

서울시 열섬현상 완화를 위한 바람길 및 녹지 입지 선정 제안: 클러스터링 및 랜덤포레스트 변수 중요도를 활용한 열섬척도 제안을 중심으로

간정현* · 남승희** · 서윤지*** · 전은지**** · 이광춘*****

A Proposal for the Location of Wind Road and Green Space for the Mitigation of Urban Heat Island Effect in Seoul: Focusing on the proposal for heat island index using clustering and variable importance by random forest

Gan, Jeong-Hyun · Nam, Seung-Hee · Seo, Yun-Ji · Jeon, Eun-Ji · Lee, Kwang-Chun

Abstract

As the climate problem becomes more serious, mitigating the heat island effect is becoming an important issue. To solve the problem, it is necessary to identify the areas where the heat island effect is more severe and present appropriate measures to those areas. In this paper, we divided Seoul into 573 grids and divided into four clusters by hierarchical clustering techniques to identify the current status of the city's heat island effect. Of these cluster, we chose the one with the most urban features. We extracted the variable importance with random forest models and developed the heat island index from 1) multiplication of the variable importance and 2) the sign of the correlation coefficient between temperature and 3) the corresponding variable. Furthermore we analyzed areas where the heat island effect is severe with the index. According to the analysis, the heat island effect was particularly severe in five grids, and we suggested ways to mitigate the heat island effect for the areas concerned.

Key Words: 열섬현상, 열섬지표, 클러스터링, 랜덤포레스트, 변수중요도, 바람길, 녹지 **영문확인** ●

I. 서 론

1.1. 연구 배경 및 목적

산업혁명 이후로 도시에 인구가 집중되면서 도시의

거대화, 고밀도, 집적화 현상이 일어나고 있다. 급속한 도시화로 인해 도시에 많은 에너지가 집중되고, 인공배열로 인해 도시의 열 환경이 변화하고 있으며, 이에 따라 여러 환경 문제가 야기되고 있다. 특히나 높은 건물들과 아스팔트, 콘크리트로 이루어진 지표면이 증가하고 화석연료 사용량이 늘어나며 녹지와 하천 등의 영역이 줄어들면서 도시 기온이 상승하고 열이 제대로 방출되지 못해 열섬현상이 발생하고 있다.

* 연세대학교 사회과학대학 사회학과 학부
** 연세대학교 문과대학 문헌정보학과 학부
*** 연세대학교 상경대학 응용통계학과 학부
**** 연세대학교 문과대학 문헌정보학과 학부
***** 연세대학교 소속확인 ●, 교신저자

〈표 1〉 분석을 위해 사용된 변수

범주	변수	설명	제공기관
기상 상태	온도	Landsat8 위성영상을 이용하여 추출한 지표면의 온도	USGS
	풍속	지상 10m풍속 및 80m 풍속	기상자료개방포털
	미세먼지	지점별 미세먼지 농도	AirKorea
토지 이용 상태	배수등급	배수등급이 수준별로 분류된 GIS 공간정보 자료	국가공간정보포털
	토지피복	토지피복 내역을 나타낸 GIS 공간정보	서울 열린데이터광장
지리적 특성	해발고도	대한민국의 해발고도를 나타낸 데이터	국토지리정보원

열섬현상은 불쾌감을 유발하고 도시민의 건강을 위협하며 에너지 소비를 증가시키는 등 여러 문제점을 동반한다. 체온 조절 능력이 미흡하고 체내 수분량이 적은 노인과 어린이, 그리고 열섬현상에 대처능력이 부족한 사회적 취약계층은 특히나 위험에 노출되어 있다. 에너지 소비 문제 역시 심각하다. 1996~2011년 1인당 전력소비량을 측정한 연구 결과에 따르면, 8월 한 달 동안 도시 열섬 현상으로 인해 전력소비량은 평균 3.6%, 최대 7.5%가 증가했다.

열섬현상에 대한 국내 연구는 주로 기상·기후학적인 측면에서 진행되어왔고, 도시와 건축적인 측면 또는 에너지 소비에 관한 연구는 많지 않다. 따라서 본 연구는 데이터 사이언스를 통해 열섬현상의 원인이 되는 도시적 특성들을 분석하고, 도시 내에서도 각 지역의 특성을 구분하여 지역 특성에 알맞은 대책을 제시하고자 한다. 이러한 맥락에서, 본 연구는 도시 설계에 있어서 공공 데이터 및 개방 데이터에 근거한 의사결정을 도입한다는 의미가 있다.

1.2. 연구 범위 및 방법

서울은 대한민국에서 가장 많은 인구가 밀집한 동시에 가장 도시화된 지역이다. 따라서 본 연구는 서울 지역을 대상으로 데이터를 수집하고 분석을 진행하였다.

1.2.1. 데이터 및 변수

본 연구에서는 시간 및 비용 상의 문제로 인해 열섬 현상 관련 데이터를 직접 수집하지 못했다. 분석에 활용된 데이터는 모두 기존에 공개된 자료들을 수집·가

공한 것이다. 최종적으로 풍속, 녹지비율, 건물밀집도, 도로 비율, 해발고도와 토양배수등급을 독립변수로, 지표 온도를 종속 변수로 설정하고 연구를 진행하였다. 최종적으로 사용한 변수들에 대한 설명은 〈표 1〉을 통해 확인할 수 있다. 데이터의 출처와 가공 방식은 다음과 같다.

온도 데이터는 USGS에서 제공하는 LANDSAT 8 위성사진 중 운량이 10% 이하인 가장 최근의 사진을 선택한 후 지표온도 변환 공식에 따라 가공하여 산출했다.¹⁾ 풍속 데이터는 기상자료개방포털에서 제공하는 1998년부터 2010년까지의 지상 10m, 지상 80m에서의 연평균 풍속 데이터를 활용하였다.²⁾ 해발고도 데이터는 국토지리정보원에서 제공하는 것을 활용하였다.³⁾ 배수등급 데이터는 국가공간정보포털에서 제공하는 것을 활용하였다.⁴⁾ 토지피복 데이터 서울 열린데이터 광장에서 제공하는 토지피복도 데이터를 활용하였다.⁵⁾ 서울시 경계 파일은 GIS Developer에서 제공하는 2017년 3월 데이터를 활용하였다.⁶⁾ 미세먼지 데이터는 AirKorea에서 제공하는 데이터를 IDW 보간법으로 가공하여 사용하였다.⁷⁾ 기상자료개방포털에서 제공하는 풍속 데이터가 1km 해상도로 제공되기 때문에, 풍속 데이터의 측정지점 좌표를 중심으로 573개 정사각형

1) <https://glovis.usgs.gov/>

2) <https://data.kma.go.kr/data/weatherResourceMap/selectWeatherResourceMapWdpw.do?pgmNo=107>

3) <http://data.nsdi.go.kr/dataset/20180927ds0069>

4) <http://data.nsdi.go.kr/dataset/12959>

5) <http://data.nsdi.go.kr/dataset/15074>

6) <http://www.gisdeveloper.co.kr/?p=2332>

7) <https://www.airkorea.or.kr/index>

그리드 벡터 레이어를 생성하고, 나머지 지도 데이터들을 그리드 레이어에 맞추어 잘라내서 사용하였다.

1.2.2. 분석 방법

서울시 안에서도 열섬현상이 심각한 지역을 찾아내기 위해서 QGIS 프로그램을 활용하여 서울시를 573개 그리드로 구분하였다. 이후 위에서 언급한 데이터를 가공하여 그리드별로 통계량을 산출하였고, 이를 바탕으로 Python 프로그램을 활용하여 클러스터링을 진행하였다. 이후 각 클러스터의 특성을 분석하고, 열섬현상이 심각한 지역에 대해서는 더욱 자세한 분석을 진행하였다.

이 때 열섬현상의 심각성을 객관적으로 측정하기 위해 도시적 특성을 고려한 열섬 지표표를 개발하였다. 랜덤포레스트 기법을 통해 앞서 설명한 변수들 중 의미 있는 변수들을 선정하고 그에 따른 가중치를 부여하여 특정 지역에서 도시적 특성에 의한 열섬현상이 얼마나 진행되었는지를 판단하였다. 그리고 최종적으로는 각 그룹의 특성을 반영한 완화방안을 제시하였다.

II. 열섬현상에 대한 이론적 논의 및 선행연구

2.1. 열섬현상의 개념

도시화 현상에 의한 열 환경의 변화는 도시 상공의 기온이 주변 지역보다 높아지는 결과를 불러왔다. 일반적으로 도시지역은 등온선에 표기하였을 때 폐쇄된 곡선의 형태로 나타나며, 주변 지역에 비해 1~2℃ 정도 기온이 높다. 주변 지역과 분리되는 모습이 마치 섬 같아 하여 도시열섬(Urban Heat Island)라고 표현한다.

낮에는 도시의 건물들로부터 뿜어져 나오는 열들이 높은 건물들 등으로 인해 방출되지 못한다. 야간에는 축적된 열에 의해 기온이 내려가지 못하고 열대야 현상

을 일으킨다. 이러한 현상은 낮보다는 밤, 여름철보다 겨울철에 더 확연하게 확인할 수 있으며 바람이 약할 때 두드러진다.

2.2. 열섬현상의 원인

열섬현상은 다양한 원인에 의해 발생하고 악화된다. 포괄적으로는 포장된 지표면, 화석연료의 사용과 건물 밀집률을 원인으로 설명할 수 있다.

우선 아스팔트, 콘크리트, 벽돌 등으로 만들어진 건축물과 도로 등은 반사율이 낮기 때문에 태양 에너지를 흡수한 후 저장한다. 따라서 낮 동안 받아들인 열은 발산되지 못하고 밤까지 축적되어 열대야 현상을 일으킨다. 또한 비도시 지역에서는 태양 에너지가 토양수분을 증발시키는 작용을 하지만, 도시에서는 급속한 지표유출로 인해 토양이 물을 저장할 수 없어 건축물과 구조물을 가열하는 작용을 하며 지표면의 온도를 높이는 결과를 불러온다.

도시지역에서는 자동차 배기가스, 공장의 가동, 가정의 연료소비 등이 증가하고 화석연료의 사용량이 증가하면서 대기에 오염물질이 발생한다. 오염물질은 먼지 지붕의 형태를 형성하기 때문에 태양 에너지가 지표면에 닿는 것을 막고 공기의 수직 상승을 방해한다. 또한 도심 주변 상공의 찬 공기에 눌러 공기는 정체되고, 대기 오염 물질 역시 그대로 체류하게 되어 오염 농도가 더 높아진다.

마지막으로 대형건물과 공장 등 건물이 밀집된 지역은 불규칙적인 지면을 형성하여 자연적인 공기의 흐름과 바람의 발생을 차단한다. 또한 인구가 밀집된 지역일수록 인공열이 많이 발생하기 때문에 열섬 현상이 더욱 심화된다.

2.3. 열섬현상 분석 선행연구

국내 도시 열환경 관련 연구의 동향은 도시화에 따른 기온변화에 대한 연구, 토지 이용에 따른 기온 특성에 대한 연구, 도시 열섬의 형성원인에 관한 연구, 도시열

섬에 대한 모델링에 대한 연구 등으로 나타난다.

이강국(2007)은 생산열량과 발열량이 일정한 열평형 상태를 유지하는 범위를 열패적 영역이라고 정의하였고, 도시에너지소비에 의한 인공배열과 토지이용상태에 의한 도시 열환경 분포의 특성에 대해 수치시뮬레이션 및 AWS와 토지피복분류지도를 통해 측정하였다.

차재규, 정응호, 류지원, 김대욱 (2007)은 대구시의 사례를 중심으로 여름철 도시열섬현상을 규명하였다. LANDSAT 5TM 자료를 통해 지표온도를 추출하고 630,000개의 그리드로 나누어 도시열섬현황을 알아보았다. 또한 토지 피복도를 9단계로 나누어 온도에 미치는 영향에 따라 가중치를 부여하였으며, KLAM_21 모델을 사용하여 지형적 특성을 고려한 바람길 시뮬레이션 분석을 수행하였다. 토지피복과의 상관관계를 규명했다는 점에서는 의의가 있었으나, 대구광역시 내에서도 특정 지역만을 대상으로 분석하였고 시계열적인 분석이 진행되어야한다는 점에서 지속적인 연구가 필요하다고 판단하였다.

서지은, 박병용, 김석기, 김태연, 이승복 (2009)은 서울시를 중심으로 열섬현상의 원인을 분석하였다. SPSS를 사용하여 회귀분석을 수행하였고, 인구 수, 건축물 연면적 합계, 자동차 수, 전기 사용량, 도시가스 사용량, 에너지 사용량을 독립변수로 설정하였다. 해당 연구는 도시 포장재질이나 녹지율, 인구분포 등을 고려하지 않았고 더욱 적합한 통계분석이 고려되어야 한다는 시사점을 남겼다.

조혜민, 하재현, 이수기(2015)는 인구, 사회, 경제적 지표들이 집계가 되어있는 행정동 단위로 국지모란계수 값을 활용하여 핫스팟 분석 실행하였다. LANDSAT 8 위성영상과 서울시 열린데이터 광장에서 인구 사회 경제적 변수 활용하여 열섬현상을 분석하였다. 분석 변수는 크게 기후특성, 인구특성, 사회경제적 특성, 토지이용 특성으로 설정하였고 토지이용특성에는 주거시설/상업시설/업무시설 연면적 비율과 용적률, 불투수면 비율, 녹지및오픈스페이스 비율, 수계면적 비율 등이 있었다. 기존에 물리적 공간적 특성만을 고려한 연구에서

더 나아가 인구적 특성에 집중한 연구를 진행하였다.

이들 연구를 바탕으로 도시화에 따른 기온 상승 현상을 확인할 수 있었고, 그중에서도 도시 에너지 소비나 토지이용상태의 변화로 인해 인공열이 증가하였으며 이러한 현상이 도시열섬현상과 밀접한 관련이 있음을 확인할 수 있었다. 또한 시뮬레이션 방식과 회귀분석이 통계적 분석방식으로 주로 사용이 된다는 것을 확인했다. 따라서 본 연구는 새로운 방식의 통계적 분석을 시도했다는 점, 타 지역에도 응용할 수 있는 새로운 열섬 지표를 개발하여 객관적인 열섬완화 위치 선정을 진행했다는 점에서 차별성을 두고 있다.

III. 서울시 열섬현상 발생 지역 분석

3.1. 클러스터링

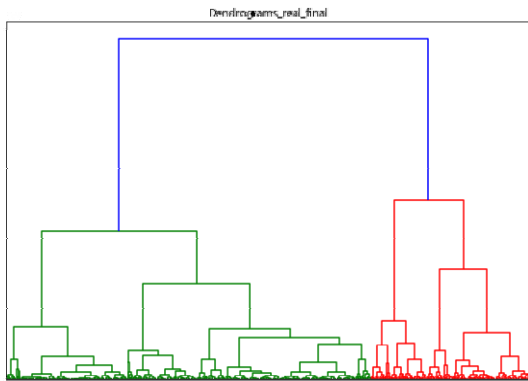
우선 데이터에 클러스터링 기법을 적용하여 각 지역을 분류하고 특성을 파악하였다. 클러스터링이란 유사성에 근거하여 데이터들을 군집으로 묶어내는 방법이다.

본 연구에서는 Python의 `scipy.cluster.hierarchy`와 `sklearn.cluster.AgglomerativeClustering` 패키지를 사용하여 클러스터 간 유사성을 와드 연결법(ward linkage method)으로 측정한 후 계층적 클러스터링을 실행하였다.

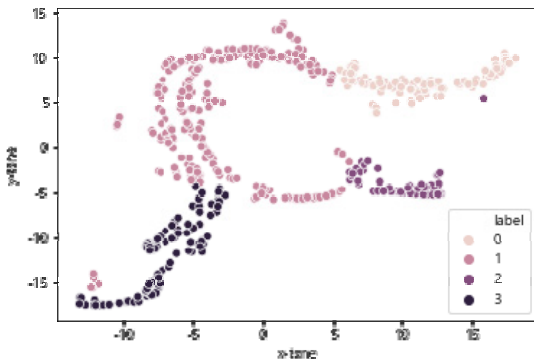
계층적 클러스터링은 사전에 군집 수를 설정하지 않으며, 하나의 데이터를 하나의 군집으로 보고 군집 간의 거리를 계산한 이후 거리가 가장 가까운 군집들끼리 통합하며 계층구조를 형성하는 방법이다. 군집 간의 거리는 와드 연결법으로 측정하였다. 와드 연결법은 군집 내의 오차제곱합에 기초하여 거리를 계산한다. 구체적으로 크기가 각각 $|v|$, $|s|$, $|t|$ 인 군집 v, s, t 가 있을 때 군집 $d(u, v)$ 의 거리 식은 다음과 같이 정의된다. 아래 식에서 군집 u 는 s, t 로 구성된 군집을 말하고 T 는 $|v|$, $|s|$, $|t|$ 의 합을 의미한다.

$$d(u,v) = \sqrt{\frac{|u|+|s|}{T}d(v,s)^2 + \frac{|v|+|t|}{T}d(v,t)^2 - \frac{|v|}{T}d(s,t)^2}$$

〈그림 1〉은 클러스터링 결과를 시각화한 덴드로그램이다. 덴드로그램을 토대로 클러스터 개수를 조정하는 과정을 거친 후, 최종 클러스터 개수는 4개로 결정하였다. 〈그림 2〉는 t-SNE 기법을 적용하여 변수들을 차원 축소한 후 클러스터링 결과를 2차원에 표현한 산점도이다.



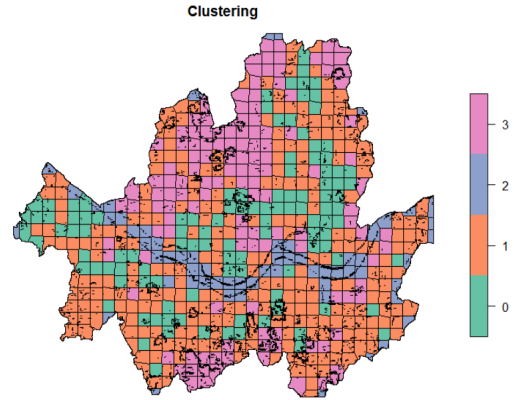
〈그림 1〉 클러스터링 덴드로그램



〈그림 2〉 클러스터링 산점도

3.2. 클러스터링 결과 및 해석

3.2.1. 클러스터링 결과



〈그림 3〉 클러스터링 결과

〈그림 3〉는 클러스터링 결과를 서울시 지도에 시각화한 것이다. 클러스터 2는 한강 부근에, 클러스터 3은 산지 부근에 분포하는 것을 확인할 수 있다. 클러스터 0과 1은 지리적으로 뚜렷한 특성을 보이지 않는다.

〈표 2〉은 각 클러스터의 특성을 평균으로 요약한 결과다. 평균 지표온도(℃)는 클러스터 0(30.68), 클러스터 1(29.54), 클러스터 2(28.74) 클러스터 3(26.71) 순으로 높다.

평균 도로도시비율은 클러스터 0(77.37%)이 가장 높았으며, 클러스터 1(69.46%)과 클러스터 2(71.73%)은 큰 차이를 보이지 않았고, 클러스터 3(49.88%)이 가장 낮았다.

평균 투수양호비율은 클러스터 0(35.71%), 클러스터 1(78.26%), 클러스터 2(79.54%), 클러스터 3(97.67%) 순으로 높아진다.

지상 80m 평균 풍속(m/s)은 클러스터 0(3.16), 클

〈표 2〉 클러스터별 특성

	온도	도로도시 비율	투수양호 비율	투수불량 비율	풍속 (80m)	풍속 (10m)	풍속 차이	평균 미세먼지(10)	평균 고도	건물 면적합	녹지 면적합
0	30.68	77.37	35.21	64.55	3.16	2.7	0.45	51.34	25.31	9.28	6.87
1	29.54	69.46	78.76	21.03	3.22	2.78	0.44	52.74	46.29	7.99	17.56
2	28.74	71.73	79.54	19.31	3.33	2.83	0.5	54.11	32.87	4.28	9.71
3	26.71	49.88	97.62	2.3	3.52	3.06	0.46	52.25	138.45	3.38	45.5

러스터 1(3.22), 클러스터 2(3.33), 클러스터 3(3.52) 순으로 높아지고, 지상 10m 평균 풍속(m/s) 역시 클러스터 0(2.7), 클러스터 1(2.78), 클러스터 2(2.83), 클러스터 3(3.06) 순으로 높아진다.

평균 미세먼지농도는 클러스터 2(54.11)에서 가장 높았고, 다음으로 클러스터 1(52.74), 클러스터 3(52.25), 클러스터 0(51.34) 순이다.

평균 고도는 클러스터 0(25.31), 클러스터 1(46.29), 클러스터 2(32.87), 클러스터 3(138.45) 순으로 높았다.

3.2.2. 클러스터링 결과 해석

〈표 2〉의 결과를 통해서, 각 클러스터의 특징을 요약할 수 있다. 먼저 클러스터 0은 ‘열섬현상이 나타나는 도심 지역’으로 정의할 수 있다. 온도는 네 군집 중 가장 높고, 풍속과 고도는 네 군집 중 가장 낮다. 즉 클러스터 0은 서울 내에서도 다른 지역들에 비해 온도가 높은 지역, 즉 열섬현상이 두드러지는 지역이다. 클러스터 0은 또한 네 클러스터 중 가장 도시적인 특성이 두드러지는 클러스터이다. 서론에서 언급했듯이 여러 연구에서 아스팔트와 도로, 건물 등으로 구성된 지표면이 열섬현상이 원인이 된다고 지적하고 있다. 클러스터 0은 토지피복에서 도로·도시가 차지하는 비중이 네 군집 중 가장 높다. 투수도가 불량 이상인 땅이 차지하는 비중 역시 네 군집 중 가장 높으며, 건물 면적은 네 군집 중 가장 넓고, 녹지 면적은 네 군집 중 가장 작다. 이러한 도시적 특성들이 열을 가두고 공기의 흐름을 막아 열섬현상을 심화시키고 있을 가능성이 크다.

클러스터 1은 특징을 요약하기 어렵다. 평균 온도는 네 클러스터 중 두 번째로 높으며 풍속은 두 번째로 낮다. 그러나 토지피복의 구성에서는 도시적인 특성이 크게 두드러지지 않는다. 건물 면적이 넓은 편이기는 하지만, 동시에 녹지의 면적 역시 다른 클러스터에 비해 넓은 편이다. [그림 2]의 산점도를 통해서 보아도, 클러스터 1은 그 분포에서 뚜렷한 특징을 드러내지 않으며, 다른 클러스터들을 연결하는 중간 위치에 있음을 알 수

있다.

클러스터 2는 [그림 3]을 통해 ‘한강 인근 지역’으로 정의할 수 있다. 온도는 네 클러스터 중 세 번째로 낮고, 풍속과 평균 고도는 두 번째로 높다. 도로도시 비율은 중간 정도이고, 토지의 투수도는 양호한 편이다. 평균 미세먼지 양이 네 클러스터 중 가장 높은데, 한강 인근에 도로가 많고 한강 바람길을 따라 미세먼지가 이동하기 때문일 가능성이 있다.

클러스터 3은 [그림 3]과 [표 2]을 통해 ‘고도가 높은 산지’임을 알 수 있다. 클러스터 3은 평균 온도가 가장 낮고, 풍속과 평균 고도는 가장 높다. 토지는 대부분 투수도 양호 이상이고, 녹지 면적합이 네 클러스터 중 가장 높다.

클러스터 0이 다른 지역에 비해서 도시적인 특성이 두드러지고 평균 온도가 높으므로 이후의 분석은 클러스터 0에 초점을 맞추어 진행하였다.

IV. 랜덤포레스트 모델 변수중요도를 활용한 열섬지표 제안

4.1. 랜덤 포레스트

4.1.1. 랜덤 포레스트

랜덤 포레스트(random forest)는 부트스트랩 표본을 다수 생성하고 의사결정나무(decision tree) 모형을 적용하여 그 결과를 종합하는 앙상블 모델 중 하나이다. 이 때 부트스트랩 표본을 생성하는 부분에서 무작위성이 도입된다. 또한 의사결정나무 모형적합시 각 마디에서의 설명변수 선택에 있어서도 무작위성이 더해진다.

랜덤 포레스트는 예측력이 높은 모형이며, 특히 설명변수가 많을 때 예측력이 매우 높으며 매우 안정적인 모형을 제공한다. 의사결정나무 모델의 단점인 성장, 가지치기 등과 관련된 조율모수가 없다는 점 또한 커다란 장점이다.

설명변수가 많을 경우 비선형 모델인 SVM (Support Vector Machine)을 사용하기도 하나, SVM의 경우 변수들의 관계에 대해 해석하거나 변수의 중요도를 계량화하기에 어려움이 있다. 또한 다범주형(multi-categorical) 자료가 주어졌을 때 숫자형태로 바꾸는 등의 전처리를 필요하고 조정해야할 매개변수들이 다른 방법론들에 비해 많기 때문에 매개변수에 따라 성능이 많이 차이날 수 있다. 따라서 본 연구는 변수들의 특성을 고려하여 랜덤 포레스트 모형을 사용하였다.

4.1.2. 랜덤 포레스트 변수 중요도 지수

본 연구에서는, 열섬의 정도를 하나의 숫자로 표현하여 지역들 간의 비교를 쉽게 하기 위해 열섬 지표라는 지표를 새로 정의하여 사용할 것을 결정하였으며, 열섬 지표 계산에 적용할 가중치 산출을 위해, 종속변수를 온도 변수로 두고, 랜덤 포레스트 모형을 적용하였다. 이후 해당 랜덤 포레스트 모형이 제공하는 변수 중요도 지수를 사용하였고, 공분산이 높은 변수들을 선택하여 제거하였다.

반응변수에 대해 중요한 설명변수를 선택하는 기법으로는 회귀분석이나 의사결정나무 등이 있다. 회귀분석의 경우 설명변수와 반응변수가 선형적인 관계에 있는 모형에서 변수 선택 비율이 높지만, 비선형적인 관계에서는 변수 선택 비율이 낮아진다. 의사결정나무는 위에서 언급하였듯 해석이 쉬우며 변수의 개수에 영향을 덜 받는다는 장점이 있지만, 그 결과가 초기의 분할에 크게 좌우되고, 변수 간의 교호작용이 지나치게 강조된다.

본 연구에서 활용된 데이터들은 선형회귀의 기본 가정들을 충족하지 못하고, 설명변수와 반응변수 간에 비선형적 관계가 많다. 즉 선형회귀 기법을 통해서 변수를 선택하는 경우 편향이 생기기 쉬운 데이터이다. 따라서 의사결정나무 기반의 변수 선택 기법을 사용하였고, 의사결정나무가 갖는 단점들을 보완하기 위해서 랜덤포레스트 모형을 통해 변수를 선택하였다. 랜덤포레

스트는 여러 개의 나무를 훈련시켜서 합치는 앙상블 기법을 사용하여 이러한 의사결정나무의 단점을 보완하였으며 우수한 정규화 성능을 가진다.

랜덤 포레스트는 의사결정나무 모형에서와 같이 직관적인 그래프가 최종모형으로 도출될 수 없기 때문에 최종 결과에 대한 해석이 어렵다. 따라서 랜덤 포레스트에서는 이러한 문제를 해결하기 위한 방안으로 변수의 중요도지수를 제시함으로써 설명변수의 반응변수에 대한 중요도(영향력)를 숫자 또는 그래프로 알아볼 수 있도록 제공한다.

랜덤 포레스트에서는 변수 중요도 지수 산출시에 원 OBB 자료의 표준오차와 수정된 OBB 자료의 분류오차 차이 평균을 표준 오차로 나눈 값을 해당 변수의 중요도로 사용하며, 변수가 음의 방향 혹은 양의 방향으로 영향을 미치는 가에는 상관없이 종속 변수에 미치는 변수의 영향력이 클수록 변수 중요도 지수는 커진다.

4.2. 열섬 지표 제안

4.2.1. 열섬현상에 대한 변수 중요도 지수

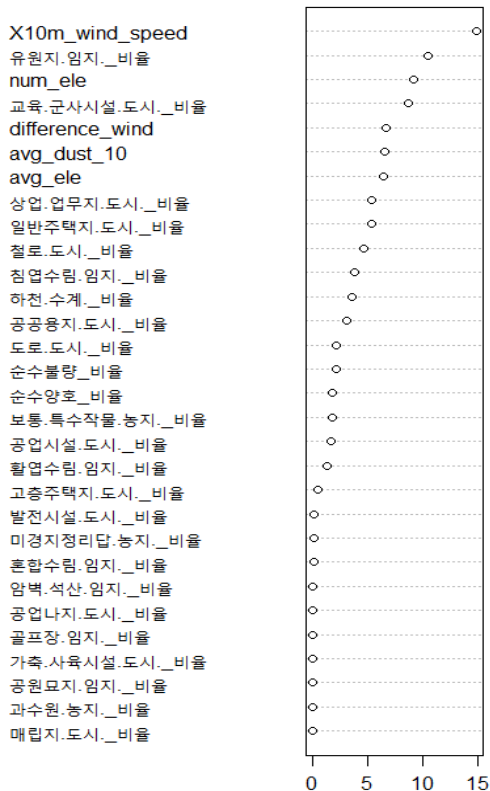
앞서 언급하였듯, 본 논문은 기존의 연구들이 열섬현상 분석을 위해 주로 회귀분석 모델을 사용한 것과는 달리 열섬 지표를 새롭게 정의함으로써 열섬현상을 측정하였다는 점에서 차별성을 가진다.

본 연구에서는 랜덤 포레스트 분석을 위해 R의 randomForest 패키지를 사용하였으며, <표 3>와 같이 파라미터를 설정하여 모델을 실행하였다. 모델의 성능을 573개 그리드에 대해 5중첩 교차검증한 결과, 평균 0.6의 결정계수와 2.38의 RMSE를 얻었다.

〈표 3〉 랜덤 포레스트 파라미터

파라미터	설명	값
ntree	트리 개수	500
mtry	변수 개수	p/3
replace	복원 추출 여부	TRUE
samplesize	표본 크기	nrow(x)
nodesize	마지막 노드 크기	5
maxnodes	최대 노드 개수	NULL(제한 없음)

랜덤 포레스트 모델 실행 결과 산출된 변수 중요도 지수를 토대로 분석하면, 고도 10m에서의 풍속이 다른 변수들에 비해 가장 큰 변수 중요도 지수를 보이고 있으며, 다음 차례로 임지 비율, 교육/군사시설의 비율, 고도 10m와 고도 80m에서의 풍속 차 등의 요소가 온도 변수에 중요하게 작용함을 알 수 있다.



〈그림 4〉 랜덤 포레스트 변수 중요도

반면, 매립지 비율, 과수원 비율, 공원묘지 비율, 가축/사육시설 비율 등의 변수는 변수 중요도 지수가 0에 가까운 값을 가지고 있으므로, 온도 변수에 거의 영향을 주지 않는 요소임을 확인할 수 있다.

4.2.2. 열섬 지표 산출

변수 중요도를 이용하여 열섬 지표를 산출하기 이전, 값의 범위가 더 큰 변수가 변수 중요도와는 상관없이

더 큰 영향을 미치게 되는 경우를 방지하기 위해 우선 연속형 변수들의 값을 Min-Max Standardization을 이용하여 0과 1사이의 숫자 값으로 스케일링 하였다.

또한 변수에 대해 부여할 가중치는, 가장 높은 중요도를 가지는 변수의 중요도로 해당 변수의 중요도를 나눈 값에 절댓값을 취한 다음, 온도와 상관계수에 따라 양수, 혹은 음수의 부호를 취하여 사용하였다.

예를 들어, 가장 높은 변수 중요도 지수를 가지고 있는 고도 10m에서의 풍속의 변수중요도 지수가 15이고, 여러 변수 중 하나인 유원지 비율의 변수 중요도가 12일 때, 유원지의 비율 변수에 대해 부여할 가중치의 절댓값은 $12/15 = 0.8$ 이며, 유원지 비율과 온도는 음의 상관관계를 가지고 있으므로 음(-)의 부호를 붙여주는 방식으로 가중치를 설정하였다.

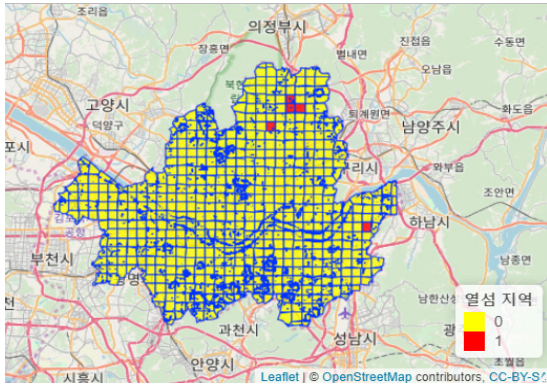
이 때, 변수 중요도가 음수인 변수들의 경우, 오히려 모델의 설명력을 떨어뜨리는 변수들이므로, 결과의 정확성을 위해 가중치를 0으로 두고 해당 변수는 반영하지 않았다.

이러한 방식으로 계산되는 열섬 지표를 서울시 그리드 각각에 적용하여 각 그리드에서의 열섬 지표를 산출하였다.

4.3. 열섬 지표를 적용한 분석

4.3.1. 열섬 지표 적용 결과

다음으로, 앞서 언급하였듯 도시적 특성이 두드러지는 클러스터 0에 열섬 지표를 적용하여 다른 지역보다 그 값이 높은 5개 지역을 선정하였다(〈그림 5〉에 표시된 지역). 결과적으로 그리드 ID 300번(둔촌동, 길동, 성내동), 381번(변동, 수유동, 미아동), 403번(상계동, 방학동, 창동), 408번(창동), 431번(상계동) 지역이 선정되었으며, 법정동 상으로 상계동, 창동, 방학동, 변동, 수유동, 미아동, 둔촌동, 길동, 성내동이 해당한다. 지리적으로는 403, 408, 431번 그리드가 서로 인접해 있고, 300번 그리드는 다른 열섬 지역으로부터 비교적 멀리 떨어져 있다.



〈그림 5〉 열섬지표 상위 5개 지역

4.3.2. 열섬 지역 특성

열섬 지역으로 분류된 그리드의 특성을 [표 3]에 정리하였다. 먼저 300번 그리드의 평균 온도는 32℃이며, 해당 지역의 열섬현상은 햇빛을 효과적으로 반사해내지 못하는 토지피복(도로 및 건물)에 기인하는 것으로 보인다. 300번 그리드는 도로의 비율이 75.58%로 높은 편이고 나머지 24.42% 지역은 건물이 차지하고 있다. 지도상으로 확인한 결과 건물은 대부분 주거단지와 상가로 이루어져 있다. 해당 지역은 또한 미세먼지 농도가 매우 높다. 도로가 차지하는 비율이 높다는 사실과 관련 지어 생각할 때, 자동차 통행량으로 인한 현상일 가능성이 크다. 즉 자동차로부터 발생하는 열 또한 해당 지역의 열섬현상을 악화시키는 원인일 가능성이 있다.

381번 그리드의 평균 온도는 31.88℃이고, 평균 고도가 42.95로 높은 편임에도 불구하고 열섬 지역으로 분류되었다. 300번 그리드와 마찬가지로 햇빛을 효과적으로 반사해내지 못하는 토지피복(도로 및 건물)이 열섬

현상의 원인인 것으로 보인다. 녹지는 거의 존재하지 않고, 풍속 역시 낮은 편이다. 건물 면적이 특히나 넓다는 사실을 고려할 때, 낮은 풍속은 공기의 흐름을 방해하는 높은 건물 밀도에 기인할 가능성이 있다.

403, 408, 431번 그리드는 각각 평균 온도가 34.51℃, 33.27℃, 32.67℃로 앞서 살펴본 300, 381번 그리드에 비해서 온도가 높은 편이다. 토양의 투수도가 비교적 불량한 편이지만 건물과 도로가 차지하는 면적은 앞의 두 지역에 비해 작은 편이다. 지도를 통해 확인한 결과 해당 지역에는 철로, 도로, 공공용지, 공업시설, 교육군사시설 등이 주로 분포하고 있었으며, 창동차량기지와 도봉 먼지시험장이 위치한다. 즉 해당 지역은 토지피복의 특성에 더해서 철도, 자동차, 공업시설 등 열을 방출하는 여러 요인들이 복합적으로 작용하여 열섬현상이 발생했을 가능성이 크다.

V. 서울시 바람길 및 녹지 입지 선정 제안

도로의 비중이 높은 경우, 도로 주변에 녹지를 조성하여 온도를 낮출 수 있다. 혹은 도로를 설계할 때 일반적인 도로중앙분리대 대신 녹지형 중앙분리대를 설치할 경우 온도를 효과적으로 낮출 수 있다. 실제로 대전시의 경우 녹지형 중앙분리대가 조성된 구간과 그렇지 않은 구간 4곳의 지표면 온도를 측정한 결과 평균 섭씨 10.86도의 차이가 났다. 이에 착안하여 열섬지표가 높

〈표 4〉 열섬지역 특성

ID	온도	도로 비율	투수양호비율	투수불량비율	풍속 (80m)	풍속 (10m)	미세먼지 (10)	평균고도	건물 면적합	녹지 면적합
300	32	75.58	51.3	48.7	2.86	2.51	61.5	23.48	24.42	0
381	31.88	59.35	21.44	78.56	2.54	2.24	52.5	42.95	40.3	0
403	34.51	55.39	15.9	84.1	2.65	2.42	52.5	24.45	24.62	3.01
408	33.27	69.11	26.2	73.8	2.57	2.32	53	24.08	17.97	0.55
431	32.67	69.26	15.2	84.8	2.62	2.31	50.5	25.77	19.24	0.04

은 위의 5곳 중 도로의 비율이 높지만 투수양호 토지의 비율 역시 높은 300번 그리드(둔촌동, 길동, 성내동)의 경우, 도로 양옆에 녹지를 조성하면 온도를 효과적으로 낮출 수 있을 뿐만 아니라 미세먼지 역시 흡수할 수 있을 것으로 기대된다. 또한, 도로를 설계할 때 녹지형 중 양분리대를 설치하면 열섬완화에 큰 역할을 할 수 있을 것으로 보인다.

하지만 건물 비중과 불투수토지의 비중이 높은 경우 녹지를 조성하기 어렵다. 실제로 300번 그리드를 제외한 나머지 4개(381번, 403번, 408번, 431번)의 그리드는 투수불량 토지의 비율이 모두 70%이상이다. 이러한 경우에는 건축물의 옥상에 관목, 교목 등의 수목을 심는 옥상녹화를 통해 효율적으로 온도를 낮출 수 있다. 투수불량 토지의 비중이 높은 위의 4곳 중 특히 투수불량이 높고 건물 면적합 역시 높은 403번 그리드와 건물 면적합이 독보적으로 높은 381번 그리드의 경우 건물 옥상에 녹지를 조성하는 것이 열섬을 완화하는 데에 큰 기여를 할 수 있을 것으로 보인다.

앞에서 열섬지표를 정의할 때 따르면 풍속은 열섬에 가장 큰 영향을 주는 요소이다. 실제로 열섬지표가 높은 5개의 그리드 모두 풍속이 약한 것을 볼 수 있다. 이 중 381번, 403번, 408번, 431번 그리드의 경우 북한산 및 인근 산지와 인접한 지역들이지만 온도가 높다. 이는 북한산 및 인근 산지의 차가운 공기가 내려오지 못하는 것이 원인이라고 추측해볼 수 있다. 산으로부터 내려오는 찬 공기가 효과적으로 주변 지역에 유입되도록 바람길을 조성한다면, 주변지역의 온도가 크게 낮아질 것이다. 독일 슈투트가르트가 효과적인 바람길 조성의 대표적인 예시라고 볼 수 있다. 독일 슈투트가르트는 도시 주변을 산이 둘러싼 분지의 형태이다. 이처럼 대기가 순환하기 어려운 구조 때문에 공업의 중심지인 슈투트가르트는 심각한 대기오염에 시달렸다. 이를 막기 위해 슈투트가르트는 주변 산지에서 생성되는 차고 신선한 바람이 유입되도록 광역종합계획과 세부지침을 수립하였다.



〈그림 6〉 독일 슈투트가르트의 바람길 사례

슈투트가르트에는 3개의 바람길이 존재하는데, 이 3곳의 경우 고층건물은 물론 주거단지, 심지어 키 큰 가로수 역시 심을 수 없도록 엄격히 규제하고 있다.

독일의 슈투트가르트의 바람길에 착안하여, 위의 4개(381번, 403번, 408번, 431번)의 그리드에 큰 나무를 밀도 있게 심어 신선하고 차가운 공기가 고이는 공기매를 만들고 강한 공기의 흐름을 확산시켜, 북한산 및 인근 산지의 공기가 효과적으로 유입되게 할 수 있을 것이다.

Ⅵ. 결 론

본 논문에서는 서울시의 도심 열섬 현황을 파악하기 위해 계층적 클러스터링 기법과 랜덤포레스트를 기법을 이용하여 열섬지표를 개발하였다. 먼저, 클러스터링을 통해 서울시를 크게 4개의 클러스터로 나누었고, 이 중 도심의 성격을 강하게 띠는 클러스터를 골랐다. 해당 클러스터를 대상으로 랜덤포레스트를 통해 변수의 중요도를 추출한 후, 추출한 중요도와 온도와 해당 변수간의 상관관계 부호를 곱하여 해당 변수의 가중치로 부여하였다. 이러한 변수들의 가중치를 바탕으로 열섬지표를 만들었다. 이 열섬지표에 따르면 5개의 그리드가 열섬현상이 심각하다.

데이터의 선형성, 결과 해석 등의 문제로 인해 열섬 현상을 모델링하는 과정에 랜덤포레스트만 모형만을 사용하였다는 것은 본 연구의 한계점이다. 열섬현상에 더욱 다양한 모델을 적용해볼 수 있다면 더욱 정확한 분석이 될 수 있었을 것이다. 사용된 데이터의 수집 시점이 모두 다르다는 것 역시 본 연구의 한계점이다. 다양한 기관에서 수집한 데이터를 이용하다 보니 수집 시점이 모두 달랐는데, 데이터의 시점이 일치했다면 더욱 정확한 분석이 되었을 것이다.

그럼에도 불구하고 본 연구는 몇 시사점을 갖는다. 첫째, 구나 동단위로 열섬현상을 분석했던 기존의 연구들과 달리 공공 데이터를 적극적으로 가공 및 활용하여 500여 개 그리드에 대해 더 자세한 분석을 하였다는 점, 그리고 도시적 특성을 고려한 열섬지표를 새롭게 제시하였다는 점에서 의미가 있다. 열섬현상은 매년 심각해지고 있으며, 이에 대해 구체적인 해결책이 요구되지만, 구나 동 단위의 분석에서는 충분히 구체적인 수준의 해결책을 마련하기 어렵다. 서울이 아닌 다른 지역에서도 열섬현상에 대해서 구체적인 해결책이 마련되어야 할 것이며, 이 과정에서 본 연구에서 제시한 모형과 지표를 참고할 수 있을 것이다.

본 연구는 또한 도시 설계에 있어서 공공 데이터의 활용 및 데이터 기반 의사결정의 가능성과 필요성을 보여준다. 정부와 지자체, 기관이 아무리 좋은 품질의 데이터를 가졌다고 하더라도 하나의 데이터로는 다양한 분석을 시도하기 어렵다. 따라서 본 연구에서는 다양한 공공 데이터를 통합하고 가공하여 더 넓은 시각에서의 분석을 진행하였고, 이를 바탕으로 구체적인 해결책을 제시할 수 있었다. 그러나 위에서 언급하였듯 본 연구에서 사용된 데이터의 시점이 모두 달라 분석에 한계가 있었다. 양질의 공공 데이터가 다양한 분야에서 더 많이 생산된다면, 이러한 문제점들을 해결할 수 있을 것이다.

참고문헌

- ‘열섬 현상’이 낳은 악순환... 에너지 소비 늘고 오염 심해져, 중앙일보, <<https://news.joins.com/article/22673702>>, 2018. 05.
- 조혜민, 하재현, 이수기 (2015). 도시열섬현상과 사회적 취약 계층의 공간적 분포 특성분석. 한국도시행정학회 학술발표대회 논문집, 129-148.
- ‘열섬 현상’이 낳은 악순환... 에너지 소비 늘고 오염 심해져, 중앙일보, <<https://news.joins.com/article/22673702>>, 2018. 05.
- 이강국 (2007). 도시 열환경 분포 특성 분석과 열섬현상 완화 방안 연구, 경북대학교 박사학위논문.
- 유기정 (2008). 바람길 계획을 통한 도시열섬현상 완화에 관한 연구. 한양대학교 석사학위논문.
- 유기정 (2008). 바람길 계획을 통한 도시열섬현상 완화에 관한 연구. 한양대학교 석사학위논문.
- 이강국(2007). 도시 열환경 분포 특성 분석과 열섬현상 완화 방안 연구. 경북대학교 박사학위논문.
- 이강국(2007). 도시 열환경 분포 특성 분석과 열섬현상 완화 방안 연구. 경북대학교 박사학위논문.
- 차재규, 정응호, 류지원, 김대욱 (2007). 도시열섬현상 완화를 위한 녹지네트워크 및 바람길 구축. 한국지리정보학회지 10권 1호, 102-112.
- 서지은, 박병용, 김석기, 김태연, 이승복 (2009). 도시지역의 열섬현상에 영향을 미치는 요소 분석. 대한건축학회 학술발표대회 논문집 - 계획계, 29(1), 581-584.
- 조혜민, 하재현, 이수기 (2015). 도시열섬현상과 사회적 취약 계층의 공간적 분포 특성분석. 한국도시행정학회 학술발표대회 논문집, 129-148.
- 손지영 (2018). word2vec을 이용한 거리 기반의 음악 가사 클러스터링 기법. 숭실대학교 석사학위논문.
- 유진은 (Yoo Jin Eun).(2015). 랜덤 포레스트. 교육평가연구, 28(2): 427-448.
- 임근옥 (2005), 변수선택을 통한 SVM의 분류성능 향상에 대한 비교연구, 서울대학교 대학원.
- Breiman, L. (2001), Random forests. Machine learning, 45(1), 5-32.

- 김지연 (2014), Multiclass classification methods 활용한 데이터 분석, 이화여자대학교 대학원.
- 권안나. (2013). 랜덤포레스트를 이용한 변수 선택. 석사학위논문. 인하대학교. 인천.
- 유진은(Yoo Jin Eun).(2015). 랜덤 포레스트. 교육평가연구, 28(2): 427-448.
- Cha, H.-J., & Lim, J.-H. (2011). The Effect of Urban Road Vegetation on a Decrease of Road Surface Temperature. Journal of the Korean Institute of Landscape Architecture, 39(3), 107116.
- 손규성, ‘녹지형 중앙분리대 도심온도 10도 낮춰’, 한겨레신문, 2008.08.
- 김예슬, 이빛나라, 이은희. (2014). 제3분과 : 하천 및 도시녹지 ; 관리조방형 옥상녹화 적용 초화류의 온도저감 효과. 한국환경생태학회 학술발표논문집, 2014(1), 62-63.
- 김영화, ‘[기획] 찾아라 에너지 블루오션, 신재생에너지가 미래다. 최악의 오염도시 슈투트가르트, 바람길로 환경도시로 우뚝’, 한산신문, 2012.01.
- 성선용, 박종순, 이상은, 김선희. (2019). 미세먼지 저감을 위한 도시 내 바람길 도입 방안. 국토정책 Brief, (), 1-6.

논문접수일: 2019년 09월 11일
1차 수정일: 2019년 10월 19일
게재확정일: 2019년 10월 20일