Understanding Congressional Bills

Adarsh Chavakula, Chi Zhang, Somya Singhal, Xiaochun Ma
Industry Mentor: Dr. Amanda Stent
Faculty Mentor: Dr. Smaranda Muresan

Introduction
In this project, we attempt to gain insights into newly introduced US Congress bills (Senate and HR) by using NLP to understand their impact. Our main focus is to know which older bills the new bills are linked to. Additionally, we extract the the topics of the bills, the financial impact they would have and the population groups affected by them.

Exploratory Data Analysis
Some interesting insights from EDA: Number of bills increase over years(Fig. 2-a), different States have different focuses(Fig. 2-b), Children, women and elderly were mentioned a lot(Fig. 2-c), spring and summer tend to be good seasons for application(Fig. 2-d), CA, NY and TX are the top 3 enthusiastic states(Fig. 2-e) and Health and Taxation are the most popular two topics around the country(Fig. 2-f).

Methodology
For the impacted population and quantity, we used SpaCy Named Entity Recognizer (NER) for information extraction. To improve recall, Regular Expressions and manual extraction using a gazetteer were employed.

To visualize the relevant bills, we calculate bill similarity based on the laws they refer to using cosine similarity. Bills are considered to be similar if the cosine similarity is greater than 0.99. We created an interactive network visualization, which clearly shows the bill clusterings and allows users to dig into each bill’s full content with one click on the nodes.

Results
Using the evaluated similarity between bills, we obtain a graph network of the bills. The network based visualization helps visualize the clusters of similar bills.

Figure 3. Visualization of relations between 500 randomly sampled bills from 2018

Using NER we obtain the most commonly cited departments in the bills. Department of Defense (DOD) and Department of Homeland Security (DHS) have the top mentions.

Figure 4. Top governmental departments mentioned in the bills

Conclusions
We were able to find the relations and created an interactive network(graph) of the congressional bills, on which you can link two bills, find a cluster of bills that focuses on a similar topic or click to find the attributes that interests you most.

Acknowledgments
Best thanks to the guidance and advices from our industry mentor Dr. Amanda Stent and faculty mentor Prof. Smaranda Muresan.

References