Zeszyty Naukowe Wydziału Informatyki
Wyższej Szkoły Informatyki Stosowanej i Zarządzania
„Informatyka Stosowana"
Nr 1/2014

LINGUISTIC SUMMARIES OF TIME SERIES: A POWERFUL TOOL FOR DISCOVERING KNOWLEDGE ON TIME VARYING PROCESSES AND SYSTEMS

Janusz Kacprzyk, Fellow of IEEE, and Sławomir Zadrożny

Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01–447 Warsaw, Poland
and
Warsaw School of Information Technology
ul. Newelska 6, 01–447 Warsaw, Poland
e-mail: {janusz.kacprzyk,slawomir.zadrozny}@ibspan.waw.pl

Abstract

We present linguistic data summarization, meant as a process for a comprehensive description of big and complex data sets via short statements in natural language represented by protoforms in the form of linguistically quantified propositions dealt with using tools and techniques of fuzzy logic to grasp an inherent imprecision of natural language. Such linguistic data summaries can provide a human user, whose only natural means of articulation and communication is natural language, with a simple yet effective and efficient means for the representation and manipulation of knowledge about processes and systems. We concentrate on the linguistic summarization of dynamic processes and systems, dealing with data represented as time series. We extend the basic, static data oriented concept of a linguistic data summary to the case of time series data, present various possible protoforms of linguistic summaries, and an analysis of their properties and ways of generation. We show two our own real applications of the new tools of linguistic summarization of time series, for the summarization of quotations of an investment (mutual) fund, and of Web server logs, to show the power of the tool. We also mention some other applications known from the literature. We conclude with some remarks on the strength of the linguistic summarization for broadly perceived data mining and knowledge discovery and some possible further research directions.

Keywords: linguistic summarization, natural language, fuzzy logic, linguistic quantifiers, data mining, knowledge discovery, big data sets

1 Introduction

An overwhelming majority of real processes and problems are dynamic and should be analyzed by taking into account their time varying behavior. This also concerns all kinds of approaches aimed at discovering information and/or knowledge about the processes and problems considered, notably those based on data mining, machine learning, etc.

A human being is more and more often a crucial element of virtually all complex processes and problem solving tasks as the research interest moves towards more and more complex systems. This implies a wider and wider gap between the human, much better in sophisticated tasks but worse in number crunching, and computer, much better in number crunching but much worse in sophisticated analyses. A human-centric systems paradigm initiated by Dertouzos (2001) and developed by, e.g., Pedrycz and Gomide (2007) can be a viable solution. In our context it would boil down to the use of natural language for the representation of that information/knowledge extracted because for the humans natural language is the only fully natural means of communication and articulation.

More specifically, we will deal with *linguistic data summaries* in the sense of Yager (1982), in their more implementable form proposed by Kacprzyk and Yager (2001) and Kacprzyk, Yager and Zadrożny (2000) in which, in the static case, we have a (relational) database that is too large to be comprehended by the human being and therefore we wish to find a short and comprehensible linguistic summary of its contents as a short linguistically quantified proposition in the sense of Zadeh (1983). For instance, in the case of a personnel database, a linguistic summary with respect to "age" and "salary" can be *most of young employees earn low salaries*.

This basic concept of a static linguistic summary has been extended to the dynamic case, specifically to *linguistic summaries of time series* by Kacprzyk, Wilbik and Zadrożny (2006). By first performing a segmentation of the time series data into trends, i.e., parts of the time series exhibiting a uniform behavior, we obtained linguistic summaries like "most of slowly decreasing trends have a large variability", "almost all of trends with a high variability are sharply decreasing", etc. This approach was extended in, e.g., Kacprzyk, Wilbik and Zadrożny (2008, 2010), showing an application to investment (mutual) fund quotations, and in Zadrożny and Kacprzyk (2007a, b) for Web log analyses. Then, many other approaches have been proposed as, e.g., Alvarez et al. (2012), Batyrshin and Sheremetov (2006, 2007), Castillo, Marín and Sánchez (2010, 2011b), Keller et al. (2009, 2011), for various application areas as, e.g., elderly care, traffic analyses, etc.

Obviously, the generation of linguistic summaries, even static not to mention dynamic ones, can be a serious problem and has rarely been addressed except for Kacprzyk and Zadrożny (2001) and their subsequent papers in which the generation is considered in the context of fuzzy database querying, Kacprzyk and Zadrożny's (2005a) use of Zadeh's protoforms representing linguistic summaries, Kacprzyk and Zadrożny's (2013) proposal to derive (generate) linguistic summaries via association rule mining, Kacprzyk and Zadrożny's (2010) proposal to generate linguistic summaries using natural language generation (NLG), and Kacprzyk and Zadrożny's (2010) proposal to generate linguistic summaries by using some elements of systemic functional linguistics (SFL).

We will now briefly present the concept of a linguistic data summary, first of static and then dynamic (time series) data. Then, we will mention two example of applications, for the summarization of investment fund quotations and Web logs.

2 Linguistic data summaries: a static and dynamic case

Yager's (1982) source concept of a *linguistic data summary* concerns: $Y = \{y_1, y_2, \ldots, y_n\}$, the set of objects (records) in the database D as, e.g., a set of employees; $A = \{A_1, A_2, \ldots, A_m\}$, the set of attributes (features) characterizing the objects from Y as, e.g., a salary or age. A linguistic (data) 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 the attribute *salary*);
- a quantity in agreement Q, i.e. a linguistic quantifier (e.g. *most*);
- a truth value (validity) $\mathcal{T}$ of the summary, i.e. a number from $[0, 1]$ yielding 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., *young* for attribute *age*).

Thus, a linguistic summary may be exemplified by the simple form

$$\mathcal{T}(most \text{ of employees earn } low \text{ salary}) = 0.7 \tag{1}$$

or by a richer (extended) form, including a qualifier (e.g. *young*), by

$$\mathcal{T}(most \text{ of } young \text{ employees earn } low \text{ salary}) = 0.82 \tag{2}$$

Basically, the core of a linguistic summary is a linguistically quantified proposition in the sense of Zadeh (1983) which for (1) may be written as

$$Qy's \text{ are } P \tag{3}$$

and for (2) may be written as

$$QRy's \text{ are } P \tag{4}$$

The truth value (validity), $\mathcal{T}$, of the above simple and extended (3) and (4) are then, respectively:

$$\mathcal{T}(Qy's \text{ are } P) = \mu_Q \left(\frac{1}{n} \sum_{i=1}^{n} \mu_P(y_i) \right) \tag{5}$$

$$\mathcal{T}(QRy's \text{ are } P) = \mu_Q \left(\frac{\sum_{i=1}^{n} \mu_P(y_i) \wedge \mu_R(y_i)}{\sum_{i=1}^{n} \mu_R(y_i)} \right) \tag{6}$$

where $\wedge$, the minimum operation, may be replaced by, e.g. a t-norm, and Q is a fuzzy set representing the linguistic quantifier like

$$\mu_Q(x) = \begin{cases} 1 & \text{for } x \geq 0.8 \\ 2x - 0.6 & \text{for } 0.3 < x < 0.8 \\ 0 & \text{for } x \leq 0.3 \end{cases} \tag{7}$$

Other methods of calculating $\mathcal{T}$ can also be used, notably those based on the OWA (ordered weighted averaging) operators (cf. Yager, 1988, 1996; Yager and Kacprzyk, 1997; and Yager, Kacprzyk and Beliakov, 2011), and the Sugeno and Choquet integrals (cf. Bosc and Lietard, 1995 or Grabisch, 1998).

The above approach can be extended to the dynamic case, i.e. to the summarization of time series, as proposed by Kacprzyk, Wilbik and Zadrożny (2006), and then considerably developed in, e.g., Kacprzyk, Wilbik and Zadrożny (2008, 2010), to name a few. Basically, in that approach the linguistic summarization of a time series is performed as the linguistic summarization of the trends (segments) extracted from the time series. First, we assume a piecewise linear representation of time series data, and we extract segments, i.e. the constituent straight lines that represent an uniform behavior of the data. This can be done by using, for instance, on-line (sliding window) algorithms, bottom-up or top-down strategies (cf. Keogh, 2004, 2001) or, as in our works, using a modification of Sklansky and Gonzalez (1980) algorithm.

The following basic features of trends in the time series are considered:

1. dynamics of change,
2. duration, and
3. variability,

meant as: the *dynamics of change* is the speed of change of the consecutive values of the time series which may be described by the slope of a line representing the trend, then represented by a linguistic variable, the *duration* is the length of a single trend, also represented by a linguistic variable and the *variability* describes how "spread out" a group of data within a segment is. The use of a small set of granulated linguistic labels as, e.g.: quickly increasing, increasing, slowly increasing, constant, slowly decreasing, decreasing, quickly decreasing, equated with fuzzy sets, is employed.

Then, we basically employ protoforms (abstract prototypes, or template, of a linguistically quantified propositions) of linguistic summaries as proposed by Kacprzyk and Zadrożny (2005a), exemplified by

- for the simple form:

$$\text{Among all segments, } Q \text{ are } P \tag{8}$$

  e.g.: "Among all segments, *most* are *slowly increasing*".

- for the extended form:

$$\text{Among all } R \text{ segments, } Q \text{ are } P \tag{9}$$

  e.g.: "Among all *short* segments, *most* are *slowly increasing*".

Moreover, we can further enhance the extended protoforms given in (8) and (9) by adding a temporal expression, E_T, like: "recently", "initially", "in the very beginning","in the early Spring of 2010", etc., which yields the *temporal protoforms*:

- for the case of the simple protoform:

$$E_T \text{ among all segments, } Q \text{ are } P \tag{10}$$

  e.g.: "*Recently*, among all segments, *most* are *slowly increasing*".

- for the case of the extended protoform:

$$E_T \text{ among all } R \text{ segments, } Q \text{ are } P \tag{11}$$

  e.g.: "*Initially*, among all *short* segments, *most* are *slowly increasing*".

The truth (validity) of those linguistic summaries is calculated similarly as in the static case, and we obtain for for the simple and extended protoform, respectively:

$$\mathcal{T}(\text{Among all } y\text{'s, } Q \text{ are } P) = \mu_Q\left(\frac{1}{n}\sum_{i=1}^{n}\mu_P(y_i)\right) \tag{12}$$

$$\mathcal{T}(\text{Among all } Ry\text{'s, } Q \text{ are } P) = \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_R(y_i)\wedge\mu_P(y_i)}{\sum_{i=1}^{n}\mu_R(y_i)}\right) \tag{13}$$

where $\wedge$ is the minimum operation or more generally a t-norm.

The computation of truth values of temporal summaries is very similar to the previous case, and – basically – we obtain for the simple temporal protoform summary (10) and for the extended temporal protoform summary (11), respectively:

$$\mathcal{T}(E_T \text{ among all } y\text{'s, } Q \text{ are } P) = \mu_Q\left(\frac{\sum_{i=1}^{n}\mu_{E_T}(y_i)\wedge\mu_P(y_i)}{\sum_{i=1}^{n}\mu_{E_T}(y_i)}\right) \tag{14}$$

where $\mu_{E_T}(y_i)$ is the degree to which a trend (segment) occurs during the time span described by E_T;

$$\mathcal{T}(E_T \text{ among all } Ry\text{'s}, Q \text{ are } P) = \mu_Q \left(\frac{\sum_{i=1}^{n} \mu_{E_T}(y_i) \wedge \mu_R(y_i) \wedge \mu_P(y_i)}{\sum_{i=1}^{n} \mu_{E_T}(y_i) \wedge \mu_R(y_i)} \right) \quad (15)$$

and for technicalities we refer the reader to our papers cited above.

Naturally, this is not the only way to calculate that degree. For instance, we may also use the OWA (ordered weighted averaging) operators (cf. Kacprzyk Wilbik and Zadrożny, 2007). One can also employ some other aggregation methods exemplified by those using the Sugeno (cf. Bosc and Lietard, 1995) or Choquet (cf. Grabisch, 1998) integrals.

Moreover, it should be noticed that the degree of truth has been assumed as the only quality criterion which is natural for Yager's setting. However, many other quality criteria are possible, for instance the degrees of: imprecision, specificity, fuzziness, covering, focus, appropriateness, informativeness, and the length of the summary. Most of them were already suggested in earlier works by Kacprzyk and Yager (2001), Kacprzyk, Yager and Zadrożny (2000), and then extended into the time series summarization, e.g., in Kacprzyk and Wilbik (2010), and details can be found therein.

3 Examples of applications

We will now show two examples of our works on the linguistic summarization of: (1) daily quotations of a mutual (investment) fund that is to invest at least 66% assets in shares listed at the Warsaw, Zagreb and Moscow Stock Exchanges, and (2) Web logs of the institute server. As we will see, different protoforms of lingusitic summaries have been employed which is implied by the specificity of the problem.

3.1 Linguistic summarization of investment (mutual) fund quotations

What concerns the daily quotations of the investment (mutual) fund, the period of January 1, 2002 until the December 31, 2010 was considered, mainly to account for some economic disturbances in Europe in recent years. The value of one share was PLN 12.06 in the beginning and PLN 40.52 at the end of the time span (PLN stands for the Polish Zloty). The minimal value recorded was PLN 9.35 while the maximal one during this period was PLN 57.85. The biggest daily increase was PLN 2.32, while the biggest daily decrease was 3.46. Using the piecewise linear segmentation we obtained 100 segments from 1 day to 191 days long. Almost 50% of segments were shorter than 10 days, with very few segments longer than 50 days.

We used the granulation of 5 labels for each attribute, like very short, short, medium and long, very long for the duration, and similarly for other criteria.

Some examples of the linguistic summaries obtained were:

- the simple summaries:
  - Among all slowly decreasing segments, most are short,
  - Among all long and constant segments, most are of very high variability,
  - Among all very long segments, most are constant,
  - Among all of moderate variability segments, most are short,
  - Among all very long segments, most are constant and of very high variability,
- the summaries which reflect a temporal aspect:
  - Recently, among all slowly decreasing segments, most are short,
  - Recently among all long and of very high variability segments, most are constant,
  - Recently among all short and of very high variability segments, most are slowly decreasing,
  - Recently among all very long and of very high variability segments, most are constant,
  - Recently among all very long segments, most are constant and of very high variability,
  - Recently among all slowly increasing segments, most are of very high variability,

3.2 Linguistic summarization of Web server logs

The second application concerns the linguistic summarization of Web server logs and was proposed by Zadrożny and Kacprzyk (2007a, b). That application was motivated by the fact that Web servers are crucial elements of virtually all IT systems in all kinds of institutions, organizations, companies, etc. and it may be very important to use some advanced tools and techniques for their design and running. For the use of soft computing, see Wang, et. al. (2002, 2005), Pal, et. al. (2002), Abraham (2003), Arotaritei and Mitra (2004), De and Krishna (2004), Asharaf and Murty (2004).

The idea of our approach, from the viewpoint of soft computing, is similar to Abraham (2002, 2003, 2005) who deals, using fuzzy clustering, evolutionary algorithm, neural networks and the Takagi-Sugeno type fuzzy systems, with access trends analysis which is similar in spirit to the dynamic linguistic summaries used by us. Moreover, Shiu et al.'s (2005) approach is also related due to their use of the fuzzy association rules used to generate a subclass of linguistic summaries (cf., our work on that topic Kacprzyk J. and Zadrożny, 2001). However, none of those works uses a human consistent natural language based approach.

Basically, each request to a Web server is recorded in one or more log files that contain the following main fields: the requesting computer name or IP address, the username of the user triggering the request, the user authentication data, the date and time of the request, the HTTP command related to the request which includes the path to the requested file, the status of the request, the number of bytes transferred as a result of the request, the software used to issue the request.

One may use an extended format with more fields but it is beyond the scope of this paper.

By using similar linguistic summarization algorithms, we obtained both static and dynamic summaries, which may be briefly exemplified as follows. First, for the static case:

- for the case of the *simple summaries*:
  - *Most* of the requests come from the Firefox browser
  - *Almost all* requested files are *small*
- for the case of the *extended summaries*:
  - *Almost all* failures concern files with an extension "ppt"
  - *Most* of the requests concerning *large* files happen in the *evening*.

What concerns the dynamic summaries, i.e. concerning time series of requests (their trends), we can quote the following examples:

- for the case of the *simple summaries*:
  - *Most* of the trends concerning the number of requests are *decreasing*
- for the case of the *simple summaries* which account for the duration:
  - Trends concerning the number of requests that took *most time* are *slowly increasing*
- for the case of the *extended summaries* which account for the frequency of event occurence:
  - *Most* of *increasing* trends concerning the number of requests are of *high* variability
- for the case of the *extended summaries* which account for the duration of event occurence:
  - *Increasing* trends concerning the total size of requested files, that took *most* of the time, are *very long*

It is easy to see that in both examples of real applications the linguistic summaries derived have provided much insight into the very essence of the set of data considered. They can be of a valuable help in making decisions.

4 Concluding remarks

We briefly presented the concept of a linguistic summary of numeric data which boils down to a linguistically quantified proposition that is dealt with using fuzzy logic tools and techniques. This makes it possible, first, to provide a short and comprehensive, yet very useful representation of information and knowledge exhibited by the (large set of) data in questions in a short natural language form. Second, via the use of fuzzy logic, an inherent imprecision of natural language can be effectively and efficiently handled. This all is clearly a considerable step towards human consistent and human centric computing that can help bridge an inherent gap between the human being and the computer. We presented linguistic summarization both for static and dynamic (time series) data sets. As a topic for a further in depth research we mentioned our proposals of using tools and techniques of natural language generation (NLG) and systemic functional linguistics (SFL).

We showed two real applications of linguistic summaries of time series data, for daily quotations of an investment mutual fund, and for Web server logs, listing some more interesting linguistic summaries of various types (protoforms) that could provide much insight and information which might be useful for the analysis of the problems considered.

5 Acknowledgments

This work was partially supported by the National Centre for Research (NCN) under Grant No. UMO-2012/05/B/ST6/03068.

References

Abraham A. (2003) Miner: A web usage mining framework using hierarchical intelligent systems, In: *Proceedings of the IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE'03*, 1129–1134.

Alvarez-Alvarez A., Sánchez-Valdes D., Trivi no G., SánchezÁ., Suárez P. D. (2012) Automatic linguistic report of traffic evolution in roads, *Expert Systems with Applications*, 39(12), 11 293–11 302.

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

Arotaritei D., Mitra S. (2004) Web mining: a survey in the fuzzy framework. *Fuzzy Sets and Systems*, 148(1), 5—19.

Asharaf S., Murty M. N. (2004) A rough fuzzy approach to web usage categorization. *Fuzzy Sets and Systems*, 148(1), 119—129.

Batyrshin I., Sheremetov L. (2006) Perception based functions in qualitative forecasting, In: Batyrshin I., Kacprzyk J., Sheremetov L., Zadeh L. A., eds., *Perception-based Data Mining and Decision Making in Economics and Finance*, Springer-Verlag, Berlin and Heidelberg.

Bosc P., Lietard L., Pivet O. (1995) Quantified statements and database fuzzy queries, In: Bosc P., Kacprzyk J., eds., *Fuzziness in Database Management Systems*, Springer-Verlag, Berlin and Heidelberg.

Castillo-Ortega R., Marín N., Sánchez D. (2011) A fuzzy approach to the linguistic summarization of time series, *Multiple-Valued Logic and Soft Computing*, 17 (2-3), 157—182.

De S. K., Krishna P. R. (2004)Clustering web transactions using rough approximation. *Fuzzy Sets and Systems*, 148(1), 131—138.

Dertouzos M. (2001) *The Unfinished Revolution: Human-Centered Computers and What They Can Do For Us*, New York, Harper Collins.

Grabisch M. (1998) Fuzzy integral as a flexible and interpretable tool of aggregation, In: Bouchon-Meunier B., ed., *Aggregation and Fusion of Imperfect Information*, Heidelberg, New York: Physica—Verlag, 51–72.

Kacprzyk J., Zadrożny S. (2001) Computing with words in decision making through individual and collective linguistic choice rules, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(Supplement), 89—102.

Kacprzyk J., Wilbik A. (2010) Linguistic summaries of time series: On some additional data independent quality criteria, In: Bouchon-Meunier B., Magdalena L., Ojeda-Aciego M., Verdegay J.-L., Yager R. R., eds., *Foundations of Reasoning under Uncertainty*, Springer Berlin / Heidelberg, 143–166.

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

Kacprzyk J., Wilbik A., Zadrożny S. (2006) Linguistic summarization of trends: a fuzzy logic based approach, In: *Proceedings of the 11th International Conference Information Processing and Management of Uncertainty in Knowledge-based Systems*, Paris, France, July 2-7, 2166–2172.

Kacprzyk J., Wilbik A., Zadrożny S. (2007) Linguistic summaries of time series via an OWA operator based aggregation of partial trends, In: *Proceedings of the FUZZ-IEEE 2007 IEEE International Conference on Fuzzy Systems*. IEEE Press, 467–472.

Kacprzyk J., Yager R. R. (2001) Linguistic summaries of data using fuzzy logic, *International Journal of General Systems*, 30, 33–154.

Kacprzyk J., Yager R. R., Zadrożny S. (2000) A fuzzy logic based approach to linguistic summaries of databases, *International Journal of Applied Mathematics and Computer Science*, 10, 813—834.

Kacprzyk J., Zadrożny S. (2001) Fuzzy linguistic summaries via association rules, In: Kandel A., Last M., Bunke H., eds., *Data mining and computational intelligence*, Heidelberg, New York: Physica-Verlag, 115–139.

Keogh E., Chu S., Hart D., Pazzani M. (2001) An online algorithm for segmenting time series, In: *Proceedings of the 2001 IEEE International Conference on Data Mining*.

Pal S. K., Talwar V., Mitra P. (2002) Web mining in soft computing framework: Relevance, *IEEE Transactions on Neural Networks*, 13(5), 1163–1177.

Pedrycz W., Gomide F. (2007) *Fuzzy Systems Engineering: Toward Human-Centric Computing*, John Wiley & Sons.

Ros M., Pegalajar M., Delgado M., Vila A., Anderson D. T., Keller J. M., Popescu M. (2011) 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*, 2080–2087.

Shiu S. C. K., Wong C. K. P. (2005) Web access path prediction using fuzzy case based reasoning, In: Khosla R., Howlett R. J., Jain L. C., eds., *KES (3)*, ser. Lecture Notes in Computer Science, Springer, 3683, 135–140.

Sklansky J., Gonzalez V. (1980) Fast polygonal approximation of digitized curves, *Pattern Recognition*, 12(5), 327—331.

Wang X., Abraham A., Smith K. A. (2002) Soft computing paradigms for web access pattern analysis. In: Wang L., Halgamuge S. K., Yao X., eds., *FSKD*, 631–635.

Wang X., Abraham A., Smith K. A. (2005) Intelligent web traffic mining and analysis, *J. Netw. Comput. Appl.*, 28(2), 147–165.

Yager R. R. (1982) A new approach to the summarization of data, *Information Sciences*, vol. 28, 69—86.

Yager R. R. (1988) On ordered weighted averaging aggregation operators in multicriteria decision making, *IEEE Transactions on Systems, Man and Cybernetics*, SMC-18, 183—190.

Yager R. R., Kacprzyk J., Beliakov G., eds. (2011) *Recent Developments in the Ordered Weighted Averaging Operators: Theory and Practice*, ser. Studies in Fuzziness and Soft Computing., Springer, 265.

Yager R. R., Kacprzyk J., eds., (1997) *The Ordered Weighted Averaging Operators: Theory and Applications*, Kluwer, Boston.

Zadeh L. A. (1983) A computational approach to fuzzy quantifiers in natural languages, *Computers and Mathematics with Applications*, 9, 149—184.

Zadrożny S., Kacprzyk J. (1996) Quantifier guided aggregation using OWA operators, *International Journal of Intelligent Systems*, 11, 49–73.

Zadrożny S., Kacprzyk J. (2004) Segmenting time series: A survey and novel approach, In: Last M., Kandel A., Bunke H., eds., *Data Mining in Time Series Databases*, World Scientific Publishing,

Zadrożny S., Kacprzyk J. (2005) Linguistic database summaries and their protoforms: toward natural language based knowledge discovery tools, *Information Sciences*, 173, 281–304.

Zadrożny S., Kacprzyk J. (2007a) From a static to dynamic analysis of weblogs via linguistic summaries, In: *IFSA-2011*.

Zadrożny S., Kacprzyk J. (2007b) Summarizing the contents of web server logs: A fuzzy linguistic approach, In: *FUZZ-IEEE*, 1–6.

Zadrożny S., Kacprzyk J. (2007c) Towards perception based time series data mining, In: Nikravesh M., Kacprzyk J., Zadeh L. A., eds., *Forging New Frontiers. Fuzzy Pioneers I*, Springer-Verlag, Berlin and Heidelberg, 217–230.

Zadrożny S., Kacprzyk J. (2010a) An approach to the linguistic summarization of time series using a fuzzy quantifier driven aggregation, *Int. J. Intell. Syst.*, 25(5), 411–439.

Zadrożny S., Kacprzyk J. (2010b) Computing with words and systemic functional linguistics: Linguistic data summaries and natural language generation, In: *Integrated Uncertainty Management and Applications*, Van-Nam Huynh Y. N., Lawry J., eds., Springer-Verlag, Heidelberg and New York, 23—36.

Zadrożny S., Kacprzyk J. (2010c) Computing with words is an implementable paradigm: Fuzzy queries, linguistic data summaries, and natural-language generation, *Fuzzy Systems, IEEE Transactions on*, 18(3), 461—472.

Zadrożny S., Kacprzyk J. (2010d) Time series comparison using linguistic fuzzy techniques, In: Hüllermeier E., Kruse R., Hoffmann F., eds., *Computational Intelligence for Knowledge-Based Systems Design, Proceedings of 13th International Conference on Information Processing and Management of Uncertainty, IPMU 2010*, ser. Lecture Notes in Computer Science, Springer, 6178, 330–339.

Zadrożny S., Kacprzyk J. (2013) Derivation of linguistic summaries is inherently dificult: Can association rule mining help? In: Borgelt C., Ángeles Gil Alvarez M., Sousa J. M. D. C., Verleysen M., eds., *Towards Advanced Data Analysis by Combining Soft Computing and Statistics*, Heidelberg: Springer, 291–303.