Towards a Unified Domain for Fuzzy Temporal Databases

M. C. Garrido
Junta de Andaluca, Spain

N. Marín and O. Pons
Dept. of Computer Science and Artificial Intelligence, University of Granada, ETSI Informática, Spain

Keywords: Fuzzy data, Temporal database, Fuzzy interval, Data manipulation.

Abstract: Temporal Databases (TDB) have as a primary aim to offer a common framework to those DB applications that need to store or handle temporal data of different nature or source, since they allow to unify the concept of time from the point of view of its meaning, its representation and its manipulation. At first sight, it may seem that incorporation of time to a DB is a direct and even simple task, but, on the contrary, it is a quite complex aim because time may be provided by different sources, with different granularities and meaning. The situation gets more complex when the time specification is not made in precise but in fuzzy terms, where together with the inherent problems of the time domain, we have to consider the imprecision factor. To deal with this problem, the first task to perform is to unify as much as possible the representation of time in order to be able to define the range and the semantics of the necessary operators to handle data of this type.

1 INTRODUCTION

1.1 Previous Concepts

Temporal Databases, in the widest sense, offer a common framework for all database applications that involve some temporal aspects when organizing data. These databases allow to unify the time concept from several points of view: the representation, the semantics and the manipulation. Database applications involving temporal data are not a new subject. In fact, they have been developed since the relational databases began to be used, but applications programmers were responsible for designing, representing, programming and managing the necessary temporal concepts.

Temporal Databases (from now on TDB) have partially solved the problem because they provide data types and operators for handling time. From the point of view of the real world, there exist two basic ways for associating temporal concepts to a fact:

1. Punctual facts: a fact is related to an only time mark that depends on the granularity and informs about the time when it happened. As instances, birthdays, the date for an order, an academic year, ...

2. Time periods: that are represented by a starting instant and an ending one, so the duration (or valid time) of the fact is implicit. Some examples are: [admission date, discharge date], [start contract date, end contract date], ...

This way of time interpretation is called valid time.

In the valid time relation EMP (see fig. 1) each tuple represents a version for the available information about an employee, and this version is valid only when used in the time interval [VST,VET]. The up-to-date version, also called valid tuple, is undefined-valued in the attribute VET (a special value).

Sometimes it is not possible for the user to give an exact but an imprecise starting/ending point for the validity period of a fact. This is the case, for example, when a patient does not exactly know when a concrete ailment or symptom started. In this case, the use of fuzzy sets theory is necessary for not missing such important information since fuzzy time values could be defined (Barro et al., 1994). This situation give...
rise to a large number of new problems (Bettini et al., 1998), and this paper is devoted to the definition of
time domain that allows the representation of different
fuzzy time specifications.

1.2 Previous Concepts on Fuzzy Sets

A fuzzy value is a fuzzy representation about the real
value of a property (attribute) when it is not precisely
known.

In this paper, according to Goguen’s Fuzzification
Principle (Goguen, 1967), we will call every fuzzy
set of the real line fuzzy quantity. A fuzzy number is a
particular case of a fuzzy quantity with the following
properties:

Definition 1.-

The fuzzy quantity A with membership function
\( \mu_A(x) \) is a fuzzy number (Dubois and Prade, 1987)
iff:
1. \( \forall \alpha \in [0,1], A_\alpha = \{ x \in R \mid \mu_A(x) \geq \alpha \} \) (\( \alpha \)-cuts of
A) is a convex set.
2. \( \mu_A(x) \) is an upper-semicontinuous function.
3. The support set of A, defined as \( \text{Supp}(A) = \{ x \in R \mid \mu_A(x) > 0 \} \), is a bounded set of R, where R is
the set of real numbers.

We will use \( \tilde{R} \) to denote the set of fuzzy numbers,
and \( h(A) \) to denote the height of the fuzzy number A.
For the sake of simplicity, we will use capital letters at
the beginning of the alphabet to represent fuzzy num-
bbers.

The interval \( [a_\alpha, b_\alpha] \) (see figure 2) is called the \( \alpha \)-
cut of A. So then, fuzzy numbers are fuzzy quantities
whose \( \alpha \)-cuts are closed and bounded intervals: \( A_\alpha = [a_\alpha, b_\alpha] \) with \( \alpha \in (0, 1] \).

If there is, at least, one point \( x \) verifying \( \mu_A(x) = 1 \)
we say that A is a normalized fuzzy number.

Sometimes, a trapezoidal shape is used to repre-
sent fuzzy values. This representation is very useful
as the fuzzy number is completely characterized by
four parameters \( (m_1, m_2, a, b) \) as shows figure 3 and
the height \( h(A) \) when the fuzzy value is not normal-
ized. We will call modal set all values in the interval
\[ m_1, m_2 \], i.e, the set \{ \( x \in \text{Supp}(A) \mid \forall y \in R, \mu_A(x) \geq \mu_A(y) \} \). The values a and b are called left and right
spreads, respectively.

In our approach, we will use trapezoidal and nor-
malized fuzzy values.

2 FUZZY TIME
REPRESENTATION

2.1 Imprecision Measure on Fuzzy
Values

As pointed out in the previous section, we are going
to translate fuzzy uncertainty into imprecision under
certain conditions. The most important of these con-
ditions is that the amount of information provided by
the fuzzy number remains equal before and after the
transformation. So then, the first step is to define an
information function for fuzzy numbers.

In (González et al., 1999) we propose an ax-
omatic definition of information, partially inspired in
the theory of generalized information given by Kampé
de Fériet (de Ferièt, 1973) and that can be related to
the precision indexes (Dubois and Prade, 1987) and
the specificity concept, introduced by Yager in (Yager,

Definition 1.-

Let \( \mathcal{D} \subseteq \tilde{R} \mid R \subseteq \mathcal{D} \); we say that the application I defined as:
I : D → [0, 1]

is an information on D if it verifies:

1. I(A) = 1, ∀ A ∈ R
2. ∀ A, B ∈ D | h(A) = h(B) and A ⊆ B → I(B) ≤ I(A).

The information about fuzzy numbers may depend on different factors, in particular, on imprecision and certainty (Chountas and Petrounias, 2000). We focus on general types of information related only to these two factors.

To compute a measure of the imprecision contained in a fuzzy number, we will consider a measure of the imprecision of its α-cuts, which are closed intervals on which the following function is defined:

∀ A ∈ R, fA(α) = \ \left\{ \begin{array}{ll} bα - aα & \text{if } α ≤ h(A) \\
0 & \text{otherwise} \end{array} \right.

From this imprecision function on the α-cuts, we define the total imprecision of a fuzzy value as a combination of the imprecision in every level α. When α = 0, we will consider that fA(0) is the length of the support set.

Definition 2.-

The imprecision of a fuzzy number is defined as follows:

f : R → R^+_0

∀ A ∈ R, f(A) = \int_0^{h(A)} fA(α) dα

That is, the imprecision function f coincides with the area below the membership function of the fuzzy value. Since we are considering that fuzzy time values are always normalized, then h(A) = 1.

2.2 Unified Domain for Temporal Data

In the introduction we have seen that, in classical TDB, the valid time is managed thanks to the extension of the tables schema by adding two new attributes (Clifford and Rao, 1987) (Elmasri and Wuu, 1990), the valid start time -VST- and the valid end time -VET- to determine the period of validity of the fact expressed by a tuple.

In this paper we are going to consider that the information provided by the VST and VET for the classical TDB is fuzzy, in the sense that we are not completely sure about when the current values of the tuple began to be valid.

The more immediate solution to this problem is to soften the VST and the VET in such a way that they may contain fuzzy dates represented by means of a fuzzy number. This means that, if we use the parametrical representation for fuzzy numbers, we need to store four values for the VST and four values for the VET, as shown in figure 4. Since the meaning of the attributes VST and VET is the period of time during which the values of a tuple are valid, it is more convenient to summarize the information given by the two fuzzy attributes in an only but fuzzy interval (from now on FVP or fuzzy validity period). This situation can be represented by the trapezoidal fuzzy set shown in figure 5 which incorporates the semantics of our problem. As can be seen in such figure, the left and right sides of the interval is the part that reflects the imprecision about the starting and ending time point of the validity time of the facts associated.

This representation has the advantage that, not only periods of time, but fuzzy dates can also be represented in a unified way. Think that a parametrical representation as (m,m,a,b) represents a central time point with some imprecision at both sides, what is interpreted as a fuzzy date.

The problem now is that the imprecision provided by the two fuzzy dates must be translated to the interval that summarizes the considered period of time. That is, all the imprecision of the starting date must be converted in the imprecision of the left side of the interval and, in the same way, all the imprecision of the ending date must be converted in the imprecision of the right side of the interval.

If we consider that a way to measure the imprecision of a fuzzy set is to compute its area, the problem we have in hands is a matter of geometrical computation.
The posed problem is shown in a graphical way in figure 6.

![Figure 6: Transformation of two fuzzy dates into a fuzzy period preserving imprecision.](image)

The resulting fuzzy interval is obtained by means of the equality $S_1 = S_2$ that obliges to maintain the same amount of imprecision after the transformation is performed.

$$S_1 = S_2 \implies \frac{(d_2 + b_2) - (d_2 - a_2)}{2} = \frac{d_1 - (d_1 - a)}{2}$$

If we assume that the data associated to this time specification are precisely known from $(d_s + b_s)$ to $(d_e - a_e)$, then $d_1 = d_2 + b_2$ and both terms become equal and $d_1 - a = d_2 - a_s$, as shown in figure 6. The same substitution should be made to obtain the right part of the interval.

As it was explained in section 1.2, it is quite easy to represent a fuzzy interval with this characteristics since only four parameters need to be stored in order to specify it. In (Medina et al., 1994) (Medina et al., 1995) is presented a generalized model of fuzzy DB that supports this representation for fuzzy data and the corresponding implementation in a classical relational DB system (Oracle).

3 CONCLUSIONS AND FUTURE WORK

In this paper we have shown how to represent different time specifications in a unified way. The representation considered is the fuzzy interval, which results very suitable for both precise and imprecise time points and periods when the time is interpreted as valid time. For the case that two fuzzy dates are provided by the user, it is necessary to perform a transformation to convert this original time information into a fuzzy interval that preserves the imprecision involved.

ACKNOWLEDGEMENTS

This work has been partially supported by research projects TIN2008-02066/TIC and P07-TIC-03175.

REFERENCES


