuzzy) data mining algorithms.**

Keywords- Similarity measure, Tversky contrast model, fuzzy attributes, fuzzy objects.

I. INTRODUCTION

Several measures of similarity among fuzzy sets have been proposed in the literature as reported in [1], [2] and [3]. The motivation behind these measures is both *geometric* and *set-theoretic*. Geometric models dominate the theoretical analysis of similarity measures [1]. Objects in these models are represented as points in a coordinate space, and the metric distance between the respective points is considered to be the dissimilarity among objects. In most cases the Euclidean distance is used to define the dissimilarity between two concepts or objects. In the set-theoretic approaches a different model is used, which is based on the concept of a non-dimensional and non-metric similarity relation [4].

Bouchon-Meunier et al. [5] conducted a study based on Tversky's feature-theoretical concepts on similarities [4] and [6]. The research proposes classification of measures that exist or have been used in previous literature to compare fuzzy characterization of objects according to their properties and their applications. The study focused on finding differences between various measures of comparisons including satisfiability, resemblance, inclusion, and dissimilarity.

Our purpose in this paper is to introduce a new family of similarity measures, used to compare fuzzy objects based on fuzzy-set-theoretical concepts, which tracked and extended the work reported in [5]. We also extend the approach to compare sets of features developed by Tversky [4] to a fuzzy measure of the similarity between fuzzy objects as detailed below.

In this approach, we will consider as fuzzy objects those objects having imprecise and uncertain attributes/features. Attributes of fuzzy objects are represented by fuzzy sets and we will make use of fuzzy set operations for processing them. Our work is also a generalization for cases where values of an object attribute are crisp values; the crisp sets are considered as particular cases of fuzzy sets that represent precise and certain attributes/features. In other words, we propose a framework where each object is characterized by a set of attributes that can be given either crisp or fuzzy value.

The structure of the paper is as follows: the next section presents the notions we are making use of. In section three, cardinality and operations of fuzzy sets are presented and studied in detail in order to define the generalised similarity measure between fuzzy sets. The similarity measure of fuzzy attributes is defined in section four using the similarity definition presented in section four. A set of aggregation operators, defined in order to calculate the similarity between fuzzy objects, is proposed in this section as well. We finally illustrate our discussion by experimental examples using this similarity approach in section five. The paper ends with conclusions and future work.

II. BASIC NOTIONS

In this section we summarize the notions which will be needed in the next sections. Let U denote the universe of discourse; $C(U)$ denotes the set of crisp subsets of U and $F(U)$ denotes the set of fuzzy subsets of U (a fuzzy set A is a subset of U that is characterized by a membership function $\mu_A: U \to [0,1]$). Two crisp sets $core(A) = \{x \in U \mid \mu_A(x) = 1\}$ and $supp(A) = \{x \in U \mid \mu_A(x) > 0\}$ core and support of A respectively, are important when characterizing the fuzzy set A. Thus, a fuzzy subset A in $F(U)$ is said that it is finite fuzzy set (*ffs)* if $supp(A)$ is a finite crisp subset of U. Accordingly, a fuzzy set A can be written as a set $A = \{(\mu_A(x)/x) \mid \forall x \in supp(A)\}$. A fuzzy singleton or a singleton is *ffs* over U, denoted by a/x for $a \in [0,1]$ and $x \epsilon U$ such that $a/x(x) = a$ and $a/x(y) = 0$ if $y \neq x$. Let the sum of a family $(A_e)_{e \in J}$ be denoted by $\bigcup_{e \in J} A_e$, where $J \neq \emptyset$ is a finite set of indices. Then $A = \bigcup_{x \in supp(A)} \mu_A(x)/x$ with a/x a singleton supported by x.

Operations between fuzzy sets $A \cup B$ and $A \cap B$ are performed in terms of their membership functions using Zadeh's definitions of union and intersection of two fuzzy sets [7], [8]: $\mu_{A\cup B} = \max_{x\in U}(\mu_A(x), \mu_B(x))$ and $\mu_{A\cap B} = \min_{x\in U}(\mu_A(x), \mu_B(x))$ respectively. The standard complement of A is denoted by A' where $\mu_{A'}(x) = 1 - \mu_A(x)$. Finally, the inclusion $A \subseteq B$ means that A is included in B (i.e. $\mu_A(x) \le \mu_B(x)\ \forall x \in U$); thus the equality of two fuzzy sets A and B is: $A = B \Leftrightarrow (A \subseteq B \text{ and } B \subseteq A)$ i.e. $\mu_A(x) = \mu_B(x) \Leftrightarrow \big(\mu_A(x) \le \mu_B(x) \text{ and } \mu_B(x) \le \mu_A(x)\big)$ $\forall x \in U$.

III. Cardinality-Based Similarity Measurements

Since our goal is to find a formal definition of the perceptive concept of similarity between objects with fuzzy attributes, we clarify the perception that the comparison of any two fuzzy sets A and B defined on a given universe of discourse U is related on one hand to their commonality (the elements of the universe which belong to both of them), and on the other hand to the difference between them (the elements belonging to A but not to B, and conversely). Therefore, the more commonality they share, the more similar they are, and the more differences they have, the less similar they are. This is the same intuition about similarity presented by Dekang Lin in [9].

The following parameterized ratio model of similarity was proposed by Tversky for comparing two objects or concepts with sets of features or attributes [4].

$$s(o_1, o_2) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)} \tag{1}$$

This similarity definition is a normalised function; in Tversky's model f is considered to be a feature/attribute-matching function that measures the degree to which two sets of attribute's values match each other rather than just measuring the distance between two points in the attribute's domain space such that f should satisfy $f(A \cup B) = f(A) + f(B)$, where A and B are disjoint crisp sets.

Nevertheless, in this paper we use the model stated above to calculate the similarity between two fuzzy attributes (attributes characterised by fuzzy sets) and then, we have used an aggregation function to compute the similarity between two objects described by such fuzzy attributes as detailed below. In the following section, we are going to use fuzzy logic and fuzzy set theory to extend the domain of applicability of the generalised Tversky's model.

A. *Scalar Cardinality of a Fuzzy Set*

In this section, we have defined the similarity between two fuzzy sets as in [5] by a mapping $s: F(U) \times F(U) \to [0,1]$ such that $s(A,B) = F_s(f(A \cap B), f(A - B), f(B - A))$ where $F_s: \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \to [0,1]$ is defined by the generalised Tversky's model and f is defined by eqs. (2) and (3) below.

In our approach, we define the cardinality of a fuzzy set for a finite universe U as a mapping $f: F(U) \to \mathbb{R}^+$ that assigns to each finite fuzzy set a single ordinary cardinal number (or non-negative real number). It can be defined as the sum of membership values that characterize a fuzzy set $A \epsilon F(U)$ [14, 15]: for finite universe U,

$$f(A) = card(A) = |A| = \sum_{x\epsilon U} \mu_A(x) \tag{2}$$

for a discrete fuzzy set A and:

$$f(A) = \int_{x\epsilon U} \mu_A(x)\, dx \tag{3}$$

for a continuous fuzzy set A.

In [10], [11] and [12] Wygralak presented the complete axiomatic theory of scalar cardinality of fuzzy sets which our work is based on. The cardinality of fuzzy sets is also understood as a convex fuzzy set of the set of natural numbers $\mathbb{N}$. The fuzzy approach with its axiomatic theory was introduced by Casanovas and Torrens in [13]. There are other approaches which define fuzzy cardinalities as fuzzy quantities (see e.g. Dubois and Prade [14], Ralescu [15]). However, in this paper we focus on using definition (2) as it is easier in calculation than the integral form.

Definition 1. A function $\sigma: F(U) \to \mathbb{R}^+ \cup \{0\}$ is called a *scalar cardinality* if the following axioms are satisfied for each $a, b \epsilon [0,1]$, $x, y \in U$, and $A, B \epsilon F(U)$ [10]:

axiom 1) $\sigma(1/x) = 1$,

axiom 2) if $a \le b$, then $\sigma(a/x) \le \sigma(b/x)$,

axiom3) if $A \cap B = \emptyset$, then $\sigma(A \cup B) = \sigma(A) + \sigma(B)$.

Proposition 1. Let $A \in F(U)$. Then the function $f(A)$ defined by (2) is a scalar cardinality of A.

Proof. Since $1/x$ for some x is a singleton supported by an element x and has a membership degree 1, then $f(1/x) = |1/x| = \sum \mu_A(x) = 0 + \cdots 0 + 1 + 0 + \cdots + 0 = 1$, $x \in U$. As a consequence of summation properties, and since $a \le b$, we get $f(a/x) = \sum_{x\epsilon U} \mu_A(x) = \sum_{x\epsilon U} a \le \sum_{x\epsilon U} b = f(b/x)$. For disjoint A and B: $f(A \cup B) = \sum_{x\epsilon U} \mu_{A\cup B}(x) = \sum_{x\epsilon U} \mu_A(x) + \sum_{x\epsilon U} \mu_B(x) = f(A) + f(B)$.

Theorem 1. Let $A, B \in F(U)$ and $A_e \in F(U)$ for each $e \in J$. The following properties are then satisfied by function f:

a) $f(\bigcup_{e\in J} A_e) = \sum_{e\in J} f(A_e)$ if $A_e \cap A_{e'} = \emptyset$ for each $e \ne e'$ (finite additivity)
b) $A \in C(U) \Longrightarrow f(A) = |supp(A)|$ (coincidence)
c) $A \subseteq B \Longrightarrow f(A) \le f(B)$ (monotonicity)
d) $|core(A)| \le f(A) \le |supp(A)|$ (boundedness)
e) $f(A) + f(B) = f(A \cup B) + f(A \cap B)$ (valuation)
f) $f(\bigcup_{e\in J} A_e) \le \sum_{e\in J} f(A_e)$ (finite subadditivity)

Proof. It is obvious that *a*) follows from axiom 3, where *b)* is a consequence of *a)* and axiom 1. For *c)*, from *a)* and axiom 2 and $A \subseteq B \implies supp(A) \subseteq supp(B) \implies A \cap B = A \neq \emptyset$ and then $|supp(A)| \leq |supp(B)| \implies f(A) = \sum_{x\epsilon supp(A)} \mu_A(x) \leq f(B)$. From *b)* and *c)* and since $core(A) \subseteq A \subseteq supp(A)$, we will get d). As a consequence of *a)* and Axiom 3 and the definition of $A \cup B$ and $A \cap B$: $f(A \cup B) + f(A \cap B) = \sum_{x\epsilon U} \mu_{A\cup B}(x) + \sum_{x\epsilon U} \mu_{A\cap B}(x) = \sum_{x\epsilon U}[\max_{x\epsilon U}(\mu_A(x), \mu_B(x))] + \sum_{x\epsilon U}[\min_{x\epsilon U}(\mu_A(x), \mu_B(x))] = \sum_{x\epsilon U}\left[\max_{x\epsilon U}(\mu_A(x), \mu_B(x)) + \min_{x\epsilon U}(\mu_A(x), \mu_B(x))\right] = f(A) + f(B)$. From e) since $f(A) + f(B) = f(A \cup B) + f(A \cap B)$, then $f(A \cup B) \leq f(A) + f(B)$ and therefore property f) is justified.

B. Fuzzy Set Operations

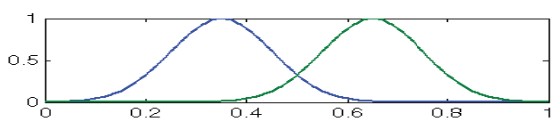
The intersection and the difference between two fuzzy sets are defined in terms of their membership functions to describe the elements belonging to *A* and *B*, and the elements belonging to only one of them. Here we are going to apply the usual definition of intersection $A \cap B$ as below:

$$\mu_{A\cap B} = \min_{x\epsilon U}(\mu_A(x), \mu_B(x)) \tag{4}$$

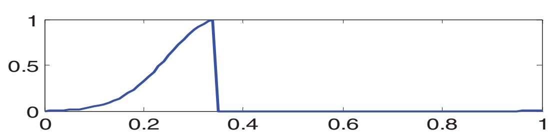
Two other examples of difference operations presented in [5] and [7] can also be listed here:

$$\mu_{A-_1B}(x) = \begin{cases} \mu_A(x) & if \;\; \mu_B(x) = 0 \\ 0 & if \;\; \mu_B(x) > 0 \end{cases}. \tag{5}$$

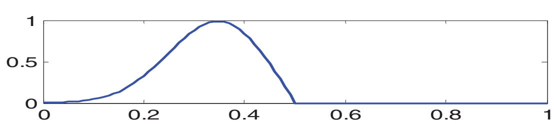
$$\mu_{A-_2B}(x) = \max_{x\epsilon U}(0, \mu_A(x) - \mu_B(x)) \tag{6}$$



(a) Two fuzzy sets *A* and *B*



(b) $A{-_1}B$



(c) $A{-_2}B$

Fig. 1 Computing the fuzzy difference between two fuzzy sets using various definitions.

According to the definitions above for the difference operation, we observe that $\mu_{A-_1B}(x) < \mu_{A-_2B}(x)$ for some $x\epsilon U$. Consequently, the similarity degree between two fuzzy sets will be different according to the difference operation that is used in its definition (1). Fig. 1 shows an example of using the difference operations that are stated above. We prefer to use the difference operation $-_2$ within the definition because it gives us the opportunity to get more information about uncertainty rather than using the other operations: it reflects more qualitative information about elements in A but not in B.

Proposition 2. For any two fuzzy *sub*sets $A, B \in F(U)$, the difference operation $-_2$ on $F(U)$ satisfies the properties:

1) if $A \subseteq B$, then $A{-_2}B = \phi$,
2) if $A \subseteq A'$, then $A{-_2}B \subseteq A'{-_2}B$. (monotonicity w.r.t. A)

Proof: For 1: since $\subseteq B \Leftrightarrow \mu_A(x) \leq \mu_B(x) \;\; \forall x \in U$, then $\mu_A(x){-}\mu_B(x) \leq 0$, and from (4) we define $A{-_2}B$ as $\mu_{A-_2B}(x) = \max(0, \mu_A(x) - \mu_B(x))$; $\forall x\epsilon U \Rightarrow \mu_A(x) - \mu_B(x){=}0 \Rightarrow A{-_2}B = \phi$. For 2: we have $A \subseteq A' \Rightarrow \mu_A(x) \leq \mu_{A'}(x) \Rightarrow \mu_A(x) - \mu_B(x) \leq \mu_{A'}(x) - \mu_B(x) \Rightarrow \max(0, \mu_A(x), \mu_B(x)) \leq \max\left(0, \mu_{A'}(x), \mu_B(x)\right) \Rightarrow \mu_{A-_2B}(x) \leq \mu_{A'-_2B}(x) \Rightarrow A{-_2}B \subseteq A'{-_2}B$.

Now, we are going to employ the cardinalities of the intersection and the difference between two fuzzy sets. This enables us to evaluate the influence of the part of the universe that is common to any two fuzzy subsets $A, B \in F(U)$ if $f(A \cap B) = card(A \cap B) = |A \cap B|$. Also, we evaluate the influence of the part that belongs to *A* but not to *B*, and conversely if we consider $f(A - B)$ and $f(B - A)$ respectively. Thus, from (2), (4) and (6), we get:

$$f(A \cap B) = card(A \cap B) = |A \cap B| \tag{7}$$

$$= \sum_{x\epsilon U} \mu_{A\cap B}(x) = \sum_{x\epsilon U}\left[\min_{x\epsilon U}(\mu_A(x), \mu_B(x))\right]$$

and

$$f(A - B) = card(A{-_2}B) = |A{-_2}B|$$

$$= \sum_{x\epsilon U} \mu_{A-_2B}(x) = \sum_{x\epsilon U} \max_{x\epsilon U}(0, \mu_A(x) - \mu_B(x)) \tag{8}$$

We can rewrite the equation (1) in order to measure the similarity between two fuzzy sets as follows:

$$s(A, B) = \frac{\alpha|A \cap B|}{\alpha|A \cap B| + \beta|A - B| + \gamma|B - A|} \tag{9}$$

for $\alpha > 0$ and $\beta, \gamma \geq 0$, where we add a new parameter α to contribute with the common part of the fuzzy sets.

IV. Fuzzy Objects Comparison

In the previous section we introduced the method to compare two fuzzy sets. Let us now apply it to see how to compare two objects whose attributes have fuzzy values characterised by using fuzzy sets.

A. Similarity Measure of Fuzzy Attributes

Suppose that o_1 and o_2 are two fuzzy objects of the same class and let $At_{o_1} = \{a_1, a_2, \dots, a_n\}$ and $At_{o_2} = \{b_1, b_2, \dots, b_n\}$ denote the sets of n attributes for each object, respectively and n stands for the number of attributes for each object. Basically, for each pair of corresponding attributes, a basic domain (a universe of discourse) should be defined in which a fuzzy domain can be built over. The fuzzy domains are defined in order to represent fuzzy object attributes. Each attribute is given a fuzzy value which is either represented as:

1) a *set* of fuzzy subsets of the universe U as follows: $a_i = \{A_{i1}, A_{i2}, \dots, A_{im_i}\}$ and $b_i = \{B_{i1}, B_{i2}, \dots, B_{im_i}\}$ where m_i stands for the number for fuzzy subsets describing the i^{th} attribute and $i = 1,2, \dots, n$. For example, the age of a person: $Age = \{(0.2/young), (0.75/middelage)\}$, or as

2) a fuzzy value that is characterised by a unique fuzzy subset: $a_i = \{A_{im_{a_i}}\}$, and $b_i = \{B_{im_{b_i}}\}$ where $m_{a_i}, m_{b_i} \in \{1,2, \dots, m_i\}$. For example, Age is *young*, where Age is a person age domain (an attribute domain) variable and *young* is a fuzzy value (an attribute value).

The first situation has been addressed in our previous work [16] where we proposed a family of similarity measures for the problem of fuzzy object comparison where a geometric model has been used for comparing this type of attribute's values. The Euclidean distance is used to calculate the dissimilarity between fuzzy sets that describe those attributes of fuzzy objects. In this paper we focus on the second situation; when each object attribute is described using a single fuzzy set and we propose another family of similarity measures based on the concepts of fuzzy set-theory and cardinality of fuzzy sets presented above.

Definition 2. A similarity measure between two attributes is a mapping $S: At_{o_1} \times At_{o_2} \to [0,1]$ such that $S(a_i, b_i) = s\left(A_{im_{a_i}}, B_{im_{b_i}}\right)$, for a given mapping $s: F(U_i) \times F(U_i) \to [0,1]$ that is defined by the equation (9). Thus:

$$S(a_i, b_i) = \frac{\alpha\left|A_{im_{a_i}} \cap B_{im_{b_i}}\right|}{\alpha\left|A_{im_{a_i}} \cap B_{im_{b_i}}\right| + \beta\left|A_{im_{a_i}} - B_{im_{b_i}}\right| + \gamma\left|B_{im_{b_i}} - A_{im_{a_i}}\right|} \quad (10)$$

for $\alpha > 0$ and $\beta, \gamma \geq 0$, where a_i and b_i are two attributes of two fuzzy objects o_1 and o_2, respectively, and U_i is the domain of i^{th} attribute. Using equation (10) allows us to determine to what extend the attributes of two objects have common points, or are different from each other. However, the validation of this similarity relation should be examined.

Now, is the similarity model maximal? (i.e., $S(a, b) \leq 1$ and $S(a, b) = 1$ if and only if $a = b$). Let us consider two objects having attributes characterised with the same fuzzy value, say for example, *Tom* is 18 years old and *John* is 27 years old. We can notice that both of them are young, so, $S(TomAge, JohnAge) = s(young, young) = 1$. But this result contradicts the crisp approach. However, it is not necessary that fuzzy perception should be practically the same as crisp perception. In our approach, two fuzzy attributes are considered to be identical if they have the same fuzzy value, even if their crisp values are different. Conversely, two different fuzzy values may be given to the same attribute of the compared objects, for example, this happens when *John* is considered to be *young* and *David* is a *middle-aged* person, at the time they are both 27 years old. So, $S(JohnAge, DavidAge) = s(young, middle-aged) \neq 1$. Consequently the assumption that S is maximal is inequitable in some cases as shown in the examples.

For symmetry, we notice that $S(a, b) \neq S(b, a)$ since $s(A, B) \neq s(B, A)$ when $\beta \neq \gamma$. The symmetry of s is satisfied only when $\beta = \gamma$ or if $f(A - B) = f(B - A)$. We further proved that function f used in the model $S(A, B) = f(A \cap B, A - B, B - A)$ is non-decreasing with respect to $A \cap B$ and non-increasing with respect to $A - B$ and $B - A$.

From these justifications, it seems that the geometric approach faces several difficulties when dealing with fuzzy data. For similarity analysis, the applicability of the dimensional hypothesis is limited, and the metric axioms are doubtful. Specifically, maximality is somewhat problematic, symmetry appears to be false in most cases, and the triangle inequality is hardly compelling as pointed out in [4].

B. Aggregation Operators for Fuzzy Object Comparison

Assume that we have a set of *m* fuzzy objects of the same class $O_F = \{o_1, o_2, \dots, o_m\}$. Each object is described by a set of *n* fuzzy attributes. In this section we define the similarity measure between any two fuzzy objects:

Definition 3. A similarity measure between two fuzzy objects $o_s, o_t \in O_F$ is a mapping $Sim: O_F \times O_F \to [0,1]$:

$$Sim(o_s, o_t) = \otimes_{Sim}\big(S(a_1, b_1), S(a_2, b_2), \dots, S(a_n, b_n)\big) \quad (11)$$

where $\otimes_{sim}: [0,1]^n \to [0,1]$ is an aggregation operation that is defined in [16]. Thus, we have chosen the following definitions among others for comparing their performance to compute the overall similarity between the objects:

1) Weighted average of the similarities of attributes (WA): Let us consider that each attribute a_i has an associated weight w_i that points out the importance that the similarity in this attribute must have when computing the similarity degree between objects of the same class. We are going to consider that $\forall i; w_i \epsilon [0,1]$:

$$Sim(o_s, o_t) = \frac{\sum_{i=1}^{n} w_i S(a_i, b_i)}{\sum_{i=1}^{n} w_i}; \quad w_j \epsilon [0,1] \quad (12)$$

2) Minimum of the similarities among attributes (Min):

$$Sim(o_s, o_t) = \min[S(a_1, b_1), S(a_2, b_2), \dots, S(a_n, b_n)] \quad (13)$$

3) The similarity ratio for the similarities among attributes (Min/Max):

$$Sim(o_s, o_t) = \frac{\min[S(a_1, b_1), S(a_2, b_2), \dots, S(a_n, b_n)]}{\max[S(a_1, b_1), S(a_2, b_2), \dots, S(a_n, b_n)]} \quad (14)$$

V. EXAMPLES AND RESULTS

For this section, we developed a MATLAB program to study the proposed method for comparing fuzzy objects. MATLAB fuzzy toolbox membership functions are used for building the fuzzy domain for each fuzzy object attribute.

Given a list of 36 student rooms described by their *quality*, *price* and distance from University *dfUni* (see Table I) we want to asses which room is closer to *a particular target described by "high quality, moderate price and close to the University"*. To achieve that, we follow the steps:

1) define the basic domain for each fuzzy attribute: For each room let $D_Q = [0,1]$ be the basic quality domain, $D_P = [0,600]$ be the basic price domain and $D_{dfUni} = [0,10]$. The units are specified as *per-cent*, *British pound*(s) and *mile*(s) respectively.

2) define fuzzy domain for each attribute by using fuzzy sets built over the attribute basic domain: We determined a fuzzy domain of room quality by defining three fuzzy subsets $FD_Q = \{low, average, high\}$ over the domain D_Q. The fuzzy domain for *price* is defined as $FD_P = \{cheap, moderate, expensive\}$. Finally, the fuzzy domain for *dfUni* is defined with two fuzzy subsets $FD_{dfUni} = \{close\_to, far\_form\}$ (see Fig. 2).

TABLE I. DUMMY DATASET OF STUDENT ROOMS

Student Rooms	Room Attributes		
	quality	*price*	*dfUni*
room1	high	expensive	close to
room2	average	expensive	close to
room3	low	Cheap	close to
room4	average	Cheap	close to
room5	high	moderate	close to
room6	low	expensive	close to
room7	low	moderate	close to
room8	high	expensive	close to
room9	low	Cheap	close to
room10	average	moderate	close to
room11	average	expensive	far from
room12	low	Cheap	far from
room13	high	expensive	far from
room14	average	Cheap	far from
room15	low	moderate	far from
room16	high	moderate	far from
room17	low	Cheap	far from
room18	average	Cheap	far from
room19	low	moderate	far from
room20	high	Cheap	far from
room21	low	Cheap	far from
room22	average	moderate	far from
room23	low	moderate	far from
room24	high	moderate	far from
room25	average	Cheap	far from
room26	average	Cheap	close to
room27	low	moderate	close to
room28	high	moderate	far from
room29	low	Cheap	far from
room30	average	expensive	close to
room31	low	Cheap	close to
room32	high	expensive	far from
room33	average	Cheap	close to
room34	low	moderate	close to
room35	high	moderate	far from
room36	low	Cheap	close to

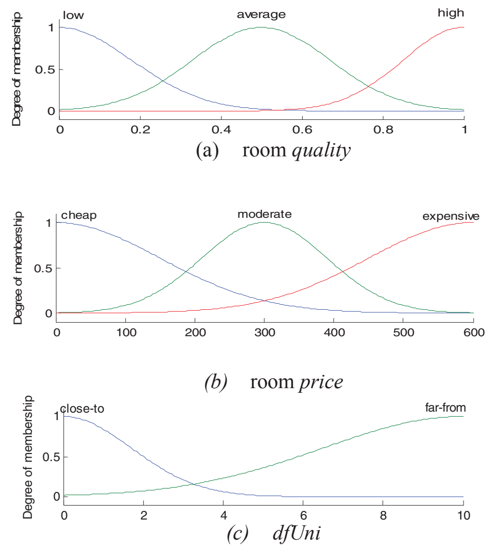


Fig 2. Fuzzy representation for (a) room *quality*, (b) room *price* and (c) distance from University (*dfUni*).

3) calculate the similarity among the corresponding attributes: Similarity matrices of the room attributes are calculated using (10) and by assuming $\alpha = \beta = \gamma = 1$, we get the result shown in the tables below. Accordingly, we have obtained similarity degrees between the requested room attributes and corresponding student rooms attributes in the above dataset (see Table I).

4) aggregate or calculate the average over all similarities: we used (12), (13) and (14) for calculating similarity values among the given rooms as shown in Table I in order to produce the final judgement to how similar the requested room and other student rooms are. Fig. 3 shows the results in each case. Of course, different weights or importance values are given to each attribute by using (12), for example, Fig. 3(a) shows the similarity results when$w_{quality} = w_{price}$ =w_{dfUni} =1; in Fig. 3(b) $w_{quality} = 0.8$, w_{price} =0.5 and $w_{dfUni} = 1$.

Obviously, more weights can be considered based on how the attributes can be determined (or which attribute is more significant). Similarity results are affected by different factors such as fuzzy representations of objects attributes, parameters α, β and γ in (10) and attributes weights in (12).

TABLE II. SIMILARITY MATRIX OF THE FUZZY ATTRIBUTE ***QUALITY***

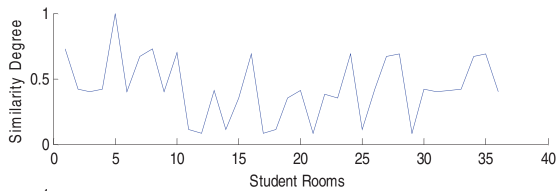
s	*low*	*average*	*high*
low	1.0000	0.1130	0.0024
average	0.1130	1.0000	0.0818
high	0.0024	0.0818	1,0000

TABLE III. SIMILARITY MATRIX OF THE FUZZY ATTRIBUTE ***PRICE***

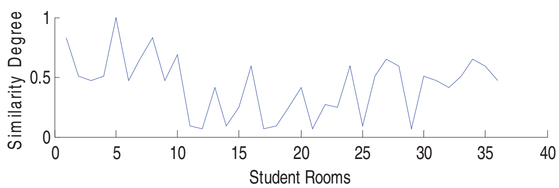
s	*cheap*	*moderate*	*expensive*
cheap	1.0000	0.1813	0.0476
moderate	0.1813	1.0000	0.1813
expensive	0.0476	0.1813	1,0000

TABLE IV. SIMILARITY MATRIX OF THE Fuzzy ATTRIBUTE ***DFUNI***

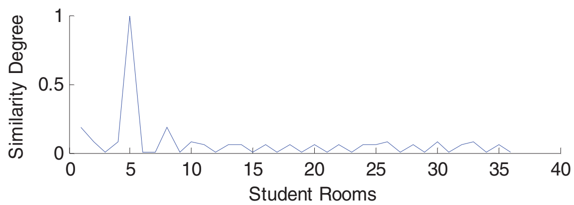
s	*close-to*	*far-from*
close-to	1.0000	0.0545
far-from	0.0545	1.0000



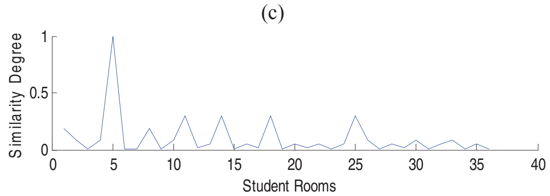
(a)



(b)



(c)



(d)

Fig. 3 Similarity degrees of the requested room and other student rooms (a) and (b) for *WA*, (c) for *Min* and (d) for *Min/Max*.

VI. CONCLUSIONS AND FURTHER WORK

In this paper, we studied the expression of similarity measure introduced by the *Tversky contrast model*, and generalised this similarity measure on fuzzy sets using the cardinality of fuzzy sets and their operations. We also emphasize the use of fuzzy difference for this aim. Taking the advantage of this similarity definition we propose a set/family of measures in order to compare fuzzy objects.

We investigated the performance and effectiveness of our approach with some experimental cases from a dummy fuzzy dataset. However the set of similarity measures that we presented can be applied in more general situations. Actually, given this set of similarity measures, there is a question that may come to the reader's mind: *Which similarity measure is most appropriate for my data mining task*? Our experimental results illustrates that *WA operator* performs better than *Min* and *Min/Max* according to the fuzzy representation of the given dataset and the given attributes weights. However, there is no one best performing similarity measure across all cases, but there may be a data driven process of choosing a suitable measure according to the exercise at hand.

Hence, one needs understanding of how a similarity measure handles the different characteristics that represent a fuzzy data set, and this certainly needs to be investigated with further research. It may be possible to construct measures that draw on the strengths of a number of measures in order to obtain superior performance or replace the difference operation used in our approach by one of those presented above for deducing new quantities (for example, difference operation $-_1$) and develop our approach in some real-world applications that belong to data mining or to information retrieval. This is also an aspect of this work that will be pursued in our future work.

ACKNOWLEDGMENT

This work is partially supported by BBSRC, TSB and Syngenta through the KTP Grant scheme.

REFERENCES

[1] R. Zwick, E. Carlstein, and D. Budescu, "Measures of similarity among fuzzy concepts: A comparative analysis," *lnternat. J. Approximate Reasoning*, vol. 1, no. 2, pp. 221-242, 1987.

[2] V. Cross, and T. Sudkamp, Similarity and compatibility in fuzzy set theory: *Assessment and Applications*: Springer, 2002.

[3] S. Santini, and R. Jain, "Similarity measures," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 21, no. 9, pp. 871-883, 2002.

[4] A. Tversky, "Features of similarity," *Psychological Review,* vol. 84, no. 4, pp. 327-352, 1977.

[5] B. Bouchon-Meunier, M. Rifqi, and S. Bothorel, "Towards general measures of comparison of objects," *Fuzzy sets and systems*, vol. 84, no. 2, pp. 143-153, 1996.

[6] A. Tversky, and I. Gati, "Studies of similarity," *Cognition and categorization*, vol. 1, pp. 79-98, 1978.

[7] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338-353, 1965.

[8] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning--I* 1," *Information Sciences,* vol. 8, no. 3, pp. 199-249, 1975.

[9] D. Lin, "An information-theoretic definition of similarity," *ICML '98 Proceedings of the Fifteenth International Conference on Machine Learning*, 1998.pp. 296–304, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA

[10] M. Wygralak, *Cardinalities of fuzzy sets*: Springer Verlag, 2003.

[11] M. Wygralak, "Fuzzy sets with triangular norms and their cardinality theory," *Fuzzy sets and systems*, vol. 124, no. 1, pp. 1-24, 2001.

[12] M. Wygralak, "A generalizing look at sigma counts of fuzzy sets," *Uncertainty in Intelligent and Information Systems*, 2000, pp. 34-45.

[13] J. Casasnovas, and J. Torrens, "An axiomatic approach to fuzzy cardinalities of finite fuzzy sets* 1," *Fuzzy sets and systems*, vol. 133, no. 2, pp. 193-209, 2003.

[14] D. Dubois, H. Prade, Fuzzy cardinality and the modelling of imprecise quantification, *Fuzzy sets and systems*, vol. 16, no 3, 1985, pp. 199–230.

[15] D. Ralescu, Cardinality, quantifiers and the aggregation of fuzzy criteria, *Fuzzy sets and systems*, vol. 69, no 3, 1995, pp. 355–365.

[16] Y. Bashon, D. Neagu, and M. Ridley, "A New Approach for Comparing Fuzzy Objects," *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications*, 13th International Conference, IPMU 2010, Dortmund, Germany, June 28–July 2, 2010. Proceedings, Part II, 2010, pp. 115-125.