Geophysical Research Abstracts
Vol. 13, EGU2011-5710, 2011
EGU General Assembly 2011

Improving Automatic Detections at IMS Seismic Stations via Machine Learning: Current Status and Future Directions

Carsten Riggelsen and Matthias Ohrnberger
Uni. Potsdam, Inst. of Earth and Environmental Science, Potsdam, Germany (riggelsen@geo.uni-potsdam.de)

We investigate and report on our progress for improving the automatic seismic detection system at the IDC. The IMS seismic network produces an abundance of time-series data, and issues such as uncertainty, noise, sensor interdependence, etc. pose great challenges for on-line processing and unbiased near real-time analysis. To this end, methods borrowed from the field of machine learning and data mining provide elegant solutions. By adhering to the multivariate statistical framework of Graphical Models combined with insights from the field of automatic speech recognition, real-time signal classification becomes viable. Using supervised learning in conjunction with the multivariate framework of Dynamic Bayesian Networks we make use of historical data obtained from the LEB bulletin to train the 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. About 100 waveform snippets of short duration (4-12 sec) are extracted from 1 week of waveform data for training both the signal and noise classes. As signal class we consider time windows enclosing the manually picked onsets from the LEB-bulletin whereas for the noise class training examples we take any possible short time window which doesn't contain any obvious transient signal contribution. As input to the classifier we use the spectral features of the raw waveform data, taking into account 6 frequency bands resembling the currently used ones at IDC. In the system evaluation we measure (binary) classification accuracy, sensitivity and specificity and compare our results with the SEL3 and LEB bulletins. Considering the limited comparability between our purely automatic station-level detector and the combined automatic network associator with subsequent manual inspection approach at the IDC we find that the classifier performs well. When testing against unseen data in time we can confirm the well-known seasonal dependency of noise characteristics. Given that the noise model is initially trained on data from a selected time period (1 week) the noise variance is small compared to the variance observed for the spectral amplitudes of the signal class. Thus the noise model needs to be adaptively adjusted over time in order to capture the station noise character at the particular period of the year (or day of week in case of anthropogenic influences of the station environment). Our (yet unconfirmed) expectation is that the adaptive updating of the noise model outperforms the classifier performance when comparing to a fixed model trained on many noise examples picked from distinct seasons although this should ideally also allow capturing the overall variability of noise characteristics.