

랜덤 포레스트를 이용한 대전시 공공 자전거('타슈') 수요 예측

민지원 문현수 이영석
충남대학교 컴퓨터공학과
miw5281@gmail.com, munhyunsu@cs-cnu.org, lee@cnu.ac.kr

Demand Forecast for Public Bicycles ('Tashu') in Daejeon using Random Forest

JiWon Min HyunSu Mun YoungSeok Lee
ChungNam National University Dept. Computer Science & Engineering

요 약

세계 주요 도시는 녹색대중교통 수단으로 무인 자전거 대여 시스템을 운영하고 있다. 대전시 또한 2012년부터 공공 자전거 시스템('타슈')을 운영하고 있으며 대중 교통 수요 분산의 효과를 이끌어 내고 있다. 본 연구는 자전거 재배치 작업의 효율성을 높이기 위해 기계 학습 방법으로 자전거 수요를 예측한다. 2013 ~ 2014 년도 자전거 대여 이력과 기상 정보, 축제 데이터를 수집하여 시민들의 자전거 이용 패턴을 파악하고 랜덤 포레스트 알고리즘에 수집한 데이터를 학습시켜 예측 모델을 구현하였다. 이를 기반으로 각 정류소마다 2015년도 시간 별 수요를 예측하였고 실제 수요와의 오차율을 구했다. 그 결과, 대부분 정류소의 예측 오차율이 최소 0.01에서 최대 0.04로 계산되었다. 이를 기반으로 대전시 무인 자전거 대여 시스템의 재배치 작업 효율에 활용할 수 있음을 보였다.

1. 서 론

세계 주요 도시는 대중 교통 수요 분산, 시민들의 운동량 증대를 목적으로 무인 자전거 대여 시스템을 도입하고 있다. 국내에서도 2008년 창원시 공공 자전거 시스템 '누비자'를 시작으로 전국적으로 확산되며 친환경 이동 수단의 역할을 해내고 있다. 대전광역시에서는 2012년부터 '타슈'라는 시스템을 운영하고 있다. 설문 조사와 무료 자전거 교실 등을 운영하며 활용도를 높이기 위한 노력을 하고 있다.

정류소의 시간별 수요를 예측하면 불균형 상황이 발생하기 전에 재배치 작업을 수행하여 원활한 운영을 도울 수 있다. 자전거 시스템이 확대될수록 재배치 작업의 규모도 커진다. 재배치 작업이 알맞은 시간에 이루어지지 않으면, 대여량이 많은 정류소에 자전거가 부족하거나 반납량이 많은 정류소에 반납할 자리가 없는 상태가 빚어진다. 즉각적인 대처가 투입되더라도 정상화될 때까지 시간이 걸리고 이용객들이 불편함을 겪는다. 각 정류소마다 재배치 작업 시점을 확보하여 작업 경로를 구하면 업무 효율성을 높이고 시민들의 편리한 이용을 도울 수 있다.

본 연구¹⁾는 기계 학습을 통해 각 정류소의 시간 별 수요 예측을 목표로 한다. 자전거 수요에 영향을 주는 요인으로 시간, 기상, 축

제 여부를 선정하였다. 2013 ~ 2015 년도의 타슈 대여 이력과 기상 데이터, 축제 데이터를 수집하였다. 정류소 마다 시간 별로 기온, 습도, 축제 여부 등에 따른 대여량을 나타낸 데이터 셋을 생성하였다. 2013 ~ 2014 년도 데이터를 학습시킨 랜덤 포레스트 예측 모델을 만들고 2015년도의 시간 별 수요를 예측하였다. 실제 수요와 비교하여 정확도를 확인한 결과, 평균 0.015592의 오차율을 확인하였다.

2. 관련 연구

해외의 경우, 항저우[1], 뉴욕, 워싱턴[2] 등에서 자전거 사용량 예측 연구가 진행된 바 있다. 각 연구마다 GBRT, Random Forest 등 다양한 기계 학습 알고리즘을 적용하여 사용량을 예측하는 모델을 제시하였다.

국내의 경우 자전거 수요를 예측한 연구 사례는 드물다. 2016년도에 타슈 데이터 분석 및 시각화 연구[3]가 진행되었다. 2013 ~ 2014 년도 데이터를 분석하여 사용량이 많은 정류소, 경로 등의 통계를 구하고 각 정류소의 위치적 특징에 따라 시민들의 사용 목적을 파악하였다. 위치, 시간, 기후 특성에 따른 자전거 사용량 변화를 시각화한 시스템을 구현하였다.

고양시에서 날씨 영향에 따른 자전거 사용량 변화를 나타낸 연구가 진행되었다[4]. 사용량에 영향을 주는 기온, 강수량의 기준선과 영향력을 수치화하여 구체적으로 나타내었다.

1) 이 논문은 2016년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임
(NRF-2016R1D1A1A09916326)

3. 자전거 수요량 예측

3.1 데이터

본 논문에서는 정류소의 시간 별 수요를 알아내기 위해 공공 데이터 포털에 공개된 2013 ~ 2015년 타슈 대여 이력을 사용한다. 2013년 1월 1일부터 2015년 12월 31일까지 자전거가 이용된 대여 장소, 반납 장소, 대여 시간, 반납 시간이 포함된 3,397,486 건의 데이터가 공개되어 있다. 그 중 오류가 있는 데이터를 제외하고 3,350,385 건의 대여 이력을 활용하였다.

예측을 위해 해당 기간 기상 데이터와 축제 데이터가 필요할 것으로 판단하였다. 기상청 국가기후 자료에서 제공하는 대전시 기상 정보를 활용하였고 대전 광역시청 홈페이지에서 안내하는 축제 데이터를 활용하였다.

본 연구에서는 2015년에 신설된 정류소를 제외한 144개 정류소만을 대상으로 수요를 예측한다.

	2013년	2014년	2015년
총 대여 수	1,033,827	1,195,795	1,120,763
하루 평균 대여 수	2,832	3,276	3,070
평균 대여 시간	114분	129분	113분

표 2 타슈의 2013, 2014, 2015 년 기초 통계

3.2 예측 및 정확도

본 논문에서는 R에서 기계 학습 라이브러리를 이용하여 자전거 사용량을 예측한다. 사용량에 영향을 주는 요인으로 시간, 기후, 축제/행사 여부를 선정하였다.

예측 요인	요소
시간	월, 시간, 요일, 계절
기후	기온, 강수량, 습도
축제 여부	축제 장소 근처의 정류소에 한함

표 3 예측 요인

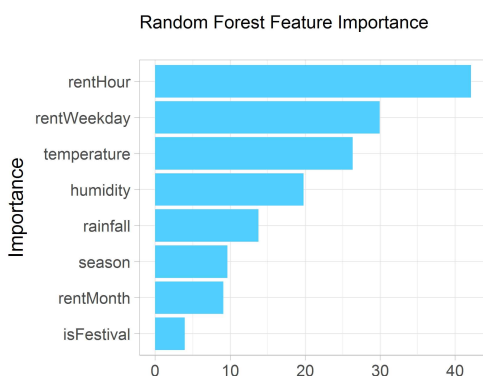


그림 1 3번 정류소 예측 요인 별 중요도

그림 1은 가장 사용량이 많은 3번 정류소(한밭 수목원)의 예측 요인별 중요도를 나타낸 그래프이다. 공원 주변에 위치한 정류소이기 때문에, 대부분 주말에 여가 활동을 목적으로 대여가 이루어진다. 시간과 요일이 수요에 가장 영향을 주는 요인이다. 날씨가 좋지 않으면 공원 이용객이 적은 것을 예상할 수 있듯이, 기상 요인 또한 정류소 수요에 영향을 준다.

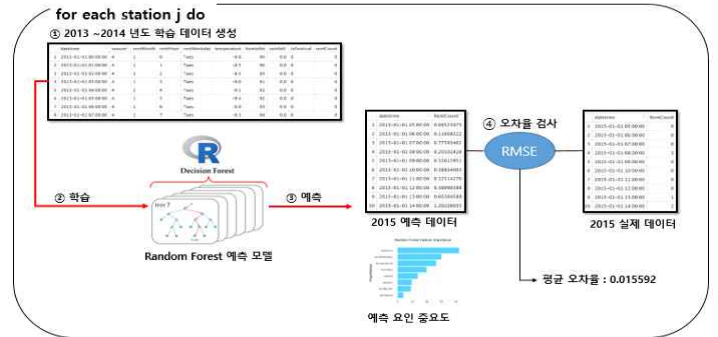


그림 2 예측 과정

그림 2는 예측 과정을 보여준다. 본 연구에서는 2013 ~ 2014 년도 데이터를 기계 학습 데이터로 활용하고 2015년도 자전거 수요를 예측하여 실제 수요와 비교하는 과정을 거친다.

- 2013 ~ 2014 년도 타슈 대여 이력 데이터를 통해 각 정류소마다 한 시간 단위로 몇 대의 자전거가 대여되었는지 나타낸 데이터 셋을 구성한다. 데이터 셋에 부가적으로 계절, 시간, 요일, 기온, 습도, 강수량, 근처 축제 여부 정보를 추가한다.
- 구성된 데이터 셋을 랜덤 포레스트 알고리즘에 학습시켜 예측 모델을 구성한다.
- 예측 모델에 2015년도 예측 요인 정보를 입력하여 수요를 예측한다. 각 정류소마다 수요에 가장 크게 영향을 주는 요인을 확인한다.
- 2015년도 실제 수요와 예측 수요를 비교한다. RMSE(Root Mean Square Error) 방법을 이용하여 평균 오차율을 구한다.

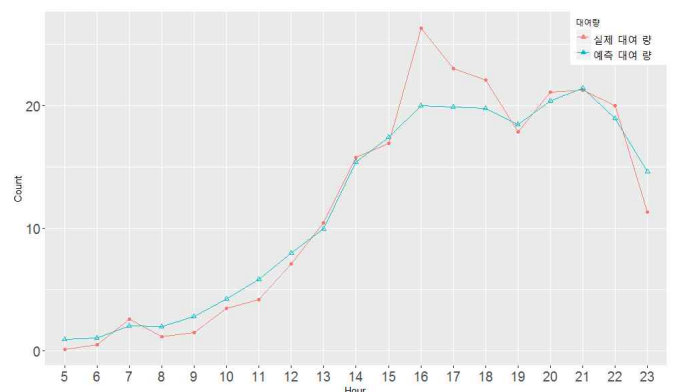


그림 3 3번 정류소 주말 대여량

그림 3은 3번 정류소의 2015년도 주말 대여량을 나타낸 그래프이다. 실제값과 예측값 모두 주말 오후에 대여량이 급증하는 패턴을 확인할 수 있다. 평균 제곱근 오차 방식으로 구한 평균 대여 오차율은 0.012로 나타났다.

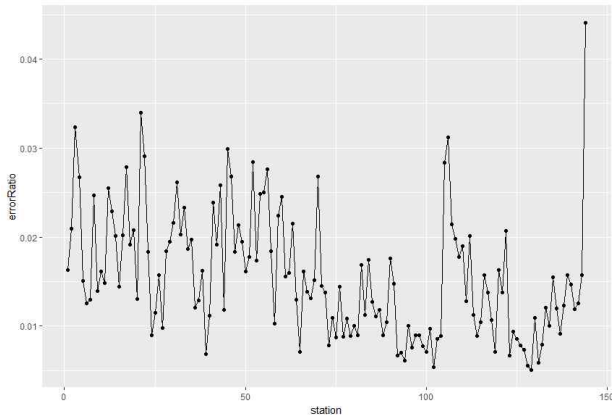


그림 4 정류소 별 예측 오차율

그림 4는 1 ~ 144 정류소의 예측 오차율을 의미한다. 144번 정류소(‘타슈 관제센터’)에서 0.044로 가장 큰 오차율을 보였고 131번 정류소(‘예코 공원앞’)에서 0.005로 가장 적은 오차율을 보였다. 평균 오차율은 0.015592로 나타났다.

3.3 재배치

정류소마다 대여 혹은 반납이 집중되는 시간이 있다. 이 때 재배치 작업이 이행되지 않으면, 대여가 많이 이루어질 때 자전거가 부족하거나 반납이 많이 이루어질 때 반납할 자리가 없는 상황이 빚어진다. 이러한 상황이 발생하면 정상화될 때까지 시간이 걸리고 이용객들이 불편함을 겪을 수 있다. 매 시간마다 수요를 예측함으로써 모든 정류소의 재배치 시각을 확보하는 방안을 제시한다.

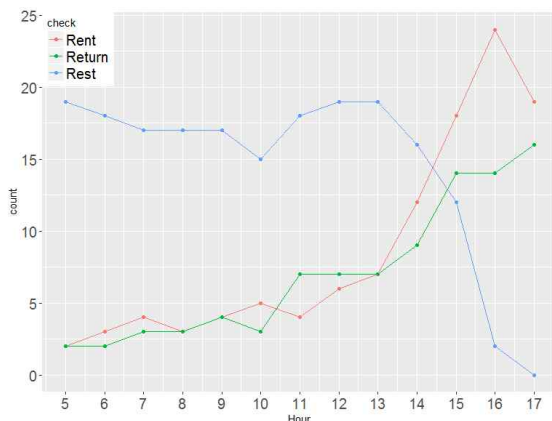


그림 5 자전거 수 변화

3번 정류소의 자전거 거치대는 총 19개이다. 타슈 시스템이 운영을 시작하는 오전 5시 자전거 점유율이 100%임을 가정한 후, 잔여 자전거 수가 0이 되는 시간대를 확보하여 재배치 작업 시간을 결정한다. 가장 사용량이 많은 5~6월, 잔여 자전거 수가 0이 되는 경우는 수요가 많아지는 저녁 시간대(오후 6시 ~ 오후 11시)에 나타났다. 해당 시간에 점유율이 높은 것으로 나타난 정류소들을 파악하여 균형을 맞추는 방안을 고려할 수 있다.

4. 결론

본 연구는 대전시 공공 자전거(‘타슈’)의 사용량을 예측하여 관리 인력의 업무 효율성을 높이는 것을 목표로 한다. 최근 몇 년 동안의 자전거 대여 이력을 이용하여 시민들의 자전거 이용 현황을 파악하였다. 자전거 사용에 영향을 주는 요인으로 시간, 기상, 축제 여부로 결정하였고 수집한 데이터를 기반으로 정류소 별 사용량을 예측하였다.

전반적으로 예측에 큰 영향을 준 요인은 ‘시간’이었다. 실제 자료와 마찬가지로 예측 결과에서도 주거, 회사 주변의 정류소에 선 출퇴근 시간에 사용량이 급증하는 양상을 보였다. 기온과 습도가 그 뒤를 이었고, 축제 여부는 생각보다 큰 영향 요인이 되지 않았다.

타슈 시스템 관리 인력은 예측 모델을 활용하여 재배치 작업이 이루어져야할 시점을 구할 수 있다. 또한 자전거 점유율이 부족한 정류소로 이동하는 시민들에겐 혜택을 주는 것을 명시하여 시민들의 자발적인 참여를 이끌어 내는 방안도 고려할 수 있다.

5. 참고 문헌

- [1] Zidong Yang, Ji Hu, Yuanchao Shu, Peng Cheng, Jiming Chen, Thomas Moscibroda, “Mobility Modeling and Prediction in Bike-Sharing System”, MobiSys’16 (2016)
- [2] Yexin Li, Yu Zheng, Huichu Zhang, Lei Chen “Traffic Prediction in a Bike-Sharing System”, SIGSPATIAL’15, (2015)
- [3] 문현수, 이영석 “대전시 공공 자전거(타슈) 공개 데이터 시각화 및 분석”, KIISE Transactions on Computing Practices 22(6), 253-267, (2016)
- [4] 김동준, 신희철, 반준식, 임형준, “날씨가 자전거 이용에 미치는 영향 분석”, 교통연구 제19권 제3호 77-88, (2012)