Application of ensemble learning noise classifier to automatically determined earthquake catalog after the 2011 Tohoku Earthquake

*Koji Tamaribuchi[1]

1. Meteorological Research Institute

A microseismic catalog is one of the big data that is important for earthquake research and disaster prevention. However, because the 2011 Mw 9.0 Tohoku earthquake caused an enormous number of aftershocks and induced earthquakes, it was impossible to process all the detectable earthquakes. Thus, the microseismic activity remains hidden. Tamaribuchi and Nakagawa [2020] applied the automatic hypocenter determination method to the continuous seismic waveform data for March 2011. The automatically determined earthquake catalog was more than twice in the JMA unified catalog.

On the other hand, it is necessary to remove the noise efficiently because the automatically determined catalog contains several percent of false detections due to noise. Therefore, Tamaribuchi et al. [2020, SSJ Fall Meeting] applied ensemble learning, a kind of supervised machine learning, to the automatically determined catalog and confirmed that we could remove about 80% of false detections. In this study, we apply the supervised ensemble learning to the aftershocks of the 2011 Tohoku earthquake and create a denoised earthquake catalog.

Tamaribuchi et al. [2020, SSJ Fall Meeting] created a training dataset of earthquakes and noise by visual inspection based on the data recorded in the automatically determined catalog from January to February 2020. We then created a noise classifier based on the supervised ensemble learning (Adaboost). The classifier uses hypocenter information such as magnitude, location, and the number of P- and S-phases at 20 stations near the epicenter for learning. In this study, we applied the classifier trained on the data from January to February 2020 to the March 2011 data. There were no living stations near the epicenter due to the lack of an observation network in some cases. Therefore, we used only alive stations when selecting the 20 stations near the epicenter.

For March 2011, the automatic process determined 122353 events all over Japan. As a result of applying the classifier, 119037 (97.3% of the total) and 3316 (2.7%) were classified as earthquakes and noises, respectively. The time taken for classification was 31 minutes (0.02 seconds per event). Most of the noises were in the aftershock areas and Ryukyu islands, and we can classify the noise appropriately in a realistic processing time. On the other hand, compared with the JMA unified catalog, 124 misclassified cases are in aftershock areas. We will consider improving the generalization performance, such as including data from 2011 in the training data.

Keywords: Earthquake catalog, Automatic hypocenter determination, The 2011 Tohoku Earthquake, Machine learning

