

# 서울 수도권 지하철 교통망에서 승객 흐름의 분석

(Analysis of Passenger Flows in the Subway Transportation  
Network of the Metropolitan Seoul)

박 종 수 †

이 금 숙 ††

(Jong Soo Park)

(Keumsook Lee)

**요 약** 서울 수도권의 지하철 교통망에서 승객들의 흐름을 찾아내는 방법을 제안하고 지하철의 주요 링크상의 승객 흐름을 분석한다. 교통망은 정점(vertex), 간선(edge), 그리고 승객 흐름으로 구성된다. 정점은 각 지하철역을 표시하고, 간선은 역과 역 사이를 연결한 지하철 링크를 표시하고, 승객이 승차역에서 하차역까지 최단 경로로 이동하면서 통과된 간선 위에 승객 흐름이 만들어진다. 본 논문에서는 지하철 승객들의 대용량 교통카드 트랜잭션 데이터베이스로부터 지하철 교통망의 각 링크의 흐름을 계산해내는 새로운 알고리즘을 개발하였다. 500만 건 이상의 지하철 승객의 교통카드 트랜잭션에서 제안된 알고리즘으로 승객 흐름을 찾아내어서, 4개의 주요 지하철 링크상의 통행 행태는 승객 흐름으로 분석되었고 모든 링크들 중에서 상위 10개 흐름을 테이블로 설명하였다.

**키워드** : 트랜잭션 데이터베이스, 통행 분석, 지하철 교통망, 승객 흐름, 최단 경로

**Abstract** We propose a method to find flows of transit users in the subway transportation network of the metropolitan Seoul and analyze the passenger flows on some central links of the network. The transportation network consists of vertices for subway stops, edges for links between two adjacent subway stops, and flows on the edges. Each subway transit user makes a passenger flow along edges of the shortest path from the origin stop to the destination stop in his trip. In this paper, we have developed a new algorithm to find the passenger flow of each link in the subway network from a large trip-transaction database of subway transit users. We have applied the algorithm to find the passenger flows from one day database of about 5 million transactions by the subway transit users. As results of the experiments, the travel behavior on 4 central subway links is analyzed in passenger flows and top 10 flows among all subway links are explained in a table.

**Key words** : Transaction database, Trip analysis, Subway transportation network, Passenger flow, Shortest path

· 본 연구는 2008학년도 성신여자대학교 학술연구조성비 지원에 의하여 연구되었음

† 종신회원 : 성신여자대학교 IT학부 교수  
jpark@sungshin.ac.kr

†† 비회원 : 성신여자대학교 지리학과 교수  
kslee@sungshin.ac.kr

논문접수 : 2009년 9월 1일

심사완료 : 2009년 11월 18일

Copyright©2010 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨팅의 실제 및 레터 제16권 제3호(2010.3)

## 1. 서론

서울시에서는 2004년 대중교통 체계를 개편하면서 대중교통을 이용할 수 있는 교통카드를 도입하여 활성화시키고 있다. 교통카드를 사용하는 대중교통 이용자들이 대량 통행 자료는 수도권 지역의 대중교통 이용자들이 움직이면서 생성하고 있는 교통흐름에 대한 데이터로서 하루 천만 건이 넘는 통행거리에 대한 실제 통행 자료가 생성되고 있다. 교통카드 트랜잭션에는 각 통행자의 통행에 대한 출발지점과 목적지, 이용 교통수단, 환승에 대한 위치와 시간 등에 대한 정확한 정보가 담겨있다. 실제 통행 자료가 대용량 데이터베이스로 관리되는 것은 세계에서 처음 있는 것으로 교통 지리 분야나 대중

교통 정책 분야 등에서 이를 활용하고자 연구를 수행하고 있다[1-6]. 본 논문에서는 대용량 교통카드 트랜잭션 데이터베이스에서 유용한 정보의 한 종류인 서울 수도권 지하철망의 각 링크의 시간대별 승객 흐름을 찾아내는 방법론을 처음으로 제안한다. 각 링크의 시간대별 승객 흐름에 대한 정보는 수도권 승객들의 실제 이동 상황을 설명하고 있기 때문에 수도권 교통정책 수립과 관련된 분야에 활용될 수 있다.

서울시를 포함한 수도권에서 대한민국 인구의 반 정도가 생활하면서 많은 교통흐름이 발생되고 있다. 수도권의 교통량 중에서 대중교통인 버스와 지하철의 승객들이 차지하고 있는 비율은 60%를 넘어서고 있고, 지하철 분담률은 36% 정도가 된다[2]. 2009년 8월 기준으로 수도권 지하철은 1호선에서 9호선까지, 분당선, 인천1호선, 중앙선으로 12개 노선이 연계 운영되고 있다[7]. 서울시 시정소식에 나타난 서울 메트로(지하철 1-4호선) 수송실적을 열람해보면, 서울 메τρό를 이용한 승객들의 교통카드 이용률은 2006년 79.5%, 2007년 80.3%, 2008년 81.6%로 계속 증가하고 있어서[8], 교통카드 트랜잭션 데이터베이스에 대한 여러 분석은 교통흐름의 개선에 크게 기여할 것이다.

지리학의 교통 분야에서 통행 흐름에 관한 연구는 주로 승객의 기종점 통행 행렬(origin-destination trip matrix) 분석을 수행하였다[6,9]. 출발역과 도착역에 관한 승객들의 이동 상황을 분석하는 것이다. 예를 들면, 그림 1은 2호선 강남역과 인접한 지하철역 사이의 통행 흐름을 보여주고 있다. 서울시 시정소식에 의하면 강남역은 2008년 승객 수송 순위가 1위인 역이다[8]. 그림에서 O-D 분석은 출발역의 점선으로 표시된 승차승객의 수와 도착역의 하차승객의 수에 대한 분석이 주를 이루었다. O-D 분석에서 간과하는 것은 강남역에 연결된 링크를 통과하는 승객들의 흐름에 대한 정보는 얻어낼 수 없다는 것이다. 본 논문에서는 역과 역사이의 링크를 지나가는 승객들의 흐름을 찾아내는 새로운 방법론을 제시하고, 교통카드 트랜잭션 데이터베이스에 알고리즘을 적용하여 수도권 지하철 교통망(subway transportation network)의 각 링크에서 승객 흐름을 계산한다. 그림 1에서 실선으로 그려진 강남역에서 교대역으로 가는 링크의 승객 흐름을 시간대별로 계산하고, 반대 방향으로 교대역에서 강남역 방향의 승객 흐름도 계산한다. 이 방법으로 수도권 지하철 교통망의 각 링크에 대한 승객 흐름을 분석할 수 있으므로 지하철 시스템의 유지 및 보수에 이용할 수 있다. 그리고 이 결과를 바탕으로 지리정보체계의 데이터베이스와 결합하여 효과적으로 분석하면 수도권지역에서 지역 간의 실질적인 기능적 연계 및 도시의 공간구조를 분석할 수 있고, 그 결과는

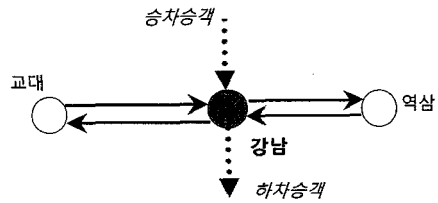


그림 1 2호선 강남역과 인접 지하철역 사이의 승객 흐름

주책정책이나 토지이용 및 시설 입지 등 각종 도시계획과 공간 계획에 기초 자료로 이용될 수도 있다[4].

본 논문의 구성은 다음과 같다. 2장에서는 지하철 네트워크에 관한 국내의 연구를 살펴보고, 3장에서는 지하철 교통망의 구성 요소들의 표현 방법과 교통카드 트랜잭션 데이터베이스에 대해 설명한다. 4장에서는 지하철 역 사이의 링크 상의 승객 흐름을 찾는 알고리즘을 설명하고, 5장에서는 실험 결과를 분석하고, 마지막으로 결론을 맺는다.

## 2. 관련 연구

대용량의 교통카드 트랜잭션 데이터베이스에서 유용한 정보를 탐사하는 첫 번째 연구는 하루 동안의 교통카드 트랜잭션 데이터베이스에서 순차 패턴을 탐사하여 승객들의 통행 패턴을 분석하였다[1]. 이 연구의 확장으로 일 년에 하루씩 2004년에서 2006년까지 3일간의 교통카드 트랜잭션 데이터베이스로부터 승객 시퀀스의 평균 정류장 개수와 환승 횟수 등을 년도별로 비교하고 시점별 통행 수요와 통행 패턴에 나타나는 공간적 특징을 연구하였다[3]. 수도권 지하철 시스템의 특성에 관한 연구로는 네트워크 구조 분석이 있고[5], O-D 통행에 의한 승객 흐름의 분포가 멱차수 법칙(power law)임을 보여주는 연구가 있다[6]. 다른 국가에서의 연구로는 보스톤 지하철 시스템과 인도의 철도 네트워크의 구조가 네트워크의 효율성(efficiency)과 클러스터링 계수를 계산한 결과로 좁은 세상(small world)임을 보여주는 것이 있다[10,11]. 지하철 시스템을 가지고 있는 세계 여러 도시에서 대중교통의 흐름을 개선하기 위하여 지하철 시스템과 버스 시스템을 연계한 교통 시스템의 네트워크 구조 분석에 많은 연구가 이루어지고 있다[12].

## 3. 지하철 교통망과 교통카드 트랜잭션 데이터베이스

수도권 지하철 교통망에서 각 링크의 승객 흐름을 분석하기 위해서, 먼저 지하철 시스템을 그래프로 표현해야 하고, 그 다음으로 교통카드 트랜잭션으로 승객의 출발지와 도착지를 연결하여 시간별로 각 링크에 승객 흐

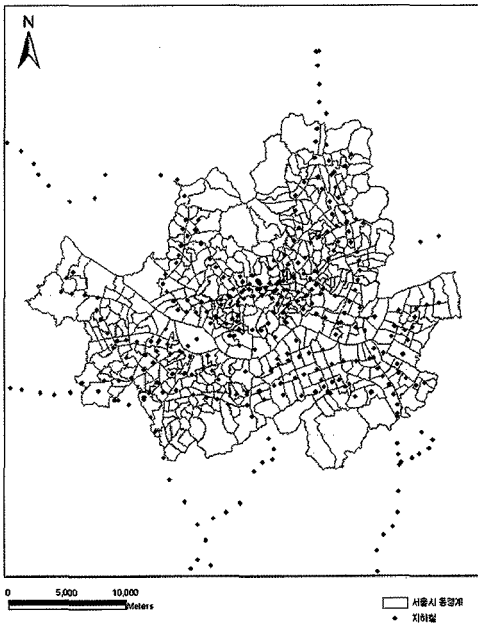


그림 2 수도권 지하철역의 위치도

름의 양을 계산한다. 수도권 지하철 시스템은 1호선을 시작으로 점진적으로 노선이 확대되어 왔다. 그림 2는 GIS(지리정보시스템)를 적용하여 수도권 지하철역의 위치를 보여주고 있다[5]. 대부분의 지하철역들은 같은 노선의 이전 역과 다음 역으로 연결되고 있지만, 나머지 일부 역들은 서로 다른 노선으로 환승할 수 있는 환승역들이 있다. 각 지하철역을 정점(vertex)으로 나타내고 역과 역 사이를 연결한 지하철 선로를 간선(edge) 또는 링크(link)로 나타내면, 수도권 지하철 시스템은 정점 집합과 간선 집합을 가지는 그래프로 표현해낼 수 있다. 이 그래프에서 승객들이 승차하여 지하철역들을 통과하는 이동 흐름을 간선에 나타내면 지하철 교통망이 구성된다.

지하철역을 정점으로 표현할 때, 환승역이 아닌 경우에는 한 정점으로 표현한다. 그렇지만, 환승역인 경우에는 어떻게 표현하느냐에 따라 정점을 표현하는 방법이 다르게 된다. 한 예로, 그림 3은 지하철 1호선, 3호선, 5호선이 만나는 종로3가역을 나타내고 있다. 그림 3에서 회색 칠이 된 원이 각 노선별 종로3가역을 나타내고 있다. 승객들은 한 노선의 종로3가역에서 내려서 다른 노선의 지하철로 갈아타서 원하는 도착역까지 갈 수 있다. 점선으로 된 원형 안의 지하철역은 지하도로 연결되어 있어 승객들이 환승하는 데 일정한 보행 시간을 필요로 한다. 서울 메트로 교통 센터에서 제공하는 노선도에서는 종로3가역을 하나의 역으로 안내하고 있다[7].

수도권 지하철 교통망을 두 종류의 네트워크로 모델

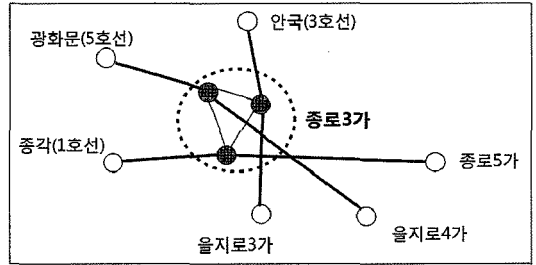


그림 3 세 노선들의 환승역인 종로3가역

링할 수 있다. 첫 번째로, 통합된 네트워크의 그래프에서는 환승역을 하나의 정점으로 나타내는 것이고, 두 번째로 환승역을 노선별로 각 지하철역을 표현하여 노선의 개수만큼의 정점들로 표시하는 분리된 네트워크의 그래프로 나타내는 것이다. 그림 3에서 종로3가역은 통합된 네트워크에서는 점선으로 된 원형처럼 한 개의 정점으로 나타내고, 분리된 네트워크에서는 노선별 지하철역에 해당하는 3개의 정점들과 이것들을 연결하는 3개의 간선들로 표현한다. 내부의 노선별 지하철역의 연결 표시인 가는 실선은 승객이 도보로 이동하여 다른 지하철 노선의 지하철역으로 환승하는 것을 나타낸다. 교통카드 트랜잭션의 정보는 한 승객의 승차역과 하차역의 자료만 제공되고 있다. 승객 흐름을 알기 위해 지하철 시스템에서 그 승객의 최단 경로를 찾아야 되는데, 통합된 네트워크에서는 널리 알려진 최단 경로 알고리즘으로 쉽게 정확한 최단 경로를 찾아낼 수 없지만, 분리된 네트워크에서는 가는 실선상의 환승시간이 포함되어 있어서 원하는 최단 시간 경로를 쉽게 구할 수 있다. 본 논문에서는 일반적인 안내표시와 최단 경로를 위해 분리된 네트워크와 통합된 네트워크에 대한 자료구조를 동시에 구축하고 있다.

수도권 지역의 승객이 지하철을 이용하면서 교통카드를 사용하는 경우에 (주)한국스마트카드(KSCC)에서 요금을 정산한다. KSCC에서 관리하는 트랜잭션은 승객의 승차와 하차에 관련된 여러 가지 속성들을 가지고 있다. 표 1은 교통카드 트랜잭션의 속성들의 일부와 실제 예제를 보여주고 있다.

표 1에서 첫 번째 행의 값은 교통 카드 트랜잭션의 속성들의 일부를 보여주고 있다. 나머지는 승객이 사용한 교통카드의 트랜잭션들을 나타내고 있다. 트랜잭션 예제 1과 2는 한 승객이 버스를 타고 내리고 난 후에 다시 5호선 지하철역인 발산역에서 승차하여 4호선인 평촌역에서 하차하는 환승 승객에 대한 자료를 보여주고 있고, 트랜잭션 예제 3과 4는 어떤 지하철 승객이 아침 5시 18분 46초에 2호선 홍대역에서 승차하여 같은 노선의 을지로입구역에서 하차하고 저녁 시간대인 오후

표 1 교통 카드 트랜잭션의 속성과 예제(2007년 기준)

속성(attribute)	교통카드번호, 승차일시, 승차역ID, 트랜잭션ID, 환승횟수, 교통수단 코드, 버스노선ID, 차량ID, 운행출발일시, 교통카드사용자구분코드, 승차교통사업자ID, 하차역ID, 하차일시, 승객수, 승차금액, 승차위반 금액, 하차금액, 하차위반금액
트랜잭션 예제 1	656, 20070516105306, 0032043, 012, 0, 105, 11111000, 111755380, 20070516094225, 01, 111520170, 0032043, 20070516105524, 1, 600, 0, 0
트랜잭션 예제 2	656, 20070516110009, 2516, 012, 1, 200, . . . , 01, 211200000, 1456, , 20070516121121, 1, 300, 0, 500, 0
트랜잭션 예제 3	658, 20070516051846, 0239, 015, 0, 200, . . . , 01, 211000000, 0202, , 20070516054159, 1, 900, 0, 0, 0
트랜잭션 예제 4	658, 20070516200409, 0424, 016, 0, 200, . . . , 01, 211000000, 2622, , 20070516202959, 1, 900, 0, 0, 0

8시 4분 9초에 4호선 명동역에서 승차하여 8시 29분 59초에 6호선의 망원역에서 하차하는 자료를 보여주고 있다. 트랜잭션 예제 1과 2는 교통카드번호 656번과 트랜잭션ID 012번으로 같고, 예제 3과 4는 같은 교통카드번호 658번과 다른 트랜잭션ID인 015번과 016번을 보여주고 있다. 예제 1과 2는 버스를 사용하고 난 후에 지하철을 사용하여 통합 환승 요금이 적용되고, 예제 3과 4와 같이 버스를 환승하지 않으면 독립적인 트랜잭션으로 간주되어 요금이 한 트랜잭션 단위로 부과된다. 본 논문에서는 지하철을 사용하는 승객들만의 트랜잭션을 추출해내어 시간과 승차차역에 관한 데이터를 이용하기 때문에, 교통카드 트랜잭션 데이터로부터 사용되는 속성은 승차일시, 승차역ID, 교통수단코드, 하차역ID, 하차일시에 대한 것이다.

#### 4. 승객 흐름 알고리즘

그림 4는 교통카드 트랜잭션 데이터베이스에서 수도권 지하철 교통망의 각 링크의 승객 흐름을 찾는 알고리즘을 간단히 설명하고 있다. 알고리즘은 지하철 교통망에 대한 자료구조를 구축하는 부분과 교통카드 트랜잭션 데이터베이스에서 지하철 승객의 트랜잭션들을 하나씩 읽어서 승객 흐름을 설정하는 부분으로 구성되어 있다. 다음 첫 번째 절에서는 교통망을 구성하는 기본 그래프에 대한 자료구조를 상세히 설명하고, 그 다음 절에서는 지하철 승객 트랜잭션으로부터 얻어진 승차역과 하차역의 정보로 승객이 통과하는 지하철 교통망의 최단경로상의 링크에 대해 상세히 설명한다.

##### 4.1 지하철 교통망 구성

승객 흐름을 찾아내려면 앞에서 언급한 두 종류의 지

```

FindPassengerFlowsFromSmartCardTransactions()
{
    BuildSplitSubwayNetwork();
    BuildMergedSubwayNetwork();
    while (smart_card_transaction in the_transaction_file)
        SettingPassengerFlow(smart_card_transaction);
}
    
```

그림 4 승객 흐름 알고리즘

하철 네트워크에 대한 정보를 가지고 있어야 한다. 일반 사용자에게 표시되는 통합된 네트워크와 승객의 최단 이동 경로를 찾기 위한 분리된 네트워크에 대한 정점과 간선들에 대한 자료구조를 만들어야 한다. 지하철역과 연결 상황에 대한 입력으로 먼저 분리된 네트워크에 대한 자료구조를 구축한다. 지하철역에 대한 입력인 지하철역ID, 이름, 노선번호, 위도, 경도 등을 입력받아서 그림 5의 지하철역 배열에 저장한다. 그림 5는 링크의 승객 흐름을 계산하기 위하여 필요한 지하철 교통망의 기본 자료구조를 설명한다. 지하철역ID로 배열에 빠르게 접근하기 위하여 지하철역ID 해시테이블을 유지한다. 지하철역은 정점에 해당되고, 이 정점들의 링크 정보를 갖는 간선에 대한 데이터를 입력받아서 지하철역ID 인접행렬(adjacency matrix)을 만든다. 간선의 입력으로는 출발역ID, 도착역ID, 간선의 가중치로 이루어지고, 지하철역ID 해시테이블에서 인접행렬의 인덱스를 구한다. 지하철역ID 인접행렬은 두 정점 사이에 간선의 유무, 두 정점 사이의 실제 거리, 두 정점 사이의 이동 시간을 포함한다. 분리된 네트워크에서는 환승역 내부의 노선별 지하철역 사이의 보행 시간도 포함한다. 한 지하철역에서 다른 지하철역 사이의 최단경로를 구하기 위하여, 이

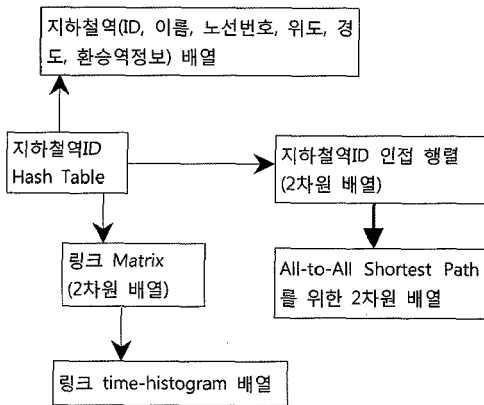


그림 5 지하철 교통망의 기본 자료 구조

동 시간 가중치 인접행렬에서 Floyd shortest path 알고리즘[13]을 적용하여 All-to-All 최단경로를 구하고, 각 최단경로의 중간 통과 지하철역들에 대한 정보를 유지한다.

다음으로 통합된 네트워크에 대한 지하철역ID 인접행렬을 만든다. 각 환승역은 지하철 노선 번호가 가장 작은 지하철역ID로 나타낸다. 통합된 네트워크의 지하철역ID 인접행렬을 기초로 하여 한 간선의 승객 흐름에 대한 time histogram 배열을 가리키는 링크 matrix를 관리한다. 예를 들어 설명하면, 그림 1에서 교대역에서 강남역으로 향할 링크의 1차원 time histogram을 접근하기 위해서 교대역ID와 강남역ID를 입력으로 하여 지하철역ID 해시 테이블에서 지하철역 배열의 인덱스를 찾는다. 두 개의 인덱스로 링크 matrix의 2차원 배열의 값은 교대→강남 링크를 나타내는 링크 time histogram 배열을 가리킨다. 타임 히스토그램은 단위 시간대별로 하루 동안의 승객들의 흐름의 양을 표시한다.

4.2 최단 경로 상의 승객 흐름

그림 6은 한 승객의 교통 카드 트랜잭션에서 그 승객이 통행한 각 링크에 승객 흐름을 추가하는 과정을 설명하고 있다. 교통카드 트랜잭션 데이터베이스에서 각 트랜잭션을 읽어서 지하철을 이용하는 고객이면 승차와 하차에 대한 데이터인 지하철역 ID와 승차 시간을 추출해낸다. 분리된 지하철 네트워크에서 그림 5의 All-to-All shortest path를 위한 2차원 배열에서 승차역과 하차역사이의 중간 통과역들로 이루어진 최단 경로를 찾아낸다. 그 다음으로 찾아진 최단 경로를 통합된 네트워크의 경로로 변환한다. 변환된 최단 경로의 간선에서의 통과 시간은 먼저 단위 길이 당 이동시간을 계산해내고 그 다음에 경로의 각 간선의 길이에 비례하여 계산해낸다. 최단 경로에 환승역이 있으면 보행 환승 시간을 배정하고 나머지 간선에 대해서는 거리에 따른 이동

```

SettingPassengerFlow(smart_card_transaction)
{
    1. 지하철 승객의 트랜잭션에서 승차역ID, 하차역ID, 승차시간, 하차 시간을 추출한다.
    2. 승차역ID와 하차역ID를 입력으로 지하철역ID Hash Table에서 지하철역의 정보를 가진 배열의 indices를 분리된 네트워크에서 결정한다.
    3. 찾은 indices로 All-to-All Shortest Path를 위한 2차원 배열을 이용하여 그 승객의 최단 경로 통과역 sequence를 얻는다.
    4. 분리된 네트워크에서의 통과역 sequence를 통합된 네트워크의 통행 sequence로 변환한다.
    5. 승차역에서 하차역 사이에 있는 통행 sequence의 각 link에 통과 시각에 해당되는 histogram의 slot에 값 1을 증가시킨다.
}
    
```

그림 6 교통 카드 트랜잭션에서 통과 링크의 승객 흐름 설정

시간을 고려하여 각 링크의 흐름에 한 승객의 이동을 시간 히스토그램에 기록한다. 각 링크의 시간 히스토그램은 오전 5시에서 다음날 오전 1시까지 1분 단위의 슬롯들로 만들어진다. 모든 지하철 승객들의 트랜잭션들을 처리하면 수도권 지하철 교통망의 각 링크의 승객 흐름은 분 단위의 시간으로 히스토그램에 기록된다.

5. 실험결과

이 장에서는 입력으로 사용되는 지하철 교통망의 범위, 교통카드 트랜잭션 데이터베이스의 크기, 실험에 사용된 컴퓨터 제원에 대해 먼저 설명하고, 그리고 제안된 방법으로 지하철 교통망의 승객 흐름을 구한 결과를 설명하고 분석한다.

지하철 시스템의 교통망의 기준 시점은 2007년 5월이고, 입력 데이터는 2007년 5월 16일 하루 동안에 지하철과 버스를 이용한 승객들의 교통카드 트랜잭션 데이터베이스이다. 기준 시점의 수도권 지하철 시스템은 지하철 1-8호선, 분당선, 인천1호선, 덕소까지의 중앙선 노선으로 구성된다. 기준 시점에서 지하철역들의 개수는 429개이고, 이 역들 사이의 간선들의 개수는 477개이다. 지하철역들 중에는 서로 다른 노선으로 환승할 수 있는 환승역들의 개수는 55개이다. 하루 동안 기록된 전체 트랜잭션들의 개수는 10,856,337개이며, 트랜잭션들의 전체 크기는 1.65GB의 CSV 파일이다. 전체 트랜잭션들 중에서 5,199,106개는 지하철을 사용한 승객들의 트랜잭션들이고 나머지는 버스를 이용한 승객들의 것이다. 지하철 승객 중에서도 승객 흐름에 기여한 트랜잭션은 5,175,067개이고, 나머지는 트랜잭션 내용 중에 승객 흐름에 기여할 수 없는 것으로 분류하였다. 이런 트랜잭션에 해당하는 경우는 승차와 하차 사이의 이동 시간이 12시간 이상인 경우도 있고, 지하철역에 들어 온 직후에 바로 그 지하철역을 나온 경우 등이다.

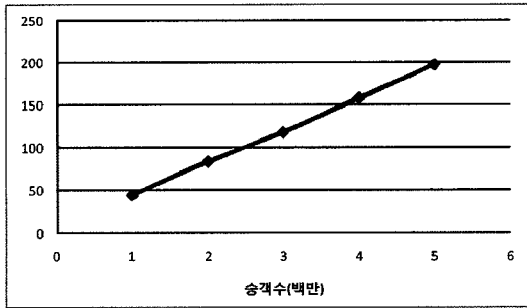


그림 7 승객 수에 따른 알고리즘의 실행 시간(단위: 초)

실험에 사용된 컴퓨터의 운영체제는 마이크로소프트 윈도우 비스타 64비트이고, 사용된 언어는 MS Visual Studio 2008 C++언어이다. 컴퓨터 하드웨어의 주요 사양은 다음과 같다: 두 개의 Intel Xeon 3.0GHz CPU, 두 개의 15Krpm의 SCSI 72GB 하드 디스크, 8GB 메인 메모리. 이와 같은 사양에서 주어진 입력 데이터베이스로부터 제안된 방법론으로 지하철 교통망의 승객흐름을 찾아내는 실행시간은 그림 7에서 표시하고 있다. 지하철 사용자의 트랜잭션들의 개수에 따라 변하는 실행시간이 선형적으로 변화함을 보여주고 있다. 지하철 교통망에 대한 자료구조를 구축하는 시간은 2.34초가 걸렸고, 전체 517만여 개의 지하철 트랜잭션들을 처리시간은

201초가 걸렸다.

그림 8은 서울 수도권 지하철 교통망의 주요 링크 중의 하나인 1호선의 서울역-시청과 2호선의 강남-교대 사이의 양 방향의 4개 링크 상에 시간대별 승객 흐름을 보여주고 있다. 승객 흐름 알고리즘에서 히스토그램의 시간 단위는 1분이지만 그래프로 데이터를 표시하기 위하여 10분 단위로 조정하여 주어진 링크에서 이동하는 승객들의 숫자를 나타내고 있다. 그래프에서 8시 40분대는 8시 40분부터 8시 49분 사이의 10분 간 이동 승객을 표시하는데 시청→서울역 방향의 승객 이동량은 1,252명이다. 같은 시간대에 교대→강남 방향의 링크 흐름은 9,416명으로 약 7.5배 정도의 차이를 나타낸다. 그림 1에서 설명된 강남역과 교대역의 양방향 링크 흐름을 분석해보면, 교대→강남 방향의 승객들의 수가 출근시간대에 반대 방향 보다 3배 정도 많음을 알 수 있다. 강남→교대 방향으로는 퇴근 시간대에 반대 방향보다 2배 정도 승객 이동량이 많음을 나타내고 있다. 그러므로 각 지하철역 사이의 링크에서 같은 시간대에 서로 다른 승객 흐름이 있음을 그래프로 알 수 있다. 승객 흐름이 많은 링크의 선로 보수가 다른 곳에 비해 더 필요함을 예측할 수 있다. 그림의 설명과 같이 제안된 방법론에 의해 수도권 지하철 교통망의 모든 링크에서 시간대별로 승객 흐름을 파악할 수 있게 된다.

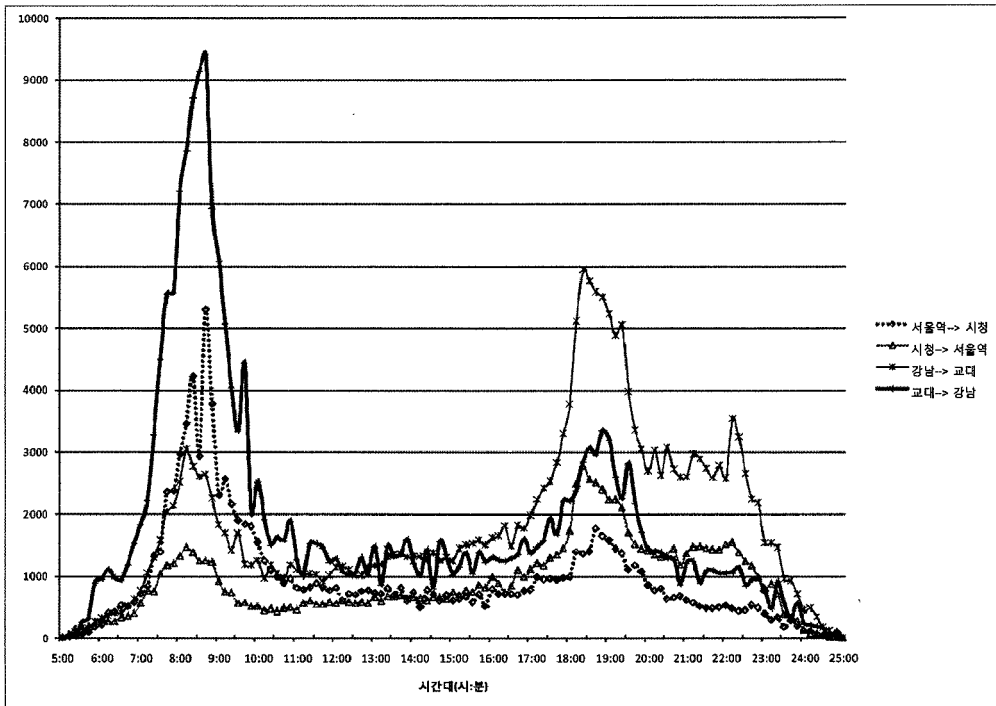


그림 8 시간대별 수도권 지하철 시스템의 링크상의 승객 이동량

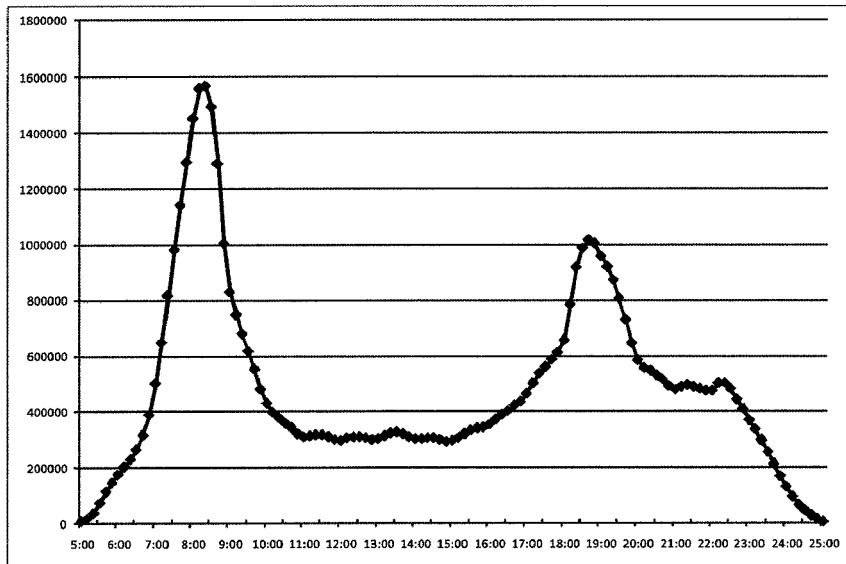


그림 9 시간대별 수도권 지하철 시스템의 링크상의 전체 승객 이동량

그림 9는 수도권 지하철 교통망의 모든 링크들에서 이동하고 있는 승객들의 숫자를 시간대별로 보여주고 있다. 그래프에서 8시 20분대의 값은 수도권 지하철 전체 구간에서 이동 승객을 표시하는데 1,566,792명을 나타내고, 오후 3시의 승객 이동량은 293,522명을 보여주고, 18시 40분의 이동량은 1,013,102명을 보여주고 있다. 이 그래프에서 수도권 지하철 승객들은 출근 시간 피크 타임인 오전 8시경에 급격하게 이용이 늘어나고 퇴근 시간대의 피크 타임인 오후 7시경에는 출근 시간대에 비해 30% 정도 승객들이 줄어드는 것을 보여준다. 업무 시간대인 오전 10시에 오후 4시 사이에는 10분당 대략 31만 명 정도가 지하철 선로 상에서 이동하고 있음을 나타내고 있다.

표 2는 수도권 지하철 교통망의 링크들 중에서 하루 동안 통과하는 승객 이동량이 많은 상위 10개 링크를 나타내고 있다. 1호선 인천선과 천안선 그리고 2호선이

만나는 구로→신도림 링크가 가장 많은 승객 흐름을 보여주고 있고, 나머지는 2호선과 4호선이 만나는 사당역에서 2호선의 강남역 사이의 링크들이 상위 순위로 보여 주고 있다.

### 6. 결론

서울 수도권 지하철 교통망은 지하철역인 정점, 지하철 역 사이에 연결된 선로인 링크, 그리고 승객들의 통행 흐름으로 구성된다. 본 논문에서는 수도권 지하철 교통망의 모든 링크에 이동하는 지하철 승객들의 흐름을 분석하는 방법론을 처음으로 제안하고, 실제 교통카드 트랜잭션 데이터베이스를 입력으로 하여 지하철역과 인접한 지하철역 사이의 각 링크 승객 흐름을 시간대별로 보여주었다. 개별 승객의 통행 경로는 최단 경로 알고리즘으로 찾아내어 교통망의 링크에 이동량으로 표시하였다. 제안된 방법으로 2007년도 하루 동안에 승객들이 사용한 약 1,100만 건의 교통카드 트랜잭션들로 이루어진 데이터베이스를 입력으로 하여 지하철 교통망의 모든 링크의 승객 흐름을 히스토그램으로 만들어 분석하였다. 각 링크의 시간대별 승객 흐름에 관한 정보는 지하철 시스템의 유지 보수에 직접 활용될 수 있고, 그리고 수도권 교통정책 수립과 수도권 지역의 토지이용계획 및 시설계획 등에 활용될 수 있다.

본 논문에서는 하루 동안 사용된 승객들의 교통카드 트랜잭션 데이터베이스가 입력으로 이용되었지만, 더 많은 입력 트랜잭션 데이터베이스로 주별, 월별, 년도별 승객 흐름을 계산하면 교통 흐름에 관한 보다 좋은 정

표 2 승객 흐름 중 상위 10개 링크

순위	링크	승객 흐름
1	구로 → 신도림	267,630
2	신도림 → 구로	256,569
3	사당 → 방배	234,678
4	방배 → 서초	233,811
5	서초 → 교대	231,623
6	교대 → 강남	229,006
7	방배 → 사당	223,556
8	서초 → 방배	223,166
9	교대 → 서초	222,455
10	강남 → 교대	221,902

보를 탐사해낼 수 있을 것이다. 대중교통의 또 다른 수송 체계인 버스를 이용하는 승객들의 흐름도 버스 노선과 버스 정류장에 대한 자료가 주어진다던 구할 수 있다. 그렇게 되면 버스와 지하철을 사용하는 대중 교통망의 승객 흐름을 구할 수 있게 되어 수도권 교통 흐름에 대한 많은 지식과 통찰력을 얻을 수 있을 것이다.

**참 고 문 헌**

[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] K. Lee and J.S. Park, "Travel Patterns of Transit Users in the Metropolitan Seoul," *Journal of the Economic Geographical Society of Korea*, vol.9, no.3, pp.379-395, Dec. 2006. (in Korean)

[3] 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)

[4] K. Lee, J. Hong, H. Min, and J.S. Park, Relationships between Topological Structures of Traffic Flows on the Subway Networks and Land Use Patterns in the Metropolitan Seoul, *Journal of the Economic Geographical Society of Korea*, vol.10, no.4, pp.427-443, Dec. 2007. (in Korean)

[5] 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)

[6] 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.

[7] Homepage of Seoul Metro Traffic Center: <http://www.seoulmetro.co.kr/>

[8] Homepage of Seoul City Hall: <http://www.seoul.go.kr/>

[9] K.W. Lim and Y.T. Lim, *Transportation Network Analysis*, Seoul National University Press, 2003. (in Korean)

[10] V. Latora and M. Marchiori, "Is the Boston subway a small-world network?," *Phys. A*, vol.314, pp.109-113, 2002.

[11] P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna, "Small-world properties of the Indian Railway network," *Physical Review E*, vol.67, issue.3, 036106, 2003.

[12] F. Chen, Q. Wu, H. Zhang, S. Li, and L. Zhao, "Relationship analysis on station capacity and passenger flow: a case of Beijing subway line 1,"

*Journal of Transportation Systems Engineering and Information Technology*, vol.9, iss.2, pp.93-99, April 2009.

[13] R. Sedgewick, *Algorithms in C++*, Third Ed., Addison-Wesley, 2002.



**박 중 수**

1981년 부산대학교 전기기계공학과(학사). 1983년 한국과학기술원 전기및전자공학과(석사). 1990년 한국과학기술원 전기및전자공학과(박사). 1983년~1986년 국방부 군무설계기과. 1994년 7월~1995년 7월 IBM Watson 연구소 객원 연구원. 1990년~현재 성신여자대학교 IT학부 교수. 관심분야는 데이터베이스, 데이터마이닝, 교통지리



**이 금 수**

1979년 성신여자사범대학교 지리학(학사) 1981년 성신여자대학교 대학원 지리학(석사). 1987년 미국보스턴대학교 지리학(지리학Ph.D.). 1988년 5월~1989년 5월 국토개발연구원 초빙책임연구원. 1989년 10월~1991년 1월 교통개발연구원 초빙 선임연구원. 1990년 9월~1992년 8월 성신여자대학교 한국 지리연구소 연구교수. 2000년 1월~2001년 2월 미국 워싱턴대학교 지리학과 객원학자. 2008년 1월~2009년 2월 미국 보스턴대학교 Center for Transportation Studies 객원학자. 1992년 9월~현재 성신여자대학교 지리학과 교수. 관심 분야는 교통지리, 최적화 입지모형, GIS