A Bridge ontology Approach for Locating Changed Objects in Data Update Applications

Chiao-Ling Kuo
Department of Geomatics,
National Cheng Kung University
No.1, University Road,
70101 Tainan City, Taiwan
chauoling@gmail.com

Jung-Hong Hong
Department of Geomatics,
National Cheng Kung University
No.1, University Road,
70101 Tainan City, Taiwan
junghong@mail.ncku.edu.tw

Abstract

Diverse geospatial data brings Geographic Information System plenty of various applications. One of the most critical issues of GIS interoperability is to clarify the semantics of data. Thereafter users could get more exactly consistent results. To enable the automatic change detection in data updating application, we propose an ontology-based mechanism to detect potential changed features (including shape changed or object changed). Relationships of bridge ontology and overlap area of data are two major judgment factors in the criteria. Four processing strategies for various types of data of different dimensionality (point, line and polygon) were concluded. Our approach aims to not only facilitate to discover candidate features for updating among heterogeneous domain efficiently, but also demonstrate the value of semantics interoperability in GIS-based applications.

Keywords: Bridge ontology, Semantics relationship, Data updating, Detection.

1 Introduction

Geospatial data is one of the most important components in GI (Geographic Information) application. As the technologies of producing geospatial data have improved and each field could always produce data to fit domain's application, abundant of various geospatial data exist nowadays. To conquer the heterogeneity of interoperability and achieve the integration in semantic level, many studies use ontology to present domain knowledge and enable the communication of heterogeneous data between different domains [1]. Ontology allows domain experts to formally present the semantics of data [2]. With appropriate encoding language (e.g. OWL[1]), it can be also readily implemented in a machine-readable environment [3, 4]. After the semantics of data are clarified, the precise and correctness of integration results could be promised. For GI field, geospatial data (resources) should be represented clearly and precisely before being involved in an application. Users can then obtain reliable results and more useful strategies to specific application.

The growth of geospatial data volume in recent years has been extremely overwhelming. As time goes by, geospatial data needs to be continuously updated to reflect the situation in reality. Despite its importance, this process is nonetheless both time and cost consuming. From the perspective of data-reuse and semantic interoperability, we intend to propose an automatic change detection mechanism using heterogeneous resources of geospatial data in this paper.

According to previous studies, the most important step of semantics interoperability is to establish communication between different ontologies and make them consistent [5]. Commonly used approaches based on similarity algorithms are ontology mapping, matching, alignment, bridge, and merge, etc. [6-10]. Instead of changing or merging ontology, we reserve the original ontology, and then use the relationship of bridge ontology to help to determine potential changed data in our procedure. The bridge ontology means bridging the semantics of two domains[10]. The development of such relationships, however, largely depends on individual application needs. Stoimenov et al. proposed four types of relationships to present the relation of local ontology and local and top-level ontology [11]. Uitermark et al. proposed semantic relationships domain ontology and application ontology, the relationships of object classes to integrate topographic data sets [12, 13].

For change detection task, photogrammetry and remote sensing are widely used to discover area of changes [14, 15].

The past discussions of changes in ontology mainly focused on comparing the difference and similarity of different versions [16-18]. About changed results recording, Kauppinen presented seven kinds of change type and change ontology for time series geospatial data in historic regions [19]. Therefore, a complete automatically detection procedure for geospatial data updating is need. In this paper, we propose an approach to combine ontology and GIS technology to detect change efficiently.

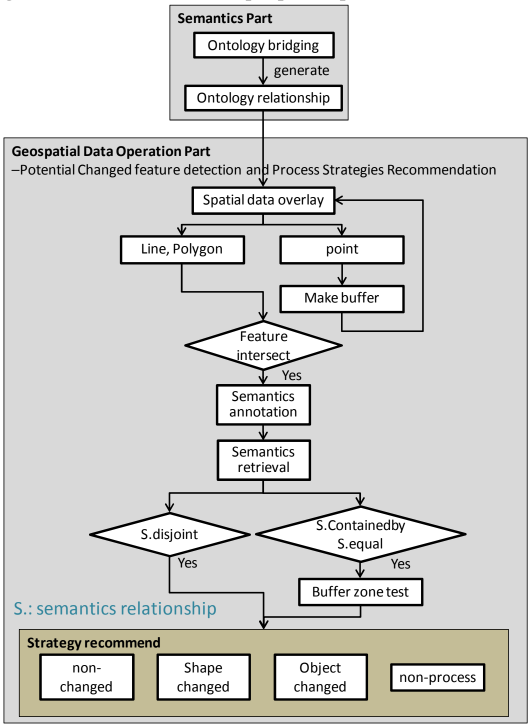
We argued that the change detection does not necessarily need to use the same type of data, a dataset from another domain may also prove to be useful for identifying possible changes. Under such circumstances, we will be able to take full advantages of the already abundant and continuously growing georesource to meet our application demands. The basic idea is to use new data from other domains (with their own concepts) to update existing data based on ontology approach. The rest of paper is organized as follows. Section 2 discusses the methodology used for identifying the changes. Section 3 implements two use cases to demonstrate our approach. Finally, we give our conclusions and future works in Section 4.

[1] http://www.w3.org/TR/owl-features/

2 Methodology

The major goals of this study are to detect potential changed data in certain domain and generate processing strategies for detected features. The comprehensive detection procedure takes both semantic and geometric constraint into consideration. From a semantic perspective, bridge ontology is used to present the semantic relationship of two heterogeneous data and those relations are capable of identifying oncoming performed features. Meanwhile, the geometric perspective considers the spatial dimensionalities and evaluates the usage of every possible scenario. Three kinds of dimensions of data will be discussed. Therefore, semantics, dimensions, and measure accuracy are three crucial factors in our methodology. In this research, the discussion will be limited in determining if a feature changes or not. That is, we will discuss about how a potential changed feature is discovered, not for performing on changing the properties and shape of features.
Fig. 1 illustrates the workflow of the proposed procedure. It consists of two parts: *Semantics part* and *Geospatial Data Operation Part*. Firstly, the concept of bridge ontology is proposed to record the relationship between concepts of two different domains. These recorded relationships serve to clarify the semantics between individual domain of geospatial data, so candidate data could have the semantic relationship (that will be processed). Secondly, the selection of the candidate data is determined by the geospatial data operation part. As data update implies the situation at a particular location changes over time, we overlay the target data (data being updated) and source data (data being used for update reference) and develop corresponding procedures according to if features from two different datasets overlap with each other.

Figure 1: The workflow of proposed procedure.



Point is a kind of GIS data representation. An object based on a coordinate reference system is abstracted in reality. In practical, it is hard to fit two points measured from time to time as measurement always brings out accuracy issue even more difficult to fit two points for the same object in two different measure systems. In order to raise the possibility of intersection, the point object should generate buffer zone, and then doing overlay testing. Once two features have intersection, that means target data has high possibility to change. The target data in non-overlaid features case caused by measure accuracy is also likely to change, but it has weaker connection than intersect case relatively. The measure accuracy issue in non-overlaid features case will be discussed in the future. In order to rapidly choose candidate features, we exclude non-overlaid features. For intersecting features, all of them will be considered as candidate data and marked semantic relationship by semantics annotation procedure [20] for later processing. Additional tests on semantic relationship and buffer zone test are needed to evaluate the type of possible changes. Different types of changes may need different data update strategies. Finally, the system shows the types of changes and prompt users to make data update decisions in the future. The following further discuss the preference of choosing target and source domain to form bridge ontology, semantic relationships of concepts of bridge ontology, feature types, the ratio of intersection, , and changed types in more detail.

2.1 Preference of Choosing Target and Source Domain to form Bridge Ontology

The success of bridge ontology largely depends on what domains of ontology are selected. The preference of ontology selection is listed below:

1. Concepts of different domains are conceptualized to present the same reality and they may have the same or similar meaning in individual domain (from concept perspective).
2. Geospatial data are related in conceptual level but they are represented in different dimensions in data level. Nevertheless, we also intend to obtain information from them based on various dimensions operation (from data dimension perspective).
3. Particular application needs exist between the two selected domains, e.g. data updating, related-data search (from application perspective)

2.2 Semantic relationships of concepts of bridge ontology

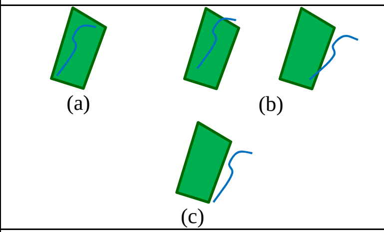
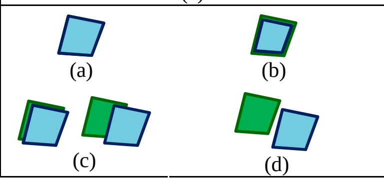
The semantic relationships between concepts of two selected domains are recorded by bridge ontology. Since these relationships are used to indicate how similar two concepts are, they play a crucial role to determine what applications can be developed between the two selected domains. Five relationships are proposed to describe the concepts' relationship, namely, *equal, contained-by, contain, overlap, and disjoint*. Assume concept A denotes the target concept and concept B represents the source concept, the meaning of the above five relationships are defined as follows:
Semantically speaking, the relationship *Equal* implies concept A has the same meaning as concept B even if they belong to two different categories and have different naming. The

relationship *contained-by* implies concept A is a special kind of B, that is, concept B has wider range than A. Its inverse relationship *contain* implies concept B is a special kind of concept A. The relationship *overlap* implies concept A and B share a common part, but each one has its own exclusive parts. Finally, the relationship *disjoint* implies concept A and B are totally different and has nothing in common. These five relationships are mutual exclusiveness, meaning the situation between two concepts can be only represented by one relationship as long as the relationship is bridged from bottom and the most detailed concept. To increase the certainly of recorded relationships, the recording priority is suggested to be equal > contained-by> contain > overlap > disjoint. For instance, when concept A is equal to concept B and contained-by concept C simultaneously, the former one (concept A is equal to concept B) has a higher priority.
At this stage, the relationships between concepts are established by human experts. Since the major focus of this paper is the use of bridge ontology in data update applications, how to analyse the difference of concepts and what comparing algorithm is used are not included in this paper.

2.3 Feature Types

Spatial dimensionality is an important factor in data update application because features of lower dimensionality cannot really "update" features to higher dimensionality. Based on the commonly used point-line-polygon structure, we summarize 15 different scenarios from six primitive combination of dimensionality (point vs. point, point vs. line, point vs. polygon, line vs. line, line vs. polygon, and polygon vs. polygon). For potential change detection, we use same evaluating approach in point vs. line and line vs. point so they could be viewed as only one primitive combination. Identically, point vs. polygon/polygon vs. point and line vs. polygon/ polygon vs. line follow the same principle as above. Fig. 2 illustrates point vs. point, line vs. polygon, polygon vs. polygon cases. The scenario of equal (e.g. I(a) and III(a)) will occur in the same dimension. Most cases are overlap (e.g. II(a), II(b), III(b), and III (c)) and disjoint(e.g. I(b), II(c), and III(d)).

Figure 2: The operation types of different dimensions.

Dimension	Situation	Note
(I) Point VS. Point		(a)equal (b)disjoint
(II) Line VS. Polygon		(a)overlap (within) (b)overlap (c)disjoint
(III) Polygon VS. Polygon		(a)equal (b)overlap (within) (c)overlap (d)disjoint

In overlap case, they will directly be chosen as candidate features. In disjoint case, the buffer zone will firstly be made for point object and then go into the estimation procedure. Others will not be processed until they find the intersecting features.

2.4 The Ratio of Intersection

Winter [21] proposed how to measure the similarity for two regular regions and he especially used ratios to represent similarity and dissimilarity for varied situations. The positional accuracy is an essential type of data quality of geospatial data. To provide a reliable data update suggestion, the influence of positional accuracy must be seriously considered. Hence, we use buffer zone to deal with accuracy problem, and then estimate the similarity by the ratio of intersection. The buffer distances depend on operated data. For instance, topographic maps' geometric accuracy is not greater than 0.3mm, so the buffer distance of 1:5000 scale map is 1.5 m for general use. After generating the buffer zone, the ratio of intersection would be calculated. If the ratio R (R_1 and R_2) is smaller than threshold T, the target object could be marked (detailed discussion of estimation is presented in section 2.5). By our experiment, threshold value T is set 75%. Equation (1) shows the ratio of intersection of two buffered objects. The domain of values of R_1 and R_2 is [0, 1]. R_1 is suited in the same dimension object (before making buffer zone), e.g. polygon vs. polygon. It means how much buffer area of *Max*(A, B) is overlapped. Once the ratio is 1 that means A equal to B in spatial. R_2 is suited in different dimension objects, e.g. line vs. polygon. R_2 is the ratio of how much buffer area of lower dimension object within a higher dimension object. The ration of R_2 is 1 that means an object A is totally within object B. And A equal to B in spatial.

$$R_1 = \frac{A \cap B}{Max(A,B)}, \quad R_2 = \frac{A \cap B}{LowerDimensionArea(A,B)}, \quad 0 \le R_1, R_2 \le 1 \tag{1}$$

where *A* and *B* are the area of feature A and feature B, R_1 and R_2 is the ratio of intersection with *Max*(A, B) and *LowerDimensionArea*(A, B). *LowerDimensionArea*(A, B) means the area of lower dimension between A and B.

2.5 Change Types

Combining semantics relationships, ratio of intersection and threshold value, we conclude four change types: *unchanged*, *shape change*, *object change*, and *non-process*. *Unchanged* means that target feature keep shape and property as it is. No more operation would be done. *Shape change* means the feature intrinsically is the same as before but its shape is changed. *Object change* is that the target feature is changed intrinsically, so its shape and attributes should be updated. As the semantic relationship of contain and overlap have complex semantics, we would not deal with these two case. Therefore we give the *non-process* type for them.

Table 1: The change types.

semantics	R<Threshold (*T*)	Threshold (*T*)<=R
equal	shape change (1)	unchanged (2)
disjoint	object change (3)	object change (3)
contained-by	shape change (1)	unchanged (2)
contain	non-process(4)	non-process(4)
overlap	non-process(4)	non-process(4)

3 Implementation

To demonstrate the proposed detection approach, we individually give two cases of shape change and object change type in river course change and new road constructed of farm, which is shown in section 3.1 and section 3.2.

3.1 Shape Change Case Study- River Course Change (equal in semantics)

B1 is source data (new data) of river course from domain B, and A1 is target data (ole data) of river course from domain A. We would like to know if A1 is changeable by using B1 to evaluate. Then, we use A1 to overlap B1. Because they are overlap, they could be annotated equal in semantic relationship that equal relationship is record in concept. Following up we retrieve the semantic relationship of them. Their semantic relationship is equal, and then put them into the buffer test procedure. The buffer area is generated by each feature's buffer distance which is decided by domain's regulation (usually considered from data's accuracy). The overlap area of buffer zone is less than the threshold T. Therefore, A1 has high possibility that the range of A1 is changed. The strategy of B1 for A1 is shape change. Fig. 3 illustrates the procedure of B1 operating with A1, both of them are polygon. Fig. 4 is the polygon versus line case, identically, the line feature (B2) generates the buffer zone and then doing the buffer and threshold test. Finally, obtaining the operation strategy is shape changed.

Figure 3: The procedure of shape change case study -river course change (polygon vs. polygon).

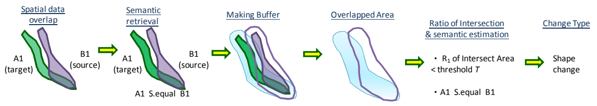
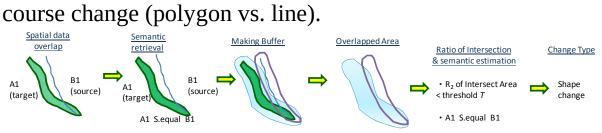


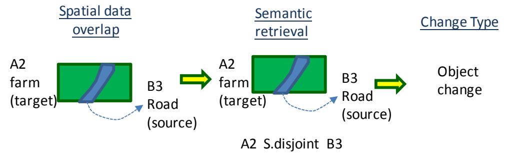
Figure 4: The procedure of shape change case study -river course change (polygon vs. line).



3.2 Object Change Case Study- New Road Construction (disjoint in semantics)

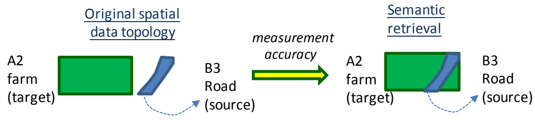
The second experiment is the semantic disjoint case. This kind of situation is common in daily life that the original place has new constructions or its content is altered. We show how a new road is potential constructed on a farm by semantic disjoint relationship detection in the following. Feature A2 is a farm region (polygon) from domain A, and as time goes by, B3 is a road which is constructed and record in domain B. A2 use B3 data to determine if it is changed. As the processing procedure shown in Fig. 5, A2 overlap B3, and their semantic relationship is disjoint. So, it is easily to present that the changed type is object change.

Figure 5: The procedure of object changed case study -new road construction.



The measurement accuracy issue of feature overlapping is not included in this paper. Like the Fig. 6 diagrams that A2 and B3 are separated in left figure, though B3 overlaps A2 in right figure which is caused by measurement accuracy. In our approach, it can not process the inconsistency but deal with the semantic detection and give the strategy for final situation of features.

Figure 6: The measurement accuracy case.



4 Conclusions and Future Works

The importance of semantic interoperability has been well recognized by GI community in recent years. How to resolve the semantic heterogeneity of data from different domains is especially interesting due to the integrated nature of GIS. In this paper, we proposed an approach to use bridge ontology to rapidly detect potential changed features by heterogeneous sources of data. Five relationships in bridge ontology and four change types for data of different dimensionality are respectively developed. We especially focus on developing procedures that can detect the potential changed data from the semantics perspectives from highly related data. In the future, we will continually refine the threshold value by different dimensions, types of data, even their measurement approaches. More works is going to be how to record the information of checked data and utilize this mechanism to perform series heterogeneous data operation.

Acknowledgement

Part of the result in this paper is from the project sponsored by National Science Council (NSC 101-2211E-006-274), Taiwan.

References

[1] Y. Bishr, "Overcoming the Semantic and Other Barriers to GIS Interoperability," *International Journal of Geographical Information Science*, vol. 12, p. 16, 1998.
[2] C. M. Keet, "Aspects of ontology integration," 2004.

[3] D. McGuinness and F. Van Harmelen, "OWL web ontology language overview," *W3C recommendation,* vol. 10, pp. 2004-03, 2004.

[4] D. R. S. V. Maniraj, "Ontology Languages – A Review," *International Journal of Computer Theory and Engineering,* vol. 2, pp. 887-891, 2010.

[5] N. F. Noy, "Semantic integration: a survey of ontology-based approaches," *SIGMOD record,* vol. 33, pp. 65-70, 2004.

[6] P. Shvaiko and J. Euzenat, "Ontology matching: state of the art and future challenges," 2012.

[7] N. Choi, I. Y. Song, and H. Han, "A survey on ontology mapping," *ACM Sigmod Record,* vol. 35, pp. 34-41, 2006.

[8] J. Euzenat, "An API for ontology alignment," *The Semantic Web–ISWC 2004,* pp. 698-712, 2004.

[9] N. F. Noy and M. A. Musen, "The PROMPT suite: interactive tools for ontology merging and mapping," *International Journal of Human-Computer Studies,* vol. 59, pp. 983-1024, 2003.

[10] B. Xu, P. Wang, J. Lu, Y. Li, and D. Kang, "Bridge ontology and its role in semantic annotation," in *Cyberworlds, 2004 International Conference on*, 2004, pp. 329-334.

[11] L. Stoimenov, A. Stanimirovic, and S. Djordjevic-Kajan, "Discovering mappings between ontologies in semantic integration process," in *Proceedings of the AGILE,* 2006, pp. 213-219.

[12] H. Uitermark, P. van Oosterom, N. Mars, and M. Molenaar, "Ontology-based geographic data set integration," 1999, pp. 60-78.

[13] H. T. Uitermark, P. J. M. van Oosterom, N. J. I. Mars, and M. Molenaar, "Ontology-based integration of topographic data sets," *International Journal of Applied Earth Observation and Geoinformation,* vol. 7, pp. 97-106, 2005.

[14] K. Green, D. Kempka, and L. Lackey, "Using remote sensing to detect and monitor land-cover and land-use change," *Photogrammetric Engineering and Remote Sensing,* vol. 60, pp. 331-337, 1994.

[15] A. Singh, "Review Article Digital change detection techniques using remotely-sensed data," *International journal of remote sensing,* vol. 10, pp. 989-1003, 1989.

[16] G. Flouris, D. Manakanatas, H. Kondylakis, D. Plexousakis, and G. Antoniou, "Ontology change: Classification and survey," *Knowledge Engineering Review,* vol. 23, pp. 117-152, 2008.

[17] P. Plessers and O. De Troyer, "Ontology change detection using a version log," *The Semantic Web–ISWC 2005,* pp. 578-592, 2005.

[18] N. F. Noy and M. Klein, "Ontology evolution: Not the same as schema evolution," *Knowledge and information systems,* vol. 6, pp. 428-440, 2004.

[19] T. Kauppinen, J. Väätäinen, and E. Hyvönen, "Creating and using geospatial ontology time series in a semantic cultural heritage portal," *The Semantic Web: Research and Applications,* pp. 110-123, 2008.

[20] E. M. Klien, "Semantic annotation of geographic information," Phd thesis, Institute for Geoinformatics, University of Muenster. Muenster, Germany, 2008.

[21] S. Winter, "Location similarity of regions," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 55, pp. 189-200, 2000.