Compound bipolar queries: combining bipolar queries and queries with fuzzy linguistic quantifiers

Janusz Kacprzyk[1] Sławomir Zadrożny[2]

[1]Systems Research Institute, Polish Academy of Sciences, email: Janusz.Kacprzyk@ibspan.waw.pl
[2]Systems Research Institute, Polish Academy of Sciences, email: Slawomir.Zadrozny@ibspan.waw.pl
[2]Kazimierz Pulaski University of Technologies and Humanities in Radom, Poland

Abstract

We propose a conceptually new type of bipolar queries meant as those in which a required condition, to be necessarily satisfied, and desired condition, to be satisfied if possible, are included. In the new type of bipolar queries, termed compound bipolar queries, both the required and desired conditions have been assumed to be queries with fuzzy linguistic quantifiers as introduced in the earlier works of the authors. We propose a logical formulation of the resulting compound bipolar queries for which a truth value can be calculated. The new type of a bipolar query proposed is a further step towards a higher human consistency of database queries to be employed by non-expert human users of an obvious limited command of database querying.

Keywords: Fuzzy query, bipolar query, query with fuzzy linguistic quantifier, fuzzy logic, De Morgan triple

1. Introduction

We are concerned with some relevant issues related to so called *flexible querying* in the context of relational (numerical) databases, to be more specific in some relevant approaches in that area, i.e. fuzzy querying, on the one hand, and bipolar querying, on the other hand.

Conceptually, our philosophy is, first, that – from the point of view of the human user – the real purpose of database querying is to retrieve information (a set of records) from a database that best reflects his or her real intention and/or preference. Second, due to the very specifics of the human being, we assume that for the human user natural language is the only fully natural means of communication and articulation of those intentions and/or preferences.

Basically, fuzzy querying boils down to providing the querying language/system employed in a particular DBMS with a facility to make it possible to use imprecise terms exemplified by values in querying conditions, like "low", "high", etc., imprecisely specified relations between some quantities in querying conditions, like "much higher than", "more or less equal", etc., and imprecisely specified aggregation operators/schemes like "most of the conditions from a set of conditions should be satisfied". These imprecise terms are represented by fuzzy sets, fuzzy relations, fuzzy linguistically quantified statements, etc.

On the other hand, we are concerned with a new, promising direction in broadly perceived flexible querying that has recently been advocated that tries to explicitly reflect bipolarity in human intention/preference articulation, and human judgments. This fact has been known for some time in social sciences, psychology, cognitive sciences, etc., that the human being while articulating his/her opinions or assessments usually tends to use some sort of a *bipolar scale* meant as that some objects, values of their features, their qualities, relations between them, etc. should rather be assessed and evaluated through:

- some degree of being *negative*, i.e., to be rejected,
- some degree of being *positive*, i.e., to be accepted.

This type of bipolarity can be dealt with using various tools and techniques,, and in this paper we will stick to those approaches that are based on fuzzy logic and possibility theory. To be more specific, we will follow the philosophy of Dubois and Prade and their collaborators, as e.g.: Benferhat, Dubois, Kaci, Prade [15], with some traces to be found in, for instance, Dubois, Prade and Smets [17], Dubois, Hájek and Prade [18] or Ughetto, Dubois and Prade [27].

There are various aspects of bipolarity that should be accounted for in the context of database querying, cf. Zadrożny, De Tré and Kacprzyk [30]. The first one is concerned with a fundamental question of which type of a bipolar scale to use. We refer here to Grabisch, Greco and Pirlot [21] who advocate two main scales: *bipolar univariate* and *unipolar bivariate*. The former assumes one scale with three levels of: negative, neutral and positive evaluation, gradually changing from one to another, while the latter assumes two independent scales separately accounting for a positive and negative eval-

uation. Technically, the intervals $[-1, 1]$ and $[0, 1]$, respectively, are predominantly employed.

It is easy to see that the inclusion of bipolarity implies some challenges which, in our context of database querying, is basically related to how to order the resulting set of tuples in the presence of bipolar evaluations.

In this paper, as well in most of our works, we deal with a special case when the positive and negative conditions are viewed as desired and required conditions, and their aggregation proceeds following an approach that has been proposed for the crisp case by Lacroix and Lavency [1] in the context of so called queries with preferences, which is then clearly extended here to the fuzzy case.

Specifically, we consider the following type of a bipolar query:

$$C \text{ and possibly } P \qquad (1)$$

exemplified by

> find a house which is *affordable* and **if possible** *conveniently located*

Notice that, obviously, the above two conditions should be aggregated but which is a proper aggregation method to reflect the tricky and nonconventional aggregation operator "...and, if possible,...", is not clear. We will adopt our approach to that aggregation (cf. Zadrożny [31]. Zadrożny and Kacprzyk [32, 34]) that follows the lines of Lacroix and Lavency, with a fuzzification and extended logical analyses.

It is obvious that a bipolar query from example (1) is certainly much more human consistent as it provides an extended representation of what a human being really wishes to retrieve.

Notice that the above bipolar query which may be viewed as a first attempt to formally represent a real human intention and preference, involves by necessity a very generally formulated desired and required condition, that is, "affordable" and "conveniently located". This is implied by the fact, which is practically assumed for any database querying problem in the case of any system that is heavily novice user focused like that for retrieving houses or apartments at a real estate agency, that the human user is just in a position to formulate his/her wishes or preferences in imprecise terms in natural language.

However, those imprecisely specified desired and required conditions are clearly not directly related to the particular attributes in the database, so that they cannot directly be employed for querying. Usually, those conditions may adequately be represented by some aggregation of conditions on attribute values. This implies that a bipolar query in question is what may be called a *compound bipolar query*. Such a type of a bipolar query will be introduced here.

Therefore, in this paper we assume that the combination of the desired and required condition is in itself a condition of a fuzzy query. For convenience, we will present this new extension in terms of our fuzzy querying interface, FQUERY for Access, which supports an extended version of SQL, cf. Kacprzyk and Zadrożny [23, 22] and Bosc et al. [7] in the sense that, first, it supports *linguistic terms* in queries exemplified by fuzzy values such as "inexpensive" and fuzzy relations (fuzzy comparison operators) such as "much less than".

This is a simple case but it makes it possible to more specifically express what the customer may wish in the case of looking for an apartment along the query of type (1). It is however not yet sophisticated enough.

In this work we propose the use in the required and desired conditions of the compound bipolar query another class of relevant linguistic terms, the so called *linguistic quantifiers* such as "most", "almost all" etc. which play the role of flexible aggregation operators. This leads to the concept of a fuzzy query with a linguistic quantifier introduced by Kacprzyk and Ziółkowski [24], and then Kacprzyk, Ziółkowski and Zadrożny [25]. Basically, this approach boiled down to the aggregation of conditions in the WHERE clause of the SQL SELECT statement as, e.g. in

> "*Most* of conditions among 'price IS *inexpensive*, floor IS *high*, ...' are to be satisfied"

Among some other approaches that could be of relevance for the future work, Bosc and Pivert's [9] use the linguistic quantifiers with subqueries or against groups of rows.

Therefore, staying for clarity in the real estate context, we have the following situation:

- we assume a bipolar query of the general type:

$$C \text{ and possibly } P \qquad (2)$$

  in which we have:

  - a required condition C, exemplified in the real estate case considered by "affordable",
  - a desired condition P, exemplified in the real estate case by "conveniently located",

- we assume that the required and desired conditions involve a linguistic quantifier driven aggregation of some partial conditions (which directly correspond to attributes in a real estate database in question!).

This boils down to the determination of a degree of truth of a linguistically quantified proposition in the sense of Zadeh.

For instance, suppose that:

- the required condition "affordable" is defined as follows:

Q of conditions among $\{c_i\}_{i=1,\ldots,n_C}$
are to be satisfied

$$(3)$$

where n_C is the number of conditions in C, exemplified by

> "*Most* of conditions among 'price IS *inexpensive*, bank loan IS *easy to get*, bank interest IS *not much higher than X%*, ...' are to be satisfied"

- the desired condition "conveniently located" is defined as follows:

Q of conditions among $\{p_i\}_{i=1,\ldots,n_P}$
are to be satisfied

$$(4)$$

where n_P is the number of conditions in P, exemplified by

> "*Most* of conditions among 'distance to railroad station IS *short*, distance to bus stop IS *much less than 5 blocks*, number of buses at stop IS *high*, ...' are to be satisfied"

Therefore, in the new compound bipolar queries proposed in this paper we have a traditional bipolar query, denoted generically as (C, P), in which both C and P are fuzzy queries themselves. Clearly, a simple bipolar query in which both C and P are composed of only fuzzy conditions imposed on attribute values may be seen as a special case of our newly introduced queries but we consider here, in particular, a non-trivial and much more realistic case when C and P are fuzzy queries with linguistic quantifiers.

We will now present a more detailed exposition of the compound fuzzy query proposed. We will start with a brief summary of the essence of fuzzy queries with linguistic quantifiers and then of bipolar queries in Sections 2 and 3, respectively. Then, we will present the essence of our new proposal.

2. Queries with linguistic quantifiers

We assume that the linguistic quantifiers are meant in the sense of Zadeh, and can be handled by two basic calculi: Zadeh's original calculus based on fuzzy logic and by using Yager's OWA operators. We will show the use of the former.

Zadeh's calculus of linguistically quantified propositions concerns natural language type expressions like

$$\text{"*Most* Swedes are *tall*"} \qquad (5)$$

where "Most" is an example of a linguistic quantifier. Other examples include "almost all", "much more than 50%" etc. These are examples of relative quantifiers we are interested in.

A linguistically quantified proposition exemplified by (5) may be written as

$$Qx\ A(x) \qquad (6)$$

where Q denotes a linguistic quantifier (e.g., *most*), $X = \{x\}$ is a universe of discourse (e.g., a set of Swedes), and $A(x)$ is a predicate corresponding to a certain property (e.g., of being *tall*).

The problem is to compute the truth value of (6). First, Q is equated with a fuzzy set defined in $[0, 1]$, and we assume that Q is a regular nondecreasing quantifier, such that

$$x \le y \Rightarrow \mu_Q(x) \le \mu_Q(y); \quad \mu_Q(0) = 0; \quad \mu_Q(1) = 1 \qquad (7)$$

where the particular $y \in [0, 1]$ correspond to proportions of elements with property A and $\mu_Q(y)$ is the degree to which a given proportion matches the semantics of Q. For example, $Q = $ "most" might be given as:

$$\mu_Q(y) = \begin{cases} 1 & \text{for } y > 0.8 \\ 2y - 0.6 & \text{for } 0.3 \le y \le 0.8 \\ 0 & \text{for } y < 0.3 \end{cases} \qquad (8)$$

The predicate A is modeled by an appropriate fuzzy set A defined in X with its membership function μ_A.

Formally, the truth degree of (6) is computed using the following formula:

$$\text{Truth}(Qx\ A(x)) = \mu_Q(\frac{1}{n} \sum_{i=1}^{n} \mu_A(x_i)) \qquad (9)$$

where $r = \frac{1}{n} \sum_{i=1}^{n} \mu_A(x_i)$ and n is the cardinality of X.

In the case of linguistically quantified statements with importance, exemplified by "*Most* (Q) of *young* (B) Swedes are *tall* (A)", written generally as

$$QBx\ A(x) \qquad (10)$$

where Q denotes a linguistic quantifier (e.g., *most*), $X = \{x\}$ is a universe of discourse (e.g., a set of Swedes), B denotes importance of the particular x's, and $A(x)$ is a predicate corresponding to a certain property (e.g., of being *tall*), we have

$$\begin{aligned} &\text{Truth}(QBx\ A(x)) \\ &= \mu_Q(\frac{\sum_{i=1}^{n} \mu_B(x_i) \wedge \mu_A(x_i)}{\sum_{i=1}^{n} \mu_B(x_i)}) \end{aligned} \qquad (11)$$

where "$\wedge$" is the minimum but may be replaced by a t-norm.

To find $\text{Truth}(Qx\ A(x))$ and $\text{Truth}(QBx\ A(x))$ one can also use Yager's *ordered weighted averaging (OWA) operators* defined as follows. Let $W \in [0, 1]^m, W = [w_1, \ldots, w_m], \sum_{i=1}^{m} w_i = 1$ be a weight vector. Then the OWA operator of dimension m and weight vector W is a function $O_W : [0, 1]^m \longrightarrow [0, 1]$ such that:

$$O_W(a_1, \ldots, a_m) = W \circ B = \sum_{i=1}^{m} w_i b_i \qquad (12)$$

where b_i is i-th largest element among a_i's and $B = [b_1, \ldots, b_m]$; $\circ$ denotes the scalar product.

The OWA operators may be used to model linguistic quantifiers. Assume that Q is a regular nondecreasing linguistic quantifier in the sense of Zadeh (7). Then the weight vector of a corresponding OWA operator is defined by Yager as follows:

$$w_i = \mu_Q \left(\frac{i}{m} \right) - \mu_Q \left(\frac{i-1}{m} \right), \qquad i = 1, \ldots, m \tag{13}$$

Then, we can easily find the corresponding truth values of linguistically quantified statements.

Since the purpose of this paper is to introduce a new class of compound bipolar queries involving as a required and desired condition a query with linguistic quantifiers, we will only assume that the truth value of a linguistically quantified proposition representing the query is calculated by using some method, and the type of the method (mainly Zadeh's calculus of linguistically quantified propositions or Yager's OWA operators) does not matter for our further considerations.

The fuzzy linguistic quantifiers may occur in various clauses of the SQL query and we will follow the approach in our FQUERY for Access package, cf. Kacprzyk and Zadrożny [23, 22], i.e., as operators aggregating conditions in the WHERE clause. For example, with "most" defined by (8), the interpretation of that aggregation may be

"*Most* of the predicates $\{A_i\}_{i=1,\ldots,n}$ are satisfied" (14)

for any number, n, of predicates.

FQUERY for Access maintains a dictionary of various user defined linguistic terms to be employed in the queries. Among them are linguistic quantifiers in the sense of Zadeh. They are all identified with a piecewise-linear membership functions and assigned a name. The linguistic quantifiers are used in a query to aggregate conditions like in (14).

Therefore, we have formal means to define and implement queries with fuzzy linguistic quantifiers.

3. Bipolar fuzzy database queries and related work

Now we will briefly show the concept of a bipolar fuzzy query as meant in this paper and discuss some properties and related aspects that will be of relevance for this paper, referring the reader to our other publications for details.

As we have already mentioned, a query is to be meant as a condition expressing what the user is looking for, and the response to a query is a list of tuples satisfying this condition(s). For simplicity we assume just for now that we have simple, atomic conditions with constraints on the values of the attributes characterizing a given relation (table), and these atomic conditions are connected using the

logical connectives of the conjunction, disjunction and negation; notice that we do not consider for now the use of linguistic quantifiers as proposed by Kacprzyk, Zadrożny and Ziółkowski [24, 25]; and they will be introduced into the bipolar queries in the next section.

Basically, such a simple fuzzy query concerning attribute X using a linguistic term modeled by fuzzy set A may be denoted as:

$$X \text{ is } A \tag{15}$$

where X in (15) may denote the attribute `price`, while the fuzzy set A represents the linguistic term "low".

A *unipolar scale* is clearly associated with (15) as $\mu_A(x)$ denotes the degree to which a given attribute value is compatible with the meaning of a given linguistic term and hence the degree to which this value satisfies the query condition. Therefore, this is a *unipolar fuzzy querying* approach.

Bipolarity in (fuzzy) database querying is essentially meant as the incorporation of *negative* and *positive* assessments/evaluations of data. In our real estate context, for a customer the location near a railroad station may be welcome (a positive assessment) while its high price is discouraging (a negative assessment). Even more, the placement near a station may be attractive (positively assessed) due to a commuting convenience and at the same time unattractive (negatively assessed) due to, e.g., the noise and social environment.

The first aspect that is crucial is related to a proper scale for expressing the bipolarity. An effective and efficient solution is to employ two types of scales (cf. Grabisch, Greco and Pirlot [21]): *bipolar univariate* and *unipolar bivariate*. In the first, an assessment is expressed as a number from $[-1, 1]$, which is divided into three parts expressing the negative (< 0), neutral (0) and positive (>0) assessments, respectively. In the second, the positive and negative assessments are expressed separately on two unipolar scales with values from $[0, 1]$. The second approach will be used here.

As for the levels of data, to which assessments are applied, they can refer to: the values of the particular attribute domains, and to the whole tuples (cf. [12])

Bipolar queries were conceptualized in the framework of broadly meant fuzzy logic by Dubois and Prade [13] (cf. Dubois and Prade [14] for a comprehensive exposition) and their basic idea is to distinguish two types of query conditions which are related to the *(flexible) constraints* representing what is required (corresponds, as the negation, to the negative condition) and what is merely *desired* (corresponds to the positive condition). Thus, in Dubois and Prade's approach there is no symmetry between negative and positive evaluations: the former are treated as more important. This is confirmed by the strategy proposed to generate an answer to such

a query which may be briefly stated as "first select (with respect to the negative condition) and then order (with respect to the positive condition)". To implement this strategy for fuzzy conditions, Dubois and Prade [13, 14, 16] propose to employ the lexicographic order of the tuples represented by vectors of two degrees of matching of the negative and positive conditions.

Research on bipolar queries in the sense adopted in this paper was triggered by the seminal paper of Lacroix and Lavency [1] who were the first to propose the use of a query with two categories of conditions (similarly to the approach of Dubois and Prade, mentioned above): C which is required (mandatory), and P which expresses just mere preferences (desires). Such a query obviously involves bivariate unipolar scale with the negative evaluation corresponding to *not* satisfying required condition C, and positive evaluation corresponding directly to satisfying desired condition P. Thus, concerning the very interpretation of negative and positive evaluations it is the same approach which was later adopted by Dubois and Prade in the above mentioned works, as well as by current authors in their works on bipolar queries.

The crucial aspect of the semantics of Lacroix and Lavency's approach is related to the "and possibly" operator which appears in (1) and was the part of this approach from the very beginning. Such an aggregation operator has been later proposed independently by Dubois and Prade but in a completely different context of default reasoning and by Yager [28, 29] in the context of multicriteria decision making for the case of so-called *possibilistically qualified criteria* that are meant as those which should be satisfied unless they interfere with the satisfaction of other criteria. This is in fact the essence of the aggregation operator "and possibly" as proposed by Lacroix and Lavency. This concept was also applied by Bordogna and Pasi [6] in information retrieval.

Lacroix and Lavency [1] consider only the case of crisp conditions C and P so that a bipolar query (C, P) may be processed using the "first select using C then order using P" strategy, i.e., the answer to the bipolar query (C, P) is obtained by, first, finding tuples satisfying C and, second, choosing from among them those satisfying P, if any. Such a very meaning was also assumed while fuzzifying the Lacroix and Lavency approach, notably in Zadrożny [31], and Zadrożny and Kacprzyk [32], and is used also here. As to other approaches, cf. Bosc and Pivert [8, 9], Dubois and Prade [14] or Lietard, Rocacher and Bosc [26].

There are various ways of making the above interpretation of bipolar queries operational as exemplified by Dubois and Prade's approach [13, 14] in which the satisfaction degrees of both conditions are considered separately and the lexicographic order on the pairs of the two satisfaction degrees is used by considering first the satisfaction degree of the

required condition.

In our works, also here, we follow another approach based on a specific formalization of the "and possibly" connective along the lines of original approach of Lacroix and Lavency [1].

Assume the queries are addressed against a set of tuples $T = \{t_j\}$ (a relation). The negative and positive assessments defining a bipolar query are identified with the predicates (fuzzy sets) that represent them, denoted as C and P, respectively (the negative assessment corresponds to the complement of C). For a tuple $t \in T$, $C(t)$ and $P(t)$ denote either that the t satisfies the respective condition (in crisp case) or degrees of this satisfaction, in a fuzzy case. Therefore, the whole bipolar query is denoted, as mentioned earlier, by (C, P).

Due to Lacroix and Lavency [1], the aggregation of C and P in "C and possibly P" proceeds as follows. A tuple t belongs to the answer set of a query (1) if it satisfies the (crisp) condition given by:

$$C(t) \ and \ possibly \ P(t)$$
$$\equiv \ C(t) \wedge \exists s(C(s) \wedge P(s)) \Rightarrow P(t) \quad (16)$$

Notice that if *there is no conflict* between P and C *with respect to the content of a database*, i.e., there are tuples satisfying both of them, then the query collapses to $C \wedge P$ while if there are no tuples satisfying both P and C, then the query collapses to C. Thus, clearly, in this interpretation of the bipolar query, the matching degree of a tuple t depends not only on t, but also on the whole set of tuples T which implies serious difficulties as the non-conventional "and possibly" aggregation cannot proceed via just the use of weights.

Zadrożny and Kacprzyk's [32] have discussed and analyzed a fuzzy version of winnow operator and its relation to bipolar queries.

The crucial issue is how to interpret the non-standard "an possibly" aggregation operator in the fuzzy context. Basically, one can derive the logical formulas expressing the matching degree of a bipolar query with the "and possibly" operator in three different main ways (cf. Zadrożny and Kacprzyk [34]):

- by a direct fuzzification of (16):

$$C(t) \ and \ possibly \ P(t)$$
$$\equiv C(t) \wedge (\exists s \ (C(s) \wedge P(s)) \Rightarrow P(t)) \quad (17)$$

- by a direct fuzzification of the winnow operator (cf. [11]):

$$C(t) \ and \ possibly \ P(t)$$
$$\equiv C(t) \wedge \neg \exists s \ ((C(s) \wedge P(s) \wedge \neg P(t))) \quad (18)$$

- by using our fuzzy version of the winnow operator (cf. Zadrożny and Kacprzyk [34]:

$$C(t) \ and \ possibly \ P(t)$$
$$\equiv C(t) \wedge \forall s \ (C(s) \Rightarrow (\neg P(s) \vee P(t))) \quad (19)$$

In the classical Boolean logic all these three formulas are equivalent but this is not true in the context of fuzzy (multivalued) logic which will be dealt with here using different versions of the the conjunction and disjunction represented by the t-norm and t-conorm, respectively [19]. We will consider the so-called De Morgan Triples $(\wedge, \vee, \neg)$ that comprise of a t-norm $\wedge$, a t-conorm $\vee$ and a negation $\neg$, where $\neg(x \vee y) = \neg x \wedge \neg y$ holds. The following three De Morgan Triples are of a special importance in fuzzy logic [19]: $(\wedge_{min}, \vee_{max}, \neg)$, $(\wedge_\Pi, \vee_\Pi, \neg)$, $(\wedge_W, \vee_W, \neg)$, called, respectively, the MinMax, Π and W triples, and the particular t-norms and t-conorms are:

$t-norms$		
$x \wedge_{min} y = \min(x, y)$		$minimum$
$x \wedge_\Pi y = x \cdot y$		$product$
$x \wedge_W y = \max(0, x + y - 1)$		$Łukasiewicz$

$t-conorms$		
$x \vee_{max} y = \max(x, y)$		$maximum$
$x \vee_\Pi y = x + y - x \cdot y$		$probabilistic\ sum$
$x \vee_W y = \min(1, x + y)$		$Łukasiewicz$

and $\neg$ in all the above De Morgan Triples is $\neg x = 1 - x$.

In fuzzy logic, the universal and existential quantifier are meant, for the finite universes, to correspond to the maximum and minimum operators, respectively.

We consider two implication operators related to the given De Morgan Triple $(\wedge, \vee, \neg)$, the so-called S-implications:

$$x \rightarrow_{S-\vee} y = \neg x \vee y \tag{20}$$

and R-implications:

$$x \rightarrow_{R-\wedge} y = \sup\{z : x \wedge z \leq y\} \tag{21}$$

Thus, for the particular De Morgan Triples, we obtain the following R-implication operators:

$$x \rightarrow_{R-min} y = \begin{cases} 1 & \text{for } x \leq y \\ y & \text{for } x > y \end{cases}$$

$$x \rightarrow_{R-\Pi} y = \begin{cases} 1 & \text{for } x = 0 \\ \min\{1, \frac{y}{x}\} & \text{for } x \neq 0 \end{cases}$$

$$x \rightarrow_{R-W} y = \min(1 - x + y, 1)$$

and the following S-implication operators:

$$x \rightarrow_{S-max} y = \max(1 - x, y)$$
$$x \rightarrow_{S-\Pi} y = 1 - x + x \cdot y$$
$$x \rightarrow_{S-W} y = \min(1 - x + y, 1)$$

Basically, in a series of our papers (cf. [33]), which culminated in Zadrożny and Kacprzyk [34], we analyzed in detail many aspects related to the choice of one of the formulas (17)–(19) to represent the bipolar queries and an appropriate modeling of the logical connectives occurring therein, i.e., the choice of one of the De Morgan Triples.

These are just some examples of more relevant properties that may be useful for the modeling of bipolar queries. For more information, we refer the reader to our recent paper [34].

Now, we will proceed to the very topic of this paper, namely the presentation of a new concept of compound bipolar queries combining traditional bipolar queries with queries with fuzzy linguistic quantifiers.

4. Compound bipolar queries: a synergistic combination of bipolar queries and queries with fuzzy linguistic quantifiers

In the discussion of the very essence and main properties of bipolar queries we have assumed up to now a simple, general form of a bipolar query, denoted by (C, P), meant as "C and possibly P". As we have already mentioned, the particular C's and P's can take on a more sophisticated form, namely can themselves be fuzzy queries with linguistics quantifiers. This gives rise to a new concept of a *compound bipolar query* proposed here as outlined in Section 1.

Since the paper is meant to propose this new concept, and due to a lack of space, we will only present in a simple and illustrative way some more relevant formulations presented for the new compound bipolar queries. For clarity, we will strongly relate our discussion to the motivating example of real estate, and this will certainly not limit the generality.

For convenience of the reader, and clarity of presentation, let us repeat now the real estate example mentioned in Section 1. Moreover, we will only use the approach based on a direct fuzzification of (16), i.e. of the source Lacroix and Lavency logical formulation of bipolar queries.

Our source bipolar query is:

> find a house which is *affordable* and **if possible** *conveniently located*

which is written as [cf. due to (1)]

$$C \text{ and possibly } P \tag{22}$$

and, due to (16), we have:

$C(t)$ *and possibly* $P(t) \equiv$
$$C(t) \wedge (\exists s\ (C(s) \wedge P(s)) \Rightarrow P(t))$$

The two types of conditions that exist in (22) are:

- a required condition C, exemplified in the real estate case considered by "affordable",
- a desired condition P, exemplified in the real estate case by "conveniently located".

We assume that the required and desired conditions involve a linguistic quantifier driven aggregation of some partial conditions (which directly correspond to attributes in a real estate database in

question!), that is, correspond themselves to queries with fuzzy linguistic quantifiers.

For instance, in our real estate context:

- the required condition "affordable" may be defined as (23), that is:

Q of conditions among $\{c_i\}_{i=1,...,n_C}$ are to be satisfied.

(23)

where n_C is the number of conditions in C, exemplified by

"*Most* of conditions among 'price IS *inexpensive*, bank loan IS *easy to get*, bank interest IS *not much higher than X%*, ...' are to be satisfied"

- the desired condition "conveniently located" may be defined as follows (cf. (4)):

Q of conditions among $\{p_i\}_{i=1,...,n_P}$ are to be satisfied.

(24)

where n_P is the number of conditions in P, exemplified by

"*Most* of conditions among 'distance to railroad station IS *short*, distance to bus stop IS *much less than 5 blocks*, number of buses at stop IS *high*, ...' are to be satisfied"

The truth values of (23) and (24) are calculated due to – obviously assuming a regular nondecreasing quantifier (7) – as:

- for the required condition C

$$\text{Truth}(Qi\ c_i(t)) = \mu_Q\left(\frac{1}{n_C}\sum_{i=1}^{n_C}\mu_{c_i(t)}\right) \quad (25)$$

where n_C is the cardinality of the set of required conditions,

- for the desired condition P

$$\text{Truth}(Qi\ p_i(t)) = \mu_Q\left(\frac{1}{n_P}\sum_{i=1}^{n_P}\mu_{p_i(t)}\right) \quad (26)$$

where n_P is the cardinality of the set of desired conditions.

Now, as we have mentioned, we will only use the approach based on a direct fuzzification of the source Lacroix and Lavency logical formulation of bipolar queries, i.e. (16).

Our query (C, P) can therefore be written as [cf. (1)]:

$$C \text{ and possibly } P \quad (27)$$

which, since C and P are queries (conditions) with fuzzy linguistic quantifiers, i.e. $C(t) \equiv Q_1 i\ c_i(t)$ and $P(t) \equiv Q_2 i\ p_i(t)$, respectively, can further be written for a notational convenience as:

$$Q_1 i\ c_i \text{ and possibly } Q_2 i\ p_i \quad (28)$$

which in turn implies due to (16)

$$Q_1 i\ c_i(t)\ and\ possibly\ Q_2 i\ p_i(t) \equiv$$
$$Q_1 i\ c_i(t)\wedge$$
$$\wedge\ (\exists s\ (Q_1 i\ c_i(s) \wedge Q_2 i\ p_i(s))$$
$$\Rightarrow Q_2 i\ p_i(t)) \quad (29)$$

We have limited our analysis, for simplicity, to C and P constituting queries with fuzzy linguistic quantifiers that are represented by linguistically quantified propositions without importance. One can clearly consider the case with importance, i.e. (10) with the truth value of the linguistically quantified proposition with importance calculated due to (11). Then, we only need to replace in (29) the respective truth values of P and C by those calculated due to (11).

A natural question that readily appears is whether a rich array of formal properties, recommendations, etc. derived for traditional bipolar queries holds for the compound bipolar queries proposed that combine the structure of the traditional bipolar queries with the inner queries with fuzzy linguistic quantifiers. This question is difficult to answer but a rich array of powerful properties of bipolar queries formulated in the literature (cf. in particular Zadrożny and Kacprzyk [34]) holds for the particular traditional bipolar queries. On the other hand, the inclusion into C and P of queries with fuzzy linguistic quantifiers does not change structurally the resulting compound bipolar query. One should therefore expect that at least most of the relevant properties of traditional bipolar queries should also be satisfied for the new compound bipolar queries proposed.

5. Concluding remarks

We have proposed a conceptually new type of bipolar queries meant as those in which a required condition, to be necessarily satisfied, and desired condition, to be satisfied if possible, are included. In the new type of bipolar queries, termed compound bipolar queries, both the required and desired conditions have been assumed to be queries with fuzzy linguistic quantifiers in the sense of the authors. We proposed a logical formulation of the resulting compound bipolar queries for which a truth value can be calculated. The new type of a bipolar query proposed is a further step towards a higher human consistency of database queries to be employed by human users of an obvious limited command of database querying.

Acknowledgments

References

[1] M. Lacroix, P. Lavency, Preferences: Putting more knowledge into queries, in: Proc. 13 Int. Conf. on Very Large Databases, Brighton, UK, 1987, 217–225.

[2] S. Benferhat, D. Dubois, S. Kaci, H. Prade, Bipolar Representation and Fusion of Preferences on the Possibilistic Logic framework, in: Proc. 8th Int. Conf. on Principles and Knowledge Representation and Reasoning (KR-02), Morgan Kaufmann, 2002, 421–448.

[3] S. Benferhat, D. Dubois, S. Kaci, H. Prade, Bipolar Possibilistic Representations, in: Proc. 18th Conference in Uncertainty in Artificial Intelligence, Morgan Kaufmann, 2002, 45–52. +

[4] S. Benferhat, D. Dubois, S. Kaci, H. Prade, Bipolar possibility theory in preference modeling: Representation, fusion and optimal solutions, Information Fusion 7, 2006 135–150.

[5] S. Benferhat, D. Dubois, S. Kaci, H. Prade, Modeling positive and negative information in possibility theory, Int. J. Intell. Syst. 23, 2008, 1094–1118.

[6] G. Bordogna, G. Pasi, Linguistic aggregation operators of selection criteria in fuzzy information retrieval, Int. J. Intell. Syst. 10 (2), 1995, 233–248.

[7] P. Bosc, O. Pivert, SQLf: A relational database language for fuzzy querying, IEEE Trans. on Fuzzy Systems 3 (1), 1995, 1–17.

[8] P. Bosc, O. Pivert, Discriminated answers and databases: fuzzy sets as a unifying expression means, in: Proc. IEEE Int. Conf. on Fuzzy Systems 1992, 745–752.

[9] P. Bosc, O. Pivert, An approach for a hierarchical aggregation of fuzzy predicates, in: Proc. 2nd IEEE Int. Conf. on Fuzzy Systems, 1993, 1231–1236.

[10] P. Bosc, O. Pivert, A. Mokhtari, L. Lietard, Extending relational algebra to handle bipolarity, in: Proc. 2010 ACM Symp. on Applied Computing (SAC), 2010, ACM, 1718–1722.

[11] J. Chomicki, Querying with intrinsic preferences, LNCS 2287, 2002, 34–51.

[12] G. De Tré, S. Zadrożny, T. Matthe, J. Kacprzyk, A. Bronselaer, Dealing with positive and negative query criteria in fuzzy database querying, LNCS 5822, 2009, 593–604.

[13] D. Dubois, H. Prade, Bipolarity in flexible querying, LNCS 2522, 2002, 174–182.

[14] D. Dubois, H. Prade, Handling bipolar queries in fuzzy information processing, in: [20], 97-114.

[15] D. Dubois, H. Prade, An introduction to bipolar representations of information and preference, Int. J. Intell. Syst. 23, 2008, 866–877.

[16] D. Dubois, H. Prade, An overview of the asymmetric bipolar representation of positive and negative information in possibility theory, Fuzzy Sets and Syst. 160, 2009, 1355–1366.

[17] D. Dubois, H. Prade, Ph. Smets, New Semantics for Quantitative Possibility Theory, in: LNCS 2143, 2001, 410–421.

[18] D. Dubois, P. Hájek, H. Prade, Knowledge-driven versus data-driven logics, J. of Logic, Lang. and Inf., 9, 2000, 65–89.

[19] J. Fodor, M. Roubens, Fuzzy Preference Modelling and Multicriteria Decision Support, Kluwer Academic Publishers, 1994.

[20] J. Galindo (Ed.), Handbook of Research on Fuzzy Information Processing in Databases, Inf. Sci. Ref., New York, USA, 2008.

[21] M. Grabisch, S. Greco, M. Pirlot, Bipolar and bivariate models in multicriteria decision analysis: Descriptive and constructive approaches. Int. J. Intell. Syst. 23, 2008, 930–969.

[22] J. Kacprzyk, S. Zadrożny, Computing with words in intelligent database querying: standalone and internet-based applications, Inf. Sci., 134 (1-4), 2001, 71–109.

[23] J. Kacprzyk, S. Zadrożny, FQUERY for Access: fuzzy querying for a windows-based DBMS, in: P. Bosc, J. Kacprzyk (Eds.), Fuzziness in Database Management Systems, Physica-Verlag, Heidelberg, 1995, 415–433.

[24] J. Kacprzyk, A. Ziółkowski, Database queries with fuzzy linguistic quantifiers. IEEE Trans. on Syst. Man and Cybern. 1986, 474–479.

[25] J. Kacprzyk, S. Zadrożny, A. Ziółkowski, FQuery III+: a "human consistent" database querying system based on fuzzy logic with linguistic quantifiers. Inf. Syst. 6, 1989, 443–453.

[26] L. Lietard, D. Rocacher, P. Bosc, On the Extension of SQL to Fuzzy Bipolar Conditions, in: Proc. NAFIPS-2009, 2009, 1–6.

[27] L. Ughetto, D. Dubois, H. Prade, Implicative and conjunctive fuzzy rules - A tool for reasoning from knowledge and examples, in: Proc. AAAI/IAAI, 1999, 214–219.

[28] R. Yager, Higher structures in multi-criteria decision making, Int. J. of Man-Machine Stud. 36, 1992, 553–570.

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

[30] S. Zadrożny, G. De Tré, J. Kacprzyk, Remarks on Various Aspects of Bipolarity in Database Querying. in: Proc. DEXA'10, International Workshops, 2010, 323–327.

[31] S. Zadrożny, Bipolar queries revisited, in: LNAI 3558, 2005, 387–398.

[32] S. Zadrożny, J. Kacprzyk, Bipolar queries and queries with preferences, in: Proc. DEXA'06, 2006, 415–419.

[33] S. Zadrożny, J. Kacprzyk, Bipolar queries using various interpretations of logical connectives, in: LNCS 4529, 2007, 181–190.

[34] S. Zadrozny and J. Kacprzyk, Bipolar queries: An aggregation operator focused perspective, Fuzzy Sets and Syst., 196, 2012, 69–81.