SPMC: Socially-Aware Personalized Markov Chains for Sparse Sequential Recommendation

Chenwei Cai, Ruining He, Julian McAuley

University of California, San Diego

{cwcai, r4he, jmcauley}@ucsd.edu

Abstract

Dealing with sparse, long-tailed datasets, and *cold-start* problems is always a challenge for recommender systems. These issues can partly be dealt with by making predictions not in isolation, but by leveraging information from related events; such information could include signals from *social relationships* or from the *sequence* of recent activities. Both types of additional information can be used to improve the performance of state-of-the-art matrix factorization-based techniques. In this paper, we propose new methods to combine both social and sequential information simultaneously, in order to further improve recommendation performance. We show these techniques to be particularly effective when dealing with sparsity and *cold-start* issues in several large, real-world datasets.

1 Introduction

Cold-start problems are a barrier to the performance of recommender systems that depend on learning high-dimensional user and item representations from historical feedback. In the one-class collaborative filtering setting, two types of techniques are commonly used to deal with such issues: *sequential* recommender systems and *social* recommender systems. The former assume that users' actions are similar to those they performed recently, while the latter assume that users' actions can be predicted from those of their friends. In both cases, these related activities act as a form of 'regularization,' improving predictions when users and items have too few observations to model them in isolation.

Two existing models of interest are *Factorized Personalized Markov Chains* (FPMC) [Rendle *et al.*, 2010a] and *Social Bayesian Personalized Ranking* (SBPR) [Zhao *et al.*, 2014]. In FPMC, the authors make use of 'personalized' Markov chains to train on sequences of users' baskets; FPMC is shown to enhance overall performance, though does little to address cold-start issues, due to the large number of parameters that need to be included to handle sequential recommendation. Among a wide variety of work that makes use of social factors, SBPR uses a ranking formulation to assume that items considered by a user's friends are more important
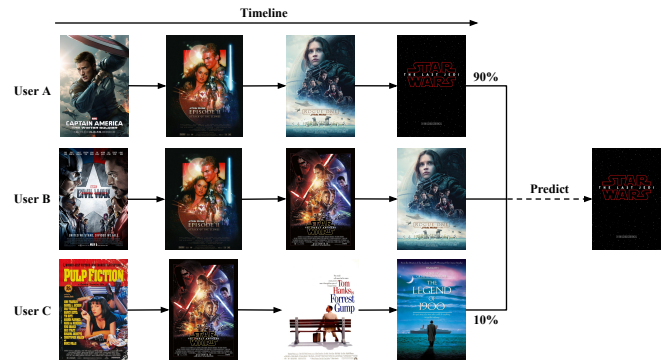

Figure 1: High-level idea of the SPMC model. By simultaneously modeling the sequential activities of users and their friends, SPMC predicts the next action of a user (e.g. B in this example) to be the combination of (1) her tastes (fantasy movies), (2) her own recent action (*Rogue One*), and (2) those recent actions from her friends (i.e., *The Last Jedi* from friend A, *The Legend of 1900* from friend C). SPMC learns to weigh different friends and make the final prediction (e.g. A's recent activities are more influential on B in this example).

than items *not* viewed by the user or their friends, but less important than items the user viewed themselves. This is a natural extension of standard ranking-based objectives, such as Bayesian Personalized Ranking (BPR) [Rendle *et al.*, 2009], and leads to increases in accuracy in cold-start settings. Similar techniques have also been used to rank and regularize based on other factors, such as groups of users who behave similarly [Pan and Chen, 2013].

In this paper we propose a new model, SPMC (Socially-aware Personalized Markov Chain) that leverages feedback from sequences, as well as social interactions in the same model. The model is based on the assumption that a user can be affected both by their own feedback sequence as well as that of their friends' (Figure 1). Our goal in doing so is to improve upon existing models, especially when dealing with user *cold-start* issues in sparse datasets.

In essence, our model is a combination of FPMC and SBPR, such that users' actions are assumed to be determined by (1) their preferences; (2) their recent activities; and (3) their friends' recent activities. We make several simplifying

assumptions to deal with the otherwise prohibitive number of parameters introduced by such a general formulation, especially when dealing with sparse datasets. Experiments on four real-world datasets reveal that the model is capable of beating state-of-the-art recommendation techniques that consider sequential and social factors in isolation.

2   Related Work

The most closely related works to ours are (1) Item recommendation methods that model user preferences in terms of latent factors; (2) Works that model sequential dynamics (and more broadly temporal information); and (3) Socially-regularized recommender systems.

Item recommendation. Item recommendation usually relies on Collaborative Filtering (CF) to learn from explicit feedback like star-ratings [Ricci *et al.*, 2011]. Although several paradigms for explicit feedback exist, of most relevance to us are *model-based* methods, including Bayesian methods [Miyahara and Pazzani, 2000; Breese *et al.*, 1998], Restricted Boltzmann Machines [Salakhutdinov *et al.*, 2007], and in particular Matrix Factorization (MF) methods (the basis of many state-of-the-art recommendation approaches such as [Bell *et al.*, 2007; Bennett and Lanning, 2007; Paterek, 2007]).

Such models have been extended in order to tackle implicit feedback data where only positive signals (e.g. purchases, clicks, thumbs-up) are observed (i.e., the so-called 'one-class' recommendation setting). Most relevant here are pair-wise methods like BPR-MF [Rendle *et al.*, 2009] that make an assumption that positive feedback instances are simply 'more preferable' than non-observed feedback. These are the same types of assumptions that have previously been adapted to handle implicit social signals, in systems like SBPR [Zhao *et al.*, 2014].

Sequential recommendation. Markov chains are powerful methods for modeling stochastic transitions; they have been leveraged to model decision processes (e.g. [Shani *et al.*, 2005]) and more generally uncover sequential patterns (e.g. [Zimdars *et al.*, 2001; Mobasher *et al.*, 2002]). In the sequential recommendation domain, Rendle *et al.* proposed FPMC that combines MF and (factorized) Markov chains to be able to capture personalization and sequential patterns simultaneously [Rendle *et al.*, 2010b]. Our work follows this thread but extends these ideas by making use of social, in addition to sequential, dynamics.

Social recommendation. In the recommender systems literature, there has been a large body of work that models social networks for mitigating cold-start issues in recommender systems, e.g. [Chaney *et al.*, 2015; Zhao *et al.*, 2014; Ma *et al.*, 2009; 2008; Guo *et al.*, 2015]. For example, *regularization*-based methods (e.g. [Jamali and Ester, 2010; Ma *et al.*, 2011]) assume that users' preferences should be similar to those of their social circles. Given the social network information $\mathcal{F}_u$ of user u, this framework uses regularization to force u's preference factors γ_u^U to be close to those users in $\mathcal{F}_u$. Finally, feedback and social regularization are optimized simultaneously. Likewise, *joint factorization*-based methods (e.g. [Tang *et al.*, 2013; Ma *et al.*, 2008]) try

Table 1: Notation

Notation	Explanation
$\mathcal{U}, \mathcal{I}$	user set, item set
u, i	user $u \in \mathcal{U}$; item $i \in \mathcal{I}$
$\mathcal{I}_u^+$	'positive' item set for user u; $\mathcal{I}_u^+ \subseteq \mathcal{I}$
$\mathcal{F}_u$	the friend set of user u; $\mathcal{F}_u \subseteq \mathcal{U}$
$\widehat{x}_{u,i,l}$	predicted score u gives to i given last item l
β_i	bias term associated with i; $\beta_i \in \mathbf{R}$
γ_u^U, γ_i^I	latent factors of u / i; $\gamma_u^U, \gamma_i^I \in \mathbf{R}^{K_1}$
θ_i^I, θ_i^L	latent factors of item i; $\theta_i^I, \theta_i^L \in \mathbf{R}^{K_2}$
M_i, N_i	latent factors of item i; $M_i, N_i \in \mathbf{R}^{K_3}$
W_u, V_u	latent factors of user u; $W_u, V_u \in \mathbf{R}^{K_4}$
K_1, K_2, K_3	dimensionality of different latent factors
Θ	parameter set
α	hyperparameter weighting the influence of the friend set
$\sigma(\cdot)$	sigmoid function; $\sigma(z) = 1/(1 + e^{-z})$
$\mathbf{1}(\cdot)$	indicator function; $\mathbf{1}(b) = 1$ *iff* b is *true*

to find a factorization of the social network matrix such that the resulting user representation can be directly used to explain users' preferences.

Our work differs from such socially-aware recommendation models mainly in that our method is sequentially-aware which not only makes it good at making predictions in a sequential manner—a desirable feature of recommender systems, but also perform well in cold-start settings. On the other hand, we propose to model the impact of the *recent* activities of a user's friends on his/her own future activities, instead of assuming the closeness amongst social circles in terms of long-term preferences. In this paper, we also empirically compare against several state-of-the-art socially-aware recommendation methods and demonstrate the effectiveness of modeling such socio-temporal dynamics.

3   The SPMC Model

In this paper, we focus on modeling implicit feedback, e.g. clicks, purchases, or thumbs-up. In addition to the feedback itself, we assume that timestamps are also available for each action, as well as the social relations (or trust relationships) of each user.

Let $\mathcal{U}$ denote the user set and $\mathcal{I}$ the item set. For each user $u \in \mathcal{U}$, we use $\mathcal{I}_u^+$ to denote the set of items toward which the user u has expressed positive feedback. $\mathcal{F}_u$ is used to denote the set of users that user u trusts, or the set of u's friends. The objective of our task is to predict the sequential behavior of users given the above information on *sparse* datasets, where dealing with user *cold-start* issues is paramount.

Notation used throughout this paper is summarized in Table 1.

3.1   Model Specifics

Our SPMC model is built on top of a state-of-the-art sequential predictor named *Factorized Personalized Markov Chains* (FPMC) [Rendle *et al.*, 2010a]. In FPMC, given a user u and

$$\widehat{x}_{u,i,l} = \underbrace{\langle \gamma_u^U, \gamma_i^I \rangle}_{\text{user preferences}} + \underbrace{\langle \theta_i^I, \theta_l^L \rangle}_{\text{sequential dynamics}} + \underbrace{\frac{2}{|\mathcal{F}_u|^\alpha} \sum_{u' \in \mathcal{F}_u, i'} \sigma\big(\langle W_u, V_{u'} \rangle\big) \cdot \langle M_i, N_{i'} \rangle}_{\text{socio-temporal dynamics}} + \underbrace{\beta_i}_{\text{item bias}} \quad (1)$$

Figure 2: The proposed socially- and sequentially-aware predictor.

the last item they interacted with $l \in \mathcal{I}_u^+$, the probability that u transitions to another item i is proportional to

$$\widehat{x}_{u,i,l} = \langle \gamma_u^U, \gamma_i^I \rangle + \langle \theta_i^I, \theta_l^L \rangle, \quad (2)$$

where the first inner product models the 'affinity' between latent user factors $\gamma_u^U \in \mathbf{R}^{K_1}$ and latent item factors $\gamma_i^I \in \mathbf{R}^{K_1}$, and the second models the 'continuity' between item i and the previous item l. $\theta_i^I, \theta_l^L \in \mathbf{R}^{K_2}$ are latent representations of item i and the last item l respectively.

The above predictor is capable of capturing both personalization and sequential dynamics. However it is unaware of the social signals in the system which are potentially important side information especially in cold-start scenarios. In particular, we argue that the recent actions of a user's friends could be influential when determining which action a user is likely to perform next.

Our SPMC model is a combination of personalization, sequential dynamics, as well as socio-temporal dynamics. The predictor of our model is simply the sum of three such components (see Eq. (1)), where i' is the item viewed by one of u's friends u' most recent to user u's feedback on item i. The inner product of the latent factors of i and i' ($M_i, N_{i'} \in \mathbf{R}^{K_3}$) models the impact from friends' recent actions.

Note that intuitively different friends could have a different amount of impact on a user. Therefore we measure the 'closeness' between two users by the inner products of their latent representations ($W_u, V_{u'} \in \mathbf{R}^{K_4}$), which are normalized via a sigmoid function ($\sigma(\cdot)$) to be between 0 (no influence) and 1 (high influence).

Finally, a bias term $\beta_i \in \mathbf{R}$ is added to the formulation to capture the overall popularity of item i. $\alpha \in \mathbf{R}$ is a parameter that balances social effects against other factors.

3.2 Merging the Embeddings

Our goal is to deal with user *cold-start* issues (i.e., users who have performed few previous actions), which requires us to reduce the number of parameters to be inferred. To this end, we merge the embeddings in Eq. (1) and obtain the following predictor:

$$\widehat{x}_{u,i,l} = \langle \gamma_u^U, \gamma_i^I \rangle + \langle \theta_i^I, \theta_l^L \rangle + \\ \frac{2}{|\mathcal{F}_u|^\alpha} \sum_{u' \in \mathcal{F}_u, i'} \sigma\big(\langle W_u, W_{u'} \rangle\big) \cdot \langle M_i, M_{i'} \rangle + \beta_i. \quad (3)$$

Note that the above equation merges (1) θ^I and θ^L into the same space θ^I, (2) W and V into the same space W, and (3) M and N into the same space M. Further merges are possible but avoided here as they empirically resulted in degraded performance, presumably because they sacrifice too much of the model's expressive power.

3.3 Learning the Model

An advantage of SPMC is that the same training framework from FPMC can be used to optimize the personalized total order $>_{u,l}$. In particular, Maximum a Posteriori (MAP) estimation of our parameters can be formulated as:

$$\arg\max_{\Theta} = \ln \prod_{u \in \mathcal{U}} \prod_{i \in \mathcal{I}_u^+} \prod_{j \neq i} p(i >_{u,l} j | \Theta) \, p(\Theta) \\ = \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}_u^+} \sum_{j \neq i} \ln p(i >_{u,l} j | \Theta) + \ln p(\Theta), \quad (4)$$

where l is the item preceding i in user u's feedback sequence. The probability that user u prefers item i over item j given u's last item l is given by

$$p(i >_{u,l} j | \Theta) = \sigma(x_{u,i,l} - x_{u,j,l}). \quad (5)$$

The full set of parameters of SPMC is $\Theta = \{\beta_{i \in \mathcal{I}}, \gamma_{u \in \mathcal{U}}^U, \gamma_{i \in \mathcal{I}}^I, \theta_{i \in \mathcal{I}}^I, W_{u \in \mathcal{U}}, M_{i \in \mathcal{I}}\}$. We uniformly sample from the dataset a user u, a positive item i and an item j that is different from the positive item, following the same negative-item selection protocol used in FPMC. Since the last item l and the relevant actions from the friends $\mathcal{F}_u$ are determined by the pair (u, i), we can apply standard stochastic gradient ascent to optimize Eq. (4) and update the parameters as

$$\Theta^{t+1} \leftarrow \Theta^t + \eta \cdot \big(\sigma(\widehat{x}_{u,j,l} - \widehat{x}_{u,i,l}) \frac{\partial(\widehat{x}_{u,i,l} - \widehat{x}_{u,j,l})}{\partial \Theta} - \lambda_\Theta \Theta \big), \quad (6)$$

where η is the learning rate and λ_Θ is a regularization hyperparameter. For completeness, we list the partial derivative of $\widehat{x}_{u,i,l} - \widehat{x}_{u,j,l}$ with respect to our parameters in Appendix A.

4 Experiments

To fully evaluate the effectiveness of our proposed model, we perform extensive experiments on a series of real-world datasets and compare against state-of-the-art sequentially- and socially-aware methods.

4.1 Datasets

We experiment on four datasets, each comprising a large corpus of user feedback, timestamps, as well as social relations (i.e. 'trusts'). All datasets are available online.

Ciao. Ciao is a review website where users give ratings and opinions on various products. This dataset was crawled by [Tang *et al.*, 2012] from Ciao's official site.[1] The feedback was given in the month of May, 2011. The dataset is available online.[2]

---

[1] http://www.ciao.co.uk/
[2] http://www.cse.msu.edu/~tangjili/trust.html

Table 2: Dataset statistics

	Ciao	Foursquare	Epnions	Flixster
#users	1,708	48,976	5,261	60,692
#items	16,485	29,879	25,996	48,549
#feedback	34,494	262,105	46,732	7,956,727
#trusts	35,753	231,506	23,915	840,141
#votes/#users	20.20	5.35	8.88	131.10
#trusts/#users	20.93	4.72	4.55	13.84

Foursquare. This dataset is from Foursquare.com[3] and consists of check-ins of users at different venues, spanning December 2011 to April 2012. While the dataset includes features other than just social and sequential signals (such as geographical data), these are beyond the scope of this paper.

Epinions. Epinions is a popular online consumer review website. Collected by [Zhao *et al.*, 2014], this dataset also contains trust relationships amongst users and spans more than a decade, from January 2001 to November 2013.

Flixster. Flixster is a social movie website where users can rate movies and share their reviews. This dataset is available online.[4]

In our experiments, we treat all observed interactions (i.e., ratings etc.) as positive instances, such that our goal is to rank items that a user would be likely to interact with.

4.2 Evaluation Protocol

The evaluation protocol we adopt is similar to previous works on sequential predictions, e.g. [He *et al.*, 2016]. From each user one positive item $\mathcal{T}_u \in \mathcal{I}_u^+$ is held out for testing, and another item $\mathcal{V}_u \in \mathcal{I}_u^+$ is held out for validation. The test item $\mathcal{T}_u$ is chosen to be the most recent item according to user u's feedback history, while the validation item $\mathcal{V}_u$ is the second most recent one. The rest of the data is used as the training set. We report the accuracies of all models in terms of the AUC (*Area Under the ROC Curve*) metric on the test set:

$$AUC = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{|\mathcal{I} \setminus \mathcal{I}_u^+|} \sum_{j \in \mathcal{I} \setminus \mathcal{I}_u^+} \mathbf{1}(\widehat{x}_{u,\mathcal{T}_u,l_{\mathcal{T}_u}} > \widehat{x}_{u,j,l_{\mathcal{T}_u}}),$$
(7)

where $\mathbf{1}(\cdot)$ is the indicator function, and $l_{\mathcal{T}_u}$ is the item preceding the test item $\mathcal{T}_u$ in u's feedback history.

4.3 Baselines

We compare against state-of-the-art recommendation models, i.e. BPR-MF and FPMC, as well as the socially-aware methods including GBPR and SBPR.

- **BPR-MF**: This model is described in [Rendle *et al.*, 2009], which is a state-of-the-art matrix factorization method that focuses on modeling user preferences.
- **FPMC**: This model is described in [Rendle *et al.*, 2010a], which is a state-of-the-art sequential prediction model. It unifies the power of matrix factorization at

---
[3]https://foursquare.com/
[4]http://www.cs.ubc.ca/~jamalim/datasets/

modeling users' preferences and the strength of Markov chains at capturing sequential continuity. Its predictor is presented in Eq. (2).

- **SBPR**: This is a state-of-the-art recommendation model that benefits from modeling social relations. Introduced by [Zhao *et al.*, 2014], the model is based on the assumption that users and their social circles should have similar tastes/preferences towards items.
- **GBPR**: A well-known model introduced by [Pan and Chen, 2013]. It makes use of group information, where users in a group have positive feedback on the same item.

Note that all methods optimize the AUC metric on the training set and the best hyperparameters are selected with grid search using the validation set.

The above baselines are selected to demonstrate that SPMC can outperform (1) state-of-the-art matrix factorization models that are unaware of sequential and social signals (i.e., BPR-MF); (2) methods that model both user preferences and sequential dynamics but no social signals (i.e., FPMC); and (3) models that are aware of social signals but ignore sequential dynamics (i.e., SBPR and GBPR).

4.4 Performance Analysis

Our goal is to mitigate the user cold-start problem which is present in many real-world datasets. To this end, for each of the four datasets introduced earlier, we obtain a series of user cold-start datasets by varying a threshold N. N is the maximum number of recent feedback instances each user can keep in his/her feedback history. In other words, if the number of user u's observed interactions exceeds the threshold N in the original dataset, only the most recent N instances will be kept.[5] The threshold N is no less than 4 since we need at least 4 feedback instances for a user to form a sequence in the training set. This protocol allows us to measure the performance improvements as the level of 'coolness' varies.

For fair comparison, we set the dimensions of latent factors in all models to 20, i.e., for SPMC $K_1 = K_2 = K_3 = 20$. Empirically, using larger dimensionality did not yield significant improvements for any of the methods being compared, presumably because these datasets are sparse and a large number of parameters would be unaffordable. We experimented with learning rates $\eta \in \{0.5, 0.05, 0.005\}$ and regularization hyperparameters $\lambda \in \{1, 0.1, 0.01, 0.001\}$, selecting the values that resulted in the best performance on the validation set. We set $\alpha = 1$ for all datasets (the sensitivity of α will be discussed later). For GBPR, the group size is set to 3 and ρ is set to 0.8. We implemented SBPR-2 as described in the original paper [Zhao *et al.*, 2014].

Table 3 shows the average AUCs of all models on the test sets as we vary the threshold N. We give the percentage improvement of our model over FPMC as well as the best performing baselines in the last two columns of this table. Figure 3 shows the trends of average AUCs with the thresholds. From these results, we find that:

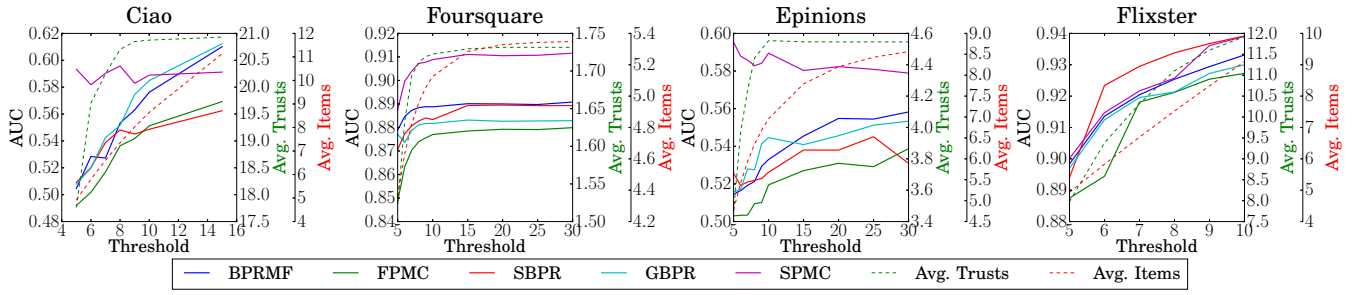[5]If the number of u's feedback instances is less than the threshold N, all the feedback will be used.

Figure 3: Average AUCs on Ciao, Foursquare, Epinions and Flixster as dataset sparsity varies. The x-axis is the threshold N controlling the amount of feedback used. Each dataset is filtered according to N, resulting in datasets where the amount of feedback per user does not exceed the threshold. Dashed green lines show the average number of social relations for the users, while dashed red lines show the average amount of feedback per user.

- As shown in Table 3, in very cold settings where thresholds are set to 5, SPMC always outperforms baselines.

- Figure 3 and Table 3 both show that SPMC can outperform other models in cold-start regions (e.g. threshold from 5 to 10) on most datasets.

- Figure 3 also shows that SPMC can significantly outperform other baselines on Foursquare and Epinions even when the threshold is set to large values. By comparing Foursquare and Epinions with the other datasets, we can see that the average number of items per user is comparatively much lower on these two datasets. This means that the majority of users on Foursquare and Epinions are actually 'cold' and thus they favor models that are strong in handling such cases.

- SPMC can always outperform state-of-the-art socially-unaware sequential method—FPMC. This means that it is important to model social signals in order to benefit from such auxiliary information especially in cold-start settings.

In conclusion, by combining personalization, sequential dynamics, and socio-temporal information carefully, our proposed model SPMC considerably outperforms all baselines that model these signals in isolation in user *cold-start* settings.

4.5 Convergence

We proceed by demonstrating comparison of convergence rates of all methods on the four datasets. Figure 4 shows the AUCs of SPMC and baselines on the test sets when the thresholds are set to 5. As we can see from this figure, the convergence efficiency of SPMC is comparable to all baselines (converging in fewer than 100 iterations),[6] although it is the integration of multiple sources of signals and is comparably complicated in its form.

4.6 Sensitivity

Next we demonstrate the sensitivity of SPMC to different hyperparameters. Figure 5 shows the changes in AUC of SPMC

---

[6]Note that each training iteration (of all methods) is a sweep of all positive feedback in the training set.
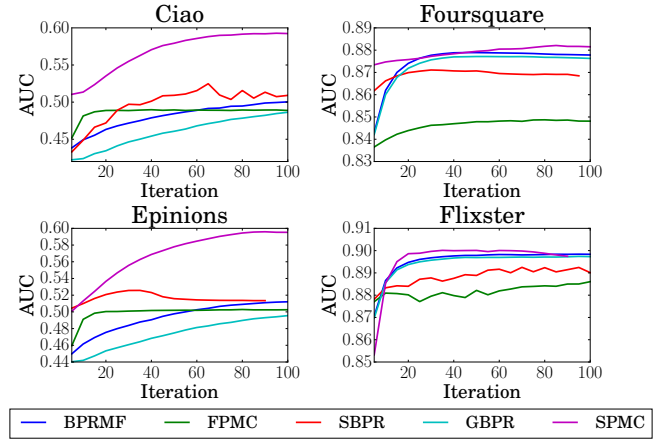

Figure 4: Learning curves of SPMC and baselines on Ciao, Foursquare, Epinions and Flixster. The threshold N is set to 5. SPMC converges roughly as fast as other methods.
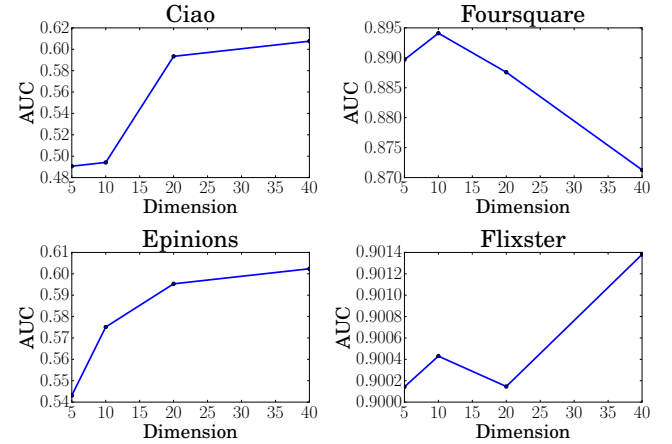

Figure 5: The AUCs of SPMC on four datasets when the dimensionality K varies. The model is more stable on larger datasets—Foursquare and Flixster. Note that AUCs are changing on a smaller scale on Foursquare and Flixster.

Table 3: AUCs of all models on the test sets as the threshold N varies. The last two columns compare SPMC to FPMC as well as the best performing baseline. The best performing method is boldfaced (higher is better).

Dataset	Threshold (N)	(a) BPR-MF	(b) FPMC	(c) SBPR	(d) GBPR	(e) SPMC	improvement e vs. b	e vs. best
Ciao	5	0.504614	0.491940	0.509185	0.507940	**0.593383**	20.62%	16.54%
	10	0.576186	0.551231	0.548679	0.584891	**0.588934**	6.84%	0.69%
	15	0.610123	0.569248	0.562466	**0.612324**	0.591179	3.85%	-3.45%
Foursquare	5	0.878899	0.848644	0.870693	0.877121	**0.887596**	4.59%	0.99%
	10	0.888771	0.876997	0.883340	0.881735	**0.908859**	3.63%	2.26%
	15	0.890104	0.878499	0.889350	0.883149	**0.911066**	3.71%	2.36%
Epnions	5	0.514541	0.503011	0.525640	0.515770	**0.595317**	18.35%	13.26%
	10	0.532994	0.519503	0.526204	0.544611	**0.589500**	13.47%	8.24%
	15	0.545447	0.527076	0.538011	0.540879	**0.580287**	10.10%	6.39%
Flixster	5	0.898370	0.887483	0.894095	0.897011	**0.900146**	1.43%	0.20%
	10	0.929493	0.927250	**0.939176**	0.929989	0.939046	1.27%	-0.01%
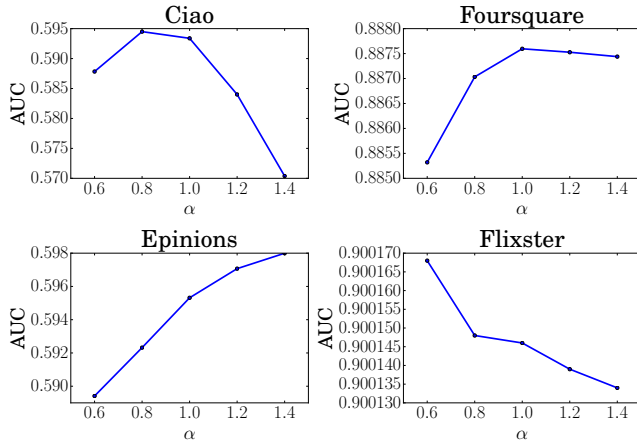
Figure 6: AUCs on the four datasets achieved by SPMC as we vary the hyperparameter α. Note that α seems to be the best when it is around 1.0, which essentially means that we are simply averaging the impacts from all friends.

on the four datasets as the dimensionality (here we adopt $K_1 = K_2 = K_3 = K$) increases from 5 to 40. We keep $\alpha = 1$, threshold $N = 5$, and set $K \in \{5, 10, 20, 40\}$. From the figure we can see that the AUC does not improve significantly when the number of dimensions is larger than 20 in most cases, which is expected as each user is only associated with around 5 activities. In addition, it seems that our model is more stable on Foursquare and Flixster, presumably because their sizes are much larger than Ciao and Epinions.

We show the variation in AUCs of SPMC as we vary the values of α in Figure 6. Empirically, it seems that the best performance can be achieved when α is around 1.0. This makes sense as it essentially means that we should take the Arithmetic Mean of the impacts from each friend.

5 Conclusion

In this paper, we proposed a new method, SPMC, to exploit both sequential and social information for recommenda-

tion. By combining different sources of signals carefully, our method beats state-of-the-art recommendation methods especially in user *cold-start* settings. We evaluated our model on four large, real-world datasets—Ciao, Foursquare, Epinions, and Flixster. Our experiments demonstrate that the model is capable of tacking different levels of *cold-start* issues.

Appendix A

Partial derivatives of $\widehat{x}_{u,i,l} - \widehat{x}_{u,j,l}$ with respect to our parameters are given by:

$$\frac{\partial}{\partial \beta_i} = 1; \quad \frac{\partial}{\partial \beta_j} = -1$$

$$\frac{\partial}{\partial \gamma_i^I} = \gamma_u^U; \quad \frac{\partial}{\partial \gamma_j^I} = -\gamma_u^U; \quad \frac{\partial}{\partial \gamma_u^U} = \gamma_i^I - \gamma_j^I$$

$$\frac{\partial}{\partial \theta_i^I} = \theta_l^I; \quad \frac{\partial}{\partial \theta_j^I} = -\theta_l^I; \quad \frac{\partial}{\partial \theta_l^I} = \theta_i^I - \theta_j^I$$

$$\frac{\partial}{\partial M_i} = \frac{2}{|\mathcal{F}_u|^\alpha} \sum_{u' \in \mathcal{F}_u, i'} \sigma\big(\langle W_u, W_{u'}\rangle\big) \cdot M_{i'}$$

$$\frac{\partial}{\partial M_j} = -\frac{2}{|\mathcal{F}_u|^\alpha} \sum_{u' \in \mathcal{F}_u, i'} \sigma\big(\langle W_u, W_{u'}\rangle\big) \cdot M_{i'}$$

$$\frac{\partial}{\partial M_{i'}} = \frac{2}{|\mathcal{F}_u|^\alpha} \sigma\big(\langle W_u, W_{u'}\rangle\big) \cdot (M_i - M_j)$$

$$\frac{\partial}{\partial W_{u'}} = \frac{2}{|\mathcal{F}_u|^\alpha} \sigma'\big(\langle W_u, W_{u'}\rangle\big) \cdot \langle M_i - M_j, M_{i'}\rangle \cdot W_u$$

$$\frac{\partial}{\partial W_u} = \frac{2}{|\mathcal{F}_u|^\alpha} \sum_{u' \in \mathcal{F}_u, i'} \sigma'\big(\langle W_u, W_{u'}\rangle\big) \cdot \langle M_i - M_j, M_{i'}\rangle \cdot W_{u'}$$

(8)

where $\sigma'(z)$ is the derivative of the sigmoid function, i.e., $\sigma(z) \cdot \sigma(-z)$.

References

[Bell *et al.*, 2007] Robert M. Bell, Yehuda Koren, and Chris Volinsky. The bellkor solution to the netflix prize, 2007.

[Bennett and Lanning, 2007] James Bennett and Stan Lanning. The netflix prize. In *KDDCup*, 2007.

[Breese *et al.*, 1998] John S Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52, 1998.

[Chaney *et al.*, 2015] Allison JB Chaney, David M Blei, and Tina Eliassi-Rad. A probabilistic model for using social networks in personalized item recommendation. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 43–50, 2015.

[Guo *et al.*, 2015] Guibing Guo, Jie Zhang, and Neil Yorke-Smith. Trustsvd: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 123–129, 2015.

[He *et al.*, 2016] Ruining He, Chen Fang, Zhaowen Wang, and Julian McAuley. Vista: a visually, socially, and temporally-aware model for artistic recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 309–316, 2016.

[Jamali and Ester, 2010] Mohsen Jamali and Martin Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 135–142, 2010.

[Ma *et al.*, 2008] Hao Ma, Haixuan Yang, Michael R Lyu, and Irwin King. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 931–940, 2008.

[Ma *et al.*, 2009] Hao Ma, Irwin King, and Michael R Lyu. Learning to recommend with social trust ensemble. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 203–210, 2009.

[Ma *et al.*, 2011] Hao Ma, Dengyong Zhou, Chao Liu, Michael R Lyu, and Irwin King. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 287–296, 2011.

[Miyahara and Pazzani, 2000] Koji Miyahara and Michael J Pazzani. Collaborative filtering with the simple bayesian classifier. In *Pacific Rim International conference on artificial intelligence*, pages 679–689. 2000.

[Mobasher *et al.*, 2002] Bamshad Mobasher, Honghua Dai, Tao Luo, and Miki Nakagawa. Using sequential and non-sequential patterns in predictive web usage mining tasks. In *IEEE International Conference on Data Mining*, pages 669–672, 2002.

[Pan and Chen, 2013] Weike Pan and Li Chen. Gbpr: Group preference based bayesian personalized ranking for one-class collaborative filtering. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 2691–2697, 2013.

[Paterek, 2007] Arkadiusz Paterek. Improving regularized singular value decomposition for collaborative filtering. In *KDDCup*, 2007.

[Rendle *et al.*, 2009] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 452–461, 2009.

[Rendle *et al.*, 2010a] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 811–820, 2010.

[Rendle *et al.*, 2010b] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 811–820, 2010.

[Ricci *et al.*, 2011] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul Kantor. *Recommender systems handbook*. Springer US, 2011.

[Salakhutdinov *et al.*, 2007] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*, pages 791–798, 2007.

[Shani *et al.*, 2005] Guy Shani, David Heckerman, and Ronen I Brafman. An mdp-based recommender system. *Journal of Machine Learning Research*, 6(Sep):1265–1295, 2005.

[Tang *et al.*, 2012] Jiliang Tang, Huiji Gao, and Huan Liu. mtrust: discerning multi-faceted trust in a connected world. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 93–102, 2012.

[Tang *et al.*, 2013] Jiliang Tang, Huiji Gao, Xia Hu, and Huan Liu. Exploiting homophily effect for trust prediction. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 53–62, 2013.

[Zhao *et al.*, 2014] Tong Zhao, Julian McAuley, and Irwin King. Leveraging social connections to improve personalized ranking for collaborative filtering. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 261–270, 2014.

[Zimdars *et al.*, 2001] Andrew Zimdars, David Maxwell Chickering, and Christopher Meek. Using temporal data for making recommendations. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 580–588, 2001.