# 8

# Case study: Personalized game recommendations on the mobile Internet

Although the interest in recommender systems technology has been increasing in recent years in both industry and research, and although recommender applications can nowadays be found on many web sites of online retailers, almost no studies about the actual *business value* of such systems have been published that are based on real-world transaction data.

As described in Chapter 7, the performance of a recommender system is measured mainly based on its accuracy with respect to predicting whether a user will like a certain item. The implicit assumption is that the online user – after establishing trust in the system's recommendations or because of curiosity – will more often buy these recommended items from the shop.

However, a shop owner's key performance indicators related to a personalized web application such as a recommender system are different ones. Establishing a trustful customer relationship, providing extra service to customers by proposing interesting items, maintaining good recommendation accuracy, and so on are only a means to an end. Although these aspects are undoubtedly important for the long-term success of a business, for an online retailer, the important performance indicators are related to (a) the increase of the conversion rate – that is, how web site visitors can be turned into buyers, and (b) questions of how to influence the visitors in a way that they buy more or more profitable items.

Unfortunately, only few real-world studies in that context are available because large online retailers do not publish their evaluations of the business value of recommender systems. Only a few exceptions exist. Dias et al. (2008), for instance, present the results of a twenty-one-month evaluation of an probabilistic item-based recommender system running on a large Swiss e-grocer web portal. Their measures include "shopper penetration", "direct extra revenue", and "indirect extra revenue". Their analysis showed several interesting points. First, a relatively small (when compared with overall sales) extra revenue can be



https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

generated directly by the recommender. The fact that direct revenues measurably increased when the probabilistic model went through a periodic update suggests that good recommendation accuracy is still important, despite some legitimate criticism of simple accuracy measures (McNee et al. 2006). The more important business value, however, comes from *indirect* revenues caused by the recommender systems. Indirect revenues include the money spent on repeated purchases of items initially recommended by the system and on items sold from categories to which the customer was newly introduced to through a recommended item. This, in turn, also supports the theory that diversity in recommendation lists is a valuable property, as "unexpected" items in these lists may help to direct users to other, possibly interesting, categories.

An earlier evaluation based on real-world data was presented by Shani et al. (2002), in which the authors performed different experiments on an online bookstore. During their experiment, visitors to the web shop received buying proposals either from a "predictive" or a new Markov decision process recommender. Thus, they were able to compare the respective profits that were generated by different techniques during the observation period. In addition, at least for a period of seven days, the recommendation functionality was fully removed from the web shop. Although this sample is statistically too small to be significant, the comparison of sales numbers of two consecutive weeks (one with and one without the recommender) showed a 17 percent drop in the recommender-free week.

Another initial study on how recommender systems influence the buying behavior of web shop visitors is presented by Zanker et al. (2006). In this work, it was shown that the recommendations of a virtual advisor for premium cigars can stimulate visitors to buy cigars other than the well-known Cohibas and thus increase sales diversity, which is interesting from up-selling and cross-selling perspectives and could also create indirect revenue as described by Dias et al. (2008); see also Fleder and Hosanagar (2007), for a discussion of the role of sales diversity in recommender systems.

In Zanker et al. (2008) and Jannach et al. (2009), a different study using the same recommendation technology was made in the tourism industry, in which it could be observed that the number of accommodation availability enquiries is measurably higher when web site visitors are guided by the virtual advisor. Another evaluation of how different information types and recommendation sources influence consumers can be found in Senecal and Nantel (2004).

Similar to these works, the case study presented in this chapter[1] focuses on evaluating the business value of recommender systems in a commercial context.

[1] The work was also presented at the 7th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems at IJCAI'09 (Hegelich and Jannach 2009); a summary of the results of the study can also be found in Jannach and Hegelich (2009).

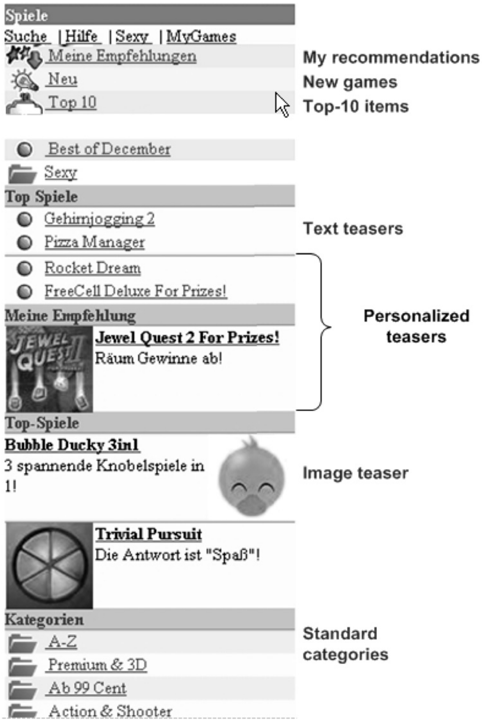


Figure 8.1. Catalog navigation and categories.

In addition, it aims to answer the question whether certain algorithms perform better than others in a certain environment and application domain in the line of the work of, for example, Breese et al. (1998) or Zanker et al. (2007).

## 8.1 Application and personalization overview

The study presented in this chapter was conducted in the context of a mobile Internet portal of a large telecommunications provider in Germany. Customers access this portal through their mobile devices and are offered a wide range of applications and games, which they can purchase directly and download to their cell phones.

Figure 8.1 shows the entry screen of the games area of the portal. Customers explore the item catalog in the following ways:

- Through manually edited *or* nonpersonalized lists such as “New items” or “Top 10 items” (top area of screen).
- Through direct text or image links (teasers) to certain items that are shown on the middle area of the start screen.

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

- Through predefined standard categories (lower area) such as "A–Z", "From 99 Cents", or "Action & Shooter".
- In addition, after a purchase, when the payment confirmation is displayed, customers are presented with a list of other, possibly interesting items (postsales recommendation).

Accordingly, the portal was extended with personalized content as follows:

(a) A new top-level link, "My Recommendations", was introduced, which leads to a personalized recommendation list ("Meine Empfehlungen" in German).
(b) The games presented in the lower two of the four text teasers and the first image teaser on the start page were personalized. Because of existing contracts, the first two text links and the two lower image links were manually predefined. The manually edited links remained the same during the whole experiment, which made it possible to analyze the effects of personalizing the other links independently.
(c) The lists in the standard categories such as "99 Cents" were personalized except for categories such as "A–Z", which have a "natural" ordering.
(d) The games presented on the postsales page were also personalized.

During the experiments, different algorithms were used to calculate the personalized recommendations. To measure the effect of personalization, members of the control group were shown nonpersonalized or manually edited lists that were based on the release date of the game.

Customers can immediately purchase and download games through the portal by choosing items from the presented lists. The relation between their navigation and buying behavior can therefore be easily determined, as all portal visitors are always logged in. Several thousand games (across all categories) are downloaded each day through the platform. The prices for the games range from free evaluation versions (demos) to "99 Cent Games" to a few euros for premium games; the amounts are directly charged to the customer's monthly invoice. In contrast to the study by Dias et al. (2008), in which users purchased the same goods repeatedly, customers in this domain purchase the same item only once – in other words, the domain is similar to popular recommender systems application areas such as books and movies.

From the perspective of the application domain, the presented game portal stands in the line of previous works in the area of recommender systems for mobile users. Recent works in the field of mobile recommenders include, for instance, Miller et al. (2003), Cho et al. (2004), van der Heijden et al. (2005), Ricci and Nguyen (2007), Li et al. (2008), and Nguyen and Ricci (2008).

Content personalization approaches for the mobile Internet are presented also by Pazzani (2002), Billsus and Pazzani (2007), and Smyth et al. (2007). In Smyth and Cotter (2002), finally, the effects of personalizing the navigational structure on a commercial Wireless Access Protocol (WAP) portal are reported.

Overall, it can be expected that this area will attract even more attention in the future because of the rapid developments in the hardware sector and the increasing availability of cheap and fast mobile Internet connections. In contrast to some other approaches, the recommender system on this platform does not exploit additionally available information such as the current geographical position or demographic and other customer information known to the service provider. Standard limitations of mobile Internet applications, such as relatively small network capacity and limited display sizes, apply, however.

## 8.2 Algorithms and ratings

During the four-week evaluation period, customers were assigned to one of seven different groups when they entered the games section of the portal. For each group, the item lists were generated in a different way. For the first four groups, the following recommendation algorithms were used:

- Item-based collaborative filtering (CF) (Sarwar et al. 2001) as also used by Amazon.com (Linden et al. 2003).
- The recent and comparably simple Slope One algorithm (Lemire and Maclachlan 2005).
- A content-based method using a TF-IDF representation of the item descriptions and the cosine similarity measure.
- A "switching" hybrid algorithm (Burke 2002b) that uses the content-based method when fewer than eight item ratings are available, and item-based collaborative filtering otherwise.

Two groups received nonpersonalized item lists, one based on the average item rating ("Top Rating") and one based on the sales numbers (top sellers). For the final group, the control group, the recommendation lists were manually edited as they were before the personalization features were introduced. Within most categories, the ordering was based on the release date of the game or chosen based on existing contracts. The top-level link "My Recommendations" was not available for the control group. During the entire evaluation period, customers remained in their originally assigned groups.

From all customers who visited the games portal during the evaluation, a representative sample of more than 155,000 was included in the experiment, so

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

each group consisted of around 22,300 customers. Only customers for which all algorithms were able to produce a recommendation were chosen – that is, users for whom a minimum number of ratings already existed. The catalog of recommendable items consisted of about 1,000 games.

A five-point rating scale from $-2$ to $+2$ was used in the experiments. Because the number of explicit item ratings was very low and only about 2 percent of the customers issued at least one rating, implicit ratings were also taken into account: both clicks on item details as well as actual purchases were interpreted as implicit ratings. When no explicit rating was given, a view on item details was interpreted as a rating of 0 (medium); several clicks on the same item were not counted. An actual purchase was interpreted as a rating of 1 (good) for the item. Explicit ratings overrode these implicit ratings.

To achieve the best possible recommendation accuracy, the item similarities and the average differences for the collaborative filtering and the Slope One techniques were computed using the full customer base and not only the 155,000-customer subsample.

## 8.3 Evaluation

The following hypotheses are in the center of the evaluation:

- H1: Personalized recommendations attract more customers to detailed product information pages (item view conversion rate).
- H2: Personalized recommendations help turn more visitors into buyers (sales conversion rate).
- H3: Personalized recommendations stimulate individual customers to view more items.
- H4: Personalized recommendations stimulate individual customers to buy more items.

The detailed evaluation will show that depending on the navigational situation of the portal visitor, different phenomena with respect to the effectiveness of recommendation algorithms can be observed. Before considering the overall effect of the use of recommendation technology on the portal, the individual results obtained for these different situations will be discussed.

### 8.3.1 Measurement 1: "My Recommendations"

The following results are related to the personalized recommendation list that is presented when the customer clicks on the "My Recommendations" link, as

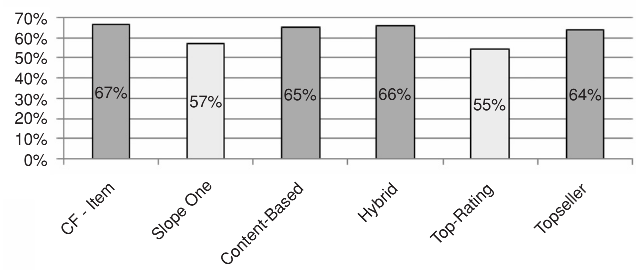


Figure 8.2. Conversion rate: item views to "My Recommendations" visits.

shown in the top area of Figure 8.1. Throughout the evaluation, different levels of gray will be used to highlight data rows in the charts that are significantly different ($p < 0.01$) from each other.

The conversion rate measurements (hypotheses H1 and H2) are given in Figure 8.2, which depicts the item view conversion rate for visitors to the "My Recommendations" list, and Figure 8.3, which shows how many of the users who visited the "My Recommendations" section actually purchased an item[2].

In Figure 8.2 it can be seen that the different algorithms fall into two groups: one in which about two-thirds of the customers actually click on at least one of the presented items and one in which only 55 percent are interested in the recommended items. Considering the actual numbers, the differences between the two groups are significant ($p < 0.01$).

From the personalized methods, only the Slope One algorithm did not attract significantly more visitors than the nonpersonalized list of top-rated items. Interestingly, the nonpersonalized top-seller list also has a good item view conversion rate – in other words, placing generally liked, top-selling items in a recommendation list seems to work quite well in the domain.

When the sales conversion rate is considered, it can be observed from Figure 8.3 that only the CF method helps to turn more visitors into buyers (Hypothesis H2).

The evidence for our hypotheses H3 (more item views per customer) and H4 (more purchases per customer) in the context of the "My Recommendations" section can be seen in Figures 8.4 and 8.5. Figure 8.4 shows that all recommendation algorithms (except for Slope One) stimulate users to click on more items. Compared with the findings with respect to the conversion rates,

[2] In Figures 8.2 to 8.5, the control group is not depicted, because the "My Recommendations" section, which was newly introduced for measuring the impact of personalization, was not available for them.

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

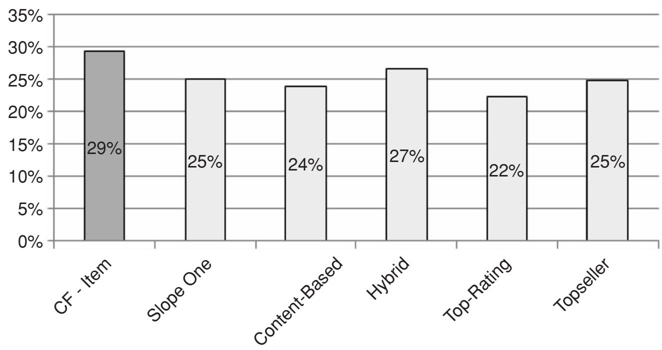


Figure 8.3. Conversion rate: buyers to "My Recommendations" visits.

this can be interpreted that personalized lists seem to contain more items that are interesting to a customer.

When it comes to actual purchases (game downloads), Figure 8.5 shows that most personalized methods, and even the simple Slope One algorithm, outperform the nonpersonalized approaches.

For some of the games provided on the mobile portal, free evaluation versions (demos) are available. If not mentioned otherwise, all numbers given with respect to conversion rates and sales figures are related to all item downloads – free demos plus actual game purchases. Figure 8.6 repeats the numbers of Figure 8.5, but also shows the fraction of demo downloads and purchased games. Because of the nature of the algorithms and the particularities of the application (see more details in Measurement 4), the recommendation lists produced by the TopRating and Slope One methods contain a relatively high portion of demo games. Given the high number of actual downloads, these demo recommendations seem to be well accepted, but unfortunately, these two techniques perform

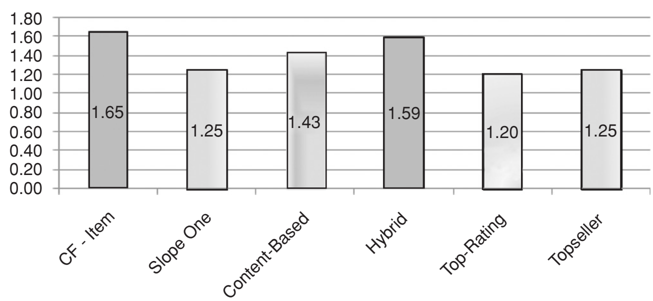


Figure 8.4. Item views per "My Recommendations" visits.

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

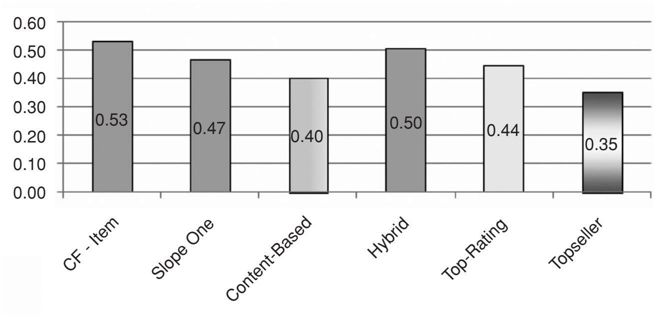


Figure 8.5. Item purchases per “My Recommendations” visits.

particularly poorly when the games are not free. The item-based, content-based, and hybrid techniques, on the other hand, not only help to sell as many items as a simple top-seller promotion but also make users curious about demo games. The TopRating method raises interest only in demo versions. The list of top-selling items is generally dominated by non-free, mainstream games, which explains the fact that nearly no demo games are chosen by the users.

### 8.3.2 Measurement 2: Post-sales recommendations

The next navigational situation in which product recommendations are made is when a customer has purchased an item and the payment confirmation has just finalized the transaction. About 90,000 customers who actually bought at least one item during the evaluation period were involved in the experiment. Overall, the evaluation sample contains more than 230,000 views of the post-sales

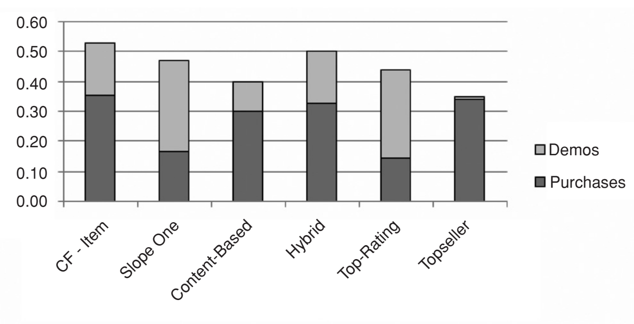


Figure 8.6. Game purchases and demo downloads in “My Recommendations” visits.

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

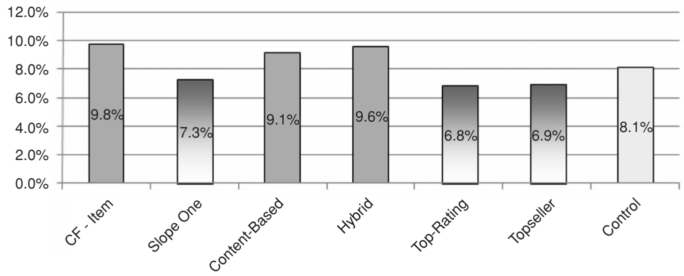


Figure 8.7. Conversion rate: item views to post-sales list views.

five-item recommendation lists, meaning that, on average, customers bought more than one item.

The experimental setup is nearly identical with that for Measurement 1; customers received their recommendations based on different recommendation algorithms. The recommendation list of the control group was manually edited and ordered by game release date. Items that the current customer had already purchased before were removed from these lists.

The same hypotheses were tested in this experiment – that is, to what extent recommender systems stimulate customers to view and buy more items. The results are shown in Figures 8.7 through 8.10.

With respect to the conversion rates, the following observations can be made. First, the manually edited list of recent items (viewed by the control group) worked quite well and raised more customer interest than the nonpersonalized techniques and even the Slope One algorithm (Figure 8.7). When it comes to actual purchases (Figure 8.8), however, the manually edited list did not help turn more visitors into buyers. Interestingly, the relative improvement caused by personalized recommendations with respect to this conversion rate is higher on the post-sales recommendation page than in the “My Recommendations”

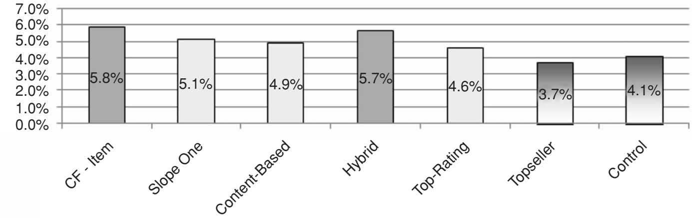


Figure 8.8. Conversion rate: Buyers to post-sales list views.

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

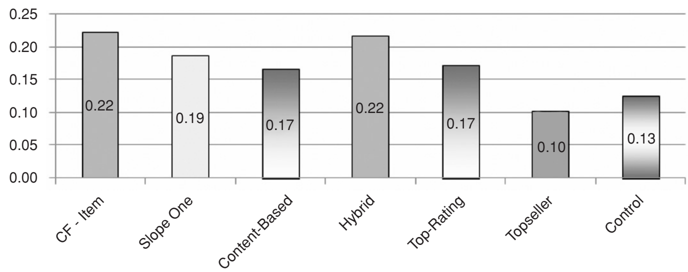


Figure 8.9. Item visits per post-sales list views.

sections. Again, the CF algorithm worked best; in absolute numbers, the differences between the various techniques are significant ($p < 0.01$).

With respect to the number of item visits and purchases per customer (Figures 8.9 and 8.10), it can again be observed that the different recommendation techniques not only stimulated visitors to view more items but actually also helped to increase sales. It can also be seen that displaying a list of top-selling items after a purchase leads to a particularly poor effect with respect to the overall number of downloads.

Another observation is that the items that are recommended by the Slope One technique and the TopRating method are also downloaded very often (see Figure 8.10), presumably because the recommendation lists again contain many free demos. Figure 8.11 therefore shows the ratio of demo downloads to game purchases, which is quite similar to the one from the "My Recommendations" section – that is, recommending top-selling or newly released items does not stimulate additional interest in free evaluation versions (demo games). The trend toward interest in demo versions seems to be a bit more amplified than

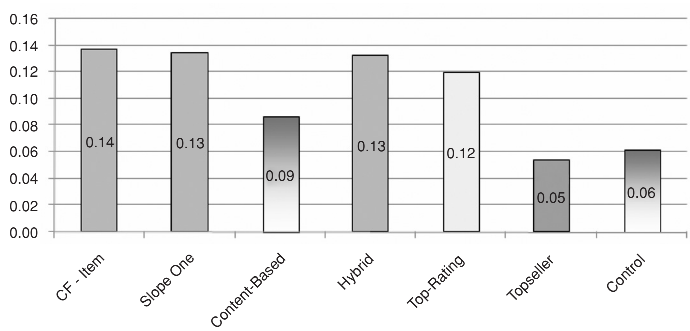


Figure 8.10. Item purchases to post-sales list visits.

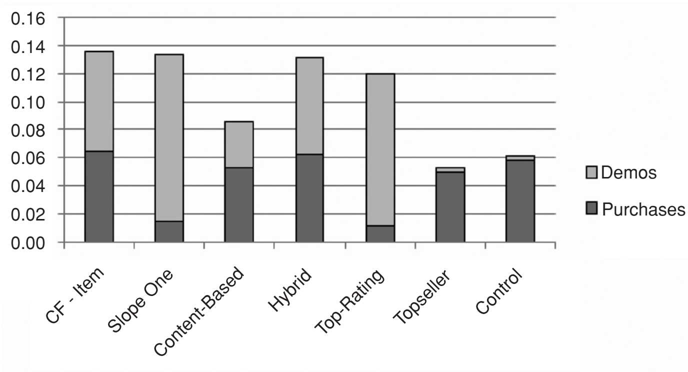


Figure 8.11. Game purchases and demo downloads on post-sales page.

in the "My Recommendations" section, which indicates that after a purchase transaction, customers first have a look at another, but free, game.

Finally, in this navigational context, the content-based method could raise some initial customer interest (Figure 8.9), perhaps because games are recommended that are quite similar to previously downloaded ones. However, although customers viewed some of the items, they had no strong tendency to purchase them, probably because the games were – according to the general tendency of content-based methods – too similar to games they already knew. The list of top-selling items again contained mostly non-free games, which explains the small fraction of demo games here; the same holds for the control group.

### 8.3.3 Measurement 3: Start page recommendations

This measurement analyzes the effect of the personalized recommendations on the start page, as shown in Figure 8.1. Remember that some elements in these lists are edited manually but were static during the experiment. Thus, item visits or purchases from these links (that could have been other banner advertisements as well) were not included in the evaluation.

During the experiment, the personalized elements of the list – the last two text teasers and the first image teaser – were determined based on the top-three list of the individual recommendation algorithms or based on the nonpersonalized lists of top-selling and top-rated items. Customers assigned to the control group received manually determined recommendations that were ranked by release date.

For this experiment, only the conversion rate figures for the different teaser elements on the start page will be shown.

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

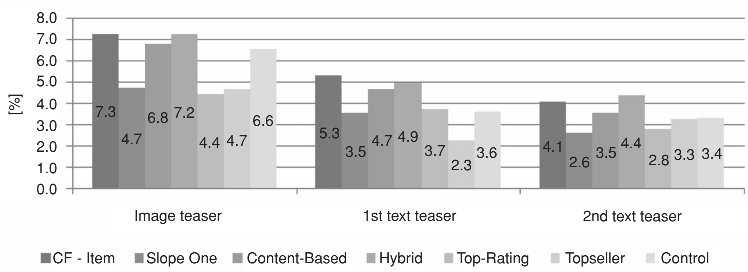


Figure 8.12. Conversion rate: item views to start page visits.

Figure 8.12 shows the percentage of portal visitors who followed one of the personalized product links on the start page. On average, the image teaser was clicked on by around 6 percent of the users. Although the image represents only the third-ranked item of the recommendation algorithms and is also positioned after the text links, its conversion rate is significantly higher than that for the text links. As this also holds for the nonpersonalized methods, the attractiveness of the third link can be attributed to its visual representation. Interestingly, however, the image teaser leads to a good conversion rate with respect to actual sales (Figure 8.13). With respect to these conversion rates, both the CF method and the content-based method lead to a significant increase of item detail clicks and purchases. It can also be observed that the conversion rates of the first text teaser can even be better than the image teaser when the text links are personalized. Thus, personalization can partially even outweigh the disadvantages of the unflashy representation.

Another particularity of this measurement on the start page is that the manually selected items used for the control group lead to comparably good conversion rates, especially with respect to item visits. A possible explanation could be that customers have no special expectations with respect to the offers on the

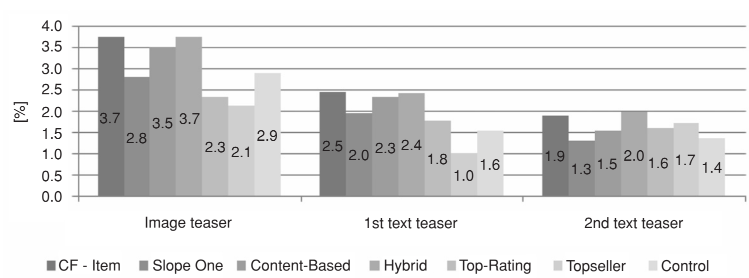


Figure 8.13. Conversion rate: purchases from start page visits.

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

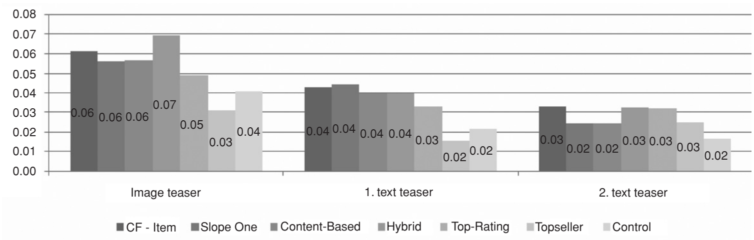


Figure 8.14. Purchases per start page visits.

start page. The fact that the manually selected items are newly released ones might further contribute to the good acceptance.

Although recommending items based on their average customer rating (as done by the Slope One and the TopRating techniques) worked well in the first two experiments, this approach does not work particularly well on the start page – customers seem to prefer either new items or items that are somehow related to their previous buying history.

Finally, when it comes to the number of purchases induced by the recommendation lists, the personalized techniques clearly outperformed the manually defined lists, at least for the first two teaser elements (see Figure 8.14). The item click and sales numbers of the other four, and statically defined image and text teasers with the personalized ones, were also compared. It could be seen that although the personalized items are partially placed lower on the screen and are thus harder to select, they received significantly more clicks and led to more sales than the nonpersonalized links.

### 8.3.4 Measurement 4: Overall effect on demo downloads

In Measurements 1 and 2, it could be seen that Slope One and the nonpersonalized technique based on item ratings led to significantly more views and downloads of demo games. In this measurement, the goal was to analyze whether this trend also exists when the entire platform is considered, including, for instance, all other personalized and nonpersonalized navigation possibilities.

No explicit category in the navigation tree for "free demos" exists. Games for which free evaluation versions exist can, however, appear in all other personalized and nonpersonalized item listings in the portal. In addition, customers are pointed to demos in two additional ways: (a) through direct-access links that are sent to them in sales promotions and (b) through pointers to other demo games that are displayed after a demo has been downloaded.

The distribution of views and downloads of demo games during the four-week evaluation period for the different recommendation groups is shown in

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

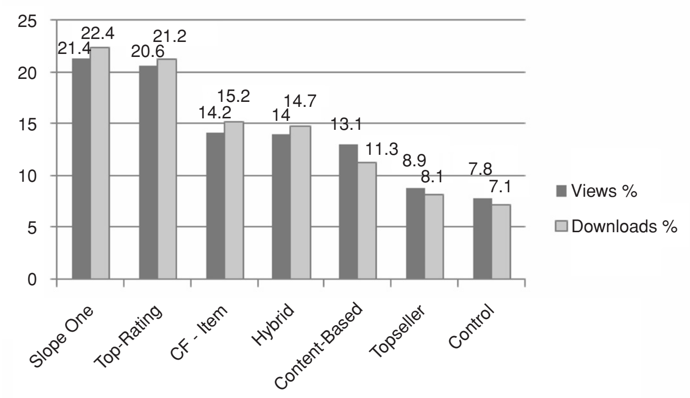


Figure 8.15. Distribution of demo game item views and downloads.

Figure 8.15. Overall, about 38,000 downloads were observed for the selected subsets of customers. When considering the actual downloads, it can be seen that the ranking of the algorithms remains the same; the differences are even amplified.

As already briefly mentioned in previous sections, this result can be explained by different facts that are related to the particular application setting and the nature of Slope One and the top-rating algorithm, which both tend to rank demo games highly in the different categories described previously, for the following reasons. First, as demo games can be downloaded at no cost and user ratings are possible on the platform only after a download, more explicit ratings are available for these games. Next, explicit ratings tend to be above average also in this domain. A similar phenomenon can also be observed in other datasets such as the MovieLens rating database. Finally, as customers receive a nonpersonalized pointer to another demo after downloading a free game, a reinforcement of the effect occurs.

An in-depth analysis of whether the downloads that were stimulated by the different algorithms led to significantly different demo-download/purchase conversion rates was not done in this case study. What could, however, be observed in a first analysis is that the demo/purchase conversion rate was significantly higher when the demo was promoted by a recommendation list (as opposed to a banner advertisement).

### 8.3.5 Measurement 5: Overall effects

In the final measurement reported in this study, the overall effect of the personalized recommendations (as an add-on to the other navigational options)

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

was evaluated. Again, the interesting figures are related to item view and sales conversion rates (H1 and H2) as well as to the question of whether more items were viewed and purchased by individual customers (H3 and H4).

With respect to the conversion rates (hypotheses H1 and H2), no significant differences between the personalized and nonpersonalized variants could be observed on the platform as a whole. On average, about 80 percent of all observed customers viewed at least one item, and around 57 percent bought at least one game, independent of the recommendation algorithm group they were assigned to. These figures are nearly identical for all seven test groups. For the item view conversion rate, for instance, the numbers only range from 79.6 percent to 80.3 percent. Thus, although slight improvements could be observed in individual (personalized) situations, as described earlier, the influence on the overall conversion rate is too small, and thus the percentage of portal visitors who view or purchase items could not be significantly increased by the additional use of personalized recommendation lists.

There could be different reasons for this non-effect. First, besides the described personalized lists, there are various other ways in which customers can access the item catalogs. Many customers, for instance, use the built-in search functionality of the portal; the ranking of search results is not personalized. The list of new items (see Figure 8.1) is also one of the most popular ways of browsing the catalog and is used by significantly more people than, for instance, the new "My Recommendations" section. An analysis showed that personalizing this particular list does not improve the conversion rates, as customers always prefer to see the latest releases at the top of such a list. Second, in this evaluation, only customers have been considered for whom a minimum number of ratings already existed – that is, users who are in generally interested in games. An evaluation of whether more *new* users can be tempted to purchase items was not in the focus of the evaluation.

With respect to hypotheses H3 and H4 (increased number of item views and sales per customer), the following observations can be made. Regarding the average number of item views per customer (H3), it could be seen that all personalized algorithms outperform the nonpersonalized top-seller list and the control group. Similar to the effect of Measurement 4, Slope One and the simple ranking based on average customer rating raised the most attention. Thus, H3 could be only partially validated at the global scale as the nonpersonalized top-rating technique was also successful.

The observations made with respect to the number of purchased/downloaded items per customer (H4) are shown in Figure 8.16.

The figure shows that the additional attention raised by Slope One and the TopRating algorithm also leads to a measurably increased number of items

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

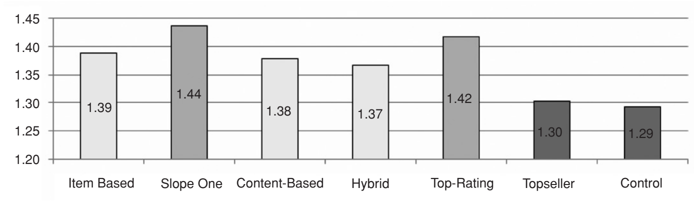


Figure 8.16. Average number of purchases, including free downloads, per customer on entire platform.

purchased and downloaded per customer. Figure 8.17 shows the number of downloaded items (including the demos) for the different algorithms. Finally, if we look at the actual sales numbers for non-free games only (Figure 8.18), it can be seen that although the Top-Rating list raised attention for free demos, it did not lead to increased sales for non-free items. Overall, all personalized techniques were more successful than the nonpersonalized one. On the global scale, however, the difference was – a bit surprisingly – significant only for the content-based method, which indicates that customers tend to spend money on items that are similar to those they liked in the past. In fact, a closer look on the performance of the algorithms in specific subcategories shows that the content-based method often slightly outperforms other methods with respect to non-free games. Although the differences were not significant in the individual situations, these slightly higher sales numbers add up to a significant difference on the global scale. Examples of categories in which the content-based method worked slightly better with respect to non-free games are the "new games", "half-price", or "erotic games" sections of the download portal.

Overall, the increase in actual sales that are directly stimulated by the recommender system is between 3.2 percent when compared to the Top-Rating

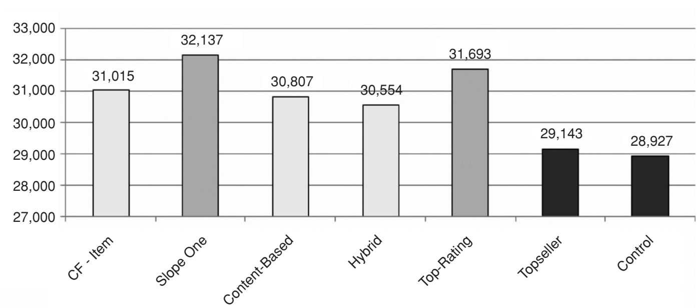


Figure 8.17. Total number of purchases and downloads.

https://doi.org/10.1017/CBO9780511763113.010 Published online by Cambridge University Press

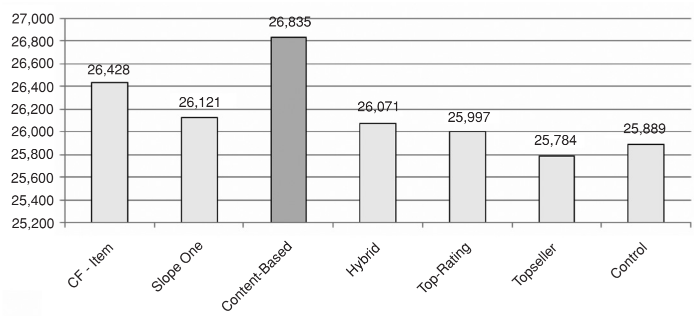


Figure 8.18. Total number of purchases (without demos).

technique, and around 3.6 percent when no personalized recommendation is available.

In general, these last observations suggest that in situations in which the user has no strong expectations on a certain genre (such as the "My Recommendations" section), collaborative methods – which also recommend items in categories that the user has not seen before – work particularly well. In many other situations, however, users tend to prefer recommendations of game subcategories that they already know. One exception is the post-sales situation, in which users are, not surprisingly, not interested in purchasing games that are very similar to the ones they have just bought.

## 8.4 Summary and conclusions

In this study, the effects of personalized item recommendation in various navigational contexts on a mobile Internet game portal were analyzed. Different standard recommendation techniques were implemented on the portal and deployed in parallel in a real-world setting for a period of four weeks. In addition, nonpersonalized techniques based on top-selling or top-rated items were used for comparison purposes.

The findings can be summarized as follows:

**Ratings in the mobile Internet.** The number of explicit item ratings was very low on the considered mobile Internet portal, and only about 2 percent of the users issued explicit ratings. Although no studies are available that compare the willingness of customers to rate items in different settings, it can be suspected that the relatively high effort for submitting an item vote using a mobile device compared with a web browser discourages users from participating in this community process.

**Recommending in a navigational context.** The effects of personalized recommendations have been measured in different navigational situations, such as the start page of the portal or the post-sales situation. In addition, a differentiation was made between the interest that was raised by the recommendations and the actual effect on the buying behavior of the customers.

With respect to the navigational context, customers seem to react slightly differently to recommendations, probably because of different expectations. In the dedicated "My Recommendations" section of the portal, classical CF and the hybrid technique are particularly good at raising customer interest, as customers view many of the recommended items. Although customers are also easily stimulated to download free games by the comparably simple Slope One and TopRating methods, these techniques do not lead to a significant increase in non-free games. A similar effect can be observed in the post-sales situation; the trend toward free demo downloads is even amplified in this situation. Thus, the item-based, content-based, and hybrid techniques that lead to a good number of purchases but also raise additional interest in demos seem to be a good choice here.

On the portal entry page, the recommendation of top-rated (or top-selling) items has a particularly poor effect, and the personalized methods lead to significantly better results. A listing of newly released items on the start page, however, also worked quite well.

In certain navigational situations, it could be observed that personalization worsens the conversion rates and sales numbers. In the section on new items, which contains games of the last three weeks, the strict chronological order, with the newest items on top, works best. Most probably, the visitors to the "New" category enter this section regularly and check only the first few lines for new arrivals.

Finally, when measuring the number of game downloads, including the demos, on the entire platform, it can be seen that naive approaches such as TopRating and the comparably simple Slope One technique work sufficiently well to raise the users' interest in individual games. The important result, however, is that with respect to actual sales, the content-based and the item-based methods were clearly better than all others. Overall, it could be demonstrated that recommender systems are capable of stimulating a measurable increase in overall sales by more than 3 percent on the entire platform.