Deriving Summaries Through an Identity-Based Approach

Kathleen Hornsby

National Center for Geographic Information and Analysis
University of Maine
Orono, Maine 04469-5711, USA
khornsby@spatial.maine.edu

Abstract

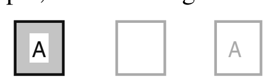
The existence of large datasets requires methods for producing simpler or summarized views of data. Typical approaches to summarization based on statistics do not capture completely the semantics associated with coarser views of spatio-temporal phenomena. In this paper, methods for deriving summaries of spatio-temporal objects are described and a set of operators is introduced for evolving summarized views. The approach uses the concept of object identity and examines the nature of deriving coarser views of objects based on changes to temporal detail as well as eliminating and combining objects.

Introduction

The existence of large datasets requires methods for producing simpler or *summarized* views of data. Summaries of large or complex scenarios improve our understanding of events or processes. Additionally, summaries are important for planning or decision-making purposes and are often used as the basis for forecasting. Methods for summarization, often based on traditional statistical approaches, have been discussed in the database literature (Johnson 1981; Sato 1981; Han and Fu 1994; Roddick *et al.* 1999) and play an important role for data mining, data warehouses, mobile databases, as well as for visualization of large datasets. Procedures developed for summarization, however, do not typically take into account the temporal aspects of phenomena. This paper describes methods for deriving summaries that capture the semantics associated with coarser views of spatio-temporal phenomena. The approach is based on modeling spatio-temporal entities as identifiable objects that experience different identity states over time. Using this model, a set of summary operators for spatio-temporal objects is introduced.

Geographic entities can be modeled as objects. A graphic *Change Description Language* (Hornsby 1999; Hornsby and Egenhofer 1997; 2000) has been developed that is based on a classification of alterations to discrete objects by tracking changes to an object's *identity*. Object identity captures the uniqueness of an object, independent of its attributes and values (Khoshafian and Copeland 1986).

The Change Description Language (CDL) uses symbols to convey the basic elements or *primitives* of a model of change as well as combinations of these primitives. The primitives are based on identity states of objects. An object can be in one of two identity states: either existing, where an identifiable object is present, conceptually (e.g., Florida) or physically (e.g., The Capital Building) (Figure 1a), or non-existing. Non-existence is further refined to distinguish between non-existing without history and non-existing with history. The first case describes the situation in which no object with identity is existing or has existed previously (Figure 1b). The term *without history* means that no previous object with that identity has existed. A non-existing object *with history* (Figure 1c) refers to the case where an object with identity previously existed but has been eliminated and no longer exists. This primitive models, for example, the Soviet Union today. Although an object's identity is unchanging, the state of an identity may change, for example, from existing to non-existing.



(a) (b) (c)

Figure 1: Basic symbols used for (a) an existing object, (b) non-existing object without history, and (c) non-existing object with history.

Temporal change, based on the order of events, is modeled qualitatively as seen for many of the typical spatio-temporal application domains (e.g., geology, archeology) (Frank 1994). The change from one identity state to another is referred to as a *transition* and is depicted with an arrow (Figure 2). Scenarios reflecting changes to object identity are developed from the left of the transition arrow to the right, where left corresponds to *before* and right to *after*. Transitions are assumed to be direct, with no intermediate states being portrayed. No quantitative measures are represented with the CDL, for example, no information on the duration of the length of time of a transition is portrayed.

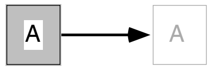


Figure 2: A transition between two identity states of an object.

Systematic combinations of the primitives result in a set of basic identity-based change operations. Beginning with a transition between an object that is non-existing without history and an existing object, the semantics of a `create` operation (Figure 3a) are captured. The operation `continue existence` reflects a transition between two states of an existing object (Figure 3b). The removal of an existing object results in a non-existing object with history (`eliminate`) (Figure 3c) or without history (`destroy`) (Figure 3d). The transition from a non-existing object with history to an existing object captures the semantics of a `reincarnate` (Figure 3e), describing the fact that the same identity has existed previously.

The remaining four are combinations of non-existing identity states: `continue non-existence with history` (Figure 3f), a `forget` operation (Figure 3g), the transition from a non-existing object without history to a non-existing object with history describes a `recall` operation (Figure 3h), and the operation `continue non-existence without history` of an object (Figure 3i).

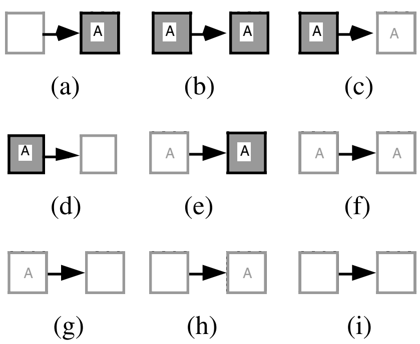


Figure 3: Operations on single objects: (a) `create`, (b) `continue existence`, (c) eliminate, (d) `destroy`, (e) `reincarnate`, (f) `continue non-existence with history`, (g) forget, (h) `recall`, and (i) `continue non-existence without history`.

Identity states can be sequenced together to form a *scenario* of change (Figure 4). A scenario reflects a conceptualization of objects over time and is composed of a set of identity states linked by transitions.

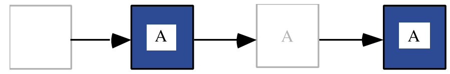


Figure 4: A scenario involving identity states for the object `A`.

Most modeling situations, however, involve more than one object. The identity-based CDL has been extended to capture transitions that link multiple objects (Hornsby 1999; Hornsby and Egenhofer 2000). Temporal co-occurrence of transitions affecting different objects is conveyed through aligning identity states vertically such that the states of more than one object can be modeled (Figure 5). In this way, scenarios can be extended to include multiple objects, such as the example shown in Figure 5, involving objects—nations, in this case—associated with the formation of the country Canada.

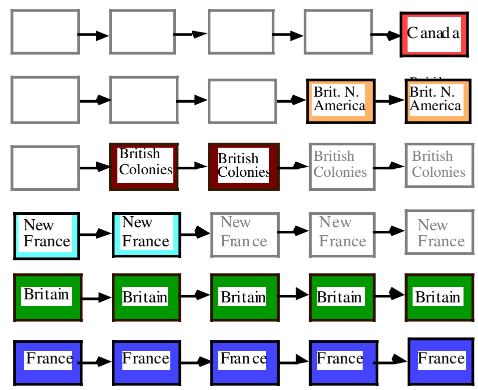


Figure 5: Co-occurring transitions affecting objects with different identities are aligned vertically. This ordering of objects and transitions models the formation of Canada.

Capturing further complexities in spatio-temporal scenarios involves modeling additional linkages that also exist among objects, such as Canada was formed from British North America (Figure 6a). In more detail, three transitions can co-occur: one between the two states of object A (e.g., British North America), another one between the two states of another object B (e.g., Canada), and a third between the first state of object A and the second state of object B—a *cross-object transition* (Figure 6b). The cross-object transition is symbolized in the CDL with a diagonal arrow (drawn either upward or downward) that links two objects of different identity and comprises many changes involving spatio-temporal objects.

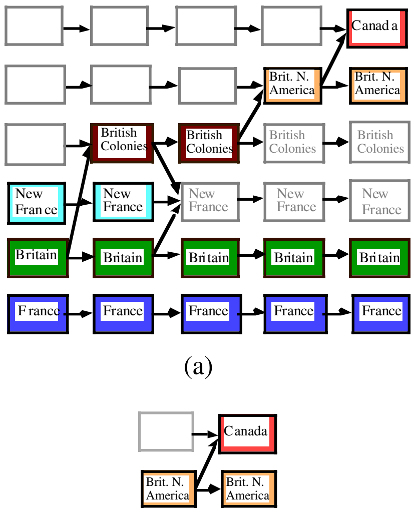


Figure 6: Spatio-temporal modeling involves (a) cross-object transitions among objects such as (b) from object `British North America` to object `Canada`.

Summarization of Spatio-Temporal Objects

A set of temporal zoom operators has been proposed for aiding spatio-temporal knowledge representation over different levels of detail (Hornsby and Egenhofer 1999). Summarization involves a particular type of shift in detail, namely, a coarsening in the level of detail. Summaries of spatio-temporal objects provide less detailed views with respect to space *and* time. Coarsening detail returns a new, less-detailed *view* of a scenario. Views here are not defined as visual representations, but rather are abstractions (i.e., less detailed views) of identifiable objects and the transitions between identity states. All scenarios have multiple views, each of them distinct from the others. Each view corresponds to a different way of perceiving the transitions among identity states. Summaries of spatio-temporal objects can be obtained through:

- reducing the detail about transitions among objects,
- collapsing transitions among identity states into fewer transitions, and
- abstracting objects from a scenario through either completely removing or combining identities.

Reducing Detail about Transitions among Objects

Simpler, summarized views of scenarios are produced by omitting the details relating to cross-linkages among entities, for example, abstracting the fact that Canada was formed from British North America to simply, Canada was created after British North America was in existence. This is effected by removing cross-object transitions from a view–a `DropCT` operation (Figure 7a). Although cross-object transitions linking identity states of different objects are removed, the transitions connecting same-identity states are left intact (Figure 7b).

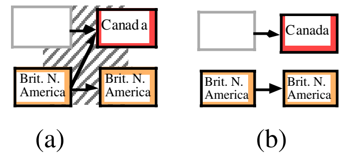


Figure 7: `DropCT`: Reducing detail relating to the transitions among objects as (a) known cross-object transitions are (b) abstracted away.

Collapsing Transitions among Identity States

Identity states of the same object can be combined through a `Collapse` operator (Figure 8a) returning a new view comprising fewer identity states (Hornsby and Egenhofer 1999) (Figure 8b). This results in a coarsening of temporal detail over transitions between objects.

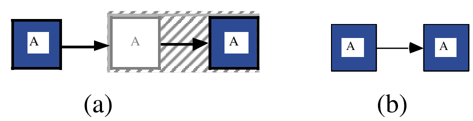


Figure 8: `Collapse` operation: (a) Two states of object `A` that are linked through a transition are collapsed into (b) a single identity state.

The `Collapse` operation uses certain rules for reasoning about combinations of identity states. Depending upon the task, a priority can be assigned to identity states determining the resulting identity state of an object after a `Collapse` (Figure 9). The priorities can be: existence ($\bullet_E$), followed by non-existing with history; non-existence ($\bullet_N$); the before-transition state ($\bullet_1$); or the after-transition state (State2 in Figure 9 ($\bullet_2$)).

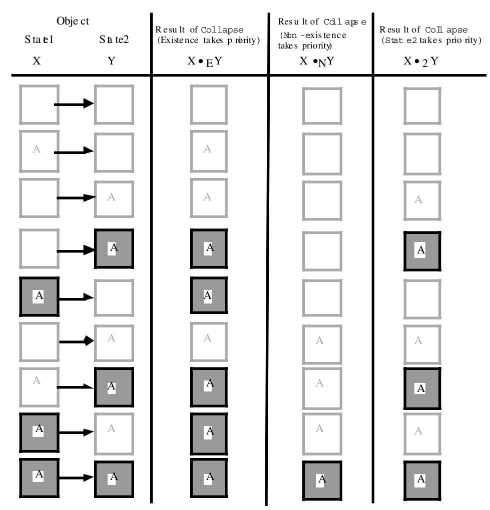


Figure 9: `Collapse` uses rules for reasoning about combinations of identity states.

Depending on the priority scheme selected for the `Collapse` operation, different results will be obtained (Figure 10).

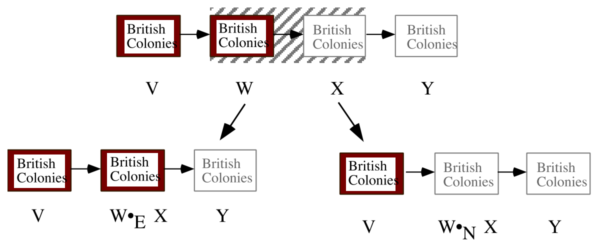


Figure 10: Result of `Collapse` depends on which priority scheme is selected.

The `Collapse` operator can be applied iteratively to obtain views at increasingly coarser levels of temporal detail. This operator must also return sequences of transitions that are valid according to a given domain ontology and therefore it tests for any invalid sequences of identities as known from combinatorial rules based on a domain ontology. From this set, for example, sequences of identities where an intermediate state that is non-existing without history may not be plausible.

For deriving summaries based on scenarios involving multiple objects, the `Collapse` operation is performed over sets of objects (Figure 11a). `Collapse` can be applied iteratively until the coarsest level of detail is reached, i.e., where only identity states at one time are modeled (Figure 11b). In this case, the summary may contain information regarding objects that are non-existing, such as former colonies, as well as existing objects.

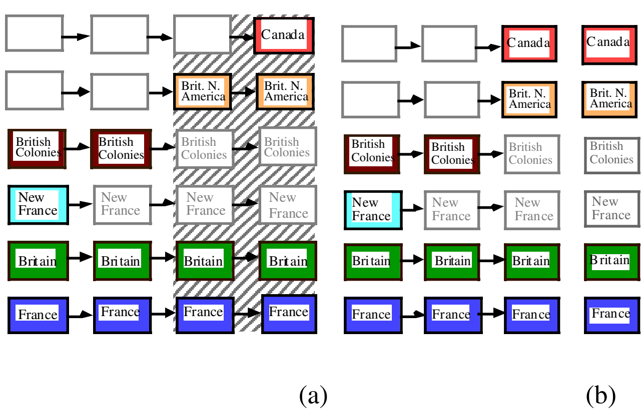


Figure 11: Summaries can be obtained through (a) a `Collapse` operation being applied to sets of objects, possibly abstracting (b) all transitions and identity states until only identities at one time remain.

Abstracting Objects from Scenarios

Summarization can also require that entire objects are abstracted from a scenario based on semantics that involve either eliminating or combining less relevant objects to remove them from a scenario.

Removing Objects

Summarized views are obtained through a `DropObject` operation where all identity states associated with one or more objects are abstracted (Figure 12a). The removal of objects is based on domain knowledge and relies on criteria such as the number of transitions involving non-existing identity states or the number of cross-object transitions. An object that goes out of existence earlier in a scenario, for instance, may be less likely to be relevant to a summary than other objects. Similarly, objects that are not linked by any cross-object transitions may be abstracted more readily than objects that are linked to others through cross-object transitions. `DropObject` must include support for reasoning about cross-object transitions after abstraction. Certain adjustments to these transitions may be necessary, as the results of eliminating objects may not be plausible given certain contexts (Figure 12b). Alterations to the transitions, such as initiating the transition from the most recent existing identity state, may be undertaken(Figure 12c).

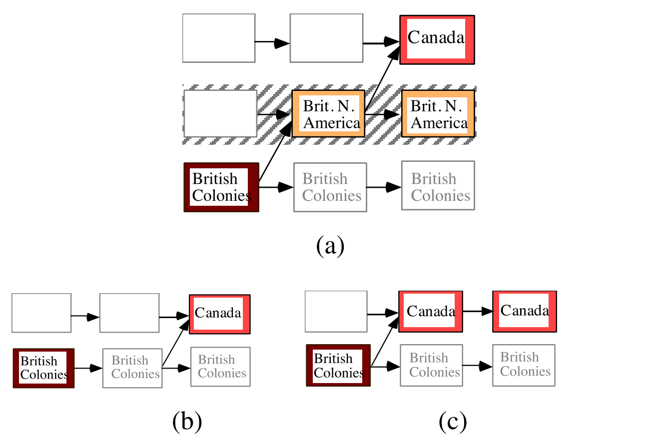
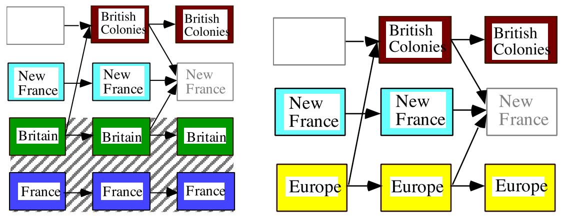


Figure 12: `DropObject` (a) abstracts objects from a scenario resulting in (b) fewer identities and transitions and (c) adjusts the cross-object transition to begin with the most recent existing object.

Combining Objects

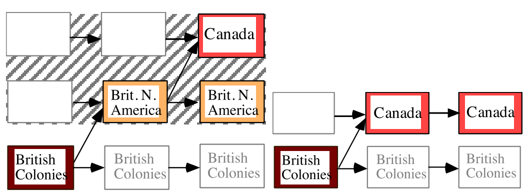
Summaries can be derived by combining or merging two or more objects (Hornsby and Egenhofer 1997) resulting in a view involving fewer objects (Figure 13). This can happen in at least two ways. In the first case, when objects are abstracted from a scenario, a smaller number of new identities take their place. `Britain` and `France`, for example, can be combined into new object, `Europe`. The new identity inherits any cross-object transitions that are linked to the identity states of predecessor objects. Similar to the `Collapse` operation, `Merge` requires a set of priority rules for combining the identity states of different objects.



(a) (b)

Figure 13: `Merge` operation: (a) combines identities into (b) a new object. Original identities are eliminated.

The second type of combine operation, `Subsume`, refers to those cases where an object subsumes another object's identity. `British North America` (BNA), for example, could be combined with `Canada` such that `Canada` continues to exist but BNA is eliminated from the scenario (Figure 14). At the level of identities, the subsumed object's identity disappears–as with a `Merge`–leaving the other object's identity to continue as an existing object. If object *properties* were also being modeled, however, the successor object would inherit the properties of the subsumed object.



(a) (b)

Figure 14: `Subsume` operation where two objects are combined such that (a) one object is subsumed into the other (existing) object and (b) like the `Merge`, a successor object continues to exist.

Conclusions

Summarized views of objects are important for providing more meaningful interpretations of data especially in the context of large databases. This paper has examined procedures for evolving summaries from an identity-based perspective. Summaries are coarser views of phenomena and, depending on the task at hand, can be derived from: reducing the detail about transitions among objects, collapsing transitions among identity states into fewer transitions, and abstracting objects from a scenario through either completely removing or combining identities. Summaries can contain details on non-existing objects as well as existing objects. A set of operators that supports the derivation of summarized views includes `Collapse`, `DropCT`, `DropObject`, `Merge`, and `Subsume`. Each of these operators can also be applied in combination with each other such that cross-object transitions *and* entire objects are abstracted from a scenario in addition to collapsing the transitions among objects. Further work is necessary, for example, on the effects of combining the operators and on support for reverting from the summarized view to the more detailed, original scenario where information about objects abstracted from the scenario will need to be retained.

Acknowledgments

This work was supported in part by grants from the National Science Foundation SBR-9700465, the National Imagery and Mapping Agency NMA202-97-1-1023, and the National Institute of Environmental Health Sciences, 1 R 01 ES09816-01.

References

A. Frank (1994) Qualitative temporal reasoning in GIS-ordered time scales. in: T. Waugh and R. Healey (Ed.), *Proceedings of the Sixth International Symposium on Spatial Data Handling*, Edinburgh, Scotland, pp. 410-431.

J. Han and Y. Fu (1994) Dynamic generation and refinement of concept hierarchies for knowledge discovery in databases. in: *AAAI'94 Workshop on Knowledge Discovery in Databases*, Seattle, WA, pp. 157-168.

K. Hornsby (1999) Identity-based reasoning about spatio-temporal change. unpublished Ph.D. Thesis, University of Maine, Orono, ME, USA.

K. Hornsby and M. Egenhofer (1997) Qualitative Representation of Change. in: S. Hirtle and A. Frank (Ed.), Spatial Information Theory—A Theoretical Basis for GIS, *International Conference COSIT '97*, Laurel Highlands, PA. *Lecture Notes in Computer Science* 1329, Springer-Verlag, Berlin, pp. 15-33.

K. Hornsby and M. Egenhofer (1999) Shifts in detail through temporal zooming. in: A. Tjoa, A. Cammelli, and R. Wagner (Ed.), *Proceedings of the Tenth International Workshop on Database and Expert Systems Applications*, *DEXA'99,* Florence, Italy, pp. 487-491.

K. Hornsby and M. Egenhofer (2000) Identity-based change: a foundation for spatio-temporal knowledge representation. *International Journal of Geographical Information Science*, Forthcoming.

R. Johnson (1981) Modelling summary data. *SIGMOD Record* : 93-97.

S. Khoshafian and G. Copeland (1986) Object identity. *SIGPLAN Notices* 21: 406-416.

J. Roddick, M. Mohania, and S. Madria (1999) Methods and interpretation of database summarisation. in: T. Bench-Capon, G. Soda, and A. Tjoa (Ed.), *Database and Expert Systems Applications, DEXA'99,* Florence, Italy, pp. 604-615.

H. Sato (1981) Handling summary information in a database: derivability. *SIGMOD Record* : 98-107.