June 9, 2017

**Proposal 1** – Reproduce the results in *Quantitative determination of technological improvement from patent data*” by Benson and Magee.

Use data and models from [“Quantitative determination of technological improvement from patent data”, C.L. Benson and C.L. Magee, PLoS one, 2015, 10 (4), e0121635], reproduce the results on some or all of the hypotheses they propose. By using patent meta-data, especially the number of patents, forward citations and the rate of forward citations, they show that it is possible to show if a particular technology is improving. They do not predict the rate of improvement. (There are attempts to do this but this is beyond the scope of this work).

They analyze the results with linear regression. The results can also be analyzed with a random forest regressor to identify which variables are most important. Natural language processing could be used to identify trends in the technology.

**Proposal 2** – Analyze the state of technology of lithium ion batteries using the same method as Proposal 1. Lithium ion batteries are the power source for portable electronics and electric cars. It would be useful to have an assessment if the technology is improving.

Using the same method as Benson and Magee, download the approximately 1000 patents dealing with lithium ion batteries. Using a previously written web scraping program, digest the patents and analyze the meta-data using one or more of the models proposed in 1.

Use linear regression, random forest or some other method(s) to analyze results. NLP may not be particularly useful in this system.

**Proposal 3** – Download the patents with the word 'alloy' in the title from the US patent office. Perform natural language processing in the abstract of these patents to determine which metals are being used in current alloys and what trends are developing. So, use NLP from 1980 to 1985 to determine the most common elements, then from 1990 to 1995, from 2000 to 2005, and 2015 to June 2017. Plot the trends.

Identify classes of alloys such as amorphous, shape memory, steel. See what trends are developing in the sub-classes of the word 'alloy'.

For the compositions that can be determined from the abstract, plot the amount of element versus its order in the alloy. The plot is expected to be a logarithmic decay function.

Apply some of the models of Proposal 1 to these results and see if the technology is advancing. There are about 10,000 patents in the database from 1976 to 2017.

**Proposal 4** – Download the patents with the word 'LED' in the title or patent classification code and apply the method of Proposal 1 to the results for the time frames 1990 to 1995, 2000 to 2005 and 2010 to 2015. See if the technology is advancing.

Use NLP to see if the word 'efficiency' can be extracted from the patent. Plot the efficiency of the LED over time. See if a trend develops.

Project proposal

Once you’ve picked your final capstone project idea, you will write a proposal. The project proposal is a short (1-2 page) document that answers the following questions:

1. What is the problem you want to solve?
2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?
3. What data are you going to use for this? How will you acquire this data?
4. In brief, outline your approach to solving this problem (knowing that this might change later).
5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.