Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2013, 5(12):565-569

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Traffic flow data mining based on cloud computing

Fang Xujian, Wu Hao, Seng Dewen and Xu Haitao

School of Software Engineering, Hangzhou Dianzi University, Hangzhou, China

---

ABSTRACT

Communication and information technologies are used by intelligent transportation systems (ITS) to manage traffic flow data and enable people to obtain the services for safe and comfortable driving. Cloud computing can help to handle the large amount of storage resources and mass traffic flow data effectively and efficiently. Intelligent traffic cloud could provide services such as autonomy, mobility, decision support and traffic management strategies, and so on. This paper analyzes the architecture of the ITS using cloud computing and proposes a new architecture that tries to improve the current architecture and reduce the limitation by using cloud computing technology. The architecture organizes, manages and goes to access traffic knowledge from data that are achieved from sensors, vehicles, roadside infrastructure and etc, by using data mining technology. The services and applications of the system are provided through the cloud computing technology.

Key words: Cloud computing; data mining; intelligent transportation systems

---

INTRODUCTION

Today, many cities are facing traffic congestion, rising accident rate and low traffic efficiency. Road capacity cannot easily be increased because of the lack of space for new roads and high development costs. Many cities are developing ITS as a new way to solve the confliction of vehicles and road infrastructure. Advanced traffic flow control methods are introduced to increase throughput and safety of existing road infrastructure. ITS have been accumulating massive and complicated road traffic data that come from wide variety of sources with different manifestation and huge amount of information. The traffic data provides important data base for the traffic management and road control [1-4]. However, there is a new demand for the large number of road traffic data storage, processing, and handling.

Transportation research and development is no longer a field of mechanical, civil and other traditional engineering and management disciplines. Rather, communication, control, computer sciences and many other emerging information sciences and engineering areas have formed the core of ITS and become an integral and important part of modern transportation engineering. Cloud computing provides an optimal solution for the control of traffic allocation process. It provides on demand computing capacity to individuals in the form of heterogeneous and autonomous services. With cloud computing, users do not need to be aware of or to understand the details of the infrastructure in the clouds. They need only know what resources they need and how to obtain appropriate services and those resources, which shields the computational complexity of providing the required services [2].

Data mining is to extract patterns from large data sets by combining methods from statistics and artificial intelligence with database management [5-10]. As an advanced data analysis method, traffic data mining will play an increasingly important role in the application of ITS. The most effective representation and analysis of complex traffic structures and data model approach is to use visualization tools such as charts, trees, cubes, and list link tables. This intuitive structure and mode will promote the understanding and discovery of the laws. Therefore, the various traffic data mining algorithms used in the design and development of ITS are an important content of the current research of ITS. To construct ITS based on data mining will provide an effective guidance tool for traffic management. Although traffic data studies have achieved certain results, and have been used to guide the application of ITS, the existing methods of road traffic data analysis also have some problems. In this paper, we will discuss the problems and propose our approach in solving the problems based on cloud computing.

TRAFFIC DATA MINING

In our research, the traffic data includes the traffic management control data, Road environmental data, and road traffic data. Intelligent traffic management and control system records a lot of traffic control and management data. Electronic police system will record the traffic violation procedure with the images and videos for the police department to provide vehicular traffic violations, including violation vehicle locations, violation date, violation time, violation parameters, violation vehicles panoramic image sequence, violation vehicle license images [2-5]. The accident alarming system provides alarm time, alarm locations, phone numbers, and related accident information. Traffic signal control system provides the relevant operating status information of the road intersection, as well as vehicle management information, driver management information.

Road environment information includes mainly about the road network topology, road surface information such as lane number, lane width, intersection drainage, traffic capacity, transport facilities such as signal, lights, speed limit signs, abnormal events information such as construction information, temporary closures and control measures, weather conditions, public transportation information. Some road environment information could not be obtained from the existing system. It will be collected by hand or from other systems.

Road traffic data is chronologically sampled numerical data sequence. It is the main management and control object of ITS [1-3]. The vehicles on the road become the traffic flow. Traffic continuous flow is formed by the motor vehicles travelling on the road within a certain time without influence or no horizontal cross sections. Typically traffic continuous flow includes the off-ramp of the supreme viaduct segments, cross-river tunnels and highways. Motor vehicles at a road intersection are controlled by traffic lights, so the traffic flow is showing non-contiguous states.

Existing traffic models have inconsistency issues between microscopic models and macroscopic models, which hinder the development and application of traffic theories and models. Macroscopic traffic model is still lack of theoretical foundation. It has only the empirical regression model. Many transportation phenomena cannot yet be explained. Because of the complexity of traffic system, complete and accurate physical model for the road traffic system cannot be established based on the traditional physics, mathematics, statistical methods [6-10]. Traffic characteristics is determined by the irrational behaviors of human, so there is a big gap between the simplified and assumed traffic model and the transportation system in real world. Real-world traffic behavior does not fully in accordance with the assumed model; therefore, it is unable to effectively analyze traffic conditions and is of significant limitations.

Data mining technology provides a new analytical **tool** for the processing of traffic data, but common data mining algorithms cannot meet the specific application requirements of ITS for traffic data mining analysis. Direct application of the existing data mining analysis methods to the analysis of traffic data cannot get a good effect. Traffic data is of time characteristics and strong spatial correlation. Various traffic information collection equipments have accumulated massive complex temporal data sets [2-4]. These data presents new data mining needs for ITS, such as complex temporal and spatial queries of road network traffic flow, association between road traffic conditions, and analysis of traffic flow congestion model.

New ITS application and analysis tasks need to study new data mining algorithms, so the new rules and features can be found in traffic data. Because of the complexity of the transportation system itself, interdependence of all sub-functions, subsystems is increasingly growing [4-6]. Isolated subsystems will be difficult to make a difference. Therefore, the integration of the various intelligent traffic information sharing platforms have been launched. However, there is currently no comprehensive road traffic data mining platform. This makes the road traffic data mining not been well studied and applied.

Demands of application and development of ITS show that the use of data mining technology to complex spatial and temporal analysis of road traffic data sets need to design suitable data mining analysis methods and new road traffic data mining models according to the characteristics of road traffic information and the application requirements of ITS, so that the decision support information can be provided for the road management, control and induction.

TRAFFIC FLOW FORECASTING

The traffic flow forecasting is an important research of dynamic traffic control and guidance. The real-time, reliable forecasting results are the basis of dynamic traffic management. The traffic flow forecasting methods can be roughly divided into four categories: the methods based on traditional statistical theory, the methods based on nonlinear theory, the methods based on knowledge discovery and the methods based on emerging technologies such as data fusion.

These methods have advantages and disadvantages and conditions. The algorithms based on the traditional methods of statistical theory are relatively simple. However, when the prediction interval is less than 5 minutes, the randomness and non-linear of the traffic flow changes will become strong, so the precision of the model will become poor. The algorithms based on nonlinear theory are well suited to the uncertainty of the changes in the traffic flow, but the calculation is more complicated, and the theory is not yet mature [1-4]. Knowledge discovery-based approaches are good at identifying the characteristics of complex nonlinear systems, but they require a lot of training programs in the learning phase.

The algorithms based on the emerging technologies such as data fusion method can coordinate of multi-source data and improve the prediction accuracy, but have not yet formed the basic theoretical framework and maturity model algorithms [5-9]. In summary, the complex features and limitations of the prediction model of the traffic flow under different conditions are difficult to obtain an accurate prediction results by the use of a single model and method.

The artificial neural network (ANN) is a system constituted by a number of simple neurons in some way. It can massively parallel processing of traffic data, and is of good adaptive ability and good noise immunity. In recent years, many scholars dedicated to research in this area. They improve the prediction accuracy by improving the neural network input layers and the network structures, and achieved fruitful results. However, it is very difficult to overcome the inherent shortcomings of the artificial neural network. The network model is the lack of transparency, when more training samples, the training time is too long and easy to fall into local minima.

The rough set (RS) theory is a new knowledge acquisition method. It is capable of expressing and processing incomplete uncertain traffic information. By the indiscernible relation reduction, RS can extract information between potential rules, and obtain the minimum expression of knowledge. At the same time, the algorithm is simple and easy to operate. However, in practical applications, RS theory is noise sensitive and poor at anti-interference ability.

It can be found through the analysis that ANN and RS theory as two methods of data processing, there are many complementarities. In our research, RS and ANN are fully integrated with the advantages of both to build traffic flow forecasting model. In our model, we use RS theory to pre-process the traffic data, eliminate the redundant data and attributes and extract the rules. Then, we use the rules to determine the number of hidden layers, hidden layer nodes and connection weights of the artificial neural network to complete the network structures. Finally, the data after reduction will be input to the artificial neural network to predict traffic flow. The model has a good transparency, and is easy to explain, and is of good generalization ability and anti-interference ability. In our approach, the estimation results of the traffic flow are show as Fig. 1. As the figure indicates, the combination of ANN and RS has got more accurate estimation.

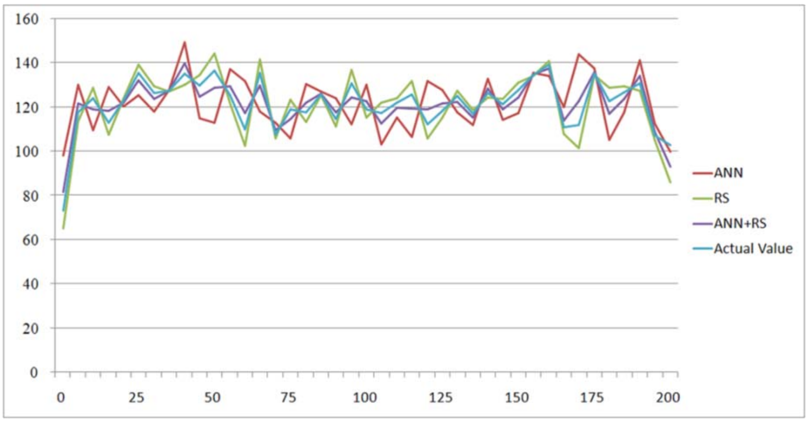

Fig. 1: The estimation results and the actual values of the traffic flow

CLOUD COMPUTING ARCHITECTURE FOR TRAFFIC DATA MINING

The architecture of a cloud includes several key modules [14-18]: user interaction interface, system resource management module with a services catalogue, and resource provisioning module. The system resource management module manages a massive network of servers running in parallel. Often it also uses virtualization techniques to dynamically allocate and release computing resources. Cloud computing system can be divided into two sections: the front end and the back end. They both are connected with each other through a network, usually the internet. Front end is what the client sees whereas the back end is the cloud of the system. Front end has the client's computer and the application required to access the cloud and the back has the cloud computing services like various computers, servers and data storage.

Monitoring of traffic, administering the system and client demands are administered by a central server. It follows certain rules, protocols and uses special software called the middleware. Middleware allows networked computers to communicate with each other. Cloud architecture has the five layers [19-21].

Client layer: a cloud client consists of computer hardware and/or computer software that relies on cloud computing for application delivery, or that is specifically designed for delivery of cloud services and that, in either case, is essentially useless without it. Examples include some computers, phones and other devices, operating systems and browsers.

Application layer: cloud application services or "Software as a Service (SaaS)" deliver software as a service over the Internet, thus eliminating the need to install and run the application on the user's own system and simplifying maintenance and support. Important characteristics include network-based access and management of commercially available software; activities that are managed from central locations rather than at each user's site, enabling users to access applications remotely via the web; application delivery that typically is closer to a one-to-many model than to a one-to-one model, including architecture, pricing, partnering, and management characteristics; centralized feature updating, which obviates the need for downloadable patches and upgrades.

CONCLUSION

Real traffic flow data from Shaoxing city of China are derived as a case study to demonstrate the effectiveness and efficiency of the data mining of ITS system architecture based on cloud computing. The traffic flow data are used to recognize the traffic flow state. We use 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 150 and 180 million bytes of traffic data and use with different Reduce nodes and Map nodes. The detection results of the traffic flow are as shown in Fig. 2. The results shows that the cloud computing can greatly enhance the efficiency of the data mining algorithms.

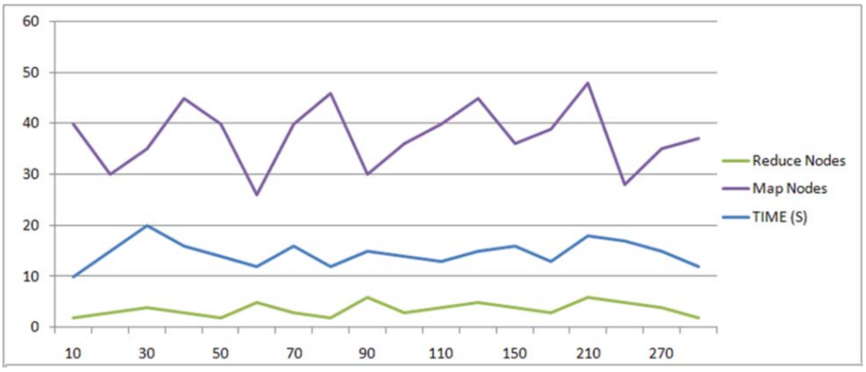

Fig.2: Results of different Reduce nodes and Map nodes

The main objective of the proposed architecture of data mining of ITS system based on cloud computing is area wide management of traffic flow aiming to increase safety of all traffic participants and maximize the traffic throughput. The architecture realized by a distributed control system based on cloud computing principles will have approximately infinite computing capabilities, scalability, and pay-per-use scheme, and can be developed independent parts concurrently with other parts. It makes performance improvement, fewer restrictions and disadvantages of previous ITS architectures, and more ease of access to the ITS services and applications.

Acknowledgments
The authors wish to thank the National Natural Science Foundation of China for contract 61100194, the foundation of education department of Zhejiang province of China for contract Y201120520 and the research foundation of Hangzhou Dianzi University for contract KYS105612008 and YB1205, under which the present work was possible.

REFERENCES

[1] M.A Vouk. *Journal of Computing and Information Technology*, v.16, no.4, pp.235-246, **2008**.
[2] Prashant T., Kavita D., Manish S. *International Journal of Soft Computing and Engineering*, V.2, July **2012**.
[3] Arora A.. *Traffic Engineering and Control*, v 53, n 10, p 375-378, November **2012**.
[4] B. Chen and H. H. Cheng. *IEEE Trans. Intelligent Transportation Systems*, vol.11, no.2, pp.485–497, **2010**.
[5] LIN T. Y. Granular Computing: From Rough Sets and Neighborhood Systems to Information Granulation and Computing in Words. European Congress on Intelligent Techniques and Soft Computing, Heidelberg, Germany, pp.1602-1606, September 8-12, **1997**.
[6] Seng Dewen. *Applied Mechanics and Materials*, v. 189, p. 482-485, **2012**.
[7] Wang C., Li S. *Advanced Materials Research*, v 664, p 1066-1071, **2013**.
[8] SENG Dewen, LI Zhongxue. *Journal of Liaoning Technical University*, Vol.27(1), pp.9-12, **2008**.02.
[9] Dikaiakos M. D.. Minersoft: *ACM Transactions on Internet Technology*, v.12, n.1, June **2012**.
[10] Seng Dewen, Shu Yueqing. *Advances in Intelligent Systems and Computing*, p. 393-400, **2013**.
[11] Du P. J, Liu S.C., Xia J. S. *Information Fusion*, v.14, n.1, pp.19-27, January **2013**.
[12] Di Lecce, V. Amato, A. *IET Intelligent Transport Systems*, v 5, n 3, p 149-158, September **2011**.
[13] Prashant T., Kavita D., Manish S. *International Journal of Soft Computing and Engineering*, v.2, July **2012**.
[14] Seng Dewen. *Applied Mechanics and Materials*, v. 189, p. 478-481, **2012**.
[15] Zadeh L A. *Fuzzy Sets and Systems*, **1997**, 90(2): 111-127.
[16] Hu T., Chen H., Huang L., Zhu, X. A survey of mass data mining based on cloud-computing. Proceedings of the International Conference on Anti-Counterfeiting, Security and Identification, August 24, **2012**.
[17] Seng Dewen. *Applied Mechanics and Materials*, v. 189, p. 457-460, **2012**.
[18] Wang J., Wan J., Liu Z., Wang P. Data mining of mass storage based on cloud computing. Proceedings - 9th International Conference on Grid and Cloud Computing, p.426-431, **2010**.
[19] Seng Dewen, Shu Yueqing. *Advances in Intelligent Systems and Computing*, p. 401-408, **2013**.
[20] Zhang Q, Ma Z.K, Wang L. *Pattern Recognition Letters*, v.34, n.2, pp.185-193, January 15, **2013**.
[21] C.J. Liau. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, V. 8, No.1, pp.93-118, **2000**.