Recommender System from Personal Social Networks

David Ben-Shimon[1], Alexander Tsikinovsky[1], Lior Rokach[2], Amnon Meisles[3], Guy Shani[3], and Lihi Naamani[2]

[1] Deutsche Telekom Laboratories at Ben-Gurion University
[2] Department of Information System Engineering, Ben-Gurion University
[3] Department of Computer Science, Ben-Gurion University

Summary. Recommender systems are found in many modern web sites for applications such as recommending products to customers. In this paper we propose a new method for recommender system that employs the users' social network in order to provide better recommendation for media items such as movies or TV shows. As part of this paper we develop a new paradigm for incorporating the feedback of the user's friends. A field study that was conducted on real subjects indicates the strengths and the weaknesses of the proposed method compared to other simple and classic methods. The system is envisioned to function as a service for recommending personalized media (audio, video, print) on mobile phones, online media portals, sling boxes, etc. It is currently under development within Deutsche Telekom Laboratories - Innovations of Integrated Communication projects.

1 Introduction

Systems that recommend items to users are becoming popular and can be found in many modern web sites for applications such as recommending products to customers in e-commerce sites, recommending TV programs to users of interactive TV and presenting personalized advertisements. There are two dominating approaches [8] for creating recommendation; Collaborative Filtering (CF) and Content-Based (CB) recommendations. CF considers the recommended items only by a unique identifier and recommends items that were purchased together, ignoring any attribute of the item. CB recommendations are based on an item profile which is commonly defined by the attributes of the item without considering acts of purchasing. Each of these methods has its pros and cons but it seems that a hybrid approach can overcome most of the disadvantages of the two methods. Another method for providing recommendations is based on Stereotypes [3]. Stereotypes are a way to define an abstract user that has general properties similar to a set of real users. Stereotypes are used in recommender systems for varying purposes, ranging from initial user profile creation to generating recommendations [8]. All methods use some type of a user profile or user model for recommendation. CF systems usually maintain a vector of rated items while CB systems maintain a rated set of item attributes.

K.M. Węgrzyn-Wolska and P.S. Szczepaniak (Eds.): Adv. in Intel. Web, ASC 43, pp. 307–315, 2007.
springerlink.com

A social network is a graph that represents connected users. The two most important characteristics of social networks are: who the people in the network are and how they are connected to each other. The usual notion of connection between people in a virtual community is related to direct social interaction [4]. This makes social networks useful for providing recommendations. Assuming that connected people have also some common interests, this paper proposes a method for recommending media items based on a personal social network of each user.

The rest of this paper is organized as follow: Section 2 surveys the current status of recommender systems and recommendations-based social network. Section 3 describes in detail the proposed method. Section 4 describes the field study and the comparative results of the proposed approach for recommending items. Section 5 presents a discussion and our conclusions.

2 Recommender Systems

With the explosion of data available online, recommender systems became very popular. While there are many types of recommender systems ranging from manually predefined un-personalized recommendations to fully automatic general purpose recommendation engines, two dominating approaches have emerged - Collaborative Filtering and Content Based recommendations.

2.1 Collaborative Filtering

Collaborative filtering stems from the idea that people looking for recommendations often ask for the advice of friends. While on the internet the population that can supply advises is very large, the problem shifts into identifying what part of this population is relevant for the current user.

CF methods identify similarity between users based on items they have rated and recommend new items that similar users have liked. CF algorithms vary by the method they use to identify similar users. Originally Nearest-Neighbor approaches based on the Pearson Correlation, computing similarity between users directly over the database of user-item ratings were implemented. Modern systems tend to learn some statistical model from the database and then use it for recommending previously rated items to a new audience. Model-based approaches usually sacrifice some accuracy in favor of a rapid recommendation generation process [5], better scaling up to modern applications. The main advantage of CF is that it is independent of the specification of the item and can therefore provide recommendations for complex items which are very different yet are often used together. The major drawback of this approach is the inability to create good recommendations for new users that have not yet rated many items, and for new items that were not rated by many users. This drawback also known as the *'cold start problem'*.

2.2 Content-Based Recommendation

The ideas of Content-Based recommendations originate in the field of information filtering, where documents are searched given some analysis of their text. Items are hence defined by a set of features or attributes. Such systems define a user using preferences over this set of features, and obtain recommendations by matching user profiles and item profiles looking for best matches.

Some researchers [8] separate methods that learn preferred attributes from rated items from methods that ask the user to specify his preferences over item attributes, but we refer to all methods that recommend based on item attribute preferences as CB recommendation. CB approaches rarely learn statistical models and usually match user profiles and item profiles directly. User and item profiles are very sensitive to profile definitions - which attributes are relevant and which attributes should be ignored. It is also difficult to create an initial profile of the user, specifying the interests and preferences of the user; Users are reluctant to provide thorough descriptions of the things they like and do not like. It is also possible that users are unaware of their preferences. For example, a user cannot know whether she likes an actor she never seen. In fact the acquisition of user preferences is usually considered a bottleneck for the practical use of these systems. Content-based recommendations may also result in very expected items and may not be able to direct the user towards items he is unaware of but may like. Nevertheless, CB systems can easily provide valid recommendations to new users, assuming that their profile is specified using a questionnaire or some other method for preferences elicitation, even if they never used the system before.

CB engines can provide recommendations for new items that were never rated before based on the item description and are therefore very useful in environments where new items are constantly added.

Hybrid approaches [10] of CF and CB can reduce the disadvantages of the methods.

2.3 Communities and Social Networks

The main idea of a Social Network (SN) is to use some relations that users sharing between them. It is the set of actors i.e. group of people, which are the nodes of the network, and ties that link the nodes by one or more relations. Social network indicates the ways in which actors are related. The tie between actors can be maintained according to either one or several relations. Moreover, the network gives not only to their actors, people that are directly connected, but also to actors of their actors, also called "friends of my friends".

The roots of link analysis predate the use of modern computers. The field of social network analysis has developed over many years as sociologists developed formal methods of studying groups of people and their relationships. The advent of computers allowed these techniques to become much more widespread and to be applied on a much larger scale in recommender systems.

Social networks can be divided into several groups in terms of different criteria:

- Dedicated - dating or business networks, networks of friends, graduates, fun clubs etc'
- Indirect - online communicators, address books, e-mails etc'
- Common activities. - co-authors of scientific papers, co-organizers of events etc'
- Local networks - people living in the neighborhood, families, employees networks etc'
- Hyperlink networks - links between home pages etc'

Recommender systems for social networks differ from the approaches which described before, since often they suggest rational human beings to other people and not just product or service. Generally the network is initiated by the users when one person initializes the relationship with another one, and the latter can respond either positively or negatively to the invitation such as in MSN messenger, ICQ etc'. After the relationship is been set up the network could be for use in some manner i.e. chatting in the MSN or ICQ examples. Such interaction is impossible with products and services. Furthermore, the relationship between people is bidirectional in opposite to the relationship between a person and product.

Most of the recommendation systems [6] which using social networks for recommendation are often need the *'trus't* parameter. That means if we are going to use the relationship between people for providing recommendation we need to know how well person A trust person B concerning the product and taste of the recommended item. So another parameter in the relations at the social network is the trust. However, there several situations where we assume the trust is high or we cannot measure i.e. ICQ or MSN messenger. There we are sure about the connection but we cannot know for sure the level of trust concerning recommendation for item of some type.

In [6] they use the trust ratings between the users in the CN as the basis for making calculations about similarity. They assume that there is correlation between trust and user similarity; that means if user A trust user B with high value, then the similarity of the preferences list of movies will also be high.

It has been shown [1] that in specific domain such as movies, users develop social connections with people who have similar preferences; however no field experiment has been provided. The results of [1] extended in [11]; their work proved that there is positive correlation between trust and user similarity in an empirical study of a real online community. There is a lot of work in the area of measuring the trust in communities and related the issue to recommender systems such as [7] and [6]. Measuring the trust is not easy task and often impossible and not accurate, thus, it is leading us to option of creating the relation on full trust and this can obtain only by explicitly order from the user i.e. in the MSN messenger and the ICQ where one cannot be related to other one unless he approved his invitation. As mentioned before we are using the suggested method for mobile application named MediaScout. In MediaScout each user can invite other users to become his friend by simply provide the mobile number; the receiver can approve or decline the invitation. Based on those direct relations the users in MediaScout can send each other recommendation about movies.

3 Personal Social Networks

3.1 Creating the Network

As described above, MediaScout is an application which provides personalized media content via mobile devices and home TV. One of the features which exist in the system is the ability to send an invitation to some other user in the system and to propose a friendship. The receiver of the invitation can accept or decline the invitation. If the second user accepts the invitation, then the two users become 'friends' of each other and can send recommendations to each other using one of the features in the application. This is a similar scenario to the one of MSN messenger and of ICQ, except for the usage goal. In the two existing applications one can chat with one's friend and in MediaScout, a pair of friends can send movie recommendations to each other. MediaScout uses a binary feedback mechanism. This means that for each media item that the user watches she can provide a feedback whether she liked it or not. This feedback helps the system to refine the profile of the user. For each user the system keeps several items that she rated positively, several items that were rated negatively and a list of friends as well.

Having this kind of information and relations about users can easily lead to the generation of a social network which involves most or all of the users of the application. Each user is a node in the network and the friendship relation is represented by edges which connect the users with a trust of 1 between each two connected nodes. For each user one can envision the first layer of friends, the users which are connected to the user by a direct link. In addition, there are friends of the friends and so on.

3.2 Constructing the Personal Network Model

Consider the network of friends, where the only absolute information we have is that the first layer of each user includes the list of the user's friends and for each user we have two lists of movies, one that she likes and one that she dislikes. We assume that there is some similarity between the preferences of a user and those of her friends and of the friends of her friends. We propose to construct recommendations that are based on the personal social network of each user. The personal social network of each user is a snapshot of the entire network which presents the relations of each user with her related friends, up to the level of 6 (friends of friends etc.). This network is constructed in the form of a social tree for each user (Ur) by using a Breadth-First Search (BFS) [9] algorithm. Figure 1 illustrates the transition between the social network into the personal network of a user. The number of the levels which we take into consideration while building the personal network will affect the recommendations later on. In the experiment at the next section, the network is constructed up to the level of six in order to cover a large range of relations among users. Denote the distance from user *Ui* to another user *Uj* by *d(Ui,Uj)* which is computed by traversing the social tree. The tree can be viewed as a set of users described by

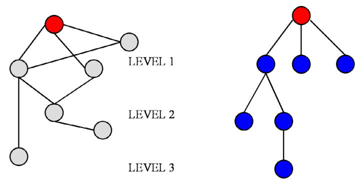

Fig. 1. The left graph describes the relations of the red user with other users in the system. The graph on the right hand side presents the resulting personal network after the run of the BFS algorithm. Note that each user appears only once in another user's personal network.

$$X(Ui, L) = \{U_j | \forall j, U_j \in \ social\ tree\ of\ user\ i\ up\ to\ the\ level\ of\ L\} \tag{1}$$

3.3 Constructing Recommendations

Once we have the personal social network of each user, and the items that each user likes and dislikes the construction of the recommendation list proceeds as follows. The list of recommendations for each user is based on the items that her 'friends' like and dislike. If some of the users in her personal network like some media item, this item is going to appear in the list of recommendations. Items that are negatively rated by friends are not going to be recommended. Recommendations are constructed by summing the impact of ratings of the personal social network, both positive and negative. We use the notations below to define the computation of the recommendation list:

- M - Media item.
- U - User in the system.
- *R(Uj,Mi)* - the rating of user j over media item i, The value 1 for positive rating, -1 for negative rating and 0 if user j did not rate media i.
- *d(Ui,Uj)* - the distance of user j from user i in the social tree of user i.
- X(Ui,L) = {Uj | ∀ j, Uj ∈ social tree of Ui computed up to the level of L}
- K - Attenuation coefficient of the social network.

The overall rating of a media item for user i, based on her personal social network computed as in (2).

$$Rank(Mi, Ui) = \sum_{\forall Uj \in X(Ui,L)} [K^{-d(Ui,Uj)} * R(Ui, Mi)] \tag{2}$$

The list of recommendations is simply a vector of sorted media items according to rank, where larger rank indicates higher order in the list.

The attenuation coefficient of the social network defines the impact of the distance between users, on the strength of recommendations. If $K = 1$ the impact is constant and the result is exactly equal to the popularity of the media items in the personal network of the user. If $K =$ n > 1 the rating of a user at level x is equivalent to n ratings of users at level of x+1.

4 Experiments

We conducted a field experiment with real users aimed to examine whether there is any correlation between user preferences regarding movies, to the preferences of her friends? Whether providing a recommendation based on a personal social network will be more effective than recommending movies by popularity, for example.

The experiment involved 50 users, all of them from the same class at Ben Gurion University, and they were all familiar with each other. The assumption is that there are groups inside this group and the entire group of the fifty users is homogenous.

Each user needed to write down two things:

1. The list of her friends - where 'friend' is defined to be a person whose recommendation she will consider, positive or negative, concerning movies.
2. To rate 108 movies on a scale of 1 to 5, where 1 means that the user did not like the movie at all and 5 means she liked the movie very much. Note that in cases where the user was not familiar with the movies, she rated them 0.

The movies were taken from the Internet Movie Data Base (IMDB) [12]. The list of 108 movies was constructed to have both popularity and diversity in the genre attribute. This information is available on IMDB. Having the rating of each user and her list of trusted friends enabled us to compare several methods on the same data. These methods included popularity, random recommendation, CF and our proposed approach.

4.1 Evaluation Metrics

To assess the performance of recommendations lists, we sort (descending) the movies for each user according to her preferences. We construct for each user her recommendation list based on the personal social network, using k=2. To assess the relevance of the resulting recommendations list, we checked the location of the movies that the user liked and used the R measure (3) which is taken from [5].

$$Ra = \sum_{t \in I} \frac{1}{2^{i/a}} \tag{3}$$

Here i is the set of locations of the movies the user liked in the recommendations list and a is the viewing half-life and in this experiment it was set to 2. This metric assumes that each successive item in the list is less likely to be viewed with an exponential decay. The grade was then divided by $Rmax$ - the maximal grade, when all the movies the user liked appear at the top of the list. The second metric we used is termed *recall*. Recall is the ratio of the number of relevant movies presented to the user to the total number of relevant movies for the user in the data. Since the users rated the items in the experiment from 1 to 5, we simply calculate the average rating for each user and considered the movies that she rated below her average as irrelevant for her and those rated above the average as relevant. Recall is quite important to measure in our case since there are a limited number of movies that we were able to present to the user as a recommendation list. We measure the recall by 10, 20, 30, and 60 based on the fact that in MediaScout users are viewing a limited number of items at once.

4.2 Comparative Results

We compare the results of the suggested method to popularity. Table 1 present the results of the popularity and the personal social network. Analysis of the results indicates that the personal social network is obtaining similar results to the popularity. In the R measure, which takes into consideration the order of the recommendation, yet all the movies which exists in the data as potential recommendations, the personal social network obtains better results. The recall measure is affected from the point of the measure. The fact that the recall at the level of sixty items (out of 108 items we got in this experiment) is close to 92% in popularity and in the social network as well is quiet impressive. It is indicating that both of the methods perform very well on this data. Generally the social network and popularity obtain similar results.

Table 1. Comparative results between popularity and the personal social network

	R measure	Recall 10	Recall 20	Recall 30	Recall 60
Popularity	111.97	0.26	0.4591	0.644	0.927
Social Network	119.734	0.2265	0.4357	0.617	0.9112

5 Conclusions

In this paper we presented a new community based recommendation method. The experimental study conducted with real subjects shows that this method can improve the recommendation performance in cases we need a long list of recommendations.

Although further experiments needed with different algorithms such as CF and different measures, we can see that the personal social network has the potential to provide very quality recommendations. One of the approaches that might be useful for improving the recommendations is to find the most appropriate attenuation factor, the K value as described in section 3, for each user. Given the list of rated item of a given user; is there any specific K which led to more quality recommendations for the user? We believe there is. It is most likely that the K will be different from user to user but we didn't manage answer the question of how to find it. Finding the personal K by some heuristic will probably improve the results of the social network in our case.

References

1. Abdul-Rahman, A. and Hailes, S. Supporting trust in virtual communities. In Proceedings of the 33rd Hawaii International Conference on System Sciences. Maui, HW, USA (2000).
2. B. Smyth and P. Cotter, Surfing the Digital Wave: Generating Personalized TV Listings Using Collaborative, Case-Based Recommendation, Lecture Notes in Computer Science, vol. 1650, pp. 561-567, (1999).
3. E. Rich. User modeling via stereotypes. Pages 329- 342, (1998).
4. Garton, L., Haythornthwaite, C., and Wellman, B. Studying Online Social Networks, in Journal of Computer-Mediated Communication, 3 (1), (1999).
5. J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In UAI'98, pages 43-52, (1998).
6. Jennifer Golbeck, James Hendler. 2006. FilmTrust: Movie recommendations using trust in web-based social networks. Proceedings of the IEEE Consumer Communications and Networking Conference, (2006).
7. Massa, P., B. Bhattacharjee. 2004. Using Trust in Recommender Systems: an Experimental Analysis. In Proceedings of iTrust International Conference (2004).
8. M. Montaner, B. Lpez, and J. L. De La Rosa. Taxonomy of recommender agents on the internet. Artificial Intelligence Review, 19:285-330, (2003).
9. M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. Phys. Rev. E, 69(2):026113, Feb, (2004).
10. R. Burke. Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12(4):331-370, (2002).
11. Ziegler, Cai-Nicolas, Georg Lausen. Analyzing Correlation between Trust and User Similarity in Online Communities" Proceedings of Second International Conference on Trust Management, (2004).
12. http://www.imdb.com/