Generating Predictive Movie Recommendations from Trust in Social Networks

Jennifer Golbeck[1]
[1]University of Maryland, College Park
A.V. Williams Building
College Park, Maryland 20742
golbeck@cs.umd.edu

Abstract. Social networks are growing in number and size, with hundreds of millions of user accounts among them. One added benefit of these networks is that they allow users to encode more information about their relationships than just stating who they know. In this work, we are particularly interested in trust relationships, and how they can be used in designing interfaces. In this paper, we present FilmTrust, a website that uses trust in web-based social networks to create predictive movie recommendations. Using the FilmTrust system as a foundation, we show that these recommendations are more accurate than other techniques when the user's opinions about a film are divergent from the average. We discuss this technique both as an application of social network analysis, as well as how it suggests other analyses that can be performed to help improve collaborative filtering algorithms of all types.

1 Introduction

Web-based social networks are growing in size and number every day. A website that maintains a comprehensive list of these networks shows 133 networks with well over 165,000,000 user accounts among them. Users spend hours maintaining personal information, blog entries, and lists of social contacts. The benefit of this time investment is vague. While a small percentage of these networks are dedicated to building business contacts, most are for entertainment purposes.

While entertainment may motivate users to maintain a presence in these web-based social networks, there is great potential to utilize the social data for enhancing end user applications. Since the networks are web-based, the information is largely publicly available. Many of these networks are beginning to output their members' profiles using FOAF, a Semantic Web vocabulary for representing social networks, means that the data is not only available but easily readable by applications.

One space that these social networks can be integrated into applications is in creating interfaces that act "intelligently" with respect to the user's social connections. This can be further refined by looking at specific features of social relationships. Nearly half of the social networks found in the aforementioned list provide some means for users to add information about their relationships with others. This could include the type of relationship (e.g. "friend", "sibling", "co-worker", etc.), the strength of the relationship (e.g. "acquaintance", "good friend", "best friend", etc.), or how much the users trust the people they know. Our research is specifically focused on this trust relationship because it has many features that make it ideal for integrating into socially intelligent interfaces.

Specifically, we are will use social trust as the basis for a recommender system. For this technique to be successful, there must be a correlation between trust and user similarity. Abdul-Rahman and Hailes [1] 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 [2] that showed a correlation between trust and user similarity in an empirical study of a real online community.

Furthermore, there is evidence to support that users will prefer systems with recommendations that rely on social networks and trust relationships over similarity measures commonly used for making recommendations. Research has shown that people prefer recommendations from friends to those made by recommender systems [3] and that users prefer recommendations from systems they trust [4]. By producing recommendations through the use of trust in social networks, both of those user preferences are addressed. Recommendations come through a network of friends, and are based on the explicit trust expressed by the user.

In this paper, we present FilmTrust, a website that integrates 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 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 correlation-based collaborative filtering algorithm. Theoretically and through a small user study, some evidence is also developed that supports a user benefit from ordering reviews based on the users' trust preferences.

2. Background and Related Work

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.

Before making any computations with trust in social networks, it is vitally important to know what trust is. Social trust depends on a host of factors which cannot be easily modeled in a computational system. Past experience with a person and with their friends, opinions of the actions a person has taken, psychological factors impacted by a lifetime of history and events (most completely unrelated to the person we are deciding to trust or not trust), rumor, influence by others' opinions, and motives to gain something extra by extending trust are just a few of these factors. For trust to be used as a rating between people in social networks, the definition must be focused and simplified. We adopt this as the definition of trust for our work: trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome. The action and commitment does not have to be significant. We could say Alice trusts Bob regarding movies if she chooses to watch a film (commits to an action) that Bob recommends (based on her belief that Bob will not waste her time).

Other work has touched on trust in recommender systems, including [9] and [10]. 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.

3. Experimental Platform: The FilmTrust Website

The FilmTrust system, at http://trust.mindswap.org/FilmTrust, is a website that combines a web-based social network and a movie rating and review system. It's membership forms the basis for our investigation.

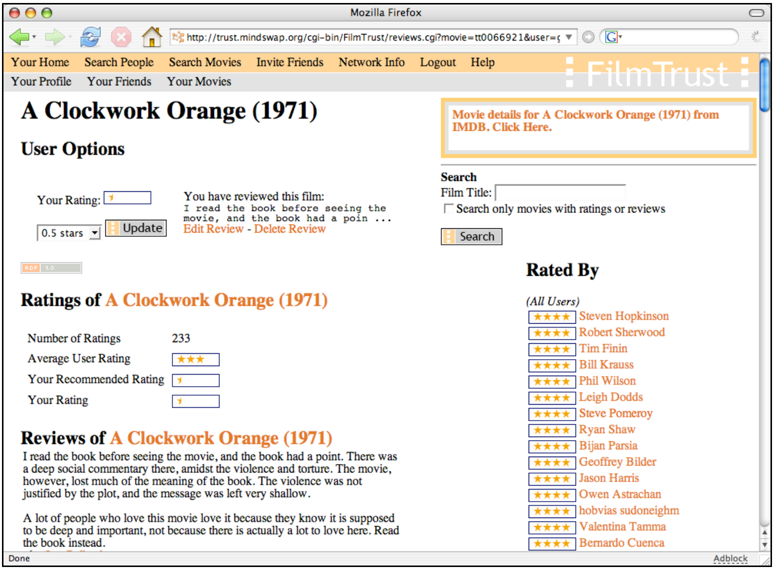

Figure 1. A user's view of the page for "A Clockwork Orange," where the recommended rating matches the user's rating, even though the average is quite different.

3.1 Social Networking with FilmTrust

The social networking component of the website allows users to maintain a list of friends who are also in the network. Our system 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. Users are advised to consider trust in this context: "...if the person were to have rented a movie to watch, how likely it is that you would want to see that film."

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.

3.2 Movie Features

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 1. 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.

3.3 Computing Recommended Movie Ratings

One of the features of the FilmTrust site that uses the social network is the "Recommended Rating" feature. As Figure 1 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 (direct or inferred) that the user has the people who have rated the film (the raters). If a user Alice has directly assigned a trust rating to another user, Bob, then the trust value is known. If Alice has not rated Bob, we need to infer how much she might trust him. Trust inference systems are a growing area of interest. In this application, we utilize TidalTrust, a breadth first search-based algorithm that outputs an inferred trust value by finding paths form Alice to Bob and composing the trust values found along those paths. Details of that algorithm are beyond the scope of this paper, but can be found in [11] and [12].

To compute the recommended movie rating, the FilmTrust system first searches for raters who the user 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 trust values 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:

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

4 Experimental Setup and Design

We are interested in knowing if the trust-based movie ratings offer a benefit to the users, and if so, in what instances. To check this, we used the data users have entered into the FilmTrust system.

4.1 Experimental Setup and Design

The FilmTrust user base was used as the foundation for our experiments. When joining the network, members were informed that their participation was part of a research project, and they consented to allow their data to be used within experiments. The system has just over 500 members.

Members were invited by friends who were already members and also found out about the website from postings in movie related forums. There is a strong Semantic Web component to the website (social network and movie information is all published in RDF), so members were frequently recruited from this circle of interest. Subjects ranged in age from 14 to 79, with an average age of 32. Subjects were 29% female and 71% male.

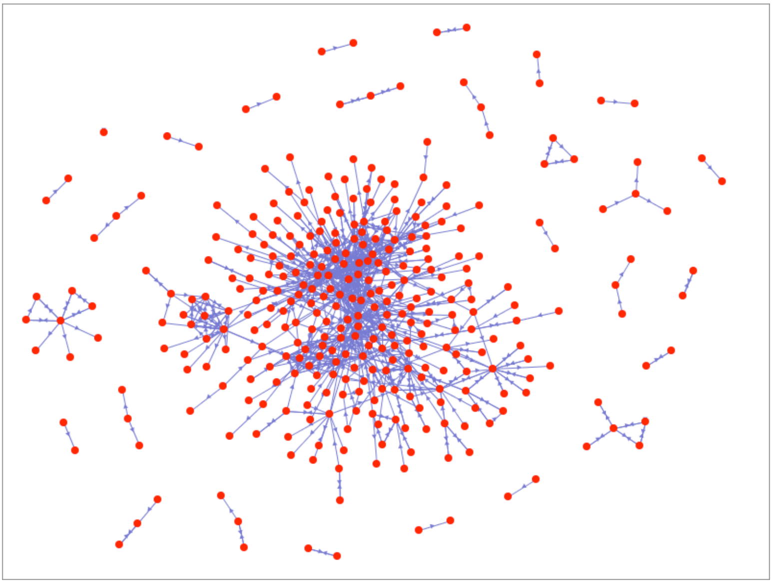

Figure 2. A visualization of the FilmTrust social network

FilmTrust users have created approximately 11,250 ratings and movie reviews for 1,250 different movies. For each movie, the average rating was computed as the simple average of all the ratings assigned to the film. To ensure that a common set of movies were rated, users were asked during the registration process to assign ratings to any movies they had seen the top 50 films AFI Top 100 Films list [13].

Not all of these members are connected into the social network. Approximately 150 of the 500 members do not have any social connections. Their participation is limited to entering data about movies. Of the members who are participating in the social network, most are connected into a strong central core, with a scattering of small groups. A spring-embedded visualization of the social network structure is shown in Figure 2

4.2 Experimental Results

To determine the effectiveness of the recommended ratings, we compare to see how closely they resemble the actual ratings a user has assigned to a film. We use the absolute difference between the recommended rating and actual rating as our measure. 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. We label these measures as follows:

- ∂r – the absolute difference between the user's rating and the trust-based recommended rating
- ∂a – the absolute difference between the user's rating and the average rating
- ∂cf – the absolute difference between the user's rating and the recommended rating from the collaborative filtering algorithm

Because the recommended ratings rely on using the trust values in the social network, we were only able to make this comparison for users with social connections, approximately 350 of the 500 total users. For each user, we selected each movie and computed the ∂ values. In the end, we made comparisons for a total of 1152 movies.

On first analysis, it did not appear that that the trust-based ratings that utilized the social network were any more accurate than average. The difference between the actual rating and the recommended rating (∂r) was not statistically different than the difference between the user's actual rating and the average rating (∂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, one of our initial motivations for creating the trust-based recommended ratings was to help 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; that is, the user's rating had to be at least ∂a stars different from the average rating. 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 3 illustrates the results of these comparisons.

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 user's actual rating than the average rating is, with $p<0.01$. For about 25% of the ratings assigned, $\partial a<0.5$, and the user's ratings are about the same as the mean. For the other 75% of the ratings, $\partial a>0.5$, and the recommended rating significantly outperforms the average.

As is shown in Figure 3, ∂cf closely follows ∂a. For $\partial a<1$, 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 $\partial a>=1$, 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 1 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, $\partial a=2.5$ – 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.

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.

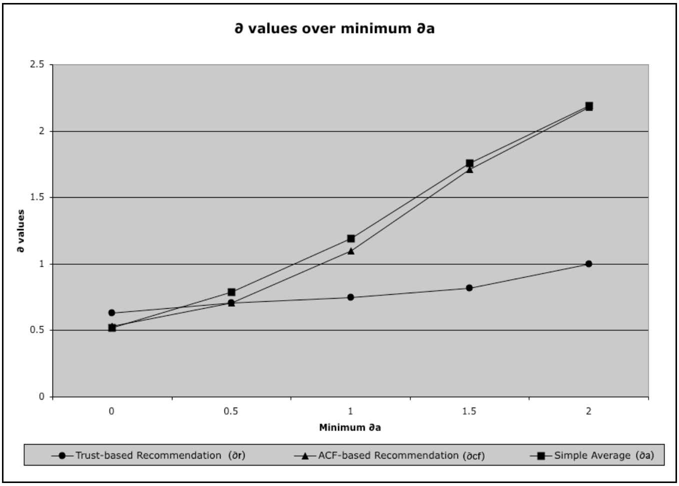

Figure 3. The increase in ∂ as the minimum ∂a is increased. Notice that the ACF-based recommendation (∂cf) closely follows the average (∂a). The more accurate Trust-based recommendation (∂r) significantly outperforms both other methods.

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 user 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.

In addition, the quality of results is dependent on users assigning accurate trust values to people in the system. If the trust ratings become too noisy, they cease to be an effective grounds for making recommendations. The FilmTrust system is still relatively small compared to other social networks, which can have tens of thousands up to millions of members. It remains to be seen how well this technique will work on larger networks. We have not yet been given access to trust values in some of the larger networks, and that analysis will be necessary to verify that user behavior will support our approach.

4.3 Presenting Ordered Reviews

In addition to presenting personalized ratings, the experience of reading reviews is also personalized. The reviews are presented to the user 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.

Figure 4. Reviews of "E.T." sorted according to the trust value that the user has for each author. Note that the ratings of the ordering also corresponds to how closely the reviewers' ratings of the film correspond with the user's rating, even though that was not considered in choosing the ordering.

For example, Figure 4 shows the reviews of "E.T." ordered for a user. The reviews from more trusted people appear at the top of the list, and less trust people are further down. Notice that the user's rating is 2 stars. Even though the reviewers' rating were not considered in the ordering, they are ordered as well; the reviewers with ratings that most closely match the user's rating are shown first, and the reviews further down in the list are different from the user. This supports the premise that ordering reviews by trust rating will show users the opinions more relevant to their own perspective 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. 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 pilot study with 9 subjects was run on the FilmTrust network. Subjects were shown the reviews for a movie and asked to order them according to how closely they matched the subject's opinion. This was frequently identical to the ordering based on trust value, and the variations the did occur were typically small. When shown the trust-based ordering, our small sample of users had a universally strong positive reaction. While these preliminary results show a strong user preference for reviews ordered by the trustworthiness of the rater, this study must be extended and refined in the future to validate these results.

5. Conclusions and Discussion

Within the FilmTrust website, trust in social networks has been used as the foundation for generating predictive movie recommendations, The accuracy of the trust-based predictive ratings in this system is significantly better than the accuracy of a simple average of the ratings assigned to a movie. The trust system also outperforms the recommended ratings from a Person-correlation based recommender system.

Overall, we believe that FilmTrust is an example of how trust and social networks can be exploited to refine the user experience. By using the social network data in computations, the efforts users are already putting to web-based socializing can be harnessed to enhance existing tools. 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.

There are many future steps for both refining this work and taking it in future directions. One step is to do a deeper comparison with the most advanced collaborative filtering algorithms. We have chosen a common, basic algorithm for comparison in this study. Since our goal was not to out perform collaborative filtering techniques, but rather to show that the trust-based recommendations were useful,

One current project we have underway is investigating how users assign trust in social networks. The results presented here show that it is not merely correlation of opinions; if that were the case, we would have seen equivalent performance between the trust-based recommendations and the collaborative filtering recommendations. We believe that users assign trust based more on agreement on outliers, rather than on overall agreement. For example, say Bob and Alice both hated the "Lord of the Rings" movies, loved "From Justin to Kelly" , but otherwise had a large variation in movies about which they are less enthusiastic. We believe that they may trust each other more than they would trust someone with a higher overall correlation but who disagreed about "Lord of the Rings" and "From Justin to Kelly". Understanding which features of user profiles correlate to higher trust values will give social insight, but it also suggests how different features of profile similarity can be incorporated into collaborative filtering algorithms to improve their accuracy even when social networks are unavailable.

6. 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.

7. References

1. 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.

2. 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.
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. Massa, P., P. Avesani. 2004. Trust-aware Collaborative Filtering for Recommender Systems. In Proceedings of the International Conference on Cooperative Information Systems (CoopIS) 2004.
10. Massa, P., B. Bhattacharjee. 2004. Using Trust in Recommender Systems: an Experimental Analysis. In Proceedings of iTrust2004 International Conference.
11. Golbeck, Jennifer. 2005. Computing and Applying Trust in Web-Based Social Networks, Ph.D. Dissertation, University of Maryland, College Park.
12. Golbeck, Jennifer. 2005. Personalizing Applications through Integration of Inferred Trust Values in Semantic Web-Based Social Networks. Proceedings of Semantic Network Analysis Workshop. Galway, Ireland.
13. American Film Institute, "100 Years, 100 Movies" http://www.afi.com/tvevents/100years/movies.aspx