MidTerm Report

2022-10-20

Bilateral Trade Flow Modelling

Team: Helios

Prateek Pangal Rao: PES1UG20CS303

Pooshpal Baheti: PES1UG20CS283

Rahul Ramesh: PES1UG20CS319

On completion of Exploratory Data Analysis, we realized that the data was extremely noisy with plenty of NaN values.

First, we performed cleaning of data where we:

- Selected an initial feature space, which includes all features currently used by the Gravity Model, by economists, and some features we decided might be useful based on limited domain knowledge.
- Removed all instances where a feature was `NaN`.
- We also cleaned up and removed any instance where flow was less than 100-200 GBP, because this was causing outlier issues.
- We applied logarithmic smoothing to achieve an even distribution of data.
- We are contemplating removing data before the 2000s in order to improve accuracy of forecasting of more recent trade flows.

Satisfied with this, we started building time-agnostic models, i.e, we do not factor time lagged values at all. This obviously does not give a great forecast, with Linear Regression and SVMs giving only 0.13 and 0.61 R^2 values. However, this provides us with a good baseline, moving forward.

Future work would include further feature selection by measuring impact of features using Gradient Boosting Models, and incorporating time-lagged values to perform estimation. We will also explore Deep Neural Networks for Time Series Analysis to improve forecast accuracy.