Exploring Choice Overload in Related-Article Recommendations in Digital Libraries

Felix Beierle[1], Akiko Aizawa[2], and Joeran Beel[2,3]

[1] Service-centric Networking
Technische Universität Berlin / Telekom Innovation Laboratories
Berlin, Germany
beierle@tu-berlin.de
[2] National Institute of Informatics (NII)
Digital Content and Media Sciences Research Division
Tokyo, Japan
{aizawa,beel}@nii.ac.jp
[3] Trinity College Dublin
School of Computer Science and Statistics
Intelligent Systems Discipline, Knowledge and Data Engineering Group
ADAPT Centre
Dublin, Ireland
joeran.beel@adaptcentre.ie

Abstract. We investigate the problem of choice overload – the difficulty of making a decision when faced with many options – when displaying related-article recommendations in digital libraries. So far, research regarding to how many items should be displayed has mostly been done in the fields of media recommendations and search engines. We analyze the number of recommendations in current digital libraries. When browsing fullscreen with a laptop or desktop PC, all display a fixed number of recommendations. 72% display three, four, or five recommendations, none display more than ten. We provide results from an empirical evaluation conducted with *GESIS*' digital library *Sowiport*, with recommendations delivered by recommendations-as-a-service provider *Mr. DLib*. We use click-through rate as a measure of recommendation effectiveness based on 3.4 million delivered recommendations. Our results show lower click-through rates for higher numbers of recommendations and twice as many clicked recommendations when displaying ten instead of one related-articles. Our results indicate that users might quickly feel overloaded by choice.

Keywords: recommendation, recommender system, recommendations as a service, digital library, choice overload

1   Introduction

More and more information is available online for academic researchers in digital libraries [18]. One way to deal with the flood of information is to utilize

recommender systems that filter information and recommend articles related to those ones a user liked previously or is currently reading. A major challenge in recommending a list of related articles is to decide how many related articles to recommend before a user becomes dissatisfied with the recommender system due to choice overload.

Developing Mr. DLib (Machine-readable Digital Library)[4] [6, 4], a recommendations-as-a-service (RaaS) provider, we currently deliver recommendations to the digital library Sowiport[5][12]. Soon, we will also deliver recommendations to JabRef[6][11]. Developing such a recommender system for digital libraries, currently, there are no information to be found about how many recommendations to deliver and display. Figure 1 shows an example of using Mr. DLib in Sowiport. While a recommender system can filter for the most relevant

Fig. 1. Screenshot of the Sowiport Digital Library showing related items on the left hand side.

content for the user, the displayed recommended items can still be overwhelming. Schwartz describes the issue as the "tyranny of choice" [19]: Confronted with too many options, participants in studies tend to not decide for any option.

---

[4] http://mr-dlib.org

[5] http://sowiport.gesis.org

[6] http://www.jabref.org

In this paper, we investigate the influence of the size of the recommendation set on choice overload in digital libraries. In order to do so, we:

1. Examine how many items other recommender systems in digital libraries recommend
2. Conduct an empirical evaluation to see how different numbers of recommendations affect the clicks on related-article recommendations[7]

2 Related Work

There have been several studies investigating choice overload with respect to consumer goods (for example [1, 10]). Based on the MovieLens dataset, Bollen et al. investigated the relationship between *item set variety*, *item set attractiveness*, *choice difficulty*, and *choice satisfaction* [8]. They suggest to diversify the recommendation set by including some lower quality recommended items in order to increase perceived recommendation variety and choice satisfaction. In another study, Willemsen et al. further analyzed the relationship of diversification, choice difficulty, and satisfaction [21]. Here, the authors also used the MovieLens dataset.

Other related studies looked into the number of search results to be displayed. Jones et al. conclude that the screen size is a determining factor with respect to how many search result items users interact with [14], which is confirmed in a newer study by Kim et al. [16]. Linden reports that although Google users claimed to want more search results, traffic dropped with the display of an increased number of search results [17]. Google suspected the extra loading time to be play a role in this. Azzopardi and Zuccon developed a cost model for browsing search results, taking into account screen size and search results page size [2]. They concluded that displaying 10 results is close to the minimum cost. Kelly and Azzopardi studied the effects of displaying different sizes of search result pages [15]. In their study, they used three, six, and ten search results. One of their main findings is that subjects of the study who were shown ten search results per page viewed and saved significantly more documents, while more time is spent on earlier search results, if the number of results per page is less. While Chiravirakul and Payne's study suggests that choice dissatisfaction happens when there is a lack of time for choosing links [9], the study by Oulasvirta et al. suggests that it is the search result page size that is causing choice overload or the "paradox of choice." Oulasvirta et al. conducted a user study with 24 participants and looked into the user's satisfaction with the results when displaying six or 24 results. For future work, they suggest also looking into objective behavior measurements like click-through rates.

In our work, we focus on a different domain. We will investigate the problem of choice overload in digital libraries – in contrast to movie recommendations

---

[7] All data relating to this paper is available on `http://datasets.mr-dlib.org`, including a table of the delivered and clicked recommendations, the information about the investigated digital libraries, and the figures presented in this paper.

and search results. The question of how many recommendations to display and choice overload has not been studied in the domain of digital libraries, to the best of our knowledge. In a recent literature survey on more than 200 articles about research paper recommender systems, no one discussed or researched this topic [5]. Furthermore, instead of user interviews or small scale studies, we investigate the real clicks logged in an actively used system. We will explore, to what extend click-through rates (CTR) reflect the findings of the cited studies.

3   Methodology

In order to investigate choice overload in digital libraries, we first examine how many items other recommender systems in digital libraries display. We investigated 63 digital libraries[8] and reference managers with search interfaces[9]. We considered recommendations that are being displayed when an item from the search results is selected. In a few cases, the number of displayed recommendations of related items was dependent on the size of the browser window. For the numbers given in the following, we assumed a full screen browser window on a laptop computer (13" display with 1280x800 resolution).

In a second step, we conducted an experimental evaluation to investigate how different numbers of recommendations affect the click rates on related-article recommendations. Click-through rates are a good way to study the users' actual behavior when displaying recommended items in real situations. We analyze data from 3.4 million recommendations. The data was obtained from users of the academic search engine Sowiport[10], which is run by *GESIS - Leibniz-Institute for the Social Sciences*[11], which is the largest infrastructure institution for the Social Sciences in Germany. Sowiport contains about 9.6 million literature references and 50,000 research projects from 18 different databases, mostly relating to the social and political sciences. Literature references usually cover keywords, classifications, author(s), and journal or conference information, and if available: citations, references, and links to full texts.

Sowiport co-operates with Mr. DLib, an open Web Service to provide scholarly literature-recommendations-as-a-service (Figure 2). This means that all computations relating to the recommendations run on Mr. DLibs servers, while the presentation takes place on Sowiports website. Our recommender system shows related-article recommendations on each articles detail page in Sowiport (see Figure 1). Whenever such a detail page is requested by a user, the recommender system randomly chooses one of four recommendation approaches to generate recommendations: 1. stereotype recommendations, 2. most popular recommendations, 3. content-based filtering (CBF), and 4. random recommendations. We measured the effectiveness of the recommendation approaches

---

[8] Most of them listed on `https://en.wikipedia.org/wiki/List_of_digital_library_projects`

[9] See `http://datasets.mr-dlib.org` for detailed results.

[10] Some explanations about Sowiport and Mr. DLib are from [3].
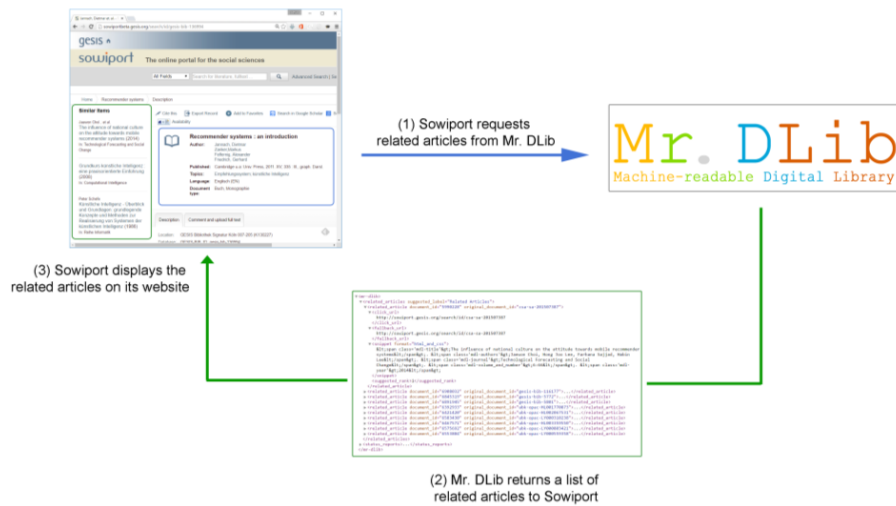
[11] `http://www.gesis.org`

Fig. 2. The recommendation process of Sowiport and Mr. DLib.

with click-through rate (CTR). CTR describes the ratio of delivered to clicked recommendations. For instance, when 1,000 recommendations were delivered, and 8.4 of these recommendations were clicked, the average CTR would be 8.4/1,000=0.84%. The assumption is that the higher the CTR, the more effective is the recommendation approach. There is some discussion to what extend CTR is appropriate for measuring recommendation effectiveness, but overall it has been demonstrated to be a meaningful and well-suited metric [13, 7, 20]. For our evaluation, we randomly displayed one to fifteen recommendations.

Our expectation is that at first, by increasing the number of displayed recommendations, there will be an increase in the CTR. By displaying more and more recommendations, we expect to reach a maximum in CTR at some point. After that, when displaying more recommendations, we expect the CTR to drop, indicating choice overload. Similarly, for the clicks, we would expect to find a maximum at a certain number of displayed recommendations. Plotting the CTR and the average clicks we show this expectation in Figure 3. Here, the CTR, the orange line, increases with the number of displayed recommendations, reaches a maximum at four, and decreases afterwards, indicating choice overload. The clicks, the gray line, reach a maximum at five displayed recommendations. In that case, we would decide for four (maximum CTR) or five (maximum clicks) recommendations.

Alternatively, we would have expected results as in Figure 4. Here, the CTR declines with increasing number of displayed recommendations. The absolute clicks have a maximum at four. The question then would be: What is better – a higher CTR or the maximum of absolute clicks? Depending on the answer, we would choose a recommendation set size of between one and four. As we will show in the following section, the results are quite different from our expectations.
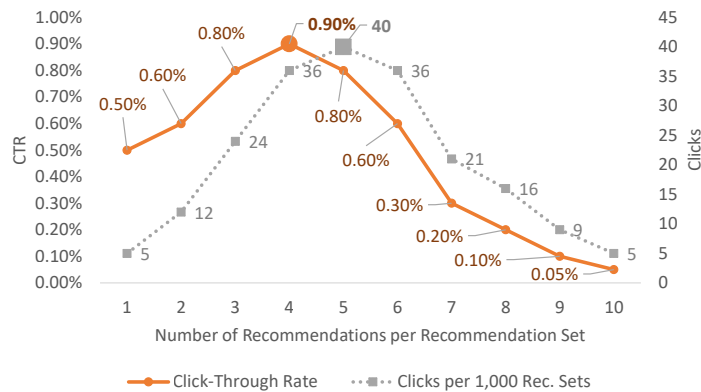
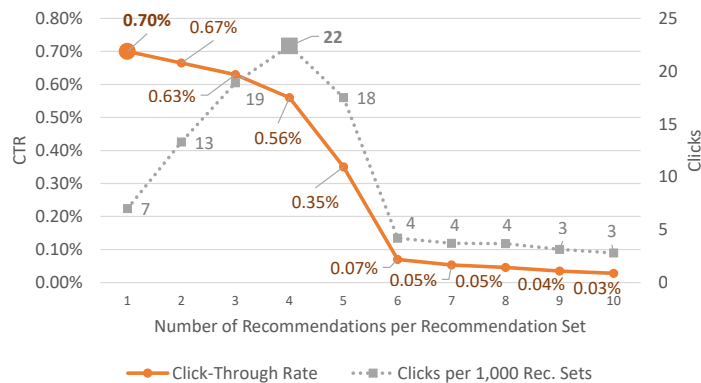**Fig. 3.** Expected click-through rate and clicks by displayed number of recommendations.

Fig. 4. Alternative expected click-through rate and clicks by displayed number of recommendations.

4 Results

4.1 Number of Recommendations in Existing Digital Libraries

19 (30%) of the 63 digital libraries displayed recommendations for related items. Figure 5 shows the distribution of recommendations for those 19 libraries. Most libraries (72% of those that display recommendations) display three, four, or five recommendations, none is displaying more than 10 or less than three. It is also notable that they all always show a fixed number of related-articles, one could also imagine displaying varying numbers. Looking for related items in the database, there can be varying scores of relevance. One way of displaying a varying amount of related-article recommendations would be to take into account such relevance scores, e.g., giving 10 recommendations if there are 10 highly
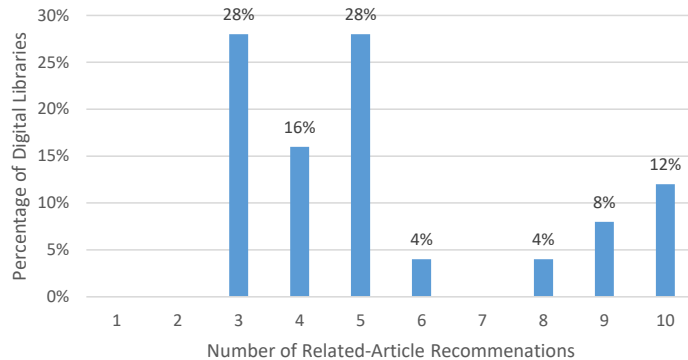
Fig. 5. Number of displayed recommendations in current digital libraries.

relevant related-articles, and giving only two recommendations, if there are only two related-articles that have a relevance score above a certain threshold.

We do not know how the numbers of related articles in the reviewed recommender systems were chosen by the operators. We assume that they either just arbitrarily chose the numbers, or did some experiments but did not publish the results. In the following, by measuring CTRs, we want to investigate and measure the effect the number of displayed recommendations has and if the CTR indicates choice overload for certain numbers of displayed recommendations.

4.2   Experiment with Varying Number of Recommendations

The solid orange line in Figure 6 shows the CTR by the number of displayed recommendations. The higher the number of recommendations, the lower the overall CTR is. The dashed line shows the average absolute number of clicked recommendations (per 1,000 recommendations). The bigger the recommendation set size is, the higher the number of absolute clicks is.

When only one recommendation was displayed, the CTR was 0.84% on average. This means, when our recommender system delivered 1,000 times one recommendation each, 8.4 recommendations were clicked. For two displayed recommendations, the CTR was only 0.49% on average. This means, when our recommender system delivered 1,000 times two recommendations each (2,000 in total), 9.8 recommendations were clicked (half clicked on the first one, half on the second one). Overall, if one or two recommendations are displayed, does not make a big difference in the absolute number of clicks, it only increases by 17% (8.4 to 9.8). When fifteen recommendations were shown, the CTR was at the minimum of 0.14% on average, while the absolute number of clicks was at the maximum of 21.4.

Comparing these results with our expectations given in Section 3, we can see that the CTR unexpectetly decreases rapidly and has a clear maximum at
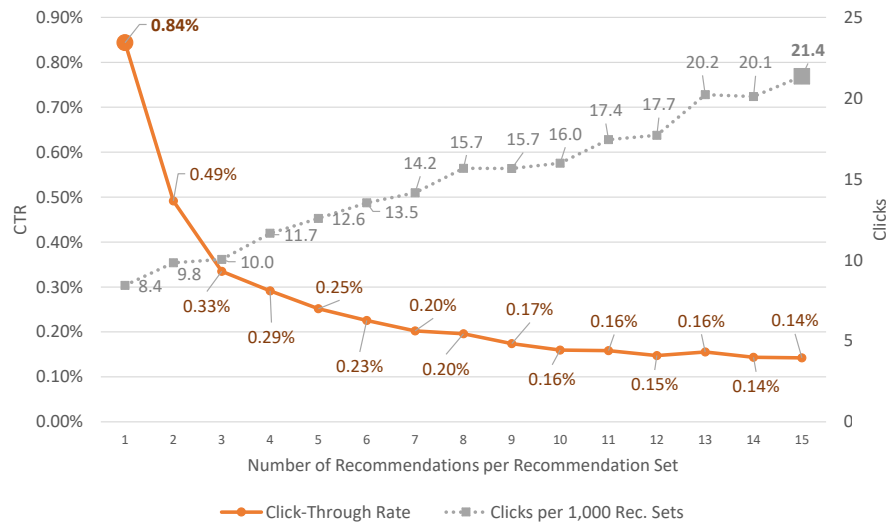
Fig. 6. Click-through rates (solid) and average absolute number of clicked recommendations (dashed) with respect to the number of displayed recommendations.

one displayed recommendation. Furthermore, contrary to our expectations, the absolute clicks keep increasing instead of having a maximum value at a few recommendations. The results show an under-proportional increase in average clicks on the displayed recommendations. Displaying twice as many recommendations does not double the clicks. In order for the absolute number of clicked recommendations (per 1,000) to double from 8.4 (for one displayed recommendation) to 17, the number of displayed related-articles has to be raised to 10 or 11. When 15 recommendations are displayed, only 2.5 as many recommendations are clicked compared to displaying a single recommendation. Regarding choice overload this implies that having more recommendations to choose from does, in general, only create a small incentive for the user to click on more of the displayed recommended items. There are some points to consider when interpreting these results. Many documents in Sowiport only have sparse information, thus they might not be interesting for the users. Another possible option why the experimental results are so unclear about the number of recommendations to give is that the relevance of the recommendations might have been too low, so that many users did not click further recommendations after clicking the first one. In that case, the research should be repeated when we are able to deliver better recommendations. The session-length is another aspect to consider. For instance, if one user visits two pages and on gets 15 recommendations on each page, we assume the CTR will be higher than for a user who looks at 10 pages and gets 15 recommendations on each page. An additional aspect of sessions is that so far, we did not filter for recommendations that have already been shown to a user. So, if a user looks at 15 detail pages and gets 10 related-article recom-

mendations on each, most likely there will be duplicate recommendations, and hence the CTR decreases the more recommendations are shown. Another aspect to consider in the results' interpretation is how users might use Sowiport. If the user clicks on a recommendation, a new tab is opened. If the recommendation was good, she might forget about the other open tab – especially if there are further good recommendations shown in the new tab.

5   Conclusion and Future Work

The average clicks on displayed recommendations under-proportionally increase with the number of displayed items. In order for the clicks to double, the size of the recommendation set has to be increased from one to 10 or 11. These numbers might imply that the users quickly feel overloaded by choice. The results differ from our expectations. Just based on these numbers, we could conclude that we should only display one recommendation because the CTR is highest – or to keep displaying even more than 15 recommendations until the number of absolute clicks does not increase anymore. Further research will be necessary to determine a good number of recommended items to display.

Our results are based on Sowiport. Further research is necessary to confirm if our findings also apply to other digital libraries. We therefore plan to repeat our research, for instance, with JabRef and the library of the Technical University of Munich. Future work could also include using other evaluation methods and metrics than CTR (e.g., a user study, or user ratings, or tracking which recommended item were actually exported or saved) or making a survey and asking the operators of the other digital libraries how they decided the number of displayed recommendations. Furthermore, a question to be discussed is what recommender systems in digital libraries should try to achieve, e.g., maximizing CTR, maximizing the number of clicked recommendations, etc.

References

1. Arunachalam, B., Henneberry, S.R., Lusk, J.L., Norwood, F.B.: An Empirical Investigation into the Excessive-Choice Effect. American Journal of Agricultural Economics 91(3), 810–825 (Aug 2009)

---

[12] http://www.dynamic-project.de

2. Azzopardi, L., Zuccon, G.: Two Scrolls or One Click: A Cost Model for Browsing Search Results. In: Advances in Information Retrieval. pp. 696–702. Springer, Cham (Mar 2016)
3. Beel, J., Dinesh, S., Mayr, P., Carevic, Z., Raghvendra, J.: Stereotype and Most-Popular Recommendations in the Digital Library Sowiport. In: Proceedings of the 15th International Symposium of Information Science (ISI). pp. 96–108. Verlag Werner Hülsbusch, Glückstadt, Germany (2017)
4. Beel, J., Gipp, B., Aizawa, A.: Mr. DLib: Recommendations-as-a-Service (RaaS) for Academia (Pre-print). In: Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL) (2017)
5. Beel, J., Gipp, B., Langer, S., Breitinger, C.: Research paper recommender systems: A literature survey. International Journal on Digital Libraries pp. 1–34 (2016)
6. Beel, J., Gipp, B., Langer, S., Genzmehr, M., Wilde, E., Nürnberger, A., Pitman, J.: Introducing Mr. DLib, a Machine-readable Digital Library. In: Proceedings of the 11th ACM/IEEE Joint Conference on Digital Libraries (JCDL‘11). pp. 463–464. ACM (2011), available at `http://docear.org`
7. Beel, J., Langer, S.: A Comparison of Offline Evaluations, Online Evaluations, and User Studies in the Context of Research-Paper Recommender Systems. In: Research and Advanced Technology for Digital Libraries. pp. 153–168. Springer, Cham (Sep 2015)
8. Bollen, D., Knijnenburg, B.P., Willemsen, M.C., Graus, M.: Understanding Choice Overload in Recommender Systems. In: Proceedings of the Fourth ACM Conference on Recommender Systems. pp. 63–70. RecSys ’10, ACM (2010)
9. Chiravirakul, P., Payne, S.J.: Choice Overload in Search Engine Use? In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems. pp. 1285–1294. CHI ’14, ACM (2014)
10. Fasolo, B., McClelland, G.H., Todd, P.M.: Escaping the tyranny of choice: When fewer attributes make choice easier. Marketing Theory 7(1), 13–26 (Mar 2007)
11. Feyer, S., Siebert, S., Gipp, B., Aizawa, A., Beel, J.: Integration of the Scientific Recommender System Mr. DLib into the Reference Manager JabRef. In: Proceedings of the 39th European Conference on Information Retrieval (ECIR) (2017)
12. Hienert, D., Sawitzki, F., Mayr, P.: Digital Library Research in Action: Supporting Information Retrieval in Sowiport. D-Lib Magazine 21(3/4) (Mar 2015)
13. Joachims, T., Granka, L., Pan, B., Hembrooke, H., Gay, G.: Accurately Interpreting Clickthrough Data As Implicit Feedback. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 154–161. SIGIR ’05, ACM (2005)
14. Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., Buchanan, G.: Improving Web interaction on small displays. Computer Networks 31(11–16), 1129–1137 (May 1999)
15. Kelly, D., Azzopardi, L.: How Many Results Per Page? A Study of SERP Size, Search Behavior and User Experience. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 183–192. SIGIR ’15, ACM (2015)
16. Kim, J., Thomas, P., Sankaranarayana, R., Gedeon, T., Yoon, H.J.: Eye-Tracking Analysis of User Behavior and Performance in Web Search on Large and Small Screens. Journal of the Association for Information Science and Technology 66(3), 526–544 (Mar 2015)
17. Linden, G.: Marissa Mayer at Web 2.0. `http://glinden.blogspot.com/2006/11/marissa-mayer-at-web-20.html` (Nov 2006)

18. Noorden, R.V.: Global scientific output doubles every nine years. Nature Blog, `http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html` (2014)
19. Schwartz, B.: The tyranny of choice. Scientific American 290(4), 70–75 (2004)
20. Schwarzer, M., Schubotz, M., Meuschke, N., Breitinger, C., Markl, V., Gipp, B.: Evaluating Link-based Recommendations for Wikipedia. In: 2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL). pp. 191–200 (Jun 2016)
21. Willemsen, M.C., Graus, M.P., Knijnenburg, B.P.: Understanding the role of latent feature diversification on choice difficulty and satisfaction. User Modeling and User-Adapted Interaction 26(4), 347–389 (Oct 2016)

Additional Information