Characterizing Cross-Domain Search Behavior*

Rishabh Mehrotra†
Dept. of Computer Science
University College London
London, UK
r.mehrotra@cs.ucl.ac.uk

Paul N. Bennett, Susan Dumais, Jaime Teevan
Microsoft Research
One Microsoft Way
Redmond, WA, USA
{pauben,sdumais,teevan}@microsoft.com

ABSTRACT

The heterogeneous nature of information needs and information sources requires techniques that efficiently combine and leverage insights from diverse data sources. Moreover, user interaction from different sources provides different perspectives on user's interests and preferences. In this work, we consider user interaction data from different verticals (news, search) and characterize behavioral differences among users. Traditional research on cross-domain methods has focused on leveraging insights for the same user from different domains. Instead, in this work, we highlight the need to consider user groups based on the cross-domain information and show that users from these groups behave differently. We investigate a number of search characteristics including re-querying behavior, topical spread of user interests and the overall popularity of queries across the different user groups, and demonstrate how considering different user groups has implications for evaluating and designing cross-domain personalization and recommendation approaches.

CCS Concepts

•**Information systems** → **Information retrieval;**

Keywords

Cross domain, web search, user groups

1. INTRODUCTION

In recent years, major search engines have extended their services to include search on specialized domains or verticals (e.g., news, books, shopping, travel, and local search) or media types (e.g., images and video). As the heterogeneity of data sources on the web increases, cross-domain personalization and recommendation is becoming an increasingly important research topic. Because verticals support particular domains of interest, the distribution of queries for each is likely to be different. This means that user interaction information from different verticals can provide different perspectives on a user's interests and preferences.

While the focus of most previous cross-domain research has been on leveraging cross-domain insights to aggregate user information for the same user, we instead analyze behavioral differences across users from the different domains. More specifically, we consider search usage data from two verticals (web search and news) and investigate a number of behavioral characteristics for users active in the two domains. We begin by investigating how the same user accesses information on different domains (4.1). Based on the volume of search activity in the different domains, we categorize users into two different groups (regular users and frequent news users) and characterize behavioral differences in search across these two groups. We focus on re-querying behavior (4.2), topical spread in interest profiles (4.3) and query popularity effects across the user groups (4.4).

Our results indicate that across different domains, users tend to click on similar *topics* which suggests user activity in one domain can inform a different domain. We show that frequent news users re-query less than regular users, have a lower topical spread in their interest profiles than regular users, and (in terms of query popularity) issue fewer head and tail but more torso queries on average. By better understanding such user-level behavioral differences, our findings can inform the evaluation and design of cross-domain personalization and recommendation systems.

2. RELATED WORK

As the amount of online information increases, personalization plays an increasingly important role in delivering the right information to people at the right time. Most personalization systems build profiles and recommend items from within the same domain (e.g., movies, music, books, web pages or news). However, people seek information across many different domains, when purchasing items, connecting with friends via social media, or seeking information in web search and news verticals. Over the last few years there has been growing interest in cross-domain recommendation [3, 5, 6, 7, 9]. The idea of cross-domain recommendation is to use information accumulated in one domain (either explicit ratings or implicit behavioral signals) to improve the quality of recommendations in another domain. Although many techniques have been used for cross-domain prediction (e.g., transfer learning, hybrid recommenders and ensemble learning), matrix factorization is the most common approach for collaborative filtering-based recommender systems both within and across domains [6]. Hu et al. [3] used a triadic matrix factorization approach to jointly model users, items, and domains. They incorporated both explicit user ratings and implicit user preferences in two different domains - Amazon (books,

*Work in progress

†Work done during an internship at Microsoft Research.

Heterogeneous Information Access (HIA) Workshop, SIGIR '16 Pisa, Italy

ACM ISBN . . . $15.00

DOI:

music, dvd and vhs) and within a social networking site (from the KDD cup in 2012). More closely related to our research, Low et al. [7] developed a model to support personalization across multiple Yahoo! properties (e.g., news, search, ads, homepage). They used a hierarchical Bayesian model and LDA features to represent users' latent interests based on the content of pages they visited. They found significant gains when predicting Yahoo! News preferences from interactions on the Yahoo! homepage (and vice versa).

Although it seems likely that the success of cross-domain personalization will depend on the nature of the source and target domains, there has not been much research in this area. Using data from the Yelp academic dataset, Sahebi & Brusilovsky [9] showed that the extent to which information in one domain is useful in predictions in another domain depends on factors such as the number of users in common across the two domains, data density and the content similarity. Data from outside of the target domain is also used to mitigate the cold-start and data-sparsity problems in collaborative filtering approaches [4, 10].

Because people use search engines as a way to connect with all types of information, search engine query logs can provide unique insight into people's interests across many domains. User profiles generated from such query logs are critical to the success of personalization. Profiles can be constructed from explicit judgments such as ratings of books, movies, hotels, or news, or implicit information such as a user's queries or browsing patterns when interacting with systems. For example, the exact query text that a user has issued before has been shown to be very successful for web search personalization [14]. The features used in constructing profiles can cover different time periods (longer term preference or current acute information need), and different levels of representation (from specific words or entities to more general topics, tags, latent variables, or demographics). Entities have been used in social and news recommendation [15], and topic models are widely used with the most common methods of deriving topics being classification into ODP topics [2], LDA [7], and tags [11]. In the domain of web search, Bennett et al. [2] showed how profiles constructed at different time scales and using different abstractions of a person's interests (specific queries and URLs, general topics from ODP) combine to produce the best personalized search rankings.

In work currently under review [8], we investigated the effects of profile domain and representation in news query recommendation, and identified and quantified the impact of several factors on the accuracy and coverage of the news query recommendation. Additionally, we found evidence of behavioral differences across user groups. This paper is a follow-up to that work in which we characterize those behavioral differences in interaction patterns across user groups.

3. DATA

In order to characterize the differences in search behaviors across user groups and domains, we use a proprietary dataset of anonymized logs of users of Microsoft's search engine, Bing, and its news vertical (www.bing.com/news). We use a sample of 6 weeks of log data gathered from February to April 2015 in the English-speaking US market. We considered the subset of active users who searched in both the web and news vertical, and had at least 300 queries in either the search or news domains. We partitioned these users into two groups - *Frequent News users* who issued at least 150 queries in the news domain and *Regular users* for which no restriction was put on the number of news domain queries. Our sample of log data contains 2 million users with over 250 million search sessions and over 1.1 billion search queries. On average, people in our sample issued 500 queries over 100 sessions in total. We partition the data to keep the training/profile building data separate from the data used for evaluation. Specifically, as shown in Figure 1, we use the first 4 weeks of user data to build user profiles and the fifth and sixth week for evaluation.

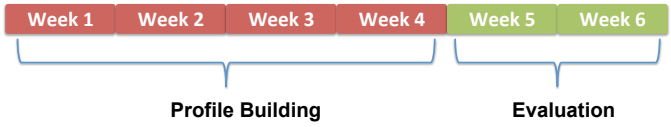


Figure 1: Timeline for profile construction and evaluation.

3.1 Profile Domain

To leverage insights across different domains, we build user profiles from user's past queries from three different sources: news search queries, general web search queries that trigger news answers, and general web search queries.

1 **News Search Queries (News)**:
In order to search for news events and stories, users issue queries in the news vertical. For each person, we construct one profile from queries they issued in the news vertical and refer to this as a user's News profile.

2 **Search Queries that Trigger News Answer (SNA)**:
People also search for news within a general web search engine. As an alternative we construct another profile using news-related queries issued to a general web search engine. News-related queries are identified as queries that trigger a news answer on the search result page. We call this profile the Search-News Answer profile (SNA).

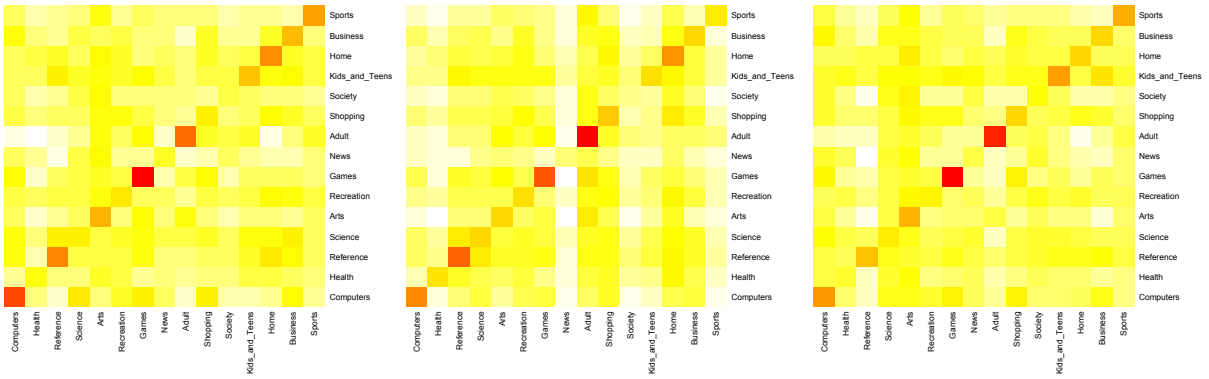
3 **General Web Search Queries (Search)**:
Finally, we construct a third profile from all of a person's search activity from a general web search engine, whether they triggered a news answer or not. This provides a general profile of a person's interests. We refer to this as a user's Search profile.

3.2 User Groups

We hypothesize that search behavior varies across users who use different verticals. We quantify behavior over two group of users:

1 **Regular Users**:
For the first group we selected users from all users who issued more than 300 queries to either the news vertical or the web search during the first four weeks of our log-data. This threshold is below the average number of queries per user over four weeks, but ensures that there is a minimal amount of evidence for making recommendations. We refer to this sample as Regular Users since it represents on average performance for a randomly selected user who *regularly* uses the search engine.

2 **Frequent News Users**:
The second group consists of those users who frequently search for news using the news vertical. Since these users consume news more frequently than the typical regular user, we hypothesized that their behavior may highlight interesting differences between general search behavior and news seeking behavior. To sample these users, we again selected users



(a) Search - News (b) Search - SNA (c) SNA - News

Figure 2: Cross-Domain Topical Effects: The ODP categories of documents clicked by the same user across different domains. The colors in the heatmaps represent the normalized point-wise mutual information (NPMI) gains, where darker means higher.

who issued more than 300 queries to either the news vertical or the web search, and additionally required that each user had issued at least 150 queries in the news vertical. Following such a user sampling approach gives us a subset of Regular Users which we refer to as Frequent News Users.

Given the user group and profile domain information, we next investigate four different research questions to characterize behavioral differences across the user groups.

4. CHARACTERIZING CROSS-DOMAIN BEHAVIOR

With search data from over 2 million users, we investigate the differences in search behavior across different domains and user groups. We begin by looking at cross-platform clicks, and proceed to quantifying various search characteristics for the different user groups.

4.1 Cross-Domain Clicks

In this section, we ask the research question, do regular users have the same set of topical interests across domains? That is, we investigate the topics of documents that users click on and compare them across different domains. We use the top level categories from the Open Directory Project (ODP) to represent the document topics. To train topic classifiers, we followed a similar approach to that used in Bennett et al. [1]. In particular, using a binary logistic regression classifier text classifier per topic we classify each clicked URL to the topic with the maximal predicted posterior probability of belonging to a topic across all binary topic predictions. To classify each query, we then aggregate topics from all clicked URLs and assign the most commonly clicked topic as the query's topic.

We look at how the same user accesses information on different domains. For each domain, we characterize a user using strongest ODP category associated with that user's search activity by aggregating hard counts over the topics of clicked URLs and selecting the top topic. For each pair (M_p, M_q) of the three domains, (Search, SNA and News), we calculate the Normalized Point-wise Mutual Information (NPMI), to measure the association of a pair of ODP categories (o_i, o_j) conditioned on the pair of domains [12]. The NPMI value is denoted as the output of function N, using the following formula:

$$N(o_i, o_j|M_p, M_q) = \frac{log\frac{P(o_i,o_j|M_p,M_q)}{P(o_i|M_p,M_q)P(o_j|M_p,M_q)}}{-logP(o_i, o_j|M_p, M_q)} \quad (1)$$

The value of N ranges between -1 and 1, where 1 indicates that the two categories completely co-occur given the two domains, -1 indicates that the two categories occur separately but not together, and 0 indicates that the two categories are independent. We visualize the cross-domain NPMI values for each pair of domains and across all top-level ODP topics as heatmaps in Figure 2. The right diagonals of the three heatmaps have the darkest colors, indicating that, overall, the same category is clicked by the user across all pairs of domains. The consistency of the topics viewed cross-domains by the user suggests that the users' search history might provide useful signals for personalizing their news domain results (and vice-versa). Additionally, we observe higher NPMI values for a common set of topics across the different pairs of domains, which indicates that certain topic categories (e.g. *Games*, *Computers* and *Adult*) witness more consistent user interest across domains than others.

4.2 Re-Querying Behavior across User Groups

In this section, we ask the research question: do regular users and frequent news users attempt to re-find related information with a similar frequency? People often repeat Web searches, both to find new information on topics they have previously explored and to re-find information they have seen in the past. In this section, we investigate a user's tendency to re-issue the same query, query terms, bigrams and ODP topic associated with a query. More specifically, for the queries, terms, bigrams and topics extracted from the first four weeks of logs, we compute the percentage of those that were observed from the same user during the fifth and sixth week. In Figure 3, we compare the re-querying behavior across the two user groups.

We observe a consistent difference in re-querying behavior among the two user groups. Across all four representations considered, regular users re-query more than frequent news users. Regular users re-query the exact query about 5% more than frequent news users, whereas the corresponding figure for topics is 3%. Across the representations considered, we observe the highest re-querying for the ODP topics, with users re-querying for the same topics 80% of the time, compared with 35-40% for queries, 40-45% for terms,

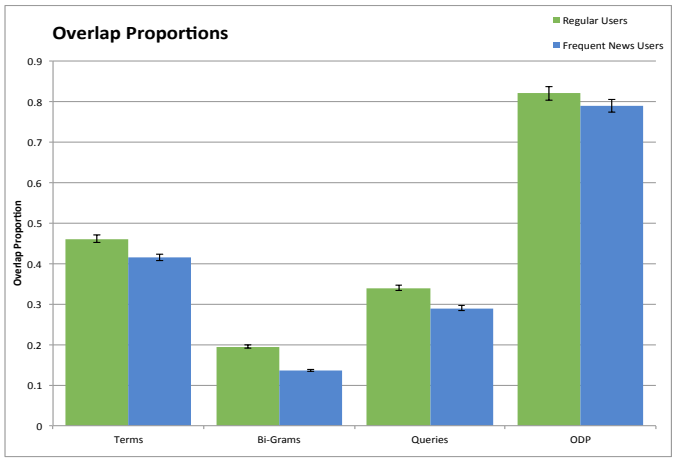


Figure 3: Overlap proportion of queries, terms, bigrams and ODP topics between a user's evaluation and profile building search histories.

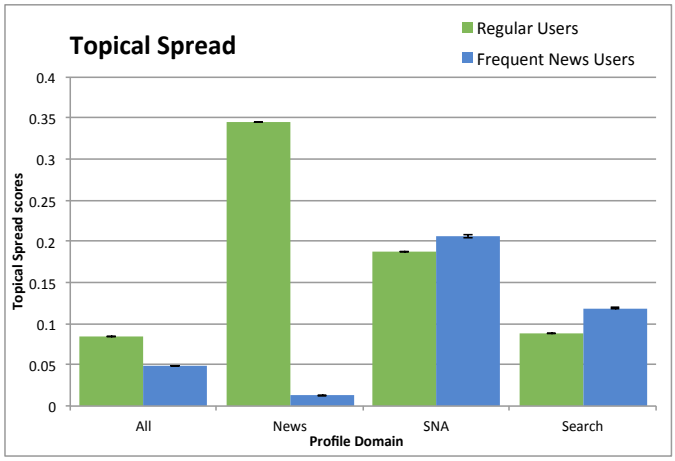


Figure 4: Analyzing the topical spread of interests across the user groups.

and 15-20% for bigrams. Our findings on re-querying behavior by query are consistent with established results on re-finding behavior of search users [13].

4.3 Topical Spread across User Groups

Next we ask the research question: do regular users and frequent news users have similar breadths of topical interests? To answer this, we examine the breadth of interests across user groups and domains by computing the average number of topics per query in their profile. We consider the level 2 of the ODP topic hierarchy (200+ categories) to measure topical differences at a finer granularity. In Figure 4, we plot the topical spread. We compute topical spread for a domain and group pair as follows:

$$spread(d, g) = \frac{1}{|U_{d,g}|} \Sigma_{u \in U_{d,g}} \frac{t_{u,d}}{q_{u,d}} \quad (2)$$

where $U_{d,g}$ is the set of users in group g and domain d, $t_{u,d}$ is the number of topics in user u's profile constructed from domain d, and $q_{u,d}$ is the number of impressions in user u's profile in domain d. Note that the ratio inside the sum will tend to zero as the number of user queries in a domain increases if all of those queries are on the same topic[1].

We observe that over all activity as well as for the News and SNA domains, the regular user (Regular) has higher topical spread in their profile than frequent news users with a slight 2% difference in SNA and a very noticeable 30% difference in the News domain. Focusing on the difference in the News domain, when a regular user issues a query in the news vertical, they search over a broad set of topics on average. In contrast a frequent news vertical user searches the news vertical over a small number of topics. This is not a given since a frequent news user could avidly pursue all news. Instead the frequent news user seems to be more of a "news hound" who is highly interested in news on very few topics rather than news-obsessed with all news.

In contrast, in the Search domain, frequent news users tend to have a wider topical spread with a 3% difference in spread over regular users. These results highlight that frequent news users tend to query about a very small set of topics in the News domain while their general search interests are broader than the regular user.

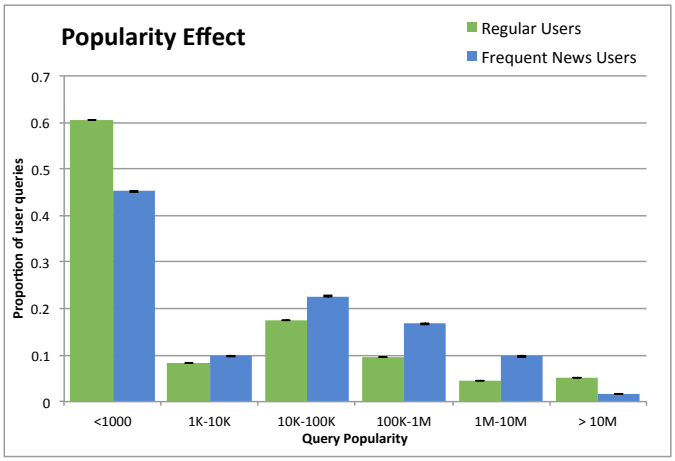


Figure 5: The proportion of user queries across the different popularity bins.

4.4 Query Popularity across User Groups

In this section, we focus on the research question: does the popularity of a query similarly influence the interests of regular users and frequent news users? Query popularity is an important indicator used to rank recommendations to users. In addition to analyzing the re-querying behavior and topical spread across the user groups, we also investigate how popular the queries from the two user groups are. We use the overall query volume in the training phase for each query as a proxy for its popularity, and we divide the popularity into 6 bins of increasing order of popularity. For each user, we then find the proportion of their queries belonging to the different popularity bins and average the scores across all users. Figure 5 presents the proportion of queries that fall into the six popularity bins for the two user groups.

We observe that regular users (Regular) issue more "tail" or niche queries (i.e., queries with frequencies <1000) than a frequent news

[1]Since the numerator is bounded by the number of overall categories, if the number of user queries in a domain is extremely large (e.g. 10× the number of categories), then an entropy measure of the user topical distribution is preferable as a measure of spread. Since that is not the case for us, this measure is a simple measure of spread that enables us to both normalize for differences in the absolute number of queries issued per user and the number of users active in a group.

users. They also issue more "head" or very popular queries (i.e., queries with a frequency of >10 million). The frequent news users issue more queries in the "torso" (i.e. intermediate popularity). This may indicate that frequent news users are already aware of head information on news topics and instead focus in on stories that have a slightly broader appeal without being of universal interest. Also, people who don't frequently search news seem to do so for very common queries (probably responding to ongoing news events) and niche interests.

5. CONCLUSION

We considered user interaction data from two different domains (web search and news) and characterize behavioral differences among user groups (Regular Users vs. Frequent News Users). We investigate a number of search characteristics including re-querying behavior, the topical spread of user interests and the overall popularity of the issued queries across the different user groups and highlight the need of considering user groups based on frequency of vertical use when personalizing or evaluating search and recommendation across domains.

6. REFERENCES

[1] P. N. Bennett, K. Svore, and S. T. Dumais. Classification-enhanced ranking. In *Proceedings of the 19th international conference on World wide web*, pages 111–120. ACM, 2010.

[2] P. N. Bennett, R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisyuk, and X. Cui. Modeling the impact of short-and long-term behavior on search personalization. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 185–194. ACM, 2012.

[3] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu. Personalized recommendation via cross-domain triadic factorization. In *Proceedings of the 22nd international conference on World Wide Web*, pages 595–606. International World Wide Web Conferences Steering Committee, 2013.

[4] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, (8):30–37, 2009.

[5] C.-Y. Li and S.-D. Lin. Matching users and items across domains to improve the recommendation quality. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 801–810. ACM, 2014.

[6] B. Loni, Y. Shi, M. Larson, and A. Hanjalic. Cross-domain collaborative filtering with factorization machines. In *Advances in Information Retrieval*, pages 656–661. Springer, 2014.

[7] Y. Low, D. Agarwal, and A. J. Smola. Multiple domain user personalization. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 123–131. ACM, 2011.

[8] R. Mehrotra, P. N. Bennett, S. Dumais, and J. Teevan. Profile domain and representation effects in personalized news query recommendation. In *Under review*.

[9] S. Sahebi and P. Brusilovsky. It takes two to tango: An exploration of domain pairs for cross-domain collaborative filtering. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 131–138. ACM, 2015.

[10] B. Shapira, L. Rokach, and S. Freilikhman. Facebook single and cross domain data for recommendation systems. *User Modeling and User-Adapted Interaction*, 23(2-3):211–247, 2013.

[11] Y. Shi, M. Larson, and A. Hanjalic. Tags as bridges between domains: improving recommendation with tag-induced cross-domain collaborative filtering. In *User Modeling, Adaption and Personalization*, pages 305–316. Springer, 2011.

[12] M. Shokouhi and Q. Guo. From queries to cards: Re-ranking proactive card recommendations based on reactive search history. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 695–704. ACM, 2015.

[13] J. Teevan, E. Adar, R. Jones, and M. A. Potts. Information re-retrieval: repeat queries in yahoo's logs. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 151–158. ACM, 2007.

[14] J. Teevan, D. J. Liebling, and G. Ravichandran Geetha. Understanding and predicting personal navigation. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 85–94. ACM, 2011.

[15] Z. Zhao, Z. Cheng, L. Hong, and E. H. Chi. Improving user topic interest profiles by behavior factorization. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1406–1416. International World Wide Web Conferences Steering Committee, 2015.