

## 5장 발표문

### 석사과정 박제욱

이번 장에서는 설명가능한 인공지능(XAI: Explainable Artificial Intelligence)을 활용한 지리공간 분석의 개념과 유형에 대해서 살펴보고 GeoShapley, LIME, GLIME 등 방법론적인 내용도 서술하고 있습니다.

전통적인 공간 모델링은 오랫동안 공간 분석의 핵심 도구로 활용되어 왔으며, 관측된 공간 패턴을 설명하고 미래의 공간적 시나리오를 예측하는 데 중요한 역할을 해왔습니다. 대표적으로 공간 지연 모형, 공간 오차 모형, 공간 더빈 모형과 같은 공간 계량경제학 모델, 지리적 가중 회귀(GWR), 크리깅 기반 모델, 베이지안 공간 모델 등이 널리 사용되어 왔습니다.

이러한 통계적 공간 모델의 가장 큰 장점은 공간 자기상관과 공간 이질성을 명시적으로 모델링할 수 있고, 회귀 계수에 대한 명확한 해석과 불확실성 추정이 가능하다는 점입니다. 즉, 공간 현상을 이론적으로 설명하는 데 매우 강력한 도구라고 할 수 있습니다.

대부분의 통계적 접근의 시작은 강한 가정을 전제로 하며, 이러한 가정은 실제 세계의 복잡한 공간 과정을 단순화할 위험이 있습니다. 또한 데이터 규모가 커질수록 계산 비용이 급격히 증가하여, 최근 증가하고 있는 대용량·고해상도 공간 데이터에는 적용이 쉽지 않습니다. 그 결과, 비선형 효과나 변수 간 복잡한 상호작용, 그리고 이들이 공간적으로 결합된 패턴을 충분히 포착하는 데 한계가 발생합니다.

이러한 배경 속에서 지난 10여 년간 머신러닝이 공간 데이터 과학 분야에서 중요한 대안으로 부상하였습니다. 머신러닝 모델은 명시적인 모형 가정 없이 데이터로부터 직접 관계를 학습하며, 비선형성, 고차원 상호작용, 대규모 데이터를 효과적으로 처리할 수 있다는 장점을 가집니다. 이 흐름은 지리 공간 문제에 인공지능을 적용하는 지리 공간 인공지능, 즉 GeoAI라는 연구 분야로 확장되었습니다. GeoAI는 날씨 예측, 공간 보간, 원격 탐사 영상 분석, 도시 분석, 사회·인문 지리 연구 등 다양한 분야에서 높은 예측 성능과 활용 가능성을 보여주고 있습니다. 앞선 장에서 많이 검토하였듯이 이런 AI기술을 활용하여 지리 공간 분석을 진행하는 GeoAI 연구 분야가 형성되었고 기상 예측, 원격탐사 영상 분석, 도시 및 사회·인문 지리 연구 등 다양한 영역에서 성과를 보이고 있습니다.

하지만 AI기술을 활용한 GeoAI연구 분야는 머신러닝 모델은 구조가 복잡하고 비선형적 특성을 가지기 때문에, 모델이 왜 특정한 예측을 했는지를 이해하기 어렵다는 한계를 지니고 있습니다. 결국 인공지능의 내부 작동 방식을 이해하기 어렵거나 불가능한 현상을 의미하는 단어로 마치 내부를 들여다볼 수 없는 검은 상자 같다는 의미에서 비롯된 블랙박스 문제를 지리 공간 분석에 있어서 간과해서는 안된다는 주의점이 있습니다. 즉 모델의 예측 정확도는 높지만, 왜 그러한 예측이 도출되었는지 설명하기 어렵다는 점은 사회적·과학적 문제를 동시에 야기합니다.

이 과정에서 고안된 것이 설명가능한 인공지능으로, 복잡한 머신러닝 모델의 결정 과정을 인간이 이해할 수 있는 방식으로 설명하고 해석할 수 있도록 하는 기술과 연구 분야입니다. 결국 XAI의 목표는 AI 시스템의 투명성과 신뢰성을 높여 사용자가 AI의 결정을 믿고 활용하며, 문제 발생 시 원인을 파악하고 개선할 수 있도록 돕는 것입니다.

이런 AI의 블랙박스문제와 설명가능한 인공지능은 지리 공간 분석에 적용할 때 더 중요하게 작용할 것이라고 생각합니다. 앞선 발표 및 토론과정에서 많은 선생님들이 언급해주셨듯 지리 공간은 사람이 살아가는 공간으로서 단순한 데이터로서 설명되기 어려운 현상이 발생하기도 하며, 때로는 정치적·경제적인 힘의 논리를 통해 분석데이터가 의도될 수 있습니다. 지리학은 단순히 "무엇이 발생했는가"를 넘어서, "왜 그 현상이 그 장소에서 발생했는가"라는 질문을 핵심적으로 다루기 때문입니다. 만약 머신러닝 모델이 높은 예측력을 보이더라도, 그 결과를 설명할 수 없다면 공간적 과정에 대한 학문적 이해로 연결되기 어렵습니다.

특히 저는 사회적·윤리적 측면에서도 관련 내용이 중요하겠다고 책을 읽으면서 생각했는데, 예를 들어 선거 분석, 범죄 예측, 재난 대응, 도시 정책과 같이 사회적 파급력이 큰 분야에서 AI 기반 예측이 활용될 경우, 특정 지역이나 집단에 불리한 결과가 나타날 수 있기 때문입니다. 이때 분석자가 모델의 판단 근거를 설명할 수 없다면, 결과에 대한 책임 소재를 명확히 하기 어렵고, 정책적 정당성 역시 약화될 수 있습니다. 따라서 AI의 예측 정확도뿐만 아니라, 설명 가능성과 투명성은 점점 더 중요한 연구 과제가 되고 있습니다.

XAI의 활용 방식은 데이터 유형에 따라 달라집니다. 텍스트 데이터에서는 주의 매커니즘을 활용해 중요한 단어나 구문을 강조하고, 이미지 데이터에서는 Grad-CAM과 같은 기법을 통해 예측에 기여한 픽셀을 시각화합니다. 여기서 머신러닝에서의 주의력이라는 개념은 시퀀스 내 각 구성 요소가 그 시퀀스 내 다른 컴포넌트들에 비해 얼마나 중요한지 결정하는 방법으로 문장 내 각 단어에 부여된 가중치로 표현된다. 즉 인간의 주의력에서 영감을 받은 매커니즘으로 순환신경망이 문장의 어느 부분에도 동등하게 접근할 수 있게 해주는 역할을 한다. 공간 분석에서 흔히 사용되는 표 형식 데이터의 경우에는 특징 중요도나 부분 의존 플롯이 널리 활용된다.

방법론적으로 XAI는 크게 모델 기반 접근법과 모델 불가지론적 접근법으로 나눌 수 있습니다. 모델 기반 접근법은 선형 회귀나 의사결정 트리처럼 본질적으로 해석 가능한 모델에 기반하며, 회귀 계수나 분기 구조 자체가 설명을 제공합니다. 반면 모델 불가지론적 접근법은 모델 구조와 무관하게 입력과 출력 간의 관계를 분석하여 설명을 생성하며, LIME이나 Shapley 값 기반 방법이 이에 해당합니다.

또한 XAI는 글로벌 설명과 로컬 설명으로 구분됩니다. 글로벌 설명은 모델 전체의 일반적인 행동을 요약하는 반면, 로컬 설명은 개별 관측값이나 특정 지역에 대한 예측 근거를 제시합니다. 공간 분석에서는 특히 로컬 설명이 중요한데, 이는 설명

결과를 지도 위에 시각화할 수 있기 때문입니다.

공간 분석에서 로컬 XAI의 대표적인 초기 방법은 LIME입니다. LIME은 특정 관측 값 주변에서 복잡한 모델을 단순한 선형 모델로 근사하여 설명을 제공합니다. 그러나 로컬 이웃의 정의가 임의적이며, 결과가 민감하다는 한계를 지닙니다. 이를 보완하기 위해 지리적 근접성을 반영한 GLIME이 제안되었지만, 여전히 MAUP 문제와 모델 충실도 문제를 완전히 해결하지는 못합니다.

이러한 한계를 보완하기 위해 큰 주목을 받은 접근법이 샤플리 값 기반 방법입니다. Shapley value는 게임이론에서 유래한 개념으로, 여러 참여자가 협력하여 얻은 성과를 각 참여자의 기여도에 따라 공정하게 분배하는 원리를 제시합니다. 이를 머신러닝에 적용하면, 각 특징이 개별 예측에 기여한 한계 효과를 공정하게 분배하며, 다양한 모델과 분야에서 널리 활용되고 있습니다. 이 개념을 체계적으로 구현한 것이 SHAP이며, 현재 가장 널리 활용되는 XAI 기법 중 하나입니다. 그러나 SHAP은 공간 효과를 명시적으로 분리하지 못하고, 회귀 계수처럼 직접적인 공간적 관계를 의미하지 않는다는 한계를 가집니다. 결국 좌표를 단순한 독립 변수로 취급할 경우, 공간적 위치가 갖는 맥락적 의미와 지역 간 상호작용을 충분히 설명하기 어렵다는 한계가 있습니다.

이러한 한계를 극복하기 위해 제안된 방법이 GeoShapley입니다. GeoShapley는 모델 예측을 위치 효과, 비위치 특성 효과, 그리고 이들의 공간적으로 변화하는 상호작용 효과로 분해함으로써 머신러닝 모델 내의 공간적 효과를 명시적으로 해석할 수 있도록 합니다. 이는 전통적인 공간 통계 모델과 머신러닝 기반 접근을 연결하는 중요한 방법론적 진전이라고 할 수 있습니다.

이러한 방법론을 실제 사례에 적용한 사례로는 2020년 미국 대통령 선거에서 카운티 수준의 민주당 득표율을 예측한 분석을 소개합니다. 카운티 단위의 민주당 득표율을 종속변수로 설정하고, 인구 구조, 사회경제적 특성, 교육 수준 등의 변수를 활용해 머신러닝 모델을 구축합니다. 이후 GeoShapley를 적용하여 각 요인이 투표 성향에 미친 영향을 공간적으로 분해하고 지도화합니다. 또한 MGWR 결과와 비교함으로써, GeoShapley가 전통적인 공간회귀 모형과 유사한 해석력을 제공하면서도, 비선형성과 복잡한 상호작용을 포착할 수 있음을 보여줍니다. 흑인 인구 비율, 학사 학위 이상 비율, 인구 밀도는 민주당 득표율과 강한 비선형 관계를 보였고, 위치 효과는 남부와 서부·북동부 지역 간의 정치적 성향 차이를 명확히 드러냈습니다. MGWR 모델과의 비교 결과, 두 접근은 유사한 공간 패턴을 보였으나, GeoShapley는 비선형 효과를 보다 유연하게 포착할 수 있다는 장점을 보여주었습니다.

제가 생각했을 때 이번 장에서는 우리가 생각하는 AI기술을 지리 공간 분석에 직접 접목시킬 수 있는 방법론이 XAI라고 설명하고 있다고 느꼈습니다. 하지만 그럼에도 인문지리학 접근방법에서 XAI를 활용하는 것에 대한 생각을 말씀드리자면, 머신러닝의 분석결과는 실제 데이터 생성 과정이나 인과관계를 직접적으로 규명하지는 못합니다. XAI를 통해 분석된 결과 값에 대한 해석은 오히려 오해를 낳을 수

있으며, 이를 해석하는 것이 연구자의 가장 큰 역할이라고 생각했습니다.

#### 토론거리 ①

어느정도의 가정과 설계를 통해서 진행되는 전통적인 지리 공간 모델의 분석값과 XAI를 활용한 지리 공간 분석이 동일하게 해석될 수 있을까?