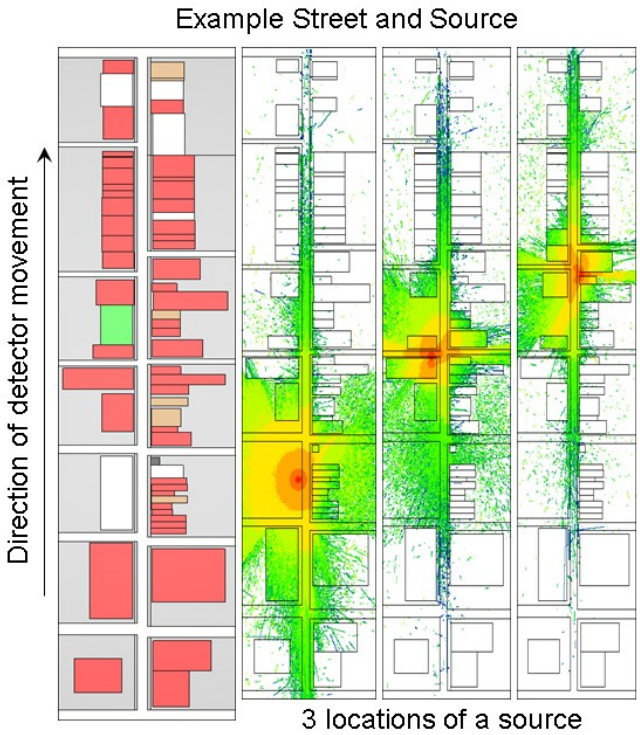
Detecting Radiological Threats in Urban Areas.



That was an interesting and hard competition. I shared code https://github.com/krunt/radthreatsmm.

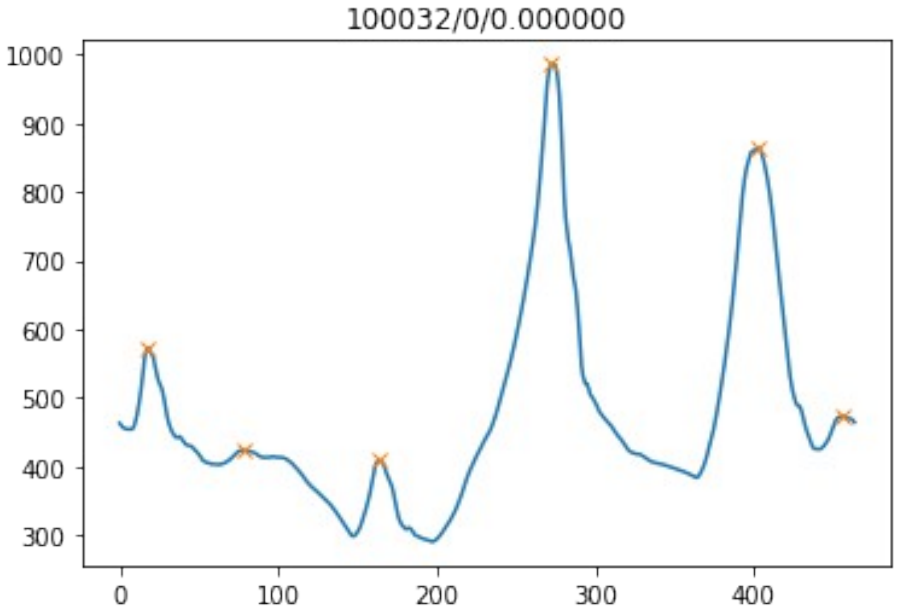
The following tasks must be performed:
1) Detection if a source located in a path
2) Localization of source position
3) Classification source isotope type

The challenges were:
1) Variable detector speed
2) Unknown source standoff
3) Unknown source strength
4) Unknown source position along street

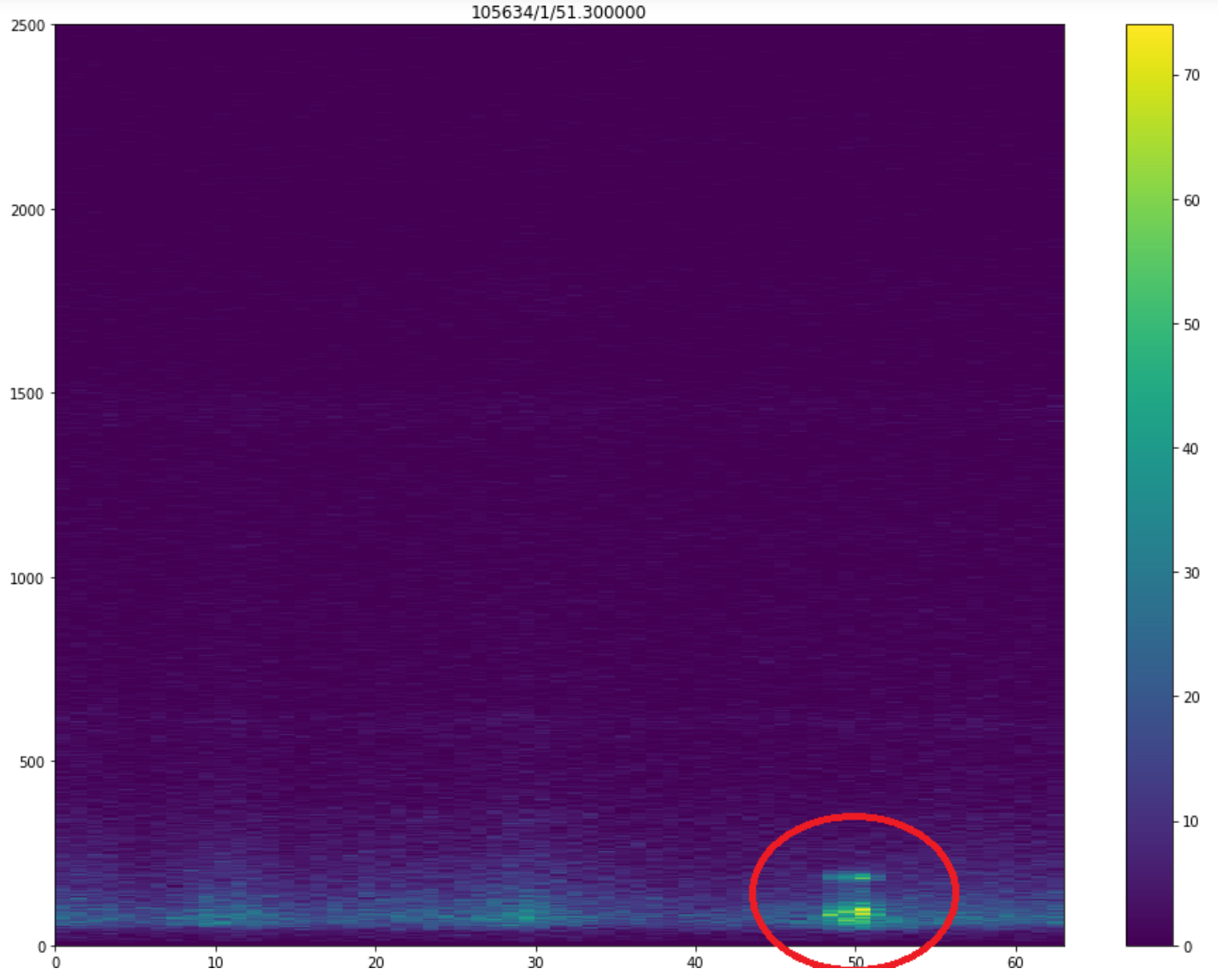
The score functionwas specially designed to penalize false alarms.

My approach
During EDA I declined peak matching approach (the resolution of NaI detector is low, peaks have big FWHM and it not improved my score).

Also I looked at peaks on time scale (xaxis – time along route, yaxis – count rate). But ground truth source positions didn't much correlate with peak positions, so I denied this approach.



Next I looked at waterfall plots, xaxis – time, yaxis – energy bins, color is count rate.

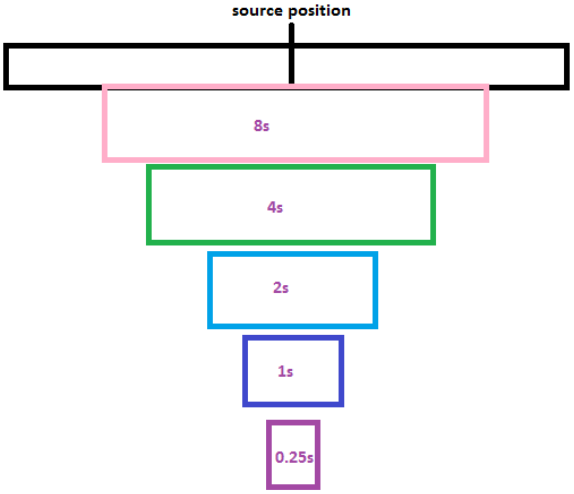


From here I designed simple approach based on waterfalls and zscore, but it gave me only 56 points on LB.

From there I started to read literature on the subject.
I based my next and final approach on [1].
1) The idea was to fit NMF on background data (NMF_bg) and then have separate NMF for each source type with source histogram injected (NMF_si).
2) The localization was performed by stepping through path by fixed step=0.5s and creating energy histograms (h_i) of several scales (8s/4s/2s/1s/0.5s/0.25s),
fit each histogram to NMF_bg and to each source NMF_si,
then for each histogram calculate goodness of fit:
3) gfit = norm(NMF_bg – h_i) / norm(NMF_si – h_i).
4) Apply fixed threshold to gfit.
5) If it is accepted take the position on path with max gfit.

That approach gave 70 points or so.

Then I tried to improve classification of source type
I designed a multiscale feature around source position.
Each scale feature consisted of
NMF weight matrix components and corresponding gfit.



|Wi_8s|gfit_8s|...|Wi_0.25s|gfit_0.25s|

I trained xgboost and source type precision was close to 90%.

The score on validation data was:

```
  Public part:
    TP =   744 | FP =     9
    FL =    76 |-----------
    FN =   147 | TN =   964

    Correct type:  675 / 744 = 90.73 %
    Average distance bonus: 80.16 %
```

I tried several approaches to improve precision, but got TP reduced to FL instead, Also I've not found a way to reduce FN without worsening TN much.

Misc hints:
In preprocessing I used wavelet-denoising on source signal, but when I tried not to use it - got much better score, SNR was improved.
Concerning shielded sources – I unified them with unshielded in corresponding source NMF.
Concrerning mixed source type 6 – I unified source type 1/5 (shielded/unshielded) in corresponding source NMF.
The problem was bad type classification between source types 1/5/6, this was improved by approach above.

Bibliography
[1] Joshi, Detection and Identification of Radiological Sources with Non-negative Matrix Factorization