Investigating Personalized Search in E-Commerce

Dietmar Jannach
TU Dortmund, Germany
dietmar.jannach@tu-dortmund.de

Malte Ludewig
TU Dortmund, Germany
malte.ludewig@tu-dortmund.de

Abstract

Personalized recommendations have become a common feature of many modern online services. In particular on e-commerce sites, one value of such recommendations is that they help consumers find items of interest in large product assortments more quickly. Many of today's sites take advantage of modern recommendation technologies to create personalized item suggestions for consumers navigating the site. However, limited research exists on the use of personalization and recommendation technology when consumers rely on the site's *catalog search functionality* to discover relevant items.

In this work we explore the value of *personalizing search results* on e-commerce sites using recommendation technology. We design and evaluate different personalization strategies using log data of an online retail site. Our results show that considering several item relevance signals within the recommendation process in parallel leads to the best ranking of the search results. Specifically, the factors taken into account include the users' general interests, their most recent browsing behavior, as well as the consideration of current sales trends.

Introduction

Recommender systems are without a doubt one of the most successful applications of artificial intelligence technology that made its way from academic research to wide-spread industrial use. Automated recommendations are today a pervasive part of our online user experience and employed to recommend, for example, things to buy on e-commerce sites, friends to connect with on social networking sites, or content to consume on media streaming sites.

Recommendations can in general serve a variety of different purposes as discussed recently by (Jannach and Adomavicius 2016). In particular on e-commerce sites like Amazon.com, one key utility of recommendations is to help consumers find items of interest within large product catalogs more quickly, for instance, when the system displays items that are similar to the one the consumer is currently inspecting. Recommendations of this latter type – showing similar items – support different possible reasons why a consumer visits the site, namely "knowledge-building", "directed buying" and "search and deliberation", see (Moe 2003). While recommenders can also support other intents like "hedonic browsing" and item discovery, the work in this paper focuses on the goal-oriented reasons of a customer visiting a site.

Automated recommendations that are made in the context of a currently inspected item are however only one possible means to help consumers build up knowledge or understand the space of options. The typical functionalities implemented by most e-commerce sites include predefined, hierarchical catalog navigation structures and, as the focus of this work, a catalog search functionality.

Typical search engines on e-commerce sites allow users to change the order in which the search results are presented with pre-defined sort criteria. Common sort orders include "by sales rank", "by average customer rating", "by price", and often "by relevance", where the relevance function in this case is typically not further explained and might be a mix of relevance for the consumer and profit for the site.

Most of today's online shops do not offer an option for users to sort the results according to their past personal preferences or shopping behavior. It is however intuitive to assume that taking the consumer's past behavior into account when ranking the results can be useful, e.g., by ranking items up in the list that correspond to the consumer's typical price preferences, by listing products of the consumer's preferred brands first, or by promoting consumables that the user has already purchased in the past.

In this work, we aim to explore and quantify the potential value of ranking the search results in a personalized way. The underlying and obvious assumption is that the search functionality is more helpful for users when the more relevant items appear higher up in the result list. Technically, we will explore the use of different common recommendation techniques as well as additional problem-specific heuristics to personalize the search results for the individual consumer. To evaluate the effectiveness of the proposed techniques, we will use an offline experimental design using log data of a German online retailer of products for babies and small children and compare the results with existing approaches from the general field of web search personalization and recommendation.

The paper is organized as follows. Next, we describe our research methodology and the dataset used for the evaluation. Then, we will summarize the tested algorithmic approaches and report the results of the empirical evaluation. The paper ends with a discussion of related works.

Research Methodology

General Setup

The overall problem setting of e-commerce search personalization is in various ways similar to the problem of web search personalization. The typical inputs to a search personalization algorithm are a search string, a collection of documents, information about the past search behavior of individual users, and possibly additional context information. The computational task of algorithms is then to filter and rank the documents in a personalized way. Finally, measures like precision or recall can be used to compare algorithms.

Our research setup is slightly different in two ways. First, we consider item filtering to be a black box, i.e., we make no assumption about how the search strings are matched with the objects. The algorithms that we investigate in this work can therefore be applied in combination with any existing search component of an e-commerce site. Second, the main input provided to the algorithms for personalization is the recorded navigation behavior of users – data that is usually available in practical settings. The task of the algorithms is then to find the best possible ranking of a given set of objects and personalization therefore only affects the ranking of the items but not their selection.

Data and Ground Truth

The dataset that we will use in our research was provided to us by a major German online retailer of goods for babies and small children. The data comprises an anonymized subset of the navigation logs of the shop's web server collected over the period of one month in early 2016.[1]

Dataset Features The time-stamped log entries of the dataset are either marked as *page requests* or as *add-to-cart* events, which we use to determine purchase events. Each entry has an accompanying URL, which gives us more details about the requested page or the item that was added to the cart. By analyzing the URLs, we can further classify the page requests into the following relevant main categories:

- item view events,
- category browsing events,
- search events (including a search string),
- shopping basket checkout events.

Each log entry furthermore has a customer ID assigned, which is approximately derived from a tracking pixel. In addition, for each item we know some basic data like the category it belongs to, the brand, and the item's textual description.

Determining the Ground Truth In order to evaluate to what extent an algorithm is successful in generating a good ranking of the search results, we need a "ground truth" (gold standard) that defines whether or not an item is relevant for a search query or not. In our case, we consider an item as relevant and the search as successful whenever the consumer actually purchased an item that was returned by the search *in the same session*. Since the results returned by the site's engine are not available in the log data, we apply the following heuristic approach to re-construct a ground truth dataset from the log data.

1. For each search action in the logs, we repeated the search on the online shop through an automated agent.
2. The agent collected all results returned by the shop, using all available search orderings that were provided by the shop (e.g., by popularity, recency etc.).
3. We then inspected all navigation actions of a user after the search action in the same session[2]. Whenever a user actually *purchased* an item that was part of the search results in this session, we consider the search to be successful.

A successful search in this interpretation does not necessarily require that a user purchases an item *immediately* after inspecting the search results. If the user inspected various other items before buying one item in the same session, we still count the search as a (potential) success. Furthermore, we chose actual purchase actions as success indicators instead of item views, because the item view events can be biased by the ordering of the results, i.e., users typically click more often on the top-ranked items even though they are not necessarily the most relevant ones in the end. In that sense, our reconstruction heuristic is very conservative.

Evaluation Approach

Algorithm Task and Metrics At the end of the process of determining successful search actions, we can apply standard IR evaluation procedures. For each search query, we are given a set of items returned by the shop's search engine and the information which item of the set was actually purchased. Furthermore, we are given different alternative rankings (by popularity, by relevance etc.) provided by the shop and can apply the common measures hit rate (recall@n) and the MRR to determine the capability of an algorithm to rank the single relevant item at the top places of the result list.

The task of the recommendation approaches that we investigate in our research is to generate an alternative ranking of the items that were retrieved by the site's search tool. The problem of the generation of a candidate set is therefore not in the scope of our work. As a result, the algorithms that we analyze in this paper can be used independently of the internal mechanisms that are used by the site for retrieving relevant products for a given search query.

Data Sampling and Measurement Procedure As usual in e-commerce environments, we observe many users who have visited the shop only once and never made any purchase. To be able to apply personalization strategies, we created a subset of 5,000 active users of the shop for which we have at least 100 log actions and who have bought at least five items during the data collection period. Table 1 shows the characteristics of the resulting dataset.

We apply an evaluation protocol for session-based log data similar to (Jannach, Lerche, and Jugovac 2015). We first

[1] The data was sampled in a way that no conclusions about the visitors or the business numbers can be drawn.

[2] We split the log actions into sessions using a 30-minute period of inactivity as an indicator that a new session started.

Table 1: Characteristics of the resulting dataset

Number of users	5,000
Number of items	23,052
Number of purchases	42,905
Number of item view events	419,945
Number of successful searches	3,300
Avg. nb. of sessions per user	13.7
Avg. nb. of views per session	6.9
Avg. nb. of purchases per session	0.6
Avg. nb. of successful searches per user	0.66

split the time-ordered log data of each user into two parts. The most recent $20\,\%$ of the sessions containing successful searches form the test set and the other sessions are used as the training set. In the test phase, we therefore only consider users for which we have determined successful search sessions in the previous step. We iterate over all these successful search sessions in the test data, compute the personalized result rankings, and apply the above mentioned IR measures, which are at the end averaged across all tested users. To avoid random effects, we apply a five-fold user-wise cross-validation procedure. Note that using a sliding-window protocol over the time axis was not meaningful in our situation as we only have log data for one month.

Empirical Results

Compared Algorithms

The online shop from which we obtained the data provides four ways of ranking the results: by sales numbers, by some unknown relevance ranking, by price, and by the average consumer rating. The set of alternative ranking strategies used in our experiments is shown in Table 2.

We evaluate existing techniques from web search personalization, collaborative filtering, context-aware personalization strategies, as well as a heuristic that considers recent sales trends. For all algorithms we tuned their parameters in a manual process to optimize the hit rate for the complete dataset.

We have furthermore tested a variety of hybrid approaches to investigate their effectiveness and to illustrate the relative importance of the different aspects. We include the results of a limited selection (including the best performing one) of the various experiments in Table 3.

Evaluation Results

Table 3 shows the detailed results of our analysis using a dataset consisting of 5,000 users who frequently visited the site during the data collection period. We report the results for various configurations regarding the minimum number of search results. The ranking of the algorithms at the end was however very consistent across all configurations.

Configurations Since the effectiveness of different algorithms might depend on the number of existing search results for a query, we report the results for different configurations. The first row of Table 3 shows the threshold regarding the minimum number of results to be returned by the shop's search engine. The lowest threshold we consider is five, which means we do not consider search tasks in our evaluation that led to less than five results, because re-ranking these results which are all displayed on one single page might not add much value for the user. Considering higher thresholds can be informative in particular in the context of the reminding strategy, which might only be able to re-rank a smaller part of the search results as only a limited amount of past interactions with the items in the result set by the individual customer might exist.

The hit rates and MRR values obviously become smaller when the search results comprise more objects. The absolute values of the hit rates are generally comparably high, e.g., at around 0.4 for the most simple baseline method, because in many cases not too many results are returned and even a random recommender would place a number of relevant objects into the top 10 lists.

Performance of the Shop's Methods The best-performing method among the ranking strategies that are currently available on the site is the ranking according to the sales numbers, i.e., the *Bestseller* strategy. This strategy is however outperformed by any of the other algorithms tested in this work as shown in Table 3. We omit the detailed results for the other current ordering strategies that are implemented on the site.

Performance of Search Personalization Methods The *PClick* method, which is based on analyzing past successful searches, only works slightly better than the *Bestseller* strategy. The comparably weak performance of this strategy is most probably caused by the limited dataset size. Even though each of the top 25 search terms in our dataset was used more than 1,000 times by site visitors during the data collection period, there is a huge long tail of rarely used search terms and only 1% of the users entered one specific query repeatedly. Considering this fact, the PClick method often cannot make any valuable prediction and then defaults to the Bestseller strategy. We however believe that with a larger dataset, the performance of PClick would improve. While it might not be better than the "winning" strategies, it might represent a valuable component in a hybrid approach.

The content-based method (CB), which combines short-term and long-term models, works relatively well and outperforms, for example, the *C-KNN* method, whose recommendations are based on the user's short-term behavior. It is, however, not as accurate as the more elaborate content-agnostic learning-to-rank method BPR, which focuses on the user's long-term preference model. The somewhat limited performance of the CB method can in parts again be attributed to the comparably short data collection period and the fact that the long-term history is limited to one month. However, since user interests in the domain of baby goods might naturally change quite fast over time, this effect might be limited and focusing on more recent interactions might even be useful. Another aspect that limits the potential of the CB method are the often very short product descriptions. We have tested Doc2Vec as an alternative representation (Le and Mikolov 2014), which however led to even worse results.

Performance of Collaborative Filtering and Session-based Approaches The *C-KNN* method, which works very well in other recommendation scenarios, does not work as well as expected in the search personalization scenario. One reason could be that there are sessions which included a successful search but no other user actions. This happens when users arrived at the site, immediately searched for a specific item (e.g., a consumable) and directly proceeded with the purchase. In such situations, the available information might be too limited to find neighboring sessions.

The modern *BPR* method and the very simple, session-based *Feature Matching* (FM) technique lead to very similar results and *BPR* leads to slightly better MRR values, possibly due to the coarse-grained re-ranking strategy of the FM method. In sum, this observation corroborates existing insights, e.g., from (Matthijs and Radlinski 2011), that both long-term models (*BPR*) and short-term interests (*FM*) should be considered in search personalization.

Performance of Recommending Reminders and Trending Items Reminding users of items that they have recently inspected is the most successful individual strategy in our experiments. This indicates that users often browse items, which they later on – e.g., in the next session – retrieve again through the search functionality before making a purchase. Reminding users of known products in recommendation lists might not necessarily be the most valuable business strategy as reminders do not help consumers discover new areas of the item catalog. Nonetheless, reminders in search results help users locate their items of interest fast (e.g., consumables), which contributes to the usability of the site. In practical settings, one might consider to use hybrids

Table 2: Summary of Compared Result-Ranking Algorithms

Web-Search Personalization Techniques	
PClick	The rationale of this method is that users often search for the same things multiple times (Dou, Song, and Wen 2007). The method ranks those items up in the list that the user has clicked on in previous search sessions with the same or a similar search term.
Content-Based (CB)	An approach based on (Matthijs and Radlinski 2011) which relies on TF-IDF representations of the textual descriptions of the shop items. We compute weighted combinations of the mean TF-IDF vectors of the long-term and short-term models of each user and rank the search results based on the cosine similarity of the item descriptions and the user model.
Collaborative Filtering and Session-Based Recommendation	
BPR	A modern learning-to-rank collaborative filtering method designed for implicit feedback (Rendle et al. 2009), which we used to learn long-term interest models. We manually fine-tuned the parameters ending up with 150 features and 150 training iterations using an optimized learning rate and regularization factors.
C-KNN	This contextualized k-nearest-neighbor method recommends items from past shopping sessions that are similar to the most recent sessions of a user. Cosine similarity is used as a distance measure for binary item vectors (see (Lerche, Jannach, and Ludewig 2016)). The short-term profiling approach proved effective for e-commerce and also music recommendation in the past (Hariri, Mobasher, and Burke 2015).We systematically optimized the parameters. Using a neighborhood size of 10 and considering a user's last two sessions led to the best results.
Feature-Matching (FM)	This method proved effective for session-based recommendation in (Jannach, Lerche, and Jugovac 2015). It compares features (category, subcategory, and brand) of the items to rank with those of items the user inspected in the last session. Items that are a better content-wise match for the current session are ranked up.
Personalized Reminders and Global Trends	
Most-Recent Reminder (MR)	Reminding users of items that were of recent interest to them can be helpful (Lerche, Jannach, and Ludewig 2016), in particular when consumables are sold on the online shop. The MR strategy ranks items up (in reverse chronological order) that the user has interacted with in recent sessions. Considering the last 6 sessions for reminding proved to lead to the best results.
Trending-N	This strategy considers the last n days before the examined session and computes a normalized popularity score for each item to be ranked. Recently trending items are moved up in the search results according to this score. The best results were achieved when considering popularity trends of the last two weeks (Trending-14).
Hybrid Approaches	
HR1(Trending-14, FM)	A basic cascading hybrid which first ranks the items based on their recent popularity and then applies the feature matching method described above. The advantage of the method is that it also works in user cold-start situations and does not require computationally expensive long-term models.
HR2(Trending-14, FM, BPR)	A weighted hybrid which in addition considers the long-term user model using the BPR method. A grid search procedure was applied to determine optimal weights w.r.t. the hit rate. The final weights were 0.6 for Trending-14 and 0.2 for FM and BPR.
HR3(Trending-14, FM, BPR, MR)	An extension of HR2 which also includes reminders through the MR strategy. The parameters were again optimized through a grid search. The final values were 0.45 for MR, 0.4 for Trending-14, and only 0.075 for BPR and FM, which emphasizes the importance of reminders and popularity aspects.

Table 3: *Hit Rate@10* and *MRR@10* results for the dataset of 5,000 frequent users. The best values are printed in bold. Differences between the best performing method and the second best which are statistically significant according to the Wilcoxon signed-rank test ($\alpha = 0.05$) are marked with a star.

Min. nb. of result items	5		10		20		50	
Metric@10	HR	MRR	HR	MRR	HR	MRR	HR	MRR
HR3(Trending-14, FM, BPR, MR)	**0.685***	**0.394**	**0.675***	**0.382***	**0.636***	**0.363***	**0.584***	**0.324**
MR	0.671	0.389	0.652	0.371	0.619	0.352	0.572	**0.324**
HR2(Trending-14, FM, BPR)	0.634	0.315	0.624	0.304	0.565	0.282	0.511	0.241
HR1(Trending-14, FM)	0.626	0.306	0.604	0.289	0.564	0.267	0.504	0.235
BPR	0.605	0.297	0.579	0.284	0.535	0.262	0.477	0.226
FM	0.603	0.284	0.580	0.265	0.536	0.242	0.475	0.209
CB	0.560	0.273	0.535	0.258	0.487	0.238	0.416	0.198
C-KNN	0.555	0.268	0.524	0.246	0.480	0.227	0.403	0.185
Trending-14	0.537	0.229	0.506	0.208	0.455	0.185	0.375	0.148
PClick	0.476	0.194	0.438	0.172	0.385	0.150	0.321	0.127
Shop baseline (Bestseller)	0.467	0.191	0.433	0.168	0.380	0.147	0.314	0.125

which recommend both already known as well as new items.

Recommending items that have been popular within the last 14 days in an unpersonalized way proved to be a significantly better strategy than recommending the bestsellers of the entire data collection period ($\alpha = 0.05$)[3]. We found this strong difference somewhat surprising, given the short data collection period and this indicates that considering not only seasonable trends but also short-term spikes in sales (e.g., due to recent discounts) can be valuable for consumers.

Performance of Hybrids Combining a variety of different signals (long-term models, short-term models, reminders, and recent trends) in the weighted approach HR3 led to the overall best results and leaving out any of these components led to performance losses. In all except two cases the hybrid approach leads to statistically significantly better results according to the Wilcoxon signed-rank test as shown in detail in Table 3. *The main practical implication of our research therefore is that a multitude of signals should be considered in the search personalization process.*

The comparison of the strategies HR1 and HR2 shows again that including long-term preferences of the individual user *and* the user community is useful. Finally, even the simplest hybrid method HR1, which is computationally cheap and which works for cold-start users, was most of the time significantly better than all non-hybrid approaches (except for the reminders). This finding emphasizes the importance of considering both short-term trends as well as the visitors' individual and collective behavior for this problem setting.

Related Work

Our work draws on various insights from existing research in search personalization and recommender systems. While search result personalization for e-commerce sites, to our knowledge, has not been investigated in this form in the literature, the problem of web search personalization has been widely explored in the past. Typical strategies of such approaches include (a) the integration of personalization features in the ranking algorithm itself (Haveliwala, Kamvar, and Jeh 2003), (b) personalized query expansion (Chirita, Firan, and Nejdl 2007; Zhou, Lawless, and Wade 2012a), and (c) re-ranking in a post-processing step (Dou, Song, and Wen 2007). The algorithms evaluated in our work fall into this last category and we adopted a corresponding evaluation scheme used, e.g., in (Matthijs and Radlinski 2011).

Generally, the personalization and adaptation of search results is typically approached by deriving short- and long-term user profiles or by modeling the context of a search action from query, click-through, and other types of data (Bennett et al. 2015). Technically, this can be achieved with the help of weighted term vectors, ontologies, language models, and various machine learning techniques (Tan, Shen, and Zhai 2006; Sieg et al. 2007; Bennett, Svore, and Dumais 2010; Matthijs and Radlinski 2011).

In terms of additional data, existing research works investigated the following aspects: (a) the value of considering different contextual factors like the potentially underlying relationships between different search actions within a session or across multiple sessions; (Luxenburger, Elbassuoni, and Weikum 2008; Kotov et al. 2011), (b) the consideration of the user's current location (Bennett et al. 2011), and (c) the incorporation of signals from social media (Zhou, Lawless, and Wade 2012b; Carmel et al. 2009).

In addition, another family of approaches works by analyzing re-occurring search actions to adapt the ranking of the search results, e.g., (Dou, Song, and Wen 2007; Shokouhi et al. 2013).

In our experiments, we included two techniques from this field, namely a content-based approach that uses weighted term vectors as short-term and long-term models as well as one approach that learns from past successful searches. In theory, also other techniques can be applied, provided that the additionally required types of information like the user's social network activity are known.

In the field of recommender systems, session-based recommendation and the problem of balancing short-term and long-term interests was recently discussed, e.g., in (Jannach, Lerche, and Jugovac 2015). Similar to their work, we use BPR for learning a long-term model and different strate-

[3]We use the Wilcoxon signed-rank test ($\alpha = 0.05$) to assess the statistical significance of differences throughout the paper.

gies (C-KNN and FM) to incorporate short-term user interests. More elaborate techniques to capture the user's interests within a session, like proposed in (Hariri, Mobasher, and Burke 2012), could in principle also be applied to our problem in case sufficient information about the items is available. The value of including reminders within recommendations was discussed, e.g., in (Plate et al. 2006) and in (Lerche, Jannach, and Ludewig 2016). From the different strategies proposed in the latter work we included a simple yet effective method in our empirical investigations (MR).

Finally, there are different works that aim to extract useful insights from e-commerce search log data, e.g., (Duan et al. 2013) or (Liu et al. 2014), but their goal is often not centered around search personalization and focused, for example, on the unpersonalized diversification of the search results as done in (Yu et al. 2014). Personalized result rankings were discussed in the work of (Parikh and Sundaresan 2011), who however propose a manual approach, where users can interactively change the relevance of different sort criteria.

Conclusions

Our work showed that a variety of different factors should potentially be considered when personalizing search results in e-commerce. Since phenomena like short-term popularity trends and repeated item consumption are also common in other domains, including music or movies, we plan to investigate the use of hybrid approaches for search personalization in other fields as part of our future works.

References

Bennett, P. N.; Radlinski, F.; White, R. W.; and Yilmaz, E. 2011. Inferring and using location metadata to personalize web search. In *SIGIR '11*, 135–144.

Bennett, P. N.; Collins-Thompson, K.; Kelly, D.; White, R. W.; and Zhang, Y. 2015. Overview of the special issue on contextual search and recommendation. *ACM Trans. Inf. Syst.* 33(1):1e:1–1e:7.

Bennett, P. N.; Svore, K.; and Dumais, S. T. 2010. Classification-enhanced ranking. In *WWW '10*, 111–120.

Carmel, D.; Zwerdling, N.; Guy, I.; Ofek-Koifman, S.; Har'el, N.; Ronen, I.; Uziel, E.; Yogev, S.; and Chernov, S. 2009. Personalized social search based on the user's social network. In *CIKM '09*, 1227–1236.

Chirita, P. A.; Firan, C. S.; and Nejdl, W. 2007. Personalized query expansion for the web. In *SIGIR '07*, 7–14.

Dou, Z.; Song, R.; and Wen, J.-R. 2007. A large-scale evaluation and analysis of personalized search strategies. In *WWW '07*, 581–590.

Duan, H.; Zhai, C.; Cheng, J.; and Gattani, A. 2013. A probabilistic mixture model for mining and analyzing product search log. In *CIKM '13*, 2179–2188.

Hariri, N.; Mobasher, B.; and Burke, R. 2012. Context-Aware Music Recommendation Based on Latent Topic Sequential Patterns. In *RecSys '12*, 131–138.

Hariri, N.; Mobasher, B.; and Burke, R. 2015. Adapting to user preference changes in interactive recommendation. In *IJCAI '15*, 4268–4274.

Haveliwala, T.; Kamvar, S.; and Jeh, G. 2003. An Analytical Comparison of Approaches to Personalizing PageRank. Technical Report 2003-35, Stanford InfoLab.

Jannach, D., and Adomavicius, G. 2016. Recommendations with a purpose. In *RecSys '16*, 7–10.

Jannach, D.; Lerche, L.; and Jugovac, M. 2015. Adaptation and evaluation of recommendations for short-term shopping goals. In *RecSys '15*, 211–218.

Kotov, A.; Bennett, P. N.; White, R. W.; Dumais, S. T.; and Teevan, J. 2011. Modeling and analysis of cross-session search tasks. In *SIGIR '11*, 5–14.

Le, Q., and Mikolov, T. 2014. Distributed representations of sentences and documents. In *ICML '14*, 1188–1196.

Lerche, L.; Jannach, D.; and Ludewig, M. 2016. On the Value of Reminders within E-Commerce Recommendations. In *UMAP '16*, 27–25.

Liu, Z.; Singh, G.; Parikh, N.; and Sundaresan, N. 2014. A large scale query logs analysis for assessing personalization opportunities in e-commerce sites. In *WSCD Workshop at WSDM '14*.

Luxenburger, J.; Elbassuoni, S.; and Weikum, G. 2008. Task-aware search personalization. In *SIGIR '08*, 721–722.

Matthijs, N., and Radlinski, F. 2011. Personalizing web search using long term browsing history. In *WSDM '11*, 25–34.

Moe, W. W. 2003. Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *J. of Cons. Psych.* 13(1):29 – 39.

Parikh, N., and Sundaresan, N. 2011. A user-tunable approach to marketplace search. In *WWW '11*, 245–248.

Plate, C.; Basselin, N.; Kröner, A.; Schneider, M.; Baldes, S.; Dimitrova, V.; and Jameson, A. 2006. Recomindation: New functions for augmented memories. In *AH '06*, 141–150.

Rendle, S.; Freudenthaler, C.; Gantner, Z.; and Schmidt-Thieme, L. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *UAI '09*, 452–461.

Shokouhi, M.; White, R. W.; Bennett, P.; and Radlinski, F. 2013. Fighting search engine amnesia: Reranking repeated results. In *SIGIR '13*, 273–282.

Sieg, A.; Sieg, A.; Mobasher, B.; Mobasher, B.; Burke, R.; and Burke, R. 2007. Web search personalization with ontological user profiles. In *CIKM '07*, 525–534.

Tan, B.; Shen, X.; and Zhai, C. 2006. Mining long-term search history to improve search accuracy. In *KDD '06*, 718–723.

Yu, J.; Mohan, S.; Putthividhya, D. P.; and Wong, W.-K. 2014. Latent dirichlet allocation based diversified retrieval for e-commerce search. In *WSDM '14*, 463–472.

Zhou, D.; Lawless, S.; and Wade, V. 2012a. Improving search via personalized query expansion using social media. *Inf. Retr.* 15(3-4):218–242.

Zhou, D.; Lawless, S.; and Wade, V. 2012b. Web search personalization using social data. In *TPDL '12*, 298–310.