Design and Implementation of Parallel Algorithm Used in Validating Topology Based on MPI

Ren Yibin, Chen Zhenjie, Li Feixue

School of Geographic and Oceanographic Sciences, Nanjing University, 210046

Telephone: 0-13182997228

Email: chenzj@nju.edu.cn

1. Introduction

With the rapid development of GIS (Geographic Information System) and spatial information obtaining technology, massive geographic data have become the main target being processed and analyzed by GIS. However, existing technologies of GIS and traditional hardware and software platforms have been unable to meet the needs of massive geocomputation. In this circumstance, using parallel computing ideas and methods to improve efficiency of massive geocomputation becomes a trend.

Topology validation plays an important role in geocomputation. For vector data, topology validation is an effective approach to ensure that data is normative and correct. But there are two problems in topology validation for massive geographic data. On one hand, not all spatial data formats have the ability to store topological information. For example, ESRI shapefile (*.shp) and AutoCAD (*.dwg,*.dxf) can't store topological information. So we can't validate these data's topology directly. We have to convert this data format to another one which can store topological information. Then we can validate topology based on topological rules. But it generates a problem. For massive geographic data, the conversion between data formats will spend a long time. What's more, it can lead to the collapse of the commonly used GIS processing software platform. On the other hand, data set whose data format has the ability to store topological information also stores lots of topological mistakes, such as: polygon overlap, line self-intersection, etc. So, validating topology based on topological information is not a good way. And for massive geographic data, it also takes a lot of time and resources.

In response to these problems, this research aims to validate topology and find topological mistakes in massive geographic data accurately and fast, regardless of whether the data stores topological information or not. This is helpful to correct topological mistakes and generate high quality data sets.

Foundation item: National High Technology Research and Development Program of China (Granted NO.2011AA120301)

Biographies: Ren Yibin (1990-),male, master student; Chen Zhenjie (corresponding author), male, doctor, associate professor, chenzj@nju.edu.cn

Address: School of Geographic and Oceanographic Sciences, Nanjing University, No.168 Xianlin Avenue, Nanjing, Jiangsu, P.R.China

2. Methods and Conclusions

Parallel computing is applied in this research. Parallel computing can improve the speed of massive geocomputation effectively. A parallel algorithm used in validating topology based on MPI is put forward. Main contents are as follows.

Firstly, parallel algorithm design. In order to improve the parallel algorithm's efficiency, the algorithm establishes the spatial index based on feature's MBR (Minimum Bounding Rectangle). Taking polygons overlap judgments for example. All polygons are sorted by the minimum X coordinate of their MBR. In this sequence, each polygon just needs to compare with the one behind it and need not to compare with the one before it. So, this spatial index can avoid repeated comparison between polygons and reduce the number of judgments effectively. Furthermore, the algorithm filters polygons' MBR with MBR's four corners' coordinates. If two polygons' MBR overlap with each other, the algorithm will make a further judgment about two polygons using polygon overlap algorithm. On the contrary, if two polygons' MBR are betwixt, there is no need for further judgment. In this way, the algorithm minimizes the number of judging whether two polygons overlap with each other. Data is partitioned by row in this algorithm. The full extent of the map is divided in several rows based on the number of processes in the Y-axis direction. Each process reads data from one of the rows and processes the data. This strategy filters the data set in Y-axis direction. Adjacent polygons in Y-axis direction are assigned to the same process. It reduces comparison times in Y-axis direction. Combining this data partition strategy with the spatial index can improve the efficiency of the algorithm obviously. This parallel algorithm use master-slave parallel programming mode. The system generates a management process and calculation processes. The management process collects information from calculation processes. Then the management process eliminates data redundancy and sorts result data by feature's ID. Calculation processes find the topological mistakes and send result data to management process.

Secondly, parallel algorithm implementation. This research selects "Must Not Overlap" as validation rule, selects C/C++ and MPI as parallel programming environment. MPI is the most important parallel programming tool. It has a variety of advantages like portability, powerful and efficient, etc. And it is the standard of message passing parallel programming mode. GDAL library is used for reading vector data sets.

Lastly, parallel algorithm experiments. This research chooses a data set which includes 1,778,093 polygons in shapefile format as test data. The amount of test data is 2GB. This parallel algorithm can check out the overlap between polygons in test data set accurately and quickly. This proves that using this parallel algorithm to validate topology is accurate and effective.

Keywords: parallel algorithm; topology validation; massive geographic data; polygons overlapping

3. Reference

Vokorokos, L. Parallel computer system utilization in geographic information systems. ICCC 2005: *IEEE 3rd International Conference on Computational Cybernetics*，2005: 333-338.

Vermeij M J . Development of a Topological DataSt ructure for On2the2Fly Map Generalisation [D] .*Netherlands : Delft University of Technology* , 2003

Wang Jiechen,Wang Bao, Hu Wei, Zhang Hui. Review on Parallel Spatial Analysis Algorithms[J]. *Geography and Geo-information Science*，2011，27: 1-5.

Lu Hao，Zhong Ershun，Wang Tianbao, Wang, Shaohua. A Polygon Data Automatic Generation Algorithm Based on Topology Information[J]. *Geography and Geo-information Science*，2012，28: 38.

XIE Zhong ,YE Zi,WU Liang. Polygon Overlay Analysis Algorithm Using the Simple Data Model[J]. *Geography and Geo-information Science*，2007(3): 19-23.

ZHAO, G, L. Xu and Q. SUN. Study and application of oriented-topology geospatial data model [J].*Science of Surveying and Mapping*, 2011. 36: p. 43-45.

ZHAO Chun-yu. Studying on the Technolgies of Storage and Processing of Spatial Vector Data in High-Performance Parallel GIS[D], 2006, *Wu Han University*.

Duan Anli. Parallel Processing Technology Research Spatial Topological Analysis Operation[D]. 2008. *Nanjing University of Aeronautics and Astronautics*.

C H EN G Chang xiu . A Multi2scale Spatial Index Method [J].*Geomatics and Information Science of Wuhan University*. 2009(5): 597-601.