Are You Influenced by Others When Rating? Improve Rating Prediction by Conformity Modeling

Yiming Liu , Xuezhi Cao , Yong Yu
Apex Data and Knowledge Management Lab
Dept. of Computer Science and Engineering
Shanghai Jiao Tong University, Shanghai, China
goblinliu,cxz,yyu@apex.sjtu.edu.cn

ABSTRACT

Conformity has a strong influence to user behaviors, even in online environment. When surfing online, users are usually flooded with others' opinions. These opinions implicitly contribute to the user's ongoing behaviors. However, there is no research work modeling online conformity yet. In this paper, we model user's conformity in online rating sites. We conduct analysis using real data to show the existence and strength of conformity in these scenarios. We propose a matrix-factorization-based conformity modeling technique to improve the accuracy of rating prediction. Experiments show that our model outperforms existing works significantly (with a relative improvement of 11.72% on RMSE). Therefore, we draw the conclusion that conformity modeling is important for understanding user behaviors and can contribute to further improve the online recommender systems.

Keywords

User Behavior, Conformity, Rating Prediction

1. INTRODUCTION

Conformity implicitly shapes people's behaviors to group norms. The phenomenon widely exists in most communities and various scenarios, from political votings to dinner choices. Plenty of psychological researches aim at studying and understanding such behaviors [9]. Experiments also prove the existence and lead to several psychological theories.

Besides traditional real-world experiments, online social networks and services now provide a new opportunity to study and to utilize the social conformity phenomenon. People are flooded with other's opinions when surfing the Internet. When making choices online, users are normally influenced by these just received opinions to some extent. All the opinions and actions are well recorded by the service providers as action logs, providing large-scale valuable data for conformity analysis. However, there are only few research works targeting at analyzing conformity in online environment.

RecSys '16, September 15-19, 2016, Boston , MA, USA

DOI: http://dx.doi.org/10.1145/2959100.2959141

Figure 1: Examples of Public Opinions in Rating Sites

One of the most suitable platforms for analyzing user's online conformity is online rating site. We have various rating sites for different domains, such as e-commerce, movie and restaurant. For most rating sites, users are aware of the public opinion towards the item before rating it. The public opinions are normally highlighted in the item's detail page, including average rating and rating distribution of the item. We show two examples in Figure 1. Such information definitely contributes to the user's rating action to some extent. For example, when user rates a 4-star movie (public opinion), even if he thinks it only worth 2-star (truthful opinion), he might still rate it as 3-star (biased opinion) due to conformity.

The benefit of modeling online conformity is multi-folded. Firstly, it gives us opportunity to better understand the user's online behaviors. Following which, we can undo the effects caused by conformity and recover the user's true preferences towards the items, leading to the improvement of personalized recommender systems. To the best of our knowledge, there is no existing work that models the online conformity jointly with traditional user preference models to improve the quality of recommender systems.

In this paper, we aim at modeling user conformity and improving rating prediction in online rating sites. We first conduct analysis using real data to show the existence of online conformity. We conclude that the conformity do have a strong influence on rating scores. Then we target at improving the accuracy of rating prediction by integrating conformity modeling and matrix factorization techniques. Specifically, we model the strength of conformity by considering the group size, unanimity and cohesion. For analysis and evaluation, we collect and publish a large-scale real data set for this task. Experiments show that our conformity model out-performs existing non-conformity approaches by 11.72% on RMSE and 9.15% on MAE.

2. RELATED WORKS

Rating prediction is one of the core tasks in recommender system. To accurately solve this task, one must model the user's preferences and capture the item's characteristic precisely. Matrix factorization (MF) is the most widely used tool for this task, which represents both users and items in common low-dimension latent factor space [4]. Researchers later extend the traditional MF approach to further employ social relations [2], cross-domain user histories [6] and etc.

Existing user modeling approaches assume user's actions always reflect their preferences truthfully. However, the assumption does not always hold. A study on review-helpfulness voting concludes that users are not always truthful judges. Whether users vote at all, as well as the polarity of their votes, depend highly on the surrounding contents [8]. Krishnan et al. design and conduct a comparative experiment in [5], where users can rate twice before and after being exposed to the public opinions. Results prove the existence of conformity in this setting. However, due to the specialty of the experiment's methodology, we can not directly extend the conclusion to general online scenarios where user actions are more freely conducted comparing to actions in the controlled experiment.

3. DATASET

Due to the special requirement of conformity modeling, there is no public dataset suitable for this task. Besides rating logs, we also need the public opinions users received right before each rating action. As existing datasets only provide action logs over sampled users, the required information can not be accurately recovered.

We collect and publish a large-scale dataset that satisfies the requirement. The dataset can be freely accessed online[1]. We collect the data from Douban[2], one of China's largest movie review sites. We randomly sample 5,000 movies with more than 50 actions, and dump ALL rating actions for these movies. In total, we obtain 17,142,351 rating actions from 1,757,451 users. We recover the public opinions user received right before each rating action, including the number of reviewers, the average rating, and also the rating distribution for the target movie (how many reviewers rates 1-5 stars respectively).

4. CONFORMITY ANALYSIS

In this section, we demonstrate the existence of conformity in online rating sites by statistical analysis.

We show the correlation between the ratings given by users and the public opinions users received according to each movie. For simplicity, we consider only the averaged rating as the public opinion for analysis. We represent each rating log instance k as (r_k, p_k), where r_k is the rating and p_k is the averaged rating right before the action. For each movie, we group its rating logs by similar p_k, and report the average r_k for each group as the indicator of the rating distribution under that specific public opinion. We plot the results for four randomly selected movies in Figure 2. As we can notice, the rating distributions (the mean) clearly correlate with the public opinions users received.

Now we demonstrate the conformity phenomenon widely exists among most movies. Note that we can only conduct

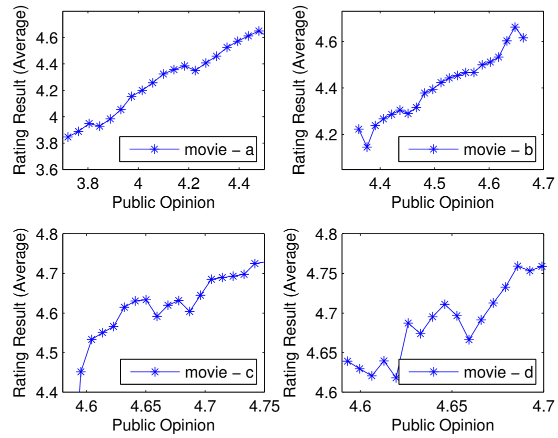

Figure 2: Influence of Public Opinion on User Ratings

the previous analysis in Figure 2 according to each movie separately. Plotting logs for all movies together is meaningless because rating distributions for different movies are not necessarily the same. Instead, we calculate the absolute correlation coefficient between $\{r_k\}$ and $\{p_k\}$ for each movie and plot the coefficient's distribution. The result is depicted in Figure 3, indicating that the correlation exists for a large portion of movies. It may raise concern that the numeric values are rather small. This is because the rating data is highly noisy (users can only rate integer values and are influenced by many other factors), thus perfect correlation is not possible.

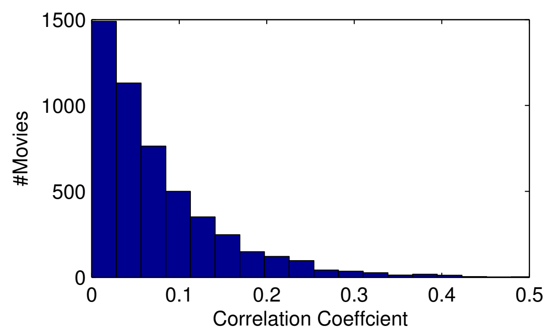

Figure 3: Correlation Coefficient Distribution

5. OUR APPROACH

We base our approach on matrix factorization [4], the most widely used technique for rating prediction. The traditional model captures user preferences and item characteristics with low-dimension latent factors. The rating user u gives to item i is modeled and estimated by:

$$\hat{r}_{ui} = t_{ui} = \mu + b_u + b_i + P_u^T Q_i \tag{1}$$

where μ, b_u, b_i are the global/user/item biases respectively, $P_u, Q_i \in \mathbb{R}^K$ are the latent factors. The loss function is:

$$\mathcal{L}(\Theta_t) = \frac{1}{2}\sum_{u,i}(r_{ui} - \hat{r}_{ui})^2 + \frac{\lambda}{2}||\Theta_t||^2 \tag{2}$$

[1]http://dataset.apexlab.org/conformity
[2]http://www.douban.com/

where Θ_t is the set of parameters. Experiments and online applications both indicate that matrix factorization can successfully capture the users' preferences towards the items.

Instead of using the preference-based estimation directly as the rating modeling ($\hat{r}_{ui} = t_{ui}$), we integrate it with conformity modeling. Specifically, we model the user's final rating as $\hat{r}_{ui} = \mathcal{C}(t_{ui}, p_{ui}, \Theta_{\mathcal{C}})$, where $\mathcal{C}$ is the conformity model, t_{ui} is the prediction of traditional preference-based model, p_{ui} is the public opinion user u receives right before she rates item i, $\Theta_{\mathcal{C}}$ is the set of parameters.

Now we discuss the design of the conformity model $\mathcal{C}$. As conformity is a group phenomenon, factors such as group size, unanimity and cohesion affect the level of conformity an individual displays. Therefore, we first propose the basic conformity model and then extend it for specific aspects in the following subsections.

5.1 Conformity Model - Basic ($\mathcal{C}_B$)

We model the user's biased rating by linear combination of pure preference-based estimation and the influence of public opinion due to conformity. For public opinions, we have $p_{ui} = (c_{ui}, a_{ui}, d_{ui})$, where c_{ui} is the number of ratings for movie i before user u rates it, a_{ui} is the average rating and d_{ui} is the rating distribution. Examples are showed in Figure 1. In the basic model we only consider the average rating a_{ui}. We use a hyper-parameter ω to tune the strength of conformity. The rating prediction model is now:

$$\begin{aligned} \mathcal{C}_B(t_{ui}, p_{ui}) &= (1-\omega)\cdot t_{ui} + \omega \cdot a_{ui} \\ &= (1-\omega)\cdot(\mu + b_u + b_i + P_u^T Q_i) + \omega \cdot a_{ui} \end{aligned} \tag{3}$$

We employ stochastic gradient descent for the parameter learning. The inference of the gradients as well as the loss function are mostly the same as traditional matrix factorization, thus we skip them here.

5.2 Conformity Model - Group Size ($\mathcal{C}_S$)

Group size is an important factor for modeling the conformity strength. Opinions from few people might be considered as coincidence, while the voice of the crowd can not be ignored. In the rating site scenario, the group size is the number of raters for the movie (c_{ui}). For example, knowing that 5 people like the movie might not affect the user a lot. However, when the group size increases to thousands, the conformity strength also boosts.

Therefore, instead of setting conformity strength with constant value as in the basic model, now we model it based on the group size. We model the conformity strength as a function of the group size. Considering the intuitive properties of conformity, we design the function to satisfies the followings: (a) non-linear; (b) monotone increasing; (c) with constrained range; (d) has a 'tipping point' to model the 'threshold' of group size for conformity. For computational convenience, we also require the function to be differentiable. To achieve the requirements, we employ the extended sigmoid function $\omega(x)$ defined as follows:

$$\sigma(x) = 1/(1+e^{-x}),\ \omega(x) = \omega_b + \omega_a \cdot \sigma(\alpha x + \beta) \tag{4}$$

where ω_a, ω_b control the overall scale of conformity strength and α, β control the shape as well as the 'tipping point'. The model is now:

$$\hat{r}_{ui} = \mathcal{C}_S(t_{ui}, p_{ui}) = (1-\omega(c_{ui}))\cdot t_{ui} + \omega(c_{ui})\cdot a_{ui} \tag{5}$$

The loss function is mostly the same with Eq. (2), with parameter set replaced by $\Theta_{\mathcal{C}_B} = (\Theta_t, \omega_a, \omega_b, \alpha, \beta)$.

We update the parameters for both traditional model and conformity model jointly using stochastic gradient decent. For parameters in traditional model, their gradients are simply the original gradients multiplied by $(1-\omega(c_{ui}))$ for each rating instance (u, i). For the newly introduced parameters, their gradients are as follows (with $e_{ui} = r_{ui} - \hat{r}_{ui}$):

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \omega_a} &= \lambda\omega_a - \sum_{u,i} e_{ui}(a_{ui} - t_{ui})\sigma(\alpha c_{ui} + \beta) \\ \frac{\partial \mathcal{L}}{\partial \omega_b} &= \lambda\omega_b - \sum_{u,i} e_{ui}(a_{ui} - t_{ui}) \\ \frac{\partial \mathcal{L}}{\partial \alpha} &= \lambda\alpha - \sum_{u,i} e_{ui}\omega_a(a_{ui} - t_{ui})\sigma(\alpha c_{ui} + \beta)(1 - \sigma(\alpha c_{ui} + \beta))c_{ui} \\ \frac{\partial \mathcal{L}}{\partial \beta} &= \lambda\beta - \sum_{u,i} e_{ui}\omega_a(a_{ui} - t_{ui})\sigma(\alpha c_{ui} + \beta)(1 - \sigma(\alpha c_{ui} + \beta)) \end{aligned} \tag{6}$$

5.3 Conformity Model - Cohesion ($\mathcal{C}_C$)

Cohesion and unanimity also affect whether and to what extent does conformity influence users. People may follow the majority for no reason when unanimity is reached. And when the cohesion is rather low (the public opinion varies), people tend to express their own opinions instead of following others.

In this scenario, the rating distribution d_{ui} is the indicator of unanimity and cohesion. We give two examples in Figure 1. Although the average ratings are both 3.5-stars, the distribution differs. Intuitively, users are more likely to rate around 4 for (a) comparing to (b) due to cohesion.

We leverage the rating distribution for the conformity model by employing Gradient Boosting Decision Tree (GBDT). GBDT can capture rather complex non-linear correlations, thus suitable for the cohesion modeling. We use XGBoost[1] as the implementation. Specifically:

$$\mathcal{C}_C(t_{ui}, p_{ui}) = xgb(\{\mathcal{C}_B(t_{ui}, p_{ui}), c_{ui}, a_{ui}, d_{ui}\}, \Theta_{xgb}) \tag{7}$$

where the first input for xgb represents the feature set and the second is the parameter set.

We need to split the training data into two disjoint sets for the inner model ($\mathcal{C}_B$) and the GBDT model respectively. If we train them using the same data set, the GBDT model would mostly use the output of the inner model as the final prediction because the inner model is well trained for these data samples. In experiments we use half training data for inner model and the rest for GBDT.

6. EXPERIMENTS

6.1 Experimental Settings

We already described the data set in Section 3, now we explain the evaluation metrics and comparing algorithms.

We employ mean absolute error (MAE) and root mean square error (RMSE) as the main metrics to evaluate the rating prediction models, defined by the following:

$$MAE = \frac{\sum_{u,i}|r_{ui} - \hat{r}_{ui}|}{\#logs},\ RMSE = \sqrt{\frac{\sum_{u,i}(r_{ui} - \hat{r}_{ui})^2}{\#logs}} \tag{8}$$

Table 1: Experimental Results, with 80% Data for Training

Metrics	UserAvg	ItemAvg	PMF	BiasedMF	SVD++	CM-B	CM-S	CM-C
MAE	0.8157	0.7754	0.7577	0.7502	0.7363	0.6798	0.6728	**0.6690**
Improvement	17.99%	13.73%	11.71%	10.83%	9.15%	1.60%	0.57%	
RMSE	1.0429	0.9952	0.9707	0.9752	0.9580	0.8595	0.8530	**0.8457**
Improvement	18.90%	15.02%	12.88%	13.28%	11.72%	1.61%	0.86%	

We evaluate our conformity-based models as well as several state-of-the-art rating prediction models in the experiments. The models are listed as follows:

UserAvg, ItemAvg: Directly predict by user/item's average rating according to training logs.

PMF: Probabilistic matrix factorization proposed in [7].

BiasedMF, SVD++: Original and revised version in [3].

CM-B, CM-S, CM-C: Conformity models proposed in this paper, according to $\mathcal{C}_B, \mathcal{C}_S, \mathcal{C}_C$ in Section 5 respectively.

All dimensions for latent factors are set to 5. Parameters are set by cross validation. We use 80% of data for training set and the rest 20% for testing.

Table 2: Experimental Results vs Training Ratio

Metric	Model	Training Ratio			
		50%	60%	70%	80%
MAE	SVD++	0.7650	0.7568	0.7650	0.7363
	CM-B	0.6845	0.6832	0.6812	0.6798
	CM-S	0.6806	0.6774	0.6749	0.6728
	CM-C	0.6883	0.6769	0.6741	0.6690
RMSE	SVD++	0.9784	0.9683	0.9829	0.9580
	CM-B	0.8640	0.8623	0.8586	0.8595
	CM-S	0.8606	0.8576	0.8551	0.8530
	CM-C	0.8645	0.8543	0.8564	0.8457

6.2 Results & Analysis

We show the experimental results in Table 1, which indicate the conformity models proposed in this paper out perform the state-of-the-art approaches that depend only on the rating matrix.

Conformity Strength. The parameter ω in the basic conformity model (CM-B) can be viewed as an indicator of conformity strength. $\omega = 0$ degenerates the model to the traditional matrix factorization model (BiasedMF), and $\omega = 1$ indicates users always follow the public opinion. By cross validation, we set ω to 0.9. The value is rather large and indicates strong conformity phenomenon in rating sites.

Conformity vs Group Size. By comparing results of CM-B and CM-S, we conclude the conformity strength does depend on the group size (number of raters for the item). By examining the parameters, we find that the conformity strength ranges over 0.4627 to 0.8868 for only 1 raters and infinite raters respectively.

Conformity vs Cohesion. Result of CM-C indicates the cohesion also contributes to the conformity-behavior. Due to the complexity of the decision trees, we do not depict them here. We notice that there are plenty of nodes in the decision trees use the normalized rating distribution (percentage) as the splitting criteria.

Effect of Training Ratio. We vary the training data ratio from 50% to 80% and show the results in Table 2. As we can notice, the performance of conformity model is rather robust to the training ratio. When training ratio drops from 80% to 50%, the performances of CM-B and CM-C only drop by less than 0.01 for both MAE and RMSE For CM-C, the performance drops slightly increases due to the learning requirement of GBRT (0.0194 and 0.0189 and for the two metrics). Note that the value is still better than traditional approaches. This is because conformity model captures user behaviors in general instead of personalized, thus suffers less from the cold-start problem comparing to traditional approaches.

7. CONCLUSION & FUTURE WORK

In this paper, we model the conformity phenomenon in online rating sites to understand the user behaviors and to improve the rating prediction accuracy. Specifically, we capture key factors for conformity strength including group size, unanimity and cohesion. We further integrate the conformity model with matrix factorization. Analysis and experiments using large-scale real data show the existence of online conformity and prove that it improves the rating prediction accuracy. For future works, we may consider having personalized conformity strength for each user and analyzing what kinds of user groups are more susceptive to public opinions.

8. REFERENCES

[1] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. *arXiv:1603.02754*, 2016.

[2] G. Guo, J. Zhang, and N. Yorke-Smith. Trustsvd: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings. In *AAAI*, pages 123–129, 2015.

[3] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *SIGKDD*, pages 426–434. ACM, 2008.

[4] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, (8):30–37, 2009.

[5] S. Krishnan, J. Patel, M. J. Franklin, and K. Goldberg. A methodology for learning, analyzing, and mitigating social influence bias in recommender systems. In *RecSys*, pages 137–144. ACM, 2014.

[6] B. Li, Q. Yang, and X. Xue. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction. In *IJCAI*, volume 9, pages 2052–2057, 2009.

[7] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. Citeseer, 2011.

[8] R. Sipos, A. Ghosh, and T. Joachims. Was this review helpful to you?: it depends! context and voting patterns in online content. In *WWW*, pages 337–348. ACM, 2014.

[9] J. W. Thibaut and H. H. Kelley. The social psychology of groups. 1959.