# Linguistic summarization of sensor data for eldercare

Anna Wilbik
Electrical and Computer Engineering
University of Missouri
Columbia, MO, USA
Systems Research Institute
Polish Academy of Sciences, Warsaw, Poland
wilbik@ibspan.waw.pl

James M. Keller
Electrical and Computer Engineering
University of Missouri
Columbia, MO, USA
KellerJ@missouri.edu

Gregory Lynn Alexander
Sinclair School of Nursing
University of Missouri
Columbia, MO, USA
AlexanderG@missouri.edu

***Abstract*—Ubiquitous passive, as well as active, monitoring of elders is a growing field of research and development with the goal of allowing seniors to live safe active independent lives with minimal intrusion. Much useful information, fall detection, fall risk assessment, activity recognition, early illness detection, etc. can be inferred from the mountain of data. Healthcare must be human centric and human friendly, and so, methods to consolidate the data into linguistic summaries for enhanced communication and problem detection with elders, family and healthcare providers is essential. Long term trends can be most easily identified using summarized information. This paper explores the soft computing methodology of protoforms to produce linguistic summaries of one dimensional data, motion and restlessness. The technique is demonstrated on a 15 month sensor collection for an elder participant.**

***Index Terms*—Linguistic summarization, fuzzy logic, computing with words, eldercare technology, sleep restlessness, electronic health records**

## I. Introduction

Adults want to remain healthy and independent during their senior years as long as possible. The Aging in Place (AIP) model enables older adults to remain in the same environment and provides services and care to meet residents' increasing needs [1]. One example of AIP is TigerPlace (www.tigerplace.net), an independent living environment, at the University of Missouri  Columbia. Tigerplace is cooperative project between by Americare and the Sinclair School of Nursing. Moreover, TigerPlace employs smart sensor technology for the continuous monitoring of residents' activity for the assessment of their "well-being". The sensor suite includes many passive sensors detecting motion, pulse, respiration, restlessness, location and activity [2], [3], [4], [5]. Many residents in TigerPlace have been monitored for functional activity levels over 5 years [2], [6], [1], providing a wealth of data and research opportunities. Additionally, much initial research has been performed, with systems slated to be installed in 2011, on multiple silhouette extraction video cameras, acoustic arrays, and radar sensors [2]. The newest technology involves multiple stereo pair cameras and the kinect sensors. The reader is referred to http://eldertech.missouri.edu for details on the many projects and technology being focused on the problem of providing privacy while preserving safe independent living for elders.

A linguistic summary of data is a concise, human consistent description in (quasi-)natural language that subsumes the very essence of the data. Within the intersection of Eldecare and soft computing, the work by Anderson et al. [7], [8] represents the best to-date contribution. It utilizes a hierarchical system of fuzzy rule bases to first produce temporal fuzzy state membership curves from features generated from a 3D silhouette-based representation, called voxel person, and then to successively combine those curves to create higher level constructs as activity summaries. The efficacy of this approach was shown in the ability to recognize falls, one of the principle drivers for equipping apartments with sensors.

However, installing multiple video sensors is still in its infancy. Useful summaries of activities can be extracted from the simpler suite of sensors that have been deployed for many years. In this paper we will apply a new approach based on the concept of protoforms [9]. We employ the approach of linguistic summarization introduced by Yager [10]. It was further extended and presented in an implementable form by Kacprzyk and Yager [11] and Kacprzyk, Yager and Zadrożny [12], [13]. According to this approach numerical data can be summarized and presented in the form of natural language like sentences as e.g., "most of employees are young" or "most of young employees earn low salary", which are easily derived and interpreted using Zadeh's fuzzy logic based calculus of linguistically quantified propositions [14]. Such constructs are called protoforms. Later this approach was successfully used for past performance analysis of mutual funds, in which segments (representing linear trends) were summarized [15].

In this paper we propose to apply the approach of linguistic summaries for the data concerning restlessness of elder residents. We use only two types of sensors: sensors detecting movements in bed, called "restlessness" and sensors detecting movements around an apartment, indicating that the person is out of the bed. Electronic signals generated by the sensor system correspond to human activity or motion around the apartment. Our data contain numbers of each type sensor firing per night in each room of the apartment. Hence, we have introduced new protoforms, which can be easily understood by the medical personnel, and which may be exemplified by the following linguistic summaries: "on most nights the resident had a medium level of restlessness" or "in last few weeks on most nights, when the resident had a medium level of restlessness, he had a low level of motion".

We will present also an example of real human activity analyzed against some medical records for that resident. This example shows how important and useful such summaries may be for providers evaluating the functional status of a resident.

## II. Linguistic summaries of data

Linguistic summaries of data are meant as usually short sentences in quasi-natural language that capture the very essence of the numeric, large set of data. We use the notation of Yager's basic approach [10], which is also used in later papers on this topic:

- $Y = \{y_1, y_2, \ldots, y_n\}$ is the set of objects (records) in the database $D$, e.g., a set of residents.
- $A = \{A_1, A_2, \ldots, A_m\}$ is the set of attributes (features) characterizing objects from $Y$, e.g., restlessness, room motion.

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. low for attribute restlessness);
- 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., high for attribute motion).

Thus, the core of a linguistic summary is a linguistically quantified proposition in the sense of Zadeh [14] which may be written, respectively, as

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

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

which may be exemplified, respectively by: "Most of nights the resident had low restlessness" $\mathcal{T} = 0.7$, or "Most of nights with high motion the resident had low restlessness", $\mathcal{T} = 0.82$.

## III. Linguistic summaries of the residents' restlessness

In our approach we summarize the number of sensor firings at a given time unit. We have two types of sensor firings: restlessness, fired if movement in bed was detected, and motion, which is fired when resident movement out of bed and in other areas of the apartment was detected.

We have the following protoforms of the linguistic summaries in order to describe the restlessness of the resident:

- simple classic summary:

$$\text{On } Q \text{ of } y\text{'s the resident had } P \tag{3}$$

e.g., "On most of the nights the resident had a medium level of restlessness".

- extended classic summary:

$$\text{On } Q \text{ of } y\text{'s, when the resident had } R, \text{ he had also } P \tag{4}$$

e.g., "On most of the nights, when the resident had a medium level of restlessness, he had a low level of motion".

- simple temporal summary:

$$E_T \text{ on } Q \text{ of } y\text{'s the resident had } P \tag{5}$$

e.g., "Recently on most of the nights the resident had a medium level of restlessness"

- extended temporal summary:

$$E_T \text{ on } Q \text{ of } y\text{'s, when the resident had } R, \text{ he had also } P \tag{6}$$

e.g., "Recently on most of the nights, when the resident had a medium level of restlessness, he had a low level of motion".

where $P$ is the summarizer, $Q$ is the linguistic quantifier, $R$ is the qualifier and $E_T$ is temporal expression. Needless to say, all of the linguistic terms and qualifiers are modeled by fuzzy sets over suitable domains. They are all represented by membership functions, labeled $\mu$ with appropriate subscripts.

We compute the truth value of the linguistic summaries using the Zadeh's [14] calculus of linguistically quantified propositions. It determines the degree to which a linguistically quantified proposition equated with a linguistic summary is true. It is computed depending on the type of the protoform as:

- simple classic summary:

$$\mathcal{T}(\text{On } Q \text{ of } y\text{'s the resident had } P) = \mu_Q\left(\frac{1}{n}\sum_{i=1}^{n}\mu_P(y_i)\right) \tag{7}$$

- extended classic summary:

$$\mathcal{T}(\text{On } Q \text{ of } y\text{'s, when the resident had } R, \text{ he had also } P) = \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_P(y_i)\wedge\mu_R(y_i)}{\sum_{i=1}^{n}\mu_R(y_i)}\right) \tag{8}$$

- simple temporal summary:

$$\mathcal{T}(E_T \text{ on } Q \text{ of } y\text{' s the resident had } P) = \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_P(y_i)\wedge\mu_{E_T}(y_i)}{\sum_{i=1}^{n}\mu_{E_T}(y_i)}\right) \tag{9}$$

- extended temporal summary:

$$\mathcal{T}(E_T \text{ on } Q \text{ of } y\text{'s, when the resident had } R, \text{ he had also } P) = \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_P(y_i)\wedge\mu_R(y_i)\wedge\mu_{E_T}(y_i)}{\sum_{i=1}^{n}\mu_R(y_i)\wedge\mu_{E_T}(y_i)}\right) \tag{10}$$

where $n$ is the number of the summarized instances, in our case nights, $Q$ is a fuzzy set representing the linguistic quantifier in the sense of Zadeh [14], i.e. regular, nondecreasing and monotone and $\wedge$ is the minimum (or a $t$-norm, cf. Kacprzyk et al [16]).

As the second quality criteria we use the degree of focus [17]. The degree of focus measures how many trends fulfill property $R$. The degree of focus makes sense for the extended protoform summaries only, and is calculated as:

$$d_{foc}(\text{On } Q \text{ of } y\text{'s, when the resident had } R, \text{ he had also } P) = \frac{1}{n}\sum_{i=1}^{n}\mu_R \tag{11}$$

$$d_{foc}(E_T \text{ on } Q \text{ of } y\text{'s, when the resident had } R, \text{ he had also } P) = \frac{\sum_{i=1}^{n}\mu_R(y_i) \wedge \mu_{E_T}(y_i)}{\sum_{i=1}^{n}\mu_{E_T}(y_i)} \tag{12}$$

The very essence of the degree of focus is to give the proportion of instances, in our case nights, satisfying property $R$ to all instances. It provides a measure that, in addition to the basic truth value, can help control and speed up the process of generating linguistic summaries. If the degree of focus is high, then we can be sure that such a summary is more general. However, if the degree of focus is low, we may be sure that such a summary describes a (local) pattern seldom occurring. More information on linguistic summaries can be found in [18].

## IV. Example

We show linguistic summaries generated over a 15 month period for a male resident, 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.

In Fig. 1 we display the plot of the nighttime sensor firings for both types of sensors: bed restlessness, and motion, which illustrates bed movement and movement around the apartment during every day. Some data are missing, like in November 2005 or from mid November till mid December 2006. Obviously sensors may create noisy, unreliable data, however by the fact that we are using fuzzy sets with low granularity to model linguistic values like “low”, “high”, etc., our method is somewhat robust to the issues of sensor reliability. Notice that in February 2006 as well as in January 2007, there are longer periods with no restlessness sensor firings. Nursing care coordinators determined the resident did not sleep in bed during these times; in fact, some of these dates he was not present and was admitted to the hospital or was staying with family. The motion sensor firings on those days could be caused by housekeeping.

We describe each attribute (one for each type of sensor) with three linguistic values low level, medium level and high level. Low level of restlessness is up to about 75 sensor firings per night. Medium level of restlessness is around 75 - 125 sensor firings per night. Higher values are described by high level of restlessness. Similarly for the motion, low level of motion is considered to be up to about 150 sensor firings per night, medium level of motion is around 150 - 200 sensor firings per night, and higher values are described by high level of motion.

We have also used several expressions, referring to some periods based on the resident’s medical history. They were in chronological order:

- initially - describing the first collected data until November 2005 (where there were no data),
- before CABG - referring to the one month period before the surgery in December 2005,
- after CABG - referring to the almost 2-month period after the surgery,
- stable time - referring to the 9-month period when no serious health events occurred,
- during the most recent health event when the resident suffered a stroke and post-stroke - referring to the timeframe when the resident experienced a stroke and post-stroke.

Some of the linguistic summaries together with their truth values ($\mathcal{T}$) and the degree of focus ($d_foc$), we have obtained are the following:

- On most of the nights the resident had a medium level of restlessness. ($\mathcal{T}$=0.85, $d_{foc}$=1.0)
- On most of the nights, when the resident had a medium level of restlessness, he had also a medium level of motion. ($\mathcal{T}$=0.88, $d_{foc}$=0.73)
- On most of the nights, when the resident had a medium level of motion, ha had also a medium level of restlessness. ($\mathcal{T}$=1.0, $d_{foc}$=0.64)

The above 3 summaries describe the most often occurring level of restlessness and motion for the resident and all fall within the medium level. Medium level of restlessness means that several movements in bed were detected, and medium level of motion may be combined with going to bed and getting up and maybe a few bath visits. However more detailed data in this respect will be needed.

- Initially on most of the nights the resident had a medium level of restlessness. ($\mathcal{T}$=0.93, $d_{foc}$=1.0)

In the period described by the linguistic term initially we could observe that the resident had a medium level of restlessness. In his medical history no special health events were discovered.

- Before CABG, on most of the nights the resident had a high level of motion. ($\mathcal{T}$=1.0, $d_{foc}$=1.0)
- Before CABG, on most of the nights the resident had a medium level of restlessness. ($\mathcal{T}$=1.0, $d_{foc}$=1.0)
- Before CABG, on most of the nights, when the resident had a high level of motion, he had also a medium level of restlessness. ($\mathcal{T}$=1.0, $d_{foc}$=0.82)

One month before the surgery we may notice that there was an increase in the level of motion detected at night. It can suggest that the patient was out of bed at nights more often or for longer periods. The restlessness stayed on the same level, i.e., medium.

- After CABG, on most nights the resident had a high level of restlessness. ($\mathcal{T}$=0.79, $d_{foc}$=1.0)

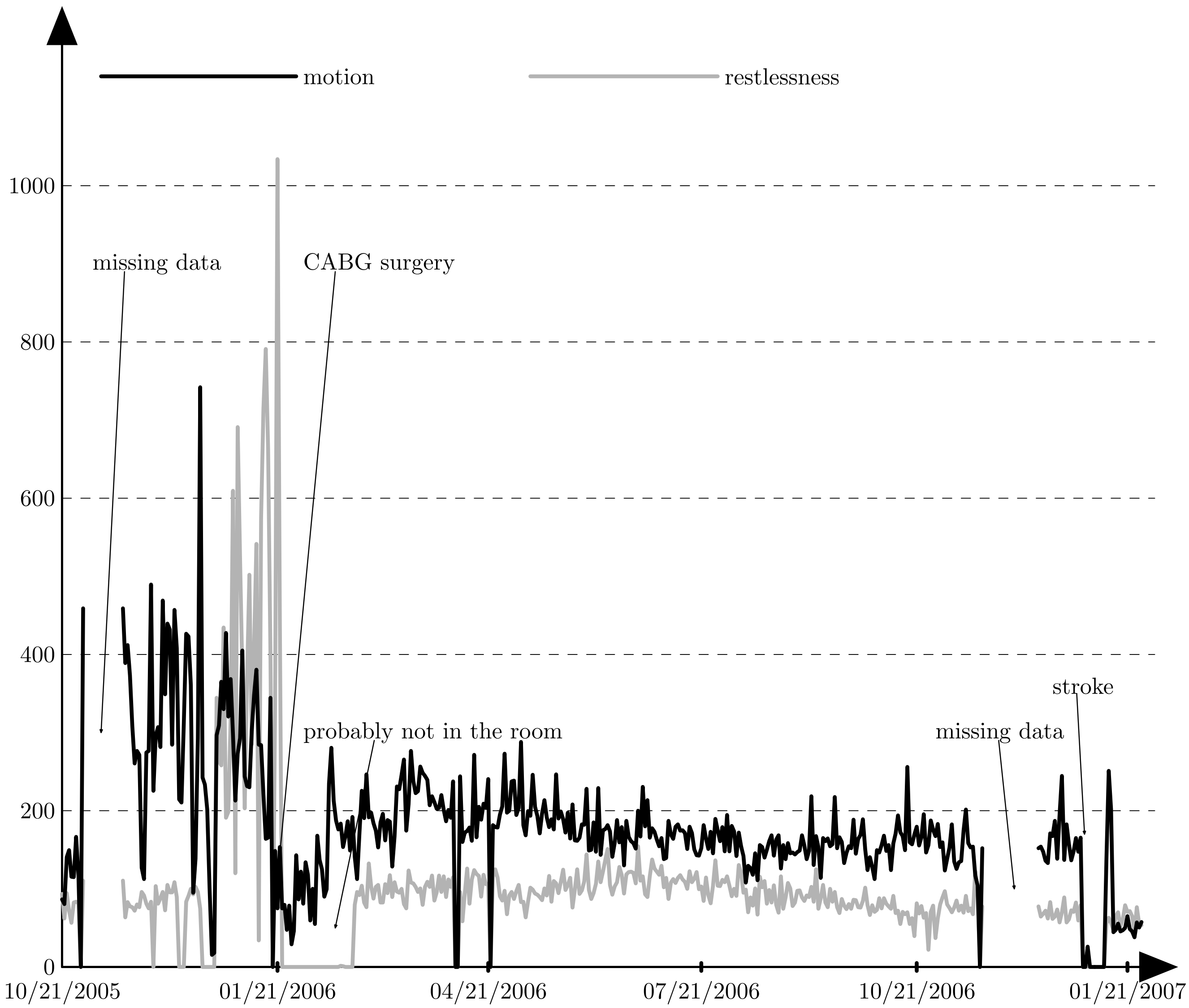


Fig. 1. The nighttime sensor firings for both types of sensors: bed restlessness, and motion

- 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)
- 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)
- 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)

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.

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

After the surgery and the rehabilitation the resident returned to his normal level of restlessness and motion detected at night.

- During the most recent health event when the resident suffered a stroke and post-stroke, on most nights the resident had a low level of motion. ($\mathcal{T}$=0.82, $d_{foc}$=1.0)

- During the most recent health event when the resident suffered a stroke and post-stroke, on most of the nights, when the resident had a low level of restlessness, he had also a low level of motion. ($\mathcal{T}$=0.99, $d_{foc}$=0.59)
- During the most recent health event when the resident suffered a stroke and post-stroke, on most of the nights, when the resident had a low level of motion, he had also a low level of restlessness. ($\mathcal{T}$=0.73, $d_{foc}$=0.71)

After the stroke the resident might have lost some part of his fitness and therefore fewer movements were detected.

## V. Concluding remarks

In this paper, we applied a novel methodology of fuzzy logic-based protoforms to create linguistic summaries from long duration sensor data in an eldercare environment. Results were successfully demonstrated on a 15 month period for an elderly resident who experienced several life threatening medical conditions.

We believe that this approach can be further extended and used for the detection of anomalies. The summaries presented above, especially of the classical protoform, can suggest what may be considered as normality, and hence the deviations from such descriptions should be looked for. Moreover this approach is employing natural language, and hence can be more comprehensible for the medical personnel.

## Acknowledgment

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

## References

[1] M. J. Rantz, M. Skubic, S. J. Miller, and J. Krampe, "Using technology to enhance aging in place," in *Proceedings of the 6th international conference on Smart Homes and Health Telematics (ICOST '08)*, S. Helal, S. Mitra, J. Wong, C. K. Chang, and M. Mokhtari, Eds. Springer-Verlag Berlin, Heidelberg, 2008, pp. 169–176.

[2] G. L. Alexander, D. Anderson, M. Popescu, J. M. Keller, M. Skubic, and M. Rantz, "Applying fuzzy logic and linguistic summarization for monitoring functional decline of elders," in *Proceedings of XXIII International Conference of the European Federation for Medical Informatics*, 2011.

[3] 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," *Computer Vision and Image Understanding*, vol. 1, no. 113, pp. 80–89, 2009.

[4] ——, "Modeling human activity from voxel person using fuzzy logic," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 17, pp. 39–49, 2009.

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

[6] M. J. Rantz, M. Skubic, G. L. Alexander, M. Popescu, M. A. Aud, B. J. Wakefield, R. J. Koopman, and S. J. Miller, "Developing a comprehensive electronic health record to enhance nursing care coordination, use of technology, and research," *Journal of Gerontological Nursing*, vol. 36(1), 2010.

[7] D. Anderson, R. H. Luke, E. Stone, and J. M. Keller, "Segmentation and linguistic summarization of voxel environments using stereo vision and genetic algorithms," in *Proceedings IEEE International Conference on Fuzzy Systems, World Congress on Computational Intelligence*, 2010, pp. 2756–2763.

[8] D. Anderson, R. H. Luke, and J. M. Keller, "Fuzzy voxel model for human activity analysis," in *Proceedings of IFSA/EUSFLAT 2009*, 2009, pp. 282–287.

[9] L. A. Zadeh, "A prototype-centered approach to adding deduction capabilities to search engines – the concept of a protoform," in *Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS 2002)*, 2002, pp. 523–525.

[10] R. R. Yager, "A new approach to the summarization of data," *Information Sciences*, vol. 28, pp. 69–86, 1982.

[11] J. Kacprzyk and R. R. Yager, "Linguistic summaries of data using fuzzy logic," *International Journal of General Systems*, vol. 30, pp. 33–154, 2001.

[12] J. Kacprzyk, R. R. Yager, and S. Zadrożny, "A fuzzy logic based approach to linguistic summaries of databases," *International Journal of Applied Mathematics and Computer Science*, vol. 10, pp. 813–834, 2000.

[13] J. Kacprzyk and S. Zadrożny, "Linguistic database summaries and their protoforms: toward natural language based knowledge discovery tools," *Information Sciences*, vol. 173, pp. 281–304, 2005.

[14] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets and Systems*, vol. 9, no. 2, pp. 111–127, 1983.

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

[16] ——, "Linguistic summarization of time series under different granulation of describing features," in *Rough Sets and Intelligent Systems Paradigms - RSEISP 2007*, M. Kryszkiewicz, J. F. Peters, H. Rybinski, and A. Skowron, Eds. Springer-Verlag, Berlin and Heidelberg, 2007, pp. 230–240.

[17] J. Kacprzyk and 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] A. Wilbik, "Linguistic summaries of time series using fuzzy sets and their application for performance analysis of investment funds," Ph.D. dissertation, Systems Research Institute, Polish Academy of Sciences, 2010.