

GM classifier - progress

Progress - Sprint 48

Done:

- Fixed FFT and SNR differences
- Custom scaled loss function
- Custom output activation function for f_{\min} (also tried for score, but doesn't work very well sample weights?)

In progress

- Pass updated predictions to Mike for usability test (especially in terms of f_{\min})

Todo

- Gridsearch
- Validation
- de-skew?

Progress - Sprint 47

Done:

- Active learning iteration f_{\min} based
- Custom output act function (+ sample weights) – (to discuss)
- custom loss function
- Added "bad" f_{\min} records back in
- SNR scaling
- CNN for SNR input

In progress

- Pass updated predictions to Mike

Todo:

- Gridsearch
- Validation
- de-skew?
- f_{\min} improvements (increase frequency range of SNR values)

Progress - Sprint 46

Done:

- Feature updates x2
- Re-ran feature extraction for updated features
- Switched to record based multi-output NN
- Fixed non-convergence issue

In progress

- Active Learning – about to pass Mike first set of new record ids to label (score-based)
- Active learning iteration f_{\min} based

Todo:

- Gridsearch
- Validation
- de-skew?
- sample weighting?
- custom loss function?

Progress - Sprint 45

Done:

- Existing functionality
- Feature extraction for the full set of records
 - Solved multiple record issues & investigated differences between new & Xavier's dataset

In Progress

- Active learning workflow
- Extracting features per component (instead of geo mean)

Todo:

- Gridsearch
- de-skew?
- Validation workflow

Progress - Sprint 44

Done:

- Re-implementation of existing functionality
 - Feature extraction (Some issues, see below)
 - Run Canterbury and Canterbury-Wellington from feature csv and records

In Progress:

- Data issues
- Exploratory work and initial gauge initial subduction performance
 - Looked at deskew vs no-deskew
 - Only use best records vs all records (with and without weighting)
 - Train model on all shallow and see how it does on subduction --baseline

Future:

- Compare trained with original on full validation dataset
- Improved validation workflow
- Gridsearch
- Active learning for subduction

Data issues:

- Some records have no buffer start date, only event start date, so should still be able to use these (currently throws an error)
- Some records in the GMR.csv from Xavier are weird, not meeting x-axis crossing condition
- malloc error, no idea?
- 3744 records instead of 3989



sub.html