

심층신경망 모형을 활용한 대중교통 이용자의 환승시간 추정에 관한 연구

A study on Estimating the Transfer Time of Transit Users Using Deep Neural Network Models

이 경 재* · 김 수 재** · 문 형 택** · 한 재 윤*** · 추 상 호****

* 주저자 : 홍익대학교 도시계획과 석사과정
 ** 공저자 : 홍익대학교 도시계획과 박사과정
 *** 공저자 : 홍익대학교 도시계획과 연구원
 **** 교신저자 : 홍익대학교 도시공학과 교수

Gyeongjae Lee* · Sujae Kim** · Hyungtaek Moon** · Jaeyoon Han***
 · Sangho Choo****

* Dept. of Urban Planning, Hongik University
 ** Dept. of Urban Planning, Hongik University
 *** Dept. of Urban Planning, Hongik University
 **** Dept. of Urban Design & Planning, Hongik University
 † Corresponding author: Sangho Choo, shchoo@hongik.ac.kr

Vol.19 No.1(2020)
 February, 2020
 pp.32~43

pISSN 1738-0774
 eISSN 2384-1729
<https://doi.org/10.12815/kits.2020.19.1.32>

Received 10 February 2020
 Revised 18 February 2020
 Accepted 20 February 2020

© 2020. The Korea Institute of
 Intelligent Transport Systems. All
 rights reserved.

요 약

환승시간은 대중교통계획 및 정책 수립에 있어서 중요한 요소이다. 이에 본 연구에서는 교통카드 이용자료를 활용하여 대중교통 이용자의 환승시간 영향요인을 규명하고, 딥러닝 기법인 심층신경망 모형을 이용한 환승시간을 추정하였으며 이를 전통적인 회귀모형과 비교 분석하였다. 먼저 환승시간 영향요인의 경우, 주변 버스의 배차간격과 버스 정류장까지의 거리가 버스 환승시간에 양의 영향을 주었으며, 버스 노선수는 반대로 음의 영향을 주었다. 또한 지하철역이 속해있는 자치구에 따라서도 환승시간에 영향을 주는 것으로 나타났다. 도출된 환승시간 영향요인을 통해 딥러닝 모형을 구축하고 성능을 비교한 결과, 회귀모형보다 딥러닝 모형의 성능이 보다 우수하였다. 본 연구의 결과는 지역별 환승허용시간의 차등 적용 등 대중교통 환승정책의 기초 자료로 활용될 수 있을 것으로 판단된다.

핵심어 : 환승시간, 딥러닝, 심층신경망, 회귀모형, 교통카드데이터

ABSTRACT

The transfer time is an important factor in establishing public transportation planning and policy. Therefore, in this study, the influencing factors of the transfer time for transit users were identified using smart card data, and the estimation results for the transfer time using the deep learning method such as deep neural network models were compared with traditional regression models. First, the intervals and the distance to the bus stop had positive effects on the subway-to-bus transfer time, and the number of bus routes had a negative effect. This also showed that the transfer time is affected by the area in which the subway station exists. Based on the influencing factors of the transfer time, the deep learning models were developed and their estimation results were compared with the regression model. For model performance, the deep learning models were better than those of the

regression models. These results can be used as basic data for transfer policies such as the differential application of transit allowance times according to region.

Key words : Transfer Time, Deep Learning, Deep Neural Network, Regression, Smart Card Data

I. 서론

도시지역의 대기오염은 지속가능한 도시성장과 삶의 질을 저해하는 문제로 인식되고 있으며, 주로 CO₂, NO_x 등의 대기오염 물질을 배출하는 자동차의 수요를 저감하는 방안이 지속적인 교통정책의 연구대상이 되어 왔다. 국내에서도 녹색교통을 실현하기 위한 대중교통 이용활성화 정책들이 도입되고 있으며, 대표적인 사례로는 수도권 대중교통 통합요금제, 중앙버스전용차로제 등이 있다. 특히 대중교통 통합요금제는 대중교통 이용자가 여러 교통수단을 환승하여 통행할 때에도 하나의 통행수단을 이용한 것처럼 요금을 부과하는 것으로, 이용한 총거리에 따라 요금을 부과하는 요금체계이다(Seoul Solution, 2017). 이러한 통합요금제가 적용되는 수도권에서는 대중교통 간의 환승이 크게 증가하였다.

그럼에도 불구하고, 대중교통 간의 환승은 환승거리 과다 및 환승수단 간의 배차 간격의 불연속성 등으로 인해 환승통행 이용 시 불편함을 야기하고 있다. 특히 환승통행에 대한 특성을 이해하고 환승시간에 영향을 미치는 요인을 규명하는 것은 대중교통계획 및 정책 수립에 있어서 매우 중요한 요소이다.

수도권의 경우 대중교통 이용 시 대부분이 AFC(Automated Fare Collection) 수단인 교통카드를 사용하고 있으며, 평일 하루 수집되는 이용건수는 2018년 기준 약 2천만 건에 달한다(Ministry of Land, Infrastructure and Transport, 2018). 이를 통해 산출할 수 있는 환승통행량, 환승거리, 환승시간 등 환승통행 특성을 분석하여 대중교통 편의성을 개선하고자하는 연구가 지속적으로 수행되고 있다.

특히 환승시간 추정은 대중교통 환승체계의 개선을 위해 수행되어야 하는 연구이며, 기존에는 주로 전통적인 모델링 방법인 선형회귀분석과 로짓모형이 활용되었다. 그러나 복잡한 요인들이 얽혀있는 도시의 교통 현상은 기존의 회귀분석이 설명하기 어려운 비선형적인 특성을 가지는 부분이 있어 최근에는 이런 비선형적인 특성을 반영하여 보다 높은 정확도로 추정이 가능한 딥러닝 기법을 활용한 연구가 증가하고 있다. 이에 본 연구에서는 대중교통의 환승체계 개선을 위해 서울시 내 환승시간을 새로운 연구방법론으로 주목받고 있는 딥러닝 기법으로 추정하고자 하며, 기존의 전통적인 방법론과 추정 결과를 비교 분석하고자 한다.

II. 선행연구검토

본 연구에서는 환승시간 추정을 위해 전통적인 통계모형과 달리 새로운 연구방법론인 딥러닝 기법을 활용하였다. 이에 환승통행 특성과 관련된 연구와 함께 교통 분야에서 딥러닝 기법을 활용한 연구에 대해 살펴 보았다.

1. 환승통행 특성 관련 연구

환승통행 특성과 관련된 선행연구들은 환승유형, 환승수단선택 등의 환승통행 실태를 분석한 연구와 환승 영향요인을 규명하는 연구로 구분할 수 있다.

Jang(2010)은 서울시 교통카드 이용자료를 활용하여 대중교통 통행시간과 환승패턴에 대해 분석하였다. 이용한 수단에 따라 환승유형을 구분한 결과, 환승통행 중에서는 버스-버스 환승이 전체 통행 중 10.15%로 가장 많았으며, 버스-지하철 8.10%, 지하철-버스 7.95%를 차지하였다. 또한 지하철역과 버스 정류장을 대상으로 5,000통행 이상의 환승이 이뤄지는 지점, 평균 환승시간이 10분 이상인 지점, 전체 통행 중 환승통행 비율이 50% 이상인 지점을 분석한 뒤 각 결과를 중첩하여 환승시설 개선이 필요한 지점을 도출하였다.

Lee et al.(2013)는 다항로짓모형을 이용하여 대구시의 환승교통수단선택모형을 구축하였다. 환승 유형은 버스-버스, 버스-지하철, 지하철-지하철, 버스-기타교통수단, 지하철-기타교통수단으로 구분하였으며, 독립변수로는 성별, 나이, 교통주체, 통행목적, 환승지역, 통행비율, 통행시간 변수를 사용하였다. 통행비용은 모든 환승유형에서 음의 영향을 주었으며, 통행시간은 지하철-지하철을 제외한 유형에서 양의 영향을 주는 것으로 나타났다. 또한 환승유형선택에 대한 한계효과 분석을 수행하였는데, 총 통행시간이 증가할수록 버스-버스, 버스-지하철 환승 이용객들이 지하철-지하철 유형을 선택하는 것으로 분석되었다.

Hwang(2014)은 대구시를 대상으로 다중회귀모형을 통해 버스와 지하철간 환승통행량에 영향을 미치는 요인을 분석하였다. 독립변수로는 지하철역별 연계버스 노선수, 환승시간, 지하철역의 공간적 위치, 지하철역과 연계되는 버스 노선의 특성을 반영한 버스노선의 연계지수를 사용하였다. 분석 결과, 버스노선의 연계지수와 지하철역 연계버스 노선수는 환승통행량에 양의 영향을 주었으나 환승시간은 음의 영향을 주었다. 또한 지하철역이 환승수요가 더 많은 것으로 나타났다.

Lim et al.(2016)은 서울시를 대상으로 환승통행의 공간적 분포 등의 환승통행행태를 분석하고, 행정동 단위의 환승통행량과 토지이용의 관계를 규명하기 위해 다중회귀분석을 수행하였다. 분석 결과, 업무시설면적 비율 및 백화점시설의 면적비가 높거나, 복합토지이용도가 높은 지역일수록 환승통행량이 많아지는 것으로 나타났다. 반면 인구밀도는 환승통행량에 음의 영향을 주는 것으로 나타났다.

Lim(2017)은 서울시 2013년 교통카드 이용자료를 활용하여 환승통행행태를 분석하였으며, 시간대별 환승시간에 영향을 미치는 요인에 대하여 규명하기 위해 다중회귀분석을 수행하였다. 환승시간에는 지하철역과 버스 정류장사이의 거리, 버스 배차간격의 변동성(CV), 상업지역 면적, 도소매업 사업체수, 판매시설 연면적이 양의 영향력을 미치고, 버스 노선수는 음의 영향력을 미치는 것으로 나타났다.

2. 딥러닝 기법 관련 연구

딥러닝의 대표적인 방법론 중 하나인 인공신경망은 인간의 뇌의 정보처리 및 전달과정을 수학적 측면에서 관찰하여 시뮬레이션한 것으로 1943년 Warren McCulloch와 Walter Pitts에 의해 처음 제안되었다. 하지만 제안된 인공신경망이 단순한 XOR 분류문제도 해결하지 못하는 것이 수학적으로 증명됨에 따라 인공신경망 연구는 침체기를 맞는다. 이후 1980~1990년도에 다층 인공신경망을 통해 XOR 분류 문제를 해결하면서 활발한 연구가 진행되었으나, 당시에는 음성, 언어, 영상과 같은 신호에 대한 식별의 어려움과 당시 학습 알고리즘은 인공신경망의 학습시간에 현실적이지 않다는 한계가 있었다(Deng and Yu, 2014). 그러나 최근에는 데이터 저장기술과 컴퓨터 계산능력의 발달로 이론적으로만 제시되었던 개념이 실제적으로 적용되는 계기가 마련되었고, 2010년에는 합성곱신경망(Convolutional Neural Network)을 활용한 이미지 인식 기술이 높은 정확도를 보여 주목을 받아 세계적으로 연구가 확산되었다.

초기 연구에서는 딥러닝이 복잡한 패턴을 인식하는데 초점이 맞춰져 있었기 때문에 이미지 인식, 음성 인식, 자연어 처리 등의 분야에 연구가 집중되어있었다. 그러나 최근에는 R, Python 등 오픈소스 프로그래밍 언어의 대중화와, 다양한 빅데이터 수집됨에 따라 많은 분야에서 딥러닝을 활용한 연구가 진행되었고, 도시 및

교통계획 분야에서도 다양한 연구가 시도되고 있다.

교통 분야에서는 차량 GPS 데이터, 교통검지기 자료를 활용하여 교통흐름의 예측(Haung et al., 2013; LV et al., 2015; Polson and Sokolov, 2017), 통행시간의 예측(Xiong et al., 2015), 통행속도의 예측(Ma et al., 2015(a)), 교통지체의 예측(Ma et al., 2015(b)), 교통정보 결측값의 대체(Duan et al., 2016), 교통사고건수 예측(Ryu et al., 2018) 등의 연구가 수행되었다. 교통카드 이용자료를 활용한 연구에서는 정류장 단위의 대중교통 수요추정(Baek and Sohn, 2017), 환승경로의 추정(Lee and Shon, 2015), 대중교통 하차지점의 추정(Jung and Sohn, 2017), 통행목적의 추정(Han and Sohn, 2016) 등의 연구가 수행되었다.

Baek and Sohn(2017)은 인공신경망 기법을 활용하여 정류장별, 기종점별 대중교통 이용량을 추정하였다. 대중교통 이용량 추정을 위해 정류장 인근의 건축물 연면적, 용도 등의 활동 관련 변수, 배차간격, 버스 노선 수, 지하철 노선수, 통행시간 등의 수단 관련 변수를 고려하였다. 추정 결과, 정류장별로 추정한 모형보다 기종점별로 추정한 모형이 더 우수한 성능을 나타냈다. 이는 정류장별로 집계된 데이터에 비해 기종점별로 집계된 데이터의 양이 많기 때문인 것으로 판단하였다.

Lee and Sohn(2015)은 Reversible-jump Markov Chain Monte Carlo 알고리즘을 기반으로 지하철 환승통행자들의 환승경로를 추정하였다. 환승경로는 이용량, 통행시간, 차내시간 등을 고려하여 다양한 경로를 추정하였으며, BIC(Bayesian Information Criterion)에 의해 추정된 기존 방법론과 비교하였다. 비교 결과, 기존모형에 비해 경로의 수, 평균 통행시간 등의 측면에서 추정력이 뛰어난 것으로 나타났다.

환승통행 관련 연구 및 딥러닝 기법을 활용한 선행연구를 검토한 결과 대부분이 환승통행량에 관한 연구가 수행되었으며, 딥러닝 기법을 활용한 교통 분야의 연구에서도 대중교통 환승에 관한 연구는 미흡한 것으로 나타났다. 따라서 본 연구에서는 딥러닝 기법을 활용하여 서울시 대중교통 이용자의 지하철-버스 환승시간을 추정하고자 한다.

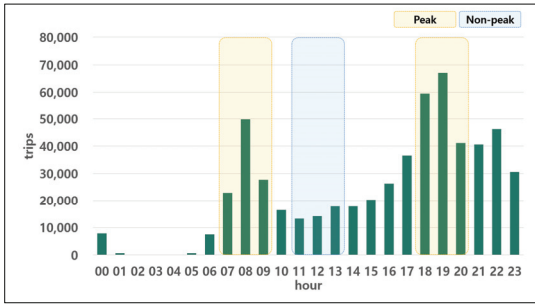
Ⅲ. 분석방법론

1. 자료 수집 및 변수 구축

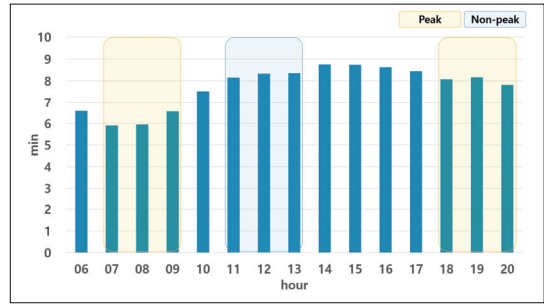
일반적으로 대중교통 환승통행은 버스 간의 환승, 지하철 간의 환승, 버스와 지하철 간의 환승 총 3가지 유형으로 구분할 수 있다. 이 중 버스간의 환승은 환승시간이 상대적으로 매우 작고 지하철간의 환승은 환승시간의 자료취득이 어려우므로 버스와 지하철간의 환승통행을 대상으로 분석하는 것이 타당하다. 그러나 버스→지하철 환승의 경우 버스 하차 후 바로 지하철을 탑승하는 통행행태로 인해 환승시간이 거의 일정하게 나타나 모형구축이 어려운 것으로 나타났다. 따라서 본 연구에서는 환승시간의 변동성이 상대적으로 높은 지하철→버스간의 환승통행을 대상으로 모형을 추정하고자 한다.

먼저 서울시의 지하철 역사별 환승시간을 추정하기 위해 환승시간 및 대중교통 운영 자료, 토지이용 및 지역 관련 자료 등을 수집하였다. 수집자료는 2017년 5월 18일 목요일 기준이며, 교통카드 이용자료, 대중교통 기반정보자료, 서울시 토지이용정보, 건축물 연면적정보를 활용하였다. 교통카드 이용자료의 경우 환승시간이 기록되지 않는 지하철 간의 통행을 제외한 대중교통 환승통행은 2,020,962건으로 나타났으며, 이중 지하철→버스 환승통행은 718,130건으로 전체의 약 35.5%를 차지한다. 수도권 통합 환승할인제에 따라 환승 유효시간은 30분(오후 9시~익일 오전 7시는 1시간)이며, 환승 유효시간과 환승 유효거리(도보 속도 4km/h 기준)를 벗어난 통행에 대해서는 이상치로 제거하였다. 또한 지하철 개찰구에서 출구로 나가 버스 정류장까지의 이동거리를

고려하여 이상치를 제거하였으며, 이상치 제거 결과 총 564,570건에 대하여 분석을 수행하였다. <Fig. 1>과 <Fig. 2>의 시간대별 환승통행량 및 환승시간을 살펴보면 출근시간대(07~09시)와 퇴근시간대(18~20시)에 환승통행이 집중되며, 출근시간대의 환승시간은 약 6.1분으로 다른 시간대에 비해 짧게 나타났다. 이는 출근시간대에는 대중교통 배차간격이 짧고 외적 요인에 영향을 덜 받기 때문인 것으로 판단된다. 이처럼 출근시간대와 퇴근시간대, 비첨두시간대에 환승시간에 영향을 미치는 요인에 차이가 있을 것으로 판단되어 본 연구에서는 일일, 출근시간대, 퇴근시간대, 비첨두시간대를 구분하여 환승시간 추정모형을 구축하고자 한다.



<Fig. 1> Distribution of transfer trips



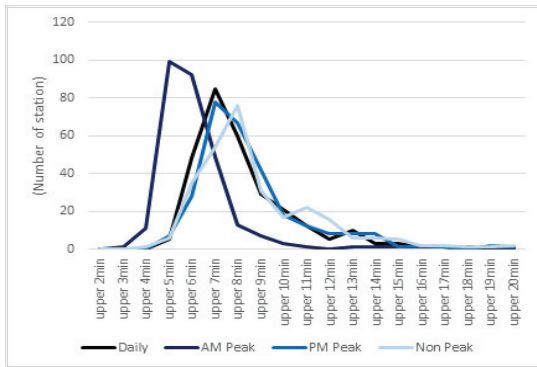
<Fig. 2> Distribution of average transfer time

환승시간은 서울시 내 284개 지하철역을 기준으로 집계하였으며, 지하철역 2km 반경 내에 위치한 버스 정류장으로의 환승시간을 산정하였다. 환승시간 산정결과는 <Table 1>과 같다. 평균 환승시간은 일일 기준 약 8.65분, 출근시간대 6.66분, 퇴근시간대 9.04분, 비첨두시간대 9.34분으로 출근시간대에 환승시간이 가장 짧은 것으로 나타났다. 표준편차 역시 출근시간대에 1.81로 다른 시간대에 비해 작아 지하철역별 차이가 크지 않은 것으로 나타났다.

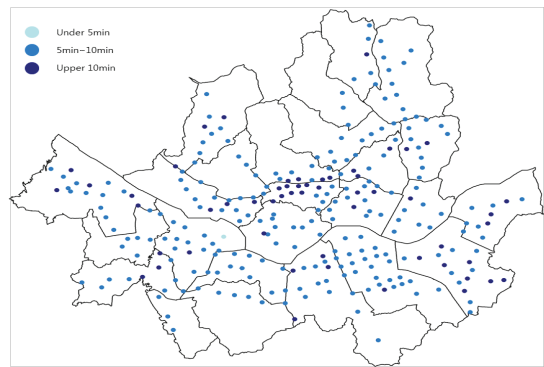
<Table 1> Result of descriptive statistics for variables

Variables		No.	Mean	Standard deviation	Minimum	Maximum
Transfer time (min)	Daily	284	8.65	2.24	5.57	21.64
	AM peak	282	6.66	1.81	3.88	18.10
	PM peak	283	9.04	2.46	5.41	22.70
	Non peak	283	9.34	2.75	4.47	25.48
Bus headway (min)	Daily	284	10.62	1.21	8.02	15.54
	AM peak	284	10.11	1.11	7.65	15.16
	PM peak	284	10.52	1.37	8.22	16.00
	Non peak	284	10.55	1.46	7.72	15.71
Distance to bus stops (m)	Daily	284	162.26	72.24	29.07	655.73
	AM peak	282	148.69	68.81	17.91	542.53
	PM peak	283	164.10	70.25	26.37	536.02
	Non peak	283	161.98	71.28	22.90	514.93
Number of bus lines		284	22.29	14.43	2	100
Commercial area (ha)		284	9.66	17.46	0.00	93.75
Floor area of retail-sales facilities (ha)		284	22.90	19.45	0.02	122.06

지하철역의 평균 환승시간 분포는 <Fig. 3>과 같이 나타난다. 환승시간이 6분 이상 12분 미만인 지하철역이 255개로 전체의 89.8%를 차지하고 있다. 시간대별 평균 환승시간 분포를 살펴보면 출근시간대에 평균 환승시간이 다른 시간대에 비해 짧게 나타났으며, 비침두시간대에는 환승시간이 더 길게 나타났다. 이는 출근 시간대에 더 많은 버스가 운행하여 배차간격이 짧게 나타나고, 반대로 비침두시간대에 배차간격이 길게 나타나기 때문인 것으로 판단된다. 평균 환승시간의 지역적 분포를 살펴보면 <Fig. 4>와 같이 나타난다. 평균 환승시간이 10분을 초과하는 지하철역은 서울역, 종각역, 용산역, 명동역 등 서울 중심부와 잠원역, 잠실나루역, 고덕역, 마천역 등 동남부 지역으로 대부분이 지하철역과 버스 정류장 간의 거리가 멀고, 유동인구가 많은 지역인 것으로 나타났다.



<Fig. 3> Distribution of average transfer time



<Fig. 4> Regional distribution(Daily)

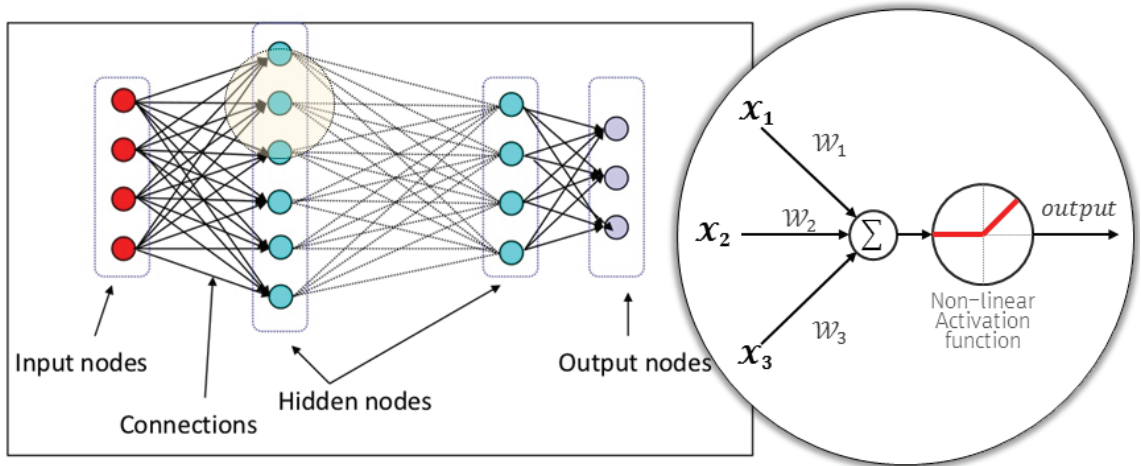
대중교통 운영 관련 변수로 버스 배차간격, 버스정류장까지의 거리, 버스 노선수를 구축하였으며, 버스 배차간격과 버스정류장까지의 거리는 환승시간과 마찬가지로 시간대별로 구분하여 산정하였다. 버스 배차간격은 지하철역 2km 반경 내에 위치한 버스 정류장을 통과하는 노선들의 평균 배차간격을 산정한 뒤 노선별 환승통행량을 가중 평균하여 산출하였다. 평균 배차간격은 일일 기준 약 10.62분, 출근시간대 10.11분, 퇴근시간대 10.52분, 비침두시간대 10.55분으로 출근시간대가 가장 짧은 것으로 나타났다. 버스 정류장까지의 거리도 지하철역으로부터 2km 반경 내에 위치한 버스 정류장까지의 거리를 환승통행량으로 가중 평균하여 산정하였다. 평균 버스 정류장까지의 거리는 일일 기준 162.26m, 출근시간대 148.69m, 퇴근시간대 164.10m, 비침두시간대 161.98m로 환승시간, 버스 배차간격과 마찬가지로 출근시간대에 가장 짧은 것으로 나타났다. 버스 노선수는 시간대별로 다르지 않기 때문에 시간대별 구분 없이 구축하였다. 각 지하철역 환승이 가능한 버스 노선수는 평균 22.29개로 나타나며, 최소 2개, 최대 100개로 나타났다.

지역 관련 변수로 지하철역 주변의 상업지역 면적, 판매 및 편의시설 연면적을 선정하였으며, 지하철역의 영향권인 반경 500m를 기준으로 면적을 산출하였다. 상업지역 면적은 평균 9.66ha, 판매 및 편의시설 연면적은 22.90ha로 나타났다. 또한 각 지하철역이 속해있는 서울시 25개 지역구를 지역 더미(dummy) 변수로 구축하였다.

2. 분석모형

본 연구에서는 서울시 내 지하철 역사별 버스 환승시간을 추정하였다. 환승시간 추정에 있어 새로운 연구 방법론인 딥러닝 기법 중 심층신경망을 활용하였으며, 전통적인 회귀모형의 추정결과와 비교하였다.

딥러닝 기법 중 하나인 심층신경망은 입력층과 출력층 사이에 2개 이상의 은닉층으로 구성되는 인공신경망으로써 비선형의 분류(Classification)와 회귀(Regression) 문제를 해결하는데 주로 사용된다. 심층신경망의 구조는 <Fig. 5>와 같으며, 은닉층의 노드에서는 앞선 입력층에서 받은 정보와 가중치를 내적한 값을 출력하며, 이 출력값은 활성화함수를 거쳐 다음 은닉층으로 전달 여부가 결정된다. 이때 은닉층의 개수, 은닉층 노드의 개수, 활성화함수 등의 매개변수를 조정하여 최적 모델을 도출한다. 반면 회귀분석은 다양한 영향요인에 따른 영향력을 계량화하여 환승시간을 추정한다.



<Fig. 5> Structure of Deep neural network

3. 모형 평가 방안

심층신경망모형과 회귀모형은 동일한 변수를 사용하였으며, 모형의 추정 성능을 평가하기 위하여 데이터를 학습 세트와 검증 데이터 세트로 분할하였다. 학습 데이터 세트와 검증 데이터세트의 구분은 임의추출법(random sampling)을 통해 이뤄졌으며, 전체 284개 역사 중 227개 역사(80%)를 학습에 사용하였고, 나머지 57개 역사(20%)를 학습된 모형의 추정성능 평가에 사용하였다. 또한, 학습된 모형의 추정성능을 평가하기 위해 RMSE(Root Mean Square Error), %RMSE, MAE(Mean Absolute Error), MAPE(Mean Absolute Percentage Error) 지표를 사용하였다.

IV. 환승시간 추정

1. 심층신경망 모델을 활용한 환승시간 추정

심층신경망 모형을 활용한 환승시간 추정은 일일, 출근시간대, 퇴근시간대, 비첨두시간대로 구분하여 총 4개의 모형을 구축하였다. 모형의 독립변수로는 배차간격, 버스정류장 거리, 버스 노선수, 상업지역 면적, 판매 및 편의시설 연면적, 지역 변수를 사용하였다.

독립변수 설정과 함께 심층신경망 모형의 구조 설계가 필요한데, 본 연구에서는 은닉층의 개수와 은닉층

노드의 개수, 과적합을 막기 위한 드롭아웃 비율을 달리하여 각 시간대별로 모형의 최적 구조를 도출하였다. 은닉층의 개수는 3, 5, 7로, 은닉층 노드의 개수는 16, 32, 64, 128, 256으로, 드롭아웃 비율은 0.2, 0.3, 0.5로 변화해가며 모형의 성능을 비교하였다.

은닉층의 개수와 은닉층 노드의 개수를 제외한 학습률(Learning rate), 학습횟수(Epoch), 활성화함수(Activation function), 손실함수(Loss function), 최적화함수(Optimizer)는 각 시간대별 모형에 동일하게 적용하였다. 이때, 활성화함수는 시그모이드(Sigmoid) 함수의 그라디언트(Gradient) 소실 문제를 해결할 수 있는 ReLU(Rectified Linear Unit) 함수를 사용하였으며, 손실함수는 회귀문제에서 주로 사용되는 MSE(Mean Square Error) 손실함수를 사용하였다. 또한 심층신경망에서 각 노드와 연결된 가중치를 조절하는 최적화함수는 Adam(Adaptive moment estimation)을 사용하였다. 심층신경망 모형의 학습에 대한 정보는 <Table 2>와 같다.

<Table 2> Configuration of Deep neural networks

Parameters	Values
Learning rate	0.01
Epoch	200
Drop-out ratio	0.2, 0.3, 0.5
Activation function	ReLU
Loss function	MSE loss
Optimizer	Adam
Number of hidden layers	3, 5, 7
Number of hidden nodes	16, 32, 64, 128, 256

실제 환승시간과 모형의 예측 환승시간 간의 MSE 값이 가장 작은 모형을 최적 모형으로 선택하였다. 시간대별로 선택된 모형을 살펴보면, 최종 모형의 [드롭아웃비율, 은닉층 개수, 은닉층 노드 개수]는 일일 모형의 경우 [0.2, 5, 64], 출근시간대 모형은 [0.3, 7, 16], 퇴근시간대 모형은 [0.2, 5, 128], 비첨두시간대 모형은 [0.3, 3, 64]로 선택되었다.

<Table 3> Result of model structures

Contents	Drop-out ratio	Number of hidden layers	Number of hidden nodes
Daily	0.2	5	64
AM peak	0.3	7	16
PM peak	0.2	5	128
Non peak	0.3	3	64

2. 회귀모형을 활용한 환승시간 추정

회귀모형을 활용한 버스 환승시간 추정결과는 <Table 4>와 같다. 각각의 버스 환승시간 추정모형을 살펴보면, 모형의 설명력을 나타내는 수정된 결정계수는 0.377~0.543으로 설명력이 충분한 것으로 분석되었다. 각 시간대별 환승시간 추정모형은 유사하게 나타났으며, 변수별 영향요인은 다음과 같다. 대중교통 운영 관련 변수 중 배차간격, 버스 정류장까지의 거리는 양의 영향력을 나타내어 지하철역 주변 버스의 배차간격이 길

수록, 버스 정류장까지의 거리가 멀수록 버스 환승시간이 증가하는 것으로 분석되었다. 반대로 버스 노선은 음의 영향력을 나타내어 지하철역 주변 버스 노선수가 많을수록 버스 환승시간이 짧아지는 것으로 분석되었다. 토지이용 관련 변수인 상업지역 면적, 판매 및 편의시설 연면적은 모두 양의 영향력을 나타내었다. 지하철역에서 버스로 환승 시 지하철역 인근이 상업지역으로 지정되어 관련 시설이 많거나 판매 및 편의시설이 많을수록 환승도중에 쇼핑 활동 등을 유발시켜 버스 환승시간이 더 증가하는 것으로 분석된다. 비첨두 시간대에는 배차간격이 통계적으로 유의하지 않게 나타났으며, 출근시간대에는 판매 및 편의시설 연면적이 통계적으로 유의하지 않게 나타났다. 이는 출근시간대의 시간제약으로 인해 타시간대에 비해 환승시 쇼핑활동을 할 여유가 없기 때문인 것으로 판단된다.

지역적 특성변수를 살펴보면 지하철역이 종로구에 위치할 경우 버스 환승시간이 증가하는 것으로 분석되었는데, 이는 종로구의 경우 유동인구가 많아 버스 환승 시 시간이 오래 걸리는 것으로 판단된다. 금천구의 경우 출근시간대에서만 양의 영향력을 나타내는 것으로 추정되었다. 이는 업무시설 밀집지역인 금천구의 특성상 업무지역으로 이동 시 버스 환승이 필요하지만 버스 배차간격이 길고, 이용량이 많기 때문에 버스 환승시간이 긴 것으로 판단된다. 강동구의 경우 출근시간대를 제외한 다른 시간대에서는 모두 양의 영향력을 갖는 것으로 나타났다. 이는 강동구가 상대적으로 주거지역이 높게 분포되어 있으며, 도심지역보다 지하철역이 적고 버스 배차간격도 길기 때문에 버스 환승시간이 증가하는 것으로 판단된다. 각 영향요인들의 표준화 계수를 살펴보면 버스 정류장까지의 거리가 가장 높게 나타나 버스 환승시간에 가장 큰 영향을 미치는 것으로 분석되었다.

<Table 4> Estimation results of regression models for bus transfer time

Variables		Daily		AM peak		PM peak		Non peak	
		Coef.	B	Coef.	B	Coef.	B	Coef.	B
Bus headway (min)	Daily	0.271**	0.147	-	-	-	-	-	-
	AM peak	-	-	0.139*	0.085	-	-	-	-
	PM peak	-	-	-	-	0.186**	0.103	-	-
	Non peak	-	-	-	-	-	-	0.127	0.068
Distance to bus stops (m)	Daily	0.020**	0.660	-	-	-	-	-	-
	AM peak	-	-	0.016**	0.593	-	-	-	-
	PM peak	-	-	-	-	0.020**	0.568	-	-
	Non peak	-	-	-	-	-	-	0.021**	0.547
Number of bus lines		-0.033**	-0.215	-0.029**	-0.228	-0.031**	-0.180	-0.038**	-0.199
Commercial area (ha)		0.022**	0.168	0.014**	0.137	0.024**	0.171	0.023**	0.144
Floor area of retail-sales facilities (ha)		0.016**	0.142	0.001	0.012	0.014*	0.107	0.020**	0.141
Jongno-gu[dummy]		1.401**	0.136	0.711*	0.086	1.429**	0.126	1.794**	0.142
Geumcheon-gu[dummy]		-0.109	-0.006	2.456**	0.161	-0.166	-0.008	0.114	0.005
Gangdong-gu[dummy]		1.789**	0.140	-0.039	-0.004	1.472**	0.105	2.201**	0.141
Constant		2.488**	-	3.343**	-	3.843**	-	4.592**	-
N		284		282		283		283	
Adj. R ²		0.543		0.377		0.409		0.400	

Note: ** significance level at $\alpha \leq 0.05$, * significance level at $\alpha \leq 0.1$.

3. 모형 추정 성능 비교

구축한 회귀모형과 심층신경망 모형의 추정 성능은 <Table 5>와 같다. 회귀모형과 심층신경망 모형 모두 출근시간대의 RMSE가 각각 1.04, 1.03으로 모형의 추정 성능이 가장 높았으며, 비첨두시간대의 RMSE가 각 2.84, 3.05로 추정 성능이 가장 낮게 나타났다.

회귀모형과 심층신경망 모형의 추정 성능을 시간대별로 비교하면, 비첨두시간대를 제외하면 모든 시간대에서 심층신경망의 RMSE, %RMSE, MAE, MAPE 지표가 가장 낮은 것으로 평가되어 심층신경망 모형이 회귀모형보다 좀 더 높은 추정 성능을 보였다.

일일 모형의 경우 회귀모형의 RMSE는 1.78, 심층신경망 모형의 RMSE는 1.47로 심층신경망 모형이 0.31 낮게 나타나 다른 시간대의 모형에 비해 추정 성능 차이가 비교적 크게 나타났다. 출근시간대와 퇴근시간대 모두 심층신경망 모형의 추정 성능이 높았으나 RMSE 차이는 각 0.01, 0.05로 매우 작게 나타났다. 반면, 비첨두시간대의 경우 회귀모형의 RMSE가 2.84, 심층신경망의 RMSE가 3.05로 회귀모형의 추정 성능이 더 높은 것으로 나타났다. 전반적으로 심층신경망 모형이 회귀모형보다 추정 성능이 더 높게 나타났으나, 그 차이가 크지 않은 것은 심층신경망 모형이 284개 역사의 한정적인 데이터만을 학습하였기 때문인 것으로 판단된다.

<Table 5> Results of model evaluation

Contents		RMSE	%RMSE	MAE	MAPE
Daily	Regression	1.78	19.5%	1.31	14.1%
	DNN	1.47	16.1%	1.04	11.0%
	Difference	0.31	3.4%	0.27	3.0%
AM peak	Regression	1.04	15.6%	0.80	12.0%
	DNN	1.03	15.5%	0.74	10.7%
	Difference	0.01	0.1%	0.07	1.2%
PM peak	Regression	2.18	23.2%	1.40	14.7%
	DNN	2.13	22.7%	1.33	13.4%
	Difference	0.05	0.5%	0.07	1.2%
Non peak	Regression	2.84	25.8%	1.94	16.0%
	DNN	3.05	27.7%	2.09	17.7%
	Difference	-0.21	-1.9%	-0.16	-1.7%

V. 결론 및 향후 연구과제

본 연구에서는 교통카드 이용자료를 활용하여 서울시의 지하철 역사별 버스 환승시간을 추정하였다. 환승시간 추정을 위해 새로운 연구방법론인 딥러닝 기법을 활용하였으며, 전통적 추정방법인 회귀분석과 추정결과를 비교분석하고자 하였다.

모형구축을 위해 2017년 5월 18일 기준의 서울시 교통카드 이용자료와 대중교통 기반정보 자료, 서울시 토지이용정보 및 건축물 연면적 자료 등을 활용하여 환승시간 및 대중교통 운영 관련 변수, 토지이용 및 지역 관련 변수를 구축하였다. 환승시간 추정을 위해서는 회귀모형과 함께 딥러닝 기법 중 심층신경망 모형을 개발하였으며, 이들 모형의 성능을 RMSE, %RMSE, MAE, MAPE 지표를 통해 비교하였다.

먼저 심층신경망 모형은 드롭아웃 비율, 은닉층의 개수, 은닉층 노드의 개수를 달리하여 각 시간대별로 최적의 모형을 구축하였다. 회귀분석 모형도 각 시간대별 수정된 결정계수는 0.377~0.543으로 설명력이 충분한 것으로 나타났다. 회귀모형과 심층신경망 모형의 추정 성능을 비교한 결과, 두 가지 모형 모두 출근시간대 모형이 추정 성능이 가장 높았으며, 비첨두시간대 모형이 가장 낮은 것으로 나타났다. 이는 타시간대 보다 상대적으로 변동폭이 적은 출근시간대의 환승통행패턴의 예측력이 높음을 시사하고 있다. 또한 비첨두시간대를 제외한 모든 시간대에서 심층신경망 모형이 회귀모형보다 추정 성능이 더 높은 것으로 나타나 심층신경망 모형을 이용하여 환승시간을 추정하는 것이 보다 예측의 정확성을 높일 수 있을 것으로 판단된다. 추가적으로 대중교통 운영, 토지이용, 지역 특성 등의 관련 변수가 환승시간에 유의한 영향을 주는 것으로 분석되어, 향후 이들 요인을 고려한 대중교통 환승정책을 수립하는 것이 보다 효율적일 것이다.

한편, 본 연구는 데이터의 제약으로 인한 몇 가지 한계점을 가지고 있다.

첫째, 연구의 범위가 제한적이다. 시간적 범위를 특정 하루를 대상으로 하였기 때문에 요일적 요인이나 계절적 요인과 같은 시계열적 특성이 반영되지 못하였다. 또한, 공간적 범위가 서울시로 한정되어 있어 시도 간을 오가는 광역버스 등의 환승패턴이 반영되었다고 보기 어렵다. 따라서 향후에는 공간적 범위를 수도권으로 확대하며, 일일 이용자료가 아닌 일주일 이상의 자료를 활용하여 시계열적 특성을 반영해야 할 것이다.

둘째, 개인에 대한 특성이 반영되지 않았다. 교통카드 이용자료에는 이용자의 연령, 성별 등에 대한 개인 속성이 기록되지 않기 때문에 모형에 개인적인 특성을 반영하기가 어려웠다. 하지만 이용자의 연령과 성별 등이 환승시간에 유의미한 영향을 줄 수 있으므로 향후 개인적인 특성을 반영할 수 있는 방안에 대한 고려가 필요하다.

셋째, 심층신경망 모형이 회귀모형에 비해 추정성능은 높았으나 그 차이가 크지는 않았는데, 이는 집계된 데이터를 활용하였기 때문이다. 본 연구에서는 서울특별시 284개 역의 집계된 환승시간을 추정하였기 때문에 전체 데이터의 개수가 284개에 불과하여 딥러닝을 통해 학습하고 검증하기에는 부족한 측면이 있다. 따라서, 환승시간 추정의 관점에서 볼 때 역별로 집계된 데이터가 아닌 개인 속성을 포함한 개별 통행 데이터를 활용한다면 보다 정밀한 추정이 가능할 것으로 판단된다.

ACKNOWLEDGEMENTS

본 연구는 2018년도 홍익대학교 학술연구진흥비의 지원을 받아 수행하였습니다. 본 연구의 초기 아이디어는 2019년도 한국ITS학회 추계학술대회에서 발표되었으며, 이를 바탕으로 수정·보완하여 본 논문을 작성하였습니다.

REFERENCES

- Baek J. H. and Sohn K. M.(2017), "Deep-Learning Architectures to Forecast Bus Ridership at the Stop and Stop-To-Stop Levels for Dense and Crowded Bus Networks," *Applied Artificial Intelligence*, vol. 30, no. 9, pp.861-885.
- Duan Y., Lv Y., Liu Y. and Wang F.(2016), "An efficient realization of deep learning for traffic data imputation," *Transportation Research Part C*, vol. 72, pp.168-181.
- Han G. I. and Sohn K. M.(2016), "Activity imputation for trip-chains elicited from smart-card data

- using a continuous hidden Markov model,” *Transportation Research Part B*, vol. 83, pp.121-135.
- Huang W., Hong H., LI M. and Hu W.(2013), “Deep Architecture for Traffic Flow Prediction,” *ADMA2013*, vol. 8347, no. 2, pp.165-176.
- Hwang J. H.(2014), “Analyzing Factors Affecting Public Transit Transfer Volume: Focused on Daegu City,” *Journal of Korean Society of Transportation*, vol. 32, no. 3, pp.179-186.
- Jang W. J.(2010), “Travel Time and Transfer Analysis Using Transit Smart Card Data,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2144, pp.142-149.
- Jung J. Y. and Sohn K. M.(2017). “Deep-learning architecture to forecast destinations of bus passengers from entry-only smart-card data,” *The Institution of Engineering and Technology*, vol. 11, no. 6, pp.334-339.
- Lee M. S. and Sohn K. M.(2015), “Inferring the Route-Use Patterns of Metro Passengers Based Only on Travel-Time Data Within a Bayesian Framework Using a Reversible-Jump Markov Chain Monte Carlo(MCMC) Simulation,” *Transportation Research Part B*, vol. 81, pp.1-17.
- Lee S. H., Kim J. S., Kim M. S. and Woo Y. H.(2013), “A Study on Users’ Travel Behavior Analysis of Transit Transfer,” *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 12, no. 12, pp.147-157.
- Li D., Dong Y.(2014), “Deep Learning Methods and Applications” *Foundations And Trends In Signal Processing*, vol. 7 no. 3-4, pp.197-387.
- Lim S. Y.(2017), *Exploring the Relationship between Transfer Trips and Land Use*, Masters dissertation, Hongik University, Korea.
- Lim S. Y., Lee H. S. and Choo S. H.(2016), “Analysis of Influencing Factors on Transfer Time of Public Transportation in Seoul-Focusing on Transfer Users from Subway to Bus,” *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 15, no. 2, pp.1-12.
- Lv Y., Duan Y., Kang W., Li Z. and Wang F.(2015), “Traffic Flow Prediction With Big Data: A Deep Learning Approach,” *IEEE Transactions on Intelligent Transportation System*, vol. 16, no. 2, pp.865-873.
- Ma X., Tao Z., Wang Y., Yu H. and Wang Y.(2015), “Long Short-Term Memory Neural Network for Traffic Speed Prediction Using Remote Microwave Sensor Data,” *Transportation Research Part C*, vol. 54, pp.187-197.
- Ma X., Yu H., Wang Y. and Wang Y.(2015), “Large-Scale Transportation Network Congestion Evolution Prediction Using Deep Learning Theory,” *PLoS One*, vol. 10, no. 3, pp.1-17.
- Ministry of Land, Infrastructure and Transport(2018), *Public Transportation Investigation 2018*, p.14.
- Polson N. G. and Sokolov V. O.(2017) “Deep learning for short-term traffic flow prediction,” *Transportation Research Part C*, vol. 79, pp.1-17.
- Ryu J. D. and Yoon I. S.(2018), *Development of Expressway Traffic Accident Prediction Model Using Deep Learning*, Doctoral dissertation, Ajou University, Korea.
- Seoul Solution, <https://www.seoulsolution.kr>, 2018.4.18.
- Xiong G., Kang W., Wang F., Zhu F., Lv Y., Dong X., Rieki J. and Pirttikangas S.(2015), “Continuous Travel Time Prediction for Transit Signal Priority Based on a Deep Network,” *IEEE 18th International Conference*, pp.524-528.