A Fuzzy Model for Representing Uncertain, Subjective and Vague Temporal Knowledge in Ontologies

Gábor Nagypál and Boris Motik

FZI Research Center for Information Technologies
at the University of Karlsruhe*
Haid-und-Neu-Str. 10-14
76131 Karlsruhe, Germany
{nagypal,motik}@fzi.de
http://www.fzi.de

Abstract. Time modeling is a crucial feature in many application domains. However, temporal information often is not crisp, but is uncertain, subjective and vague. This is particularly true when representing historical information, as historical accounts are inherently imprecise. Similarly, we conjecture that in the Semantic Web representing uncertain temporal information will be a common requirement. Hence, existing approaches for temporal modeling based on crisp representation of time cannot be applied to these advanced modeling tasks. To overcome these difficulties, in this paper we present fuzzy interval-based temporal model capable of representing imprecise temporal knowledge. Our approach naturally subsumes existing crisp temporal models, i.e. crisp temporal relationships are intuitively represented in our system. Apart from presenting the fuzzy temporal model, we discuss how this model is integrated with the ontology model to allow annotating ontology definitions with time specifications.

1 Introduction

Time modeling is a crucial feature in many application domains, such as medicine, history, criminal and financial applications. Its importance is shown by the numerous works in the area of temporal databases [1, 2] and temporal reasoning [3]. Our experience from the EU-IST sponsored VICODI project[1] supports this claim fully. The main aim of this project is to develop an ontology of European history used for semantical indexing of historical documents. The goal of the VICODI system is to demonstrate benefits of semantics by improving searching and navigation in historical databases. We note that the system is not intended to be used by the general public, but is aimed at power-users, such as historians and

* This work was partially supported by the EU in the framework of the VICODI project (EU-IST-2001-37534)

[1] http://www.vicodi.org/

librarians. Therefore, the expressivity and accuracy of the modeling capabilities are crucial to the success of the project.

It is quite obvious that time modeling is a fundamental issue for modeling historical information, since almost every historical statement is time dependent. Moreover, we identified several specific features that make capturing this information a challenge. For one, time information in history is often uncertain or ill-defined. It is usually extracted from historical documents written in an imprecise and inherently vague style. Even worse, important documents are often missing or contain contradictory information, so temporal information about historical events is uncertain. Apart from the uncertainty of temporal information, historical events are often abstract, so their definition is inherently subjective. For example, we have found it impossible to precisely define the time span of the 'Middle Ages'.

To the best knowledge of authors, there is no existing approach for temporal modeling which is capable of capturing such uncertain, subjective and vague temporal information. The vast majority of related work concentrates on modeling crisp and definite temporal information (for which we use synonyms 'traditional' or 'classical' in this paper). Moreover, existing approaches for modeling imprecise temporal information focus on handling either uncertainty or subjectivity, but not both. Therefore, we decided to propose a new model capable of fulfilling our requirements. We designed our model based on a rigorous requirements analysis. Although this analysis is driven by our application scenario, we believe that the temporal model is general enough to be applied in other settings demanding advanced temporal modeling as well. Also, temporal model alone is not worth much – a mechanism for embedding it into an ontology model is needed. Hence, in this paper we present an approach for integrating two orthogonal models, namely the temporal and the ontology model, into a unified modeling framework.

This paper is organized as follows. Section 2 introduces the requirements posed by temporal knowledge in history. Section 3 formally introduces our temporal model. Section 4 shows how to extend the classical temporal interval relations introduced by J. F. Allen [4] to our model. Section 5 discusses various properties of the new temporal model, and Section 6 discusses how to integrate the temporal model in ontologies. Section 7 analyzes related work and Section 8 concludes the paper and provides some outlook about future work.

2 Requirements

In this section we discuss requirements we gathered in the course of the VICODI project, based on which we design our temporal model. Although these requirements come from the domain of modeling historical information, they are still valid in many other application settings.

2.1 Unique Features of Historical Temporal Knowledge

History, as a human science, is different from natural sciences, mainly because information sources are historical documents written by people using vague natural languages, using vague and subjective concepts, such as 'revolution' or 'golden age'. Therefore, the concepts found in these sources, as well as their temporal specifications, are often imprecise.

Based on the accounts of history experts in the VICODI project, debate and disagreement over temporal specifications in history is rather the norm than an exception. This makes modelling this type of knowledge extremely difficult. In this subsection we discuss the nature of these challenges and give some examples.

Uncertainty. Sometimes information about a historical event can only be deduced from documents reporting about related events. Often several documents state contradictory facts about some event. As an example consider Stalin's birthdate. Officially for the USSR it is 21 December 1879 but according to church records his birth was registered in 6 December 1878. Historians are in disagreement over which time specification is the right one.

In such a case we say that the temporal specification of the event is uncertain. In natural language such uncertain temporal specifications are often written informally as ? - 1640 (meaning that the beginning time of the event is unknown) or as ca. 1801 (meaning circa, or around this year).

Subjectivity. Many historical events are not exactly defined, but are subjective. For example, the 'early renaissance', 'Russian revolution' or 'industrial revolution' do not have a clear definition, so it is impossible to clearly state exactly when these events occurred. In this case it is intuitive for historians to talk about 'beginning or end phases', 'process, development and core periods' or of 'transition periods', which clearly indicates that the traditional model of having temporal intervals with definite start and end points does not meet the reality in this case.

Vagueness. Historical time specifications are given at different granularity (years, months, days) and are often defined fuzzily (early morning, spring etc.). Hence, the temporal specification is not known precisely, but is vague. The reader may also note that any temporal specification made in a natural language will become vague if we refine the granularity of temporal axis sufficiently.

There are also events exhibiting a combination of the afore mentioned properties, so the temporal model should be capable of representing all of them in a unified manner.

2.2 Temporal Relations

Since the goal of the VICODI project is to allow creating powerful temporal queries, expressive relations among temporal primitives are of paramount importance. The model should at least be able to represent the temporal interval relations defined by Allen in [4]. Apart from these relations, we have found out that the relation `intersects`, checking whether two time intervals have a common point, is crucial in the historical context. Hence, Table 1 summarizes all of the operations required. For an interval i, i^- denotes the starting point of the interval, and i^+ denotes the ending point of the interval.

It is a natural requirement that our temporal model should, if applied to traditional temporal specifications, yield the same results as in classical temporal models. With other words, our approach should naturally subsume the classical case.

2.3 Temporal Specifications vs. General Theory of Time

Our requirements differ from those found usually in temporal reasoning literature (eg. [5–8]). Our goal is not to develop a general axiomatization of time using which one can reason about time. For example, in temporal logic one can axiomatize that event A occurred before event B, and event B occurred before event C. Then one can derive that A occurred before C, even without knowing the exact time when either of the events occurred.

In our application field we are dealing primarily with concrete temporal specifications, which may be imprecise, but still use absolute dates. In such setting, axiomatizing total order of the time dimension is not necessary, since it follows naturally from the total order of dates. We have found out that many application domains share this fundamental feature: rather than requiring a general theory of time allowing arbitrary inferences, they deal with numerous concrete temporal facts associated domain entities. In such setting a general theory of time is an overkill, requiring inferencing capabilities of significant computational complexity. We can replace these with more efficient mechanisms, implemented outside the logical framework.

3 Fuzzy Temporal Model

3.1 Basic Decisions

Intervals or Points? The most fundamental question in any temporal model is the choice of the basic temporal primitive. Literature mentions two usual primitives [3, 9]: time instants (or time points, chronons) and time intervals. A temporal model can be based on either or on both of them. In literature there there is an ongoing debate about which primitive is more appropriate, with no clear winner. While in the temporal database research community the time instant is more commonly used [2], in the artificial intelligence community time interval or mixed approaches are more popular [3]. In [3] it has been argued that the choice of the basic primitive mostly depends on the application requirements. We believe that time intervals are closer to human intuitive perception of time. Instants can always be viewed as time intervals if the granularity of time dimension is sufficiently increased. Further, intervals lend themselves to intuitive generalization to the fuzzy case.

Continuous vs. Discrete Temporal Model. Although there are some good arguments in the temporal database literature (e.g. [10]) for using a discrete time model, most of the approaches in AI use the continuous model, as it fits well with the choice of intervals as the basic primitive. Under a continuous temporal model, each natural language 'time point' translates to an interval in the temporal model, which is not necessarily the case for the discrete model. Therefore, we choose the unbounded, continuous time line $\mathcal{T}$ as the basis for defining time intervals, which is isomorphic to the set of real numbers, i.e. there exists a natural total ordering among elements of $\mathcal{T}$. Elements of $\mathcal{T}$ are termed as 'time points' in this paper. To anchor this abstract time line to a real calendar system, we choose the zero time point t_0 to match to the zero point of the Gregorian calendar, measured by the Greenwich Mean Time (GMT).

3.2 Time Intervals as Fuzzy Sets

We base the rest of this paper extensively on fuzzy set theory [11]. In Appendix we give a brief overview of fuzzy set theory, along with the pointers to relevant introductory literature.

In the following presentation, we assume that we need to represent a crisp interval i when a historical event happens. We denote with i^- its crisp starting point and with i^+ its crisp ending point. As discussed in Section 2.1, although the interval is crisp (after all, historical events really did occur at some precise time), the interval's start and ending points may not be known precisely. We call such an interval *imprecise* and model it by means of a fuzzy set $\tilde{I}$, defined by its membership function $\tilde{I} : \mathcal{T} \rightarrow [0, 1]$. $\tilde{I}(t)$ represents our confidence level that t is in i. If $\tilde{I}(t) = 0$, we are completely confident that t is not in i; if $\tilde{I}(t) = 1$, we are completely confident that t is in i. We term such a fuzzy set representing an abstract interval as a *fuzzy interval*. We denote the set of all fuzzy intervals as $\mathcal{I}$.

Fuzzy intervals are capable of representing imprecision caused by all of the three special properties of historical knowledge (vagueness, uncertainty or subjectivity) in a unifying way. Indeed, the possibility to express partial confidence of the membership of some time point t in i allows us the express the imprecision of the accounts about the interval i, regardless of the actual nature of imprecision.

We do not pose any constraints on the fuzzy intervals except from the requirement that they should be convex, thus reflecting our requirement that the abstract interval i which is represented by the fuzzy interval should be continuous. Although some events occur at time intervals which are not continuous (e.g. 'Poland exists as a country'), they can be represented as a set of subevents denoting continuous parts of the original event (e.g. 'Poland exists for the first time', 'Poland is divided among the Russian Empire, the Habsburg Empire and Prussia/Germany' and 'Poland exists again').

In the case of convex fuzzy sets, their support and core are continuous. Further, all time points of i must be in the support of $\tilde{I}$ (denoted as $S_{\tilde{I}}$). I.e., if $t \notin S_{\tilde{I}}$, then we are certain that $t \notin i$. We also assume that all of the time points in the core of $\tilde{I}$ (denoted as $C_{\tilde{I}}$) are really members of i. I.e., if $t \in C_{\tilde{I}}$, then we are certain that $t \in i$. This can also be written as $C_{\tilde{I}} \subseteq i \subseteq S_{\tilde{I}}$.

4 Fuzzy Temporal Relations

In this section we show how to realize relations from Table 1 in our fuzzy model. We start with the observation that, since our intervals are not crisp, our relations will also not be crisp. After all, since the intervals are not exact, we cannot exactly determine whether one interval precedes the other one. Hence, given two imprecise crisp intervals i and j and a crisp temporal relation θ, the fuzzy temporal relation $\tilde{\theta}$ will take fuzzy sets $\tilde{I}$ and $\tilde{J}$ and produce a number $c \in [0,1]$, giving the confidence that the classical temporal θ relation holds between i and j.

Extending classical temporal relations to fuzzy sets is not easy, since classical relations in Table 1 are defined using interval endpoints. However, for fuzzy intervals the notion of endpoints is meaningless. Therefore, we define fuzzy temporal relations in an alternative way, compatible with the crisp case. We do this in several steps: first we reformulate the definition of crisp temporal relations based on the set operations on intervals, thus eliminating references to interval endpoints. In doing so, we introduce several auxiliary unary operators on intervals (e.g. 'before extension'), representing intervals with particular relationship to the original interval (e.g. interval additionally including the time before the interval). After that we extend the definitions of temporal relations to the fuzzy case by providing a fuzzy counterpart of auxiliary operators and reusing the usual fuzzy set operations.

4.1 Defining Crisp Temporal Relations using Set Operations

The basic idea for eliminating references to interval endpoints is the following. Firstly, if $t_1 < t_2$, then the interval between t_1 and t_2 is not equal to the empty set. This we can be written as

$$t_1 < t_2 \quad \Leftrightarrow \quad (t_1, t_2) \neq \emptyset \tag{1}$$

Secondly, if $t_1 = t_2$, then we have to make sure that both intervals (t_1, t_2) and (t_2, t_1) are empty sets, thus expressing that neither t_1 is after t_2 or vice versa. This we can be written as

$$t_1 = t_2 \quad \Leftrightarrow \quad (t_1, t_2) = \emptyset \ \wedge \ (t_2, t_1) = \emptyset \tag{2}$$

Further, we define several auxiliary unary operators on intervals. The role of these operators is to construct the intervals commonly used in definitions of temporal relations. The following eight operators $<-, \leq-, >-, \geq-, <+, \leq+, >+, \geq+$ take an interval and construct an interval containing all of the time points which are (strictly) before or (strictly) after the starting or ending point of the original interval. E.g. $<-(i) = (-\infty, i^-)$. The definition of these operators is given in Table 2.

Now we are ready to redefine the temporal relations using the ideas presented in (1) and (2) and the unary operators from Table 2. We explain how this is done for the `starts` relation, since its definition uses both endpoint equality and inequality. Other relations are defined similarly and are given in Table 3. We did not redefine the `after`, `overlapped-by`, `contains`, `met-by`, `started-by` and `finished-by` relations, as they are simply the inverse of other relations.

We start the redefinition of the `starts` relation by repeating the definition from Table 1:

$$i \, \texttt{starts} \, j \;\equiv\; i^- = j^- \;\wedge\; i^+ < j^+ \tag{3}$$

The constraint $i^- = j^-$ can be expressed as (cf. (2))

$$i^- = j^- \;\Leftrightarrow\; (i^-, j^-) = \emptyset \;\wedge\; (j^-, i^-) = \emptyset \tag{4}$$

which can be written with the help of auxiliary unary operators as

$$>\!-(i) \cap <\!-(j) \;=\; \emptyset \;\wedge\; >\!-(j) \cap <\!-(i) \;=\; \emptyset \tag{5}$$

because we know that

$$(t_1, t_2) = (t_1, +\infty) \cap (-\infty, t_2) \tag{6}$$

This last step is needed to eliminate all references to interval endpoints in the definition.

Similarly, the constraint $i^+ < j^+$ can be expressed using (1) and (6) as

$$i^+ < j^+ \;\Leftrightarrow\; >\!+(i) \cap <\!+(j) \neq \emptyset \tag{7}$$

Hence, the `starts` relation can be defined by means of set operations on intervals as

$$i \, \texttt{starts} \, j \;\equiv\; \begin{array}{l} >\!-(i) \cap <\!-(j) \;=\; \emptyset \;\wedge \\ >\!-(j) \cap <\!-(i) \;=\; \emptyset \;\wedge \\ >\!+(i) \cap <\!+(j) \;\neq\; \emptyset \end{array} \tag{8}$$

Finally, we note that the `intersects` relation has not been derived from the definition in Table 1. Instead, simply the fact was used that it expresses the constraint that the intersection of i and j is not empty.

4.2 Extending Auxiliary Interval Operators to Fuzzy Intervals

In this section we extend the auxiliary interval operators to operate on fuzzy intervals. We denote the extended operators with the same symbols, i.e. as $<\!-, \leq\!-, >\!-, \geq\!-, <\!+, \leq\!+, >\!+, \geq\!+$, exactly as in the crisp case. Each operator $\tilde{\theta}$ will be function $\tilde{\theta} : \mathcal{I} \rightarrow \mathcal{I}$, i.e it will take a fuzzy interval and yield another fuzzy interval. The semantics of $\tilde{\theta}(\tilde{I})$ should be understood as follows: $\tilde{\theta}(\tilde{I})(t)$ gives our confidence that t is in $\theta(i)$. For example, $<\!-(\tilde{I})(t)$ represents our confidence that t is in $<\!-(i)$. In order to make our notation simpler, we will sometimes write $\tilde{\theta}(\tilde{I})$ as $\tilde{I}_{\tilde{\theta}}$.

In the rest of this section we will show how to extend the operators $\geq\!-$ and $<\!-$ to the fuzzy case. The other operators can be extended in a similar manner, and their definitions are shown in Table 4.

The operator $\geq\!-$ should, for some fuzzy interval $\tilde{I}$ representing interval i, give a fuzzy interval $\tilde{I}_{\geq -}$, representing the interval $\geq\!-(i)$. Let $s^-_{\tilde{I}}$ and $s^+_{\tilde{I}}$ denote

Table 1. Required Temporal Relations

Interval Relation	*Definition*
i `before` j	$i^+ < j^-$
i `after` j	j `before` i
i `overlaps` j	$i^- < j^- \ \wedge\ i^+ > j^- \ \wedge\ i^+ < j^+$
i `overlapped-by` j	j `overlaps` i
i `during` j	$i^- > j^- \ \wedge\ i^+ < j^+$
i `contains` j	j `during` i
i `meets` j	$i^+ = j^-$
i `met-by` j	j `meets` i
i `starts` j	$i^- = j^- \ \wedge\ i^+ < j^+$
i `started-by` j	j `starts` i
i `finishes` j	$j^- < i^- \ \wedge\ i^+ = j^+$
i `finished-by` j	j `finishes` i
i `equals` j	$i^- = j^- \ \wedge\ i^+ = j^+$
i `intersects` j	$i^+ > j^- \ \wedge\ i^- < j^+$

Table 2. Auxiliary Unary Operators on Intervals

Operator	*Result*
$<-$	$<-(i) = (-\infty, i^-)$
$\leq-$	$\leq-(i) = (-\infty, i^-]$
$>-$	$>-(i) = (i^-, +\infty)$
$\geq-$	$\geq-(i) = [i^-, +\infty)$
$<+$	$<+(i) = (-\infty, i^+)$
$\leq+$	$\leq+(i) = (-\infty, i^+]$
$>+$	$>+(i) = (i^+, +\infty)$
$\geq+$	$\geq-(i) = [i^+, +\infty)$

Table 3. Transcribed Temporal Relations

Temporal Relation	*Definition*
i `before` j	$>+(i) \cap <-(j) \neq \emptyset$
i `overlaps` j	$>-(i) \cap <-(j) \neq \emptyset \ \wedge$ $<+(i) \cap >-(j) \neq \emptyset \ \wedge$ $>+(i) \cap <+(j) \neq \emptyset$
i `during` j	$<-(i) \cap >-(j) \neq \emptyset \ \wedge$ $>+(i) \cap <+(j) \neq \emptyset$
i `meets` j	$>+(i) \cap <-(j) = \emptyset \ \wedge$ $<+(i) \cap >-(j) = \emptyset$
i `starts` j	$>-(i) \cap <-(j) = \emptyset \ \wedge$ $<-(i) \cap >-(j) = \emptyset \ \wedge$ $>+(i) \cap <+(j) \neq \emptyset$
i `finishes` j	$<-(i) \cap >-(j) \neq \emptyset \ \wedge$ $>+(i) \cap <+(j) = \emptyset \ \wedge$ $<+(i) \cap >+(j) = \emptyset$
i `equals` j	$>-(i) \cap <-(j) = \emptyset \ \wedge$ $<-(i) \cap >-(j) = \emptyset \ \wedge$ $>+(i) \cap <+(j) = \emptyset \ \wedge$ $<+(i) \cap >+(j) = \emptyset$
i `intersects` j	$i \cap j \neq \emptyset$

the starting and ending points of $S_{\tilde{I}}$ (i.e. of the the support of $\tilde{I}$). By assumptions from Section 3.2, we known that $i \subseteq S_{\tilde{I}}$. Therefore, $\tilde{I}_{\geq-}(t)$ should be 0 for each $t < s^-_{\tilde{I}}$, and should be 1 for each $t > s^+_{\tilde{I}}$, as we know that each time point before $s^-_{\tilde{I}}$ are before i^-, and we know that each time point after $s^+_{\tilde{I}}$ is after i^+ and therefore also after i^-. For a $t \in S_{\tilde{I}}$, we can tell that our confidence that $t \in \geq-(i)$ should be as big as our confidence that $s \in \geq-(i)$ for any $s \leq t$. Therefore we define the operator $\geq- : \tilde{I} \to \tilde{I}$ as follows:

$$\tilde{I}_{\geq-}(t) = \begin{cases} 0 & \text{if } t < S^-_{\tilde{I}} \\ \sup_{s \leq t} \tilde{I}(s) & \text{if } t \in S_{\tilde{I}} \\ 1 & \text{if } t > S^+_{\tilde{I}} \end{cases} \tag{9}$$

The definition of the operator $<-$ is easy if we already defined the $\geq-$ operator. We note that, if $t \in <-(i)$, then $t \notin \geq-(i)$. Therefore we simply define the $\tilde{I}_{<-}$ fuzzy interval as the fuzzy complement of the $\tilde{I}_{\geq-}$ fuzzy interval:

$$\tilde{I}_{<-}(t) = \tilde{I}^C_{\geq -}(t) = 1 - \tilde{I}_{\geq -}(t) \tag{10}$$

The results of applying the $\geq -$ and $<-$ operators on a fuzzy interval are shown in Figure 1.

4.3 Constraints using Comparison with Empty Set

Before we can finally extend the definitions of the temporal relations to fuzzy intervals, we must extended constraints of the form $a \cap b \neq \emptyset$ and $a \cap b = \emptyset$ to fuzzy intervals. We use the following intuition: the value $\sup_t \tilde{I}(t)$ (i.e. the maximum confidence of membership of any time point in the interval) gives the confidence that some t is in i, i.e. our confidence that i is not empty.

Since fuzzy intersection is expressed using min operator (cf. Appendix), our confidence that $a \cap b \neq \emptyset$ can be represented as

$$\sup_t \min(\tilde{A}(t), \tilde{B}(t)) \tag{11}$$

Since $a \cap b = \emptyset$ is simply the negation of $a \cap b \neq \emptyset$, our confidence in that this constraint is fulfilled is given as

$$1 - \sup_t \min(\tilde{A}(t), \tilde{B}(t)) = \inf_t \max(\tilde{A}^C(t), \tilde{B}^C(t)) \tag{12}$$

4.4 Temporal Relations on Fuzzy Intervals

Now we are ready to extend the definition of the temporal relations to fuzzy intervals based on the transformed definitions from Table 3. We define a fuzzy temporal relation γ as a function $\gamma : \tilde{I} \times \tilde{I} \to [0, 1]$. In another words, a temporal relation takes two fuzzy intervals and gives the confidence that the crisp temporal relation holds between the abstract intervals represented by the respective fuzzy intervals. We denote the fuzzy temporal relations with big letters to distinguish them from their crisp counterparts.

We discuss the definition of relation $\texttt{STARTS}(\tilde{I},\tilde{J})$. Other relations can be defined in a similar way, and they are shown in Table 5. We start from the definition of the crisp relation `starts`, which was defined in Section 4.1 as

$$i \,\texttt{starts}\, j \;\equiv\; \begin{array}{l} >-(i) \cap <-(j) \;=\; \emptyset \;\wedge \\ >-(j) \cap <-(i) \;=\; \emptyset \;\wedge \\ >+(i) \cap <+(j) \;\neq\; \emptyset \end{array} \tag{13}$$

After applying the rules for transcribing the constraints (cf. subsection 4.3) we get:

$$\begin{array}{l} \texttt{STARTS}(\tilde{I}, \tilde{J}) = \min(\\ \qquad \inf_t \max(\tilde{I}_{\leq -}(t), \tilde{J}_{\geq -}(t)), \\ \qquad \inf_t \max(\tilde{I}_{\geq -}(t), \tilde{J}_{\leq -}(t)), \\ \qquad \sup_t \min(\tilde{I}_{>+}(t), \tilde{J}_{<+}(t))) \end{array} \tag{14}$$

Table 4. Auxiliary Operators on Fuzzy Intervals

$\tilde{\theta}$	$\tilde{\theta}(\tilde{I})(t)$
$<-$	$1 - \tilde{I}_{\geq -}(t)$
$\leq -$	$1 - \tilde{I}_{> -}(t)$
$>-$	$\begin{cases} 0 & \text{if } t < S_{\tilde{I}}^- \\ \sup_{s<t} \tilde{I}(s) & \text{if } t \in S_{\tilde{I}} \\ 1 & \text{if } t > S_{\tilde{I}}^+ \end{cases}$
$\geq -$	$\begin{cases} 0 & \text{if } t < S_{\tilde{I}}^- \\ \sup_{s\leq t} \tilde{I}(s) & \text{if } t \in S_{\tilde{I}} \\ 1 & \text{if } t > S_{\tilde{I}}^+ \end{cases}$
$<+$	$\begin{cases} 0 & \text{if } t < S_{\tilde{I}}^- \\ \sup_{s>t} \tilde{I}(s) & \text{if } t \in S_{\tilde{I}} \\ 1 & \text{if } t > S_{\tilde{I}}^+ \end{cases}$
$\leq +$	$\begin{cases} 0 & \text{if } t < S_{\tilde{I}}^- \\ \sup_{s\geq t} \tilde{I}(s) & \text{if } t \in S_{\tilde{I}} \\ 1 & \text{if } t > S_{\tilde{I}}^+ \end{cases}$
$>+$	$1 - \tilde{I}_{\leq -}(t)$
$\geq +$	$1 - \tilde{I}_{< -}(t)$

Table 5. Temporal Relations on Fuzzy Intervals

Relation	*Definition*
`BEFORE`$(\tilde{I},\tilde{J})$	$\sup_t \min(\tilde{I}_{>+}(t), \tilde{J}_{<-}(t))$
`OVERLAPS`$(\tilde{I},\tilde{J})$	$\min(\ \sup_t \min(\tilde{I}_{>-}(t), \tilde{J}_{<-}(t))\ ,\ \sup_t \min(\tilde{I}_{<+}(t), \tilde{J}_{>-}(t))\ ,\ \sup_t \min(\tilde{I}_{>+}(t), \tilde{J}_{<+}(t))\ )$
`DURING`$(\tilde{I},\tilde{J})$	$\min(\ \sup_t \min(\tilde{I}_{<-}(t), \tilde{J}_{>-}(t))\ ,\ \sup_t \min(\tilde{I}_{>+}(t), \tilde{J}_{<+}(t))\ )$
`MEETS`$(\tilde{I},\tilde{J})$	$\min(\ \inf_t \max(\tilde{I}_{\leq +}(t), \tilde{J}_{\geq -}(t))\ ,\ \inf_t \max(\tilde{I}_{\geq +}(t), \tilde{J}_{\leq -}(t))\ )$
`STARTS`$(\tilde{I},\tilde{J})$	$\min(\ \inf_t \max(\tilde{I}_{\leq -}(t), \tilde{J}_{\geq -}(t))\ ,\ \inf_t \max(\tilde{I}_{\geq -}(t), \tilde{J}_{\leq -}(t))\ ,\ \sup_t \min(\tilde{I}_{>+}(t), \tilde{J}_{<+}(t))\ )$
`FINISHES`$(\tilde{I},\tilde{J})$	$\min(\ \inf_t \max(\tilde{I}_{\leq +}(t), \tilde{J}_{\geq +}(t))\ ,\ \inf_t \max(\tilde{I}_{\geq +}(t), \tilde{J}_{\leq +}(t))\ ,\ \sup_t \min(\tilde{I}_{>-}(t), \tilde{J}_{<-}(t))\ )$
`EQUALS`$(\tilde{I},\tilde{J})$	$\min(\ \inf_t \max(\tilde{I}_{\leq +}(t), \tilde{J}_{\geq +}(t))\ ,\ \inf_t \max(\tilde{I}_{\geq +}(t), \tilde{J}_{\leq +}(t))\ ,\ \inf_t \max(\tilde{I}_{\leq -}(t), \tilde{J}_{\geq -}(t))\ ,\ \inf_t \max(\tilde{I}_{\geq -}(t), \tilde{J}_{\leq -}(t))\ )$
`INTERSECTS`$(\tilde{I},\tilde{J})$	$\sup_t \min(\tilde{I}(t), \tilde{J}(t))$

5 Discussion

In this section we discuss the model and relations presented in Sections 3 and 4 and demonstrate how they fulfill the requirements from Section 2.

Representing Imprecise Information. As discussed briefly in Section 3.2, fuzzy intervals are capable of representing all three causes of impreciseness in history. They represent the net confidence of the history expert about the statement $t \in i$, where the lack of confidence can be caused by any combination of vagueness, uncertainty and subjectivity.

Of course, in a realistic application scenario historical experts will not specify fuzzy intervals by encoding their confidence about each $t \in \mathcal{T}$, but will apply some heuristics on their intuitive temporal knowledge. Finding the best heuristics in the specific cases is subject of further research. As an example, we show a possible interpretation of the intervals 'late twenties – early thirties', '320? B.C. – 280 B.C.' and 'Russian Revolution happens' in Figure 1, each of them showing one of the special characteristics of historical knowledge.

Compatibility with Crisp Case It is easy to see that fuzzy relations are natural extensions of the classical ones, as they give the same results on crisp intervals as the classical ones. This is because we defined the fuzzy relations based on

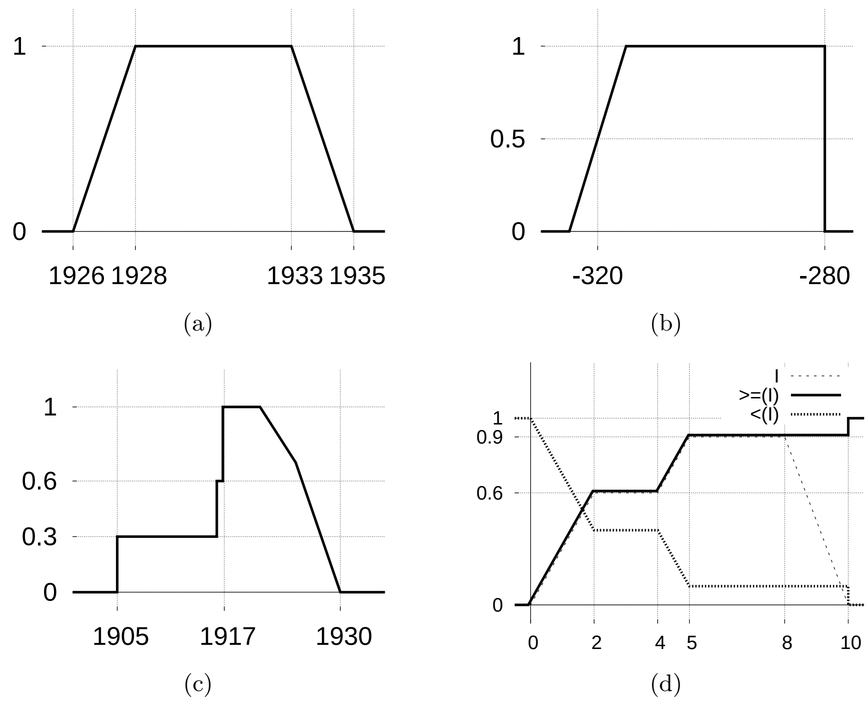

Fig. 1. Fuzzy intervals (a) 'late twenties' – 'early thirties' (b) 320? B.C. – 280 B.C. (c) Russian Revolution happens (d) $\tilde{I}_{\geq -}$ and $\tilde{I}_{<-}$

the original definitions of the classical relations. Hence, the requirements on the compatibility with the crisp case is fulfilled completely.

Intuitiveness of Fuzzy Relations. We believe that our fuzzy relations yield intuitive result, where 'intuitive' for us means that the relation gives 1 as result if we are completely certain that the classical relation exists between the classical abstract intervals represented by the fuzzy intervals, and 0 if we are certain that this is impossible. A result between 0 and 1 is given if neither of these possibilities are sure. E.g. in case of the `BEFORE`($\tilde{I}$,$\tilde{J}$), it should yield 0 if $S_{\tilde{I}}$ `before` $S_{\tilde{J}}$ holds (i.e. we are sure that i `before` j holds) and it should yield 1 if $C_{\tilde{I}}$ `before` $C_{\tilde{J}}$ does not hold (i.e. we are sure that i `before` j does not hold). Intuitiveness can similarly be checked for other fuzzy relations as well.

6 Connecting Temporal and Ontology Models

In this section we discuss how the temporal model described previously is used in ontology modeling. Our approach for integrating temporal and ontology models is general and not tied to any particular model. However, to make the discussion more concrete, we briefly describe the designated ontology model first, after which we show how model integration is actually done.

6.1 Target Ontology Model

Out target ontology model is that of KAON[2] - an ontology management framework developed by FZI and AIFB at University of Karlsruhe. The model [12] is based on RDF(S) [13], allowing modeling of concepts, properties, instances and relations between instances. It extends RDF(S) with several useful features, such as symmetric, transitive and inverse properties. Unlike RDF(S), however, KAON does not support reification of statements, due to the fact that formal semantics of the KAON language is much more similar to OWL [14] (actually, KAON ontology model is currently being extended with OWL primitives using results from logic programming [15]). As it is the case in all description logics based languages, formal semantics of OWL is based on the usual first-order theory. However, contrary to OWL, KAON semantics is based on HiLog [16], providing a clean mechanism for second-order flavor of the ontology language. For example, in KAON it is possible to interpret a symbol as a concept or as an instance, depending on the symbol's context. In KAON information about concepts and instances is structured in OI-models (ontology-instance models). Figure 2 shows a fragment of the VICODI OI-model.

6.2 Integrating Temporal and Ontology Model

Our approach for integrating the temporal model into ontological definitions follows a pattern of modular semantics, which we believe will gain importance in the near future as the complexity of domains being modeled increases. This pattern is based on the observation that particular formalism may be good for some modeling tasks, but totally inappropriate for other ones. Trying to apply the most general formalism (e.g. first-order logic) to all modeling tasks usually results in cumbersome systems with inadequate performance. Rather, a more promising approach is to combine various formalisms in a modular way, thus harvesting the best of each of them. In this paper we apply this principle to time modeling. However, we could imagine a spatial algebra being orthogonally added to the temporal and description logics formalisms in a similar manner.

Our approach may schematically be described as in Figure 3. On the left-hand side of the figure is the ontology model with its HiLog semantics, whereas on the right-hand side is the fuzzy temporal model with the semantics as described in Sections 3 and 4. These two heterogeneous semantics are orthogonal and need to be integrated at the syntactical and at the semantic level.

Integration at the syntactical level defines how to physically connect elements from one model with another. We have found out that data types provide an excellent mechanism for this purpose. Many ontology languages (e.g. OWL) offer the capability of modeling atomic objects whose semantics is out of scope of the logical theory. In semantic interpretation of the ontology, instances of data types are interpreted as members of some concrete domain (often denoted as Δ_D). On the other hand, ontology instances are interpreted as

[2] http://kaon.semanticweb.org/

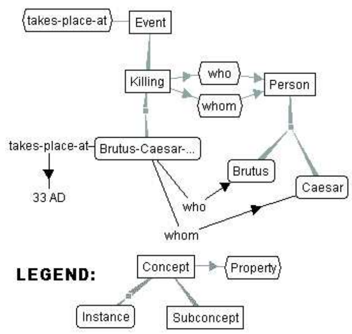

Fig. 2. Fragment of VICODI OI-model

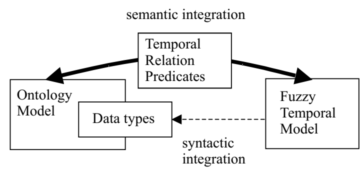

Fig. 3. Integrating Ontology and Fuzzy Temporal Models

members of the abstract domain (often denoted as Δ^I). The concrete and abstract domains must be disjoint, thus causing the semantics of data types and of the ontology model to be separated. In our case, we introduced a separate TemporalSpecification data type which is responsible for representing temporal information. Currently, TemporalSpecification has only one sub-data type called FuzzyTimeInterval.

Integration at the semantic level defines how properties of one model semantically relate to the other model. In our case this means we need to specify how the temporal relations from Section 4 is represented in the first-order setting. This is done by introducing for each temporal relation from Table 5 a many-sorted predicate reflecting the logical properties of the relation. One can think of these predicates as built-ins: the arguments of the predicate are fuzzy intervals whose content is opaque to the logic infrastructure. The predicates serve as a gate between the two worlds, providing an abstract interface to the interval model. One must observe that the semantics of the predicates is not axiomatized in first-order logic, but reasoning may still be performed on the arguments and results of the predicates. Each predicate has an additional argument receiving the fuzzy value of the relation. For example, if fuzzy interval $\tilde{I}$ is before interval $\tilde{J}$ with confidence level 0.8, then the first-order formula $\texttt{BEFORE}(\tilde{I},\tilde{J},0.8)$ is true. Note that constraints on the confidence level can be expressed by using variables:

$$\texttt{BEFORE}(\tilde{I}, \tilde{J}, X) \quad \wedge \quad X > 0.8 \tag{15}$$

6.3 Annotating Ontology Definitions with Temporal Specifications

The previous subsection discussed how to integrate ontology and temporal model at the generic level. However, an important question remains open: how to attach the temporal specifications to the ontology definitions? For example, how to represent the fact that French Revolution lasted from 1789 to 1794? Before answering this question, we examine the types of ontology definitions to which temporal specifications could be applicable at all.

Annotating Ontology Elements Themselves. At the first glance, it seems to be useful annotating ontology definitions, such as concept or property declarations, themselves, with time information in which the concept or property definition exists. For example, one might be tempted to state that the concept EUCountry exists only after 1992. Before this date it does not make much sense to talk about EU countries anyway. Another useful example is annotating the subconcept relationship with time information, thus allowing representation of classification which varies over time. Although useful at the first glance, such annotations are extremely difficult to manage in a logical setting. Concept and property declarations are semantically represented as predicates. In a first-order theory one cannot attach additional information to predicates - one would need a full second-order language where predicates can be treated as instances of meta-predicates. Second-order logic is extremely difficult to manage, and introducing such features in HiLog or first-order logic is extremely messy, so we decided not to support such annotations. Further, ontology axioms stating that A is a subclass of B are tantamount to first-order formulas $\forall x(A(x) \rightarrow B(x))$. Annotating formulas is not possible in any logic and the semantic interpretation of such annotations has not been proposed yet.

After some considerations, one may see that such annotations often do not make sense. One may argue that the concept of EU countries existed before 1992, it just did not have any instances. Therefore, this is in practice not really a limitation. Thus, we have dismissed temporal annotations for ontology definitions, but we allow temporal annotations for instances and relationships among them.

Semantics of Temporal Specifications. When building ontologies containing temporal information, a fundamental question about the meaning of temporal specifications for ontology entities arises. Most approaches in temporal databases [10] use only one type of time. Namely, each tuple in the database is annotated with the validity time, thus specifying the period when the information represented by the tuple is valid in the real world. In temporal reasoning the usual approach is to associate an additional temporal argument with each predicate [3].

We argue that such approaches are too simple for many applications, as they allow specifying only one type of time. For example, one may say that Allan Turing lived from 23 June 1912 until 7 June 1954, which certainly may be represented as his validity time. On the other hand, one may associate with him additionally the interval from 1931 until 1935, which represents the time when he studied at King's College. Hence, we see that different types of time information are needed for expressive ontology modeling. If needed, one may additionally axiomatize the lived relation as validity time, but this is, according to our opinion, domain-specific.

Annotating Relationships between Instances. A unique feature of ontologies containing temporal information is that temporal information is not only associated to instances (e.g. Person was-born-at certain time), but also to relationships between instances (e.g. Brutus kills Caesar with annotation that this event

TAKES-PLACE-AT certain time). Capturing this type of information presents new challenges to our approach for integrating temporal and ontology models. In particular, our ontology model allows connecting two instances through a property instance (i.e. at the logical level all relationships between instances are represented as two-place predicates). This restriction is fundamental to our ontology model, since it is known that ontology languages with predicates of arbitrary arity may easily become undecidable.

We could solve this problem using reification approaches of RDFS. However, as we explained in subsection 6.1, the semantics of statement reification is not clean in RDFS and does not match well with the usual first-order semantics. Therefore, we solve the problem by reifying relationships annotated by temporal specifications into first-class objects. In another words, the example problem presented above is solved by introducing the concept KILLING as a subconcept of the EVENT concept, with properties WHO, WHOM and TAKES-PLACE-AT. This solution is presented in Figure 2.

Although cumbersome at the first glance, this approach has an advantage that all time information is represented in a uniform way through a taxonomy of temporal properties. In this way selecting all events taking place at certain point in time becomes very easy.

7 Related work

There is a significant amount of approaches for representing temporal information in the areas of temporal reasoning and temporal databases. Most of these, however, consider only classical time intervals (or time points) and do not deal with any form of imprecisions.

There are some approaches, however, which provide some solutions for handling uncertain temporal information. Most of them model uncertainty with possibility or probability distributions on interval endpoints.

In the area of temporal databases [2], the approaches [17, 18] define the probability distribution of each endpoint of crisp time intervals. However, it is generally debated whether probability distributions are appropriate for representing subjective information at all, as objective statistics, that probability distributions are based on, are often missing when defining subjective information [19].

Because of that, most approaches in temporal reasoning for modeling subjective temporal knowledge use probability distributions expressed as fuzzy sets to represent uncertainty. Dubois and Prade [5] propose an approach where endpoints of a fuzzy interval are modeled as fuzzy numbers. Further, they use possibility theory [20, 19] to calculate time points which are possibly or necessarily between the two fuzzy endpoints. They also provide fuzzy extensions of Allen's interval algebra and some basic inference mechanisms.

Kurutach [21] also proposes using fuzzy numbers representing interval endpoints similarly to Dubois and Prade. Moreover, he imposes constraints on the length of intervals. Godo and Vila [7] propose using fuzzy sets constraining the length of the time period between intervals. Although this approach is adequate for some problems in the health-care domain, we found it quite inadequate for modeling historical imprecise intervals as it is not possible to specify absolute dates for intervals.

Almost all of the approaches for representing uncertain intervals is based on the notion of uncertain interval endpoints. However, as it was discussed in Section 2.1 it is not always intuitive to assume that there is a (possibly ill-known) definite starting and ending point for an interval. There are some events, where it makes sense of speaking of 'transition periods' in addition of a 'core period'. E.g. consider the case of the Russian revolution, whose interval is shown in Figure 1. In this case the transtition periods are much longer as the core period of the event. Using an endpoint-based approach one has to decide which one of the possible endpoints to take, which means loosing information. With our approach it is possible to model not only the core period of an event, but also transtition, development etc. periods, which are only partially relevant for a specific event. This is posisble because we define intervals directly, without referencing the endpoints.

An interesting approach for representing uncertainty about time-dependent events which follows a different idea as the works introduced so far is that of Dutta's [6]. He uses the set of known intervals as the universe for fuzzy sets. A fuzzy set representing an event e shows the possibility for each interval i that that event occurs in it. Although this approach is different from the other described approaches in the sense that it is capable of representing fuzzily defined events, it views intervals themselves only as abstract, crisp entities without any further temporal specification, therefore it is not applicable in our application scenario.

8 Conclusion and Outlook

In this paper we presented a fuzzy temporal model that is capable to represent all aspects of historical temporal knowledge in a common formal framework. Although the model's design was based on the requirements from the application domain of history, it can also be used in other application domains that require representing uncertain, subjective and vague temporal knowledge (such as health-care). Apart from defining the temporal model based on fuzzy sets, we also generalized Allen's temporal relations on intervals. We did this by providing a definition of crisp interval relations based on set theory and then generalized them to the fuzzy case.

Our temporal model is intended for use in ontology modeling. Hence, we provided an approach for integrating temporal and ontology models. The approach follows the modular semantics pattern, which tries to keep the semantics of each model separate and to provide clean interfaces between them.

We are presently working on a methodology for capturing fuzzy temporal knowledge from experts. There are specifically two challenges we plan to address. Firstly, we want to provide an easy, intuitive strategy for capturing temporal specifications without the need to specify fuzzy intervals directly. In this way we hope to make our system usable by non-IT experts. Secondly, a conversion mechanism between different temporal granularities is needed since temporal granularity of historical facts usually differs (e.g. some accounts are specified in years and some in days).

We are also examining the different properties of our new fuzzy temporal relations (like transitivity) which work will allow us to make basic inferences even in case of fuzzy intervals (e.g. $\text{BEFORE}(\tilde{I}, \tilde{J}) = 0.8 \wedge \text{BEFORE}(\tilde{J}, \tilde{K}) = 0.7 \Rightarrow \text{BEFORE}(\tilde{I}, \tilde{K}) \geq 0.7$).

References

1. Jensen, C.S., Dyreson, C.E., Böhlen, M., Clifford, J., Elmasri, R., Gadia, S.K., Grandi, F., Hayes, P., Jajodia, S., Käfer, W., Kline, N., Lorentzos, N., Mitsopoulos, Y., Montanari, A., Nonen, D., Peressi, E., Pernici, B., Roddick, J.F., Sarda, N.L., Scalas, M.R., Segev, A., Snodgrass, R.T., Soo, M.D., Tansel, A., Tiberio, P., Wiederhold, G.: The consensus glossary of temporal database concepts — february 1998 version. Lecture Notes in Computer Science **1399** (1998) 367–405
2. Etzion, O., Jajodia, S., Sripada, S., eds.: Temporal databases: research and practice. In Etzion, O., Jajodia, S., Sripada, S., eds.: Temporal databases: research and practice. Volume 1399 of Lecture Notes in Computer Science., New York, NY, USA, Springer-Verlag Inc. (1998)
3. Vila, L.: A survey on temporal reasoning in artificial intelligence. AICOM (Artificial Intelligence Communications) **7** (1994) 4–28
4. Allen, J.F.: Maintaining knowledge about temporal intervals. Communications of the ACM **26** (1983) 832–843
5. Dubois, D., Prade, H.: Processing fuzzy temporal knowledge. IEEE Transactions of Systems, Man and Cybernetics **19** (1989) 729–744
6. Dutta, S.: An event-based fuzzy temporal logic. In: Proc. 18th IEEE Intl. Symp. on Multiple-Valued Logic, Palma de Mallorca, Spain (1988) 64–71
7. Godo, L., Vila, L.: Possibilistic temporal reasoning based on fuzzy temporal constraints. In Mellish, C., ed.: IJCAI'95: Proceedings International Joint Conference on Artificial Intelligence, Montreal (1995)
8. DARPA Agent Markup Language project: DAML-Time Homepage. (2002) Accessible from the URL http://www.cs.rochester.edu/~ferguson/daml/.
9. Chomicki, J.: Temporal query languages: A survey. In Gabbay, D.M., Ohlbach, H.J., eds.: Proceedings of the 1st International Conference on Temporal Logic. Volume 827 of LNAI., Berlin, Springer (1994) 506–534
10. McKenzie, E., Snodgrass, R.: An evaluation of relational algebras incorporating the time dimension in databases. ACM Computing Surveys **23** (1991) 501–543
11. Zadeh, L.A.: Fuzzy sets. Information and Control **8** (1965) 338–353
12. Motik, B., Maedche, A., Volz, R.: A conceptual modeling approach for semantics-driven enterprise applications. In: Proc. 1st Int'l Conf. on Ontologies, Databases and Application of Semantics (ODBASE-2002). (2002)
13. Brickley, D., Guha, R.: RDF Vocabulary Description Language 1.0: RDF Schema. W3C. (2000) Accessible from the URL http://www.w3.org/TR/rdf-schema.
14. Patel-Schneider, P.F., Hayes, P., Horrocks, I., van Harmelen, F.: Web Ontology Language (OWL) Abstract Syntax and Semantics. W3C. (2002) Accessible from the URL http://www.w3.org/TR/owl-semantics/.
15. Grossof, B., Horrocks, I., Volz, R., Decker, S.: Description logic programs: Combining logic programs with description logic. In: Proceedings of WWW 2003, Budapest, Hungary (2003)

16. Chen, W., Kifer, M., Warren, D.S.: Hilog: A foundation for higher-order logic programming. Journal of Logic Programming **15** (1993) 183–230
17. Dekhtyar, A., Ross, R., Subrahmanian, V.S.: Probabilistic temporal databases, I: algebra. ACM Transactions on Database Systems (TODS) **26** (2001) 41–95
18. Dyreson, C.E., Snodgrass, R.T.: Supporting valid-time indeterminacy. ACM Transactions on Database Systems **23** (1998) 1–57
19. Dubois, D., Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York (1986)
20. Zadeh, L.: Fuzzy sets as a basis for possibility. Fuzzy Sets and Systems **1** (1978) 3–28
21. Kurutach, W.: Modelling fuzzy interval-based temporal information: a temporal database perspective. In: Proceedings of 1995 IEEE International Conference on Fuzzy Systems, Yokohama, Japan (1995) 741–748
22. Dubois, D., Prade, H.: Fundamentals of Fuzzy Sets. Volume 7 of The handbooks of fuzzy sets series; FSHS 7. Kluwer Academic Publishers (2000)
23. Nguyen, H.T., Walker, E.A.: A First Course in Fuzzy Logic. CRC Press, New York (1997)

Appendix: Fuzzy Set Basics

In this appendix we briefly recapitulate the fundamental notions about the fuzzy sets and fuzzy logic. Further details can be found in any textbook of fuzzy sets or fuzzy logic (e.g. [22, 23]).

Fuzzy sets generalize the notion of classical sets (also called as crisp sets as a counterpart of fuzzy). A subset A of the set $\mathcal{U}$ (the universe of discourse) can be specified using the *characteristic function* $A : \mathcal{U} \rightarrow \{0, 1\}$. $A(x) = 1$ if $x \in A$ and $A(x) = 0$ if $x \notin A$. Similarly, a *fuzzy subset* $\tilde{A}$ of $\mathcal{U}$ can be characterized with a *membership function* $\tilde{A} : \mathcal{U} \rightarrow [0, 1]$. For each $x \in \mathcal{U}$ $\tilde{A}(x)$ represents the membership grade of x in $\tilde{A}$. Hence, x can be a member of a $\tilde{A}$ partially. We call fuzzy subsets (of $\mathcal{U}$) simply as fuzzy sets from now on and assume that the universe of discourse is understood from the context.

Similarly as in the crisp case, logical connectives $\wedge$, $\vee$ and $\neg$ may in the fuzzy case be identified with the set operations $\cap, \cup$ and $\tilde{A}^C$ (the former is the fuzzy complement of $\tilde{A}$). The usual definition of the fuzzy set operations (which we will also use in this paper) are the following:

$$(\tilde{A} \cap \tilde{B})(x) = \min(\tilde{A}(x), \tilde{B}(x)) \tag{16}$$

$$(\tilde{A} \cup \tilde{B})(x) = \max(\tilde{A}(x), \tilde{B}(x)) \tag{17}$$

$$\tilde{A}^C(x) = 1 - \tilde{A}(x) \tag{18}$$

The *core* of $\tilde{A}$ is the crisp set $C_{\tilde{A}} = \{x \in \mathcal{U} : \tilde{A}(x) = 1\}$, i.e. the set of elements which completely belong to $\tilde{A}$ and the *support* of $\tilde{A}$ is the crisp set $S_{\tilde{A}} = \{x \in \mathcal{U} : \tilde{A}(x) > 0\}$, i.e. the set of elements which somewhat belong to $\tilde{A}$.

A fuzzy set $\tilde{A}$ is called convex if the following holds:

$$\forall x \forall x_1 \forall x_2 \ \ x \in [x_1, x_2] \Rightarrow \tilde{A}(x) \geq \min(\tilde{A}(x_1), \tilde{A}(x_2)) \tag{19}$$