Gregory Knapp

Feb 19, 2022

Final Project Proposal

Measuring collection diversity and representation in a library setting has become an important task in recent years. A previous study showed a pattern of measuring the cultural diversity of the local area, manually evaluating a sample of books from the collection across different subject areas, then adjusting holdings to more closely reflect the demographics of the area.[1] As part of my work as a library research assistant in the Psychology section, the idea of doing similar work in BYU's own Harold B. Lee Library was brought up to me. However, we quickly realized that a diversity audit using the previously described demographic-centric method would be difficult in a state that is 86.43% white.[2]

As we brainstormed ideas, the idea to diversity classification by topic viewpoint came up. We wondered if there was a way to determine if books in the collection covered a diverse range of viewpoints on a topic, especially controversial ones like gender, sexuality, etc. We then realized that previous studies hired teams of 15 or more students and employees to manually evaluate and rate books for their criteria, something that we lacked resources for. This led me to the question of my final project: Can a sentiment analysis neural network be used to judge or score the diversity of viewpoints contained in our library collection for one specific topic?

The approach of training a Sentiment Analysis network was inspired by research done regarding scoring political points of view regarding the 2016 presidential election.[3] In that study, Recurring Neural Network was trained on specifically selected political news articles that were associated with different viewpoints, and was tested on Twitter data from the timeline of users.

I intend on taking a similar approach with my project; rather than use a pre-existing sentiment analysis model, I plan on implementing my own network from scratch. This process will involve experimenting with Bi-Directional layers, LSTM vs. GRU, 1D Convolutional networks and more to find out which implementation has the best results. I will source data from the both the web (get articles about the selected topic reflecting the different viewpoints for training) and via the Library IT department (to get book data that is not publicly available on the HBLL website for testing). I am not sure if the model will be able to function well with just a book description text or if it needs full-text information to predict more accurately; I hope to experiment with that as part of my research as well.

Lastly, when training and testing is complete, I hope to score my models rating against a panel of library professionals' manual evaluation of which viewpoints are reflected by the books to determine if my model is accurately recognizing the content or not. Ideally, my model would be able to classify each text as representing a viewpoint on the topic. If this is the case, then we try to use it on the entire topic section of books, not just a random sample, and see what the model says about the viewpoint diversity in that part of our library collection.

---

[1] Jensen, K. (October 22, 2018). Diversity Auditing 101: How to Evaluate Your Collection. School Library Journal. https://www.slj.com/?detailStory=diversity-auditing-101-how-to-evaluate-collection.

[2] World Population Review (2022). Utah Population 2022. https://worldpopulationreview.com/states/utah-population

[3] Alzhrani, K. M. (November 15, 2021). Political Ideology Detection of News Articles Using Deep Neural Networks. Intelligent Automation & Soft Computing. https://www.techscience.com/iasc/v33n1/46167/pdf