

머신러닝기반 쇠퇴도시 예측 연구: 그래디언트 부스팅(Gradient Boosting) 모형을 중심으로*

Predicting Urban Decline Using Gradient Boosting Algorithm

정지은**
Jeong, Ji-Eun

Abstract

This study aims to analyze the spatial patterns of urban decline and develop a model for predicting urban decline using machine learning algorithms. Using Census data from the 2005 and 2010 Jipgyegu levels (called Micro Census data), we employ the Getis-Ord G_i^* , a spatial measurement tool, to detect the spatial patterns of urban decline. This study applies a gradient boosting algorithm to an urban decline prediction model using aggregated Eup-Myeong-Dong data from Jipgyegu levels. The 2005-2010 data are used to train, and the 2010-2015 data are used to test the machine learning model.

Key findings from our empirical analyses are as follows. First, we observe different spatial distributions between Jipgyegu and Eup-Myeong-Dong levels, which implies that Jipgyegu level data would better explain the current situation for determining the policy. Second, an urban decline prediction model based on a gradient boosting algorithm shows that interregional migration, number of birth, and the ratio of dilapidated dwellings are the highest contribution to urban decline, which indicates the important factors predicting an urban decline. Third, the performance of a prediction model reveals that Accuracy is 0.69, Precision is 0.63, Recall is 0.99, and f1-score is 0.77.

This study will provide an understanding of the urban decline prediction model based on machine learning algorithms and the necessity to elaborate the prediction model by adding a variety of variables to explain the urban decline. We expect it will be used as a basis for urban regeneration projects and policy priorities if we add more sophisticated urban decline indicators to improve the prediction model.

주제어: 쇠퇴도시, 머신러닝, 그래디언트 부스팅 알고리즘, 공간계량분석, 인구쇠퇴

Keywords: Urban Decline, Machine Learning, Gradient Boosting Algorithm, Spatial Statistics Analysis, Population Decline

* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A5B5A07107324)

** 중앙대학교 도시계획부동산학과 박사(주저자: jjeun1223@gmail.com)

I. 서론

1. 연구의 배경 및 목적

도시정책의 패러다임이 개발과 성장 중심에서 삶의 질 향상으로 전환되면서 도시개발보다는 도시재생정책에 대한 수요가 증가하고 있다. 2013년 정부는 「도시재생 활성화 및 지원에 관한 특별법」을 제정하고, 도시재생을 위한 다양한 사업들을 계획, 운영하고 있다. 또한 도시정보 종합체계를 구축해서 지자체 및 연구기관, 그리고 민간에 도시쇠퇴와 관련한 다양한 데이터를 제공함으로써 연구를 활성화 하고 지방정부의 정책 의사결정을 지원하는 역할을 하고 있다.

도시쇠퇴 문제는 저출산과 고령화 등 인구 구조적 변화와 맞물리면서 더욱 중요한 문제로 인식되고 있다. 도시쇠퇴는 인구, 산업 등과 같은 도시 구성요소들의 축소를 의미하며, 이는 다양한 경제적, 물리적, 사회적 문제들을 양산시키고 이러한 문제들이 악순환 하는 과정을 거친다. 일반적으로 도시쇠퇴는 도시의 일자리 감소, 인구유출, 젊은 인력의 유출로 인한 출산율 저하와 고령화 양산, 도시 내 토지이용의 둔화와 시설물 낙후, 경제활동의 침체 등이 악순환 되면서 심화되는 양상을 보인다(서준교, 2014).

지방정부는 도시쇠퇴 문제를 해결하기 위해 다양한 재생정책을 마련하고 있지만 지역적 특성과 상황이 다르고 도시환경이 계속 변화하기 때문에 지속적이고 체계적인 대응책을 마련하는데 한계가 있다. 재생정책 수립 시 가장 기초가 되는 쇠퇴지표 역시 도시재생 종합정보체계를 통해 도시쇠퇴진단지표가 매년 발표되지만 단순히 지표만으로 각 지방정부가 다양한 원인으로 발생하는 복잡한 도시의 쇠퇴현상을 진단하고 대응하기에는 한계가 있다. 더군다나 계속해서 변화하는 도시환경속에서 체계적으로 대응하기 위해서는 도시의 쇠퇴현상을 미리 예측할 필요가 있다.

최근 머신러닝방법이 도입되면서 사회과학 분야에도 활용되고 있으며 대표적으로 범죄발생예측 모형, 지하철 승하차 인원 예측모형, 부동산 가격 및 지가변동 예측모형 등이 있다. 이러한 모형 분석결과는 상당히 높은 예측력을 보여주며 더 다양한 사회과학 분야의 연구에 도입되어 활용할 수 있는 가능성을 보여준다.

본 연구는 우리나라 전국을 대상으로 쇠퇴도시 지표와 관련된 GIS기반 도시 빅데이터를 구축하여 쇠퇴도시의 공간적 분포패턴을 분석하고 머신러닝을 활용하여 쇠퇴도시 결정요인 분석 및 예측을 목표로 한다.

2. 연구의 범위 및 방법

본 연구는 도시쇠퇴현상의 공간적 분포 패턴 분석 및 머신러닝을 활용한 도시쇠퇴 예측모형 구축을 위해 공간적 범위를 전국으로 설정하며, 시간적 범위는 2000년-2015년을 대상으로 한다. 본 연구는 크게 네 단계로 이루어진다. 첫째, 쇠퇴도시와 관련된 이론과 기존문헌 검토를 통해 쇠퇴도시를 정의하고 쇠퇴도시 진단지표를 도출한다. 둘째, 기존문헌을 통해 선정된 변수를 기준으로 GIS기반 DB를 구축하고 공간통계기법을 이용하여 쇠퇴도시의 공간적 분포패턴을 분석한다. 공간적 분포패턴 분석은 인구센서스의 최소 집계단위인 집계구 자료를 이용하여 집계구를 분석단위로 한다. 셋째, 머신러닝을 활용하여 쇠퇴도시 결정요인을 분석하고 예측모형을 구축한다. 그래디언트 부스팅 알고리즘을 이용한 모형을 구축하고 예측 모형의 검증 및 결과를 해석한다. 넷째, 분석결과를 바탕으로 정책적 시사점을 제시한다.

II. 이론 및 선행연구 검토

1. 도시쇠퇴의 정의

도시쇠퇴는 상대적이고 추상적이며 주관적이어서 이에 대한 정의를 분명하게 내리기는 쉽지 않다. 초기에는 도시쇠퇴의 중요 요소를 물리적 요소와 경제적 요소로 접근한 반면, 시간이 흐름에 따라 환경 및 사회적 요소들이 포함되어 왔다. 이에 따라 최근에는 도시쇠퇴를 인구·사회, 산업·경제, 물리·환경적 측면에서 활력을 잃어가는 상태로 용어의 정의가 구체화 되고 있다(전호진 외, 2018).

정부는 무분별한 공간적 확산 및 도시쇠퇴현상을 해결하기 위해 2013년 「도시재생 활성화 및 지원에 관한 특별법」(이하, 도시재생법)을 제정하고, 도시재생 활성화지역을 지정하려는 경우 인구 감소, 사업체 감소, 노후주택 증가 중 2개 이상의 요건을 충족해야 함을 기준으로 명시하고 있다. 그 세부 기준은 도시재생법 시행령 제17조에 제시되어 있으며, 이는 도시쇠퇴 현황 및 진단에도 활용되고 있다. 도시재생법 지정 이후의 다수 연구에서는 아래의 지표를 기준으로 쇠퇴지역을 판단하고 있다.

표 1. 도시재활성화지역 지정의 세부 기준(도시재생법 시행령 제17조)

기준	지표
인구 감소	최근 30년간 인구가 가장 많았던 시기와 비교하여 20% 이상 인구가 감소한 지역 최근 5년간 3년 이상 연속으로 인구가 감소한 지역
사업체 감소	최근 10년간 총 사업체 수가 가장 많았던 시기와 비교하여 5%이상 총 사업체 수가 감소한 지역 최근 5년간 3년 이상 연속으로 총 사업체 수가 감소한 지역
노후주택증가 (주거환경악화)	전체 건축물 중 준공된 후 20년 이상 지난 건축물이 차지하는 비율이 50%이상인 지역

출처: 국가법령정보센터 내용을 토대로 표로 작성함

도시쇠퇴를 측정하는 방법은 하나의 개별지표를 통한 측정방법과 각 지표를 표준화시켜 복합쇠퇴지수를 산출하는 방식이 있다(엄현태·우명제, 2014). 복합쇠퇴지수는 해당 도시가 얼마나 쇠퇴했는지에 대한 정량적 수치를 제공해 준다는 장점이 있으나, 지수만으로는 도시에 어떤 재생전략을 제시 할지 제언이 어렵다(허재완·이삼수, 2015). 따라서 최근 다수의 연구에서는 개별지표를 이용한 방식을 사용하고 있으며, 측정지표로 도시재생법에서 지정하고 있는 도시재생활성화지역 지정기준인 인구, 사업체수, 준공된 후 20년이 지난 노후주택 비율을 활용하고 있다(엄현태·우명제, 2016; 허재완·이삼수, 2015; 김예슬·김형보, 2015).

허재완·이삼수(2015)의 연구는 기존문헌검토를 통해 쇠퇴의 원인이 다양하며 여러 원인이 복합적으로 나타날 수 있기 때문에 다양한 지표를 활용하고 있음을 밝혔으며, 1) 인구사회부문, 2) 산업경제부문, 3) 물리환경부문에 구분하여 각 부문에서 대표 지표들의 사용빈도를 분석하였다.

본 연구에서는 도시재생법에 따른 도시재생 활성화지역 지정 기준인 인구감소, 사업체감소, 노후주택증가(주거환경악화)를 기준으로 도시쇠퇴를 정의한다. 이에 따른 도시쇠퇴 지표를 이용해 공간적 분포패턴을 분석하고 도시쇠퇴 예측 모형을 구축 하고자 한다.

2. 도시쇠퇴 관련 연구동향

우리나라는 성남, 일산 등 1기 신도시 지역으로의 대규모 인구이동이 시작된 1990년대 후반부터 도시쇠퇴에 대한 논의가 본격적으로 시작되었다(이상대, 1996). 초기의 연구들은 한국적 맥락의 도시쇠퇴 원인과 특성을 파악하는 연구들이 수행되었으며(김광중, 2010; 오윤표·강진학, 2002; 장희순·송상열, 2006; 김광중·박현영·김예성·안현진, 2010; 서울대학교 산학협력단, 2010a), 이후 도시쇠퇴를 어떻게 진단할 것인가에 대한 논의 속에 도시쇠퇴 지표 및 지수개발에 관한 연구가 다수 진행되었다(조진희·이동건·황희연, 2010; 이영성·김예지·김용욱, 2010; 조윤애, 2014; 서울대학교 산학협력단, 2010b).

또한 도시쇠퇴의 지표에 관한 연구가 진행됨과 동시에 도시재생방안 및 정책 지원을 위한 쇠퇴지표 적용 및 이에 따른 도시쇠퇴 양상을 분석한 연구가 다수 진행되었으며, 이는 지방 대도시, 지방 중소도시, 지방 소도시, 중심도시 및 구시가지 등 다양한 도시단위별 지역들을 대상으로 하고 있다(조수희·정재호, 2010; 엄상근·남윤섭, 2014; 박소연·오덕성, 2015; 엄현태·우명제, 2014; 강인호, 2019; 장문현·이민석·김화환, 2016).

그리고 최근에는 빅데이터의 접근이 상대적으로 용이해지면서 GIS를 활용한 공간통계분석으로 도시쇠퇴의 공간적 패턴을 분석하는 연구도 증가하고 있으며, 이는 행정구역 단위뿐만 아니라 근린지구, 필지 및 건축물, 집계구 등 다양한 하위레벨의 공간적 차원을 연구 대상으로 하고 있다(장문현, 2016; 임현성·김충호, 2019; 엄현태·우명제, 2016; 이민석, 2018; 이종휘·이태희, 2020).

임현성·김충호(2019)는 도시쇠퇴 진단체계와 도시재생전략 수립절차를 연계하는 과정을 제안하면서 기존의 행정동 단위의 쇠퇴도시 진단이 아니라 시가화구역의 변천 현황 등 도시구조를 우선 파악하고, 그러한 맥락 하에서 점도조건 필지형상, 노후건축물, 빈집분포 등 위치 기반의 필지 및 건축물 차원에서 도시쇠퇴지역을 도출할 것을 제안하였다. 부산시 부산진구의 사례로 실증분석한 결과는 기존의 쇠퇴도시 진단과 도시 구조 및 위치기반의 지표들을 적용한 결과가 다르게 나타남을 보여주었다. 장문현(2016)은 광주광역시를 대상으로 기존 쇠퇴지표를 공간데이터로 구축 및 시각화 하고, 공간통계기법을 적용하여 쇠퇴지역이 밀집한 공간적 범역과 군집의 강도를 통계적으로 도출 하였다. 그리고 이러한 연구결과는 대도시권 수준에서의 도시쇠퇴 진단에 기여할 것으로 제안하였다.

이처럼 다수의 연구가 도시쇠퇴 지표 적용을 통해 도시쇠퇴 실태 분석과 도시별 재생관리방안을 도출하는데 집중하는데 반해 전경구·전형준(2016)은 도시쇠퇴가 도시관리비용 증가에 영향이 크다는 점을 실증 분석하였고, 권오규·강은택·마강래(2019)는 도시쇠퇴로 인해 주민들의 삶의 만족감이 감소함을 실증 분석을 통해 밝혀냈고, 특히 이는 도시 규모별로 차이가 있음을 확인하였다.

3. 기계학습 모형 연구동향

머신러닝(Machine Learning), 또는 기계학습은 인공지능의 한 분야로 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야를 일컬으며, 머신러닝에 대한 연구는 최근 4차 산업에 대한 관심 증가와 빅데이터의 활용과 맞물려 증가하고 있는 추세이다. 국내에서는 금융분야(윤종문, 2019; 허주성·권도형·김주봉·한연희·안채현, 2018), 보건·환경·복지분야(한은정·김동건, 2015; 오미애 외, 2017; 한국환경정책·평가연구원, 2018) 등 다양한 분야에 적용되고 있다. 그리고 정책적으로 활용된 사례로 보건복지 분야에서는 빅데이터를 활용한 기계학습 기반 예측모형을 통하여 동절기 복지 사각지대 위기아동 발굴 시스템을 개발, 지원한 사례가 있다(오미애 외, 2017).

도시와 관련한 분야는 부동산 가격 및 지가 예측 연구(배성완·유정석, 2018; 원석환·이창규·박지만, 2017) 외에도 범죄발생 위험지역 예측(허선영 외 2018), 지하철 승차인원 예측(조수진 외, 2019),

보행 만족도 예측(이제승·이현희, 2019; 박근덕·이수기, 2018), 토지이용변화 예측(Jun, 2021; 노승철·김태현·이석준·임재욱·고진수, 2020) 등 부동산, 주택, 교통, 안전 등 다양한 연구가 있으며, 이러한 연구결과는 도시계획 분야에도 머신러닝 기법을 적용하고 활용할 수 있음을 보여준다.

이제승·이현희(2019)의 연구는 서울시 유동인구 조사자료를 이용하여 보행만족도 예측모형을 개발하였으며, 로지스틱, 랜덤 포레스트, 인공신경망 알고리즘 기반의 예측 모형을 테스트한 결과 랜덤 포레스트 모형이 정확도 79.8%로 가장 우수한 것으로 분석되었다. 또한 랜덤 포레스트 모형의 변수중요도를 이용하여 연령에 따라 보행만족도에 영향을 미치는 주요 변수에 차이가 있음을 확인하였다.

Jun (2021)의 연구는 Landsat TM 위성영상자료와 필지단위의 토지관리정보시스템(LMIS, Land Management Information System)자료를 이용하여 수도권 토지이용변화 예측모형을 개발하였으며, 그래디언트 부스팅, 인공신경망, 랜덤포레스트 모형을 각각 구축해 예측 성능을 비교·평가 하였다. 분석결과 그래디언트 부스팅 알고리즘과 랜덤포레스트 모형이 인공신경망 보다 예측 성능이 높은 것으로 분석되었으며, 이를 토대로 토지이용변화 예측모형의 경우 트리기반의 앙상블 모형이 높은 예측력을 갖는다고 평가하였다.

이에 본 연구에서는 이제승·이현희(2019), Jun(2021)의 연구에서 예측력이 높은 것으로 평가되는 트리기반의 앙상블 모형 중 그래디언트 부스팅 알고리즘을 이용해 쇠퇴도시 예측모형을 구축한다.

4. 선행연구와의 차별성

본 연구는 머신러닝 알고리즘을 이용하여 도시쇠퇴를 예측하기 위해 기존연구에서 도시쇠퇴를 어떻게 정의하는지를 살펴보고 도시쇠퇴와 기계학습모형에 관한 연구동향을 살펴보았다. 최근 쇠퇴도시와 관련한 연구들은 대도시, 중소도시, 농촌도시 혹은 소도시, 근린지구 등 다양한 규모별 대상 도시의 쇠퇴현상을 진단하고 지역특성에 맞는 지역재생정책의 방향을 도출하고 있다. 이러한 기존 연구들은 시군구 혹은 읍면동 단위로 제공되는 통계자료에 의존하기 때문에 현실에서 국지적으로 발생하는 쇠퇴현상을 분석하는데 한계가 있으며, 이에 실제 위치기반의 건축물 단위 등의 자료를 활용한 필지 및 그리드 단위의 연구도 수행되었다. 하지만 이 역시 데이터 확보 문제, 통계의 지속성 측면에서 모든 지역으로 일반화하기에는 한계가 있다.

본 연구는 통계청에서 제공하는 최소 통계구역 단위인 집계구 데이터와 공간분석기법을 활용하여 마이크로레벨에서 도시쇠퇴현상의 공간적 분포 특성을 분석한다. 집계구 자료는 인구주택총조사와 전국사업체조사 자료를 포함하기 때문에 주요한 쇠퇴도시 지표를 확보할 수 있고, 공간적 단위도 행정동의 1/30 크기이기 때문에 도시쇠퇴 현상의 국지적 양상을 파악할 수 있다.

그리고 쇠퇴도시의 진단 및 지표에 관한 연구가 다수 축적되었지만 그럼에도 불구하고 이러한 연구들은 현상 진단에만 그친다는 한계 등이 여전히 존재하고, 도시쇠퇴를 예측하는 연구는 전무한 실정이다.

이에 본 연구는 최근 다양한 연구 분야에 적용되고 있는 머신러닝 기법을 활용해 쇠퇴도시 예측모형을 구축하고자 한다. 전통적인 회귀모형은 종속변수와 모형에 투입된 설명변수간의 선형 관계를 나타내는 회귀계수의 추정 및 검정에 초점을 맞추고 있어, 해당 분석 결과를 실제 종속변수의 예측값으로 보기에 어려움이 있다(오미애 외, 2017). 머신러닝 알고리즘은 전통적인 통계모형과 달리 데이터의 분포에 대한 가정이 없고 변수들 간의 상관관계가 머신러닝의 접근법에 영향을 미치지 않기 때문에 다양한 변수들을 고려할 수 있으며, 비선형관계인 변수들도 사용할 수 있다(Reades et al., 2019; Jun, 2021). 본 연구에서 사용하는 그래디언트 부스팅 알고리즘(Gradient Boosting Algorithm)은 최근 머신러닝 대회에서도 최고의 예측력을 가지는 모형으로 평가되고 있다.

Ⅲ. 도시쇠퇴 공간분포 특성 분석

1. 분석자료

본 연구는 공간통계기법을 활용하여 도시쇠퇴의 공간적 분포 패턴을 분석하기 위해 통계청에서 통계를 제공하는 최소 통계구역 단위인 집계구를 공간단위로 설정하여 도시쇠퇴의 국지적 양상을 분석한다. 기존 도시쇠퇴 지표 적용에 관한 연구들에 따르면 시군구, 읍면동 등 행정구역 단위를 비롯해 필지, 격자망 등 보다 세부적인 공간단위에서도 연구가 이루어졌다. 이러한 연구결과들을 종합해보면 시군구 및 읍면동의 경우 통계의 지속성은 확보되지만 실제 도시쇠퇴가 발생하는 국지적 양상을 반영하는데 한계가 있으며, 위치기반의 필지 및 그리드 단위의 연구는 데이터 확보 문제, 통계의 지속성 측면에서 모든 지역으로 일반화하기에는 한계가 있다.

따라서 본 연구에서는 통계자료가 제공되는 공간적 최소단위인 집계구를 분석단위로 설정하였다. 집계구 통계자료는 크게 인구, 가구, 주택, 사업체로 구성되어 있으며, 인구와 가구, 주택자료는 5년 주기로 시행하는 인구주택총조사 결과를 집계구 단위로 집계하여 제공한다. 사업체 통계는 1년 주기로 시행하는 전국사업체조사 결과를 집계구 단위로 집계하여 제공한다. 집계구 자료는 2000년 자료부터 제공되기 때문에 2000년과 2015년을 시간적 범위로 설정한다.

2. 분석모형

도시쇠퇴현상의 공간적 분포패턴을 분석한다는 것은 공간적 자기상관성 분석을 의미하며, 전체 공간을 대상으로 자기상관관계가 존재하는지를 판단하는 Moran's I와 Getis-Ord General G, 특정지역과 그 주변 지역들을 비교해 국지적 공간 자기상관성(Local Indicator of Spatial

Association, LISA)을 분석하는 Local Moran's I, Getis-Ord G_i^* 등이 있다.

본 연구에서는 핫스팟 분석으로 잘 알려진 Getis-Ord의 G_i^* 분석을 통해 도시쇠퇴현상의 공간적 분포 패턴을 확인한다. 핫스팟 분석은 알고자 하는 변수에 대해 Z-score를 계산하여 통계적으로 유의한 핫스팟과 콜드스팟을 구분한다. Z값이 양(+)의 값이면 공간적으로 높은 값이 군집되어 있는 핫스팟을 의미하고, Z값이 음(-)의 값이면 공간적으로 낮은 값이 군집되어 있는 콜드스팟을 의미한다. 따라서 도시쇠퇴현상의 국지적 양상을 통계적으로 판단하고, 분석결과를 도면화해서 직관적으로 확인 할 수 있다는 장점이 있다(Getis & Ord, 1992).

$$G_i^*(d) = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{SD \sqrt{\frac{[n \sum_{j=1}^n w_{i,j}^2 - (\sum_{j=1}^n w_{i,j})^2]}{n-1}}}, \text{ all } j: \bar{X} = \frac{\sum_{j=1}^n x_j}{n}, SD = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (1)$$

여기서, i, j 는 분석의 공간단위인 집계구, x_i, x_j 는 i 또는 j 지역의 속성데이터로 도시쇠퇴의 유무, $w_{i,j}$ 는 i 와 j 지역 간 공간 가중치, n 는 분석단위인 집계구의 수를 의미한다.

도시쇠퇴 유무는 도시재생법에 따른 도시재생 활성화지역 지정 기준인 인구감소, 사업체감소, 노후주택증가(주거환경악화)를 기준으로 2개 이상을 만족하는 경우를 도시쇠퇴로 정의하였다. 도시재생법에 따른 인구감소 기준은 최근 30년 동안 최대치 인구 대비 현재 인구가 20%이상 감소하거나 최근 3년 동안 인구가 지속적으로 하락 한 경우를 도시쇠퇴로 판단하지만, 본 연구에서 사용하는 집계구 자료는 2000년 인구센서스부터 조사되었기 때문에 2000년부터 2015년까지 15년 동안의 변화로 인구감소를 판단한다. 따라서 15년 동안 인구가 10% 이상 감소한 경우 인구감소로 정의하였다. 사업체수 감소 역시 최근 10년 동안 최대치 사업체수 대비 5%이상 감소하였거나 최근 5년 동안 3년 연속 사업체수가 감소한 경우 사업체수 감소로 판단하지만 본 연구에서는 15년 동안 5%이상 감소하였을 경우 사업체수 감소로 정의하였다. 또한 도시재생법에서는 전체 건축물 중 준공된 후 20년 이상 지난 건축물이 차지하는 비율이 50% 이상인 지역을 노후주택 증가 등 주거환경 악화 기준으로 정하고 있으나, 본 연구에서 노후주택비율(준공 후 20년이 지난 주택)이 50% 이상인 지역을 기준으로 사용하였다.

3. 분석결과

그림 1은 2000년부터 2015년 까지 15년 동안의 인구, 사업체, 주택노후도 3가지 지표를 기준으로 정의한 도시쇠퇴현상의 핫스팟 분석 결과를 보여준다. 2015년 기준 집계구 수는 102,219개로 그 중 45.9%가 도시쇠퇴 현상의 핫스팟으로 분석되었다. 시도별로는 부산광역시 85.2%, 대전광역시

69.4%, 대구광역시 68.0%, 인천광역시 66.3%, 충청북도 65.9%, 전라북도 65.2%, 전라남도 59.6%, 광주광역시 58.7%, 경상북도 55.4%, 울산광역시 51.4%, 경상남도 45.4%, 서울특별시 41.3%, 강원도 35.7%, 충청남도 25.8%, 경기도 21.6%, 제주특별자치도 6.1%, 세종특별자치시 4.4% 순으로 높게 나타났다. 수도권은 서울의 동대문구, 중랑구, 강북구, 노원구, 도봉구 등 북동부 지역과 남서부 지역의 금천구를 시작으로 경기도 광명시, 안양시 만안구, 인천 방면으로는 부천시와 부평구 등이 핫스팟 지역으로 나타났다.

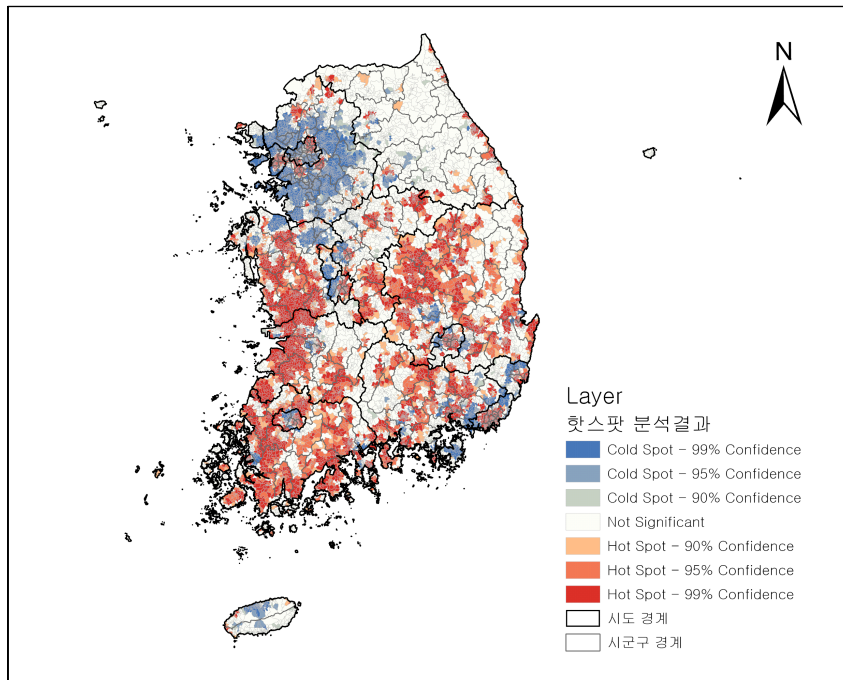


그림 1. 도시쇠퇴 공간분포 현황(2000-2015)

핫스팟 분석은 공간가중치를 결정짓는 기준에 따라 분석결과에 차이가 나타날 수 있으며, 본 연구는 집계구 중심점을 기준으로 반경 5km를 기준으로 분석하였다.²⁾ 서울을 비롯하여 광역시 등 인구밀도가 높은 지역은 집계구간의 거리가 훨씬 짧고 개수도 많기 때문에 거리 기준을 더 작게 해야 국지적인 패턴을 분석할 수 있으나 본 연구는 전국적인 공간분포패턴을 보는 것이기 때문에 해석하는데 주의가 요구된다.

그럼에도 불구하고 광역시 중에서는 부산광역시가 도시쇠퇴 핫스팟 비율이 가장 높은 것으로 나타났다. 부산광역시는 도시재생 종합정보체계에서 매년 발표하는 읍면동 단위의 도시쇠퇴현황 (2015년 12월 기준)³⁾에서도 도시쇠퇴 읍면동 비율이 가장 높은 것으로 분석되었다. 그림1을 보면

2) 적절한 거리 밴드값을 찾기 위해 Optimized Hotspot Analysis를 수행하였으며 결과를 토대로 5km로 정함.

도시쇠퇴 핫스팟이 충청북도를 시작으로 서해안을 따라 전라남도까지 고르게 분포되어 있고, 경상북도 지역에 넓게 분포되어 있는 것을 직관적으로 확인 할 수 있다. 실제로 분석결과를 시군단위로 집계해보면 각 시군별 집계구의 50% 이상이 도시쇠퇴 핫스팟으로 분석된 지역이 경상북도가 70.8%로 가장 높게 나타나났으며, 전라남도가 63.6%, 전라북도가 60.0, 충청북도가 57.1% 순으로 집계되었다.

IV. 도시쇠퇴 예측모형 분석

1. 분석모형

1) 그래디언트 부스팅 알고리즘

부스팅(Boosting)이란 비교적 부정확한 약한 학습기(Weak Learner)를 묶어서 보다 정확하고 강한 학습기(Strong Learner)를 만드는 방식을 뜻한다. 일단 정확도가 낮더라도 첫 번째 트리 모델을 만들고, 드러난 약점(예측 오류)은 두 번째 트리 모델이 보완한다. 이와 같은 방법으로 다음 트리 모델에서 약점을 계속 보완하여 결국에는 강한 학습기를 구축한다(허주성 외, 2018). 그래디언트 부스팅은 손실 함수의 기울기를 바탕으로 여러 개의 약한 예측 모형을 단계적으로 생성한 뒤 앙상블 방법으로 결합하여 강한 예측력을 가지도록 하는 부스팅 기법을 의미한다(강수연 외, 2015).

손실함수(Loss Function, J)는 예측 모델의 오류를 정량화하며, 이러한 손실함수 값을 최소화하는 모델 내 파라미터를 찾기 위하여 일반적인 기계학습 모델은 경사 하강(Gradient Descent) 방식을 사용한다. 그래디언트 부스팅은 이러한 파라미터 손실함수 최소화 과정을 모델 함수(f_i) 공간에서 수행하며, 손실함수를 모델 파라미터가 아니라 다음과 같은 수식에 의해 현재까지 학습된 트리 모델 함수로 미분한다(허주성 외, 2018; Friedman, 2001, 2002; Ding et al., 2016).

$$f_{i+1} = f_i - \rho \frac{\delta J}{\delta f_i} \quad (2)$$

즉, 그래디언트 부스팅 모델에서 트리 모델 함수 미분값은 현재까지 학습된 모델의 약점을 나타내는 역할을 하며, 다음 트리 모델의 피팅을 수행할 때 그 미분값을 사용하여 약점을 보완하여 성능을 Boosting한다(허주성 외, 2018; Friedman, 2001, 2002; Ding et al., 2016).

2) 변수 중요도(Feature importance)

그라디언트 부스팅 알고리즘의 장점은 사용된 변수의 상대적 중요도를 산출 가능하다는 것이다. 변수들의 상대적 중요도는 트리기반 모델에서 사용되며, 특정 변수가 트리를 분할하는데 얼마나 기여를 했는지에 따라 결정된다.

트리를 분할 할 때 지니 불순도(Gini Impurity) 혹은 엔트로피를 이용하여 변수가 결정되고 정보획득량(Information Gain)을 산출 할 수 있다. 이때 개별 트리의 각각 변수들이 획득한 정보량 증가분의 평균으로 각 변수들의 상대적 중요도가 산출되며 수식은 다음과 같다(Hastie et al., 2009).

$$I_{\ell}^2 = \frac{1}{M} \sum_{m=1}^M I_{\ell}^2(T_m) \tag{3}$$

3) 모형 성능 평가

머신러닝에서 분류모형의 성능 평가는 일반적으로 혼동행렬(Confusion Matrix)과 그로부터 도출되는 정확도, 정밀도, 재현율, f1-score로 판단할 수 있다. 혼동행렬은 실제 클래스와 모형이 예측한 클래스의 일치, 불일치 여부를 개수로 세어 표로 나타낸 것이다. 실제 양성 클래스에 속하는데 모형도 양성 클래스로 예측할 경우 TP(True Positive), 실제 음성 클래스에 속하는데 모형도 음성 클래스로 예측할 경우 TN(Ture Negative)로 카운트 된다. 반면 실제 양성 클래스에 속하는데 모형이 음성 클래스로 예측하면 FN(False Negative), 실제 음성 클래스에 속하는데 모형이 양성 클래스로 예측하면 FP(False Positive)로 카운트 된다.

표 2. Confusion Matrix 예시

실제값 \ 예측값	양성 (Positive)	음성 (Negative)
양성 (Positive)	TP (True Positive)	FN (False Nagative)
음성 (Negative)	FP (False Positive)	TN (True Negative)

위의 Confusion Matrix에서 정확도, 정밀도, 재현율, f1-score 지표를 산출할 수 있다. 정확도 (accuracy)는 전체 샘플 수 대비 TP와 TN의 비율, 즉 정확하게 예측된 샘플 수의 비율을 의미하며, 높을수록 좋은 모형이고 일반적으로 학습에서 최적화 목적함수로 사용한다.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

정밀도(precision)는 모형이 양성 클래스에 속한다고 예측한 샘플 수 대비 실제 양성 클래스에 속하는 샘플 수의 비율을 의미하며 수식으로 표현하면 다음과 같다. 정밀도 역시 높을수록 좋은 모형이다.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

재현율(recall)은 양성검출율(TPR, True Positive Rate) 또는 민감도(Sensitivity)로 불리기도 하며, 실제 양성클래스에 속하는 샘플 수 대비 모형이 양성클래스에 속한다고 예측한 샘플 수의 비율을 의미한다. 재현율 역시 높을수록 좋은 모형이고 수식으로 표현하면 다음과 같다.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

마지막으로 f1-score는 클래스별 샘플 수가 고르지 못할 때, 즉 예측하고자 하는 변수의 클래스가 비대칭(imbalanced data)일 때 모형의 성능을 평가할 수 있는 지표로 정밀도(precision)와 재현율(recall)의 조화평균으로 산출된다. f1-score 역시 높을수록 좋은 모형으로 판단하며, 정밀도와 재현율이 골고루 클 때 값이 높아진다.

$$f1 \text{ score} = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (7)$$

2. 분석자료

본 연구는 도시쇠퇴의 가장 중요한 지표인 인구를 기준으로 1) 인구증가 지역, 2) 인구감소 지역을 타겟 변수로 하는 Gradient boosting binomial model을 구축한다. 도시의 물리적 쇠퇴나 지역일자리 감소 등의 경제적 쇠퇴는 결국 인구유출로 인한 인구감소로 이어진다. 따라서 인구의 감소는 도시쇠퇴의 원인이면서 동시에 도시쇠퇴의 결과이기도 하며(김광중, 2010), 본 연구에서는 인구의 감소를 도시쇠퇴의 대리 변수로 사용한다.

쇠퇴도시를 예측하는데 사용한 변수는 도시재생법의 도시재생 활성화 지역 지정을 위한 도시쇠퇴 현황 및 진단에 사용하는 지표와 앞서 서술한 기존의 도시쇠퇴 지표에 관한 연구들을 바탕으로 인구, 경제, 물리적 환경 부문으로 구성하였다. 인구부문은 기준년도 인구, 출생자수, 사망자수, 순인구이동자수를 변수로 구성하였으며, 경제부문 변수는 총 사업체수, 그리고 물리적 환경 부문 변수는 노후주택비율과 신규주택비율을 사용하였다.

표 3. 분석자료 기초통계량

변수		레이블	N	평균	표준편차	최솟값	최댓값
2005~ 2010년	Target	인구감소=1, 인구증가=0	3,494	0.7	0.4	0	1
	POP05	인구수(2005년)	3,494	13,262	11,120	0	88,449
	Birth	출생자수(2006~2010년)	3,494	665	686	0	5,060
	Death	사망자수(2006~2010년)	3,494	354	194	0	1,788
	Movepop	순인구이동(2006~2010년)	3,494	0	4,244	-14,367	97,762
	Firm05	총사업체수(2005년)	3,494	832	942	0	17,035
	NewHs_r	신규주택비율	3,494	18.9	17.0	0.0	100.0
	OldHs_r	노후주택비율	3,494	21.4	17.3	0.0	100.0
2010~ 2015년	Target	인구감소=1, 인구증가=0	3,494	0.5	0.5	0	1
	POP10	인구수(2010년)	3,494	13,550	11,884	0	84,464
	Birth	출생자수(2011~2015년)	3,494	648	721	3	6,416
	Death	사망자수(2011~2015년)	3,494	382	220	9	2,341
	Movepop	순인구이동(2011~2015년)	3,494	0	3,976	-14,871	71,768
	Firm10	총사업체수(2010년)	3,494	873	952	0	15,257
	NewHs_r	신규주택비율	3,494	10.3	13.2	0.0	100.0
	OldHs_r	노후주택비율	3,494	35.4	21.1	0.0	100.0

출처: 인구, 총사업체수, 신규주택비율, 노후주택비율은 센서스 집계구 자료를 읍면동으로 집계하여 사용하였으며, 순인구이동은 연도별 국내인구이동통계를 사용함

일반적으로 기계학습모형을 구축할 때 모형을 학습시킬 훈련데이터(train data)와 평가데이터(test data)가 필요하다. 연구자에 따라서는 훈련데이터를 다시 일정 비율로 분할해서 하이퍼파라메타 조정 및 평가를 위한 검증데이터(validation data)로 사용하기도 한다. 본 연구에서는 2005-2010년 자료를 75:25로 분할하여 75%는 훈련데이터로 학습을 시키고, 25%는 최적의 학습모형을 선택하는 검증데이터로 사용하였다. 이를 토대로 최종 선택된 모형으로 2010-2015년 도시쇠퇴를 예측하는 평가데이터로 사용하였다.

분석단위는 전국 3,494개 읍면동이며, 이는 2015년 기준 3,502개에서 민간인통제구역 8개 지역⁴⁾을 제외한 것이다. 2005~2010년의 인구감소 지역은 70%, 2010~2015년의 인구감소 지역이 50%로 나타났다.

3. 분석결과

1) 모형성능평가

본 연구에서는 도시쇠퇴를 예측하기 위해 인구감소, 인구증가를 타겟 변수로 하는 그래디언트 부스팅 분류모형을 구축하였다. 2005-2010년 자료로 모형을 학습시킨 후 최종 선택된 모형으로 2010-2015년 인구 감소 지역을 예측하였다. 다음으로 모형이 예측한 결과와 실제 데이터를 비교해 정확도, 정밀도, 재현율, f1-score 등의 지표를 이용해 모형의 성능을 평가하였다. 아래 표4는 모형성능 결과를 보여준다.

표 4. 모형성능평가

실제값 \ 예측값	인구증가(Y=0)	인구감소(Y=1)	Recall
인구증가(Y=0)	638	1047	0.38
인구감소(Y=1)	26	1,783	0.99
Precision	0.96	0.63	-
f1-score	0.54	0.77	-
Accuracy	0.69		

먼저 전체 데이터 중에서 모형이 정확하게 분류한 데이터의 비율로 산출되는 정확도(Accuracy)는 0.69로, 이는 모형이 예측한 인구감소지역과 인구증가지역이 실제 인구감소지역 및 증가지역과 69% 일치함을 의미한다. 실제로 인구가 감소한 지역 중에서 모형이 정확하게 예측한 비율을 나타내는 재현율(recall)은 0.99로 높게 나타났으며, 모형이 예측한 인구감소 지역 중에 실제로 인구가 감소한 지역의 비율을 나타내는 정밀도(Precision)는 0.63로 다소 낮게 나타났다. 불균형 데이터의 경우 정밀도와 재현율의 가중조화평균으로 산출되는 f1-score 값이 모형의 전체 성능을 더 잘 평가한다고 할 수 있는데 본 도시쇠퇴 예측모형은 인구감소의 경우 0.77로 모형이 인구감소지역을 잘 예측한다고 평가 할 수 있다.

4) 경기도 파주시 장단면, 진동면, 진서면, 강원도 철원군 근동면, 원동면, 원남면, 임남면, 강원도 고성군 수동면, 이상의 8개 지역은 분석에서 제외함

2) 변수중요도

그라디언트 부스팅 모형은 사용된 변수가 트리분할시 노드 불순도를 얼마나 개선하는가에 따라서 모형에 기여한 상대적 중요도를 산출한다. 즉 변수 중요도는 모형을 예측하는데 각 변수가 상대적으로 얼마만큼 기여했는지를 의미한다. 다음의 그림 2는 도시쇠퇴 예측 모형의 결과로 산출된 변수의 상대적 중요도를 보여준다.

인구감소를 예측하는데 가장 중요하게 기여한 변수는 인구이동으로 34.5%로 압도적으로 높은 것으로 분석되었다. 다음으로 출생자수가 12.4%, 노후주택비율이 12.0%로 중요한 것으로 나타났다. 노후주택비율은 인구변화율과 함께 도시재생법에 따른 도시쇠퇴 진단지표 중 하나이면서 동시에 본 연구 결과는 인구 증감을 도시쇠퇴의 대리변수로 사용하였을 경우 노후주택비율이 중요한 예측변수로 기여함을 확인 할 수 있었다. 한편 총사업체수의 변화율은 10.0%로 상대적으로 영향이 작은 것으로 분석되었다.

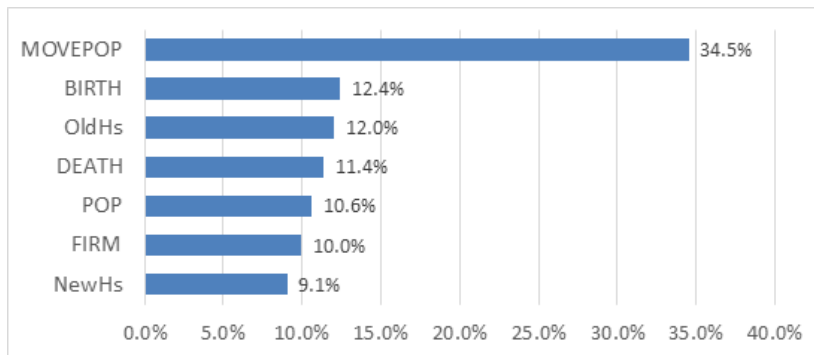


그림 2. 변수의 상대적 중요도

3) 실제 도시쇠퇴현황과 모형 예측결과 비교

아래 그림 3은 모형을 통한 도시쇠퇴를 예측 한 후 혼동행렬을 바탕으로 결과를 지도화 한 것이다. 혼동행렬은 실제 인구감소 지역을 모형도 인구감소 지역으로 분류한 경우(TP), 실제 인구증가 지역을 모형도 인구증가 지역으로 분류한 경우(TN), 실제 인구증가 지역이나 모형은 인구감소 지역으로 분류한 경우(FP), 그리고 실제 인구감소 지역이나 모형은 인구증가 지역으로 분류한 경우(FN)로 구분되며 공간적 분포를 보면 아래 그림과 같다. 수도권을 비롯하여 지방 광역도시를 중심으로 인구 감소 및 인구 증가를 모형이 잘 예측 한 것으로 보이는 반면, 그 외 지방 중소도시의 경우 앞선 핫스팟 분석결과에서 도시쇠퇴 군집지역으로 분류된 경상북도와 전라남도, 전라북도 일부 지역을 제외하면 실제로는 인구증가 지역이나 모형이 인구감소지역으로 잘못 분류하고 있는 것으로 보인다. 이는 쇠퇴가 심각한 지역의 경우 모형에 의해 잘 분류되는 것으로 볼 수

있지만 모형의 정확도를 높이기 위해 추후 도시재생법에서 정한 도시쇠퇴지표 외에 더 다양한 도시쇠퇴를 반영할 수 있는 변수들을 추가하여 모형을 구축해야 할 것으로 판단된다.

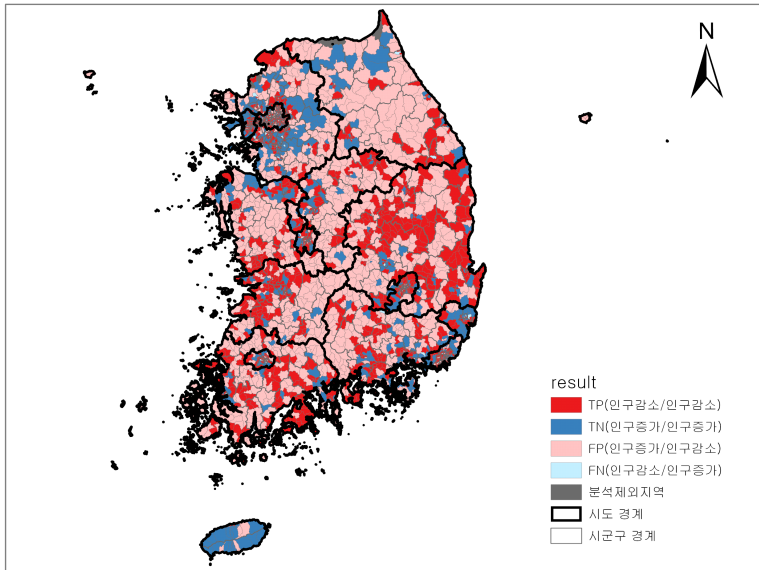


그림 3. 도시쇠퇴 예측 결과의 공간적 분포

V. 결론 및 시사점

본 연구는 공간통계기법을 통해 도시쇠퇴의 공간적 군집 정도를 통계적으로 확인하고, 도시쇠퇴 지표를 이용하여 그래디언트 부스팅 알고리즘 기반의 도시쇠퇴 예측모형을 개발 및 예측하는 것을 목적으로 한다. 기존의 도시쇠퇴 관련 연구들이 다수 축척 되었으나 현상 진단에만 그친다는 한계 등이 여전히 존재하였다. 이에 본 연구는 최근 도시계획 분야에도 활발히 도입되고 있는 기계학습 기법을 활용한 도시쇠퇴 예측모형의 개발을 위한 기초 연구로서의 의미를 갖는다.

본 연구의 주요 결과는 다음과 같다. 첫째, 공간통계기법을 통해 집계구 단위로 도시쇠퇴의 공간적 핫스팟을 확인하였다. 그 결과 읍면동 단위로 본 도시쇠퇴의 공간적 패턴과 다름을 확인하였으며 광역시도 단위로 집계한 결과는 패턴이 일치함을 확인하였다. 이는 도시쇠퇴의 전반적인 추세를 확인하는 목적으로는 현재의 읍면동 단위의 집계로도 충분하지만 집계구 단위의 통계로 작성된다면 정책을 결정하는데 현장의 상황을 더 잘 반영할 수 있음을 의미한다.

둘째, 그래디언트 부스팅 알고리즘 기반의 도시쇠퇴 예측모형을 개발하고 예측 성능을 평가하였다. 모형의 전체 정확도는 0.69로 분석되었으며, 재현율은 0.99로 높지만 정밀도는 0.63으로 다소

낮게 나타났다. 이는 실제 인구감소지역 중에 모형이 99%는 정확하게 인구감소지역을 판단하지만, 모형이 예측한 인구감소지역 중에서 실제로 인구가 감소한 지역은 63%인 것을 의미한다. 모형의 목적에 따라 정밀도가 중요한 평가지표일 수도 있고, 재현율이 중요한 평가지표가 될 수도 있다. 본 연구에서는 재현율과 정밀도 차이가 다소 높게 나타났으며 향후 도시재생법에 따른 도시쇠퇴지표 뿐만 아니라 다양한 도시의 쇠퇴현상을 반영할 수 있는 지표들을 변수로 더 확보하여 예측모형을 정밀하게 할 필요가 있다.

셋째, 도시쇠퇴 예측모형의 분석결과 인구이동 변수가 예측에 기여하는 정도를 의미하는 상대적 변수 중요도에서 가장 높은 것으로 나타났다. 또한 출생률 다음으로 물리적 노후도를 반영하는 노후주택비율도 중요한 요인인 것으로 분석되었다.

이상의 연구결과는 도시재생 부분에 적용될 수 있는 시사점을 제시할 수 있다. 첫째, 모형예측에서 산출된 상대적 변수의 중요도는 인구이동이 쇠퇴도시를 예측하는데 가장 크게 기여함을 의미한다. 이는 지표들 간의 도시쇠퇴에 대한 영향력이 다르다는 것을 암시하며, 따라서 도시재생 활성화 지역 지정을 위한 도시쇠퇴 현황 및 진단에 사용하는 지표를 판단할 때 지표마다 서로 다른 가중치를 고려해 볼 수 있다.

둘째, 도시쇠퇴 예측 결과의 공간적 분포는 수도권 및 지방 광역도시, 그리고 핫스팟 분석에서 도시쇠퇴 군집지역으로 분류된 지역의 경우 도시쇠퇴 현상을 잘 예측한 반면 그 외의 지역은 모형이 잘 예측하지 못하는 것을 보여준다. 이는 인구가 상대적으로 작고 도시쇠퇴현상이 분산, 다발적으로 발생하는 지역의 경우 예측모형으로 잘 예측되지 않음을 의미한다. 따라서 향후 도시재생 연구는 이러한 지방 중소도시를 대상으로 좀 더 집중될 필요가 있다.

본 연구결과는 쇠퇴도시 예측을 위한 기초연구로서의 의미를 갖지만 데이터 가용성 측면에서의 한계점은 추가적인 분석과 연구의 필요성을 제시한다. 도시재생법에서 정의하는 도시쇠퇴 기준의 시간적 범위와 본 연구에서의 시간적 범위가 다르기 때문에 연구 자료의 시간적 범위를 반영하여 산정방식을 조정하였다. 또한 연구의 시간적 범위가 2000~2015년으로 쇠퇴진단의 결과가 최신의 경향을 반영하는데 한계가 있을 수 있으므로 해석상에 주의가 요구된다. 따라서 후속 연구과제로 이러한 한계점을 보완하고, 추가적으로 도시쇠퇴지표를 더욱 정밀하게 구축하여 도시쇠퇴 예측모형을 보완할 수 있으며, 이는 도시재생 사업 및 정책적 우선순위 결정을 위한 근거자료로 사용할 수 있을 것으로 판단된다.

인용문헌

- 강수연·전희정·김지혜·송중우, 2015. “국내 드라마 시청률 예측 및 영향요인 분석”, 「응용통계연구」 28(5): 933-949.
- 강인호, 2019. “지방중소도시의 도시축소 및 실태에 관한 실증적 연구”, 「한국지역개발학회지」 31(5): 45-71.
- 권오규·강은택·마강래, 2019. “도시의 쇠퇴가 지역주민의 삶의 만족감에 미치는 영향—도시 규모의 차이를 중심으로”, 「지역연구」 35(2): 33-47.
- 김광중, 2010. “한국 도시쇠퇴의 원인과 특성”, 「한국도시지리학회지」 13(2): 43-58.
- 김광중·박현영·김예성·안현진, 2010. “도시 내 지구차원(地區次元)의 쇠퇴실태와 양상” 「한국도시지리학회지」, 13(2): 27-42.
- 김예슬·김형보, 2015. “도시재생과 활성화를 위한 구도심 쇠퇴분석연구”, 「GRI 연구논총」 17(3): 341-361.
- 노승철·김태현·이석준·임재욱·고진수, 2020. “기성시가지 필지별 토지이용 변화 예측을 위한 기계학습 모델 개발에 관한 연구: 서울시 신축 발생 예측을 중심으로”, 「감정평가학 논집」 19(1): 27-52.
- 박근덕·이수기, 2018. “가로공간 보행만족도 예측을 위한 딥러닝 모형의 적용과 검증”, 「한국도시설계학회지 도시설계」 19(6): 19-34.
- 박소연·오덕성, 2015. “쇠퇴지역의 유형에 따른 도시재생사업의 평가 :대전광역시 쇠퇴지역을 중심으로”, 「한국산학기술학회 논문지」 16(7): 4984-4991.
- 배성완·유정석, 2018. “머신 러닝 방법과 시계열 분석 모형을 이용한 부동산 가격지수 예측”, 「주택연구」 26(1): 107-133.
- 서울대학교 산학협력단, 2010a. 한국 도시쇠퇴의 실태 및 유형화, 도시재생사업단.
- 서울대학교 산학협력단, 2010b. 도시쇠퇴 및 잠재력 진단지표 개발. 도시재생사업단.
- 서준교, 2014. “도시쇠퇴(urban decline)와 수축(shrinkage)의 원인과 대응전략 연구 :리버풀(Liverpool)과 라이프치히(Leipzig)의 사례를 중심으로”, 「한국지방자치학회보」 26(1): 97-115.
- 엄상근·남윤섭, 2014. “도시재생 정책 수립을 위한 지방중소도시의 도시쇠퇴지표 적용 -제주시를 중심으로”, 「한국도시지리학회지」 17(3): 111-122.
- 엄현태·우명제, 2014. “교외지역 신시가지 개발이 중심도시의 구시가지 쇠퇴에 미치는 영향 분석” 「국토계획」 49(5): 51-66.
- 엄현태·우명제, 2016. “도시쇠퇴의 공간적 확산현상과 쇠퇴확산 요인에 대한 실증분석”, 「국토계획」 51(2): 5-18.
- 오미애·최현수·김수현·장준혁·진재현·천미경, 2017. 기계학습 (Machine Learning) 기반 사회보장 빅데이터 분석 및 예측모형 연구, 한국보건사회연구원.
- 오윤표·강진학, 2002. “우리나라 도시의 인구성쇠패턴에 관한 연구”, 「국토계획」 37(1): 45-56.
- 윤종문, 2019. “딥러닝 신경망을 이용한 신용카드 부도위험 예측의 효용성 분석”, 「금융연구」 33(1): 151-183.
- 원석환·이창규·박지만, 2017. “머신러닝 기법을 적용한 지가 예측 연구”, 「국토지리학회지」 51(4): 347-355.

- 이민석, 2018. 빅데이터를 활용한 쇠퇴도시 주거지역 진단평가 및 사업유형 개발, 전남대학교.
- 이상대, 1996. “서울시 내부기성시가지의 시가지특성과 도시계획적 과제”, 「국토계획」 31(6): 9-23.
- 이영성·김예지·김용욱, 2010. “도시차원의 쇠퇴실태와 경향”, 「한국도시지리학회지」 13(2): 1-11.
- 이재승·이현희, 2019. “기계학습 알고리즘을 이용한 보행만족도 예측모형 개발”, 「국토계획」 54(3): 106-118.
- 이종휘·이태희, 2020. “공간적 범위의 차이에 의한 도시재생 활성화지역 지정기준 적용 결과에 관한 연구”, 「한국산학기술학회논문지」 21(10): 567-573.
- 임현성·김충호, 2019. “도시쇠퇴의 공간적 실태분석 및 정책개선방향 고찰:부산시 부산진구의 사례를 중심으로”, 「국토계획」 54(1): 186-198.
- 장문현, 2016. “GIS와 공간통계기법을 활용한 도시쇠퇴 특성 분석 :광주광역시를 중심으로”, 「한국지역지리학회지」 22(2): 424-438.
- 장문현·이민석·김화환, 2016. “지방 대도시의 도시재생정책 지원을 위한 쇠퇴지표 적용 연구”, 「국토지리학회지」 50(4): 455-473.
- 장희순·송상열, 2006. “비성장형도시의 쇠퇴원인 분석과 활성화 방안”, 「국토연구」 50: 39-57.
- 전경주·전형준, 2016. “인구감소지역의 스마트 쇠퇴를 위한 축소도시계획에 관한 연구”, 「한국지역개발학회지」 28(1): 1-28.
- 전호진·김윤필·우명재, 2018. “도시쇠퇴와 스프롤의 유형 및 상관성에 관한 연구:전국 지역노동시장권(LLMAs)을 중심으로”, 「국토계획」 53(2): 183-200.
- 조수진·김보경·김나현·송종우, 2019. “데이터마이닝 기법을 이용한 서울시 지하철역 승차인원 예측”, 「응용통계연구」 32(1): 111-128.
- 조수희·정재호, 2010. “쇠퇴지방도시 유형화와 재생과제”, 「부동산학보」 40: 152-165.
- 조윤애, 2014. “도시 쇠퇴지수 개발에 관한 연구”, 「국토계획」 49(7): 117-134.
- 조진희·이동건·황희연, 2010. “도시쇠퇴 수준 및 특성 유형화”, 「국토지리학회지」 44(1): 35-50.
- 한국환경정책평가연구원, 2018. 빅데이터를 활용한 기후재해(가뭄, 폭설 및 집중호우) 원인과 현황 상관관계 분석, 인공지능(Machine Learning-인공신경망) 기반 미래 기후재해 예측 및 사회기반시설 영향 분석 기법 개발, 국토교통부 국토교통과학기술진흥원.
- 한은정·김동건, 2015. “노인장기요양보험 이용지원 상담 대상자 선정모형 개발”, 「응용통계연구」 28(6): 1063-1073.
- 허선영·김주영·문태현, 2018. “머신러닝기반 범죄발생 위험지역 예측”, 「한국지리정보학회지」 21(4): 64-80.
- 허재완·이삼수, 2015. 도시진단 및 중장기 로드맵(1), 국토연구원 도시재생실증연구단.
- 허주성·권도형·김주봉·한연희·안채현, 2018. “그래디언트 부스팅을 활용한 암호화폐 가격동향 예측”, 「정보처리학회논문지.소프트웨어 및 데이터 공학」 7(10): 387-396.
- Ding, C., Wu, X., Yu, G., and Wang, Y., 2016. “A gradient boosting logit model to investigate driver’s stop-or-run behavior at signalized intersections using high-resolution traffic data,” Transportation research part C: emerging technologies, 72: 225-238.
- Friedman, J.H., 2001. “Greedy function approximation: a gradient boosting machine,” Annals of Statistics, 29 (5): 1189-1232.
- Friedman, J.H., 2002. “Stochastic gradient boosting,” Computational Statistics & Data Analysis, 38 (4): 367-378.

- Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H., 2009. The elements of statistical learning: data mining, inference, and prediction, 2nd edition. New York: springer.
- Getis, A, and Ord, J. K, 1992. “The Analysis f Spatial Association by Use of Distance Statistics,” *Geographical Analysis*, 24(3): p189-206.
- Jun, M. J., 2021. “A comparison of a gradient boosting decision tree, random forests, and artificial neural networks to model urban land use changes: The case of the Seoul metropolitan area,” *International Journal of Geographical Information Science*, 35(11): 2149-2167.
- Reades, J., De Souza, J., and Hubbard, P., 2019. “Understanding urban gentrification through machine learning,” *Urban Studies*, 56(5): 922-942.

< 투고 2022.04.08., 심사 2022.04.16, 게재확정 2022.04.20. >