

# 실제 운용 환경에서의 배터리 용량 추정을 위한 최적 회귀 모델 선정 연구

신승화\*, 김종훈\*

충남대학교 에너지저장변환연구실\*

## A study on the selection of the optimal regression model for the estimation of battery capacity in a realistic operation environment

Seunghwa Sin\*, Jonghoon Kim\*

Energy Storage Conversion Lab., Chungnam National University\*

### ABSTRACT

배터리의 열화는 비선형적으로 발생하기 때문에, 부정확한 용량 추정으로 인한 안전사고가 발생할 수 있다. 정확한 배터리의 용량 추정을 위한 다양한 연구에서 회귀 모델이 사용된다. 대부분의 연구에서는 실험을 통해 취득된 데이터를 기반으로 회귀 모델의 설계와 검증이 진행된다. 따라서 측정되는 데이터의 노이즈가 크지 않으며 발생하는 이상치는 데이터 전처리 과정을 통해서 제거될 수 있다. 하지만 실제 운용 환경에서는 어플리케이션에 따라 노이즈가 크게 발생할 수 있으며, 미처 제거하지 못한 이상치가 포함될 가능성이 있다. 따라서 이러한 상황을 고려하여 널리 사용되는 회귀 모델 간의 성능을 비교하고 최적의 회귀 모델을 선정한다.

### 1. 서론

세계적으로 기후변화에 대응하기 위해 배터리가 주목받고 있다. 환경 보호 측면에서 배터리는 화석 연료를 대체할 수 있는 청정 에너지 저장 방법으로, 탄소 배출을 줄이고 기후 변화를 완화하는 데 중요한 역할을 한다. 특히 전기차(Electric Vehicles; EV)의 보급이 급증하면서 고성능 배터리의 수요가 폭발적으로 증가하고 있다. 배터리가 적용되는 범위가 확장됨에 따라, 배터리의 안전하고 효율적인 운용을 위한 방안들이 탐색되고 있다. 대표적으로 배터리의 용량을 추정하기 위해 다양한 연구들이 진행되고 있다. 배터리의 용량은 비선형적으로 감소하기 때문에, 정확한 용량이 추정되지 않는다면 갑작스럽게 어플리케이션의 동작이 정지되거나 각종 안전사고가 발생할 수 있다. 따라서 배터리의 용량을 추정하기 위해 회귀 모델은 다양한 연구에서 사용된다. 회귀 모델은 주로 실험을 통해 취득된 데이터를 통해 설계 및 검증이 진행된다. 하지만 실험 데이터는 발생하는 노이즈의 크기가 작으며, 이상치가 거의 발생하지 않는다. 또한 데이터 전처리를 통해 발생한 이상치는 제거될 수 있다. 실제 운용 환경에서는 어플리케이션에 따라 노이즈가 크게 발생할 수 있으며, 이상치를 미처 제거하지 못할 수 있다. 따라서 본 논문에서는 노이즈와 이상치가 발생한 상황을 고려하여, 용량 추정을 위한 최적의 회귀 모델을 선정한다. 일반적으로 회귀 모델 학습에 사용되는 데이터셋을 구성한 후, 노이즈와 이상치를 포함시킨다. 노이즈와 이상치가 포함되어 있는 데이터셋에 대해 회귀 모델의 성능을 변화를 확인하고, 최적의 모델을 선정한다.

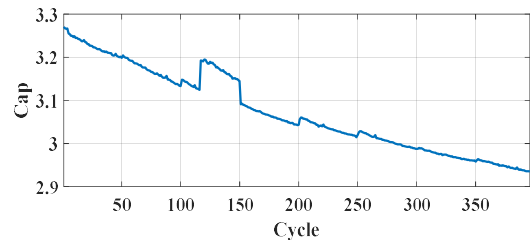


그림 1 배터리 열화 실험 결과  
Fig. 1 Result of Battery degradation test

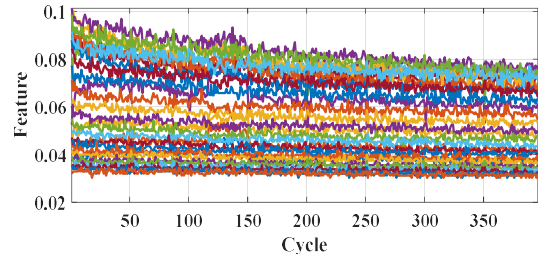


그림 2 열화 실험 데이터로부터 추출한 IC  
Fig. 2 Extracted IC from degradation experiment data

### 2. 배터리 열화 실험 및 Feature 추출

본 논문에서 회귀 모델의 학습 및 검증에 사용할 데이터를 취득하기 위해, LiNiMnCoO<sub>2</sub> 계열 배터리에 대해 열화 실험을 진행한다. 열화 실험은 완전 충전과 완전 방전을 반복하며 배터리를 열화 시킨다. 취득된 데이터의 방전 구간에서의 전류를 누적하여, 배터리의 용량을 계산하며 그림 1과 같다.

열화 실험 데이터로부터 회귀 모델들에 사용할 데이터셋을 생성한다. 데이터셋은 추정하고자 하는 용량과 용량의 변화를 간접적으로 나타낼 수 있는 Feature로 구성된다. 본 논문에서는 충전 구간에서 고정된 전압 간격에서의 전류를 적산한 용량을 나타내는 증분용량(Incremental Capacity; IC)을 Feature로 사용하였다<sup>[1]</sup>. IC는 수식 (1)과 같이 계산되며, 계산 결과는 그림 2와 같다.

$$\frac{dQ}{dV} = \frac{\Delta Q}{\Delta V} = \frac{Q_i - Q_{i-1}}{V_i - V_{i-1}} \quad (1)$$

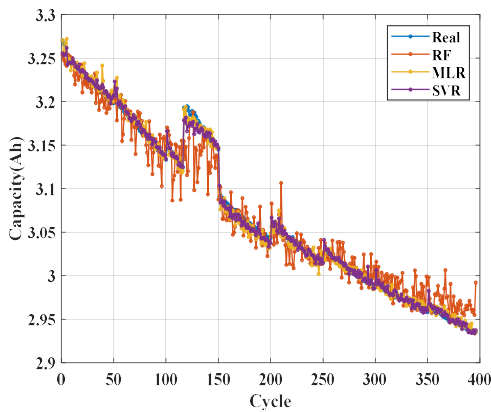


그림 3 회귀 모델 별 용량 추정 결과  
Fig. 3 Capacity estimation results by regression model

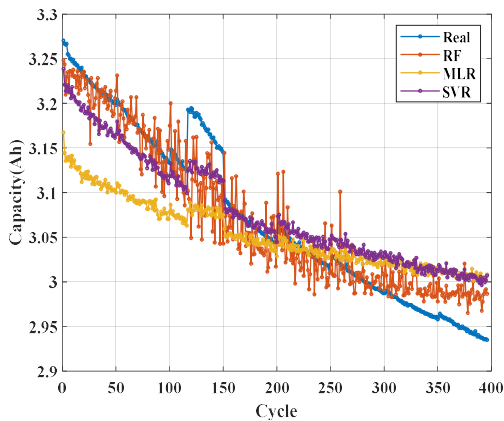


그림 4 실제 운용 환경에서의 회귀 모델 별 용량 추정 결과  
Fig. 3 Capacity estimation results by regression model under realistic operation environment

Feature를 추출한 구간은 3.7V에서 4.0V이며, 수식 (1)에서 전압 간격에 해당하는  $\Delta V$ 는 0.01V로 설정하였다. 따라서 IC는 1 사이클의 충전 구간에서 총 30개가 추출될 수 있다.

### 3. 최적 회귀 모델 선정

최적 회귀 모델을 선정하기 위해, 데이터셋에 실제 운용 환경에서 발생할 수 있는 종류의 노이즈와 이상치를 추가한다. 노이즈를 추가하기 위해, Feature인 IC에 -50에서 50까지의 값을 랜덤으로 생성하여 그림 2의 결과에 추가하였다. 또한 이상치를 추가하기 위해, 약 0.14의 값을 각 Feature마다 랜덤한 위치에 추가하였다.

비교를 위한 회귀 모델은 많은 연구에서 사용되는 랜덤 포레스트(Random forest; RF), 다중 선형 회귀(Multiple linear regression; MLR), Support Vector Regression(SVR)를 사용하였다. RF는 앙상블 모델로서 훈련 과정에서 생성된 다수의 결정 트리들의 결과를 종합하여 최종 추정값을 결정한다<sup>[2]</sup>. MLR은 다수의 독립변수의 변화에 따라 종속변수의 변화를 선으로 추정하는 기법이며, 별도의 하이퍼파라미터는 없이 동작 가능한 모델이다. SVR은 Support Vector Machines (SVM)를 기반으로 한 회귀 분석 기법이며, 비선형적인 데이터들을 처리하기 위해 커널이 사용된다.

그림 3은 데이터셋에 노이즈 및 이상치를 포함하지 않았을 때, 회귀 모델 별 용량 추정 결과를 나타낸다. 추정 결과에 대한 회귀 모델 별 정확도는 표 1과 같다.

표 1 회귀 모델 별 추정 정확도 비교

Table 1 Comparison of estimation accuracy by regression model

	RF	MLR	SVR
MAE	0.0145	0.0052	0.0042
RMSE	0.0197	0.0067	0.0059
R2	0.9576	0.9950	0.9962

그림 4는 데이터셋에 노이즈 및 이상치를 포함한 회귀 모델 별 용량 추정 결과를 나타낸다. 추정 결과에 대한 회귀 모델 별 정확도는 표 2와 같다.

표 2 실제 운용 환경에서의 회귀 모델 별 추정 정확도 비교

Table 2 Comparison of estimation accuracy by regression model under realistic operation environment

	RF	MLR	SVR
MAE	0.0253	0.0526	0.0328
RMSE	0.0336	0.0634	0.0371
R2	0.8765	0.5610	0.8494

노이즈 및 이상치가 포함되지 않았을 경우, MLR과 SVR이 RF보다 높은 추정 정확도를 나타낸다. 하지만 노이즈 및 이상치를 포함하였을 경우, RF의 추정 정확도가 가장 높게 나타났다.

## 4. 결론

본 논문은 실제 운용 환경을 고려하여 최적의 회귀 모델 간 비교를 진행하였다. 실제 운용 환경을 고려하였을 때, 노이즈 또는 이상치가 발생할 수 있어 구성된 데이터셋에 포함하였다. 회귀 모델 별 용량 추정 결과를 확인하였을 때, 노이즈 및 이상치가 포함되지 않을 경우 MLR과 SVR가 높은 정확도를 나타냈다. 하지만 노이즈 및 이상치가 포함되었을 때, RF가 가장 높은 정확도를 나타냈다. 특히 MLR의 정확도가 크게 떨어졌다. 이는 RF와 SVR은 하이퍼파라미터의 최적화 과정을 통해 모델 성능을 향상시킬 수 있지만, MLR은 고려할 수 있는 하이퍼파라미터가 없기 때문이다. 따라서 실제 운용 환경을 고려했을 때, 하이퍼파라미터를 반영할 수 있는 모델을 선정해야 하고, 노이즈 및 이상치를 고려한 환경에서의 검증이 필요하다.

## 참고 문헌

- [1] ZHOU, Ruomei, et al. State of health estimation for fast-charging lithium-ion battery based on incremental capacity analysis. *Journal of Energy Storage*, 2022, 51: 104560.
- [2] LI, Yi, et al. Random forest regression for online capacity estimation of lithium-ion batteries. *Applied energy*, 2018, 232: 197-210.