The Realization of Decision Support System for Cross-border Transportation based on the Multi-dimensional Database

Zhihong Jin
College of Transportation Management, Dalian Maritime University, Dalian, China
Email: jinzhihong@dlmu.edu.cn

Qi Xu
College of Transportation Management, Dalian Maritime University, Dalian, China
Email: cogi@163.com

***Abstract*—To provide a more effective data processing system for cross-border transportation for Yunnan Province in China, this paper first analyzed the cross-border transportation data requirements from different perspectives. Then, a systematic structure for the system which is based on the multi-dimensional database was designed. The structure includes three levels: source data level, data processing level and application level. The structure of constellation was designed and adopted to construct the database for the system. The main application of the system includes OLAP and multi-dimensional analysis reports, GIS visualization, and data mining. Later, the paper displayed how the system is realized. The application show proves the effectiveness of the system.**

***Index Terms*— cross-border transportation, multi-dimensional database, decision support system, OLAP, Data Mining, GIS**

I. INTRODUCTION

It is doubtless that cross-border transportation is an engine to promote the foreign trade of a port city and its economic hinterland. As China's 'big province of ports', Yunnan Province has an urgent demand on a new-typed data processing system. The functions of this system should at least include: (1) displaying the port logistics data from different perspectives; (2) providing intelligent aided decision support like potential knowledge discovery for the management departments. Both of that can be realized by constructing a decision support system (DSS) based on the multi-dimensional database (MDDB).

Erik [1] proposed the OLAP solutions to build a multi-dimensional information system; Brobst et al [2] explained the five stages of an active data warehouse evolution; Shi et al [3] made research on some technology of real-time data warehouse; Lin et al [4] explored how the OLAP applies to the multi-dimensional data warehouse; Liu [5] looked into the requirements on constructing the Data Warehouse and introduced the logic model by the example of rail sales system; Meng et al. [6] constructed a multi-dimensional analysis model for a certain port, and discussed the role, constructing mode and application prospect of the DW technology in port logistics systems. Many other scholars made research on the application of data warehouse and multi-dimensional database to different fields [7-15].

The main tasks of this paper are to: (1) make the analysis of the data requirements for the cross-border transportation decision support system; (2) design a systematic structure for the DSS based on the multi-dimensional database; (3) the realization of the DSS and the application show of that.

II. THE CROSS-BORDER TRANSPORTATION DATA REQUIREMENTS

Usually the port cross-border transportation will experience process including port inspection, quarantine, customs, border identity identification. Departments related to these processes are: transportation management department (represented by Yunnan Communications Department), foreign trade supervision department (Department of Commerce of Yunnan Province), customs and other inspection departments. The data all these departments focus are mainly different while intersected to some degree (see Figure 1).

As to the transportation management department, it is necessary to understand the actual situation of difference

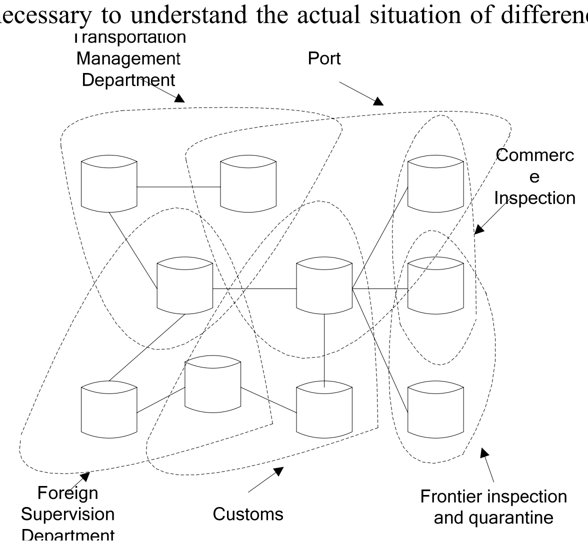


Figure 1. Illustration of different perspectives from related departments of cross-border transportation system

doi:10.4304/jsw.7.5.974-981

between the service ability and demand of port infrastructure; identify what are the bottlenecks among the infrastructure channels that restrict the cross-border transportation development; understand the proportion of cross-border transportation amount that it entails, and when there is traffic congestion there in a certain transportation channel; judge the traffic development situation between Yunnan and its adjacent countries by forecasting the cross-border transportation amount.

What the foreign trade supervision department need is to recognize the general situation of cross-border trade in Yunnan, so as to provide decision support for foreign regulators when judging cross-border trade development trend of Yunnan; accurately predict the prospective new economic growth point through the advantage of the data mining function in the cross-border transportation information management system.

The customs and other inspection departments need to consider how to implement data tracking and monitoring on the whole movement process about goods, personnel and vehicles; analyze how one can increase the convenience of customs clearance, and reduce the proportion that the clearance process time takes up in the whole cross-border transportation; consider the influence of existing tax mechanism on foreign trade, so that customs can adjust accordingly to promote foreign trade business.

. THE SYSTEMATIC STRUCTURE FOR CROSS-BORDER TRANSPORTATION DSS

In this chapter, the modeling method for multi-dimensional database was discussed first. The method is illustrated as shown in Figure 2.

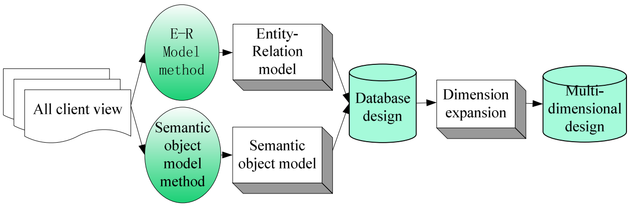


Figure 2. The modeling method for multi-dimensional database

As shown from Figure 2, we see that the process for the modeling for multi-dimensional database can be concluded as follows:

(1) Requirements analysis. During investigation of the specific operation process, acquire what the related departments require, and collect the data media of the business processing—client view.

(2) Classify the client view and establish the object model using both the two methods for data modeling(E-R model and semantic object model). The object model is the cluster analysis results on all the data, and its basic method is to identify the semantic object, entity, relation and other classes.

(3) Standardize all the relation sheets in the database model using the technology of database design, so as to meet the customers' need on the degree of standardization. The process of standardization has to balance between the efficiency of storage and that of processing from the perspective of application.

(4) Operate the dimension expansion based on the object model. By the analysis on the storage structure of the data object model, expanse on the data model according to the difference of 'dimension', and construct the multi-dimensional database model through the dimension expansion.

(5) Execute the mapping from the multi-dimensional database model to the operation.

Based on the process figure and steps illustrated above, a systematic structure for the DSS based on the multi-dimensional database is designed. There are three levels in the structure: source data level, data processing level and application level. Figure 3 illustrated the structure for the DSS in detail.

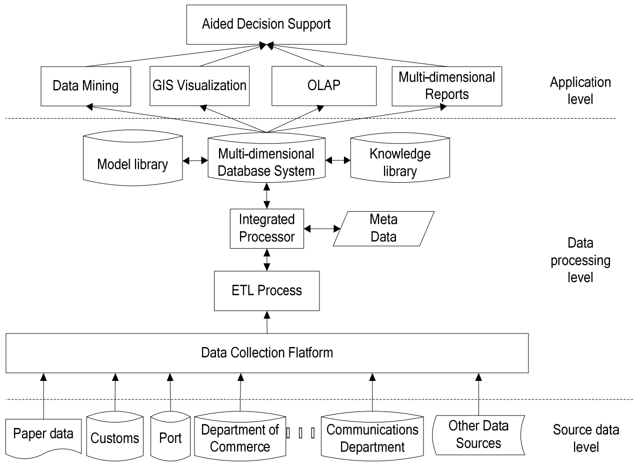


Figure 3. The systematic structure of cross-border transportation decision support system

A. Source data level

In the source data level, integrate all the data that are scattered in different departments, including customs clearance data, trade data from department of commerce, transportation data from the communications department, all departments of inspection and quarantine of specific business data and other relevant data, etc. By the data collection platform, these data are interrelated, integrated and stored in the multi-dimensional database system. These data are used as the data source for the multi-dimensional database.

B. Data processing level

In data collection platform of the data processing level, all the data coming from the source data level are collected. After the 'extraction-transformation–loading' (ETL) process, source data are transferred to clean data, which are ready to be integrated and input into the multi-dimensional database system.

The multi-dimensional database system is the core of the whole structure. The data organization form in the multi-dimensional database system is different from that in traditional database system. In this paper, the structure of the constellation is adopted for constructing the database. The structure is illustrated in Figure 4.There are two fact sheets in the structure: freight fact sheets and passenger fact sheets. Dimension sheets of port, time, transportation mode, vehicle, freight information, customs broker information are connected to these two fact sheets by foreign keys.

Besides, the model library and the knowledge library are also the important parts in the whole structure. There are three basic functional models in the model library, including statistical model, predictive model and simulation model. Among them, the main function of statistical model is finishing statistical historical passenger traffic volume and OD flow, customs clearance data. The main function of the prediction model is forecasting the future development trend of cross-border transportation business based on historical data in the statistical module. The simulation model is the simulation to the actual situation of ports, such as the visualization of the comprehensive transportation network, historical OD flow, etc. From the perspective of application, model libraries mainly include network model, transport model, location model and inventory model. By using models in the model library to operate computation, problems such as location for port logistics node can be solved. On the other hand, knowledge library interacts with the model in model library and data in the multi-dimensional database, so as to realize the combination of qualitative and quantitative analysis, and help policymakers to clarify decision objects, establish and modify decision model. The knowledge and rules provided in the system mainly include empirical knowledge rules, such as customs clearance knowledge, goods knowledge, port knowledge and some rules in the process of model building and choosing. Besides, they also include knowledge and rules obtained in the process of OLAP.

C. Application level

In the application level, four main functions of OLAP, multi-dimensional reports, GIS visualization and data mining are to be realized.

Therein, by following client's thinking mode from different perspectives, OLAP can establish a multi-dimensional data model in advance. Once the model is established, clients can retrieve data from various perspectives quickly, namely multi-dimensional reports. Through multi-dimensional reports, one can analyze data sets after cutting from different angles. The basic action includes slice, dice, roll up and drill down, pivot, etc.

By the connection with the MapInfo software, the system can realize the function of GIS visualization so as to give its customers a visual display on relevant contents.

Another application is the data mining based on data from the multi-dimensional database. Different from displaying historical data, data mining is inclined to automatically search for mode and useful information hidden in historical data, to mine knowledge from the the database, and put them into knowledge library. The results of OLAP multi-dimensional analysis can be used as the basis of data mining, and the data mining is a deeper level of knowledge discovery based on the multi-dimensional analysis.

IV. Illustration of the DSS Application

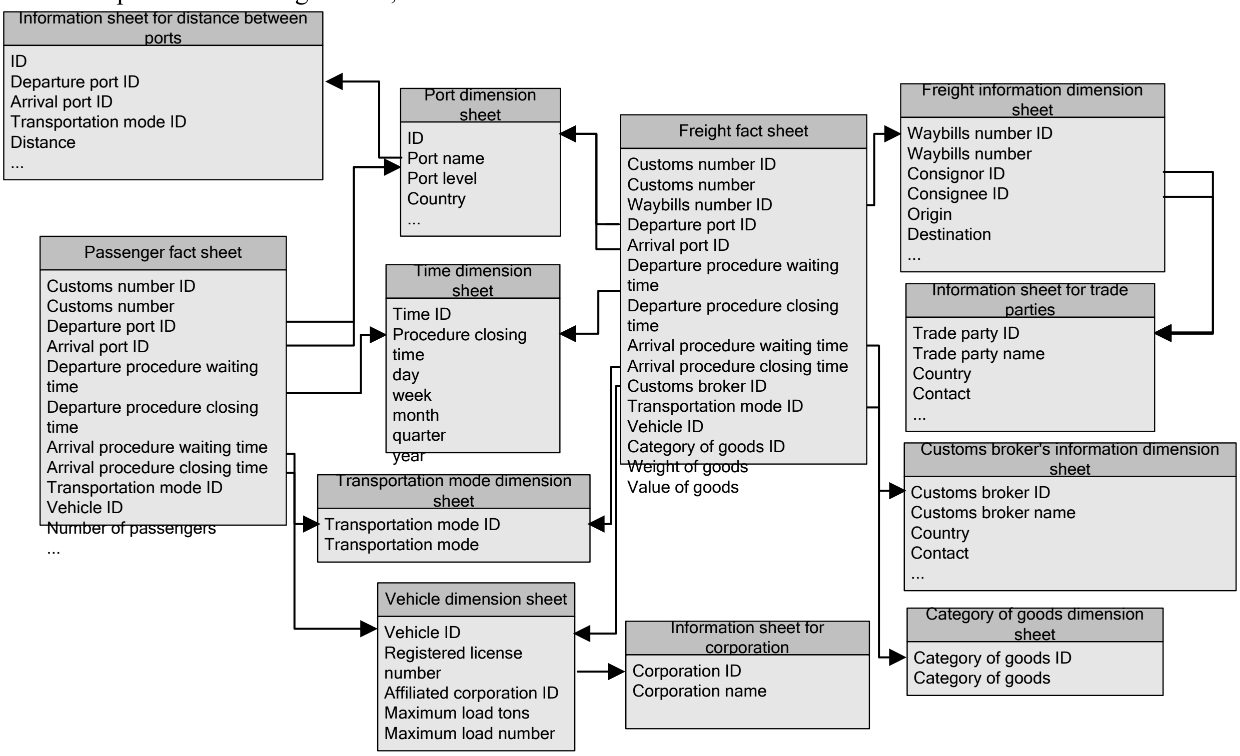


Figure 4. The structure of the constellation to construct the database for cross-border transportation decision support system

This paper will establish multi-dimensional set using MS SQL Server 2005, and design the application interface on Visual Studio 2005. Based on that, the decision support system for cross-border transportation can be realized. The following sections illustrate the application of the decision support system.

A. OLAP and multi-dimensional reports

For example, in order to inquire the transportation data of ports in GMS from the perspective of category and transportation mode, one just need to edit the fields required, and then the results will show up in the form of Figure 5.

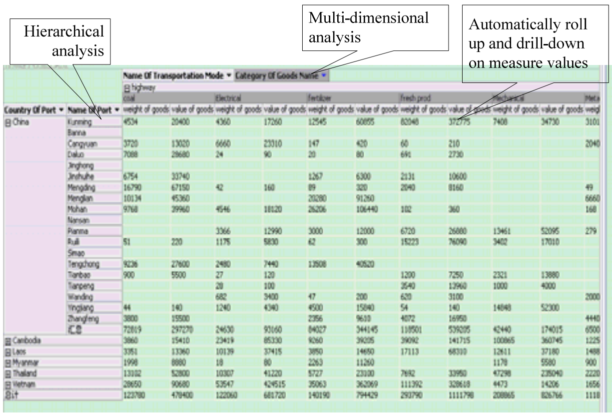


Figure 5. Illustration of the OLAP multi-dimensional analysis

As shown in Figure 5, once 'country of port' is unfolded, you can see that the weight and value of goods of Yunnan's 20 ports are displayed. The effect that Fig 3 shows must have the technical support of a DW system, as traditional database can't reach the purpose of multi-dimensional analysis well. By folding and unfolding the field selected, one can easily realize the operation of slice, dice, roll up and drill down, pivot, so that the target of specific data query is achieved.

B. Application of Model library and GIS visualization

For example, we want to make a planning for cross-border transportation channel. We use the fuzzy *c*-means cluster analysis method in the model library.

Fuzzy *c*-means cluster analysis method is a method that aims at minimizing the variance of all classes. Its mathematical model is as follows.

$$\min z(\tilde{U}) = \sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^m \| x_k - v_i \|^2 \tag{1}$$

$$s.t.\quad v_i = \frac{\sum_{k=1}^{n}(\mu_{ik})^m x_k}{\sum_{k=1}^{n}(\mu_{ik})^m} \qquad \forall i \tag{2}$$

$$\tilde{U} \in M_{fc} \tag{3}$$

Therein, the universal set $X = \{x_1, x_2, ..., x_k, ..., x_n\}$, the i th class is $\tilde{S}_i$, $i = 1, 2, ..., c$, μ_{ik} is the degree of membership that x_k belongs to the i th class $\tilde{S}_i$,

$$\mu_{ik} = \left(\frac{1}{\| x_k - v_i \|_G^2}\right)^{\frac{1}{m-1}} \Bigg/ \left(\sum_{j=1}^{c}\frac{1}{\| x_k - v_i \|_G^2}\right)^{\frac{1}{m-1}},$$

$$i = 1, 2, ..., c; k = 1, 2, ..., n \tag{4}$$

v_i is the cluster center of the i th class, M_{fc} ={fuzzy *c*-partition on all the X }, $m \geq 1$.

We choose seven indexes in the 20 port city as the fuzzy cluster indexes, which are GDP value of the region where the port city locates(hundred million RMB), fixed asset investment(hundred million RMB), all industrial output(hundred million RMB), passenger flow(ten thousand people), total cargo value(ten thousand RMB), communications environment coefficient, policy coefficient. Therein, the communications environment coefficient was calculated by conversion using a fixed proportion from numbers of highways, national main roads, provincial roads, mainline railways, regional railways and inland waterway capability. The policy coefficient was decided by the planning of Yunnan province to each region. In this research, the value of policy coefficient for the first-class ports was set 2, and 1 for the second-class ports.

The values for all these indexes (see table I.) are stored in the database. Using these data stored, we get the cluster results by computation as follows.

Class 1- Kunming;

Class 2- Ruili, Hekou, Mohan;

Class 3 - Wanding, Tengchong, Jinshuihe, Tianbao, Simao, Jinghong, Banna, Mengding Qingshuihe, Daluo;

Class 4 - Pianma, Yingjiang, Zhangfeng, Nansan, Menglian, Cangyuan, Tianpeng.

We can get conclusion from the cluster results that three cross-border transportation channel can be constructed. The three channels are centered by Kunming, distributed by cities that are along the highway, ended by border port cities that are class 2.

TABLE I.
THE FUZZY *C*-CLUSTER ANALYSIS INDEXES VALUE

Port	GDP	Fixed Asset Investment	All industrial output	Passenger flow
Kunming	1837.46	1600.66	632.36	53142
Ruili	115.71	100.92	25.94	808.88
Wanding	115.71	100.92	25.94	35.89
Hekou	560.88	404.53	244.4	342.58
Mohan	138.64	88.86	28.09	61.22
Jinshuihe	560.88	404.53	244.4	7.06
Tianbao	284.9	188.13	69.53	20.06
Simao	211.7	172.29	44.95	0.11
Jinghong	138.64	88.86	28.09	4.3
Banna	138.64	88.86	28.09	1.44
Tengchong	221.66	167.3	45.91	22.08
Mengding Qingshuihe	181.33	114.5	41.01	25.09
Daluo	138.64	88.86	28.09	46.42
Pianma	48.05	72.15	16.66	18.36
Yingjiang	115.71	100.92	25.94	104.61
Zhangfeng	115.71	100.92	25.94	49.23
Nansan	181.33	114.5	41.01	68.48
Menglian	211.7	172.29	44.95	47.32
Cangyuan	181.33	114.5	41.01	27.03
Tianpeng	284.9	188.13	69.53	8

Port	Total cargo value	Communications environment coefficient	Policy coefficient
Kunming	145148	42	2
Ruili	79956	40	2
Wanding	5042	38	2
Hekou	60653	42	2
Mohan	35910	42	2
Jinshuihe	500	36	2
Tianbao	12690	30	2
Simao	2228	29	2
Jinghong	20030	32	2
Banna	31419	32	2
Tengchong	5181	36	2
Mengding Qingshuihe	6401	34	2
Daluo	3872	30	2
Pianma	677	28	1
Yingjiang	9399	29	1
Zhangfeng	14277	25	1
Nansan	2661	26	1
Menglian	7360	34	1
Cangyuan	3553	27	1
Tianpeng	1628	25	1

The three channels can be visualized by connecting the system to the GIS software. (see Figure 6).

Figure 6. Yunnan cross-border transportation channel

After the computation, we want to choose a city from Kunming, Ruili, Hekou, Mohan to establish international logistics hub. We use the rate-of-flow method location model in the model library. The basic process is as shown in Figure 7.

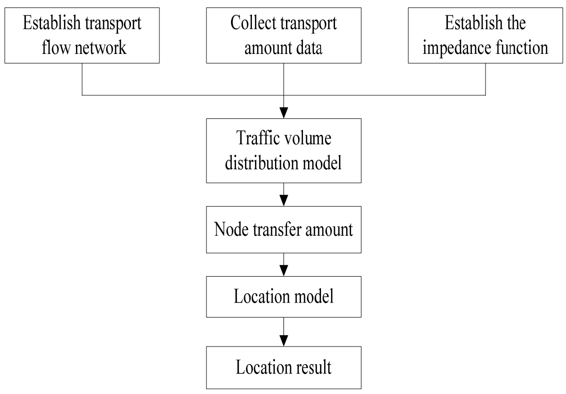


Figure 7. The flowchart of the rate-of-flow method model

During the process, the system is connected to the MapInfo software, and then the cross-border transportation network model of ports in Yunnan is constructed as Fig. 8. Data including transportation speed, cost, distance and capacity between each port of different transportation mode are included in this model.

The impedance function is as equation (5) shows.

$$c_i(x) = \gamma_d\left(k_i + op \cdot L_i\right) + \gamma_t t_i \left[1 + \alpha_i \left(\frac{x_i}{C_i}\right)^{\beta_i}\right] vot \tag{5}$$

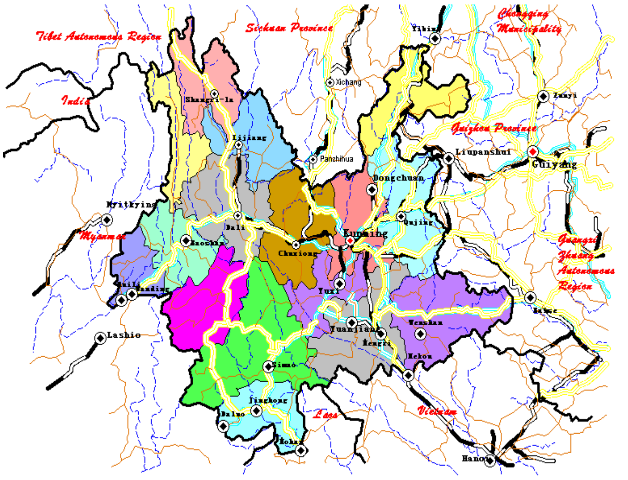


Figure 8. The network model of the cross-port transportation for Yunnan.

In the equation, $c_i(x)$ is the generalized cost for section i of a channel; γ_d is the weight of distance cost; γ_t is the weight of time cost; k_i means the fixed cost for section i; op stands for the operational cost of a unit length of channel; L_i is the distance of section i; t_i means the free transportation time on section i; C_i is the capacity of section i; vot is the constant for time value; α_i, β_i are the coefficients of the impedance function.

The traffic volume distribution model is shown in (6)-(9).

$$min \quad \sum \int_0^{x_i} c_i(x)dx \tag{6}$$

$$s.t. \quad \sum_k f_k^{rs} = q_{rs} \quad \forall r,s \tag{7}$$

$$x_i = \sum_r \sum_s \sum_k f_k^{rs} \delta_{i,k}^{rs} \quad \forall i \tag{8}$$

$$f_k^{rs} \geq 0 \quad \forall r,s \quad \forall k \tag{9}$$

In the traffic volume distribution model, x_i means the traffic volume on section i; $c_i(x)$ is the generalized cost for section i; f_k^{rs} means the traffic volume on the k th route where the origin is r and destination is s; C_k^{rs} is the generalized cost on the k th route where the origin is r and destination is s; $\delta_{i,k}^{rs}$ is a binary variable, when section i is on the k th route where the origin is r and destination is s, $\delta_{i,k}^{rs}$ equals 1, otherwise $\delta_{i,k}^{rs}$ equals 0.

Last, we use the rate-of-flow method location model, as in (10)-(12).

$$max \quad \sum_i \sum_j Q_{ij} x_i \tag{10}$$

$$s.t. \quad \sum_i x_i = 1 \tag{11}$$

$$Q_{ij} = q_{ijo} \tag{12}$$

Therein, Q_{ij} stands for the traffic flow of the i th candidate node on the j th channel; x_i is a binary variable, if the i th node is chosen as the hub, x_i equals 1, otherwise, it equals 0; q_{ijo} is the quantity of shipments that is assigned to the i th node on the j th channel.

We choose Kunming, Ruili, Hekou, Mohan as the candidate port for the international logistics hub. Using the model we proposed before, the annual transfer amount and the location result are illustrated as Tab II. Therein, 1 means chosen for hub, 0 means not.

TABLE II. THE ANNUAL AMOUNT OF TRANSFER FOR EACH CANDIDATE NODES AND THE LOCATION RESULTS

Candidate node	Kunming	Ruili	Hekou	Mohan
Transfer amount(ten thousand tons)	16,278	11,352	13,682	9,204
x_i	1	0	0	0

By using the rate-of-flow method location model, we propose choosing Kunming as the international logistics hub.

C. Data Mining

For example, we use the decision tree algorithm to do the data mining on the historical freight data in the Greater Mekong Subregion(GMS), so as to see whether there is any undiscovered information. We choose 'weight of goods' as the analysis object, and include transportation mode, category of goods etc. as possibly related columns. After processing, the dependency network is shown as Figure 9.

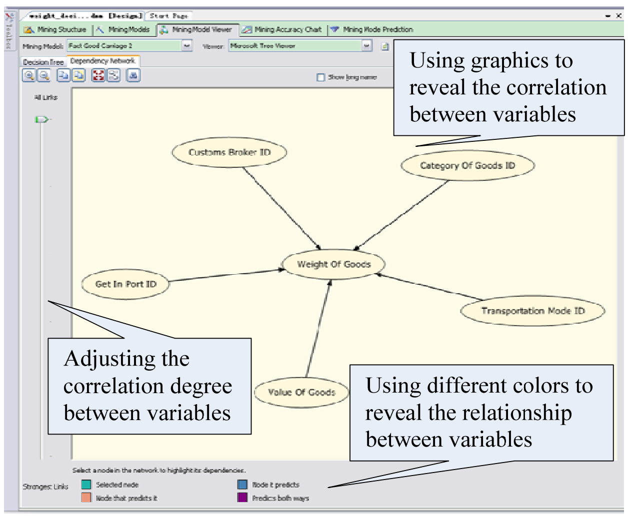


Figure 9. Illustration of DM using the decision-tree algorithm(1)

As Figure 9 shows, all of the relevant columns point to 'weight of goods'. That is to say, they will affect the 'weight of goods' to a certain extent. Then we drag the slider to the bottom, namely the strongest link, we get Figure 10. At this time, only the 'transportation mode ID' points to 'weight of goods'. This means that transportation mode has the greatest influence on traffic volume. This is quite consistent with our inspiration from the fact that the highway traffic volume takes up a great proportion of the total volume.

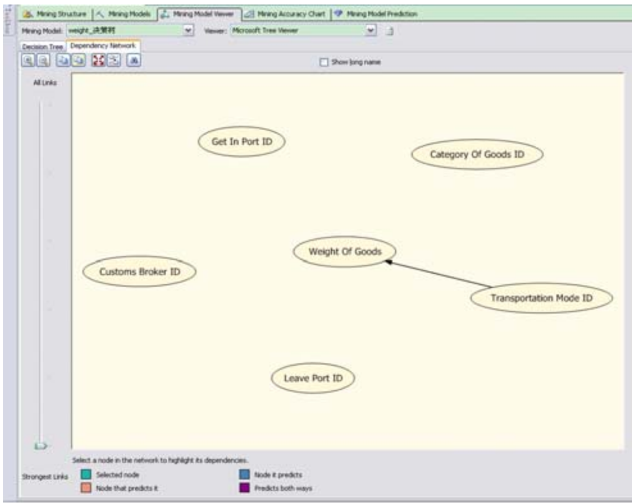


Figure 10. Illustration of DM using the decision-tree algorithm(2)

On the other hand, we also need to analyze the accuracy of DM by electronic interface. Thus, we need to use DM accuracy charts. There are two kinds of accuracy charts, one is lift chart, and the other one is classification matrix. Figure 11 is the display of lift chart for the DM accuracy analysis.

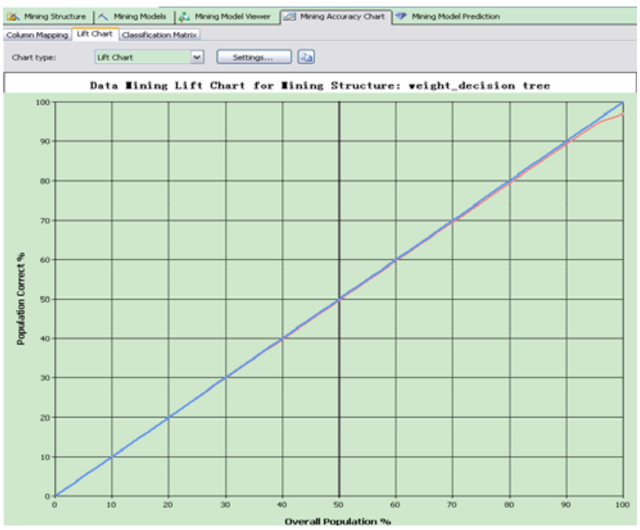
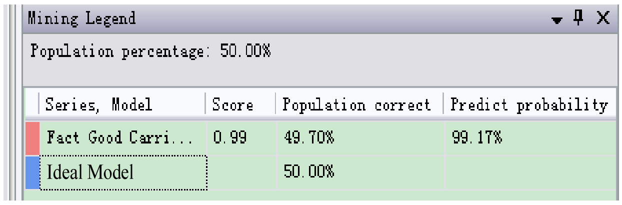


Figure 11. The mining accuracy chart for testing the data mining results (1)

As shown in Figure 11, the diagonal line represents results an ideal model can produce, with exactly perfect predictions; the curve line is the result for the Data Mining. The closer the two lines are, the closer the effect of Data Mining to the ideal model.


Mining Legend
Population percentage: 50.00%

Series, Model	Score	Population correct	Predict probability
Fact Good Carri...	0.99	49.70%	99.17%
Ideal Model		50.00%	

Figure 11. The mining accuracy chart for testing the data mining results(2)

Figure 12 shows the relationship of the two lines, and provides the degree of satisfactory for the mining results. In the illustration, the mining results obtain 99 points, which is a very high score.

V. Conclusion

Compared with traditional database, multi-dimensional database has more advantages as an emerging data processing technology. The main functions of that (that is OLAP and data mining) can greatly raise the ability for processing and application of data information. Based on the multi-dimensional database technology and Yunnan's participation in the GMS, this paper made a data requirement analysis on cross-border transportation, and designed a systematic structure for the decision support system for Yunnan cross-border transportation. The system can meet demands from different relative departments. Through the application of the system, more reliable and comprehensive decision support information can be provided. Application show proves the effectiveness of the system. However, the technology in China is still in the stage of research and preliminary application. Building a more perfect port cross-border decision support system need more theoretical support and practical experience.

Acknowledgment

This paper was partly supported by the National Nature Science Foundation of China (70572104); the Doctoral Foundation of Ministry of Education of China (20070151002); Transportation construction projects in Western China(2009318791007).

References

[1] Erik Thomsen. OLAP Solutions: Building multi-dimensional Information Systems . John Wiley & Sons, 2004.

[2] Brobst S, Rarey J. "The five stages of an active data warehouse evolution". *Teradata Magazine*. vol. 3, pp. 38-44, 2001.

[3] Shi Jingang et al. "QoS-based updates and queries scheduling in real-time data warehouse". *Journal of Chinese Computer Systems*. vol. 32, pp. 801-806, 2011.

[4] Lin Zhichao, Sun Lei. "Implementation of OLAP based on multi-dimensional database". *Computer Technology and Development*. vol. 20, pp. 75-82, 2010.
[5] Liu Jingrong. "Application of Data Warehouse technology in Chinese Railway Transportation". *Railyway transport and economy*. vol. 24, pp. 24-25, 2002.
[6] Meng Yanping et al. "Applying data warehousing technology to port logistics analysis". *Journal of Shanghai Maritime University*. vol. 29, pp. 64-69, 2008.
[7] Wang Yunpeng et al. "Logistics Decision Support System for Multimodal Transportation Based on Data Warehouse". *Journal of Jilin University of Technology(Natural Science Edition)*. vol. 35, pp. 641-645, 2005.
[8] Gao Yan et al. "The decision support system studies based on the data warehouse and data mining's accountant". *Henan Science*. vol. 29, pp. 211-213, 2011.
[9] Yang Shan. "Design and realization of data warehouse to life insurance company". *Software Guide*.vol. 10, pp. 129-131, 2011.
[10] Liu Zhaobin. "Design and implementation of airport cargo data warehousing system". *Computer Systems & Applications*. vol. 18, pp.105-107, 2009.
[11] Yuan Chen et al. "Sales data visualization analyzing system based on DW". Journal of Jilin University. Vol. 28, pp. 61-67, 2010.
[12] Chen Chen. "The Study on Application of Multi-dimensional Database in Container Multi-modal Transportation". Dalian Maritime University, 2010.
[13] Hugh J. Watson, Celia Fuller, Thilini Ariyachandra. "DW governance: best practices at Blue Cross and Blue Shield of North Carolina". *Decision Support Systems*. vol. 38, pp. 435-450, 2008.
[14] Lei Jian. "Research on application of DW system to the road passenger transportation". *Technoeconomics & Management Research*. 2, pp. 54-55, 2007.
[15] Wang Jia. "Design and application of integrated data warehouse in petroleum exploration and production". *Office Informatization*. vol. 5, pp. 15-22, 2011.

Jin Zhihong was born in Shenyang, China in 1963. He is a doctoral tutor in Dalian Maritime University, Dalian, China. He earned his doctor degree in Nagoya Institute of Technology in 2000. His research fields are logistics system optimization and simulation, supply chain management.

Xu Qi was born in Guilin, China in 1984. She is a doctor candidate on logistics engineering and management in Dalian Maritime University, Dalian, China. Her interests of research fields include logistics system optimization and simulation.