Available online at www.sciencedirect.com

SciVerse ScienceDirect

Fuzzy Sets and Systems 208 (2012) 79–94

FUZZY sets and systems

www.elsevier.com/locate/fss

A distance metric for a space of linguistic summaries

Anna Wilbik[a,b], James M. Keller[a,*]

[a] *Electrical and Computer Engineering, University of Missouri, 307 Engineering Building West Columbia, MO, USA*
[b] *Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland*

Received 16 July 2011; received in revised form 14 March 2012; accepted 17 March 2012
Available online 27 March 2012

Abstract

Producing linguistic summaries of large databases or temporal sequences of measurements is an endeavor that is receiving increasing attention. These summaries can be used in a continuous monitoring situation, like eldercare, where it is important to ascertain if the current summaries represent an abnormal condition. It is therefore necessary to compute the distance between summaries as a basis for such a determination. In this paper, we propose a dissimilarity measure between summaries based on fuzzy protoforms, and prove that this measure is a metric. We take into account not only the linguistic meaning of the summaries, but also two quality evaluations, namely the truth values and the degrees of focus. We present examples of how the distance metric behaves and show that it corresponds with intuition.

Keywords: Linguistic modeling; Linguistic summarization; Distance metric

1. Introduction

As data sets from sensors and other sources become increasingly large, methods to summarize the data and to analyze these summaries are more and more important. Acknowledging the uncertainty in summarizing large amount of data, the fuzzy set community has investigated several approaches for linguistic summarization [1–4,15,25–27,31]. Many times, summaries are produced with the understanding that people will read them and take appropriate action. For example, summaries of nighttime motion and restless of an elder resident can offer indications to the health care professionals of potential abnormal conditions [30]. However, as the nature and magnitude of sensor and other data collection grows, so does the complexity and quantity of the set of linguistic descriptions. Hence, it is necessary to perform some automated analysis of this condensed information. In order to make decisions such as “today is normal for Mr. Smith” or “Mrs. Jones is much better this week than she was last week”, some form of distance is required between, in the above case, temporal summaries of sensor data. We can even consider clustering summaries instead of the raw or processed data, as was done in [28]. In the case of static databases, for instance, information about employees of companies, we might want to determine that company A is closer to company B than to company C with respect to employee makeup and benefits (as derived though summarization). While there are myriad of applications of linguistic summarization, many focus on time series analysis. In [4], such temporal linguistic summarization

[*] Corresponding author. Tel.: +1 573 882 7339; fax: +1 573 882 0397.
E-mail addresses: wilbik@ibspan.waw.pl (A. Wilbik), kellerj@missouri.edu (J.M. Keller).

0165-0114/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fss.2012.03.010

was applied to patient inflow in a medical center. Both [5,18] consider comparison of time series from a summarization standpoint. The latter approach, for example, was applied to the analysis of investment fund quotations. Whatever the application, having a metric between individual linguistic summaries is the first step towards building automated analysis algorithms for linguistically aggregated data.

In this paper, we create protoform-based linguistic summaries [19,20,30] in the sense of [31]. We propose a similarity/dissimilarity measure for those linguistic summaries, and prove that the proposed dissimilarity measure is a metric on the space of protoform summaries. The distance measure is studied analytically when the quantifiers are *L–R* fuzzy numbers, examined through simulation, and applied to summaries from both a toy example and real data. One simple example we use here is a collection of balls. The linguistic summaries can be exemplified by "most balls are big", "most of heavy balls are big", etc. Sets of these summaries are computed, distances are calculated and we show that they match our intuition. As a real world example, we apply the new metric to linguistic summaries generated from sensor data for a resident of TigerPlace, an "aging-in-place" facility in Columbia, MO [24]. Here also, we establish the potential of distance calculations on linguistic summaries of nighttime sensor firing.

2. Linguistic summaries of data

Linguistic summaries of data are usually short (quasi-)natural language sentences that capture the very essence of the set of data, that is numeric, large, and because of its size, not comprehensible to a human.

We use the following notation from [31]: $Y = \{y_1, y_2, \dots, y_n\}$ is the set of objects (records), e.g., night times for an elder or balls in a collection; $A = \{A_1, A_2, \dots, A_m\}$ is the set of attributes (features) characterizing objects from Y.

A linguistic summary includes:

- a summarizer P, i.e., an attribute together with a linguistic value (fuzzy predicate) defined on the domain of attribute A_j (e.g., *large* for attribute *size*);
- a quantity in agreement Q, i.e., a linguistic quantifier (e.g., *most*);
- truth (validity) $\mathcal{T}$ of the summary, i.e., a number from the interval [0, 1] assessing the truth (validity) of the summary (e.g., 0.7);
- optionally, a qualifier R, i.e., another attribute together with a linguistic value (fuzzy predicate) defined on the domain of attribute A_k determining a (fuzzy) subset of Y (e.g., *heavy* for attribute *weight*).

Basically the core of a linguistic summary is linguistically quantified and may be written as a simple protoform:

$$Qy\text{'s are } P \tag{1}$$

or as an extended protoform:

$$QRy\text{'s are } P \tag{2}$$

Here, Q, P and R are modeled by fuzzy sets over appropriate domains. The truth (validity), $\mathcal{T}$, of a linguistic summary directly corresponds to the truth value of (1) and (2). The truth may be calculated using either original Zadeh's calculus of quantified propositions (cf. [32]) or other interpretations of linguistic quantifiers. In the former case the truth values of (1) and (2) are calculated, respectively, as

$$\mathcal{T}(Q\ y\text{'s are } P) = \mu_Q\left(\frac{1}{n}\sum_{i=1}^{n}\mu_P(y_i)\right) \tag{3}$$

$$\mathcal{T}(QR\ y\text{'s are } P) = \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_P(y_i)\wedge\mu_R(y_i)}{\sum_{i=1}^{n}\mu_R(y_i)}\right) \tag{4}$$

where $\wedge$ is the minimum operation (more generally it can be another appropriate operator, notably a *t*-norm), and Q is a fuzzy set representing the linguistic quantifier that is normal and monotone.

Note that in most applications, both the fuzzy predicates P and R are assumed to be of a rather simplified, atomic form referring to just one attribute. They can be extended to cover more sophisticated summaries involving some confluence of various attribute values as, e.g., "*heavy* and *red*" balls. To combine more than one attribute value, like P_1 and P_2, we will use *t*-norms (for instance, the minimum or product) for conjunction. Actually, since the domains of P_1 and P_2

for such a compound attribute are different, the conjunction must take the form of the cylindrical closure in the cross product space [21]. Theoretically we may use a corresponding *s*-norm (for instance, the maximum or probabilistic sum, respectively) for disjunction of attributes, but such summaries appear to be not as meaningful as conjunctions for the domain experts.

Another very useful quality criterion of the linguistic summary is the degree of focus, introduced by Kacprzyk and Wilbik [17]. The very purpose of a degree of focus is to limit the search for the best linguistic summaries by taking into account some additional information in addition to truth values. The extended protoform linguistic summaries (2) does limit by itself the search space as the search is performed in a limited subspace of all (most) objects that fulfill an additional condition specified by qualifier R. The very essence of the degree of focus is to give the proportion of objects satisfying property R to all objects. The role of the degree of focus is similar to the concept of support in association rules in a data mining context [29]. It provides a measure that, in addition to the basic truth value, can help control the process of discarding non-promising linguistic summaries.

$$d_{foc}(QR\ y\text{'s are } P) = \frac{1}{n}\sum_{i=1}^{n}\mu_R(y_i) \tag{5}$$

The degree of focus in this form exists only for summaries of extended protoforms. We fix it to 1 for simple protoform linguistic summaries.

If the degree of focus is high, then we can be sure that such a summary concerns many objects (records), so that it is more general. When the degree of focus is low, such an abstraction describes a (local) pattern that occurs seldom.

Those two criteria are not the only ones. More on linguistic summaries, especially in the context of time series, can be found in [29]. In [23] there is a novel approach for generating these linguistic summaries, as they provide some rules on how to discard non-interesting or not promising summaries.

3. Similarity evaluation

In this section, we describe how to compute the proposed degree of similarity or distance between two linguistic summaries.

Consider two summaries Q_1R_1 y's are P_1 and Q_2R_2 y's are P_2 with the truth values $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively. The qualifiers R_1 and/or R_2 may be empty in case of simple protoform summaries.

The degree of similarity is computed as the minimum of the following four elements:

- Similarity of the summarizers P_1 and P_2. The formula for calculating this similarity depends if the summarizers describe the same attributes or not. Namely,

$$sim(P_1, P_2) = \min\left(\frac{a}{b}, \frac{\int(\mu_{P_1}\cap\mu_{P_2})}{\int(\mu_{P_1}\cup\mu_{P_2})}\right) \tag{6}$$

  where a is the number of joint (common) attributes for summarizers P_1 and P_2 and b is the number of distinct attributes used in summarizer P_1 or P_2, i.e., a/b is the Jaccard measure [16] of the sets of attributes for the summarizers P_1 and P_2. Also,

$$\frac{\int(\mu_{P_1}\cap\mu_{P_2})}{\int(\mu_{P_1}\cup\mu_{P_2})}$$

  is the Jaccard measure of the summarizers where μ_{P_1} and μ_{P_2} are the membership functions of fuzzy sets used to model the summarizers P_1 and P_2. Note that if the summarizers contain more than one attribute we operate on their cylindrical extensions [21].
- Similarity of the quantifiers Q_1 and Q_2 computed using the Jaccard measure, i.e.,

$$sim(Q_1, Q_2) = \frac{\int(\mu_{Q_1}\cap\mu_{Q_2})}{\int(\mu_{Q_1}\cup\mu_{Q_2})} \tag{7}$$

  where μ_{Q_1} and μ_{Q_2} are the membership functions of fuzzy sets used to model the linguistic quantifiers Q_1 and Q_2.

- Similarity of the truth values $\mathcal{T}_1$ and $\mathcal{T}_2$, of the summaries $Q_1 R_1$ y's are P_1 and $Q_2 R_2$ y's are P_2, respectively, computed as

$$sim(\mathcal{T}_1, \mathcal{T}_2) = 1 - |\mathcal{T}_1 - \mathcal{T}_2| \tag{8}$$

  where $|\cdot|$ denotes the absolute value of the difference.
- Similarity of the qualifiers R_1 and R_2. Note that in case of simple protoform summaries R is absent, and so, R should be treated as being the fuzzy set that characterizes the whole universe, Y, i.e., the set whose membership is 1 for all y. The similarity of the qualifiers, $sim(R_1, R_2)$, is defined as the minimum of two elements:
  - ○ Similarity computed using the Jaccard measure of fuzzy sets R_1 and R_2,

$$\frac{\int(\mu_{R_1} \cap \mu_{R_2})}{\int(\mu_{R_1} \cup \mu_{R_2})}$$

    where μ_{R_1} and μ_{R_2} are the membership functions of fuzzy sets used to model the qualifiers R_1 and R_2. Note that if the qualifiers contain more than one attribute we consider their cylindrical extensions.
  - ○ Similarity between the values of the degree of focus for those two summaries, computed as $1 - |d_{foc}(Q_1 R_1$ y's are $P_1) - d_{foc}(Q_2 R_2$ y's are $P_2)|$.

  Hence,

$$sim(R_1, R_2) = \min\left(\frac{\int(\mu_{R_1} \cap \mu_{R_2})}{\int(\mu_{R_1} \cup \mu_{R_2})}, 1 - |d_{foc}(Q_1 R_1 \text{ } y\text{'s are } P_1) - d_{foc}(Q_2 R_2 \text{ } y\text{'s are } P_2)|\right) \tag{9}$$

  Thus, if the two protoforms are simple, $sim(R_1, R_2) = 1$, since the degrees of focus are zero and the fuzzy sets are the universe set.

Finally, the total similarity between two protoform linguistic summaries is defined as

$$sim(Q_1 R_1 \text{ } y\text{'s are } P_1, Q_2 R_2 \text{ } y\text{'s are } P_2) = \min(sim(P_1, P_2), sim(Q_1, Q_2), sim(\mathcal{T}_1, \mathcal{T}_2), sim(R_1, R_2)) \tag{10}$$

Alternatively we may use the notion of dissimilarity, which is defined as

$$d(Q_1 R_1 \text{ } y\text{'s are } P_1, Q_2 R_2 \text{ } y\text{'s are } P_2) = 1 - sim(Q_1 R_1 \text{ } y\text{'s are } P_1, Q_2 R_2 \text{ } y\text{'s are } P_2) \tag{11}$$

We now show that the proposed dissimilarity in (11) is a metric over the space of protoform summaries. Note that the dissimilarity measure produces values in the range [0,1].

Lemma 1. *The Jaccard distance is defined as $d(A, B) = 1 - |A \cap B|/|A \cup B|$ and if $|A|$ denotes the cardinality of the set A (standard cardinality for crisp sets, Sigma-count for fuzzy sets) is a metric.*

Proof. The fact that the Jaccard distance is a metric and is crucial to the proof of our main result, and so, we give a proof here that comes from results in the literature. A direct algebraic alternative proof is given in the Appendix.

There is an equivalence between the triangle inequality of a dissimilarity measure d and the T_L-transitivity (with T_L the Lukasiewicz t-norm) of the corresponding similarity measure (obtained through complementation). So, $d(A, B) + d(B, C) = d(A, C)$ if and only if $\max(sim(A, B) + sim(B, C) - 1, 0) \leq sim(A, C)$ [12,13].

Moreover the Jaccard measure on crisp and on fuzzy sets (using sigma-count) is T_L-transitive [9,8]. The latter is proven more generally using meta-theorems on sigma-counts and for a broad class of similarity measures [10,11].

Thus, the Jaccard measure is a metric. □

Lemma 2. *Suppose that $d_1, d_2 : M \times M \rightarrow [0, 1]$ are two metrics on the same space. Then $d(x, y) = \max(d_1(x, y), d_2(x, y))$ is also a metric.*

Proof. Clearly, $\forall x, y\; d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$ and that $d(x, y) = d(y, x)$.

We show the triangle inequality: $\forall x, y, z\; d(x, z) \leq d(x, y) + d(y, z)$. Suppose that $d_1(x, z) \geq d_2(x, z)$. Then $d(x, z) = d_1(x, z) \leq d_1(x, y) + d_1(y, z) \leq \max(d_1(x, y), d_2(x, y)) + \max(d_1(y, z), d_2(y, z)) = d(x, y) + d(y, z)$. Similarly, the results hold when $d_2(x, z) \geq d_1(x, z)$. Thus, the max of two (or a finite number) of metrics is a metric. □

We are now ready to prove the main result.

Theorem 1. *The dissimilarity*

$$\begin{aligned} &d(Q_1 R_1 y\text{'s } \mathit{are}\ P_1, Q_2 R_2 y\text{'s } \mathit{are}\ P_2) \\ &\quad = 1 - \min(sim(P_1, P_2), sim(Q_1, Q_2), sim(\mathcal{T}_1, \mathcal{T}_2), sim(R_1, R_2)) \\ &\quad = \max(1 - sim(P_1, P_2), 1 - sim(Q_1, Q_2), 1 - sim(\mathcal{T}_1, \mathcal{T}_2), 1 - sim(R_1, R_2)) \end{aligned}$$

is a metric on the space of protoform summaries.

Proof. Note that

$$1 - sim(P_1, P_2) = 1 - \min\left(\frac{a}{b}, \frac{\int(\mu_{P_1} \cap \mu_{P_2})}{\int(\mu_{P_1} \cup \mu_{P_2})}\right) = \max\left(1 - \frac{a}{b}, 1 - \frac{\int(\mu_{P_1} \cap \mu_{P_2})}{\int(\mu_{P_1} \cup \mu_{P_2})}\right)$$

Now, $1 - a/b$ as well as

$$1 - \frac{\int(\mu_{P_1} \cap \mu_{P_2})}{\int(\mu_{P_1} \cup \mu_{P_2})}$$

are both Jaccard metrics by Lemma 1, the first over the crisp set of attributes and the second between the fuzzy sets that define the summarizers. Hence, $1 - sim(P_1, P_2)$ is a metric by Lemma 2.

Additionally, we have that $1 - sim(Q_1, Q_2)$ is a metric.

Clearly, $1 - sim(\mathcal{T}_1, \mathcal{T}_2) = 1 - 1 + |\mathcal{T}_1 - \mathcal{T}_2| = |\mathcal{T}_1 - \mathcal{T}_2|$, the 1-norm, is a metric.

For the fourth factor, we have

$$\begin{aligned} 1 - sim(R_1, R_2) &= 1 - \min\left(\frac{\int(\mu_{R_1} \cap \mu_{R_2})}{\int(\mu_{R_1} \cup \mu_{R_2})}, 1 - |d_{foc}(Q_1 R_1 y\text{'s are } P_1) - d_{foc}(Q_2 R_2 y\text{'s are } P_2)|\right) \\ &= \max\left(1 - \frac{\int(\mu_{R_1} \cap \mu_{R_2})}{\int(\mu_{R_1} \cup \mu_{R_2})}, |d_{foc}(Q_1 R_1 y\text{'s are } P_1) - d_{foc}(Q_2 R_2 y\text{'s are } P_2)|\right) \end{aligned}$$

Again

$$1 - \frac{\int(\mu_{R_1} \cap \mu_{R_2})}{\int(\mu_{R_1} \cup \mu_{R_2})}$$

is a Jaccard distance and $|d_{foc}(Q_1 R_1 y\text{'s are } P_1) - d_{foc}(Q_2 R_2 y\text{'s are } P_2)|$ is the 1-norm.

Therefore, all four elements $1 - sim(P_1, P_2)$, $1 - sim(Q_1, Q_2)$, $1 - sim(\mathcal{T}_1, \mathcal{T}_2)$, $1 - sim(R_1, R_2)$ are metrics, and so, their maximum is also a metric. □

This is not the only possibility. Instead of the Jaccard distance we could use other similarities (cf. [6,7,22]). If the corresponding dissimilarity between the fuzzy sets is a metric, then the dissimilarity between summaries will also fulfill the metric properties. Moreover it is also possible to combine the four elements not with the minimum but, for instance, with the weighted average, and it would also satisfy the metric properties. In the case of weighted average, there is a problem on how to choose weights. Additionally, if the regular average is used, the differences between summaries are not as visible as in case of using the minimum. The other possibilities will be investigated in a subsequent paper.

4. Numerical examples

In this section we present several examples from two domains: linguistic summaries of two boxes containing a certain number of balls of various sizes, weights, condition, age, colors, etc., and a real world example from ongoing work in eldercare. We begin with the simple toy problem for the purpose of analyzing the metric.

Every attribute (feature) is described with some linguistic values, which are modeled as fuzzy sets with trapezoidal membership functions. An example of such a function, Trap[a,b,c,d], is shown in Fig. 1.

We use the following linguistic values:

- size expressed in inches from 0 to 10—*small* (Trap[0,0,2,4]), *medium-size* (Trap[2,4,6,8]), *large* (Trap[6,8,9,10]), *huge* (Trap[8,9,10,10]);
- weight in pounds (also from 0 to 10)—*light* (Trap[0,0,2,4]), *medium-weight* (Trap[2,4,6,8]), *heavy* (Trap[6,8,10,10]);
- age in years—*new* (Trap[0,0,0.5,1]), *old* (Trap[3,5,10,10]);
- color, e.g., based on the Hue value—*red* (Trap[-30,-20,20,30]), *green* (Trap[90,100,140,150]), *blue* (Trap[210,220,260,270]), *reddish* (Trap[-120,-60,60,120]).

For the example, the linguistic quantifiers are: *almost all* (Trap[0.9,0.95,1,1]), *most* (Trap[0.5,0.8,1,1]), *many* (Trap[0.3,0.6,1,1]), *a few* (Trap[0.01,0.1,0.3,0.5]) and *none* [4](Trap[0,0,0.01,0.02]).

We start with a few general observations. Assume that a summary "*Many* balls are *large*" with truth value of $\mathcal{T}_1$ describes the first box, while a summary "*Many* balls are *large*" with truth value of $\mathcal{T}_2$ describes the second one. In Fig. 2 we show how the distance of those two summaries changes with the change of the values of the truth values $\mathcal{T}_1$ and $\mathcal{T}_2$. If those truth values are equal the distance equals 0, as expected.

Now let us analyze the following two summaries:

- *Many* balls are *large*; $\mathcal{T} = 1$.
- *Many* balls are *huge*; $\mathcal{T} = 1$.

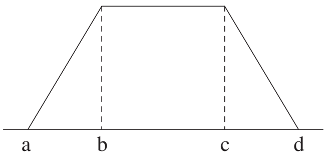

Fig. 1. Trapezoidal fuzzy membership function used in the numeric examples, denoted by Trap[a,b,c,d].

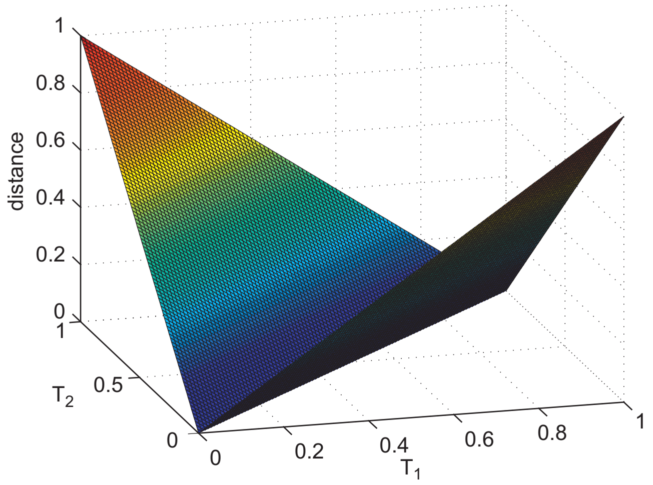

Fig. 2. Protoform distance between two identical summaries with differing truth values.

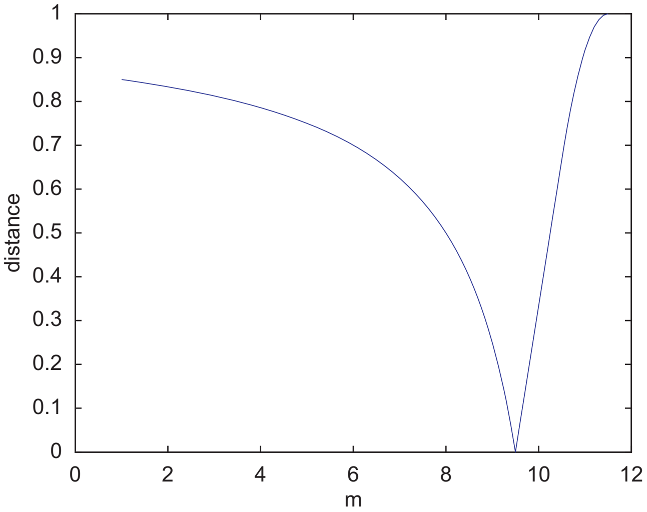

Fig. 3. Distance between *large* and *huge* as *large* changes size.

We examine the effect of the fuzzy set definitions on the distance value. In the Appendix, we develop the equations for the Jaccard distance (the foundation of our metric) in the case that the fuzzy sets are defined as *L–R* fuzzy numbers [14] to study the behavior of this metric. We further specify this formulation to the trapezoidal numbers we use here. Assume *huge* is defined as a fuzzy set with a trapezoidal membership function Trap[8,9,10,10], that can be represented as *L–R* fuzzy set Trap[$m - a$,$m - b$,$m + c$,$m + d$], where m is the center of the trapezoid and a, b, c and d are the distances to the change points left and right reference functions (see Appendix) and $m = 9.5$. *Large* is defined as [$m - 1.5$, $m - 0.5$, 10, 10]. Fig. 3 depicts the distance between the two summaries as the definition of large changes, depending on the value of m.

For small values of m the distance is *large*, since *large* is much bigger than *huge*. As the values of m increase the distance becomes smaller, and it is equal 0 for $m = 9.5$. In this case both fuzzy sets *large* and *huge* are identical. Then the distance becomes bigger since area of *large* is getting smaller and smaller. For m bigger than 10.5 there is only a small triangle left of *large*, until $m = 11.5$, where *large* is equal 0 everywhere. Fig. 3 thus depicts graphically the analysis found in the Appendix.

In Fig. 4 we show the distance of the two summaries: "*many* balls are *huge*" and "*many* balls are *large*" with respect to the area of intersection of the fuzzy sets *large* and *huge*. In this case the definition of *huge* (fuzzy set) is fixed and we modify and shift the membership function of *large*. The solid line represents a case when the area of *large* is the same as the area of *huge* (equal in the case depicted to 1.5). The dashed lines are examples of *larg*e sets with area smaller than that of *huge*, and the dotted lines represent *large* sets with area bigger than that of *huge*. The numbers define the area of the *large* set.

Now we consider some other linguistic summaries related to the example:

- s_1: *most* of balls are *large*, $\mathcal{T} = 0.75$;
- s_2: *many* balls are *large*, $\mathcal{T} = 0.9$;
- s_3: *many* balls are *huge*, $\mathcal{T} = 0.8$;
- s_4: *many* balls are *new*, $\mathcal{T} = 0.8$;
- s_5: *many* balls are *large* and *new*, $\mathcal{T} = 0.7$;
- s_6: *many new* balls are *large*, $\mathcal{T} = 1$, $d_{foc} = 0.6$;
- s_7: *many red* balls are *large*, $\mathcal{T} = 0.85$, $d_{foc} = 0.3$;
- s_8: *a few* balls are *large*, $\mathcal{T} = 0.1$;
- s_9: *a few* balls are of *medium size*, $\mathcal{T} = 0.8$;
- s_{10}: *a few* balls are *small*, $\mathcal{T} = 0.85$;
- s_{11}: *a few* of *red* and *new* balls are *large*, $\mathcal{T} = 0.7$, $d_{foc} = 0.2$.

The matrix of the distances for those pairs of summaries is presented in Table 1.

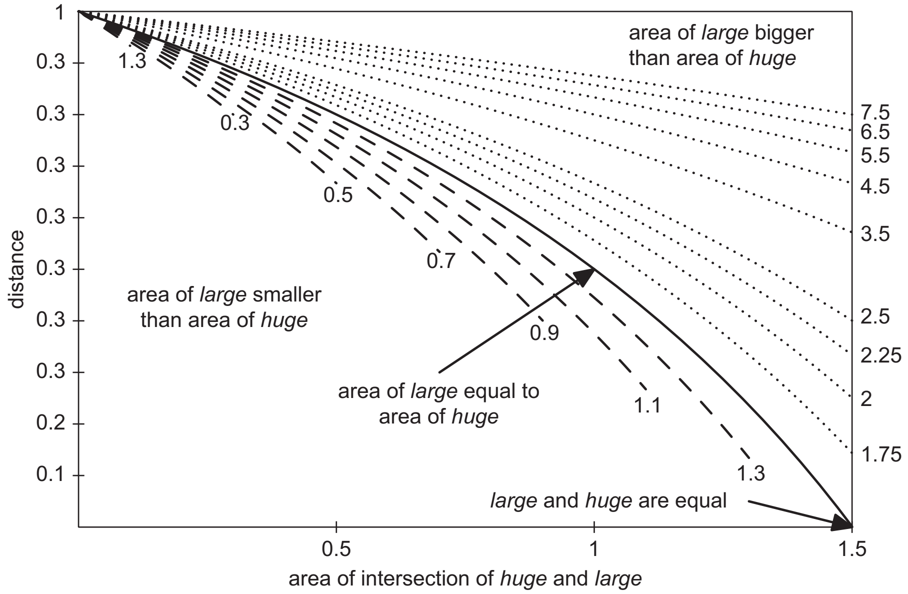

Fig. 4. Effects of shifting the definition of *large* across that of *huge*. The different curves represent different areas under the trapezoidal definition of *large*.

Table 1
The matrix of distances between summaries.

Summaries	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}
s_1	0.00	0.36	0.50	1.00	0.93	0.93	0.86	1.00	1.00	1.00	1.00
s_2	0.36	0.00	0.50	1.00	0.93	0.93	0.86	0.95	0.95	1.00	0.99
s_3	0.50	0.50	0.00	1.00	0.96	0.93	0.86	0.95	1.00	1.00	0.99
s_4	1.00	1.00	1.00	0.00	0.70	1.00	1.00	1.00	1.00	1.00	1.00
s_5	0.93	0.93	0.96	0.70	0.00	0.93	0.93	0.95	0.99	1.00	0.99
s_6	0.93	0.93	0.93	1.00	0.93	0.00	0.99	0.95	0.95	1.00	0.95
s_7	0.86	0.86	0.86	1.00	0.93	0.99	0.00	0.95	0.95	1.00	0.95
s_8	1.00	0.95	0.95	1.00	0.95	0.95	0.95	0.00	0.92	1.00	0.99
s_9	1.00	0.95	1.00	1.00	0.99	0.95	0.95	0.92	0.00	0.92	0.99
s_{10}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.00	1.00
s_{11}	1.00	0.99	0.99	1.00	0.99	0.95	0.95	0.99	0.99	1.00	0.00

Note that all the elements on the main diagonal are equal to 0 as in this case we compare two identical summaries, and of course, the matrix is symmetric. Other summaries are far away from each other. We analyze a few examples.

Summaries s_2 and s_3: *Large* and *huge* are the summarizers in s_2 and s_3, both referring to the attribute size. They are quite similar notions with similar fuzzy sets describing those two notions. The Jaccard similarity of the fuzzy sets representing those linguistic terms is 0.5; hence $sim(P_1, P_2) = 0.5$. In both summaries the same quantifier *many* was used; therefore $sim(Q_1, Q_2) = 1$. The similarity of the truth value is $sim(\mathcal{T}_1, \mathcal{T}_2) = 1 - |0.9 - 0.8| = 0.9$. Thus the degree of the similarity of s_2 and s_3 is equal to 0.5 and the distance to 0.5, a reasonable value.

Summaries s_1 and s_2: The difference in the above two summaries is the quantifier *most* and *many*. The Jaccard similarity of those two fuzzy sets is equal to 0.64. The similarity of the truth values is equal to 0.85. Hence, the similarity is equal to 0.64 and the distance to 0.36. It is intuitive that s_2 is closer to s_1 than it is to s_3.

Summaries s_2 and s_4: The above two summaries are not similar, i.e., their distance is equal to 1 because the summarizers refer to different attributes, i.e., *large* describes the size, while *new* refers to age.

Summaries s_2 and s_5: In this case, the above summaries are similar to a small degree of 0.2 and their distance is equal to 0.8. It is so because the Jaccard similarity of the fuzzy values of the summarizers: *large* and "*large* and *new*"

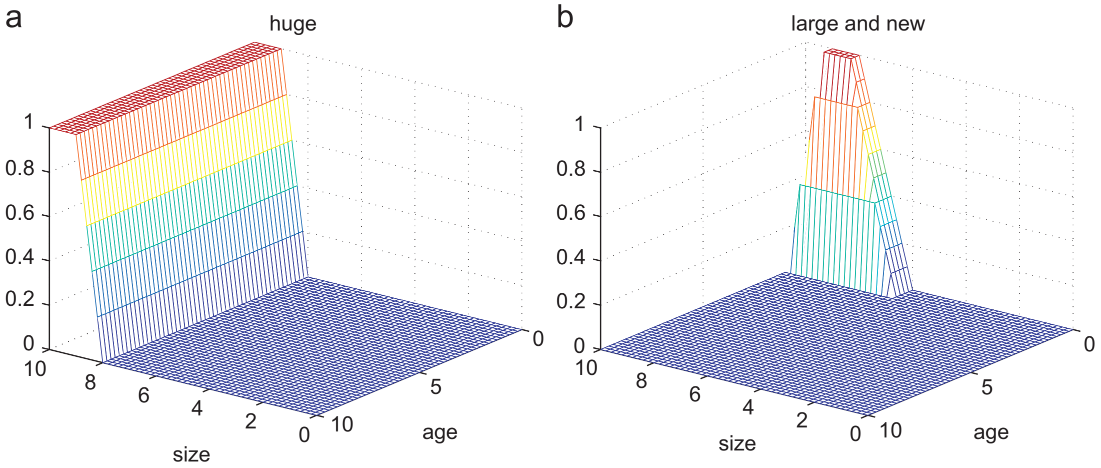

Fig. 5. Membership functions of *huge* (a) and "*large* and *new*" (b).

is equal 0.2. Note also that the similarity of attributes in the summarizer is equal to 0.5 because the sets of attributes of the summarizers are partially different. The summary s_5 contains additional information in comparison to the summary s_2 and this is the reason why they are not very similar.

Summaries s_2 *and* s_6: In this case we compare a simple protoform summary with an extended one. The similarity of the qualifier *new* and "empty set of qualifiers" denoting whole space is equal to 0.2 and as it is the smallest value of all similarities, the distance of these two summaries is equal to 0.8. Similarly as in the previous example summary s_6 contains additional information, hence their similarity is low.

Summaries s_3 *and* s_5: In Fig. 5 we show the membership functions of the summarizers of the summaries s_3 and s_5.

Note that their intersection will be very small, hence the similarity of *huge* and "*large* and *new*" is low. Hence the overall similarity of those two summaries is low.

Summaries s_6 *and* s_{11}: In those two summaries only summarizers are identical. Hence we need to compare quantifiers, qualifiers and the truth values. Qualifiers *many* and *a few* are not similar with the Jaccard similarity equal to 0.05. Also qualifiers *new* and "*light* and *new*" are of low similarity. Note there is a significant difference in values of the degrees of focus. The truth values are quite similar. As we can see those two summaries are not similar with the distance of 0.95.

Summaries s_2 *and* s_{10}: The distance of this pair of summary is equal to 1, since the large and small are linguistic values with different meanings. This result is clearly intuitive.

Now we present examples from a real world application. In [30] we showed linguistic summaries generated over a 15 month period for a male resident, of TigerPlace about 80 years old. He had a past history of syncope, bradycardia with pacemaker placement in 2002. He suffered from stenosis of carotid arteries, hypertension and probable transient ischemic attacks. He had a bypass surgery (CABG) in December 2005 and a stroke in December 2006. Based on the nurses notes describing his medical history, we distinguished some periods like *after CABG* or *stable time*. We used the nighttime sensor firings for two types of sensors: bed restlessness and motion in the apartment, which illustrates bed movement and movement around the apartment during every day. We described each attribute (one for each type of sensor) with three linguistic values *low* level, *medium* level and *high* level.

For the time after CABG (surgery) we obtained the following linguistic summaries:

- C_1: After CABG, on *most* nights the resident had a *high* level of restlessness; $\mathcal{T} = 0.79$, $d_{foc} = 1.0$.
- C_2: After CABG, on *most* of the nights, when the resident had a *high* level of motion, he had also a *high* level of restlessness; $\mathcal{T} = 1.0$, $d_{foc} = 0.58$.
- C_3: After CABG, on *most* of the nights, when the resident had a *low* level of motion, he had also a *low* level of restlessness; $\mathcal{T} = 1.0$, $d_{foc} = 0.22$.
- C_4: After CABG, on *most* of the nights, when the resident had a *low* level of restlessness, he had also a *low* level of motion; $\mathcal{T} = 0.83$, $d_{foc} = 0.27$.

Table 2
Distances between the linguistic summaries from stable time and after CABG period.

	C_1	C_2	C_3	C_4
S_1	0.99	0.99	0.95	1
S_2	1	1	1	0.94

After the surgery we may observe two types of behavior. The first behavior (top two summaries) is characterized with high level of restlessness, sometimes together with high level of motion. It means that he was not sleeping well, moving a lot in the bed, and also getting up often. Nurses' notes confirmed that the resident was suffering pain during this time period which could contribute to increased restlessness and lack of sleep.

The second behavior (last two summaries) occurred with low level of restlessness and low level of motion around the apartment; however the degree of focus of those two summaries are not very high, so it means it did not last a very long time. Nurses' notes confirmed that the resident was visiting his family around that time.

After the surgery and the rehabilitation the resident returned to his normal level of restlessness and motion detected at night, what can be seen on the following summaries:

- S_1: During stable time, on *most* nights the resident had a *medium* level of restlessness; $\mathcal{T} = 1.0$, $d_{foc} = 1.0$.
- S_2: During stable time, on *most* nights the resident had a *medium* level of motion; $\mathcal{T} = 1.0$, $d_{foc} = 1.0$.

The distance between the individual summaries for these periods are shown in Table 2.

The large distances in Table 2 clearly demonstrate the dissimilarity between the two periods.

Sometimes more detailed description may be required, for example of a single night. In this case we summarize the number of sensor firings in 15-min-slots during the nighttime, defined here as from 9pm till 7am. In this small example we are using the sensor data for the same resident. However, our data come from two sensors only: bed restlessness and bedroom motion, which illustrates the bed movement while lying in the bed and movement around the bedroom during the night. As an example we analyze one night after CABG period and two nights during the stable time, and we show both the linguistic summaries obtained and the distances between them. We use only three linguistic values for each attribute: *low* (Trap[0,0,2,5]), *medium* (Trap[2,5,12,15]) and *high* (Trap[12,15,50,50]).

The set of linguistic summaries generated for a night from after CABG time (January 12, 2006) is

- s_{11}: *almost all* 15-min-slots are *low* motion, $\mathcal{T} = 1.0$, $d_{foc} = 1.0$;
- s_{12}: *about a half* of the 15-min-slots are *high* restlessness, $\mathcal{T} = 1.0$, $d_{foc} = 1.0$;
- s_{13}: *a few* 15-min-slots are *low* restlessness, $\mathcal{T} = 1.0$, $d_{foc} = 1.0$;
- s_{14}: *a few* 15-min-slots are *medium* restlessness, $\mathcal{T} = 1.0$, $d_{foc} = 1.0$

and the linguistic summaries generated for a night from the stable time (August 15, 2006) are

- s_{21}: *almost all low* restlessness 15-min-slots are *low* motion, $\mathcal{T} = 1.0$, $d_{foc} = 0.72$;
- s_{22}: *most* 15-min-slots are *low* motion, $\mathcal{T} = 1.0$, $d_{foc} = 1.0$;
- s_{23}: *most low* motion 15-min-slots are *low* restlessness, $\mathcal{T} = 0.94$, $d_{foc} = 0.87$;
- s_{24}: *many* 15-min-slots are *low* motion and *low* restlessness, $\mathcal{T} = 0.91$, $d_{foc} = 1.0$;
- s_{25}: *many* 15-min-slots are *low* restlessness, $\mathcal{T} = 1.0$, $d_{foc} = 1.0$;
- s_{26}: *a few* 15-min-slots are *medium* restlessness, $\mathcal{T} = 1.0$, $d_{foc} = 1.0$.

The distance matrix of linguistic summaries for the two nights described above is shown in Table 3. This matrix compares two nights that are dissimilar. As intuitively expected, the linguistic summaries for those two nights are different, and we observe many distance values close to 1.

In Table 4 we present the distances of two nights from the stable time. The three linguistic summaries describing this night are

- s_{31}: *almost all low* restlessness 15-min-slots are *low* motion; $\mathcal{T} = 1.0$, $d_{foc} = 0.66$;
- s_{32}: *most* 15-min-slots are *low* motion; $\mathcal{T} = 1.0$, $d_{foc} = 1.0$;
- s_{33}: *many low* motion 15-min-slots are *low* restlessness; $\mathcal{T} = 1.0$, $d_{foc} = 0.88$.

Table 3
Distance matrix of linguistic summaries for the nights of January 12, 2006 and August 15, 2006.

	s_{11}	s_{12}	s_{13}	s_{14}
s_{21}	0.96	1.0	1.0	1.0
s_{22}	0.7	1.0	1.0	1.0
s_{23}	1.0	1.0	1.0	1.0
s_{24}	0.96	1.0	1.0	1.0
s_{25}	1.0	1.0	1.0	1.0
s_{26}	1.0	0.95	0.93	0.0

Table 4
Distance matrix of linguistic summaries for the night of March 18, 2006 and August 15, 2006.

	s_{31}	s_{32}	s_{33}
s_{21}	0.06	0.96	1.0
s_{22}	0.96	0.0	1.0
s_{23}	1.0	1.0	0.29
s_{24}	0.96	0.96	0.96
s_{25}	1.0	1.0	0.96
s_{26}	1.0	1.0	1.0

All three of the summaries for the second night are similar (small distance value) to one of the summaries from the first night.

As we see from this small example, the proposed similarity/distance metric provides the potential for automatically determining baseline activity and for detecting anomalous nights.

In the eldercare domain, computational complexity is not a big issue. Although the apartment of a resident may contain 30 different sensors, not all combinations make sense or are interesting for the experts (nurses). Hence in this framework only one set of summaries will be generated per day, which will be compared with those from some recent time frame, say the last two weeks (a good window span for comparison according to the health care providers).

5. Conclusions

A distance metric for linguistic summaries will enable many automated analysis activities to be performed on this compact representation of large data sets. In this paper, we defined a similarity measure for protoform summaries. The measure involves the linguistic summarizers, the truth values, the linguistic quantifiers and, in the case of extended protoforms, the linguistic qualifiers and degrees of focus. We then proved that the corresponding dissimilarity measure is a metric over this space of summaries. The properties of the metric were examined through two examples, a toy problem and data from sensor summarization in Eledercare where the results matched (our) intuition on the closeness of summaries.

Summaries are not, in general, isolated events. There may (should) be several that describe various aspects of the data. For example, in a sensor-based monitoring of elders, we can easily have a dozen or more summaries to describe the health related events of a given night: different levels of bed restlessness, motion in the bedroom, wandering in other rooms, number and frequency of bathroom visits, and so on. We want to compare last night with the previous night or with last week, this week with a week last month, this week with a “normal” week, etc. We now can measure the distance of each summary of the first group to each of the second. We are currently investigating methods of aggregating these distances to provide a global assessment. Future work involves temporal clustering of sets of summaries to define a possibly moving normal state as well as the efficacy of groups of linguistic summaries and their distances with respect to health care professionals.

Acknowledgments

This work was partially supported by NSF CPS program Grant CNS-0931607.

Appendix

In this Appendix we provide an algebraic alternative proof of Lemma 1 and more analysis of the properties of Jaccard distance of two fuzzy sets. The proof of Lemma 1 was already published [8–13]. Here we present a more direct one.

Proof.

(1) $\forall A, B\ d(A, B) \geq 0$ and $d(A, B) = 0$ if and only if $A = B$. Since $|A \cap B| \leq |A \cup B|$, $|A \cap B|/|A \cup B| \leq 1$, and so, $d(A, B) = 1 - |A \cap B|/|A \cup B| \geq 0$. Also, $d(A, B) = 1 - |A \cap B|/|A \cup B| = 0$ if and only if $|A \cap B|/|A \cup B| = 1$ if and only if $|A \cap B| = |A \cup B|$ if and only if $A = B$.

(2) $\forall A, B\ d(A, B) = d(B, A)$. $d(A, B) = 1 - |A \cap B|/|A \cup B| = 1 - |B \cap A|/|B \cup A| = d(B, A)$.

(3) $\forall A, B, C\ d(A, C) \leq d(A, B) + d(B, C)$ Suppose that this is not true, i.e., there are sets A, B, and C such that $d(A, C) > d(A, B) + d(B, C)$. Then

$$1 - \frac{|A \cap C|}{|A \cup C|} > 1 - \frac{|A \cap B|}{|A \cup B|} + 1 - \frac{|B \cap C|}{|B \cup C|}$$

or

$$\frac{|A \cap B|}{|A \cup B|} + \frac{|B \cap C|}{|B \cup C|} - \frac{|A \cap C|}{|A \cap C|} > 1$$

Since we postulate that the triangle inequality does not hold, we want the left hand side of this last inequality to be as small as possible. It is sufficient only to consider the cases where $A \subseteq B$ or $B \subseteq A$, since, if not, for $B' = A \cap B$ we have

$$\frac{|A \cap B'|}{|A \cup B'|} = \frac{|A \cap B|}{|A \cup (A \cap B)|} \geq \frac{|A \cap B|}{|A \cup B|}$$

and so, we can replace either A or B with $A \cap B$ and obtain

$$\frac{|A \cap B'|}{|A \cup B'|} + \frac{|B \cap C|}{|B \cap C|} - \frac{|A \cap C|}{|A \cup C|} \geq \frac{|A \cap B|}{|A \cup B|} + \frac{|B \cap C|}{|B \cup C|} - \frac{|A \cap C|}{|A \cup C|}$$

Similarly, either $C \subseteq B$ or $B \subseteq C$. Now we consider four cases. Case 1. $A \subseteq B$ and $C \subseteq B$. Then

$$\begin{aligned}\frac{|A \cap B|}{|A \cup B|} + \frac{|B \cap C|}{|B \cup C|} - \frac{|A \cap C|}{|A \cup C|} &= \frac{|A|}{|B|} + \frac{|C|}{|B|} - \frac{|A \cap C|}{|A \cup C|} = \frac{|A| + |C|}{|B|} - \frac{|A \cap C|}{|A \cup C|} \\ &= \frac{|A \cup C| + |A \cap C|}{|B|} - \frac{|A \cap C|}{|A \cup C|} \leq \frac{|A \cup C|}{|B|} + \frac{|A \cap C|}{|B|} - \frac{|A \cap C|}{|B|} \\ &= \frac{|A \cup C|}{|B|} \leq 1\end{aligned}$$

Case 2. $A \subseteq B$ and $B \subseteq C$. Then

$$\frac{|A \cap B|}{|A \cup B|} + \frac{|B \cap C|}{|B \cup C|} - \frac{|A \cap C|}{|A \cup C|} = \frac{|A|}{|B|} + \frac{|B|}{|C|} - \frac{|A|}{|C|} \leq \frac{|A|}{|C|} + \frac{|B|}{|C|} - \frac{|A|}{|C|} = \frac{|B|}{|C|} \leq 1$$

Case 3. $B \subseteq A$ and $C \subseteq B$. Then

$$\frac{|A \cap B|}{|A \cup B|} + \frac{|B \cap C|}{|B \cup C|} - \frac{|A \cap C|}{|A \cup C|} = \frac{|B|}{|A|} + \frac{|C|}{|B|} - \frac{|C|}{|A|} \leq \frac{|B|}{|A|} + \frac{|C|}{|A|} - \frac{|C|}{|A|} = \frac{|B|}{|A|} \leq 1$$

Case 4. $B \subseteq A$ and $B \subseteq C$. Then

$$\frac{|A \cap B|}{|A \cup B|} + \frac{|B \cap C|}{|B \cup C|} - \frac{|A \cap C|}{|A \cup C|} = \frac{|B|}{|A|} + \frac{|B|}{|C|} - \frac{|A \cap C|}{|A \cup C|} \leq \frac{|A \cap C|}{|A|} + \frac{|B|}{|C|} - \frac{|A \cap C|}{|A|} = \frac{|B|}{|C|} \leq 1$$

All four cases produce a contradiction to the statement that the expression had to be strictly bigger than 1. Hence, that assumption is not valid, and so, the triangle inequality holds.

Thus, the Jaccard measure is a metric. □

Additional analysis of the properties of the distance metric of Theorem 1.

As a more general case we replace the trapezoidal fuzzy sets with *L–R* fuzzy numbers (c.f. [14]) to define linguistic values. We will start with reminding the definition of *L–R* fuzzy numbers. *L* is a piecewise continuous non-increasing mapping from $[0, \infty)$ to [0,1] such that $L(0) = 1$, called reference function. Using two reference functions *L* and *R* a fuzzy set is defined such that

$$A(u) = \begin{cases} L\left(\frac{m-u}{\alpha}\right) & \text{if } u \leq m \\ R\left(\frac{u-m}{\beta}\right) & \text{if } u \geq m \end{cases} \tag{A.1}$$

where *m* is the mean value of *A*, and α and β are the right and left spreads. And the fuzzy number can be denoted as $A = (m, \alpha, \beta)_{L,R}$. Then the intersection (minimum) of two *L–R* fuzzy sets $A = (m_A, \alpha, \beta)_{L,R}$ and $B = (m_B, \alpha, \beta)_{L,R}$ is defined as

$$\min(A, B) = \begin{cases} 0 & \text{if } |m_A - m_B| \leq \alpha + \beta \\ L\left(\frac{\max(m_A, m_B) - u}{\alpha}\right) & \text{if } u \leq p \text{ and } |m_A - m_B| > \alpha + \beta \\ R\left(\frac{u - \min(m_A, m_B)}{\beta}\right) & \text{if } u > p \text{ and } |m_A - m_B| > \alpha + \beta \end{cases} \tag{A.2}$$

where *p* is the point of intersection of the left reference function of the set *B* and right reference function of the set *A*, such that, i.e.,

$$L\left(\frac{\max(m_A, m_B) - p}{\alpha}\right) = R\left(\frac{p - \min(m_A, m_B)}{\beta}\right) \tag{A.3}$$

Then the area of min(*A*,*B*) can be calculated as

$$\int \min(A, B) = \begin{cases} 0 & \text{if } |m_A - m_B| \leq \alpha + \beta \\ \int_{\max(m_A, m_B)-\alpha}^{p} L\left(\frac{\max(m_A, m_B) - u}{\alpha}\right) du & \\ \quad + \int_{p}^{\min(m_A, m_B)+\beta} R\left(\frac{u - \min(m_A, m_B)}{\beta}\right) du & \text{if } |m_A - m_B| > \alpha + \beta \end{cases} \tag{A.4}$$

Also, the area of the union (maximum) is calculated as

$$\begin{aligned} \int \max(A, B) &= \int A + \int B - \int \min(A, B) \\ &= \int_{m_A-\alpha}^{m_A} L\left(\frac{m_A - u}{\alpha}\right) du + \int_{m_A}^{m_A+\beta} R\left(\frac{u - m_A}{\beta}\right) du \\ &\quad + \int_{m_B-\alpha}^{m_B} L\left(\frac{m_B - u}{\alpha}\right) du + \int_{m_B}^{m_B+\beta} R\left(\frac{u - m_B}{\beta}\right) du - \int \min(A, B) \end{aligned} \tag{A.5}$$

In this simple case α and β were the same for both sets. We obtain more complicated formulas, with more intersection points, when they are different.

Hence the Jaccard similarity,

$$sim(A, B) = \frac{\int \min(A, B)}{\int \max(A, B)}$$

can be directly computed from (A.4) and (A.5).

For the specific case of trapezoids functions (Trap[$m-a$,$m-b$,$m+c$,$m+d$]), where m is the center of the trapezoid and a, b, c and d are the distances to the change points left and right reference functions are defined as

$$L(u)=\begin{cases}0 & \text{if } u \le m-a\\ \dfrac{u-m-a}{a-b} & \text{if } m-a \le u \le m-b\\ 1 & \text{if } m-b \le u \le m\end{cases} \tag{A.6}$$

$$R(u)=\begin{cases}0 & \text{if } u \ge m+d\\ \dfrac{m+d-u}{d-c} & \text{if } m+c \le u \le m+d\\ 1 & \text{if } m<u<m+c\end{cases} \tag{A.7}$$

In the case of two trapezoidal sets Trap[$m-a$,$m-b$,$m+c$,$m+d$]), the area of the intersection can be expressed as

$$\min(A,B)=\begin{cases}0 & \text{if } |m_A-m_B|>a+d\\ \dfrac{(\min(m_A,m_B)-\max(m_A,m_B)+d-a)^2}{2(a+d-b-c)} & \text{if } b+c<|m_A-m_B|<a+d\\ \min(m_A,m_B)-\max(m_A,m_B)+\dfrac{a+b+c+d}{2} & \text{if } |m_A-m_B|<b+c\end{cases} \tag{A.8}$$

And hence the Jaccard similarity is

$$sim(A,B)=\begin{cases}0 & \text{if } |m_A-m_B|>a+d\\ \dfrac{\dfrac{(\min(m_A,m_B)-\max(m_A,m_B)+d-a)^2}{2(a+d-b-c)}}{(a+b+c+d)-\dfrac{(\min(m_A,m_B)-\max(m_A,m_B)+d-a)^2}{2(a+d-b-c)}} & \text{if } b+c<|m_A-m_B|<a+d\\ \dfrac{\min(m_A,m_B)-\max(m_A,m_B)+\dfrac{a+b+c+d}{2}}{\dfrac{a+b+c+d}{2}-\min(m_A,m_B)+\max(m_A,m_B)} & \text{if } |m_A-m_B|<b+c\end{cases} \tag{A.9}$$

In the case presented in Fig. 3 we deal only with L fuzzy sets, the R side is equal to 1. Hence A=Trap[m_A-a,m_A-b,X_{max},X_{max}] and similarly B= Trap[m_B-a,m_B-b,X_{max},X_{max}], where X_{max} is the maximal value of the range. Then the area of intersection is equal to

$$\int \min(A,B)=\begin{cases}\dfrac{2X_{max}-\max(m_A,m_B)+a+b}{2} & \text{if } \max(m_A,m_B)-b \le X_{max}\\ \dfrac{(X_{max}-\max(m_A,m_B)+a)(X_{max}-\max(m_A,m_B)-a)}{2(a-b)} & \text{if } \max(m_A,m_B)-b > X_{max}\end{cases} \tag{A.10}$$

And the area of union as

$$\int \max(A,B)=\frac{2X_{max}-\min(m_A,m_B)+a+b}{2} \tag{A.11}$$

And Jaccard distance as

$$d(A,B) = \begin{cases} 1 - \dfrac{2X_{max} - \max(m_A, m_B) + a + b}{2X_{max} - \min(m_A, m_B) + a + b} & \text{if } \max(m_A, m_B) - b \leq X_{max} \\ \dfrac{1 - (X_{max} - \max(m_A, m_B) + a)(X_{max} - \max(m_A, m_B) - a)}{(a-b)(2X_{max} - \min(m_A, m_B) + a + b)} & \text{if } \max(m_A, m_B) - b > X_{max} \end{cases} \tag{A.12}$$

Eq. (A.12) is used to obtain Fig. 3.

References

[1] D. Anderson, R.H. Luke, J.M. Keller, M. Skubic, M. Rantz, M. Aud, Linguistic summarization of video for fall detection using voxel person and fuzzy logic, Comput. Vision Image Understanding 1 (113) (2009) 80–89.

[2] D. Anderson, R.H. Luke, J.M. Keller, M. Skubic, M. Rantz, M. Aud, Modeling human activity from voxel person using fuzzy logic, IEEE Trans. Fuzzy Syst. 1 (17) (2009) 39–49.

[3] P. Bosc, D. Dubois, O. Pivert, H. Prade, M.D. Calmes, Fuzzy summarization of data using fuzzy cardinalities, in: Proceedings of the IPMU 2002 Conference, 2002, pp. 1553–1559.

[4] R. Castillo-Ortega, N. Marín, D. Sánchez, Linguistic summarization for business intelligence using the time dimension in data warehouses, in: H. Weghorn, P.T. Isaías (Eds.), Proceedings of the IADIS International Conference Applied Computing 2009, IADIS AC, vol. 1, 2009, pp. 19–26.

[5] R. Castillo-Ortega, N. Marín, D. Sánchez, Time series comparison using linguistic fuzzy techniques, in: E. Hüllermeier, R. Kruse, F. Hoffmann (Eds.), Computational Intelligence for Knowledge-Based Systems Design, Proceedings of the 13th International Conference on Information Processing and Management of Uncertainty, IPMU 2010, Springer, 2010, pp. 330–339.

[6] S.-M. Chen, M.-S. Yeh, P.-Y. Hsiao, A comparison of similarity measures of fuzzy values, Fuzzy Sets Syst. 72 (1995) 79–89.

[7] V. Cross, T. Sudkamp, Similarity and Compatibility in Fuzzy Set Theory: Assessment and Applications, Springer-Verlag, Heidelberg and New York, 2002.

[8] B. De Baets, H. De Meyer, Transitivity-preserving fuzzification schemes for cardinality-based similarity measures, Eur. J. Oper. Res. 160 (3) (2005) 726–740.

[9] B. De Baets, H. De Meyer, H. Naessens, A class of rational cardinality-based similarity measures, J. Comput. Appl. Math. 132 (1) (2001) 51–69.

[10] B. De Baets, S. Janssens, H. De Meyer, Meta-theorems on inequalities for scalar fuzzy set cardinalities, Fuzzy Sets Syst. 157 (11) (2006) 1463–1476.

[11] B. De Baets, S. Janssens, H. De Meyer, On the transitivity of a parametric family of cardinality-based similarity measures, Int. J. Approximate Reasoning 50 (1) (2009) 104–116.

[12] B. De Baets, R. Mesiar, Pseudo-metrics and t-equivalences, J. Fuzzy Math. 5 (1997) 471–481.

[13] B. De Baets, R. Mesiar, Metrics and t-equalities, J. Math. Anal. Appl. 267 (2) (2002) 531–547.

[14] D. Dubois, H. Prade, Operations on fuzzy numbers, Int. J. Syst. Sci. 6 (9) (1978) 613–626.

[15] D. Dubois, H. Prade, Gradual rules in approximate reasoning, Inf. Sci. 61 (1992) 103–122.

[16] P. Jaccard, The distribution of the flora in the alpine zone, New Phytol. 11 (2) (1912) 37–50.

[17] J. Kacprzyk, A. Wilbik, Towards an efficient generation of linguistic summaries of time series using a degree of focus, in: Proceedings of the 28th North American Fuzzy Information Processing Society Annual Conference—NAFIPS 2009, 2009.

[18] J. Kacprzyk, A. Wilbik, Using fuzzy linguistic summaries for the comparison of time series: an application to the analysis of investment fund quotations, in: Proceedings of IFSA/EUSFLAT 2009, 2009, pp. 1321–1326.

[19] J. Kacprzyk, A. Wilbik, S. Zadrożny, Linguistic summarization of time series using a fuzzy quantifier driven aggregation, Fuzzy Sets Syst. 159 (12) (2008) 1485–1499.

[20] J. Kacprzyk, A. Wilbik, S. Zadrożny, An approach to the linguistic summarization of time series using a fuzzy quantifier driven aggregation, Int. J. Intelligent Syst. 25 (5) (2010) 411–439.

[21] in: G.J. Klir, B. Yuan (Eds.), Fuzzy Sets and Fuzzy Logic, Theory and Applications, Prentice Hall, 1995.

[22] C.P. Pappis, N.I. Karacapilidis, A comparative assessment of measures of similarity of fuzzy values, Fuzzy Sets Syst. 56 (2) (1993) 171–174.

[23] D. Pilarski, Linguistic summarization of databases with quantirius: a reduction algorithm for generated summaries, Int. J. Uncertainty Fuzziness Knowl.-Based Syst. 18 (3) (2010) 305–331.

[24] M.J. Rantz, R.T. Porter, D. Cheshier, D. Otto, C.H. Cervey, R.A. Johnson, M. Aud, M. Skubic, H. Tyrer, Z. He, G. Demiris, G.L. Alexander, G. Taylor, TigerPlace, a state-academic-private project to revolutionize traditional long term care, J. Housing Elderly 22 (1) (2008) 66–85.

[25] G. Raschia, N. Mouaddib, SAINTETIQ: a fuzzy set-based approach to database summarization, Fuzzy Sets Syst. 129 (2002) 137–162.

[26] D. Rasmussen, R.R. Yager, Finding fuzzy and gradual functional dependencies with SummarySQL, Fuzzy Sets Syst. 106 (1999) 131–142.

[27] M. Ros, M. Pegalajar, M. Delgado, A. Vila, D.T. Anderson, J.M. Keller, M. Popescu, Linguistic summarization of long-term trends for understanding change in human behavior, in: Proceedings of the IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2011, 2011, pp. 2080–2087.

[28] I.J. Sledge, J.M. Keller, G.L. Alexander, Emergent trend detection in diurnal activity, in: Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2008, 2008, pp. 3815–3818.

[29] A. Wilbik, Linguistic Summaries of Time Series Using Fuzzy Sets and Their Application for Performance Analysis of Investment Funds. Ph.D. Thesis, Systems Research Institute, Polish Academy of Sciences, 2010.

[30] A. Wilbik, J.M. Keller, G.L. Alexander, Linguistic summarization of sensor data for eldercare, in: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC 2011), 2011, pp. 2595–2599.

[31] R.R. Yager, A new approach to the summarization of data, Inf. Sci. 28 (1982) 69–86.

[32] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets Syst. 9 (2) (1983) 111–127.