地理学报(英文版) 2001年第11卷第5期

Adding the value of NAVTECH road database: an implementation of spatial data mining techniques

作者: Lichun Sui Liqiu Meng

This paper reports a spatial data mining prototype system developed at the Technical University of Munich in cooperation with NavTech. The system serves the purpose of value-adding the road database maintained by NavTech. In the original database, each road element is described by more than 110 attributes. A number of algorithms on the basis of entropy theory, rough-set modeling have been implemented to rank the individual attributes and detect the dependencies among attributes based on their values in an arbitrarily selected region. Other algorithms are developed on the basis of road geometry and devoted to the quantitative description of spatial patterns such as routes and urban structures. With the knowledge of relative importance of the individual attributes, users are given the flexibility to buy a local road database with truncated attribute list. By observing the ranking list and correlation matrix calculated for different regions, information that reflects the regional differences of a road network can be extracted. Likewise, the changes in ranking list and correlation matrix of the same region after removing or adding a route imply the relative importance of this particular route.

Adding the value of NAVTECH road database: an implementation of spatial data mining techniques Lichun Sui, Liqiu Meng (Inst. of Photogrammetry and Cartography, Technical University of Munich, München, Germany) Abstract: This paper reports a spatial data mining prototype system developed at the Technical University of Munich in cooperation with NavTech. The system serves the purpose of value-adding the road database maintained by NavTech. In the original database, each road element is described by more than 110 attributes. A number of algorithms on the basis of entropy theory, rough-set modeling have been implemented to rank the individual attributes and detect the dependencies among attributes based on their values in an arbitrarily selected region. Other algorithms are developed on the basis of road geometry and devoted to the quantitative description of spatial patterns such as routes and urban structures. With the knowledge of relative importance of the individual attributes, users are given the flexibility to buy a local road database with truncated attribute list. By observing the ranking list and correlation matrix calculated for different regions, information that reflects the regional differences of a road network can be extracted. Likewise, the changes in ranking list and correlation matrix of the same region after removing or adding a route imply the relative importance of this particular route. Key words: data mining; information gain; rough set; semantic analysis CLC number: P208 1 Introduction Since years, geo-data suppliers have been concentrating themselves on the tasks of constructing portal sites to attract buyers, filling holes and removing redundancy in their data warehouses, updating and versioning the data items, developing compression methods as well as data structures for efficient transmission and so on. However, the availability of a seamless digital earth is not our ultimate goal, rather the starting point of personal geo-services. The effectiveness of geo-services is strongly influenced by the accessibility and transparency of the available spatial data warehouses on the Internet. The accessibility requires that (1) the database as a whole be well-tagged with a summary containing the relevant key words; and (2) the individual data items be explicitly indexed with attributes and metadata. The transparency requires further methods to (1) discover the spatial concepts that are otherwise hidden in the database; and (2) describe the discovered concepts using an easily understandable language. An accessible and transparent database allows flexible aggregation and segregation, hence the personal division of the information space. However, personalizing large data inventories is complex and unintuitive. Spatial data suppliers would go insane trying to determine what to offer to whom, especially when they themselves have lost an overview of their ow

n databases (Meng, 2001). Therefore, such tasks should be performed by automatic spatial data mining systems. 2 Navig ation and semantic analysis of road objects For a vehicle driver, a driver assistant means a navigation system. It co nsists of the digital map, GPS and other additional components. If the driver wants to drive from a place to another place in a city. He needs at first to determine the general orientation to the destination. This means a coarse navig ation for which a map showing a coarse road structure may be sufficient. When the driver approaches his destination, however, he must have the detailed street connections nearby, which triggers a refined navigation. This example revea ls the necessity of value-adding a road database so as to provide adaptive road information required by users in diff erent navigation stages. In the road database delivered by NavTech, each individual road segment is recorded as an ob ject attached with up to 110 attributes describing its qualities and quantities. The semantic analysis of the attribu te values covering a selected region can reveal, e.g. (1) the relative importance of each attribute, (2) the correlat ion between two attributes and (3) the structural character of the local road network. 2.1 Semantic analysis based o n information gains 2.2 Semantic analysis based on rough set modeling 3 Future plans Road network can be mined not on ly by processing the semantic attributes, but also through the analysis of the spatial relationships within it. Howev er, there are still very few spatial data mining methods that can process the geometric and semantic attributes in a well-integrated manner. Though the separate treatment of geometry and semantic data of spatial objects might be suffi cient for many GIS applications, it is not able to create a reasonable description of spatial patterns that occur as outcome of the interplay between geometric and semantic attributes. Keeping this challenge in mind, the authors will make their further experiments concerning the integration of non-spatial data mining methods into a spatial data mini ng procedure. References

关键词： data mining; information gain; rough set; semantic analysis

