

# Introducing the aggregate "count" into flexible queries

Patrick Bosc  
IRISA/ENSSAT  
Technopole ANTICIPA BP 447  
22305 Lannion Cédex France  
[bosc@enssat.fr](mailto:bosc@enssat.fr)

Ludovic Liétard  
IRISA/IUT  
Rue Édouard Branly BP 150  
22302 Lannion Cédex France  
[lietard@iut-lannion.fr](mailto:lietard@iut-lannion.fr)

## Abstract

Flexible querying aims at expressing preferences inside queries instead of strict requirements. We consider ordinary relational databases and flexible querying calling on fuzzy sets. The topic of this paper is to define new types of flexible conditions: the ones involving the count operator. The main difficulty is to handle cardinalities of fuzzy sets.

**keywords** : databases, flexible querying, fuzzy sets, cardinalities of fuzzy sets

## 1 Introduction

This paper concerns querying of relational databases which mainly relies on the expression of conditions allowing for preferences. The data stored in the database remains precisely known but conditions may be vague which means that their satisfaction is gradual.

After a flexible querying, each answer is provided with a degree expressing its fulfillment with respect to the query. In so doing, the user is provided with a more discriminated set of answers. Let us consider the example of a user who is looking for "*economical* restaurants *not too far* from downtown". The Boolean query: "retrieve the restaurants having a menu less than \$20 located less than 2 miles from the city hall" delivers an empty answer if no restaurant of the database satisfies both conditions and a restaurant with a menu at \$22, located at 1.5 mile from the city hall, cannot be an answer. In case of a flexible query, the two vague predicates *economical* and *not too far from*

*downtown* are available and the query is turned into: "retrieve the restaurants having an *economical* menu and *not too far from downtown*". In that case, the previous restaurant is presented as a *somewhat acceptable* element.

We consider relational databases where it has been shown [1] that fuzzy set theory [7] provides a general sound framework for expressing flexible queries. In this context, a vague predicate (such as *economical*) is defined by a fuzzy set and its satisfaction is expressed by a degree of membership. It is also possible to consider fuzzy connectors expressing preferences or compensation between vague predicates and all these components of flexible queries have been widely studied until now. However, flexible queries calling on cardinalities have not been studied and they constitute the topic of this paper.

In a Boolean query, it is possible to take into consideration the cardinality of a set defined by a Boolean condition. An example is given by the query "retrieve the firms where the number of employees having an age under 20 is larger than the number of employees with a salary over \$15,000". This query needs to compare the cardinality of two sets of employees. In the context of flexible querying, the previous query becomes "retrieve the firms where the number of *young* employees is larger than the number of *well-paid* employees", which leads to compare the cardinalities of two fuzzy sets: *young* employees on the one hand and *well-paid* employees on the other hand. Another example is given by the query "retrieve the firms where the number of *young* employee is *much greater than 4*". This case leads to determine the extent to which the cardinality of *young* employees is *much greater than 4*.

This paper proposes a theoretical framework to define the semantics (and interpretation) of such queries. More precisely, we would like to allow vague conditions calling on cardinalities in flexible queries. So, we propose an interpretation to determine (in terms of a degree of satisfaction) the extent to which these conditions are fulfilled.

In the remainder, the cardinality of any fuzzy set  $A$  is denoted  $C_A$  and that of the empty set  $C_\emptyset$ . Section 2 introduces two kinds of vague conditions involving cardinalities. In section 3, some approaches to evaluate the degree of truth of such conditions are suggested and section 4 concludes the paper.

## 2 Vague conditions and cardinalities

This section shows two types of vague conditions where cardinalities can be introduced. Vague conditions of type I are such that the cardinality of a fuzzy set satisfies a vague condition, whereas in vague condition of type II two cardinalities are compared.

### 2.1 Type I conditions

If  $A$  and  $V$  are denoting two vague conditions,  $X$  a regular set of elements, a condition of type I is written:

$$C_A \text{ is } V$$

Such a condition means: "the cardinality of the fuzzy set made of  $A$  elements from  $X$  satisfies condition  $V$ ". Condition  $V$  is a constraint defined by a fuzzy subset of naturals and each natural (i.e. cardinality) is associated with a degree expressing its adequation with the constraint. Consequently, this constraint can be interpreted as a linguistic quantifier [8]. In addition, as a type I condition means:

"among  $X$  elements, a quantity  $V$  satisfies  $A$ ",

it has same meaning then the quantified statement " $\forall X$  are  $A$ " and it is possible to define " $C_A$  is  $V$ " by the quantified statement " $\forall X$  are  $A$ " [8].

**Example 1.** Let  $A$  be the vague predicate *much greater than 4* given by figure 1 and the query: "retrieve the firms where the number of *young* employee is *much greater than 4*".

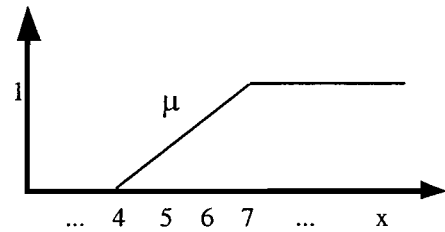


Figure 1. Fuzzy predicate *much greater than 4*

A firm  $X$  is an answer to the query at a degree given by the flexible condition "the cardinality of *young* employees from  $X$  is *much greater than 4*". This condition:

$$C_{\text{young}} \text{ is } V,$$

is of type I with  $V$  the constraint *much greater than 4*. An interpretation of  $V$  in terms of linguistic quantifier is *much more than 4*. The quantified statement "*much more than 4* employees are *young*" has same meaning than the type I condition "the cardinality of *young* employees from  $X$  is *much greater than 4*" ♦

### 2.2 Type II conditions

If  $A$  and  $B$  are two vague conditions, a type II condition is written:

$$C_A \theta C_B$$

where  $\theta$  is taken into  $\{\leq, >, \geq, <, =, \neq\}$ . This condition means that the cardinality of fuzzy set  $A$  is in relation  $\theta$  with that of fuzzy set  $B$ . A fuzzy operator  $\theta$  could also be used but this case is not considered in this preliminary paper.

**Example 2.** We consider the query "retrieve the firm where the number of *young* employees is larger than the number of *well-paid* employees". A firm  $X$  is an answer at a degree given by the condition: "the cardinality of *young* employees is larger than the cardinality of *well-paid* employees". This condition is of type II and is written  $C_{\text{young}} > C_{\text{well-paid}}$  ♦

The study of type II conditions can be limited to the case where  $\theta$  is  $\leq$  since other formulations can be derived from that case.

## 3 Evaluation of type I or type II conditions

A type I condition is equivalent to a quantified statement and it has been shown [2, 3] that, in the context of flexible querying, such statements can be evaluated by the Choquet [5] or the Sugeno [6] fuzzy integral.

This section is then limited to type II conditions and proposes a way to determine the extent to which the condition  $C_A \leq C_B$  is satisfied.

If A and B are two regular sets (not fuzzy), it is then possible to define  $C_A \leq C_B$  by (N being the set of naturals):

$$\begin{aligned} \forall i \in N, \\ \text{"at least } i \text{ elements in A"} \\ \Rightarrow \text{"at least } i \text{ elements in B"} \end{aligned}$$

This definition serves as a basis for our proposition, i.e., when A and B are two fuzzy sets. In that case, the expression "at least i elements in A" (resp. "at least i elements in B") is more or less satisfied and we denote  $\text{atleast}(i,A)$  the degree of truth of statement "at least i elements in A", for any natural i and any fuzzy set A. We obviously have :

$$\begin{aligned} \text{atleast}(0,A) &= 1 \\ \text{atleast}(i,A) &= \text{the } i^{\text{th}} \text{ highest membership degree in A and 0 otherwise.} \end{aligned}$$

In practise, if membership grades are sorted:  $\mu_{A_1} \geq \mu_{A_2} \geq \dots \geq \mu_{A_n}$  then  $\text{atleast}(i,A)$  equals  $\mu_{A_i}$ . If the universal quantifier is interpreted as a generalized conjunction, the degree of satisfaction of  $C_A \leq C_B$  is given by:

$$d(C_A \leq C_B) = \min_i \text{atleast}(i,A) \rightarrow \text{atleast}(i,B)$$

where  $\rightarrow$  stands for a fuzzy implication. Many implications are available [4] (Gödel, Goguen, Kleene-Dienes, ...) and we suggest to choose Lukasiewicz's implication ( $a \rightarrow_{Lu} b = 1$  if  $a \leq b$  and  $1-a+b$  otherwise) since it takes into account the difference between the antecedent and the conclusion which is meaningful.

If  $\text{atleast}(i,A) \rightarrow_{Lu} \text{atleast}(i,B)$  equals 1 it means that each  $i^{\text{th}}$  highest degree of membership in B is larger than (or equal to) the  $i^{\text{th}}$  highest membership degree in A. It is then certain that B has more element than A.

**Example 3.** Let A and B be two fuzzy sets with their respective degrees of membership:  $\{1, 0.89, 0.01\}$  and  $\{1, 0.9\}$ . We get:

$$\begin{aligned} d(C_A \leq C_B) \\ = \min(1 \rightarrow_{Lu} 1, 1 \rightarrow_{Lu} 1, 0.89 \rightarrow_{Lu} 0.9, 0.01 \rightarrow_{Lu} 0) \\ = \min(1, 1, 1, 0.99) = 0.99. \end{aligned}$$

Due to the comparison between 0.01 and 0, it is impossible to obtain  $d(C_A \leq C_B) = 1$ . The difference between 0.01 and 0 being extremely low, the result is however very close to 1 (0.99)♦

Lukasiewicz's implication has the following properties:

P1 It is certain that any fuzzy set has at least as many elements as the empty set :

$$\forall A, d(C_{\emptyset} \leq C_A) = 1;$$

P2 The lower the height of a fuzzy set, the more its cardinality tends towards that of the empty set :

$$\forall A \neq \emptyset, d(C_A \leq C_{\emptyset}) = 1 - \max \mu_A(x);$$

P3 If A is included in (or equal to) B, the cardinality of A is smaller than (or equal to) that of B:

$$\text{if } A \subseteq B (\forall x \mu_A(x) \leq \mu_B(x)) \text{ then } d(C_A \leq C_B) = 1;$$

Let B be a regular set of cardinality n (represented as n degrees equal to 1) and A a fuzzy set. We get:

P4 The degree  $d(C_B \leq C_A)$  expresses the extent to which A contains at least n elements:

$$d(C_B \leq C_A) = \mu_{A_n},$$

$\mu_{A_n}$  being the  $n^{\text{th}}$  highest  $\mu_A(x)$ ,

and conversely, the degree  $d(C_A \leq C_B)$  expresses the extent to which A contains at most n elements:

$$d(C_A \leq C_B) = 1 - \mu_{A_{n+1}},$$

$\mu_{A_{n+1}}$  being the  $(n+1)^{\text{th}}$  highest  $\mu_A(x)$ .

#### 4 Conclusion

In this paper, we have considered vague conditions involving cardinalities in the context of flexible queries. Two types of conditions have been distinguished: those where the cardinality is checked against a fuzzy condition (e.g., "retrieve the firms where the number of *young* employee is *much greater than 4*") and those where two cardinalities are compared ("retrieve the firms where the number of *young* employees is larger than the number of *well-paid* employees").

We have recalled that a condition of first type is equivalent to a quantified statement and, consequently it can be evaluated by the Choquet [5] or the Sugeno [6] fuzzy integral. Concerning the conditions of the second type we have focused on the comparison  $C_A \leq C_B$  ( $C_A$  and  $C_B$  being the cardinalities of two fuzzy sets A and B), since other formulations (based on  $\{>, \geq, <, =, \neq\}$ ) can be derived from that case.

The contribution of the paper relies in the proposal for determining the extent to which the proposition  $C_A \leq C_B$  is true (in terms of a degree  $d(C_A \leq C_B)$ ). A definition of  $C_A \leq C_B$  when A and B are regular sets is: for any integer i, "at least i elements in A"  $\Rightarrow$  "at least i elements in B" and it has been extended to fuzzy sets using a fuzzy implication (Lukasiewicz's one) and a generalized conjunction (min). The closer to 1  $d(C_A \leq C_B)$ , the more the cardinality of A reaches that of B. Some properties related to that degree have also been pointed out, in particular properties available in the crisp case (e.g.,  $\forall A d(C_\emptyset \leq C_A) = 1$  and if  $A \subseteq B$  then  $d(C_A \leq C_B) = 1$ ). The originality of the contribution resides in the fact that the approach taken here does not lead to compute the explicit cardinality of fuzzy sets involved in flexible conditions.

involving a fuzzy comparison operator between two cardinalities would also be of interest. Last, this paper can be seen as a first attempt to introduce aggregate operations over fuzzy sets in a flexible query framework.

## References

- [1] P. Bosc, O. Pivert (1992), Some Approaches for Relational Databases Flexible Querying, *Journal of Intelligent Information Systems*, 1, 323-354.
- [2] P. Bosc, L. Liétard and O. Pivert (1995), Quantified statements and database fuzzy querying, in *Fuzziness in Database Management Systems*, (P. Bosc, J. Kacprzyk eds), Physica-Verlag, 275-308.
- [3] P. Bosc, L. Liétard (1994), On the comparison of the Sugeno and the Choquet fuzzy integrals for the evaluation of quantified statements, *Proceedings of the third European Congress on Fuzzy and Intelligent Technologies (EUFIT'95)*, Aachen (Allemagne), 709-716.
- [4] D. Dubois, H. Prade (1985), A review of fuzzy set aggregation connectives, *Information Sciences*, 36, 85-121.
- [5] T. Murofushi, M. Sugeno (1989), An interpretation of fuzzy measure and the Choquet integral as an integral with respect to fuzzy measure, *Fuzzy Sets and Systems*, 29, 201-227.
- [6] M. Sugeno (1974), Theory of fuzzy integrals and its applications, *Ph. D Thesis, Tokyo Institute of Technology*, Tokyo (Japan).
- [7] L.A. Zadeh (1965), Fuzzy sets, *Information and Control*, 8, 338-353.
- [8] L.A. Zadeh (1983), A computational approach to fuzzy quantifiers in natural languages, *Computer Mathematics with Applications*, 9, 149-183.

This work has to be continued along several lines among which the study (interest) of other fuzzy implications. Moreover, the case of conditions