KISDI

정책자료 25-12-05

AI기반의 인구이동 예측모형 개발

국토연구원

장요한·손재선·김수지

KISDI 정보통신정책연구원
KOREA INFORMATION SOCIETY DEVELOPMENT INSTITUTE

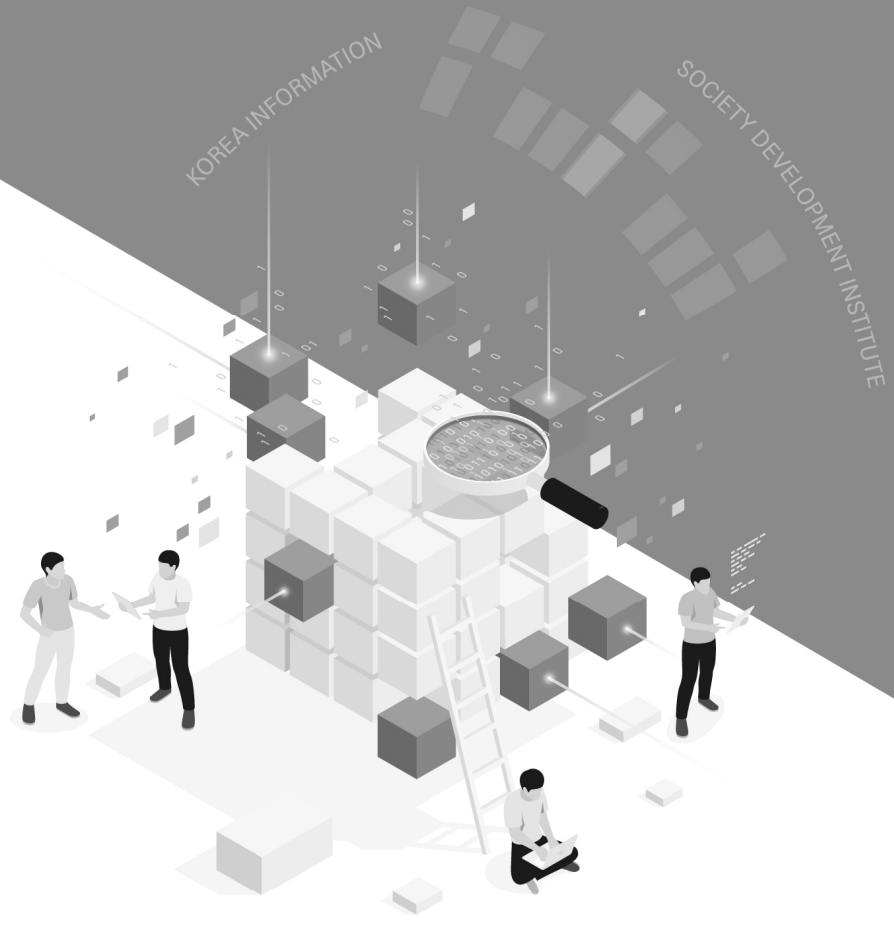
KISDI

정책자료 25-12-05

AI기반의 인구이동 예측모형 개발

국토연구원

장요한·손재선·김수지



KISDI 정보통신정책연구원
KOREA INFORMATION SOCIETY DEVELOPMENT INSTITUTE

2025 데이터 기반 미래예측·정책지원 사업

□ 연구보고총서 시리즈

연구보고총서 일련번호	연구보고서명	연구기관
25-12-01	2025 데이터 기반 미래예측·정책지원 사업	정보통신정책연구원
25-12-02	인구구조 변화를 고려한 ICT 산업 고용 분석 및 전망 모형 연구	정보통신정책연구원
25-12-03	LLM을 통한 AI 직업 노출도 측정 연구	정보통신정책연구원
25-12-04	데이터 기반 헌법상 기본권 입법수요 예측 모델 고도화	한국법제연구원
25-12-05	AI기반의 인구이동 예측모형 개발	국토연구원

□ 「AI기반의 인구이동 예측모형 개발(정책자료 25-12-05)」 참여연구진

연구기관	총괄책임자	참여연구원
국토연구원	장요한 부연구위원	손재선 부연구위원 김수지 부연구위원

목 차

CONTENTS

• 국문요약 ····· 1

• Abstract ····· 3

• 정책제언 ····· 5

제1장 연구의 개요

제 1 절 연구의 배경 및 목적 ····· 7

1. 연구의 배경 ····· 7
2. 연구의 목적 ····· 10

제 2 절 연구의 범위 및 방법 ····· 11

1. 연구의 범위 ····· 11
2. 연구의 방법 ····· 11

제 3 절 연구의 차별성 및 기대효과 ····· 14

1. 연구의 차별성 ····· 14
2. 연구의 기대효과 ····· 14

제2장 AI기반 인구이동 예측모형 개발의 필요성

제 1 절 도시화, 산업구조 변화와 인구이동의 현실 ····· 15

CONTENTS

제 2 절 기존 인구이동예측 모형의 한계 ····· 17

제 3 절 공간 기반 분석과 인공지능 기술의 필요성 ····· 19

제 4 절 정책 수요와 시의성 ····· 21

제 5 절 소결 ····· 22

제3장 선행연구검토

제 1 절 인구이동 요인에 관한 선행연구검토 ····· 23

제 2 절 도시구조 변화와 인구 분포 관련 연구 ····· 27

제 3 절 인구 예측 방법론 AI 기반 연구 동향 ····· 29

제 4 절 소 결 ····· 32

제4장 공간분석 기반 변수설계 및 AI 예측모형 개발

제 1 절 데이터 수집 및 전처리 ····· 33

1. 데이터 수집 ····· 33
2. 데이터 전처리 ····· 35

제 2 절 격자 기반 공간 네트워크 분석 ····· 37

제 3 절 변수 정의 및 주요 입력 구조 설계 ····· 38

제 4 절 예측모형 알고리즘 구성 ····· 41

1. 학습 방법에 따른 구분 ····· 41

2. 인공신경망 구조의 특징 ····· 44
3. 구동 방식에 따른 구분 ····· 45
4. 딥러닝 및 혼합모델 ····· 47
5. 결정계수(hyper parameter) 최적화 ····· 51

제5절 예측 결과 및 성능 검증 ····· 52

제6절 변수 영향도 및 민감도 분석 ····· 54

제5장 AI기반 인구이동 예측 모형 시범 분석

제1절 개요 ····· 56

제2절 데이터 전처리 ····· 57
1. 인구이동 데이터 ····· 57
2. 기타 데이터 ····· 61

제3절 비교모델군 선정 ····· 63

제4절 시범 분석 결과 ····· 64

제5절 정책 시나리오 구성 및 적용 방향 ····· 71
1. 정책 시나리오 구성 방향 ····· 71
2. 정책 시나리오의 적용 방향 ····· 72

제6절 소결 ····· 73

제6장 결론 및 향후 과제

제1절 연구 요약 및 핵심 성과 ····· 74

CONTENTS

제 2 절 기술적·정책적 시사점 ····· 75

제 3 절 한계점 및 향후 연구 방향 ····· 76

• **참고문헌** ····· **78**

표 목 차

〈표 4-1〉 본 연구에서 활용한 주요 데이터 항목 및 개요 ····· 34
〈표 4-2〉 본 연구에서 활용한 주요 변수 개요 ····· 40
〈표 5-1〉 예측 목표연도에 따른 각 인구이동 예측모형의 예측 결과 비교 ····· 64
〈표 5-2〉 예측 목표연도에 따른 직전 주소지(읍면동) 예측 결과 비교 ····· 69

그림목차

[그림 1-1] AI 기반 인구이동 예측모형 개발을 위한 본 연구의 틀 ····· 13
[그림 2-1] 연도별 주요 도시 인구이동 시각화 예시 ····· 16
[그림 2-2] 공간정보를 활용한 인공지능(Geo-AI) 소개 예시 ····· 20
[그림 3-1] 연령대별 인구이동 유형 ····· 24
[그림 3-2] 순환적·누적적 인과과정 모형도 ····· 24
[그림 3-3] 인구성장률과 지역의 인구학적 사회적 요인들 간의 상관관계 ····· 25
[그림 3-4] 주요 권역별 인구 유출입(순이동: 1970~2014) ····· 26
[그림 3-5] 인구이동의 인구학적, 사회학적 파급효과(인구이동에 의한 종합적 악순환 구조) ····· 28
[그림 3-6] 선행연구에서 활용된 주요 AI 모형 도식화 ····· 29
[그림 3-7] CNN-LSTM 구조 도식화 예시 ····· 30
[그림 3-8] AI를 활용한 1km 격자 단위 장래인구 예측(2025~2050년) ····· 31
[그림 4-1] 본 연구의 읍면동 및 격자 단위 인구이동 집계방식 도식화 ····· 35
[그림 4-2] 그래프 이론의 연결중심성(a)과 매개중심성(b) 개념도 ····· 36
[그림 4-3] Cellular Geography 개념 도식화 ····· 37
[그림 4-4] AI-PriSM(가칭)의 주요 학습 및 예측 방식 도식화 ····· 39
[그림 4-5] 지도학습(Supervised learning) 개념도 ····· 41
[그림 4-6] 비지도학습(Unsupervised learning) 개념도 ····· 42
[그림 4-7] 준지도학습(Semi-supervised learning) 개념도 ····· 43
[그림 4-8] 강화학습(Reinforcement learning) 개념도 ····· 43
[그림 4-9] 인공신경망의 일반적인 구조 ····· 45
[그림 4-10] 분류 기반 모형의 개념도 ····· 46
[그림 4-11] 딥러닝 인공신경망의 일반적인 구조 ····· 47
[그림 4-12] 합성곱 신경망(CNN)의 일반적인 구조 ····· 48
[그림 4-13] CNN-LSTM의 일반적인 구조 ····· 48
[그림 4-14] CPU와 GPU의 차이 도식화 예시 ····· 49
[그림 4-15] GAT의 작동 방식 도식화 ····· 50
[그림 4-16] ST-GAT의 작동 방식 도식화 ····· 50
[그림 4-17] Tabu-search(a)와 Hill-climbing(b), 유전자알고리즘(c) 개념 도식화 예시 ····· 51
[그림 4-18] 모형의 정산 및 보정 개념도 ····· 52
[그림 4-19] SHAP 예시 ····· 54

CONTENTS

[그림 5-1] 주요 연도별 주요 인구이동 패턴(Top 30 지역) 시각화 ····· 57
[그림 5-2] 정보 누락 연도의 보간법 적용 예시 ····· 57
[그림 5-3] 주요 전입 읍면동(a)과 전출 읍면동(b) ····· 58
[그림 5-4] 최다 인구이동 월(a)과 최다 인구이동 연도(b) ····· 59
[그림 5-5] 주요 전출 사유 분포 시각화 ····· 59
[그림 5-6] 주요 연도별·시도별 인구이동 시각화 ····· 60
[그림 5-7] 지방인허가 데이터 예시: 보건의료(a)와 음식점(b) ····· 61
[그림 5-8] 주요 열차역·지하철역 위치 및 5km 버퍼 시각화 예시: 2010년(a)과 2023년(b) ····· 62
[그림 5-9] 모형별 RMSE와 MAE 비교 ····· 65
[그림 5-10] SHAP 도식화(2025년 예측모형 예시, 상위 20개 변수) ····· 66
[그림 5-11] 예측 목표연도에 따른 각 모형의 인구에 따른 기타 주요 변인과의 SHAP 비교 ····· 67
[그림 5-12] 예측 연도에 따른 인구이동 패턴 ····· 68
[그림 5-13] 과거 연도 인구이동 패턴 그래프 시각화(상위 30 읍면동 지역) ····· 70
[그림 5-14] 예측 연도에 따른 인구이동 패턴 그래프 시각화(상위 30 읍면동 지역) ····· 70

국문요약

인구이동(migration)은 “일정한 지역에 살던 사람이 읍면동 경계를 벗어나 다른 곳으로 거주지를 옮긴 경우”로 정의되며, 더 넓게는 “정치·행정의 경계를 넘는 거주지의 변화”로 이해된다. 이는 단순한 주거지의 ‘이전(移轉)’과는 구분되며, 일정 거리를 넘어 삶의 기반이 되는 생활공간이 장기적 또는 영구적으로 변화하는 개념을 포함한다. 인구이동은 출생이나 사망과 달리 단기간에 반복적으로 발생할 수 있으며, 정책변화나 경제적 여건 등 사회경제적 요인에 민감하게 반응하는 특성을 가진다. 우리나라에서는 국토의 관리와 중장기적 계획수립의 목적으로 인구이동 집계 통계를 주기적으로 생산하고 있으나, 그 단위가 행정구역 중심으로 이루어져 변화하는 행정구역과 생활권이나 교통망 변화 등 도심지 공간 패턴 분석에 한계가 있었다. 또한, 전통적인 코호트요인법은 미래 시나리오 가정에 의존하여 급변하는 사회·경제적 환경과 비선형적 요인을 충분히 반영하지 못하는 단점이 있었다.

본 연구에서는 이러한 한계를 극복하기 위해 1km 격자 단위의 미시적 공간분석과 인공지능(AI)을 결합한 인구이동 예측모형(AI-based Policy-oriented Spatio-temporal Prediction for Population Migration, AI-PriSM, 가칭)을 개발하는 것을 목표로 한다. 이를 통해 인구이동의 시공간적 패턴을 정밀하게 예측하고, 정책 시뮬레이션을 통한 의사결정 지원에 활용하고자 한다.

AI-PriSM(가칭)에서는 인구이동 예측모형과 직전 주소지 예측모형을 구분하여 개발하였다. 먼저, 인구이동 예측모형에서는 인구의 이동에 상관관계가 높을 것으로 예상되는 상권정보, 교통접근성 등 정주 여건과 사업체 수 및 종사자 수 등을 함께 고려하여 Cellular Geography 개념을 적용한 격자 재배분 과정을 거쳤고, XGBoost, 다층인공신경망(Multilayer Neural Network), 강화학습(Reinforcement Learning) 등을 활용하여 비교 분석하였다. AI 모형의 접목 과정에서는 Tabu-search, 유전자알고리즘 등 다양한 휴리스틱(Heuristic) 방법론을 통한 하이퍼파라미터(Hyper parameter) 최적화를 함께 구성하였다. 직전 주소지 예측모형을 위해서는 그래프 네트워크(Graph Network) 기반의 혼합 딥러닝 모형인 Spatio-temporal Graph Attention

Network(ST-GRAT) 모형을 활용하여 다수의 기점과 종점 간의 패턴을 학습하도록 구성하였다. 분석 결과, 인구 규모(총인구, 가구수) 자체가 인구이동의 가장 큰 요인으로 확인되었으며 철도 및 지하철 접근성, 사업체 밀도 등이 주요 영향을 미치는 것으로 나타났다. 이는 수도권과 비수도권 간의 인구집중 현상을 설명하는 주요 근거가 될 것이다. 직전 주소지 예측모델(ST-GRAT)은 지역 간 이동 패턴 학습에는 효과적이었으나, 이동량의 정량적 예측에는 일부 한계를 보여 향후 고도화의 필요성이 확인되었다.

본 연구는 격자 단위의 세밀한 데이터와 AI 기술을 결합하여 기존 통계적 방법론의 한계를 기술적으로 극복하고, AI 모형에서 주요 문제로 대두되는 '블랙박스(blackbox)' 문제를 완화할 수 있는 설명 가능한(Explainable) 예측 프레임워크를 제시하였다는 데 기술적인 의의가 있다. 개발된 모형은 교통 인프라 확충, 택지 개발, 산업 단지 조성 등 정책 개입에 따른 인구 변화를 사전에 시뮬레이션하는 도구로 활용될 수 있을 것으로 기대한다. 이를 통해 정책 실패 위험을 줄이고 예산 배분의 효율성을 높여 지방소멸 대응 및 국토 균형발전 전략 수립에 기여할 것으로 기대한다.

향후 연구에서는 AI 알고리즘의 고도화 및 추가 적용과 정책 시나리오 고도화 작업 등을 통한 효용성 제고가 필요할 것이다.

Abstract

Population migration is defined as “a case where a person living in a certain area moves their residence beyond the boundaries of an Eup, Myeon, or Dong,” and more broadly, it is understood as “a change of residence crossing political or administrative boundaries.” This is distinct from a simple “move” of housing; it encompasses the concept of a long-term or permanent change in the living space that forms the basis of life, exceeding a certain distance. Unlike birth or death, population migration can occur repeatedly over a short period and is characterized by its sensitivity to socio-economic factors such as policy changes and economic conditions. In Korea, population migration statistics are periodically produced for the purpose of national land management and medium-to-long-term planning. However, since these statistics are compiled based on administrative districts, there have been limitations in analyzing urban spatial patterns, such as changing administrative boundaries, living spheres, and transportation networks. Furthermore, the traditional Cohort-Component Method has had shortcomings in that it relies on future scenario assumptions and fails to sufficiently reflect rapidly changing socio-economic environments and non-linear factors.

To address these limitations, this study aims to develop an AI-based Policy-oriented Spatio-temporal Prediction for Population Migration(AI-PriSM, tentative name) model that combines 1km grid-based micro-spatial analysis with Artificial Intelligence(AI). The goal is to precisely predict spatio-temporal patterns of population migration and utilize them to support decision-making through policy simulations.

The AI-PriSM(tentative name) framework was developed by distinguishing between a Population Migration Prediction Model and a Previous Address Prediction

Model. First, the Population Migration Prediction Model incorporated settlement conditions expected to have a high correlation with migration, such as commercial information and transportation accessibility, along with the number of businesses and employees. It underwent a grid redistribution process applying the concept of Cellular Geography and was comparatively analyzed using XGBoost, Multilayer Neural Networks(MLP), and Reinforcement Learning(RL). In the process of integrating AI models, hyperparameter optimization was configured using various heuristic methodologies such as Tabu-search and Genetic Algorithms(GA). For the Previous Address Prediction Model, Spatio-temporal Graph Attention Network(ST-GRAT), a hybrid deep learning model based on Graph Networks, was utilized to learn patterns between multiple origins and destinations.

The analysis results confirmed that population size(total population, number of households) itself is the most significant factor in population migration. Additionally, rail and subway accessibility and business density were found to have a major influence. These findings provide key evidence explaining the phenomenon of population concentration in the Seoul metropolitan area compared to non-metropolitan areas. While the Previous Address Prediction Model(ST-GRAT) was effective in learning inter-regional movement patterns, it showed some limitations in the quantitative prediction of migration volume, confirming the need for future advancement.

This study holds technical significance in that it technically overcomes the limitations of existing statistical methodologies by combining grid-based data in conjunction with AI technology. It also presents an Explainable AI(XAI) prediction framework capable of mitigating the “black box” problem often cited as a major issue in common AI models. The developed model is expected to be utilized as a tool to simulate population changes in advance according to policy interventions such as transportation infrastructure expansion, housing land development, and industrial complex creation. Through this, it is anticipated to contribute to reducing the risk of policy failure, improving the efficiency of budget allocation, and establishing strategies for responding to local extinction and balanced national development.

Future research will require enhancing utility through the advancement and additional application of AI algorithms, as well as the refinement of policy scenarios.

정책제언

본 연구는 기존 행정구역 중심의 통계적 추계 방식에서 벗어나, 각종 공간 기반의 공공 빅데이터와 AI 기술을 결합한 1km 격자 단위 인구이동 예측모형(AI-PriSM)을 개발하고 그 유효성을 입증하였다. 연구 과정에서 도출된 변수 중요도(SHAP) 분석 결과와 시범 시뮬레이션 성과를 바탕으로, 향후 효과적인 국가 균형발전 및 인구 정책 수립을 위한 몇 가지 정책제언을 제시한다.

1. 데이터 기반의 '마이크로 공간' 정책 패러다임 전환

기존의 시·군·구 단위 인구 정책은 행정 경계 내의 총량적 접근에 그쳐, 실제 인구의 생활권과 활동 반경을 반영하지 못하는 한계가 있었다. 본 연구의 시범 분석 결과, 인구이동은 행정구역 경계보다는 철도역, 상권, 직장 밀집도 등 미시적인 공간 특성에 따라 비선형적으로 반응함이 확인되었다. AI-PriSM(가칭)을 활용하여 인구 소멸 지역과 성장 잠재 지역 등을 1km 단위 등으로 정밀 진단하고, 교통축 등 행정 경계를 넘어서는 광역적 생활권에 기반한 지원 정책 수립이 필요하다.

2. '정주 여건' 중심의 지방소멸 대응 전략 재편

본 연구의 변수 민감도 분석(SHAP) 결과, 인구 규모 자체뿐만 아니라 철도 및 지하철 접근성, 사업체 및 종사자 수가 인구 유입에 결정적인 영향을 미치는 것으로 나타났다. 이는 단순한 주택 공급 위주의 정책보다는 교통 인프라와 일자리가 결합된 복합적 정주 여건 개선이 인구 유입의 핵심임을 시사한다. 수도권과 비수도권의 격차가 교통 접근성과 직장 밀집도에 기인하는 부분이 높은 것을 확인한 만큼, 비수도권 중심의 거점도시 확산과 광역교통망 등을 고려할 필요가 있다.

3. 사전 시뮬레이션을 통한 '증거 기반(Evidence-based)' 의사결정 체계 도입

대규모 예산이 투입되는 택지개발, 산업단지 조성, 교통망 확충 사업은 높은 매몰 비용으로 사후 교정이 불가능에 가깝다. 신규 정책 고려 과정에서 AI-PriSM(가칭)을 이용한 개입 시나리오(Policy intervention)를 확인하여 인구 유입 및 유출 효과나 주변 지역의 인구 빨대 효과 등 부작용을 사전에 검증하는 절차를 고려할 필요가 있다. 이는 과학적 의사결정 체계의 중요 축으로 작용할 것으로 기대한다.

제1장
연구의 개요

제 1 절 연구의 배경 및 목적

1. 연구의 배경

인구이동(migration)은 “일정한 지역에 살던 사람이 읍면동 경계를 벗어나 다른 곳으로 거주지를 옮긴 경우”로 정의되며(국가데이터처 지표서비스), 더 넓게는 “정치·행정적 경계를 넘는 거주지의 변화”로 이해된다(이상림·최인선, 2023). 이는 단순한 주거지의 ‘이전(移轉)’과는 구분되며, 일정 거리를 넘어 삶의 기반이 되는 생활공간이 장기적 또는 영구적으로 변화하는 개념을 포함한다(남형권, 2020; 이상림·최인선, 2023). 이러한 정의는 인구학적 접근에서도 유지되는데, 인구학에서는 인구이동을 국내(internal) 및 국제(international) 이동 모두를 포괄하는 개념으로 보고, 출생 및 사망과 함께 지역 간 인구 규모 및 구조 변동에 영향을 미치는 주요 인구학적 요소로 간주한다(오은열·문채, 2016; 이상림·최인선, 2023).

인구이동의 원인에 대해서는 다양한 논의가 존재한다. 통계적 관점에서는 대규모 신규 아파트 입주, 재개발·재건축 등 토지이용 행태의 변화, 부동산 경기나 고용 여건 변화와 같은 사회·경제적 요인, 출산율 저하 및 고령화에 따른 인구구조 변화, 도시화·산업화로 인한 교통·통신 인프라의 발달 등이 주요 원인으로 제시된다(국가데이터처 지표서비스). 그러나 인구이동은 출생이나 사망과 달리 단기간에 반복적으로 발생할 수 있으며, 정책변화나 경제적 여건 등 사회경제적 요인에 민감하게 반응하는 특성을 가진다. 이러한 특성은 인구이동이 지역 인구변화에 가장 직접적이고 강력한 영향을 미치는 요인이며, 동시에 성별이나 연령 등 생물학적 요인의 영향은 상대적으로 미미하다는 점에서 사회적 요인에 기인한 결과로 해석될 수 있다(이상림·최인선, 2023).

정보 관점에서 볼 때, 인구의 출생이나 사망은 생성 또는 소멸로 인구집계에 직접적인 영향을 미치는 반면, 인구의 이동은 지역 간 유출(전출, 轉出)과 유입(전입, 轉入)이라는 방향성을 가지며,

지리적 특성과 더불어 '흐름(flow)'이라는 동태적 속성을 내포한다는 점에서 차별화된 특징을 지닌다. 인구이동 현상은 주택, 일자리, 교육 등 개인적 요인에 의한 결과이기도 하지만, 지역 환경 변화에 따른 집단적 요인이 반영된 결과로도 해석된다(최대식 외, 2024). 이 때문에 인구이동 연구에서는 '누가-왜-어디로' 이동하는지를 분석하는 것이 핵심적인 접근으로 간주된다, 즉, '누구'는 이동 주체의 인구학적 특성, '왜'는 경제적·사회적·환경적 요인과 정책적 요인을 포함한 정주여건, '어디로'는 유출지와 유입지 간의 공간적 관계와 방향성을 의미한다(최대식 외, 2024; 이상림·최인선, 2023).

우리나라 국가데이터처에서 주기적으로 발표하고 있는 인구이동 집계는 시군구 등 행정구역 중심이고,[1] 전국을 대상으로 집계된 인구이동 통계 중에서 학술 및 연구 등 특정 목적에 따라서 전출·입 사유, 이전 주거지, 가족구성원에 관한 개략적인 정보 등 구체적이고 다양한 정보를 함께 제공[2]하고 있다. 현재 국가데이터처가 제공하는 인구이동 통계는 시군구 등 행정구역 단위를 중심으로 집계되며, 전출·입 사유, 이전 주거지, 가족구성원 등에 관한 다양한 부가 정보도 포함되는데, 이러한 공간 단위는 5년마다 수행되는 인구주택총조사 등과의 정보 연계를 염두에 둔 행정적 활용성에 기반한다(이태직, 2010; 이상림·최인선, 2023). 국가 및 지방자치단체의 중장기 계획 수립에서도 행정구역 단위가 기본 공간단위로 활용되어 왔으며, 이에 따라 해당 통계 단위가 가장 적절하다는 인식이 일반화되어 왔다(박승용, 2022; 이보경 외, 2022).

인구 고령화 및 출산율 감소로 인한 인구구조 변화와 국토 전반의 인구감소로 인해 대두되고 있는 지방소멸 대응을 위해서 다양한 지역균형발전 정책이 거론되고 있으나, 행정구역이라는 공간 단위로는 변화하는 인구구조에 대응하기가 쉽지 않아 보다 세밀한 공간 단위 집계 및 대응 필요성이 요구되고 있다(이보경 외, 2022). 생활권이나 교통망, 상권변화 등 공간의 기능적 측면에서 행정구역이라는 공간 단위는 그 효용성의 측면에서 전혀 다른 성질을 나타내고 있다. 한편으로, 국가데이터처에서 제공하는 행정구역 단위의 정보가 개인정보 노출을 염두에 둔 결과로, 개인정보가 노출되지 않는 최소 단위의 접근으로 시군구를 다루고 있는 점 또한 더욱 세밀한 공간 단위 집계를 어렵게 만들고 있는 이유 중의 하나라고 할 수 있다. 그러나 행정구역 경계는 고정된 불변의 단위가 아니라, 행정 편의에 따라 구분된 것으로서 지역 통폐합이나 택지 개발 등에 따라 변경될 수 있는 불확실성을 내포하고 있다. 이러한 점에서 행정구역을 기반으로 한 정책 수립이나 중장기 계획은 본질적인 한계를 지닌다.

한편, 인구는 모든 도시와 지역의 가장 기본적인 주체이자, 공간과 연계된 모든 계획의 기초 지표이므로 지역 인구변화를 예측하는 일은 매우 중요하다(오은열·문채, 2016; 이상림·최인선,

1) KOSIS 국가통계포털. 국가데이터처.(https://kosis.kr/index/index.do, 2025년 7월 31일 접속)
2) 마이크로데이터 통합서비스. 국가데이터처(https://mdis.kostat.go.kr/index.do, 2025년 7월 31일 접속)

2023). 특히 인구이동에 대한 정확한 예측은 미래 인구 규모와 구조를 파악하고, 이에 기반한 적절한 계획을 수립하는 데 필수적이다(이보경 외, 2022; 최대식 외, 2024). 그러나 국가데이터처에서 제공하는 인구이동 관련 정보는 예측보다는 집계 성격이 강하다. 장래인구 추계 정보가 장래 인구이동을 가늠하는 데 참고가 될 수 있으나, 이는 시나리오 기반 추계로 보수적 가정에 의존하기 때문에 불확실성이 크고, 정확도 측면에서도 한계가 있다. 최근 지속되는 도심지 인구집중은 수도권과 비수도권, 대도시와 중소도시 간의 격차를 심화시키며, 일부 지방자치단체의 존립에도 영향을 미칠 수 있다. 이러한 문제를 이해하기 위해서는 인구이동의 패턴뿐 아니라 지역 간 관계 구조에 대한 종합적 분석이 필요하다. 행정 측면에서도 제한된 자원을 효율적으로 배분하고, 장래인구와 인구이동 흐름을 기반으로 사회기반시설 계획이나 인구집중으로 인한 혼잡비용에 유연하게 대응할 수 있어야 한다. 인구이동 예측은 정책 효과를 과학적으로 검증하고, 지역 및 개인 특성별 결정요인을 파악하여 인구 문제에 대한 실질적 정책 방안을 마련하는 데 핵심적이다(이찬영, 2018; 이보경 외, 2022).

기존의 인구이동 예측 모형은 인구변화의 복잡성과 예측의 정확성을 저해하는 여러 구조적 한계를 지니고 있다. 첫째, 국가데이터처의 장래인구 추계는 미래 가정을 전제로 작성되는 시나리오 기반이어서, 가정이 변경될 경우 예측 결과의 불확실성이 크게 증가한다(박승용, 2022). 특히 시·도 단위 인구추계는 총인구 추계에 비해 시나리오의 다양성이 부족하고, 지역 간 잦은 인구이동과 지역별 인구 구성 변화 등 다수의 요인을 동시에 고려해야 하므로 복잡성이 높다(박승용, 2022; 이보경 외, 2022). 이러한 가정 의존성은 이해관계에 따라 추계 데이터가 조정될 수 있는 편향성을 초래한다. 둘째, 일부 지방자치단체는 사회적 유입 인구 증가를 전제로 도시계획을 수립하는 경향이 있어, 계획인구가 실제보다 과다하게 산정되고, 결과적으로 재정 부담을 가중시키는 부정적 결과를 낳기도 한다(박승용, 2022). 셋째, 기존 추계 방식은 인구 이동성이 점차 낮아진다는 가정을 전제로 하기 때문에, 총이동량이 감소하더라도 지역 간 인구 비대칭성이 심화되는 최근 경향을 반영하지 못한다[3](이태직, 2010; 이상림·최인선, 2023). 이로 인해 급변하는 지역 상황에 다양한 시나리오로 대응하기 어렵다. 넷째, 현재 국가데이터처가 사용하는 코호트요인법은 사망률·출생률·순이동률 등의 시계열 추세를 반영하더라도 공간적 특성을 고려하지 않기 때문에, 인구 총량 예측에만 머무르며 토지이용 변화에 따른 세부 공간 분포를 전망할 수 없다(이보경 외, 2022). 마지막으로, 기존 통계 기반 접근에서는 인구이동과 관련한 다양한 요인과 변수를 동시에 고려하기 어렵다. 특히 여러 변수 간 상호작용이 복잡하게 얽히는 경우 데이터 구조가 비선형관계를 이루게 되는데, 이러한 비선형 데이터 구조를 통계 기반 모형으로 분석하고자 하는 경우 적절한 학습과 예측이 어렵고 결과적으로 모형의 성능과 예측 정밀도가 저하된다.

3) 2021년 정기 장래인구추계 공표 시기에도 이전 자료가 급변하는 인구변화를 제대로 반영하지 못해 2019년에 '장래인구 특별 추계'를 공표하기도 했다(박승용, 2022).

최근 다양한 분야에서 접목되고 있는 인공지능(Artificial Intelligence, AI) 기반 접근방법은 다양한 비선형 데이터 모델링에 효과적이며, 예측 정확도와 민감도 분석 등에 강점을 가지는 것으로 알려져 있다. AI 모형은 인구이동에 내재한 사회적 영향과 변수 간 상호작용을 학습하고 탐색할 수 있어, 정책 의사결정을 지원하는 데 높은 활용성과 실효성을 보인다. 또한, 충분히 학습이 완료된 AI 모형은 수 초 내에 시뮬레이션 수행이 가능하고, 복잡한 요인 간 인과관계를 효과적으로 분석할 수 있다는 강점이 있다. 이러한 배경을 고려할 때 전통적인 통계적 접근방법에 국한된 인구이동 예측의 한계 극복을 위해 AI를 활용한 고도화가 시의적절하게 필요하며, 나아가 실시간 분석과 시나리오 기반 예측이 가능한 고성능 정책지원 도구의 개발이 요구된다.

2. 연구의 목적

본 연구는 행정동 등 기존 인구이동 통계의 시공간 제약 한계를 극복하고 생활권, 교통망, 상권 등 공간의 다양한 기능적 특성을 반영한 1km 격자 기반 AI 인구이동 예측 모형을 개발하는 것을 목표로 한다. 이를 위해 시계열 데이터와 공간분석을 결합하고, 머신러닝(Machine Learning, ML)과 딥러닝(Deep Learning, DL) 기법을 적용하여 인구이동의 미시적 시공간 패턴을 효과적으로 예측한다. 개발된 모형은 다양한 시나리오 설정과 특정 정책 변수 변화에 대응할 수 있는 시뮬레이션 기능을 포함하며, 정책의 효과를 사전에 검증하고 다양한 시뮬레이션 결과를 기반으로 최적의 대응 방안을 도출할 수 있는 틀을 제시할 수 있도록 구성한다.

본 연구를 통해 격자 단위의 공간적 해상도 고도화, 정책 시뮬레이션 및 예측의 정확도 개선, 지역 기반 맞춤형 정책 설계 등을 지원하며, 궁극적으로는 국토 및 각 지방자치단체의 선제적 계획수립과 전략적 의사결정, 자원 배분의 효율화, 지방소멸 대응 전략 수립 등에 기여할 수 있을 것으로 기대한다. 기술적으로는 각종 시공간 데이터를 1km 공간 단위 연도별 데이터로 집계 및 재정립할 수 있는 시도를 제시하며, 유사한 시공간 예측모형 분석을 위한 선험적인 사례를 제공하고자 한다.

제 2 절 연구의 범위 및 방법

1. 연구의 범위

인구이동 예측을 위해 본 연구의 주제적 범위는 행정구역 단위로 수행되던 기존 통계적 접근방식 외에 격자라는 새로운 공간단위 분석방법론을 제시하고 생활권, 교통, 상권 등 다양한 공간의 특성을 반영한 격자 기반 AI 예측모형을 개발하는 데 방점을 둔다. 본 연구에서 제시된 격자 기반의 AI 인구이동 예측모형은 인구이동의 미시적 패턴 학습과 분석, 주요 요인 도출, 장래 시나리오 기반 정책 시뮬레이션 방향 제시 등을 포함한다.

본 연구에서 다루는 공간적 범위는 대한민국 전역을 대상으로 하되, 1km 격자 단위를 기본 분석 단위로 설정한다. 격자별 데이터는 인구, 가구, 주택, 사업체, 교통망, 상권 등의 변수를 중심으로 구성한다. 시간적 범위는 국가데이터처에서 공개하고 있는 2000년부터 2023년 또는 2024년까지 최신의 시계열 데이터를 포함하며, 이를 바탕으로 시계열 AI 모형에 학습하여 활용할 수 있도록 구성한다. 예측의 범위는 2025년부터 2030년까지 1년 단위의 단·중기 예측을 목표로 하며 더 나아가 완성된 AI 모형을 바탕으로 정책 시나리오별 영향 분석 방향 제언을 포함한다.

내용적 범위는 인구이동 데이터의 수집·전처리 및 공간 네트워크 분석과 인구이동 영향요인 변수 설계 및 데이터 통합을 포함하며, AI 기반 예측모형 개발 및 검증, 정책시나리오 구성과 민감도 분석, 정책 적용 가능성 및 활용 방안제시까지를 본 연구의 범위로 한다.

2. 연구의 방법

본 연구는 AI 기반 인구이동 예측모형 개발의 필요성 검토, 선행연구 분석, 데이터 수집 및 전처리, 변수 정의와 입력 구조 설계, 예측모형 개발, 모형 검증과 성능평가, 정책 시뮬레이션 및 활용 방안 도출의 순으로 진행된다. 먼저, AI 기반 인구이동 예측모형 개발의 필요성 검토 단계에서는 도시화와 산업구조 변화 등 환경 변화로 인한 인구이동의 현실을 진단하고, 기존 모형의 한계를 검토한다. 더 나아가 이를 보완하기 위한 공간분석과 AI 기술 도입의 시급성을 검토한다. 선행연구 분석 단계에서는 인구이동 요인을 중심으로 한 연구, 도시구조 변화와 인구 분포 관계를 다룬 연구, 인구(이동) 예측 방법론 등을 다각도로 검토하여 본 연구의 차별성과 연계성 방향 등을 도출한다.

데이터 수집 및 전처리 단계에서는 인구이동 데이터, 공간 특성 데이터, 보조 데이터를 포괄적으로 확보한다. 인구 및 인구이동 데이터는 국가데이터처(MDIS),[4] SGIS,[5] 인구주택총조사 등에

서 전출·입, 인구·가구·주택 통계를 수집하고, 공간 특성 데이터는 인구이동에 영향을 미칠 가능성이 높은 교통망(도로·철도), 토지이용 패턴, 상권·소상공인 분포, 주요 시설(학교, 병원 등)을 포함한다. 장기적인 보조 데이터로는 경제지표, 지역개발사업 정보, 민원 데이터 등 정책적·환경적 변수를 활용하여 예측모형의 설명력을 높일 수 있는 방향을 검토한다.

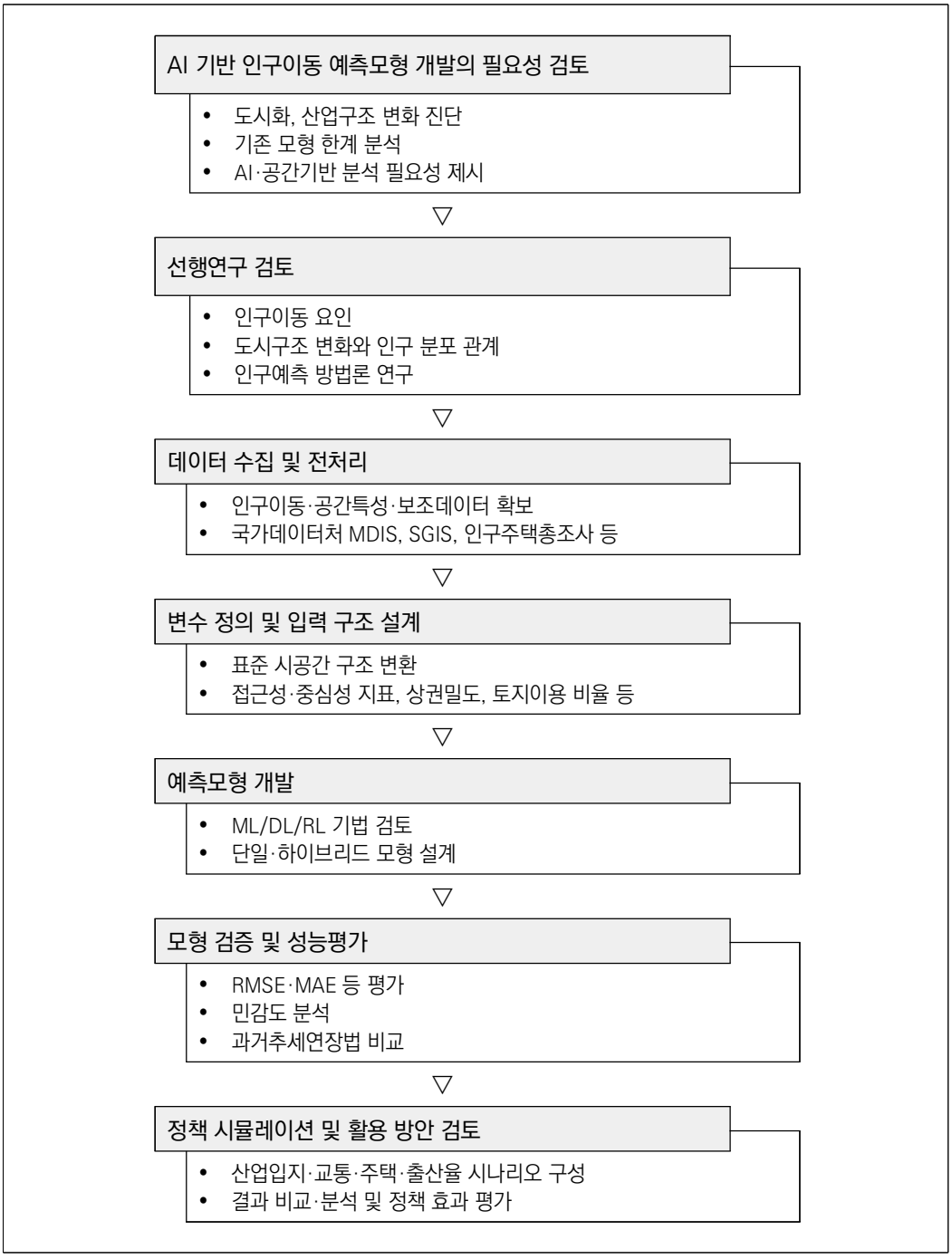
변수 정의와 입력 구조 설계 단계에서는 수집된 데이터를 표준화된 시공간 구조로 변환하고, 격자별로 접근성 지표, 중심성 지표(Betweenness·Degree Centrality), 상권 밀도, 토지이용 비율 등을 직간접적으로 산출한다. 인구 변수로는 격자 단위의 인구 규모, 연령·성별 구조, 가구 특성, 인구변동 특징 등을 포함하며, 시계열 변수로는 연도별 순이동자 수, 출생 수·사망 수, 지역별 경제·사회 지표 변화 등을 정의하여 모형의 주요 입력값으로 활용한다.

머신러닝(ML), 딥러닝(DL), 강화학습(Reinforcement Learning, RL) 등 다양한 AI 기법을 예측모형 개발 단계에서 검토하고, 방법론별 주요 특징과 데이터 특성을 고려한 단일 또는 하이브리드(hybrid) 형태의 최적(Optimized) 모형을 검토한다. 모형 검증과 성능평가 단계에서는 평균제곱근오차(Root-mean-squared-error, RMSE), 평균절대오차(Mean-absolute-error, MAE), 평균제곱오차(Mean-squared-error, MSE) 등을 활용하여 예측 정확도를 평가하고, 추가로 모형 학습과정 중에 파생되는 민감도 분석 과정을 통해 변수별 설명력과 민감도 등을 확인한다. 본 연구에서 개발된 AI 기반 모형의 객관적인 성능 비교를 위해, 과거 추세연장법을 이용한 회귀(Regression) 기반의 기본모형(Baseline)을 제시하고 비교를 통한 성능 비교를 병행한다.

마지막으로, 정책 시뮬레이션 및 활용 방안 검토 단계에서는 앞선 단계에서 확인한 주요 변인별 민감도 결과를 바탕으로 산업입지 변화, 교통 인프라 변화, 출산율 변화 등 다양한 시나리오를 구성하고, 이에 따른 시나리오별 인구이동 예측 결과를 비교·분석하여 정책적 영향을 평가할 수 있는 방향을 검토한다. 이를 바탕으로 지방소멸 대응, 지역 균형발전, 사회기반시설 및 인프라 등의 국가 및 지방자치단체의 전략적 의사결정을 지원할 수 있는 정책도구로써의 활용 방안을 제고한다. 특히, 정책 시뮬레이션 및 활용 방안 검토 부분은 후속 연구를 염두에 둔 부분으로, 본 연구에서 개발하는 격자단위 인구이동 모형의 고도화와 함께 정책활용도 제고 측면에서 그 활용성과 효용성을 발굴하는 것이 목적이다.

4) 마이크로데이터 통합서비스. 국가데이터처(https://mdis.kostat.go.kr/index.do, 2025년 7월 31일 접속)
5) SGIS 통계지리정보서비스. 국가데이터처(https://sgis.kostat.go.kr/view/index, 2025년 7월 31일 접속)

[그림 1-1] AI 기반 인구이동 예측모형 개발을 위한 본 연구의 틀



출처: 저자 작성

제 3 절 연구의 차별성 및 기대효과

1. 연구의 차별성

AI 기반 인구이동 예측모형 개발 연구는 기존 인구이동 예측연구가 행정구역 단위 통계와 단순 시계열 추정에 의존해 온 한계를 극복하기 위해, 공간의 유연한 기능적 특성을 반영한 1km 격자 기반 AI 예측모형을 개발한다는 점에서 차별성을 가진다. 첫째, 행정경계 중심의 분석을 넘어 생활권·교통망·상권 등 실제 인구활동의 공간적 패턴을 반영함으로써 미시적 수준의 예측 정밀도를 높인다. 둘째, 통계 기반의 선형 추세 예측을 넘어 머신러닝(ML)·딥러닝(DL)·강화학습(RL) 및 하이브리드(Hybrd) AI 모형을 적용하여 비선형·고차원 데이터 구조를 효과적으로 학습한다. 셋째, 공간 네트워크 분석(중심성·접근성 지표)과 시계열 AI 분석을 융합하여 인구이동의 구조적 요인과 시계열 변화를 동시에 고려한다. 넷째, 정책 시나리오별 모의실험이 가능한 정책지원형 예측 플랫폼으로 구현하여 단순 예측을 넘어 의사결정 지원도구로 확장한다.

2. 연구의 기대효과

본 연구를 통해 다음과 같은 기대효과를 도출할 수 있다.

먼저, 국가 및 지방자치단체는 단·중기 인구이동 변화를 고해상도로 예측하여 지방소멸 대응 전략, 지역균형발전 지원, 인프라 파급력 확인 등 전략적 정책 수립의 근거자료로 활용할 수 있다.

둘째, 마이크로 공간 단위의 AI 예측 결과를 통해 소지역 단위 및 생활권 중심의 교통·주택·산업 민감도를 파악하여 자원 배분의 효율성과 정주여건 개선 판단 자료로 활용할 수 있다.

셋째, 시나리오별 정책 영향 분석을 통해 출산·이주·산업정책 등 다양한 분야의 정책 민감도 평가 및 사전 영향 분석 등을 통한 시뮬레이션 기반의 선제적 분석이 가능할 것으로 기대한다.

넷째, 데이터 수집, 전처리, 변수 설계, 모형 개발까지 일원화된 본 연구의 AI 기반 인구이동 예측 프레임워크를 바탕으로 향후 택지개발, 주택공급, 생활 인프라 지원 등으로 확장할 수 있는 선험적 활용이 가능하다.

다섯째, 그동안 보수적으로 다뤄온 인구이동 예측 분야에 시공간 데이터를 융복합하고 AI의 새로운 접근방법을 접목한 본 연구의 프레임워크는 학문적 기여가 높아 정책적 측면뿐만 아니라 관련 연구의 국내외 학술적 위상 제고에도 기여할 수 있을 것으로 기대한다.

제2장
AI 기반 인구이동 예측모형 개발의 필요성

제 1 절 도시화, 산업구조 변화와 인구이동의 현실

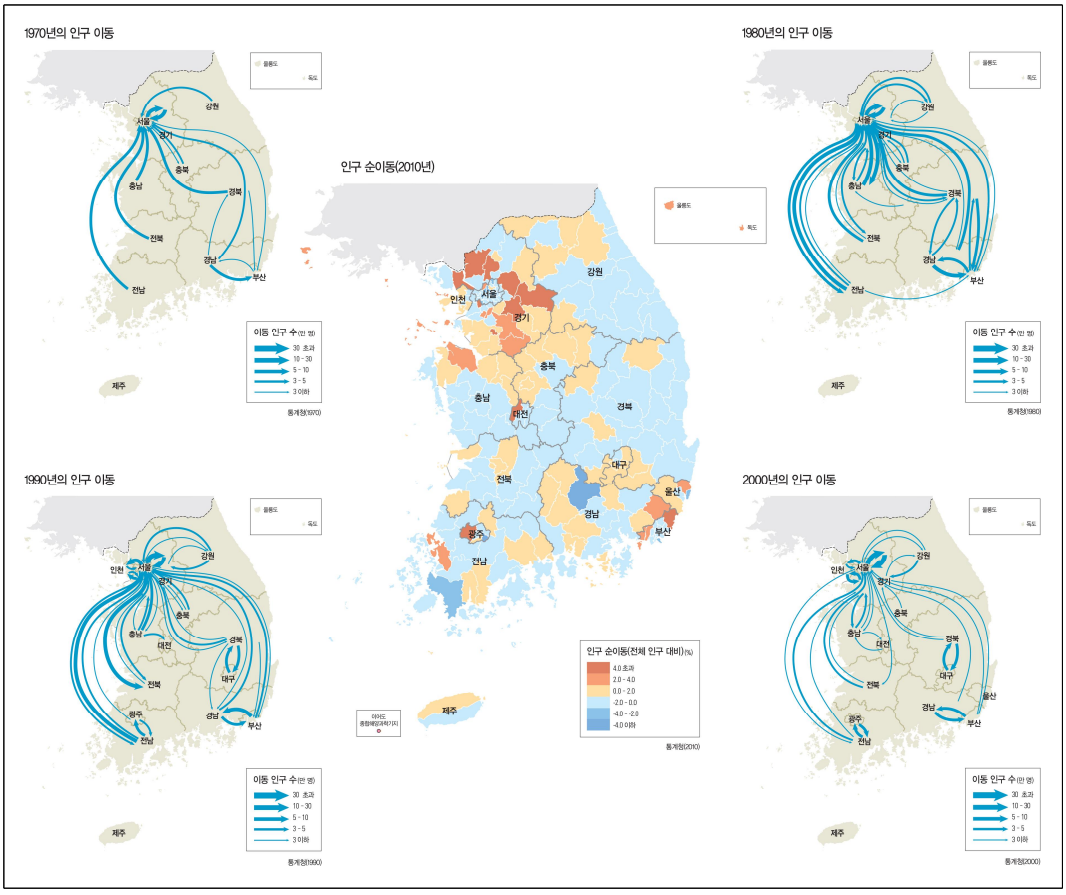
한국의 인구이동 양상은 도시화의 가속화와 산업구조의 변화에 따라 지속적으로 재편되어 왔다. 급격한 산업화·도시화가 진행되었던 1960~1980년대에는 농촌에서 대도시로의 이동이 주를 이루었으며, 특히 수도권으로의 집중 현상이 두드러졌다. 제조업 중심의 산업구조가 대규모 고용을 창출하면서 서울·인천·경기권의 순유입 인구는 꾸준히 증가하였다. 이후 1990년대에 접어들면서 서비스 산업의 비중이 확대되고, 대도시권 내부 재배치, 신도시 개발, 광역교통망 확충에 따른 서울 및 경기도 외곽 이전 등 새로운 형태의 이동 패턴이 나타났다(서창원·양진홍, 1998; 국가데이터처 「국내인구이동통계」).

2000년대부터는 수도권 과밀화와 부동산 가격 상승, 교통 인프라 확충이 맞물리면서 인구이동의 중심이 '대도시 집중'에서 '대도시권 재편'으로 전환되었다. 수도권 내부에서는 중심부에서 외곽 및 인근 도시로의 이동이 늘어났고, 지방의 대도시권에서도 비슷한 분산 경향이 관찰되었다. 반면, 비수도권 농촌 지역과 산업 기반이 취약한 중소도시는 청년층 유출과 고령화 심화로 인한 인구 감소가 장기적으로 고착되는 양상을 보였다(김대용, 2019; 정민수 외, 2023; 임태경, 2023).

산업구조 변화 측면에서는 제조업에서 첨단 서비스·지식 기반 산업으로의 전환이 빨라지면서 고학력 인력과 청년층이 산업·교육·문화 등의 인프라가 밀집된 지역에 집중되는 현상이 두드러졌다. 이에 비해 전통 제조업이나 1차 산업 의존도가 높은 지역은 일자리 축소와 청년 인구 이탈로 인해 인구 구조 불균형이 심화되었다. 이러한 변화는 4차 산업혁명, 디지털 전환, 원격근무 확산 등 새로운 경제·사회 여건에 따라 향후 인구이동 패턴에도 추가적인 변화를 불러올 것으로 예상된다(정민수 외, 2023; 임태경, 2023).

특히 최근 10여 년간의 인구이동은 절대 이동량 측면에서 감소 추세를 보이나, 이동의 방향성 측면에서는 수도권과 일부 혁신도시·광역도시로의 쏠림이 더욱 강화되는 지역 간 비대칭성이 뚜렷하게 나타나고 있다(국가데이터처 「국내인구이동」; 민보경, 2023). 이러한 비대칭적 인구이동은 지방소멸 위험을 가속화하고, 국가 차원의 균형발전 정책 수립과 실행을 위한 정밀한 예측의 필요성을 한층 높이고 있다.

[그림 2-1] 연도별 주요 도시 인구이동 시각화 예시



출처: 대한민국 국가지도집 1권, 인구이동

제 2 절 기존 인구이동 예측모형의 한계

기존의 인구이동 예측모형은 국가데이터처 등 공공기관에서 제공하는 장래인구 추계와 지방자치단체별 자체 추계가 주를 이루며, 주로 코호트요인법(Cohort-Component Method)과 같은 전통적인 통계기법에 기반한다. 이러한 접근은 비교적 단순하고 계산이 용이하다는 장점이 있으나 급변하는 사회·경제 환경과 비선형적 요인을 충분히 반영하지 못하는 구조적 한계를 지니고 있다. 이러한 구조적 한계는 추가적으로 다양한 어려움과 직면하게 된다.

첫째, 시나리오 가정에 대한 높은 의존성이 문제로 지적된다. 장래 인구 추계는 특정한 미래 조건을 전제로 산출되기 때문에, 출산율·사망률·이동률 등 주요 지표가 가정과 다르게 변화하는 경우 예측 결과의 불확실성이 크게 확대된다. 특히 시도 단위 추계는 총인구 추계에 비해 시나리오의 다양성이 제한적이며, 빈번한 지역 간 이동과 인구구조 변화 등 복합적인 요인을 동시에 고려해야 하는 복잡성이 존재하기 때문에 이해관계에 따른 추곗값 조정 위험이 수반될 수 있다.

둘째, 계획인구의 과다 산정에 따른 왜곡 가능성이 있다. 일부 지방자치단체는 인구 유입 증가를 전제로 도시계획을 수립하는 경향이 있으며, 이 과정에서 실제보다 높은 계획인구를 설정하는 사례가 발생한다. 이러한 과대 추계는 재정 부담을 키우고, 장기적으로 인프라 과잉 공급과 재정 건전성 악화를 초래할 수 있는 위험성을 내포하고 있다(감사원, 2021; 김명한, 2020).

셋째, 변화 추세 반영의 한계가 존재한다. 기존 추계 방식은 인구 이동성이 점차 감소한다는 전제를 기반으로 하여, 절대 이동량이 줄어드는 상황에서도 지역 간 인구 불균형이 심화하는 상황이 벌어질 수 있어, 최근 시계열 경향을 충분히 반영하지 못한다는 한계가 있다(이보경 외, 2022; 조대헌·이상일, 2022). 실제로 2019년 국가데이터처가 '장래인구 특별 추계[6)·7)]'를 별도로 발표한 사례에서도 확인할 수 있듯이, 기존의 주기적 추계 방식이 급격한 인구구조 변화를 적시에 포착하기는 어렵다.

넷째, 공간적 특성 반영의 부족이 한계로 지적된다. 현행 코호트요인법은 사망률·출생률·순이동률 등 인구학적 요인을 고려하더라도 공간구조, 토지이용 변화, 교통망 확충과 같은 입지·환경 요인은 적절하게 반영하지 못한다(이태직, 2010; 이보경 외, 2022; 박승용, 2022). 결과적으로

---

6) 국가데이터처, 「2019년 장래인구 특별추계를 반영한 내·외국인 인구전망: 2017~2040년」 (https://www.kostat.go.kr/board.es?mid=a10301020100&bid=207&tag=&act=view&list_no=385624&ref_bid=, 2025년 11월 27일 접속)

7) 국가데이터처 보도자료. 「장래인구특별추계: 2017~2067년」 (https://www.kostat.go.kr/synap/skin/doc.html?fn=e34f89db4c82501b96c8212e2579544fb0e2ce6a81a61c449051c04cc04d7c11&rs=/synap/preview/board/207/, 2025년 11월 27일 접속)

총량 예측 수준에 그치며, 생활권·산업권·상권 등 세부 공간 단위에서의 인구 재배치 전망 등에는 원론적인 한계가 있다.

마지막으로, 비선형·다변량 데이터 처리의 제약이 존재한다. 인구이동은 경제·사회·환경·정책 등 다양한 요인이 복합적으로 작용해 발생하는 결과이다(이보경 외, 2022; 박승용, 2022). 이러한 상호작용은 데이터의 관점에서는 비선형적 특성을 지닌다. 그러나 전통적 통계모형은 이러한 복잡한 데이터 구조를 충분히 학습하거나 예측하기 어렵다. 특히 대규모 시·공간 데이터의 통합, 네트워크 특성, 시나리오별 변동성 등을 종합적으로 반영하기 위해서는 기존의 접근법을 넘어서는 새로운 분석 시도가 필요하다.

제 3 절 공간 기반 분석과 인공지능 기술의 필요성

인구이동은 단순히 행정구역 경계 내에서 발생하는 현상에 그치지 않고, 생활권·교통망·상권 등 다양한 공간의 범위 및 기능적 특성과 긴밀히 맞물려 나타난 결과이다. 그러나 기존 행정구역 단위 통계는 이러한 공간 속성을 충분히 반영하지 못해, 실제 거주 및 활동 패턴을 설명하거나 예측하는 데 제약이 있다. 이에 따라 행정구역을 넘어 실질적인 생활·경제 활동권을 반영할 수 있는 미시적 단위의 격자 기반 공간분석 필요성이 커지고 있다(이보경 외, 2022; 박승용, 2022).

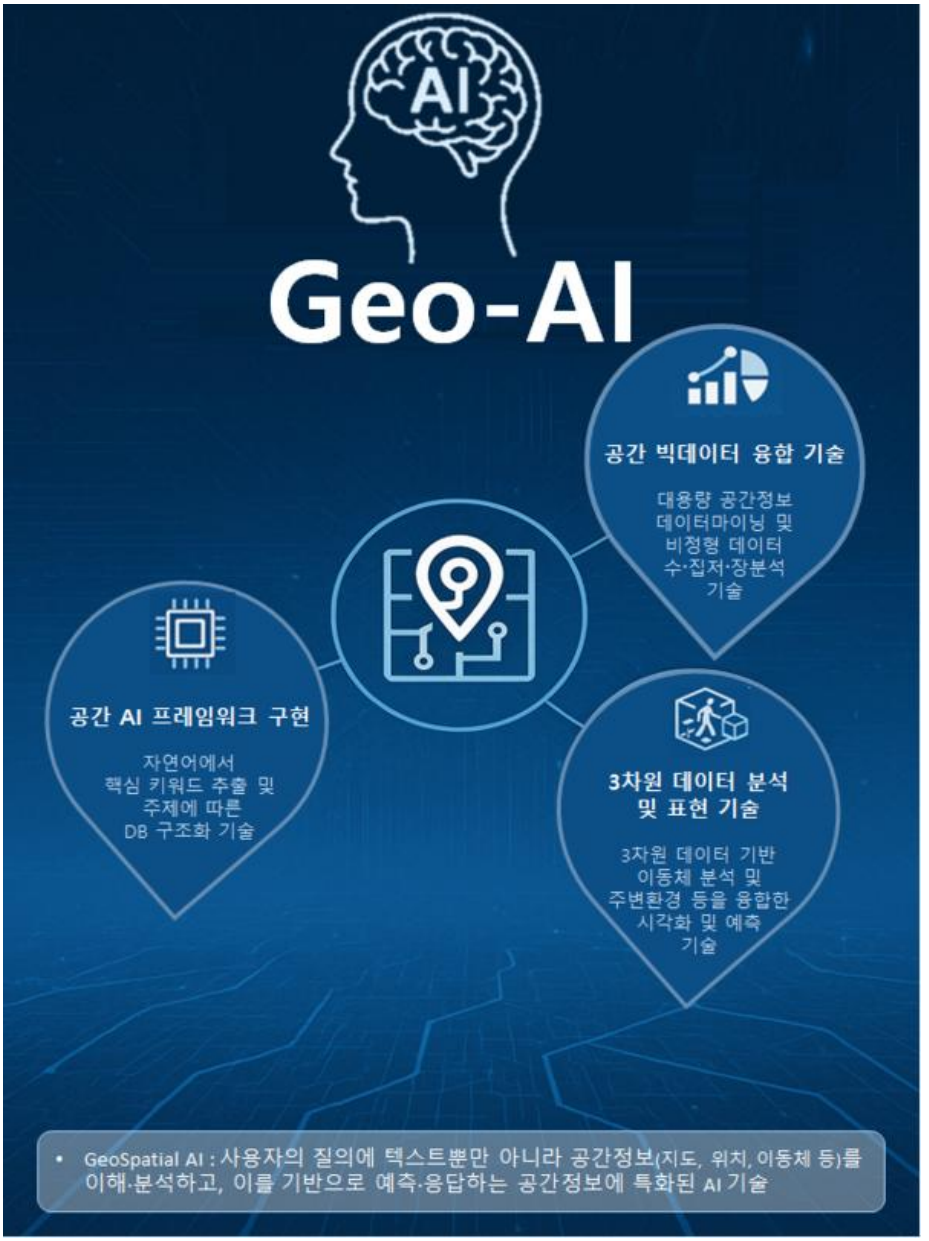
다양한 정보가 적절하게 고려된 격자기반 접근법은 단순 인구 규모뿐 아니라 연령별·성별 구조, 가구 특성, 교통 접근성, 상권 밀도 등 다양한 입지 요인을 동시에 고려할 수 있는 강점이 있다. 이는 단순 총량적 예측 수준을 넘어 인구의 세부 공간 분포를 이해하고, 재배치를 위한 근거자료를 도출하고, 도시계획·교통정책·상권 활성화 등과의 이해도 및 연계성을 한층 강화할 수 있다.

다양한 시공간 인과관계 파악과 대규모 비선형 데이터 구조를 활용한 정밀 분석을 구현하기 위해서는 전통 통계기법의 한계를 보완할 수 있는 인공지능(AI) 기반분석 기법의 시도가 필수적이다. AI는 대규모 시·공간 데이터와 다차원 변수를 효과적으로 처리할 수 있으며, 변수 간의 비선형적 관계를 학습해 복잡한 인구이동 패턴을 미세하게 모델링할 수 있는 강점이 있다. 특히 머신러닝(ML)과 딥러닝(DL) 등 인공지능 기법은 방대한 데이터에서 그 예측력을 극대화하고, 기존 분석에서 드러나지 않던 요인 간 민감도와 직관적 상관관계를 발굴하는 데 강점을 지닌다.

또한, AI 기반 분석은 민감도 분석과 시나리오 기반의 시뮬레이션을 가능하게 한다. 예컨대 산업단지 조성, 교통 인프라 확충, 주택 공급 확대, 출산율 변화와 같은 정책·환경 변화를 가정했을 때의 인구이동 변화를 단기간 내 모의 실험할 수 있고, 이러한 정책·환경 변화의 복합적인 시나리오 구성도 가능하다. 이는 전통적인 통계 접근이 제공하지 못했던 정책 의사결정 지원 역량을 강화하며, 정책 입안자나 시민 모두에게 정책적 공감대 제공 및 전략적 가치를 제공한다.

결과적으로, 공간 기반 분석과 AI 기술의 결합은 행정구역 중심의 단편적 총량 예측을 넘어, 생활권 단위의 정밀하고 동적인 인구이동 전망을 가능하게 한다. 더 나아가 이는 지역 맞춤형 정책 설계, 효율적 자원 배분, 그리고 지방소멸 대응 등 국가적 현안 해결에 효율적으로 기여할 수 있을 것으로 기대된다.

[그림 2-2] 공간정보를 활용한 인공지능(Geo-AI) 소개 예시



출처: 《국토교통부 보도자료》, "일상에서 만나는 공간정보 AI, 지도로 답하다" (2025년 7월 29일자)

제 4 절 정책 수요와 시의성

우리 사회는 최근 급속한 인구구조 변화와 지역 불균형 심화라는 이중 과제에 직면해 있다. 저출산과 고령화, 청년층의 수도권 집중, 그리고 지방소멸 위기는 국가의 지속가능성을 위협하는 핵심 변수로 부상하고 있다. 이러한 이슈는 비단 국내에 한정된 내용은 아닐 것이다. 때문에, 이러한 변화에 효과적으로 대응하기 위해서는 중장기적으로 인구이동을 정밀하게 예측하고, 이를 실시간에 가깝게 모니터링할 수 있는 체계 구축이 필수적이며, 이는 중앙정부와 지방자치단체의 전략적 의사결정과 직결된다고 할 수 있다.

정책 수요 관점에서 보면, 지역 맞춤형 정책 설계를 위한 생활권이나 경제권 단위에서 인구변화를 예측하고 분석할 수 있는 도구가 필요하다. 이를 통해 산업입지 조정, 정주여건 개선(예를 들어, 교통 인프라 확충, 주택 공급, 교육·의료 시설 배치) 등 다양한 정책 영역에서 기민한 대응이 가능해진다. 두 번째로, 지방소멸에 전략적인 대응을 위해서는 과거로부터 확인할 수 있는 인구유출 요인을 사전에 진단하고, 청년층 유입이나 고령층 정주 지원과 같은 '핀셋형' 지원 정책을 설계할 필요가 있다. 세 번째로, 국가 균형발전 전략 수립에는 권역별·생활권별 인구 흐름을 기반으로 지역 간 불균형 완화를 위한 과학적 근거가 요구되는데, 현재 국가데이터처의 인구추계에만 의존하는 방식은 시나리오 및 정책 개입 등의 영향을 견주어 보기에는 무리가 따른다.

현재 총인구이동 규모는 줄어드는 반면 인구구조 쏠림현상으로 인한 지역 간 비대칭성은 심화되고 있다. 일부 지역에서는 고령화와 청년층 유출이 동시에 진행되는 반면, 어떤 지역은 산업 개발이나 관광자원 발굴, 인프라 확충 등으로 인구 유입의 양극화가 현상이 뚜렷하다. 더불어 기후위기, 재난, 산업구조 변화 등 돌발적 환경 요인이 인구이동 패턴에 즉각적인 영향을 미치고 있어, 기존의 장기적이고 정적인 예측 방식을 고수하는 방법만으로는 면밀한 대응이 어려워지고 있다.

이러한 맥락에서 AI 기반 인구이동 예측모형은 데이터 중심의 동적(dynamic) 예측과 시나리오 기반의 분석이 가능하여, 단기 및 장기 등 전주기 전략 수립 모두에 수렴될 수 있는 핵심 수단이 될 수 있을 것으로 기대된다. 특히, 격자 기반 공간분석과 결합된 AI 모형은 정책 효과의 시공간적 확산 범위를 탄력적이고 정량적으로 산출할 수 있어, 정책의 사전·사후 영향 평가를 파악하고 자원 배분 효율성을 극대화하는 데 기여할 수 있을 것이다.

현재의 인구 및 지역 문제는 단순 통계분석을 넘어, 실시간성·정밀성·정책 연계성을 고르게 갖춘 고도화된 종합적 예측 체계가 필요하며, 이는 AI 기반 인구이동 예측모형 개발의 필요성을 뒷받침한다.

제 5 절 소결

본 장에서는 AI 기반 인구이동 예측모형의 개발 필요성을 다양한 방면으로 검토하였다. 먼저, 도시화의 가속화와 산업구조의 변동에 따라 인구이동 양상이 과거보다 복합적이고 비선형적인 형태로 전개되고 있음을 확인하였다. 이러한 변화는 수도권 집중 심화, 지방소멸 위험 확대, 지역 간 불균형 심화 등 국가 차원의 주요 현안과 직결되며, 이에 대응하기 위한 과학적 예측과 선제적 정책 마련의 필요성이 강조된다.

기존 인구이동 예측 체계의 분석을 통해 행정구역 중심의 집계 방식, 고정된 시나리오 가정, 공간 특성 반영 방식의 부재, 비선형 다변량 데이터 처리의 한계 등이 정책 활용성 제약 및 활용 정확도 저하 등의 핵심 요인을 확인하였다. 이러한 구조적 제약은 장기 인구 전망뿐 아니라 인구이동 예측, 또 이와 관련된 각종 단기 및 중기 정책 대응에도 한계를 초래한다.

공간분석과 인공지능의 결합 필요성을 검토한 결과, AI가 비선형 데이터 처리, 다차원 변수 분석, 시뮬레이션 수행에 있어 두드러진 강점이 있음을 확인하였다. 특히 격자 기반 공간분석과 연계된 AI 모형은 생활권, 교통망, 상권 등 다양한 기능적 공간 요인을 반영하여 미시 단위의 인구이동을 정밀하게 예측할 수 있으며, 정책 효과의 공간적 파급범위를 사전에 파악하고 시뮬레이션 분석에도 유용할 것으로 기대한다.

마지막으로, 인구 총량 감소와 지역 간 비대칭성이 동시에 심화되는 현시점에서 실시간성·정밀성·정책 연계성을 갖춘 예측 체계의 도입이 정책 수요와 시의성 측면에서도 필수적임을 확인하였다. 이는 지방소멸 대응, 국가 균형발전, 인프라 투자, 산업 입지 조정 등 다양한 분야에서 공통적으로 요구되는 핵심 역량이다.

종합하면, 인구이동 예측의 정확성과 정책 활용성을 동시에 달성하기 위해서는 기존 통계 기반의 한계를 극복하고, 시공간 분석과 고도화된 AI 기술을 접목한 격자 기반 AI 인구이동 예측모형을 개발하는 것이 필요하다. 이러한 모형은 변화하는 인구·공간 환경 속에서 국가와 지방자치단체가 장단기 정책 목표를 능동적으로 달성할 수 있는 핵심 전략 도구가 될 것으로 기대한다.

제3장
선행연구검토

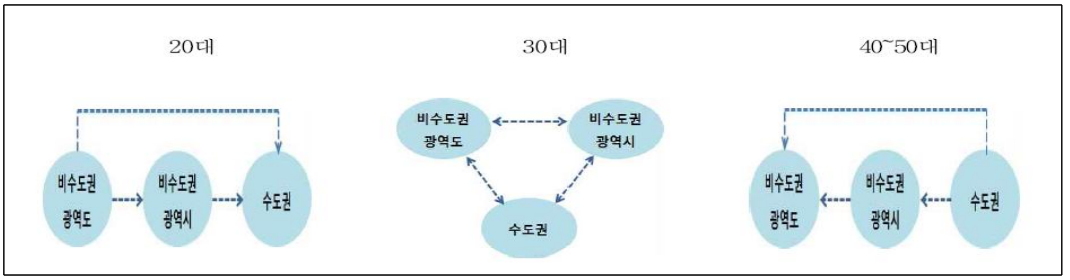
본 장에서는 인구이동의 요인에 관한 연구를 시작으로, 도시구조 변화와 인구 분포에 관한 연구, 그리고 인구 예측 방법론과 AI 기반 접근 동향에 관한 선행연구를 체계적으로 검토한다. 인구 변화는 단기간에 급격히 발생하기보다는 장기간에 걸쳐 점진적으로 축적·진행되는 특성을 지니므로, 관련 논의와 연구 성과는 장기적인 시계에서 이해할 필요가 있다. 이에 따라 본 장에서는 상대적으로 오래된 문헌에서부터 최근 연구에 이르기까지의 학문적 흐름을 연대기적으로 살펴보고, 이를 통해 인구이동 예측 연구의 역사적 맥락과 기술적 진화를 함께 조망하고자 한다.

제 1 절 인구이동 요인에 관한 선행연구 검토

도시 및 지역 간 인구이동은 한국 사회의 경제적·사회적 역동성에 복합적으로 영향을 미쳐왔으며, 그 요인은 시대적 흐름에 따라 변화해 왔다. 1990년대 이전 산업화와 도시화 시기에 농촌에서 도시로, 비수도권에서 수도권으로의 이동이 주를 이루었다고 분석한 사례가 있는데. 이는 주로 양질의 일자리 확보와 임금 격차 해소를 위한 경제적 기회 추구가 핵심 동인이었다고 바라보는 관점이다(이찬영, 2018; 오은열·문채, 2016). 또한, 제조업 비중이 높은 지역에서 인구 유입이 활발하게 나타났으며, 대규모 개발사업과 중심지 서비스 확충이 인구를 흡수하는 주요 요인으로 작용했다고 연구된 바 있다(이정록, 2017; 남형권, 2020).

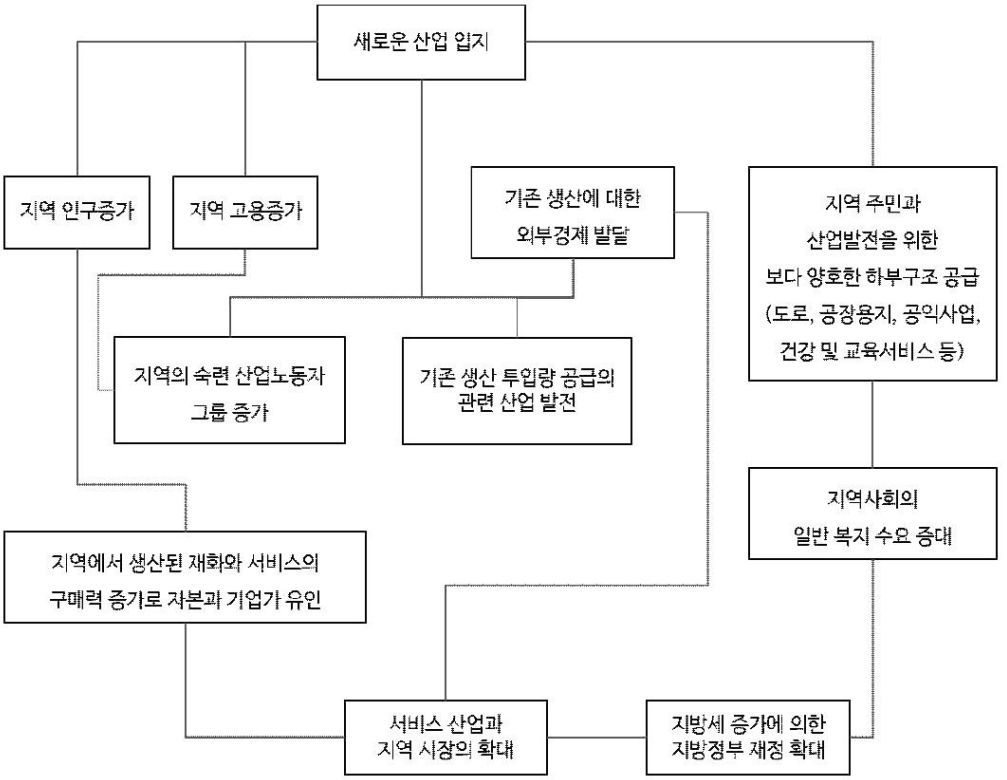
2000년대 이후의 인구이동은 총량 감소 경향 속에서 지역 간 불균형이 심화되는 양상이 나타났다. 이상림·최인선(2023)은 수도권 집중이 일시적으로 완화되는 듯했으나, 2010년대 중반 이후 청년층을 중심으로 다시 강화되었다고 지적하였다. 서울의 높은 주택가격은 경기·인천 지역으로의 전출을 유발하였으며, 반면 비수도권은 일자리 부족과 정주 여건 취약성으로 인해 청년층 유출과 고령화가 가속화되었다고 하였다(박승규, 2020; 이상림·최인선, 2023).

[그림 3-1] 연령대별 인구이동 유형



주: 화살표 방향은 인구순이동 흐름을 나타냄
출처: 이찬영(2018), p. 717

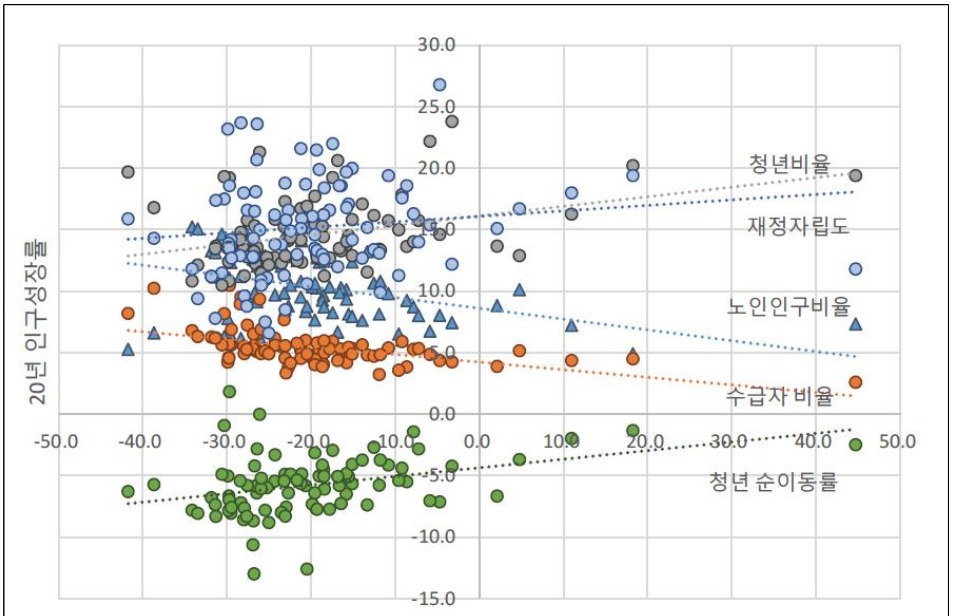
[그림 3-2] 순환적·누적적 인과과정 모형도



주: 새로운 산업의 입지가 인구를 포함한 지역에 미치는 순환적·누적적 효과를 의미(남형권, 2020. p. 30)
출처: Keeble(1967); 남형권(2020), p. 30에서 재인용

[그림 3-3] 인구성장률과 지역의 인구학적 사회적 요인들 간의 상관관계

(단위: %)



출처: 통계청(KOSIS), 주민등록연앙인구통계(2002년, 2022년)
통계청(KOSIS), 인구이동통계(검색일자: 2023. 10. 11.)
통계청(KOSIS), e-지방지표, 재정자립도 통계(검색일자: 2023. 10. 11.)
보건복지부(2023) 2022년 국민기초생활 수급자 현황

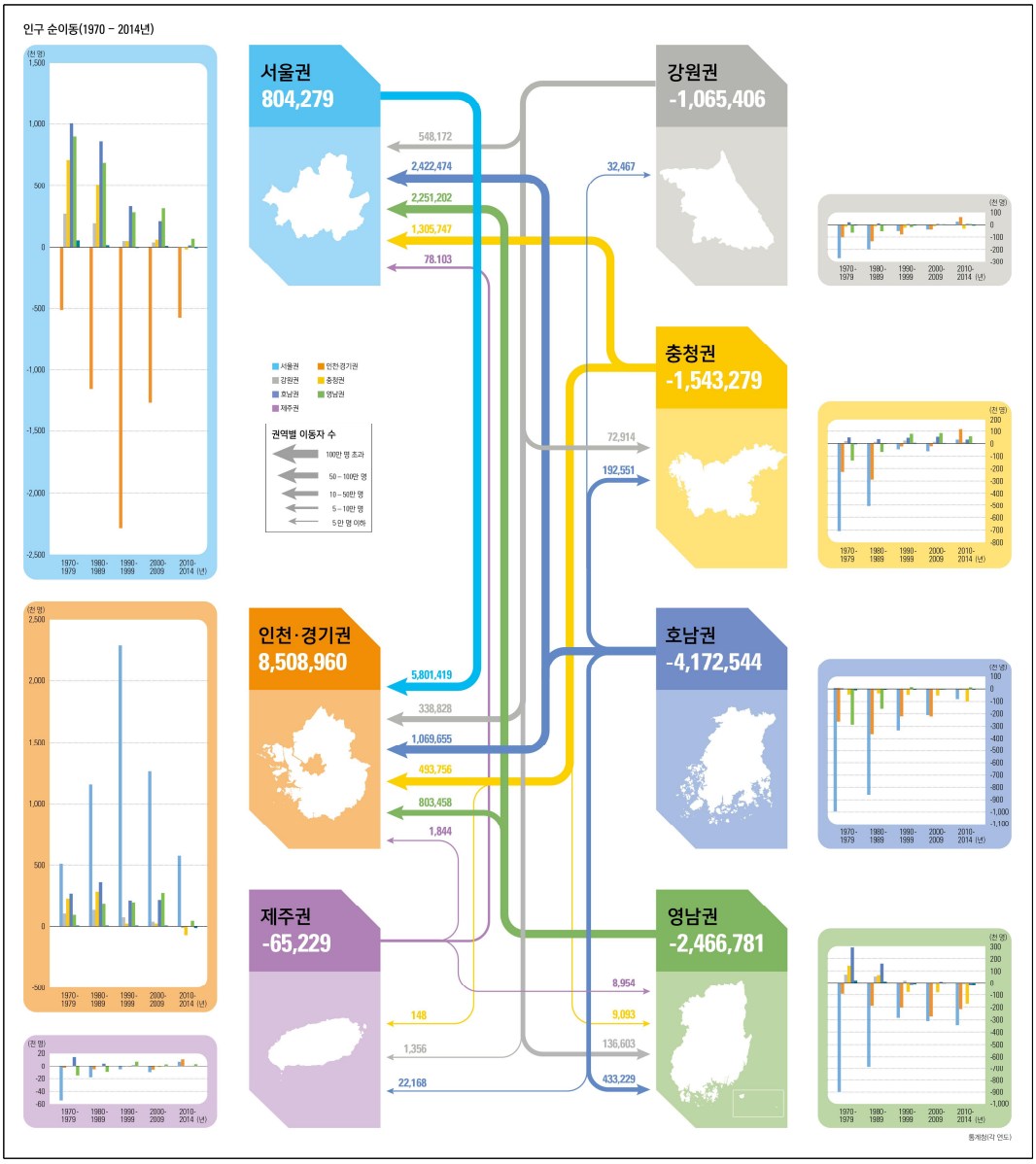
출처: 이상림·최인선(2023), p. 88 재인용

인구 특성별 이동 요인에 관한 연구도 다수 진행되었다. 청년층의 경우, 일자리와 교육 기회, 주거비, 삶의 질 요인이 주요 결정요인으로 작용하는데, 특히 수도권 대학 진학과 취업 기회가 강한 유입 요인이라는 점이 지적되었다(이찬영, 2018; 최대식 외, 2024). 남형권(2020)은 고학력자가 산업구조 변화에 민감하게 반응하여 원거리 이동 가능성이 크다고 분석한 바 있고, 이상림·최인선(2023)은 성별 측면에서 여성의 수도권 순유입 비중이 높게 나타나는 현상을 '수도권 이주의 여성화'로 정의하며, 지방 제조업 중심 산업구조의 한계를 주요 원인으로 지적하였다. 중장년층과 고령층의 경우, 퇴직 후 귀농·귀촌, 의료 인프라 접근성, 주거비 부담 완화 등이 주요 이동 요인으로 제시되었으며, 고령층 경제활동 참여가 일부 지역에서 긍정적 효과를 가져온다는 분석도 이루어졌다(이찬영, 2018).

그 밖에도, 산업구조 변화, 광역교통망 확충, 주택 공급, 생활 인프라 등은 인구이동의 중요한 결정요인으로 연구되었는데, 남형권(2020)은 광역교통망과 대규모 주거단지 조성이 인구 유입 및 분산에 긍정적인 효과를 준다고 바라보았으며, 이찬영(2018)과 최대식 외(2024)는 의료·문

화·교육 인프라가 인구 유출 억제와 유입 증대에 긍정적인 효과가 있음을 강조하였다. 이러한 요인들은 상호 복합적으로 작용하며, 인구이동의 패턴과 규모를 결정짓는 중요한 기반으로 자리매김해 왔다.

[그림 3-4] 주요 권역별 인구 유출입(순이동: 1970~2014)



출처: 대한민국 국가지도집

제 2 절 도시구조 변화와 인구 분포 관련 연구

도시 및 지역의 인구 분포와 이동 패턴에 관한 연구는 시대별 사회·경제적 변화와 인구이동 양상에 따라 그 초점과 분석 방법이 다양하게 다루어져 왔다.

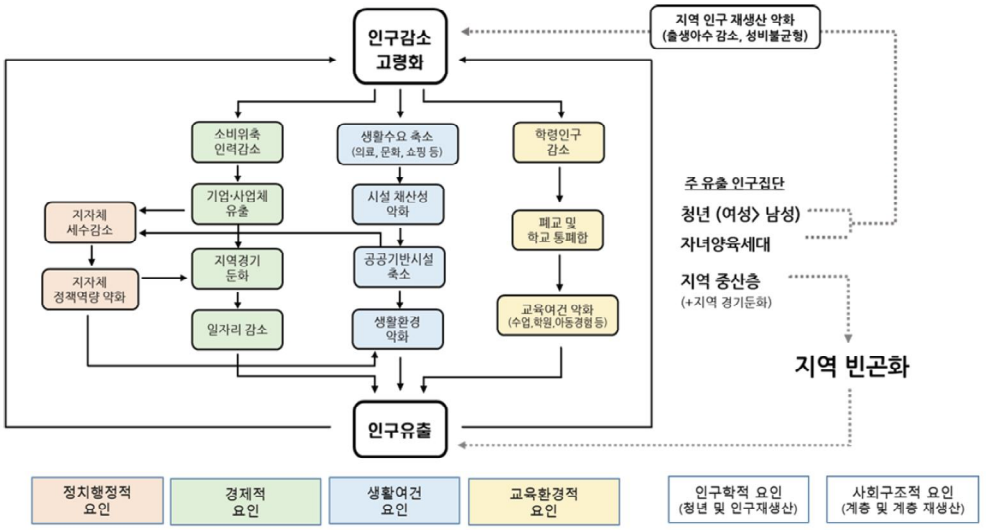
1990년대 이전 산업화 및 도시화 시기의 연구들은 주로 농촌에서 도시로, 비수도권에서 수도권으로의 집중 현상을 분석하였다(이찬영, 2018). 이러한 이동은 양질의 일자리와 임금 격차 해소를 위한 경제적 기회 추구가 핵심 동인이었으며(오은열·문채, 2016; 이태직, 2010), 제조업 성장과 대규모 개발사업이 인구 유입의 주요 요인으로 작용하였다(이정록, 2017; 남형권, 2020). 예를 들어, 광양제철소 건설과 같은 대규모 산업단지 조성이 순천시로의 노동력 유입을 촉진하였으며, 중심지 서비스 확충이 인근 지역 인구를 흡수하는 역할을 했다고 바라보았다(남형권, 2020). 이 시기의 연구들은 Thünen의 농업 입지론, Weber의 공업 입지론, Christaller의 중심지 이론 등 전통적 입지이론을 적용하여 교통·통신 인프라 수준과 수송비 구조가 인구 분포에 미치는 영향을 설명하였다(남형권, 2020).

2000년대 이후에는 인구이동 총량이 감소하는 가운데 지역 간 불균형이 심화하는 양상이 두드러졌는데, 2000년대 중반 수도권 유입세가 약화되었으나, 2010년대 중반 이후 청년층을 중심으로 수도권 집중이 재강화하였고(이상림·최인선, 2023), 서울의 높은 주택가격이 경기·인천 지역으로의 전출을 유발하는 한편, 경기 지역의 고용기회 증가는 인구 유입을 촉진하는 요인으로 작용한다고 지적하였다(최대식 외, 2024). 반면 비수도권 지역은 일자리 부족과 정주 여건 취약으로 청년층 유출과 고령화가 동시에 심화하고 있음이 연구되었다(이상림·최인선, 2023).

연령·성별 특성별 연구에서는 청년층 이동이 일자리, 교육 기회, 주거비 부담, 삶의 질 등 복합요인의 영향을 받는 것으로 나타났다(이찬영, 2018; 최대식 외, 2024). 특히 이상림·최인선(2023)은 '수도권 이주의 여성화' 현상을 지적하며, 지방 제조업 중심 산업구조가 여성 친화적 일자리 부족을 초래한다고 지적한 바 있다. 중장년층과 고령층의 경우, 퇴직 후 귀농·귀촌, 의료 인프라 접근성, 주거비 완화 등이 주요 요인으로 나타났으며, 일부 지역에서는 고령층 경제활동 참여가 지역 경제에 긍정적인 영향을 미친다고 강조하였다(이찬영, 2018; 남형권, 2020).

산업구조 변화의 영향에 대해서는, 고차 제조업으로의 전환이 인구 유입을 촉진하거나 유출을 억제하는 긍정적 효과를 가진다고 해석한 반면, 서비스업 중심 구조로의 전환은 인구 유입 억제 및 유출 촉진 경향을 보인다고 지적하였다(남형권, 2020). 광역 교통망 확충이 수도권 인구 분산과 산업 균형 개발에 양의 영향을 주며, 신규 주택 공급과 생활 인프라 확충이 인구 유입을 유도한다는 연구 결과도 제시되었다(이찬영, 2018; 남형권, 2020; 최대식 외, 2024).

[그림 3-5] 인구이동의 인구학적, 사회학적 파급효과(인구이동에 의한 종합적 악순환 구조)



출처: 이상림·최인선(2023), p. 91

연구 방법론 측면에서, 오은열·문채(2016)는 2000년대 이후 지역패널 분석을 활용하여 지역별 인구이동 현황과 연령대별 이동 결정요인을 정량적으로 분석하였다. 최대식 외(2024)는 2010년대 들어 인구이동의 기점과 종점을 군집화하여 공간적 의미를 해석하는 네트워크 분석을 시도하였고, 최근에는 이동찬·김기환(2021), 이보경 외(2022) 등의 연구에서 모바일 기지국 데이터, SGIS 등 공간 빅데이터를 활용하여 1km 격자 단위의 인구 추정·예측을 시도하고, 전통적 코호트 요인법(Cohort Component Method)[8]과 딥러닝(DL) 기반 인공지능 기법을 결합하여 행정구역 경계를 넘어선 정책 단위 설정 가능성을 모색하는 연구를 소개한 바 있다.

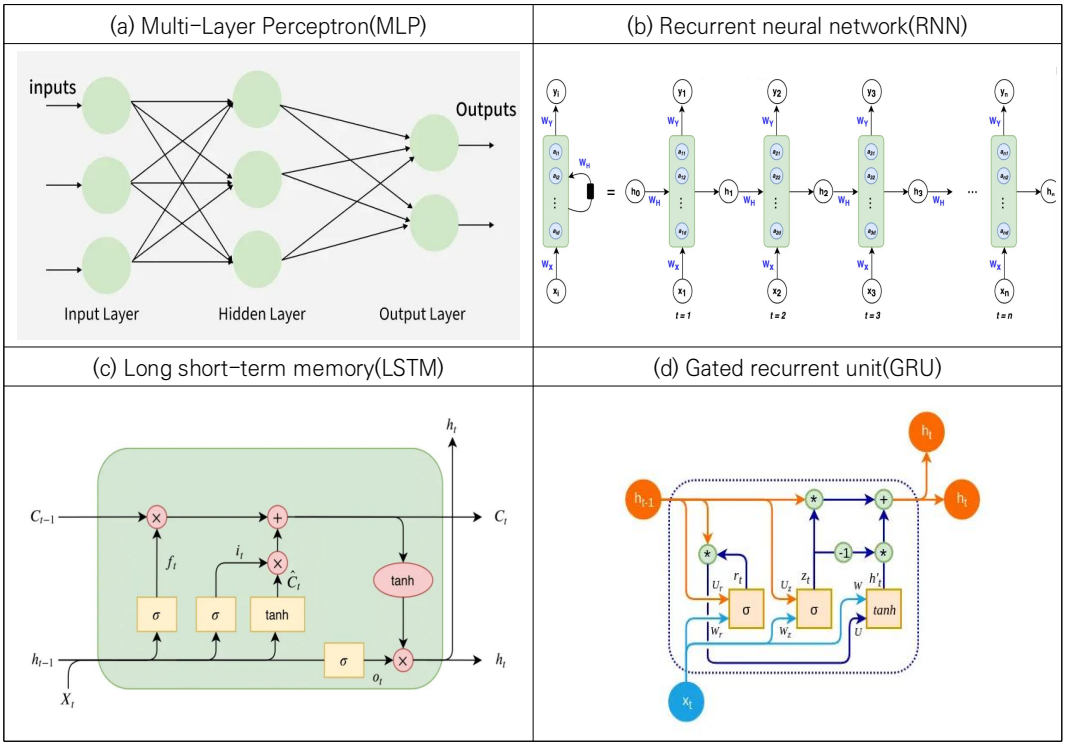
도시구조 변화와 인구 분포에 관한 연구는 시대별 사회·경제적 맥락과 데이터·분석 기법의 발전에 따라 지속적으로 확장되어 온 것으로 확인된다. 이러한 연대기별·주제별 연구 흐름은 전통적 인구 예측 방법론의 한계를 보완하며, 공간 기반 AI 인구이동 예측모형 개발의 실질적 설계 방향에 중요한 시사점을 제공한다.

8) "코호트요인법이란 인구변동요인별(출생, 사망, 인구이동) 미래 수준을 각각 예측한 후, 추계의 출발점이 되는 기준인구에 출생아 수와 순이동자 수는 더하고, 사망자 수는 빼는 인구균형방정식을 적용하여 다음 해 인구를 반복적으로 산출해 나가는 인구추계 방법"(국가데이터처, 통계조사 통계별 질문, https://kostat.go.kr/board.es?mid=a10502010000&bid=3207&tag=&act=view&list_no=161814&ref_bid=, 2025년 7월 31일 접속)

제 3 절 인구 예측 방법론 AI 기반 연구 동향

도시구조 변화와 인구 분포 관련 연구의 확장과 함께, 인공지능(AI)을 활용한 인구 예측 방법론은 최근 들어 간헐적으로 시도되고 있다. 이는 전통적인 통계적 추계 방식이 가진 구조적 한계를 보완하고, 복잡한 대규모의 공간 빅데이터를 활용하여 예측의 정확성과 설명력을 높이기 위한 노력을 일환으로 전개되고 있다. 본 절에서는 전통적 방법론의 한계 인식, AI·빅데이터 기반 연구의 등장, 그리고 하이브리드 모델로의 발전 과정을 연대기에 따라 개략적으로 살펴본다.

[그림 3-6] 선행연구에서 활용된 주요 AI 모형 도식화

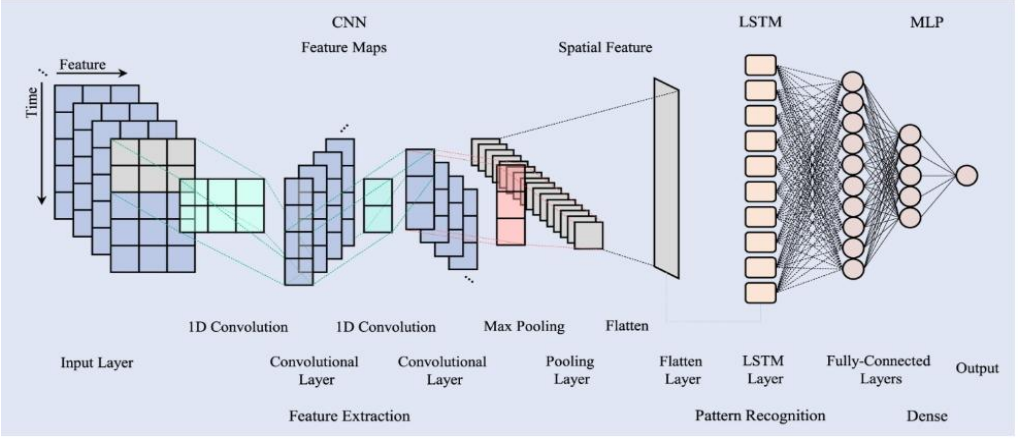


출처: (a) Multi-Layer Perceptron Learning in Tensorflow, https://www.geeksforgeeks.org/deep-learning/multi-layer-perceptron-learning-in-tensorflow/(2025년 11월 29일 접속)
(b) The Basics of Recurrent Neural Networks(RNNs), https://pub.towardsai.net/whirlwind-tour-of-rnns-a11effb7808f(2025년 11월 29일 접속)
(c) Long Short-Term Memory(LSTM), https://medium.com/@saba99/long-short-term-memory-lstm-fffc5eaebfdc(2025년 11월 29일 접속)
(d) Understanding Gated Recurrent Unit(GRU) in Deep Learning, https://medium.com/@anishnama20/understanding-gated-recurrent-unit-gru-in-deep-learning-2e54923f3e2(2025년 11월 29일 접속)

국가데이터처의 코호트요인법(Cohort Component Method)과 같은 전통적인 인구추계 기법이 인구 예측에서 주류를 이루는 사실은 부정할 수 없는 사실이다. 이 방법은 기준인구에 출생과 전입을 더하고, 사망과 전출을 제하는 인구 균형 방정식을 바탕으로 연속적인 연도별 인구를 산출하는 구조를 가진다(박승용, 2022). 그러나 미래 가정에 의존한다는 점에서 내재적인 불확실성이 크고, 시도 단위 이상의 거시적 예측에 치중하여 세밀한 공간 단위 인구 분포 변화를 설명하기 어렵다는 한계가 있다. 특히 소지역에 가까워질수록 예측의 정확도가 떨어지고, 토지이용 변화나 생활권 변화와 같은 공간적 맥락을 반영하지 못하는 약점이 지적되었다(박승용, 2022; 이보경 외, 2022).

최근에 이르러서는 기존 접근방식의 한계를 극복하기 위해 (공간) 빅데이터와 AI 기법을 접목하려는 시도가 간헐적으로 소개되었다. 초기 연구에서는 모바일 기지국 데이터, 토지 피복 정보, 건물 위치, 대중교통 승하차 인원 등 다양한 변수를 결합하여 격자 단위 인구를 추정하는 모형이 소개되었다(이동찬·김기환, 2021). 이 과정에서 푸아송 회귀와 랜덤 포레스트(Random Forest) 등 기계학습 기법이 활용되었다. 주로 현시점의 공간적 분포 추정에 초점을 두었으나 AI와 공간데이터 결합의 가능성을 입증하였다는 점에서 구분되는 특징이 있다. 이어 일부 연구에서는 다층 퍼셉트론(Multilayer Perceptron, MLP), 순환신경망(Recurrent neural network, RNN), 장단기 기억망(Long short-term memory, LSTM), 게이트 순환 유닛(Gated recurrent unit, GRU) 등 딥러닝 기반 시계열 모델을 적용하여 미래 인구변화를 직접 예측한 사례도 소개되었다(박승용, 2022). 이러한 모델들은 비선형 패턴과 복잡한 상호작용을 포착할 수 있다는 점에서 전통적 선형모델 대비 유리함을 보였으나, 충분히 긴 시계열 데이터 확보와 변수 확장의 어려움이 한계로 지적되었다.

[그림 3-7] CNN-LSTM 구조 도식화 예시

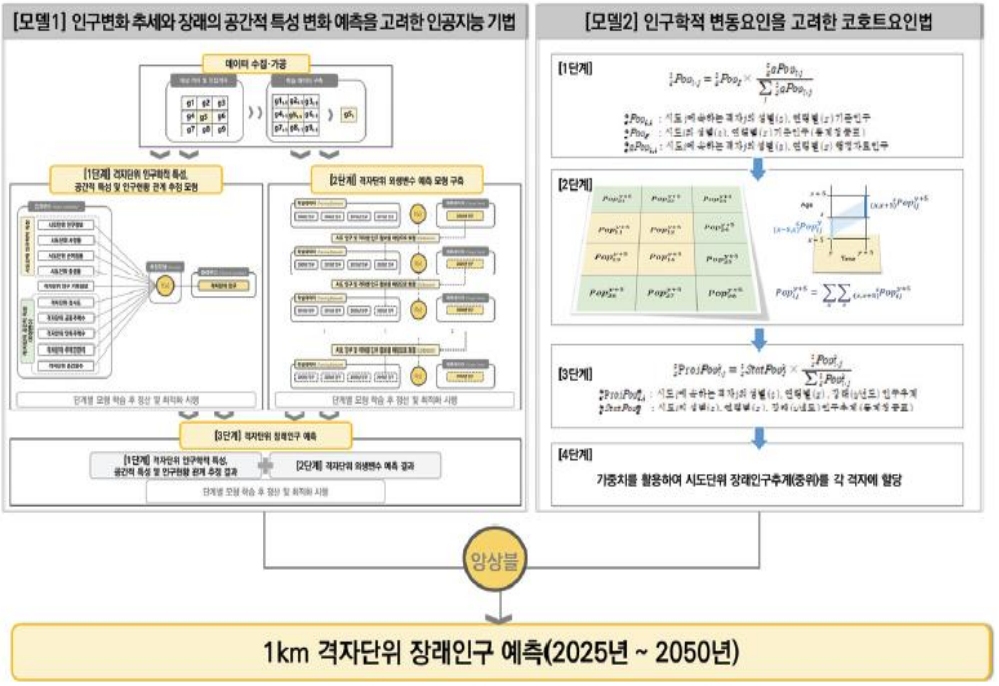


출처: Eggebrecht, P. and Eva Lutkebohmert(2023), p. 599

전통적 인구학 모델과 AI 기법의 장점을 결합한 하이브리드 접근도 소개되었다. 대표적으로 이보경 외(2022)에서 제시한 GPAI(Gridded Population Forecast with AI) 모델은 코호트요인법에 기반한 장래인구 추계와 건축물 수·주거면적·경사도 등 공간 특성을 학습한 AI 모델을 앙상블 방식으로 결합하여 1km 격자 단위의 장래 인구 예측모형을 대표한다(이보경 외, 2022). 이 시도에서 랜덤 포레스트로 인구와 공간 변수 간 관계를 학습하고, CNN-LSTM[9] 결합모형으로 공간 특성의 변화를 학습한 뒤, 두 결과를 통합하여 시도 단위 통계와 정합성을 제고하였다. 이러한 방식은 다양한 정책 단위에서 활용할 수 있는 공간 정합적 예측 결과를 제공하며, 행정 경계에 구애받지 않는 맞춤형 계획수립을 가능하게 했다는 데 주안점이 있다.

최근에는 데이터 융합, 격자 단위 공간 통계 생산, 토지이용 변화 예측, 기후변화와 같은 환경요인의 반영 등 모형의 고도화가 진행되고 있다. 특히 일본과 같이 인구감소 추세가 선행적으로 나타난 국가의 사례를 학습하여 한국의 미래 변화 패턴을 보완적으로 예측하거나 주택 공급·산업입지·교통망 확충 효과를 반영하는 모형 설계가 시도되고 있다.

[그림 3-8] AI를 활용한 1km 격자 단위 장래인구 예측(2025~2050년)



출처: 이보경 외(2022), p. 68

9) Convolutional neural network - Long short-term memory, CNN-LSTM

제 4 절 소 결

본 장에서는 인구이동의 결정 요인, 도시 구조 변화와 인구 분포, 그리고 AI를 활용한 인구 예측 기법에 관한 선행연구를 시간 흐름을 중심으로 검토하였다. 인구 예측 연구는 오랫동안 국가데이터처의 코호트요인법과 같은 전통적 통계기법을 중심으로 수행되어 왔으나, 이러한 방식은 세부 공간 단위에서의 예측이나 변화 요인의 다차원 분석 측면에서 제약이 있음을 확인할 수 있었다. 특히 행정구역 단위의 총량 예측 중심 구조는 토지이용 변화, 산업입지, 교통망 확충과 같은 공간적·기능적 요소를 충분히 반영하지 못하는 한계를 드러낸다.

최근, 빅데이터와 인공지능 기술이 접목되면서 모바일 기지국, 토지 피복, 건물 위치, 대중교통 이용 데이터 등 다양한 데이터의 공간 변수를 활용한 격자 단위 인구 추정이 가능해진 것으로 소개되고 있다. 초기 연구는 주로 현재 시점의 인구 분포를 산정하는 데 그쳤으나, 딥러닝 기반 시계열 예측 모델이 도입되면서 미래 변화를 직접 예측하는 연구가 확산하는 추세이다. 최근에는 전통적인 인구학적 모델과 AI 기법을 결합한 하이브리드 접근이 등장하여, 1km 격자 단위 공간으로 정합성 있는 장래인구 예측이 가능해졌으며, 이는 행정 경계를 넘어선 맞춤형 정책 수립으로 이어지고 있다.

이러한 발전 흐름은 전통적 방법론의 신뢰성과 AI의 예측 역량을 결합하는 새로운 연구 패러다임을 제시한다는 점에서 그 효용성이 높다. 동시에 데이터 융합, 격자 단위 통계 생성, 토지이용 변화와 환경 요인 반영 등 향후 고도화 방향을 제시한다. 이는 단순 인구 예측에 그치지 않고 더 나아가 인구이동의 시계열 주기 또한 수년 후의 인구이동뿐만 아니라 수 시간 내의 유동인구 예측에까지 그 가능성이 기대되고 있다.

제4장
공간분석 기반 변수설계 및 AI 예측모형 개발

본 장에서는 앞서 제시한 AI 기반 인구이동 예측모형 개발의 필요성과 선행연구에서 확인한 주요 한계점 극복을 위해, 시공간 분석을 기반으로 한 정책 지향형 예측모형인 AI-PriSM(가칭, AI-based Policy-oriented Spatio-temporal Prediction for Population Migration)의 설계와 구현에 관련된 내용을 제시한다. 이를 위해 시공간 특성을 반영한 변수 선별과 데이터 전처리 과정을 거쳐, 인구이동과 관련된 핵심 요인을 식별하였다. 나아가 시범적 데이터 탐색 분석(Exploratory Data Analysis, EDA)을 통해 변수 간 인과관계를 검증하고, 이를 토대로 정책 시뮬레이션에 적용할 수 있는 예측모형을 구현하였다. 이러한 접근은 정밀한 공간 단위의 인구이동 예측을 가능하게 하여, 지역 맞춤형 정책 수립의 실효성을 극대화하는 데 그 목적이 있다.

제 1 절 데이터 수집 및 전처리

1. 데이터 수집

본 연구에서 제시하는 AI 기반 인구이동 예측모형 AI-PriSM(가칭) 개발을 위해, 인구이동과 지역 특성을 반영할 수 있는 다양한 데이터를 수집하였다. 먼저, 인구이동 변화를 확인하기 위해 국가데이터처(舊통계청, MDIS)[10]에서 제공하는 읍·면·동 단위의 인구이동 데이터를 활용하였다. 해당 데이터에는 다양한 속성정보가 포함되어 있으나, 본 연구에서는 인구이동 집계, 전입 사유, 직전 주소지 등 모형과 직접적으로 관련된 항목을 중심으로 분석하였다.

또한, 격자 단위 인구 정보는 핵심 변수로서 국가데이터처(舊통계청, SGIS)[11]의 1km 격자 총인구 집계를 활용하였다. 이 데이터는 2000년부터 2023년까지 제공되며, 2000~2015년은 5년 주

10) 마이크로데이터 통합서비스. 국가데이터처, https://mdis.kostat.go.kr/index.do(2025년 7월 31일 접속)
11) SGIS 통계지리정보서비스. 국가데이터처, https://sgis.kostat.go.kr/view/index(2025년 7월 31일 접속)

기, 2016년 이후는 1년 주기로 생산되는 특징이 있다. 동일 시·공간 해상도 기준으로 총주택수, 총가구수 등 인구 관련 변수와 대분류 기준 총사업체수 등 지역 경제활력도를 나타내는 지표도 함께 사용하였다.

정주 여건 및 상권 특성을 반영하기 위해 지방행정인허가[12] 데이터를 수집하였으며, 국가교통DB[13]에서 확보한 교통망 데이터 역시 정주 여건 변수로 포함하였다. 이를 통해 인구이동에 영향을 미칠 수 있는 물리적·사회경제적 환경 요인을 종합적으로 반영할 수 있도록 하였다.

이상 서술한 본 연구에서 수집되고 활용한 데이터의 개괄적인 속성정보 및 개요는 〈표 4-1〉에 제시하였다.

〈표 4-1〉 본 연구에서 활용한 주요 데이터 항목 및 개요

구분	데이터 항목	주요 속성	시간 단위	공간 단위	출처
1	인구이동	인구이동 집계, 전입 사유, 직전 주소지 정보 등	2001~2024	전국 읍면동	MDIS
2	인구	격자 단위 전체 및 연령별 인구	2000, 2005, 2010, 2015~2023	전국 1km 격자	SGIS
3	총주택수	격자 단위 총주택수			
4	총가구수	격자 단위 총가구수			
5	총사업체 종사자수 (대분류)	격자 단위 대분류 기준 총사업체 종사자수			
6	총사업체수 (대분류)	격자 단위 중분류 기준 총사업체수			
7	사망자수	시군구별 사망자수	2001~2024	시군구	국가데이터처
8	신생아수	시군구별 신규 출생아수	2001~2024	시군구	국가데이터처
9	상권정보	문화, 식품, 생활 등 지역의 상권과 연관된 각종 위치정보 데이터	등록 시~현재	전국 위치정보	지방행정 인허가 데이터
10	도로망	전국 도로망 정보	2009~2024	전국	KTDB
11	KTX역	전국 KTX역 위치 정보			

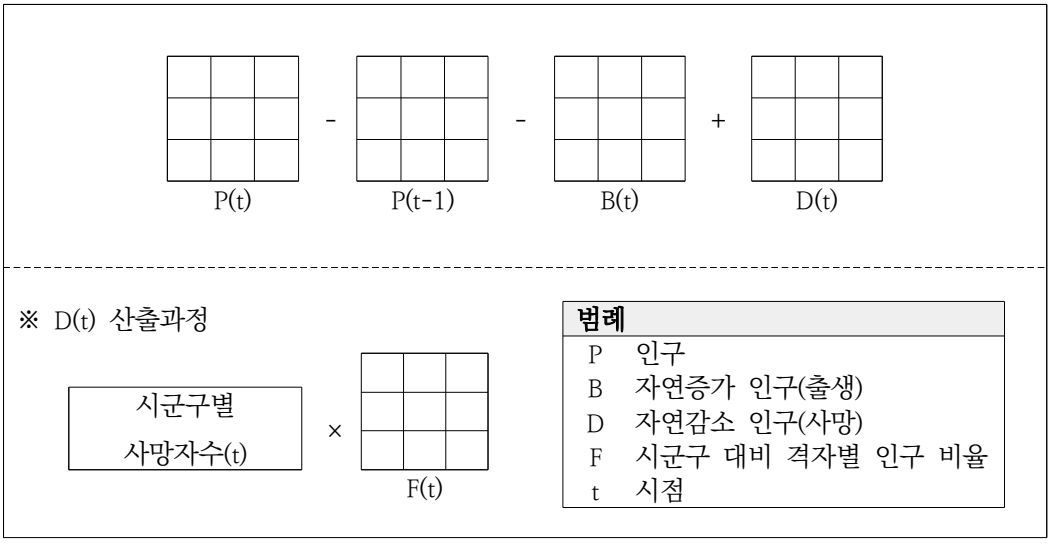
출처: 저자 작성

12) 지방행정인허가데이터, https://www.localdata.go.kr/(2025년 7월 31일 접속)
13) 국가교통DB, https://www.ktdb.go.kr/(2025년 7월 31일 접속)

2. 데이터 전처리

이렇게 수집된 데이터는 시·공간 단위 분석과 예측 정확도 향상을 위해 일원화 과정을 거쳤다. 먼저, 읍·면·동 단위로 집계된 인구이동 데이터는 격자 단위로 변환하였다. 인구이동의 증감을 반영하기 위해 직전연도(t-1) 대비 당해연도(t)의 총인구 차이를 계산하고, 여기에 유아 인구(자연증가분) 차를 차감한 후, 시·군·구별 사망자수(자연감소분)를 격자에 따라 비율 배분하였다. 이러한 변환 과정은 읍·면·동 인구를 격자 단위로 할당하는 과정에서 일부 내재적 오차가 발생할 수 있으나, 격자 비율 배분과 집계 총량 정합성을 유지할 수 있다는 점에서 시도되었다(그림 4-1 참조).

[그림 4-1] 본 연구의 읍면동 및 격자 단위 인구이동 집계방식 도식화

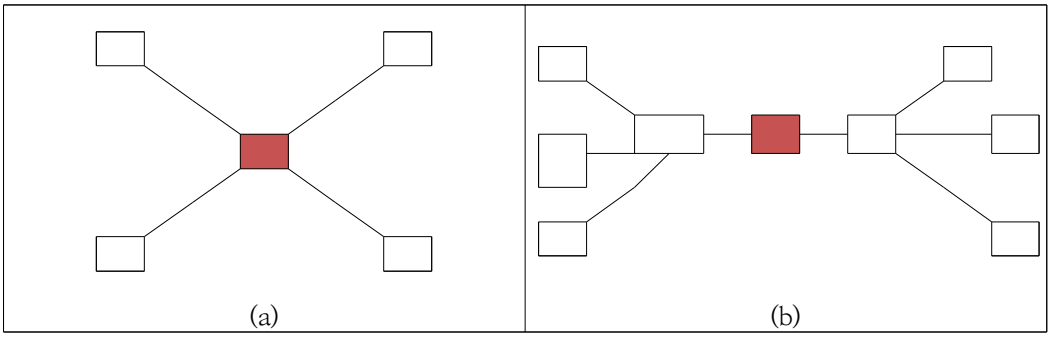


출처: 저자 작성

지역상권 정보는 위치 기반으로 집계된 지방행정인허가 데이터를 활용하여 격자 단위 밀도를 산출하고, 이를 0~1 범위로 정규화하여 사용하였다. 상권의 세부 분류는 21개 산업 대분류를 기준으로 구분하여 반영하였다.

도로망 접근성의 경우, 국가교통DB의 선형(Polyline) 데이터 특성을 고려하여 그래프이론을 적용하였다. 도로망을 노드(node)와 링크(link)로 구성된 네트워크로 간주하고, 연결중심성(Degree centrality)과 매개중심성(Betweenness centrality)을 산출한 뒤 이를 격자별 평균값으로 변환하였다. 연결중심성은 가장 많은 접점의 중심이 되는 노드를, 매개중심성은 주요 경로상에 위치해 다른 노드 간 이동에 중요한 역할을 하는 노드를 의미한다. 실제 도로망에서 두 지표가 높을수록 접근성과 편의성이 높다고 해석할 수 있다(그림 4-2 참조).

[그림 4-2] 그래프 이론의 연결중심성(a)과 매개중심성(b) 개념도



주: (a)는 연결중심성(degree centrality)을, (b)는 매개중심성(betweenness centrality)을 도식화
출처: 저자 작성

철도망의 경우, 네트워크 구조보다는 역세권 접근성을 중심으로 분석하였다. 기존 연구에 따르면 KTX 역세권 거리는 정주여건, 인구유입, 생활편의 등에 중요한 영향을 미친다(허재완 외, 2018; 윤준희·진혁, 2024; 하미란·진장익, 2024). 이에 따라 KTDB에서 제공하는 기차역 위치를 기준으로 5km 거리 버퍼를 설정하여 활용하였고 버퍼의 포함 여부를 격자에 배분하여 활용하였다.

마지막으로 각 제시된 값은 그 간극이 다양하기 때문에 Min-Max Normalization 등의 방법을 통해 0과 1 사이의 값으로 치환하여 분석에 활용한다. 아래는 일반화 식을 표현한 것으로, 관측값 X가 그룹의 최댓값과 최솟값을 통해 0과 1 사이의 값으로 치환하게 되는 방식이다.

$$X_{normalization} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

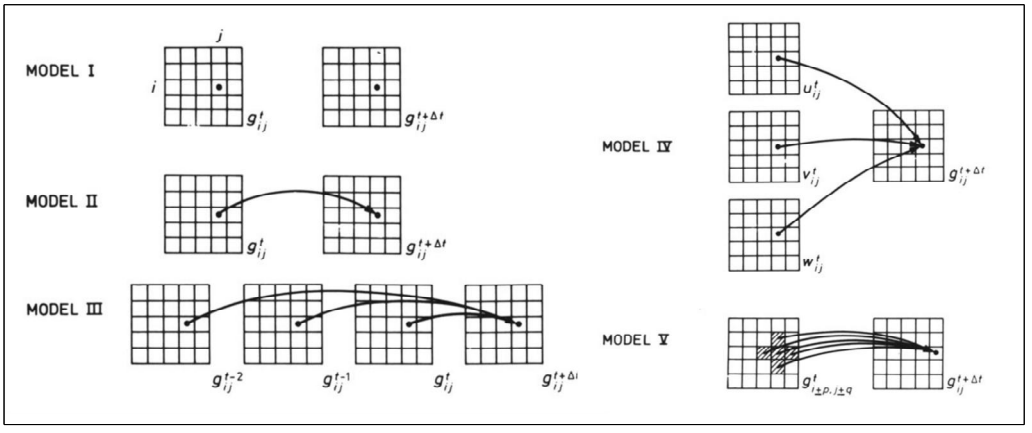
이러한 전처리 과정을 통해 서로 다른 출처와 공간 단위를 가진 데이터를 1km 격자라는 공통 분석 단위로 통합함으로써, 시·공간적으로 정합성이 높은 예측 데이터를 구축할 수 있다. 이는 인구이동의 세밀한 공간 분포와 주요 영향 요인을 정밀하게 분석할 수 있는 기반이 되며, 향후 정책 시뮬레이션과 의사결정 지원에 활용 가능성을 높인다. 다만, 읍·면·동 단위 인구를 격자 단위로 변환하는 과정에서 발생할 수 있는 할당 오차와, 일부 변수의 시계열·공간 해상도 차이로 인한 데이터 보정 한계는 향후 모형 해석 시 고려해야 할 점으로 남는다.

제 2 절 격자 기반 공간 네트워크 분석

본 연구에서 제시하는 AI 기반 인구이동 예측모형 AI-PriSM(가칭)은 격자 기반의 공간 네트워크를 분석한다는 데 그 의의가 있다. AI-PriSM(가칭) 구현을 위해서 본 연구는 Tobler(1979)가 제시한 Cellular Geography 개념을 적용하였으며, 이는 이미 다양한 선행연구에서 공간 단위 네트워크 분석 방법론으로 활용된 바 있다(이보경 외, 2022; 장요한 외, 2024).

Cellular Geography는 격자 단위 시공간 데이터의 인과관계를 추론하기 위해 설명되는 이론으로, 시점 t의 특정 격자 G(t)가 과거 시점 G(t-1)과 독립적인 관계를 가지거나 종속적인 관계를 가질 수 있음을 가정한다. 더 나아가 이러한 종속성은 단일 시점의 과거 데이터뿐만 아니라 G(t-2), G(t-3), …, G(t-n)과 같은 장기 시계열의 누적 효과에 기인한 결과로 설명되기도 한다. 또한, 한 격자의 상태는 해당 격자 자체의 변화뿐 아니라 인접 격자의 종합적인 영향의 결과일 수 있다. 이때 인접 격자의 영향 또한 단기적 변화뿐 아니라 시계열적으로 누적된 결과일 수 있다. 본 연구에서는 이러한 Cellular Geography의 개념에 입각하여, 인접 격자와 시계열 관계를 동시에 학습할 수 있는 데이터 구조를 설계하여 AI가 학습할 수 있도록 구성하였다. 이를 AI-PriSM 모형에 적용함으로써, 모델이 인구이동의 시공간적 연속성과 공간 상호작용을 최대한 정교하게 포착할 수 있도록 구성하였다.

[그림 4-3] Cellular Geography 개념 도식화



출처: Tobler(1979) pp. 379-380

제 3 절 변수 정의 및 주요 입력 구조 설계

AI 모형에서 다양한 데이터의 구성은 모형의 예측력에 도움이 될 수 있다는 것은 자명한 사실이다. 하지만, 현재의 시점을 기준으로 장래의 인구이동을 예측해야 한다는 점을 고려할 때 다양한 데이터의 활용에는 다양한 제약이 수반될 수밖에 없다. 회귀모형을 예로 들면, 장래 특정 시점의 예측값을 추정하기 위한 종속변수(dependent variable)가 있고, 이 종속변수에 영향을 주는 독립변수(independent variable)가 있다. 여기서 장래 시점의 종속변수를 얻어내기 위해서는 장래에도 독립변수가 계속 유지되거나, 아니면 현재 시점으로 장래의 종속변수 상태를 예측할 수 있도록 모형을 구성해야만 현재 시점에서 구축된 회귀모형이 장래에도 성립될 것이다. 다시 말해, 독립변수가 장래에도 활용할 수 있는 특성을 가졌다면 모형의 인과관계 구성은 현재의 독립변수 구성이 장래에도 변하지 않을 것이라 가정하고 종속변수 간 인과관계를 학습할 수 있도록 구성하면 되고 그렇지 않다면, 현재의 독립변수 구성을 바탕으로 장래의 종속변수를 추정하거나 제외하는 방식으로 모형을 구성하는 식이다.

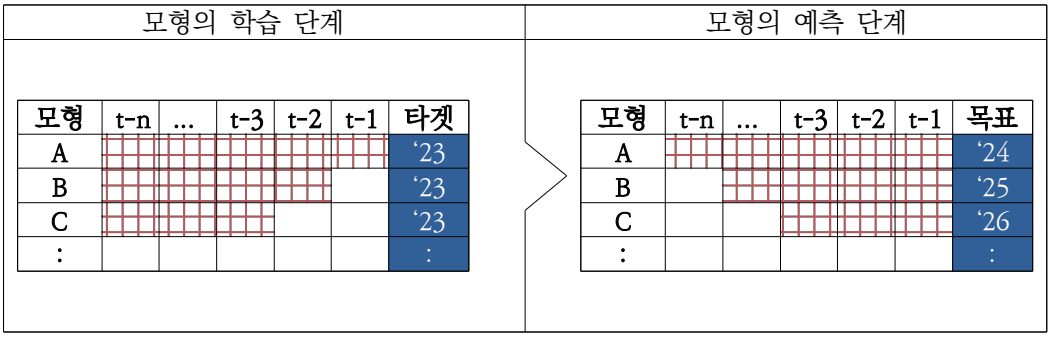
현재를 기준으로 각종 데이터를 학습한 모형을 장래에 활용하기 위해서 독립변수를 장래에도 불변한다고 가정하거나 장래의 독립변수를 함께 예측하는 방식도 충분히 가능한 방법임에는 틀림이 없으나, 예측모형이 먼 장래로 갈수록 그 신뢰도가 저하될 수 있는 한계가 있다(이보경 외, 2022). 또한 복잡한 독립변수의 가정을 제외한 뒤 단순히 주요 변인만을 활용하여 장래를 추정하는 경우 제한적인 독립변수 때문에 미래 정책 시뮬레이션이 어려운 한계가 있다(이태리 외, 2020). 예측 시점이 가까운 미래에 수렴하는 경우 본 연구에서 활용하고자 하는 정책 시뮬레이션의 기능을 고려하여, 현재의 독립변수 조합을 최대한 유지하면서 장래 연도 종속변수 예측을 유도할 수 있는 모형이 필요할 것이다.

본 연구에서 제시하는 인구이동 예측에 있어서 장래 연도에 특별한 가정 없이 활용할 수 있는 정보는 연도를 제외하면 전무한 상황이기 때문에, AI-PriSM(가칭)에서는 현재 또는 과거 시점의 독립변수 정보를 바탕으로 미래의 인구이동 정보와 관련한 정보를 예측할 수 있도록 구성하였다. 이를 위해서 목표 예측 기간인 2025~2030년을 가정하고, 보유한 최신 데이터를 활용하여 'N년 후'를 예측하는 다중 시차 학습 모형을 구성하였다. 즉, [모형 A]는 2022년까지의 정보를 기준으로 1년 후인 2023년을 예측하도록 구성하였고, [모형 B]는 2021년까지의 정보를 기준으로 2년 후인 2023년을, [모형 C]는 2020년까지의 정보를 기준으로 3년 후인 2023년을, [모형 D]는 2019년까지 정보를 기준으로 4년 후를, [모형 E]는 2018년까지를 기준으로 5년 후를, 마지막으로 [모형 F]는 2017년까지의 정보를 기준으로 6년 후인 2023년을 예측하도록 구성하였다.

- 모형 A(1년 후 예측): 2000년부터 2022년까지 데이터 →(학습) → 2023년 예측
- 모형 B(2년 후 예측): 2000년부터 2021년까지 데이터 →(학습) → 2023년 예측
- 모형 C(3년 후 예측): 2000년부터 2020년까지 데이터 →(학습) → 2023년 예측
- 모형 D(4년 후 예측): 2000년부터 2019년까지 데이터 →(학습) → 2023년 예측
- 모형 E(5년 후 예측): 2000년부터 2018년까지 데이터 →(학습) → 2023년 예측
- 모형 F(6년 후 예측): 2000년부터 2017년까지 데이터 →(학습) → 2023년 예측

이렇게 학습된 모형은 장래 연도 예측 시에 순차적으로 적용된다. 예를 들어, [모형 A]는 2023년까지의 데이터를 활용해 2024년을, [모형 B]는 2023년까지의 데이터를 활용해 2025년을 예측하는 식이다. 이러한 순차 예측(Rolling Forecast) 접근법을 통해, AI-PriSM(가칭)은 단·중기인 5~6년 후까지의 인구이동을 예측할 수 있도록 아키텍처를 설계하였다.

[그림 4-4] AI-PriSM(가칭)의 주요 학습 및 예측 방식 도식화



출처: 저자 작성

설정된 주요 입력 구조에 따라서 입력 변수는 각 격자의 총인구수, 가구수, 주택수, 전입-전출 인구, 총사업체수 및 종사자수(대분류), 도로망 접근성 지수, 철도역 영향권 여부, 출산 및 사망자수, 상권정보 등이 사용되며, 시공간 격자별로 배분된 값을 입력 구조에 따라서 Cellular Geography 방식으로 구현하여 사용한다(표 4-2 참조).

또한 본 연구에서 제시하는 AI-PriSM(가칭) 모형은 인구이동의 정량적 예측뿐 아니라 그 배경요인 분석을 목적으로 두 가지 세부 모형으로 구분하여 설계하였다. 첫째, 격자별 인구이동 규모를 예측하는 인구이동 예측 모형, 둘째, 전입 사유별 비율을 예측하는 전입사유 예측모형, 셋째,

전입자의 직전 주소지(읍면동)를 예측하는 직전주소지 예측모형이 그것이다. 세 모형 모두 동일한 학습 데이터와 예측 구조를 기반으로 하나, 목푯값으로 설정되는 종속변수가 각각 다르며, 이를 통해 예측뿐 아니라 인구이동의 인과관계까지 통합적으로 분석할 수 있도록 하였다.

〈표 4-2〉 본 연구에서 활용한 주요 변수 개요

구분	변수명	설명	단위/범위
1	총인구수	격자 단위의 총인구	명/격자
2	총가구수	격자 단위의 총가구수	가구/격자
3	총주택수	격자 단위의 총주택수	호/격자
4	전입인구	해당 격자로 전입한 총인구수*	명/격자
5	직전 주소지	직전 주소지(읍면동)	명/읍면동
6	전출인구	해당 격자에서 전출한 인구수*	명/격자
7	총사업체 종사자수	대분류 기준 격자 단위의 총사업체 종사자수	명/격자
8	총사업체수	대분류 기준 격자 단위의 사업체수	개소/격자
9	도로망 접근성 지수	연결·매개중심성 기반 접근성 지수**	0~1
10	기차/지하철역 위치	기차역으로부터 반경 5km 이내의 포함 여부**	개소/위치
11	신생아수	해당 지역의 연간 신생아수	명/시군구
12	사망자수	해당 지역의 연간 사망자수	명/시군구
13	상권정보	업종 대분류별 상권정보	개소/주소

주: * 읍면동 정보와 격자별 인구 분포 관계로부터 유추한 정보임
** 추가 분석을 통해 유추한 정보임
출처: 저자 작성

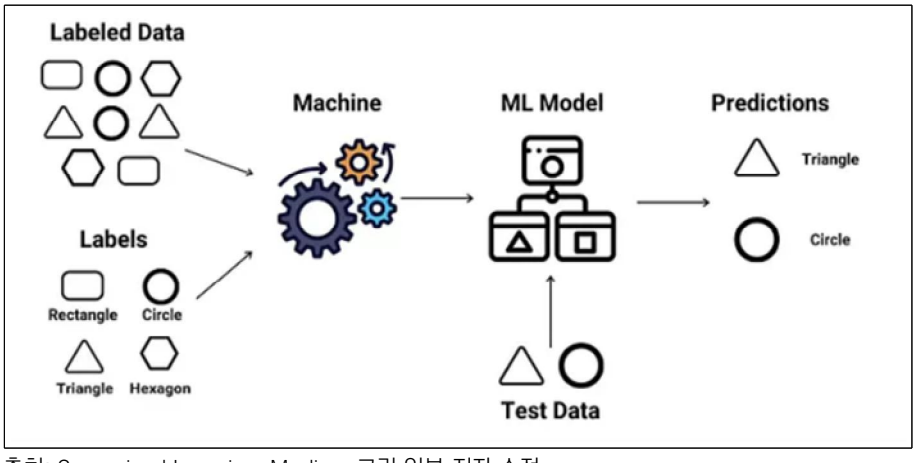
제 4 절 예측모형 알고리즘 구성

예측모형의 알고리즘은 AI-PriSM(가칭)의 핵심 요소로, 최종 예측 성능과 정책적 활용성을 결정한다. 본 연구에서는 다양한 알고리즘의 장단점과, 적용에 요구되는 적합성을 균형 있게 평가하기 위해 AI 학습 방식별로 주요 후보 모형을 선별하였다. 이후 설정된 학습 구조를 기반으로 개별 모형의 성능을 비교·분석한 뒤, 예측 정확도와 안정성을 고르게 충족하는 최적 알고리즘을 채택하였다.

1. 학습 방법에 따른 구분

최근 인공지능(AI)에 대한 관심은 전 세계적으로 지속해서 확대되고 있으며, 다양한 응용 시도가 이루어지고 있다. 그러나 모든 분야에 동일한 AI 모형이 적용되는 것은 아니며, 상황이나 문제 특성에 맞는 머신러닝(ML) 유형을 선택적으로 활용하는 것이 일반적이다. 머신러닝은 학습 방식에 따라 크게 지도학습(Supervised Learning), 비지도학습(Unsupervised Learning), 준지도학습(Semi-supervised Learning), 그리고 강화학습(Reinforcement Learning)으로 구분된다.

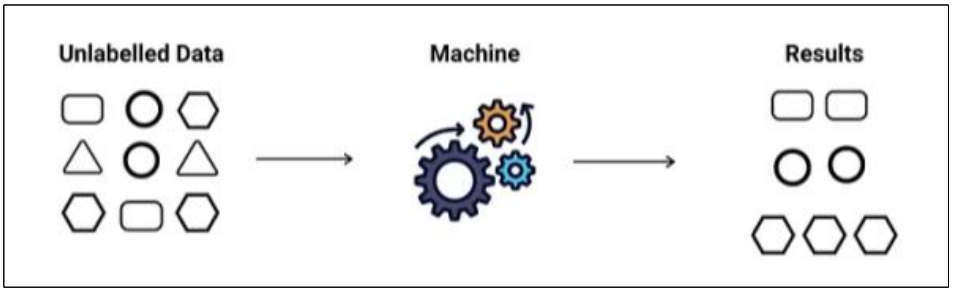
[그림 4-5] 지도학습(Supervised learning) 개념도



출처: Supervised Learning. Medium 그림 일부 저자 수정, https://medium.com/@metehankozan/supervised-and-unsupervised-learning-an-intuitive-approach-cd8f8f64b644(2025년 7월 31일 접속)

지도학습은 사전에 정의된 목푯값(표적 또는 레이블)을 기반으로, 입력 데이터와 목푯값 간의 관계를 학습하여 새로운 입력에 대한 결과를 확률적으로 유추하는 방식이다. 비지도학습은 목푯값이 주어지지 않은 상태에서 데이터 자체의 패턴을 찾아 군집화하거나 특성을 도출하는 방법이며, 예를 들어 추천시스템에서 유사도 기반 콘텐츠 분류 등에 활용된다. 준지도학습은 일부 데이터만 레이블이 존재하거나 학습할 수 있는 데이터가 제한적인 상황에서, 기존의 레이블 데이터로 초기 학습을 진행한 뒤, 미라벨 데이터에 예측값을 부여하고 이를 다시 학습에 반영하는 방식이다. 마지막으로 강화학습은 특정 환경에서 주어진 목표를 달성하기 위해, 에이전트(agent)가 반복적인 시도와 피드백(보상, reward)을 통해 최적의 행동 전략을 학습하는 방법이다.

[그림 4-6] 비지도학습(Unsupervised learning) 개념도

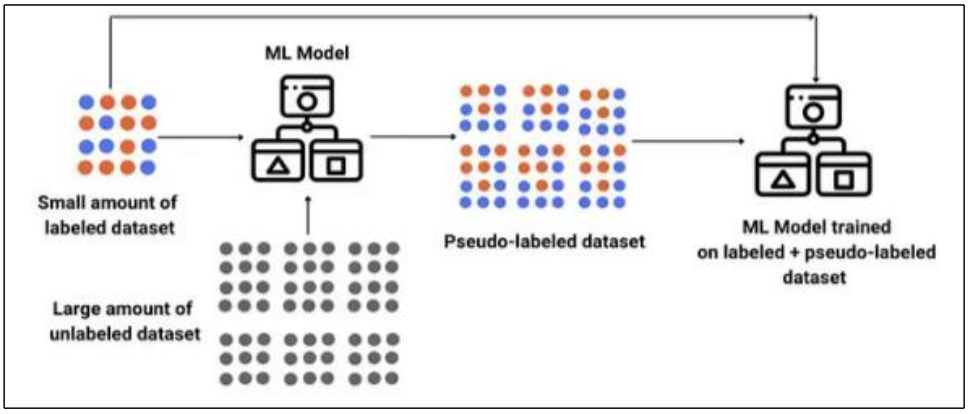


출처: Unsupervised Learning. Baoyu. 그림 일부 저자 수정,
https://baoyu.io/translations/ai-reports/stanford-hai-ai-index-report-2024-chapter1-en
(2025년 7월 31일 접속)

머신러닝(ML)은 유형별로 서로 다른 특성을 가지며, 이에 따라 각 모델의 장단점 또한 뚜렷하게 구분된다. 예를 들어, 지도학습은 다른 유형에 비해 상대적으로 높은 정확도(약 90% 이상)를 달성하는 경우가 많다는 강점이 있다. 그러나 학습 과정에서 접하지 않은 데이터에 대해서는 패턴 인식에 실패하여 예측 성능이 급격히 떨어질 수 있다는 한계를 지닌다. 비지도학습은 사전에 목푯값(target)을 정의하지 않고 데이터의 분포나 내재한 데이터 패턴을 기반으로 군집화(clustering)하는 방식이 주를 이루는데, 일반적 정확도는 70~80% 수준에 머무르지만, 데이터 오차(noise)나 이상치(outlier)에 대한 민감도가 낮아 예측 경험이 없는 데이터에 대해서도 위험을 완화(robust)할 수 있는 장점이 있다. 준지도학습은 지도학습과 비지도학습의 장점을 결합한 방식으로, 레이블이 없는 데이터에 임곗값(threshold)을 설정해 조건에 부합하는 경우 새로운 레이블을 부여하고 이를 학습에 반영한다. 이는 데이터의 양이 적을 때 효과적이지만, 지도학습과 마찬가지로 표본 데이터의 대표성이 결과의 정확도를 크게 좌우하므로 사전 검토가 필수적이다. 또한, 추가 데이터의 비율이 지나치게 높으면 과적합(over-fitting) 가능성이 커지며, 시계열 예측 모델의 경우 시간이 지남에 따라 오차 범위가 걷잡을 수 없이 확대될 수 있다. 강화학습은 특정 환경과

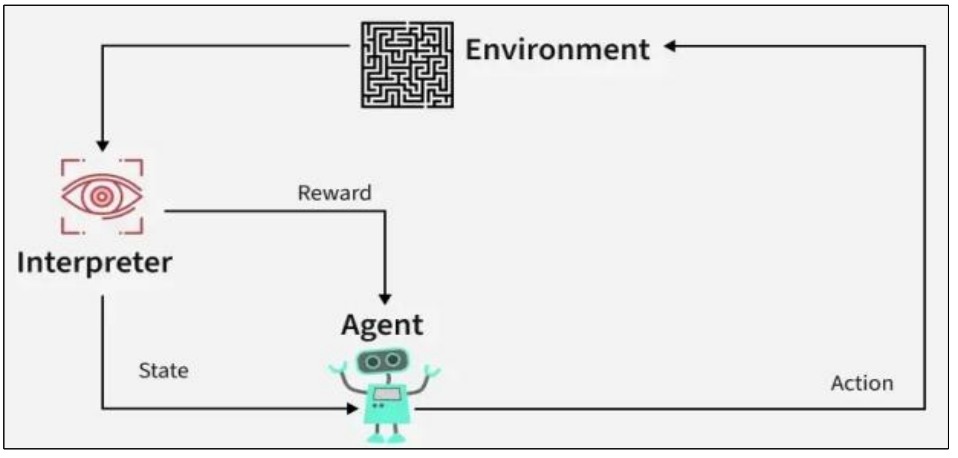
상황에 맞춰 설계되며, 비선형 데이터 구조 처리에 강점을 가진다. 반복적인 상호작용과 보상(reward) 메커니즘을 기반으로 높은 정확도와 강인성(robustness)을 보이지만, 병렬 연산이 어렵고 학습 시간에 상당한 시간이 요구되는 하드웨어 성능에 종속적인 제약이 있다.

[그림 4-7] 준지도학습(Semi-supervised learning) 개념도



출처: Semi-supervised Learning. Medium. 그림 일부 저자 수정,
https://medium.com/@gayatri_sharma/a-gentle-introduction-to-semi-supervised-learning-7afa5539beea(2025년 7월 31일 접속)

[그림 4-8] 강화학습(Reinforcement learning) 개념도



출처: GeeksforGeeks, 그림 일부 저자 수정,
https://www.geeksforgeeks.org/what-is-reinforcement-learning/(2025년 7월 31일 접속)

한편, 딥러닝은 머신러닝 유형별 장점을 계승하면서도 이미지·텍스트 등 비정형 데이터 처리 능력이 뛰어난 접근법이다. 머신러닝과 딥러닝 모두 적용 시에는 해결하려는 문제의 성격, 데이터 구조와 특성을 종합적으로 고려해 모델을 설계해야 한다. 특히, 데이터가 단순한 선형(linear)

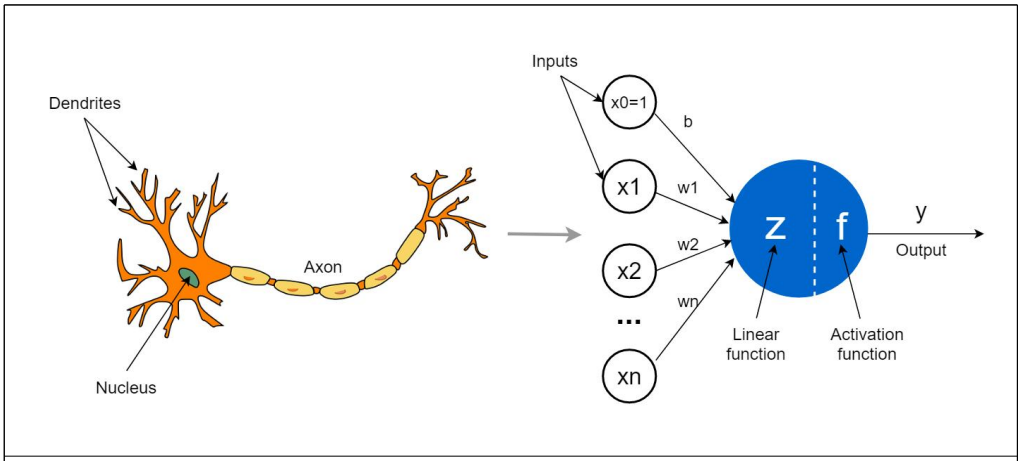
구조일 경우 복잡한 AI 기법을 적용해도 효과가 제한적일 수 있으나, 비선형(nonlinear) 구조를 띠거나 예측 과정에서 불규칙적 충격(shock)과 특이 변수가 빈번히 작용하는 문제 환경에서는 비선형 데이터 처리에 특화된 모델을 선택해 구현하는 것이 필요하다.

2. 인공신경망 구조의 특징

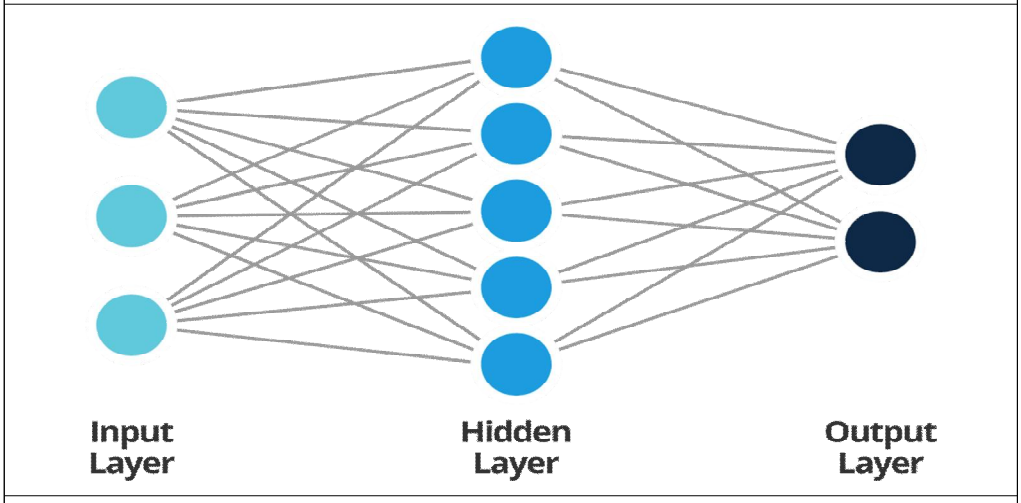
비선형적 데이터 구조를 처리하는 데 적합한 접근방식으로 지도학습군에 속하는 인공신경망(Artificial Neural Network, ANN) 기반의 머신러닝 또는 딥러닝 기법이 널리 활용된다. ChatGPT와 같은 대규모 언어모델 사례에서도 그 효용성이 입증된 바 있는데(그림 4-9 참조), 인공신경망은 인간 두뇌의 신경 신호 전달 과정을 모사한 퍼셉트론(Perceptron) 구조로, 입력층(Input Layer)-은닉층(Hidden Layer)-출력층(Output Layer)의 3단계 계층으로 구성된다. 각 계층은 다수의 뉴런(neuron)으로 이루어지며, 각 뉴런은 이전 및 다음 계층의 뉴런과 가중치(weight)를 매개로 연결된다. 입력층에서는 외부의 다양한 신호를 입력층의 입력 뉴런을 통해 감지하고, 이를 은닉층으로 보내 적절한 종합적인 신호를 출력층으로 전달하는 구조라 할 수 있다. 학습 과정에서는 출력층의 목푯값(target)에 근접하도록 가중치를 반복적으로 조정하여 최적의 네트워크 아키텍처(architecture)를 형성한다. 특히 은닉층은 입력층에서 전달된 다양한 신호를 종합·변환하여 다음 계층으로 전달하는 핵심 역할을 하며, 은닉층이 다수로 확장될 경우 이를 다층신경망(Multi-layer Neural Network) 또는 딥러닝(Deep Learning)이라고 한다. 이러한 구조 확장은 뉴런 간 결합이 단순 선형 관계를 넘어 복잡한 비선형 상호작용을 반영할 수 있도록 하는데, 이로 인해 인공지능 모델이 '블랙박스(black box)'로 불리는 이유가 되기도 한다.

앞서 선행연구에서 간략하게 다루어진 바 있는 LSTM, RNN 모두 이 인공신경망 구조를 기반으로 아키텍처의 변형을 통해 적용된 모델이며, ChatGPT의 대규모 언어모델도 이 구조를 근간으로 발전된 모형이라 할 수 있다. 인공신경망 모형의 학습은 매우 중요한데, 입력층과 은닉층, 은닉층과 출력층 간의 연결된 뉴런별 가중치를 어떻게 다루는지가 핵심이며, 어느 만큼의 반복적인 학습을 통해 충분한 학습을 유도해 내는지가 모델의 예측에서 큰 비중으로(critical) 작용한다.

[그림 4-9] 인공신경망의 일반적인 구조



(a) 사람의 신경망과 인공신경망의 기본인 퍼센트론(percentron) 구조 예시



(b) 일반적인 인공신경망의 구조

출처: (a) Skyflow,
https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc/(2025년 7월 31일 접속)
(b) Almabetter 그림 일부 저자 수정,
https://www.almabetter.com/bytes/articles/artificial-neural-network(2025년 7월 31일 접속)

3. 구동 방식에 따른 구분

한편, AI 모형은 학습 방식뿐 아니라 구동 방식에 따라 통계 확률 기반 모형, 분류 기반 모형, 인공신경망 구조 모형, 시계열 특화 모형, 혼합형 모형 등으로 세분할 수 있다.

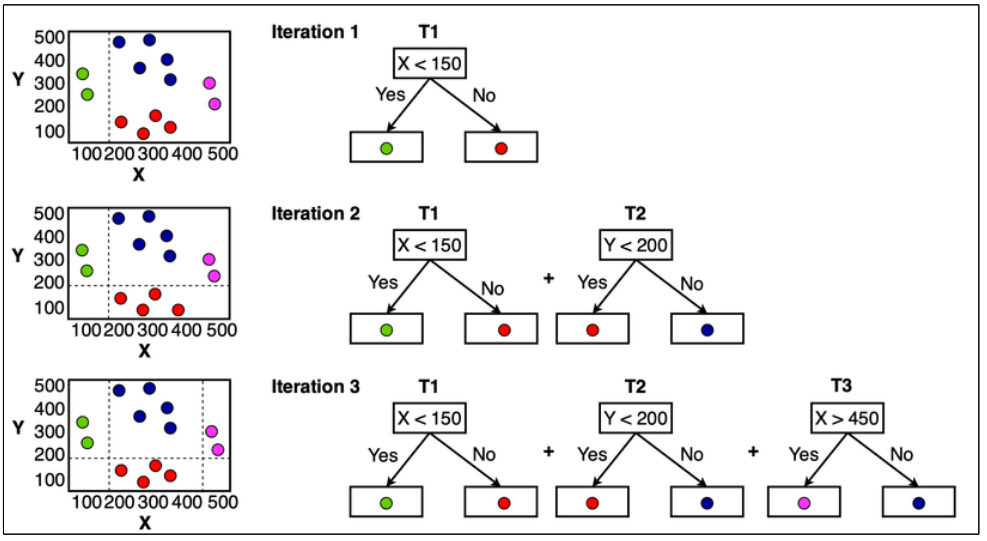
먼저, 통계 기반 모형의 대표 사례로는 마르코프 연쇄(Markov Chain) 모형이 있다. 이는 변수

의 시간적 연속성을 전제로 조건부 확률을 계산하여 미랫값을 예측하는 방식으로, 인과관계 파악에 있어서 통계적 해석이 쉽다는 장점이 있다. 그러나 조건부 확률 계산에 필요한 연산량이 방대하여 대규모 데이터 환경에서는 효율성이 떨어지는 한계가 있다.

분류 기반 모형은 데이터의 독립변수 간 경곗값을 찾아 종속변수를 구분하는 방식이며, 대표적으로 랜덤 포레스트(Random Forest)와 Extreme Gradient Boosting(xGBoost) 등이 있다(그림 4-10 참조). 두 기법 모두 표본과 임곗값 추정을 반복하여 평균 회귀에 근접하는 개념이지만, xGBoost는 분산형 계산 구조를 채택하기 때문에 대규모 데이터에 효과적으로 알려져 있다. 이들 모형이 대규모 데이터에 효과적으로 적용이 가능한 장점이 있지만, 어디까지나 선형(linear) 기반의 모형이기 때문에 학습에 사용되는 데이터의 분포가 충분하지 않거나 비선형(non-linear) 구조의 문제라면 그 효용이 제한적인 한계가 있다.

시계열 특화 모형으로는 LSTM(Long Short-Term Memory)이 널리 사용된다. 이는 RNN(Recurrent Neural Network)의 진화된 모형으로, 과거부터 현재까지의 시계열 패턴을 점진적으로 학습(batch learning)하면서 불필요한 메모리 사용을 줄이고 시계열 특성을 유지하는 구조를 가지며 모형의 아키텍처는 인공신경망(ANN) 구조로 되어 있다.

[그림 4-10] 분류 기반 모형의 개념도



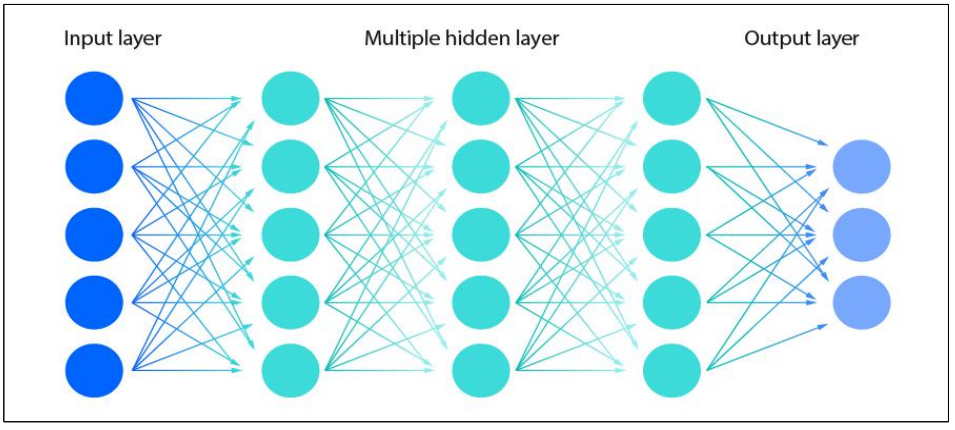
주: 독립변수군과 종속변수군 간의 다수 샘플링 및 확률 분포를 반복적으로 수행하여 예측모형을 구성하는 형태임

출처: KDnuggets, https://www.kdnuggets.com/wtf-is-the-difference-between-gbm-and-xgboost(2025년 7월 31일 접속)

4. 딥러닝 및 혼합모델

최근 고성능 컴퓨팅 자원 활용이 가능해지면서 딥러닝(DL) 등에 관한 노력도 다양하게 이루어지고 있는데, 딥러닝 계열에서는 일차적으로 인공신경망에서 은닉층의 학습을 반복적으로 추가하여 깊이 있는(Deep) 학습을 하도록 구성한 딥러닝 인공신경망(Multilayer Neural Network 또는 Deep Neural Network)이 있다. 또한 딥러닝 인공신경망 구조를 간직하면서 이미지 등의 픽셀 구조 격자를 학습할 수 있도록 고안된 합성곱 신경망(Convolutional Neural Network, CNN), 합성곱 신경망과 LSTM을 결합한 형태의 CNN-LSTM 모형 등이 있다(그림 4-12, 그림 4-13 참조). 특히 CNN-LSTM은 격자 단위의 집계라는 점에서 공간 단위 학습에 유리하여 이미 유사한 예측 문제에서도 활용된 바 있다(이보경 외, 2022; 장요한 외, 2024). 딥러닝 계열의 모델은 비선형 문제에 효과적으로 적용될 수 있는 것으로 알려져 있으나, 일반적으로 주어진 문제에서 일정 패턴을 추출하기 위한 충분한 학습 시간과 계산 자원이 필수적으로 요구되는 한계가 있다. 또한 기존 머신러닝 모델에서 CPU[14] 중심으로 이루어지던 계산 방식과는 달리 방대한 양의 데이터세트를 효과적으로 훈련하기 위한 GPU[15] 등 분산형 처리장치의 성능에 종속적으로 작용한다는 어려움이 있다(그림 4-14 참조).

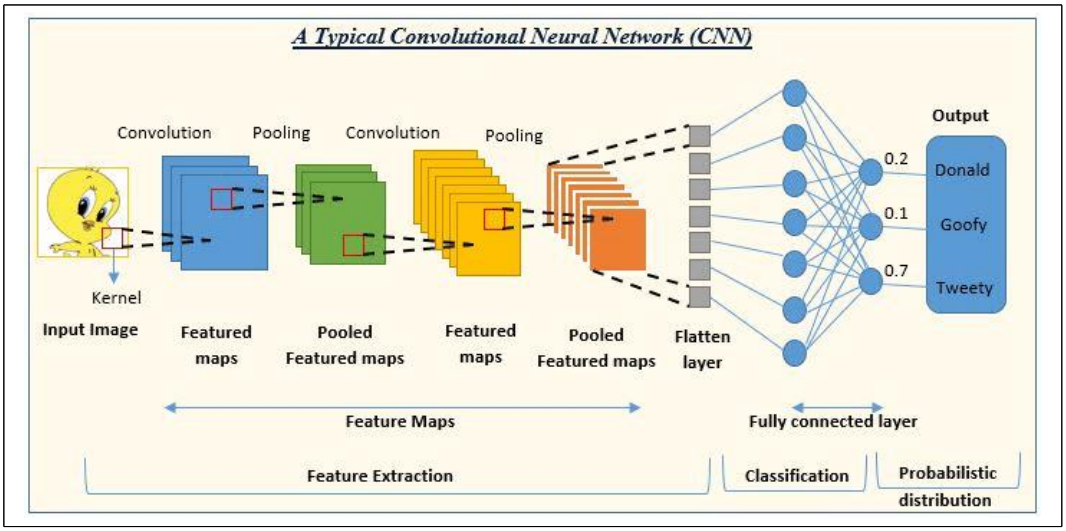
[그림 4-11] 딥러닝 인공신경망의 일반적인 구조



주: 입력층(input layer)- 다중 은닉층(hidden layer) - 출력층(output layer)으로 구성되어 있고, 각 층에 속한 뉴런(neuron)은 앞뒤로 복잡하게 연결된 형태임
출처: IBM, 그림 일부 저자 수정, https://www.ibm.com/think/topics/neural-networks(2025년 7월 31일 접속)

14) 중앙처리장치(Central Processing Unit)
15) 그래픽처리장치(Graphic Processing Unit)

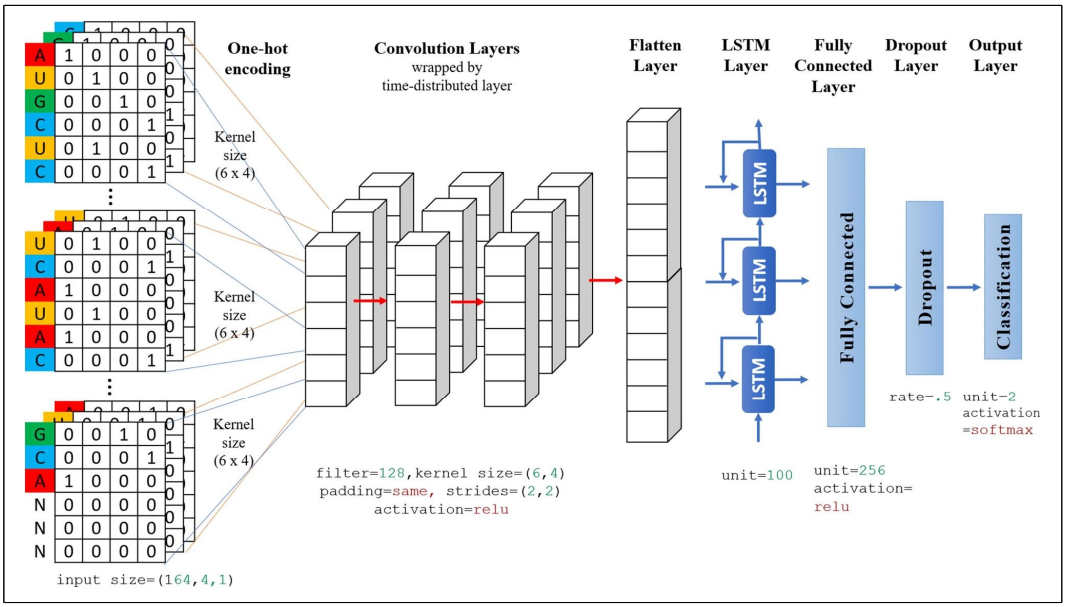
[그림 4-12] 합성곱 신경망(CNN)의 일반적인 구조



주: 주어진 이미지에서 특징을 추출(feature extraction)하는 단계에서는 이미지의 단순 특징 추출(featured maps) 및 데이터의 차원을 축소하는 단계(pooled fatured maps)를 반복한 뒤, 데이터의 차원이 충분히 줄어들면 나열하여(flatten layer) 신경망에 입력하는 형태로 구성됨

출처: Almabetter, https://www.almabetter.com/bytes/articles/convolutional-neural-networks (2025년 7월 31일 접속)

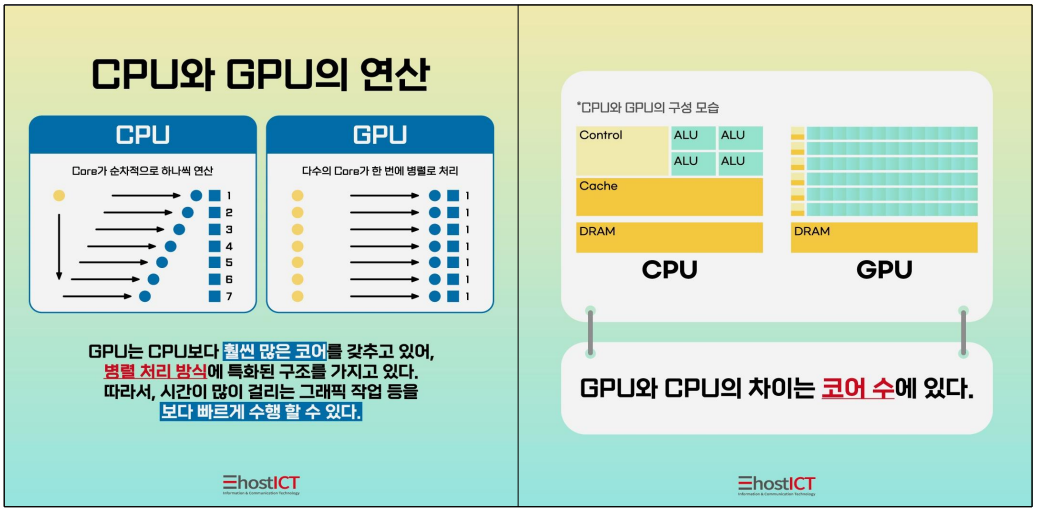
[그림 4-13] CNN-LSTM의 일반적인 구조



주: 다수의 매트릭스 값을 [그림 4-12]에서 설명된 CNN의 방식과 동일하게 추출한 뒤, 나열된(flatten layer) 데이터를 LSTM의 시계열 상관관계 학습을 통해 목푯값을 학습하고 예측하는 방식으로 구성됨

출처: Tasdelen and Baha(2021) p. 5

[그림 4-14] CPU와 GPU의 차이 도식화 예시

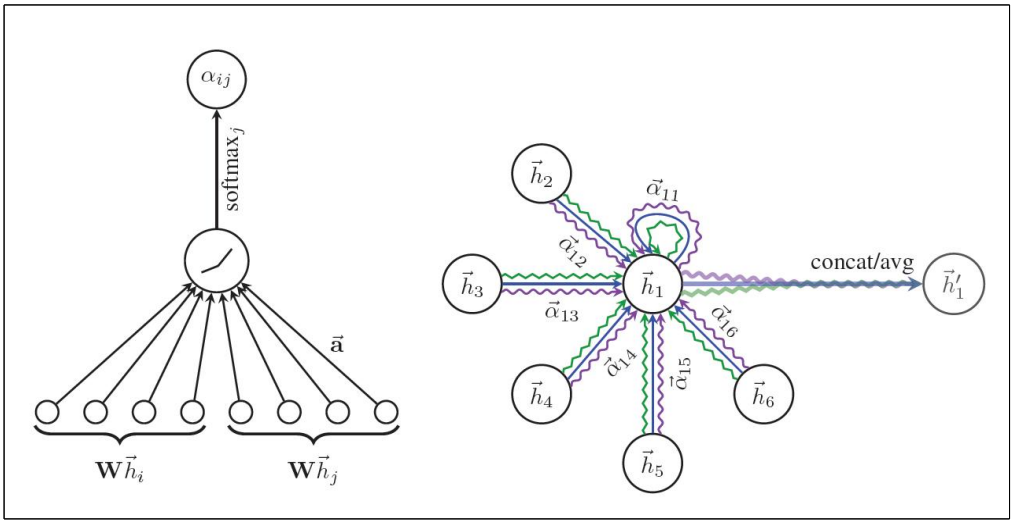


주: CPU와 GPU의 주된 차이는 병렬계산 및 계산 자원의 차이에 있음
출처: 이호스트 ICT 공식 블로그, https://webzero.tistory.com/3012(2025년 7월 31일 접속)

앞서 살펴본 인공신경망(ANN) 구조가 다수의 독립변수를 통한 소수 종속변수의 확률을 추론하는 방식이라면, 다수의 기점과 종점 간의 관계를 추출하여 그 패턴 및 기·종점 간 선택 확률을 추론하는 방식에 특화된 모형도 있다. 대표적인 예가 그래프이론(Graph theory)을 접목한 방식이 가능한데, 인공신경망 구조를 변형한 형태의 Graph Attention Networks(GATs)는 특정 입·출력(기·종점) 노드 쌍 간의 계수(weight)를 집중(attention)하여 학습함으로써 그 인과관계 패턴을 강화하는 방식으로, 다중(멀티헤드) 어텐션(attention) 학습을 통해 기·종점의 인과관계 도출에 효과적인 것이 특징이다(Velickovic et al., 2017. 그림 4-15 참조).

다수의 기·종점 관계가 시계열로 이루어졌을 때 유리한 모형도 소개된 바 있다. Park et al., 2020)에서는 도로 교통 속도 예측을 위한 Spatio-temporal Graph Attention Network (ST-GRAT)을 소개하였는데, 공간과 시간 어텐션을 구분하여 학습하고 기·종점 패턴을 집중하여 학습한 것이 특징이다. 먼저, 공간 어텐션은 diffusion prior와 방향성 헤드(inflow/outflow)를 통해 그래프 구조를 반영하고, 도로 상태에 따라 동적으로 공간 상관관계를 조정한다. 시간 어텐션은 RNN 없이 긴 입력 시퀀스의 과거 정보에 직접 접근하여 급격한 속도 변화를 포착하고, 마지막으로 Spatial Sentinel 벡터는 이웃 노드 정보가 불필요할 때 기존 특징을 유지하도록 하여 불필요한 어텐션을 방지하는 것이 ST-GRAT의 핵심 구성 요소라 할 수 있다(Park et al., 2020. 그림 4-16 참조).

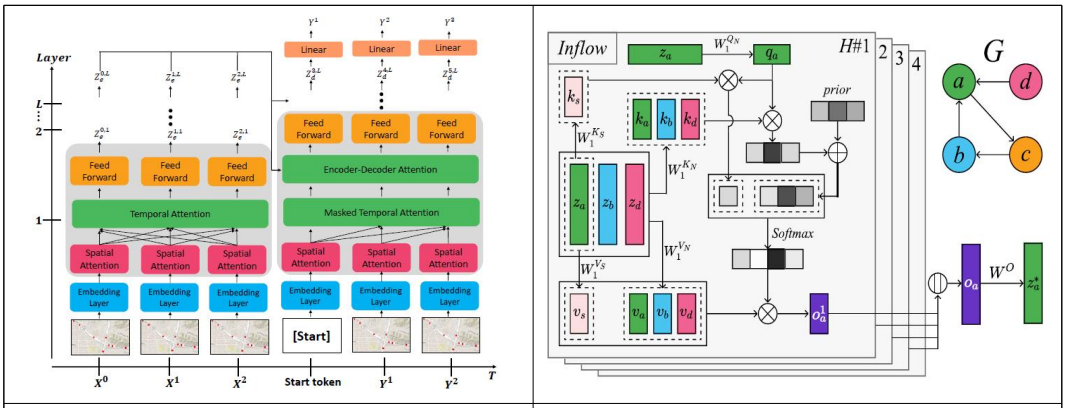
[그림 4-15] GAT의 작동 방식 도식화



주: 왼편의 그림은 단일 어텐션 메커니즘으로 노드 간 어텐션(강조) 계수 a_{ij}를 계산하는 과정을 나타내고, 오른편의 그림은 다중(멀티헤드) 어텐션 메커니즘으로, 이웃노드(h_1~h_6)에 대해서 다중 어텐션을 계산하는 과정을 나타냄

출처: Velickovic et al.(2017), p. 4

[그림 4-16] ST-GAT의 작동 방식 도식화



(a) (b)

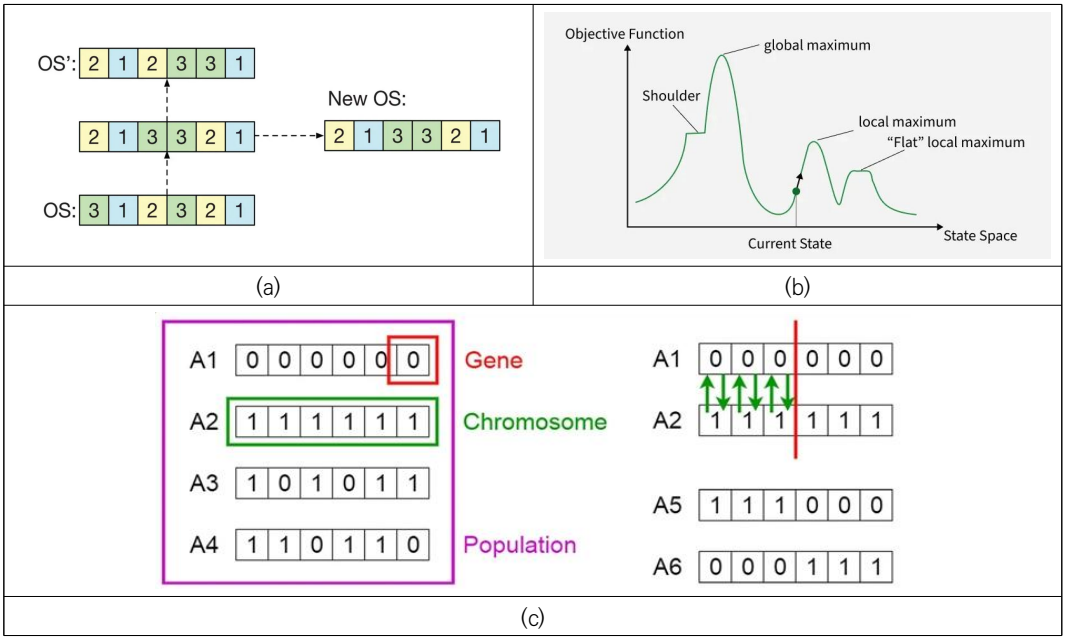
주: (a)는 ST-GRAT의 전반적인 아키텍처를 나타내는 그림으로 기본적으로 Transformer 구조를 가짐. x축은 시간을 y축은 레이어 수를 나타내는데, Embedding → Spatial Attention → Temporal Attention → Feed Forward 방식으로 진행. [Start] 토큰으로 시작하여 미래 시퀀스를 예측하는 방식 (b)는 공간 어텐션 메커니즘을 나타내는 그림으로, 그래프 구조 정보(거리, 연결성)를 추출하여 유사도와 인접 정보(diffusion prior)를 확인하고 sentinel 벡터를 통해 상관관계를 비교한 뒤 가중합을 추출하는 방식으로 진행

출처: Park et al.(2020), pp. 3-4.

5. 결정계수(hyper parameter) 최적화

머신러닝 및 딥러닝 구동을 위해서 필수적으로 요구되는 다양한 결정계수(hyper parameter)가 있는데, 그 경우의 수가 다양하고 예측이 쉽지 않기 때문에 휴리스틱(heuristic) 접근방식으로 해결하는 경우가 많다. 특히, 목푯값이 주어지지 않는 비지도학습의 경우 분석가의 결정계수 선택에 따라 절대적인 성능 차이가 발생하며, 이는 순차적인 학습을 통해 최적의 값을 학습하는 인공신경망 기반의 모형에서 더욱 두드러진다. 다양한 최적화 방법론이 현재 활용되고 있고, 본 연구에서는 반복적인 계산과 최적값 비교를 통해 최적의 결정계수 조합을 찾아가는 Tabu-search, Hill-climbing, 유전자알고리즘(genetic algorithm, GA)[16] 등의 기법을 활용하여 AI 예측 모형별 최적값을 탐색하는 방식으로 설계하였다(그림 4-17 참조).

[그림 4-17] Tabu-search(a)와 Hill-climbing(b), 유전자알고리즘(c) 개념 도식화 예시



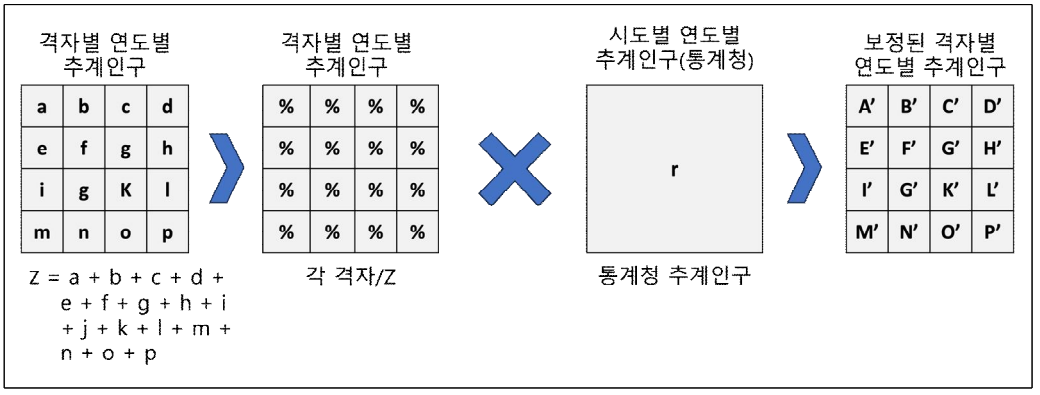
주: (a) Tabu-search: 최적의 조합을 찾는 과정에서 적절한 조합이 아닌 값들은 금기(Tabu)하여 최적 조합에 수렴할 수 있도록 구성
(b) Hill-climbing: 최적의 조합을 찾기 위해서 순차적으로 값을 바꾸어 가면서 계산하는 방식
출처: (a) fiveable, https://fiveable.me/combinatorial-optimization/unit-9/tabu-search/study-guide/grDB u3l BmHF5V9 Ya(2025년 11월 28일 접속)
(b) GeeksforGeeks, https://www.geeksforgeeks.org/artificial-intelligence/introduction-hill-climbing-artificial-intelligence/(2025년 11월 28일 접속)
(c) JanBask Traning, https://www.janbasktraining.com/tutorials/genetic-algorithm/(2025년 11월 28일 접속)

16) 유전자 알고리즘은 인류의 진화과정을 모방한 휴리스틱 최적화 접근방식으로 교배, 돌연변이 등의 결정계수 조합 시도를 통해 최적의 결정계수 조합을 찾아내는 방식이다(위키백과, 검색어: 유전 알고리즘)

제 5 절 예측 결과 및 성능 검증

예측 결과의 정합성과 정확성을 검증하는 것은 예측모형의 신뢰성을 확보하기 위한 핵심 과정이다. AI-PriSM(가칭) 모형은 다양한 데이터를 활용하여 장래 인구이동과 그 인과관계를 도출하는 데 중요한 강점을 가지고 있으나, 이러한 결과가 국가데이터처에서 제시하는 공식 추계치와 큰 차이가 발생할 경우 설득력이 저하될 수 있다는 한계가 있다. 따라서 본 연구에서는 분석 프레임 제시에 그치지 않고 예측값의 신뢰성을 높이기 위해 일련의 추가 보정 과정을 포함하였다. 보정은 국가데이터처의 시도별·연도별 추계인구를 기준으로, 모형이 예측한 격자별 인구를 시도 단위로 합산하여 양자의 차이를 비율로 산정하고, 그 비율만큼 격자별 예측값을 가감하는 방식으로 수행하였다. 이때 앞서 정의한 격자 단위 인구이동 집계방식을 적용하였으며, 출산율과 사망률은 국가데이터처 추계치를 반영하여 인구 변동량을 산출하였다. 이러한 절차를 통해 격자별 장래인구 추계의 불확실성을 최소화하고, 예측 결과의 정확성을 높여 정책적 활용 가능성을 제고할 수 있을 것으로 기대한다.

[그림 4-18] 모형의 정산 및 보정 개념도



출처: 저자 작성

모형의 효과척도(Measure of Effectiveness, MoE)는 기계학습 분야에서 널리 활용되고 있는 접근법 중 평균제곱근편차(RMSE)와 평균절대오차(MAE)를 활용한다. 추가로 직전 주소지 예측 모형에 대해서는 평균제곱오차(MSE)를 활용한다. RMSE는 참값(예: 2023년 격자 인구이동, y)과 예측값($\hat{y}$)의 차이를 제곱하여 모두 합산한 뒤, 이를 전체 표본 수 n으로 나누고 다시 제곱근을 취하는 방식으로 산출된다. 반면, MAE는 두 값의 차이에 대한 절댓값을 합산하고 이를 전체 표본 수로 나누어 계산한다. MSE는 RMSE와 유사한 형태를 가지지만, 제곱근을 적용하는 단계가 생략되어 RMSE보다는 상대적으로 그 민감도가 높게 도출되는 것이 특징이다. 세 지표 모두 값이 0

에 가까울수록 예측력이 우수한 모형임을 의미한다.

RMSE의 장점은 계산 결과가 직관적으로 해석 가능하다는 점이지만, 1 미만의 오차에는 상대적으로 둔감하며 값의 범위가 넓은 경우 결과의 신뢰성이 낮아질 수 있다는 한계가 있다. MAE 역시 직관성이 높고 참값 및 예측값과 동일한 단위를 갖는 장점이 있으나, RMSE와 유사하게 값의 분포 특성에 따라 신뢰성의 제약을 받을 수 있다. 따라서 합리적인 모형 비교를 위해서는 두 가지 이상의 성능 지표를 병행하여 다각적으로 검토하는 접근이 필요하다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y-\hat{y})^2}{n}}$$

$$MAE = \frac{\sum_{i=1}^{n}|y-\hat{y}|}{n}$$

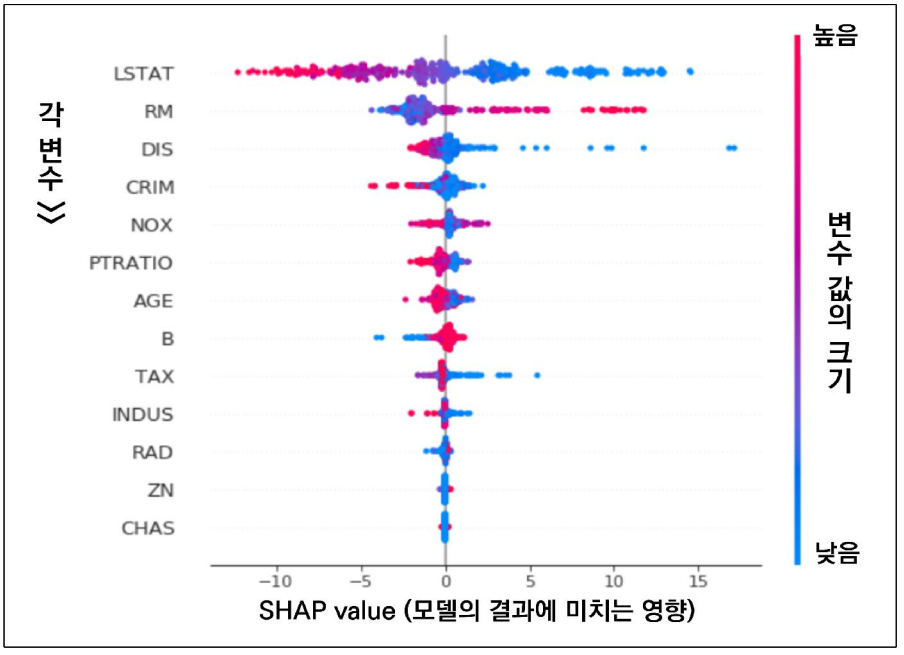
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i-\hat{Y})^2$$

마지막으로, 비교 예측모형의 결과 및 성능의 객관적인 비교를 위해 기본모델(baseline)을 추가한다. 이 기본모델(baseline)은 단순 회귀분석(regression)을 도입하여 과거추세 연장법을 적용한 방법으로, '현재까지 제공된 정보가 앞으로도 계속된다'라는 사실을 가정으로 연도별·격자별 목푯값의 R^2값이 최소가 되는 최적의 회귀식을 도출한 뒤 사용한다. 2000년부터 2023년까지의 시계열 정보를 바탕으로 전 국토인 10,744개의 격자별로 10,744개의 회귀식을 도출한 뒤 2030년까지 추정하여 산출하였다.

제 6 절 변수 영향도 및 민감도 분석

변수의 영향도와 민감도 분석은 AI-PriSM(가칭) 예측모형의 해석 가능성을 높이는 핵심 과정 중 하나이다. 본 연구에서는 비교 대상 모델 중 하나인 xGBoost의 변수 중요도 산출 기능을 활용하여, 모델 예측에 대한 각 변수의 기여도를 정량적으로 평가할 수 있는 SHapley Additive exPlanations(SHAP) 기법, 변수 중요도 등을 접목하여 활용하였다. SHAP는 대규모 데이터에 대해서도 개별 예측값에 대한 입력 변수별 영향을 직관적으로 시각화할 수 있어, 복잡한 구조의 예측 모형에서도 논리적인 설명력을 확보할 수 있는 장점이 있다.

[그림 4-19] SHAP 예시



출처: 저자 작성

분석 결과는 향후 정책 시나리오 설계와 정책 우선순위 결정에 있어 핵심적인 참고 자료로 활용될 수 있다. 특히, 영향도가 높게 나타난 변수는 인구이동 예측과 관련된 핵심 요인으로 간주되어, 해당 요인에 대한 정책적 개입이나 자원 배분의 우선순위를 조정하는 데 기여할 것으로 기대된다. 또한, 민감도 분석을 통해 특정 변수 변화가 예측값에 미치는 영향 범위를 파악함으로써, 정책 실행 시 발생할 수 있는 파급효과와 잠재적 위험 요소를 사전에 식별할 수 있다.

[그림 4-15]는 SHAP 기법의 예시로, x축은 0을 중심으로 예측값을 증가하거나 미치는 양의

영향과 음의 영향으로 나뉘고, y축은 중요도가 높은 순서대로 나열된 변수를 의미한다. 붉은색과 푸른색으로 표현되는 것은 해당 변수가 종속변수에 미치는 영향력이 증가(붉은색) 또는 감소(푸른색)의 역할을 의미하고, 0을 중심으로 산포되어 있는 각 점은 모형에서 관측된 종속변수에 대한 각 독립변수 분포(observation)를 의미한다. 예를 들어, 제일 상단에 자리를 잡은 첫 번째 변수인 LSTAT값은 모델에서 그 중요도가 가장 높음을 의미하며, 좌우로 붉은색과 푸른색으로 펼쳐진 산포도의 의미는 모델의 증가하거나 감소한 결괏값에 미치는 영향이 높으나, 붉은색이 집중된 음의 영역에서 특징은 높은 LSTAT값이 많을수록 결괏값의 감소에 영향을 주는 것으로 유추해 볼 수 있다. 또한, x축을 따라서 산포되는 형태가 아닌 y축 방향으로도 분산된 점들의 분포는 해당 x축의 지점에서 분산된(abundant) 관측값의 분포를 설명할 수 있다.

제5장
AI기반 인구이동 예측 모형 시범 분석

제 1 절 개요

본 연구에서 제시한 AI-PriSM(가칭) 모형을 기반으로 한 인구이동 예측모형과 직전 주소지 추정 모형을 시범 분석하였다.

인구이동 예측 모형은 현재 시점의 인구 및 정주 여건을 학습한 뒤, 장래에 발생할 수 있는 인구이동을 예측하는 모형으로 구성하였고, 공간 단위는 1km 격자로 추진하였다. 본 연구에서 검토한 XGBoost, 강화학습(Reinforcement Learning), 다층인공신경망(Multilayer Neural Network)을 비교모형으로 선택하여 활용하였다. XGBoost는 분류 기반의 대표적인 머신러닝(ML) 모델 중 하나라는 점에서 선택되었고, 다층인공신경망은 비선형 데이터에 특화된 딥러닝 및 인공신경망(ANN) 구조의 대표성을 가진다는 점에서 선택되었다. 마지막으로 강화학습은 기존의 머신러닝 분류에서 독특한 학습 방식과 우수한 예측력을 나타낸다는 점에서 시범적으로 본 연구에 선택되어 사용되었다.

직전 주소지 추정 모형은 2001년부터 2024년까지 읍면동 단위로 집계된 데이터를 기·종점 단위로 변환하여 인구이동량(flow)을 학습한 ST-GRAT 모형을 사용하였다. ST-GRAT 모형은 그래프 네트워크를 적용한 혼합(hybrid) 방식의 딥러닝 모델로, 직전 주소지 추정과 같은 이동 쌍 간의 이동량 예측에 적합한 접근으로 판단되어 본 시범 분석에 활용되었다.

각종 시계열 데이터는 연도 및 1km 격자 공간으로 재취합하여 각 모형의 학습에 활용되었고 RMSE, MAE, MSE 등 본 연구에서 선택한 효과척도(MOE)를 기반으로 모형이 구동되었다. 다만, 본 연구에서 수행하고자 하는 AI-PriSM(가칭)의 프레임워크 및 방법론의 시범 분석이라는 점을 고려하여 각 모형의 학습은 간단하게 구성하여 활용하였으며, 자세한 내용은 이후 모형의 구동 과정에서 다시 기술하기로 한다.

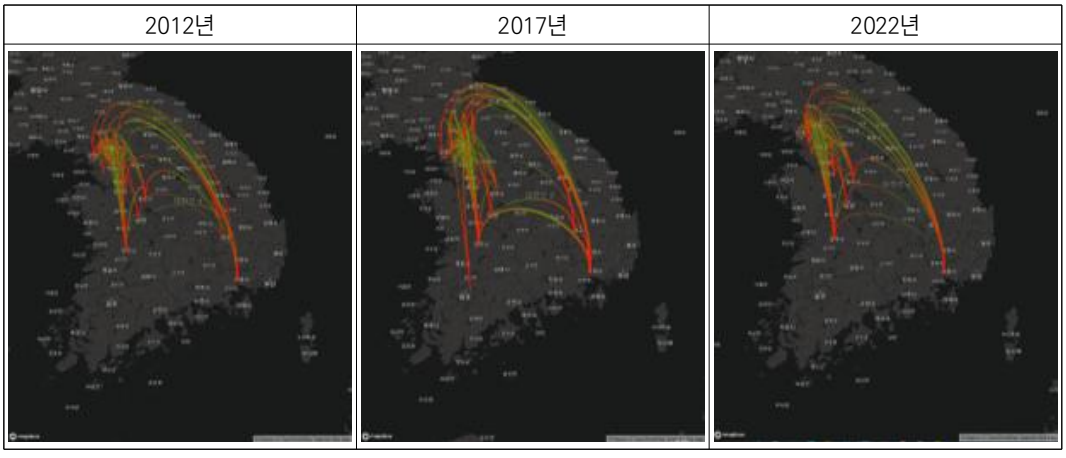
제 2 절 데이터 전처리

1. 인구이동 데이터

1km격자 단위로 치환된 전 국토의 격자는 107,442개로, 각 격자마다 AI-PriSM(가칭)에서 다루고자 하는 사회경제지표, 교통, 토지 이용패턴 등을 종합적으로 배분하여 데이터세트를 구성하였다. 본 절에서는 데이터세트 구성과정에서 인구, 교통 등 몇 개의 데이터군에 대해서 선별적인 시각화와 설명을 기술하고자 한다.

[그림 5-1]은 주요 연도별 인구이동 패턴 중 상위 30개 지역을 선별하여 도식화한 내용으로, 연도에 따라 수도권으로 집중되는 패턴은 유사하나, 연도에 따라 지역별 차이가 간헐적으로 발생하는 것을 확인할 수 있다.

[그림 5-1] 주요 연도별 주요 인구이동 패턴(Top 30 지역) 시각화



출처: 저자 작성

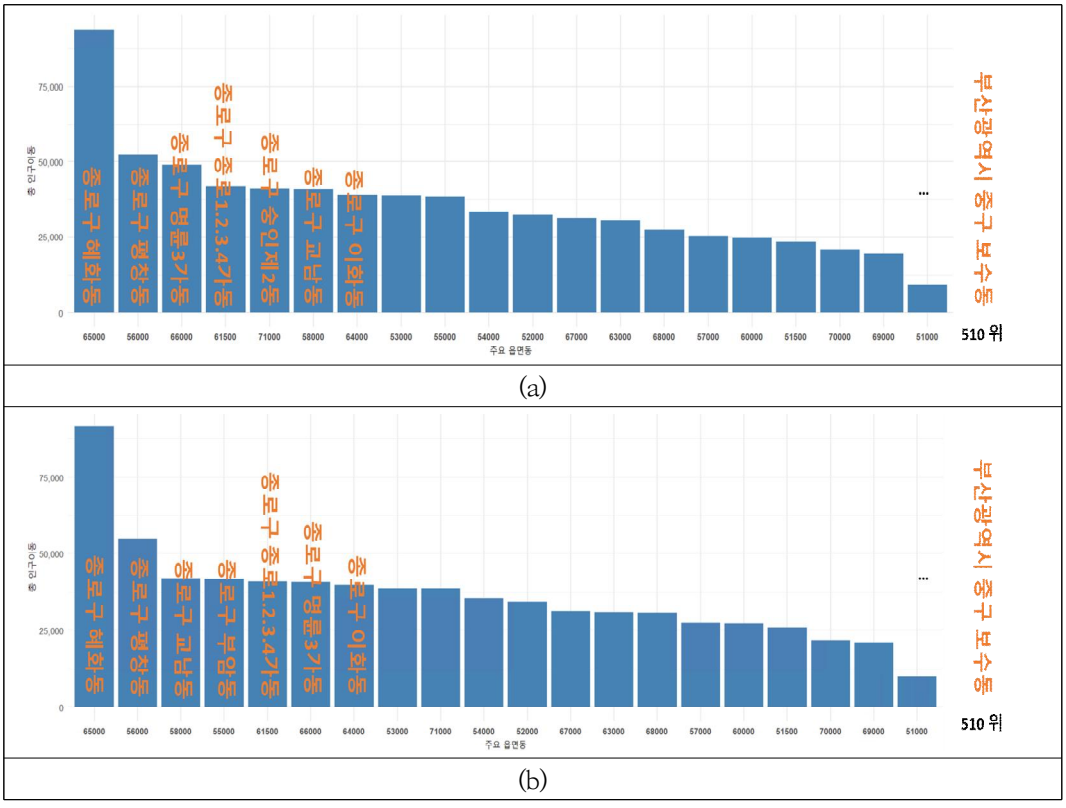
[그림 5-2] 정보 누락 연도의 보간법 적용 예시

출처: 저자 작성

한편, 2001년부터 2015년까지는 5년 단위로 격자별 인구정보가 구축되어 있다는 점을 고려하여 중간 연도인 2002년부터 2004년, 2006년부터 2009년, 2011년부터 2014년의 값들은 앞뒤 연도의 데이터를 활용하여 보간법(interpolation)을 이용한 보정을 수행하였다(그림 5-2 참조).

지난 20여 년 동안의 읍면동 단위 인구이동 데이터에서 확인된 주요 패턴을 살펴보았다. 전입 지역의 주요 지역은 서울특별시 종로구 혜화동, 평창동, 명륜 3가동 등의 순서로 집계되었으며, 서울을 제외한 지역 중 첫 번째 지역으로는 510위에 부산광역시 중구 보수동이 확인되었다. 주요 전출 지역도 전입 지역의 패턴과 크게 다르지는 않았다. 서울특별시 종로구 혜화동, 평창동, 교남동 순으로 주요 전출 지역이 확인되었으며, 부산광역시 중구 보수동이 510위에서 처음으로 서울 이외 지역으로 확인되었다. 전출과 전입 모두 1위와 2위의 차이가 크게 확인되어 해당 지역의 전출입이 두드러졌음을 확인할 수 있다(그림 5-3 참조).

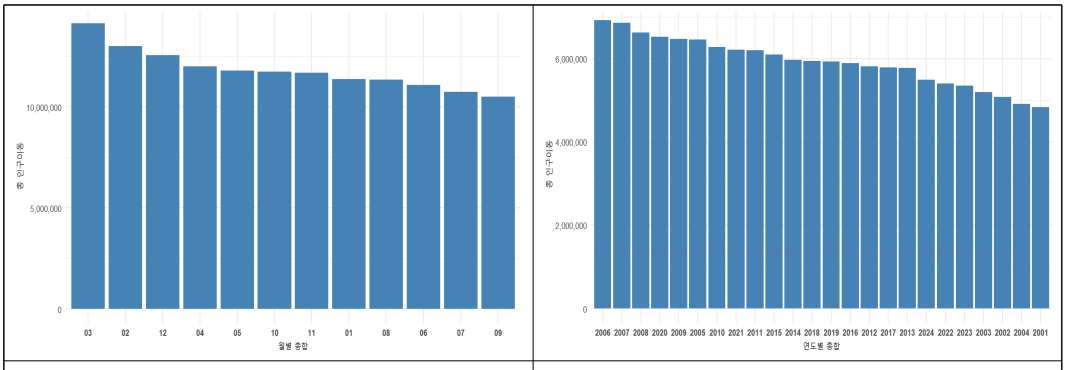
[그림 5-3] 주요 전입 읍면동(a)과 전출 읍면동(b)



출처: 국가데이터처의 인구이동 통계를 바탕으로 저자 작성

시기적으로는 3월에 인구이동이 가장 많았던 것으로 나타나며, 그 이후로는 2월, 12월, 4월, 5월 순으로 집계되었다. 또한 지난 2001년부터 2024년까지 가장 많은 인구이동이 발생했던 시기는 2004년으로 집계되었으며, 2007년이 두 번째, 2008년, 2020년 등의 순서로 집계되었다(그림 5-4 참조). 주요 전출 사유로는 주택, 직업, 가족 및 기타, 교육 순으로 집계되었으나 주택, 직업, [가족-기타-교육]의 구분이 대다수의 전출 사유로 집계된 것으로 미루어 볼 때 인구이동과 관련한 특징은 확인하기 쉽지 않았다(그림 5-5 참조).

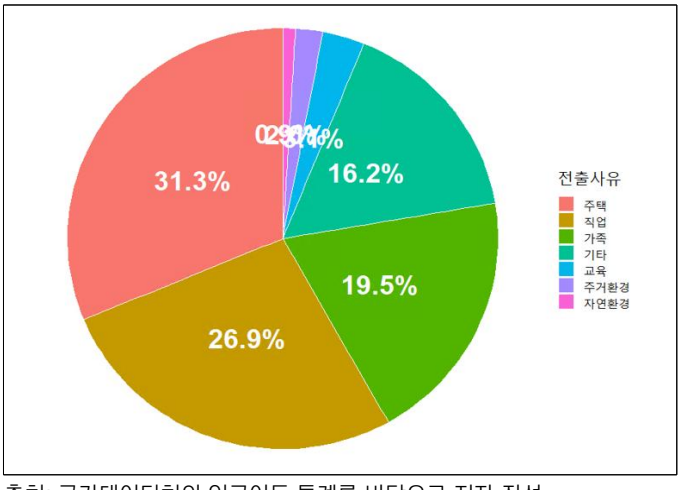
[그림 5-4] 최다 인구이동 월(a)과 최다 인구이동 연도(b)



(a) (b)

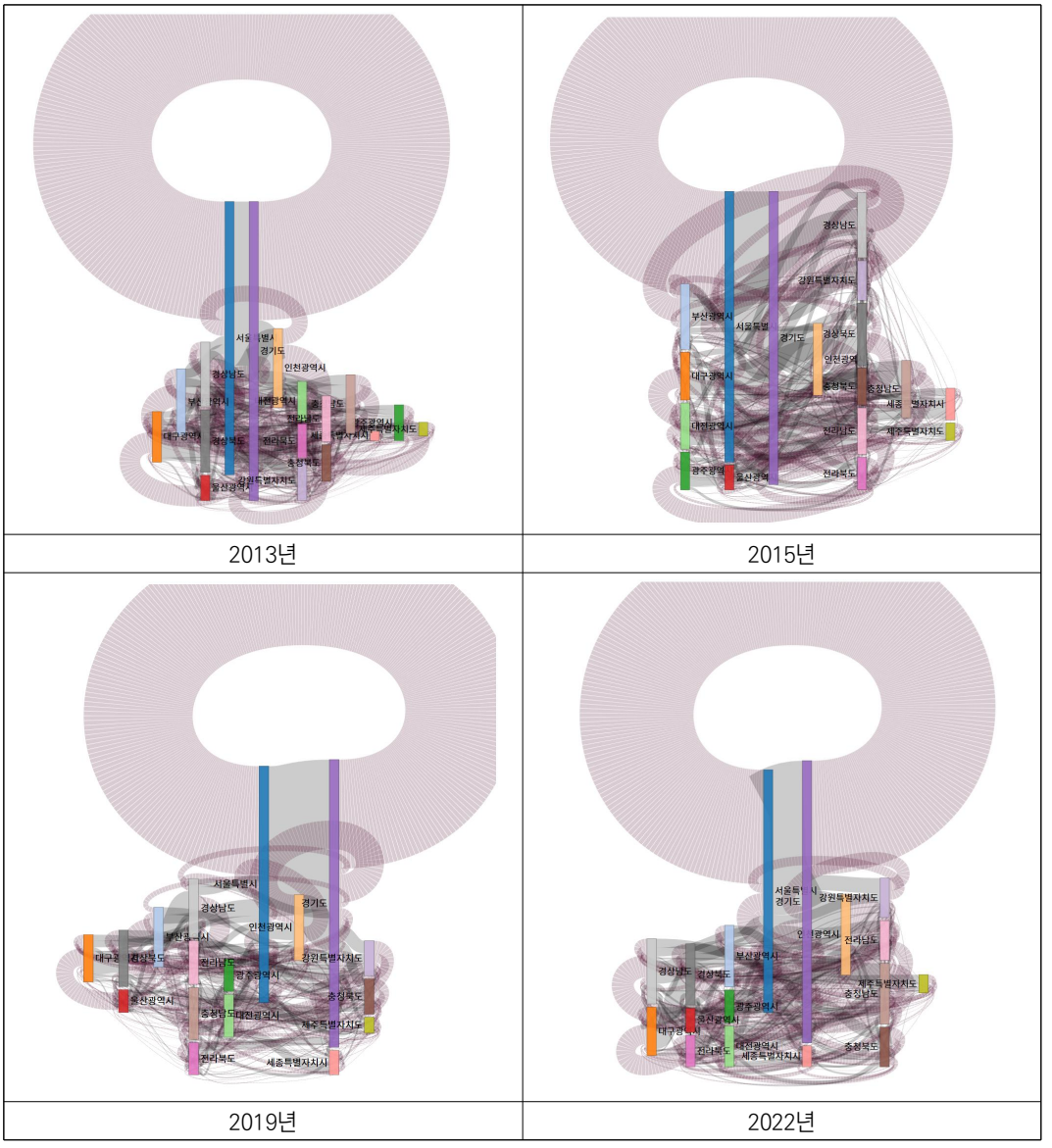
출처: 국가데이터처의 인구이동 통계를 바탕으로 저자 작성

[그림 5-5] 주요 전출 사유 분포 시각화



출처: 국가데이터처의 인구이동 통계를 바탕으로 저자 작성

[그림 5-6] 주요 연도별·시도별 인구이동 시각화



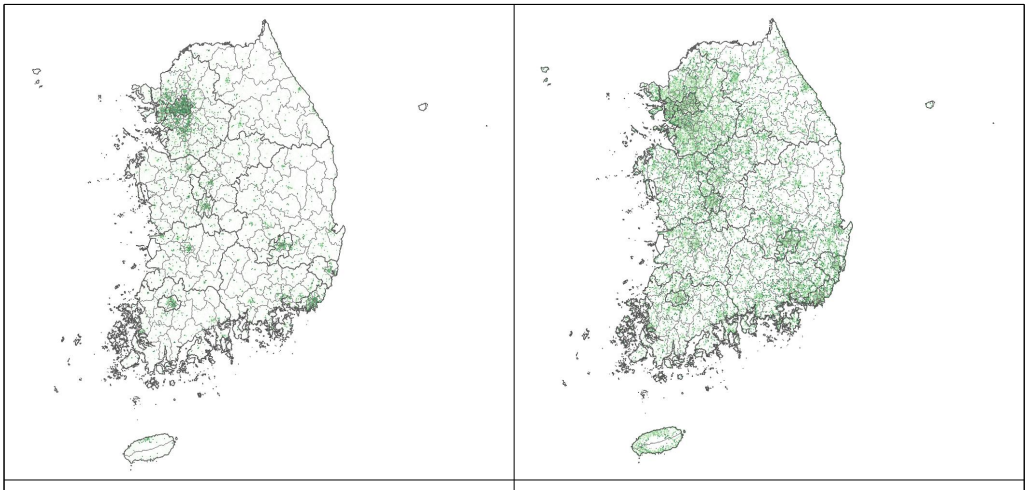
출처: 국가데이터처의 인구이동 통계를 바탕으로 저자 작성

[그림 5-6]은 주요 연도별·시도별 인구이동 패턴을 도식화한 것으로, 바(bar)의 각 색은 각 시도를 의미하며 각 바의 왼편으로는 해당 시도에서 발생한 총전출인구를 오른편으로는 해당 시도로 전입한 총전입인구를 나타낸다. 연도별로 시군구별 인구이동의 양에 어느 정도 차이는 있으나, 서울(파란색) 및 경기도(보라색) 간의 인구이동량이 월등하게 높게 나타나 수도권 중심의 인구이동 패턴을 한눈에 확인할 수 있다.

2. 기타 데이터

지방인허가 데이터에서 확인되는 각종 산업분류별 위치정보를 1km 격자에 배분하여 활용하였다. 등록 일자를 기준으로 연도별 구분을 하였는데, 2018년부터 최근까지 다양한 분포가 확인되었으며, 대체로 2018년 이전의 데이터에 관해서는 확인이 어려워 모형의 학습 과정에서 과적합 및 과소식별의 어려움이 예상되었다.

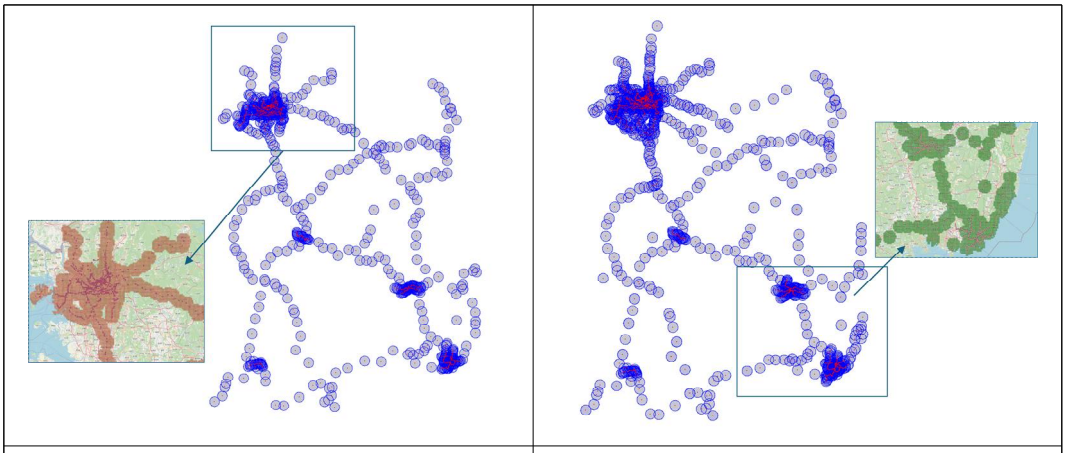
[그림 5-7] 지방인허가 데이터 예시: 보건의료(a)와 음식점(b)



출처: 지방인허가 데이터를 바탕으로 저자 작성

주요 열차역 및 지하철역 위치에 5km 버퍼를 설정하여 1km 격자에 배분하였다. 수도권 및 도심지의 경우 거의 토지 전체가 영향권에 포함되는 것으로 나타났고, 비수도권 및 지방에서는 그 범위가 역사 주변에만 간헐적으로 포함되는 것으로 나타났다. 2010년과 2023년을 시각적으로만 비교해도 대도시권은 그 범위가 더욱 확장되는 반면, 비수도권 및 지방은 열차 접근성에 개선이 확인되나 그 범위는 여전히 일부 지역에만 국한하는 것으로 나타난다.

[그림 5-8] 주요 열차역·지하철역 위치 및 5km 버퍼 시각화 예시: 2010년(a)과 2023년(b)



출처: KTDB 데이터를 바탕으로 저자 작성

마지막으로, 본 시범 분석에서는 각종 데이터의 이격 범위를 고려하여 Min-Max Normalization를 통한 일반화 과정을 거쳤다. Min-Max 일반화 방법은 머신러닝/딥러닝 등 AI기반의 모형 구동뿐만 아니라 전통적인 통계분석에도 널리 활용되는 방식으로, 입력 데이터 범위의 이격으로 발생할 수 있는 데이터의 편향성과 모델의 영향을 최소화하여 입력 데이터 분포를 고르게 일반화하는 데 있다. Min-Max 일반화를 거친 모든 값은 0부터 1 사이의 값을 갖게 된다.

$$X_{normalization} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

제 3 절 비교모델군 선정

모형의 비교 및 판단을 위해서 네 가지 모델을 선별적으로 활용하였다. 먼저 기본모델(baseline)은 단순 과거 추세를 연장한 방법으로, '현재까지 제공된 정보가 앞으로도 계속된다'는 사실을 가정으로 연도별·격자별 목푯값의 R^2값이 최소가 되는 최적의 회귀식을 도출한 뒤 사용하였다. 두 번째로 xGBoost 방법을 활용하였는데, 본 시범 분석의 실험 결과 랜덤 포레스트(Random forest) 방법론은 각종 데이터 조합에 상당한 학습 시간이 요구되어 상대적으로 계산 속도가 빠른 xGBoost 방법을 활용하였다. 세 번째는 강화학습(Reinforcement Learning)으로, 격자별로 객채(agent)를 구성하여 324개에서 약 600개의 격자별 시공간 환경을 학습하도록 구성하였다. 107,442개 객체의 환경에 따른 보상(reward)는 타깃값과 예측값 간의 차가 적을수록 높은 보상을, 차이가 클수록 상대적으로 낮은 보상을 가질 수 있도록 구성하여 학습하였다. 마지막으로 다층인공신경망(Multilayer Neural Network)을 활용하였다. 전통적인 인공신경망 모형에서 역전파전이(backpropagation) 방식의 학습을 구성한 뒤, 은닉층(hidden layer)을 추가로 배치하여 [입력층 1 - 은닉층 3 - 출력층 1]의 방식으로 아키텍처를 구성하였다. 과적합(overfitting) 방지를 위해서 20%의 파라미터 생략(dropout)과 대략 200회의 iteration 학습 및 relu 활성화 함수 선택을 통한 최적화를 구성하였다.

특히, xGBoost에서는 Tabu-search 최적화 방법을 통해 최적의 분기점과 샘플링 수를 결정할 수 있도록 구성하였고, 강화학습에서는 확률적 탐색 및 Hill-climbing을 통한 결정계수의 최적화를 설정하였다. 모든 모형의 학습에 사용된 목적함수는 절대평균오차(MAE)를 이용하여 0 이하의 값도 면밀하게 검지하고 오차를 줄일 수 있도록 구성하였다.

직전주소지 예측모형인 ST-GRAT에서는 유전자 알고리즘(GA)을 통한 최적 파라미터를 탐색 및 도출하였는데, 공간 인코더는 GAT 모형을, 시간 인코더는 LSTM 모형을, 예측 단계에서는 다층인공신경망을 채택하여 구성하였다. 추가로 256 은닉면을 중심으로 0.001 학습률, 150 분할(epoch) 학습 및 LSTM 단계에서는 3년 주기(look-back interval)의 과거 데이터 학습을 활용하여 모형으로 구축하였다.

제 4 절 시범 분석 결과

〈표 5-1〉과 [그림 5-9]는 목표 연도별로 각 예측모형의 성능을 비교한 결과로, 학습 데이터의 구성에 따라 훈련 데이터세트와 목표 데이터세트 간의 오차를 최소화할수록 모델의 성능이 우수하다고 볼 수 있다. 예를 들어, 2030년을 예측하는 경우와 같이 2000~2017년 인구 데이터를 기반으로 6년 이후인 2023년을 예측하도록 구성된 방식은 비교적 장기 예측에 해당하므로, 근시일내를 예측하는 모델보다 오차가 증가하는 것은 자연스러운 결과로 해석할 수 있다.

〈표 5-1〉 예측 목표연도에 따른 각 인구이동 예측모형의 예측 결과 비교

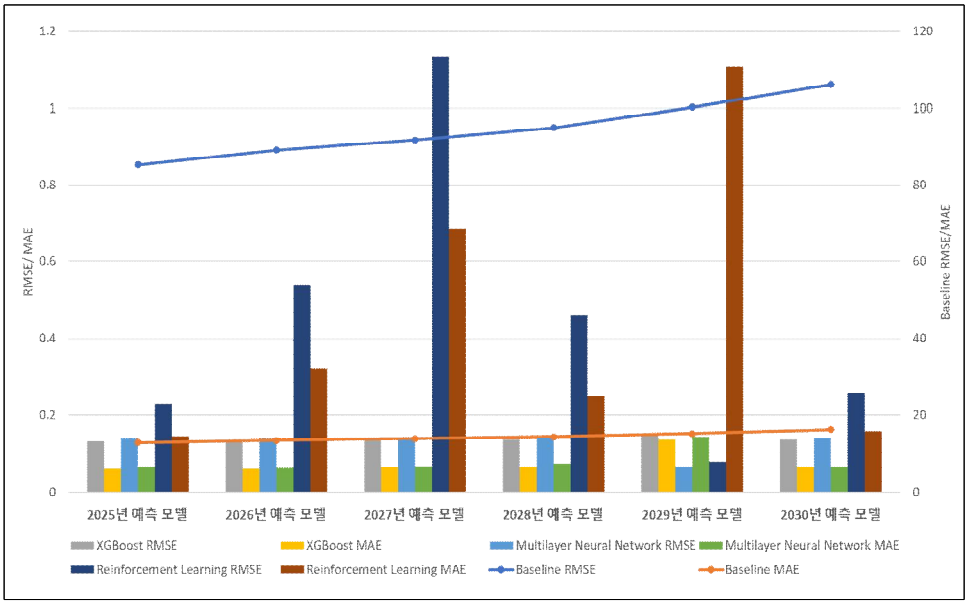
예측모형	과거추세연장법		XGBoost		다층인공신경망		강화학습	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
2025년 예측	85.3671	12.7682	0.1310	0.0621	0.1376	0.0671	0.2298	0.1446
2026년 예측	89.1269	13.3555	0.1313	0.0624	0.1379	0.0639	0.5366	0.3196
2027년 예측	91.7444	13.8264	0.1339	0.0643	0.1395	0.0678	1.1347	0.6852
2028년 예측	94.9971	14.3344	0.1365	0.0661	0.1419	0.0742	0.4609	0.2481
2029년 예측	100.2281	15.1339	0.1363	0.0668	0.1432	0.0790	1.1078	0.4181
2030년 예측	106.1129	16.0948	0.1368	0.0668	0.1387	0.0667	0.2590	0.1574

출처: 저자 작성

예측 목표연도별 결과를 살펴보면, 과거추세연장법 대비 XGBoost, 다층인공신경망, 강화학습 모두 RMSE와 MAE 지표 모두에서 월등히 우수한 성능이 확인되었다. 특히, XGBoost와 다층인공신경망은 그 오차가 매우 유사하게 나타났다. 비선형 데이터에 강한 다층인공신경망에 견줄 만큼 XGBoost의 RMSE와 MAE 오차도 두드러지게 예측되었는데, 이는 Cellular Geography 방식으로 구성된 데이터를 바탕으로 각 모델 아키텍처의 최적화된 하이퍼파라미터 조합이 비선형적인 시공간 패턴과 인구구조를 효과적으로 학습한 결과로 유추해 볼 수 있다. 또한, 장기 예측일수록 RMSE와 MAE 모두에서 정확도 저하가 나타났으나, 그 차이가 크지는 않았다. 강화학습의 경우 Baseline인 과거추세연장법보다는 월등한 성능을 나타냈으나, 기타 비교모형인 XGBoost와 다층인공신경망에 비해서는 그 정확도가 다소 떨어지며 시계열에 따른 일관성도 다소 불안정하게 확인되어 적절한 학습이 이루어지지는 못한 결과로 해석된다.

이러한 결과에 대해서는 몇 가지 해석이 가능하다. 첫째, 인구가 '0'으로 예측되는 격자의 비율이 높아 평균값 중심의 성능 지표에 영향을 미쳤을 가능성이 있다. 그러나 일부 격자는 인구 변화가 미미하거나 소규모 인구가 서서히 증가·감소하는 특성을 보이며, 이에 따라 특정 구간에서 예측 패턴이 달라지는 양상이 나타났으므로 이에 대한 후속 분석이 필요하다.

[그림 5-9] 모형별 RMSE와 MAE 비교

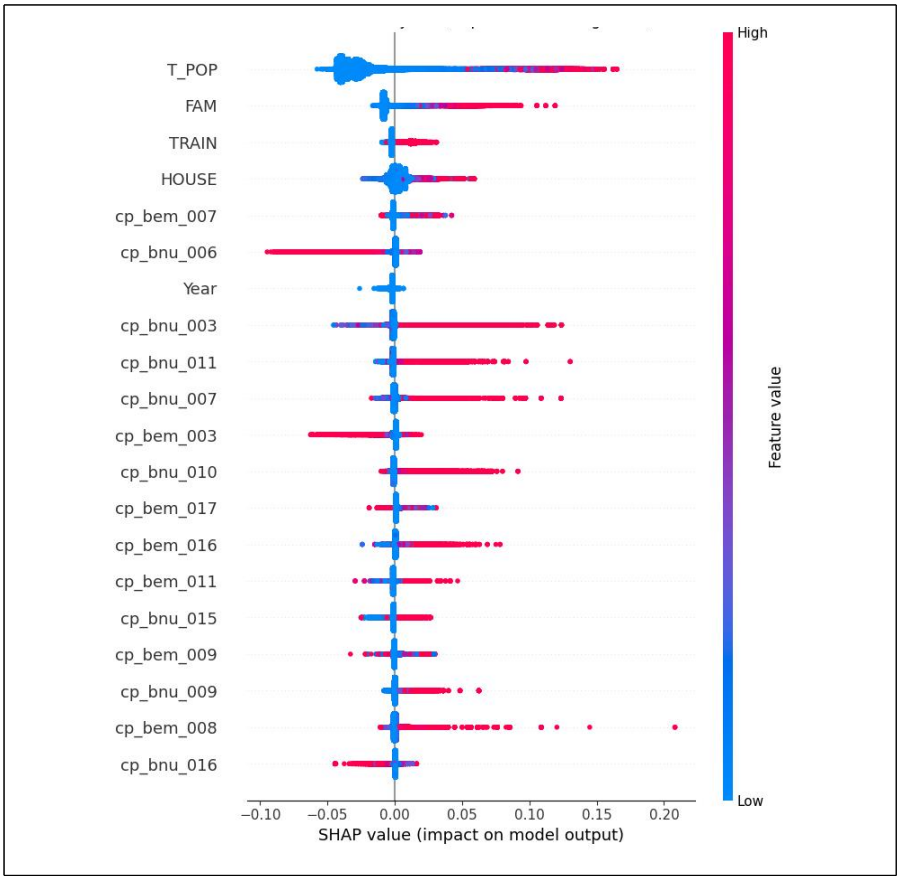


출처: 저자 작성

변수 중요도에서는 인접 격자의 인구정보가 가장 중요한 근거가 되는 것으로 나타났는데, 변수별로 객관적이고 전반적인 비교를 위해서 격자 정보를 제외한 변수별(group) 변수 중요도를 산출하여 확인하였다(그림 5-10 참조). 변수 민감도에서 가장 높은 비중을 차지하는 변수는 총인구(T_POP)로 확인되었고, 뒤를 이어 총가구수(FAM), 철도 및 지하철 접근성(TRAIN), 총주택수(HOUSE), 도소매업 종사자수(cp_bem_007), 건설업 사업체수(cp_bnu_006) 순으로 집계되었다. 특히, 인구가 많을수록(x-축에서 오른쪽) 인구이동에 있어 증가(붉은색)의 영향이 강한 것으로 나타났고, 적을수록(x-축에서 왼쪽) 인구이동의 감소(푸른색) 영향이 강한 것으로 나타났다. 특히 인구(T_POP) 변수 왼편에 집중되어 군집화된 결과는 인구이동 예측에 있어서 소수의 인구그룹이 모델에 분산되는 영향까지 미치고 있음을 나타내고 있다. 총가구수(FAM) 또한 이와 유사한 영향을 나타냈으며, 특히 특정 가구수 단위에서는 위-아래로 분산되는 독특한 음의 영향을 보이는 것으로 나타났다.

철도 및 지하철 접근성(TRAIN)에서 SHAP의 중심축인 0을 기준으로 모형의 양적인 영향과 음의 영향이 두드러지게 구분되는 현상은, 앞선 철도 및 지하철 버퍼 시각화에서 확인된 것과 같이 수도권 및 도심지와 지방 및 비도심지 간의 차이가 극명하게 구분됨을 나타낸다. 건설업 사업체수(cp_bnu_006)와 광업 종사자수(cp_bem_003)가 기타 변수와 반대의 경향을 보이는 현상도 독특한데, 이는 건설업 사업체수와 광업종사자수가 적은 지역일수록 인구이동에 주는 영향이 높음을 의미해, 도심지로 집중되는 이촌향도의 패턴을 보여주는 현상이라 할 수 있다.

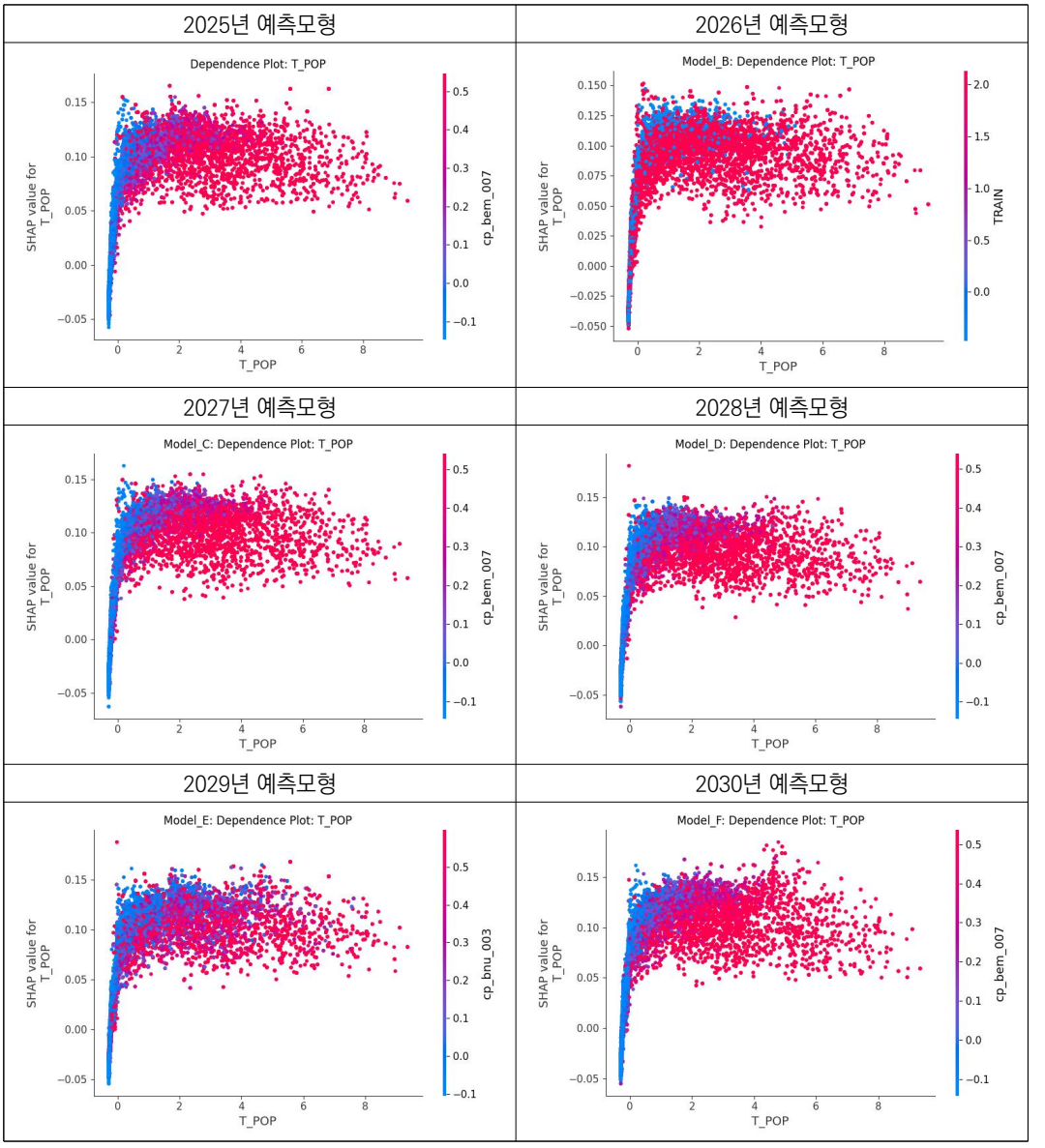
[그림 5-10] SHAP 도식화(2025년 예측모형 예시, 상위 20개 변수)



주: 각 변수는 모두 격자별 집계 정보임. T_POP는 총인구, FAM은 총가구수, TRAIN은 철도/지하철 접근성, HOUSE는 총주택수, cp_bem_007은 도소매업 종사자수, cp_bnu_006은 건설업 사업체수, cp_bnu_003은 제조업 사업체수, cp_bnu_011은 금융 및 보험업 사업체수, cp_bnu_007은 건설업 도매 및 소매업 사업체수, cp_bem_003은 광업 종사자수, cp_bnu_010은 통신업 사업체수, cp_bem_017은 보건업 및 사회복지 서비스업 종사자수, cp_bem_016은 교육 서비스업 종사자수, cp_bem_011은 금융 및 보험업 종사자수, cp_bnu_015은 교육 서비스업 사업체수, cp_bem_009은 숙박 및 음식점업 종사자수, cp_bnu_009은 운수업 사업체수, cp_bem_008은 운수업 종사자수, cp_bnu_016은 보건 및 사회복지사 사업체수 등임

출처: 저자 작성

[그림 5-11] 예측 목표연도에 따른 각 모형의 인구에 따른 기타 주요 변인과의 SHAP 비교



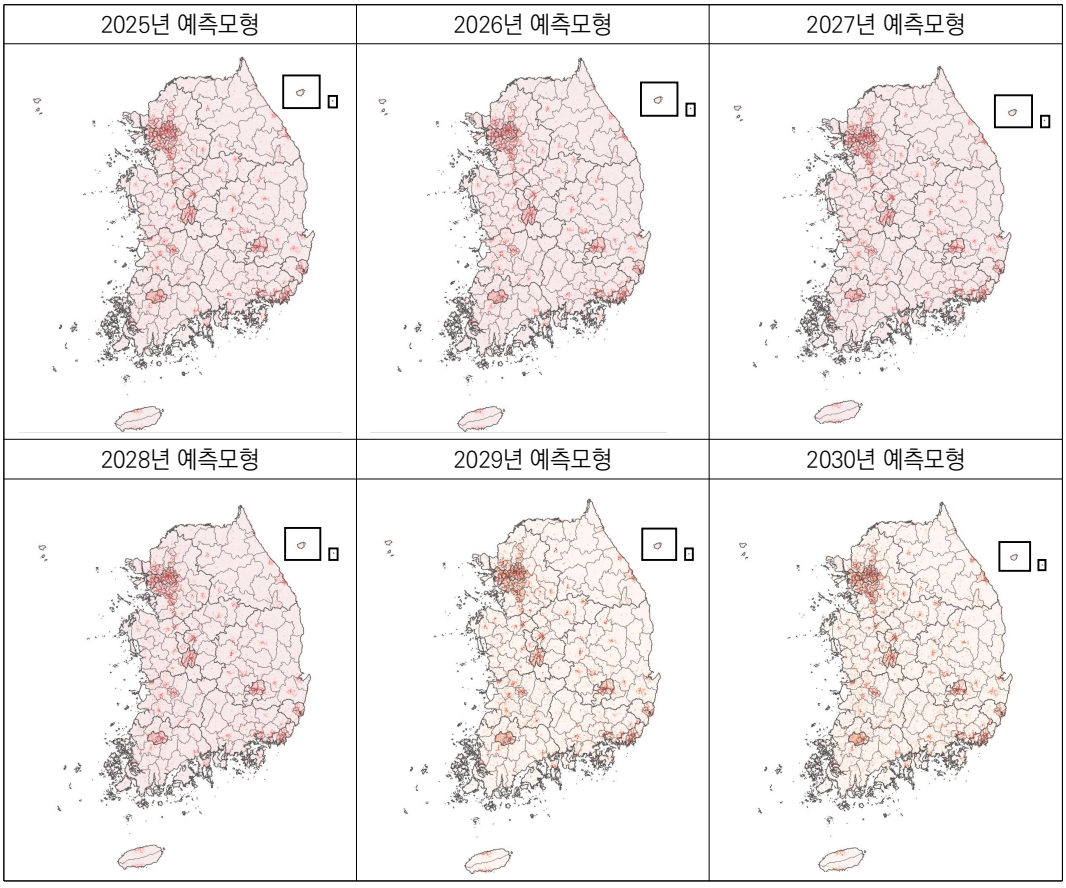
주: 각 산포도의 X축은 총인구 변수의 정도를, Y축은 연도별 모형의 주요 변인을 의미
출처: 저자 작성

인구 변수의 강한 종속성은 연도별 모형 모두에서 두드러지게 확인되었다. [그림 5-11]은 인구 변수와 차순위 주요 변수 간의 SHAP 시각화로, 주요 변수였던 도소매업 종사자수, 철도/지하철 접근성, 제조업 사업체수 등과의 관계를 나타낸다. 철도/지하철 접근성을 제외하면 일정 수준의 인구규모 이상에서 각종 변수 간의 높은 영향이 확인되었다. 철도/지하철 접근성이 낮은 인구밀

도에서도 양의 영향을 나타내는 점은, 수도권 지역을 중심으로 확산하고 있는 1인 가구 등의 사회현상과도 높은 연관성이 있을 것으로 유추해 볼 수 있는데, 이는 추가연구를 통한 확인이 필요하다.

[그림 5-12]는 예측 연도별 인구이동 예측 도식화 결과로, 색의 농도는 각 격자로 배분된 인구이동의 많고 적음을 나타낸다. 예측 시점이 2030년까지 단기라는 점을 고려하면, 연도에 따라 단기 시점에서는 큰 차이가 발생하지 않았음을 알 수 있다. 또한 읍면동 단위로 집계된 이동인구를 1km 격자 단위로 배분하는 단계에서 소수점으로 환산되는 과정 때문에 인구 이동의 증감이 공간적으로도 큰 특이점을 확인할 수 없는 것으로 보인다.

[그림 5-12] 예측 연도에 따른 인구이동 패턴



주: 색의 농도는 인구이동의 많고 적음을 의미
출처: 저자 작성

〈표 5-2〉 예측 목표연도에 따른 직전 주소지(읍면동) 예측 결과 비교

예측모형	과거추세연장법		ST-GRAT		
	RMSE	MAE	Training Loss(MSE)	RMSE	MAE
2025년 예측	7.9532	1.5788	23.3369	23.0948	2.9677
2026년 예측	9.6988	1.7440	23.9456	23.0961	2.9797
2027년 예측	10.4888	1.8240	22.3817	23.1019	2.9941
2028년 예측	10.9135	1.8646	20.9842	23.1087	3.0109
2029년 예측	12.6732	1.9828	19.5214	23.1164	3.0328
2030년 예측	14.0870	2.0995	18.1139	23.1263	3.0549

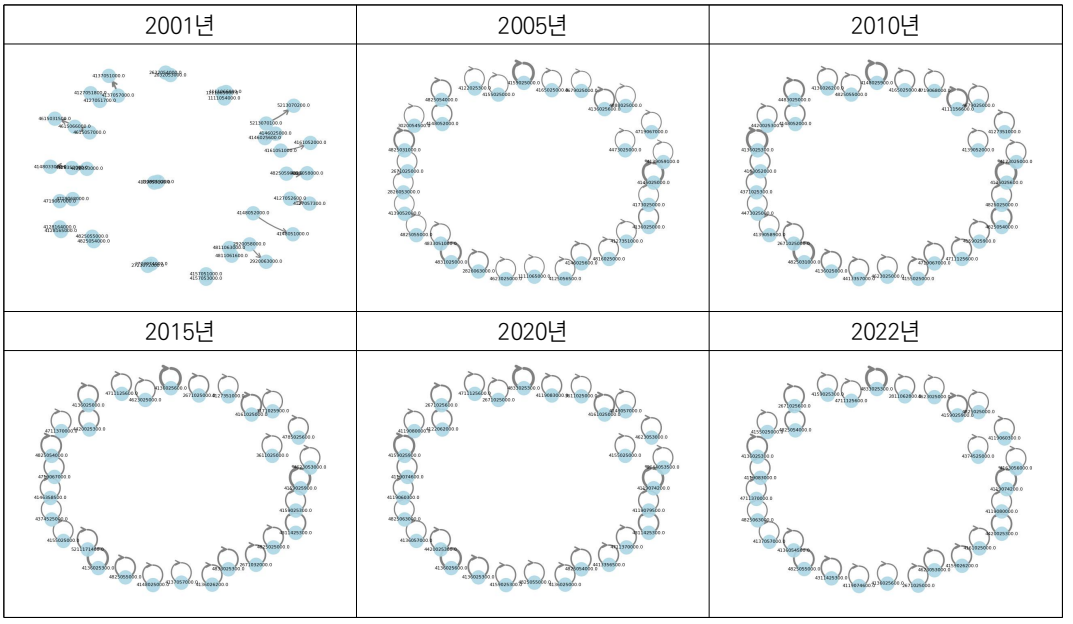
출처: 저자 작성

〈표 5-2〉는 예측 목표연도에 따른 직전 주소지(읍면동) 예측 결과를 표로 정리한 내용이다. 인구이동 예측모형과 마찬가지로 과거추세연장법을 기준 모델로 제시하였고, 본 연구에서 시범적으로 적용하는 ST-GRAT 모형을 함께 비교하였다.

앞선 인구이동 예측모형과는 달리, 과거추세연장법이 ST-GRAT 모형보다 RMSE와 MAE에서 우수한 결과를 나타내었는데, 이 현상은 [그림 5-13]과 [그림 5-14]를 함께 비교하면 간단하게 납득이 가는 결과이다. 과거 인구이동 패턴의 상위 30 읍면동 지역을 시각화하였을 때 2001년을 제외하면 모두 지역 내 인구이동이 주를 이루는 것으로 확인된다(그림 5-13 참조). 이는 앞선 인구이동 주요 전입지역과 전출지역 비교에서도 그 순위가 크게 달라지지 않았던 현상과 높은 연관성이 있는데, [그림 5-14]의 ST-GRAT에서 학습하여 예측한 결과에서는 기·종점 간 연계와 이동을 집중적으로 학습한 결과가 확인된다(그림 5-14 참조). 즉, ST-GRAT의 메커니즘에서 지역 간 이동에 집중한 모델이 지역 내 이동(self-loop)에 대해서는 적절한 학습이 이루어지지 않은 결과로 해석된다. 이는 향후 연구에서 본 시범 분석에서 확인한 ST-GRAT의 한계점을 숙지하여 지역 내 이동을 함께 학습할 수 있는 모형을 개발할 필요가 있다.

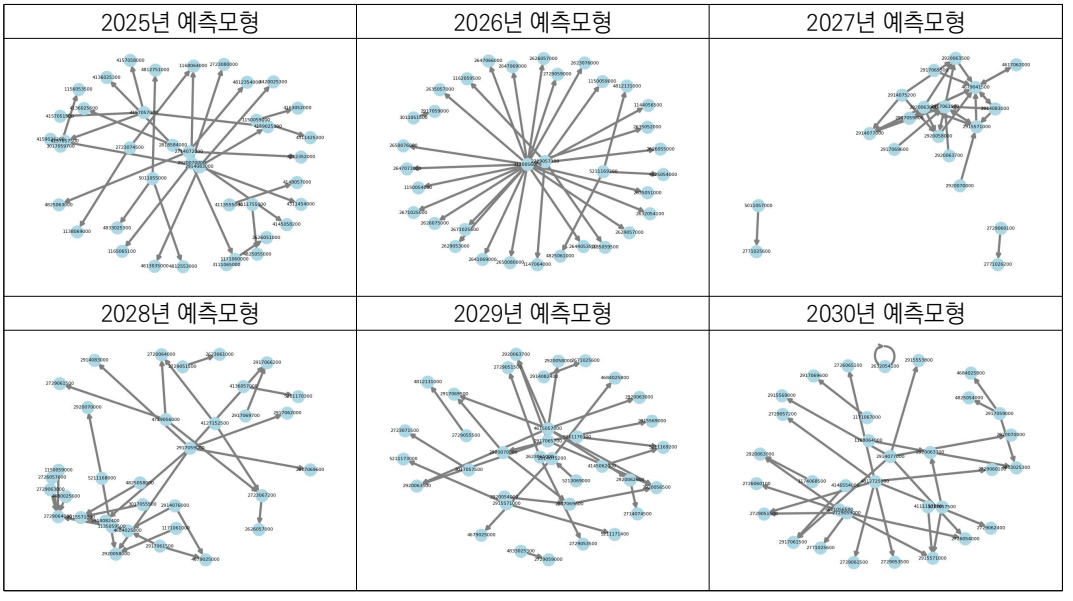
특이한 점은, 연도의 변화에 따라서 어느 정도 RMSE와 MAE 값이 달라지는 과거추세연장법과는 달리, ST-GRAT에서는 RMSE와 MAE 값의 큰 변화가 확인되지 않아, 인구이동의 패턴만큼은 적절하게 학습한 결과로 유추해 볼 수 있다.

[그림 5-13] 과거 연도 인구이동 패턴 그래프 시각화(상위 30 읍면동 지역)



출처: 저자 작성

[그림 5-14] 예측 연도에 따른 인구이동 패턴 그래프 시각화(상위 30 읍면동 지역)



출처: 저자 작성

제 5 절 정책 시나리오 구성 및 적용 방향

1. 정책 시나리오 구성 방향

정책 시나리오 분석은 AI-PriSM(가칭) 예측모형의 실질적 정책 연계성을 검증하기 위한 핵심 절차일 것이다. 본 시범 분석에서 확인된 내용을 바탕으로, AI-PriSM(가칭) 고도화 과정에서 정책 관련 시나리오를 구성하고 다양한 인과관계를 시험할 필요가 있다. 다음은 고도화 과정에서 시나리오 구성 시에 고려해야 할 방향성을 개괄적으로 기술하였다.

고도화 과정에서 시나리오를 구성하고자 할 때, ① 변수 조정의 범위, ② 정책 실행 기간, ③ 지역별 적용 차등의 세 가지 축을 중심으로 설계할 필요가 있는데, 시나리오 구성은 다음과 같이 구체화할 수 있다.

- 기준 시나리오(Baseline): 현재 정책 및 환경 변수를 유지한 상태에서의 예측값
- 개입 시나리오(Policy Intervention): 핵심 변수(예: 산업특화 시나리오, 택지개발 시나리오, 출생률 변화 시나리오 등)를 변화시켜 가상의 정책 개입에 따른 지역별 인구이동 효과를 반영
- 복합 시나리오(Mixed Intervention): 2개 이상의 개입 시나리오의 요소를 결합하여 다차원적인 정책 효과를 검토하고 구체적인 영향을 확인

각 시나리오는 AI-PriSM(가칭) 모형에 동일한 구조로 적용하되, 변수 조정 전후의 결괏값을 비교함으로써 시·공간상으로 확산하는 차이를 바탕으로 실질적인 정책 효과를 정량적으로 산출할 수 있다는 특징이 있다. 이러한 접근방식은 기존에 정성적으로만 이루어지던 전문가 자문이나 시나리오 플래닝(Scenario Planning), 델파이 기법(Delphi Technique), 시스템 다이내믹스(System Dynamics) 등의 방법론에 증거 기반의 시뮬레이션 방법론이라는 새로운 가능성의 축을 더할 수 있을 것이다.

특히, 데이터 기반 시뮬레이션 과정을 통해 단일 예측에 의존하기보다는 다수의 가능성을 고려한 접근방법으로 위험을 최소화하며 미래에 대한 유연한 대처 방향을 제시할 수 있다는 점에서 필요한 부분일 것이다. 더 나아가 전략적인 정책 수립에 있어서 다양한 정책지원 시스템으로 역할을 기대해 볼 수 있다.

2. 정책 시나리오의 적용 방향

이렇게 고도화된 AI-PriSM(가칭)의 시나리오 분석 결과는 다음과 같은 분야에서 직접 활용할 수 있을 것으로 기대한다.

- 도시 계획 – 인구 유입이 집중되는 지역에 기반시설 확충 및 주택 공급 전략 수립
- 교통 인프라 – 전입 증가 지역에 대중교통 노선 확대 및 접근성 개선
- 지역 균형발전 – 특정 지역의 인구 유출 방지를 위한 산업·고용 유치 정책 설계
- 주거·복지 정책 – 인구구조 변화에 따른 맞춤형 주거 지원 및 복지 자원 배분

활용 전략 측면에서는 특정 정책 시행 전 예측 시뮬레이션을 통해 사전 효과 검증을 통한 사전 진단 도구로서의 활용이 가능할 것으로 기대해 볼 수 있다. 물론 이러한 시뮬레이션이 가능하기까지 특정 정책에 집중된(fine-tunning) 모델을 특화할 필요가 있고, 과거에 발생한 적 있는 현상에 대해 집중적인 구조화를 구축할 수 있다면 더욱 정교한 모델로 발전할 수 있을 것이다. 하지만, 주기적인 데이터 업데이트와 정책 방향별로 전략적 모델의 세분화 및 경량화 과정을 통해서 지속 가능한 과학적 정책지원 도구의 노력이 수반되어야 할 것이다. 이렇게 설계되는 AI-PriSM(가칭)은 정책 맞춤형 고도화라는 측면에서 그 효용성이 높을 것으로 기대한다.

제 6 절 소결

본 연구에서 제시한 프레임워크인 AI-PriSM(가칭) 모형의 시범 분석 결과, 전통적인 선형회귀 기반의 추세연장 기법의 인구이동 예측은 안정성 측면에서 한계를 확인할 수 있었다. 반면, 비선형 데이터 구조 처리에 강점을 지닌 다층인공신경망 및 강화학습 등 딥러닝 모델은 전반적으로 우수한 예측 정확도를 보였으며, 중장기 전망에서도 비교적 일관된 성능을 유지하는 것으로 나타났다. 특히 Cellular Geography 패턴으로 학습된 데이터를 통해 선형 중심의 분류 모델인 XGBoost도 우수한 성능을 나타내었다. 직전 주소지 예측 모형에서는 인구이동의 양적인 예측에서는 다소 아쉬움이 남는 결과를 도출하였으나, 읍면동 간의 관계는 두드러지게 학습하여 향후 고도화 과정에서 지역 간 이동뿐만 아니라 지역 내 이동까지 수렴할 수 있는 필요성을 확인할 수 있었다.

한편, 공간적인 예측에서 현재의 수도권 집중형 인구이동 패턴이 가까운 장래에도 변하지 않을 것으로 예측되었으며, 수도권 중에서도 주요 지역의 인구 쏠림 현상은 계속될 것으로 예상된다. 인구이동 예측모형의 요인 민감도 분석에서는 인구의 많고 적음이 곧 인구이동의 주요 요인으로 작용하는 것으로 확인되어, 공간적으로 수도권과 비수도권의 인구집중 현상의 주요 원인으로 유추해 볼 수 있는 부분이다. 그밖에 정주 여건으로 간주해 볼 수 있는 기차 및 지하철 등 대중교통의 접근성 여부, 주요 사업체(직장)의 밀도 여부 등이 주요 요인으로 확인되어 인구이동의 과학적 원인을 유추하기 위해서는 정주 여건의 면밀한 분석이 함께 요구됨을 시사한다.

AI 기반 예측 방법의 장점은 다양한 요인을 다차원적으로 통합 분석할 수 있다는 것이다. 최근 눈부시게 발전하는 컴퓨팅 기술의 도움으로 고려 대상이 될 수 있는 요인의 범주는 계속해서 확장하고 있으며, 하드웨어뿐만 아니라 딥러닝 등 고급화된 AI 알고리즘 또한 이러한 가능성을 견인하고 있다. 본 시범 분석에서는 인구·공간통계, 지리정보, 이동 관련 지표 등을 결합하여 격자 단위 인구변화를 산정함으로써, 향후 정책 수립 시 공간적 타기팅의 정밀성을 높일 수 있는 가능성을 확인하였다. 또한 변수 중요도 분석을 통해 예측에 기여하는 핵심 요인을 식별하여, 정책 설계 시 우선 고려해야 할 요인의 실증적 근거를 제시하였다. 다만, 데이터 최신성과 가용성의 한계, 미래 변숫값의 불확실성, AI 모델의 과적합 가능성 등은 여전히 개선이 필요한 영역이다.

제6장
결론 및 향후 과제

제 1 절 연구 요약 및 핵심 성과

본 연구는 AI-PriSM(가칭, AI-based Policy-oriented Spatio-temporal Prediction for Population Migration) 모형을 기반으로 시공간 격자 단위에서의 인구이동과 직전 주소지 등을 예측할 수 있는 모형을 제시하였다. 인구정보를 비롯하여 정주 여건으로 간주할 수 있는 교통 접근성, 상권, 업종별 직장 밀집도 여부, 종사자 수 등을 다양하게 고려하여 모형에 포함할 수 있도록 구성하였다. 특히 1km 격자 단위로 배분된 각종 정보는 Cellular Geography 기반의 시공간 배분 과정을 거쳐 AI 학습에 최적화할 수 있는 구조로 변환하였고, 지도(Supervised)·강화학습(Reinforcement Learning) 등 다양한 머신러닝 및 딥러닝 등의 AI 알고리즘의 특성을 고려한 모형의 성능 비교와 최적화(Optimization) 등을 병행하였다. 더 나아가 모형의 예측력에 영향을 주는 변수 민감도(SHAP)를 함께 산출하여 향후 정책 시뮬레이션의 활용 가능성을 확인하였다.

본 연구에서 제시한 AI-PriSM(가칭) 모형에 대한 일련의 과정을 거쳐 기존 시군구 기반의 인구이동 추계(Projection)의 한계를 극복할 수 있는 새로운 예측 방법론을 제시하였고, 격자라는 세밀한 공간 단위에서의 분석 프레임워크를 구현하였다. AI-PriSM(가칭) 모형의 프레임워크에서는 격자별로 누적된 비선형 데이터 구조를 각종 AI 모델이 효과적으로 학습하여 중장기 예측의 일관성 및 우수성을 확인할 수 있었다. 직전 주소지 예측 모형에서는 기존 모형으로는 다루기 어려웠던 다수의 기·종점 쌍 간의 관계를 학습하여 예측하는 혼합 딥러닝 기반의 모형인 ST-GRAT을 적용하여 향후 직전 주소지 예측 모형의 발전 가능성을 제안하였다.

본 1차 연도 연구에서 마무리하지 못한 정책 시뮬레이션 부분은 모형의 고도화와 함께 다양하게 시도될 수 있으며, 출생률 및 사망률 변동, 택지개발 등 미래 여건의 개입을 통한 시뮬레이션 또한 본 AI-PriSM(가칭) 모형에서 충분히 가능한 영역이다.

제 2 절 기술적·정책적 시사점

기술적 측면에서 본 연구는 AI와 공간정보 기반 예측모형의 결합을 통해 공간 데이터 활용의 확장성과 정밀도를 높였으며, 다양한 유형의 머신러닝·딥러닝 모형의 시범 분석을 통해 각 모형의 장단점을 통합적으로 검증할 수 있는 구조를 제시하였다. 특히 SHAP 기반의 변수 영향도 분석을 적용함으로써 모델 해석 가능성을 강화하고 '블랙박스(blackbox)'로 우려하던 변수의 민감도 문제를 완화하였으며, Tabu-search, Hill-climbing, 유전자 알고리즘(GA) 등 고급 최적화 기법을 활용하여 대규모 데이터 환경에서도 안정적이고 효율적인 하이퍼파라미터(hyper parameter) 탐색 가능성을 확인하였다. 공간 데이터의 배분 및 융복합 과정에서는 인구, 교통, 사회기반 시설 등 각종 공공 데이터 기반의 효율적인 융복합 가능성을 제시하였으며, 모형별 비교분석에서는 머신러닝 모형과 딥러닝 및 혼합 딥러닝 모형 간의 특징을 함께 고려한 변수설정 및 시범 분석을 제시하였다.

정책적 측면에서는 시공간 격자 단위의 인구이동 분석과 직전 주소지 예측 모형 개발을 통해 데이터 기반의 지역균형발전, 교통계획, 주거정책 등 다양한 분야에서 맞춤형 전략을 수립할 가능성을 제시하였다. 특히, 민감도 분석에서 제시한 SHAP는 신규 교통망 공급, 택지개발, 지역 산업 유치 등 정부의 정책 개입을 통한 전후 효과를 시뮬레이션할 수 있는 기반을 마련하였다는 데에 그 의의가 있는데, 정책 시뮬레이션 과정을 통해 정책 실패와 부작용 위험을 최소화하고 제한된 자원의 배분 과정에서도 효율성을 극대화하는 데 효과적으로 활용될 수 있을 것으로 기대한다. 또한, 고도화 연구를 통해 인구구조 변화에 따른 중장기 사회 및 경제적 영향 분석에도 적용할 가능성을 제시할 수 있을 것으로 기대한다.

제 3 절 한계점 및 향후 연구 방향

본 연구에서 제시하고 시범적으로 분석한 AI-PriSM(가칭) 모형은 높은 기술적·정책적 효용성이 확인되었으나, 실질적 활용 및 고도화 과정에서 몇 가지 한계점이 존재한다.

첫 번째로, 인구 및 상권정보 등 일부 변수의 경우 최신 자료 확보가 제한적이거나, 민감정보 보호를 위해 데이터의 세밀도를 임의로 조정해야 하는 한계가 있었다. 예를 들어, 국가데이터처에서 제공하는 인구이동 데이터의 최소 공간 단위가 읍면동인 데 반해 본 모형에서 분석하고자 하는 최소 공간 단위는 1km이기 때문에, 이를 보완하기 위해 직전 연도 대비 인구변화와 출생·사망자 수를 종합한 접근방식을 적용하였다. 하지만, 보다 정확한 모형의 설계를 위해서는 동등한 시공간 단위 집계를 활용한 분석방법론이나 논리적인 치환 방법이 필수적으로 요구될 것이다.

두 번째로, 국가데이터처에서 생산한 각종 데이터의 행정동 및 법정동 연계 등 시계열 데이터 융복합 과정에서 상당한 노력이 요구되었다. 시점상으로 20여 년 전의 데이터를 현재 시점에서 분석하는 과정에서 행정구역의 통폐합과 행정구역 코드의 변화, 행정구역의 승격 등 다양한 변화를 모두 면밀하게 반영하기에는 연구 기간 및 투입 인력의 물리적 한계가 있었다. 고도화 과정에서는 이러한 과정에서 누락되거나 잘못 처리된 행정 데이터의 수정과 면밀한 고려가 필요할 것이다.

세 번째로, 알고리즘 확장성 측면에서 현 단계에서는 선정된 알고리즘 범위 내에서 최적화를 수행하였으나, 연구 과업의 수행 시간과 하드웨어의 제한 등으로 외부 최신 기법과의 충분한 비교·검증을 수행하지 못하였다. 특히, 이러한 제약은 인구이동 예측모형의 비교 모형으로 삼았던 강화학습(Reinforcement Learning)의 시계열 예측의 일관성 확보에 어려움을 초래하였다. 향후에 보다 충분한 자원과 시간을 바탕으로 다양한 비교 모형의 제시를 통해 인구이동 관련 모형의 최적화와 고도화가 필요할 것이다.

네 번째로, 정책 시나리오 구성을 시범적으로 구현하지 못한 점은 향후 고도화 과정에서 반드시 보완해야 할 과제이다. 본 연구에서 제시한 AI-PriSM(가칭)이 정책 시뮬레이션과의 선순환 체계를 구축했을 때 비로소 그 효용성이 극대화될 수 있을 것으로 기대한다.

다섯 번째로, 변수 및 정책 시나리오 구성 시 실제 정책 환경의 복잡성을 모두 반영하기에는 변수와 관계 설정이 단순화된 측면이 있다. 단순히 한 가지 정책으로 단기간에 인구이동이 발생하는 경우는 드물며, 경험한 적 없는 이벤트에 대해서 그 영향을 예측하는 고도화된 AI 모델을 기대하는 것 또한 기술적으로 해결해야 할 난제일 것이다.

여섯 번째로, 본 시범 분석에서 제시한 AI-PriSM(가칭)은 전 국토 차원에서 모형의 비교와 분

석이 이루어졌다. 지역별 체감과 효용성을 극대화하기 위해서는 수도권과 특정 지방 등을 집중적으로 분석하는 사례 분석 등을 통해 모형의 실효성을 입증할 필요가 있다.

마지막으로, 향후 연구에서는 민간과 공공 데이터를 연계하여 인구이동과 밀접하게 연관된 생활, 경제, 환경 변수를 포함하는 통합 분석 체계를 확립할 필요가 있다. 더불어 최신 AI 기법인 Transformer 기반 시계열 예측 모형이나 CNN-LSTM 등과 같은 유사 혼합 알고리즘을 적용하고, 추가적인 머신러닝 및 딥러닝 혼합모형 설계를 통해 성능을 고도화할 수 있을 것이다.

본 연구에서 제안한 AI 기반 예측 모형은 인구이동 변화의 중장기 전망을 지원하고, 한정된 자원을 효율적으로 배분하기 위한 과학적 근거를 제공할 수 있을 것으로 기대한다. 특히 제한된 재원을 효율적으로 배분해야 하는 상황에서, AI 기반 정밀 예측은 정책 효과 극대화를 위한 유용한 도구가 될 것이다.

참고문헌

▌국내문헌

감사원(2021),『감사보고서: 인구구조변화 대응실태 I(지역)』.

국토지리정보원(2014),『대한민국 국가지도집 1권』.

김대용(2019), “산업구조 변화가 경제성장에 미치는 영향에 관한 연구: 성장수렴 가설 검정을 중심으로”,《산업연구(JIET)》, 3(1), pp. 1-30.

김명한(2020),『도시·군기본계획의 계획인구 추정 실태 및 시사점』, 국토연구원: WP20-12.

남형권(2020), “지역 산업구조 변화와 인구이동에 관한 연구: 유출입 인구의 규모 및 특성을 중심으로”, 중앙대학교 대학원 박사학위 논문.

민보경(2023), “지방소멸위기 대응 방향,” 국회미래연구원 *Futures Brief*, 제 23-03호.

박승규(2020), “산업별 고용구조변화에 따른 노동자 인구이동의 지역경제효과에 관한 연구”,《지역산업연구》, 43(2), pp. 231-261.

박승용(2022), “지역인구 예측을 위한 딥러닝 모델 연구: 부산지역 인구 기준,” 서울과학종합대학원 대학교 석사학위 논문.

서창원·양진홍(1998),『도표로 본 국토공간구조의 변화』, 국토개발연구원.

오은열·문채(2016), “지역인구이동 변화에 미치는 결정요인분석과 정책적 시사점: 전라남도를 중심으로”,《도시행정학보》, 29(4), pp. 67-92.

윤준희·진혁(2024), “비수도권 지역의 KTX역이 주변 지역 성장에 미치는 영향: 도심 외곽형을 중심으로”, 서울시립대학교 도시공학과 제38회 도시과학 공동작품전.

이동찬·김기환(2021), “모바일 기지국 데이터를 이용한 격자인구 추정”,《한국자료분석학회》, 23(4), pp. 1535-1548.

이보경·장요한·정예진·홍사흠·김동근(2022),『인공지능 기법을 적용한 소지역단위 장래인구 예측 방법론 개발 기초 연구』, 국토연구원.

이상림·최인선(2023), 『국내 인구이동의 변동 양상과 정책적 함의』, 보건사회연구원.

이정록(2017), “지방중심도시의 인구이동과 공간구조 변화: 전남 순천시를 사례로”, 《대한지리학회지》, 52(5), pp. 579-594.

이찬영(2018), “연령대별 인구유출입 결정요인 분석”, 《산업경제연구》, 31(2), pp. 707-729.

이태리·조정희·장요한·최진도·유승동·주승민(2020), 『4차 산업혁명 시대의 상업용 부동산 수요 및 이용행태 변화 연구』, 국토연구원.

이태직(2010), 『국내인구이동 추계모형 개선방안 연구』, 국가통계연구원.

임태경(2023), “연령별 청년인구 유입에 미치는 영향요인 분석:수도권-비수도권 간 차이분석을 중심으로”, 《국가정책연구》, 37(4), pp. 1 - 24.

장요한·손재선·최혜림(2024), 『민생현안 모니터링을 위한 민원지도 개선방안 연구』, 국토연구원.

정민수·김의정·이현서·홍성주·이동렬(2023), “지역간 인구이동과 지역경제”, 《BOK 이슈노트》, 한국은행.

조대헌·이상일(2022), “우리나라 소지역 인구추계를 위한 방법론 비교 분석”, 《한국지도학회지》, 22(2), pp. 51-67.

최대식·송영일·박미규·성태영(2024), 『지역특성을 고려한 수도권 내 인구이동 패턴 변화 분석 연구』, 한국토지주택공사 토지주택연구원(연구관리 2024-012호).

하미란·진장익(2024), “KTX 개통이 인구변화에 미치는 영향에 관한 연구”, 《한국지역개발학회지》, 36(1), pp. 117-138.

허재완·김갑성·유예진(2018), “고속철도(KTX)의 입지효과 및 개통으로 인한 수도권 인구집중 완화효과 분석”, 《국토계획》, 53(4), pp. 107-122.

▌해외문헌

Eggebrecht, P. and Eva Lütkebohmert(2023). “A hybrid convolutional neural network with long short-term memory for statistical arbitrage.” *Quantitative Finance* 23. no. 4. pp. 595-613.

Keeble, David E.(1967). “Models of economic development.” *In Socio-Economic Models in Geography*(Routledge Revivals). pp. 243-302.

Park, Cheonbok, Chunggi Lee, Hyojin Bahng, Yunwon Tae, Seungmin Jin, Kihwan Kim, Sungahn Ko, and Jaegul Choo(2020). “ST-GRAT: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed.” *In Proceedings of the 29th ACM international conference on information & knowledge*

management. pp. 1215-1224.

Tasdelen, Abdulkadir, and Baha Sen(2021). "A hybrid CNN-LSTM model for pre-miRNA classification." *Scientific reports.* 11. no. 1. 14125.

Tobler, Waldo R.(1979). "Cellular geography." *Philosophy in geography*. pp. 379-386. Dordrecht: Springer Netherlands.

Velickovic, Petar, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio(2017). "Graph attention networks." stat 1050, no. 20. 10-48550.

▌인터넷문헌

국가교통DB(https://www.ktdb.go.kr/, 2025년 7월 31일 접속).

국가데이터처, 「국내인구이동통계」
(https://mods.go.kr/statDesc.es?act=view&mid=a10501010000&sttr_cd=S001003, 2025년 11월 27일 접속).

국가데이터처, 지표서비스 「국내인구이동」
(https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx_cd=1013, 2025년 7월 37일 접속).

국가데이터처, 통계조사 통계별 질문
https://kostat.go.kr/board.es?mid=a10502010000&bid=3207&tag=&act=view&list_no=161814&ref_bid=, 2025년 7월 31일 접속).

국가데이터처, 「2019년 장래인구 특별추계를 반영한 내·외국인 인구전망: 2017~2040년」
(https://www.kostat.go.kr/board.es?mid=a10301020100&bid=207&tag=&act=view&list_no=385624&ref_bid=, 2025년 11월 27일 접속).

국가데이터처 보도자료, 「장래인구특별추계: 2017~2067년」
(https://www.kostat.go.kr/synap/skin/doc.html?fn=e34f89db4c82501b96c8212e2579544fb0e2ce6a81a61c449051c04cc04d7c11&rs=/synap/preview/board/207/, 2025년 11월 27일 접속).

국토교통부 보도자료, 「일상에서 만나는 공간정보 AI, 지도로 답하다」
(https://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp?lcmspage=1&id=95091096, 2025년 11월 27일 접속).

국토정보플랫폼, 국토지리정보원
(https://map.ngii.go.kr/ms/map/NlipMap.do?tabGb=statsMap, 2025년 7월 31일 접속).

마이크로데이터 통합서비스. 국가데이터처(https://mdis.kostat.go.kr/index.do, 2025년 7월 31일 접속).

위키백과, 검색어: 유전 알고리즘(https://ko.wikipedia.org/wiki/유전_알고리즘, 2025년 7월 31일 접속).

이호스트ICT 공식 블로그(https://webzero.tistory.com/3012, 2025년 7월 31일 접속).

지방행정인허가데이터(https://www.localdata.go.kr/, 2025년 7월 31일 접속).

Almabetter(https://www.almabetter.com/bytes/articles/artificial-neural-network, 2025년 7월 31일 접속).

Almabetter(https://www.almabetter.com/bytes/articles/convolutional-neural-networks, 2025년 7월 31일 접속).

fiveable(https://fiveable.me/combinatorial-optimization/unit-9/tabu-search/study-guide/grDBu3lBmHF5V9Ya, 2025년 11월 28일 접속).

GeeksforGeeks(https://www.geeksforgeeks.org/artificial-intelligence/introduction-hill-climbing-artificial-intelligence/, 2025년 11월 28일 접속).

GeeksforGeeks(https://www.geeksforgeeks.org/what-is-reinforcement-learning/, 2025년 7월 31일 접속).

IBM(https://www.ibm.com/think/topics/neural-networks, 2025년 7월 31일 접속).

JanBask Traning(https://www.janbasktraining.com/tutorials/genetic-algorithm/, 2025년 11월 28일 접속).

KDnuggets(https://www.kdnuggets.com/wtf-is-the-difference-between-gbm-and-xgboost, 2025년 7월 31일 접속).

KOSIS 국가통계포털, 국가데이터처(https://kosis.kr/index/index.do, 2025년 7월 31일 접속).

Long Short-Term Memory(LSTM)
(https://medium.com/@saba99/long-short-term-memory-lstm-fffc5eaebfdc, 2025년 11월 29일 접속).

Multi-Layer Perceptron Learning in Tensorflow
(https://www.geeksforgeeks.org/deep-learning/multi-layer-perceptron-learning-in-tensorflow/, 2025년 11월 29일 접속).

Semi-supervised Learning
Medium(https://medium.com/@gayatri_sharma/a-gentle-introduction-to-semi-supervised-learning-7afa5539beea, 2025년 7월 31일 접속).

SGIS 통계지리정보서비스, 국가데이터처(https://sgis.kostat.go.kr/view/index, 2025년 7월 31일

접속).

Skyflow(https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc/, 2025년 7월 31일 접속).

Supervised Learning, Medium
(https://medium.com/@metehankozan/supervised-and-unsupervised-learning-an-intuitive-approach-cd8f8f64b644, 2025년 7월 31일 접속).

The Basics of Recurrent Neural Networks(RNNs)
(https://pub.towardsai.net/whirlwind-tour-of-rnns-a11effb7808f, 2025년 11월 29일 접속).

Understanding Gated Recurrent Unit(GRU) in Deep Learning
(https://medium.com/@anishnama20/understanding-gated-recurrent-unit-gru-in-deep-learning-2e54923f3e2, 2025년 11월 29일 접속).

Unsupervised Learning, Baoyu
(https://baoyu.io/translations/ai-reports/stanford-hai-ai-index-report-2024-chapter1-en, 2025년 7월 31일 접속).

정책자료 25-12-05

AI기반의 인구이동 예측모형 개발

인 쇄: 2025년 12월
발 행: 2025년 12월
발행인: 이 상 규
주 소: 27872 충청북도 진천군 덕산읍 정통로 18
전 화: 043-531-4114
팩 스: 043-535-4695~6
I S B N: 979-11-7000-448-6 / 94320