Dual Graph Attention Networks for Deep Latent Representation of Multifaceted Social Effects in Recommender Systems

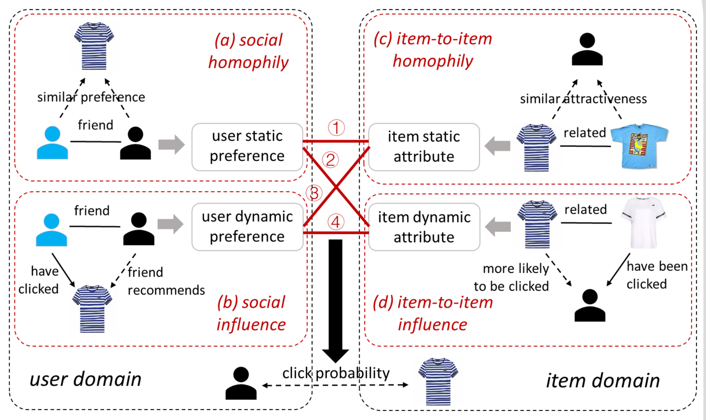
Qitian Wu[1], Hengrui Zhang[1], Xiaofeng Gao[1], Peng He[2], Paul Weng[1], Han Gao[2], Guihai Chen[1]

[1]Shanghai Jiao Tong University, [2]Tencent Inc.

Introduction

1. Motivation

- A user's preference isn't influenced by his friends equally.
- Social effects may be dynamic and dependent of specific contexts.
- Previous methods lack interpretability



two-fold social effects in both user domain and item domain

- Social homophily
- Social influence

2. Contribution

- **General Aspects:** We distinguish the social homophily and social influence notions in view of static and dynamic effects
- **Novel Methodologies:** We are one of the first to use GAT for social recommendation task, and the dual GATs can collaboratively model four social effects in both user domain and item domains.
- **Multifaceted Experiments:** We deploy DANSER on one benchmark dataset and a commercial dataset. Experiment results verify the superiority of DANSER over state-of-the-art techniques.

Training strategies

Mini-Batch Training

For each iteration, consider B user-item pairs .

For each pair (u,i), The neighbor nodes $F_U(u), F_I(i)$ in two networks as well as the rated history $R_I(v), v\in \Gamma_U(u)$, and $R_U(j), j\in\Gamma_I(i)$ will be together input into the model to compute the gradient

Local-Graph Aware Regularization

Ll regularization $\mathcal{L}_2=\sum_u(||p_u||+||x_u||)+\sum_i(||q_i||+||y_i||)$.

constrain the embedding parameters of users, items as well as their neighbor nodes in social network and item implicit network

$$\mathcal{L}_2=\frac{1}{2}\sum_{(u,i)}[||\mathbf{p}_u||+||\mathbf{x}_u||]+|\sum_{v\in\Gamma_U(u)}\frac{1}{|F_U(v)|}(||\mathbf{p}_v||+||\mathbf{x}_v||)$$

$$+||\mathbf{q}_i||+||\mathbf{y}_i||+\sum_{j\in\Gamma_I(j)}\frac{1}{|F_I(j)|}(||\mathbf{q}_j||+||\mathbf{y}_j||)$$

final loss function

$$\mathcal{L}=\mathcal{L}_1+\lambda\mathcal{L}_2$$

Policy Gradient

the active probability given by l-th policy network $p_l(\gamma|\mathbf{p}_u,\mathbf{q}_i)$

draw $\gamma\sim Multi(p_l(\gamma|\mathbf{p}_U,\mathbf{q}_i))$, and feed $\mathbf{s}_\gamma$ into output layer

define the reward as subtraction of loss generated by selected interacted feature $\mathcal{R}(p_u,q_i,\gamma)=-\mathcal{L}(\mathbf{s}_\gamma)$

$\mathcal{L}(\mathbf{s}_\gamma)$ denotes the loss when we input $\mathbf{s}_\gamma$ into the output layer

maximize the expected reward

$$\mathbb{E}_{\gamma\sim p(\gamma|p_u,q_i)}(\mathcal{R}(p_u,q_i,\gamma))$$

calculate gradient

$$\nabla_\theta\mathbb{E}_{\gamma\sim p_\theta(\gamma|\mathbf{p}_u,\mathbf{q}_i)}(\mathcal{R}(\mathbf{p}_u,\mathbf{q}_i,\gamma))$$

$$\simeq\frac{1}{4}\sum_\gamma\nabla_\theta\log p_\theta(\gamma|\mathbf{p}_u,\mathbf{q}_i)\mathcal{R}(\mathbf{p}_u,\mathbf{q}_i,\gamma)$$

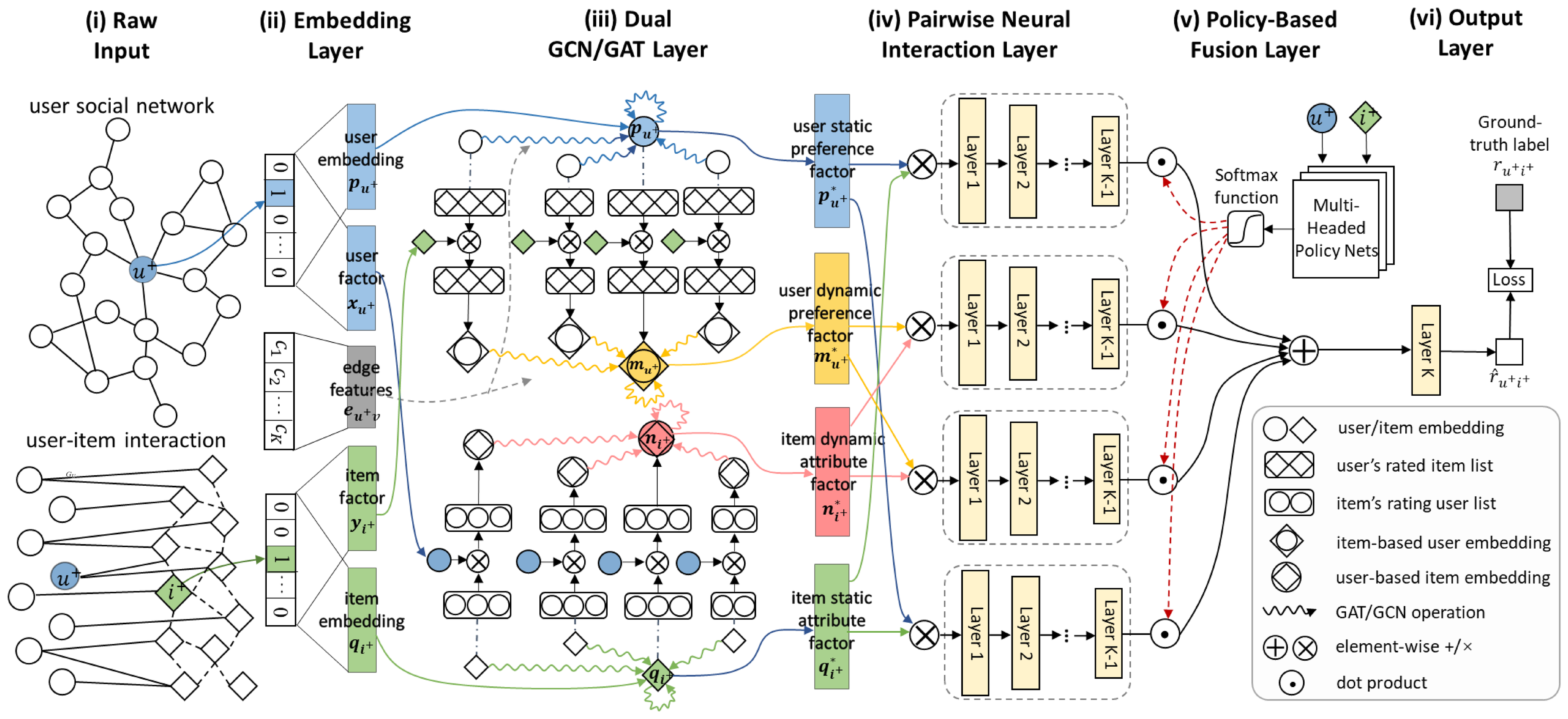
DANSER model

Raw Input and Item Implicit Network

- user-item interaction matrix R
- user social network G_U
- similarity coefficient s_{ij} of item i and item j : number of users who clicked both items
- item i is related to item j if $s_{ij}>\tau$
- define item implicit network as graph $G_I=(V_I,E_I)$

Embedding Layer

- user embedding $\mathbf{P}=\{\mathbf{p}_u\}_{D\times M}$
- item embedding $\mathbf{Q}=\{\mathbf{q}_i\}_{D\times N}$
- user factor $\mathcal{X}_i=\{\mathbf{x}_v|v\in R_U(i)\}$
- item factor $\mathcal{Y}_u=\{\mathbf{y}_j|j\in R_I(u)\}$
- edge features $e_{u,v}$



Dual GCN/GAT Layer

GAT to capture social influence (user domain)

$$\mathcal{Y}_u^{i+}=\{\mathbf{y}_j\otimes\mathbf{y}_{i^+}|j\in R_I(u)\}$$

$$m_{ud}^{i+}=\max_{j\in R_I(u)}\{y_{jd}\cdot y_{i^+d}\},\forall d=1,...,D$$

GAT to capture social homophily (user domain)

$$\mathbf{P}^*=\sigma(\mathbf{A}_P(G_U)\mathbf{P}\mathbf{W}_P^T+\mathbf{b}_p)$$

$$\alpha_{uv}^P=\frac{attn_U(\mathbf{W}_p\mathbf{p}_u,\mathbf{W}_p\mathbf{p}_v,\mathbf{W}_E\mathbf{e}_{uv})}{\sum_{w\in\Gamma_U(u)}attn_U(\mathbf{W}_p\mathbf{p}_u,\mathbf{W}_p\mathbf{p}_w,\mathbf{W}_E\mathbf{e}_{uv})}$$

$$M_{i^+}^*=\sigma(\mathbf{A}_M(G_U)\mathbf{M}\mathbf{W}_M^T)+\mathbf{b}_\mathbf{M},\mathbf{A}_M(G_U)=\{\alpha_{uv\hat{M},i^+}\}_{M\times M}\}$$

$$\alpha_{uv,i^+}^M=\frac{attn_U(\mathbf{W}_M\mathbf{m}_u^{i+},\mathbf{W}_M\mathbf{m}_v^{i+},\mathbf{W}_M\mathbf{E}\mathbf{e}_{uv})}{\sum_{w\in\Gamma_U(u)}attn_U(\mathbf{W}_M\mathbf{m}_u^{i+},\mathbf{W}_M\mathbf{m}_w^{i+},\mathbf{W}_M\mathbf{E}\mathbf{e}_{uv})}$$

Formulas of GAT in item domain are similar

Pairwise Neural Interaction Layer

k-layers fully-connected neural network (tanh activation)

Policy-Based Fusion Layer

$$e_\gamma=\mathbf{W}_F^2\tanh(\mathbf{W}_F^1(\mathbf{p}_u||\mathbf{q}_i)+\mathbf{b}_F^1)+\mathbf{b}_F^2.\quad p(\gamma|\mathbf{p}_u,\mathbf{q}_i)=\frac{\exp(e_\gamma)}{\sum_{a=1}^4\exp(e_a)}$$

$$\mathbf{s}=\mathbb{E}_{\gamma\sim p(\gamma|\mathbf{p}_u,\mathbf{q}_i)}(\mathbf{s}_\gamma)=\sum_{\gamma=1}^4(p(\gamma|\mathbf{p}_u,\mathbf{q}_i)\cdot\mathbf{s}_\gamma)$$

Output Layer

a fully-connected layer without activation function

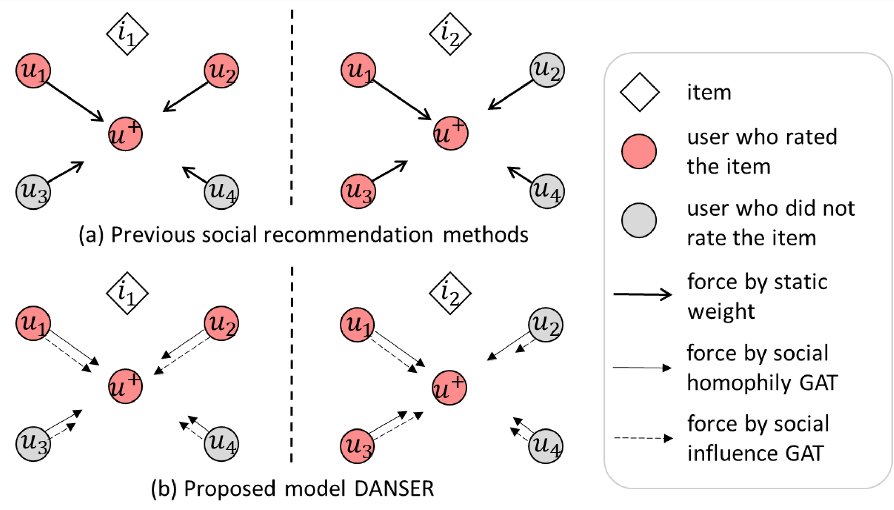
Loss Function

implicit feedback: $\mathcal{L}_1=-\sum_{(u,i)}r_{ui}\log\hat{r}_{ui}+(1-r_{ui})\log(1-\hat{r}_{ui})$

explicit feedback: $\mathcal{L}_1=\sum_{(u,i)}||\hat{r}_{ui}-r_{ui}||^2$

Justification of Dual GATS

Figure: Illustration of effectiveness of dual GATs and comparison with previous social recommendation methods



DANSER contains two different attention weights, where one captures user-specific static effect (for homophily) and the other can dynamically change under different contexts (i.e., for influence). The Global and local views of DANSER work collaboratively for selecting the 'right' neighbors that are similar to the targeted user under a specific context, which improves model expressiveness.

Algorithm: Training for DANSER

REQUIRE: R, observed user-item interaction. G_U, user social network. G_I, item implicit network.
REQUIRE: η, ζ, learning rates.
while *not converged* **do**
  for $i=1,\cdots,n_p$ **do**
    Sample B user-item pairs.;
    For each pair (u,i), uniformly sample F neighbors from $F_U(u)$ and $F_I(i)$, respectively;
    Draw $\gamma_l\sim Multi(p_{\theta l}^l|\mathbf{p}_u,\mathbf{q}_i)$, $l=1,\cdots,L$;
    $w\leftarrow\eta\nabla_w\sum_l L(\mathbf{s}_{\gamma_l})$;
  $\theta_l\leftarrow\zeta\nabla_{\theta_l}\mathbb{E}_{\gamma\sim p_{\theta_l}^l(\gamma|\mathbf{p}_u,\mathbf{q}_i)}(\mathcal{R}(\mathbf{p}_u,\mathbf{q}_i,\gamma)), l=1,\cdots,L;$

Experiments

Table1: Comparative results for Epinions and Wechat. For MAE, RMSE, the smaller value is better and vice versa for P@10, AUC.

	Epinions		WeChat	
	MAE	RMSE	P@10	AUC
SVD++ [15]	0.8321	1.0772	0.0653	0.7304
DELF [2]	0.8115	1.0561	0.0752	0.7818
TrustPro [37]	0.9130	1.1124	0.0561	0.6482
TrustMF [36]	0.8214	1.0715	0.0625	0.7005
TrustSVD [10]	0.8144	1.0492	0.0664	0.7325
NSCR [31]	0.8044	1.0425	0.0736	0.7727
SREPS [16]	0.8014	1.0393	0.0725	0.7745
DANSER	**0.7781**	**1.0268**	**0.0823**	**0.8165**
Impv.[1]	2.87%	1.25%	9.33%	4.48%

[1] The improvement compares DANSER with the best competitor (underlined).

Table2: Ablation study of components in proposed method

	Epinions		WeChat	
	MAE	RMSE	P@10	AUC
DualEMB	.7920(1.7%)[1]	1.0363(0.7%)	.0794(3.6%)	.7992(2.2%)
DualGCN	.7840(0.7%)	1.0335(0.4%)	.0814(1.1%)	.8102(0.8%)
userGAT	.7858(0.9%)	1.0364(0.7%)	.0813(1.2%)	.8136(0.4%)
itemGAT	.7919(1.7%)	1.0335(0.4%)	.0813 (1.2%)	.8138(0.3%)
DANSER-w	.8191(4.9%)	1.0659(3.4%)	.0820(0.4%)	.8151(0.2%)
DANSER-m	.8211(5.2%)	1.0681(3.6%)	.0815(1.0%)	.8144(0.3%)
DANSER-a	.8232(5.4%)	1.0710(3.9%)	.0814 (1.1%)	.8140(0.3%)
DANSER-c	.8091(3.8%)	1.0659(3.4%)	.0809(1.7%)	.8118(0.6%)
DANSER	**0.7787**	**1.0292**	**0.0823**	**0.8165**

[1] The ratio indicates the impv. comparing DANSER with corresponding variant.

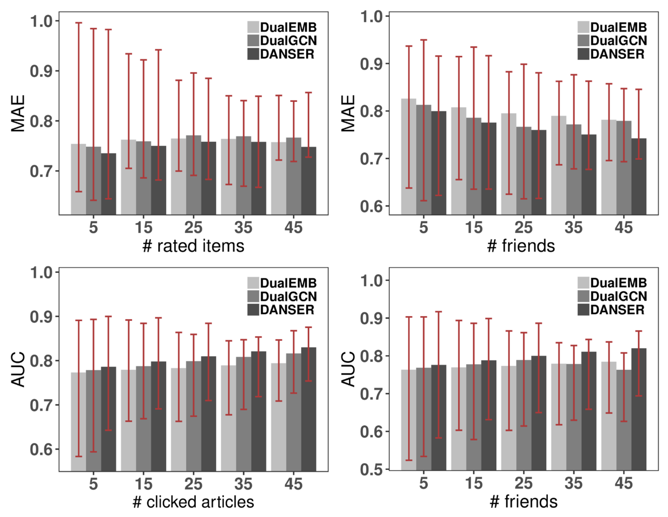


Figure1: MAE/AUC of DANSER, DualGCN, DualEMB on Epinions/Wechat w.r.t users with different number of clicked items and friends.

Figure2: Visualization of the four GAT's weights for a case study. There are four user item pairs: (a) (u,i), (b) (u,j), (c) (v,i), (d) (v,j).

Row (i) and row (ii) are for social homophily and social influence GAT's weights, respectively, while row (iii) and row (vi) are for item-to-item homophily and item-to-item influence, GAT's weights. Respectively, Grey nodes are negative samples while orange nodes are positive ones.

Available sources

Paper: https://arxiv.org/abs/1903.10433 **Codes:** https://github.com/echo740/DANSER

Qitian Wu: echo740@sjtu.edu.cn