Learning to Recommend with Social Trust Ensemble

2021 Spring
Special Lectures on Databases (Recsys)

2021-06-02
한민희

Social Trust Ensemble?

“the formulation of the social trust restrictions on the recommender systems.”

“we always turn to friends we trust for book, music, or restaurant recommendations, and our favors can easily be affected by the friends we trust.”

이 paper의 주요 목표

- RSTE(Recommendation with Social Trust Ensemble) 제안
- user-item matrix 외에 user 간 trust list를 함께 써서 추천 정확도를 제고

“Based on the intuition that every user’s decisions on the Web should include both the user’s characteristics and the user’s trusted friends’ recommendations, we propose a novel, effective and efficient probabilistic matrix factorization framework for the recommender systems.”

기존 CF Recommender System들의 문제

1) Data Sparsity

   “density of the available ratings in commercial recommender systems is often less than 1%”

2) Social connection / trust relation의 고려 부족

   “Therefore, traditional recommender systems, which purely mine the user-item rating matrix for recommendations, do not provide realistic output.”

Assumptions (intuitions)

1) “Users have their own characteristics, and they have different tastes on different items, such as movies, books, music, articles, food, etc.”
2) “Users can be easily influenced by the friends they trust, and prefer their friends’ recommendations.”
3) “One user’s final decision is the balance between his/her own taste and his/her trusted friends’ favors.”

-> “we infer and formulate the recommendation problem purely based on their trusted friends’ favors.”

Problem Description

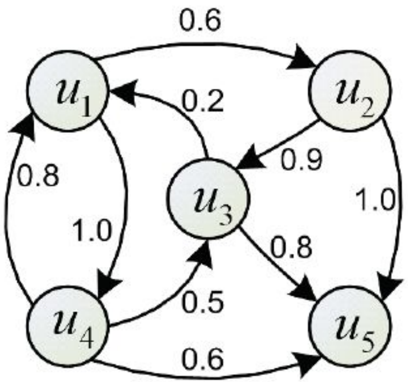
오른쪽과 같은 정보가 주어졌을 때

특정 유저 u의,

특정 아이템 v에 대한

rating R을 추정하는 것

(단, 실제로는 Social Trust Graph가 아니라 trust list로 주어짐)



(a) Social Trust Graph

	v_1	v_2	v_3	v_4	v_5	v_6
u_1		5	2		3	
u_2	4			3		4
u_3			2			2
u_4	5			3		
u_5		5	5			3

(b) User-Item Rating Matrix

Figure 1: Example for Trust based Recommendation

Recommendations by Trusted Friends

$$\widehat{R}_{ik} = \sum_{j \in \mathcal{T}(i)} R_{jk} S_{ij}.$$

$$\begin{pmatrix} \widehat{R}_{i1} \\ \widehat{R}_{i2} \\ \dots \\ \widehat{R}_{in} \end{pmatrix} = \begin{pmatrix} R_{11} & R_{21} & \dots & R_{m1} \\ R_{12} & R_{22} & \dots & R_{m2} \\ \dots & \dots & \dots & \dots \\ R_{1n} & R_{2n} & \dots & R_{mn} \end{pmatrix} \begin{pmatrix} S_{i1} \\ S_{i2} \\ \dots \\ S_{im} \end{pmatrix}$$

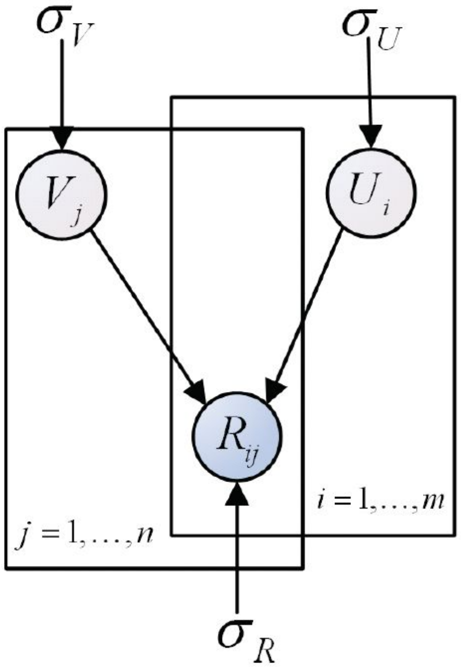
Features Learning

학습 대상:

*l*차원 feature representation 2D matrix

U(user latent matrix) V(item latent matrix)

$$
\begin{aligned}
p(U,V|R,\sigma_R^2,\sigma_U^2,\sigma_V^2) &\propto p(R|U,V,\sigma_R^2)p(U|\sigma_U^2)p(V|\sigma_V^2) \\
&= \prod_{i=1}^{m}\prod_{j=1}^{n}\left[\mathcal{N}\left(R_{ij}|g(U_i^T V_j),\sigma_R^2\right)\right]^{I_{ij}^R} \\
&\times \prod_{i=1}^{m}\mathcal{N}(U_i|0,\sigma_U^2\mathbf{I}) \times \prod_{j=1}^{n}\mathcal{N}(V_j|0,\sigma_V^2\mathbf{I}).
\end{aligned}
$$



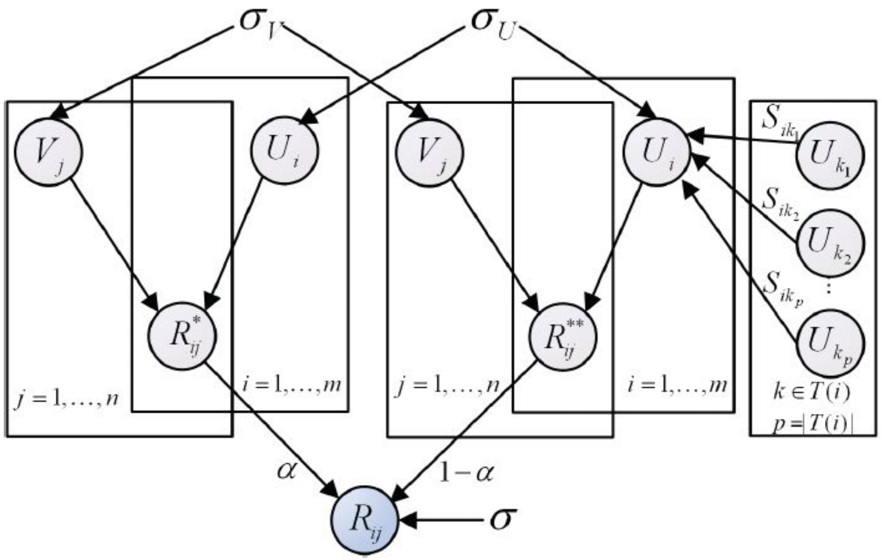
(a) Factorization of User-Item Matrix

The fusion of two information

- f(α(user-item rating matrix) , (1-α)(trust graph))
- α가 클수록 user의 취향을 더 많이 반영
- α가 작을수록 trusted friends의 취향을 더 많이 반영
- 실험에서는, α=0.4가 최고의 결과를 보였음
    - 두 데이터를 fuse하는 것이 좋은 성능을 보임
    - 단, trusted friends의 취향을 더 반영하는 것이 좋은 결과를 보인다
    - 이것은 user-item matrix의 sparsity 때문

Ratings(R) + Social Trust Matrix(S)

$$
\begin{aligned}
& p(U, V | R, S, \sigma^2, \sigma_U^2, \sigma_V^2) \\
&= \prod_{i=1}^{m} \prod_{j=1}^{n} \left[\mathcal{N}\left(R_{ij} | g(\alpha U_i^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(i)} S_{ik} U_k^T V_j), \sigma^2 \right) \right]^{I_{ij}^R} \\
&\times \prod_{i=1}^{m} \mathcal{N}(U_i | 0, \sigma_U^2 \mathbf{I}) \times \prod_{j=1}^{n} \mathcal{N}(V_j | 0, \sigma_V^2 \mathbf{I}). \qquad (11)
\end{aligned}
$$



(c) Recommendations with Social Trust Ensemble

Scalability

- 학습을 위한 gradient 계산의 complexity는
- Ratings matrix에서 관찰된 non-zero 값의 수(실제 rating 수)에 비례
- 따라서 큰 dataset에도 scalable하게 적용 가능

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial U_i} = {} & \alpha \sum_{j=1}^{n} I_{ij}^{R} g'(\alpha U_i^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(i)} S_{ik} U_k^T V_j) V_j \\
& \times (g(\alpha U_i^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(i)} S_{ik} U_k^T V_j) - R_{ij}) \\
& + (1-\alpha) \sum_{p \in \mathcal{B}(i)} \sum_{j=1}^{n} I_{pj}^{R} g'(\alpha U_p^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(p)} S_{pk} U_k^T V_j) \\
& \times (g(\alpha U_p^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(p)} S_{pk} U_k^T V_j) - R_{pj}) S_{pi} V_j + \lambda_U U_i,
\end{aligned}
$$

$$O(\rho_R \overline{p}\, l + \rho_R \overline{p}\, \overline{k} l)$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial V_j} = {} & \sum_{i=1}^{m} I_{ij}^{R} g'(\alpha U_i^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(i)} S_{ik} U_k^T V_j) \\
& \times (g(\alpha U_i^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(i)} S_{ik} U_k^T V_j) - R_{ij}) \\
& \times (\alpha U_i + (1-\alpha) \sum_{k \in \mathcal{T}(i)} S_{ik} U_k^T) + \lambda_V V_j, \qquad (14)
\end{aligned}
$$

$$O(\rho_R l + \rho_R \overline{k} l)$$

$$O(\rho_R \overline{p}\, l + \rho_R \overline{p}\, \overline{k} l)$$

Experiments

- 비교 대상

1) PMF: Probabilistic Matrix Factorization, user-item matrix만을 활용
2) Trust: 오직 trusted friends의 rating만을 활용
3) SoRec: 동일한 user latent space에 user-item rating matrix와 social trust network를 반영

Experiments: Data

- 실험 데이터: Epinions.com의 2009년 1월 데이터
- user-item ratings + user reputation
- users: 51,670, item: 83,509, ratings: 631,064, trust statements: 511,799
- user-item matrix density가 0.015% 이하로, Movielens(4.25%), Eachmovie(2.29%)보다 더 sparse한 데이터

Table 1: Statistics of User-Item Rating Matrix of Epinions

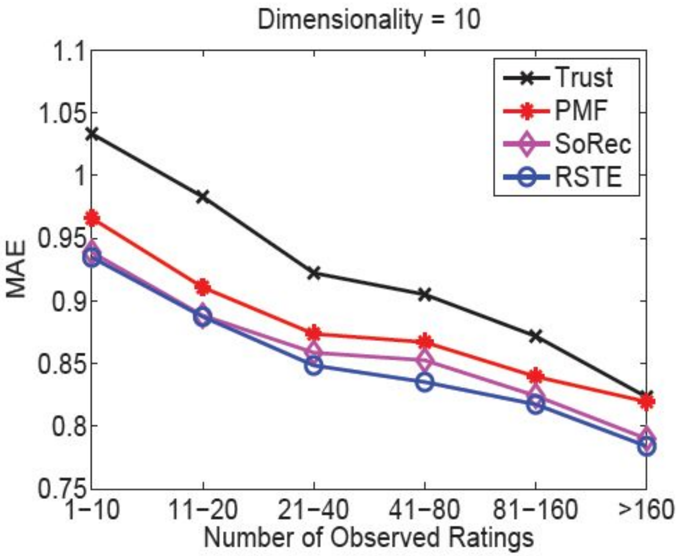
Statistics	User	Item
Max. Num. of Ratings	1960	7082
Avg. Num. of Ratings	12.21	7.56

Table 2: Statistics of Social Trust Network of Epinions

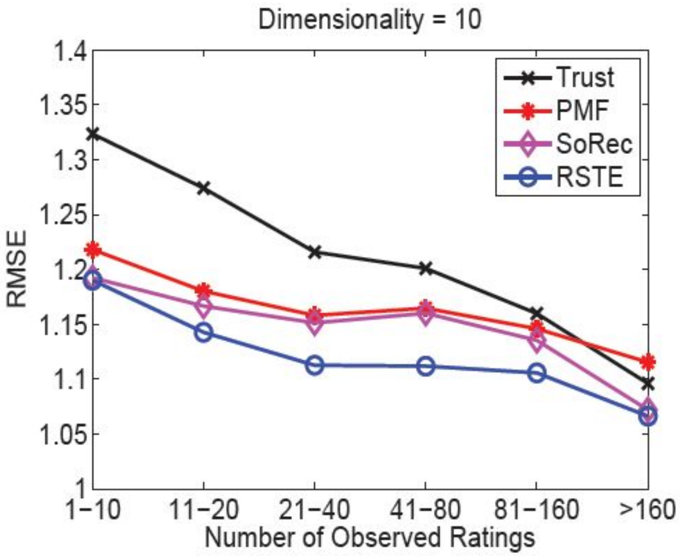
Statistics	Trust per User	Be Trusted per User
Max. Num.	1763	2443
Avg. Num.	9.91	9.91

Experiments

MAE: Mean Absolute Error, RMSE: Root Mean Square Error



(b) MAE Comparison on Different User Rating Scales (90% as Training Data)



(c) RMSE Comparison on Different User Rating Scales (90% as Training Data)

Experiments

- RSTE > SoRec > PMF > Trust
- 참고: 해당 데이터에 대한 2019년 SOTA 수치 MAE:0.7781, RMSE:1.0268

Table 3: Performance Comparisons (A Smaller MAE or RMSE Value Means a Better Performance)

Training Data	Metrics	Dimensionality = 5				Dimensionality = 10			
		Trust	PMF	SoRec	RSTE	Trust	PMF	SoRec	RSTE
90%	MAE	0.9054	0.8676	0.8442	**0.8377**	0.9039	0.8651	0.8404	**0.8367**
	RMSE	1.1959	1.1575	1.1333	**1.1109**	1.1917	1.1544	1.1293	**1.1094**
80%	MAE	0.9221	0.8951	0.8638	**0.8594**	0.9215	0.8886	0.8580	**0.8537**
	RMSE	1.2140	1.1826	1.1530	**1.1346**	1.2132	1.1760	1.1492	**1.1256**

Questions

“Epinions is thus an ideal source for experiments on social trust recommendation.”

- 현실적으로 “trust list”가 주어지는 경우가 얼마나 될까?
- 예를 들어 SNS와 연동한다 하여도, ‘특정 품목에 대해 누군가를 신뢰하는 정도’를 알기는 힘들 것
- ‘trust list’가 존재하지 않는 경우에 이것을 어떤 식으로 응용할 수 있을 것인가?