


## 1차년도 기술문서

(과제명) 대규모 분산 에너지 저장장치 인프라의 안전한 자율운영  
및 성능 평가를 위한 지능형 SW 프레임워크 개발  
(과제번호) 2021-0-00077

- 기술문서명 : 인공지능 기반 배터리 데이터 분석기법 기술 보고서
- 작성일자 : 2021년 12월 30일

과학기술정보통신부 SW컴퓨팅산업원천기술개발사업  
“기술문서”로 제출합니다.

수행기관	성명/직위	확인
서울대학교 산학협력단	강명주/교수	

정보통신기획평가원장 귀하

# 목차

• 배터리 데이터 선행연구 분석 -----	1
• 배터리 데이터 전처리 모델 -----	27
• 배터리 SOH 추정 모델 -----	34
• 배터리 이상탐지 모델 -----	40

---

# 배터리 데이터 선행연구 분석

---

# 배터리 초기 사이클 데이터를 이용한 배터리 수명 예측 기법

## 1. 개요

주어진 운영 조건에 대하여 리튬-이온 배터리 수명을 정확하게 예측하는 기술의 개발은 그 활용도가 매우 높으며, 배터리 성능 최적화 기술의 핵심에 있다. 하지만 배터리 성능 저하가 선형적으로 이루어지지 않고 매우 가변적이기 때문에 운영 조건이 동일한 상황에서조차도 이를 잘 예측하는 것은 매우 어렵다. 그림 1에서는 본 논문에서 사용한 데이터에 대한 각 배터리별 방전 용량 그래프를 표시하였는데, 운영 조건에 따라 방전 용량 그래프가 매우 달라지는 것을 확인할 수 있으며, 초기 주기의 감소량에 비해 후반이 훨씬 큰 것을 확인할 수 있다.

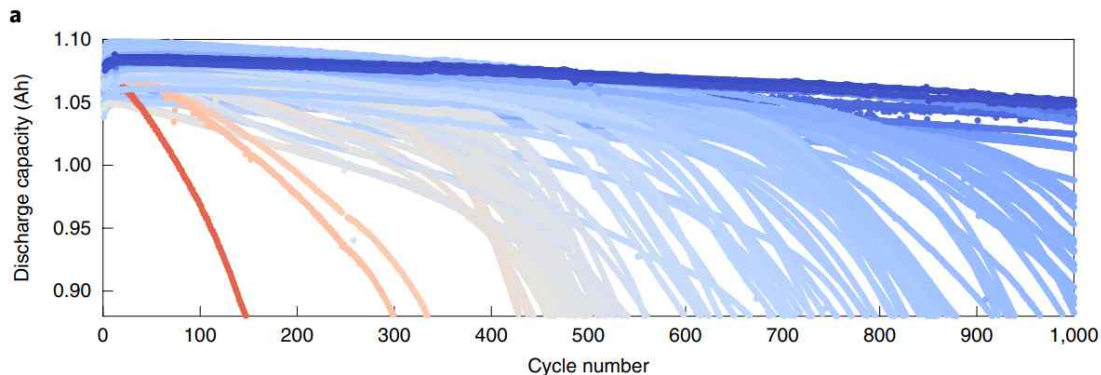


그림 2 다양한 운영 조건 하에서 리튬 인산 철/흑연 배터리의 주기별 방전 용량 그래프

실제로 전체 셀들 중 약 81% 정도는 100번째 주기에서 방전 용량이 초기에 비해 오히려 증가하였으며, 이는 초기 주기의 용량 감소 곡선만으로는 배터리의 수명을 예측하기 어려울 것임을 시사한다. 그림 2에서는 각각 2번째 주기의 방전 용량 (d), 100번째 주기의 방전 용량 (e), 그리고 95~100번째 주기 사이 방전 용량 그래프의 기울기 (f)와 배터리의 수명 간의 상관관계를 확인하였는데, 이들은 전부 낮은 상관관계를 보였다.

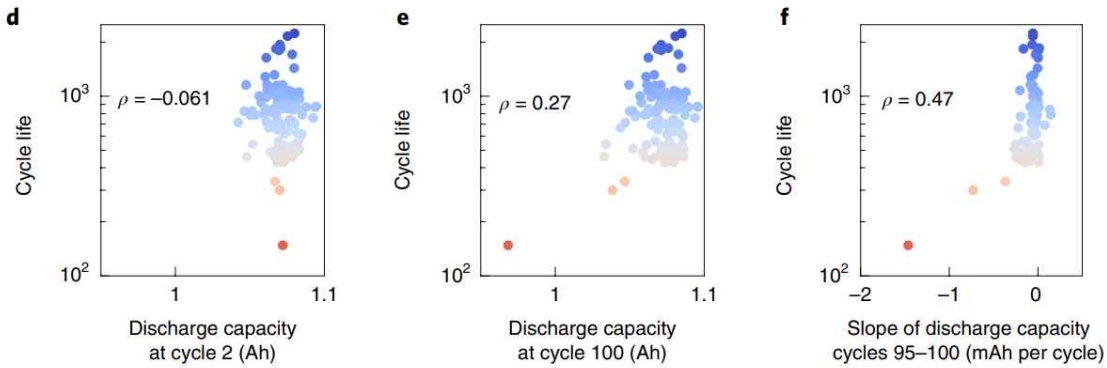


그림 3 배터리의 초기 방전 용량과 배터리 수명 간의 상관관계

배터리의 성능 저하 진단에 보다 효과적인 데이터는 방전 시 전압 변화에 따른 배터리 용량 변화 그래프를 포괄적으로 봄으로써 얻을 수 있다. 같은 운영 조건 하에서는 전압의 변화 정도가 일정하며, 따라서  $N$ 번째 주기에서의 전압 별 용량 그래프를  $Q_N(V)$ 로 표현할 수 있다. 그림 3의 (b)에서는  $\Delta Q_{100-10}(V) = Q_{100}(V) - Q_{10}(V)$ 의 그래프를 표시하였는데, 이를 보면 전압의 양 끝에서는 용량의 변화가 적더라도 그 사이에서 변화가 큰 것을 확인할 수 있으며, 이 변화는 수명이 짧은 배터리일수록 더 커진다. 실제로 (c)에서  $\Delta Q_{100-10}(V)$ 의 분산과 배터리 수명 간의 상관관계를 보면 매우 높은 음의 상관관계를 가지는 것을 확인할 수 있다.

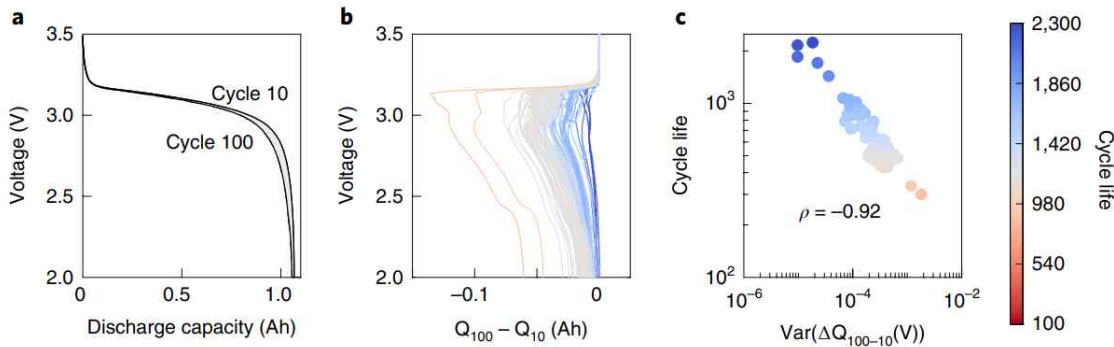


그림 4 10번째, 100번째 주기의  $Q(V)$  그래프 차이의 분산과 배터리 수명 간의 상관관계

본 논문에서는 수명 감소 메커니즘에 대한 수학적 지식 없이 초기 주기의 배터리 성능 저하 데이터만으로 배터리의 수명을 정확하게 예측하는 데이터 기반 모델을 개발하고자 하였다. 이를 위하여 배터리의 성능 저하에 영향을 끼칠 것으로 생각되는 다양한 특징들을 추출한 이후 이들의 선형 결합 형태로 배터리의 수명을 예측하였다. 학습에는 기계 학습 중 선형 모델에 해당하는 elastic net이 사용되었다.

## 2. 기법

### ○ 특징 추출

본 논문에서는 배터리의 수명을 예측하기 위해 배터리 수명과 관련이 있을 것으로 생각되는 다양한 특징들을 추출해낸 이후에, 이들의 선형 결합 형태로 배터리 수명 주기의 로그 값을 근사하였다. 즉  $\hat{y}_i$ ,  $\hat{w}$ ,  $x_i$ 를 각각  $i$ 번째 셀의 예상 로그 주기, 모델의 가중치,  $i$ 번째 셀의 특징들이라 할 때  $\hat{y}_i = \hat{w}^T x_i$ 의 형태를 가지는 모델을 사용하였다.

본 논문에서는 총 20종류의 특징들이 연구되었는데, 이중 6종류는  $\Delta Q_{100-10}(V)$  곡선에 대한 통계량들이며(평균, 분산 등) 7종류는 방전 용량 그래프에 대한 통계량들이고(2~100주기 사이 곡선의 기울기 등) 나머지 7종류는 그 외의 평균 충전 시간, 셀 온도, 측정된 내부 저항 값 등을 포함한다. 실제 모델에는 이들을 전부 사용하는 대신 이들 중 일부만이 사용되었다.

### ○ Elastic net

선형 모델의 가중치를 학습할 때 과적합 문제를 해결하기 위한 다양한 regularization 방법이 존재한다. 이들 중 대표적인 것은 L2 norm을 사용하는 ridge regression, L1 norm을 사용하는 lasso regression, 그리고 둘의 혼합 버전인 elastic net이다.

Lasso의 경우 필요 없는 특징들의 가중치를 0으로 만들어 버리는 특성을 가져 학습 결과 나오는 가중치에 sparsity를 주어 모델이 단순해진다는 장점이 있지만 여러 특징들 사이에 상관관계가 있는 경우 이들 중 하나만을 선택하여 정확도가 떨어진다는 단점이 있으며, ridge regression의 경우 상관관계가 있는 특징들 사이에 적절히 가중치를 분배한다는 장점이 있지만 모델이 복잡해지는 단점이 있다.

Elastic net은 둘의 장점을 모두 지니기 위해 고안된 방법으로, 전체 특징들 중 일부만을 선택하여 이들에게 적절한 가중치를 부여한다. Elastic net을 이용하여 모델의 가중치를 다음과 같이 구할 수 있다.

$$\hat{w} = \arg \min_w \|y - Xw\|_2^2 + \lambda P(w), \quad P(w) = (1 - \alpha)\|w\|_2^2 + \alpha\|w\|_1$$

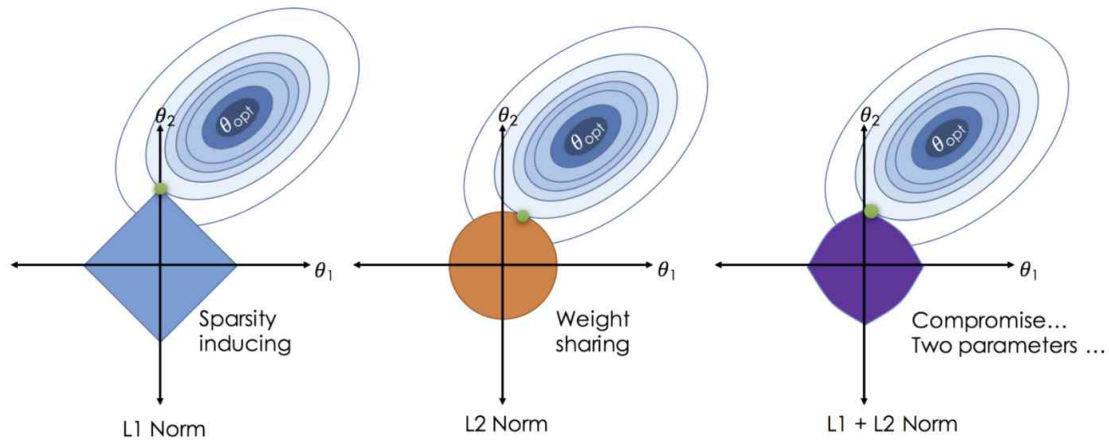


그림 5 Lasso regression, ridge regression, elastic net의 비교

### 3. 실험

#### ○ 데이터

본 실험에서는 서로 다른 124개의 상용 리튬 인산 철/흑연 셀들을 30도로 유지되는 챔버 내에서 다양한 급속 충전 조건으로 운영하여 이를 학습 및 검증 데이터로 활용하였다. 이때 챔버의 온도가 조절되더라도 각 셀의 온도는 충/방전시의 열에 의해 변화할 수 있는데, 이를 기록하여 기계학습에서 사용할 특징으로 활용할 수 있도록 하였다.

사용한 셀들의 규격 용량은 1.1Ah이며 (즉 1C는 1.1A로 충전함을 의미한다) 규격 전압은 3.3V이고 제조업체에서 권장하는 급속 충전 조건은 3.6C로 유지하는 CC-CV 충전이다. 실험 중 모든 셀들에 대하여 방전 조건은 동일하게 두었는데, 우선 4C를 유지하여 2V까지 떨어뜨렸으며, 이후에 전류가 0.02C까지 떨어질 때까지 이를 유지하는 CC-CV 방식을 사용하였다.

충전의 경우 서로 다른 72 종류의 충전 방식을 도입하였다. 우선 state-of-charge(SOC)가 0~80%까지 증가하는 동안은 3.6~6C 사이의 충전율을 유지하였는데, 이때 충전 규약은 서로 같거나 다른 두 종류의 충전율을 순차적으로 사용하는 방식을 사용하였다. (예시: 0~50%까지는 6C의 충전율, 50~80%까지는 4C의 충전율을 사용) 이후에는 모든 셀들에 대해 동일하게 1C를 유지하여 3.6V까지 충전하였으며 방전 때와 마찬가지로 전류가 0.02C에 다다를 때까지 이를 유지하는 CC-CV 방식을 사용하였다.

각 셀의 수명은 용량이 규격 용량의 80%까지 줄어든 시점의 주기로 정의하였는

데, 실험 결과 최종적으로 얻어지는 셀들의 수명은 150~2300 주기 사이로, 평균은 약 806 주기, 표준편차는 약 377 주기로 나타났으며 총 데이터 셋의 주기는 96,700 주기에 달한다.

이들은 각각 학습 데이터와 두 종류의 검증 데이터로 나뉘는데, 41개 셀로 이루어진 학습 데이터와 43개 셀로 이루어진 1차 검증 데이터는 2017년도 5~6월에, 40개 셀로 이루어진 2차 검증 데이터는 모델 개발이 끝난 이후인 2018년도 4월에 만들어져 추가적인 검증 용도로 활용하였다.

#### ○ 실험 방법

본 논문에서는 총 두 종류의 실험이 진행되었는데, 첫 번째는 초기 100개 주기의 배터리 방전 데이터로부터 배터리의 수명을 예측하는 회귀 모델 실험이고, 두 번째는 초기 5개 주기의 데이터만을 이용하여 수명이 550 주기보다 짧은 지 긴 지 여부만을 판별하는 분류 모델 실험이다.

회귀 모델 실험을 위해 총 세 종류의 선형 모델이 고안되었다. 첫 번째 모델은 ‘variance’ 모델이라 명명하였으며 오로지  $\Delta Q_{100-10}(V)$ 의 분산의 로그 값만을 이용하여 수명을 예측한 모델이다. 앞서 해당 특징이 수명과 높은 상관관계를 띄는 것을 확인했는데, 실제로 이 모델만으로도 상당히 높은 정확도를 얻을 수 있음을 확인하였다. 두 번째 선형 모델은 ‘discharge’ 모델이라 명명하였고 방전 용량 그래프에 대한 통계량까지 같이 사용한 것이다. Elastic net에 의해 총 13종류의 특징 중 6종류가 선택되었다. 마지막 ‘full’ 모델의 경우 모든 특징들을 사용하였으며, elastic net에 의해 총 20종류의 특징 중 9종류가 선택되었다.

분류 모델 실험은 ‘variance’와 ‘full’의 두 모델이 학습되었다. ‘variance’ 모델의 경우  $\Delta Q_{5-4}(V)$ 만이 사용되었으며, ‘full’ 모델의 경우 총 18종류의 특징들 중 4종류가 선택되었다.

검증 결과 표시를 위한 성능 지표로는 회귀 모델의 경우 root mean squared error (RMSE)와 mean absolute percent error (MAPE)가 사용되었으며, 분류 모델의 경우 정확도가 지표로 사용되었다.

#### ○ 실험 결과

표 4에는 회귀 모델 실험의 결과가 나타나 있다. 1차 검증 데이터 중 하나의 셀은 80%의 SOH까지 너무 빠른 속도로 떨어져 다른 셀에 비해 관측된 패턴이 매우 달랐다. 따라서 괄호 안에는 해당 셀을 제외시켰을 때의 지표를 표기하였다. 이때 놀랍게도 단 하나의 특징만을 사용한 ‘variance’ 모델도 높은 정확도를 보여주었는데, 1차 검증에서 13.2%, 2차 검증에서 11.4%의 검증 오류를 보였다. ‘Variance’ 모델에서 사용한 특징인  $\Delta Q_{100-10}(V)$ 의 로그 분산 정보는 ‘discharge’ 모델과 ‘full’ 모델



에서도 elastic net에 의해 선택됨으로서 그 중요성을 입증하였다. ‘Full’ 모델의 경우 각 검증 데이터에 대해 7.5%, 10.7%의 검증 오류를 보이며 총 9.1%의 낮은 검증 오류를 띄었다.

본 논문에서 사용한 배터리 정보가 초기 100주기만을 사용했다는 점도 주목할 점인데, 이는 기존의 논문들이 최소 25% 정도의 용량 저하 데이터를 이용하여 예측한 것에 비해 매우 적은 주기만을 사용한 것이다. 그럼에도 높은 정확도를 얻어낸 것은 각 주기의 전압-용량 곡선을 활용했기 때문이다. 실제로 초기 300주기의 용량 저하 곡선에 대한 특징들만 사용하여 학습한 경우 1차, 2차 검증 데이터에 대해 검증 오류는 각 27%, 46%로 매우 저조한 정확도만을 얻을 수 있었다.

배터리 열화 모드 중 하나인 LAM<sub>deNE</sub>의 경우 용량의 손실 없이 방전 전압이 이동하는데, 이는 리튬 인산 철/흑연 배터리와 같이 음극 용량이 양극 용량보다 클 때 일어난다. 이 경우 용량은 변화하지 않더라도 이 영향이 전압-용량 곡선에 나타나는데, 이로부터 용량 저하 곡선 뿐 아니라 전압-용량 곡선의 특징들도 같이 사용한 본 논문의 모델이 성공한 이유를 설명할 수 있다.

성능지표	RMSE (주기)			MAPE (%)		
	학습 데이터	1차 검증 데이터	2차 검증 데이터	학습 데이터	1차 검증 데이터	2차 검증 데이터
‘Variance’ 모델	103	138 (138)	196	14.1	14.7 (13.2)	11.4
‘Discharge’ 모델	76	91 (86)	173	9.8	13.0 (10.1)	8.6
‘Full’ 모델	51	118 (100)	214	5.6	14.1 (7.5)	10.7

표 4 회귀 모델 실험 결과 성능 지표. 괄호 안 수치는 SOH가 급격히 떨어진 데이터 1개를 제외한 것

표 5에서는 분류 모델 실험의 결과를 확인할 수 있다. 이 경우 ‘variance’ 모델은 88.1%, ‘full’ 모델은 95.1%의 높은 정확도를 얻을 수 있었다. 본 실험의 경우 5주기 이하의 매우 적은 주기만을 이용하였음에도  $\Delta Q(V)$ 에 배터리 예후에 대한 정보가 들어있음을 시사하며, 본 논문에서 제안한 데이터 기반 모델이 다양하게 활용될 수 있을 것으로 기대된다.

성능지표	분류 정확도 (%)		
	학습 데이터	1차 검증 데이터	2차 검증 데이터
‘Variance’ 모델	82.1	78.6	97.5
‘Full’ 모델	97.4	92.7	97.5

표 5 분류 모델 실험 결과 성능 지표

$\Delta Q_{i-j}(V)$ 만 사용한 ‘variance’ 모델에서  $i$ 와  $j$ 를 변화시킬 때 RMSE 값이 어떻게 변화하는지도 검증하였다. 이때  $i$ 가 60보다 큰 경우 결과의 변화는 둔감한 편이었

으며, 이는 본 논문에서 사용한 주기 (100-10)보다 더 이른 주기를 사용한 모델로도 충분히 예측이 가능함을 시사한다. 주기를 고정시키지 않고 SOH를 기준으로 상대적으로 정하는 방법도 고려하였지만 결과가 좋아지지 않는하였으며, 초기 100주기 동안 SOH가 오히려 증가했다는 점에서 고정 주기 방법보다 훨씬 많은 주기를 사용해야 한다는 단점을 가지고 있다.

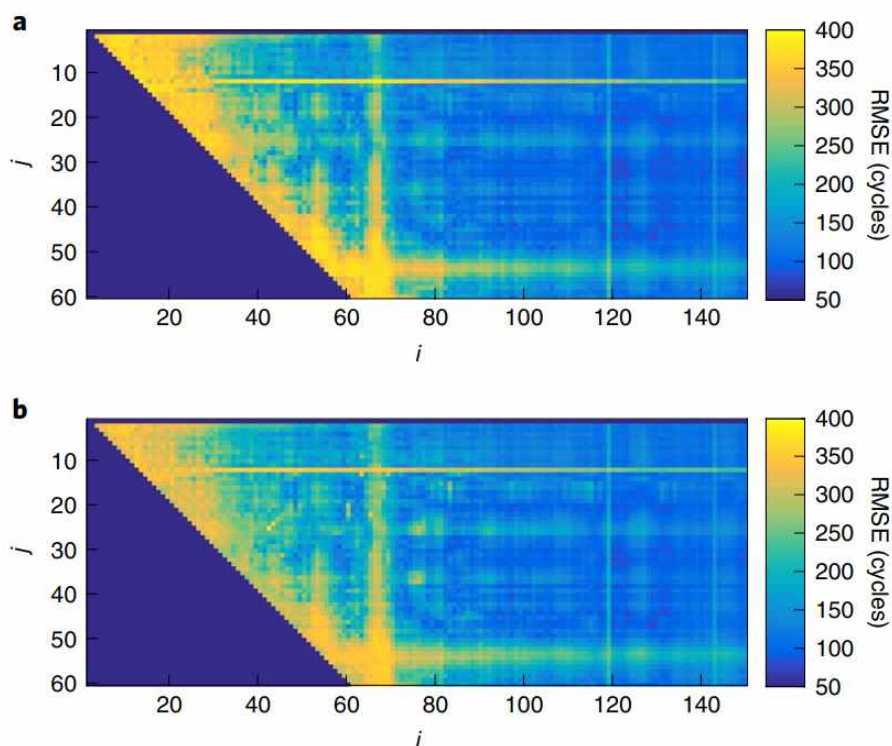


그림 6 용량 차이 계산 주기를 변화시킬 때 학습(a), 검증(b) 데이터에서의 RMSE 변화

## 4. 결론

데이터 기반 모델은 리튬-이온 배터리의 진단 및 예후를 위해 유망한 방법이며, 정확한 배터리 수명의 예측 모델 개발은 배터리의 제조 및 최적화에 있어 중요한 수단이 된다. 본 논문에서는 배터리에 대한 화학적 지식이나 배터리 성능 저하의 자세한 원리에 대한 지식 없이 높은 정확도를 가지는 배터리 수명 예측 모델을 개발하였고, 이는 회귀 문제에서는 처음 100주기의 데이터만을 사용하여 9.1%의 검증 오류를, 분류 문제에서는 처음 5주기의 데이터만을 사용하여 4.9%의 검증 오류를 얻었다. 이러한 정확도는 배터리의 용량 감소 곡선 뿐 아니라 방전 시 전압-용량 곡선에서 추출한 특징들을 사용하였기에 얻어진 것이다.

이 모델의 성공은 초기 사이클에서 용량 감소가 일어나지 않은 셀들에 대해서도 전압

곡선에서는 그 영향이 일관되게 들어난다는 점에서 설명된다. 본 논문에서 사용한 데이터 기반 모델은 물리적, 경험적 기반 방법론들과 결합되어 이들을 보완할 수 있으며, 본 논문에서 연구한 데이터 생성 및 데이터 기반 모델의 학습 방법론들은 리튬-이온 배터리와 같은 매우 복잡한 시스템을 이해하고 개발하는 수단이 될 것이다.

## 5. 참고자료

- Kristen A. Severson. et al. Data-driven prediction of battery cycle life before capacity degradation – Nature Energy, 2019
- Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. J. R. Stat. Soc. B 67, 301-320 (2005).

# 리튬이온 배터리 전기차의 건강지표 예측 기법

## 1. 개요

리튬 이온 배터리의 정전용량(capacity)을 정확하게 측정하는 것은 배터리의 상태를 진단하기 위해 매우 필수적이다. 초기 정전용량 대비 현재 정전용량의 비율은 흔히 SOH(State Of Health)라고 불리는데, 처음에는 100%로 시작하는 SOH가 배터리의 충/방전을 지속함에 따라 점점 감소하는 것이 알려져 있다. 많은 기업들은 SOH 80%가 되면 해당 배터리의 수명이 다했다고 보고, 이를 EOL(end of life)라고도 부른다. 80%라는 기준은 제조사나 사용처에 따라 변경될 수 있으나, 일반적으로 그 시점을 넘어서면 배터리의 노화가 더욱 가속된다는 사실이 알려져 있다.

본 논문에서 수행하고자 하는 목표는 크게 두 가지로, 하나는 정전용량 추정이고, 다른 하나는 잔존수명(RUL, Remaining Useful Life) 예측이다. 정전용량의 '측정'이 아닌 '추정'이라고 표현한 이유는 원하는 대로 매 순간 측정할 수 있는 물리량이 아니기 때문이다. 즉, 배터리의 정전용량은 전압과 전류 같은 물리량처럼 아무 때나 잴 수 있는 대상이 아니며 정확히 계측하기 위해서는 완전충전과 완전방전을 반복해야만 하고, 이를 매번 수행하는 것은 비현실적이기 때문에(특히, 운행 중인 전기차 등) 수학적 모델링 기법을 통해 특정 시점의 정전용량을 추정하고자 하는 것이다. 잔존수명 예측 또한 수학적 모델링을 통해 이루어지는데, 본 논문에서는 특별히 moving window 방식과 최소제곱법을 사용하는 linear aging model을 적용하였다.

## 2. 기법

### ○ Capacity Labeling 방법

주기적으로 계측을 수행하고, 이를 Linear interpolation함으로써 Capacity curve를 얻었다.

### ○ HI extraction을 이용한 Capacity Estimation

HI(Health Indicator)는 CC(Constant Current) 충전 시기 중 3.6V ~ 4.2V 구간 동안 흐른 전류의 적분으로 정의된다. 이를 중요한 지표로 삼은 이유는 CC 충전 시기의 충전용량이 배터리의 건강 상태를 가장 잘 대표하는 지표 중 하나일 뿐 아니라, 대부분의 충전 cycle들이 3.6V 이상 4.2V 이하 구간을 포함하고 있기 때문이다. 이 때 HI의 단위는 Ah가 된다.

HI-Capacity 그래프를 HI에 대한 3차식으로 fitting하면 다음 그림 1과 같은 모습이 나오게 된다. 즉, HI를 측정하는 것만으로 해당 시점의 Capacity를 추정하는 것이 가

능해지는 것이다.

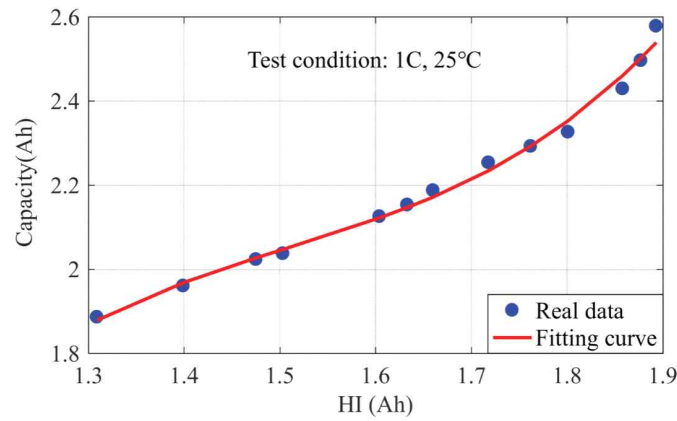


그림 7 HI-Capacity curve

#### ○ RUL prediction

배터리의 수명은 SOH가 80%가 되는 시점이다(EOL).

Baumhofer et al. [26]에 따르면 SOH 85% 이하에서는 SOH가 선형 감소, 그 이전 시기에는 더 빠른 속도로 감소한다. 이 기준으로 배터리의 생애를 Stage 1 (빠른 노화) / Stage 2 (선형적 노화)로 구분할 수 있다.

Gao et al. [29]에 의하면 배터리의 생애를 Stage 1, 2, 3으로 구분하였으며, 단계가 나뉘는 시점은 각각 SOH 95%, 87%이다. Stage가 바뀔 때마다 SOH의 감소 속도가 점차 완화된다고 설명하고 있다.

본 논문에서는 3개의 Stage로 구분을 하였고, 단계가 나뉘는 시점을 따로 고정하지는 않았다. (25도/40도 환경에서 stage 구분 기준이 다르게 나옴. 그림 2 참조) 또한 잔존수명 예측을 실시간 환경에서 잘 수행할 수 있도록 moving-window 방식으로 linear regression 기법을 적용하였다. (그림 3 참조)

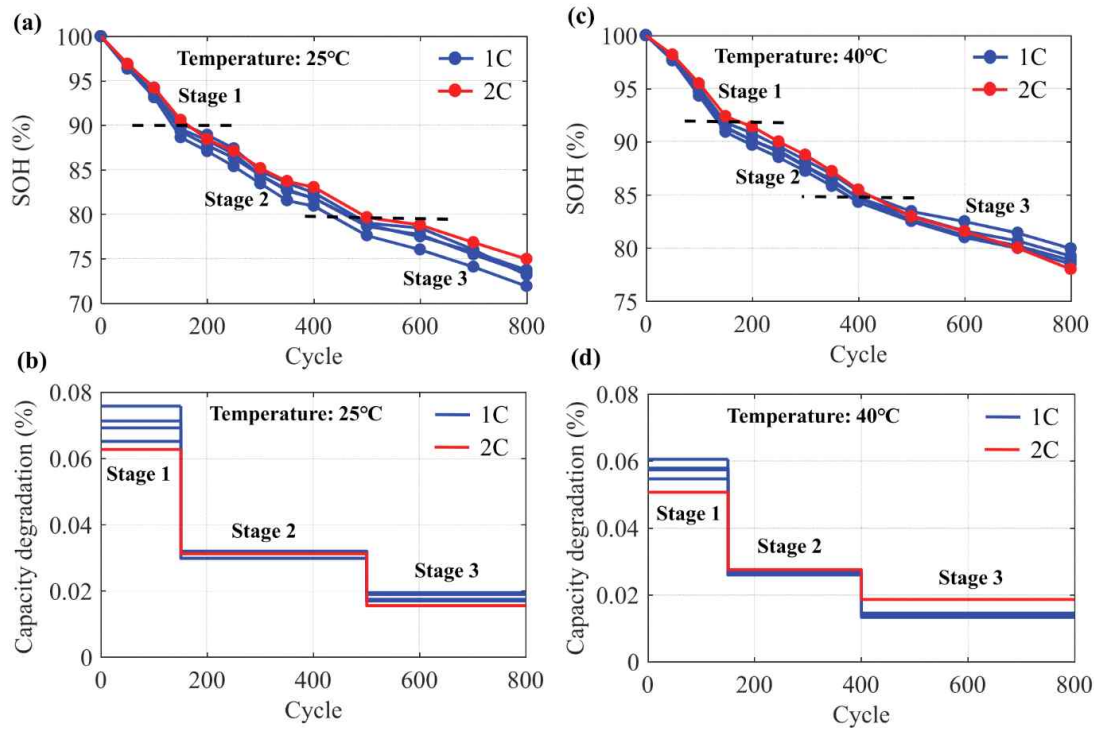


그림 8 온도에 따른 배터리의 수명감소 양상

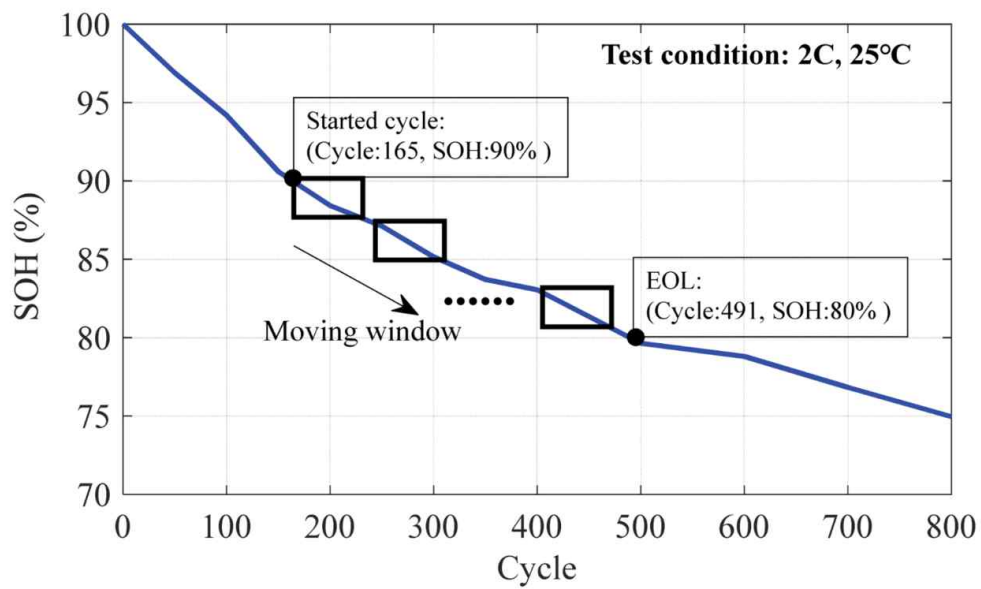


그림 9 Moving-window-based RUL prediction 기법

### 3. 실험

○ 데이터

Panasonic사 제조 NCR18650RF Battery를 이용하여 직접 실험 데이터 추출

○ 실험 방법

방전 속도(C-rate) 1C, 2C 및 온도 25℃, 40℃를 조합, 총 4가지의 부하 조건으로 실험을 진행하였다. 기타 실험 조건은 아래 그림 4를 참조.

TEST CONDITIONS FOR THE NCR18650PF BATTERY

Test Category	Test Condition 1	Test Condition 2
Cycle test	Charge: CCCV, charged at 0.5C rate up to 4.2 V	Charge: CCCV, charged at 0.5C rate up to 4.2 V
	0.05C rate-cutoff	0.05C rate-cutoff
	Charge rest: 30 min	Charge rest: 30 min
	Discharge: 1C rate and 2C rate, 2.5 V cutoff	Discharge: 1C rate and 2C rate, 2.5 V cutoff
	Discharge rest: 30 min	Discharge rest: 30 min
	Temperature: 25 °C	Temperature: 40 °C
	Capacity calibration-OCV test-DST test	Capacity calibration-OCV test-DST test
Characterization tests		

그림 10 논문에서의 실험 조건

○ 실험 결과 및 해석

[1] 부분 충전 그래프로부터 HI 추출 성공 및 추출된 HI와 Capacity 간의 관계를 3차식 fitting으로 묘사하는데 성공하였다. 이 때 RMSE는 1% 전후로 나왔다. 계측이 아닌 부하 cycle에서 Capacity 측정은 어려워도 HI 측정은 쉽게 가능하므로, 이를 통해 cycle에 따른 Capacity estimation curve를 만들 수 있다(그림 5 참조). 이로써 부분 충전의 정보만으로 완전 충전의 정보를 추정할 수 있음을 확인할 수 있다.

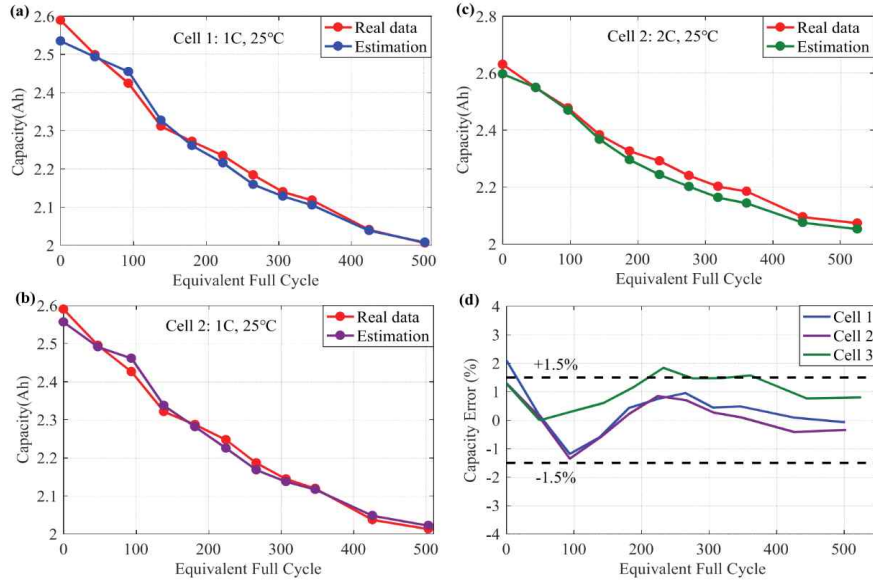


그림 11 Capacity estimation results

[2] RUL prediction의 경우 EOL 직전 마지막 20% 구간 동안 예측에 있어서 RMSE 값이 15~20 cycle 전후로 나왔다. RMSE 계산법은 다음과 같다.

$$RMSE = \sqrt{\sum_{i=k_s}^L (RUL_{p,i} - RUL_{r,i})^2 / N}$$

( $k_s$  : 시작 cycle,  $L$  : EOL에 해당하는 cycle,  $RUL_{p,i}$  :  $i$ 번째 cycle에서 RUL 예측 값,  $RUL_{r,i}$  :  $i$ 번째 cycle에서 RUL 실제 값,  $N$  :  $k_s \sim L$  cycle 동안 예측된 RUL 데이터의 개수)

예시로 25°C 조건에서의 예측 결과를 그림 6에서, 40°C 조건에서의 예측 결과를 그림 7에서 확인할 수 있다(RUL의 단위 : cycle). 여기서 붉은 실선은 실제 RUL 값이고, 붉은 점선은 위아래로 50 cycle씩 오차를 허용하는 구간을 표시한 것이며, 파란 점들은 moving-window size에 따른 각각의 예측 결과들을 찍은 것이다. 일반적으로 moving window size가 클수록 새로운 변화에 대해 보수적으로 받아들이고, 작을수록 새로운 변화에 대해 민감하게 받아들인다.

한편 논문에서는 잔존수명 예측이 배터리의 생애 전체에서 잘 이루어지지 않는다고 밝힘과 동시에 EOL 직전 마지막 20% 구간에서의 예측 정확도를 성과 지표로 제시했다. 생애 초반부터 잔존수명을 예측하는 것이 실제 시스템에서 크게 필요하지 않다는 점을 상기하면 충분히 납득 가능한 부분이다.



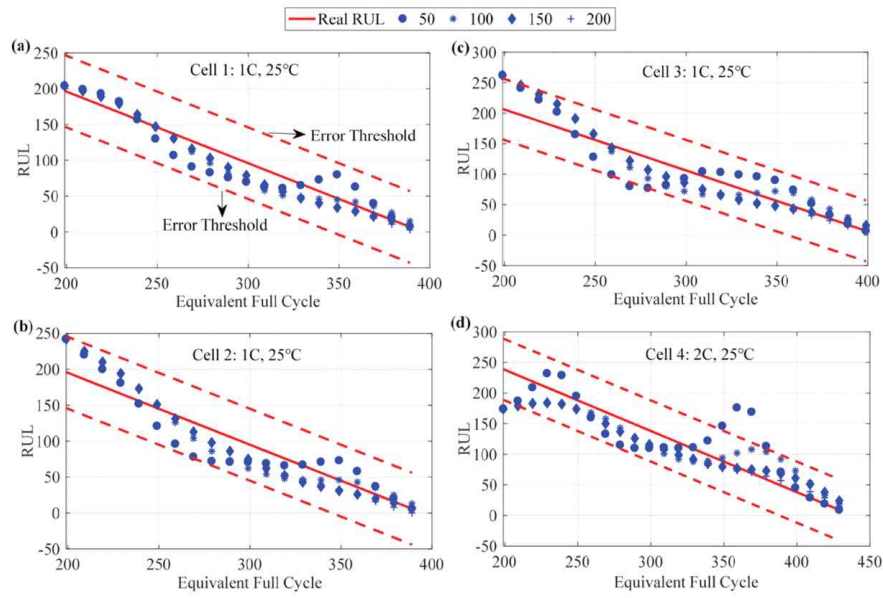


그림 12 RUL 예측 결과 (25°C)

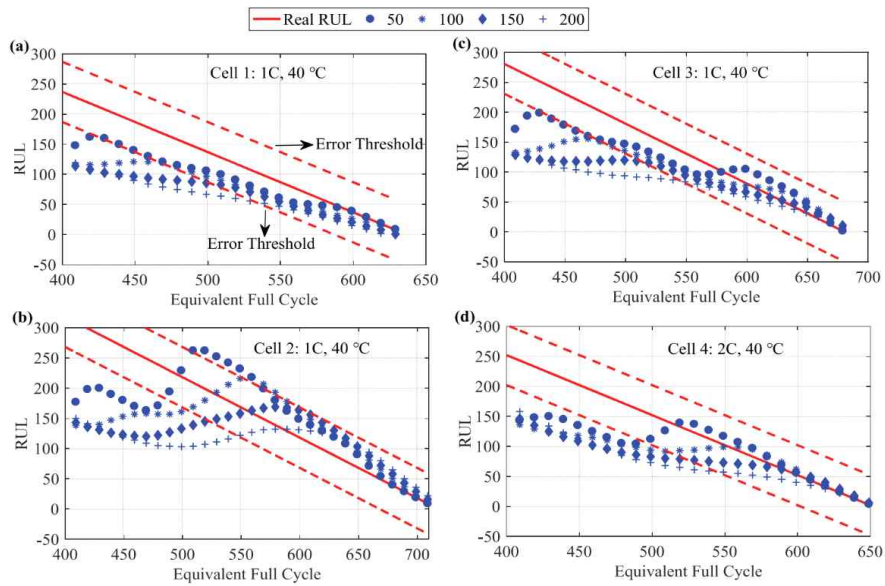


그림 13 RUL 예측 결과 (40°C)

## 4. 결론 및 고찰

[결론]

3.6V ~ 4.2V 구간에 해당하는 부분 충전 그래프만으로 HI 계산을 시행하고, 이를 기반으로 3차식 fitting을 통해 원하는 시점의 Capacity를 추정할 수 있다. 또한 moving-window based linear aging model을 이용하여 잔존수명(RUL)을 예측할 수 있다.

[개선점]

- (1) 전기차 사용의 경우, 사용패턴에 따라 3.6V ~ 4.2V 부분 충전조차 정보를 완전히 얻지 못 할 수 있다. 따라서 더 좁은 구간에 대해서도 적용 가능한 기법을 탐구할 필요가 있다.
- (2) HI와 Capacity 간의 관계가 3차식 fitting으로 충분히 설명 가능한지 검토해야 한다. 추가로 가능한 방법으로는 더 높은 차수의 polynomial이나 다른 함수를 이용한 fitting 혹은 머신러닝 기법 등이 가능하다.
- (3) 본 논문은 실험실에서 미리 설정된 환경 하에 실험이 진행되었으므로 임의 사용패턴의 데이터에 대해 robust하지 않을 확률이 높다. 우선 그 부분을 검토한 후, 변화무쌍한 배터리 사용 패턴(ex: 전기차 실사용 충/방전 데이터)에 대해서도 잘 적용될 수 있도록 접근 방법을 개선하는 것이 새로운 도전과제의 핵심이다.

## 5. 참고자료

- Xiong, Rui, et al. "Lithium-ion battery health prognosis based on a real battery management system used in electric vehicles." IEEE Transactions on Vehicular Technology 68.5 (2018): 4110-4121.

# 전압의 증분값에 기반한 리튬이온 배터리의 수명 예측 기법

## 1. 개요

전기차가 상용화되는 데에 있어 가장 중요한 요소 중 하나는 전기차에 대한 안정성일 것이다. 즉, 사용되는 전기차 배터리의 상태를 판단하는 것이 중요하다. SOC를 기반으로 계산된 주행 가능 거리와 배터리의 안전 상태는 운전자가 신경 쓸 만한 가장 큰 두 요소인데, 이는 모두 배터리의 SOH(state of health)를 통해 계산할 수 있다. 따라서 SOH의 정확한 계산이 매우 유의미한 것이다.

그런데 현재까지의 SOH 계산 방식에는 잘 다뤄지지 않은 문제 요소가 많은데, 요약하자면 다음과 같다.

- 1) 현재 논문상에 구현된 대부분의 SOH의 계산은 완충/완방의 상황을 기본으로 하여 구성된 알고리즘이다. 그런데, 사용자의 배터리는 사용자의 차량 사용 습관에 의해 불규칙하므로 완충/완방이 잘 이루어지지 않는다.
- 2) 배터리가 중간중간 충전이나 방전이 이루어지므로 battery cycle number를 계산하기 쉽지 않다.
- 3) 학습 데이터의 양이 많아야 하고, 배터리 노화 시 모형의 파라미터가 부정확해짐
- 4) 보통 Offline estimation에 관해서만 연구가 이루어졌다.

본 논문에서는 이에 초점을 맞추어, 다음과 같은 특징점을 보이고 있다.

- 1) 10 - 20분 정도의 부분적인 충전 데이터로도 SOH를 계산할 수 있다.
- 2) SOH를 계산하기 위해 미분가능한 voltage based feature vector를 새로이 구현하였다.
- 3) 400 cycle(최장 45일 정도)에 해당하는 충방전 cycle 데이터만으로 계산 모형을 학습할 수 있다.
- 4) 훈련용 배터리와 평가용 배터리가 동일한 스펙이 아니어도 좋다. 본 논문의 실험에서는, 3.0 Ah 배터리를 학습에 사용하였고, 3.5 Ah 배터리를 평가하는 데에 사용하였다. (전기화학적으로 비슷한 성능이라는 가정)

## 2. 데이터 구성

논문에서 활용할 데이터를 구성하는 과정을 다룬다.

Lithium Cobalt Oxide (LCO) cathode와 graphite anode 2종의 배터리가 데이터 구축하는 데에 활용되었으며, 모두 3.85 - 4.4V nominal voltage의 pouch cell geometry이다. 3.0 Ah의 1종 배터리 10개와 3.5 Ah의 2종 배터리 8개가 사용되었다. 이 배터리들은 각각 훈련 데이터와 평가 데이터를 구성하게 된다. 배터리들은 45°C와 25°C의 두 가지 온도 환경에서 충방전 cycle이 진행되었으며, 온도는 thermal chamber 안에서 일정하게 유지되도록 하였다. 충방전은 표에 제시된 충방전 프로토콜을 따랐다.

0.2C rate 방전 데이터는 target SOH를 예측하는 데에 활용되었다. 0.8C rate 데이터는 1분, 1.2C rate 데이터는 10초 단위로 추출되었고, cubic polynomial을 통해 모든 데이터는 1초마다 샘플링된 것으로 단위 시간 이내의 데이터를 계산하였다.

단계	프로토콜 설명
1	일정한 전류로 충전 (0.8C / 1.0C / 1.2C rate)
2	일정한 전압으로 충전 (4.4V)
3	일정한 전류로 방전 (1.0C / 1.2C)
4	50 / 125 cycle마다 CC-CV 충전과 0.2C rate의 CC 방전 cycle을 조사

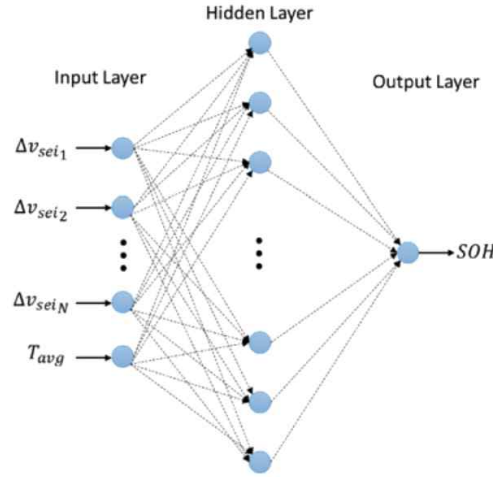
### 3. 모형의 구성

우선 배터리의 내부저항, SOC,  $V_{sei}$ 를 이전 cycle들의 정보를 통해 계산한다. 계산은 각각 다음과 같이 구성된다.

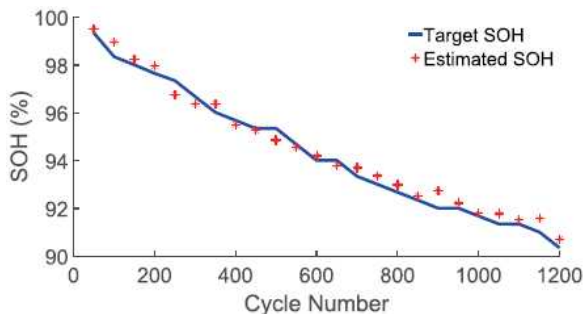
내부 저항	$R_i = R_f + \Delta R_{sei}$ <p>(<math>R_f</math> : fixed internal resistance of the battery which does not change with aging)</p>
SOC	$SOC(k) = SOC(k-1) + \frac{i(k) T_s}{C_{max}}$ <p>(<math>T_s</math> : sampling time, <math>C_{max}</math> : rated capacity of the battery)</p>
$V_{sei}$	$V_{sei}(k) = V(k) - R_f i(k)$

계산 후, 각 sampling time마다의  $V_{sei}$  값의 차와 충전 시 배터리의 온도 값으로 feature vector  $x = [\Delta V_{sei_1}, \Delta V_{sei_2}, \dots, \Delta V_{sei_9}, T_{avg}]$ 를 구성한다. 이 feature vector를 input으로 하는 ANN(인공신경망)을 구성하여, feature vector와 SOH의

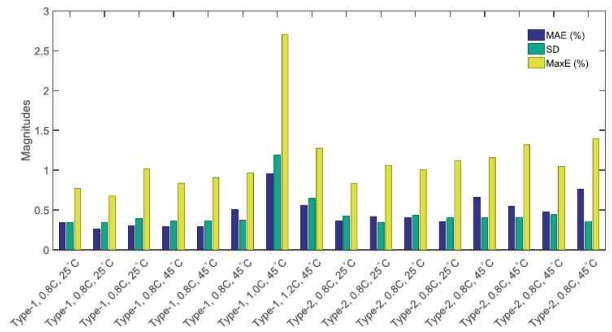
연관성을 찾아 SOH의 값을 예측하게 한다.



[그림]  $V_{sei}$ 를 이용해 만든 feature vector를 통해 SOH를 계산하는 ANN



[그림] SOH 예측 결과



[그림] 여러 환경에서의 실험 결과 비교

#### 4. 결론

부분적인 battery cycle 데이터를 사용하기 위해 online solution을 개발하였고, 약 15분 가량의 배터리 사용 및 충전 데이터만 가지고도 SOH를 예측해낼 수 있게 되었다. 논문에서 제시한 모형은 현재 사용되고 있는 BMS에서 계산할 수 있는 전류, 전압, 온도를 사용하고 있으므로, 이를 구하는 것은 어렵지 않은 일이다. 모형을 훈련시키기 위해서는 충방전 cycle이 400회 가량이 필요한데, the SOH 예측 알고리즘이 약 45일 이내에 사용 가능해짐을 알 수 있다. 모델에서 제시한 모형을 다른 조건에서 실험해보았을 때, 대부분의 자료에서  $MAE < 1\%$ 로 관찰되었고, 환경에 robust한 것을 확인할 수 있었다. 또한, 첫 cycle에서 계산한 내부저항 값만으로도 잔여 수명동안의 SOH를 계산할 수 있다는 점이 특징이다.

논문에서 제시한 기법은 데이터가 더 많아질 경우 딥 러닝 기법 등을 적용하여 더 정교한 모형을 구현할 수 있을 것이다. 또한, 전기차 배터리 뿐 아니라 배터리를 사용하는 다른 전자기기에도 동일한 방식으로 적용할 수 있을 것이라 기대된다.

## 5. 참고자료

- Arunava Naha. et al. An incremental voltage difference based technique for online state of health estimation of Lo-ion batteries – Scientific reports, 2020

# 리튬이온 배터리의 내부 단락 검출 기법

## 1. 개요

Li-ion batteries(LIB)의 사용이 늘어나면서, LIB의 안전성에 대한 관심이 높아졌다. 그중에서도 Internal short circuit(ISC)은 내부 발열을 일으키고, 폭발과 화재 등의 사고원인이 된다. 내부 온도는 문제가 생기기 직전이나 급격하게 증가하기 때문에, 더 이른 ISC탐지가 필요하다. 여기서는 배터리의 통상 충방전 전류와 전압 데이터를 기반으로 스마트폰 배터리의 ISC여부를 진단할 수 있는 random forest classifier를 소개한다. 이를 통해 배터리의 특정 부분의 온도 데이터가 필요하던가, 특별한 충방전을 해야 하는 기존 기법의 한계를 극복했다.

## 2. 기법

- ISC detection을 위해 random forest classifier를 사용했다. Random forest classifier는 앙상블 학습 방법의 일종으로, 훈련 과정에서 구성한 다수의 결정 트리로부터 투표를 실시하여 가장 많이 득표한 결과를 최종 분류 결과로 선택하는 machine learning 기법이다. Tree 개수( $N_t$ )와 max\_features는 grid search를 통해 구했다.

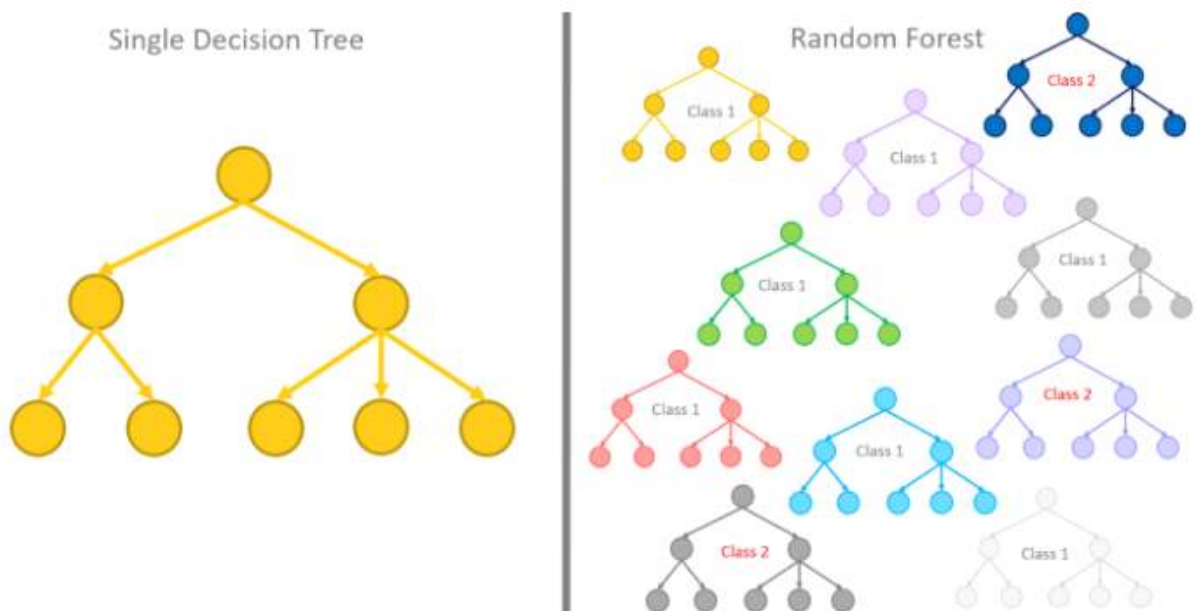


그림 17 결정트리의 구조(왼쪽)와 다수의 결정트리들로 구성된 random forest 구조(오른쪽)

다수의 결정 트리를 사용함으로써 single classifier보다 강건하고 정확도도 높다.

몇몇 다른 기법들에 비해 필요 계산량이 적다. 실제 사용에 깊이 D인  $N_t$ 개의 trees를 사용하므로  $O(N_t * D)$ 의 계산량이 필요하다.

배터리의 실 사용 데이터(전류와 전압)에서 추출한 8개의 features를 학습에 사용하고, ISC 여부를 타겟으로 했다. 8개의 features는 다음과 같다.

1. End of charge SOC( $SOC_{max}$ ) : 충전 마지막의 SOC(state of charge,  $SOC(k+1) = SOC(k) + \frac{I_t(k)T_s}{C_{max}}$ ). ISC인 경우 전류에 누수가 생겨서 평소보다 많음.
2. OCV at low SOC level( $V_{cmin}$ ) : 방전중에 SOC가 p%(방전 SOC level)일 때의 OCV(open circuit voltage). ISC인 경우 평소보다 더 낮음.
3. CV time( $T_{cv}$ ): CV(constant voltage) 충전이 완료될 때 까지의 시간. CC(constant current)에 걸리는 시간보다 뚜렷하게 차이남. ISC인 경우 평소보다 오래 걸림.
4. Energy loss( $E_L$ ) : 충전과 방전간의 에너지 차이(충전과 방전 Q-V곡선의 내부 면적). ISC인 경우 평소보다 큼.
5. Slope of OCV near the end of discharge( $Slope_{ocv}$ ) : 방전 마지막의 OCV와 SOC가 5%차이나는 시점의 OCV간의 기울기. ISC인 경우 평소보다 가파름.
6. SOC vs.  $V_{ocv}$  slope( $a_1$ ) : OCV를 SOC에 대한 5차 다항식으로 표현했을 때, 다항식의 1차 계수( $V_{ocv}(k) = a_0 + a_1SOC(k) + a_2SOC^2(k) + a_3SOC^3(k) + a_4SOC^4(k) + a_5SOC^5(k)$ )
7.  $\Delta R$  : (충전 마지막의 저항) - (방전 시작의 저항)
8. meanR : 방전 중 SOC가 0.3~0.8 구간의 평균 저항

### 3. 실험



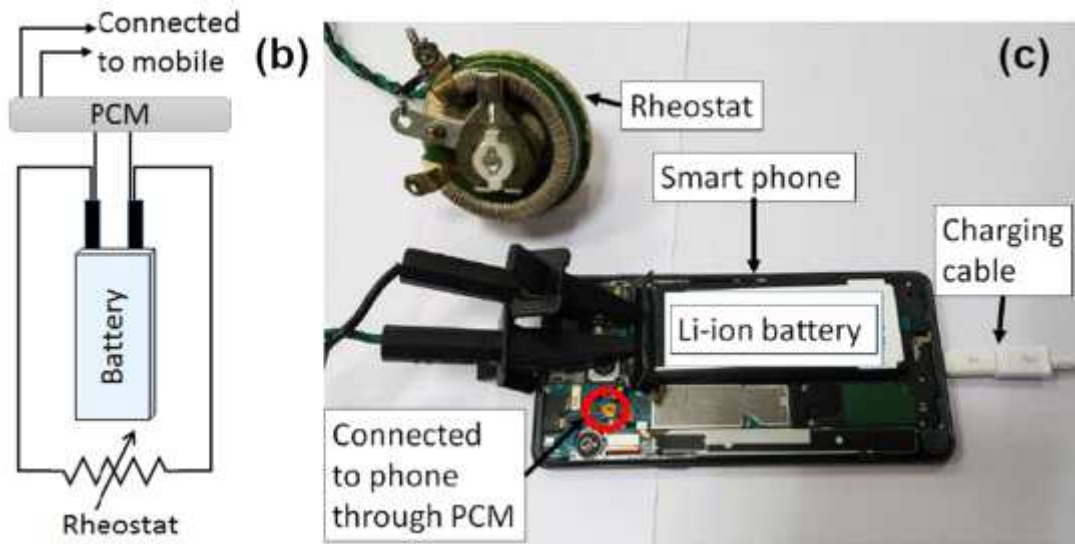


그림 18 ISC 조건의 배터리를 모방하기 위해 사용한 실험 세팅

#### ○ 데이터

정상 배터리의 단자를 외부에서 작은 저항으로 연결해서 ISC 상태의 데이터를 모방했다. 이렇게 얻은 ISC 배터리 데이터와 정상 배터리 데이터를 training dataset 으로 사용했다. 사용자들이 매번 배터리를 매우 낮은 SOC까지 사용하지 않기 때문에, 방전 마지막 SOC level을 다르게 한 데이터들을 사용했다. ISC 290 cycles, 정상 53 cycles이다.

Testing dataset은 정상과 abused battery data로 구성됐다. 5개의 다른 배터리에서 얻은 abused 129 cycles, 정상 148 cycles이다. 5개의 정상 배터리에서 정상 배터리 데이터를 얻은 뒤, 4ft높이에서 단단한 지표에 낙하시켜서 abused 배터리 데이터를 얻었다. (ISC가 없을 수도 있으니, 낙하시킨 후 완전 충전해서 3일 기다린다. 3일 뒤에 OCV가 정상에 비해 0.4V이상 떨어질 경우 ISC로 간주한다. 그렇지 않을 경우, 조건을 만족할 때까지 반복한다.)

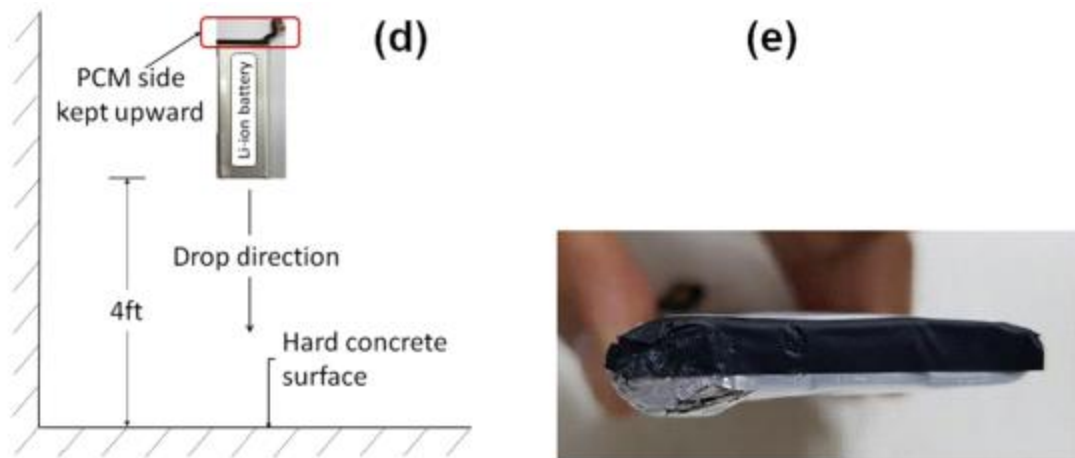


그림 19 검증에 사용하기 위한 ISC 데이터를 얻기 위한 낙하 실험(d)과 그 결과(e)

#### ○ 실험 방법

충방전의 전류와 전압에서 파생된 8개의 features를 사용하는 random forest classifier를 training dataset으로 학습하고, testing dataset으로 검증했다.

#### ○ 실험 결과

RF classifier를 통해 각 feature의 중요도를 계산한 결과는 다음과 같다.

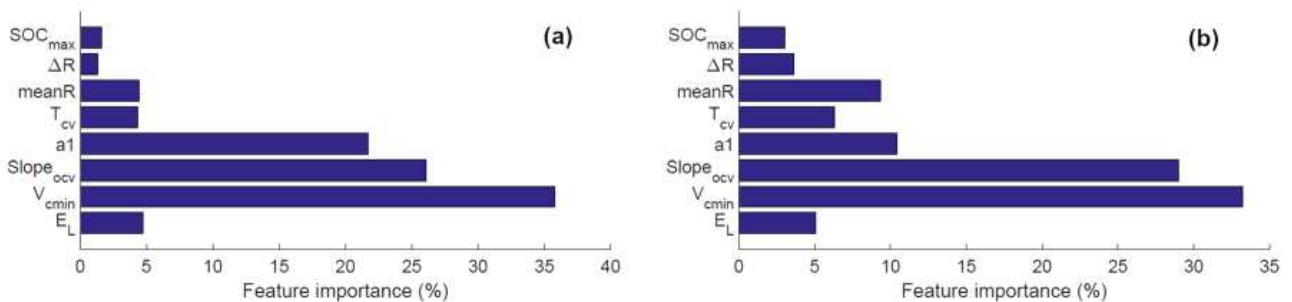


그림 20 Feature 중요도 그래프, 방전 SOC level 5%(a), 방전 SOC level 15%(b)

Feature의 중요도는 방전 마지막의 SOC값에 따라 달랐다. 정상과 ISC 배터리를 구별하는데에  $V_{cmin}$ ,  $Slope_{ocv}$ ,  $a_1$ , meanR,  $E_L$ 가 가장 중요한 features로 보인다. Features의 수를 제한할 경우 정확도가 떨어졌으므로, 나머지 features도 어느정도 분류에 기여하고 있음을 알 수 있다.

방전 SOC level 5%에서 classifier의 training performance는 Normally

Normal(NiN) = 100%, Faulty in Fault(FiF) = 99.66%를 기록했다. 다른 방전 SOC level에 대해서도 training performance는 크게 다르지 않았다.

테스트 결과는 다음과 같다.

			NiN = 100%				NiN = 97.97%
<b>SOC = 5%</b>	Faulty	Healthy	FiF = 98.45%	<b>SOC = 15%</b>	Faulty	Healthy	FiF = 97.67%
Faulty	TN = 127	FP = 2	False Alarm = 0.0%	Faulty	TN = 126	FP = 3	False Alarm = 2.03%
Healthy	FN = 0	TP = 148	Miss-detection = 1.55	Healthy	FN = 3	TP = 145	Miss-detection = 2.33%
			NiN = 97.97%				NiN = 93.92%
<b>SOC = 30%</b>	Faulty	Healthy	FiF = 97.67%	<b>SOC = 50%</b>	Faulty	Healthy	FiF = 97.67%
Faulty	TN = 126	FP = 3	False Alarm = 2.03%	Faulty	TN = 126	FP = 3	False Alarm = 6.08%
Healthy	FN = 3	TP = 145	Miss-detection = 2.33	Healthy	FN = 9	TP = 139	Miss-detection = 2.33%

그림 21 Testing dataset을 사용한 테스트 결과

방전 SOC level이 높을수록 더 낮은 정확도를 보인다.

Classifier	RF	ANN	SVM
NiN	100.0	97.97	94.59
FiF	98.45	100.0	87.60

그림 22 다른 classifiers의 성능

같은 training dataset을 사용한 Artificial Neural Networks, Support Vector Machine의 결과이다. Random forest classifier의 NiN값이 가장 높음을 알 수 있다.

#### 4. 결론

낮은 방전 SOC level에서 매우 높은 정확도를 보였지만, 높은 방전 SOC level에서는 비교적 낮은 정확도를 보인다. 스마트폰과 전기차는 보통 완전 방전하지 않으므로, 높은 방전 SOC level에서의 정확도를 높이는 것이 하나의 과제이다. Random forest 이외의 machine learning 기법을 사용하거나, 다른 features를 추가하는 등의 추가적인 실험이 필요하다. ISC 조건의 배터리 데이터가 주어지지 않았을 경우에는 어떤 학습 방법을 취할 수 있을지도 고민할 필요가 있다.

#### 5. 참고자료

- Naha, Arunava, et al. "Internal short circuit detection in Li-ion batteries using supervised machine learning." Scientific reports 10.1 (2020): 1–10.

---

## 배터리 데이터 전처리 모델

---

## 1. 개요

최근 전기차 등의 수요가 커지면서, 배터리의 수명 예측 및 이상 반응 탐지에 대한 관심이 커지고 있다. 그중에서 SOH(State of Health)는 SOC(State of Charge) 예측과 배터리의 안전 상태를 확인할 수 있는 대표적인 물성 지표이다. 따라서 배터리의 SOH 예측을 위한 학습 데이터를 구축하는 것이 중요한데, 배터리 관련 오픈 데이터셋 중 하나인 NASA 데이터셋을 사용하여 이를 구축하였다.

또한, 인셀에서 제공받는 태양광 ESS 데이터를 이용한 실험 및 예측이 이루어질 수 있게끔 데이터에 알맞은 전처리 기법을 구축하였다.

## 2. 데이터셋 소개

NASA 데이터셋은 실험방식에 따라 크게 7가지 시나리오가 있는데, 충전이 일정한 데이터셋인 RW03, RW04, RW05, RW06을 사용한다. 이 데이터셋은 4.2V까지 충전한 후, 3.2V까지 random walk 과정을 통해 방전을 실시한다. 매 50 cycle의 Random Walk 방전 사이클을 실시한 후, reference cycle을 둬으로써 배터리의 SOH를 계산할 수 있게 구성되어 있다.

NASA 데이터셋은 다양한 실험 환경에 따른 결과를 제공하여 다양한 분석이 가능하나, 측정 간격이 일정하지 않거나 너무 긴 시간동안 기록이 비어있는 데이터도 여럿 존재한다. 이러한 데이터는 바로 분석에 활용하기 어려우므로, 별도 전처리 모델을 활용해 이를 다듬을 필요가 있다.

반면 인셀 데이터셋은 일정한 시간 단위로 데이터를 수집하므로 별도로 결측치가 없이 양호한 상태이다. 따라서 전처리 시 별도의 함수를 구성하지 않고 일반적인 물리량을 계산할 수 있다.

## 3. 전처리 모델 소개

○ NASA\_preprocess\_basic 모델

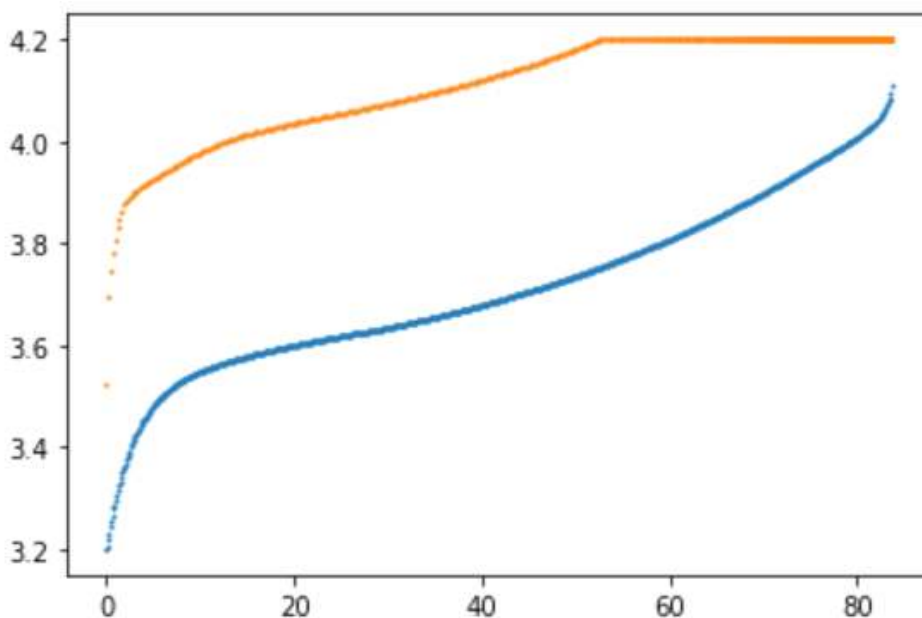
- 1) 기존 데이터는 기록 간격이 일정하지 않아 안정적인 추가 물리량 계산이 어려웠다. 이를 해소하기 위해 1초 간격으로 시간 간격을 통일하였다. scipy library의 interpolate 모델을 활용하여, 배터리의 capacity, V, I의 값을 1초 단위로 정제하였

다.

- 2) 10000초 이상 결측되어 있는 구간이 몇십 개 있었는데, 그 구간 이후 전압이 4.2V가 되어있었으므로, 배터리가 완전히 충전되었다고 가정하고 결측치를 채웠다. 그러나 전류의 값을 interpolate하는 과정에서 음수가 나오는 문제가 있어, 이는 advanced 모델에서 대처한다.
- 3) reference cycle의 방전이 4.2V에서 3.2V로 잘 방전시킴을 확인하였다. 즉, 완전한 방전이 이루어졌다고 볼 수 있으므로, 이때를 기준으로 capacity를 계산한다. 여기서, reference cycle의 방전은 총 2회 발생하는데, 실제로는 이를 평균 낸 값을 이용하고 0.05%만큼의 수치를 조절하여 자연스러운 plot이 되도록 하였다.

#### ○ NASA\_preprocess\_advanced 모델

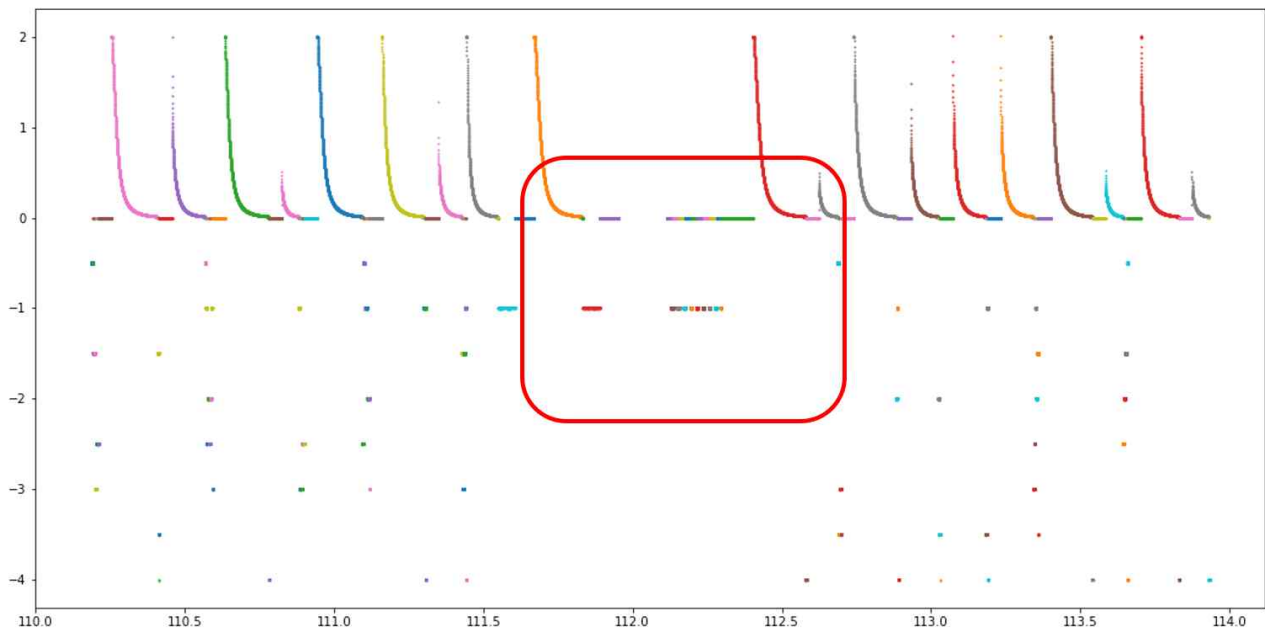
- 1) 전류의 값을 통해 전하량 Q를 계산하였다. 이는 전류를 적분한 값을 이용한 것으로, 수치적으로 불안정할 수는 있으나 전처리 결과 약 1350만 라인 중 대부분이 정상범위에 들어있으므로 큰 문제는 없는 것으로 보이므로 이를 그대로 사용한다. 또한, 일부의 비정상 구간(해당 시점의 capacity보다 높거나, 음수인 경우)은 정상 범위에 들 수 있게끔 최대/최솟값을 이용해 범위 이탈을 제한하였다.



Q-V Plot

- 2) basic 모델에서는 장기간 결측된 구간을 interpolate하는 과정에서 charge 중 음수가 계산되는 경우가 있었다. 이는 위의 전하량 계산을 개선함으로써 문제를 개선할 수 있다.

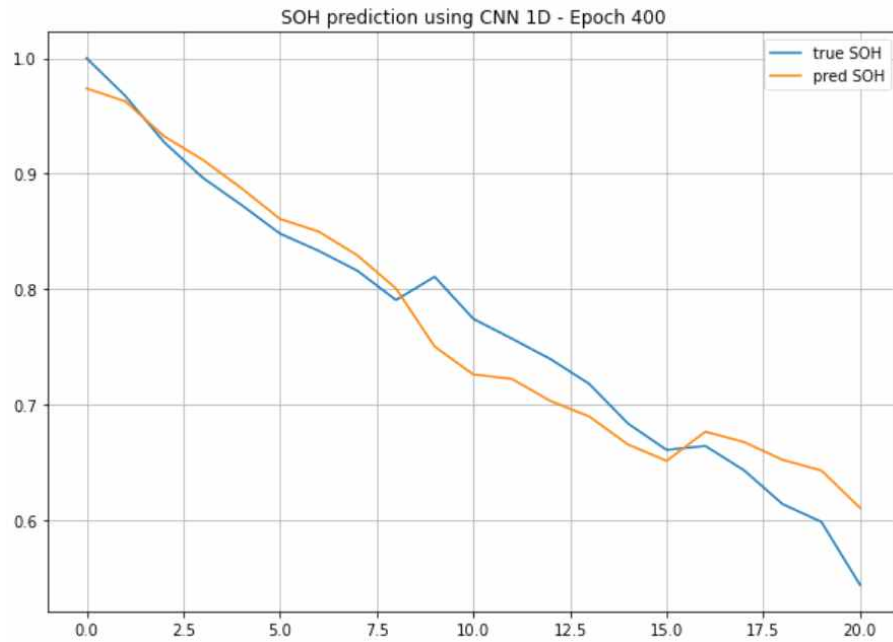
- 3) random walk cycle과 reference cycle의 번호를 정제하였다. random walk cycle의 경우, '충전'과 '방전과 휴식'을 각각 하나의 과정으로 보고 번호를 매겼다. 또한, reference cycle은 2회의 충전과 방전을 하나의 cycle로 보고 reference cycle 과정이 아닌 경우 -1을 부여한다.



Reference cycle 추출

- 4) SOC를 계산한다. 배터리의 현 시점 capacity 대비 현재 배터리의 전하량  $Q$ 의 비율을 계산한다. 즉,  $SOC = \frac{Q}{capacity}$  이다.
- 5) SOH를 계산한다. reference cycle이 매번 있으므로, 배터리의 현 시점 capacity를 매번 계산할 수 있으므로, 이를 통해 SOH를 계산할 수 있다. SOH는 가장 높은 capacity 값 대비 현재 capacity 수치의 비율을 계산한다. 즉,  $SOH = \frac{capacity}{\max(capacity)}$  이다.





SOH 추정

### ○ 인셀 데이터셋 전처리 모델

- 1) 전류, 전압은 PCS / Bank / Rack의 조건에 따라 전압과 전류 값을 뽑아서 사용한다.

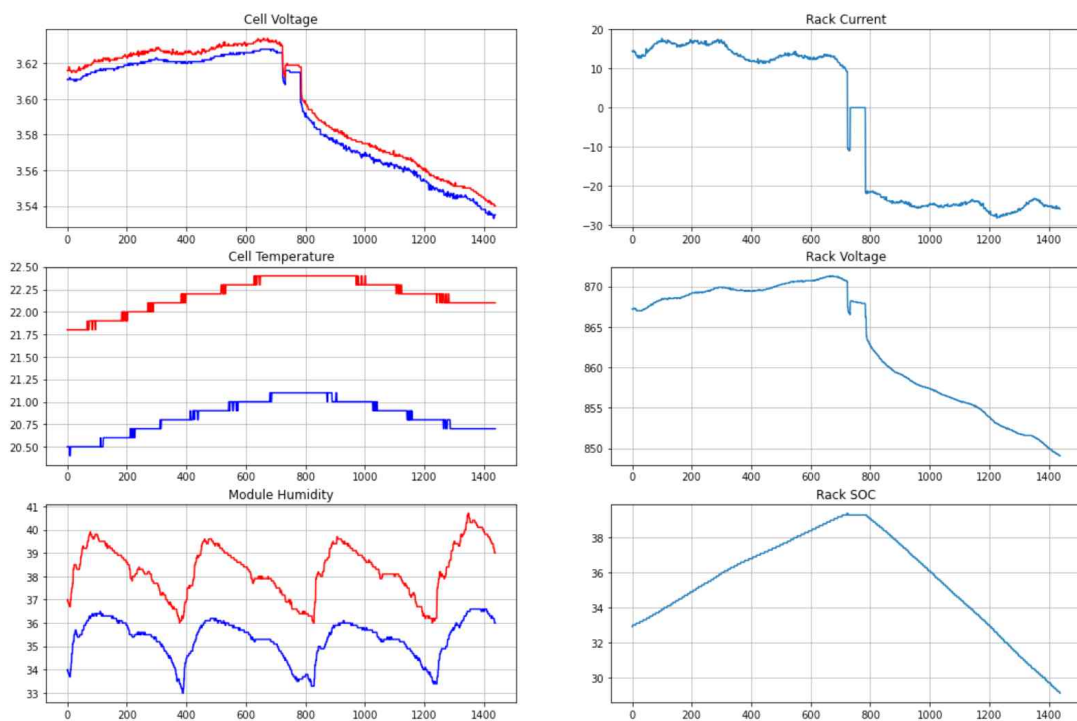
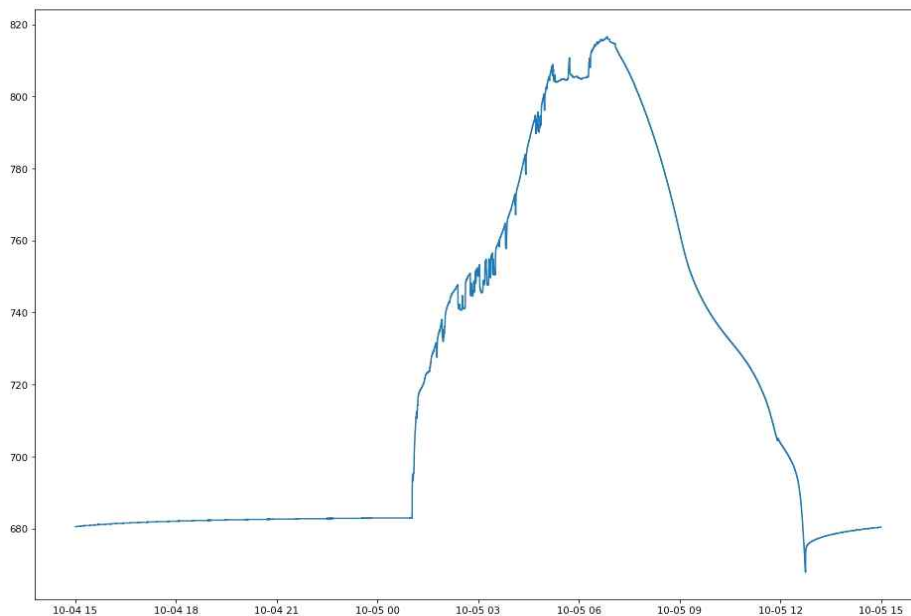


그림 26 인셀 데이터 시각화

2) 전하량  $Q$ 는 NASA 데이터셋과 동일하게 전류를 적분한 값을 사용하며, 전력량 또한 전류와 전압을 통해 직접 계산할 수 있다.



하루(00시~24시) 내 전류 데이터



하루(00시~24시) 내 전압 데이터

#### 4. 결론

배터리의 안전 상태를 확인할 수 있는 가장 대표적인 물성 지표인 SOH를 예측하기 위한 모델을 구축하기 위해 필요한 데이터를 전처리하는 모델을 개발하였다. NASA 데이터셋의 결측치 보강 및 주요 물리량의 보정을 통해 데이터의 안정성을 도모하였다. 이제 전처리된 NASA 데이터를 통해 여러 가지 방법으로 모델을 학습하여 SOH 예측에 활용할 수 있다. 또한 학습한 모델을 충분한 기간 동안 누적된 인셀 데이터셋에 활용하여 SOH 예측이 잘 이루어지는지 비교 분석할 수 있겠다.

---

## 배터리 SOH 추정 모델

---

## 1. 개요

State of health(SOH)는 배터리의 상태를 이상적인 조건과 비교하여 백분율로 나타내는 수치로, 배터리 상태를 진단하는 데 있어서 중요한 지표이다. 따라서 SOH를 정확하게 예측하는 것은 배터리의 안전 상태를 확인하는 데 있어서 매우 중요한 요소이며 본 실험에서는 전기화학적 지식 없이 배터리 운영 상의 데이터를 통해 SOH를 정확히 추정하는 것을 목표로 하고 있다. 본 실험에서는 시계열 데이터 분석에 자주 사용되는 LSTM (Long-Short Term Memory)을 사용하여 SOH 추정을 하고자 한다. cycle당 연속된 시간에서의 전압, 전류, 온도, 시간 차를 입력받아 SOH값을 추정하고 있으며 이 방법은 동일한 cycle 내 충분한 양의 데이터만 있으면 추정이 가능하다는 이점을 가지고 있다.

## 2. 기법

데이터 추출 과정은 sliding window 개념을 적용하여 각 cycle 마다 연속된 시계열 데이터를 추출하는 방식으로 진행하였다.. 동일한 cycle 내의 데이터는 같은 SOH값을 가지기 때문에 cycle 단위로 데이터를 추출하였다. SOH에 영향을 끼치는 것으로 알려진 V(전압), I(전류) 등의 요소가 동일한 SOH에 대해서도 다른 값을 가지고 있다.. 시간을 기준으로 정리된 데이터의 특성을 고려한다면 연속된 시간에서 이들(V,I등의 요건)의 변화가 중요하다고 판단하여 이러한 점을 모델에 참고하고자 같은 SOH값을 가지는 연속된 데이터를 추출하였다. 본 실험에서는 sliding window는 일정 개수의 연속된 시계열 데이터의 블록을 지칭한다. 또한 overfitting을 막기 위하여 cycle 내의 임의의 시간대의 데이터를 랜덤하게 추출하였다..

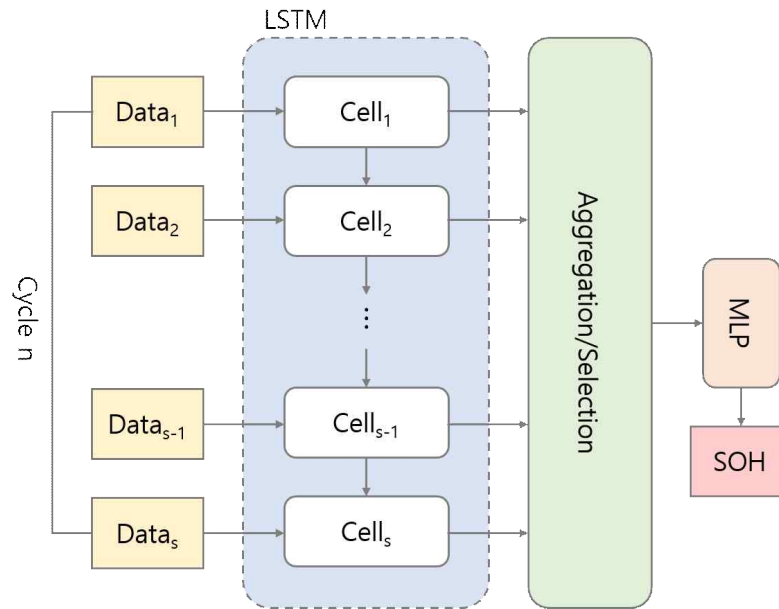


그림 29 : LSTM 모델 구조도

본 연구의 최종적인 목표는 배터리 운영 데이터의 SOH를 추정하는 것이므로 시계열 데이터 처리에 주요하게 사용되는 RNN(recurrent Neural Network) 계열의 모델 사용하였다. 실험에 사용한 모델은 vanilla RNN과 LSTM이며, test 시의 최종적인 추정은 모델을 통해 50번 추측하여 이들의 평균을 최종적인 추정치로 활용하였다.

### 3. 실험

#### ○ 데이터

전처리된 NASA 데이터를 사용하였다. 이 중, 3, 4, 5번 셋을 training data로, 6번 셋을 test data로 사용하였다.

#### ○ 실험 방법

다음과 같은 model hyperparameter와 training hyperparameter를 변경하며 진행하였다.

모델	SRN	LSTM
입력값	V, I	V, I, T, dt
마지막 layer 처리	select last	feature aggregation
sliding window	32	64

#### 1) 모델

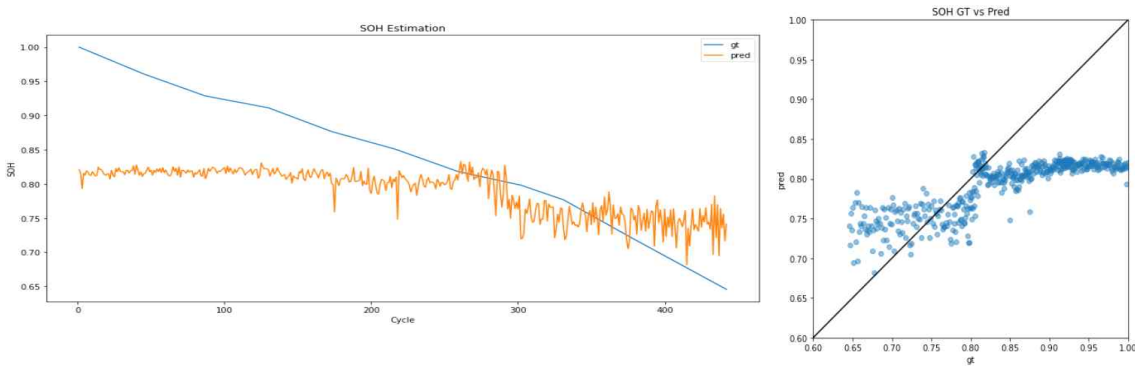


그림 30 SRN 이용한 SOH 추정 (1)cycle에 따른 SOH추정 (2)SOH 추정값과 실제값 비교

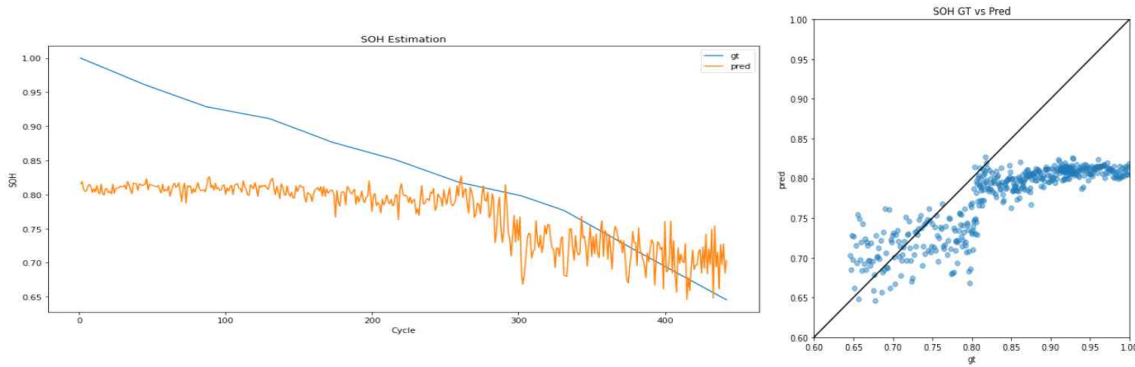


그림 31 LSTM 이용한 SOH 추정 (1)cycle에 따른 SOH추정 (2)SOH 추정값과 실제값 비교  
 SRN(Simple Recurrent Network)는 pytorch에서 제공하는 가장 기본적인 RNN모델이며 LSTM(Long-Short Term Memory) 모델은 기본적인 RNN 모델에서 사용하고자 하는 정보와 활용하고자 하는 정보 간의 간격이 멀더라도 이를 보완하는 모델이다. 실험에서도 SOH를 추정하는 데 있어서 바로 직전에 학습한 cycle의 정보가 아니라 그 이전에 학습한 cycle의 정보가 필요해 RNN에 비해 LSTM의 추정 성능이 좋은 것으로 나타났다고 판단하였다.

## 2) 입력값

‘2.기법’에서 소개한 바와 같이 각 cycle마다 연속된 데이터를 랜덤하게 추출하는 데 있어서 데이터의 V(전압), I(전류), T(온도), dt(해당 데이터와 바로 이전 데이터의 시간 차)를 우선적으로 사용하고, 추가적으로 V, I 만을 사용해 실험을 진행하였다.

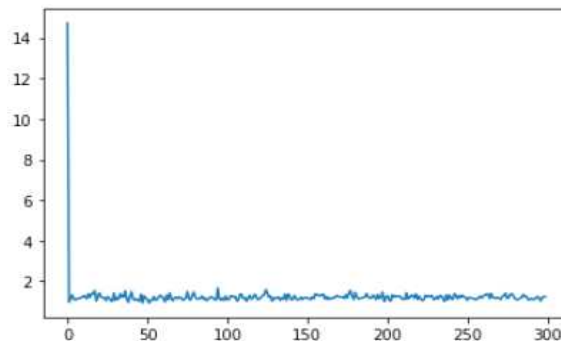


그림 32 V,I 활용 (0.108)

### 3) 마지막 layer 처리

모델의 마지막 layer(그림1. 모델 구조도 ‘Aggregation/Selection’)에서의 추정을 위해 실험에서는 다음과 같은 두 가지 방식의 처리를 적용하였다. 1)LSTM의 마지막 output feature 값을 선택하여 모델의 추정치로 사용하는 방법(select last)과 2)LSTM을 거쳐 나온 feature들은 self-attention mechanism을 이용해 하나의 feature로 모은 aggregation 방식(feature aggregation layer), 두 가지를 거쳐 최종적인 MLP를 통과하여 SOH를 추정하였다. 실험 결과, Aggregation 방식으로 feature를 추출한 경우 위의 selection 방식보다 두 방식의 loss는 비슷한 경향으로 감소하는 것에 비해 training/inference time이 더 길기 때문에 selection 방식을 선택하였다.

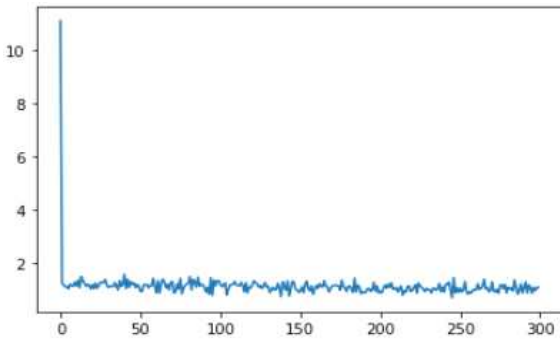


그림 33 select last 활용 (0.103)

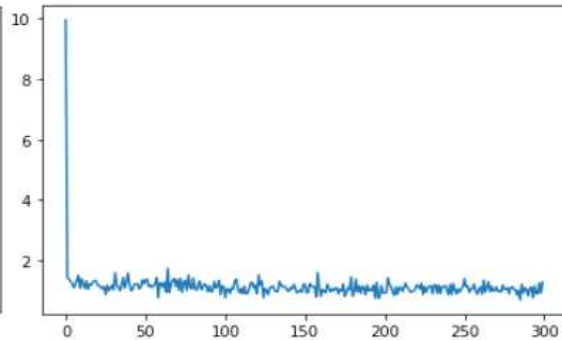


그림 34 aggregation layer 활용 (0.112)

이 외에도 training hyperparameter로 learning rate, sliding window의 크기 등을 고려하였다.

### ○ 실험 결과

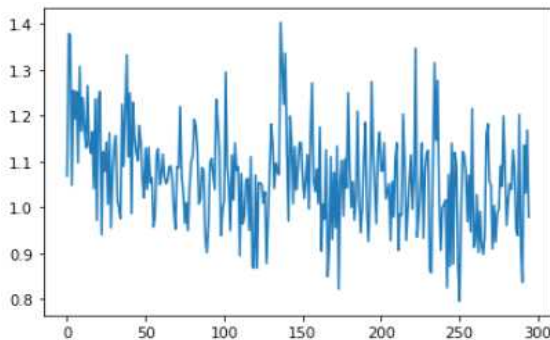


그림 36 sliding window 32

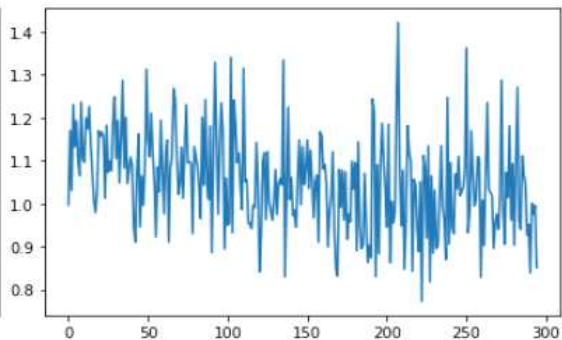


그림 35 sliding window 64

loss는 작은 epoch 안에서 수렴하고 있으나 학습이 계속되는 것에 대한 의미는 전체 시간에서의 예측과 비교하며 알 수 있다. epoch를 늘리더라도 loss는 크게 감소하지는 않지만 SOH의 감소를 더 잘 표현한다. 또한 target이 되는 SOH 값이 작을 때, (lstm



기준) 비교적 더 안정적으로 추정하였다. 대체적으로 10% 정도의 오차를 가지는 성능을 보이고 있다.

#### 4. 결론

배터리 운영 과정에서의 cycle을 기준으로 한 연속된 데이터로 LSTM을 통해 SOH를 추정하는 방법을 제시하였다. 대체적으로 10% 정도의 오차를 가지는 성능을 보이고 있다. 하지만 SOH의 80%를 배터리 수명이 다했다고 판단하는 것에 비추어 본다면 비교적 큰 오차이다. 이는 데이터의 일부 시점만으로 추정값을 판단하는 현재 방식에 비해 SOH는 배터리의 전체적인 history가 반영되어 나타나는 지표이기 때문이다. 따라서 현재의 방식에서 전체 배터리 데이터의 history를 반영하는 모델을 설계하여 더 정밀한 SOH를 추정하는 연구를 진행할 예정이다.

#### 5. 참고자료

<태양광 시스템 배터리 review paper>

Tian, Jinpeng, Rui Xiong, and Weixiang Shen. "A review on state of health estimation for lithium ion batteries in photovoltaic systems." *ETransportation*2 (2019): 100028.

<태양광 시스템 배터리 패턴>

Narayan, Nishant, et al. "Estimating battery lifetimes in Solar Home System design using a practical modelling methodology." *Applied energy*228 (2018): 1629-1639.

<신경망을 활용한 health 분석>

Dai, Houde, et al. "A novel estimation method for the state of health of lithium-ion battery using prior knowledge-based neural network and Markov chain." *IEEE Transactions on Industrial Electronics*66.10 (2018): 7706-7716.

Wu, Yitao, et al. "State of health estimation for lithium-ion batteries based on healthy features and long short-term memory." *IEEE Access*8 (2020): 28533-28547.

Li, Penghua, et al. "State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network." *Journal of power sources*459 (2020): 228069.

Chen, Cheng, et al. "State-of-charge estimation of lithium-ion battery using an improved neural network model and extended Kalman filter." *Journal of Cleaner Production*234 (2019): 1153-1164.

---

## 배터리 이상탐지 모델

---

# 시계열 이상치 데이터 분류 및 생성

## 1. 개요

시계열 데이터를 위한 이상 탐지 기법은 제조, 엣지디바이스, 배터리 데이터 등 많은 분야에 널리 응용되며 시계열 이상 탐지를 위한 알고리즘으로는 auto-regression, recurrent neural network(RNN), isolation forest 등 수많은 기법이 개발되었다. 이러한 노력에도 불구하고 아직 이러한 알고리즘들을 어떤 기준으로 평가할 것인지에 관한 연구는 미비하다. 현실의 데이터를 벤치마크로 사용하는 방안도 있으나 다양한 종류의 이상치가 혼재되어 있어 공정한 평가가 어렵다는 문제가 존재한다.

이상 데이터의 예시로, 다음 그림의 NYC 택시 데이터를 살펴보면 회색, 붉은색, 하늘색으로 표시한 부분이 각각 서로 다른 형태의 이상치임을 확인할 수 있다. 또한 실제 데이터에서 이상치를 라벨링하는 과정은 번거롭고 비용의 문제가 있으므로 연구자들은 실제 데이터보다, 직접 정의한 데이터에 이상치를 합성하는 방식을 선호한다.

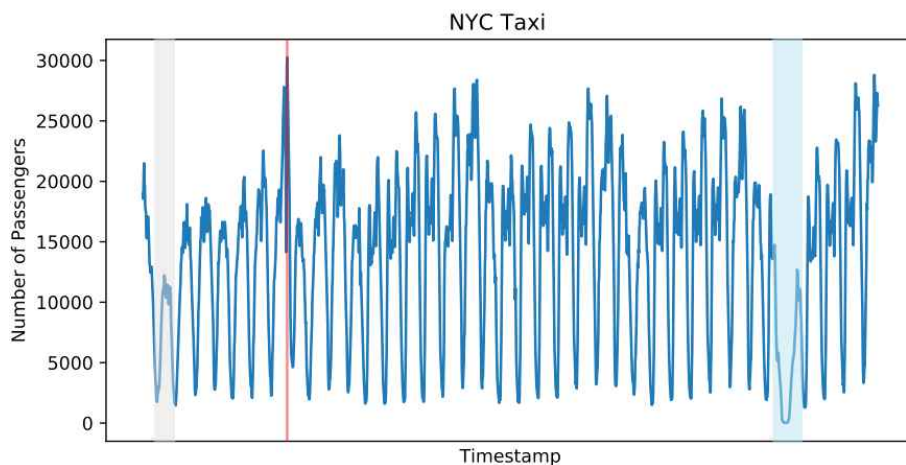


그림 37 NYC taxi 데이터의 서로 다른 종류의 이상치

하지만 아직 시계열 데이터에 존재하는 이상치에 대한 정의가 불명확하며 합의가 이루어지지 않은 상황인데, 어떤 정의를 채택하느냐에 따라 모델의 성능이 크게 달라질 수 있다. 따라서 본 논문에서는 서로 다른 형태의 이상치를 분류하는 기준을 제시하고, 이 기준에 따라 현존하는 알고리즘의 성능을 평가한다. 또한, 일반적인 데이터에 이상치를 합성하는 방법을 제시한다.

## 2. 기법

### ○ 시계열 데이터를 표현하는 방법

시계열 데이터를 모델링하는 가장 일반적인 방법은 데이터를 시간에 따른 점의 나열로 보는 것이다. 시간 1부터  $t$ 까지의 시계열 데이터  $X$ 는 다음과 같이 나타낼 수 있다.  $X = (x_1, x_2, \dots, x_t)$ . 선행 연구들은 이상치를 point, contextual, collective라는 세 특징으로 나누는데, 일반적인 연속적 표현은 이러한 특징을 잘 반영하지 못한다.

더 나은 방법은 시계열 데이터를 스펙트럼 분석(spectral analysis) 측면에서 바라보는 것이다. 스펙트럼 분석에서는 일반적으로 시계열을 삼각함수의 일차결합인  $X = \sum_n A \sin(2\pi\omega_n T) + B \cos(2\pi\omega_n T)$ 로 표현하여 모양 정보를 반영한다. 이 때  $A$ 와  $B$ 는 값의 범위를 결정하고,  $\omega_n$ 은  $n$  wave의 주기를 결정한다. 이러한 모델링은 시계열의 모양과 계절성은 효과적으로 표현하지만 추세는 표현하지 못하므로,  $\tau(T)$ 를 더 함으로써 이를 보완한다. 예를 들어, 선형적으로 증가하는 square sine 파형을 표현하기 위해서는  $A = \frac{1}{2n+1}, B = 0, \omega_n = 2n+1, \tau(T) = T$ 로 정의하면 된다.

아래 그림은 각각 point, contextual, collective 이상치의 예시를 나타낸다.

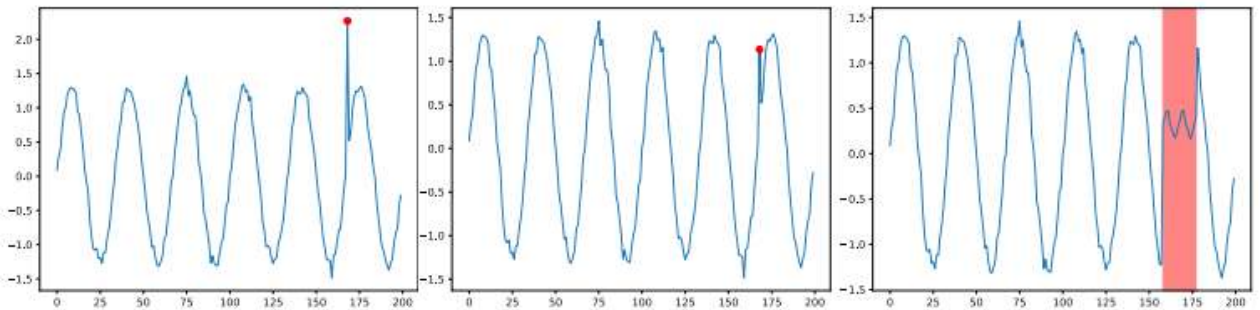


그림 38 point, contextual, collective 이상치의 예시

#### ○ 이상치 정의

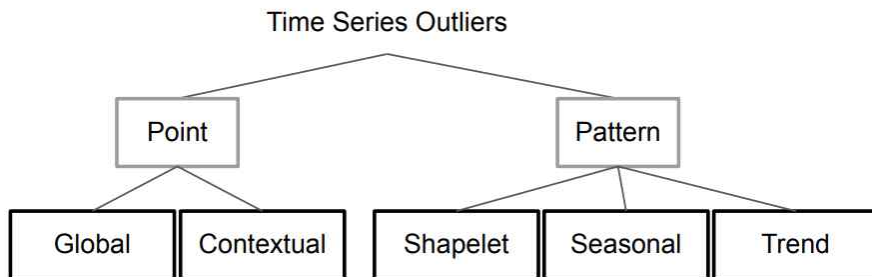


그림 39 이상치 분류도

점 이상치는 주변의 점들과 다르게 극단적인 값을 갖는 점을 말한다. 직관과 일치하게, 시계열  $X = (x_1, x_2, \dots, x_t)$ 가 주어지면 두 가지 이상치 유형은 다음 한계점을 넘는 값으로 정의할 수 있다.

$$|x_t - x'_t| > \delta$$

여기서  $x'_t$ 는 시간  $t$ 에서 regression이나 전체 평균, 혹은 주변의 점들에 의해 예상 가능한 값을 의미한다.

1) global outlier는 다른 모든 점들로부터 확연히 떨어져 있는 점을 말하며, 한계 점  $\delta$ 는 다음과 같이 정의된다.

$$\delta = \lambda \cdot \sigma(X)$$

여기서  $\sigma(X)$ 는 시계열의 표준편차이고  $\lambda$ 는 범위를 조정한다.

2) contextual outlier는 주변의 점들로부터 확연히 떨어져 있는 점을 말하며, 한계점  $\delta$ 는 다음과 같이 정의된다.

$$\delta = \lambda \cdot \sigma(X_{t-k, t+k})$$

여기서  $X_{t-k, t+k} = (x_{t-k}, x_{t-k+1}, x_{t-k+2}, \dots, x_{t+k})$ 이고  $\lambda$ 는 범위를 조정한다.

다음으로 패턴 이상치는 점이 아닌 시계열의 어떤 구간이 극단적인 값을 갖는 점들을 말한다. 이는 세 가지로 분류되는데, 각각 shaplet, seasonality, trend anomaly라 한다. 특히 시계열 데이터  $X$ 가 주어지고 시간  $i$ 부터  $j$ 까지인 부분 시계열  $X_{i,j}$ 은 다음과 같이 shapelet, seasonality, trend를 이용해 표현할 수 있다.

$$X_{i,j} = \rho(2\pi\omega T_{i,j}) + \tau(T_{i,j})$$

여기서  $\rho$ 는 기본적인 모양을,  $\omega$ 는 계절성을,  $\tau$ 는 추세를 나타낸다.

3) shaplet outlier는 기본적인 모양과 다른 모양을 갖는 부분 시계열을 말하며, 다음과 같이 정의된다.

$$s(\rho(\cdot), \hat{\rho}(\cdot)) > \delta$$

여기서  $s$ 는 두 부분열의 비유사성을 측정하는 함수이고,  $\hat{\rho}$ 는 기본적인 모양이며,  $\delta$ 는 한계점이다.

4) seasonal outlier는 전체의 계절성과 비교해 일반적이지 않은 계절성을 가진 부분시계열을 말하며, 다음과 같이 정의된다.

$$s(\omega, \hat{\omega}) > \delta$$

여기서,  $\hat{\omega}$ 는 기본적인 계절성이며,  $\delta$ 는 한계점이다.

5) trend outlier는 전체의 추세와 비교해 일반적이지 않은 추세를 가진 부분시계열을 말하며, 다음과 같이 정의된다.

$$s(\tau(\cdot), \hat{\tau}(\cdot)) > \delta$$

여기서,  $\hat{\tau}$ 는 기본적인 계절성이며,  $\delta$ 는 한계점이다.

다음 그림은 패턴 이상치의 예시를 나타낸다.

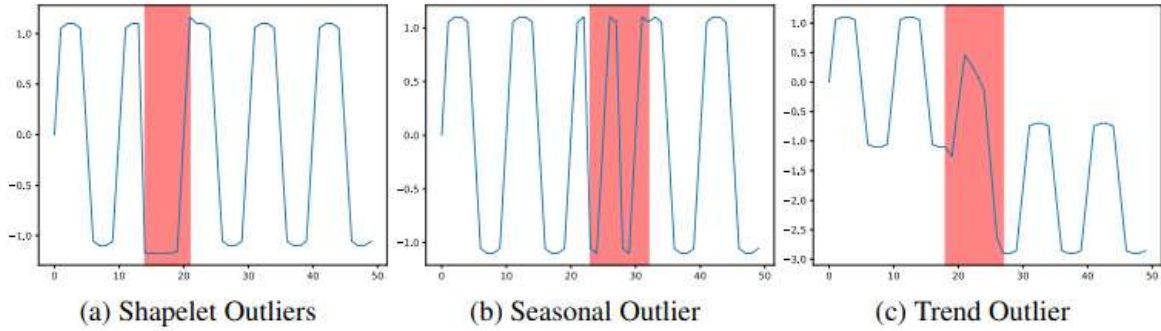


그림 40 패턴 이상치의 예시

#### ○ 이상치 합성

(1) global outlier는  $x_t = \mu(X) \pm \lambda \cdot \sigma(X)$ 로 합성할 수 있다. 여기서  $\mu(X)$ 는 평균을,  $\sigma(X)$ 는 표준편차를 나타내며,  $\lambda$ 는 범위를 조절한다.

(2) contextual outlier는  $x_t = \mu(X_{t-k,t+k}) \pm \lambda \cdot \sigma(X_{t-k,t+k})$ 로 합성할 수 있다.

(3) shapelet outlier는  $X_{i,j} = \rho(2\pi\hat{\omega}T_{i,j}) + \hat{\tau}(T_{i,j})$ 로 합성할 수 있는데,  $\hat{\omega}$ 은 기본적인 계절성,  $\hat{\tau}$ 는 일반적인 추세,  $\rho$ 는 합성할 모양을 나타낸다.

(4) seasonal outlier는  $X_{i,j} = \hat{\rho}(2\pi\omega T_{i,j}) + \hat{\tau}(T_{i,j})$ 로 합성할 수 있는데,  $\omega$ 는 합성할 계절성을 나타낸다.

(5) trend outlier는  $X_{i,j} = \hat{\rho}(2\pi\hat{\omega}T_{i,j}) + \tau(T_{i,j})$ 로 합성할 수 있는데,  $\tau$ 는 합성할 추세를 나타낸다.

### 3. 실험

#### ○ 데이터

합성 데이터는 20개의 univariate과 15개의 multivariate 데이터셋을 사용한다. 구체적으로, sinusoidal wave에 앞서 제시한 5가지 이상치를 합성하여 데이터를 생성한다.

실제 데이터는 credit card fraud detection, IoT for drinking water monitoring, server attack monitoring, extreme space weather detection을 사용한다.

#### ○ 실험 방법

평가할 모델로는 다음을 사용한다. Autoregression(AR)은 각 점이 지난 몇 개의 점과 선형적인 상관관계가 존재한다고 가정한다. Gradient boosting regression은 시계열 데이터를 windowed-fashion으로 바라본 segmented subsequences 기반 회

귀모형이다. recurrent neural network with long short term memory unit(LSTM-RNN)은 비선형적인 딥러닝 기반 모델이다. One-class SVM(OCSVM)은 일반적인 데이터와 이상치를 구분하는 decision boundary를 학습하는 모델이다. Isolation forest(IForest)는 트리 기반의 모델이다. Autoencoder(AE)는 데이터를 낮은 차원의 잠재 공간으로 매핑한 후 재현이 잘 되는지 측정하여 이상치를 찾아내는 모델이다. Generative adversarial network(GAN)는 generator와 discriminator를 이용한 생성 모델이다.

## ○ 실험 결과

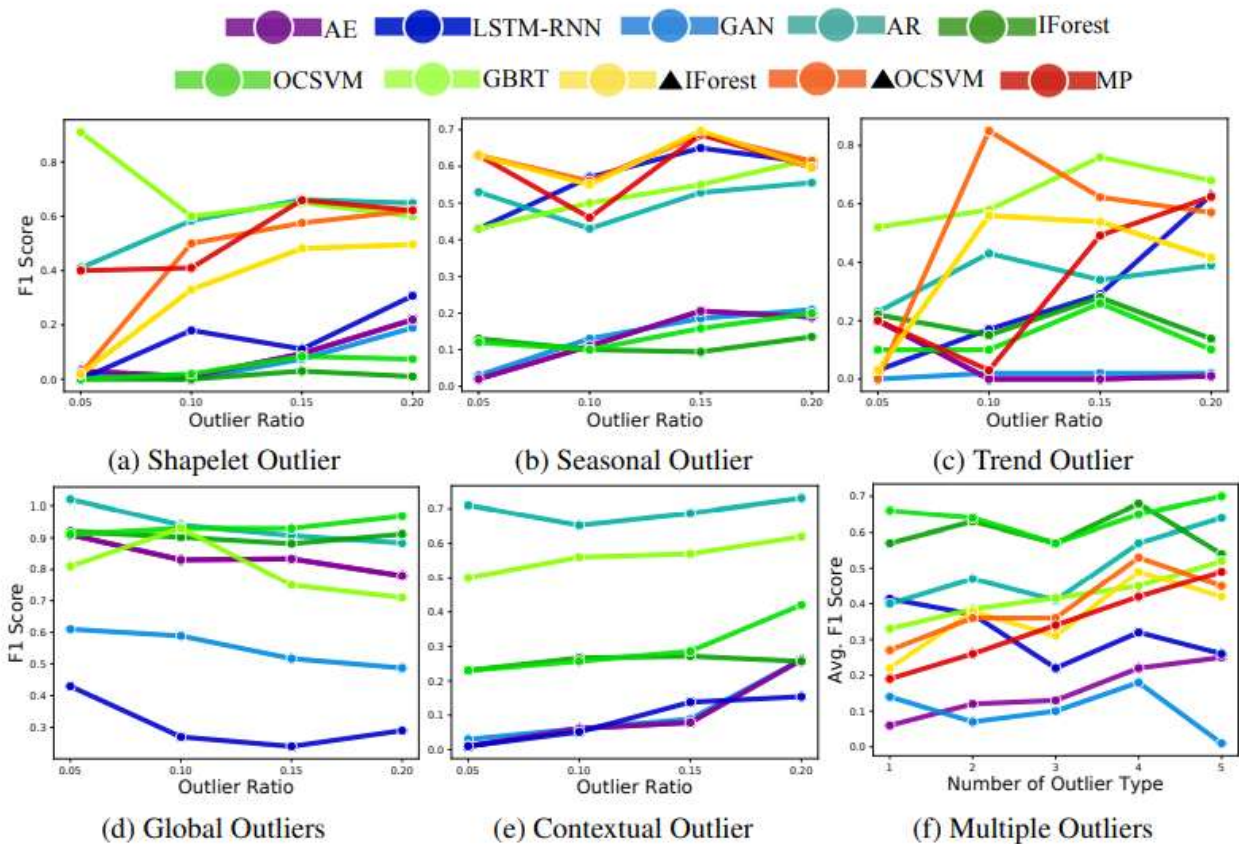


그림 41 대표적인 시계열 이상 탐지 모델 평가

F1 score를 측정했을 때 각 모델의 성능을 측정한 결과이다. 첫 번째로, 딥러닝 기반 방법보다 전통적인 방법이 모든 합성 데이터에서 더 뛰어난 성능을 보인다. 특히, AR이 contextual outlier와 shapelet outlier에서 압도적인 성능을 보인다. 둘째로, contextual outlier를 판별하는데 AR을 제외한 다른 모델은 모두 어려움을 겪는다. 셋째로, 예측 기반의 알고리즘들은 point-wise outlier를 판별하기 위해 설계되었지만 pattern-wise outlier 역시 잘 판별한다. 마지막으로, RNN과 GAN같은

딥러닝 방법론들은 제한된 형태의 이상치만을 잘 판별한다.

## 4. 결론

본 논문에서는 시계열 데이터의 이상치를 5가지로 분류하고 정의하였다. 이에 따라 35개의 합성 데이터를 생성하여 4개의 실제 데이터와 함께 9개의 기존 알고리즘의 성능을 평가하였다. 향후 더 많은 데이터셋을 추가하고 이상치의 정의를 다듬은 후 더 많은 알고리즘의 성능을 평가할 수 있을 것으로 기대한다.

## 5. 참고자료

- Lai, Kwei-Herng, et al. "Revisiting Time Series Outlier Detection: Definitions and Benchmarks." (2021).



# 배터리 이상탐지 모델

## 1. 개요

배터리 이상 데이터 검출 연구에 있어서 이상 데이터의 확보 여부는 매우 중요하다. 충분히 다양하고 많은 이상 데이터를 획득했을 경우 모델 학습을 위해 사용이 가능하고, 소량의 이상 데이터라도 획득했을 경우엔 모델의 검증을 위해 활용할 수 있다. 하지만 배터리 연구 환경의 특성상 소량의 이상 데이터도 확보할 수 있을지 장담하기 어렵다. 이에 대한 대책 중 하나로서, 정상 배터리 데이터를 기반으로 이상 데이터를 생성하는 방법이 있다. 시계열 데이터에 대한 이상치를 분류하고 정의하는 방법에 대해서 이미 알려진 부분이 어느정도 존재하지만, 배터리 데이터에 적용하기 위해선 배터리에서 발생하는 이상상황을 구체적으로 알아야한다. 본 논문에서는 배터리 수집과정에서 발생하는 4가지 이상상황(constant offset fault, stuck fault, time-delay fault, replay fault)에 주목했다. 정상 배터리 데이터를 기반으로 위의 4가지 이상 배터리 데이터를 생성하는 방법을 소개했다. 또한 이를 학습에 활용, 이상 배터리 데이터를 검출하고 분류하는 모델을 제안했다.

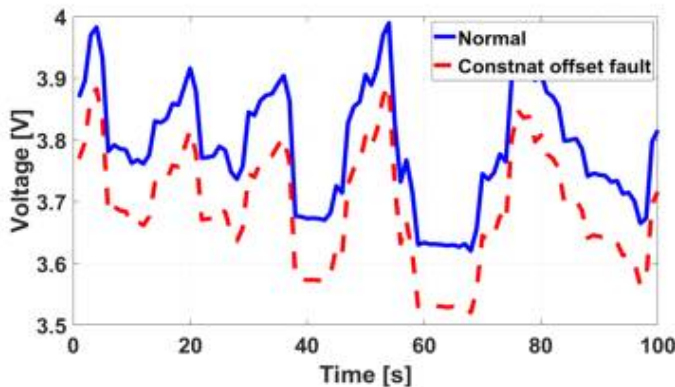
## 2. 기법

### ○ 배터리 이상 데이터

Cyber-Physical Layer	Battery Data	Vulnerabilities	Battery Data Faults	Comments	Fault Cases
Sensor Board	$V, I, T$	Aging and breakage, Temperature variation and error, EMI noises	<ul style="list-style-type: none"> <li>- Offset fault (bias, noises)</li> <li>- Scaling fault (gain)</li> <li>- Stuck fault</li> <li>- Drift fault</li> <li>- Intermittent fault</li> </ul>	These cases have been already reported a lot.	[24], [28], [33]-[35]
BMS Controller	$V, I, T, SOC, SOH, SOP, Capacity$	Algorithm manipulation attacks Data manipulation attacks (FDIA)	<ul style="list-style-type: none"> <li>- Inaccurate on-board health monitoring data (e.g., SOC, SOH, SOP)</li> <li>- Delay fault</li> <li>- Offset fault (bias, noises)</li> <li>- Stuck fault</li> <li>- Replay fault</li> </ul>	Since these cases have been less reported in literature, we explore by reverse engineering based on literature review in other applications.	[10], [25], [26], [27], [36]-[39]
Network	$V, I, T, SOC, SOH, SOP, Capacity$	Network failures Network attacks (FDIA and DoS attacks)	<ul style="list-style-type: none"> <li>- Stuck fault</li> <li>- Time delay fault</li> <li>- Packet drop</li> <li>- Delay data fault</li> <li>- Offset fault (bias, noises)</li> <li>- Stuck fault</li> <li>- Replay fault</li> </ul>	Since these cases have been less reported in literature, we explore by reverse engineering based on literature review in other applications.	[10], [19], [31], [32], [33]

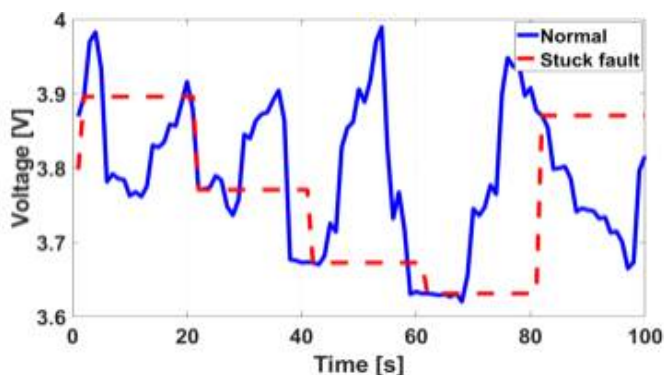
위는 Battery Energy Storage System(BESS)에서 발생하는 이상 데이터들의 종류를 정리한 표이다. 이 중에서 굵은 글씨로 표기한 4가지 이상상황에 주목, 각각 정상 데이터와 어떤 차이가 있는지를 분석했다.

1. Constant offset fault ( $V_{of}(t) = V(t) + \alpha$ )



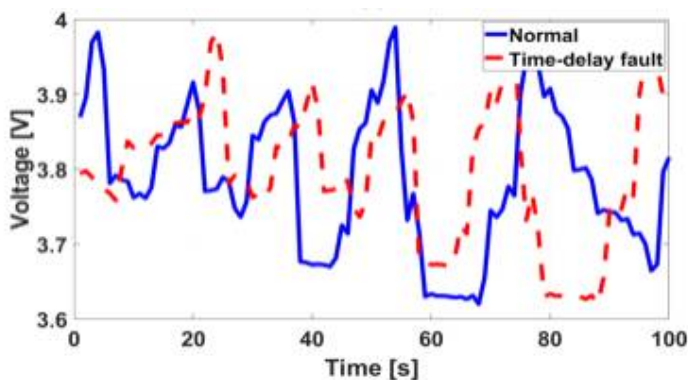
$V(t)$ 는 정상 전압 데이터,  $\alpha$ 는 constant offset value로서 본 논문에서는  $+0.1V$ ,  $-0.1V$ ,  $+0.2V$ ,  $-0.2V$ 의 값을 사용했다. 위의 그래프는 정상 데이터 (파란 실선)와  $\alpha$ 가  $-0.1V$ 일 때의 이상 데이터(빨간 점선)이다.

2. Stuck fault ( $V_{sf}(t) = V(t_{sf})$ )



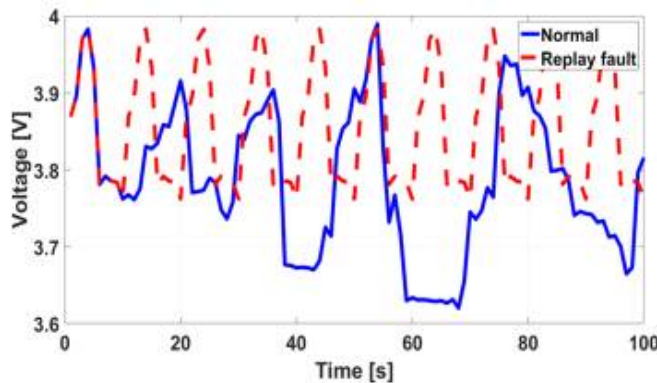
$t_{sf}$ 는 stuck fault가 발생했을 경우의 sampling time이고, 본 논문에서는 10초 혹은 20초의 값을 사용했다. 위의 그래프는 정상 데이터(파란 실선)와  $t_{sf}$ 가 10초일 때의 이상 데이터(빨간 점선)이다.

3. Time-delay fault ( $V_{df}(t) = V(t - t_{df})$ )



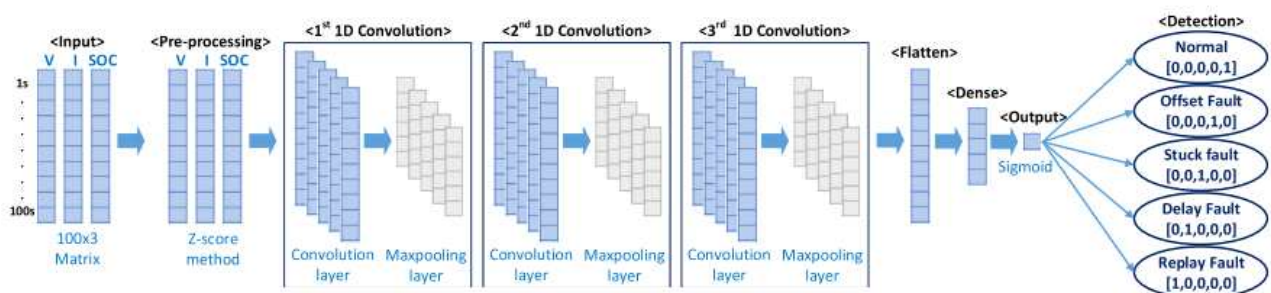
$t_{df}$ 는 delay time이고, 본 논문에서는 10초 혹은 20초의 값을 사용했다. 위의 그래프는 정상 데이터(파란 실선)와  $t_{df}$ 가 10초일 때의 이상 데이터(빨간 점선)이다.

#### 4. Replay fault ( $V_{rf}(t) = V(t_{rf})$ )



$t_{rf}$ 는 range for replay voltage data의 길이이고, 본 논문에서는 10초 혹은 20초의 값을 사용했다. 위의 그래프는 정상 데이터(파란 실선)와  $t_{rf}$ 가 10초일 때의 이상 데이터(빨간 점선)이다.

#### ○ CNN 기반의 배터리 이상 데이터 검출 및 분류 모델



일정 구간의 시계열 데이터를 입력으로 하는 1D-CNN, Long Short Term Memory(LSTM), Gated Recurrent Units(GRU), Support Vector Classifier(SVC), Decision Tree(DT), K-Nearest Neighbor(KNN), Random Forest(RF) 모델들에 대해 실험하고 결과를 비교했다. 위의 그림은 가장 성능이 뛰어났던 1D-CNN 분류 모델의 구조이다.

#### 1D-CNN 모델 구성

3개의 Convolution layer(CL), 3개의 Max-pooling layer(MPL)을 거치고, 최

종적으로 dense layer와 Sigmoid function을 거쳐서 5개의 클래스에 대한 분류 예측값을 출력한다.

각 CL와 MPL의 설정은 아래 테이블에 정리되어있다.

Layers	Number of Filters	Size of Filters (Kernel) <sup>a</sup>	Stride
1 <sup>st</sup> CL	32	4	1
1 <sup>st</sup> PL	32	3	1
2 <sup>nd</sup> CL	64	4	1
2 <sup>nd</sup> PL	64	3	1
3 <sup>rd</sup> CL	128	4	1
3 <sup>rd</sup> PL	128	3	1

### 3. 실험

#### ○ 데이터

노화 정도와 용량이 다른 3개의 배터리 셀을 4가지 전류 프로파일(Dynamic Stress Test(DST), DST with discharge-only, Urban Dynamometer Driving Schedule(UDDS), UDDS with discharge-only)로 실험해서 데이터의 다양성을 확보했다.

배터리의 전압, 전류, SOC를 1초간격으로 측정해서 길이 100에 3채널의 시계열 데이터를 생성했다. 이러한 시계열 데이터를 정상 클래스 데이터 100개, 4가지 이상 클래스 데이터 100개씩, 총 500개의 데이터를 학습에 사용했다. 또한 추가적으로 정상 클래스 데이터 50개, 이상 클래스 데이터 50개씩, 총 250개의 데이터를 모델 성능 검증에 사용했다.

전압, 전류 SOC의 스케일이 다르기 때문에 Z-Score 정규화 기법( $z_t = \frac{x_t - \mu}{\sigma}$ ,  $\mu$  : 평균,  $\sigma$  : 표준 편차)을 사용해서 전처리했다.

○ 실험 결과

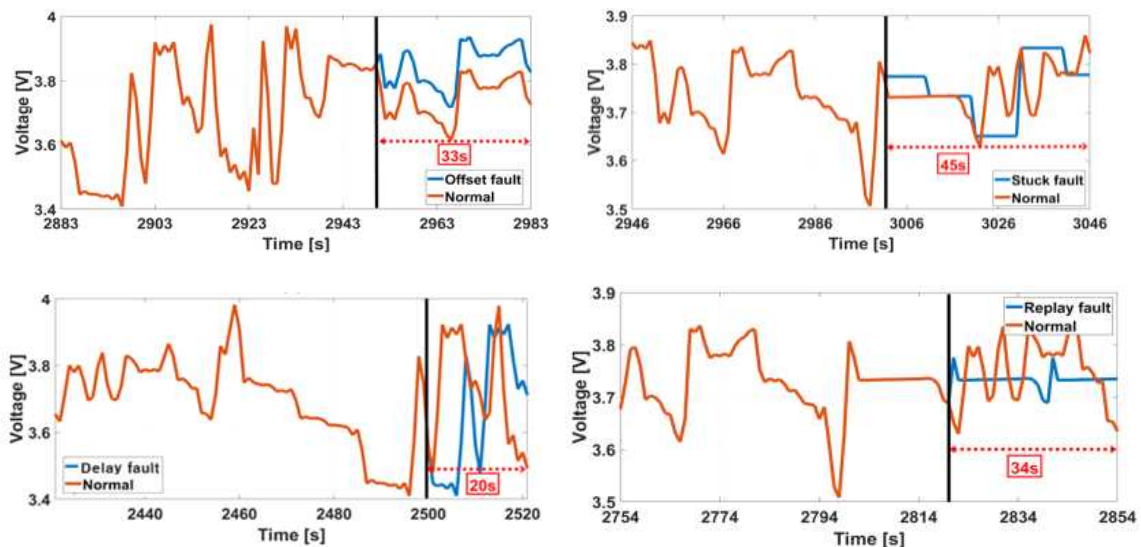
Class	Normal	Offset Fault	Stuck Fault	Delay Fault	Replay Fault
Normal	50	0	0	0	0
Offset Fault	0	50	0	0	0
Stuck Fault	0	0	47	3	0
Delay Fault	0	0	2	48	0
Replay Fault	0	0	0	0	50

위는 1D-CNN 분류 모델에 의한 250개의 검증용 데이터의 분류 결과이다. 총 250개의 데이터 중 245개를 맞춰, 98%의 정확도를 보였다.

LSTM, SVC 등 다른 모델과의 성능 비교는 아래와 같다.

Model	Precision	Recall	F-1	Accuracy
LSTM	0.83	0.82	0.82	0.82
GRU	0.87	0.86	0.86	0.86
SVC	0.79	0.77	0.78	0.78
DT	0.56	0.55	0.56	0.56
KNN	0.64	0.64	0.63	0.64
RF	0.73	0.72	0.72	0.72
Proposed CNN	0.98	0.97	0.98	0.98

또한 본 논문에서는 이상 데이터의 발생 시점으로부터 모델이 이상 데이터로 판별하기까지 어느 정도의 시간이 걸리는지를 분석했다. 그 결과 이상 데이터가 발생한 후 약 30초에서 50초의 시간이 지나면, 학습된 모델은 직전의 100초 길이의 시계열 데이터를 이상 데이터라고 판별했다.



## 4. 결론

### [결론]

BESS에서 발생하는 이상 상황을 정리하고, 이들 중 4가지 이상 상황을 분석했다. 배터리 시뮬레이션을 통해 얻은 정상 데이터를 기반으로 4가지 이상 데이터를 생성, 이를 학습하는 1D-CNN 모델을 개발했다. 250개의 검증용 데이터에 대해서 98%의 정확도를 달성했고, 이상 데이터의 발생 시점으로부터 30초 ~ 50초만에 이상 상황을 검출해냈다.

### [본 연구로의 적용]

ESS 데이터에서의 이상 상황에 대한 구체적인 예시와 설명이 주어진다면, 이를 분석해서 이상 상황에 대한 분류를 진행할 수 있을 것이다. 그 후 ESS의 정상 데이터를 기반으로 노화 정도와 전류 프로필이 다른 이상 데이터들을 여럿 생성한 뒤, 이 이상 데이터들을 판별하는 모델을 개발할 수 있을 것이다.

## 5. 참고자료

- Hyun-Jun Lee. et al. Convolutional Neural Network-Based False Battery Data Detection and Classification for Battery Energy Storage Systems - IEEE, 2021