

베이지안 트리 앙상블 기반 전기차 주행 가능 거리 예측 및 성능 분석

전혜원, 홍참길
한동대학교 전산전자공학부

{weon0329, charmgil}@handong.ac.kr

Bayesian Additive Trees for Electric Vehicle Range Prediction and Performance Evaluation

Hyewon Jeon, Charmgil Hong
School of Computer Science and Electrical Engineering, Handong Global University

요 약

본 연구는 전기차 주행 가능 거리 예측의 정확도를 향상시키기 위해, 실제 대규모 주행 데이터를 기반으로 Bayesian Additive Regression Trees(BART) 모델을 활용한 예측 프레임워크를 소개한다. 데이터 내 존재하는 분산과 비선형성을 효과적으로 반영하기 위해 베이지안 트리 앙상블 구조를 채택하였으며, 성능 비교를 위해 트리 기반 모델과 베이지안 선형 회귀 모델을 기준 모델로 설정하였다. 실험 결과, BART는 모든 평가 지표에서 가장 우수한 성능을 나타냈으며, 비선형 관계 학습과 예측 불확실성 추정이 주행 거리 예측 성능 향상에 기여함을 확인할 수 있었다. 확보된 실험 결과는 향후 주행거리 불안 해소와 전기차 보급 확대에 기여할 가능성을 제시한다.

1. 서론

최근 전기차에 대한 사회적 관심이 높아지면서 수요도 급격히 증가하고 있다[1]. 이와 함께, 운전자들이 크게 체감하는 문제는 주행 가능 거리에 대한 불확실성으로, 이는 ‘주행거리 불안(Range Anxiety)’이라는 용어로 알려져 있다[2]. 이러한 불안은 전기차 선택을 가로막는 요인 중 하나로, 이를 해소하기 위한 방안으로 주행 가능 거리 예측(Range Prediction)이 주목받고 있다. 지금까지 다양한 예측 모델이 제안되었지만, 대부분 단일 예측 값만을 제공하며 예측의 불확실성을 반영하지 못한다는 한계가 있다[3].

본 연구는 실제 대규모 전기차 주행 데이터를 활용하여, 주행 가능 거리를 보다 정밀하게 예측할 수 있는 모델을 개발하는 것을 목표로 한다. 이를 위해, 트리 기반 앙상블 모델인 Light Gradient Boosting Machine(LGBM)[4]과 불확실성을 정량적으로 반영하는 Bayesian Linear Regression(BLR)[5], Bayesian Additive Regression Trees(BART)[6] 모델을 적용하고, 성능을 체계적으로 비교·분석한다. 연구를 통해 운전자에게 높은 신뢰도와 정확도를 갖춘 주행 가능 거리 정보를 제공하여 전기차 운행의 실용성과 편의성을 향상시키는 데 기여하고자 한다.

2. 방법

2.1. 데이터 수집 및 전처리

전기차 주행 가능 거리 예측을 위해, 2022년 9월

부터 2024년 10월까지 실제 운행된 6종의 차량(총 479대)에서 약 1,591만 km 규모의 주행 데이터를 수집하였다. 이 데이터는 2초 간격의 Battery Management System(BMS) 정보와 1분 간격의 텔레매틱스 데이터를 결합하여 주행 사이클 단위로 구성된다. 각 사이클은 10분 간격으로 리샘플(resample)되며, 해당 구간 및 전체 주행 구간에 대한 통계량(최대, 최소, 평균, 표준편차)과 과거 주차(week) 및 주행 기반의 시계열 특성이 포함된 요약된 vector 형태로 변환된다. 이렇게 구성된 각 행(row)은 하나의 주행 사이클에 대응되며, 모델 학습의 입력 값으로 사용된다.

2.2. 학습 및 예측 프레임워크

데이터 수집을 통해 확보한 1,594개의 변수 중 래퍼 방법(wrapper method)[7]을 활용하여 예측에 유의미한 34개를 선정하였다. 예측 대상은 주행 중 임의의 시점(t 시점)부터 종료 시점까지의 전비(단위 전력당 주행 거리, km/kWh)이며, 실제 데이터에는 배터리를 완전히 소모한 사례가 드물어 전비 예측 값에 해당 구간의 실제 전력 사용량을 곱해 주행 가능 거리를 산출한다. 이후 이를 실제 주행 거리와 비교하여 모델의 전비 예측 성능을 평가한다.

2.3. 예측 모델

전기차 주행 가능 거리 예측에서는 동일한 차량 조건과 유사한 운행 환경에서도 실제 주행 거리가 크게

달라지는 경우가 빈번히 발생한다. 이러한 데이터 내 분산(variance)은 모델링 과정에서 단일 예측 값만으로는 충분히 설명되기 어렵고, 결과적으로 예측 정보의 신뢰도를 저하시킬 수 있다. 이러한 문제를 해결하기 위해, 본 연구에서는 BART를 주요 예측 모델로 채택하였다. BART는 다수의 회귀 트리를 결합하는 방식에 베이지안 추론을 접목하여, 예측 값뿐만 아니라 불확실성까지 정량적으로 추정할 수 있다. 특히, 베이지안 모델은 파라미터를 고정된 값이 아닌 확률 분포로 가정하고, 관측된 데이터를 사후분포로 갱신함으로써 예측의 신뢰구간을 제공한다. BART는 이러한 구조를 바탕으로 비선형 관계를 효과적으로 포착할 뿐만 아니라, 운전 조건 간 내재된 예측 값의 변동성을 흡수할 수 있어 실질적인 예측 신뢰도 향상에 기여한다. 결과적으로, 단일 수치 기반 예측보다 정밀한 주행 가능 거리 정보를 제공할 수 있다.

3. 실험

3.1. 실험 세팅

BART 모델의 성능을 효과적으로 검증하기 위해 비교 모델로 LGBM과 BLR을 선정하였다. LGBM은 트리 기반 앙상블 모델로, BART와 구조적으로 유사하지만 베이지안 추론이 적용되지 않는다는 점에서 두 모델 간 구조적 특성의 영향을 구분하는 데 적합하다. 반면 BLR은 BART와 동일한 베이지안 프레임워크를 공유하지만, 단일 선형 모델 구조를 갖고 있어 앙상블 구조가 성능에 미치는 영향을 비교할 수 있는 기준점으로 활용된다. 이를 통해 BART의 전기차 주행 가능 거리 예측 성능을 모델 구조와 추론 방식이라는 두 측면에서 정량적으로 평가할 수 있다.

실험은 데이터의 시계열 특성을 반영하여 5-fold 시계열 교차 검증 방식으로 진행하였으며, 각 모델의 하이퍼파라미터는 Optuna[8] 최적화 프레임워크로 자동 조정하였다. 모델 성능 평가는 Mean Absolute Percentage Error(MAPE), Root Mean Squared Error(RMSE), Mean Absolute Error(MAE), Coefficient of Determination(R^2) 기준으로 수행하였다.

3.2. 결과 분석

표 1은 차량 계기판 내의 주행 가능 거리와 실험을 통해 적용한 세 예측 모델의 성능을 비교한 결과이다. 이 중 BART 모델이 모든 평가 지표에서 가장 우수한 성능을 보였으며, 특히 MAPE 기준 차량 계기판 대비 약 60.3%, LGBM 및 BLR 대비 각각 7.1%와 6.2% 향상된 수치를 기록하였다. MAPE 기준 t-검정 결과, BART는 LGBM과 BLR 모두에 대해 통계적으로 유의한 차이($p < 0.05$)를 보였으며, 이는 예측 정밀도 측면에서의 우위를 시사한다.

실험 결과, BART를 활용한 예측이 모든 평가 지표에서 가장 뛰어난 성능을 보였다. 이는 트리 기반 앙상블 구조가 복잡한 비선형 관계를 효과적으로 학습하고, 다양한 주행 조건에 따른 내재된 데이터 분산을 흡수한 결과로 해석된다. 특히 베이지안 앙상블 방식은 예측을 보다 안정적으로 만들고, 신뢰 구간을 함께 제공함으로써 운전자에게 보다 신뢰도 높은 정

	MAPE	RMSE	MAE	R^2
차량 계기판	24.551	16.278	10.489	0.867
LGBM	10.481±0.402	8.217±0.318	5.066±0.181	0.965±0.002
BLR	10.380±0.312	7.755±0.653	4.729±0.234	0.967±0.004
BART	9.738±0.239	7.159±0.191	4.432±0.118	0.972±0.001

〈표 1〉 차량 계기판 및 모델 별 성능 지표

보를 전달할 수 있다. 종합적으로, BART는 전기차 주행 가능 거리 예측에 있어 구조적·실용적 측면 모두에서 가장 적합한 모델로 평가된다.

4. 결론

본 연구는 전기차 479대에서 수집한 약 1,591만 km 규모의 데이터를 기반으로, 주행 가능 거리 예측을 위한 BART 중심 프레임워크를 개발하였다. 차량별 주행 사이클을 구성하고 전처리를 통해 예측에 적합한 형태로 데이터를 정제하였으며, 데이터 내 분산을 반영할 수 있도록 베이지안 앙상블 구조를 적용하였다. 실험 결과, 제안한 BART 모델은 차량 계기판 및 비교 모델(LGBM, BLR)보다 높은 예측 정확도를 기록하였고, 모든 평가 지표에서 일관된 우수성을 보였다. 이러한 결과는 비선형 관계 학습과 불확실성 추정을 함께 수행하는 구조적 특성에 기인한다. 향후에는 해당 프레임워크를 실시간 응용 시스템에 적용하여, 주행거리 불안을 완화하고 전기차 보급 확대에 기여할 수 있을 것으로 기대된다.

Acknowledgements

- 본 연구는 2025년 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업 지원을 받아 수행되었음 (2023-0-00055).
- 본 연구는 2024년도 중소벤처기업부 창업진흥원의 초격차스타트업 1000+ 프로젝트 지원을 받아 수행되었음 (20226466).

참고문헌

- [1] J. A. Sanguesa et al., "A Review on Electric Vehicles: Technologies and Challenges," *Smart Cities*, vol. 4, no. 1, pp. 372–404, 2021.
- [2] L. Noel et al., "Fear and loathing of electric vehicles: The reactionary rhetoric of range anxiety," *Energy Research & Social Science*, vol. 48, pp. 96–107, 2019.
- [3] A. Bolovinou et al., "Online prediction of an electric vehicle remaining range based on regression analysis," *Proceedings of the 2014 IEEE Vehicle Power and Propulsion Conference (VPPC)*, Coimbra, Portugal, 2014, pp. 1–6.
- [4] G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *Proceedings of the 31st Conference on Neural Information Processing*.
- [5] G. E. P. Box, et al., *Bayesian Inference in Statistical Analysis*, New York, Wiley, 1973.
- [6] H.A. Chipman, et al., "BART: Bayesian Additive Regression Trees," *Ann. Appl. Statistics.*, vol. 4, no. 1, pp. 266–298, 2010.
- [7] R. Kohavi et al., "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1–2, pp. 273–324, 1997.
- [8] T. Akiba et al., "Optuna: A Next-generation Hyperparameter Optimization Framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining (KDD '19)*, Anchorage, AK, USA, Aug. 2019, pp. 2623–2631.