A Proposition for Implementation of priority Queries into FSQL

Aleksandar Takači

Faculty of Technology, University of Novi Sad
Bulevar Cara Lazara 1, 21000 Novi Sad, Serbia and Montenegro
E-mail: atakaci@infosky.net

Abstract: The concept of priority is often used in real time systems. In this paper a system called pFCSP is given that captures and implements the concept of priority. Using this system we can broaden the query language FSQL with queries that can handle priority. FSQL is a query language for fuzzy relational databases (FRDB). In our model priority reasoning which has theoretical background is incorporated giving us yet another dimension in knowledge acquisition through data mining.

Keywords: priority t-norm, fuzzy relational databases, FSQL, pFCSP

1 Introduction

Priority is most often viewed as the importance level of an object. The concept of priority is often used in real time systems.

FSQL is a query language which implements fuzzy queries and it is an upgrade of standard SQL. Most often it is implements on fuzzy databases and supports fuzzy notions i.e. high, low, large small etc. (see [2, 9])

In this paper a proposition for including priority into FSQL will be given. The theoretic background will be the axiomatic framework of pFCSP's (prioritized fuzzy constraint satisfaction problems). pFCSP's were introduced by Dubois, and an axiomatic framework was given in [7]. pFCSP is actually a FCSP (fuzzy constraint satisfaction problem) in which the notion of priority is introduced. One of the key factors in that implementation are priority t-norms. They are introduced in such a way that the smallest value (usually the value with the biggest priority) has the largest impact on the result given by priority t-norm.

We will give an axiomatic framework for pFCSP and also some systems which satisfy this framework thus justifying its cause.

Finally, we will show what type of queries can be included into FSQL and sketch the implementation idea avoiding technicalities.

2 Priority t-norms

First, we will give a definition of t-norms. For more details on t-norms see [4].

Definition 1 *A mapping* $T : [0,1]^2 \Rightarrow [0,1]$ *is called a t-norm if the following conditions are satisfied for all* $x, y, z \in [0,1]$*:*
(T1) $T(x,y) = T(y,x)$
(T2) $T(x, T(y,z)) = T(T(x,y), z)$
(T3) if $y \leq z$ *then* $T(x,y) \leq T(x,z)$
(T4) $T(x,1) = x$.

The four basic t-norms are:
$T_M(x,y) = \min(x,y)$
$T_P(x,y) = xy$
$T_L(x,y) = \max(x+y-1, 0)$
$T_D(x,y) = \begin{cases} 0, & \text{if}(x,y) \in [0.1[; \\ \min(x,y) & \text{otherwise.} \end{cases}$

Since every t-norm is associative (condition (T3)) it can easily be extended to n-ary operation:

$$T(x_1, x_2, \ldots, x_n) = T(x_1, T(x_2, \ldots T(x_{n-1}, x_n)) \ldots)$$

If $x_1 = x_2 = \cdots = x_n = x$ we denote $T(x, \cdots, x) = x_T^{(n)}$.

Now, we will give a short definition of copulas. For more details on copulas see [8].

Definition 2 *A two-dimensional copula (or briefly 2-copula or simply copula) is a binary operation* C *on the unit interval* $[0,1]$*, i.e., a function* $C : [0,1]^2 \Rightarrow [0,1]$*, such that for all* $x \leq x^*$ *and* $y \leq y^*$ *we have:*
(CO1) $C(x,y) + C(x^*, y^*) \geq C(x, y^*) + C(x^*, y)$
(CO2)$C(x,0) = C(0,x) = 0$
(CO3) $C(x,1) = C(1,x) = x$

In to order to introduce priority into FCSP (fuzzy constraint satisfaction problem) priority t-norms are used. When it is necessary to satisfy all the constrains a t-norm is used to aggregate the degree of satisfaction of each one of the constrains. PFCSP (prioritized fuzzy constraint satisfaction problem) is designed in such a way that the constraint with the largest priority is more likely to have a smaller value. Definition of priority t-norms is such that the increase of the smallest value (most often the value of the constraint with the highest priority) has a bigger impact i.e. gives larger values of the general satisfaction degree than an increase on any other components. More details on PFCSP, axioms, instantiation and validation is given in [7]. Priority t-norms are defined in the next definition.

Definition 3 *A t-norm is called a priority t-norm if the following condition holds for all* $a_1 \leq a_2$ *in* $[0,1]$ *and* δ *such that* $a_2 + \delta \leq 1$*:*

$$T(a_1 + \delta, a_2) \geq T(a_1, a_2 + \delta)$$

Proposition. Three basic t-norms, T_M, T_P and T_L are priority t-norms.

Proposition. The fourth basic t-norm (non-continuous), T_D is not a priority t-norm.

The following theorem gives a characterization of priority t-norms.

Theorem 1 *The following propositions are equivalent:*

(i) T is a continuous priority t-norm

(ii) T is an ordinal sum of Archimedean associative copulas, T_α, $\alpha \in A$, where A is the index set.

3 pFCSP - Prioritized Fuzzy Constraint Satisfaction Problem

Definition 4 *A fuzzy constraint satisfaction problem (FCSP) is defined as a 3-tuple (X, D, C^f) where:*

1. *$X = \{x_i | i = 1, 2, \cdots, n\}$ is a set of variables.*
2. *$D = \{d_i | i = 1, 2, \cdots, n\}$ is a finite set of domains. Each domain d_i is a finite set containing the possible values for the corresponding variable x_i in X.*
3. *C^f is a set of fuzzy constrains. That is,*

$$C^f = \{R^f | \mu_{R_i^f} : (\prod_{x_j \in var(R_i^f)} d_j) \Rightarrow [0,1], i = 1, 2, \cdots n\}$$

Definition 5 *A label of a variable x is an assignment of a value of the variable, denoted as v_x. A compound label $v_{x'}$ of all variables in the set $X' = x'_1, x'_2, \cdots, x'_n$ is a simultaneous assignment of all variables in the set X', that is,*

$$v_{X'} = (v'_{x_1}, v'_{x_2}, \cdots, v'_{x_n}).$$

The membership degree of each constraint indicates the local degree to witch the constraint is locally satisfied with a compound label. In order to obtain the global satisfaction degree local degrees are aggregated.

Definition 6 *In a FCSP (X, D, C^f), given a compound label v_x of all variables X, the global satisfaction degree for the compound label v_x is defined as*

$$\alpha(v_X) = \oplus\{\mu_{R_i^f}(v_{var(R^f)}) | R^f \in C^f\}$$

where $\oplus$ is an aggregation operator on the unit interval. A solution of FCSP is a compound label v_X such that

$$\alpha(v_X) \geq \alpha_0$$

where α_0 is called a solution threshold which is usually predetermined.

If all constrains in FCSP have to be satisfied a t-norm is used for $\oplus$. If at least one of the constrains has to be satisfied a t-conorm is used.

In order to introduce the concept of priority into FCSP the following axiomatic framework is given:

Definition 7 *When given (X, D, C^f, ρ), where $\rho : R^f \Rightarrow [0, \infty[$ and a compound label v_x of all variables in X, and $\bigoplus : [0,1]^n \Rightarrow [0,1]$, $g : [0, \infty[\times[0,1] \Rightarrow [0,1]$ and a satisfaction degree $\alpha_\rho(v_X)$ which is calculated in the following way:*

$$\alpha_\rho(v_X) = \bigoplus\{g(\rho(R^f), \mu_{R_i^f}(v_{var(R^f)}))|R^f \in C^f\},$$

it is a PFCSP if the following axioms are satisfied:

1. *If for the fuzzy constraint R^f_{max}, $\rho_{max} = \rho_{max}(R^f_{max}) = \max\{\rho(R^f)|R^f \in C^f\}$ then*

$$\mu_{var(R^f_{max})} = 0 \Rightarrow \alpha_\rho(v_X) = 0$$

2. *If $\exists\rho_0 \in [0,1]$, $R^f \in C^f$, $\rho(R^f) = \rho_0$, then*

$$\alpha(v_X) = \oplus\{\mu_{R_i^f}(v_{var(R^f)})|R^f \in C^f\}$$

   where $\oplus$ is a t-norm.

3. *For R_i^f, $R_j^f \in C^f$, suppose $\rho(R_i^f) \geq \rho(R_j^f)$, $\delta > 0$ and there are two different compound labels v_X and v'_X such that*

   - *when $R^f \neq R_i^f$ and $R^f \neq R_j^f$ then we have $\mu_{R^f}(v_{var(R^f)}) = \mu_{R^f}(v'_{var(R^f)})$*
   - *when $R^f = R_i^f$ then we have $\mu_{R^f}(v_{var(R^f)}) = \mu_{R^f}(v'_{var(R^f)}) + \delta$*
   - *when $R^f = R_j^f$ then we have $\mu_{R^f}(v'_{var(R^f)}) = \mu_{R^f}(v_{var(R^f)}) + \delta$*

   Then the following property holds:

$$\alpha(v_X) \geq \alpha(v'_X).$$

4. *For two different compound labels* v_X *and* v'_X *such that* $\forall R^f \in C^f$

$$\mu_{R^f}(v_{var(R^f)}) \geq \mu_{R^f}(v'_{var(R^f)})$$

then the following holds

$$\alpha(v_X) \geq \alpha(v'_X)$$

5. *If there exists a compound label such that* $\forall R^f \in C^f,\ \mu_{R^f}(v_{var(R^f)}) = 1$ *then* $\alpha_\rho(v_X) = 1$.

The first axiom states that if the constraint with the maximum priority has a zero value of the local satisfaction degree then the global satisfaction degree should be also zero.

The second axiom states that if all priorities are equal PFCSP should be a FCSP.

The third axiom is the most important one. It captures the notion of priority i.e. if one constraint has a larger priority then, the increase of the value on that constraint should result in a bigger increase of the global satisfaction degree then when the value with the smaller priority has the same increase. This rather complicated explanation is the result of the if clause in the third axiom which will be removed later in the paper yielding to new pFCSP systems. The fourth axiom is the monotonicity property, and the fifth is the upper boundary condition.

In order to obtain an actual PFCSP we will introduce several types of operators.

Definition 8 *An operator* $\oslash : [0,\infty[\times]0,\infty[\Rightarrow [0,1]$ *is called a general division operator if the following conditions are satisfied:*

- $\forall a \in]0,\infty[,\ a \oslash a = 1$
- $\forall a \in]0,\infty[,\ 0 \oslash a = 0$
- $\forall a_1,\ a'_1 \in [0,\infty[,\ a_2 \in (0,\infty),\ a_1 \leq a'_1 \Rightarrow a_1 \oslash a_2 \leq a'_1 \oslash a_2$
- $\forall a_1, \in [0,\infty[,\ a_2,\ a'_2 \in (0,\infty),\ a_2 \leq a'_2 \Rightarrow a_1 \oslash a_2 \geq a_1 \oslash a'_2$

It is easy to see that the classical division $a_1 \oslash a_2 = \dfrac{a_1}{a_2}$ satisfies the previous conditions.

Definition 9 *An operator* $\diamond : [0,1] \times [0,1] \Rightarrow [0,1]$ *is called a priority operator if the following conditions are satisfied:*

- $\forall a \in [0,1],\ 1 \diamond a = a$
- $\forall a \in [0,1],\ 0 \diamond a = 1$

- $\forall a_1,\ a_2,\ a_2' \in [0,1],\ a_2 \leq a_2' \Rightarrow a_1 \diamond a_2 \leq a_1 \diamond a_2'$
- $\forall a_1,\ a_1',\ a_2 \in [0,1],\ a_1 \leq a_1' \Rightarrow a_1 \diamond a_2 \geq a_1' \diamond a_2$

Actually these conditions are designed in such a way that for any every t-conorm S, the operator $a_1 \diamond a_2 = S(1-a_1, a_2)$ is a priority operator.

Definition 10 *A t-norm is called a priority t-norm if the following condition holds for all $a_1 \leq a_2$ in $[0,1]$ and δ such that $a_2 + \delta \leq 1$:*

$$T(a_1+\delta, a_2) \geq T(a_1, a_2+\delta)$$

The characterization of priority t-norms is given in the following section. The most common priority t-norm is the $T_M(a_1, a_2) = \min(a_1, a_2)$.

Now we will give a concrete pFCSP which satisfies previously given axioms.

Theorem 2 *Let (X, D, C^f, ρ) be defined in the following way:*

1. *$X = \{x_i | i = 1, 2, \cdots, n\}$ is a finite set of variables.*
2. *$D = \{d_i | i = 1, 2, \cdots, n\}$ is a finite set of domains. Each domain d_i is a set containing the possible values for the corresponding variable x_i in X.*
3. *C^f is a set of fuzzy constrains. That is,*

$$C^f = \{R^f | \mu_{R_i^f} : (\prod_{x_j \in var(R_i^f)} d_j) \Rightarrow [0,1], i = 1, 2, \cdots n\}$$

4. *$\rho : [0, \infty[\Rightarrow [0,1]$*

Then if
is a priority t-norm then the following system is a PFCSP:

$$\alpha_\rho(v_X) = \bigoplus_L \{\frac{\rho(R^f)}{\rho_{max})} \diamond_P \mu_{R_i^f}(v_{var(R^f)}) | R^f \in C^f\},$$

where $\bigoplus_L(x,y) = T_L(x,y)$, and $\diamond_P(x,y) = S_P(1-x,y)$.

It is interesting that the system based on the two most common operators min and max does not satisfy is not a pFCSP which is a surprising result.

Proposition. The following system **does not satisfy** the third axiom given in this section:

$$\alpha_\rho(v_X) = \bigoplus_L \{\frac{\rho(R^f)}{\rho_{max})} \diamond_P \mu_{R_i^f}(v_{var(R^f)}) | R^f \in C^f\},$$

Where $\bigoplus_L(x,y) = \min(x,y)$, and $\diamond_P(x,y) = \max(1-x,y)$.

Now we will try to expand our set of pFCSP with some other combinations of operators. The Frank t-norms are given by the following formula:

$$T_\lambda^F(x,y) = \begin{cases} T_M(x,y), & \text{if}\lambda = 0; \\ T_P(x,y), & \text{if}\lambda = 1; \\ T_L(x,y), & \text{if}\lambda = \infty; \\ \log_\lambda(1 + \frac{(\lambda^x-1)(\lambda^y-1)}{\lambda-1}) & \lambda \in]0,1[\cup]1,\infty[. \end{cases}$$

Also, Frank t-conorms are given by the following formula:

$$S_\lambda^F(x,y) = \begin{cases} S_M(x,y), & \text{if}\lambda = 0; \\ S_P(x,y), & \text{if}\lambda = 1; \\ S_L(x,y), & \text{if}\lambda = \infty; \\ 1 - \log_\lambda(1 + \frac{(\lambda^{1-x}-1)(\lambda^{1-y}-1)}{\lambda-1}) & \lambda \in]0,1[\cup]1,\infty[. \end{cases}$$

Theorem 3 *The following system satisfies axioms 1,2,4,5 and the changed third axiom:*

$$\alpha_\rho(v_X) = \bigoplus_L \{ \frac{\rho(R^f)}{\rho_{max})} \diamond_P \mu_{R_i^f}(v_{var(R^f)}) | R^f \in C^f \},$$

if $\bigoplus_L(x,y) = T_\lambda^F(x,y)$, *and* $\diamond_P(x,y) = S_\lambda^F(1-x,y)$, *for* $\lambda > 1$.

4 FSQL

4.1 Fuzzy Relational Databases and FSQL

In order to expand classical relational databases to model impression fuzzy relational databases (FRDB) are introduced. Many models dating up to 1982 of FRDB have been studied.

Classical databases store only precise information. For example if a persons height is not known but it is known that a person is a tall person classical databases cannot store this information. On the other hand FRDB through the concept of linguistic labels can have the value *tall* for attribute *height*. Besides storing imprecise values FRDB through FSQL can answer a broader set of queries. Take for example queries: *I need students of medium height that have good marks in PE* or *Give me people with average salaries and small housing capacity* etc.

Now we will see how data is stored in FRDB. In the model introduced in [2] two types of data are introduced:

1) Traditional database - traditional data with crisp and/or fuzzy attributes

2) Fuzzy meta knowledge base - information about attributes and their relations is stored in a relational format

Now we will give a short preview of FSQL. The main extensions of classical SQL are:

- linguistic labels: each label has a possibility distribution or a similarity relation with other labels. Mostly, trapezoidal possibility distributions are used.
- fuzzy comparators: fuzzy equal, fuzzy greater or lower etc. that can be either necessity or possibility comparators.
- fulfillment threshold: a value in the unit interval indicating that the condition should be fulfilled to a minimum degree
- fuzzy constants: unknown, undefined, trapezoidal values, approximately n, range [m,n] and null.

Now we will give an idea how to implement priority queries into FRDB.

4.2 Incorporating priority queries into FSQL

In the previous section we have introduced pFCSP who can handle priority. Also, the a concrete pFCSP which uses T_L and S_p or any Frank t-norm and t-conorm can easily be implemented. For example, our system can handle calculations of type: Is it better to have fair GPA and excellent physical condition or good GPA and good physical condition if you want to teach PE in a high school. This concept can be easily incorporated into FRDB. For example queries of type:

SELECT * FROM Students
WHERE GPA FEQ Good WITH priority HIGH
AND MathGrade GEQ 7 WITH priority MEDIUM
AND Sex EQ Male WITH priority LOW
SORT DESC

These type of queries add a new dimension to FRDB. They allow us to choose based not only on aggregated values of all attributes but also taking onto account the importance (priority) of a particular attribute. Someone would argue that instead of priority we can use crisp thresholds for each attributes giving higher thresholds to the more important attributes but this would lead to a crisp join of α cuts which could not lead us to the proper answer. In our model priority reasoning which has theoretical background is incorporated giving us yet another dimension in knowledge acquisition through data mining.

There are many technical and implementations details to be done in order for this concept to be implemented but we hope that they will be solved and we will have a pFSQL i.e. FSQL that can handle priority queries.

5 Acknowledgements

This paper was written with the support of the Serbian Ministry of Science and Technology project No. 1866 and CEEPUS SK-42 network.

References

[1] T. Calvo, A. Kolesarova, M. Komornikova, R. Mesiar, Agregation operators: Properties, Classes and Construction Methods , Fuzziness and Soft Computing 97, (2000) 1-103.

[2] J. Galando, M. C. Aranda, J. L. Caro, A. Guevara, A. Aguayo Applying fuzzy databases and FSQL to the managmnet of rural acomodation, Tourism Managment 24, (2003) 457-463.

[3] D. Dubois, E. Kerre, R. Mesiar, H, Prade, Fuzzy Interval Analysis, Fundamentals of Fuzzy Sets, Handbook Series of Fuzzy Sets Vol.1 (eds. D. Dubois and H. Prade), Kluwer, Dodrecht, (2000) 483-582.

[4] E. Klement, R. Mesiar, E. Pap, Triangular norms", Series: Trends in Logic, Kluwer Academic Publishers, , Vol. 8, Dordrecht, 2000.

[5] E. P. Klement, R. Mesiar, E. Pap, "Generated triangular norms", Kybernetika 36 (2000) 363-377.

[6] X. Luo, N. R. Jennigs, N. Shadbolt , H. Leung, J. H. Lee, A fuzzy constraint based model for bilateral multi isssue negotiations in semi competitive enviroments, Artificial Intelligence 148 (2003) 53-102.

[7] X. Luo, J. H. Lee, H. Leung, N. R. Jennigs , Prioritised fuzzy constraint satisfaction problems: axioms, instantiation and validation, Fuzzy sets and systems 136 (2003) 151-188.

[8] R. W. Nelsen, An introduction to copulas Springer Verlag, Berlin, 1999.

[9] H. Prade, C. Testemale, Fuzzy Realtional databases: Representational issues and reduction using similarity measures, Journal of the American Society of Informational Sciences 38(2), (1987), 118-126

[10] H. Schweizer , A. Sklar, Probabilistic metric spaces North Holand, New York, 1983.

[11] L. A. Zadeh "Fuzzy Sets", Information and Control 8 (1965) 338-353.