Integrating Trust and Similarity to Ameliorate the Data Sparsity and Cold Start for Recommender Systems

Guibing Guo
School of Computer Engineering
Nanyang Technological University, Singapore
gguo1@e.ntu.edu.sg

ABSTRACT

Our research aims to tackle the problems of data sparsity and cold start of traditional recommender systems. Insufficient ratings often result in poor quality of recommendations in terms of accuracy and coverage. To address these issues, we propose three different approaches from the perspective of preference modelling. Firstly, we propose to merge the ratings of trusted neighbors and thus form a new rating profile for the active users, based on which better recommendations can be generated. Secondly, we aim to make better use of user ratings and introduce a novel Bayesian similarity measure by taking into account both the direction and length of rating vectors. Thirdly, we propose a new information source called *prior ratings* based on virtual product experience in virtual reality environments, in order to inherently resolve the concerned problems.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation/methodology*

Keywords

Similarity, Trust, Similarity measure, Prior ratings

1. INTRODUCTION

Collaborative filtering (CF) [1] is an effective technique for recommender systems to provide users with high quality items of interest, tackling the problem of *information overload*. It makes use of the heuristic that the preferences of like-minded users are similar. However, CF inherently suffers from two issues, namely *data sparsity* and *cold start* [1]. The former issue refers to the difficulty in finding reliable similar users, given the fact that users only rate a small portion of items, while the latter refers to the difficulty in modelling user preference due to only few items rated by the active users (known as *cold users*). Our research aims to resolve these issues by better modelling user preferences from both user behaviors (i.e., ratings) and social connections (i.e., trusted friends).

RecSys'13, October 12–16, 2013, Hong Kong, China.

ACM 978-1-4503-2409-0/13/09 ...$15.00.
http://dx.doi.org/10.1145/2507157.2508071.

Trust has been extensively exploited to improve the predictive performance and ameliorate the concerned issues, given the strong and positive correlation with preference [21]. Trust is defined as one's belief towards the ability of others in providing valuable ratings. For example, users in Epinions.com can specify other users as trustworthy (to the 'web of trust') or untrustworthy (to the 'block list). Massa and Anesani [16] substitute similarity with trust in predicting items' ratings. Trust is allowed to propagate through the trust networks in order to incorporate more trusted neighbors and hence alleviate the data sparsity and cold start problems. It shows that more robust recommendations are obtained without significant loss in accuracy. Chowdhury et al. [5] suggest to evaluate a value for the similar users who have not rated the given item, according to the ratings of their trusted neighbors. In this way, more similar users can be involved in generating prediction. However, it is hard for cold users to find many reliable similar users in the first place. Ray and Mahanti [17] argue that trusted neighbors may have different preferences and by removing the trust links with low similarity can further improve the performance. The drawback is that only few items can be predicted. Although empirical studies [5, 17] show that better performance can be obtained, it seems not far that we have achieved in trust-aware recommender systems [20].

Another line of research is to design new similarity measures to make better use of existing user ratings, given the ineffectiveness of the traditional approaches [14] such as Pearson correlation coefficient and cosine similarity. It has been shown that they cannot function well when only few ratings are available or the length of rating vectors is ignored [15]. Lathia et al. [13] find the proportion of agreement and estimate the correlation between two users based on the number of concordant, discordant and tied pairs of common ratings. However, user similarity is not computable in the cases where ratings are flat or only a single rating pair is available. Ahn [2] suggests to highlight the importance of semantic differences between rating scales. The author introduces three semantic heuristics through which user similarity is obtained. The drawback is that the obtained value is not bounded (often greater than 1) and lack of meaningful indications. Bobadilla et al. [4] identify *singularity* as a critical element to distinguish user similarity, based on the intuition that users with high singularity from the majority are more

Table 1: The predictive performance on the Flixster data set

Views	Approaches measured in MAE, RC and F1									
	CF	MT1	MT2	MT3	RN	TCF1	TCF2	Merge1	Merge2	Merge3
All Users	0.928	1.060	0.932	0.862	0.858	0.870	**0.850**	0.890	0.877	**0.875**
	68.56%	12.36%	71.37%	90.71%	0.38%	80.92%	**85.23%**	89.64%	94.90%	**95.04%**
	0.7357	0.2128	0.7512	0.8549	0.0076	0.8079	**0.8312**	0.8466	0.8690	**0.8719**
Cold Users	1.153	1.127	1.005	0.934	NaN	1.047	**0.923**	1.008	0.960	0.949
	3.27%	8.11%	52.69%	**79.55%**	0.00%	12.97%	21.41%	63.03%	83.11%	**85.15%**
	0.0626	0.1464	0.6279	**0.7939**	NaN	0.2219	0.3373	0.6956	0.8083	**0.8191**

similar than those whose ratings are consistent with the others. However, the length of rating vectors is ignored, and the effectiveness is only verified on a single data set.

In addition, most current research focus on building advanced recommendation algorithms to exploit existing (behavioral or contextual) data, but only a few works have attempted to elicit more user ratings from the perspective of user interfaces. We argue that eliciting more user ratings can inherently mitigate the concerned issues. The emergence of 3D virtual reality (VR) environments such as SecondLife.com offers more useful information (than traditional website environments) that can be used to model user preference. In particular, products could be represented in the form of 3D models which allow users to interact efficiently and effectively and thereby gain intuitive, direct and concrete first-hand experience prior to purchase. Hence, it has great potential to better model user preference in VR. However, the research on recommender systems is still in its infancy. Eno et al. [6] analyze the content of 3D objects and the locations that users visited in VR environments from which user preference can be modelled. Shah et al. [18] propose a location recommender system by analyzing users' login data, and thus help users navigate in VR. Hu and Wang [11] implement a controlled prototype of a system to recommend virtual furniture according to users' interest and requirements. However, the features of VR are not well exploited to elicit more user ratings.

2. RESEARCH PROGRESS TO DATE

Following these research lines described in the previous section, we have conducted three different approaches to resolve the data sparsity and cold start problems. We will elaborate them in subsequent sections.

2.1 The Merge Method

Our work [8] aims to form a new rating profile for the active users, especially for the cold users who have only rated few items, by merging the ratings of trusted neighbors. Specifically, the ratings for a given item reported by trusted neighbors will be aggregated by:

$$\tilde{r}_{u,i} = \frac{\sum_{v \in TN_u} t_{u,v} r_{v,i}}{\sum_{v \in TN_u} t_{u,v}}, \tag{1}$$

where $\tilde{r}_{u,i}$ is the merged rating for the active user u on a given item i, and $r_{v,i}$ is the rating reported by a trusted user v. User u is also regarded as a trusted neighbor in the trust neighborhood TN_u, and $t_{u,v} \in [0,1]$ is the trustworthiness of user v from u's point of view (hence $t_{u,u} = 1$). The quality of a merged rating $\tilde{r}_{u,i}$ is measured as the certainty by taking into account two factors, namely the number of ratings and the conflicts between positive and negative opinions:

$$c_{u,i} = \frac{1}{2} \int_0^1 \left| \frac{x^{p_{u,i}}(1-x)^{n_{u,i}}}{\int_0^1 x^{p_{u,i}}(1-x)^{n_{u,i}} dx} - 1 \right| dx, \tag{2}$$

where $c_{u,i}$ is the certainty of merged rating $\tilde{r}_{u,i}$ as a function of $p_{u,i}$ and $n_{u,i}$, referring to the number of positive and negative ratings used for the merging, respectively. A rating is defined as *positive* if its value is greater than the median rating scale; otherwise, it is *negative*. Hence, the more number of and the less conflicts among ratings, the more certain that the merged rating represents the collective opinion. Only the ratings whose certainty is greater than a predefined threshold will be adopted as a useful merged rating. This procedure is continued until all the ratings of the items rated by the trusted neighbors are merged. As a result, a new rating profile is generated to represent the preferences of the active users. Based on the new rating profile, a traditional CF is then employed to make recommendations.

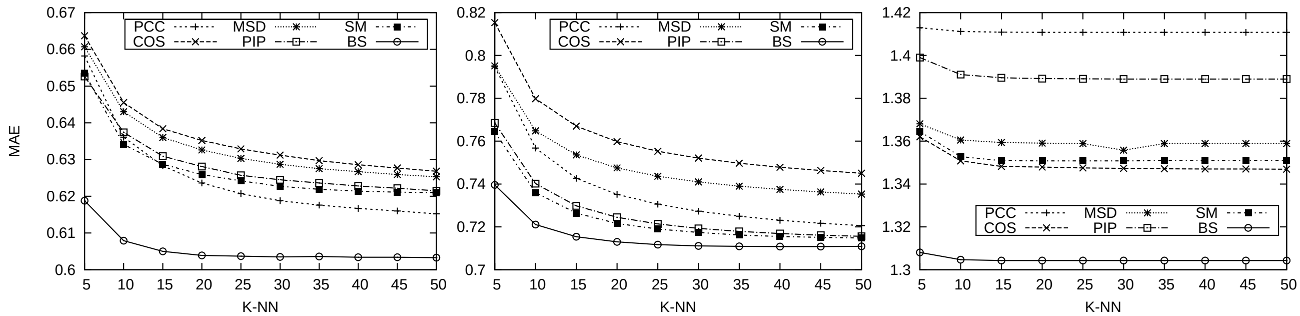
The effectiveness of our Merge approach is evaluated on three real-world data sets and compare with a batch of comparative methods as shown in Table 1[1]: **MT**x is the method proposed by Massa and Anesani [16] where x is the length of trust propagation; **RN** is the method of Ray and Mahanti [17]; **TCF**x is reported by Chowdhury et al. [5] where x is the number of iterations; **Merge**x is our method where x denotes the length of trust propagation. Table 1 shows that our method generally and consistently achieves the best performance: the relatively low *mean absolute errors* (MAE) and the highest *rating coverage* (RC) as well as the best F-measure (F1) of MAE and RC, especially in the view of *cold users*. Principally, the Merge method is capable of handling two extreme cold scenarios where (1) only trust is available; and (2) only ratings are available. Our method can survive in either case[2] to form an effective rating profile for the cold users. Although it fails to function if neither trust nor ratings are present, this is beyond the scope of our discussion. In summary, our method shades light on a new way to utilize social trust and alleviate the concerned problems.

2.2 Bayesian Similarity Measure

In the work of [10], we propose a novel Bayesian similarity measure by taking into account both the direction and length of rating profiles (vectors), with the aims to solve the issues of traditional approaches. Our approach consists of three components: *overall similarity* ($s'_{u,v}$), *chance correlation* ($s''_{u,v}$) and the *user bias* (δ). Formally, the user similarity between users u and v is computed by removing the

[1] The other two data sets are Epinions and FilmTrust, where similar results are obtained.

[2] In the second case, our method will be equivalent with CF since the only trusted neighbors are active users themselves.



(a) FilmTrust (b) MovieLens 1M (c) BookCrossing

Figure 1: The predictive accuracy of comparative approaches

chance correlation and user bias from the overall similarity:

$$s_{u,v} = \max(s'_{u,v} - s''_{u,v} - \delta, 0). \quad (3)$$

For the overall similarity, the Dirichlet distribution is adopted to accommodate the multinomial rating distances. The underlying assumption is that rating evidences (i.e., a pair of ratings from two users) towards different items are not equally useful in computing similarity. We also claim that realistic value can be derived from consistent item ratings, which is determined by two factors: (1) the standard deviation of ratings on a specific item; and (2) the rating tendency of all users. As new evidence arrives, the posterior probability density is updated. Therefore, we define the *user distance* as the average of rating distances weighted by their respective importance weights which reflect the extent to which evidences fall in such a rating distance. Lastly, the overall similarity is computed by inversely normalizing user distance. In practice, two users may be regarded as 'similar' if they happen to have small rating distances, especially when there are only a small number of user ratings. Hence, the chance correlation is defined as the probability that the amount of evidences fall in different distance levels independently. In addition, by investigating the nature (mean and standard deviation) of traditional and our approaches, we note that our method will generally hold a limited (yet much smaller) user bias, i.e., $\delta = 0.04$. We empirically demonstrate that removing chance correlation and user bias is helpful for similarity computation.

Six real-world data sets are used to evaluate the effectiveness of our approach (denoted by **BS**) in comparison with a number of benchmarks: three traditional approaches (PCC, COS, MSD) and two recently proposed approaches (PIP, SM). In particular, **PCC**, **COS**, **MSD** [19] refer to Pearson correlation coefficient, cosine similarity and mean square distance, respectively; **PIP** is proposed by Ahn [2] based on three semantic heuristics; **SM** is the singularity similarity introduced by Bobadilla et al. [4]. Due to space limitation, the results on three data sets are illustrated in Figure 1 which are similar with other data sets. It is concluded that our approach consistently outperforms the other counterparts in terms of MAE, i.e., better accuracy.

2.3 Prior Ratings

Our recent works [7, 9] introduce a new direction of research to inherently resolve the concerned problems by eliciting more kinds of user ratings. Traditional ratings are usually reported by users who purchased a product according to their after-use experience. However, users are not well-motivated to provide their ratings and hence result in the data sparsity problem. In contrast, we propose a new information source called *prior ratings* which are based on virtual product experience gained via real-time interactions with virtual products—the second presence of products in VR environments which allow users to view, rotate, zoom in and out, and virtually try on, etc. Accordingly, we term the traditional ratings as 'posterior ratings'.

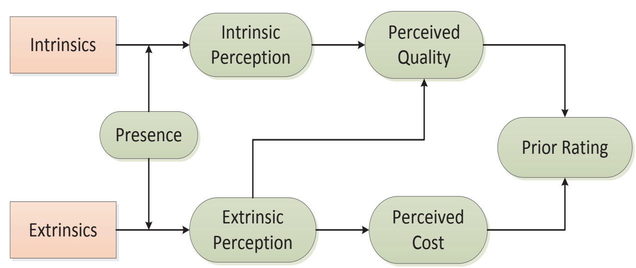


Figure 2: The conceptual model of prior ratings

Prior and posterior ratings differ in reflecting different forms of user experience. Posterior ratings are usually post-purchase ratings whereas prior ratings are likely to be prior-purchase ratings. The major difference lies in the confidence that users have. Specifically, posterior ratings may have higher confidence due to the experience derived from tangible products. However, prior ratings are more dense since users are more willing to provide their ratings partially due to the fact more types of media and interactions are supported in VR environments and hence more pleasant experience could be received. In addition, it is well understood that users browse or experience more products than they actually purchase, especially in the virtual environments where product information is easy to reach. Therefore, prior ratings could be greatly complementary to posterior ratings, and by leveraging both prior and posterior ratings, the concerned problems can be inherently alleviated.

To provide a principled foundation, we propose a conceptual model for prior ratings, as shown in Figure 2. For a specific product, a number of intrinsic and extrinsic *attributes* are associated. In different environments, the perceptions of these attributes could vary according to the types of media and interactions that deliver information about them. The intrinsic and extrinsic *perceptions* indicate the quality of products as perceived directly and indirectly, respectively.

In contrast, the *perceived cost* (e.g., time, price) refers to the cost that users have to bear in order to obtain the products. A prior rating is an overall evaluation of preference of products in terms of both perceived quality and cost. Five hypotheses are proposed regarding the relations among the components in the conceptual model.

A user study was designed in order to validate the proposed conceptual model and hypotheses. In particular, two user interfaces—a 2D website (WS) and 3D virtual store—were designed to represent two different kinds of virtual environments with distinct sense of presence. A number of subjects were recruited in the user study. The results demonstrated the validity of the conceptual model. Specifically, we found that presence has positive influence on the perceptions of some intrinsic and extrinsic attributes; perceived quality in VR mainly depends on intrinsic attributes while in WS it rests more on extrinsic attributes; both perceived quality and cost have a positive impact on prior ratings, if the price is acceptable. Further, the stronger sense of presence a virtual environment is, the more confident prior ratings will be and the closer to posterior ratings. Hence, it indicates that the design of virtual environments should enhance the sense of presence by increasing the media richness or the effectiveness of user interactions.

3. FUTURE RESEARCH

One potential drawback of our current research is the dependency on explicit trust. In most applications, users may not share or be willing to expose such information due to the concerns of, for example, privacy. Thus, we intend to infer users' trust from historical user behaviors and ensure that such modelling can approach explicit trust as close as possible. For this purpose, we may need to model and learn the trust factors from explicit trust information. On the other hand, dimension reduction-based approaches (e.g., matrix factorization [12]) have also been proposed to investigate the utility of trust. Hence, we intend to focus on model-based approaches to further resolve the concerned problems. In particular, clustering-based approaches will be our main focus due to its efficiency and scalability. The main issues for clustering-based approaches are the relatively low accuracy and coverage. However, recent research [3] has indicated that superior accuracy can be achieved if more sophisticated clustering methods are applied. To our knowledge, no dedicated approaches have been proposed to resolve the cold start problem for clustering-based approaches. We plan to incorporate both trust and similarity to better cluster users and improve the recommendation performance by alleviating the cold start problem.

4. REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[2] H. Ahn. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences*, 178(1):37–51, 2008.

[3] A. Bellogín and J. Parapar. Using graph partitioning techniques for neighbour selection in user-based collaborative filtering. In *Proceedings of the 6th ACM Conference on Recommender Systems*, RecSys, pages 213–216, 2012.

[4] J. Bobadilla, F. Ortega, and A. Hernando. A collaborative filtering similarity measure based on singularities. *Information Processing & Management*, 48(2):204–217, 2012.

[5] M. Chowdhury, A. Thomo, and B. Wadge. Trust-based infinitesimals for enhanced collaborative filtering. In *Proceedings of the 15th International Conference on Management of Data*, COMAD, 2009.

[6] J. Eno, G. Stafford, S. Gauch, and C. Thompson. Hybrid user preference models for Second Life and opensimulator virtual worlds. In *Proceedings of the 19th International Conference on User Modeling, Adaption and Personalization*, UMAP, pages 87–98, 2011.

[7] G. Guo and M. Elgendi. A new recommender system for 3D e-commerce: An EEG based approach. *Journal of Advanced Management Science*, 1(1), 2013.

[8] G. Guo, J. Zhang, and D. Thalmann. A simple but effective method to incorporate trusted neighbors in recommender systems. In *Proceedings of the 20th International Conference on User Modeling, Adaptation and Personalization*, UMAP, 2012.

[9] G. Guo, J. Zhang, D. Thalmann, and N. Yorke-Smith. Prior ratings: A new information source for recommender systems in e-commerce. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys, 2013.

[10] G. Guo, J. Zhang, and N. Yorke-Smith. A novel bayesian similarity measure for recommender systems. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, IJCAI, 2013.

[11] X. Hu and K. Wang. Personalized recommendation for virtual reality. In *Proceedings of the 2010 International Conference on Multimedia Technology*, ICMT, pages 1–5, 2010.

[12] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[13] N. Lathia, S. Hailes, and L. Capra. Private distributed collaborative filtering using estimated concordance measures. In *Proceedings of the 2007 ACM Conference on Recommender Systems*, RecSys, pages 1–8, 2007.

[14] N. Lathia, S. Hailes, and L. Capra. The effect of correlation coefficients on communities of recommenders. In *Proceedings of the 23rd Annual ACM Symposium on Applied Computing*, SAC, pages 2000–2005, 2008.

[15] H. Ma, I. King, and M. Lyu. Effective missing data prediction for collaborative filtering. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR, pages 39–46, 2007.

[16] P. Massa and P. Avesani. Trust-aware recommender systems. In *Proceedings of the 2007 ACM Conference on Recommender Systems*, RecSys, pages 17–24, 2007.

[17] S. Ray and A. Mahanti. Improving prediction accuracy in trust-aware recommender systems. In *Proceedings of the 43rd Hawaii International Conference on System Sciences*, HICSS, pages 1–9, 2010.

[18] F. Shah, P. Bell, and G. Sukthankar. A destination recommendation system for virtual worlds. In *Proceedings of the 21st International Florida Artificial Intelligence Research Society Conference*, FLAIRS, pages 475–476, 2010.

[19] U. Shardanand and P. Maes. Social information filtering: algorithms for automating âĂIJword of mouthâĂİ. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, SIGCHI, pages 210–217, 1995.

[20] Y. Shi, M. Larson, and A. Hanjalic. How far are we in trust-aware recommendation? In *Proceedings of the 33rd European Conference on Advances in Information Retrieval*, ECIR, pages 704–707. Springer, 2011.

[21] P. Singla and M. Richardson. Yes, there is a correlation: from social networks to personal behavior on the web. In *Proceedings of the 17th International Conference on World Wide Web*, WWW, pages 655–664, 2008.