Blended Recommending: Integrating Interactive Information Filtering and Algorithmic Recommender Techniques

Benedikt Loepp
University of Duisburg-Essen
Duisburg, Germany
benedikt.loepp@uni-due.de

Katja Herrmanny
University of Duisburg-Essen
Duisburg, Germany
katja.herrmanny@uni-due.de

Jürgen Ziegler
University of Duisburg-Essen
Duisburg, Germany
juergen.ziegler@uni-due.de

ABSTRACT
We present a novel approach that integrates algorithmic recommender techniques with interactive faceted filtering methods. We refer to this approach as *blended recommending*. It allows users to interact with a set of filter facets representing criteria that can serve as input for different recommendation methods including both collaborative and content-based filtering. Users can select filter criteria from these facets and weight them to express their preferences and to exert control over the hybrid recommendation process. In contrast to hard Boolean filtering, the method aggregates the weighted criteria and calculates a ranked list of recommendations that is visualized and immediately updated when users change the filter settings. Based on this approach, we implemented an interactive movie recommender, *MyMovieMixer*. In a user study, we compared the system with a conventional faceted filtering system that served as a baseline to obtain insights into user interaction behavior and to assess recommendation quality for our system. The results indicate, among other findings, a higher level of perceived user control, more detailed preference settings, and better suitability when the search goal is vague.

Author Keywords
Recommender Systems; Interactive Recommending; Information Filtering, User Interfaces

ACM Classification Keywords
H.3.3. Information Storage and Retrieval: Information Search and Retrieval – information filtering;
H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces – evaluation/methodology, graphical user interfaces (GUI), user-centered design.

INTRODUCTION
Recommender Systems (RS) are a well-established technology that aims at assisting users with finding relevant information in large information spaces [25], thus reducing or avoiding the information overload problem. The goal of RS is to select from a large set of items—such as products, documents or movies—those items that best match the user's interests and preferences, and to present them in a suitable manner [25]. For this purpose, most RS rely on user profiles and use fully automated recommendation techniques. Successful RS minimize the interaction effort the user needs to invest to find relevant items, produce well-fitting recommendations, and may also reduce cognitive load [24].

On the downside, 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 [29]. As a consequence, acceptance of the recommendations and trust in the system may be reduced [33]. Since 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. It is also possible that users do not want their preferences to be stored in the system, or their existing profile to be applied as it may differs from their current interests. All the issues mentioned may result in reduced usability, trustworthiness and user acceptance of RS [16,24,29,33].

While RS represent the upper end of the automation dimension, conventional information filtering systems [14] are positioned at the opposite end. Interactive search and filter techniques, including hierarchical navigation and faceted filtering [14,34], allow users to flexibly explore large item spaces while providing a high level of user control and transparency. However, the user's effort for searching and navigating is typically much higher compared to accepting recommended items. In addition, users may over-constrain their search in a filtering system, when strict logical query processing is applied [26]. Manual filtering also requires users to mentally form a more or less specific search goal which may be difficult in large or unknown domains. Alleviating this cognitive task is one of the strengths of RS.

Considering the respective advantages and shortcomings of fully interactive filtering and automatic recommending suggests that a closer integration of interactive approaches with recommender techniques may alleviate some of the issues described. Recommender systems research has up to now been mainly focused on optimizing the accuracy of recommender algorithms, i.e., how well the recommendations fit the user preferences. It is often overlooked, however, that the impact of improved algorithms on the overall success of a RS is decreasing [17]. Since existing algorithms are already very mature, only small and ever decreasing additional precision can be expected from further optimizing the algorithms [24]. Besides, the accuracy of recommender algorithms has typically been evaluated offline, without considering the actual user interaction with the system [16].

Only more recently, research has started to focus more on user aspects, e.g., the user's interaction behavior, user interfaces for RS, and the resulting user experience [16,24]. 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 [33]. 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 [17]. Considering these aspects in the development of RS seems an important research goal.

To overcome the limited extent of interaction, user control and transparency in conventional RS, we propose in this paper a novel approach we refer to as *blended recommending*, that combines the advantages of interactive information filtering and automated RS. Our approach expands the concept of hybrid RS, usually defined as combining different recommender algorithms to increase the accuracy [6], by additionally integrating several interactive filter techniques. However, our point of departure is faceted filtering [34] which has been shown to be an intuitive and efficient technique [18,28] for filtering and browsing large item spaces [14]. Users select concrete filter criteria from different facets which in our approach can represent hard (Boolean) as well as soft (fuzzy) constraints. In addition, facet values may serve as input for collaborative and content-based recommender techniques in a hybrid fashion.

To demonstrate the concept of blended recommending we implemented a movie recommender called *MyMovieMixer* (MMM). The system allows users to directly manipulate both setting and degree of influence of the different filters and recommender algorithms while the resulting set of recommendations is immediately updated and visualized. Based on the different filter criteria, the recommendation methods' output, and the weights set by the user, the system calculates each item's overall relevance, trying to optimally satisfy all user preferences. The system can cope with cold start situations or when conventional filtering may not yield useful results at all. By varying the settings and weights (with corresponding sliders) of a filter, users can easily explore the effect of each facet, thus increasing interactivity and transparency of the recommendation process.

In the following, we first discuss relevant related work regarding interactive RS and information filtering. Next, we describe the components of MMM and the methods used, its interaction concept and some algorithmic details. Then, we present a user study we conducted to compare MMM with a more standard filter interface. Finally, we conclude by discussing results and providing an outlook on future work.

INTERACTIVE RECOMMENDATION AND INFORMATION FILTERING APPROACHES

Popular recommenders like the ones used by Amazon [21] or Netflix [3] often produce recommendations that fit well the user's interest and can thus contribute to reducing the user's interaction effort and cognitive load [24]. Nonetheless, in many cases users are dissatisfied with the results since they may currently have a different search goal or feel too much dominated by the system. Current RS do generally not allow the user (or only in a limited fashion) to influence or control the recommendation process. More recently, the potential of interactive approaches to recommending has been discussed as a means to overcome these usage-related issues. Some systems, for instance, allow users to refine the result set by applying the Relevance Feedback principle [27]. While employing this principle in RS increases the perceived user control, basically, the already existing user interest profiles are just refined. Such long-term user preference data are central for most recommender algorithms, although they make it difficult to react to situational needs [9] and may lead to filter bubble effects [22]. Moreover, the necessary profile information is often not available, or the profile size is too small to generate precise recommendations. While several approaches exist to overcome such cold start problems algorithmically [15,37], a promising alternative can be seen in methods that capture the user's preferences interactively.

Critique-based RS [8] increase the interactivity by allowing users to criticize certain properties of the current recommendations. Users can iteratively refine the result set towards their search goal, e.g., by requesting longer movies or films by a different director. This is based on the assumption that users find it is easier to criticize recommended items with respect to certain product features than to formulate a search goal up-front. Visual support and direct manipulation of the criticized features can have a positive effect in terms of comprehensibility, user-friendliness and interaction effort [36]. Dynamically suggesting one or more features to be criticized can increase the efficiency [8]. However, as critique-based RS typically depend on predefined product dimensions, this principle is not applicable in all situations. Other approaches use, for instance, latent factors automatically determined with Matrix Factorization [19] to let users choose between item characteristics [20]. *MovieTuner* [31] is a system based on user-defined tags. The tags are automatically weighted by the system, and the most important ones are presented to the users. Users can then explicitly indicate a preference for movies with, e.g., more humor or less violence. While expressing preferences through tags or latent factors in these systems can be useful, there is no integration with other content information such as data on actors or movie genres. Users can thus not select and weight their preferences from a wider range of different feature types (e.g. predefined content information, tags and latent factors).

Only few approaches combine recommenders with interactive visualizations. *SmallWorlds* [13] is one example that embeds a graph-based interactive visualization in Facebook to simplify preference elicitation and to increase transparency. Whereas hybrid RS [6] are typically not controllable by users, *TasteWeights* [4], an interactive hybrid music recommender, is one of the few exceptions, providing possibilities to directly manipulate graphically connected widgets. By weighting the influence of different information types and social data sources, users perceived a higher recommendation quality and better understood how the results were produced. *SetFusion* [23] shows how a RS based on a common hybridization strategy [6] can be influenced by the user. Here, users can weight the influence of three different recommender algorithms individually. While several interactive features are provided (e.g. a Venn diagram visualizing the result set), user manipulations do not always have immediate impact on the results and explicitly selecting and weighting individual content-related filter criteria is not possible. The system also requires a persistent user profile. Nonetheless, user studies showed a high degree of perceived control and increased user engagement. Visualizing the results as well as explanations how they were generated also improved perceived transparency. Another example of increased interactivity in a hybrid RS is the browser plugin *MovieBrain* [10], that enhances the Internet Movie Database (IMDb)[1] with interactive settings and filters to generate movie recommendations that better match the user's situational needs. But, apart from filtering out particular genres, it also does not take further content information into account.

With information retrieval as their point of departure, there exist a broad range of information filtering techniques outside the RS field that are based on explicit user queries. Faceted filtering [34] is one of the most prominent and successful examples for iterative query refinement. It supports exploration and discovery [14,34] of large item spaces by selecting values from a set of facets, so that the item space is iteratively constrained until the desired result is found. Faceted search is also used to enhance conventional keyword search and to support more flexible navigation [14], e.g. in digital libraries or on commercial websites such as Amazon or eBay. While most simple variants use fixed filters and values from which the user selects in a stepwise manner, more dynamic approaches with complex filter settings were already introduced a long time ago. *FilmFinder* [1], for instance, established the concept of dynamic queries, supporting continuous manipulation of filters with immediate feedback in a visualization of the item space.

Early interactive filtering techniques mostly rely on predefined sets of filter attributes (facets) and typically implement only hard Boolean filtering. Also, it has been criticized that faceted search systems often only allow for conjunctive queries and consider all facets equally important for the user [26,28,30,32]. Furthermore, most approaches perform an exact matching to determine the result set. *DocuBrowse* [12] is one example that also employs fuzzy matching. The system supports faceted browsing of a large document collection. Document genres are automatically identified and when a corresponding facet value is selected, different coloring is used to express a document's relevance. The system also offers recommender functionality, but without integrating it with the filter methods. To deal with lacking metadata, the system utilizes information about the file hierarchy. Other approaches use, for instance, semantic [7] or social [30] data sources for that purpose. Most of these newer works also focus on automatic extraction and adaptive selection of facets and facet values [7,18,30]. Still, as facets and facet values are automatically inferred relying on user profiles (e.g. extracted from semantically enriched Twitter tweets [7] or social network data [30]) and observed user behavior, the user's influence on the current filter setting is limited. In addition, from a user's perspective, items may not be sufficiently described by content-related attributes [14].

The aforementioned approaches focus on ranking facets and facet values to suggest those that are likely to be most relevant for the user's search goal. *VizBoard* [32] is one of the few systems that allow users to change the influence of the selected criteria on the result set. The authors emphasize the importance of prioritizing facets to order the results appropriately, and to avoid excluding relevant items. Recent work has also concentrated on user experience of faceted search and has combined it with different visualizations. *IVEA* [28] uses a matrix visualization to display documents and their relevance according to the selected facets based on TF-IDF heuristics. The facets derived from a user-built ontology can be ordered by the users—although here, the main purpose is to compare document sets in large collections, and the item relevance is not considered for sorting the results.

Overall, the review of related work shows that blending interactive information filtering with recommender techniques seems to be a promising approach. Combining the benefits of hybrid recommending with incremental filtering may make recommenders more transparent and user-controllable while retaining the high level of usability and comprehensibility of faceted filtering [14]. Increasing the degree of user control in RS in general and improving interactivity of the

[1] http://www.imdb.com

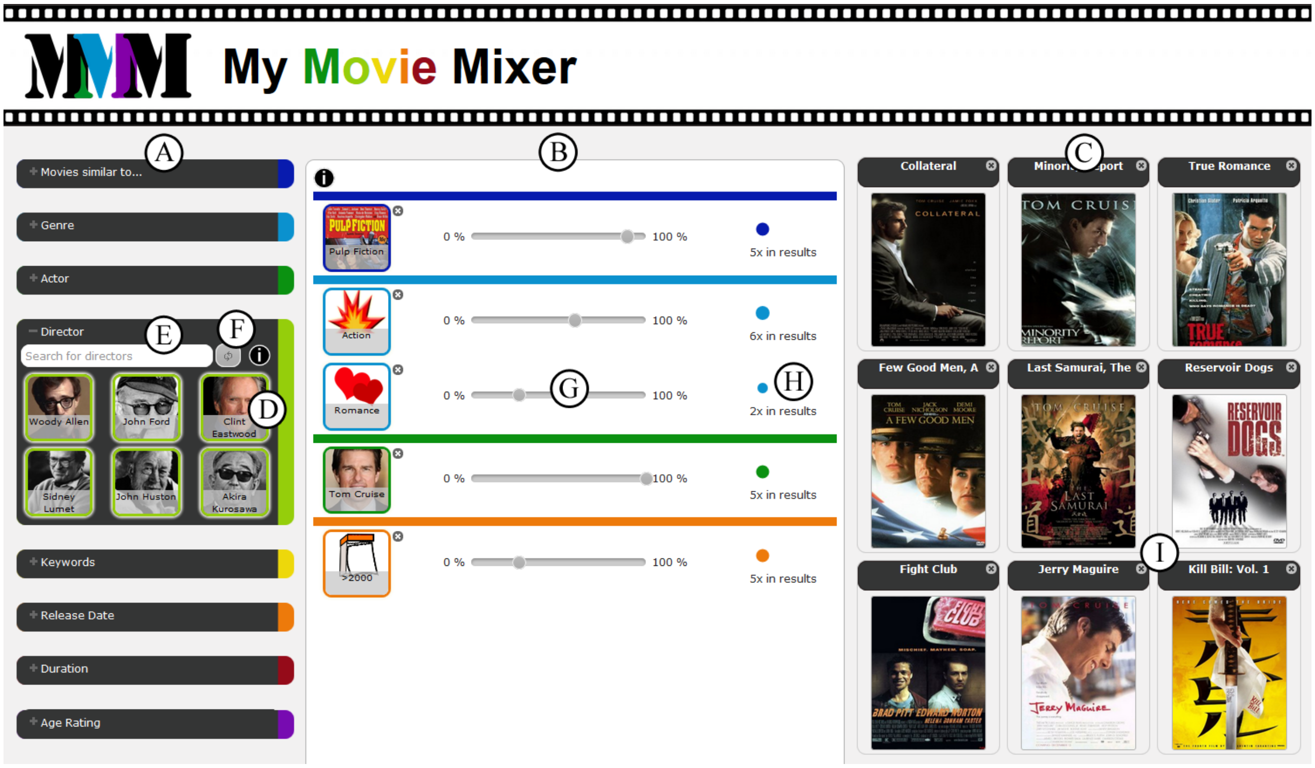

Figure 1. The MyMovieMixer application: widget area (A), work area (B), result area (C), tile representing a facet value (D), input field to search values (E), shuffle button to receive a new set of suggested tiles (F), slider to adjust a tile's weight for the recommendations (G), visualization of the number of movies fulfilling the criterion (H), button to dismiss a recommendation (I).

recommendation process have been described as important design goals [17] but are still not optimally realized in existing systems. It has also been requested that preference elicitation for RS should be more entertaining and attractive to motivate users to interact with the system [24]. Finally, fuzzy filtering with increased user influence in combination with intelligent recommender algorithms may overcome several drawbacks of conventional information filtering systems.

MYMOVIEMIXER: AN EXAMPLE APPLICATION OF BLENDED RECOMMENDING

In this section, we present *MyMovieMixer* (MMM, Figure 1), a web-based application we developed to demonstrate the concept of blended recommending. MMM integrates methods from the different approaches discussed before to recommend movies from the MovieLens dataset[2], a widely used collection of movie data with user ratings. To make the system flexible and useful in different contexts of use (e.g., for different moods, presence of different people or cold start situations), the recommendation process is entirely based on explicit user input given during a session. 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 incrementally specify their preferences by selecting filter criteria from a set of facets. For facets with large number of values, a search function is provided. Users can also weight each criterion to change its influence on the resulting recommendations. The system calculates an overall relevance score for each movie by aggregating the movie's relevance values with respect to each criterion selected. Depending on the respective facet, Boolean and fuzzy filtering, as well as conventional collaborative and content-based recommendation algorithms are employed. For example, a user may indicate (as seen in Figure 1) that he or she would like to see movies similar to Pulp Fiction that also contain elements of the genres Action and Romance, although the latter criterion might be less important. In addition, the user may, for instance, love movies with Tom Cruise and would also prefer (at least to some extent) a movie from the last decade.

MMM combines the benefits of information filtering interfaces with the intelligence of RS. It allows to directly manipulate the different filters and weights and immediately shows the effect on the resulting recommendations, thus increasing user control and making it easy to understand the effect of the different settings. In the following, we first describe 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.)

different kinds of facets, and the corresponding filtering and recommendation methods used. Next, we elaborate on the interaction concept of MMM. Finally, we explain algorithmic details regarding the calculation of movie relevance scores.

Facet Types and Corresponding Filtering Methods

MMM offers a range of different facets, labeled: *Movies similar to…, Genre*, *Actor*, *Director*, *Keywords*, *Release Date*, *Duration*, and *Age Rating*. Depending on the facet type different methods are used to determine the relevance of movies with respect to the user preferences. In particular, we use Boolean filtering (for movie genres, directors and age ratings), fuzzy filtering (for a movie's release year and duration) as well as collaborative (for similar movies) and content-based (for actors and tags) recommender techniques:

- Boolean filtering: If the user selects a criterion from a facet such as movie genre, each movie with this genre will be considered in the results while the other ones will not be taken into account. The same procedure is used for directors and the age rating chosen by the user.
- Fuzzy filtering: We use Fuzzy Logic [35] 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 would also include, although with linearly decreasing relevance, movies released some years before or after that decade. This also applies to the length of a movie, where users can choose between "short", "normal" and "overlength" (corresponding to <80, 90–120 and >130 minutes duration). Movies falling within these time spans receive the full weight, movies in between receive lower relevance.
- 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 integrated Collaborative Filtering (CF), the overall most common recommendation method [25]. To determine similar movies, we utilize the ratings given by other users in the MovieLens 10M dataset and calculate the similarity between the selected movie and all other movies by means of their latent factor vectors (which are determined by a common Matrix Factorization [19] recommender[3]) using an 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 [14,28,30].
- Content-based Filtering: For the actor and keyword facet we use conventional content-based recommender methods [25]. For instance, we calculate the relevance of a movie with respect to a certain keyword the user selects via TF-IDF heuristics [2]. Inspired by *MovieTuner* [31], 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.

While the choice of a concrete method depends on the underlying data, conceptually it is not important which method is used—any filtering method or recommender algorithm might be employed.

User Interaction Concept

The workspace of MMM consists of three main parts (Figure 1): The area on the left-hand side (A) presents the different facets from which the user can choose filter criteria. The work area (B) shows the selected criteria and the 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). We will now describe the interaction elements of these three components as well as the feedback mechanisms we integrated additionally in more detail.

Facets (A) are rendered as menu-like widgets, which in the uncollapsed mode show a number of rectangular tiles (D) representing possible criteria (facet values), visualized with images where possible. Users can drag tiles into the work area as input to the recommendation process. Since the number of values is typically large, users can add tiles by using a search box (E) with auto-completion [14]. Moreover, users can request a new set of values by pressing the shuffle button (F). The system then suggests tiles based on the frequency with which the corresponding criterion occurs in the current results. For example, when the user has dragged the genre tile Action into the work area, the result set is updated and actors most frequently occurring in action movies are shown in the actor facet. This allows the user to further filter the results with the most frequent values of other facets.

The work area (B) contains all tiles the user has dragged there as filter criteria. The weight of each criterion can be manipulated with the associated slider (G). Changing the weight of a criterion immediately updates the result set. In this way, users can interactively explore the effect of their preference settings on the recommendations. Since it may not be possible to fulfill all criteria specified, the system provides textual and graphical feedback (H) how often the criterion occurs in the current recommendations.

The ranked set of recommendations is presented on the right-hand side (C). Movies users are not interested in (e.g., because they do not like them or have already seen them), can be removed from the result set for the current session by clicking the x-button (I).

[3] *FactorWiseMatrixFactorization* from the MyMediaLite [11] recommender library.

Besides its explorative characteristic, MMM offers several means supporting the user's understanding of the recommendation generation. For example, users can open a dialog with detailed information about a recommended movie that also explains why it was recommended (i.e., which criteria were satisfied for this recommendation). In addition, recommendations related to a criterion are highlighted in color when the user moves the mouse over a tile in the work area or changes a slider's value, showing the correspondence of a criterion to the items in the result set. Considering how important it is for users to know the source of the recommendations in complex hybrid settings [23], this increases transparency and enables users at the same time to give the system feedback as in critique-based RS. Hence, they might use these hints to increase or decrease the weight of particular criteria to obtain better matching results.

Aggregating Facets for Relevance Calculation

Internally, MMM acts like a weakly coupled hybrid recommender [6]. This means, that it handles all criteria separately at first. We now describe the method used in MMM for aggregating the specified facet values, i.e., all criteria applied by the user by dragging tiles into the work area, for calculating the result set. 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 the fulfillment degree is done in different ways: As described earlier, for a criterion such as "Movie should be similar to movie x", a recommender based on Matrix Factorization [19] determines the similarity between the movies m and x by means of their latent factors. Other criteria, for instance, make use of Fuzzy Logic [35]. Assuming that a user prefers movies from a specific decade (e.g. the 1990s), a movie released in 1993 entirely fulfills this criterion and receives a 1. However, a movie from 1989 is not completely ignored—and therefore not mapped to 0, as it would be the case in Boolean filtering. In line with Fuzzy Set Theory [35] we apply a fuzzy membership function, so that a movie which only partly satisfies a criterion is (although not as strongly) considered in the results, too (and receives, e.g., a value of 0.5). In addition, we integrated a number of content-based recommending approaches [25], for example to determine how relevant a tag is with respect to a certain movie. In this case, we use TF-IDF heuristics [2] as described earlier. Boolean filtering as in conventional faceted search is applied when no other method can be meaningfully used, in particular for tiles from the genre or the director widget.

However, in some cases, especially where using Boolean filtering, a large number of items may receive a value of 1. Consequently, these items are ranked equally regarding their fulfillment degree. To avoid this, we apply an artificial ordering on these items based on the movies' average rating and the number of ratings they received. For this, we rely on the formula the Internet Movie Database (IMDb) uses for its top 250 movie charts[4]:

$$f(\mathrm{i}) = \frac{k * \bar{r} + |r_i| * \overline{r_i}}{k + |r_i|}$$

The function f basically corrects the average rating of an item towards the global mean. Therefore, it considers both the number of ratings $|r_i|$ for an item i, its average rating $\overline{r_i}$, the mean rating across all items $\bar{r}$, and a constant k we set to 100 as the result of earlier experiments.

As a consequence, the list of values for each tile can be sorted, e.g., by item similarities, fuzzy values, TD-IDF scores or at least by values returned by the formula described above. For each movie m and each criterion c_i we thus can determine the relevance value $rel_i(m, c_i) \in [0; 1]$ according to the movie's position in the sorted result list.

Subsequently, to get a relevance score rel for a movie with respect to all criteria, we aggregate the relevance values from all n tiles by means of 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 these overall relevance values and the movies with the highest resulting values are presented to the user. Table 1 illustrates the calculations with a small example, where a user searches for a movie directed 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. We also dispense ordering the movies in case of equal relevance scores rel_i.

Movie	$\mathbf{rel_1(m, c_1)}$ *(Director)*	$\mathbf{rel_2(m, c_2)}$ *(Release)*	$\mathbf{rel(m, c_1, c_2, 100, 50)}$ *(Overall relevance)*
Indiana Jones 3 *(Spielberg, 1989)*	1.0	0.5	0.833
Jurassic Park *(Spielberg, 1993)*	1.0	1.0	1.000
Pulp Fiction *(Tarantino, 1994)*	0.0	1.0	0.333

Table 1. Relevance calculation for some example movies.

By applying the ranking technique described, we avoid the conjunctive application of filter criteria as it is used in most information filtering approaches [26,28], and allow users to fluently explore the results of different facet combinations. 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 by means of their latent factor values.

[4] http://www.imdb.com/chart/top

EVALUATION

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. We dismissed the idea of including a conventional RS in this comparison because we were particularly interested in the user interaction, whereas RS are usually based on existing user profiles and lack interactive features for expressing user preferences. Due to its high level of interactivity and controllability, a filter interface therefore appears to be a more natural competitor.

Thus, we developed a faceted filtering system (FFS, Figure 2) as an alternative condition that uses the same facets (except *Movies similar to…*), values and dataset as MMM. However, facets and values were fixed, users were not allowed to weight the different filters, and the filtering used only Boolean AND operations. Thus, in this condition, it was possible to over-constrain the search leading to empty result sets. While this general issue in faceted search can be circumvented by various means [14,26] we were interested in obtaining deeper insights into user interaction behavior. However, just as in MMM, the results are updated immediately. The system presents on the left-hand side the facets for filtering the result set, shown on the right-hand side. Additionally, the results can be sorted with respect to several criteria (e.g. movie title, release year, average rating). To further increase the fairness of the comparison, we adopted MMM's interface design, and implemented all features as similarly as possible. For the purpose of the study, we extended both interfaces with functions to add or remove items to/from a virtual shopping cart.

We hypothesized that users interacting with MMM would have a stronger feeling of control while the quality of the resulting item set would be at least as good as for the filter interface. On the other hand, the richer functionality in MMM might have led to lower usability. We therefore also evaluated usability for both systems. Moreover, we expected an influence of different situations and contexts of use. In particular, we hypothesized that MMM would be preferred in situations when users do not yet have a clear search goal or only a vague idea of what they want, which is often the case in large domains and, especially, for experience products such as movies. In contrast, we assumed that the filter interface would be preferred when the search target is known.

Method

We recruited 33 participants (20 male, 13 female, average age of 27, $\sigma = 6.46$) for the user study. The study was conducted over two weeks, and was designed as an experiment under controlled conditions. Participants used a desktop PC with a 24" LCD-display (1920×1200 px resolution) and a common web browser. The two different conditions (MMM and FFS) were tested in a between-subject design. We decided against the within-subjects option because searching

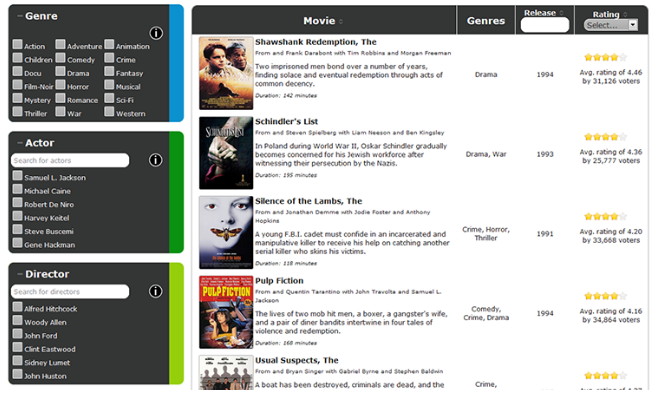

Figure 2. Screenshot of the conventional faceted filtering system we implemented to compare MMM against it.

movies the participants liked with the first system might have too much influenced their behavior with the second. In particular, participants might already develop a more specific search goal in the first trial, lowering the validity of the study for the intended usage scenarios of MMM. Thus, we randomly assigned each participant to one of the 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 [16,24] for evaluating interaction and recommendation quality, using a positive 5-point Likert scale (1–5). Furthermore, we used SUS [5] to assess the systems' usability. Finally, we collected demographic data and asked about the participants' interests in movies, their familiarity with this domain, and their knowledge regarding movie portals and web product search.

Results

From the questionnaire data, we could identify significant differences between the two conditions: Participants reported a significantly higher feeling of control (*Control* construct from [24]) for MMM based on the impression that they were able to more flexibly adjust their preferences (MMM: $M = 4.43$, $\sigma = .50$; FFS: $M = 3.85$, $\sigma = .99$; $t(22) = 2.10$, $p < .05$). Also, the interaction (*Interaction Adequacy* construct from

[24]) was seen as significantly more adequate for generating recommendations in MMM (MMM: $M=3.94$, $\sigma=.53$; FFS: $M=3.13$, $\sigma=1.00$; $t(22)=2.90$, $p<.01$).

With regard to the constructs *Perceived Recommendation Quality*, *Perceived System Effectiveness* and *Perceived Recommendation Variety* from [16] the results for both conditions did not differ significantly (MMM: $M=3.73$, $\sigma=.46$; FFS: $M=3.74$, $\sigma=.38$). Also, the usability assessment with SUS [5] did not reveal any significant differences between the two systems (MMM: $M=82.35$, $\sigma=14.80$; FFS: $M=83.59$, $\sigma=12.35$).

We also asked participants to rate the suitability of the systems for different situations of use on a positive Likert scale (1–5). The results show that they found both systems suitable when users have an approximate search goal idea in mind (MMM: $M=4.24$, $\sigma=.66$; FFS: $M=4.31$, $\sigma=.70$). In situations where users are looking for a specific movie, they found the filter system significantly more appropriate (MMM: $M=2.47$, $\sigma=1.46$; FFS: $M=3.50$, $\sigma=1.27$; $t(31)=-2.16$, $p<.05$). In contrast, for situations with no clear search direction, participants rated MMM as more suitable (MMM: $M=4.13$, $\sigma=1.09$; FFS: $M=2.80$, $\sigma=1.27$; $t(29)=3.13$, $p<.01$).

For both systems, 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. The perceived interaction effort was rated highly acceptable on a positive Likert scale (1–5) for both conditions (MMM: $M=4.47$, $\sigma=.73$; FFS: $M=4.25$, $\sigma=.68$) without significant differences.

From the screencast, we extracted the number of criteria participants selected in task 2, including values which were used multiple times. In MMM the mean number was 8.21 ($\sigma=2.91$) and in FFS 9.92 ($\sigma=3.73$), with no significant difference between the conditions. However, the mean number of facet values selected when a movie was added to the shopping cart was significantly higher for the MMM condition (MMM: $M=4.21$, $\sigma=2.51$; FFS: $M=2.22$, $\sigma=.83$; $t(24)=2.61$, $p<.05$). The interaction analysis based on the screencasts also showed that in FFS participants often selected combinations of values that lead to empty result sets due to the hard filtering used. Then, users typically deselected and reselected some of the criteria in order to explore the results with different combinations of criteria. One participant even explicitly expressed the need for a Boolean OR operation. Analogously, also the participants in the MMM condition selected several criteria. But, as MMM always calculates the best matching rank order of items, users could not over-constrain their search as in FFS. Instead, they were able to immediately explore the results and adjust their preferences with the sliders. Typically, after exploring the result set and possibly adding movies to the cart, users started a new “iteration” with new or additional facet values.

Regarding further questionnaire data, in particular with respect to confounding factors such as the users’ domain knowledge or prior experience with RS and filter techniques, we did not find any significant effects on our findings.

Discussion

A main finding of the study is that users feel more in control with MMM than with a standard faceted filtering system. While one might expect the level of control to be higher in the fully manual approach, the possibility to weight criteria, the soft ranking technique always leading to results ranked by how well they match the user preferences, and other interactive features of MMM seem to be the main contributors to this finding. Whereas the perceived overall quality of the results did not differ significantly, there were marked differences for different situations: 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.

Despite the larger range of functionality in MMM, there were no significant differences in the usability ratings. Both interfaces received a high SUS score, indicating that the new concept is as easy to comprehend and interact with as the more well-known filtering approach. Also, task time did not differ significantly in the main task, following the short introductory task. Since also the number of items put into the shopping cart did not differ significantly this indicates that users spent a similar amount of time collecting the desired items (there was also no significant difference in the time to select the first item). Similarly, the perceived effort did not differ and was rated highly acceptable for both systems. This lack of differences on the efficiency dimensions will require some further, more in-depth analysis of the interaction behavior. An initial (weak) assumption of ours was that users would spend more time in the MMM condition since it has more functionality and might engage users more to explore the options before finally selecting a movie. Despite the lack of time differences this may still be true however. A qualitative analysis using screencasts of the user interaction with the filter system showed that users often over-constrained their search, forcing them to backtrack and change their filter settings. The time needed to correct the settings may have compensated the time gained due to the simpler filter technique. Forcing backtracking could of course be eliminated, e.g., by integrating query previews [14] or using dynamic taxonomies [26]. However, due to the underlying logic of filtering, even more advanced features would probably not have avoided that users had to try more combinations of filter settings to obtain appropriate results. In contrast, MMM always provides users with a ranked list of recommendations that match the criteria currently specified best, while they do not have to observe the logical implications of their query. However, this issue needs further investigation. Furthermore, while comparing MMM against a conventional filtering system served as a useful baseline to get insights into the user interaction behavior and the general quality of the results, it should be complemented with further comparisons against more advanced systems, both RS and filter interfaces.

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. This is also supported by the fact that the total number of criteria set in the whole process was not different in both conditions. As this variable also includes changing and resetting criteria due to empty result sets which happened only in FFS, users of MMM may actually have applied more criteria to obtain the final result. These observations, together with the perceived high recommendation quality, indicate that users 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.

CONCLUSIONS AND OUTLOOK

In this paper, we have presented *blended recommending*, an approach that integrates conventional recommendation techniques with interactive faceted filtering methods. In blended recommending, the user is provided with a set of filter facets, the values of which can influence different recommendation methods including both collaborative and content-based filtering techniques in a hybrid fashion. Fixed filter facets as well as fuzzy filters are also used. User rating data may be used for Collaborative Filtering, supporting users even in situations where items are from a user's perspective not sufficiently described with content-related attributes. However, no stored preference profile for the current user is required, thereby circumventing the cold start problem. In principle, considering long-term interests is possible in our approach and will be subject of future work. In contrast to hard Boolean filtering, the user can select any combination of criteria and specify their degree of influence from which the system calculates a ranked recommendation list. The results of filter updates are shown immediately and can in turn be used to further refine the result set.

Besides the integration of (hybrid) recommender techniques with interactive filtering, a major goal of this work has been to provide the user with a higher level of influence and interactive control over how recommendations are generated. Active user involvement in the recommendation process has been identified by prior research as one of the major aspects in terms of user satisfaction. We achieve this by allowing users to interactively explore the results of different recommenders and filter settings by incrementally specifying preferences and manipulating their weights as well as by providing immediate visual feedback.

To demonstrate the approach, we implemented the movie recommendation system *MyMovieMixer*. We compared the system with a conventional information filtering interface in an empirical study in order to evaluate the interaction and recommendation concept of MMM. As MMM seems close to typical faceted filtering from a user perspective, such a comparison appears most natural as a baseline for a first evaluation, especially when focusing on HCI concerns. Nevertheless, we plan to also compare the system with other actual RS, in particular with interactive approaches. The results of the study show that allowing users to specify their preferences as input to different recommenders and filter techniques in combination with varying the influence of each criterion indeed leads to a high level of perceived user control. As a further result, MMM can be considered suitable for situations when the user already has a vague idea of the desired items as well as when no search goal has been formed yet. A conventional filter interface seems not very helpful in the latter case. However, as expected, the benefit of our approach is limited when the user already has a specific item in mind. In conclusion, the blended recommending approach allows users to interactively control the filtering of large item spaces while benefiting from the power of state-of-the-art recommender techniques. The approach gives users a strong feeling of control and is both effective in terms of recommendation quality as well as usability.

In future work, we plan to integrate more sophisticated visualizations, e.g., showing the result set in its entirety. In addition, we aim to investigate more intelligent facet selection techniques and different algorithmic implementations of the recommendation calculation. In this respect, it is worth mentioning that until now, our evaluation was focused on the user's perception of the interaction concept and the quality of the results, and thus should be complemented with more objective measurements. Moreover, we plan to compare MMM against other interactive recommending approaches and to implement the concept for different domains, in particular not only for experience products such as movies, but also for commercial search products.

REFERENCES

1. Ahlberg, C., and Shneiderman, B. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proc. CHI '94*, ACM (1994), 313–317.
2. Baeza-Yates, R., and Ribeiro-Neto, B. *Modern Information Retrieval*. ACM, 1999.
3. Bennett, J., and Lanning, S. The netflix prize. In *Proc. KDD Cup and Workshop '07*, ACM (2007), 3–6.
4. Bostandjiev, S., O'Donovan, J., and Höllerer, T. Taste-Weights: A visual interactive hybrid recommender system. In *Proc. RecSys '12*, ACM (2012), 35–42.
5. Brooke, J. SUS – A quick and dirty usability scale. In *Usability Evaluation in Industry*. Taylor & Francis, 1996, 189–194.
6. 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.
7. Celik, I., Abel, F., and Siehndel, P. Towards a framework for adaptive faceted search on twitter. In *Proc. DAH '11* (2011).

8. Chen, L., and Pu, P. Critiquing-based recommenders: Survey and emerging trends. *User Mod. and User-Adapted Interaction 22*, 1-2 (2012), 125–150.
9. Chi, E. H. Transient user profiling. In *Proc. Workshop on User Profiling* (2004), 521–523.
10. Dooms, S., de Pessemier, T., and Martens, L. Improving IMDb movie recommendations with interactive settings and filters. In *Proc. RecSys '14*, ACM (2014).
11. Gantner, Z., Rendle, S., Freudenthaler, C., and Schmidt-Thieme, L. MyMediaLite: A free recommender system library. In *Proc. RecSys '11*, ACM (2011), 305–308.
12. 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.
13. Gretarsson, B., O'Donovan, J., Bostandjiev, S., Hall, C., and Höllerer, T. SmallWorlds: Visualizing social recommendations. *Computer Graphics Forum 29*, 3 (2010), 833–842.
14. Hearst, M. A. *Search User Interfaces*. Cambridge University Press, 2009.
15. 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.
16. 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.
17. 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.
18. Koren, J., Zhang, Y., and Liu, X. Personalized interactive faceted search. In *Proc. WWW '08*, ACM (2008), 477–485.
19. Koren, Y., Bell, R. M., and Volinsky, C. Matrix factorization techniques for recommender systems. *IEEE Computer 42*, 8 (2009), 30–37.
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. Linden, G., Smith, B., and York, J. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing 7*, 1 (2003), 76–80.
22. Pariser, E. *The Filter Bubble: What the Internet is Hiding From You*. Penguin Press, 2011.
23. 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.
24. 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.
25. Ricci, F., Rokach, L., and Shapira, B. *Recommender Systems Handbook*. Springer, 2010, ch. Introduction to Recommender Systems Handbook, 1–35.
26. Sacco, G. M., and Tzitzikas, Y. *Dynamic Taxonomies and Faceted Search*. Springer, 2009.
27. Salton, G., and Buckley, C. Improving retrieval performance by relevance feedback. In *Readings in Information Retrieval*. Morgan Kaufmann, 1997, 355–364.
28. 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.
29. Tintarev, N., and Masthoff, J. *Recommender Systems Handbook*. Springer, 2010, ch. Designing and Evaluating Explanations for Recommender Systems, 479–510.
30. 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.
31. Vig, J., Sen, S., and Riedl, J. Navigating the tag genome. In *Proc. IUI '11*, ACM (2011), 93–102.
32. 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.
33. Xiao, B., and Benbasat, I. E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly 31*, 1 (2007), 137–209.
34. 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.
35. Zadeh, L. Fuzzy sets. *Information and Control, 8*, 3 (1965), 338–353.
36. Zhang, J., Jones, N., and Pu, P. A visual interface for critiquing-based recommender systems. In *Proc. EC '08*, ACM (2008), 230–239.
37. Zhao, X., Zhang, W., and Wang, J. Interactive collaborative filtering. In *Proc. CIKM '13*, ACM (2013), 1411–1420.