

통행사슬 구조를 이용한 교통카드 이용자의 대중교통 통행종점 추정

신강원*

경성대학교 건설환경도시공학부

Inferring the Transit Trip Destination Zone of Smart Card User Using Trip Chain Structure

SHIN, Kangwon*

School of Civil, Urban, and Environmental Engineering, Kyungsoong University, Busan 48434, Rep. Korea.

*Corresponding author: kangwon@ks.ac.kr

Abstract

Some previous researches suggested a transit trip destination inference method by constructing trip chains with incomplete(missing destination) smart card dataset obtained on the entry fare control systems. To explore the feasibility of the transit trip destination inference method, the transit trip chains are constructed from the pre-paid smart card tagging data collected in Busan on October 2014 weekdays by tracing the card IDs, tagging times(boarding, alighting, transfer), and the trip linking distances between two consecutive transit trips in a daily sequences. Assuming that most trips in the transit trip chains are linked successively, the individual transit trip destination zones are inferred as the consecutive linking trip's origin zones. Applying the model to the complete trips with observed OD reveals that about 82% of the inferred trip destinations are the same as those of the observed trip destinations and the inference error defined as the difference in distance between the inferred and observed alighting stops is minimized when the trip linking distance is less than or equal to 0.5km. When applying the model to the incomplete trips with missing destinations, the overall destination missing rate decreases from 71.40% to 21.74% and approximately 77% of the destination missing trips are the single transit trips for which the destinations can not be inferable. In addition, the model remarkably reduces the destination missing rate of the multiple incomplete transit trips from 69.56% to 6.27%. Spearman's rank correlation and Chi-squared goodness-of-fit tests showed that the ranks for transit trips of each zone are not significantly affected by the inferred trips, but the transit trip distributions only using small complete trips are significantly different from those using complete and inferred trips. Therefore, it is concluded that the model should be applicable to derive a realistic transit trip patterns in cities with the incomplete smart card data.

Keywords: missing alighting stop, public transport, smart card, trip chain, trip destination

초록

본 연구는 선행연구에서 제시하고 있는 통행기점 정보만을 제공하고 있는 불완전한 대중교통카드 자료로부터 대중교통 통행의 종점을 통행사슬 구조를 이용하여 추정할 수 있는 모형의 국내 자료 적용 가능성을 살펴보고 모형 적용 결과를 제시하였다. 이를 위해 본 연구는 부산에서 2014년 10월 주중에 수집된 선불 교통카드 승·하차 태그 원시자료 1,846,252건을 대상으로 하루 동안 한 대중교통 이용자가 발생시킨 일련의 통행들을 시·공간적으로 연계시켜 통행사슬을

J. Korean Soc. Transp.
Vol.34, No.5, pp.437-448, October 2016
<http://dx.doi.org/10.7470/jkst.2016.34.5.437>

pISSN : 1229-1366
eISSN : 2234-4217

Received: 31 May 2016

Revised: 11 August 2016

Accepted: 27 October 2016

Copyright ©
Korean Society of Transportation

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

형성하고, 대중교통 이용자의 결측 종점을 연속된 다음 통행의 승차지점 또는 최초 승차지점이 속한 교통존으로 추정하였다. 모형 검증에 위해 대중교통 통행종점이 관측된 자료에 모형을 적용한 결과 실제 통행종점과 추정 통행종점의 일치도는 82.4%로 나타났으며 이 때 통행종점으로 추정된 정류장과 실제 하차 정류장간 거리의 오차는 최소가 되는 것으로 나타나 제안모형의 유용성은 높은 것으로 분석되었다. 통행사슬 구조를 이용한 통행종점 추정 모형을 종점결측 통행에 적용했을 때 종점결측 통행의 비율은 적용 전 71.40%(718,915통행)에서 21.74%(218,907통행)로 감소하였으며 종점추정이 불가능한 218,907통행의 대부분은 모형 적용이 불가능한 일일 통행횟수 '1회'인 통행(169,359통행, 77.37%)인 것으로 나타났고, 일일 통행횟수가 '2회 이상'인 통행의 종점결측 비율은 69.56%에서 모형 적용 후 6.27%로 크게 감소하였다. 한편 통행종점 추정 모형 적용에 따른 중간 통행 및 존내 통행분포의 변화를 비교하기 위해 순위상관계수 및 카이제곱 적합도 검정을 수행하였으며, 분석 결과 통행종점 추정 모형 적용에 따라 각 종준별 통행량의 순위는 변화하지 않으나 통행량 분포는 유의한 변화를 보였다. 따라서 통행사슬 구조를 이용한 교통카드 이용자의 통행종점 추정 모형 적용은 통행종점이 결측된 불완전 대중교통카드 자료가 수집되고 있는 도시의 대중교통 통행패턴을 보다 현실적으로 반영할 수 있게 도움을 줄 것으로 판단된다.

주요어: 하차결측, 대중교통, 교통카드, 통행사슬, 통행종점

서론

대중교통 운임 전자 결제용 교통카드는 1996년 서울특별시의 시내버스 카드인 U-Pass와 1998년 부산광역시의 하나로카드(시내버스와 도시철도 운임 통합 결제)를 시작으로 전국으로 확산되어 사용되고 있다. 2014년도를 기준으로 서울의 교통카드 이용률은 99.02%로 100%에 육박하고 있으며 부산의 교통카드 이용률은 93%로 한 해 동안 약 6.12억 통행이 생성되었다(Lee, 2015; Busan Transport Bureau, 2015). 이처럼 우리나라 대도시권의 교통카드 이용률은 90%를 상회하고 있어 많은 도시에서 교통카드 자료를 이용한 다양한 교통정책 수립 연구들이 수행되고 있다(Park et al., 2008; Shin and Choi, 2014; Lee, 2015). 그러나 각 도시에서 수집되고 있는 교통카드 자료는 운영 요금제에 따라 차이를 보이고 있다. 대중교통 이용자의 하차지점 결측이 대표적인 차이인데 7대도시 중 서울과 인천에서는 2009년 거리비례제 기반의 수도권 통합요금제 도입에 따라 각 승객의 승차지점 뿐 아니라 하차지점도 수집되고 있어 이를 활용한 교통정책 발굴이 활발하게 이루어지고 있으나 부산, 대구, 광주, 대전, 울산은 균일요금제를 기반 운임체제로 채택하고 있어 교통카드 이용자의 하차지점 정보가 누락되어 대중교통 통행의 종점(destination)은 알 수 없는 실정이다.

구체적으로 본 연구의 대상지역인 부산의 시내버스와 마을버스의 운임체계는 균일요금제이며 도시철도와 부산·김해경전철은 각각 이동구간 요금제(이동거리가 10km 이상일 때 2구간 요금 부과)와 구역 요금제가 적용되고 있으며 부산 대중교통 수단간 일반 환승 할인제도와 인접 도시인 김해시 및 양산시의 대중교통 수단간 광역 환승 할인제도가 병행되고 있다. 전술한 균일요금제 기반의 운임체제로 인해 부산의 시내버스 및 마을버스 이용자들 중 타 대중교통 수단 환승을 계획하지 않은 이용자는 차내 하차 단말기에 교통카드를 별도로 인식시킬 필요가 없다. 따라서 2014년 한 해 동안 부산에서 수집된 약 6.12억 통행 중 최종 탑승수단이 시내버스 또는 마을버스인 통행 4.06억 통행(66.4%)의 하차 위치는 결측되고 있다고 가정할 수 있다. 이처럼 부산에서 수집되고 있는 교통카드 자료는 수도권과 달리 하차지점이 결측된 불완전 구조를 형성하고 있어 교통카드 수송실적 중 승차정보(승차객수, 환승객수)만이 여객 OD 현행화에 단편 활용되고 있다.

따라서 대중교통 이용자(교통카드)의 통행종점(하차위치)를 추정하여 교통카드 수송실적 기반의 대중교통 OD를 구축하고 이를 이용한 대중교통 OD 현행화 방법론의 개선이 요구된다. 본 연구에서는 교통카드 수송실적을 이용한 대중교통 OD 현행화 방법론 개선을 위한 첫 번째 단계로 국외 선행연구에서 제시한 '통행사슬 구조를 이용한 교통카드 이용자의 통행종점 추정' 모형의 적용 가능성을 살펴보고, 모형을 이용한 대중교통 OD 구축결과를 제시하고자 한다.

선행연구

통행사슬(trip chain)이란 ‘집 또는 특정장소에서 출발하여 다시 그 장소로 되돌아오는 동안 발생한 일련의 목적 통행 집합’을 의미한다(Ji, 1999; Choo et al., 2008; Bin, 2011; Currie and Delbosc, 2011). 하루 동안 발생한 한 통행자의 통행들이 사슬구조를 이룬다는 것은 연속된 통행들이 발생 순서대로 연계되어 있음을 나타내므로 국외에서는 한 통행자의 연속된 교통카드 거래내역을 시·공간적으로 연계하여 대중교통 이용자들의 통행종점을 추정하는 연구가 비교적 활발하게 이루어져왔다.

Barry et al.(2002)은 미국 뉴욕시 메트로카드의 ID, 승차 시각 및 승차 위치 자료를 이용한 역간 OD 추정 방법을 제안하였는데, 이 연구는 ‘한 통행자의 하루 동안 연속된 지하철 통행의 기점은 직전 통행의 종점 인근이며, 최종 지하철 통행의 종점은 최초 지하철 통행의 기점 인근’이라는 통행사슬 구조 하에서 수행되었다. 이 연구는 통행사슬 구조 하에서 추정된 OD를 통행조사(travel diary survey) OD와 비교하였으며 분석 결과 약 90%가 일치한다고 보고했다. 또한 제안 방법을 2002년 9월에 수집한 자료에 적용한 결과 전체 통행 중 83%의 종점 추정이 가능함을 제시하였으며 이 추정 방법은 2004년 4월에 수집된 뉴욕시 대중교통수단(지하철, 버스, 페리, 트램)의 교통카드의 ID, 승차 시각, 승차 위치, 대중교통 배차간격 자료 등을 이용하여 개선되기도 하였다(Barry et al., 2009).

Zhao et al.(2007)는 미국 시카고 철도 자동요금징수(AFC, automatic fare collection)시스템을 통해 수집된 통행종점 결측 개별 거래 자료를 이용하여 철도 이용객들의 OD를 추정하였다. 이 연구 또한 철도 이용객의 종점 추정 시 통행사슬 구조를 가정하였고 선행연구에서 다루지 못했던 철도-버스 간 환승 통행의 종점을 추정하기 위해 연속된 대중교통 통행의 연계는 최소 0.4km(도보시간 5분)이내에서 이루어진다고 가정하였다(연속된 통행의 승차지점과 잠재 하차지점간 거리가 0.4km 이내일 때만 종점 추정). 2004년 1월에 수집된 자료에 제안 방법을 적용한 결과 전체 통행의 71.2%의 종점 추정이 가능했으며 종점 추정이 불가능한 나머지 28.8%의 대부분(18.6%)은 일일 통행횟수가 1회인 통행자의 통행인 것으로 나타났다.

통행사슬 구조를 이용한 대중교통 통행종점 추정 연구들은 약간의 개선을 거듭해왔는데, Trépanier et al.(2007)은 캐나다 퀘벡 주 가티뉴(Gatineau)시에서 2003년 7월과 10월에 수집된 버스 교통카드 자료(ID, 승차지점, 승차시각 등)를 이용하여 버스 통행의 종점을 추정하였다. 이 연구 또한 버스 통행의 종점을 통행사슬 구조(대중교통 통행 이력자료를 이용한 통행사슬 구축 포함)와 연속된 대중교통 통행의 연계는 2km 이내에서 이루어진다는 전제하에 추정하였다. 이 연구에서는 제안 방법 적용 시 전체 통행의 약 66%의 종점을 추정할 수 있고 침투 시간대의 통행 종점 추정비율은 약 80%로 다소 증가함을 제시하였다. Munizaga and Palma(2012)는 칠레 산티아고시의 도시철도와 버스 이용객의 종점을 전술한 대중교통 통행사슬 구조와 수단 간 통행 연계거리 1km를 이용하여 추정하였다. 이 연구에서는 제안 방법을 2009년 3월과 2010년 6월에 수집된 74백만 건의 자료에 적용하였으며, 그 결과 80% 이상의 통행종점을 추정할 수 있었으며 2개의 자료(2009년 3월, 2010년 6월)에서 추정된 OD 분포는 유사성을 갖고 있다고 보고하였다. He and Trépanier(2015)는 기점 정보만 제공되고 있는 캐나다 퀘벡 주 가티뉴시 버스 교통카드 자료(2009년 10월 자료)에 대중교통 통행연계거리 2km를 적용하여 통행 종점을 추정하였다. 이 연구에서는 통행 연계거리가 2km를 넘는 통행을 비연결통행(unlinked trip)으로 정의하고 해당 통행자의 이력자료를 이용하여 커널 밀도함수를 추정한 후 통행종점을 재추정하였다. 커널밀도함수를 이용한 통행종점 재추정 결과 통행사슬 구조만을 이용했을 때 약 80%였던 종점 추정률이 약 91%로 향상되었음을 제시하였다.

한편 몇몇 연구들은 통행사슬 구조를 이용한 대중교통 통행종점 추정 시 적정 통행연계거리에 대한 분석을 수행하기도 했다. He et al.(2015)은 호주 브리즈번시에서 수집된 40,431건의 대중교통 교통카드 기종점 자료를 이용하여 대중교통 통행의 종점 추정 시 적정 대중교통 통행연계거리를 분석하였다. 이 연구에서는 실제 하차지점과 통행사슬 구조를 이용한 추정 하차지점간 거리가 1km 이상이면 종점 추정률에 큰 변화가 없음을 밝혔다. Alsger et al.(2016)은 통행사슬 구조 기반의 대중교통 통행 종점 추정 알고리즘에 적용되는 통행연계거리가 연구자별로 상이함을 지적하고 호주 퀸즐랜드에서 2013년 3월에 수집된 버스, 철도, 페리 기종점 교통카드 자료를 이용하여 적정 통

행연계거리를 분석하였다. 이 연구에서는 대중교통 통행연계거리(allowable walking distance)를 0.4km, 0.8km, 1.0km, 1.1km로 구분하여 통행중점을 추정한 후 이를 실제 통행 기종점과 비교하였으며 통행연계거리가 0.8km 이상일 때 추정 통행량에 큰 변화가 없음을 밝혔다.

이처럼 국외에서는 통행사슬 구조를 이용한 대중교통 통행의 중점 추정연구가 비교적 활발하게 이루어지고 있으나 국내 연구사례는 그리 많지 않다. Park et al.(2008)은 서울특별시에서 2007년 3월에 수집된 교통카드 자료를 분석하여 하차 시 교통카드를 단말기에 접촉하지 않아 발생하는 ‘하차결측’이 전체 자료의 6.2-6.7% 수준에 이르고 있음을 밝혔고 이를 보정하기 위한 개인별, 노선별, 총량적 보정 방법의 개념을 제시하였다. 이 연구에서는 각 승차 정류장별 하차 정류장 이력자료를 이용하여 하차결측을 보정하는 노선별 ‘하차결측’ 보정방법을 1개 노선에 대해 적용하였는데 이는 무작위 결측(missing at random) 발생 메커니즘(하차결측 보정은 각 승차지점별 하차지점 비율에 의존)에 기반한 것으로 균일 요금제를 채택하고 있는 지역에서는 적용이 불가하다는 단점이 있다. Cho(2015)은 2013년 9월 한 달간 청주시에서 수집된 교통카드 자료를 이용하여 하차위치를 추정하고, 텍스트마이닝 기법인 잠재 디리클레 할당법(latent Dirichlet allocation, LDA)을 이용해 청주시 시내버스 이용자들의 이동패턴을 분석하였다. 이 연구에서는 교통카드 이용자들의 하차위치 추정을 위해 교통카드 원시자료와 버스정보시스템 DB를 비교하여 시간차가 적은 레코드에 정류장명을 매칭하여 노선정보를 추가한 후 통행사슬 구조를 이용하여 각 통행자의 하차위치를 단순 추정하였다. 추정 결과 전체 카드 데이터에서 약 68.4%의 데이터의 하차 위치를 추정할 수 있는 것으로 보고하였으나 각 승객의 정확한 탑승 노선정보 부재로 연속된 통행의 공간적 연계성에 대한 고려가 없었다는 한계점을 갖고 있다.

선행연구 분석 결과 국내에는 연속된 통행의 시공간적 연계성을 고려한 대중교통 중점 추정 연구가 거의 없으나 국외 여러 도시에서는 관련 연구가 활발하게 이루어지고 있음을 알 수 있다. 따라서 본 연구에서는 부산에서 수집되고 있는 통행중점이 결측됐거나 관측된 교통카드 자료를 이용하여 국외 선행연구에서 제시하고 있는 통행사슬 구조를 이용한 통행중점 추정을 수행한 후 이 모형의 적용 가능성을 검증하고 모형을 적용하여 구축한 대중교통 OD를 모형 적용 전 대중교통 OD와 비교·분석하였다.

대중교통 통행중점 추정 방법

1. 대중교통 통행

본 연구에서는 대중교통 통행(trip)을 ‘하나의 목적을 갖고 대중교통수단에 승차하여 필요 시 환승을 거쳐 목적지에 하차하기까지의 이동 행위’로 정의하였다. 따라서 한 통행자가 하루 동안 형성할 수 있는 대중교통 통행의 개수 범위는 ‘0-n’이며 대중교통 통행이 ‘0개’인 통행자는 ‘하루 동안 대중교통을 전혀 이용하지 않은 통행자’이다. 한편 각 대중교통 통행은 최소 ‘1회’ 또는 최대 ‘환승할인 허용 횟수+1회’만큼의 대중교통 부분 통행(trip segment)들로 구성된다. 본 연구의 대상지인 부산광역시 환승할인 허용 횟수가 최대 2회인데, 한 사람이 2회의 대중교통 환승을 거쳐 목적지에 하차했다면 그 사람의 대중교통 통행은 총 3개의 대중교통 부분 통행으로 구성된다.

2. 대중교통 통행중점 추정 모형

한 대중교통 이용자가 하루 동안 발생시킨 n개의 대중교통 통행들이 서로 연계되어 있다면 그 통행자의 i번째 통행의 종점과 i+1번째 통행의 기점은 매우 근접해있을 것이며 그 통행자의 최종 통행(n번째 통행)의 종점과 최초 통행의 기점은 매우 근접해 있을 것이다. Figure 1은 하차 결측이 발생한 3개의 대중교통 통행으로 구성된 대중교통 통행사슬 구조를 나타내는데, 각 통행의 기점(승차지점)은 B1, B2, B3이나 각 통행의 종점(하차지점)은 알려져 있지 않다. 각 통행의 하차지점은 선행연구(Barry et al., 2002; Zhao et al., 2007; Trépanier et al., 2007; Munizaga and Palma, 2012; He and Trépanier, 2015; Alsgjer et al., 2016)에서 제시한 통행사슬 구조를 이용하면 각 통행의

최종 탑승 노선의 정류장(버스 정류소 및 역을 통칭) 중 다음 통행 또는 최초 통행의 승차지점에서 가장 가까운 정류장인 A_1, A_2, A_3 로 추정할 수 있다.

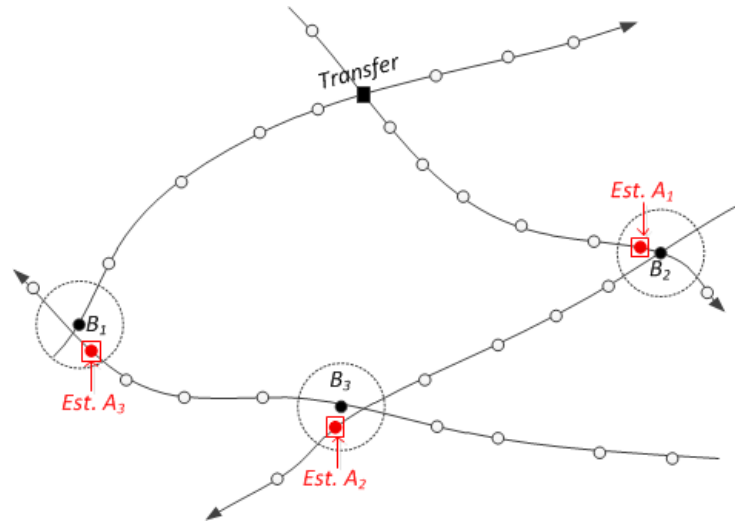


Figure 1. Concept for inferring alighting stops using transit trip chain structure

그러나 대중교통 이용자의 다음 통행 또는 최초 통행의 승차지점과 잠재 하차지점(최종 탑승 노선의 정류장 집합) 간 최소 거리가 수락할 수 없을 정도로 길다면 하차지점을 추정하고자 하는 통행은 연속된 다음 통행과 공간적 연계성을 갖고 있지 않으므로 추정된 하차지점은 불확실성을 갖는다. 따라서 연속된 통행들의 승차지점과 잠재 하차지점 간 최소 거리는 대중교통 통행의 종점 추정 시 반드시 검토되어야 하며 본 연구에서는 이를 ‘통행연계거리(trip linking distance)’로 정의하고 통행연계거리가 수락할만한 거리(수락 통행연계거리 : acceptable trip linking distance)내에 있을 때 해당 통행의 하차지점을 통행사슬 구조를 이용하여 추정하였다(통행연계거리 증가 시 연속된 두 통행이 대중교통수단만으로 연계될 가능성은 감소함을 반영). 한편 전술한 정류장 수준(stop-level)의 통행 종점 추정 방법을 본 연구의 목적에 부합하는 교통존 수준(TAZ-level)의 통행 종점 추정 방법으로 단순화시키면 각 통행의 통행종점(교통존)은 연속된 통행의 통행연계거리가 수락할만한 거리 내에 있을 때 다음 통행 또는 최초 통행의 승차지점이 속한 교통존으로 추정되며 이는 아래와 같이 정리된다.

한 통행자가 하루 동안 발생시킨 연속된 n 개의 대중교통 통행의 하차지점이 결측되어 있을 때 i 번째 통행과 $i+1$ 번째 통행 간 통행연계거리는 Equation 1과 같이 i 번째 통행의 최종 탑승노선의 정류장 집합 \mathcal{S}_i 와 $i+1$ 번째 통행 승차지점인 B_{i+1} 간 최소 거리(km)로 정의된다.

$$d_{i,i+1} = \min \{d(\mathcal{S}_i, B_{i+1})\}, \quad (1)$$

여기서 \mathcal{S}_i 는 i 번째 통행에 최종 사용된 노선이 경유하는 정류장 집합(i 번째 통행의 최종 승차 정류장 이후 하차 가능 정류장 집합)이며, $d(\mathcal{S}_i, B_{i+1})$ 는 i 번째 통행에 최종 사용된 노선의 정류장 집합 \mathcal{S}_i 와 $i+1$ 번째 통행 승차지점인 B_{i+1} 간 유클리드 거리 집합, $d_{i,i+1}$ 는 i 번째 통행과 $i+1$ 번째 통행의 통행 연계거리로 $d(\mathcal{S}_i, B_{i+1})$ 의 최솟값을 의미한다. 또한 최종 통행(n 번째 통행)과 최초 통행 간 통행 연계거리 $d_{n,1}$ 는 Equation 2와 같이 n 번째 통행에 최종 사용된 노선이 경유하는 정류장 집합(n 번째 통행의 최종 승차 정류장 이후 하차 가능 정류장 집합) \mathcal{S}_n 과 최초 통행 승차지점인 B_1 간 유클리드 거리의 최솟값으로 산출된다.

$$d_{n,1} = \min \{d(\mathcal{S}_n, B_1)\}. \quad (2)$$

본 연구에서는 각 통행의 통행종점(교통존)을 Equation 1과 2에서 제시한 통행연계거리가 수락 통행연계거리 (d_a , 본 연구에서는 통상적인 역세권 반경인 0.5km 적용) 보다 짧거나 같을 때 연계된 통행의 승차지점이 속한 교통존으로 추정하였다. 이는 각 통행이 0.5km 이내에서 연계되어 있다면 전담 통행의 목적은 후담 또는 최초 통행의 기점에서 이루어졌다는 가정을 전제로 하고 있다. 구체적으로 i번째 통행의 종점(교통존)은 Equation 3과 같이 통행연계거리가 수락 통행연계거리보다 짧거나 같을 때 i+1번째 통행의 승차지점 B_{i+1} 가 속한 교통존 O_{i+1} 로 추정된다.

$$\hat{D}_i = O_{i+1} \quad \text{if } d_{i,i+1} \leq d_a, \quad (3)$$

여기서 \hat{D}_i 은 통행사슬 구조 하에서 추정된 i번째 통행의 종점(교통존)이며, O_{i+1} 는 i+1번째 통행의 승차지점 B_{i+1} 이 속한 교통존, $d_{i,i+1}$ 는 i번째 통행과 i+1번째 통행의 통행연계거리(km), d_a 는 수락 통행연계거리(km)이다. 즉 i번째 통행의 종점은 i번째 통행과 연속된 i+1번째 통행의 연계거리가 충분히 짧아 공간적으로 연계되었다고 판단됐을 때 i+1번째 통행의 승차지점이 속한 교통존으로 교환 추정된다. 최종 통행(n번째 통행)의 종점(D_n) 또한 Equation 4와 같이 최종 통행과 최초 통행 간 통행연계거리 $d_{n,1}$ 이 d_a 보다 짧거나 같을 때 최초 통행의 승차지점 B_1 이 속한 교통존 O_1 으로 추정하였다.

$$\hat{D}_n = O_1 \quad \text{if } d_{n,1} \leq d_a. \quad (4)$$

대중교통 통행종점 추정

1. 자료

통행사슬 구조를 이용한 교통카드 이용자의 통행종점 추정을 위해 본 연구에서는 2014년 10월 주중(10/14-10/17)에 부산에서 수집된 선불 교통카드 승·하차 태그 원시자료 2,936,538건 중 단순오류(중복태그, 통행자의 각 통행의 첫 번째 부분 통행의 승·하차 정보가 하차 또는 환승통행인 경우, 각 부분 통행의 승차 지점 ID가 누락된 경우, 통행수단이 도시철도 또는 경전철임에도 하차 지점 ID가 누락된 경우 등)를 포함하는 교통카드 이용자와 정류장 좌표정보 자체가 없는 마을버스(부산, 김해, 양산) 및 시내버스(김해, 양산)를 탑승한 교통카드 이용자의 모든 통행을 제외한 1,846,252건을 사용하였다(자료의 정제, 가공, 분석은 Stata MP 13과 TransCAD 6.0을 이용).

부산의 선불 교통카드 이용률은 45%(2014년도 기준)로 비교적 높은 편이며 승·하차 태그 건별 원시자료는 각 승차자의 카드 ID, 승차 수단 등 총 13개의 변수를 포함하고 있다. 본 연구에서는 해당 원시자료를 각 통행자의 카드 번호, 통행의 승·하차 및 환승승차 시각, 환승 유무 등을 이용하여 총 1,006,836 통행으로 전환시켰다. Table 1은 각 통행을 탑승수단별로 분류한 결과로 시내버스만 탑승한 통행이 569,645 통행으로 가장 많았으며 도시철도 단독탑승(179,323 통행), 시내버스 2회 환승통행(155,241 통행)의 순으로 나타났다. 전체 자료 중 종점결측이 있는 통행은 총 718,915 통행으로 전체 통행의 71.40%를 차지하는 것으로 나타났으며 각 탑승유형을 주수단으로 변경시켜 종점결측 비율을 살펴보면 시내버스는 92%, 도시철도는 12%로 나타났다.

Table 1. Trips with missing destination by boarding type

Type	Boarding Mode			Total trips	Trips with missing destination	
	1 st Boarding	2 nd Boarding	3 rd Boarding		Trips	Percent
1	Bus			569,645	523,001	91.81
2	Metro			179,323	0	0.00
3	Bus	Bus		155,241	144,009	92.76
4	Bus	Metro		26,477	0	0.00
5	Metro	Bus		25,985	24,438	94.05
6	Bus	Bus	Bus	21,369	20,214	94.59
7	LRT			11,571	0	0.00
8	Bus	Metro	Bus	3,844	3,651	94.98
9	LRT	Metro		3,171	0	0.00
10	Metro	LRT		3,122	0	0.00
11	Metro	Bus	Bus	1,563	1,438	92.00
12	LRT	Bus		1,455	1,349	92.71
13	Bus	LRT		1,450	0	0.00
14	Bus	Bus	Metro	1,143	0	0.00
15	LRT	Metro	Bus	438	396	90.41
16	Bus	Metro	LRT	346	0	0.00
17	LRT	Bus	Bus	229	212	92.58
18	Bus	Bus	LRT	170	0	0.00
19	Bus	LRT	Bus	142	138	97.18
20	Metro	LRT	Bus	71	69	97.18
21	Bus	LRT	Metro	70	0	0.00
22	LRT	Bus	Metro	8	0	0.00
23	Metro	Bus	LRT	3	0	0.00
Main mode : Bus(Type 1 + Type 3 + Type 6)				746,255	687,224	92.09
Main mode : Metro(all types excluding Type 1, 3, 6)				260,581	31,691	12.16
Total				1,006,836	718,915	71.40

통행사슬 형성을 위해 본 연구에서는 먼저 교통카드 ID, 승차 및 환승시각 정보를 이용하여 한 교통카드 이용자가 하루 동안 발생시킨 각 통행의 시간적 연계(temporal linking)를 수행하였다. 그 결과 Table 2에 제시한 바와 같이 일일 통행횟수가 '1회'인 통행은 총 216,770통행이며 통행사슬 구조 적용이 가능한 일일 통행횟수가 '2회 이상'인 통행은 총 790,066통행으로 나타났고 이 중 549,556 통행(69.6%)에서 종점결측이 발생한 것으로 나타났다. 한편 분석자료에서 총 교통카드 통행자 수(교통카드 ID 기준)는 568,465명이고 일일 통행횟수가 2회 이상인 통행자는 총 351,695명으로 나타났다. 이 중 모든 통행의 하차 정보가 관측된 통행자는 64,080명(18%)이며, 하차 정보가 한 번이라도 결측된 통행을 가진 통행자는 총 351,695명 중 287,615명(82%)으로 나타났다.

Table 2. Trips with missing destination by daily trip frequency

Daily trip frequency	Tagging frequency	Total trips (A)	Trips with missing destination	
			Trips(B)	Percent : (A/B)*100
1	368,987	216,770	169,359	78.13
2 over	2	1,061,552	408,206	71.33
	>2	415,713	141,350	64.90
Sub-total		1,477,265	549,556	69.56
Total		1,846,252	718,915	71.40

통행사슬 구조를 이용한 대중교통 통행의 종점 추정을 위해서 본 연구에서는 각 통행의 시간적 연계 뿐 아니라 각 통행의 통행연계거리(연속된 다음 통행의 승차지점 또는 최초 승차지점과 탑승 노선의 정류장 집합인 잠재 하차지점 간 최소 거리)와 수락 통행연계거리(0.5km)를 이용한 시간적으로 연속된 대중교통 통행의 공간적 연계(spatial linking)도 고려하였다. 이 때 한 대중교통 이용자가 하루 동안 발생시킨 통행들의 통행연계거리는 총 14,734개의 부산 시내버스 정류장, 도시철도 역, 경전철 역간 유클리드 거리(217,090,756 셀 = 14,734 × 14,734)를 이용하여 산출하였다.

2. 통행종점 추정 모형 검증

통행사슬 구조를 이용한 통행종점 추정모형의 검증을 위해 일일 통행횟수가 2회 이상이고 각 통행의 통행종점 결측이 없는 64,080명의 138,581통행을 이용하였다. 통행종점 결측이 없는 통행 중 대부분(91.35%, 126,587통행)은 주수단이 도시철도로 본 연구의 추정대상인 시내버스 통행과의 통행연계특성이 다를 것으로 판단되나 동일 지역 내 대중교통 통행이라는 공통점이 있으므로 이를 검증대상으로 삼았다. 모형 검증은 제안 모형을 통해 추정된 소존 단위의 통행종점과 실제 통행종점의 일치성과 통행종점으로 추정된 정류장(다음 통행의 승차 정류장 및 최초 승차 정류장)과 실제 하차 정류장(역)간 거리를 이용하여 수행하였다.

Table 3은 통행연계거리(d)별 통행종점 일치성과 추정 정류장과 실제 하차 정류장간 거리에 대한 분석 결과로 전체 자료의 78.9%에 달하는 109,395통행의 실제 통행종점과 추정 통행종점이 동일한 것으로 나타났다. 추정된 통행종점과 실제 통행종점의 일치성은 통행연계거리가 길어질수록 감소하는 경향을 보이며 통행종점으로 추정된 정류장과 실제 하차 정류장간 거리는 통행연계거리가 길어질수록 증가하는 경향을 보였다. 본 연구에서 수락 통행연계거리로 설정한 0.5km 이내의 통행종점 일치성은 82.4%로 가장 높게 나타났으며 이 때 통행종점으로 추정된 정류장과 실제 하차 정류장간 거리의 평균절대오차(MAE, mean absolute error)와 평균제곱근오차(RMSE, root mean squared error)는 각각 0.62km와 2.28km로 가장 짧게 나타났다. 한편 추정된 하차 정류장과 실제 하차 정류장간 거리가 '0'인 통행(정류장 수준의 추정결과가 일치하는 통행)은 총 99,432통행으로 이 통행은 모두 수락 통행연계거리(0.5km) 구간에서 나타났으며 이는 본 모형 적용 시 정류장 수준의 모형 정확도가 75.23%에 달하는 것을 의미한다.

Table 3. Trip destination estimating model validation results

d	Total trips	Estimation accuracy		Difference in distance b/w real and estimated alighting stops	
		Trips	Percent	MAE	RMSE
d≤0.5	132,172	108,874	82.4	0.62	2.28
0.5<d≤1.0	2,157	320	14.8	1.81	3.81
1.0<d≤1.5	1,071	112	10.5	2.14	4.20
1.5<d≤2.0	661	19	2.9	2.94	5.29
2.0<d≤2.5	393	14	3.6	3.16	4.99
d>2.5	2,127	56	2.6	3.35	5.39
All	138,581	109,395	78.9	0.72	2.45

전술한 대중교통 통행종점 추정모형 검증결과 수락 통행연계거리를 통상적인 역세권 설정반경인 '0.5km'로 설정했을 때 모형의 정확도(교통존 수준)는 82.4%인 것으로 나타났다. 모형의 정확도는 현재 가용할 수 없는 대중교통망 자료(노선별 GIS 자료 및 배차간격 등)를 통행종점 추정의 제약조건으로 추가하면 보다 향상될 수 있을 것으로 기대되며, 하차결측이 발생한 자료를 기종점 추정을 위해 전혀 사용되지 못하고 있는 현실에서 대중교통 통행의 기종점 추정을 위한 현 모형의 유용성은 매우 높다고 판단된다.

3. 통행중점 추정 결과

1) 통행중점 추정 통행량

통행사슬 구조를 이용한 통행중점 모형을 일일 통행횟수가 2회 이상이고 종점결측이 발생한 총 549,556통행에 적용한 결과 이 중 500,008통행(90.98%)의 통행중점을 모형을 이용하여 추정 가능한 것으로 나타났다. 이러한 결과는 종점결측이 발생한 대중교통 통행의 90.98%가 연속된 다음 통행의 승차지점 또는 최초 승차지점을 중심으로 0.5km이내에서 연계되고 있는 것을 의미한다. 결과적으로 제안 모형 적용 시 전체 분석 자료 중 종점결측 통행의 비율은 Table 4에 제시한 바와 같이 71.40%(718,915통행)에서 21.74%(218,907통행)로 감소하였으며 종점추정이 불가능한 218,907통행의 대부분은 모형 적용이 불가능한 일일 통행횟수가 '1회'인 통행(169,359통행, 77.37%)인 것으로 나타났다. 한편 일일 통행횟수가 '2회 이상'인 통행의 종점결측 비율은 모형 적용 후 69.56%에서 6.27%로 크게 감소하였으며 일일 통행횟수가 '3회 이상'인 통행의 종점결측 비율은 6.74%로 일일 통행횟수가 '2회'인 통행의 종점결측 비율(6.09%)에 비해 다소 높았다.

Table 4. Trip destination estimation results

Daily trip frequency	Total trips	Trips with missing destination		Percent of trips with missing destination	
		Before	After	Before	After
1	216,770	169,359	169,359	78.13	78.13
2 over	2	572,262	408,206	71.33	6.09
	>2	217,804	141,350	64.90	6.74
Sub-total	790,066	549,556	49,548	69.56	6.27
Total	1,006,836	718,915	218,907	71.40	21.74

5) 중존별 유출입 통행량

본 소절에서는 모형 적용을 통해 추정된 500,008통행의 종점 정보를 이용하여 구축된 주수단 개념의 대중교통 통행 기종점 통행량 중 부산광역시 내부 통행량을 중존(16개 구·군)단위로 집계하여 이를 모형 적용 전 기종점 유출입 통행량과 비교한 결과를 제시하였다. Table 5는 통행중점 추정 모형 적용 전과 후의 각 중존별 주수단 버스 및 도시철도의 유출입 통행량을 비교한 표로 버스 통행량은 56,707통행(추정 전)에서 516,450통행(추정 후)으로 도시철도 통행량은 202,937통행(추정 전)에서 223,981통행(추정 후)으로 증가했다. 통행중점 추정 모형 적용 시 버스 통행 증가율은 810.73%(459,743통행)로 매우 높았으나 도시철도 통행 증가율은 10.37%(21,044통행)에 그쳤는데, 이는 통행중점 추정 모형이 적용된 대부분의 통행은 하차결측이 발생하고 있는 버스통행이기 때문이다.

통행중점 추정 모형 적용에 따른 중간 통행 및 존내 통행분포의 변화를 비교하기 위해 모형 적용 전·후의 소존 통행량을 중존단위로 집계한 후 16개 구·군의 수단별 유출통행량과 유입통행량의 차이를 순위상관계수(Spearman's rho or Spearman rank correlation coefficient)와 카이제곱 적합도(χ^2 Goodness-of-fit) 검정을 통해 비교하였다. 순위상관계수는 통행중점 추정 모형 적용 전·후 각 중존의 유출입 통행량의 순위(rank) 상관도를 나타내는 것으로, 분석 결과 통행중점 추정 모형 적용 전·후의 각 구·군별 통행량 순위의 상관관계수는 0.77-1.00로 강한 양의 상관관계를 갖는 것으로 나타났고 통행중점 추정 모형 적용은 통행량 순위에 영향을 주지 않는 것으로 나타났다($\alpha = 0.05$). 카이제곱 적합도 검정은 통행중점 추정 모형 적용 전 각 중존의 유출입 통행량 분포가 통행중점 추정 모형 적용 후 유출입 통행량의 분포를 따르는지를 살펴보기 위해 수행한 것으로, 분석 결과 추정 모형 적용 전 중존별 유출입 통행량의 분포는 모형 적용 후 중존별 유출입 통행량의 분포와 유의한 차이를 보이는 것으로 나타났다. 이러한 분석결과는 통행중점 추정 모형 적용에 따라 각 중존별 통행량의 순위는 변화하지 않으나 통행량 분포는 달라짐을 나타낸다.

Table 5. OD trip comparison results by zone

Zone	Bus				Metro			
	O		D		O		D	
	Before	After	Before	After	Before	After	Before	After
Gangseo-gu	508	3,422	268	3,405	4,067	4,761	4,032	4,282
Geumjeong-gu	2,555	26,112	2,341	25,608	20,439	22,956	21,877	22,882
Gijang-gun	1,571	12,454	1,074	12,526	1,569	1,756	613	1,420
Nam-gu	3,972	35,437	2,922	35,051	14,683	16,006	13,536	15,838
Dong-gu	2,696	26,226	3,625	27,095	11,006	12,056	11,489	12,130
Dongnae-gu	3,422	34,245	4,324	33,963	17,319	18,887	18,574	19,961
Busanjin-gu	7,441	72,797	8,828	72,177	28,326	30,994	30,038	32,547
Buk-gu	3,439	29,658	3,472	29,800	16,467	17,950	16,330	17,656
Sasang-gu	5,160	48,718	4,531	48,730	14,257	15,479	13,643	15,190
Saha-gu	4,780	43,287	4,646	44,128	12,065	13,788	10,629	12,104
Seo-gu	3,646	32,238	3,332	32,284	5,778	6,530	5,382	6,431
Suyeong-gu	2,564	22,105	2,665	21,833	9,344	9,972	9,634	10,182
Yeonje-gu	3,303	28,417	3,082	27,609	18,806	20,764	19,802	21,384
Yeongdo-gu	4,348	31,986	3,115	31,829	2,694	3,014	172	2,315
Jung-gu	2,266	25,174	4,349	25,780	8,459	9,595	10,163	10,741
Haeundae-gu	5,036	44,174	4,133	44,632	17,658	19,473	17,023	18,918
Total	56,707	516,450	56,707	516,450	202,937	223,981	202,937	223,981
Spearman's rho	0.9618(<0.001)		0.7706(<0.001)		1.0000(<0.001)		0.9941(<0.001)	
χ^2	527.298(<0.001)		1903.08(<0.001)		69.8103(<0.001)		2381.41(<0.001)	

결론 및 향후 연구과제

본 연구는 균일요금제 채택에 따라 하차위치 정보 결측이 다수 발생하고 있는 도시에서 수집되고 있는 불완전 대중교통카드 자료(trip origin-only data)에 대중교통 통행사슬 구조를 적용하여 대중교통 통행의 종점을 추정하기 위한 목적으로 수행되었다. 이를 위해 본 연구에서는 부산에서 2014년 10월 주중에 수집된 선불 교통카드 승·하차 태그 원시자료 1,846,252건을 대상으로 하루 동안 한 대중교통 이용자가 발생시킨 일련의 통행들을 시·공간적으로 연계시켜 통행사슬을 형성하였다. 구체적으로 한 대중교통 이용자가 발생시킨 대중교통 통행들의 시간적 연계는 교통카드 ID, 승차 및 환승시각 정보를 이용하여 발생 순서대로 연계하였으며, 탑승 시각 순으로 연계된 통행들의 공간적 연계성은 각 통행의 통행연계거리(연속된 다음 통행의 승차지점 또는 최초 승차지점과 탑승 노선의 정류장 집합인 잠재 하차지점 간 최소 거리)와 수락 통행연계거리(0.5km)를 비교하여 검토하였다. 정제된 대중교통 통행사슬 자료를 이용하여 대중교통 이용자의 결측 종점은 연속된 다음 통행의 승차지점 또는 최초 승차지점이 속한 교통존으로 추정하였으며 본 연구를 통해 도출된 결론은 아래와 같이 요약할 수 있다.

- 부산의 대중교통카드 자료 분석 결과 전체 자료 중 종점결측(하차결측)이 있는 통행은 전체 통행의 약 71% (718,915 통행)를 차지하고 있고 특히 시내버스 주수단 통행의 종점결측 비율은 92%로 매우 높아 현재 교통카드 자료 기반의 대중교통 OD는 왜곡된 대중교통 통행행태를 반영할 가능성이 높음.
- 대중교통 통행종점이 관측된 자료(일일 통행횟수 ≥ 2 회)에 제안 모형을 적용한 결과 실제 통행종점과 추정 통행종점의 일치도는 82.4%로 나타났으며 이 때 통행종점으로 추정된 정류장과 실제 하차 정류장간 거리의 오차는 최소가 되는 것으로 나타나 제안모형의 유용성은 매우 높은 것으로 나타남.

- 통행사슬 구조를 이용한 통행종점 추정 모형 적용 시 종점결측 통행의 비율은 적용 전 71.40%(718,915통행)에서 21.74%(218,907통행)로 감소하였으며 종점추정이 불가능한 218,907통행의 대부분은 모형 적용이 불가능한 일일 통행횟수가 '1회'인 통행(169,359통행, 77.37%)인 것으로 나타났고, 일일 통행횟수가 '2회 이상'인 통행의 종점결측 비율은 69.56%에서 모형 적용 후 6.27%로 크게 감소하였음.
- 통행종점 추정 모형 적용에 따른 중간 통행 및 존내 통행분포의 변화를 비교하기 위해 순위상관계수 및 카이제곱 적합도 검정을 수행하였으며, 분석 결과 통행종점 추정 모형 적용에 따라 각 중존별 통행량의 순위는 변화하지 않으나 통행량 분포는 유의한 변화를 보였음.
- 따라서 통행사슬 구조를 이용한 교통카드 이용자의 통행종점 추정 모형 적용은 통행종점이 결측된 불완전 대중교통카드 자료가 수집되고 도시의 대중교통 통행패턴을 보다 현실적으로 반영할 수 있게 도움을 줄 것으로 판단됨.

이처럼 본 연구는 통행사슬 구조를 이용하여 통행기점 정보만을 제공하고 있는 불완전한 교통카드 자료로부터 대중교통 통행의 종점을 추정할 수 있는 방법론을 정리하고 국내 자료를 이용한 추정모형의 정확도와 모형 적용에 따른 OD 분포 변화 분석 결과를 제시하였다. 그러나 본 연구는 대중교통 상세자료(모든 노선의 정류장 위치정보 및 배차간격 등)의 부재로 정류장 수준의 통행종점 추정 결과를 상세 비교분석하지 못한 한계가 있다. 따라서 향후에는 구축 중인 대중교통 상세자료를 이용하여 수락 통행연계거리 내 정류장의 개수와 차내통행시간 등을 반영한 정류장 수준의 통행종점 추정 모형의 비교·분석이 요구된다. 또한 좀 더 많은 통행종점 관측자료를 이용한 수락 통행연계거리 특성 분석을 통해 통행수단, 통행 시간대, 통행 지역별 수락 통행연계거리 도출도 필요할 것으로 판단된다. 끝으로 선별 교통카드 자료 뿐 아니라 후불 교통카드 자료를 이용하여 일일 통행횟수가 1회인 통행자의 종점을 추정할 수 있는 빅데이터 기반의 대중교통 통행종점 추정모형의 개선이 필요하며, 이를 통한 대중교통 OD 현행화 방법론의 개선과 대중교통 정책 발굴이 뒤따라야 할 것이다.

ACKNOWLEDGEMENT

This research was supported by Kyungshung University Research Grants in 2016.

알림 : 본 논문은 대한교통학회 제74회 학술발표회에서 발표된 내용을 수정·보완하여 작성된 것입니다.

REFERENCES

- Alsger A., Assemi B., Mesbah M., Ferreira L. (2016), Validating and Improving Public Transport Origin-Destination Estimation Algorithm Using Smart Card Fare Data, *Transportation Research Part C: Emerging Technologies*, 68, 490-506.
- Barry J., Freiner R., Slavin H. (2009), Use of Entry-only Automatic Fare Collection Data to Estimate Linked Transit Trips in New York City, *Transport. Res. Rec.: J. Transport. Res. Board* 2112, 53-61.
- Barry J.J., Newhouser R., Rahbee A., Sayeda S. (2002), Origin and Destination Estimation in New York City With Automated Fare System Data, *Transportation Research Record* 1817, 183-187.
- Bin M. (2011), A Study on Trip Chain Typed Selection Behavior, *J. Korean Soc. Transp.*, 29(3), Korean Society of Transportation, 7-19.
- Busan Transport Bureau (2015), Busan 2014 Transportation Report, Busan Metropolitan City.
- Cho A. (2015), Mobility Pattern Analysis of Bus Passengers With LDA. Master's Degree Dissertation, Chungbuk National University.

- Choo S., Kwon S., Kim D. (2008), Exploring Characteristics on Trip Chaining: The Case of Seoul, *J. Korean Soc. Transp.*, 26(4), Korean Society of Transportation, 87-97.
- Currie G., Delbosc A. (2011), Exploring the Trip Chaining Behaviour of Public Transport Users in Melbourne, *Transport Policy*, 18, 204-210.
- He L., Nassir N., Trépanier M., Hickman M. (2015), Validating and Calibrating a Destination Estimation Algorithm for Public Transport Smart Card Fare Collection Systems, CIRRELT-2015-52.
- He L., Trépanier M. (2015), Estimating the Destination of Unlinked Trips in Public Transportation Smart Card Fare Collection Systems, Annual meeting of the Transportation Research Board, Washington, D.C.
- Ji W., Han B. (1999), A Study on Travel Patterns in Kyonggi-Do, Research Report, 99-03, Kyonggi Development Institute.
- Lee M. (2015), Analyzing Transit Trip Pattern in the Seoul Metropolitan Area Using Smart Card Data, KRIHS Policy Brief, 536, 1-6.
- Munizaga M.A., Palma C. (2012), Estimation of a Disaggregate Multimodal Public Transport Origin-Destination Matrix From Passive Smartcard Data From Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, 24, 9-18.
- Park J., Kim S., Cho J., Heo M. (2008), The Study on Error, Missing Data and Imputation of the Smart Card Data for the Transit OD Construction, *J. Korean Soc. Transp.*, 26(4), Korean Society of Transportation, 109-119.
- Shin K., Choi K. (2014), Analyzing the Relationship Between Precipitation and Transit Ridership Through a Seemingly Unrelated Regression Model, *J. Korean Soc. Transp.*, 32(2), Korean Society of Transportation, 83-92.
- Trépanier M., Tranchant N., Chapleau R. (2007), Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System, *Journal of Intelligent Transportation Systems* 11, 1-14.
- Zhao J., Rahbee A., Wilson N. (2007), Estimating a Rail Passenger Trip Origin-destination Matrix Using Automatic Data Collection Systems, *Computer-Aided Civil and Infrastructure Engineering* 22, 376-387.