Geophysical Research Abstracts
Vol. 12, EGU2010-4887, 2010
EGU General Assembly 2010

Improving the Accuracy of Automatic Detections at Seismic Stations via Machine Learning

Carsten Riggelsen (1) and Matthias Ohrnberger (2)

(1) Uni. Potsdam, Inst. of Geosciences, Potsdam, Germany (riggelsen@geo.uni-potsdam.de), (2) Uni. Potsdam, Inst. of Geosciences, Potsdam, Germany (mao@geo.uni-potsdam.de)

We present a Machine Learning approach aiming for improving the accuracy of automatic detections of noise and signal at 3-component seismic stations. Using supervised learning in conjunction with the multivariate framework of Dynamic Bayesian Networks (DBNs) we make use of historical data obtained from the LEB bulletin to train a classifier to capture the intrinsic characteristics of signal and noise patterns appearing in seismic data streams. On a per station basis this yields generative statistical models that essentially summarize and generalize the information implicitly contained in the LEB allowing for classifying future an previously unseen seismic data of the same kind. Also, the system provides a numerical value reflecting the classification confidence potentially aiding the analyst is correcting or identifying events that are non-typical. The system has the potential for being implemented in real time: both feature computation/extraction as well as classification work on data segments/windows and seismic patterns of varying length, e.g., 12 sec.

Various features are considered including spectral features, polarization information and statistical moments and moment ratios. All features are derived from a time-frequency-(amplitude) decomposition of the raw waveform data for each component, taking the 6 frequency bands currently in use at IDC into account. These different feature sets give rise to different DBN structures (model-feature scenarios) that probabilistically relate the features to each other depending on empirical observations and physical knowledge available. 1 week of waveform data is considered for training both the signal and noise classes. The performance of the classifier is measured on a separate test set from the same week of data but also on a 1-month data set, where 4 weeks of data is distributed over a one year period. In the system evaluation both a static approach as well as a sliding-window approach is taken. Binary classification accuracy, sensitivity and specificity is measured as well as a comparison to the SEL3 and LEB bulletins. Our results suggest that a model-feature scenario with spectral features performs well, with a relatively high accuracy. Moreover, when using the DBN trained on 1 week data to classify noise throughout the year, our analysis suggests that there is a seasonal noise variation/dependence (the degree depending on the station in question). Spectral analysis also confirms this observation. In effect this means that an adaptive approach to training the generative DBN for seismic noise needs to be employed in order to correctly identify noise at any given time.