Semantics of quotient operators in fuzzy relational databases(*)

Didier Dubois – Henri Prade

Institut de Recherche en Informatique de Toulouse (I.R.I.T.) – C.N.R.S.
Université Paul Sabatier, 118 route de Narbonne
31062 Toulouse Cedex – France
Email: {dubois, prade}@irit.irit.fr – Fax: (+33) 61.55.62.39

Abstract

The quotient operator plays an important role for query evaluation in relational databases. Its extension to fuzzy databases raises the question of the intended interpretation of the intermediary degrees attached to the items in the fuzzy relations. This note points out that the fuzzy quotient operator should be defined in slightly different ways depending on the possible interpretations of the degrees: levels of satisfaction of a gradual property, levels of importance of a required property, uncertainty pertaining to the membership of an element to a subset.

(*) This short paper is a revised and slightly expanded version of a note entitled "Quotient operators in fuzzy relational databases" presented at the Second European Congress on Intelligent Techniques and Soft Computing (EUFIT'94), Aachen, Germany, September 20-24, 1994 and which appears pp. 357-360 in the Proceedings.

1. Introduction

Quotient operations aim at finding out the sub-relation $R \div S$ of a relation R, containing sub-tuples of R which have for complements in R *all* the tuples of a relation S, as in the example below (used in Yager (1991)) where we look for the persons who have the skills numbered I, II and III altogether.

R:

Name	Skill-type
Jean	I
Jean	II
Jean	III
Barbara	I
Barbara	II
Debbie	I
Debbie	II
Debbie	III
Debbie	IV
Tina	II
Kathy	IV

÷ S:

Skill-type
I
II
III

= R ÷ S:

Name
Jean
Debbie

In mathematical terms the quotient operation is defined by

$$R \div S = \{t, \forall\, u \in S, (t,u) \in R\} \tag{1}$$

where u is a tuple of S and t a sub-tuple of R such that (t,u) is a tuple of R. Viewing the relation R as inducing a multiple-valued mapping Γ which associates to u the set $\Gamma(u) = \{t, (t,u) \in R\}$, $R \div S$ is nothing but the *lower image* of S by Γ. In terms of Γ, (1) can be written equivalently as

$$R \div S = \{t, S \subseteq \Gamma^{-1}(t)\} \text{ with } \Gamma^{-1}(t) = \{u, (t,u) \in R\} \tag{2}$$

or as

$$R \div S = \bigcap_{u \in S} \Gamma(u). \tag{3}$$

The upper image is defined by changing '$\forall$' into '$\exists$' in (1), namely $\{t, \exists\, u \in S, (t,u) \in R\} = \{t, S \cap \Gamma^{-1}(t) \neq \emptyset\} = \bigcup_{u \in S} \Gamma(u)$. It corresponds in terms of database operations to an equijoin followed by a projection. In the above example, this yields the unary relation <Name>: {Jean, Barbara, Debbie, Tina}.

2. Fuzzy Quotient Operations

Upper and lower images have been extended to fuzzy relations; see Dubois and Prade (1992). Let us assume now that R and S are fuzzy relations, i.e., that their tuples are weighted by a number between 0 and 1. Then a natural extension of the quotient operation (1)-(2)-(3) is of the form

$$\mu_{R \div S}(t) = \inf_u \mu_S(u) \to \mu_R(t,u) \tag{4}$$

where $\mu_A(a)$ denotes the weight associated with the tuple a in the relation A, and where $\to$ denotes some multiple-valued logic implication. The use of generalized lower images for defining quotient operations in databases was first suggested in (Dubois and Prade, 1994). Expression (4) is in fact one of the fuzzy relational products introduced by Bandler and Kohout (1980) whose properties have been more recently studied by De Baets and Kerre (1993).

There are three main families of fuzzy implication connectives (see, e.g., Dubois and Prade, 1991): the S-implications, the R-implications and the reciprocals of R-implications. S-implications are of the form $a \to b = n(a * n(b))$ where n is an involutive negation operation, and $*$ is a conjunction operation modelled by a triangular norm. R-implications are obtained by residuation, namely $a \to b = \sup\{c \in [0,1], a * c \le b\}$. While $a \to b = n(b) \to n(a)$ holds for S-implications, it is not generally the case for R-implications where the reciprocal $n(b) \to n(a)$ gives birth to a third kind of implications. In an ordinal setting where we use $*$ = min and the ordering-reversing operation $n(a) = 1 - a$, we respectively get Dienes implication $a \to b = \max(1 - a, b)$ as a S-implication, Gödel implication as a R-implication, namely $a \to b = 1$ if $a \le b$, $a \to b = b$ if $a > b$, and $n(b) \to n(a) = 1$ if $a \le b$, $n(b) \to n(a) = 1 - a$ if $a > b$. When operations other than min and max are meaningful in practice on the scale [0,1], we may take $*$ as the product, or use $a * b = \max(0, a + b - 1)$; in this latter case the three implications reduce to Lukasiewicz implication.

The proper choice of the implication in (4) depends on the intended meaning of the weights. We distinguish in the following between three possible intended meanings:

– the weight expresses to what extent a gradual property is fulfilled, e.g., to what extent Jean has the skill I (if we consider that there exists several levels in the mastering of skill I);

– the weight expresses uncertainty, e.g., μ_R(Jean, skill I) = 0.8 then may express that we are not completely sure that Jean has skill I. The skill is then an all-or-nothing notion and 0.8 is a certainty degree;

– the weight express importance, e.g., we are looking for people having skills I, II and III, but for instance it is more important to have skill I than skill II (i.e., μ_S(skill I) > μ_S(skill II)), which is of interest to express if the database would contain nobody having both skill I and II (then we shall be glad to retrieve people having at least skill I which is the more important).

It is clear that in practice $\mu_S(u)$ might be either a degree of fulfilment or a degree of importance, since it seems to make no sense to be uncertain on the elements of the set S, which in our example gathers the *required* skills. Besides, $\mu_R(t,u)$ might be either a degree of fulfilment or a degree of (un)certainty, but not a degree of importance (since R corresponds to the available information, and S to the requirement). This is why we are going to consider the three following cases to be the most common ones (for simplicity we do not put degrees of fulfilment and degrees of uncertainty together).

	$\mu_R(t,u)$	$\mu_S(u)$
Case 1	degree of fulfilment	degree of fulfilment
Case 2	degree of uncertainty	{0,1}-valued degree
Case 3	degree of fulfilment (or {0,1}-valued degree)	degree of importance

Case 1: When dealing with degrees of fulfilment, it seems natural to consider that a tuple completely satisfies the query, i.e., belongs to R ÷ S with the degree 1 as soon as the degree of required fulfilment $\mu_S(u)$ is less or equal to the degree of fulfilment $\mu_R(t,u)$ of the stored item. Formally, we should have

$$\mu_{R \div S}(t) = 1 \Leftrightarrow \forall u,\ \mu_S(u) \leq \mu_R(t,u). \tag{5}$$

Otherwise, if $\mu_S(u) > \mu_R(t,u)$, i.e., the tuple (t,u) satisfies the property to a degree less than the required one, the most natural proposal is to keep for t the same weight as for (t,u), which leads to $\mu_{R \div S}(t) = \mu_R(t,u)$ which evaluates the level of performance of the tuple. In other words, we should use Gödel implication in (4): $a \rightarrow b = 1$ if $a \leq b$ and $a \rightarrow b = b$ otherwise. Indeed the only other alternative (in the ordinal setting(*)) would be the contraposition of Gödel implication that leads to $\mu_{R \div S}(t) = 1 - \mu_S(u)$ if $\mu_S(u) > \mu_R(t,u)$. But this would be strange: for instance if $\mu_S(u) = 1$ (i.e., we want the skill corresponding to u to be completely mastered in our example) and $\mu_R(t,u) = .9$ (t almost masters u) it would not be satisfactory to claim that $\mu_{R \div S}(t) = 0$ instead of .9. Provided that the *relative* level of satisfaction of the property by the tuple in R with respect to the level required in S is meaningful, we can use Goguen implication, a R-implication generated by $*$ = product and defined by $a \rightarrow b = 1$ if $a = 0$, $a \rightarrow b = \min(1, b / a)$ if $a \neq 0$, i.e.,

(*) If the membership grades belong to a scale which is only supposed to be a totally ordered set, the only available (symmetric) conjunction is minimum and the only possible implications are Gödel implication, its contraposition, and Dienes implication (used later in Case 3).

$$\mu_{R \div S}(t) = \min\left(1, \frac{\mu_R(t,u)}{\mu_S(u)}\right) \text{ if } \mu_S(u) \neq 0.$$

Case 2: In this case, only the tuples in R are weighted in terms of certainty. Such a weighted relation may be the result of a query to a fuzzy information database asking for items which necessarily satisfy some property (see, e.g., Dubois and Prade, 1988, Chapter 6). Since S is non-fuzzy, (4) can be written

$$\mu_{R \div S}(t) = \inf_{u \in S} \mu_R(t,u). \tag{6}$$

Here $\mu_R(t,u)$ is viewed as a necessity degree, the necessity that t satisfies u. So, (6) is in full agreement with the characteristic axiom of necessity measures (e.g., Dubois and Prade, 1988) which expresses that the certainty of a conjunction is the minimum of the certainty degrees attached to members of the conjunction.

Case 3: In this case the complete satisfaction of the query by a tuple seems to demand that all the requirements in S, whatever their importance, be included in the set of tuples with maximum fulfilment, i.e.,

$$\mu_{R \div S}(t) = 1 \Leftrightarrow (\forall u, \mu_S(u) > 0 \Rightarrow \mu_R(t,u) = 1). \tag{7}$$

Moreover, it seems natural that $\mu_{R \div S}(t) = 0$ *only if both* $\mu_S(u) = 1$ (the requirement has the maximum level of importance) and $\mu_R(t,u) = 0$ (the tuple does not at all fulfil the requirement). But, $\mu_{R \div S}(t)$ is allowed to be strictly positive if $\mu_R(t,u) = 0$ provided that $\mu_S(u) < 1$, i.e., the requirement is not completely important (since then the requirement can be forgotten to some extent). These desired behavior leads to define the quotient operation for this case by using Dienes implication $a \rightarrow b = \max(1 - a, b)$:

$$\mu_{R \div S}(t) = \inf_u \max(1 - \mu_S(u), \mu_R(t,u)) \tag{8}$$

where S is a normalized fuzzy set ($\exists u, \mu_S(u) = 1$) in order to have an appropriate scaling of the levels of importance. This expression can be viewed as a conjunction of the degrees of fulfilment $\mu_R(t,u)$ weighted by the levels of importance $\mu_S(u)$ (see Dubois and Prade, 1986).

By delocalizing the weights in (8), we can turn (8) into a (fuzzily) quantified conjunction, corresponding to an evaluation of a requirement that the tuples t are associated with 'at least k', or more generally 'most' elements in a non-fuzzy set S (rather than 'all' elements in S as in lower images). This can be done in the following way (see, e.g., Dubois, Prade and Testemale, 1988; Dubois and Prade, 1992):

i) let I be a fuzzy integer, i.e., a fuzzy subset of the set of integers {0, 1, 2, …} such that

$$\mu_I(0) = 1,\ \mu_I(i) \geq \mu_I(i + 1).$$

For instance, "at least k" will be modelled by $\mu_I(i) = 1$ if $0 \leq i \leq k$ and $\mu_I(i) = 0$ for $i \geq k + 1$. More generally, I will be a fuzzy set;

ii) rank-order the degree $\mu_R(t,u_i)$ decreasingly for $u_i \in S$ and $|S| = n$, i.e.,

$$\mu_R(t, u_{\sigma(1)}) \geq \mu_R(t, u_{\sigma(2)}) \geq \ldots \geq \mu_R(t, u_{\sigma(n)})$$

where σ is a permutation of {1, …, n};

iii) the quotient operation is then defined by

$$\mu_{R \div S(I)}(t) = \inf_{i=1,n} \max(1 - \mu_I(i), \mu_R(t, u_{\sigma(i)})). \tag{9}$$

When $\mu_I(i) = 1$ for $0 \leq i \leq n$, (9) reduces to $\mu_{R\div S}(t) = \mu_R(t, u_{\sigma(n)}) = \inf_{u\in S} \mu_R(t,u)$ as expected. When $\mu_I(1) = 1$ and $\mu_I(2) = \ldots = \mu_I(n) = 0$, (9) reduces to $\mu_R(t, u_{\sigma(1)}) = \sup_{u\in S} \mu_R(t,u)$, which corresponds to the upper image.

The expression (9) which involves the absolute quantifier I, can be easily modified for accommodating relative quantifiers like 'most', by changing $1 - \mu_I(i + 1)$ into $\mu_Q\left(\frac{i}{n}\right)$ for i = 0, n – 1 and $\mu_Q(1) = 1$, i.e.,

$$\mu_{R\div S(Q)}(t) = \inf_{i=1,n} \max(\mu_Q\left(\frac{i-1}{n}\right), \mu_R(t, u_{\sigma(i)})). \tag{10}$$

Again if Q means 100 %, i.e., 'all', $\mu_Q(1) = 1$ and $\mu_Q\left(\frac{k}{n}\right) = 0$ if $k < n$, then (10) reduces to $\mu_{R\div S}(t) = \inf_{u\in S} \mu_R(t,u)$.

What has been computed here is an *ordered weighted minimum* operation, or if we prefer, the median of the set of numbers made by the $\mu_R(t,u_{\sigma(i)})$'s and the $1 - \mu_I(i)$'s. See Dubois, Prade and Testemale (1988).

3. Concluding Remarks

This note has proposed to define fuzzy quotient operators for relational databases, in terms of lower images of fuzzy mappings. Moreover, the complete definition of the appropriate quotient operator requires a clear understanding of the intended semantics conveyed by the weights. The extension of the quotient operator to fuzzy relational databases has been recently investigated by various authors. Yager (1991) has proposed an approach based on Ordered Weighted Averages (OWA) to extend the quotient operator to a fuzzy database where each tuple of a relation is weighted by a number between 0 and 1. His motivations were mainly to soften the universal quantifier 'for all', as done at the end of the preceding section, but using an averaging operation rather than a median. Mouaddib (1993) also proposes a new definition of fuzzy quotient operator based on an extension of its classical decomposition into intersection, projection and selection operators. His proposal corresponds to state that $\mu_{R \div S}(t) = \inf_{u \in S} \mu^*_R(t,u)$ where $\mu^*_R(t,u)$ is the result of the modification of $\mu_R(t,u)$ for taking into account the fuzzy semantics of the attribute values appearing in R and S (in terms of approximate synonymy). Bosc and Pivert (1994) follow the lower image definition, suggested by Dubois and Prade (1994) and discussed here, and use Dienes' implication in (4) in a SQL-like environment.

References

Bandler W. and L.J. Kohout (1980). Fuzzy relational products as a tool for analysis and synthesis of the behavior of complex natural and artificial systems. In: Fuzzy Sets — Theory and Applications to Policy Analysis and Information Systems (P.P. Wang, S.K. Chang, eds.), Plenum Press, New York, 341-367.

Bosc P. and O. Pivert (1994). Quelques aspects de la sémantique et des équivalences de requêtes dans le langage SQLf. Journées Bases de Données Avancées, Clermont-Ferrand, Aug. 30-Sept. 2, 1994.

De Baets B. and E. Kerre (1993). Fuzzy relational compositions. Fuzzy Sets and Systems, 60, 109-120.

Dubois D. and H. Prade (1986). Weighted minimum and maximum operations in fuzzy set theory. Information Sciences, 39, 205-210.

Dubois D. and H. Prade (with the collaboration of H. Farreny, R. Martin-Clouaire and C. Testemale) (1988). Possibility Theory — An Approach to Computerized Processing of Uncertainty. Plenum Press, New York.

Dubois D. and H. Prade (1991). Fuzzy sets in approximate reasoning — Part 1: Inference with possibility distributions. Fuzzy Sets and Systems, 40, 143-202.

Dubois D. and H. Prade (1992). Upper and lower images of a fuzzy set induced by a fuzzy relation: application to fuzzy inference and diagnosis. Information Sciences, 64, 203-232.

Dubois D. and H. Prade (1994). A note on the quotient operator in fuzzy relational databases. BUSEFAL (I.R.I.T., Univ. P. Sabatier, Toulouse, France), 58, 23-34.

Dubois D., H. Prade and C. Testemale (1988). Weighted fuzzy pattern matching. Fuzzy Sets and Systems, 28, 313-331.

Mouaddib N. (1993). The nuanced relational division. Proc. of the 2nd IEEE Inter. Conf. on Fuzzy Systems (FUZZ-IEEE'93), San Francisco, CA, March 28-April 1st, Vol. II, 1419-1424.

Yager R.R. (1991). Fuzzy quotient operators for fuzzy relational data bases. In: Fuzzy Engineering toward Human Friendly Systems, Vol. 1 (Proc. of the Inter. Fuzzy Engineering Symp. (IFES'91), Yokohama, Japan, Nov. 13-15, 1991) (T. Terano, M. Sugeno, M. Mukaidono and K. Shigemasu, eds.), 289-296. Available from IOS Press, Amsterdam.