

<https://doi.org/10.7236/JIIBC.2018.18.4.13>

JIIBC 2018-4-2

기계학습 클러스터링을 이용한 승하차 패턴에 따른 서울시 지하철역 분류

Classification of Seoul Metro Stations Based on Boarding/ Alighting Patterns Using Machine Learning Clustering

민미경*

Meekyung Min*

요약 본 연구에서는 기계학습을 이용하여 서울시 지하철역의 승하차 패턴에 따라 지하철역을 분류한다. 대상 데이터는 공공데이터 포털에서 제공하는 2008년부터 2017년까지 서울 지하철 233개 역에서의 매일 매시간별 승차객 숫자와 하차객 숫자이다. 기계학습 기법으로는 가우시안 혼합 모델(GMM)과 K-평균 클러스터링을 사용한다. 이용객의 승차시간과 하차시간의 분포는 가우시안 혼합 모델로 모델링할 수 있으며, 이를 K-평균 클러스터링을 이용하여 비지도 학습시킨다. 학습결과 서울시 지하철역은 승하차 패턴에 따라 4개의 그룹으로 분류되었다. 본 연구의 결과는 서울시 지하철역의 특성을 파악하여 경제, 사회, 문화적으로 분석하기 위한 주요 기반 지식으로 활용될 수 있다. 본 연구의 방법은 클러스터링이 필요한 모든 공공데이터나 빅데이터에 적용할 수 있다.

Abstract In this study, we classify Seoul metro stations according to boarding and alighting patterns using machine learning technique. The target data is the number of boarding and alighting passengers per hour every day at 233 subway stations from 2008 to 2017 provided by the public data portal. Gaussian mixture model (GMM) and K-means clustering are used as machine learning techniques in order to classify subway stations. The distribution of the boarding time and the alighting time of the passengers can be modeled by the Gaussian mixture model. K-means clustering algorithm is used for unsupervised learning based on the data obtained by GMM modeling. As a result of the research, Seoul metro stations are classified into four groups according to boarding and alighting patterns. The results of this study can be utilized as a basic knowledge for analyzing the characteristics of Seoul subway stations and analyzing it economically, socially and culturally. The method of this research can be applied to public data and big data in areas requiring clustering.

Key Words : Machine Learning, Public Data, Seoul Metro Station, GMM, K-means Clustering

I. 서론

공공데이터 포털은 정부부처 및 산하기관에서 생성하는 다양한 공공데이터를 화일과 오픈 API 형태로 제공하

고 있다^[7]. 그동안 국내 공공데이터나 빅데이터의 활용에 관한 여러 연구가 있었다^[1,2,4,5]. 공공데이터 중에는 서울시 지하철역과 관련된 많은 데이터들이 존재한다. 이러한 데이터를 이용한 연구 사례로 지하철 혼잡도 예측^[4],

*정회원, 서경대학교 컴퓨터과학과
접수일자 2018년 6월 14일, 수정완료 2018년 7월 14일
게재확정일자 2018년 8월 10일

Received: 14 June, 2018 / Revised: 14 July, 2018 /

Accepted: 10 August, 2018

*Corresponding Author: mkmin@skuniv.ac.kr

Dept. of Computer Science, Seokyeong University, Seoul, Korea

지하철역 승차인원 예측^[5] 등이 있으나, 아직까지 연구가 활발히 진행되지는 않았다. 서울시에서 지하철을 이용하는 이용객의 숫자는 10년간 100만 명 이상 꾸준히 증가하고 있으므로, 지하철 이용객에 관해 여러 각도에서의 분석이 필요하다. 이러한 분석 연구는 효율적인 지하철 운영을 위해서만이 아니라, 경제, 사회, 문화 등 다양한 분야에서 활용될 수 있을 것이다.

기계학습은 인공지능을 구현하는 핵심기술로서, 고전적으로는 지도 학습, 비지도 학습, 강화 학습 등으로 구분한다. 최근에는 딥러닝이 기계학습의 주류가 되었다. 클러스터링은 특성이 비슷한 데이터를 그룹화해주는 기계 학습 기법으로서, 레이블이 없는 데이터를 그룹화 하는 경우가 일반적이기 때문에 비지도 학습으로 분류한다. 클러스터링 알고리즘으로는 K-Means, DBSCAN, 가우시안 혼합(GMM)을 사용한 최대 기대치 클러스터링, 계층적 클러스터링, Mean-shift 클러스터링 등이 대표적이다^[6].

K-평균 클러스터링은 데이터를 입력 받아 이를 소수의 그룹으로 묶는 알고리즘이다. n개의 중심점을 정한 후에 이 중심점에서 각 점간의 거리의 합이 가장 최소가 되는 중심점 n을 찾고, 이 중심점에서 가까운 점들을 그룹으로 묶는다. K-평균 클러스터링은 간단하면서도 실행 속도가 빠르고 특정 형태의 데이터에 대해 매우 좋은 성능을 보이므로 널리 이용된다^[3].

가우시안 모델은 데이터가 가우스 분포라고 가정하고 평균과 분산 두 개의 파라미터를 이용한다. 가우시안 혼합 모델(GMM)은 이러한 가우스 분포가 여러 개로 구성되는 경우로서, 현실에 존재하는 여러 개의 복잡한 확률 분포를 혼합하여 표현하는 것이 GMM의 기본 아이디어이다. 데이터가 각 분포의 어디에 속하는가를 확률적으로 계산하여 가장 큰 확률을 갖는 정규분포로 클러스터링한다.

본 연구에서는 GMM 모델과 K-평균 알고리즘을 이용하여 서울시 지하철역을 분류한다. 데이터는 공공데이터 포털에서 제공하는 2008년부터 10년간의 서울시 지하철역의 승차객 숫자, 하차객 숫자이다. 본 연구는 이 데이터로부터 시간에 따른 승차, 하차 패턴을 발견하고 발견된 패턴을 기반으로 기계학습을 시켜서 지하철역들을 분류한다. 학습 모듈 및 시각화 프로그램은 파이썬으로 구현하였다.

다음 제2장에서는 지하철 승하차 원본 데이터와 그 특

징을 살펴본다. 제3장에서는 GMM과 K-평균 클러스터링을 이용한 기계학습 과정을 설명한다. 제4장은 학습을 통해 서울시 지하철역을 분류한 결과를 설명한다. 제5장에서는 결론을 맺는다.

II. 지하철 승하차 데이터

1. 원본 데이터

공공데이터 포털에서 제공받을 수 있는 원본 데이터는 서울시 각 지하철역에서 매시간별 승차, 하차 인원이며, 2008년 1월1일부터 2017년 9월30일까지 서울시 233개 역에서의 승하차 데이터이다. 이는 그림 1과 같다.

	A	B	C	D	E	F	G	H
1	날짜	역번호	역명	구분	05-06	06-07	07-08	08-09
2	2008-01-01	150	서울역(150)	승차	379	287	371	
3	2008-01-01	150	서울역(150)	하차	145	707	689	
4	2008-01-01	151	시청(151)	승차	131	131	101	
5	2008-01-01	151	시청(151)	하차	35	158	203	
6	2008-01-01	152	종각(152)	승차	1287	867	400	
7	2008-01-01	152	종각(152)	하차	74	147	261	
8	2008-01-01	153	종로3가(153)	승차	484	290	192	
9	2008-01-01	153	종로3가(153)	하차	28	80	116	
10	2008-01-01	154	종로5가(154)	승차	89	77	86	
11	2008-01-01	154	종로5가(154)	하차	14	76	93	
12	2008-01-01	155	동대문(155)	승차	190	149	152	
13	2008-01-01	155	동대문(155)	하차	23	111	120	
14	2008-01-01	156	신설동(156)	승차	65	107	113	
15	2008-01-01	156	신설동(156)	하차	15	98	81	
16	2008-01-01	157	계기동(157)	승차	156	129	115	
17	2008-01-01	157	계기동(157)	하차	16	84	84	
18	2008-01-01	158	정왕리(지하)(158)	승차	230	292	351	
19	2008-01-01	158	정왕리(지하)(158)	하차	56	428	510	

그림 1. 서울시 지하철 승하차 원본 데이터

Fig. 1. Boarding/ Alighting Raw Data of Seoul Metro Stations

2. 승하차 데이터의 시각화

승하차 패턴에 따른 지하철역 분류 연구를 위해서, 먼저 원본 데이터를 시각화하여 보았다. 시각화를 위해 10년간 시간별 승차객 숫자와 하차객 숫자를 입력하여 이의 평균값을 구한 후 그래프를 그리는 파이썬 프로그램을 작성하였다. 하루 중 05시~24시까지 1시간 단위로 10년치 승차객 수와 하차객 수의 평균을 그래프로 나타낸 것이다. 1시간 단위로 데이터가 구성되어 있으므로 시간은 중간 값을 사용하였다. 그림 2는 233개 중 8개 역의 예를 보여주고 있다. 8개 역은 가산디지털단지, 여의도, 까치산, 상계, 강동구청, 마포, 동작, 대치역이다. 가로축은 지하철이 운행되는 05시~24시이며 세로축은 승차객 수(점선), 하차객 수(실선)이다.

다음 4장에 설명되어 있듯이 본 연구의 기계학습에 의

한 분류 결과, 가산디지털단지역과 여의도역은 동일 그룹으로 분류된다. 까치산과 상계역이 동일 그룹이며, 강동구청과 마포, 대치와 동작이 동일 그룹이다. 이들은 원본 데이터에서도 직관적으로 비슷한 승하차 패턴을 보임은 알 수 있다.

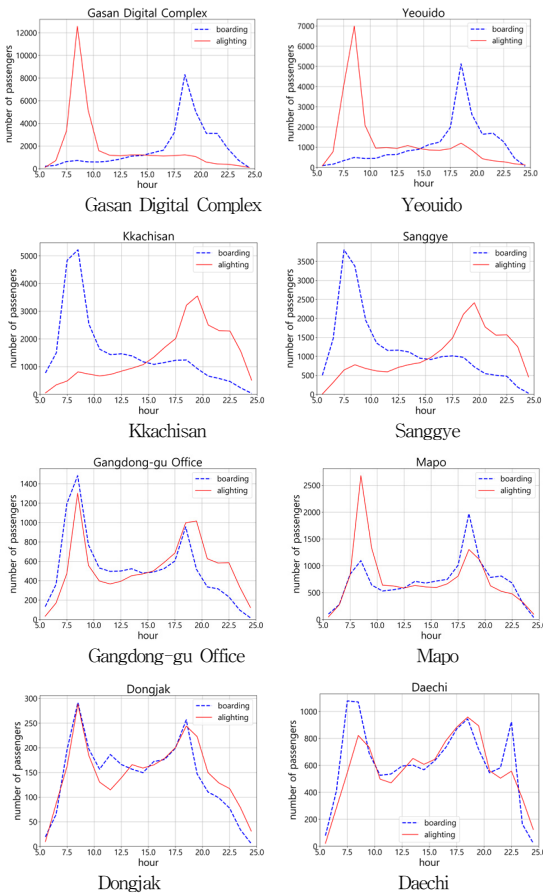


그림 2. 지하철역별 시간별 승하차객 수 시각화
 Fig. 2. Visualization of Hourly Boarding/ Alighting Data of Seoul Metro Stations

3. 승하차 데이터의 특징

그림 2의 예에서 볼 수 있듯이 원본 데이터의 시간별 지하철 승하차객 수에서는 다음과 같은 특징이 나타남을 알 수 있다.

첫째, 각 지하철역의 시간별 승차객 그래프와 하차객 그래프에서는 일반적으로 몇 개의 피크점이 나타난다. 또한 이 피크를 중심으로 좌우로 감소하는 형태의 분포를 보이고 있다. 예를 들어 강동구청 역의 승차 그래프는

8.5시에 첫 번째 피크를 보이며 이를 중심으로 좌우로 감소하는 형태이다. 또한 18.5시에 두 번째 피크를 보이며 이를 중심으로 좌우로 감소하는 형태를 보인다. 하차 그래프에서도 마찬가지이다.

둘째, 피크를 중심으로 좌우로 감소하는 형태는 정규 분포 곡선과 유사하다. 이로부터 가우시안 모델이 적합하다는 직관을 얻을 수 있다. 피크의 개수는 역마다 다르다. 1-2개인 경우도 있고 3-4개인 경우도 있으나, 2개의 피크를 갖는 경우가 가장 일반적이다. 따라서 2차의 가우시안 혼합 모델을 사용할 수 있다. 원본 데이터는 정확한 정규분포를 이루는 것이 아니므로, 이를 보정하면 가우시안 혼합 모델을 이용할 수 있다.

III. 학 습

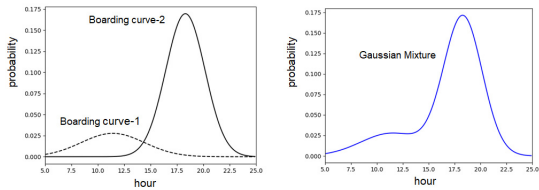
1. 가우시안 혼합

원본 데이터를 시각화 해 본 결과 가우시안 분포에 적합하다는 점에 착안하여 본 연구를 진행하였다. 2차 가우시안 혼합 모델(GMM)에 의해 각 233개 역에서의 시간별 승차객수를 2개의 정규분포로 보정할 수 있다. 이 2개의 정규분포는 혼합되어 가우시안 혼합을 형성한다. 2008년 1월1일부터 2017년까지 9월30일까지의 데이터를 모두 반영하였다.

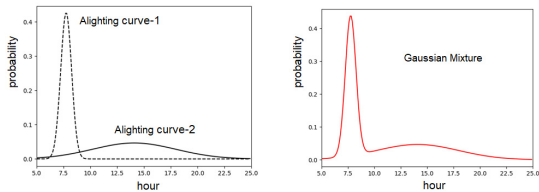
그림 3은 가우시안 혼합 그래프로 표현한 가산디지털단지역과 동작역의 예이다. (a)는 가산디지털단지의 승차 패턴이다. 좌측 그래프는 승차 패턴을 나타내는 2개의 가우시안 커브로서 정규분포 형태의 확률밀도함수(PDF)이다. 우측 그래프는 이 좌측 2개의 커브를 혼합하여 최종 승차 패턴을 나타내는 가우시안 혼합 그래프이다. 마찬가지로 (b)는 가산디지털단지의 하차 패턴, (c)는 동작역의 승차 패턴, (d)는 동작역의 하차 패턴이다. 그래프의 가로축은 지하철이 운행되는 05시~24시이다. 데이터가 1시간 단위이므로 시간은 중간 값을 사용한다. 세로축은 확률이며, 그 시각에 승차객이 발생할 확률, 하차객이 발생할 확률을 의미한다.

가우시안 혼합으로부터 각 역마다 승하차 패턴을 모델링한 가우시안 커브의 피크시간을 구할 수 있다. 각 역의 첫 번째 가우시안 커브의 승차 피크 시간을 on1, 두 번째 가우시안 커브의 승차 피크 시간을 on2, 첫 번째 가우시안 커브의 하차 피크 시간을 off1, 두 번째 가우시안 커

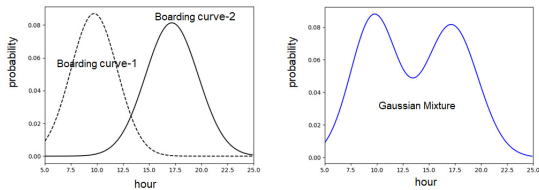
브의 하차 피크 시간을 off2라 한다. 가우시안 혼합 모델링 결과 233개의 각 역은 [on1, on2, off1, off2]로 이루어진 4차원의 데이터를 갖는다. 가산디지털단지의 경우, on1은 11.4시, on2는 18.3시, off1은 7.74시, off2는 14.08시이다. 동작역의 경우, on1은 9.71시, on2는 17.16시, off1은 9.25시, off2는 17.25시이다.



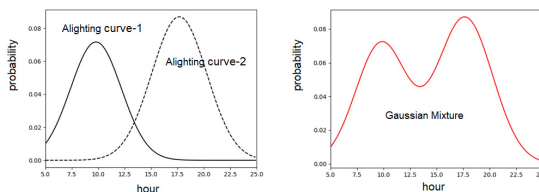
(a) Boarding Pattern of Gasan Digital Complex Station



(b) Alighting Pattern of Gasan Digital Complex Station



(c) Boarding Pattern of Dongjak Station



(d) Alighting Pattern of Dongjak Station

그림 3. 가우시안 혼합에 의한 승차 패턴과 하차 패턴
Fig. 3. Boarding/ Alighting Patterns Based on Gaussian Mixture

2. K-평균 클러스터링

원본 데이터로부터 얻은 직관과 가우시안 혼합으로부터 얻은 데이터를 볼 때, 지하철역들은 4개의 그룹이라고 가정한다. 가우시안 혼합으로 생성된 4차원의 데이터를 입력으로 K-평균 클러스터링을 이용하여 다음과 같이 비지도 학습을 시킨다.

입력) 233개의 4차원 벡터: 4차원 벡터는 [on1, on2, off1, off2]이며, 233개는 역의 개수이다.

단계1) 초기화: 그룹 역그룹0, 역그룹1, 역그룹2, 역그룹3에 해당하는 초기 중심 값 c0, c1, c2, c3을 무작위로 설정한다. c0, c1, c2, c3는 각각 4차원 벡터이다.

단계2) 할당: 233개의 입력 데이터에 대하여 가장 가까운 거리에 있는 c0, c1, c2, c3을 찾아서 이에 해당하는 분류 그룹 역그룹0, 역그룹1, 역그룹2, 역그룹3을 배정한다.

단계3) 보정: 같은 분류 그룹 내의 모든 데이터들의 평균 중심 값을 다시 계산하여 c0, c1, c2, c3을 재 설정한다.

단계4) 학습이 끝날 때까지 단계2로 돌아가 반복한다. K-평균 클러스터링은 몇 가지 제약에도 불구하고, 간단하면서 성능이 좋아 널리 사용되며, 본 연구의 지하철역 분류에서는 우수한 결과를 보였다.

IV. 분 류

1. 클러스터링 결과

학습에 의한 클러스터링 결과 233개의 지하철역은 초기 가정대로 4개의 그룹으로 분류되었다. 이 클러스터링 결과는 4개의 변수, on1, on2, off1, off2 에 따라 분류된 그룹이므로, 4차원이 된다. 따라서 형상화하기가 불가능하다. 가시적인 확인을 위해 이를 2차원으로 나누어 분포도를 그릴 수 있다. 분포도의 가로축과 세로축은 모두 가우시안 커브의 피크 시간을 나타낸다. 4개의 값 on1, on2, off1, off2가 있으므로 16개의 조합으로 그래프를 그릴 수 있다. 이 중에서 의미 있는 조합은 가우시안 커브의 승차 피크 시간과 가우시안 커브의 하차 피크 시간과의 관계를 그린 8개뿐이며, 이 중 절반은 가로와 세로만 바뀐 것이므로 실제로 의미 있는 2차원 분포도 그래프는 다음 4가지 경우이다. off1(세로축)-on1(가로축) 그래프, off1(세로축)-on2(가로축) 그래프, off2(세로축)-on1(가로축) 그래프, off2(세로축)-on2(가로축) 그래프이다.

다음 그림 4는 본 연구로 생성된 클러스터링 결과들이 4가지 그래프의 형태로 보여준다. 클러스터링 그룹에 따라 4가지 색으로 표시하였다. 가로축과 세로축의 눈금은 모두 시간을 나타낸다. 의미는 각각 가우시안 승차 피크, 가우시안 하차 피크이다. 그룹 G0은 on1, on2나 off1,

off2가 하루 중 넓게 분포하며, 이른 시간에는 나타나지 않는다. 저녁 시간에는 승차와 하차가 모두 발생한다. 따라서 업무지역은 아니나 사람들의 이동이 많은 역임을 알 수 있다. 환승지역이나 상업지역이 많을 것으로 예상되는 역이다. 그룹 G1은 승차와 하차가 이른 아침이나 저녁보다는 상대적으로 낮 시간대에 주로 발생한다. 그룹 G1은 주거, 업무, 상업 어느 쪽으로도 구분하기 어려운 그룹이다. 그룹 G2는 on1이 오전 이른 시간, off2가 저녁 시간에 집중되어 있으므로 주거지역임을 알 수 있다. 그룹 G3은 off1이 오전 이른 시간에, on1이 저녁 시간에 집중되어 있으므로 업무지역임을 알 수 있다.

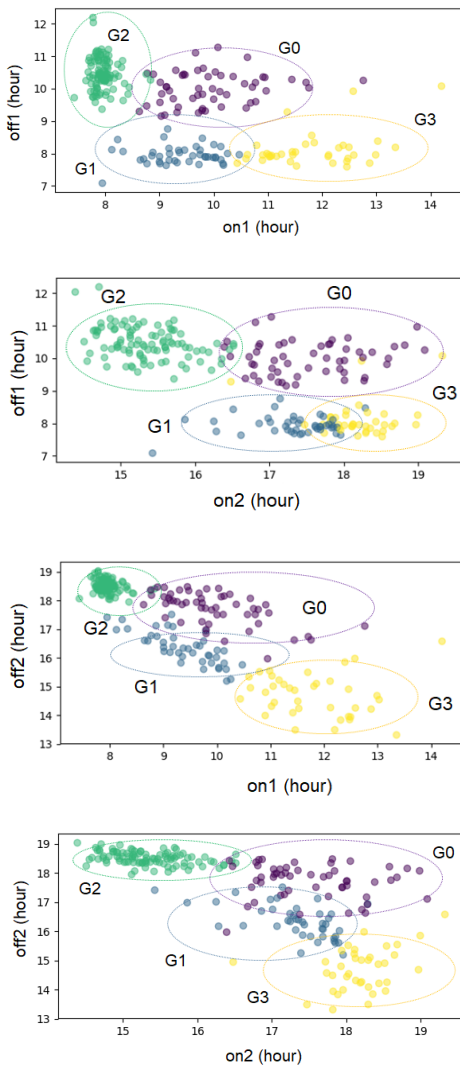


그림 4. 기계학습에 의한 클러스터링 결과
 Fig. 4. Clustering Results by Machine Learning

2. 지하철역 분류

서울시 지하철역은 학습 결과에 의해 4개의 그룹으로 분류되었으며, 역그룹0은 54개, 역그룹1은 43개, 역그룹2는 98개, 역그룹3은 38개의 역으로 구성되었다. 분류결과는 표 1과 같다.

표 1에서 동일 그룹에 속해 있는 지하철역들의 승하차 패턴을 그림 3과 같은 가우시안 혼합과 비교해 보면 그 유사점을 찾을 수 있다. 그림 4에 나타난 바와 같이 역그룹0은 환승지역 또는 상업지역으로 볼 수 있다. 역그룹1은 기타그룹으로 볼 수 있으며, 역그룹2는 주거지역, 역그룹3은 업무지역으로 볼 수 있다.

표 1. 서울시 지하철역 분류
 Table 1. Classification of Seoul Metro Stations

분류	역 이름	개수
역 그룹 0	강변 개롱 개화산 건대입구 고덕 고려대 고속터미널 김포공항 노원 녹사평 당산 대치 독립문 동대문 동대문운동장 동작 마장 마포구청 망원 모란 발산 사당 상수 성신여대입구 송정 신당 신도림 신천 신촌 안암 약수 어린이대공원 영등포시장 오목교 올림픽공원 왕십리 용두 월드컵경기장 이대 이수 이촌 이태원, 일원 잠실 장지 제기동 종합운동장 천호 청량리 충신대입구 학여울 한성대입구 혜화 홍대입구	54
역 그룹 1	가락시장 강동구청 공덕 광흥창 구로디지털단지 남구로 남태령 내방 대흥 도곡 도림천 디지털미디어시티 똑섬 마곡 마포 배봉 동촌토성 문래 문정 반포 방배 방이 버티고개 보라매 복정 상왕십리 서울역 석촌 성수 송파 수서 수대입구 신길 신철동 애오개 양평 영등포구청 오금 장한평 청구 충정로 합정 효창공원앞	43
역 그룹 2	강동 거여 공릉 광나루 광명사거리 구산 구의 구파발 군자 굽은다리 금호 길동 길음 까치산, 낙성대 남성 남한산성입구 녹번 단대오거리 답십리 당고개 대림 대청 도봉산 독바위 돌곶이 둔촌동 똑섬유원지 마들 마천 먹골 면목 명일 목동 무악재 미아 미아삼거리 방화 보문 봉천 봉화산 불광 사가정 산성 상계 상도 상봉 상월곡 상일동 새철 서울대입구 석계 성내 수락산 수유 수진 송실대입구 신금호 신담 신대방 신대방삼거리 신림 신정 신정네거리 신흥 신흥쌍문 아차산 아현 압사 양천구청 역촌 연신내 옥수 운수 용답 용마산 우장산 월곡 응암 잠원 장승배기 잠암 중계 중곡 중화 증산 지축 창동 창신 철산 태릉입구 하계 행당 홍제 화곡 화랑대	98
역 그룹 3	가산디지털단지 강남 강남구청 경복궁 광화문 교대 남부터미널 논현 동대문역사문화공원 동대입구 동묘앞 명동 삼각지 삼성 서대문 서초 선릉 시청 신사 신용산 안국 압구정 양재 여의나루 여의도 역삼 을지로3가 을지로4가 을지로입구 종각 종로3가 종로5가 청담 충무로 학동 한강진 한양대 회현	38

V. 결 론

본 연구에서는 233개 서울시 지하철역을 승하차 패턴에 따라 분류하였다. 구현은 파이썬 언어를 사용하였다. 대상 데이터는 공공데이터 포털에서 제공하는 233개 지하철 각 역에서의 매일 05시부터 24시까지 1시간 단위의 승차인원과 하차인원이다. 기간은 2008년 1월1일부터 2017년 9월30일까지이다.

지하철역의 분류는 승하차가 발생하는 패턴을 기반으로 한다. 분류방법으로는 기계학습 방법 중, 가우시안 혼합 모델과 K-평균 클러스터링 알고리즘을 사용하였다. 각 역의 시간별 승차객 수와 하차객 수의 원본 데이터를 시각화해 본 결과 이는 2차 가우시안 혼합 모델에 적합하였다. 따라서 원본 데이터를 가우시안 혼합 모델로 보정하여 2개의 가우시안 커브의 승차 피크 시간과 2개의 가우시안 커브의 하차 피크 시간을 구하였다. 이 4차원 데이터를 기반으로 K-평균 클러스터링에 따라 지하철역을 분류하였으며, 클러스터링 결과 4개의 그룹으로 분류할 수 있었다.

첫 번째 그룹의 지하철역들은 이용도가 매우 높거나 환승지역, 상업지역에 속한 역으로 볼 수 있었다. 두 번째 그룹은 업무, 상업, 주거 중 어느 한 쪽의 특정한 지역으로 지정하기 어려운 그룹이었다. 세 번째 그룹의 지하철역들은 주거지역으로 볼 수 있으며, 네 번째 그룹의 지하철역들은 업무지역으로 볼 수 있었다.

본 연구의 결과는 서울시 지하철역의 특성을 파악하는 데 의미가 크며, 지하철역 주변의 특성을 경제, 사회, 문화적으로 분석하기 위한 주요 기반 지식으로 활용될 수 있다. 향후 본 연구에서 제시한 지하철역의 승하차 패턴뿐만 아니라, 주변의 상업시설이나 인구분포, 혼잡도 등의 요소와 본 연구의 결과를 통합하여 지하철역을 분류한다면 보다 다양한 정책이나 연구에 활용될 수 있을 것이다.

References

- [1] BongHyun Back, Il-Kyu Ha, "A Method for Selective Storing and Visualization of Public Big Data Using XML Structure", Journal of the Korea Institute of Information and Communication Engineering, Vol. 21, No. 12, pp. 2305-2311, Dec 2017.
- [2] Jae-Young Chang, "An Experimental Evaluation of Box office Revenue Prediction through Social Bigdata Analysis and Machine Learning", The Journal of the Institute of Internet, Broadcasting and Communication(JIIBC), Vol. 17, No. 3, pp. 167-173, Jun 2017.
DOI: <https://doi.org/10.6109/jkiice.2017.21.12.2305>.
- [3] Min-Soo Kang, Yong-Gyu Jung, Du-Hwan Jang, "A Study on the Search of Optimal Aquaculture farm condition based on Machine Learning", The Journal of the Institute of Internet, Broadcasting and Communication(JIIBC), Vol. 17, No. 4, pp. 135-140, Apr 2017.
DOI: <https://doi.org/10.7236/JIIBC.2017.17.3.167>.
- [4] Jin-su Kim, "Subway Congestion Prediction and Recommendation System using Big Data Analysis", Journal of Digital Convergence, Vol. 14, No. 11, pp. 289-295, Nov 2016.
DOI: <https://doi.org/10.14400/JDC.2016.14.11.289>.
- [5] Minwoo Kim, "Predicting Subway Passengers Flows By Spatio-Temporal Modeling", Master Thesis, Seoul National University, Aug 2017.
- [6] R. S. Michalski, J. G. Carbonell, T. M. Mitchell, Machine Learning: An Artificial Intelligence Approach, Springer Science & Business Media, 2013.
- [7] <https://www.data.go.kr>

저자 소개

민 미 경(정회원)



- 1987년 : 서울대학교 계산통계학과 학사
- 1989년 : 서울대학교 계산통계학과 전산과학전공 석사
- 1993년 : 서울대학교 계산통계학과 전산과학전공 박사
- 1994년 ~ 현재 : 서경대학교 컴퓨터 과학과 교수

<주관심분야 : 데이터베이스, 인공지능, 지능정보>