An Analysis of Privacy-Aware Personalization Signals by Using Online Evaluation Methods

Arjumand Younus¹ and M. Atif Qureshi²

¹ Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland
² Centre for Applied Data Analytics Research, University College Dublin, Dublin, Ireland

{arjumand.younus@ucd.ie, muhammad.qureshi@ucd.ie}

Abstract. Personalization despite being an effective solution to the problem of information overload remains tricky on account of multiple dimensions to consider. Furthermore, the challenge of avoiding overdoing personalization involves estimation of a user’s preferences in relation to different queries. This work is an attempt to make inferences about when personalization would be beneficial by relating observable user behavior to his/her social network usage patterns and user-generated content. User behavior on a search system is observed by means of team-draft interleaving whereby results from two retrieval functions are presented in an interleaved manner, and user clicks are utilised to infer preference for a certain retrieval function. This improves upon earlier work which had limited usefulness due to reliance on user survey results; our findings may aid real-time personalization in search systems by detecting a user-related and query-related personalization signals.

1 Introduction

In recent years personalized Web search has emerged as a promising way to improve the search quality through customization of search results for people with different information interests and goals [5,11]. Personalization however is a tricky endeavour due to the involvement of various factors mainly across the user dimension or the query dimension [10]. Of these, the user dimension in particular imposes privacy-related concerns adding a significant amount of complexity on account of users unwilling to share their search engine usage data [7,8]. To further complicate matters, there is the challenge to avoid overdoing personalization, which may harm the user experience; so far work on achieving a balance between producing personalized and non-personalized results considers the query dimension or the user dimension in isolation with each other [10]. This paper argues for a framework that takes into consideration user-related personalization signals in combination with query-related personalization signals; we utilise characteristics of a user’s social network usage data for these signals thereby alleviating the privacy considerations that arise due to personalization.
We conducted a similar study previously \cite{8,10} whereby a user survey was conducted to gather data about users’ personalization preferences and their social network usage habits. However, the reliance on user survey responses limits the usefulness of those studies and in this work we attempt to make useful and reliable inferences through utilization of online evaluation methods.

Online evaluation methods essentially comprise controlled experiments over a user population in order to reach a valid conclusion as to which system is preferred by the users. The two available choices are AB testing and interleaving. We utilize interleaving due the fact that it requires few interactions. To the best of our knowledge, no prior work utilizes online evaluation for detecting signals of personalization adjustments. Such signals can essentially aid the development of privacy-aware personalization that can adaptively adjust the level of personalization for different users. To the aim of identifying these signals, we conduct a user-study whereby users are presented with interleaved search results from two retrieval functions of which one is a personalized version. Within the study, we measure the degree to which personalization is preferred by means of a newly proposed “Personalization Entropy” both by utilising observable clicks across real information needs, and by analysing correlations between various aspects of social network usage behavior/data and personalization preferences. Our results demonstrate that users who engage in high levels of communication over Twitter (in the form of mentions) prefer personalization to a larger degree. Furthermore, with respect to different topical aspects of a query the users prefer personalization for query topics across which their Twitter network contains information in the form of related keywords and keyphrases.

2 Experimental Methodology

As the core of this work is investigating the link between social network usage patterns and personalization preferences, we focus on an efficient personalization mechanism that simply reranks the retrieved documents using a similarity measure between the returned document \(t\) and a user \(u\)'s profile \(p\). Note that we utilise users’ tweets for constructing their user profile in line with previous work \cite{9}.

The experimental procedure is setup as follows. Users were asked to deploy a browser plugin that detects Google Web searches and sends the search query and the unique user identifier to the server, which in turn fetches top 50 search results from Google. The plugin then presents interleaved results to the user whereby the original and personalized rankings are used to produce a combined ranking. We utilise the Team-Draft Interleaving (TDI) \cite{4} method on account of its tendency to minimize a bias towards rankings. A total of 137 users participated in our study who were asked to perform regular day-to-day searches with real information needs, and the study was conducted over a period of two months from 4th March, 2017 to 4th May, 2017.

\footnote{This is essentially on account of the need for methods that are fast enough to operate in real-time which is crucial for online evaluation.}
In order to study the correlations between various aspects of social network usage and personalization potential, we have performed our study by analysing the following aspects of a user’s microblog behavior:

- **Tweeting Behavior:** This aspect captures the amount of user engagement over Twitter in the form of number of tweets, mentions and retweets. We also include tweeting frequency as a mechanism to measure Twitter usage frequency \[8,10\].

- **Twitter Network:** This aspect captures statistics about a user’s Twitter network such as number of followers, number of Twitterers followed, number of Twitterers retweeted, number of Twitterers mentioned etc.

- **Twitter Topical Densities:** This aspect captures statistics about topical content within a user’s tweets such as keywords and their associated frequencies. Note that this is the aspect that captures both the user and query dimension, and is hence crucial to understanding a user’s personalization preferences.

- **Twitter Network Topical Densities:** This aspect is similar to “Twitter Topical Densities” with the difference that it takes into account topical content of a user’s Twitter network such as those he follows, retweets, and mentions.

More specifically, the topical densities mentioned above are computed using the approach proposed by Zhao et al. \[12\] where we first extract topical keywords and topical keyphrases \[4\] by combining tweets of the user under consideration and his/her Twitter network (i.e., those followed, retweeted and mentioned) and these are then used in relation to issued topics within user queries. A detailed description of how topical densities are computed from query topics follows below. To study the effect of social network usage on personalization preferences of users, we devise a measure that captures the personalization entropy measure inspired from the click entropy measure \[2\]:

\[
PersonalizationEntropy(u_{q_{topic}}) = p(\text{per}_u | q_{topic}) \ast \log p(\text{per}_u | q_{topic}) \quad (1)
\]

Here, \(p(\text{per}_u | q_{topic})\) represents the probability that user \(u\) clicks a personalized URL query topic \(q_{topic}\). This measure essentially captures the amount of uncertainty associated with personalization preferences for a given user corresponding to various query topics: a lower personalization entropy value implies that many personalized urls were clicked whereas a higher personalization entropy value implies many non-personalized urls were clicked. Our measure is unique in that it models both the user and the query topic simultaneously. Note that our personalization entropy measure is particularly suited to online evaluation methods and is helpful in the correlation analysis as outlined in Section 3.

A variant of personalization entropy sums over all query topics of user \(u\), and thereby indicates his/her personalization entropy:

\[4\] We extract both unigram and bigram keywords for the purpose of topical densities’ computation.
IV

\[
\text{PersonalizationEntropy}(u) = \sum_{q_{\text{topic}}} p(\text{per}_u | q_{\text{topic}}) \ast log p(\text{per}_u | q_{\text{topic}})
\] (2)

We merge related queries into a single query topic by means of utilisation of Wikipedia categories using a variant of the approach by Gabrilovich and Markovitch [3]; through this method the queries “brexit”, “theresa may europe plan”, and “scientific mobility europe and uk” will be classified as a single query topic. For computation of topical densities from Twitter profiles, we utilise the words (unigrams and bigrams) appearing within search result snippets and check for their occurrence in the Twitter topical keywords; and density scores are then computed from keyword frequencies within Twitter-based profiles. As an example, consider the query topic “brexit” with queries “brexit”, “theresa may europe plan”, and “scientific mobility europe and uk” with the set \( W \) representing words returned in search results and the set \( V \) representing all the keywords and keyphrases in the Twitter profile of the considered user; a topical density score corresponding to this query topic is computed as follows:

\[
\text{TopicalDensity}_{q_{\text{topic}}} = \sum_{w \in W} \frac{n(w, t)}{\sum_{w' \in V} n(w', t)}
\] (3)

Here, \( n(w, t) \) denotes the no. of tweets containing word ‘\( w \)’. Topical density scores essentially capture the normalized term frequencies for terms appearing in search results of a given query topic.

3 Experimental Results and Discussion

The analysed correlations between “Tweeting Behavior” and “personalization entropy” are shown in Table 1, whereas the analysed correlations between “Twitter Network” and “personalization entropy” are shown in Table 2. Note that we apply a threshold to the number of tweets and to the number of Twitterers followed above which statistically significant correlations are observed; here we report the results of the user-level personalization entropy by using the formula in Equation (2).

Table 3 on the other hand shows the correlations between “Personalization Entropy” measure across a user and across different query topics (shown in Equation (1) of Section 2), and topical densities from within query topics.

According to Tables 1 and 2, the number of mentions and the number of Twitterers mentioned have a high, negative correlation with personalization entropy thereby implying that users with a high volume of mentions and a high number of Twitter users mentioned prefer personalized search results. Intuitively, mentions account for highly specific conversations and represent specific user interests. Note that this is in contrast to our previous findings where mentions were found to be negatively correlated with a preference for personalization [8,10]; and
| Analysed Correlations | Dependent Variables       | No. of Tweets | No. of Mentions | No. of Retweets |
|-----------------------|---------------------------|---------------|----------------|----------------|
| **Pearson**           |                           | 0.464         | -0.822**       | 0.261          |
| **Spearman**          |                           | 0.367         | -0.876*        | 0.341          |
| **Regression Analysis** |                          | 0.612         | -0.724***      | 0.387          |

Note *p<.05, **p<.01, ***p<.001

Table 1: Analysis of Correlations between “Tweeting Behavior” Features and “Personalization Entropy”

| Analysed Correlations | Dependent Variables       | No. of Followers | No. of Followed | No. of Mentioned | No. of Retweeted |
|-----------------------|---------------------------|------------------|----------------|------------------|-----------------|
| **Pearson**           |                           | 0.451            | 0.567          | -0.834**         | 0.461           |
| **Spearman**          |                           | 0.345            | 0.472          | -0.921***        | 0.493           |
| **Regression Analysis** |                          | 0.612            | 0.556          | -0.864**         | 0.513           |

Note *p<.05, **p<.01, ***p<.001

Table 2: Analysis of Correlations between “Twitter Network” Features and “Personalization Entropy”

| Analysed Correlations | Dependent Variables       | Topical Density_{user} | Topical Density_{f} | Topical Density_{m} | Topical Density_{r} |
|-----------------------|---------------------------|------------------------|---------------------|---------------------|---------------------|
| **Pearson**           |                           | -0.751***              | -0.567              | -0.876**            | -0.9214**          |
| **Spearman**          |                           | -0.658*                | -0.431**            | -0.913***           | -0.721**           |
| **Regression Analysis** |                          | -0.612**              | -0.556*             | -0.775***           | -0.822***          |

Note *p<.05, **p<.01, ***p<.001

Table 3: Analysis of Correlations between “Twitter Topical Densities” Features and “Personalization Entropy”
this further proves the need for a more reliable means of observing user search behavior instead of reliance on user survey results.

Finally, Table 3 depicts a significant outcome of a user’s Twitter content being highly reflective of his/her information-seeking needs, and personalization entropy is low (meaning high preference for personalized urls). Moreover, topical densities of a user’s Twitter network showing a negative, high correlation with personalization entropy depicts the significance of a user’s associations with a query topic in the form of his/her network and/or his/her content. This association may help in inference of a personalization signal in real-time search systems and this could alleviate the need for utilising private user content such as desktop data, search history etc.

References

1. F. Cai, S. Wang, and M. de Rijke. Behavior-based personalization in web search. Journal of the Association for Information Science and Technology, 68(4):855–868, 2017.
2. Z. Dou, R. Song, and J.-R. Wen. A large-scale evaluation and analysis of personalized search strategies. In Proceedings of the 16th international conference on World Wide Web, pages 581–590. ACM, 2007.
3. E. Gabrilovich and S. Markovitch. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In IJcAI, volume 7, pages 1606–1611, 2007.
4. F. Radlinski, M. Kurup, and T. Joachims. How does clickthrough data reflect retrieval quality? In Proceedings of the 17th ACM conference on Information and knowledge management, pages 43–52. ACM, 2008.
5. D. Sontag, K. Collins-Thompson, P. N. Bennett, R. W. White, S. Dumais, and B. Billerbeck. Probabilistic models for personalizing web search. In Proceedings of the fifth ACM international conference on Web search and data mining, pages 433–442. ACM, 2012.
6. J. Teevan, S. T. Dumais, and D. J. Liebling. To personalize or not to personalize: modeling queries with variation in user intent. In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, pages 163–170. ACM, 2008.
7. Q. Wang and H. Jin. Exploring online social activities for adaptive search personalization. In Proceedings of the 19th ACM international conference on Information and knowledge management, pages 999–1008. ACM, 2010.
8. A. Younus, C. O’Riordan, and G. Pasi. An investigation into the correlation between willingness for web search personalization and sns usage patterns. In IIR, pages 80–84, 2014.
9. A. Younus, C. O’Riordan, and G. Pasi. A language modeling approach to personalized search based on users’ microblog behavior. In ECIR, pages 727–732. Springer, 2014.
10. A. Younus, C. ORiordan, and G. Pasi. Predictors of users willingness to personalize web search. In International Conference on Flexible Query Answering Systems, pages 459–470. Springer, 2013.
11. H. Zamani, M. Bendersky, X. Wang, and M. Zhang. Situational context for ranking in personal search. In Proceedings of the 26th International Conference on World
Wide Web, pages 1531–1540. International World Wide Web Conferences Steering Committee, 2017.

12. W. X. Zhao, J. Jiang, J. He, Y. Song, P. Achananuparp, E.-P. Lim, and X. Li. Topical keyphrase extraction from twitter. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1, pages 379–388. Association for Computational Linguistics, 2011.