

이종 데이터 연계를 통한 공공자전거 대여규모 예측기술기반의 연합디지털트윈 서비스

안세희^{1, 2}, 백명선^{2, 3}

¹국립 한밭대학교, ²한국전자통신연구원, ³한국과학기술대학원대학교

dkstpgml7@gmail.com, sabman@etri.re.kr

Federated Digital Twin service based on Prediction Technology of the public bicycle rental size through heterogeneous data linkage

An Se Hee^{1, 2}, Myung-Sun Baek^{2, 3}

¹Hanbat National University, ²ETRI, ³UST

요 약

디지털트윈은 현실 세계를 디지털 세계로 복제하고, 다양한 데이터를 연계하여 현실세계의 문제를 해결하거나 최적화할 수 있는 기술이다. 연합 디지털트윈은 여러 개의 디지털트윈을 유기적으로 연합하여 복잡한 시스템의 최적화를 수행할 수 있다. 본 논문에서는 이종 데이터의 유기적 연계를 통한 연합 디지털 트윈 기반의 서비스를 설계한다. 이종의 데이터 연계를 위해 기상 데이터와 공공자전거 대여수 데이터를 활용하였으며, 공공자전거 대여 규모를 예측하기 위한 딥러닝 모델을 설계하였다. 개발된 이종데이터 연계기반의 공공자전거 대여규모 예측 기술을 통해 공공자전거 배치 및 관리 서비스 등에 활용할 수 있다.

I. 서 론

디지털 트윈은 현실 세계의 다양한 객체를 디지털 세계로 복제하고, 이를 통해 다양한 모의실험을 수행하여, 현실 세계의 문제들을 해결하거나 최적화하기 위해 개발된 기술이다 [1]. 연합 디지털 트윈 기술을 활용하면, 목적에 따라 다수 개의 디지털 트윈을 선택하여 연합 및 운용하는 것이 가능하다. 그러나 디지털 트윈은 다양한 최신의 기술들이 복합되어 구현된 기술이므로, 오류에 매우 민감하며, 각 기술들의 동작 및 데이터의 유효성 여부의 검증이 매우 중요하다 [2]. 본 논문에서는 연합디지털트윈 서비스 제공을 위한 이종 데이터 연계를 통한 공공자전거의 대여 규모를 예측하는 기술을 제안한다. 사용되는 데이터는 서울시 방재 기상관측데이터와 공공자전거 이용현황 데이터이다. 이종데이터 연계 기반의 연합디지털트윈 서비스 제공을 위해 입력데이터의 유효성을 검증하고 개별 데이터의 이상징후를 파악하는 전처리 절차를 수행하였다. 또한 공공자전거 대여 규모 예측을 위해 DNN 기반 딥러닝 모델을 사용하였다. 이번 연구를 통해 도출된 결과물은 공공자전거의 대여나 반납 등 공공자전거 관리 서비스에 효과적으로 활용될 수 있다.

II. 서비스 개발을 위한 이종 데이터 연계 및 전처리

1. 데이터 분석

본 논문에서는 기상 데이터와 공공자전거 데이터 자료를 분석한다. 두 가지 데이터 모두 서울시의 데이터를 사용하였으며, 서울시 행정구 단위로 제공되는 데이터를 특정구만 추출하여 데이터분석에 활용하였다. 2017년 1월 1일 자부터 2017년 12월 31일까지의 데이터를 사용한다.

1.1 기상 데이터 분석

기상 데이터는 기상청에서 제공하는 서울시 방재 기상관측 데이터를 활용하였다. [그림1]은 서울특별시 마포구의 최저기온, 최고기온, 강수량, 평균 풍속이다. 시간 단위로 제공되는 데이터를 일별로 취합하여 데이터 분석을 진행했다.

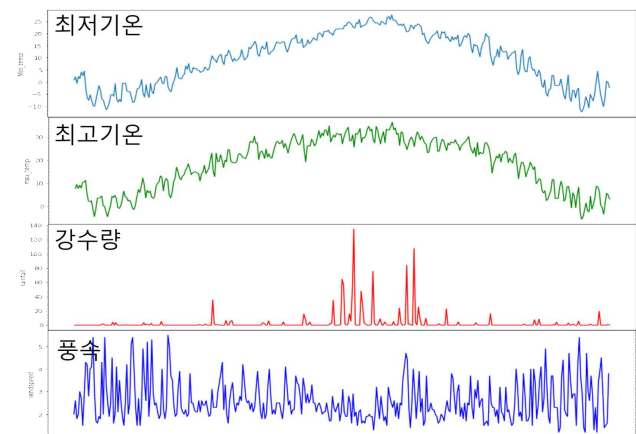


그림1. 2017.01.01.~ 2017.12.31. 서울특별시 마포구 일별 최저기온, 최고기온, 강수량, 풍속

1.2 자전거 데이터 분석

공공자전거 데이터는 '서울시 열린 데이터 광장'에서 제공하는 공공자전거 이용현황 데이터를 활용하였다. 대여소별 대여량을 일별로 취합하여 데이터 분석을 진행했다. [그림2]는 서울특별시 마포구의 대여량 분포도이다. 2017년 일평균 2,160대의 대여량을 가지지만, 일별로 최소 7대부터 최대 5,050대까지 다양한 분포를 확인할 수 있다.

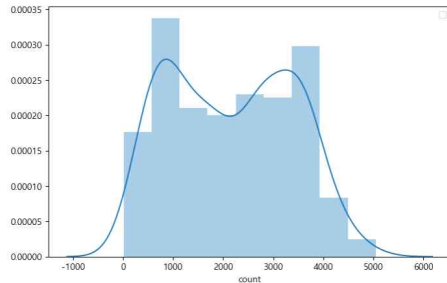


그림2. 공공자전거 대여량 분포도

1.3 상관관계 분석

[표1]은 공공자전거 대여량과 각각의 기상 요소와의 상관계수를 소수 셋째 자리까지 나타낸 것이다.

최저기온	최고기온	강수량	풍속
0.678	0.745	-0.202	-0.270

표1. 공공자전거 대여량과 기상 요소와의 상관관계

기상 요소에는 최저기온, 최고기온, 강수량, 풍속이 해당한다. 최저기온과 최고기온의 상관계수가 1에 가까운 값을 보이므로 공공자전거 대여량에 가장 많은 영향을 미치는 것을 확인할 수 있다.

이중 데이터 연계를 통해 딥러닝 모델에 사용되는 최종 데이터 세트는 [표2]와 같다. 공공자전거 대여량은 2017년 1월 1일부터 2017년 12월 31일까지 일별로 구성되었다. 본 연구에서는 다중 분류(Multi-class Classification)모델을 사용하므로 자전거 대여량 레이블을 추가하였고, 최종적으로 365개의 행과 6개의 열을 가지는 데이터 세트를 구성하였다.

	min_temp	max_temp	rainfall	windspeed	count	count_label
0	0.8	7.7	0	2	527	0
1	2.1	9.4	0	2.6	786	0
2	-0.6	7.6	0	1.8	854	0
..
362	0.3	5.6	0	1.5	1087	0
363	-0.1	4.9	1	1.6	557	0
364	-2.3	3.1	0	3.8	462	0

표2. 최종 데이터 세트

III. 예측모델 및 성능검증

본 논문에서는 공공자전거의 대여량을 클래스로 분류하기 위한 딥러닝 모델을 구축하였다. 3개의 레이블을 갖는 다중 분류(Multi-class Classification)모델을 설계하였다. 딥러닝 모델의 레이블은 일별 대여량을 3개의 구간으로 나누어, 각 클래스에 해당하는 학습 데이터의 수가 동일하도록 기준을 설정했다. 모델이 불균형한 데이터를 학습하여 오차가 커지는 것을 막기 위하여 대여량을 오름차순으로 정렬하여 각 122개, 121개, 122개로 나누어 '적음(0)', '보통(1)', '많음(2)'으로 범주화하였다. 본 논문에서는 심층 신경망(Deep Neural Networks, DNN) 기반 모델을 사용한다. DNN은 구조적으로 간단하며 시뮬레이션에 비해 연산량을 크게 줄일 수 있다는 장점이 있다. 예측을 위한 입력 데이터는 입력층(Input Layer)을 통해 모델에 입력되고, 예측하고자 하는 목표 데이터는 출력층(Output Layer)로 계산된다. 입력층과 출력층 사이에 존재하는 1개 이상의 층은 은닉층(Hidden Layer)이라고 하며 각 층은 다수의 노드(Node)로 구성되어 있으며, 각 노드 간의 연결은 인공신경망의 신경 세포에 해당하는 퍼셉트론들의 다층 연결에 의해 이루어진다 [3]. 출력층의 활성화 함수는 Softmax를 활용한다. 전체 데이터 세트에서 10%를 랜덤으로 샘플링하여

Test 데이터로 사용하였고, 90%는 Train 데이터로 사용하였다. [그림3]은 기상 요소에 따른 자전거 대여량을 분류하기 위해 개발한 딥러닝 모델의 구조이다. 2개의 Hidden Layer는 각각 128, 64개의 노드를 가진다. Output Layer를 제외한 모든 Layer에는 활성화 함수로 ReLU를 사용한다. Output Layer에는 3개의 클래스로 분류하기 위해 Softmax 함수를 사용한다. 또한, Optimizer는 Adam을 사용하였으며 Batch Size는 16, Epochs는 100으로 설정하여 모델을 학습했다. 설계한 딥러닝 모델의 학습 결과 정확도는 약 84%이며, 손실함수로는 교차 엔트로피를 사용했으며 값은 0.479이다. [그림4]는 설계 모델의 학습 결과를 나타낸다.

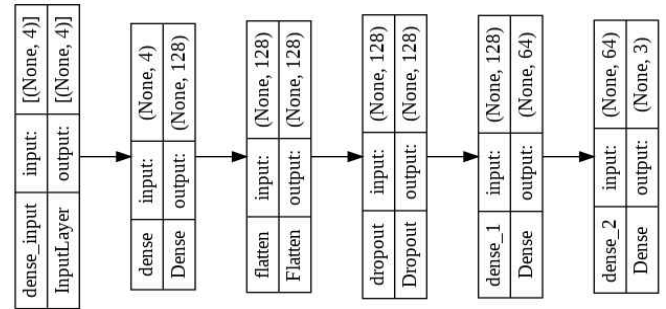


그림3. 설계한 딥러닝 모델 구조

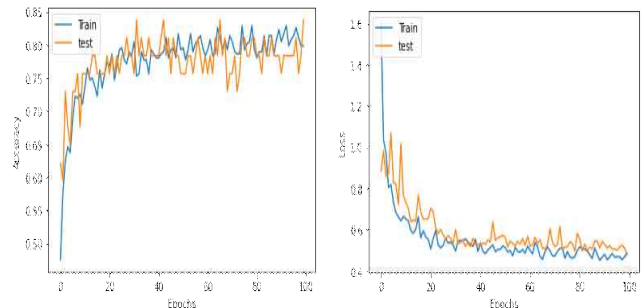


그림4. 설계 모델 결과 (왼쪽부터 정확도, 손실)

ACKNOWLEDGMENT

이 논문은 2023 년도 정부 (과학기술정보통신부) 의 재원으로 정보통신기획평가원의 100% 지원을 받아 수행된 연구임(No. 2022-0-00545, (2 세 부) 지능형 디지털 트윈 연합 객체 구성 및 데이터 프로세싱 기술 개발)

참 고 문 헌

- [1] D. -Y. Jeong et al., "Digital Twin: Technology Evolution Stages and Implementation Layers With Technology Elements," *IEEE Access*, vol. 10, pp. 52609–52620, May, 2022.
- [2] 백명선, 정득영, and 이용태. "연합 디지털 트윈을 위한 데이터 유효성 검증 방법 연구." 한국통신학회 학술대회논문집 pp. 421–422, 2022.
- [3] 김경곤, 박찬수, 김우영, 전지연, 전미연, 배충식. "진동신호를 이용한 CNG-디젤 이중연료 엔진 연소압력 예측 DNN 모델에서 천연가스 대체율 및 디젤 분사시기가 예측 정확도에 미치는 영향." 한국자동차공학 회논문집, 29(10), pp 909–919, 2021.