

Project Proposal

Zitong

April 2023

In my initial exploration, I want to focus on the government funding that Chinese universities received in STEM disciplines. The government-funded projects received by each university can be clustered and analysed using machine learning (text mining, topic analysis, hierarchical clustering), and similar projects can be grouped into the same category to find similarities and differences in funded projects on universities and subjects' level. This helps to examine the focus of Chinese government-funded projects to determine if collaboration with Chinese universities is appropriate.

I found the National Natural Science Foundation of China (NSFC), a Chinese website on funding research in the subject, which has many reports on the specific allocation of funds. This site has the total amount of funding for the subject categories (I have attached the translated version). This summary data included are the annual funding received by sub-discipline and the funding received by each university. I can use these data to test the efficiency of using machine learning for classification (Using Bayesian: $P(t|D)$, MCMC to test and improve the efficiency of machine learning method). However, there is an issue that it is indicated on this site that it is not permissible to crawl every project data, so I do not know how to get the dataset (and I do not know if it is feasible to buy these dataset).

I can also do the efficiency test with another available country's raw dataset and the official summary data.