ONLINE SOCIAL INFLUENCE AND PERSONALIZED RECOMMENDATION: AN EMPIRICAL STUDY

(Full Paper Submission)

Xiaoyun He
Auburn University at Montgomery
xhe@aum.edu

Amir Karami
University of South Carolina
karami@mailbox.sc.edu

ABSTRACT

Recent studies suggest that online social relations influence users' both product choices and ratings, which provides empirical support for incorporating online social influence into the recommender models. To incorporate such influence, we consider the ratings from a user's social circle as a source of implicit feedback from the user. We then integrate such feedback into the representative recommendation model of collaborative filtering to improve personalized recommendations. The evaluation results show that our proposed approach performs better than the pure ratings-based methods in terms of accuracy and coverage.

KEYWORDS: social influence, online recommendation, personalized list, implicit feedback

INTRODUCTION

In the current digital age, online retailing and content providers are empowered to offer a huge selection of products and services to meet a variety of consumers' needs and tastes. As a result, consumers are inundated with a wealth of information and choices. For example, if we type in "digital camera" in the search box at Amazon's web site, the search result shows over 1 million related items. To mitigate such information overload, recommender systems have been widely used to recommend products that are most likely to be of interest to users (Jannach et al. 2010). According to a Forrester Research report, one third of online customers that notice these recommendations wind up purchasing the recommended product (Mulpuru 2006). By matching consumers with most appropriate products, companies could potentially enhance customer satisfaction and retention, thus boosting sales and the bottom line. Furthermore, with the trend of consumers spending more time and money on online shopping and increasingly intensive competition in online retailing, improving the quality of product recommendations (e.g., in terms of predicting accuracy and coverage) becomes a critical factor for companies to gain sustainable advantage.

In a nutshell, a basic recommendation method seeks to predict the 'preference' or 'rating' that a user would give to an item, such as music, books, movies, etc. that has not been seen by the user. A rating indicates how a particular user liked a particular item, e.g., Jane Betty gave the book ``Thinking, Fast and Slow'' the rating of 4 (out of 5). These ``particular'' recommendations are often referred as personalized recommendations (Kim et al. 2003). The task of personalized recommendation requires the ability to predict which items will be considered interesting by the user. The personalized recommendations have been viewed as an important source to assist and augment the natural social recommendation process - in our everyday life, we rely on recommendations from our social circles by word of mouth (Shardanand & Maes 1995). A central concept for the social recommendation process is social influence, which refers to the change in one's attitudes, behavior, or opinions to external pressure that is real or imagined (Cialdini & Goldstein 2004).

The literature on social influence and consumer decision making suggests that social influence is an important factor of affecting people's decision making and consumer decisions are best understood in the social contexts in which these decisions are made (e.g., Deutsch & Gerard 1955; Simpson et al. 2012). For instance, we tend to be influenced by our friends in terms of products purchase. Social influence not only occurs among the directly connected acquaintances, also propagates via the chain of online connections in a social network (Richardson & Domingos 2002). In addition, recent studies suggest that online social relations influence users' both product choices and ratings (e.g., He et al. 2017). Within an informal community of users and social context, it is therefore natural to incorporate the factor of social influence derived from online social networks into recommendation model to make personalized recommendations.

In this study, we utilize a data set that has been collected from an online Web 2.0 site. The website integrates both online social networking and online product rating functions.
In particular, we consider the ratings from a user's social circle as a source of implicit feedback from the user, thus incorporating the factor of social influence into the recommendation model. Specifically, we integrate the implicit feedback into the representative recommendation model of collaborative filtering to improve personalized recommendations. The evaluation results show that our proposed approach performs better than the pure ratings-based methods in terms of both accuracy and coverage.

LITERATURE REVIEW AND RELATED WORK

Collaborative Filtering based Recommendation Techniques

Over the past two decades, collaborative filtering techniques have received much attention in the area of recommender systems (Adomavicius & Tuzhilin 2005). These techniques rely on the past user behavior (e.g., their previous purchase transactions or product ratings) to predict what a user may like. The collaborative filtering methods can be generally classified into two categories: neighborhood-based methods and model-based methods.

The neighborhood-based approaches are centered on finding the relationship between users or, alternatively, between items. The initial form of neighborhood-based approaches is user-oriented (e.g., Herlocker et al. 1999). In the user-oriented case, the system predicts a user's rating based on the known ratings of similar users, i.e., the like-minded users. Later, an analogous item-oriented approach was proposed (e.g., Deshpande & Karypis 2004). In this case, a rating is computed based on past ratings made by the same user on similar items.

The model-based approaches typically involve two steps. First, training data consisting of past user-item ratings is used to train a predefined model. Then, the model is used to predict the unknown ratings. In this category, various models have been proposed, including clustering model (Merialdo 1999), latent semantic model (Hofmann 2004), ranking-based model (Liu & Yang 2008), Bayesian hierarchical model (Zhang & Koren 2007), to name a few. In particular, in order to address the efficiency problem in large datasets, several matrix factorization methods have been proposed for collaborative filtering (Mnih & Salakhutdinov 2008; Srebro & Jaakkola 2003; Koren et al. 2009; Agarwal & Chen 2009). These methods focus on fitting the user-item rating matrix with low-rank approximations and use the learned latent user / item factors to make predictions of the ratings. Such low-rank matrix factorizations are based on the assumption that only a small number of factors characterize user preferences, and that a user's preference vector is thus determined by how each of those factors applies to that user. Indeed, low-rank matrix approximations based on minimizing the sum-squared errors can be easily solved using Singular Value Decomposition (SVD). Srebro and Jaakkola have proposed a simple yet efficient Expectation Maximization (EM) algorithm for solving weighted low-rank approximation (Srebro & Jaakkola 2003). In addition, these factorization methods are generally effective at estimating the overall structure that relates simultaneously to most or all items (Koren 2008).

Recommendation Based on Social Relations

Exploiting social networks to enhance recommendation performance has been studied by the researchers (e.g., Kautz et al. 1997; Guy et al. 2009; Liu & Lee 2010; Ma et al. 2011). Kautz et al. (1997) propose to build an interactive ReferralWeb system that uses social network to make recommendation / search more focused and effective. They argue that the referral chain arising from a social network enables the utilization of experts' expertise and the evaluation of trustworthiness of the expert. The explorative study in Guy et al. (2009) demonstrates a clear superiority of users' familiar network (e.g., being connected in a social network) over similarity network (e.g., having similar tastes as expressed in item ratings) as a basis for recommendation. Liu and Lee point out that their evaluation results on recommendations indicate that more accurate prediction algorithm can be developed by incorporating social network information into collaborative filtering (Liu & Lee 2010). Ma et al. (2011) propose a matrix factorization framework to include the social information as regularization terms into recommendation systems. However, the similarity between a pair of user in the social network is still based on the past ratings of the users. That is, the social information is based on users' similarity network. Rather, our paper exploits user's familiar network to get implicit feedback that is then integrated into the recommendation method.

RESEARCH CONTEXT AND METHODOLOGY

Research Context and Data

The data was obtained from a Web 2.0 site, which wishes to remain anonymous. The website offers the integrated social networking and online rating services to the users. It provides a unique platform for users to establish online social relations and to participate in rating and reviewing the products that they have consumed. The data of users’ relationships establishment and rating participations generated from this platform enables us to empirically leverage the effects of online social relations so as to improve personalized product recommendation. With this particular site and the focus of our study, a crawling software program was used to collect the related data from the site.

At the web site, a user may use a valid email address to register and create a screen name as the user identifier that is shown to the other registered users. The screen name does not have to be a user's real name, thus it offers certain level of anonymity. Over 95% of the registered users chose to use a screen name that is different from their real name. Registered users can rate and write reviews for specific product items. These ratings are available to be viewed by all other registered users. The products fall into one of three product categories - books, music, and movies. The rating scale is from 1 to 5 stars with 1 being the lowest and 5 being the highest level. In addition, the web site also supports online social network service that a registered user can choose to follow another registered user by simply clicking a "Follow" button. Unlike other well-known social networking sites such as Facebook or LinkedIn, the establishment of "following" relations does not need to be consented / invited by the followed user. Based on a survey of the registered users, about 87% of the respondents stated that they follow other users whom they do not know in person. In other words, the online "following" relations in this context are largely different from those in the physical offline world.

The final sample data span the period from January 2008 to August 2011. The data that we used contain the following data sets: (1) Books: the rating data for books; (2) Music: the rating data for music; (3) Movies: the rating data for movies; (4) OnlineSN: the online social network that describes that "following" relationships among users including user identifiers and date when the relation was established. For each of the categories of items, the rating data include the following attributes: itemID (the specific item that was rated); userID (the user who rated the item); rating (an integer number between 1 and 5 assigned by the user to rate the item); date (when the rating was provided).

Personalized Recommendations Incorporating Social Influence

In this section, we first provide background knowledge related to personalized recommendation methods. Then, we discuss to consider the influence from user's social circle as a type of implicit feedback on user's preference. We adapt an integrative recommendation model proposed in Koren (2008) to incorporate such type of implicit feedback.

Preliminaries and Baseline Recommendation model

For clarity and simplicity, special indexing letters are reserved for distinguishing users from items: for users, we use u, v; and for product items, use i, j. A rating r_{ui} denotes the preference by user u for item i, where higher ratings indicate stronger preference. We use $\hat{r}_{ui}$ to denote the predicted value of r_{ui}. The notation $\check{W}$ is used to denote the set of (u, i) pairs for which r_{ui} is provided, that is, $\check{W} = \{(u, i) \mid r_{ui} \text{ is known}\}$. In general, the rating data is sparse with significant large number of items unrated. To overcome the problem of overfitting the sparse ratings, models are typically regularized so that predictions are shifted towards baseline values. Regularization is controlled by constants which are denoted as: λ_1, λ_2, ……. Exact values of these constants are determined by cross validation.

A simple and straightforward approach for recommendation is to estimate $\hat{r}_{ui}$ as the mean rating of item i across all users. On the other hand, by applying personalization, we could obtain more accurate estimations. Yet, one common characteristic existing in the rating data is the disparity in terms of rating tendency. Specifically, some users tend to give higher ratings than others; and some items tend to receive higher ratings than others. In order to account for such disparities, baseline prediction model is used Koren (2008). A baseline prediction, denoted by b_{ui}, for an unknown rating $\hat{r}_{ui}$ is estimated as follows:

$$b_{ui} = \mu + d_u + d_i \qquad (1)$$

where μ denotes the overall average rating, and d_u and d_i are the observed deviation of user u and item i from the average rating respectively. For example, suppose we would like to estimate the baseline prediction for the rating of movie *Star Wars* by user Jane. Let's say that the average rating over all movies, μ, is 3.6 stars. The movie *Star Wars* tends to be rated 0.4 stars above the average while Jane tends to rate 0.5 lower than the average. Then, the baseline prediction for movie *Star Wars* 's rating by Jane would be 3.5 (= 3.6 - 0.4 + 0.5) stars. In order to estimate d_u and d_i, the following least squares problem needs to be solved:

$$\min_{d^*} \Sigma_{(u,i) \in W} (r_{ui} - \mu - d_u - d_i)^2 + \lambda_1 (\Sigma_u d_u^2 + \Sigma_i d_i^2)$$

In the above formula, the first term tries to find d_u and d_i that fit the given ratings data and the second one is a regularizing term used to overcome overfitting.

Implicit Feedback

Traditional collaborating filtering approaches primarily rely on explicit feedback, such as product ratings provided by users. Recent studies have recognized the importance of considering also implicit feedback, which indirectly reflects users' interest by observing user behaviors (Oard & Kim 1998; Koren 2008). Koren (2008) has shown that prediction accuracy is improved by integrating implicit feedback into the recommendation model. Sources of implicit feedback may include rental / purchase history, search patterns, or browsing history. Yet, implicit feedback is typically not available. To overcome the problem of unavailability, Koren used a less obvious kind of implicit data - reducing the ratings matrix into a binary matrix, where ``1'' stands for ``rated'' and ``0'' for ``not rated''. However, such binary data is very limited in signaling users' implicit preferences, because we do not have much independent implicit feedback for the typical ratings dataset.

Based on the factor of social influence and the above empirical findings, we subsequently include the ratings from a user's social circle as a source of his / her implicit feedback. That is, given a user u, the product ratings provided by the user's social circle, N(u), indirectly reflect an implicit preference of user u. In the setting of online rating and social network, this type of implicit data are handily available. We explicitly incorporate such kind of implicit feedback into the recommendation model, as shown next.

Integrative Recommendation Model

Since we would like to incorporate both explicit product ratings and implicit feedback from users into our recommendation model, we adapt one such integrative model proposed by Koren (2008). The model is latent factor based. Below, we briefly review the approach and adapt it to our context by incorporating the implicit feedback that reflects social influence.

One popular approach to latent factor models is induced by Singular Value Decomposition (SVD) on the user-item ratings matrix. Specifically, the model associates each user u with a user-factors vector $p_u \in \mathbb{R}^f$ (f is the number of factors), and each item i with an item-factors vector $q_i \in \mathbb{R}^f$. The prediction of an unknown rating is attained through an inner product:

$$\hat{r}_{ui} = b_{ui} + p_u^T q_i \qquad (2)$$

Replacing b_{ui} by the baseline estimation in equation (1), we have the following prediction:

$$\hat{r}_{ui} = \mu + d_u + d_i + p_u^T q_i$$

Again, in order to overcome the overfitting problem in the sparse rating data, regularization is applied to the above prediction. Then, we get the following adequate regularized model:

$$\min_{\{p^*,q^*,d^*\}} \Sigma_{(u,i) \in W} (r_{ui} - \mu - d_u - d_i - p_u^T q_i)^2 + \lambda_2 (d_u^2 + d_i^2 + ||p_u||^2 + ||q_i||^2) \quad (3)$$

where $||.||^2$ denotes the usual Frobenius/L2-norm; a simple gradient descent technique was used to solve the above squared error function (see also Salakhutdinov et al. (2007) for more detailed discussion on it).

Koren (2008) has extended the above model to avoid explicitly parameterizing each user and incorporate implicit feedback. In order to specifically account for the implicit feedback derived from social network mentioned earlier, we adapt Koren's extended model to be the following:

$$\hat{r}_{ui} = b_{ui} + q_i^T (|R(u)|^{-½} \Sigma_{j \in R(u)} (r_{uj} - b_{uj}) x_j + |S(u)|^{-½} \Sigma_{j \in S(u)} y_j) \quad (4)$$

where R(u) is the set of items explicitly rated by user u, S(u) is the set of items for which u provided an implicit feedback, each item i is associated with three factor vectors q_i, x_i, $y_i \in \mathbb{R}^f$. Since, in this model, users are represented through the items they rate, the user factor p_i in the model (2) is replaced by the term $|R(u)|^{-½} \Sigma_{j \in R(u)} (r_{uj} - b_{uj}) x_j + |S(u)|^{-½} \Sigma_{j \in S(u)} y_j$.
In our case, since the product ratings provided by user u's social circle, N(u), are considered to reflect an implicit preference of user u, S(u) contains the set of items that the users in N(u) have provided the ratings.

Similar to solve the model (3), the parameters in equation (4) are learnt by minimizing the associated squared error function as shown below:

$$\min_{\{q^*,x^*, y^*,d^*\}} \Sigma_{(u,i) \in W} (r_{ui} - \mu - d_u - d_i - q_i^T (|R(u)|^{-½} \Sigma_{j \in R(u)} (r_{uj} - b_{uj}) x_j + |S(u)|^{-½} \Sigma_{j \in S(u)} y_j))^2 + \lambda_3 (d_u^2 + d_i^2 + ||q_i||^2 + \Sigma_{j \in R(u)}||x_j||^2 + \Sigma_{j \in S(u)} ||y_j||^2)$$

Then, a simple gradient descent technique can be used to solve the above function. For our ratings data, we used 26 iterations with step size of 0.001 and λ_3 of 0.05.

We call the above adapted model SocialSVD. In the model, the implicit preference becomes more prevalent as |S(u)| increases. Specifically, when the number of ratings provided by user u's social circle increases, more implicit feedback are received from u. In other words, the higher the value of |S(u)| is, the stronger signals of implicit feedback user u sends out. On the other hand, the explicit preference becomes more prominent when |R(u)| is growing, that is, the user u is providing more ratings. As noted in Koren (2008), a single explicit input would typically be more valuable to a recommender system than a single implicit one. The right conversion ratio between these two types of inputs, which indicates how many implicit inputs are as important as one single explicit input, is automatically learnt from the data through setting the relative values of parameters x_j and y_j.

EXPERIMENTAL EVALUATION AND DISCUSSION

In this section, we describe the metrics used for performance evaluation, present the evaluation results, and discuss the main findings from our experimental evaluations. Evaluation metrics are essential in order to measure the quality and performance of recommendation approaches. We consider both accuracy measurement and non-accuracy evaluation metrics such as coverage.

Prediction accuracy metrics in recommender systems measure how close estimated ratings come to actual user ratings. In this category, the Root Mean Square Error (RMSE) is widely used to measure the statistical accuracy of predictions. The RMSE is defined as:

$$RMSE = sqrt(\Sigma_{(u,i) \in TestSet} (r_{ui} - \hat{r}_{ui})/|TestSet|)$$

where |TestSet| denotes the number of ratings in the test set.

There has been considerable understanding that high recommendation accuracy alone does not necessarily provide users of recommender systems satisfying experience (Herlocker et al. 2004; Ziegler et al. 2005). Among the proposed non-accuracy evaluation metrics, coverage has been the most frequently used. The coverage of a recommender system is a measure of the domain of the items over which the system can form predictions or make recommendations. System with lower coverage may be less valuable to users. Coverage typically should be measured in combination with accuracy, so recommender system are not tempted to boost one over the other. In particular, the most common measure for coverage is defined as the percentage of <user, item> pairs in the test set for which a recommendation prediction can be made.

We perform 5-fold cross validation in our experimental evaluations. In each fold, we use 80% of randomly selected data as training set and the remaining 20% as the test set. For the Books, Music, and Movies data, the observations for relative rankings are 1,137, 1,412, and 972 respectively.

In order to demonstrate the effectiveness of our proposed SocialSVD, we compare it with other benchmark methods including SVD (Sarwar et al. 2000) and AsymmetricSVD (Koren 2008). In addition, we also examine the results when SocialSVD does not consider implicit feedback derived from social influence, which is referred as SocialSVD w/o (without) Implicit.

Table 1. Experimental Results

Category	Method	20 factors		40 factors	
		RMSE	Coverage %	RMSE	Coverage %
Books	SVD	0.8034	81.72	0.8031	81.84
	SymmetricSVD	0.7941	82.06	0.7903	82.70
	SocialSVD w/o Implicit	0.7983	80.65	0.7880	81.29
	SocialSVD	0.7862	84.31	0.7859	84.54
Music	SVD	0.8531	79.32	0.8502	80.10
	SymmetricSVD	0.8428	81.47	0.8411	81.82
	SocialSVD w/o Implicit	0.8510	80.91	0.8499	81.04
	SocialSVD	0.8229	82.52	0.8213	82.88
Movies	SVD	0.8792	78.86	0.8719	78.97
	SymmetricSVD	0.8602	80.26	0.8543	81.34
	SocialSVD w/o Implicit	0.8744	79.03	0.8702	80.16
	SocialSVD	0.8623	81.94	0.8608	82.15

As shown in Table 1, we can see that our approach SocialSVD performed better in almost all the data sets in terms of accuracy and coverage. This provides encouraging support of our approach that leverages implicit feedback derived from online social influence. We note that one exception is the movies dataset, in which AsymmetricSVD performed better in terms of accuracy. This might be due to the data characteristics in movies data, for instance, for some users who have not provided much explicit product ratings, their friends did not provide sufficient product ratings either. Nevertheless, the relative improvements of AymmetricSVD over SocialSVD are in very small margin.

As we use the ratings from a user's online social circle as a source of implicit feedback in the recommendation model, the above results suggest that leveraging the ratings data from online social circles can help to improve the personalized recommendation. In the context of online social networks, the ratings from a user's online social circle can be considered as a proxy for online social influence. Thus we believe that our findings have practical implications for online retailers to take advantage of online social relations.

CONCLUSIONS

In this study, we propose to incorporate the factor of social influence derived from online social networks into recommendation model to improve personalized recommendations. Specifically, we consider the ratings from a user's online social circle as a source of implicit feedback from the user. We then integrate the implicit feedback into the representative recommendation model of collaborative filtering to improve personalized recommendations. The evaluation results show that our proposed approach performs better than the pure ratings-based methods in terms of both accuracy and coverage. In the future, we plan to examine the diversification issue of personalized recommendation list, which may realistically reflect a user's spectrum of interests and thus enhance users' satisfaction.

REFERENCES

References available upon request.