



Evidence Summary

Machine-learning Recommender Systems Can Inform Collection Development Decisions

A Review of:

Xiao, J., & Gao, W. (2020). Connecting the dots: reader ratings, bibliographic data, and machine-learning algorithms for monograph selection. *The Serials Librarian*, 78(1-4), 117-122.
<https://doi.org/10.1080/0361526X.2020.1707599>

Reviewed by:

Kristy Hancock
Evidence Synthesis Coordinator
Maritime SPOR SUPPORT Unit
Halifax, Nova Scotia, Canada
Email: Kristy.Hancock@nshealth.ca

Received: 29 Feb. 2024

Accepted: 26 March 2024

© 2024 Hancock. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

DOI: 10.18438/eblip30521

Abstract

Objective – To illustrate how machine-learning book recommender systems can help librarians make collection development decisions.

Design – Data analysis of publicly available book sales rankings and reader ratings.

Setting – The internet.

Subjects – 192 New York Times hardcover fiction best seller titles from 2018, and 1,367 Goodreads ratings posted in 2018.

Methods – Data were collected using Application Programming Interfaces. The researchers retrieved weekly hardcover fiction best seller rankings published by the New York Times in 2018 in CSV file format. All 52 files, each containing bibliographic data for 15 hardcover fiction titles, were combined and duplicate titles removed, resulting in 192 unique best seller titles. The researchers retrieved reader

ratings of the 192 best seller titles from Goodreads. The ratings were limited to those posted in 2018 by the top Goodreads reviewers.

A Bayes estimator produced a list of the top ten highest rated New York Times best sellers. The researchers built the recommender system using Python and employed several content-based and collaborative filtering recommender techniques (e.g., cosine similarity, term frequency-inverse document frequency, and matrix factorization algorithms) to identify novels similar to the highest rated best sellers.

Main Results – Each recommender technique generated a different list of novels.

Conclusion – The main finding from this study is that recommender systems can simplify collection development for librarians and facilitate greater access to relevant library materials for users. Academic libraries can use the same recommender techniques employed in the study to identify titles similar to highly circulated monographs or frequently requested interlibrary loans. There are several limitations to using recommender systems in libraries, including privacy concerns when analyzing user behaviour data and potential biases in machine-learning algorithms.

Commentary

The study was assessed using a critical appraisal tool developed for library and information research (Glynn, 2006). Recommender systems use machine-learning algorithms to predict user choice and recommend items based on user characteristics and behaviour. In the study, the authors describe a method for building a recommender system and suggest that librarians can use the system to identify library materials that will be appealing to users.

In a literature review, the authors provide a brief overview of the use and impact of recommender systems in libraries and other settings. In e-commerce, recommender systems can improve sales and user experience. In libraries, catalogues with built-in recommendation features can improve collection usage and increase the discoverability of relevant library materials for users. Even though there is a dearth of evidence demonstrating the use and impact of recommender systems in a collection development context, there are several relevant studies that could have been included in the literature review. For example, a 2019 study compared the accuracy of different machine-learning models in making predictions about demand-driven acquisition e-book purchasing patterns in a university library (Walker & Jiang, 2019).

There is a lack of clarity around the study methods, but the larger issue is to do with the study design. The authors position their study as an opportunity to leverage library data such as circulation statistics, user characteristics, and borrowing patterns, but they omit library data from their analysis. Instead, they use a list of highest rated best seller novels as the basis for their recommender system, reasoning that "books that are more popular will have a higher probability of being preferred by other readers." That assumption is logical, but it also impacts the usefulness of the approach. Given the skills required to build a recommender system, librarians with collections responsibilities may be left wondering what the complex approach adds to their existing practices, especially if they already refer to existing collection development tools that reflect sales metrics or reader ratings when selecting materials for purchase.

While the added value of the described approach is unclear, the authors contribute to the small body of evidence exploring the use of recommender systems in a collection development context. They also draw attention to interesting datasets that librarians could use as part of collections analyses. Researchers can build on this study by integrating library data into the analysis, to demonstrate an evidence-based approach to developing library collections. For example, librarians with programming

skills could start with a list of their own library's most circulated titles and build a recommender system using the same content-based and collaborative filtering techniques from the study.

References

- Glynn, L. (2006). A critical appraisal tool for library and information research. *Library Hi Tech*, 24(3), 387–399. <https://doi.org/10.1108/07378830610692154>
- Walker, K. W., & Jiang, Z. (2019). Application of adaptive boosting (AdaBoost) in demand-driven acquisition (DDA) prediction: A machine-learning approach. *The Journal of Academic Librarianship*, 45(3), 203–212. <https://doi.org/10.1016/j.acalib.2019.02.013>
- Xiao, J., & Gao, W. (2020). Connecting the dots: Reader ratings, bibliographic data, and machine-learning algorithms for monograph selection. *The Serials Librarian*, 78(1–4), 117–122. <https://doi.org/10.1080/0361526X.2020.1707599>