Towards a Logic-based Framework for Representing and Reasoning about Geographic Phenomena

Cláudio E. C. Campelo

[1]Departamento de Sistemas e Computação
Universidade Federal de Campina Grande (UFCG)
Rua Aprígio Veloso, 882 – Bairro Universitário
58.429-140 – Campina Grande – PB – Brazil

campelo@dsc.ufcg.edu.br

Abstract. *This paper presents a logic-based approach to representing and reasoning about geographic phenomena. This approach is divided into two major parts. The former consists of a logical model of the Earth surface, which considers its spatial, temporal and thematic dimensions. This model provides improved integration of heterogeneous spatio-temporal data and enables the derivation of implicit data by means of logical inferences. The latter is a logical framework for representing and reasoning about geographic phenomena. This framework is based on the concepts of events and processes and on the relationships that may hold between them. The paper also presents a system prototype implemented to evaluate the applicability of the proposed theory. A case study about deforestation in Brazilian Amazon rainforest was conducted using this prototype, where events and processes are described in terms of changing spatial extensions of geographic features.*

1. Introduction

In the last decades, many different approaches have been proposed to modelling distinct kinds of geographic phenomena. During this period, the role that knowledge representation plays in developing modern Geographic Information Systems (GIS) have been increasingly recognised by the geographical information science community. Knowledge-based approaches can provide ways of deriving implicit data by means of logical inference; can enable spatio-temporal reasoning to help interpret complex phenomena; and can provide efficient querying mechanisms over spatio-temporal data.

This paper presents a logic-based approach to representing and reasoning about geographic phenomena. This approach can be separated into two principal parts. The former is a logical representation of the Earth surface that takes into account its spatial, temporal and thematic dimensions. This model provides improved integration of heterogeneous spatio-temporal data and enables the derivation of implicit data by means of logical inferences. It also includes a method of individuating geographic features and of grounding an spatio-temporal geographic ontology upon the data.

The latter, that works in conjunction with the former, is a logical framework that can be applied to represent and reason about a variety of geographic phenomena. This framework is based on the concepts of events and processes and on the relationships that may hold between them. Representing geographic phenomena in terms of *events* and *processes* has been suggested by many authors [Claramunt and Theriault 1996, Worboys and Hornsby 2004, Galton and Mizoguchi 2009, Devaraju and Kuhn 2010], and such concepts appear to be significant in the way humans reason about changes affecting geographic space. Many issues and challenges are encountered to develop an appropriate representation of these concepts. Some of these challenges are discussed in Section 2, whilst the developed framework is presented in Section 4.

Section 5 presents a system prototype implemented to evaluate the applicability of the proposed theory. The kind of phenomena that can be interpreted by the framework depends on the approach taken to represent geographic features, and on the variety of changes that such features are able to undergo over time. In order to evaluate the framework using the developed prototype, special focus has been placed on the representation of geographic phenomena that can be described in terms of changes affecting the spatial extension of geographic features. An example is the phenomenon of ‘urbanisation’, which can be described in terms of spatial changes affecting ‘built-up’ areas (e.g., appearance or expansion of these features). A case study about deforestation in Brazilian Amazon conducted using our prototype is also described.

2. Issues and Challenges

Establishing a suitable representation of geographic events and processes requires dealing with many different issues and challenges[1]. Challenges relate to the relationship between these concepts, between their types and particular instances, and also between them and their participant material objects. This is an unresolved field and therefore there has been many disagreements in the literature. Controversies relate to issues on the classification of events and processes as *endurants* or *perdurants* entities; to questions of whether they can be affected by temporal gaps; and to the discussion of whether they possess the characteristic of undergoing change over time. The debate also covers issues on the way events and processes relate to each other (e.g., whether one is a subclass of the other).

Entering into the existing debate was not the objective while developing the framework described here. Rather, the goal was to propose a representational approach which is in agreement with the semantic analysis discussed in previous work, and then evaluate the applicability of such a formalism for processing real geographic data, as a call for further research approaching the *application* of spatio-temporal formalisms to support the development of new generation of GIS with strong foundations in theory.

Geographic information can be affected by different kinds of vagueness, leading to considerable representational difficulties [Bennett 2010]. Such a representation task becomes particularly challenging when the temporal dimension is considered. Thus associating specific spatial and temporal boundaries with instances of events and processes requires an appropriate method of handling spatio-temporal vagueness. Although many approaches have been proposed to dealing with vagueness in geography, it seems that methods of handling spatio-temporal vagueness for representing and reasoning about geographic events and processes have not yet been sufficiently investigated. Our framework comprises an approach to handling spatial and temporal vagueness that is based on *Standpoint Semantics* [Bennett 2011].

[1]For a comprehensive review on the issues and challenges for representing geographic events and processes, see [Campelo and Bennett 2012].

Formal theories for modelling spatial changes, events and processes have been proposed. Nonetheless, most approaches are not particularly related to the geographic domain and their applicability to geographic space would require further developments. In addition, although some works provide important directions, most of them are not yet implemented, and therefore their suitability for processing real-world data is not often discussed. Implementing a system to evaluate such a logical framework with real data requires establishing a method of grounding the symbols upon elements of data. This requires work at multiple levels, both to select the appropriate set of predicates to be grounded and to formulate a suitable representation for the data. Methods of grounding geographic ontologies upon the data have been already proposed; however, approaches to developing an ontology grounded upon spatio-temporal data have not been sufficiently discussed in the literature, and therefore further investigations are still required.

3. Representing Spatio-temporal Data and Geographic Features

Our logic-based approach to modelling the Earth surface was named *STAR* (Spatio-temporal Attributed Regions)[2]. In this model, spatio-temporal data are stored as triples of the form $\langle a,\ g,\ s\rangle$, which corresponds to the fact that attribute a holds for geometry g at time instant denoted by timestamp s. A broad range of attributes can be associated with geometries. They can be used to describe either types of region coverage[3] (e.g., 'forested', 'arid', 'water covered') or types of geographic features (e.g., 'ocean', 'desert', 'forest').

Those triples are represented at the logical level by using the predicate $\mathsf{Star}(a, g, s)$. In addition, sortal predicates are employed to distinguish different types of attributes. Examples are the predicates $\mathsf{CAtt\text{-}Hom}(a)$ and $\mathsf{CAtt\text{-}Het}(a)$, which assert that a is an attribute representing a homogeneous and a heterogeneous type of coverage, respectively. Additionally, different logical relations are employed between types of attributes. For example, the relation *Can Contain* $\mathsf{CC}(a_1, a_2)$ determines that the part-hood relation *can* hold between *Stars* associated with attributes a_1 and a_2 (e.g. $\mathsf{CP}(urbanised, paved)$. Moreover, a set of axioms is specified to determine inference rules for deriving implicit data and to specify data storage constraints.

The STAR approach also comprises a method of individuating geographic features. This includes identifying the set of regions that denote the spatial extension of an individual feature f at a particular time instant; and the set of regions that denote the extension of f during a time interval in which the feature is said to exist. Of particular interest here are *geographic features* that can be modelled as the maximal connected region of some particular coverage, such as deserts (which can be defined based on the level of precipitation measured in distinct portions of the Earth surface). This can be inferred as the maximal well-connected region[4] of some particular coverage. Geographic features are regarded as a particular kind of *endurant entity*. They are discrete individuals and are able to undergo change while keeping their identity (which include loosing some of their parts).

[2]Additional details about this approach can be found in [Campelo et al. 2012].

[3]A type of *region coverage* is *not* restricted to types of land coverages. This can also denote *qualities* which can be measured (e.g., by sensors or human observation) and associated with a certain portion of the earth surface, such as 'hot' or 'arid'.

[4]The term 'well-connected region' is used here in agreement with the discussion and definitions given in [Cohn et al. 1997].

4. A Framework to Representing and Reasoning about Geographic Events and Processes

We developed a logical framework named *REGEP* (REasoning about Geographic Events and Processes), comprising formal descriptions of space, time, events, processes, geographic features and their related aspects. The syntax and semantics of this framework is described in detail in [Campelo and Bennett 2013]. In this framework, *events* are conceived as perdurant entities, that is, entities which are not subject to change over time. On the other hand, we regard a *process* as an entity which is subject to change over time, and therefore a process is not considered as a perdurant entity as defended by some authors. An event is usually associated with *precise temporal boundaries*, which may be denoted by the *culmination of a process* (i.e., when the goal in initiating it is realised) [Galton 2006]. We agree with the view that events and processes can be related by different forms. Hence, in this framework, we provide ways of representing events as a chunking of a process, whilst processes can be conceived as constituted of events.

The logical language $\Re$ used within the framework incorporates other existing formalisms, such as the *Allen's Interval Algebra* [Allen 1983] to represent time, the *Region Connection Calculus – RCC* [Randell et al. 1992] as the theory of space, and the *Standpoint Semantics* [Bennett 2011] to handle spatial and temporal vagueness. This language, specified using *First-Order Logic*, includes the predicates to represent event and process *classifiers*, *types* and *tokens*. Classifiers identify general categories of events and processes, independently of particular occurrences or participants. That is, a classifier describes something that might happen in space and time without specifying any temporal information or relating any type of geographic feature. *Natural language verbs* are usually applied to name these classifiers. Examples of such verbs are 'to fall', 'to expand' and 'to shrink'.

On the other hand, event and process types denote spatial changes involving a particular geographic feature, that is, types associate classifiers with individual features. An example of an event type is the expansion of the Atlantic Ocean. Finally, event and process tokens denote instances of event/process types, that is, they associate types with specific time intervals on which events are said to occur and processes are said to proceed. An example of an event token is the shrinkage of the Amazon rainforest from 01/01/2004 to 31/12/2004. Events and processes types are treated as complex nominals (i.e. functional terms). For example, an event type e is represented by $e = \mathsf{event}(v, f)$, where v is an event classifier and f a geographic feature which participates in this event. In addition, logical relations are provided to assert explicit relations which may hold between events and processes associated with given classifiers. For example, a fact of the form $\mathsf{Is\text{-}Chunk\text{-}Of}(v_1, b_1)$ relates an event classifier v_1 with a process classifier b_1. Asserting a fact using this relation means that the occurrence of an event (classified by v_1) on a given time interval i, is determined by the fact that a process (classified by b) proceeds on i.

The following principal predicates are used to reason about events and processes:

- $\mathsf{Occurs\text{-}On}(e, i)$, which asserts that an event of type e occurs on a time interval i.
- $\mathsf{Active\text{-}At}(p, t)$, meaning that a process of type p is active at a given time point t.
- $\mathsf{Proceeds\text{-}On}[a_{th}](p, i)$, meaning that a process of type p proceeds over a time interval i. This is true if the process is active at all time instants of sub-intervals i' of i, unless i' is shorter than a given activeness threshold a_{th}.

A number of different first-order logic queries can be formulated using the logical language $\Re$. Below are some examples of queries which can be easily specified using that language:

- Did any event of type e occur in region r after time instant t, whose duration was greater then 10 weeks?
- What are the sub-regions of region r where a process of type p has proceeded over the last month but it has not yet reached its culmination point?
- Was any process of type p active in region r at time instant t?
- Where and when did events of type e occur between years 2006 and 2008 (inclusive) whose affected regions do not overlap region r?

Beyond the use as a mechanism for directly querying spatio-temporal data, the REGEP framework can be used as the basis for the development of other systems applied to reason about different aspects of geographic phenomena. For example, systems which model causal relations between different geographic phenomena are often interested in identifying specific patterns of event occurrences, and this framework can help identify such a kind of information within a spatio-temporal dataset.

4.1. Handling Vagueness

Our framework provides a method of handling spatial and temporal vagueness based on *standpoint semantics* [Bennett 2011], which proposes a parametrised logic where parameter values denote different possible precisification of a vague predicate. In standpoint semantics, the syntax for defining a predicate allows additional arguments to be attached to it corresponding to semantic variation parameters. Specifically, where a vague n-ary predicate V depends on m parameters we write it in the form:

$$\mathsf{V}[p_1, ..., p_m](x_1, ..., x_n).$$

The following example illustrates the use of this syntax, where the threshold $tallness_{thresh}$ is employed to specify whether a certain height is positively relevant to classify a person x as tall.

$$\mathsf{Tall}[tallness_{thresh}](x) \equiv_{def} \mathsf{height}(x) > tallness_{thresh}$$

Geographic processes are affected by various different kinds of vagueness, regarding both their spatial and temporal aspects. Defining the temporal boundary of a process depends on many variables, such as the sort of process examined (e.g. deforestation) the agents involved (e.g. human action or wildfire originated from spontaneous combustion), the purpose (e.g. deforestation caused by human actions with purpose of wood trading). To illustrate, suppose a forest that has been observed for a year, and suppose it has been noticed that the forest has been deforested during 4 consecutive months, every day, except on Mondays. Thus the judgements on whether a single process has proceeded over the whole 4-month period or many different processes proceeded from Tuesdays to Sundays relies on the standpoint of an expert (e.g., an ecologist), and therefore both interpretations might be admissible according with the problem at hand.

The framework provides precise definitions for predicates to represent event occurrences, and well as processes which are said to proceed on specific time intervals.

Such intervals determine explicit temporal boundaries for instances of particular events and processes types. However, in order to allow different instances to be determined based on individual viewpoints, we also model the notion of process activeness. That is, a certain process can be regarded as active or inactive (while keeping its identity) within the interval on which it proceeds. Since this approach is based on standpoint semantics, different values for an *activeness threshold* parameter can be set according to different properties of the geographic phenomenon under examination. For example, in the predicate $\mathsf{Proceeds\text{-}On}[a_{th}](p, i)$, the activeness threshold a_{th} is used as a standpoint semantics parameter, so that the interpretation of the predicate depends on the values assigned to this parameter. These values can be specified by experts or obtained by using machine learning algorithms.

Moreover, to accommodate distinct viewpoints regarding the spatial boundaries of events and processes (which in fact correspond to the boundaries of their participant geographic features), the notion of spatial 'connectivity' used in RCC theory is relaxed. This means that the extension of a feature can be determined by the aggregation of regions which are in fact disconnected, where the distance between them is limited by a standpoint semantics parameter, called the *aggregation factor*. For example, a given forest can be composed by forested regions which are not necessarily connected (these regions can be separated by rivers, for example).

5. Application

We have implemented a *system prototype* (named PROGRESS) to evaluate the applicability of the proposed formalism to process real geographic data. The system takes temporal series of topographic data as an input and allows logical queries to be formulated about the data, returning information on events, processes, and the geographic features which participate in them. Experiments using this prototype have been conducted in the form of a case study, investigating the phenomenon of deforestation in Amazon between 2004 and 2011. The dataset used consists of 47,459 polygons, each of which representing a different region in Amazon which is known to be deforested at a particular time.

Queries are specified at a high level of abstraction, using the logical language employed within the theoretical framework (by adopting a Prolog-like syntax). To illustrate the way queries can be formulated and how the prototype can be interacted with, we now describe examples of logical queries and the results returned by the system.

Query 1: *Where was Amazon being deforested between 15/09/2005 and 30/04/2006?*

As deforestation is characterised here in terms of the expansion of features of type 'deforested', this query can also be described as *"show the geographic features of type 'deforested' that were expanding between 15/09/2005 and 30/04/2006?"*. Hence, for a given activeness threshold of 3 months, this query's equivalent representation in first-order logic is as follows:

$$\exists f p[\mathsf{Feature}[\mathsf{50Km}](f) \wedge (\mathsf{f\text{-}type}(f) = \mathsf{deforested}) \wedge (p = \mathsf{process}(\mathsf{extending}, f)) \\ \wedge \mathsf{Proceeds\text{-}On}[\mathsf{3months}](p, [\mathsf{15/09/2005,30/04/2006}])] \text{ ?}$$

Logical queries submitted to PROGRESS are specified using Prolog syntax, so that they can be processed by the interpretation engine. Figure 1 shows one way in which

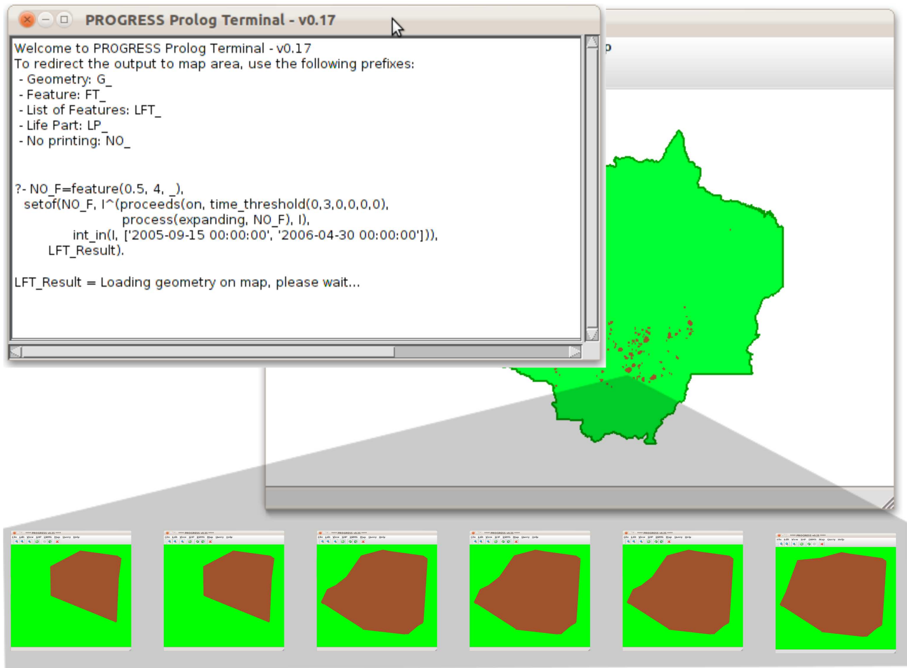

Figure 1. Example of query formulation and system interaction.

Query 1 could be written to be input to the system, as well as the results it returns. The predicate `time_threshold`(0,3,0,0,0,0) represents the activeness threshold, and its parameters denote, respectively, the number of years, months, days, hours, minutes and seconds; `int_in`(...) corresponds to the temporal relation $\mathsf{In}(i_1, i_2)$ between time intervals. Variables of a query can named using special prefixes handled by PROGRESS's Terminal to help the system control the result output. In the query of this illustration, `LFT_` is used to inform to the system that the variable is a list of features and that its value must be shown on the map (rather than on the Terminal), whilst `NO_` commands the system to hide the variable's value from the output.

Figure 1 also exhibits the results displayed on the map area. The system provides a navigation mechanism which allows the user to verify the extension of the features at different time instants. On the bottom of Figure 1, it is shown the extension of a particular feature for 6 consecutive months within the specified interval. From these illustrations, it can be seen that the feature's extension remained unchanged for a certain period of time, however this period is shorter than the 03 months activeness threshold and therefore the feature was still regarded as a process participant.

The query shown in Figure 1 uses a built-in predicate *setof* to collect solutions together by repeatedly backtracking and generating alternative values for a *result set*, corresponding to different instantiations of the free variables of the goal. As the only values of interest were those of F variable, the *Interval* variable is existentially quantified *I*. The following query illustrates a different scenario where time intervals are also of interest.

Query 2: *Where and when was Amazon deforested before 2011?*

This query could be rewritten as *"show the geographic features of type 'defor-*

ested' which expanded before 2011 and the respective time intervals on which these expansions occur". This query could be specified in our prototype as follows.

```
?- FT_F=feature(0.5,4,_), NO_E=event(expands, FT_F),
   occurs(on, NO_E, I), int_before(I,
   ['2011-01-01 00:00:00', '2011-01-01 00:00:00']).
```

In the query above, it can be seen that `FT_F` and `I` are both free variables, and therefore their values are gradually displayed on the map and on the terminal (respectively), corresponding to different solutions for the query.

6. Discussion and Conclusion

The representation of events and processes is still the subject of considerable controversy in the literature. For this reason, some previous work have avoided providing precise definitions for certain concepts. On the other hand, the formalism of our framework is specified in terms of precise logical definitions, with the aim of processing real topographic spatio-temporal data.

The REGEP framework was tested by considering geographic phenomena that can be described in terms of certain spatial changes affecting geographic features, but there are many other phenomena that do not meet this assumption. However, this framework can be easily extended to deal with other kinds of phenomena, without much modification to the rest of the semantics, including the formal apparatus for modelling of temporal aspects events and processes, to determine the relationships between them and the method of handling vagueness.

We consider the contribution described here as an important step towards the application of formal theories of space and time to solve real problems affecting geographic space. We believe that a more comprehensive theory should provide a number of additional characteristics, including the representation of more complex relationships between events and processes; and a method of modelling certain properties of processes (e.g. a process may be said to accelerate).

Acknowledgements

Thanks are due to Dr. Brandon Bennett, Dr. Vania Dimitrova, Dr. Anthony Cohn and Dr. Antony Galton for their contributions to the development of this work.

References

Allen, J. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.

Bennett, B. (2010). Spatial vagueness. In Jeansoulin, R., Papini, O., Prade, H., and Schockaert, S., editors, *Methods for Handling Imperfect Spatial Information*. Springer.

Bennett, B. (2011). Standpoint semantics: a framework for formalising the variable meaning of vague terms. In Cintula, P., Fermuller, C., Godo, L., and Hajek, P., editors, *Understanding Vagueness — Logical, Philosophical and Linguistic Perspectives*. College Publications.

Campelo, C. E. C. and Bennett, B. (2012). Geographical processes representation: Issues and challenges. In Podobnikar, T. and Ceh, M., editors, *Universal Ontology of Geographic Space: Semantic Enrichment for Spatial Data*. IGI Global, USA.

Campelo, C. E. C. and Bennett, B. (2013). Representing and reasoning about changing spatial extensions of geographic features. In *Conference on Spatial Information Theory (COSIT'13)*, Scarborough, UK.

Campelo, C. E. C., Bennett, B., and Dimitrova, V. (2012). From polygons and timestamps to dynamic geographic features: Grounding a spatio-temporal geo-ontology. *Advances in Conceptual Modeling*, 7518:251–260.

Claramunt, C. and Theriault, M. (1996). Toward semantics for modelling spatio-temporal processes within GIS. *Advances in GIs Research I*, pages 27–43.

Cohn, A. G., Bennett, B., Gooday, J., and Gotts, N. (1997). RCC: a calculus for region-based qualitative spatial reasoning. *GeoInformatica*, 1:275–316.

Devaraju, A. and Kuhn, W. (2010). A Process-Centric ontological approach for integrating Geo-Sensor data. In *FOIS 2010*.

Galton, A. (2006). On what goes on: The ontology of processes and events. In Bennett, B. and Fellbaum, C., editors, *Proceedings of FOIS Conference 2006*, volume 150. IOS Press.

Galton, A. and Mizoguchi, R. (2009). The water falls but the waterfall does not fall: New perspectives on objects, processes and events. *Applied Ontology*, 4(2):71–107.

Randell, D., Cui, Z., and Cohn, A. (1992). A spatial logic based on regions and connection. *KR*, 92:165–176.

Worboys, M. and Hornsby, K. (2004). From objects to events: GEM, the geospatial event model. *Geographic Information Science*, pages 327–343.