Extraction and construction algorithm of traffic road network model based on OSM file

Hangong Wang¹, Jiakang Wang¹, Peiyao Yu¹, Xudong Chen¹ and Ziyang Wang²

¹ School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China
² School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Email:18301020@bjtu.edu.cn

Abstract. In order to solve the problems that are not conducive to the automatic analysis and processing of related applications in the initial data of the road network, the main data is extracted from it to construct a traffic road network model. We designed an algorithm for extracting and constructing a traffic network model based on OSM files. The algorithm is mainly divided into preprocessing of OSM geographic data, road network extraction, road recognition, intersection recognition and merging, providing data basis for traffic simulation needs.

1. Introduction
At present, the common method of generating road network is to reprocess the Internet map or the drawings of the surveying and mapping department as the base map. Internet maps are born for navigation. The road network data model is organized according to the business logic of navigation. Connectivity is more concerned, and professional information such as the classification of roads is less concerned. To solve the above problems, the algorithm first processes and analyzes OSM geographic data files, then constructs a traffic network model according to the road level, and finally generates the intersection topology structure, road coordinates and other data that can be automatically analyzed and processed by the program.

2. Functional design
2.1. OSM geographic data file reading and crossing identification
The pre-processing of the first stage of this algorithm includes extracting various information needed to build a traffic network model from the original geographic information data, eliminating redundant information, and secondly dividing the extracted data according to road types and according to certain Regularly identify crossings between roads.
2.2. Crossings merge

Large-scale crossings of urban road networks are often interlaced with multiple roads, and this kind of crossing will lead to multiple crossings and small short roads at the crossing. Too many instances will adversely affect the program, and the computer cannot fully utilize the advantages of computing power and memory. Therefore, it is necessary to perform a merge operation on complex crossings. This algorithm identifies complex crossings, abstracts its topological model, and replaces it with a single large crossing. The replaced crossing is equivalent to a black box, our program only needs to care about input and output, and no longer judge its internal action logic.
2.3. side road identification
In the real road network, the side road is often a whole with the main road. However, in the data obtained, the main and side roads are respectively represented by two polylines, which increases the burden of the program and reduces the efficiency. Here through the side road identification, the side road will be deleted. In order to ensure the reliability and correctness of the data, the number of lanes of the corresponding main road is increased, and the side road is taken as one or some lanes of the main road, which is also in line with the actual road situation. If the side road carries other functions at the same time, the side road will be reserved to ensure the correctness of the program.

3. Technical realization

3.1. Technical realization of OSM geographic data file reading and crossing identification
The pre-treatment of the first stage of the algorithm includes extracting all kinds of information needed to construct the traffic network model from the original geographic information data, eliminating the redundant information, and then classifying the extracted data according to the road types and initially identifying the crossing between roads according to certain rules.

3.1.1. OSM geographic data file reading
The original geographical information data (as shown in Figure 1) usually contains various information, and this algorithm extracts only six basic Road data according to Road types: Arterial Road, Collector Road, Interstate Highway system, Major/US Highway, Ramp and State Highway. Railroads, rivers and other data with little reference significance were excluded. According to the actual situation of the real road network, the extracted 6 kinds of roads are classified into three levels. The first level is Interstate, Major/US Highway, which is generally only connected with RAMP. The second level is the Ramp, and the two ends of the road must be intersections, which are generally used to connect the first and third level roads. The third is Arterial Road, Collector Road and State Highway, which can be connected to each other.

3.1.2. crossing identification
The road pre-processed in the first stage is very primitive, in which each road is a list composed of a number of latitude and longitude points. The two ends of the list do not correspond exactly to the crossing of the real road, but may be just a complete road part. Therefore, the algorithm splices roads according to their types and directions to ensure the integrity of each road. In addition, the algorithm will also cut according to the crossing points between different types of roads to avoid the problem of
excessive road splicing. After the above processing each road corresponding to the two ends of the list is initially identified to the crossing. R tree spatial index is also introduced in the algorithm to search for roads and intersections.

3.2. Merging of complex intersections
After obtaining the directed graph of the road network through the relationship between roads and crossings, we use the knowledge of graph theory to color the crossings. The implementation method is given below:

Step 1: Determine a constant distance \( d \).
Step 2: Traverse all the nodes, and apply BFS algorithm on the nodes.
Step 3: Calculate the distance \( D(n1, n2) \), compare \( D \) with \( d \) to determine whether the two points belong to the same crossing.
Step 4: If \( D < d \), color them with the same color (serial number).
Step 5: Mark the short roads between the two nodes as a state to be deleted.

After a round of BFS algorithm on a node, the nodes that belong to the same complex crossing with it will be marked with the same color. After BFS traverses all points, the short roads are deleted, and only one crossing of the same color remains. Finally, extend the opened road to the reserved crossing.

Figure 5. Merge algorithm diagram.
3.3. side road identification implementation

To accurately identify the side road, side road and the corresponding main road has the following properties: Side road and main road connected to the same two crossings; The side road and the main road are at the same level height, that is, there is no height difference between the side road and the main road. By comprehensive analysis of the data, it is found that the side road and the “Collector Road” completely overlap in reality, but the main Road may be “Major Highway”, “State Highway”, “Interstate”, and “Arterial Road”. Considering that the main and side roads should be at the same height level and the elevated highways such as “State Highway” and “Interstate” are excluded. The implementation method of side road identification is given below:

Step 1: Get one road
Step 2: The roads that meet the conditions are connected to the same two crossings:
① If one path is Collector Road type, the other path is State Highway type or Arterial Road type.
② Delete side road.
③ Increase the number of lanes of the corresponding main road.
Step 3: If all the roads have been traversed, go to the end; otherwise, get the next road and go back to step 2.

4. Conclusion

Through the construction of traffic network model, the algorithm generates the road topology structure that can be analyzed automatically by the program, thus providing effective data support for traffic simulation, path planning and other requirements.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 71631002)

References

[1] Simplified model of road network based on hierarchical classification and its application [J]. Cao Jiandong, Zhang Aomuhan, Li Bing. Ordnance Industry Automation. 2015(03)
[2] Cui Xia. Research on Topology Reconstruction of Community Road Network Based on GPS Data of Slow Traffic [D]. North China University of Technology, 2020.

[3] Zhou Shuigeng, Zhou Aoying, Cao Jing. DBSCAN algorithm based on data partition [J]. Computer Research and Development, 2000(10):1153-1159.

[4] Deng Hongyan, Wu Fang, Zhai Renjian, Liu Weiwei. A comprehensive model of road network based on genetic algorithm [J]. Journal of Wuhan University (Information Science Edition), 2006(02):164-167.