Mapping global STEM funding using AI

Research proposal

Author

Hanyi Jiang

April, 2023

Supervisor: Dr Samraat Pawar

Department of Life Sciences

1   Keyword

Funding organization, Research grant

2   Introduction

Hand in hand with the rise of New Public Management and expanding global techno-economic competition, increasing prominence has been given to the idea that countries have been paying more attention on research project over the last 20 years.[1] The two most important components of any research project is idea and execution. The successful execution of the research project depends not only on the effort of the researcher but also on the available infrastructure to conduct the research. The conduct of a research project entails expenses on man and material and funding is essential to meet these requirements.[2] Therefore, the funding largely influences the project's progress. In this project, we will focus on the direction of funding to find the change in the focus of the academic committee over time in UK and US. Then the results are used to analyze the differences in subject priorities between the two countries.

This project aims to answer the following questions:
1. Whether and if so, exactly how the core subjects and STEM priorities of governments have changed over time?
2. Do the STEM priorities of governments in the UK and the US overlap at the same point in time?

3   Method

I already have the pre-processed clean data which contains information such as project title, abstract, and funding amount, so I can spend a short time double-checking to make the data specifically fit my focus.

Topic models provide a simple approach to analyze huge volumes of unlabeled text.[3]Through topic analysis, we can analyze the content of successful grant proposals to find common themes in large collections of text. Then fit and train LDA model to model research topics to get the result of topic probabilities for each document, which is a quantitative measure of dominant themes for each document. The hyperparameters α (alpha) and β

(beta) are involved in this step.[4] Finally, after identifying topics across all the funded projects, we can quantify how funding has been allocated across the topics in each year in UK and US. This analysis consisted of four steps: (1) calculating total funding in each year in UK and US (2) comparing received funding in each topic (3) listing the top 10 subject priorities in each year in both UK and US and labeling all topics using the Library of Congress Classification Outline as a guide.

4   Anticipated Outcomes

For the first question, the best-case scenario is that some specific subjects consistently at the top of the rankings, some subjects from the top change to the middle of the list, or some subjects list position change from bottom to the top. However, sometimes it may not be such easy, if we cannot find the change of the subjects, some other analysis may need.

For the second question, the top 10 subject priorities of UK and US should overlap to a large extent. If time permits, we can analyze in a more in-depth way to quantify in a statistical perspective.

5   Time Line

Month 1: Reading and finding resources, Writing Introduction
Month 2-3: Parameterizing models
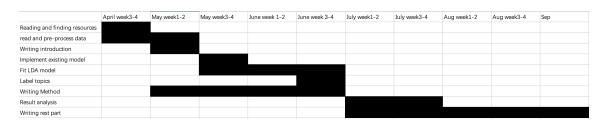Month 4-5: Analysis and writing

	April week3-4	May week1-2	May week3-4	June week 1-2	June week 3-4	July week1-2	July week3-4	Aug week1-2	Aug week3-4	Sep
Reading and finding resources	■									
read and pre-process data		■								
Writing introduction										
Implement existing model			■	■						
Fit LDA model										
Label topics					■					
Writing Method		■	■							
Result analysis						■	■			
Writing rest part								■	■	■

Figure 1. *Estimate Timetable*

References

[1] Sargent, J. F. *Global research and development expenditures: Fact sheet* (Congressional Research Service, 2017).

[2] Neema, S. & Chandrashekar, L. Research funding—why, when, and how? *Indian Dermatology Online Journal* **12**, 134 (2021).

[3] Jelodar, H. *et al.* Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications* **78**, 15169–15211 (2019).

[4] MonkeyLearn. Topic analysis: The ultimate guide. `https://monkeylearn.com/topic-analysis/`.