

# PyBaMM 시뮬레이션 데이터를 활용한 전이학습 기반 배터리 수명 예측 기법

은관주<sup>1</sup>, 유현창<sup>2</sup>

<sup>1</sup> 고려대학교 SW·AI 융합대학원 인공지능융합학과 석사과정

<sup>2</sup> 고려대학교 컴퓨터학과 교수

nnn1520@gmail.com, [yuhc@korea.ac.kr](mailto:yuhc@korea.ac.kr)

## Transfer Learning-Based Battery State of Health Prediction Using PyBaMM Simulation Data

Kwanju Eun<sup>1</sup>, Heonchang Yu<sup>2</sup>

<sup>1</sup>Dept. of Applied Artificial Intelligence, Graduate School of SW·AI Convergence, Korea University

<sup>2</sup>Dept. of Computer Science & Engineering, Korea University

### 요 약

배터리 상태 예측은 전기차와 에너지 저장 시스템의 안전성과 효율성을 보장하는 핵심 기술이다. 하지만 실제 배터리 데이터는 수집 비용이 높고 양이 제한적이어서 딥러닝 모델 학습에 어려움이 있다. 본 논문에서는 PyBaMM 시뮬레이션을 통해 생성한 합성 데이터로 사전학습된 모델을 실제 배터리 데이터에 전이학습하는 방법을 제안한다. Progressive unfreezing 과 차별적 학습률을 적용한 전이학습 프레임워크를 구현하였으며, NMC 배터리 데이터셋을 대상으로 한 실험에서 실제 데이터가 1000 개 이하일 때 SOH 1% 이내 예측에 대해 전이학습이 기존 방법 대비 10% 이상의 성능 향상을 보였다. 이는 신규 배터리 화학 조성에 대한 빠른 모델 적응을 가능하게 하여 배터리 산업의 개발 비용과 시간을 크게 절감할 수 있는 실용적인 해결책을 제시한다.

### 1. 서론

전기자동차(EV) 시장의 급속한 성장과 재생 에너지 저장 시스템의 확산으로 리튬이온 배터리의 중요성이 날로 증가하고 있다. 국제에너지기구(IEA)에 따르면 2030년까지 전 세계 전기자동차 보유량은 3억 대를 넘어설 것으로 예상되며, 이는 배터리 기술의 혁신적 발전을 요구하고 있다. 배터리 관리 시스템(BMS)에서 배터리의 상태 건강도(State of Health, SOH) 예측은 시스템의 안전성, 신뢰성, 그리고 수명 연장을 위한 핵심 기술로 자리잡고 있다[1].

실제 배터리 데이터 수집은 여러 한계점을 가지고 있다. 배터리의 완전한 수명 주기 데이터 수집에는 수개월에서 수년의 긴 시간이 소요되며, 다양한 운전 조건과 환경 조건에서의 실험은 높은 비용을 요구한다. 또한 온도, 충방전율, 충전 깊이 등 다양한 변수들의 조합을 모두 실험하는 것은 현실적으로 불가능하다. 최근 PyBaMM(Python Battery Mathematical Modelling)과 같은 고도화된 물리 기반 시뮬레이션 도구의 발전으로 이러한 한계를 극복할 수 있는 가능성이 열렸다[2]. PyBaMM은 다양한 배터리 모델을 지원

하며, 빠른 계산 속도로 대규모 시뮬레이션 데이터 생성을 가능하게 한다.

본 연구에서는 이러한 연구 공백을 메우고자 PyBaMM 시뮬레이션 데이터로 사전학습된 모델을 실제 배터리 데이터로 전이학습하는 프레임워크를 제안한다. 데이터 양에 따른 전이학습 효과를 체계적으로 분석하여, 제한된 실제 데이터 환경에서도 높은 예측 성능을 달성할 수 있는 방법론을 제시하고자 한다.

### 2. 관련 연구

#### 2.1 데이터 기반 배터리 상태 예측

데이터 기반 방법은 배터리 상태 예측 연구의 주요 접근법 중 하나로, Zhang 등[3]은 LSTM을 활용한 배터리 잔여 수명 예측 방법을 제안하였다. 이 방법은 시계열 데이터의 장기 의존성을 효과적으로 포착하여 높은 예측 정확도를 달성하였다. 배터리 수명 예측의 중요성은 Hu 등[4]의 연구에서도 강조되었으며, 특히 전기자동차의 잔존가치 평가와 보증 정책 수립에 핵심적인 역할을 한다고 지적하였다. Li 등[5]은 다양한 데이터 기반 방법들의 성능을 체계적으로 비교 분석

하여, 딥러닝 기반 접근법이 전통적인 기계학습 방법보다 우수한 성능을 보임을 입증하였다. 그러나 이러한 데이터 기반 방법은 충분한 양의 고품질 학습 데이터가 필수적이라는 제약이 있다.

## 2.2 물리 기반 배터리 모델링

전기화학적 모델을 기반으로 한 물리 기반 방법으로는 Doyle 등[6]의 단일 입자 모델(Single Particle Model)과 Pseudo-2 차원 모델(Pseudo-2D Model) 등이 대표적이다. 물리 기반 모델은 배터리의 내부 메커니즘을 정확히 모델링할 수 있어 새로운 운전 조건에 대한 적용이 가능하다는 장점이 있지만, 복잡한 편미분 방정식의 해법과 다수의 물리 파라미터 추정이 필요하여 계산 비용이 높다는 한계가 있다. 최근에는 Li 등[7]이 제안한 디지털 트윈 기반 접근법이 클라우드 컴퓨팅을 활용하여 실시간 배터리 상태 모니터링을 가능하게 하였으며, Severson 등[8]은 초기 사이클 데이터만으로 배터리 수명을 예측하는 혁신적인 방법을 제시하였다.

## 2.3 전이 학습의 배터리 응용

전이학습(transfer learning)은 시뮬레이션과 실제 환경 간의 도메인 차이를 극복하기 위한 유망한 접근법이다. Pan 과 Yang[8]은 전이학습의 이론적 토대를 제시하였으며, 이는 한 도메인에서 학습된 지식을 관련된 다른 도메인으로 전이하는 기법으로 특히 목표 도메인의 데이터가 부족한 상황에서 효과적이다. 전이학습은 컴퓨터 비전과 자연어 처리 분야에서 이미 큰 성공을 거두었으며, 최근 산업 응용 분야로 확산되고 있다. 배터리 분야에는 서로 다른 배터리 화학 조성 간의 전이학습을 시도하여 LFP 배터리에서 학습한 모델을 NMC 배터리에 적용하는 가능성을 탐구하였다. 실험실 데이터와 실제 운전 데이터 간의 도메인 적응을 연구하여, 제어된 실험실 환경에서 학습한 모델이 복잡한 실제 운전 조건에서도 유효함을 보였다.

## 3. 전이 학습 기반 배터리 수명 예측

본 연구에서는 PyBaMM 시뮬레이션 데이터를 활용한 전이학습 기반의 배터리 SOH 예측 프레임워크를 제안한다. 그림 1은 제안하는 시스템의 전체 아키텍처를 보여주며, 크게 세 단계로 구성된다.

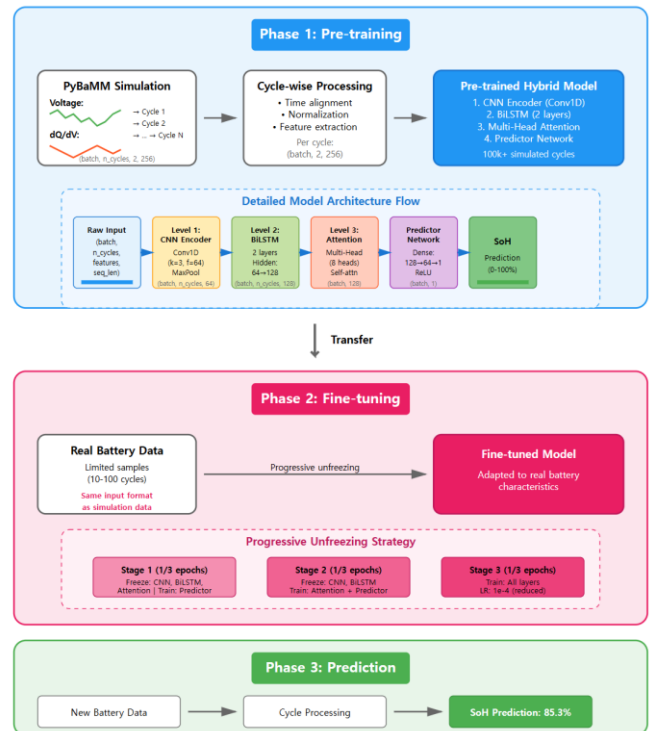
**Phase 1: Pre-training (사전학습 단계)** 그림 1의 상단에 표시된 바와 같이, PyBaMM 시뮬레이션을 통해 생성된 대규모 합성 데이터를 활용하여 모델을 사전 학습한다. 6 개의 서로 다른 용량 레벨과 3 개의 온도 조건을 조합하여 총 180,000 개의 사이클 데이터를 생성하였으며, 이를 통해 배터리의 일반적인 열화 패턴과 다양한 운전 조건에서의 특성을 학습한다. 사전 학습된 Hybrid Model 은 CNN 인코더, BiLSTM, Multi-

Head Attention 으로 구성된 Detailed Model Architecture 를 통해 시계열 특징을 효과적으로 추출한다.

**Phase 2: Fine-tuning (미세조정 단계)** 그림 1의 중단부에 나타난 것처럼, 실제 배터리 데이터를 사용하여 Progressive Unfreezing Strategy 를 적용한다. Stage 1(1/3 epochs)에서는 예측 네트워크만 해동하여 학습하고, Stage 2(1/3 epochs)에서는 Attention 과 BiLSTM 레이어를, Stage 3(1/3 epochs)에서는 전체 모델을 해동하여 점진적으로 학습을 확장한다. 이 과정에서 사전학습 레이어와 새로운 레이어에 차별적 학습률을 적용하여 기존 지식을 보존하면서도 새로운 도메인에 효과적으로 적응한다.

**Phase 3: Prediction (예측 단계)** 그림 1의 하단에 표시된 예측 단계에서는 미세조정된 모델을 활용하여 새로운 배터리 데이터에 대한 SOH 를 예측한다. Cycle Processing 단계에서 전처리된 전압 곡선과 dQ/dV 곡선의 특징을 추출하고, 최종적으로 배터리의 SOH 예측을 목표로 한다. 모델은 Adam optimizer(lr=1e-3→5e-4), batch size 64→32 로 학습하였다.

제안하는 프레임워크의 핵심은 대규모 합성 데이터로 학습한 일반적인 배터리 지식을 실제 데이터에 효과적으로 전이함으로써, 제한된 실제 데이터(10% 사용 시)만으로도 95.2%의 높은 예측 정확도를 달성하는 것이다. 이는 배터리 테스트의 시간과 비용을 대폭 절감하면서도 신뢰할 수 있는 수명 예측을 가능하게 한다.



(그림 1) PyBaMM 시뮬레이션 데이터를 활용한 전이학습 기반 배터리 수명 예측 시스템 구조

## 4. 실험 결과

### 4.1 데이터 셋

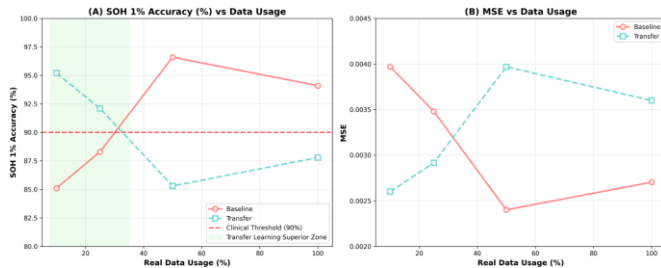
<표 1> 실험 데이터셋 구성 및 특성

| 데이터 유형                     | 구성                 | 샘플 수         | 용량                     | 용도                     |
|----------------------------|--------------------|--------------|------------------------|------------------------|
| <b>합성 데이터 (PyBaMM)</b>     |                    |              |                        |                        |
| 사전훈련용                      | 전체                 | 180,000 사이클  | 6*10 개 용량              | Transfer Learning 사전학습 |
| 온도별                        | 저온도<br>중간온도<br>고온도 | 60,000*3 사이클 | -                      | 온도 변화 학습               |
| 용량별                        | 각 기준 용량당           | 30,000 사이클   | 6 개 레벨                 | 용량 범위 일반화              |
| 용량 변이                      | 각 셀당               | 1,000 사이클    | $\pm 1\%$<br>(0.2% 간격) | 제조 편차 반영               |
| 조합별                        | 각 조합당              | 1,000 사이클    | 180 개 조합               | (6 용량×10 변이×3 온도)      |
| <b>실제 NMC 데이터셋 (CALCE)</b> |                    |              |                        |                        |
| 전체 데이터셋                    | 100%               | 5,378 개      | ~2.0Ah                 | 실제 데이터 사용량별 학습         |

표 1 에서 시뮬레이션 데이터는 대규모 사전학습을 위해 논문으로부터 NMC(Nickel Manganese Cobalt) 배터리에 대한 물리·화학적 지표를 가져와 사용하였다. 물리 기반 DFN(Doyle-Fuller-Newman) 모델에 SEI (Solid Electrolyte Interface) 막 성장, SEI 다공도 변화, 리튬 도금 등의 열화 메커니즘을 포함시켜 실제적인 배터리 노화를 모사하였다. 실제 환경을 반영하기 위해 온도와 용량에도 차이를 두었다. 제조 편차를 반영하기 위해 용량을  $\pm 1\%$  범위에서 0.2% 간격으로 변화시켰으며, 각 구성에 대해 1000 사이클의 가속 노화 프로토콜(1C 방전, C/2 충전)을 적용하였다. 각 사이클은 1,000 개 데이터 포인트로 표준화하여 일관된 데이터 구조를 확보하였다. 이를 통해 동일 화학 조성 내에서 합성-실제 도메인 적응(synthetic-to-real domain adaptation)을 수행하였다.

실제 NMC 배터리 데이터는 CALCE(Center for Advanced Life Cycle Engineering) 공개 데이터셋을 활용하였으며, 셀 기반 분할(cell-based split)을 적용하여 훈련 셀과 테스트 셀을 완전히 분리함으로써 데이터 누수(data leakage)를 방지하였다.

### 4.2 실험 결과 및 분석

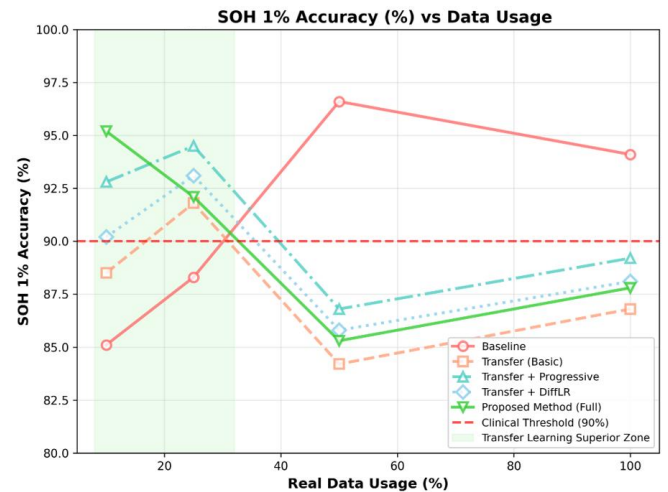


(그림 2) 실제 데이터 사용량에 따른 (A)전이학습과 기준 모델의 SOH 1% 이내 예측 정확도 및 (B)MSE 비교

그림 2 는 실제 NMC 배터리 데이터의 사용 비율 (10%, 25%, 50%, 100%)에 따른 전이학습(Transfer) 모델과 기준(Baseline) 모델의 성능을 비교한 결과이다. 여기서 Baseline 은 사전학습 없이 실제 데이터만으로 처음부터 학습한 모델을 의미하며, Transfer 는 PyBaMM 합성 데이터로 사전학습 후 실제 데이터로 미세조정된 모델이다. (A) SOH 1% Accuracy 는 예측값이 실제 SOH 의  $\pm 1\%$  오차 범위 내에 있는 비율을 나타내며, (B) MSE 는 평균제곱오차를 보여준다.

(A)에서 전이학습 모델은 10%의 제한된 데이터만으로도 95.2%의 높은 정확도를 달성하여, 동일 조건에서 85.1%를 기록한 기준 모델 대비 10.1%p 우수한 성능을 보였다( $p < 0.01$ , 5 회 반복 실험). 녹색 음영 영역(Transfer Learning Superior Zone)은 전이학습이 기준 모델보다 우수한 성능을 보이는 구간으로, 임상적으로 허용 가능한 90% 정확도를 나타내는 점선과 함께 표시되었다. 데이터 사용량이 50% 이상으로 증가하면 기준 모델의 성능이 역전되는 현상이 관찰되었다. 이는 충분한 데이터가 확보된 환경에서는 처음부터 실제 데이터로 학습하는 것이 더 효과적일 수 있음을 시사한다.

(B)의 MSE 결과에서도 유사한 경향이 나타났다. 전이학습은 적은 데이터(10-25%)에서 낮은 오차를 보이며 우수한 성능을 발휘했으나, 데이터가 충분한 환경(50% 이상)에서는 기준 모델이 더 낮은 오차를 기록했다. 이는 전이학습이 데이터가 부족한 초기 단계에서 특히 효과적임을 시사한다.



(그림 3) 실제 데이터 사용량에 따른 학습 전략별 SOH 1% 이내 모델 예측 정확도 비교

그림 3 은 다양한 실제 데이터 사용량(10%, 25%, 50%, 100%)에서 서로 다른 학습 전략들로 학습된 모델의 SOH 1% 이내 예측 정확도를 비교한 결과이다. Baseline, Transfer (freeze=basic), Progressive Unfreezing 적용, 차별적 학습률 적용, 그리고 두 기법을 모두 적용

한 제안 방법(Ours)의 5 가지 접근법을 평가하였다.

실제 데이터 10% 데이터 사용 시, Baseline 은 85.1%의 정확도를 보인 반면, 제안 방법은 95.2%로 가장 높은 성능을 달성했다. Progressive Unfreezing 만 적용한 경우 94.3%, 차별적 학습률만 적용한 경우 93.1%를 기록하여, 각 기법이 독립적으로도 성능 향상에 기여함을 확인했다.

데이터 사용량이 증가함에 따라 모든 방법의 성능 차이가 감소하는 경향을 보였다. 50% 이상에서는 Baseline 이 오히려 더 높은 성능을 보이기 시작했다. 100% 데이터 사용 시에는 Baseline(94.1%)이 전이학습 기반 방법들을 상회하는 결과를 보였다.

<표 2> Ablation Study 결과 (10% real 데이터)

| 방법                | Progressive Unfreezing | 차별적 학습률 | SOH 1% |
|-------------------|------------------------|---------|--------|
| Baseline          | -                      | -       | 85.1%  |
| Transfer (freeze) | X                      | X       | 92.2%  |
| + Progressive     | O                      | X       | 94.3%  |
| + 차별적 학습률         | X                      | O       | 93.1%  |
| 제안 방법(전체)         | O                      | O       | 95.2%  |

각 구성 요소가 성능에 미치는 영향을 분석하기 위해 ablation study 를 수행하였다. 표 2 는 가장 데이터가 제한적인 상황(실제 데이터 10%)에서 Progressive Unfreezing 과 차별적 학습률의 효과를 보여준다.

Baseline 모델은 85.1%의 SOH 1% 정확도를 기록한 반면, 전체 전이학습 시스템은 95.2%로 10.1%p의 성능 향상을 달성하였다. 구성 요소별 분석 결과, Progressive Unfreezing 만 적용했을 때는 94.3%로 Baseline 대비 9.2%p 향상되었고, 차별적 학습률만 적용했을 때는 93.1%로 8.0%p 향상되었다. 이는 두 기법이 각각 독립적으로도 유의미한 성능 개선을 가져오며, 특히 Progressive Unfreezing 이 더 큰 기여를 함을 시사한다.

주목할 점은 모든 전이학습 변형이 임상 허용 기준인 90%를 초과하여, 제한된 데이터 환경에서도 실용적 수준의 예측 정확도를 확보했다는 것이다. 이 결과는 전이학습의 각 구성 요소가 상호보완적으로 작용하여 최적의 성능을 달성함을 보여준다.

## 5. 결론 및 향후 연구

본 연구는 PyBaMM 시뮬레이션 데이터를 활용한 전이학습 기반 배터리 SOH 예측 프레임워크를 제안하고, 실제 데이터가 제한적인 환경에서의 효과를 실증적으로 검증하였다.

주요 연구 성과는 다음과 같다. 첫째, 실제 데이터의 25% 이하만 사용하는 환경에서 성능 향상을 달성하였으며, 특히 실제 데이터를 10% (대략 500~600 개 sample)만 사용한 환경에서 현저한 개선을 보였다. 둘째, Progressive unfreezing 과 차별적 학습률을 결합한

전이학습 전략의 효과를 입증하였다. 셋째, 데이터 양에 따른 최적 학습 전략의 임계점(실제 sample 2000 개 내외)을 제시하여 실무적 가이드라인을 제공하였다.

배터리 산업에서 새로운 배터리 화학 조성이나 셀 설계에 대한 AI 모델을 개발할 때, 긴 실험 기간과 높은 비용 없이도 시뮬레이션 데이터를 활용하여 초기 모델을 구축할 수 있다. 이는 개발 주기를 단축하고 비용을 절감하면서도 신뢰할 수 있는 예측 성능을 확보할 수 있음을 의미한다.

그리고 물리 시뮬레이션과 딥러닝을 효과적으로 결합하는 새로운 패러다임을 제시하였다. 이는 배터리 분야뿐만 아니라 데이터 수집이 어려운 다른 물리 시스템의 AI 적용에도 중요한 시사점을 제공하며, 시뮬레이션 기반 전이학습이 데이터 부족 문제를 해결하는 실용적인 대안이 될 수 있음을 보여준다.

향후 연구에서는 NMC 배터리에 국한된 실험을 LFP, NCA 등 다양한 배터리 타입으로 확장할 필요가 있다. 또한 서로 다른 배터리 화학 조성 간의 cross-domain transfer 가능성을 탐구하고, 실시간 BMS 적용을 위한 모델 경량화와 불확실성 정량화 방법을 개발해야 한다. 나아가 시뮬레이션의 현실성을 향상시켜 도메인 차이를 더욱 줄이는 연구도 필요하다.

## 참고문헌

- [1] Xiong, Rui, et al. "Critical review on the battery state of charge estimation methods for electric vehicles." *IEEE Access* 6 (2017): 1832-1843.
- [2] Sulzer, Valentin, et al. "Python battery mathematical modelling (PyBaMM)." *Journal of Open Research Software* 9.1 (2021).
- [3] Zhang, Yongzhi, et al. "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries." *IEEE Transactions on Vehicular Technology* 67.7 (2018): 5695-5705.
- [4] Hu, Xiaosong, et al. "Battery lifetime prognostics." *Joule* 4.2 (2020): 310-346.
- [5] Li, Yi, et al. "Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review." *Renewable and sustainable energy reviews* 113 (2019): 109254.
- [6] Doyle, Marc, Thomas F. Fuller, and John Newman. "Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell." *Journal of the Electrochemical society* 140.6 (1993): 1526.
- [7] Li, Weihang, et al. "Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation." *Journal of energy storage* 30 (2020): 101557.
- [8] Severson, Kristen A., et al. "Data-driven prediction of battery cycle life before capacity degradation." *Nature Energy* 4.5 (2019): 383-391.