

지하철 혼잡도 개선을 위한 데이터 분석 및 시각화

Data Analysis and Visualization to Improve Subway Congestion

강정현

(건국대학교 인공지능학과, 석사과정)

김학수

(건국대학교 컴퓨터공학부, 교수)

Key Words : 빅 데이터, 데이터 시각화, 지하철 혼잡도

목 차

- I. 서론
- II. 데이터 분석 및 시각화
- III. 결론

I. 서론

지속적인 도시화와 인구 증가는 현대 도시의 교통 시스템에 대한 새로운 과제를 제기하고 있다. 수도권 인구의 계속적인 증가로 인해 대중교통 이용객도 함께 증가하고 있는데 특히, 지하철은 환경 친화적이면서도 많은 인원을 수용할 수 있는 교통수단으로 교통 체증도 줄여주는 역할도 하고 있다. 하지만 도시 인구의 증가와 더불어 지하철 이용객 수도 급격하게 늘어나면서, 이로 인한 혼잡도 증가는 지하철 운영 및 승객의 편의성 및 안전을 위협하고 있다. 또한 혼잡도가 높아질수록 원활한 지하철 운영이 어려워 지연과 시스템 고장의 위험성도 증가한다. 따라서 이러한 문제를 해결하기 위해 지하철 혼잡도에 대한 심층적인 연구와 분석이 필요하다.

본 논문에서는 지하철 혼잡도에 대한 분석을 위해 지하철 운행현황에 대해 조사하고, 지하철 이용객에 관한 데이터를 시각화하고 이를 분석한 결과를 제시한다.

4월에 업데이트된 데이터를 사용했고, 역별, 시간대별 승하차 인원에 대한 데이터는 2023년 7월 한 달간의 데이터를 사용했다. 또한 지하철 혼잡도가 가장 높은 출퇴근 시간에 대해 중점적으로 분석하기 위해 주로 평일 데이터를 사용했다.



<그림 1> 서울 지하철 운행 현황

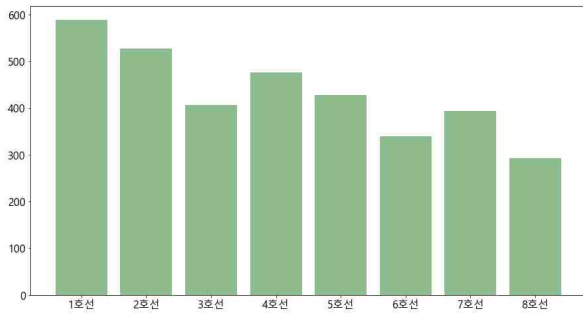
II. 데이터 분석 및 시각화

1. 지하철 운행 현황

지하철 혼잡도 분석을 위한 데이터는 서울교통공사에서 제공하는 서울 지하철 운행현황 데이터, 역별 시간대별 승하차 인원 정보 데이터 등에 대한 시각화 및 분석을 진행했다.

데이터는 최신 데이터를 사용하였는데 이는 도시 교통 시스템의 현재 상황을 잘 보여주기 때문이다. 본 논문에서는 지하철 운행현황은 가장 최신 데이터인 2023년

<그림 1>은 2023년 4월에 업데이트된 서울 지하철 운행 현황으로 각 호선 마다 편성된 열차 수와 운행거리를 나타낸다. 5호선에 편성된 열차 수가 가장 많고 그 뒤를 이어 2호선, 7호선의 열차 편성 수가 많음을 확인할 수 있고, 1호선의 열차 수가 가장 적는데 이는 서울에서 운행되는 1호선 운행 거리가 가장 짧기 때문인 것으로 예상된다.



〈그림 2〉 서울 지하철 각 호선 별 운행횟수(평일)

〈그림 2〉는 각 호선별 평일에 운행되고 있는 총 횟수를 나타낸다. 1호선이 현재 가장 많이 운행되고 있음을 알 수 있고 이어서 2호선, 4호선의 운행횟수가 많다는 것을 확인할 수 있다.

2. 지하철 혼잡도 분석

지하철 혼잡도 분석을 위해 지하철 역 별 승하차 인원 데이터가 아닌 혼잡도 데이터를 사용하였는데 지하철은 민자 노선과 같이 환승게이트가 따로 있는 경우를 제외하고 대부분 환승게이트 없이 환승하기 때문에 지하철간의 환승이 제대로 파악되지 않아 승하차 인원만으로는 혼잡도를 정확히 예측하기에 어려움이 따른다(김진수, 2016).

2023년 7월 역별, 시간대별 혼잡도 데이터에서 혼잡도가 높은 상위 10개를 추출하였다. 〈표1〉은 해당 역에 대한 정보를 정리하였다.

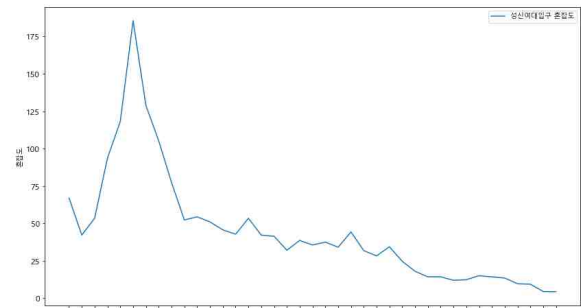
〈표1〉 역별 시간별 전체 혼잡도에서 상위 10개 추출

출발역	호선	상하구분	시간	혼잡도
성신여대입구	4	하선	08:00	185.5
한성대입구	4	하선	08:00	184.5
길음	4	하선	08:00	176.2
사당	2	외선	08:30	172.3
혜화	4	하선	08:00	171.5
방배	2	외선	08:30	164.6
미아삼거리	4	하선	08:00	161.9
어린이대공원	7	하선	08:00	160.6
중곡	7	하선	08:00	160.4
군자	7	하선	08:00	159.4

국토교통부내 도시철도 건설과 지원에 관한 기준에 대한 정보에 따르면 혼잡도가 150%를 넘는 경우 해당 노선의 증편을 요구하고 있다.

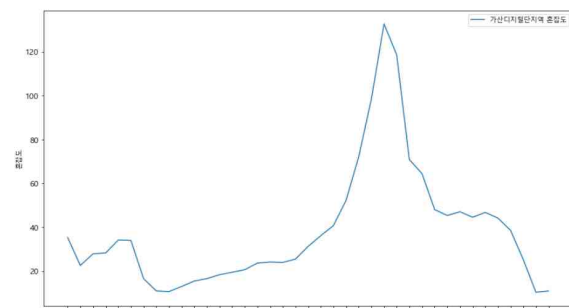
현재 출근 시간대에 혼잡도가 150이 넘는 역들이 존재하는 것을 확인할 수 있고, 〈표 1〉로부터 출근 시간대인 오전 8시의 지하철 혼잡도가 높은 것을 알 수 있다. 오전 8시의 평균 혼잡도가 60.167인데 반면 혼잡도가 가장 높은 성신여대입구역은 3배 가까이 높은 185.5로 상당히 높은 것을 알 수

있다.



〈그림 3〉 성신여대입구역(하선)의 시간대별 혼잡도

〈그림 3〉은 혼잡도가 가장 높았던 성신여대입구(하선)의 전체 시간대별 혼잡도를 시각화한 것이다. 출근 시간대에 혼잡도가 180을 넘길 만큼 높지만 퇴근 시간대에는 오히려 평균 혼잡도보다 낮은 혼잡도를 보이고 있다.



〈그림 4〉 가산디지털단지역(하선)의 시간대별 혼잡도

〈그림 4〉는 가산디지털 단지역의 전체 시간대별 혼잡도를 시각화한 자료이다. 해당 그래프는 〈그림 3〉의 그래프 양상과는 반대로 출근시간대의 혼잡도에 비해 퇴근시간대의 혼잡도가 월등하게 높다. 이는 가산디지털단지역 근처에 회사가 밀집해 있고 출퇴근에 이용객이 많은 방향이 달라 이러한 양상을 보이는 것으로 예상된다.

III. 결론

본 논문에서는 지하철 혼잡도 개선을 위한 연구에 도움이 되고자 가장 최신 서울 지하철 혼잡도 데이터를 이용해 시각화하고 결과를 분석하였다.

역별, 시간대별 혼잡도 상위 10개에 4호선에 있는 역이 절반을 차지하고 있는 것에 비해 현재 4호선의 운행횟수나 편성된 열차의 수가 그만큼 충분하지는 않다는 것을 확인할 수 있다.

또한 해당 역의 시간대별 혼잡도를 확인해 봤을 때 출근 시간대에 가장 높은 혼잡도를 보였던 성신여대입구역(하선)이 퇴근시간대에는 평균에 미치지 못하는 낮은 혼잡도를 보이고

있고 반면 출근 시간대에 낮은 혼잡도를 보이는 가산디지털단지(하선)이 퇴근시간대에는 아주 높은 혼잡도를 보이는 것을 확인할 수 있었다.

서사

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재 4.0사업의 연구결과로 수행되었음 (IITP-2023-2020-0-01834)

참고문헌

1. 김진수, “빅데이터 분석을 이용한 지하철 혼잡도예측 및 추천시스템”, 디지털융복합연구, 제14권, 제11호, 2016, 289-295.
2. 안재홍, 강인성 비용효과분석을 통한 서울시 지하철 9호선 혼잡도 개선방안에 관한 연구, 「한국공공관리학보」 제31권제2호(2017.6):105~12