Are You a Social Conformer?

Priyanka Garg, Irwin King, and Michael R. Lyu

The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
{priyanka,king,lyu}@cse.cuhk.edu.hk

Abstract. Social recommendations have been found to increase the product adoption probability. However, very few studies have considered the impact of social opinions on the users' evaluation of the product. In social networks, many times users' opinions are not completely independent from their friends and users tend to change their rating such that they are more similar to the social opinions. Understanding this behavior is important for developing accurate recommendation systems, precise information flow models and to launch effective viral marketing campaigns. In order to understand this phenomenon, we propose a novel formulation for the users ratings where every expressed rating is considered as a function of the social opinion along with the user preference and item characteristics. The proposed method helps in improving the prediction accuracy of users' rating by more than 2% in presence of social influence. Additionally, the learned model parameters reveal the degree of conformity of users. Detailed analysis of user social conformity show that more than 76% of users tend to conform to their friends to some extent. On an average, user ratings become more positive in presence of the social influence.

Keywords: Pattern Mining, Social Conformers, Recommender System.

1 Introduction

Social networks play a fundamental role in spreading information, ideas and technologies among their members. Often the decision to adopt a product is influenced by one's social connections. For example, positive friends reviews about a book encourages us to read it. Numerous studies have indicated that social recommendations result in an increase in the sales volume [2]. As a result, a large amount of research efforts have been devoted to understand the intricacies involved [5] and coming up with interesting applications like viral marketing [5], personalized recommender systems [6], etc.

However, most of the existing models have largely ignored the effect of social opinions on the *posterior* users evaluation of products i.e. the opinion the user form after experiencing the product. They either assume that the expressed opinion is same as the influencing opinion [5] or they are assumed to depend strictly on the product quality [1]. However many times, user's evaluation of the product, is not completely independent of her social circle and she tends to conform with social opinions. For example, a user reads a book and does not

T. Huang et al. (Eds.): ICONIP 2012, Part III, LNCS 7665, pp. 694–701, 2012.

like it much. However lots of friends praise it and call it a really insightful book or "5/5", then this might change the user's opinion slightly and user might rate the book as "3". Had she not interacted with her friend, she might have given a rating of "2". This behavior usually arise because of the presence of social pressure and the innate difficulty involved in providing an absolute numerical rating to a product [8]. In such cases, social opinions can act as a reference rating and calibrates the user ratings such that they are not very different from the prevalent social opinion. We call this behavior as **social conformity** and the users who changes their rating as *social conformers*. Recently, this effect has been shown to exist on Goodreads and Douban [4].

Quantifying this behavior is important not just from the point of curiosity, but it is also crucial in improving the accuracy of personalized recommender systems and in developing better information flow models. The recommendation systems can boost the quality of recommendation by removing the social conformity bias, thus making the recommendation better tailored to users' preference. While the information flow models can more accurately predict the further information cascade by accurately predicting the users' opinions. However, it is a very **difficult task** to quantify the social conformity as for a given user and product we never get to know the two ratings, one under the social influence and one without it. All that is known is a single opinion expressed by the user. Thus, the key challenge is to identify what component of any rating corresponds to the user's preference and what component corresponds to the social conformity.

In this paper, we account for social conformity and propose a novel formulation for the user's ratings. Contrary to homophily based recommender systems [6], which try to learn user preference based on her friends' preference, we focuse on the change of ratings at item level *caused* by the social influence. The proposed formulation represents every user rating as a function of social conformity and social opinion along with user's preference and item's characteristics. The social conformity down-weighs the user's preference such that as the number of influential friends increases, the user's rating become more similar to the social opinion. Further, the model parameters provide an intuitive interpretation of the social conformity behavior which reflect the degree a user conforms to her friend. It is important to note that different from the homophily based recommendation systems, we focus on the change of ratings at item level. Using this model, we explore the presence of social conformity on a real large scale dataset, Goodreads[1].

The key contributions of this paper are following.

1. We propose a novel formulation for user ratings that explicitly considers the social conformity. The proposed model improves the prediction accuracy of users' ratings by more than 2% in presence of social influence.
2. The learned social conformity parameters are also verified by qualitatively comparing the discovered most influential users with the authoritative and most socially active users.
3. Based on the learned users' degree of conformity, we find various interesting patterns on Goodreads that underline the impact of social conformity.

[1] `http://www.goodreads.com/`

2 Conformity Rating Model (CRM)

Notations. Let $G = (V, E)$ be a directed graph where every node $u \in V$ corresponds to a user in the social network and edge $(u, f) \in E$ exists if node f is a friend of node u. The user ratings for the set of items I, are stored in user-item matrix R, such that every element $r_{u,i}$ represents the rating for item i given by user u. Let the set of *active neighbors* who have posted their ratings for item i before user u be $A(u, i)$.

Problem Definition. The task is to predict the rating $r_{u,i}$ for item i given by user u, given the user-item matrix R and the set of active neighbors $A(u, i)$.

Social opinions calibrate user's inner rating $r^0_{u,i}$ such that they are not very different from them. To account for such social behavior, we propose the following social conformity based rating model CRM as

$$\hat{r}_{u,i} = r^0_{u,i} + conf \cdot (social\_opinion - r^0_{u,i}) \tag{1}$$

$$= \Big(1 - conf\Big) r^0_{u,i} + conf \cdot social\_opinion, \tag{2}$$

where $conf$ represents the degree by which a user conforms to the social opinion and $social\_opinion$ is the social opinion about the item i before the user u rates it. The rewritten form in Eq. (2) can also be seen as down-weighing the user's personal preference and giving higher weight to the friends' opinions. That is, if the user u has extremely high degree of social conformity then user u will change her rating such that it becomes same as the social opinion. Now we define each of the quantity $conf$, $social\_opinion$ and $r^0_{u,i}$ one by one.

- **User's Conformity** $conf$**.** We expect the degree of conformity $conf$ to take large values as the number of friends who have already rated the item increases. This phenomenon is known as the **bandwagon effect** in social sciences [3]. According to the bandwagon effect, as the number of individuals who believe in something increases, others tend to disregard their own opinions and also "hop on the bandwagon". That is, the social conformity is directly proportional to the number of friends with similar opinions. Thus, $conf = |A_{u,i}|$, because only active friends can affect the user's rating for the item. However, one can expect that users do not conform to all their friends equally. The friends who are regarded highly in the user's eyes, tend to affect their rating more. Hence, we introduce a parameter $\eta_{f,u}$ corresponding to every user and her friend pair. This parameter defines the degree by which user u conforms to the rating of its friend f. As the number of friends with high $\eta_{f,u}$ increases, the $conf$ can be expected to increase. Thus, we write

$$conf = \sum_{f \in A_{u,i}} \eta_{f,u}. \tag{3}$$

Since $conf$ can take maximum value of 1, we constraint $\eta_{f,u}$ such that $\sum_f \eta_{f,u} \leq 1$. Such linear forms of social influence have also been used in Linear threshold model [5] where the adoption probability of a product depends linearly on the active friends' influence.

- **Social Opinion** *social_opinion***.** We write the *social_opinion* as the sum of friends' opinions weighted according to $\eta_{f,u}$. This is because the opinions of friends with high $\eta_{f,u}$ affect the user's rating by the most amount. Thus, we have
$$social\_opinion = \frac{\sum_{f \in A_{u,i}} \eta_{f,u} \cdot r_{f,i}}{\sum_{f \in A_{u,i}} \eta_{f,u}}. \tag{4}$$
- **User's Inner Rating** $r^0_{u,i}$**.** To represent the user's inner rating $r^0_{u,i}$, we user one of the state of art recommendation models, Probability Matrix Factorization (PMF) method [7]. PMF model uses a small number of factors to represent the preference of users and item characteristics. The preference of users $q_u \in R^K$ and item characteristics $p_i \in R^K$ are represented by low dimensional vectors in latent space of dimensionality K. Then every rating is written as
$$\hat{r}^0_{u,i} = \mu + b_i + b_u + q_u^T \cdot p_i, \tag{5}$$
where μ is average user-item rating, b_i is item bias and b_u is user bias.

Thus, we finally have

$$\hat{r}_{u,i} = \Big(1 - \sum_{f \in A_{u,i}} \eta_{f,u}\Big)(\mu + b_i + b_u + q_u^T \cdot p_i) + \sum_{f \in A_{u,i}} \eta_{f,u} \cdot r_{f,i}.$$

Parameter Estimation. To estimate the model parameters b_i, b_u, q_u, p_i, $\eta_{f,u}$, we construct the objective function such that it minimizes the square of difference between observed user rating $r_{u,i}$ and estimated rating $\hat{r}_{u,i}$. Additionally, all parameters are regularized to avoid over fitting on the train dataset. Thus, our objective function is

$$\min \sum_{u,i} (r_{u,i} - \hat{r}_{u,i})^2 + \lambda_1 \Big(\sum_u b_u^2 + \sum_i b_i^2 + \sum_u ||q_u||^2 + \sum_i ||p_i||^2 \Big) + \lambda_2 \sum_{u,f} \eta_{f,u}^2$$
$$\text{s.t. } \eta_{f,u} \geq 0 \ \forall u, f; \ \sum_f \eta_{f,u} \leq 1 \ \forall u,$$

where λ_1 and λ_2 are the hyper-parameters which control the amount of regularization. The objective function is minimized by using the alternating minimization. In every first alternating step, we minimize the function with respect to the PMF model parameters b_i, b_u, q_u and p_i, using the steepest gradient decent method. Then in the second alternating step, we minimize the function with respect to $\eta_{f,u}$. Given the estimate $\hat{r}^0_{u,i}$ from first step, the objective function in the latter step can be written as the sum of small subproblem, each corresponding

Table 1. Goodreads data statistics

Users	Edges	Items	Ratings	Number of Authors
55,654	1,757,568	120,703	9,462,016	5,078

to one user. Since the set of parameters $\eta_{f,u}$ of every subproblem are different from the others, the objective function can be minimized by minimizing each of the sub problems separately. Thus, each of the sub problem can be minimized efficiently in parallel, using the gradient descent method.

3 Empirical Evaluation

We evaluate the effectiveness of CRM, both in terms of its ability to predict user ratings and its ability to identify the social influencers.

3.1 Goodreads Dataset

Goodreads is an online social books cataloging website, which permits users to rate books on $0-5$ scale (with 5 being the best) and share their reviews with friends. We use the dataset crawled by authors in [4]. The items and users are filtered such that every item has at least 10 ratings and every user has rated at least 5 books rated on 5 different dates and have at least 10 friends. This is to make sure the selected users are active users. In addition, we also crawl the profile pages of all the selected users. Users who have also authored books are marked as the authors. The statistics of the data is summarized in Table 1.

3.2 Prediction Accuracy

We evalute the ability of CRM to predict the users' ratings and compare its accuracy with the PMF method. The model parameters of both the PMF and the CRM, are trained on a train set and their performance is calulated on a test set. The train and test sets are constructed by splitting the user-item ratings in 4:1 ratio.

Performance Measure. The Root Mean Square Error (RMSE) metric is used for measuring the prediction accuracy. It is defined as $\sqrt{\frac{\sum_{u,i}(r_{u,i}-\hat{r}_{u,i})^2}{N}}$, where N is the number of ratings in the test set.

Observations. The RMSE values obtained using the CRM and PMF when $\lambda_1 = 1, \lambda_2 = 0.1$ are presented in Table 2. We can observe following.

- RMSE improves by more than 0.3% when social conformity is taken into consideration. Further, if we calculate RMSE value only for ratings who are potentially affected by the social influence ($conf > 0.1$), the RMSE improves by **more than** 2%.

Table 2. RMSE when $\lambda_1 = 1, \lambda_2 = 0.1$

		Number of test ratings	**PMF**	**CRM**
$K = 10$	All ratings	1,789,663	0.8556	**0.8520**
	Ratings with $conf > 0.1$	208,852	0.8476	**0.8254**
$K = 5$	All ratings	1,789,663	0.8472	**0.8441**
	Ratings with $conf > 0.1$	196,855	0.8471	**0.8280**

– **Sensitivity to K.** The RMSE value increases for both CRM and PMF model when K increase from 5 to 10. However, the drop in RMSE value is larger for PMF model than for CRM. This might be because a large value of latent space dimensionality K, can lead to over fitting on the training set. However, smaller impact on CRM underlines its robust performance.
– **Sensitivity to λ_2.** Effect of λ_2 (hyper-parameter to control the regularization) on the RMSE values is as per the expectations and is shown in Figure 1. The performance gets hurt if λ_2 is too large (≥ 1) or when it is too small (≤ 0.01). Higher value of λ_2 forces the selection of small social conformity factors $\eta_{f,u}$ and thereby under fits the model. While very small value leads to over-fitting on the training set. The best RMSE value is achieved when $\lambda_2 = 0.1$, though performance is reasonably robust around this value.

3.3 Influencers Quality

We evaluate the quality of the learned $\eta_{f,u}$ parameters by analyzing the properties of most influential users. In our setting, the users who have maximum effect on their friends' ratings are the social influencers. Formally, we define the social influence of a user u as $\sum_f \eta_{u,f}$ (Note that it is defined by reversing the conformity $\eta_{f,u}$ direction). We expect that the top influencers should have higher authority and higher number of social connections. Hence, we rank the users based on their social influence and study their degree and authority.

– **Average degree of top influential users.** The average degree (number of friends) of top x most influential users as function of x is plotted in Figure 2. It can be noticed that top 200 influential users have the highest average degree and as we consider more and more top users as the influential ones, their average degree starts to fall.
– **Authority of top influential users.** We validate the authority of the influential users by checking if they have also authored the books. This is a reasonable criterion as the book authors have higher perceived authority among their friends. The plot of percentage of authors among the top x influential users is shown in Figure 3. It can be noted that percentage of authors is very high among the top influencers. More than 45% authors appear in the top 200 influencers and there are only 12% authors among the top 5000 users, while the entire dataset has approximately 9% authors. Thus, as we keep widening the value of x, the ratio of authors to non-authors approaches to the ratio of entire dataset.

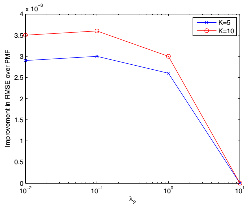


Fig. 1. Improvement over PMF as $\lambda 2$ varies

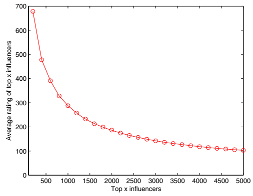


Fig. 2. Average degree of top influencer

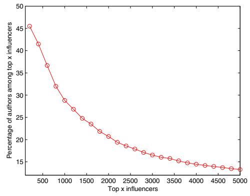


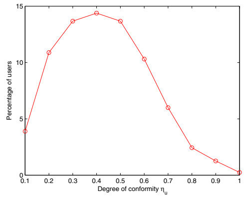
Fig. 3. Authors among top influencer

Both the observations show that CRM is able to **identify the social influencers** accurately.

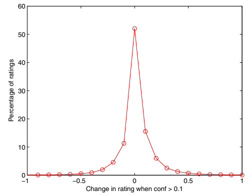
4 Social Conformity Analysis

Having verified CRM model, in this section we explore the nature of social conformity. We seek to answer following the questions.

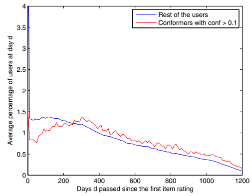
- How many users conform to their friends?
  The distribution of user conformity $\eta_u = \sum_f \eta_{f,u}$ is presented in Figure 4(a). It can be noticed that more than 76% users have $\eta_u > 0$. Among these users, most of them belong to the 0.2 to 0.6 interval. That is, most of the users in the network display some sort of conformity to their friends. In gerernal, we find that most of the users conform to only one friend and less than 9% users conform to more than 14 friends.
- By how much amount the user ratings change because of social opinions?
  We plot the distribution of change in ratings caused by the social opinions for ratings with $conf > 0.1$. Results are presented in Figure 4(b). It can be noted that, most of the ratings change is between -0.5 and 0.5. Additionally, it is interesting that more ratings change by positive factor than by the negative factor. 15% rating changes by +0.1 amount while only 12% ratings changes to -0.1 amounts.
- When does the conformity prevail the most?
  For each item, we find the percentage of social ratings with $conf > 0.1$ that appeared in between day d and day $d + 10$ since first rating is posted. Then, we calculate their average over all the items and plot them against day d. Similarly we plot the ratings with $conf \leq 0.1$. Results are presented in Figure 4(c). It can be seen that two kinds of ratings follow different pattern. The ratings with $conf \leq 0.1$ have maximum presence during the start of the information cascade and their percentage decays slowly as the time passes by. While the ratings with $conf > 0.1$ have relatively smaller presence at the start. As the time passes by their percentage increases and peaks at around 300 days passed. After that, their percentage falls with time and follows similar pattern as the other ratings. In general, we find that users with higher value of η_u tend to post their ratings late.



(a) Distribution of user conformity



(b) Change in ratings caused by social opinions



(c) % of conformers as the item cascade unfolds

Fig. 4. Patterns of social conformity

5 Conclusion

We propose a novel formulation for the user ratings CRM that explicitly considers the social opinions. The CRM is shown to be effective in both improving the prediction accuracy of user's rating and in accurately identifying the social influencers. Further, several interesting patterns have emerged. We find that more than 76% of users show some degree of conformity with their friends. To our surprise, our friends opinion makes our posterior evaluation of the product more positive then negative, which is certainly a good news for the viral marketing strategy. Similar to the item product adopters, the users with high conformity tend to post their rating during the later part of the information cascade. We hope that the patterns found in this paper, would help in developing better recommendation systems and information propagation models.

References

1. Chen, W., Colin, A., Cumming, R., Ke, T., Liu, Z., Rincon, D., Sun, X., Wang, Y., Wei, W., Yuan, Y.: Influence maximization in social networks when negative opinion may emerge and propagate. In: SDM (2011)
2. Chevalier, J.A., Mayzlin, D.: The Effect of Word of Mouth on Sales: Online Book Reviews. Journal of Marketing Research 43(3), 345–354 (2005)
3. Gisser, M., McClure, J.E., Okten, G., Santoni, G.: Some anomalies arising from bandwagons that impart upward sloping segments to market demand. Econ. Journal Watch 6(1), 21–34 (2009)
4. Huang, J., Cheng, X.Q., Shen, H.W., Zhou, T., Jin, X.: Exploring social influence via posterior effect of word-of-mouth recommendations. In: WSDM (2012)
5. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: KDD (2003)
6. Ma, H., Zhou, D., Liu, C., Lyu, M.R., King, I.: Recommender systems with social regularization. In: WSDM (2011)
7. Salakhutdinov, R., Mnih, A.: Probabilistic matrix factorization. In: Advances in Neural Information Processing Systems, vol. 20 (2008)
8. Hwang, S.W., Lee, M.W.: A uncertainty perspective on qualitative preference. In: UAI (2009)