

교통카드자료를 이용한 통행패턴분석과 정책활용방안 연구 -경기도를 중심으로-

빈미영* · 문주백** · 조창현***

요약 : 본 연구는 교통카드 데이터를 이용하여 대중교통 이용과 관련하여 통행패턴을 분석하였으며 교통정책에 활용할 수 있는 방안을 제시하였다. 교통카드 데이터는 경기도권역을 대상으로 하였고 활용방안으로 교통정책 의사결정자가 버스정류소 시설을 개선할 때 교통카드데이터에서 얻어질 수 있는 여러 변수를 이용하여 대상지를 선정한다는 시나리오를 설정하여 분석하였다. 분석결과, 의사결정방법론인 K평균 군집분석과 CHAID(Chi-squared automatic interaction detection)를 이용하였으며, 유의수준 $p < 0.01$ 에서 정책에 유용하게 이용될 수 있는 결과를 얻었다. 또한 본 연구에서는 이러한 결과들을 근거로 교통카드데이터를 실제로 정책에 활용되기 위해서 개선되어야 할 정책적 함의를 제시하였다.

주요어 : 교통카드데이터, K평균군집법, CHAID분석, 통행패턴

1. 서론

최근 수도권 대중교통정책에서 두드러진 변화 중 하나는 이용요금의 전자지불화가 정착되었다는 점이다. 교통카드 이용은 2000년대 초반부터 수도권을 중심으로 부분적으로 확산되다가 2007년에는 수도권 환승할인요금제가 도입되어 본격적으로 이용되었으며, 2011년 현재에는 교통카드 이용건수가 하루 평균 2천 4십만 건(2011년 4월 기준)에 달하고 있다((주)한국스마트카드). 이 수치는 대중교통 이용건수의 98%를 차지하는 것으로 나타났다. 즉, 수도권의 대중교통 통행패턴을

98%나 파악할 수 있다.

그러나 이러한 변화에도 불구하고 아직 교통카드 데이터를 이용하여 수행된 연구는 다양하지 않고 최근에 와서야 활발해지기 시작했는데, 주로 정류장 승하차 이용수요, 기종점 통행량 산출 등 한정된 분야에서만 활용되고 있다(빈미영 외, 2012). 대량의 교통카드 데이터는 교통정책 의사결정자가 단기간에 현장조사를 수행하지 않고도 효율적으로 의사결정을 할 수 있는 정보로 활용될 수 있다.

본 연구는 이러한 배경 하에서 교통정책 의사결정자가 교통카드 데이터를 이용하여 정책을 수립할 때 활용할 수 있는 방안을 제시하였다. 교통정

이 논문은 경기개발연구원의 정책과제로 수행된 연구임.

* 경기개발연구원 교통연구부 연구위원

** 경기개발연구원 교통연구부 연구원

*** 경희대학교 지리학과 부교수

책 의사결정자가 정류장시설을 개선하기 위하여 제한된 예산집행 환경에서 정류장의 대상지를 선정한다고 가정하고 시나리오를 구상하였다. 본 연구에서 시나리오는 3개를 구상하였는데, 첫째, 환승거점 정류장을 선정하는 방안, 둘째, 교통약자(어린이)가 많이 이용하는 정류장의 시설개선을 위하여 대상정류장을 선정하는 방안, 마지막으로 관광지주변 정류장의 시설개선을 위해 정류장을 선정하는 방안을 제안하였다.

본 연구는 경기도 버스에서 이용된 교통카드데이터를 수집하고 활용하였다. 시간적 범위는 2011년을 기준으로 하였으며, 분석날짜는 일반적인 통행패턴을 고려하기 위하여 학생들의 방학이나, 직장인의 휴가철을 제외하여 통행패턴 변화가 가장 적게 일어나는 시기로 결정하여 아래와 같이 선정하였다.

-주말 : 2011년 4월 24일(일)

-주중 : 2011년 4월 27일(수)

본 연구에서는 경기도 교통카드데이터를 이용하였기 때문에 서울에서 서울인허가 버스만 이용한 '서울순수통행'과 인천에서 인천인허가 버스만 이용한 '인천순수통행'은 누락되어있으나, 경기도의 대표성은 나타낼 수 있을 것으로 판단하였다. 교통카드데이터 수집범위를 통행기준으로 정리하면 표 1과 같다.

하루 이용건수 6백만 건이 넘는 대량의 교통카드데이터를 이용하였는데 이를 분석하기 위해서 MS-Access를 이용하였으며, 시나리오별 분석을 위해서는 군집분석과 CHAID(Chi-Squared Automatic Interaction Detection)를 이용하였다.

표 1. 교통카드데이터 수집범위(통행기준)

기점 \ 종점	서울	인천	경기
서울	×	×	○
인천	×	×	○
경기	○	○	○

계산은 SPSS ver. 18을 이용하였다. 마지막으로 도출된 결과들을 이해하기 쉽게 도식화하기 위해서 ArcGIS ver 10.0을 활용하였다.

2. 문헌연구 및 방법론

1) 문헌연구

교통카드를 이용한 연구로는 국내에서 박진영·김동준(2007; 2008a, 2008b)이 처음으로 체계적인 분석을 시도하였다. 교통카드 데이터를 이용하여 교통수단별 이용자, 환승 승객수, 탑승시간, 교통수단별 통행시간을 산출하였다.

성현곤 외(2008)는 자가용 승용차의 이용수요를 줄이고 대중교통수단의 이용을 촉진할 수 있는 TOD에 미치는 토지이용 및 도시설계 특성의 영향을 파악하고자 하였으며, 교통카드자료는 역세권의 이용수요를 산출하는데 이용되었다.

김기혁·이승철(2011)은 대구시의 교통카드자료를 이용하여 환승정류소의 특성을 분석하였다. 대구시의 교통카드자료와 BMS(Bus Management Systems)를 매칭하여 환승통행 데이터를 구축하였다. 환승통행의 행태를 분석하기 위하여 노선수(개), 평균 대기시간(분), 승차인원(명/일), 도심으로부터 거리(Km), 환승량(명/일), 평균접근거리(Km), 평균접근시간(분) 등의 환승통행 변수를 이용하였다. 군집분석을 이용하여 대구 시내 버스정류소를 3개로 그룹화 하였으며 입지 조건에 따라 시계유출입권, 시내외곽, 도심·부도심권, 광역외곽권 환승센터로 구분하였다.

이와 같이 교통카드데이터의 활용과 관련된 연구는 최근에 와서 활발히 진행되고 있으며, 국내에서 이용이 비교적 확산된 수도권 중심을 중심으로 진행되고 있다. 2007년 수도권 통합환승 할인제가 시행되면서, 서울, 인천, 경기도가 교통카드와 관

표 2. 수도권교통카드관련 연구사업 현황

지역	연구사업명	카드관련 활용내용
서울	교통카드 데이터를 활용한 OD 추정 및 활용(김순관, 2008)	- 교통카드 데이터 기반의 대중교통OD구축 및 활용방안 제시 • 총 이용객수, 교통수단별 이용객수, 사용자 구분에 따른 이용객수, 환승횟수별 이용객수, 환승횟수별 교통수단별 이용객수, 시간대별 이용객수
	교통카드 Data를 활용한 대중교통 평가체계 구축 방안 모색(신성일 외, 신성일, 2008)	- 대중교통 평가체계 구축 방안 모색 • 평균통행시간, 평균통행속도, 총 환승횟수, 평균 환승시간, 평균 승차요금, 지역간 통행특성
	대중교통 서비스 지표 산출연구 (1,2단계)(이신혜, 2010)	- 대중교통 서비스지표 산출 • 운행횟수, 정시성, 통행속도, 재차인원, 혼잡률, 중앙버스전용차로의 유무관련 지표평가, 노선개수, 정류소개수
인천	인천시 대중교통 정책에 따른 교통수요 변화 연구(김중형, 강미혜, 2010)	- 수요패턴 변이 분석 • 교통카드 이용객수, 교통카드 사용자 계층별 이용현황, 교통수단별 이용현황, 시간대별 이용현황, 승하차지역별 이용현황 • 환승지점·수단별 이용현황, 기타분석(버스유형별 통행거리와 통행시간, 버스노선별 이용객규모)
경기	경기도 버스서비스 개선방향 - 교통카드 자료 분석결과를 중심으로(빈미영, 2009)	- 버스서비스 개선을 위한 서비스 지표 산출 • 기·종점 통행량, 평균 버스 차내 혼잡도, 도로구간별 버스중복도, 정류소 정차노선 중복도, 도로구간별 이동통행, 환승통행(환승횟수별 승객수, 평균환승소요시간), 환승통행 정류소위치, 시·군별 버스이용 분포도, 정류소별 버스이용 분포도 • 노선별 승차통행수, 승차 유형별 통행, 1통행당 이동관련 지표(버스이용 1통행당 평균이동거리(분), 버스이용 1통행당 평균이동거리(m), 버스이용 1인당 평균통행횟수(회)), 차량 1대당 평균운송량

련한 연구사업을 추진하기 시작하였다. 현재는 (주)한국스마트카드가 수도권의 교통카드시스템을 통합 정산하고 있다. 서울, 인천, 경기도가 포함된 수도권에서 추진된 교통카드 관련 연구를 정리하면, 표 2와 같다. 대중교통서비스의 지표를 산출하는 기초연구가 대부분이다.

Pelletier 외(2011)는 교통카드 연구현황을 종합적으로 정리하였다. 교통카드 활용을 3가지로 분류하였는데 장·단기 정책(long-term planning), 도로 네트워크 개발과 서비스 조정(service adjustments and network development), 그리고 이용자 통계와 운영지표(ridership statistics and performance indicators)로 구분하여 기대효과와 함께 정리하였다.

2) 의사결정방법론

(1) K평균 군집분석

군집분석은 실험의 결과나 표본 등에서 얻어진 자료들을 어떤 성질에 따라 분류하여 동질적인 몇 개의 군집(cluster)으로 나누고자 할 경우에 사용한다. 아래내용은 이기영,장명순(2005)의 연구내용을 토대로 인용하였다. 이 기법은 n 개의 개체를 대상으로 p 개의 변수들을 측정하였을 때, 이 변수들을 이용하여 n 개의 개체들 사이의 유사성 또는 비유사성의 정도를 측정하여 개체들을 가까운 순서대로 연결해 나간다. MacQueen(1967)은 가장 가까운 중심점을 갖는 군집에 각 항목을 할당하는 알고리즘을 설명하면서 K평균법이란 용어를 처음으로 사용하였다. K평균법은 계층적 군집분석과는 달리 개체가 어느 한 군집에만 속하도록 하는 상호 배반적 군집방법이다. 이 방법은 군집의 수를 미리 정하고, 각 개체가 어느 군집에 속하는지를 분석하는 방법으로서 대규모의 자료에 대한 분석에 유용하게 이용된다. 분석의 절차는 아래와 같다.

- step 1: 전체 개체를 K군집수로 초기화한다. 이러한 군집의 초기 분류는 분석자가 지정할 수도 있고, 프로그램에서 제공할 수 있다.
- step 2: 각 군집의 각각의 변수에 대해서 중심점들을 계산한다.
- step 3: 주어진 케이스에 대해서 중심점과의 거리를 계산한다. 만약 현재 속하고 있는 군집의 중심점과 가까우면 그대로 놔두고, 그렇지 않으면 다른 군집으로 배정한다.
- step 4: 각 케이스에 대하여 step 3번 절차를 시행한다.
- step 5: 어느 케이스도 다른 군집에 재배정되지 않을 때까지 step 2, step 3, step 4과정을 반복한다.

K평균법에서는 자료의 상호 연관성을 판단하는 지표로 유클리드 거리를 적용한다. 즉 p 차원 상의 두 관찰값 $X=(x_1, x_2, \dots, x_p)$ 와 $Y=(y_1, y_2, \dots, y_p)$ 간의 유클리드 거리는 식 1과 같이 정리된다(이기영·장명순, 2005).

$$d(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

$$= \sqrt{(X - Y)'(X - Y)} \quad (\text{식 1})$$

(2) CHAID분석

본 연구에서는 정류소시설 개선을 위한 대상지를 분류하여 선정하는 방법으로 CHAID방법을 적용하였다. 정류소의 이용수요, 평균대기시간 등을 이용하기 위해서는 연속변수도 이용할 수 있어야 하며, 분류된 변수가 재사용되지 않는다는 점 등을 고려하여 여러 의사결정분류법 중에서 CHAID방법론을 선정하였다(조창현 외, 2009).

CHAID 알고리즘은 카이제곱 통계량을 통해

비율이 유지되는 정도를 파악하는데, 여러 변수 중 비율을 가장 많이 깨뜨리는 변수가 결과변수에 영향을 가장 많이 미치는 변수가 된다. 비율이 깨진 정도는 카이제곱에서 $r \times c$ 분할표(Contingency Table)로부터 계산된다. 이때, Pearson의 카이제곱 통계량은 식 2와 같이 정의된다(김신곤·박성용, 1999).

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \quad (\text{식 2})$$

여기서, f_o 는 관찰치, f_e 는 예측치.

이 통계량은 자유도가 $(r-1)(c-1)$ 인 카이제곱 분포를 따른다. 카이제곱통계량이 자유도에 비해 매우 작다는 것은 입력변수의 각 범주에 따른 결과변수의 분포가 동질적이라는 것을 의미하며, 입력변수가 결과변수의 분류에 영향을 주지 않는다고 말할 수 있다. 자유도에 대한 카이제곱 통계량의 크고 작음은 p 값으로 표현될 수 있는데, 카이제곱 통계량이 자유도에 비해서 작으면 p 값은 커지게 된다. 결국 노드는 p 값이 가장 작은 변수를 기준으로 가지가 형성된다.

3. 교통카드기반 통행패턴 분석

1) 개요

본 연구에서는 평일데이터는 4월 27일 수요일 자료를 이용하였으며, 주말데이터는 4월 24일 일요일 데이터를 이용하였다. 승·하차의 패턴을 파악하기 위해서는 버스 하차 시 카드를 단말기에 대지 않았던 하차 미태그 정보를 제거해야하는데, 이를 제거하여 집계하면, 표 3과 같이 6,746,651건으로 나타났으며, 하차 시 단말기에 카드를 대지 않는 비율은 약 2.8%로 나타났다. 통행인은

표 3. 표본수

구분	표본수
통행인수(ID수)(인)	2,632,484
하차시 미테크 카드데이터 제거 (건)	6,746,651
연계통행수(건)	4,311,217

주 : 평일자료, 4월 27일 수요일 일일데이터

2,632,484명으로 집계되었다.

2) 기초분석

표 4는 하루에 이용하는 대중교통 승차횟수 분포를 나타냈는데, 최다 이용회수는 12회로 나타났다. 하루에 1회 이용하는 이용자가 전체의 46%이며, 2회 통행이 46%를 차지하고 있다. 따라서 2회까지 누적비율을 보면, 92%로 대부분이 하루에 2회 대중교통을 이용하고 있는 것으로 나타났다. 본 연구에서는 승차횟수 2회까지의 통행수 2,434,652건을 가지고 분석하였다.

이를 토대로 기점에서 종점까지 이동하는데 얼마정도의 대중교통을 환승하는지를 나타내면 표 5와 같다. 환승하지 않고 대중교통을 한 번 이용하는 건수가 전체의 52.4%로 나타났다. 평균적으로

표 4. 승차횟수 분포

승차횟수	통행자수	통행자수 비율		통행(건)
		비율	누적	
1	1,213,556	46%	46%	1,213,556
2	1,217,326	46%	92%	2,434,652
3	153,129	6%	98%	459,387
4	40,638	2%	100%	162,552
5	6,301	0%	100%	31,505
6	1,250	0%	100%	7,500
7	231	0%	100%	1,617
8	37	0%	100%	296
9	12	0%	100%	108
10	2	0%	100%	20
12	2	0%	100%	24
합계	2,632,484	100%	-	4,311,217

표 5. 승차횟수별 통계지표

승차횟수	환승횟수	이용자 분포(인, (%))	평균이동 시간(분)	평균이동 거리(m)
1회	0	1,274,123(52.4)	19.66	7,502.24
2회	1	912,539(37.6)	53.96	20,859.54
3회	2	215,226(8.9)	72.93	26,661.5
4회	3	23,514(1.0)	92.76	29,924.19
5회	4	4,492(0.2)	122.36	32,486.80
종합	-	2,429,894(100.0)	38.16	14,478.71

38.16분 걸려 14.5km를 이동하는 것으로 나타났다.

경기도 교통카드 데이터에는 이용자가 총 16개의 코드로 분류되어있으나, 현재 데이터가 수집되는 것은 표 6과 같이 일반, 청소년, 초등학생, 국가유공자의 4개로 분류되고 있으며, 각각 동반탑승한 것도 기록된다. 가장 많은 것이 일반이용건수이며, 전체의 87.5%를 차지하고 있고 그 다음으로는 청소년으로 나타나고 있다.

표 6. 이용객 유형별 분포

승차유형	이용건수	비율(%)
일반	3,409,720	87.5
청소년	452,204	1.6
초등생	20,851	0.5
일반+초등생	9,324	0.2
일반+청소년	2,500	0.1
국가유공자	1,540	0.0
청소년+초등생	258	0.0
일반+청소년+초등생	220	0.0
합계	3,896,617	100.0

주: 이용건수합계는 총 이용건수 합계로써 이용건수와 승차횟수를 곱하여 산출한 결과임.

표 7. 수도권 통행 평균 이동시 소요된 환승횟수

O \ D	서울시	인천시	경기도	합계(건)
서울시	0.64	1.07	0.94	461,730
인천시	1.12	0.41	1.17	39,945
경기도	0.98	1.17	0.35	1,863,248
합계	470,222	39,484	1,855,217	2,364,923

표 7은 수도권 시·도간 대중교통을 이용하여 이동시 소요된 환승횟수를 나타냈다. 가장 낮은 값을 나타내는 곳은 경기도에서 경기도로 이동하는 경우이며, 이는 경기도내에서 통행시 직결노선을 주로 이용하는 것으로 해석된다. 또한, 인천시와 경기도 간의 평균 환승횟수가 1.17회로 가장 높은 값을 나타내는데 이것은 인천과 경기도간의 노선 체계가 직결노선은 적고 비효율적으로 되어있어 있음을 시사하고 있다.

4. 환승정류소 유형구분과 활용방안

1) 시나리오 설정

교통정책의 일환으로 정류소 시설개선을 위하여 대상지점을 선정하는 방법을 시나리오를 설정하여 분석하였다. 시나리오는 버스정책 의사결정자가 버스정류소의 시설개선을 위해 예산의 한계 내에서 우선순위를 선정한다고 가정하였다. 3가지 활용방안을 검토하였는데, 첫째, 환승거점 정류소를 선정하는 방안이다. 정류소의 이용자 수요와 차량수요, 평균대기시간의 변수를 이용하여 선정하였다. 둘째, 교통약자를 위한 버스정류소 시설개선 방안이다. 교통약자의 대표적인 대상은 노인을 들 수가 있으나, 본 연구에서 카드데이터에서 얻을 수 있었던, 초등학교를 대상으로 버스정류소 시설개선을 위한 대상 정류소를 선정한다고 가정하였다. 셋째, 관광지주변의 정류소 시설개선이다. 관광지를 이용하는 이용자들을 위하여 막차정보와 환승정보를 제공하기 위한 것이다. 시나리오 1과 시나리오 2는 의사결정나무 방법론으로 CHAID (Chi-squared automatic interaction detection)론을 이용하였으며, 시나리오3은 군집분석으로 수행하였다. 정리하면, 표 8과 같다.

표 8. 시나리오 및 변수

시나리오		변수
1	환승거점 정류장 선정	역 이용수요, 버스 평균 대기 시간
2	교통약자(어린이)를 위한 정류소 시설 개선 사업	초등학교 수요, 일반 이용, 평균 대기 시간
3	관광지 대중교통 이용	관광지 역에서의 이용자 수요

(1) 시나리오1: 환승거점 정류장 선정

환승거점을 선정하기 위해서 종속변수는 환승정류소 이용수요, 설명변수는 환승정류소에 정착하는 차량수요, 각 환승정류소에서 환승했던 이용자들이 대기했던 평균 대기시간을 선정하였으며, CHAID분석을 수행하였다. 환승거점을 선정하기 위한 기본 룰을 정의하면, 식 3과 같다.

$$\begin{aligned}
 & \text{if 임의의 정류소에서 이용자수요} > \text{기준치,} \\
 & \qquad \qquad \qquad \text{차량수요} > \text{기준치,} \\
 & \qquad \qquad \qquad \text{평균대기시간} > \text{기준치}
 \end{aligned}$$

$$\text{then 환승거점 정류소} \qquad \qquad \qquad \text{(식 3)}$$

(2) 시나리오 2: 교통약자(어린이)를 위한 정류소 시설 개선 사업

두 번째 시나리오는 교통약자를 위한 정류소 시설개선 사업 시 우선대상 정류소를 선정하는 것이다. 버스정류소에 어린이의 안전한 승·하차를 위하여 방법시설을 설치하려고 할 때 우선 대상이 되는 정류소를 파악하고자 한다. 여러 기준이 있을 수 있으나, 식 4와 같은 범위로 선정할 수 있다. 임의의 정류소에서 초등학교 수요가 어느 정도 있으나, 주변의 대기하는 일반 수요가 적고 정류소에서 대기하는 시간이 기준치보다 긴 정류소를 선정할 수 있다.

$$\begin{aligned}
 & \text{if 임의의 정류소 초등학교 수요} > \text{기준치} \\
 & \qquad \qquad \qquad \text{주변 일반 수요} < \text{기준치}
 \end{aligned}$$

평균대기시간 > 기준치

2) 분석결과

then 시설개선 대상 정류장 (식 4)

(3) 시나리오 3: 관광지 대중교통 이용

관광지는 주중에 출퇴근을 목적으로 이용하는 수요와는 다르게 익숙하지 않은 통행행태를 보이는 수요가 많다. 관광지에 대중교통을 이용하여 접근하게 되는 경우, 노선이나 환승정보의 접근이 어려울 수 있다. 따라서 관광지 정류소에서 환승 거점으로 안내할 수 있는 노선안내와 환승정보를 추가적으로 설치하는 방안을 검토할 수 있다.

if 임의의 관광지 정류소 수요 > 기준치

then 시설개선 대상 정류장 (식 5)

(1) 환승거점 정류장 선정결과

임의의 정류소에서 이용수요, 버스차량 대수가 많으면 매우 혼잡한 정류소라고 볼 수 있으며, 평균 차량도착 대기시간이 길면, 수송체계 서비스 질이 낮은 정류소라고 볼 수 있다. 환승거점 정류장 선정은 이러한 정류장을 선정한다고 가정하였다. 해석의 편의상 의사결정나무의 노드수를 차량수요는 3개, 평균대기시간은 2개 그룹을 사전에 정하여 실행하였다.

CHAID 분석결과를 나타내면 그림 1과 같다. 정류소 이용수요에 영향을 미치는 변수는 차량수요로 나타났으며, 차량수요에 영향을 미치는 변수는 평균대기시간으로 나타났다. 각 노드로 분류된

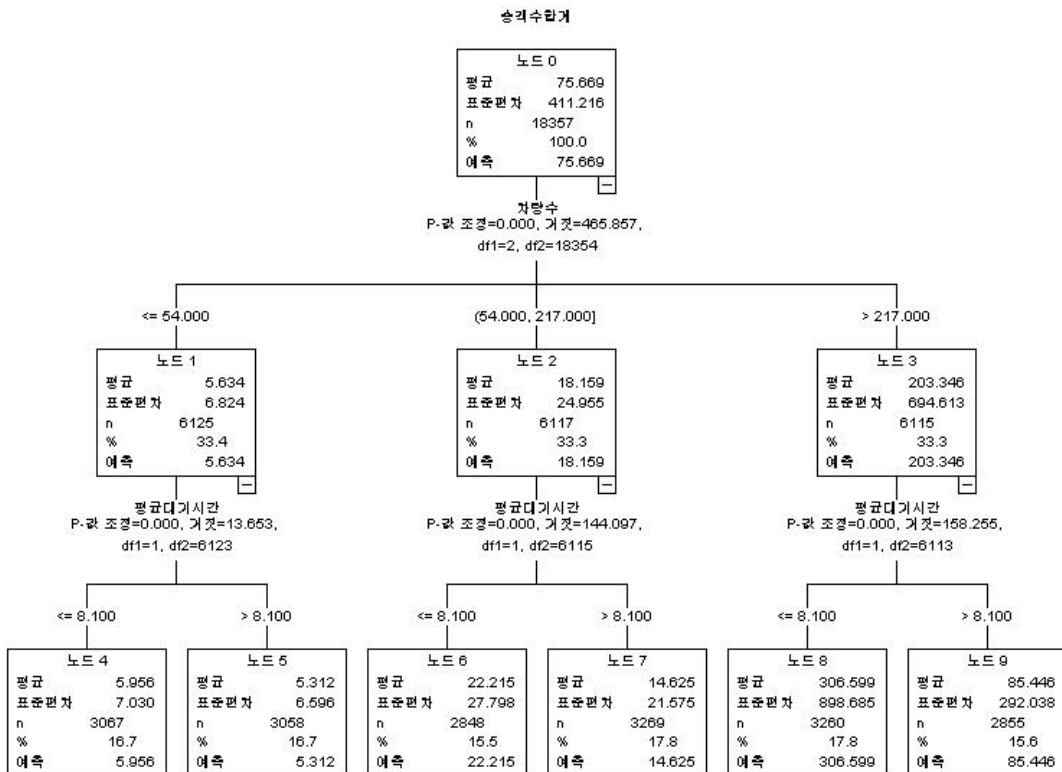


그림 1. 환승거점 정류장 선정을 위한 CHAID 분석결과

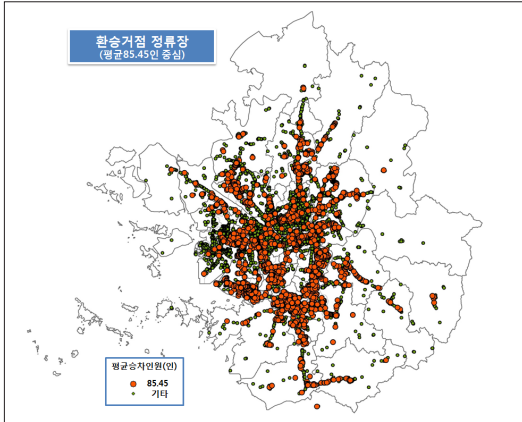


그림 2. 환승거점으로 선정된 정류소 위치

결과값들은 유의수준 $p < 0.01$ 에서 통계적으로 유의미한 것으로 나타나 분류된 집단은 차이가 있는 것으로 나타났다. 차량수요는 한 정류소에 정차하는 대수가 3개 그룹으로 분류되었는데, 54대 미만, 54대에서 217대 이하, 217대 이상으로 구분되었다. 평균대기시간은 8.1분 이하와 8.1분 이상으로 구분되었다. 노드8과 노드9가 해당되며, 노드8은 평균대기시간이 8.1분 미만이며, 노드9는 평균대기시간이 8.1분 이상인 정류소를 의미한다. 그중 노드9번에 해당되는 정류장을 그림 2에 다른 노드를 기타로 차별화하여 나타낸 정류장 위치분포도이다. 이들은 차량이 많아 혼잡하고 차량도착

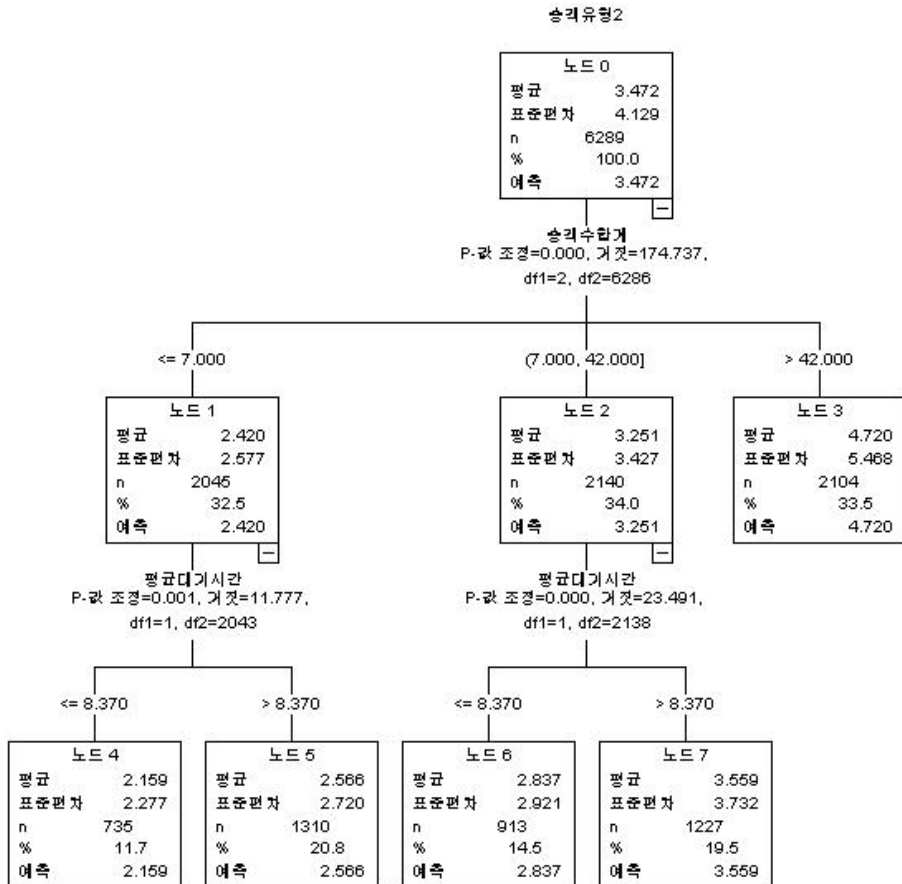


그림 3. 정류소 치안시설 확보 대상 선정을 위한 트리 다이어그램

대기시간이 길어 환승거점 시설 설치가 필요한 정류장이다.

노드9에 해당하는 정류장을 이용한 승하차 인원은 평균 85.45인으로 나타났다. 환승거점 정류장의 시설개선 사업은 정류장의 쉘터 크기와 버스베이의 적절한 규모설치 등의 사업을 고려할 수 있다.

(2) 교통약자(어린이)를 위한 정류소 시설개선사업

중속변수를 초등학생 수요로 하고 독립변수를 일반승객과 평균대기시간으로 하여 CHAID 분석을 수행하면 그 결과가 그림 3과 같다. 노드의 수는 마찬가지로 해석의 편의상, 일반승객수요(승객수)는 3개로 평균대기시간은 2개로 정하였다.

같은 정류소에 대기하는 이용객이 평균 7.0인 이하이며, 평균대기시간이 8.4분 이상인 정류소가 1,310개로 선정되었고 이것은 노드 5에 해당된다. 이와 같은 정류소에는 어린이 범죄를 예방하기 위해 방범시설을 설치할 수 있다. 그림 4는 노드 5에 해당하는 정류소 위치를 나타냈다. 이들은 이용객수가 적고 평균대기시간이 길어 어린이 보호를 위한 시설 보완이 필요한 정류장들을 나타낸다.

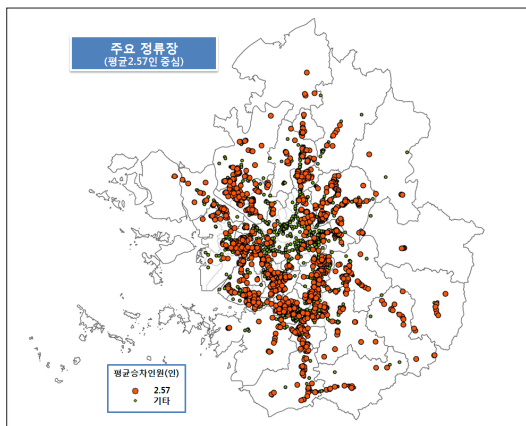


그림 4. 교통약자를 위한 시설보완이 필요한 정류소

(3) 관광지 정류장 주변 개선사업

관광지는 경기도 280개 지점을 사전에 선정하여 각 지점주변에 도보로 접근할 수 있는 정류장을 GIS를 이용하여 선정하였다. 이를 바탕으로 4월 24일 일요일에 해당 정류장에서 하차한 이용객의 통행자료를 만들었다. 그 결과, 총 127개의 정류소에서 평균 11.84명이 관광지 주변의 정류장에 하차한 것으로 나타났다. 이를 하차건수를 기준으로 군집분석을 수행하였다. 그 결과를 표 9에 제시하였다.

총 3개의 그룹으로 분류되었으며, 가장 많은 곳은 239명이 하차하였다. 그 위치는 그림 5에 나타난 지점으로 경기도 수원 팔달문에 위치한 정류소로 나타났다. 이들은 하차건수 기준으로 가장 많은 이용객들이 이용하는 정류장으로서 관광지 정류소 기능을 갖도록 정류소 개선사업을 하는 것이 타당한 것으로 보인다.

표 9. 관광지 평균승차건수 군집분석(승차기준)

군집 번호	평균승차 건수(건)	표준편차	최소값 (건)	최대값 (건)	N (개소)
1	6.48	8.38	1.00	42.00	120
2	81.33	38.54	50.00	155.00	6
3	239.00	.	239.00	239.00	1
합계	11.84	28.15	1.00	239.00	127

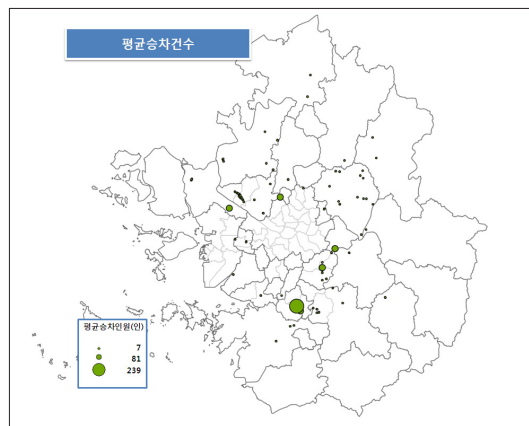


그림 5. 승차인원 기준 그룹별 관광지 정류소 위치

정류장시설개선을 위하여 교통카드데이터를 이용하여 대상 사업지를 분류하는 방안을 시나리오 별로 분류하여 나타냈다. 교통정책 의사결정자는 이와 같이 선정된 정류장의 시설을 개선하기 위해서 정류소 유형별 시설설치 가이드라인이 필요하다. 정류소 수요에 따른 버스정류소 적정길이를 산정하고 정류소에 설치되어야 하는 시설에 대한 표준모델과 가이드라인이 제공되어야 할 것이다. 또한, 관광지주변의 정류소 개선사업으로는 막차, 환승정보 제공방안을 검토할 수 있다.

5. 결론 및 정책적 시사점

본 연구는 경기도 교통카드 자료를 이용하여 정책시나리오를 설정하여 활용방안을 검토하였다. 본 연구를 통해 교통카드 자료를 이용하는 것은 기존의 대중교통 서비스지표를 도출하는 것뿐만 아니라, 다양한 정책에 활용될 수 있는 가능성을 제시하였다. 교통카드 데이터를 이용하여 기초 분석으로써 통행패턴을 분석하였는데, 하루에 대중교통을 이용하는 이용건수는 최대 12회까지 나타났으며, 대중교통을 연속해서 이용하는 행태는 이용건수의 92%가 환승을 포함하여 2회이하인 것으로 나타났다. 이 중 하루에 2회 대중교통을 이용한 이용자들을 대상으로 분석하였다. 활용방안으로는 정류소 개선사업을 위해 대상지를 선정하는 시나리오를 3개 제안하여 분석하였다. 시나리오로는 환승거점 선정, 교통약자 시설개선 우선대상 정류소 선정, 관광지 정류소 환승정보 제공을 위한 우선 대상 정류소 선정사업으로 가정하여 제시하였다. 방법은 CHAID와 군집분석을 적용하였으며, 환승거점을 선정하기 위해서는 정류소이용 수요, 차량수요, 평균대기시간 변수를 적용하였다. 교통약자를 위한 정류소시설개선의 대상정류소 선정에는 초등학교가 주로 이용하는 정류소를

대상으로 하였으며, 초등학교 이용수요, 일반이용 수요, 평균대기시간을 고려하였다. 관광지 정류소는 이용수요만을 가지고 분석하였다.

분석 결과, 환승거점으로는 차량이 많아 혼잡하고 차량도착 대기시간이 긴 정류장에 필요한 시설을 설치하는 것이 타당함을 보였으며, 어린이를 위한 시설 보완이 필요한 정류소로서는 이용객 수가 적고 평균대기시간이 긴 정류장들을 선정할 필요가 있음을 확인하였다. 또한 관광지 정류소로서 주변 개선사업이 필요한 정류소로서는 하차건수 기준으로 가장 많은 이용객들이 이용하는 정류장을 선정하는 것이 타당함을 확인하였다. 결론적으로 정책의사 결정자의 직관적인 판단을 지원하는 의사결정 방법으로는 유용할 것으로 판단된다.

이러한 방법이 실제로 적용되기 위해서 선행적으로 개선해야 할 것들이 있는데 정책건의로 함께 아래와 같이 제시하였다. 첫째, 수도권 교통카드 데이터 활용을 위한 공유체계가 필요하다. 공유체계란, 서울, 인천, 경기도의 각 지역별로 수집, 운영하는 교통카드데이터를 물리적, 행정적으로 하나로 통합하여 수집, 관리하며, 정보를 공유할 수 있는 체계를 의미한다. 현재는 수도권의 교통카드 자료를 (주)한국스마트카드에서 통합 정산하고 있다. 그러나 수도권에서 이용된 교통카드데이터는 각 시·도에서 관리하며, 정산업체가 다르다. 따라서 교통카드 데이터를 수집하기 위해서는 1차적으로 서울, 인천, 경기도로 각각 정보공개 요청을 하고 2차적으로 각 기관의 정산 위탁업체인 (주)한국스마트카드와 (주)이비카드로부터 자료를 수집해야 한다. 이에 대한 행정적, 시간적 손실이 막대하다고 할 수 있다. 교통카드 데이터를 분석 시 나올 수 있는 의문점들을 해결하기 위해서 일일이 확인하는 작업도 어렵다. 이와 관련하여 철도공사는 자체 정산시스템이 없고 (주)한국스마트카드에 의존하고 있기 때문에 요금정산을 위한 데이터 검증조차 어려워 2011년 독립정산 계획을 수립한 바 있다.

둘째, 다양한 지표산출을 위하여 원시데이터를 활용할 수 있는 환경이 조성되어야 한다. 당초 교통카드는 정산을 목적으로 하여 수집하고 운영되고 있다. 따라서 정산에 필요하지 않은 데이터는 수집, 관리가 되지 않는다. 일례로 노인통행은 향후 고령화 사회에 대중교통서비스의 현황을 파악하고 정책을 제시하는데 중요한 지표이나, 현재 교통카드 데이터로는 수집할 수가 없다. 노인카드데이터가 구분되어 수집된다면, 고령화 사회에 대응할 수 있는 노인 대중교통정책에 이를 활용할 수 있을 것이다. 이를 위해서는 정산단계가 아닌 원시데이터 수집단계에서 데이터를 수집하고 분석할 수 있는 환경이 조성되어야 할 것이다.

셋째, 다양한 정책활용을 위하여 교통카드데이터 이용메뉴얼 작성이 필요하다. 지금까지 교통카드 자료는 정산을 위하여 수집, 관리되고 있다. 그러나 교통카드데이터가 교통정책에 활용되기 위해서는 정산위주의 데이터에서 정책활용 데이터의 변환이 필요하다. 예로, 현재 교통카드데이터는 관할지역과 요금위주로 교통수단이 분류되어 있다. 이를 여객자동차운수사업을 근거로 한 교통수단 분류체제로 구분할 수 있는 단계가 필요하다.

넷째, 대중교통기본계획 수립시 교통카드데이터를 적극 활용할 수 있도록 제도화해야 한다. 예로, 경기도는 대중교통의육성및이용촉진에관한법률에 의해 31개 시·군이 5년마다 지방대중교통기본계획을 수립해야 한다. 기본계획을 수립하기 위해서는 시·군의 대중교통 운행과 이용현황을 조사해야 한다. 현재까지 대부분이 현장조사를 수행하는데 많은 인력이 동원되며 시간이 소모되고 있기도 하다. 최근에는 일부 시·군에서 교통카드 데이터를 이용하여 계획수립에 반영하는 노력을 보이고는 있으나, 교통카드 데이터를 이용한 사례가 많지 않기 때문에 데이터의 수집과 이용방법의 이해가 어려워 활용하고 있지 못하고 있다. 1년 365일 24시간동안 수집, 관리되고 있는 교통

카드 데이터는 경기도 대중교통이용자의 이용패턴을 알 수 있는 모집단 자료이다. 중앙정부는 대중교통기본계획과 지방대중교통기본계획 수립시 교통카드 데이터가 적극활용되어 불필요한 현장동원 조사를 지양하여 조사비 등을 절감할 수 있도록 제도화 혹은 인증화를 추진해야 할 것이다.

다섯째, 국가단위 교통카드활용연구사업의 실행이 필요하다. 국토해양부는 2008년 12월에 수립한 교통카드전국호환기본계획(2009~2013)에서 교통카드 사용처 확대사업과 교통카드 이용자 활용체계 구축방안을 제안한 바 있다. 교통카드 정보 활용체계 구축사업으로 2009년부터 2012년까지 중기사업으로 추진과제로 제시한 바 있다. 그러나 현재 추진되지 않고 있는 실정이다. 이 사업이 추진된다면, 1차적 목표로 교통카드 정보 활용체계가 구축될 수 있으며, 앞서 통합 카드데이터를 활용할 수 있는 공유체계까지도 마련될 수 있을 것이다.

본 연구의 한계를 정리하면 첫째, 시간상의 한계로 경기도 통행만을 고려하여 서울 순수통행, 인천 순수통행은 제외되었다. 따라서 경기도에서 경기도 혹은 서울과 인천으로 가는 통행은 모집단의 자료를 이용하였다고 할 수 있으나, 서울내부통행, 인천내부통행에 대해서는 다루지 못하였다.

둘째, 대량의 정보로 교통카드 활용의 가능성과 잠재성을 중점적으로 고려하였기 때문에 환승정류소의 유형에 대한 기준을 고려하지 못했다는 점이다. 즉, 환승수요를 구분함에 있어 한 정류소의 최대 허용수요는 승객의 도착분포 등을 고려하여 정해진다. 그러나 이와 같은 점은 고려하지 못하였다.

셋째, 본 연구에서 이용된 교통카드 데이터는 정산 후 데이터를 이용하였기 때문에 데이터 신뢰성 측면에서는 별도로 검토하지 않았다. 본 연구에서 언급한 바와 같이 개인정보를 제외한 원시데이터가 공유된다면, 경기도 대중교통정책을 위해서 신뢰성 검토를 수행할 필요가 있으며, 교통카

드 활용 측면을 다양하게 확대할 수 있을 것이다.

참고문헌

김기혁 · 이승철, 2011, 교통카드자료를 이용한 환승정류장의 유형별 입지특성에 관한 연구, 대한토목학회지논문집 31(4), pp. 519-526.

김순관, 2008, 교통카드 데이터를 활용한 OD추정 및 활용, 서울시정개발연구원.

김신근 · 박성용, 1999, Data Mining: 의사결정트리 알고리즘의 성과 비교에 관한 연구, 한국경영정보학회추계학술대회.

김종형 · 강미혜, 2010, 인천시 대중교통정책에 따른 교통수요변화연구, 인천발전연구원.

박진영 · 김동준, 2007, 서울시 교통카드자료의 특성 및 신뢰성 분석, 서울도시연구 8(4), pp. 127-138.

박진영 · 김동준, 2008a, 해외교통카드 현황 및 시사점: 옥토스카드와 이지링크카드, 월간교통 2008년 1월호.

박진영 · 김동준, 2008b, 일본의 교통카드 운영 현황 및 시사점, 월간교통 2008년 5월호.

빈미영, 2009, 경기도 버스서비스 개선방향, CEO Report 30, 경기개발연구원.

빈미영 · 이원도 · 정의석 · 조창현, 2012, 통행연계 변수를 중심으로 한 경기도 버스정류장 유형 구분, 한국경제지리학회지 15(2), pp. 331-341.

이기영 · 장명순, 2005, K-평균법을 이용한 고속도로 사고분석구간 분할기법 개발, 한국도로학회논문집 7(4), pp. 211-219.

이신해 · 김순관 · 김원호, 2010, 2009 대중교통 서비스지표 산출연구, 서울시정개발연구원.

성현곤 · 김동준 · 박지형, 2008, 서울시 역세권에서의 토

지이용 및 도시설계특성이 대중교통이용증대에 미치는 영향 분석, 대한교통학회지 26(4), pp135-147.

신성일, 2008, 교통카드Data를 활용한 대중교통 평가체계 구축방안, 서울시정개발연구원.

신성일 · 이창주 · 조용찬, 2007, 대중교통 카드 자료를 이용한 지역 간 대중교통 서비스 평가 체계, 2007년 한국ITS학회 추계학술대회.

조창현 · 이백진 · 빈미영, 2009, 도시 대중교통정보 이용 행동 특성 연구, 한국경제지리학회지 12(1), pp.55-66.

MacQueen, J. B., 1967, Some Methods for classification and Analysis of Multivariate Observations, Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press. pp. 281-297.

Pelletier, M. T. Martin and M. Catherine, 2011, Smart card data use in public transit: A literature review, Transportation Research Part C 19(4), pp. 557-568.

교신: 조창현, 서울시 동대문구 회기동 1번지, 경희대학교 이과대학 동관 504호, 전화: 02-961-9264, 이메일: bwchjoh@khu.ac.kr

Correspondence: Chang-Hyeon Joh, Room 504, Donggwan Bldg., College of Science, Kyung Hee University, 1 Hoegi-dong, Dongdaemun-gu, Seoul, 130-701, Korea, Tel: +82-2-961-9264, e-mail: bwchjoh@khu.ac.kr

최초투고일 2012년 11월 10일

최종접수일 2012년 11월 25일

Journal of the Economic Geographical Society of Korea
Vol.15, No.4, 2012(615-627)

A Study on Travel Pattern Analysis and Political Application using Transportation Card Data: In Gyeonggi-Do Case

Miyoung Bin* · Juback Moon** · Chang-Hyeon Joh***

Abstract : This study analyzed the travel pattern with respect to use of public transportation by using transportation card data and presented the measures that can be used in a traffic policy. Transportation card data targeted Gyeonggi-Do area and as a utilization plan, a scenario that when a traffic policy decision maker improves bus stop facilities, the person selects a target site by using several variables that can be obtained from transportation card data was set and analyzed. The analysis result showed that K means cluster analysis which is decision making methodology and CHAID(Chi-squared automatic interaction detection) were used and it can be used usefully in policies in significance level of $p < 0.01$. Also, based on these results, this study presented policy implications to be improved to actually use transportation card data in policies.

Key Words : Traffic Card data, K-means clustering, CHAID, travel patterns

This work was conducted as part of a policy research project, Gyeonggi Research Institute.

* Research Fellow, Department of Transportation Policy, Gyeonggi Research Institute

** Associate Research Fellow, Department of Transportation Policy, Gyeonggi Research Institute

*** Associate Professor, Department of Geography, Kyung Hee University