Flexible Querying
via Intuitionistic Fuzzy Sets

Przemysław Grzegorzewski
Systems Research Institute,
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
pgrzeg@ibspan.waw.pl

Edyta Mrówka
Systems Research Institute,
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
mrowka@ibspan.waw.pl

Abstract

The traditional query languages used in database management systems require precise and unambiguous queries only. Fuzzy querying were introduced to relax this rigidity and allow the user for a more natural information retrieval. In this paper we suggest how to enrich fuzzy querying by the use of intuitionistic fuzzy values.

Keywords: Intuitionistic fuzzy sets, Fuzzy querying, Querying databases.

1   Introduction

The traditional query languages, used in the database management systems, require a precise and unambiguous specification of a query. It seems to be a serious limitation since a typical user often formulates his requirements in a natural language using imprecise expressions and vague terms. For this reason several approaches have been proposed to relax the rigidity of the conventional queries and make possible to use queries that allow for a more intelligent and human consistent information retrieval (see, e.g. [4]).

The FQUERY for Access by Kacprzyk and Zadrożny [9], [10], [11] is an example of a computer program that enables to create different kinds of fuzzy queries. Using such fuzzy queries we deal no longer with binary outputs – whether a record fulfil given requirement or not – but we get an information on the degree the record complies with the requirement.

However, fuzzy logic sometimes is not sufficient for modelling statements expressed in natural language.

It may happen that we cannot accept one-to-one correspondence between a fuzzy set representing given term and its complement, representing negation of that term. For example, if a fuzzy set is used for modelling a term "warm" then its logical negation does not necessarily identifies with its linguistic negation "cold". Moreover, sometimes it seems to be more natural to describe a fuzzy condition not only by its membership function. It is due to the fact, that in some situations it is easier to describe our negative feelings than the positive attitude. Even more, quite often one can easily specify objects or alternatives he dislikes, but simultaneously, he cannot specify clearly what he really wants. For example, it may happen that a person asked about his favorite district in Warsaw cannot definitely choose whether it is Ochota, Mokotów or Żoliborz, but he feels sure that he hates Praga.

It seems that querying which enables the user to specify both membership and nonmembership functions, such that they not necessarily sum up to 1, would be more human-consistent and human-friendly. A formal tool for such querying is provided by the theory of intuitionistic fuzzy sets proposed by Atanassov [1], [2].

In the present paper we suggest how to extract information from a conventional crisp database using queries that utilize not only membership functions but which admit nonmembership functions as well. This problem was discussed for the first time by Grzegorzewski and Mrówka [7]. Below we propose another method for computing matching degrees based on the notion of metric.

2 Basic notions

An intuitionistic fuzzy set A in a universe of discourse $\mathcal{X}$ is given by an ordered triple

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle : x \in \mathcal{X}\}, \qquad (1)$$

where $\mu_A, \nu_A : \mathcal{X} \to [0,1]$ such that

$$0 \le \mu_A(x) + \nu_A(x) \le 1 \qquad \forall x \in \mathcal{X}. \qquad (2)$$

For each x the numbers $\mu_A(x)$ and $\nu_A(x)$ represent the degree of membership and degree of nonmembership of the element $x \in \mathcal{X}$ to $A \subset \mathcal{X}$, respectively. It is easily seen that an intuitionistic fuzzy set $\{\langle x, \mu_A(x), 1 - \mu_A(x)\rangle : x \in \mathcal{X}\}$ is equivalent to the classical fuzzy set (i.e. each fuzzy set is a particular case of the intuitionistic fuzzy set). We will denote a family of all intuitionistic fuzzy subsets of $\mathcal{X}$ by $IFS(\mathcal{X})$.

Let us assume that the universe of discourse is finite, i.e. $\mathcal{X} = \{x_1, \ldots, x_n\}$. In this present paper we apply distances between intuitionistic fuzzy sets to calculate matching degrees. Namely, we consider *the normalized Hamming distance*

$$
\begin{aligned}
d_h(A,C) \;=\; & \frac{1}{n} \sum_{i=1}^{n} \max\left\{ |\mu_A(x_i) - \mu_C(x_i)|, \right. \\
& \left. |\nu_A(x_i) - \nu_C(x_i)| \right\} \qquad (3)
\end{aligned}
$$

and *the normalized Euclidean distance*

$$
\begin{aligned}
d_e(A,C) \;=\; & \left(\frac{1}{n} \sum_{i=1}^{n} \max\left\{ (\mu_A(x_i) - \mu_C(x_i))^2, \right. \right. \\
& \left. \left. (\nu_A(x_i) - \nu_C(x_i))^2 \right\} \right)^{1/2}. \qquad (4)
\end{aligned}
$$

These two metrics were obtained as based on the Hausdorff metric generalizations of well-known Hamming distance and Euclidean distance used in the classical fuzzy set theory. For more details we refer the reader to [6]. It is worth noting that using normalized distances between two intuitionistic fuzzy set as a result we obtain the number from the unit interval $[0,1]$.

3 Querying via IFS

A query may be treated as a set of searching criteria conceived by a user. A typical query expressed in SQL is written in a following form

$$
\begin{array}{lll}
\text{SELECT} & < \text{list of attributes} > & \\
\text{FROM} & < \text{list of tables} > & (5) \\
\text{WHERE} & < \text{condition} > .
\end{array}
$$

Its role is to select records (rows) that satisfy given condition. Each record from the table either satisfies or does not satisfy the condition and as a result we obtain a crisp set of database records that come up to query. However, as it was mentioned above, a traditional query syntax requires very rigid formulation of the constraints, while for a human being a common language is a natural medium to form and express his thoughts. Now we will try to construct a query that enables a direct use of linguistic terms modelled by intuitionistic fuzzy sets, i.e. a query with a following syntax:

$$
\begin{array}{ll}
\text{SELECT} & < \text{list of attributes} > \\
\text{FROM} & < \text{list of tables} > \\
\text{WHERE} & < \text{intuitionistic fuzzy condition} > .
\end{array}
$$

$$(6)$$

Let us consider a crisp relational database with a set of attributes $\mathcal{A} = \{A_1, \ldots, A_n\}$ and a set of records $\mathcal{R} = \{r_1, \ldots, r_m\}$. Let $\mathcal{X}_j$ denote the universe of discourse for the attribute A_j. Moreover, let $Z : \mathcal{R} \to \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ denote a function that determines a vector of values of all attributes corresponding to each record, i.e.

$$Z(r_i) = [z_{i1}, \ldots, z_{in}], \qquad (7)$$

where z_{ij} is a value of the attribute A_j for the record r_i.

To construct an ifs-query, a suitable intuitionistic fuzzy set must be defined for each attribute used in WHERE clause. Thus, actually, our ifs-query is an operator T which transforms each attribute A_j to the corresponding intuitionistic fuzzy set A_j^T

$$A_j^T = \{\langle x, \mu_{A_j}^T(x), \nu_{A_j}^T(x)\rangle : x \in \mathcal{X}_j\}, \qquad (8)$$

where $\mu_{A_j}^T(x), \nu_{A_j}^T(x) : \mathcal{X}_j \to [0,1]$ are the membership and nonmembership function of the defined by the intuitionistic fuzzy term T for the attribute A_j, respectively.

As soon as we accept vague terms in queries we also have to modify our meaning of *matching* between the query and a record of database. It would be unreasonable to require the answer for a ifs-query to be completely precise, adhering to the classical yes-no logic.

Now we expect the system to produce a list of records matching a query to a degree higher than a specified threshold and to list the records according to the linear semiordering. However, in our approach utilizing intuitionistic fuzzy sets we do not have such natural linear ordering, because we have to look on two functions $\mu_{A_j}^T$ and $\nu_{A_j}^T$. Therefore, we will construct a desired semiordering using distances mentioned in Sec. 2.

Let us define a function $U : \mathcal{R} \to IFS(\mathcal{R})$ which determines an intuitionistic fuzzy set R_i for each record r_i in a following way

$$
\begin{aligned}
R_i &= U(r_i) = \{\langle A_1, \mu_{R_i}(A_1), \nu_{R_i}(A_1)\rangle, \\
&\quad \dots, \langle A_n, \mu_{R_i}(A_n), \nu_{R_i}(A_n)\rangle\}, \quad (9)
\end{aligned}
$$

where $\mu_{R_i}(A_j) = \mu_{A_j}^T(z_{ij})$ and $\nu_{R_i}(A_j) = \nu_{A_j}^T(z_{ij})$. In other words

$$
\begin{aligned}
R_i &= \{\langle A_1, \mu_{A_1}^T(z_{i1}), \nu_{A_1}^T(z_{i1})\rangle, \\
&\quad \dots, \langle A_n, \mu_{A_n}^T(z_{in}), \nu_{A_n}^T(z_{in})\rangle\}. \quad (10)
\end{aligned}
$$

It is obvious that an intuitionistic fuzzy set B corresponding to the best record, i.e. the record satisfying perfectly all requirements of the query, would have a following form

$$
B = \{\langle A_1, 1, 0\rangle, \dots, \langle A_n, 1, 0\rangle\}, \quad (11)
$$

while an intuitionistic fuzzy set W corresponding to the worst record, i.e. the record that does not satisfy any requirements of the query, would look like

$$
W = \{\langle A_1, 0, 1\rangle, \dots, \langle A_n, 0, 1\rangle\}. \quad (12)
$$

We will apply intuitionistic fuzzy sets B and W in our method of calculating matching degrees. They would simply constitute the upper horizon and the lower horizon, respectively. The idea of calculating distances with respect to such horizons in the classical fuzzy set theory is given in [8] and [5].

Hence $d(R_i, B)$ and $d(R_i, W)$ denote the Hamming or Euclidean distance (3) of the intuitionistic fuzzy set R_i from the upper and lower horizon, respectively. These two numbers show how close is the record r_i to the best and to the worst possible record, respectively. Of course, while querying database we are looking for records with possibly low $d(\cdot, B)$ and possibly high $d(\cdot, W)$. Therefore, let us define

$$
\begin{aligned}
\underline{S}_i &= 1 - d(R_i, B), \quad (13) \\
\overline{S}_i &= d(R_i, W). \quad (14)
\end{aligned}
$$

It is clear that a desired record should have both values $\underline{S}_i$ and $\overline{S}_i$ as high as possible. An easy computation shows that for the Hamming distance we obtain:

$$
\begin{aligned}
\underline{S}_i &= 1 - d_h(R_i, B) = \frac{1}{n}\sum_{k=1}^{n}\mu_{R_i}(A_k), \quad (15) \\
\overline{S}_i &= d_h(R_i, W) = 1 - \frac{1}{n}\sum_{k=1}^{n}\nu_{R_i}(A_k). \quad (16)
\end{aligned}
$$

Similarly, we can consider the Euclidean distances $e(R_i, B)$ and $e(R_i, W)$ and corresponding values

$$
\begin{aligned}
\underline{S}_i &= 1 - d_e(R_i, B) \\
&= 1 - \sqrt{\frac{1}{n}\sum_{k=1}^{n}(1 - \mu_{R_i}(A_k))^2}, \quad (17) \\
\overline{S}_i &= d_e(R_i, W) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(1 - \nu_{R_i}(A_k))^2}. \quad (18)
\end{aligned}
$$

Now the question is how to apply (15), (16), (17) and (18) in matching degrees computation. We suggest here three basic methods for determining matching degrees. Namely, we can calculate the matching degree for the i-th record either as an average of $\underline{S}_i$ and $\overline{S}_i$, i.e.

$$
S_i^{AV} = \frac{\underline{S}_i + \overline{S}_i}{2}, \quad (19)
$$

or as a maximum of these two values

$$
S_i^{MAX} = \max\left(\underline{S}_i, \overline{S}_i\right), \quad (20)
$$

or as the minimum

$$
S_i^{MIN} = \min\left(\underline{S}_i, \overline{S}_i\right). \quad (21)
$$

It is easily seen that $\underline{S}_i \le \overline{S}_i$. Thus we get $S_i^{MAX} = \overline{S}_i$ and $S_i^{MIN} = \underline{S}_i$. Hence using S_i^{MIN} we restrict our consideration to the distance from the record which fits best, while using S_i^{MAX} we consider the distance from the worst possibility only. Thus S_i^{MIN} gives us an optimistic matching degree, S_i^{MAX} a pessimistic one and S_i^{AV} is a balanced one. We can also consider a natural family of operators for matching degree computation. Suppose $q \in [0, 1]$ is a constant that characterizes the subjective weight attributed to the distance from the upper and the lower horizon. Then, for given q, let us define the matching degree for the record i-th as follows

$$
S_i^q = q\underline{S}_i + (1 - q)\overline{S}_i. \quad (22)
$$

One can see easily that this operators discussed above are particular members of the family $\{S_i^q : q \in [0,1]\}$. Namely, $S_i^{AV} = S_i^{0.5}$, $S_i^{MIN} = S_i^1$ and $S_i^{MAX} = S_i^0$.

Whatever method for calculating matching degrees (note it briefly as S_i) we choose, this method induces a semiordering on a set of records. Hence we may say that a record r_i precedes record r_j (or is – in some sense – better) if and only if the matching degree S_i is not smaller than S_j, i.e.

$$r_i \succ r_j \Leftrightarrow S_i \geq S_j. \tag{23}$$

Of course, this semiordering strongly depends on the method used for calculating matching degree.

We expect the system to reject the records with matching degree lower than a specified threshold. Therefore we reject the i-th record if $S_i \leq \xi$, where ξ is a fixed number from the interval $[0,1]$. Hence we obtain a following algorithm of querying via intuitionistic fuzzy values:

1. Take the record from the database.

2. Calculate S_i.

3. Accept the record if $S_i \geq \xi$ (where $\xi \in [0,1]$), otherwise reject.

4. If there are more records go to Step 1, otherwise go to Step 5.

5. List all accepted records from the 'best' to the 'worst' according to (23).

4 Conclusions

In the present paper we have shown how to enrich fuzzy querying by the use of intuitionistic fuzzy values. Since a condition in the clause WHERE may involve not only imprecise values but also such linguistic terms as fuzzy relations, and linguistic quantifiers, some other generalizations seem natural. In further work we would try to apply intuitionistic fuzzy sets for modelling relations and in defining quantifiers too. However, we believe that even limited, our method enables the user to construct queries in a more flexible way.

References

[1] Atanassov K. (1986), *Intuitionistic fuzzy sets*, Fuzzy Sets and Systems **20**, 87–96.

[2] Atanassov K. (1999), *Intuitionistic Fuzzy Sets: Theory and Applications*, Physica-Verlag.

[3] Buckles B.P., Petry F.E. (1982), *A fuzzy representation of data for relational databases*, Fuzzy Sets and Systems **7**, 213–226.

[4] Bosc P., Kacprzyk J., Eds. (1995), *Fuzziness in Database Management Systems*, Physica-Verlag, Heidelberg.

[5] Grzegorzewski P. (1998), *Metrics and orders in space of fuzzy numbers*, Fuzzy Sets and Systems **97**, 83-94.

[6] Grzegorzewski P. (2002), *Distances between intuitionistic fuzzy sets based on the Hausdorff metric*, submitted to Fuzzy Sets and Systems.

[7] Grzegorzewski P., Mrówka E. (2002), *Soft querying via intuitionistic fuzzy sets*, Proceedings of the 9th International conference on Information Processing and management of Uncertainty in Knowledge-Based Systems IMPU'2002, Annecy, July 1-5, 2002, pp. 1845–1849.

[8] Heilpern S. (1995), *Comparison of fuzzy numbers in decision making*, Tatra Mountains Math. Publ. **6**, 47–53.

[9] Kacprzyk J., Zadrożny S. (1995), *FQUERY for Access: fuzzy querying for a Windows-based BDMS*, In: Bosc P., Kacprzyk J. (Eds.): Fuzziness in Database Management Systems, Physica-Verlag, Heidelberg, pp. 415–433.

[10] Kacprzyk J., Zadrożny S. (1997), *Flexible querying using fuzzy logic: an implementation for Microsoft Access*, In: Andreasen T., Christiansen H., Larsen H.L. (Eds.): Flexible Query Answering Systems, Kluwer, Boston, pp. 247–275.

[11] Kacprzyk J., Zadrożny S. (2000), *On combining intelligent querying and data mining using fuzzy logic concepts*, In: Bordogna G., Pasi G. (Eds.): Recent Issues on Fuzzy Databases, Physica-Verlag, Heidelberg, pp. 67–81.