Results of crowdsourcing for first arrival picking of earthquakes: lessons learned

Serafim Grubas, Mirko van der Baan
University of Alberta

Summary

In August 2022, a two-day datathon was organized with the challenge of detecting and picking earthquakes, which are crucial steps in understanding the characteristics of earthquakes such as origin time, location, magnitude, and source mechanism. We synthesized full waveform data with surface acquisition geometry and generated 25 datasets. Challengers were asked to submit the first arrival times of P-waves for all principal detected earthquakes (maximum 5) for a test dataset. Five teams took part in the challenge, and contestants were encouraged to use machine-learning approaches to beat traditional methods. Three teams were able to detect the right number of earthquakes, and only two of them were able to pick the first arrivals close to the true solution. From the reports of the teams, we found that the successful strategy was to split the challenge into the detection and picking aspects and solve them separately by combining traditional and machine-learning approaches. However, we also discovered that the assessment metric we used caused undesirable behavior in the scores and provided scope for the identification of the true number of earthquakes using the trial-and-error approach.

Introduction

Detection and traveltime picking of earthquake first arrivals are important steps towards forecasting, but both often require tedious manual inspection to verify and validate automatic results. Improved automatization could speed up the processing of large data volumes significantly, with machine learning methods showing much promise. We proposed this problem as a challenge in an open Datathon competition with the intention of using crowdsourcing to generate fresh views on these long-standing problems. We encouraged contestants to beat existing and well-established traditional methods with modern machine learning approaches.

One of the most popular traditional methods is Short-Term Average and Long-Term Average (STA/LTA) (Allen, 1978) which is used both for detection and picking by comparing energies in two temporal windows of different lengths; however, parameter setting is often challenging, producing often either high false alarm rates (false positives) that need to be manually corrected, or high numbers of missed events (false negatives), that lead to missing information (Vaezi and Van der Baan, 2015).

Provided datasets

The data were synthesized by calculating the full waveforms (Dietrich, 1988). Seismic events had the following parameters: frequency, location, origin time, and mechanism. The corresponding first arrivals of P waves were approximated by Neural Eikonal Solver (Grubas, 2023). We considered a 2D model in XZ-plane with 15 layers (see Figure 1.A). The acquisition geometry is a line of 127 evenly spaced receivers on the flat surface. We simulated ~250 events with random parameters which made up 25 datasets. Each dataset contained 127 receivers, 3 components, and 30000 time samples. We assigned from 0 to 5 earthquakes to each dataset and blended it with 7% white gaussian noise and 500 low magnitude events. The true first arrivals are calculated for each receiver.

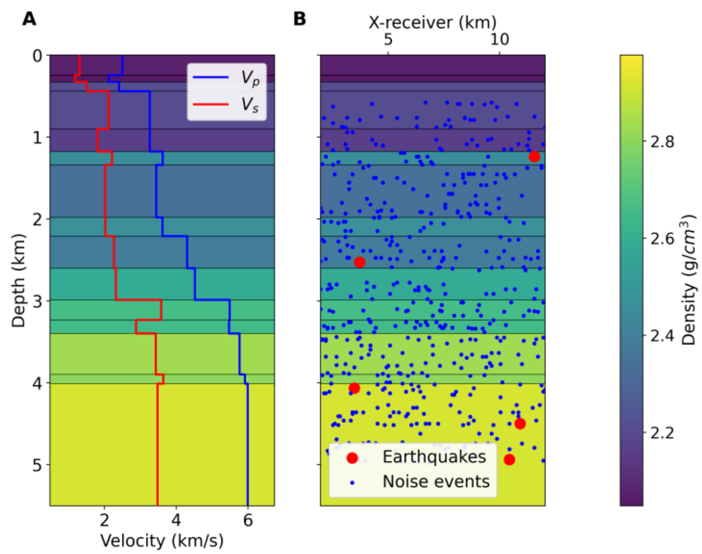

*Figure 1. Velocity model (**A**) and sample locations of earthquakes (**B**).*

Competition

In the challenge, contestants had to pick the first arrivals for all principal earthquakes they could detect in a dataset (min 0, max 5). 24 datasets were provided for training, and one dataset was given without first arrivals for final scoring only. The scoring had public and private leaderboards, which consisted of different parts of the test dataset. The public score was visible during the competition, the private was used for the final ranking (Figure 2). After the competition ended, we analyzed the results of the top 5 teams, who all kindly disclosed their solution strategies.

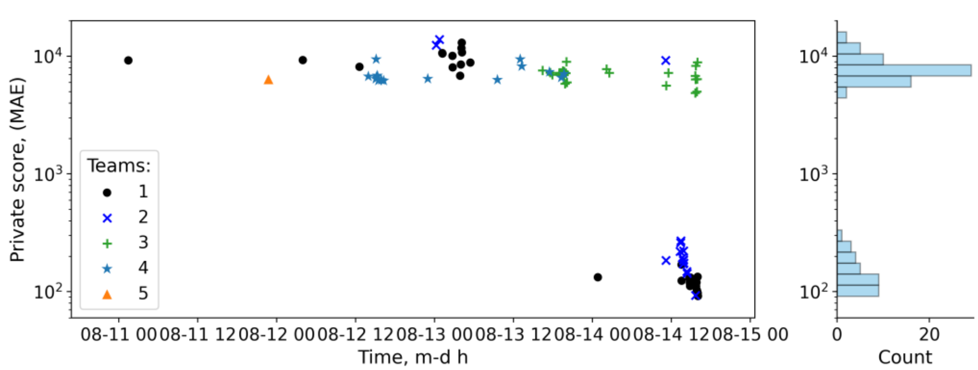

Figure 2. Submission history of the private leaderboard over time by different teams. The right panel illustrates the distribution of score values.

The assessment metric was a relative mean-absolute error. The solution consisted of arrivals times for 5 earthquakes for 127 receivers. If the number of earthquakes is less than 5, `$-1$` is used. The first arrivals must be sorted by the time they occur. It was discovered that this metric caused undesirable behavior in the scores and provided scope for identification of the true number of earthquakes using trial-and-error approach (see abrupt score transition from 10^4 to 10^2 in Figure 2). Only the top 3 teams could detect the right number of earthquakes and only the top 2 could pick the first arrivals close to the true solution.

The solutions on test dataset with 4 earthquakes are shown in Figure 3. *Team 3* (thin green lines) produced random numbers near the first three earthquakes. *Team 5* (orange) produced 5 earthquakes with the first arrivals far from true. *Team 4* (pale blue) also detected 5 earthquakes, but their first arrivals are close to the truth. *Team 2* (blue lines) detected 4 earthquakes correctly and picked the first arrivals very well making their solution outstanding. *Team 1* (top-1, black lines) detected 4 events correctly, but the first arrivals are simple straight lines.

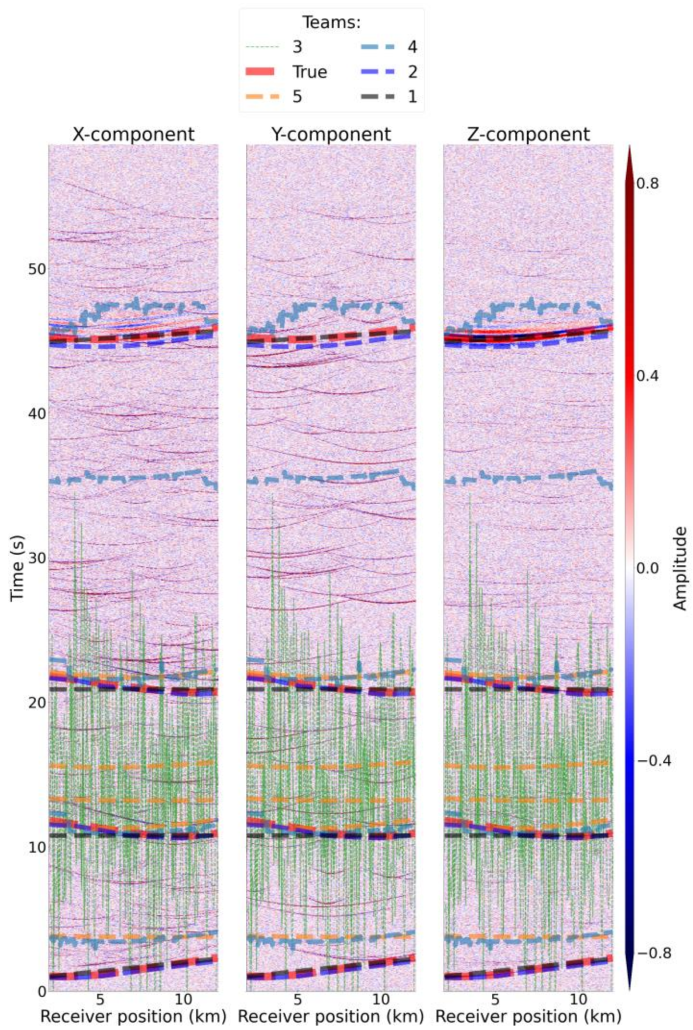

Figure 3. The test dataset and solutions. Teams are indicated by colored lines. Team 1 is indicated by black color; Team 2 is blue; Team 3 is green; Team 4 is pale blue; Team 5 is orange; true is red.

Conclusions - what did we learn?

The challenge was two days open for submissions, plus seven days prior for an opportunity to familiarize teams with datasets and the exact problem. Five teams took part in the challenge, three teams were fully involved till the end of the competition, two of which obtained meaningful results. Both of these teams relied on visual inspection of the predictions obtained either by machine learning or standard methods. The most successful teams split the problem into detection-picking aspects, using different algorithms for each step, contrary to the more usual approach in seismology to use the same algorithm for both detection and picking (most typically STA/LTA). From the given reports, we found the solution of *Team 2* outstanding (blue lines, Figure 3) because they picked the first arrivals with correct curvature with respect to receiver location. They combined machine learning (Logistic Regression) and traditional (STA/LTA) approaches.

In hindsight, it is likely that teams were hindered by the non-smooth behavior of the metric for judging both the number of events and the event arrival times. The problem was not formulated properly, and the way of assessment was not fully applicable due to the format of the first arrivals for five earthquakes. The metric was extremely sensitive to the order of earthquakes and did not accurately represent the real quality of the solution. To avoid this issue, earthquake arrivals could be compared to the closest true arrivals so that the order does not matter. Unfortunately, custom metrics are not supported by the platform we used. Tests of the order-insensitive metric showed that it could significantly improve the smoothness of the metric showing real proximity to the true solution.

Acknowledgements

We would like to thank the sponsors of the datathon, AI4Society and the Microseismic Industry Consortium, for their financial support, and all participants for their hard work and willingness to share their methodologies.

References

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. Bulletin of the seismological society of America, 68(5), 1521–1532.

Dietrich, M. (1988). Modeling of marine seismic profiles in the tx and τ -p domains. Geophysics, 53(4), 453–465.

Grubas, S., Duchkov, A., & Loginov, G. (2023). Neural Eikonal solver: Improving accuracy of physics-informed neural networks for solving eikonal equation in case of caustics. Journal of Computational Physics, 474, 111789.

Vaezi Y. and Van der Baan M. (2015) Comparison of the STA/LTA and power spectral density (PSD) methods for microseismic event detection. Geophysical Journal International, 203(3), 1896-1908