Second peer review report (MMDF)

Polina Pilyugina, Alexey Voskoboinikov

May 13, 2021

1 Code review summary

Structure In general, repository is well-structured. The 2 main files are $format_brat.py$ that contains utilities to work with brat annotations, and *parser.py*, with functional for transforming dataframe. In the report actors mentioned, that they extracted P wave using a discrete wavelet method. Unfortunately, we could not find this part of code in the repository.

The repository has 3 IPython Notebooks. The first one is about transformation of Brat annotations into *.csv* file and dataframe. The second notebook notebook shows how to apply parser rules to the dataframe, and third - investigation of different models. Unfortunately, due to lack of the data, it was unclear, how to properly run this notebooks. Notebooks did not show of the main aspects of the code in a clear way.

Possible improvements The file *mistakes_handler.py* has a little chaotic structure, so we applied some refactoring to this code to make it more readable. Moreover, we would suggest using some spelling correction libraries for this purposes.

Changes to the code are available in this Pull Request `https://github.com/adasegroup/MMDF-multimodal-medical-features/pull/4`

As for content, in the function parser.thyroid_diag (function determines if the patient has a symptom in the diagnosis) there is no split of the diagnosis into hyperthyreosis and hypothyreosis, which may have impact onto the prognosis.

2 Project readiness

We see that the team has implemented and done experiments for several main steps of the project, contained in Text features part of the repository. This includes text feature extraction and benchmarking and ECG feature extraction. Overall, the team is on the good pace with their project creation and has done a big amount of work already.