Anne Sawyer
KM 60370-001
Final Assignment
2 May 2018

MODELING A SEMANTIC SOLUTION:

A Hybrid Recommender System for Semantic Library Services

Books, Music, and Movies: Library Recommendation Services

People come to the Cincinnati Public Library to fulfill their information needs, whether they be related to academic research, accessing the internet, attending story times and other events, or finding something new to read, watch, or listen to. For the latter, patrons currently rely on library staff to make recommendations in-person, over the phone, and online, or they may use the many e-resources provided via the Library's website. While helpful, online databases, curated reading lists, and outside websites often require patient exploration and cross-referencing with the library's OPAC in order to get one's hands on a really good reading recommendation. Alternatively, customers may opt to receive personalized reading recommendations from human library professionals through the Library's *Book Hookup* service or be automatically placed on hold for copies of the latest popular books by signing up for *Hot Authors*. The Library also offers similar services for music and movies (*CD of the Month Club* and *Hot Tickets*, respectively).

These services provide a unique way to reach our customers and provide the personalized, virtual reader's advisory assistance they desire. However, the efficacy of highly-personalized services like *Book Hookup* and *CD of the Month Club* is often undermined by the limited availability of library staff to meet the daily demand of new requests, which are often time-consuming to process and may only be fulfilled during normal working hours. Due to these and other constraints placed on human recommenders, many customers fail to receive a new CD every month as advertised, or may not receive book recommendations for up to 48-72 hours after submitting their request. *Hot Authors* and *Hot Tickets* face a different set of constraints. Because these services require customers to select from among a predefined set of criteria, fulfilling their requests is much simpler. A member of ILS staff places blanket-level holds on new titles for customers based on their selections of movie genres or authors, which customers may pick up from any library location upon delivery.

Unfortunately, a sizable number of these *Hot Author/Ticket* holds end up expiring on the hold shelf, at which point they are re-routed to fill the hold of the next person in line, or they are re-shelved in the main collection. At first blush, this doesn't seem like such a big deal, but the Library makes purchasing decisions based on an optimum ratio of holds to copies of these items, so it is crucial that we are able to accurately predict demand for popular titles. If enough people place holds but don't actually check items out, then libraries end up with a glut of James Patterson, Stuart Woods, and Danielle Steel books that gather dust in the stacks. *Hot Authors/Tickets* have the potential to produce less waste and provide more value to library users if they could be redesigned to make decisions about which items to recommend to users and which ones to filter out, based on data about customers' actual use.

Hybrid Recommender Engines: a Semantic Solution

Recommender engines attempt to predict which items or products will be of interest to consumers, using a variety of filtering techniques to make recommendations. Collaborative filtering analyzes demographic information and community-generated ratings of items and builds a profile for a given user based on her similarity to other users. Recommendations are made using a neighborhood approach that assumes that a given user with similar preferences to other users will rate items similarly. In contrast, content-based filtering attempts to match a user to new items based on the similarity of their properties and by comparing them with other items the user has rated in the past. This approach relies on robust knowledge engineering to accurately classify new recommendations. Hybrid recommenders combine these and other approaches to produce better results than may be achieved by any one method alone, which makes them an obvious choice for an automated reader's advisory tool for libraries.

The current limitations of human reader's advisory services may be enhanced with the implementation of a recommender system that could learn user interests and preferences, using an optimized set of machine learning methods to automate the recommendation process and help library staff provide more standardized and reliable service. A feature augmentation approach that leverages a content-boosted collaborative filter would be ideal for an organization like the Cincinnati Public Library, which has a vast knowledge base from which user preferences may be inferred, but a relative sparsity of explicit ratings of items by users. A hybrid approach that uses the features of a contributing recommender as input to the actual recommender would help offset the common cold-start problem as well as information sparsity related to less popular items and new cardholders by virtue of the complementary strengths of different filters.

Costs, Risks, and Benefits of a Semantic Recommender

COSTS

In addition to the cost of the technology itself, implementation of a hybrid recommender system would necessarily need to coincide with updates to the Library's data architecture, as this technology would rely on linked data across multiple sources to operate as intended. This will mean a larger up-front investment to map KOS structures to each other as well as restructuring data sources to be meet RDF standards for enhanced interoperability, extensibility, and flexibility of our knowledge architecture. Bandwidth and data storage requirements as well as potential disruptions to current workflows will also need to be accounted for before moving forward. Assuming that the Library chooses to purchase rather than build the necessary technology for a recommender system means that librarians and IT staff will need to be trained in the product's use and maintenance, which will be part on the recurring costs associated with the project. Finally, there may be opportunity costs associated with making the system scalable for application in new services and business processes. If the Library intends to meet its goals for innovation and maintain its reputation as a good steward of taxpayer funds, it will need to prepare its enterprise architecture for a semantic future.

RISKS & CHALLENGES

Both collaborative and content-based recommenders are at risk for problems associated with scalability, cold start, and sparsity. While a hybrid feature augmentation structure may help to mitigate these risks, there are several others that will need to be assessed for project viability:

- The effectiveness of the recommender will have a positive correlation with the density of the Library's user profiles, so it will not work as well for users who don't log into the system with their library accounts. However, unregistered guests may still enjoy the benefits of enhanced search and information retrieval during their in-house use of library computers and/or library e-resources.
- A user's reading history plays a very important role in their user profile, which the recommender would use to generate more accurate and personalized results. Currently, the Library does not automatically track the reading histories of its patrons. Many users want and find value in this option, but they don't realize it's available. The Library will need to decide whether to change their data collection policy to one that is "opt in" by default.
- In an information economy that is increasingly data-driven, the Library will need to be able to respond to customers' concerns about their privacy and the Library's data collection practices. Constraints on the recommender's ability to use customer data will have a direct and salient influence on its effectiveness. Users may also need to agree to a new privacy policy.
- There may be unforeseen risks in overhauling the Library's current technology and data architectures for open data and enhanced ontological capabilities. Web-based systems will need to be carefully analyzed to determine the level of interoperability that is possible within

the constraints of our technology budget. This will of course impact the recommender system, as well as the adoption of any other semantic technology.

- The quality of reading recommendations may be confounded by multiple customers using the same account, e.g., kids using grandma's card.
- Finally, there is risk associated with disturbing customers by pushing recommendations or interrupting their research. The system will need to be designed to be relatively unobtrusive so that customers will be delighted by the recommender rather than annoyed by it.

BENEFITS & CAPABILITIES

A recommender system would have many capabilities for improved reader's advisory library services, and a hybrid system in particular would enhance the quality of recommendations by virtue of its complementary components. The collaborative component would be able to exploit behavioral similarities among users, while the content-based component would be able to compare item properties with user preferences to match the reader to her book and the book to its reader. In addition to recreational needs, the recommender could also potentially be trained to bundle research materials from multiple sources and in multiple formats for patrons looking to expand their knowledge about a particular topic. Users could also critique recommendations in real time, which would serve the dual purpose of increasing the accuracy of the recommender while giving patrons more control over the preferences that feed into their user profiles.

The adoption of a semantic recommender system would help the Library achieve its organizational goals and sustain its current high level of customer service to the community. Specifically, the recommender will:

- improve customer experience by enhancing access to information and materials.
- improve staff ability to assist customers with speedier location of materials and processing of reference queries.
- empower the Library to make better, data-driven business decisions through keener knowledge and insight into customer behavior and needs.

- help meet the Library's strategic goals of meeting customers at their point of need and being an innovative and dynamic force in the community.

Online forms & Reference Interviews: How Humans Recommend

IN-PERSON READER'S ADVISORY

For live help, library staff will perform a standard reference interview to ascertain the customer's preferences, tastes, and specific needs. Once this information has been clarified and verified, staff will find available items on the shelf, or place holds for the customer as necessary. The quality of interaction with the customer will be a primary determinant in continued use of library services. The human connection here is all-important: recommendations are conducted relatively quickly, so they may not be as precise, but the level of care and consideration given by staff will outweigh this in the final outcome of user satisfaction. In this context, staff may also directly influence customer interest and participation in related library services, activities, and events.

BOOK HOOKUP & CD OF THE MONTH CLUB

Book Hookup is an on-demand service that customers may use at their pleasure to receive three books in any format. CD of the Month Club is a recurring service that customers sign up for once, and receive a music-CD from the Library's collection once a month until they request to be removed from the service. Both services are highly-personalized and delivered based on user-submitted preferences. Initial requests are received via a web-based form, which library staff use to build their recommendations. Staff supplement their own domain knowledge with a variety of resources to provide high-quality recommendations to customers. Depending on the complexity of the request, staff might individually search various web resources such as Goodreads, Gnooks, Novelist, Whichbook, et al., or ask knowledgeable coworkers for guidance. Ideas for new items are then cross-referenced with the catalog and the request form for availability, accuracy, and appropriateness. Final recommendations are input into a separate form on the Intranet, then transmitted via email to the customer.

HOT AUTHORS & HOT TICKETS

These services are more automated forms of recommendation. Customers use the Library's website to sign up for one or both services, indicating their preferences for movie genres in the case of *Hot Tickets*, or indicate the names of authors and format preferences in the case of *Hot Authors*. From there, holds are placed automatically on a recurring basis as new titles are released. The criteria here is obviously much simpler than that of more personalized services, and little interaction between staff and customers is required for recommending. The downside is that customers may be more likely to reject recommendations, or simply not bother to pick up holds for authors that they may no longer be interested in. Since delivery of materials is the most expensive service provided by the Library, this is obviously problematic.

Inputs, Processes, & Outputs: How the Recommender System Would Work

Data inputs to the Library's hybrid recommender would come from a variety of explicit and implicit access points that can be used to provide a denser user profile and content-based filtering. The existing knowledge domain will act as a data source to combat cold-start problems associated with new users and more obscure items in the catalog. Specific data inputs include:

- ILS: provides basic user information as well as items currently checked out and placed on hold.
- Public website: holds additional user account data such as reading history, preferred searches, reading lists, and electronic checkouts (ebooks, eaudio).
- Downloadable providers: include details about movies and music streamed.

- Services users have signed up for: *Book Hookup*, *CD of the Month Club*, *Hot Authors*, *Hot Tickets*.
- Demco Event tool: tracks user data in the form of program attendance and signups.
- OPAC: can track browsing history, click paths, etc.
- Catalog records: include detailed technical and descriptive metadata for every item in the Library's collection, physical or digital.
- Databases and other outside resources: Novelist and Goodreads could provide a broader knowledge base from which to link recommendations, such as community ratings, reviews, and read-alike lists.

The content-boosted collaborative recommender uses a feature augmentation method involving two components. At a high-level, the content-based secondary recommender is used to compute a set of features that are then used to support the primary collaborative recommender. Like all recommendation systems, feature augmentation is first trained on a set of existing users and items. Once the recommender has been trained, new data is introduced to which the content-based filter contributes its own features. Bayesian (content-based) algorithms are used to generate new ratings for previously unrated items using the recommendation logic of the Library's existing knowledge domain. The collaborative filter then uses these classifications in its K-nearest neighbor algorithms to calculate similarity between user profiles. The resultant candidates are then scored by the system to produce its final recommendations, which may then be refined via user-contributed ratings before the learning process starts anew.

As output for general reader's advisory and on-demand personalized services, the semantic recommender would produce a mosaic of recommended items and/or resources on the Library's website that the user could select to place a hold, view more information, or rate the item to reinforce their profile for future recommending. If a user rates an item positively, they may receive other similar recommendations; if rated negatively, the system will learn this and recast its recommendations. For services like *Book Hookup*, human reference experts can use the recommender to quickly generate a list of titles for customers from which they may place holds for the user, use as inspiration to find other titles, or simply reject in favor of completely different titles.

To further customize automated services like *Hot Authors* and *Hot Tickets*, customers would sign up for the service as normal, indicating their prefered genres or authors. By analyzing customer behavior over time, the recommender could generate output in the form of predictions for new arrivals that the customer is more or less likely to actually check out. ILS staff could use this information to only place on hold items above a certain threshold of probability. For example, a 65-year old grandmother who watches family comedies and Hallmark romances will probably not be interested in *Office Christmas Party*, even though she may have selected Comedy as a genre that she generally enjoys.

Model Evaluation

TRUST

The level of trust experienced by customers with the recommender (and the Library generally) is a performance indicator that will be influenced by how well the Library explains the service to its users and clarifies how their information is used. Also, users are more likely to trust recommendations made by other people than those made by machines. Humanizing the recommender system through use of natural language and names/faces of library staff may encourage interaction with the recommender.

USER EXPERIENCE

Patron satisfaction will be the primary determinant for successful recommendation, of which back-end precision and front-end usability will play equal parts. For precision, success may be measured explicitly, through circulation of materials, downloads and use of streaming services, and recommendation service registration, or implicitly, by analyzing click-paths, labelling, browsing behavior, and searches performed. Levels of user satisfaction may be measured by supplementing self-reporting via periodic surveys with user research performed "in the wild." Studying patrons in the environment of their use (i.e., at a library branch) would be a practical way to evaluate the effectiveness of front-end recommender interface and performance.

SERENDIPITY & DIVERSITY

Machine learning techniques may be enhanced by combining collection metadata with outside data sources that provide more detailed genre information and keywords from album/book reviews. The ability of the recommender to generate pleasant surprises is especially important for *CD of the Month Club* and *Book Hookup* recommendations. Part of the success of the system will be measured by how well it can recommend items that are outside the user's typical repertoire but are nonetheless welcome additions. This may be seen in increased circulation of more obscure items.

Conclusion

Adopting a hybrid recommender system will facilitate the evolution of the Cincinnati Public Library from a well-run information organization to a more optimized knowledge organization that is prepared for a semantic future. A system-wide shift from data silos towards an interoperable knowledge architecture will be costly in the short term, but will result in long-term cost savings in the form of greater efficiency, adaptability, and competitiveness in an information economy. Customers will experience improved service turn-around and increased access to relevant information, while staff will be better able to manage their time by serving as editors of the recommender's work. By combining content-based and collaborative techniques in a feature augmentation process, the Library's recommender system will be able to contend with scalability, cold-start, and sparsity challenges. Continuous machine learning of users and their preferences will result in the delivery of better recommendations.

References

Burke, R. (2007). Hybrid Web recommender systems. In *The adaptive web* (pp. 377-408). Springer Berlin Heidelberg. PDF retrieved from http://www.dcs.warwick.ac.uk/~acristea/courses/CS411/2010/Book%20-%20The%20Adaptive%20Web/HybridWebRecommenderSystems.pdf

Dolan, S. (2018). Modeling a solution. PDF of lecture slides retrieved from Blackboard Learn https://learn.kent.edu/bbcswebdav/pid-7781589-dt-content-rid-69604720_1/courses/15370.201810/4A_modeling_solution%281%29.pdf

Dolan, S. (2018). Recommender systems. PDF of lecture slides retrieved from Blackboard Learn https://learn.kent.edu/bbcswebdav/pid-7781596-dt-content-rid-69604702_1/courses/15370.201810/15_recommender_systems%281%29.pdf

Konstan, J. A., & Riedl, J. (2012). Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22 (1-2), 101-123. Retrieved from *http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.371.8751&rep=rep1&type=pdf*

Melville, P., & Sindhwani, V. (2011). "Recommender systems". In *Encyclopedia of machine learning* (pp. 829-838). Springer US. PDF retrieved from http://www.prem-melville.com/publications/recommender-systems-eml2010.pdf

Zhai, C., & Massung, S. (2016). *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. [Yuzu ebook version]. Retrieved from https://reader.yuzu.com/#/books/9781970001174/cfi/241!/4/2@100:0.00