

Traffic Flow Forecasting Model Based on Data Mining

Xin Guo

School of Software Service Outsourcing
Jishou University
Zhangjiajie, China
jianghai079@126.com

Abstract—The research and application of Intelligent Transportation System has developed rapidly due to the demand on safe, convenient, comfortable and information-based modern transportation. It is important part of the research of Intelligent Transportation System to study different forms and operation rules on traffic flow and establish rapid, stable and effective traffic flow model. With the development of Intelligent Transportation System, mass traffic flow data have been accumulated in Intelligent Transportation System. More and more researchers have started to analyze the information of traffic flow by use of advanced data-mining technique, and discover hidden transportation mode and regulation amongst the information of traffic flow. Data mining technology provides a powerful analysis and processing function of mass traffic data. This paper analyzes the characteristics of traffic data in intelligent transportation system, and puts forward the system model and the hierarchical architecture of traffic data mining system.

Keywords-ITS; data mining; model; system structure; technical analysis; traffic flow

I. INTRODUCTION

Data mining technology is a kind of data processing technology developed in recent years. It provides a powerful and flexible data analysis and processing function in large scale data. It has been applied in decision support system (DSS)[1-3]. Safety, convenience, comfort and the traffic demand of the information, the research and application of intelligent transportation system has achieved rapid development. Various advanced information technology has been widely used in intelligent transportation system, intelligent transportation system has accumulated huge and complex traffic data, complex traffic data has put forward new requirements to the management and processing of information. The paper proposed the use of data warehouse, data integration platform, etc., to organize and manage complex intelligent traffic information, and to carry out data fusion, data compression, data standardization, data mining, data analysis and processing[4]. Data mining technology is a kind of data analysis and processing technology that is produced in the application and is oriented to the application[5]. It can be used to analyze the massive traffic information quickly and effectively, and to excavate a large amount of traffic data. In this paper, the traffic flow forecasting and traffic flow prediction based on fuzzy logic is studied. The traffic flow forecasting method based on

ARIMA, neural network and non parametric regression is studied[6].

The data mining technology is used to mine the traffic system's real-time traffic model and the comprehensive traffic model. It can be used in the traffic management and control, improve the service level of the intelligent transportation system[7-8]. This paper analyzes the characteristics of traffic data in the intelligent transportation system, and puts forward the main models and methods of data mining. The paper is structured as follows: the second part analyzes the traffic data in ITS; the third part presents the system model and the hierarchical architecture of traffic data mining system; the fourth part introduces the main models and methods of traffic data mining, including traffic forecasting model, classification model and correlation model.

II. TRAFFIC DATA ANALYSIS OF ITS

Intelligent transportation system has a wide range of traffic data sources, including the dynamic traffic flow data and intelligent transportation subsystem management control data, as well as static road environment data. Intelligent transportation system management and control of the object is the traffic flow, traffic flow data is sampled by the time sequence of a series of numerical data sequence, is the most important data in the transport system[9-10].

Intelligent transportation system (ITS) recorded a lot of traffic information, such as electronic police system the vehicle traffic violations of images and data are recorded, providing information for traffic violations, including vehicle illegal sites, illegal date and time of illegal, illegal, illegal parameter, illegal vehicle panoramic image sequence, illegal vehicle license plate image; traffic accident alarm system provides the alarm time and alarm location, alarm phone number and related traffic accident information; traffic signal control system with intersection of running state, color step progressive information[11-12].

Road environment information has the road network topology, road surface information, capacity, main line speed limit signs, abnormal events (such as construction, etc.), weather conditions, public transportation information, vehicle information, driver information, certain road environment information can not be obtained from the existing system, need to be integrated through artificial means or other systems.

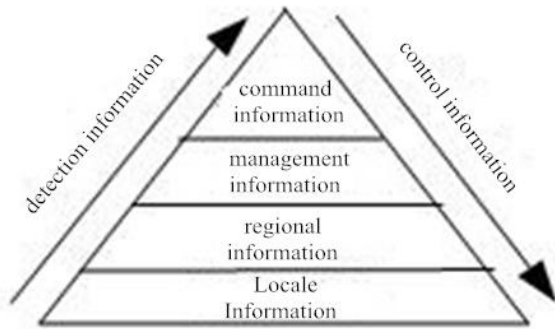


Figure 1. ITS hierarchical distribution

III. TRAFFIC DATA MINING SYSTEM STRUCTURE

A. System model

Data mining process is divided into three main stages: data preparation, pattern discovery, result expression and interpretation.

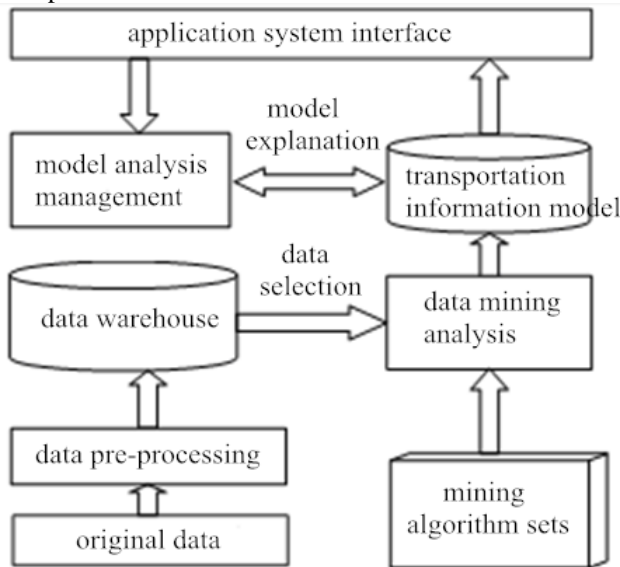


Figure 2. Traffic information mining system function diagram

ETL and data pretreatment model for traffic information that provide a clean, consistent, integrated and reduction (reduction) of the data set of traffic information data warehouse. Data mining task management in data mining algorithm to focus on the completion of the task of mining algorithm, the selection of data mining algorithm in the traffic data warehouse, the implementation of the corresponding mining operations, the mining model to the traffic information. Model analysis management is the application interface of traffic data mining system and other intelligent transportation system, and the feedback information of the application system is used to explain and evaluate the model of traffic information model.

B. Traffic data integration

The system model of traffic data mining system provides a multi tier application architecture. It can be divided into application layer, logic layer, algorithm layer and data layer. The application layer is the entrance to the

analysis function which is set up by the user call analysis logic, and the logic of the application is the analysis ability of the application system. Multi tier architecture can be used in cross platform and network environment, and the application system can be used in a flexible way, such as B/S, C/S, etc..

In traffic data mining, analysis of the function of the extraction and response, the selection of data mining algorithms, design is a problem, the need to use the knowledge and data mining technology to the close combination of traffic system. The analysis logic layer is extracted from the actual needs of the analysis model, and the function of a certain independent analysis is achieved by the algorithm of one or more data mining algorithms, each of which is an independent functional unit.

The algorithm and related computing tools, such as data mining algorithm, statistical method, similarity measure method, are the collection of all independent algorithms. In addition to data mining algorithms, the algorithm needs to be based on data mining algorithm, such as clustering algorithm, similarity measure or distance function is the core problem of clustering quality.

IV. MINING MODEL

Traffic forecast includes many levels: the prediction of traffic flow data based on data layer, the traffic flow characteristics, traffic incident and event type, traffic incident and event type, road congestion prediction, the service level of road network, forecast of events and event impact assessment. In intelligent transportation system, the identification and prediction of traffic flow can be carried out effectively to monitor and manage the traffic flow. It has very important significance for the traffic signal control, traffic assignment, route guidance, automatic navigation, accident detection and so on.

A. Forecasting model

Traffic forecast includes many levels: the prediction of traffic flow data based on data layer, the traffic flow characteristics, traffic incident and event type, traffic incident and event type, road congestion prediction, the service level of road network, forecast of events and event impact assessment. In intelligent transportation system, the identification and prediction of traffic flow can be carried out effectively to monitor and manage the traffic flow. It has very important significance for the traffic signal control, traffic assignment, route guidance, automatic navigation, accident detection and so on.

Nineties of last century. With the rapid development of database system wide application and network technology, database technology has entered a new phase, that is from the past management of the development of only a few simple data to be managed by a variety of computer generated graphics, images, audio, video, electronic many types of complex data files, Web pages, etc., and the amount of data is also growing. Giving us a wealth of information, but also reflect the significant characteristics of the mass of information. Era of information explosion. Vast amounts of information gave rise to many negative effects, the most important is difficult to extract useful information. Too much useless information would have a loss of information and useful knowledge distance. This is John Neisibote (John Nalsbert) called "information rich

and knowledge poor" dilemma. Therefore, it is eager to carry out in-depth analysis of the massive amounts of data, find and extract information hidden in them. In order to make better use of these data. But only entry, query, statistics and other functions of database systems, can not be found present in the data relationships and rules, we can not predict future trends based on existing data. But the lack of data mining behind means knowledge. It is under such conditions, data mining technology is introduced.

Data mining is a relatively new database technology, which is based on the database by the daily accumulation of large amounts of data posed, and found potentially valuable information - called knowledge to support decision making. Data mining is a database application technology, this paper, an overview of data mining, data mining clarify the background, what steps and basic data mining techniques are then introduced algorithms and main application areas of data mining, development and current situation at home and abroad development trend.

At present, many research results have been made in the traffic flow forecasting of basic data layer, such as ARIMA, neural network, non parametric regression model. In the prediction model of the characteristics and the status of the description layer, the prediction model of the development and change of the traffic system is also needed.

B. Classification model

Classification model is a widely used data model in data mining. Classification is based on the characteristics of the data set to construct a classifier, the use of the classification of the unknown class of objects to give a class of technology. Assuming the traffic flow data set R , traffic flow classification feature set P , $P = \{p_1, \dots, p_K\}$, R to f : P to f to make each of the traffic flow object and the only one P , then f is called the traffic flow data classification model, or simply as a classifier.

Classification problem to be solved is categorized as an event or object. In use, both can use this model to analyze the existing data, it can also be used to predict future data. For example, use classification to predict which customers are most likely to respond to sales of electronic mail, and what the client might change his cell phone service provider, or in the medical field when faced with a case to determine what classification from which drugs started better.

Classification technology has applications in many fields, such as to construct a classification model can be classified by customers to carry out a risk assessment on bank lending; the current marketing is a very important feature is emphasizing customer segments. Customer group function is also here, using data mining classification techniques, customers can be divided into different categories, such as call center design can be divided into: Call frequent customer, customer occasionally large number of calls, the customer stable call, other help call centers to find out characteristics of these different types of customers, such a classification model allows users to understand the distribution of the different categories of customer behavior; other classification applications such as document retrieval and search engine automatic text classification; the security field intrusion detection based classification technology and so on.

Researchers machine learning, expert systems, neural networks, and other areas of statistics and has put forward a number of specific classification prediction method.

The establishment of classification models generally has training (or learning) and testing two steps. In the training stage, the classification model is created by analyzing the characteristics of the training data set. In the testing stage, the classification model is tested and the classification accuracy of the model is tested by using the test data set. To achieve the requirements of accuracy, the model is established successfully, and the classification model is used to classify the unknown data tuple or object. Classification is an important data mining technology, and the research of classification algorithm has also made a lot of results. Such as decision tree classification algorithm ID3, C4.5, based on the probability of Bayes classification method NB, TAN, BP neural network method, based on association rules classification method CBA, based on fuzzy logic classification method, etc..

The classification model is usually expressed in the form of rules, decision tree, or mathematical formula. Decision tree classification model is a tree structure similar to the flow chart, which is the most popular model in the classification analysis. It can be easily used to show the results of mining. Each branch of a decision tree corresponds to a classification rule, so the decision tree classification algorithm can eventually output an easy rule set. Using decision tree to classify the unknown data object, the attribute values of the data object are tested on the decision tree. Each branch represents a test output, each internal node represents a test on one attribute.

Traffic flow can have different classification characteristics P , so the establishment of traffic flow classification model also need to choose the different relevant attributes, delete the training process and classification tasks are not related to the property. Different classification algorithms can also establish different traffic flow data classification models, which have been studied by using neural network method and fuzzy logic method. The method of fuzzy logic can be easily understood by a group of if... The classification rules of then, and the neural network method is used to train the neural network model, which can be used to classify the new data.

C. Correlation model

The association rules are shaped like the pattern, and the general use of the support and the degree of the two indicators to describe an association rules. Association rules can find out the correlation between the data items. Traffic flow information is the time correlation and spatial correlation, with time and space characteristics, for example, traffic data only in a certain moment and when the intersection is meaningful. Spatial data is mainly refers to the digital map, remote sensing data, medical images, traffic control, environment and other areas of large numbers of space related data, including the spatial information has a distance, topology, etc.. Analysis of the spatial data of the association may be "30 of the hospital door 88% meters within the scope of a flower shop".

The data mining model of spatio-temporal data, including temporal and spatial, temporal and spatial, temporal and spatial clustering and association, evolution

rules, spatial association rules are defined according to spatial predicates and not according to terms. Intelligent traffic information is an important application value for the prediction of intelligent transportation system. For example, the traffic accident in the A site of T1 to T2 time period X to Y site has unusual high traffic flow in T2+2 to T2+3 time B."

V. CONCLUSIONS

Intelligent traffic data mining is found to be valuable in a large amount of traffic information. Data driven approach is used to analyze the traffic conditions. The analysis, evaluation and forecasting models of intelligent transportation system are established. The real-time traffic control and information management of the intelligent transportation system can be improved.

ACKNOWLEDGMENT

A Project Supported by Scientific Research Fund of Hunan Provincial Education Department.
Project number: 14B143.

REFERENCES

- [1] DaimlerChrysler Corporation. Cross Industry Standard Process for Data Mining[EB/OL]. <http://www.crisp-dm.org>, 1999
- [2] Ichiro Masaki. A brief History of ITS[R]. USA: Massachusetts Institute of Technology, 1999.
- [3] Qin Xiao-hu, Liao Chuan-jin, Huang Xi-Yue, Qin Guo-qiang. ITS Information Architecture, Management And Processing. IEEE, 2003
- [4] Ruimin Li, Qixin Shi, Wei Shen. Development of Integrated Information Platform for Intelligent Transportation Systems. IEEE, 2003
- [5] National proseminar of ITS Shared Information Platform Technique, Chongqing, 2002
- [6] Dougherty, M., Cobbett, M. Short-term inter-urban traffic forecasts using neural networks. International Journal of Forecasting, 13, 21-31.
- [7] Hai Yang, Fengxiang Qiao. Neural network approach to classification of traffic flow states. Journal of Transportation Engineering, 1998
- [8] Jia Lu, Li Cao. Congestion evaluation from traffic flow information based on fuzzy logic. IEEE, 2003
- [9] Voort, M., Dougherty, M., Watson, S. Combining Kohonen maps with ARIMA time series models to forecast traffic flow. Transportation Research C. 1996, 4(5):307-318.
- [10] Chao Han, Su Song. A review of some main models for traffic flow forecasting. IEEE, 2003
- [11] Brian L. Smith, Michael J. Demetsky. Traffic Flow Forecasting: Comparison Of Modeling Approaches. Journal Of Transportation Engineering. July/August, 1997, 261-266.
- [12] Der-Horng Lee, Shin-Ting Jeng. Applying data mining techniques for traffic incident analysis. Journal of The Institution of Engineers, 2004.