OLAP-based Analysis on Passengers' Bus Usage Patterns

G. Y. Kim, I. S. Jeoung, W. S. Cho, K. H. Lee, I. H. Cho, and Y. M. Lee

***Abstract*—** As traffic-related big data have long been accumulated through computer systems, a more scientific way to organize traffic service routes based on big data analysis becomes more necessary. In this context, many researches are being made on traffic data to find more values in them. In this research, the bus information system (BIS) data and traffic card data are collected to examine the usage patterns of the passengers. For the purpose of this research, a multidimensional analysis model was built herein, which made possible to identify passengers' bus transit patterns and examine them from various angles. The analysis results can be used to design or re-organize bus service routes in order to save transit and limited bus operational resources.

***Keywords*—** Big data, OLAP, Pattern Mining, Transit Patterns.

I. Introduction

As car ownership rate has increased every year in South Korea, public transportation use has reduced, causing more serious traffic congestion. The country's government has pursued policies to stimulate public transportation use continuously to deal with the rising demand for passenger car use. Bus is a major public transportation mode. The more bus service routes coincide with passengers, the more convenience the passengers would feel and try to use public transportation more frequently. Travel pattern refers to a travel flow of a passenger moving from a starting point to a destination. If this travel pattern is not well reflected in a bus route, the passenger would have to transit more because the person cannot go to his or her destination directly at once. By analyzing such travel patterns in transit cases, we can find out and realign less reflective service routes. Transportation card data contain every public transportation user's travel information who uses the card. And such data collection is easy and accurate as they are saved by systems.

G.Y. Kim is with the Dept. of MIS(BK Plus Team), Chungbuk National University, Cheong-ju, KOREA (e-mail wscho@chungbuk.ac.kr).

I. S. Jung is with the Dept. of MIS (BK Plus Team), Chungbuk National University, Cheong-ju, KOREA (e-mail: 517insun@gmail.com).

W. S. Cho is with the Dept. of MIS(BK Plus Team), Chungbuk National University, Cheong-ju, KOREA (e-mail: wscho63@gmail.com).

K. H. Lee is with the Dept. of Business Data Convergence, Chungbuk National University, Cheong-ju, KOREA (e-mail: lee.kyunghee@gmail.com).

I. H. Cho is with the Dept. Transportation Admin, Chungbuk Cheongju City Government, Cheong-ju, KOREA (e-mail: cih6138@gmail.com).

Y. M. Lee is with the Dept. Transportation Admin, Chungbuk Cheongju City Government, Cheong-ju, KOREA (e-mail: leeym@korea.kr).

Analyzing these transportation card use data would generate diversified and significant outcomes. To examine transit patterxs, we collected 5,350,000 sets of transportation card transaction data of using buses in one region for one month.

The materials were processed in line with this research purpose and a star schema-form of data warehouse was built. Based on the data warehouse, the OLAP was performed and transiting passengers' transit patterns were searched. This study identifies transit travel flows to look in the transit pattern between bus service lines. Also, multidimensional analysis on transit patterns is proposed herein as that can give multi-angled examination on transit. Such analytical outcomes herein can be useful for bus operators to reflect bus users' travel patterns in their bus service routes. By doing so, passengers will rely on more convenient public transportation services and operators will manage bus resources more efficiently, contributing to less traffic congestion.

The paper is organized as follows. Chapter 2 looks at researches on analytical techniques. Chapter 3 proposes an analytical system and explains analysis process. Chapter 4 examines analysis results based on the system built herein, and Chapter 5 discusses this research limitations or future study agenda and concludes.

II. Related Work

A. OLAP(Online Analytical Processing)

The online analytical processing (OLAP) is a real time analysis technique where an end user accesses multidimensional information directly to examine information in a conversation manner to use for his or her decision making. It is also called as the multidimensional data analysis. The OLAP is an idea as opposed to the online transaction processing (OLTP). When the data warehouse saves and manages data, the OLAP converts data of the data warehouse into a more strategic information. The OLAP system, depending upon ways to save multidimensional data, is classified into MOLAP(Multidimensional OLAP), ROLAP(Relational OLAP), and a combination of both called HOLAP (Hybrid OLAP)[4][11][12].

B. Data Warehouse

Data warehouse is a database where enterprises integrate and separately save large accumulated data (time variant) in a subject-oriented manner in line with users' perspectives and demand to support their decision making process. The phase of data warehouse data loading is one of the most essential steps in building a data warehouse. Operational data are not just simply

http://dx.doi.org/10.15242/IIE.E0214080

extracted and replicated into a warehouse but integrated based on an enterprise-wide model. This phase requires cleansing and transformation procedures. That is, duplicated data should be removed, wrong values have to be corrected and data in different formats have to be standardized in a unified form. If necessary, data are compiled and transformed through a calculation process[6].

III. Analysis System

In this chapter, we will propose an analysis system for traffic data examination and explain the analysis process thereof.

A. Transit sequence building

In the case of transit, bus get-on, get-off and transit data are generated under the identical card number. These three transactions of get-on, get-off and transit forms a set of single transit sequence for each transportation card number. Each transaction has information on bus stop and route in order of the bus stop to get on→ route of get-on → bus stop to get off → route of get-off → bus stop to transit → transit route. These 6 items structure a transit sequence in order. There must be all of the 6 items and only one single item change means different thing. BIS (Bus Information System) data and transportation card data were collected and refined. By combining the refined data, we process transit sequence database.

B. Data warehouse building

A data warehouse should be designed to draw analysis results in line with the analytical purpose. In this research, transit sequences were complied in a multidimensional manner to study transit patterns. That is, the purpose is to compile transiting passenger data in various dimensions. A transit sequence database was built for the multidimensional analysis of transit travel flows and based on this, a data warehouse in a star schema structure was established in this research. It is consisted of 5 dimensional tables and one fact table. The dimensional tables are on bus service route date, bus stop, bus service route, bus information and bus type.

C. Multidimensional analysis

For multidimensional analysis herein, we use MS SQL Server's SSAS (MS SQL Server 2012 Analysis Service) supporting the OLAP function. Figure 1 is a cube structure for multidimensional analysis connecting tables in the data warehouse.

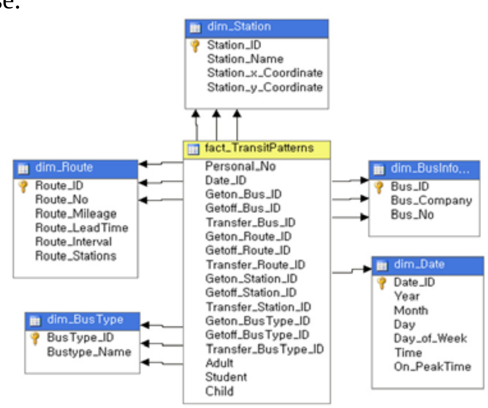

Fig. 1 OLAP cube structure of data warehouse

Based on this cube structure, the fact table values can be compiled and examined multidimensionally.

As a result of the analysis, it was found that the OLAP cube data can be brought to Excel by using Excel's function of 'get external data'. If necessary, the result data can be easily put into charts with the 'pivot table/pivot chart' function.

IV. Expreiments And Results

In this chapter, we will compile and analyze transit data in diverse dimensions and come up with transit patterns by using the data warehouse built based on transit sequence.

A. Frequency analysis for Bus Route dimension

Transit patterns can be found through analysis on route-dimensional frequencies. Table 1 shows 5 most frequent transit patterns found. The most frequently appearing transit pattern is to transit from bus no. 823 to no. 502.

TABLE I
10 MOST FREQUENT TRANSIT PATTERNS BEEN BUS ROUTES

Rank	Transit Patterns		
	Line no. to get on	Line no. to get off	Transit Line no.
1	823	823	502
2	502	502	105
3	105	105	105
4	105	105	502
5	832	832	105

Rush hour and closing hour means school/office-going hours and school/office-leaving hours, respectively. Time period-specific analysis is also possible on transit pattern frequency. Figure 2 is the result of peak time-specific analysis of the most frequent transit patters in Table 1. Transit from no. 823 to no. 502 occurs more during the morning peak time than the afternoon peak time. And of the morning peak time, from 9 to 10 am is found to be the busiest.

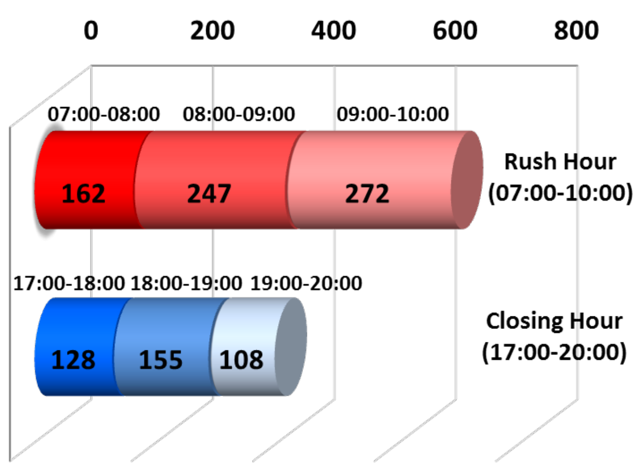

Fig. 2 Transit frequency in peak times

B. Frequency analysis for Bus Route, Station dimension

Most frequent place of transit can be found if the transit patterns are examined for frequency by considering the bus stop to get on in the bus stop-dimension table. The green dots in Fig.

3 represent the no. 823's bus stops. Red dots are no. 502's stops. The larger the green circles are, the more passengers transiting at the corresponding stops.

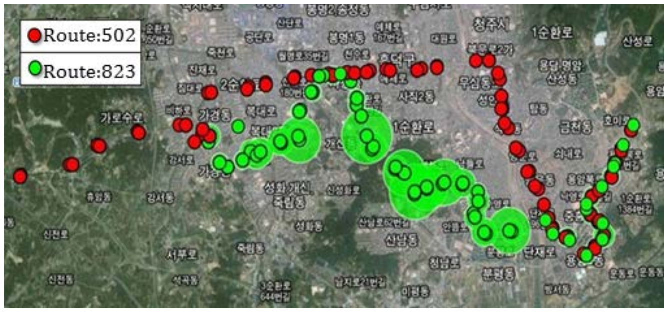

Fig. 3 Transit pattern frequency in bus stops to get on

Bus stops of frequent re-board (transit) can be found if transit patterns are examined for frequency in consideration of the transit bus stop in the bus stop-dimension. The larger the red circles marked on no.502 route are in Fig. 4, the more passengers transit at the corresponding stops.

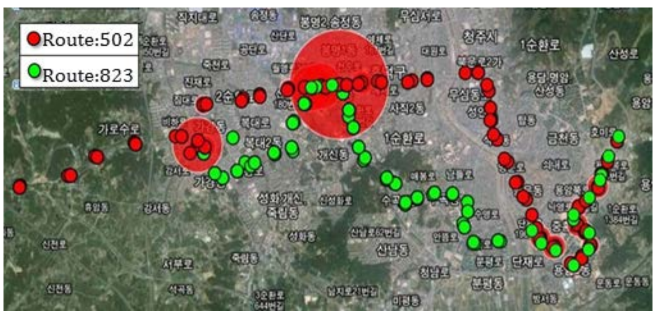

Fig. 4 Transit pattern frequency in transit bus stops

Figure 5 combines the two analysis outcomes together in a single picture. It is well shown that which bus stops are most often used to get on no. 823 to transit to no. 502 and in the same case, which stops are the busiest transiting place. In addition, the multidimensional analysis herein discovered that transit cases moving in the arrow-marked direction accounted for 83% of the entire transit between the two buses. This finding also indicates which direction the passengers wanted to move by boarding no. 823 then transiting to no. 502. Such an understanding would be helpful for bus route realignment.

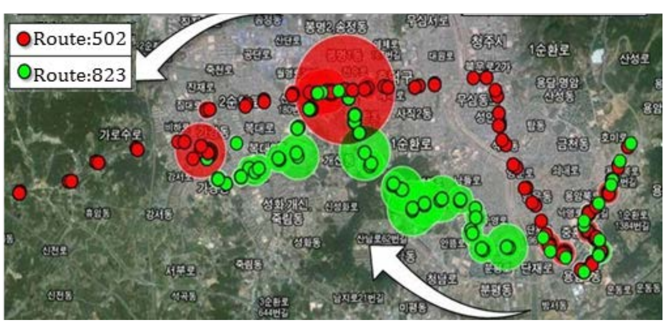

Fig. 5 Transit pattern frequency of bus stop of get-on and transit

V. Conclusions

In this research, transportation big data were analyzed not simply for transit volume calculation but for multi-angled analysis on transit. By doing so, this research has proposed a multidimensional analysis technique through which we can reflect passengers' travel patterns in bus route realignment, reducing transition and improving bus route service efficiency. As a result of this research examination, bus stops and routes with frequent transit were identified along with their characteristics. And *n* most frequent routes of transit were found and the transit patterns between them were further analyzed. This research analysis has found bus routes to be reorganized for reduced transit and enhanced route operational efficiency. Other more diverse analyses are also possible than that described herein. The analytical method in this research will be more useful as it helps users reach their desired level of multidimensional analytical conclusion based on their questions. The present research used data of only one month but future studies are expected to use more data for better outcomes with higher reliability.

References

[1] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
[2] J. Cho and I. Seo, "Implementation of the OLAP-based Subway Passenger Transit Pattern Analysis System," *Information Systems Review*, Vol.7, No.1, 2005.03.
[3] S. Whang and M. Cho, "Analysis of the Travel Characteristics of Transit Users in the Seoul Metropolitan Area Using Public Transportation Card Data," In *Proc. of Korean Society of Transportation*, vol.2011, No.3, 2011.
[4] J. Cho and S. Park, *OLAP Technology*, Sigma Insight, 2003.
[5] Nak-Guun Joo, Young-Gil Lim, "The Study on the Transfer Bus Stop Selection," Honam University Industrial Technology, vol.18 no.1 ,2012.06.
[6] O. Kwon, *OLAP Solutions + α SQL Server Analysis Services*, Daelim, 2001.
[7] S. Lee, *Estimation of Transfer Volume Between Public Transport Modes Using Traffic Card and BMS Data*, Ph. D. thesis, Keimyung University, 2012.06.
[8] J. S. Park and G. S. Lee., "Mining Trip Patterns in the Large Trip-Transaction Database and Analysis of Travel Behavior," *Journal of the Economic Geographical Society of Korea*, Vol.10, No.1 2007.
[9] J. H. Hwang, G. S. Kim, and J. H. Jeon, "Analysis of User's Travel Pattern and Bus Service Satisfaction Index for Public Transportation Reform in Daegu," *Journal of Korean Society of Transportation*, Vol.24 No.7 2006.
[10] C. M. Choi., *Design and implementation of Data WareHouse-based for the Efficient Analysis of Data*, Master thesis, Hanyang University, 2007.08.
[11] R. Kimball, "A Dimensional Modeling Manifesto," *DBMS*, 1997.
[12] S. Youness, *Professional Data Warehousing With SQL Server 7.0 and OLAP Services*, Wrox Press, Birmingham (UK), 2000.