Analysing Imperfect Temporal Information in GIS Using the Triangular Model

Yi Qiang1, Delafontaine Matthias1, Tijs Neutens1, Birger Stichelbaut2, Guy De Tré3, Philippe De Maeyer1 and Nico Van de Weghe1

1Department of Geography, Ghent University, Ghent, Belgium. 2Department of Archaeology, Ghent University, Ghent, Belgium. 3Department of Telecommunications and Information Processing, Ghent University, Ghent, Belgium. Email: yi.qiang@ugent.be

Rough set and fuzzy set are two frequently used approaches for modelling and reasoning about imperfect time intervals. In this paper, we focus on imperfect time intervals that can be modelled by rough sets and use an innovative graphic model [i.e. the triangular model (TM)] to represent this kind of imperfect time intervals. This work shows that TM is potentially advantageous in visualizing and querying imperfect time intervals, and its analytical power can be better exploited when it is implemented in a computer application with graphical user interfaces and interactive functions. Moreover, a probabilistic framework is proposed to handle the uncertainty issues in temporal queries. We use a case study to illustrate how the unique insights gained by TM can assist a geographical information system for exploratory spatio-temporal analysis.

Keywords: rough sets, imperfect time interval, the triangular model, geographical information system, spatio-temporal analysis

INTRODUCTION

Every entity has an extent in time, such as the lifetime of an object or the duration of an event. These temporal extents are usually described by crisp time intervals bounded by a well-defined start and end point. However, under some circumstances, the temporal extent of an entity is imperfect, and cannot be adequately modelled by a crisp time interval. On the one hand, some events may start or end gradually and therefore their start and end points cannot be pinned to exact time stamps. For example, it is difficult to decide when the industrial revolution started and finished. Intervals of this kind of events are usually modelled by fuzzy sets (Zadeh, 1965) through the quantification of the graded truth of whether a time point is in the interval, bringing the concept of fuzzy time interval. On the other hand, in some other cases, it is only known that the start and end of a crisp interval are within certain ranges, but no extra information or assumptions about the distribution of the start and the end is available. Modelling this kind of imperfect time intervals with fuzzy sets would induce extra overhead and unnecessary complexity. In these cases, the alternative approach, i.e. rough sets (Pawlak, 1982), can excellently suit the modelling and handling of time intervals. Currently, a lot of disciplines are faced with the issue of imperfect time intervals, which is reflected in many contributions in modelling intervals by fuzzy sets (De Caluwe et al., 1997, 1999; Nagypál and Motik, 2003; Ohlbach, 2004; Schockaert et al., 2008; Garrido et al., 2009) and rough sets (Bittner, 2002; Bassiri et al., 2009). However, while most of this work focuses on modelling and reasoning about imperfect time intervals, techniques and tools for visualizing and analysing imperfect time intervals are still lacking. This probably stems from the conventional representation of time intervals (e.g. Gantt chart, time table and historical timeline) which represents time intervals as linear segments along a one-dimensional time line. The second dimension is often exploited merely to differentiate intervals of entities and has no temporal meaning. In this representation, the visual distribution of intervals is variable, according to the application of different ordering rules in the second dimension, for example, ordering intervals from the shortest to the longest, or from the earliest started to the latest started. The characteristics of the distribution of intervals cannot be observed in one single display. This is not convenient for visual pattern detection of time intervals, let alone imperfect time intervals with more complex structure.

This issue also exists in geographical information science (GIScience), which considers time as one of the most
important component of geographical information (Maceachren et al., 1999; Peuquet, 2002; Li and Kraak, 2008; Neutens et al., 2008). In the recent development of GIScience, considerable effort has been made in handling the temporal aspect of geographical data (Andrienko et al., 2003, Smith et al., 2007, Neutens et al., 2007). Due to the limitations of data acquisition techniques, spatial data are often linked to imperfect temporal information. Dealing with imperfect temporal information becomes an increasingly significant issue in spatio-temporal data analysis, particularly in exploratory spatio-temporal data analysis (ESTDA) which greatly relies on graphical representations and visualisations. Since most prevalent techniques and tools of ESTDA represent time in the linear form, their ability in dealing with temporal imperfection in spatio-temporal data is still not satisfactory.

To address these issues, attempts have been made to represent time in a two-dimensional (2D) space. For example, Keim et al. (2006) arranged time series of financial investments in a 2D representation called the growth matrix. This matrix is able to display the growth rates of investments in all possible sub-intervals in the time series. The TT-plot introduced by Imfeld (2000) applied a similar idea to analyse movement patterns. Besides these approaches, Kulpa (1997) proposed the ER diagram that represents time intervals as 2D points, and investigated the use of this diagram for interval reasoning and arithmetic. Later, Van de Weghe et al. (2007) named this representation the triangular model (TM) and applied it in an archaeological context. Recently, Qiang et al. (2009, 2010) have extended TM to represent time intervals modelled by rough sets and fuzzy sets. This work majorly focussed on the use of TM in temporal reasoning. The practical value of TM in visualizing and analysing imperfect time intervals is still yet to be exploited.

To fill this gap, this paper investigates the use of TM in visualizing and analysing imperfect time intervals. The focus is on imperfect time intervals that can be modelled by rough set theory. A probabilistic framework is proposed to model the uncertainties in the temporal relations of such roughly described intervals. In order to better demonstrate the use of TM, a prototype tool is introduced, which implements TM in a geographical information system (GIS). This tool support analysis of geographical entities with imperfect intervals. We show that the advantages of TM in visualizing and querying roughly described intervals can be better exploited through a computer application with a graphical user interface (GUI) and interactive functionalities. In addition, we use a case study to illustrate how the unique insights gained by TM can assist a GIS for ESTDA involving imperfect time intervals.

In the remainder of the paper, we first introduce the basic concept of TM and how rough approximations of intervals are represented in TM. In Section on ‘Probabilities of Rough Approximations’, we apply probability theory to model uncertainties in temporal queries of imperfect intervals. Section on ‘Implementation’ presents GeoTM, including its GUI, functionalities and supported data model. In Section on ‘Case Study’, the use of GeoTM is demonstrated in a concrete use case. The paper ends with a brief conclusion and an outline of avenues for future work.

Figure 1. The transformation from the linear representation to triangular model (TM): (a) the linear representation of time intervals, (b) the construction of an interval point in TM, (c) the TM representation of time intervals

All notations and symbols used in this paper are summarized in the table in the Appendix.

TRIANGULAR MODEL

Representation of time intervals

A time interval $I$ is usually modelled as a convex set of real numbers, i.e. $[I^-, I^+]$ with $I^- < I^+$. $I^-$ and $I^+$ respectively denote the start and end of $I$. In the traditional linear representation, a time interval is represented by a finite linear segment bounded by $I^-$ and $I^+$ (see Figure 1a). This linear representation of time intervals is widely used in our daily life, for example time tables and historical time lines. The transformation from the linear representation of a time interval to TM can be achieved by constructing two lines
Allen (1983) specified 13 possible relations between two time intervals (see Table 1), which are referred to as Allen relations. In TM, every Allen relation can be represented as a specific zone (Kulpa, 1997). Given a study interval $I_0 = [0, 100]$, in TM all examined intervals are located within the isosceles triangle formed by $I_0$, $I_1$, and $I$. Let us consider, for example, a reference interval $I_2 = [33, 66]$. Any intervals (e.g. $I_{1a}$, $I_{1b}$ and $I_{1c}$) before $I_2$ (Figure 2a) are located in the triangular area in the left corner of the study area (Figure 2b). Therefore, it is easy to deduce that all intervals $I_1$ before $I_2$ must be located in the black zone in Figure 2c, which is called the before zone of $I_2$. Likewise, all Allen relations with respect to an interval can be represented by zones in $\mathbb{R}^1$ (Figure 3), which are called relational zones. For each relation in Figure 3, the reference interval $I$ has been chosen in the centre of the study period in order to avoid visual bias. Each relational zone represents the set of intervals that are in a specific relation to the reference interval $I$, which are denoted as $\text{Rel}(I)$. For example, the duration zone of $I$ represents the set of intervals that are during $I$ and is denoted as $\text{during}(I)$. On the other hand, $\text{Rel}(I_1, I_2)$ expresses the statement that $I_1$ is in the relation $\text{Rel}$ to $I_2$. For example, $\text{overlaps}(I_1, I_2)$ expresses that $I_1$ overlaps $I_2$.

**Representation of rough approximation**

Incomplete information may result in uncertainties about the exact start and end of a time interval. This may happen in many observation activities in which data are acquired at discrete time stamps, such as images or photographs in remote sensing. From a sequence of images, for example, one can determine whether a feature exists at specific time stamps. However, the feature’s status between two time stamps is unknown. With these discrete snapshots, the interval of the feature’s existence is thus imperfect. If there is no prior knowledge about the distribution of the start and end, modelling this kind of imperfect time intervals with membership functions of fuzzy sets would induce extra overhead and unnecessary complexity. In these cases, rough sets can be considered as an appropriate and adequate solution (Bittner, 2002; Bassiri et al., 2009).

In the rough set approach, an imperfect time interval $I$ is described by an upper approximation $\bar{I}$ and a lower approximation $\underline{I}$, with $\underline{I} \subseteq \bar{I}$ (Figure 4). We call such a pair of $\bar{I}$ and $I$ the rough approximation of $I$, which is denoted as $\text{R}(I)$. Time points in $\bar{I}$ are definitely in $I$, whereas all time points not in $\bar{I}$ are definitely not in $I$. $I$ is bounded by the earliest possible start $I^-$ and the latest possible end $I^+$, while $\bar{I}$ is bounded by the latest possible start $I^-_\bar{I}$ and the earliest possible end $I^+_\bar{I}$. In between $\bar{I}$ and $I$, there are two rough boundaries $\text{R}(I^-)$ and $\text{R}(I^+)$ gathering the time points that are possibly in $I$. Unlike fuzzy time intervals which define the extent to which the time points are possibly in $I$, the rough approximation of $I$ classifies time points into definitely in, definitely not in and possibly in $I$.

In the linear representation, the rough approximation of an interval is usually represented as a tripartite linear segment (Figure 5a). However, in TM, a rough approximation of an interval is represented by a convex 2D

---

**Table 1.** Thirteen Allen relations (Allen, 1983)

<table>
<thead>
<tr>
<th>Relation</th>
<th>Allen Notation</th>
<th>Linear Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_1$ equal $I_2$</td>
<td>$I_1 = I_2$</td>
<td>$I_1^- = I_2^-$ and $I_1^+ = I_2^+$</td>
</tr>
<tr>
<td>$I_1$ starts $I_2$</td>
<td>$I_1 &lt; I_2$</td>
<td>$I_1^+ &lt; I_2^-$ and $I_1^+ &lt; I_2^+$</td>
</tr>
<tr>
<td>$I_1$ starts-by $I_2$</td>
<td>$I_1 &lt; I_2$</td>
<td>$I_1^+ &lt; I_2^+$ and $I_1^+ &lt; I_2^+$</td>
</tr>
<tr>
<td>$I_1$ finishes $I_2$</td>
<td>$I_1^+ &gt; I_2$</td>
<td>$I_1^- &gt; I_2^-$ and $I_1^- &gt; I_2^+$</td>
</tr>
<tr>
<td>$I_1$ finishes-by $I_2$</td>
<td>$I_1^+ &gt; I_2$</td>
<td>$I_1^- &gt; I_2^-$ and $I_1^- &gt; I_2^+$</td>
</tr>
<tr>
<td>$I_1$ meets $I_2$</td>
<td>$I_1 = I_2$</td>
<td>$I_1 = I_2$</td>
</tr>
<tr>
<td>$I_1$ meets-by $I_2$</td>
<td>$I_1 = I_2$</td>
<td>$I_1 = I_2$</td>
</tr>
<tr>
<td>$I_1$ overlaps $I_2$</td>
<td>$I_1 &gt; I_2$ and $I_1 &lt; I_2$</td>
<td>$I_1^+ &lt; I_2^-$ and $I_1^- &lt; I_2^+$</td>
</tr>
<tr>
<td>$I_1$ overlapped-by $I_2$</td>
<td>$I_1 &gt; I_2$ and $I_1 &lt; I_2$</td>
<td>$I_1^+ &lt; I_2^-$ and $I_1^- &lt; I_2^+$</td>
</tr>
<tr>
<td>$I_1$ during $I_2$</td>
<td>$I_1 &gt; I_2$</td>
<td>$I_1^+ &lt; I_2^+$ and $I_1^- &lt; I_2^+$</td>
</tr>
<tr>
<td>$I_1$ contains $I_2$</td>
<td>$I_1 &gt; I_2$</td>
<td>$I_1^+ &lt; I_2^+$ and $I_1^- &lt; I_2^+$</td>
</tr>
<tr>
<td>$I_1$ before $I_2$</td>
<td>$I_1 &lt; I_2$</td>
<td>$I_1 &lt; I_2$</td>
</tr>
<tr>
<td>$I_1$ after $I_2$</td>
<td>$I_1 &lt; I_2$</td>
<td>$I_1 &lt; I_2$</td>
</tr>
</tbody>
</table>
To construct a rough approximation \( R(I) \) in TM, two parallel lines are projected from \( I^- \) and \( I^+ \) with angle \( \alpha \) to the horizontal axis, and the other two lines are projected from \( I^+ \) and \( I^- \) with angle \( -\alpha \) to the horizontal axis. These four lines form a rectangle (Figure 5b) which indicates a zone where the exact interval \( I \) can be found. In other words, this zone represents the set of intervals that have an uncertain start and an uncertain end, temporal queries about such rough time intervals may result in uncertain answers. To handle the uncertainties in temporal queries, a probabilistic framework is needed. Given a rough approximation \( R(I) \), \( R(I^-) \) and \( R(I^+) \) define the ranges where the exact start \( I^- \) and the exact end \( I^+ \) can be found (equations (1) and (2)). As there is no further knowledge about the distribution of \( I^- \) and \( I^+ \) in \( R(I^-) \) and \( R(I^+) \), we assume that the probability distributions in \( R(I^-) \) and \( R(I^+) \) are uniform. That is to say, every time point in \( R(I^-) \) has the same probability of being equal to \( I^- \) (i.e. equation (3)). Therefore, the probability density of \( x=I^- \) [where \( x \in R(I^-) \)] is the quotient of 1 and the cardinality of \( R(I^-) \) (equation (4)). In the same way, every time point in \( R(I^+) \) has the same probability of being equal to \( I^+ \) (equation (5)), and therefore the probability density of \( x=I^+ \) [where \( x \in R(I^+) \)] is the quotient of 1 and the cardinality of \( R(I^+) \) (equation (6)). Based on these prerequisites, it can be deduced that every interval in \( R(I) \) has the same probability of being equal to \( I \) (equation (7)).

Because \( R(I) \) defines the interval set that includes \( I \) (equation (8)), the probability density of \( x=I \) [where \( x \in R(I) \)] is the quotient of 1 and the cardinality of \( R(I) \) (equation (9)). Therefore, \( R(I) \) can be considered as an interval set in which every interval has the same probability of being equal to \( I \). Because TM maps time intervals into points in a 2D space, the cardinality of an interval set is proportional to its area, and therefore the probability density of \( x=I \) [where \( x \in R(I) \)] is inversely proportional to the area of \( R(I) \) in TM (see equation (9)).

According to equation (9), it can be deduced that, given a set of intervals \( A \) and a rough approximation \( R(I) \) (such as Figure 6), the probability that \( I \) is in \( A \) is the ratio between the cardinality of \( A \cap R(I) \) and the cardinality of \( R(I) \) (equation (10)). In the TM representation, this probability is the ratio between the area of \( A \cap R(I) \) and the area of \( R(I) \) (equation (10)). If \( R(I) \) is totally contained in \( A \) (i.e. \( R(I) \subseteq A \)), the probability that \( I \) is in \( A \) is 1. If \( R(I) \) and \( A \) do not intersect [i.e. \( R(I) \cap A = \emptyset \)], the probability that \( I \) is in \( A \) is 0. The interval set \( A \) can also be a relational zone of a certain interval, e.g. \( \text{before}(I) \). In this case, the probability that \( I \) is in \( A \) can be interpreted as the probability that \( I \) is \( \text{before}(I) \). With respect to the principles of probability theory (Jaynes and Brethorst, 2003), the probabilities of multiple rough approximations can be deduced. For example, given an interval set \( A \) and two independent rough approximations \( R(I_1) \) and \( R(I_2) \) (Figure 7), the probability that only one of \( I_1 \) and \( I_2 \) is in \( A \) can be obtained from equation (11), and the probability that both \( I_1 \) and \( I_2 \) are in \( A \) can be obtained by equation (12). Analogously, given an interval set \( A \) and \( n \) independent rough approximations, i.e. \( R(I_1), R(I_2), \ldots, R(I_n) \), the probability that \( m \) intervals \( 0 \leq m \leq n \) are in \( A \) can be obtained by equation (13). Since interval sets and rough approximations of intervals are represented as 2D geometries.
in TM, the probability that an interval is in an interval set is expressed by the overlap ratio of their geometries. This feature of TM is similar to Venn diagrams (Venn, 1881) in which the areas represent the occurrences of a certain event and the intersection of areas expresses the coincidence of events. Compared with mathematic formulas, we contend that such graphic representation is more intuitive for human beings. Hence, TM can be considered as a promising platform for visual queries of intervals described by rough approximations.

\[
\sum_{x \in R(I^-)} P(I^- = x) = 1 \\
\sum_{x \in R(I^+)} P(I^+ = x) = 1 \\
P(I^- = x_1) = P(I^- = x_2), \quad x_1, x_2 \in R(I^-) \\
f_I^-(x) = P(I^- = x) = \frac{1}{|R(I^-)|}, \quad x \in R(I^-) \\
P(I^+ = x_1) = P(I^+ = x_2), \quad x_1, x_2 \in R(I^+) \\
f_I^+(x) = P(I^+ = x) = \frac{1}{|R(I^+)|}, \quad x \in R(I^+) \\
P(I = I_1) = P(I = I_2), \quad I_1, I_2 \in R(I) \\
\sum_{x \in R(I)} P(I = x) = 1 \\
f_I(x) = P(I = x) = \frac{1}{|R(I)|} \times \frac{1}{\text{Area}(R(I))}, \quad x \in R(I) \\
P(I \in A) = \sum_{x \in A} \frac{1}{|R(I)|} = \frac{|A \cap R(I)|}{|R(I)|} = \frac{\text{Area}(A \cap R(I))}{\text{Area}(R(I))}
\]
unique identifiers (ID) of entities. GeoTM is built on other attributes. The spatial locations, temporal extents and intervals described by rough approximations. Besides spatial and temporal extents, these geographical entities may have other attributes. The spatial locations, temporal extents and other attributes are linked to geographical entities by unique identifiers (ID) of entities. GeoTM is built on top of ArcGIS\textsuperscript{TM} which is a desktop GIS produced by ESRI (Redlands, CA, USA). Within its object model (ArcObjects), developers can call on existing functions and components of ArcGIS\textsuperscript{TM} to develop customized applications. Consequently, GeoTM is compatible with the supported data formats in ArcGIS, such as shapefile and dBASE files. In the representation of geographical entities, spatial locations and geometries of time intervals are stored in two shapefiles, while the attributes are stored in a database table (i.e. a dBASE file). Figure 8 gives an overview of how geographical entities are modelled and represented in GeoTM.

\textbf{GUI of GeoTM}

The user interface of GeoTM consists of a map view, a TM view and controls that can trigger specific functions or user forms (Figure 9). The map view is used to display spatial locations of geographical entities, which are modelled as points, lines or polygons. The TM view displays time intervals of entities, using the TM representation. Rough approximations of intervals are represented as polygons. In the TM view, there may exist overlaps of polygons. In order to display the pattern of overlapped polygons, gradual colours are assigned to areas according to the numbers of overlaps: a darker colour is assigned to the areas with more overlaps. In Figure 9, a dataset of military features in the World War 1 (WWI) is displayed in GeoTM. The spatial locations of these features are displayed in the map view, while their lifetimes (represented by rough approximations of intervals) are displayed in the TM view. In this section, we use this dataset to illustrate the functionalities of GeoTM. The entire scenario of this dataset will be described in Section on ‘The Case Study’. The temporal resolution of the dataset is one day. Thus, in this context, a time interval is the period between two specific dates formatted as year/month/day.

Users can interactively browse the map view and TM view by zooming and panning. In the map view, objects can be selected by dragging a rectangle. In the TM view, special selection tools have been designed to select intervals according to specific temporal queries, which will be described in detail in the next sub-section. Other attributes of entities are stored in the attribute table that can pop up when the user clicks the corresponding button. ‘Linked brushing’ is supported among the map view, the TM view and the attribute table. This function allows selecting objects from any of these three views and dynamically updating the other two views to highlight the corresponding objects. With this function, geographical entities in GeoTM can be queried with spatial, temporal and attribute constraints. Additionally, many common functions of a conventional GIS are supported in GeoTM. For example, several datasets can be loaded into GeoTM as multiple layers.

\textbf{Temporal queries in GeoTM}

As introduced in Section on ‘Representation of temporal relations’, intervals that satisfy a certain temporal relation are located within a relational zone. Thus, queries based on temporal relations can be modelled as specific zones in TM,
A query zone can be a relational zone or combinations (e.g. intersection and union) of multiple relational zones. Whether an interval satisfies the temporal query depends on whether the interval point lies within the query zone. However, when intervals are described by rough approximations, in TM the rough approximation can be partially in the query zone. If this is the case, temporal queries can be answered with probability thresholds. For example, within a dataset of rough approximations of intervals, one can select all intervals that have more than 90% probability of being during an indicated interval. In TM, this query is expressed by selecting rough approximations that have more than 90% probability of being in the before zone of the indicated interval. Therefore, in the TM view of GeoTM, querying tools are available to define temporal queries by creating query zones. On the one hand, users can indicate an interval by moving the cursor to a specific position in the TM view. When right-clicking on this position, a drop-down menu of Allen relations appears. Next, by clicking an Allen relation in the menu, all intervals in this relation are selected, according to a pre-defined probability threshold. Figure 10 shows an example of selecting intervals that are probably before [1916/06/25, 1917/05/31]. In this example, the lower probability threshold is 0.6 and the upper probability threshold is 1. As a result, intervals with more than 60% probability before [1916/6/25, 1917/5/31] are selected.

Besides temporal queries of Allen relations, some other queries can be made by dragging specific geometries in the TM view. For example, a convex set of intervals can be selected by dragging a rectangle whose sides are in z or −z angle to the horizontal axis (Figure 11a). A convex interval set is defined as the set of time intervals in-between two different time intervals (Kulpa, 2006). We developed this query tool because convex interval sets can be easily interpreted by Allen relations or combinations of Allen relations. \( \ll I_1, I_2 \gg \) is used to denote a convex interval set in-between \( I_1 \) and \( I_2 \). The formal definition of \( \ll I_1, I_2 \gg \) can be found in equation (14). Moreover, two parallel lines can be dragged in z or −z angle with the horizontal axis, in order to select intervals that start-within or end-within a certain interval (e.g. Figure 11b and c). start-within \( (I_1) \) is defined as a set of intervals whose starts are in \( I_1 \) (equation (15)). Analogously, end-within \( (I_1) \) is defined as a set of intervals whose ends are in \( I_1 \) (equation (16)). Though the start-within and end-within can be expressed by unions of Allen relations, we design such query tools for people that are more accustomed to the expressions of ‘starts within’ and ‘ends within’ than unions of Allen relations. Furthermore, by dragging a range along the vertical axis, users can select intervals whose durations are in a specific range (Figure 11d). In GeoTM, all queries are carried out with respect to pre-defined probability thresholds. In the examples in Figure 11, the lower threshold is
set to 0.6 and the upper threshold is set to 1, meaning that the selected intervals have more than 60% probability of satisfying these queries.

\[
\begin{align*}
\ll I_1, I_2 \gg & \equiv \{ I \mid \min(I^{-}, I_2^{-}) < I^{-} \\
& < \max(I^{-}, I_2^{-}) \wedge \min(I^+, I_2^+) < I^+ < \max(I^+, I_2^+) \} \tag{14} \\
\text{starts-within}(I_1) & \equiv \{ I \mid I^{-} < I^{-} < I_1^{+} \} \tag{15} \\
\text{ends-within}(I_1) & \equiv \{ I \mid I_1^{-} < I^+ < I_1^{+} \} \tag{16}
\end{align*}
\]

Based on equation (13), when a temporal query is defined, users may obtain the probability that a given number of intervals satisfy this query. For example, GeoTM can return the probabilities that a certain number of intervals are \textit{before} [1915/11/21, 1916/08/30]. In this case, 67 rough approximations have common parts with (intersect or within) the query zone, while 43 out of these 67 rough approximations are totally within the query zone. This means that there are minimum 43 intervals \textit{before} [1915/11/21, 1916/08/30] and maximum 67 intervals \textit{before} [1915/11/21, 1916/08/30]. The probability that \( n \) intervals \((43 < n < 67)\) are \textit{before} [1915/11/21, 1916/08/30] is between 0 and 1. GeoTM can automatically generate the probability that \( n \) intervals are \textit{before} [1915/11/21, 1916/08/30], from \( n = 43 \) to \( n = 67 \) (Figure 12). In this way, one can know how many intervals satisfy the query, with respect to a confidence level. According to Figure 12, there are 59 intervals \textit{before} [1915/11/21, 1916/08/30], at 0.9 confidence level. This function is also useful in distinguishing ‘real clusters’ and ‘fake clusters’ of intervals. Because the colours in TM represent the number of overlaps, which is actually the maximum number of intervals within the area, an area with a darker colour indicates a potential cluster of intervals. However, the probability that intervals are clustered in this area varies. By dragging a query zone over the dark area, one can be aware of the numbers of intervals clustered in this area with respect to specific probabilities.

**CASE STUDY**

Having introduced basic functionalities of GeoTM in the previous section, this section will illustrate the use of TM in supporting GIS for ESTDA by means of a case study.

**Dataset**

During WWI, a large number of aerial photos in West-Flanders (Belgium) were taken at discrete time stamps by all combating nations as an intelligent tool to collect information on the enemy’s intentions. These aerial photographs are preserved in archives all over the world. The largest collections are held at the Belgian Royal Army Museum, the Imperial War Museum, the Australian War Memorial and the Bavarian Military archive. From these aerial photos, one can observe whether a military feature (e.g. a fire trench, gun position or barrack) was not yet present, present, or destroyed (Stichelbaut and Bourgeois, 2009). Although the state of a feature is uncertain between two snapshots, we assume that it does not change between two snapshots which show similar states. Certainly, this assumption relies on our knowledge that snapshots are dense enough to capture most of features’ changes. When more volatile entities are considered, an appropriate temporal resolution will be required. In this context, rough sets are excellently suited for temporal modelling, because no knowledge is available about the state of a feature between two snapshots. Thus, we can consider a period of
snapshots showing similar states as the lower approximation, its neighbouring uncertain intervals as boundary regions, and all of them form the upper approximation (Figure 13). Thus, a feature’s lifetime can be meaningfully represented by a rough approximation.

More specifically, four photos determine the lifetime of a feature (Figure 13): (1) the last photo in which the feature is not yet present; (2) the first photo in which the feature is present; (3) the last photo in which the feature is present; and (4) the first photo in which the feature is destroyed or abandoned. The interval between the dates of photos (2) and (3) is the lower approximation of the feature’s lifetime, while the interval between the dates of photo (1) and photo (4) is the upper approximation. Intervals between the dates of photos (1) and (2), and intervals between the dates of photos (3) and (4) are the boundary regions, which indicate respectively the range of the feature’s construction and destruction/abandonment dates. There are a few exceptions, where a feature was not yet present in one photo and already destroyed in the following photo. In these cases, photos (2) and (3) are missing, and therefore the rough approximation has an empty lower approximation. As described in the Section on ‘Representation of rough approximation’, such rough approximations are represented as triangles on the horizontal axis. In this case study, a rectangular area (around 3 \times 3.5 \text{ km}) is selected near Ypres (Belgium) as the study area, containing 2466 military features (Figure 14). This study area near Ypres is one of the most important battlefields of WWI. From 1914 to 1918, there was constant artillery fighting going on. And more importantly, Ypres was the scene of several large offensive actions by both Germans and Allies. Therefore, it is an ideal area to test the potential of TM in exploratory analysis.
For analysis, we also take account of spatio-temporal information of the frontlines during the war. There are 11 snapshots of the states of the German and Allied frontlines during WWI. From these snapshots, we have observed that the frontlines were stable in $I_2 \{1915/5/25, 1916/6/14\}$ (we format a date as year/month/date), but had significant shifts in three time intervals, i.e. $I_1 \{1915/5/23, 1915/5/25\}$, $I_3 \{1916/6/14, 1917/10/1\}$ and $I_4 \{1917/10/1, 1918/4/15\}$ (Figure 15). Note that the frontline shifts in $I_3$ and $I_4$ might only take a few days. But from snapshots of frontlines, it is only known that the shifts happened in these three intervals and the exact dates of the shifts are unavailable. Figure 15 displays the locations and shifts of frontlines in these four intervals. Note that the German army was always on the east side of the Allied army. If the positions of frontlines were out of the study area, such as in $I_3$ and $I_4$, we use arrows to indicate the direction of the shifts.

Analysing rough time intervals in GeoTM

First, the dataset of the military features is imported to GeoTM. The spatial locations of these features are displayed in the map view, while the rough approximations of their intervals are displayed in the TM view (Figure 16). To identify the intervals of frontline shifts, the interval points and during zones of $I_1$, $I_2$, $I_3$ and $I_4$ are added to the TM view (Figure 16). The zone with solid boundary is the during zone of the interval in which the frontlines were stable, i.e. $I_2$, while the zones with dashed boundaries are the during zones of the intervals in which the frontlines had significant shifts, i.e. $I_1$, $I_3$, and $I_4$. As $I_1$ is very short, its during zone is invisible in this scale. Using the selection tool, we find that 99% of features (2433 out of 2466) in the study area were built after $I_1$. Considering the frontline shift in $I_1$, we can infer that the major military activities in the study area started after $I_1$ when the frontline moved from the edge of the study area to the centre. Three dark areas that indicate potential clusters of intervals can be observed in TM (Figure 16). In the next three sub-sections, these potential clusters will be analysed with respect to frontline states in $I_2$, $I_3$ and $I_4$. Note that in the following three sub-sections, we set the lower probability threshold of all temporal queries as 0.9 and the upper probability threshold as 1, in order to select intervals that satisfy the query with more than 90% probability.

Stable period of frontlines

Most intervals in Cluster 1 and Cluster 2 start-within $I_2$, when the frontlines were relatively stable. When selecting intervals probably start-within $I_2$ (more than 0.9

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure12.png}
\caption{The line diagram of probabilities that at least \( n \) intervals are before \([1915/11/21, 1916/08/30]\).}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure13.png}
\caption{The rough approximation of the lifetime of a military feature.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure14.png}
\caption{The location of the study area.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure15.png}
\caption{Figure 15. The location of the study area}
\end{figure}
Figure 15. The states of German and Allied frontlines in World War I. The map in the solid box shows the period during which the frontlines were relatively stable, while the maps in the dashed boxes show periods during which the frontlines had significantly shifted.

Legend:
- Allied frontline (earlier):
- Allied frontline (later):
- German frontline (earlier):
- German frontline (later):
- Military features (German or Allied):

$I_1$: [1915/5/23, 1915/5/25]
$I_2$: [1915/5/25, 1916/6/14] (Stable period)
$I_3$: [1916/6/14, 1917/10/1]
$I_4$: [1917/10/1, 1918/4/15]
Figure 16. The rough approximations of the feature lifetimes in the triangular model view, with the *during* zones of $I_1$, $I_2$, $I_3$ and $I_4$ displayed.

Figure 17. The selection of intervals that *end-within* $I_2$. 

*Lower probability: 0.9, Upper probability: 1.0*
probability), we obtain 1531 intervals (62% of all features in the study area), which reflects that a majority of military features were constructed during this stable period. There are also some intervals that end-within $I_2$, which are lifetimes of features that were destroyed within $I_2$. By selecting intervals that end-within $I_2$, we obtained 200 intervals (8% of all features) and the map view shows that all these features were close to the frontlines (Figure 17). This is reasonable because military features close to the frontline were easier to be destroyed by artillery attacks or minor offensives. Therefore, it can be inferred that during $I_2$ there were no significant military actions, and the armies at both sides were building military features in order to keep their positions. Only a small portion of features close to the frontlines were destroyed.

Shift of frontlines in $I_3$

During $I_3$, the frontlines shifted from their positions to the east of the study area, towards the German side. In the TM view, it is observable that the intervals in Cluster 2 end-within $I_3$. More specifically, the intervals in Cluster 2 end-within a quite short interval [1917/07/10, 1917/08/01], which results in a slim shape (see Cluster 2 in Figure 16). When selecting intervals that end-within [1917/07/10, 1917/08/01], we obtain 1050 features, which is 43% of all features (2466 features) in the study area. This finding reflects that a large number of features were intensively destroyed during [1917/07/10, 1917/08/01]. The map view shows that all these features were distributed in the eastern side of the old frontlines, which were earlier occupied by the German army (Figure 18). By checking the attribute table, one can see that 99% of these 1050 features were German features, which was 83% of all German features (1260 features) in the study area. With these findings, it becomes possible to infer, without consulting any historical documents, that an overwhelming and intensive destruction happened to German features in July 1917 in the study area. Probably due to this destruction, the German army lost its area and was pushed to the east. Referring to historical literature (Barton, 2005; Verbeke, 2006), these findings reflect the fact that during this period, the Allies intensively destroyed German features using artillery fire, in a battle which became known as the Third Battle of Ypres. During the battle, Allied army took Passchendaele (Belgium), pushing the frontline towards the east.

Shift of frontlines in $I_4$

Through observation in the TM view, most intervals in Cluster 1 and Cluster 3 end-within $I_4$. When selecting intervals probably end-within $I_4$ (with more than 90% probability), 817 features are obtained and the attribute table shows that all these features are Allied. This means that 68% Allied features (817 out of 1206) in the study area were probably destroyed within $I_4$. Considering that within $I_4$ the frontlines have moved from east to west towards the
Allied side, it is straightforward to infer that these features were destroyed due to the offensive by the German army. Features in Cluster 1 were built earlier (i.e. during \( I_2 \)), when the frontlines were relatively stable. By selecting features in Cluster 1, one can see that these features are mostly located in the eastern part of the study area, which was occupied by the Allied army during \( I_2 \). Features in Cluster 3 were built later (i.e. during \( I_4 \)), and were evenly distributed over the entire study area (Figure 19). These findings reveal that after the Third Battle of Ypres, the Allied army controlled the whole study area and built military features over the area (i.e. features in Cluster 3). Later on, the German army attacked back and destroyed all Allied features (features in both Cluster 1 and Cluster 3) in the study area. After this military action, the frontlines shifted from the east of the study area to the west of the study area. Known from historical literatures (Howard, 2002), these findings probably reflect the Battle of Lys (a part of Spring Offensives of Germany) in April 1918, during which the German army attacked the Allied army and pushed the frontline back to the west.

**CONCLUSIONS AND FUTURE WORK**

This paper has investigated the use of TM in visualizing and analysing time intervals modelled by rough sets. Compared with the fuzzy set approach, the rough set approach is excellently suited to imperfect time intervals with no prior knowledge and assumptions about the distribution of their starts and ends, which broadly exist in discrete data acquisitions. In TM, rough approximations of intervals are represented as polygons in a 2D space. We contend that this representation is advantageous in visualizing the distribution of imperfect time intervals, because the patterns (e.g. clusters) in the distribution can be explicitly displayed. If an interval is described by rough approximation, its relation with other intervals may have uncertainties. Due to these uncertainties, temporal queries may generate uncertain answers. Therefore, we have proposed a probabilistic framework to model uncertainties in temporal relations of roughly described intervals. Since TM represents temporal relations as 2D geometries, the probability of temporal relations can be expressed by the overlap rate of the corresponding geometries, which is analogous to Venn diagrams. Compared with mathematical expressions, such graphical representation is potentially more intuitive to human beings and can offer a promising basis for visual temporal queries.

In order to evaluate the capabilities of TM, we have implemented it into a prototype tool (i.e. GeoTM) which incorporates TM within a GIS. GeoTM shows that the advantages of TM in visualisation and querying can be better exploited when it is implemented in a computer program. Instead of providing a static TM visualisation, GeoTM supports interactive functionalities that enable flexible manipulations of the TM display. These functionalities have potential of assisting visual observation and
pattern detections. Besides visualisation, GeoTM also offers possibilities to make temporal queries by creating 2D geometries. Whether intervals satisfy the query depends on the extent to which the intervals are part of the query zone. Moreover, query zones can be directly created on top of the visualisation of intervals so that users can select an observed cluster of intervals by designing a proper temporal query. In order to handle the uncertainties in temporal relations of roughly described intervals, special functions based on the proposed probabilistic framework have been applied in GeoTM, allowing users to make temporal queries with consideration of probabilities. On the one hand, temporal queries can be defined with probability thresholds, which ensure that every retrieved interval satisfies the query in a certain probability. On the other hand, once a query is made, GeoTM can return the probability that a certain number of intervals satisfy the query. With this feature, users can be aware of how many intervals satisfy the query, with respect to a certain confidence level. With these functionalities, GeoTM is able to support ESTDA of geographical entities with reference to roughly described intervals. As shown in the case study of WWI features, GeoTM can be used to explore the military features from spatial, temporal and attribute aspects. From the detected patterns and clusters, users can discover interesting phenomena in the war without consulting historical literatures.

Not only spatio-temporal data, we believe that TM can be applied in other contexts that involve roughly described intervals for purposes of information visualisation, exploratory analysis and data mining. In future work, the applicability of TM needs to be further assessed by more use cases, preferably covering different research contexts. Also, future extensions will improve the usability and interactivity of the implementation. Before the implementation is released to a broader community, its scalability needs to be systematically evaluated. Furthermore, we plan to investigate the possibility of representing and analysing fuzzy time interval by TM. The fuzzy set approach is more applicable in other contexts that involve roughly described intervals for purposes of information visualisation, exploratory analysis and data mining. In future work, the applicability of TM needs to be further assessed by more use cases, preferably covering different research contexts. Also, future extensions will improve the usability and interactivity of the implementation. Before the implementation is released to a broader community, its scalability needs to be systematically evaluated. Furthermore, we plan to investigate the possibility of representing and analysing fuzzy time interval by TM. The fuzzy set approach is more suitable for modelling imperfect time intervals when assumptions or knowledge of the distributions of starting points and end points are available. This extension may rely on the application of more advanced visualisation techniques such as 3D visualisation.

BIOGRAPHICAL NOTES

Mr Yi Qiang (1983) is born in China. He started his study on geographical information systems from Beijing Normal University. After the 4-year undergraduate study, he has obtained his bachelor degree in GIS. And then, he got his master degree in geographical information science in the University of Edinburgh. In this master programme, he initiated the research about the application of the Semantic Web in environmental modelling. After the master programme, he worked in the UK National e-Science Institute in Edinburgh for 3 months. During this period, he was involved in a research of building a general knowledge infrastructure using the technologies of Semantic Web. Afterwards, he went to Belgium and started to work in the Department of Geography in Ghent University as a PhD candidate till present. In Ghent, he has been working on a research project funded by the Research Foundation – Flanders (FWO). Yi Qiang is interested in temporal reasoning, information visualisation and spatio-temporal analysis.

ACKNOWLEDGEMENT

The research work of Yi Qiang, Matthias Delafontaine and Tjits Neuteuns is funded by the Research Foundation – Flanders (FWO).

REFERENCES


### APPENDIX: THE SUMMARY OF NOTATIONS USED IN THE MANUSCRIPT

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Name of symbols</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>I−</td>
<td>A crisp time interval</td>
<td></td>
</tr>
<tr>
<td>I+</td>
<td>The start of I</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>The upper approximation of I</td>
<td></td>
</tr>
<tr>
<td>I−</td>
<td>The start of I</td>
<td></td>
</tr>
<tr>
<td>I+</td>
<td>The end of I</td>
<td></td>
</tr>
<tr>
<td>R(I)</td>
<td>The rough approximation of I</td>
<td></td>
</tr>
<tr>
<td>R(I−)</td>
<td>The earlier boundary region or the rough start of I</td>
<td></td>
</tr>
<tr>
<td>R(I+)</td>
<td>The later boundary region or the rough end of I</td>
<td></td>
</tr>
<tr>
<td>fI(x)</td>
<td>The probability density function of X</td>
<td>fI(x) is the probability density function of I−.</td>
</tr>
<tr>
<td>P(X)</td>
<td>The probability that the statement X is true</td>
<td>P(1I=I) denotes the probability that I1=I is true.</td>
</tr>
<tr>
<td></td>
<td>The cardinality of the set A, i.e. the total number of elements in A. A can be a set of real numbers or a set of intervals.</td>
<td></td>
</tr>
<tr>
<td>Rel(I)</td>
<td>A relational zone of I or the set of intervals that are in a certain relation to I.</td>
<td>before(I) is the before zone of I and denotes the set of intervals that are before I.</td>
</tr>
<tr>
<td>Rel(I1, I2)</td>
<td>The statement that I1 and I2 are in a certain relation.</td>
<td>overlap(I1, I2) expresses the statement that I1 overlaps I2.</td>
</tr>
<tr>
<td>I1 &lt; I2</td>
<td>The set of intervals that are in-between I1 and I2.</td>
<td>The formal definition of in-between can be found in equation (14).</td>
</tr>
</tbody>
</table>