

2025년 추계학술발표대회 : 일반부문

기상 데이터에 대한 인과구조 식별 및 경로 분석

- 기상 데이터에서 건물 에너지로 -

Identification of causal structure and analysis of causal path

- from weather data to building energy-

○추 한 경*

Chu, Han-Gyeong

김 혜 기*

Kim, Hye-Gi

김 덕 우*

Kim, Deuk-Woo

Abstract

Conventional building energy analyses tend to focus primarily on input-output correlations, which risks conflating spurious inter-variable associations. Although causal inference approaches that consider cause-effect relations among variables have been introduced to address this limitation, it remains impractical for researchers to explicitly specify all causal relations among the measured variables. Accordingly, this study applies causal discovery to identify causal relations between building and climatic variables and proposes a method for identifying potential biasing paths that can create spurious associations between causes and outcomes. The results indicate that the PC algorithm provides meaningful insights for the weather data, and analysis of the extended causal structure reveals identifiable biasing paths.

키워드 : 인과발견, PC 알고리즘, 편향 경로

Keywords : Causal discovery, Peter-Clark algorithm, Biasing path

1. 서론

건물에너지 분야는 탄소 배출에 있어 주요한 요인 중 하나로 여겨지고 있으며, 이에 따라 건물에너지 사용량을 예측하고 그린리모델링의 효과를 분석하려는 다양한 연구들이 수행되고 있다. 예측 모델로는 비교적 적은 노력으로 정확한 예측을 수행 가능한 기계학습 방법이 주로 이용되고 있으나, 이러한 접근방법은 입력변수와 출력변수 간 변화하는 관계, 즉 데이터 간 상관관계에만 주로 집중되어 있기 때문에 우연히 발생한 사건을 유의미한 연관으로 오인할 수 있으며, 제실자 행동과 같은 설명하기 어려운 요인이나 변수 간 상호관계가 출력변수에 미치는 영향을 분석하는데는 한계가 존재한다.

이를 극복하기 위해 변수 간 원인-결과 관계, 즉 인과관계를 고려하는 접근 방법을 통해 변수 간 설명력을 증가시키는 인과추론 연구가 수행되고 있다. 그러나 고려해야 하는 변수의 수가 증가할수록 경험적 지식이나 전통적인 수식에 기반한 접근의 한계가 분명해지기 때문에, 인과발견 기법을 이용한 기초적인 인과구조 식별을 수행할 필요가 있다(Nogueira et al., 2022). 따라서 본 연구에서는 기상 요인과 건물 냉방에너지 간 인과관계에 대한 파일럿

테스트의 일환으로, 여름철 기상데이터에 대하여 인과발견 알고리즘을 적용하고, 발견한 구조에 대하여 확장 및 해석하는 과정을 소개한다.

2. 연구방법

2.1 인과발견

인과발견은 주어진 데이터셋을 바탕으로 변수간의 인과관계를 식별하고 해석하기 위한 접근 방식으로(Spires et al., 2000), 본 연구에서는 다양한 인과발견 방법 중 변수 간 조건부 독립성 검정을 통해 인과구조를 탐색하는 제약 기반 방법(Constraint-based)에 기반하며 널리 사용되는 Peter-Clark(PC) 알고리즘을 활용하였다. 해당 알고리즘은 관측되지 않은 교란변수가 존재하지 않는다는 가정 하에 인과구조를 식별하는 방법으로, “→과” “—” 두가지 표기를 사용하여 인과정보를 전달한다. 예를 들면, “외기 온도→냉방에너지”의 관계는 외기 온도가 냉방에너지에 영향을 미치는 명확한 방향성을 드러내며, “상대습도—강수량”의 관계는 연관성은 존재하나 명확한 방향성은 파악하기 힘든 상태를 나타낸다.

2.2 편향경로

편향 경로(biasing path)는 인과 구조에서 원인과 결과 사이에 허위 연관성을 만들어 내거나 둘 모두에 공동으로 영향을 미쳐 연구 결과에 편향을 도입하는 경로를 말한다. 따라서 인과발견을 통해 인과 그래프(Directed Acyclic Graph, 이하 DAG) 비순환성 DAG를 구축하고 잠재적 편

* 한국건설기술연구원 건축에너지연구본부

(Corresponding author : Korea Institute of Civil Engineering & Building Technology, deukwookim@kict.re.kr)

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(과제번호 RS-2023-00244769)

향 경로를 식별하면, 이후 분석에서 문제가 될 가능성이 높은 교란변수(confounders) 및 관련 요인에 대한 적절한 보정을 통해 이러한 편향 경로를 차단하거나 허위 연관관 발생을 방지함으로써 인과관계 해석 정확도를 향상시킬 수 있다.

3. 사례연구를 통한 인과구조 식별 및 인과경로 분석

3.1 기상 데이터 및 준 정답 구조

냉방에너지에 대한 기상 요인의 인과적 영향을 분석하기에 앞서, 기상 요인 간 인과관계를 분석하기 위하여 한국 서울의 2024년 여름철 (7월과 8월) 일일 측정값으로 인과발견용 데이터셋을 구성하였다. 총 10개의 변수가 선정되었으며, 이는 풍속(Wind), 수평면 일사량(Global horizontal irradiation, 이하 GHI), 구름량(Cloud), 증발량(Evar), 해면기압(Avg.sea), 상대습도(Re humi), 일일 강수량(Rain), 평균 증기압(Avg.sp), 이슬점 온도(Dew point), 건구 온도(Dry bulb)이다. 강수량의 결측치는 0 mm/day로 대체하였으며, 그 외 변수의 결측치는 전일과 후일의 평균값을 사용하였다. 본 연구에서는 습공기 선도를 통해 관계를 추론 가능한 상대습도, 이슬점온도, 건구온도 이외의 변수들 간 인과관계에 대한 합의된 정답 구조는 존재하지 않은 상황을 가정하고, 저자의 경험적 지식을 바탕으로 10개 변수 간의 준 정답구조(Quasi-ground truth)를 설계하였다(그림 1(a)).

3.2 인과구조 식별 및 인과경로 분석

그림 1(b)는 기상 데이터셋에 PC 알고리즘을 적용해 얻은 인과발견 결과를 제시한다. 그림 1(a)에서 저자의 기대와는 달리, 7-8월 자료에서는 이슬점 온도(Dew point)와 평균 수증기압(Avg. sp)이 다른 변수들과의 상호작용에서 고립되어 있음을 보여주며, 수평면 일사량(GHI)은 구름량(Cloud)에 영향을 미치지 않으며, 증발량(Eva)은 GHI와 상대습도(Re humi)와의 연관성을 보이는 것으로 나타난다. 아울러 저자들은 상대습도에서 나오는 정보가 건구온도(Dry bulb)로만 흐른다고 가정했으나, 발견된 구조는 상대습도가 윤량과 건구온도 모두에 영향을 미치는 것으로 나타났다.

이후 기상 데이터가 냉방에너지에 미치는 영향을 분석

하기 위해 상기의 발견된 인과 구조를 바탕으로, 건구 온도(Dry bulb)와 수평면 일사량(GHI)이 냉방 에너지에 직접적인 영향을 미친다고 가정하여 인과 구조를 그림 1(c)와 같이 확장하였다. 이러한 인과 구조에서 건구 온도를 원인으로, 냉방 에너지를 결과로 설정하였을 때, 건구 온도가 냉방 에너지에 미치는 직접 인과효과(direct causal effect)를 추정하고자 하는 경우를 가정하였다(그림 1(d)). 이 때, 수평면일사량, 증발량, 상대습도 사이의 방향성이 미확정 상태이므로 특정한 환경에서는 수평면일사량이 건구 온도와 냉방 에너지 모두에 영향을 미치는 교란변수로 작용하여 편향 경로가 생성될 수 있음을 확인할 수 있다.

4. 결론

본 연구에서는 냉방 에너지에 대한 기상 요인의 인과적 영향을 분석하기 위해, 변수 간의 인과관계를 식별하고, 원인과 결과 사이에 허위 연관성을 만들어낼 수 있는 잠재적 편향 경로를 식별하는 방법을 사례연구를 통해 제안하였다. 대한민국의 여름철 기상 데이터에 대하여 PC 알고리즘으로 기초 인과 구조를 식별한 뒤, 식별된 구조에 대하여 냉방에너지 사용량을 포함해 이를 확장하였다. 그 결과, PC 알고리즘은 기상 데이터에 대해 의미 있는 통찰을 제공하였으며, 확장된 인과 구조의 분석을 통해 식별 가능한 편향 경로가 드러냄으로써, 인과발견을 이용한 DAG 구축 및 분석은 인과효과를 보다 신뢰성 있게 식별하고 추정할 수 있음을 확인하였다.

참고문헌

1. Nogueira, A. R., Pugnana, A., Ruggieri, S., Pedreschi, D., & Gama, J. (2022). Methods and tools for causal discovery and causal inference. Wiley interdisciplinary reviews: data mining and knowledge discovery, 12(2), e1449.
2. Spirtes, P., Glymour, C., Scheines, R., Kauffman, S., Aimala, V., & Wimberly, F. 2000. Constructing Bayesian network models of gene expression networks from microarray data.

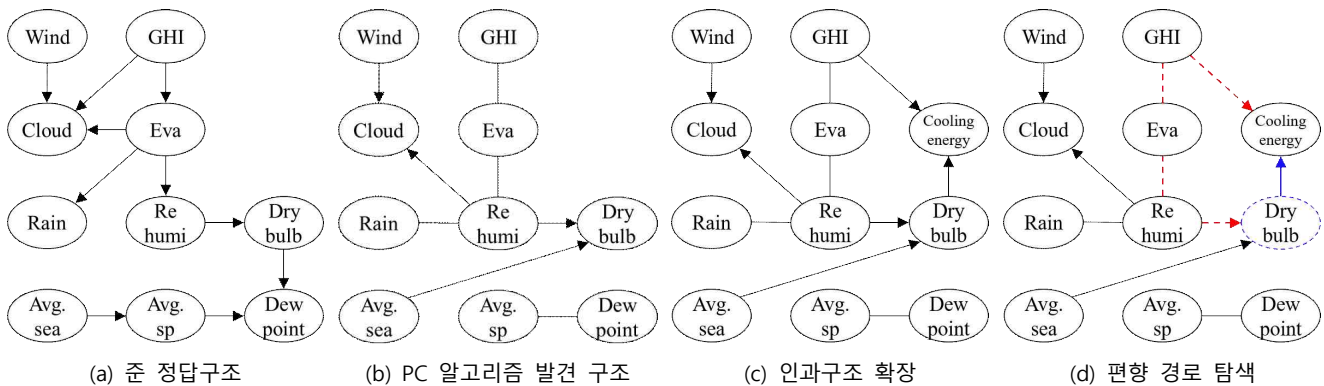


그림1. 기상 데이터에 대한 인과구조 식별 및 경로 분석 과정