2020년 한국ITS학회 추계학술대회
그린뉴딜 정책에 따른 ITS의 확대 추진 및 고도화

GRU를 이용한 서울시 지하철 구간별 혼잡도 예측 모델 연구

권효승, 최창준, 정효석, 송재인, 강민희, 황기연

To cite this article : 권효승, 최창준, 정효석, 송재인, 강민희, 황기연 (2020) GRU를 이용한 서울시 지하철 구간별 혼잡도 예측 모델 연구, 2020년 한국ITS학회 추계학술대회, pp.185-190

GRU를 이용한 서울시 지하철 구간별 혼잡도 예측 모델 연구

A Study on the Prediction Model of Congestion by Subway Sections in Seoul Using GRU

권효승*, 최창준*, 정효석*, 송재인, 강민희***, 황기연****

(*홍익대학교 도시공학과 학사과정

**홍익대학교 과학기술연구소 연구교수

***홍익대학교 산업융합협동과정 스마트시티 박사과정

****홍익대학교 도시공학과 교수)

Key Words : 서울시 지하철, 혼잡도, 지하철 최단경로, R seoul subway package, GRU, 혼잡도영향데이터

목 차

Ⅰ.서론

1. 연구의 배경 및 목적

서울특별시의 지하철 연간 이용 인원은 약 28억명으로 1일 이용자수는 약 746만명이며, 전체 통행수단 중 약 40.7%의 수단분담률[1]을 차지하고 있다. 이와 같이 다수가 이용하는 서울시 지하철은 출·퇴근 시간 등 첨두시에 수송량이 집중되고 있으며, 특히 업무중심 · 주거중심지역 등이 밀집한 특정 구간에서 혼잡도가 증가하여 심각한 문제로 나타나고 있다. 9호선의 경우 오전 첨두시간의 혼잡도가 당산~여의도 구간 219%, 염창~당산구간 234%[2] 로 타 노선 및 시간대 대비 낮은 서비스 수준을 보이고 있다. 향후 3기 신도시 개발 및 지하철 노선 연장에 따라 수도권에 인구가 집중 될 경우, 특정 구간의 혼잡도 증가로 인한 서비스 수준의 하락은 심화될 것으로 예상된다.

첨두시간대의 높은 혼잡도는 낮은 서비스 수준 뿐만 아니라 전염병 감염에 대한 취약성을 높일 수 있다. 최근 전세계적으로 확산되고 있는 코로나 바이러스의 경우 2m 이내 접촉시 높은 감염률을 보인다. 특히 카페, 클럽, 대중교통 등 사람이 밀집할 수 있는 폐쇄된 공간에서 강한 전염성을 나타내고 있다. 이와 관련하여 영국 사우스 햄튼 대학의 연구진은 중국 고속철 사례를 분석하였으며, 동일 고속철 내 평균 감염율은 0.32%였지만 밀접 접촉 공간 감염률은(확진자가 앉은 자리에서 앞뒤 5칸, 좌우 3칸) 최대 10.3%까지 높아진다는 결과를 도출하였다. 이로 보아 고속철보다 좁은 공간에 밀집되어 거리유지가 불가능한 첨두시간의 지하철의 경우 감염 위험이 더욱 높을 것으로 예상된다. 실제로 이용객들도 지하철 내 감염에 대해 불안감을 느낀다는 것을 코로나 확진자 수가 크게 증가한 2020년 1~3월의 월별 지하철 이용객이 약 1억 3,800만명에서 약 9,400만명으로 30% 이상 크게 감소하였다는 지표로 확인[3]할 수 있다.

1) 수도권 여객 기·종점통행량(O/D) 현행화 공동사업(2019)

2) 서울시정개발연구원

3) 서울교통공사 2020년 1월~7월 수송수입 실적

위에 상기된 서비스 수준의 하락 및 코로나 감염 취약성을 개선하기 위해서는 이용객과 운용자가 대비할 수 있도록 첨두 시간대에 특정구간의 혼잡도를 사전에 예측할 필요가 있다고 판단된다. 이에 본 연구에서는 코로나 감염상황 등을 반영한 30분 단위의 지하철 혼잡도 예측 모델을 연구하고자 한다. 이를 통해 대중교통 이용자 및 운영 기관에 혼잡도 정보 제공하고 이를 대응할 수 있는 정책 방안을 모색하고자 한다.

2. 연구의 범위 및 수행방법

연구의 범위는 공간적으로는 서울시 지하철 1~9호선을 대상으로 하였으며 시간적으로는 2019년 10월부터 2020년 7월까지의 데이터를 활용하였다.

연구 수행방법은 다음과 같다. 우선 지하철 혼잡도를 분석 및 예측한 기존 연구들을 고찰한 후 연구의 차별성을 도출하고자 한다. 이후 서울시 빅데이터 캠퍼스에서 구득이 가능한 30분단위의 서울시 지하철 이용객 승하차 데이터를 통해 구간별 승하차량을 산출하여 데이터 셋을 구성하고자 한다. 마지막으로 가공데이터 이외에 지하철 혼잡도에 영향을 미칠 것으로 예상되는 Context 데이터를 결합하여 딥러닝 기반의 GRU 모델을 통해 예시로 가장 혼잡한 구간의 혼잡도를 예측하고자 한다.

Ⅱ.선행연구 고찰

본 장에서는 지하철 혼잡도 추정 및 예측에 관련된 국내·외 선행연구를 고찰하였으며, 본연구와의 차별성을 도출하였다. 신성일(2011)은 대중교통카드자료를 활용하여 도시철도의 차량 및 환승역의 혼잡도를 추정하는 방법론을 제안하였다. 이때 동적 최적경로탐색 알고리즘을 활용하였으며 호선별 도시철도 네트워크, 환승네트워크, 스케쥴 기반의 네트워크를 구축하여 첨두시 혼잡도를 추정하였다. 또한 해당연구를 통해 기존 도시철도 혼잡도 추정방법 개선, 도시철도 운영정책 개선 등 정책방안을 제안하였다. 신성일(2019)의 연구에서는 지하철 승강장의 동적 혼잡도 추정모형을 개발하고, 스마트카드 자료를 기반으로 정확도를 검증하였다. 모형의 정확도 분석결과 최적경로 및 유사경로 탐지비율이 99.3%로 모형의 정확도가 현실을 충분히 반영한 것으로 도출하였다. LENG Biao et al.(2013)은 2009년 11월부터 2010년 1월의 지하철 OD데이터를 활용하여 확률적 배정을 통해 베이징 지하철 이용량을 추정하는 방법을 제시하였다. 분석결과 제안한 방법으로 출구승객의 흐름을 예측하는 것에는 기존에 비해 성능이 다소 떨어졌으나, 지하철 시스템 전체에서 승객의 이동을 예측할 수 있다는 점에서 의의가 있음을 도출하였다. 방준아(2020)는 수도권 지역 지하철 최단 경로 패키지인 seoulsubway를 이용하여 지하철 이동에 대한 경로와 혼잡도를 추정하였고 Sf-R과 노선 shapefile을 이용하여 구간 혼잡도의 효율적인 시각화 방법을 제시하였다.

이경재(2020)은 대중교통통행사슬 데이터를 가공한 후 LSTM 모형에 그래프 기반의 합성곱 레이어를 추가한 GC-LSTM(Graph Convolutional - Long Short)모형을 활용하여 특정 시점의 지하철 역사의 5분 단위의 승하 차량을 예측하는 딥러닝 모형을 구성하여 다른 추정 모형들과 추정력을 비교하였다. 추정값을 실제 승하차량과 비교했을 때 피크지점에서 LSTM과 GC-LSTM의 추정값의 차이가 큰 것으로 도출되었으며, 주중에 비해 주말에 차이가 큰 것을 알 수 있었다. 이정훈(2018)은 부산광역시의 교통카드데이터와 기온, 습도, 풍속, 강우량을 포함한 기상데이터를 분위회귀모형을 통해 기상특성과 대중교통 이용량과의 관계를 분석하였다. 분석결과 기상조건이 대중교통 통행량과 밀접한 관련이 있음을 확인하였고, 통행수단별, 요일별(주말-평일) 유의미한 차이를 확인 할수 있었다.

김진 외(2010)은 지하철역 400m 반경 내의 토지이용 및 건축물 용적률과 수도권의 지하철 역의 연간 이용량의 관계를 선형회귀모형을 통해 분석하였다. 분석결과 고밀개발이나 토지이용의 혼합, 도시설계가 대중교통 이용을 유도하는 것에 효과가 있음을 도출하였다. 김재익(2013)은 오전 첨두시간대의 지하철 승하차량과 토지이용간의 관계를 분석하였다. 분석결과 역주변의 거주인구 및 주거지 비율이 높을수록 승차량이 높게 나타났으며, 역주변의 고용수, 학교의 정원수, 상업용지 및 공업용지의 비율이 높을수록 하차량이 높은 것으로 도출되었다. 성현곤(2017)은 2011년부터 6년 동안의 대중교통 이용실적을 통하여 월별 대중교통 수단별 이용량의 시계열 변동 발생 원인을 분석하였다. 연구결과 인구의 감소에도 불구하고 수도권 철도망의 개선 및 접근성 향상에 따라 대중교통 이용이 증대되는 것을 확인하였다.

지하철 혼잡도와 관련한 연구를 고찰한 결과, 혼잡도의 지표개발 및 시각화, 혼잡도 추정, 요인분석 등의 연구가 주로 수행되었다. 대다수의 연구에서 지하철 이용객 데이터와 기존 모델링 방법을 이용하여 혼잡도를 추정하였으며 이경재(2020) 등 최근 연구에서 LSTM 등과 같은 인공지능 기법을 활용하였다. 다만 LSTM 이외의 인공지능 방법론을 활용한 연구는 부재한 것으로 나타났다. 본 연구의 경우 정확성 향상을 위하여 코로나 확진데이터 및 기상조건 등 외부 요인을 기반으로 30분 단위의 혼잡도를 추정하고자 하며, 기존에 활용되지 않은 GRU기법을 활용하는 점에서 기존 연구와 차별성을 갖는다.

Ⅲ.연구 방법

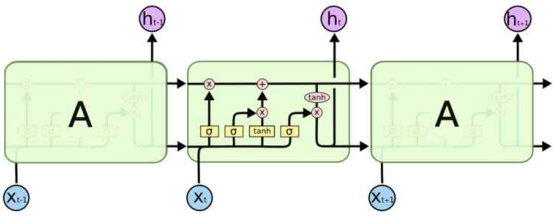
본 장에서는 지하철의 구간별 혼잡도 예측을 위한 연구의 방법론 고찰 및 분석환경설정에 대해 기술하고자 한다.

1. GRU 알고리즘

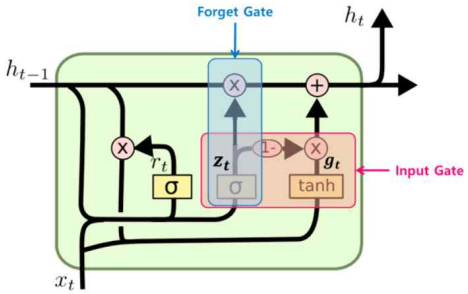
GRU 방법론은 LSTM 셀을 간호화한 방법론으로 볼 수 있다. LSTM은 기존 RNN이 갖고 있는 장기 의존성 문제를 해결한 모델로 여러 게이트로 이루어진 셀(cell)의 형태이며, 게이트별 정보 저장 및 삭제 여부를 결정하고 다음셀로 전달하

는 구조를 갖고 있다[4].

GRU는 업데이트 게이트와 리셋 게이트 총 두 개의 게이트만 가지고 있어 3개의 게이트를 가지고 있는 LSTM보다 연산 속도가 빠르고, 정확도 또한 LSTM과 유사하거나 높은 결과 값을 도출하는 것으로 나타났다[5]. 따라서 속도와 성능 측면에서 더 좋은 결과를 가지는 GRU 모델을 사용하여 지하철 구간 혼잡도를 예측 연구를 진행하였다.



<그림 1 > LSTM 구조



<그림 2 > GRU구조

2. 분석환경설정

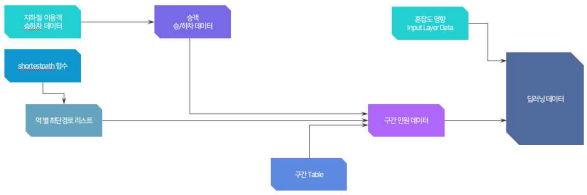
1) 활용 데이터

본 연구에는 지하철 이용객 승하차 데이터, 기상자료 데이터, 코로나 확진자 수 데이터를 사용하였다. 이용객 승하차 데이터는 2019년 10월부터 2020년 7월까지의 30분 단위 자료로 각 구간별 지하철의 인원 수를 도출하는데 사용하였다. 기상자료 데이터와 코로나 확진자 수 데이터는 2019년 10월부터 2020년 7월까지 1일 단위의 자료로 혼잡도에 영향을 미칠 것으로 예상되는 Context 데이터로 사용하였다.

2) 전처리 및 분석과정

(1) 데이터 전처리

GRU를 통한 혼잡도 예측을 하기 위해서는 지하철 이용객 승하차 데이터를 각 구간별 인원 데이터로 가공하는 데이터 전처리 과정이 필요한데 그 과정은 다음과 같다.



<그림 3 > 데이터 전처리 과정

우선, 지하철 이용객 승하차 데이터에서는 승하차 데이터만 제공되어 이것만으로는 각 구간별 지하철 인원을 알 수 없기 때문에 승객이 지나간 경로를 추정할 필요가 있다. 이에 가상의 이동경로를 생성하는 작업을 수행하였다. 승객이 이용한 경로를 승/하차역의 최단경로로 설정하였고, Seoulsubway package의 shortestpath함수를 이용해 최단경로를 산정하였다. 해당 과정에서 함수의 연산속도에 의해 대량의 데이터 처리에 의한 시간 소모를 줄이기 위해 모든 역별 최단경로를 미리 생성한 path_exist list를 통해 해결하고자 하였다. Seoul subway package의 shortestpath함수는 역에서 역까지 가는 시간과 환승시간을 고려한 최소 시간을 기준으로 최단경로를 구성하였다.

	path_exist	list [69777]	List of length 69777
	가락시장-가산디지털...	list [24 x 1] (S3: data.frame)	A data.frame with 24 rows and 1 column
	가락시장-가양	list [28 x 1] (S3: data.frame)	A data.frame with 28 rows and 1 column
	가락시장-강남	list [10 x 1] (S3: data.frame)	A data.frame with 10 rows and 1 column
	가락시장-강남구청	list [10 x 1] (S3: data.frame)	A data.frame with 10 rows and 1 column
	가락시장-강동	list [7 x 1] (S3: data.frame)	A data.frame with 7 rows and 1 column
	station	factor	Factor with 7 levels: "가락시장", "강동", "경찰병원", "둔...
	가락시장-강동구청	list [6 x 1] (S3: data.frame)	A data.frame with 6 rows and 1 column
	가락시장-강변	list [6 x 1] (S3: data.frame)	A data.frame with 6 rows and 1 column
	가락시장-개롱	list [4 x 1] (S3: data.frame)	A data.frame with 4 rows and 1 column
	가락시장-개화	list [34 x 1] (S3: data.frame)	A data.frame with 34 rows and 1 column

<그림 4 > shortestpath함수를 이용한 최단경로

다음으로 생성된 최단경로에 따라 지하철 이용객 수 추정하기 위해 데이터를 정제하였다. 정제 과정에서 본 연구의 공간적 범위인 1~9호선을 제외한 노선을 제거하였다.

YEAR	MONTH	DAY	HOUR	HALF_HOUR	GETON_LINE_NM	GETON_STATION_NM	GETOFF_LINE_NM	GETOFF_STATION_NM	PASSN_CNT
2020	2	5	0	0	1호선	서울역	1호선	서울역	3
2020	2	5	0	0	1호선	서울역	4호선	회현(남대문시장)	3
2020	2	5	0	0	1호선	종로3가	3호선	구파발	3
2020	2	5	0	0	1호선	종로3가	5호선	상일동	3
2020	2	5	0	0	1호선	종로5가	1호선	서울역	3
2020	2	5	0	0	1호선	종로5가	3호선	연신내	3
2020	2	5	0	0	1호선	종로5가	6호선	봉곡(동묘여대)	3
2020	2	5	0	0	1호선	동대문	3호선	연신내	3
2020	2	5	0	0	1호선	제기동	1호선	서울역	3
2020	2	5	0	0	1호선	청량리(서울시립대입구)	1호선	동묘앞	3
2020	2	5	0	0	1호선	동묘앞	1호선	서울역	3
2020	2	5	0	0	2호선	을지로입구	5호선	마장	3
2020	2	5	0	0	2호선	을지로3가	2호선	아현	3
2020	2	5	0	0	2호선	을지로3가	3호선	홍제	3
2020	2	5	0	0	2호선	을지로4가	2호선	상왕십리	3
2020	2	5	0	0	2호선	동대문역사문화공원	2호선	상왕십리	3
2020	2	5	0	0	2호선	동대문역사문화공원	2호선	신림	3

<그림 5 > 기존 승하차 데이터

C.Olach(2015)
J. Chung et al.(2014)

ID	up_Line	up_Name	down_Line	down_Name	Time	Count
ID1	1	서울	1	서울	1	3
ID2	1	서울	4	회현	1	3
ID3	1	종로3가	3	구파발	1	3
ID4	1	종로3가	5	상일동	1	3
ID5	1	종로5가	1	서울	1	3
ID6	1	종로5가	3	연신내	1	3
ID7	1	종로5가	6	월곡	1	3
ID8	1	동대문	3	연신내	1	3
ID9	1	제기동	1	서울	1	3
ID10	1	청량리	1	동묘앞	1	3
ID11	1	동묘앞	1	서울	1	3

<그림 6 > 불필요한 데이터가 제거된 1~9호선 승하차 데이터

이후 각 구간들의 이용 인원을 파악하기 위해 link table을 생성하였으며, 행은 1~9호선의 모든 역간의 구간, 열은 30분 단위의 시간으로 가상의 이동경로를 통해 구간별 인원을 산출하였다.

link	X0	X30	X430	X500	X530	X600	X630
가락시장-경찰병원	0	0	3	188	602	536	969
가락시장-문정	0	0	15	548	900	1295	1435
가락시장-송파	0	0	0	311	1505	921	1418
가락시장-수서	0	0	0	561	627	958	907
가산디지털단지-구로	0	0	0	30	288	153	260
가산디지털단지-남구로	0	0	0	1093	1736	2340	1529
가산디지털단지-철산	6	0	3	677	1301	1177	1872
가양-양천향교	0	0	3	314	705	607	821
가양-증미	0	0	0	481	987	1077	979
가좌-디지털미디어시티	0	0	0	45	93	99	168
가좌-홍대입구	0	0	0	75	31	96	89
가천대-복정	0	0	0	6	9	6	9

<그림 7 > 산출된 구간별 인원 데이터

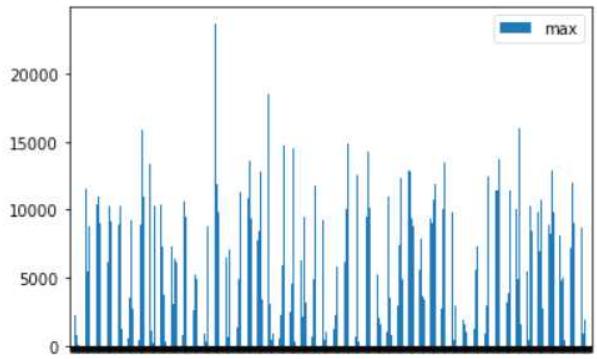
	가락시장-경찰병원	가락시장-문정	가락시장-송파	가락시장-수서	가산디지털단지-구로	가산디지털단지-남구로	가산디지털단지-철산	가양-양천향교	가양-증미	가좌-디지털미디어시티	...	회기-외대앞	회기-중랑	회기-청량리	회현-명동	회현-서울	효창공원앞-공덕	효창공원앞-삼각지	효창공원앞-용산	흑석-노량	흑석-동작
0	82.0	32.0	48.0	0.0	0.0	0.0	56.0	21.0	0.0	3.0	...	6.0	3.0	0.0	0.0	21.0	50.0	0.0	0.0	158.0	0.0
1	237.0	606.0	329.0	569.0	78.0	1111.0	713.0	376.0	507.0	51.0	...	81.0	79.0	427.0	472.0	797.0	498.0	478.0	60.0	505.0	1075.0
2	557.0	1024.0	1338.0	853.0	281.0	1470.0	1327.0	595.0	849.0	90.0	...	323.0	268.0	455.0	1922.0	1906.0	1561.0	1123.0	121.0	1848.0	2098.0
3	495.0	1346.0	1052.0	1027.0	161.0	2473.0	1196.0	709.0	1119.0	128.0	...	163.0	162.0	621.0	1161.0	1999.0	1148.0	1193.0	132.0	1030.0	2263.0
4	990.0	1767.0	1484.0	1129.0	325.0	2014.0	2335.0	649.0	437.0	215.0	...	408.0	482.0	574.0	2848.0	3143.0	2373.0	1610.0	261.0	1913.0	1467.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10682	1030.0	1676.0	1357.0	647.0	142.0	1274.0	2840.0	931.0	680.0	236.0	...	525.0	430.0	266.0	2008.0	1642.0	2328.0	1319.0	203.0	3323.0	1726.0
10683	608.0	1168.0	1605.0	978.0	295.0	2070.0	1484.0	722.0	1110.0	132.0	...	453.0	385.0	485.0	1401.0	1223.0	1456.0	1299.0	180.0	2030.0	2431.0
10684	860.0	1286.0	1319.0	487.0	238.0	1594.0	2208.0	917.0	721.0	241.0	...	356.0	335.0	108.0	1701.0	1178.0	1601.0	1128.0	155.0	3610.0	1903.0
10685	488.0	889.0	1189.0	708.0	242.0	1369.0	1251.0	265.0	419.0	86.0	...	133.0	174.0	357.0	1413.0	1527.0	1392.0	1052.0	151.0	1083.0	898.0
10686	107.0	104.0	176.0	3.0	9.0	105.0	314.0	77.0	61.0	29.0	...	70.0	6.0	6.0	130.0	241.0	206.0	177.0	21.0	186.0	110.0

<그림 8 > 구간별 지하철 탑승인원

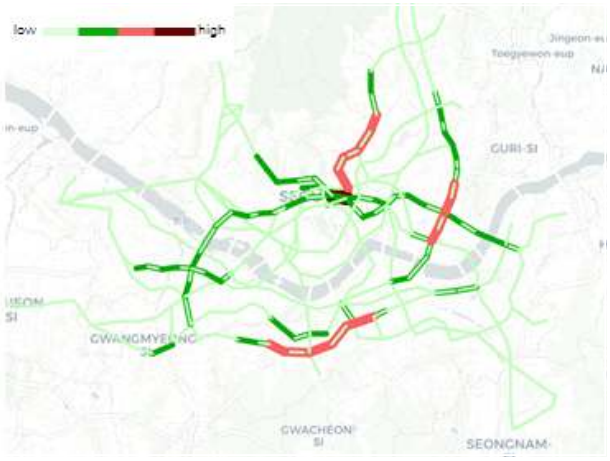
데이터 전처리 과정을 통해 얻어낸 2019년 10월부터 2020년 7월까지의 link_count 승객 수 파일의 배열을 변경한 후 merging을 하였으며 <그림 9>와 같이 나타낼 수 있다. 행렬의 구성은 10,959행 × 718열로 구성되어 있으며 1행은 1일당 39행(1행 당 30분)으로 281일에 대한 정보이고, 열은 역과 역 사이의 구간의 개수이다.

구간별 승객수에 대해 승객수가 높은 구간을 추출하였을 때 동대문역사공원 → 을지로4가의 구간이 23,774명으로 가장 높게 나타났으며, 특히 오전 첨두시에 가장 높은 5개 구간으로 '동대문역사문화공원 → 을지로4가' '사당 → 방배' '낙성대→ 사당' '방배 → 서초' '한성대입구 → 혜화'구간이 도출되었다.

GRU 분석을 위해 구간 승객수가 가장 높게 나타난 '동대문역사문화공원 → 을지로4가'를 대상으로 연구 방법의 실험 환경설정에 따라 모델학습을 진행하고자 하였다.



<그림 9 > 구간 max 승객수 그래프



<그림 10 > 구간 혼잡도 시각화

(2) 실험 환경설정

딥러닝 학습을 위해 승객수 데이터와 코로나 확진자 데이터, 기온, 강수 등 데이터를 통합 하였으며, 원활한 학습을 위해 데이 정규화 과정을 진행하였다.

	동대문역사문화공원-을지로4가	일일확진자	평균기온(℃)	강수량(mm)	평균풍속(m/s)	평균 상대습도(%)
0	3.0	0.0	16.7	0.0	2.4	58.0
1	2205.0	0.0	16.7	0.0	2.4	58.0
2	3805.0	0.0	16.7	0.0	2.4	58.0
3	5352.0	0.0	16.7	0.0	2.4	58.0
4	5887.0	0.0	16.7	0.0	2.4	58.0
...	...	...	...	...	...	...
10682	4030.0	11.0	22.5	2.9	4.6	74.3
10683	4644.0	11.0	22.5	2.9	4.6	74.3
10684	2685.0	11.0	22.5	2.9	4.6	74.3
10685	3269.0	11.0	22.5	2.9	4.6	74.3
10686	278.0	7.0	24.3	0.0	3.1	69.3

10687 rows × 6 columns

<그림 11 > 구간별 인원과 혼잡도 영향 변수들 예시

다음으로, lookback 과 delay의 값을 각 273, 39로 설정하였다. lookback은 입력층에 들어가는 데이터 셋의 크기로 모델이 그 크기만큼을 보며 학습을 진행한다. 이에 일주일 단위의 학습을 위해 하루당 39 row x 7일을 기준으로 하여 273으로 지정하였다. delay는 예측하고자하는 지점을 의미하며, 하루 뒤를 예측하기 위해 39로 설정하였다.

train_set과 validation_set과 test_set은 7 : 2 : 1의 비율로

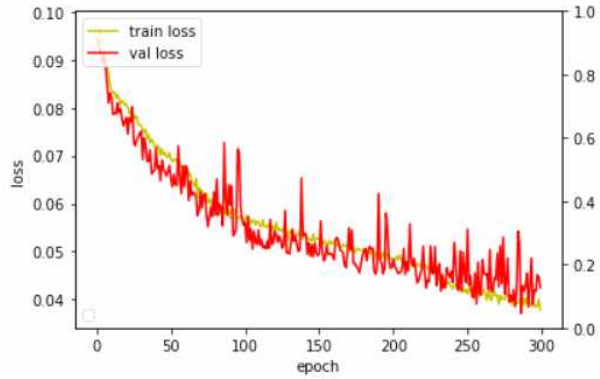
나누었으며 train_set으로 학습을 진행하고 validation_set으로 모델의 과적합 여부를 판단하고자 하였다. 또한 test_set으로 모델의 예측 성능을 확인하고자 하였다.

본 연구는 python을 기반으로 하였으며 keras 라이브러리를 활용하였다. 모델은 GRU를 사용하였고 activation 함수는 Relu를 사용하였으며, optimizer는 RMSprop를 사용하였다. 모델의 성능 평가 척도는 평균 절대 오차(MAE)로 확인하였으며, 오차의 최소화를 위해 각 파라미터를 수정하며 실험을 진행하였다. MAE의 공식은 다음과 같다.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - x| \qquad (1)$$

수식(1)의 의미는 모델의 예측값과 실제 데이터값 차이의 절댓값 합의 평균이다. MAE 결과값이 작을수록 실제 데이터값과 가깝게 모델이 예측함을 의미한다. batch_size는 한번에 학습되는 데이터의 개수로 512로 설정하였다. epoch는 데이터 전체를 학습한 횟수로 1,000회 실시하였며, early stopping을 적용하여 최적 결과를 도출하고자 하였다.

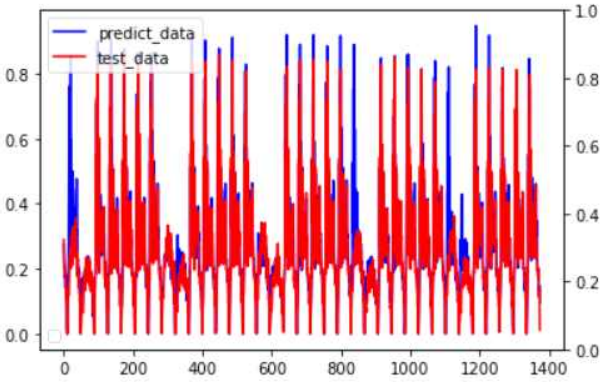
Ⅳ.연구결과



<그림 12 > 학습결과

학습 결과, train_loss와 val_loss가 비슷한 형태로 수렴하다가 epoch가 약 200부터 val_loss가 train_loss에 수렴하지 않으며, loss값이 더 이상 감소하지 않은 것을 확인하였다. 이에, keras의 내장 함수인 early stopping을 사용하여 모델이 train_set에 과적합 되지 않는 epoch을 확인 하고자 하였으며, val_loss가 30회 연속으로 이전 loss보다 떨어지지 않을 경우로 설정 했을 때 epoch=230에서 val_loss값 0.0431으로 오차가 가장 낮게 도출되어 해당 모델로 test를 진행하였다.

학습된 모델을 바탕으로 test_set을 입력시켜 예측한 결과값과 실제 데이터값을 시각화 했을 때 <그림 13>과 같이 나타낼 수 있다. 지하철 데이터 특성상 평일(5일)과 주말(2일)의 주기성을 모델이 예측할 수 있으며 출퇴근 시간에 혼잡도가 높아지는 현상까지 작은 오차로 예측할 수 있었다.



<그림 13 > GRU의 예측 결과

Ⅴ.결론

본 연구는 첨두시간대의 특정 구간의 혼잡도가 높아지면서 발생하는 지하철의 서비스 수준 하락 및 코로나 확산에 대한 위험성 상승 문제를 해결하기 위하여 첨두 시간에 혼잡한 지하철 구간 혼잡도를 예측하였다. 이를 위하여 서울 지하철 1~9호선의 승하차 인원 데이터를 가공하여 각 구간별 지하철 탑승자 수 데이터를 만들고 기상데이터와 코로나 확진자 현황을 반영하여 30분 단위로 혼잡도를 예측하는 방법으로 진행하였다. 가공한 데이터를 통해 '동대문역사문화공원 → 을지로4가''사당 → 방배' '낙성대 → 사당' '방배 → 서초' ' 한성대입구 → 혜화 5개 구간에서 첨두시간 혼잡도가 가장 높게 나타나는 것을 확인하였고 위 구간 중 '동대문역사문화공원 → 을지로4가'의 혼잡도를 30분 단위로 예측한 결과 구간 혼잡도 모델의 loss값은 0.0431로 도출되었다.

본 연구의 시사점은 지하철의 혼잡도를 GRU기법으로 기상상황 및 코로나 확진 현황을 반영하여 혼잡도를 예측하는 모델을 만들었다는 점에서 의의를 가진다. 이를 통해 혼잡도가 높은 특정 시간의 특정 구간을 예측해 방역 거점 조성 , 사전 고지 후 통행제한 , 배차 간격 조절 , 출퇴근 시간 조정, 구간 임시 급행화 등의 정책들을 제시할 수 있을 것으로 판단된다.

본 연구의 한계는 다음과 같다. 우선, 30분 단위의 승하차 데이터 이외의 데이터 부재와 개인정보 보호로 인한 데이터의 한계이다. 또한, 시간 및 재원의 한계로 인한 학습 데이터의 부족과 실제 이동경로가 아닌 최단경로를 이용한다는 가정에 의한 오차가 있다. 향후 연구에서는 이를 보완하고 환승경로의 다양성을 고려하여 우선 순위를 확률적으로 부여하고, 심층적으로 혼잡도에 영향을 끼치는 요소들을 고려한다면 보다 정확한 결과를 도출할 수 있을 것으로 사료된다.

참고문헌

1. 신성일. (2011). 대중교통카드를 활용한 도시철도 혼잡도 지표개발연구 (pp. 55-63). 서울연구원, Working Paper 2011-BR-04
2. 신성일, 이상준, & 이창훈. 스마트카드자료를 활용한 지하

철 승강장 동적 혼잡도 분석모형. 한국 ITS 학회논문지, 18(5), 49-63.
3. Leng, B., Zeng, J., Xiong, Z., Lv, W., & Wan, Y. (2013). Probability tree based passenger flow prediction and its application to the Beijing subway system. Frontiers of Computer Science, 7(2), 195-203.
4. 방준아, (2020), 지하철 이용객 개별 승하차 데이터를 통한 인구흐름 추정 및 시각화, 성균관대학교 대학원 석사학위 논문
5. 이경재, (2020) 딥러닝을 활용한 지하철 이용량 추정에 관한 연구, 홍익대학교 대학원 석사학위 논문.
6. 이정훈, & 정헌영. (2018). 분위 회귀를 활용한 기상조건이 대중교통 수단별 통행량에 미치는 영향에 대한 연구. 국토계획, 53(4), 95-106.
7. 김진, & 이민석. (2010). 지하철 이용수요와 역세권도시구조특성과의 관계분석연구: 수도권 역세권 지역을 중심으로. 대한건축학회 논문집-계획계, 26(10), 305-312.
8. 김재익. (2013). 아침 첨두시간대 지하철 이용수요의 결정요인에 관한 연구: 대구 지하철 역세권 토지이용을 중심으로. 교통연구, 20(1), 15-25.
9. 성현곤. (2017). 서울시 대중교통 수단별 월별 이용수요의 변동에 영향을 미치는 요인 분석.Journal of Korea Planning Association-Vol, 52(2), 81-97.
10. C.Olach, “Understanding LSTM networks,” 2015
11. J. Chung, C. Gulcehre, K. Cho and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” in Proc. 2014 NIPSWorkshop on Deep Learning, pp.1-9, Montreal, Canada, Dec. 2014