

# 통행시간 분포 기반의 전철역 클러스터링

## Metro Station Clustering based on Travel-Time Distributions

공인택(InTaek Gong)\*, 김동윤(DongYun Kim)\*\*, 민윤홍(Yunhong Min)\*\*\*

### 초 록

스마트교통카드 데이터는 대표적인 모빌리티 데이터로 이를 이용하여 대중교통 이용행태를 분석하고 정책 개발에 활용할 수 있다. 본 논문은 이러한 연구의 하나로 전철 이용패턴을 이용하여 전철역들을 분류하는 문제를 다룬다. 전철역의 클러스터링을 다룬 기존 논문들은 이용행태 중 통행량만을 고려하였기에 본 논문은 이에 대한 보완적인 방법의 하나로 통행시간을 고려한 클러스터링을 제안한다. 각 역의 승객들을 출근 시간 출발, 출근 시간 도착, 퇴근 시간 출발, 퇴근 시간 도착 승객들로 분류한 다음 각각의 통행시간을 와이블 분포로 모형화하여 추정된 형상모수를 역의 특성값으로 정의하였다. 그리고 특성 벡터들을 K-평균 클러스터링 기법을 사용하여 클러스터링하였다. 실험결과 통행시간을 고려하여 역의 클러스터링을 수행하면 기존 연구의 클러스터링 결과와 유사한 결과가 나올 뿐만 아니라 더 세분화 된 클러스터링이 가능함을 관찰하였다.

### ABSTRACT

Smart card data is representative mobility data and can be used for policy development by analyzing public transportation usage behavior. This paper deals with the problem of classifying metro stations using metro usage patterns as one of these studies. Since the previous papers dealing with clustering of metro stations only considered traffic among usage behaviors, this paper proposes clustering considering traffic time as one of the complementary methods. Passengers at each station were classified into passengers arriving at work time, arriving at quitting time, leaving at work time, and leaving at quitting time, and then the estimated shape parameter was defined as the characteristic value of the station by modeling each transit time to Weibull distribution. And the characteristic vectors were clustered using the K-means clustering technique. As a result of the experiment, it was observed that station clustering considering pass time is not only similar to the clustering results of previous studies, but also enables more granular clustering.

**키워드** : 네트워크 분석, 스마트카드 데이터, 전철역 클러스터링  
Network Analysis, Smart-Card Data, Station Clustering

본 연구는 인천대학교 교내연구 (2019-0083)의 지원으로 수행되었음.

\* First Author, Professional, Procurement DX Team, LG CNS(itgong78@naver.com)

\*\* Co-Author, Graduate Student (Master Course), Graduate School of Logistics, Incheon National University (dongyunkim@inu.ac.kr)

\*\*\* Corresponding Author, Associate Professor, Graduate School of Logistics, Incheon National University (yunhong.min@inu.ac.kr)

Received: 2022-04-18, Review completed: 2022-04-27, Accepted: 2022-04-28

## 1. 서 론

스마트 교통카드 데이터는 스마트교통카드를 사용한 대중교통 사용 이력을 기록한 모빌리티 데이터 중 하나로 카드 이용자의 출발 전철역(또는 버스정거장), 출발역 입장시간(또는 버스 승차시간), 도착 전철역(또는 버스정거장), 도착역 퇴장시간(또는 하차시간) 등의 정보를 포함하고 있다. Kim[12]에 따르면 국내 대중교통 이용자의 95% 이상이 스마트 교통카드를 이용하고 있다고 파악되고 있으며 이는 스마트교통카드 데이터가 대중교통을 이용하는 모빌리티와 관련하여 대표성을 갖고 있는 데이터라는 것을 의미한다.

모빌리티 데이터는 사람들의 이동과 관련된 모든 데이터를 통칭하는 개념으로 운전자에게 혼잡한 도로를 미리 알려주는 것과 같이 운행을 도와주는 개념부터 국민들의 사회·경제 활동을 분석하여 정책개발에 활용하는 것까지 다양한 용도로 활용될 수 있다[2, 3, 4, 8, 11, 23]. 스마트교통카드 데이터도 이러한 모빌리티 데이터의 일종으로 이를 이용하는 다양한 연구가 존재한다. 초기의 스마트교통카드 데이터 연구는 데이터 자체의 불완전성을 보완하는 연구가 주를 이루었다. 버스의 경우, 카드의 승차 태깅은 대부분 이루어지나 하차 태깅이 이루어지지 않는 경우가 많아 하차장소 및 시간에 대한 정보가 누락 된 경우가 많다. 전철의 경우에는 환승 태깅이 이루어지지 않는 경우 환승에 대한 정보를 데이터로부터 알기 어려운 문제가 있다. 따라서, 이러한 불완전성을 해결하는 연구가 초기 스마트교통카드 데이터의 주를 이루었다 [5, 8].

스마트교통카드 데이터의 등장 이전에는

가구통계 조사 등의 설문 조사와 인터뷰를 활용하여 다양한 교통 관련 정보를 추출해 왔었다. 스마트교통카드 데이터로부터 추출된 정보는 이전의 방식으로 추출된 정보와 상당한 일관성을 갖고 있으며[6] 인터뷰나 설문 조사보다 훨씬 많은 데이터의 확보가 가능하므로 이 데이터로부터 의미 있는 정보를 추론하여 다양한 목적으로 활용하는 방법들이 제안되었다. 이러한 연구에는 통행패턴을 추출하여 이에 영향을 주는 요인을 분석하는 연구[13], 승객들의 통행패턴에 대한 클러스터링을 수행하는 연구[10, 16] 등이 있다.

본 논문에서는 스마트교통카드 데이터를 이용하여 전철역을 클러스터링하는 문제를 고려한다. 전철역 클러스터링은 Min[15]과 Park and Lee[19]의 연구에서도 다룬 문제로 역들을 여러 개의 클러스터 혹은 군집으로 분류한 다음 동일 군집에 속한 전철역들의 유사 속성을 파악하는 것이 목적이다. Min[15]은 스마트교통카드 데이터로부터 시간대별 역에 입장하고 퇴장하는 승객수를 기반으로 역들을 클러스터링했으며 Park and Lee[19]는 전철역 네트워크에서 역 간 승객들의 흐름을 스마트교통카드 데이터로 추출하여 클러스터링을 수행하였다. 본 논문이 제안한 클러스터링 방법은 통행량 기반의 클러스터링을 수행한 이전 연구들에서 고려하지 않은 통행시간이라는 측면을 고려한 클러스터링이다. 통행시간은 통행량과 함께 교통 정책 수립에 있어서 중요한 역할을 하는 지표로 역들을 통행시간을 이용하여 클러스터링함으로써 통행량 기반의 클러스터링 결과에 대한 보완적인 역할을 할 것을 기대할 수 있다. 따라서, 본 연구에서는 출·퇴근 시간 각 역에서 출발하고 도착하는 승객들의 통행시간을 스

마트교통카드 데이터로부터 추출한 다음 이를 각 역의 특성 벡터로 정의하여 클러스터링을 수행하였다.

본 논문의 구성은 다음과 같다. 제2장에서는 본 연구와 관련 있는 기존 연구를 소개하고 제3장에서는 통행시간을 이용한 전철역 클러스터링의 방법을 소개할 것이다. 제4장에서는 스마트교통카드 데이터를 이용해 제3장에서 제안한 클러스터링 방법을 적용하여 역들을 클러스터링한 후 그 결과를 분석할 것이다. 마지막으로 제5장에서는 본 논문의 내용을 요약하고 연구의 한계점 및 추후 연구 방향을 소개할 것이다.

## 2. 문헌 연구

초기 스마트교통카드 데이터에 관한 연구는 데이터 자체에 누락된 정보를 보완하는 연구가 주를 이루었다. 버스의 경우 하차 태깅이 누락되는 문제와 전철의 경우 환승 정보가 누락되는 문제를 해결하기 위한 연구가 이에 해당한다. Bagchi and White[1], Utsunomiya et al.[21], Hofmann and O'Mahony[7], Seaborn et al.[20], Ma et al.[14], Hong et al.[8]은 버스와 지하철을 대상으로 교통카드 데이터의 둘 이상의 트랜잭션을 연결하는 방식으로 환승역 혹은 환승 정거장을 추론하는 연구를 수행하였다. Trepanier et al.[20], Munizaga and Palma[18], Gordon et al.[5]은 버스의 트랜잭션 데이터에서 누락되어 있는 하차 정거장에 관한 정보를 추론하는 방법을 제안하였다.

스마트교통카드 데이터는 개인의 교통 이력을 모두 담고 있기에 이를 이용할 경우 기존의

가구통계 조사보다 훨씬 정확한 통행 관련 정보를 얻을 수 있을 것으로 기대되었다. Munizaga and Palma[17]는 여러 트랜잭션을 하나의 통행 여정을 의미하는 체인으로 묶은 다음 이를 기반으로 OD-행렬(Origin/Destination-matrix)을 만드는 방법을 제안하여 기존의 가구통계 조사로부터 추출한 OD-행렬을 스마트교통카드 데이터로부터 얻을 수 있음을 확인하였다. 그뿐만 아니라, Ha and Lee[6]은 한주 분량의 교통카드 데이터로부터 통근 패턴 및 통근 네트워크의 구조를 분석하여 가구통계 조사 자료와 비교를 통해 두 데이터 사이의 일관성을 확인하였다.

스마트교통카드 데이터로부터 추출한 개인의 모빌리티 정보는 대중교통 패턴을 이해하고 이를 기반으로 더 좋은 교통서비스를 개발하는 등의 정책 제안에 유용할 것으로 기대된다. 예를 들어 대중교통 패턴을 이해하기 위해 Lee et al.[13]은 주요 승하차 지역의 통행패턴과 이에 영향을 주는 요인에 대한 회귀분석을 수행하였다. 그리고 Morency et al.[15]과 Kieu et al.[10]은 각각 K-평균 기법과 K-평균++ 클러스터링 알고리즘을 적용하여 통행패턴을 몇 개의 클러스터로 나누는 분석을 시도하였다. 이러한 통행패턴의 분석결과는 교통 정책 수립에 도움을 줄 수 있다. 예를 들어, Jun et al.[9]과 Zhou et al.[23]은 도시의 토지 이용과 교통 시스템 사이의 영향을 분석하였다. 따라서, 역 중심의 통행패턴을 분석한 Mahrsi et al.[3]를 토대로 각 역이 전체 도시의 교통 시스템에 어떤 역할을 수행하는지를 파악하는 것이 가능하며 이는 더 좋은 교통서비스를 위한 정책 계획에 활용될 수 있다.

본 논문은 전철역들의 클러스터링을 다룬

연구로 같은 문제를 다룬 연구로는 Min[15]과 Park and Lee[19]의 연구가 있다. Min[15]의 연구는 스마트교통카드 데이터로부터 매 시간대 역에 입장하고 퇴장하는 승객들을 추출하여 시간대별 승객수로 구성된 특성 벡터를 역마다 정의했다. 그리고, 역들의 특성 벡터들이 가우시안 혼합 모델을 따른다고 가정하고 추출된 데이터로 이를 추정하였다. 가우시안 혼합 모델에서는 여러 개의 가우시안 분포가 도출되는데 각 가우시안 분포가 하나의 클러스터를 설명한다고 해석하였다. Min[15]은 4개의 클러스터를 가정한 가우시안 혼합 모델(Gaussian Mixture Model)을 도출하였으며 이를 통해 4개의 클러스터들을 각각 환승지역(상업지역), 주거지역, 업무지역, 기타지역으로 분류하였다. Park and Lee[19]의 연구에서는 역을 노드로 표현한 도시철도 네트워크를 정의한 다음 두 노드 사이의 통행량을 호(arc 또는 edge)의 가중치로 정의하였다. 이때, 두 역 사이에 둘 이상의 경로가 존재하는 경우 별도의 방법으로 최단경로를 계산하여 호의 가중치를 결정하는 데 사용하였다. 이러한 방법으로 가중치를 갖는 그래프를 정의한 다음 그래프 분할을 이용하여 그래프의 노드(전철역)들을 클러스터링하였다. Park and Lee[19]는 전철역들이 지리적으로 동서남북으로 분류되는 것을 보였다. 하지만 두 연구 모두 각 역에서 출발하거나 도착하는 통행량만을 기준으로 역들을 클러스터링 했기 때문에 통행량 이외의 통행특성을 반영하지 못한다는 한계가 있다. 본 논문은 통행시간을 고려한 클러스터링 방법을 제안함으로써 기존 연구가 고려하지 못한 통행시간을 전철역 분류에 활용할 수 있는 방법을 제공한다.

### 3. 확률분포 파라미터 기반의 역 클러스터링

전철역을 클러스터링하기 위해서는 1) 첫 번째로 각 전철역을 설명하는 특성 벡터를 정의해야 하고 2) 클러스터링에 사용할 기법을 개발 또는 선택해야 한다. 본 논문에서는 각 역에 출발 및 도착하는 승객들의 통행시간을 와이블 분포로 모형화하고 데이터로부터 추정된 와이블 분포의 형상모수를 역의 특성 벡터로 사용한다. 와이블 분포[17]는 제품의 수명을 확률적으로 모형화하여 신뢰성 분석에 널리 사용된다. 본 논문에서 통행시간을 와이블 분포로 모형화한 이유는 데이터로부터 통행시간의 분포를 그린 결과 와이블 분포의 파라미터를 적절하게 조절할 경우 통행시간의 분포가 와이블 분포와 유사한 모습을 갖는 것을 관찰했기 때문이다. 그리고 클러스터링 기법으로는 잘 알려진 클러스터링 기법인 K-평균 클러스터링(K-means clustering) 기법을 사용한다.

특성 벡터는 총 4개의 값으로 구성되어 있으며 이들은 모두 특정 통행시간 데이터를 이용하여 와이블 분포의 형상모수를 추정한 값이다. 이들 4개의 값은 다음과 같은 통행시간 데이터들이 각각 독립적인 와이블 분포를 따른다고 가정하여 얻은 값이다.

- 출근 시간 역에서 출발한 승객들의 통행시간
- 출근 시간 역에 도착한 승객들의 통행시간
- 퇴근 시간 역에서 출발한 승객들의 통행시간
- 퇴근 시간 역에 도착한 승객들의 통행시간

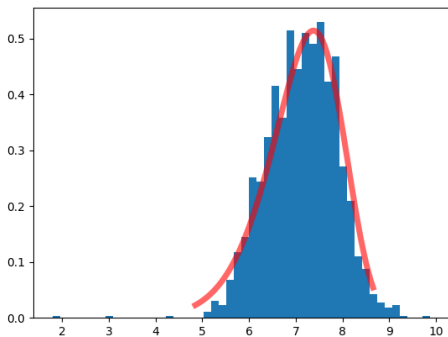
출근 시간과 퇴근 시간에는 승객들의 통행패턴이 비교적 정기적이지만 그 외의 시간대에는 승객들의 통행패턴이 비정기적이라는 가정에

서 출근 시간과 퇴근 시간만을 고려하였다. 그리고, 전철역의 속성에 따라 해당 시간대에 역에서 출발하는 승객과 도착하는 승객들의 통행시간에는 차이가 있을 것이라는 가정에서 출발과 도착 통행시간을 구분하였다. 승객들의 통행시간은 스마트교통카드 데이터의 “INVTIME” 항목에 기록이 되어 있는 초로 환산된 통행시간(퇴장 시간-입장 시간)을 사용하였으며 이 값을 그대로 사용하는 대신 로그변환(log transformation)을 적용한 값을 사용하였다 통행시간은 초 단위로 작성되어 긴 통행시간의 경우 3600 이상의 값까지 가질 수 있어서 값의 스케일 조정의 목적으로 로그변환을 사용하였다.

본 논문에서는 통행시간이 와이블 분포를 따른다고 가정한다. 와이블 분포는 다음의 확률 밀도함수를 갖는 분포로 변수의 값이 0보다 같거나 큰 경우를 가정한다[15].

$$f(x;\lambda;k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \quad (1)$$

<그림 1>는 출근 시간(오전 7시에서 오전 9시) 강남역에서 출발한 승객들의 통행시간 도수분포와 이를 와이블 분포로 추정한 결과이다.



<그림 1> 출근시간 강남역 출발 승객들의 도수분포와 추정된 와이블 분포

와이블 분포는 제품의 수명을 모형화하는 데 주로 사용하는 분포로 이를 결정하는 모수에는 형상모수(k)와 척도모수(λ)가 있다. 본 논문에서는 출·퇴근 시간 출발·도착 승객의 통행시간을 각각 와이블 분포로 추정하는 데 사용하여 4개의 형상모수를 얻었다. 그리고 출근 시간 출발에 해당하는 형상모수 값을 기준으로 형상모수들을 정규화하였다. 따라서, 모든 역은 출근 시간 출발에 해당하는 형상모수 값으로는 1의 값을 갖고 나머지 형상모수들은 출근·출발 형상모수에 대한 상대적인 크기에 해당하는 값을 갖는다.

각 역에 대해 정의한 특성 벡터를 이용하여 클러스터링을 수행한다. 본 논문에서는 K-평균 클러스터링 기법을 사용하여 클러스터링을 수행하였다. 형상모수의 값에 따라 와이블 분포의 모양이 달라지므로 형상모수들의 벡터로 구성된 공간은 유클리디안 공간으로 가정하였다. 따라서, 각 벡터 사이의 거리로 유클리디안 거리를 사용하였다.

## 4. 실험결과

### 4.1 실험 데이터

스마트교통카드 데이터로부터 추출한 승객들의 통행시간으로부터 와이블 분포의 형상모수를 추출한 다음, 이를 기반으로 역들을 클러스터링하는 방법론을 2016년 11월 13일자 스마트교통카드 데이터를 이용하여 실증 분석하였다. 이 데이터는 총 5,121,911건의 수도권 도시철도(인천지역 포함) 이용고객들의 데이터를 포함하고 있으며 코드 오류를 갖는 데이터를 삭제하면 총 5,028,425건의 통행실적이다. 출근

시간은 오전 7시부터 9시까지 2시간, 퇴근 시간은 오후 6시부터 8시까지의 2시간으로 가정하였다. 각 역에서 특정 시간대에 출발한 데이터는 해당 역의 입장 시간이 해당 시간대에 속하는 승객들로 구성되어 있고 각 역에 도착한 데이터는 해당 역의 퇴장 시간이 해당 시간대에 속하는 승객들로 구성되어 있다. 클러스터링에 충분한 수의 역과 승객들을 갖는 호선들을 대상으로 하기 위해 2호선부터 7호선까지의 6개 호선만을 대상으로 분석을 수행하였다. 각 승객의 통행시간은 로그변환을 통해 전처리한 값을 사용하였다.

각 역에 대해서 출근 시간 출발 승객과 도착 승객, 퇴근 시간 출발 승객과 도착 승객의 4가지 경우 각각에 대해 통행시간에 대한 와이블 분포의 형상모수를 추정하였다. 따라서, 각 역에 대해서 4개의 특성값이 정의된다. 다음으로 출근 시간 출발 승객의 형상모수 값을 기준으로 정규화를 수행하였다. 따라서, 출근 시간 출발 승객의 형상모수는 정규화 이후 항상 1의 값을 갖게 되며 나머지 형상모수의 경우 1보다 크면 출근 시간 출발 승객에 대한 형상모수보다 큰 값을 갖는다는 것을 의미하고 1보다 작으면 출

근 시간 출발 승객에 대한 형상모수보다 작은 값을 갖는다는 것을 의미한다.

본 논문의 실험은 하루의 스마트교통카드 데이터에 기반하기 때문에 요일별/계절별 등 시간에 따라 달라지는 통행시간의 패턴은 반영하지 못한다. 하지만, 실험결과에서 확인할 수 있듯이 하루 데이터만으로도 전철역들의 의미있는 클러스터가 형성됨을 관찰할 수 있었다.

#### 4.2 클러스터링 결과 분석

각 역에 정의된 4개의 값을 각 역의 특성벡터로 정의한 다음 K-평균 클러스터링(K-means clustering)을 수행하였다. 파이썬으로 구현된 scikit-learn으로 구현하였으며 Min[15]의 결과와의 비교를 위해 클러스터의 개수를 4로 고정하였다. 호선별로 역의 수와 역 간 이동시간, 다른 호선과의 환승 연결 여부 등이 다르기에 호선 간에 통행시간의 근본적인 차이가 발생할 수 있다. 따라서 이러한 요인들의 영향을 최소화하기 위해 클러스터링은 전체 역을 한 번에 분석하는 대신 호선별로 수행하였다.

먼저 2호선의 클러스터링 결과는 <표 1>과

<표 1> 2호선 클러스터링 결과 (을지로입구역 제외)

클러스터	클러스터 중심 정보				역 목록
	출근시간		퇴근시간		
	출발	도착	출발	도착	
1	1.00	1.25	0.79	0.71	서초, 신촌, 역삼, 을지로3가, 충정로(경기대입구)
2	1.00	1.15	1.69	1.74	교대(법원, 경찰청), 양천구청, 한양대
3	1.00	1.02	0.97	1.01	강변(동서울터미널), 건대입구, 구의(광진구청), 낙성대, 당산, 봉천, 사당, 상왕십리, 서울대입구(관악구청), 선릉, 신당, 신대방, 신설동, 아현, 영등포구청, 왕십리(성동구청), 용답, 이대, 잠실(송파구청), 잠실나루, 합정
4	1.00	1.12	1.28	1.21	강남, 구로디지털단지, 뚝섬, 문래, 방배, 삼성(무역센터), 성수, 시청, 신정네거리, 신천, 을지로4가, 홍대입구

같다. <표 1>의 두 번째부터 네 번째까지의 열은 클러스터 중심지에 대한 정보로 출근 시간과 퇴근시간대의 출발 승객과 도착 승객의 이동시간을 대상으로 한 와이블 분포의 형상모수 추정값이다. 이들 값은 소수점 세 번째 자리에서 반올림한 결과이다. 클러스터 1은 출근 시간의 형상모수 값들이 퇴근 시간의 형상모수 값들보다 큰 중심을 갖는 클러스터이고 클러스터 2와 4는 클러스터 1과 반대 성질의 중심을 갖는 클러스터들이다. 특히, 클러스터 2의 경우, 출근 시간과 퇴근 시간의 형상모수의 차이가 크다. 반면, 클러스터 3은 출근 시간과 퇴근 시간의 형상모수 값이 거의 같은 중심을 갖는 클러스터이다. 클러스터 3에 속하는 역들은 낙성대, 봉천과 같이 주거 구역이 밀집된 지역의 전철 역들이 다수 포함된 것을 관찰할 수 있다. 하지만, Min[15]이 도출한 역 분류에서 상업지역, 업무지역에 포함된 역들도 일부 포함되어 있다는 점에서 순수하게 주거지역에 해당하는 역들만으로 구성되어 있다고 결론 지을 수는 없다. 2호선 역들을 대상으로 Min[15]의 클러스터링 결과와 본 논문의 클러스터링 결과를 비교한

결과는 <표 2>와 같다.

<표 2>에서 확인할 수 있듯이 본 연구의 클러스터링 결과에서 클러스터1과 클러스터2는 모두 Min[15]에서의 업무지역에 해당하는 클러스터에 해당하지만, 클러스터 3은 상업지역과 주거지역이 혼재되어 있으며 클러스터 4는 기타지역과 업무지역이 혼재된 클러스터이다. 하지만 Min[15]의 클러스터링 결과에서 각 지역으로 분류된 역들이 반드시 하나의 지역에만 속한다고 결론지을 수 없다는 점에서 Min[15]과 본 논문의 클러스터링 결과를 종합적으로 해석할 경우, 각 역을 더 세분화하여 특징을 정의할 수 있다. 예를 들어, 강남역과 서초역의 경우 Min[15]에서는 모두 업무지역으로 분류되었으나 본 논문의 클러스터링 결과에 따르면 두 역을 이용하는 시민들의 통행시간 분포는 다른 양상을 갖는다는 점에서 두 역은 동일 업무지역 중심의 역이지만 다른 특성을 갖는다고 말할 수 있다.

<표 3>은 2호선을 포함하여 나머지 호선들인 3호선부터 7호선에 대해서 각각 클러스터링을 수행한 다음 도출된 클러스터들의 중심을

<표 2> 2호선 클러스터링 결과와 Min(14)의 결과 비교

	클러스터1	클러스터2	클러스터3	클러스터4
상업지역	신촌		강변(동서울터미널), 건대입구, 당산, 사당, 신당, 왕십리(성동구청), 이대, 잠실(송파구청)	신천, 홍대입구
기타지역	충정로		상왕십리, 신설동, 영등포구청, 합정	구로디지털단지, 독섬, 문래, 방배, 성수
주거지역		양천구청,	구의(광진구청), 낙성대, 봉천, 서울대입구(관악구청), 신대방, 아현, 용답, 잠실나루	신정네거리
업무지역	서초, 역삼, 을지로3가	교대(법원, 경찰청), 한양대	선릉	강남, 삼성(무역센터), 시청, 을지로4가

〈표 3〉 호선별 클러스터 중심

호선	클러스터	출근시간		퇴근시간		호선	클러스터	출근시간		퇴근시간	
		출발	도착	출발	도착			출발	도착	출발	도착
2	1	1.00	1.25	0.79	0.71	5	1	1.00	1.07	1.28	1.28
	2	1.00	1.15	1.69	1.74		2	1.00	0.93	1.83	2.07
	3	1.00	1.02	0.97	1.01		3	1.00	0.89	0.91	0.96
	4	1.00	1.12	1.28	1.21		4	1.00	1.01	0.63	0.52
3	1	1.00	0.94	0.69	0.77	6	1	1.00	0.99	1.37	1.50
	2	1.00	1.27	1.58	1.55		2	1.00	0.86	0.44	0.52
	3	1.00	1.16	1.02	1.00		3	1.00	0.92	0.84	0.93
	4	1.00	1.14	2.42	2.31		4	1.00	1.13	2.14	2.15
4	1	1.00	0.81	0.75	0.85	7	1	1.00	0.92	0.89	0.96
	2	1.00	1.35	2.26	2.15		2	1.00	0.93	1.90	1.93
	3	1.00	1.20	1.16	1.13		3	1.00	1.13	1.30	1.29
	4	1.00	0.97	1.75	1.62		4	1.00	0.83	0.53	0.60

정리한 결과이다. 2호선의 클러스터 3과 유사하게 출근 시간과 퇴근 시간의 형상 모수 값이 거의 같은 중심을 갖는 클러스터들은 3호선의 클러스터 3, 5호선의 클러스터 3, 6호선의 클러스터 3, 7호선의 클러스터 1이 있다. 이들 클러스터도 Min[15]이 도출한 역 분류에서 주거지역에 해당하는 역들이 상당수 포함되어 있으나 상업지역과 업무지역에 속하는 역들 또한 일부 포함되어 있다.

### 5. 결 론

본 논문은 대표적인 모빌리티 데이터인 스마트교통카드 데이터로부터 승객의 통행시간을 추출하여 전철역들을 클러스터링하는 문제를 고려하였다. 각 역의 승객들을 출근 시간 출발, 출근 시간 도착, 퇴근 시간 출발, 퇴근 시간 도착 승객들로 분류한 다음 각각의 통행시간을 와이

블 분포로 모형화하여 형상모수를 추정하여 역별로 4개의 특성값을 정의하였다. 그리고 특성 벡터들을 K-평균 클러스터링 기법을 사용하여 클러스터링하였다. 역들의 클러스터들을 분석한 결과 통행량만을 고려한 Min[15]의 클러스터링 결과에 통행시간을 고려할 경우 클러스터들이 더 세분화 될 수 있음을 관찰하였다.

하지만, 본 연구의 클러스터링 결과는 다음과 같은 한계점을 갖고 있다. 첫 번째는 도출된 클러스터들의 해석에 있어 체계적이고 엄밀한 분석이 부족하다. Min[15]의 클러스터 분류 기준인 상업, 주거, 업무, 기타의 네 가지 분류에 대해서만 비교를 수행했는데 각 전철역의 입지 분석, 인구 분포 등을 수행할 경우 더욱 체계적인 클러스터 해석이 가능할 것이다. 두 번째는 Min[15]의 연구와의 비교를 위해 클러스터의 개수를 4개만 고려했는데 개수를 증가시킬 경우 더욱 세분화 된 클러스터링이 가능할 수 있다. 따라서, 각 역에 대한



체계적인 입지, 인구 분포 분석 등과 함께 다양한 개수의 클러스터를 갖도록 클러스터링한 결과를 함께 고려하는 것은 후속 연구를 위한 중요한 연구 방향이라 할 수 있다. 세 번째로 실험에서 하루의 데이터만을 고려했는데 여러 날짜의 데이터를 고려할 경우 와이블 분포의 형상모수를 더욱 정확하게 추정할 수 있을 뿐 아니라 클러스터의 시간에 대한 변화도 분석할 수 있다. 이 연구 방향은 후속 연구로 남긴다.

---

## References

---

- [1] Bagchi, M. and White, P. R., “The potential of public transport smart card data,” *Transport Policy*, Vol. 12, No. 5, pp. 464-474, 2005.
- [2] Ebrahimpou, Z., Wan, W., Cervantes, O., Luo, T., and Ullah, H., “Comparison of main approaches for extracting behavior features from crowd flow analysis,” *ISPRS International Journal of Geo-Information*, Vol. 8, No. 10, p. 440, 2019.
- [3] El Mahrsi, M. K., Come, E., Oukhellou, L., and Verleysen, M., “Clustering smart card data for urban mobility analysis,” *IEEE Transactions on Intelligent Transportation Systems*, Vol. 18, No. 3, pp. 712-718, 2017.
- [4] Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L., “Understanding individual human mobility patterns,” *Nature*, Vol. 453, No. 7196, pp. 779-782, 2008.
- [5] Gordon, J., Koutsopoulos, H., Wilson, N., and Attanucci, J., “Automated inference of linked transit journeys in London using fare-transaction and vehicle location data,” *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2343, No. 1, pp. 17-24, 2013.
- [6] Ha, J. and Lee, S., “The estimation of commuting patterns and the analysis of the commuting network structure using smart card data: Focused on the possibility of application through the validation process with household travel survey data,” *Journal of Korea Planning Association*, Vol. 51, No. 4, pp. 123-143, 2016.
- [7] Hofmann, M. and O’Mahony, M., “Transfer journey identification and analyses from electronic fare collection data,” In the *Proceedings of IEEE Intelligent Transportation Systems Conference*, pp. 34-39, 2005.
- [8] Hong, S.-P., Min, Y.-H., Park, M.-J., Kim, K. M., and Oh, S. M., “Precise estimation of connections of metro passengers from smart card data,” *Transportation*, Vol. 43, pp. 749-769, 2016.
- [9] Jun, M. J., Choi, K., Jeong, J. E., Kwon, K. H., and Kim, H. J., “Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul,” *Journal of Transport Geography*, Vol. 48, pp. 30-40, 2015.
- [10] Kieu, L. M., Bhaskar, A., and Chung, E., “Passenger segmentation using smart

- card data,” *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 3, pp. 1537-1548, 2015.
- [11] Kim, K., Oh, K., Lee, Y., and Jung, J., “Discovery of travel patterns in Seoul Metropolitan subway using big data of smart card transaction systems,” *The Journal of Society for e-Business Studies*, Vol. 18, No. 3, pp. 211-222, 2013.
- [12] Kim, S. K., “Plans for raising the utilization of smart card data,” *KRIHS Monthly Magazine*, Vol. 205, pp. 18-24, 2015.
- [13] Lee, M., Han, J., and Lee, H., “Analysis of the transit ridership pattern using transportation card data: Focusing on Ganghwa,” *The Journal of Korea Institute of Intelligent Transportation Systems*, Vol. 17, No. 2, pp. 58-72, 2018.
- [14] Ma, X. L., Wu, Y. J., Wang, Y. H., Chen, F., and Liu, J. F., “Mining smart card data for transit riders’ travel patterns,” *Transportation Research Part C: Emerging Technologies*, Vol. 36, pp. 1-12, 2013.
- [15] Min, M. K., “Classification of seoul metro stations based on boarding/alighting patterns using machine learning clustering,” *The Journal of the Institute of Internet, Broadcasting and Communication*, Vol. 18, No. 4, pp. 13-18, 2018.
- [16] Morency, C., Trepanier, M., and Agard, B., “Analysing the variability of transit users behaviour with smart card data,” In *Proceedings of IEEE Intelligent Transportation Systems Conference*, pp. 44-49, 2006.
- [17] Mudholkar, G. S. and Srivastava, D. K., “Exponentiated Weibull family for analyzing bathtub failure-rate data,” *IEEE Transactions on Probability*, Vol. 42, No. 2, pp. 299-302, 1993.
- [18] Munizaga, M. and Palma, C., “Estimation of a disaggregate multi-modal public transport origin-destination matrix from passive smartcard data from Santiago, Chile,” *Transportation Research Part C: Emerging Technologies*, Vol. 24, pp. 9-18, 2012.
- [19] Park, J. S. and Lee, K., “Classification of the seoul metropolitan subway stations using graph partitioning,” *Journal of the Economic Geographical Society of Korea*, Vol. 15, No. 3, pp. 343-357, 2012.
- [20] Seaborn, C., Attanucci, J., and Wilson, N., “Analyzing multimodal public transport journeys in London with smart card fare payment data,” *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2121, No. 1, pp. 55-62, 2009.
- [21] Trepanier, M., Tranchant, N., and Chapleau, R., “Individual trip destination estimation in a transit smart card automated fare collection system,” *Journal of Intelligent Transportation Systems*, Vol. 11, No. 1, pp. 1-14, 2007.
- [22] Utsunomiya, M., Attanucci, J., and Wilson, N., “Potential uses of transit smart card

registration and transaction data to improve transit planning,” *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1971, No. 1, pp. 119–126, 2006.

[23] Zhou, Q., Liu, S., and Wang, Y., “A study on the coordinative relation of land use and transport around the metro station,” *Railway Transport and Economy*, Vol. 40, No. 4, pp. 100–106, 2018.

## 저 자 소개



공인택

2013년

2015년

2021년

2021년~2022년

2022년~현재

관심분야

(E-mail: itgong87@naver.com)

인천대학교 무역학과 (학사)

인천대학교 동북아물류대학원 (석사)

인천대학교 동북아물류대학원 (박사)

인천대학교 국제물류유통BK교육연구팀 박사 후 연구원

LG CNS 책임

물류 및 SCM, 빅데이터, 수요예측



김동윤

2021년

2021년~현재

관심분야

(E-mail: dongyunkim@inu.ac.kr)

인하대학교 아태물류학부 (학사)

인천대학교 동북아물류대학원 석사과정 재학 중

물류 및 SCM, 빅데이터, 최적화



민윤홍

2006년

2012년

2012년~2017년

2017년~현재

관심분야

(E-mail: yunhong.min@inu.ac.kr)

포항공과대학교 산업경영 (학사)

서울대학교 산업공학 (박사)

삼성종합기술원

인천대학교 동북아물류대학원 교수

최적화, 인공지능