

COMPUTING WITH WORDS, PROTOFORMS AND LINGUISTIC DATA SUMMARIES: TOWARDS A NOVEL NATURAL LANGUAGE BASED DATA MINING AND KNOWLEDGE DISCOVERY TOOLS

Submitted: 15th May 2014; accepted: 24th June 2014

Janusz Kacprzyk, Sławomir Zadrozny

DOI: 10.14313/JAMRIS_4-2013/27

Abstract:

We show how Zadeh's idea of computing with words and perceptions, based on his concept of a precisiated natural language (PNL), can lead to a new direction in the use of natural language in data mining, linguistic data(base) summaries. We emphasize the relevance of Zadeh's another idea, that of a protoform, and show that various types of Yager type linguistic data summaries may be viewed as items in a hierarchy of protoforms of summaries. We briefly present an implementation for a sales database of a computer retailer as a convincing example that these tools and techniques are implementable and functional. These summaries involve both data from an internal database of the company and data downloaded from external databases via the Internet.

Keywords: computing with words, linguistic summaries, protoform

1. Introduction

The purpose of this article is to shortly present our opinion on what might be considered to be the most influential and far reaching idea conceived by Zadeh, i.e. *computing with words* (CWW), and – on a more technical level – *protoforms*. We do not mention here his "grand inventions" like fuzzy sets and possibility theories or foundations of the state space approach in systems modeling, which has been probably more relevant in a general sense, for various fields of science. To follow the spirit of this volume, our exposition will be concise and comprehensible. This article is an extended version of a short reserach note by Kacprzyk and Zadrozny [34]

Computing with words (and perceptions), introduced by Zadeh in the mid-1990s, and best and most comprehensively presented in Zadeh and Kacprzyk's books [48], may be viewed to be a new "technology" in the representation, processing and solving of various real life problems when a human being is a crucial element, one that makes it possible to use natural language, with its inherent imprecision, an an effective and efficient way.

To formally represent elements and expressions of natural language, Zadeh proposed to use the so-called PNL (precisiated natural language) in which statements about values, relations, etc. between variables are represented by constraints. In PNL, statements -- written " x is R " -- may be different, and correspond to numeric values, intervals, possibility dis-

tributions, verity distributions, probability distributions, usuality qualified statements, rough sets representations, fuzzy relations, etc. For our purposes and in most our works, the usuality qualified representation has been be of special relevance. Basically, it says " x is usually R " that is meant as "in most cases, x is R ". PNL may play various roles among which crucial are: description of perceptions, definition of sophisticated concepts, a language for perception based reasoning, etc. Notice that the usuality is an example of modalities in natural language. Clearly, the above tools are meant for the representation and processing of perceptions.

Another concept that Zadeh has subsequently introduced is that of a *protoform*. In general, most perceptions are summaries, exemplified by "most Swedes are tall" which is clearly a summary of the Swedes with respect to height. It can be represented in Zadeh's notation as "most A s are B s". This can be employed for reasoning under various assumptions. One can go a step further, and define a protoform as an abstracted summary. In our case, this would be " Q A s are B s". Notice that we now have a more general, deinstantiated form of our point of departure (most Swedes are tall), and also of "most A s are B s". Needless to say that most human reasoning is protoform based, and the availability of such a more general representation is very valuable, and provides tools that can be used in many cases.

Basically, the essence of our work over the years boiled down to showing that the concept of a precisiated natural language, and in particular of a protoform, viewed from the perspective of CWW, can be of use in attempts at a more effective and efficient use of vast information resources, notably through linguistic data(base) summaries which are very characteristic for human needs and comprehension abilities.

We will briefly discuss an approach based on the concept of a *linguistic data(base) summary* that has been originally proposed by Yager [43,44] and further developed mainly by Kacprzyk and Yager [19], and Kacprzyk, Yager and Zadrozny [20]. The essence of such linguistic data summaries is that a set of data, e.g., concerning employees, with (numeric) data on their age, sex, salaries, seniority, etc., can be summarized linguistically with respect to a selected attribute or attributes, say age and salaries, by linguistically quantified propositions, e.g., "almost all employees are well qualified", "most young employees are well paid", etc. which are simple, extremely human consistent and intuitive, and do summarize in a concise yet very informative form what we may be interested in. This will be

done from the perspective of Zadeh's CWW paradigm (cf. Zadeh and Kacprzyk [48]), and we will in particular indicate the use of Zadeh's concept of a protoform of a fuzzy linguistic summary (cf. Zadeh [47], Kacprzyk and Zadrozny [23]) that can provide an easy generalization, portability and scalability.

We will mention both the classic static linguistic summaries, notably showing that a class of summaries of interest is mined via Kacprzyk and Zadrozny's [22, 25] FQUERY for Access, and that by relating various types of linguistic summaries to fuzzy queries, with various known and sought elements, we can arrive at a hierarchy of protoforms of linguistic data summaries. Moreover, we will also briefly mention new protoforms of linguistic summaries of time series as proposed by Kacprzyk, Wilbik and Zadrozny [17, 18].

2. Linguistic Data Summaries via Fuzzy Logic with Linguistic Quantifiers

The linguistic summary is meant as a sentence [in a (quasi)natural language] that subsumes the very essence (from a certain point of view) of a set of data. Here this set is assumed to be numeric, large and not comprehensible in its original form by the human being. In Yager's approach (cf. Yager [43], Kacprzyk and Yager [19], and Kacprzyk, Yager and Zadrozny [20]) we have:

- $Y = \{y_1, \dots, y_n\}$ is a set of objects (records) in a database, e.g., the set of workers;

- $A = \{A_1, \dots, A_m\}$ is a set of attributes characterizing objects from Y , e.g., salary, age, etc. in a database of workers, and $A_j(y_i)$ denotes a value of attribute A_j for object y_i .

A linguistic summary of data set D consists of:

- a summarizer S , i.e. an attribute together with a linguistic value (fuzzy predicate) defined on the domain of attribute A_j (e.g. "low salary" for attribute "salary");
- a quantity in agreement Q , i.e. a linguistic quantifier (e.g. most);
- truth (validity) T of the summary, i.e. a number from the interval $[0, 1]$ assessing the truth (validity) of the summary (e.g. 0.7); usually, only summaries with a high value of T are interesting;
- 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, the linguistic summary may be exemplified by

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

A richer form of the summary may include a qualifier as in, e.g.,

$$T(\text{most of young employees earn low salary}) = 0.7 \quad (2)$$

The core of a linguistic summary is a *linguistically quantified proposition* in the sense of Zadeh [46],

the one corresponding to (1) written as

$$Qy's \text{ are } S \quad (3)$$

and the one corresponding to (2) written as

$$QRy's \text{ are } S \quad (4)$$

The T , i.e., the truth value of (3) or (4), may be calculated by using either original Zadeh's calculus of linguistically quantified statements (cf. [46]), or other interpretations of linguistic quantifiers (cf. Liu and Kerre [38]), including Yager's OWA operators [45] and Dubois et al. OWmin operators [6], or via generalized quantifier, cf. Hájek and Holeňa [13] or Glöckner [12].

Recently, Zadeh [47] introduced a relevant concept of a *protoform* which is defined as a more or less abstract prototype (template) of a linguistically quantified proposition. The most abstract protoforms correspond to (3) and (4), while (1) and (2) are examples of fully instantiated protoforms. Thus, evidently, protoforms form a hierarchy, where higher/lower levels correspond to more/less abstract protoforms. Going down this hierarchy one has to instantiate particular components of (3) and (4), i.e., quantifier Q and fuzzy predicates S and R . The instantiation of the former one boils down to the selection of a quantifier. The instantiation of fuzzy predicates requires the choice of attributes together with linguistic values (atomic predicates) and a structure they form when combined using logical connectives. This leads to a theoretically infinite number of potential protoforms. However, for the purposes of mining of linguistic summaries, there are obviously some limits on a reasonable size of a set of summaries that should be taken into account. These results from a limited capability of the user in the interpretation of summaries as well as from the computational point of view.

The concept of a protoform may provide a guiding paradigm for the design of a user interface supporting the mining of linguistic summaries. It may be assumed that the user specifies a protoform of linguistic summaries sought. Basically, the more abstract protoform the less should be assumed about summaries sought, i.e., the wider range of summaries is expected by the user. There are two limit cases, where:

- a totally abstract protoform is specified, i.e., (4),
- all elements of a protoform are totally specified as given linguistic terms,

and in the former case the system has to construct all possible summaries (with all possible linguistic components and their combinations) for the context of a given database (table) and present to the user those verifying the validity to a degree higher than some threshold. In the second case, the whole summary is specified by the user and the system has only to verify its validity. Thus, the former case is usually more interesting from the point of view of the user but at the same time more complex from the computational point of view. There is a number of intermediate cases that may be more practical. In Table 1 basic types

of protoforms/linguistic summaries are shown, corresponding to protoforms of a more and more abstract form.

Basically, each of fuzzy predicates S and R may be defined by listing its atomic fuzzy predicates (i.e., pairs of "attribute/linguistic value") and structure, i.e., how these atomic predicates are combined. In Table 1 S (or R) corresponds to the full description of both the atomic fuzzy predicates (referred to as linguistic values, for short) as well as the structure. For example: " Q young employees earn a high salary" is a protoform of Type 2, while "Most employees earn a "?" salary" is a protoform of Type 3. In the first case the system has to select a linguistic quantifier (usually from a predefined dictionary) that when put in place of Q makes the resulting linguistically quantified proposition valid to the highest degree, and in the second case, the linguistic quantifier as well as the structure of summarizer S are given and the system has to choose a linguistic value to replace the question mark ("?") yielding a linguistically quantified proposition as valid as possible.

Thus, the use of protoforms makes it possible to devise a uniform procedure to handle a wide class of linguistic data summaries so that the system can be easily adaptable to a variety of situations, users' interests and preferences, scales of the project, etc.

Usually, most interesting are linguistic summaries required by a summary of Type 5. They may be interpreted as fuzzy IF-THEN rules, and many interpretations are proposed (cf., e.g., Dubois and Prade [8]) there are considered many possible interpretations for fuzzy rules), and some of them were directly discussed in the context of linguistic summaries later on.

There are many views on the idea of a linguistic summary, for instance a fuzzy functional dependency, a gradual rule, even a typical value. Though they do reflect the essence of a human perception of what a linguistic summary should be, they are beyond the scope of this paper which focuses on a different approach.

3. Mining of Linguistic Data Summaries

In the process of mining of linguistic summaries, at the one extreme, the system may be responsible for both the construction and verification of summaries (which corresponds to Type 5 protoforms/summaries given in Table 1). At the other extreme, the user proposes a summary and the system only verifies its validity (which corresponds to Type 0 protoforms/summaries in Table 1). The former approach seems to be more attractive and in the spirit of data mining meant as the discovery of interesting, unknown regularities in data. On the other hand, the latter approach, obviously secures a better interpretability of the results. Thus, we will discuss now the possibility to employ a flexible querying interface for the purposes of linguistic summarization of data, and indicate the implementability of a more automatic approach.

3.1. A fuzzy querying add-on for formulating linguistic summaries

In Kacprzyk and Zadrozny's [24, 29] approach, the interactivity, i.e. a user assistance, in the mining of linguistic summaries is a key point, and is in the definition of summarizers (indication of attributes and their combinations). This proceeds via a user interface of a fuzzy querying add-on. In Kacprzyk and Zadrozny [22, 25, 30], a conventional database management system is used with a fuzzy querying tool, FQUERY for Access. An important component of this tool is a dictionary of linguistic terms to be used in queries. They include fuzzy linguistic values and relations as well as fuzzy linguistic quantifiers. There is a set of built-in linguistic terms, but the user is free to add his or her own. Thus, such a dictionary evolves in a natural way over time as the user is interacting with the system. For example, an SQL query searching for *troublesome orders* may take the following WHERE clause:

```
WHERE Most of the conditions are met out of
PRICE*ORDERED-AMOUNT IS Low
DISCOUNT IS High
ORDERED-AMOUNT IS Much
Greater Than ON-STOCK
```

Obviously, the condition of such a fuzzy query directly correspond to summarizer S in a linguistic summary. Moreover, the elements of a dictionary are perfect building blocks of such a summary. Thus, the derivation of a linguistic summary of type (3) may proceed in an interactive (user-assisted) way as follows:

- the user formulates a set of linguistic summaries of interest (relevance) using the fuzzy querying add-on,
- the system retrieves records from the database and calculates the validity of each summary adopted, and
- a most appropriate linguistic summary is chosen.

Referring to Table 1, we can observe that Type 0 as well as Type 1 linguistic summaries may be easily produced by a simple extension of FQUERY for Access. Basically, the user has to construct a query, a candidate summary, and it is to be determined which fraction of rows matches that query (and which linguistic quantifier best denotes this fraction, in case of Type 1). For Type 3 summaries, a query/summarizer S consists of only one simple condition built of the attribute whose typical (exceptional) value is sought. For example, using: $Q = \text{"most"}$ and $S = \text{"age=?"}$ we look for a typical value of "age". From the computational point of view Type 5 summaries represent the most general form considered: fuzzy rules describing dependencies between specific values of particular attributes. The summaries of Type 1 and 3 have been implemented as an extension to Kacprzyk and Zadrozny's [26–28] FQUERY for Access.

The discovery of general, Type 5 rules is difficult, and some simplifications about the structure of fuzzy predicates and/or quantifier are needed, for instance to obtain association rules which have been initially

Tab. 1. Classification of protoforms/linguistic summaries

Type	Parameters		
	Protoform	Given	Sought
0	QRy 's are S	All	validity T
1	Qy 's are S	S	Q
2	QRy 's are S	S and R	Q
3	Qy 's are S	Q and structure of S	linguistic values in S
4	QRy 's are S	Q , R and structure of S	linguistic values in S
5	QRy 's are S	Nothing	S , R and Q

defined for binary valued attributes as (cf. Agraval and Srikant [1]):

$$A_1 \wedge A_2 \wedge \dots \wedge A_n \longrightarrow A_{n+1} \quad (5)$$

and note that much earlier origins of that concept are mentioned in the work by Hájek and Holeňa [13]).

The use of fuzzy association rules to mine linguistic summaries through a fuzzy querying interface was proposed by Kacprzyk and Zadrozny [26–28,31] advocated the use of fuzzy association rules for mining linguistic summaries in the framework of flexible querying interface.

In particular, *fuzzy association rules* may be considered:

$$A_1 \text{ IS } R_1 \wedge A_2 \text{ IS } R_2 \wedge \dots \wedge A_n \text{ IS } R_n \longrightarrow A_{n+1} \text{ IS } S \quad (6)$$

where R_i is a linguistic term defined in the domain of the attribute A_i , i.e. a qualifier fuzzy predicate in terms of linguistic summaries (cf. Section 2) and S is another linguistic term corresponding to the summarizer. The confidence of the rule may be interpreted in terms of linguistic quantifiers employed in the definition of a linguistic summary. Thus, a fuzzy association rule may be treated as a special case of a linguistic summary of type defined by (4). The structure of the fuzzy predicates R_i and S is to some extent fixed but due to that efficient algorithms for rule generation may be employed. These algorithms are easily adopted to fuzzy association rules. Usually, the first step is a preprocessing of original, crisp data. Values of all attributes considered are replaced with linguistic terms best matching them. Additionally, a degree of this matching may be optionally recorded and later taken into account. Then, each combination of attribute and linguistic term may be considered as a Boolean attribute and original algorithms, such as Apriori [1], may be applied. They, basically, boil down to an efficient counting of support for all conjunctions of Boolean attributes, i.e., so-called itemsets (in fact, the essence of these algorithms is to count support for as small a subset of itemsets as possible). In case of fuzzy association rules attributes may be treated strictly as Boolean attributes -- they may appear or not in particular tuples -- or interpreted in terms of fuzzy logic as in linguistic summaries. In the latter case they appear in a tuple to a degree and the support counting should take that into account. In our context we employ basically the approach by Lee and Lee-Kwang [37] and Au and Chan [2], Hu *et al.* [14] who

simplify the fuzzy association rules sought by assuming a single specific attribute (class) in the consequent. Kacprzyk, Yager and Zadrozny [20, 26–28, 31, 36] advocated the use of fuzzy association rules for mining linguistic summaries in the framework of flexible querying interface. Chen *et al.* [5] investigated the issue of generalized fuzzy rules where a fuzzy taxonomy of linguistic terms is taken into account. Kacprzyk and Zadrozny [32] proposed to use more flexible aggregation operators instead of conjunction, but still in context of fuzzy association rules. More information on fuzzy association rules, from various perspectives, may be found later in this volume.

As to some other approaches to the derivation of fuzzy linguistic summaries, we can mention the following ones. George and Srikant [10], [11] use a genetic algorithm to mine linguistic summaries in which the summarizer is a conjunction of atomic fuzzy predicates. Then, they search for two linguistic summaries: the most specific generalization and the most general specification, assuming a dictionary of linguistic quantifiers and linguistic values over domains of all attributes. Kacprzyk and Strykowski [15, 16] have also implemented the mining of linguistic summaries using genetic algorithms. In their approach, the fitting function is a combination of a wide array of indices: a degree of imprecision (fuzziness), a degree of covering, a degree of appropriateness, a length of a summary, etc. (cf. also Kacprzyk and Yager [19]). Rasmussen and Yager [41, 42] propose an extension, SummarySQL, to SQL to cover linguistic summaries. Actually, they do not address the mining linguistic summaries but merely their verification. The SummarySQL may also be used to verify a kind of fuzzy gradual rules (cf. Dubois and Prade [7]) and fuzzy functional dependencies. Raschia and Mouaddib [40] deal with the mining of hierarchies of summaries, and their understanding of summaries is slightly different than here because they consider them as a conjunction of atomic fuzzy predicates (each referring to just one attribute). However, these predicates are not defined by just one linguistic value but possibly by fuzzy sets of linguistic values (i.e., fuzzy sets of higher levels are considered). The mining of summaries (a whole hierarchy of summaries) is based on a concept formation (conceptual clustering) process.

An interesting extension of the concept of a linguistic summary to the linguistic summarization of time series data was shown in a series of works by

Kacprzyk, Wilbik and Zadrozny [17, 18]. In this case the array of possible protoforms is much larger as it reflects various perspectives, intentiones, etc. of the user. Just to give an examples, the protoforms used in those works may be exemplified by: "Among all y 's, Q are P ", exemplified by "among all segments (of the time series) most are slowly increasing", and "Among all R segments, Q are P ", exemplified by "among all short segments almost all are quickly decreasing", as well as more sophisticated protoforms, for instance temporal ones like: " E_T among all y 's Q are P ", exemplified by "Recently, among all segments, most are slowly increasing", and " E_T among all Ry 's Q are P ", exemplified by "Initially, among all short segments, most are quickly decreasing"; they both go beyond the classic Zadeh's protoforms.

It is easy to notice that the mining of linguistic summaries may be viewed to be closely related to natural language generation (NLG) and this path was suggested in Kacprzyk and Zadrozny [33]. This may be a promising direction as NLG is a well developed area and software is available.

A very relevant issue of comprehensiveness of linguistic data summaries, in Michalski's sense, that is related to how well they can be understandable to an average user is considered in a recent paper by Kacprzyk and Zadrozny [35].

4. Concluding Remarks

We have shown how Zadeh's idea of computing with words, often called computing with words and perceptions, based on his concepts of a precisiated natural language (PNL) and linguistically quantified propositions can lead to a new direction in the use of natural language in data mining and knowledge discovery, namely a linguistic data(base) summary. We have in particular focused our attention on the relevance of Zadeh's another idea, that of a protoform, and show that various types of linguistic data summaries may be viewed as items in a hierarchy of protoforms of linguistic data summaries. We have briefly presented an implementation of linguistic data summaries for a sales database of a computer retailer as a convincing example that these tools and techniques are implementable and practically functional. These summaries can involve both data from a company database and data downloaded from external databases via the Internet.

ACKNOWLEDGEMENTS

This research has been partially supported by the National Centre of Science under Grant No. UMO-2012/05/B/ST6/03068.

AUTHORS

Janusz Kacprzyk* – Systems Research Institute Polish Academy of Sciences, ul. Newelska 6, 01-447 Warszawa, Poland, e-mail: kacprzyk@ibspan.waw.pl.

Sławomir Zadrozny – Systems Research Institute Polish Academy of Sciences, ul. Newelska 6, 01-447

Warszawa, Poland, e-mail: zadrozny@ibspan.waw.pl.

*Corresponding author

REFERENCES

- [1] Agrawal R., Srikant R., "Fast algorithms for mining association rules". In: *Proceedings of the 20th International Conference on Very Large Databases*, Santiago de Chile, 1994.
- [2] Au W.-H., Chan K.C.C., "FARM: A data mining system for discovering fuzzy association rules". In: *Proceedings of the 8th IEEE International Conference on Fuzzy Systems*, Seoul, Korea, 1999, 1217–1222.
- [3] Berzal F., Cubero J.C., Marín N., Vila M.A., Kacprzyk J. and Zadrozny S., "A General framework for computing with words in object-oriented programming", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 15, 2007, 111–131. DOI: <http://dx.doi.org/10.1142/S0218488507004480>.
- [4] Bosc P., Dubois D., Pivert O., Prade H., de Calmes M., "Fuzzy summarization of data using fuzzy cardinalities". In: *Proceedings of IPMU 2002*, Anancy, France, 2002, 1553–1559.
- [5] Chen G., Wei Q., Kerre E., "Fuzzy data mining: discovery of fuzzy generalized association rules". In: G. Bordogna and G. Pasi (Eds.): *Recent Issues on Fuzzy Databases*. Springer-Verlag, Heidelberg and New York, 2000, 45–66. DOI: <http://dx.doi.org/10.1007/978-3-7908-1845-1>.
- [6] Dubois D., Fargier H., Prade H., "Beyond min aggregation in multicriteria decision: (ordered) weighted min, discri-min, leximin". In: R.R. Yager and J. Kacprzyk (Eds.): *The Ordered Weighted Averaging Operators. Theory and Applications*, Kluwer, Boston, 1997, 181–192.
- [7] Dubois D., Prade H., "Gradual rules in approximate reasoning", *Information Sciences*, vol. 61, 1992, 103–122.
- [8] Dubois D., Prade H., "Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions", *Fuzzy Sets and Systems*, vol. 40, 1991, 143–202. DOI: [http://dx.doi.org/10.1016/0165-0114\(91\)90050-Z](http://dx.doi.org/10.1016/0165-0114(91)90050-Z).
- [9] Bosc P., Dubois D., Prade H., "Fuzzy functional dependencies – an overview and a critical discussion". In: *Proceedings of 3rd IEEE International Conference on Fuzzy Systems*, Orlando, USA, 1994, 325–330. DOI: <http://dx.doi.org/10.1109/FUZZY.1994.343753>.
- [10] George R., Srikanth R., "Data summarization using genetic algorithms and fuzzy logic". In: F. Herrera, J.L. Verdegay (Eds.): *Genetic Algorithms and Soft Computing*, Springer-Verlag, Heidelberg, 1996, 599–611.
- [11] George R., Srikanth R., "A soft computing approach to intensional answering in databases",

- Information Sciences*, vol. 92, no. 1–4, 1996, 313–328. DOI: [http://dx.doi.org/10.1016/0020-0255\(96\)00049-7](http://dx.doi.org/10.1016/0020-0255(96)00049-7).
- [12] Glöckner I., "Fuzzy quantifiers, multiple variable binding, and branching quantification". In: T. Bilgic *et al.*: *IFSA 2003. LNAI 2715*, Springer-Verlag, Berlin and Heidelberg, 2003, 135–142.
- [13] Hájek P., Holeňa M., "Formal logics of discovery and hypothesis formation by machine", *Theoretical Computer Science*, vol. 292, 2003, 345–357. DOI: http://dx.doi.org/10.1007/3-540-49292-5_26.
- [14] Hu Y.-Ch., Chen R.-Sh., Tzeng G.-H., "Mining fuzzy association rules for classification problems", *Computers and Industrial Engineering*, vol. 43, no. 4, 2002, 735–750. DOI: [http://dx.doi.org/10.1016/S0360-8352\(02\)00136-5](http://dx.doi.org/10.1016/S0360-8352(02)00136-5).
- [15] Kacprzyk J., Strykowski P., "Linguistic data summaries for intelligent decision support". In: R. Felix (Ed.): *Fuzzy Decision Analysis and Recognition Technology for Management, Planning and Optimization - Proceedings of EFDAN'99*, Dortmund, Germany, 1999, 3–12.
- [16] Kacprzyk J., Strykowski P., "Linguistic summaries of sales data at a computer retailer: a case study". In: *Proceedings of IFSA'99*, vol. 1, 1999, Taipei, Taiwan R.O.C, 29–33.
- [17] Kacprzyk J., Wilbik A., Zadrożny S., "Linguistic summarization of time series using a fuzzy quantifier driven aggregation", *Fuzzy Sets and Systems*, vol. 159, no. 12, 2008, 1485–1499. DOI: <http://dx.doi.org/10.1016/j.fss.2008.01.025>.
- [18] Kacprzyk J., Wilbik A., Zadrożny S., "An approach to the linguistic summarization of time series using a fuzzy quantifier driven aggregation", *International Journal of Intelligent Systems*, vol. 25, no. 5, 2010, 411–439. DOI: <http://dx.doi.org/10.1002/int.20405>.
- [19] Kacprzyk J., Yager R.R., "Linguistic summaries of data using fuzzy logic", *International Journal of General Systems*, vol. 30, no. 2, 2001, 33–154. DOI: <http://dx.doi.org/10.1080/03081070108960702>.
- [20] Kacprzyk J., Yager R.R., Zadrożny S., "A fuzzy logic based approach to linguistic summaries of databases", *International Journal of Applied Mathematics and Computer Science*, 10, 2000, 813–834.
- [21] Kacprzyk J., Yager R.R., Zadrożny S., "Fuzzy linguistic summaries of databases for an efficient business data analysis and decision support. In W. Abramowicz and J. Zurada (Eds.): *Knowledge Discovery for Business Information Systems*, pp. 129–152, Kluwer, Boston, 2001.
- [22] Kacprzyk J., Yager R.R., Zadrożny S., "FQUERY for Access: fuzzy querying for a Windows-based DBMS". In: P. Bosc and J. Kacprzyk (Eds.): *Fuzziness in Database Management Systems*, Springer-Verlag, Heidelberg, 1995, 415–433. DOI: http://dx.doi.org/10.1007/978-3-7908-1897-0_18.
- [23] Kacprzyk J., Zadrożny S., "Protoforms of linguistic data summaries: towards more general natural-language-based data mining tools". In: A. Abraham, J. Ruiz-del-Solar, M. Koeppen (Eds.): *Soft Computing Systems*, pp. 417 - 425, IOS Press, Amsterdam, 2002.
- [24] Kacprzyk J., Zadrożny S., "Data Mining via Linguistic Summaries of Data: An Interactive Approach". In: T. Yamakawa and G. Matsumoto (Eds.): *Methodologies for the Conception, Design and Application of Soft Computing. Proc. of IIZUKA'98*, Iizuka, Japan, 1998, 667–668.
- [25] Kacprzyk J., Zadrożny S., "The paradigm of computing with words in intelligent database querying". In: L.A. Zadeh and J. Kacprzyk (Eds.): *Computing with Words in Information/Intelligent Systems. Part 2. Foundations*, Springer-Verlag, Heidelberg and New York, 1999, 382–398. DOI: <http://dx.doi.org/10.1007/978-3-7908-1872-7>.
- [26] Kacprzyk J., Zadrożny S., "Computing with words: towards a new generation of linguistic querying and summarization of databases". In: P. Sinčák and J. Vaščák (Eds.): *Quo Vadis Computational Intelligence?*, Springer-Verlag, Heidelberg and New York, 2000, 144–175, .
- [27] Kacprzyk J., Zadrożny S., "On a fuzzy querying and data mining interface", *Kybernetika*, vol. 36, 2000, 657–670.
- [28] Kacprzyk J., Zadrożny S., "On combining intelligent querying and data mining using fuzzy logic concepts". In: G. Bordogna and G. Pasi (Eds.): *Recent Research Issues on the Management of Fuzziness in Databases*, Springer-Verlag, Heidelberg and New York, 2000, 67–81.
- [29] Kacprzyk J., Zadrożny S., "Data mining via linguistic summaries of databases: an interactive approach". In: L. Ding (Ed.): *A New Paradigm of Knowledge Engineering by Soft Computing*, World Scientific, Singapore, 2001, 325–345. DOI: <http://dx.doi.org/10.1142/4606>.
- [30] Kacprzyk J., Zadrożny S., "Computing with words in intelligent database querying: standalone and Internet-based applications", *Information Sciences*, vol. 134, no. 1–4, 2001, 71–109. DOI: [http://dx.doi.org/10.1016/S0020-0255\(01\)00093-7](http://dx.doi.org/10.1016/S0020-0255(01)00093-7).
- [31] Kacprzyk J., Zadrożny S., "On linguistic approaches in flexible querying and mining of association rules". In: H.L. Larsen, J. Kacprzyk, S. Zadrożny, T. Andreassen and H. Christiansen (Eds.): *Flexible Query Answering Systems. Recent Advances*, Springer-Verlag, Heidelberg and New York, 2001, 475–484.
- [32] Kacprzyk J., Zadrożny S., "Linguistic summarization of data sets using association rules". In: *Pro-*

- ceedings of The IEEE International Conference on Fuzzy Systems*, St. Louis, USA, 2003, 702–707.
- [33] Kacprzyk J., Zadrożny S., "Computing with words is an implementable paradigm: fuzzy queries, linguistic data summaries, and natural language generation", *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 3, 2010, 461–472. DOI: <http://dx.doi.org/10.1109/TFUZZ.2010.2040480>.
- [34] Kacprzyk J., Zadrożny S., "Computing with words and protoforms: powerful and far reaching ideas". In: Rudolf Seising, Enric Trillas, Claudio Moraga, and Settimo Termini (Eds.): *On Fuzziness*. Springer-Verlag, Berlin Heidelberg 2013, 255–270. DOI: <http://dx.doi.org/10.1007/978-3-642-35641-4>.
- [35] Kacprzyk J., Zadrożny S., "Comprehensiveness of Linguistic Data Summaries: A Crucial Role of Protoforms". In: Ch. Moewes and A. Nürnberger (Eds.): *Computational Intelligence in Intelligent Data Analysis*. Springer-Verlag, Berlin Heidelberg 2013, 207–221. DOI: <http://dx.doi.org/10.1007/978-3-642-32378-2>.
- [36] Kacprzyk J., Zadrożny S., "Derivation of Linguistic Summaries is Inherently Difficult: Can Association Rule Mining Help?" In: Borgelt Ch., Gil M. A., Sousa J. M. C., Verleysen M. (Eds.): *Towards Advanced Data Analysis by Combining Soft Computing and Statistics*, Springer-Verlag, 2013, 291–303. DOI: <http://dx.doi.org/10.1007/978-3-642-30278-7>.
- [37] Lee J.-H., Lee-Kwang H., "An extension of association rules using fuzzy sets". In: *Proceedings of the Seventh IFSA World Congress*, Prague, Czech Republic, 1997, 399–402.
- [38] Liu Y., Kerre E.E., "An overview of fuzzy quantifiers. (I). Interpretations", *Fuzzy Sets and Systems*, vol. 95, 1998, 1–21.
- [39] Mannila H., Toivonen H., Verkamo A.I., "Efficient algorithms for discovering association rules". In: U.M. Fayyad and R. Uthurusamy (Eds.): *Proceedings of the AAAI Workshop on Knowledge Discovery in Databases*, Seattle, USA, 1994, 181–192.
- [40] Raschia G., Mouaddib N., "SAINTETIQ: a fuzzy set-based approach to database summarization", *Fuzzy Sets and Systems*, vol. 129, no. 2, 2002, pp. 137–162. DOI: [http://dx.doi.org/10.1016/S0165-0114\(01\)00197-X](http://dx.doi.org/10.1016/S0165-0114(01)00197-X).
- [41] Rasmussen D., Yager R.R., "Fuzzy query language for hypothesis evaluation". In: Andreasen T, H. Christiansen and H. L. Larsen (Eds.): *Flexible Query Answering Systems*, pp. 23 - 43, Kluwer, Boston, 1997. DOI: http://dx.doi.org/10.1007/978-1-4615-6075-3_2.
- [42] Rasmussen D., Yager R.R., "Finding fuzzy and gradual functional dependencies with SummarySQL", *Fuzzy Sets and Systems*, vol. 106, no. 2, 1999, 131–142. DOI: [http://dx.doi.org/10.1016/S0165-0114\(97\)00268-6](http://dx.doi.org/10.1016/S0165-0114(97)00268-6).
- [43] Yager R.R., "A new approach to the summarization of data", *Information Sciences*, vol. 28, no. 1, 1982, 69–86. DOI: [http://dx.doi.org/10.1016/0020-0255\(82\)90033-0](http://dx.doi.org/10.1016/0020-0255(82)90033-0).
- [44] Yager R.R., "On linguistic summaries of data". In: W. Frawley and G. Piatetsky-Shapiro (Eds.): *Knowledge Discovery in Databases*. AAAI/MIT Press, 1991, 347–363.
- [45] Yager R.R., Kacprzyk J. (Eds.), *The Ordered Weighted Averaging Operators: Theory and Applications*, Kluwer, Boston, 1997.
- [46] Zadeh L.A., "A computational approach to fuzzy quantifiers in natural languages", *Computers and Mathematics with Applications*, vol. 9, no. 1, 1983, 149–184. DOI: [http://dx.doi.org/10.1016/0898-1221\(83\)90013-5](http://dx.doi.org/10.1016/0898-1221(83)90013-5).
- [47] Zadeh L.A., *A prototype-centered approach to adding deduction capabilities to search engines -- the concept of a protoform*, BISC Seminar, 2002, University of California, Berkeley, 2002. DOI: <http://dx.doi.org/10.1109/IS.2002.1044219>.
- [48] Zadeh L.A., Kacprzyk J. (Eds.), *Computing with Words in Information/Intelligent Systems. Part 1. Foundations. Part 2. Applications*, Springer - Verlag, Heidelberg and New York, 1999. DOI: <http://dx.doi.org/10.1007/978-3-7908-1873-4> (part 1) and DOI: <http://dx.doi.org/10.1007/978-3-7908-1872-7>.
- [49] Zadrożny S., Kacprzyk J., "Computing with words for text processing: an approach to the text categorization", *Information Sciences*, vol. 176, no. 4, 006, 415–437. DOI: <http://dx.doi.org/10.1016/j.ins.2005.07.017>.
- [50] Zadrożny S., Kacprzyk J., "Issues in the practical use of the OWA operators in fuzzy querying", *Journal of Intelligent Information Systems*, vol. 33, no. 3, 2009, 307–325. DOI: <http://dx.doi.org/10.1007/s10844-008-0068-1>.