Study of Traffic Information Analysis and Decision Support System Based on Grid Computing

Liu Chong, and Lu Huapu

Institute of Transportation Engineering, Tsinghua University, Beijing, China 100084

Liudeer98@mails.tsinghua.edu.cn
Luhp@tsinghua.edu.cn

Abstract

Intelligent transportation system is becoming one of the important means to solve the traffic problem, especially the traffic management problem. The information analysis and decision support are the key points of intelligent transportation system. Because of the deficiencies of the information analysis and decision support system at present, a system model of information analysis and decision support is proposed based on analyzing the concept of grid computing and data mining. It is easier to identify the useful information in the data and provide valuable advice with this system, and this system makes the information analysis and decision support intelligent by using the grid computing and data mining technologies. Finally, the function part of the system is expounded in detail and operation course of the system is analyzed.

Keyword: Traffic information analysis, Decision support, Grid computing

I. Introduction

With the intelligent transportation system's application in many cities in China, the traffic management departments have set up some information management systems to meet their own need, such as: the traffic flow detecting system, closed circuit TV (CCTV) system, traffic signal control system, etc. These systems provide abundant traffic data and information resources which are very valuable for intelligent transportation system application systems. In order to optimize the strategy of traffic management and guide the behavior of the traffic participant so as to make the traffic system highly efficient, some decisions must be made based on the state of traffic system which can be obtained through the analysis of traffic information and on the basis of the traffic engineering knowledge. So the information analysis and decision support system are becoming a kind of important subject in the transportation field study day by day.

Challenges the traffic information analysis faces

As valuable resource of the intelligent transportation system, traffic information has several characteristics as follows [1].

(1)The source is very wide, the style is various, and the amount is very large. The traditional analytical methods based on statistics do not have the ability to deal with this information as it is not number type or structured data. At the same time, the traditional decision support system, based on relational database system, is difficult to support the analysis of the huge amount of information.

(2)The information has very strong space-time dependence. Traffic information is always produced in a certain time and at certain place, if the information is not associated with the time and space, it will no longer be easily understood and utilized, and the traffic information is not real information, just data.

(3)The information has obvious subject dependence. Traffic information can be divided into several parts according to the subject such as: traffic flow information, the signal control information, traffic accident information, vehicle information, driver information, transit dispatcher information, traffic infrastructure information, etc.

(4)The traffic information is always real-time and dynamic. As the traffic system is real-time and dynamic, therefore, all decisions made by the traffic management department must be based on the real-time information. If the time limit is exceeded, the decision made on it will not be reasonable.

As a result in future, we must face huger and huger traffic information, and the key is how to get the most important information, how to find out the representative's essential rules from the huge information seeming disorderly and unsystematic, and how to put forward the scheme consulted within the shortest time. The classical mathematics method and traditional modeling method can't meet the requirement of quality and time of the information analysis, so the new generation information analysis and decision support techniques are needed to solve those problems, represented by the data mining and artificial intelligence. Meanwhile, the research results in related fields such as traffic engineering, system engineering, and computer science must be applied in the information analysis and decision support system, then we can make a comprehensive, intelligent and fast analysis, and make full use of the information resources of existing systems. Based on the system, we can offer the accurate analysis result in a limited time and give a scientific decision for the traffic management department.

Limitation of decision support system (DSS)

At present, in most traffic management DSS, the complicated problems are always tried to be made simple, and the key idea is mainly as follows: the traffic system is supposed to accord with a certain physics system model (its inputs accord with a certain probability distribution), then such methods as classical mathematics, statistics, calculus, probability theory and fuzzy mathematics are used to solve the problem and seemingly optimum results are provided to the system users. In fact, the received result like this can't reflect the inherent characteristics of the traffic system in most time, because the traffic system will not run according to the design and supposition of the traffic engineer. So the objectivity and usability of the result must take a great discount to some extent. With the development of such artificial intelligence technologies as machine studying, fuzzy reasoning, etc., the probability to adopt these techniques to solve the qualitative, approximate or inaccurate traffic management problem is becoming higher and higher in the DSS. Meanwhile, the application of the artificial intelligence technique will also offer such functions as reasoning, machine studying, etc. which are more like the way human decide.

II. Idea of the System

Because of the understanding in the traffic information analysis and DSS now, the main idea of how to develop the information analysis and DSS is:

1, to integrate the traffic engineering knowledge and the information system technique

How to obtain (gather) the primitive information, with which method and technique to deal with the primitive information, what information to provide to the certain users, in what scope (both space and time) to distribute the information, in what frequency the information should be upgraded, how the users' response will be after the information is put about, and according to the response how to adjust the information content, upgrade frequency and the scope put about. All these things can not be solved only by the information systems [2].

2, to make the traffic system intelligent

The key to improve the efficiency of the traffic system is to be intelligent, including the intelligence on information processing itself, which would make use of the information most effectively. For this reason, we have adopted the relevant theories and technology of data mining and artificial intelligence, on the basis of grid computing, the new generation of information analysis and DSS model is put forward.

III. Case-Based Reasoning

Case-Based Reasoning (CBR) is a knowledge-based, problem-solving paradigm that resolves new problems by adapting the solutions used to solve problems of a similar nature in the past.[3,4] A further advantage of this approach is that it allows consolidation of rule knowledge and provides a reasoning engine that is capable of probabilistic-based matching. With CBR technology, development has taken place in an incremental fashion facilitating rapid prototyping of an initial system. The development of robust strategies for integration of multiple health information sources is achieved using reasoning algorithms of progressively increasing complexity.

In contrast to the conventional search engines, CBR systems contain a knowledge model of the application domain in which it operates on. It is therefore not universal but specifically designed for the domain. Hence, it is possible to develop intelligent search abilities, which even show reasonable results when fuzzy or incomplete requests are given. Moreover, the results are ranked and complemented by variants and alternatives, thus, not only matches are given but also information is valued with "more suitable" or "less suitable".

IV. Grid Computing

The Grid computing concept [5, 6], first developed in the scientific community, was initially aimed to address the problems of sharing and working with large datasets. Grid computing is now moving towards a mainstream challenge of creating reliable, robust, and secure distributed systems. The Grid [7] is an aggregation of geographically dispersed computing, storage and network resources, coordinated to deliver improved performance, higher quality of service, better utilization and easier access to data. It enables collaboration across “virtual organizations”, enabling the sharing of applications and data in an open, heterogeneous environment. Services previously considered to be host-centric now can be distributed throughout a network of powerful computers improving quality

of service while also offering enhanced and improved capabilities. The emergence of Grid software such as the Globus Toolkit [8, 9] provides the necessary middleware to implement a Grid system and includes services that tackle issues such as accessibility, security and resource management.

V. Data Mining

Data mining, also called knowledge discovery, simply to say, is to find implicit knowledge and rules from a large amount of data. Different with the traditional knowledge discovery technology (such as statistical method, neural network), data mining technology generally combines closely with the relevant database technology, and it particular emphasis on finding and concluding many knowledge modes and rules which hide in huge amount of data automatically. With this characteristic, data mining has more extensive application prospects.

The course of data mining can be divided into three stages: data preparing, mode finding and result expressing. According to the mode and rule, the analytical method of data mining can be divided into four kinds: Classify and predict analysis, relativity analysis, cluster's analysis and unusual things finding.

VI. System Structure of the Traffic Information Analysis and Decision System

The system is consisted of five main components as shown in Fig.1: central control console, information analysis and data mining, data centre, decision support and business application [10].

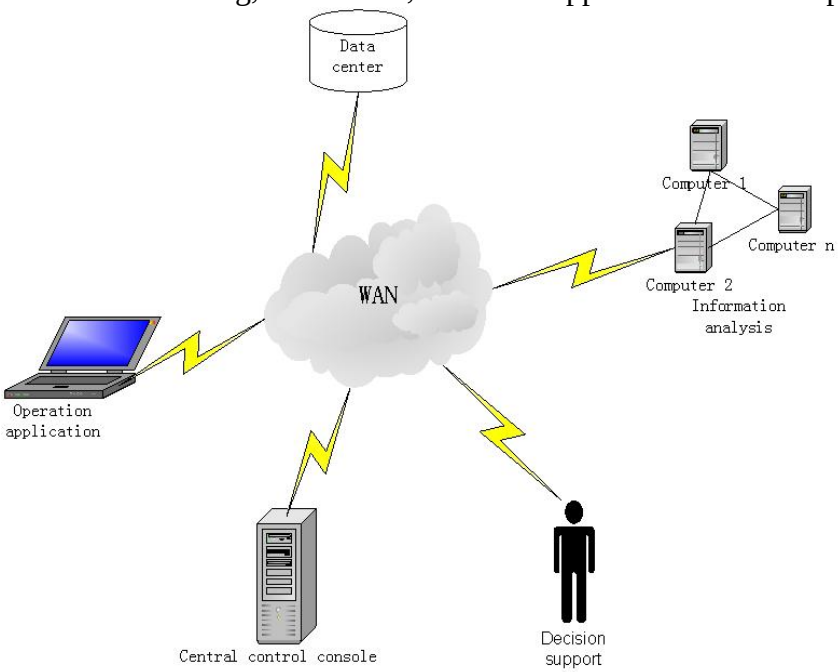


Fig. 1. System structure of the information analysis and decision support system

Central console

This module is mainly used to the assignment of task and resource, and it also controls other parts of the whole system.

Decision support

The main goal of this module is to give the rational solution according to the demand of the users by utilizing the information that the traffic data centre and the information analysis module offered, and the methods and rules that are stored and produced in this module, in an acceptable time as small as possible. For example, it can give the optimization control parameters and control strategy for regional signal control and route guidance; it can also optimize the command scheme of leading. This module is mainly consisted of these components on logic: the knowledge base, model base, method base, reasoning machine, analyzing machine, consulting / explaining system, the problem-analyze system, the solution simulation/appraising/optimization system.

Knowledge base, model base and method base obtain and store various kinds of knowledge, models (including the case), algorithms and rules to be used in reasoning, which are produced by the information analysis module and offered by the expert. Reasoning machine, analyzing machine and consulting and explaining module deal with the data and information and give the preliminary solution, according to the problem model given by the problem-analyze system. The problem-analyze system is responsible for constructing and the concrete problem-solving model supported by knowledge base, model base and method base.

The solution simulation/appraising/optimization system uses the data offered by data centre to simulate the corresponding traffic running environment, then let the preliminary solution implement in the simulation environment, and record the parameter value in the implement course, then uses the recorded data to appraise the preliminary solution according to the appraise index. If the appraising result is not as good as needed, then carry on further optimization according to the appraising result and optimizing criterion. The whole module can not only give the solution, but also appraise and optimize the solution, and it can also predict the emergence, possible reason, and consequence of the traffic incident as well as provide the corresponding solution automatically. The decision support process is showed in figure2.

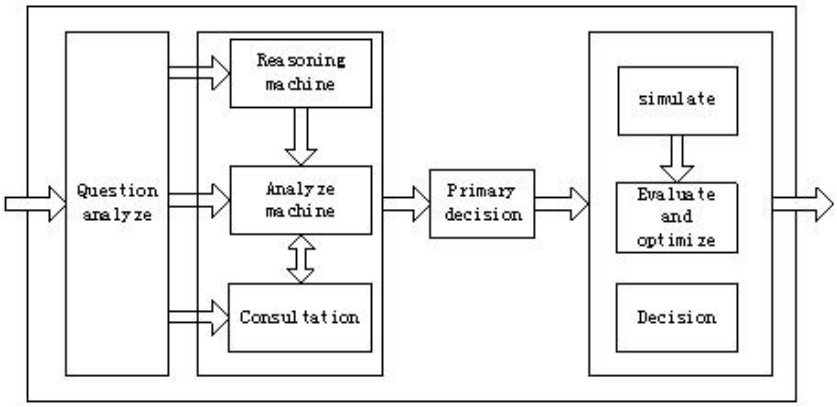


Fig. 2. The decision support process of the system

Information analysis

This module is formed by one or several calculated resources (such as supercomputer, parallel computer cluster), which is responsible for data mining and knowledge discovery. Traditional information analysis is mainly based on the data inquiry in the database, or the useful information found through gathering function to calculate, so that the rule hidden in the data can not be found. But data mining is not merely simple information analysis. It includes data

gathering and comparing, and carrying out the tasks such as associating, classifying, predicting, clustering, time series analysis and other data analysis task, and then the implied mode and interesting knowledge hidden in a large number of data can be found automatically. For example, the neural network can study and be trained by the input data directly, then produce the rules and knowledge, and finally it can set up the calculating model automatically. Adopting time series analysis and array mode analysis to analyze and predict the various kinds of traffic flow information. Applying classify and cluster method to check the traffic accident section and judge the state of traffic jam; Adopting the isolated data mining technique to detect the unusual incident of traffic. Space data mining with GIS assists in getting the space traffic mode model and rule. Through these traffic analysis methods we can obtain the inherent rule and trend of traffic state development and change, and offer effective basis for decision support of traffic management department.

Data center

Data center is consisted of one or several storages which are used to store data sources that the information analysis needed for the traffic application system, simulation data, middle results and final results. In order to transparently visit the distributed traffic data sources, the data center of this system is a virtual database, which can conveniently integrate the information of intelligent transportation system fields such as traffic flow information collection system, signal control system, CCTV (closed circuit television) monitoring system, GPS vehicle navigation system, variable message system, etc. Fig. 2 shows the system structure of the virtual database. Users' inquiry is sent out by Web, and connected with the database through the middleware such as ODBC and JDBC, etc., then it visits the virtual database with SQL. When the VDBMS (virtual database management system) receives the inquiring request, the inquire processors resolve it into different parts , then send them to the corresponding data source to carry on the real data extract, and make up the extractive result, finally send back to the users. Results through this method are like the single database offered to the user. In order to improve the performance, the cache is needed to store the extractive result at first. Besides this, one publishing system is needed to set up physics image for the virtual form, and the image is stored in the related local data memory to improve the speed [11].

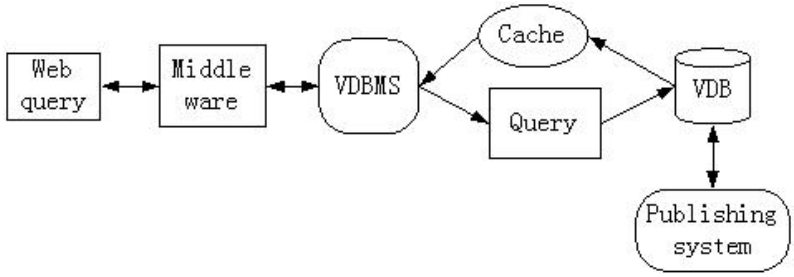


Fig. 3. System structure of the virtual database

The operation application

The operation application is consisted of the traffic management application system. These systems can visit the data center, obtain the middle result data calculated in real time, and display the real-time progress that calculate, and it also can obtain the final result after calculating to summarize and analyze.

In this system, the connection between the module and within the module is linked to each other through LAN or WAN, and the communication and security mechanism of grid is operated on their local operating system, thus a grid computing system is formed.

The information analysis module is made up of several compute resources to meet the computing capability demand for information analysis function. In central console, the multiple tasks submit mechanism of grid is utilized to submit the task to several compute resources in one time, and the co-allocation mechanism of the grid is utilized to coordinate and control the compute tasks running in the distributed computing resource.

While the central console submits the task of computing, it also points out the data which is needed to be moved from the data center to the computing site before computing and the data which is needed to be sent back to data center from the compute site. By this, the data movement mechanism of grid starts, and it can finish the above-mentioned data moved in and moved out safely and reliable in backstage supporter without intervention from users. The compute procedure running at the compute resource site realizes the conveying of the middle result of the computational process in real time through transferring the data visiting service interface of the grid, and storing the data in data center.

VII. The operation course

The operation course of the system mainly includes the following steps:

(1)The question is input by the user or the question is found through the decision support module, and the decision support module is driven.

(2) The central console submits the information analysis task through the multiple task summit mechanism and co-allocation mechanism of the grid.

(3) The grid resource management and monitor procedure in the information analysis module receive the information analysis request. It first makes an authentication, and then to carry on the concrete information analysis request.

(4) Before carrying on the information analysis task, the backstage supporter obtains the necessary data for information analysis from the data center through the data transmission service of the grid.

(5)The local parallel computation environments of the information analysis module start the local application program and data mining algorithms to carry on the information analysis task after receiving the analysis task.

(6)During the process of carrying on information analysis task, the middle result is sent to the data center in real time through the data transmission mechanism of grid, and the data of diagnosing is sent in real time to the control console.

(7) During the process of carrying on information analysis task, the application system can obtain the real-time middle result through visiting data center.

(8) When the information analysis is finished, the backstage supporter transfers the result to the data center through the data transmission service of the grid.

(9)The decision support module carries on the reasoning, analysis, consults according to the result of information analysis, and gets the preliminary solution of the problem.

(10)The decision support module notifies the central console to start the information analysis module.

(11)Information analysis module gets the simulate data from data center, and sets up the simulation environment.
(12)The information analysis module conveys the characteristic parameter value to data center according to the simulate result of the solution.
(13)Decision support module obtains the characteristic parameter value from the data center and appraises the value, through the appraised result to optimize, thus getting the further solution.
(14)Do 10-13 until the evaluation index value is in the range that is established beforehand.
(15) The whole course can be played back to understand the information and decision support course by the record of the central console.

VIII. Conclusions

In order to effectively analyze the huge amount of information which is coming down with intelligent transportation application system and to tackle the challenges faced by the traffic management of changing from simple static behavior to the real-time dynamic management, in this paper, a system model and system structure of traffic information analysis and decision support are proposed. This model adopts the intelligent information analysis methods, decision support theory of new generation, and research results of the fields about knowledge engineering, artificial intelligence and data mining to analyze the information, so the decision is reasonable for the dynamic traffic management. As an organic component of the intelligent transportation system, this model is suitable in an extensive range. It can be used independently, or be imbed into the key subsystem of ITS such as traffic signal control, traffic guidance, incident detect, command management, traffic information service, etc. The application of this system can improve the performance and expand the depth and scope of intelligent transportation system.

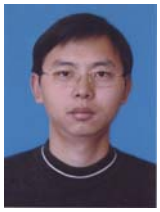
References

[1] Xie Jiameng, Peng Hong, Zhou Bin and Yang Donghui, “Analysis of Intelligent Traffic Information and Research on Decision Support Based on Data Mining Techniques,” *Road*, 2004, 4
[2] He Guoguang, “Some Basic Problems about ITS,” *Journal of Intelligent Transportation System*, 2003,2
[3] J. Kolodner, *Case-Based Reasoning*. Morgan Kauffman, 1993.
[4] D. B. Leake, *Case-Based Reasoning: Experiences, Lessons & Future Directions.* The MIT Press, 1996.
[5] I. Foster, C. Kesselman, and S. Tuecke, “The Anatomy of the Grid : Enabling Scalable VirtualOrganizations,” *International J. Supercomputer Applications*, 15(3), 2001.
[6] I. Foster, C. Kesselman, J. Nick, and S. Tuecke, “The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration, Open Grid Service Infrastructure WG”, *Global Grid Forum*, June 22, 2002.
[7] I. Foster and C. Kesselman, *The Grid: Blueprint for a New Computing Infrastructure.* Morgan Kauffman, 2003.
[8] I. Foster and C. Kesselman, “Globus: A Metacomputing Infrastructure Toolkit,” *In Intl J. Supercomputer Applications*, 11(2): 115-128, 1997.

[9] I. Foster, C. Kesselman and C. Lee, R, Lindell, K. Nahrstedt, and A. Roy. "A Distributed Resource Management Architecture that Supports Advance Reservations and Co-Allocation," *Intl Workshop on Quality of Service*, 1999.

[10] Wang Bin and Xu Zhouqun, "A Grid-Computing-Based solution to High Performance Computer and its Implementation", *Computer Engineering and Applications*,2003,36

[11] Tan Hansong and Hu Chunhua, "Virtual Database Querying System Based on XML," *Information technology*, 2003,4



Liu Chong was born in China, 1981. He received the B. Sc degree in Civil Engineering from Tsinghua University, China, in 2002. Since 2002, he has been a Ph. D. degree candidate in Transportation Engineering from Tsinghua University, China. His current research interests include ITS and Transportation Planning.

Lu Huapu was born in China, 1957. He is a professor of Tsinghua University, China. His main research interests are Transportation Planning and new technology application in transportation management.