Control and Cybernetics

vol. **35** (2006) No. 2

Interval-valued linguistic summaries of databases

by

A. Niewiadomski[1]**, J. Ochelska**[1] **and P. S. Szczepaniak**[1,2]

[1] Institute of Computer Science
Technical University of Łódź, Poland
[2] Systems Research Institute, Polish Academy of Sciences
Warsaw, Poland

Abstract: The so-called *linguistic summaries of databases* are the semi-natural language sentences that enable distilling the most relevant information from large numbers of tuples, and present it in the human consistent forms. Recently, the methods of constructing and evaluating linguistic summaries have been based on Zadeh's fuzzy sets, which represent uncertain data. The main aim of the paper is to enhance and generalize the Yager's approach to linguistic summarization of data. This enhancement is based on interval-valued fuzzy sets. The newly presented methods enable handling fuzzy concepts, whose membership degrees are not given by real values explicitly, but are approximated by intervals in $[0, 1]$. From now on, the Yager's approach can be viewed as a special case of the method presented in this paper. Finally, illustrative examples are presented.

Keywords: interval-valued fuzzy set, interval-valued linguistic variable, interval-valued linguistic quantifier, linguistic summary of database, interval-valued linguistic summary of database.

1. Introduction

1.1. Motivation

The amount of data, stored and processed electronically, is growing exponentially. People's natural capabilities to grasp all information which is necessary to manage and control various processes (business, scientific, medical, etc.) are naturally limited, therefore the need for computational support is well visible. In particular, tools which enable extracting information and knowledge from large number of figures, as well as of presenting the extracted data in natural languages can be very helpful. In this study, we intend to focus on *linguistic summarization of databases* according to the Yager's approach (Yager, 1982; Yager, Ford, and Canas, 1990, 1991) in which knowledge obtained from a database is presented in the sample form of MANY YOUNG *girls are* VERY TALL, where MANY, YOUNG, and VERY TALL are linguistic expressions handled by fuzzy sets (Zadeh, 1965, 1975, 1983).

The main motivation for extending this approach by the use of interval-valued fuzzy sets is that memberships of properties/phenomena/facts may not be expressible in terms of real values, as in Zadeh's membership functions. Fuzzy sets appear insufficient when determining the terms objective meanings in summaries is the necessary condition to provide a relevant linguistic description of numeric data. Membership functions for these terms should be constructed on the basis of at least a few experts' knowledge. Even if done so, final membership levels used in summaries are usually the average (arithmetic or weighted) or the median of levels given by experts. Naturally, this causes the loss of differences which appear when experts propose memberships according to their knowledge. Therefore, the use of interval-valued fuzzy sets and of their interval-valued membership functions is expected to provide better and more natural handling of experts' propositions for membership levels. Thus, the article presents the methods for obtaining *Interval-Valued Linguistic Summaries of Databases* which are an extension of the Yager's linguistic summaries, and include them as a special case. Some basic ideas for interval-valued linguistic summaries of databases have already been given in Niewiadomski (2005a, d).

1.2. Linguistic summarization of data

Summarizing data is the process which allows to grasp and shortly describe global tendencies appearing in a set of stored data without doing manual "record-by-record" analysis. Summarization is defined by Mani and Maybury (1999) as

> Summarization is the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks)

Linguistic information and knowledge can be obtained from databases via many different algorithms, computational methods, and under many assumptions. This is the subject of wider considerations about fuzziness and its connections with human perceptions and natural language. The various points of view are presented by Bosc and Pivert (1992), Bosc et al. (2002), Raschia and Mouaddib (2002), Rasmussen and Yager (1997, 1999), and by Srikanth and Agrawal (1996). However, they are all based on the assumptions and – in general – on philosophy which differs from the basic ideas given by Yager (1982).

It can be easily noticed that automated generation of sentences containing crisp qualifications does not seem to be a problem for statistical methods. The summarization via statistical tools may apply means, medians, standard deviations, and other well known indices. However, such a manner of interpreting data is understandable and practicable for rather small and specialized group of people, such as analysts, managers, etc. According to Yager, Ford and Canas (1990):

> ... summarization would be especially practicable if it could provide us with summaries that are not as terse as the mean, as well as treating the summarization of non-numeric data

Naturally, people express information in natural language. Hence, the main assumption for user friendly summarization is to give results formulated linguistically, not numerically. Therefore, the effects in the form of *Many people have bought cheap cars in last years* rather than *Between 1998 and 2004 3.8% of customers have bought a car of the average price of 10,376.99 Euro* are expected, not only by an average user, but also by qualified personnel frequently needing approximated and compact information, instead of detailed figures and thousands of raw tuples.

Thus, fuzzy sets, which provide computational support and semantics for linguistic summaries of databases, are employed in modeling the linguistic terms describing features of objects, e.g. VERY TALL, LOW PRICE, or HIGH SPEED, and amounts of records/objects/tuples satisfying given properties, e.g. MUCH LESS THAN 1000, ABOUT 10, ABOUT HALF, or ALMOST ALL. These elements of summaries are called *summarizers* and *linguistic quantifiers*, respectively, and the main idea presented here is to apply interval-valued fuzzy sets to model them, in order to obtain more universal and satisfactory linguistic summaries of databases.

1.3. Basic definitions

1.3.1. Fuzzy sets

A *fuzzy set* A in a non-empty universe of discourse $\mathcal{X}$ is the classical set of ordered pairs

$$A =_{df} \{< x, \mu_A(x) >: x \in \mathcal{X}\} \tag{1}$$

where $\mu_A \colon \mathcal{X} \to [0,1]$ is the *membership function of* A, whose values express the membership level of x in A (Zadeh, 1965). Each function μ_A may be seen as a generalization of the characteristic function of the crisp set A – in this sense, crisp sets are special cases of fuzzy sets.

A fuzzy set A in $\mathcal{X}$ is *normal* if and only if

$$\sup_{x \in \mathcal{X}} \mu_A(x) = 1. \tag{2}$$

An α-*cut* of A, $\alpha \in [0,1]$, is the crisp set $A_\alpha \subseteq \mathcal{X}$ having the characteristic function

$$\chi_{A_\alpha}(x) = \begin{cases} 1, & \text{if } \mu_A(x) \geq \alpha \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

A is *convex* if and only if $\forall \alpha \in [0,1] \;\; A_\alpha$ is convex in the classical sense.

Cardinality of a fuzzy set A in a finite $\mathcal{X}$ can be represented as

$$card(A) = \sum_{x \in \mathcal{X}} \mu_A(x). \tag{4}$$

The method is called σ-*count*, and, in contrary to the *fuzzy cardinalities* of fuzzy sets, e.g. FG-count, see, e.g., Zadeh (1983), $card(A)$ is a real number.

Fuzzy sets are mostly applied to formalize linguistic and imprecise but understandable statements which express both properties of objects, e.g. FAST CAR, BIG HOUSE, and amounts, e.g. VERY FEW, ABOUT 3/4, MUCH MORE THAN 20,000, etc. In particular, the idea of *linguistic variable*, based on fuzzy sets, is applied to the former, and the so-called *fuzzy quantification*, as the model of *linguistic quantification* – to the latter.

1.3.2. Linguistic variables

A linguistic variable (Zadeh, 1975) is an ordered quintuple $< L, H, \mathcal{X}, G, M >$, where:
L is the name of the variable,
H or $H(L)$ is the term-set (linguistic values of L)
$\mathcal{X}$ is the universe of discourse,
G is a syntactic rule which generates values (labels) of L,
M is a semantic rule which associates a term from L with a fuzzy set in $\mathcal{X}$.
A linguistic variable is exemplified by: L="temperature", $H(L) = \{$low, medium, acceptable, high, very high$\}$, $\mathcal{X} = [-50°C, +50°C]$, in which M associates to e.g. "very high" a non-decreasing monotonic and continuous membership function in $\mathcal{X}$, etc. Values of the membership functions of fuzzy sets in $\mathcal{X}$ are interpreted as *compatibility levels*, i.e. the degrees in which labels given are relevant to x's, e.g. the compatibility level of $39°C$ with "low temperature" is 0 and with "very high" – 0.9.

1.3.3. Linguistic quantifiers

The predicate calculus in the two-valued logic is extended by the use of *existential*, $\exists$, and *general*, $\forall$, quantifiers. Similarly, linguistic predicates can be quantified by the linguistic quantifiers which are natural language statements expressing amounts or numbers of objects, e.g. LESS THAN HALF. There are two (canonical) forms of linguistically quantified propositions

$$Q \text{ objects are } S_1 \tag{5}$$

denoted also as Q^I, and

$$Q \text{ objects being } S_2 \text{ are } S_1 \tag{6}$$

or Q^{II} (Zadeh, 1983; Liu, Kerre, 1998). In terms of fuzzy logic, S_1 and S_2 are the labels associated with fuzzy sets, and Q is a linguistic pronouncement of quantity represented by a normal and convex fuzzy set in a non-negative universe of discourse $\mathcal{Y} \subseteq \mathbb{R}^+ \cup \{0\}$. In particular, two types of the Zadeh fuzzy quantifiers can be distinguished: *absolute*, e.g. ABOUT 1000, BETWEEN 3 AND 6, which are fuzzy sets in $\mathbb{R}^+ \cup \{0\}$, and *relative*, e.g. ABOUT HALF, VERY FEW OF, which are fuzzy sets in $[0, 1]$.

The degree of truth of proposition (5), T, is computed via the membership function of a fuzzy quantifier Q and via cardinalities of fuzzy sets in $\mathcal{X}$ associated to S_1, S_2.

$$T\,(\ Q \text{ objects are } S_1) = \mu_Q\left(\frac{card(S_1)}{M}\right) \tag{7}$$

where $M = card(\mathcal{X})$ if Q is relative, and $M = 1$ if Q is absolute. And for (6):

$$T\,(\ Q \text{ objects being } S_2 \text{ are } S_1) = \mu_Q\left(\frac{card(S_1 \cap S_2)}{card(S_2)}\right) \tag{8}$$

Moreover, in (6), S_2 can be interpreted as *importance*.

2. Classical linguistic summaries

This section is intended to present the most fundamental information on linguistic summarizing of databases by Yager and on its selected improvements. Crucial terms and definitions are introduced, on the basis of which the interval-valued-fuzzy-set-based approach is then explained in detail in Section 4.

2.1. The Yager approach: the point of departure

”To summarize a database linguistically” means – according to Yager – to build a natural language sentence which describes amounts of elements that have the chosen properties (Yager, 1982). In general, a linguistic summary of a database by Yager is in the form of

$$Q\ P \text{ are/have } S\ [T] \tag{9}$$

where the symbols are interpreted: Q is a determination of amount (a quantity in agreement), or a *linguistic quantifier*, e.g. ABOUT HALF, FEW, MORE THAN 150. P is a subject of summary; it is determined as a set of objects from the summarized database. These objects manifest the attributes with values written in the fields of records. S is a feature of interest, the so-called *summarizer*, e.g. LOW TEMPERATURE, HIGH SALARY. T is a quality measure for the summary, *a degree of truth* or *a truth of a summary* which describes the reliability of the quantity pronouncement Q for a given feature S. T is a real number from the interval $[0, 1]$, and it is interpreted as *the level of confidence* for a given summary.

If y_1, y_2,..., y_m are the objects which manifest an attribute V, and the value of V for y_i is denoted as $V(y_i)$, then T is computed as a value of the membership function of a quantifier Q:

$$T = \mu_Q\left(\frac{r}{m}\right) \tag{10}$$

if Q is relative or

$$T = \mu_Q(r) \tag{11}$$

if Q is absolute, where

$$r = \sum_{i=1}^{m} \mu_S\big(V(y_i)\big). \tag{12}$$

A sample summary, constructed in this way, is:

ABOUT HALF *of my friends have* BIG HOUSES $[0.65]$

where ABOUT HALF and BIG HOUSE are the linguistic quantifier and the summarizer, respectively, both handled by fuzzy sets.

2.2. Extensions of linguistic summaries

Yager's idea of linguistic summarization was extended by George and Srikanth (1996). Apart from original application of genetic algorithms, they formulated the linguistic summary which concerns more than one attribute, and these attributes are joined by the 'AND' connective, e.g. TALL AND VERY YOUNG.

Let us define a set of objects $\mathcal{Y} = \{y_1, y_2, ..., y_m\}$, a set of attributes $V = \{V_1, V_2, ..., V_n\}$. Let $\mathcal{X}_1$, $\mathcal{X}_2$,..., $\mathcal{X}_n$ be the domains of V_1, V_2,..., V_n, respectively. The attributes from V describe objects from $\mathcal{Y}$; this is denoted as $V_j(y_i)$ – a value of the attribute V_j for the object y_i, i=1, 2,..., m, $j = 1, 2, ...n$. Hence, the database $\mathcal{D}$, which collects information about elements from $\mathcal{Y}$, is in the form of

$$\begin{aligned} \mathcal{D} &= \{< V_1(y_1), V_2(y_1), ..., V_n(y_1) >, < V_1(y_2), V_2(y_2), ..., V_n(y_2) >, \\ &\quad ...., < V_1(y_m), V_2(y_m), ..., V_n(y_m) >\} = \\ &= \{d_1, d_2, ..., d_m\} \end{aligned} \tag{13}$$

where $d_1, d_2, ..., d_m$ are the records describing objects y_1, y_2,..., y_m, respectively, such that $d_i \in \mathcal{X}_1 \times \mathcal{X}_2 \times ... \times \mathcal{X}_n$. Let S_1, S_2,..., S_n be the labels associated to fuzzy sets in $\mathcal{X}_1$, $\mathcal{X}_2$,..., $\mathcal{X}_n$, respectively. Let Q be a linguistic quantifier. The expected summary is

$$Q \text{ objects from } \mathcal{Y} \text{ are/have } S_1 \text{ AND } S_2 \text{ AND ... } S_n \ [T] \tag{14}$$

where the summarizer S is expressed as the family of fuzzy sets $\{S_1, S_2, ..., S_n\}$. μ_S is the membership function determined as

$$\mu_S(d_i) = \min_{j=1,2,\ldots,n} \left\{ \mu_{S_j}\left(V_j(y_i)\right) \right\}, \ i = 1, 2, ..., m \tag{15}$$

where $\min$ is the t-norm. This or another t-norm is a model of the operator AND which connects linguistic descriptions of features of d_i's. Computation of T has not been changed since the Yager's approach; it is still a real number from $[0, 1]$, and it is interpreted as the level of confidence for a given summary.

When the number of records is relatively large and each of them is described by several attributes, computation of T may be costly and/or may take much time. For instance, when a database contains $m = 6000$ records, described by $n = 12$ attributes each, it is necessary to compute $m \times n = 72000$ membership degrees. Experience shows that usually most of these degrees equal 0, hence the computation of them may seem to be pointless. Therefore, the limitations which help to decrease the computational cost should be determined. One of propositional modifications is presented by Kacprzyk and Yager (2001) and Kacprzyk, Yager, and Zadrożny (2000, 2001) and is based on limiting classes of admissible summary by defining qualifications of a summarizer[1]. While in Yager's and George and Srikanth's summaries, the first canonical form of a quantification (5) is used, this version is based on the second canonical form (6) – summaries are constructed only for these objects which manifest a preselected property w_g, named "a query", at a non-zero level. As the result, the search process is significantly shorter and quicker, since computation of minima (or t-norms, see (15)) of all attributes values for all records is not necessary. Moreover, obtained summaries are much more interesting, informative, and close to natural language.

Let us preselect a qualification $w_g = S_g$ from among S_1, S_2,..., S_n. The general form of such a summary is

$$Q \text{ objects from } \mathcal{Y} \text{ being/having } w_g \text{ are/have } S\ [T] \tag{16}$$

and finding the degree of truth is a bit different in this case; the membership function of the summarizer must be reformulated from (15) to

$$\mu_S(d_i) = \min_{j=1,2,\ldots,n} \left\{ \mu_{S_j}\big(V_j(y_i)\big)\ t\ \mu_{w_g}\big(V_g(y_i)\big) \right\}, i = 1, 2, ..., m \tag{17}$$

where the cofactor $\mu_{w_g}\big(V_g(y_i)\big)$ means that only the tuples with the non-zero memberships to S are considered in final results; other records are not considered. It must be explained that it is necessary to preselect a database $\mathcal{D}' \subseteq \mathcal{D}$ consisting of those records d_i only for which $\mu_{w_g}(d_i) > 0$, and $|\mathcal{D}'| = m'$; otherwise, the computation via (17) would be more, instead of less, complicated. The total membership r is

$$r = \frac{\sum_{i=1}^{m'} \mu_S(d_i)}{\sum_{i=1}^{m'} \mu_{w_g}\big(V_g(y_i)\big)} \tag{18}$$

which is similar to the total membership (12) divided by m, but differs in the form of the denominator that is the sum of memberships to the w_g feature for all objects in $\mathcal{D}$. Notice that in this case, only the relative quantification is possible, which is suggested by the specific form of r that relates the total membership to the sum of memberships to w_g.

Not only numerical data can be summarized; an original approach to textual database mining and summarizing is presented by Ochelska, Niewiadomski, and Szczepaniak (2001), and by Ochelska, Szczepaniak, and Niewiadomski (2004). The characteristic point there is a summarizer in the textual form, whose membership function is computed according to its similarity to a given textual pattern. Moreover, the approach is enriched by the application of intuitionistic fuzzy sets (Atanassov, 1999).

[1]Another possible manner is to seek these records only, for which a membership function takes the greatest values, or at least equals an assumed threshold.

2.3. Quality measures for linguistic summaries

The method of determining a quality measure T for a linguistic summary in the basic form by Yager, depends essentially on membership functions of summarizers and/or quantifiers. When the summarizer or quantifier is determined without sufficient expert knowledge, e.g. the membership function of YOUNG MAN takes 1 on the whole $[0, 120]$ interval, the informativeness of a summary is, in fact, none, even if its degree of truth equals 1.

That fact was noticed very early – Yager (1982), Yager, Ford, and Canas (1990), George and Srikanth (1996), Kacprzyk and Yager (2001), and Kacprzyk, Yager, and Zadrożny (2000, 2001) defined various modifications and improvements, which enable to eliminate, at least partially, the problem of subjective quality measures for linguistic summaries. For instance, Yager, apart from his fundamental T index, defines also the *informativeness of a summary*, I, which is computed on the basis of relations between a summary and its "complement" in the form of (Q^c, S^c, T), where Q^c and S^c are the complements $(1 - \mu(\cdot))$ for Q and for S, respectively (Yager, Ford, and Canas, 1991).

Other two quality measures are defined by George and Srikanth (1996) and named *constraint descriptor* and *constituent descriptor*. Both of them are summaries of a given database, and the former one is the summary that concerns as large as possible number of records with attributes meeting at least an assumed threshold of membership, while the latter is the most specific summary that grasps the largest number of records in a database.

The quality indices of knowledge mined from databases are defined by Traczyk (1997); they can express degrees of truth due to lengths of sentences expressing some facts, or due to the fuzzy set modeling properties, shape and "behaviour". These indices are reformulated and applied by Kacprzyk, Yager, and Zadrożny (2000, 2001) to determine *imprecision*, *covering*, *appropriateness*, and *length* for a linguistic summary; these qualities are expressed with real numbers from $[0, 1]$. Furthermore, the method of finding the optimum summary for a given database is also presented as an optimization task.

3. Interval-valued fuzzy sets

3.1. Basic concepts

The main idea of an interval-valued fuzzy set is based on two, instead of one, membership mappings (Turksen, 1986; Gorzałczany, 1987, 1989). They are called, analogously to ordinary fuzzy sets, *the lower membership function* and *the upper membership function*. Both are established on a universe of discourse $\mathcal{X}$ as a domain, and map each element from $\mathcal{X}$ to a real number in the $[0, 1]$ interval.

DEFINITION 3.1 *An interval-valued fuzzy set A in $\mathcal{X}$ is a (crisp) set of ordered triples*

$$A =_{df} \{< x, \underline{\mu}_A(x), \overline{\mu}_A(x) >: x \in \mathcal{X}; \underline{\mu}_A, \overline{\mu}_A \colon \mathcal{X} \to [0, 1]\} \tag{19}$$

where: $\underline{\mu}, \overline{\mu}$ *are the lower and the upper membership functions, respectively, satisfying the following condition*

$$0 \leq \underline{\mu}_A(x) \leq \overline{\mu}_A(x) \leq 1 \ \ \forall x \in \mathcal{X}. \tag{20}$$

With respect to the name of this kind of fuzzy sets, *Interval-Valued*, values of $\underline{\mu}_A$ and $\overline{\mu}_A$, computed for any $x \in \mathcal{X}$ have the interpretation of the lower and upper bounds of the interval number which is the membership degree for x to the set A. That interval is included in $[0, 1]$ and closed on both ends.

EXAMPLE 3.1 *Let* $\mathcal{X} = \{36.0, 36.5, 37.0, 37.5\}$*. The interval valued fuzzy set* A *in* $\mathcal{X}$ *which is a model of the predicate "regular temperature of a healthy human body" is defined as*

$$\begin{aligned} A \ \ = \{ \ \ & < 36.0, [0.3, 0.6] >, < 36.5, [0.8, 1.0] >, \\ & < 37.0, [0.0, 0.5] >, < 37.5, [0.0, 0.0] > \} \end{aligned} \tag{21}$$

A sample interpretation of the element $< 36.5, [0.8, 1] >$ *is: the minimal grade of acceptability that temperature of 36.5°C suitably characterizes a healthy human body is* 0.8*; the maximal acceptability is* 1.0*.*

This method of data representing is very promising when it is impossible to determine membership degrees as real-valued. In data summarization, interval-valued fuzzy sets may be a very useful tool when applied as models for linguistic statements expressing both amounts and properties of objects described by records. They could be especially fruitful when:

1. summarized information is of the interval character, as frequently occurs in technical and engineering data, e.g. *air pressure in tires: 220–250 kPa* or *Device powered with 220–230V*, in medical and economical diagnosis, expert opinions, measurements, and reducing these intervals to reals may cause the loss of information (i.e. presenting it too tersely or laconically)[2], or
2. membership functions of quantifiers and/or summarizers are constructed according to a few different sets of data (e.g. experts opinions; compare Example 3.1) and it is required to maintain this uncertainty rather than computing average values.

Two interval-valued fuzzy sets A, B in $\mathcal{X}$ are *equal* if and only if their lower and upper membership functions take the same values on $\mathcal{X}$:

$$A = B \longleftrightarrow \underline{\mu}_A(x) = \underline{\mu}_B(x) \wedge \overline{\mu}_A(x) = \overline{\mu}_B(x) \ \ \forall x \in \mathcal{X}. \tag{22}$$

An interval-valued fuzzy set (IVFS) A in $\mathcal{X}$ is empty iff $\underline{\mu}_A(x) = \overline{\mu}_A(x) = 0 \ \ \forall x \in \mathcal{X}$.

From the point of view of using interval-valued fuzzy sets as summarizers it is crucial to define *cardinality* of an IVFS:

[2] Such data must be, due to their semantics, collected and modelled with some "margin of safety", and, in fact, they always consist of intervals expressing memberships and other quantities. Very intuitive and convincing explanations, and motivating examples of processing medical interval data are presented by Chen, de Korvin, and Hu (2002).

DEFINITION 3.2 *Let A be an interval-valued fuzzy set in $\mathcal{X}$. The* cardinality *of A is the interval number*

$$card(A) = [\underline{card}(A), \overline{card}(A)] = \left[\sum_{x \in \mathcal{X}} \underline{\mu}_A(x_i), \sum_{x \in \mathcal{X}} \overline{\mu}_A(x_i)\right]. \tag{23}$$

Naturally,

$$0 \leq \underline{card}(A) \leq \overline{card}(A) \leq card(\mathcal{X}) \tag{24}$$

Some arithmetic operations on interval-numbers, defined by Hu (1997), Hu, Xu, and Yang (2002), Moore and Lodwick (2003), and by Sengupta, Pal, and Chakraborty (2001), have been recalled since they are useful in operating on cardinalities of interval-valued fuzzy sets. Let $a = [\underline{a}, \overline{a}], b = [\underline{b}, \overline{b}]$ be intervals in $\mathbb{R}$, and $r \in \mathbb{R}^+$. The arithmetic operations '$+$', '$-$', '$\cdot$' and power are defined

$$[\underline{a}, \overline{a}] + [\underline{b}, \overline{b}] = [\underline{a} + \underline{b}, \overline{a} + \overline{b}] \tag{25}$$

$$[\underline{a}, \overline{a}] - [\underline{b}, \overline{b}] = [\underline{a} - \overline{b}, \overline{a} - \underline{b}] \tag{26}$$

$$[\underline{a}, \overline{a}] \cdot [\underline{b}, \overline{b}] = \left[\min\{\underline{a} \cdot \underline{b}, \underline{a} \cdot \overline{b}, \overline{a} \cdot \underline{b}, \overline{a} \cdot \overline{b}\}, \max\{\underline{a} \cdot \underline{b}, \underline{a} \cdot \overline{b}, \overline{a} \cdot \underline{b}, \overline{a} \cdot \overline{b}\}\right] \tag{27}$$

$$([\underline{a}, \overline{a}])^r = [\underline{a}^r, \overline{a}^{\,r}] \quad \text{for non-negative } \underline{a}, \overline{a}. \tag{28}$$

In addition, the division of a by b, $b \neq 0$ is defined as the multiplication of a by $\left[\frac{1}{\overline{b}}, \frac{1}{\underline{b}}\right]$. The variant of (27) – the operation of multiplying/dividing an interval by a positive real number $r \in \mathbb{R}^+$ is needed frequently. If r is treated as the degenerated interval $[r, r]$, (27) is in the form of

$$[\underline{a}, \overline{a}] \cdot r = [\underline{a} \cdot r, \overline{a} \cdot r]. \tag{29}$$

It must be emphasized that if operations (25)–(28) are to be used in processing membership values in interval-valued fuzzy sets, then some additional restrictions must be taken to ensure that the set of all intervals in $[0, 1]$ (denoted as $Int([0, 1])$) is closed under these operations.

The complement of an interval-valued fuzzy set A in $\mathcal{X}$ is denoted as A^c and its membership function has the form

$$\mu_{A^c}(x) = 1 - \mu_A(x) = \left[1 - \overline{\mu}_A(x), 1 - \underline{\mu}_A(x)\right] \quad \forall x \in \mathcal{X}. \tag{30}$$

It may be noticed that

$$\underline{card}(A^c) = \sum_{x \in \mathcal{X}} \left(1 - \overline{\mu}_A(x)\right) \tag{31}$$

$$\overline{card}(A^c) = \sum_{x \in \mathcal{X}} \left(1 - \underline{\mu}_A(x)\right) \tag{32}$$

$$\underline{card}(A) + \overline{card}(A^c) = \overline{card}(A) + \underline{card}(A^c) = card(\mathcal{X}) \tag{33}$$

and

$$\left(A^c\right)^c = A. \tag{34}$$

The operations of union and intersection for interval-valued fuzzy sets are defined by triangular norms. Let A, B be interval-valued fuzzy sets in $\mathcal{X}$, t – a t-norm and s – an s-norm (t-conorm). The union of A and B is the interval-valued fuzzy set $A \cup B$ with the membership function

$$\mu_{A\cup B}(x) = \left[\underline{\mu}_A(x)\ s\ \underline{\mu}_B(x), \overline{\mu}_A(x)\ s\ \overline{\mu}_B(x)\right] \tag{35}$$

and the intersection of A and B is the interval-valued fuzzy set $A \cap B$ in which

$$\mu_{A\cap B}(x) = \left[\underline{\mu}_A(x)\ t\ \underline{\mu}_B(x), \overline{\mu}_A(x)\ t\ \overline{\mu}_B(x)\right]. \tag{36}$$

Thus, de Morgan laws for interval-valued fuzzy sets A, B in $\mathcal{X}$ are

$$(A \cup B)^c = A^c \cap B^c \tag{37}$$

$$(A \cap B)^c = A^c \cup B^c. \tag{38}$$

Moreover

$$A \cup A = A \tag{39}$$

$$A \cap A = A \tag{40}$$

but usually

$$A \cup A^c \neq \mathcal{X} \tag{41}$$

$$A \cap A^c \neq \emptyset. \tag{42}$$

3.2. Type-reduction

The operations that enable converting an interval-valued fuzzy set into an ordinary fuzzy set and maintaining, at least partially, pieces of information stored in the former one, are frequently needed. This kind of operation is called, in some wider sense, *type-reduction* by Karnik and Mendel (1998, 1999) and Mendel (2001). Here, the field of interest is limited only to obtaining ordinary fuzzy sets from interval-valued fuzzy sets.

DEFINITION 3.3 *Let A be an interval-valued fuzzy set in $\mathcal{X}$, and $\underline{\mu}_A(x)$, $\overline{\mu}_A(x)$ be its lower and upper membership functions, respectively. The following operations, which transform A into an ordinary fuzzy set are type-reductions*

$$TR_{opt}(A) = \{< x, \overline{\mu}_A(x) >: x \in \mathcal{X}\} \tag{43}$$

$$TR_{pes}(A) = \{< x, \underline{\mu}_A(x) >: x \in \mathcal{X}\} \tag{44}$$

$$TR_{re}(A) = \left\{< x, 0.5 \cdot \left(\underline{\mu}_A(x) + \overline{\mu}_A(x)\right) >: x \in \mathcal{X}\right\}. \tag{45}$$

Here, lower indices mean: opt – *optimistic*, pes – *pessimistic*, and re – *realistic*. ”Optimistic” means that in the ordinary fuzzy set which is the result of (43), membership degrees for x's, are taken as the highest (the most optimistic) degrees in A and (44), (45) are defined accordingly. It is also possible to modify (45) with a weighted average of $\underline{\mu}_A(x)$ and $\overline{\mu}_A(x)$, if there are premises that one of membership levels could be more influential on a description of phenomenon than the other. In that case, (45) is in the form of

$$TR_{rew}(A) = \big\{ < x, w_1 \cdot \underline{\mu}_A(x) + w_2 \cdot \overline{\mu}_A(x) >: x \in \mathcal{X} \big\} \tag{46}$$

provided that $w_1 + w_2 = 1$. Formula (46) is a generalization of the definitions above. In particular, the optimistic variant, (43), may be written as (46) with $w_1 = 0$ and $w_2 = 1$. Analogously, the pessimistic variant, (44), is computed via (46) for $w_1 = 1$ and $w_2 = 0$. In the case of the realistic variant, (45), $w_1 = w_2 = 0.5$. It can be noticed that

$$card(TR_{pes}(A)) = \underline{card}(A) \tag{47}$$

$$card(TR_{opt}(A)) = \overline{card}(A) \tag{48}$$

where $card(TR_{opt}(A))$, $card(TR_{pes}(A))$ are cardinalities of ordinary fuzzy sets.

The operations of type-reduction can be applied when a crisp value must be extracted as a final result of a computational or thinking process, but the only accessible data are of interval character.

3.3. Interval-valued linguistic variables

3.3.1. Preliminaries

The concept of linguistic variable exists, thanks to Zadeh (1975), in scientific literature since 1975. This simple and useful construction is described in Section 1.3. The idea of *interval-valued linguistic variable* is presented here as the enhancement of ordinary linguistic variable in which ordinary fuzzy sets are replaced by interval-valued fuzzy sets. The construction of interval-valued linguistic variables has already been mentioned in Mendel (2001) and Niewiadomski (2005a):

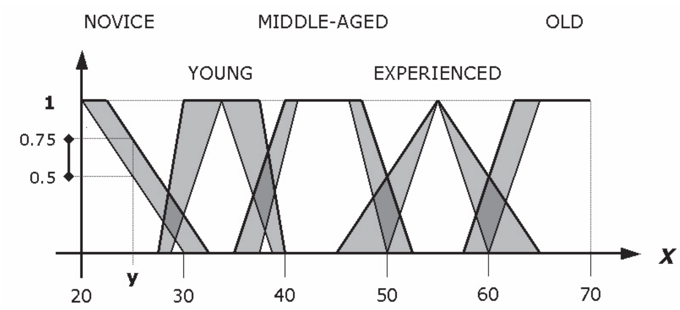

Figure 1. Interval-valued linguistic variable

DEFINITION 3.4 *An interval-valued linguistic variable is an ordered quintuple* $< L, H, \mathcal{X}, G, M >$*, where:*
L *is the name of the variable,*
H *or* $H(L)$ *is the set of linguistic values of* L *(the term-set),*
$\mathcal{X}$ *is the universe of discourse,*
G *is a syntactic rule which generates values (labels) of* L*,*
M *is a semantic rule which associates a term to an interval-valued fuzzy set in* $\mathcal{X}$*.*

As the consequence, compatibility levels of linguistic statements modelled by interval-valued linguistic variables are interval numbers in the $[0, 1]$ interval.

EXAMPLE 3.2 *Let* $\mathcal{X}$ *= [20,70]. The interval-valued linguistic variable* AGE *which describes the age of workers, is defined by the set of its values* $H = \{$NOVICE, YOUNG, MIDDLE-AGED, EXPERIENCED, OLD $\}$. *The values of* AGE *are modelled by the interval-valued fuzzy sets* S_1*,...,* S_5 *in* $\mathcal{X}$*, respectively, and their membership functions are given in Fig. 1. Let* y *be an employee who is 25. The compatibility level of the sample statement "*y *is* a NOVICE*" is the interval* $[0.5, 0.75]$.

As it can be concluded from Definition 3.4 and Example 3.2, each ordinary linguistic variable L' is a special case of an interval-valued linguistic variable L, it is assumed that crisp values of ordinary membership functions are equivalent to the degenerated intervals – values of an interval-valued membership function.

3.3.2. Ranking interval-valued compatibility levels

The methods, which enable comparing intervals are necessary to determine the total order, or, at least, a partial order among the values of membership functions of interval-valued fuzzy sets. The need for such comparison is visible when two intervals, corresponding to greater/smaller compatibility value, should be ranked. In other words, an answer to the following question is sought: *which of two (or more) different intervals shows greater degree of truth?* or, simply, *is interval* a *more/less than interval* *b?* The following partial order relation for intervals $a = [\underline{a}, \overline{a}], b = [\underline{b}, \overline{b}]$ is defined (see Lin, 2002; Ishibuchi and Tanaka, 1990; Sengupta, Pal, and Chakraborty, 2001):

$$a < b \leftrightarrow \overline{a} < \underline{b} \tag{49}$$

$$a \leq b \leftrightarrow \underline{a} \leq \underline{b} \wedge \overline{a} \leq \overline{b}. \tag{50}$$

Ishibuchi and Tanaka (1990) defined another ordering relation for intervals in $\mathbb{R}$. The number $m(a) = \frac{\underline{a}+\overline{a}}{2}$ is termed *the mid-point of* a, and the number $w(a) = \frac{\overline{a}-\underline{a}}{2}$ is termed *the half of width of* a. Indices $m(a)$ and $w(a)$ are an alternative representation for an interval a. They may define the partial order relation on $Int([0,1])$:

$$a \leq b \longleftrightarrow m(a) \leq m(b) \wedge w(a) \geq w(b). \tag{51}$$

In case of a pair of intervals that cannot be compared via (49)–(51) (e.g. $a = [0.1, 0.9]$ and $b = [0.4, 0.5]$), some other definitions must be given. The method of ranking such intervals is defined in Sengupta, Pal, and Chakraborty (2001). The term *grade of acceptability* of the sentence *interval* a *is less then* b is introduced and formulated as

$$\mu_{(\leq)}(a,b) = \frac{m(b) - m(a)}{w(a) + w(b)} \quad \forall a, b \in Int([0,1]). \tag{52}$$

Formula (52) can be easily converted to the form of the membership function of the fuzzy relation on $\big(Int([0,1])\big)$ which represents *Interval* a *is more/less than interval* b, if only all the values of $\mu_{(\leq)}$ exceeding 1 are reduced to 1 and all the values under 0 are treated as 0.

3.3.3. Operations on interval-valued linguistic variables

Interval-valued linguistic variables can be applied to represent the composite terms which contain the connectives AND, OR, and NOT. In case of interval-valued fuzzy sets – models of interval-valued linguistic variable labels – the new methods (with respect to Definition 3.4) are necessary.

DEFINITION 3.5 *Let L be an interval-valued linguistic variable, and $\mathcal{X}$ be its universe of discourse. Let S_1, S_2,..., S_n be the labels of L which are modelled by interval-valued fuzzy sets in $\mathcal{X}$. The* AND, OR *and* NOT *connectives are modelled via the intersection, union, and complement operations for IVFSs, respectively. In particular, the compatibility values for the composite terms:*

1. *x is S_i* AND *S_j, $i, j \leq n$*
2. *x is S_i* OR *S_j, $i, j \leq n$*
3. *x is* NOT *S_i, $i \leq n$*

where $x \in \mathcal{X}$, are computed as intervals a, b, c:

$$a = [\underline{a}, \overline{a}] = \Big[\min\{\underline{\mu}_{S_i}(x), \underline{\mu}_{S_j}(x)\}, \min\{\overline{\mu}_{S_i}(x), \overline{\mu}_{S_j}(x)\}\Big] \tag{53}$$

$$b = [\underline{b}, \overline{b}] = \Big[\max\{\underline{\mu}_{S_i}(x), \underline{\mu}_{S_j}(x)\}, \max\{\overline{\mu}_{S_i}(x), \overline{\mu}_{S_j}(x)\}\Big] \tag{54}$$

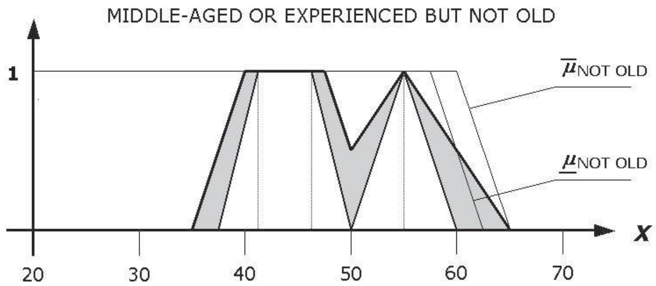

Figure 2. The interval-valued membership function for MIDDLE-AGED OR EXPERIENCED BUT NOT OLD in Example 3.3.

$$c = [\underline{c}, \overline{c}] = \left[1 - \overline{\mu}_{S_i}(x), 1 - \underline{\mu}_{S_i}(x)\right] \tag{55}$$

respectively, where $i, j \leq n$.

This definition is an extension of the analogous definition for ordinary linguistic variables (Zadeh, 1975).

EXAMPLE 3.3 *Let* L *be the linguistic variable as in Example 3.2 and Fig. 1. We construct the interval-valued membership function for the composite term* MIDDLE-AGED OR EXPERIENCED BUT NOT OLD. *The properties* MIDDLE-AGED, EXPERIENCED, *and* OLD *are modelled by* S_3, S_4, *and* S_5, *respectively, and* OR *is in the form of (54) and* BUT=AND *suggests the construction (*S_3 OR S_4*)* AND NOT S_5, *where* AND *is given via (53). Hence*

$$\underline{\mu}_{S_3 \text{ OR } S_4 \text{ BUT NOT } S_5}(x) = \min\left\{\max\{\underline{\mu}_{S_3}(x), \underline{\mu}_{S_4}(x)\}, 1 - \overline{\mu}_{S_5}(x)\right\} \tag{56}$$

and

$$\overline{\mu}_{S_3 \text{ OR } S_4 \text{ BUT NOT } S_5}(x) = \min\left\{\max\{\overline{\mu}_{S_3}(x), \overline{\mu}_{S_4}(x)\}, 1 - \underline{\mu}_{S_5}(x)\right\} \tag{57}$$

The function is depicted in Fig. 2.

Now, the enhancement of Definition 3.5 is introduced; it can be applied to build sentences based on two or more linguistic variables.

DEFINITION 3.6 *Let* L_1, L_2 *be interval-valued linguistic variables, and* $\mathcal{X}_1$, $\mathcal{X}_2$ *be their universes of discourse. Let* S_1, S_2 *be the labels modelled by interval-valued fuzzy sets in* $\mathcal{X}_1$, $\mathcal{X}_2$, *respectively, such that* $S_1 \in H(L_1)$, $S_2 \in H(L_2)$. *Let* y *be an object described by crisp values* $\{x_1, x_2\}$, *such that* $x_i \in \mathcal{X}_i$, $i \in \{1, 2\}$. *Let* t *be a* t*-norm, and* s *be an* s*-norm. The intervals* $a, b \in Int([0, 1])$ *– the compatibility levels for the propositions:*

a) y *is* S_1 AND y *is* S_2

b) y *is* S_1 OR y *is* S_2
are computed with formulae (58), (59):

$$a = [\underline{a}, \overline{a}] = \left[\underline{\mu}_{S_i}(x_1)\ t\ \underline{\mu}_{S_j}(x_2), \overline{\mu}_{S_i}(x_1)\ t\ \overline{\mu}_{S_j}(x_2)\right] \tag{58}$$

$$b = [\underline{b}, \overline{b}] = \left[\underline{\mu}_{S_i}(x_1)\ s\ \underline{\mu}_{S_j}(x_2), \overline{\mu}_{S_i}(x_1)\ s\ \overline{\mu}_{S_j}(x_2)\right]. \tag{59}$$

Other two operations on interval-valued linguistic variables, analogous to the known in fuzzy logic *concentration* and *dilation*, are defined. They are extensions of these operations for ordinary fuzzy sets.

DEFINITION 3.7 *Let* A *be an interval-valued fuzzy set in* $\mathcal{X}$. *The membership function of interval-valued fuzzy set* A_{con} *(*A *concentrated) is*

$$\mu_{A_{con}}(x) = [\underline{\mu}_{A_{con}}(x), \overline{\mu}_{A_{con}}(x)] = [\underline{\mu}_A^2(x), \overline{\mu}_A^2(x)], \forall x \in \mathcal{X} \tag{60}$$

and the membership function of interval-valued fuzzy set A_{dil} *(*A *dilated) is*

$$\mu_{A_{dil}}(x) = [\underline{\mu}_{A_{dil}}(x), \overline{\mu}_{A_{dil}}(x)] = [\underline{\mu}_A^{0.5}(x), \overline{\mu}_A^{0.5}(x)], \forall x \in \mathcal{X}. \tag{61}$$

In general, the indices 2 and 0.5 in (60) and (61), respectively, may be replaced by any positive real index, $r > 1$ and $r < 1$, respectively, if only "the strength" of the operation is to be modified. The proposition of modeling the linguistic statements like EXTREMELY or SLIGHTLY by the values of the r index in the concentration/dilation operations is discussed by Chen and Liu (2003).

In addition, the fact that the mapping $(\cdot)^k$, $k \in \mathbb{R}^+$ is strictly increasing on $[0,1]$ assures that $\underline{\mu}_A(x) \leq \overline{\mu}_A(x) \to \underline{\mu}_A^k(x) \leq \overline{\mu}_A^k(x)$, thus the intervals $[\underline{\mu}_{A_{con}}(x), \overline{\mu}_{A_{con}}(x)]$ and $[\underline{\mu}_{A_{dil}}(x), \overline{\mu}_{A_{dil}}(x)]$ are well determined.

3.3.4. Fuzzy quantification and interval-valued linguistic variables

The linguistic quantification of imprecise statements modelled by interval-valued fuzzy sets is presented in this subsection. Degree of truth of a statement in the form of (5), when Q is an ordinary fuzzy quantifier and S_1 is an interval-valued fuzzy set in $\mathcal{X}$, is computed as an interval number in [0,1]

$$T = [\underline{t}, \overline{t}] = \left[\mu_Q\left(\underline{card}(S_1)\right), \mu_Q\left(\overline{card}(S_1)\right)\right] \tag{62}$$

if Q is absolute, or

$$T = [\underline{t}, \overline{t}] = \left[\mu_Q\left(\frac{\underline{card}(S_1)}{card(\mathcal{X})}\right), \mu_Q\left(\frac{\overline{card}(S_1)}{card(\mathcal{X})}\right)\right] \tag{63}$$

if Q is relative. The degree of truth of a linguistically quantified proposition in the form of (6), in which Q is a relative ordinary fuzzy quantifier, S_1 – as given above and S_2 is an ordinary fuzzy set in $\mathcal{X}$, is computed as an interval number in [0,1]:

$$T = [\underline{t}, \overline{t}] = \left[\mu_Q\left(\frac{\underline{card}(S_1 \cap S_2)}{card(S_2)}\right), \mu_Q\left(\frac{\overline{card}(S_1 \cap S_2)}{card(S_2)}\right)\right] \tag{64}$$

where the intersection of the interval-valued S_1 with the ordinary S_2 is computed via (36), and the membership function μ_{S_2} is treated as the degenerated interval-valued membership function $[\underline{\mu}_{S_2}, \overline{\mu}_{S_2}]$.

Formulae (62)–(64) are valid only for the quantifiers with non-decreasing membership functions, otherwise the results could be in the form of an irregular interval T, i.e. in which $\underline{t} \geq \overline{t}$, or even irrelevant, if, e.g. Q has a local maximum in $[\underline{card}(S_1), \overline{card}(S_1)]$. To make them useful also for other shapes of quantifier memberships, it is needed to reformulate them as

$$T = \left[\inf_{r \in [\underline{card}(S_1), \overline{card}(S_1)]} \mu_Q(r) , \sup_{r \in [\underline{card}(S_1), \overline{card}(S_1)]} \mu_Q(r) \right] \tag{65}$$

if Q is absolute, or

$$T = \left[\inf_{r \in \left[\frac{\underline{card}(S_1)}{card(\mathcal{X})}, \frac{\overline{card}(S_1)}{card(\mathcal{X})}\right]} \mu_Q(r) , \sup_{r \in \left[\frac{\underline{card}(S_1)}{card(\mathcal{X})}, \frac{\overline{card}(S_1)}{card(\mathcal{X})}\right]} \mu_Q(r) \right] \tag{66}$$

if Q is relative. For the statements constructed according to Q^{II} it is

$$T = \left[\inf_{r \in \left[\frac{\underline{card}(S_1 \cap S_2)}{card(S_2)}, \frac{\overline{card}(S_1 \cap S_2)}{card(S_2)}\right]} \mu_Q(r) , \sup_{r \in \left[\frac{\underline{card}(S_1 \cap S_2)}{card(S_2)}, \frac{\overline{card}(S_1 \cap S_2)}{card(S_2)}\right]} \mu_Q(r) \right] . \tag{67}$$

Interval-valued linguistic variables are applied in linguistic summaries to express properties with respect to which the database is summarized (called *interval-valued summarizers*) and/or quantities, i.e. quality expressions (called *interval-valued fuzzy linguistic quantifiers*).

3.4. Interval-valued fuzzy quantifiers

Interval-valued fuzzy quantifiers are introduced as an original extension of ordinary fuzzy quantifiers by Zadeh (1983). They model the natural language quantifiers, i.e. natural statements which pronounce quantities of objects as imprecise numbers (e.g. ABOUT 15) and/or ratios (e.g. MUCH LESS THAN 1/4). Interval-valued fuzzy quantifiers are very similar – with respect to their construction, usage, and intuitions – to ordinary fuzzy quantifiers. The only difference is the application of interval-valued fuzzy sets, instead of ordinary, to express quantities. Interval-valued fuzzy quantifiers include ordinary fuzzy quantifiers as special cases, i.e. those with interval-valued but degenerated membership functions.

The following propositions relate selected properties of ordinary fuzzy sets to the analogous properties of interval-valued fuzzy sets:

PROPOSITION 3.1 *An interval-valued fuzzy set A in $\mathcal{X}$ is normal iff $TR_{pes}(A)$ and $TR_{opt}(A)$ are normal.*

PROPOSITION 3.2 *An interval-valued fuzzy set A in $\mathcal{X}$ is convex iff $TR_{pes}(A)$ and $TR_{opt}(A)$ are convex.*

Therefore, the definition of the interval-valued fuzzy quantifier is

DEFINITION 3.8 *Recall equations (5) and (6). Let $\mathcal{Y} \subseteq \mathbb{R}^+ \cup \{0\}$ be a universe of discourse. A normal and convex interval-valued fuzzy set in $\mathcal{Y}$ which is a model of the quantity pronouncement Q in (5) or (6) is an* interval-valued fuzzy quantifier. *Q is* absolute *if $\mathcal{Y} = \mathbb{R}^+ \cup \{0\}$. Q is* relative *if $\mathcal{Y} = [0, 1]$.*

The evaluation of the quantified linguistic statements in the form of (5), (6), in which the properties are modelled by ordinary fuzzy sets and the quantifier – via interval-valued fuzzy sets, proceeds as follows: let Q be an interval-valued fuzzy quantifier, and S_1, S_2 – ordinary fuzzy sets in $\mathcal{X}$. The degree of truth for (5) is computed as the interval number in [0,1]

$$T = [\underline{t}, \overline{t}] = \left[\underline{\mu}_Q \left(card(S_1)\right), \overline{\mu}_Q \left(card(S_1)\right)\right] \tag{68}$$

if Q is absolute, or

$$T = [\underline{t}, \overline{t}] = \left[\underline{\mu}_Q \left(\frac{card(S_1)}{card(\mathcal{X})}\right), \overline{\mu}_Q \left(\frac{card(S_1)}{card(\mathcal{X})}\right)\right] \tag{69}$$

if Q is relative. The degree of truth for a proposition in the form of (6) is

$$T = [\underline{t}, \overline{t}] = \left[\underline{\mu}_Q \left(\frac{card(S_1 \cap S_2)}{card(S_2)}\right), \overline{\mu}_Q \left(\frac{card(S_1 \cap S_2)}{card(S_2)}\right)\right] \tag{70}$$

for a relative Q.

The properties of ordinary relative quantifiers described in Zadeh (1983), Yager (1993) and Liu and Kerre (1998) may be extended to the analogous definitions for interval-valued (also absolute) fuzzy quantifiers.

DEFINITION 3.9 *Let Q be an interval-valued fuzzy quantifier in $\mathbb{R}^+ \cup \{0\}$. Let a_l, a^u, b_l, b^u, c_l, c^u, d_l, $d^u \in \mathbb{R}^+ \cup \{0\}$ be such that $a_l \le b_l \le c_l \le d_l$ and $a^u \le b^u \le c^u \le d^u$. Let $TR_{pes}(Q)$ and $TR_{opt}(Q)$ be convex and normal ordinary fuzzy sets.*
Q is regular non-decreasing (see Fig. 3a) if

1. $\forall x \le a_l \ \underline{\mu}_Q(x) = 0 \wedge \forall x \le a^u \ \overline{\mu}_Q(x) = 0 \wedge a_l \ge a^u$ (71)
2. $\forall x \ge b_l \ \underline{\mu}_Q(x) = 1 \wedge \forall x \ge b^u \ \overline{\mu}_Q(x) = 1 \wedge b_l \ge b^u$ (72)
3. $\forall x_1, x_2 \in [a_l, b_l] \ x_1 \le x_2 \to \underline{\mu}_Q(x_1) \le \underline{\mu}_Q(x_2)$ (73)
4. $\forall x_1, x_2 \in [a^u, b^u] \ x_1 \le x_2 \to \overline{\mu}_Q(x_1) \le \overline{\mu}_Q(x_2).$ (74)

Q is regular non-increasing (Fig. 3b) if

1. $\forall x \le c_l \ \underline{\mu}_Q(x) = 1 \wedge \forall x \le c^u \ \overline{\mu}_Q(x) = 1 \wedge c_l \le c^u$ (75)
2. $\forall x \ge d_l \ \underline{\mu}_Q(x) = 1 \wedge \forall x \le d^u \ \overline{\mu}_Q(x) = 1 \wedge d_l \le d^u$ (76)
3. $\forall x_1, x_2 \in [c_l, d_l] \ x_1 \le x_2 \to \underline{\mu}_Q(x_1) \ge \underline{\mu}_Q(x_2)$ (77)
4. $\forall x_1, x_2 \in [c^u, d^u] \ x_1 \le x_2 \to \overline{\mu}_Q(x_1) \ge \overline{\mu}_Q(x_2).$ (78)

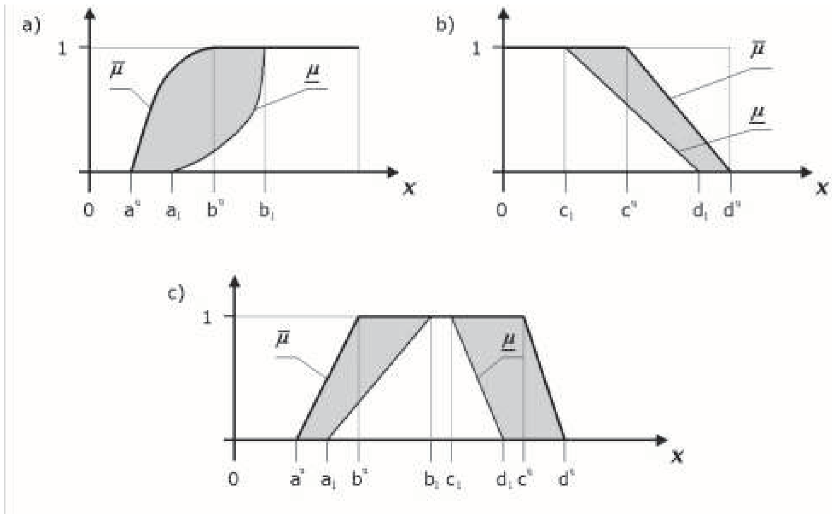

Figure 3. Interval-valued fuzzy quantifiers: a) non-decreasing, b) non-increasing, c) unimodal.

Q is regular unimodal (Fig. 3c) if

1. $\forall x \leq a_l \; \underline{\mu}_Q(x) = 0 \wedge \forall x \geq d_l \; \underline{\mu}_Q(x) = 0$ (79)
2. $\forall x \leq a^u \; \overline{\mu}_Q(x) = 0 \wedge \forall x \geq d^u \; \overline{\mu}_Q(x) = 0$ (80)
3. $\forall x \in [b_l, c_l] \; \underline{\mu}_Q(x) = 1 \wedge \forall x \in [b^u, c^u] \; \overline{\mu}_Q(x) = 1 \wedge b_l \geq b^u \wedge c_l \leq c^u$ (81)
4. *Conditions (73), (74), (77), and (78) are satisfied*. (82)

The properties of non-decreasing, non-increasing, and unimodal quantifiers may be expressed in terms of the analogous properties for ordinary fuzzy quantifiers using the type-reduction operations.

PROPOSITION 3.3 *Let Q be an interval-valued linguistic quantifier.*

a) Q is non-decreasing (non-increasing) iff $TR_{pes}(Q)$ and $TR_{opt}(Q)$ are non-decreasing (non-increasing);
b) Q is unimodal iff $TR_{pes}(Q)$ and $TR_{opt}(Q)$ are unimodal.

The examples of non-decreasing, non-increasing, and unimodal interval-valued fuzzy quantifiers are shown in Fig. 3 a)–c). The original methods applying interval-valued fuzzy quantifiers to data summarization are presented in Section 4.

4. Interval-valued linguistic summaries

When data used for constructing membership functions of quantifiers and/or summarizers, are of the interval form, e.g. experts have determined membership levels with interval values instead of reals, there are at least two possibilities for handling this uncertainty: 1) computing average (arithmetic or weighted) membership levels and use them in ordinary linguistic summarizing, or 2) applying interval-valued memberships when building summaries. Since the latter option includes the former one (e.g. via type-reduction operations which can be applied during the summarization process), let us introduce the concept of *interval-valued linguistic summary of a database*:

DEFINITION 4.1 *An interval-valued linguistic summary of a database is a quasi-natural language sentence*

$$Q\ P\ \textit{are/have}\ S\ [\underline{t},\overline{t}] \tag{83}$$

where the symbols Q, P, *and* S *are interpreted as in (9), Section 2.1, but at least one of* Q, S *is represented by an interval-valued fuzzy set, and* $T=\left[\underline{t},\overline{t}\right]\subseteq[0,1]$ *is an interval-valued degree of truth of the summary.*

The definition extends the classical Yager's approach with the use of interval-valued fuzzy sets and interval-valued linguistic variables as quantifiers and summarizers, respectively.

4.1. Summaries with interval-valued fuzzy quantifiers

According to the assumption made by Yager, the goal is to find a quality index for a given summary in the form of "Q P are/have S". Naturally, if Q is modelled by an interval-valued fuzzy set, then the T index, which is a value of $\mu_Q(r)$, is also an interval: $T=[\underline{t},\overline{t}]$. The semantics of r is the same as in the classical case, see (12), and the computation is based on (10) or on (11). Since S is an ordinary fuzzy set, r is a real number.

Let $\mathcal{Y}=\{y_1,y_2,...,y_m\}$ be a set of objects, and V be an attribute describing objects from $\mathcal{Y}$ with crisp values from $\mathcal{X}$, such that $V(y_i)=x$, $x\in\mathcal{X}$, $i=1,2,...,m$. Hence, the database modeling $\mathcal{Y}$ is represented as $\mathcal{D}=\{V(y_1),V(y_2),...,V(y_m)\}$, which describes the dependence between $\mathcal{X}$ and $\mathcal{Y}$. Let Q be an interval-valued fuzzy quantifier modelled by an interval-valued fuzzy set given by the membership functions $\underline{\mu}_Q$ and $\overline{\mu}_Q$. Constructing and evaluating an interval-valued linguistic summary proceeds:

Step 1 Compute r – the total membership of objects from $\mathcal{Y}$ to the feature S as in (12).

Step 2 Compute the lower and upper bounds of T via membership degrees of r to Q:

$$T=\left[\underline{t},\overline{t}\right]=\left[\underline{\mu}_Q(r),\overline{\mu}_Q(r)\right] \tag{84}$$

if Q is absolute, or

$$T = \left[\underline{t}, \overline{t}\right] = \left[\underline{\mu}_Q\left(\frac{r}{m}\right), \overline{\mu}_Q\left(\frac{r}{m}\right)\right] \tag{85}$$

if Q is relative.

Step 3 Hence, the final form of the summary is

$$Q\ y\text{'s are/have } S\ \left[\underline{t}, \overline{t}\right]. \tag{86}$$

The crucial fact should be noticed: if the interval-valued fuzzy quantifier Q is determined by the degenerated membership function, i.e. $\forall x \in \mathcal{X}\ \underline{\mu}_Q(x) = \overline{\mu}_Q(x)$, an interval-valued linguistic summary is equivalent to an ordinary linguistic summary in sense of Yager, as in Section 2.1.

4.2. Summaries with interval-valued summarizers

We introduce the use of interval-valued linguistic variables in modeling of features of objects in databases. An interval-valued linguistic variable allows to assign intervals as membership values of attributes manifested by the subject of a summary, i.e. the records in a database. Hence, interval-valued summarizers are presented in this section.

Let us define $\mathcal{Y}$ – a set of objects, and a set of attributes $V = \{V_1, V_2, \ldots, V_n\}$, which describe objects from $\mathcal{Y}$. Let $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ be the domains of $V_1, V_2, \ldots, V_n$, respectively, and the symbol $V_j(y_i)$, $i = 1, \ldots, m$, $j = 1, 2, \ldots, n$ denotes a value of attribute V_j for object y_i. Hence, the denotation of the database is adequate to (13). Let $S_1, S_2, \ldots, S_n$ be interval-valued fuzzy sets in $\mathcal{X}_1, \ldots, \mathcal{X}_n$ respectively, representing the features expressed by attributes $V_1, \ldots, V_n$, respectively. Let Q be an ordinary fuzzy quantifier. The general form of this variant of summary, when the summarizer is composed of several single features $S_1, S_2, \ldots, S_n$, is analogous to the form due to George and Srikanth (14):

$$Q \text{ objects from } \mathcal{Y} \text{ are/have } \underbrace{S_1 \text{ AND } S_2 \text{ AND} \ldots S_n}_{S} \left[\underline{t}, \overline{t}\right]. \tag{87}$$

The problem is to determine the construction of an interval-valued multi-featured summarizer S. Since in the original approach, (15), the membership level in S is computed as minimum (in general: as a t-norm) of compatibility levels in S_1,..., S_n, hence, the interval-valued extension must follow this. The definition of the interval-valued form of function (15) is

$$\underline{\mu}_S(d_i) = \min_{j=1,2,\ldots,n} \left\{\underline{\mu}_{S_j}\left(V_j(y_i)\right\}\right\} \tag{88}$$

and

$$\overline{\mu}_S(d_i) = \min_{j=1,2,\ldots,n} \left\{\overline{\mu}_{S_j}\left(V_j(y_i)\right)\right\} \tag{89}$$

where d_i – as in (13). In general, the minima in (88) and (89) may be replaced by t-norms. Another manner of computing the summarizer S built on the basis of several attributes is

$$\mu_S(d_i) = \left[\underline{\mu}_S(d_i), \overline{\mu}_S(d_i)\right] = \overset{*}{\min_{j=1,\ldots,n}} \left\{\left[\underline{\mu}_{S_j}\left(V_j(y_i)\right), \overline{\mu}_{S_j}\left(V_j(y_i)\right)\right]\right\} \tag{90}$$

where the $\overset{*}{\min}$ operation is the choice of the smallest interval via one of operations (49)–(52).

The sum of memberships of all objects in S, r, is computed via the extended formula (12). Having an interval-valued membership function given with (88) and (89), the interval-valued form of the r index is

$$r = [\underline{r}, \overline{r}] = \sum_{i=1}^{m} \left[\underline{\mu}_S(d_i), \overline{\mu}_S(d_i)\right] \tag{91}$$

where the addition operation for interval-numbers is given via (25). Constructing and evaluating an interval-valued linguistic summary with the composite summarizer S proceeds as follows:

Step 1a Compute the interval-valued $r = [\underline{r}, \overline{r}]$ via (91).

Step 1b IF Q is relative, THEN substitute $r: = \frac{r}{m}$.

Step 2 Determine T – the quality of the summary – as

$$T = \left[\inf_{r\in[\underline{r},\overline{r}]} \mu_Q(r), \sup_{r\in[\underline{r},\overline{r}]} \mu_Q(r)\right] \tag{92}$$

(see the note at the end of the procedure).

Step 3 Hence, the final form of the linguistic summary is:

$$Q\ y\text{'s are/have } S\ \left[\underline{t}, \overline{t}\right] \tag{93}$$

where $T = \left[\underline{t}, \overline{t}\right]$ is the interval-valued degree of truth for the summary.

Note Step 2 should be additionally commented upon. Formula (92) determines the interval $[\underline{r}, \overline{r}]$ through μ_Q, even if it is not monotonic and has – as unimodal fuzzy quantifiers have – regular or irregular maxima in this interval. Notice that in case of monotonicity of μ_Q on $[\underline{r}, \overline{r}]$ (92) may be simplified to

$$\mu_Q\left([\underline{r}, \overline{r}]\right) = \begin{cases} [\mu_Q(\underline{r}), \mu_Q(\overline{r})], & \text{if } \mu_Q \text{ increases monotonically on } [\underline{r}, \overline{r}] \\ [\mu_Q(\overline{r}), \mu_Q(\underline{r})], & \text{if } \mu_Q \text{ decreases monotonically on } [\underline{r}, \overline{r}]. \end{cases} \tag{94}$$

4.3. Interval-valued summaries with a query w_g

Analogously to the approach presented by Kacprzyk, Yager and Zadrożny (2000, 2001), it is possible to construct interval-valued linguistic summaries which are built according to the second canonical form of linguistically quantified proposition, i.e. Q^{II}, see (6). The summaries obtained this way are more interesting and informative. Moreover, as mentioned in Section 2.2, the use of the w_g preselection prevents from computing many unnecessary values, mostly the zero membership levels, see (16). The limitation introduced for the classic linguistic summaries, is also applied here, but in the form which is compatible with interval-valued forms of summarizer.

In order to handle queries in the form of w_g, the interval-valued form of the membership function of S should be redefined from (88) and (89) to

$$\underline{\mu}_S(d_i) = \min_{j=1,2,\ldots,n} \left\{ \underline{\mu}_{S_j}(V_j(y_i)) \; t \; \mu_{w_g}(V_g(y_i)) \right\}, \; i = 1, 2, ..., m \tag{95}$$

and

$$\overline{\mu}_S(d_i) = \min_{j=1,2,\ldots,n} \left\{ \overline{\mu}_{S_j}(V_j(y_i)) \; t \; \mu_{w_g}(V_g(y_i)) \right\}, \; i = 1, 2, ..., m. \tag{96}$$

respectively. Computing r requires determining the database $\mathcal{D}' \subseteq \mathcal{D}$ consisting only of records d_i for which $\mu_{w_g}(d_i) > 0$ (otherwise, the computational cost would increase). Assuming that w_g is a real-valued (non-interval-valued) membership function, the interval-valued form of r is

$$r = [\underline{r}, \overline{r}] = \frac{\left[\sum_{i=1}^{m} \underline{\mu}_S(d_i), \sum_{i=1}^{m} \overline{\mu}_S(d_i) \right]}{\sum_{i=1}^{m} \mu_{w_g}\left(V_g(y_i)\right)}. \tag{97}$$

The given method makes it possible to apply the relative quantification only, since r is expressed by the ratio of the total membership to the sum of memberships in the w_g query.

The following observation must be noted here: the method presented does not handle the cases in which a query is described by an interval-valued membership function. It would require defining intervals with interval-valued bounds (see Wu and Mendel, 2002). Suitable methods and the semantics for them are currently being developed.

4.4. Examples

The described methods of summarization are applied as a part of a larger experiment in the field of e-testing (distance and automated testing in the process of learning over the Internet). The subject of the experiment is to run and score grammar tests in German for e-students. Since the experiment requires multidisciplinary knowledge, the additional task, solved by the use of linguistic summaries, is to support the natural (or close-to-natural) language data interpretation for experts in different domains, such as philology, computer science, methodology, etc. Further details can be found in Niewiadomski, Bartyzel, and Szczepaniak (2005), Niewiadomski (2005c), and Niewiadomski et al. (2005).

The set of $m = 243$ correct, partially correct and incorrect answers to 50 questions was collected. The answers were then scored by three experts separately; the scale used in scoring is 0 – incorrect, 0.25, 0.5, 0.75, and 1 – totally correct. The results – the scores given to answers – are stored in the vectors E_1, E_2, E_3, where $E_1 = \{e_{1,1}, e_{1,2}, ..., e_{1,243}\}$ collects scores by Expert 1, and E_2, E_3 – analogously for Experts 2 and 3. Thus, the database to be summarized is:

ID	**Score**	**Expert**
1	0.75	1
2	0.5	3
3	1	2
4	0.5	med
...	...	...
n	0	med

(98)

where "med" means the median of scores given by Experts 1–3 to an answer, and ID is the key of the table. The sample linguistic quantifiers used in the experiment are[3]:

$$\mu_{\text{FEW}}(x) = \exp\left(-\left(\frac{x-0.24}{0.13}\right)^2\right) \tag{99}$$

$$\mu_{\text{MANY}}(x) = \exp\left(-\left(\frac{x-0.76}{0.13}\right)^2\right) \tag{100}$$

$$\mu_{\text{ALMOST ALL}}(x) = \exp\left(-\left(\frac{x-1}{0.1}\right)^2\right). \tag{101}$$

The sample interval-valued summarizers (describing scores) are used in the experiment in the following form:

$$\underline{\mu}_{\text{HIGH}}(e) = \begin{cases} e, & \text{if } e \in [0.5, 1] \\ 0, & \text{otherwise} \end{cases} \tag{102}$$

$$\overline{\mu}_{\text{HIGH}}(e) = \begin{cases} \frac{e+1}{2}, & \text{if } e \in [0.5, 1] \\ 0, & \text{otherwise} \end{cases} \tag{103}$$

$$\underline{\mu}_{\text{LOW}}(e) = \begin{cases} -e+1, & \text{if } e \in [0.5, 1] \\ 0, & \text{otherwise} \end{cases} \tag{104}$$

$$\overline{\mu}_{\text{LOW}}(e) = \begin{cases} \frac{-e+1}{2}, & \text{if } e \in [0.5, 1] \\ 0, & \text{otherwise} \end{cases} \tag{105}$$

where e is a score from the "Score" column in the database (98).

Constructing and evaluating the sample summary MANY *scores given by experts are* HIGH proceeds as given in Section 4.2: let $m = 972$ (number of records), $j = 1$

[3] As it is seen, the quantifiers presented here are built with ordinary fuzzy sets; such solution has been chosen to present the new material as clearly as possible. The summarizers used in the experiment are interval-valued; the interval-valued fuzzy quantifiers have also been considered but finally omitted as irrelevant from the point of view of expected results.

(number of summarized attributes), S = HIGH [score] – the summarizer, Q = MANY – the quantifier.

Step 1

$$r = \sum_{i=1}^{m} \left[\underline{\mu}_S(d_i), \overline{\mu}_S(d_i)\right] = [676.75, 755.38] \tag{106}$$

thus

$$\frac{r}{m} = \frac{[676.75, 755.38]}{972} = [0.70, 0.78] \tag{107}$$

according to (91).

Step 2

$$\inf_{r \in [0.70, 0.78]} \mu_{\mathrm{MANY}} = 0.79 \tag{108}$$

and

$$\sup_{r \in [0.70, 0.78]} \mu_{\mathrm{MANY}} = 1 \tag{109}$$

for relative Q.

Step 3 Hence, the final form of the summary is

MANY scores are HIGH $[0.79, 1]$ (110)

so its degree of truth is relatively high.

Another sample summary: FEW *scores by Expert 1 are* LOW is obtained according to the method described in Section 4.3. The additional element is the query BY EXPERT 1 which means that not all the tuples in the database are considered, but only those in which the field ”Expert”, see (98), takes the ”1” value, which means ”by Expert 1”. Let $m = 972$ (number of records), $j = 1$ (number of summarized attributes), S = LOW [score] – the summarizer, w_g = BY EXPERT 1 – the query, Q = FEW – the quantifier.

Step 1

$$r = \frac{\sum_{i=1}^{m} \left[\underline{\mu}_S(d_i), \overline{\mu}_S(d_i)\right]}{\sum_{i=1}^{m} \mu_{w_g}(d_i)} = \frac{[35, 84]}{243} = [0.17, 0.35] \tag{111}$$

according to (95)–(97).

Step 2

$$\inf_{r \in [0.17, 0.35]} \mu_{\mathrm{MANY}}(r) = 0.52 \tag{112}$$

and

$$\sup_{r \in [0.17, 0.35]} \mu_{\mathrm{MANY}}(r) = 1. \tag{113}$$

Step 3 Hence, the final form of the summary is

$$\text{FEW scores BY EXPERT 1 are LOW } [0.52, 1]. \tag{114}$$

5. Conclusions and future work

This paper presents the concept of interval-valued linguistic summary as a tool for distilling the most important information from a large number of tuples and presenting obtained results in a linguistic form. The approach is an extension of the Yager approach in the same sense as an interval-valued fuzzy set extends the idea of an ordinary fuzzy set, and the former includes the latter as a special case. The interval-valued forms of summarizers and fuzzy quantifiers are defined, exemplified, and applied in summarizing a sample database. The direct consequence of using intervals as membership values is the interval form of the T index – the degree of truth of summary. Finally, two illustrative examples have been provided.

At least two additional concepts related to the introduced summarization methods should be examined in the nearest future: 1) the interval-valued form of the w_g query, which is considered here as an ordinary fuzzy set only (see Section 4.3), and 2) interval-valued-based extensions of quality measures. Till now, only the indices by Kacprzyk, Yager, and Zadrożny have been extended to interval-valued forms and described in Niewiadomski (2005d).

Currently, the authors are working on a further extension of linguistic summarizing in the sense of Yager – it is based on type-2 fuzzy sets (Karnik, Mendel, 1998, 1999; Mendel, 2001), which are a promising field for modeling uncertain data by fuzzy membership levels in fuzzy sets. The very first concepts have been given already in Niewiadomski (2005b). The crucial observation that interval-valued fuzzy sets are the equivalent of interval type-2 fuzzy sets, and therefore the equivalence of interval-valued linguistic summaries to interval type-2 linguistic summaries may be observed, has been made. The type-2-based extension is a further generalization of interval-valued linguistic summaries presented here, and, in consequence, includes also the Yager approach.

References

ATANASSOV, K.T. (1999) *Intuitionistic Fuzzy Sets. Theory and Applications*. Springer Verlag.

BOSC, P. and PIVERT, O. (1992) Fuzzy querying in conventional databases. In: L.A. Zadeh, J. Kacprzyk, eds., *Fuzzy Logic for the Management of Uncertainty*. Wiley, New York, 645–671.

BOSC, P., DUBOIS, D., PIVERT, O., PRADE, H. and DE CALMES, M. (2002) Fuzzy summarization of data using fuzzy cardinalities. *Proceedings of IPMU'2002*, Annecy, France, 1553–1559.

CHEN, P., DE KORVIN, A. and HU, C. (2002) Association Analysis with Interval Valued Fuzzy Sets and Body of Evidence. *Proceedings of the 2002 IEEE International Conference on Fuzzy Systems*, Honolulu, HI, 518–523.

CHEN, C.-Y. and LIU, B.-D. (2003) Linguistic Hedges and Fuzzy Rule Based Systems. In: J. Cassillas, O. Cordon, F. Herrera, L.Magdalena, eds., *Accuracy Improvement in Linguistic Fuzzy Modelling*. Physica-Verlag, c/o Springer-Verlag, Heidelberg, New York.

GEORGE, R. and SRIKANTH, R. (1996) Data Summarization Using Genetic Algorithms and Fuzzy Logic. In: F. Herrera, J.L. Verdegay, eds., *Genetic Algorithms and Soft Computing*, Physica–Verlag, Heidelberg, 599–611.

GORZAŁCZANY, M.B. (1987) A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems* **21**, 1–17.

GORZAŁCZANY, M.B. (1989) An interval-valued fuzzy inference method in approximate reasoning. *Fuzzy Sets and Systems* **31**, 243–251.

HU, C. (1997) Reliable Computing with Interval Arithmetic. *Proc. of the International Workshop on Computational Science and Engineering*. Press of University of Science and Technology of China, 17–22.

HU, C., XU, S. and YANG, X. (2002) A Review on Interval Computation – Software and Applications. *Int. J. of Computational and Numerical Analysis and Applications* **1** (2), 149–162.

ISHIBUCHI, H. and TANAKA, H. (1990) Multiobjective programming in optimisation of the interval objective function. *European Journal of Operational Research* **48**, 219–225.

KACPRZYK, J. and YAGER, R.R. (2001) Linguistic summaries of data using fuzzy logic. *International Journal of General Systems* **30**, 133–154.

KACPRZYK, J., YAGER, R.R. and ZADROŻNY, S. (2000) A fuzzy logic based approach to linguistic summaries of databases. *International Journal of Applied Mathematics and Computer Sciences* **10**, 813–834.

KACPRZYK, J., YAGER, R.R., ZADROŻNY, S. (2001) Fuzzy linguistic summaries of databases for an efficient business data analysis and decision support. In: W. Abramowicz, J. Żurada, eds., *Knowledge Discovery for Business Information Systems*. Kluwer Academic Publisher B. V., Boston, 129–152.

KARNIK, N.N. and MENDEL, J.M. (1998) *An Introduction to Type-2 Fuzzy Logic Systems*. University of Southern California, Los Angeles.

KARNIK, N.N. and MENDEL, J.M. (1999) Type-2 Fuzzy Logic Systems. *IEEE Transactions on Fuzzy Systems* **7**, (6), 643–658.

LIN, F.-T. (2002) Fuzzy Job-Shop Scheduling Based on Ranking Level ('1) Interval-Valued Fuzzy Numbers. *IEEE Transactions On Fuzzy Systems* **10** (4), 510–522.

LIU, Y. and KERRE, E.E. (1998) An Overview of Fuzzy Quantifiers, Part I: Interpretations. *Fuzzy Sets and Systems* **95**, 1–21.

LODWICK, W.A. and JAMISON, K.D. (2003) Special issue on interfaces between fuzzy set theory and interval analysis. *Fuzzy Sets and Systems* **135**, 1–3.

LOSLEVER, P. and BOUILLAND, S. (1999) Marriage of fuzzy sets and multiple correspondence analysis. Examples with subjective interval data and biomedic signals. *Fuzzy Sets and Systems* **107**, 255–275.

MANI, I. and MAYBURY, M.T., EDS., (1999) *Advances in Automatic Text Summarization*. The MIT Press, Cambridge Massachusetts, USA.

MENDEL, J.M. (2001) *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice-Hall, Upper Saddle River, NJ.

MOORE, R. and LODWICK, W. (2003) Interval analysis and fuzzy set theory. *Fuzzy Sets and Systems* **135**, 5–9.

NIEWIADOMSKI, A. (2005a) Interval-valued linguistic variables. An application to linguistic summaries. In: O. Hryniewicz, J. Kacprzyk, J. Koronacki, S.T. Wierzchoń, *Issues in Intelligent Systems. Paradigms*. The Academic Press EXIT, Warsaw, 167–184.

NIEWIADOMSKI, A. (2005b) On Two Possible Roles of Type-2 Fuzzy Sets in Linguistic Summaries. *Lecture Notes in Artificial Intelligence* **3528**, 341–347.

NIEWIADOMSKI, A. (2005c) Interval-valued data structures and their applications in e-learning. *Lecture Notes in Computer Science* **3381**, 403–407.

NIEWIADOMSKI, A. (2005d) Interval-valued quality measures for linguistic summaries. In: P. Grzegorzewski, M. Krawczak, S. Zadrożny, eds., *Issues in Soft Computing. Theory and Applications*. EXIT Academic Press, Warszawa, 211-224.

NIEWIADOMSKI, A., BARTYZEL, M. and SZCZEPANIAK, P.S. (2005) Linguistic summaries of databases in evaluating algorithms for automated distance testing. *Proceedings of XV KKA Conference, Warsaw, Poland, June 27–30* (in Polish), vol. 3, 81–86.

NIEWIADOMSKI, A., RYBUSIŃSKI, B., SAKOWSKI, K. and GRZYBOWSKI, R. (2005) An application of multivalued similarity relations to automated evaluation of grammar tests. In: J. Mischke, ed., *Academy On-Line*, 149-154, in Polish.

OCHELSKA, J., NIEWIADOMSKI, A. and SZCZEPANIAK, P.S. (2001) Linguistic Summaries Applied To Medical Textual Databases. *Journal of Medical Informatics & Technologies* **2**, 125–130.

OCHELSKA, J., SZCZEPANIAK, P.S. and NIEWIADOMSKI, A. (2004) Automatic Summarization of Standarized Textual Databases Interpreted in Terms of Intuitionistic Fuzzy Sets. In: P. Grzegorzewski, M. Krawczak, S. Zadrożny, eds., *Soft Computing: Tools, Techniques and Applications*. The Academic Press EXIT, Warsaw, 204–216.

RASCHIA, G. and MOUADDIB, N. (2002) SAINTETIQ: a fuzzy set-based approach to database summarization. *Fuzzy Sets and Systems* **129**, 137–162.

RASMUSSEN, D. and YAGER, R.R. (1997) A fuzzy SQL summary language for data discovery. In: D. Dubois, H. Prade, R.R. Yager, eds., *Fuzzy Information Engineering: A Guided Tour of Application's*. Wiley, New York, 253–264.

RASMUSSEN, D. and YAGER, R.R. (1999) Finding fuzzy gradual and functional dependencies with SummarySQL. *Fuzzy Sets and Systems* **106**, 131–142.

SENGUPTA, A., PAL, T.K. and CHAKRABORTY, D. (2001) Interpretation of inequality constraints involving interval coefficients and a solution to interval linear programming. *Fuzzy Sets and Systems* **119**, 129–138.

SRIKANTH, R. and AGRAWAL, R. (1996) Mining quantitative association rules in large relational databases. *The 1996 ACM SIGMOD International Conference on Management of Data*, 1–12.

SZMIDT, E. and KACPRZYK, J. (1997) Intuitionistic Fuzzy Linguistic Quantifiers. *Notes on IFS* **3**, 111–122.

TRACZYK, W. (1997) Evaluation of Knowledge Quality. *System Science* **23**.

TURKSEN, I.B. (1986) Interval-valued fuzzy sets based on normal forms. *Fuzzy Sets and Systems* **20**, 191–210.

WU, H. and MENDEL, J.M. (2002) Uncertainty Bounds and Their Use in the Design of Interval Type–2 Fuzzy Logic Systems. *IEEE Transactions on Fuzzy Systems* **10** (5), 622–639.

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

YAGER, R.R., FORD, M. and CANAS, A.J. (1990) An Approach To The Linguistic Summarization of Data. *Proceedings of 3rd International Conference, Information Processing and Management of Uncertainty in Knowledge-based System, Paris, France*, 456–468.

YAGER, R.R., FORD, M. and CANAS, A.J. (1991) On linguistic summaries of data. In: G. Piatetsky-Shapiro, W.J. Frawley, eds., *Knowledge discovery in databases*. AAAI Press, the MIT Press, 347–363.

YAGER, R.R. (1993) Families of OWA operators. *Fuzzy Sets and Systems* **59**, 125–148.

ZADEH, L.A. (1965) Fuzzy Sets. *Information and Control* **8**, 338–353.

ZADEH, L.A. (1975) The concept of a linguistic variable and its application to approximate reasoning (I). *Information Science* **8**, 199–249.

ZADEH, L.A. (1983) A computational approach to fuzzy quantifiers in natural languages. *Computers and Maths with Applications* **9**, 149–184.