

포트폴리오

한양대학교 산업경영공학과

석사 3기 김태이



About me



Taeyi Kim

M.S. Student
IODA Lab
Hanyang Univ.

Email: taeyii3056@hanyang.ac.kr
Homepage: <https://taeyii.github.io/>

EDUCATION

- 한양대학교 석사 산업경영공학과 (2024 - 현재) 지도교수: Barde Stephane
- 한양대학교 ERICA 학사 산업경영공학과 (2022 - 2024)

RESEARCH INTERESTS

- Prognosis and Health Management (PHM)
- Time-Series Forecasting
- Edge AI & Model Compression (Knowledge Distillation, Quantization, Pruning)

SKILLS

Language: Python, C++
Package: Pytorch, Pandas, Numpy, ONNX

EXPERIENCE

- Research Intern, KITECH (2024.07-08)- CNC 공구 마모 예측 알고리즘 개발 Research Intern,
- EdgeCross AI (2023.07-08)- OPC UA 기반 데이터 수집 펌웨어 개발

HIGHLIGHTS

3 Awards 🏆 4 Papers 📄 1 Patent 📑

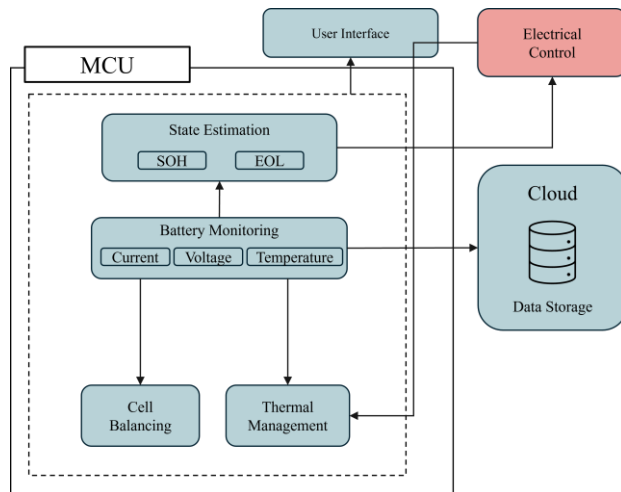
- 2025.12 Factory Hack Korea 2025 최우수상 (산업통상자원부)
- 2025.11 Journal of Energy Storage (IF:9.8, JCR기준 상위 10%) 논문 게재 (1저자)
- 2025.10 Energy (IF:9.4, JCR기준 상위 4%) 논문 게재 (2저자)
- 2025.10 IEEE Global Reliability & PHM Conference Best Paper Award
- 2025.10 배터리 예지용 통합 압축 파이프라인 특허 출원
- 2025.06 한국신뢰성학회 춘계학술대회 우수발표 논문상

KEY ACHIEVEMENTS

- **83.5x** 모델 압축 달성 (451KB → 5.4KB)
- **18.7x** 처리량 향상
- Raspberry Pi 실시간 추론 구현

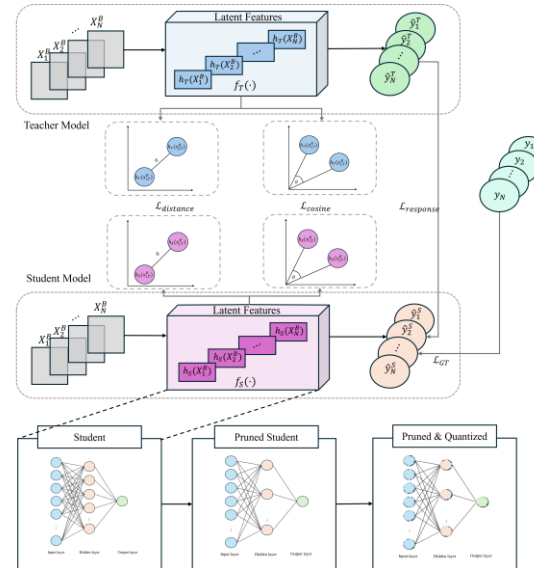
Edge-compatible Battery SOH Estimation

Problem definition



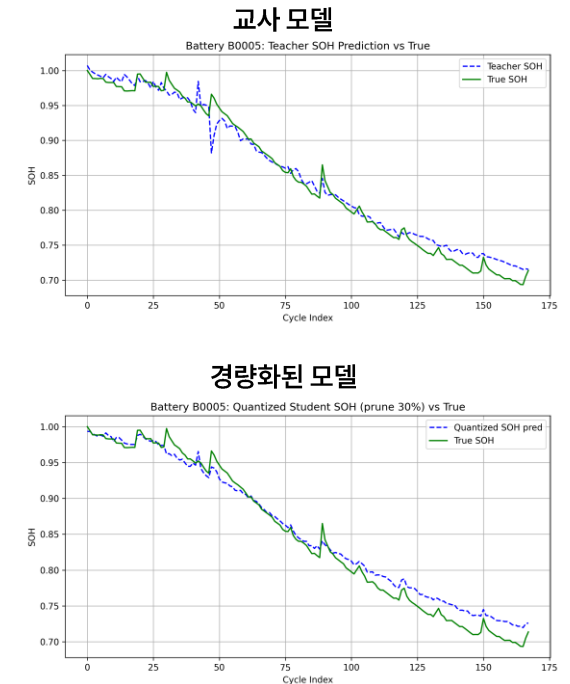
- 전기차 배터리 관리 시스템에 고성능의 AI 모델을 적용하는데 컴퓨팅 자원의 한계로 문제가 있음

Proposed method



- Knowledge Distillation을 사용해 큰 교사 모델을 작은 학생 모델로 압축
- 이후 압축된 학생 모델에 대해 프루닝과 양자화를 추가적으로 적용

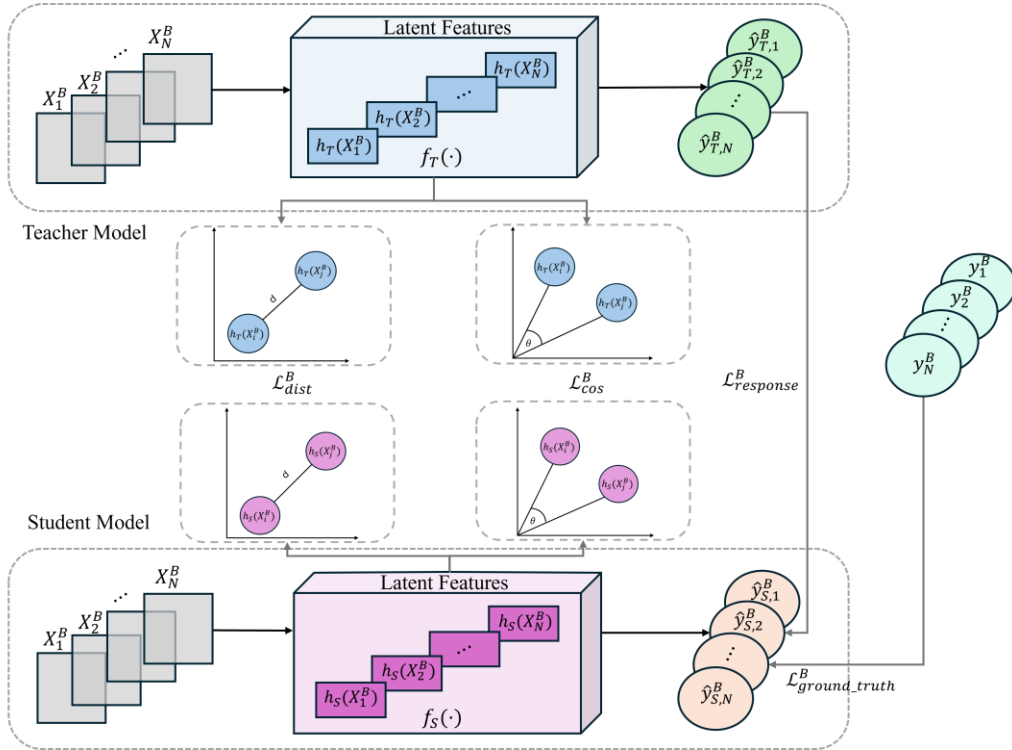
Result



- 모델 용량을 99%까지 줄이면서 성능은 유지

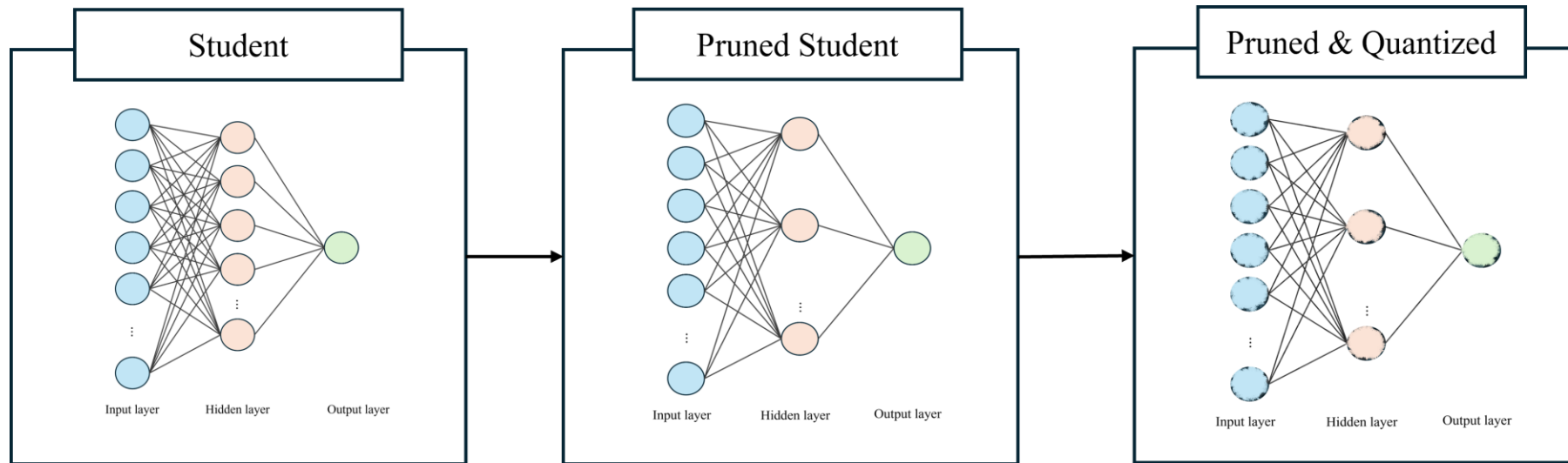
Edge-compatible Battery SOH Estimation

Main idea



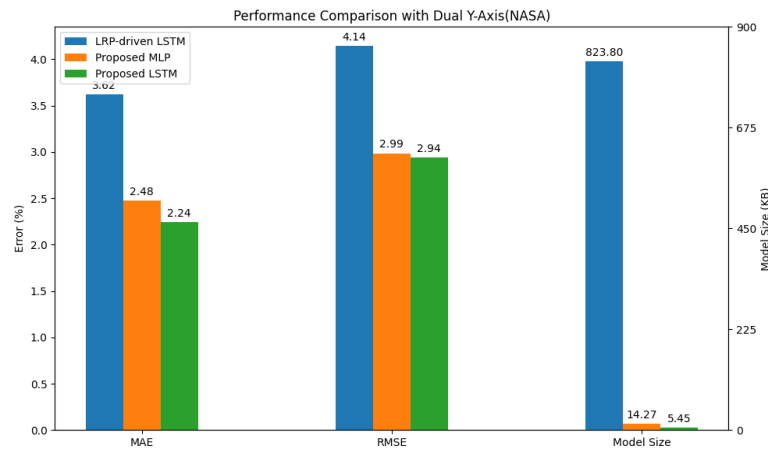
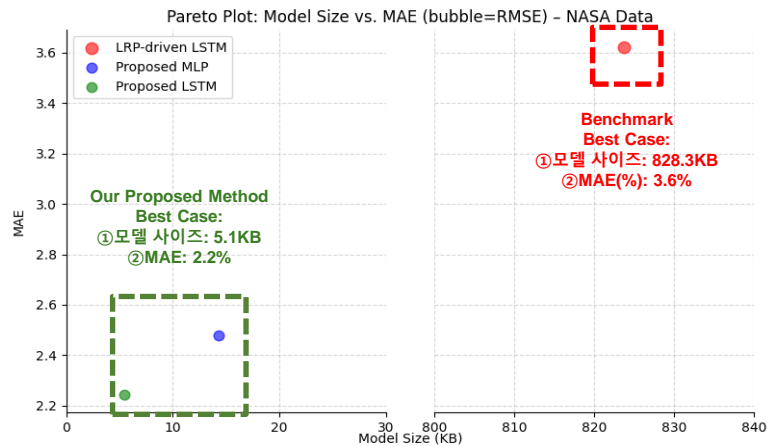
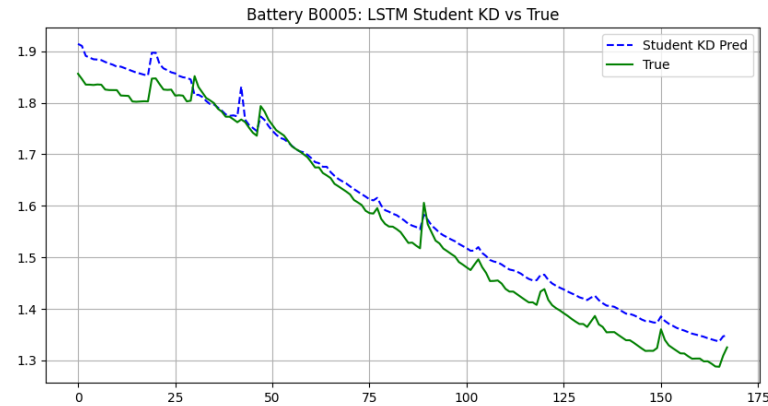
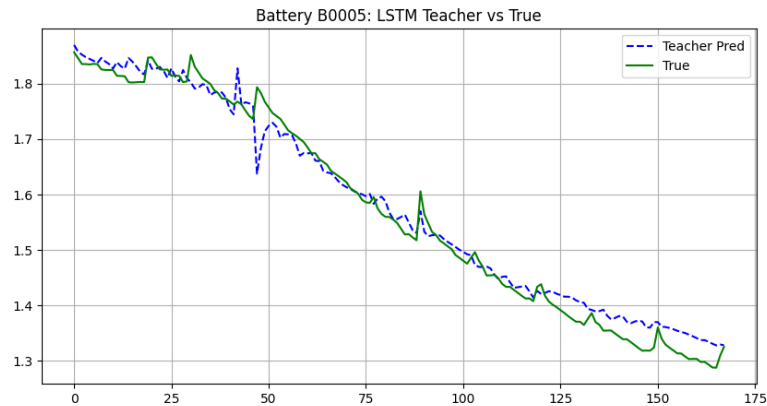
- Response-based Knowledge Distillation(KD) 뿐만 아니라 교사 모델의 latent feature 간의 관계를 모방하는 Relation-based Knowledge Distillation(KD)을 수행
- 학생 모델은 교사모델의 hidden feature간의 거리, 코사인 유사도를 모방
- $D_T(i, j) = \|h_T(X_i) - h_T(X_j)\|^2, D_S(i, j) = \|h_S(X_i) - h_S(X_j)\|^2$
- $S_T(i, j) = \frac{h_T(X_i) \cdot h_T(X_j)}{\|h_T(X_i)\| \|h_T(X_j)\|}, S_S(i, j) = \frac{h_S(X_i) \cdot h_S(X_j)}{\|h_S(X_i)\| \|h_S(X_j)\|}$
- $\mathcal{L}_{dist} = \frac{2}{M} \sum_{(i,j) \in P} (D_T(i, j) - (D_S(i, j)))^2$,
- $\mathcal{L}_{cos} = \frac{2}{M} \sum_{(i,j) \in P} (S_T(i, j) - (S_S(i, j)))^2$,
- $\mathcal{L}_{total} = \mathcal{L}_{ground_truth} + \lambda_{response} \mathcal{L}_{response} + \lambda_{dist} \mathcal{L}_{dist} + \lambda_{cos} \mathcal{L}_{cos}$.

Compression technique



- 모델을 더욱 더 압축하기 위해, 순차적으로 두 가지 경량화 기법을 추가
- KD를 끝낸 학생 모델에 구조적 프루닝 적용 후, 동적 양자화 까지 적용
- 총 3단계 압축 프레임워크를 통해 최대 99%까지 모델 압축 가능

Model comparison

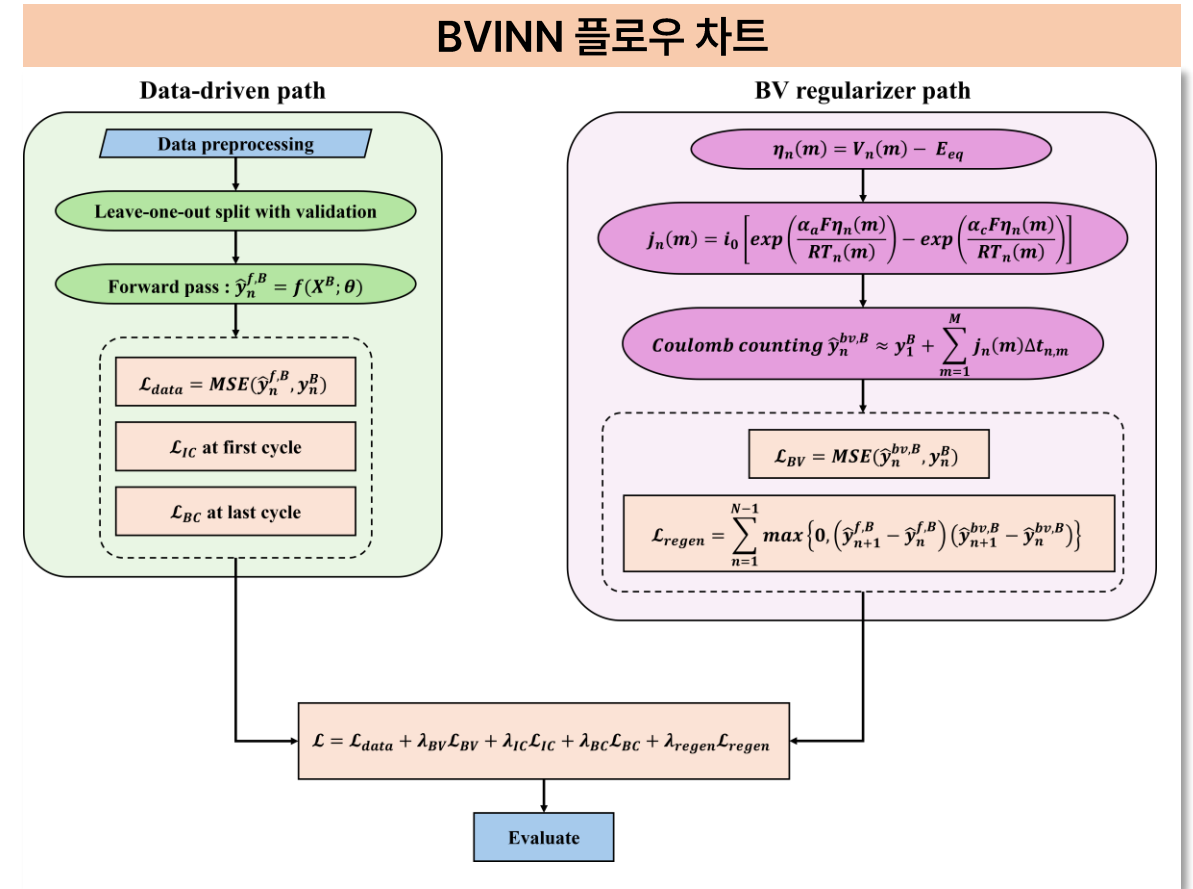


SCI급 저널 모델 대비,
더 가벼우면서 우수한
성능을 달성함

방법론

본 연구에서는 BVINN (Butler-Volmer Informed Neural Network)을 제안하여, 전기화학적 원리를 신경망 학습 과정에 직접 반영함. BVINN은 Butler-Volmer 방정식을 기반으로 한 물리 정보 정규화 항을 손실 함수에 포함시켜, 모델이 물리적으로 타당한 예측을 학습하도록 유도함.

- 데이터 손실(\mathcal{L}_{data}): 실제 데이터와 예측 값의 차이를 최소화
- Butler-Volmer 손실(\mathcal{L}_{BV}): Butler-Volmer 방정식 기반 전류-용량 관계와 일치하도록 제약
- 초기 조건 손실(\mathcal{L}_{IC}): 첫 cycle에서의 초기 값 보장
- 경계 조건 손실(\mathcal{L}_{BC}): 마지막 cycle에서의 경계 조건 반영
- 재생 손실(\mathcal{L}_{regen}): 배터리의 일시적 용량 회복(Capacity Regeneration)을 반영



Hybrid Compression for EOL Prediction on Edge BMS

방법론

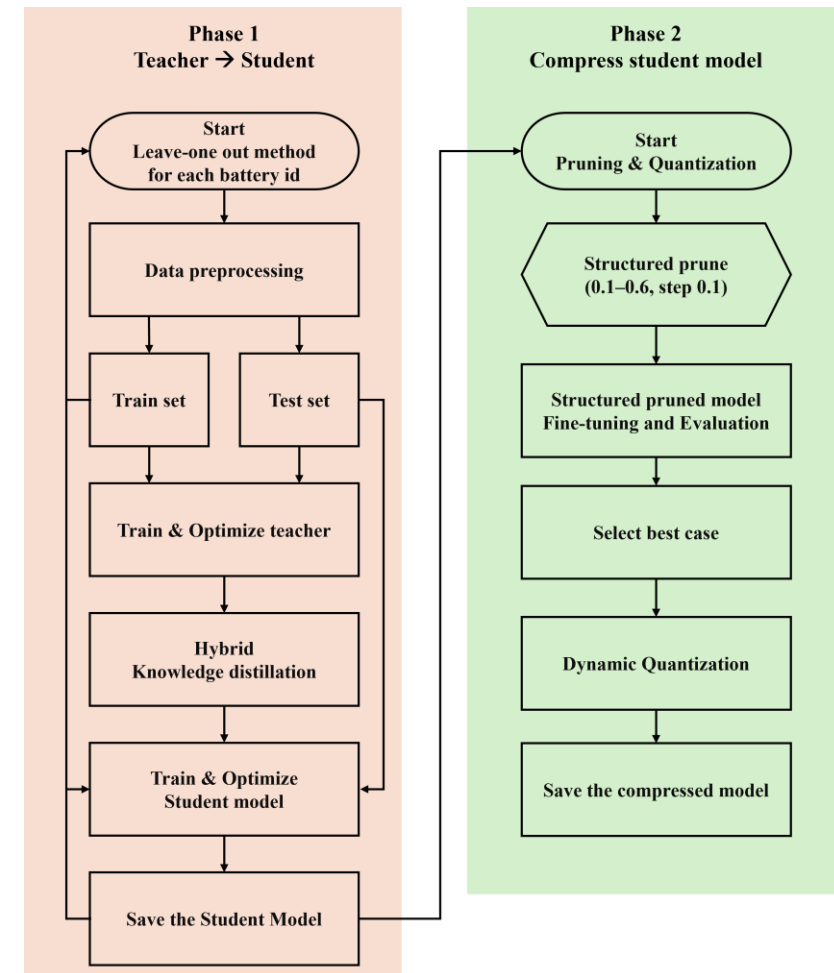
본 연구에서는 Knowledge Distillation (KD), Structured Pruning, Dynamic Quantization을 결합한 3단계 압축 파이프라인을 제안하여, 고용량 Teacher 모델을 엣지 디바이스에 배포 가능한 경량 Student 모델로 변환함

Phase 1: Teacher → Student (Knowledge Distillation)

- Leave-one-out 교차검증 기반으로 Teacher LSTM 모델 학습
- Response-based KD와 Relation-based KD를 결합한 Hybrid KD 적용
- Teacher의 출력값뿐 아니라 latent feature 간의 거리, 코사인 유사도까지 Student가 학습

Phase 2: Compress Student Model

- Structured Pruning (10%~60%): 중요도가 낮은 hidden unit을 구조적으로 제거하여 모델 경량화
- Fine-tuning 후 최적 pruning ratio 선택
- Dynamic Quantization: 32-bit → 8-bit 변환으로 추가 압축



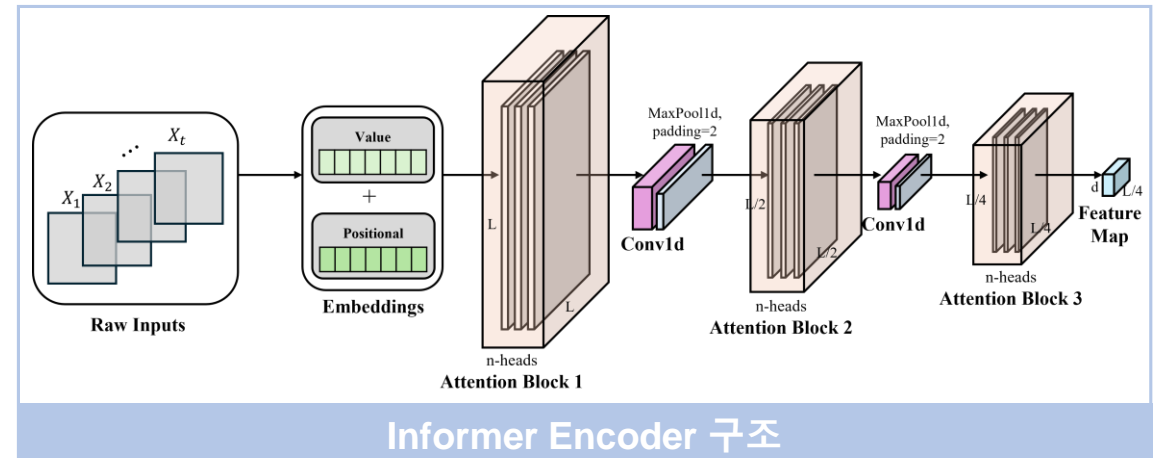
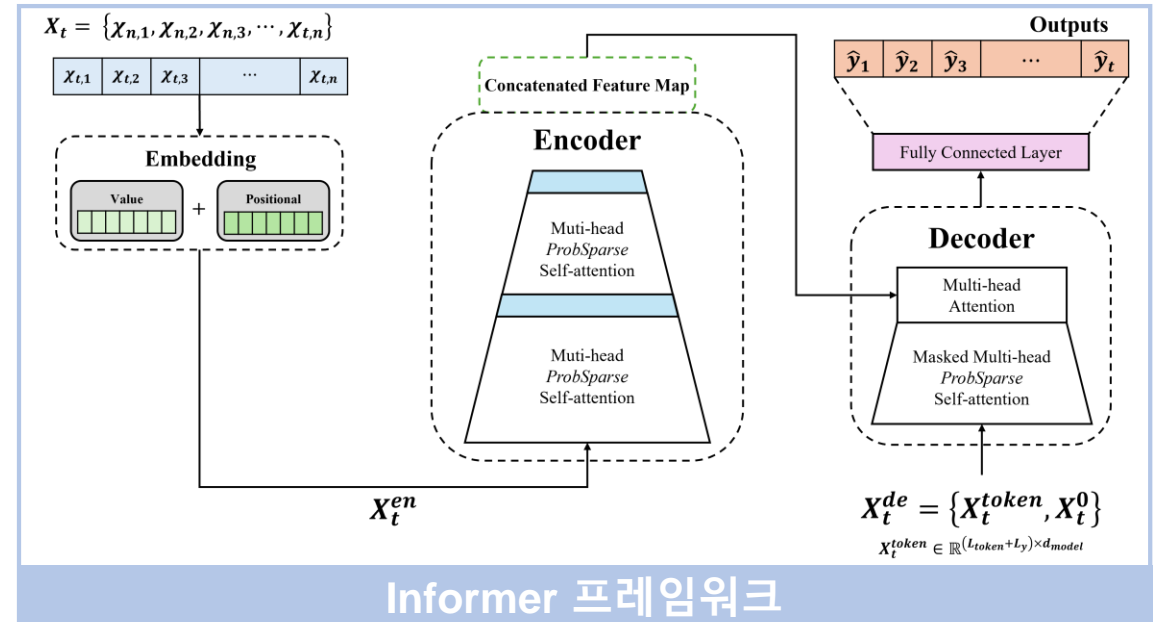
모델 경량화 플로우차트

SOH Prediction via ProbSparse Informer Architecture

방법론

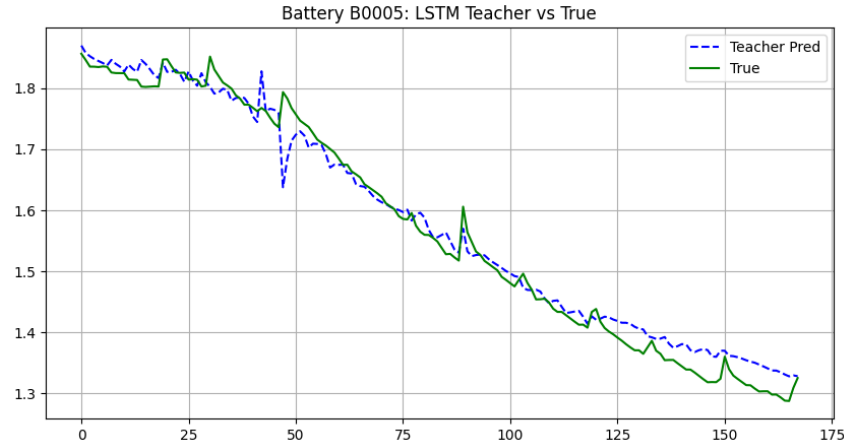
본 연구에서는 Informer 기반 SOH 추정 프레임워크를 제안하여, 긴 시계열 데이터를 효율적으로 학습할 수 있도록 ProbSparse Attention과 Self-Attention Distilling 기법을 신경망 구조에 직접 반영함.

- ProbSparse Self-Attention: 일반 Transformer의 계산 복잡도를 줄이고 긴 시퀀스에서도 효율적인 학습을 위해 ProbSparse Attention을 적용하여, 중요한 시점 정보에 집중할 수 있도록 설계함.
- Informer Encoder: 다중 distilling stack을 병렬로 배치하여 세밀한 단기 패턴과 장기 의존성을 동시에 학습함. 이를 통해 긴 시퀀스에서도 안정적으로 특징을 추출할 수 있도록 함.
- Informer Decoder: Masked ProbSparse Attention을 적용해 미래 시점 정보를 차단하면서, Encoder의 출력을 결합하여 SOH 추정 값을 생성함.



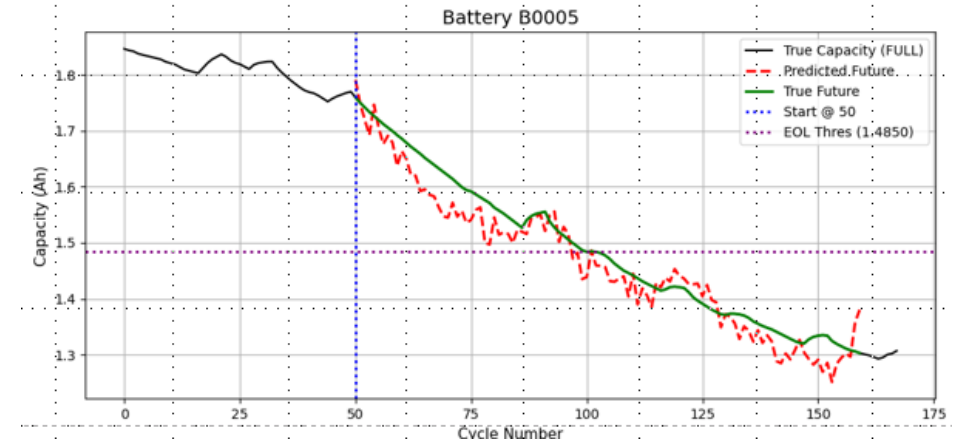
Ongoing Paper: RUL Forecasting

Prediction



- $p(y_{1:t}|X_{1:t}) = p(y_1, y_2, \dots, y_t | X_1, X_2, \dots, X_t)$
- 현재까지 연구한 SOH prediction 태스크에서는 현재 시점의 센서 값들을 바탕으로 현재 시점의 타겟 값(배터리 capacity)을 예측함

Forecasting



- $p(y_{t+1}, y_{t+2}, \dots, y_T | y_1, y_2, \dots, y_t, X_1, X_2, \dots, X_t)$
- 진행중인 연구에서는 더 난이도 높은 task인 현재 시점까지의 센서 값과 타겟값을 바탕으로 미래 시점의 타겟값을 예측하는 time-series forecasting을 연구할 예정임