

# 공공데이터를 활용한 도시 내 공공체육시설 위치 추천

임주영<sup>1</sup>, 팽소연<sup>2</sup>, 이가은<sup>3</sup>, 이찬녕<sup>4</sup>, 구재성<sup>5</sup>, 안서현<sup>6</sup>, 강민지<sup>7</sup>, 김진<sup>8\*</sup>, 이지항<sup>8</sup>

<sup>1</sup>숙명여자대학교 통계학과, <sup>2</sup>서울시립대학교 통계학과  
<sup>3</sup>숙명여자대학교 컴퓨터과학과, <sup>4</sup>단국대학교 응용통계학과  
<sup>5</sup>상명대학교 컴퓨터과학과, <sup>6</sup>삼육대학교 컴퓨터메카트로닉스공학부  
<sup>7</sup>한국외국어대학교 생명공학과, <sup>8</sup>상명대학교 지능데이터융합학부

wn\_young@naver.com, paengsy@naver.com, xmasge424@gmail.com, icn9713@naver.com,  
 gjs5448@gmail.com, hellomstars@daum.net, rkdalsw11999@daum.net, jinkim@smu.ac.kr,  
 jeehang@smu.ac.kr

(\*: 교신저자)

## A Location Recommendation Model for Public Sports Facilities

Joo-Young Lim<sup>1</sup>, So-Yeon Paeng<sup>2</sup>, Ga-Eun Lee<sup>3</sup>, Chan-Nyoung Lee<sup>4</sup>,  
 Jae-Sung Koo<sup>5</sup>, Seo-Hyun Ahn<sup>6</sup>, Min-Ji Kang<sup>7</sup>, Jin Kim<sup>8</sup>, Jee Hang Lee<sup>8</sup>

<sup>1</sup>Dept. of Statistics, Sookmyung Women's University

<sup>2</sup>Dept. of Statistics, University of Seoul

<sup>3</sup>Dept. of Computer Science, Sookmyung Women's University

<sup>4</sup>Dept. of Information Statistics, Dankook University

<sup>5</sup>Dept. of Computer Science, Sangmyung University

<sup>6</sup>Division of Computer·Mechatronics, Samyook University

<sup>7</sup>Dept. of Bioscience and Biotechnology, Hankuk University of Foreign Studies

<sup>8</sup>School of AI and Data Science, Sangmyung University

### 요 약

본 논문에서는 서울시를 대상으로 2020년 기준 자치구별 공공체육시설의 개수를 분석하고, 도출된 서비스 지역 적정 개소 수를 기준으로 추가 설치가 필요한 자치구 내 입지를 예측하였다. 기존 공공체육시설 수와 선행연구의 입지 지표를 활용해 회귀분석을 바탕으로 유의한 입지요인을 도출하고, 이를 변수로 한 k-means 군집화를 통해 자치구별 입지 후보군이 될만한 행정구역상 동을 구분하였다. 이후 선정된 행정구역 내 기준 인구 당 공공체육시설 비율이 같아지도록 공공체육시설 설치 개수를 결정한 다음 각 구역의 중심점으로부터 가까운 동 순으로 공공체육시설의 추가 설치가 필요한 동을 선정하였다.

### 1. 서론

다양한 이동 수단 사용과 업무, 교육, 오락을 위한 컴퓨터 사용이 증가함에 따라 현대인의 운동량이 현격하게 줄어들고 있는 바, 심혈관 질환, 당뇨병 등 성인병 환자와 비만 인구가 증가하는 실정이다. 이러한 질병들은 규칙적인 신체활동을 통해 예방할 수 있는데, 국가와 지역사회는 국민의 신체활동을 격려하고 촉진하는 방법 및 관련 시설을 추가하고 장려하는 움직임을 보이고 있다.

이에 발맞추어, 최근에는 증가하는 국민 소득에 따라 규칙적인 체육활동 참여도가 높아지고 있다. 연구에 따르면, 거리상 가까운 곳에 위치하고 저렴하게 이용할 수 있는 공공체육시설의 선호도와 이용률이 가장 높다.[1] 여기서 공공체육시설이란 국가나

지방 자치단체 또는 공공단체가 국민의 체육활동에 제공하기 위하여 설치 및 관리, 운영하는 체육시설을 말한다.[2]

체육시설의 수요 증가에 따라 공공스포츠시설의 보급률은 개선되었지만, 선진국 수준에는 여전히 미치지 못하고 있다. 인구밀도가 높은 도시지역에서도 접근성은 확보되나 서비스 반경 내 거주 인구가 많아 적정 서비스가 이루어지지 못하는 경우가 다수 발생하므로 추가적인 시설 공급이 필요하다.[3] 이에 따라, 본 연구는 공공빅데이터를 활용, 서울시 내 공공체육시설 확충이 필요한 지역을 분석하고, 군집 분석을 통해 서울시 자치구별 공공체육시설의 입지 선정 예측에 대해 논하고자 한다.

2. 방법

2.1 사용 모델

**다중회귀분석** 다중회귀분석은 2개 이상 복수의 독립변수들을 이용한 회귀분석으로, 독립변수가 종속변수에 미치는 영향력의 크기를 파악하고 이를 통하여 독립변수의 일정한 값에 대응하는 종속변수 값을 예측하는 모형을 산출하는 방법이다.[4] 수식은 아래와 같다.

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k, \quad (1)$$

여기서  $y$ 는 종속변수,  $\beta_k$ 는  $k$ 번째 회귀계수,  $x_k$ 는  $k$ 번째 독립변수를 의미한다. 여기서는 선행연구[5]의 공공서비스 입지요인 지표들을 독립변수로 사용했으며, 공공체육시설의 수를 종속변수로 하였다.

**k-means clustering** k-means clustering[6]은 중심점과 각 개체 간의 유클리디안 거리를 계산해 가까운 군집으로 개체를 분류하는 군집화 비지도 학습 알고리즘이다. 개체를 각 군집으로 분류한 후 평균값으로 중심점을 갱신하며, 더이상 갱신이 되지 않을 때까지 알고리즘을 반복한다. 유클리디안 거리 계산식은 아래와 같다.

$$dist(x_i, c_k) = \|x_i - c_k\| = \sqrt{\sum_{j=1}^p (x_{ij} - c_{kj})^2}. \quad (2)$$

군집의 개수는 하이퍼 파라미터로 사용자가 제시하나, 보통은 elbow 방법과 silhouette 방법을 활용하여 추정한다. 여기서는 elbow 방법을 활용하였는데, 이것은 군집 수에 대한 SSW(Sum of Squares Within the clusters)의 플롯(plot)으로부터 최적 군집 개수를 추정한다.

2.2 입지 요인 도출 및 추가 설치 입지 추정

클러스터링을 위한 입지요인 선택을 위해 다중회귀분석을 진행하였다. 독립변수 간 다중공선성 제거를 위해 상관계수의 크기가 0.7 이상인 변수를 제거하였다. 나머지 변수를 사용해 backward stepwise 회귀분석을 적용한 결과는 Table 1과 같다. backward stepwise 다중회귀분석 결과 인구수, 재산세, 수급인구비율을 입지요인으로 선정하였다.

서울시 수요 불충족 지역구 K-means clustering 결과를 지도 시각화한 결과는 Fig. 1과 같다.

Table 1. backward stepwise 다중회귀분석 결과

	step 1	step 2	step 3	step 4	step 5
Adjusted R Squared	0.7188	0.7297	0.7394	0.7380	0.7325
Variable	인구수	인구수	인구수	인구수	인구수
	재산세	재산세	재산세	재산세	재산세
	수급인구비율	수급인구비율	수급인구비율	수급인구비율	
	투표율	투표율	투표율		
	재정수입	재정수입			
중사업체수					

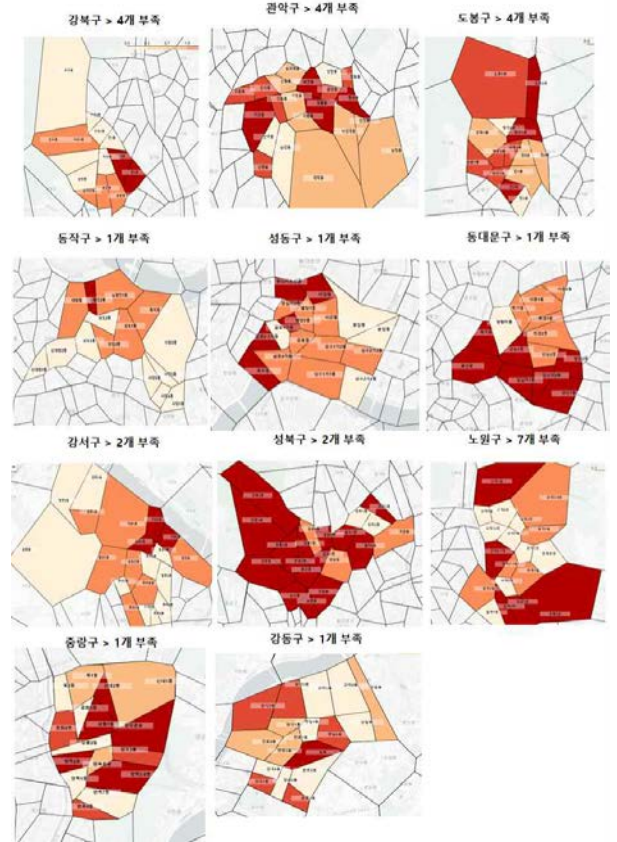


Fig. 1. 서울시 k-means clustering 결과

2.3 공공체육시설 추가 설치 입지 예측

공공체육시설의 최종 입지 예측을 위해 클러스터별 인구 만 명당 체육시설 수의 비율을 구했다. 수식은 아래와 같다.

$$p_0 = \frac{(\text{기존개수}) + x}{(\text{인구수})} \times 10000, \quad (3)$$

여기서  $x$ 는 추가 설치 개수를 말한다.

이 비율이 동일하도록 각 cluster에 추가로 필요한 체육시설 개수를 결정하였다. 실제로는 비율이 완벽하게 동일하도록 개수 선정이 불가능한 바, 여기서 각 cluster 비율의 오차제곱합이 가장 작아지도록 하는  $x$ 값을 선정하였다.

이후, 각 cluster의 중심점을 구해 그 중심점으로부터 가까운 동부터 시설을 하나씩 설치하는 방향으로 위치를 선정했다. 선정된 추가 설치 입지는 Fig. 2와 같다.

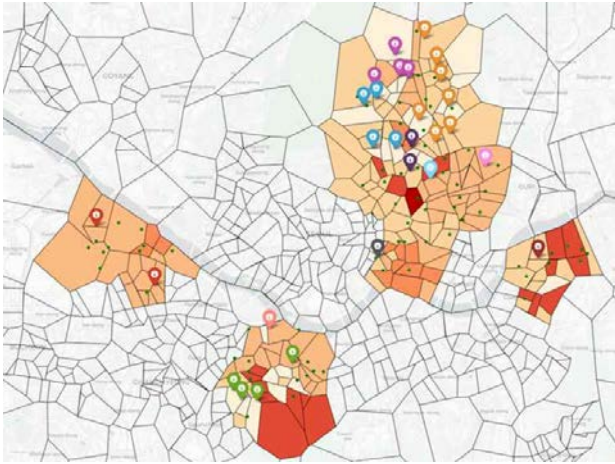


Fig. 2. 서울시 공공체육시설 추가 설치 입지

### 3. 결론

본 연구의 목적은 지역 주민들의 접근성과 기대 수요를 파악하여 적절한 입지 위치를 제안하는 위치 선정모델을 만드는 것이다. 먼저 자료 수집이 용이한 서울시 내에서 자치구별 공공체육시설 수를 확인하고 서비스 지역 적정 개소수를 기준으로[7] 공공체육시설의 확충이 필요한 지역을 선별하여 clustering 하였다.

기존 공공체육시설 현황과 선행연구의 입지분석 지표를 활용한 회귀분석을 통해 인구수, 재산세, 수급인구 비율을 입지요인으로 도출하였다. 다음으로 서울시 내 수요 불충족 지역구를 대상으로 k-means clustering 분석을 진행하였다. 앞에서 도출한 입지요인을 변수로 cluster를 나누고 행정동을 배분하여 cluster 별 만 명당 공공체육시설 설치 비율이 같아 지도록 공공체육시설 설치 개수를 결정한 뒤 각 cluster의 중심점으로부터 가까운 동 순으로 공공체육시설의 추가 설치가 필요한 동을 선정, 시각화하였다.

본 모델을 적용하여 공공체육시설 추가 설치에 대해 예측했을 때 Fig. 3에서 볼 수 있듯이, 공공체육시설이 더 균등하게 배분되어 있음을 알 수 있다. 공공체육시설의 균등하게 배분된 정도를 분산으로 정량화하면, 이전 (Fig. 3-가)는 0.0083, 추가 설치 경우 (Fig. 3-나)는 0.0040으로, 추가 설치 시 각 자치구 별 편차가 더 줄었음이 확인되었다. 이는 더 많은 자치구에서 더 균등한 조건으로 체육시설을 이용할 수 있음을 의미하며, 결국 전 국민 신체 건강 증진 가능성을 높이는 계기가 될 수 있다.

최근 국민 건강이 중요한 이슈로 부상함에 따라 체육활동의 수요가 늘어나면서 공공체육시설 증설의

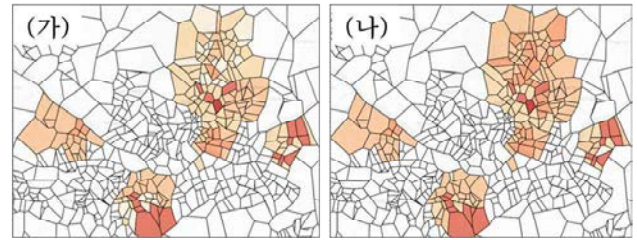


Fig. 3. 공공체육시설 설치 현황 및 추가 설치 예측 사례. (가) 공공체육시설 기존 설치 현황. (나) 공공체육시설 추가 설치 예측 결과. (나)에서 공공체육시설이 더 균등하게 배분되어 있음을 알 수 있음. 균등하게 배분된 정도를 분산으로 나타내었을 때, (가)는 0.0083, (나)는 0.0040으로, 추가 설치 시 각 자치구 별 편차가 더 줄었음이 확인됨.

필요성이 높아졌다. WHO에 따르면[8], 신체활동은 심혈관 질환, 암, 당뇨 등 비전염성 질환 예방에 기여한다. 실제로 연구 과정에서 비교해본 결과, 공공체육시설 필요 개수가 충족된 자치구의 당뇨, 고혈압, 비만율이 그렇지 않은 자치구에 비해 눈에 띄게 우세함을 보였다 ( $p < 0.05$ ). 본 모델을 통해 체육시설이 적절한 위치에 균등하게 배분됨에 따라 더 많은 사람의 신체활동 증진을 기대할 수 있으며, 이를 통해 국민 건강 수준 향상 및 의료비 감소 등 사회적 파급 효과를 불러일으킬 것으로 기대한다.

본 논문은 한국데이터산업진흥원이 진행하는 2022 데이터 청년 캠퍼스 지원을 받았습니다.

### 참고문헌

- [1] 국민생활체육 참여실태 조사, 문화체육관광부,75-77 ,2021
- [2] 이태신, 체육학대사전, 민중서관, 2000
- [3] 생활 SOC 추진단, 생활SOC 3개년 계획(안), 국무조정실,9 , 2019
- [4] 이선희, 윤동현, 고준환, 다중회귀분석을 이용한 서울시 1인 가구의 연령별 분포요인에 관한 연구, Spatial Information Research, 통권, 84호, 11-21, 2015
- [5] 김현민, 김희영, 도시 공공서비스 시설의 공급 결정요인, 사회과학연구논총, 제 12권, 109-128, 2004
- [6] 신지은, "RHadoop 플랫폼을 이용한 빅데이터 K-평균 클러스터링.", 국내석사학위논문, 경상대학교 대학원 ,10 , 2016
- [7] 체육진흥과, 생활밀착형 국민체육센터 모델개발 및 추진 방안 연구, 문화체육관광부,1-5 , 2018
- [8] WHO, Physical activity 2020.11. <https://www.who.int/news-room/fact-sheets/detail/physical-activity>