International Journal on Digital Libraries - preprint

Choice Overload and Recommendation Effectiveness in Related-Article Recommendations

Analyzing the Sowiport Digital Library

Felix Beierle · Akiko Aizawa · Andrew Collins · Joeran Beel

Received: 16 March 2018 / Revised: 15 February 2019 / Accepted: 7 May 2019

This work has received funding from project DYNAMIC (http://www.dynamic-project.de) (grant No 01IS12056), which is funded as part of the Software Campus initiative by the German Federal Ministry of Education and Research (BMBF). This work was also supported by a fellowship within the FITweltweit programme of the German Academic Exchange Service (DAAD). This publication also has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number 13/RC/2106.

Felix Beierle (corresponding author)
Service-centric Networking
Telekom Innovation Laboratories
Technische Universität Berlin
Berlin, Germany
E-mail: beierle@tu-berlin.de
Telephone: +49 30 8353 54265

Akiko Aizawa
National Institute of Informatics (NII)
Digital Content and Media Sciences Research Division
Tokyo, Japan
E-mail: aizawa@nii.ac.jp

Andrew Collins
Trinity College Dublin
School of Computer Science and Statistics
ADAPT Centre
Dublin, Ireland
E-mail: andrew.collins@adaptcentre.ie

Joeran Beel
Trinity College Dublin
School of Computer Science and Statistics
ADAPT Centre
Dublin, Ireland
E-mail: beelj@tcd.ie
and
National Institute of Informatics (NII)
Digital Content and Media Sciences Research Division
Tokyo, Japan
E-mail: beel@nii.ac.jp

Abstract Choice Overload describes a situation in which a person has difficulty in making decisions due to too many options. We examine choice overload when displaying related-article recommendations in digital libraries, and examine the effectiveness of recommendation algorithms in this domain. We first analyzed existing digital libraries, and found that only 30% of digital libraries show related-article recommendations to their users. Of these libraries, the majority (74%) displays 3–5 related articles; 28% of them display 6–10 related articles; and no digital library displayed more than ten related-article recommendations. We then conducted our experimental evaluation through *GESIS*' digital library *Sowiport*, with recommendations delivered by recommendations-as-a-service provider *Mr. DLib*. We use four metrics to analyze 41.3 million delivered recommendations: click-through rate (CTR), percentage of clicked recommendation sets (clicked set rate, CSR), average clicks per clicked recommendation set (ACCS), and time to first click (TTFC), which is the time between delivery of a set of recommendations to the first click. These metrics help us to analyze choice overload and can yield evidence for finding the ideal number of recommendations to display. We found that with increasing recommendation set size, i.e., the numbers of displayed recommendations, CTR decreases from 0.41% for one recommendation to 0.09% for 15 recommendations. Most recommendation sets only receive one click. ACCS increases with set size, but increases more slowly for six recommendations and more. When displaying 15 recommendations, the average clicks per set is at a maximum (1.15). Similarly, TTFC increases with larger recommendation set size, but increases more slowly for sets of more than five recommendations. While CTR and CSR do not indicate choice overload, ACCS and TTFC point towards 5-6 recommendations as being optimal for *Sowiport*. Content-based filtering yields the highest CTR with 0.118%, while stereotype recommendations yield the highest ACCS (1.28). Stereotype recommendations also yield the highest TTFC. This means that users take more time before clicking stereotype recommendations when compared to recommendations based on other algorithms.

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

1 Introduction

An increasing number of research articles are available in digital libraries (Van Noorden 2014; Bornmann and Mutz 2015). This makes it more difficult for users to find articles relevant to their needs (Bawden and Robinson 2009). Recommender systems can help users by suggesting articles that are related to ones they liked previously or are currently reading.

One challenge of recommending related-articles is in deciding how many articles to display to users before they might become dissatisfied with the recommender system, due to possible choice overload. While a recommender system can choose the most relevant content for a user, the displayed recommendations may still be overwhelming. Schwartz describes this issue as "tyranny of choice" (2004): confronted with too many options, participants in some studies tend to not make any choice. Another challenge in the recommendation process is choosing an algorithm that is most effective, in order to deliver recommendations that are appropriate for users.

As developers of the recommender system Mr. DLib (Machine-readable Digital Library)[1] (Beel et al 2017; Beel et al 2011) we have faced the challenge of both choice overload and recommendation effectiveness. Mr. DLib is a recommendations-as-a-service (RaaS) that delivers related-article recommendations for academic papers. One partner of Mr. DLib is the digital library Sowiport[2] (Hienert et al 2015; Stempfhuber et al 2008). Figure 1 shows an example of using Mr. DLib in Sowiport. More details about the recommendation process will be given in Section 4.

In this paper, we examine choice overload and recommendation effectiveness in Sowiport. Our goal is to obtain insights about the usage and effectiveness of related-article recommendations in Sowiport, and more generally in digital libraries. The research questions we address are:

1. How many recommendations are typically displayed currently in digital libraries?
2. How many recommendations should be displayed?
3. How much does the recommendation algorithm influence a user's click behavior?

To answer these questions, we conducted empirical evaluations.

1. We examined 63 digital libraries to identify how many recommendations they display (research question 1).
2. We conducted an experimental study with Mr. DLib and Sowiport. In that study, we
   (a) randomly varied the recommendation set size, i.e., the number of recommendations to be displayed, to see how different numbers of recommendations influence user behavior (research question 2).
   (b) In addition, we varied the recommendation algorithm to see how this affects user behavior (research question 3).

This paper is an extended version of Beierle et al (2017). In this extension, we analyze a dataset 12.1 times larger than the dataset from the previous analysis. Furthermore, in addition to the click-through rate (CTR), we use three further metrics to assess choice overload and recommendation effectiveness: the percentage of clicked sets (clicked set rate, CSR), the average clicks per clicked recommendation set (ACCS), and the time between recommendation delivery and first click (time to first click, TTFC). Additionally, we investigate the effectiveness of different recommender algorithms.

In Section 2, we detail the results of our pre-study in which we investigate the number of recommendations displayed in existing digital libraries. In Section 3, we will give an overview of related work. After detailing the methodology in Section 4, we will give the results in Section 5, structured by applied metric, before concluding and describing future work in Section 6.

2 Pre-Study: Analysis of the Number of Recommendations in Digital Libraries

First, we examine how many recommendations other digital libraries display (research question 1). We investigated the websites of 63 digital libraries[3] and reference managers with search interfaces. We only consider related-article recommendations that are displayed on items' detail pages. In a few cases, the number of

[1] `http://mr-dlib.org`
[2] `http://sowiport.gesis.org`
[3] Most of them listed on `https://en.wikipedia.org/wiki/List_of_digital_library_projects`.

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

displayed recommendations depended on the size of the browser window. For our analysis, we used a full-screen browser on a laptop computer (13" display with 1280x800 resolution).

Only 19 (30%) of the 63 digital libraries have a recommender system that suggests related-articles to their users. Five of the 19 libraries (26%) displayed more than one recommendation set at a time on their website. This means that when viewing an article's detail page, more than one box with related-article recommendations is visible. For example, there may be one box next to the title of the selected document and one box at the bottom of the page.

Figure 2 shows the distribution of recommendation set sizes for the 25 displayed recommendation sets in those 19 libraries. 74% of libraries (14/19 of those that displayed recommendations) display three, four, or five recommendations; none display more than 10, or less than three.

3 Related Work

In a meta-study about choice overload, Scheibehenne et al. summarized that all definitions of choice overload have in common that it is about "the notion of adverse consequences due to an increase in the number of options to choose from" (Scheibehenne et al 2010). Those adverse consequences include a decrease in the motivation to choose or to make any choice at all.

Choice overload has been widely researched in the field of e-commerce (Moser et al 2017; Arunachalam et al 2009; Fasolo et al 2007; Haynes 2009; Park and Jang 2013; Reed et al 2011; Besedeš et al 2015; Dellaert et al 2017; Buturak and Evren 2017). A meta analysis of 63 studies found a mean effect size near zero (Scheibehenne et al 2010). However, variance among the studies was high, ranging from choice overload to more-is-better. Another meta study, looking at 99 individual studies, identifies four key factors that affect choice overload (Chernev et al 2015). These factors are choice set complexity, decision task difficulty, preference uncertainty, and decision goal.

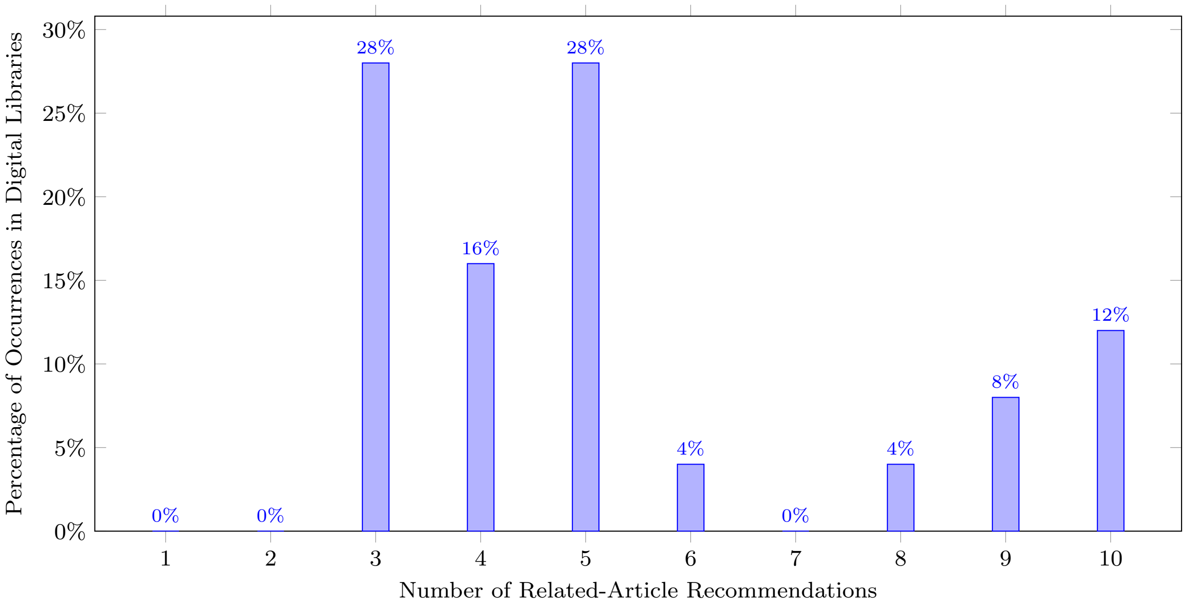

Fig. 2 Number of displayed recommendations in current digital libraries

In the context of recommender systems for digital libraries, there is no related work on choice overload, besides our own work (Beierle et al 2017), to the best of our knowledge. In a recent literature survey, we analyzed more than 200 research articles on recommender systems in digital libraries, and none of these articles dealt with choice overload (Beel et al 2016).

In the field of recommender systems, there is a significant amount of work on various aspects of user behavior in recommender systems (Jugovac and Jannach 2017; Zheng et al 2016; Knijnenburg et al 2012; Motajcsek et al 2016). However, with respect to choice overload, there is only little research. Bollen et al. (2010) conducted tests using the Movielens dataset. They experimented with recommendation set sizes of 5 and 20 and suspect the ideal number of recommendations to be between 7 and 10. Such a set size would have enough variety (compared to a set that is too small) while still being manageable (in contrast to a set size too big). Willemsen et al. (2016) analyzed the relationship between diversification, choice difficulty, and user satisfaction. They also used the MovieLens dataset and their findings suggest that a diverse, small set of recommendations – as few as five items – may be best to deliver to the users.

There is a notable amount of research on choice overload in the field of search engines. Jones et al. found that screen size is a determining factor with respect to how many search result items users interact with (Jones et al 1999). This is confirmed in a later study by Kim et al. (2015). Linden reports that although Google users claimed to want more search results, traffic dropped when increased numbers of search results were displayed (2006). Extra loading time was suspected to play a role in this decrease in traffic, rather than choice overload being predominant. Azzopardi and Zuccon developed a cost model for browsing search results, taking into account screen size and search results page size (2016). They concluded that displaying 10 results is close to the minimum cost. Kelly and Azzopardi studied the effects of displaying different numbers of search results on a page to users (2015). In their study, they used three, six, and ten search results. One of their main findings is that subjects who were shown ten search results per page viewed and saved significantly more documents, however reported a greater workload. Search results are more closely examined if the number of results per page is less. This may indicate choice overload by users when interacting with larger lists of results. Chiravirakul and Payne find that the strongest influencer of choice overload is the amount of time-pressure that users are under (2014). However, Oulasvirta et al. find that it is the search result page size that causes choice overload, even when users are under time-pressure (2009).

In our work, we focus on choice overload in recommender systems for digital libraries, and the question of whether or not choice overload exists, and how many recommendations to display.

4 Methodology

To answer the second and third research question, we conducted an experimental study. In this study, we randomly varied the numbers of displayed recommendations and the recommendation algorithms. The goal was to see how the different variations would affect the users' behavior, i.e., whether or not users clicked on recommendations and after what time. To the best of our knowledge, we are first to investigate choice overload in recommender systems with click data. Other research was based on user studies (cf. Section 3).

4.1 Study with Mr. DLib and Sowiport[4]

Sowiport is the largest social science repository in Germany (Hienert et al 2015) and is operated by *GESIS – Leibniz-Institute for the Social Sciences*[5]. Sowiport's corpus 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, authors, journal or conference information, and sometimes citations, references, and links to full texts.

Sowiport shows related-article recommendations on each article's detail page (see Figure 1). Whenever such a detail page is browsed by a user, Sowiport requests a list of related-article recommendations from Mr. DLib's recommendations-as-a-service (cf. Figure 3). When Mr. Dlib receives a request for recommendations, Mr. DLib's recommender system randomly chooses one of four recommendation algorithms to generate recommendations: 1. Content-based filtering (CBF), 2. Stereotype recommendations, 3. Most Popular recommendations, and 4. Random recommendations.

Content-based filtering recommendations are calculated with Apache Lucene/Solr. The *stereotype* recommender assumes the stereotype of a researcher and recommends items related to academic writing, research methods, and research evaluation. The *most popular* recommender recommends the top viewed items and the top exported items. For *random* recommendations, random items form the digital library are recommended. For more details regarding the stereotype and most popular recommender, see Dinesh et al (2017). Note that the computation of the recommendation is based on the item that is currently being viewed. So far, the recommendation process is not personalized.

Mr. DLib further chooses randomly how many recommendations to return to Sowiport (a number between 1 and 15 is randomly chosen).

[4] Some explanations about Sowiport and Mr. DLib are copied from (Dinesh et al 2017).

[5] `http://www.gesis.org`

By having the two random variables *number of displayed recommendations* and *recommendation approach*, we can analyze with click data how users react to the different number of recommendations and recommendation approaches.

4.2 Data Collection

We analyze data from recommendations delivered in the three-month time period from 4 November 2016 to 8 February 2017. During this time, there were 5,436,080 requests for recommendations, i.e., users visited 5,436,080 articles' detail pages. For these requests, 5,436,080 recommendation sets were returned containing overall 41,434,861 recommendations (1-15 recommendations per set). All data of our study is available at `http://data.mr-dlib.org` to enable other researchers to replicate our calculations and perform further analyses (Beel et al 2018).

4.3 Metrics

The metrics with which we assess choice overload and recommendation effectiveness are:

- Click-through rate (CTR)
- Clicked set rate (CSR)
- Average clicks per clicked recommendation set (ACCS)
- Time to first click (TTFC)

In the remainder of this section, we will describe the metrics, their usefulness, and our expectations.

Click-through rate (CTR) describes the ratio of clicked to displayed recommendations. For instance, if 1,000 recommendations are displayed, and 9 of these recommendations are clicked, the CTR would be $9/1,000 = 0.9\%$. The assumption is that the higher the CTR, the more effective the recommendation approach is. There is some discussion as to what extent CTR is appropriate for measuring recommender effectiveness, but overall it has been demonstrated to be a meaningful and well-suited metric (Joachims et al 2005; Beel and Langer 2015; Schwarzer et al 2016).

For an increasing set size, the number of recommended items, the denominator of the CTR formula, increases. Because of this, there is the likelihood that CTR will always decrease. However, even with a decreasing CTR, the number of clicks per 1,000 recommendation sets can increase: For example, a CTR of 0.7% with one displayed recommendation results in $0.7\% \cdot 1 \cdot 1,000 = 7$ clicks per 1,000 recommendation sets.

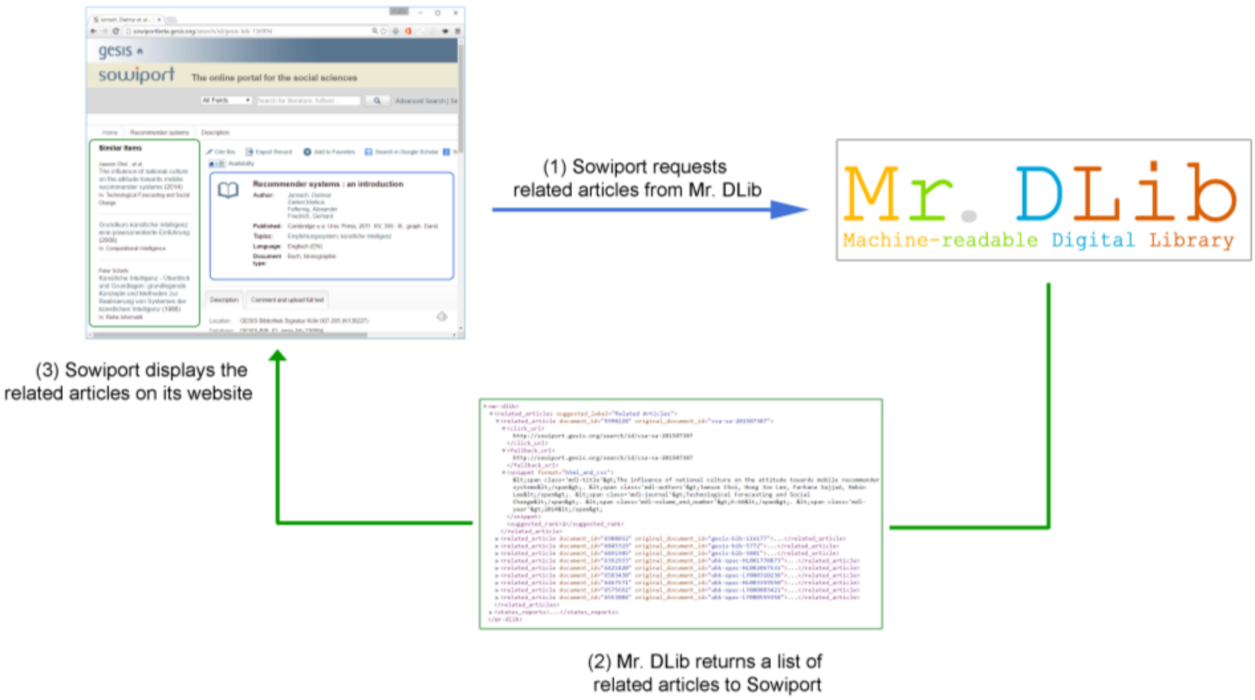

Fig. 3 The recommendation process of Sowiport and Mr. DLib

If we have a lower CTR, 0.56% with four displayed recommendations, the overall number of clicks per 1,000 recommendations sets is $0.56\% \cdot 4 \cdot 1,000 = 22.4$.

One possible outcome of our evaluation regarding CTR is illustrated in Figure 4. In this scenario, CTR would decrease as more recommendations are displayed in each recommendation set. At some point, we would expect users to be confronted with too many option, resulting in less motivation to make a choice, i.e., we would expect them to click on fewer recommendations or on none. However, while CTR would decrease, the absolute number of clicks would initially increase, and only level out or decrease when choice overload occurs. If the results were as in Figure 4, then the question would be: What is better – a higher CTR or the maximum of the number of absolute clicks? For instance, in Figure 4, the maximum number of absolute clicks is 22 for four displayed recommendations, but CTR is only 0.56%. In contrast, the maximum CTR (0.7%) is achieved for displaying one recommendation per set, but the absolute number of clicked recommendations is only 7.

Another potential outcome of our evaluation is illustrated in Figure 5. It could be that when the number of recommendations increases from a small number (e.g., one or two) to a somewhat larger number (e.g., four or five), CTR would increase. When the number of recommendations further increases (e.g., to 15), CTR may decrease. The rationale is as follows. If only few recommendations are displayed, the user may just not see them because they use only little space on the screen. Displaying more recommendations increases the chance that the user becomes aware of them. If, however, too many recommendations are displayed, choice overload would occur. Here, the CTR, the orange line, increases with the number of displayed recommendations, reaches a maximum at recommendations set size of four, and decreases afterwards. The clicks, the dashed gray line, reach a maximum at five displayed recommendations. In that case, we would decide to use a recommendation set size based on the maximum CTR or the maximum clicks when delivering recommendations to digital libraries.

Regarding CTR for the different recommendation algorithms, in general, we expect that the better the recommendation algorithm, the higher the CTR. This correlation could give us an indication which recommendations are most effective for users of Sowiport.

Clicked set rate (CSR) expresses the percentage of recommendation sets with at least one click. Our rationale for including this metric is that we consider a set of recommendations to be effective as long as at least one recommendation is clicked. Displaying more recommendations per set, the chances increase that at least one recommendation is perceived as relevant and is clicked. Our expectation would be that CSR increases when the number of recommendations per set increases,

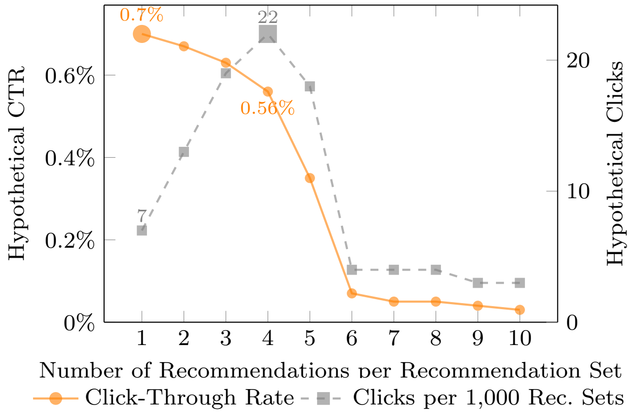

Fig. 4 Hypothetical CTR and clicks by displayed number of recommendations

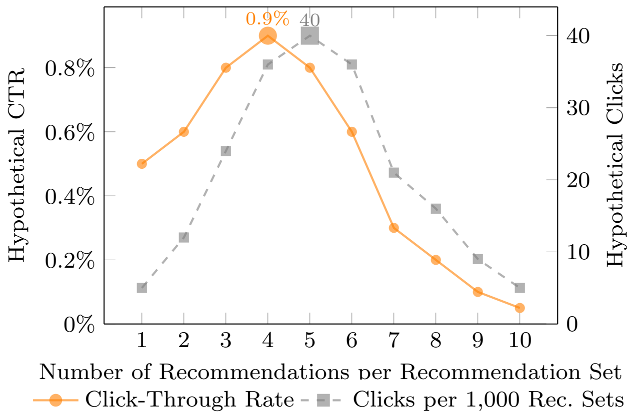

Fig. 5 Alternative hypothetical CTR and clicks by displayed number of recommendations

and potentially saturates after a certain number of recommendations per set, or even decreases. If CSR would decrease for a certain amount of recommendations, this would be a strong indicator for choice overload.

Average clicks per clicked recommendation set (ACCS) measures how many recommendations were clicked on average per set size, for those sets in which at least one recommendation was clicked. Users who may have felt choice overload, leading to not clicking any recommendation, are not reflected in this metric. However, this metric can help answer the second research question regarding the optimal number of recommendations to display. For recommendation sets with one recommendation, ACSS is always 1 because in a recommendation set of size 1, not more than one recommendation can be clicked. Using this metric we can get more insight into users that have clicked on recommendations. We can measure how many recommendations users click, and if average clicks decline when displaying too many recommendations. Additionally, we can assess which recommendation algorithm results in more clicks and therefore, possibly, display the most suitable recommendations for this scenario. Regarding the displayed number of recommendations, we expect a curve similar to the gray click lines in Figure 4 and 5. With more recommendation to choose from, the user might click on more, increasing the average clicks. If there are too many recommendations displayed, the user might not bother looking through them all and only click on one. This could indicate choice overload, limited effort, or a dissatisfaction with the recommendations. Regarding the recommendation algorithm, we expect the indication that the higher the number of clicks, the more relevant that algorithm for Sowiport users.

Time to first click (TTFC) is an additional metric we use to get an indication of choice overload, as well as to assess the effectiveness of the recommendations. TTFC is the time that has passed between delivering the recommendations and the first click on one of them. This is the time that the user probably looked at the website before deciding to click on one of the recommendations.

We expect TTFC to increase with each additional recommendation that we display, because we expect users to spend more time reading the titles and deciding whether to click on any. At some point though, a lack of a further increase in TTFC could indicate that the users feel choice overload, do not read all remaining options, and choose a recommendation from those already examined. Note that this is not strong evidence for choice overload but could also be because of limited effort or general dissatisfaction with the recommendations. We visualize this expectation in Figure 6. Note that users that experiences choice overload to an extent that they did not click on any item would have a TTFC of infinity and thus are not reflected in this metric. TTFC still can help answer our second research question regarding the ideal number of recommendations to display.

Our expectations regarding TTFC, given choice overload, are similar to our expectations for the CTR: at some point, there are too many recommendations displayed and the user will not examine all of them, but rather the first few. In the example given in the figure, we would recommend operators of digital libraries to display six recommendations as, when displaying more, the described point of too many displayed recommendations is reached. An alternative potential user behavior could be that the user examines recommendations sequentially, and for each one decides whether to click or not. In this case, the TTFC should increase linearly with increased recommendation set size. In this case, we could not determine whether choice overload exists, based on TTFC alone.

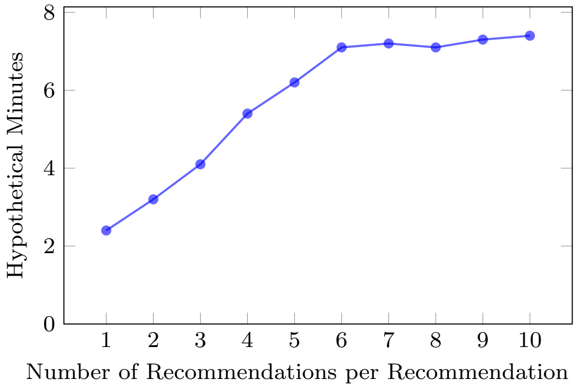

Fig. 6 Expectation regarding the average time between delivery of recommendation set and first click (TTFC), with respect to the number of displayed recommendations

Looking at the different TTFCs with respect to the recommendation algorithm, it will be hard to decide if a larger TTFC indicates thorough examination of good recommendations, and therefore indicates user satisfaction, or whether a larger TTFC indicates dissatisfaction with all results. Furthermore, users that felt so overloaded by choice that they did not click at all, are not reflected in this metric. When combined with our analysis of multiple clicks (ACCS), TTFC might give us an indication of what algorithm is most useful in the context of the Sowiport digital library: a combination of higher ACCS and a higher TTFC could indicate satisfaction with the results.

Additional Notes Regarding Relevance and Diversity. The results computed by the CBF recommender are always sorted by text relevance, which is calculated by Apache Lucene. Unfortunately, the relevance cannot be compared across queries (Langer and Beel 2017), so we cannot fully determine the relationship between relevance of the recommendation and the applied metrics like CTR, etc.

In our experiment, we did not take into consideration the diversity of the results. Extending the recommendation sets to consider this is left for future work.

Some of the results that we present may appear rather low. For example, CTR may be as low as 0.1%. In other recommender systems, CTR is often significantly higher; CTR is approximately 5% in Docear (Langer et al 2013). There are a number of potential reasons for this. On the one hand, bots crawling the Sowiport website increase the number of views, but not clicks (Beel and Dinesh 2017). In addition, when a user clicks on a recommendation, the recommended document opens in a new tab. For a recommendation set to have more than one click, a user would have to switch back to the previous tab to click a second recommendation. This is a rather high burden that may prevent many users from clicking more than one recommendation.

5 Experimental Study with the Sowiport Digital Library

5.1 Click-Through Rate

Analyzing the results of our experimental evaluation, we first look at CTR. The solid orange line in Figure 7 shows the CTR by the number of displayed recommendations. The more recommendations are displayed, the lower CTR becomes. The dashed gray line shows the average number of clicked recommendations (per 1,000 recommendation sets). The more recommendations per set that are displayed, the higher the number of absolute clicks becomes.

When only one recommendation per set was displayed, the CTR was 0.41% on average. This means for 1,000 sets delivered with 1 recommendation each, a total of 4.07 recommendations were clicked on average. For 1,000 sets delivered with 2 recommendations each, the CTR was only 0.25%. This means, 5.06 recommendations (out of $2 \cdot 1,000 = 2,000$) were clicked. For recommendation set size 2, half of the clicks were made on the first recommendation, and the other half on the second recommendation. For 1,000 sets delivered with 15 recommendations each, CTR was at a minimum of 0.09%, while the absolute number of clicks was at the maximum of 13.65.

Comparing these results with our expectations given in Section 4, we can see that the CTR decreases consistently and has a maximum at one displayed recommendation. However, contrary to our hypothetical expectations for absolute clicks in the case of choice overload, clicks continue to increase with increasing set size. They do not decrease after a certain set size. It may be that the absolute count would decrease if more than 15 recommendations per set were displayed.

The results show an under-proportional increase in average clicks on the displayed recommendations. This means, displaying twice as many recommendations does not double the clicks. For instance, doubling the number of recommendations from 1 to 2 increases absolute clicks per 1,000 sets by 24% (4.07 to 5.06). In order for the absolute number of clicked recommendations (per 1,000 sets) to double from 4.07 (for one displayed recommendation) to 8.1, the number of displayed related-articles has to be increased to 6. When 15 recommendations are displayed per set, only 3.35 times as many recommendations are clicked compared to displaying a single recommendation. Regarding choice overload this implies that having more recommendations to choose

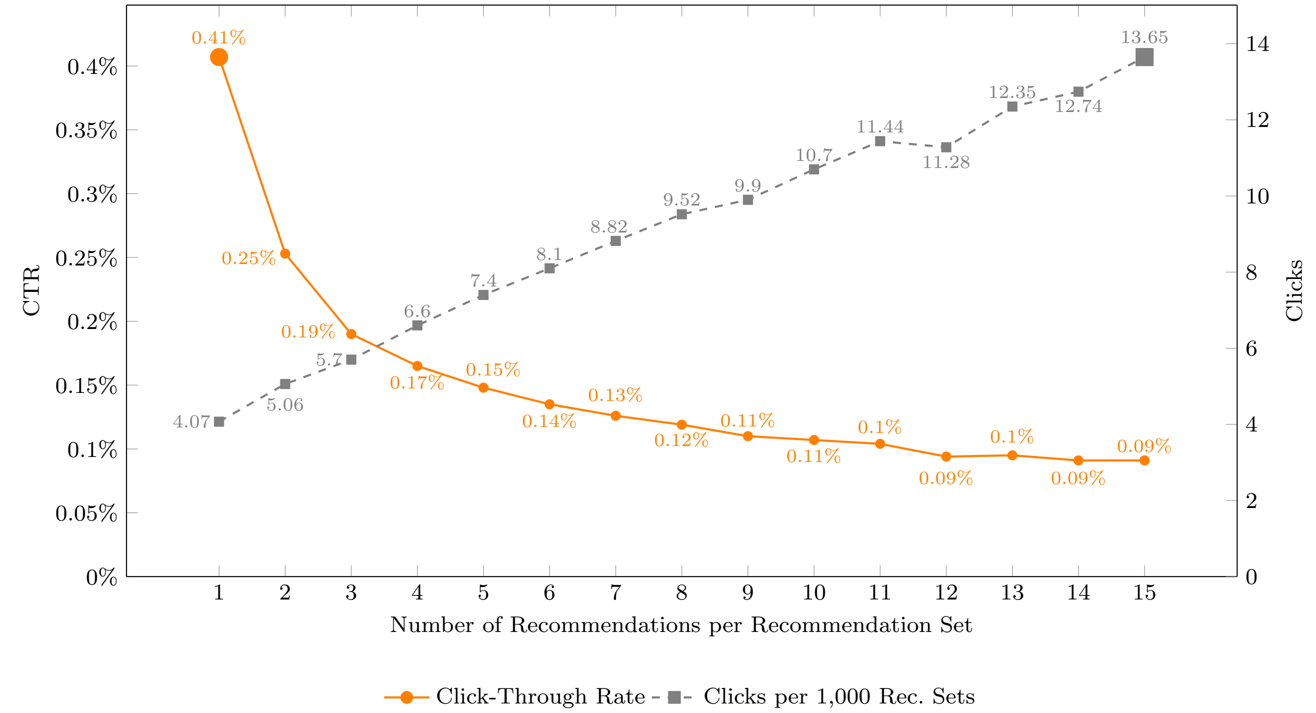

Fig. 7 CTRs (solid) and average absolute number of clicked recommendations (dashed) with respect to the number of displayed recommendations

from does, in general, only create a small incentive to click more items.

There are some points to consider when interpreting these results. Many documents in Sowiport only have sparse information, e.g., there is no author information available. Such documents may not be interesting for the users, and hence they may not be clicked. The session is another aspect to consider. For instance, if one user visits two pages and 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 consideration is that we do not filter recommendations that have already been shown to a user. If a user looks at 15 detail pages and gets 10 related-article recommendations on each, there could be duplicate recommendations, and hence the CTR may decrease if the same recommendations are repeatedly shown.

Compared to our previous work in Beierle et al (2017), the CTR curve shows the same trend, although in this paper, CTRs are lower. This could be because of an increased usage of bots that crawl Sowiport. Some differences to our previous results are that about 10 recommendations were needed in order to double the clicks (now 6). Displaying 15 recommendations yielded 2.5 times as many clicks (now 3.35). In general, we regard the current results as a confirmation of the trend we described in our previous work.

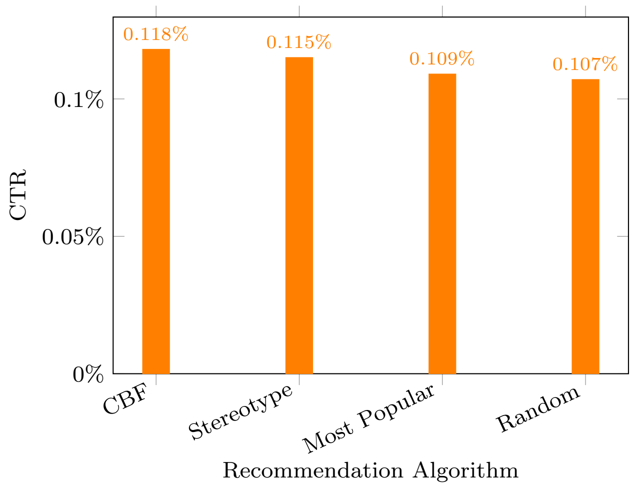

Fig. 8 CTR with respect to the recommendation algorithm

Looking at CTR with respect to the different recommendation algorithms, all values are similar and range from 0.107% to 0.118% (Figure 8). This indicates that the different recommendation algorithms do not differ very much from each other in terms of effectiveness, as measured by CTR. The biggest difference is between random and CBF – the CTR of the CBF recommendations is 10% higher than that of the random recommender. A chi-squared test with pairwise post hoc comparisons indicates that only the differences between CBF and most popular and between CBF and ran-

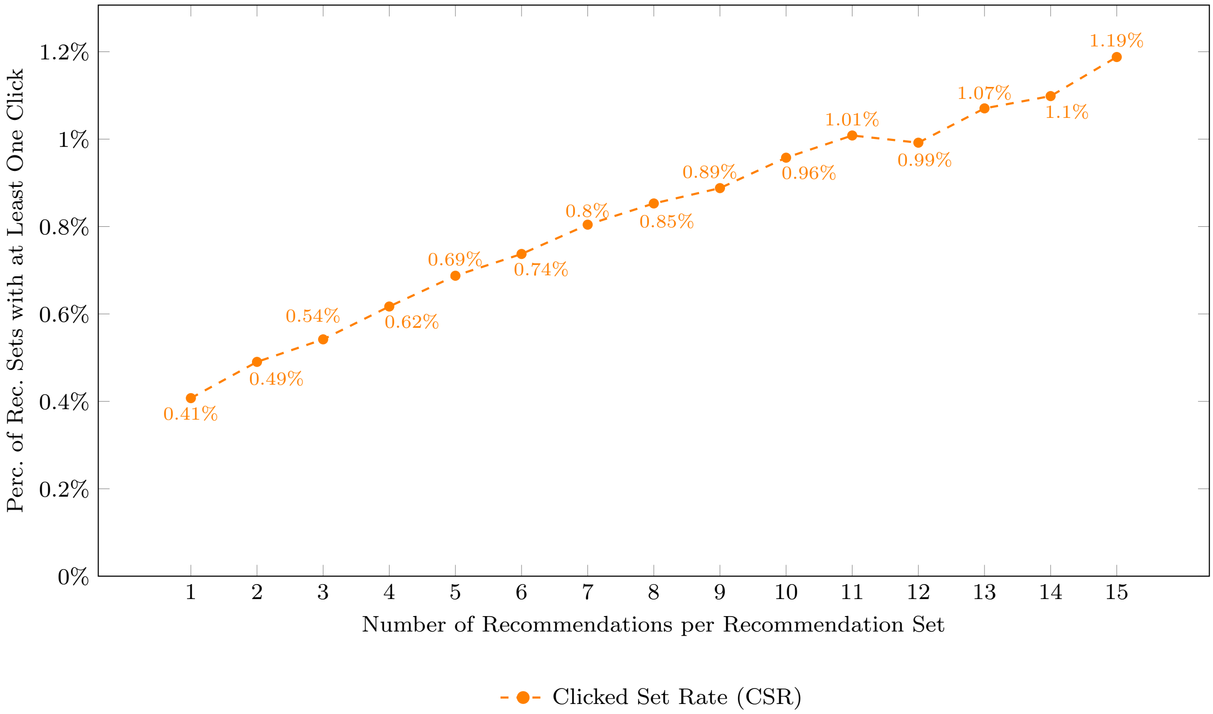

Fig. 9 Clicked Set Rate (CSR) indicating the percentage of recommendation sets with at least one click

dom are significant ($p < 0.05$). In Dinesh et al (2017), there was an analysis based on a smaller dataset. The main differences are that in this paper, with the bigger dataset, the range of the CTR is from 0.107% to 0.118%, while with the smaller dataset, the range was 0.11% to 0.145%. The second key difference is that most popular recommendations were slightly less effective than the baseline random recommender, while our data suggest a slightly higher effectiveness of the most popular recommender. For a deeper analysis on the effectiveness of stereotype and most popular recommendations in the Sowiport digital library, see Dinesh et al (2017). So far, the recommendation process is not personalized. This could lead to recommendations that are not of the highest quality, which in turn might contribute to the overall low level of CTRs. Improving the recommendations with personalization, we expect higher quality recommendations and a clearer differences between those recommendations and, e.g., the random baseline recommender. Looking at the combination of the different recommendation algorithms and the varying number of recommendations, the same general trend for CTR can be observed.

5.2 Clicked Set Rate

When applying the CSR metric that considers the percentage of recommendation sets with at least one click, we get the graph as depicted in Figure 9. The percentage keeps increasing, except for a slight drop-off at a recommendation set size of 12. Delivering 100,000 recommendation sets with one recommendation, 410 are clicked (0.41%). Delivering 100,000 recommendation sets with 15 recommendations, 1190 are clicked (1.19%). If we consider a set of recommendations to be effective if at least one recommendation is clicked, this metric would indicate that using a recommendation set size of at least 15 is sensible. Our expectation, given choice overload, was that at some point when displaying increased numbers of recommendations, CSR would decrease (cf. orange line in Figure 5). This does not happen, and as such, we have no indication of choice overload when considering CSR.

5.3 Average Clicks per Clicked Recommendation Set

In this section, we analyze the average number of clicks for recommendation sets with at least one click (ACCS). We want to find out more about users that click on recommendations: how many recommendations do they click on on average, and do they click less frequently if we display too many recommendations. In Figure 10, we plotted the means of ACCS by recommendation set size including standard error. Generally speaking, the average clicks increase when the recommendation set size is increased. For two displayed recommendations,

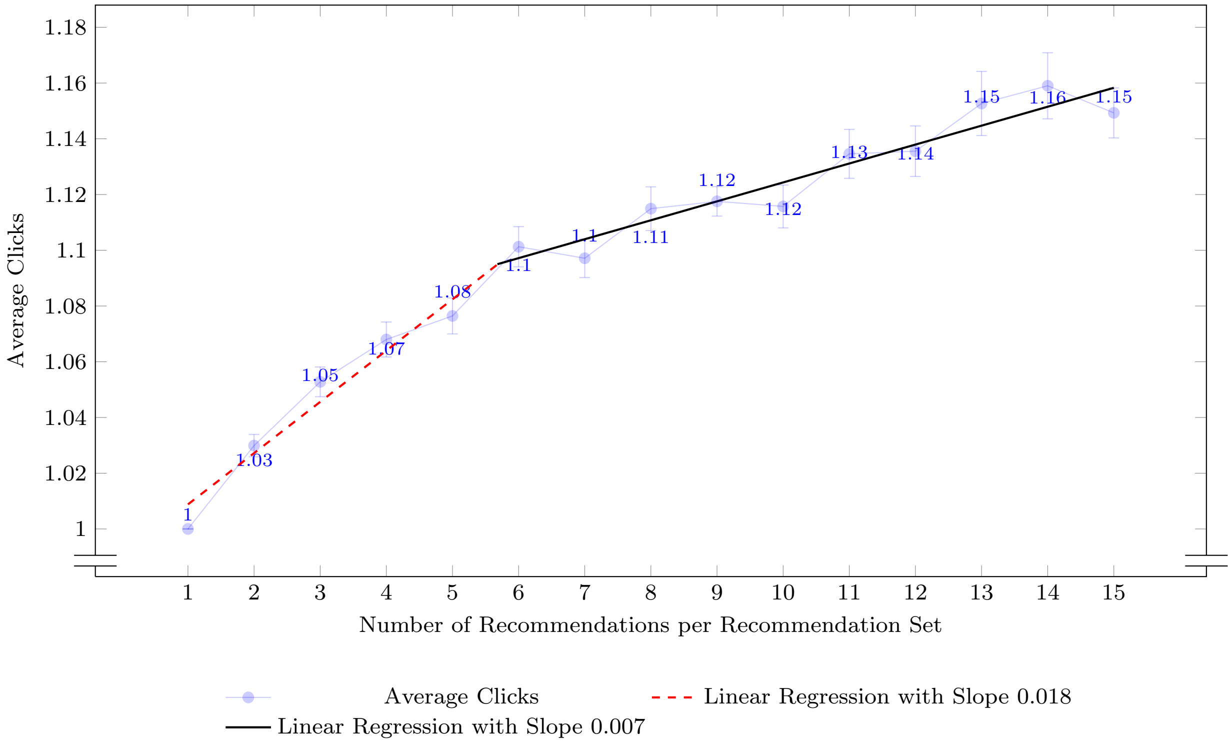

Fig. 10 Average Clicks for Clicked Recommendation Sets (ACCS), with respect to the number of displayed recommendations. Also plotted are two fitted linear regressions with breakpoint at 5.68

ACCS was 1.03, which means that almost none of the users clicked on more than one of the two recommendations. The highest value is just 1.16, for 14 displayed recommendations.

Looking at the plot, the slope seems to be not the same throughout. Utilizing a metric for finding breakpoints (Muggeo 2003, 2008), we performed a statistical analysis that suggests that the plot can be segmented and two linear regressions can be fitted with different slopes. There is a breakpoint between those two linear regressions at 5.68 with an error of ±0.86. A Chow-test confirms the significance of the breakpoint at 5.68 with $p = 0.00003$. In the figure, we can see that the slope of the fitted linear regression between 1 and 5-6 recommendations is 0.018 (error of 0.003). The fitted linear regression between 5-6 and 15 has a slope 61% smaller (0.007, error of 0.001). Given the breakpoint and its error, we can establish that after 5 to 6 recommendations, ACCS does not increase as much as for fewer recommendations. One part of our first expectation from Section 4, that ACCS may increase as recommendation set size increases, is apparent. However, we do not see a negative slope after the initial increase.

There are several things to consider when interpreting these results. We observe that overall, displaying more recommendations only creates a small incentive for the user to click on more than one of the recommended items. 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 this case, the user might not exactly experience choice overload but might just not be willing to engage in looking into further recommendations. If this is the case, the research should be repeated when we are able to deliver better recommendations. Another aspect to consider in the results' interpretation is how users might use Sowiport. If the users clicks on a recommendation, a new tab is opened. If the recommendation was good, they might forget about the other open tab – especially if there are further good recommendations shown in the new tab. For this study however, we did not track the sessions of the user, i.e., which clicks are from which user. To further investigate the implications of sessions, new tabs opening, etc., we need to repeat the experiments after tracking user sessions.

Investigating the different recommendation algorithms, we want to see which ones result in the most clicks and which, therefore, likely display the most appropriate recommendations for this scenario. In Figure 11, we see a range of average clicks from 1.08 to 1.28. The data is non-normally distributed and we performed a Kruskal-Wallis test. The test indicates that we can dismiss the null hypothesis that there is no difference between the algorithms ($p < 0.05$). Furthermore,

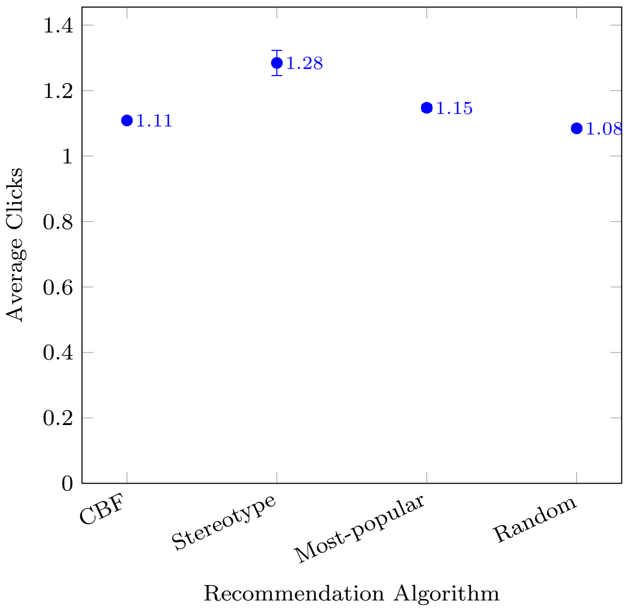

Fig. 11 Average Clicks for Clicked Recommendation Sets (ACCS), with respect to the recommendation algorithm

Dunn's multiple comparison test indicates that all of the algorithms are significantly different ($p < 0.05$). Because the stereotype recommender attracts the most clicks on average, we regard it to be the most effective. The nature of the stereotype recommender in the system is based on the assumption that the users are interested in academic writing, research methods, and research evaluation. The stereotype recommender will produce recommendations based on these categories. Having the highest number of average clicks is evidence that our assumption about the Sowiport users may be correct and that they indeed fit the assumed stereotype.

5.4 Time to First Click

TTFC may give us evidence of choice overload. It might, for example, show that users spend more time before clicking recommendations when more recommendations are shown. If TTFC stops increasing at some point as more recommendations are displayed, this may indicate the point at which choice overload tends to increase. Analyzing TTFC with respect to the varying number of recommendations, in general, TTFC increases with increasing recommendation set size, see Figure 12[6]. For one displayed recommendation, the average TTFC is 3.8 minutes. We assume that the first minutes after receiving the recommendations are spend looking at the search result and users in general afterwards looked at the recommendations. At 15 displayed recommendations, the average TTFC is at a maximum of 4.57 minutes.

As with the average clicks in Figure 10, we can see that the slope does not seem to be constant. Statistical analysis yields two linear regressions with different slopes with a breakpoint at 4.5 with an error of ± 1. A Chow-test confirms the significance of that breakpoint with $p = 0.005$. We plotted the two differing slopes, 0.113 (error of 0.034) and 0.031 (error of 0.006). Displaying 5 instead of 1 recommendation, the average time until the first click increases from 3.8 to 4.28 minutes (about 13%). Displaying 15 instead of 5 recommendation increases the average TTFC about 7% (from 4.28 to 4.57 minutes).

Given the breakpoint and its error, we can establish that after 4 to 6 recommendations, the TTFC does not increase as much as for fewer recommendations. The first part of our expectation from Section 4, increasing TTFC for increasing recommendation set size, is apparent. We consider this a confirmation of the expected effect of users spending more time looking at the recommendations when offered more choice. After displaying a certain number of recommendations, TTFC does not increase as much. This could indicate choice overload – too many recommendations are displayed and not much additional time is spend to look at all the results.

One important aspect to consider when interpreting these results is that we do not know much about the behavior of Sowiport users. There might be different types of users, for example users that click quickly when they are interested in something and users that wait and spend more time looking at the recommended documents. In general, a shorter time might indicate that the user quickly found a fitting recommendation to click, while taking a longer time for the first click might indicate that the displayed recommendations are equally good or equally bad.

Looking at the different recommendation algorithms we see a range of average TTFCs from 4.26 to 4.58 minutes (Figure 13). We performed a Kruskal-Wallis test on the non-normally distributed data. We can dismiss the null hypothesis that there is no difference between the algorithms ($p < 0.05$). Dunn's multiple comparison test furthermore indicates that all groups are significantly different from each other ($p < 0.05$), except when comparing most-popular and random ($p = 0.2$). The expectation could be that random recommendations would irritate the users and thus yield the highest TTFC. One possible explanation could be that a lot of users tend to click on the first item of the recommendation set as previous work regarding position bias has

[6] Here, as a data cleaning process, we considered recommendations as clicked if they were clicked within 30 minutes of delivery of the recommendation set. This is true for 90.77% of all first clicks in the dataset (39,709 of 43,748).

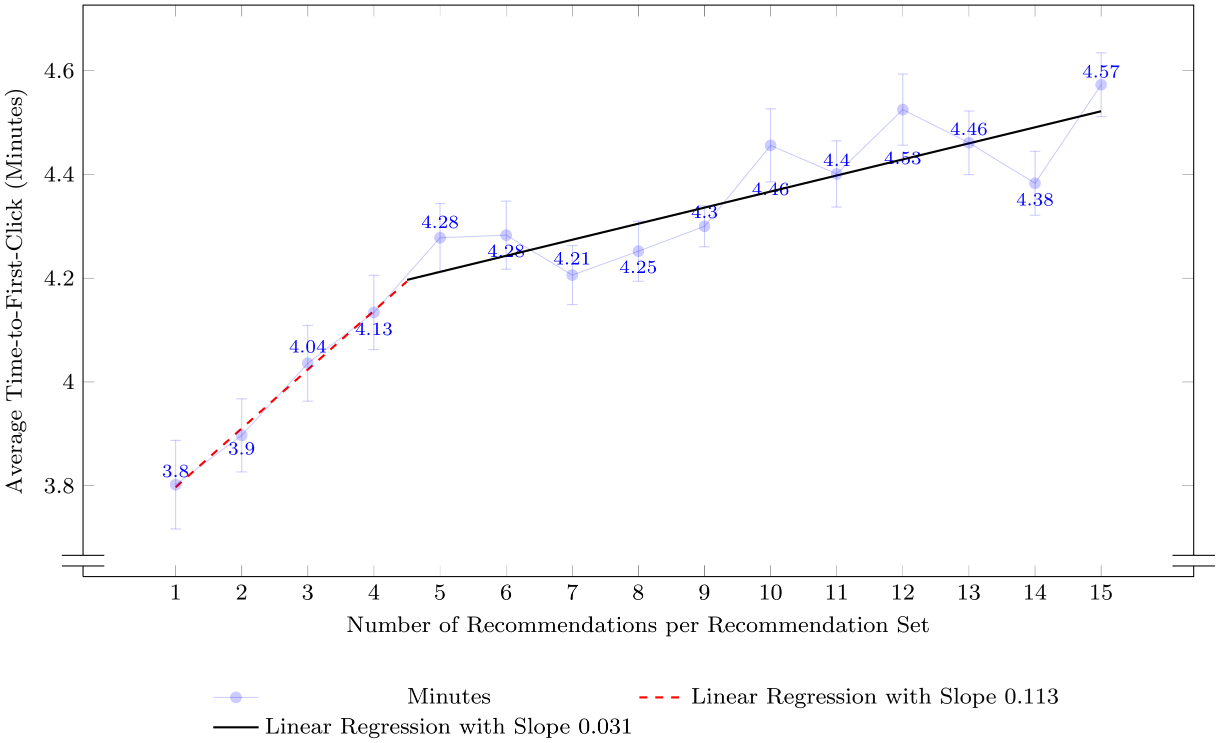

Fig. 12 Average time between delivery of recommendation set and first click (TTFC), with respect to the number of displayed recommendations. Also plotted are two fitted linear regressions with breakpoint at 4.5

shown (Collins et al 2018). In this case, users might assume that the recommendation displayed at the top might be good and might click on it without inspecting its relevance, relative to other recommendations in the set.

One reason for the stereotype recommender yielding the highest TTFC could be that the recommendations are too similar to each other, thus resulting in a longer time the users spends to choose which one to click. In combination with Figure 11, we can see that users spend both more time until the first click, and more average clicks on stereotype recommendations. Users most likely look at the recommendations longer before clicking more of them. As we described in Section 4, the combination of having more clicks and more time spend before the first click might indicate satisfaction with the results. This could be additional evidence that Sowiport users are interested in stereotypical research recommendations. However, this is not very strong evidence for user satisfaction and needs to be validated in future work. The results from the TTFC metric help us for our second research question regarding the ideal number of recommendations to display. They points towards a recommendations set size of 4–6.

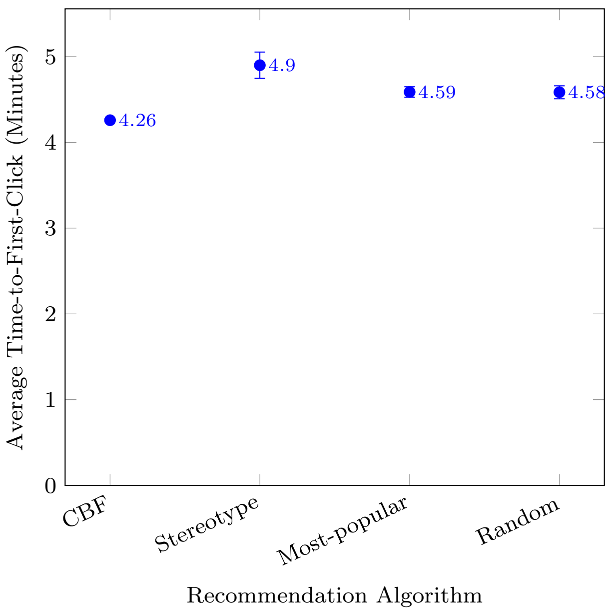

Fig. 13 Average time between delivery of recommendation set and first click (TTFC), with respect to the recommendation algorithm

5.5 Discussion

We can draw several conclusions about choice overload for users of Sowiport, by considering the evaluated metrics in tandem.

CTR decreases consistently with increasing set sizes, and does not provide us with evidence of choice overload. Conversely, both CSR and Clicks per 1,000 sets show an under-proportional increase in clicks for increasing set sizes; doubling the number of recommendations does not double the number of clicks. This may suggest choice overload. Despite their under-proportional increases however, they do both increase linearly up to a set size of 15.

We see from Average Clicks per Clicked Set that, on average, users tend to click just one recommendation at most. For example, increasing the set size from 2 to 15 increases average clicks from 1.03 to just 1.15. Given that users tend to click one recommendation at most, and given that CSR and Clicks per 1,000 sets increases linearly up to a maximum set size, we can say that there is no strong evidence for choice overload for users of Sowiport according to these results. The more recommendations you give to users in sets, the more the absolute number of clicks will increase, up to the largest set size. Considering ACCS, Clicks per 1,000 sets, and CSR together, it would be sensible to display as many relevant recommendations as possible.

In contrast, TTFC and ACCS together provide small evidence for user disengagement with increasing set size. The slope of both metrics decreases when more than 5–6 recommendations are shown. We cannot state firmly that this effect is due to choice overload, or whether some other factor affects user engagement when more than 5–6 recommendations are shown. It does give us guidance regarding the optimal number of recommendations to display to users of Sowiport, however. Our pre-study showed that 74% of digital libraries displayed between 3 and 5 recommendations, with 28% displaying just 3. In the case of Sowiport, we suggest that users will engage with 5–6 recommendations per set before slight disengagement becomes apparent on average.

6 Conclusion and Future Work

In this paper, we investigated the challenges of choice overload and recommendation effectiveness in related-article recommendations in digital libraries. In a first step, we analyzed existing digital libraries. 74% of those that display recommendations display 3–5 recommendations. It remains unclear how those numbers are determined by their operators. We analyzed 41.4 million recommendations delivered to users of the digital library Sowiport. These recommendations were delivered in sets of 1 to 15. We applied four different metrics to analyze the data: CTR, CSR, ACCS, and TTFC.

With increasing recommendation set size, CTR consistently decreases, from 0.41% for 1 recommendation to 0.09% for 15 recommendations. CTR does not provide evidence for choice overload. CSR, the percentage of recommendation sets with at least one click, increases with increasing recommendation set size. There is a maximum of 1.19% at 15 recommendations. Choice overload cannot be confirmed with this metric.

We find that users are not inclined to click on more than one recommendation. ACCS does increase with increasing recommendation set size, however this increase is slight. This increase is higher when displaying up to 5–6 recommendations. There is a maximum of 1.15 clicks on average for 15 displayed recommendations. We use the TTFC metric that measures the time between the delivery of the recommendation set and the first click. We observe an increase in TTFC for increasing recommendation set sizes, with a maximum of 4.57 minutes for 15 displayed recommendations. The TTFC increases more for 1 to 4–6 displayed recommendations, and increases less for more recommendations. The results for ACCS and the results for TTFC show some evidence that delivering 5–6 recommendations to users of Sowiport may be most appropriate.

Looking at the different recommendation algorithms, content-based filtering yields the highest CTR with 0.118%. The stereotype recommender generates recommendations that attract the highest ACCS (1.28). Stereotype recommendations also yield the highest TTFC. That means that users spend more time with items from the stereotype recommender before clicking on any item. Future work has to show if this is because the items are too similar or because the recommendations are perceived as especially useful.

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 – both the cloud (Kopp et al 2018) and desktop version (Feyer et al 2017) – and the library of the Technical University of Munich.

Future work could also include using other evaluation methods and metrics, for example, user ratings or tracking which recommended items were actually exported or saved by users. Also, we plan on conducting a user study to further study the effect of choice overload. The user study could involve interviews with users on why they clicked the recommendations they clicked, and how they felt about the quality and amount of recommendations. It may also be useful to survey operators of digital libraries to understand their decision making process for deployment of a recommender system, for example, discovering why they chose to use certain algorithms, or why they display certain numbers of recommendations.

Other forms of bias may affect a user's level of choice overload, such as production biases (Spillane et al 2017) or presentation biases (Radlinski and Joachims 2006; Collins et al 2018). Choice overload could be examined with these factors varied.

Lastly, it would be useful to assess what recommender systems in digital libraries should aim to achieve, for instance, whether it is beneficial to maximize CTR, or maximize the number of clicked recommendations.

Acknowledgements We are grateful for the support provided by Jan Blechschmidt, Boris Lorbeer, Sophie Siebert, Stefan Feyer, Siddharth Dinesh, Zeljko Carevic, and Mr. DLib's partners. We furthermore thank the reviewers for their valuable comments.

References

Arunachalam B, Henneberry SR, Lusk JL, Norwood FB (2009) An Empirical Investigation into the Excessive-Choice Effect. American Journal of Agricultural Economics 91(3):810–825, DOI 10.1111/j.1467-8276.2009.01260.x

Azzopardi L, Zuccon G (2016) Two Scrolls or One Click: A Cost Model for Browsing Search Results. In: Advances in Information Retrieval, Springer, Cham, pp 696–702, DOI 10.1007/978-3-319-30671-1_55

Bawden D, Robinson L (2009) The dark side of information: overload, anxiety and other paradoxes and pathologies. Journal of information science 35(2):180–191

Beel J, Dinesh S (2017) Real-world recommender systems for academia: The pain and gain in building, operating, and researching them. In: BIR@ ECIR, pp 6–17

Beel J, Langer S (2015) 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, Springer, Cham, pp 153–168, DOI 10.1007/978-3-319-24592-8_12

Beel J, Gipp B, Langer S, Genzmehr M, Wilde E, Nürnberger A, Pitman J (2011) Introducing Mr. DLib, a Machine-readable Digital Library. In: Proceedings of the 11th ACM/IEEE Joint Conference on Digital Libraries (JCDL'11), ACM, pp 463–464, DOI 10.1145/1998076.1998187

Beel J, Gipp B, Langer S, Breitinger C (2016) Research paper recommender systems: A literature survey. International Journal on Digital Libraries pp 1–34, DOI 10.1007/s00799-015-0156-0

Beel J, Aizawa A, Breitinger C, Gipp B (2017) Mr. DLib: Recommendations-as-a-Service (RaaS) for Academia. In: 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL), pp 1–2, DOI 10.1109/JCDL.2017.7991606

Beel J, Smyth B, Collins A (2018) RARD II: The 2nd Related-Article Recommendation Dataset. In: arXiv:1807.06918 [cs.IR]

Beierle F, Aizawa A, Beel J (2017) Exploring Choice Overload in Related-Article Recommendations in Digital Libraries. In: Proceedings of the 5th International Workshop on Bibliometric-enhanced Information Retrieval (BIR2017), co-located with the 39th European Conference on Information Retrieval (ECIR), CEUR-WS, pp 51–61, URL `http://ceur-ws.org/Vol-1823/paper5.pdf`

Besedeš T, Deck C, Sarangi S, Shor M (2015) Reducing choice overload without reducing choices. Review of Economics and Statistics 97(4):793–802

Bollen D, Knijnenburg BP, Willemsen MC, Graus M (2010) Understanding Choice Overload in Recommender Systems. In: Proceedings of the Fourth ACM Conference on Recommender Systems, ACM, RecSys '10, pp 63–70, DOI 10.1145/1864708.1864724

Bornmann L, Mutz R (2015) Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology 66(11):2215–2222

Buturak G, Evren Ö (2017) Choice overload and asymmetric regret. Theoretical Economics 12(3):1029–1056

Chernev A, Böckenholt U, Goodman J (2015) Choice overload: A conceptual review and meta-analysis. Journal of Consumer Psychology 25(2):333–358

Chiravirakul P, Payne SJ (2014) Choice Overload in Search Engine Use? In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, ACM, CHI '14, pp 1285–1294, DOI 10.1145/2556288.2557149

Collins A, Tkaczyk D, Aizawa A, Beel J (2018) Position bias in recommender systems for digital libraries. In: Chowdhury G, McLeod J, Gillet V, Willett P (eds) Transforming Digital Worlds, Springer International Publishing, Cham, pp 335–344

Dellaert B, Baker T, Johnson E (2017) Partitioning sorted sets: Overcoming choice overload while maintaining decision quality. Columbia Business School Research Paper 18

Dinesh S, Beel J, Mayr P, Carevic Z, Raghvendra J (2017) Stereotype and Most-Popular Recommendations in the Digital Library Sowiport. In: Proceedings of the 15th International Symposium of Information Science (ISI), Verlag Werner Hülsbusch, Glückstadt, Germany, pp 96–108

Fasolo B, McClelland GH, Todd PM (2007) Escaping the tyranny of choice: When fewer attributes make choice easier. Marketing Theory 7(1):13–26, DOI 10.1177/1470593107073842

Feyer S, Siebert S, Gipp B, Aizawa A, Beel J (2017) Integration of the Scientific Recommender System Mr. DLib into the Reference Manager JabRef. In: Proceedings of the 39th European Conference on Information Retrieval (ECIR)

Haynes GA (2009) Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction. Psychology & Marketing 26(3):204–212

Hienert D, Sawitzki F, Mayr P (2015) Digital Library Research in Action: Supporting Information Retrieval in Sowiport. D-Lib Magazine 21(3/4)

Joachims T, Granka L, Pan B, Hembrooke H, Gay G (2005) Accurately Interpreting Clickthrough Data As Implicit Feedback. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, SIGIR '05, pp 154–161

Jones M, Marsden G, Mohd-Nasir N, Boone K, Buchanan G (1999) Improving Web interaction on small displays. Computer Networks 31(11–16):1129–1137, DOI 10.1016/S1389-1286(99)00013-4

Jugovac M, Jannach D (2017) Interacting with recommenders—overview and research directions. ACM Transactions on Interactive Intelligent Systems (TiiS) 7(3):10

Kelly D, Azzopardi L (2015) 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, ACM, SIGIR '15, pp 183–192

Kim J, Thomas P, Sankaranarayana R, Gedeon T, Yoon HJ (2015) 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, DOI 10.1002/asi.23187

Knijnenburg BP, Willemsen MC, Gantner Z, Soncu H, Newell C (2012) Explaining the user experience of recommender systems. User Modeling and User-Adapted Interaction 22(4-5):441–504, DOI 10.1007/s11257-011-9118-4, URL http://dx.doi.org/10.1007/s11257-011-9118-4

Kopp O, Breitenbücher U, Müller T (2018) Cloudref – towards collaborative reference management in the cloud. In: Proceedings of the 10th Central European Workshop on Services and their Composition

Langer S, Beel J (2017) Apache Lucene as Content-Based-Filtering Recommender System: 3 Lessons Learned. In: Proceedings of the 5th International Workshop on Bibliometric-enhanced Information Retrieval (BIR2017), co-located with the 39th European Conference on Information Retrieval (ECIR), CEUR-WS, pp 85–92, URL http://ceur-ws.org/Vol-1823/paper8.pdf

Langer S, Genzmehr M, Beel J, Nürnberger A (2013) Introducing Docear's Research Paper Recommender System. In: Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'13), ACM, pp 459–460, DOI 10.1145/2467696.2467786

Linden G (2006) Marissa Mayer at Web 2.0. http://glinden.blogspot.com/2006/11/marissa-mayer-at-web-20.html

Moser C, Phelan C, Resnick P, Schoenebeck SY, Reinecke K (2017) No such thing as too much chocolate: Evidence against choice overload in e-commerce. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, CHI '17, pp 4358–4369, DOI 10.1145/3025453.3025778

Motajcsek T, Le Moine JY, Larson M, Kohlsdorf D, Lommatzsch A, Tikk D, Alonso O, Cremonesi P, Demetriou A, Dobrajs K, Garzotto F, Göker A, Hopfgartner F, Malagoli D, Nguyen TN, Novak J, Ricci F, Scriminaci M, Tkalcic M, Zacchi A (2016) Algorithms aside: Recommendation as the lens of life. In: Proceedings of the 10th ACM Conference on Recommender Systems, ACM, New York, NY, USA, RecSys '16, pp 215–219, DOI 10.1145/2959100.2959164

Muggeo VMR (2003) Estimating regression models with unknown break-points. Statistics in Medicine 22:3055–3071, DOI 10.1002/sim.1545

Muggeo VMR (2008) Segmented: An R package to fit regression models with broken-line relationships. R News 8/1:20–25

Oulasvirta A, Hukkinen JP, Schwartz B (2009) When More is Less: The Paradox of Choice in Search Engine Use. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, SIGIR '09, pp 516–523

Park JY, Jang SS (2013) Confused by too many choices? choice overload in tourism. Tourism Management 35:1–12

Radlinski F, Joachims T (2006) Minimally invasive randomization for collecting unbiased preferences from clickthrough logs. In: Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2, AAAI Press, AAAI'06, pp 1406–1412

Reed DD, Reed FDD, Chok J, Brozyna GA (2011) The "tyranny of choice": Choice overload as a possible instance of effort discounting. The Psychological Record 61(4):547–560

Scheibehenne B, Greifeneder R, Todd PM (2010) Can there ever be too many options? a meta-analytic review of choice overload. Journal of Consumer Research 37(3):409–425

Schwartz B (2004) The tyranny of choice. Scientific American 290(4):70–75

Schwarzer M, Schubotz M, Meuschke N, Breitinger C, Markl V, Gipp B (2016) Evaluating Link-based Recommendations for Wikipedia. In: 2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL), pp 191–200

Spillane B, Lawless S, Wade V (2017) Perception of bias: the impact of user characteristics, website design and technical features. In: Proceedings of the International Conference on Web Intelligence, ACM, pp 227–236

Stempfhuber M, Schaer P, Shen W (2008) Enhancing visibility: Integrating grey literature in the sowiport information cycle. In: Conference Papers: International Conference on Grey Literature, vol 9, pp 23–29

Van Noorden R (2014) Global scientific output doubles every nine years. Nature Blog. URL http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html

Willemsen MC, Graus MP, Knijnenburg BP (2016) Understanding the role of latent feature diversification on choice difficulty and satisfaction. User Modeling and User-Adapted Interaction 26(4):347–389, DOI 10.1007/s11257-016-9178-6

Zheng Y, Mobasher B, Burke R (2016) Emotions in context-aware recommender systems. In: Tkalčič M, De Carolis B, de Gemmis M, Odić A, Košir A (eds) Emotions and Personality in Personalized Services, Springer, Cham, pp 311–326, DOI 10.1007/978-3-319-31413-6_15

Additional Information

Bibliographic Data

F. Beierle, A. Aizawa, A. Collins, and J. Beel, “Choice Overload and Recommendation Effectiveness in Related-Article Recommendations. Analyzing the Sowiport Digital Library,” *International Journal on Digital Libraries* (2019).
https://doi.org/10.1007/s00799-019-00270-7

Pre-print from

https://beierle.de

Online at

https://doi.org/10.1007/s00799-019-00270-7 and https://rdcu.be/bEDnM

Authors

Felix Beierle

Akiko Aizawa

Andrew Collins

Joeran Beel

BibTeX

```
@article{BeierleIJDL2019,
title = {{Choice Overload and Recommendation Effectiveness in Related-Article Recommendations. Analyzing the Sowiport Digital Library}},
author = {Beierle, Felix and Aizawa, Akiko and Collins, Andrew and Beel, Joeran},
journal = {{International Journal on Digital Libraries}},
publisher = {Springer},
year = {2019},
doi = {10.1007/s00799-019-00270-7}
}
```