Identifying a Potential Cause of Inefficiency in ABA Free Legal Answers Systems with Natural Language Processing

Tiantian Li, James Park, Alexia Witthaus, Anusha Kurapati
Department of Quantitative Theory and Methods, Emory University

Background

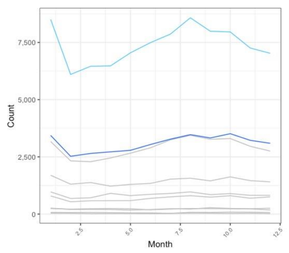
The American Bar Association (ABA) Free Legal Answers is an important part of making our justice system both equitable and accessible. By matching clients with attorneys on a pro-bono basis, low-income clients can obtain the services that their wealthier counterparts are able to afford, ultimately preventing them from legal issues that may arise. Thus, we understand that filling out accurate information and efficiently matching these clients to the right attorneys is a significant process that can help a plethora of individuals. The purpose of our approach is to examine why the category "Other" is the second most selected category, and accurately depict how we can make this matching process more efficient through machine learning.

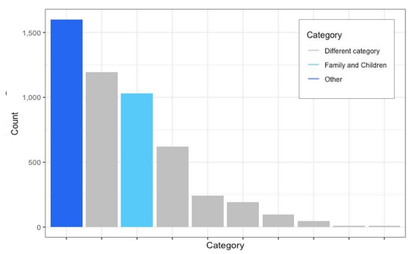
Summary Statistics

- Average Attorney Matching time: 10d 21h
- Average Time to Close: 8d 21h
- ½ of the questions with legal deadline are closed past the deadline

Family and Children	88753
Other	37053
Housing and Homelessness	34755
Consumer Financial Questions	17273
Work, Employment and Unemployment	10187
Individual Rights	8312
Health and Disability	2707
Income Maintenance	2688
Education	794
Juvenile	357

Name: Category, dtype: int64





Methodology

Keyword extraction → Word embedding → Unsupervised Learning

Keyword extraction – YAKE [2]

- Unsupervised model for identifying key phrases from documents based on statistical features extracted from text

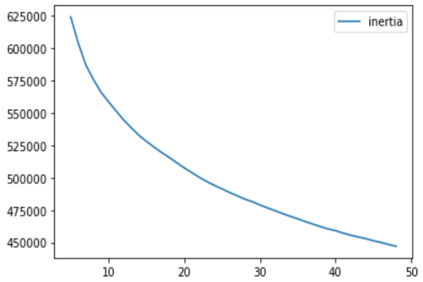
Word embedding – GloVe [1]

- Text vectorizing methods that make it possible to represent each word by a unique numerical vector
- Ability to capture semantic relationships between words and phrases found in the conversations between the clients and attorneys
- Pre-trained on Wikipedia 2014 and Gigaword 5 (news articles before 2010)

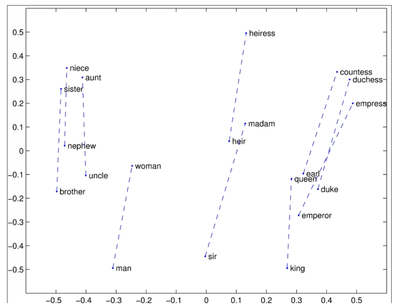
Unsupervised Learning

- Perform unsupervised clustering: K-means clustering on the average of the word embedding vectors of the top three key phrases (3-gram), with K = 15 (15 clusters)
- Each cluster showcases words that fit under a certain category from the Free Legal Answers categorizing system

K-means Clustering Elbow Point

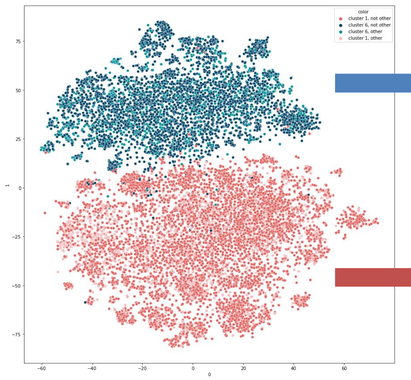


Global Vector for Word Representation (GloVe)



Results & Conclusions

Results



Family and Children	2558
Other	1490
Housing and Homelessness	643
Work, Employment and Unemployment	497
Health and Disability	387
Individual Rights	371
Consumer Financial Questions	197
Income Maintenance	166
Education	34
Juvenile	23

Consumer Financial Questions	2921
Housing and Homelessness	1945
Other	1881
Family and Children	1501
Work, Employment and Unemployment	780
Income Maintenance	273
Individual Rights	237
Health and Disability	149
Education	14
Juvenile	2

Conclusions

The category "Other" is the second most selected category among clients, but the clusters show that these conversations can be filled out under the given specializations.

Re-assess current categorization method to more effectively match the right attorneys to the right clients.

- Semi-automate the categorization system: Category recommendation based on keywords
- Add features for attorneys to re-classify question posts

References

1. Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation.
2. Campos, R., Mangaravite, V., Pasquali, A., Jatowt, A., Jorge, A., Nunes, C. and Jatowt, A. (2020). YAKE! Keyword Extraction from Single Documents using Multiple Local Features. In Information Sciences Journal. Elsevier, Vol 509, pp 257-289.