Collaborative Topic Regression with Social Network Factorization for Recommendation Systems

Ming Hsieh Department of Electrical Engineering

USC Viterbi School of Engineering

Sanjay Purushotham, Yan Liu, C.-C. Jay Kuo

Introduction

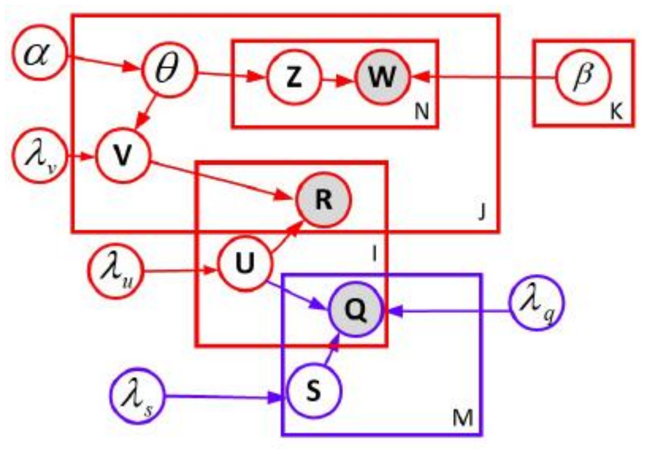
- Motivation: Social Networks have become important platform for content, opinion sharing and provide rich information to study social circle's influence on user's decision process
- Problem: How can social networks help improve the recommendation systems? Does information leak occur in social recommendation systems?
- Solution: We propose a novel hierarchical Bayesian model which jointly incorporates content topic modeling and matrix factorization of social networks. Empirical experiments on two large-scale datasets show that our algorithm outperforms the state-of-the art approaches. Our results reveal interesting insight that the social circles have more influence on people's decisions about the usefulness of information (e.g., bookmarking preference on Delicious) than personal taste (e.g.,music preference on Lastfm)

Related Work

- Collaborative Filtering (CF) approaches predict user preferences based on collective rating records of similar users or items. However, CF-based methods suffer from sparsity problem and imbalance of rating data, especially for new and infrequent users
- Latent factor methods such as Probabilistic Matrix Factorization (PMF) incorporate user interests into the CF-model. However, users are assumed to be *i.i.d* and additional information is ignored
- Collaborative Topic Regression (CTR)[1] incorporates content information via LDA into collaborative filtering framework. CTR represents users with topic interests and assumes that items are generated by a topic model
- Social recommendation (SoRec)[2], based on matrix factorization, uses social network information and user ratings to recommend items
- Current approaches cannot predict ratings for new/unseen items and new/infrequent users of social network

Our Approach

Our model is generalized hierarchical Bayesian model which jointly learns the user, item and social factor latent spaces. We use LDA to capture item's content information in latent topic space, and we use matrix factorization to derive latent feature space of user from his social network graph.



Red: CTR, Blue: Social Matrix Factorization

Conditional distribution over social network relationships:

$$P(Q|U,S,\sigma_Q^2) = \prod_{i=1}^{m}\prod_{k=1}^{m} N(qij|g(U_i^T S_k),\sigma_Q^2)^{I_{ij}^Q}$$

Through Bayesian inference, combining LDA with social matrix factorization, we get:

$$p(U,V,S|Q,R,\sigma_Q^2,\sigma_R^2,\sigma_U^2,\sigma_V^2,\sigma_S^2) \propto p(R|U,V,\sigma_R^2)\, p(Q|U,S,\sigma_Q^2) \times p(U|\sigma_U^2)p(V|\sigma_V^2)\, p(S|\sigma_S^2)$$

Learning Parameters

Expectation-Maximization Algorithm

Maximization of posterior is equivalent to maximizing the complete log-likelihood of $U,V,S,\theta_{1:J},R\ and\ Q$ given $\lambda_U,\lambda_V,\lambda_S,\lambda_Q\ and\ \beta$

$$L = -\frac{\lambda_U}{2}\sum_i u_i^T u_i - \frac{\lambda_V}{2}\sum_j (v_j-\theta_j)^T(vj-\theta_j) + \sum_j\sum_n log\left(\sum_k \theta_{jk}\beta_{k,\omega_{jn}}\right) - \sum_{i,j}\frac{c_{ij}}{2}(r_{ij}-u_i^T v_j)^2 - \frac{\lambda_Q}{2}\sum_{i,m}\frac{d_{im}}{2}(q_{im}-u_i^T s_m)^2 - \frac{\lambda_S}{2}\sum_k s_k^T s_k$$

Update Equations

$$u_i \leftarrow (VCiV^T+\lambda_Q SD_i S^T+\lambda_U I_K)^{-1}(\mathrm{VCiR_i}+\lambda_Q SD_i Q_i)$$
$$v_j \leftarrow (UC_jU^T+\lambda_V I_K)^{-1}(UCjR_j+\lambda_V\theta_j)$$
$$s_m \leftarrow (\lambda_Q UD_m U^T+\lambda_S I_K)^{-1}(\lambda_Q UD_m Q_m)$$

Optimize this function by gradient ascent approach by iteratively optimizing the collaborative filtering and social network variables u_i, v_j, s_m and topic proportions θ_j

Given U and V, we can learn topic proportions θ_j. We cannot optimize θ_j analytically, so we use projection gradient approaches

Experiments

Dataset Description

Dataset	Lastfm	Delicious
Users	1892	1867
Items	17632	69226
tags	11946	53388
User-user links	25434	15328
User-tags-items	186479	437593
User-items	92834	104799

Both the datasets are highly sparse (99.7% and 99.91% respectively)

Evaluation

In-matrix prediction: Item rated by at-least one user

$$\varepsilon[rij|D] \approx \varepsilon[u_i|D]^T(\varepsilon[\theta_j|D]+\varepsilon[\epsilon_j|D])$$
$$r_{ij}^* \approx (u_i^*)^T v_j^*$$

Out-matrix prediction: New item (no ratings)

$$\varepsilon[rij|D] \approx \varepsilon[u_i|D]^T(\varepsilon[\theta_j|D])$$
$$r_{ij}^* \approx (u_i^*)^T\theta_j^*$$

We consider 'recall' as evaluation metric

$$recall@M = \frac{number\ of\ items\ user\ likes\ in\ top\ M}{total\ number\ of\ items\ the\ user\ likes}$$

Results

Comparisons

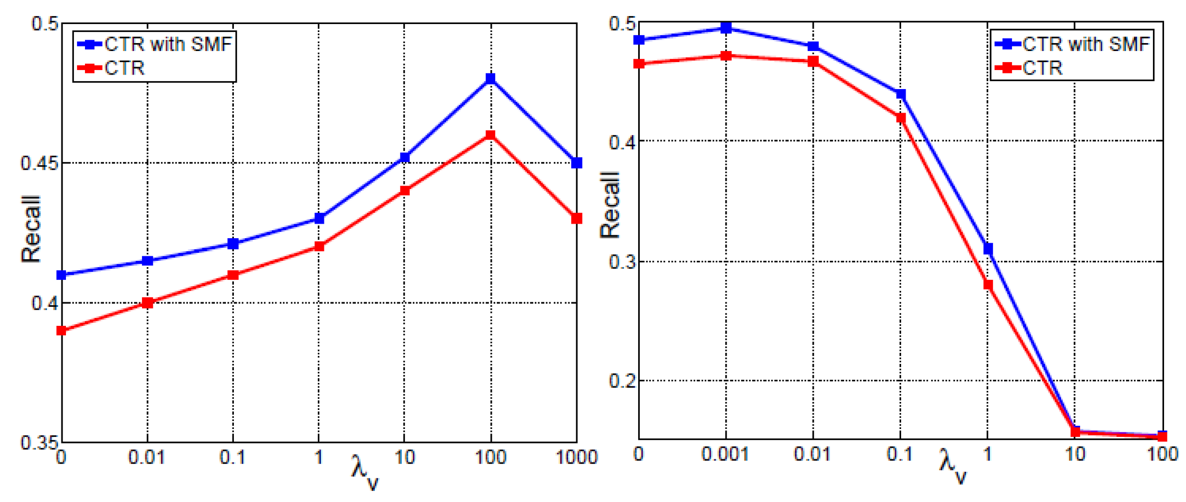


Fig 1: In-matrix recall comparisons (Left: Lastfm, Right: Delicious)

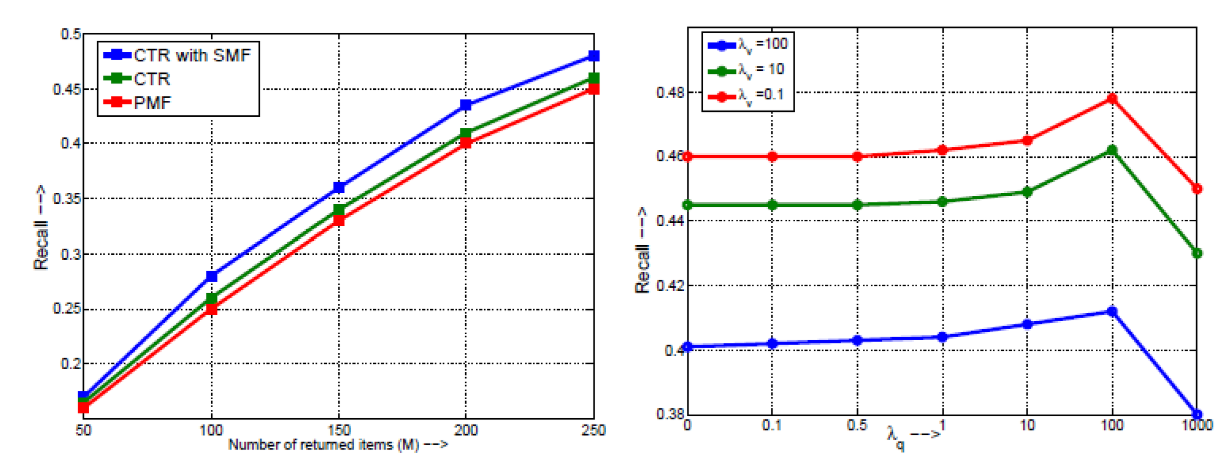


Fig 2: In-matrix prediction recall. Left: Comparing our model with state-of-the-art approaches. Right: Impact of social network parameter λ_q by varying content parameter λ_v

Impact of content and social network parameters

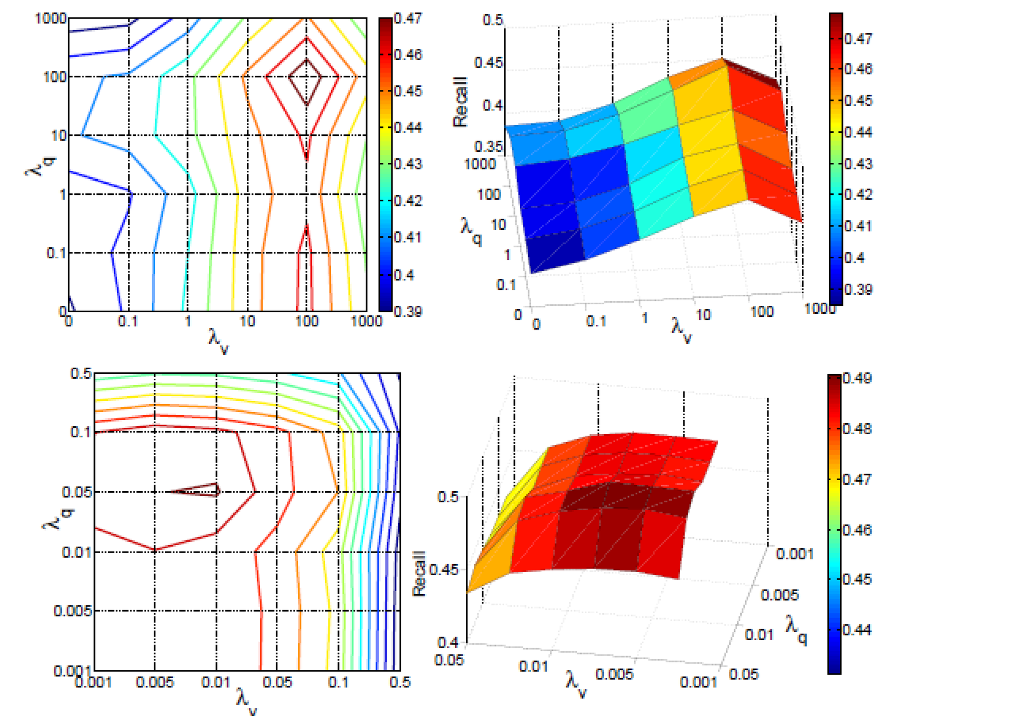


Fig 3: In-matrix prediction recall by varying content parameter λv and social network parameter λv at M=250. Top: Lastfm, Bottom: Delicious dataset

Complexity Analysis

Model	Avg. time
CTR	9.46 hr
Our model	9.53 hr

K	Time	Avg. recall
50	14.3 min	0.472
200	10.5 hr	0.510

Social Information Leak

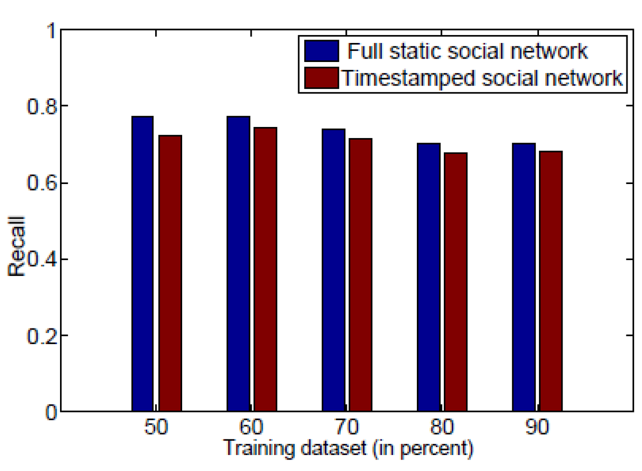


Fig 4: Recall of proposed model by varying social network structure. Dataset used: Delicious

Using final 'static' social network in recommendation systems could lead to potential information leak. Our model could be making better predictions using future social network information. Fig. 4 confirms information leak. We observe that smaller training dataset (sparser social graph) results in larger information leak.

Conclusions

Contributions

- Demonstrating the usefulness of social network information to improve recommendation systems
- Providing effective tradeoff techniques to improve recommendation accuracy when both social and content information are both known
- Starting discussion on new problem (social information leak) in recommendation systems

Future work

- Parallel implementation of our algorithms and capturing social network dynamics into our model

[1] Wang, Chong and Blei, David M. Collaborative topic modeling for recommending scientific articles. KDD, 2011
[2] Ma, Hao, Yang, Haixuan, Lyu, Michael R., and King, Irwin. Sorec: social recommendation using probabilistic matrix factorization. CIKM, 2008
[3] Hu, Yifan, Koren, Y., and Volinsky, C. Collaborative filtering for implicit feedback datasets. ICDM, 2008

Contact information: Sanjay Purushotham, Email: spurusho@usc.edu

Ming Hsieh Institute
Ming Hsieh Department of Electrical Engineering