Quantifying nitrogen oxides and ammonia via frequency modulation in gas sensors

Master thesis - Progress report

Marcos F Mourão

March 21, 2021

Most of the progress so far was on the writing of the thesis report: introduction, theory and data (partially) sections are fairly advanced. I also participate in regular meetings with Annika and other students (Erik and Mudith) to discuss writing and the content of our thesis. I have been also meeting with my external supervisor, Mike Andersson, for lab visits and discussions on the experiments.

As far as implementations of the methods goes, progress was hindered by a considerable delay on the data acquisition experiments. Delays were due to equipment fault, power cuts and other problems with the measurement system. As of now (March 21, 2021), the experiments are being run (during the weekend) and should be available this week. This new data is expected to be more complete, with a well defined measurement layout: more samples, more levels (possible concentrations), and the shape features (slope and average) are being directly measured in the lab, i.e. the raw sensor response will not be available.

Nonetheless, a few weeks ago some preliminary experiments were run and I managed to use the data from it to create a dummy dataset. The data, however, is significantly different from the real one, with very few data points. Additionally, the low sample rate from the lab equipment made the shape features extraction difficult. I go in a bit more detail as of why this data is different in my presentation.

I tried some basic regression methods, namely linear, principal components, partial least squares and ridge regression. Hyperparameters were chosen via cross-validation with root mean squared error (RMSE) scoring. In my presentation I display some very crude results of the regression predictions. The results are poor. I did not look deeper into it given the aforementioned limitations. The most promising method is Partial Least Squares Regression, as it is wildly used for this kind of

data/field.

Regardless, code is (mostly) ready when the real data comes for these initial models. I also want to look into some non parametric regression techniques and compare. I have been thinking on support vector regression or a deep learning alternative.

I am aware that I am at risk of being late with my thesis, but I am trying / will try my best to meet the first submission. I am in regular contact with both Annika and Mike. Meanwhile, I am focused on other parts of the thesis work, namely reading literature and writing my report. I deeply believe I will make significant progress in the following weeks.

As soon as I have the real data, in addition to applying the methods, I intend to do a more thorough analysis, specially on linearity, possible outliers and collinearity of the predictors.

APPENDIX: SAMPLE TEXT