Flexible Database Querying and the Division of Fuzzy Relations

P. Bosc[1], L. Lietard[1] and O. Pivert[1]

The role and properties of the division are well-known in the context of queries addressed to regular relations. On the other hand, Boolean queries are sometimes too limited to cope with user needs and it is desirable to weaken the queries in order to enlarge the set of answers and/or rank them. In this paper, the extension of the division is investigated in the framework of gradual relations (i.e. relations made of weighted tuples). Several interpretations of the extended division can be envisaged which depend especially on the meaning of the grades tied to the tuples (degree of fulfillment of gradual properties or level of importance). These different interpretations and the algebraic properties of the associated operator are considered in this paper.

INTRODUCTION

The database domain is currently a matter of research and development so that a new database generation will appear which will extend the capabilities of presently available relational systems. If object-oriented data models are an important research topic, other areas are also worthy of interest, especially those aspects connected with the comfort of DBMS's users. It is often said that commercial DBMS's suffer from a lack of flexibility (even if this term has different meanings).

In this paper, the term flexible will concern DBMS's support of discriminated answers, in particular thanks to imprecise queries (whose interpretation is flexible). In this context, the problem is no longer to decide whether an element satisfies (or not) a condition but rather to what extent it satisfies this condition, which, as a matter of course, implies an order over the responses. Consequently, in such a system, two points must be fulfilled:

- Allowance of a qualitative distinction between the selected elements.
- Introduction of imprecise conditions inside queries, which is especially useful in the two following situations: 1) the user is not able to define his need in a crisp way, 2) a prespecified number of responses is desired and, therefore, a some slack is allowed to interpret the query.

It can be noticed that in so far intending to avoid empty answers through the flexibility of conditions, our view presents a connection with the works related to cooperative answers.

Several approaches for the support of imprecision in user queries can be envisaged and some of them have been proposed and implemented in the context of research prototypes.

1. IRISA/ENSSAT BP 447, 22305 Lannion Cédex, France.

Scientia Iranica, Vol. 2, No. 4, © Sharif University of Technology, January 1996.

The first idea is to consider queries made of two components: a usual one aiming at tuple selection and another to specify how to rank the previously selected elements. A second way is based on distances and resemblances. Queries involving imprecise conditions are first translated into Boolean conditions referring to intervals of acceptance rather than single values. Then, local distances (related to each elementary condition) are aggregated to determine the rank of the selected elements. Finally, a third solution is founded on fuzzy sets for the interpretation of the imprecise conditions. Here again, one can roughly consider that a distance is calculated for each tuple concerned by the query. The expression of the first two kinds of imprecise queries in terms of fuzzy predicates was the basis for showing that fuzzy sets theory provides a general setting for database flexible querying [1] since, 1) any query can be translated and 2) fuzzy predicates are capable of representing a wider range of imprecise queries, such as imprecise queries using linguistic terms, or non standard aggregations.

In order to express imprecise queries with fuzzy sets in the relational model of data [2], some work [3–5] has already been done. Selection, projection, Cartesian product, join and set operations have been deeply investigated, but the relational division has received little attention (perhaps because of its non-primitivity). If, for example, the price of products (relation S) and the quantities ordered by different stores (relation R) are known, the query looking for the stores having ordered at least 10 pieces of products priced over \$15 will be a matter of division (namely: R divided by S). At this point, one may imagine changing the previous query into: "find the stores having ordered a moderate number of pieces of all medium-priced products", which intuitively should call on a division involving fuzzy relations, R' (resp. S'), expressing the extent to which the number of pieces ordered is moderate (resp. the price of a product is medium). In this paper, relational DBMS's are considered where it is assumed that the stored data are precisely known. Concentration is focused on the division of gradual relations derived from base relations with the aid of fuzzy predicates. The main objective here is to give some semantics and meanings of this operator in the fuzzy case, since the previous introduced extensions [6–9] do not emphasize this aspect. The possible meaning(s) of the division of fuzzy relations is introduced as well as their associated expressions in an extended relational algebra. It is important to note that this approach is not connected to the relational technology but rather to the general concept of relation. Thus, it can be applied in an object-oriented framework where an object may be linked to several other objects through references.

The paper is organized as follows. In the second section, the definition of the relational division is recalled and fuzzy sets, vague predicates and fuzzy relations are briefly presented. Some possible semantics for the division of gradual relations are proposed in the third section and the next section is devoted to a study of some algebraic properties of these interpretations. In the conclusion, some future work concerning the evaluation of queries involving this extended operator are pointed out.

RELATIONAL DIVISION, FUZZY SETS AND VAGUE PREDICATES

In the framework of the relational model of data, a universe is modelled as a set of relations (in a mathematical sense i.e. a relation R is a subset of the Cartesian product of some domains) which can be manipulated with the help of specific operators known as the relational algebra. In the following, R (resp. S) is denoted as a relation defined on the set of attributes X (resp. Y).

The Relational Division

The division of $R(A, X)$ by $S(A, Y)$ denoted by $R[A \div A]S$, where A is a set of attributes common to R and S, aims at determining the X-values connected in R with all the A-values appearing in S. Formally, this operation can be defined in several ways:

- $x \in R[A \div A]S \Leftrightarrow \forall a \in S[A], (x, a) \in R$, (1)

- $x \in R[A \div A]S \Leftrightarrow a \in S[A] \Rightarrow (x, a) \in R$, (2)

- $x \in R[A \div A]S \Leftrightarrow S[A] \subseteq \Gamma^{-1}(x)$. (3)

In Expression 3, $\Gamma^{-1}(x) = \{a|(x,a) \in R\}$ and the relation R is viewed as inducing a multiple-valued mapping Γ which associates to a value a the set $\Gamma(x) = \{x|(x,a) \in R\}$; in other words, the division $R[A \div A]S$ is nothing but the lower image of $S[A]$ by Γ. The division can also be equivalently defined in terms of other relational operators (which shows its non-primitivity):

$$R[A \div A]S = R[X] - (R[X] \times S[A] - R)[X]. \quad (4)$$

In this formula, the expression $(R[X] \times S[A] - R)$ determines the tuples that are missing in R, X-values which are present in this set must be discarded from the final result, which is done by the outermost difference.

Example

Referring to the example given in the introduction, the relations PRODUCT(p#, price) and ORDERS(store, p#, quantity) were assumed to be available, therefore the query: "find the stores which have ordered at least 10 pieces of products priced over $ 15" can be expressed as O[p# ÷ p#]P with O = (ORDERS : quantity > 10) [store, p#] and P = (PRODUCT : price ≥ 15).

Fuzzy Sets

The concept of a fuzzy set, introduced by L.A. Zadeh [10], aims at extending the notion of a regular set in order to express classes with ill-defined boundaries (corresponding in particular to linguistic values, e.g. tall, young, well-paid, important, etc). Within this framework there is a gradual transition between non-membership and total membership. A degree of membership is associated to every element x of a referential X. It takes its values in the interval [0,1] instead of the pair $\{0,1\}$. When F is a finite and discrete fuzzy set, it is denoted as:

$$F = \Sigma \mu_F(x_i)/x_i = d_1/x_1 + \ldots + d_n/x_n,$$

where d_i/x_i stands for an element x_i with the degree $d_i(\mu_F(x_i) = d_i)$ and $+$ is the union operation.

It is important to clarify that, beyond these degrees themselves, the most important idea is to order the elements of the universe. It is shown in the second section that this notion is used to define imprecise predicates. The combination of such predicates will rely in particular on the set operations: intersection, union, complement and difference which are applicable to fuzzy sets according to the following definitions where A and B stand for two fuzzy sets defined over the universe X (these definitions meet the usual ones when the arguments are regular sets):

- Intersection:
  $\forall x \in X, \mu_{A \cap B}(x) = op_1(\mu_A(x), \mu_B(x))$, where op_1 is a triangular norm (associative, commutative, monotonic operator such that $op_1(a, 1) = a$).
- Union:
  $\forall x \in X, \mu_{A \cup B}(x) = op_2(\mu_A(x), \mu_B(x))$, where op_2 is a triangular co-norm (associative, commutative, monotonic operator such that $op_2(a, 0) = a$).

Among the pairs norm/co-norm of operators op_1/op_2, let us mention: $op_1(x,y) = \min(x,y); op_2(x,y) = \max(x,y)$ which will be assumed later. The complement relies on the notion of strong negation (application f from $[0,1]$ into $[0,1]$ which is involutive, decreasing and continuous such that $f(0) = 1$) and thus will retain:

$$\forall x \in X, \mu_{\bar{A}}(x) = 1 - \mu_A(x) .$$

It is easy to see that according to these definitions the double property (Morgan laws) holds:

$$A \cap B = \overline{\bar{A} \cup \bar{B}} \text{ and } A \cup B = \overline{\bar{A} \cap \bar{B}}.$$

The set difference may be defined as: $A - B = A \cap \bar{B}$ and thus:

$$\forall x \in X, \mu_{A-B}(x) = op_1(\mu_A(x), \mu_{\bar{B}}(x)).$$

In the context of usual sets, the intersection, union and complement operations are similar to AND, OR and NOT in the Boolean algebra over the pair $\{0,1\}$. However, it has been shown that the interval $[0,1]$ could not be provided with a Boolean algebra structure. Consequently, whatever the definitions retained, some of the usual properties of the set-oriented operators are no longer valid in the context of fuzzy sets. If the pair (min, max) for (op_1, op_2) and the complement to 1 for the negation are chosen, the excluded-middle law ($\bar{A} \cup A = X$) and the non-contradiction principle ($\bar{A} \cap A = \varnothing$) are lost.

The cardinality of a fuzzy set A, of universe X, has been proposed by De Luca and Termini [11] as the value:

$$\Sigma\text{Count } (A) = \sum_{x \in X} \mu_A(x).$$

Other operations specific to fuzzy sets (i.e. having no counterpart in usual sets) have been defined, such as the support of a fuzzy set defined as the regular set made of all the elements having a membership degree strictly over 0, or such as averaging operators [12, 13].

Fuzzy Implications

A fuzzy implication is an operator ($\rightarrow$) defined from $[0,1] \times [0,1]$ to $[0,1]$ which, when restricted to the values 0 and 1, must have the same behavior as the regular implication: $(0 \rightarrow 1) = 1, (0 \rightarrow 0) = 1, (1 \rightarrow 0) = 0, (1 \rightarrow 1) = 1$. These properties are supposed to be sufficient in [14]. However, fuzzy implications can also be defined according to the following five axioms [15]:

1. $(a \rightarrow 1) = 1$,
2. $(0 \rightarrow a) = 1$,
3. $(1 \rightarrow a) = a$,
4. if $b \geq c, (a \rightarrow b) \geq (a \rightarrow c)$ ($\rightarrow$ is increasing for the second argument),
5. if $a \leq c, (a \rightarrow b) \geq (c \rightarrow b)$ ($\rightarrow$ is decreasing for the first argument).

Among the diverse implications that can be found in the literature, the following three are presented which will play a further role:

- From $A \Rightarrow B = \bar{A} \cup B$, one can derive: $x \rightarrow y = \max(1-x, y)$ (Dienes implication).
- From the theorem of deduction, one can derive in particular:

  $$x \rightarrow y = \sup\{\beta | \min(x, \beta) \leq y\} = 1 \ ,$$

  if $x \leq y$ and y otherwise (Gödel implication), and:

  $$x \rightarrow y = \sup\{\beta | x^* \beta \leq y\} = 1 \ ,$$

  if $x \leq y$ and y/x otherwise (Goguen implication).

The Gödel and Goguen implications are specific cases of residuations induced by a t-norm (min for Gödel and $*$ for Goguen).

Inclusion of Two Fuzzy Sets

As illustrated above, an initial way to define the inclusion of two fuzzy sets A and B defined on the same referential X is based on:

$$A \subseteq B \Leftrightarrow \forall x, \mu_A(x) \leq \mu_B(x).$$

A more gradual view consists in defining a degree of inclusion $d(A \subseteq B)$ by using a fuzzy implication to express it:

$$A \subseteq B \Leftrightarrow \forall x((x \in A) \Rightarrow (x \in B))$$
$$\Leftrightarrow \forall x(\mu_A(x) \rightarrow \mu_B(x)),$$

and then, the indice is obtained:

$$d(A \subseteq B) = \min_x(\mu_A(x) \rightarrow \mu_B(x)). \qquad (5)$$

Another way to define a degree of inclusion is to consider:

$$A \subseteq B \Leftrightarrow \text{ Card } (A \cap B) = \text{ Card } (A) \ ,$$

and then:

$$d(A \subseteq B) = \frac{\Sigma\text{Count}(A \cap B)}{\Sigma\text{Count}(A)} = \frac{\sum_{x \in X} op_1(\mu_A(x), \mu_B(x))}{\sum_{x \in X} \mu_A(x)} \ , \qquad (6)$$

where op_1 is a triangular norm.

Fuzzy Relations and Fuzzy Predicates

In this paper, a fuzzy relation R is built over a set of domains $D_1, \ldots, D_n$ as a fuzzy subset of the Cartesian product and each element r of the relation R is provided with a membership degree $\mu_R(r)$ which expresses the extent to which r belongs to R. It is obvious that a regular relation is just a special case of a fuzzy relation where $\mu_R(r)$ equals 1. In the context of this paper, fuzzy relations will be derived from regular relations belonging to a database with the aid of fuzzy predicates.

A fuzzy predicate is similar to a fuzzy set as it expresses the extent to which the arguments fit the predicate. Elementary predicates (single or multi-variable) allow for the comparison between variables and constants or between variables and they can be combined by means of connectors such as conjunction, disjunction and mean operators [13]. Figure 1 illustrates the fuzzy predicate: "age = young".

For example, one may consider that if John is 26 years old, he satisfies the fuzzy predicate "young" with a degree of 0.55 (μ_{young} (26) being equal to 0.55). The value 0.55 is also the membership degree of John to the fuzzy set of young people.

The extensions of the usual relational operators to the case of two fuzzy relations R and S are:

- Union:

  $\forall x \in X, \mu_{R \cup S}(x) = \max(\mu_R(x), \mu_S(x))$.

- Intersection:

  $\forall x \in X, \mu_{R \cap S}(x) = \min(\mu_R(x), \mu_S(x))$.

- Difference:

  $\forall x \in X, \mu_{R-S}(x) = \min(\mu_R(x), 1 - \mu_S(x))$.

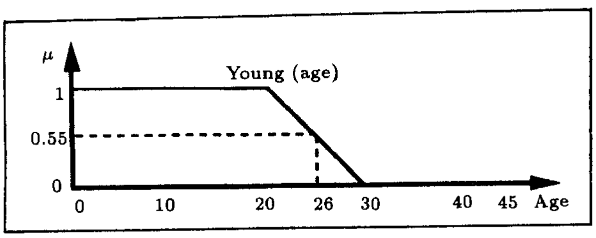

Figure 1. The fuzzy predicate: "age = young".

- Cartesian product:

  $\mu_{R \times S}(x) = \min(\mu_R(x), \mu_S(x))$.

- Selection:

  $\mu_{R:\varphi}(x) = \min(\mu_R(x), \varphi(x))$

  φ is a fuzzy predicate.

- Projection:

  $\mu_{R[\gamma]}(u) = \sup_R(\mu_R(uv))$,

  where, $u \in \gamma$, $\exists v \in (X - \gamma)$ (γ is a subset of X),

- Join:

  $$\mu_{R[A\theta B]S}(rs) = \min(\mu_R(r), \mu_S(s), \mu_\theta(r[A], s[B])).$$

  where A is a subset of X; B is a subset of Y; θ is a fuzzy comparator (i.e. a binary fuzzy relation between $r[A]$ and $s[B]$) and $r[A]$ (resp. $s[B]$) stands for the A-field value in the tuple r (resp. the B-field value in the tuple s).

All these extensions are written using the minimum (resp. maximum) for the norm (resp. co-norm). Let us detail a specific operator, denoted $R]X[$ returning the support of the projection of the relation R on the set of attributes X, defined as follows:

$$\mu_{R]X[}(x) = \begin{cases} 1 & \text{if } x = r[X] \text{ and } \mu_R(r) > 0, \\ 0 & \text{otherwise} \end{cases}.$$

SOME SEMANTICS FOR THE EXTENDED DIVISION

Introduction

Let us assume now that R and S are gradual relations, i.e. their tuples are weighted by a number between 0 and 1. The extension of the relational division to fuzzy sets naturally defines the result of the division as a fuzzy set. Then, a natural extension stems from Expression 3 where the usual set inclusion operator is changed into some multiple-valued one:

$$\mu_{R[A\div A]S}(x) = d(S[A] \subseteq \Gamma^{-1}(x)), \qquad (7)$$

$\Gamma^{-1}(x)$ being a fuzzy set defined by $\{< a/\mu > \mid < (x,a)/\mu > \in R\}$. Several definitions are possible for the degree of inclusion and a proper choice depends in particular on the intended meaning of the weights associated with tuples in relations R and S.

Fuzzy Relations and Grades

At least two possible intended meanings of the grades exist [16]:

a) Fulfillment: the weight expresses to what extent a gradual property is fulfilled, e.g. the extent to which a quantity is moderate or a price is medium.

b) Importance: the weight expresses importance, e.g. looking for stores having ordered products $p1, p4$ and $p5$, for instance it is more important to order $p4$ than $p1$ (i.e. $\mu_{PR}(p4) > \mu_{PR}(p1)$), which is of interest if the database would contain no store having ordered both $p1$ and $p4$ (then it will be pleasant to retrieve stores having ordered at least $p4$ which is more important).

The Extended Division as a Set Inclusion Operator

In the remainder of this paper, paying particular attention to the context of regular databases, only those fuzzy relations issued from regular ones by means of fuzzy predicates will be considered. It is clear that regarding the division $R[A \div A]S, \mu_{S[A]}(a)$ may be either a degree of fulfillment (i.e. a kind of threshold that should be attained) or a degree of importance of the considered value $s[A]$. Besides, $\mu_R(x,a)$ may only be a degree of fulfillment, but not a degree of importance (since R corresponds to the information available, and S to the requirement). This is why the two cases presented in Figure 2 are being considered. Two kinds of interpretation are possible for defining the inclusion of two fuzzy sets: one is based on a fuzzy implication while the other involves cardinalities. Thus the semantics of the relational

	$\mu_R(x,a)$	$\mu_{S[A]}(a)$
case 1	degree of fulfillment	degree of fulfillment
case 2	degree of fulfillment	degree of importance

Figure 2. The two considered arrangements of degrees.

division for each of these interpretations have to be defined.

Inclusion Seen as a Fuzzy Implication

Interpreting the inclusion as a fuzzy implication in the definition of the extended division, as suggested in [16], consists in extending Expression 1 in such a way that the usual implication is changed into some multiple-valued one and the universal quantifier is interpreted as a generalized conjunction:

$$\mu_{R[A\div A]S}(x) = \inf_{a\in S]A[}\ \mu_{S[A]}(a) \rightarrow \mu_R(x,a)\ . \qquad (8)$$

Case 1

When dealing with degrees of fulfillment, it seems natural to consider that a value x completely satisfies the query, i.e. belongs to $R[A \div A]S$ with the degree 1 as soon as the degree of required fulfillment $\mu_{S[A]}(a)$ is less than or equal to the degree of fulfillment $\mu_R(x,a)$ of the stored item for each s of S. Formally, it is known that:

$$\mu_{R[A\div A]S}(x) = 1 \Leftrightarrow \forall a, \mu_{S[A]}(a) \leq \mu_R(x,a).$$

Otherwise, if $\mu_{S[A]}(a) > \mu_R(x,a)$, i.e. the tuple (x,a) satisfies the property to a degree less than the required one, a convenient solution can be to return a result which reflects the proportion $\mu_R(x,a)/\mu_{S[A]}(a)$ and consequently, to use Goguen implication:

$$\mu_{R[A\div A]S}(x) = \inf_{a\in S]A[}\ \mu_{S[A]}(a) \rightarrow_G \mu_R(x,a), \qquad (9)$$

with $a \longrightarrow_G b = 1$ if $a \leq b, b/a$ otherwise.

An alternative solution consists in using Gödel implication, i.e. when $\mu_R(x,a)$ is less than $\mu_{S[A]}(a)$, one keeps for x the same weight as for (x, a):

$$\mu_{R[A\div A]S}(x) = \inf_{a\in S]A[} \mu_{S[A]}(a) \longrightarrow_{G'} \mu_R(x,a), \quad (10)$$

with $a \longrightarrow_{G'} b = 1$ if $a \leq b, b$ otherwise. This behavior is somewhat absolute since the result does not depend at all on $\mu_{S[A]}(a)$ when this degree is greater than $\mu_R(x, a)$.

Case 2

In this case, the complete satisfaction of the query by a value demands that all the values in $S]A[$, whatever their importance, are present in each tuple whose X-value is x and which completely belongs to R, i.e.:

$$\mu_{R[A\div A]S}(x) = 1 \Leftrightarrow (\forall a, \mu_{S[A]}(a) > 0 \Rightarrow \mu_R(x,a) = 1).$$

Moreover, it seems natural that $\mu_{R[A\div A]S}(x) = 0$ only if for at least one a of $S]A[$ both $\mu_{S[A]}(a) = 1$ (the requirement has the maximum level of importance) and $\mu_R(x, a) = 0$ (the tuple does not at all fulfill the requirement). However, $\mu_{R[A\div A]S}(x)$ is allowed to be strictly positive if $\mu_R(x, a) = 0$ provided that $\mu_S(s) < 1$, i.e. the requirement is not completely important (since then, the requirement can be forgotten to some extent). This desired behavior leads to the definition of the quotient operation for this case by using Dienes implication $a \longrightarrow_D b = \max(1 - a, b)$ and to the adoption of:

$$\mu_{R[A\div A]S}(x) = \inf_{a\in S]A[} \mu_{S[A]}(a) \longrightarrow_D \mu_R(x,a) = \inf_{a\in S]A[} \max(1 - \mu_{S[A]}(a), \mu_R(x,a)), \quad (11)$$

where S is a normalized fuzzy relation $(\exists u, \mu_S(u) = 1)$ in order to have an appropriate scaling of the levels of importance.

Inclusion Based on Cardinalities

The definition of the inclusion may also be based on cardinalities (cf Formula 6).

Case 1

The degrees $\mu_{S[A]}(a)$ are supposed to act as thresholds on the values $\mu_R(x, a)$. If the norm op_1 is chosen as the minimum in Formula 6, the following is obtained:

$$\mu_{R[A\div A]S}(x) = \frac{\sum_{S]A[} \min(\mu_R(x,a), \mu_{S[A]}(a))}{\sum_{S]A[} \mu_{S[A]}(a)}. \quad (12)$$

In this case, $S[A]$ is seen as a reference set of values.

Case 2

The degrees $\mu_{S[A]}(a)$ are supposed to act as weights assigned to the values $\mu_R(x, a)$. If the norm op_1 is chosen as the product, the following is obtained:

$$\mu_{R[A\div A]S}(x) = \frac{\sum_{S]A[} (\mu_R(x,a)^{\star} \mu_{S[A]}(a))}{\sum_{S]A[} \mu_{S[A]}(a)}, \quad (13)$$

which can be seen as a weighted average where $\mu_{S[A]}(a)$ expresses a degree of importance.

Conclusion

As stated above, two visions of inclusion can be used to define the extended division operator. The first one is logical and conjunctive since: 1) it is based on an implication, 2) the universal quantifier is interpreted as a generalized conjunction. A consequence of this conjunctive aspect is an "absorption effect": the division operator only retains the smallest degree of implication between $S[A]$ and R. The second vision of inclusion is compensatory. It aggregates all the values $op_1(\mu_R(x, a), \mu_{S[A]}(a))$, thus no absorption is observed.

In each case, the S-grades can express either fulfillment or importance. Recalling that only those fuzzy relations issued from regular ones by means of fuzzy predicates are considered, in this context, the meaning of the degrees (fulfillment or importance) is not an intrinsic property of a given fuzzy relation. It is in fact the choice of the division operator by the user which will confer, as explained above, a particular meaning on these degrees.

An Example

Considering Figure 3, the fuzzy relation R contains tuples which assign candidates to a position. There exist 4 skills, from I to IV, that may be achieved by each candidate. The value $\mu_R(x,u)$, ranked between 0 and 1, indicates how competent the candidate x is in regard to the skill u (the minimal value 0 means that he is not skilled, whereas the maximal value 1 means that he is fully skilled). The set of important skills is the fuzzy set $S1$. The fuzzy set $S2$ describes the skills of a particular candidate, taken as a reference.

The division R[SKILL ÷ SKILL]S1 contains the names of the candidates having all the important skills. Using Definition 11 based on Dienes implication, the following are obtained:

$$\mu_{R[SKILL \div SKILL]S1}(\text{ JOHN }) = 0.2,$$

$$\mu_{R[SKILL \div SKILL]S1}(\text{ PETER }) = 0.6,$$

whereas using Definition 13 based on cardinality/product, it is discovered that:

$$\mu_{R[\text{ SKILL} \div \text{ SKILL}]S1}(\text{ JOHN }) = 0.7,$$

$$\mu_{R[\text{ SKILL} \div \text{ SKILL}]S1}(\text{ PETER }) = 0.71 \ .$$

Relation R

NAME	SKILL	μ
JOHN	I	1
JOHN	II	0.9
JOHN	III	0.1
JOHN	IV	0.2
PETER	I	0.7
PETER	II	0.6
PETER	III	0.3
PETER	IV	0.8

Set S1

SKILL	μ
I	1
II	0.5
IV	0.8

Set S2

SKILL	μ
I	0.5
II	0.6
III	0.2
IV	0.4

Figure 3. Relation PERSON (R) and fuzzy sets of skills ($S1, S2$).

The division R[SKILL ÷ SKILL]$S2$ contains the names of the candidates having all the skills of fuzzy set $S2$, which means having at least all the skills of the reference candidate. Using Definition 9 based on Goguen implication, the following are obtained:

$$\mu_{R[\text{ SKILL} \div \text{ SKILL}]S2}(\text{ JOHN}) = 0.5,$$

$$\mu_{R[\text{ SKILL} \div \text{ SKILL}]S2}(\text{ PETER}) = 1 \ .$$

whereas, according to Definition 10 based on Gödel implication, it is found that:

$$\mu_{R[\text{ SKILL} \div \text{ SKILL}]S2}(\text{ JOHN}) = 0.1,$$

$$\mu_{R[\text{ SKILL} \div \text{ SKILL}]S2}(\text{ PETER}) = 1 \ .$$

With Definition 12 based on cardinality/minimum, it is found that:

$$\mu_{R[\text{ SKILL} \div \text{ SKILL}]S2}(\text{ JOHN}) = 0.82,$$

$$\mu_{R[\text{ SKILL} \div \text{ SKILL}]S2}(\text{ PETER}) = 1 \ .$$

In comparison with the reference candidate, Peter completely fulfills the condition and belongs fully to R[SKILL ÷ SKILL]S2 which is not the case for John. Actually, it can be checked that for each skill Peter outranks the reference candidate which is not the case for John. One can check that John has got at least 50% of each skill of the reference candidate. This is the reason why the values of $\mu_{R[SKILL \div SKILL]S2}(\text{JOHN})$ is 0.5 when using Definition 10. One can notice that John obtains a greater satisfaction degree when Formula 12 is used than when 9 or 10 are used. This is due to the fact that Formula 12 expresses a kind of compensatory behavior (an aggregation of the degrees is performed) while Formulas 9 and 10 only retain the smallest degree, since the universal quantifier is interpreted as a generalized conjunction.

SOME ALGEBRAIC PROPERTIES OF THE EXTENDED DIVISION

In this section, Expression 4 of the relational division is extended to the fuzzy case and some results concerning the algebraic properties of the different interpretations presented in the preceding section are set out.

At that point, if it is assumed that the algebraic operations are defined for fuzzy relations (cf second section), the status of Expression 4, with respect to its ability to express an extended division, must be examined. First, let us note that in this expression the term $R[X]$ is used to define a referential of X-values. Consequently, if this expression is considered when the relations R and S are possibly fuzzy, an extended projection cannot be used:

$$\mu_{R[\gamma]}(u) = \sup_{r \in R \text{ and } r[\gamma]=u} \mu_R(r),$$

and the operator $R]X[$ is required for returning the support of the projection of the relation R on the set of attributes X.

Result 1

According to the definitions given above, if R (resp. S) is a fuzzy (resp. fuzzy normalized) relation, the equality: $R[A \div A]S = R]X[-(R]X[\times S - R)[X]$ holds if the division is based on Dienes implication, i.e.: $\mu_{R[A \div A]S}(x) = \inf_S \mu_S(s) \rightarrow_G \mu_R(x, s[A]) = \inf_S \max(1 - \mu_S(s), \mu_R(x, s[A]))$.

Proof

R_x is denoted as the tuples of R whose X-value is x (each tuple has the form $< x, a_i > / \mu_i$). Similarly, $< y_i, a_i > / \mu'_i$ is written as an S-tuple. Then, the evaluation of the degree of x according to Formula 4 ($\mu_{R]X[-(R]X[\times S[A]-R)[X]}(x)$) requires calculation of:

- $R_x]X[\times S = \cup_i(< x, a_i > / \mu'_i)$,
- $R_x]X[\times S - R_x = \cup_i(< x, a_i > / \min(\mu'_i, 1 - \mu_i))$,
- $(R_x]X[\times S - R_x)[X] = < x > / \sup_i \times \min(\mu'_i, 1 - \mu_i)$,
- $R_x]X[-(R_x]X[\times S - R_x)[X] = < x > /1 - \sup_i \min(\mu'_i, 1 - \mu_i) = < x > / \inf_i \max(1 - \mu'_i, \mu_i)$.

This value is exactly that returned by Formula 11•

This result generalizes Formula 4 that is valid for regular relations.

Now some properties of the interpretations of the extended division depending on the chosen implication are pointed out. The result of the usual division of $R(A, X)$ by $S(A, Y)$ is a set of X-values included in those appearing in R (see Expression 4).

Result 2

If R and S are fuzzy relations, the inclusion: $R[A \div A]S \subseteq R[X]$ is not valid in general if the division $R[A \div A]S$ is based on:

1. Goguen implication (except when S is normalized),
2. Gödel implication (except when S is normalized),
3. cardinality/minimum.

On the other hand, the inclusion $R[A \div A]S \subseteq R[X]$ is valid if the division is based on:

1. Dienes implication (in this case S must be normalized) or
2. cardinality/product.

Proof

Consider the following extensions: $R = < x, a1 > /.2$ and $S = < a1 > /.1$.

1. With Formula 9, i.e. Goguen implication, $R[A \div A]S = < x > /1$ is obtained which is not included in $R[X](< x > /.2)$.
2. With Formula 10, i.e. Gödel implication, the same result is obtained.
3. With Formula 12, i.e. cardinality/minimum, $R[A \div A]S = < x > /1$ is also obtained.
4. Consider the definition of the division given in Formula 11. Due to the normalization of S, there exists a tuple s_0 in S such that:

$\mu_S(s_0) = 1$ and:

$$\mu_{R[A\div A]S}(x) = \inf_S \max(1-\mu_S(s), \mu_R(x, s[A]))$$
$$\leq \max(1-\mu_S(s_0), \mu_R(x, s_0[A]))$$
$$\max(1-\mu_S(s_0), \mu_R(x, s_0[A]))$$
$$= \mu_R(x, s_0[A])$$
$$\leq \sup_{r\in R\wedge r[X]=x} \mu_R(r)$$
$$\sup_{r\in R\wedge r[X]=x} \mu_R(r)$$
$$= \mu_{R[X]}(x) \ ,$$

then:

$$\mu_{R[A\div A]S}(x) \leq \mu_{R[X]}(x) \ .$$

5. Formula 13 is the expression of a weighted mean. Since the result of a mean is less than the maximum of the arguments, $\mu_{R[A\div A]S}(x) \leq \mu_{R[X]}(x)$ is obtained•

Another interesting property valid for the regular division is the commutativity of the division $R[A \div A]S$ and a selection bearing on X by means of a condition denoted by φx. In other words, when R and S are regular relations, the following formula is obtained:

$$R[A \div A]S : \varphi_x = (R : \varphi_x)[A \div A]S.$$

The question is whether this holds true for fuzzy relations and will be examined for each of the five interpretations presented before.

Result 3

Let R and S be fuzzy relations; the equality: $R[A \div A]S : \varphi_x = (R : \varphi_x)[A \div A]S$ does not hold in general if the division $R[A\div A]S$ is based on:

1. Goguen implication.
2. Gödel implication.
3. Cardinality/minimum.
4. Cardinality/product.

On the other hand, if R (resp. S) is a fuzzy (resp. fuzzy normalized) relation and the division $R[A \div A]S$ is based on Dienes implication, the equality holds.

Proof

First consider a value x such that $\mu_{\varphi_x}(x) = .5$ and the extension $\{< x, a1 > /.8, < x, a2 > /.6\}$ for $R(X, A)$, $\{a1/.4, a2/.2\}$ for $S(A)$. This yields:

1. $\mu_{(R\div S):\varphi_x}(x) = \min(.5, \min(.4 \rightarrow_G .8, .2 \rightarrow_G .6)) = .5$,

   whereas:

   $\mu_{(R:\varphi_x)\div S}(x) = \min(.4 \rightarrow_G \min(.8, .5), .2 \rightarrow_G \min(.6, .5)) = 1$.

2. $\mu_{(R\div S):\varphi_x}(x) = \min(.5, \min(.4 \rightarrow_{G'} .8, .2 \rightarrow_{G'} .6)) = .5$,

   whereas:

   $\mu_{(R:\varphi_x)\div S}(x) = \min(.4 \rightarrow_{G'} \min(.8, .5), .2 \rightarrow_{G'} \min(.6, .5)) = 1$.

3. $\mu_{(R\div S):\varphi_x}(x) = \min(.5, (\min(.4, .8) + \min(.2, .6))/.6 = .5$,

   whereas:

   $\mu_{(R:\varphi_x)\div S}(x) = (\min(.4, \min(.8, .5)) + \min(.2, \min(.6, .5)))/.6 = 1$.

4. Suppose that $\mu_{\varphi_x}(x) = .7$. It is shown that:

   $\mu_{(R\div S):\varphi_x}(x) = \min(.7, (.4^\star .8 + .2^\star .6)/.6) = .73$,

   whereas:

   $\mu_{(R:\varphi_x)\div S}(x) = (.4^\star \min(.8, .7) + .2^\star \min(.6, .7))/.6 = .67$.

5. As there is no room here for an exhaustive demonstration only the main lines of the proof are given:

   $$\mu_{R[A\div A]S:\varphi_x}(x) = \min(\mu_{\varphi_x}(x), \inf_S \max(1-\mu_S(s), \mu_R(x, s[A]))).$$

Case 1

$\mu_{R[A \div A]S:\varphi_x}(x) = \alpha = \mu_{\varphi_x}(x)$ and $\forall s \in S \max(1 - \mu_S(s), \mu_R(x, s[A])) \geq \alpha$. Let the following denote :

$$S_0 = \{s \in S \mid (1 - \mu_S(s)) \geq \alpha \text{ and } \mu_R(x, s[A])) \geq \alpha\},$$

$$S_1 = \{s \in S \mid (1 - \mu_S(s)) \geq \alpha \text{ and } \mu_R(x, s[A])) < \alpha\},$$

$$S_2 = \{s \in S \mid (1 - \mu_S(s)) < \alpha \text{ and } \mu_R(x, s[A])) \geq \alpha\}.$$

It is shown that:

$$\mu_{(R:\varphi_x)[A \div A]S_0}(x) \geq \alpha, \mu_{(R:\varphi_x)[A \div A]S_1}(x) \geq \alpha,$$

and:

$$\mu_{(R:\varphi_x)[A \div A]S_2}(x) = \alpha,$$

therefore it can be concluded that:

$$\mu_{(R:\varphi_x)[A \div A]S}(x) = \alpha,$$

since s_0 such that $\mu_S(s_0) = 1$ belongs to S_2.

Case 2

$$\mu_{R[A \div A]S:\varphi_x}(x) = \alpha = \max(1 - \mu_S(s_1), \mu_R(x, s_1[A])) .$$

$$\forall s \neq s_1 : \max(1 - \mu_S(s), \mu_R(x, s[A])) \geq \alpha,$$

and:

$$\mu_{\varphi_x} \geq \alpha .$$

It is possible to show that:

1. $\forall s \neq s_1 : \max(1 - \mu_S(s), \min(\mu_{\varphi_x}(x, s[A]), \mu_R(x, s[A]))) \geq \alpha,$
2. $\max(1 - \mu_S(s_1), \min(\mu_{\varphi_x}(x, s_1[A]), \mu_R(x, s_1[A]))) = \alpha,$

and here again, it can be concluded that: $\mu_{(R:\varphi_x)[A \div A]S}(x) = \alpha$•

CONCLUSION

This paper is mainly concerned with the extension of the division of two relations R and S, denoted as $R[A \div A]S$, in the context of relational database management systems. In the regular case, this operation can be seen in different ways (implication or set containment especially) and its extension to the case has been investigated where its arguments are fuzzy relations(for instance intermediate relations obtained from base relations with the help of fuzzy predicates), where a grade is tied to each tuple.

One interesting point in this approach concerns the meaning of the extended operation (of prime importance for the user) and four of them have been pointed out which are based on multiple-valued implications (Dienes, Gaines/Gödel) and on aggregation of cardinalities involving a triangular norm (minimum or product). These meanings depend on the semantics of the grades in relation S. Grades in relation R are assumed to express degrees of fulfillment of a condition whereas grades in S may express either degrees of fulfillment or importance. A formal definition of these meanings has been proposed.

The expression of the extended divisions in the framework of SQLf, a language supporting fuzzy querying [5], has also been investigated, but this aspect has not been presented here for the sake of brevity. It has been shown that one classical expression of the division in terms of "in" and "exists" nestings in SQL led to the algebraic interpretation based on Dienes implication provided that S is normalized. The interpretation relying on Goguen/Gödel implication can be attained only with a special construct of SQLf (set containment along with a grouping). As a matter of fact, this construct is very powerful since it allows coping with each one of the four proposed interpretations (refer to [5] for more details).

The evaluation of queries involving an extended division (whatever its interpretation) has not yet been addressed and should be a matter for future work as well as the case where the universal quantifier underlying the division is weakened into "almost all" as suggested in [6, 16].

REFERENCES

1. Bosc, P. and Pivert, O. "Some approaches for relational databases flexible querying", *J. of Intelligent Information Systems*, **1**, pp 323–354 (1992).

2. Codd, E.F. "A relational model of data for large shared data banks", *Communication of the ACM*, **13**, pp 377–387 (1970).

3. Tahani, V. "A conceptual framework for fuzzy query processing; a step toward very intelligent database systems" *Information Processing and Management*, **13**, pp 289–303 (1977).

4. Baldwin, J.F. and Zhou, S.Q. "A fuzzy relational inference language", *Fuzzy Sets and Systems*, **14**, pp 155–174 (1984).

5. Bosc, P. and Pivert, O. "SQLf: a relational database language for fuzzy querying", *IEEE Transactions on Fuzzy Systems*, **3**, pp 1–17 (1995).

6. Yager, R.R. "Fuzzy quotient operators for fuzzy relational databases", *International Fuzzy Engineering Symposium (IFES'91)*, Yokohama, Japan, pp 289–296 (1991).

7. Mouaddib, N. "The nuanced relational division", *2nd IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'93)*, San Francisco, USA, pp 419–1424 (1993).

8. Cubero, J.C., Medina, J.M., Pons, O. and Vila, M.A. "The generalized selection: an alternative way for the quotient operations in fuzzy relational databases", *5th IPMU Conference*, Paris, France, pp 23–30 (1994).

9. Umano, M. and Fukami, S. "Fuzzy relational algebra for possibility-distribution-fuzzy-relational model of fuzzy data", *J. of Intelligent Information Systems*, **3**, pp 7–28 (1994).

10. Zadeh, L.A. "Fuzzy sets", *Information and Control*, **8**, pp 338–353 (1965).

11. De Luca, A. and Termini, S. "A definition of non-probabilistic entropy in the setting of fuzzy sets theory", *Information and Control*, **20**, pp 301–312 (1972).

12. Dubois, D. and Prade H. "A review of fuzzy set aggregation connectives", *Information Sciences*, **36**, pp 85–121 (1985).

13. Yager, R.R. "Connectives and quantifiers in fuzzy sets", *Fuzzy Sets and Systems*, **40**, pp 39–75 (1991).

14. Bandler, W. and Kohout, L.J. "Fuzzy power sets and fuzzy implication operations", *Fuzzy Sets and Systems*, **4**, pp 13–30 (1980).

15. Yager, R.R. "An approach to inference in approximate reasoning", *International Journal of Man-Machine Studies*, **13**, pp 323–328 (1980).

16. Dubois, D. and Prade H. "Quotient operators in fuzzy relational databases", *2nd European Congress on Fuzzy and Intelligent Techniques (EUFIT'94)*, Aachen, Germany, pp 357–360 (1994).