

스마트카드 빅데이터를 이용한 서울시 지역별 대중교통 이동 편의성 분석

Analysis of Regional Transit Convenience in Seoul Public Transportation Networks Using Smart Card Big Data

저자 (Authors)	문현구, 오규협, 김상국, 정재윤 Hyunkoo Moon, Kyuhyup Oh, SangKuk Kim, Jae-Yoon Jung
출처 (Source)	대한산업공학회지 42(4) , 2016.8, 296-303(8 pages) Journal of the Korean Institute of Industrial Engineers 42(4) , 2016.8, 296-303(8 pages)
발행처 (Publisher)	대한산업공학회 Korean Institute Of Industrial Engineers
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06748689
APA Style	문현구, 오규협, 김상국, 정재윤 (2016). 스마트카드 빅데이터를 이용한 서울시 지역별 대중교통 이동 편의성 분석. 대한산업공학회지, 42(4), 296-303
이용정보 (Accessed)	제주대학교 223.194.***.198 2020/09/04 01:13 (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.

스마트카드 빅데이터를 이용한 서울시 지역별 대중교통 이동 편의성 분석

문현구 · 오규협 · 김상국 · 정재윤*

경희대학교 산업경영공학과

Analysis of Regional Transit Convenience in Seoul Public Transportation Networks Using Smart Card Big Data

Hyunkoo Moon · Kyuhyup Oh · SangKuk Kim · Jae-Yoon Jung

Department of Industrial and Management Systems Engineering, Kyung Hee University

In public transportation, smart cards have been introduced for the purpose of convenient payment systems. The smart card transaction data can be utilized not only for the exact and convenient payment but also for civil planning based on travel tracking of citizens. This paper focuses on the analysis of the transportation convenience using the smart card big data. To this end, a new index is developed to measure the transit convenience of each region by considering how passengers actually experience the transportation network in their travels. The movement data such as movement distance, time and amount between regions are utilized to access the public transportation convenience of each region. A smart card data of five working days in March is used to evaluate the transit convenience of each region in Seoul city. The contribution of this study is that a new transit convenience measure was developed based on the reality data. It is expected that this measure can be used as a means of quantitative analysis in civil planning such as a traffic policy or local policy.

Keywords: Smart Card, Big Data, Transit Convenience Index, Civil Planning, Seoul City

1. 서 론

대중교통이란 시민들이 이용하는 교통시설과 수단을 의미하며, 정해진 시간계획과 노선을 따라서 제공되는 교통 서비스이다(Han, 2010). 대중교통에 대한 서비스 평가 및 개선을 위하여 다양한 연구들이 진행되었다(Chien *et al.*, 2004; Friman, 2004; Andreassen, 2005; Lee *et al.*, 2015). 그중에서도 대중교통 이용객에 대한 편의성 연구는 교통 시설물 중심의 정적인 데이터와 이용객에 대한 설문조사 등을 활용하여 분석이 진행되었다(TRB, 2003; Cantwell *et al.*, 2009; Dell'Olio *et al.*, 2011). 이러한 연구들은 편의성을 측정할 수 있는 지표를 개발하였으

나, 지역에 기반을 둔 이동 편의성 부분이 고려되지 않았고 이용객의 사용 내역 수집의 어려움으로 지표를 적용하여 대중교통을 평가하지는 못하였다.

최근 대중교통 운임지불이 편리한 스마트카드로 변경됨에 따라 스마트카드 거래 내역을 기반으로 한 다양한 연구가 등장하였다(Bagchi *et al.*, 2005; Park *et al.*, 2008). 이러한 거래 내역을 이용하면 이용객의 이동을 추출하여 다양한 평가 지표에 반영하여 계산할 수 있게 되었다.

본 논문에서는 대중교통 이용객의 이동 데이터에 기반하여 대중교통 이동 편의성을 측정하는 측도들을 정리 및 개발하였고, 제시된 지표들을 활용하여 서울시 지역별 대중교통 이동 편

제11회 석사논문경진대회 수상논문.

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2013R1A2A2A03014718).

* 연락처 : 정재윤 교수, 17104 경기도 용인시 기흥구 덕영대로 1732 경희대학교 산업경영공학과, Tel : 031-201-2537, Fax : 031-204-8114,
E-mail : jyjung@khu.ac.kr

2016년 1월 15일 접수; 2016년 6월 9일 수정본 접수; 2016년 7월 14일 게재 확정.

의성을 측정하였다. 데이터 전처리로는 스마트카드에 기록된 대중교통 승하차 데이터를 버스 승강장 및 지하철 역 기준으로 정리하고, 사각 격자로 구분된 서울시 각 지역별로 결합하여 지역별 이동량 및 이동거리, 이동속도, 이동 편의성 등을 계산하였다. 지역별 결합을 위하여 버스 승강장 및 지하철역들의 GPS 위치를 기준으로 GeoHash 코드라는 geo-coding 기법을 사용하여 지리적 데이터로 변환하여 기초 데이터로 사용하였다.

본 연구에서는 제안된 기법은 서울시를 비롯한 각 행정단위들의 지역별 대중교통 이동 편의성을 분석하는 데 활용할 수 있으며, 앞으로 증가하게 될 스마트카드 빅데이터를 기반으로 도시개발 및 교통서비스 개선을 위하여 활용될 수 있을 것으로 기대한다(Lee and Sohn, 2015).

2. 관련 연구

2.1 대중교통 스마트카드 데이터 분석

대중교통 스마트카드 데이터 분석은 이용자들의 이동 정보를 추출하여 이동패턴 및 행동패턴 등의 유용한 정보를 발견하거나 패턴정보를 다른 요소들과 결합하여 대중교통 수요예측 및 지역별 노선 개선 등에 활용되었다(Park *et al.*, 2008; Jang, 2010). 대중교통 스마트카드 데이터를 이용하여 각 승객의 이동 경로를 생성하고, BRT(Bus Rapid Transit) 이용객과 비 BRT 이용객을 비교 분석한 후, BRT 승객의 이동 흐름지도와 설문조사로 작성된 교통 편의지역 지도를 비교 분석하였다(Tao *et al.*, 2014). 또한 스마트카드 데이터를 이용한 서울시 지하철 이동패턴 분석 연구에서는 유사한 행태의 Zone을 발견하고 두 Zone 간의 이동패턴을 도출하는 MZP(Movement Pattern between Zones)를 제시하였다(Kim *et al.*, 2013). 대중교통 평가 지표를 혼잡도와 이용거리를 기반으로 제안하고 제안된 지표를 바탕으로 스마트카드 데이터를 활용하여 버스 회사에 대해 평가하였다(Trepanier *et al.*, 2009).

2.2 대중교통 편의성 분석

대중교통 편의성이란 대중교통 서비스 평가를 위한 요소들 중 대중교통 이용객의 이용편의성 요소이다(TRB, 2003). TCQSM (Transit Capacity & Quality of Service Manual)는 대중교통 서비스 수준을 평가할 수 있는 지표로 이용 가능성(availability)과 편의성(comfort and convenience)을 제시하고 있다. TCQSM에서는 경쟁 대중교통 간에 서비스 수준을 비교하기 위한 기준으로 <Table 1>과 같이 대중교통의 혼잡도, 편의시설, 신뢰성, 통행시간, 안전도 및 환승 등을 서비스 품질 요소로 포함하였다(TRB, 2003).

대중교통을 이용 통근하는 사람들의 만족도 조사 연구에서는 이용객의 만족도는 혼잡도, 대기시간, 신뢰성, 이동시간 순으로 영향을 받는 것으로 조사되었다(Cantwell *et al.*, 2009). 대

Table 1. Service Quality Measures of TCQSM

	Service Measures		
	Transit Stop	Route Segment	System
Availability	Frequency	Hours of Service	Service Coverage
Comfort & Convenience	Passenger Load	Reliability	Travel Time

중교통 이용객의 증가를 위한 대중교통 서비스 품질 평가 지표를 개발 연구에서는 잠재 이용객이 제일 중요하게 생각하는 요소는 이동 시간이다(Dell'Olio *et al.*, 2011). 따라서 대중교통 이동시간은 이용객의 편의성에 영향을 미치는 요소이지만 기존 연구에서는 크게 반영되지 않았다.

본 연구에서는 대중교통 이용객의 이동행태를 분석하여, 지역별 이동거리, 이동시간, 이동량 등 대중교통 이용자들이 체감할 수 있는 요소들을 반영한 편의성 지표를 제안한다.

3. 대중교통 빅데이터 처리 프레임워크

<Figure 1>은 본 연구의 전체 프레임워크를 나타낸다. 스마트카드 데이터 중 본 연구에 필요한 승객의 이동 데이터(승차일시, 승차 정류장ID, 하차일시, 하차 정류장ID)를 Sqoop을 통하여 추출한다. 추출된 데이터를 Hadoop의 HDFS(Hadoop Distributed File System)에 저장한다. Apache Tajo를 이용하여 지역별 이동 데이터로 결합 후 추출한다. Tajo를 통해 추출된 데이터를 Lucene에 저장한다. 저장된 데이터를 작성된 Java 프로그램을 사용하여 지역별 편의성 지표를 계산한다.

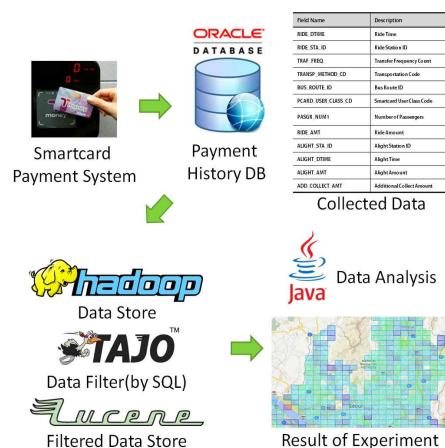


Figure 1. A Framework for Public Transportation Convenience Analysis

본 연구를 위하여 채택된 빅데이터 분석 프레임워크의 주요 구성요소의 역할은 다음과 같다. Sqoop은 관계형 데이터베이스 시스템에서 데이터를 추출하여 HDFS로 저장하는 역할을 한다. Hadoop은 대용량 파일을 저장할 수 있는 분산 파일

시스템(HDFS)을 제공하고 데이터 저장소로 프로그램을 배포하여 병렬 처리하여 속도를 개선한 MapReduce를 지원한다. Tajo는 Hadoop 기반의 대용량 데이터 웨어하우스 시스템으로 SQL을 사용하여 HDFS 파일을 읽을 수 있으며 MapReduce를 활용하여 빠른 데이터 결합 처리가 가능하다. Lucene은 고성능 정보 추출(IR : Information Retrieval) 라이브러리이다. 필요한 데이터를 빠르게 저장 및 추출할 수 있다.

4. 지역적 데이터 분석

4.1 지역 데이터 전처리

대중교통 이용객을 지역별로 묶어서 분석하기 위해서는 각각의 이용객의 승하차 정류소의 위치를 수집하는 것이 필요하다. 버스 정류소는 지하철과 달리 정류소 위치의 추가 및 변경되는 경우가 있으므로 분석을 위해서는 이용객이 이용한 시기의 정확한 정류소의 위치가 필요하다. 정류소 위치는 서울시 교통정보과에 요청하여 해당 시기의 정류소의 위치 정보를 수집하였다. 수집된 정류소의 위치는 중부원점 TM(Transverse Mercator) 좌표계를 기반으로 작성되어 있다. 지도 서비스에서는 GPS에서 사용하는 좌표계인 WGS84가 널리 쓰이고 있으므로, 중부원점 TM 좌표인 정류소의 위치를 WGS84 좌표로 변환하였다. 변환된 위치 데이터를 주소지와 맵핑하기 위하여 geo-coidng을 진행하였다. 정류소의 WGS84 위치와 지역별 주소지를 기반으로 지역을 분할하여 다양한 분석에 활용한다.

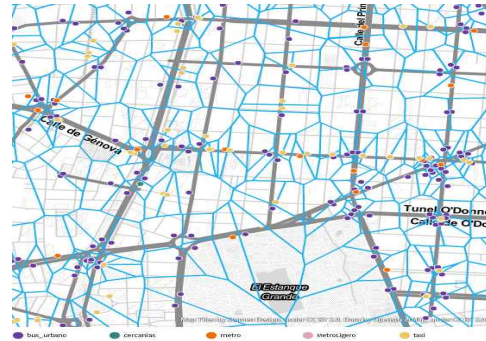
4.2 지역 분할 기법

지역 분할 기법은 보로노이 다이어그램, hexagon 방식, 격자 방식 등이 있다. 보로노이 다이어그램(Voronoi Diagram)은 두 개의 지점을 잇는 선분과 직교하는 직선을 그리는 과정을 반복하여 지역을 분할한다. 보로노이 다이어그램은 통신 네트워크상에 최적의 센서 노드 배치 등에 종종 이용되어 불필요한 지역크기를 최소화할 수 있다. <Figure 2>(a)는 마드리드 지역의 대중교통 정류소를 기준으로 보로노이 다이어그램으로 지역을 분할한 예시이다(<http://www.ekonlab.com/?p=995>).

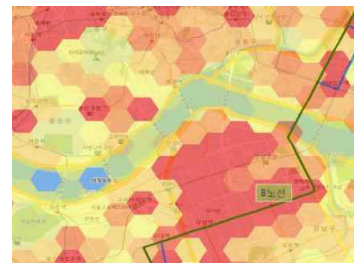
hexagon(Hexagon) 방식은 육각 모형으로 지역을 분할하는 기법으로 사각 격자 기법과 다르게 접경 지역들에 대한 지역 인접성이 우수하다. hexagon 방식의 예로는 서울시 심야버스 노선 선정 사례가 있다. 심야버스 노선 선정을 위하여, <Figure 2>(b)과 같이 서울시 전역을 약 1km 반경의 1,250개 hexagon으로 나누어 유동인구 밀집도와 이동수요를 분석하였다(<http://news.join.com/article/11961889>).

마지막으로 사각 격자(Grid) 방식은 hexagon 방식과 마찬가지로 동일한 형태의 지역으로 분할한다는 점에서 유사하지만, 필요에 따라 계층화하기 용이하다. <Figure 2>(c)는 서울을 격자 방식인 GeoHash 기법으로 분할한 예시이다(<http://geohash>.

gofreerange.com). 본 연구에서는 지역의 결합과 분할이 편리한 GeoHash 기법을 도입하여 교통 데이터 분석 연구를 진행하였다.



(a) Voronoi Diagram



(b) Hexagon Tile



(c) GeoHash Grid

Figure 2. Geograph Partitioning Techniques

4.3 GeoHash 기반 지역 분할 기법

GeoHash는 전 세계의 위도와 경도로 이루어진 위치(2차원)를 유일한 문자열(1차원)로 인코딩 하는 해싱(Hashing) 기법이다. GeoHash는 지구를 4행×8열의 32개의 격자(bucket)로 분할한 뒤 각 격자별로 z-order 방식과 base32 인코딩을 통하여 코드 한 자리를 생성한다(Anthony et al., 2013). 분할된 하나의 격자를 32개의 격자로 다시 분할하고, 코드 뒤에 GeoHash 코드를 한 자리 더 추가하는 방식으로 최대 12회 분할하여 최대 12자리 코드로 지구의 위치를 표현할 수 있다.

GeoHash 관련 라이브러리를 사용하면, 1차원으로 표현된 GeoHash 지점 간 거리, 인접 유무, 소속 관계 등을 빠르게 확인할 수 있다. 단, 동일한 GeoHash 코드로 표현된 지점 간의 최대 오차거리는 GeoHash 코드에 대응되는 격자의 대각선 길이와 같다. 예를 들어, 6자리 GeoHash 코드는 $1.2\text{km} \times 609.4\text{m}$ 크기의 사각형 지역을 표현하며, 두 지점의 최대 거리는 약 1.3km이다.

본 연구에서는 대중교통 정류소의 인접 정류소를 빠르게 추출하고, 다양한 크기의 격자를 활용하여 분석 목표인 지역의 크기를 변화시키기 위하여 GeoHash 코드를 사용하였다.

정류장 간의 인접유무를 계산하기 위해서는 모든 정류장의 거리를 직접 계산을 해야 하고 정류장이 하나만 추가 되더라도 해당 정류장과 모든 정류장의 거리를 계산할 필요가 있다. 하지만 GeoHash 코드를 사용하면 GeoHash 코드의 자리수의 일치하는 길이의 크기로 인접 유무를 쉽게 파악 가능하다.

GeoHash는 코드의 자리 수에 따라 해당하는 격자의 크기를 쉽게 변환 할 수 있으므로, 분석 목표에 따라 격자의 크기를 변환하여 분석을 시도할 수 있다.

5. 대중교통 편의성 지표 개발

본 연구에서는 대중교통 이용자의 이동거리 및 이동시간과 지역별 대중교통 이용량을 결합하여 이용객이 체감할 수 있는 편의성 지표를 제안한다. 서울시 도시철도 이용객의 편의성 설문 조사 결과에 따르면 편의성에 대한 주요 요인이 대기시간, 접근 시간, 이동시간, 요금 등의 순으로 나타났다(Byun *et al.*, 2014).

특정 지역의 대중교통 편의성은 다양한 측면에서 바라볼 수 있다. 예를 들어, 평균 이동거리가 짧다는 것은 지역민들의 타 지역 접근성이 우수하다는 의미이며, 평균 표정속도가 빠르다는 것은 지역민들이 원하는 지역까지 신속하게 접근할 수 있다고 해석 가능하다. 본 논문에서는 “해당 지역 거주자들이 자신이 실제로 원하는 목적지까지 얼마나 빠르게 접근할 수 있는가”를 기준으로 대중교통 편의성 지표를 개발하였다. 이를 평가하기 위하여 실 데이터를 기반으로 지역민들의 타 지역으로의 이동량을 가중평균 후 반영하여 평균 표정속도를 측정함으로써, 해당 지역 거주자들의 체감 편의성을 반영하고자 하였다.

5.1 지역별 평균 이동거리

본 절에서는 GeoHash 기법을 이용하여 생성된 Zone G_i 의 평균 이동거리를 나타내는 계산식을 제시한다. 두 지역간의 거리는 Zone의 중점 간 직선거리로 측정하며 Zone G_i 의 평균 이동거리 D_i 는 각 지역으로 이동하는 이용객의 비율을 이동 거리에 가중평균하여 계산한다.

$$D_i = \sum_{j=1}^N \left(\frac{|G_i \rightarrow G_j|}{|G_i \rightarrow G_j|} d(G_i \rightarrow G_j) + \frac{|G_j \rightarrow G_i|}{|G_j \rightarrow G_i|} d(G_j \rightarrow G_i) \right) / 2 \quad (1)$$

G_i 는 GeoHash를 통해 생성된 Zone 중에서 i 번째 Zone을 의미한다. $G_i \rightarrow G_j$ 는 i 번째 Zone에서 j 번째 Zone으로의 이동을 나타내며 $G_i \rightarrow G_j$ 는 i 번째 Zone에서 다른 모든 Zone으로의 이동량을 나타낸다. 또한 $d(G_i \rightarrow G_j)$ 는 Zone G_i 에서 Zone G_j 까지의 이동거리를 나타낸다. Zone G_i 에서 다른 지역으로 나가는 경우를 G_i 의 유출이라 하고, Zone G_i 로 들어오는 경우를 G_i 의 유입이라고 한다. 즉, $\frac{|G_i \rightarrow G_j|}{|G_i \rightarrow G_j|}$ 는 Zone G_i 에서 G_j 로 유출된 대

중교통 이용객의 비율을 나타내며, $\frac{|G_j \rightarrow G_i|}{|G_j \rightarrow G_i|}$ 는 Zone G_j 에서 G_i 로 유입된 비율을 나타낸다. 식 (1)에 표현된 Zone G_i 의 평균 이동거리는 Zone G_i 로 유입 이동거리와 유출 이동거리를 각각 실제 이동량으로 가중평균한 후 다시 평균하여 계산된다.

5.2 지역별 평균 표정속도

일반적으로 대중교통 평균속도는 시종점간의 주행거리를 중간 정차역 등에서 정차할 때 발생하는 시간을 포함하지 않은 총소요시간으로 나눈 양, 즉 운행거리를 순수 운행시간으로 나눈 값을 의미한다. 반면에, 대중교통 표정속도(schedule speed)는 시종점간의 주행거리를 중간 정차역 등에서 정차한 시간을 포함한 총소요시간으로 나눈 양으로 계산한다. 따라서 표정속도가 대중교통 이용객들이 출발지로부터 목적지에 도달하는 체감속도라고 할 수 있다. Zone G_i 의 평균 표정속도 S_i 는 다음과 같이 계산한다.

$$S_i = \sum_{j=1}^N \left(\frac{d(G_i \rightarrow G_j)}{t(G_i \rightarrow G_j)} + \frac{d(G_j \rightarrow G_i)}{t(G_j \rightarrow G_i)} \right) / 2 \quad (2)$$

$t(G_i \rightarrow G_j)$ 는 Zone G_i 에서 Zone G_j 까지의 각 정차역에서의 시간을 포함한 총 이동시간이다. 표정속도 S_i 는 Zone G_i 에서 타 지역으로 유출되는 이용객의 경우와 타 지역에서 Zone G_i 로 유입되는 경우를 고려한다. 즉, $\frac{d(G_i \rightarrow G_j)}{t(G_i \rightarrow G_j)}$ 는 Zone G_i 에서 Zone G_j 로 유출된 대중교통 이용객의 평균 표정속도이며, $\frac{d(G_j \rightarrow G_i)}{t(G_j \rightarrow G_i)}$ 는 Zone G_j 에서 Zone G_i 로 유입된 대중교통 이용객의 평균 표정속도이다. 최종적인 Zone G_i 의 평균 표정속도는 유출 및 유입 표정속도의 평균으로 계산된다.

5.3 지역별 이동 편의성

본 절에서는 표정속도에 지역 거주자들의 실제 이동량을 결합함으로써 대중교통 이용객이 체감할 수 있는 이동 편의성(transit convenience) 측도를 개발하였다. 먼저 Zone G_i 에서 N 개의 G_j 지역으로 이동하는 경우, 지역 G_i 의 유출 편의성 CV_i^{out} 은 다음과 같이 계산된다.

$$CV_i^{out} = \sum_{j=1}^N \frac{|G_i \rightarrow G_j|}{|G_i \rightarrow G_j|} \frac{d(G_i \rightarrow G_j)}{t(G_i \rightarrow G_j)} \quad (3)$$

CV_i^{out} 은 Zone G_i 의 유출 표정속도 $\frac{d(G_i \rightarrow G_j)}{t(G_i \rightarrow G_j)}$ 와 Zone G_i

에서 G_j 로 유출된 대중교통 이용객의 비율 $\frac{|G_i \rightarrow G_j|}{|G_i \rightarrow G_j|}$ 을 결합하여 N개의 지역으로 이동하는 유출 편의성을 계산한다. 마찬가지로 지역 G_i 의 유입 편의성 CV_i^{into} 은 다음과 같이 계산된다.

$$CV_i^{into} = \sum_{j=1}^N \frac{|G_j \rightarrow G_i|}{|G_j \rightarrow G_i|} \frac{d(G_j \rightarrow G_i)}{t(G_j \rightarrow G_i)} \quad (4)$$

CV_i^{into} 은 Zone G_i 의 유입 표정속도 $\frac{d(G_j \rightarrow G_i)}{t(G_j \rightarrow G_i)}$ 와 Zone G_j

에서 G_i 로 유입된 대중교통 이용객의 비율 $\frac{|G_j \rightarrow G_i|}{|G_j \rightarrow G_i|}$ 을 결합하여 계산된다.

최종적으로 Zone G_i 의 대중교통 편의성 CV_i 는 G_i 의 유출 편의성과 유입 편의성의 평균으로 계산된다.

$$CV_i = (CV_i^{out} + CV_i^{into})/2 \quad (5)$$

제안된 지표는 해당지역에서 모든 지역으로 이동하는 표정속도를 실제 이동량으로 가중평균함으로써 전체 이용객이 체감한 속도들을 편의성 지표에 반영하였다는 특징이 있다.

6. 서울시 대중교통 편의성 평가

본 연구에서는 제안된 대중교통 평가 지표들을 이용하여 서울시 대중교통 이용객 데이터에 적용 분석하였다. 서울시의 지역별 편의성 평가를 위하여, GeoHash를 이용하여 서울시를 사각격자로 나누어 각 지역별 대중교통 편의성을 분석하였다.

6.1 실험데이터 및 실험목록

제안된 지표를 통한 실험을 위하여 2012년 3월 12일부터 16일까지 서울시 버스 및 지하철의 주중 이용객 데이터를 사용하였다. 또한 서울시 간의 접근성 평가에 집중하기 위하여 서울시 내부 이동 데이터만을 추출하여 사용하였으며, 마을버스 이동데이터는 데이터 처리량 감소를 위해 본 분석에서는 제외하였다.

서울시를 GeoHash 코드의 길이를 활용하여 다양한 크기의 격자(Zone)로 쪼갤 수 있지만, 그중에서도 지역별 특성을 반영할 수 있는 적합한 크기인 6자리 GeoHash 코드로 1.2km×609.4m 크기(약 0.73km²)의 격자를 이용하여 738개의 Zone으로 구분하였다.

구분된 Zone에 속하는 모든 정류소의 이동은 해당 Zone의 이동으로 반영하였다. 이용객의 이동은 모든 환승을 포함하여 목적지, 도착시간을 도출하였다. 도출된 이동과 시간을 바탕으로 앞 절에서 제시된 이동거리, 표정속도, 편의성을 계산하였다.

계산된 데이터를 활용하여 지역별 대중교통 평균 이동량, 평균 이동거리, 평균 표정속도, 대중교통 편의성을 평가하였으며, 나아가 출근, 퇴근시간의 대중교통 편의성을 지역별로 비교 평가해보았다.

6.2 지역별 평균 이동량

<Figure 3>은 738개 Zone으로 서울시를 분할하여 해당 기간의 평일 하루 평균 이동량을 순위별로 나타낸 그래프이다. 평균 이동량이 많은 상위 5개 지역은 <Table 2>와 같이 강남구 역삼1동을 비롯하여, 종로구 종로1-4가동, 마포구 대흥동, 중구 명동, 강남구 대치4동이 포함되었으며, 이동량이 적은 지역은 <Table 3>과 같이 마포구 용강동, 서초구 양재1동, 성북구 장위3동 등으로 나타났다. 서울시 각 지역별 평일 하루 평균 이동량은 12,420명으로 나타났다. 이는 각 지역에서 하루동안 출발한 이용객 수와 최종적으로 도착한 이용객 수의 합들을 평균한 값이다. 환승 이용객들은 최초 출발지와 최종 목적지만 반영하였다. 이동량 상위권 지역은 주로 회사가 밀집한 지역들에서 도출되었다.

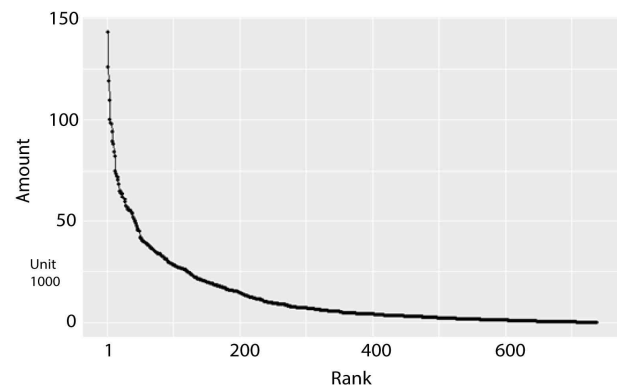


Figure 3. Distribution of Average Daily Movement Amounts of Zones in Seoul

Table 2. Top 5 Average Daily Movement Amount Zones in Seoul

Rank	Zone (GeoHash)	Avg. amount	Zone center
1	wydm6d	143,139	Yeoksam1, Gangnam
2	wydm9x	125,937	Jongno 1.2.3.4 ga, Jongno
3	wydm8s	119,178	Daeheung, Mapo
4	wydm9w	109,499	Myeong, Jung
5	wydm75	100,064	Daechi 4, Gangnam

Table 3. Bottom 5 Average Daily Movement Amount Zones in Seoul

Rank	Zone (GeoHash)	Avg. amount	Zone center
1	wydkgr	1	Yangjae 2, Seocho
2	wydm88	3.2	Yonggang, Mapo
3	wydm4s	3.2	Yangjae 1, Seocho
4	wydlq5	10	Jangwi 3, Seongbuk
5	wyfdm47	10.4	Yangjae 1, Seocho

6.3 지역별 평균 이동거리

<Figure 4>는 지역별 대중교통 평균 이동거리를 이동거리가 긴 지역부터 짧은 지역까지 순위에 따라 나타낸 그래프이다. 대중교통 이용객의 1회 평균 이동거리는 10.41km로 나타났으며, 강서구 공항동이 18.05km로 가장 긴 1회 이동거리를 나타냈다. 이 거리는 앞서 언급한 바와 같이, 환승을 고려하여 출발 지로부터 최종목적지까지의 거리를 의미한다. 이동거리 상위 5개 지역은 <Table 4>와 같이 모두 강서구에서 나타났으며, 강서구 공항동, 방화2동, 방화3동으로 도출되었다. 반대로 이동거리가 가장 짧은 지역은 <Table 5>와 같이 서초구 양재2동으로 나타났다.

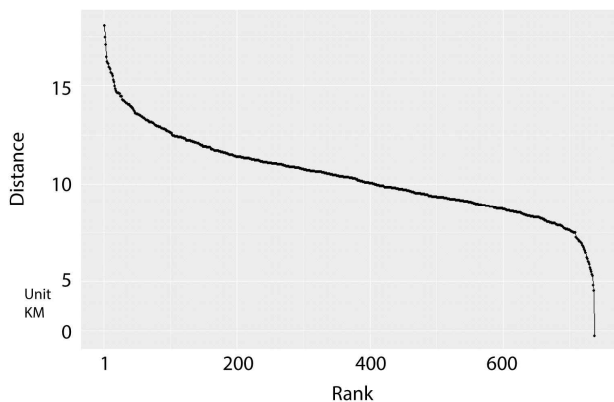


Figure 4. Distribution of Average Movement Distances of Zones in Seoul

Table 4. Top 5 Average Movement Distance Zones in Seoul

Rank	Zone (GeoHash)	Distance (km)	Zone center
1	wydtjm	18.05	Gonghang, Gangseo
2	wydlv2	17.46	Banghwa 2, Gangseo
3	wydlvtq	17.07	Banghwa 2, Gangseo
4	wydlv8	16.46	Banghwa 3, Gangseo
5	wydlvtx	16.21	Banghwa 2, Gangseo

Table 5. Bottom 5 Average Movement Distance Zones in Seoul

Rank	Zone (GeoHash)	Distance (km)	Zone center
1	wydkgr	2.20	Yangjae 2, Seocho
2	wydm88	4.53	Yonggang, Mapo
3	wydm84	4.80	Seogang, Mapo
4	wydm9g	5.27	Jangchung, Jung
5	wydm4s	5.35	Yangjae 1, Seocho

6.4 지역별 평균 표정속도

환승대기시간이 포함된 이용객의 이동시간을 반영한 평균 표정속도를 계산하였다. <Figure 5>는 지역별 대중교통 평균 표정속도를 순서대로 정렬한 그래프이며 <Figure 6>은 지역별 표정속도를 지도에 표시한 것이다.

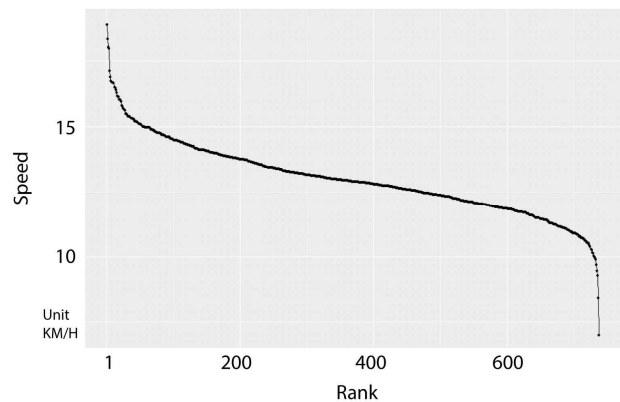


Figure 5. Distribution of Average Schedule Speeds of Zones in Seoul

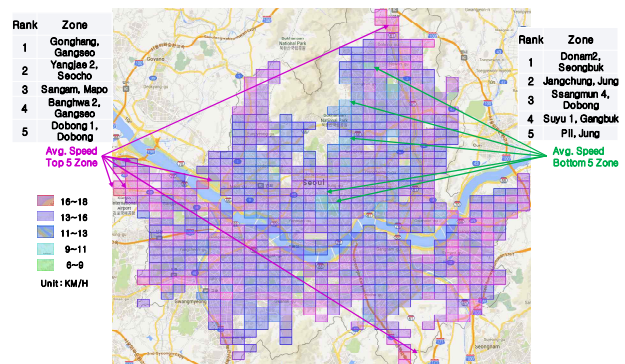


Figure 6. Average Schedule Speeds of Zones on the Map

서울시의 평균 표정속도는 12.90km/h로 측정되었고, 강서구 공항동이 18.60km/h로 가장 높게 나타났으며 상위권 지역은 서초구 양재2동, 마포구 상암동, 강서구 방화2동, 도봉구 도봉1동 등도 포함되었다. 표정속도가 높게 측정된 지역은 비교적 이동거리가 긴 지역들이 많이 도출되었다. 이와 반대로 표정

속도가 가장 낮게 측정된 지역은 성북구 돈암2동으로 9.34km/h로 나타났으며, 표정속도 하위권 지역에는 중구 장충동, 도봉구 쌍문4동, 강북구 수유1동, 중구 필동 등 강북지역이 주로 포함되었다. 이는 중구를 포함한 도봉구, 강북구 지역의 상대적으로 밀집되고 정류장 간 거리가 짧게 배치되어 있기 때문으로 예상된다.

6.5 지역별 이동 편의성

서울시의 대중교통 이동 편의성 점수를 정렬한 그래프는 <Figure 7>이며 평균 점수는 13.80로 나타났다. 이는 서울시 거주자들이 현재 서울시내 타지역으로 이동할 때의 경험하는 평균속도라고 해석할 수 있다. <Figure 8>은 서울시 지역별 편의성 점수를 지도에 표현한 것이다. 편의성 점수가 가장 높은 지역은 강서구 공항동으로 22.08로 측정이 되었다. 편의성 점수 결과 표정속도가 가장 높았던 강서구 공항동이 편의성 점수도 가장 높게 측정이 된 것을 알 수 있다. 편의성 점수 상위권 지역은 서초구 내곡동, 강남구 세곡동, 마포구 상암동 등이 포함되었으며, 편의성 점수가 가장 낮은 지역은 노원구 월계2동으로 4.74로 측정이 되었다. 이 외에 서초구 양재1동, 성북구 돈암2동, 중랑구 신내1동 등이 하위권 지역에 포함되었다.

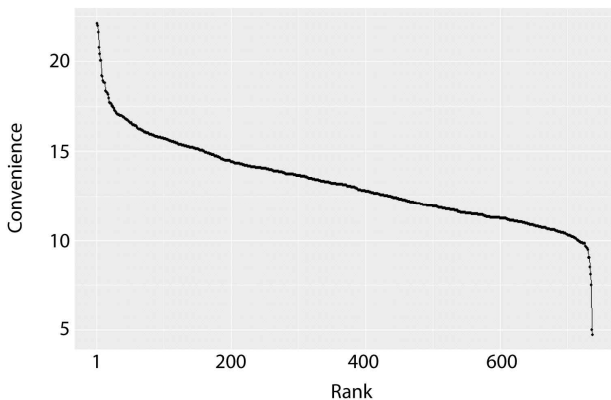


Figure 7. Distribution of Average Transit Convenience of Zones in Seoul

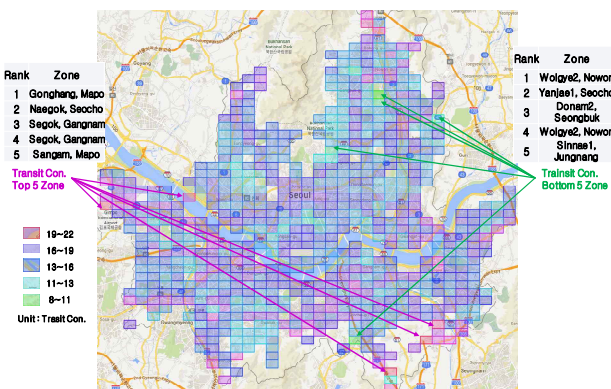


Figure 8. Average Transit Convenience of Zones on the Map

6.5 출퇴근시간 이동 편의성 분석

서울시내의 이동이 가장 많은 출근시간과 퇴근시간의 대중교통 편의성을 비교 분석하였다. 실험 데이터는 동일 기간의 6시~9시(3시간)의 이동 데이터를 출근시간 데이터로, 17시~20시(3시간)의 이동 데이터를 퇴근시간 데이터로 사용하였다. 사전 연구에 따르면 하루 이동량의 50%가 이 시간대에 분포하고 있음에 근거하였다(Kim *et al.*, 2014).

출근시간 편의성 점수는 <Figure 9>와 같이 금천구 시흥2동에서 25.24로 가장 높게 측정이 되었으며 상위권 지역은 강남구 세곡동, 강서구 방화2동, 서초구 양재2동, 도봉구 도봉1동이 포함되었다. 도봉구 도봉1동은 강북지역에서도 끝에 위치한 지역으로 강북의 중심부까지 이동하는 속도가 빠르기 때문에 편의성이 높게 측정되었을 것이라고 예상할 수 있으며, 강서구 방화2동 역시 강서구 지역에서 서울 중심지역으로 이동하는 이동속도가 빠르게 나타났기 때문에 편의성이 높게 나타난 것으로 예상된다. 전반적으로 출근시간 편의성 점수가 높게 나온 지역은 서울 도심으로 진입하기 위해 이동거리가 길게 측정되었던 지역을 중심으로 관측되었다. 출근시간 편의성 점수가 가장 낮게 나온 지역은 서초구 양재1동으로 4.93로 측정되었다. 하위권 지역은 서울 도심과 인접한 지역인 강북의 노원구 월계2동, 도봉구 쌍문4동 등이 포함되었다.

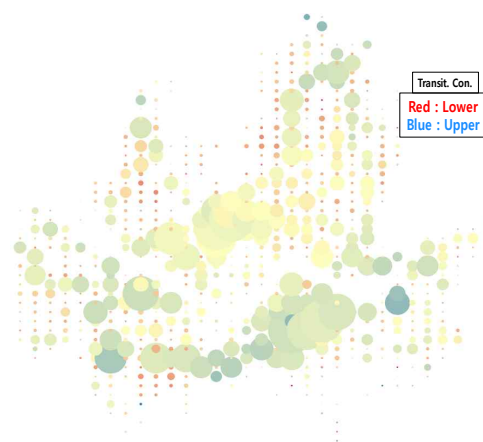


Figure 9. Average Transit Convenience of Zones at Morning Rush Hour

퇴근시간 편의성 점수는 <Figure 10>과 같이 서초구 내곡동이 23.89로 가장 높게 측정이 되었다. 편의성 점수 상위권 지역은 강서구 공항동을 포함하여 도봉구 도봉1동, 서초구 양재2동에서 측정이 되었다. 퇴근시간 평균 편의성 점수는 출근시간 평균 편의성 점수에 비해 2점 정도 낮은 것을 확인할 수 있었다. 이는 퇴근시간 표정속도가 출근시간 표정속도보다 전반적으로 낮기 때문이라고 예상된다. 편의성 점수 하위권 지역은 노원구 월계2동, 성북구 돈암2동, 중랑구 신내1동 등에서 나타났으며 노원구 월계2동이 편의성 점수가 가장 낮은 4.60으로 측정되었다.

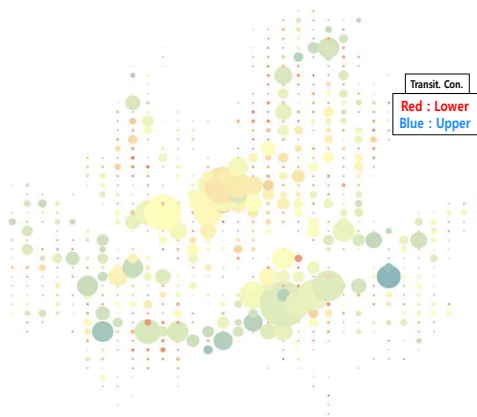


Figure 10. Average Transit Convenience of Zones at Evening Rush Hour

7. 결 론

본 논문에서는 대중교통 이용자의 실제 이동 데이터에 기반하여 지역별 대중교통 편의성 평가 측도들을 제시하고 서울시를 대상으로 수행하였다. 대중교통 편의성을 평가하기 위하여, GeoHash 기법을 이용하여 지역을 사각 격자로 분할하여 정거장 및 이동데이터를 매핑하였고, 생성된 Zone 간의 대중교통 이용자의 출발지와 도착지를 기준으로 거리, 이동시간 및 속도를 측정하였다. 또한 환승을 고려하여 실제 최종 목적지까지 이동 데이터로 결합하고, 정차시간까지 포함한 표정속도를 계산한 후, 지역별 이동량을 반영하여 최종적인 이동 편의성을 측정하였다.

제시된 측도들을 이용하여 서울시 대중교통(버스 및 지하철)의 평일 이동 데이터를 대상으로 6자리 GeoHash 격자(약 0.73km²)로 분할된 738개 지역들의 지역별 대중교통 편의성을 평가하였다. 한 지역의 일평균 이동량은 12,420명이고, 1회 평균 이동거리는 10.41km이며, 타지역으로의 평균 표정속도는 12.90km/h였으며, 이용객들이 경험한 이동 편의성은 13.80이었다. 또한, 출근시간대의 이동 편의성은 25.24이고 퇴근시간대의 이동 편의성은 23.89로 퇴근시간대가 조금 더 느리게 체감할 것으로 분석되었다.

본 연구는 실제 수집된 대중교통 스마트카드 빅데이터를 기반으로 복합적 교통수단(버스 및 지하철), 전 지역 및 노선의 실제 데이터를 이용하여 분석을 진행하였다는 점에서 의미가 있다. 또한 지역 분할을 위하여 사용된 GeoHash 기법은 서울시 외에도 지리적 범위와 분석 목적에 따라 격자 크기를 변화하여 빠르게 적용할 수 있다는 장점이 있다.

실험 결과, 제시된 이동 편의성 측도가 표정속도에 많은 영향을 받는 것으로 나타났다. 편의성 점수가 높은 지역들은 주로 시 외곽 지역에 위치하고 있었다. 이는 시 외곽에서 시내로 들어오는 경우에는 대중교통 전체의 영향을 적게 받아 표정속도가 증가한 것으로 사료된다. 이들 지역은 주요 도심과의 거리가 멀지만 다른 지역에 비해 상대적으로 도심의 접근 속도가 빠르다. 목적지에 접근하는 거리보다는 여러 지점으로 이동하는 평균시간을 기반으로 대중교통 편의성을 측정하였기 때문에 장거리 직행 노선이 존재하는 경우가 상대적으로 편의

성이 우수하다고 측정되었다.

추후 연구로는 사용자 이동 시간대, 사용 편의성, 접근성, 혼잡도, 안락함 등을 반영한 대중교통 편의성 지표개선, 시설물과 지리적 특성 데이터와 결합하여 이용 및 거주 목적에 따른 지역별 편의성 지표로 확장할 수 있을 것으로 기대한다.

참고문헌

- Andreassen, T. W. (2005), (Dis) satisfaction with public service : the case of public transportation, *Journal of Service Marketing*, **9**(5), 30-41.
- Anthony, F., Chris, E., James, H., and Skylar, L. (2013), Spatio-temporal Indexing in Non-relational Distributed Databases, *Proc. IEEE Int. Conf. on Big Data*, 291-299.
- Bagchi, M. and White, P. R. (2005), The potential of public transport smart card, *Transport Policy*, **12**(5), 464-474.
- Byun, J. H. and Jang, K. T. (2014), User Convenience Analysis of Urban Railway Using Survey, *Proc. 10th Autumn Conf. of the Korean Society for Railway*, 839-844.
- Cantwell, M., Caulfield, B., and O'Mahony, M. (2009), Examining the factors that impact public transport commuting satisfaction, *Journal of Public Transportation*, **12**(2), 1-21.
- Chien, S. I. and Qin, Z. (2004), Optimization of bus stop locations for improving transit accessibility, *Transportation Planning and Technology*, **27**(3), 211-227.
- Dell'Olio, L., Ibeas, A., and Cecin, P. (2011), The quality of service desired by public transport users, *Transport Policy*, **18**(1), 217-227.
- Friman, M. (2004), Implementing quality improvements in public transport, *Journal of Public Transportation*, **7**(4), 49-65.
- Han, J. S. (2010), *Geography of transportation*, Hanul Academy, Paju, Korea.
- Jang, W. (2010), Travel time and transfer analysis using transit smart card data, *Journal of the Transportation Research Board*, **2144**, 142-149.
- Kim, K. H., Oh, K. H., Lee, Y. K., Kim, S. H., and Jung, J. Y. (2014), An Analysis on Movement Patterns between Zones Using Smart Card Data in Subway Networks, *International Journal of Geographical Information Science*, **28**(9), 781-1801.
- Konjar, M., Lisec, A., and Drobne, S. (2010), Method for delineation of functional regions using data on commuters, *Proc. 13th AGILE Int. Conf. on Geographic Information Science*, 1-10.
- Lee, S. J., Lee, Y., and Shin, H. (2015), Empirical Study on the Determinants of Service Quality Deterioration of Taxi Industry : Based on Traffic Accidents by Taxi in Seoul, *Journal of the Korean ORMS Society*, **40**(4), 83-105.
- Lee, W. S. and Sohn, S. Y. (2015), Topic Model Analysis of Research Trend on Spatial Big Data, *Journal of Korean Institute of Industrial Engineers*, **41**(1), 64-73.
- Park, J. Y. and Kim, D. J. (2008), The potential of using the smart card data to define the use of public transit in Seoul, *Journal of the Transportation Research Board*, **2063**, 3-9.
- Tao, S., Corcoran, J., Mateo-Babiano, M., and Rohde, D. (2014), Exploring Bus Rapid Transit passenger travel behaviour using big data, *Applied Geography*, **53**, 90-104.
- Trepanier, M., Morency, C., and Agard, B. (2009), Calculation of transit performance measures using smartcard data, *Journal of Public Transportation*, **12**(1), 79-96.