CLASSIFICATION AND REPRESENTATION OF CHANGE IN SPATIAL DATABASE FOR INCREMENTAL DATA TRANSFER

Huaji ZHU[1,2], Jun CHEN[2], Jie JIANG[2]

1 The Institute of Remote Sensing Application of CAS, Datun Road, Beijing, China, 100101

zhuhuaji@sohu.com

2 National Geomatics Center of China, 1 Baishengcun, Zizhuyuan, Beijing, China, 100044

Key words: incremental data transfer, change, Geographical Object, update, Classification, representation

ABSTRACT: Nowadays, more and more spatial databases are used in various fields, so the demands for keeping spatial database "fresh" are growing rapidly. Generally, the end-users get the basic spatial data from professional spatial data producers. There are mainly two kinds of way for transfer the changed data. One is to batch transfer; the other is incremental data transfer. With the batch method, the whole up-to-date database is delivered. This process is time-consuming and might induce significant risks of errors occurrence and information loss.. Recent years, more and more people begin to study incremental updating of spatial database. This method of updating makes it possible to transfer change-only information to the end-users, namely to transfer the incremental data. In order to find out the changed objects in a database, firstly we need to identify uniquely all of the objects. In this paper, we define a geographical object as a new 4-tuple {semantic descriptor, thematic descriptor, spatial descriptor, and temporal descriptor}. Spatial descriptor consists of geometric component, position component and topologic component. What change happens to an object? In order to answer this question, firstly, we have to identity and classify changes of geographical objects. Based on the 4-tuple model of geographical object, we propose taxonomy of the change of geographical object and describe these changes with data/knowledge packets.

1. INTRODUCTION

Geographic Information Systems (GIS) are increasingly used in many domains, such as traffic and water sections. Since, implementing such systems are complex, they are always based on basic spatial data which are provided by special institutions. Users, generally, gain reference geographic data from producers in order to establish their GIS. For instance, a realty company purchases from a producer a geographic database representing the land usage of a given region for its company planning application. Users regard the data provided by producers as spatial reference data. Geographic data producers are responsible for producing and maintaining up-to-date databases and delivers new information to users. In this way, there are many advantages. But many new problems appear. How to update and maintain user database, especially maintain the spatial reference data, is one of the main challenges. When the databases created by the data producer are updated, the changed data must be transferred to the end-user to keep their client-databases current. This includes the management of geographical changes. A more explicit focus on change has considered the semantics associated with change, (Claramunt and Theriault 1995, 1996). These studies, however, have omitted certain kinds of change and so far no systematic treatment of change has been undertaken and they have no given formal expression. The objective of our research is the identification of a model for the classification and description of changes in topological database.

2. TRANSFER OF UPDATES

There are several solutions available for the transfer of updated dataset (Spéry,2001).

- Bulk data transfer
- Incremental data transfer: Data producers only provide those objects that have been changed since the last transfer. In this case, the producer has to identify geographical changes in order to generate relevant data sets to each user context (Wijngaarden et. Al., 1997).

Typically, the whole new versions of a data set are currently disseminated to end users even if only 10% of the objects stored in the database have changed (this is a common estimation of the rate of evolution per year for geographic databases (Raynal, 1996). By bulk transfer, the users either ignore the update (if it is not significant enough) or manually (and selectively) update their database and rebuild their value-added data. In some cases, users can't accomplish their tasks. So, we have to adopt new transfer mode. Incremental Data Transfer is the preferable solution that is a data transfer system that enables suppliers to supply only information that has been affected by change between two versions of a dataset and for users to be able to incorporate those changes into their data. Many researchers proposed various spatio-temporal data models describing the geographical changes(Peuquet, 1994; Frank, 1994; Langran, 1992; Worboys, 1994; Jiangjie, 2000). But up to now, no single spatio-temporal model in a GIS has been adopted. Therefore, a description and classification of geographical changes is required. The analysis of geographical changes has been one of the important directions of the GIS research agenda over the past ten years (Peuquet, 1994). Several classifications of geographical changes have been proposed, changes at different levels of abstraction (Hornsby and Egenhofer, 1997) and the description of spatio-temporal processes that generate these changes (Claramunt and Theriault 1995, 1996; Hornsby and Egenhofer, 2000). And these researches describe a geographical change with the comparison of two pictures that denote the two states of the objects. This way implicitly contains people's cognitive process, that is to say, this description has limited information about geographical changes.

The objective of this research is the identification of a model for the classification and description of geographical changes. We explicitly describe people's cognitive process about real world changes and express these changes as a set of rules. A geographical entity is represented as a four-element group {semantic component, thematic component, temporal component, and spatial component}. Based on this model, we propose taxonomy of geographical changes.

3. REPRESENTATION OF GEOGRAPHICAL OBJECT

Geographical objects are always presented as a 3-tuple {thematic attribute, geometric attribute, ID}. This model defines semantics by thematic attributes. Semantics in this paper is defined as the relationship among the computer representations and the corresponding real world feature within a certain context (Bishr , 1998). The major problem of defining semantics by attributes, is to identify which attributes to choose. In general, it is extremely difficult and often impossible to find a complete set of characteristic attributes for a Real World object (Kuhn, 1995). In order to transfer more information about geographical object and its change to users, we prefer a model with more semantics to represent the geographical object. We replace the ID with semantics description and add temporal description in the classical representation of geographical object , i.e. a 4-tuple{ semantic descriptor, thematic descriptor, spatial descriptor, and temporal descriptor }. Spatial descriptor consists of geometric component, position component and topologic component .

4. CLASSIFICATION AND REPRESENTATION OF SPATIO-TEMPORAL CHANGES

4.1 The definition of spatio-temporal change

What's the change? Let's see some examples of changes, spread of forest fire, building collapse by earthquake, alternation of land access, and so on. Object is the database representation of real feature. By oriented-object method, we can represent the real feature as the object. Object is the basic semantic unit for us to understand real world. The definition spatio-temporal change is a process of the transformation from one state to another state of the geographical features. Spatio-temporal change influences one or more geographical features and their attributes. So, there are several factors related to change, object set, object, object component and component attribute. We give a 4-tuple to represent the composition of a change {ObjectSetb, ObjectSeta, name of a change, rule set }, ObjectSetb represents object set before a change, ObjectSeta represents object set after a change, rule set contains many rules by which we use to identify the change.

4.2 Capturing semantics

Capturing semantics of change is the identifying process of a change. The identification of a real world change is a heuristic process. Heuristics are common sense knowledge, or rules of thumb, that originate from the expert's past experience (Avelino et al., 1993). There are three categories of heuristic knowledge: associational, motor skills, and theoretical. Associational knowledge is mostly acquired through observation. It is always represented as rules(for instance IF—THEN). Hence, representing such knowledge as a set of rules is a viable solution. Data, geometric and thematic, and knowledge can be encapsulated into an abstract object type with sufficient discipline knowledge to be able to interpret the appropriate meaning of an object(Bishr,1998). This is known as data/knowledge packets that allow both knowledge and data to be represented in a unified model (Doyle and Kerschberg, 1991). A data/knowledge packet could look like this

{
ChangeClass C
ObjectSetb
ObjectSeta
Rules□IF—THEN expression□
}

A object set can contains n objects (n>=0), firstly we classify change by comparison the cardinality of ObjectSetb and ObjectSeta there are 6 types of change: 0 to 1 (there is zero object in ObjectSetb, there is only one object in ObjectSeta); 1 to 0; 1 to 1; 1 to n; n to 1; n to m.

The first three types of changes influence single geographical feature, the other influence several geographical features. We classify the spatio-temporal changes based on the following aspects:

1)Levels of change
Object
Object component
Component attribute

2) factors related to change
Object set and its attributes
Object
The relationship of objects in set
Object components and their attributes

3) Only focus on single change

4) change impact on simple object.

Change has three levels, i.e. change about object, change about object component, change about component attribute. For example, disappearance is the object change, distortion is change about object component and component attribute. We take into account four factors that are used to identify changes, they have hierarchy relationship, object set contains object, object is composed of components, and component is described by attributes. By these hierarchy factors, we can shorten the identifying process of changes. For example , if we identify the disappearance of an object, it is dispensable to judge whether the shape of the object changed. Here, we only focus on the changes affecting simple objects not include changes affecting complex objects. We only define the single change and not consider the multiple changes. For example, an object changes its shape, its area may change. We can determine the object has two types of change, distortion and area change. We don't make up a complex change by the two simple changes.

4.3 Classification and Representation of changes impact on a single object

We regard object set, attributes of object set, object, object component and attributes of components as the factors to identify spatio-temporal changes. By comparing the value before change and that of after change of the same factor, we give the taxonomy of the spatio-temporal changes impact on a single geographical object. There are many expressions used in the description of changes(as show in table 1). Here, we detailedly explain the function of semantic component of a geographical object. Given a scenario of the change of a geographical object, if the value of the semantic component of the object has not changed, we can announce the object is the same object.

Table 1 expression and their meanings

Expression	**meanings**	**Expression**	**meanings**
ObjectSet	Object Set	Position(OO)	Position
ONum(Object Set)	Cardinality of set	Direction(OO)	Direction
O	Geographical object	GType(OG)	Geometry type
OS	Semantic component	Length(OG)	Length of line object
OO	Orientation component	Area(OG)	Area of area object
OG	Geometric component	NumItems(OA)	Number of attributes
OA	Thematic component	ItemVal(OA_i)	Value of attribute A_i
OR	Topology component	X_b	The state before change
OT	Temporal component	X_a	The state after change

Now , we use the data/knowledge packets to describe changes. Table 2 gives the taxonomy of spatio-temporal changes related to a single geographical object. The taxonomy has hierarchy. These changes are basic. In the real world, a geographical feature may have several changes at the same time. By combining these basic changes, we can model scenarios of the real world changes.

Table 2 the taxonomy of spatio-temporal changes impact on a geographical object

Name of change	**data**	**Rules**
Appearance	$ObjectSet_b=\varphi$ $ObjectSet_a=\{O_a\}$	IF $ObjectSet_b=\varphi$ AND $ObjectSet_a=\{O_a\}$
Disappearance	$ObjectSet_b=\{O_b\}$ $ObjectSet_a=\varphi$	IF $ObjectSet_b=\{O_b\}$ AND $ObjectSet_a=\varphi$
Stability	$ObjectSet_b=\{O_b\}$ $ObjectSet_a=\{O_a\}$	IF $ObjectSet_b$ □$ObjectSet_a$
Mutation	$ObjectSet_b=\{O_b\}$ $ObjectSet_a=\{O_a\}$	IF $O_b \neq O_a$ AND $OS_b \neq OS_a$

Change of thematic attribute	$ObjectSet_b=\{O_b\}$
	$ObjectSet_a=\{O_a\}$
		IF $O_b \neq O_a$ AND $OS_b=OS_a$ AND $OA_b \neq OA_a$ THEN
		IF NumItems(OA_b)> NumItems(OA_a) THEN Number of attributes decreases
		IF NumItems(OA_b)< NumItems(OA_a) THEN Number of attributes increases
		IF NumItems(OA_b)= NumItems(OA_a) THEN
		IF ItemVal(OA_{ib})> ItemVal(OA_{ia}) THEN Value of an attribute minishes
		IF ItemVal(OA_{ib})> ItemVal(OA_{ia}) THEN Value of an attribute goes up
		IF ItemVal(OA_{ib})≠ItemVal(OA_{ia}) THEN Value of an attribute alters.

Change of orientation	$ObjectSet_b=\{O_b\}$ $ObjectSet_a=\{O_a\}$	IF $O_b \neq O_a$ AND $OS_b=OS_a$ AND $OO_b \neq OO_a$ THEN IF Position(OO_b) ≠ Position(OO_a) THEN object moves IF Direction(OO_b)≠Direction (OO_a)THEN object rotates

Change of geometry	$ObjectSet_b=\{O_b\}$
	$ObjectSet_a=\{O_a\}$
		IF $O_b \neq O_a$ AND $OS_b=OS_a$ THEN
		IF GType(OG)=POLYGON AND $OG_b \neq OG_a$ THEN
		IF Area(OG_b)= Area(OG_a) THEN Area object deforms
		IF Area(OG_b)> Area(OG_a) THEN Area object contracts
		IF Area(OG_b)< Area(OG_a) THEN Area object expanses
		IF GType(OG)=LINE AND $OG_b \neq OG_a$ THEN
		IF Length(OG_b)=Length(OG_a) THEN Line object deforms
		IF Length(OG_b)>Length(OG_a) THEN Line object shortens
		IF Length(OG_b)<Length(OG_a) THEN Line object extends

4.4 Classification and Representation of changes impact on many objects

Many complex scenarios of change involve several objects. They have impact on the result of the changes. In this case, there are three catalogs of change: 1:n, n:1 and n:m. Here, we only take into account the first two catalogs. Table 3 gives the taxonomy of spatio-temporal changes impact on many geographical objects.

Table 3 the taxonomy of spatio-temporal changes impact on many geographical objects

Name of change	**data**	**Rules**
spawn	$ObjectSet_b=\{O_b\}$ $ObjectSet_a=\{O_{1a}$□ O_{2a} □…$O_{na}\}$	IF $O_b \neq O_{ia}$ { for $\forall O_{ia}$ ($1 \le i \le n$) }THEN IF there is only one O_{ia} ($1 \le i \le n$) OS_b□OS_{ia} THEN O_b spawns new objects.
fission	$ObjectSet_b=\{O_b\}$ $ObjectSet_a=\{O_{1a}$□ O_{2a} □…$O_{na}\}$	IF $O_b \neq O_{ia}$ { for $\forall O_{ia}$ ($1 \le i \le n$)}THEN IF $OS_b \neq OS_{ia}$ { for $\forall O_{ia}$ ($1 \le i \le n$)}THEN O_b fissions into several objects.

merge	$ObjectSet_b = \{O_{1b}$□ O_{2b} □… O_{nb} \} $ObjectSet_a = \{O_a\}$	IF $O_{ib} \neq O_a$ { for $\forall O_{ib}$ ($1 \le i \le n$) }THEN IF $OS_{ib} \neq OS_a$ { for $\forall O_{ib}$ ($1 \le i \le n$) }THEN $O_{1b}, O_{2b}, \ldots O_{nb}$ merge into O_a
identify	$ObjectSet_b = \{O_{1b}$□ O_{2b} □… O_{nb} \} $ObjectSet_a = \{O_a\}$	IF $O_{ib} \neq O_a$ { for $\forall O_{ib}$ ($1 \le i \le n$) }THEN IF there is only one O_{ib} ($1 \le i \le n$) $OS_{ib} = OS_a$ THEN $O_{1b}, O_{2b}, \ldots O_{nb}$ identify to O_a

5. AN EXAMPLE

In section 4 we give the taxonomy of spatio-temporal changes impact on geographical objects. The data/knowledge packets explicitly describe the people's cognitive process of the change. Here is a scenario of real world change(Figure 5). From time T1 to T4 the parcels have a sequences of changes. By the method proposed in this paper, we describe the changes impact on land parcels by a formalized way.

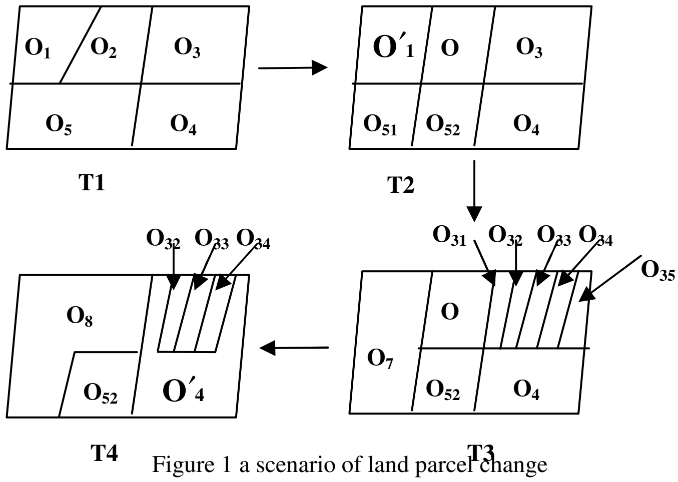

Figure 1 a scenario of land parcel change

Table 4 T1to T2□ O_1 deformed□ O_2 Mutated□ O_5 fission into O_{51}, O_{52}

Type of change	objects	rules
deform	$ObjectSet_b = \{O_1\}$ $ObjectSet_a = \{O'_1\}$	$O_1 \neq O'_1$ AND $OS_1 = OS'_1$ AND $OG_1 \neq OG'_1$ AND AREA(OG_1)=AREA(OG'_1)
Mutate	$ObjectSet_b = \{O_2\}$ $ObjectSet_a = \{O_6\}$	$O_2 \neq O_6$ AND $OS_2 \neq OS_6$
fission	$ObjectSet_b = \{O_5\}$ $ObjectSet_a = \{O_{51}, O_{52}\}$	$OS_5 \neq OS_{51}$ AND $OS_5 \neq OS_{52}$

The example indicates that the method proposed in this paper can save steps judging a change and give clear expressions. For instance, the semantic component of O2 changed, according to this, we can affirm O2 mutated. We are unnecessary to compare the other components of the object.

In order to collect this change, we need to execute some database operations. For this scenario, some operations need execute:

```
UPDATE tableN SET spatailField=… Where ObjectID = O1
DELETE FROM tableN WHERE ObjectID = O2
INSERT INTO tableN VALUES(O6 □……)
DELETE FROM tableN WHERE ObjectID = O5
INSERT INTO tableN VALUES(O 51 □……)
INSERT INTO tableN VALUES(O 52 □……)
```

6. CONCLUSION AND FUTURE WORKS

If we want to publish and transfer incremental data about the master topology database (producer's database) to users, firstly we should get incremental data by comparing the two versions of the master database. This process is executed based on the types of spatio-temporal changes. Therefore, Classification and representation of spatio-temporal changes is the precondition. Based on the object-oriented idea, we give taxonomy of spatio-temporal changes. How to represent the geographical object is the premise for analyzing the classification of changes. In this research, we regard the geographical object as a 4-tuple {semantic descriptor, temporal descriptor, themantic descriptor, and spatial descriptor}. Spatial descriptor contains three parts, i.e. Shape and Size, Position and orientation and Topology relation.

Based on the change impact on the component, we propose the taxonomy of change impact on the geographical object.
The definition and classification of change is a people's cognitive process. People's cognitive knowledge is important to this process. So we propose a model to represent the change based on the data/knowledge packets. A packet contains three basic elements, change name, objects affected by the change, cognitive knowledge. Cognitive knowledge is formalized by a series of rules. A change influences the states of one or many objects. Before change, there is a set of object related to the change. After change, there is another set of object resulted from the change. Object set contains many objects, object is composed of components, and component is described by many attributes. Based on this idea, we explicitly describe all kinds of spatio-temporal changes. The example indicates this way is more self-contained and suitable for automatic disposal in computer. The expression for every change clearly describes the people's cognitive process.
There are several areas for further research in this context. We only focus on single change and change impact on the simple object. The change of complex object and complex change is not considered in the research. Here, we only give 4 kinds of change impact on many objects. We need to analyze other kinds of change impact on many objects.

ACKNOWLEDGMENTS

This work was supported by the National 863 Project of China, under grant No.002AA131040.

REFERENCES

Avelino, J. G. and Douglas, D. D., 1993. The Engineering of Knowledge-Based Systems, Theory and Practice (Englewood Cli. s, NJ, Prentice Hall).

Claramunt, C. and M. Thériault ,1995. Managing Time in GIS An Event-Oriented Approach. Recent Advances on Temporal Databases. J. Clifford and A. Tuzhilin. Zurich, Switzerland, Springer-Verlag: 23-42.

Claramunt, C. and M. Thériault, 1996. Toward Semantics For Modeling Spatio-Temporal Processes Within GIS. Advances in GIS II(SDH96). M. J. Kraak and M. Molenaar. Delft, the Netherlands, Taylor and Francis: 47-64.

Doyle, W., and Kerschberg, L., 1991. Data/knowledge packets as a means of supporting semantic heterogeneity in multidatabase systems. SIGMOD Record, 20, 69- 73.

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

Hornsby, K. and M. J. Egenhofer 1997. Qualitative Representation of Change. Spatial Information Theory: A Theoretical Basis for GIS, Proceedings of the International Conference COSIT'97, Lecture Notes in Computer Science 1329, Laurel Highlands: Pennsylvania, USA, Springer-Verlag, pp15-33

Hornsby, K. and M. J. Egenhofer, 2000. Identity-based change: a foundation for spatio-temporal knowledge representation. International Journal of Geographical Information Science 14(3): pp207- 224

Jiang Jie 2000. Research on Event Based Spatio-Temporal Database. Ph.D. Dissertation of The China University of Mining and Technology Beijing

Langran, G. 1992. Time in Geographic Information Systems, Taylor&Francis

Peerbocus, A., G. Jomier, et al., 2002. A Methodology for Updating Geographic Databases using Map Versions. 10th International Symposium on Spatial Data Handling(SDH2002), Ottawa, Canada

Peuquet,D., 1994. It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. Annals of the Association of American Geographers, 84, pp441–461.

Raynal, L. 1996. Some elements for modelling updates in topographic databases. In the proceedings of GIS/LIS'96,Annual Exposition and Conference. Denver, Colorado, USA, pp1223-1232.

Spéry, L., C. Claramunt, et al.,2001, A Spatio-Temporal Model for the Manipulation of Lineage Metadata. Geoinformatica 5(1): pp51-70

Wijngaarden F. van, J. van Putten, P. van Oosterom, and U.H. 1997 . Map integration–Update propagation in a multi-source environment. in Proceedings of the 5th International Workshop on Advances in Geographic Information Systems (GIS'97), Las Vegas, Nevada, pp. 71-76, 1997.

Worboys, M., 1994. A unified model of spatial and temporal information. Computer Journal, 37, 26–34.