A Survey of Music Recommendation Aids

Pirkka Åman and Lassi A. Liikkanen
Helsinki Institute for Information Technology HIIT
Aalto University and University of Helsinki
Tel. +358 50 384 1514
firstname.lastname@hiit.fi

ABSTRACT

This paper provides a review of explanations, visualizations and interactive elements of user interfaces (UI) in music recommendation systems. We call these UI features "recommendation aids". Explanations are elements of the interface that inform the user why a certain recommendation was made. We highlight six possible goals for explanations, resulting in overall satisfaction towards the system. We found that the most of the existing music recommenders of popular systems provide no explanations, or very limited ones. Since explanations are not independent of other UI elements in recommendation process, we consider how the other elements can be used to achieve the same goals. To this end, we evaluated several existing music recommenders. We wanted to discover which of the six goals (transparency, scrutability, effectiveness, persuasiveness, efficiency and trust) the different UI elements promote in the existing music recommenders, and how they could be measured in order to create a simple framework for evaluating recommender UIs. By using this framework designers of recommendation systems could promote users' trust and overall satisfaction towards a recommender system thereby improving the user experience with the system.

Categories and Subject Descriptors

H5.m. Information interfaces and presentation: Miscellaneous. H.5.5 Sound and Music Computing.

Author Keywords

Recommendation systems, music recommendation, explanations, user experience, UI design.

1. INTRODUCTION

Recommender systems are a specific type of information filtering technique that aims at presenting items (music, news, other users, etc.) that user the might be interested in. To do this, information about the user is compared to reference characteristics, e.g. information on the other users of the system (collaborative filtering) or content features, such as genre in the case of books or music (content-based filtering). In its most common formulation, the recommendation task is reduced to the problem of estimating relevance of the items that a user has not encountered yet, and then presenting the items that have the highest estimated ratings [6]. The importance of recommender systems lies in their potential to help users to more effectively identify items of interest from a potentially overwhelming set of choices [7]. The importance of these mechanisms has become evident as commercial services over the Internet have extended their catalogue to dimensions unexplorable to a single user. However, the overwhelming numbers of content create a constant competition and can reduce the usefulness of recommendations unless they can persuade the user to try the suggested content. Explanations and other recommendation aiding UI features are examined in this paper as a way to increase the satisfaction towards recommenders among users.

The first interactive systems to have explanations were expert systems, including legal and medical databases [4]. Their present successors are commercial recommendation systems commonly found embedded in various entertainment systems such as iTunes [9] or Last.fm [12]. Explanations can be described as textual information telling e.g. why and how a recommendation was produced to the user. Earlier research shows that even rudimentary explanations build more trust towards the systems than the so-called "black box" recommenders [13]. Explanations also provide system developers a graceful way for handling errors that recommender algorithms sometimes produce [6].

The majority of previous recommendation system research has been focused on the statistical accuracy of the algorithms driving the systems, with little emphasis on interface issues and the user experiences [13]. However, it has been noted lately that when the new algorithms are compared to the older ones, both tuned to the optimum, they all produce nearly similar results. Researchers have speculated that we may have reached a level where human variability prevents the systems from getting much more accurate [7]. This mirrors the human factor: it has been shown that users provide inconsistent ratings when asked to rate the same item several times [14]. Thereby an algorithm cannot be more accurate than the variance in the user's ratings for the same item.

An important aspect for the assessment of recommendation systems is to evaluate them subjectively, e.g. how well they can communicate their reasoning to users. That's why user interface elements such as explanations, interactive elements and visualizations are increasingly important in improving user experience. In the past years subjectively perceived aspects of recommendations systems have accordingly gained ground in their evaluation.

In this paper we want to illustrate the possibilities of user-evaluation of recommendation supporting features in recommendation systems. We do this by performing a review on several publicly available music recommenders. Music is today one of the most ubiquitous commodities and the availability of digital music is constantly growing. Massive online music libraries with millions of tracks are easily available in the Internet. However, finding new and relevant music from those vast collections of music becomes similarly increasingly difficult. One approach to tackle the problem of finding new, relevant music is developing better (reliable and trustworthy) recommendation systems. Music recommenders are also easy to access and music has reasonably short process in determining the quality of recommendation results.

WOMRAD 2010 Workshop on Music Recommendation and Discovery, colocated with ACM RecSys 2010 (Barcelona, SPAIN)

2. GOALS FOR RECOMMENDATION AIDS

Tintarev and Masthoff [16] present a taxonomy for evaluating goals for explanations. Those are shown slightly modified in the Table 1 below. We argue that satisfaction towards a recommendation system is an aggregate of the six other dimensions, more a goal of itself than the other dimensions. In addition, we noticed that the dimensions are not so straightforward as Tintarev and Masthoff present them. Some of them cannot be evaluated using objective measures, and therefore framework for evaluation recommendation aids must be drawn from user research. In the following we describe each dimension and give examples of how they could be evaluated and measured.

Table 1. Dimensions for recommendation explanations.

Goal	Definition
Transparency	Explain how the system works
Scrutability	Allow users to tell the system it is wrong
Effectiveness	Help users make good decisions
Persuasiveness	Convince users to try or buy
Efficiency	Help users make decisions faster
Trust	Increase users' confidence in the system
Resulting in →	
Satisfaction (increasing the ease of use or enjoyment towards the system)	

1. An explanation may tell users how or why a recommendation was made, allowing them to see behind the UI and thus making recommendation *transparent*. Transparency is also a standard usability principle, formulated as a heuristic of 'Visibility of System Status' [13]. Transparency can be measured objectively, using binary scale (yes/no), e.g. if a UI provides some kind of explanation how a recommendation was made transparency gets a vote. However, evaluating transparency subjectively may involve users to be asked if they understand how the recommendation was made using e.g. Likert scale.

2. *Scrutability* means that users are allowed to provide feedback for the system about the recommendations. Scrutability is related to the established usability principle of 'User Control' [13]. Scrutability can be measured objectively by finding out if there is a way to tell the system it is wrong. To evaluate scrutability subjectively, users may be given a task to find a way to tell how to stop receiving e.g. recommendations of Elvis songs. If users feel they can control the recommendations by changing their profile, the UI has the possibility to scrutinize.

3. *Effectiveness* of an explanation help users make better decisions. Effectiveness is 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 [16]. This would increase the likelihood that the user discards irrelevant options while helping them to recognize useful ones. Unlike travel or film recommenders, in the case of music recommenders the process of deciding the goodness of a recommendation is done quite quickly.

4. *Persuasiveness*. Explanations may convince users to try or buy recommended items. However, persuasion may result in an adverse reaction towards the system, if users continuously decide to choose bad recommendations. Persuasion could be measured according to how much the user actually tries or buys items compared to the same user in a system without an explanation facility [16] and what kind of persuasion techniques are utilized. Persuasion could also be measured by applying click-through rates used in measuring online ads.

5. *Efficient* explanations help users to decide faster which recommendation items are best for their current situation. Efficiency can be improved by allowing the user to understand the relation between recommended options [12]. A simple way to evaluate efficiency is to give users tasks and measure how long it takes to find e.g. an artist that is novel and pleasing to the user.

6. Increasing users' confidence in the system results in *trust* towards a recommender. Trust is in the core of any kind of recommendation process, and it is perhaps the most important single factor leading to better user satisfaction and user experience with the interactive system. 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 [2]. The interface design of a recommender affects its credibility and earlier research has shown that in user evaluation of web page credibility the largest proportion of users' comments referred to the UI design issues [5]. Trust needs to be measured using subjective scales over multiple tasks or questions about the recommendation aiding features of a recommender UI.

The ease of use or enjoyment results finally in more *satisfaction* towards a system. Descriptions of recommended items have been found to be positively correlated with both the perceived usefulness and ease of use of the recommender system [6], enhancing users' overall satisfaction. Even though we see satisfaction as an aggregate of the dimensions presented above, satisfaction with the process could be measured e.g. by conducting a user walk-through for a task such as finding a satisfactory item.

3. RELATED EMPIRICAL RESEARCH

It is widely agreed that expert systems that act as decision-support systems need to provide explanations and justifications for their advice [13]. However, there is no clear consensus on how explanations should be designed in conjunction with other UI elements or evaluated by users. Studies with search engines show the importance of explanations. Koenmann & Belkin [11] found that greater interactivity for feedback on recommendations helped search performance and satisfaction with the system. Johnson & Johnson [10] note that explanations play a crucial role in the interaction between users and interactive systems. According to their research, one purpose of explanations is to illustrate the relationship between cause and effect. In the context of recommender systems, understanding the relationship between the input to the system (ratings and choices made by user) and output (recommendations) allows the user to interact efficiently with the system. Sinha and Swearingen [15] studied the role of transparency in recommender systems. Their results show that users like and feel more confident about recommendations that they perceive transparent. Explanations allow users to meaningfully revise the input in order to improve recommendations, rather than making "shots in the dark."

Herlocker and Konstan [6] suggest that recommender systems have not been used in high-risk decision-making because of a lack of transparency. While users might take a chance on an opaque movie recommendation, they might be unwilling e.g. to commit to a vacation spot without understanding the reasoning behind such a recommendation. Building an explanation facility into a recommender system can benefit the user in various ways. It removes the "black box" around the recommender system, providing transparency. Some of the other benefits include *justification*. If users understand the reasoning behind a recommendation, they may decide how much confidence to place in the suggestion. That results in greater *acceptance* or *satisfaction* of the recommender system as a decision aide, since its limits and strengths are more visible and its suggestions are justified.

4. RECOMMENDATION AIDS IN EXISTING MUSIC RECOMMENDERS

We conducted an expert walkthrough of six publicly available music systems with recommendation functionalities in order to find out which of the six goals explanations, visualizations and interactive UI elements promote in the existing music recommenders, and how they can be measured in order to create a simple framework for evaluating recommenders. The walkthrough was conducted by authors listing the UI features capable of promoting the goals mentioned above. The reviewed systems include Pandora, Amazon, Last.fm, Audiobaba, Musicovery and Spotify. We wanted to include the most popular online music services, and on the other hand, include a variety of different UIs. Each of the evaluated systems provides recommendations but not necessarily explanations. Systems without textual explanations were also included in order to find out what kind of goals or functions similar to verbal explanations other recommendation aids provide.

Table 2. The occurrences of recommendation aids in a selection of music recommenders

	Trans.	Scrt.	Effect.	Pers.	Effic.	Trust	
Amazon	1	2	2	3	1	3	12
Last.fm	-	2	2	1	2	2	9
Audiobaba	1	1	2	1	1	2	8
Musicovery	2	2	2	2	2	1	11
Spotify	-	-	1	1	1	1	4
Pandora	2	2	3	3	2	3	15
	6	9	12	11	9	12	

If a recommender has a possibility to promote a goal with explanations, visualizations or interactive elements, it gets a vote in Table 2. For example, persuasiveness promoted through visualizations is potentially possible in all of the interfaces that have visualizations, even rudimentary ones, such as an album cover. A single user might be persuaded to try or buy by presenting a subjectively compelling album cover. From Table 2 we can see that Pandora, Amazon and Musicovery have the greatest number of UI elements able to provide users support for sense-making of recommendations. Effectiveness, persuasiveness and trust are the most commonly promoted goals. In each recommender, each UI element has the potential to increase trust towards the systems, but for more accurate measurement, it remains to be evaluated by empirical user research, to which extent each elements in certain recommender interface really promote trust. This applies to most of the six goals: without empirical data, it is almost impossible to decide, whether the potential for promoting effectiveness, persuasiveness and efficiency actually realizes. Only transparency and scrutability can be measured using objective binary scale of yes/no, but they can be evaluated also using subjective (Likert style) scales. We argue that by measuring these goals for UI elements together with a set of usability guidelines, it is possible to evaluate and design better user experiences for recommendation systems.

Some of the dimensions are easy to connect to certain UI elements. For instance, scrutability is usually designed as a combination of explanation and interactivity, whereas other, more general level dimensions depend strongly on subjective experience and are hard to connect with specific UI elements. For example, satisfaction or trust towards a system is usually combination of different experienced UI dimensions. Therefore the most common dimensions promoted in the evaluated systems were trust and satisfaction. Those, together with persuasiveness, are experienced very subjectively, which means that empirical user evaluation is needed for more reliable and comparable evaluations of those dimensions.

Obvious example of an explanation providing transparency is Amazon's "Customers with Similar Searches Purchased…", with up to ten albums' list. Pandora tells a user: "This song was recommended to you because it has jazzy vocals, light rhythm and a horn section." Transparency is very hard to achieve without textual, explicit explanations. Of the reviewed systems, only Musicovery's UI with several interactive elements, graphical visualization of the recommendations and the relations between them give users clear clues of why certain pieces of music were recommended, without providing explanations.

Last.fm offers users scrutability in many ways, e.g. with its music player (Figure 1). One of the system's more sophisticated scrutinizing tactics is a social one. Last.fm allows users to turn off the registering (called scrobbling) of the listened music. The system's users can perform identity work by turning scrobbling off, if they feel they do not want to communicate what they have listened to the other users. Amazon provides "Fix this recommendation" option for telling the system to remove recommended item from the users browsing history.

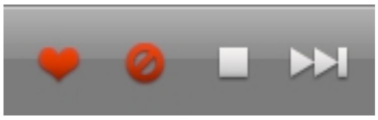

Figure 1: Example of scrutable interactivity: Last.fm player's love, ban, stop and skip buttons give users a tool to control their profiles and thereby affect recommendations.

Users can be helped in efficiency and effectiveness, i.e. making better and faster decisions by offering appropriate controls with interactive elements. For instance, Musicovery's timeline slider is presented in Figure 2. It works in real time with the system's graphical presentation of recommended items.

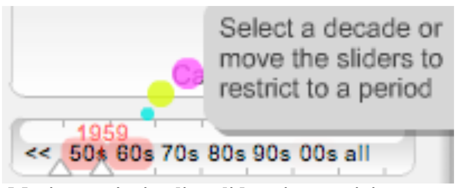

Figure 2: Musicovery's timeline slider: interactivity promoting efficiency, scrutability, and effectiveness, resulting in more trust and satisfaction towards the system.

5. DISCUSSION AND CONCLUSIONS

We reviewed dimensions of explanations in six music recommendation systems and found out that most of the reviewed commercial music recommendation systems are "black boxes", producing recommendations without any, or very limited explanations. Most of the dimensions are poorly promoted by textual explanations, but can be promoted by other means, namely by visualizations and interactive elements, and further, by user-generated content and social facilities. From the expert walkthrough of the selected music recommendation systems we can draw a tentative conclusion that if UI elements can fulfill similar functions as explanations, there is necessarily no need for textual descriptions. By using non-verbal recommendation aids as "implicit" explanations and using them in recommendation system design, we can promote better user experience. This is the case especially when the user has enough cultural capital and therefore competence for "joining the dots" between recommended items without explicit explanations. On the other hand, if the recommender is used e.g. for learning about musical genre, textual explanations may be indispensable.

As an example of the dimensions that UI elements other than verbal explanations can promote is the overall satisfaction or trust towards the systems that can be achieved by conversational interaction such as in UI example presented in Figure 3, where users are given a chance for optional recommendations based on their situational desires and needs.

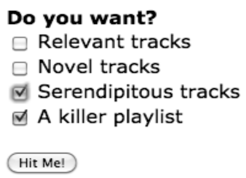

Figure 3: A recommendation aid with optional inputs.

Last.fm is an example of recommendation system with no explanations. However, it has an abundance of other elements such as user created biographies, genre tags and pictures of artists, not to mention advanced social media features that together effectively work towards the same goals as the dimensions of explanations. Furthermore, Spotify, a popular European music service with very simple recommendation facility, does not provide any explanations whatsoever. Its popularity relies on providing users a minimalistic UI with effective search facility and a functional, high-quality audio streaming. Spotify's usability and functionality work effectively towards overall satisfaction of the system, making explanations, visualizations or advanced interactivity redundant. Obviously, Spotify's abilities for helping to find new music are limited, because of very simple recommendation facility, but it can be used as an example of the argument that user trust and satisfaction can be promoted by diverse means depending on the different users' various needs and desires.

The next step of our research is to conduct an empirical user evaluation of the importance and functions of different UI elements in music recommenders. We are looking for feasible scales of measurement that are drawn from user evaluation of the goals for UI elements in recommenders. User evaluation could be done with modified music recommender UIs where users are given tasks and comparing e.g. how much taking away a UI feature such as an explanation effects to the time the task is completed. It would also be interesting to explore how different goals can be promoted by combining various UI elements, and by assigning unconventional roles for UI elements, e.g. creating visualizations that would reveal the logic behind a recommendation and at the same time give a user a tool to scrutinize.

REFERENCES

[1] Adomavicius G., Tuzhilin, A. 2005. 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*, 17(6), 734-749.

[2] Buchanan, B., Shortcliffe, E. 1984. *Rule-Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison Wesley Publishing Company.

[3] Chen, L., P. Pu. 2002. Trust building in recommender agents. In Proc. of *International Workshop on Web Personalisation, Recommender Systems and Intelligent User Interfaces '02*.

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

[5] Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J, Tauber, E.R. 2003. How do users evaluate the credibility of web sites? In Proc. of *Designing for User Experiences '03.* Pages 1-15.

[6] Herlocker J. L., Konstan, J. A. 2000. Explaining collaborative filtering recommendations. Proc. of *Computer Supported Collaborative Work '00*. Pages 241-250.

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

[8] Hill, W., Stead, L., Rosenstein, M., Furnas, G. 1995. "Recommending and Evaluating Choices in a Virtual Community of Use", Proc. of *Conference on Human Factors in Computing Systems '05*.

[9] iTunes, http://www.apple.com/itunes.

[10] Johnson, J. & Johnson, P. 1993. Explanation facilities and interactive systems. In Proceedings of *Intelligent User Interfaces '93*. (159-166).

[11] Koenemann, J., Belkin, N. 1996. A case for interaction: A study of interactive information retrieval behavior and effectiveness. In Proc. *of Conference on Human Factors in Computing Systems '96*, ACM Press, NY.

[12] Last.fm, http://www.last.fm.

[13] Nielsen, J. and R. Molich. Heuristic evaluation of user interfaces. In Proc. of *Conference on Human Factors in Computing Systems '90.*

[14] Pu, P., Chen, L. 2006. Trust building with explanation interfaces. In Proc. of *Intelligent User Interfaces '06*, pages 93-100.

[15] Sinha, R., Swearingen, K. 2002. The role of transparency in recommender systems. In Proc. of *Conference on Human Factors in Computing Systems '02*.

[16] Tintarev, N., Masthoff, J. 2007. Survey of explanations in recommender systems. In Proc. of *International Workshop on Web Personalisation, Recommender Systems and Intelligent User Interfaces '07*.