Bipolar Queries and Queries with Preferences (Invited Paper)

Sławomir Zadrożny
Systems Research Institute
Polish Academy of Sciences
Warszawa, Poland
Slawomir.Zadrozny@ibspan.waw.pl
and
Warsaw School of Information Technology
Warszawa, Poland

Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
Warszawa, Poland
Janusz.Kacprzyk@ibspan.waw.pl

Abstract

The concepts of bipolar queries and queries with preferences are studied. Various interpretations of the former, recently defined by Dubois and Prade, are discussed. The latter were defined by Chomicki together with a new relational algebra operator winnow. The fuzzy version of the winnow operator is proposed. It is shown how it may be used to express a selected interpretation of the bipolar queries.

1. Introduction

Basically, a database query may be identified with a *condition* that the data sought should satisfy. Such a condition often has a complex structure and comprises some atomic conditions combined using logical connectives. Various extensions has been proposed in the literature in the framework of so-called *flexible querying*. For example the use of *fuzzy predicates* modelling *linguistic terms* in conditions has been advocated (cf., e.g., [9]). Also the assignment of *importance weights* to particular parts of the condition has been proposed and studied (cf., e.g., [7]). The aim is to provide for a greater flexibility in expressing the user requirements as to the data sought.

In 2002 Dubois and Prade [8] introduced a concept of *bipolar query*. There is a pair of (fuzzy) conditions corresponding to *rejected* and *desired* (*preferred*) sets of tuples, respectively. Thus the reason for referring to these queries as *bipolar*. Here we assume an equivalent definition of bipolar queries where a pair of the *required* and *preferred* conditions is considered. The former corresponds to the complement of the set of rejected tuples. In [8] the authors use both definitions interchangeably.

The general interpretation of the bipolar query is as follows. The tuples sought have to satisfy the *required* condition unconditionally, while the *preferred* condition is of somehow secondary importance. The interrelation between these two types of conditions may be meant in various ways and leads to different interpretations of the bipolar queries. Basically, Dubois and Prade adopt a lexicographic order based view of this interrelation.

More generally, the way these conditions are jointly taken into account may be treated as the question of a definition/selection of an appropriate aggregation operator to be applied. It turns out, that in the literature this problem has been studied under different names, and sometimes in slightly different contexts, by many authors. In the framework of database querying the paper by Lacroix and Lavency [10] was first to propose such type of queries and triggered the interest of the other researchers. This lead to the development of a more general concept of a *query with preferences* and a corresponding new relational algebra operator *winnow*, introduced by Chomicki [5, 6].

Both Lacroix and Lavency's as well as Chomicki's approach deal with crisp conditions only. In [15] we compare various interpretations of bipolar queries with and advocate one of them that is a direct "fuzzification" of the approach by Lacroix and Lavency. In this paper we propose a fuzzy version of the *winnow* operator and we show that the bipolar query, interpreted as the fuzzy counterpart of the Lacroix and Lavency approach, is a special case of a *query with preferences*, i.e., a query using the fuzzy *winnow* operator.

2. Bipolar queries

The following notation will be used: $T = \{t_j\}$ is a set of tuples to be queried; $C(\cdot)$ and $P(\cdot)$ are, *fuzzy* in general, predicates corresponding to the *required* and *preferred* conditions, respectively. We will identify these predicates with fuzzy sets and $C(t)$ and $P(t)$ will denote their membership function values. A query of this type may be exemplified with:

$$\text{\textit{Find a house cheaper than 150 000 \$ and possibly located not more than two blocks from a railway station}} \tag{1}$$

Thus, the required, strict condition concerns the price and the preferred condition refers to the distance to a railway station. A fuzzy version of such a query may be exemplified with:

$$\text{\textit{Find a cheap house preferably near a railway station}} \tag{2}$$

Basically, Dubois and Prade [8] adopt an interpretation of bipolar queries leading to the *lexicographic ordering*, $\preceq$, of the tuples, i.e.,

$$\begin{aligned} t_1 \preceq t_2 \Longleftrightarrow (C(t_1) < C(t_2)) \vee \\ \vee ((C(t_1) = C(t_2)) \wedge (P(t_1) \leq P(t_2))) \end{aligned} \tag{3}$$

The second, *possibilistic interpretation* of (1) is as follows. We are looking for a house that:

- has to be less expensive than 150 000 $
- if there is a house closer than three blocks from a railway station then other houses satisfying just the first condition are of no interest

The concept of the queries with such an interpretation has been originally proposed by Lacroix and Lavency [10] in 1987. These authors considered such queries only in the crisp context. This interpretation may be more formally expressed by the following description of the set of tuples sought ([10]):

$$\{t \in T : C(t) \wedge (\exists s(C(s) \wedge P(s)) \longrightarrow P(t))\} \tag{4}$$

In order to cover also fuzzy predicates C and P we will express the characteristic/membership function of this set, $\gamma_1(C, P, t, T)$, in the following form:

$$\begin{aligned} &\gamma_1(C, P, t, T) = \\ &\min(C(t), \max(1 - \max_{s \in T} \min(C(s), P(s)), P(t))) \end{aligned} \tag{5}$$

In this formula $\min$ and $\max$ correspond to logical conjunction and disjunction in (4), while $\neg x = 1 - x$ is used for negation and Kleene-Dienes version of fuzzy implication operator is assumed. The formula (5) has been proposed earlier by Yager [13, 12, 14] in the context of the multicriteria decision making. It refers to the concept of the conditional *possibility* of satisfying P when C is satisfied - that is why we refer to this interpretation as *possibilistic*. The approach proposed by Yager was employed by Bordogna and Pasi [2] for information retrieval tasks.

Dubois and Prade [8] also considered a formula similar to (5) which they found unsatisfactory. It was characterized by an arbitrary parameter that, in light of (5), obtains a meaningful interpretation; cf. [15] for details.

The characteristic feature of the interpretation represented via (5) is that the value of matching degree, $\gamma_1(C, P, t, T)$, for a tuple t depends not only on t, but also on the whole set of tuples T (what is appropriately accounted for by introducing the fourth parameter, T, of γ_1).

The third interpretation of bipolar queries comes from Bosc and Pivert [3, 4]. They proposed a fuzzy interpretation of the Lacroix and Lavency approach. However they assume a slightly different point of view interpreting (1) as a request to:

- find houses that are less expensive than 150 000$ (or more generally, satisfy C),
- order them in the answer in such a way that those closer than three blocks from a railway station go first (or more generally, order them according to their satisfaction of the condition P).

Thus, for them the question of combining the satisfaction of C and P is secondary. They are mainly concerned with various, non-Boolean forms of the condition P. Anyway, in [4] Bosc and Pivert propose the following formula for the aggregation of both types of conditions:

$$\begin{aligned} &\gamma_2(C, P, t, T) = \\ &\min(C(t), \max(P(t), \min_{s \in T}(P(s) : P(s) > 0) - \epsilon)) \end{aligned} \tag{6}$$

where ϵ is some small value. Basically, this formula corresponds to the conjunction, $C(t) \wedge P(t)$, modified so as to avoid 0 matching degree for a tuple satisfying C but not P. This is obviously quite different from the original concept of Lacroix and Lavency.

Thus we have three interpretations of bipolar queries:

I. possibilistic, cf., (5),

II. Bosc and Pivert's, cf. (6),

III. based on the lexicographic order, cf. (3)

It may be shown (cf. [15]) that only the *possibilistic interpretation* is an appropriate fuzzy counterpart of the original approach of Lacroix and Lavency.

3. Fuzzy *winnow* operator and bipolar queries

The paper by Lacroix and Lavency [10] started the research on queries with *preferences*. The concept of bipolar queries discussed in the previous sections may be interpreted as the development of the original idea of Lacroix and Lavency in the framework of *fuzzy logic*. In the literature a more general approach is also proposed, namely the preferences directly represented by *preference relations*, as they are used in the domain of *decision making*. In this respect the works of Chomicki [5, 6] are of a special interest. He proposes an extension to the *relational algebra* as well as to the SQL language by a new operator *winnow*. This unary operator selects from a set of tuples[1] those which are *non-dominated* with respect to a given preference relation. Chomicki defines this operator for *crisp* preference relations and *crisp* sets of tuples. In this paper we propose a *fuzzy* version of the *winnow* operator and we show how it may be used to represent *bipolar queries*, according to their *possibilistic interpretation* expressed by (5).

3.1. The *winnow* operator

In Chomicki's approach [5, 6] the *preference relation* is any binary relation R defined on the set of tuples T.

Definition 1. *The preference relation* on the set of tuples T over the scheme $\{A_1 : D_1, \ldots, A_n : D_n\}$, where A_i denotes an attribute and D_i its domain, is a binary relation R

$$R \subseteq (D_1 \times D_2 \times \ldots \times D_n) \times (D_1 \times D_2 \times \ldots \times D_n)$$

If two tuples $t, s \in T$ are in relation R, i.e., $R(t, s)$, then it is said that the tuple t *dominates* the tuple s with respect to the relation R.

The *winnow* operator is defined for certain preference relation R as a *relational algebra* operator. For a given set of tuples it yields a subset of the *non-dominated* tuples with respect to R.

Definition 2 ([5])**.** Let T be a set of tuples and R a preference relation defined on T. Then the *winnow* operator ω_R is defined as follows

$$\omega_G(T) = \{t \in T : \neg\exists_{s \in T}\, G(s, t)\} \tag{7}$$

A (relational algebra) query employing the *winnow* operator is referred to as *the query with preferences*. Chomicki [5] shows that the *winnow* operator may be expressed within the classical relational algebra. However he also points out that this operator has a great practical potential. Distinguishing the *winnow* operator makes it possible to study its behavior depending on the assumed preference relation properties. Moreover, this makes it possible to come up with specialized methods of its execution and the optimization of the execution of the whole query containing it.

The semantics of the *winnow* operator is illustrated with the following example.

[1]Here we have to refer to two types of relations: relations representing data in a database and preference relations. In order to avoid misunderstandings we will refer in this sections to the relations in the former sense as to the *sets of tuples*.

Example 1. Let us consider a database of a real-estate agency with a table (relation) HOUSES containing the details of the real-estate properties offered by the agency. The schema of the relation HOUSES contains, among other, the attributes city and price. Let us assume that we are interested in the list of *cheapest* houses in each city. We define the preference relation as follows

$$R(t, s) \Leftrightarrow (t.\texttt{city} = s.\texttt{city}) \wedge (t.\texttt{price} < s.\texttt{price})$$

Then the *winnow* operator $\omega_R(\texttt{HOUSES})$ will select the houses that are sought.

A bipolar query with *crisp* conditions: *required* C and *preferred* P may be expressed using the *winnow* operator in the following way. Let us start with the example (1). Let us define the following preference relation

$$R(t, s) \Leftrightarrow (t.\texttt{to\_station} \leq 2) \wedge (s.\texttt{to\_station} > 2)$$

Then, the following composition of the relational algebra operators yields the required results

$$\omega_R(\sigma_{\texttt{price} \leq 150000}(\texttt{HOUSES}))$$

where σ_ϕ is the classical *selection* operator that selects from a set of tuples those for which ϕ holds.

It may be easily verified that such a composition preserves the semantics of *bipolar queries* assumed in this paper. If there are houses cheaper than 150 000 $ and located closer than two blocks from the station then only them will be selected. Otherwise all houses cheaper than 150 000 $ will be selected if such exist.

In general, for a bipolar query given by a pair of required and preferred conditions (C, P), a corresponding *query with preferences* is determined as follows. The *preference relation* R is defined as

$$R(t, s) \Leftrightarrow P(t) \wedge \neg P(s) \tag{8}$$

and then the composition of the relational algebra operators

$$\omega_R(\sigma_C(T))$$

yields the same result as the bipolar query with *crisp* conditions (C, P).

3.2. The fuzzy *winnow* operator

In order to make the *winnow* operator able to express bipolar queries with *fuzzy* conditions C and P its definition has to be extended. In particular the following aspects have to be taken into account:

– R should be assumed to be a *fuzzy preference relation*,

– a fuzzy counterpart of the *non-dominance* concept has to be employed,

– the set of tuples T should be assumed to be a *fuzzy set*.

We will introduce a definition of the *fuzzy winnow operator* using the Świtalski's [11] approach to the *fuzzy choice functions*. In this approach an expression of the set of non-dominated elements is introduced that is convenient for our purposes. It has the following form

$$N(T,R) = T \cap \bigcap_{s \in T} \neg R^{-}(s) \tag{9}$$

where

$$R^{-}(s) = \{u : R(s,u)\} \tag{10}$$

and $N(T,R)$ denotes the set of non-dominated elements of the set T with respect to the preference relation R, while $R^{-}(s)$ is a set of elements *dominated by the element* s with respect to the relation R, and $\neg A$ denotes the complement of the set A.

Using (9) we may define the *winnow* operator in the following way, equivalent to (7)

$$\omega_R(T) = N(T,R) \tag{11}$$

Now let us define what will be meant here by the *fuzzy preference relation*. In fact the research on fuzzy preference relations has a very long history in the framework of the decision making theory - cf. [1] for one of the first papers on that topic. Here we will assume, following def. 1, a rather simplified definition.

Definition 3. *A fuzzy preference relation* on a *crisp* set of tuples T over the scheme $\{A_1 : D_1, \ldots, A_n : D_n\}$, is any *fuzzy* binary relation $\tilde{R}$

$$\tilde{R} \in \mathcal{F}((D_1 \times D_2 \times \ldots \times D_n) \times (D_1 \times D_2 \times \ldots \times D_n))$$

that is identified with its membership function $\mu_{\tilde{R}}$.

A set of elements non-dominated with respect to a fuzzy preference relation $\tilde{R}$ is assumed to be a fuzzy set. In order to define this set we use (9), assuming standard operations of the *intersection* and *complement* of fuzzy sets, and a following generalization of (10)

$$\mu_{\tilde{R}^{-}(s)}(u) = \mu_{\tilde{R}}(s,u) \tag{12}$$

defining the membership function of the *fuzzy set* $\tilde{R}^{-}(s)$ of elements *dominated* by the element s with respect to the relation $\tilde{R}$.

Finally, we have to adapt (9) to cover the case when the set of tuples T is a *fuzzy* set. Let us first rewrite (9) as a predicate calculus formula

$$N(T,\tilde{R})(t) \Leftrightarrow T(t) \wedge \forall_{s \in T} \neg \tilde{R}^{-}(s)(t) \tag{13}$$

where particular fuzzy predicates are denoted with the same symbols as corresponding fuzzy sets. Then applying (12) we transform (13) to

$$N(T,\tilde{R})(t) \Leftrightarrow T(t) \wedge \forall_{s \in T} \neg \tilde{R}(s,t) \tag{14}$$

Taking into account the fuzziness of the predicate T, which will be denoted as $\tilde{T}$, we replace the restricted quantifier $\forall_{s \in T}$ appearance in (14) with an equivalent non-restricted form, thus leading to

$$N(\tilde{T},\tilde{R})(t) \Leftrightarrow \tilde{T}(t) \wedge \forall_s \, (\tilde{T}(s) \rightarrow \neg \tilde{R}(s,t)) \tag{15}$$

Using the above transformations we propose the following definition.

Definition 4. Let $\tilde{T}$ be a fuzzy set of tuples and $\tilde{R}$ a fuzzy preference relation, both defined on the same set of tuples T. Then the *fuzzy winnow operator* $\omega_{\tilde{R}}$ is defined as follows

$$\omega_{\tilde{R}}(\tilde{T})(t) = N(\tilde{T},\tilde{R})(t) \qquad t \in T \tag{16}$$

where the fuzzy predicate $N(\tilde{T},\tilde{R})$ is determined by (15).

Assuming in (15) various interpretations of the fuzzy logical connectives of *conjunction*, *negation* and *implication* one obtains variants of the fuzzy *winnow* operator. Now we will show how one of these variants may be used to express bipolar queries with fuzzy conditions meant according to their *possibilistic interpretation* (5).

Let us consider a bipolar query defined by a pair of *fuzzy* conditions (C,P). We will identify these conditions with fuzzy predicates, which will be denoted with the same symbols. Let $\tilde{R}$ be a *fuzzy preference relation* of a following form (cf., (8))

$$\tilde{R}(t,s) \Leftrightarrow P(t) \wedge \neg P(s) \tag{17}$$

Then we may express a bipolar query as a combination of the selection and fuzzy *winnow* operator

$$\omega_{\tilde{R}}(\sigma_C(T)) = N(C(T),\tilde{R}) \tag{18}$$

where $C(T)$ is a fuzzy set of the elements of T satisfying (to a degree) the condition C, i.e.,

$$\mu_{C(T)}(t) = C(t)$$

Using (15) we can define the predicate (set) $N(C(T),\tilde{R})$ in (18) as follows

$$N(C(T),\tilde{R})(t) \Leftrightarrow C(t) \wedge \forall_s \, (C(s) \rightarrow \neg(P(s) \wedge \neg P(t))) \tag{19}$$

Now we will show that for a certain selection of the logical connectives in (19) the fuzzy set of tuples obtained using (18) is identical with the fuzzy set defined by (5). Let us assume the *conjunction*, *negation* and *implication* interpreted by the operators of the minimum, $n(x) = 1 - x$ and Kleene-Dienes implication, respectively. Then

$$N(C(T), \tilde{R})(t) = \tag{20}$$
$$\min(C(t), \min_{s \in T} \max(1 - C(s),$$
$$1 - \min(P(s), 1 - P(t))) =$$
$$\min(C(t), \min_{s \in T} \max(1 - C(s), \max(1 - P(s), P(t)))) =$$
$$\min(C(t), \min_{s \in T} \max(1 - C(s), 1 - P(s), P(t))) =$$
$$\min(C(t), \max(\min_{s \in T} \max(1 - C(s), 1 - P(s)), P(t))) =$$
$$\min(C(t), \max(\min_{s \in T}(1 - \min(C(s), P(s)), P(t))) =$$
$$\min(C(t), \max(1 - \max_{s \in T} \min(C(s), P(s)), P(t)))$$

Hence we obtain

$$N(C(T), \tilde{R})(t) = \gamma_1(C, P, t, T)$$

where $\gamma_1(C, P, t, T)$ is defined by (5) and expresses the *possibilistic interpretation* of the bipolar queries.

4. Concluding remarks

We propose a fuzzy version of the *winnow* operator introduced by Chomicki. We show that the proposed definition is consistent with possibilistic interpretation of bipolar queries proposed by Dubois and Prade.

References

[1] Blin J.M. Fuzzy relations in group decision theory. *Journal of Cybernetics*, 4:17–22, 1974.

[2] Bordogna G. and Pasi G. Linguistic aggregation operators of selection criteria in fuzzy information retrieval. *International Journal of Intelligent Systems*, 10(2):233–248, 1995.

[3] Bosc P. and Pivert O. Discriminated answers and databases: fuzzy sets as a unifying expression means. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 745–752, San Diego, USA, 1992.

[4] Bosc P. and Pivert O. An approach for a hierarchical aggregation of fuzzy predicates. In *Proceedings of the Second IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'93)*, pages 1231–1236, San Francisco, USA, 1993.

[5] Chomicki J. Querying with intrinsic preferences. *Lecture Notes in Computer Science*, 2287:34–51, 2002.

[6] Chomicki J. Preference formulas in relational queries. *ACM Transactions on Database Systems*, 28(4):427–466, 2003.

[7] Dubois D. and Prade H. Using fuzzy sets in flexible querying: why and how? In Andreasen T., Christiansen H., and Larsen H.L., editors, *Flexible Query Answering Systems*, pages 45–60. Kluwer Academic Publishers, 1997.

[8] Dubois D. and Prade H. Bipolarity in flexible querying. In Andreasen T. et al., editor, *FQAS 2002*, volume 2522 of *LNAI*, pages 174–182. Springer-Verlag, Berlin, Heidelberg, 2002.

[9] Kacprzyk J. and Zadrożny S. Computing with words in intelligent database querying: standalone and internet-based applications. *Information Sciences*, (134):71–109, 2001.

[10] Lacroix M. and Lavency P. Preferences: Putting more knowledge into queries. In *Proceedings of the 13 International Conference on Very Large Databases*, pages 217–225, Brighton, UK, 1987.

[11] Świtalski Z. Choice functions associated with fuzzy preference relations. In Kacprzyk J. and Roubens M., editors, *Non-Conventional Preference Relations in Decision Making*, pages 106–118. Springer-Verlag, Berlin, 1988.

[12] Yager R.R. Fuzzy sets and approximate reasoning in decision and control. In *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 415–428, San Diego, USA, 1992.

[13] Yager R.R. Higher structures in multi-criteria decision making. *International Journal of Man-Machine Studies*, 36:553–570, 1992.

[14] Yager R.R. Fuzzy logic in the formulation of decision functions from linguistic specifications. *Kybernetes*, 25(4):119–130, 1996.

[15] Zadrożny S. Bipolar queries revisited. In Torra V., Narukawa Y., and Miyamoto S., editors, *Modelling Decisions for Artificial Intelligence (MDAI 2005)*, volume 3558 of *LNAI*. Springer-Verlag, Berlin, Heidelberg, 2005.