EGU24-7091, updated on 23 Jul 2024
https://doi.org/10.5194/egusphere-egu24-7091
EGU General Assembly 2024

The Development of precipitation model modifed with ECMWF IFS and XGBoost and its performance verification

Eunju Cho, Yeon-Hee Kim, Seungbum Kim, and Young Cheol Kwon
Korea Meteorological Administration, Numerical Modeling Center, Korea, Republic of (eunju424@korea.kr)

This study was conducted to develop a modified precipitation model for its amount and existence by combining machine learning method, Extreme Gradient Boosting(XGBoost), with ECMWF IFS(Integrated forecasting system) and, finally, estimate the related performance.

According to the analysis of regional precipitation characteristic, prior to its development, the ratio of precipitation existence was various on a basis of a forecast's district and its season. These different patterns on each district makes it necessary to develop the regional and seasonal model respectively.

And, the first attempt at the machine learning showed the importance of each feature as input-variables, as a result of which cloud physics-related features, for example large-area precipitation, total precipitation, visibility and what not, proved so significant. However, the insufficient amount of these feature's data seemed to result in overfitting. And therefore, the feasible features, except for cloud physics-related things, of IFS data were used. In addition, auxiliary features and their gradient for every lead-time were calculated and added: relative vorticity, divergence, equivalent potential temperature, main 6 patterns for Korean summer and so on. The number of features amounted to around 144 with which for the 9-year training set, 2013~2021, based learning to be conducted regionally, followed by using validation-set of 2022.

As a result of validation for precipitation existence and its amount up to 135 hours ahead on the 10 regions at 00UTC in summer of 2022, Critical Success Index(CSI) was more improved by 10.3% than before. Accuracy(ACC) for each lead-time rose by 6% and its fluctuation also decreased. And the correction by this machine learning alleviated the overfitting trend of precipitation forecast amount produced by the original model, and improved correlation and linearity between observation and forecast. In particular, while the machine learning prevailed over the original model up to 100 hours ahead, from then on, both of them showed similar performance or that of the former was downward slightly. If the above-mentioned cloud physics features are used to further sharpen machine learning technique, its performance should be enhanced more and more.