sExploring China’s Government Media on Central Asia

* Data: Xinwen Lianbo (i.e., Network News Broadcast <https://tv.cctv.com/lm/xwlb/>) transcript web scrapping (June 1, 2017 – May 31, 2022). I first search the keyword “中亚” (Central Asia) and then scraped all the transcripts that mentioned this keyword.
  + Notes: People’s Daily Database is authorization required. I applied for authorization but still waiting.
* Key findings: The term “Central Asia” is highly associated with the term “energy”, and “Shanghai Cooperation Organization”. I suppose (by reading the transcript) the term “Central Asia” is also highly associated with “Belt and Road Initiative”, and “A community with a shared future,” but I encountered the coding issues when I tried to conduct an analysis. Still fixing the issues.

Exploring Weibo Comments

* Data: Weibo Comments web scrapping. I search “Central Asia” on Weibo, and it displayed 757 related posts (38 pages), I scrapped all of them (but deleted unrelated posts).
  + Notes 1: Weibo API is not very useful since Sina (Weibo’s owner) set many limitations on Weibo API. Therefore, I conduct web scrapping.
  + Notes 2: As mentioned above, I got 757 posts related to “Central Asia,” but all of them were posted in 2022. Weibo hides posts earlier than 2022. I need to use “Advance Search” to get earlier posts.
* Key Findings: By exploring the word-correlation matrix, when Weibo users mention “Central Asia,” they are most likely talking about soccer, history (“silk road,” “Khwarazm”), Xinjiang, and politics (“Russia”, “Russian”, “USA”, “China”, “Ukraine”, “Nation”, “Afghanistan”)

What’s Next

* Further analyzing Xinwen Lianbo and Weibo data.
  + N-gram: Exploring the relationship between words. I tried R package “tidytext” to make n-gram, but it seems not compatible with Chinese. So I need to find other packages for making n-gram.
  + Sentiment analysis. We might conduct sentiment analysis on Weibo comments. But I need to find an R package that is compatible with Chinese characters.
* Web scrapping: People’s Daily
  + Since I may not get the authorization of the People’s Daily Database, I can try to do web scrapping on the People’s Daily’s website.