Intelligent Tagging Interfaces: Beyond Folksonomy

Jesse Vig
Department of Computer Science and Engineering
University of Minnesota
jvig@cs.umn.edu

ABSTRACT

This paper summarizes our work on using tags to broaden the dialog between a recommender system and its users. We present two tagging applications that enrich this dialog: *tagsplanations* are tag-based explanations of recommendations provided by a system to its users, and *Movie Tuner* is a conversational recommender system that enables users to provide feedback on movie recommendations using tags. We discuss the design of both systems and the experimental methodology used to evaluate the design choices.

ACM Classification: H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces; H.5.2 [Information Interfaces and Presentation]: User Interfaces

General terms: Design, Experimentation, Human Factors

Keywords: tagging, recommender systems, explanations, conversational recommenders

1 Introduction

Recommender systems and their users engage in a form of dialog: users communicate their preferences to the system and the system provides recommendations in returns. The channel of communication between user and system is generally narrow; in a typical recommender system, users articulate their preferences through numeric ratings of items, and the system communicates its recommendations as lists of items and predicted ratings. While this limited form of communication offers benefits such as efficiency and low cognitive load, it omits information that can yield a more informed recommendation process. For example, what are the qualities of a particular item that led the system to recommend it to a particular user? And why did a user accept or reject a particular recommendation?

The goal of my research is to use tags to widen the channels of communication between a recommender system and its users. We have developed two applications with this goal in mind. *Tagsplanations* are tag-based explanations of recommendations provided by the system to its users, for example, "We recommend *Fargo* because it is tagged with *quirky* and we believe you like quirky movies." *Movie Tuner* is a conversational recommender system in which users give feedback on recommendations using tags, for example, "I want a movie like *Pulp Fiction* but more *action*".

One design challenge is to provide the underlying intelligence needed by these applications. For example, how does the system know that Alice likes *quirky* movies, or that *Kill Bill Vol. 1* has more *action* than *Pulp fiction*? Rather than rely on domain experts for this knowledge, we design systems that generate this knowledge automatically. We develop machine learning algorithms that infer properties of items and users based on user-contributed content including ratings, tags, and user reviews.

2 TAGSPLANATIONS: EXPLAINING RECOMMENDATIONS USING TAGS

While much of the research in recommender systems has focused on improving the accuracy of recommendations, recent work suggests a broader set of goals including trust, user satisfaction, and transparency [11, 1, 8, 5]. A key to achieving this broader set of goals is to *explain* recommendations to users. While recommendations tell users what items they might like, explanations reveal *why* they might like them. An example is the "Why this was recommended" feature on Netflix[1]. Netflix explains movie recommendations by showing users similar movies they have rated highly in the past. Research shows that explanations help users make more accurate decisions [1], improve user acceptance of recommendations [5], and increase trust in the recommender system [8].

While many different types of explanation facilities exist, they all seek to show how a recommended item relates to a user's preferences. As Figure 1 illustrates, a common technique for establishing the relationship between user and recommended item is to use an *intermediary entity*. An intermediary entity is needed because the direct relationship between user and item is unknown, assuming that the user has not yet tried the item. In the Netflix example above, the intermediary entity is the list of previously rated movies shown in the explanation. The relationship between the user and these movies is that he or she has rated them positively. The relationship between these movies and the recommended movie is that other users who liked these movies also liked the recommended movie.

2.1 Related Work

Existing explanations of recommendations fall into one of three categories: *item-based*, *user-based*, and *feature-based*,

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

UIST'10, October 3-6, 2010, New York, NY, USA.

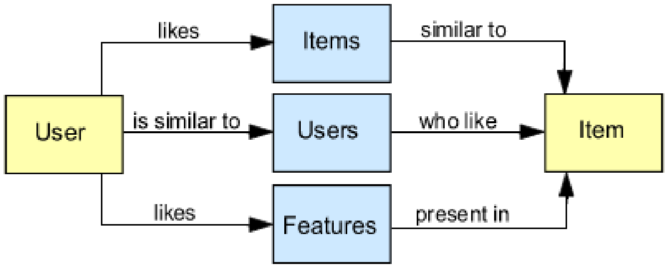

Figure 1: Intermediary entities (center) relate user to recommended item.

depending on the type of intermediary entity used to relate the user to the recommended item. In item-based explanations like the Netflix example, a set of items serves as the intermediary entity. User-based explanations utilize other users as intermediary entities. For example, Herlocker et al. designed an explanation that shows a user how other users with similar taste rated the recommended item [5]. Feature-based approaches use features or characteristics of the recommended item as intermediary entities. For example, one movie recommender prototype uses movie features including genre, director, and cast to justify recommendations [10].

2.2 Design

We present a new type of explanation based on tags, called a *tagsplanation* [12]. The intermediary entity for tagsplanations is a tag or a set of tags. Tagsplanations identify the relationship between user and tag, which we call *tag preference* and the relationship between tag and recommended item, which we call *tag relevance*. Tag preference represents a user's sentiment towards a given tag, e.g. "Maria likes quirky", while tag relevance describes how strongly a tag describes an item, e.g. "Fargo is a very quirky movie". The tagsplanation connects tag relevance and tag preference to explain a recommendation, e.g. "We recommend Fargo because it is very quirky and Maria likes quirky movies."

We have developed novel algorithms for estimating both tag preference and tag relevance. Tag preference is computed based on a user's ratings for movies with a particular tag. For example, if Maria has given 5-star ratings to all movies tagged with *action*, one can infer that Maria has a strong preference for *action*. Tag relevance is computed based on the correlation between tag preference and item preference. For example, if users with a strong preference for *action* tend to rate "Die Hard" highly, one can infer that *action* is highly relevant to "Die Hard". We chose to use preference correlation to represent tag relevance, because it directly relates tag preference (the relationship between user and tag) with item preference (the relationship between user and item).

We deployed tagsplanations within a survey conducted on MovieLens[2], a movie recommender system in which users rate movies and receive recommendations in return. Figure 2 shows an example of the tagsplanation interface. Tag relevance is represented as a bar of varying length, while tag preference is depicted as a number of stars rounded to the nearest half. An arrow indicates sort order of the tags.

[2]http://www.movielens.org

Figure 2: Tagsplanation for *Eraserhead*.

2.3 Evaluation

We conducted a within-subjects study of four variations of the interface. The interfaces varied on whether they showed just tag relevance, just tag preference, or both, and on how they sorted the results. We asked users how well the tagsplanations helped them (1) understand why an item was recommend, (2) decide if they would like the item, and (3) determine if an item fit their current mood. Survey results showed that both tag relevance and tag relevance played a role in achieving these three goals. Further, over 80% of subjects evaluating the best-performing interface agreed that the tagsplanation helped them achieve the three goals.

We also asked users to evaluate specific tags shown in the tagsplanations. We found that users preferred subjective tags like *funny*, *poignant*, *witty* to factual ones. Finally, we asked users to evaluate the accuracy of tagsplanations for movies they had already seen. Over 80% of respondents agreed with the statement "Overall this is a good explanation given my knowledge of the movie", when shown tagsplanations for movies they had seen in the past.

3 MOVIE TUNER: A CONVERSATIONAL RECOMMENDER BASED ON TAGS

Conversational recommenders enable users to provide multifaceted feedback on recommendations [6, 2, 3, 9]. A classic example is the Entree restaurant recommender, in which users critique restaurant recommendations by asking for restaurants that are, for example, "less expensive", "more quiet", or "less traditional" [2]. In response, the system provides new recommendations that satisfy the user's critiques. The recommendation process thus becomes a conversation between system and user.

Conversational recommenders typically offer the user a relatively compact set of dimensions for critiquing recommendations, and these dimensions are generally chosen by designers of the system. For example, in one conversational recommender, users critique recommendations for digital cameras based on manufacturer, zoom level, memory, weight, resolution, size, case, and price [7].

We have designed a new type of conversational recommender system based on community tags. In this system, the set of critique dimensions correspond to the tags that users have applied in the system. For example, on MovieLens users have applied tags such as *funny*, *violent*, and *action* to various movies. In a conversational recommender system based on tags, users may ask for a movie that is "more *funny*", "less *violent*", or with "more *action*". In contrast to the compact set of system-engineered dimensions typically provided by conversational recommender systems, tags provide a broad range of feedback that may be useful in subjective domains such as movies, music, or literature.

3.1 Related Work

Recent work has explored using tags to "steer" recommendations in the music domain [4]. Our work differs from [4] in several ways. First, we predict tag relevance using a training set of tag relevance values collected in a survey, and we evaluate a variety of regression models against this data set. Second, we design an interface that is considerably different from the tag cloud visualization used in [4]. Third, we compare multiple algorithms for choosing candidate critiques to show to users. Finally, we evaluate our design in a live user study involving thousands of users.

3.2 Design

In this section, we describe the design of Movie Tuner, a tag-based conversational movie recommender system that we deployed inside of MovieLens. Figure 3 shows the Movie Tuner interface, discussed in greater detail below.

Tag Relevance One challenge is to determine each movie's position along the various critique dimensions. For example, how *violent* is "Forrest Gump", or how much *action* is in "Titanic"? We refer to the relevance of a particular tag to a particular movie as *tag relevance*. Tag relevance values are used in two ways: (1) to show users the relevance of a tag to a movie, and (2) to provide an ordering of movies with respect to each tag in order to respond to a critique such as "more action than Pulp Fiction".

We developed several algorithms that predict tag relevance for an arbitrary $(movie, tag)$ pair. In order to train these algorithms, we first collected a gold standard data set of tag relevance values from a user survey. In the survey, MovieLens users rated the tag relevance values of over 50,000 $(movie, tag)$ pairs on a scale of 1 to 5. We constructed features from several types of community-contributed content for predicting tag relevance, and we compare four different regression algorithms that predict tag relevance based on these features. The best performing regression model achieved a mean absolute error of 0.85 on the tag relevance values from the survey (scale of 1-5), based on 10-fold cross validation.

Selecting Candidate Critiques Because there are over 1,500 dimensions (i.e. tags), we wished to display a smaller set of candidate dimensions to reduce the users' cognitive load

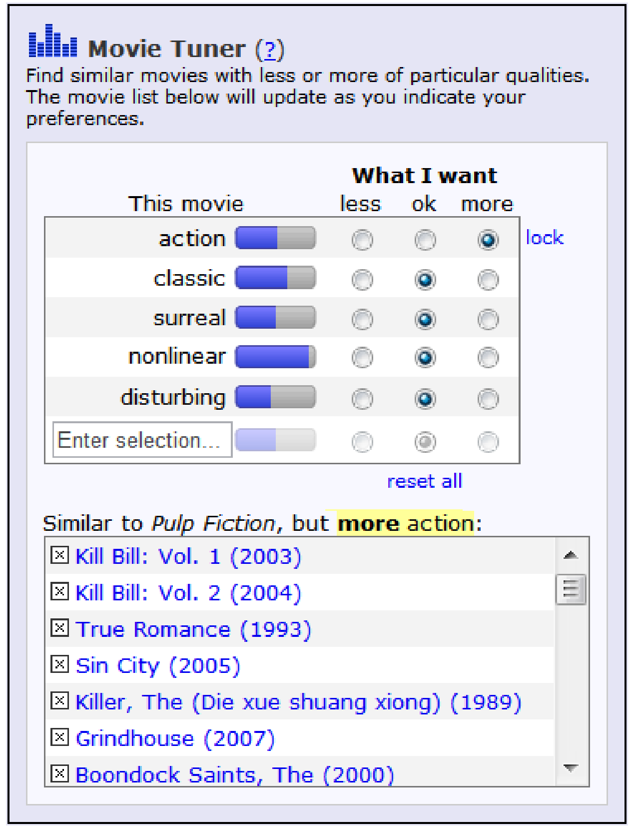

Figure 3: Movie Tuner interface for *Pulp Fiction*.

when choosing critiques. We developed two algorithms to select a subset of tags to display to users. The *entropy-based* approach chooses the subset of tags that most evenly divides up the space of neighboring movies when used in critiques. The *relevance-based* approach chooses tags that have highest relevance to the recommended movie. In order to generate a diverse set of tags, we also enforce a maximum pairwise similarity between the tags displayed.

Critique Satisfaction We designed an algorithm for choosing new recommendations in response to users' critiques. The algorithm chooses movies based on two criteria. First, the new movies must be sufficiently different along the dimension being critiqued. For example, when a user asks for a movie less violent than "Pulp Fiction", they expect that the new recommendation will be notably less violent. Second, the new recommendation should be somewhat similar to the original recommendation. If the user starts with "Pulp Fiction" and asks for less violence, they will likely be unhappy if the new recommendation is for, say, "The Sound of Music."

Interface As shown in Figure 3, we visualize tag relevance as a bar of varying length. Tag relevance values are displayed for each of the candidate critique dimensions selected by the system, in this case *action*, *classic*, *surreal*, *nonlinear*, and *disturbing*. Users may inquire about the relevance of any of the other 1,500 tags in the system by entering them in the auto-complete text box ("Enter selection"). Once entered in the text box, they are displayed at the top along with the system-selected tags.

Users apply critiques by clicking the *less* or *more* radio button next to any of the tags displayed. Users may combine critiques (e.g. "less violence and more action") by clicking *lock* next to the initial critique. The movies that satisfy the critique(s) are then displayed at the bottom. In the example in Figure 3, the user has selected the critique "more action". In response the system displays a list of movies at the bottom that are similar to "Pulp Fiction", but with more action. The user may then click on any one of these movies to be brought to its Movie Tuner.

3.3 Evaluation

We began a field study of Movie Tuner on 7/14/2010. The goal of the study was to answer two research questions: (1) What is the best way to choose new recommendations in response to users' critiques? (2) What is the best method for selecting a subset of critique dimensions to display to users? To answer these questions we designed two variations of the algorithm for choosing new recommendations as well as two variations of the algorithm for choosing a subset of tags to show. We randomly assign each MovieLens user to one of the two algorithms, resulting in 4 experimental groups.

In the 3 weeks since Movie Tuner was deployed, 1,600 users have viewed Movie Tuner, and 700 users have applied a total of 7,000 critiques. Once we collect additional data, we will compare user activity between each of the 4 experimental conditions. We will measure dependent variables including number of critiques applies and number of click-throughs on resulting recommendations. Finally we will conduct a survey measuring user satisfaction with Movie Tuner and compare the results between each of the experimental conditions.

4 FUTURE WORK

I plan to continue to develop tag-based applications that help users and recommender systems communicate better with one another. Specifically, I am interested in the following research directions:

- Using tags to visualize a recommender system's model of a user's preferences and to enable the user to correct this model. For example, a recommender system might think that Fred likes zombie movies because he has rated many zombie movies highly in the past, whereas in fact Fred saw these movies as a child and is no longer interested in zombies. With an application that supports explicit feedback on the user model through tags, the Fred could simply tell the system that he is no longer interested in movies with a *zombie* tag.
- Designing a graphical front end for Movie Tuner. Movies could be visualized as nodes on a graph and critiques as connecting edges. Alternatively, the set of movies could be projected onto a 2-dimensional map with critiques as "directions" in that space.
- Extending the logic of Movie Tuner to other types of user feedback that are traditionally unilateral. For example, if a user writes a review of "Pulp Fiction" and uses the word *violent* repeatedly, the system might respond by showing the user movies similar to "Pulp Fiction" that are less violent.

5 REFERENCES

1. M. Bilgic and R. J. Mooney. Explaining recommendations: Satisfaction vs. promotion. In *Proceedings of Beyond Personalization Workshop, IUI*, 2005.
2. R. D. Burke, K. J. Hammond, and B. C. Young. The findme approach to assisted browsing. *IEEE Expert*, 12:32–40, 1997.
3. B. Faltings, P. Pu, M. Torrens, and P. Viappiani. Designing example-critiquing interaction. In *IUI '04: Proceedings of the 9th international conference on Intelligent user interfaces*, pages 22–29. ACM, 2004.
4. S. J. Green, P. Lamere, J. Alexander, F. Maillet, S. Kirk, J. Holt, J. Bourque, and X.-W. Mak. Generating transparent, steerable recommendations from textual descriptions of items. In *RecSys '09: Proceedings of the third ACM conference on Recommender systems*, pages 281–284. ACM, 2009.
5. J. Herlocker, J. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 2000. CHI Letters 5(1).
6. G. Linden, S. Hanks, and N. Lesh. Interactive assessment of user preference models: The automated travel assistant. In *In Proceedings of the Sixth International Conference on User Modeling*, pages 67–78. Springer, 1997.
7. K. McCarthy, J. Reilly, L. McGinty, and B. Smyth. Experiments in dynamic critiquing. In *IUI '05: Proceedings of the 10th international conference on Intelligent user interfaces*, pages 175–182. ACM, 2005.
8. R. Sinha and K. Swearingen. The role of transparency in recommender systems. In *CHI '02: CHI '02 extended abstracts on Human factors in computing systems*, pages 830–831. ACM, 2002.
9. B. Smyth, L. McGinty, J. Reilly, and K. McCarthy. Compound critiques for conversational recommender systems. In *WI '04: Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 145–151. IEEE Computer Society, 2004.
10. N. Tintarev. Explanations of recommendations. In *RecSys '07: Proceedings of the 2007 ACM conference on Recommender systems*, pages 203–206. ACM, 2007.
11. N. Tintarev and J. Masthoff. A survey of explanations in recommender systems. In *IEEE 23rd International Conference on Data Engineering Workshop*, pages 801–810, 2007.
12. J. Vig, S. Sen, and J. Riedl. Tagsplanations: Explaining recommendations using tags. In *IUI '09: Proceedings of the 13th International Conference on Intelligent User Interfaces*, pages 47–56. ACM, 2009.