Geophysical Research Abstracts
Vol. 12, EGU2010-6269-2, 2010
EGU General Assembly 2010

Supervised Classification Methods for Seismic Phase Identification

Jeff Schneider (1), Jeff Given (2), Ronan Le Bras (2), and Misrak Fisseha (2)

(1) Carnegie Mellon University School of Computer Science Heinz College, (2) CTBTO, IDC, Vienna, Austria (ronan.lebras@ctbto.org)

The Comprehensive Nuclear Test Ban Treaty Organization (CTBTO) is tasked with monitoring compliance with the CTBT. The organization is installing the International Monitoring System (IMS), a global network of seismic, hydroacoustic, infrasound, and radionuclide sensor stations. The International Data Centre (IDC) receives the data from seismic stations either in real time or on request. These data are first processed on a station per station basis. This initial step yields discrete detections which are then assembled on a network basis (with the addition of hydroacoustic and infrasound data) to produce automatic and analyst reviewed bulletins containing seismic, hydroacoustic, and infrasound detections.

The initial station processing step includes the identification of seismic and acoustic phases which are given a label. Subsequent network processing relies on this preliminary labeling, and as a consequence, the accuracy and reliability of automatic and reviewed bulletins also depend on this initial step. A very large ground truth database containing massive amounts of detections with analyst-reviewed labels is available to improve on the current operational system using machine learning methods. An initial study using a limited amount of data was conducted during the ISS09 project of the CTBTO. Several classification methods were tested: decision tree with bagging; logistic regression; neural networks trained with back-propagation; Bayesian networks as generative class models; naive Bayse classification; support vector machines. The initial assessment was that the phase identification process could be improved by at least 13% over the current operational system and that the method obtaining the best results was the decision tree with bagging. We present the results of a study using a much larger learning dataset and preliminary implementation results.