Research on Traffic Information Intelligent Analysis And Decision Support System Modeling

Jiameng Xie [1,2], Hong Peng [1], Donghui Yang [2], Yansheng Tang [2]
[1] College of Computer Science and Engineering South China University of Technology, Guangdong Guangzhou, 510641, China
[2] Traffic Administration Bureau of Guangdong Province, Guangdong Guangzhou, 510440, China

Abstract

Intelligent Transportation System (ITS) is a promising approach for dealing with the traffic problems. One of the pivotal problems of ITS is the analysis and processing of traffic information. This paper provides an integrated model of ITS information intelligent analysis and decision support system--IADSS, which is based on computing intelligence techniques and decision support theory of the new era. As a realization of the model, an application, named "Traffic Accident Information Analysis and Intelligent Liability Judgment System" is built. It is used by the traffic administration department of a province in south China, and achieves a good effect.

Keywords: Traffic Information; Intelligent analysis; Decision support; Modeling

1. Introduction

Intelligent Transportation System (ITS), as its name implies, means the integration of transportation and intelligence technology. With the widely applications of ITS in recent years, the mass traffic information has become the most precious deposit. It is one of the important problems and challenges for ITS researchers to comprehend the inner meaning of the abundant information and make full use of them.

According to our researches on the popular ITS applications, we find that the precondition of the analysis module in most of these applications is to assume the traffic system conforming to certain physical model, i.e. the input accords with some specifically probabilities. Then, such traditional approaches as classical mathematics, calculus, and statistic methods etc are used to figure out the optimum solutions. However, in order to achieve the optimum solution, every traffic participator is supposed to act according to the system's premise (this hypothesis is usually made by sampling statistics), which is impossible in the real world. As a result, the outcome is difficult to reflect the changeable world well. Therefore new modeling methods and analysis techniques are requested to meet the demands of intelligent and dynamic controls, and it is the work that we concentrate our attention upon.

This paper is organized as follows. Section 2 discusses the characters of traffic information and the demands of analysis. Section 3 expounds the theories and techniques, which are used for building the intelligent information analysis and decision support system model--IADSS. Section 4 presents an application in traffic accident information analysis as a practical example of IADSS. Section 5 provides a conclusion.

2. Characters of Traffic Information and the Demands of Analysis

On the view of analysis, traffic information has four main characters[2]:

Firstly, the information, which comes from many independent ITS applications, is varied widely in category and scale. They include well-structured data, semi-structured and non-structured data. The quantities of records vary from MB degree to TB degree. Therefore, the basic problem of data management is to design the storage mechanism, the reconstruction strategy, and the display format of data.

Secondly, traffic information (i.e. traffic flow data, route guidance information etc) has highly correlations of time and space. Modern information management techniques, such as time series analysis, spatial data mining, machine learning techniques etc, are quite adept to analyze this kind of information.

Thirdly, traffic information is "subject-oriented". According to different subjects, traffic information can be divided into different sorts, such as traffic flow information, traffic control information, traffic incident information, traffic inducement information, and so on. The data structures and analysis methods of these sorts are also different from each other. Based on the definition of data warehouse[3]: data warehouse is a subject-oriented, integrated, time-variant and nonvolatile collection of data in support of management's decision making process. It provides the facility for integration in a world of unintegrated application systems, which need by traffic information analysis and its decision support.

Fourthly, the traffic information has lifecycle character. Just like the natural creature, the ITS information has alike characters of self-reproduce, self-evolution and decease. The procedures of data collection, combination, processing, and the finally abandon reflect the biological evolutionism of genetic, selection, mutation and evolution. Therefore, the evolutionism, genetic algorithm (GA), and artificial neural network (ANN), and so on are fit for processing such kinds of information.

Based on above researches, we can see that, the traffic information has the characters such as dynamic, uncertainty, time-space relativity, subject relativity and lifecycle. So traditional statistical principles and data modeling methods are not the best way to deal with them. That's main reason why we change the analysis foundation from model-driven to data-driven, and use artificial intelligence, knowledge engineering and intelligent decision support theory. Our goal is to bring forward an integrated platform of traffic information analysis and intelligent decision support system for ITS, and give the user an objective and useful help when make a decision by analyzing the traffic information intelligently.

3. Traffic Information Intelligent Analysis and Decision Support System (IADSS) Model Design

3.1. Architecture Design

The basic principles of IADSS model design are listed below.

Firstly, we regard each ITS application as an intelligent system. So, we use intelligent analysis and the advanced decision support theory (which based on data warehouse, OLAP, and data mining) to break through the limitations of the traditional DSS theory (which based on database, knowledge base and model base).

Secondly, considering the analysis methods, we use computational intelligence, called soft computing, including neutral computing, fuzzy computing, evolution computing and so on, to analyze the traffic information intelligently. Nowadays, such research branches attract more and more attentions and are widely used in practice.

Thirdly, since the model is a general platform, we use software modular programming technique to encapsulate each function into a module. Therefore, it can be easily embedded into the existing ITS application systems. On the other hand, our model can run as an independent system, with the input coming from different ITS application systems through the standard interface.

Fourthly, the model is environment-sensitive. That means the traffic models and its parameters, which used in our system, are not invariable. Contrariwise, they can be adjusted automatically in order to fit the real traffic conditions. Furthermore, since the traffic data are different from areas (or regions), the traffic models and rules, which generated by our system, may vary correspondingly. It ensures the output of IADSS can reflect the diverse traffic characteristic in certain location objectively.

Fig.1 demonstrates the architecture of IADSS.

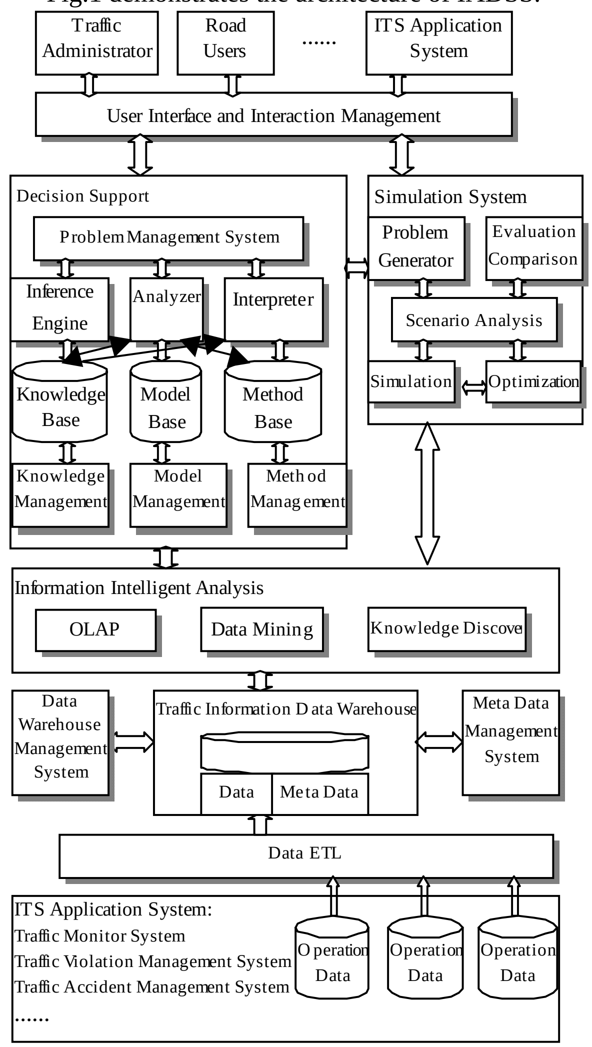

Fig. 1. Architecture of IADSS

3.2. Main Functions and Realization Approaches

IADSS includes 5 main processing modules. They are data warehouse of traffic information, information intelligent analysis module, decision support module, simulation module and user interface.

3.2.1 Data warehouse of traffic information

Data warehouse is kernel of the system architecture. It brings large amounts of data from several sources together and gives a general platform of integrated data. Since data warehouse is physically separate store of data transformed from the application data found in the operational environment, in the period of data collection, we acquire data from multifarious ITS applications. And then do the extracting, transforming and loading. After integration, data enter warehouse. Because data is "subject-oriented", we create corresponding "subjects" for analyzing according to different demands. We choose multidimensional data model, and use snowflake schema or fact constellation schema as data structure. Every subject composes of one or more snowflake (or fact constellation) schema structures, and every snowflake schema structure makes up of a fact table and some dimensional tables. The fact table is the center of every subject. To improve performance of data warehouse, we create 3 assistant parts, they are data collection and ETL (Extract, Transform, Load) subsystem, data warehouse management subsystem and metadata management subsystem.

3.2.2 Information intelligent analysis module

In order to do complex analysis, such intelligent information management techniques as OLAP, data mining, knowledge discovery etc are provided based on data warehouse. It can no matter react the complicated query quickly and agilely, but also discover the hidden patterns and rules without a predetermined idea or hypothesis about what the patterns may be. In IADSS, on one hand, with the OLAP technique, the multidimensional logical view is mapped directly to data cube structure. And we use partial materialization strategy to do pre-computation. Through the operations of drill-down, roll-up, slice and dice, pivot etc on data cubes, we can analyze the data from different points of view and hierarchies. On the other hand, the module offers many algorithms to mine the data and discover the knowledge, such as association, classification, predication, aggregation, time-series and so on. For instance, the time-series analysis and sequence pattern mining are used to analyze the traffic flow information for predicting traffic trend. The artificial neural network can be used on dynamic traffic patterns classification. Furthermore, the methods such as fuzzy reasoning and genetic algorithms can be used in traffic incident detection. Hence, with the help of such intelligent analysis techniques, the system can mine very large datasets much more effectively, find significant patterns automatically, and give the powerful support for the next step of decision-making.

3.2.3 Decision support module

The goal of this module is to give users the reasonable solutions within an acceptable time. The solutions are generated by the information, patterns, models and rules, which come from the data warehouse and intelligent analysis module. The logical structure of this module consists of knowledge base, model base, method base, reference engine, consultation/explanation subsystem, and problem management subsystem. The main functions are: (1) The knowledge base, model base and method base store various kinds of knowledge, models (including traffic cases), algorithms, rules and data match relationships, which summed up by the information intelligent analysis module and traffic experts. The management subsystems of these three bases take charge of the structure definition, optimization and information maintenance. (2) The inference engine, analyzer and consultation/explanation subsystem take the responsibility for controlling the system's execution and explain the results. (3) The duty of the problem management subsystem is to construct the concrete problem models using reasoning rules, models and algorithms, which come from knowledge base, model base and method base. All in all, this model can not only resolve problems, but also discover the accident contributing factors, predict the occurrence of traffic incident and provide the corresponding countermeasures under the user's guidance.

3.2.4 Simulation module

The functions of this model are assisting the decision support module to ensure feasibility and reliability of solutions. The main operations of this model are doing the compare, optimization and simulation on decision projects. It is made up of 5 parts, including problem generator, scenario analyzer, simulation, and optimization and evaluation subsystems. Their functions are: (1) The operations of problem generator and scenario analyzer are recognizing and explaining the problem, and then put it into the specified information environment to do the feasibility and validity analysis. (2) The main target of simulation and optimization subsystems is providing viable (or un-inferior) solutions for decision maker to choose through series of simulation algorithms and optimization strategies. (3) The operations of evaluation and compare subsystems are judging viable solutions through interacting with user, then feedback the information to decision support and simulation modules, finally, revise the solution until user accept it.

3.2.5 User interface

With visual technique, this module provides a friendly interface, which can be easily understood and manipulated. Also, it can translate the analysis results into control instructions according to the instruction format of such ITS application system.

4. An Application Example

In order to diminish accident rates, traffic administration departments need to do the causal analysis of traffic accidents and make the countermeasures. We build the “Traffic Accident Information Intelligent Analysis and Liability Judgment System” as one of the applications of IADSS. This system has already been used by the traffic administration of one province in south China. It comprises 4 main parts: traffic accident information data warehouse, intelligent analysis module, automatic liability judgment module and user interface. According to the software design standard, which commended by the traffic administration department of China, we choose ORACLE products to build the database and data warehouse. The analysis tools are ORACLE OLAP and BUSINESS OBJECTS. The realization scheme is discussed below.

4.1. Traffic Accident Information Data Warehouse

Data warehouse is the center of architecture of the system. It brings large amounts of data from several operational systems, such as traffic accident management system, vehicle management system, driver management system, traffic violation management system, traffic monitor system etc together. After extracting, transformation, cleaning, loading and summarization, data enter data warehouse. Fig.2 shows the data schema of data warehouse.

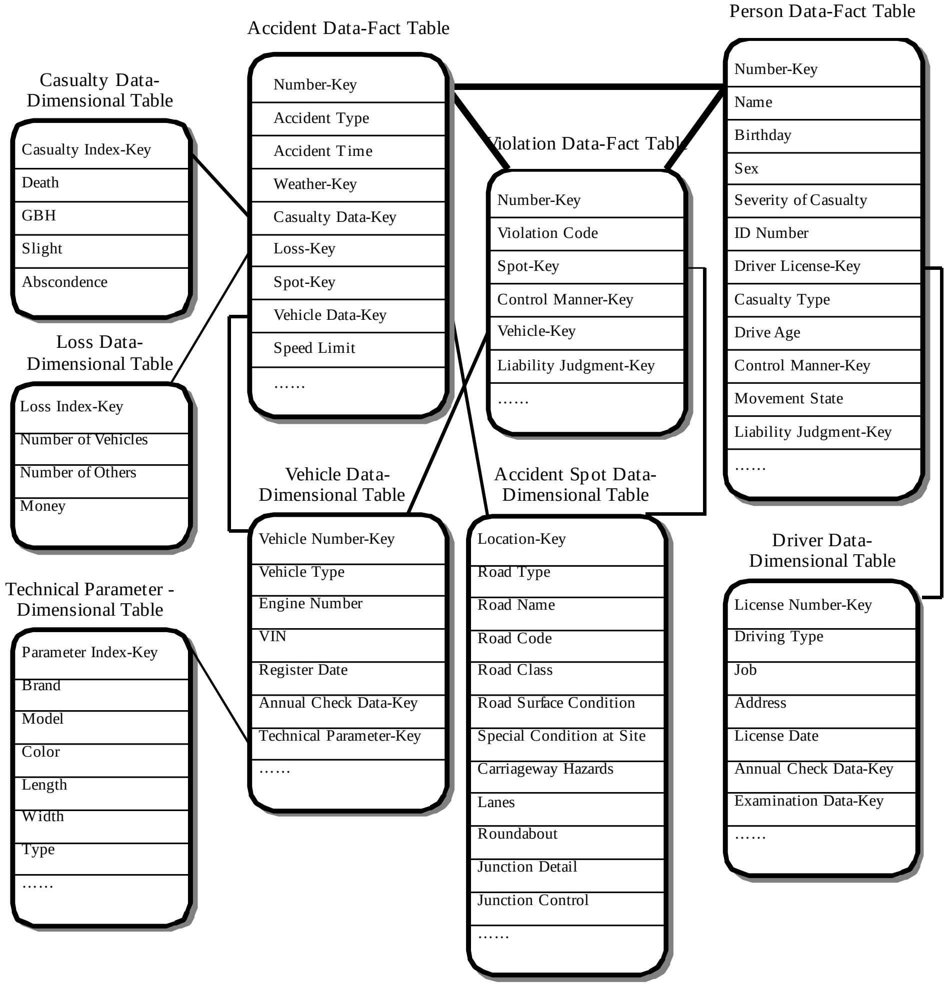

Fig 2. The Fact Constellation Schema of Traffic Information Data Warehouse

This schema can be defined with the Data Mining Query Language (DMQL). The syntax of DMQL is below [4].

(1) Definition of data cube

define cube <cube_name> [<dimension_list>]:<measure_list>

(2) Definition of dimension

define dimension <dimension_name> as (<attribute_or_subdimension_list>)

according to the syntax the definitions of Fig.2. are listed below.

define cube accident-data [number-key, weather, casualty-data, loss, spot, vehicle…]:accident-type, accident-time, speed-limit…

define dimension casualty-data **as** (casualty –index-key, death, GBH, slight, abscondence)

define dimension loss-data **as** (loss-index-key, number-of -vehicles, number-of-others, money)

define dimension vehicle **as** (vehicle-number-key, vehicle-type, engine-number, VIN, register-date, annual-check-date, technical-parameter(parameter-index-key, brand, model, color, length, width, type…)…)

define dimension accident-spot-data **as** (location-key, road-type, road-name, road-code, road-class, road-surface-condition, special-condition-at-site, carriageway-hazards, lanes, roundabout, junction-detail, junction-control…)

define cube violation-data [number-key, spot, control-manner, vehicle, liability-judgment…]: violation-code…

define dimension spot **as** spot **in cube** accident-data

define dimension vehicle **as** vehicle **in cube** accident

define dimension liability-judgment **as** (violation-key, violation-type, fact, reduce-mark)

define dimension control-manner **as** (traffic-key, straight, swerve…)

define cube person-data [number-key, driver-license, control-manner, liability-judgment…]: name, birthday, sex, severity-of-casualty, ID-number, casualty-type, drive-age, movement-state…

define dimension driver-data **as** (license-number-key, driving-type, address, license-date, annual-check (license-number-key, year, check-flag), examination-data (license-number-key, subject1-mark, subject2-mark,…)…)

define dimension liability-judgment **as in cube** violation-data

4.2. Traffic Information Intelligent Analysis Module

This module provides various useful methods and algorithms to search for patterns of information in data warehouse. Furthermore, we not only offer regular algorithms to mine the data, but also do some ameliorations and optimizations on them to ensure that they can fit different traffic status objectively. For example, improved classification and clustering techniques are applied to reveal traffic accident “black spots”. Isolated point mining techniques are used in traffic incident detection. The spatial data mining, which based on Graphic Information System (GIS), is used to discover accident patterns. Apart from data mining, this module also offers many knowledge acquisition algorithms, such as decision tree, rough set theory, Bayesian etc. With these methods, we can dispose imprecise and amphibolous information effectively, grasp main characters of each traffic accident sample, discover accident contributing factors, and facilitate decision-making process.

As an example, Fig.3 shows the result of causal analysis with the traffic accidents, which occurred in a city of south China between January to August in 2002, using OLAP technique. Fig.4 demonstrates the distribution of ages and drive years of vehicle drivers, whom involved in the accident, which occurred in a province of south China in 2002. The color bar represents quantity of traffic accidents. Furthermore, the analysis result using K-Means aggregation algorithm is listed in table 1. Based on it, we can do farther thorough researches on the drivers’ characters for making the pertinent measures on them. Since the analysis degree of this system is deeper and more extensive than before, it satisfies user’s demands in practice.

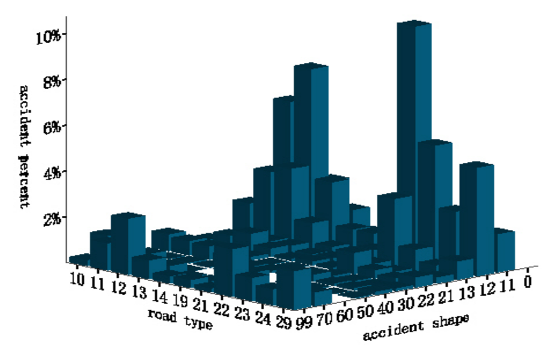

Fig 3. Accident causal Analysis Using OLAP

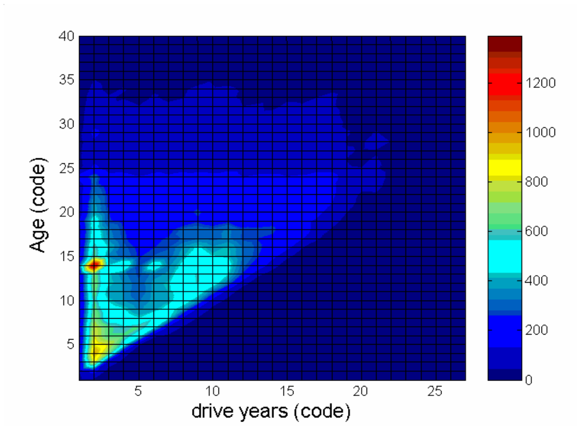

Fig. 4. Distribution of Driver's Age and Drive Years in Traffic Accidents

Table 1. Clustering of Driver's Age-Drive Years in Traffic Accidents

Cluster	Center age-code drive-year-code	Numbers of cases in each cluster	Distance from cluster center					
			1	2	3	4	5	6
1	8 4	30950		8.11	11.12	16.69	23.66	28.07
2	13 9	21606	8.11		7.59	8.69	17.90	20.16
3	19 4	18096	11.12	7.59		10.14	12.64	19.60
4	21 14	15152	16.69	8.69	10.14		12.60	11.47
5	31 7	8783	23.66	17.91	12.64	12.60		12.71
6	31 20	5474	28.07	20.16	19.60	11.47	12.71	

Note 1.The total number of cases is 100061.
2. The number of iterations is 20.

4.3. Liability Judgment Module

This module composes of 4 parts; they are inference engine, rule base, case base, and their management systems. The realization approaches are listed below.

(1) Analyzing cases. The first step of this period is to investigate large scale of traffic accident cases, and summarize experiences of traffic experts. Then gather influencing factors and rules, which need to be considered during the liability judgment. Finally, create a basic two- dimension Relational rule table, as the rudiment of rule base.

(2) Creating rule base. Based on the rule table, in order to form conditional items, we use Apriori algorithm to do association analysis with mass accident samples. Then use production rules to describe all combinations of conditions and their consequences. Finally, generate an integrated traffic liability judgment rule base. In this base, every judgment rule corresponds with existing law, which makes inference result to be accountable.

(3) When a new case coming, the module firstly uses production rules to calculate values of this new case, then tries to match the knowledge base using forward reasoning method. If successful, liability judgment result is generated, or else step (4) is executed.

(4) When the coming case cannot match anyone of the case base, the most similar 5 liability judgments are provided for reference. The attribute-based induction, OLAP and case-based reasoning method are used to search such cases in the case base.

In a word, the main function of this module is to study traffic violation facts and give the automatic liability judgment based on the traffic laws and regulations (see Fig.5). At last, it generates the formal traffic accident liability judgment document. On one hand, judgment result is given to the traffic administrators and persons whom concerned with the accident. On the other hand, it is the important aspect for intelligent analysis in next step. After we finish the programs of this module, we use 1000 real cases to test it. Comparing with the results of manual liability judgment, correctness of this automatic liability judgment is over 95%. This module has already been used by the traffic administration of a province in south China to do traffic accident liability judgment of highway. The future work is to expand the application field to other types of road.

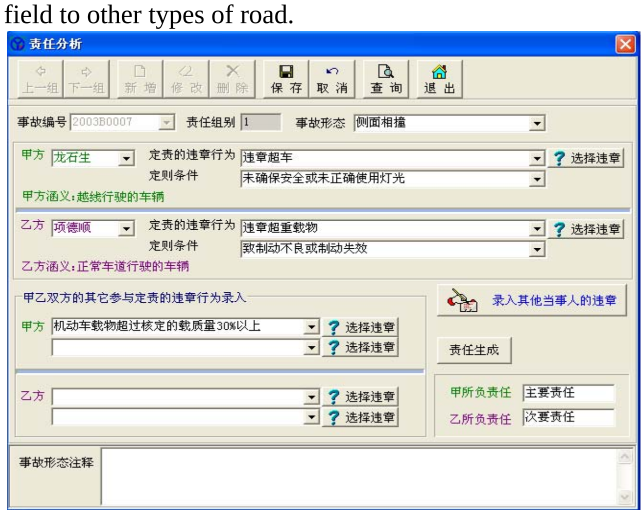

Fig.5. Interface of liability judgment module

5. Conclusions

Traffic information analysis is an important research area of ITS. Because of characters of traffic information, i.e., large scale, highly time-space relativity, subject-oriented, traditional approaches, which based on classical mathematics and statistics, meet many difficulties and challenges. In this paper, a new integrated traffic information intelligent analysis and decision support model--IADSS is proposed, which combines the theories and techniques of decision support system, artificial intelligence, knowledge engineering, data warehouse, data mining etc. As a part of ITS, IADSS not only can run independently, but also can be embedded into such application systems as traffic control, traffic inducement, incident detection, traffic schedule etc to improve performance of these systems.

Furthermore, we give a realization of IADSS in the realm of traffic accident management, and build up an application system, named "Traffic Accident Information Intelligent Analysis and Liability Judgment System". This system has been put into use in the traffic administration of a province of south China. The practice shows that our model is reasonable and effective.

6. References

[1] Yin Hong-bin, Xu Jian-min. "Road Traffic Control Techniques". *South China University of Technology Press*, 2000.

[2] Ren Jiang-tao Zhang Yi Li Zhi-heng Hu Dong-cheng. "ITS Information Characters and the Relational Problems". *Information and Control*, pp.550-554, 2001, 30(6).

[3] W.H.Inmon. "Building the Data Warehouse". *New York: John Wiley & Sons*, 1996.

[4] Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques". *Morgan Kaufmann Publishers, Inc.* 2001.

[5] Chen Xiao-hong. "Decision Support System Theory and Its Applications". *Tsinghua University Press.*2000.