

Use of Library Loan Records for Book Recommendation

Keita Tsuji[†], Erika Kuroo[‡], Sho Sato[†], Ui Ikeuchi[†], Atsushi Ikeuchi[†], Fuyuki Yoshikane[†]
and Hiroshi Itsumura[†]

[†] Graduate School of Library, Information and Media Studies, University of Tsukuba, 1-2 Kasuga, Tsukuba-city, Ibaraki-ken 305-8550, Japan.

{keitalmin2flyloulatsushilfuyukihits}(at)slis.tsukuba.ac.jp

[‡] College of Knowledge and Library Sciences, School of Informatics, University of Tsukuba, 1-2 Kasuga, Tsukuba-city, Ibaraki-ken 305-8550, Japan.
s0811586(at)u.tsukuba.ac.jp

Abstract: To show the effectiveness or limitation of using library loan records for book recommendation, we implemented the collaborative filtering system (henceforth LLR system) which is similar to that of Harada & Masuda (2010) and compared its performance with Amazon. It was found that LLR system could not outperform Amazon. Library loan records contain users' privacy and collaborative filtering usually requires much computer resources. We should pay attention to whether LLR system is really effective enough to worth its cost.

Keywords: Library loan records, Book recommendation, Collaborative filtering, Amazon, privacy.

I. INTRODUCTION

Library loan records are attracting people's attentions who are trying to produce new library services. Among them is the book recommendation system using collaborative filtering based on loan records (henceforth LLR system). However, library loan records contain users' privacy and collaborative filtering usually requires much computer resources. Is it really effective enough to worth its cost? Against this background, we implemented LLR system which was similar to that of Harada & Masuda (2010) and compared its performance with Amazon. If the latter outperformed the former, we might be able to say that Amazon is an inexpensive alternative for book recommendation system. The loan records we used for our experiment were those of one Japanese university library.

II. METHODS

A. Recommendation Algorithm

In this paper, we call the user to whom we would like to recommend books as "active user". Similarly, we call users whose library loan records are used for recommendation as "sample users".

In brief, our system calculates the similarity of library loan records between active user and sample users and recommends books that several sample users who are similar to active user have borrowed (and active user has not borrowed). For instance, if a sample user exists who borrowed books {a, b, c, d, e} and the active user has borrowed books {a, b, c, d}, our system recommend book e to the active user. The detail is as

follows:

(a) Our system calculates the similarity score p_{ai} between active user a and sample user i as follows:

$$p_{ai} = \frac{\sum_{k \in Y_{ai}} (s_{ak} - \bar{s}'_a)(s_{ik} - \bar{s}'_i)}{\sqrt{\sum_{k \in Y_{ai}} (s_{ak} - \bar{s}'_a)^2} \sqrt{\sum_{k \in Y_{ai}} (s_{ik} - \bar{s}'_i)^2}} \quad (1)$$

where Y_{ai} represents a set of books that the active user a and sample user i evaluated in common (in fact, Y_{ai} represents all the books the library holds as we mention later). s_{ij} represents a score the user i gave to the book j . \bar{s}'_i represents the average score that the user i gave to the books.

(b) Then our system recommends book j whose score \hat{s}_{aj} is high:

$$\hat{s}_{aj} = \sum_{i \in X} p_{ai}(s_{ij} - \bar{s}'_i) \quad (2)$$

where X represents all the sample users. The score we defined is slightly different from that of Kamishima (2008) and Resnick et al. (1994). In order to adopt score definition by Harada & Masuda (2010), we regarded the books which were not borrowed were scored 0 by the user. Therefore, all books the library holds were regarded to be scored by both active user and sample users (otherwise, we cannot calculate p_{ai}).

According to Harada & Masuda (2010), the scores are defined as follows: (1) the score the sample user gave to book is either 0 (he/she has not borrowed that book) or 1 (borrowed that book). (2) the score the active user gave were determined based on the subject of the book. The books which have the most popular Japanese decimal classification number (henceforth NDC. It ranges from 0 to 9) among the set of books the user borrowed are given score 1. The second, third and fourth most popular books are given scores 0.5, 0.25, 0.125,..., respectively. The loan records are divided by month according to Harada & Masuda (2010). Therefore, for instance, sample user X's loan record in April 2009 and in May 2009 are regarded as different loan records of different sample users.

B. Library Loan Records

The library loan records of one Japanese university library were used for our experiment. Each record consists of user ID, book ID, the dates the book was

borrowed/returned, NDC of the book, etc. These records were collected from 2006 to 2010. The number of them is approximately two million.

C. Evaluation

Four students were asked to be subjects of our experiment. Two of them are Ph.D. course students (henceforth students A and B) and the others are undergraduates of fourth grade (students C and D). They all belong to the library and information science course. Student A and B borrowed 58 and 207 books (tokens) during the previously mentioned period. Students C and D borrowed 82 and 11 books, respectively.

The subjects were asked to indicate one book R which they were interested in and the library holds. Then:

- (1) We put it into LLR system. The system recommends six books based on R and loan records of each subject.
- (2) We put R into Amazon.co.jp and obtain six books that the Amazon recommends as "The customer who bought R also bought these six books". Then the books that the library holds among these six were chosen for the final recommendation.

These two sets of recommended books were mixed and shown to the subjects. They were asked to rate them 3, 2, 1, 0 and x as follows:

- 3: I already have that book (the best recommendation).
- 2: I am strongly interested in that book (second best).
- 1: I am interested in that book.
- 0: I am not interested in that book.
- x: I have no idea.

III. RESULTS & DISCUSSIONS

The rates given to each recommended book by subjects A, B, C and D are shown in Tabs. 1, 2, 3 and 4, respectively. We can see in Tab. 1, for instance, that graduate student A gave "1" to four books (i.e., "Libraries in Finland & Sweden", "Introduction to Cultural Informatics",...) which were recommended by LLR system and its ratio is 67% ($=4/6$). Likewise, she gave "3" to three books (i.e., "Orientalism Vol.1", "Orientalism Vol.2", ...) which were recommended by Amazon and its ratio is 60% ($=3/5$). By comparing the rates given to books which were recommended by LLR and Amazon, we can see that Amazon obtained relatively better scores than LLR and thus showed better performance. For instance, while undergraduate student D showed no interest in books recommended by LLR, she gave rates "2" and "1" to one and two books recommended by Amazon, respectively (in Tab. 4).

As we previously mentioned, undergraduate student D borrowed only 11 books from library (although she is in the fourth grade, i.e., she has been in university three years and a half). To such students, LLR system cannot

recommend adequate books. Unfortunately, we think she is not an exceptional case in Japan. While the difference in recommendation performance between LLR and Amazon to student B (who borrowed 207 books) is small, those to students A and B are not negligible.

The loan records in university libraries have inherent weakness compared to Amazon purchase records. Borrowing books in libraries is free of charge and some users carelessly borrow books which have no relation to each other and to their needs. That brings noise to recommendation system. Besides, students borrow various books together which are unrelated especially in the end of the semester to prepare term reports and term exams. That might bring another noise to recommendation.

IV. CONCLUSIONS

To show the effectiveness or limitation of using library loan records for book recommendation, we implemented the collaborative filtering system which is similar to that of Harada & Masuda (2010) and compared its performance with Amazon. As we were concerned about, the LLR system could not outperform Amazon. We do not claim that libraries should not adopt new technology but library loan records contain users' privacy and collaborative filtering usually requires much computer resources. We should pay attention to whether LLR system is really effective enough to worth its cost.

REFERENCES

- Harada, T., "The book recommendation system using library loan records," *Digital Libraries*, **36**, 22-31 (2009). (Text in Japanese)
- Harada, T. and K. Masuda, "A trial approach of weighting for library loan records for developing a book recommendation system," *Digital Libraries*, **38**, 54-66 (2010). (Text in Japanese)
- Hwang, S. Y. and S. M. Chuang, "Combining article content and Web usage for literature recommendation in digital libraries," *Online Information Review*, **28**(4), 260-272 (2004).
- Kamishima, T., "Algorithms for Recommender Systems (2)," *Journal of Japanese Society for Artificial Intelligence*, **23**(1), 89-103. (2008) (Text in Japanese)
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J., "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," *Proceedings of the Conference on Computer Supported Cooperative Work*, 175-186 (1994).
- Saura, T. and K. Tsuji, "Attitude Survey on Utilizing the Public Library Use Records," *Proceedings of the 57th Japan Society of Library and Information Science*, 1-4 (2009). (Text in Japanese)
- Whitney, C. and L. Schiff, "The Melvyl Recommender Project: Developing library recommendation services," *D-Lib Magazine*, **12**(12), (2006).

Books Recommended by Library Loan Record		
Rate	(%)	Title and Other Bibliographic Data
3	17	Incidents related to Intellectual Freedom in Libraries (「図書館の自由に関する事例集」日本図書館協会, 2008)
1	67	Libraries in Finland & Sweden (図書館計画施設研究所, 図書館流通センター, 1994)
		Introduction to Cultural Informatics (「文化情報学入門」村上征勝ら, 勉誠出版, 2006)
		Libraries in Norway & Sweden (図書館計画施設研究所, リブリオ出版, 1996)
		Cultural History of Mimeograph (「ガリ版文化史...」田村紀雄ら, 新宿書房, 1985)
0	17	Design of Interactive Video (「インタラクティブ...」Iuppaら, パーソナル..., 1994)

Books Recommended by Amazon		
Rate	(%)	Title and Other Bibliographic Data
3	60	Orientalism Vol.1 (「オリエンタリズム(上)」サイード, 平凡社, 1993)
		Orientalism Vol.2 (「オリエンタリズム(下)」サイード, 平凡社, 1993)
		Media Control: Democracy and International Society... (「メディア・コントロール...」チョムスキー, 集英社, 2003)
1	40	What is Palestine? (「パレスチナとは何か」サイード, 岩波書店, 1995)
		Said that Said Tells (「サイード自身が...」サイードら, 紀伊國屋書店, 2006)

Table 1. Books recommended to graduate student A

Books Recommended by Library Loan Record		
Rate	(%)	Title and Other Bibliographic Data
3	17	Resource Organization Practice (「資料組織演習」岡田靖ら, 樹村房, 2007)
1	17	Utilizing Namazu Full-Text Search Engine (「Namazuシステムの...」馬場肇, ソフトバンク..., 2003)
0	50	Skills for Information Retrieval (「情報検索のスキル...」三輪眞木子, 中央公論新社, 2003)
		Cognitive Psychology for Sentence Understanding (「文章理解についての...」邑本俊亮, 風間書房, 1998)
		Strategies of Discourse Comprehension (Teun A. van Dijk et al., Academic Press, 1983)
x	17	Symbolic System which Produces Meanings (「意味を生み出す記号...」加藤雅人, 世界思想社, 2005)

Books Recommended by Amazon		
Rate	(%)	Title and Other Bibliographic Data
3	20	Amusing Ruby (「たのしいRuby」高橋征義ら, ソフトバンク..., 2010)
1	20	Meta Programming Ruby (「メタプログラミングRuby」Perrottaら, アスキー..., 2010)
0	40	Developing Web Application by Ruby on Rails 2 (「Ruby on Rails 2」...」北野光一ら, 工学社, 2009)
		Agile Web Application Development by Rails (「Railsによるアジャイル...」Sam Rubyら, オーム社, 2009)
x	20	Ruby Handbook (「Ruby逆引きレシピ...」島田浩二ら, 翔泳社, 2009)

Table 2. Books recommended to graduate student B

Books Recommended by Library Loan Record		
Rate	(%)	Title and Other Bibliographic Data
3	17	Classifying Books (「本を分類する」緑川信之, 勁草書房, 1996)
1	33	Understanding and Utilizing Pathfinder, LCSH and Metadata (「パスファインダー...」鹿島みづきら, 紀伊國屋書店, 2005)
		Introduction to Book Classification (「図書分類法要説 改訂版」加藤宗厚, 理想社, 1950)
0	17	Approach to Subject Information (「主題情報へのアプローチ」丸山昭二郎ら, 雄山閣出版, 1990)
x	33	Song of Ariran: A Korean Communist in... (Nym Wales et al., Ramparts Press, 1941)
		Metadata and its applications in the digital library... (Jia Liu, Libraries Unlimited, 2007)
Books Recommended by Amazon		
Rate	(%)	Title and Other Bibliographic Data
2	50	Survey Tips of Librarians (「図書館のプロが伝える調査...」高田高史ら, 柏書房, 2009)
		Using Libraries 100%: Tips to Find Resources (「図書館を使い倒す...」千野信浩, 新潮社, 2005)
1	50	How to Get Ideas in Libraries (「図書館が教えてくれた発想法」高田高史, 柏書房, 2007)
		Ask Librarians! (「図書館に訊け!」井上真琴, 筑摩書房, 2004)

Table 3. Books recommended to undergraduate student C

Books Recommended by Library Loan Record		
Rate	(%)	Title and Other Bibliographic Data
0	100	School Management and School Library (「学校経営と学校図書館」古賀節子ら, 樹村房, 2002)
0		How to Read Science Books (「科学の本の読み方すすめ方」板倉聖宣ら, 仮説社, 1993)
0		Forty Years of School Librarian Pursueing Dream (「夢を追い続けた学校司書...」五十嵐絹子, 国土社, 2006)
0		Introduction to Linear Algebra (「線型代数入門」斎藤正彦, 東京大学出版会, 1966)
0		What the School Library should be (「学校図書館経営の在り方」尾原淳夫, 駸々堂, 1950)
0		Introduction to Curriculum Study (「カリキュラム研究入門 新版」安彦忠彦, 勁草書房, 1999)
Books Recommended by Amazon		
Rate	(%)	Title and Other Bibliographic Data
2	25	Perl Quick Reference (「Perlクイックリファレンス」Ellen Sieverら, オーム社, 2000)
1	50	CGI Programming (「CGIプログラミング」Scott Guelichら, オーム社, 2001)
		Perl Cook Book (「Perlクックブック」Tom Christiansenら, オーム社, 2004)
0	25	make make 改訂版」Andrew Oramら, オーム社, 1997)

Table 4. Books recommended to undergraduate student D