

통행시간 기반 혼합분포모형 분석을 통한 도시철도 승객의 급행 탑승 여부 추정 연구

Estimation of Mass Rapid Transit Passenger's Train Choice Using a Mixture Distribution Analysis

장진원* · 윤호상** · 박동주***

* 주저자 : 서울시립대학교 교통공학과 & 스마트시티학과 석박사통합과정

** 공저자 : 서울시립대학교 교통공학과 & 도시빅데이터융합학과 석사과정

*** 교신저자 : 서울시립대학교 교통공학과 & 도시빅데이터융합학과 교수

Jinwon Jang* · Hosang Yoon** · Dongjoo Park***

* Dept. of Transportation Engineering & Dept. of Smart Cities

** Dept. of Transportation Engineering & Dept. of Urban Big Data Convergence

*** Dept. of Transportation Engineering & Dept. of Urban Big Data Convergence

† Corresponding author : Dongjoo Park, djpark@uos.ac.kr

Vol.20 No.5(2021)

October, 2021

pp.1~17

pISSN 1738-0774

eISSN 2384-1729

[https://doi.org/10.12815/kits.](https://doi.org/10.12815/kits.2021.20.5.1)

2021.20.5.1

Received 21 July 2021

Revised 14 August 2021

Accepted 8 September 2021

© 2021. The Korea Institute of
Intelligent Transport Systems. All
rights reserved.

요약

대부분의 도시철도 시스템은 승객의 탑승열차 및 탑승열차종을 정확히 알 수 없다. 다수의 선행연구에서는 교통카드데이터와 열차시각표를 매칭하여 탑승열차를 추정하였으나, 추정이 불가능한 승객 또한 다수 존재한다. 본 연구의 9호선 사례분석 결과 교통카드데이터-열차시각표 매칭만으로는 약 28% 승객의 탑승열차종을 추정할 수 없음을 확인할 수 있었다. 이에 교통카드데이터-열차시각표 매칭과 본 연구에서 정의한 통행시간 기반 혼합확률분포분석을 순차적으로 적용하여 급행운영 도시철도노선 승객의 탑승열차종을 추정하는 방법을 개발하였다. 분석 결과, 298개 OD pair에서 본 연구의 검증 기준을 만족하는 합리적인 급행이용/비이용 승객 분류기준점을 도출할 수 있었다.

핵심어 : 도시철도, 급행, 탑승열차종, 통행시간분포, 교통카드

ABSTRACT

Identifying the exact train and the type of train boarded by passengers is practically cumbersome. Previous studies identified the trains boarded by each passenger by matching the Automated Fare Collection (AFC) data and the train schedule diagram. However, this approach has been shown to be inefficient as the exact train boarded by a considerable number of passengers cannot be accurately determined. In this study, we demonstrate that the AFC data - diagram matching technique could not estimate 28% of the train type selected by passengers using the Seoul Metro line no.9. To obtain more accurate results, this paper developed a two-step method for estimating the train type boarded by passengers by applying the AFC data - diagram matching method followed by a mixture distribution analysis. As a result of the analysis, we derived reasonable express train use/non-use passenger classification points based on 298 origin-destination pairs that satisfied the verification criteria of this study.

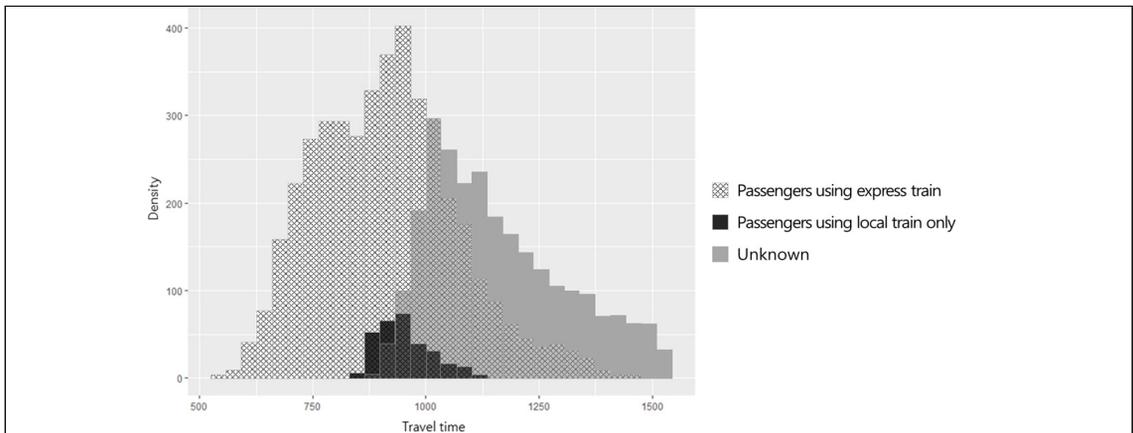
Key words : Subway, Express train, Train type, Travel time distribution, AFC data

I. 서론

전세계의 도시철도 네트워크는 점점 더 복잡해지고 길어지고 있다. 도시철도는 출발지와 목적지 간 모든 역에서 정차하기 때문에, 네트워크가 길어질수록 속도 경쟁력이 저하되는 것이 문제점으로 지적되고 있다.(Kim et al., 2013) 이를 극복하기 위해 급행 시스템이 도입되었다. 문제는 대부분의 도시철도 시스템은 역 구내에서 개집표를 하고 열차 탑승 시에는 별도의 개집표를 하지 않아 승객의 탑승열차종, 즉 급행열차 이용 여부를 알 수 없으며 정확한 탑승열차 또한 알 수 없다는 것이다. 탑승열차를 알 수 있다면 열차종별 이용비율이나 통행거리 등 도시철도 내 다양한 통행특성에 대한 분석이 가능하다(Kim et al., 2015a)는 점에서 이는 중요하다. 통행특성에 대한 분석 결과는 급행열차 정차패턴·배차간격 변경 등 도시철도 운영 개선에도 활용 가능하다.

다양한 선행연구가 도시철도 승객의 탑승열차종을 추정하기 위해 노력하였다. 우선 탑승열차종 추정을 교통공학의 수단선택 개념으로 접근하여 로짓모형(Logit Model)을 활용한 연구가 있다. Baek and Sohn(2016)은 서울지하철 9호선 승객 255명을 대상으로 설문조사를 수행하여 열차종별 이용비율과 그것의 영향변수를 분석하였다. 선행연구가 가장 많이 활용한 것은 교통카드데이터-열차시각표 매칭 방법이다.(Kim et al., 2015a 등) 이는 교통카드데이터를 통해 알 수 있는 승객의 입퇴장역 및 입퇴장시각과 정보와 열차시각표 정보를 대조하여 탑승가능열차를 추정하는 방법이다. 수단선택 결과가 확률 형태로 산출되는 로짓모형과 달리 탑승가능열차 추정결과가 확정적(deterministic)이고, 원리가 단순하다는 장점이 있어 많이 활용되었다.

그러나 본 연구 결과, 교통카드데이터-열차시각표 매칭 방법론을 통해 탑승열차종을 명확히 밝혀낼 수 없는 경우 또한 다수 존재한다는 것을 알 수 있었다. <Fig. 1>은 교통카드데이터-열차시각표 매칭을 통해 노랑진-고속터미널 간 승객의 탑승열차를 추정한 결과이다. 약 40%에 해당하는 승객들은 입퇴장시각을 만족하는 복수의 열차가 존재하여 탑승열차종을 특정할 수 없는 사례이다.



<Fig. 1> A case of estimation results of the train type boarded by passengers (Noryangjin - Express Bus Terminal)

급행을 운영하는 도시철도노선 승객의 통행시간분포는 급행이용승객의 통행시간분포와 급행비이용승객의 통행시간분포가 혼합되어있는 혼합확률분포라고 볼 수 있다. 따라서 급행이용승객의 통행시간분포와 급행비이용승객의 통행시간분포의 혼합 형태를 분석할 수 있다면, 각 승객의 통행시간을 통해 해당 승객의 탑승열차종을 확률적으로 산출할 수 있다. 이를 ‘통행시간기반 혼합분포모형 분석’ 방법으로 정의한다. 본 연구는

통행시간기반 혼합분포모형 분석의 신뢰성을 검증한 후, 교통카드데이터-열차시각표 매칭 방법을 통해 탑승 열차종을 특정할 수 없는 승객 데이터를 대상으로 해당 방법론을 순차적으로 적용하여 탑승열차종 추정 성공률을 제고하는 것을 목적으로 한다. 이를 통해 비교적 단순한 방법으로 교통카드데이터-열차시각표 매칭 방법의 단점을 보완할 수 있을 것으로 기대된다.

II. 선행연구고찰

본 연구는 급행이 운영되는 도시철도 노선에서 승객의 탑승열차종을 추정하는 것을 목적으로 한다. 이때 도시철도 급행시스템은 크게 복복선 방식과 대피선 방식, skip-stop 방식으로 구분된다.(Lee et al., 2018) 본 연구는 대피선 방식 급행운영 노선에 초점을 맞춘다. 대피선 방식은 일정 간격의 역에 대피선 건설이 필요하기 때문에 최적 급행정차역 및 대피선 위치 선정, 급행운행방안 수립과 관련된 연구가 꾸준히 이루어졌다. 이 연구들은 열차 통행시간 최소화(Kim et al., 2013; Lee et al., 2018; Soo et al., 2006; Sohn, 2007), 총 승객 통행시간 최소화(Katori et al., 2004; Luo et al., 2012) 등을 목적함수로 하였다.

탑승열차 추정을 위한 교통카드데이터-열차시각표 매칭 연구는 크게 두가지 방향으로 진행되고 있다 (Table 1). 하나는 복수의 대안경로가 존재하는 도시철도 네트워크 OD pair에서 승객의 통행경로선택비율을 추정하는 연구이다. Zhu et al.(2017), Zhu and Xu(2016)은 승객의 통행경로를 추정한 대표적인 연구로, 복수의 대안경로가 존재하는 상하이 도시철도 구간에서 경로별 선택비율을 산출하였다. Sun and Xu(2012)는 통행과정을 입장도보시간, 대기시간 등으로 세분화하여 분석을 수행하였으며, 플랫폼 대기시간 및 열차용량부족으로 인한 열차탑승실패확률 등을 산출하여 이를 통해 경로선택비율을 도출하였다. 그리고 Sun and Schonfeld(2016)는 Sun and Xu(2012) 연구를 기반으로 열차탑승실패(Fail to Board) 확률 문제를 더욱 자세히 분석하였다. Othman et al.(2015)은 교통카드데이터-열차시각표 매칭과 동일한 방법론을 적용하였으나, 실제 열차시각표가 아닌 Agent Based Model에 기반한 싱가포르 도시철도 시뮬레이션 내에서 생성된 열차운행시간을 적용하여 승객의 경로선택확률을 시뮬레이션화하였다. Hong et al.(2016)은 Kusakabe et al.(2010)에서 개념화한 퇴장시각-입장시각 기준 승객 데이터 군집화 방법을 적용하여 서울지하철 네트워크에서 승객의 경로선택비율을 산출하였다. Kim et al.(2015b)은 이를 기반으로 혼잡이 경로선택에 미치는 영향을 분석하였다.

교통카드데이터-열차시각표 매칭 연구의 다른 한 방향은 본 연구의 목적과 같이 급행운영노선에서 승객의 탑승열차종을 추정하는 연구이다. Kusakabe et al.(2010)은 일반열차와 다종의 급행열차, 특급열차가 대피선 방식으로 운행되는 도시철도노선을 대상으로 탑승열차를 추정하는 연구를 하였다. 이때 퇴장시각을 x 축, 입장시각을 y 축으로 승객 데이터를 위치시키면 열차시각을 중심으로 군집화된다는 점을 활용하여 군집별 탑승열차를 추정하였다. Kim et al.(2015a; 2016)은 이러한 군집화 방법을 적용하여 서울 지하철 9호선을 대상으로 탑승열차를 추정하였다.

한편 승객의 통행시간분포를 통해 탑승열차 또는 통행경로를 추정하는 방법 또한 교통카드데이터-열차시각표 매칭 방법의 보완수단으로서, 또는 독립적인 방법으로서 연구되어 왔다. Sun and Xu(2012)와 Sun and Schonfeld(2016)은 실제 승객들의 도보시간과 속도분포를 현장관측을 통해 산출하는 등 세부통행과정별 시간분포를 세밀하게 산출하였다. 이를 통해 게이트-플랫폼간 도보 등 세부통행과정에 고정적인 소요시간이 아닌 확률적인 소요시간을 적용하였다.

통행시간을 이용하여 도시철도 네트워크 이용자의 통행행태를 분석하는 연구는 두 방향으로 진행되고 있다(Table 2). 하나는 통계적인 모형을 이용하여 승객의 통행행태를 추정하는 연구이다. Lee et al.(2017)은

통행시간을 Tag-in부터 열차 탑승 플랫폼까지의 도보시간, 열차 대기시간, 차내시간, 환승이동시간, 환승대기시간, 열차 하차 플랫폼부터 Tag-out까지의 도보시간으로 분리하고, 각각의 시간이 승객의 통행행태에 미치는 영향을 분석하였다. Zhou et al.(2015)은 각 시간으로부터 확률분포를 생성한 뒤, 해당 확률분포들을 이용하여 개인의 이용경로를 추정하였다. Cheng et al.(2019)의 연구에서는 데이터 마이닝을 통해 비합리적인 통행행태를 나타내는 승객을 제외하였고, 승객들의 통행시간분포가 로그정규분포로 나타나는 것을 확인하였다. 또한, 승객의 통행시간에 따라 어떤 경로선택을 하는지 확률분포모형을 통해 추정하였다.

통행시간을 이용하여 통행행태를 분석하는 다른 방향은 빅데이터의 머신러닝 기법을 적용하는 것이다. Hong et al.(2017)은 Tag-in, Tag-out 정보와 열차시간표를 매칭하여 승객의 통행시간으로부터 순수 이동시간을 추정한 뒤, 군집분석을 적용하여 통행배정을 수행하는 방안을 제시하였다. Wu et al.(2019)은 같은 OD pair를 통행하는 승객들의 통행시간 집합에 군집분석을 적용하여 여러 계층으로 구분한 뒤, 각 계층의 경로선택을 Fuzzy matching을 통해 추정한 방법론을 제안하였다.

<Table 1> Literature review : AFC data - diagram matching method

Author	Purpose	Walking time	Location	Sample size	Remarks
Zhu et al.(2017)	Path Choice Analysis	deterministic	Shanghai	57	-
Zhu and Xu(2016)	Path Choice Analysis	deterministic	Shanghai	150	-
Sun and Xu(2012)	Path Choice Analysis	stochastic	Beijing	-	Segmenting the travel process
Sun and Xu(2016)	Path Choice Analysis	stochastic	Beijing	3,760	-
Othman et al.(2015)	Path Choice & Congestion Analysis	stochastic	Singapore	-	Singapore MRT simulation based on Agent-Based Model
Hong et al.(2015)	Path Choice Analysis	stochastic	Seoul	47,618,710	Clustering
Kim et al.(2015)	Path Choice Analysis	stochastic	Seoul	1,898,397	
Kusakabe et al.(2010)	Train Type Choice Analysis	-	Osaka	662,419	
Kim et al.(2015)	Train Type Choice Analysis	-	Seoul metro No.9	86,518	
Kim et al.(2016)	Train Type Choice Analysis	-	Seoul metro No.9	411,032	

<Table 2> Literature review : travel time distribution analysis

Author	Purpose	Location	Sample size	Methodology
Lee et al.(2017)	Travel Behavior Analysis	Shenzhen	15,466,305	Travel time decomposition,
Zhou et al.(2015)	Path Choice Analysis	Shanghai	10,000	Travel time decomposition, probability model
Cheng et al.(2019)	Path Choice Analysis	Chengdu	31,714,394	Data mining, mixture probability model
Hong et al.(2017)	Traffic Assignment	Shanghai	-	Travel time decomposition, clustering
Wu et al.(2019)	Path Choice Analysis	Beijing	-	Clustering, fuzzy matching algorithm

III. 방법론

1. 분석과정

교통카드데이터 - 열차시각표 매칭 방법은 승객의 탑승열차까지 추정이 가능하나, <Fig. 1>과 같이 복수의 탑승가능열차가 존재하여 탑승열차를 추정할 수 없는 경우가 다수 존재한다. 반면 본 연구에서 정의한 통행시간기반 혼합분포모형 분석을 통해 알 수 있는 정보는 급행 이용/비이용 여부에 한정되나, 보다 확률이 높은 쪽을 명확히 알 수 있다. 따라서 본 연구에서는 통행시간기반 혼합분포모형 분석 방법이 교통카드데이터 - 열차시각표 매칭 방법을 보완해줄 수 있을 것으로 기대하였다. 탑승열차종 추정을 위한 본 연구의 과정은 다음과 같다.

① 교통카드데이터 - 열차시각표 매칭을 통한 1차 추정

교통카드데이터를 통해 승객의 입퇴장역과 입퇴장시각을 추출한 후, 각 승객이 입장시각-퇴장시각 이내에 탑승할 수 있는 열차를 열차시각표에서 탐색하는 방법을 통해 1차적으로 탑승열차종 추정

② 통행시간기반 혼합분포모형 분석

각 OD pair의 통행시간분포를 급행열차(ET, Express Train) 이용 승객의 통행시간분포와 급행열차 비이용 승객 (일반열차(LT, Local Train)만 이용한 승객)의 통행시간분포가 혼합된 분포로 가정하고, 급행열차 이용/비이용 승객의 통행시간분포를 분리, 급행열차 이용/비이용 승객 분류 기준점 설정

③ 통행시간기반 혼합분포모형 분석 결과의 검증

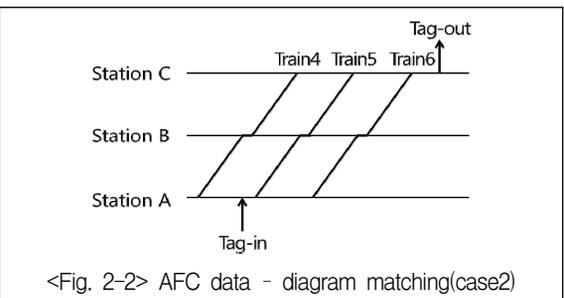
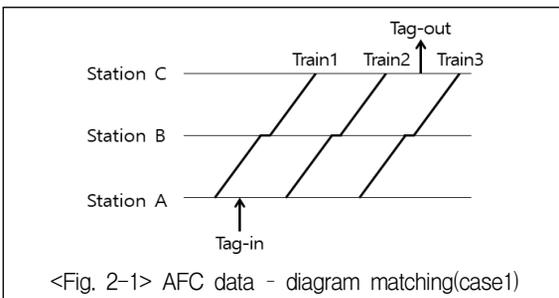
통계적 검정 및 교통카드데이터 - 열차시각표 매칭 결과 대입을 통한 오차율 검증

④ 통행시간기반 혼합분포모형 분석 결과를 통한 2차 추정

1차 추정으로 탑승열차종 추정이 불가능한 승객을 대상으로 재추정

2. 교통카드데이터 - 열차시각표 매칭

교통카드데이터 - 열차시각표 매칭 방법론은 다수의 연구에서 탑승열차를 추정하기 위해 활용되었다. 이는 교통카드데이터를 통해 승객의 입퇴장역과 입퇴장시각을 추출한 후, 각 승객이 입장시각-퇴장시각 이내에 탑승할 수 있는 열차를 열차시각표에서 탐색하는 방법론이다. 이 방법론의 기본적인 개념은 <Fig. 2-1>과 같다. 한 승객이 Station A에 입장한 후 Station C에서 퇴장하였다고 가정한다. 이 때 승객이 탑승가능한 열차는 [입장시각 + 입장도보시간] 이후에 Station A를 출발하고 [퇴장시각 - 퇴장도보시간] 이전에 Station C에 도착한 열차이다. 위 예시의 탑승가능열차는 Train2이다.



매칭 방법론은 복수의 탑승가능열차가 존재하는 승객이 발생하며(Fig. 2-2), 이 경우 실제 탑승열차를 특정할 수 없다는 문제가 존재한다. 탑승열차를 특정하기 위해 선행연구의 경우 승객이 차외시간을 최소화하려 한다는 가정을 추가하기도 하고(Kusakabe et al., 2010), 입퇴장시각이 비슷한 승객끼리 군집화하여 탑승열차를 추정하기도 하였다(Kim et al., 2015a; Kim et al., 2016). 그러한 가정이나 군집화의 경우 검증이 불가능하고, 승객의 탑승가능열차 추정과정에서 실제 탑승열차가 누락될 가능성이 존재한다는 한계를 가진다.

본 연구는 세가지 기초적인 가정외에 다른 가정을 적용하지 않았다.

- ① 승객은 퇴장역 방향과 반대방향으로 통행하지 않는다.
- ② 승객은 비합리적인 환승통행을 하지 않는다. 예를 들어 입장역과 퇴장역이 모두 급행정차역인 경우 급행-일반-급행과 같은 환승통행을 하지 않는다.
- ③ 환승을 통해 급행을 이용할 경우, 승객은 일반열차 이용 거리가 최소화되는 역에서 환승한다.

대피선 방식의 급행시스템은 일부 역만 급행이 정차하기 때문에, 입퇴장역의 급행정차여부에 따라 급행이 용특성이 상이하다. 예를 들어 입퇴장역 모두 급행이 정차할 경우 환승 없이 급행을 이용할 수 있는 반면, 입퇴장역 모두 급행이 정차하지않을 경우 일반-급행-일반열차 이용의 과정을 거쳐야 한다. 따라서 <Table 3> 과 같이 입퇴장역의 급행정차여부에 따라 데이터를 4개로 분류하고, 급행이용특성에 맞게 탑승가능열차를 탐색하였다.

<Table 3> Data classification and express train use characteristic by OD type

OD type		Express train use characteristic	conceptual diagram
O station	D station		
Y (ET stop)	Y (ET non-stop)	No transfer required	
Y	N	ET - LT transfer required	
N	Y	LT - ET transfer required	
N	N	ET - LT - ET transfer required	

note : ET=Express Train / LT=Local Train

3. 통행시간기반 혼합분포모형 분석 방법

통행시간기반 혼합분포모형 분석 방법은 한 OD pair를 통행하는 승객의 통행시간분포가 급행 이용/비이용 승객 통행시간이 혼합된 분포인 것으로 가정하고, 혼합확률분포를 산출하여 각 승객이 급행열차를 이용했을 확률을 분석하는 방법이다.

한 OD에서 복수의 대안경로가 존재할 때, 혼합확률분포를 산출하여 각 경로별 이용확률을 분석하는 방법은 선행연구에서도 적용한 방법이다.(Sun and Xu, 2012; Cheng et al., 2019) 선행연구의 경우 통행시간을 통행과정별로(도보시간, 대기시간 등) 나누고 각 과정별 시간분포를 별개의 방법으로 추정한 후, 이를 다시 결합하는 방식으로 경로별 통행시간분포를 산출한다. 반면 본 연구는 전체 통행시간분포로부터 EM알고리즘을 통해 급행 이용/비이용 승객별 통행시간분포를 산출한다는 차이점이 있다. 선행연구는 급행 비운영노선을 대

상으로 한다는 차이점 또한 존재한다.

식(1)은 특정 OD pair를 통행하는 승객의 통행시간에 대한 혼합확률분포모형이다.

$$f^{OD}(x) = \sum_{k=1}^2 \pi_k f_k^{OD}(x|\Theta_k) \dots\dots\dots (1)$$

where,

O : origin station

D : Destination station

k : groups according to the type of train on board(1 = express, 2 = local)

π_k : probability that the k^{th} group is selected

$f_k^{OD}(x|\Theta_k)$: mixture probability density function of travel time at a given OD pair conditional on the group k

Θ_k : parameter space for the k^{th} group

통행시간분포를 혼합확률분포모형으로 나타내기 위해 적합한 분포모형을 설정해야 한다. 후보분포로 웨이블분포, 로그정규분포, 정규분포, 감마분포를 설정하였다. 해당 분포들로부터 혼합확률분포모형을 생성 후, 각 모형의 적합도를 분석하였다. <Table 4>은 후보분포들의 PDF와 모수를 나타낸 것이다.

<Table 4> Distributions and parameter of distributions

Distribution	Weibull	Lognormal	Normal	Gamma
PDF	$\frac{k}{\lambda} \left(\frac{y}{\lambda}\right)^{k-1} e^{-(y/\lambda)^k}$	$\frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln y - \mu)^2}$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$	$\frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$
Parameter	k, λ	μ, σ	μ, σ	α, β

후보분포들의 모수를 추정하기 위해 EM(Expectation-Maximization)알고리즘을 이용하였다. EM 알고리즘은 기댓값 최대화 알고리즘이며, 관측되지 않는 잠재변수에 의존하는 확률모형에서 최대가능도나 최대사후확률을 갖는 모수의 추정값을 찾는 반복적인 과정을 수행하는 알고리즘이다. 즉, 최대가능도 추정을 통해 관측된 데이터에 알맞은 분포의 모수를 추정하는 방법론이다. EM 알고리즘은 E Step인 추정단계, M Step인 최대화 단계로 이루어져 있다. E Step은 추정단계로 각 관측값으로부터 각 확률분포의 잠재변수 발생 확률값을 추정하는 단계이며, M Step은 최대화 단계로 E step에서 추정된 확률값을 이용하여 혼합분포모형 내 각 확률분포의 최대가능도값이 최대가 되는 확률분포의 모수를 추정하는 단계이다. 일정 수렴 기준을 만족할 때까지 알고리즘은 반복 수행된다. 아래의 식(2)는 EM 알고리즘을 식으로 나타낸 것이다.

$$\begin{aligned}
 [Estep] \quad Q(\theta|\theta^{(t)}) &= E_{Z|X,\theta^{(t)}}[\log L(\theta; X, Z)] = \sum_Z p(Z|X, \theta^{(t)}) \log L(\theta; X, Z) \\
 [Mstep] \quad \theta^{(t+1)} &= \underset{\theta}{arg\max} Q(\theta|\theta^{(t)}) \dots\dots\dots (2)
 \end{aligned}$$

where,

X : observed data

Z : latent variable

θ : a vector of unknown parameters

t : number of trials

$L(\theta; X, Z)$: the maximum likelihood estimate of the unknown parameters is determined by maximizing the marginal likelihood of the observed data

$Q(\theta|\theta^{(t)})$: expected value of the log likelihood function of θ

후보분포모형의 모수를 추정이 완료되면 가장 적합한 분포를 선택해야 한다. 적합 분포의 선택은 AIC(Akaike Information Criterion)를 기준으로 하였다. AIC는 Kullback-Leibler(KL) divergence라 부르는 값에 대한 추정치이며, 실제 데이터들로부터 생성된 분포와 통계모형으로부터 생성된 분포의 차이를 수량화한 값이다. AIC가 작을수록 적합한 모형으로 판단할 수 있다. 따라서, OD pair별로 AIC가 가장 작은 분포를 적합분포로 선정하였고, 전체 OD pair에서 적합분포 선정 수가 가장 많은 후보분포를 최적 적합분포로 설정하였다. 선정된 최적 적합분포를 이용하여 통행시간분포를 혼합확률분포모형을 생성하였다.

최종적으로 산출된 혼합확률분포모형을 이용하여 급행열차 이용 승객 통행시간분포와 급행열차 비이용 승객 통행시간분포의 교점을 산출한 뒤, 이를 분류 기준점으로 활용하여 급행열차 이용/비이용 승객을 분류하였다.

4. 통행시간기반 혼합분포모형 분석 결과의 검증

OD pair별 통행시간분포 모형화 결과의 적합도를 직접 분석할 수 있는 관측 데이터가 존재하지 않기 때문에, 세가지 방법으로 통행시간기반 혼합분포모형 분석 결과의 합리성을 검증하였다. 첫 번째로, K-S(Kolmogorov-Sminorov) 검정을 수행하였다. K-S 검정은 주어진 표본분포와 이론적으로 기대되는 분포의 유사도를 검정할 때 사용할 수 있는 방법이다. 표본분포의 누적확률과 이론적으로 기대되는 분포의 누적확률의 최대거리를 검정통계량으로 사용하며, 이를 귀무가설의 채택여부를 결정할 시 이용한다. 식(3)은 표본분포의 누적확률과 이론분포의 누적확률 간의 거리를 수식으로 나타낸 것이다.

$$D = \sup |F_0(x) - F_n(x)| \dots\dots\dots (3)$$

where,

S : maximum distance between the cumulative sample distribution and the hypothesized cumulative distribution

$F_0(x)$: cumulative sample distribution

$F_n(x)$: cumulative hypothesized distribution

두 번째로, 산출된 분류 기준점과 일반열차 탑승 시 최소 통행시간을 비교하였다. 급행열차를 비이용할 경우, 일반열차 최소 통행시간보다 빨리 통행할 수 없다. 따라서 급행열차 이용/비이용 승객 분류 기준점 시간이 해당 OD pair의 일반열차 최소 통행시간보다 짧게 산출되면, 급행열차 비이용이 불가능한 일부 승객을 급행열차 비이용으로 분류하게 되는 것이므로 비합리적인 분석결과로 판단할 수 있다.

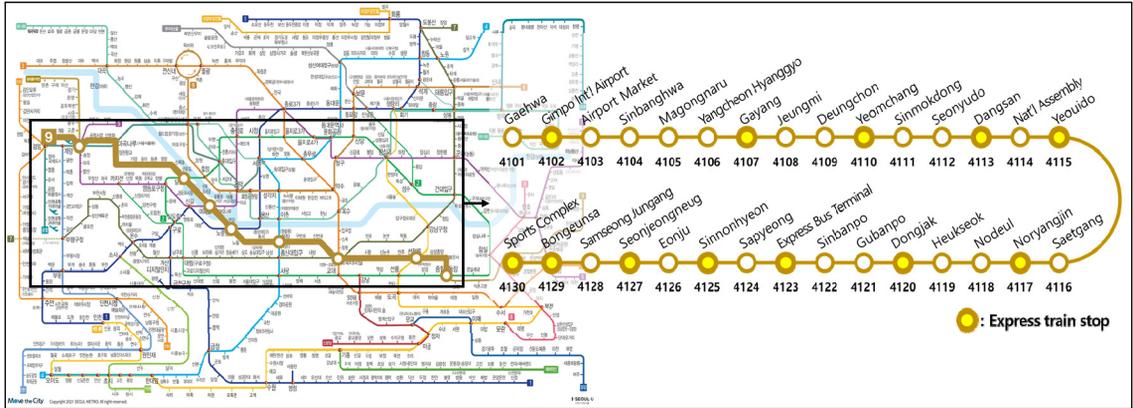
셋째, 교통카드데이터 - 열차시각표 매칭을 통해 탑승열차종을 파악한 데이터를 분류 기준점 기준으로 재분류하여 정확도(Accuracy)를 산출하였다. 이를 통해 통행시간기반 혼합분포모형 분석 결과의 합리성을 평가하였다.

IV. 분석 대상 및 분석 결과

1. 분석과정

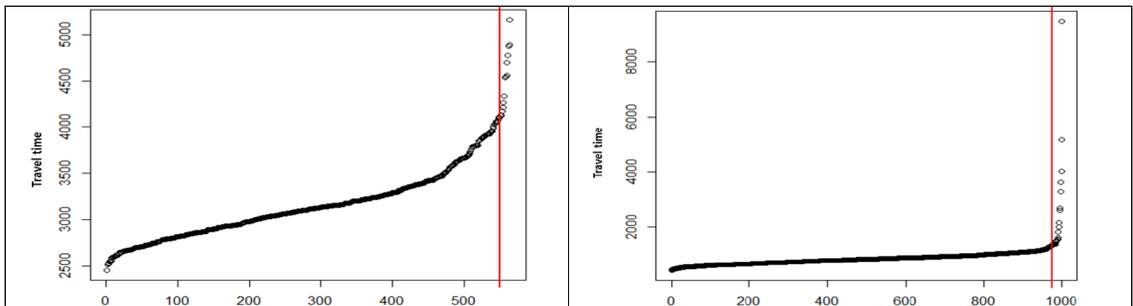
본 연구의 분석대상구간은 서울지하철 9호선 2단계(개화~종합운동장) 구간의 30개역(급행정차역 12개)이다. 방법론을 검증하기 위해 활용한 데이터는 2017년 11월 22일 수요일 9호선 이용 승객의 스마트카드 데이

터이다. 비정상적인 데이터라고 할 수 있는 동일역 승하차 데이터를 제거한 통행수는 461,859통행이다.



<Fig. 3> Seoul Metro Line 9 route map

정확한 분석을 위해서는 또다른 비정상적 데이터인 ‘만짓’을 하는 승객의 통행데이터를 제거해야 한다. 이러한 승객은 통행시간이 지나치게 길어 탑승한 열차를 특정할 수 없기 때문이다. 이를 위해 스마트카드 데이터를 통해 주요 OD pair의 각 승객 통행시간을 산출한 뒤, 오름차순 정리하여 그래프로 나타내보았다. 그 결과, <Fig. 4>와 같이 통행시간 상위 97.5%이상의 데이터(빨간선 오른쪽)는 통행시간이 타 데이터보다 급격히 증가하는 비정상적 데이터로 판단되었다. 이에 본 연구에서는 통행시간 상위 97.5% 이상의 데이터를 제외한 449,836통행을 대상으로 탑승열차를 추정하였다.



<Fig. 4> A case of travel time plot in ascending order (Gimpo Int'l Airport - Gayang(left), Gimpo Int'l Airport - Sports Complex(right))

열차시각표는 스마트카드 데이터에 맞게 2017년 11월 22일의 시각표를 활용하였다. 급행 : 일반 배차비율은 약 1:1로, 배차횟수 등은 <Table 5>와 같다.

<Table 5> Train operation schedule of Seoul Metro Line 9

Train Types	Operation Time	Train Frequency(times)		Train Interval (min/bound)	ET : LT Ratio
		Up Train	Down Train		
Express Train	05:30~25:00	119	123	10min	1 : 1.09
Local Train		132	132	9min	

2. 교통카드데이터 - 열차시각표 매칭 결과

교통카드데이터 - 열차시각표(운행이력)를 매칭한 결과를 정리하면 <Table 6>과 같다. 전체 승객 중 23.5%가 일반열차만 이용하여 통행하며(일반 \geq 1, 급행=0), 48.3%는 급행열차를 이용하여(일반=0, 급행 \geq 1) 통행하는 것으로 나타났다. 단, 급행이용을 위해 1회 이상 환승해야하는 OD pair의 경우 일반열차만 이용한 통행이 33.5~90.1%, 급행열차를 이용한 통행이 4.9~28.9%로 일반열차만 이용한 통행의 비율이 더 높았다. 한편 탑승가능열차를 탐색하지 못한(일반=0, 급행=0) 승객이 0.2% 존재했는데, 이는 교통카드데이터의 오류 또는 보행속도의 2배 이상으로 달려 열차에 탑승한 일부 승객으로부터 기인한 것으로 보인다.

교통카드데이터 - 열차시각표를 매칭한 결과를 입퇴장역간 거리별로 분석하였다. 이 때 역간 거리는 실제 물리적인 거리가 아닌 역간 급행정차역수를 의미한다. 이는 도시철도의 통행시간은 물리적인 거리가 아닌 정차역수에 영향을 많이 받을 뿐만 아니라, 입퇴장역간 급행정차역수에 따라 일부 OD pair의 경우 급행열차 이용이 불가능하거나 비합리적이어서 일반열차만 이용 가능하기 때문이다. 예를 들어 입퇴장역 모두 급행비정차역이면서 입퇴장역간 급행정차역수가 2 미만일 경우 급행 이용이 불가능하다. 또한 급행 이용이 가능한 OD pair이더라도 급행열차가 일반열차를 추월하지 못해 급행열차 이용시 통행시간이 급행열차 비이용시 통행시간보다 길거나 같을 경우 비합리적 환승통행이므로 일반열차만 이용 가능하다.

<Table 6> AFC data - diagram matching result (by OD type)

OD types		Number of passengers	Number of available trains							
			LT \geq 1, ET=0			LT=0, ET \geq 1			LT \geq 1 ET \geq 1	LT=0 ET=0
O	D		LT=1 ET=0	LT \geq 2 ET=0		LT=0 ET=1	LT=0 ET \geq 2			
Y	Y	230,569	3.7%	0.0%	3.8%	61.9%	10.2%	72.1%	24.0%	0.1%
Y	N	95,838	29.6%	3.9%	33.5%	19.7%	2.6%	22.3%	44.1%	0.2%
N	Y	98,612	40.9%	2.1%	43.0%	27.3%	1.5%	28.9%	27.6%	0.5%
N	N	24,817	84.6%	5.5%	90.1%	4.5%	0.4%	4.9%	4.9%	0.2%
Total		449,836	21.8%	1.6%	23.5%	42.2%	6.1%	48.3%	28.0%	0.2%

<Table 7>은 입퇴장역간 거리별 분석결과이다. 입퇴장역간 거리가 증가할수록 일반열차 대비 급행열차의 시간 경쟁력이 증가하기 때문에, 급행열차 비이용 통행의 비율은 감소하는 반면(60.9→0.0%)급행열차 이용 통행의 비율은 증가(21.0→96.2%)하는 것을 알 수 있다.

교통카드데이터 - 열차시각표 매칭 결과에서 중요한 시사점은, 약 28%의 승객은 탑승가능열차종이 복수(일반 \geq 1, 급행 \geq 1)라는 것, 즉 급행열차 이용/비이용 통행이 모두 가능해 탑승열차종을 알 수 없다는 것이다. 입퇴장역간 거리별로 보면, 거리가 멀 경우 탑승가능열차종이 복수인 통행의 비율이 다소 낮으나(입퇴장역간 급행정차역수 10일 때 3.0%), 입퇴장역간 거리가 가까운 경우 탑승가능열차종이 복수인 통행의 비율이 높다.(입퇴장역간 급행정차역수 1일 때 42.4%) 단, 입퇴장역간 급행정차역수가 0일 때는 1일 때에 비해 비율이 낮았는데, 입퇴장역 중 하나 이상의 역이 급행비정차역이면서 입퇴장역간 급행정차역수가 0인 OD pair의 승객은 급행을 이용할 수 없어 일반열차만 이용가능하기 때문이다.

교통카드데이터 - 열차시각표 매칭 방법의 문제를 보완하기 위해, 본 연구에서는 OD pair별 통행시간분포 분석을 통해 급행열차 이용/비이용 승객 분류 기준점을 산출하여 탑승열차종을 알 수 없는 28% 승객의 탑승열차종을 추정하고자 한다.

<Table 7> AFC data - diagram matching result (by OD distance)

Distance	Number of passengers	Number of available Train							
		LT ≥ 1, ET = 0			LT = 0, ET ≥ 1			LT ≥ 1 ET ≥ 1	LT = 0 ET = 0
		LT = 1 ET = 0	LT ≥ 2 ET = 0		LT = 0 ET = 1	LT = 0 ET ≥ 2			
0	107,012	55.3%	5.6%	60.9%	19.8%	1.3%	21.0%	17.8%	0.2%
1	96,929	24.4%	0.5%	24.9%	29.7%	2.8%	32.5%	42.4%	0.2%
2	80,382	11.8%	0.4%	12.2%	45.2%	6.1%	51.3%	36.3%	0.2%
3	53,317	6.7%	0.4%	7.1%	53.5%	8.2%	61.6%	31.1%	0.2%
4	37,521	4.9%	0.3%	5.2%	61.2%	9.7%	70.8%	23.8%	0.2%
5	28,153	1.5%	0.1%	1.7%	65.5%	11.2%	76.7%	21.4%	0.2%
6	21,214	0.5%	0.0%	0.5%	71.3%	15.6%	86.9%	12.3%	0.3%
7	11,537	0.2%	0.0%	0.2%	72.4%	14.4%	86.8%	12.7%	0.3%
8	6,591	0.1%	0.0%	0.1%	74.8%	15.8%	90.6%	8.9%	0.3%
9	3,966	0.0%	0.0%	0.0%	71.0%	19.3%	90.3%	9.2%	0.5%
10	2,772	0.0%	0.0%	0.0%	75.0%	21.4%	96.5%	3.0%	0.6%
11	442	0.0%	0.0%	0.0%	75.6%	20.6%	96.2%	3.6%	0.2%
합계	449,836	21.8%	1.6%	23.5%	42.2%	6.1%	48.3%	28.0%	0.2%

note : Distance means the number of ET stop station between O-D

3. 통행시간기반 혼합분포모형 분석 결과

통행시간기반 혼합분포모형 분석대상 OD pair를 분류하였다. 우선 급행 이용이 불가능하거나 급행 이용이 비합리적인 OD pair를 분석대상에서 제외하였다. 또한 통행수가 30 이하인 OD pair를 분석대상에서 제외하였다. 표본이 30개 이하일 경우 EM 알고리즘을 적용할 수 없기 때문이다. 이에 총 870개의 OD pair중 445개가 분석대상으로 선정되었다(Table 8).

<Table 8> Summary of subject OD pairs of mixture distribution analysis

Distance	OD types				Distance	Express train stop(O-D)			
	Y-Y	Y-N	N-Y	N-N		Y-Y	Y-N	N-Y	N-N
0	21(22)	0(35)	0(35)	0(20)	6	10(10)	13(16)	13(16)	7(30)
1	20(20)	3(31)	4(31)	0(58)	7	8(8)	12(13)	12(13)	6(20)
2	17(18)	17(28)	17(28)	0(46)	8	6(6)	11(11)	11(11)	7(14)
3	16(16)	24(25)	23(25)	1(38)	9	4(4)	7(8)	8(8)	5(10)
4	14(14)	22(23)	21(23)	7(34)	10	2(2)	5(5)	4(5)	1(2)
5	12(12)	19(20)	19(20)	14(34)	11	-	1(1)	1(1)	-
Total - 445(870)						130(132)	134(216)	133(216)	48(306)

Note : Total number of OD pairs in ()

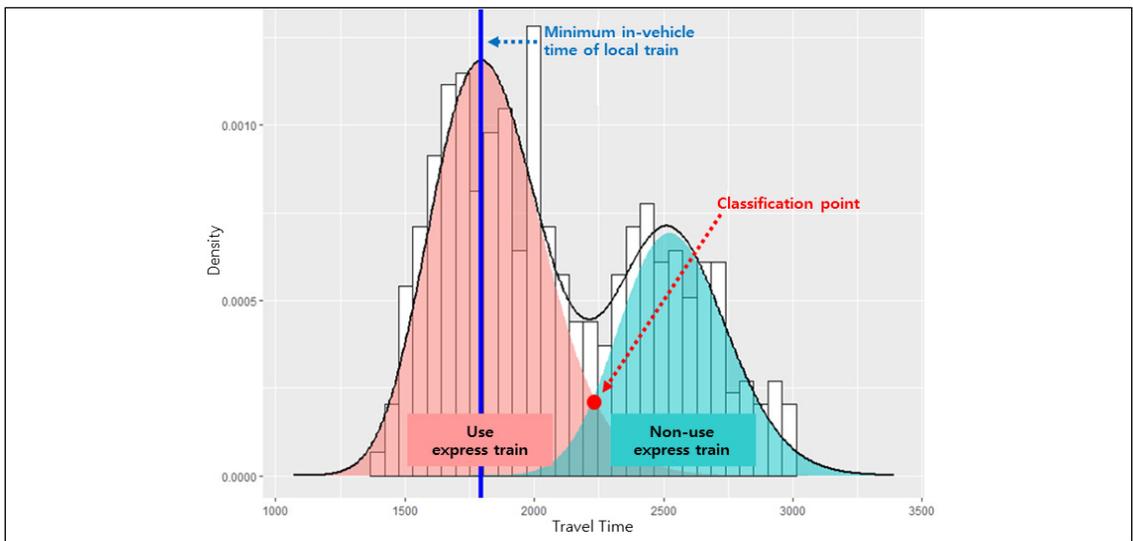
후보분포인 웨이블분포, 로그정규분포, 정규분포, 감마분포로부터 생성된 OD pair별 통행시간분포의 모수

를 EM알고리즘을 통해 추정하고, 적합분포를 선정하기 위해 AIC값을 산출하였다. EM 알고리즘 수행 시 통계 R 프로그램의 “mixR” package에 내장된 함수를 이용하였다. 그리고 OD pair별로 AIC값이 가장 낮은 분포모형을 적합모형으로 선정하였다. <Table 9>는 적합모형 선정 개수를 후보분포모형별로 정리한 것이다. 그 결과 로그정규분포가 전체 445개 OD pair 중 347개(78%)에서 가장 적합한 분포인 것으로 나타나, 통행시간 분포를 가장 적절하게 모형화할 수 있는 최적 적합분포로 결정하였다.

<Table 9> The number of best-fit OD pairs obtained form the distribution model

Distribution	Weibull	Lognormal	Normal	Gamma	Total
No. of best-fit OD	0(0%)	347(78%)	87(20%)	11(2%)	445(100%)

이에 각 분석대상OD pair의 승객 통행시간분포를 혼합로그정규분포로 모형화하고, 급행열차 이용 승객 분포와 급행열차 비이용 승객 분포의 교점을 산출하였다. <Fig. 5>는 언주역(O)-당산역(D)을 통행하는 승객들의 통행시간을 분포로 나타낸 것이다. 검은 선이 생성된 혼합로그정규분포이며, 급행열차 이용 승객들의 통행시간분포(빨간선)와 급행열차 비이용 승객들의 통행시간분포(초록선) 및 그 교점(빨간점)으로 구성되어 있다. 본 연구에서는 이 교점을 급행열차 이용/비이용 승객 분류 기준점으로 정의한다.



<Fig. 5> Mixture distribution analysis concept(a case of Eonju - Dongsan)

4. 통행시간기반 혼합분포모형 분석 검증

혼합로그정규분포로 모형화한 통행시간기반 혼합분포모형 분석 결과의 합리성을 검증하기 위해 우선 K-S(Kolmogorov-Sminorov) 검정을 수행하였다. 유의수준 $\alpha = 0.01$ 로 설정하여 분석을 수행한 결과 전체 445개의 OD pair 중 2개의 OD가 적합하지 않은 것으로 나타났다.

또한 OD pair별 일반열차 최소 차내시간과 분류 기준점을 비교하였다. 55개 OD pair의 분류 기준점이 해당 OD pair의 일반열차 최소 차내시간보다 짧은 것으로 나타나, 부적합한 것으로 분석되었다.

마지막으로, 교통카드데이터 - 열차시각표 매칭을 통해 탑승열차종을 파악한 데이터를 분류 기준점 기준으로 재분류하여 정확도(Accuracy)를 산출하였다. 이 때 분석대상OD pair 중 K-S검정 및 분류 기준점이 부적합한 57개 OD pair의 데이터를 제외하고, 388개 OD pair에서 탑승열차종이 파악된 181,219통행 데이터를 활용하였다. 그 결과 전체 정확도는 81.4%, 정확도가 80% 이상인 검증통과 OD pair는 298개로 산출되었다(Table 10).

<Table 10> Result of mixture distribution analysis

Category	Total OD pairs	Non-subject OD pairs from mixture distribution analysis		Subject OD pairs from mixture distribution analysis			
		Under 30 passengers	Only LT available	K-S test unfit	Shorter classification point than in-vehicle time of LT	Under Accuracy 80%	Over Accuracy 80%
Num. of OD pairs	870	96	329	2	55	90	298
Num. of passengers	449,836	1,518	129,681	15,514	36,517	61,114	205,492

5. 통행시간기반 혼합분포모형 분석 결과를 통한 2차 추정

예시 데이터인 선정릉-고속터미널은 교통카드데이터 - 열차시각표 매칭 결과 59.5% 승객의 탑승가능열차종이 복수인 OD pair이다. 통행시간분포 분석 검증 결과 K-S검정과 분류기준점 적합성 검증을 만족하고 정확도 90.1%로 분석되었다(Table 11).

<Table 11> Boarding train estimation result (before reclassification)

O	D	OD type	Distance	Number of passengers	AFC Data - diagram matching result				Mixture distribution analysis result	
					LT≥1, ET=0	LT=0, ET≥1	LT≥1, ET≥1	LT=0, ET=0	classification point	Accuracy
Seonjeon-neung	Exp. Bus Terminal	Y-Y	1	1,465	0.5%	40.1%	59.5%	0.0%	955.5sec	90.1%

교통카드데이터 - 열차시각표 매칭 결과 탑승가능열차종이 복수인 승객을 분류기준점을 기준으로 재분류하였다. 그 결과 871통행 중 급행이용통행이 318통행, 급행비이용통행이 553통행으로 재분류되어, 최종적으로 선정릉-고속터미널 승객 중 61.8%가 급행을 이용하여 통행한 것으로, 38.2%가 급행을 이용하지 않고 일반열차만 탑승하여 통행한 것으로 추정되었다(Table 12).

<Table 12> A case of boarding train estimation result (after reclassification)

Reclassification	Total trips	Boarding train type		
		ET use	ET non-use	unknown
Before Reclassification	1,465(100.0%)	587(40.1%)	7(0.5%)	871(59.5%)
After Reclassification		905(61.8%)	560(38.2%)	-

6. 결과 분석

통행시간기반 혼합분포모형 분석 검증 결과 본 연구의 기준을 만족하는 OD pair는 298개로 분석대상OD pair 445개 중 약 2/3수준이다. 검증을 만족하지 못하는 원인은 3가지로 추정된다. 첫째, 표본수가 부족하여 실제 통행분포를 충분히 반영하지 못하는 경우이다. 표본수가 부족할 경우 분포는 실제 형태와 왜곡되어 나타날 가능성이 높다. 둘째, OD간 거리가 가까워 실제 급행이용승객 통행시간분포와 급행비이용승객 통행시간분포가 대부분 중첩되어 나타나 혼합확률분포로 모형화하기 어려운 경우이다. 셋째, 반대로 OD간 거리가 먼 것으로 인해 급행비이용승객이 매우 적어, 승객 통행시간분포가 급행이용승객 통행시간분포와 급행비이용승객 통행시간분포의 혼합 분포라는 가정에 기반한 혼합확률분포 모형화가 합리적으로 산출되지 않는 경우이다.

<Table 13>은 봉은사역을 중심으로 OD pair의 거리별 통행시간분포 분석 결과를 비교한 것으로, OD간 거리가 가깝거나 멀 때 비합리적인 분석 결과가 도출되는 것을 보여준다. 입퇴장역간 급행정차역수가 4일 경우 분석결과가 합리적인 반면, 0 또는 9인 경우 비합리적인 분석결과가 산출되었다.

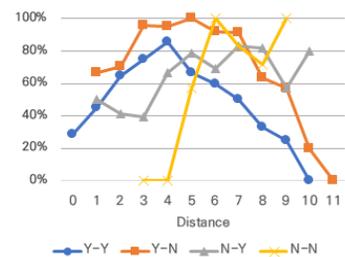
<Table 13> Comparison of mixture distribution analysis results by OD distance (the case of Bongeunsa station)

OD pair	Bongeunsa - Seonjeongneung	Bongeunsa - Noryangjin	Bongeunsa - Gimpo Int'l Airport
Distance	0	4	9
Accuracy	61.2%	88.2%	Shorter classification point than in-vehicle time of LT
Graph			

<Table 14> OD pairs over 80% accuracy

Dist.	OD types				Dist.	OD types			
	Y-Y	Y-N	N-Y	N-N		Y-Y	Y-N	N-Y	N-N
0	6(21)	-	-	-	6	6(10)	12(13)	9(13)	7(7)
1	9(20)	2(3)	2(4)	-	7	4(8)	11(12)	10(12)	5(6)
2	11(17)	12(17)	7(17)	-	8	2(6)	7(11)	9(11)	5(7)
3	12(16)	23(24)	9(23)	0(1)	9	1(4)	4(7)	4(7)	5(5)
4	12(14)	21(22)	14(21)	0(7)	10	0(2)	1(5)	4(5)	1(1)
5	8(12)	19(19)	15(19)	8(14)	11	-	0(1)	1(1)	-
Total - 298(445)					71(130)	112(134)	84(133)	31(48)	

Note : Total number of OD pairs for mixture distribution analysis in ()



<Fig. 6> Rate of OD pairs over 80% accuracy

OD거리에 따른 통행시간기반 혼합분포모형 분석 결과의 합리성은 입퇴장역 급행정차여부에 따라 상이하였다(<Table 14> 및 <fig. 6>). 입퇴장역 급행정차여부 [Y-Y]의 경우 입퇴장역간 급행정차역수가 4일 때 검증 통과OD pair 비율이 가장 높은 반면(분석대상OD pair 14개 중 11개, 85.7%), 1이하 및 7이상일 때는 50% 이하였다. [Y-N]는 입퇴장역간 급행정차역수 5일 때(분석대상OD 19개 중 19개, 100%) 검증통과OD pair수 비율

이 가장 높고 10이상일 때 50%를 하회하였다. [N-Y]는 3이하, [N-N]은 4이하일 때 검증통과OD pair 비율이 50% 이하였다. 입퇴장역 급행정차여부별로 결과가 상이한 것은 환승횟수에 영향을 받기 때문인 것으로 판단된다.

결론적으로, 본 연구의 추정 방법은 입퇴장역간 급행정차역수가 약 2~9일 때 검증통과OD pair 비율이 50%를 상회하며 합리적인 탑승열차종 추정 결과를 산출할 수 있는 것으로 분석되었다.

V. 결 론

본 연구에서는 도시철도 승객의 탑승열차종을 추정하기 위해 기존 연구에서 사용되었던 교통카드데이터 - 열차시각표 매칭 방법의 한계를 보완하고자 하였다. 이에 매칭 방법과 통행시간기반 혼합분포모형 분석 방법을 순차적으로 적용하여 급행운영 도시철도 승객의 탑승열차종을 추정하고 분석 결과를 검증하였다. 교통카드데이터 - 열차시각표 매칭 결과에서는 28% 승객의 탑승열차종을 추정할 수 없음을 알 수 있었다. 통행시간기반 혼합분포모형 분석 결과 승객 통행시간분포가 혼합로그정규분포의 형태임을 밝혔다. 그리고 298개 OD pair에서 정확도 80% 이상의 합리적인 급행이용/비이용 승객 분류기준점을 도출했다. 이때 이 검증통과OD pair의 비율은 OD거리별, 입퇴장역 급행정차여부별로 상이하며, 입퇴장역간 급행정차역수가 2~9개 거리일 때 검증통과OD pair의 비율이 높아 해당 구간에서 본 연구의 탑승열차종 추정 방법이 합리적으로 적용될 수 있음을 알 수 있었다.

기존의 탑승열차종 추정은 분석자 임의의 가정을 덧붙이거나 분석 방법이 지나치게 복잡한 반면, 본 연구의 방법은 비교적 간단하고 임의의 가정 없이 급행운영 도시철도 승객의 탑승열차종을 추정할 수 있다는 장점이 있다.

그러나 몇가지 한계점을 지적하지 않을 수 없다. 첫째, 본 연구는 OD pair별 전체 통행시간분포로부터 급행 이용/비이용 승객 통행시간분포를 추정하는 과정에서, 타 제약조건 없이 EM알고리즘을 통해 혼합분포모형을 추정하는 것만을 고려하였다. 이는 검증통과 OD pair수 감소에 영향을 미쳤다. 급행 비이용 승객 통행시간분포의 경우 해당 OD의 일반열차 최소차내시간보다 우측에 추정되도록 보완이 필요하다. 이를 통해 검증통과 OD pair수 증가 및 전반적인 정확도 향상, 특히 OD거리가 일정 이상인 pair의 정확도 향상을 기대해볼 수 있을 것으로 생각된다.

둘째, OD거리가 일정 이하인 OD pair는 급행 이용/비이용 승객의 통행시간 차이가 크지 않아 통행시간을 기준으로 탑승열차종을 분류할 경우 정확도가 상대적으로 낮은 것으로 나타났다. 이러한 OD pair의 경우 추정 정확도 향상을 위해서는 선행연구 등을 참고한 보완적인 추정 방법론의 적용이 필요할 것으로 판단된다.

셋째, 신뢰할 수 있는 자료와의 비교를 통한 검증이 부족하였다. 그러나 이는 탑승열차 추정과 관련된 모든 연구의 근본적인 한계이며, 향후 통신데이터 등 승객의 실제 탑승열차를 확인할 수 있는 신뢰성 높은 데이터가 축적될 경우 검증이 가능할 것으로 판단된다.

ACKNOWLEDGEMENTS

본 논문은 2021년 한국철도학회 춘계학술대회에서 발표되었던 논문을 수정·보완하여 작성하였습니다.

REFERENCES

- Baek J. H. and Sohn K. M.(2016), “An investigation into passenger preference for express trains during peak hours,” *Transportation*, vol. 43, pp.623-641.
- Cheng G., Zhao S. and Xu S.(2019), “Estimation of passenger route choices for urban rail transit system based on automatic fare collection mined data,” *Transactions of the Institute of Measurement and Control*, vol. 41, no. 11, pp.3092-3102.
- Hong L., Li W. and Zhu W.(2017), “Assigning Passenger Flows on a Metro Network Based on Automatic Fare Collection Data and Timetable,” *Discrete Dynamics in Nature and Society*, vol. 2017, pp.1-10.
- Hong S. P., Min Y. H., Park M. J. and Kim K. M. et al.(2016), “Precise estimation of connections of metro passengers from Smart Card data,” *Transportation*, vol. 43, pp.749-769.
- Katori T., Takahashi Y. and Izumi T.(2004), “Determination of stations where rapid trains stop or pass to local ones using a genetic algorithm to shorten total trip time,” *Computers in Railways*, vol. 9, pp.113-121.
- Kim K. M., Hong S. P., Ko S. J. and Kim D. W.(2015), “Does crowding affect the path choice of metro passengers?,” *Transportation Research Part A*, vol. 77, pp.292-304.
- Kim K. M., Oh S. M. and Rho H. L.(2016), “Express Train Choice and Load Factor Analysis as Line Extension in Seoul Metro 9,” *Journal of the Korean Society for Railway*, vol. 19, no. 5, pp.663-671.
- Kim K. M., Oh S. M., Hong S. P. and Ko S. J.(2015), “Tracing a Logical Path of Passengers: A Case study of Seoul Metro Line 9,” *Journal of the Korean Society for Railway*, vol. 18, no. 6, pp.586-595.
- Kim M. S., Kim J. T., Kim T. S. and Park S. S. et al.(2013), “Study of the Metropolitan Rapid Transport System to Minimize Sidetrack Construction,” *Journal of the Korean Society for Railway*, vol. 16, no. 5, pp.402-409.
- Kusakabe T., Iryo T. and Asakura Y.(2010), “Estimation method for railway passengers’ train choice behavior with smart card transaction data,” *Transportation*, vol. 37, pp.731-749.
- Lee H. J., Zhang D., He T. and Son S. H.(2017), “Metro Time: Travel Time Decomposition under Stochastic Time Table for Metro Networks,” *2017 IEEE International Conference on Smart Computing*, Hong Kong, pp.1-8.
- Lee K. H., Lee T. G., Kim H. G. and Lee H. Y.(2018), “A Study on Express Train Operation for Reducing Travel Time of Urban Railway,” *Journal of the Korean Society for Urban Railway*, vol. 6, no. 2, pp.103-110.
- Luo Q., Hou Y., Li W. and Zhang X.(2012), “Stop Plan of Express and Local Train for Regional Rail Transit Line,” *Journal of Advanced Transportation*, vol. 2018, 3179321.
- Othman N. B., Legara E. F., Selvam V. and Monterola C.(2015), “A data-driven agent-based model of congestion and scaling dynamics of rapid transit systems,” *Journal of Computational Science*, vol. 10, pp.338-350.
- Sohn K. M.(2007), “Mixed Operation with Express Train for Urban Railways in Seoul Metropolitan

- Area,” *Journal of Korean Society of Transportation*, vol. 25, no. 5, pp.195-207.
- Soo P. J., Hee L. H. and Mu W. J.(2006), “A Development of Optimum Operation Models for Express-Rail Systems,” *Journal of the Korean Society for Civil Engineers D*, vol. 26, no. 4D, pp.679-686.
- Sun Y. and Schonfeld P. M.(2016), “Schedule-Based Rail Transit Path-Choice Estimation using Automatic Fare Collection Data,” *Journal of Transportation Engineering*, vol. 142, no. 1.
- Sun Y. and Xu R.(2012), “Rail Transit Travel Time Reliability and Estimation of Passenger Route Choice Behavior,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2275, pp.58-67.
- Wu J., Qu Y., Sun H. and Yin H.(2019), “Data-driven model for passenger route choice in urban metro network,” *Physica A*, vol. 524, pp.787-798.
- Zhou F., Shi J. and Xu R.(2015), “Estimation Method of Path-Selecting Proportion for Urban Rail Transit Based on AFC Data,” *Mathematical Problems in Engineering*, vol. 2015, pp.1-9.
- Zhu W. and Xu R.(2016), “Generating route choice sets with operation information on metro networks,” *Journal of Traffic and Transportation Engineering(English Edition)*, vol. 3, no. 3, pp.243-252.
- Zhu W., Wang W. and Huang Z.(2017), “Estimating train choices of rail transit passengers with real timetable and automatic fare collection data,” *Journal of Advanced Transportation*, vol. 2017, 5824051.