

지속적 방향 제한성을 고려한 Markov-CA와 LightGBM 하이브리드 모델 기반 공장 입지 시공간 패턴 예측*

박정우^{1*}·남광우²

Predicting Spatiotemporal Patterns of Factory Locations Using a Hybrid Markov-CA and LightGBM Model with Continuous Directional Constraints*

Jeongwoo Park^{1*}·Kwangwoo Nam²

요 약

본 연구는 공장 입지 결정 과정에서 발생하는 시공간적 패턴을 효과적으로 예측하기 위해 Markov-Cellular Automata(Markov-CA)와 LightGBM을 결합한 하이브리드 모델을 제안한다. 기존의 공장 입지 연구들은 과거 데이터 기반의 단편적 분석에 치중되어 있어, 시계열 데이터와 공간 데이터를 통합적으로 고려하지 못하는 한계가 있었다. 이를 해결하기 위해 본 연구에서는 공간적·시간적 의존성을 동시에 반영할 수 있는 Markov-CA 모델에 LightGBM의 고성능 이진 분류 알고리즘을 접목하였다. 특히 "지속적 방향 제한성" 개념을 도입하여 공간적 확장의 편향성을 제어하고, Optuna를 활용한 하이퍼파라미터 최적화를 통해 모델의 성능을 극대화하였다. 연구 결과, 과거 공장 입지의 좌표 데이터만으로도 입지 결정에 영향을 미치는 외부 요인들의 영향력을 효과적으로 모델에 반영할 수 있었으며, 10개 주요 요인에 대한 최적화된 파라미터 값을 도출하였다. 또한 "지속적 방향 제한성"의 적용을 통해 기존 CA 모델이 가진 무분별한 공간 확장 경향을 제어할 수 있었다. 본 연구에서 제안한 하이브리드 모델은 지방자치단체의 산업단지 계획 수립, 기업의 입지 선정, 정부의 산업 정책 수립 등에 실질적인 도움을 줄 수 있는 의사결정 지원 도구로 활용될 수 있을 것으로 기대된다.

주요어 : 시설입지예측, Markov-CA, LightGBM, 하이브리드 모델, 지속적 방향 제한성,
하이퍼파라미터 최적화

2024년 12월 9일 접수 Received on December 9, 2024 / 2024년 12월 17일 수정 Revised on December 17, 2024 / 2024년 12월 18일 심사완료 Accepted on December 18, 2024

* 본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(과제번호 RS-2022-00143404).

1 경성대학교 도시공학과, 학술연구교수 / Kyungsoo University, Department of urban engineering, Research Professor

2 경성대학교 도시공학과, 교수 / Kyungsoo University, Department of urban engineering, Professor

※ Corresponding Author E-mail: polemic21@gmail.com

ABSTRACT

This study proposes a hybrid model combining Markov-Cellular Automata (Markov-CA) and LightGBM, with the objective of developing an effective method for predicting the spatiotemporal patterns that occur during the factory location decision process. The existing literature on factory location has concentrated on piecewise analysis based on historical data, which is constrained by the inability to integrate time series and spatial data. To address this issue, we have developed a hybrid model that combines a Markov-CA model, which is capable of reflecting spatial and temporal dependencies simultaneously, with LightGBM's high-performance binary classification algorithm. In particular, the concept of "continuous directional restrictiveness" is introduced as a means of controlling the bias of spatial scaling and of optimising the performance of the model through hyperparameter optimisation using Optuna. The model was thus able to effectively reflect the influence of external factors affecting the location decision, utilising solely the coordinate data of past factory locations. Furthermore, optimised parameter values for 10 major factors were derived. Furthermore, the implementation of the "continuous directional constraint" can regulate the expansion tendency of the existing CA model. It is anticipated that the hybrid model proposed in this study will serve as a decision-support tool, offering practical assistance to local governments in the planning of industrial parks, companies in the selection of locations, and governments in the formulation of industrial policies.

KEYWORDS : Facility Location Prediction, Markov-CA, LightGBM, Hybrid Model, Continuous Directional Constraint, Hyperparameter Optimization

서론

현대 산업 사회에서 공장의 입지 결정은 지역 경제 발전과 도시 계획의 핵심 요소로 작용한다. 공장의 입지는 토지 이용 계획, 환경 규제, 교통 접근성, 인프라 공급, 토지 가격, 정책 지원 등 다양한 외부 요인들의 복잡한 상호작용에 의해 결정되며(Choi and Lee, 2008), 이는 시간의 흐름에 따라 독특한 공간적 패턴을 형성한다.

기존 연구들은 주로 공장의 입지 특성을 분석하거나 개별 요인이 입지 결정에 미치는 영향을 규명하는 데 중점을 두었다. 그러나 이러한 접근은 과거 데이터에만 의존하여 입지 결정 과정에서 발생하는 외부 요인의 동태적 영향을 충분히 반영하지 못하는 한계를 보였다. 특히 시계

열 데이터와 공간 데이터를 통합적으로 분석하지 못해, 입지 변화의 시공간적 동향을 효과적으로 예측하는 데 제약이 있었다.

이러한 한계를 극복하기 위해 Wu and Webster(1998)는 공간적 패턴과 시간적 변화를 동시에 고려할 수 있는 Markov-Cellular Automata(Markov-CA) 모델을 제안하였다. 그러나 이 모델은 전이 규칙 설정이 모델의 성능을 좌우하는 핵심 요소임에도 불구하고, 공간적·시간적 의존성을 최적화하는 데 어려움이 있었다.

이러한 배경에서 본 연구는 다음과 같은 연구 목적을 가진다.

첫째, 시계열적 공장 현황 자료에 내재된 개별 공장의 입지 결정에 영향을 미치는 외부 요인들을 파악하고자 한다. 둘째, 시계열 형태로 구축된 공장 위치 데이터(X, Y 좌표)만을 활용

하여 입지 결정에 영향을 주는 외부 요인의 도출 가능성을 검증하고자 한다. 셋째, 도출된 요인들을 기반으로 향후 공장 입지의 예측 가능성을 탐색하고자 한다.

위 연구 목적을 달성하기 위해, 본 연구는 Markov-CA 모델과 LightGBM 모델을 결합한 하이브리드 접근법을 제안한다. 특히 "지속적 방향 제한성" 개념을 도입하여 현실적인 공간 확장의 제약을 모델에 반영하고, Optuna를 활용한 하이퍼파라미터 최적화를 통해 모델의 예측 정확도를 높이고자 한다.

본 연구의 궁극적 목표는 과거의 시계열 위치 데이터만으로도 시설 입지에 영향을 미치는 외부 요인을 효과적으로 파악하고, 이를 통해 미래의 시설 입지를 정확하게 예측할 수 있는 통합적 방법론을 제시하는 것이다. 이는 기존 연구의 한계를 극복하고 더욱 실용적인 입지 예측 모델을 제시함으로써, 효율적인 산업단지 계획 및 도시 개발에 기여할 것으로 기대된다.

연구방법 및 선행연구 고찰

1. 연구의 범위와 방법

본 연구는 2008년부터 2023년까지의 시계열 자료를 바탕으로, 울산광역시를 공간적 범위로 설정하여 공장입지 결정 패턴을 분석하고자 한다. 이를 위하여 공장설립 온라인 지원시스템(Factory Establishment Management Information System, FEMIS)에서 제공하는 전국 등록공장 데이터를 활용하였다. FEMIS 데이터는 기업의 공장 설립·이전 정보, 등록일자, 위치좌표 등 공장입지 의사결정 과정과 직결되는 핵심 정보를 포함하고 있어 본 연구의 목표 달성에 적합하다.

분석 방법론으로는 공간-시간적 패턴을 동시에 반영할 수 있는 하이브리드 예측 모델을 제안한다. 본 모델은 마르코프-셀룰러 오토마타(Markov-Cellular Automata, 이하 Markov-CA) 모델과 LightGBM(Light Gradient Boosting Machine)을 결합한 형태로

구성된다. Markov-CA 모델은 시·공간적 종속성을 반영하여 특정 시점의 입지 패턴이 이후 시점의 패턴 형성에 영향을 미침을 모형화하는 한편, LightGBM은 고성능 이진 분류 알고리즘을 바탕으로 각 격자 셀의 시설 입지 발생 여부를 효율적으로 판별한다.

2. 선행연구 및 이론적 고찰

본 연구는 시설 입지의 시공간적 변화를 예측하기 위한 새로운 방법론을 제시한다. 구체적으로, Markov-CA(Cellular Automata) 모델과 LightGBM을 결합한 하이브리드 모델을 제안하여 공간적 패턴과 시간적 변화를 동시에 고려하고자 한다.

Markov-CA 모델은 Markov 체인과 셀룰러 오토마타를 결합한 모델로, 공간적 패턴과 시간적 변화를 동시에 고려할 수 있는 특징을 가진다. Markov 체인은 상태 전이 확률을 기반으로 시간적 변화를 모델링하지만, 공간적 의존성을 반영하지 못하는 한계가 있다. 반면, 셀룰러 오토마타는 인접한 셀들과의 상호작용을 통해 공간적 패턴을 반영할 수 있다. 이러한 두 모델의 결합은 시설 입지 변화와 같은 복잡한 공간-시간적 현상을 모형화하는 데 적합하다. 특히, 셀 기반의 각 상태에서 다음 시간으로의 변화 확률을 예측하여 전체적인 방향을 설정할 수 있다는 점에서 큰 장점을 가진다.

표 1과 같이 Markov-CA 모델의 전이 규칙을 최적화하기 위해 다양한 머신러닝 기법이 활용됐다. Li와 Yeh(2002)는 인공신경망을 활용하여 전이 규칙을 학습하였으나, 학습 속도가 느리고 과적합의 위험이 있다는 한계를 지적하였다. Huang et al.(2009)은 랜덤 포레스트를 사용하여 토지 이용 변화를 예측하였으나, 대용량 데이터 처리에 한계가 있었다고 보고하였다. Yang et al.(2008)은 SVM을 활용하였지만, 대용량 데이터에서 학습 속도가 저하되는 문제가 있었다. Zhang et al.(2018)은 딥러닝을 적용하여 복잡한 패턴을 학습하였으나, 많은 데이터와 계산 자원이 필요하다는 한계를 보고하였다.

Table 1. A comparison of transition rule optimisation techniques with related research

Researchers (Year)	Method	Advantages	Disadvantages
Li & Yeh (2002) Gu, C., Wang, X., & Zhang, Y(2014)	Artificial Neural Network (ANN)	Capable of learning complex nonlinear relationships	Slow learning speed, risk of overfitting
Huang et al. (2009)	Random Forest	High accuracy and stability	Limited in processing large-scale data
Yang et al. (2008)	Support Vector Machine (SVM)	Effective with high-dimensional data	Learning speed deteriorates with large-scale data
Ke et al. (2017)	LightGBM	Fast learning speed, high accuracy, easy interpretation	Potential limitations in learning complex nonlinear patterns

이러한 선행연구들의 한계를 극복하기 위해, 본 연구에서는 Ke et al.(2017)이 개발한 LightGBM을 활용하여 Markov-CA 모델의 전이 규칙을 최적화하고자 한다. LightGBM은 Gradient Boosting Decision Tree를 기반으로 한 머신러닝 알고리즘으로, 빠른 학습 속도와 높은 예측 정확도를 제공한다. 특히, 여러 개의 결정 트리를 순차적으로 학습시키면서 각 트리가 이전 트리의 오차를 줄이는 방식으로 동작하며, 최종 예측은 각 트리의 예측값을 결합하여 얻는다.

LightGBM은 이진 및 다중클래스 분류에 효율적이며, 빠른 속도, 고성능, 메모리 효율성 등의 장점을 가진다. 또한, 특성 중요도를 통해 모델의 해석 가능성을 높여준다는 점에서 본 연구의 목적에 매우 적합하다. 다만, 다소 복잡한 매개변수 조정이 필요하며, 트리 기반 모델이므로 이미지나 텍스트와 같은 비정형 데이터에는 부적합하다는 특징이 있다.

본 연구에서 제안하는 하이브리드 모델은 Markov-CA의 공간-시간적 모델링 능력과 LightGBM의 효율적인 학습 능력을 결합함으로써, 보다 정확하고 해석 가능한 시설 입지 예측을 가능하게 한다. 이는 기존 연구들이 가진 한계를 극복하고, 대규모 공간-시간 데이터에 대한 효율적인 처리와 예측을 가능하게 한다는 점에서 중요한 의미를 가진다.

분석방법 설정 및 모델 학습

1. 연구 수행 절차

본 연구는 크게 세 단계로 구성된다. 첫째, 데이터 전처리 단계에서는 FEMIS 시설등록 데이터를 수집하고 정제하여 격자 단위로 표준화하였다. 이를 위해 시설등록 데이터에 포함되어 있는 주소정보를 좌표화하여 공간데이터로 변경하였다. 둘째, 모델 구축 단계에서는 Markov-CA와 LightGBM을 결합한 하이브리드 모델을 설계하고, 지속적 방향 제한성 개념을 적용하였다. 셋째, 모델 최적화 단계에서는 Optuna를 활용한 하이퍼파라미터 최적화를 수행하고, 재현율, 정밀도, F1-Score를 기반으로 모델의 성능을 평가하였다.

본 연구에서는 울산광역시의 2008년부터 2023년까지의 공장입지 데이터를 활용하여 분석자료를 생성하였다. 생성 데이터는 공장 위치의 좌표 정보와 연도별 설치 현황 만을 포함하고 있다. 연구 지역을 100m x 100m 크기의 격자로 분할하고, 각 그리드 셀에 고유한 ID를 부여하였다.

각 셀의 공장 유무를 파악하기 위해 공장 위치 데이터를 해당 그리드 셀에 존재하지 않으면 0, 존재하면 1의 값으로 매핑하였다. 2008년부터 2022년까지의 데이터를 학습에 사용하고, 2023년 데이터를 검증하였으며 각 셀의 8방향 (Moore 이웃)에 위치한 이웃 셀들의 공장 유무

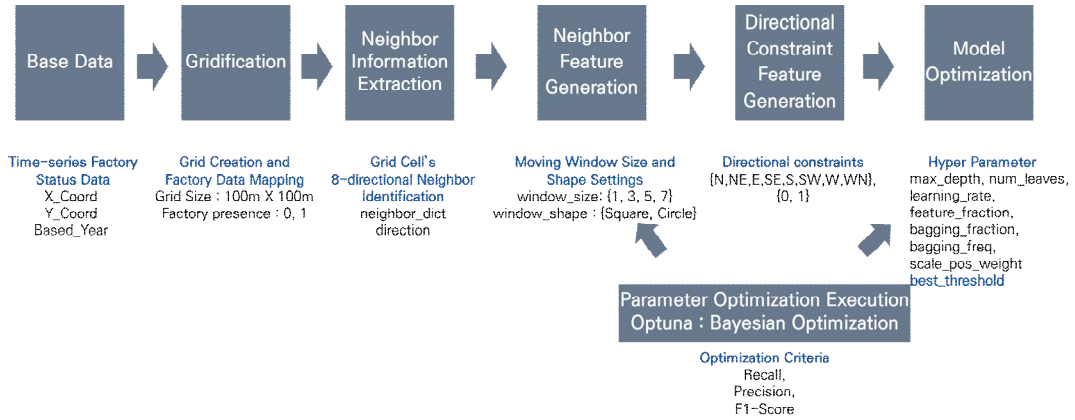


FIGURE 1. Research procedures

정보를 추출하여 입력 특징으로 활용하였다. 또한 과거 데이터에서 특정 방향으로 공장이 지속해서 입지하지 않은 패턴을 분석하여 방향별 제약 지표를 생성하였다.

입력 특징으로는 현재 셀의 공장 유무, 이웃 셀들의 공장 유무, 방향별 제약 지표 등을 사용하였으며, 타겟 변수는 다음 해의 셀 공장 유무로 설정하였다. LightGBM 모델을 학습하기 위해 데이터의 80%를 학습용으로, 20%를 검증용으로 분할하였다. 하이퍼파라미터 튜닝과 조기 종료율을 통해 최적의 모델을 선정하였다.

하이퍼파라미터의 최적화를 통해 주요 변수들의 상수값을 최적화하였으며, 최적화된 변수의 상수값을 통해 Markov-CA와 LightGBM을 결합한 하이브리드 모델의 전이 규칙을 구성하였다.

해당 프로세스는 파이썬으로 코드를 작성하여 그림 1과 같이 분석을 수행하였다.

2. 외부 효과성 배제

본 연구는 시설 입지 예측의 정확도 향상을 위해 "지속적 방향 제한성" 기법을 도입하여 공간적·시간적 의존성을 통합적으로 고려하였다. 이는 Markov-CA와 LightGBM 모델을 결합함으로써 외부 요인의 영향을 체계적으로 분석하는 새로운 방법론을 제시한다.

시계열 데이터에 나타난 공장의 입지 분포를 분석한 결과(그림 2), 산지나 하천과 같은 자연적 제약요소가 있는 지역이나 용도지역 규제로 인해 공장 설립이 제한된 구역에서는 지속적으로 공장이 부재한 것으로 나타났다. 이러한 패턴은 시계열 데이터 상에서 공장이 지속적으로 입지하지 않는 지역이 다양한 외부 요인에 의한 입지 제한 구역임을 시사한다. 따라서 이러한 특성을 모델에 반영함으로써, 추가적인 외부 요

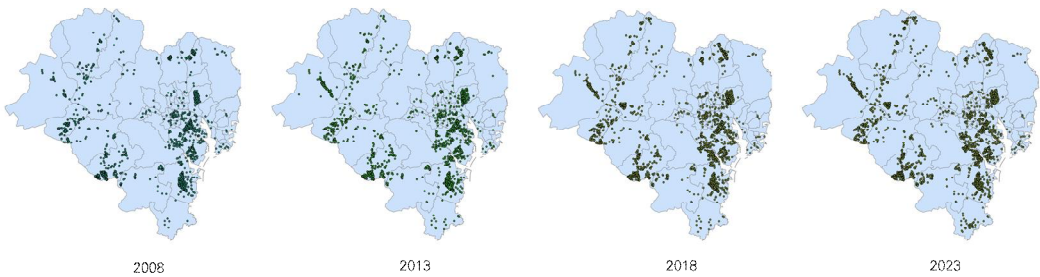


FIGURE 2. Time series data: factory locations

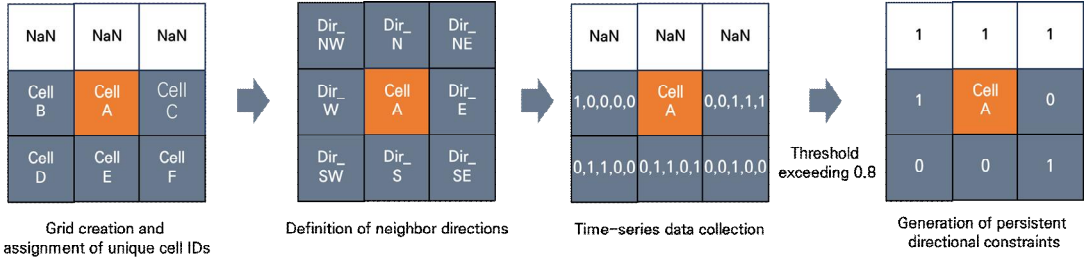


FIGURE 3. Analytical procedures for excluding externalities

인 데이터 없이도 입지 제한 요인을 효과적으로 고려할 수 있을 것으로 판단된다.

이를 구현하기 위해 그림 3과 같이 대상 지역을 X, Y 좌표 기반의 그리드 셀로 분할하고, 2008년부터 2022년까지의 시계열 데이터를 구축하여 각 셀의 상태 변화를 분석하였다. 특정 방향에서 반복적으로 상태 변화가 확인된 경우, 해당 방향에 대한 제약 조건을 설정하였다. 이러한 제약 조건은 외부 요인의 영향을 체계적으로 분석하는 데 핵심적 역할을 한다.

Markov-CA 모델은 이러한 지속적 방향 제한성을 입력 변수로 활용하여 이웃 셀 간의 관계와 시간적 의존성을 동시에 반영한다. 이를 통해 각 셀의 상태 전이를 더욱 정확하게 예측할 수 있게 되었다. LightGBM 모델은 Markov-CA의 결과를 기반으로 최종 예측을 수행하며, 지속적 방향 제한성으로 생성된 특성을 주요 입력 변수로 활용하여 예측의 정확도와 재현율을 향상시켰다. 이러한 방법론은 외부 요인의 영향을 체계적으로 반영하면서도 모델의

예측 성능을 극대화하는 데 주안점을 두었다.

3. 모델 학습 및 학습 결과

본 연구에서는 모델의 최적 성능을 도출하기 위해 베이지안 최적화 기법을 활용하였다. Optuna 프레임워크를 통해 총 100회의 반복 학습을 수행하여 최적의 하이퍼파라미터를 표 2와 같이 도출하였다. 최적화 결과, 시공간 윈도우 크기는 7로 설정되었으며, 정방형(Square) 윈도우 형태가 선정되었다. LightGBM 모델의 주요 파라미터로는 max_depth가 4, num_leaves가 41로 설정되어 과적합을 방지하면서도 충분한 모델 복잡도를 확보하였다. learning_rate는 0.016으로 비교적 낮은 값을 채택하여 모델의 안정적인 학습을 도모하였다.

특히 주목할 만한 점은 feature_fraction이 0.959로 설정되어 대부분의 특성이 학습에 활용되었다는 것이다. 이는 본 연구에서 도입한 지속적 방향 제한성 기법을 통해 생성된 특성들

Table 2. Optimized values for variables via hyperparameters

Model	Parameter	Value
Markov-CA	window_size	7
	window_shape	Square
	max_depth	4
	num_leaves	41
	learning_rate	0.016291
LightGBM	feature_fraction	0.959424
	bagging_fraction	0.75014
	bagging_freq	2
	scale_pos_weight	3.455489
	best_threshold	0.48

이 모델 성능 향상에 실질적으로 기여했음을 시사한다. 또한 bagging_fraction은 0.750, bagging_freq는 2로 설정되어 모델의 안정성과 일반화 성능을 향상시켰다. 불균형 데이터 처리를 위한 scale_pos_weight는 3.455로 조정되었다. 하이퍼파라미터로 도출된 최적의 예측 임

계값(best_threshold)은 0.48로, 이를 기준으로 셀별 공장 입지 여부를 결정하였다. 구체적으로, 각 셀의 공장 입지 확률이 0.48 이상인 경우는 공장 존재로, 0.48 미만인 경우는 공장 부재로 후처리를 수행하였다.

본 모델의 성능 검증을 위해 2008년부터

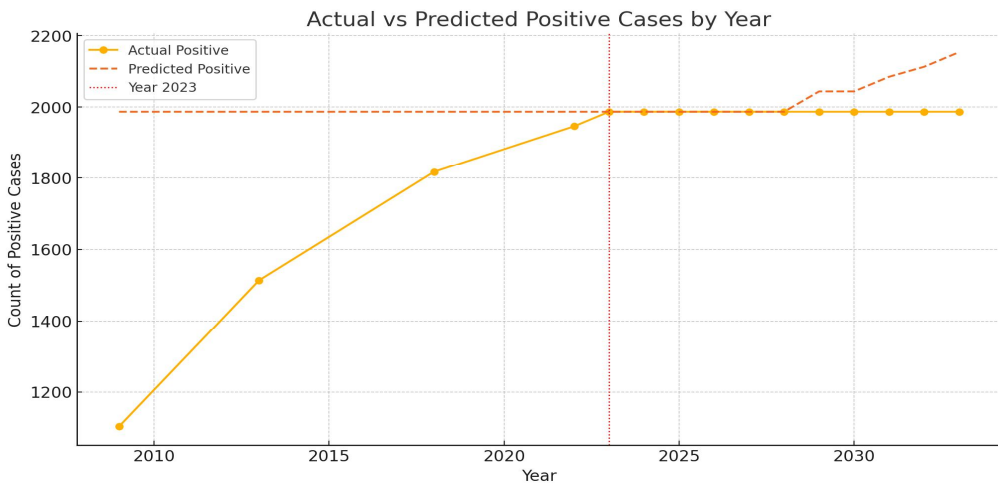
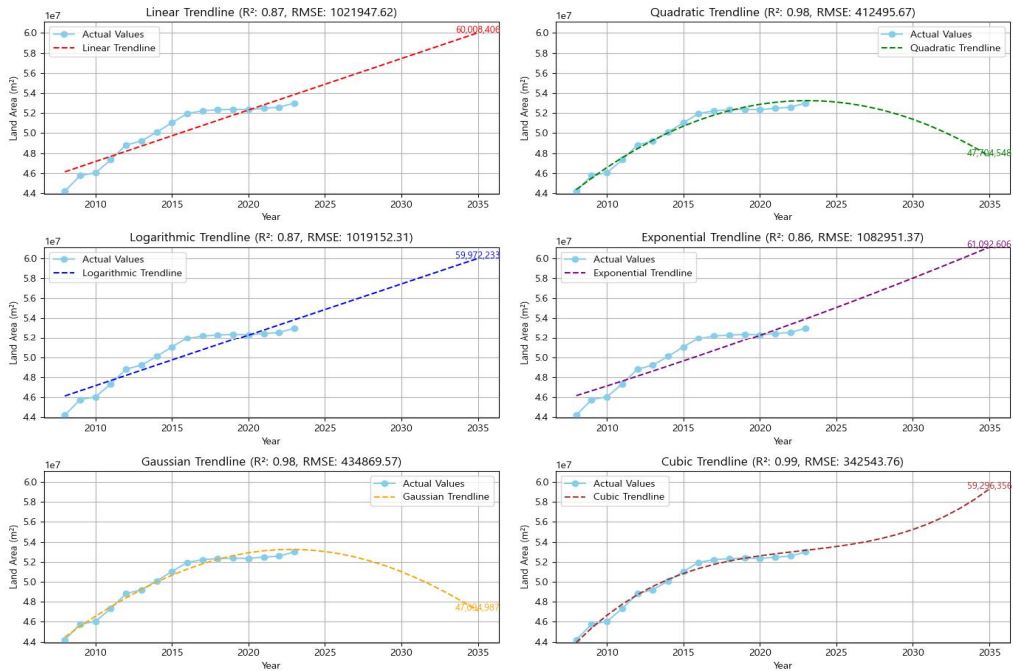


FIGURE 4. Future demand forecast graph(Top) and learning model results(Bottom)

2022년까지의 데이터를 학습에 활용하고, 2023년 데이터를 통해 검증을 수행하였다. 혼동 행렬 분석 결과, True Positives 1,979건, False Negatives 5건, False Positives 3건, True Negatives 152,580건으로 나타났다. 이를 기반으로 산출된 성능 지표는 정밀도 (Precision) 0.9985, 재현율(Recall) 0.9975, F1 Score 0.9980으로, 모델이 우수한 예측 정확도를 보여주었다. 특히 False Positives와 False Negatives의 낮은 발생률은 본 모델이 시설 입지 예측에 있어 높은 신뢰성을 제공할 수 있음을 시사한다.

본 모델의 높은 재현율은 과적합 가능성을 제기할 수 있으나, 이는 본 연구에서 활용된 셀룰러 오토마타의 특성에 기인한 것으로 판단된다. 구체적으로, 입력되는 기준 셀들의 값이 안정적으로 유지되며 외부적 확장 방향으로 분석이 이루어지는 모델의 특성상, 기준년도와 예측결과 간 높은 적합도는 타당한 것으로 해석된다.

그림 4에서는 모델 예측의 정확성을 검증하기 위해 선형, 로그, 지수, 다항식 등 6가지 미래 추정식을 비교 분석하였다. 분석 결과, 3차 다항식이 R^2 값 0.99, RMSE 342,543.76으

로 가장 우수한 적합도를 보였다. 학습된 모델을 활용하여 2008년 기준 데이터로부터 2009년 이후의 시설 입지를 순차적으로 예측한 결과, 예측된 시계열 패턴이 3차 다항식의 추세와 높은 유사성을 보였다. 다만, 본 연구의 모델은 격자 단위에서 시설의 입지 여부만을 이진 분류하는 특성상, 실제 시설의 절대적 개수와 직접적인 비교가 제한되는 한계가 있다. 따라서 시계열 패턴의 전반적인 추세 분석을 통해 모델의 예측 성능을 간접적으로 검증하였으며, 이를 통해 모델이 시설 입지의 시간적 변화를 효과적으로 포착하고 있음을 확인하였다.

그림 5는 "지속적인 방향 제한성" 개념이 적용된 공장 입지 예측 모델의 2029년부터 2033년까지의 예측 결과를 보여주며, 기존 시설 입지와의 관계를 효과적으로 반영하여 무분별한 확장을 방지하고 균형 잡힌 입지 패턴이 도출되었음을 시각적으로 나타내고 있다. 특히 기존 공장 입지 셀 주변으로의 과도한 확장이 제한되었으며, 기존 시설 입지 셀 사이에 신규 시설 입지가 예측되는 경향이 관찰되었다. 이는 모델이 특정 방향으로의 편향된 확장을 억제하면서도 입지 요인들을 효과적으로 반영하고 있음을

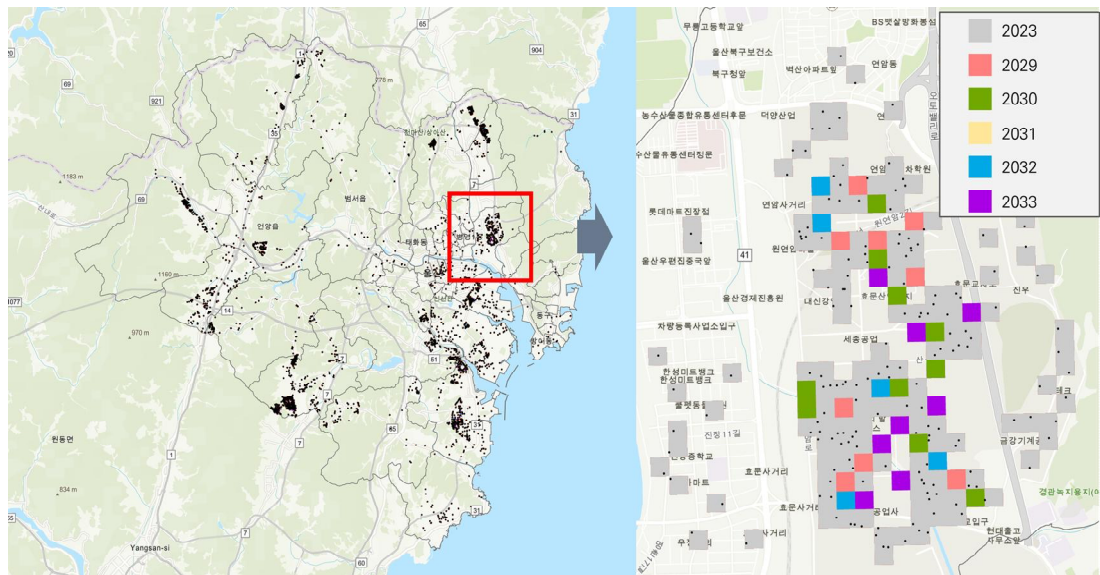


FIGURE 5. Results of location analysis forecast for 2029–2033 based on 2023 baseline

입증한다. "지속적인 방향 제한성" 기법은 도시 계획 규제와 자연환경적 제약 등 현실적인 외부 요인들을 모델에 반영함으로써, 환경 변화 및 제약 조건을 고려한 신뢰성 있는 예측 결과를 제공하였다. 결과적으로 본 연구의 예측 모델은 다양한 입지 요인을 통합적으로 고려하면서도 외부 요인의 배제를 통한 균형적인 입지 결정이 가능함을 시사한다.

결 론

본 연구는 공장의 시계열적 입지 데이터를 활용하여 외부 요인의 간접적 영향을 분석하고, 입지 결정 요인을 도출하며, 미래 입지를 예측하는 새로운 방법론을 제시하였다. 연구 결과를 통해 다음과 같은 주요 결론을 도출할 수 있었다.

첫째, 본 연구에서 제안한 "지속적 방향 제한성" 개념은 도시계획 규제와 자연환경적 제약과 같은 외부 요인의 간접적 영향을 효과적으로 포착하였다. LightGBM 모델의 특성 중요도 분석을 통해 이러한 외부 요인들이 공장 입지 결정에 미치는 영향을 정량적으로 평가할 수 있었으며, 이는 기존 연구에서 간과되었던 중요한 발견이다.

둘째, 시계열적 공장 위치 데이터(X, Y 좌표)만으로도 입지 결정의 주요 요인을 효과적으로 도출할 수 있음을 입증하였다. 특히, 인접 셀의 공장 입지 여부와 지속적 방향 제한성과 같은 파생 변수들이 모델의 예측 성능 향상에 크게 기여하였다. 이는 추가적인 외부 데이터 없이도 시공간적 패턴 분석만으로 의미 있는 결과를 도출할 수 있다는 점에서 실무적 활용도가 높다.

셋째, Markov-CA와 LightGBM을 결합한 하이브리드 모델은 2023년 검증 데이터에서 정밀도 99.85%, 재현율 99.75%, F1 스코어 99.80%를 달성하여 연구질의에 부합하는 예측 성능을 보여주었다. 이는 기존의 단일 모델 접근법이 가지고 있던 한계를 극복하고, 보다 정확하고 신뢰성 있는 예측이 가능함을 시사한다.

본 연구의 학술적 의의는 다음과 같다. 첫째, 공장 입지 예측 분야에서 새로운 방법론적 프레

임워크를 제시하였다. 둘째, "지속적 방향 제한성" 개념을 통해 외부 요인의 간접적 영향을 모델링하는 혁신적인 접근법을 개발하였다. 셋째, 시계열적 위치 데이터만으로도 외부효과요인을 반영하여 미래 예측을 수행할 수 있음을 알 수 있었다.

본 연구의 주요 한계는 공장 유형의 세분화된 분석이 불가하였다는 점에 있다. 이는 도시 전역에 분포한 공장의 입지 수가 제한적인 상황에서, 공장 유형을 세분화할 경우 유형별 학습 데이터의 부족 현상이 발생하기 때문이다. 이러한 데이터의 한계로 인하여 본 연구는 공장 유형별 세부 입지 분석을 수행하지 못하는 방법론적 한계점을 내포하고 있다.

향후 연구에서는 다음과 같은 방향으로 연구를 확장하고자 한다. 첫째, 격자 데이터 생성 시 현재의 공장 존재 여부뿐만 아니라 격자 내 공장의 빈도값을 추가하여 공장 밀집도가 미치는 영향을 보다 정교하게 분석할 계획이다. 이를 통해 공장 군집화 현상이 미래 입지 선정에 미치는 영향을 더욱 정확하게 파악할 수 있을 것으로 기대된다. 둘째, Transformer 모델을 적용하여 장기간에 걸친 시간적 변동성을 심층적으로 분석하고자 한다. 이는 시계열 데이터의 장기 의존성을 효과적으로 포착하여 예측 모델의 성능을 더욱 향상시킬 수 있을 것으로 예상된다.

본 연구는 데이터 기반의 공장 입지 예측 모델을 제시함으로써, 산업 입지 계획 및 정책 수립 과정에서 실질적으로 활용 가능한 과학적 근거를 제공하였다는 점에서 의의가 있다. 향후 본 연구의 방법론이 다양한 실무 환경에서 적용되고 발전되어, 보다 효율적이고 합리적인 산업 입지 의사결정에 기여할 수 있기를 기대한다.

KAGIS

사 사

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(과제번호 RS-2022-00143404).

REFERENCES

- Anas, A., R. Arnott and K.A. Small. 1998. Urban spatial structure. *Journal of Economic Literature* 36(3):1426–1464.
- Choi, C.G. and W.Y. Lee. 2008. Defining factors of unplanned factories' location in the suburbs of Seoul Metropolitan Area by using AHP. *Journal of the Korean Regional Development Association* 20(3): 21–38 (최창규, 이원영. 2008. 수도권 교외 지역 개별입지공장 입지 요인에 대한 연구 – 경기도 화성시내 개별입지공장을 대상으로 AHP를 이용한 실증 분석-. *한국지역개발학회지* 20(3):21–38).
- Choi, H.H. and S.B. Kim. 2011. The characteristics of changes of spatial structure in Ulsan. *Journal of the Architectural Institute of Korea* 27(3): 261–268 (최호현, 김선범. 2011. 울산시 공간 구조의 변화특성. *대한건축학회논문집* 27(3): 261–268).
- Gu, C., X. Wang and Y. Zhang. 2014. Simulation of urban growth and form with cellular automata: A case study of Shenzhen, China. *Environmental Systems Research* 3(5). <https://doi.org/10.1186/s40068-014-0026-6>.
- Huang, B., L. Zhang and B. Wu. 2009. Spatiotemporal analysis of rural-urban land conversion. *International Journal of Geographical Information Science* 23(3): 379–398.
- Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye and T.Y. Liu. 2017. LightGBM: A highly efficient gradient boosting decision tree. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp.3146–3154.
- Li, X. and A.G. Yeh. 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science* 16(4):323–343.
- Wu, F. and C.J. Webster. 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B: Planning and Design* 25(1): 103–126.
- Yang, Q., X. Li, and X. Shi. 2008. Cellular automata for simulating land use changes based on support vector machines. *Computers & Geosciences* 34(6):592–602. 