Evanston Public Library Book Recommender

Justin Aronson

Northwestern University, justinaronson2025@u.northwestern.edu

Adam Chen

Northwestern University, adamchen2024@u.northwestern.edu

Sung Park

Northwestern University, joonpark2024@u.northwestern.edu

Andres Rojas

Northwestern University, andresrojas2024@u.northwestern.edu

William Wu

Northwestern University, williamwu2024@u.northwestern.edu

Given the Evanston Public Library’s (EPL) large selection of books, it could be difficult for patrons to choose what to read next. To remedy this, we developed a book recommendation system using a variant of k-nearest neighbors (KNN) with cosine similarity. Patrons can input books they like into our program to get recommendations. Furthermore, we engineered our program to promote authors of diverse backgrounds and stories of people of color. We also built our recommendation system with the hopes of benefitting EPL’s contributions to the maker movement by easing the process of finding materials related to making. Our project is a significant step for the EPL, since, to our knowledge, there are no prior machine-learning based recommendation systems available to patrons. Although this is just a starting point, as our recommendation system grows, we hope it can help tackle issues of inclusivity and making inaccessibility, while also just being a generally useful tool for patrons.

CCS CONCEPTS • Computing methodologies Artificial intelligence • Natural language processing

1. Introduction

Over this past spring quarter, our group had the wonderful opportunity to make multiple visits to our local public library: the Evanston Public Library (EPL). Even though there is a clear demarcation of different areas in the library that serve different topics of interest, there is still room for improvement in terms of catering visitors to a wider array of books by authors from diverse backgrounds. For example, if I wanted to read books addressing a view of American capitalism and its socioeconomic effects from an immigrant perspective, it would likely involve numerous unsuccessful round-trips of manually searching through bookshelves before needing to seek a librarian for help.

Given the context of the problem, our group proposed an accessible way to allow underrepresented ideas in books to flourish and ultimately become more streamlined in the search process. Hence, our group developed an Evanston Public Library focused book recommender system that leverages ML algorithms (most notably K-Nearest-Neighbor) in order to provide greater efficiency and convenient search mechanism for people who are seeking to obtain books written by minority-background authors as well as collaborate on such expounded ideas. Even though our recommender system currently works with Evanston Public Library’s specific books/video/audio offerings, we hope that our initial prototype can ultimately serve as an ideal model to expand and scale to meet the same needs of other libraries outside of EPL.

1. positionality statement

As everybody in our team has a background that is rooted in a minority group, each of us has experienced a point in our lives where we felt that certain ideas we were exposed to didn’t fully resonate with us or could have benefited from a more diverse perspective. Inspired by this, we all sought out ways to bring more attention to perspectives and opinions from people that come from underrepresented backgrounds like us. We came up with the idea of creating a library recommendation system that promotes stories written by people of color because we felt like most patrons at the library are people that have limited access to technology or the internet. As a result, they are not always aware of the plight that people from minority backgrounds face when it comes to sharing ideas. By creating a recommendation system with diverse backgrounds in mind, these patrons can learn from our unique perspectives, have an opportunity to thoughtfully engage with them, and hopefully contribute to a more diverse community and future.

1. Overview of project “Deliverable”

In terms of our project deliverables, we wanted to ensure that our recommender system’s generated book recommendations take account demographic factors into similarity scores that are utilized in our k-nearest-neighbor algorithm (our primary ml algorithm). The most utilized factors are race and ethnicity, where we accentuate similarity scores for books written by authors possessing minority background by scaling by a parity factor of 1.1.

 Overall, our current state of project prototype relies on data being imported locally via normal csv files. If given more time, we hope to ideally streamline and build out a data pipelined extract, transform, load (etl) process so that data is cleaned and ready for performing ML computation such as similarity matrix computation or knn-algorithm based recommendation logic. Additionally, our resulting output is stored locally in an output file, but we would ideally like to present info in a summative, visual format by building out a front-end data analytics powered dashboard of useful metrics that are utilized to drive specific recommendations to inform users of the types of recommendations they are getting.

1. Design Process and Preliminary Evaluation

After the initial meeting with the EPL, our team composed a problem statement and initial design ideas. The first problem statement was ‘Library patrons often struggle with decision paralysis when picking books to read. EPL wants to develop a book recommendation system for patrons based on books that they have checked out in the past. Potential users include students, parents, and teachers.’ Due to a miscommunication, our team believed that we would have access to anonymized user checkout history. From this assumption, we designed the following algorithm:

Say a user has checked out n books. We assume that the user likes all of the books they have checked out. We then have a sequential list of books where books 1…n-x indicate that the user could like books (n-x)+1…n. The value of x would be chosen by hyperparameter tuning. Our training set is essentially input vectors of length n-x and output vectors of x. By passing books 1… n-x into an LSTM model, which can accept sequential inputs of varying lengths, we could create a book recommendation system ‘personalized’ for each library patron based on their checkout history. We were looking into LSTM models because our application is pretty small scale - we had no need for the scalability of transformers. Since checkout history would likely be pretty short for most patrons, we also do not need the infinite attention of transformer models.

Our initial design had a few issues. First, we are assuming that a user enjoys every book they check out. This is very likely a false assumption, and we have no way of determining which books a user liked from their checkout history. We thought about parsing user reviews for books, but came to the conclusion that there were far too few user reviews for them to be a reliable source of how much a user enjoyed a book. Second, when we presented this initial design to the library team, we went over the data requirements and realized that we had misunderstood what data the library was able to give us. Because EPL data is tied to the government, they were not allowed to give us any user specific data, even if it was completely anonymized. Instead, they gave us the transaction history for all books, with no user data. As such, we proposed to build our knn inspired algorithm. This decision was backed by the EPL, and we decided to move forward with the idea.

Since we do not have any user specific data, we do not have any data that indicates whether a user who has checked out x books will like y book. Thus, we cannot test the quality of our recommendations, since we do not have a labeled dataset. Instead, we tested the algorithm for demographic parity. The EPL Adult Sci-Fi catalog is very not diverse, with White authors making up 92% of the book selection (Table 1). Note that we are assuming all of the authors with 'Unkown' race in the dataset are White. This assumption is backed by the EPL, who said that volunteers would often write 'Unknown' for White authors.

Table 1: Percentage of author ethnicities in the EPL Adult Sci-Fi catalog.

|  |  |
| --- | --- |
| Race | Percentage |
| White | 90.7 |
| Black | 4.1 |
| Asian | 3.7 |
| Latinx | 0.94 |
| Middle Eastern | 0.4 |
| Indigenous | 0.17 |

Without any tuning, our algorithm is somewhat close to the population level percentages for the EPL catalog (Table 2). However, we want our algorithm to promote authors from historically underrepresented backgrounds. As such, we implemented a demographic parity tuning mechanism, where we scale up the similarity score of underrepresented authors. We found a demographic parity factor of 1.1 to be the most successful (Table 3), though scaling with higher factors further promotes underrepresented authors over White authors.

Table 2: Percentage of author ethnicities with demographic parity factor of 1 over 50 randomized trials with 3 input books each.

|  |  |
| --- | --- |
| Race | Percentage |
| White | 90.2 |
| Black | 3.7 |
| Asian | 4.3 |
| Latinx | 0.9 |
| Middle Eastern | 0.6 |
| Indigenous | 0.3 |

Table 3: Percentage of author ethnicities with demographic parity factor of 1.1 over 50 randomized trials with 3 input books each.

|  |  |
| --- | --- |
| Race | Percentage |
| White | 39.7 |
| Black | 22.6 |
| Asian | 28.9 |
| Latinx | 4.3 |
| Middle Eastern | 0 |
| Indigenous | 4 |

We were able to successfully promote historically underrepresented groups with our demographic parity tuning. However, I would argue that our system can only do so much. The percentage of white sci-fi authors in the EPL database is enormous. Following, there are very few books in the collection by nonwhite authors. For this reason, when we select for underrepresented authors, our book recommendation system often recommends the same books. For our algorithm to have material to work with, the EPL needs to become more proactive about adding books by people of color into their collection. We will share this feedback with them when we present our algorithm.

1. Conclusions and future work

The development of the book recommendation system for EPL marks a significant step towards enhancing accessibility and fostering a more inclusive community of readers and Makers. We were able to create a tool that not only personalizes the reading experience, but also democratizes access to the library's vast resources for all authors. This project embodies the principles of empathy and accessibility, as it opens up new avenues for library patrons to discover books that resonate with their interests, which is especially important for those who may feel overwhelmed by the sheer volume of available options and no good way to search for books. In conclusion, EPL has no current recommendation system in place, but if they were to adopt ours, many library goers would be able to find and access the books that most match their interests, while also providing fair recommendations that promote authors of all backgrounds.

Looking ahead, there are several exciting avenues for expanding and refining the recommender system. One key area for future work is broadening the scope of the system beyond the current focus on the adult sci-fi collection. By incorporating data from the entire library book catalog, including Maker devices such as EPL’s 3D printers, and other media like music CDs and movies, we can provide a more comprehensive and versatile recommendation service. This will not only enhance the utility of the system but also ensure that all aspects of the library's offerings are equally accessible to patrons.

Furthermore, exploring additional machine learning models beyond KNN can help improve the accuracy and efficiency of the recommendations. Techniques such as collaborative filtering, content-based filtering, and hybrid models could offer new insights and potentially superior performance.

Another innovative future direction is the integration of social features into the recommender system. By connecting individuals based on their reading preferences and recommendations, we can facilitate the formation of small online book clubs. These virtual communities would provide a platform for patrons to engage in meaningful discussions, share their thoughts on books, and foster a sense of camaraderie centered around their love for reading.

In addition to benefiting individual library patrons, our recommender system has the potential to significantly support the Maker movement. By suggesting relevant resources such as books, tutorials, and guides related to various maker activities like 3D printing, we can empower makers with the knowledge and tools they need to pursue their creative projects. This aligns with the library's mission to serve as a hub for learning, creativity, and innovation.

In conclusion, the book recommender system is more than just a tool for suggesting books, it is a gateway to a more connected, informed, and engaged community. By focusing on accessibility and empathy, we have laid the foundation for a system that not only meets the needs of today’s library patrons but also anticipates and adapts to the evolving demands of the future. The continued development and expansion of this system will undoubtedly enhance its impact, making the library a more inclusive and valuable resource for everyone.