FilmTrust: Movie Recommendations using Trust in Web-based Social Networks

Jennifer Golbeck, James Hendler
University of Maryland, College Park
College Park, Maryland
{golbeck, hendler}@cs.umd.edu

Abstract

In this paper, we present FilmTrust, a website that integrates Semantic Web-based social networks, augmented with trust, to create predictive movie recommendations. We show how these recommendations are more accurate than other techniques in certain cases, and discuss this technique as a mechanism of Semantic Web interaction.

1. Introduction

Trust in social networks on the Semantic Web is a topic that has gained increased interest in the last few years. Using FOAF as the basis for the social network, trust has been encoded using the FOAF Trust Module[1] or computed from other information. With these trust values as a starting place, several algorithms for inferring trust relationships have been introduced. This analysis of Semantic Web-based social networks has produced results, but their usefulness in the space of user interaction has not been fully addressed.

In this paper, we present FilmTrust, a website that integrates Semantic Web-based social networking into a movie recommender system. We begin with a description of the FilmTrust website, followed by an analysis of its features. TidalTrust, a trust network inference algorithm, is used as the basis for generating predictive ratings personalized for each user. The accuracy of the recommended ratings is shown to outperform both a simple average rating and the ratings produced by a common recommender system algorithm. Theoretically and through a small user study, some evidence is also established that supports a user benefit from ordering reviews based on the users' trust preferences.

[1] http://trust.mindswap.org/ont/trust.owl

1.1. Background and Related Work

Social Network data, represented using the FOAF Vocabulary[1], is some of the most prevalent data on the Semantic Web. TidalTrust[2] is an algorithm for inferring trust relationships. Using a recursive search with weighted averages, it can take two people in the network and generate a recommendation about how much one person should trust the other, based on the paths that connect them in the network, and the trust ratings on those paths. Part of our recommender system relies on inferred trust ratings, and this is the algorithm that is used there.

Recommender systems help users identify items of interest. These recommendations are generally made in two ways: by calculating the similarity between items and recommending items related to those in which the user has expressed interest, or by calculating the similarity between users in the system and recommending items that are liked by similar users. This latter method is also known as collaborative filtering.

Collaborative filtering has been applied in many contexts, and FilmTrust is not the first to attempt to make predictive recommendations about movies. MovieLens [5], Recommendz [6], and Film-Conseil [7] are just a few of the websites that implement recommender systems in the context of films. Herlocker, et al. [8] present an excellent overview of the goals, datasets, and algorithms of collaborative filtering systems. However, FilmTrust is unlike the approach taken in many collaborative filtering recommender systems in that its goal is not to present a list of good items to users; rather, the recommendations are generated to suggest how much a given user may be interested in an item that the user already found. For this to work, there must be a measure of how closely the item is related to the user's preferences.

In this work, the aim is to use trust ratings within the social network as the basis for making calculations about similarity. For this technique to be successful, there must be a correlation between trust and user similarity. Abdul-Rahman and Hailes [9] showed that in a predefined context, such as movies, users develop social connections with people who have similar preferences. These results were extended in work by Ziegler and Lausen [10]. Their work showed a correlation between trust and user similarity in an empirical study of a real online community.

Other work has touched on trust in recommender systems, including [11] and [12]. These works address the use of trust within systems where the set of commonly rated items between users is sparse. That situation leads to a breakdown in correlation-based recommender system algorithms, and their work explores how incorporating even simple binary trust relationships can increase the coverage and thus the number of recommendations that can be made.

2. The FilmTrust Website

The social networking component of the website requires users to provide a trust rating for each person they add as a friend. When creating a trust rating on the site, users are advised to rate how much they trust their friend about movies. In the help section, when they ask for more help, they are advised to, "Think of this as if the person were to have rented a movie to watch, how likely it is that you would want to see that film."

Part of the user's profile is a "Friends" page,. In the FilmTrust network, relationships can be one-way, so users can see who they have listed as friends, and vice versa . If trust ratings are visible to everyone, users can be discouraged from giving accurate ratings for fear of offending or upsetting people by giving them low ratings. Because honest trust ratings are important to the function of the system, these values are kept private and shown only to the user who assigned them.

The other features of the website are movie ratings and reviews. Users can choose any film and rate it on a scale of a half star to four stars. They can also write free-text reviews about movies.

Social networks meet movie information on the "Ratings and Reviews" page shown in Figure 2. Users are shown two ratings for each movie. The first is the simple average of all ratings given to the film. The "Recommended Rating" uses the inferred trust values, computed with TidalTrust on the social network, for the users who rated the film as weights to calculate a weighted average rating. Because the inferred trust values reflect how much the user should trust the opinions of the person rating the movie, the weighted average of movie ratings should reflect the user's opinion. If the user has an opinion that is different from the average, the rating calculated from trusted friends – who should have similar opinions – should reflect that difference. Similarly, if a movie has multiple reviews, they are sorted according to the inferred trust rating of the author. This presents the reviews authored by the most trusted people first to assist the user in finding information that will be most relevant.

3. Site Personalization

3.1. Computing Recommended Movie Ratings

One of the features of the FilmTrust site that uses the social network is the "Recommended Rating" feature. As figure 7.3 shows, users will see this in addition to the average rating given to a particular movie.

The "Recommended Rating" is personalized using the trust values for the people who have rated the film (the *raters*). The process for calculating this rating is very similar to the process for calculating trust ratings presented in chapter 6. First, the system searches for raters that the source knows directly. If there are no direct connections from the user to any raters, the system moves one step out to find connections from the user to raters of path length 2. This process repeats until a path is found. The opinion of all raters at that depth are considered. Then, using TidalTrust, the trust value is calculated for each rater at the given depth. Once every rater has been given an inferred trust value, only the ones with the highest ratings will be selected; this is done by simply finding the maximum trust value calculated for each of the raters at the selected depth, and choosing all of the raters for which that maximum value was calculated. Finally, once the raters have been selected, their ratings for the movie (in number of stars) are averaged. For the set of selected nodes S, the recommended rating r from node s to movie m is the average of the movie ratings from nodes in S weighted by the trust value t from s to each node:

$$r_{sm} = \frac{\sum_{i \in S} t_{si} r_{im}}{\sum_{i \in S} t_{si}}$$

This average is rounded to the nearest half-star, and that value becomes the "Recommended Rating" that is personalized for each user.

As a simple example, consider the following:

- Alice trusts Bob 9
- Alice trusts Chuck 3
- Bob rates the movie "Jaws" with 4 stars
- Chuck rates the movie "Jaws" with 2 stars

Then Alice's recommended rating for "Jaws" is calculated as follows:

$$\frac{t_{Alice->Bob} * r_{Bob->Jaws} + t_{Alice->Chuck} r_{Chuck-Jaws}}{t_{Alice->Bob} + t_{Alice->Chuck}} = \frac{9*4+3*2}{9+3} = \frac{42}{12} = 3.5$$

3.2. Determining the Accuracy of Recommended Ratings

For each movie the user has rated, the recommended rating can be compared to the actual rating that the user assigned. In this analysis, we also compare the user's rating with the average rating for the movie, and with a recommended rating generated by an automatic collaborative filtering (ACF) algorithm. There are many ACF algorithms, and one that has been well tested, and which is used here, is the classic user-to-user nearest neighbor prediction algorithm based on Pearson Correlation [5]. If the trust-based method of calculating ratings is best, the difference between the personalized rating and the user's actual rating should be significantly smaller than the difference between the actual rating and the average rating.

On first analysis, it did not appear that that the personalized ratings from the social network offered any benefit over the average. The difference between the actual rating and the recommended rating (call this ∂r) was not statistically different than the difference between the user's actual rating and the average rating (call this ∂a). The difference between a user's actual rating of a film and the ACF calculated rating (∂cf) also was not better than ∂a in the general case. A close look at the data suggested why. Most of the time, the majority of users actual ratings are close to the average. This is most likely due to the fact that the users in the FilmTrust system had all rated the AFI Top 50 movies, which received disproportionately high ratings. A random sampling of movies showed that about 50% of all ratings were within the range of the mean +/- a half star (the smallest possible increment). For users who gave these near-mean rating, a personalized rating could not offer much benefit over the average.

However, the point of the recommended rating is more to provide useful information to people who *disagree* with the average. In those cases, the personalized rating should give the user a better recommendation, because we expect the people they trust will have tastes similar to their own [10].

To see this effect, ∂a, ∂cf, and ∂r were calculated with various minimum thresholds on the ∂a value. If the recommended ratings do not offer a benefit over the average rating, the ∂r values will increase at the same rate the ∂a values do. The experiment was conducted by limiting ∂a in increments of 0.5. The first set of comparisons was taken with no threshold, where the difference between ∂a and ∂r was not significant. As the minimum ∂a value was raised it selected a smaller group of user-film pairs where the users made ratings that differed increasingly with the average. Obviously, we expect the average ∂a value will increase by about 0.5 at each increment, and that it will be somewhat higher than the minimum threshold. The real question is how the ∂r will be impacted. If it increases at the same rate, then the recommended ratings do not offer much benefit over the simple average. If it increases at a slower rate, that means that, as the user strays from the average, the recommended rating more closely reflects their opinions. Figure 1 illustrates the results of these comparisons.

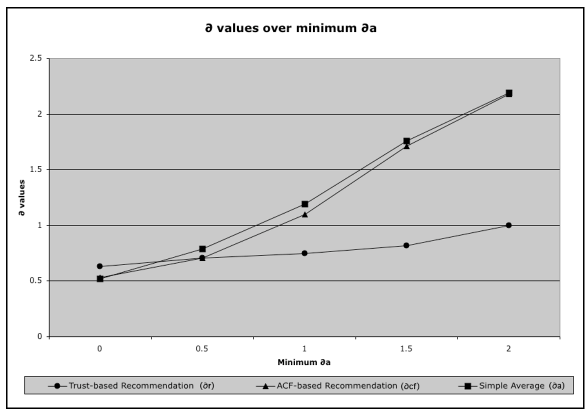

Fig. 1. The increase in ∂ as the minimum ∂a is increased. Notice that the ACF-based recommendation (∂cf) closely follows the average (∂a). The more accurate Trust-based recommendation (∂r) significantly outperforms both other methods.

Notice that the ∂a value increases about as expected. The ∂r, however, is clearly increasing at a slower rate than ∂a. At each step, as the lower threshold for ∂a is increased by 0.5, ∂r increases by an average of less than 0.1. A two-tailed t-test shows that at each step where the minimum ∂a threshold is greater than or equal to 0.5, the recommended rating is significantly closer to the actual rating than the average rating is, with $p<0.01$. For about 25% of the ratings assigned, ∂a<0.5, and the user's ratings are about the same as the mean. For the other 75% of the ratings, ∂a>0.5, and the recommended rating significantly outperforms the average. As is shown in Figure 1, ∂cf closely follows ∂a. For ∂a<1, there was no significant difference between the accuracy of the ACF ratings and the trust-based recommended rating. However, when the gap between the actual rating and the average increases, for ∂a>=1, the trust-based recommendation outperforms the ACF as well as the average, with p<0.01. Because the ACF algorithm is only capturing overall correlation, it is tracking the average because most users' ratings are close to the average.

Figure 2 illustrates one of the examples where the recommended value reflects the user's tastes. "A Clockwork Orange" is one of the films in the database that has a strong collective of users who hated the movie, even though the average rating was 3 stars and many users gave it a full 4-star rating. For the user shown, ∂a=2.5 – a very high value – while the recommended rating exactly matches the user's low rating of 0.5 stars. These are precisely the type of cases that the recommended rating is designed to address.

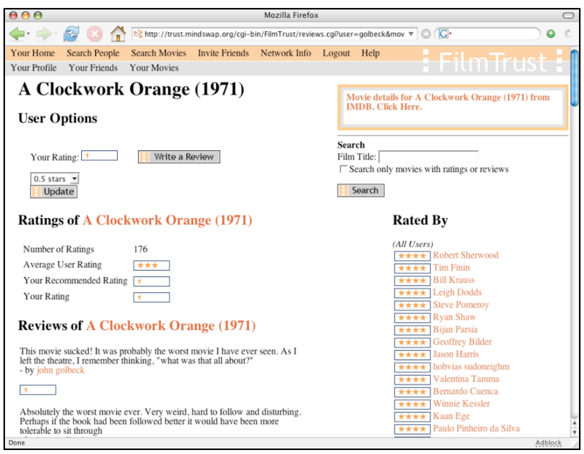

Fig. 2. A user's view of the page for "A Clockwork Orange," where the recommended rating matches the user's rating, even though ∂a is very high (∂a = 2.5).

Thus, when the user's rating of a movie is different than the average rating, it is likely that the recommended rating will more closely reflect the user's tastes. When the user has different tastes than the population at large, the recommended rating reflects that. When the user has tastes that align with the mean, the recommended rating also aligns with the mean. Based on these findings, the recommended ratings should be useful when people have never seen a movie. Since they accurately reflect the users' opinions of movies they have already. Because the rating is personalized, originating from a social network, it is also in line with other results [3,4] that show users prefer recommendations from friends and trusted systems.

One potential drawback to creating recommendations based solely on relationships in the social network is that a recommendation cannot be calculated when there are no paths from the source to any people who have rated a movie. This case is rare, though, because as long as just *one* path can be found, a recommendation can be made. In the FilmTrust network, when the user has made at least one social connection, a recommendation can be made for 95% of the user-movie pairs.

The purpose of this work is not necessarily to replace more traditional methods of collaborative filtering. It is very possible that a combined approach of trust with correlation weighting or another form of collaborative filtering may offer equal or better accuracy, and it will certainly allow for higher coverage. However, these results clearly show that, in the FilmTrust network, basing recommendations on the expressed trust for other people in the network offers significant benefits for accuracy.

3.3. Presenting Ordered Reviews

In addition to presenting personalized ratings, the experience of reading reviews is also personalized. The reviews are presented in order of the trust value of the author, with the reviews from the most trustworthy people appearing at the top, and those from the least trustworthy at the bottom. The expectation is that the most relevant reviews will come from more trusted users, and thus they will be shown first.

Unlike the personalized ratings, measuring the accuracy of the review sort is not possible without requiring users to list the order in which they suggest the reviews appear. Without performing that sort of analysis, much of the evidence presented so far supports this ordering. The definition of trust has been used to support many of the calculations made throughout this dissertation. That definition also supports the ordering of reviews. Trust with respect to movies means that the user believes that the trusted person will give good and useful information about the movies. The analysis also suggests that more trusted individuals will give more accurate information. It was shown there that trust correlates with the accuracy of ratings. Reviews will be written in line with ratings (i.e. a user will not give a high rating to a movie and then write a poor review of it), and since ratings from highly trusted users are more accurate, it follows that reviews should also be more accurate.

A small user study with 9 subjects was run on the FilmTrust network. Preliminary results show a strong user preference for reviews ordered by the trustworthiness of the rater, but this study must be extended and refined in the future to validate these results.

4. Conclusions and Discussion

Within the FilmTrust website, trust in social networks has been used to personalized the user experience. Trust took on the role of a recommender system forming the core of an algorithm to create predictive rating recommendations for movies. The accuracy of the trust-based predicted ratings in this system is significantly better than the accuracy of a simple average of the ratings assigned to a movie and also the recommended ratings from a Person-correlation based recommender system.

Overall, we believe that FilmTrust is an example of how the Semantic Web, and Semantic trust networks in particular, can be exploited to refine the user experience. By using the Semantic Web data in computations, interaction with the Semantic Web becomes integrated into common tasks, and enhances existing tools.

Acknowledgements

This work, conducted at the Maryland Information and Network Dynamics Laboratory Semantic Web Agents Project, was funded by Fujitsu Laboratories of America -- College Park, Lockheed Martin Advanced Technology Laboratory, NTT Corp., Kevric Corp., SAIC, the National Science Foundation, the National Geospatial-Intelligence Agency, DARPA, US Army Research Laboratory, NIST, and other DoD sources.

References

[1] FOAF Vocabulary: http://xmlns.com/foaf/0.1/

[2] Golbeck, Jennifer (2005) "Computing and Applying Trust in Web-Based Social Networks," Ph.D. Dissertation, University of Maryland, College Park.

[3] Sinha, R., and Swearingen, K. (2001) "Comparing recommendations made by online systems and friends." *In Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries* Dublin, Ireland.

[4] Swearingen, K. and R. Sinha. (2001) "Beyond algorithms: An HCI perspective on recommender systems," *Proceedings of the ACM SIGIR 2001 Workshop on Recommender Systems*, New Orleans, Louisiana.

[5] Herlocker , Jonathan L., Joseph A. Konstan , John Riedl, Explaining collaborative filtering recommendations, *Proceedings of the 2000 ACM conference on Computer supported cooperative work,* p.241-250, December 2000, Philadelphia, Pennsylvania, United States.

[6] Garden, Matthew, and Gregory Dudek (2005) Semantic feedback for hybrid recommendations in Recommendz. *Proceedings of the IEEE International Conference on e-Technology, e-Commerce, and e-Service (EEE05*), Hong Kong, China, March 2005.

[7] Perny, P. and J. D. Zucker. Preference-based Search and Machine Learning for Collaborative Filtering: the ``Film-Conseil'' recommender system. *Information, Interaction , Intelligence,* 1(1):9-48, 2001.

[8] Herlocker , Jonathan L., Joseph A. Konstan , Loren G. Terveen , John T. Riedl, (2004) Evaluating collaborative filtering recommender systems, *ACM Transactions on Information Systems (TOIS)*, v.22 n.1, p.5-53, January 2004.

[9] Abdul-Rahman, A. and Hailes, S. 2000. Supporting trust in virtual communities. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*. Maui, HW, USA.

[10] Ziegler, Cai-Nicolas, Georg Lausen (2004) Analyzing Correlation Between Trust and User Similarity in Online Communities" *Proceedings of Second International Conference on Trust Management*, 2004.

[11] Massa, P., P. Avesani. 2004. Trust-aware Collaborative Filtering for Recommender Systems. In *Proceedings of the International Conference on Cooperative Information Systems (CoopIS) 2004*.

[12] Massa, P., B. Bhattacharjee. 2004. Using Trust in Recommender Systems: an Experimental Analysis. In *Proceedings of iTrust2004 International Conference*.