The Impact of Demographics (Age and Gender) and Other User-Characteristics on Evaluating Recommender Systems

Joeran Beel[1], Stefan Langer[1], Andreas Nürnberger[2], and Marcel Genzmehr[1]

[1] Docear, Germany
{beel,langer,genzmehr}@docear.org
[2] Otto-von-Guericke University,
Dpt. of Computer Science, DKE Group, Magdeburg, Germany
andreas.nuernberger@ovgu.de

Abstract. In this paper we show the importance of considering demographics and other user characteristics when evaluating (research paper) recommender systems. We analyzed 37,572 recommendations delivered to 1,028 users and found that elderly users clicked more often on recommendations than younger ones. For instance, 20-24 years old users achieved click-through rates (CTR) of 2.73% on average while CTR for users between 50 and 54 years was 9.26%. Gender only had a marginal impact (CTR males 6.88%; females 6.67%) but other user characteristics such as whether a user was registered (CTR: 6.95%) or not (4.97%) had a strong impact. Due to the results we argue that future research articles on recommender systems should report detailed data on their users to make results better comparable.

Keywords: recommender systems, demographics, evaluation, research paper.

1 Introduction

There are more than one hundred research articles on research paper recommender systems, and even more on recommender systems in general. Many of them report on new recommendation approaches and their effectiveness. For instance, *Papyrus* is supposed to have a precision around 20% [1]; Quickstep's approach is supposed to have a precision around 10% [2]; and Jomsri et al. claim an accuracy of 91.66% for their research paper recommender system [3]. Unfortunately, results cannot be compared with each other because researchers used different evaluation methods, metrics, and data sets.

We believe there is another factor influencing the comparability which has received too little attention: users' demographics and characteristics. In other disciplines it is well known that results from one study cannot be used to draw conclusions for a population if the study's user sample differs too much from that population. For instance, in marketing you cannot draw reliable conclusions about how elderly people in Germany will react to a product if a study about that product was conducted in France with university students. Evaluations of recommender systems widely ignored

T. Aalberg et al. (Eds.): TPDL 2013, LNCS 8092, pp. 400–404, 2013.

differences in user samples. Some studies report to have asked their participants for demographic data, but they do not report on them in their papers [4]. Another paper reports that age and gender had no impact on the accuracy of recommendations but test subjects were all students [5]. With students typically being all in the same age-range, it is no surprise that the study could not find any differences between different ages.

We analyzed empirical data collected with Docear's research paper recommender system [6] to find out whether users' demographics and characteristics influence the outcome of the recommender system evaluation.

2 Methodology

Docear users can register an account and provide demographic information such as year of birth and gender if they like. They may also opt-in for receiving research paper recommendations (even without registration). Recommendations are shown on request or automatically every three days of use, ten at a time. During March and Mai 2013 1,028 users received 37,572 recommendations. Details on the recommendation process may be found in [6]. For the evaluation we used click-through rate (CTR) which expresses how many out of the displayed recommendations were clicked. For instance, when 37,572 recommendations were shown, and 2,361 were clicked, CTR is 6.28%. CTR is a common measure in online advertisement and equivalent to "precision" in information retrieval.

3 Results

From a total of 1,028 users who received recommendations, 38.62% did not register and 61.38% registered. 21.79% registered but did not provide information about their gender, 33.17% registered and were males, and 6.42% registered and were females (Figure 1, left pie). Looking only at those users who specified their gender, 83.79% were male, and 16.22% were female (Figure 1, right pie). Among the genders there is only a marginal difference in CTR with 6.88% for males and 6.67% for females (Figure 2). However, there is a significant difference between registered users (6.95%) and unregistered users (4.97%). Interestingly, those users who registered and did not specify their gender have the highest CTR with 7.14%. Another interesting difference between genders relates to the willingness of accepting recommendations. From all male users, 38.09% activated recommendations while only 34.74% of women did and even less (28.72%) of the users who did not specify their gender during registration (Table 1). This might indicate that these users are concerned about privacy issues when receiving recommendations [7].

From the registered users, 39.62% did not specify their age. From those who did, around one quarter (24.15%) were 25 to 29 years of age (Figure 3, bar chart). 11.29% were between 20 and 24 years and only two users were younger than 20, namely 17 and 18. The vast majority (88.19%) was older than 25 years. 4.46% of the users were 60 or older. The mean age was 36.56 years, the median was 33. Of course, it might be that some users did not provide their correct age and the true ages slightly differ from the ones presented.

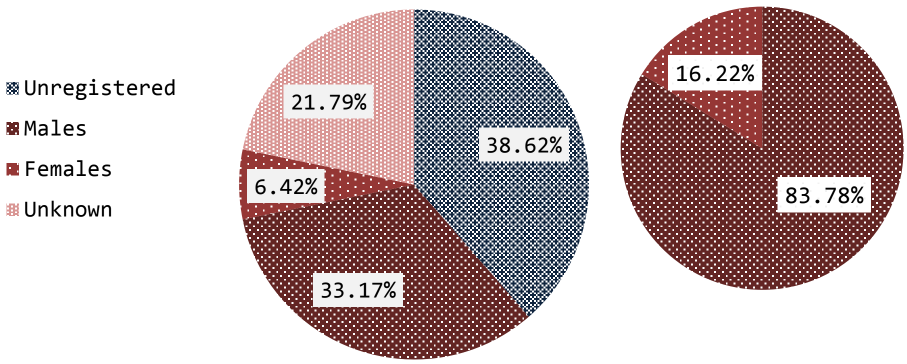


Fig. 1. Gender and user type (registered/unregistered) distribution

Looking at click-through rate by age shows that the older a user is the higher CTR becomes (Figure 3, dotted line). While younger users (20-24 years) have the lowest CTR of only 2.73% on average, CTR for users older than 60 is the highest with 9.92%. Overall, a clear linear trend is recognizable (Figure 3, dotted line). CTR for users who registered but did not provide their age was 7.66% on average (not shown in Figure 3).

Table 1. Percentage of activated recommendations by gender

	Male	Female	n/a
Recs. Activated	38.09%	34.74%	28.72%
Recs. Deactivated	61.91%	65.26%	71.28%

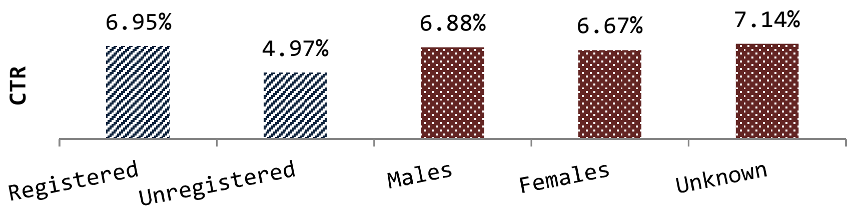
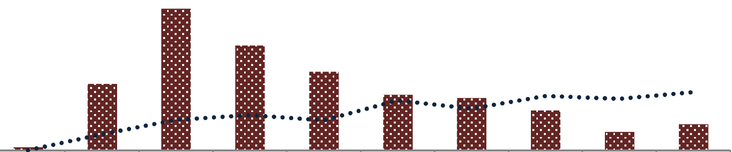


Fig. 2. Click-through rate (CTR) by user type and gender



	0-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60+
# of users	0.52%	11.29%	24.15%	17.85%	13.39%	9.45%	8.92%	6.82%	3.15%	4.46%
CTR	0.00%	2.73%	5.17%	5.99%	5.10%	8.46%	7.10%	9.26%	8.77%	9.92%

Age

Fig. 3. Age distribution and click-through rate (CTR) by age

The analysis also indicates that the number of days on which a user started Docear impacts CTR (Figure 4). For the first 20 times a user starts Docear, CTR increases. For instance, users who started Docear on one to five days had a CTR of 5.62% on average while users having started Docear on 11-20 days had a CTR of 7.30% on average. This is not surprising assuming that the more often users start Docear, the more information they enter, the better the user models become, and hence the recommendations. However, for users having started Docear on more than 20 days, CTR decreased. For instance, users having started Docear on more than 100 days achieve a CTR of 4.92% on average.

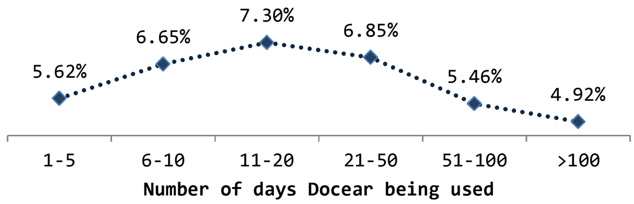
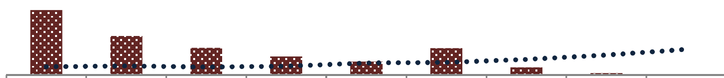


Fig. 4. Click-through rate by the number of days Docear being used

Another analysis brings even more confusion. We analyzed how CTR changes based on the number of recommendations a user received. Based on the above results we assumed that the more recommendations a user received, the lower the CTR would become because users starting Docear often also receive more recommendations. Our assumption was not correct. There is a trend that the more recommendations users see, the higher the CTR becomes (Figure 5, dotted line). Users who received only one recommendation set (i.e. typically ten recommendations) had a CTR of 4.13% while users who saw 21-50 sets had a CTR of 9.91% on average.



	1	2	3	4	5	6-10	11-20	21-50	51-100
# of users	32.00%	19.46%	13.52%	9.34%	6.81%	13.42%	3.99%	1.17%	0.29%
CTR	4.13%	4.56%	4.03%	4.39%	6.07%	6.32%	7.74%	9.91%	12.67%

Number of delivered recommendation sets (10 recs per set)

Fig. 5. User distribution and CTR by number of delivered recommendation sets

4 Conclusion

The analysis showed that demographics and user-characteristics may have a significant impact on click-through rates on (research paper) recommender systems. Although gender had only a marginal impact, age impacted CTR strongly. It made also a difference for CTR whether users were registered or not, how many recommendations they had seen before and how often users had started Docear. However, to fully understand the effects and correlations between the last two factors, more research is required.

We suggest that future evaluations should report on their users' demographics and characteristics in order to create valid and comparable results of recommender systems. Some of these are registered vs. unregistered; intensity of the software being used; and amount of previously shown recommendations. There are certainly further demographics and characteristics that might impact an evaluation such as nationality, field of research, and profession, whose impact should be researched.

Open Data. Due to space restrictions, some data and graphs were omitted in this paper. For those being interested in more details (or validating our research), we publish our data on http://labs.docear.org.

References

1. Naak, A., Hage, H., Almeur, E.: A multi-criteria collaborative filtering approach for research paper recommendation in papyres. E-Technologies: Innovation in an Open World (2009)
2. Middleton, S.E., Shadbolt, N.R., De Roure, D.C.: Ontological user profiling in recommender systems. ACM Transactions on Information Systems (TOIS) 22, 54–88 (2004)
3. Jomsri, P., Sanguansintukul, S., Choochaiwattana, W.: A framework for tag-based research paper recommender system: an IR approach. In: 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, WAINA (2010)
4. Bonhard, P., Harries, C., McCarthy, J., Sasse, M.A.: Accounting for taste: using profile similarity to improve recommender systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1057–1066. ACM (2006)
5. Parsons, J., Ralph, P., Gallagher, K.: Using viewing time to infer user preference in recommender systems. In: Proceedings of the AAAI Workshop on Semantic Web Personalization Held in Conjunction with the 9th National Conference on Artificial Intelligence (2004)
6. Beel, J., Langer, S., Genzmehr, M., Nürnberger, A.: Introducing Docear's Research Paper Recommender System. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, JCDL (2013)
7. Stober, S., Steinbrecher, M., Nürnberger, A.: A Survey on the Acceptance of Listening Context Logging for MIR Applications. In: Proceedings of the 3rd Workshop on Learning the Semantics of Audio Signals (LSAS), pp. 45–57 (2009)