JOURNAL OF APPLIED
COMPUTER SCIENCE
Vol. 21 No. 1 (2013), pp. 89-105

Fuzzy Querying with the Use of Interval-Valued Fuzzy Sets

Krzysztof Myszkorowski, Izabela Superson

Technical University of Łodz
Institute of Information Technology
Wolczańska 215 90-924 Łodz
krzysztof.myszkorowski@p.lodz.pl

Abstract. *The paper deals with flexible queries in relational databases. Conditions included in queries are modeled with the use of interval-valued fuzzy sets. Each value returned by a query is associated with a subinterval of [0,1] which expresses a membership degree. The bounds of membership intervals have been determined for different operations of relational algebra and different SQL operators.*
Keywords: *fuzzy sets, databases, relational algebra, query languages.*

1. Introduction

Conventional database systems are designed with the assumption of precision of information collected in them. Every element (attribute values, integrity conditions etc.) must be precisely defined. Imprecise details are either precluded or receive precise interpretation in somehow artificial way. Elements of attribute domains are mutually unrelated i.e. they can be either identical or completely distinct. There is no partial membership of tuples in relations which means that the degree of membership equals 1. Conditions included in queries are required to be formulated precisely. They must be satisfied by the data contained in the result set. Thus the answer is defined uniquely. However, in some cases the precise definition of searching conditions may be difficult to formulate. Questions expressed in the natural language often contain imprecise elements. The lack of preciseness is a property of the natural language. When query conditions are defined imprecisely, the matching becomes a matter of degree. In such cases one has to apply tools for describing imprecise notions [1]. One of them is the fuzzy set theory [2]. Fuzzy notions can be modeled as fuzzy sets in the appropriate universe of discourse. An imprecise query returns a fuzzy relation. Its tuples are associated with values which determine a membership degree or in other words a fulfilment degree of query conditions.

However, in some circumstances a traditional fuzzy set (FS) may appear an insufficient tool. Values of membership degrees are not always unique as it is required by the concept of the type-1 fuzzy set. To capture this kind of uncertainty one can apply its extension, known as an interval-valued fuzzy set (IVFS). The idea was proposed in 1975 [3, 4]. A place of IVFSs among other extensions of the concept of fuzzy sets was shown in [5]. In the presented paper IVFSs are used as a base for the definition of fuzzy conditions in queries.

One of the first works devoted to the use of fuzzy sets in queries was the paper of Tahani [6]. The notion of an imprecise query was defined in the monograph [7]. The author described a grammar of imprecise queries which was used in the system FQUERY for Access [8]. Translation of fuzzy queries to classical forms was described in [9]. The authors applied fuzzy numbers for defining fuzzy conditions. Bosc and Pivert proposed a database language (SQLf) for fuzzy querying [10, 11]. They extended the classical language SQL by fuzzy elements. Another extension of SQL has been proposed in [12, 13]. The authors applied two major approaches concerning fuzzy data representation: possibility-based approach [14] and similarity-based approach [15]. A query language for fuzzy databases was also proposed in [16].

The paper presents modeling of flexible queries by means of interval-valued fuzzy sets. The structure of the article is the following. The next section presents queries with fuzzy selection conditions which are defined by means of interval-valued fuzzy sets and interval-valued fuzzy comparators. Section 3 is devoted to interval-valued fuzzy operators of relational algebra. In section 4 we show SQL constructs extended by fuzzy interval-valued fuzzy conditions. For answers to the presented queries degrees of matching have been determined.

2. Interval-valued fuzzy sets and fuzzy queries

In classical set theory one can define a characteristic function which indicates membership of elements in sets. It is a mapping $\Re \rightarrow \{0, 1\}$, where $\Re$ denotes the universe of discourse. The characteristic function of the set A takes the value 1 if the element e belongs to A and 0 in the opposite case. However, if there are no sharp boundaries of membership the unique qualification of elements is not always obvious.

A fuzzy set is a generalization of an ordinary set. The set $\{0, 1\}$ has been replaced with the interval [0,1]. The definition contains a membership function which is a mapping $\Re \rightarrow [0, 1]$.

Definition 1 *Let $\Re$ be a universe of discourse. A fuzzy set A in $\Re$ is defined as a set of ordered pairs:*

$$A = \{< x, \mu_A(x) >: x \in \Re, \quad \mu_A(x) : \Re \rightarrow [0, 1]\} \ , \tag{1}$$

where $\mu_A(x)$ is a membership function.

Numerical fuzzy values can be expressed by trapezoidal fuzzy numbers with the following membership function:

$$\mu(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } b \leq x \leq c \\ \frac{d-x}{d-c} & \text{if } c < x \leq d \\ 0 & \text{if } x > d \end{cases} . \tag{2}$$

A trapezoidal fuzzy number is often represented by a 4-tuple (a, b, c, d).

The precise determination of the membership grade is not always possible. In such cases it can be expressed by means of FSs in the interval [0,1]. Such extension requires modification of the above definition by replacing the mapping $\Re \rightarrow [0, 1]$ with $\Re \rightarrow F([0, 1])$, where $F([0, 1])$ denotes the set of all FSs that can be defined over the interval [0, 1]. This leads to the definition of type-2 fuzzy sets. A particular case of this notion is the concept of an interval-valued fuzzy set. The elements of the IVFS are assigned with closed subintervals of [0,1]. The idea of the IVFS extends the notion of the ordinary fuzzy set. The assigned intervals approximate the correct value of membership degrees.

Definition 2 *Let $\Re$ be a universe of discourse. An interval-valued fuzzy set A in $\Re$ is defined as:*

$$A = \{< x, \mu_A(x) >: x \in \Re, \quad \mu_A(x) : \Re \rightarrow \text{Int}([0,1])\} \ , \tag{3}$$

where $\mu_A(x) = [\ \mu_{A_L}(x),\ \mu_{A_U}(x)]$ is an interval-valued membership function and Int$([0,1])$ stands for the set of all closed subintervals of $[0,1]$: Int$([0,1]) = \{[a,b] : a, b \in [0,1]\}$.

An interval-valued fuzzy set such that $\exists$ x $\mu_{A_L}(x) = \mu_{A_U}(x) = 1$ is said to be normalized. Membership functions - lower $\mu_{A_L}(x)$ and upper $\mu_{A_U}(x)$ - satisfy the following condition:

$$0 \le \mu_{A_L}(x) \le \mu_{A_U}(x) \le 1 \ . \tag{4}$$

Each value which belongs to $[\mu_{A_L}(x), \mu_{A_U}(x)]$ is fully possible. If $\mu_{A_L}(x) = \mu_{A_U}(x)$ for every x then A is an ordinary fuzzy set. Membership functions $\mu_{A_L}(x)$ and $\mu_{A_U}(x)$ determine two border ordinary fuzzy sets A_L and A_U:

$$A_L = \{< x, \mu_{A_L}(x) >: x \in \Re, \quad \mu_{A_L}(x) : \Re \rightarrow [0,1]\} \ , \tag{5}$$

$$A_U = \{< x, \mu_{A_U}(x) >: x \in \Re, \quad \mu_{A_U}(x) : \Re \rightarrow [0,1]\} \ . \tag{6}$$

The basic characteristics of IVFSs are defined with the use of the border sets. They have been presented in works of Niewiadomski [17, 18].

An interval-valued fuzzy set is said to be an interval-valued fuzzy number, if its border fuzzy sets are fuzzy numbers. Membership functions of an interval-valued fuzzy number are defined by means of the following formulas:

$$\mu_L(x) = \begin{cases} 0 & \text{if } x < a_L \\ \frac{x-a_L}{b_L-a_L} & \text{if } a_L \le x < b_L \\ 1 & \text{if } b_L \le x \le c_L \\ \frac{d_L-x}{d_L-c_L} & \text{if } c_L < x \le d_L \\ 0 & \text{if } x > d_L \end{cases} , \quad \mu_U(x) = \begin{cases} 0 & \text{if } x < a_U \\ \frac{x-a_U}{b_U-a_U} & \text{if } a_U \le x < b_U \\ 1 & \text{if } b_U \le x \le c_U \\ \frac{d_U-x}{d_U-c_U} & \text{if } c_U < x \le d_U \\ 0 & \text{if } x > d_U \end{cases} . \tag{7}$$

Parameters of an interval-valued fuzzy number must satisfy the following conditions:

$$a_L \ge a_U, \qquad b_L \ge b_U, \qquad c_L \le c_U, \qquad d_L \le d_U. \tag{8}$$

A fuzzy condition FC occurring in database queries consists of n elementary components:

$$FC = FC_1\ \Theta_1\ FC_2\ \Theta_2\ ...\ FC_{n-1}\ \Theta_{n-1}\ FC_n\ , \tag{9}$$

where Θ_i denotes an AND/OR fuzzy connective. One can distinguish two kinds of elementary fuzzy search conditions [19]:

$$X \# a \quad \text{or} \tag{10}$$

$$X \# Y\,, \tag{11}$$

where X and Y denote attributes, $\# \in \{=, <, \leq, >, \geq, \neq\}$ and a is a fuzzy value. Let us assume that a is defined by means of an interval-valued fuzzy number (7). Degrees of matching for different comparators are as follows:

$$\mu_{<_L}(x) = \begin{cases} 1 & \text{if } x \leq a_U \\ \frac{b_U - x}{b_U - a_U} & \text{if } a_U < x < b_U \\ 0 & \text{if } x \geq b_U \end{cases}, \quad \mu_{<_U}(x) = \begin{cases} 1 & \text{if } x \leq a_L \\ \frac{b_L - x}{b_L - a_L} & \text{if } a_L < x < b_L \\ 0 & \text{if } x \geq b_L \end{cases}, \tag{12}$$

$$\mu_{>_L}(x) = \begin{cases} 0 & \text{if } x \leq c_U \\ \frac{x - c_U}{d_U - c_U} & \text{if } c_U < x < d_U \\ 1 & \text{if } x \geq d_U \end{cases}, \quad \mu_{>_U}(x) = \begin{cases} 0 & \text{if } x \leq c_L \\ \frac{x - c_L}{d_L - c_L} & \text{if } c_L < x < d_L \\ 1 & \text{if } x \geq d_L \end{cases}, \tag{13}$$

$$\mu_{\leq_L}(x) = \begin{cases} 1 & \text{if } x \leq c_L \\ \frac{d_L - x}{d_L - c_L} & \text{if } c_L < x < d_L \\ 0 & \text{if } x \geq d_L \end{cases}, \quad \mu_{\leq_U}(x) = \begin{cases} 1 & \text{if } x \leq c_U \\ \frac{d_U - x}{d_U - c_U} & \text{if } c_U < x < d_U \\ 0 & \text{if } x \geq d_U \end{cases}, \tag{14}$$

$$\mu_{\geq_L}(x) = \begin{cases} 0 & \text{if } x \leq a_L \\ \frac{b_L - x}{b_L - a_L} & \text{if } a_L < x < b_L \\ 1 & \text{if } x \geq b_L \end{cases}, \quad \mu_{\leq_U}(x) = \begin{cases} 1 & \text{if } x \leq a_U \\ \frac{b_U - x}{b_U - a_U} & \text{if } c_U < x < d_U \\ 0 & \text{if } x \geq b_U \end{cases}, \tag{15}$$

$$\mu_{\neq_L}(x) = \begin{cases} 1 & \text{if } x \leq a_U \\ \frac{b_U - x}{b_U - a_U} & \text{if } a_U < x < b_U \\ 0 & \text{if } b_U \leq x \leq c_U \\ \frac{x - c_U}{d_U - c_U} & \text{if } c_U < x < d_U \\ 1 & \text{if } x \geq d_U \end{cases}, \quad \mu_{\neq_U}(x) = \begin{cases} 1 & \text{if } x \leq a_L \\ \frac{b_L - x}{b_L - a_L} & \text{if } a_L < x < b_L \\ 0 & \text{if } b_L \leq x \leq c_L \\ \frac{x - c_L}{d_L - c_L} & \text{if } c_L < x < d_L \\ 1 & \text{if } x \geq d_L \end{cases}. \tag{16}$$

The degree of matching for "=" is given by (7).

Fuzziness in selection condition can also occur if an attribute value is compared with a precise value by means of a fuzzy comparator such as "approximately equal", "much larger than" or "much smaller than" [19]. A fuzzy comparator $\#_f$ can be represented by a piecewise linear membership function defined for a difference of the compared values. Let a_L, a_U, b_L, b_U, c_L, c_U, d_L, d_U be positive real numbers. Membership functions $\mu_{\#_{f_L}}$ and $\mu_{\#_{f_U}}$ of interval-valued fuzzy comparators $=_f,\ \neq_f,\ <_f,\ \leq_f,\ >_f,\ \geq_f$ are as follows:

$$\mu_{=_{f_L}}(x)=\begin{cases} 0 & \text{if } x<-a_L \\ \frac{x+a_L}{a_L-b_L} & \text{if } -a_L\leq x<-b_L \\ 1 & \text{if } -b_L\leq x\leq c_L, \\ \frac{d_L-x}{d_L-c_L} & \text{if } c_L<x\leq d_L \\ 0 & \text{if } x>d_L \end{cases}\quad \mu_{=_{f_U}}(x)=\begin{cases} 0 & \text{if } x<-a_U \\ \frac{x+a_U}{a_U-b_U} & \text{if } -a_U\leq x<-b_U \\ 1 & \text{if } -b_U\leq x\leq c_U, \\ \frac{d_U-x}{d_U-c_U} & \text{if } c_U<x\leq d_U \\ 0 & \text{if } x>d_U \end{cases} \tag{17}$$

where $a_L\leq a_U, b_L\leq b_U, c_L\leq c_U, d_L\leq d_U$,

$$\mu_{\neq_{f_L}}(x)=\begin{cases} 1 & \text{if } x<-a_L \\ \frac{-b_L-x}{a_L-b_L} & \text{if } -a_L\leq x<-b_L \\ 0 & \text{if } -b_L\leq x\leq c_L, \\ \frac{x-c_L}{d_L-c_L} & \text{if } c_L<x\leq d_L \\ 1 & \text{if } x>d_L \end{cases}\quad \mu_{\neq_{f_U}}(x)=\begin{cases} 1 & \text{if } x<-a_U \\ \frac{-b_U-x}{a_U-b_U} & \text{if } -a_U\leq x<-b_U \\ 0 & \text{if } -b_U\leq x\leq c_U, \\ \frac{x-c_U}{d_U-c_U} & \text{if } c_U<x\leq d_U \\ 1 & \text{if } x>d_U \end{cases} \tag{18}$$

where $a_L\geq a_U, b_L\geq b_U, c_L\geq c_U, d_L\geq d_U$,

$$\mu_{\leq_{f_L}}(x)=\begin{cases} 1 & \text{if } x<a_L \\ \frac{b_L-x}{b_L-a_L} & \text{if } a_L\leq x\leq b_L, \\ 0 & \text{if } x>b_L \end{cases}\quad \mu_{\leq_{f_U}}(x)=\begin{cases} 1 & \text{if } x<a_U \\ \frac{b_U-x}{b_U-a_U} & \text{if } a_U\leq x\leq b_U, \\ 0 & \text{if } x>b_U \end{cases} \tag{19}$$

where $a_L\leq a_U, b_L\leq b_U$,

$$\mu_{<_{f_L}}(x)=\begin{cases} 1 & \text{if } x<-a_L \\ \frac{-x}{a_L} & \text{if } -a_L\leq x\leq 0, \\ 0 & \text{if } x>0 \end{cases}\quad \mu_{<_{f_U}}(x)=\begin{cases} 1 & \text{if } x<-a_U \\ \frac{-x}{a_U} & \text{if } -a_U\leq x<0, \\ 0 & \text{if } x>0 \end{cases} \tag{20}$$

where $a_L\geq a_U$,

$$\mu_{\geq_{f_L}}(x)=\begin{cases} 0 & \text{if } x<-a_L \\ \frac{x+a_L}{a_L-b_L} & \text{if } -a_L\leq x\leq -b_L, \\ 1 & \text{if } x>-b_L \end{cases}\quad \mu_{\geq_{f_U}}(x)=\begin{cases} 0 & \text{if } x<-a_U \\ \frac{x+a_L}{a_U-b_U} & \text{if } -a_U\leq x\leq -b_U, \\ 1 & \text{if } x>-b_U \end{cases} \tag{21}$$

where $a_L\leq a_U, b_L\leq b_U$,

$$\mu_{>_{f_L}}(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{a_L} & \text{if } 0 \le x \le a_L, \\ 1 & \text{if } x > a_L \end{cases} \quad \mu_{\ge_{f_U}}(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{a_U} & \text{if } 0 \le x \le a_U, \\ 1 & \text{if } x > b_U \end{cases} \tag{22}$$

where $a_L \ge a_U$.

Interval-valued fuzzy connectives are represented by the following operators:

$$\text{AND:} \quad \wedge (x, y)_L = \min(x_L, y_L), \quad \wedge(x, y)_U = \min(x_U, y_U)\,, \tag{23}$$

$$\text{OR:} \quad \vee (x, y)_L = \max(x_L, y_L), \quad \vee(x, y)_U = \max(x_U, y_U)\,, \tag{24}$$

$$\text{NOT:} \quad \neg(x)_L = 1 - x_U, \quad \neg(x)_U = 1 - x_L\,. \tag{25}$$

Based on the formulas given in this section one can determine degrees of matching for corresponding query conditions.

3. Algebra of fuzzy relations with interval membership of tuples

A fuzzy relation $R(X_1, X_2, \ldots, X_n)$ is a fuzzy set defined on the Cartesian product of attribute domains. In our considerations we will assume that the membership degree of a tuple r is expressed by subinterval $[\mu_{R_L}(r), \mu_{R_U}(r)] \subseteq [0, 1]$. Classical operations of relational algebra must be extended by defining membership degrees for final relations. Let R or S be fuzzy relations with membership intervals μ_R and μ_S, respectively.

Let us assume that R and S are relations with the same schemes. Union of R and S is a fuzzy relation $R \cup S$ containing tuples which belong to at least one of them. The membership degree of a tuple t in $R \cup S$ equals $\mu_{R\cup S}(t) = [\mu_{R\cup S_L}(t), \mu_{R\cup S_U}(t)]$, where $\mu_{R\cup S_L}(t) = \max\,(\mu_{R_L}(t), \mu_{S_L}(t))$ and $\mu_{R\cup S_U}(t) = \max\,(\mu_{R_U}(t), \mu_{S_U}(t))$. Intersection of R and S is a fuzzy relation $R \cup S$ containing tuples which belong to both of them. The membership degree of a tuple t in $R \cap S$ equals $\mu_{R\cap S}(t) = [\mu_{R\cap S_L}(t), \mu_{R\cap S_U}(t)]$, where $\mu_{R\cap S_L}(t) = \min\,(\mu_{R_L}(t), \mu_{S_L}(t))$ and $\mu_{R\cap S_U}(t) = \min\,(\mu_{R_U}(t), \mu_{S_U}(t))$. Difference: $R - S$ is a fuzzy relation containing tuples which belong to R and do not belong to S. The membership degree of a tuple t in $R - S$ equals $\mu_{R-S}(t) = [\mu_{R-S_L}(t), \mu_{R-S_U}(t)]$, where $\mu_{R-S_L}(t) = \min\,(\mu_{R_L}(t), 1 - \mu_{S_U}(t))$ and $\mu_{R-S_U}(t) = \min\,(\mu_{R_U}(t), 1 - \mu_{S_L}(t))$. All these three operations are showed in Fig. 1.

R

A	B	μ
a_1	b_1	[0.5, 0.7]
a_2	b_1	[1, 1]
a_2	b_2	[0.7, 0.9]
a_3	b_2	[0.2, 0.4]

S

A	B	μ
a_1	b_1	[0.1, 0,4]
a_1	b_2	[0.2, 0.6]
a_2	b_2	[0.5, 0.8]
a_2	b_3	[0.7, 0.9]

$R \cup S$

A	B	μ
a_1	b_1	[0.5, 0.7]
a_1	b_2	[0.2, 0.6]
a_2	b_1	[1, 1]
a_2	b_2	[0.7, 0.9]
a_2	b_3	[0.7, 0.9]
a_3	b_2	[0.2, 0.4]

$R \cap S$

A	B	μ
a_1	b_1	[0.1, 0,4]
a_2	b_2	[0.5, 0.8]

$R - S$

A	B	μ
a_1	b_1	[0.5, 0,7]
a_2	b_1	[1, 1]
a_2	b_2	[0.2, 0,5]
a_3	b_2	[0.2, 0,4]

Figure 1. Union, multiplication and difference of fuzzy relations

Fuzzy selection, denoted as $\delta_{FC}(R)$, is a fuzzy relation containing tuples of R which satisfy a fuzzy condition FC. Let $\mu_{FC}(t) = [\mu_{FC_L}(t), \mu_{FC_U}(t)]$ be an interval expressing the fulfilment degree of FC. Thus the membership degree of a tuple t in $\delta_{FC}(R)$ equals $\mu_{\delta_{FC}(R)}(t) = [\mu_{\delta_{FC}(R)_L}(t), \mu_{\delta_{FC}(R)_U}(t)]$, where $\mu_{\delta_{FC}(R)_L}(t) = \min\ (\mu_{R_L}(t), \mu_{FC_L}(t))$ and $\mu_{\delta_{FC}(R)_U}(t) = \min\ (\mu_{R_U}(t), \mu_{FC_U}(t))$.

Projection of relation R on attribute set X, denoted as $\Pi_X(R)$, is a fuzzy relation on X with the membership degree $\mu_{\Pi_X(R)}(t) = [\mu_{\Pi_X(R)_L}(t), \mu_{\Pi_X(R)_U}(t)]$, where $\mu_{\Pi_X(R)_L}(t) = \sup_{r(X)=t(X)} \mu_{R_L}(r)$ and $\mu_{\Pi_X(R)_U}(t) = \sup_{r(X)=t(X)} \mu_{R_U}(r)$. The operation removes duplicates. For instance relation $\Pi_B(R)$ (Fig. 1) contains two tuples: (b_1, [1,1]) and (b_2, [0.7,0.9]).

Natural join $R(X, Y) * S(Y, Z)$ of fuzzy relations R and S is a fuzzy relation of the scheme $SCH = \{X, Y, Z\}$ (Fig. 2). The membership degree of a tuple t in $R * S$ equals $\mu_{R*S}(t) = [\mu_{R*S_L}(t), \mu_{R*S_U}(t)]$, where $\mu_{R*S_L}(t) = \min\ (\mu_{R_L}(r),\ \mu_{S_L}(r))$ and $\mu_{R*S_U}(t) = \min\ (\mu_{R_U}(r),\ \mu_{S_U}(r))$, where r and s are tuples of relations R and S, respectively. A quotient of fuzzy relations $R(X, Y)$ and $S(Y)$, denoted as $R \div S$, contains values of attribute X which satisfy the requirements defined by S in non-zero degree (Fig. 3). In order to determine tuples of $R \div S$ one should apply interval-valued fuzzy implication which is an extension of the fuzzy impli-

R

A	B	μ
a_1	b_1	[0.3, 0.5]
a_2	b_1	[0.6, 0.9]
a_2	b_2	[0.4, 0.7]
a_3	b_3	[1, 1]

S

B	C	μ
b_1	c_1	[0.1, 0,4]
b_1	c_2	[0.5, 0.8]
b_2	c_2	[0.5, 0.8]
b_3	c_3	[0.5, 0.8]

$R * S$

A	B	C	μ
a_1	b_1	c_1	[0.1, 0,4]
a_1	b_1	c_2	[0.3, 0.5]
a_2	b_1	c_1	[0.1, 0,4]
a_2	b_1	c_2	[0.5, 0,8]
a_2	b_2	c_2	[0.4, 0,7]
a_3	b_3	c_3	[0.5, 0,8]

Figure 2. Natural join of fuzzy relations

cation. An interval-valued fuzzy implicator is a mapping Int([0, 1]) × Int([0, 1]) → Int([0, 1]). Its arguments are subintervals of [0, 1]. In further considerations it will be applied the following implicator [20]:

$$I_V([a,b],[c,d]) = \begin{cases} [c,d] & \text{if } a > c \text{ and } b > d \\ [c,1] & \text{if } a > c \text{ and } b \leq d \\ [1,1] & \text{if } a \leq c \text{ and } b \leq d \\ [d,d] & \text{if } a \leq c \text{ and } b > d \end{cases} \tag{26}$$

which is the extension of the Gödel implicator:

$$I_G(a,b) = 1 \quad \text{if } a \leq b \quad \text{and} \quad I_G(a,b) = b \quad \text{if } a > b \ . \tag{27}$$

Let us consider tuples $r \in R$ of the form (a, b), where a is a fixed value and the set of b-values contains all values existing in S. The membership degree of a in $R \div S$ equals: $\mu_{R\div S}(a) = [\mu_{R\div S_L}(a), \mu_{R\div S_U}(a)]$, where $\mu_{R\div S_L}(a) = \min_b I(\mu_S(b), \mu_R(a,b))_L$ and $\mu_{R\div S_U}(a) = \min_b I(\mu_S(b), \mu_R(a,b))_U$, $\mu_R(a,b)$ and $\mu_S(b)$ denote membership intervals of R and S, respectively and I is the interval-valued fuzzy implicator defined by (26).

Example 3 *Let us consider several queries on the database consisting of the following relations:*

PROJECTS($\underline{P-id}$, *PNAME*, *BUDGET*, *TOWN*),
EMPLOYEES($\underline{E-id}$, *ENAME*, *PROFESSION*, *AGE*),
CONTRACTS($\underline{P-id}$, $\underline{E-id}$, *MONEY*).

R

A	B	μ
a_1	b_1	[0.4, 0.6]
a_1	b_2	[0.3, 0.5]
a_1	b_3	[0.5, 0.8]
a_2	b_1	[0,4, 0.7]
a_2	b_2	[0.6, 0.9]
a_3	b_1	[0.5, 0.7]
a_3	b_2	[1, 1]
a_3	b_3	[0.8, 0.9]

S

B	μ
b_1	[0.3, 0,5]
b_2	[0.5, 0.8]
b_3	[0.7, 0.9]

$R \div S$

A	μ
a_1	[0.3, 0,5]
a_3	[1, 1]

Figure 3. Division of fuzzy relations

Let us assume that linguistic terms "around 5000", "Young" and "High" are defined by suitable interval-valued fuzzy sets.

1. Get numbers of employees who have contracts of around 5000 for participation in the project 'P1'.

$R1 = \Pi_{E-id}(\delta_F(CONTRACTS))$ *, where*

$F = P - id\ =\ 'P1'\ AND\ AMOUNT\ =\ 'Around\ 5000'$.

Degrees of matching are as follows:

$\mu_{R1_L}(t[E - id]) =$

$\min\ (\mu_{CONTRACTS_L}(r[E - Id,' P1', AMOUNT]), \mu_{Around\ 5000_L}(r[AMOUNT]))$,

$\mu_{R1_U}(t[E - id]) =$

$\min\ (\mu_{CONTRACTS_U}(r[E - Id,' P1', AMOUNT]), \mu_{Around\ 5000_U}(r[AMOUNT]))$,

where r is a tuple of $CONTRACTS$ *such that* $r[P - id] = 'P1'$. *In case of full belongingness of tuples to relation* $CONTRACTS$, $\forall_r\ \mu_{CONTRACTS}(r) = [1, 1]$, *the matching degree depends on the value of the attribute* $AMOUNT$. *Operator* sup *is not applied because projection is performed on the key attribute of* $CONTRACTS$. *There can exist only one tuple* $(E - Id, 'P1')$ *for each value of* $E - Id$.

2. Get professions of young employees who have contracts of around 5000 for participation in the project 'P1'.

$R2 = \Pi_{PROFESSION}(\delta_{AGE\ =\ 'Young'}(EMPLOYEES) * \delta_F(CONTRACTS))$,

Degrees of matching are as follows:

$\mu_{R2_L}(t[PROFESSION]) =$

$\sup_{t[PROFESSION]=s[PROFESSION]}\min(\mu_{EMPLOYEES_L}(s[E-Id, ENAME,$
$PROFESSION, AGE]), \mu_{Young_L}(s[AGE]), \mu_{CONTRACTS_L}(r[E-Id,'P1',$
$AMOUNT]), \mu_{Around\ 5000_L}(r[AMOUNT]))$,
$\mu_{R2_U}(t[PROFESSION]) =$
$\sup_{t[PROFESSION]=s[PROFESSION]}\min(\mu_{EMPLOYEES_U}(s[E-Id, ENAME,$
$PROFESSION, AGE]), \mu_{Young_U}(s[AGE]), \mu_{CONTRACTS_U}(r[E-Id,'P1',$
$AMOUNT]), \mu_{Around\ 5000_U}(r[AMOUNT]))$.

3. Get numbers of young employees who do not have any contract of around 5000.
$R3 = R3' - R3''$, *where* $R3' = \Pi_{E-Id}(\delta_{AGE\ =\ 'Young'}(EMPLOYEES)$ *and*
$R3'' = \Pi_{E-Id}(\delta_{AMOUNT\ =\ 'Around\ 5000'}(CONTRACTS))$.
Degrees of matching are as follows:
$\mu_{R3_L}(t[E-Id]) =$
$\min(\mu_{EMPLOYEES_L}(s[E-Id, ENAME, PROFESSION, AGE]), \mu_{Young_L}(s[AGE]),$
$(1-\sup_{t[E-ID]=r[E-Id]}\min(\mu_{CONTRACTS_U}(r),\ \mu_{Around\ 5000_U}(r[AMOUNT]))$,
$\mu_{R3_U}(t[E-Id]) =$
$\min(\mu_{EMPLOYEES_U}(s[E-Id, ENAME, PROFESSION, AGE],\ \mu_{Young_U}(s[AGE),$
$(1-\sup_{t[E-ID]=r[E-Id]}\min(\mu_{CONTRACTS_L}(r),\ \mu_{Around\ 5000_L}(r[AMOUNT]))$.

4. Get numbers of employees who take part in all projects with high budgets and amount of each their contract concerning such projects equals around 5000.
$R4 = R4' \div R4''$, *where*
$R4' = \Pi_{E-Id,P-Id}(\delta_{AMOUNT\ =\ 'Around\ 5000'}(CONTRACTS)*$
$\delta_{BUDGET='High'}(PROJECTS))$
$R4'' = \Pi_{P-Id}(\delta_{BUDGET='High'}(PROJECTS))$
Degrees of matching are as follows:
$\mu_{R4_L}(t[E-Id])_L = \min_{tp} I_G(\mu_{R4''}(tp),\ \mu_{R4'}(tk))_L,$
$\mu_{R4_L}(t[E-Id])_U = \min_{tp} I_G(\mu_{R4''}(tp),\ \mu_{R4'}(tk))_U,$
where tp and tk are tuples of $R4'$ *and* $R4''$, *respectively and*
$\mu_{R4'}(tp[E-Id, P-Id])_L =$
$\min(\mu_{CONTRACTS_L}(r[E-Id, P-Id, AMOUNT]),\ \mu_{Around\ 5000_L}(r[AMOUNT]))$,
$\mu_{R4'}(tp[E-Id, P-Id])_U =$
$\min(\mu_{CONTRACTS_U}(r[E-Id, P-Id, AMOUNT]),\ \mu_{Around\ 5000_U}(r[AMOUNT]))$,
$\mu_{R4''}(tk[P-Id])_L =$
$\min(\mu_{PROJECTS_L}(p[P-Id, PNAME, BUDGET, TOWN]),$
$\mu_{High_L}(p[BUDGET]))$,
$\mu_{R4''}(tk[P-Id])_U =$

$\min(\mu_{PROJECTS_U}(p[P-Id, PNAME, BUDGET, TOWN]),$
$\mu_{High_U}(p[BUDGET]))$.

4. Imprecise queries in extended SQL

The main operation of SQL is a block SELECT-FROM-WHERE:

SELECT $A1, A2, ..., An$
FROM $R1, R2, ..., Rm$
WHERE W;

where $A1, A2, ..., An$ denote attributes, $R1, R2, ..., Rm$ denote relations and W is a selection condition. It can be treated as a composition of algebraic selection and projection without elimination of duplicates.

In imprecise queries W is a fuzzy condition. Let us denote it by Wf. Each element $f = (a_1, a_2, ..., a_n)$ of the resulting set F is associated with a membership interval $\mu_F(f) = [\mu_{F_L}(f),\ \mu_{F_U}(f)]$. The bounds μ_{F_L} and μ_{F_U} are as follows:

$$\mu_{F_L}(f) = \min(\mu_{R_L}(r),\ \mu_{Wf_L}(r))\ ,$$
$$\mu_{F_U}(f) = \min(\mu_{R_U}(r),\ \mu_{Wf_U}(r))\ ,$$
$$R = R1 \times R2 \times ... \times Rm, \quad r[A_i] = f.a_i, i = 1, 2, ..., n\ . \tag{28}$$

If duplicates are eliminated (operator DISTINCT) formula (28) takes the following form:

$$\mu_{F_L}(f) = \sup_{f.a_i = r[A_i],\ i=1,2,...n} \min(\mu_{R_L}(r),\ \mu_{Wf_L}(r))\ ,$$
$$\mu_{F_U}(f) = \sup_{f.a_i = r[A_i],\ i=1,2,...n} \min(\mu_{R_U}(r),\ \mu_{Wf_U}(r))\ . \tag{29}$$

Formulas (28) and (29) concern the general form of the block SELECT-FROM-WHERE. They constitute a base for determining membership intervals for different kinds of queries.

For the most simple query (with the use of only one relation) "SELECT A FROM R WHERE Wf" the bounds of the membership interval are as follows:

$$\mu_{F_L}(a) = \min(\mu_{R_L}(r),\ \mu_{Wf_L}(r))\ ,$$
$$\mu_{F_U}(a) = \min(\mu_{R_U}(r),\ \mu_{Wf_U}(r))\ ,$$
$$r[A] = a\ . \tag{30}$$

The join operation of relations $R(A, B)$ and $S(C, D)$ can be expressed by the following query: "SELECT A, C FROM R, S WHERE Wf1 AND Wf2 AND B = D", where $Wf1$ and $Wf2$ are fuzzy conditions imposed on R and S, respectively. Based on (28) we have:

$$\mu_{F_L}(a, c) = \min(\mu_{R_L}(r),\ \mu_{Wf1_L}(r),\ \mu_{S_L}(s),\ \mu_{Wf2_L}(s)),$$
$$\mu_{F_U}(a, c) = \min(\mu_{R_U}(r),\ \mu_{Wf1_U}(r),\ \mu_{S_U}(s),\ \mu_{Wf2_U}(s)),$$
$$r[B] = s[D], \quad r[A] = a, \quad s[C] = c. \tag{31}$$

The above query contains a sharp condition of joining: $B = D$. In general this condition may be fuzzy and so it can be expressed by means of a fuzzy comparison operator e.g. "more or less equal", "a little greater" etc. The degree of its fulfilment should also be taken into account in (31).

Searching conditions can be expressed by means of a subquery defining a fuzzy set of values. Let us denote it by SQf. In order to check whether an attribute value a belongs to SQf one should apply the operator "IN"'. The fulfilment degree of the following condition: "A IN SELECT B FROM R WHERE Wf" is an interval $\mu_{IN}(f) = [\mu_{IN_L}(f),\ \mu_{IN_U}(f)]$ with the following bounds:

$$\mu_{IN_L}(a) = \sup_{b=a}\min(\mu_{R_L}(r),\ \mu_{Wf_L}(r)) \ ,$$
$$\mu_{IN_U}(a) = \sup_{b=a}\min(\mu_{R_U}(r),\ \mu_{Wf_U}(r)) \ ,$$
$$r[B] = b \ . \tag{32}$$

A subquery may also define a set the existing of which is to be checked. The relevant condition is created by means of the operator EXISTS: "EXISTS (SELECT A FROM R WHERE Wf)". Let us denote the defined set by Z. The fulfilment degree is an interval $\mu_{EXISTS}(Z) = [\mu_{EXISTS_L}(Z),\ \mu_{EXISTS_U}(Z)]$ with the following bounds:

$$\mu_{EXISTS_L}(Z) = \sup_a\min(\mu_{R_L}(r),\ \mu_{Wf_L}(r)) \ ,$$
$$\mu_{EXISTS_U}(Z) = \sup_a\min(\mu_{R_U}(r),\ \mu_{Wf_U}(r)) \ ,$$
$$r[A] = a \ . \tag{33}$$

Example 4 *Let us consider queries from the previous example expressed by SQL statements:*

1. Get numbers of employees who have contracts of around 5000 for participation in the project 'P1'.

SELECT $E-Id$ FROM $CONTRACTS$
WHERE $P-Id='P1'$ AND $AMOUNT='Around\ 5000'$
Degrees of matching are as follows:
$\mu_L(eid)=$
$\min_{r[E-Id]=eid\ \wedge\ r[P-Id]='P1'}(\mu_{CONTRACTS_L}(r), \mu_{Around\ 5000_L}(r[AMOUNT]))$
$\mu_U(eid)=$
$\min_{r[E-Id]=eid\ \wedge\ r[P-Id]='P1'}(\mu_{CONTRACTS_U}(r), \mu_{Around\ 5000_U}(r[AMOUNT]))$

2. Get professions of young employees who have contracts of around 5000 for participation in the project 'P1'.
SELECT $PROFESSION$ FROM $EMPLOYEES$
WHERE $AGE='Young'$ AND $E-Id$ IN
(SELECT $E-Id$ FROM $CONTRACTS$
WHERE $P-Id='P1'$ AND $AMOUNT='Around\ 5000'$)
Degrees of matching are as follows:
$\mu_L(prof)=\sup_{s[PROFESSION]=prof}\min(\mu_{EMPLOYEES_L}(s),\ \mu_{Young_L}(s[AGE]),$
$\sup_{r[E-Id]=s[E-Id]\ \wedge\ r[P-Id]='P1'}\min\ (\mu_{CONTRACTS_L}(r), \mu_{Around\ 5000_L}(r[AMOUNT])),$
$\mu_U(prof)=\sup_{s[PROFESSION]=prof}\min(\mu_{EMPLOYEES_U}(s),\ \mu_{Young_U}(s[AGE]),$
$\sup_{r[E-Id]=s[E-Id]\ \wedge\ r[P-Id]='P1'}\min(\mu_{CONTRACTS_U}(r), \mu_{Around\ 5000_U}(r[AMOUNT])).$

3. Get numbers of young employees who do not have any contract of around 5000.
SELECT $E-Id$ FROM $EMPLOYEES$
WHERE $AGE='Young'$ AND NOT EXISTS
(SELECT $*$ FROM $CONTRACTS$
WHERE $E-Id=EMPLOYEE.E-ID$ AND $AMOUNT='Around\ 5000'$)
Degrees of matching are as follows:
$\mu_L(eid)=\min_{r[E-Id]=eid}(\mu_{EMPLOYEES_L}(s),\ \mu_{Young_L}(s[AGE]),$
$1-\sup_{r[E-Id]=s[E-id]}\min\ (\mu_{CONTRACTS_U}(r), \mu_{Around\ 5000_U}(r[AMOUNT]))),$
$\mu_U(eid)=\min_{r[E-Id]=eid}(\mu_{EMPLOYEES_U}(s),\ \mu_{Young_U}(s[AGE]),$
$1-\sup_{r[E-Id]=s[E-id]}\min\ (\mu_{CONTRACTS_L}(r), \mu_{Around\ 5000_L}(r[AMOUNT]))).$

4. Get numbers of employees who take part in all projects with high budgets and amount of each their contract concerning such projects equals around 5000.
SELECT DISTINCT $E-ID$ FROM $CONTRACTS\ C$
WHERE $AMOUNT='Around\ 5000'$ AND NOT EXISTS
(SELECT $P-Id$ FROM $PROJECTS\ P$
WHERE $BUDGET='High'$ AND $P-Id$ NOT IN

(SELECT $P - Id$ FROM $CONTRACTS$
WHERE $C.E - Id = E - Id$ AND $AMOUNT$ $='$ $Around$ 5000))

Degrees of matching are as follows:

$\mu_L(eid) = \sup_{r[E-Id]=eid} \min(\mu_{CONTRACTS_L}(r),\ \mu_{Around\ 5000_L}(r[AMOUNT]),$
$1 - \sup_s \min\ (\mu_{PROJECTS_U}(s), \mu_{High_U}(s[BUDGET]),\ 1 -$
$\sup_{r'[Eid]=eid\ \wedge\ r'[Pid]=s[Pid]} \min(\mu_{CONTRACTS_L}(r'),\ \mu_{Around\ 5000_L}(r'[AMOUNT])))),$
$\mu_U(eid) = \sup_{r[E-Id]=eid} \min(\mu_{CONTRACTS_U}(r),\ \mu_{Around\ 5000_U}(r[AMOUNT]),$
$1 - \sup_s \min\ (\mu_{PROJECTS_L}(s), \mu_{High_L}(s[BUDGET]),\ 1 -$
$\sup_{r'[Eid]=eid\ \wedge\ r'[Pid]=s[Pid]} \min(\mu_{CONTRACTS_U}(r'),\ \mu_{Around\ 5000_U}(r'[AMOUNT])))).$

5. Conclusions

Ability to extract data in the language close to natural increases the possibilities of database management system. Flexible queries were often modeled with the use of type-1 fuzzy sets. However, in many cases users cannot unambiguously determine not only sharp boundaries of relevant data but also precise values of membership degrees. Therefore we propose applying of interval-valued fuzzy sets in which the correct value of the membership grade is approximated by the membership interval. In the paper we studied extended algebraic operators and SQL constructs. The relevant formulas for membership intervals have been derived. Fuzzy queries may also contain conditions dealing with sets of tuples. In such cases one should apply fuzzy aggregation functions, fuzzy quantifiers and linguistic summarization of data [7, 18]. These topics will be a subject of future works.

References

[1] Motro, A., *Imprecision and Uncertainty in Database Systems*, In: Studies in Fuzziness: Fuzziness in Database Management Systems, edited by P. Bosc and J. Kacprzyk, Physica Verlag, 1995, pp. 3–22.

[2] Zadeh, L., *Fuzzy Sets*, Information and Control, Vol. 8, No. 3, 1965, pp. 338–353.

[3] Sambuc, R., *Fonctions Φ-floues. Application à l'aide au diagnostic en pathologie thyroidienne*, Ph.D. thesis, Univ. Marseille, France, 1975, (in French).

[4] Zadeh, L., *The concept of a linguistic variable and its application to approximate reasoning*, Information Sciences, Vol. 8, No. 3, 1975, pp. 199–249.

[5] Deschrijver, G. and Kerre, E., *On the relationship between some extensions of fuzzy set theory*, Fuzzy Sets and Systems, Vol. 133, No. 2, 2003, pp. 227–235.

[6] Tahani, V., *A conceptual framework for fuzzy query processing: a step toward very intelligent database systems*, Information Processing and Management, Vol. 13, No. 5, 1977, pp. 289–303.

[7] Zadrożny, S., *Zapytania nieprecyzyjne i lingwistyczne podsumowania baz danych*, Akademicka Oficyna Wydawnicza EXIT, 2006, in Polish.

[8] Kacprzyk, J., Zadrożny, S., and Ziółkowski, A., *FQUERYIII+: a "human consistent" database querying system based on fuzzy logic with linguistic quantifiers*, Information Systems, Vol. 6, No. 2, 1989, pp. 443–453.

[9] Podstawny, J. and Małysiak, B., *Metodyka przetwarzania rozmytych zapytań SQL na klasyczne zapytania*, In: Bazy danych. Nowe technologie. Architektura, metody formalne i zaawansowana analiza danych, edited by S. Kozielski, B. Małysiak, K. P., and D. Mrozek, Wydawnictwa Komunikacji i Łączności, 2007, pp. 167–179, in Polish.

[10] Bosc, A. and Pivert, O., *Fuzzy queries and relational databases*, In: SAC '94 Proceedings of the 1994 ACM Symposium on Applied Computing, 1995, pp. 170–174.

[11] Bosc, A. and Pivert, O., *SQLf: A Relational Database Language for Fuzzy Querying*, IEEE Transactions on Fuzzy Systems, Vol. 3, No. 1, 1995, pp. 1–17.

[12] Dembczyński, K., Hapke, M., and Przybył, D., *SQLfj a fuzzy querying language and application*, Tech. Rep. RA - 010/04, Poznań University of Technology, 2004.

[13] Dembczyński, K., Hapke, M., and Przybył, D., *Fuzzy querying with fuzzy projection*, In: Advances in Web Intelligence, Lecture Notes in Artificial Intelligence LNAI 3528, edited by P. Szczepaniak, J. Kacprzyk, and A. Niewiadomski, Springer Verlag, 2005, pp. 121–125.

[14] Prade, H. and Testemale, C., *Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries*, Information Sciences, Vol. 34, No. 2, 1984, pp. 115–143.

[15] Buckles, B. P. and Petry, F., *A Fuzzy Representation of Data for Relational Databases*, Fuzzy Sets and Systems, Vol. 7, No. 3, 1982, pp. 213–226.

[16] Galindo, J., Urutia, A., and Piattini, M., *Fuzzy Databases. Modeling, Design and Implementation*, Idea Group Publishing, 2005.

[17] Niewiadomski, A., *Interval-Valued and Interval Type-2 Fuzzy Sets: A Subjective Comparison*, In: Proceedings of FUZZ-IEEE'07VII ITKR's Session, London, 2007, pp. 1198–1203.

[18] Niewiadomski, A., *Methods for the linguistic summarization of data: applications of fuzzy sets and their extensions*, Akademicka Oficyna Wydawnicza EXIT, 2008.

[19] De Tre, G., Zadrożny, S., and Bronselaer, J., *Handling Bipolarity in Elementary Queries to Possibilistic Databases*, IEEE Transactions on Fuzzy Systems, Vol. 18, No. 3, 2010, pp. 599–611.

[20] Alcade, C., Burusco, A., and Fuentes-Gonzales, R., *A constructive method for the definition of interval-valued fuzzy implication operators*, Fuzzy Sets and Systems, Vol. 153, No. 2, 2005, pp. 211–227.