Explanation Interfaces
in Recommender Systems

Ning Xu

s0542814

Information and Communication Technology in Business

Supervisor: Drs. Peter van der Putten

Dr. Walter Kosters

Leiden Institute of Advanced Computer Science

Leiden University

August 2007

Abstract

In this thesis, we try to address several aspects of the explanation interface in a recommender system and create a design framework that can help guide future designers to successfully implement a user-friendly and highly effective explanation interface for a recommender system. This framework will also play an indicating role to assist us in analyzing existing explanation interfaces of recommender systems.

By carrying out a review on previous research papers and a survey on existing commercial recommender systems, we aim to find out the academic and commercial best practices and hence formulate a set of draft design principles. Furthermore, by following our proposed analytical framework, we try to build a prototype system and perform two experiments to justify our research hypotheses and also test the usability of different explanation related functions and features. Finally, a survey will be used to collect users' degrees of satisfaction towards these explanation related functions and their visualization techniques.

After a full analysis of data collected from our experiments and survey, an explanatory statement of the proposed design framework and a refined set of design principles are concluded as the final academic outcome of this thesis. The major results can be found at the end of Chapter 4 and a summarized short conclusion is presented in Chapter 5.

Keywords: explanation interface, recommender system, presentation, explanation, interaction

Acknowledgement

First, I would like to express my thanks to Drs. Peter van der Putten for his invaluable guidance and advices during the process of my final research. Without his help, I would not have had the opportunity of finishing this thesis on time.

I would also like to thank my second supervisor Dr. Walter Kosters who gave me valuable comments to improve my thesis.

Many thanks to Prof. H.P. Borgman who granted me the admission to come study here in the first place.

Thanks to our coordinator Ms. Esther van Eijk who has been giving great support during the past two years to our class as a whole.

Also a million thanks to my family for their understanding and continuous support during my 20-year long study.

Thanks to all my classmates from the ICT in Business 2005 Class. I had a lot of fun learning everything with you guys and gals!

Table of Contents

Chapter 1: Introduction

During the past decades, the information technology has been developing at an unprecedented pace and influenced almost every aspect of our way of life. Especially after the introduction of Internet and its increasing popularity in the 21st century, the quantity of information has been increasing exponentially. It is much more convenient nowadays for people to get access to new information through news portals like CNN.com and BBC.co.uk. And search engine giants such as Yahoo and Google are also providing efficient searching services for both laymen and expert users alike to quickly retrieve their needed information from Internet. Furthermore, Internet also brings life into a new domain of business – e-commerce. Old school brick-and-mortar stores can now be replaced by online e-shops. And with very little effort, customers can explore these online business presences and make comparisons and purchases by simply clicking their mouse buttons. E-business on the web enables retailers to have the power to reach virtually every customer who would like to surf on the Internet. And the population of so called “netizens” around the world, just like the fast growing information volume, is also increasing at a stunning speed. We have already seen many success stories about major e-commerce websites such as Amazon.com and Walmart.com. Besides its B2C (Business to Consumer) aspect, Internet also equips individuals with the capability to reach out to the same amount of potential customers as any large and professional online vendor. And that’s also why C2C (Consumer to Consumer) auctioning sites like eBay.com and Chinese Taobao.com are also among the most-visited in the cyber space.

Sufficient amount of information is good but overloaded information can become inconvenient for a user to retrieve and digest. The fast-expanding Internet hosts too much information and has caused a well-know problem addressed by computer scientists as “information overload”, which means that too much information to choose from leads to the feeling of losing control and anxiety (Edmunds et al., 2000). Like searching in Google by a single keyword to reach a target webpage may sometimes seems like finding a needle in a haystack; online shoppers can also meet with the same dilemma when facing too many purchase choices on an e-commerce website. And that is where the recommender system comes to the rescue.

A recommender system is a system that facilitates a customer to make the right purchasing decisions in an information overload environment. It has become a widely-implemented technology to assist online users to select their interested products from e-commerce sites. These products include books, music CDs, movie DVDs, flight tickets, hotel reservations etc. Successful e-commerce sites like Amazon.com and IMDb.com are leading recommender system implementers and beneficiaries as well. Enhanced sales results from these e-commerce sites by utilizing recommender systems have proved that a recommender system is effective in guiding shoppers to buy the products that fulfill their needs.

This thesis aims at addressing the overall “information overload” problem by adapting a recommender system in the perspective of its explanation interface, which conveys the recommender system’s internal logic and explanations effectively to the users, and also engages the users to more interactions. Different from many traditional researches focusing on the technical details of recommender algorithms, our focal point will be the explanation interface between the recommender system and its users. At the end of this thesis, we will promote a design framework for building the explanation interface of a collaborative filtering based recommender system after summarizing all the research efforts put into this research project.

But first things first, we are going to explore and diagnose existing problem situations in the next section before we expanding our research into a wider perspective. As said by the ancient Chinese general Sun Tzu, “if you know the enemy and know yourself, you need not fear the result of a hundred battles”, let us get to know our “enemy” here.

1.1 Problem Description

The most popular technique employed in the modern recommender system is collaborative filtering (CF) (Resnick et al., 1997; Resnick et al., 1994) which focuses on identifying the active user with similar interests as “neighbors” and recommend each active user with other like-minded neighbors’ interested items. However, most of recommender systems operate in a “black box” mode and users do not often have reasonable explanations from why the recommended items are presented for them especially when some recommendations are unreasonably far away from their mind. Jerry Zaslow of the Wall Street Journal wrote an article about recommender systems entitled "If TiVo thinks you are gay, here's how to set it straight - Amazon.com knows you, too, based on what you buy; why all the cartoons?" (Zaslow, 2002), in which Zaslow described the phenomenon that sometimes recommender systems can go extremely wrong on their recommendations and even under normal functioning they sometimes still make users feel hard to understand their choices.

As mentioned above, recommender system can go wrong and make stupid mistakes. But under a “black box” model, users have no chance of understanding why theses recommendations are presented and how these mistakes are made. (Tintarev et al., 2007) addresses this problem as loss of scrutability for users. Previous research (Buchanan et al., 1984) on expert system in the 1980s has already shown that explanation and transparency are indispensable parts of the expert system especially when an expert’s decision guide is made, follow-up explanation and justifications are necessary. Johnson et al. stated in their findings (Johnson et al., 1993) that based upon previous studies on expert systems, explanations play a crucial role in the interaction between users and complex systems. And Koenemann et al. pointed out in their paper (Koenemann et al., 1996) that better interactivity and larger visibility for relevance feedback helped search performance and users' satisfaction with information retrieval system. While in 2000, Herlocker et al. suggested that collaborative filtering recommender systems are not feasible for high-risk product or service recommendation because of their lack of explanation functions (Herlocker et al., 2000). Later, more research efforts towards scrutability (Kay et al., 2006) in personalized interface and trust (Pu et al., 2006) in recommender systems and the recommender and human interaction (McNee, 2006) began to surface.

However, the question of how to build transparency still remains unanswered for recommender system designers when they try to incorporate an explanation interface into their used-to-be “black box” approach. But among different types of recommendation technologies (Burke, 2002), e.g., collaborative filtering, content-based and demographic etc, how to maximize the best effect of recommendation explanation becomes not as clear as recommendation algorithms themselves. Furthermore, by “effect” in the aforementioned sentence, we are meaning from both the recommender system’s perspective and the user’s as well, while from the user side, a better expression would be “trust” (Pu et al., 2006).

Even due to the recent surge in the amount of research papers published about the user interaction with recommender systems and their explanation interfaces (Bonhard, 2004; McNee et al., 2006; Tintarev et al., 2007); a complete and thorough theoretical model of how to design an explanation interface and its supporting techniques is still missing in the academia. Therefore, based upon a review of existing literature and best practices in the modern e-commerce domain and a well-crafted prototype experiment, how to build a design framework for the recommender system’s explanation interface has become the main topic of this research project.

1.2 Research Objective

To solve the problem mentioned in the last section, we formalize our research objective here as follows and this objective is threefold:

1. We aim to put forward a framework for explanation interface design based on a full survey of existing literature and commercial systems' explanation functions to guide future development of explanation functions for recommender systems.

2. Based upon the experiment results collected from a hybrid recommender system prototype and post-experiment survey feedbacks, we will provide insight from both promotion and satisfaction (Bilgic, 2005) perspectives into the added value of an explanation interface for a recommender system by clarifying to what extent user acceptance of recommendation and trust of the system can be enhanced.

3. We will also analyze which explanation visualization techniques are preferred by actual users and how product content and user demographic data can be combined with a collaborative filtering recommender engine to achieve a better visualization effect for explanation.

The first aspect of our research objective will be achieved partially by a full review of research publications on the recommender system's explanation interface. Previous research papers have already shed some light on the design principles, but no single view on a systematic design framework has been brought up yet. For the second and third aspects, they will be mainly investigated in an empirical way by experiments performed on a collaborative filtering-based hybrid recommender system prototype designed and implemented by the author himself. The user experiences of two groups of candidates (explanation "haves" and "have-nots") will be tested and the corresponding data will be collected and analyzed. Furthermore, post-experiment survey provides another important source for us to collect.

1.3 Research Framework

All the research objectives formulated in the previous chapter can be transformed into a research framework in which the following research steps to fulfill these research objectives will be visualized. Following the steps, the objectives will be successfully achieved on time.

1. The first step in this research project is to gain a good understanding of the theoretical concepts and foundations that will be used. For example, the two closely related "direct marketing" and "Internet marketing" theories are definitely playing a guiding role in our explanation interface design.

2. Secondly, a review on existing recommendation technologies and mainstream recommender

system's explanation functions will be carried out. Each recommendation technology has its unique core algorithm and data structure while each commercial recommender system also has its own unique way of explaining their recommendations. To achieve a valid experiment on designing the explanation interface, a carefully selected and well-crafted recommender system platform needs to be chosen beforehand. This step's outcome will become the input of the experiment phase of this research.

3. The third step will be analyzing existing explanation interfaces among major e-commerce websites such as Amazon.com and Netflix.com to have an overview of the current application of recommender system's explanation functions. And this overview will support our hypothesis that leading e-commerce sites have not fully adopted explanation to their applications yet and validate our research's necessity.

4. The forth step will be involving the analysis of well-researched explanation interfaces in existing popular and research literatures and to identify some best practices. This analysis will be concluded with some selected explanation interface design principles and directives. Based on these previous research results, we will come up with a better systematic design framework and put it to test in the experiment. So this step's output will also become the input in the experiment phase.

5. The fifth step will be the experiment phase itself. First and foremost, a requirement analysis session will be performed to collect users' perceptions and needs for an explanation interface. And then, under the guidance from learned best practices and design principles, a full functioning explanation interface will be built upon a fine-tuned recommender system prototype. Experiment participants will be invited to actually interact with this system to validate or disagree with our hypotheses. For example, whether a recommender system with explanation interface is really able to generate more sales revenue than those without and how big the margin is are among the questions to be answered.

6. Then, a post-experiment survey will be carried out to let all the participants to fill out an online questionnaire to express their user experiences and views about our recommender system and its explanation interface. In this part, participants can rate each explanation feature and also give text feedbacks on what their overall opinions on our prototype system's explanation interface.

7. After all the above 6 phases are finished, collected experiment data and survey data will be put into our careful analysis scope and a final conclusion with some discussion will be drawn to conclude the whole thesis.

1.4 Research Question

Now that the boundaries of this research have been laid out, the research question can be formulated as follows:

How can we design an effective explanation interface for a recommender system?

To fully answer this research question, 5 sub questions need to be answered first.

1. *Which design principles are the most effective based on previous best practices in published research papers and current e-commerce practices?*
   Previous researchers have already explored some aspects of the design principles of a recommender system's explanation interface. By a thorough literature review in Chapter 2, we will summarize a draft design framework to guide our own prototype design and experiment. Furthermore, practical user oriented e-commerce websites may also demonstrate some best practices which have not been covered by preceding research. Therefore, a survey on several major e-commerce websites with recommender functions is also needed.

2. *What are the user requirements for the explanation interface of a recommender system?*
   Before the actual development of our experiment prototype system, an interview session will be carried out to collect the initial user requirements and perceptions for the explanation interface of a recommenders system. We will take users' initial thoughts into our design consideration.

3. *How much can an explanation interface enhance recommender system's promotion effect in terms of website online sales revenues?*
   We will provide some online purchase functions to enable two groups of users ("haves" and "havenots") to actually spend some virtual money to buy the items in our experiment system. In order to create a ground for further discussion, the "sales volume" from the "haves" group will be compared with the controlled "havenots" group's results.

4. *How can the limited content data and demographic data be utilized to contribute to a better recommendation algorithm and more convincing explanation visualization?*

Because our prototype system is based upon the "movielens" dataset (MovieLens, 2007) with only limited content and demographic data, and according to some previous research, these data can play a significant role in conveying the explanation information to the active user. Here in this research project, we'll try to incorporate these limited data to our explanation interface design and see how effectively we can utilize them.

5. *What are the preferred visualization techniques that can retain customers to an e-commerce site and improve its trust?*

   A picture might not be always worth a thousands words, but is definitely more convincing than pure text. Different visualization techniques will also be assessed in this research to rank them in different scenarios to test their effect in promoting items and satisfying users.

1.5 Research Materials

Source	**Explanation**	**Access**
For sub question 1:		
Research papers and literature	Research on the explanation interface of recommender systems	Content analysis
For sub question 2:		
An interactive design session	During the preparation phase of the experiment, an interactive design session is carried out to collect users' requirements and perceptions about recommender system's explanation interface	Observation and data collection
For sub question 3:		
Difference on virtual sales revenues based on experiment on "haves" and "have-nots" users	On the experimental prototype system, a virtual online shopping function will be provided to assess users' willingness to purchase; the sales figure will demonstrate the explanation interface's promotion effect	Observation and data collection
For sub question 4 and 5:		
Post-experiment questionnaire on users feedbacks about different explanation visualization techniques	Post-experiment survey will focus on users' personal feelings (both positive and negative ratings) against	

	several visualization techniques built upon the CF algorithm itself, content data and demographic data	

Table 1.1 Research materials for research sub questions

1.6 Research Strategy

This research project on one hand aims to review and validate existing explanation interface design principles and best practices, while on the other hand it also focuses to discover user-preferred explanation interface design methods and visualization techniques. Therefore, an empirical experiment based on a prototype recommender system is necessary and essential to the collection of quantitative virtual sales revenue. Prior to the experiment, a user requirement analysis will also be performed in the form of an interview. Moreover, a post-experiment survey for collecting qualitative user feedbacks is also a required step. Furthermore, a literature review definitely involves a desk research strategy as well.

To sum up, four research strategies will be adopted and intertwined throughout our research project:

1. Experiment
2. Interview
3. Survey
4. Desk research

1.7 Academic Relevance

Although recommender systems and collaborative filtering technology have been in the research focus for several years, yet the “white box” approach with the support from explanation interfaces has only been attracting researchers’ attention in the recent couple of years. And there has not been a thorough and complete study on how to design explanation interfaces for a recommender system in a general purpose way. Also, limited existing research approaches are not systematic and consistent enough to guide a system administrator to implement and develop a recommender system with convincing explanation, effective visualization techniques and professional design principles.

Upon finishing this thesis, the main contribution to the academia is that we will successfully draw a blueprint that a system developer can consult to build explanation interfaces. E-commerce websites and e-business operators with the intention of commercializing their recommender systems' explanation functions can also find empirical study proof to validate their investment decisions. Also, the design framework and principles put forward in this thesis will guarantee a better return on investment than purely starting up from trial and error.

1.8 Thesis Structure

The 1st chapter is the introduction of the research background, research topic in general. After the problem has been identified, we also formulate our research objective and questions here too. And the last part of this warm-up section is our claim of this research's academic relevance.

The 2nd chapter will present a review of existing literature about both recommender systems and their explanation interfaces. By such a coverage, readers will develop a general overview of the current research depth in this explanation interface design field. Furthermore, well-proved effective design principles and best practices will also be transformed and integrated into our design framework for further testing.

In the 3rd chapter, we will carry out a survey on existing major e-commerce websites to see how they are implementing the explanation functions for their recommender systems. Our hypothesis is that in the business domain, the explanation interface has not yet been well-discovered and well-recognized and the status quo will validate our suggestion that further investment into this direction shall be considered by real-world e-business operating companies.

By summarizing our review on literature and the survey on existing e-commerce websites, we will bring forward a design framework in the 4th chapter and carry on some testing on our prototype recommender system. Also the analysis on the collected date will be shown to the readers to not only test our hypotheses but also to validate our design framework.

Finally, in the 5th chapter, our design framework on the explanation interface will be finalized with a conclusion drawn from the previous four chapters' efforts, and further discussion on this research topic will also play a roll in guiding future research attempts following our work here.

In sum, the major results of this thesis can be found in the last section of the 4th chapter and a conclusion is drawn in the first section of Chapter 5.

Chapter 2: Related Work and Literature Review

In this chapter, we will present some of the most important concepts about recommender system and review existing research work about a recommender system in general and its explanation interface in much more detail. By doing so, a sound theoretical foundation will be laid out to base all our future research efforts upon. In the following scetions, we will respectively address the concept of recommender system and some major recommender techniques; theories on marketing, information visualization and human-recommender interaction; recommender related privacy issues; and finally we will present a draft design framework for the explanation interface.

2.1 Recommender System

To put it in a very simple and intuitive way, a recommender system is one that gives suggestions and recommendations to users when they are making a decision while facing different choices. The most common scenario about the application of a recommender system is on an e-commerce website and during a live user shopping session, the site provides a list of recommended merchandise to the active user with the help of the underlying recommender system. But formal definitions can be more precise and appropriate for this chapter here.

Resnick and Varian (1997) describe a recommender system as a system which can acquire users' opinions about different items and also use these opinions to direct users to those items that might be interesting to them. Herlocker (2000) gives a much more concise definition that a recommender system is one that predicts what items a user might find interesting or suitable to his or her needs. As can be seen from the previous two definitions, Resnick and Varian's approach is more related to our real-world recommendation concept while Herlocker's description emphasizes more the prediction side of the recommender system. Later, Burke (2002) put forward his definition that a recommender system is any system that can produce individualized recommendations and have the ability to guide users in a personalized manner to find interesting information items in a large space of possible options. Obviously, Burke's definition adds more new elements like individualization and personalization into our old perception of the recommender system. Also, as can been seen from later chapters, individualization and personalization also play vital and guiding roles in the explanation interface's design framework.

2.2 Recommendation Techniques

Several classifications (Resnick et al., 1997; Schafer et al., 1999; Schafer et al., 2001) of recommendation techniques have already been published. For the interest of discussion in this research project, we only give some insight into the three most recent classification approaches and use them as a corner stone to support our next research step.

Robin Burke (2002) presents a very thorough classification of existing recommendation techniques by identifying each application's background data, input data and its inner algorithm and then comes up with the following five types of recommendation techniques:

1. Collaborative filtering
2. Content-based
3. Demographic
4. Utility-based
5. Knowledge-based

Mark van Setten (2005) brings forward another classification for recommendation techniques, i.e.,, a social-based technique and an information-based technique. A complete taxonomy is shown below:

1. Social-based recommendation techniques:
    a) User-user collaborative filtering
    b) Item-item collaborative filtering
    c) Stereotypes and demographics
    d) Popularity
    e) Average
2. Information-based techniques:
    a) Information filtering
    b) Case-based reasoning (CBR)
    c) Attribute-based prediction techniques

Riedl et al. (2002), in his book "Word of Mouse: the Marketing Power of Collaborative Filtering", also presents "a complete list of recommenders" as follows:

1. Manual recommender
2. Searchable database
3. Segmentation technique

4. Statistical summarization technique
5. Social navigation technique
6. Custom proprietary recommender system
7. Machine-learning recommendation
8. Information-filtering techniques
9. Collaborative filtering
10. Combination recommenders

The research angle of Burke's (2002) classification is the underlying data used by a recommender system. Mark van Setten (2005) formulates the two big categories from both social and information aspects prior to breaking them down into more detailed sub-categories. However, Reidl et al. (2002) simply takes a very commercial approach in their taxonomy of recommender system. Of course, quite some overlap can be easily found from these classifications. And why we deliberately present these three categorizations as above is because, firstly Burke's (2002) classification is among the most well-accepted and thorough classifications up till now; secondly, the classification from van Setten (2005) and Riedel (2002) both have some suitable elements to guide a design framework for developing an explanation interface and also a good yardstick for the later survey on the different explanation interfaces of several commercial recommender systems. Later in this thesis, we will see that many design elements for an explanation interface corresponds with the aforementioned techniques such as stereotypes and demographics, popularity and average etc. In the next several sections, we will pick some of the above recommendation techniques and have a deeper look into them.

2.2.1 User-user Collaborative Filtering

The central idea of user-user collaborative filtering is that users who have similar rating records and show the same interest in the same items will probably have similar tastes (Resnick et al., 1994; Shardanand et al., 1995). If the recommender system knows about this kind of similar rating patterns, it can predict whether a user will find an unseen item interesting to his or her needs. Usually a typical user-user collaborative filtering process comprises of three steps (Herlocker, 2000):

1. Similarity calculation

   The system calculates the similarities between the active user and the other users who have rated the same items. This calculation step might be the most well-researched area in the field of user-user collaborative filtering recommender system. Mainstream algorithms like Cosine similarity, distance measurers and Pearson correlation (Herlocker, 1999) and their derivatives are among the most popular applications. Since this research project does not do much with the algorithmic part of a recommender system, we do not go into more detail in this aspect.

2. Neighborhood formation

   A subset of similar users will be selected to form the neighborhood of the active user. Most of the time, the most similar users calculated from the 1st step will be included in this neighborhood. Moreover, these neighbors' rating and purchasing records will be used to find out the most interesting items for the active user.

3. Prediction or recommendation generation

   The final step's output can be twofold. One is that the system will use the active user's neighborhood to generate a list of recommended items which usually takes the form of "top-N" recommendations in a real-world ecommerce website; the other is that the system can use the similarity information and neighbors' rating data to predict the active user's rating for specific items. Herlocker et al. in their survey (Herlocker et al., 2002) presents several algorithmic choices for this step as well.

The most compelling benefit of user-user collaborative filtering is its domain independent character. Because only rating or purchasing data are needed to calculate the similarity between users, any type of products can be associated pretty quickly with users' tastes. Furthermore, our later experiment will be performed on a collaborative filtering recommendation engine where most of our hypotheses will be tested. Therefore, we give more focus to this recommendation technique to back up our further discussion.

2.2.2 Item-item Collaborative Filtering

Item-item collaborative filtering is similar with user-user collaborative filtering and can be viewed as the same approach but from the item's angle. The idea behind this technique also sounds familiar now: items that have been rated in the same way are likely to share some similar attribute or features, so people who like one of them will probably like the others that are similarly rated (Herlocker, 2001). In the e-commerce domain, Amazon.com might be the most well-known implementer and investor of the item-item collaborative filtering technique (Linden et al., 2003). And not surprisingly, Amazon.com also claims large gains through its well-functioning and highly effective item-item recommendation technologies (Linden et al., 2003).

Due to the fact that it also requires only the rating data, item-item collaborative filtering shares the same strength of domain independence with user-user collaborative filtering.

2.2.3 Segmentation Technique

Stereotyping and demographics are the most important two elements that form the basis for the segmentation technique. Rich (1998) gives a good overview on user stereotyping techniques by which people model different users into different stereotypes containing a group of unique characteristics, e.g., behaviors, hobbies, interests etc. In a typical stereotype-based recommender system, users are grouped into stereotypes based on either their explicit or implicit online activities. After collecting sufficient characteristical data, usually demographic data such as gender, occupation and education, the recommender system can formulate different recommended items for each stereotype. This process clearly requires some expert knowledge and manual fine-tuning. In his TV program recommender system, Ardissono et al. (2004) forms different stereotypes for TV viewers and recommends programs accordingly, and his main judgment criteria are users' age groups and explicit interests in different genres.

Usually, stereotype-based and demographics-based recommendation techniques don't perform totally independently. For example, the TV recommender system mentioned in (Ardissono et al., 2004) also employs many features from knowledge-based recommendation techniques. Stereotype-based and demographics-based recommendation techniques are very easy to be combined with other techniques, e.g., collaborative filtering, which will also be the case in our experiment. In addition to this hybrid combination, demographic information can also play an important role in conveying user similarity in a recommender system's explanation interface. As argued by Herlocker (2000), some of his experiment participants explicitly express the willingness to view their neighbors' demographic backgrounds and also like to tag them in a kind of social networking manner. Therefore, we can conclude that people are inclined to be convinced with knowing recommender's demographic data and shape up a sense of common community. And if a recommender system's user interface can achieve such an effect, its general promotion capability will be surely boosted. We leave an open thread on this perspective and will come back later in the design framework discussion part.

2.2.4 Statistical Summarization

People have a natural intention to trust statistical information. And we believe that the most popular statistical summarization techniques, i.e., popularity and average, have compelling power in explaining to users the current system's authority. This will be one of our hypotheses to be tested later on.

1. Popularity

The idea behind a popularity-based recommendation technique is pretty simple and straight-forward. Many e-commerce websites present some top selling items on their entry page or title banner. Herlocker (2000) puts forward a recommender system based on the average ratings calculated from all users to select the most welcome products. As a matter of fact, people are inclined to be attracted to popular items when visiting an e-commerce website. And a popularity-based technique might come much earlier than more sophisticated collaborative filtering techniques. In later chapters, we will see that an item's popularity information can be a very persuasive element for an explanation interface.

2. Average

Using the average, alongside with popularity, may be another simple but practical recommendation technique. The average rating of any item, to some extent, reflects its inner quality and its popularity among users as well. For example, IMDb.com provides for each movie in its database the average rating from all its registered users to convey a relative general positive or negative opinion about a specific movie. Also, here we put a little attention to another type of average, which is the average rating from all neighbors in a user-user collaborative filtering recommender system; the presentation of neighbors' average rating is also believed to have a strong persuasive effect (Herlocker et al., 2000).

2.2.5 Social Navigation Technique

According to Riedl et al. (2002), social navigation means "making the behavior of other customers visible". This idea agrees with the finding from Herlocker et al. (2000) that recommender system users would like to know more about their like-minded peers and some experiment participants even expressed the willingness to tag a neighbor to see his/her future rating records for further recommendation guidance (Herlocker et al., 2000). Beside the further social networking combination approach suggested in (Bonhard, 2004), a social navigation function in the explanation interface will make some positive contributions in familiarizing the active users with his or her potential recommenders (neighbors). So we also suggest that a social navigation technique shall be employed as a new design principle for a recommender system's explanation interface.

2.2.6 Information Filtering

Information filtering technique stems from the domain of text retrieval. And the most widely used information filtering algorithm bases itself on the well-known term-frequency-inverse-frequency (td-idf) algorithm. Pazzini et al. (1997) has already adopted several information filtering algorithms, such as PEBLS, decision trees etc, into the recommender system research area.

The reason we present information filtering here as one of the recommendation techniques is that it has quite a strong relation with information visualization concept which will be covered later in this chapter, as any visualization process involves some sort of information filtering beforehand.

2.2.7 Hybrid Recommender System

All the above mentioned techniques can also be combined together to form a hybrid recommender system, because usually a single technique may have its own shortcomings while combining different techniques together, their weaknesses can be neutralized and performance can be greatly enhanced as well (Burke, 2002).

In effect, major commercial recommender systems such as Amazon.com already employ more than one recommendation technique to generate more accurate predictions and also to let users have much more purchase choices to make during their online session (Linden et al., 2003).

2.3 Theories on Marketing

Since the recommender system is a vital piece of equipment for modern online vendors to implement their marketing efforts, various theories on marketing are playing different roles in designing a recommender system as a whole and especially its explanation interface. In addition, as far as marketing is concerned, the importance of the recommender system's explanation function and interface has been raised to an unprecedented level, because without sufficient explanation, people are not likely to be persuaded to buy a product, let alone to start a viral "word-of-mouth" about it.

2.3.1 Mass Customization and Internet Marketing

"If I have 3 million customers on the Web, I should have 3 million stores on the Web."

-- Jeff Bezos, CEO of Amazon.com

In a physical world, it is impossible for everyone owning a specially tailored brick-and-mortar store right in his or her neighborhood, but in the online world, businesses are marching towards developing “multiple products that meet multiple needs of multiple consumers” (Schafer et al., 2001). In the book “Mass Customization” (Pine, 1993), Joe Pine pointed out that modern companies need to shift their focus from “mass production” (standardized products with long product life for various market segments) to “mass customization” (personalized products and services tailored toward different market segments or even individuals). And one way to achieve “mass customization” is through recommender systems. Although for low-risk commodities such as DVDs and books, it is hard to say these products can be further customized individually, but a user’s taste usually involves collecting of different products. Each collection of non-customizable products can be customized again and tailored to individual user’s needs accordingly. A recommender system is unable to literally customize its products but can customize the collection of interested items for a particular user. Furthermore, Pine et al. (1999) extends his original scope of mass customizing the products to the customization of user experiences. This idea definitely blows some new breeze into the design methodology of recommender system, especially concerning the user-recommender interaction aspect which highly relates to our discussion on the explanation interface here in this thesis. Nowadays, people not only look into a product or service for higher quality, but also for satisfactory user experience. Recommender systems, under a “black box” approach, seldom present sufficient and satisfactory explanation to their users (Bilgic et al., 2005). Bilgic et al. (2005) brings forward two aspects to focus recommender system’s explanation capability, namely “promotion” and “satisfaction”. We also believe these two important measurements shall no doubt be included in our proposed design principles, but we do not totally agree with the authors’ experiment results and will come up with our own evidence from our prototype system in later chapters.

Internet marketing, “the practice of using all facets of internet advertising to generate a response from your audience” (Wikipedia, 2007), is another closely related concept that has always been accompanying the development and evolution of recommender system. According to (Wikipedia, 2007), Internet marketing associates with three major business models which are the traditional business-to-business (B2B) model, business-to-consumer (B2C) model and the more recent peer-to-peer (P2P) model. A recommender system has been helping enhancing the Internet marketing return on investment of all three types of business for years, but what kind of role is being played by the recommender system’s explanation interface has not been well-studied yet.

Therefore, in later chapters, we will present our own survey on existing major online vendors' websites and see how they implement their recommender system's explanation functions to enhance their Internet marketing results.

2.3.2 Word-of-Mouth Marketing and Viral Marketing

World-of-mouth marketing, also called buzz marketing, is another concept currently playing a guiding role in recommender system's design. The aforementioned user-user collaborative filtering technique can be considered as a computerized word-of-mouth marketing practice, because the underlying idea of user-user collaborative filtering is a carbon copy of assessing the "buzz" (ratings or purchase records from a set of similar neighbors) for a particular product and then recommend it to the active user.

If current recommender systems are playing the role of virtual sales people dedicated to collecting the word-of-mouth about a particular product, the next step for a recommender system's development is to engage users to spread the buzz and thus creates a marketing "virus" for a particular product favored by a virtual community of like-minded people. One hot discussed future research direction of the recommender system is its combination with social networking functions to elicit users' participation in rating, reviewing, discussing and purchasing (Bonhard, 2004) and Riedl et al. (2002) also suggests that hosting an "online community can lead to viral-marketing and more recommender opportunities". Therefore, two aspects concerning the design of the explanation interface need to be fully considered. The first aspect is how to use an explanation interface to encourage users to participate in community activities e.g., writing a review, giving a rating to another reviewer and joining a discussion. And the other aspect will be from the system's viewpoint about how community data can be used to provide more compelling evidence for an explanation interface to justify recommendations or satisfy users' requests.

2.3.3 The Long Tail Theory

Anderson (2004) coined the term "Long Tail" in one article with the same name in the *Wired* magazine. It basically means that online vendors, unlike bricks-and-mortar counterparts in the physical world, has unlimited "shelf space", which enables them to "display" more products to customers. Also through effective recommendations with inter-links among all these products, customers can quickly explore other similar or related products when they buy one of them. Eventually, a phenomenon that many more less popular products that are sold in small quantities altogether can outweigh those bestsellers and blockbusters sold in large volumes will take shape.

It can be illustrated by the graph from Figure 2.1. The x-axis can be seen as different types of products with a declining popularity while the y-axis can be seen as the amount of products being sold. As the long tail goes on, the area of the yellow part of the graph will be bigger than the green part.

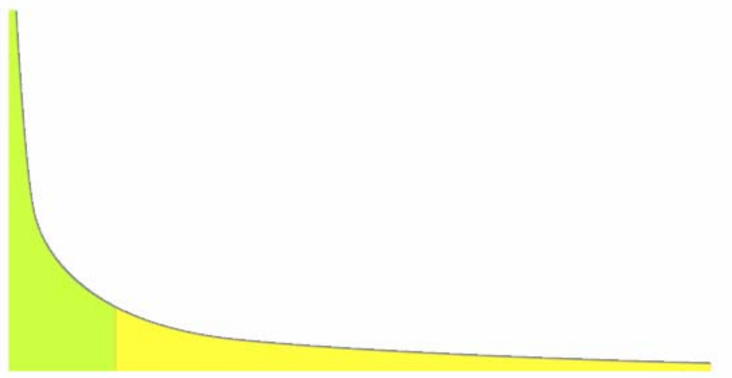


Figure 2.1 The Long Tail

In this case, the role that a recommender system is playing is to help customers to recognize their needs which are beyond what those popular products can fulfill. Unpopular products might be in small demand, but they may also be interesting to some specific small user groups. And one efficient means that an online vendor can recognize these individuals with specific needs is through a recommender system.

2.4 The Reasons Why to Explain

Historically, recommender system research has only been focusing on enhancing the accuracy of recommendations, which researchers used to believe corresponds with higher user satisfaction and better promotion effect. In (Herlocker et al., 1999), two types of accuracy metrics are discussed i.e., statistical accuracy and decision-support accuracy. Popular measures such as Mean Absolute Error (MAE) (Shardanand et al., 1995) and Coverage (Herlocker et al., 1999) belong to the former group and so do Recall and Precision (Sarwar et al., 2000). Herlocker et al. (1999) only lists ROC sensitivity as one type of decision-support accuracy metric, however, we believe when concerning decision-support accuracy, design elements such as promotion and satisfaction are very difficult to quantify unless with rich user interaction and feedbacks. And this is also why we take a more empirical approach in our experiment to measure users' feelings toward different design principles discussed later in this thesis.

Since the human analogue of asking friends and families for recommendations is a totally transparent social process, the traditional “black box” approach of recommender systems seems to be contradictive to our intuitive perception. Preceding research results on expert systems have shown that explanation and justification for experts’ advice are extremely important design guidance (Buchanan et al., 1984). Johnson et al. (1993) pointed out that explanation plays an essential part between human and complex information systems. Also, Koenmann et al. (1996) found out that more visibility and interactivity could significantly enhance search engine’s performance. Furthermore, McNee et al. (2006) argues that after years of fine-tuning and evolution, most well-known recommender algorithms score similarly on various accuracy metrics, so how to enhance the quality of the total human-recommender interaction experience becomes the new way forward. Based on these existing research publications, we can conclude that enabling users to understand the inner logic between their inputs like ratings and reviews and the system’s outputs like recommended item lists allow users to interactively revise their inputs and consequently gaining more satisfactory and accurate outputs. Therefore, we argue that the role of interaction shall also be taken serious consideration when designing a recommender system’s explanation interface.

2.5 Human Recommender Interaction

Human Computer Interaction (HCI) has been a well-researched, complex and interdisciplinary area for ages. It requires the bridging between multiple science fields such as computer science, artificial intelligence, computer graphics, anthropology and psychology as well (Wikipedia, 2007b). Although HCI as a whole has a very solid theoretical backbone, the Human Recommender Interaction (HRI), especially in the explanation interface design perspective, is relatively lacking sufficient theoretical guidance (Haynes, 2001). Johnson et al. (1993) pointed out that there is a serious weakness in the research of computer-generated explanations: a general underlying theory is missing. Haynes (2001) tried to formulate a new theory based on the idea of “design rationale” to shrink the gap between theory and practice mentioned in Johnson et al. (1993).

However, when we look into the specific design of human recommender interaction, these preceding research efforts are becoming insufficient again. Therefore, the work described in this thesis does not draw too much from previous theoretical achievements and maintains quite preliminary and empirical. But some recent publications can still shed light on our research endeavors. McNee et al. (2006) brings forward an analytical model for human-recommender interaction with three supporting pillars namely, recommendation dialogue, recommender personality and end user’s information seeking task. (McNee et al., 2006)’s HRI theory aims at solving an average user’s three major concerns, e.g., trust in the recommender, understanding the recommender and believing that he/she is looking for the right information. This HRI framework definitely refreshes our thoughts on the design of an effective explanation interface.

2.6 Information Visualization

As an old saying goes that a picture is worth a thousand words, human eyes interpret visualized data much faster and more effectively than interpreting plain texts. More academic evidence can be found in (Ware, 2000) human eyes can process many visual cues simultaneously while understanding a chunk of text takes much more cognitive efforts. A fair and generic demonstration is the use of map. The left panel of Figure 2.2 shows the textual description of how you should drive to the destination while the right panel presents a bird's eye view of the route that you should take. Obviously, only reading the textual guidance without the visualized routing on the right hand side will be time-consuming and difficult to understand.

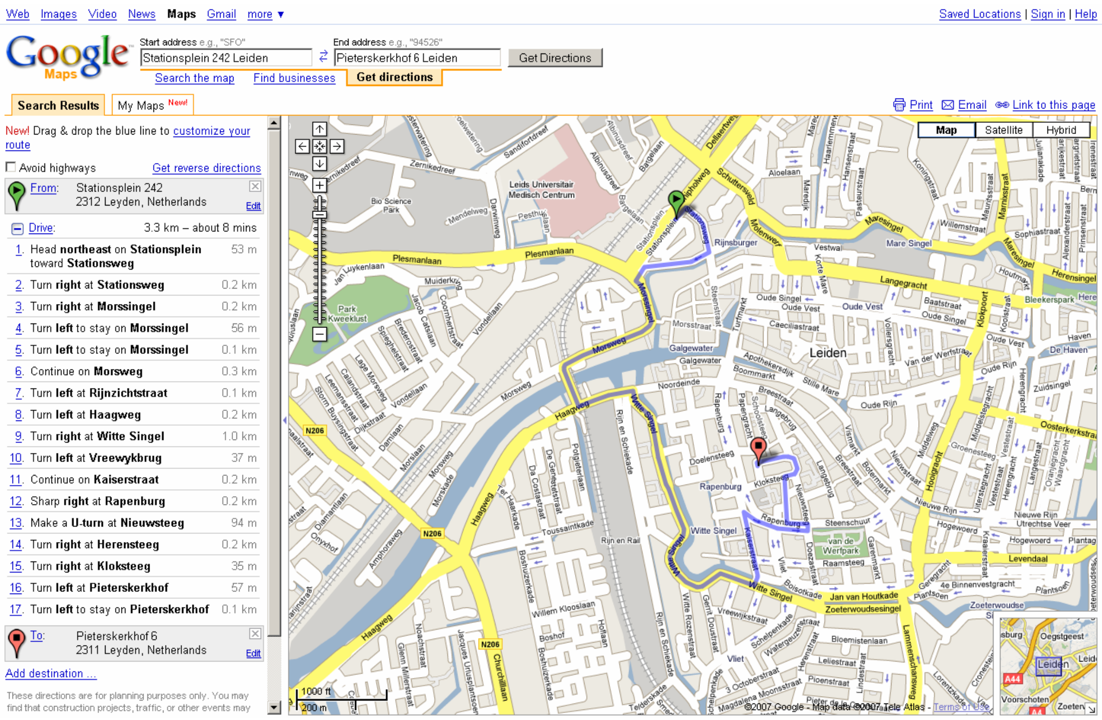


Figure 2.2 Finding the way in Leiden from my dormitory in Stationsplein to LUSM on the Google Maps

An immediate advantage of information visualization is to assist people to see trends and anomalies in data which might otherwise be very difficult to detect. In this thesis, as we are addressing the explanation interface of a recommender system, the visualization techniques are an indispensable part. Xerox's Palo Alto Research Center (PARC), a research leader in the information visualization field, has defined the term of "visualization" as "An external representation that makes it easy to see certain patterns in data". However, Wikipedia (2007c) chooses a different angle to define information visualization: "Any technique for creating images, diagrams or animations to communicate a message". Pack (1998) also argues that what is being communicated in visualization is not only the information itself, but also its structure; users can quickly assimilate and comprehend the information when it is presented in a virtually appropriate relationship with a quick overview. In light of the above definitions and descriptions, we can see that various techniques including tables, images, diagrams and even animations can be used to convey explanations to recommender users. Therefore, in the following paragraph, different visualization techniques which have been addressed in previous research will be discussed.

The research topic about how to use graphs, diagrams, tables and other visualization techniques to enhance the explanation effect has been firstly probed in (Herlocker et al., 2000). Herlocker et al. (2000) test several visualization methods to explain the collaborative filtering based recommendations to about 80 users of the MovieLens site. He concluded that the histogram of neighbors' ratings has the most persuasive effect to an average user. Later in (Bilgic et al., 2005), Bilgic et al. argues that there are three types of explanation, namely keyword style explanation (KSE), neighbor style explanation (NSE) and influence style explanation (ISE). For KSE and ISE, Bilgic et al. (2005) presents them in a table-like visualized manner while on NSE he does not fully agree with (Herlocker et al., 2000) as the grouped three-category (bad, neutral and good) histogram is proved to be more effective in promotion. Also in (Herlocker et al., 2000) he found that text-highlighted statement of the system's performance can also be convincing. Pu et al. (2006) put forward another organization-based explanation visualization structure to highlight different groups of recommendations instead of showing them all in a plain list. In addition, the approached mentioned by Pu et al. (2006) is further confirmed by McCarthy et al. (2004): the conversation-based recommendation with a grouped overview with trade-off information to the top recommended item can intrigue users to give feedbacks and locate for themselves the items that they truly like.

From existing research, we can summarize the following popular visualization techniques that have been or can be implemented on a recommender system:

1. text highlighting
2. categorization in table
3. histogram
4. bar chart
5. rating display in stars

6. color highlighting in various schemes

We will implement the above mentioned visualization techniques in our later experiment and also another popular cloud visualization style tagging (Kaser et al., 2007) technique will be tested for saving users' genre navigation efforts and creating a better user experience.

2.7 Privacy Issue

As every recommender system deals with users' personal data, the privacy issue is definitely a significant concern for both system builders and interface designers. A typical recommender system, before making any recommendations, often requires users to register their basic demographic background information like age, gender, occupation and education, and then it will usually collect users' rating data and monitor their browsing behaviors as well. But do users really like the idea that a recommender system is monitoring their online browsing history? Some privacy-savvy users even prefer shutting off the cookie function in their browsers because they don't feel very comfortable of being traced online. And the worries from this group of "careful" users' are not without reasons. Riedl (2001) states that these sensitive data are prone to be abused and hence recommender systems can become a good means for people like marketers to invade users' privacy.

In this privacy related research field, a new industrial standard which is called the platform for privacy preferences (P3P), has been developed with the objective to assist future Internet users to gain a full control of their personal information over the whole cyberspace by "expressing their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents" (P3P, 2007). However, this P3P project is currently under suspension and lack of momentum for further development.

In this thesis, as we are talking about the explanation interface of recommender systems, users' feedback results on surveys from previous research (Herlocker et al., 2000) suggests that the means of showing liked-minded users' profiles is of better persuasive power. Therefore, how willing ordinary users are to share some of their private demographic data and how to make effective presentations to achieve better explanation results will be later tested in our experiment.

Another aspect is concerned with creating online communities via recommender systems. Amazon.com provides a function called purchase circles which enable users to view the top-selling books in a specific organization or geographic region. And according to (Riedl et al., 2002), this function arose quite some disputes about customer information misuse, but Riedl et al.

(2002) also believe that building online communities through recommender system can significantly enhance an e-commerce site's viral marketing effect. Therefore, in our later chapters, we will put a deeper insight into how explanation interface can facilitate community building without hurting users' privacy rights.

2.7 Draft Design Framework

2.7.1 Beyond Explanation

After the review of related work and previous research publications, we cautiously put forward some of our own thoughts about the design principles and the general design framework for the explanation interface of a recommender system. But before we start our discussion, we want to expand our research scope about the term "explanation interface" a little bit here. The "explanation interface" literally means the interface that explains why the recommender system presents the recommended items or the predicted ratings of particular items. Unfortunately, if we kept our research scope into such a narrow space like this, our entire research project would be of little value. Therefore, the term "explanation interface" is beyond only explanation, but also includes two other important elements as well, namely "presentation" and "interaction".

Presentation

The reason why we put "presentation" into consideration here is because even the most effective explanation requires some presentation styles such as content-based keyword style, community-based influence style and traditional collaborative filtering-based neighborhood style (Bilgic, 2005). Moreover, presentation is where our visualization techniques are put into practice for, and how to present the explanation in a visually appealing or convincing way is one of our research objectives. Also, a successful recommender system not only presents recommendation information, but also presents rich product descriptions to demonstrate product expertise which is believed by (Riedl, 2002) to be one of the most persuasive features that an e-commerce website must possess. Therefore, presentation has a really close relationship with explanation, especially on the promotion side.

Interaction

A vivid example acts as a better illustration for the next step in our discussion. Imagine a user called Mr. Buff visits a movie recommender site – "mymovies.com" and the recommender system on that site presents a list of movie related news and recommendations as follows:

1. 2007 summer blockbuster “Transformers” has been released in the theatres near your home, click to see the locations and ticket discount info.
2. Ziyi Zhang says she learns English from rapper Eminem, but later realizes how rude his lyrics are.
3. Buy the DVD of the movie TMNT (Teenage Mutant Ninja Turtles) from Amazon.co.uk starting from €3.75.

In this scenario, Mr. Buff might have some questions in his mind:

1. How did the recommender decide that I would be interested in “Transformers”?
2. Why is a piece of news about Ziyi Zhang presented to me?
3. What is the plot of the movie “TMNT” and who star the characters?
4. How can I let the recommender system know that I am pleased with the ticket discount information?

As we said before, a traditional “black box” recommender system is probably able to provide the plot description for the 3rd question and will hardly answer any of the remaining three above, but we also believe the next generation recommender system with a well-designed explanation interface will address all of them. For the 1st question, it is because the recommender system’s user model for Mr. Buff shows that he is an animation movie fan and a regular movie-goer to local cinemas. And for the 2nd question, Mr. Buff also highly rated all the movies, such as “Crouching Tiger Hidden Dragon”, “Rush Hour 2” and “House of Flying Daggers”, starring Chinese actress Ziyi Zhang indicating he might favor her as a reason to choose these movies. This too explains why TMNT is recommended: it is a highly appraised animation movie with an average of 3.5 stars from all the neighbors who the system considers to have similar tastes with Mr. Buff, plus that one of TMNT’s characters “Karai” is voiced by Ziyi Zhang. For the 4th question, Mr. Buff is pretty happy with the recommendations and prefers to give a positive feedback to the system which can help it to better build a more complete user model for himself. And after seeing the neighbor-style explanation, Mr. Buff would probably like to view his neighbors’ profiles and follow some of their movie reviews to decide the next movie to see and even add them to his own friend circle to have a real-life conversation.

We can see from the above example that the interaction with the system (e.g., feedbacks, viewing neighbors’ profiles) and the interaction with other users (e.g., friend circle, conversation) are required by a common user. This hypothesis will also be confirmed by our interviews in the interactive prototype design session. Actually in the CBR-based recommender research field, the emerging concept of interaction (Swearingen, 2002) with the recommender system itself and with other users in a more social networking oriented recommender system, has become increasingly popular, therefore we take "interaction" as our third design element for our general design framework.

Furthermore, previous researches neither formulate any design framework nor do they still regard the explanation phase as merely a successive part (see Figure 2.2) after product information and recommendation presentation (Tintarev, 2007).

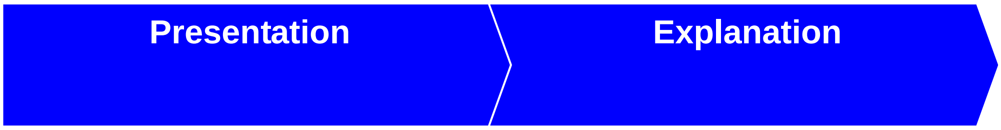


Figure 2.2 Linear relations between presentation and explanation

Therefore, after collecting "interaction" as the missing element from the puzzle, here comes a complete overview for designing an explanation interface. In addition, we also believe that these three design aspects, unlike only presentation and explanation forming a two-step linear process (Figure 2.2), can form a positive cycle as shown in Figure 2.3. We argue that future explanation interface design shall not only focus on how and what to present and explain, but also take a deeper look in the users' interactive participation, which will generate more rating and review data for further presentation and eventually more satisfactory recommendations and explanations alike.

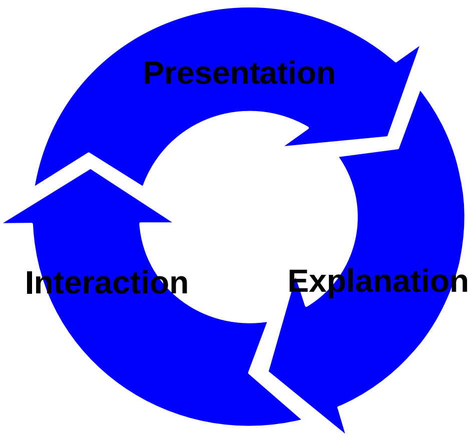


Figure 2.3 Three design aspects forming a positive cycle

2.7.2 Explanation of the Draft Design Framework

To make the concepts in this framework more precise, we try to clarify each aspect and provide some concrete descriptions in the next few paragraphs. Further research efforts will largely be based upon our discussion below.

What to present

1. Product expertise

   This seems generic, but a successful e-commerce site is definitely good at presenting rich data about a specialized product domain. If we consider “explanation” in a broader sense, a well-built product information profile page acts exactly as the most important source of explanation as long as users are familiar with the domain and context background. From our user requirement interview results (see Chapter 4) we can see that only when users are unfamiliar with specific domain knowledge, they tend to heavily rely on other types of explanation to justify a recommendation.

2. Recommendations

   Recommendations themselves are also needed to be presented to users before they can even be explained. Usually the page presenting recommendations shall also show explanations or at least provide quick access to explanations when needed.

3. System’s prediction on how much the user will like the current product

   Users can be intrigued to see the explanation behind a statement such as “you will like this movie” asserted on the product profile page. Also, prediction assertion shows that the recommender system is thinking on behalf of the user, which creates a much friendlier user experience.

What to explain

For different recommendation techniques, different explanations are needed. Here is a list of design elements that a recommender system shall use as explanation sources.

1. User-to-user similarity

   If the recommender system implements user-user collaborative filtering, then neighbor-style explanation like displaying neighbors’ ratings is proved to be well-accepted (Herlocker et al., 2000; Bilgic et al., 2005) as a good explanation.

2. Item-to-item similarity

   Item-item collaborative filtering technique is also widely used by major e-commerce sites like Amazon.com. According to previous research, key-word style explanation from (Bilgic et al., 2005) is a good approach in explaining item-to-item similarity. In later chapters, we will try to discover whether more suitable and effective means to explain this type of similarity exists.

3. Recommendation algorithm's inner logic

   This looks also very generic as well, but some websites do provide plain explanation on its recommendation algorithms on how they predict a user's likes and dislikes. Usually, a help page besides the recommendations contains such kind of information.

4. User-generated content

   Existing recommendation algorithms put more weights on users' historical rating and purchasing data while neglecting more humanized feedbacks like comments and reviews. We will in later interviews and surveys discover how important a role that user-generated content besides ratings and purchasing history is playing in explaining a recommendation.

Whom to interact with

1. To the system itself

   The "cold start" problem is common for any type of recommender systems. Nearly all recommenders encourage people to give as many ratings and demographic data as possible to cope with the inaccuracy of the early recommendations because of lacking users' information. A good explanation interface, as depicted in our proposed design framework, shall not stop right after mere explanation, but engage users to interact more with the system. One of our initial thoughts on this aspect is to let users to create tags about his/her interested items. These tagged keywords can become a new source of a user's personal data that reflect his/her general interests.

2. To other users

   The combination of online community features with pure recommender systems has lead to hot discussion in recent years (Bonhard, 2004). Konstan (2005) points out that "in psychology theory, if somebody believes that they have a unique contribution to make they are more likely to contribute than if they think that other people can make the same contribution and therefore they are reluctant". "Do you show them how much other people are benefiting from their work; or how much they've been benefited from the work of others" are the two questions raised by Konstan, J. (2005). In our design framework, the interaction between users is definitely an important element and in later chapters, we will find out existing best practices and test some of them in our experiment and survey.

Chapter 3: Explanation Interfaces of Major E-Commerce Sites

After our in-depth discussion and review on the theoretical side of the recommender system's explanation interface, we will in this chapter further explore its practical use in modern major e-commerce websites. Our selection includes: Amazon.com, IMDb.com, Hollywoodvideo.com, and Last.fm. Our aim of this survey is to evaluate the role that the explanation interface is playing in these recommender systems and also to identify existing best practices and generate guidelines for future designers.

Our analysis model roughly follows the draft design framework put forward in the 2nd chapter. We will examine each website's recommender system in three aspects, namely presentation, explanation and interaction. For each aspect, we will ask different questions and demand answers accordingly.

3.1 Amazon.com

3.1.1 Review of Explanation Related Functions

As a well-known recommender system implementer, Amazon's success depends a lot on its successful recommendations, especially by the item-item recommender techniques (Linden et al., 2003). The following part reflects a personal shopping experience, by which we will try to explore the explanation interface and related functions about Amazon's recommender system. On its homepage (Figure 3.1), every visitor is encouraged to "sign in and get personalized recommendations". Therefore, we can see that Amazon really prefers user engagement from its subscribers and puts significant weight on user interactions.

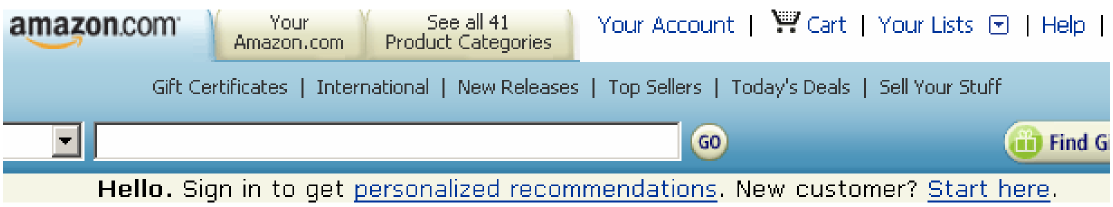


Figure 3.1 Amazon.com homepage

After signing in, because a new user like myself does not have any ratings or reviewing records, Amazon's recommender system simply says "sorry, we have no recommendations for you in this category today", but it encourages me to browse the top-selling items selected according to popularity (Figure 3.2).

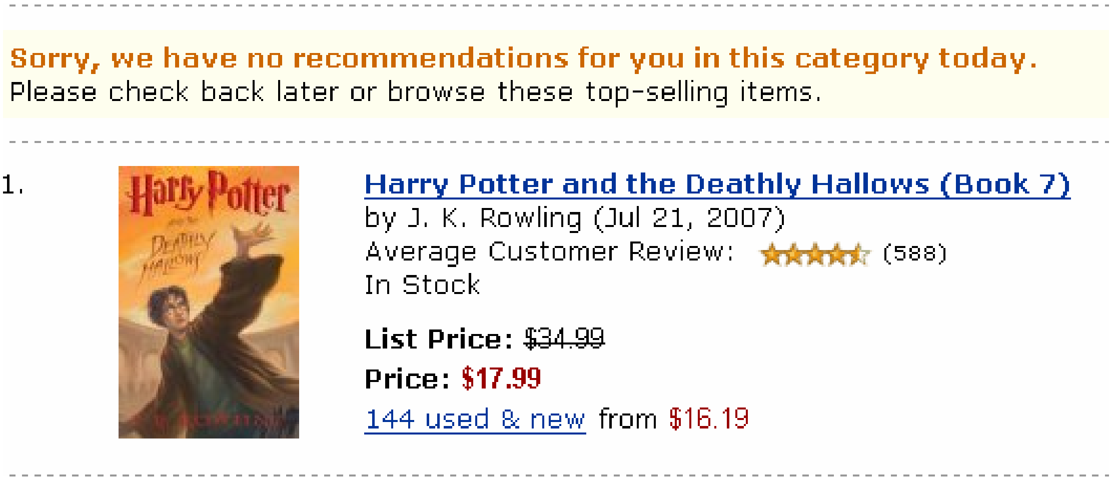


Figure 3.2 Popularity recommendation technique on Amazon.com

For each product's profile page, the average rating is presented with a bar chart to let user to view the rating breakdowns from all customers who have reviewed it on five-star scale (Figure 3.3). Furthermore, a user can even to read different reviews given by previous customers (Figure 3.5) and fully assess the quality and the buzz about the current product. Also, Amazon presents a list of products that are similar to the current one with a text-lighted assertion that "customers who bought items like this also bought".

Figure 3.3 Viewers' rating bar chart on Amazon.com

Figure 3.4 Amazon's item-item collaborative filtering recommendation

If we go further to browse all customers' reviews, Amazon.com provides even more interaction functions, such as comments on a specific review, report to the system whether the review is helpful, how many customers consider it helpful or even read all the reviews written by a specific customer (Figure 3.5). If a user is really interested in a particular customer's shopping tastes, he/she can surely follow that customer's reviews to help himself/herself to make purchasing decision. Moreover, Amazon's "badge" function (the subtitle like "Top 1000 Reviewers", "Real Name" under each reviewer's name) provides a measurement for users to assess a reviewer's credibility.

Figure 3.5 Customer review function on Amazon.com

A user can again navigate to a reviewer's profile page (Figure 3.6); further explore his/her tastes by tags and tagged products he/she is interested in (Figure 3.7); judge this reviewer's similarity to the user and even invite this reviewer as "Amazon Friend" or "Interesting People".

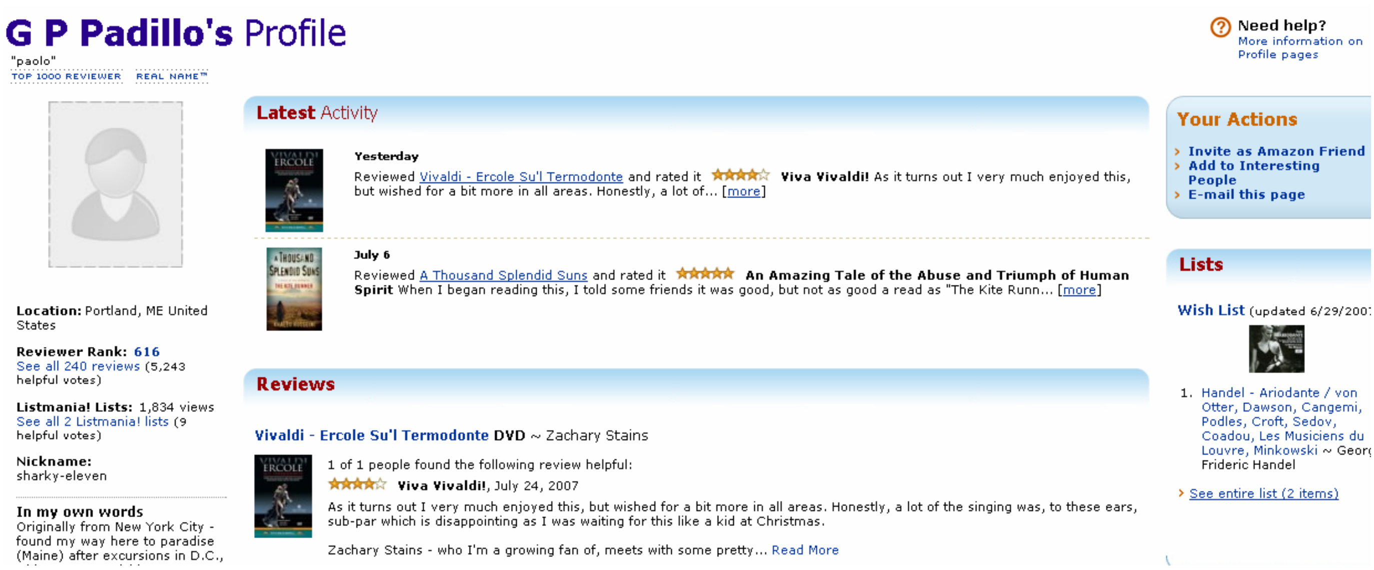


Figure 3.6 Profile page of a reviewer or customer

Figure 3.7 Tags and tagged products on reviewer's profile page

From the previous discussion, we can see that Amazon is really good at engaging its customers to interact with each other and form a mutual-beneficial online community. The best recommendations always come with real-world human voice. A buzz created by the users themselves is much more persuasive than the recommender system generated one.

After rating some movies and buying a book "Word of Mouse: The Hidden Marketing Power of Collaborative Filtering" (Riedl, 2001) from Amazon, we have a list of recommended products (Figure 3.8). It can be seen that only the explanation stating the recommender system selects these products based on my purchase history ("based on items you own") and ratings ("because you rated **Borat** and more"). One thing also worth noting is that Amazon provides the interactive functions to let users to adjust which purchase or rating records can used by the recommender system (Figure 3.9). A user can explicitly specify which items can reflect their interest to the system. This function is a good example of system transparency that lets the user take control.

Figure 3.8 Amazon's recommendation list

Figure 3.9 Amazon's function of specifying whether a rating can be used for recommendation

3.1.2 Summarization

Presentation

In the perspective of presentation, Amazon.com is doing a very impressive job. Firstly, its sophisticated item profile pages successfully demonstrate the product expertise. Secondly, different visualization techniques have been utilized such as text highlighting, five-star rating scale etc. Thirdly, functions like categorization of product display in tabular form and built-in search engine enables its users with quick and easy navigation. And fourthly, different recommendations are also classified according to different recommendation mechanisms like popularity-based recommendation and item-item based recommendation etc.

Explanation

Amazon takes various explanation measures. Firstly, the most outstanding one is the bar chart of reviewers' rating breakdowns which give a clear overview of how this item is liked by all interested customers. Secondly, Amazon puts great significance on its peer review system and let the customers' reviews to be the most compelling evidence to persuade other customers to buy a product. Thirdly, Amazon does not give any explanation on its famous item-item collaborative filtering technique, but roughly states that the recommended items are based on either the current user's purchase history, browsing history or rating records.

Interaction

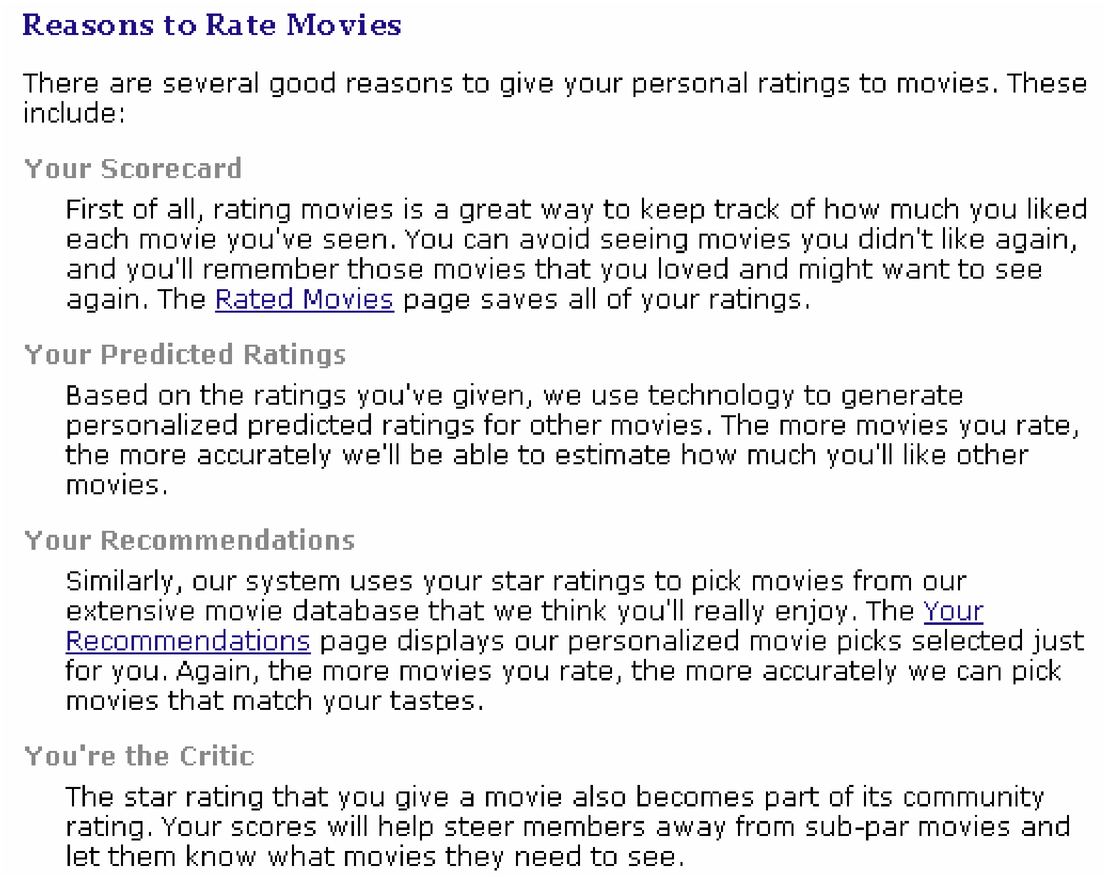
This is the very part that Amazon.com distinguishes itself as the one of the top recommender systems implementers from all the other major e-commerce websites. Firstly, on every product's profile page, a user can easily provide his/her opinions by either giving a rating or writing a review. Secondly, a user has full control of what type of personal data (ratings, purchases etc) can be collected for recommendation use. Thirdly, through Amazon's badge function, users can quickly identify top reviewers who usually have high credibility and tag them as "Amazon friends" or "interesting people"; these tagged customers' shared purchase history, rating records and reviews become a new source of explanation for a user to justify his/her purchase decision.

Amazon.com is specialized in both item-item collaborative filtering and customer community building. For the explanation of its collaborative filtering technique, Amazon is still using a "black box" approach; the only so called "explanation" or "hint" to be more precise is that Amazon states some customer related products as the source of its similarity comparison (Figure 3.8).The latter feature of community building significantly generates user stickiness to the website and enhances users' overall trust to the system and also mutual trust to each other. From the example of Amazon.com, our later experiment will test what type of explanations, between collaborative filtering neighbor style explanation and customers' positive reviews, can generate higher user satisfaction.

3.2 HollywoodVideo.com

3.2.1 Review of Explanation Related Functions

HollywoodVideo.com provides a general explanation of its recommender system's inner logic in its help page (Figure 3.10). Also on the same help page, the system gives four reasons to encourage users to rate more movies, namely, "your scorecard", "your predicted ratings", "your recommendations" and "you're the critic". The 1st reason is to track the user's own interests on different movies and this has barely anything to do with recommendation explanation; the 2nd and 3rd reasons are in a similar way to explain that more ratings can result in better predicted ratings and better recommendations; the 4th reason therefore puts more efforts on recruiting the user as a critic in order to help other users to assess the quality of a recommended movie. Moreover, on the same help page, it states that "Innovative technology analyzes the star ratings you've provided to figure out what your movie tastes are. Once you've rated several movies, we'll be able to start suggesting other movies from our extensive movie catalog that you might enjoy. The more movies you give ratings to, the more accurate our picks for you will be." Obviously, HollywoodVideo.com hides from the user the technical details of its recommender system, but asserts that it is using an advanced technology to assist users to select interesting products.



Reasons to Rate Movies

There are several good reasons to give your personal ratings to movies. These include:

Your Scorecard

First of all, rating movies is a great way to keep track of how much you liked each movie you've seen. You can avoid seeing movies you didn't like again, and you'll remember those movies that you loved and might want to see again. The Rated Movies page saves all of your ratings.

Your Predicted Ratings

Based on the ratings you've given, we use technology to generate personalized predicted ratings for other movies. The more movies you rate, the more accurately we'll be able to estimate how much you'll like other movies.

Your Recommendations

Similarly, our system uses your star ratings to pick movies from our extensive movie database that we think you'll really enjoy. The Your Recommendations page displays our personalized movie picks selected just for you. Again, the more movies you rate, the more accurately we can pick movies that match your tastes.

You're the Critic

The star rating that you give a movie also becomes part of its community rating. Your scores will help steer members away from sub-par movies and let them know what movies they need to see.

Figure 3.10 HollywoodVideo.com's help page on reasons to rate movies

On each movie's or DVD's profile description page, two pieces of reviews from Reel.com are presented (Figure 3.11). Also a section called "critics' perspectives" also shows general ratings from ratings from 3 major media (Figure 3.12). All the reviews are written by professional editors and critics; therefore, HollywoodVideo.com relies partially on manual recommendations and does not encourage users' participation into the site themselves.

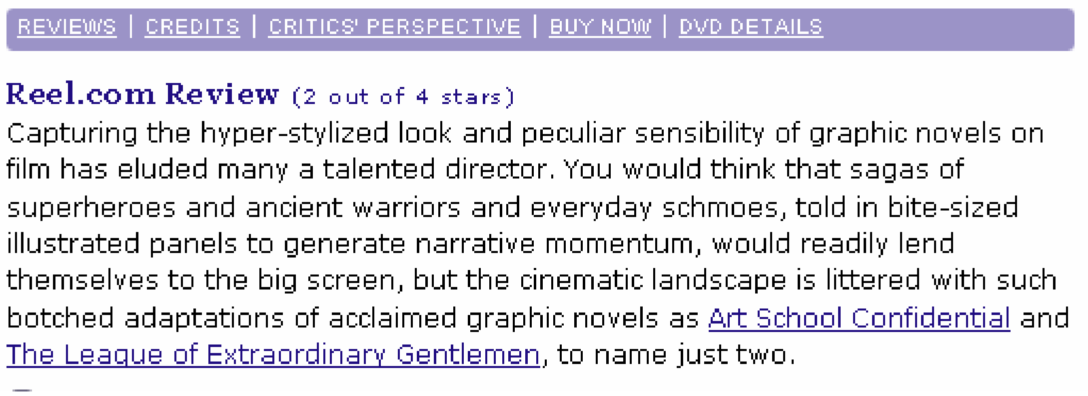


Figure 3.11 Reviews from Reel.com on HollywoodVideo.com

Figure 3.12 Critics' Perspectives on HollywoodVideo.com

On the "your recommendations" page, only a plain list of recommended items is presented and no further explanation is provided (Figure 3.13).

Figure 3.13 HollywoodVideo.com's recommendation page

3.2.2 Summarization

Presentation

Besides detailed movie/TV show descriptions, HollywoodVideo.com uses different color schemes to present three types of ratings in a five-star rating scale, i.e., user's rating, average rating and predicted rating. Also a plain list of recommendations is accessible from a navigation menu, but no classification is shown. Users can quickly navigate through different products with their predicted rating information highlighted when recommended.

Explanation

HollywoodVideo.com is still almost functioning under a "black box" model. Except a help page with the system logic description which also encourage users to rate as many movie as possible, very limited concrete evidence can a user find through each recommendation process. Average community rating and system asserted prediction are the only two means to justify a recommendation.

Interaction

HollywoodVideo.com's users interact with its recommendation system in an isolated way. The only inputs to the system are wish list, ratings and purchase records while the outputs are a recommendation list. There are no user-user interactions supported on HollywoodVideo.com.

We conclude that HollywoodVideo.com is still following the "black box" model and the three essential aspects of an explanatory recommender system have not been put together to form a positive cycle yet.

3.3 IMDb.com

3.3.1 Review of Explanation Related Functions

IMDb.com does not explicitly establish its name by its recommendation function, but users can be shown a recommendation list when viewing one movie's profile (Figure 3.14). These recommendations are partially generated by an item-item collaborative filtering technique and can also be contributable to users' inputs.

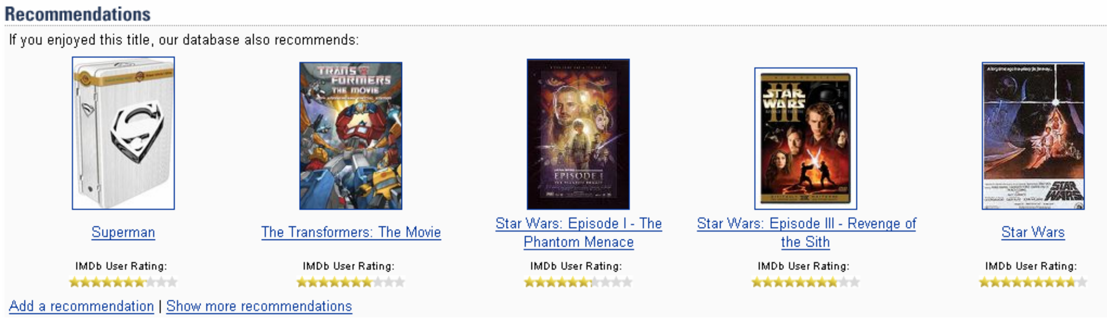


Figure 3.14 Recommendations from IMDb.com

IMDb.com's "add a recommendation" function encourages users who disagree with the recommended items to suggest the movies they think are more appropriate to be recommended. These users' inputs are claimed by IMDb.com to improve the results generated by recommendation algorithms (Figure 3.15). Furthermore, a quick recommendation engine is also available to generate recommendations according to any keywords that users type in.

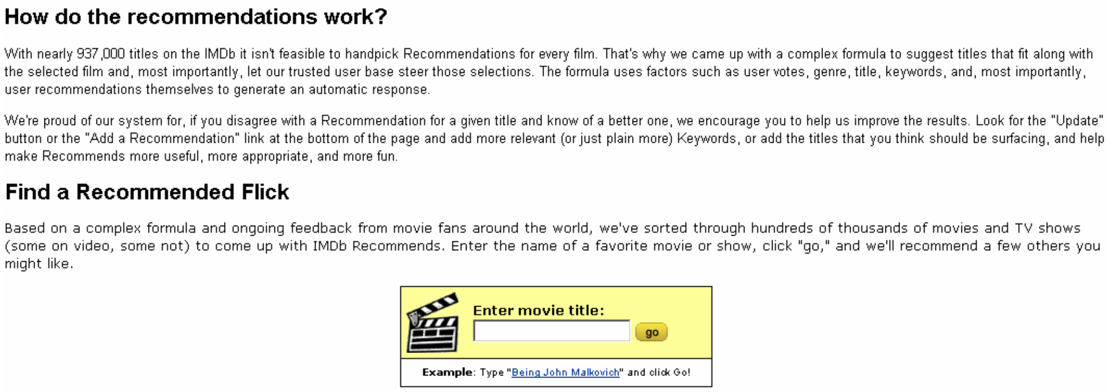


Figure 3.15 Recommendation logic explanation and quick recommendation engine

IMDb.com also utilizes users' comments and message board to engage users' participation in comment and discussion (Figure 3.16 and figure 3.17). Another notable function is to let users to decide whether a shown comment is "helpful" or not.

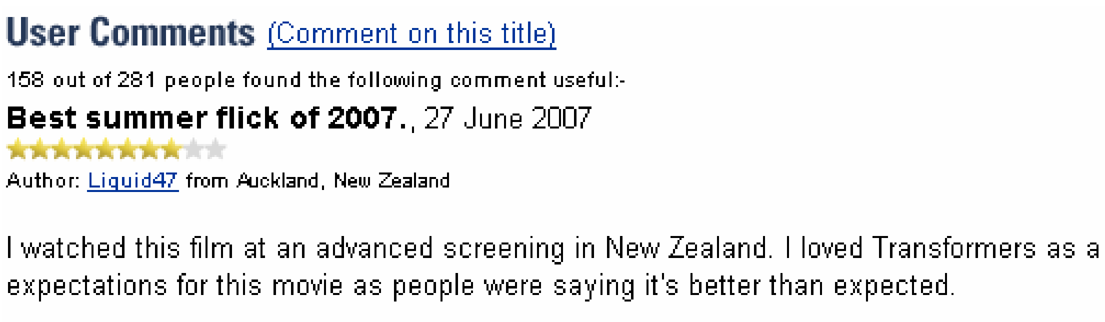


Figure 3.16 IMDb.com's user comments

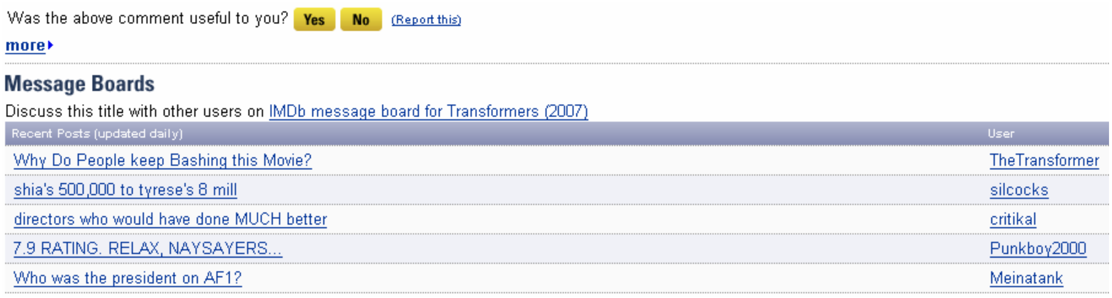


Figure 3.17 IMDb.com's message board

Each movie is given a ten-star scale rating and detailed rating information can be retrieved in different breakdowns such as the one from demographic angle (Figure 3.18).

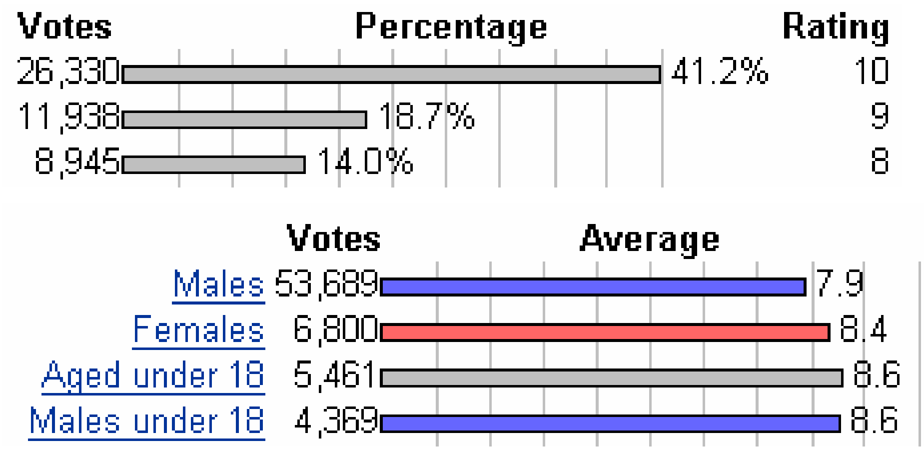


Figure 3.18 IMDb.com's rating breakdowns

3.3.2 Summarization

Presentation

In the perspective of presentation, IMDb.com, to be claimed the "Earth's Biggest Movie Database", shows unmatchable expertise in the domain of the movie entertainment industry. Different visualization techniques have been utilized such as text highlighting, ten-star rating scale etc while recommended films are presented in an item-to-item style.

Explanation

IMDb.com takes various explanation approaches. Firstly, the most compelling two are the bar charts of reviewers' rating breakdown and demographic breakdown, which provide a clear overview of how this item is liked by customers in different demographic groups. Secondly, IMDb.com puts great emphasis on the user comments and the message board system to generate a customized buzz about a movie. Thirdly, although IMDb.com's item-to-item recommendation remains "black box", it argues that users' inputs are playing a significant role in improving recommender system's overall accuracy: "the (recommendation) formula uses factors such as user votes, genre, title, keywords, and, most importantly, user recommendations themselves to generate an automatic response"(http://www.IMDb.com/title/tt0418279/recommendations#explanation).

Interaction

One of the reasons that IMDb.com becomes the ultimate movie database is its active user participation. Besides discussions on message boards, actor profiles and plot synopsis can both be updated by regular users with the moderation of professional editors. A user's comments with his/her rating information play another significant role for both recommendation and explanation. Most importantly, users can submit their recommendations when they are viewing a specific movie profile page and think some keywords or titles might be associated with the current movie. This is an outstanding feature compared with all the other reviewed e-commerce recommender system, because it explicitly invites and engages users into the recommendation process.

3.4 Last.fm

3.4.1 Review of Explanation Related Functions

Last.fm is one big step further toward a social approach to recommendations. It gives every user a quick start and friendly interface with multiple points of entry such as charts, tags, listeners etc. After a quick search and locating any interested artist, Last.fm presents a "similar artists" list right besides the profile page (Figure 3.19). A user can easily recommend one track to another user by clicking the  icon or express his/her love to the song by clicking the  icon.

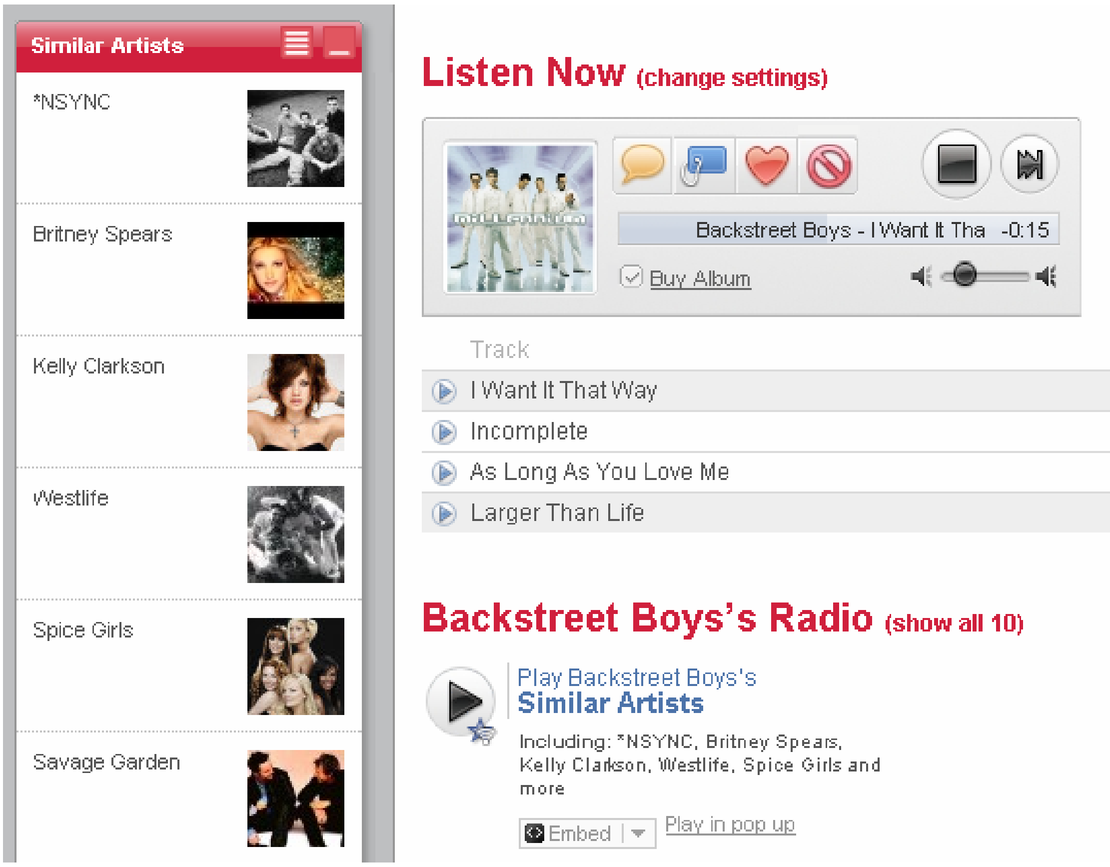


Figure 3.19 Similar Artists list from Last.fm

Obviously, Last.fm employs a recommendation algorithm based on item-item similarities. A user can also assess a full list of similar artists with similarity statistics in a bar-chart visualization style (Figure 3.20).

Figure 3.20 Full Similar Artists list in bar charts from Last.fm

Last.fm also fully combines computer generated recommendation with social network generated recommendations. As shown in figure 3.21, three types of recommendations are given to serve the listeners. As also can be seen from Figure 3.21, the system logic of automatic generated recommendations is expressed but without further explanation. While the other two types of recommendation are explained as coming from social networks and friend circles.

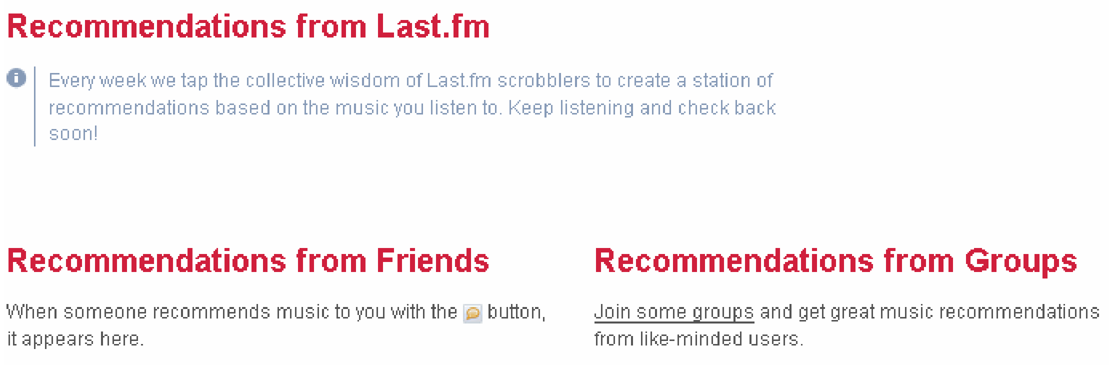


Figure 3.21 Three types of recommendations from Last.fm

Furthermore, Last.fm also widely uses the tagging function (Figure 3.22) to create more links inside the user-item space. And another strong explanation feature from Last.fm is users are allowed to trace “neighbors”, the other users with similar tastes. A track, a radio station or an artist can be tagged with keywords by users while users are tagged by each other to create different groups. Detailed music information plus users’ tracking records all act as a new source of explanation for both computer generated and social network generated recommendations.

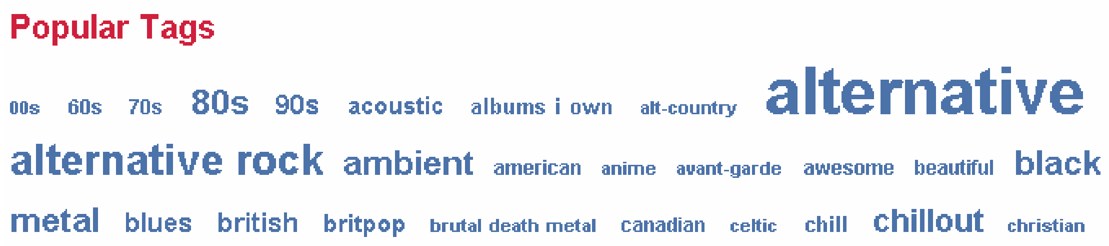


Figure 3.22 Tagging function in a cloud visualization style from Last.fm

3.4.2 Summarization

Presentation

Firstly, in the perspective of the presentation aspect, Last.fm can quick generate item-item recommendation by only a single query search; a list of “similar artists” and a group of keywords tagged to the target are the main output of Last.fm’s item-item based recommendation. Secondly, three types of recommendations are presented to every user, namely “recommendations from Last.fm”, “recommendations from Friends” and “recommendations from Groups”; the first type is from the “collective wisdom” which we believe is a user-user collaborative filtering technique while the other two are purely based on social network data, probably with the help with some data mining algorithms as well. Thirdly, the tagging function is heavily used to give users convenient access to any item in the product domain, which is also a good way to express the usability or the site and its expertise in the field of music.

Explanation

As for the recommended list of “similar artists”, users can demand a more detailed view with similarity scores depicted in bar charts. Also, every user is allowed to trace his or her like-minded “neighbors” and justify system’s recommendation by these similar users’ ratings, listening historical records and the most importantly tags. Most of the time, music lovers with the same interest also have a similar collection of tagged keywords.

Interaction

This is the most interesting aspect concerning this site. Last.fm, at a first glance, is more like a social networking site for all music fans to build communities while its recommendation function has been comparatively less outstanding. Users tend to join groups to share music (this site also supports music uploading) and share recommendations to each other as well. Personal recommendations are extremely popular and widely facilitated by various functions inside Last.fm’s online communities.

3.5 Conclusion of the Comparison

After reviewing the above four major e-commerce websites supported by various recommender systems and recommendation techniques. It is can be seen that the user participation and the online community building are setting up a new trend. How a recommender system’s explanation interface facilitates e-commerce sites to follow this trend is essential for future system designers. Clearly, Amazon.com and Last.fm provide the most convincing best practices for the design of the explanation interface while IMDb.com is also putting great efforts in engaging users for giving more inputs. However, HollywoodVideo.com still follows a traditional “black box” model with neither sufficient explanation nor users’ participation to be presented.

If we apply our design framework onto these four e-commerce websites, they score differently on each of the three design aspects. A figure like the following (Figure 3.23) is our effort to illustrate how good each of the 4 major websites are in a three-dimensional space with the design aspects as the three axes. Due to the fact that HollywoodVideo.com is still following the “black box” approach, it has only a position on the presentation axis. Also can be seen from the figure is that Amazon.com and IMDb.com are much more balanced with Amazon.com more fully-fledged in all the three dimensions. And the biggest winner in the direction of interaction is Last.fm.

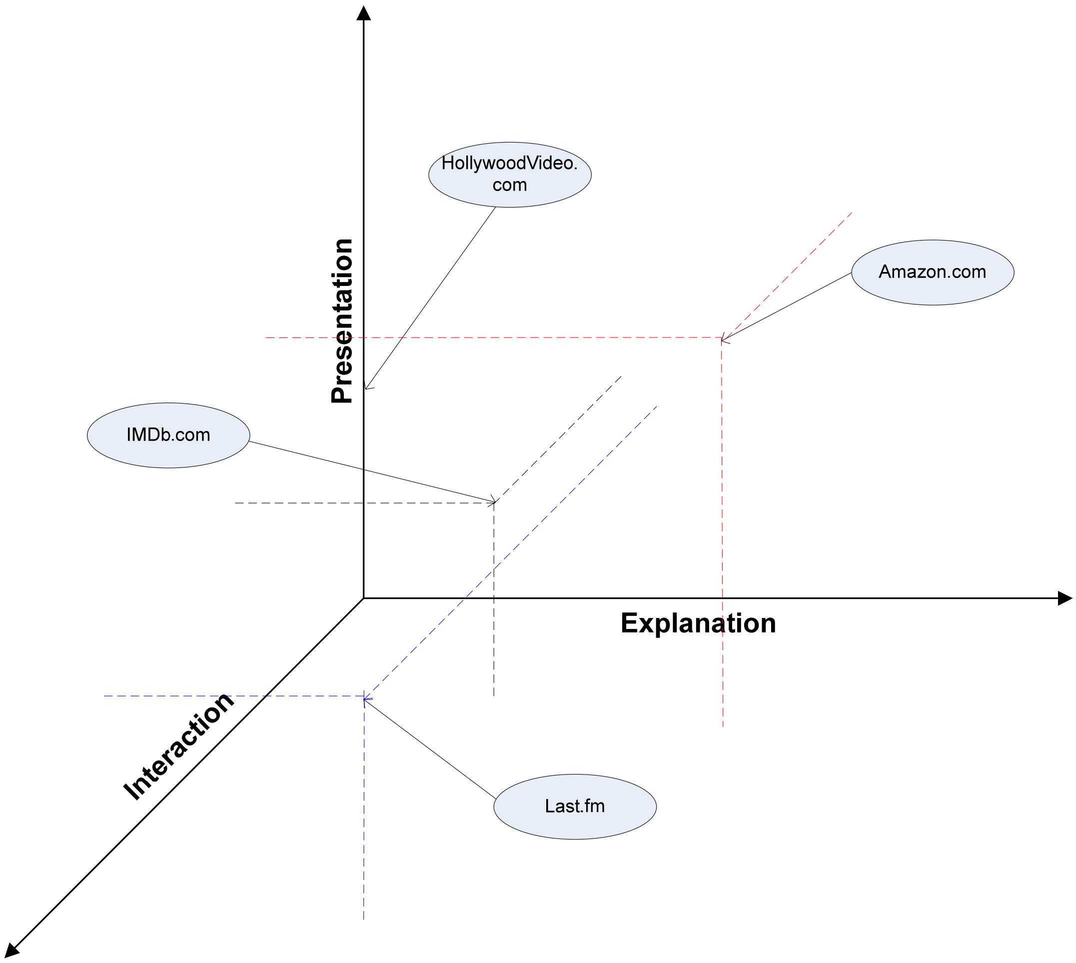


Figure 3.23 the positioning of four major e-commerce websites in the three aspects of an explanation interface

3.6 Draft Design Principles

Based on our observations from this chapter, we can summarize a set of several draft design principles from the existing best practices:

1. In the aspect of presentation, instant recommendation is the key to demonstrate product expertise, right after users give the first input, item-item recommendations shall be displayed; keyword tagging functions about product content shall be fully implemented and users shall be enabled to collect and create the tags that interest them; also clear paths to detailed product information and community feedbacks such as reviews, comments and similar users' ratings shall be mapped out.

2. In the aspect of explanation, user (neighbor) rating breakdowns in bar charts can give clear overviews over a large set of users' (neighbors') opinions; item-item (user-user) similarity

displayed in a bar chart also presents a quick snapshot of comparison; user-generated contents such as comments, reviews and tags shall be used as a new source of explanation data besides traditional rating records or demographic prototype information; a user's ranking can also be a convincing factor to support his/her credibility when comment and review data are used for explanation. We also propose that future recommendation algorithms shall expand their scope from only purchasing and rating data to harness more implicit community feedback data.

3. In the aspect of interaction, the explanation interface shall provide an advanced social navigation function that allows users to tag each other or create online communities; by sharing more personal information and joining discussions concerning a similar topic, users can profit from mutual contribution and be encouraged to contribute and generate more content to the site; a user's personal profile page shall be customized according to his/her privacy preference, while purchasing records and rating history shown to public or friend circles can create more purchasing opportunities for the e-commerce website.

Chapter 4: Prototype Design and Experiment

In this chapter, we will map out the process of how our experiment and survey are designed and carried out. Starting from interviewing five users of recommender systems, we collect some of the basic user requirements and perceptions about a recommender system's explanation interface. Later, we choose the Vogoo CF engine as our experiment platform and implement a user-user recommendation technique based on the MovieLens dataset. Based upon both user experience data and survey answers collected during and after the experiment, we try to summarize a set of design principles for the explanation interface of a future recommender system.

4.1 User Requirements Analysis

4.1.1 Pre-experiment Interview

For our prototype system, we take a user-centered design methodology starting from an interview session with 5 (potential) recommender system users to collect some general design requirements (the 10 interview questions and the participants' demographic information can be found in Appendix I). During the interview process, we also demonstrate our first prototype with simple interface and no explanation to demonstrate as an example when participants do not recall or never realize what a recommender system looks like.

4.1.2 Results from the Interview

Presentation

Out of the 5 interviewees, 4 of them have had online shopping experience and 2 of the 4 online shoppers deliberately claimed that they benefited from the recommendations from Amazon.com and Taobao.com respectively. In the perspective of presentation, 5 interviewees all agree that a plain list of recommendations is insufficient and well-built product profile page is commonly used to justify a recommended item. Surprisingly, all of them have an overwhelming dubious feeling toward recommendations when answering the question "what's your general point of view of these recommendations". These responses totally justify the necessity of implementing an explanation function besides the "black box" recommendations.

Explanation

Their general response to the next question “do you like to know why these recommendations are made for you” is that presenting explanations can be useful, but these explanations may become as dubious as recommendations themselves because they might end up to be an implicit promotion channel manipulated by the website, unless this website has an established brand name such as Amazon.com. This answer is also astonishing in a way that an e-commerce website’s brand image plays an unparalleled role in justifying both recommendations and explanations. Concerning what types of explanation they would like, we presented them with neighbor-style, keyword style and influence style recommendations from (Bilgic et al., 2005); they all agree that neighbor-style explanation has the most satisfactory effect while the other two look more confusing. Furthermore, 4 of the 5 participants express a preference toward a friend-made recommendation; only 1 of them firmly believes more in computer-generated recommendations. Concerning user-generated content, all the 5 interviewees agree that positive reviews or comments are valuable evidence and 1 participant even gives an example when booking a hostel on the website of worldhostel.com, previous tenants’ reviews are considered as the ultimate go or no-go explanation.

Interaction

Talking about interaction and user participation with the recommender system, 4 interviewees express the willingness to rate a product or post comments and reviews to an recommender-based online community, while 1 feels reluctant, but to post or rate only under the circumstance that the recommended product is either impressively good or extremely bad in quality. Also about their own privacy, personal profiles like ratings, purchase history and reviews can be shared in a controllable way, but only 2 interviewees are willing to be directly contacted by other users who might add him/her into a friend circle or watch list. But when we ask the question whether they would like to trace other users with system-qualified credibility and base their own purchases on these credible users’ ratings and reviews, the 5 of them all agree with this social navigation function.

In summary, we can conclude from our interview that besides recommendation and explanation, users requires deeper interactions which elicit more user engagement and stickiness to the site. And furthermore, users are also counting on user-generated content like reviews to justify their purchasing decisions.

4.2 Prototype and Experiment Design

4.2.1 System Structure

Our prototype is developed based on the Vogoo collaborative filtering engine which is an open source PHP library. Its current version is 1.8.1 (Vogoo, 2007) with its core k-NN algorithm based upon user-user similarity. Besides user-user collaborative filtering, the item-based Slope one algorithm (Lemire et al., 2005) is also provided. Although compared to other back ends written in Java or C++, Vogoo suffers more from the scalability problem, yet for the greater good of our experiment, it can quickly deploy a decent web-based user interface with a user similarity display function which is a major part of our explanation interface. By expanding its existing library functions, more explanation features have been added such as registration, user profile viewing, rate, purchase, predict etc and various visualization functions supported by JPGraph, another open source PHP library which provides rich functions on visualization. In addition, we use the popular MovieLens dataset (MovieLens, 2007) from the GroupLens research group as our movie, rating and user information repository. The following diagram depicts the main structure of our recommendation prototype system with two separate interfaces.

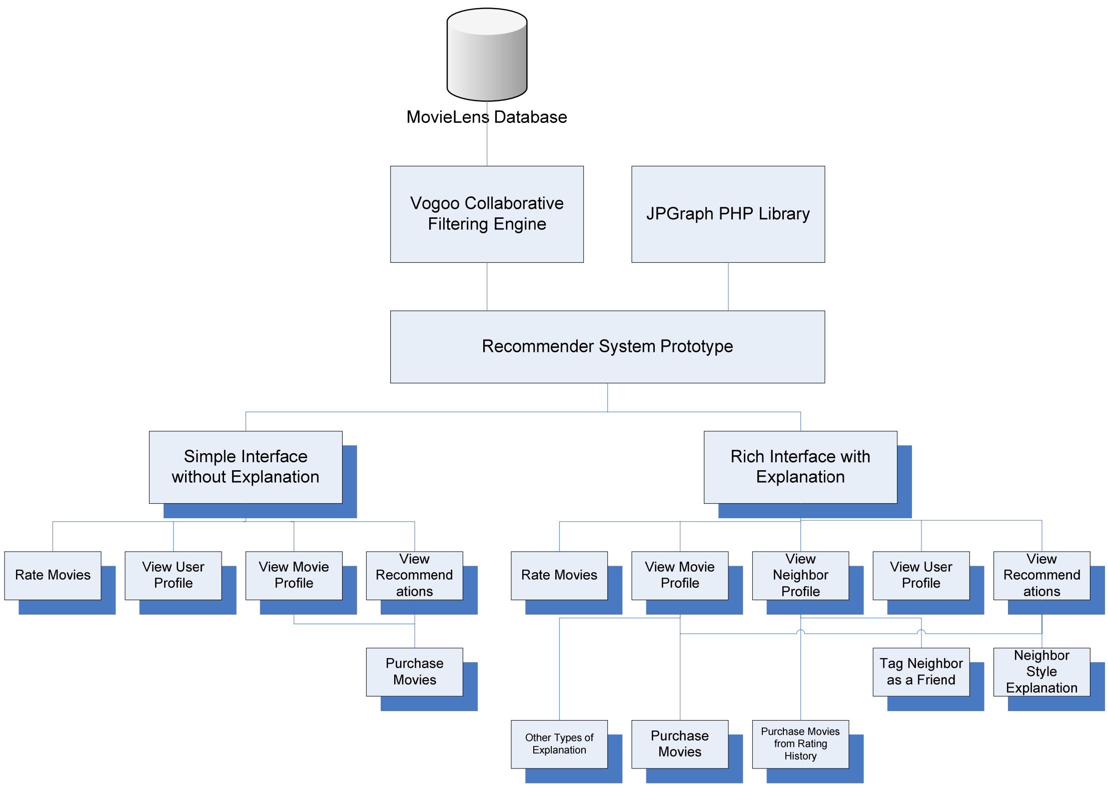


Figure 4.1 Recommender system prototype structure

4.2.2 The Simple Interface

The simple interface basically follows the original built-in functions of the Vogoo collaborative filtering engine. A user, after logging on, needs to rate some movies to warm up the system and get recommendations based on similar neighbors' rating records. A user can also view his profile including rating history. On each movie's profile page (Figure 4.2), user can choose to purchase it. A general overview of the Top-10 recommendations (Figure 4.3) can be easily accessed from the navigation menu. The aim of developing such an interface is to test our explanation interface's promotional effect as the simple interface serves like a traditional "black box" approach. Besides, the only user data to be collected are users' purchasing records and they will be compared with the purchasing data from the user group who test the rich interface with explanation and interaction functions.

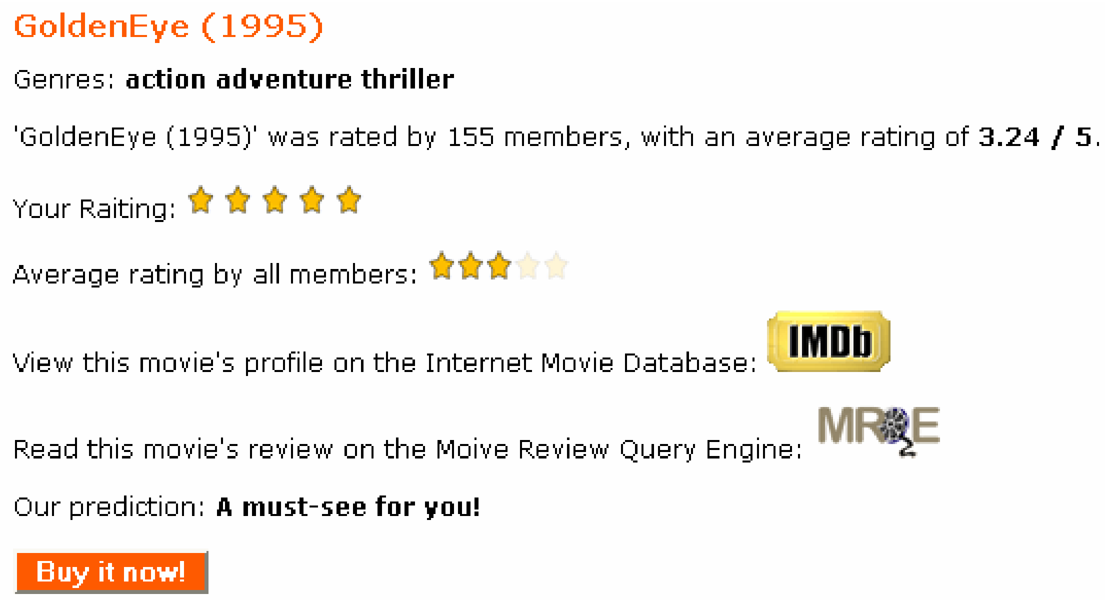


Figure 4.2 Movie profile page

Figure 4.3 Recommendations on the simple interface

4.2.3 The Rich Interface

Besides all the features from the simple interface, this rich interface is built under the guidance of our proposed design framework. It combines not only explanation functions from research literature such as histogram of neighbors' ratings (Herlocker et al., 2000), but also those explanation methods that come from existing e-commerce websites such as reviewer's ranking which is one of Amazon.com's features. Furthermore, some interaction-oriented functions such as social navigation (view neighbors' profiles) and friend tagging are also included in our rich interface. We present four major system interface screenshots below to give readers a general overview of our prototype system.

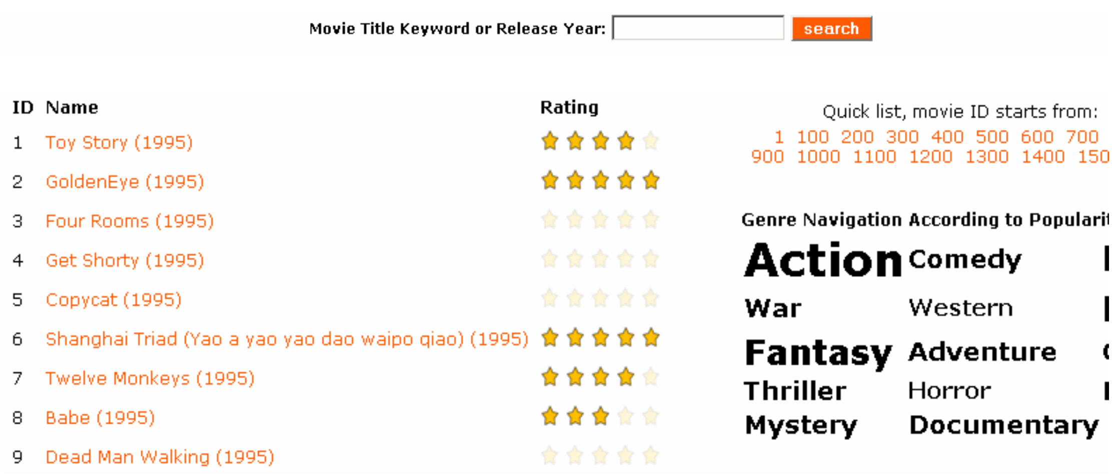


Figure 4.4 The movie rating page

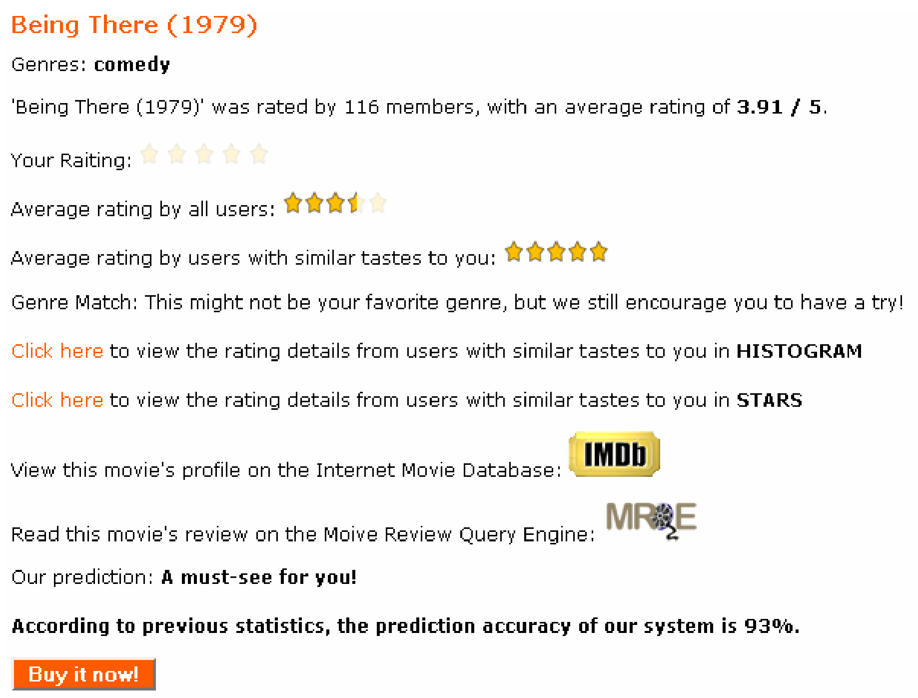


Figure 4.5 A movie's profile page

Figure 4.4 shows the movie rating page, in which the cloud style genre tags are presented to enable users with the quick genre navigation function. We also adopt the Ajax technology to let users roll over the stars to give a movie a rating without submitting a form and refreshing a whole page. A simple search engine is positioned on the top of the page to facilitate a quick search for users as well. If the active user clicks on any movie title, he/she will be linked to the actual movie profile page (Figure 4.5) with many explanations presented such as the average rating by all (similar) users, the rating details in histogram or stars, external links to IMDb.com and MRQE.com, prediction assertion and past performance declaration. And of course, if the active user is convinced to buy this product, a “Buy it now” button is handy to press on.

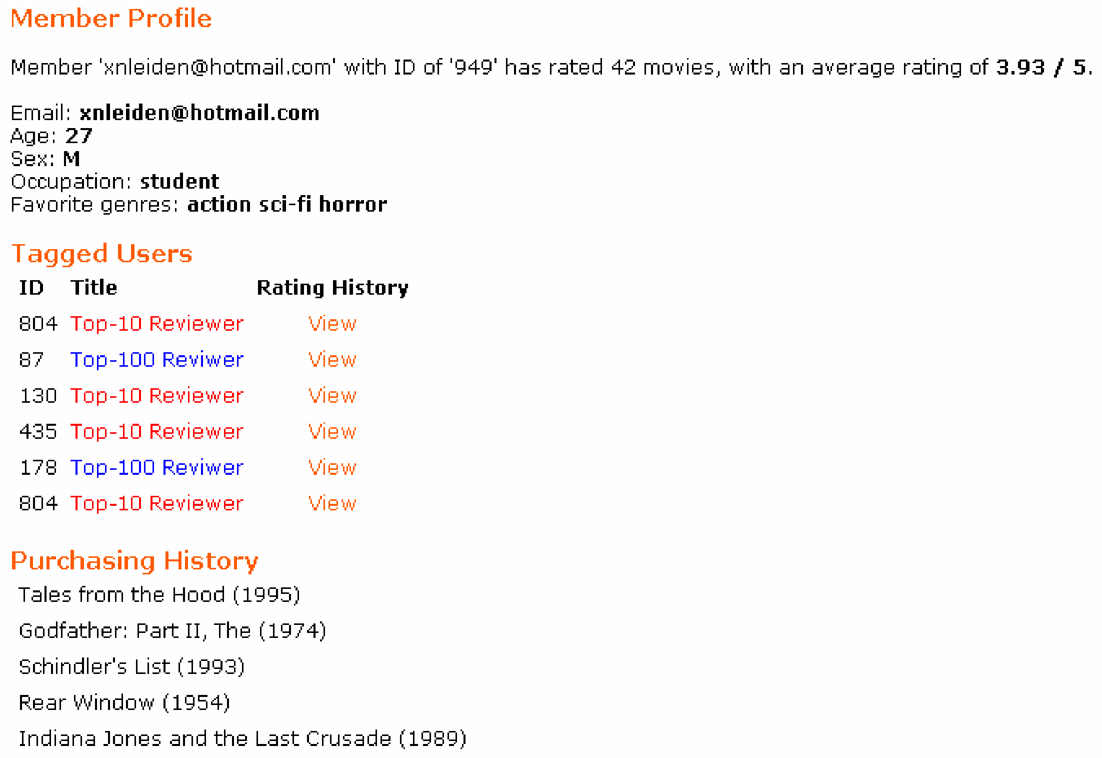


Figure 4.6 Active user’s own profile page

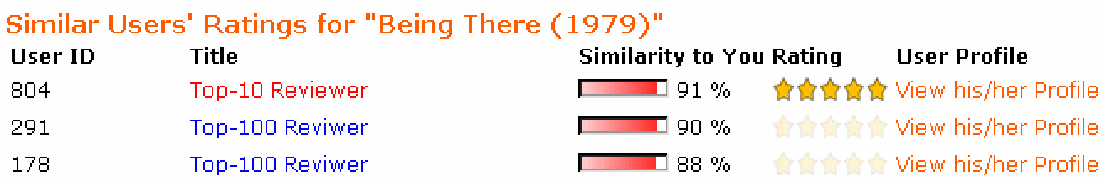


Figure 4.7 The explanation page showing neighbors’ ratings in stars

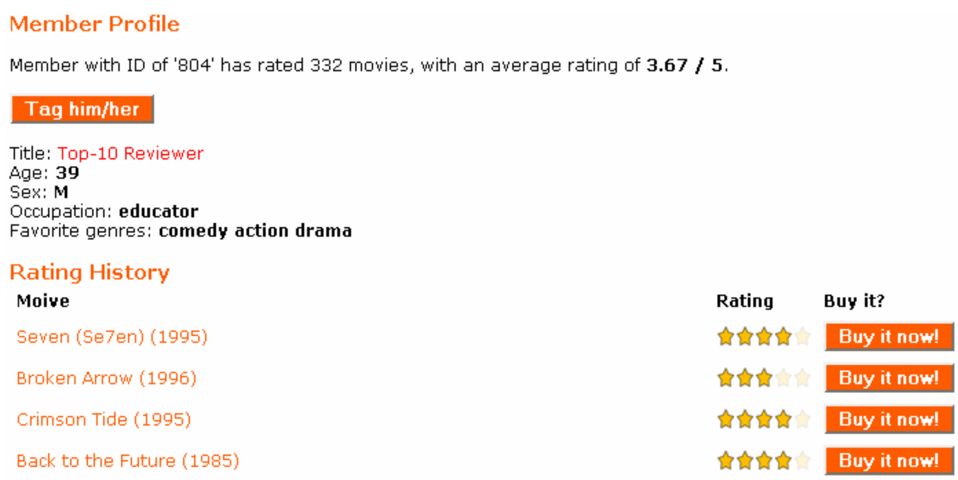
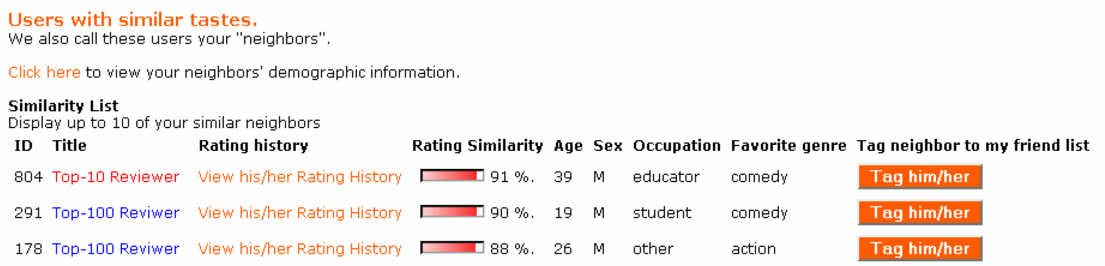


Figure 4.8 A like-minded user’s profile page

The active user can access to his/her own personal profile page (Figure 4.5); from this page, he/she can view his/her rating and purchasing histories and most importantly, the tagged friends list. This is part of the social navigation function realized in this prototype. Through the tagged friends, the active user will be encouraged to find explanations and recommendations from others' rating and purchasing records (Figure 4.6) and more purchasing opportunities are too given on the social navigation page as can be seen from Figure 4.8. Another notable feature is many neighbor-style explanation pages like Figure 4.7 also provide links to like-mined users' profiles which generate extra social navigation effect.


Users with similar tastes.
We also call these users your "neighbors".

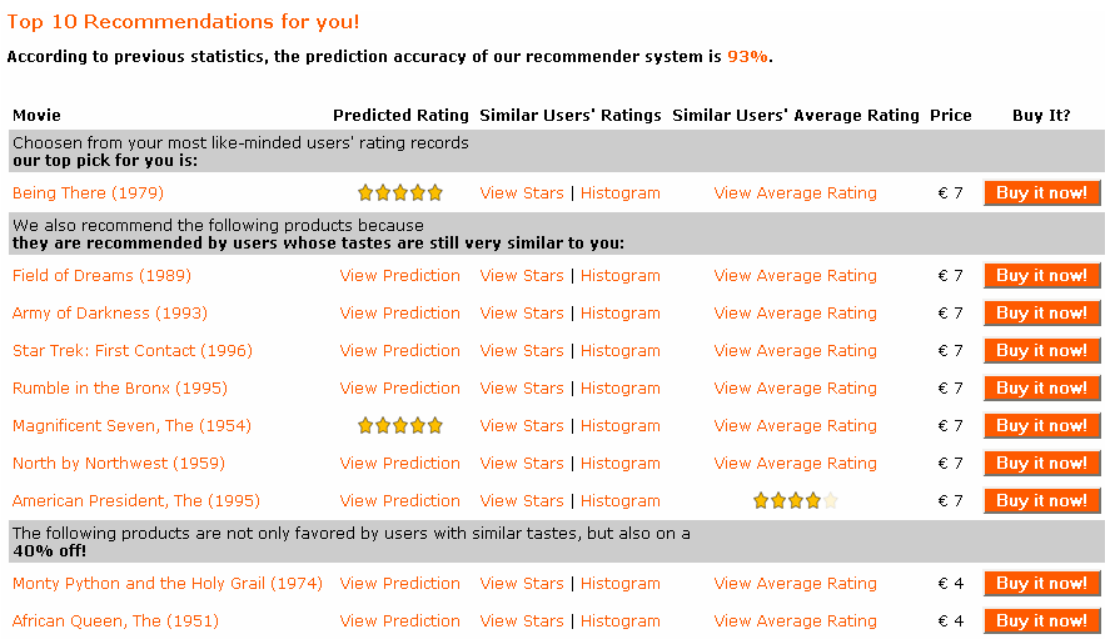
Click here to view your neighbors' demographic information.

Similarity List
Display up to 10 of your similar neighbors

ID	Title	Rating history	Rating Similarity	Age	Sex	Occupation	Favorite genre	Tag neighbor to my friend list
804	Top-10 Reviewer	View his/her Rating History	91 %.	39	M	educator	comedy	Tag him/her
291	Top-100 Reviwer	View his/her Rating History	90 %.	19	M	student	comedy	Tag him/her
178	Top-100 Reviwer	View his/her Rating History	88 %.	26	M	other	action	Tag him/her

Figure 4.9 The interface showing like-minded users' profiles

Figure 4.9 is the interface showing all the active user's neighbors including their demographic data and rating behavior similarities to the active user. We also present reviewer's ranking info as another clue to explain a neighbor's recommendation and community credibility. The tagging function presented above keeps up with our design principle of engaging users in as much social navigation as possible.


Top 10 Recommendations for you!
According to previous statistics, the prediction accuracy of our recommender system is 93%.

Movie	Predicted Rating	Similar Users' Ratings	Similar Users' Average Rating	Price	Buy It?
Choosen from your most like-minded users' rating records **our top pick for you is:**					
Being There (1979)		View Stars | Histogram	View Average Rating	€ 7	Buy it now!
We also recommend the following products because **they are recommended by users whose tastes are still very similar to you:**					
Field of Dreams (1989)	View Prediction	View Stars | Histogram	View Average Rating	€ 7	Buy it now!
Army of Darkness (1993)	View Prediction	View Stars | Histogram	View Average Rating	€ 7	Buy it now!
Star Trek: First Contact (1996)	View Prediction	View Stars | Histogram	View Average Rating	€ 7	Buy it now!
Rumble in the Bronx (1995)	View Prediction	View Stars | Histogram	View Average Rating	€ 7	Buy it now!
Magnificent Seven, The (1954)		View Stars | Histogram	View Average Rating	€ 7	Buy it now!
North by Northwest (1959)	View Prediction	View Stars | Histogram	View Average Rating	€ 7	Buy it now!
American President, The (1995)	View Prediction	View Stars | Histogram		€ 7	Buy it now!
The following products are not only favored by users with similar tastes, but also on a **40% off!**					
Monty Python and the Holy Grail (1974)	View Prediction	View Stars | Histogram	View Average Rating	€ 4	Buy it now!
African Queen, The (1951)	View Prediction	View Stars | Histogram	View Average Rating	€ 4	Buy it now!

Figure 4.10 The Top-10 recommendation page

Figure 4.10 is the most important Top-10 recommendation page, on which four types of explanations are presented, namely, system prediction assertion, neighbors' ratings in stars, neighbors' ratings in histograms and average rating from all users. We trace users' clicking behaviors to these links and also the "Buy it now" button to get an overview of how implicitly liked these explanations are by the participants and their actual purchasing records as well. In our post-experiment survey, we will also explicitly ask the participants to rate these explanation and interaction functions on five-star scale.

Besides the above major interfaces from the prototype system, we summarize in the he following table all the explanation and interaction functions with different underlying visualization techniques.

Explanation Functions	**Visualization Technique**
Past accuracy performance assertion	Text highlighting
Average rating by all users in stars	Star
Average rating by all users in plain text	None
Average rating by users with similar tastes in stars	Star
Reviewer's ranking	Text highlighting
Genre match display	Text highlighting
Help page with the description of system inner logic	None
Similar users' ratings in stars	Star
Similar users' ratings in histograms	Histogram
Similar users' demographic backgrounds in pie chart	Pie chart
Similar users' demographic backgrounds in tables	Table
Similar users' similarity to you in bar chart and percentage	Bar chart
Interaction Functions	**Visualization Technique**
View neighbor's ratings	None
Friend tagging	None

Table 4.1 Explanation and interaction functions

4.3 Experiment and Survey Design

Since every explanation interface deals with the two major aspects, namely promotion and satisfaction, they become the first two targets for our experiment and survey. Then the other two parts from our design framework, interaction and presentation, will also be assessed.

4.3.1 Promotion Effect

Two groups of participants are invited to test our prototype system, i.e., the "haves" group and the “havenots” group. The “havenots” group is asked to test our simple interface without explanation functions in one login session. During this session, they are suggested to rate as many movies as they would like to and view the Top-10 recommendations and decided whether to buy a recommended movie or not. Likewise, the “haves” group will be asked to go through the same process on the rich interface. The system will record any purchasing action made by participants. These purchasing records from the two groups will be compared to each other. This comparison is a direct answer to one of our research questions: “How much can an explanation interface enhance a recommender system's promotion effect in terms of website online sales revenues?”

4.3.2 Satisfaction Effect

Explanation functions and their visualization techniques can result in different user acceptance. A good method to assess satisfaction effect for a recommender system has been mentioned in (Bilgic et al., 2005): a user is asked to rate a book and rate it again after reading it. Since this method is not applicable under our experiment environment, we use a different approach by letting users fill out an online survey to express their satisfaction toward different explanation functions including their visualized forms by a 1 to 5 rating scale. Our academic reference is from (Herlocker et al., 2000) which used the same approach. The explanation function presented in Table 4.1 are all to be rated by users to see how satisfied they are.

4.3.3 Interaction Effect

Since our prototype is not a full-fledged system with all possible interaction functions due to lack of data also because of the time constraint, we realized two social navigation functions in our system, i.e., viewing neighbors’ ratings and friend tagging. Participants on the rich interface will also rate on these two functions according to their user experiences. From their ratings, we can have a brief overview of how well the reception of these two functions is. We also summarize 4 other interaction-oriented functions from e-commerce best practices which we do not implement in our prototype. Therefore, we design four corresponding survey questions to ask users to give their opinions on the mock-ups of these functions. These four questions are:

1. How much would you like a function that a recommender system preferentially selects movies based on your tagged friends prior to other less familiar users?
2. How much would you like a function that let you trace the ratings and reviews from tagged friends and base your movie-going decision on these data?
3. How much would you like a function that enables you to participate in the discussion with other users and write reviews about the movies you watched?

4. How much would you like a function that enables you to rate others' movie reviews and let other users to rate your reviews too?

4.3.4 Presentation Effect

Five survey questions are aiming at the presentation effect. According to our design framework, a well-expressed product profile is the first front line of good explanation. Quick product domain navigation also facilitates the system to present product expertise to users. Better visualization techniques like clear categorization of recommended items presents a clearer overview of users' potential choices and hence save cognitive efforts. Therefore, the following five questions are:

1. Links to a movie's profile on IMDb.com and its critique on MRQE.com.
2. Prediction assertion.
3. A cloud style quick genre navigation function.
4. Recommendations in categorization which provides different groups of choices.
5. Recommendations in a plain list without categorization (screenshot from the simple interface of this prototype).

A full list of survey questions can be found in Appendix II.

4.4 Data Analysis

4.4.1 Comparison of Promotion Effect

We invite 24 participants to have a test tour on our simple interface. The only possible actions they can perform are rating movies, viewing movie profiles, viewing recommendations and "buy" an interesting one. Their average amount of purchased movies is 2.09, meaning that a common user in one session "buys" about 2.09 item. We also invite 36 participants to test our rich interface. The purchasing records from this "haves" group are divided into two sections, one is from either a movie's profile page or the recommendation page (same with the simple interface), the other one is from the social navigation page (when a user is viewing a neighbor's rating records, he/she can directly purchase from them, see Figure 4.8). Surprisingly, the purchasing data from the "haves" group in the traditional section is averaged at 1.42, lower than that of the "havenots" group. However, we also see that the "buy" behavior contributes another 1.0 from the social navigation section. Therefore, the total purchasing average from the "haves" group is 2.42, a significant increase compared with the "havenots" group's results. Our argument about the "haves" group's lower number in the same “buy” interface is that the “havenot” group has limited navigation capability, so users are tempted to press the “buy” button more; while the “haves” group are presented with rich explanation and interaction functions, purchasing itself becomes less tempting. However, due to the exposure to extra purchasing opportunities on social navigation pages, the final sales figure is boosted up in total.

Furthermore, we also perform a significance test on the two groups of collected data by using Excel’s t-test function.

t-Test: Two-Sample Assuming Equal Variances

	Havenots Group	Haves Group
Mean	2.086956522	2.416666667
Variance	2.719367589	12.76428571
Observations	23	36
Pooled Variance	8.887299771	
Hypothesized Mean Difference	0	
df	57	
t Stat	-0.41432063	
P(T<=t) one-tail	0.340097191	
t Critical one-tail	1.672028889	
P(T<=t) two-tail	0.680194383	
t Critical two-tail	2.002465444	

Table 4.2 The significance test on two samples of purchasing data

Because we are only interested in whether the rich interface has a better promotion effect than the simple interface, the one-tail test is relevant here. However, the p-value observed here is 0.34 which is larger than the alpha of 0.05. This proves that statistically, these two samples are not significant enough. So this result indicates that we need to expand our test to a larger audience to get a more concrete conclusion. However, we also present the detailed frequency of the purchasing records in histograms to see whether any interesting patterns exist. From Figure 4.11 and Figure 4.12 below, a trend that can be easily recognized is that a majority of users from both the haves groups and the “havenots” group only “buy” 2 or less products; however, there are a small portion of users from the “haves” group that “buy” more than 8 products while none of the “havenots” go to that far. We can then speculate that a number of these “far-reach buyers” are actually converted from the previous less willing “buyers”, because they now are exposed to more purchasing opportunities and more intriguing explanations that easily justify their extra purchasing decisions.

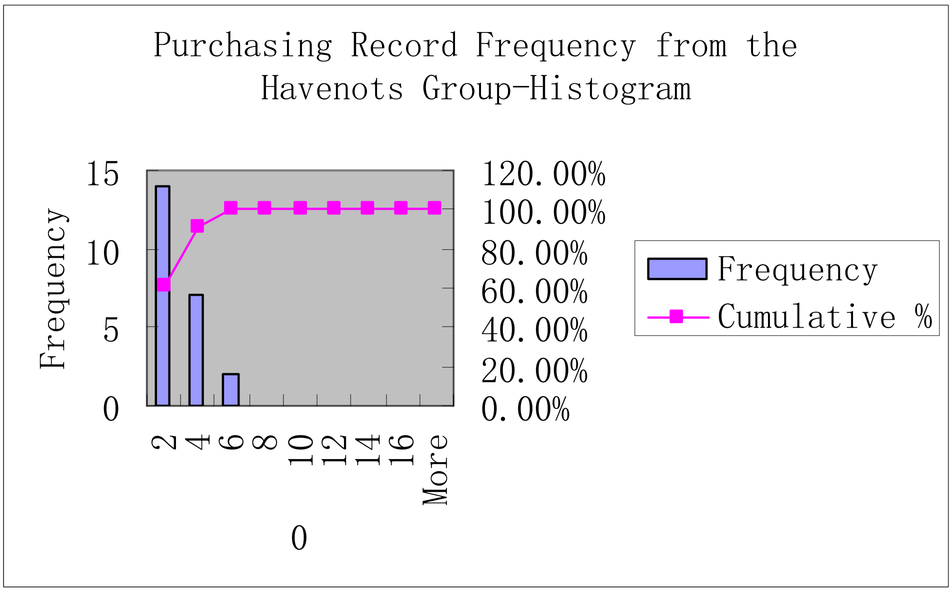


Figure 4.11 Histogram of the frequency of purchasing records from the “havenots” group

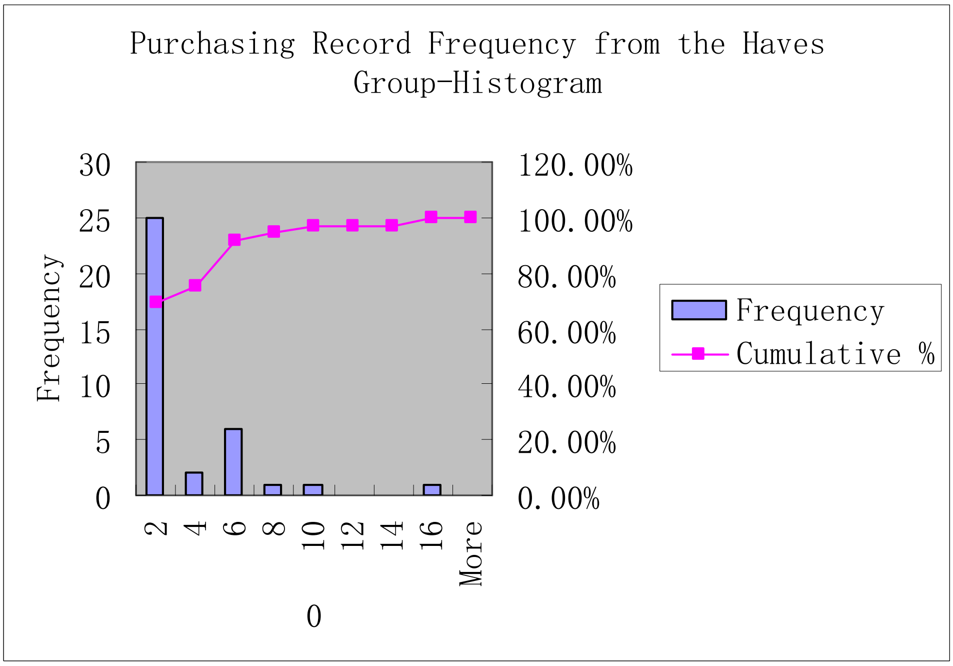


Figure 4.12 Histogram of the frequency of the purchasing records from the “haves” group

Under our speculation, this result shown in the difference of the average mean of purchased products totally supports our hypothesis that a rich interface can generate more promotion than a simple interface. A much more valuable conclusion can be drawn from the purchasing records which are generated from the social navigation pages is that we see a phenomenon that some users tend to be more convinced to buy other products highly rated by a like-minded user when viewing his/her rating history. So this also justifies another hypothesis of our experiment: more interaction between users induces more promotion.

4.4.2 Users' Behavior Data

Besides the purchasing data, we also collect 8 types of user behavior data on the 36 participants (30 out the 36 finished the post-experiment survey) from the "haves" group in order to assess the usability of different presentation, explanation and interaction functions (Table 4.3).

Functions	**Mean Average Views per Session**	**Standard Deviation**
Predicted rating on the recommendation page	0.61	0.84
Neighbors' ratings in stars and table form	0.58	0.96
Neighbors' ratings in histogram	3.83	4.31
Neighbors' average rating	0.44	0.93
Neighbors' demographic breakdown in pie chart	0.14	0.35
System's help page with description of user-user collaborative filtering logic	0.33	0.53
Friend tagging clicks	0.47	1.68
Social navigation page views when a user checks out a neighbor's profile	1.0	1.47

Table 4.3 Users' online behavior data from the "haves" group

Most obviously, participants favor neighbors' rating in histogram to serve as the most important evidence to assess an explanation with 3.83 views per session. The least interesting explanation is neighbors' demographic breakdown pie chart, which indicates that a common user does not care too much about neighbors' demographic similarity to him/her. Another interesting fact is the ratio of friend tagging clicks against social navigation page views which is about 47% (0.47/1.0). This is a good argument that social navigation function can significantly enhance user-user interactions.

4.4.3 Users' Satisfaction towards Presentation

We judge users' satisfaction the presentation effect of our prototype by the first five survey questions and the average ratings from survey users are:

Presentation functions and features	**Visualization Technique**	**Mean Average Rating**	**Standard Deviation**
1. Links to a movie's profile on IMDb.com and its critique on	None	3.69	0.99

	MRQE.com.			
2.	Prediction assertion.	Text highlighting	3.45	0.87
3.	Cloud style quick genre navigation function.	Cloud	3.42	1.12
4.	Recommendations in categorization which provides different groups of choices.	Categorization in tabular form	3.46	0.79
5.	Recommendations in a plain list without categorization (screenshot from an old version of this prototype).	Tabular form	3.27	1.03

Table 4.4 Average ratings to presentation functions and features

Another weakness of our prototype is the lack of professional movie information and that's why we provide links to well-known movie sites like IMDb.com and movie critique site MRQE.com. This approach serves an alternative means to demonstrate our expertise in the movie product domain. And unsurprisingly, users give a relatively high rating (3.69) toward this function and this justifies our hypothesis that an all-round product profile is our first explanation effort. Accompanying professional product knowledge, an easy navigation function like the tag-cloud style genre navigation scores average rating of 3.42 which is relatively lower and against our hypothesis. We believe that the reason behind this lower average rating is because this function is not completely realized and users are not tempted to fully utilize it either. However, plain prediction assertion with explanation may cause users' suspect and that's why it has a lower score of 3.45. Categorization of recommendation presentation in a conversation-like style is argued to be a better way to convey recommended product information to users (Pu et al., 2006). Our experiment results agree with this statement (average of 3.46 for categorized view in a tabular form against average of 3.27 for uncategorized view in a plain list). Therefore, we can say a categorized view of recommendations shall be included in our proposed design principles.

4.4.4 Users' Satisfaction towards Explanation

We also build some explanation functions with different visualizations based on our previous research and the industrial best practices. Besides tracing users' behavior data, in the post-experiment survey, we deliberately ask our participants to rate all these functions and some unrealized mock-ups to have another overview on their satisfaction toward our rich interface

(Table 4.5).

Explanation Functions and Features	**Visualization Technique**	**Mean Average Rating**	**Standard Deviation**
6. Past accuracy performance assertion.	Text highlighting	3.68	0.98
7. Average rating by all users in stars.	Star	4.2	0.76
8. Average rating by all users in plain text.	Plain text	3.55	0.94
9. Average rating by users with similar tastes.	Star	4.1	0.92
10. Reviewer's ranking.	Text highlighting	3.27	0.94
11. Genre match display (keyword style explanation).	Text highlighting	3.29	1.81
12. Help page with the description of system inner logic.	Plain text	3.32	1.21
13. Similar users' ratings in stars.	Star	3.85	0.86
14. Similar users' ratings in histograms.	Histogram	3.62	1.38
15. Similar users' demographic backgrounds in pie chart.	Pie chart	2.93	1.36
16. Similar users' demographic backgrounds in tables.	Tabular form	2.8	1.16
17. Similar users' similarities to you in bar chart and percentage.	Bar chart	3.7	0.95
18. Generally speaking, how satisfactory are you with the explanation function and its visualization as a whole?	None	3.8	0.55

Table 4.5 Average ratings to explanation functions and features

As can be seen from the above table, the most convincing four explanations are "the average rating by all users" (4.2), "the average rating by all users with similar tastes" (4.1), "similar users' ratings in stars" (3.82) and "similar users' similarities to you in bar chart and percentage" (3.7). Also, people consider the "star" illustration is much more compelling than the plain text, as "average rating by all users in plain text" is only rated as a relatively lower 3.55, definitely not eye-catching enough compared to its "star" expression. Moreover, histograms have been proved to be very popular by user behavioral data, but users' rating toward it is slightly lower than expected (3.62). Users' rating on neighbors' demographic information either in pie chart (2.93) or tables

(2.8) is among the lowest indicating that a user does not care much about demographic similarity. Reviewer's ranking also fails to prove to be very convincing with a relatively low rating of 3.27. Another interesting explanation function, which is developed based on key-word style explanation with the consideration of the tagging function, is the "genre match display". Because every user is asked to select three different genres as their favorites (a mock-up for the tagging function) before their testing begins, therefore, our prototype can explicitly display whether the movie shown to the user is a fit to his favorites. The motive behind that we build this function is inspired by Last.fm (see Chapter 3) which enables every user to have a collection of tagged keywords and we think each user's tag collection is a new source of explanation data. However, the average rating for this function is only 3.29. The last survey question for this section is a rating of participants' overall satisfaction feeling which shows a mean average of relatively higher 3.8 with the smallest deviation of only 0.55.

4.4.5 Users' Satisfaction towards Interaction

In this section of the post-experiment survey, we ask users to rate on two social navigation functions realized in our prototype (the first two questions). We also collect some opinions on unrealized mock-up functions which we think that a future recommender system shall have. The following table shows all the responses from the participants.

Questions about Interaction Functions	**Average Rating or Answers**	**Standard Deviation**
19. How much do you like the function that enables you to navigate through the users with similar tastes to you and tag a like-minded one?	3.79	0.90
20. How much do you like the function that enables you to view the rating histories of the users with similar tastes?	3.67	0.96
21. How much would you like a function that the recommender system preferentially selects movies based on your tagged friends prior to other less familiar users?	3.93	0.86
22. How much would you like a function that let you trace the ratings and reviews from tagged friends and base your movie-going decision on these data?	4.0	0.83
23. How much would you like a function that enables	3.47	0.97

you to participate in the discussion with other users and write reviews about the movies you watched?		
24. How much would you like a function that enables you to rate others' movie reviews and let other users to rate your reviews too?	2.87	1.00
25. How much would you like your limited profiles (like reviews, ratings and demographic background) to be viewed by other users or tagged friends?	80% Yes 20% No	

Table 4.6 Average ratings to interaction functions and features

The results from Table 4.6 indicate that social navigation functions like tagging "friends" and view each other's rating history (3.79) is highly welcomed by users. And the recommendation algorithm improvement on how to harness tagging information and user-generated review or comment data is also much more preferred, while simply enabling users to discussion participation (3.47) is rated relatively lower. An even worse rating (2.79) comes from the function mock-up to let users to rate each other's reviews. About the privacy issue, we also ask user whether they would like to be in full control of their public profile, 80% give positive answers, which is a strong proof that the majority of users would like to share personal information which can hence become another important source of explanation.

4.5 Limitations of the Experiment

The first and foremost outstanding limitation of our experiment is its limitation of testing time. After almost two weeks of running, around 70 participants took the test and about 30 of them finished the post-experiment survey, which prove to be a bit insufficient to draw a statistically significant conclusion from the purchasing data to compare the promotion effect. If we had a more flexible test period, more concrete data from a larger audience could be collected and a more accurate statistical analysis could be made.

The second limitation is from the technical side. The underlying MovieLens database is relatively old with most of its movies released before the year 1999, which affects our system's friendliness to movie goers who are not quite familiar with film made before the last decade. Another aspect often complained about by participants is the hosting service's limited bandwidth and relatively lower response time especially when Ajax enabled functions are involved.

The third limitation comes partially from the time constraint too and hence only two interaction-oriented functions are implemented with many other best practices yet remaining to be tested. Because of this, we make some mock-up functions based on the existing platform and collect participants' viewpoints by directly asking them to rate these mock-ups in the final survey. By doing so, unlike the ratings on explanation functions are supported with user behavior data, the credibility of interaction-related questions in the survey lack the back-up from participants' online behavioral statistical data.

4.6 Summarized Design Principles

With strong statistical support from our experiment and survey, we summarize the following design principles plus corresponding visualization techniques under the three design pillars in the proposed design framework.

4.6.1 Presentation Aspect

Design Principle 1: Demonstrate product expertise

Best practices include:

1. A recommender system shall be useful not only because of its recommendation and good explanatory support, but also be useful before any recommendation is made. Complete product profile which serves as the first explanation for a recommendation (industrial example: IMDb.com) is highly preferred.
2. Instant item-item recommendation is presented right after a user gives his or her first input (industrial example: Last.fm).
3. Each product can be tagged by keywords created by users (industrial example: Last.fm).

Design Principle 2: Easy navigation support in the product domain

Best practices include:

1. A clear path shall be built to guide users to detailed product profile with related reviews and comments displayed (industrial example: Amazon.com).
2. Cloud style tagging display enables users to quickly navigate through the product domain (industrial example: Last.fm).

Design Principle 3: Recommendation display in categorization

Best practices include:

1. A categorized display of the recommendation list in a tabular form can give some

recommended products a salient appearance by showing product attribute tradeoffs and positioning toward different segments of user needs; this is also beneficial to engage a user into a conversational interaction with the system (Pu et al., 2006) (academic example: the CBR-based recommender prototype from [Pu et al., 2006] and our prototype system).

4.6.2 Explanation Aspect

Design Principle 1: Display neighbor style explanation

Best practices include:

1. Users' or neighbors' rating breakdown in a histogram is the most popular and convincing explanation so far proved by previous literature (Herlocker et al., 2000; Bilgic et al., 2005) and also our prototype system (academic example: the prototype systems from [Herlocker et al., 2000; Bilgic et al., 2005] and our prototype system).
2. All users' or neighbors' average rating in stars and its breakdowns form in a bar chart are also well-received explanation functions in our experiment (industrial example: IMDb.com and Amazon.com).

Design Principle 2: Display item-item similarity

Best practices include:

1. Product similarity percentage display in bar chart (industrial example: Last.fm).

Design Principle 3: Use user-generated content as a way of mutual explanation

Best practices include:

1. Not only reviews, comments and reviewers' rankings are available for consultation, but other users' rating history and purchasing records can also serve as another source of solid evidence (industrial example: Amazon.com).

4.6.3 Interaction Aspect

Design Principle 1: Enable social navigation

Best practices include:

1. Users can tag each other and create groups with common interests; users in the same group can socially navigate through each other's shared personal information including ratings, reviews, tags and comments on rated or purchased products (industrial example: Amazon.com, Last.fm).

Design Principle 2: Make system participatory and sociable

Best practices include:

1. Online community building is essential in making a recommender system more sociable (industrial example: Amazon.com and IMDb.com); enable users or communities to directly recommend products to their friends or their members (industrial example: Last.fm).
2. Expose meta data through tags about both product and users; engage users to contribute more than only ratings, reviews and comments; enable users to create new tags and build their own collections of linkable tags (industrial example: Last.fm).

Design Principle 3: Provide balance between public and private

Best practices include:

1. User profiles and photos help to enhance the human element in a recommender system (industrial example: Amazon.com).
2. Controllable public profile for users is another necessity in a sociable recommender system which relies heavily on users' interaction to produce better recommendation and explanation alike (industrial example: Amazon.com and Last.fm).

Chapter 5: Conclusion and Discussion

5.1 Conclusion

We have presented in this thesis a novel design framework in Chapter 2 and tested it through an experiment prototype built upon it. From our survey on existing industrial best practices and collected user behavioral data and feedbacks from our experiment and survey, we can conclude that our design framework sheds some new light on the issue of how to solve some of the most important design problems about the recommender system's explanation interface. Hence by looking from three different angles, namely presentation, explanation and interaction, a recommender system designer is equipped with a well-mapped blueprint to help overcome the difficulties of how to position an e-commerce website's recommender application.

Clearly, the "black-box" model (e.g., HollywoodVideo.com) is losing its charm in both the modern e-business setting and the academic research field. Therefore, pure concentration on the presentation perspective is against the current trend forward. Explanation functions, though already implemented by many commercial recommender systems like IMDb.com, have not been performing very impressively on engaging users to generate more interest-reflecting data or give a full play to the community power which makes "mutual explanation" possible. By connecting the "black-box" presentation and the transparent explanation, we extend our research work to a new level which is the "interaction" aspect. Both pre-experiment interview feedbacks and the post-experiment survey results highly support our hypothesis that explanations based on social navigation and community are much more favored by the new generation of online shoppers. And hence, the interaction part of a recommender's explanation interface harnesses users' own motivations to find clues to justify system's recommendations and support their own purchasing decisions; while at the mean time users motivated to contribute to other users with their own shared personal interest-reflecting data such as reviews, comments and tags.

In addition to our proposed design framework, we also summarize in Chapter 4, from literature review and post-experiment survey, some design principles with corresponding visualization techniques to guide recommender system designers. By selecting some of the best practices, future recommender system can either follow a much balanced approach like Amazon.com or an extremely social networking oriented example such as Last.fm as these two websites are the most impressive industry leaders in the sense of designing an all-round explanation interface.

5.2 Discussion and Future Work

Since social networking sites have recently been booming in numbers, grassroots participation is a typical feature in the era of Web 2.0. Recommender system used to be playing a role of technical proxy for a social recommendation process, but now it has to shift its design focus in a more sociable direction (Sinha, 2007). Sinha (2007) also points out the following four challenges facing future recommender system designers:

1. Motivating participation
2. Giving users fine-grained control
3. Making item information available
4. Making recommendations transparent

Seen from the above summarized challenges, we feel glad that our design framework of explanation interface fits well with the recommender system's general development direction. An explanation interface in its interaction perspective shall be able to motivate users to share more, review more, rate more and tag more. Interaction also deals with fine-grained control over users' personal data and profiles, as we suggest in our design principles that users shall have full control of their public profile while hiding sensitive privacy data. The presentation aspect of our design framework makes sure that product information is easily accessible when the most important linking part, the explanation itself, plays a role of making the whole recommendation process transparent and hence increases users' trust toward the system. By adapting such a social and participatory design concept, recommendations are no longer being "pushed" to users. However, most of the time, users will feel enjoyable interacting with the recommender and "pull" product information, recommendations and explanations by themselves.

However, still some problems are facing future explanation interface designers and we believe the most significant reason is that this time the explanation function shall be the driving force for the underlying algorithms to go forward. As in the future, the recommender system's users will tend to generate more content like reviews, comments and tags beside the traditional rating and purchasing data, the recommendation algorithms shall fully take advantage of the additional interest-reflecting information while at the same time mapping out a transparent route to explain the recommender system's underlying logic. And which might be even better is that after the recommendation algorithms help to locate those like-minded peers for the active user, the explanation interface then shall take over and automatically generate social networking choices for the active user. Furthermore, another possible issue might rise after a sophisticated online community is built. The trust issue used to be placed upon recommendations themselves, but later on how to guarantee inter-user trust inside a mutual-explanatory community may become a new concern.

Appendix I

Pre-experiment interview questions

1. Do you have any online shopping experience?

2. Have you ever used any recommender system? (Show IMDb.com and the simple interface of our prototype if the respondent has not.)

3. What's your general viewpoint on the recommender system?

4. Do you like the way it presents the recommendation, such as a Top-N list?

5. Do you like to know why these recommendations are made for you?

6. What type of explanation do you prefer? (item-item similarity, user-user similarity, neighbors' rating, statistical summarization, social navigation)

7. Would you like to give your ratings or reviews to the system in order to improve its accuracy?

8. If some friends from a reliable online community recommend something for you, will you be more likely to accept it than if it was suggested by a recommender system?

9. Which recommendation is more stratifying? A machine-generated one or a personal one given by a trustworthy friend?

10. Would you like to make part of your online profile (like ratings, purchase history, reviews) public so that people can assess your credibility when the recommender system is recommending something based on your tastes?

Background information of the five interviewees

Age	Gender	Major	Recommender system experience
26	Female	Spanish	Yes (Amazon.com)
32	Male	Computer Science	Yes (Taobao.com)
26	Female	EU Study	No
26	Female	ICT in Business	No
24	Female	Law	No

Appendix II

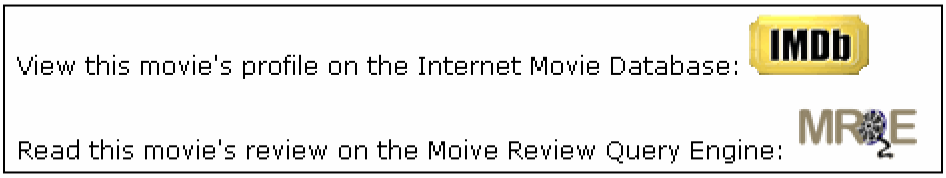
Post-experiment survey questionnaire

*Note:

1. "v" means "visualization technique"
2. Please give your ratings from 1 (very bad) to 5 (very good) to the following functions that appear in this prototype recommender system for question No.1 to No. 17.

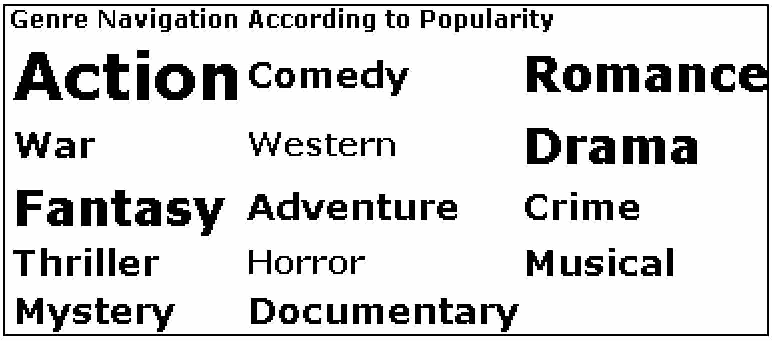
Presentation

1. Links to a movie's profile on IMDb.com and its critique on MRQE.com.

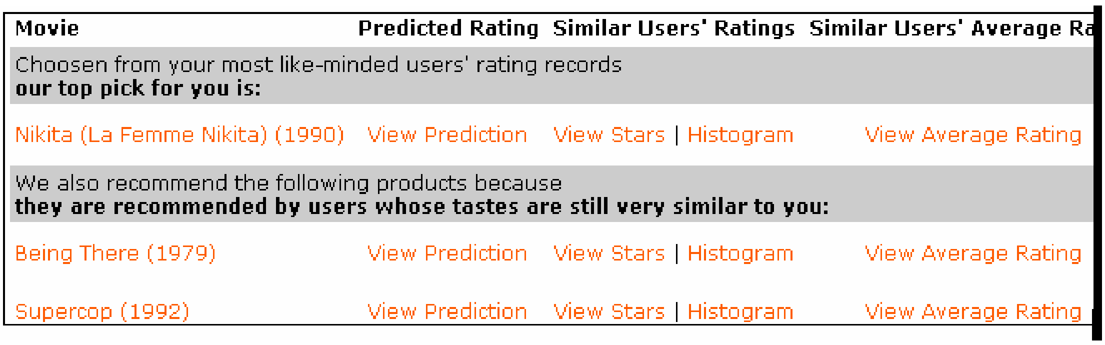


2. Prediction assertion (v: text highlighting).

3. Cloud style quick genre navigation function (v: cloud).



4. Recommendations in categorization which provides different groups of choices (v: categorization).



5. Recommendations in a plain list without categorization (screenshot from an old version of this prototype) (v: table).

Top 10 recommended movies for you!

Movie	Price	Buy It?
Nikita (La Femme Nikita) (1990)	€7	Buy it now!
Being There (1979)	€7	Buy it now!
Supercop (1992)	€7	Buy it now!
Field of Dreams (1989)	€7	Buy it now!
Army of Darkness (1993)	€7	Buy it now!
Star Trek: First Contact (1996)	€7	Buy it now!
Rumble in the Bronx (1995)	€7	Buy it now!
39 Steps, The (1935)	€7	Buy it now!
Magnificent Seven, The (1954)	€7	Buy it now!
North by Northwest (1959)	€7	Buy it now!

Explanation

6. Past accuracy performance assertion (v: text highlighting).

According to previous statistics, the prediction accuracy of our recommender system is 93%.

7. Average rating by all users in stars (v: star)

Average rating by all users: 

8. Average rating by all users in plain text (v: plain text)

'Toy Story (1995)' was rated by 458 members, with an average rating of **3.87 / 5**.

9. Average rating by users with similar tastes (v: star)

Average rating by users with similar tastes to you: 

10. Reviewer's ranking. (v: text highlighting)

Similarity List
Display up to 10 of your

ID	Title
804	Top-10 Reviewer
291	Top-100 Reviwer
256	Top-100 Reviwer
532	Top-100 Reviwer
216	Top-1000 Reviwer

11. Genre match display (v: text highlighting).

Genre Match: This movie's genre-- **action** matches one of your favorites!

12. Help page with the description of system inner logic (v: plain text).

Things you Need to Know about our Recommender System

What technology are we using?

By comparing your rating records with those of all the registered users, our recommender system employs collaborative filtering technology to find like-minded people (called "neighbors") for you.

How do we choose recommended items?

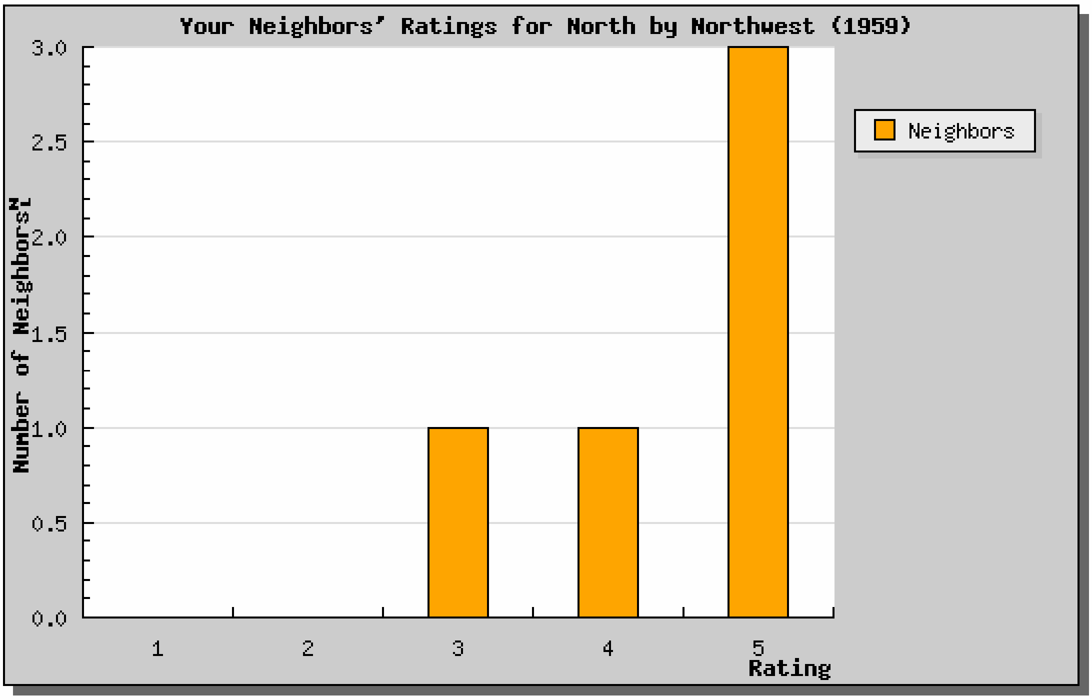
From your neighbors' viewing records, our system selects the best movies which are lilely to interest you the most as our recommendations!

13. Similar users' ratings in stars (v: star).

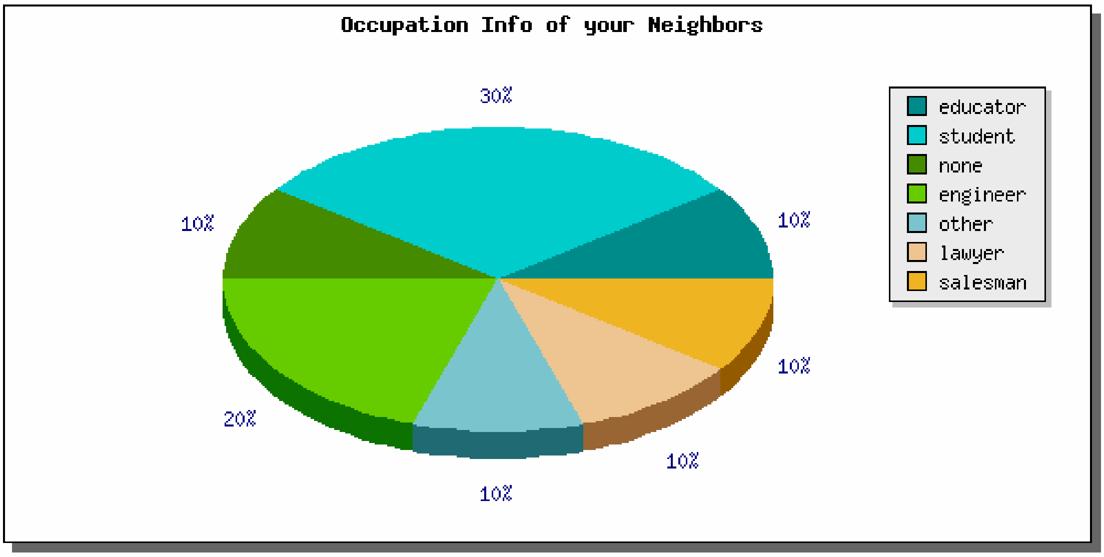
Similar Users' Ratings for "Supercop (1992)"

User ID	Title	Similarity to You	Rating	User Profile
804	Top-10 Reviewer	93 %		View his/her Profile
291	Top-100 Reviwer	91 %		View his/her Profile
256	Top-100 Reviwer	90 %		View his/her Profile
532	Top-100 Reviwer	89 %		View his/her Profile
216	Top-1000 Reviwer	88 %		View his/her Profile
654	Top-1000 Reviwer	88 %		View his/her Profile
178	Top-100 Reviwer	87 %		View his/her Profile
267	Top-1000 Reviwer	87 %		View his/her Profile
339	Top-100 Reviwer	87 %		View his/her Profile
457	Top-100 Reviwer	87 %		View his/her Profile

14. Similar users' ratings in histograms (v: histogram).



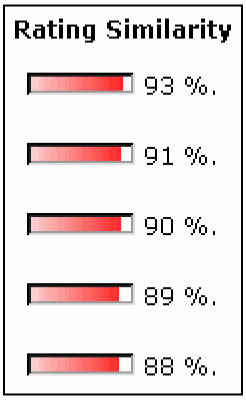
15. Similar users' demographic backgrounds in pie chart (v: pie chart).



16. Similar users' demographic backgrounds in tables (v: table).

Age	Sex	Occupation	Favorite genre
39	M	educator	comedy
19	M	student	comedy
35	F	none	action
20	M	student	comedy

17. Similar users' similarity to you in bar chart and percentage. (v. bar chart)



18. Generally speaking, how satisfactory are you with the explanation function and its visualization as a whole?
    a) Very unsatisfied, hard to understand
    b) Unsatisfied, unhelpful
    c) Neutral
    d) Interesting and helpful
    e) Very satisfied

Interaction

19. How much do you like the function that enables you to navigate through the users with similar tastes to you and tag a like-minded one?
    a) It is useless
    b) Not very useful
    c) Neutral
    d) Good function, nice to have
    e) Excellent, a must-have for a recommender system
20. How much do you like the function that enables you to view the rating histories of the users with similar tastes?
    a) It is useless
    b) Not very useful
    c) Neutral
    d) Good function, nice to have
    e) Excellent, a must-have for a recommender system
21. How much would you like a function that the recommender system preferentially selects movies based on your tagged friends prior to other less familiar users?
    a) It is useless
    b) Not very useful
    c) Neutral
    d) Good function, nice to have
    e) Excellent, a must-have for a recommender system
22. How much would you like a function that let you trace the ratings and reviews from tagged friends and base your movie-going decision on these data?
    a) It is useless
    b) Not very useful

c) Neutral
d) Good function, nice to have
e) Excellent, a must-have for a recommender system

23. How much would you like a function that enables you to participate in the discussion with other users and write reviews about the movies you watched?
    a) It is useless
    b) Not very useful
    c) Neutral
    d) Good function, nice to have
    e) Excellent, a must-have for a recommender system
24. How much would you like a function that enables you to rate others' movie reviews and let other users to rate your reviews too?
    a) It is useless
    b) Not very useful
    c) Neutral
    d) Good function, nice to have
    e) Excellent, a must-have for a recommender system
25. How much would you like your limited profiles (like reviews, ratings and demographic background) to be viewed by other users or tagged friends?
    a) No, I am very cautious with my privacy
    b) OK, but I shall have full control to decide what part of my profile can be public to whom
26. Any other comments are extremely welcome!

References

Anderson, C. (2004). "The Long Tail". The Wired Magazine, Issue 12.10 - October 2004 http://web.archive.org/web/20041127085645/http://www.wired.com/wired/archive/12.10/tail.html

Ardissono, L., Gena, C., Torasso, P., Bellifemine, F., Chiarotto, A., and Difino, A. (2004). User Modeling and Recommendation Techniques for Personalized Electronic Program Guides. In L. Ardissono, A. Kobsa and M.T. Maybury (Eds.), Personalized Digital Television: Targeting Programs to Individual Viewers. Dordrecht, the Netherlands: Kluwer Academic Publishers, Pages 5-7.

Bilgic, M. and Mooney, R.J. (2005). Explaining Recommendations: Satisfaction vs. Promotion. In Proceedings of the Workshop Beyond Personalization 2005, pages 13-18.

Bonhard P. (2004). Improving Recommender Systems with Social Networking. In Proceedings Addendum of CSCW 2004, Nov. 6-10, Chicago, IL.

Buchanan, B. and Shortcliffe, E. (1984). Rule-Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project. Reading, MA: Addison Wesley Publishing Company. 1984, pages 331-338.

Burke, R. (1999). The Wasabi Personal Shopper: A Case-Based Recommender System. In Proceedings of the 11th Conference on Innovative Applications of Articificial Intelligence (pages. 844-849). American Association for Artificial Intelligence.

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction. 12(4), pages 331-370.

Edmunds, A. and Morris, A. (2000). The Problem of Information Overload in Business Organisations: A Review of the Literature. International Journal of Information Management, pages 17-28.

Kaser, O. and Lemire, D. (2007). Tag-Cloud Drawing: Algorithms for Cloud Visualization. In proceedings of Tagging and Metadata for Social Information Organization (WWW 2007), 2007.

Haynes, S. R. (2001) Explanation in Information Systems: A Design Rationale Approach. PhD thesis, The London School of Economics and Political Science, Department of Information Systems and Department of Social Psychology

Herlocker, J., Konstan, J., Borchers and A., Riedl, J. (1999). An Algorithmic Framework for Performing Collaborative Filtering. In proceedings of the 1999 Conference on Research and Development in Information Retrieval. Aug. 1999, pages 241-250.

Herlocker, J., Konstan, J.A., and Riedl, J. (2000). Explaining Collaborative Filtering Recommendations. ACM 2000 Conference on CSCW.

Herlocker, J., Konstan J. (2001). Content-Independent, Task-Focused Recommendation, IEEE Internet Computing, pages 5-6

Herlocker, J, Konstan, J, and Riedl, J. (2002). An Empirical Analysis of Design Choices in Neighborhood-based Collaborative Filtering Algorithms, Information Retrieval, 5 (2002), Kluwer Academic Publishers, pages 287-310.

Johnson, J. and Johnson, P. (1993). Explanation Facilities and Interactive Systems. In Proceedings of Intelligent User Interfaces ’93.pages 159-166.

Kay, J. and Kummerfeld, B. (2006). Scrutability, User Control and Privacy for Distributed Personalization, in Kobsa, A., R. K. Chellappa and S. Spiekermann, (eds.) Proceedings of PEP06, CHI 2006 Workshop on Privacy-Enhanced Personalization, Montreal, Canada

Koenemann, J. and Belkin, N. (1996). A Case for Interaction: A Study of Interactive Information Retrieval Behavior and Effectiveness. In Proceedings of the Human Factors in Computing Systems Conference. ACM Press, NY, pages 205-212

Konstan, J. (2005). Joseph Konstan on Human-Computer Interaction: Recommender Systems, Collaboration, and Social Good, Ubiquity, Volume 6, Issue 10 (March 24 - April 4, 2005)

Lemire, D. and Maclachlan, A. (2005). Slope One Predictors for Online Rating-Based Collaborative Filtering. In Proceedings of SIAM Data Mining (SDM'05), 2005, pages 471-475.

Linden G., Smith B. and York J. (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering, IEEE Internet Computing, v.7 n.1, pages 76-80, January 2003

McCarthy, K., Reilly, J., McGinty, L. and Smyth, B. (2004) Thinking Positively - Explanatory Feedback for Conversational Recommender Systems. In: P. Gervas and K. M. Gupta (eds) Proceedings of the Workshop on Explanation in CBR at the Seventh European Conference on Case-Based Reasoning (ECCBR-04), pages 115-124. Madrid. Spain.

McNee, S.M., Riedl, J and Konstan J.A. (2006). "Making Recommendations Better: An Analytic Model for Human-Recommender Interaction". In the Extended Abstracts of the 2006 ACM Conference on Human Factors in Computing Systems (CHI 2006), Montreal, Canada, April 2006, pages 1103-1108

MovieLens dataset (2007) http://www.grouplens.org/node/73

P3P Project. (2007). Platform for Privacy Preferences (P3P) Project. http://www.w3.org/P3P/

Pazzani, M.J. and Billsus, D. Learning and Revising User Profile: The Identification of Interesting Web Sites. Machine Learning, 27, pages 313-331. 1997

Pine, J. (1993). Mass Customization. Harvard Business School Press. Boston, Massachusetts

Pine, J. and Gilmore, H. (1999). The Experience Economy. Harvard Business School Press. Boston, Massachusetts

Pu, P. and Chen, L. (2006). Trust Building with Explanation Interfaces. In Proceedings of the 11th International Conference on Intelligent User Interfaces, pages 93-100.

Pack, T. (1998). Visualizing Information: Visualization Systems Data Management. Database Magazine, 21(1), Feb, 1998, pages 47-49

Resnick, P. and Varian, H.R. (1997). Recommender Systems. Communications of ACM 40, 3 pages 56-58.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. GroupLens: An Open

Architecture for Collaborative Filtering of Netnews. In Proceedings of CSCW ’94, Chapel Hill, NC. 1994, pages 175-186.

Rich, E. (1998). User Modeling via Stereotypes. In M.T. Maybury and W. Wahlster (Eds.), Readings in Intelligent User Interfaces. San Francisco: Morgan Kaufmann Publishers Inc. pages 329-341.

Riedl, J. (2001). Personalization and Privacy. IEEE Internet Computing, 5, pages 29-31.

Riedl, J. and Konstan, J. A. (2002). Word of Mouse: The Hidden Marketing Power of Collaborative Filtering. Warner Business Books, 2002.

Riesbeck, C.K. and Schank, R. (1989). Inside Case-Based Reasoning. Northvale, NJ: Lawrence Erlbaum Associates.

Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl, J. (2000). Analysis of Recommender Algorithms for E-Commerce. In proceedings of the ACM E-Commerce 2000 Conference. Oct. 17-20, 2000, pages 158-167.

Schafer, J.B., Konstan, J.A., and Riedl, J. (2001), “E-Commerce Recommender Applications,” Data Mining and Knowledge Discovery, vol 5 nos.1/2, pages 115-152.

Shardanand, U. and Maes, P. (1995). Social Information Filtering: Algorithms for Automated "Word of Mouth". In Proceedings of Human factors in computing systems 1995 (pages 210-217). New York: ACM.

Sinha, R. (2007). Design Strategies for Recommender Systems. UIE Web App Summit 2007

Swearingen, K. and Sinha, R. (2002). Interaction Design for Recommender Systems. In Designing Interactive Systems 2002.

Tintarev, N. and Masthoff, J. (2007). A Survey of Explanations in Recommender Systems G Uchyigit (ed), Workshop on Recommender Systems and Intelligent User Interfaces associated with ICDE'07 (Istanbul, Turkey), IEEE van Setten, M. (2005). Supporting People in Finding Information-Hybrid Recommender Systems and Goal-Based Structuring. PhD Thesis. Telematica Instituut Fundamental Research Series, vol. 016. Enschede, the Netherlands: Telematica Instituut

Vogoo (2007). Vogoo CF Engine, http://www.Vogoo-api.com

Ware, C. (2000). Information Visualization: Perception for Design. San Francisco: Morgan Kaufmann (Academic Press), 2000

Wikipedia (2007a). Internet marketing. Retrieved July 18, 2007, from Wikipedia.org Web site: http://www. wikipedia.org /topic/internet-advertising

Wikipedia (2007b). Human-computer interaction. Retrieved July 19, 2007, from Wikipedia.org Web site: http://www. wikipedia.org /topic/human-computer-interaction-1

Wikipedia (2007c). Visualization. Retrieved July 30, 2007, from Wikipedia.org Web site: http://www. wikipedia.org/topic/visualization

Zaslow, J. (2002). “If TiVo thinks you are gay, here’s how to set it straight – Amazon.com knows you, too, based on what you buy; why all the cartoons?” The Wall Street Journal, sect. A, p. 1, November 26, 2002.