Research Journal of Applied Sciences, Engineering and Technology 5(2): 563-567, 2013
DOI:10.19026/rjaset.5.4990
ISSN: 2040-7459; E-ISSN: 2040-7467

Submitted: May 14, 2012 Accepted: June 12, 2012 Published: January 11, 2013

Research Article

A Study on Theory and Method of Spatial Vector Data Conflation

H.E. Dongcai
Taiyuan University of Technology, Taiyuan 030024, China

Abstract: In this study, we set up the framework of vector spatial data conflation in multi-source vector space and discuss the flow of the special conflation of attribute data and geometry data. The constraints in map conflation are first derived and the coordinates displacements of the conflated objects are then obtained by topological correlation computation. The results of the map conflation test show that the proposed conflation model and algorithm are effective and has a high precision and better ability to maintain the characteristics of the conflated objects compared with the exist methods.

Keywords: Adjusting, attributes transfer, geometric matching, multi-scale, vector data conflation

INTRODUCTION

With the wide application of spatial information technology in the national economy, spatial data has been applied in many fields like environment, territorial resources, disaster prevention, city planning, municipal administration, scientific research and national defense education, etc. However, since a unified standard for it is lacked, there is repeated collection and production of spatial data by different departments, which causes the omni-dimentional and multi-tiered difference of existing spatial data and the huge waste of various limited resources. Furthermore, the scope, precision and timeliness requirements on spatial data from various fields improve increasingly with the constant deepening of the application of spatial data. Therefore, it has become an important research subject for GIS currently on how to have an effective integration of multi-source spatial data so as to obtain a spatial database meeting the application requirements (He *et al*., 2010a).

From the research, it's found that the effective technical way to solve the issue mentioned above is to use conflation technology of spatial vector data. For this reason, this paper studies the theory and method of conflation which refers to extracting the information needed from the spatial vector data the similar or same areas with different databases, data precision and data models to form new vector data. It has important theoretical significance and application value in quickening the updating of spatial information, reducing the production cost of spatial data and improving the quality of existing spatial data.

THE ARCHITECTURE OF MULTI-SOURCE VECTOR DATA CONFLATION

Vector space vector data conflation includes space data preprocessing (data integration) and vector space data conflation. Data integration and conflation are not isolated processing and integration is the basis of conflation and the first step of conflation. Data conflation is a processing that integrates the various vector space data from different sources, formats, scales, multi-projecting types and world coordinate systems in logic or physics to realize a share of vector information. The integrated vector space data still remains the original attributes and has no essential change. Data integration first needs to realize the union of data model, space reference and semantic coding (He *et al*., 2010b; Kovalerchuk *et al*., 2005; Spaccapietra, 2008). On the basis of integration, the data conflation mainly includes the making of data conflation strategy, the matching and recognition of homonymic entities and the combination of graphic data and attribute data.

Before the data conflation, the first step needs to do is to analyses and evaluate the multi-source data, decide the used degree of various data source by combining actual needs and make effective data conflation strategies to real applications. Homonymous entity matching is the key technology in data conflation and the emphasis and difficulty in this study. Matching algorithm decides the difference of entitative sizes according the entitative space information and attribute information and resolves not only the recognition of one-to-one homonymic entities, but also the identifications for one-to-many, many-to-one and many-to-many homonymic entities. Matching algorithm can select the optimal method to match data

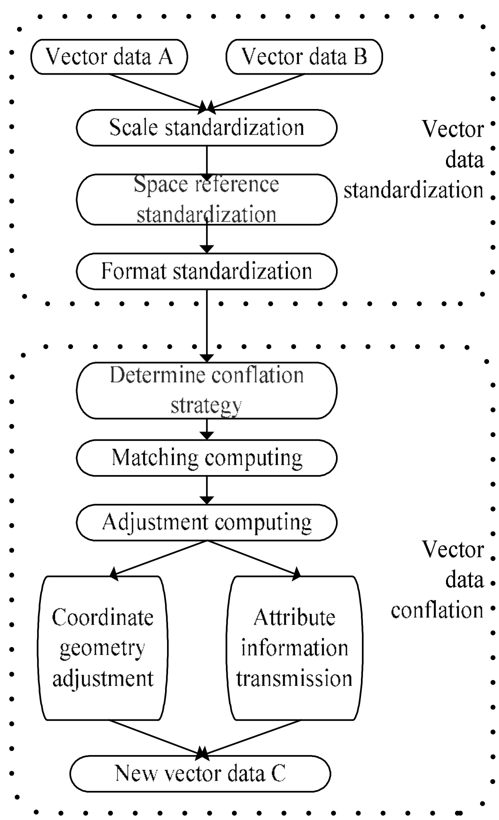

Fig. 1: Architecture and flow of vector space vector data conflation

from many matching methods according to different features of objects. The common matching algorithms are geometrical matching, semantic matching and matching strategy. Data conflation includes the conflation of geometrical data (such as image conflation) and the conflation of attribute data. And the conflation's aim is to produce the new and higher-quality data. The architecture of vector space vector data conflation is shown in Fig. 1.

VECTOR DATA STANDARDIZATION

The integration of multi-source vector space data is a basis of vector space data conflation. Space vector data integration and conflation first require data model and semantic coding of entities. Multi-source space data integration is to use various mathematic transform tools to integrate space data from different sources into a GIS, forming the recognizable data to GIS. Its core task is to ignore the difference between various data models and integrate the different data sources together, with which users can access transparently multi-source and heterogeneous vector space data (Sholokhov, 2007; Hoseok, 2004).

The union of space reference: GIS space reference involves reference ellipsoid, coordinate system, level origin, map projection, zoning and many other factors. Therefore, the GIS space reference is a complex problem. To integrate space data, a unified space reference is required greatly to ensure the consistency, compatibility and convertibility of space data. Without unite space reference, there is no integration of space data or realization of data's share and application. The GIS space reference can be divided into level reference and elevation reference according to technology. The main factors that affect level reference are reference ellipsoid, coordinate origin, map projection and plane control network and the main factors that affect elevation reference are level origin and level network. The union of space reference needs the union of coordinate system, map projection transform and elevation reference.

The union of data model: In recent years, many experts and scholars have done a great deal studies on GIS space data model. With the development of GIS, computer technology, database theory and technology, GIS space data model goes through the following types: cartography, topological relation data model, entity-oriented data model and object-oriented data model.

In order to cast space relationship to data structure that is fit for the computer to process, the united model is necessary to be built. This united data model must be compatible with various data models belonging to different data source, so that various data source can be transformed to the united data model.

The method of data standardization: Different from general data of business administration, vector space data can't implement integration easily due to the different understands of space phenomenon and difference in the definitions, representations and storement types of space objects. Solving multi-format data transform has still been an important problem in development of GIS application system. At present, the methods of data integration probably can be divided into three types: external data transform mode, direct data access mode and space data interoperation mode. External data transform is to read and write internal format, external format or some standard format transformed by itself directly. It is an indirect data transform format and other data formats are copied into the database or files of current systems through being transformed to corresponding data formats by special format transform programs. This is the main method of data transform in current GIS systems. Direct data access is to access other software's data format directly in the GIS system, that is to say, it transforms an internal data file in a system into another data file in another system. Users may use single GIS software to access multiple data formats. Direct data access not only avoids complex data transform, but also didn't require users to possess the Parasitifer software of this data format and even not requires running this software, so its substance is still a method of data transform. Space data interoperation is a regulation that Open GIS Consortium makes. GIS interoperation means to acquisition some information that needed transparently under the conditions of Isomerous databases or distributed computation and the understanding of each other between GIS users. Data interoperation provides a new train of thought and regulation. It makes GIS open so that space data can realize centralized management and distributed storage as well as sharing.

FEATURE MATCHING

The spatial graph data processing in GIS to a large extent is one kind of topology transformation, that means the relationships of graphic data remains unchanged in continuous transformation. Feature matching actually is special processing for space graph data .Ideally, two groups of matched graph should keep the topology consistent .Feature matching algorithm can be divided into three types according to the characteristic of different space objects : Dot entity matching, line entity topology matching ,plane entity topology matching.

Dot entity topology matching: Dot entity matching is a process that recognizes the data in different datasets but reflecting the same feature point in the same place in a real world. Saalfeld points out that two factors in homonymous entity matching can be considered as follows: one is the line entity numbers that linked to a dot, namely, the degree of the dot, the other is the line entity's direction that linked to the dot. Saalfeld designed a coding method called spider coding to realize a homonymous entity matching when matching road nodes in block map. As there is a great quantity of leak match and wrong match in dot entity matching methods such as traditional spider coding method, this study introduces an optimal spider coding method. The improved spider coding method is as follows:

- Make a node original node to set up X-Y coordinate axis and make Y axis original direction and divide the line direction angle into 16 angle areas that not across
- Use a 32 bits long code to represent a node's structure and make each two bits to correspond to one angle area
- If a dot's line direction is in i-th angle area and the dot's line number is one, the corresponding area's coding value is 01 and if the dot's line number is two, the corresponding area's coding value is 11, or else 00

Based on the representation of spider coding, the comparing between two node's structures only requires the 'AND' operation according to bits of their codes and then count the numbers whose value is 1.

When matching dots, the following scheme is feasible. According to distance threshold, the candidates of matching dots are determined. If only there is only one candidate matching dot, we consider it matching dot, otherwise we compare the similarity of matching dot and waiting matching dot and make the maximum the final matching dot.

Line entity topology matching: In general, we use topology similarity to match line entities. Before matching, node matching is usually preferential. If the starting points of two line entities are a match for their terminal points and the shapes of the two line entities are similar, then we may think that the two lines are consistent with each other.

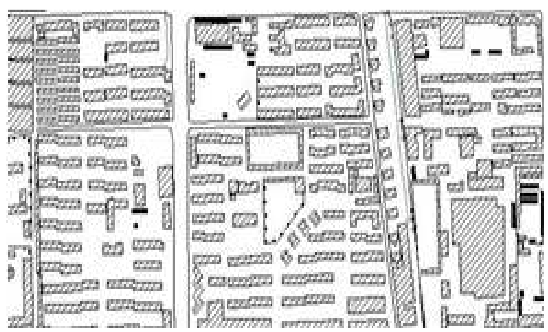

Fig. 2: 1:10000 orthophoto map

Fig. 3: 1:50000 orthophoto map

Fig. 4: Conflation result

Plane entity topology matching: The concepts that measure the topology relationship of plane objects are split ratio, overlay ratio, compact joint degree and middle degree. This study adopts the algorithm of same entity matching based on comprehensive weighted value, improving the traditional algorithm that makes entity matching according to a single index by introducing the theory of human brain visual graphic matching into matching calculation and determines the crucial link of the algorithm in accordance with the visual graphic theory and the rational threshold from repeated experiments: the selection of all parameters and their weight values in the matching calculation. The result has proved the algorithm greatly improves the computing accuracy compared to existing algorithms.

FEATURE ADJUSTING

In the process of feature matching, there is large amount of uncertainty, so some entities can't be matched. Aiming at those unsuccessfully matched entities, we adopts feature merging algorithm based on topological relationship to work out the optimal space position of unmatched features. The algorithm is as follows: all the features are decomposed into dot entity. Unmatched dot is viewed as unadjusted and the matched dot as matching dot. We start from every unadjusted dot, search and record the entire matching dots joint with it and the route length between the matching dot and the unadjusted. The route length is a weight value that is to say, when the route is short, the influence is enormous; otherwise and the influence is tiny. For any unadjusted dot P, its coordinate adjustment is decided by Qi (i = 1,…, N) the N dots. The route between P and Qi is Li (i = 1,…, N). The coordinate adjustment of is Qi and P is decided by (4), the weight value Wi equals 1/Li:

$$\begin{cases} \Delta X_p = \sum_{i=1}^{n} \Delta X_{Q_i} \dfrac{W_i}{\sum_{j=1}^{n} W_j} \\ \Delta Y_p = \sum_{i=1}^{n} \Delta Y_{Q_i} \dfrac{W_i}{\sum_{j=1}^{n} W_j} \end{cases}$$

A map conflation example: It takes the basic vector map of a city in 1:10,000 (Fig. 2) and 1:5,000 (Fig. 3) as a computing case to make implementation on the components of the combined computing strategy and algorithm under Arcgis 9.3 with Microsoft Visual C++ 6.0 and MATLAB 7.0 and validates the availability and reliability of the algorithm. The conflation result is as shown in Fig. 4.'

COLCLUSION

In this study, we studied detailed how to build the vector space vector data conflation framework and algorithm by abstracting the required information from different data source, precision and data models in vector space data according to user's new application. The example quoted in the conflation reflects that both the map conflation model and the algorithm proposed in this study is effective, It removes the data difference between distinct spatial vector maps, improves the precision of dot position and meets the demand of map conflation.

REFERENCES

Dechter, R. and I. Rish, 1994. Directional resolution: the davis-putnam procedure. Proceeding of 4th International Conference on Principles of KR&R, Bonn, Morgan Kaufmann, Germany, pp: 134-145.

He, D.C., J.J. Chen and J. Zhang, 2010a. A study on method of vector spatial data fusion based on topological relationship [C]. 3rd International Conference on Computer and Electrical Engineering, Chengdu, China, 9: 277-280.

He, D.C., J. Zhang and J.J. Chen, 2010b. A study on method of map conflation based on topological relationship [C]. International Conference on Electrical Engineering and Automatic Control, Zibo, China, 7: 293-296.

Hoseok, K., 2004. Topographically consistent map conflation for federal and local government [J]. J. Korean Geogr. Soc., 39(5): 804-818.

Kovalerchuk, B., M. Kovalerchuk, W. Sumner and A. Haase, 2005. Image conflation and change detection using area ratios. NASA, 5806: 418-429.

Sholokhov, A.V., 2007. Evaluation of the reliability of determining coordinates in navigation systems corrected for road maps [J]. Meas. Tech., 50(5): 507-514.

Spaccapietra, S., 2008. On Spatial Database Integration [J]. Int. J. Geogr. Inf. Sci., 8(3): 334-353.