

대용량 교통카드 트랜잭션 데이터베이스에서

통근 패턴 탐사

박종수[○]

성신여자대학교 IT학부

jpark@sungshin.ac.kr

Mining Commuter Patterns

from Large Smart Card Transaction Databases

Jong Soo Park[○]

School of Information Technology, Sungshin Women's University

수도권 대중교통 이용자는 2004년 서울시의 대중교통 체계 개편에 따라 교통 카드를 사용하여 버스 와 지하철을 이용하게 되었다. 교통 카드를 사용하는 각 승객의 승차와 하차에 관한 데이터가 하나의 트랜잭션으로 구성되고, 하루 천만 건 이상의 트랜잭션들로 구성된 대용량 교통카드 트랜잭션 데이터베이스가 만들어지고 있다. 대중교통을 이용하는 승객들의 승차와 하차에 관한 여러 정보를 담고 있는 교통카드 트랜잭션 데이터베이스에서 유용한 패턴이나 정보를 탐사해내는 연구가 계속 진행되고 있다. 이런 연구 결과는 수도권 대중교통 정책을 입안하는 데 중요한 기초 자료가 되고, 수도권 승객들에게 대중교통을 보다 잘 이용할 수 있는 정보로 제공된다. 교통카드 이용률은 2006년 79.5%, 2007년 80.3%, 2008년 81.6%로 점차적으로 증가하고 있다.

대용량의 교통카드 트랜잭션 데이터베이스에 대한 연구를 살펴보면, 하루 동안의 교통카드 트랜잭션 데이터베이스에서 순차 패턴을 탐사하는 알고리즘을 연구하였고[1], 승객들의 통행 패턴에 대한 분석연구를 확장하여 일 년에 하루씩 2004년에서 2006년까지 3일간의 교통카드 트랜잭션 데이터베이스로부터 승객 시퀀스의 평균 정류장 개수와 환승 횟수 등을 연도별로 비교하였다[2]. 수도권 지하철 시스템의 특성에 관한 연구로는 네트워크 구조 분석이 있었고[3], 승객의 기종점 통행 행렬(Origin-Destination trip matrix)에 의한 승객 흐름의 분포가 멱함수 법칙(power law)임을 보여주는 연구가 있었고[4], 지하철 교통망에서 모든 링크상의 승객들의 흐름을 찾아내는 연구가 있었다[5].

본 논문에서는 교통카드 트랜잭션 데이터베이스에서 지하철 승객들의 통근 패턴을 탐사해내는 방법을 연구하였다. 수도권 지하철 네트워크에 대한 정보를 입력하고, 하루치의 교통카드 트랜잭션 데이터베이스에 연구된 방법을 적용하여 8가지 통근 패턴들을 탐사해내고 분석하였다. 탐사된 패턴들 중에서 많은 승객들이 지지하는 출퇴근 패턴에 대해서는 시간대별로 승객수를 그래프로 보여주었다.

1. 통근 패턴

수도권 대중교통 이용자들 중에서 출·퇴근하는 통근자들의 하루 동안의 통근 패턴들 중에서 먼저 많은 승객들이 이용할 것이라고 예측되는 8가지 패턴들을 탐사하고, 이 패턴들의 각 경우에 대한 승객들의 수와 승·하차 시간을 교통카드 트랜잭션 데이터베이스에서 추출하도록 한다. 다음은 고려하고 있는 8가지 통근 패턴에 대한 설명이다. 여기서, a, b, c, d, y, z는 지하철역 이름을 나타내고, 화살표(\Rightarrow)는 하나의 트랜잭션을 나타낸다.

pattern 0: $a \Rightarrow b$ (출발지 \Rightarrow 도착지)

pattern 1: $a \Rightarrow b, b \Rightarrow a$ (출발지에서 도착지로 가서 다시 출발지로 되돌아오는 유형)

pattern 2: $a \Rightarrow b, b \Rightarrow c$ (출발지에서 도착지로 가서 다시 다른 곳으로 가는 유형)

pattern 3: $a \Rightarrow b, c \Rightarrow a$ (출발지에서 도착지로 가서 다른 곳에서 출발지로 되돌아오는 유형)

pattern 4: $a \Rightarrow b, c \Rightarrow d$ (출발지에서 도착지로 가서 다른 곳에서 또 다른 곳으로 가는 유형)

pattern 5: $a \Rightarrow b, \dots, b \Rightarrow a$ (처음 도착지에서 출발지로 되돌아오는 유형)

pattern 6: $a \Rightarrow b, \dots, y \Rightarrow a$ (처음 도착지와 다른 곳에서 출발지로 되돌아오는 유형)

pattern 7: $a \Rightarrow b, \dots, y \Rightarrow z$ (처음 출발지로 되돌아오지 않는 유형)

통근 패턴을 탐사하기 위해서는 먼저 지하철 네트워크를 구성하는 지하철역에 관한 정보를 읽고 그 다음에 지하철역 사이의 연결 정보를 읽는다. 지하철역에 해당되는 정점과 지하철 역 사이의 링크 자료구조를 해시 테이블로 저장하여 빠르게 원하는 데이터를 접근하도록 한다. 그런 후에 교통카드 트랜잭션 데이터베이스에서 트랜잭션을 하나씩 읽어서 카드번호가 동일한 한 승객의 트랜잭션들을 묶어내어 8가

지 패턴들 중의 하나로 분류하고 이 승객의 첫 번째 지하철과 마지막 지하철의 승차와 하차 시간을 얻어서 히스토그램으로 만든다.

2. 실험 결과 및 결론

제안된 방법을 사용하여 2005년 6월 24일 하루 동안 지하철과 버스를 이용한 승객들의 교통카드 트랜잭션 데이터베이스를 입력으로 하여 수도권 지하철 승객들의 8가지 통근 패턴들을 탐사하였다. 하루 동안 기록된 전체 트랜잭션들의 개수는 10,667,518개이고, 이 중에서 4,930,039개는 지하철을 사용한 승객들의 트랜잭션들이고 나머지는 버스를 이용한 승객들의 것이다. 지하철 승객 중에서도 통근 패턴에 기여한 트랜잭션들은 4,909,316개이고, 나머지는 트랜잭션 내용 중에서 정확한 속성을 가지고 있지 않은 경우 등으로 통근 패턴에 사용할 수 없는 것으로 간주되었다. 8가지 통근 패턴들 중의 하나로 분류된 트랜잭션들을 가진 승객들의 숫자는 2,746,517명으로 나타났다. 실험에서 승객들의 지하철 탑승회수, 8가지 패턴들의 분포 비율, 통근 패턴별로 시간에 따른 승객들의 숫자 등을 찾아내는 것이 이루어졌다. 그 중에서 표 1은 통근 패턴별 승객수를 나타내고, 그림 1은 주로 출·퇴근을 하는 승객들의 대표적인 패턴인 pattern 1의 시간대별 승객수를 보여준다. 첫 번째 지하철 승·하차시간은 출근 시간대인 오전 8시경이 최고치이고, 두 번째 지하철 승차는 퇴근 시간대인 오후 7시경에 최고치를 보여준다.

본 연구에서 지하철 승객들의 통근 패턴을 8가지로 나누어 각 패턴별 사용자들의 승객수를 시간대별로 찾아서 그래프로 보여주어 지하철 승객들의 통근 패턴을 이해할 수 있게 하고 이를 기반으로 수도권 교통 정책을 입안할 때 중요한 참고 자료로 사용할 수 있다. 앞으로 연구과제는 수도권 대중교통 이용자인 버스 승객, 지하철 승객, 환승 승객들에 대한 전체 승객들의 통근 패턴을 탐사하는 것이다.

표 1 : 지하철 통근 패턴별 승객수

통근패턴	승객수	비율(%)
pattern 0	1,032,187	37.5817
pattern 1	748,883	27.2666
pattern 2	352,991	12.8523
pattern 3	151,874	5.5296
pattern 4	132,895	4.8386
pattern 5	23,052	0.8393
pattern 6	137,475	5.0054
pattern 7	167,160	6.0862

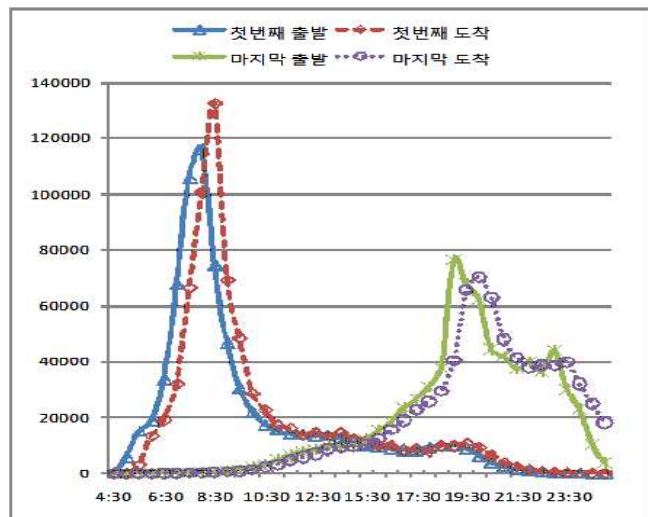


그림 1 : pattern 1의 경우 시간에 따른 승객수

참고문헌

[1] K. Lee and J.S. Park, "Traversal pattern analysis of transit users in the Metropolitan Seoul", *Proceedings of International Forum on the Public Transportation Reform in Seoul*, Seoul, July 7-8, 2005.

[2] J.S. Park and K. 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, pp.44-63, Mar. 2007. (in Korean)

[3] J.S. Park and K. Lee, "Network Structures of the Metropolitan Seoul Subway Systems", *Journal of the Economic Geographical Society of Korea*, vol.11, no.3, pp.459-475, Sep. 2008. (in Korean)

[4] K. Lee, W.-S. Jung, J.S. Park, and M.Y. Choi, "Statistical analysis of the Metropolitan Seoul Subway System: Network structure and passenger flows", *Physica A: Statistical Mechanics and its Applications*, vol.387, iss.24, pp. 6231-6234, Oct. 2008.

[5] J.S. Park and K. Lee, "Analysis of Passenger Flows in the Subway Transportation Network of the Metropolitan Seoul", *Journal of KIISE : Computing Practices and Letters*, 16 (3), pp. 316-323, 2010. (in Korean)