

Book Recommender Prototype based on Author's Writing Style

Paula Cristina Vaz
INESC-ID/IST
Rual Alves Redol, 9
1000029 Lisbon, Portugal
paula.vaz@inesc-id.pt

Ricardo Ribeiro
ISCTE-IUL/INESC-ID
Rual Alves Redol, 9
1000029 Lisbon, Portugal
ricardo.ribeiro@inesc-id.pt

David Martins de Matos
INESC-ID/IST
Rual Alves Redol, 9
1000029 Lisbon, Portugal
david.matos@inesc-id.pt

ABSTRACT

In this work we present a prototype that calculates similarity between authors and books. The application allows the user to choose the author and the system retrieves books from other authors that are similar to the chosen author's work.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Digital Libraries; H.3.7 [Digital Libraries]: System issues

General Terms

Algorithms, Recommendation

Keywords

Author recommendation, Book recommendation, Content-book recommendation

1. INTRODUCTION

Literary reading is an important and rewarding activity for people who engage in it. Very often readers choose books by their author. Readers that have a good experience reading a book from an author, most likely will try to read more books from the same author. They like to connect and know the author through her work.

Readers choose books by its author because they like the subject. This is the case, for example, of collections, where several books with same characters, but different stories, are available, e.g., "Harry Potter" by J. K. Rowling or "The Famous Five" by Enid Blyton. Readers may also choose books from the same author, because they like her writing style.

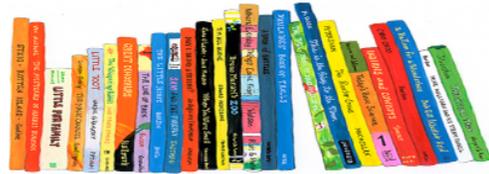
When readers choose books from the same author, because they had a good experience reading books from this author, they develop certain expectations. If these expectations are not met, readers often became disappointed with the choice made.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

OAIR '13, May 22-24, 2013, Lisbon, Portugal.
Copyright 2013 CID 978-2-905450-09-8.

To choose new books, readers typically go to libraries or bookstores, both real and virtual, and spend considerable amounts of time browsing shelves in search of something to read next. Both libraries and bookstores, even small ones, have enough books to make the selection time-consuming. In this scenario, book recommendation systems [4, 9, 10] are useful to help readers find what they want.

Two main techniques are used to develop recommendation systems [1]. Content-based techniques [5] in which users will be recommended items similar to those the user liked in the past; and collaborative filtering techniques [8, 6] in which users will be recommended items that were preferred together.



Similar Book Recommender

Books similar in subject and style to books by J. K. Rowling:

A Christmas Carol by Charles Dickens
The Lion, the Witch and the Wardrobe by C. Lewis
Prince Caspian by C. Lewis

Search for books similar to

Figure 1: Book recommender initial page.

Previous works [3, 2, 10] have shown that author ranking can improve book recommendation quality. In this work we present a hybrid content-based recommender prototype (Figure 1) that allows readers to retrieve books that are similar to their favorite authors. The prototype allows the user to choose an author that is in the database and retrieves books from other authors that are similar in style and topic to the chosen author's work.

2. THE BOOK RECOMMENDER

The book recommender prototype is a web-based application. Figure 1 shows our prototype. The prototype presents a list of authors to the user. The user chooses an author and the type of similarity: subject, style, or subject and style. When the user clicks on the “Search button”, the recommender searches for books that match the criteria and presents a top of the 20 most similar books to the user.

3. OUR ALGORITHM

The algorithm implemented in this work is based on the k NN item-based collaborative filtering algorithm proposed by [6]. Our prototype first calculates a $book \times book$ similarity matrix, where $cell_{i,j}$ contains the similarity between $book_i$ and $book_j$. Similarity is calculated using the Pearson correlation in Equation 1, where B_1 and B_2 are the vectors representing books.

$$sim(B_1, B_2) = \frac{\sum_{i=1}^n (B_{1i} - \overline{B_1})(B_{2i} - \overline{B_2})}{\sqrt{\sum_{i=1}^n (B_{1i} - \overline{B_1})^2} \sqrt{\sum_{i=1}^n (B_{2i} - \overline{B_2})^2}} \quad (1)$$

Authors are represented as classes of books that they have written. For each book that was not written by the author, the algorithm ranks the book calculating how close the book is to the author using Equation 2, where a is the author for whom we are calculating the top-20; b_i is the book not written by author a that we are ranking; A is author a book set; and b_{aj} is a book written by author a .

$$rank_{a,b_i} = 20 \times \frac{\sum_{b_{aj} \in A} sim(b_i, b_{aj})}{||A||} \bigwedge b_i \notin A \quad (2)$$

After calculating the rank for all books not written by the author, the algorithm returns the top-20 ranked books for which the rank is greater than 1.

Books are represented as feature vectors. The similarity *in subject* is calculated using vectors of content-word frequencies and the similarity *in style* is calculated using vectors of stylometric features [7]. For this work we used vectors of stop-word frequencies. The final rank is calculated using the proper similarity matrix.

To find books both similar *in subject and style*, first, ranks are calculated separately and, then, the algorithm averages both ranks, as shown in Equation 3.

$$rank_{b_i} = \frac{rankInSubject(b_i) + rankInStyle(b_i)}{2} \quad (3)$$

4. CONCLUSION & FUTURE WORK

In this work, we have presented a content-based book recommendation prototype that searches for books similar to books written by a given author. The prototype represents books as feature vectors and uses content-word frequencies to retrieve books with similar subjects. To retrieve books with similar writing styles, the prototype uses stylometric features. Stylometric features present a promising ground in the book recommendation field, because they are subject-independent and allow the retrieval of books that are more similar to the type of books the user likes to read. This is an important feature in literary book recommendation. This work is part of our ongoing study of the use of these features in book recommendation.

5. REFERENCES

- [1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749, 2005.
- [2] S. D. Gollapalli, P. Mitra, and C. L. Giles. Ranking authors in digital libraries. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, JCDL '11, pages 251–254, New York, NY, USA, 2011. ACM.
- [3] J. Kamps. The impact of author ranking in a library catalogue. In *Proceedings of the 4th ACM workshop on Online books, complementary social media and crowdsourcing*, BooksOnline '11, pages 35–40, New York, NY, USA, 2011. ACM.
- [4] T. Kuroiwa and S. Bhalla. Book recommendation system for utilisation of library services. *Int. J. Comput. Sci. Eng.*, 5(3/4):207–213, Dec. 2010.
- [5] M. J. Pazzani and D. Billsus. Content-based recommendation systems. In *The Adaptive Web*, pages 325–341, 2007.
- [6] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, WWW '01, pages 285–295, New York, NY, USA, 2001. ACM.
- [7] E. Stamatatos. A survey of modern authorship attribution methods. *J. Am. Soc. Inf. Sci. Technol.*, 60(3):538–556, Mar. 2009.
- [8] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Adv. in Artif. Intell.*, 2009:4:2–4:2, Jan. 2009.
- [9] P. C. Vaz, D. Martins de Matos, and B. Martins. Stylometric relevance-feedback towards a hybrid book recommendation algorithm. In *Proceedings of the fifth ACM workshop on Research advances in large digital book repositories and complementary media*, BooksOnline '12, pages 13–16, New York, NY, USA, 2012. ACM.
- [10] P. C. Vaz, D. Martins de Matos, B. Martins, and P. Calado. Improving a hybrid literary book recommendation system through author ranking. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*, JCDL '12, pages 387–388, New York, NY, USA, 2012. ACM.