A Survey of Explanations in Recommender Systems

Nava Tintarev, Judith Masthoff
University of Aberdeen, Scotland, U.K.
ntintare,jmasthoff@csd.abdn.ac.uk

Abstract

This paper provides a comprehensive review of explanations in recommender systems. We highlight seven possible advantages of an explanation facility, and describe how existing measures can be used to evaluate the quality of explanations. Since explanations are not independent of the recommendation process, we consider how the ways recommendations are presented may affect explanations. Next, we look at different ways of interacting with explanations. The paper is illustrated with examples of explanations throughout, where possible from existing applications.

1 Introduction

The history of explanations in intelligent systems began with expert systems which were predominantly based on heuristics [7], but also on case-based reasoning (CBR) [12], and model based approaches [13]. In recent years their more commercial or entertainment inclined successors - recommender systems - have begun to offer explanations as well [5, 18, 24]. These systems represent user preferences for the purpose of suggesting items to purchase or examine, i.e. recommendations.

In the recommender systems community it is increasingly recognized that accuracy metrics such as mean average error (MAE), precision and recall, can only partially evaluate a recommender system [23]. User satisfaction, and derivatives thereof such as serendipity [23], diversity [39] and trust [9] are increasingly seen as important. This paper aims to provide a systematic overview on the open question of what makes a *good* explanation, and surveys existing approaches.

2 Why explanations are the best thing since sliced bread...

Table 1. Aims

Aim	Definition
Transparency (Tra.)	Explain how the system works
Scrutability (Scr.)	Allow users to tell the system it is wrong
Trust	Increase users' confidence in the system
Effectiveness (Efk.)	Help users make good decisions
Persuasiveness (Pers.)	Convince users to try or buy
Efficiency (Efc.)	Help users make decisions faster
Satisfaction (Sat.)	Increase the ease of usability or enjoyment

Among other things, good explanations could help inspire user trust and loyalty, increase satisfaction, make it quicker and easier for users to find what they want, and persuade them to try or purchase a recommended item. Table 1 defines seven possible aims of explanation facilities in recommender systems. Some of these aims are similar to the reasons for explaining reasoning in expert systems, c.f. [7]. In Table 2 we summarize which of these aims a number of academic recommender systems strive for. Systems for which no clear aims are stated are omitted from this table.

2.1 Explain how the system works: Transparency

An anecdotal article in the Wall Street Journal titled "*If TiVo Thinks You Are Gay, Here's How to Set It Straight*" describes users' frustration with irrelevant choices made by a video recorder that records programs it assumes its owner will like, based on shows the viewer has recorded in the past. For example, one user, Mr. Iwanyk, suspected that his TiVo thought he was gay since it inexplicably kept

Table 2. Aims of academic systems

System	Tra.	Scr.	Trust	Efk.	Per.	Efc.	Sat.
[2]			X	X			
[5]				X			
[6]		X		X			
[7]	X			X			
[10]					X	X	
[11]		X				X	
[18]				X	X		X
[20]				X		X	
[21]						X	
[24]	X					X	
[28]			X				
[31]			X				
[35]				X		X	
[37]			X		X		

recording programs with gay themes. This user clearly deserved an explanation and the chance to put things straight.

An explanation may clarify **how** a recommendation was chosen. In expert systems, such as in the domain of medical decision making, the importance of transparency has long been recognized [7]. Transparency or the heuristic of 'Visibility of System Status' is also an established usability principle [25], and it's importance has also been confirmed by user studies of recommender systems [31].

2.2 Allow users to tell the system it is wrong: Scrutability

Explanations may help isolate and correct misguided assumptions or steps. When the system collects and interprets information in the background, as is the case with TiVo, it becomes all the more important to make the reasoning available to the user. Following transparency, a second step is to allow a user to correct reasoning, or make the system *scrutable* [11]. Explanations should be part of a cycle, where the user understands what is going on in the system and exerts control over the type of recommendations made, by correcting system assumptions where needed [32]. Scrutability is related to the established usability principle of User Control [25]. See Figure 1 for an example of a scrutable holiday recommender.

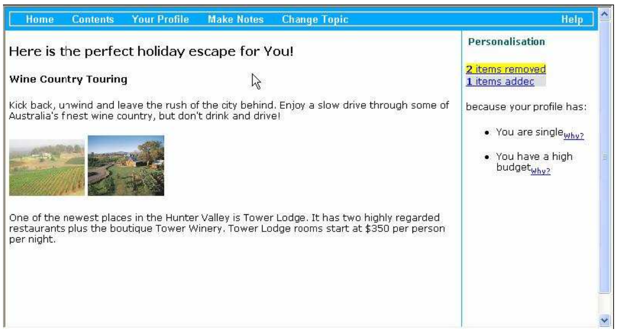

Figure 1. Scrutable adaptive hypertext, [11]

2.3 Increase users' confidence in the system: Trust

Trust is sometimes linked with transparency: previous studies indicate that transparency and the possibility of interaction with recommender systems increases user trust [14, 31]. Trust in the recommender system could also be dependent on the accuracy of the recommendation algorithm [22]. A study of users' trust (defined as perceived confidence in a recommender system's *competence*) suggests that users intend to return to recommender systems which they find trustworthy [9].

We do not claim that explanations can fully compensate for poor recommendations. On the other hand, a user may be more forgiving, and more confident in recommendations, if they understand why a bad recommendation has been made and can prevent it from occurring again. A user may also appreciate when a system is "frank" and admits that it is not confident about a particular recommendation.

In addition, the interface design of a recommender system may affect its credibility. In a study of factors determining web page credibility, the largest proportion of users' comments (46.1%) referred to the "design look" [16]. Likewise the perceived credibility of a Web article was significantly affected by the presence of a photograph of the author [15]. So design is a possible confounding factor and it is one to be seriously considered while aiming to increase the overall trustability of a recommender system.

2.4 Convince users to try or buy: Persuasiveness

Explanations may increase user evaluation of the system or the given recommendations [18]. This may qualify as persuasion, as it is an attempt to gain benefit for the system rather than for the user. In fact, it has been shown that users can be manipulated to give a rating closer to the system's prediction, whether this prediction is accurate or not [10]. In this study participants were persuaded to rent a movie, and it is possible that users may be less influenced by incorrect predictions for items that require greater investment, such as purchase of a personal computer. It is also important to consider that too much persuasion may backfire once users realize that they have tried or bought items that they do not really want.

2.5 Help users make good decisions: Effectiveness

Rather than simply persuading users to try or buy an item, an explanation may also assist users to make *better* decisions. Effectiveness is by definition highly dependent on the accuracy of the recommendation algorithm. An effective explanation would help the user evaluate the quality of suggested items according to their own preferences. This would increase the likelihood that the user discards irrelevant options while helping them to recognize useful ones. For example, a book recommender system with effective explanations would help a user to buy books they actually end up liking. Bilgic and Mooney emphasize the importance of measuring the ability of a system to assist the user in making accurate decisions about which recommendations to utilize [5].

Effective explanations could also serve the purpose of introducing a new domain, or the range of products, to a novice user, thereby helping them to understand the full range of options [14, 28].

2.6 Help users make decisions faster: Efficiency

Explanations may make it *faster* for users to decide which recommended item is best for them. Efficiency is another established usability principle, i.e. how quickly a task can be performed [25]. This criteria is one of the most commonly addressed in the recommender systems literature (See Table 2) given that the task of recommender systems is to find needles in haystacks of information.

Efficiency may be improved by allowing the user to understand the relation between competing options [20, 24, 28]. In the domain of digital cameras, competing options may for example be described as ”Less Memory and Lower Resolution and Cheaper” [20]. This way users are quickly able to find something cheaper if they are willing to settle for less memory and lower resolution.

2.7 Make the use of the system fun: Satisfaction

Explanations may increase user satisfaction with the system [34], although poor explanations are likely to decrease a user’s interest [31], or acceptance of a system [18]. The presence of longer descriptions of individual items has been found to be positively correlated with both the *perceived* usefulness and ease of use of the recommender system [31]. This can be seen as improving users’ overall satisfaction. Also, many commercial recommender systems such as those seen in Table 3 are primarily sources of entertainment. In these cases, any extra facility should take notice of the effect on user appreciation.

3 How we know an explanation is good

The advantages above can also be described as *criteria*, which can be used to assess how *good* an explanation is. Below we describe previous evaluations of explanation facilities, supplemented with a description of how existing measures could be adapted to evaluate the explanation facility in a recommender system.

In addition we would like to draw attention to the fact that some attributes of explanations may contribute toward achieving multiple goals. For instance, one can measure how *understandable* an explanation is, which can contribute to e.g. transparency, user trust, as well as satisfaction.

3.1 Transparency

There is not much previous work on evaluating transparency, and often its evaluation is coupled with scrutability [11]. It is however possible to ask if users understand how the personalization works, for example, if they believe recommendations are based on similarity with other items, or on similarity with other users etc.

Users can also be given the task of influencing the system so that it ”learns” a preference for a particular type of item, e.g. comedies in a movie recommender system. Task correctness and time to complete such a task would then be relevant quantitative measures.

3.2 Scrutability

It is important to recognize that users do not scrutinize often [11]. In an evaluation setting it is therefore important to supply users with task-based scenarios where they are more likely to scrutinize, e.g. stop receiving recommendations of Disney movies.

Quantitative measures such as time to complete a scrutinization task and task correctness were found to be misleading when interface issues (e.g. not finding the scrutability tool) arose [11]. Once interface issues were resolved, the results of questionnaires showed that users could appreciate that adaptation in the system was based on their personal attributes stored in their profile; that their profile contained information they volunteered about themselves and information that was inferred through observations made about them by the system; and that they could change their profile to control the personalization [11].

3.3 Trust

Questionnaires can be used to determining the degree of trust a user places in a system. An overview of trust questionnaires can be found in [26] which also suggests and validates a five dimensional scale of trust. Note that this validation was limited to celebrities and product endorsements, additional validation may be required to adapt this scale to a particular recommender system. Although questionnaires are very focused, they suffer from the fact that explicit preferences are not always consistent with implicit user behavior, as has been shown in regard to e.g. recommendation accuracy [27].

It is also possible to measure trust indirectly through desirable bi products, such as user loyalty and increased sales. In a study of interface effects on user opinions, loyalty was measured in terms of the number of logins and interactions with the system [22]. Among other things, this study found that allowing users to independently choose which items to rate did affect user loyalty. It has been speculated that Amazon's conservative use of recommendations, mainly recommending familiar items, enhances user trust and has led to increased sales [33].

3.4 Persuasiveness

Persuasion can be measured as the difference in likelihood of selecting an item. It is possible to measure if the evaluation of an item has changed, i.e. if the user rates an item differently after receiving an explanation. Another possibility would be to measure how much the user actually tries or buys items compared to the same user in a system without an explanation facility.

In a study of a collaborative filtering- and ratings-based recommender system for movies, participants were given different explanation interfaces [18]. This study inquired how likely users were to see one particular movie for 21 different explanation interfaces. The best response was for a *histogram* of how similar users had rated the item, with the "good" ratings clustered together and the "bad" ratings clustered together.

A persuasive explanation may also influence previous evaluations of items [10]. In this study persuasive ability was calculated as the difference between two ratings. One being a previous rating, and the second a re-rating for the same item but with an explanation interface. Naturally this also requires a baseline interface without explanations for re-rating, to control for intra-user differences over time.

Similarly to trust, the overall persuasiveness of a recommender system can also be measured quantitatively in terms of average increase in sales.

3.5 Effectiveness

One way to evaluate how effective explanations are would be to measure the liking of the recommended item prior to and after consumption. For example, in a previous study, users rated a book twice, once after receiving an explanation, and a second time after reading the book [5]. If their opinion on the book did not change much, the system was considered effective. This study explored the effect of the whole recommendation process, explanation inclusive, on effectiveness. Another possibility would be to test the same system with and without an explanation facility, and evaluate if subjects who receive explanations are on average happier with the items they selected.

Effectiveness is also the criteria that is most closely related to accuracy measures such as precision and recall [35]. In systems where items are easily consumed, e.g. internet news, these can be directly translated into recognizing relevant items and discarding irrelevant options respectively.

3.6 Efficiency

Efficiency is often used in the evaluation of so-called conversational recommender systems, where users continually interact with a recommender system, refining their preferences. A study of a personalized restaurant recommender found a significant decrease in the total amount of time, and number of interactions needed to find a satisfactory item [35]. Pu and Chen, compared completion time for two interfaces [28]. Completion time was defined as the amount of time it took a participant to locate a desired product in the interface. The results in this particular study were however not significant.

Indirect measures also include number of inspected explanations, and number of activations of repair actions [14, 30].

3.7 Satisfaction

One can directly ask users whether they prefer the system with or without explanations, and if the system is fun to use. Satisfaction can also be measured indirectly, measuring user loyalty [22, 14] (see also Section 3.3).

In measuring explanation satisfaction it is important to differentiate between satisfaction with the recommendation process, and the recommend products [14]. One (qualitative) way to measure satisfaction with the process would be to conduct a user walk-through for a task such as finding a satisfactory item. It in such a case it is possible to identify usability issues and even apply quantitative metrics such as the ratio of positive to negative comments; the number of times the evaluator was frustrated; the number of times the evaluator was delighted; the number of times and where the evaluator worked around a usability problem etc.

3.8 Choosing criteria

It is hard to create explanations that do well on all our criteria, in reality it is a trade-off. For instance, an explanation that offers great transparency may impede efficiency as the user may spend time taking in explanations, thereby increasing overall search time. An explanation that has great persuasive power might convince the user to buy books they later do not like, thereby reducing effectiveness.

When designing explanations one has to bear in mind the system goal. For instance, when building a system that sells books one might decide that user trust is the most important aspect, as it leads to user loyalty and increases sales. For selecting tv-shows, user satisfaction is probably more important than effectiveness. That is, it is more important that a user enjoys the service, than they are presented the best available shows.

4 Presenting recommendations

Some ways of presenting recommendations affect the explanation more than others. In fact, some ways of offering recommendations, such as the organizational structure we will describe shortly (See Section 4.5), can be seen as an explanation in itself.

The way a recommendation is presented may also show how good or relevant the item is considered to be. Relevance can be represented by the order in which recommendations are given. In a list, the best items are at the top. When a single item is recommended, it tends to be the best one available. Relevance can also be visualized using e.g. different colors and font sizes, or shown via ratings. Ratings can use different scales and different symbols such as numbers or stars.

Below we mention ways of offering recommendations in more detail, and illustrate how explanations may be used in each case. The content of this section is referred to as "Presentation" in the summary tables of commercial (Table 3) and academic (Table 4) recommender systems with explanations facilities.

4.1 Top item

Perhaps the simplest way to present a recommendation is by offering the user the best item. The way in which this item is selected could then be used as part of the explanation. Let us imagine a user who is interested in sport items, and appreciates football, but not tennis or hockey. The recommender system could then offer a recent football item, for example regarding the final in the world cup. The generated explanation may then be the following:

"You have been watching a lot of sports, and football in particular. This is the most popular and recent item from the world cup."

Note that this example uses the user's viewing history in order to form the explanation. A system could also use information that the user specifies more directly, e.g. how much they like football.

4.2 Top N-items

The system may also present several items at once. In a large domain such as news, it is likely that a user has many interests. In this case there are several items that could be highly interesting to the user. If the football fan mentioned above is also interested in technology news, the system might present several sports stories alongside a couple of technology items. Thus, the explanation generated by the system might be along the lines of:

"*You have watched a lot of football and technology items. You might like to see the local football results and the gadget of the day.*"

The system may also present several items on the same theme, such as several football results. Note that while this system should be able to explain the relation between chosen items, it should still be able to explain the rational behind each single item.

4.3 Similar to top item(s)

Once a user shows a preference for one or more items, the recommender system can offer *similar* items. The recommender system could present a single item, a list of preferred items, or even your most currently viewed items. For each item, it can present one or more similar items, and may show explanations similar to the ones in Sections 4.1 and 4.2. For example, given that a user liked a book by *Charles Dickens* such as Great Expectations, the system may present a recommendations in the following way; *"You might also like...Oliver Twist by Charles Dickens"*.

A recommender system can also offer recommendations in a social context, taking into account users that are similar to you. For example a recommendation can be presented in the following manner; *"People like you liked..Oliver Twist by Charles Dickens"*.

Table 3. A selection of commercial recommender systems with explanation facilities.

System	**Item type**	**Presentation** (Section 4)	**Explanation**	**Interaction** (Section 5)
Amazon	e.g. Books, Movies	Similar to top item(s)	Content-based	Rating, opinion
Findory	News	Similar to top item(s)	Preference-based	(Implicit) rating
LibraryThing	Books	Similar to top item(s)	Collaborative-based	Rating
LoveFilm	Movies	Top-N, Predicted ratings	Content-based	Rating
OkCupid	People to date	Top-N, Predicted ratings	Preference-based	Specify reqs.
Pandora	Music	Top item	Preference-based	Opinion
StumbleUpon	Web pages	Top item	Preference-based	Opinion
Qwikshop [20]	Digital cameras	Top item, Similar to top item	Preference-based	Alteration

Table 4. A selection of academic recommender systems with explanation facilities.

System	**Item type**	**Presentation** (Section 4)	**Explanation**	**Interaction** (Section 5)
LIBRA [5]	Books	Top-N, Predicted ratings	Content-based, Collaborative-based	Rating
News Dude [6]	News	Top-N items	Preference-based	Opinion
MYCIN [7]	Prescriptions	Top item	Preference-based	Specify reqs.
MovieLens [10, 18]	Movies	Top-N, Predicted ratings	Collaborative-based	Rating
SASY [11]	E.g. holiday	Top item	Preference-based	Alteration
Sim [21]	PCs	Top-N	Preference-based	(varied)
Top Case [24]	Holiday	Top-item, Similar to top item	Preference-based	Specify reqs.
"Organizational Structure" [28]	Digital camera, notebook computer	Structured overview	Preference-based	(None)
ADAPTIVE PLACE ADVISOR [35]	Restaurants	Top item	Preference-based	Specify reqs.
ACORN [37]	Movies	Structured overview, Top-N	Preference-based	Specify reqs.

4.4 Predicted ratings for all items

Rather than forcing selections on the user, a system may allow its users to browse all the available options. Recommendations are then presented as predicted ratings on a scale (say from 0 to 5) for each item. A user may then still find items with low *predicted* ratings, and can counteract predictions by rating the affected items, or directly modifying the user model, i.e. changing the system's view of their preferences. This allows the user to tell the system when it is wrong, fulfilling the criteria of Scrutability (see Section 2.2). Let us re-use our example of the football and technology fan. This type of system might on average offer higher predictions for football items than hockey items. A user might then ask why a certain item, for example local hockey results, is predicted to have a low rating. The recommender system might then generate an explanation like:

"This is a sports item, but it is about hockey. You do not seem to like hockey!".

If the user is interested in local hockey results, but not in results from other countries, they might modify their user model to limit their interest in hockey to local sports.

4.5 Structured overview

Pu and Chen [28] suggest a structure which displays trade-offs between items. The best matching item is displayed at the top. Below it several categories of trade-off alternatives are listed. Each category has a title explaining the characteristics of the items in it, e.g.

"[these laptops]...are cheaper and lighter, but have

lower processor speed".

The order of the titles depends on how well the category matches the user's requirements.

Yee et al. [38] used a multi-faceted approach for museum search and browsing. This approach considers *several* aspects of each item, such as location, date and material, each with a number of levels. The user can see how many items there are available at each level for each aspect. Using multiple aspects might be a suitable approach for a large domain with many varying objects.

Although not yet used in recommender systems, the "treemap" structure (see Figure 2) allows a different type of overview [4]. Here it is possible to use different colors to represent topic areas, square and font size to represent importance to the current user, and shades of each topic color to represent recency.

The advantage of a structured overview is that the user can see "where" they are in the search space, and possibly how many items can be found in each category. This greatly facilitates both navigation and user comprehension of the available options.

Figure 2. Treemap visualization of news, [1]

4.6 Recommender "Personality"

The choice of recommended items, or the predicted rating for an item can be angled to reflect a "personality" of the recommender system [23]. The recommender may have an *affirming* personality, supplying the user with recommendations of items they might already know about. This could inspire a user's *trust* (see Section 2.3) in the system's ability to present relevant or accurate items. Or, on the contrary, it may aim to offer more *novel* and positively surprising (serendipitous) recommendations in order to increase user *satisfaction*.

When a recommendation is made, it is operating along two often conflicting dimensions [19]. The first dimension is the strength of the recommendation: how much does the recommender system think the user will like this item. The second dimension is the confidence of the recommendation: how sure is the recommender system that its recommendation is accurate. A recommender system can be *bold* and recommend items more strongly than it normally would, or it could simply state its true *confidence* in its own recommendation [18].

If such factors are part of the recommendation process, the criteria of Transparency (see Section 2.1) suggests that they should be part of the explanations as well.

5 Interacting with the recommender system

Justifying recommendations to the user is only half of the solution, the second half is making the system *scrutable* (see Section 2.2) by allowing the user to make changes. So called conversational systems allow users to elaborate their requirements over the course of an extended dialog [29]. This contrasts with standard "single-shot" recommender systems, where each user interaction is treated independently of previous history.

There are different ways in which a user can give feedback to the system to let it know how it is doing. Here we expand on four ways suggested by [17], supplying examples of current applications. Note that although unobtrusive elicitation of user preferences, via e.g. usage data [27] or stereotypes [2] can be sometimes be more effective [27], this section focuses on *explicit* feedback from users.

The content of this section is referred to as "Interaction" in the summary tables of commercial (Table 3) and academic (Table 4) recommender systems with explanations facilities .

5.1 The user specifies their requirements

By allowing a user to directly specify their requirements it is possible to circumvent the type of faulty assumptions that can be made by a system where the interests of a user are based on the items they decide to see, or how they interact with them. An indirect approach depends on the way it collects data, and is sometimes hard to change in the intended direction. Remember Mr. Iwanyk? This user tried to tame TiVo's gay fixation by recording war movies and other "guy stuff". As a result it started to give him documentaries on Joseph Goebbels and Adolf Eichmann. It stopped thinking he was gay and decided he was a crazy guy reminiscing about the Third Reich!

One movie recommender system allows users to specify their preferences in plain English [36], p. 77:

” **User:** I feel like watching a thriller.
System: Can you tell me one of your favorite thriller movies.
User: Uhm, I'm not sure
System: Okay. Can you tell me one of your favorite actors or actresses?
User: I think *Bruce Willis* is good
System: I see. Have you seen *Pulp Fiction*?
User: No
System: Pulp Fiction is a thriller starring *Bruce Willis*”

This dialog does not make use of the user's previous interests, nor does it explain *directly*. That is, there is no sentence that claims to be a justification of the recommendation. It does however do so indirectly, by reiterating (and satisfying) the user's *requirements*. The user should then be able to interact with the recommender system, and give their opinion of the recommendation, thus allowing further refinement.

5.2 The user asks for an alteration

A more direct approach is to allow users to explicitly ask for alterations to recommended items, for instance using the structured overview described in Section 4.5. This approach helps the users to find what they want quicker. Users can see how items compare, and see what other items are still available if the current recommendation should not meet their requirements. Have you ever put in a search for a flight, and been told to try other dates, other airports or destinations? This answer does not explain which of your criteria needs to be changed, requiring you to go through a tiring trial-and-error process. If you can see the trade-offs between alternatives from the start, the initial problem can be circumvented.

Some feedback facilities allow users to see how criteria affect their remaining options. One such system explains the difference between a selected camera and remaining cameras. For example, it describes competing cameras with ”Less Memory and Lower Resolution and Cheaper” [20].

The user can ask for a more detailed explanation of the alternative criteria, and have a look at the cameras which fulfill these criteria. Instead of simply explaining to a user that no items fitting the description exist, these systems show what types of items *do* exist. These methods have the advantage of helping users find good enough items, even if some of their initial requirements were too strict.

5.3 The user rates items

To change the type of recommendations they receive, the user may want to correct predicted ratings, or modify a rating they made in the past. Rating may be explicitly inputted by the user, or inferred from usage patterns. In a book recommender system a user could see the influence (in percentage) their previous ratings had on a given recommendation [5]. The *influence based explanation* showed which rated titles influenced the recommended book the most (see Figure 3). Although this particular system did not allow the user to modify previous ratings, or degree of influence, in the explanation interface, it can be imagined that this functionality could be implemented. Note however, that ratings are often easier to modify than the degree of influence which is likely to be computed.

BOOK	YOUR RATING Out of 5	INFLUENCE Out of 100
Of Mice and Men	4	54
1984	4	50
Till We Have Faces : A Myth Retold	5	50
Crime and Punishment	4	46
The Gambler	5	11

Figure 3. Influence of ratings on recommendation,[5]

5.4 The user gives their opinion

A common usability principle is that it is easier for humans to recognize items, than to draw them from memory. Therefore, it is sometimes easier for a user to say what they want or do not want, when they have options in front of them. The options mentioned below can be simplified to be mutually exclusive, e.g. either a user likes an item or they do not. It is equally possible to create an explanation facility using a sliding scale.

We discuss some options inspired by previous work in recommender systems [33, 6]. For example, a user should be able to specify whether they think an item is interesting or not, if they would like to see more similar items, or if they have already seen the item previously [6].

- **More like this:** If a user likes an item they are seeing, they can ask to receive more items of this type. Assuming the user likes an item, it does not necessarily mean they want another one just like it next.
  - *”More later!”* - The user may have had enough, but would like to specify that it is the type of item they would generally like to receive. For example, the user might like a certain author and

want to know about any books they write in the future.

  – *"Give me more!"* - The user may be recommended an item which they like, and want to see even more items. For example, if the user discovers a new genre, they may want to be recommended more books in this genre.

- **No more like this:** Likewise, there are different situations in which a user would request not to receive recommendations of a certain type. For example, a recommender system may recommend books that a user has already read, or is planning to read.

  – *"I already know this!"* - The user may already be familiar with the content of an item. This feedback is not necessarily negative; this depends on the rating the user gives the item as well! It is possible that this recommendation is of the type a user likes, and would have sought out on their own. In a way the user is specifying that the recommendation is correct, but is asking not to increase the likelihood for receiving similar recommendations.

  – *"No more like this!"* - The user may dislike, or be disinterested in an item.

  It can also be imagined that there are certain aspects of an item that a user likes more than others. In that case, it should be possible for the user to specify this in greater detail. A feedback facility could allow users to specify how they relate to specific aspects of an item. For example, rather than specifying that they do not like a specific sports item, the user may want to say they like the sport, but not that the game took place at a distant location.

- **Surprise me!** Often recommender systems do not have enough information about a user. Analogously, a user may wish to broaden their horizon of possibilities. That is, a user may ask a system to *"Surprise me!"*. The system may also mark on a sliding bar, to which extent it offers random recommendations. It may also inform the user how much information the system gets out of the user rating the item. This information can consequently be used to generate explanations, and allow the user to modify the settings accordingly.

6 Conclusion

When choosing and comparing explanation techniques, it is very important to agree on what the explanation is trying to achieve. For example, although the study by [18] measured user satisfaction with *recommendations* (persuasion), this is not the same as measuring satisfaction with actual *items* (effectiveness) [5]. As shown in this paper, explanations are intrinsically linked with the way recommendations are presented (Section 4), and with the degree of interactivity offered (Section 5). Tables 3 and 4 offer an overview of explanation facilities in existing commercial and academic recommender systems respectively. The column "explanation" reflects the content of the explanation regardless of the underlying algorithm, e.g. *"We have recommended X because you liked Y"* (content-based explanation), *"People who liked X also liked Y"* (collaborative-based explanation), *"Your interests suggest that you would like X"* (preference-based explanation),

There are a number of future research directions we would like to explore. One direction is to define similarity measures which are easily understood by users, and investigate how these measures can be adapted to each user. A system that can explain to the user *in their own terms* why items are recommended is likely to increase user trust, as well as system transparency and scrutability.

Although beyond the scope of this paper, a second direction is to extend existing research on modalities of explanations [3, 8, 9, 18], but rather than assuming that either text or images are preferable, see how they can compliment each other.

We hope that the framework and survey presented in this paper will lead to more systematic research on explanations in recommender systems.

References

[1] http://www.marumushi.com/apps/newsmap/index.cfm, retrieved 02 September, 2006.

[2] L. Adrissono, A. Goy, G. Petrone, M. Segnan, and P. Torasso. INTRIGUE: Personalized recommendation of tourist attractions for desktop and handheld devices. *Applied Artificial Intelligence*, 17:687–714., 2003.

[3] J. V. Barneveld and M. V. Setten. *Personalized digital television*, chapter 10, pages 259–285. Kluwer Academic Publishers, 2004.

[4] B. Bederson, B. Shneiderman, and M. Wattenberg. Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. *ACM Transactions on Graphics*, 21(4):833–854., 2002.

[5] M. Bilgic and R. J. Mooney. Explaining recommendations: Satisfaction vs. promotion. In *Beyond Personalization Workshop, IUI*, 2005.

[6] D. Billsus and M. J. Pazzani. A personal news agent that talks, learns, and explains. In *Proceedings of the Third International Conference on Autonomous Agents*, 1999.

[7] B. G. Buchanan and E. H. Shortliffe, editors. *The Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, chapter 30-35, pages 571–665. Addison-Wesley Publishing Company, 1985.

[8] G. Carenini and J. J. Moore. An empirical study of the influence of argument conciseness on argument effectiveness. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, 2000.

[9] L. Chen and P. Pu. Trust building in recommender agents. In *WPRSIU'02*, 2002.

[10] D. Cosley, Lam, S. K., I. Albert, Konstan, J. A., and J. Riedl. Is seeing believing?: how recommender system interfaces affect users' opinions. In *ACM CHI*, volume 1 of *Recommender systems and social computing*, pages 585–592, 2003.

[11] M. Czarkowski. *A Scrutable Adaptive Hypertext*. PhD thesis, University of Sydney, 2006.

[12] D. Doyle, A. Tsymbal, and P. Cunningham. A review of explanation and explanation in case-based reasoning. Technical report, Department of Computer Science, Trinity College, Dublin, 2003.

[13] M. J. Druzdzel. Qualitative verbal explanations in bayesian belief networks. *Artificial Intelligence and Simulation of Behaviour Quarterly, special issue on Bayesian networks*, pages 43–54, 1996.

[14] A. Felfernig and B. Gula. Consumer behavior in the interaction with knowledge-based recommender applications. In *ECAI 2006 Workshop on Recommender Systems*, 2006.

[15] B. Fogg, J. Marshall, T. Kameda, J. Solomon, A. Rangnekar, J. Boyd, and B. Brown. Web credibility research: A method for online experiments and early study results. In *CHI 2001*, 2001.

[16] B. J. Fogg, C. Soohoo, D. R. Danielson, L. Marable, J. Stanford, and E. R. Tauber. How do users evaluate the credibility of web sites?: a study with over 2,500 participants. In *Proceedings of DUX'03: Designing for User Experiences*, number 15 in Focusing on user-to-product relationships, pages 1–15, 2003.

[17] L. M. Ginty and B. Smyth. Comparison-based recommendation. *Lecture Notes in Computer Science*, 2416:575, 2002.

[18] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *ACM conference on Computer supported cooperative work*, 2000.

[19] J. L. Herlocker, J. A. Konstan, L. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53, 2004.

[20] K. McCarthy, J. Reilly, L. McGinty, and B. Smyth. Thinking positively - explanatory feedback for conversational recommender systems. In *Proceedings of the European Conference on Case-Based Reasoning (ECCBR-04) Explanation Workshop,*, 2004.

[21] L. McGinty and B. Smyth. Extending comparison-based recommendation: A review. In *Poster acceptance for the British Computer Society's Specialist Group on Artificial Intelligence (AI-03)*, 2003.

[22] S. M. McNee, S. K. Lam, J. A. Konstan, and J. Riedl. Interfaces for eliciting new user preferences in recommender systems. *User Modeling*, pages pp. 178–187, 2003.

[23] S. M. McNee, J. Riedl, and J. A. Konstan. Being accurate is not enough: How accuracy metrics have hurt recommender systems. In *Extended Abstracts of the 2006 ACM Conference on Human Factors in Computing Systems (CHI 2006)*, 2006.

[24] D. Mcsherry. Explanation in recommender systems. *Artificial Intelligence Review*, 24(2):179 – 197, 2005.

[25] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *ACM CHI'90*, 1990.

[26] R. Ohanian. Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, 19:3:39, 1990.

[27] D. O'Sullivan, B. Smyth, D. C. Wilson, K. McDonald, and A. Smeaton. Improving the quality of the personalized electronic program guide. *User Modeling and User-Adapted Interaction*, 14:pp. 5–36, 2004.

[28] P. Pu and L. Chen. Trust building with explanation interfaces. In *IUI'06*, Recommendations I, pages 93–100, 2006.

[29] R. Rafter and B. Smyth. Conversational collaborative recommendation - an experimental analysis. *Artif. Intell. Rev*, 24(3-4):301–318, 2005.

[30] J. Reilly, K. McCarthy, L. McGinty, and B. Smyth. Dynamic critiquing. In P. Funk and P. A. González-Calero, editors, *ECCBR*, volume 3155 of *Lecture Notes in Computer Science*, pages 763–777. Springer, 2004.

[31] R. Sinha and K. Swearingen. The role of transparency in recommender systems. In *Conference on Human Factors in Computing Systems*, 2002.

[32] F. Sørmo, J. Cassens, and A. Aamodt. Explanation in case-based reasoning perspectives and goals. *Artificial Intelligence Review*, 24(2):109 – 143, 2005.

[33] K. Swearingen and R. Sinha. Interaction design for recommender systems. In *Designing Interactive Systems*, 2002.

[34] K. Tanaka-Ishii and I. Frank. Multi-agent explanation strategies in real-time domains. In *38th Annual Meeting on Association for Computational Linguistics*, 2000.

[35] C. A. Thompson, M. H. Göker, and P. Langley. A personalized system for conversational recommendations. *J. Artif. Intell. Res. (JAIR)*, 21:393–428, 2004.

[36] P. Wärnestål. Modeling a dialogue strategy for personalized movie recommendations. In *Beyond Personalization Workshop*, 2005.

[37] P. Wärnestål. User evaluation of a conversational recommender system. In *Proceedings of the 4th Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, 2005.

[38] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *ACM Conference on Computer-Human Interaction*, 2003.

[39] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *WWW'05*, 2005.