

## 교통카드 데이터를 활용한 사전태그 행태 분석

The Analysis of Passengers Tag Behavior using Transportation Card Data

---

저자 (Authors)	김경태, 민재홍, 이인목 Kyoung Tae Kim, Jae Hong Min, In Mook Lee
출처 (Source)	<a href="#">한국철도학회 학술발표대회논문집</a> , 2014.10, 635-638(4 pages)
발행처 (Publisher)	<a href="#">한국철도학회</a> The Korean Society For Railway
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE02502749">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE02502749</a>
APA Style	김경태, 민재홍, 이인목 (2014). 교통카드 데이터를 활용한 사전태그 행태 분석. 한국철도학회 학술발표대회논문집, 635-638
이용정보 (Accessed)	제주대학교 223.194.***.198 2020/09/04 01:05 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

## 교통카드 데이터를 활용한 사전태그 행태 분석

### The Analysis of Passenger's Tag Behavior using Transportation Card Data

김경태<sup>\*†</sup>, 민재홍<sup>\*</sup>, 이인묵<sup>\*\*</sup>

Kyoung Tae Kim<sup>\*†</sup>, Jae Hong Min<sup>\*</sup>, In Mook Lee<sup>\*</sup>

**Abstract** Transportation cards are widely used to travel by public transit in the Seoul metropolitan area. And, passengers have no resistance to transfer from one mode to the other mode because the public transportation fare is only proportional to the travel distance. Passengers have to tag when they get off the bus, but many passengers tag farther in advance. In this study, we analyzed the abnormal tag behavior of bus passengers by utilizing transport card data. It is analyzed that the ratio of tags in advance is equivalent to about 40.3% according to the suggested criteria. When you analyze average transfer distance or transfer behavior using transportation card data these factors should be considered sufficiently.

**Keywords :** Transportation card, Tag, Transfer, Public Transit, Big Data

**초 록** 수도권 대중교통을 이용하는 승객들은 거의 대부분이 교통카드를 사용하고 있다. 수도권의 대중교통요금이 통합거리비례요금제로 운영되고 있기 때문에 정확한 승하차 정보가 필요하다. 그러나 버스 이용객들은 하차역에서 내릴 때 교통카드를 태그하는 것이 아니라 그 이전에 태그하는 현상이 많이 발생하고 있다. 본 연구에서는 교통카드 데이터를 활용하여 버스 승객들이 비정상적으로 태그하는 행태를 분석하고자 하였으며, 본 연구에서 설정한 기준에 따르면 약 40.3%에 해당하는 승객이 사전에 태그하는 것으로 분석되었다. 따라서 교통카드 데이터를 활용한 대중교통 이용객들의 평균환승거리 산정 시에는 이러한 요인을 충분히 검토하여야 한다.

**주요어 :** 교통카드, 태그, 환승, 대중교통, 빅데이터

## 1. 서 론

최근 빅데이터에 대한 관심이 높아지면서 교통부문의 주요 빅데이터 중의 하나인 교통카드 데이터에 대한 분석이 활발하게 이루어지고 있다[1]. 빅데이터의 속성 상 수많은 정보를 가지고 있기도 하지만 자료의 생성 과정에서 포함된 오류가 있다고 하더라도 쉽게 찾기 힘든 경우가 많이 발생한다. 이러한 오류에는 단순히 기계적인 오류뿐만 아니라 물리적으로는 정상적으로 생성된 정보라고 하더라도 이 자료를 생성시키는 이용자가 왜곡된 정보를 생성시키는 경우도 포함된다. 그 대표적인 오류 중의 하나가 버스 이용객들의 사전 태그로서 이로 인해서 실제로 이용객이 탑승한 거리, 환승거리 등과 저장된 정보 사이에 차이가 발생하여 이 자료를 활용한 분석 결과의 신뢰도를 떨어뜨릴 수 있다.

† 교신저자: 한국철도기술연구원 교통체계분석연구팀(ktkim@krri.re.kr)

\* 한국철도기술연구원 교통체계분석연구팀

본 연구에서는 교통카드 자료를 활용하여 사전 태그 행태를 분석하고, 이렇게 왜곡된 자료가 환승거리에 어떠한 영향을 주는 지를 살펴보았다.

## 2. 본 론

### 2.1 분석자료

2013년 3월 18일 기준 서울시 버스와 지하철 이용객들의 교통카드 자료를 활용하였다. 전체 통행수는 10,032,903통행인 것으로 파악되었다. 지하철의 경우 게이트를 통과하는 방식이기 때문에 사전 태그의 문제가 발생하지 않으며, 사전 태그의 문제는 버스를 이용하고 있는 승객인 경우에 발생한다. 버스이용 승객이 사전에 태그를 하는 경우는 2가지로 구분해 볼 수 있다. 현재 이용하고 있는 버스에서 하차하는 것으로 통행이 끝나는 경우와 타 버스 또는 지하철로 환승하는 경우로 구분된다. 현재 버스에서 하차하는 것으로 통행이 끝나는 경우에는 사전 태그 여부를 파악하기가 곤란하지만 환승이 이루어진 경우에는 버스노선정보를 이용하여 추정할 수 있다.

따라서 지하철 내 환승은 환승횟수에 포함시키지 않고 통행을 환승횟수로 구분해 보면 직결통행이 전체의 76.6%, 1회 환승 20.4%, 2회 이상 환승이 3.0% 수준을 차지하고 있다. 전체 자료에서 본 분석에 활용한 자료는 2회 이상 환승인 통행에 대해서 버스를 이용하다가(선통행구간) 지하철 또는 버스로 환승하는(후통행구간) 통행을 선별하여 분석하였으며, 대상자료의 수는 476,966개에 달한다.

### 2.2 분석기준 및 결과

대상자료에서 사전에 태그한 경우와 정상적으로 태그한 경우의 구분은 다음의 기준을 적용하였다. 첫째, 각 선통행구간에서 하차한 정류장과 후통행구간의 승차정류장 사이의 직선거리(D)를 구한다. 둘째, 선통행구간에서 이용한 노선의 운행정보를 활용하여 하차한 정류장 이후의 정류장들과 후통행구간의 승차정류장과의 직선거리를 구하여, D보다 더 짧은 경우가 있을 때 사전에 태그한 통행으로 간주한다.<sup>1</sup>

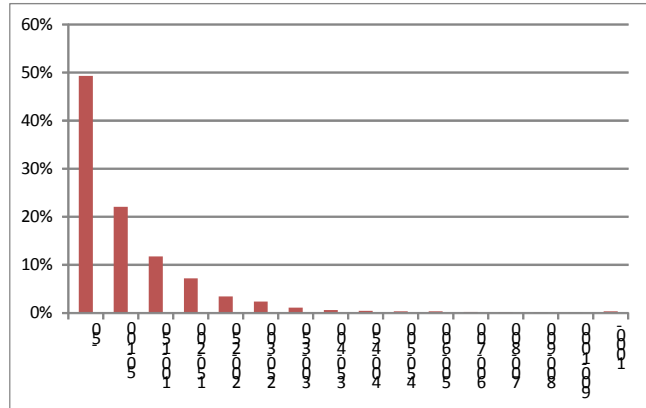
이상의 방법을 통해서 도출한 사전 태그 비율은 40.3%로서 192,042개에 달하는 것으로 분석되었으며, 정상 태그 비율은 59.7%인 284,924개이다.

정상적인 태그로 간주되는 통행의 경우(Table 1 참조) 환승거리가 50m 미만인 경우가 전체의 약 절반 수준인 49.3%를 차지하고 있으며, 200m 미만 90.3%, 300m 미만 96.2%, 400m 미만 98.0%, 500m 미만 98.8%를 차지하는 것으로 분석되었다. 500m 이상인 경우도 약 1.2% 차지하는 것으로 분석되었으며, 상식적으로 이해하기 힘든 환승거리가 1km 이상인 통행도 있었으나 그 비율은 매우 낮은 0.4% 수준인 것으로 분석되었다. 환승거리가 길어질수록 통행비율이 점진적으로 줄어드는 형태를 보여 직관적으로도 합리적인 결과인 것으로 판단할 수 있다.

<sup>1</sup> 본 연구에서 언급하는 환승거리는 실제 도로망에서의 도보거리가 아닌 정류장과 정류장 사이의 직선거리를 나타내는 것으로서, 일반적으로 실제 도보거리에 비해서는 더 짧은 거리임.

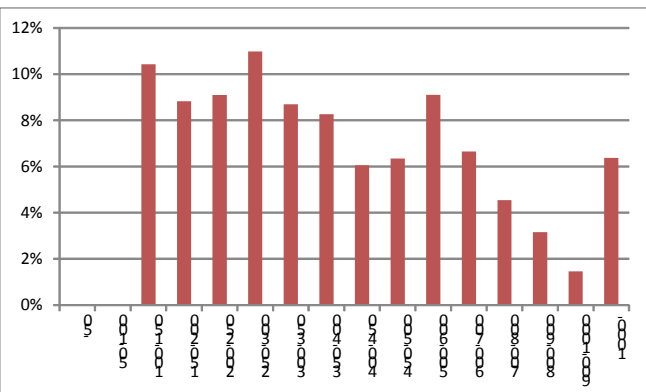
**Table 1** The number of trips by transfer distance using normal tag data

Transfer distance(m)	No. of trips	Ratio (%)	Cumulative Ratio (%)
d < 50	140,393	49.3	49.3
50 ≤ d < 100	62,956	22.1	71.4
100 ≤ d < 150	33,463	11.7	83.1
150 ≤ d < 200	20,506	7.2	90.3
200 ≤ d < 250	9,878	3.5	93.8
250 ≤ d < 300	6,810	2.4	96.2
300 ≤ d < 350	3,207	1.1	97.3
350 ≤ d < 400	1,871	0.7	98.0
400 ≤ d < 450	1,332	0.5	98.4
450 ≤ d < 500	1,014	0.4	98.8
500 ≤ d < 600	1,076	0.4	99.2
600 ≤ d < 700	676	0.2	99.4
700 ≤ d < 800	349	0.1	99.5
800 ≤ d < 900	196	0.1	99.6
900 ≤ d < 1000	133	0.0	99.6
1000 ≤ d	1,064	0.4	100.0
Sum	284,924	100.0	



**Table 2** The number of trips by transfer distance using pre-tag data

Transfer distance(m)	No. of trips	Ratio (%)	Cumulative Ratio (%)
d < 50	-	0.0	0.0
50 ≤ d < 100	-	0.0	0.0
100 ≤ d < 150	20,036	10.4	10.4
150 ≤ d < 200	16,951	8.8	19.3
200 ≤ d < 250	17,465	9.1	28.4
250 ≤ d < 300	21,090	11.0	39.3
300 ≤ d < 350	16,696	8.7	48.0
350 ≤ d < 400	15,867	8.3	56.3
400 ≤ d < 450	11,634	6.1	62.4
450 ≤ d < 500	12,198	6.4	68.7
500 ≤ d < 600	17,488	9.1	77.8
600 ≤ d < 700	12,760	6.6	84.5
700 ≤ d < 800	8,736	4.5	89.0
800 ≤ d < 900	6,060	3.2	92.2
900 ≤ d < 1000	2,809	1.5	93.6
1000 ≤ d	12,252	6.4	100.0
Sum	192,042	100.0	

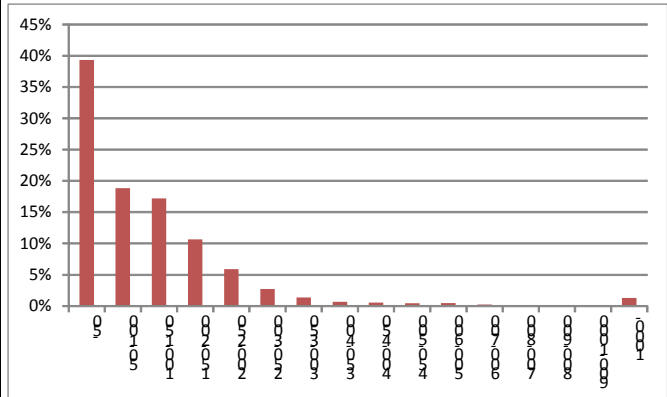


그러나 사전 태그로 간주된 통행의 경우에는(Table 2 참조) 이러한 형태를 보이지 않는다. 환승거리가 1km 이상인 통행이 약 6.4%나 차지하고 있으며, 구간대별로 통행량 비율이 불규칙적으로 나타나는 것을 알 수 있다. 구간별 비율을 보면 250m 이상 300m 미만인 경우가 가장 높은 11.0% 수준을 보이고 있으며, 환승거리가 길어질수록 통행수가 줄어드는 경향을 찾아볼 수 없다.

Table 1의 자료를 기준으로 평균환승거리를 계산해 보면 98m가 되지만, Table 2의 자료를 기준으로 평균환승거리는 577m로서 5.9배의 차이가 난다. 비정상적인 값이라고 판단되는 1km 이상의 자료를 제외하고 평균환승거리를 계산하더라도 각각 79m, 386m로서 차이는 4.9배에 이른다.

**Table 3** The number of trips by transfer distance using adjusted pre-tag data

Transfer distance(m)	No. of trips	Ratio (%)	Cumulative Ratio (%)
d < 50	75,534	39.3	39.3
50 ≤ d < 100	36,177	18.8	58.2
100 ≤ d < 150	33,004	17.2	75.4
150 ≤ d < 200	20,425	10.6	86.0
200 ≤ d < 250	11,309	5.9	91.9
250 ≤ d < 300	5,274	2.7	94.6
300 ≤ d < 350	2,632	1.4	96.0
350 ≤ d < 400	1,351	0.7	96.7
400 ≤ d < 450	1,031	0.5	97.2
450 ≤ d < 500	840	0.4	97.7
500 ≤ d < 600	952	0.5	98.2
600 ≤ d < 700	470	0.2	98.4
700 ≤ d < 800	273	0.1	98.6
800 ≤ d < 900	178	0.1	98.7
900 ≤ d < 1000	119	0.1	98.7
1000 ≤ d	2,473	1.3	100.0
Sum	192,042	100.0	



비정상 태그로 간주된 자료를 노선운행정보를 이용하여 보정하면 Table 3과 같은 결과를 얻을 수 있다. 표에서 보는 바와 같이 Table 1과 매우 유사한 형태를 보이는 것을 알 수 있다. 평균환승거리는 171m로 개선되어 정상적인 통행에 비해서 1.7배 수준이며, 1km 이상의 자료를 제외하고 계산하면 98m로서 1.2배 수준을 보인다.

### 3. 결 론

본 연구에서는 교통카드 데이터를 활용하여 환승거리를 산정할 경우 승객의 사전 태그 행태에 따라 어떠한 영향을 주는 지를 살펴보았다. 비록 전체통행 중에서 약 3%를 차지하고 있는 2회 이상의 환승에 한정하여 분석한 것이기는 하지만 단순히 교통카드 데이터 원자료를 활용하여 환승거리를 산정할 경우 오류가 발생할 가능성이 있음을 확인하였다. 1회 환승에 대해서도 동일한 분석을 시행하더라도 같은 결론을 얻을 수 있을 것으로 판단되지만 향후 확인할 필요성은 있다.

최근 빅데이터에 대한 관심이 높아지면서 이를 활용, 분석하는 연구가 활발하게 진행되고 있다. 본 연구에서 분석한 사례에서 알 수 있듯이 빅데이터를 활용하여 유용한 정보를 얻기 위해서는 빅데이터의 구조 및 특성 등에 대한 충분한 연구를 통해서 왜곡된 정보를 제외하거나 보정하는 것이 선행되어야 할 것이다.

### 참고문헌

- [1] I.M. Lee, J.H. Baek, J.H. Min(2013) The status of Research and Policy using Transportation Card Data, Railway Journal, 16(5), pp. 27-32.