IARPA monthly report

Reed College team activities during month 13, May 2020

- team leader – Mark Bedau
- student 1 – Tobias Rubels
- student 2 – James McCaull (sidelined in May due to Covid-19 restrictions)
- student 3 – Ananthan Nambiar (now a graduate student at UIUC)

Activities during May

In May the Reed team completed the first step in measuring the across-industry variation (AIV) model success metric (Hoberg and Phillips 2010) for the product spaces of semantic vector models. Due to Compustat license restrictions, the second step is for Reed's partners at Oxford to calculate the AIV metric from model firm classifications produced by Reed. For AIC calculation Reed delivered to Oxford firm classifications from a suite of product space models, including, among others, models TNIC-2018, Boolean 10k 1.0|0.06, Boolean 10k 0.2|0.06, doc2vec 10k 1.0|0.0, and doc2vec (word2vec) 10k 1.0|0.0. Reed also delivered firm similarity matrices for each of these models, so that Oxford can measure their relative effect on sales forecasts of each firm similarity matrix. Reed also delivered to Oxford for dynamic community analysis the patent vectors from a doc2vec model of technology space, with the vectors indexed by patent date and ID number.

Activities planned for June

Reed plans to train the most important word vector and doc2vec models on a 10k training corpus that includes only non-financial firms with Compustat gvkeys (following Hoberg and Phillips 2010), and then to deliver to Oxford for AIV measurements the models' firm classifications. In June Reed plans to analyze the results of the AIV calculations when they are received from Oxford, by comparison with the current industry standard (TNIC-2018). Reed also plans to analyze the results of the models' relative sales forecast improvement when they are received from Oxford. Reed also plans to deliver in the first week of June a report documenting the milestones achieved in project Task 2.2.