Journal of Knowledge Information Technology and Systems(JKITS), Vol. 14, No. 6, pp. 615~623, December 2019
https://doi.org/10.34163/jkits.2019.14.6.004

Journal of Knowledge Information Technology and Systems
ISSN 1975-7700
http://www.kkits.or.kr

Anaysis for Public Bike Utilization Status of Seoul City

Jangwoo Park*, Young-Yun Cho
Dept. of Information and Communication, Sunchon NationalUniversity

ABSTRACT

Bicycles have played a great role not only as eco-friendly transportation, but also as exercise equipments for leisure and health. Cities in many countries offer public bicycle sharing services to help solve the health and traffic jams of citizens and to solve environmental problems caused by automobiles. To serve the efficient bike sharing, the knowledge of the status of bike use will be necessary. We have analyzed the data set of public bicycle service in Seoul to determine the temporal and spatial characteristics of bicycle use status in Seoul. The travel time and the latitude and longitude of the bike rental place and return places were transformed into 50 clusters using K-means clustering method. Spartial and temporal characteristics have been obtained and visualized. We can know the bicycle retal and return status according to the location of the City of Seoul. In addition, these clusters were classified and visualized for year, such as date, day of the week, time, etc. From the analysis, we could show the bike use patterns of weekday and weekend. During the weekday, the bikes are being guessed to be used for commute devices and for working assistance. On weekends, bicycles are slowly increasing from morning to high, peaking around 6pm. It is confirmed that during the week, there are many short-distance uses, and on weekends various travel distance patterns were seen.

KEYWORDS: PCA, Kmeans, Data Science, Haversine distance, Public bike service, Kaggle

ARTICLE INFO: Received 8 October 2019, Revised 30 October 2019, Accepted 7 December 2019.

*Corresponding author is with the Dept of Information and Communication Engineering, Sunchon National University, 255 Jungang-ro, Sunchon, Jeollanamdo, KOREA.
E-mail address: jwpark@sunchon.ac.kr

1. 서 론

자전거는 친환경 교통수단으로써 뿐만 아니라 레저 및 건강을 위한 운동 장비로서도 훌륭한 역할을 하고 있다. 여러 나라의 여러 도시들에서 시민의 건강과 교통 정체를 해소하고 자동차 등으로 인한 환경 문제를 해결하기 위하여 공공 자전거 서비스를 실시하고 있다[1-3]. 우리나라에서도 서울시를 비롯하여 여러 지자체들에서 자전거 공유 서비스를 제공한다[4,5]. 서울시는 1500곳 넘는 자전거 대여소를 서울시 전역에 설치하여 운영하고 있고 이로 부터 발생하는 데이터를 제공하고 있다[6].

서울시에서 발생하는 각종 자료를 열린 데이터 광장[7]을 통하여 일반인들에게 공개하고 있다. 다양한 자료들을 여러 가지 형태로 제공하고 있고 다양한 형태로 사용될 수 있다. 본 연구에서 공공 자전거 대여 서비스를 이용하여 서울시의 자전거 이용 실태를 분석해 보려고 한다. 자전거 이력 데이터 셋을 분석하면 서울시에서 자전거를 이용하는 주요한 시간, 자전거의 주요 이용 형태 등을 살펴 볼 수 있다. 또한 간접적으로 서울시만의 생활 형태도 예측해 볼 수 있다.

Kaggle[8]은 데이터 사이언스를 위한 다양한 데이터를 제공하고, 데이터를 활용하는 여러 연구자들의 결과를 커널 형태로 제공하고 있다. Kaggle에서 실시하고 있는 여러 데이터 분석 경연은 새로운 데이터 분석의 방법을 익히는데 도움이 된다. 또한 다양한 주제의 데이터들과 분석 방법이 제공되기 때문에 이를 활용하여 새로운 데이터 분석에 사용할 수 있다. “New York City Taxi Trip Duration” [9-11]은 택시를 타는 위치, 내리는 위치, 승객의 수 등의 데이터를 제공하고 이들 자료를 이용하여 이용 거리를 예측하는 것이다. 이들 자료를 분석하여 뉴욕 시의 생활 패턴을 간접적으로 분석해 볼 수 있다. “New York City Taxi Fare Prediction” [12]는 택시를 타는 위치, 내리는 위치, 승객의 수 등을 이용하여 택시 운임을 예측한다. “Bike Sharing Demand” [13]는 날짜, 기상 정보를 바탕으로 자전거의 수요를 예측하고 있다.

서울시 공공 자전거의 일반적인 분석은 거치대 분포, 이용건수, 대여 및 반납 분석 등에 대하여 이루어져 있다[14,15]. 본 연구에서는 서울시의 공공자전거의 일반적인 현황을 알아보고[14,15], 자전거 이용에 대한 시간적, 공간적 특성을 파악하고자 한다. 자전거 대여소의 위치(위도와 경도), 이용시간, 이동 거리 등을 바탕으로 Kmeans 클러스터링 방법[16]을 사용하여 그룹화 하였다. 각 클러스터들의 나름대로의 특징과 의미를 갖고 분류 되었다. 각 클러스터 들은 이용거리, 이용시간, 그리고 공간적 특징에 따라서 분류된다. 공통적인 특징을 갖는 클러스터를 이용하면 이용 현황을 시각화하는데 도움이 된다. 다음으로는 클러스터들을 시간, 요일, 날짜 등에 따라서 재 정렬하고 사각화하여 자전거의 이용 실태를 분석한다. 또한, 클러스터들의 차원을 PCA[16]를 이용하여 낮추고 다시 시간별 요일별 특성을 살펴보았다.

논문의 구성은 다음과 같다. 2절에서는 공공 자전거 데이터에서 얻은 자전거 이용 현황을 살펴본다. 3절에서 데이터 셋에 있는 이상 데이터를 살펴보고 제거 방법에 대하여 설명한다. 4절은 데이터 셋을 분석하여 얻은 자전거 운행의 공간적 이동 패턴을 살펴보고 5절 에서는 시간적인 이동 현황을 살펴본다. 끝으로 6절에서 끝마친다.

2. 2018년 공공 자전거 이용 현황

연구에 사용되는 2018년도 자전거 이력 데이터는 여덟 개로 나뉘어 있고 파일의 형태도 csv 와 xlsx 로 두 가지로 되어 있다. 각 파일은 “자전거번호“, “대여일시“, “대여소번호“, “대여소명“, “대여

거치대“, 반납일시“, “반납대여소번호“, “반납대여소명“, “반납거치대“, “이용시간(분)“, 이용거리(M)“ 등 총 11개의 항목으로 되어 있다. 이들 8개의 파일을 모두 합치는 경우 약 1000만 샘플 이상이 된다. 우선 이들 파일을 합쳐서 하나의 파일을 만들었다. 자전거 이력 데이터셋은 자전거 대여소를 아이디 형식으로 갖고 있다. 대여 시각과 반납 시각은 월, 일, 시, 분 등으로 분리하여 시계열 데이터 처리를 위한 준비를 하였다. 준비한 데이터 셋은 위도, 경도와 같은 공간적 데이터와 월, 일, 시, 분 등의 시간적 데이터로 구성되었다.

자전거 대여소의 위치 정보는 “자전거 대여소 정보“에서 얻었다. 자전거 대여소 정보는 “구분“, “대여소번호“, “대여소명“, “위도“, “경도“ “거치대수“로 이루어져 있다. 이 파일에서 “구분“은 자전거 대여소가 위치한 ’구’를 의미한다. 대여소 번호는 정수로 되어 있다. 2019년 5월 현재 총 대여소의 수는 1538개이다. 대여소의 지역별 분포는 <그림 1>와 같다. 강남구에 자전거 대여소가 가장 많고 그 다음이 송파구이다. 가장 적은 곳은 도봉구 강북구, 용산구 순이다.

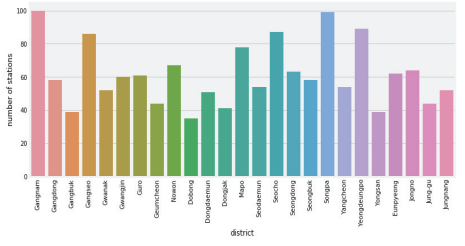


그림 1. 구별 자전거 대여소 현황(2019년 5월)
Figure 1. Location counts of districts in Seoul

자전거 대여 대여소의 위치와 반납 대여소를 분석하여 얻은 자전거 대여 현황을 <그림 2>에 나타내었다. 붉은 색으로 표시한 곳은 자전거 대여 및 반납이 활발한 곳이다. 자전거 많이 대여되고 있는 곳은 여의나루역, 뚝섬 유원지, 홍대입구, 롯데 월드 타워 부근 등이다.

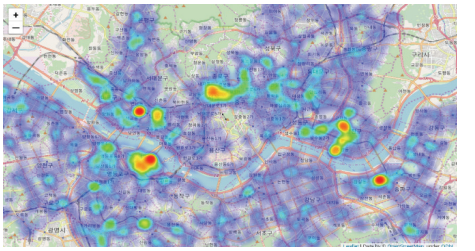


그림 2 자전거 대여 현황(대여가 활발할수록 붉은 색에 가깝다)
Figure 2. Bike rental-return status

데이터 샘플들 중에서 이동거리가 100m 이하이거나 운행 시간이 0분인 데이터들은 분석에서 제거하였다. 한편 운행 시간이 지나치게 길거나 이동거리가 지나치게 긴 경우도 분석에서 제외하였다. 이상 샘플을 제거한 후 데이터의 분포를 <그림 3>에 나타내었다. 이동거리는 거리의 로그 값을 취하여 계산하였다. 이동거리의 로그 값은 정규 분포에 가깝다. 이용 시간은 일반적으로 30분 이내이다.

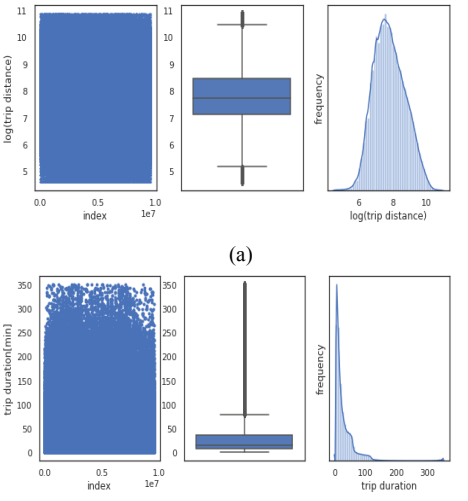
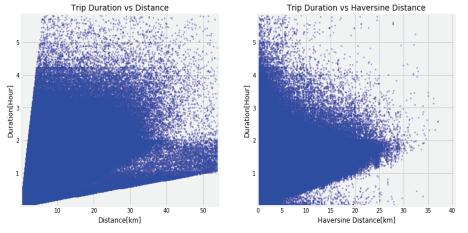


그림 3 이동거리(a)와 이용시간(b)의 분포
Figure 3. The distribution of (a) bike riding duration and (b) bike riding distance

데이터 중에서 이상 데이터를 알아보기 위하여 이동거리와 이동 시간을 이용하여 속력을 계산하였다. 계산한 속력은 km/hour로 하였는데 매우 큰 값이 얻어지는 경우들이 있다. 이렇게 큰 값들은 올바른 측정 값일 수 없다. 시속 50km 이상인 경우와 시속 1km 이하인 데이터 들은 제거하였다.



(a) (b)

그림 4. 이동 거리와 대여시간의 관계 (a) 기록 거리 (b) Haversine 거리

Figure 4. Riding distance and riding duration (a)recored distance (b) Haversine distance

데이터 셋에 주어진 자전거 이용거리와 자전거 대여소의 위도 경도를 이용하여 계산한 거리를 비교해 보았다. 위도 경도를 이용한 거리 계산은 Haversine 공식을 이용하였다. Haversine 공식[17]을 이용하여 계산한 거리에는 약 10% 정도의 영 값이 있다. 이것은 대여 장소와 반납 장소가 같은 것을 의미한다. 그림 4(b)에서 보듯이 Haversine 거리가 영인 경우에도 실제의 대여시간은 상당히 긴 경우를 볼 수 있다. 그에 비해 그림 4(a)의 그림은 실제의 측정 거리와 대여시간을 그린 것이다. 이용시간과 이동거리는 비례관계를 보이지 않는다.

Haversine 거리는 유클리드 거리를 지구의 구면을 고려한 것이다. 따라서 매우 넓은 지역이 아니면 유클리드 거리나 Haversine 거리는 큰 차이를 보이지 않는다. Haversine 거리가 영 근처 값을 보임에도 불구하고 실제 기록된 거리는 매우 큰 경우들이 있다. 즉, 자전거 대여한 곳과 반납한 곳이 같거나 근처에 있음을 의미한다. 한편 측정된 거리가 Haversine 거리에 비하여 작은 경우가 전체 샘플 중 10% 정도가 된다. 이들은 측정 오차 혹은 기록 오차라고 생각한다.

3. 이동 패턴 해석

연구에서 사용한 방법은 Kmeans 클러스터링 방법과 PCA(Principal Component Analysis)이다[9, 10, 16]. Kmeans 클러스터링 알고리즘은[18] 클러스터링 방법 중 분할법에 속한다. 분할법은 주어진 데이터를 여러 파티션 (그룹) 으로 나누는 방법이다. 데이터를 한 개 이상의 데이터 오브젝트로 구성된 k개의 그룹으로 나누는 것이다. 이 때 그룹을 나누는 과정은 거리 기반의 그룹간 비유사도 (dissimilarity) 와 같은 비용 함수 (cost function) 을 최소화하는 방식으로 이루어진다. KMeans 알고리즘은 각 그룹의 중심 (centroid)과 그룹 내의 데이터 오브젝트와의 거리의 제곱합을 비용 함수로 정하고, 이 함수값을 최소화하는 방향으로 각 데이터 오브젝트의 소속 그룹을 업데이트 해 줌으로써 클러스터링을 수행하게 된다.

PCA(Principal component analysis)[19]은 고차원의 데이터를 저차원의 데이터로 환원시키는 기법이다. 서로 연관 가능성이 있는 고차원 공간의 표본들을 선형 연관성이 없는 저차원 공간(주성분)의 표본으로 변환하기 위해 직교 변환을 사용한다. 주성분의 차원수는 원래 표본의 차원수보다 작거나 같다. 주성분 분석은 데이터를 한개의 축으로 사상시켰을 때 그 분산이 가장 커지는 축을 첫 번째 주성분, 두 번째로 커지는 축을 두 번째 주성분으로 놓이도록 새로운 좌표계로 데이터를 선형 변환한다. 이와 같이 표본의 차이를 가장 잘 나타내는 성분들로 분해함으로써 여러 가지 응용이 가능하다.

자전거를 빌린 대여소의 위도, 경도와 반납한 대여소의 위도, 경도를 자전거의 운행 시간, 운행 거

리를 포함하여 Kmeans 클러스터링 방법을 이용하여 50개의 클러스터로 분류하였다. 각 클러스터는 위치 정보와 이용시간, 이용거리의 특성에 따라서 분류되었고 이들 클러스터의 의미 및 특징을 파악하였다. 다음으로 자전거 이용특성의 공간적인 특징인 클러스터들을 시간(월, 요일, 일, 시간)등에 따라서 정렬하고 재배치하였고 그 결과들을 이용하여 시간에 따른 여러 가지 특징들을 클러스터와 함께 시각화 하였다. 시각화 한 이미지는 공간의 특성과 시간의 특성을 잘 표시할 수 있다. <그림 5>는 클러스터의 색인과 각 클러스터에 포함된 샘플의 수를 나타낸 것이다.

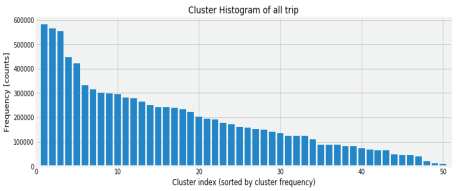


그림 5. 클러스터 색인과 빈도수
Figure 5. frequencies of clusters

클러스터 중 22개는 이동거리가 5km이하이고, 이동시간도 30분 이하이다. 클러스터들은 대여소의 위치에 따라서 그룹화 된 경향이 있다. 한편 이동시간이 매우 길거나(19, 47 클러스트) 혹은 이동거리가 매우 긴(14번) 경우 들은 서울시 전역에 걸쳐서 분포하고 있다. 가장 빈도수가 높은 세 개의 클러스터들은 평균 운행시간이 15분 내외이고 이동거리는 1.5km 정도이다. 상당히 가까운 거리를 이동하는 것이 특징이다.

자전거 이용에 있어서 시간은 중요한 요인이고 시간에 따른 패턴을 분석하는 것은 중요하다. 자전거는 단지 레저를 위해서 사용되는 것이 아님을 분석으로 부터 알 수 있다. 앞에서 분류한 클러스터는 50차원을 갖기 때문에 시각화를 용이하도록 PCA를 이용하여 3차원으로 낮춘다[9]. 3차원 PCA 계수들은 공간적 정보를 포함하고 있다. 이들 결과를 요일, 하루 중 시간, 연간 날짜 등 시간별로 분리하였다. 요일에 따른 이용 패턴, 시간에 대한 이용패턴, 1년간의 전체 이용 모습을 살펴본다.

요일 별 자전거 이용 현황을 시각화 하였다. <그림 6>은 앞서 구한 클러스터들에 대하여 각 요일의 이용 횟수를 그림으로 표시하였다. 노란색에 가까울수록 자전거 이용이 활발한 것을 나타낸다. 클러스터의 번호가 작을수록 가까운 거리를 짧은 시간 동안 이용한 것이다. 이동거리가 짧고 이용시간이 작은 경우는 월요일에서 금요일까지에 집중되어 있다. 이것은 대체로 출퇴근을 비롯한 업무에 이용하는 경우라고 생각할 수 있다. 한편 클러스터의 번호가 높은 중, 장거리 혹은 긴 시간 이용하는 경우에는 요일에 따른 특성이 보이지 않는다.

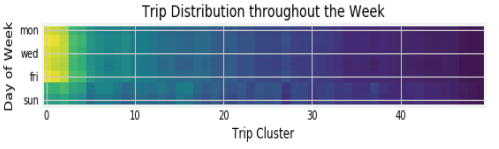


그림 6. 요일 별 이용 현황
Figure 6. The bike rental status according to day of week

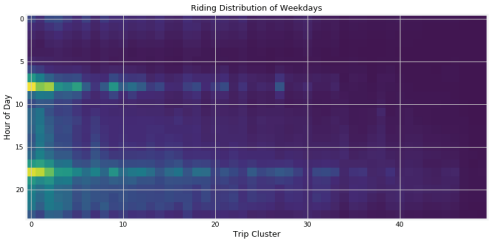


그림 7. 주중 시간에 따른 대여와 반납 횟수
Figure 7 rental and return frequencies of weekday

시간별 자전거 이용을 분석하는 것은 자전거의 이용 특징을 이해하는데 있어서 중요한다. 시간별 자전거 이용 패턴을 주중과 주말로 나누어 살펴본다. <그림 7>은 월요일에서 금요일까지 시간에 따

른 이용 패턴을 그렸다. 그림에서 밝은 노란 색에 가까울수록 이용 횟수가 높은 것이다. 클러스터의 번호의 의미는 앞에서 설명한 것과 같다. 출근과 퇴근 시간의 이용이 활발하다는 것을 알 수 있다. 짧은 거리 이용자들의 경우 출근시간(8시경)과 퇴근시간(18시경)에 특히 높은 이용 횟수를 보이고 있다. 출근시간은 한 시간 정도에 집중되어 있는 것에 비하여 퇴근 시간은 좀 더 넓게 분포되어 있다. 이러한 출퇴근 시간에 자전거를 이용하는 패턴은 넓은 클러스터 범위에 걸쳐서 있다. 물론 숫자가 줄어들기는 하지만 장거리 출퇴근에 자전거를 이용하는 것으로 생각할 수 있다. 이러한 시간에 따른 이용 특성은 단순한 대여 및 반납 횟수에서도 확인할 수 있다.

주말의 이용 형태는 주중과 다르다. <그림 8>은 토요일, 일요일의 이용 모습을 나타낸다. 주말의 경우는 출근-퇴근 시간에서 피크인 특징이 보이지 않는다. 오전 8시 경부터 서서히 이용이 증가하고 오후에 적극적인 이용 패턴을 보인다. 이와 같은 모습은 전체 클러스터에서 거의 비슷하게 보인다. 즉, 많은 이용자들이 주말에 장거리를 이용하는 운동 및 레저에 자전거를 이용하고 있다고 생각된다.

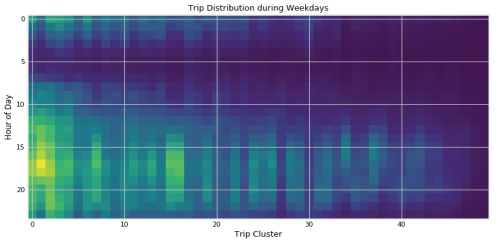


그림 8. 주말의 시간에 따른 자전거 이용
Figure 8. bike retal status of weekend

2018년 1년간의 자전거 이용 형태 <그림 9>에 표시하였다. 앞의 그림들과 달리 y-축이 클러스터이며 x-축은 1월 1일 1부터 계산한 1년간의 날자이다. 겨울에는 자전거 이용이 저조하고 봄과 가을에 이용 늘고 있음을 확인할 수 있다. 여름철은 봄이나 가을 보다는 활동이 작아지는 것도 볼 수 있다. 봄과 가을이 되면 오랜 시간 장거리 활동도 늘고 있다. 또한 중간에 활동이 매우 저조한 날들이 끼어 있다 아마도 기상 상황이 좋지 않은 영향이 크다고 생각한다.

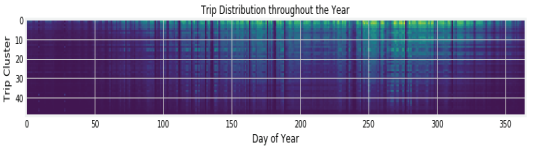


그림 9. 2018년 일년간 이용 패턴
Figure 9. The bike retal patten during a whole year of 2018

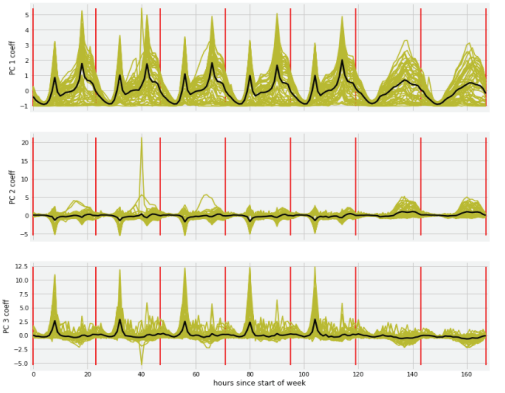


그림 10. PCA 계수 주간 평균
Figure 10. CA coefficients of day of week

PCA 계수를 1주일간의 평균값으로 다시 정리하였다. <그림 10>는 PCA 세 개의 계수를 그림으로 표현하였다. PCA 의 첫 번째 계수는 주중에 출, 퇴근 시간의 이용 현황을 잘 묘사하고 있다. 주중 패턴은 거의 비슷하게 출근 시간과 퇴근 시간에 계수가 값이 커진다. 주말은 주중과는 많이 다른 형태를 보이며 오후로 갈수록 점점 더 자전거 이용량이 높아진다. PCA를 이용하여 차원을 축소하여

그림으로 표시함으로써 주중의 시간에 따른 주기적인 특징을 알 수 있으며 주중과 주말의 차이를 명확히 알 수 있다.

끝으로 자전거의 속력을 살펴보았다. 〈그림 11〉는 하루 중 시간별, 요일별, 월별 평균 속력을 그린 것이다. 첫 번째 그림에서 보듯이 자전거의 평균 속력은 아침 출근 시간에 가장 빠르다. 한편 요일별로 보면 주중이 주말보다 빠른 것을 알 수 있다. 확실히 주말이 시간적인 여유가 있기 때문인 듯하다. 그리고 월별로 보면 겨울철이 가장 빠르게 달리고 있으며 봄과 가을은 속력이 느린 것을 확인할 수 있다.

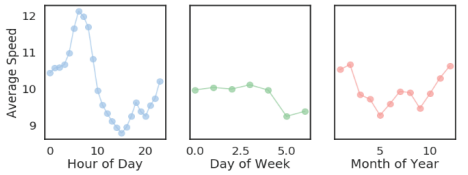


그림 11. 자전거 속도
Figure 11. the speed of bike

4. 결 론

우리나라는 여러 지방 자치 단체들에서 공공 자전거 대여 서비스를 제공하고 있다. 자전거는 이동 수단뿐만 아니라 레저 및 운동 등 다양한 용도로 이용되고 있다. 특히, 친환경 이동 수단으로 인식되고 있어서 이용이 늘어나고 있다. 본 연구에서는 서울시의 공공 데이터 서비스에서 제공하고 있는 자전거 이력 데이터 셋을 이용하여 서울시 자전거 이용특성을 분석한다. 주말과 주중의 특성이 명확히 구분됨을 알 수 있다. 자전거의 이용 특성 분석을 통하여 자전거가 단순히 레저에 이용되는 것이 아니라 주중에는 출퇴근을 비롯한 업무에 이용하고 있음을 알 수 있다. 출근 시간에 비하여 퇴근 시간에 이용이 많다.

분석 방법은 자전거 대여소의 위치 정보와 이동거리, 이용시간을 KMeans 크러스터링 기법으로 유사 그룹으로 나누고 클러스터들을 이용하여 시간에 따른 특성을 분석하였다. 특히, 요일, 시간에 따른 특성을 살펴보았고 2018년 전체의 특징도 살펴보았다. 또한 PCA를 이용하여 클러스터의 차원을 줄여서 클러스터의 특징을 살펴보았다.

자전거 이용 시간과 이동 거리는 비례관계에 있지 않다. 또한 이동시간, 이동거리는 자전거 이용 형태 분석에서 중요하다. 자전거의 이용 특성은 특히 온도, 습도, 미세 먼지 등 여러 환경적 용인에 영향을 받을 것이라고 생각한다. 차기 연구에서는 공간, 시간 특성에 환경적 요인을 고려하여 이동시간과 이용거리 등을 예측하는 것이 필요하다.

References

[1] velib, https://www.velib-metropole.fr/, Aug. 2019.
[2] Bixi, https://montreal.bixi.com/, Aug. 2019.
[3] Divvy, https://www.divvybikes.com/, Aug. 2019.
[4] Daejeon citizen public bike Tashu, https://www.tashu.or.kr/mainPageAction.do?process=mainPage, Aug. 2019.
[5] Sejong citizen public bike Eouling, https://www.sejongbike.kr/mainPageAction.do?process=mainPage, Aug. 2019.
[6] Seoul metropolitan public bike information, http://data.seoul.go.kr/dataList/datasetView.do?infId=OA-15245&srvType=F&serviceKind=1&currentPageNo=1, Aug. 2019.
[7] Seoul metropolitan public data information, http://data.seoul.go.kr/, Aug. 2019.
[8] kaggle, https://www.kaggle.com/, Aug. 2019.
[9] Yellow Cabs tell The Story of New York City, https://www.kaggle.com/selfishgene/

yellow-cabs-tell-the-story-of-new-york-city, 2017.

[10] NYCT - from A to Z with XGBoost (Tutorial), https://www.kaggle.com/karelrv/nyct-from-a-to-z-with-xgboost-tutorial, 2017.

[11] New York City Taxi Trip Duration, https://www.kaggle.com/c/nyc-taxi-trip-duration/overview, 2017.

[12] New York City Taxi Fare Prediction,https://www.kaggle.com/c/new-york-city-taxi-fare-prediction, Aug. 2019.

[13] Bike Sharing Demand, https://www.kaggle.com/c/bike-sharing-demand, 2015.

[14] Seoul metropolitan public bike rental data analysis, https://colab.research.google.com/drive/189HhcbleCnFGuoC4vnHxnEbMxDejmFDr#scrollTo=eqrwc7e5cj-T, Aug. 2019.

[15] Seoulbike EDA, https://github.com/miningful/seoulbike, Nov. 2018.

[16] Aurelien Geron, Hands-on Machine Learning with Scikit-Learn & Tensorflow, O'Reilly, 2017.

[17] Haversine formula, https://en.wikipedia.org/wiki/Haversine_formula, Aug. 2019.

[18] k-meas clustering, https://en.wikipedia.org/wiki/K-means_clustering, Jul. 2019.

[19] Principal component analysis, https://en.wikipedia.org/wiki/Principal_component_analysis, Oct. 2019.

서울시 공공 자전거의 이용 현황 분석

박장우, 조용윤

순천대학교, 정보통신 멀티미디어 공학부 교수

요 약

자전거는 친환경 교통수단으로써 뿐만 아니라 레저 및 건강을 위한 운동 장비로서도 훌륭한 역할을 하고 있다. 여러 나라의 여러 도시들에서 시민의 건강과 교통 정체를 해소하고 자동차 등으로 인한 환경 문제를 해결하기 위하여 공공 자전거 서비스를 실시하고 있다. 효과적인 자전거 공유 서비스를 제공하기 위해서 자전거 활요 현황을 파악하는 것이 필요하다. 서울시의 공공 자전거 데이터 셋을 분석하여 서울시 자전거 이용 형태의 시간적, 공간적 특성을 알아 보았다. 자전거 대여 장소와 반납 장소의 위도, 경도 그리고 이동 시간 등을 Kmeans 클러스터링을 이용하여 50개의 클러스터로 나누고 이들을 공간적 특성을 표현할 수 있도록 시각화 하였다. 이를 통하여 서울시 위치에 따른 자전거 이용 패턴을 확인하였다. 또한 클러스터를 년간 날짜, 요일, 시간 등으로 분류하고 시각화하였다. 주중의 시간에 따른 이용 패턴 및 주말의 이용 패턴을 확인할 수 있다. 주중에는 자전거를 출근과 퇴근에 이용하고 있는 것을 확인할 수 있다. 주말에는 자전거의 이용이 오전부터 서서히 늘어서 오후 6시 경에 최고치를 기록하고 있다. 주중에는 단거리 이용이 많으며 주말에는 다양한 이동 거리 패턴을 보임을 확인하였다.

Jangwoo Park received the B.S., M.S. and Ph.D. degrees in Electronic engineering from Hanyang University, Seoul, Korea in 1987, 1989 and 1993, respectively. In 1995, he joined the faculty member of the Sunchon National University, where he is currently a professor in the Department of Information & Communication engineering. His research focuses on Data Science, Machine learning and Deep learning.

E-mail address: jwpark@sunchon.ac.kr

Yongyun Cho received the Ph.D. degree in computer engineering at Soongsil University. Currently, he is an assistant professor of the Department of Information & communication engineering in Sunchon National University. His main research interests include System Software, Embedded Software and Ubiquitous Computing.

E-mail address: yycho@sunchon.ac.kr