Determining Characteristics of Successful Recommendations from Log Data – A Case Study

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

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

ABSTRACT

Academic research in recommender systems largely focuses on the problem of predicting the relevance of (long-tail) items that the individual user presumably does not know yet. Many real-world systems however also recommend items that users have inspected in the past, items that are popular at the moment, and items currently on sale. In this work we investigate the value of including such items in recommendation lists based on an analysis of the web logs of a large online retailer. An examination of the features of successful item suggestions reveals that the chances of a recommendation leading to a purchase increase when the item is recently trending, on sale, or was recently viewed by the user. Offline simulation experiments furthermore show that considering those success factors that were identified from log data in the ranking algorithms can help to increase the prediction accuracy of recommender systems.

CCS Concepts

•**Information systems** → **Recommender systems;**
•**Applied computing** → **Electronic commerce;**

Keywords

Recommender Systems; Success Factors; Case Study

1. INTRODUCTION

Automated recommendations of the type "You may also be interested in" are common on today's e-commerce sites, and ample evidence exists that such personalized or item-related recommendations can measurably impact businesses [2, 4, 5, 6, 9, 10]. Recommender systems are correspondingly an active area of research, and significant advances have been made in recent years in terms of accurately predicting the relevance of individual items for each user.

The most common research approach in the field is to consider only items for recommendation that the user has not seen, consumed, or purchased before. In addition, recommending items which are unknown to the user, but very popular in general, is often considered to be of limited value for users, and sometimes such items are explicitly excluded from the evaluation as done, for example, in [13].

Real-world e-commerce sites like Amazon.com, however, include items in their recommendation lists that the users have just recently inspected and are therefore at least aware of or already familiar with. Furthermore, recommending popular items like top sellers or recently trending items is quite common in this domain. Finally, an additional aspect of real-world online shops barely investigated in the literature is that some recommended items can be on sale and correspondingly labeled with discount tags.

In a previous work [10], the authors analyzed the first of these aspects by examining the value of including *reminders*, i.e., items that the user has inspected in the past, into recommendation lists. For these analyses, the authors relied on specific log data of a large online retailer, which contains information about (a) which items were recommended to customers and (b) which items were actually purchased from the recommendation list. The analyses showed that many of the recommendations that were clicked on and later on purchased were already known to the users.

In this work we investigate – using the same log data – if there are other seldom-researched factors in e-commerce settings that contribute to the success of a recommendation. To identify such factors, we conduct a broader systematic analysis of the given user interaction data. Based on the results, we will particularly focus on the effectiveness of including *trending items* and *items on sale* into recommendations. Both types of recommendations are not uncommon in practice as mentioned above. We additionally propose first algorithms that take these aspects into account in the item ranking process and test if the inclusion of this additional information helps to increase the ranking accuracy of recommendation algorithms.

Obviously, recommending *only* discounted or currently trending items might in most cases be less effective in terms of the business value. This was shown, e.g., in [6], where the recommendation of top sellers was not very effective in terms of sales. However, we show that including *some* known or trending items can be valuable and also that many users actually inspect the recommended on-sale items. Overall, our work therefore contributes to a better understanding of what makes a recommendation successful in practice, at least in the domain of fashion products that we comprehensively analyze in this paper.

SAC 2017, April 03 - 07, 2017, Marrakech, Morocco

ISBN 978-1-4503-4486-9/17/04…$15.00

DOI: http://dx.doi.org/10.1145/3019612.3019757

2. DATA ANALYSIS

We base our analyses on a dataset of user navigation logs of the large European fashion retailer Zalando. In the following, we first introduce characteristics of the dataset itself and the subsets we use. Then, we investigate properties of successful recommendations with the help of basic statistical analyses. Finally, we present the results of a systematic feature analysis.

2.1 Dataset Properties

The raw dataset contains website interaction events of about 3.5 million users. As in [8], the logged actions include typical events such as item views, purchases, and add-to-cart actions. Each action relates to one of over 400,000 catalog items. For each item basic information is available to us, e.g., the brand, the category, or if the item was on sale at that time. The dataset is anonymized and distorted in a way that no inference about business figures of the shop is possible.

An additional unique feature of our dataset is that it contains (a) information about *which recommendations were made* to users when they inspected the details of an item[1] and (b) click events on these recommended items. This information allows us to investigate *which features make recommendations successful in this domain.*

The raw dataset is very sparse and contains many users who have only visited the shop once or never made a purchase. We therefore created two data samples as in [8] and [10]. One consists of 3,000 *heavy* users who visited the shop frequently. The other one consists of 3,000 *occasional* users that were randomly selected from a subset of users for which we recorded at least 10 purchase events. Over the period of one year, *heavy* users visited the shop about two times a week, resulting in about 114 sessions per year. *Occasional* users on average interacted with the shop every second week (about 28 sessions).

2.2 Properties of Successful Recommendations

General Conversion Rate. We use *conversion rates* as the usual *success* measure for our analysis. Using the dataset of *heavy* users as a basis, we observe that about every 100th recommendation list attracted a click. Note that the absolute number depends on several factors, e.g., on the visual layout of the shop, and cannot be compared across shops. Our analysis in addition shows that in as many as 14% of the cases when a recommended item was actually clicked-on (viewed), it was added to the shopping cart in the current or next two sessions. Furthermore, about 7% of the recommended-item clicks actually led to a purchase of the item.

This indicates that while recommendation lists are generally not a main navigation mechanism for users on the site, they lead to high "click-to-buy" conversions once the users have found a potentially relevant item within the recommendations.

Recommendation of Familiar and New Items. Continuing the research in [10], we examined to what extent recommending items that users already know can be a successful strategy. From all recommended items, about 10% were goods that the users had inspected on the site before. This percentage depends on the specific inner workings of the used algorithms on the site, which are unknown to us. It is therefore more interesting that nearly half of the *successful* recommendations (44%) were not new to the users, which highlights the potential value of including reminders in recommendation lists [10].

On the other hand, this also means that more than half of the recommendation-induced purchases were related to items that the consumers have not seen before. This can be seen as additional evidence of the capabilities of recommenders to help users discover relevant items. The main practical implication of the combined findings is however that including at least some already known (but not yet purchased) items can represent a promising strategy to increase the overall effectiveness of the recommender system.

The Importance of Short-Term Shopping Intents. The investigations in [8] showed – using simulation experiments on real data – that adapting to the user's immediate shopping intent is important to increase the prediction accuracy of recommendation algorithms[2]. To understand to what extent online consumers are focused on a goal when they visit the site, we analyzed their browsing behavior. On average, users inspected about 9 different items from 2.7 (of more than 330 available) categories, and considered 2.5 different colors and 3.6 different brands during one session. These numbers suggest that users indeed often have a specific shopping goal in each session. Thus, recommending items that are similar to the estimated shopping intent seems promising.

To quantify this aspect, we compared the recommend-to-purchase conversion rate when items were recommended that were similar to the most recent item to the conversion rate when this was not the case. The results are shown in Table 1 and we can see that recommending items that, e.g., have the same color as the currently viewed one, leads to a substantial increase of the conversion rate.

Table 1: Recommend-to-purchase conversions for similar-item and different-item recommendations

Item feature	Different value	Same value	Difference
Brand	0.950 %	4.227 %	345 %
Price level	1.403 %	3.624 %	158 %
Category	1.207 %	2.844 %	135 %
Color	1.521 %	2.701 %	77 %

The difference in the conversion rates lie between about 77% (color attribute) and more than 300% for the brand. The dominating importance of the brand is not particularly surprising in the fashion domain. However, consumers also seem to have a strong preference for certain price ranges (which were predefined categories in our dataset). Also, recommending items from the same sub-category and with the same color led to high increases of the conversion rates.

In sum, recommendations were more successful when the recommended items were very similar to the currently inspected ones. This is consistent with the observations regarding the focused shopping behavior of many users discussed above. Also, this underlines that in the examined domain the recommendation of substitute products is clearly

[1]Sets of three recommendations were recorded.

[2]A similar approach has shown to work well for news recommendations [11].

advantageous when compared to the recommendation of alternatives, which corroborates some observations from [3]. The accuracy results from the experiments in [8], which included a recommendation component that considers the features of recently viewed items, also confirm this hypothesis.

Considering Popular and Trending Items. Recommending popular items is generally a "safe" strategy, even though it might not lead to the highest business value as shown, e.g., in [6]. The fact that not all items are equally popular in a shop was also reflected in the consumer's adoption of the recommendations. When one of the three recommendations of a list was inspected by a user, the chances that it was the most popular one among the three – measured in terms of clicks and purchases by the entire user community – were at 43%, i.e., measurably higher than the theoretical 33% random chance.

Since seasonal trends are common in the fashion domain, we furthermore looked at the most recent popularity of the items. When looking for example at the popularity of the items on the day of the recommendations, we found that recommendations were particularly successful when they concerned these recently trending items. Using a normalized popularity score, the daily average popularity of all recommended items was at 0.024, whereas the average of those which were actually selected afterwards was at 0.088, i.e., three times higher.[3]

The Role of Discounts. The pricing of items usually has a direct impact on demand levels and sales. It therefore stands to reason that items in recommendation lists which are marked as being on sale and which are discounted are more attractive for users than other items. On the other hand, recommendations that contain too many or only discounted items, might raise the impression that the item suggestions are biased and the presented list is rather an advertisement than a recommendation.

In our dataset, we know for each recommended item if it was on discount at the time of the recommendation or not. We could therefore compare the recommendation-to-buy conversion rates for on-sale items and regular-prized items. The observed differences were huge. While recommended items with regular prices only lead to a conversion rate of 0.45%, the rate for discounted items was at least 18 times higher with a value of 8.12%. This clearly indicates that recommending discounted items can be a promising strategy in this domain.

2.3 Systematic Feature Analysis

The statistical analysis from the previous section showed that certain item features (popularity and discount status) can be indicators for the success of recommendations. In this section, we report the results of a more comprehensive and systematic analysis, which should help us understand what makes a recommendation successful. To assess the importance of different potentially relevant factors, we first used the available data to frame a classification problem. Then, we applied different feature weighting methods to numerically estimate the importance of the factors.

Each line of the resulting classification dataset corresponds to an item recommendation and is labeled as being successful or not. We then engineered 95 different features, including simple ones – like the popularity of an item during the last n days – as well as more complex ones that combine item characteristics with contextual or user-specific aspects[4]. An example for a more complex feature is the ratio of clicks by a user on items that have the same brand as the recommended one during the last n sessions. Since the dataset is very imbalanced and there are comparably few successful recommendations, we applied random downsampling to obtain an equal number of samples for each class.

Table 2: Gain Ratio

Feature	Weight
Discount level	0.439
Current popularity (day)	0.371
Discount flag	0.325
Viewed before	0.286
Current popularity (week)	0.242
Distance to first view (in days)	0.232
Distance to last view (in days)	0.217
Distance to first view (in sessions)	0.214
Distance to last view (in sessions)	0.210
Current popularity (month)	0.201

Table 3: Chi Squared

Feature	Rel. weight
Current popularity (day)	1.000
Current popularity (week)	0.785
Current popularity (month)	0.563
Discount flag	0.556
Discount level	0.556
Distance to last view (in sessions)	0.443
Distance to last view (in days)	0.441
Views count	0.435
Distance to first view (in days)	0.428
Distance to first view (in sessions)	0.428

Table 2 and 3 show the 10 most relevant features using the *Gain Ratio* and the *Chi Squared* method, respectively. For this measurement we sampled 2.000 *occasional* users and their successful recommendations from our dataset, leading to about 3.800 positive samples. The results of both methods are comparable and confirm the observations from the previous section. The most important features to predict the success of a recommendation are the recent popularity of an item and the fact that an item is on sale. Furthermore, when the user has recently viewed the item (multiple times), the chances are good that a recommendation will be successful. This result is also in line with the observation that reminding users of known items through recommendation lists can be useful.

A correlation analysis of the relevant features over all positive and negative samples showed that while most features are not related, a measurable correlation between the discount level and the current item popularity exists (0.47 for popularity/day, 0.34 for popularity/week). This suggests that discounts in this domain can have a positive effect on sales and confirms that recommending on-sale items can be valuable.

[3]The normalized score was computed by dividing the number of events (clicks and purchases) of an item by the maximum number of events recorded for an item in the dataset.

[4]A detailed list of all features is can be found at http://ls13-www.cs.tu-dortmund.de/homepage/sac2017-cosr

3. CONSIDERING TRENDS AND DISCOUNTS WHEN RECOMMENDING

Previous work [8, 10] has shown that considering short-term interests and recently viewed items can help to increase the accuracy of the recommendation algorithms. In the following, we will now aim to assess the value of incorporating popularity and discount information into a recommender.

3.1 Experimental Setup

We apply the offline measurement protocol proposed in [8, 10] to numerically assess the prediction accuracy of different algorithms. The protocol is based on time-stamped user interaction logs which are organized in shopping sessions. For each user, we split the time ordered list of sessions into training and test set where the latter one contains 20% of the user's purchases. The goal then is to predict every item an online visitor will purchase in each session of the test set in which a purchase was made. Since the consideration of short-term shopping intents is crucial for the e-commerce domain, we always reveal the interactions of the last p sessions including the current one that precede the tested purchase as proposed in [10]. Finally, we apply user-wise five-fold repeated random subsampling to account for random effects. The evaluation was conducted on the two datasets described above (*heavy* and *occasional*), which we sampled from the raw data. As accuracy measures we use the hit rate and the Mean Reciprocal Rank, each at list length ten, i.e., *HitRate@10* and *MRR@10*. The latter measure not only counts for how many of the purchases in the test sessions the algorithm made a correct guess, but also considers the position of the "hit" in the list. Since we hide and predict each purchase event in the test data individually, Precision is proportional to the hit rate (Recall) and not reported here.

3.2 Algorithms

Baseline Algorithms.

We use the following collaborative filtering (CF) recommendation algorithms as baselines to assess the value of incorporating additional features in the process:

- BPR: Bayesian Personalized Ranking [12] is a learning-to-rank method designed for one-class (implicit feedback) collaborative filtering problems that optimizes an AUC-like performance criterion. The method has been shown to outperform other techniques like Item-to-Item CF as well as different matrix factorization approaches for this and similar problem setups [7, 8].

- C-CoOcc: A baseline method that uses patterns of item co-occurrences in the users' recent navigation histories to implement a functionality of the form "Users who bought ..., also bought ..." (see [10]). The co-occurrence patterns are then applied to the user's recent interaction history to make recommendations.

- C-KNN. This method takes a user's most recent interaction history as an input and looks for past sessions in the training set related to the same items. The cosine similarity is used to determine the distance between such sessions and the recommendations are ranked in a weighted scoring process as described in [10].

In contrast to the BPR method, both C-CoOcc and C-KNN take the user's most recent shopping intent as a contextual factor into account.

Reranking Schemes.

We implemented several schemes that take the relevance-ranked list of a baseline algorithm as an input and apply different strategies to adapt the ranking based on additional information, e.g., about item properties or current discounts.

- RV: This simple method, also from [8], called *Recently Viewed* assumes that recent user interest in an item is a good indicator for an upcoming purchase. The method recommends items from the user's interaction history in reverse order.

- FM: *Feature Matching* was described as a successful strategy in [8] and moves those items up the list that have characteristics similar to those that the user has recently inspected, e.g., the same color or brand. The ranking is determined by the number of item characteristics that were also found in the user's current session. The ranking of the baseline strategy is kept for items that have the same number of matching features.

- DR: *Discount Reranking* ranks items higher that are currently on sale. In our data, a limited set of discrete discount levels is available for each purchase, and the method ranks items with the highest discounts first. Again, the order of the underlying baseline technique is kept for items with the same discount level.

- RPOP: The *Recent Popularity* method computes a normalized popularity value for each item during a recent period. The popularity is determined by counting the interactions with the items in the entire user community. Here, we will report the popularity of the day of the purchase, as this was the strongest success indicator according to our analysis above. When determining the popularity value, we only counted interactions that happened *before* the purchase to be predicted on the same day.

- Hybrids: Since the different features according to our analysis in the previous section are often not strongly related, we tested a number of different combinations of the reranking strategies. In particular, we combined the well-performing content-based FM ranking strategy with the methods proposed in this paper.

In the next section, we will particularly focus on the value of combining methods that have shown to work well in [8] (FM and RV) with techniques that additionally leverage information about discounts (DR) and trends in the recommendation process (RPOP).

3.3 Results

Table 4 and 5 show the results for the *heavy* and *occasional* datasets. In both cases, the parameter p of the protocol, corresponding to the number of revealed recent sessions, was set to 6. Different parameter values did not lead to significantly better results. The best values for each column are printed with a grey background. The overall best values are printed in bold face.

Table 4: *Hitrate@10* and *MRR@10* results for the Zalando *heavy* user subset.

Dataset	Zalando *heavy* users subset					
Baseline	C-KNN		C-CoOcc		BPR	
Metric@10	HR	MRR	HR	MRR	HR	MRR
WR(RPOP,DR,0.5)-FM	**0.382**	**0.220**	0.262	0.121	0.225	0.100
RPOP-FM	0.361	0.187	0.233	0.103	0.216	0.096
DR-FM	0.316	0.177	0.242	0.120	0.168	0.094
RV-FM	0.306	0.097	0.266	0.096	0.262	0.111
FM	0.281	0.093	0.145	0.052	0.119	0.046
No postprocessing	0.268	0.091	0.123	0.046	0.062	0.021

Table 5: *Hitrate@10* and *MRR@10* results for the Zalando *occasional* user subset.

Dataset	Zalando *occasional* users subset					
Baseline	C-KNN		C-CoOcc		BPR	
Metric@10	HR	MRR	HR	MRR	HR	MRR
WR(RPOP,DR,0.5)-FM	**0.376**	**0.213**	0.225	0.115	0.273	0.129
RPOP-FM	0.359	0.192	0.212	0.108	0.277	0.132
DR-FM	0.275	0.152	0.197	0.095	0.206	0.111
RV-FM	0.296	0.097	0.198	0.076	0.244	0.097
FM	0.257	0.091	0.129	0.049	0.192	0.076
No postprocessing	0.241	0.087	0.109	0.043	0.152	0.060

Baseline Results.

The last rows of the tables show the results when *no post-processing* is applied to the ranked lists returned by the baselines BPR, C-CoOcc, and C-KNN. For both datasets, C-KNN substantially outperforms the non-contextualized BPR method and the more simple contextualization method C-CoOcc on both measures.[5] This clearly shows that considering individual items of interest in the current browsing session *in isolation* (as C-CoOcc does) is not sufficient. C-KNN considers all recent interactions of a session and is thereby more effective in finding additional relevant items from past sessions. On the *occasional* dataset, even the BPR method, which only relies on long-term user models and no context information, is better than the C-CoOcc method. This indicates that the C-CoOcc method could not find a sufficient number of frequent co-occurrence patterns in this smaller dataset, which comprises the same number of users but has fewer interactions per user.

Adding Feature Similarity and Interaction Recency.

Moving up from the bottom-most row of the tables, we see that focusing the recommendations on items that have similar features as those inspected in the current session – as done by the FM strategy – helps to improve the hit rates and the MRR in every case. Also, considering the recency of item views as a ranking criterion (RV-FM strategy) leads to another accuracy increase as was shown in [8]. Note that RV-FM specifies a cascading hybrid strategy where the items are first pre-filtered based on their features and then re-ranked based on the time-stamp of the last interaction with it, in case the user has interacted with the item before.

Including mostly recently viewed items (reminders) in recommendation lists can help to achieve high prediction accuracy as shown in [10]. The recommendations might however be too obvious and prevent the recommender from pointing users to other potentially relevant items unknown to them. In the following, we will therefore focus on the value of recommending discounted and trending items.

Adding Discounts and Recent Popularity Information.

The effects of putting more emphasis on items that are currently on sale are given in the row DR-FM in Table 4 and Table 5. Across all configurations, ranking the items according to their level of discount is measurably better than ranking only based on their feature similarity (FM strategy).

A similar increase in prediction accuracy can be obtained by recommending recently trending items. The row labeled with RPOP-FM shows the results obtained when pushing items up the list according to their popularity on the day of the session in which the purchase was made. We varied the popularity time span to consider, e.g., the trending items of the week or month. The results show that recommending items that are trending over a longer period is also helpful; the strongest effects can however be obtained by looking at the most recent trends.

The success of recommending currently trending items can, according to our correlation analysis, at least in part be explained by the fact these items might be discounted. Nevertheless, since other factors might also increase the popularity of the offered items, we tested the hybrid strategy WR(RPOP,DR,0.5)-FM which combines popularity and discount information in a weighted approach.[6] This novel approach, in combination with the strongest baseline C-KNN, led to the *best overall results* for both datasets. The difference between the best-performing method and any other method was statistically significant at $p < 0.01$ according to a Student's t-test with Bonferroni correction.

[5]The C-KNN method was in similar form applied to the music playlist continuation problem in [1] and exhibited competitive performance also in this domain.

[6]The value 0.5 indicates that both aspects received equal weights, which was the best configuration in our tests.

Implications.

The analyses in Section 2 indicate that recommendations that are adopted by consumers are often related to trending and on-sale items. The results presented in this section show that these insights can be successfully operationalized in recommendation algorithms, e.g., through post-processing. At least for the fashion domain analyzed in this paper this in particular suggests that consumers do not consider items that are labeled as being discounted as potentially biased advertisements. In contrast, pointing consumers to on-sale items through recommendation lists seems to be an effective strategy, which, to our knowledge, has not been investigated in the recommender systems research literature before. An additional analysis showed that recommending on-sale items is not only effective for typical "bargain hunters", who have a tendency to mostly purchase discounted items. In contrast, discount recommendations were effective also for users for whom discounts did not play a role in the past.[7]

At the same time, recommending trending items has also proven to be effective in our experiments. At least to some extent, items became trending when they were put on sale, but not all trending items were actually on sale. Given the data that is available to us, we could not identify with certainty what caused the short-term popularity effects. It could be the result of a marketing campaign, it could be due to the recent addition of an item to the shop, or some other external event. We tested if some of the trending items appeared unproportionally often in recommendation lists, causing a "rich-get-richer" effect, which was however not the case. Also, the trending items were usually not the first items consumers looked at in a session, which would have indicated a deep link to the shop item from a marketing campaign. Nonetheless, a general strategy for a recommender in practice could be to constantly monitor the current popularity of the items and include (some of the) short-term bestsellers into the recommendation lists more often. Adding "a good dose of popularity" was also previously mentioned in [5] as a successful strategy for the domain of movie recommendations.

4. CONCLUSIONS

With our work we aim to contribute to a better understanding of what makes recommendations successful in practice. In this paper we focused on e-commerce recommendations and analyzed the recommendation logs of a large fashion retailer. The results show that besides the recommendation of recently viewed items, recommending on-sale items and currently trending items can lead to higher adoption rates of the recommendations in this domain. Our future work includes the analysis to what extent our findings generalize to other product domains and recommendation scenarios.

Generally, our work also shows that automatically identifying the factors that contribute to the success of recommendations by modelling it as a classification problem can help to generate better recommendations for a given application domain. The most important features in the fashion domain according to our analysis were recent item popularity, discounts, recent item views, and content-wise similarity to previously inspected items. The conducted experiments in [8, 10] and this paper showed that considering these aspects helped to improve the accuracy of our domain-specific algorithms. Generally, we therefore argue that a systematic analysis of successful recommendations can be one possible way to approach the largely open problem of predicting the online success of a recommender from offline experiments, as discussed, e.g., in [4, 5].

[7]More sales of discounted items usually leads to higher revenue but not necessarily more profit. These aspects have to be balanced by the service provider.

5. REFERENCES

[1] G. Bonnin and D. Jannach. Automated generation of music playlists: Survey and experiments. *ACM Computing Surveys*, 47(2):26:1–26:35, 2014.

[2] P. Cremonesi, F. Garzotto, and R. Turrin. Investigating the persuasion potential of recommender systems from a quality perspective: An empirical study. *ACM TIST*, 2(2):11, 2012.

[3] K. Diehl, E. van Herpen, and C. Lamberton. Organizing products with complements versus substitutes: Effects on store preferences as a function of effort and assortment perceptions. *Journal of Retailing*, 91(1):1–18, 2015.

[4] F. Garcin, B. Faltings, O. Donatsch, A. Alazzawi, C. Bruttin, and A. Huber. Offline and online evaluation of news recommender systems at swissinfo.ch. In *RecSys '14*, pages 169–176, 2014.

[5] C. A. Gomez-Uribe and N. Hunt. The Netflix recommender system: Algorithms, business value, and innovation. *ACM TMIS*, 6(4):13:1–13:19, 2015.

[6] D. Jannach and K. Hegelich. A case study on the effectiveness of recommendations in the mobile internet. In *RecSys '09*, pages 205–208, 2009.

[7] D. Jannach, L. Lerche, F. Gedikli, and G. Bonnin. What recommenders recommend - an analysis of accuracy, popularity, and sales diversity effects. In *Proc. UMAP '13*, pages 25–37, 2013.

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

[9] E. Kirshenbaum, G. Forman, and M. Dugan. A live comparison of methods for personalized article recommendation at Forbes.com. In *ECML/PKDD '12*, pages 51–66, 2012.

[10] L. Lerche, D. Jannach, and M. Ludewig. On the Value of Reminders within E-Commerce Recommendations. In *UMAP '16*, 2016.

[11] J. Liu, P. Dolan, and E. R. Pedersen. Personalized news recommendation based on click behavior. In *IUI '10*, pages 31–40, 2010.

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

[13] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, N. Oliver, and A. Hanjalic. CLiMF: Learning to Maximize Reciprocal Rank with Collaborative Less-is-more Filtering. In *RecSys '12*, pages 139–146, 2012.