

NGSI-LD 인터페이스 활용 머신러닝 파이프라인 구축 방법

정승명*, 김성운, 안일엽

*한국전자기술연구원

*sm.jeong@keti.re.kr, chosun@csu.ac.kr, @keti.re.kr

A deployment method on the machine learning pipeline based on NGSI-LD

SeungMyeong Jeong*, Seongyun Kim, Il-yeup Ahn

*Korea Electronics Technology Institute

요약

유럽의 ICT 표준화 기구인 ETSI는 링크드 데이터를 지원하는 데이터 수집 및 공유 표준 인터페이스인 NGSI-LD를 정의하였다. NGSI-LD 인터페이스는 국내 및 유럽의 다양한 데이터 플랫폼에 적용되고 있으며 이와 연계한 머신러닝 기반 서비스에도 활용 사례가 늘어나고 있다. 본 논문에서는 NGSI-LD 플랫폼에 수집된 이력 데이터를 활용해 머신러닝 모델을 생성하고 이를 통해 예측 서비스를 제공하기 위한 NGSI-LD 기반 머신러닝 파이프라인 구축 방법을 제안한다. 구축한 머신러닝 파이프라인은 도시 수자원 데이터 플랫폼에 연계하여 물수요 등 물정보 예측 서비스 개발에 활용하여 해당 시스템 구축 사례 또한 본 논문에서 소개한다.

I. 서론

머신러닝을 활용한 예측 또는 자동화 서비스가 늘어나고 있는 가운데 머신러닝 모델 생성, 배포 그리고 실행(serving)하는 머신러닝 파이프라인을 구현하기 위한 다양한 기술이 개발되고 있다. 본 논문에서는 NGSI-LD 표준을 적용한 데이터 플랫폼에 적용할 수 있는 머신러닝 파이프라인 구축 방안을 제시한다.

모델 생성에 있어서 대규모 학습 데이터셋 구축이 중요하며 oneM2M과 같은 IoT 표준 인터페이스를 확장하여 데이터셋 구축 기능을 플랫폼에서 지원하는 방안도 연구된 바 있다[1].

NGSI-LD는 oneM2M과 같은 IoT 플랫폼, 공공 데이터 포털 등에서 데이터를 수집하여 링크드 데이터로 표현하며 데이터를 저장하고 조회할 수 있는 REST API를 제공한다[2]. 이러한 NGSI-LD를 기반으로 구축한 데이터 플랫폼에서 표준 API를 활용한 데이터 연계 및 분석 예측 모델을 생성하여 예측 결과를 저장하고 다른 서비스에 활용하는 사례도 표준 인터페이스 활용의 장점을 보여준다[3].

본 논문에서 제안하는 방안은 기존 연구에서 주로 다른 표준 인터페이스를 활용한 학습 데이터 구축이나 내재화된 알고리즘에 의한 모델 생성보다는 다양한 요구사항을 가지는 전처리나 학습 모델 실행 관리를 유연하게 수행하기 위한 Node-RED 활용 파이프라인 구축에 중점을 둔다.

II. 본론

본 논문에서는 NGSI-LD 표준 인터페이스로 상호 연동하는 머신러닝 파이프라인을 구축하고 이를 도시 수자원 예측 서비스로 시범 적용한 사례를 소개한다. 시스템은 그림 1과 같이 NGSI-LD를 적용한 데이터 플랫폼 및 플랫폼과 연동하는 데이터 소스, 모델 개발자, 모델 실행 플랫폼 및 데이터 처리를 위한 Node-RED로 구성한다. 도면1에서 NGSI-LD 인터페이스는 실선으로 표시되어 파이프라인 구축에 표준 인터페이스 적용의 중요성을 확인할 수 있다.

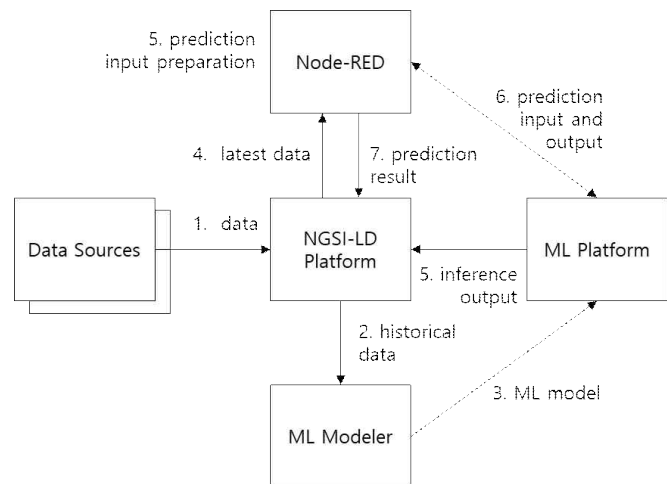


그림1 NGSI-LD 기반 머신러닝 파이프라인 구성 및 데이터 흐름도

그림1은 NGSI-LD 플랫폼을 중심으로한 전체 시스템 구성도 및 데이터의 흐름을 개략적으로 나타낸다. 수집된 데이터는 플랫폼으로 저장되고 머신러닝 모델 개발자는 플랫폼에서 이력 데이터를 조회하여 학습 데이터셋을 구축하고 모델을 생성하고 생성한 모델은 머신러닝 플랫폼에 등록한다. 머신러닝 플랫폼은 tensorflow 기반으로 개발되어 있으며 모델 실행을 위한 REST API를 제공한다.

예측 수행 요청은 Node-RED로 구성하였으며, 예측에 필요한 입력 데이터는 NGSI-LD 플랫폼에 수집되는 최신 데이터를 주기적으로 조회하든지 이벤트 기반 데이터는 구독/통지 인터페이스를 이용해서 입력 데이터를 준비한다. 이후 해당 데이터를 머신러닝 플랫폼의 예측 API 호출에 사용하여 예측 결과를 회신받고, 예측 결과는 다시 NGSI-LD 플랫폼에 저장한다.

머신러닝 플랫폼은 모델 개발자에게 모델 등록 관리를 위한 웹서비스를 제공하며 예측 결과를 가시화하기 위한 대시보드를 제공한다.

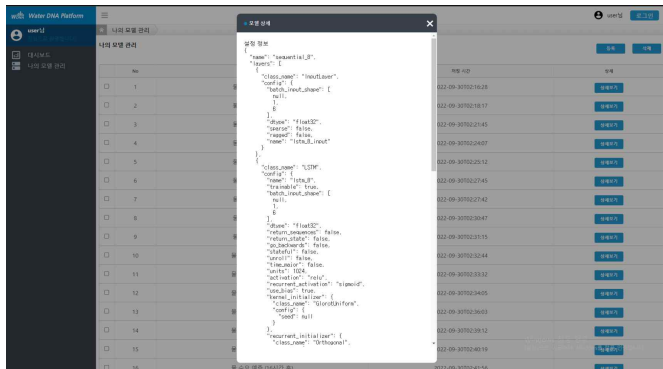


그림2 머신러닝 플랫폼의 사용자 모델 관리 화면

앞서 설명한 바와 같이 Node-RED는 머신러닝 파이프라인 구축에 기본적인 기능을 제공하면서도 사용자가 원하는 프로그램을 작성하여 서비스 특성에 맞는 상세 기능을 자유로이 구현할 수 있다. 그림3은 본 논문의 머신러닝 파이프라인에서 그림1과 같이 예측용 데이터 생성 및 예측 요청 그리고 예측 결과 저장을 수행하는 Node-RED의 구현 예시를 나타낸다.

그림3의 Node-RED 플로우에서는 주기적으로 복수의 NGSI-LD 엔티티 인스턴스로 저장된 유량 데이터를 NGSI-LD 인터페이스로 조회하여 이를 전처리하고 머신러닝 플랫폼의 모델 예측 API를 호출하여 응답받은 결과를 저장한다.

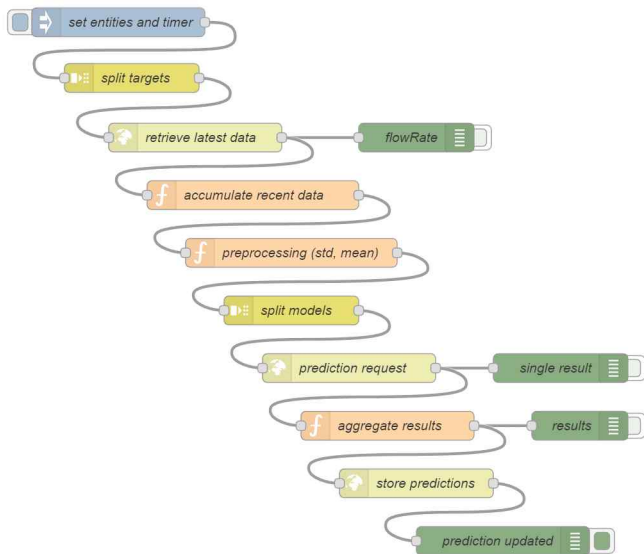


그림3 Node-RED 활용 머신러닝 추론 데이터 파이프라인 플로우

좀 더 상세하게는 해당 예측 모델은 현재를 기준으로 1시간 단위로 향후 24시간의 물수요량 예측 정확도 향상을 위해 최신 유량 데이터를 포함한 최근 복수의 시계열 데이터를 입력으로 사용한다. 또한 표준편차 및 평균값으로 전처리를 수행하고 이후에 예측 입력 데이터로 사용한다. 성능 향상을 위해 추가로 시간 단위별로 머신러닝 모델을 별도로 생성하여 실행 관리하기 위해 입력 데이터를 개별 모델에 추론 입력으로 요청하고 수신한 응답을 다시 병합하여 하나의 시계열 예측 결과 데이터를 생성한다. 이렇게 생성된 결과는 다시 NGSI-LD 플랫폼이 제공하는 REST API로 저장하여 대시보드 가시화 등에 사용한다.

III. 결론

본 논문에서는 NGSI-LD 표준 인터페이스 기반의 플랫폼에 적용할 수 있

는 머신러닝 파이프라인 구축 방법과 수자원 데이터 예측 서비스 구현 예시를 제시하였다. 제안하는 방법은 유량과 같은 단일 데이터뿐만 아니라 기상, 수질 등 다른 환경 데이터를 연계한 머신러닝 모델 기반 서비스에도 적용할 수 있다.

이후에는 관련 연구의 사례와 같이 학습용 데이터셋 구축 및 학습에 사용한 NGSI-LD 엔티티 데이터 속성과 같은 추론 데이터 생성을 NGSI-LD에서 지원하는 인터페이스를 설계하고 개발 검증해볼 수 있다. 이러한 기술이 개발되면 모델 개발자는 데이터 전처리 과정을 플랫폼이 제공하는 인터페이스를 통해 좀 더 자동화 할 수 있고 시스템 관리자는 서비스 요구사항이나 예측 성능 향상을 위한 파이프라인 구축에 중점을 두고 시스템 개발을 수행할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

본 결과물은 환경부의 재원으로 한국환경산업기술원의 지능형 도시수자원 관리사업의 지원을 받아 연구되었습니다.
(RE201903069)

참 고 문 헌

- [1] Jeong, S., et al, "Machine Learning dataset creation common service for real-time data platform," International conference on Internet 2021,
- [2] ETSI, "GS CIM 009 Context Information Management (CIM); NGSI-LD API," v1.6.1
- [3] Jeong, S, Kim S, Kim J. "City data hub: Implementation of standard-based smart city data platform for interoperability." Sensors 20.23 (2020): 7000.