

Heterogeneous Embedding Propagation for Large-scale E-Commerce User Alignment

Vincent W. Zheng, Mo Sha, Yuchen Li, Hongxia Yang,
Yuan Fang, Zhenjie Zhang, Kian-Lee Tan, Kevin Chen-Chuan Chang



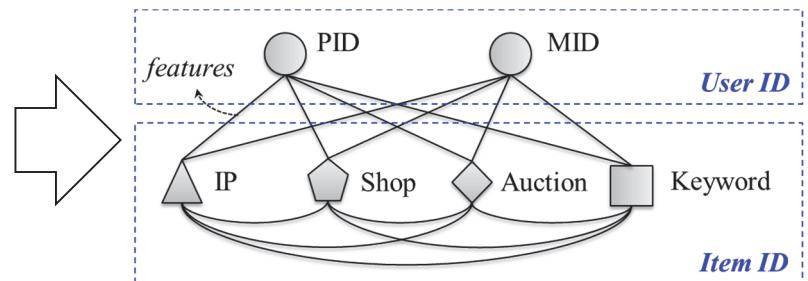
Problem: Matching users across devices (PC vs. Mobile)

E-commerce user activity log

Time	User	IP	Keywords	Auction	Shop
04/05/2017 16:21	PID1	IP2	toys	-	Shop3
04/05/2017 22:12	MID3	IP2	lego	Auction1	Shop2
...

To determine if PID1 (a PC identifier) is the same user as MID3 (a mobile device identifier).

Heterogeneous interaction graph



Approach: Heterogeneous Embedding Propagation (HEP)

Challenges

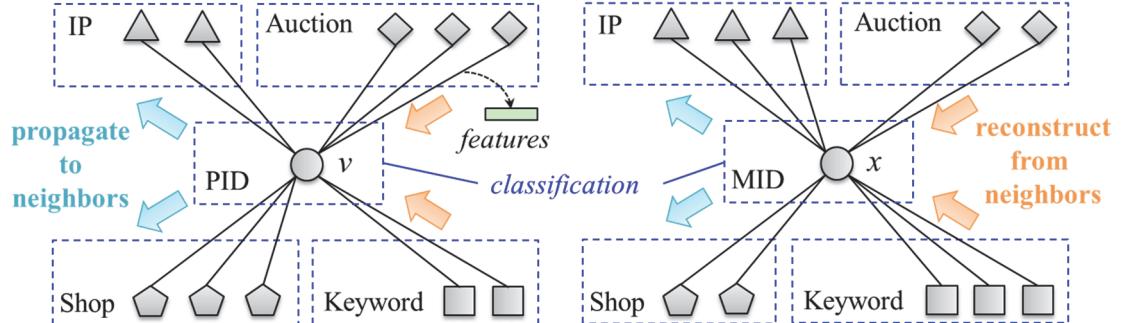
Heterogeneity

- Different node types / semantics

Edge features

- Historical interaction patterns

Overall framework



Classification loss

Semi-supervised learning: some PID-MID pairs are known to be positive or negative.

$$P(y_i|v_i, u_i) = \sigma(y_i \cdot \mathbf{h}_{v_i}^T W \mathbf{h}_{u_i})$$

$$L_1 = -\frac{1}{m} \sum_{i=1}^m \log P(y_i|v_i, u_i)$$

Reconstruction loss

Reconstructed node embeddings from neighbors need to be similar to the target embeddings.

$$\tilde{\mathbf{g}}_v^{(c)} = \sum_{u \in N_v^{(c)}} \frac{s_{v,u}}{\sum_{u \in N_v^{(c)}} s_{v,u}} \mathbf{h}_u \quad \tilde{\mathbf{g}}_v = \text{CONCAT}(\tilde{\mathbf{g}}_v^{(c_1)}, \dots, \tilde{\mathbf{g}}_v^{(c_{n_1})})$$

$$\tilde{\mathbf{h}}_v = \sigma(W'_{\phi(v)} \tilde{\mathbf{g}}_v + \mathbf{b}''_{\phi(v)}) \quad \ell(v, u) = [\gamma + \pi(\tilde{\mathbf{h}}_v, \mathbf{h}_v) - \pi(\tilde{\mathbf{h}}_v, \mathbf{h}_u)]_+$$

$$L_2 = \frac{1}{|V|} \sum_{v \in V} \sum_{u \sim P_n(u)} \ell(v, u)$$

Experimental Results

Datasets:

- Taobao's one-week user activity log in a city
- TB-Top: top 10% active users
- TB-Top: random 10% users

Baselines:

- FEM: feature engineering
- LDA: latent Dirichlet allocation
- GRU: gated recurrent unit
- Metapath2vec: meta-path based embedding
- EP: embedding propagation
- HEP-: HEP without edge features

	TB-Top			TB-Rnd		
	Precision	Recall	F1	Precision	Recall	F1
FEM	60.3	3.4	6.4	68.7	1.9	3.7
LDA	70.4	10.6	18.5	68.3	6.1	11.3
GRU	51.8	26.2	34.8	52.6	22.1	31.2
Metapath2vec	1.7	62.9	3.4	2.3	58.7	4.4
EP	34.3	6.7	11.2	35.0	6.1	10.4
HEP-	32.9	31.3	32.1	34.7	25.0	29.0
HEP	36.5	39.2	37.8	44.5	40.5	42.4