

MLOps 플랫폼 기반 ESS 데이터 분석 및 예측 애플리케이션 서비스에 관한 연구

김창우, 최효섭

한국전자기술연구원

cwkim@keti.re.kr, hschoi@keti.re.kr

A Study on the ESS Data Analysis and Prediction Application Services Based on the MLOps Platform

Changwoo Kim, Hyo-sub Choi

Korea Electronics Technology Institute (KETI)

요약

배터리 제조 시 발생하는 결함, 배터리 셀 기반 충·방전 운영에서의 결함, 통합 관리 시스템의 부재 등 ESS(에너지 저장 장치)에 관한 안전사고에 따라 사회 전반에 걸쳐 안전을 위한 연구 및 개발이 활발히 진행 중이다. ESS 운영 및 실험 데이터를 기반으로 안전 원인을 분석 및 예측한 애플리케이션이 다양하며, 시계열 데이터 특성에 맞춰 주기적 업데이트 및 모듈 변경 등 워크플로 관점에서 반복되는 상황이 발생하고 있다. 본 논문에서는 반복되는 작업을 최소화하고 다양한 연구자들에게 머신러닝 모델링에 집중할 수 있는 환경을 제공하기 위해 MLOps 플랫폼을 활용한 ESS 데이터의 분석 및 예측 애플리케이션 서비스 방법론을 설명한다. 이를 통해 연구자가 정의한 데이터셋을 기반으로 데이터 전처리에서부터 데이터 학습까지 분할된 처리 모듈로 구성된 머신러닝 파이프라인 워크플로가 어떠한 흐름으로 동작하는지 데이터 분석 및 학습 처리의 구현 과정을 보여준다.

I. 서론

최근 ESS(에너지 저장 장치) 및 EV(전기차)에서 사용되는 리튬이온 등 배터리 활용이 오용 등으로 인해 안전사고로 이어지고 있다. 구체적으로 배터리 제조 시 발생하는 배터리 내부 결함, 배터리 셀 기반 충·방전 운영 상황에서의 결함, BMS 및 EMS 등 통합 관리 시스템의 부재 등 다양한 요소가 사고의 원인으로 분석되고 있다.[1] 이 중 소프트웨어 측면의 안전 사고를 극복하기 위해 배터리 충·방전 운영 사이클, 노화 및 이상 실험 등을 토대로 데이터 분석 및 예측으로 예지보전을 수행하는 연구가 활발히 진행 중이다.[2] 또한, 최근 데이터 과학 분야에서 데이터 분석 및 학습을 통해 인공지능 모델을 개발함에 있어 다수의 연구자가 공동 개발 작업을 하고 있으며 머신러닝 모델링 개발에 보다 집중하기 위해 MLOps 플랫폼을 활용한 ML 파이프라인 자동화 처리 기술이 활발히 활용되고 있다.[3] 본 논문에서는 ESS 배터리 데이터를 기반으로 데이터 분석 및 예측하는 애플리케이션에 효과적인 개발 및 공개 서비스를 위해 MLOps 플랫폼을 활용한 방법론을 소개한다. 그림 1처럼 연구자가 정의한 데이터셋을 토대로 컨테이너 이미지로 구성된 머신러닝 파이프라인 워크플로가 어떠한 흐름으로 동작하는지 데이터 분석 및 학습 처리의 구현 과정을 보여준다.

II. 본론

MLOps 플랫폼은 머신러닝 모델 개발과 머신러닝 운영 워크플로에서 발생하는 문제 및 반복을 최소화하기 위한 분야로 데이터 분석가에서 ML 모델링에 집중할 수 있는 인프라를 제공하고 피드백에 주기적으로 반영되도록 만드는 환경을 제공하는 플랫폼이다. 구체적으로 데이터 분석 및 예측 연구개발 단계에서 원천 데이터의 수집을 토대로 데이터의 전처리 및 분석하여 학습 모델을 위한 데이터셋을 구성하고 이를 특정 ML 모델이 학습 및 검증으로 모델의 하이퍼파라미터 및 가중치 등을 구한다. 그리고 새로운 데이터 및 이상 상황에 대한 데이터를 이용할 때마다 모델의 재학습, 재배포를 반복하여 수행한다. ML 파이프라인을 구성하기 위한 처리 과정에서 반복되는 SW 모듈들은 단일 또는 복합 파이프라인으로 생성하여 CT(지속적 학습) 처리함으로써 보다 빠른 적용 및 공유를 할 수 있다.

표 1에서 보듯 다양한 MLOps 플랫폼을 분석하였고 본 논문에서는 데이터 관리, ML 모델 훈련, ML 모델 배포, 모니터링, ML 모델 추적 및 버전 관리, 파이프라인 오케스트레이션 등 다양한 MLOps 플랫폼의 요소를 고려하여 오픈소스 기반인 Kubeflow 플랫폼을 선정하였다.[4]

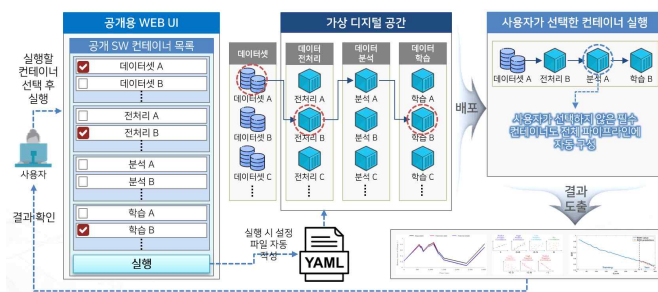


그림 1. MLOps 플랫폼 기반 데이터 분석 및 예측 애플리케이션 서비스 구조

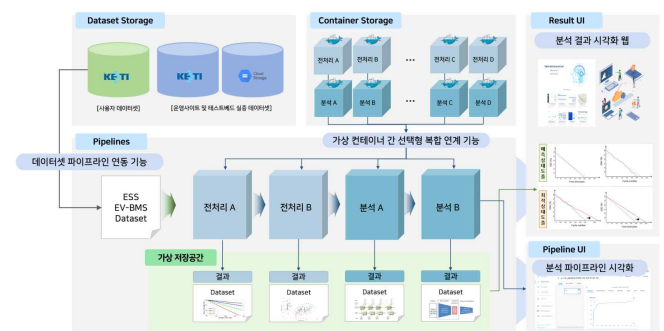


그림 2. Kubeflow 플랫폼의 데이터 수집 및 저장 스토리지 구조

표 1 MLOps 플랫폼의 종류

프로젝트명	설명	분야	중점
AWS Sagemaker	완전 관리형 인프라, 도구 및 워크플로를 사용하여 모든 사용 사례에 대해 기계 학습(ML) 모델을 구축, 교육 및 배포	추적 및 버전 관리, 파이프라인 오케스트레이션, 모델 배포	AWS
ClearML	2줄의 코드만 있는 MLOps. 대규모 ML 워크플로를 쉽게 개발, 조정 및 자동화	추적 및 버전 관리, 파이프라인 오케스트레이션, 모델 배포	실험, 정형 데이터
cnvrg.io	AI 모델을 대규모로 구축 및 배포하기 위한 엔드 투 엔드 머신 러닝 플랫폼	추적 및 버전 관리, 파이프라인 오케스트레이션, 모델 배포	Technology agnostic
Dataiku	Dataiku는 데이터 액세스를 민주화하고 기업이 인간 중심적인 방식으로 AI에 대한 자체 경로를 구축할 수 있도록 하는 플랫폼	추적 및 버전 관리, 파이프라인 오케스트레이션, 모델 배포	엔터프라이즈, 데이터 분석, 비즈니스 인텔리전스
Kubeflow	Kubeflow 프로젝트는 Kubernetes에서 기계 학습(ML) 워크플로를 간단하고 이식 가능하며 확장 가능하게 배포	추적 및 버전 관리, 파이프라인 오케스트레이션, 모델 배포	커뮤니티, 확장성
MLflow	MLflow는 중단 간 기계 학습 수명 주기를 관리하기 위한 오픈 소스 플랫폼	추적 및 버전 관리	실험, Spark

Kubeflow 플랫폼을 활용하기 위해서는 GCP, AWS, Azure 등의 상용 클라우드 플랫폼을 활용하거나 온프레미스 형태로 자체 구축한 사이트를 이용할 수 있다. 본 논문에서는 온프레미스 클라우드 형태로 Kubernetes 플랫폼 위에 설치된 Kubeflow 플랫폼을 대상으로 기술하였다.[5]

또한, 본 논문에서 ML 파이프라인은 데이터 수집, 전처리, 분석, 예측 등과 같은 머신러닝 요소를 이용하여 ML 파이프라인을 단순화하고 모듈 단위로 수정 및 삽입할 수 있도록 각 처리 요소는 컨테이너 이미지로써 구성을 하고 이미지들을 컨테이너 저장소에서 불러와서 사용할 수 있도록 설정 파일을 구성하였다. YAML 파일로 작성된 Kubeflow의 ML 파이프라인 설정 파일을 작성하기 위해 Python 3.X 기반의 SDK를 활용하여 Kubeflow Pipeline(KFP)를 작성하였다.

ESS 데이터를 기반으로 특정 컬럼 값을 예측하는 ML 파이프라인 구성 예제는 한국전자기술연구원에서 운영 중인 Github 저장소에 수록되어 있다. (https://github.com/kei-dp/OpenESS/blob/main/MLOps/Onprem_pipeline_test_v3.ipynb) 위 파이썬 소스코드는 기 작성된 컨테이너 이미지를 불러와서 사용하는 파이프라인 정의 코드로 KFP에서 필요한 파이프라인 패키지 파일을 생성하여 동작시킬 수 있도록 구성되어 있다. 해당 파이프라인 정의 코드는 사전에 생성된 PVC(Persistent Volume Client) 연동 이후 데이터 선택, 데이터 분리, 데이터 전처리, AI 모델(분류, 회귀, 전이학습 등) 등 4단계의 컨테이너 이미지를 활용하도록 구성되어 있다.

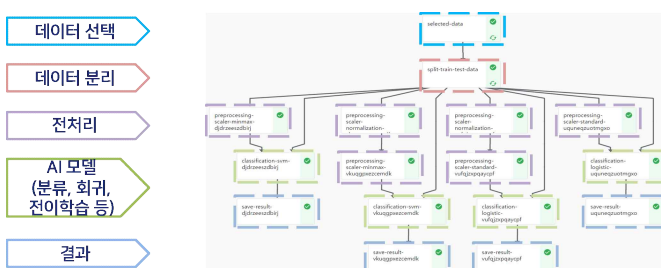


그림 3. Kubeflow 파이프라인 설계 및 구현 (예시)

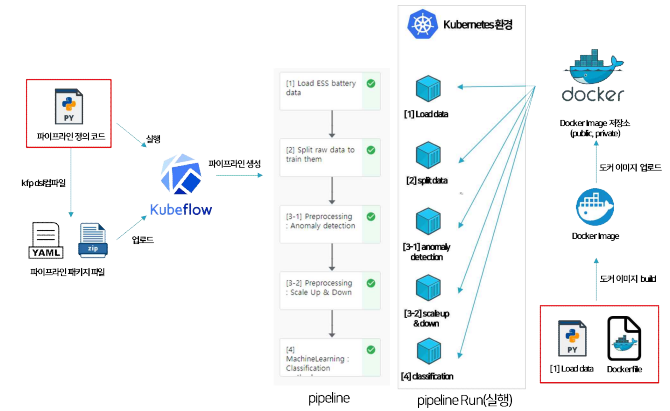


그림 4. Kubeflow 파이프라인 정의 코드 및 동작 흐름 (예시)

III. 결론

본 논문에서는 ESS 배터리 데이터를 기반으로 데이터 분석 및 예측하는 애플리케이션에 효과적인 개발 및 공개 서비스를 위해 MLOps 플랫폼을 활용한 방법론을 소개하였다. 연구자가 정의한 데이터셋을 기반으로 데이터 전처리에서부터 데이터 학습까지 분할된 처리 모듈로 구성된 머신러닝 파이프라인 워크플로가 어떠한 흐름으로 동작하는지 데이터 분석 및 학습 처리의 구현 과정을 보여주었다. 인공지능 모델을 개발하는 데 있어 이러한 데이터 분석 및 학습 처리 방법을 통해 다수의 연구자가 공동 개발 작업을 능동적으로 수행할 수 있고 머신러닝 모델링 연구 및 개발에 집중할 수 있다.

ACKNOWLEDGMENT

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00077, 대규모 분산 에너지 저장장치 인프라의 안전한 자율운영 및 성능 평가를 위한 지능형 SW 프레임워크 개발)

참고 문헌

- [1] Komarnicki, P., Lombardi, P., and Styczynski, Z.A. "Electric Energy Storage System," Electric Energy Storage Systems. Springer, Berlin, Heidelberg, pp. 37-95, 2017.
- [2] R. Schainker, "Executive overview: Energy storage options for a sustainable energy future," in Proc. IEEE Power Eng. Soc. General Meeting, pp. 2309 - 2314, 2004.
- [3] D. Kreuzberger, N. Kühl, and S. Hirschl, "Machine learning operations (mlops): Overview, definition, and architecture," arXiv preprint arXiv:2205.02302, 2022.
- [4] "Kubeflow." Kubeflow. 2023년 1월 6일 접속, <https://www.kubeflow.org/>
- [5] "Kubernetes." Kubernetes. 2023년 1월 6일 접속, <https://kubernetes.io/>