

# 기호 회귀의 환경경제학적 응용에 관한 연구 동향

양민호, 최태중\*

전남대학교

likepeople99@jnu.ac.kr, \*ctj17@jnu.ac.kr

## Research Trends on Environmental Economic Applications of Symbolic Regression

Min Ho Yang, Tae Jong Choi\*

Chonnam National Univ.

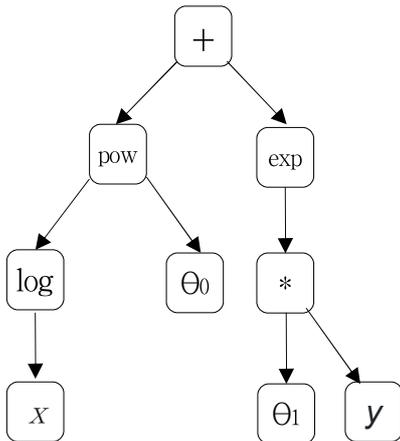
### 요약

본 논문은 환경경제학 분야에서 기호 회귀의 적용 사례와 연구 동향을 소개한다. 기호 회귀는 데이터로부터 수학적 모델의 구조와 파라미터를 탐색하여 해석 가능한 최적의 수식을 도출하는 기법이다. 환경경제학은 경제학적 관점에서 환경 문제를 분석하고, 개선 방안을 모색하는 학문이다. 기호 회귀를 통해 환경 요인과 경제 변수 간의 상관관계를 수식화함으로써 환경 정책의 설계와 평가에 유용하게 활용될 수 있다. 본 연구는 에너지 소비 및 대기 오염 요인 분석, 탄소 배출 요인 규명, 폐기물 가스화 모델 구축 등의 사례를 통해 환경경제학에서 기호 회귀가 적용된 다양한 연구 동향을 보여준다. 이를 통해 지속 가능한 발전을 위한 기호 회귀의 역할을 강조할 수 있다.

### I. 서론

기호 회귀(Symbolic Regression, SR)는 데이터로부터 수학적 모델의 구조와 파라미터를 탐색하는 방법이다. 일반적인 인공지능 알고리즘은 많은 파라미터를 포함하여, 모델의 의사 결정 과정을 명확히 설명하기 어려운 '블랙박스' 문제를 안고 있다. 그러나 기호 회귀는 간결한 수식을 통해 변수 간의 관계를 설명함으로써 이 문제를 극복한다. 이러한 장점으로 인해 기호 회귀는 금융, 생물학, 물리학, 임상 정보학 등 다양한 분야에서 응용되고 있다[1].

그림 1. 기호 회귀를 통해 트리 구조로 표현된 수식, [2]에서 발췌



기호 회귀는 여러 수식을 무작위로 생성한 뒤, 데이터에 대한 적합도를 평가하고, 이를 바탕으로 진화 알고리즘을 사용하여 최적의 수식을 찾는 방식으로 작동한다. 작동 과정에서 수식은 트리 구조로 표현되며, 노드는 연산자와 피연산자로 구성된다. 예를 들어  $(\log(x))^{\theta_0} + \exp(\theta_1 x)$ 와 같은 수식은 그림 1과 같은 트리 구조로 나타낼 수 있다[2]. 이와 같은 방식으로 데이터로부터 해석할 수 있는 수학적 모델을 도출할 수 있다.

환경경제학은 경제학적 원리와 방법론을 적용하여 환경 문제를 연구하는 학문이다[3]. 기호 회귀는 환경 요인과 경제 변수 간의 관계를 명확히

게 수식화할 수 있어, 환경 정책 수립 시 유용한 도구로 활용될 수 있다. 본 논문은 기호 회귀가 환경경제학 분야에 응용된 여러 연구 사례를 제시하고, 환경 정책 수립 과정에서 기여할 가능성을 논의한다.

### II. 기호 회귀의 환경경제학 응용 연구 사례

#### 1. 에너지 소비와 대기 오염 요인 분석

베이저안 기호 회귀(Bayesian Symbolic Regression, BSR)는 수식의 복잡성을 고려한 베이저안 통계 방법을 통해 최적의 모델을 선택하는 방법이다[4]. 연구[4]에서는 BSR을 활용하여 25년간 168개국의 1인당 GDP, 총인구, 활동인구, 인구밀도, 도시화율, 평균 기온 등 6가지 사회 경제 자료를 수집하고, 이산화탄소 배출량(CDE), 에너지 소비(EC), 아산화질소 배출량(NOE), 메탄 배출량(ME) 등 환경적 요인에 미치는 영향을 예측하는 모델을 개발하였다. 기존에는 각 변수의 환경적 영향을 일정 비율로 설명하는 STIRPAT 회귀분석 모델이 사용되었으나, 이 모델은 탄력성이 고정되어 있으며 비선형적 상호작용을 설명하는 데 한계가 있었다. 표 1은 BSR이 STIRPAT보다 더 높은 결정계수( $R^2$ )와 더 낮은 평균제곱오차(MSE)를 기록한 결과를 보여준다.

표 1. 환경적 영향 측정 지표 성능 비교, [4]에서 발췌

환경적 영향 측정 지표	평가 지표	STIRPAT	베이저안 기호 회귀
CDE	$R^2$	0.858	0.869
	MSE	0.677	0.626
EC	$R^2$	0.861	0.870
	MSE	0.627	0.588
ME	$R^2$	0.809	0.825
	MSE	0.625	0.572
NOE	$R^2$	0.809	0.817
	MSE	0.649	0.620

이 연구는 기존 회귀분석 모델의 한계를 기호 회귀를 통해 극복할 수 있음을 보여준다. 기호 회귀는 다양한 수식을 탐색하여 데이터에 가장 적합한 모델을 찾는 과정에서 더욱 유연하고 정확한 결과를 도출한다.

## 2. 탄소 배출 요인 탐구

연구[5]는 유전 프로그래밍(Genetic Programming, GP) 기반 기호 회귀 기법을 활용하여 OECD 회원국들의 탄소 배출 요인을 탐구하였다. 30년간 OECD 34개 회원국들의 국내총생산(GDP), 산업화 수준(ISG), 기술 혁신(TI), 도시화 수준(UP), 인구(TP), 외국인 직접 투자(FDI) 등의 주요 경제 변수가 각국의 탄소 배출 강도에 미치는 영향을 파악하였다. GP 기반 기호 회귀로 생성된 다수의 후보 모델 중 낮은 복잡성과 오차를 보이는 모델들을 파레토 프런티어 기법을 통해 선별하였다. 선별된 모델들에서 가장 빈번하게 등장하는 변수들을 탄소 배출에 영향을 미치는 주요 요인으로 판단하였다. 표 2는 연구 결과, 17개국에서 GDP가 탄소 배출에 영향을 미치는 가장 중요한 요인으로 확인되었음을 보여준다. 또한 산업화 수준과 기술 혁신이 4개국에서, 도시화 수준, 인구, 외국인 직접 투자가 3개국에서 가장 중요한 요인으로 나타났다. 또한 뒷순위로 갈수록 국가별 로 탄소 배출 요인은 더 다양한 양상을 보였다.

표 2. OECD 국가 탄소 배출 요인별 중요도 순위, [5]에서 발췌

중요도/요인	A	B	C	D	E	F	미미함
GDP	17	11	3	2	1	-	-
ISG	4	6	4	9	5	5	1
TI	4	4	9	3	6	5	3
UP	3	9	4	7	6	4	1
TP	3	2	9	9	5	5	1
FDI	3	2	5	4	11	5	4

이 연구는 GDP가 탄소 배출에 영향을 미치는 가장 중요한 요인은임을 확인하였으나, 뒷순위로 갈수록 국가별로 중요한 요인이 서로 다를 수 있다. 따라서 OECD가 하나의 탄소 감축 정책을 입안하면 모든 국가에 효과적이지 않을 수 있으며, 각국이 자국의 상황에 맞는 탄소 감축 전략을 수립해야 한다고 주장하였다.

## 3. 폐기물 가스화 모델 구축

폐기물 가스화는 고형 폐기물을 연료로 사용하여 합성 가스를 생산하는 기술이다[6]. 이 연구는 물리학에서 영감을 받은 기호 회귀 기법인 AI Feynman을 사용하여 고온(750°C ~ 1,100°C)에서 합성 가스의 핵심 성분인 수소와 이산화탄소의 비율을 예측하였다. AI Feynman은 물리학에서 발견되는 대칭성, 분리성, 저차 다항식 등의 특징을 활용하여 복잡한 비선형 데이터에서 수식을 도출한다[7]. 합성 가스에서 이산화탄소의 비율이 낮을수록 온실가스 배출을 줄여, 환경 문제를 예방할 수 있다. 연구에서는 AI Feynman을 사용하여 다양한 후보 모델을 생성한 후, 성능을 평균제곱 오차(MSE)와 피어슨 상관 계수(PCC)로 측정하였다. 또한 모델의 복잡성을 평가하기 위해, 근사 엔트로피와 샘플 엔트로피와 같은 동적 시스템 기준을 적용하였다. 이러한 과정을 통해 높은 복잡성을 가진 모델을 평가에서 제외하고, 우수한 성능과 낮은 복잡성을 가진 모델을 선택하였다. 연구는 동적 시스템을 적용해 복잡한 모델을 효과적으로 제거함으로써, 환경경제학 분야뿐만 아니라 다양한 모델링에서도 활용할 가능성을 제시하였다.

## III. 결론

본 논문에서는 기호 회귀를 환경경제학적 분석에 적용한 연구 동향을 검

토하였다. 에너지 소비와 대기 오염 요인 분석, 탄소 배출 요인 탐구, 그리고 폐기물 가스화 모델 구축 사례를 통해, 기호 회귀가 기존의 통계적 모델링 접근법의 한계를 극복하면서도 더 정밀하고 유연한 예측 능력을 제공하였음을 확인하였다. 특히, 기호 회귀는 비선형적인 경제 변수와 환경 요인 간의 관계를 효과적으로 모델링할 수 있어, 환경경제학 분야에 새로운 시각을 제시할 수 있다.

에너지 소비와 대기 오염 분석에서 베이지안 기호 회귀는 기존의 STIRPAT 모델에 비해 뛰어난 예측 정확도를 보여주었다. 유전 프로그래밍 기반 기법을 통해 탐구한 탄소 배출 요인 연구에서는 국가별 특성을 고려한 탄소 감축 정책 개발의 필요성을 강조하였다. 또한, 폐기물 가스화 모델 구축 사례에서는 물리학적 원리에 기반한 AI Feynman 기법으로 비선형적 데이터에서 복잡성이 낮고 해석 가능한 모델을 도출하였다. 본 연구는 기호 회귀가 환경 정책의 수립 및 평가 과정에서 유용하게 적용될 수 있음을 보여준다.

## ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2023-00214326 및 RS-2023-00242528).

## 참고 문헌

- [1] LA CAVA, William, et al. Contemporary symbolic regression methods and their relative performance. *Advances in neural information processing systems*, 2021, 2021.DB1: 1.
- [2] BARTLETT, Deaglan J.; DESMOND, Harry; FERREIRA, Pedro G. Exhaustive symbolic regression. *IEEE Transactions on Evolutionary Computation*, 2023.
- [3] SUHJUNGWOO and 최영, "Development Process and Prospects of Environmental Economy in China," *경영사연구*, vol. 38, no. 4, pp. 81-108, 2023.
- [4] VÁZQUEZ, Daniel, et al. Automatic modeling of socioeconomic drivers of energy consumption and pollution using Bayesian symbolic regression. *Sustainable Production and Consumption*, 2022, 30: 596-607.
- [5] PAN, Xiongfeng, et al. Influential factors of carbon emissions intensity in OECD countries: evidence from symbolic regression. *Journal of Cleaner Production*, 2019, 220: 1194-1201.
- [6] PRAKS, Pavel, et al. Selection of appropriate symbolic regression models using statistical and dynamic system criteria: Example of waste gasification. *Axioms*, 2022, 11.9: 463.
- [7] UDRESCU, Silviu-Marian; TEGMARK, Max. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 2020, 6.16: eaay2631.