Explanations of Recommendations

Considering Human Factors in Recommender Systems

Muhammet Tugberk Isyapar
Computer Engineering
Middle East Technical University
Ankara, Turkey
tugberk.isyapar@metu.edu.tr

***Abstract*— Traditional evaluation metrics based on statistical formulas employed to assess the performance of a recommender system are now considered to be inadequate when utilized solely. New metrics considering the quality of user-system interaction alongside with traditional ones have been proposed in evaluation process in order to arrive at more adequate results. Generating explanations for recommendations is a research topic that has emerged as a way to evaluate the system with respect to various criteria considering users' opinions and feelings. This paper presents the state-of-the-art with respect to the explanation of a recommendation.**

Keywords*— Recommender systems; recommendation algorithms; design and evaluation of recommender systems; explanations; human factors.

I. Introduction

The objective of recommendation technology is to help the end-user making sense of large and growing amounts of data. Recommender systems have been developed for various problem domains including automatic movie recommendation. People keen on cinema would like to discover yet-unseen movies that suit their tastes and avoid the ones that they would probably regret watching. Before deciding to watch a movie, they may also like some kind of prediction about the item, since there are usually many alternatives to choose and dedicating a considerably long time and other resources to a bad movie is likely to be annoying. Movie recommender systems for the domain of cinema including MovieLens [7] and Netflix [8] have been devised to provide the demanded facilities. Other problem domains for which recommender systems have been developed include online shopping, news filtering, academic paper discovery, and social networking.

The recommendation problem, as commonly formulated, is the problem of estimating ratings for yet-unseen items by a user. The estimation is called *prediction* and is based on the ratings given by the user to seen items. As soon as predictions for unseen items are generated, the system can recommend to the user several items with the highest ratings, which is commonly named in the literature as *top-N recommendations* [1][3]. The user of the recommender system is referred to the *active user* in this paper.

There are various methods for estimating ratings for the yet-unrated items and recommender systems are classified according to the approach chosen for prediction generation. Current research poses three categories [1]:

- *Collaborative recommendations:* The active user is recommended items similar to the ones that are liked by other users with similar tastes without considering item contents.
- *Content-based recommendations:* The active user is recommended items that are similar to the ones he/she liked in the past depending on the contents of the items.
- *Hybrid approaches:* These imply developing methods that combine benefits yet avoid disadvantages of collaborative and content-based methods.

Recommender systems technology proposes several statistical metrics to measure the coverage, accuracy and precisions of generated recommendations. It is currently thought that these metrics can only partially evaluate the systems [2][3][4]. User satisfaction, serendipity, diversity and trust are now considered among important evaluation criteria [3] since recommender systems are deployed with well-designed user interfaces and the quality of recommendations tends to increase by providing better user-system interaction through interfaces. Generating explanations for the recommendations made via the interface to the user has emerged as an idea to compensate for the new evaluation criteria. Explanations have several aims determined by the characteristics of the problem domain and these aims could be utilized in evaluating explanations. Throughout the rest of the paper, definitions, aims, evaluation, types, design and usage of explanations will be described at sufficient detail to show the current trends in the technology and the effects of explaining recommendations.

II. Aims and Evaluations of Explanations

Explanation facilities propose several aims to be attained [3]. Good explanations tend to increase user satisfaction, give users trust about the system and inspire loyalty, persuade them to buy or use the recommended item or correctly guess the user's possible rating about the item, and make it easier and quicker for users to find what they want. Distinct aims may hold only in particular domains since

 ISBN: 978-1-61208-247-9

coexistence of particular aims cannot be true by definition. Seven possible aims of employing explanations, evaluation of explanations with respect to these aims, ways of presenting recommendations with explanations, and facilities providing user-system interaction will be described in the rest of this section of the paper.

A. Aims of Explanations

Transparency is the first aim of explanations. Explanations making a system transparent help the user understand how a system works, i.e., why a particular recommendation is generated instead of others. The user receiving the explanation can then understand the mechanisms of the system and act accordingly.

The second aim of explanations is *flexibility*. A flexible system can correct itself whenever the user spots an incorrect assumption made about them. Transparency and clarity could be regarded as two aims to be realized in a cyclic fashion. Explanations should provide insights into the system as a first step and should next allow the user to correct reasoning, that is, explanations should make the system flexible.

The third aim of explanations is inspiring *trust*. Good explanations tend to increase users' confidence in the system and users are loyal to systems that they regard as trustworthy. Trust as an aim has bounds with transparency. Whenever the system is unsure about the recommendation it generates, it should state that with appropriate explanations. Users' trust for the system increases whenever they are provided information about the quality of recommendations they receive. Moreover, as most commercial recommender systems come with user interfaces, the design of the interface is an important factor that affects users' trust.

Persuasiveness is the fourth aim of explanations. Appropriate explanations of recommendations could contribute to the system's persuasiveness by convincing user to try or buy the recommended item. If the quality of generated explanations is adequate, users may be affected by the rating predicted by the system and may believe that they would give the same rating even if not provided recommendations. However, a balance should be taken into account as too persuasive explanations carry the risk of convincing the user to buy a bad item which may end up in a decrease of user's trust and loyalty to the system.

Effectiveness is the fifth aim of explanations. An explanation may help the user to make better decisions in the sense that the predicted rating of the item will actually reflect the user's own preferences and tastes. To put it in other words, a recommender system with effective explanations assists a user to reach the items that they will like in the end. Therefore effectiveness has bounds with the accuracy of the recommendation algorithm and could be thought as a user-based extension to the statistical accuracy utilized in the literature.

The sixth aim of explanations is *efficiency*. Explanations may make it quicker for users to decide which recommended item fits their needs best. As recommendation process itself is finding the most valuable information out of huge amounts of data, it is also important that the user could reach what they are looking for in the least possible amount of time. Explanations should provide interaction with user in order to reduce the search time and make the system efficient.

The seventh and the last aim of explanations is *satisfaction*. Satisfaction is a broad and abstract category yet in the context of recommender systems it could be considered as making the use of system fun. Explanations could serve as means that increase users' satisfaction with the system. The quality of explanations is crucial since poor explanations are likely to decrease the user's interest or acceptance of the system. Moreover, explanations are deployed as an integral part of the user interface of the recommender system. It is known [3][4] that users tend to like more features included in the interface; therefore, including an explanation facility in commercial recommender systems is an important contribution to users' overall satisfaction with the system.

B. Evaluation of Explanations

In this section, we explore the criteria used to evaluate a recommendation. The aims of explanations could be utilized as criteria to evaluate how good an explanation is. More criteria based on combinations of the seven aims could be generated. The appropriate evaluation technique regarding each criterion will be described below.

To evaluate an explanation with respect to the *transparency* criterion, one could ask users if they believe the recommendations they have acquired are based on similar tastes with other users or items to discover if the users could understand the insides of the recommendation process. An implicit way of evaluating how transparent an evaluation is could be conducting tests based on particular tasks involving users. An example could be affecting the system in a particular way to see if the behavior changes as expected. To be concrete, one can set up a task in which the user affects the system by giving ratings only to items with a particular characteristic to see whether the recommended items will also have the same characteristic or not.

The second way of evaluating explanations is checking how good they are with respect to the *flexibility* criterion. To measure the performance of an explanation according to this criterion, one has to make use of task-based scenarios in which users give feedback via the user interface to the system stating, for example, that they no longer want to get recommendations about items with particular characteristics. The time for the system to complete such a task could be utilized as a quantitative measure unless the user interface is problematic when providing feedback.

Explanations could be evaluated according to the *trust* criterion. To measure how explanations affect the trust of users' one can use questionnaires with users. Yet, such explicit tools could be misleading and it could be a better idea to also keep track of variables related with the trust including users' loyalty (for how long and at which frequencies the user has been using the system, etc.) and sales profile if the recommender system was deployed for commercial purposes.

As the fourth criterion, explanations could be evaluated with respect to *persuasiveness*. Persuasion by evaluations could be measured as the difference in likelihood of selecting a recommended item before and after the recommendation has been delivered to the user. If the user rates the item more highly following the recommendation, one can deduce that the user is convinced by the system with the help of explanations. Another way to measure the degree of persuasiveness due to explanations could be observing if the user tends to buy or try more recommended items by using a recommender system with explanations than a system without explanations facility. The overall persuasiveness of the system could also be measured implicitly by analyzing sales profile to see if there is a significant increase.

In order to evaluate explanations according to the *effectiveness* criterion, one could measure how much a recommended items is liked by the user before and after receiving the system's prediction about the item. If the degree to which the user likes the item does not change significantly, then one can conclude that the system has been effective by providing the user the accurate recommendation. In order to analyze the role of explanations in effectiveness, one could set up tests with two recommender systems, one with and the other without an explanations facility and see how the user's degree of liking changes in both cases.

As the sixth criterion, one can evaluate explanations with respect to *efficiency*. The time to spend until the desired recommendations is delivered to the user through the interface, namely the completion time, could be used as a quantitative measure. Indirect measures including the number of inspected explanations could also be devised.

The last criterion with respect to which explanations could be evaluated is *satisfaction*. To measure users' satisfaction, one can form questionnaires to investigate if the users tend to like the system with or without explanations. One can also measure satisfaction indirectly by keeping track of users' loyalty. A qualitative way to measure users' satisfaction with the recommendation process could be observing characteristics of the users' experiences with the system until they eventually locate the desired item(s) in the interface.

Choosing the criteria exhibits certain tradeoffs. As one can observe from the definitions of the criteria, some contradict with each other like persuasiveness and effectiveness. Another example could be that the systems providing high degrees of transparency may lack having much efficiency. In design of explanations, the goal of the system should be taken into consideration together with certain properties of the problem domain. As an example, persuasiveness (balanced by trust) could be a more important criterion in online shopping than in a movie recommender system since for the latter effectiveness (together with user satisfaction) may be regarded as crucial to reach the goal of introducing the user with items both unseen and also similar to their tastes.

C. Presenting Recommendations and Explanations

The way the recommendations are presented affect the explanations and some particular recommendation representations combine the recommendations and the explanations altogether.

1) *Top item:* The best item is presented to the user with an explanation. For example, a user who is keen on sports and swimming, in particular, could be recommended recent news about the results of a swimming contest together with an explanation like *"You have been following a lot of sports news, and swimming in particular. This is the most popular and recent item from the championship."*

2) *Top-N items:* The top-N items with highest predicted ratings are presented to the user. Suppose that the user in the previous example is also interested in politics but not at as much as sports. Therefore, the system might present sports news together with a couple of recent political analysis to the user with an explanation like *"You have watched a lot of sports and politics news. You might like to see the results of the local swimming contests and the featured article of the day about the intervention in Libya."*

3) *Similar to top item(s):* The system might list similar items to the already listed ones with an explanation. This approach is generally adopted in online shopping. Customers who bought a number of items could be recommended to buy other items by presenting an explanation such as *"People who bought these also bought …"* or *"You might also like to buy … which is similar to the ones you have already bought."*.

4) *Predicted ratings for all items:* Instead of presenting the user a limited number of items as recommendations, the system may allow them to see the predicted ratings for all items, i.e. the items with low predicted ratings as well with explanations. This way the user may also receive explanations about why an item is predicted to have a low rating. If the user of the previous examples does not like football, they could receive an explanation like *"This is a sports item, but it is about football. You do not seem to like football!"* about a football story.

5) *Structured overview:* In order to allow displaying trade-offs between recommended items, the best item suiting user's needs and/or characteristics could be listed at the top and below it other alternatives having particular trade-offs could be listed certainly with explanations. This representation combines recommendations and explanations integrally. A user of an online shopping system could be recommended a camera that best fits their needs, and the rest of the cameras could be listed as *"[this camera]... is cheaper but has less resolution and poorer zooming capacity."* by explicitly stating the trade-off. Structured overview presents users several items of a particular category and increases efficiency by easing navigation and user comprehension of available options.

The recommender system may present the user with recommendations they might already know about to inspire trust, or may supply them more serendipitous recommendations to increase user satisfaction. Recommender systems could be bold in the sense that they know that the user will like to item to a certain degree, or they could state that they are sure about the recommendation they have made. These factors are part of the recommendation process and should be taken into account while presenting explanations as well.

D. Interacting with The System

There are various ways in which a user can give feedback to the system to take part in the recommendation flow.

- *The user specifies their requirements directly:* By implicitly participating in data collection process the user could tell the system what they demand. Another way of doing that could be providing the required facilities in the user interface of the system so that the user could succeed in interacting with the system. Such a facility could even allow natural language processing and the user could specify their requirements through conversations with the system.
- *The user asks for an alteration:* Like in the structured overview presentation, the user might demand the system recommends them another item having more or less a particular characteristic than the already recommended item. For example the user might desire to be recommended a similar laptop to the one they are already presented but one that is cheaper and that could have less processing power via an online shopping system interface.
- *The user rates items:* This is the most typical way of giving feedback to the system. Most movie recommender systems require that users have to rate a certain number of movies in order to start receiving recommendations about unseen items.
- *The user states their opinion:* If the user likes an item they are trying, they could ask through the interface to be recommended more items of this type. The interface could provide additional options including that the user could specify if they would like receiving more recommendations about similar items currently or later. In the same way, the user could specify that they do not want items like the recommended one. They could ask for a diversification or could state total rejection for the particular type. Finally the user could demand to be recommended a serendipitous item as well.

The way the recommendations and explanations are presented could further be extended with facilities providing user-system interaction as indicated above. Following the general framework outlined up to this point, we will continue with mechanisms that generate explanations in the next section.

III. TYPES OF EXPLANATION

Recommendations provide user the items they might like or predictions about items that the user queries about. Explanations of recommendations deliver the user the adequate information why they might like the recommended items or why they are given a particular prediction about an item. As shown in the previous section, there are various ways in which the user could receive explanations.

There are various types of explanations with relations to the mechanisms that generate recommendations and explanations [5][6]. The direct relations between users and items are unknown. In generating explanations particular *intermediary entities* are utilized to understand the relations between the active user and the item of interest. This technique is illustrated in Figure 1.

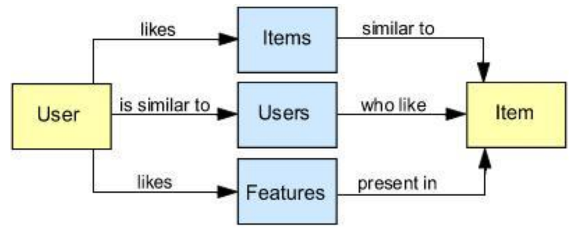

Figure 1. The user is related to the item via intermediary entities.

Explanations are categorized according to the intermediary entities they utilize:

- *Item-based explanations:* Other items rated by the user which are similar to the item for which the prediction is generated are utilized as intermediary entities to form explanations, as depicted on the first line in Figure 1. The active user receives explanations like *"(...) because you rated similarly [items used as intermediary entities]"*. This approach is adapted by particular movie recommender systems such as Netflix.
- *User-based explanations:* Other users similar to the active user who rated the item for which the prediction generated are utilized as intermediary entities to form explanations like *"(...) because [users used as intermediary entities] like you rated the item similarly"* as depicted on the middle line in Figure 1. This kind of explanations is adapted in certain scientific researches.
- *Feature-based explanations:* Particular features of the recommended item that the active user likes are utilized as intermediary entities to form explanations like "(…) because you like [features used as intermediary entities] present in the item.", as depicted on the last line in Figure 1. This approach could be utilized in the movie recommendation domain where users may receive why they have been delivered particular recommendations on the basis of their degree of liking the director, the genre, the actors and other characteristics of recommended movies.

There are several benefits and shortcomings of each approach. Item-based explanations improve users' satisfaction with the recommendation and help users to make more accurate decisions, yet users receiving this kind of explanations may not understand the relations between the recommended item and the explaining items. User-based explanations contribute to persuasiveness, yet they are less effective in helping users make accurate decisions. Feature-based explanations pose several challenges due to limited content analysis whenever multimedia items are recommended, and the results of the content analysis could occasionally be regarded as too low level since there are utilized particular techniques to extract the features by keeping track of frequencies of keywords contained in the item.

 ISBN: 978-1-61208-247-9

IV. EXPLAINING RECOMMENDATIONS USING TAGS

Vig et al. [5] propose utilizing tags formed by users to generate explanations. Their approach called *tagsplanation* makes use of a tag or a set of tags as intermediary entities. This type of explanations is developed for MovieLens, the online movie recommender system which has been in use since the second half of 1990's. The system includes millions of ratings for thousands of movies by thousands of users. MovieLens has been one of the pioneer systems developed through the recommender systems research.

Tag usage is currently popular online. Most web systems allow their users to tag items they present. However, tags pose certain challenges:

- *Tag relevance:* The relationship of the tag with the item is a key component of tagsplanations. Tag relevance specifies the degree to which the tag can represent the item.
- *Tag preference:* The relationship of the tag with the user is the other key component of tagsplanations. Tag preference implies how much the user likes the category marked by the tag considering the item which has the tag.

Tag-based explanations are inspired by certain benefits of the existing approaches and the adapted approach tries to abstain from the shortcomings described previously. Tag-based approach adapts the rating scale utilized in the item-based approach to be applied on tags by users yet through making use of user tags about items it avoids confusions of users about the explanations generated. Tagsplanations try to address a solution for the lack of effectiveness in user-based explanations. Tag-based explanation generation is similar to feature-based approach by utilizing tags with ratings and frequencies as features. However, it is different from the latter in the sense that it also deals with eliminating low quality or redundant tags and user tags can be said to have better quality than the keywords obtained through limited content analysis of items. Tags offer the possibility to generate explanations having more cognitive values since they reflect users' understandings of movies by their nature.

The aim of tagsplanations is providing *justifications* rather than *descriptions*. Descriptions reveal the actual mechanism that generates recommendations and are means of ensuring high degrees of transparency, yet they may be irrelevant, confusing, or too complex for the purposes of users. On the other hand justifications convey a conceptual model that may differ from the insides of the algorithm. Although adapting that way one has to keep considerations over transparency low, choosing justifications versus descriptions provides a degree of freedom in designing the mechanisms that generate explanations than the recommendation algorithm. This is especially useful as designing explanations can be performed as a module and integrating the module to the rest of the recommender system can be managed without increasing complexities of the recommendation algorithms. Moreover explanations gain more importance as they are more meaningful than crude descriptions of algorithmic mechanisms when they are generated to provide justifications about the generated recommendations.

Most popular tags are presented with recommendations on MovieLens website, as depicted in Figure 2.

Figure 2. MovieLens lists recommended movies with popular tags.

Users could vote for or against the adequacies of the tags presented under each recommended movie. Users' votes determine interactively the *tag popularity*. They could insert new tags for movies through the interface depicted in Figure 2 as well.

Tag preference could be measured by directly asking users their opinions about the tags presented in particular movies. Yet practical rejections may be raised against adapting this kind of approach because even though adequate features were provided by the interface, users do not have to use them and even when they use them they may not rate a substantial numbers of tags.

Tag preference could be inferred based on the ratings each user has given to movies. First a weighted average of the user's ratings of movies with the particular tag is computed. Next *tagshare* of the tag, the number of time the tag is applied to a movie divided by the total number of tags of the movie, is calculated to be used as the weight factor. User's preference of the tag is computed according to a formula which returns a manipulation of the tagshare of the tag and the user's ratings for the movies with the tag in (0, 5) interval similar to the rating scale. If the user has not rated any movies with the tag, then the tag preference is unknown.

Tag relevance is computed by calculating the similarity between the tag preference of the user and the rating of the movie by the user via applying Pearson correlation formula, which is a well-known and extensively-utilized similarity metric in CF algorithms [1]. This approach enables the employment of a continuous scale rather than a binary one such as <relevant, not relevant> since it is more meaningful to concern the degree of relevance in a more detailed fashion.

Tagpslanations differ from traditional feature-based explanation techniques in the sense that tag filtering is an important component of their designs. Filtering tags is realized based on the quality of the tag, tag redundancy, and the usefulness of the tag for explanation. To deduce the quality of a tag, it is checked against particular constraints including adequate popularity, and a minimum threshold related to total number of times it is rated by users. If these constraints do not hold for the tag, it is eliminated. In order to understand whether a tag is redundant, it is checked against its possible synonyms such as (film, movie) pair and different words for the category it implies such as (violence, violent) pair. One of these tags is eliminated and the user forming the eliminated tag is supposed to form the other tag as a consequence. Lastly, tags whose preferences are undefined or relevance is very small are eliminated as they are not useful for generating explanations.

Tag-based explanations have been evaluated by conducting experiments involving users of MovieLens [7]. Users are asked questions about the explanations presented in distinct interfaces depicted in Figure 3, Figure 4, Figure 5, and Figure 6. The questions asked users to rate three proposals about each interface to evaluate the system with respect to the following criteria:

- The proposal to be rated by the participants of the experiment in order to evaluate the system with respect to *justifiability* is *"This explanation helps me understand my predicted rating."*
- To evaluate the system with respect to *effectiveness*, the users are asked to rate the proposal *"This explanation helps me determine how well I will like this movie."*
- In order to evaluate the system according to the *mood-compatibility*, which measures how well the generated explanation fits with the user's temporal feelings, situation, etc, the participants are asked to rate the proposal *"This explanation helps me decide if this movie is right for my current mood."*

In Figure 3, tags of recommended movie *Rushmore* utilized in the generated explanation are sorted with respect to relevance and for each tag user's preference is depicted using a 5-star representation. The interface is called *RelSort*. In Figure 4, tags used in the explanations for movie *Rear Window* are sorted according to preference and the relevance is also included in the interface called *PrefSort*. In Figure 5, tags for movie *The Bourne Ultimatum* are shown only according to relevance in *RelOnly* interface and in Figure 6, only the preferences of the tags for movie *The Mummy Returns* are depicted in the interface *PrefOnly*.

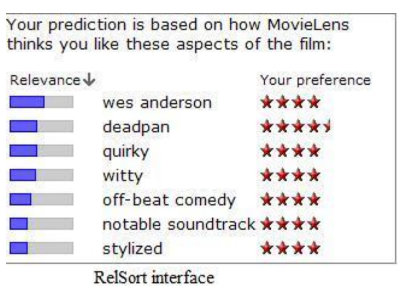

Figure 3. RelSort interface for movie *Rushmore*.

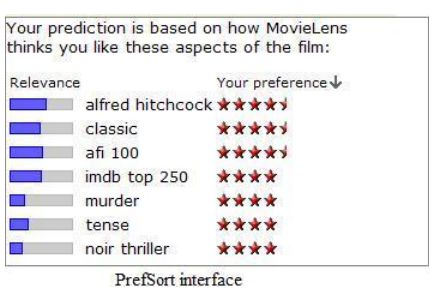

Figure 4. PrefSort interface for movie *Rear Window*.

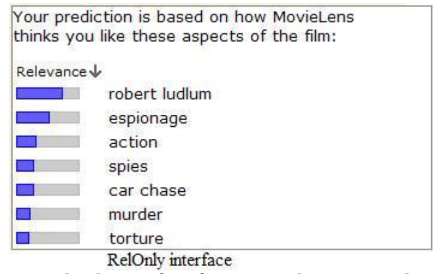

Figure 5. RelOnly interface for movie *The Bourne Ultimatom*.

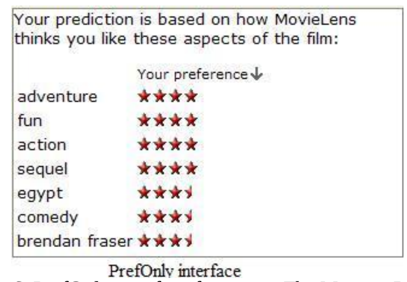

Figure 6. PrefOnly interface for movie *The Mummy Returns*.

Results are listed in Figure 7 based on the percentages to which the users either *strongly agree* or *agree* with the proposals given for the criteria.

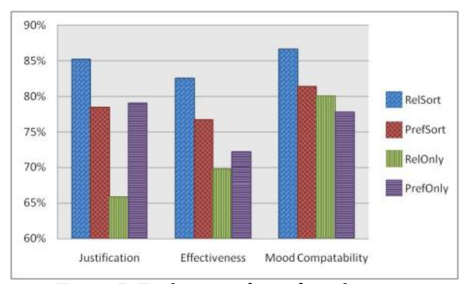

Figure 7. Evaluation of interfaces by users

Users' evaluation of explanations imply that tag preference is more important than tag relevance for justifying recommendations since the interface PrefOnly attains higher percentages than RelOnly according to the justification criterion. However, users preferred the tags to be sorted by relevance as RelSort has a higher percentage than other interfaces. According to effectiveness criterion, tag preference and tag relevance appear to have roughly equal importance as PrefOnly and RelOnly interfaces are evaluated to have close percentages. Users evaluated RelSort interface as the most effective one. Although tag relevance and tag preference appear to be equally important according to mood-compatibility criterion, it can be deduced that relevance plays its most important role in mood-compatability since RelOnly attains its highest percentages in that evaluation metric. Participants of the experiment rated the RelSort interface better than others according to three criteria.

J. Vig et al. also conducted an experiment to evaluate which kind of tags the users tend to like most in generated explanations. The results have shown that users find subjective tags more important than factual tags in all categories. However, in certain cases factual tags outperformed subjective ones such as the users preferred the factual tag *sexuality* more important than the subjective tag *sexy*. Users consider general factual tags like *World War II* more important than specific factual tags like *Manhattan*, and descriptive subjective tags like *surreal* and *dreamlike* more important than subjective tags with sexual themes or tags with opinions without descriptions like *magnificent* and *brilliant*.

Almost 82% of the participants rated the generated explanations as good overall. Thus it can be concluded that tagsplanation as an extension of the existing explanation generation approaches is successful in fulfilling its aims of justifiability and effectiveness and brings additional value to MovieLens.

J. Vig (2010) discusses the requirement of certain future work in order to extend tagsplanations to achieve other aims including scrutability [4]. MovieLens develops a new interface called the Movie Tuner that allows users to change the recommendations they receive. A sample screenshot is included in Figure 8.

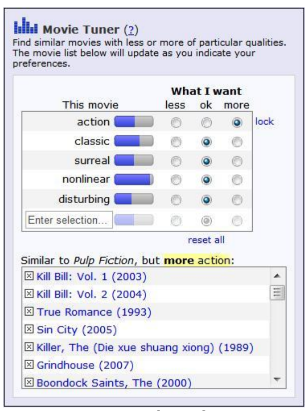

Figure 8. Movie Tuner interface for movie *Pulp Fiction*.

Movie Tuner interface injects a conversational aspect to MovieLens. Users are presented tags that are relevant to them as explanations. Tag relevance is also used to sort movies in order to answer queries like *"more action than Pulp Fiction"* more quickly. Movie Tuner chooses and lists tags utilizing a regression-based machine learning technique. This is an improvement achieved over the approach based on similarity computation to calculate the tag relevance previously. In order to select candidate tags to be displayed in user critiques (in the interface), an *entropy-based approach* to divide the space of neighbouring movies when used in critiques and a *relevance-based approach* to choose the tags that have the highest relevance to the recommended movie. A new algorithm to provide scrutability is developed to ensure that newly recommended items will significantly have the characteristics implied by the tag more or less as specified by the user.

In Figure 8, tags selected by the system are depicted to the user for movie *Pulp Fiction*. Users could query about other tags existing in the system through entering them in the text box *"Enter selection"*. The interface enables users to demand other movies having *more* or *less* of the category implied by a particular tag to be recommended to them. They could also combine critiques by clicking *lock* in the interface to demand for example *"more classic and less surreal"* movies than Pulp Fiction to be recommended.

Movie Tuner interface has brought additional aims that could be achieved by utilizing explanations, it particularly allows users to criticize the recommendations they have received and to ask for an alteration based on the particular criteria they supply. This is an achievement of tag-based explanations as it has been shown that they could unite important aims to be accomplished together. The degree of success of Movie Tuner has not been announced yet, as it is being evaluated by users according to various criteria.

V. JUSTIFIABLE AND ACCURATE RECOMMENDATIONS

P. Symeonidis et al. (2009) develop a new movie recommender system which they call MoviExplain [2]. MoviExplain combines collaborative filtering with content-based filtering to adapt a hybrid recommendation algorithm and similarly generates explanations by combining *influence* (user-based or item-based) and *keyword* (feature-based) explanation techniques.

MoviExplain relies on user's ratings of movies. Through ratings it infers users' possible votes about particular features of the rated movies. By using these features, it builds *feature profiles* for users. The clusters for users such as *users that prefer comedies* are generated to reason about collective preferences of whole communities. The generated explanations of MoviExplain are of the form *"Movie X is recommended because it contains features a, b ... which are also included in movies Z, W ... you have already rated"*. If these features occur frequently in the user's feature profile, than it could be utilized as evidence for justifying recommendations. Feature extraction is performed by making use of the Internet Movie Database (IMDB) as the knowledge-base.

The recommendation algorithm applies in stages. First user groups are created. Next the feature-weighting is performed and the neighborhood is formed. Lastly the recommendation and justification (explanation) lists are generated. These lists are presented in MoviExplain's interfaces online.

MoviExplain is evaluated with respect to statistical precision and recall and the results are compared to the ones obtained by evaluating particular hybrid recommender systems which proved to be successful previously. MoviExplain is claimed to attain better precision than similar systems [2] as it uses a clustering approach and detects particular matches among the preferences of users. Furthermore, MoviExplain achieves better explain coverage values than other systems because it is based on the notion of groups of users whilst other systems work on individual users.

Explanations generated by the system are evaluated by surveys conducted by users as well. The participants are asked to rate five movies before receiving recommendations. Then they are asked to rate each recommendation based on

 ISBN: 978-1-61208-247-9

influence, keyword and hybrid explanation approaches. Lastly they are asked to rate each recommended movie after receiving the explanation. Results obtained through the survey indicate that the hybrid explanation enables both accuracy and justifiability.

Developing hybrid approaches for generating recommendations has been discussed to improve overall accuracy. The study over the system MoviExplain has shown that hybridizing existing approaches could help in explanation technology since it is possible to attain both accurate and justifiable recommendations through generating hybrid explanations.

VI. DIVERSIFICATION BASED ON RECOMMENDATIONS

Recommender systems are occasionally faced with problems of *overspecialization*, that is recommended items are too similar to each other and the aim of introducing users with yet-unseen items that they may not encounter themselves is not satisfied. In order to overcome this problem a “flavor” of *diversity* should be added to the list of recommended items.

The goal of *recommendation diversification* is to recommend items that are dissimilar with each other but still fit with user’s tastes and preferences. Therefore certain trade-offs are taken into account so that diversification will not result in recommending users irrelevant items. C. Yu et al. (2009) show that explanations could be utilized in performing diversification [6].

The notion of similarity of explanations between distinct items can be conceptualized as the *diversity distance*. Certain correlations between the recommendation algorithm and explanation-based diversity for a list of recommended items exist. Applying the known similarity metrics utilized in the recommendation algorithm to the explanations for distinct items, one can calculate the diversity distance since explanations consist of a list of similar items and similar users.

Based on the notion of diversity distance between items, Yu et al. develop efficient algorithms for generating recommendations which achieve a good balance between relevance and diversity. Details of the algorithms and evaluation techniques are described in [6]. The results of their evaluation indicate that the proposed method indeed achieves its goals.

VII. CONCLUSIONS

Traditional statistical methods to evaluate the performance of recommender systems are not considered to be adequate. Since recommender systems are widespread and deployed with well-designed user interfaces, user-centered criteria should be devised to test the user’s relations with the system.

Generating explanations for recommendations has emerged to compensate for the need of increasing human-system interaction and bringing cognitive aspects to the recommender systems. Explanations provide several aims and distinct problem domains for which recommender systems could be developed. In order to evaluate explanations, surveys will be utilized to be carried out with a critical mass of real users.

Diversification could be attained using algorithms based on the notion of similarity between explanations or allowing flexibility by designing interfaces which include facilities through which users can state their opinions and ask for alterations to the recommended items. Interfaces providing flexibility have additional benefits as they increase the importance of cognitive aspects in recommender systems.

Explanation technology is open to contributions including new approaches such as tag processing, hybrid approaches and the notion of similarity between explanations to solve problems arising from recommendation algorithms. The state-of-the-art implies that explanations will be an integral part of all large scale commercial recommender systems both to increase results obtained by users’ evaluation of the system and provide material to improve particular problems exhibited by extensively-used recommendation algorithms.

REFERENCES

[1] G. Adomavicius and A. Tuzhilin, “Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions,” IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, 2005, pp. 734-749.

[2] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, “MoviExplain: a recommender system with explanations,” Proceedings of the 3rd ACM Conference on Recommender Systems, 2009, pp. 317-320.

[3] N. Tintarev and J. Masthoff, “A survey of explanations in recommender systems,” Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering: Workshop on Recommender Systems and Intelligent User Interfaces, 2007, pp. 801-810.

[4] J. Vig, “Intelligent tagging interfaces: beyond folksonomy,” Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology, 2010, pp. 371-374.

[5] J. Vig, S. Sen, and J. Riedl, “Tagsplanations: explaining recommendations using tags,” Proceedings of the 13th International Conference on Intelligent User Interfaces, 2009, pp. 47-56.

[6] C. Yu, L. Lakshmanan, and S. Amer-Yahia, “It takes variety to make a world: diversification in recommender systems”, Proceedings of the 12nd International Conference on Extending Database Technology: Advances in Database Technology, 2009, pp. 368-378.

[7] MovieLens web site, http://www.movielens.org, [retrieved: 01, 2013].

[8] Netflix web site, http://www.netflix.com, [retrieved: 01, 2013].