Submission for PhD position in computational social sciences at ETH Zürich

March 2, 2020

Marius Herget

Contents

Task 1

My first experience with research was within my practical training for my B.Sc. employer at the IBM Almaden Research Center in 2016. I experienced first hand the enthusiasm of the scientist in asking and pursuing groundbreaking questions. After these three months, I decided to follow my inner craving to change the world. Especially the scientific approach with no prejudices, religion, or political agenda matches my inner conviction of just listening to facts. These experiences convinced me to achieve the highest university degree.

Within my life so far, I always set myself the standard to obtain and deliver the best possible outcome. Nevertheless, I . . . .

My ambitions for my Ph.D. position is a mix of a great team and flexibility. Exploring new areas, receiving founded advice, and the willingness to surpass oneself should be the daily business. Nevertheless, the essential aspect of any future career step is the working environment. The personal and professional climate has to be friendly and giving me a strong impetus. Research wise I am very open-minded where the journey goes. I love the challenge of familiarizing myself with new facts and theories.

Task 2

test

Task 3

Since the dataset is limited to only one specific research field, I would like to ask the following research question as a base for my data analysis:

Exist unnoticed differences in wording within various research groups of a similar research area?

My analysis would take several steps:
1. Identify research groups based on the connection of *author*, *affiliations* with references of their publication to other papers and their authors (*citations*).
2. Identify significant words within the *title*, *abstract*, and *full text*. This includes a data cleanup of common words like *is*, *are*, *they*, *and*, etc.
3. Connect both results of steps 1. and 2. to recognize which research group applies which wording (= wording bubbles).
4. Classify word bubbles and review their significance.
5. Examine significant words by connecting synonyms within the dataset. Compare wording bubbles and identify via step 5. how various research groups differ in phrasing and content.

The goal of this analysis is to recognize whether scientists miss relevant research publications in consequence of their **wording filter bubble**.

This could be quite interesting for identifying other important papers while minimizing research expenses. On the one hand, in a practical application, a reduction of reading the current state of the art can be quite impelling. On the other hand, a broadening of the horizon can enhance a publication's quality.

Task 4

This task is centered around the fact of finding the most important member of the House. Firstly it is crucial to define what importance means in this context:

An important person is an individual who has the most power within the House. There are several aspects to consider:

Money An influential person needs financial stability and has a high budget.

Reelections Besides financial stability, a politician depends on its position in his party and its district's population (voters). Another aspect is that a member of the House gains more importance with his experience and the broader network.

Bills Another aspect besides the politicians standing is the importance of his sponsored bills. In this assignment, this is shrunk down to the number of cosponsors and their party. It is more challenging to receive cosponsors from the opposite party.

This is only a small sample of a lot of different aspects to recognize the most important person.

The following pseudo equations give a quick overview how the importance is calculated

$$Importance_{Bill} = (1.1 * \frac{CS_{otherParty}}{CS_{All}} + 0.9 * \frac{CS_{ownParty}}{CS_{All}}) * Normalized(CS_{Total}) \tag{1}$$

$$Importance_{Representative} = \frac{1}{2} * Normalized(\sum Money_{Individual}) + \frac{1}{2} * \frac{Periods\ in\ House}{4} \tag{2}$$

$$Importance = \frac{1}{2} * Importance_{Representative} + \frac{1}{2} * \sum_{Sponsor=Individual} Importance_{Bill} \tag{3}$$

To achieve the correct analysis, the given data had to be supplemented by [1] and [2].

References

[1]   *Bills - ProPublica Congress API.* [Online; accessed 24. Feb. 2020]. Jan. 2020. URL: `https://projects.propublica.org/api-docs/congress-api/bills/#get-cosponsors-for-a-specific-bill`.

[2]   *House Office Expenditure Data.* [Online; accessed 24. Feb. 2020]. June 2019. URL: `https://projects.propublica.org/represent/expenditures`.