Reasons of the Incremental Information in the Updating Spatial Database

Huaji Zhu, Huarui Wu, Xiang Sun

Reasons of the Incremental Information in the Updating Spatial Database

Huaji Zhu Huarui Wu* Xiang Sun

National Engineering Research Center for Information Technology in Agriculture, Beijing, 100097, PR China

ARTICLE INFO	ABSTRACT
	In order to quickly extract and distribute the incremental information of spatial database, this work analyses the reasons of incremental information of spatial database with the instance of residential feature. Firstly, a new concept of increment information is proposed based on the spatio-temporal change type and data delta. The change type describes the semantics of increment information and data delta describes the content of increment information. Based on this concept, the types of residential change and relevant data deltas are put forward. There are two catalogs of change type, i.e. subjective change and objective change. The subjective change can be detailedly divided into 3 small change types and the objective change can be detailedly divided into 12 small change types. Additionally, the incremental information resulted from the objective change is formalized. Based on the above formalized changes, the automatic extraction of incremental information can be expediently programmed. The classifying result was applied to the incremental information automatic extracting system of 1.250,000 topographic databases.

1. Introduction

How to distribute the new data in the updated topological database is becoming a hot topic in GIS domain [1, 3]. Currently, the new data of topological database is distributed by the batch method which has such problems as the data volume distributed is very big and users are difficult to integrate the distributed data in their database [2, 4, 6]. According to the Stat. by Raynal [3], the change rate every year of the geographical objects in a geographical database is about 10%. For instance, comparing to the 1998 version of the 1.250000 topographical database (a kind of special geographical database), the change rates of all kinds of topographical feature in 2002 version of the same database are respective 9.7% for water feature (2.4% every year), 27.9% for district boundary feature (7% every year), 20.9% for road feature (5.2% every year), 34.4% for point residential feature (8.6% every year) and 75.3% for area residential feature (18.8% every year). According to above percentages, we can find that most features' change rates every year are less than 10% except for the area residential feature. If the batch method is used, about 90% distributed data are invalid. In practical, users only focus on the changed data in the updated database, i.e. the data delta between old version and new version of the same database. So providing data delta not the whole new updated database can overcome above mentioned problems in the batch method [8, 9, 10].

Data delta result from the change in real world. Different change results in different data delta, that is to say, every change type is corresponding to certain data delta. If only data delta is provided, the users sometimes don't know how to integrate the data delta in their database without knowing the reasons (change type) of the data

*Corresponding author. Tel.: +86 10 51503594; fax: +86 10 51503594.
E-mail address: zhuhuaji@126.com (Huarui Wu).

delta. For instance, data delta $\triangle$ = -[OID, O.A.V, O.S.G] denotes a deleted residential feature O, OID denotes the ID of O, O.A.V denotes the attribute set of O, O.S.G denotes the graphics of O. Several reasons can result in the data delta $\triangle$. For instance, one reason is the disappearance of a residential feature in the real world. The second reason is that the area of a residential feature is less than the area threshold of certain scale because of shrinking. In the certain scale database, a feature can be collected in the database if its area is more than or equal to the area threshold, otherwise the feature should be abandoned[11, 12]. However, the problem will completely vanish if data delta and its change type are provided together. In this work, we call the combination of data delta and change type as increment.

2. Relationship between change and increment

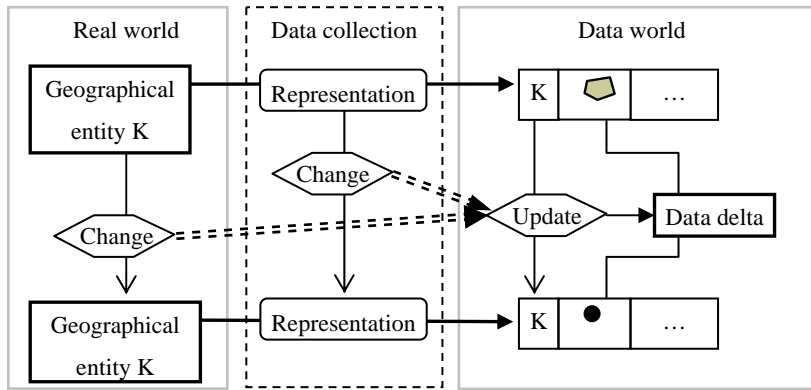The relationships among change, data delta and increment are showed in Fig.1.

Fig.1. The relationship among change, data delta and increment

Firstly, we explain the concept "geographical feature". The concept commonly is called as geographical entity in the real world and as geographical object in the data world. The changes of geographical entities in the real world or the changes of data collection standards may result into update operations which modify the content of the database and produce the data delta. The data delta resulted from every update operation is exclusive and certain. However, different change types correspond to different update operations. So the different change produces different data delta. In this paper, we define 4 kinds of update operations and relative data delta. The 4 update operations are object deletion, object creation, graphics update and attribute update. The data delta of the object deletion is $\triangle=\{$-[OID, O.A.V, O.S.G] $\}$. The data delta of the object creation is $\triangle =\{$+[OID, O.A.V, O.S.G] $\}$. The data delta of the graphics update is $\triangle=\{$O.S.-G/+G' $\}$, G is the value before change, G' is the value of after change. The data delta of the attribute update is $\triangle=\{$O.A.-V/+V' $\}$, V is the value before change, V' is the value of after change. By establishing the certain relationship between the change type and the set of update operations, the data delta of every change can be exclusively decided. So the increment which is composed of change type and data delta can be decided.

3. Reasons of increment

There are two reasons to generate the increment information, subjective reason and objective reason. The objective change is the changes of geographical features in the real world. The subjective changes result from the change of data model, the defined attribute set of every feature and so on.

3.1 Subjective changes

There are three kinds of subjective changes, the changes coming from the errors amendment, the changes coming from the change of correction rules and the change of database scheme. The subjective changes of spatial data are ruleless and have no uniform expression.

The errors amendment may generate the data delta. For example, the grade of a road is amended from the national road to provincial road. The data delta is 200542.Grade.-"national road"/+"provincial road". The 200542 is the ID number of road. The increment information is [error amendment,update, 200542.Grade.-"national road"/+"provincial road"].

The change of correction rules may generate data delta. For example, the district feature has the attribute of district code. The rule of the district code is different between the old database and new database. For example, the old district code of the "Zhuo zhou city" is 121024, but the new district code is 132402. In fact, the "Zhuo zhou city" has no change. But because the change of the code rules, the feature generates some increment information. Another example, in the old data of the 1.250000 topological database, the database field 'FName' denotes the name of the feature. The values of 'Fname' don't include the district grade. But in the new data of the 1.250000 topological database, the values of 'Fname' include the district grade. For instance, a village is called as "Chang Cao" in old data but is called as "Chang Cao Cun" in the new data.

The change of database scheme also may generate increment information. The topological database is established based on a data scheme. The database can be established based many kinds of data scheme. In the scheme, the attribute set of every feature can be changes according to the real world. So, the scheme of a topological database often changed. For example, the residential feature has a field "PYNAME" in the old data of the 1.250000 topological database. But in the new data, the field "PYNAME" is deleted.

3.2 Objective changes

There are many kinds of objective change in the real world. [4, 5] defined the event types expressing the changes. In the real world, different geographical feature with the same event may generate different changes. So, we give a new change expression based on the event type and the data delta. In this work, we only consider the simple events, that is to say, the event which affects an attribute of the geographical feature. But in the real world, there are many complex events. We can compound the changes generated from complex events based on the changes of the simple events. According to the alterant extent, the event affected the change is divided into the evolution event and the death/birth event. The death/birth event can cause the appearance of new feature or the disappearance of old feature. The evolution event may cause the change of the characteristics of feature. The evolution event can be future divided into spatial evolution event and theme evolution event.

[4, 5] defined eight kinds of death/birth event, including appearance, disappearance, split, unite, reallocate, replace and produce. Based on the practical facts, we define six kinds of death/birth event, including appearance, disappearance, divide, split, unite, and merge into, as showed in Fig.2, the &n in figure is the identification ID.
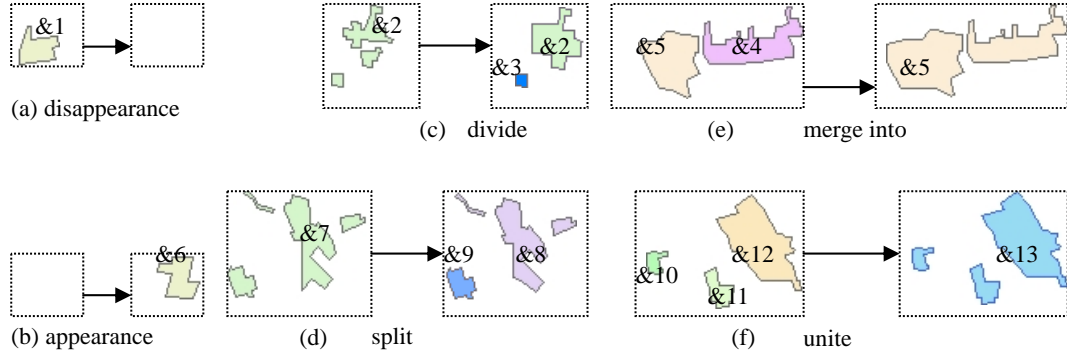
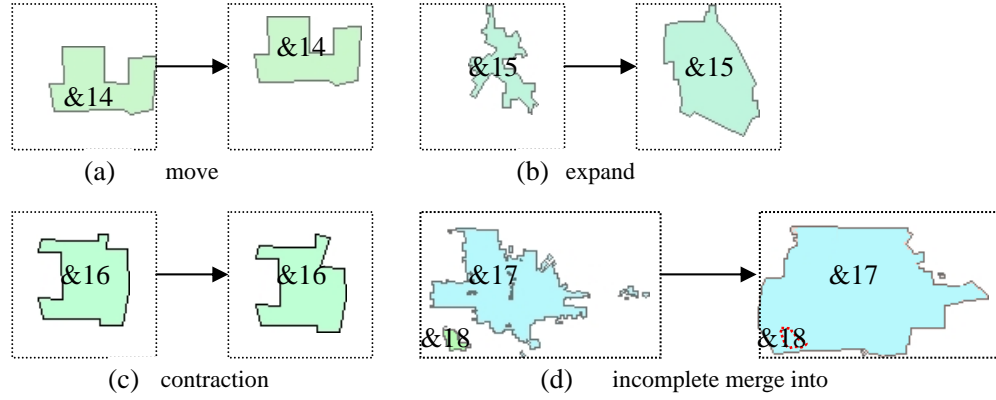**Fig.2.** The change process resulted from the birth or death events

Different death/birth processes correspond to different update operations. This means the different death/birth processes generate different increment information. The Fig.2 shows six kinds of increment information. In Table 1, the increment information and update operations resulted from the death/birth changes are given. For instance, Fig.2 (a) is the disappearance process, the update operation is "Delete" and the data delta is $\triangle = -[\&1, \&1.A.V, \&1.S.G]$. The increment information is Increment=[disappearance,delete &1,-&1]. Fig.2 (b) is the appearance process, the update operation is "Create" and the data delta is $\triangle = +[\&6, \&6.A.V, \&6.S.G]$. The increment information is Increment=[appearance,create &6,&6.A.V,&6.S.G]. Table 1 shows the increment information and update operation of death/birth evenets.

Table 1

The increment resulted from the birth or death evenets.

Change type	The instance of change	Update operation	Increment information
Appearance	&6 Appearance	{Create &6}	Appearance,Create &6,+[&6,&6.A.V,&6.S.G]
Disappearance	&1 Disappearance	Delete &1	Disappearance,Delete &1,－&1
Unite	&10 、 &11 、 &12 Unite as &13	Create &13,Delete &10,Delete &11,Delete &12	Unite,{ Create &13,Delete &10,Delete &11,Delete &12},{ － &10, － &11, － &12,+[&13, &13.A.V, &13.S.G]}
Merge into	&4 Merge into &5	Update the Graphics of &5,Delete &4	Merge into,{ Update the Graphics of &5,Delete &4},{－&4,&5.S.－G/+G'}
Split	&7Split into &8 and &9	{Create &8 and &9, Delete&7}	Split,{Create &8,Create &9,Delete &7},{+[&8, &8.A.V1, &8.S.G1],+[&9, &9.A.V2, &9.S.G2], － &7}
Divide	&3 is divided form &2	{Create &3,Update the Graphics of &2}	Divide,{Create &3,Update the Graphics of &2},{+[&3, &3.A.V1, &3.S.G1],&2.S.－G/+G'}

[4, 5] defined seven kinds of evolution event, including move, deformation, contraction, rotation, discoloration, density increase and texture change. Based on the practical facts, we define four kinds of evolution event, including move, contraction, expand and incomplete merge into, as showed in Fig.3, the &n in figure is the identification ID. Table 2 shows the increment information and update operation of evolution evenets.
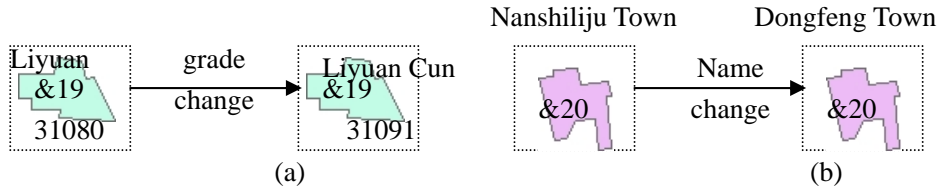
Fig.3. The change process resulted from the spatial evolution events

Table 2

The increment resulted from the evolution events.

Change type	The instance of change	Update operation	Increment information
expand	The area of Geographical feature &15 increases. (&15 doesn't exist in the old database)	Create &15	Expand, Create &15,+[&15, &15.A.V, &15.S.G]
expand	The area of Geographical feature &15 increases. (&15 exists in the old database)	Update the graphics of &15	Expand, Update the graphics of &15, &15.S.－G/+G'
contraction	The area of Geographical feature &16 decreases. (&16 exists in the new database)	Update the graphics of &16	contraction, Update the graphics of &16, &16.S.－G/+G'
contraction	The area of Geographical feature &16 decreases. (&16 doesn't exist in the new database)	Delete &16	Contraction , Delete &16,-&16
move	The position of a geographical feature &14 moves east 600m.	Update the graphics of &14	move, Update the graphics of &14 ,&14.S.－G/+G'
incomplete merge into	&18 incomplete merge into &17	Update the graphics of &17, Update the graphics of &18	incomplete merge into, { Update the graphics of &17, Update the graphics of &18},{ O1.S.－G1/+G1',O2.S.－G2/+G2' }

Current researches can't give the change type of attribute event. Based on the practical facts, we define two kinds of attribute change, i.e. the relative attribute change and the generic attribute change, as showed in Fig. 4. Table 3 gives the increment information of the attribute change.

Fig.4. The change process resulted from the attribute evolution event

Table 3

The increment resulted from the attribute evolution processes

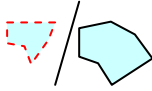| Change type	The instance of change	Update operation	Increment information
grade change	The grade of &19 changed from town to village	Update the Grade and Name Field of &19	The grade change, Update the Grade and Name Field of &19,{&19.Grade. － 31080/+31091,&19.Name. － Liyuang Town/+Liyuan Village}
Name change	The Name of &20 changed from 'Nanshiliju Town' to 'Dongfeng Town'	Update the Name Field of &20	Name change, Update the Name Field of &20 , O.Name.－Nanshiliju Town /+ Dongfeng Town

4. Application to the Case Study

Based on the above method, we developed software to extract the increment information. The core component of the software is the rule database. The rules are established according to the change type and the corresponding increment information. Every change type can be expressed as one or more rules. For example, the "Expand" of the residential feature can be expressed two rules. Fig. 5 gives some instances of increment information.

Rule 1. IF feature R has the change "Expand" And R exists in the old database Then increment information=[Expand, Update the graphics of R, R.S.－G/+G'].

Rule 2. IF feature R has the change "Expand" And R doesn't exist in the old database Then increment information=[Expand, Create R, +[R, R.A.V, R.S.G]]

The flow of the software is divided into 6 steps.

1) Load data. Load the old and new data of the same feature and same area.

2) Data pretreatment. check the invalid data and amend it.

3) Compute change types. based on the old and new data, extract the change types of every feature.

4) Correspond rules. find the rules from database based the change types.

5) Extract increment information

Old data	New data	Increment information
TN: 245781,GB:31090, Name: Dongfeng Town,	TN: 245781,GB:31090, Name: Dongfeng Town, Change type: Expand	Expand， Update the graphics of 245781
	TN: 240241,GB:31090, Name:Dali village， Change type: Expand	Expand， Create 240241， [TN:240241,GB:31091, Name: Dali village,]

Fig.5. The instances of increment information

5. Conclusions

The focus of this work was on how to help the exchange of updating information between a spatial data producer and a user. In order to more clearly distinguish the increment change of geographical object, we analyses the increment information of the geographical feature.

A new model of increment information is proposed based on the spatio-temporal change type and data delta. The model not only clearly represents the semantics of change but also describes the data delta of change. Based on the model, we can get the reason of the change. In this work, we define 15 kinds of change type, including 3 kinds of subjective change and 12 kinds of objective change. And the increment information expressions of every change are given. The method can be used to extract the increment information. But the method proposed in this work can't be used to express the increment information of subjective change.

Acknowledgment

6. Reference

[1] T. Badard, D. Richard, Using XML for the exchange of updating information between geographical information systems, Computers, Environment and Urban Systems, (2001) 17-31.

[2] T. Badard, Towards a generic updating tool for geographic databases, In Proceedings of GIS/LIS'98, Annual Conference and Exposition, Fort Worth, TX., 1998.

[3] L. Raynal, Some elements for modelling updates in topographic databases, In Proceedings of GIS/LIS'96, Annual Conference and Exposition, Denver, Colorado, USA, 1996.

[4] C. Claramunt and M. Thériault, Toward Semantics For Modeling Spatio-Temporal Processes, Within GIS. Advances in GIS II(SDH96), the Netherlands, Taylor and Francis,1996.

[5] H. Shu, Principles of Conceptual, Formal and Logical Spatio-temporal Data Modeling, Wuhan, Docter Dissertation of Wuhan Technical University of Surveying and Mapping, 1998 (in Chinese).

[6] X.G. Zhou, J. Chen, J.J. Zhu, Event-based Incremental Updating of Spatio-Temporal Database, Journal of Image and Graphics, 11(10)(2006) 1431-1438 (in Chinese).

[7] Y. Xia, X.Y. Zhu, W. Guo, Research on spatial data version management based on ArcSDE, Computer Engineering and Applications, 43(14)(2007)14-16 (in Chinese).

[8] L.Spéry, C. Claramunt, et al., A Spatio-Temporal Model for the Manipulation of Lineage Metadata, Geoinformatica, 5(1)(2001) 51-70.

[9] T. Badard, D. Richard, Using XML for the exchange of updating information between geographical information systems, Computers, Environment and Urban Systems, 25(2001)17-31.

[10] Y. Lin, W.Z. Liu, J. Chen, Modeling spatial database incremental updating based on base state with amendments, Proceedings of the International Conference on Mining Science & Technology (ICMST2009), Xuzhou, China, 2009.

[11] N. Wang, H.B. Xu, N.B. Wang, Incremental Maintenance of Global Templates in Heterogeneous Data Integration System, Journal of Software, 10(4)(1999)382-389(in Chinese).

[12] C. Claramunta, C. Parent, Modelling concepts for the representation of evolution constraints, Computers, Environment and Urban Systems, 27 (2003) 225-241.