

## Prophet 모델을 사용한 기상데이터 예측

김준석, 김성희, 윤주상, 강재환\*

동의대학교 산업 ICT 공학과, \*동의대학교 인공지능그랜드 ICT 연구센터

[junsuk.kim; sh.kim; jsyoun; \*jh.kang]@deu.ac.kr

## Meteorological Data Forecasting using Facebook Prophet Library

Junsuk Kim, Sung-Hee Kim, Joosang Youn, Jae-Hwan Kang\*

Dong-Eui Univ., \* Dong-Eui Univ. Grand ICT Research Center

### 요 약

본 논문은 시계열 데이터 예측에서의 훌륭한 성능으로 주목받고 있는 Facebook 의 Prophet 모델을 사용하여 기상 데이터 (온도, 습도, 풍속)를 학습시키고 시계열 데이터 변화양상을 예측하는 연구를 진행하였다. Prophet 의 예측 성능은 다양한 모델 성능 평가 기법 (Index of Agreement; IOA, Mean Error, Normalized Root Mean Squared Error; NRMSE)을 사용하여 수치화하였다. 그 결과, Prophet 모델은 온도 데이터 예측에는 탁월한 성능을 보여주는 반면 습도와 풍속 데이터 예측에는 좋은 성능을 보여주지 못하였다. 이 연구결과는 향후 인공지능 알고리즘 또는 다양한 시계열 예측 기술을 사용한 기상 데이터 예측 연구에 기초 자료가 될 것으로 생각된다.

### I. 서 론

최근 회귀분석, 심층 신경망 등 다양한 기술을 활용한 시계열 예측 연구가 활발히 진행되고 있다 [1,2]. 이 연구에서는 시계열 예측 분석 대회에서 특별한 전략을 사용하지 않고 데이터만을 활용하여 훌륭한 예측 성능을 보여주었던 Facebook 의 시계열 분석 라이브러리 Prophet 을 사용하여 기상 변화 (온도, 습도, 풍속) 를 예측해 보았다 [3]. 이 모델을 통해 예측된 기상데이터 값을 실제 데이터와 비교하여 예측 정확도를 평가하였다.

### II. 본론

본 논문에서는 2018년 Taylor S.J. 와 Letham B. 이 발표한 Facebook 의 시계열 분석 라이브러리 Prophet 을 사용하여 기상 인자들의 시계열 예측 모델을 구현하였다 [3].

분석에 사용된 기상데이터는 기상청 기상자료 개방포털 (<https://data.kma.go.kr>) 에서 제공하는 온도, 습도, 풍속 데이터를 사용하였다. 특히, 강원도 고성군 지역의 3 년간 (2016.01.01. ~ 2018.12.31.) 의 관측치를 트레이닝에 사용하고 그 이후의 4 개월간 (2019.01.01. ~ 2019.04.30.) 의 관측치를 성능평가를 위한 테스트에 사용하였다. 입력 데이터로는 각 시간별로 제공되는 온도, 습도, 풍속 관측치를 사용하였으며, 각각 기상데이터 종류에 따라 시계열 예측 모델 입력으로 트레이닝 데이터와 테스트 데이터를 가진 두 개의 데이터 시퀀스가 사용되었다.

Prophet 모델의 예측 성능을 정량적으로 평가하기 위해, 즉, 모델의 예측 결과를 실제 관측 값과 비교하기 위하여 각각의 기상데이터 조건에 대하여 IOA (Index of Agreement), ME (bias), NRMSE (Normalized Root Mean Squared Error) 를 계산하였다. 이 평가 방법들은 인공지능 알고리즘을 활용하여 시계열 예측력을 평가하는 용도로 많은 연구에서 사용되고 있다 [2,4].

### III. 결론

온도 데이터 예측의 경우, IOA는 0.78, ME는 0.50,

NRMSE는 14.35% 의 결과를 얻었다. 또한, 습도 데이터 예측의 경우 IOA는 0.38, ME는 -5.02, NRMSE는 35.95% 의 결과를, 풍속 데이터 예측의 경우 IOA는 0.36, ME는 0.12, NRMSE는 35.41% 의 결과를 얻었다. 결과적으로 세 종류의 입력데이터 중에서 Prophet 모델은 온도 데이터의 예측 능력이 습도, 풍속 데이터 예측력 보다 우수한 것으로 나타났다. 추가로 모델의 예측 결과와 실제 관측 값 간의 상관관계를 분석하였으며, 그 결과 온도의 경우  $r = 0.64$ ,  $p < .01$ , 습도의 경우  $r = 0.29$ ,  $p < .01$  로서 통계적으로 유의미한 유사성을 확인하였으며, 풍속의 경우  $r = 0.12$ ,  $p = .19$  로 통계적으로 유사성을 찾지 못했다.

### ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 Grand ICT 연구지원센터 지원사업의 연구결과로 수행되었음 (IITP-2020-0-01791).

### 참 고 문 헌

- [1] Yu, P-S., Chen, S-T., Chang, I-F. (2006) Support vector regression for real-time flood stage forecasting, Journal of Hydrology, vol. 328, no. 3-4, pp. 704-716.
- [2] Kim, D., Kim, J., Kwak, J., Necesito, I. V., Kim, J., Kim, H. S. (2020) Development of Water Level Prediction Models Using Deep Neural Network in Mountain Wetlands, Journal of Wetlands Research, vol. 22, no. 2, pp. 106-112.
- [3] Taylor, S. J., Letham, B. (2018) Forecasting at scale, The American Statistician, vol. 72, no. 1, pp. 37-45.
- [4] Cho, K., Lee, B-Y., Kwon, M., Kim, S. (2019) Air quality prediction using a deep neural network model, Journal of Korean Society for Atmospheric Environment, vol. 35, no. 2, pp. 214-225.