Integrating Machine Learning into Real Time Processing of Earthquakes

Andrew Reynen, Kit Chambers, Sepideh Karimi, Dario Baturan
Nanometrics

Summary

The accuracy of real time event locations are often negatively affected by the misclassification of the phase type and/or introduction of noise picks. Here a machine learning technique is used to discriminate between different types of phase arrivals and noise on results of a real time earthquake processing system. Events with reclassified arrivals are then reported in near real time. Inputs to the regression contain the spectral content as well as downsampled characteristic functions (CFs). Preliminary results show a significant reduction in the distance between automatic solutions and their corresponding manually reviewed counterparts.

Introduction

Seismic event catalogs generated in real time are important inputs into risk management and public safety strategies. However, due to the vast amount of seismic data being collected, the construction of such catalogs can become time and labor intensive. Hence an automated process is sought to reduce the workload involved in catalog production. Conventionally events are detected by applying an autopicker to seismic data streams (hereafter referred to as autopicking) and grouping picks into sets corresponding to potential events (pick association). In autopicking picks are generated when a threshold is surpassed on a characteristic function (CF) for a given channel (ex. STA/LTA (Allen 1978), Kurtosis (Saragiotis et al. 2002), etc.). In the scenario of a small magnitude event it is likely only a few channels have visible arrivals and the threshold must be set low to increase the chances of detecting these signals. However,

reducing the threshold also increases the number of false positives produced by the detection system and this mixture of true and false picks in any given solution can introduce significant errors into the corresponding event location. In order to maintain the number of true positives (earthquakes) while reducing the effect of false positives (picks on noise) we utilize a classifier leveraging machine learning techniques to better distinguish phase arrivals than conventional CFs. The technique assigns a likelihood to a windowed waveform as to which among the potential phases the coda belongs. Windows which are more likely to be noise are down weighted to reduce their effect on the final automatic solution.

Method

The procedure begins with the construction of a training dataset comprised of previously recorded automatic solutions and their associated manual solution. Windows are selected about phase arrival times from the automated solutions, if the time aligns well with a predicted arrival time from the manual solution it is assigned to the P or S arrival class; otherwise it is added to the noise class. Additional observations are added to the noise class by randomly selecting time windows prior to the predicted P-arrival and after the predicted S-arrival. For each of the time window we compute features using spectral content and various CFs, which are then downsampled (Fig. 1).

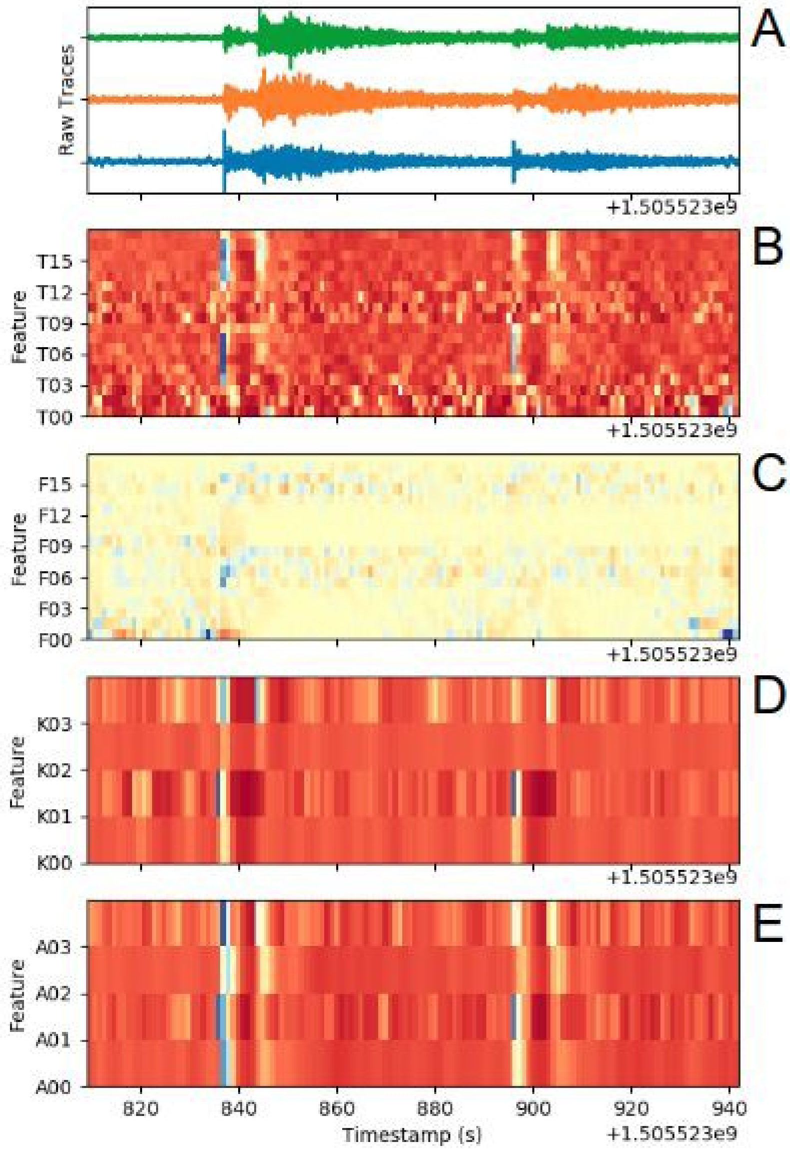

Figure 1. Conversion of raw trace data (A) into example features. Four groups of features are shown: Time normalized spectrogram (B), band normalized spectrogram (C), kurtosis (D), arias intensity (E).

An SVM (Cortes & Vapnik 1995) is used to find the optimal boundary separating the classes (P, S and Noise), using the features from the waveforms. Once the model has been trained, when new automatic solutions are obtained the features are computed for the corresponding waveforms and the arrivals are classified. This results in a phase likelihood for the 3 categories at a given time and station (Fig 2). In decreasing noise likelihood, arrivals are removed if the noise likelihood exceeds 50%. If the number of observations is less than 6, the noisy arrival is instead down weighted to 0.25 (as opposed to 1.00).

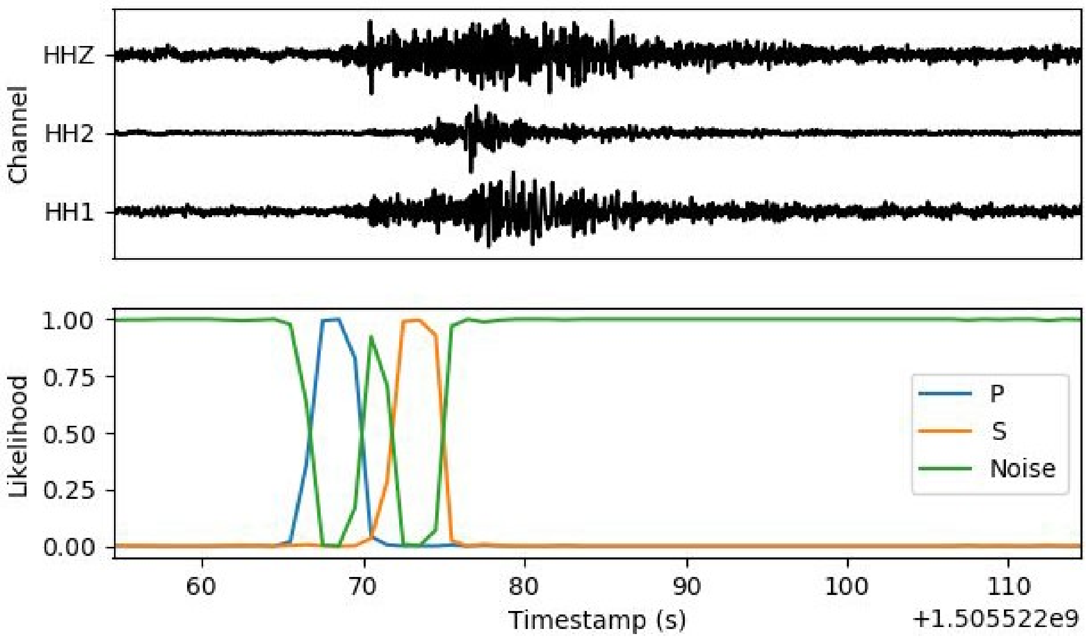

Figure 2. Likelihood of class (P arrival, S arrival, Noise) over time for one station. Strong peaks in the likelihood for real phase arrivals can be see at the correct times even when the signal to noise ratio is low.

Results

A test sample of 200 events from a regional seismic array were relocated using the new method (Fig 3.). Event locations show significantly less scatter, and are on average 50% closer to their corresponding manual location.

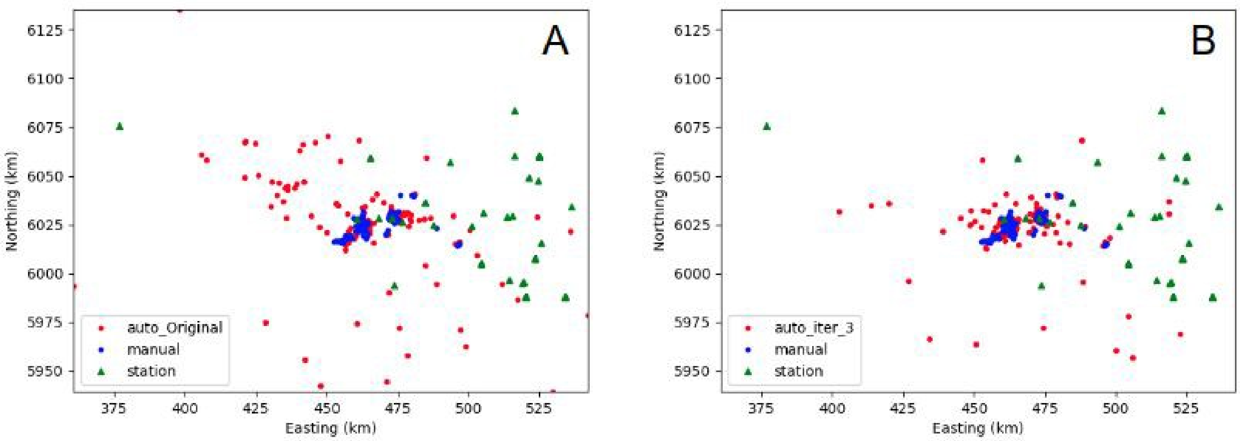

Figure 3. Comparison of automated event locations (red) along with the associated manual locations (blue). Improvement in location can be seen after applying the method (B) as compared to the original automatic locations (A).

Impact

A machine learning technique has been shown to provide good discrimination between different phase types and noise. The method allows for a high dimensional feature space to be utilized objectively. Preliminary results show a significant improvement in the automatic location after reclassifying arrivals, reducing the distance between the automatic and manually reviewed location by 50%.

References

Allen, R.V., 1978. Automatic earthquake recognition and timing from single traces, Bull. seism. Soc. Am., 68(5), 1521–1532.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

Saragiotis, C.D., Hadjileontiadis, L.J. & Panas, S.M., 2002. PAI-S/K: a robust automatic seismic P phase arrival identification scheme, IEEE Trans. Geosci. Remote Sens., 40(6), 1395–1404.