Journal of Computing and Electronic Information Management

ISSN: 2413-1660

Research on Uncertainty Modeling and Quality Control for Spatial Data

Xueying Wu [1, 2, 3, 4], Hailan Shi[1, 2, 3, 4], Jianglong Shen [1, 2, 3, 4,]

[1]Shaanxi Provincial Land Engineering Construction Group Co.,Ltd. Xi'an 710075, Chin;

[2]Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of Natural Resources, Xi'an 710021, China;

[3]Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co.,Ltd. Xi'an 710021, China;

[4]Shaanxi Provincial Land Consolidation Engineering Technology Research Center, Xi'an 710075, China

Abstract: This article introduces the concept of spatial data such as location of road signs and traffic lights on Google Map, explains the uncertainty and types of spatial data, analyzes their sources, and briefly explains the impact. The methods of controlling and quantifying the uncertainty and error of spatial data are introduced, and the research trends of quality control for spatial data are discussed.

Keywords: Spatial data, quality control, uncertainty modeling.

1. Introduction to Spatial Data Quality - Take Google Map as an Example

The quality of spatial data refers to the uniformity of accuracy, consistency, and completeness of spatial data in expressing the spatial position, characteristics, and time of the entity, and the ability of the data to be suitable for different applications. The core of spatial data is quality. The production and quality control of data is an interactive process. The production of data is for application. The quality of the data will directly affect the reliability of the analysis results of Google Map and the realization of application goals. Spatial data of quality conforming to the standard is the basis for a comprehensive analysis system.

The data included in the Google Map is very jumbled, such as the identification and marking of road traffic signs, the location of traffic lights and so on. The process of obtaining spatial data is mainly to integrate the aerial photos, satellite images and GIS attribute information on the 3d model, so as to make the map data information very detailed [1].

2. Types of Spatial Data Errors and Uncertainties

For Google Map, data is done by interpreting, measuring, entering data, processing spatial data, and representing data in the real world. Each of these processes has the potential to generate errors, leading to a considerable amount of error accumulation. Uncertainty can be considered as the degree to which the true value cannot be affirmed. It is an inherent property of various entities or phenomena in nature. In form, it generally includes a range of true values. The larger this range, the greater the uncertainty of the data [2].

2.1 Errors in Spatial Data

Spatial data error can be divided into three types: original error, processing error and application error [3]. Original error refers to the inherent error of data and the error generated in the process of data acquisition, including the data error of aerial photograph or satellite photograph, the quality of topographic map itself, and the measurement error in the process of data acquisition. Error index data processing error generated in the process of data processing, including computer data processing mathematical model and calculation accuracy, data format conversion. The application error refers to the error in the process of using the spatial data, including the completeness of the data and the correctness of the topological relation [4].

2.2 Uncertainties in Spatial Data

Spatial data quality is the ability of data to be used in different applications. Generally speaking, it includes the following aspects: accuracy of position, attribute accuracy, logical consistency and completeness.

3. Sources and Effects of Errors and Uncertainties

3.1 Sources and Effects of Errors

3.1.1 Original Error

Data errors in aerial and satellite photographs. Aerial survey and remote sensing are one of the important methods for basic geospatial data collection. The observation range is large, it can provide instant static images, and it can perform large-area repetitive observations. The quality problems of aerial and satellite films mainly come from the errors and quality problems introduced by remote sensing observation, remote sensing image processing and interpretation process, the influence of geometric time-varying on data quality, image and image correction matching, interpretation and classification.

Map digitization error. The main factors that affect the digital precision are the inherent error of the original map, the distortion error of the drawing, the density, width and complexity of the map elements, the influence of the digitizer or scanner instrument error operator and its operation mode [5].

Direct measurement of data error. Measurement errors usually include systematic errors, operating errors and accidental errors. Measurement data errors include errors in human-operated instruments, inadequate instruments, lack of calibration, uncorrected, environmental climate, and signal interference.

Error of attribute data. Attribute data is important data, and its errors are mainly caused by data entry and database operations. As a result, the flexibility of attribute combination and the complexity of attribute data make the operator make mistakes.

Error propagation problem. Errors in the data are propagated in a certain way and affect reliability and the quality of the application results. There are many ways of error propagation, including not only the error propagation under arithmetic relation, but also the error propagation under logic relation and the error propagation under imprecise inference relation.

3.1.2 Processing Error

The error caused by the coordinate transformation. The coordinate transformation includes the changes of the coordinates of the scanned image and the geodetic coordinates, as well as the changes of the coordinates of two different geodetic coordinate systems. The errors include the computer word length representation accuracy and rounding errors.

Errors due to projection transformation. The projection transformation from ellipsoid to plane is carried out on the geographical coordinates of the spatial entity. Under different projection forms, the geographical position, area and direction of the entity will be different.

Edit error of geometric data. The error produced by the editing of geometric data are mainly related to the logical consistency of geometric data, the rationality of the entity's mutual relationship, the errors generated by the editing of the data, the errors caused by the collation of the data when the data is scaled, the superposition operation and the data Errors generated during the update, and errors generated during the interoperation of various types of data sets of different sources and types.

Error caused by mutual conversion of vector and raster data. In the transformation between the vector and the grid, serious errors may occur.

3.1.3 Application Error

For the same spatial data, different users may have different interpretations and understandings of APP data content, which is the concept of application error.

3.2 Sources and Effects of Uncertainties

3.2.1 Accuracy of Position

The accuracy of the data location will directly affect the application of the data. It is very important to strictly check and analyze the position accuracy of the data in the spatial data quality control inspection [6]. Positioning accuracy includes deviation and accuracy. Deviation is the systematic difference between the real position and the expressed position. The ideal deviation should be zero, indicating that there is no systematic deviation between the position on the graph and the expressed position. The deviation is usually measured by the mean position error of the sample points. The accuracy is the description of the position error of data elements. The accuracy is usually estimated by calculating the standard dispersion difference of the selected detection points.

3.2.2 Attribute Accuracy

Attribute data includes geometric type, classification level, data feature, quality description, geographical name, update date of spatial entity data. The quality characteristics of the system's attribute data include the following aspects. The definition of attribute items that describe spatial data, including name, type and length, must be correct. The value of attribute and its unit of each data item in the attribute sheet must not be abnormal. Flag code is the code to distinguish the identity of spatial data, must be unique and effective, not repeated. The various attribute data that describe the characteristics of each geographical entity should be correct. Spatial data and attribute data must be correctly correlated.

3.2.3 Logical Consistency

The logical consistency of data is the reliability of the expression of the logical relationship between the description features and the maintenance of good logical relationship between the data elements [7]. The following aspects are required to maintain the logical consistency of spatial data. The definition of point, line and surface types of spatial entities must be correct and should be stored in different types. Ensure the closure of polygonal space entities. Ensure the connectivity of linear spatial entities. To ensure compliance with the relationship principles or constraints between the combined entities and the underlying graphical elements. To ensure that there are no hanging nodes and pseudo- nodes, all the line elements intersection, should establish a node library.

3.2.4 Completeness

Completeness refers to the same accuracy and precision of the data within the scope of the specific space is complete, it includes the classification of the integrity of the data layer, data completeness and completeness of the entity type, attribute data completeness and the integrity of the note.

The above original errors and uncertainties can cause changes in the location of spatial data and errors in attributes, resulting in a series of problems such as inaccurate positioning of Google Map, and bringing a bad application experience to users.

4. Methods for Controlling and Measuring Errors and Uncertainties

In the process of spatial quality data formation, although there will be a lot of errors, some errors can be ignored. For example, the error magnitude requirement for digital map projection is a few thousandths; The error magnitude of the null decision analysis requires a few tenths. The systematic error caused by measurement can be simulated and estimated by mathematical model, the consistency of operating error can be verified by simple geometric relation or algebraic examination, or the operating error can be eliminated by repeated observation, and the random error can be estimated and processed by using random function. The error generated in the propagation process can be controlled and quantified by double calculation of checksum [8].

The measurement of position accuracy usually uses the measured root mean square error, and the estimated root mean square error of the data is very valuable for effective use of the data. A small discrete difference indicates that the value range of the position error is relatively small. The higher the accuracy of the measured data, the higher the reliability of the measurement. The coordinate values of the outline points, the kilometer grid, the cross points of the latitude and longitude network, and the control points should be accurate, and all measurement errors should meet the requirements of the corresponding scale mapping specifications. Due to the error of manual operation, the correctness of the logical relationship between spatial objects will inevitably be destroyed. Therefore, before the data is stored in the database, the topological check of the spatial data must be performed according to a certain relationship between the spatial objects.

5. Future Research Trends on Quality Control for Spatial Data

Quality inspection based on spatial entities and topological planes is an important aspect of quality control of spatial data, so it is necessary to further study the diverse types of spatial data in spatial entities. Currently, spatial data products are mainly represented as "4D" products, and corresponding data quality models should be established respectively. Based on the characteristics of spatial data and the requirements of data users, a unified spatial data quality model is established to provide a basis for the formulation of quality standards.

References

[1].Lei Wang. (2015), "The Application of Google Earth in Railway Engineering Geological Mapping.", Resources, Environment and Engineering, 04, 476-479.

[2].Juqing Zhang. (2003), "Representation and Propagation of Attribute Data Uncertainty.", Gold Geology, 1, 56-60.

[3].Chun Liu. (2001), "Research Status of Spatial Data Uncertainty and Quality Control.", Surveying and Mapping Engineering, 10, 6-10.

[4].Yongling Weng. (2001). "Error Source Analysis and Quality Control of Spatial Data.", Railway Aerial Photogrammetry, 3, 10-12.
[5].Xiaohong. (2001). "Error Analysis of Map Scanning Digitization.", Science of Surveying and Mapping, 26, 49-52.
[6].Yanwei Zheng. (2004). "Spatial Data Quality Control and Evaluation Methods and Implementation Techniques.", Journal of Wuhan University Information Science Edition, 29, 686-690.
[7].Tinghua Ai. (2000). "Correction of the Uniformity of Shared Boundaries of Adjacent Polygons.", Journal of Wuhan University of Surveying and Mapping Technology, 25, 426-431.
[8].Xiaolie Gu. (2006). "Third Edition of Surveying.", Shanghai Tongji University Press, 226-243.