

Benedikt Loepp, Katja Herrmann, Jürgen Ziegler

Merging Interactive Information Filtering and Recommender Algorithms – Model and Concept Demonstrator

Abstract: To increase controllability and transparency in recommender systems, recent research has been putting more focus on integrating interactive techniques with recommender algorithms. In this paper, we propose a model of interactive recommending that structures the different interactions users can have with recommender systems. Furthermore, as a novel approach to interactive recommending, we describe a technique that combines faceted information filtering with different algorithmic recommender techniques. We refer to this approach as blended recommending. We also present an interactive movie recommender based on this approach and report on its user-centered design process, in particular an evaluation study in which we compared our system with a standard faceted filtering system. The results indicate a higher level of perceived user control, more detailed preference settings, and better suitability when the search goal is vague.

Keywords: Models, Recommender Systems, Interactive Recommending, Information Filtering, User Interfaces.

1 Introduction

With the ever growing amount of information on the Web, recommender systems have come to play an important role in supporting users when searching for information items or products they are interested in [31]. In the domain of electronic commerce, recommender systems (RS) fulfill different, equally important roles: They act as a tool supporting the user's search process and as a marketing instrument on the part of the information provider. While existing RS often produce recommendations that match the user's interests and goals well, most RS afford little or no user interaction, and, in particular, lack options to control how recommendations are produced. A further problem is the lack of transparency that may hinder users in comprehending why a particular item is recommended [36]. As a consequence, acceptance of the recommendations

and trust in the system may be reduced [40]. Since most RS require the availability of a user-specific preference profile, they suffer from the cold start problem when no information about the current user's preferences is available. Also, users often do not want their preferences to be stored due to privacy concerns. Furthermore, a long-term user profile may differ from the user's current interests, not taking into account the situational, context-dependent aspects of the user's search and decision process. All these issues may result in reduced usability, trustworthiness and user acceptance of RS [15,29,36,40].

While RS research has traditionally been focused on optimizing the used recommendation algorithms, there is an increasing awareness that this endeavor has its limitations since further incremental improvements of existing recommender algorithms may not lead to a commensurate increase in user satisfaction. This may be due to the observation that variability in user goals or product valuations is often much larger than the additional precision gained by an improved algorithm [16,23]. Only more recently, several researchers have suggested to focus more on user aspects of RS, including the user's interaction behavior, user interface design, and the resulting user experience [15,29]. It has been shown, for instance, that users are not only interested in receiving precise recommendations and in lowering their search effort, but also in having a more active role in the entire recommendation process [40]. Users may be willing to invest more effort and even accept less accurate system recommendations if they are able to exert more influence over the system [16]. Thus, providing users with more interactive control over the recommendation process is an important goal for RS research.

The contribution we make in this paper is twofold. As a first contribution, we describe a model of interactive recommending that structures the different types of interactions users can have with a recommender. The model describes three interaction cycles according to whether users interact 1) with the application in which the RS is embedded, 2) with

explicit representations of their preferences, or 3) with the generated recommendations themselves. The model assumes that user interaction is tightly interwoven with the generation of recommendations, typical dialog-based recommenders with sequential question-answer steps are not in the focus. The model can serve as a basis for classifying the different phenomena involved in interactive recommending, and, at the same time, provides a design space for exploring the different interactive functions that may be made available in a RS.

As a second contribution, we describe a novel approach to interactive recommending which we call *blended recommending* [19] that merges interactive faceted filtering techniques [11,41] with algorithmic recommender functions. As a proof-of-concept, we present the interactive movie recommender *MyMovieMixer*, that was initially introduced in [12]. This demonstrator system employs different recommender techniques and can thus be described as a hybrid recommender [4] in conventional RS terminology. However, it integrates the algorithmic recommender functions with fuzzy techniques and interactive information filtering methods. We also report on the user-centered design process we applied and, in particular, a comparative evaluation study we performed to assess the system's utility and usability.

The present paper is an adapted and extended version of work we have reported in [19]. We extend the prior publication by introducing the novel model of interactive recommending that generalizes concepts and design considerations we developed in context of the prototype interactive recommender *MyMovieMixer*. The system is described in a more focused manner as one instance of this model, which, however, opens up a much wider design space for interactive recommending than is covered in this application. The description of related work is extended to cover these additional aspects. We also describe the user-centered design process we followed and, in particular, report data unpublished so far that we obtained in the evaluation of the system.

2 Interactive Recommendation and Information Filtering Approaches

Recommender systems aim at suggesting items that match the user's interests and preferences, typically represented in a long-term user model. Well-matched recommendations can contribute to reducing interaction effort and cognitive load [29]. However,

since the search goal may vary in different situations users might be dissatisfied or feel too much dominated by the system because influencing the recommendation process is mostly not (or only partially) possible. Relying (only) on long-term user profiles makes it difficult to react to situational needs [7] and may lead to filter bubble effects [26]. More recently, interactive recommending approaches have been proposed to overcome these usage-related issues. For example, applying the *relevance feedback* principle [34] in RS allows users to refine the results, which may increase perceived user control. However, in this case, the existing user profile is just modified. Moreover, the required profile information is often not available, or not sufficiently detailed to generate accurate recommendations. While several works try to solve such cold start problems algorithmically [14,44], capturing user preferences interactively can be seen as a promising alternative.

Critique-based RS [6] allow users to criticize features of the currently recommended items, based on the assumption that this is easier than formulating a search goal up-front. Users can thus iteratively refine the result set towards their search goal, e.g., by requesting longer movies or films by a different director. Visual support and direct manipulation of the criticized features can have positive effects on comprehensibility, user-friendliness and interaction effort [43]. Efficiency can also be increased by dynamically suggesting one or more features to be criticized [6] as well as by taking into account interaction histories from previous sessions of similar users to adapt the critiquing process [21]. However, critiquing usually requires predefined product attributes which are often not available. Recently, interactive *preference elicitation techniques* have been proposed that do not require pre-specified product attributes but use, for instance, latent factors automatically derived from other users' ratings [20] or depend entirely on user-defined tags [38]. *MovieTuner* [38], for example, automatically weights tags and presents users with the most important ones. Users can then explicitly indicate a preference for movies with, e.g., less humor and more violence. While expressing preferences in such a way can be useful, there is typically no integration with other feature types, thus, users cannot simultaneously select and weight their preferences from a wider range of, e.g., predefined content information, tags and latent factors.

Only a few systems use *interactive visualizations*, and especially hybrid approaches [4] are typically not controllable by users. *TasteWeights* [2], an interactive hybrid music recommender, is one of the few exceptions. Here, users can directly manipulate graphically connected widgets and weight the influence of different information types and social data sources, which lead to higher perceived recommendation quality and understanding of how results were generated. *SetFusion* [27] employs a common hybridization strategy [4], but allows users to change the influence of the different recommender algorithms individually. Several interactive features are provided (e.g. a Venn diagram visualizing the result set), but the system still requires a persistent user profile and does not allow to explicitly select and weight individual content-related filter criteria. Another example of a more interactive hybrid RS is the browser plugin *MovieBrain* [8] that enhances the Internet Movie Database (IMDb)¹ with interactive filters to generate movie recommendations matching the user's situational needs. But, apart from filtering out particular genres, it also does not take further content information into account.

While RS can be helpful tools to support a user's search process, there is also a broad range of manual information filtering techniques outside the RS field that have proven to be effective in helping users find the items they want. Faceted filtering [41] is one of the most prominent and successful examples. It supports exploration and discovery [11,41] of large item spaces by selecting values from a set of facets, thus iteratively constraining the item space until the desired result is found. Faceted search is also used to enhance conventional keyword search and to support more flexible navigation [11], e.g. in digital libraries or online shops. Early filtering approaches often rely on predefined sets of filter attributes, typically implement only hard Boolean filtering, allow just for conjunctive queries and consider all facets equally important [33,35,37,39]. While most approaches perform an exact matching of the facet values, a few systems apply fuzzy matching to deal with misspellings and similar values [10]. A number of more recent systems automatically extract facets and facet values, and apply adaptive techniques to faceted search, based on, for instance, semantic [5] or social [37] data sources to facilitate the user's selection of suitable filter criteria and to deal with lack of

metadata. However, the user's influence on the current filter setting is still limited and, from a user's perspective, items may not get sufficiently described this way [11]. *VizBoard* [39] is one of the few systems that not only suggests facets or facet values but also allows users themselves to prioritize the selected criteria. Thus, results can be ordered more appropriately while excluding relevant items is avoided. Recent work also investigated user experience of faceted search as well as integrating visualizations. *IVEA* [35], for instance, uses a matrix visualization to display documents and their relevance according to the selected facets based on TF-IDF heuristics. While research in faceted filtering has produced a range of promising methods, including intelligent methods to extract and adapt filters, they have thus far, to our knowledge, not been combined with recommender functions.

Against this background, blending interactive information filtering with recommender techniques seems to be a promising approach to overcome limitations of the individual approaches. Generally, increasing user control and interactivity in RS as well as improving user experience have been described as important design goals [16,29] but are still not optimally realized in existing systems. Introducing more interactivity in RS, however, can be achieved in many different ways which raises the question how the various options can be mapped out in a more systematic manner. Several authors have proposed ways to classify and structure the different aspects that pertain to user interaction with a RS. One perspective relates to the process and the cognitive activities that users perform when moving from an initial intention to the final selection of an item. The need for a better understanding of users' information seeking behavior has been stated, for example, in [24], where it is argued that "recommender systems need a deeper understanding of users and their information seeking tasks to be able to generate better recommendations". While several models of users' information seeking behavior have been proposed in the field of information retrieval (cf. [18,22]) these models seem not directly applicable to recommenders due to their focus on document collections. With a somewhat similar intention, a model showing different phases of a recommendation process with feedback cycles has been proposed in [30], identifying the four main phases: preference

¹ <http://www.imdb.com>

specification, recommendation generation, revision of preferences, and final decision. In other work, for instance, more general models of interaction in conversational [32] or critique-based [6] recommenders are presented. Another distinction can be made with respect to the methods by which user preferences are elicited. Here, explicit preference specification is often distinguished from implicit methods [13]. While in explicit preference specification a user consciously states desired properties of items or rates them, implicit methods attempt to learn the user’s preferences from a range of behavioral parameters such as clicking on an item to view details, or how long the user views the description of an item. Overall, however, a more general model of interactive recommending, representing the different objects a user may be interacting with and the interactive processes involved, is still missing.

3 A Model of Interactive Recommending

As outlined in the previous section, there are several useful models that either structure the user’s search process or that distinguish the different methods by which user preferences and feedback can be captured. Regardless of these proposals, a model that explicates the different interactions which may be tightly interwoven with the generation of recommendations, and that integrates goal-driven search behavior with reactive, response-driven user interactions, is still missing. To structure, detail and illustrate the different aspects of interactive recommendation processes we therefore propose a model that distinguishes different levels of user interactions as well as system components that take part in the process (Figure 1). The model presents three different interaction cycles that may be involved in interactive recommending. It must be noted, though, that not all interactive approaches to recommending will comprise all features shown in the model. In this sense, the model can also serve as a design space that helps to identify useful functions not yet present in current systems.

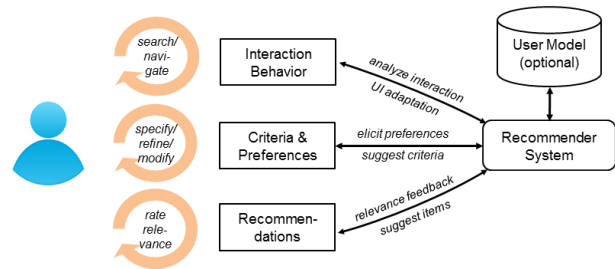


Figure 1: A model of interactive recommending.

The model distinguishes three interaction cycles through which the user can potentially influence the recommendation process as well as the learning of a user-specific preference model. This user model, however, is an optional component which may not be present in all instantiations of the model. The three cycles describe the user’s interaction at three different levels: First, users usually interact with the application in which the recommender is embedded without necessarily interacting with the recommended items. At a second level, users can explicitly specify or modify their preferences, either as input for a (long-term) user model or in an interactive preference elicitation process that directly influences the recommendations produced. Thirdly, users can interact with the recommendations themselves by providing feedback on the relevance of the suggested items or by selecting them for subsequent actions, for instance, when buying a product.

Conventional recommenders essentially use only the last one of these interaction cycles: the system calculates items that match an existing user model and presents them to the user. In most systems, users can only view the recommended items or select them if they fit their needs, in some cases users can also provide feedback on the relevance of the items presented or exclude items they do not want. Providing relevance feedback may lead to an update of the user model. In either case, this standard approach is very limited in terms of interactivity. The approach is somewhat extended in critique-based RS where users can select an item that is close to meeting their wishes, and request changes in one or more properties of that item, thus partially expressing their preferences based on characteristics of the suggested items. Both cases can be useful when the user has not yet mentally formed a search goal or the preferences are unclear.

The approach can be further extended in the second interaction cycle by letting users explicitly

MMM My Movie Mixer

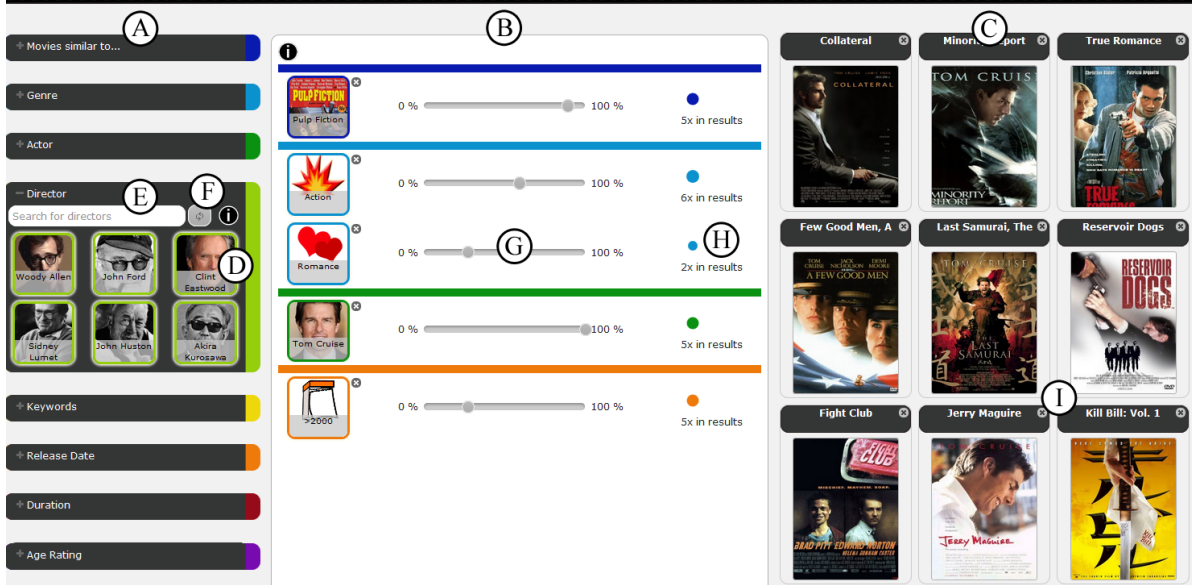


Figure 2: The MyMovieMixer application: widget area (A), work area (B), result area (C), tile representing a facet value (D), input field to search values (E), shuffle button to receive a new set of suggested tiles (F), slider to adjust a tile's weight for the recommendations (G), visualization of the number of movies fulfilling the criterion (H), button to dismiss a recommendation (I).

indicate their preferences independently of specific items. Stating preferences in advance, for example by rating a set of items prior to recommending, has been frequently applied but cannot be classified as interactive recommending because preference elicitation and recommending happen in two separate phases. Online preference elicitation on the other hand, i.e., specifying preferences by selecting and possibly weighting desired item properties in parallel to the recommendation process, is a much less explored area. The system *MyMovieMixer* we present in this paper aims at filling this gap by relying on the concept of blended recommending, i.e., integrating algorithmic recommenders with interactive information filtering methods. From a cognitive perspective, allowing users to explicitly specify, refine and modify their preferences supports them in situations when they have formed their preferences to some extent as well as when they react to proposed items (and possibly to item features suggested by the system) in a situated manner to incrementally develop their search goal.

The top-level cycle in Figure 1 refers to the general interaction with the application, e.g. an online shop, where the user's interaction behavior, such as navigating between item categories or viewing the details of an item can be used to learn user

preferences. While the other two interaction cycles can provide either explicit or implicit feedback directly linked to preferences or the recommendations, this cycle can only be used for deriving preferences implicitly. While users may mostly not be aware of the fact that their preferences are learned from their general interaction with the system, system feedback and explanations could be provided to inform users of the effects of their interactions and, thus, to increase transparency.

With the presented model, we aim at shedding some light on the different options that can be used for making recommendation processes more interactive and user-controllable. It also indicates possibilities for supporting goal-directed search and reactive, situated behavior in a more integrated manner. Finally, the model can serve for defining functional components of interactive RS that support the three interaction cycles presented.

4 MyMovieMixer: An Example Application of Blended Recommending

MyMovieMixer (MMM, Figure 2) is a web-based application we developed to demonstrate the concept

of blended recommending [19]. It combines the benefits of hybrid RS with the ones of information filtering interfaces by integrating the respective methods to recommend movies from the MovieLens dataset². For flexible use in different contexts (e.g., various moods, presence of different people, cold start situations), the recommendation process is entirely based on explicit user input given during the current session. For example, a user may indicate (as shown in Figure 2) interest in watching movies similar to *Pulp Fiction* that also contain elements of the genres Action and—somewhat less relevant—Romance. In addition, the user likes the actor Tom Cruise and would to some extent prefer a movie from the last decade. Although it would be possible to consider a user’s long-term profile as well, this is not required for the approach.

MMM allows users to directly manipulate the different filters and their corresponding weights, and immediately visualizes the effects on the resulting recommendations, thus increasing user control and making different settings easy to understand. In the following, we describe the interaction concept of MMM, the different kinds of facets, as well as the algorithmic details of calculating movie relevance scores.

4.1 User Interface and Feedback Mechanisms

The workspace of MMM consists of three main parts (Figure 2): The area on the left-hand side (A) presents facets from which the user can choose filter criteria. The work area (B) shows the selected criteria and sliders by which users can change their degree of influence, while the resulting recommendations are shown on the right-hand side of the screen (C).

Facets (A) are represented by menu-like widgets, which when expanded show a number of rectangular tiles (D) representing possible criteria (facet values), visualized with images where possible. For facets with many values, users can add tiles by using a search box (E) with auto-completion [11]. Moreover, users can request a new set of values by pressing the shuffle button (F). The system then suggests tiles based on the values that occur most frequently in the current results allowing users to further refine the results. To specify their preferences as input to the recommendation process, users can drag tiles into the

work area (B). The weight of each corresponding criterion can then be manipulated with the associated slider (G) to change its influence on the resulting recommendations. Adding criteria or changing their weights immediately updates the result set, so that users obtain instant feedback on their preference settings. Since it may not be possible to fulfill all criteria specified, MMM provides textual and graphical feedback (H) how often the criterion occurs in the current recommendations. In correspondence with the second cycle of our model of interactive recommending, users can thus specify and refine their preferences supported by the system through different feedback mechanisms, including suggestion of criteria. To generate the ranked list of recommendations (C), an overall relevance score is calculated for each movie by aggregating the movie’s relevance values with respect to each selected criterion and also considering the respective weights. Users can drag movies from the resulting recommendation list into the work area to further refine their preferences. In addition, they can remove recommended movies they are not interested in (e.g., because they do not like them or have already seen them) for the current session by clicking the x-button (I). This way, the model’s third cycle is addressed.

Besides its explorative interaction concept, MMM offers a range of additional means for better comprehension of the recommendations. For example, users can open a details view for each recommended item that also explains why it was recommended, i.e., which criteria were satisfied for this recommendation. In addition, recommendations that fulfill a criterion are highlighted when the user hovers over a tile or changes its weight. The system thus combines features for preference specification and refinement as well as critiquing and helps to understand the sources of the recommendations in the hybrid setting.

4.2 Facet Types, Filtering Methods and Relevance Calculation

MMM offers a range of different facets, labeled: *Movies similar to...*, *Genre*, *Actor*, *Director*, *Keywords*, *Release Date*, *Duration*, and *Age Rating*. While different methods are used to calculate item relevance depending on the specific facet type and the

² <http://grouplens.org/datasets/movielens/> (The MovieLens 10M dataset contains about 10 million ratings and 95 000 tags from more than 70 000 users for over 10 000 movies.)

underlying data, the approach is flexible and can also be used with other facet types or methods. Internally, MMM acts like a weakly coupled hybrid recommender [4], i.e., it handles all criteria separately at first. We now describe the different facet types and the method used to calculate the result set by aggregating the specified facet values and their weights.

For each movie m and each criterion c_i a value between 0 and 1 is determined. This value represents the degree with which m fulfills a criterion. Depending on the type of criterion, the calculation of this fulfillment degree is done in different ways:

- Boolean filtering: If the user selects a criterion from a facet such as movie genre, director or age rating, each movie with this value will be considered in the results while the other ones will not be taken into account. This may lead to large number of items receiving a value of 1, i.e., these items would be ranked equally regarding their fulfillment degree. To avoid this, we assume that the more popular of these items are also the important ones for the users and thus apply an artificial ordering on these items based on the movies' average rating and the number of ratings they have received (for more details, see [19]).
- Fuzzy filtering: We use Fuzzy Logic [42] to implement a soft filtering for criteria such as a movie's release year to avoid the need for exact matches as in most filtering systems. For instance, selecting a specific decade (e.g. the 1990s) would also include, although with linearly decreasing relevance, movies released some years before or after (e.g. a movie from 1989 will not be completely ignored as it would be the case in Boolean filtering). This also applies to the length of a movie, where users can choose multiple time spans. Using a fuzzy membership function, movies falling within these time spans receive full weight while movies in between are considered to be less relevant.
- Collaborative Filtering: From the *Movies similar to...* facet, users can select movies they like. Movies rated similarly by other users are then considered for the recommendations with increased relevance. For this purpose, we integrate the most popular recommendation method, Collaborative Filtering (CF) [31]. To

determine similar movies, we utilize the ratings given by other users in the MovieLens 10M dataset and calculate similarities between the selected movie m and all other movies by means of their latent factor vectors using a common Matrix Factorization [17] recommender³ and a Euclidean distance metric. This item-based CF approach allows users to take more than just content-related metadata of the items into account, what is often problematic or even not possible in information filtering systems [11,35,37].

- Content-based Filtering: For the actor and keyword facet we use conventional content-based recommender methods [31]. For instance, we calculate the relevance of a movie with respect to a certain keyword the user selects via TF-IDF heuristics [1]. Inspired by *MovieTuner* [38], we consider tags as terms and the set of tags associated with a movie as a document, and calculate the relative importance of each tag for this movie. This allows us to give those movies a high relevance value that are very specific for a certain keyword. Regarding the actor facet, relevance is determined based on the actor's importance (a value given by the dataset) in the particular movie.

Finally, the items can be sorted with respect to each criterion, e.g., by fuzzy values, item similarities or TD-IDF scores. For each movie m and each criterion c_i we determine the relevance value $rel_i(m, c_i) \in [0; 1]$ according to the movie's position in this sorted result list. An overall relevance score rel for each movie m is subsequently calculated in accordance with Multi Attribute Utility Theory (MAUT), an approach frequently used in critique-based RS [6]. With respect to all criteria, this score aggregates the relevance values from all n tiles and the weights w_i the user has expressed by using the sliders with a weighted arithmetic mean:

$$rel(m, c_1, \dots, c_n, w_1, \dots, w_n) = \frac{\sum_{i=1}^n w_i \cdot rel_i(m, c_i)}{\sum_{i=1}^n w_i}$$

Finally, the movies are sorted in descending order with respect to their overall relevance score and the movies with the highest values are presented to the user. Table 1 illustrates the calculations with a small example, where a user searches for a movie directed

³ *FactorWiseMatrixFactorization* from the MyMediaLite [9] recommender library.

by Steven Spielberg (criterion c_1 with weight $w_1 = 100$) from the 1990s (c_2 with $w_2 = 50$). For demonstration purposes, we assume that the dataset consists of only three movies and dispense ordering the movies in case of equal relevance scores rel_i .

Table 1: Relevance calculation for some example movies.

| Movie | $rel_1(m, c_1)$ <i>(Director)</i> | $rel_2(m, c_2)$ <i>(Release)</i> | $rel(m, c_1, c_2, 100, 50)$ <i>(Overall relevance)</i> |
|--|--------------------------------------|-------------------------------------|---|
| Indiana Jones 3 <i>(Spielberg, 1989)</i> | 1.0 | 0.5 | 0.833 |
| Jurassic Park <i>(Spielberg, 1993)</i> | 1.0 | 1.0 | 1.000 |
| Pulp Fiction <i>(Tarantino, 1994)</i> | 0.0 | 1.0 | 0.333 |

By applying the ranking technique described, we avoid the conjunctive application of filter criteria as it is used in most information filtering approaches [33,35], and are thus always able to provide users with a ranked recommendation list matching their stated interests best. Nevertheless, there still may be filter settings that lead to too few results. In these cases, we extend the recommendation set dynamically with movies similar to the recommended ones based on their latent factor values.

5 Evaluation

The development of MMM followed a user-centered design process with multiple user studies. First, we performed a preliminary study (n=22) to evaluate several layout aspects (e.g. tile design, ordering of facets inside the widget area). Second, we implemented a basic prototype to assess the users’ visual impression of the interface in a follow-up study (n=30). Using the VisAWI questionnaire [25] we evaluated aesthetic aspects of the user interface. Overall, participants gave positive ratings and valuable feedback that contributed to further development. Third, we implemented a first running version of MMM (already quite similar to the one described in this paper) and performed a user study (n=30) focusing on usability aspects, usage-related problems and general acceptance of the blended recommending approach [12]. The participants—all not involved in the previous studies—stated a high usability and responded very positively to questionnaire items regarding ease of use and comprehensibility of MMM’s specific interaction elements such as tiles and sliders. In particular,

participants seemed to enjoy using MMM because of its novel and intuitive interaction concept leading to meaningful recommendations. Nonetheless, feedback given in this study led to further improvements. For instance, tiles shown in the widgets were initially randomly chosen instead of considering the corresponding values’ frequency in the current result set. We also extended the feedback mechanisms provided to improve understanding of, e.g., how different slider settings affect the recommendations. Further modifications were introduced regarding the interaction concept (e.g. drag and drop was not so extensively used before), widget and tile handling as well as the score aggregation. Finally, we used the revised system (described in this paper and in [19]) to perform another user study comparing MMM against a standard filter system. Most of the results can already be found in [19], but in the following we will briefly describe the study again and report further results to reveal additional insights into users’ interaction and their perception of the blended recommending approach.

5.1 Goals and Setting

Since blended recommending can be seen as an integration of faceted filtering and recommender techniques, we compared MMM with a conventional filtering interface to evaluate the effectiveness and the interaction quality of the system. Due to its high level of interactivity and controllability, a filter interface appears to be a useful baseline and a more natural competitor than conventional RS which typically require existing user profiles and lack interactive features for expressing user preferences. We hypothesized that users interacting with MMM would have a stronger feeling of control while the quality of the results and the usability of the system would be at least as good as for a standard filter interface. Moreover, we expected a better suitability for varying situational needs. In particular, we assumed that a filter interface would be preferred when users are aware of their search target whereas they would be in favor of MMM when they have no or only a vague search goal, which is often the case in large domains and, especially, for experience products such as movies.

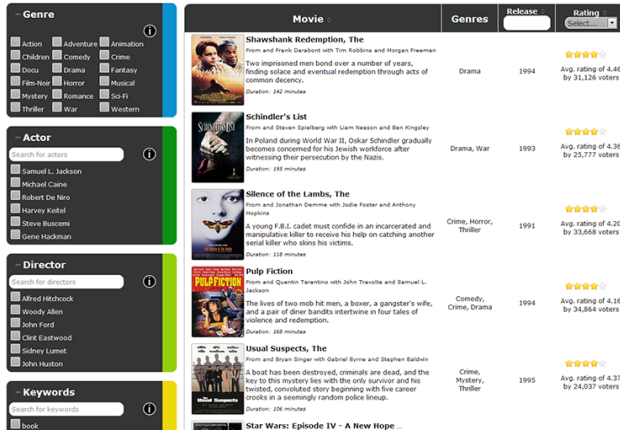


Figure 3: Screenshot of the alternative faceted filtering system we implemented to compare MMM against it.

For the purpose of the study, we thus implemented a faceted filtering system (FFS, Figure 3) as an alternative condition and extended both interfaces with typical shopping cart functionality. The FFS used the same facets (except *Movies similar to...*, as this is a recommender-specific feature), values and dataset as MMM. We further adopted the interface design and implemented all features as similarly as possible. Initially, the items were ordered with respect to their overall popularity, but could also be sorted differently by the user. However, FFS did not allow to weight criteria and used only Boolean AND operations, as is typical in standard faceted filtering.

5.2 Method

We recruited 33 participants (20 male, 13 female, average age of 27, $\sigma=6.46$) for the user study, which was conducted over two weeks designed as an experiment under controlled conditions. Participants used a desktop PC with a 24" LCD-display (1920×1200px resolution) and a common web browser. The two different conditions (MMM and FFS) were tested in a between-subject design as in a within-subjects option participants' use of one system might have too much influenced their behavior with the other. To avoid lowering the validity of the study for the intended usage scenarios, we thus randomly assigned each participant to different groups (MMM: $n=17$, FFS: $n=16$).

After a brief introduction by the moderator to the experiment and the system used, participants were asked to perform two tasks subsequently, which were equal for both conditions:

1. The first task can be seen as a training trial for the respective system, allowing participants to learn using its interface. Users were asked to assume that they want to buy a DVD as a gift for a friend who prefers movies from the genres Action and Romance, and especially likes the actor Brad Pitt.
2. The main task involved finding items matching the participants' personal interests. Therefore, they were allowed to use all features of the respective interface and were not restricted in time. While freely interacting with the system, they were asked to add movies (at least one) they actually would like to watch to the shopping cart.

We recorded the interaction as a screencast for later evaluation and measured task times as well. After performing the tasks, participants filled in a questionnaire comprising items we gathered from [15,28] for evaluating interaction and recommendation quality, using a positive 5-point Likert scale (1–5). Furthermore, we used SUS [3] to assess the systems' usability, asked participants further questions specific to MMM as well as regarding their familiarity with the movie domain, their knowledge about movie portals and web product search (again using a positive 5-point Likert scale), and collected demographic data.

5.3 Results

Among others, the questionnaire data led to the results shown in Table 2, which are for the most part already reported in [19]. However, it is worth mentioning that MMM performed significantly better in terms of control and interaction adequacy while the interface adequacy and the usability of MMM are on the same level as FFS with its limited interaction possibilities. In addition, our assumptions regarding the systems' suitability for different situations of use were confirmed by the results.

Table 2: Results regarding interaction, recommendations, usability, and the suitability of the respective system for different situations. Significant differences are marked with *.

| | MMM | | FFS | | |
|---|----------|----------|----------|----------|------------------------|
| | <i>M</i> | σ | <i>M</i> | σ | |
| Control [28] | 4.43 | 0.50 | 3.85 | 0.99 | $t(22)=2.10, p<.05^*$ |
| Interaction Adequacy [28] | 3.94 | 0.53 | 3.13 | 1.00 | $t(22)=2.90, p<.01^*$ |
| Interface Adequacy [28] | 4.07 | 0.40 | 3.86 | 0.60 | $t(31)=1.21, p>.05$ |
| Perceived Rec. Quality [15] | 3.99 | 0.45 | 4.15 | 0.48 | $t(31)=-0.96, p>.05$ |
| Perceived System Effectiveness [15] | 3.66 | 0.51 | 3.45 | 0.45 | $t(31)=1.29, p>.05$ |
| Perceived Rec. Variety [15] | 3.15 | 0.88 | 3.41 | 0.96 | $t(31)=-0.81, p>.05$ |
| Usability (SUS [3]) | 82.35 | 14.80 | 83.59 | 12.35 | $t(31)=-0.26, p>.05$ |
| Suitability when looking for a specific movie | 2.47 | 1.46 | 3.50 | 1.27 | $t(31)=-2.16, p<.05^*$ |
| Suitability with an approximate search goal | 4.24 | 0.66 | 4.31 | 0.70 | $t(31)=-0.32, p>.05$ |
| Suitability with no clear search direction | 4.13 | 1.09 | 2.80 | 1.27 | $t(29)=3.13, p<.01^*$ |

Users also felt to be able to influence the recommendation process (MMM: $M=4.06, \sigma=.90$; FFS: $M=3.75, \sigma=.93$) while the perceived interaction effort was rated highly acceptable for both conditions (MMM: $M=4.47, \sigma=.73$; FFS: $M=4.25, \sigma=.68$), without any significant differences. Overall, users were satisfied with both systems (MMM: $M=3.76, \sigma=1.03$; FFS: $M=3.69, \sigma=.87$).

The number of selected movies (MMM: $M=7.18, \sigma=5.81$; FFS: $M=7.21, \sigma=6.02$), the duration of the main task (MMM: $M=6.18$ min, $\sigma=2.25$; FFS: $M=5.37$ min, $\sigma=2.28$) and the time per selected movie (MMM: $M=1.25$ min, $\sigma=0.78$; FFS: $M=1.25$ min, $\sigma=1.23$) did not differ significantly between the two conditions. With respect to the number of criteria participants selected in task 2 (including values which were used multiple times), the mean number in MMM was 8.21 ($\sigma=2.91$) and in FFS 9.92 ($\sigma=3.73$) showing no significant difference. However, the average number of facet values selected when a movie was added to the shopping cart was significantly higher in MMM (MMM: $M=4.21, \sigma=2.51$; FFS: $M=2.22, \sigma=.83$; $t(24)=2.61, p<.05$). Nonetheless, there were almost no considerable differences with respect to the relative amount values from each facet were selected. But, the option to select similar movies—that was only available in MMM—was the second most used facet, and was

together with the Genre facet (which was used most often), much more frequently used in relation to other facets in FFS.

The interaction analysis based on the screencasts showed that users in MMM made extensive use of the sliders to adjust their preferences after selecting criteria, immediately explored the results, and, after possibly adding movies to the cart, started a new “iteration” with new or additional facet values. However, we did not find any effects over time, i.e., users typically selected the same number of values when adding an item to the shopping cart. They also stayed with using the same types of facets. While individual user behavior seems constant, we found differences between users. Few participants used approximately two values on average when settling for an item, the majority used about four, and a small number of participants even more. But, as participants were not required to add a specific number of items to the shopping cart, these results have to be treated with caution.

Interestingly, nonetheless, there seem to be differences using the tiles, i.e. facet values, with respect to domain knowledge. Besides being faster and adding more items to the shopping cart, users with higher domain knowledge created more than twice as many new tiles, i.e., they used the search functionality to create new tiles in order to formulate their search goal more specifically. In contrast, users with less domain knowledge seemed to prefer choosing from a broader range of facets and selected tiles that were more spread out across the widgets. In particular, they also had on average 44% more criteria activated when settling for an item. Nevertheless, despite highly positive ratings independent of domain knowledge, users with less domain knowledge found Recommendation Quality [15] significantly inferior (low: $M=3.63, \sigma=.36$; high: $M=4.24, \sigma=.40$; $t(12)=-2.81, p<.05$). Also System Effectiveness [15] was rated significantly lower by the users with low domain knowledge (low: $M=3.27, \sigma=.65$; high: $M=3.87, \sigma=.37$; $t(12)=-2.24, p<.05$). The same tendency was observed for perceived effort (low: $M=4.00, \sigma=.71$; high: $M=4.89, \sigma=.33$; $t(12)=-3.25, p<.01$). In terms of Usability (SUS [3]), Control and Interaction/Interface Adequacy [28], however, we did not find significant differences, so that the interaction concept in general seems to be perceived as highly appropriate independent of domain knowledge. This is also supported by the generally positive assessment of the

particular interaction features, e.g., sliders ($M=3.41$, $\sigma=1.23$) and visualizations how many of the resulting items fulfill a criterion ($M=3.13$, $\sigma=1.41$). We also asked participants about their understanding of these features. Regarding the sliders, we presented three predefined answers. All participants chose the correct answer out of these alternatives. Also, 88% of the participants explained the visualizations correctly using their own words (the rest also seem to have understood the visualizations, but their explanations were not clear enough to conclude that).

5.4 Discussion

The study shows that MMM users felt more in control than with the faceted filtering system. While one might expect the level of control to be higher in the manual approach, the possibility to weight criteria, the soft ranking technique, and other interactive features of MMM seem to contribute to this finding. Whereas the perceived overall quality of the results did not differ significantly, there were marked differences between varying situations of use: The filtering system seems to be useful for more targeted searches whereas the blended RS is considered more appropriate when the user has no specific goal or the direction of the search is only vaguely known. Also, users appreciated being able to specify not only content-related features but also additional recommender-related ones, e.g. stating the preference to see movies similar to the one selected.

Possibly supported by the preceding design and usability studies, the larger range of functionality in MMM did not result in significant differences in terms of usability. Both systems received high scores on the SUS and for interface adequacy. Interaction adequacy of MMM was assessed even superior, and the new interaction features appear to be comprehensible and useful. While the task time for the main task did not differ significantly between the two conditions, the longer time needed with MMM for the introductory task suggests that the learning phase of using the novel interface is quite short. Also, perceived interaction effort (which was rated highly acceptable) and the number of items put into the shopping cart did not differ significantly.

A further advantage of MMM seems to lie in the fact that users were not required to deal with Boolean filtering logic like in FFS. An indication that users expressed their preferences more extensively in MMM can be seen in the fact that significantly more

criteria were active when an item was added to the shopping cart.

While the total number of criteria set during the entire process was not different in both conditions, we can assume that users had to change or reset criteria more often in FFS due to hard filtering while more criteria were used ‘productively’ in MMM for making the final decision. Overall, there are several indications that users are more engaged to explore the options and tend to specify their preferences in more detail—provided they have the option to do so—even if not all of them can be satisfied for each recommended item. A further finding is that users with higher domain knowledge appear to specify their preferences more precisely. They tend to use less criteria, but still rate the quality of the received recommendations higher than users with lower domain knowledge who seem to need specifying more criteria before settling for an item. This indicates that additional adaptive mechanisms might be helpful. As suggested by our model of interactive recommending, tailoring the interface based on the user’s interaction behavior might further improve the recommendation process. In line with the other interaction cycles in our model, more intelligent techniques for automatically suggesting filter criteria may support users, especially with less domain knowledge, to obtain better recommendations while stating fewer preferences. However, although we found significant differences in this scale, all ratings are already in a very positive range.

6 Conclusions and Outlook

In this paper, we have presented a model of interactive recommending as well as one instance of this model, the prototype movie recommender *MyMovieMixer* that implements the concept of *blended recommending*, initially described in [19]. While the model comprises aspects not yet realized in this proof-of-concept demonstrator, it seems to be useful for exploring the larger design space of interactive recommending. In particular, blended recommending, and thus also MMM, is specifically focused on merging faceted filtering, and retaining its high level of usability and comprehensibility, with intelligent recommender techniques. The results of our evaluation indicate that allowing users to select any combination of criteria and to specify their weights leads to a high level of perceived control and recommendation quality. In addition, users rated

MMM more suitable than the filtering system when they have not yet formed a clear search goal, and tended to describe their preferences with more criteria when not being required to observe the logical implications, in particular, to avoid over-constraining the search. The usage of filter criteria in our approach with respect to specifying, refining, resetting and changing criteria will require more empirical investigation to identify typical interaction patterns which are likely to be dependent on personal characteristics and decision strategies.

The permanent availability of a ranked list of recommendations that matches the criteria currently specified best always provides a cognitive anchor, supporting reactive search behavior, and motivating refinement and critiquing of features. The recommender techniques applied include both collaborative and content-based methods in a hybrid fashion. This allows users to apply different strategies in their search, using Collaborative Filtering based on user rating data when unsure about content-based properties, and Content-based Filtering when they were already aware of preferred item features. While the approach helps in overcoming several drawbacks of conventional information filtering systems, it also does not require the prior availability of a user preference model, thus circumventing the cold start problem and accommodating users who do not wish to share their preferences due to privacy reasons. However, profile data could be easily incorporated in the approach when available. This, however, is subject of future work.

We also plan to more completely cover the different interaction cycles described in the model. For this purpose, we will investigate how filter facets and values can be made more user-adaptive in the context of recommending, suggesting criteria in a way that would reduce the number of actions needed to finally decide which item to choose. Furthermore, we aim at developing and incorporating methods for deriving preference data from the user's general interaction behavior, thus also addressing the uppermost interaction cycle shown in our model. In conclusion, we believe that the presented model opens up a design space that bears the potential of making recommender systems more user-controllable and transparent and that may in consequence lead to better and more trustworthy recommendations.

References

1. Baeza-Yates, R., and Ribeiro-Neto, B. *Modern Information Retrieval*. ACM, 1999.
2. Bostandjiev, S., O'Donovan, J., and Höllerer, T. Taste-Weights: A visual interactive hybrid recommender system. In *Proc. RecSys '12*, ACM (2012), 35–42.
3. Brooke, J. SUS – A quick and dirty usability scale. In *Usability Evaluation in Industry*. Taylor & Francis, 1996, 189–194.
4. Burke, R. Hybrid web recommender systems. In *The Adaptive Web. Methods and Strategies of Web Personalization*, P. Brusilovsky, A. Kobsa and W. Nejdl, Eds., Springer, 2007, 377–408.
5. Celik, I., Abel, F., and Siehdnel, P. Towards a framework for adaptive faceted search on twitter. In *Proc. DAH '11* (2011).
6. Chen, L., and Pu, P. Critiquing-based recommenders: Survey and emerging trends. *User Mod. and User-Adapted Interaction* 22, 1-2 (2012), 125–150.
7. Chi, E. H. Transient user profiling. In *Proc. Workshop on User Profiling* (2004), 521–523.
8. Dooms, S., de Pessemer, T., and Martens, L. Improving IMDb movie recommendations with interactive settings and filters. In *Proc. RecSys '14*, ACM (2014).
9. Gantner, Z., Rendle, S., Freudenthaler, C., and Schmidt-Thieme, L. MyMediaLite: A free recommender system library. In *Proc. RecSys '11*, ACM (2011), 305–308.
10. Girgensohn, A., Shipman, F., Chen, F., and Wilcox, L. DocuBrowse: Faceted searching, browsing, and recommendations in an enterprise context. In *Proc. IUI '10*, ACM (2010), 189–198.
11. Hearst, M. A. *Search User Interfaces*. Cambridge University Press, 2009.
12. Herrmann, K., Schering, S., Berger, R., Loepf, B., Günter, T., Hussein, T., and Ziegler, J. MyMovieMixer: Ein hybrider Recommender mit visuellem Bedienkonzept. In *Proc. Mensch & Computer '14*, De Gruyter Oldenbourg (2014), 45–54.
13. Jawaheer, G., Weller, P., and Kostkova, P. Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. *ACM Trans. Interact. Intell. Syst.* 4, 2 (2014), 8:1–8:26.
14. Karimi, R., Freudenthaler, C., Nanopoulos, A., and Schmidt-Thieme, L. Exploiting the characteristics of matrix factorization for active learning in recommender systems. In *Proc. RecSys '12*, ACM (2012), 317–320.
15. Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., and Newell, C. Explaining the user experience of recommender systems. *User Mod. and User-Adapted Interaction*, 22, 4-5 (2012), 441–504.
16. Konstan, J. A., and Riedl, J. Recommender systems: From algorithms to user experience. *User Mod. and User-Adapted Interaction* 22, 1-2 (2012), 101–123.

17. Koren, Y., Bell, R. M., and Volinsky, C. Matrix factorization techniques for recommender systems. *IEEE Computer* 42, 8 (2009), 30–37.
18. Kuhlthau, C. C. Inside the search process: Information seeking from the user's perspective. *J. Am. Soc. Inf. Sci.* 42, 5 (1991), 361–371.
19. Loepp, B., Herrmann, K. and Ziegler, J. Blended recommending: Integrating interactive information filtering and algorithmic recommender techniques. In *Proc. CHI '15*, ACM (to appear).
20. Loepp, B., Hussein, T., and Ziegler, J. Choice-based preference elicitation for collaborative filtering recommender systems. In *Proc. CHI '14*, ACM (2014), 3085–3094.
21. Mandl M. and Felfernig, A. Improving the Performance of Unit Critiquing. In *Proc. UMAP '12*, Springer (2012), 176–187.
22. Marchionini, G. *Information Seeking in Electronic Environments*. Cambridge University Press, 1995.
23. McNee, S. M., Riedl, J. and Konstan, J. A. Being accurate is not enough: How accuracy metrics have hurt recommender systems. In *Ext. Abstracts CHI '06*, ACM (2006), 1097–1101.
24. McNee, S. M., Riedl, J. and Konstan, J. A. Making recommendations better: An analytic model for human-recommender interaction. In *Ext. Abstracts CHI '06*, ACM (2006), 1103–1108.
25. Moshagen, M. and Thielsch, M. T. Facets of visual aesthetics. *Int. J. Hum.-Comput. St.* 68, 10 (2010), 689–709.
26. Pariser, E. *The Filter Bubble: What the Internet is Hiding From You*. Penguin Press, 2011.
27. Parra, D., Brusilovsky, P., and Trattner, C. See what you want to see: Visual user-driven approach for hybrid recommendation. In *Proc. IUI '14*, ACM (2014), 235–240.
28. Pu, P., Chen, L., and Hu, R. A user-centric evaluation framework for recommender systems. In *Proc. RecSys '11*, ACM (2011), 157–164.
29. Pu, P., Chen, L., and Hu, R. Evaluating recommender systems from the users perspective: Survey of the state of the art. *User Mod. and User-Adapted Interaction* 22, 4-5 (2012), 317–355.
30. Pu, P., Faltings, B., Chen, L., Zhang, J., and Viappiani, P. *Recommender Systems Handbook*. Springer, 2010, ch. Usability Guidelines for Product Recommenders Based on Example Critiquing Research, 511–545.
31. Ricci, F., Rokach, L., and Shapira, B. *Recommender Systems Handbook*. Springer, 2010, ch. Introduction to Recommender Systems Handbook, 1–35.
32. Smyth, B., and McGinty, L. An analysis of feedback strategies in conversational recommenders. In *Proc. AICS '03* (2003).
33. Sacco, G. M., and Tzitzikas, Y. *Dynamic Taxonomies and Faceted Search*. Springer, 2009.
34. Salton, G., and Buckley, C. Improving retrieval performance by relevance feedback. In *Readings in Information Retrieval*. Morgan Kaufmann, 1997, 355–364.
35. Thai, V., Rouille, P.-Y., and Handschuh, S. Visual abstraction and ordering in faceted browsing of text collections. *ACM Trans. Intell. Syst. Technol.* 3, 2 (2012), 21:1–21:24.
36. Tintarev, N., and Masthoff, J. *Recommender Systems Handbook*. Springer, 2010, ch. Designing and Evaluating Explanations for Recommender Systems, 479–510.
37. Tvarožek, M., Barla, M., Frivolt, G., Tomša, M., and Bieliková, M. Improving semantic search via integrated personalized faceted and visual graph navigation. In *Proc. SOFSEM '08*, Springer (2008), 778–789.
38. Vig, J., Sen, S., and Riedl, J. Navigating the tag genome. In *Proc. IUI '11*, ACM (2011), 93–102.
39. Voigt, M., Werstler, A., Polowinski, J., and Meißner, K. Weighted faceted browsing for characteristics-based visualization selection through end users. In *Proc. EICS '12*, ACM (2012), 151–156.
40. Xiao, B., and Benbasat, I. E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly* 31, 1 (2007), 137–209.
41. Yee, K.-P., Swearingen, K., Li, K. and Hearst, M. Faceted metadata for image search and browsing. In *Proc. CHI '03*, ACM (2003), 401–408.
42. Zadeh, L. Fuzzy sets. *Information and Control*, 8, 3 (1965), 338–353.
43. Zhang, J., Jones, N., and Pu, P. A visual interface for critiquing-based recommender systems. In *Proc. EC '08*, ACM (2008), 230–239.
44. Zhao, X., Zhang, W., and Wang, J. Interactive collaborative filtering. In *Proc. CIKM '13*, ACM (2013), 1411–1420.