[17] R. K. H. Galvao and T. Yoneyama, "A competitive wavelet network for signal clustering," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 34, no. 2, pp. 1282–1288, Apr. 2004.
[18] X. L. Li, S. K. Ts, and J. Wang, "Real-time tool condition monitoring using wavelet transforms and fuzzy techniques," *IEEE Trans. Syst., Man Cybern. C, Appl. Rev.*, vol. 30, no. 3, pp. 352–357, Aug. 2000.
[19] G. D. Wu, Z. W. Zhu, and P. H. Huang, "A TS-type maximizing-discriminability-based recurrent fuzzy network for classification problems," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 2, pp. 339–352, Apr. 2011.
[20] W. Pedrycz, V. Loia, and S. Senatore, "Fuzzy clustering with viewpoints," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 2, pp. 274–284, Apr. 2010.
[21] X. W. Liu and J. M. Mendel, "Connect Karnik–Mendel algorithms to root-finding for computing the centroid of an interval type-2 fuzzy set," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 4, pp. 652–665, Aug. 2011.
[22] J. M. Mendel, F. L. Liu, and D. Y. Zhai, "A-plane representation for type-2 fuzzy sets: Theory and applications," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 5, pp. 1189–1207, Oct. 2009.
[23] D. Y. Zhai and J. M. Mendel, "Computing the centroid of a general type-2 fuzzy set by means of the centroid-flow algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 3, pp. 401–422, Jun. 2011.
[24] S. Barkat, A. Tlecami, and H. Nouri, "Noninteracting adaptive control of PMSM using interval type-2 fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 5, pp. 925–936, Oct. 2011.
[25] S. Jafarzadeh, M. Sami Fadali, and A. H. Sonbol, "Stability analysis and control of discrete type-1 and type-2 TSK fuzzy systems—Part I: Stability analysis," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 6, pp. 989–1000, Dec. 2011.
[26] J. M. Mendel, R. I. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 6, pp. 808–821, Dec. 2009.
[27] N. N. Karnik, J. M. Mendel, and Q. L. Liang, "Type-2 fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 6, pp. 643–658, Dec. 1999.
[28] R. John and S. Coupland, "Type-2 fuzzy logic: A historical view," *IEEE Trans. Comput. Intell. Mag.*, vol. 2, no. 1, pp. 57–62, Feb. 2007.
[29] H. Hagras, "Type-2 FLCs: A new generation of fuzzy controllers," *IEEE Trans. Comput. Intell. Mag.*, vol. 2, no. 1, pp. 30–43, Feb. 2007.
[30] F. J. Lin, P. H. Chou, P. H. Shieh, and S. Y. Chen, "Robust control of an motion control stage using an adaptive interval type-2 fuzzy neural network," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 1, pp. 24–38, Feb. 2009.
[31] F. J. Lin and P. H. Chou, "Adaptive control of two-axis motion control system using interval type-2 fuzzy neural network," *IEEE Trans. Ind. Electron.*, vol. 56, no. 1, pp. 178–193, Jan. 2009.
[32] J. R. Castro, O. Castillo, P. Melin, and A. Rodríguez-Díaz, "A hybrid learning algorithm for a class of interval type-2 fuzzy neural networks," *Inf. Sci.*, vol. 179, no. 13, pp. 2175–2193, 2009.
[33] C. S. Lee, M. H. Wang, and H. Hagras, "A type-2 fuzzy ontology and its application to personal diabetic-diet recommendation," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 2, pp. 374–395, Apr. 2010.
[34] A. Niewiadomski, "On finity, countability, cardinalities, and cylindric extensions of type-2 fuzzy sets in linguistic summarization of databases," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 3, pp. 532–545, Jun. 2010.
[35] Q. Wu and R. Law, "Complex system fault diagnosis based on a fuzzy robust wavelet support vector classifier and an adaptive Gaussian particle swarm optimization," *Inf. Sci.*, vol. 180, no. 23, pp. 4514–4528, 2010.
[36] S. J. Yoo, J. B. Park, and Y. H. Choi, "Indirect adaptive control of nonlinear dynamic systems using self recurrent wavelet neural networks via adaptive learning rates," *Inf. Sci.*, vol. 177, no. 15, pp. 3074–3098, 2007.
[37] S. Arafat and M. Skubic, "Modeling fuzziness measures for best wavelet selection," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 5, pp. 1259–1270, Oct. 2008.
[38] H. X. Li, X. X. Zhang, and S. Y. Li, "A three-dimensional fuzzy control methodology for a class of distributed parameter system," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 3, pp. 470–481, Jun. 2007.
[39] Y. Li, H. L. Wei, and S. A. Billings, "Identification of time-varying systems using multi-wavelet basis functions," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 3, pp. 656–663, May 2011.
[40] X. X. Zhang, H. X. Li, and C. K. Qi, "Spatially constrained fuzzy-clustering-based sensor placement for spatiotemporal fuzzy control system," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 5, pp. 946–957, Oct. 2010.
[41] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 121–131, Jan. 2011.
[42] J. P. S. Catalao, H. M. I. Pousinho, and V. M. F. Mendes, "Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 137–144, Feb. 2011.
[43] C. Y. Yeh, W. H. R. Jeng, and S. J. Lee, "An enhanced type-reduction algorithm for type-2 fuzzy sets," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 2, pp. 227–240, Apr. 2011.
[44] D. R. Wu and J. M. Mendel, "On the continuity of type-1 and interval type-2 fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 1, pp. 179–192, Feb. 2011.
[45] Y. J. Liu and W. Wang, "Adaptive fuzzy control for a class of uncertain nonaffine nonlinear systems," *Inf. Sci.*, vol. 177, no. 18, pp. 3901–3917, 2007.
[46] Y. J. Liu, W. Wang, S. C. Tong, and Y. S. Liu, "Robust adaptive tracking control for nonlinear systems based on bounds of fuzzy approximation parameters," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 40, no. 1, pp. 170–184, Jan. 2010.

# A Fuzzy Measure Similarity Between Sets of Linguistic Summaries

Anna Wilbik and James M. Keller, *Fellow, IEEE*

***Abstract*—In this paper, we consider the problem of evaluating the similarity of two sets of linguistic summaries of sensor data. Huge amounts of available data cause a dramatic need for summarization. In continuous monitoring, it is useful to compare one time interval of data with another, for example, to detect anomalies or to predict the onset of a change from a normal state. Assuming that summaries capture the essence of the data, it is sufficient to compare only those summaries, i.e., they are descriptive features for recognition. In previous work, we developed a similarity measure between two individual summaries and proved that the associated dissimilarity is a metric. Additionally, we proposed some basic methods to combine these similarities into an aggregate value. Here, we develop a novel parameter free method, which is based on fuzzy measures and integrals, to fuse individual similarities that will produce a closeness measurement between sets of summaries. We provide a case study from the eldercare domain where the goal is to compare different nighttime patterns for change detection. The reasons for studying linguistic summaries for eldercare are twofold: First, linguistic summaries are the natural communication tool for health care providers in a decision support system, and second, due to the extremely large volume of raw data, these summaries create compact features for an automated reasoning for detection and prediction of health changes as part of the decision support system.**

***Index Terms*—Anomaly detection, fuzzy measure, linguistic summaries, similarity, Sugeno integral.**

Manuscript received February 1, 2012; revised May 19, 2012; accepted July 13, 2012. Date of publication August 20, 2012; date of current version January 30, 2013. This work was supported in part by the National Science Foundation Cyber-Physical Systems program under Grant CNS-0931607.

A. Wilbik is with Systems Research Institute, Polish Academy of Sciences, 00-956 Warsaw, Poland, and also with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211 USA (e-mail: wilbik@ibspan.waw.pl).

J. M. Keller is with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211 USA (e-mail: kellerj@missouri.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2012.2214225

TABLE I
AMOUNT OF DATA LOGGED IN ONE APARTMENT AT TIGERPLACE IN A WEEK

| Sensors | Amount of weekly data in one apartment |
|---|---|
| PIR motion plus bed sensor | 0.2 GB |
| Doppler radar | 1.0 GB |
| Hydraulic bed sensor | 1.0 GB |
| Vision sensors (compressed silhouettes) | 7.0 GB |
| Kinect depth imagery (compressed) | 14 GB |
| Total | 23.2 GB |

## I. INTRODUCTION

Evaluating similarity between two objects, properties, etc., is a very common and important problem in automated decision making. Consider, for example, the problem of evaluating similarity of the health-related sensor time series in the context of eldercare. Elderly people want to remain active and independent for as long as possible. To fulfill those needs, the concept of "aging in place" has been developed and realized, for instance, in TigerPlace [1], which is an independent living environment, in Columbia, MO. In TigerPlace, integrated sensor networks have been installed in apartments of the residents, some for many years [2]. They produce huge amounts of data. Table I displays the volume of data captured by the sensor network in one apartment in TigerPlace. It is impossible for health care providers to shift through the sensor outputs, even with the graphical interfaces available. In addition, conditions, such as gait, are extracted features at different time scales compared with say, motion and bed sensor data, and therefore, some type of aggregation is necessary. Linguistic summarization provides a person-friendly mechanism to compact this mountain of data into relevant reports about health conditions and changes of an elderly resident. Additionally, given the sheer volume of data, reduced sets of significant features must be generated to feed any automated decision making process. In [3], we demonstrated that linguistic summaries are meaningful condensations of large amounts of sensor readings for our Nursing collaborators. Our hypothesis, then, is that these summaries can and should be employed in pattern recognition approaches, such as early illness detection.

Of course, there are many approaches to pattern recognition using raw data, but linguistic summarization of those data seems to be a natural solution, since natural language is the only fully comprehensive means of communication for humans. Several approaches of linguistic summaries were proposed and investigated [4]–[12]. Using the format proposed in [11], we generate protoform-based summaries for eldercare health monitoring [3]. In particular, we construct a set of linguistic summaries for every night for motion sensor firings within 15-min intervals e.g., "*many* 15-min intervals have *low* restlessness" or "*almost all* 15-min intervals of *medium* bedroom motion have *medium* restlessness." Our aim is to find a method of comparison of those linguistic descriptions with another set of linguistic summaries representing a "typical night."

In [13], we built a dissimilarity measure between individual protoform summaries, and proved that it is a metric over the space of such summaries. In any real application, many such synopses will be produced in a given time frame, for example, nighttime activity in a senior's apartment. If we want to compare one night with the next to perform change detection, or to compare one night with a set of "normal night" summaries, we need the ability to aggregate the distances between the almost assuredly unequal sets of summaries representing the two time periods. It is not possible, therefore, to line up the distances in a vector and do standard vector space comparisons. This seemingly simple operation is the key to utilization of this new people-friendly feature set for pattern recognition. The individual distances are the building blocks, but a good aggregation scheme that is sensitive to real health changes but somewhat insensitive to minor fluctuations of activity patterns is crucial.

In [14], we considered the "soft degree" of consensus between sets of summaries. While that approach gave reasonable results, it is based on defining a quantifier in the calculation, like "*most* summaries are similar." It would be advantageous to build a parameter free comparison measure directly from individual distances. We had success in comparing sets of ontology annotations for gene products within the framework of fuzzy measure theory [15], and therefore, here, we propose a similarity measure of two sets of linguistic summaries that also utilizes fuzzy measures. More on measure guided aggregation methods can be found in [16]. The proposed methodology can be applied to a wide range of time-series analysis problems, but we demonstrate its efficacy for eldercare in this paper. This method of comparison can distinguish abnormal nights from a typical one and can detect changes in behavior of an elderly resident that may indicate a possible health problem.

## II. BACKGROUND

We are using the concept of linguistic protoform summaries that was proposed in [11], [17], and [18]. A linguistic summary is a prototype (template)-based quasi-natural language sentence of a simple form

$$Q\ y\text{'s are } P \tag{1}$$

or in extended form

$$QR\ y\text{'s are } P \tag{2}$$

where $Q$ is a quantifier, $P$ is a summarizer, $R$ is a qualifier, and $y$'s represent the objects that are summarized. In our task of elder activity monitoring, the objects are 15-min intervals of a time series representing some sensor firings of nighttime activity of the resident. Examples of quantifiers include *most*, *many*, *almost all*, etc.; summarizers can be terms like *high restlessness*, *low bedroom motion*, and so on; in addition, qualifiers are used to restrict summaries to a subset of objects, for example, in "*almost all* 15-min intervals of *medium* bedroom motion have *medium* restlessness," the phrase "of *medium* bedroom motion" is the qualifier. For every summary, we calculate the truth value, which is the basic criterion to evaluate the quality of the linguistic summaries. Truth values for simple protoforms are defined by

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

and for extended protoforms, including a qualifier, as

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

where $n$ is the number of objects ($y_i$) that are summarized, and $\mu_P$, $\mu_R$, and $\mu_Q$ are the membership functions of the summarizer, qualifier, and quantifier, respectively.

Another very useful quality criterion for a protoform summary is the degree of focus. The essence of the degree of focus is to specify the proportion of all objects satisfying property $R$. The role of the degree of focus is similar to the concept of support in association rules in a data mining context [19]. It provides a measure that, in addition to the basic truth value, can help control the process of discarding nonpromising linguistic summaries. Its calculation is given by

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

The degree of focus exists only for extended protoform summaries. We set 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, and hence, it is more general. When the degree of focus is low, such an abstraction describes a (local) pattern that occurs infrequently. More about the linguistic summaries, their quality evaluation, and methods for their generation can be found in [19].

In [13], we developed a similarity between the two linguistic summaries that is calculated as

$$\begin{aligned}&\mathrm{sim}(Q_1 R_1 \ y\text{'s are } P_1, Q_2 R_2 \ y\text{'s are } P_2)\\ &= \min\Bigg(\min\left(\frac{a}{b}, \frac{\int(\mu_{P_1} \cap \mu_{P_2})}{\int(\mu_{P_1} \cup \mu_{P_2})}\right)\\ &\qquad \min\left(\frac{\int(\mu_{R_1} \cap \mu_{R_2})}{\int(\mu_{R_1} \cup \mu_{R_2})}, 1 - |d_{foc1} - d_{foc2}|\right)\\ &\qquad \frac{\int(\mu_{Q_1} \cap \mu_{Q_2})}{\int(\mu_{Q_1} \cup \mu_{Q_2})}, 1 - |\mathcal{T}_1 - \mathcal{T}_2|\Bigg) \qquad (6)\end{aligned}$$

where $\frac{a}{b}$ is the Jaccard measure of the sets of attributes for the summarizers $P_1$ and $P_2$, i.e., $a$ is the number of common attributes in summarizers $P_1$ and $P_2$, and $b$ is the total number of attributes in $P_1$ and $P_2$. In (6), $\int(\mu_{P_1} \cap \mu_{P_2})$ is the area of the intersection of the membership functions of fuzzy sets representing sets $P_1$ and $P_2$. Note that if the summarizers contain more than one attribute, such as $low$ restlessness and $medium$ bedroom motion, we operate on their cylindrical extensions [20]. Additionally, $\mathcal{T}_1$ and $\mathcal{T}_2$ are the truth values, and $d_{\mathrm{foc}1}$ and $d_{\mathrm{foc}2}$ are the degrees of focus of the summaries $Q_1 R_1 \ y$'s are $P_1$ and $Q_2 R_2 \ y$'s are $P_2$, respectively. Hence, the similarity between two, possibly extended, protoform summaries is generated as the minimum of the agreement of the summarizers, the quantifiers, the truth values, and if available, the qualifiers and their degrees of focus. The numeric parts of the summaries are compared with the 1-norm and the fuzzy set parts by the Jaccard measure [21].

We proved in [13] that the associated dissimilarity measure

$$\begin{aligned}&d(Q_1 R_1 \ y\text{'s are } P_1, Q_2 R_2 \ y\text{'s are } P_2)\\ &\quad = 1 - \mathrm{sim}(Q_1 R_1 \ y\text{'s are } P_1, Q_2 R_2 \ y\text{'s are } P_2)\end{aligned}$$

is a metric over the space of protoform summaries. This distance satisfies the properties of any metric (nonnegatitvity, identity of indiscernibles, symmetry, and the triangle inequality). The main proof utilizes results concerning Jaccard measures and Minkowski norms, but a direct proof that it satisfies the triangle inequality is found in [13, App.].

## III. Similarity of Sets of Summaries With Fuzzy Measures

Given two sets of linguistic summaries $C_1 = \{S_1^1, S_2^1, \ldots, S_n^1\}$ and $C_2 = \{S_1^2, S_2^2, \ldots, S_m^2\}$, we can now compute the pairwise similarities between all members of $C_1$ and $C_2$. Our proposed fusion methodology for the sets is based on fuzzy measures. Let $X = \{x_1, x_2, x_3, \ldots, x_n\}$ be a finite set. A fuzzy measure [22] on $X$ is a function $g : 2^X \to [0, 1]$ such that

1) $g(\oslash) = 0$;
2) $g(X) = 1$;
3) if $A \subseteq B$, then $g(A) \le g(B)\ \forall A, B \in 2^X$.

The measure $g$ satisfies the λ-rule if and only if there exists $\lambda \in (-1, \infty)$ such that $g(E \cup F) = g(E) + g(F) + \lambda g(E) \cdot g(F)$ whenever $E \cap F \neq \oslash$. Such a function $g$ is called a λ-fuzzy measure on $2^X$.

When $g(X) = \sum_{i=1}^n g(\{x_i\})$, the parameter $\lambda = 0$; otherwise, λ can be uniquely determined by the following equation [23]:

$$g(X) = \frac{1}{\lambda}\left[\prod_{i=1}^n (1 + \lambda \cdot g(\{x_i\})) - 1\right]. \qquad (7)$$

This equation comes directly from the λ-rule and the fact that $X = \bigcup_{i=1}^n \{x_i\}$. The values $\{g(\{x_i\}) | i = 1, 2, \ldots, n\}$ are called the fuzzy densities of $g$, and they uniquely define the λ-measure.

In our context, the set $X$ is a set of linguistic summaries $C = \{S_1, S_2, \ldots, S_n\}$. The densities that we choose for this application, $g_\lambda(S_i) = (\mathcal{T}_i \cdot d_{\mathrm{foc}\ i})/n$, are then used to compute the corresponding λ-fuzzy measure. Note that if all degrees of truth and degrees of focus are 1, then the densities add up to one, and $\lambda = 0$, generating a probability measure. Otherwise, as usually the case, several of the densities will be less than $1/n$, and therefore, their sum will be less than 1, producing a λ that is bigger than 0. Intuitively, subsets of $C$ that contain summaries whose degrees of truth are close to 1 and whose foci include most of the objects (degree of focus close to 1) have large measure. Hence, those subsets are more important in describing the overall condition under evaluation. Certainly, other choices of densities could be made, but these are the natural values for protoform-based linguistic summaries.

Now, suppose we have two sets of summaries $C_1 = \{S_1^1, S_2^1, \ldots, S_n^1\}$ and $C_2 = \{S_1^2, S_2^2, \ldots, S_m^2\}$. Let $\mathcal{T}_j^i$ be the truth value, and $d_{\mathrm{foc}\ j}^i$ be the degree of focus of the summary $S_j^i$. We then compute two λ-fuzzy measures, one over $C_1$ and the other over $C_2$ by defining the densities

$$g_\lambda^1(S_i^1) = \frac{\mathcal{T}_i^1 \cdot d_{\mathrm{foc}\ i}^1}{n} \qquad (8)$$

$$g_\lambda^2(S_i^2) = \frac{\mathcal{T}_i^2 \cdot d_{\mathrm{foc}\ i}^2}{m} \qquad (9)$$

using (7) to calculate $\lambda_1$ and $\lambda_2$, and, hence, the two fuzzy measures $g_1$ and $g_2$.

These fuzzy measures are used to compare the two sets of summaries. For each $\alpha \in [0, 1]$, define sets

$$C_{1\alpha} = \left\{S_i^1 \in C_1 : \min_{j=1,\ldots,m} d(S_i^1, S_j^2) \le \alpha\right\}$$

and

$$C_{2\alpha} = \left\{S_j^2 \in C_2 : \min_{i=1,\ldots,n} d(S_i^1, S_j^2) \le \alpha\right\}.$$

Here, $C_{1\alpha}$ represents the set of summaries in $C_1$ that are $\alpha$-close to some summary in $C_2$ and conversely for $C_{2\alpha}$. The more "good" summaries from a given set that have counterparts (small distances) in the other set, the stronger similarity is between the sets.

Next, we define a fuzzy measure on $C_1 \times C_2$ as the function $g : 2^{C_1} \times 2^{C_2} \to [0, 1]$, such that if $A \subseteq 2^{C_1}$ and $B \subseteq 2^{C_2}$, then

$$g(A, B) = \frac{g_1(A) + g_2(B)}{2}. \qquad (10)$$

Instead of averaging, any t-norm or s-norm may be used, providing a different sense to the aggregation.

The function $g$ is a fuzzy measure since

1) $g(\oslash, \oslash) = 0$;
2) $g(X) = 1$ because $g(C_1, C_2) = \frac{1+1}{2} = 1$;
3) $g$ is monotonic, because $\forall A, B \subseteq 2^{C_1} \times 2^{C_2}$ if $A = (A_1, A_2) \subseteq B = (B_1, B_2)$; then, $A_1 \subseteq B_1$ and $A_2 \subseteq B_2$, and hence, $g(A) = \frac{g_1(A_1)+g_2(A_2)}{2} \le \frac{g_1(B_1)+g_2(B_2)}{2} = g(B)$.

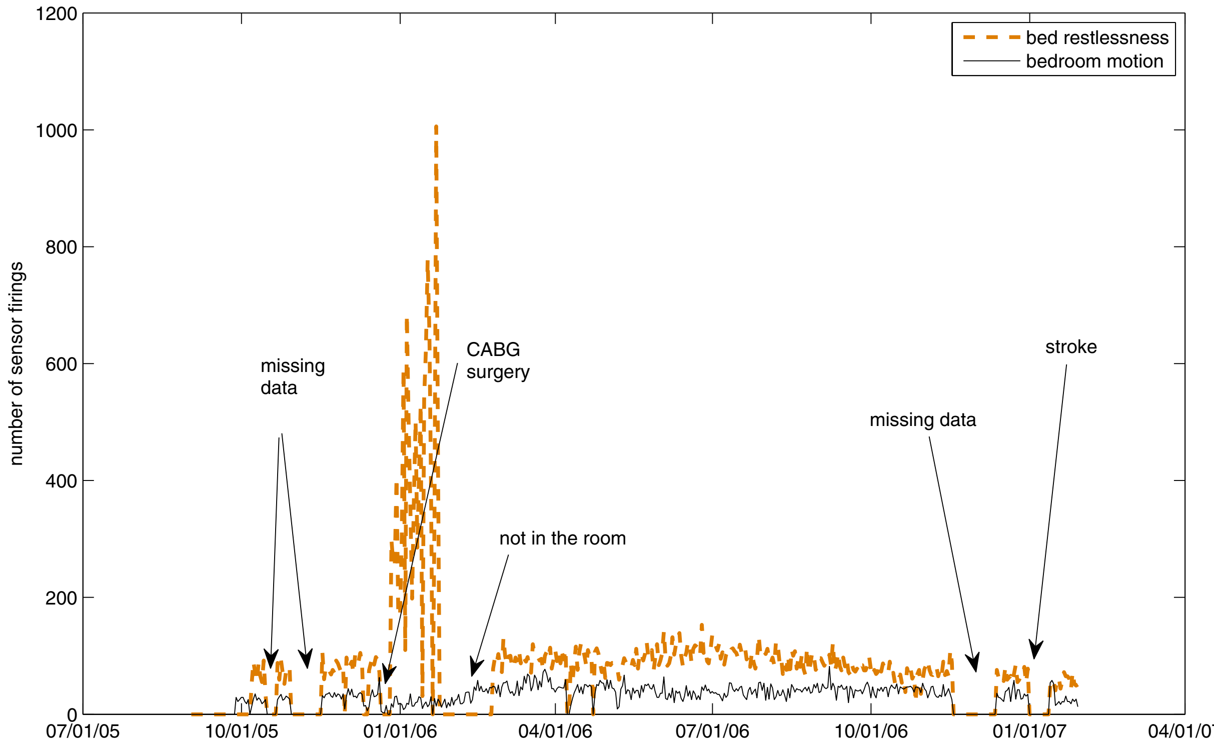


Fig. 1. Nighttime sensor firings for two types of sensors: bed restlessness and bedroom motion summed over each night for display.

We could just compute $g(C_{1\alpha}, C_{2\alpha})$ for some $\alpha < 1$ as the similarity of the two sets of summaries, but that introduces a parameter that needs to be determined. Instead, we define the similarity by integrating the measures across all $\alpha$.

Hence, we define the similarity of the two sets of protoform summaries $C_1$ and $C_2$ as the Sugeno integral of the function $f(x) = 1 - x$ with respect to the measure $g$

$$\mathrm{sim}_S(C_1, C_2) = \max_{\alpha \in [0,1]} (\min(1-\alpha, g(C_{1\alpha}, C_{2\alpha}))). \tag{11}$$

Note that the function to be integrated is obviously sorted in descending order, and the measure is evaluated on the corresponding level set.

## IV. Example

As an example, we show real data from a resident's apartment in TigerPlace. Our data come from an almost 80 year old (in 2005) resident who had a past history of syncope and bradycardia, with a pacemaker placement in 2002. He suffered from stenosis of his carotid arteries, hypertension, and probable transient ischemic attacks. He had a bypass surgery [coronary artery bypass graft (CABG)] in December 2005 and a stroke in December 2006.

Note that a basic tenet of our research in eldercare, which is continually reinforced by the Nursing component of our team, is that, while there may be some common baselines, processing must be individualized for the conditions of each resident. The overall goal of resident monitoring is to perform change detection to signal the early onset of health problems. Hence, the approach to compare summaries from one time period with another certainly generalizes to other residents. In fact, this framework is not restricted to our particular application. It is a general method to create sets of human-centric features from complex time series and to compare the different sets for overall changes in the underlying conditions that generate the sensor readings.

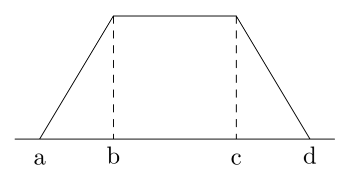


Fig. 2. Trapezoidal fuzzy membership function used in the numeric examples denoted by Trap$[a, b, c, d]$.

The time series analyzed here come from two sensors only: bed restlessness and bedroom motion during the night (from 9 P.M. to 7 A.M.). They are plotted in Fig. 1. Some data are missing, like in November 2005 or from mid November to mid December 2006. Notice that in February 2006 as well as in January 2007, there are longer periods with no restlessness sensor firings. Nursing care coordinators determined that the resident did not sleep in bed during these times; in fact, for some of these dates, he was not present either in the hospital or staying with family. The motion sensor firings on those days could be caused by housekeeping.

For every night, we generate a set of linguistic summaries of the sensor firings in 15-min slots [13], [14].

All the linguistic values are modeled with trapezoidal membership functions, since they are sufficient in most applications. Moreover, they can be very easily interpreted and defined by a person not familiar with fuzzy sets and fuzzy logic, for instance, a health care provider.

To represent a fuzzy set with a trapezoidal membership function, we need to store four numbers: $a$, $b$, $c$, and $d$. An example of such a function Trap$[a, b, c, d]$ is shown in Fig. 2.

We use five linguistic quantifiers: *almost all* (Trap$[0.9, 0.95, 1, 1]$), *most* (Trap$[0.7, 0.8, 1, 1]$), *many* (Trap$[0.6, 0.7, 1, 1]$), *about a half* (Trap$[0.3, 0.45, 0.55, 0.7]$), and *a few* (Trap$[0.1, 0.2, 0.3, 0.45]$). To

TABLE II
DISTANCE BETWEEN LINGUISTIC SUMMARIES FROM THE PROTOTYPES AND A SINGLE NIGHT, JULY 30, 2006

| | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ | $p_7$ |
|---|---|---|---|---|---|---|---|
| $s_1$ | 1 | 1 | 1 | 0.96 | 0.1 | 0.96 | 1 |
| $s_2$ | 1 | 1 | 1 | 0.96 | 0.96 | 0 | 1 |
| $s_3$ | 0.12 | 0.93 | 0.96 | 0.96 | 1 | 1 | 0.99 |
| $s_4$ | 0.96 | 1 | 0.96 | 0.25 | 0.96 | 0.96 | 1 |
| $s_5$ | 0.99 | 1 | 1 | 1 | 1 | 1 | 0 |

describe motion and restlessness, we use three linguistic values: *low* (Trap$[0, 0, 2, 5]$), *medium* (Trap$[2, 5, 12, 15]$), and *high* (Trap$[12, 15, 50, 50]$). After generating all possible linguistic summaries, we include only those with truth value higher than 0.7 and degree of focus higher than 0.1.

Every night is compared with a "prototype" set of summaries for this resident representing a typical night from his stable time. A method to automatically generate the prototypes was proposed in [24]. For this resident, the prototypes were produced from linguistic summaries over the period from May 11 to June 10, 2006 and consist of the following:

1) $p_1$: *many* 15-min intervals have *low* restlessness; $\mathcal{T} = 1.0$ and $d_{\mathrm{foc}} = 1.0$;
2) $p_2$: *most* 15-min intervals of *medium* bedroom motion have *medium* restlessness; $\mathcal{T} = 1.0$ and $d_{\mathrm{foc}} = 0.12$;
3) $p_3$: *most* 15-min intervals of *low* bedroom motion have *low* restlessness; $\mathcal{T} = 0.96$ and $d_{\mathrm{foc}} = 0.9$;
4) $p_4$: *many* 15-min intervals have *low* restlessness and *low* bedroom motion; $\mathcal{T} = 1.0$ and $d_{\mathrm{foc}} = 1.0$;
5) $p_5$: *almost all* 15-min intervals of *low* restlessness have *low* bedroom motion; $\mathcal{T} = 1.0$ and $d_{\mathrm{foc}} = 0.72$;
6) $p_6$: *most* 15-min intervals have *low* bedroom motion; $\mathcal{T} = 1.0$ and $d_{\mathrm{foc}} = 1.0$;
7) $p_7$: *a few* 15-min intervals have *medium* restlessness; $\mathcal{T} = 1.0$ and $d_{\mathrm{foc}} = 1.0$.

We examine in detail the evaluation of similarity between prototype summaries (set $P$) and summaries from the night of July 30, 2006, as well as from his stable time (set $S$). This night is described by the following five linguistic summaries:

1) $s_1$: *almost all* 15-min intervals of *low* restlessness have *low* bedroom motion; $\mathcal{T} = 1.0$ and $d_{\mathrm{foc}} = 0.63$;
2) $s_2$: *most* 15-min intervals have *low* bedroom motion; $\mathcal{T} = 1.0$ and $d_{\mathrm{foc}} = 1.0$;
3) $s_3$: *many* 15-min intervals have *low* restlessness; $\mathcal{T} = 0.88$ and $d_{\mathrm{foc}} = 1.0$;
4) $s_4$: *many* 15-min intervals have *low* restlessness and *low* bedroom motion; $\mathcal{T} = 0.75$ and $d_{\mathrm{foc}} = 1.0$;
5) $s_5$: *a few* 15-min intervals have *medium* restlessness; $\mathcal{T} = 1.0$ and $d_{\mathrm{foc}} = 1.0$.

For the set of summaries $S$, $\lambda_S = 0.59$, and for the set of prototypes $P$, $\lambda_P = 0.48$. The distances between the individual summaries are shown in Table II. Thus, for instance, $S_{\alpha=0} = \{s_2, s_5\}$, $S_{\alpha=0.12} = \{s_1, s_2, s_3, s_5\}$, and $P_{\alpha=0.25} = \{p_1, p_4, p_6, p_7\}$.

Fig. 3 depicts the measure $g$ [see (10)] between set of prototypes $P$ and set of linguistic summaries $S$ with respect to $\alpha$.

The Sugeno integral in this case is equal to 0.75, which is a reasonable value of similarity between these two sets of summaries since all five summaries in $S$ have close (or exact) counterparts in $P$, although $P$ contains two additional summaries.

Repeating those calculations for every night, we obtain the graph in Fig. 4, which shows the fuzzy measure similarity between the prototype set and summaries from each night during the investigated time.

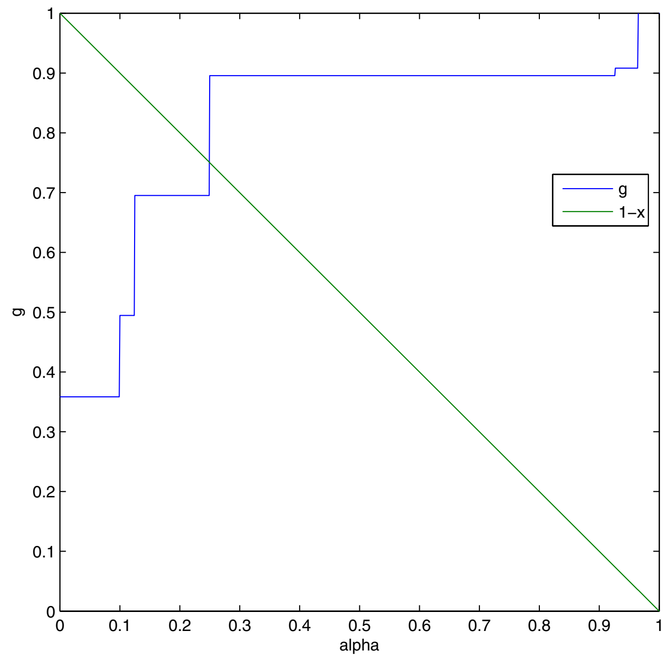


Fig. 3. Fuzzy measure $g$ between prototype summaries and those of July 30, 2006 with respect to $\alpha$. The Sugeno integral is the intersection of the fuzzy measure curve and the function $1 - \alpha$.

We divided those similarities into three crisp categories: different nights (with similarity value less than 0.5), a bit similar (with similarity value between 0.5 and 0.7), and similar (with similarity value higher than 0.7). In Fig. 5, we plotted those category assignments. We clearly see that the nights after the CABG period are different from those in the stable time. In addition, the nights after November 2006 start differing more frequently from the normal and could indicate a future health problem. In this retrospective case, we know that the "future" event was a stroke.

As a comparison, we used the maximum and minimum similarity between the groups of summaries as a method of aggregation. Neither was able to capture the known conditions in our examples: The maximum almost always overestimated the similarity of the two sets (it just needs one summary in each group that is the same), while minimum underestimates for the opposite reason. The fuzzy measure/integral method allows us to capture the richness of the two sets of summaries and is insensitive to the fact that there are usually different numbers of summaries in each group.

## V. CONCLUDING REMARKS

Linguistic summaries are an excellent tool for human/machine interfaces. However, beyond that, linguistic summaries are compact feature representations of large amounts of sensor data. In this paper, we have considered a fuzzy measure-based similarity evaluation between the sets of linguistic summaries as a basis for automated decision making. The underpinning of the calculation is a distance metric between individual fuzzy protoform summaries. The fuzzy measure similarity itself could be used directly, but since the initial set similarity was a function of a level cut, an aggregation operator was proposed: the Sugeno fuzzy integral. We demonstrated the utility of our method in an eldercare environment to distinguish between "good nights" and "bad nights" for a resident, showing the potential for this method to detect patient anomalies. Of course, there is much more to be done. Future work

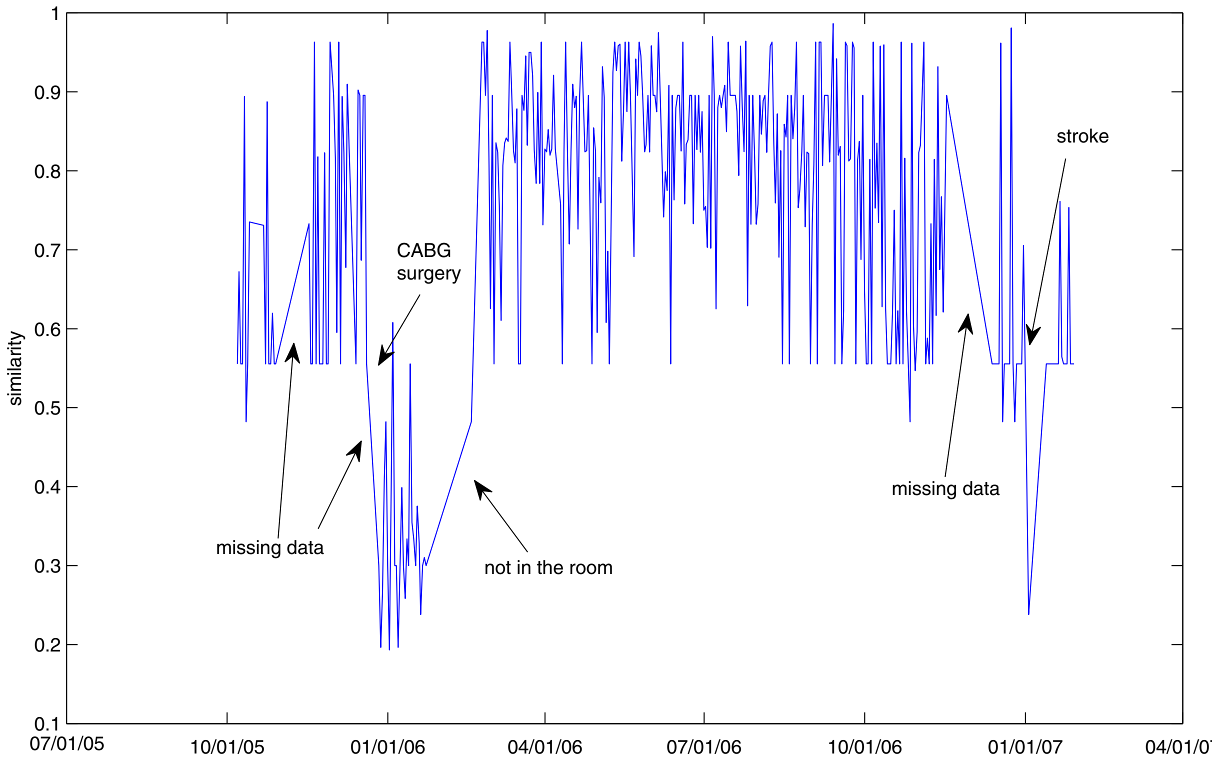


Fig. 4. Similarity between each night's description and the prototype set.

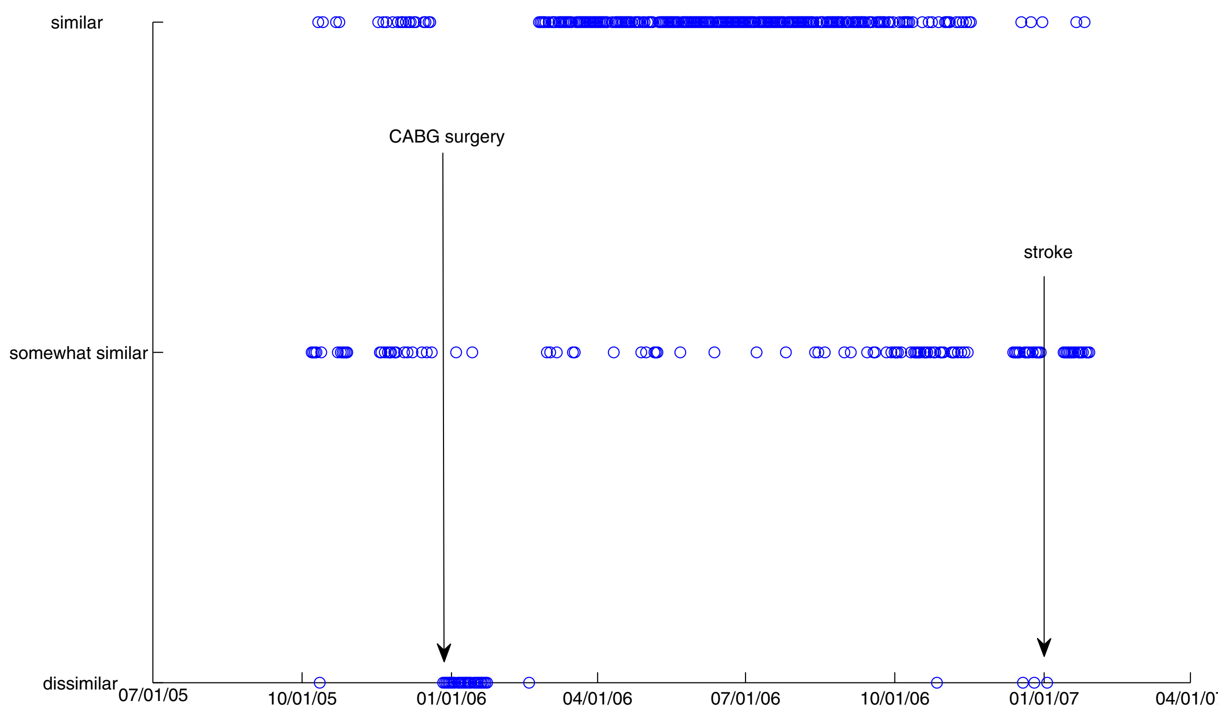


Fig. 5. Category assignment for summaries from each night as compared with the prototype set of summaries.

will include, for instance, incorporating other sensors, generating better, more meaningful summaries for health care providers, automatic anomaly detection, and temporal clustering for long-term monitoring. Since the aggregation method turned out to be the Sugeno integral, it opens the way to use automated techniques to learn the best measure for particular questions about time-series similarities.

## REFERENCES


[1] M. J. Rantz, R. T. Porter, D. Cheshier, D. Otto, C. H. S. III, R. A. Johnson, M. Aud, M. Skubic, H. Tyrer, Z. He, G. Demiris, G. L. Alexander, and G. Taylor, "Tigerplace, a state-academic-private project to revolutionize traditional long term care," *J. Housing Elderly*, vol. 22, no. 1, pp. 66–85, 2008.

[2] M. Skubic, G. L. Alexander, M. Popescu, M. Rantz, and J. M. Keller, "A smart home application to eldercare: Current status and lessons learned," *Technol. Health Care*, vol. 17, no. 3, pp. 183–201, 2009.

[3] A. Wilbik, J. M. Keller, and G. L. Alexander, "Linguistic summarization of sensor data for eldercare," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2011, pp. 2595–2599.

[4] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, and M. Aud, "Linguistic summarization of video for fall detection using voxel person and fuzzy logic," *Comput. Vis. Image Understanding*, vol. 1, no. 113, pp. 80–89, 2009.

[5] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, and M. Aud, "Modeling human activity from voxel person using fuzzy logic," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 17, pp. 39–49, Feb. 2009.

[6] P. Bosc, D. Dubois, O. Pivert, H. Prade, and M. D. Calmes, "Fuzzy summarization of data using fuzzy cardinalities," in *Proc. Int. Conf. Inf. Process. Manage. Uncertainty Knowl.-Based Syst.*, 2002, pp. 1553–1559.

[7] D. Dubois and H. Prade, "Gradual rules in approximate reasoning," *Inf. Sci.*, vol. 61, pp. 103–122, 1992.

[8] G. Raschia and N. Mouaddib, "SAINTETIQ: A fuzzy set-based approach to database summarization," *Fuzzy Sets Syst.*, vol. 129, pp. 137–162, 2002.

[9] D. Rasmussen and R. R. Yager, "Finding fuzzy and gradual functional dependencies with SummarySQL," *Fuzzy Sets Syst.*, vol. 106, pp. 131–142, 1999.

[10] M. Ros, M. Pegalajar, M. Delgado, A. Vila, D. T. Anderson, J. M. Keller, and M. Popescu, "Linguistic summarization of long-term trends for understanding change in human behavior," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2011, pp. 2080–2087.

[11] R. R. Yager, "A new approach to the summarization of data," *Inf. Sci.*, vol. 28, pp. 69–86, 1982.

[12] D. Wu and J. M. Mendel, "Linguistic summarization using if–then rules and interval type-2 fuzzy sets," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 1, pp. 136–151, Feb. 2011.

[13] A. Wilbik and J. M. Keller, "A distance metric for a space of linguistic summaries," *Fuzzy Sets Syst.*, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.fss.2012.03.010

[14] A. Wilbik, J. M. Keller, and G. L. Alexander, "Similarity evaluation of sets of linguistic summaries," *Int. J. Intell. Syst.*, vol. 27, pp. 926–938, Oct. 2012.

[15] M. Popescu, J. M. Keller, and J. A. Mitchell, "Fuzzy measures on the gene ontology for gene product similarity," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, vol. 3, no. 3, pp. 263–274, Jul.–Sep. 2006.

[16] R. R. Yager, "Set measure directed multi-source information fusion," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 6, pp. 1031–1039, Dec. 2011.

[17] J. Kacprzyk, A. Wilbik, and S. Zadrożny, "An approach to the linguistic summarization of time series using a fuzzy quantifier driven aggregation," *Int. J. Intell. Syst.*, vol. 25, no. 5, pp. 411–439, 2010.

[18] J. Kacprzyk, A. Wilbik, and S. Zadrożny, "Linguistic summarization of time series using a fuzzy quantifier driven aggregation," *Fuzzy Sets Syst.*, vol. 159, no. 12, pp. 1485–1499, 2008.

[19] A. Wilbik, "Linguistic summaries of time series using fuzzy sets and their application for performance analysis of investment funds," Ph.D. dissertation, Syst. Res. Inst., Polish Acad. Sci., Warsaw, Poland, 2010.

[20] G. J. Klir and B. Yuan, Eds., *Fuzzy Stes and Fuzzy Logic, Theory and Applications*. Englewood Cliffs, NJ: Prentice–Hall, 1995.

[21] P. Jaccard, "The distribution of the flora in the alpine zone," *New Phytologist*, vol. 11, no. 2, pp. 37–50, 1912.

[22] M. Grabisch, T. Murofushi, and M. Sugeno, Eds., *Fuzzy Measures and Integrals. Theory and Applications*. Heidelberg, Germany: Physica Verlag, 2000.

[23] M. Sugeno, "Theory of fuzzy integrals and applications," Ph.D. dissertation, Tokyo Inst. Technol., Tokyo, Japan, 1974.

[24] A. Wilbik, J. M. Keller, and J. C. Bezdek, "Generation of prototypes from sets of linguistic summaries," in *Proc. IEEE World Congr. Comput. Intell.*, 2012, pp. 472–479.