Connecting CMU Researchers to Current Grant Opportunities

By: Divya Rao, Pam Sun, Lisa Zhang

Advisor: Peter Freeman

Carnegie Mellon University Libraries

Aim

- In Fall 2020, a group of CMU Statistics & Data Science undergraduates developed a data exploration tool to explore and analyze data on grant funding involving CMU researchers. As an extension to this work, we are interested in building a data recommendation system that matches previously funded CMU researchers (as documented in one database) to future grant solicitations (as provided in another).
- We utilize the `Dimensions` (previously funded) and `GrantForward` (future solicitations) databases.

Data

- The data were provided by Neelam Bharti, Huajin Wang, and Sarah Young of the CMU Libraries.
- The `Dimensions` dataset is a subset of 6,326 previously funded research grants affiliated with CMU from the Dimensions AI database. This dataset consists of information about the grants such as title, abstract, funding amount, start/end date, researchers, funder/funding group, resulting publication, grant URL, and categorizations. We leverage the data from six categorization systems to generate the subject list we use when matching across datasets.
- The `GrantForward` dataset is a subset of 2,856 currently available research grant opportunities with solicitations from September to October 2021. This dataset consists of information such as title, description, sponsors, funding per grant, grant URL, and subject.

Matching Opportunities to Researchers

Input: Name of Grant Opportunity

- Use `subject` column to obtain specific topics, and more generalized ones with fuzzy matching
- Create a list of related researchers through the matching of obtained topics to category columns in dataset
- For each related researcher, quantify the correlation between the words in the grant `description`, publication `description`, and researcher `abstract`
- Rank the list of related researchers based on word-wise correlation and return the top ten matching researchers, along with related information

Output: Researchers

Future Work

- As mentioned above, our recommendation system currently uses a subset of data from the `GrantForward` database. In the future, we aim to tie our recommendation system to the `GrantForward` database by utilizing its API.

Grant Opportunities by Category

List of relevant main and subcategories

Grant opportunities filtered by input subcategories and submission date

Information of specific grant opportunity (source URL, amount, and description)

In Progress: Matching Researchers to Opportunities

Input: Name of Researcher

- Use category columns to obtain specific topics and generalized ones with fuzzy matching
- Obtain a list of related grant opportunities based on the matching of topics
- For each related grant opportunity, quantify the correlation between the words in the grant `description` and researcher `abstract`
- Rank the list of related grant opportunities based on word-wise correlations obtained and return the titles, URLs, and correlation values for reference

Output: Grant Opportunities

References

- *Dimensions*, https://www.dimensions.ai/.
- "GrantForward." *GrantForward Search Engine | Search for Federal Grants, Foundation Grants, and Limited Submission Opportunities*, https://www.grantforward.com/index.
- Robinson, Julia Silge and David. "Welcome to Text Mining with r: Text Mining with R." *Welcome to Text Mining with R | Text Mining with R*, https://www.tidytextmining.com/index.html.
- The Fall 2020 Project: https://www.youtube.com/watch?v=ijZXBx82ExA