Research on Spatial Large Data Mining Technology Based on Network Optimization

Liu Yuxuan, Yan Guanghui, Ye Jianyun, Li Zongren

Lanzhou Jiaotong University, Lanzhou, China

Keywords: spatial data, data mining, cloud computing

Abstract: In the process of long-term network renewal, network construction has become more and more mature, and people rely more and more on network maintenance, rather than network construction. Increasing people are increasingly demanding mobile networks, which requires to take necessary optimization measures to improve the quality of network and the overall level of business. In recent years, many cloud computing platforms make full use of the advantages of distributed and cluster advantages to achieve good mining results by implementing parallel algorithms. In this paper, the current large spatial data mining technology is sorted out, and the direction of technology development is systematically summarized.

1. Introduction

The purpose of mobile network optimization is to optimize the network as soon as possible when the network quality is declining or the key traffic indicators can not meet the requirements. Many uncertainties will have a significant impact on the wireless network, and operators may introduce new services or change the types of services. These factors will cause changes in traffic. Therefore, although in the network construction period, various factors affecting network performance will be fully considered, after all, planning can not solve all problems. Network optimization is a process requiring long-term research, in which more new technologies and methods need to be innovated to adapt to the rapid development. In view of the rapid change of data service rate, huge throughput and dynamic real-time change of coverage and so on, it is necessary to introduce large data mining technology in the field of network optimization.

2. Overview of Spatial Large Data Mining Platform

2.1 Spatial Hadoop

Spatial Hadoop, proposed by the University of Minnesota in combination with the technical characteristics of cloud computing platform, is based on Hadoop and extends it accordingly, which is a technical platform specially used for spatial data mining. Because it is based on Hadoop, its framework is essentially an extension of MapReduce. The platform applies pigeon language. This original language reduces the difficulty of programming and processing problems. Spatial index is a prominent feature of the platform and its structure is double-level. The index system is completed by five steps that are partition, calculation of partition boundaries, physical partition, single physical partition and all partitions. Meanwhile, the platform provides researchers with a variety of native algorithms, which basically meets all kinds of spatial data mining and analysis. These native algorithms can provide machine learning for model training, and parallel computing can be carried out through distributed framework in specific research work, so as to obtain effective research results. The emergence of Spatial Hadoop platform means the great progress of independent large-scale spatial data mining technology. Although there is no essential difference between the core idea and architecture of the platform with the Hadoop platform used by researchers, the original programming language, parallel algorithms with machine learning characteristics and perfect storage index system all bring great convenience to spatial data mining.

  DOI: 10.25236/iwmecs.2019.057

2.2 Integration and Development of Spark, Hadoop Platform and ArcGIS

Geographic information system researchers still rely on ArcGIS platform when facing a large number of vector data, raster data and elevation data. In order to meet the needs of large spatial data research, the integration of Spark platform and ArcGIS has become an important research direction. The most representative one is GIS tools forHadoop. The core idea of the fusion is to make full use of the advantages of cluster and realize parallel operation on the basis of serial mining algorithm, so as to improve the effect. Some research results have proved that the integration of Spark, Hadoop and ArcGIS can realize infrastructure resource optimization, source data processing, analysis and calculation, and a large number of spatial data storage. Especially by using RDD provided by Spark, data can be elastically distributed to ensure data integrity in the specific analysis process. Using K-means, Dbscan and other algorithms on the above two platforms will also effectively remove noise points.

2.3 Application of MATLAB in Spatial Data Mining

In the application of Matlab platform in large spatial data mining, parallel algorithm is realized by parallel toolbox. Meanwhile, compared with Spark and Hadoop, it is more convenient to design parallel algorithm by Matlab. The FOR loop can be directly replaced by Porfor in specific programming. The greatest advantage of Matlab is that it can directly obtain visual results and display the research results directly with visual results without the aid of map tools. Spark and Hadoop platforms, as general platforms for large data mining and analysis applications, can not directly provide visualization results, which is one of the reasons that restrict their applications.

3. Mobile Network Optimization and Data Mining

3.1 Mobile Communication Network Optimization

Network optimization is divided into three basic modules: wireless network optimization module, core network optimization module and transmission network optimization module. Because there are not many network elements in the actual transmission network and core network, and its operating environment is relatively sTable, the focus of mobile communication network optimization is wireless network optimization.

Network optimization is mainly to collect and analyze the data of existing GSM network, and adjust its parameters, antenna, network configuration and network routing by means of technology and engineering. By adjusting, network optimization can be improved, play its greatest role, bring more benefits to enterprises, and provide users with better service. Generally, network optimization is operated as a practical project, which is usually implemented in three stages.

3.1.1 Data collection and familiarity with the network stage

The main work of this stage is to find out where the needs are, communicate with customers, understand the needs of customers and find a starting point for the whole optimization process. There is also the need to import the data into the application tools, including the base station information Table, which mainly conclude the latitude and longitude and inclination of the main information. After collecting these information, it is imported into the analysis software to prepare for the next analysis and processing. The main goal of the first stage is to collect and complete basic data, complete the groping road survey and the audit of parameters.

3.1.2 Network adjustment and implementation optimization stage

In the stage of adjusting mobile network and executing optimization, the main problems of dropping calls, signaling and user complaints are solved. There are some problems that can not be solved by network optimization. Detailed design and reason analysis are made to improve the KPI index of network statistics. In this stage, new functions of equipment need to be enabled to solve the above problems. According to the actual situation of the network, in the second stage of the optimization process, the topic is optimized to form a thematic optimization report, leaving

materials for future projects, which can be consulted.

3.1.3 Maintain performance and summarize optimization phase

In the above two stages, network performance will be significantly improved, and this improvement is limited. The work of adjusting the network and executing the optimization stage has made the network performance reach the best state. Maintaining the performance and summarizing optimization is to keep this state all the time. At the same time, it also needs to enter the summarizing stage of the whole project execution. The third stage is to summarize and summarize the first two stages, so as to keep the network optimization performance good.

3.2 Data Mining Technology

The so-called data mining technology is based on the needs of individuals and enterprises, extracted and analyzed from large databases, so as to obtain valuable and usable information of data analysis technology, which can guide business behavior or provide reference value for scientific research based on the data information obtained. Due to the accelerated process of informationization, information is in the era of big explosion. A large amount of information needs to be understood but information is too messy to accurately extract useful information, so data mining technology is produced. It is a new field in database research and can be combined with various disciplines, so it is also applicable to various fields.

3.3 Common Clustering Algorithms in Data Mining

3.3.1 Clustering algorithm

Clustering analysis is a process of assembling some actual or Abstract objects together and judging whether they can be classified into a group according to their similarities. This is an important human analysis behavior. This system uses the shortest distance clustering method to judge whether they can be classified according to the distance of individual data points in geometric space.

3.3.2 Complex mining of network community

The concept of community structure was first put forward by Newman in 2002. The so-called “community“ is commonly referred to as “clustering“. A community can be roughly described as a subgraph consisting of some vertices. Within the subgraph described, the connections between vertices are very close, but they are connected with vertices other than this subgraph. Because many network structures present such community structure, it is of great practical significance to detect and describe this community structure.

3.3.3 Improved CNM algorithm

In order to accurately classify the network structure and some important weighted networks, a hierarchical data clustering method called improved Newman greedy algorithm (CNM algorithm) should be applied. In this CNM algorithm, the concepts of point weight and edge weight are introduced, and the community modularity of the new Q function is redefined. The final result of community division is determined by the number peak of Q function. After using the improved algorithm, it is found that this algorithm has obvious advantages in the accuracy of partition, the complexity of algorithm and so on. In view of the existing understanding of complex network mining, the improved CNM algorithm has great advantages in clustering effect and time complexity compared with other similar algorithms.

3.4 Application of Data Mining in Mobile Network Optimization

The main role of data mining technology in network optimization system is reflected in two aspects: prediction analysis and statistical analysis. Prediction analysis includes traffic equilibrium analysis while statistical analysis includes interference analysis, handover analysis, coverage analysis, access analysis, drop-call analysis and so on. According to the topological structure of GSM network, the work of GSM network optimization can be divided into two aspects: system

optimization and wireless optimization. Wireless optimization is mainly aimed at parameter setting of base station, base station sector and so on, while system optimization is aimed at MSC and above.

4. Analysis of the Development Direction of Spatial Large Data Mining Technology

There are two main directions for the future development of large spatial data mining technology:

Firstly, with the help of more and more mature cloud computing platform, the integration of ArcGIS or the application of a separate platform can be realized. Further utilizing the advantages of distributed framework and cluster will improve the efficiency of mining and reduce the time spent on specific research. Fully considering on how to visualize is also one of the research directions of platform selection and platform optimization. The main research direction is to use large data analysis platform to provide feasible solutions for spatial data mining.

Secondly, it is possible to realize parallel operation of specific algorithms through distributed framework, so as to achieve technical optimization with the help of related platforms. The algorithm model training is implemented to avoid problems,so that all kinds of large spatial data can be oriented to meet the application requirements of all spatial data analysis. The optimization direction of algorithm tends to be more in-depth, and the scope of research will gradually expand. Machine learning, in-depth learning, including the combination of some previously unused algorithms, are the key research areas in the future.

5. Conclusion

Spatial data with huge data scale can no longer be studied with the original mining technology and mining platform. It must be integrated with cloud computing platform to optimize and combine with parallel algorithm to achieve technological innovation. This research summarizes and collates the existing research results, and specifically analyses the research direction of platform selection and algorithm optimization. To improve the effect of large spatial data mining, it is necessary to achieve good results through double optimization. The era of large spatial data has already arrived. To ensure the validity of research, it is essential to explore three aspects together: research ideas, cloud computing platform and parallel algorithm.

References

[1] Guan Xuefeng, Zeng Yumei. Progress and Trend of Parallel Data Processing, Analysis and Mining under the Background of Large Spatiotemporal Data [J]. Progress in Geographic Science, 2018,37(10): 1314-1327.

[2] Liu Dongbiao. Parallelization of Spatial Data Mining Clustering Algorithms Based on SPARK [D]. Jiangxi University of Technology, 2018.

[3] Pan Tingting. Research on Key Technologies of Knowledge Discovery and Information Extraction for Large Geospatial Data [D]. China University of Geosciences (Beijing), 2018.

[4] Xu Yanqiang. Application of Data Mining Based on Large Database [J]. China New Communications, 2018, 20 (08): 121.

[5] Bai Jinsong, Yang Guangting. Research on Spatial Data Mining Technology in Network Optimization [J]. Computer and Network, 2017, 43(22): 68-70.

[6] Xue Anrong. Research on Spatial Outlier Mining Yechnology [D]. Jiangsu University, 2008.