Interactive Visualization of Recommender Systems Data

Christian Richthammer
Department of Information Systems
University of Regensburg
Universitätstraße 31
Regensburg, Germany
christian.richthammer@uni-r.de

Johannes Sänger
Department of Information Systems
University of Regensburg
Universitätstraße 31
Regensburg, Germany
johannes.saenger@uni-r.de

Günther Pernul
Department of Information Systems
University of Regensburg
Universitätstraße 31
Regensburg, Germany
guenther.pernul@uni-r.de

ABSTRACT
Recommender systems provide a valuable mechanism to address the information overload problem by reducing a data set to the items that may be interesting for a particular user. While the quality of recommendations has notably improved in the recent years, the complex algorithms in use lead to high non-transparency for the end user. We propose the usage of interactive visualizations for presenting recommendations. By involving the user in the information reduction process, the quality of recommendations could be enhanced whilst keeping the system’s transparency. This work gives first insights by analyzing recommender systems data and matching them to suitable visualization and interaction techniques. The findings are illustrated by means of an example scenario based on a typical real-world setting.

CCS CONCEPTS
•Information systems → Recommender systems; Personalization •Human-centered computing → Visualization techniques;

KEYWORDS
Recommender systems, interactive visualization, data types

ACM Reference format:
DOI: 10.1145/3099012.3099014

1 INTRODUCTION
According to IMDb\(^1\), 228,414 movie and TV productions have been published in 2016. On Twitter, 500 million tweets\(^2\) are sent every day. Facebook\(^3\) users post 510,000 comments, update 293,000 statuses, and upload 136,000 photos every 60 seconds. Which ones would you like to see?

Information overload has nowadays become a common problem. It describes the difficulty of effectively making decisions when too much information is available [15]. Reasons for this phenomenon are eclectic but can primarily be attributed to the advances in computer and information technology in the recent decades. The mounting digitalization allows to easily and rapidly create new information and share it over various channels.

In this context, recommender systems have proven themselves to provide a helpful mechanism for information reduction. They aim to reduce an initial data set to those items that may be most interesting for a particular user [8, 9]. When looking for a new movie on Netflix\(^4\), for instance, Netflix’s recommender system tries to support the search by presenting a list of movies the user may like.

In the recent years, the quality of recommendations has considerably improved. At the same time, the complexity of the algorithms used for the computation of a recommended set of items has notably increased. For example, the first progress prize as part of the Netflix Prize\(^5\) was awarded to a team of researchers combining more than 100 algorithms [4]. Consequently, the computation appears as a “black box” to the end users because they cannot comprehend how recommendations are computed. This, in turn, leads to high non-transparency [10]. Higher transparency and understandability, however, may result in a higher level of trust towards recommendations [32].

In this work, we address an alternative approach to improve the quality of recommendations by not providing an enhanced algorithm but by involving the user in information reduction using interactive visualizations. Visualizations allow to transparently depict a wide range of information in one picture. Adding interaction techniques, users may dive into the data through filtering or zooming into particular subsets, for instance. In this way, the quality of a recommendation may be enhanced whilst keeping the system’s transparency.

The remainder of the paper is organized as follows. First we give a short overview of the relevant and related work on recommender systems and visualizations in Section 2. We thereby identify the challenges and clarify the research gap. As the selection of suitable visualization techniques is strongly dependent on the data that are visualized, we conduct a detailed analysis and classification of recommender systems data in Section 3. Based on this classification, we then introduce suitable visualization and interaction techniques for the corresponding data. In Section 4, we demonstrate how interactive visualizations could be used in recommender systems in a real-world environment by means of a practical example scenario. Section 5 concludes the paper by summarizing the most important points and discussing aspects for future work.

\(^1\)http://www.imdb.com/search/title?release_date=2016
\(^2\)http://www.dsayce.com/social-media/tweets-day/
\(^3\)https://zephoria.com/top-15-valuable-facebook-statistics/
\(^4\)https://www.netflix.com/
\(^5\)http://www.netflixprize.com/
2 BACKGROUND

2.1 Recommender Systems

The main motivation for the use of recommender systems is to tackle the information overload problem by providing people with only those information and items they are most likely to be interested in [8, 9]. This task has come to be known as the recommendation problem [2]. Over the years, several ways of estimating people's preference values have been proposed, with collaborative, content-based, demographic-based, knowledge-based, and context-based being the basic and generally distinguished recommendation models [3, 8].

Collaborative filtering [13] assumes that people tend to agree with people they agreed with in the past. Thus, users are recommended those items that users similar in terms of their rating behaviors also like. By contrast, content-based filtering [21] assumes that people will like items similar to the ones they liked in the past. It does not rely on any ratings of other users but is solely based on the ratings of the user in focus and the similarities of items in terms of their item features. Demographic-based recommender systems [24] classify users according to their demographic attributes (e.g. age, gender) and generate recommendations based on stereotypes. Knowledge-based recommender systems [7] allow the user to focus to specify preferences in the form of explicit constraints (e.g. “display resolution of at least 1920x1080 pixels”) and provide recommendations matching these criteria. Context-aware recommender systems [1] use various situation-specific factors as additional input data for recommendation generation. By incorporating preferences the users have had in the past when similar contextual conditions have been present, context-aware recommender systems are able to adapt to the users' situational needs.

2.2 Data Visualization and Data Types

Visualizations such as charts or graphs have been successfully used for hundreds of years in order to make sense of rather abstract data [25]. Visualizations help us leverage the human's ability to see patterns and spot trends or outliers [14]. In this way, the understanding of data can notably be supported.

Meanwhile, the “visualization zoo” [14] accommodates a large variety of different visualization techniques. When trying to visualize a data set, however, only a very limited range of visualizations may fit the specific purpose. Thus, the selection of suitable visualizations is a crucial task in information visualization. To give a structured overview of the many types of visualizations available, several taxonomies have been introduced in literature in the recent years. Thereby, the data type of the input data is considered the main characteristic. Based on the classifications of Shneiderman [31] and Keim [19], we here distinguish between three linear data types (one-, two-, and multi-dimensional) as well as four additional data types that can be clearly differentiated from one another (text and hypertext, hierarchies, graphs, and algorithms and software). Unlike Shneiderman [31], however, who considers temporal data as an own data type, we argue that every data set can be viewed in both a temporal and a non-temporal context.

3 MATCHING RECOMMENDER SYSTEMS DATA AND VISUALIZATIONS

3.1 Data Classification

As introduced in Section 2.1, input data and recommendation processes may vary strongly across different recommendation approaches. Nevertheless, the entities and interactions on online platforms with embedded recommender system can be described in a generic way (cf. Figure 1).

![Figure 1: Entities and interactions on online platforms with embedded recommender system](image-url)
### Table 1: Classification of common recommender systems data

<table>
<thead>
<tr>
<th>Information Block</th>
<th>Data</th>
<th>Data Type</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td></td>
<td>Multi-dimensional</td>
<td><em>genre, release year, and age rating are exemplary chosen as item features</em></td>
</tr>
<tr>
<td></td>
<td>genre</td>
<td>One-dimensional</td>
<td></td>
</tr>
<tr>
<td></td>
<td>release year</td>
<td>One-dimensional</td>
<td></td>
</tr>
<tr>
<td></td>
<td>age rating</td>
<td>One-dimensional</td>
<td></td>
</tr>
<tr>
<td>Opinions</td>
<td></td>
<td>Multi-dimensional</td>
<td><em>opinions typically include ratings, tags, and textual reviews</em></td>
</tr>
<tr>
<td></td>
<td>rating</td>
<td>One-dimensional</td>
<td></td>
</tr>
<tr>
<td></td>
<td>tag</td>
<td>Text</td>
<td></td>
</tr>
<tr>
<td></td>
<td>review</td>
<td>Text</td>
<td></td>
</tr>
<tr>
<td>Actors</td>
<td>user in focus</td>
<td>Multi-dimensional</td>
<td></td>
</tr>
<tr>
<td></td>
<td>peer user</td>
<td>Multi-dimensional</td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td>time of day</td>
<td>One-dimensional</td>
<td><em>time of day and mood are exemplary chosen as contextual conditions, price and genre are exemplary chosen as user constraints</em></td>
</tr>
<tr>
<td></td>
<td>mood</td>
<td>One-dimensional</td>
<td></td>
</tr>
<tr>
<td></td>
<td>price</td>
<td>One-dimensional</td>
<td></td>
</tr>
<tr>
<td></td>
<td>genre</td>
<td>One-dimensional</td>
<td></td>
</tr>
</tbody>
</table>

Depending on the application area, users – including both the user in focus and his peer users on the online platform – may consume, visit, use, or just search for different kinds of items such as books, movies, smartphones, restaurants, and web services. For the sake of generality, we refer to this as a transaction. Each transaction happens in a particular context specified by various contextual dimensions (e.g., time of day), and users are able to provide their opinions on transactions and the respective items. This may happen explicitly (e.g., writing a review after a transaction) and/or implicitly (e.g., buying an item). Note that users may not only provide feedback on items but also on other users, for example by assessing the quality of their textual reviews.

From this generalization, the following four information blocks can be derived: items, opinions, actors (including the user in focus and his peer users), and context. These constitute the sources of all input data and are similar to the knowledge sources described by Burke et al. [9] and Felfernig and Burke [11]. Since the selection of suitable visualization techniques depends on the data type of the input, the data considered in each information block need to be further analyzed. The results of the analysis are summarized in Table 1. Subsequently, we discuss the classification in more detail and show how the respective data types can be visualized.

### 3.2 Visualization Techniques

While the pool of suitable visualization techniques for each information block is limited to those fitting the identified data type, the selection of the most useful and usable technique is hard to measure and a quite subjective estimate. Therefore, we introduce visualization techniques commonly used to illustrate data of the identified types.

#### 3.2.1 Items: item features

The items that are recommended to the user are the central entity of a recommender system. Items can be characterized by several item features such as genre, release year, or age rating. This constitutes multi-dimensional data when taken as a whole, but the single features are one-dimensional. Typical visualization techniques commonly used for multi-variate data are parallel coordinates [16, 17] or star plots [5]. In such visualizations, each attribute (or dimension) is depicted by an axis.

Taking parallel coordinates as an example, the axes are laid out in parallel side by side. The items of a data set are represented by polylines connecting vertices on the parallel axes. Figure 2 shows an example. Obviously, the recommended items refer to the four genres “horror”, “thriller”, “western”, and “animation”. The colored polylines furthermore reveal that nearly every movie that has an age rating of 18 is attributed to the genre “horror”, whereas animated movies are all intended for a younger audience. In order to show which polyline belongs to which movie title, a mouse hover effect could be added.

![Figure 2: Parallel coordinates visualization of an example item set with one movie highlighted](image)

#### 3.2.2 Opinions: feedback data

For many recommender systems (especially collaborative filtering), opinions constitute the most important data source when calculating a recommendable item
set. Opinions may include numerical ratings (e.g. five-star ratings) as well as textual reviews or tags. Numerical ratings, on the one hand, are a typical example for one-dimensional or univariate data. Here, only one attribute is dependent on one or several independent attributes. Visualization techniques commonly used for one-dimensional data are bar charts or pie charts. Figure 3 shows an example bar chart visualizing five-star ratings.

Reviews and tags, on the other hand, are typical examples for textual data. Textual comments are a helpful alternative to numerical ratings which cannot reveal anything about the reasons for a specific rating. A visualization technique to depict the content of textual data in one picture that has become popular in the recent years is the tag cloud. Tag clouds display the most important tags or words of a text. Thereby, the text size of a tag varies in accordance to its importance (mostly based on the frequency of usage). Figure 4 depicts an example tag cloud for feedback towards a horror movie. Here, the color of the tags refers to the related rating value.

3.2.3 Actors: user in focus & peer users. The information block “actors” includes the user in focus and his peer users. The users involved in recommender systems are characterized by several demographic attributes such as age, gender, and occupation (multi-dimensional). As actors may rate each other, the information block as a whole can be considered a network with actors as nodes and rating relations as unidirectional edges. Using this network structure as a basis for computing recommendations, clusters of users related to each other may easily be identified.

When trying to visualize network data, one can choose between a node-link representation, a matrix visualization, or a hybrid approach. While node-link representations are particularly useful for small graphs, matrix visualizations allow to display large graphs and dense networks. By applying node-ordering, structures and sub-communities can transparently be revealed. Figure 5 shows a rating matrix, where colored fields display a rating relation. Obviously, the cluster on the upper-left side only provided positive ratings for each other, which suggests the assumption that those users share the same taste.

3.2.4 Context: conditions & constraints. The information block “context” fills a special role in this analysis as it helps to specify the boundaries of a presented data set. Users, for instance, might be interested in movies of a particular genre or may only want to consider ratings that have been created by people having a good mood. Thus, context information is particularly valuable if integrated in item-related or opinion-related visualizations. Figure 6 shows an example parallel coordinates visualization of a set of recommended movies (items). Here, the constraint “category = thriller” has been added by the user through clicking on the corresponding category.

\[^6\text{Note that numerical ratings may not only be provided explicitly but also be derived from user behavior such as buying an item, which would be mapped to 1 for “has bought” as opposed to 0 for “has not bought.”}\]
4 EXAMPLE SCENARIO

Having introduced suitable visualization and interaction techniques for the different data types relevant in connection with recommender systems, we now illustrate how interactive visualizations may support the information reduction process in a typical real-world setting.

Suppose it is Saturday evening and Bob wants to watch a movie together with a group of friends. In general, Bob is not very knowledgeable in the movie domain and does not have any clear preferences in terms of favorite genres. Because of the increasingly large number of movie options available through various video on demand providers, finding a movie to watch thus constitutes a problem of information overload for Bob.

Recommender systems are able to determine those movies out of the sheer endless set of available alternatives that Bob is most likely to enjoy. However, the decisions of traditional “black box” recommender systems are hard to follow because they do not provide any information on the nature of the input data. This is especially relevant if Bob is unsatisfied with the final recommendations. Therefore, our concept of interactive visualizations for recommender systems data uses traditional “black box” computations only to reduce the set of all alternatives to a manageable size (instead of directly reducing it to the final set of recommendations). This preparatory step is needed because including all available options in one picture would make the corresponding visualizations unreadable.

The resulting set of candidate items can then be visualized in terms of the features of the items. As introduced in Section 3.2.1, parallel coordinates are a suitable visualization technique for this task. An example is depicted in Figure 7, where the item features of the movies with the highest predicted ratings are visualized in terms of their genre, release year, age rating, and running time. Moreover, the movies are listed according to their predicted rating (in descending order) below the parallel coordinates. The predicted rating says how Bob would assess a movie according to the recommender algorithm.

Through this first visualization, Bob gains a quick overview of the items with the highest predicted ratings and their respective item features. In the example, he realizes that all candidate items belong to one of the following genres: action, drama, romance, or music. Even though Bob indeed likes all of these genres, he thinks that an action movie or a drama is most appropriate for an evening with a group of friends. Moreover, he decides that the movie should be from the last ten years, have an age rating of 16, and have a running time of at least 150 minutes.

Therefore, he interacts with the parallel coordinates visualization to reduce the set of polylines to those fitting these criteria (cf. Section 3.2.4). We argue that this approach is more meaningful than specifying such criteria upfront, because in the beginning Bob has no idea about how many items are satisfying which constraints. As Figure 8 shows, applying the criteria leads to changes regarding the top recommendations because the ones with the highest predicted ratings do not satisfy the specified constraints. In particular, the list of candidate items is reduced to six movies, which both have a high predicted rating and fit Bob’s criteria.

5 CONCLUSION

In the past, researchers have focused on improving recommender systems by developing and optimizing increasingly complex algorithms that are supposed to reduce an unmanageably large amount of information to those information the users are most likely to be interested in. Even though these efforts have been successful in terms of higher quality of recommendations, they have led to high non-transparency because end users are no longer able to comprehend the processes underlying the recommendation generation. Therefore, this paper did not aim at develop yet another recommendation technique to improve the quality of recommendations but at involving the end user in the information reduction process using interactive visualizations.
As opposed to existing works that are limited to one specific use case and one suitable visualization technique, we wanted to introduce this approach in a comprehensive manner and on a general level. To this end, we first classified different kinds of recommender systems data into information blocks and commonly accepted data types. We then matched these data types to suitable visualization and interaction techniques and introduced several example visualizations for each information block. Moreover, we discussed an example scenario describing how interactive visualizations can support the user’s entire decision process.

Even though this work constitutes only a starting point, we hope that discussing interactive visualization for recommender systems from the ground up serves researchers as a basis for future work in this area. After all, recommender systems will continue to be an active research field but its focus may be about to turn away from algorithm optimization and towards other ways of enhancing the user experience.

ACKNOWLEDGMENT

The research leading to these results was supported by "Bavarian State Ministry of Education, Science and the Arts" as part of the FORSEC research association (http://www.bayforsec.de/).

REFERENCES


Figure 9: Bar chart visualizing the overall rating distribution of “The Hateful Eight”