

## An Road Traffic Information Analysis in Data Mining Study

Putturu Janaki<sup>[1]</sup>, P Karthikeyan<sup>[2]</sup>

<sup>[1]</sup>PG Scholar, Dept of MCA, SIETK PUTTUR.

<sup>[2]</sup>Asst. Professor, Dept of MCA, SIETK, PUTTUR, AP.INDIA 517583.

### Abstract

According to the characteristics and requests of data analysis of traffic information, data mining techniques are put forward to apply to the management and decision-making of traffic information. Based on this foundation, an integrated model of the traffic information intelligent analysis and decision support is further given. This model has data mining techniques and data warehouse at its core, and fully applies computer intelligence techniques and the theory of the decision support system. As main parts of this article, the model's main components, key techniques and implementing plan are elaborated.

**Keywords:** Data Mining; the Association Rules; Intelligent Traffic; Traffic Information; Integration

### 1. Introduction

Transportation is the foundational industry of a nation's economy. In the past several years, transportation system has already accumulated a great deal of data. Currently, these data, however, are only used in conventional inquiring and statistics service to get some skin-deep information. And the information which can describe these data's overall characteristics and can forecast these data's developing tendency are not mined out, because these information are deeply hidden in those enormous data. How to make use of these data to discover knowledge and to enhance traffic management effectively and scientifically is a current instant problem to be dealt with. Data Mining, one of the hottest topics in fields such as database, artificial intelligence, statistics etc., can discover some useful data associating mode and developing trend by mining a great deal of data in the data warehouse. By data mining technique, the road traffic data can be transformed to information which is helpful to decision-making, and decision-makers can be supported with more accordance to make valid decision<sup>[1]</sup>.

### 2. Analysis of the structural characteristics of road traffic information

In the whole road traffic system, both information and information sources are extensive, diversified and numerous.

(1) Subjects-related data. According to traffic subjects, all the road traffic information can be divided into categories such as road environment information, road facilities information, traffic signal controlling information, traffic accident information and so on. And each of these categories can be also subdivided. Thus, traffic sub-systems can obtain amount of various information through different ways, but each system was independent and the data from each system was used independently. As a result, problems such as data isolation and data waste come out.

(2) Isomeric data. According to various subjects, information database presents corresponding isomeric data environment. In addition, types of data and ways of obtainment are also isomeric. For example, Access data source, Sql server data source, Excel data source etc. The data's isomerism can produce difficulties when the data are integrated.

(3) Timespace-related data. Timespace-related traffic information, on one hand, provides strong support for traffic control, estimation, research etc., and on the other hand, it puts forward new requests for the information processing technique.

(4) Data with life characteristics. From collection, fusion, process, application to abjuration, the information possesses the living creatures' growing characteristics such as heredity, variation, choice and evolution. This kind of dynamic state characteristics brings dynamic indeterminacy to the traffic information.

In a word, it can be seen that traffic information has the characteristics such as dynamic state, indeterminacy, timespace-relatedness, subject-relatedness, life characteristic and so on. Thus the classic mathematical formula and traditional mold method can no longer satisfy the request of information analysis, so it is urgently necessary to adopt a new generation of data fusion, decision support and analysis techniques which make use of result of data mining research to carry on intelligence analysis, to discover potential capacity of each system and further to

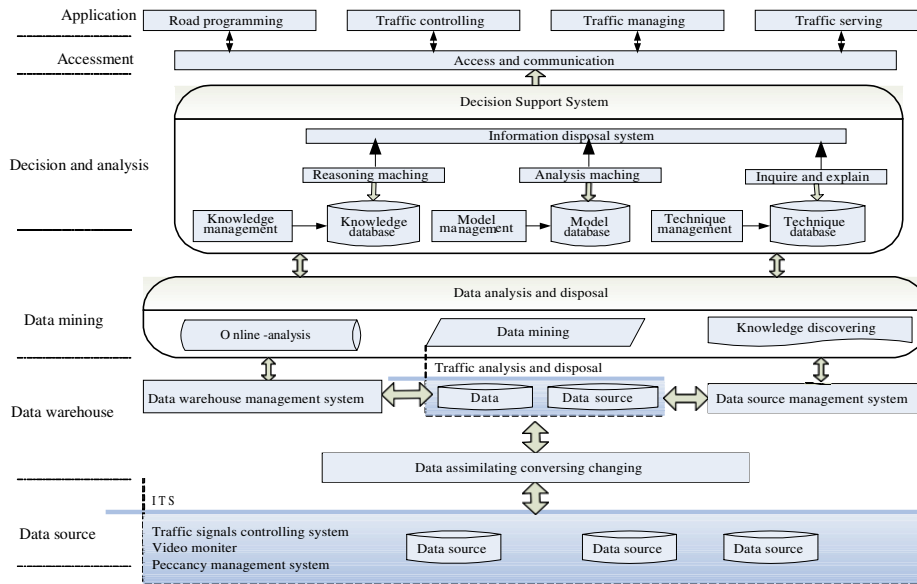


Figure 1. Model of analysis and decision of traffic information

provide scientific and accurate support for dynamic traffic management.

### 3. Model design

The model of analysis and decision-making of traffic information is shown in Figure 1. It includes data source module, data warehouse module, data mining module, decision and analysis module, access and communication module, and application and management module. Its key techniques include data mining and data warehouse.

#### Data warehouse module

The basic principle of data warehouse's technique is to draw out a dataset from the database by a series of processing such as cleaning, transforming and loading. This dataset can be applied to support decision-making. According to the information integration and data mining, the data warehouse is the core of traffic information analysis and decision-making. Figure 2 shows the process of traffic information data warehouse building, whose implementation includes data collection, data extraction, data transformation, data loading, information integration, communication protocol etc. Considering extensively realms of traffic, it is required in the building process that the data warehouse should be flexible and traffic information should be collected from each sub-system [2].

**Data extraction and transformation.** With information techniques collecting bottom layer data

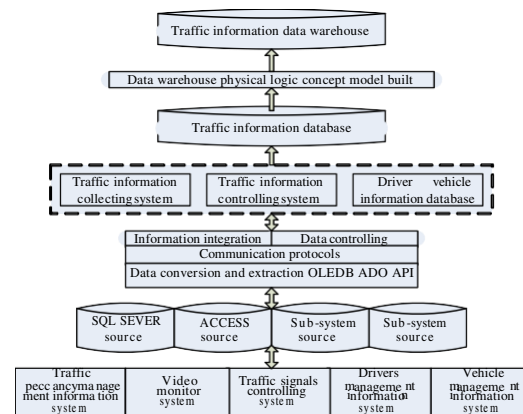


Figure 2. Traffic informationData warehouse building

from databases, the Data warehouse can be built. These bottom data are from traffic stream collecting system, signal controlling system, video supervision system, peccancy management information system etc. Through data extraction and transformation, these collected data can be got into following information integration. Certain of rules and opportunities should be followed and caught in the data extraction and transformation. For example, the car stream data is meaningful only when it is related to street crossings at certain time. Otherwise, it is hard to be comprehended and applied [3].

In the data extraction, the coming rules should be followed. The first one is to keep data's integrity. That is, the data should be extracted on certain standards so that and extracted data can reflect panorama of objects; Secondly, attention should be paid to the relevancy

among the data. The higher the relevancy is, the more accurate the information is, and the stronger information's adaptability is. The third one is information's approachability. In other words, the information drawn out should be put into certain approachable form so that can be applied smoothly<sup>[4]</sup>.

**Integration.**Integration refers to a particular way to accumulate related data in various data source systems. With a unitive way of data storing, it provides a data foundation for searching and building data warehouse. Its main function is to ensure that target data be drawn out and integrated and the discrepancy of different systems be shielded, so that the data are offered as a support for the establishment and operation of data warehouse. This makes the data warehouse more efficient. In this way, the data in each traffic sub-systems can be integrated intentionally and further, "one-station –search service" can be realized across different systems<sup>[5]</sup>.

### Data analysis module

**Online analysis.**Data analysis module consists of three main sections including online analysis, data mining and knowledge discovery. As is shown in Figure 3, online analysis is divided into four parts: the interface of data obtaining, the memorizer of online analysis, the application server of online analysis, the tool of online analysis. This kind of data analysis is an intelligent analysis technique based on data warehouse. The main function of online analysis includes data processing and searching and generating report. Especially it is used to support complicated analysis operation, and pays attention to the decision support for senior administrator and decision-maker. With online analysis, information searching among deals of complicated data can be achieved quickly and flexibly, so that relative analysts can read the information rapidly, congruously and interactively, examine the information from different angles and obtain a deeper understanding about these data which is transformed from the basic data. The result of analysis will be provided for decision makers in a clear, intuitionistic

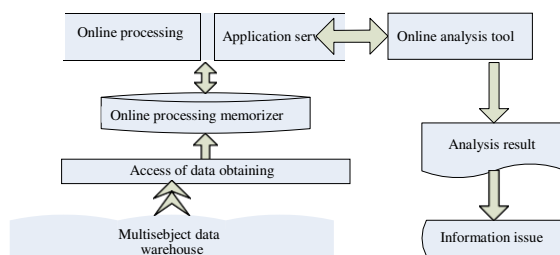


Figure 3. Online processing

and approachable form.

**Data mining.**Data mining is the critical technique in this module. It refers to a kind of method, tool and process of decision support. What are mined out from a great deal of traffic information data include valuable relation, mode and tendency that are implicit, unknown and used for decision-making. These knowledge and established rules are used for model of traffic analysis and decision-making and provide forecasting information as decision support for the traffic administrator. For example, the information of traffic stream can be analyzed and forecasted by adopting time sequence analysis and sequence mode mining. And one more example, relative model and rules are built by data mining in model of traffic combined with GIS, so the law and tendency inside the traffic development can be discovered to provide groundwork for scientific decision. The Figure 4 shows the data mining procedure, in which the following points should be stressed.

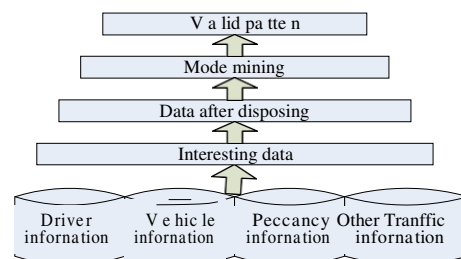


Figure 4. Traffic information data mining procedure

Firstly, a great deal of real data is the foundation and precondition for implementing data mining. Generally, the data used for data mining has two kinds of source: data warehouse or database. The data in data warehouse is organized with different subjects, which provides convenience for data mining to choose the suitable data source. In database, a great deal of data is accumulated and integrated so that a user can see clearly data sources and usages.

Secondly, mining target should be explicit. It is hard to achieve success by mining blindly. Analyzing aimed data which user is interested in is necessary to data processing.

Thirdly, participants in the work of data mining should consist of professional experts, data managing personnel, data analyst and data mining experts. They are assigned jointly to look for a set of method for gradually establishing mining models of database.

Finally, evaluation on the mining modes is necessary. Result from data mining is provided for decision. The mining mode's validity must be evaluated to choose an optional mode which the

administrator supports, approves and participates for traffic administrant decision-making.

**Knowledge discovery.** The knowledge discovery is a repeated process to identify new and valid mode in the data. The modes have potential usability, and can be comprehended in the end. The process of the knowledge discovery can be divided into three main stages: data preparation, data mining and result expression and comprehension. The result from traffic information analysis comes out as knowledge, which is the potential information mined out and used for decision-making by expressing.

### Decision-making analysis module

Decision-making analysis module is used to set up a decision support system, which includes knowledge database, model database, and method database. The data of knowledge database comes from the reasoned data after information processing. After the data is analyzed continuously, the model of traffic management is set up. The data in the mode database is applied in management actually. By offering information data, simulative operation is carried on. After being compared, optimized and put into simulate operation, the optional scheme that is gained through the system can be thus handed over to the user as a feasible solution. This solution provides more credible guarantee for decision-making. The method database are provided for decision-making in the course of continuous consultation and explanation after being permitted.

### Access and communication module

The mode of access is divided into the database interview, the Socket according to TCP/IP, Web-service, and FTP etc. Access modules of dissimilar types run independently. And the amount of interfaces can expand continuously while communicating with more data source. By using the video technique, the module can provide a system window which is easy to be comprehended and used. What's more, intelligent correcting and self-learning is the virtues of the interface system. It also possesses the function of carrying on transformations between natural language as well as mankind's mode of thinking and computer. When the system makes decision, corresponding instruction is provided for decision maker to use based on relevant instruction format. It is convenient for traffic service and controlling.

## 4. Model applications

The analysis and decision-making model of traffic system serves the traffic administrator and participants. Its application mainly involves road network programming, road facilities service construction, transportation control, traffic service, traffic safety etc.

In the field of road network programming, by analyzing various traffic information road constructions in cities synthetically, effective approach to the network programming can be made to treat with the problem of traffic jam and to decrease the number of traffic accidents. By making use of the data mining technique, association rules are used to mine the data from traffic signals controlling system and GIS. This can make it come true that transportation stream is forecasted and decision is brought to the traffic control. For a better traffic service, it provides valid information by choosing the optional route and analyzing road environment, and also provides the service of transportation navigating. The model of analysis and decision of traffic information can realize regional traffic management. It can forecast and analyze the traffic safety to avoid accident happening.

## 5. Conclusions

For the sake of validly analysis of great deals of isomorous data which are accumulated continuously in the application of traffic information system and on order to deal with the challenge from simple static management to intelligent dynamic management, this paper puts forward a model of traffic information intelligence analysis and decision support and further provides a objective and scientific basis for dynamic traffic management.

## References

- [1] Dong Liyan, "Based on the data mining application in the analysis of the traffic accident", Journal of JILING University (science edition) 2006 Vol.44 No.6
- [2] Song Zhongshan and Wu Lifeng, "Application of Incremental Data Mining Algorithm LIUA in AVIM", Journal of WUHAN Science University(traffic society and project edition), 2006, Vol.30 No.3
- [3] Zhang Xiao-li and Chen Li "An application of data mining in intelligent traffic system", Journal of Northwest University (society edition), 2005, Vol.35
- [4] Zhang Gang, "Estimation of Vehicle Queue Length with Data Mining", computer application, 2003, Vol.3 No.1
- [5] Yang Yonghu, "Application of Fuzzy Data Mining Technique in Traffic Forecast", Exploitation and Application in computer, 2006, Vol.16 No.4

### About Authors:



**Putturu Janaki** <sup>[1]</sup>,

PG Scholar, Dept of MCA, SIETK PUTTUR.



**P Karthikeyan, MCA., M.E.** <sup>[2]</sup>,

Asst.Professor, Dept of MCA, SIETK,PUTTUR.