Tropical Cyclone Genesis Detection using COMS Data with Machine Learning Approaches

Seongmun Sim (1), Jungho Im (1)

[1] Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulju-gun, Ulsan-si, 44919, Korea

Email: smsim@unist.ac.kr; ersgis@unist.ac.kr

Abstract: Tropical cyclones (TCs) are one of the hazardous atmospheric phenomena, resulting enormous socio-economic losses. The western North Pacific (WNP) region is the most vulnerable area to TCs, especially, equatorial countries (e.g., Philippine, Indonesia) only has little time to prepare for TC approaching from the TC prediction/warning. Many institutes are developing models to detect TC as early as possible, not only numerical weather predictions (NWPs) based, but also satellite-based models are studied actively. In this study, TC early detection models were suggested using the Communication, Ocean, and Meteorological Satellite (COMS) Meteorological Imager (MI) products with machine learning (ML) approaches such as random forest (RF), logistic regression (LR), and support vector machine (SVM). The Joint Typhoon Warning Center (JTWC), one of the biggest TC warning institute, serves archive of past TC tracks, but the most of these are recorded from the TC detection timing (maximum sustained wind speed > 25 knot). Thus, tropical cluster center (TCC) method from the Hennon et al. 2011 were adopted to get the disturbance (early stage of TC) track information. Totally 100 developing and 614 non-developing tracks were extracted from 2011 to 2017. On every disturbance center, 8.1-degree by 8.1-degree box area images were cut and various spatial informatic variables were extracted. With those variables, auxiliary variables were extracted to considering diurnal and seasonal characteristics. The extracted variable set was separated in 5-bunch conducting 5-fold cross validation method. Four evaluating indices such as probability of detection (POD), probability of false detection (POFD), false alarm (FAR), and Heidke skill score (HSS), were used to assess model performance. In 5-fold cross validation result, all ML models exceeding 80% of POD, but RF serves slightly higher POFD and FAR than the LR and SVM model. In best lead-time record, LR and SVM is identifying developing TCs 47 hours before the TC detection, while RF recognizing it 10 hours later. In conclusion, ML based TC genesis detection models show high detection rate and low false alarm rate. Among the ML models, LR and SVM which are regressing-based model has higher suitability than rule-based model, RF.

Keywords: Tropical cyclone, cyclone genesis, western North Pacific, COMS, machine learning,