

머신러닝과 실주행 데이터를 이용한 자동차 LCA 사용 단계 온실가스 배출량 보정에 대한 연구

조 성 인¹⁾ · 정 환 수²⁾ · 박 수 한³⁾

건국대학교 미래모빌리티융합전공¹⁾ · 국립환경과학원 모빌리티환경연구센터²⁾ · 건국대학교 기계로봇자동차공학부³⁾

A Study on Correcting Use-phase Greenhouse Gas Emissions in Automotive LCA Using Machine Learning and Real-world Driving Data

Seongin Jo¹⁾ · Hwan Soo Chong²⁾ · Suhan Park³⁾

Convergence Major of Future Mobility, Konkuk University¹⁾

Mobility Environmental Research Center, National Institute of Environmental Research²⁾

School of Mechanical, Robotics and Automotive Engineering, Konkuk University³⁾

Key words : Real-Driving Emissions(실도로 주행 배출량), Discrepancy Factor(불일치 계수), Machine Learning (머신러닝), SHAP(SHapley Additive exPlanations)

* 교신저자, E-mail: suhanpark@konkuk.ac.kr

자동차 전과정평가(Life Cycle Assessment, LCA)는 차량의 환경적 영향을 제조, 사용, 폐기 단계를 포괄하여 정량화하는 방법론으로, 탄소중립 정책 수립 및 친환경 차량 기술 평가의 핵심 도구로 활용되고 있다. 화석연료를 사용하는 내연기관 차량(ICEV)의 경우 사용 단계 CO₂ 배출은 전체 생애주기 환경 부하의 70~80%를 차지하는 핵심적인 요소이며, 이 단계에서 입력되는 데이터의 정확성은 분석 결과 전체의 신뢰성을 좌우하는 결정적인 척도가 된다. 그럼에도 불구하고 현재 대부분의 자동차 LCA 연구와 정책 평가는 사용 단계의 입력값으로 공인 인증값을 그대로 사용하는 관행에 의존하고 있다. 이는 단순히 실도로 배출량이 인증값보다 높게 나타나는 과소추정의 문제를 넘어서 더욱 근본적인 한계를 내포한다. 인증값과 실도로 배출량 사이의 불일치 계수(Discrepancy factor)가 차량의 유형, 물리적 제원, 그리고 주행 국가의 도로 환경에 따라 비균질적으로 분포한다는 점이 핵심이다. 예를 들어, 소형 경차와 고배기량 대형 SUV의 불일치 계수가 각각 다를 때, 두 차량을 동일한 인증값 기반으로 비교하는 LCA의 상대적 결론은 실도로 조건 기반 비교와 달라질 수 있다. 특히 내연기관과 전기차 간의 탄소 배출 역전 지점(Break-even point) 추정이나 하이브리드(HEV)의 실제 저감 효과 평가 역시 이러한 비균질적 편향으로 인해 왜곡될 가능성이 크다.

본 연구는 이러한 방법론적 한계점에 착안하여 차량 제원과 국가별 도로 환경 조건을 입력 변수로 하는 실도로 CO₂ 불일치 계수 예측 모델을 구축하고, 이를 자동차 LCA 사용 단계 배출량 보정에 통합하는 방법론적 프레임워크를 제안한다. 연구의 목표는 머신러닝을 활용하여 변수 간 상관관계를 규명하고, 차량 유형과 사용 환경에 따라 차별화된 실도로 보정 계수를 LCA에 반영하는 방법론적 근거를 마련하는 것이다. 분석에는 EU 유럽환경청(EEA)의 OBFCM(On-Board Fuel and Energy Consumption Monitoring) 데이터를 활용하였다. 이는 2021년부터 2023년까지 EU 회원국에서 운행된 승용차 실도로 연료 소비 기록을 포함한 대규모 데이터세트이다.

데이터 가공을 위해 네 가지 정보 층위를 하나의 레코드로 결합하는 데이터 매핑을 수행하였다. 차량 센서로 측정된 실제 연료 소비량 및 주행 거리 정보를 바탕으로 해당 차량의 인증값, 물리 제원, 주행 국가의 외부 환경 변수를 연결하였다. ICEV와 (P)HEV는 연료 소비 메커니즘이 다르므로 분석 모델을 분리하여 구축하였다. 국가별 도로 환경 분석 시에는 국가 명칭 대신 고속도로 비율, 도시 주행 비율, 연평균 기온, 인구 밀도 등의 물리적 수치를 연속형 변수로 산출하여 모델에 입력하였다. 이는 특정 국가의 불일치 계수 특성을 기후나 도로 인프라 관점에서 해석하기 위함이다.

예측 모델로는 트리 기반 앙상블 알고리즘인 LightGBM을 적용하였다. 동일 차량 모델이 반복 등장하는 데이터 구조에서 발생할 수 있는 성능 과장 문제를 방지하기 위해, 특정 차량 모델을 검증셋으로 분리하여 학습에 사용되지 않은 차량에 대한 예측력을 평가하는 그룹 기반 교차검증(Group

K-Fold CV)을 수행하였다. 또한 SHAP(SHapley Additive exPlanations) 분석을 도입하여 차량 무게, 배기량, 주행 지역 기온 등의 변수가 불일치 계수에 미치는 기여도를 정량적으로 산출하였다. 구축된 모델은 기존 LCA의 사용 단계 배출량을 실도로 보정값으로 대체하며, 민감도 분석이나 불확실성 정량화 연구에 활용할 수 있도록 설계하였다.

본 연구는 자동차 LCA에서 사용 단계 배출량 입력값의 비균질적 편향이라는 방법론적 한계에 주목하고, 대규모 실도로 데이터와 머신러닝을 결합하여 이를 체계적으로 다루는 프레임워크를 제안하였다. 이는 차량 제원과 사용 환경에 따라 차별화된 실도로 보정값으로 사용 단계 CO₂ 배출량을 대체하는 접근은 단일 인증값 사용 관행이 유발하는 비교 LCA 결과의 왜곡을 개선한다. 실측 데이터를 LCA 방법론 개선에 직접 연결한 본 연구는 데이터 기반 LCA 방법론 연구의 새로운 방향성을 제시한다. 동일한 차량 제원을 가진 차량이라도 사용 국가의 도로 환경에 따라 불일치 계수가 체계적으로 달라짐을 입증함으로써, 사용 단계 배출량이 차량 설계와 사용 환경의 상호작용으로 결정된다는 함의를 도출한다. 또한 제안된 LCA 통합 프레임워크는 자동차 LCA에서 실도로 보정 계수 도입의 정량적 근거를 제공한다. 특히 동일 차량의 공인 인증값을 사용하더라도 도로 환경에 따른 별도 보정이 필요함을 논리적으로 제시한다. 한국은 에너지소비효율 평가에 2-사이클 방식을 적용하고 있는데, 본 연구의 프레임워크는 한국형 보정 모델 개발 및 국내 자동차 실도로 탄소 성능 연구로의 확장 가능성을 시사한다.

Acknowledgement

이 논문은 한국연구재단 중견연구후속지원사업(RS-2023-00208502)과 국립환경과학원 연구용역사업(LCA 기반 자동차 온실가스 배출량 평가 방법 연구(III))의 지원으로 수행되었습니다. 지원기관에 감사드립니다.