A Heterogeneous Information Network based Cross Domain Insurance Recommendation System for Cold Start Users

Ye Bi Liqiang Song Mengqiu Yao
Zhenyu Wu Jianming Wang Jing Xiao
Ping An Technology Shenzhen Co., Ltd

Speaker: **Ye Bi**

About the Speaker

- Senior Algorithm Engineer in Ping An Technology Shenzhen Co., Ltd
- **Major Research Direction:** Deep Learning, Recommendation System, Knowledge Graph

1 Introduction

Background

Online insurance is becoming more and more popular, thought it is in its growth bottleneck. A special recommendation system for online insurance domain is in demand.

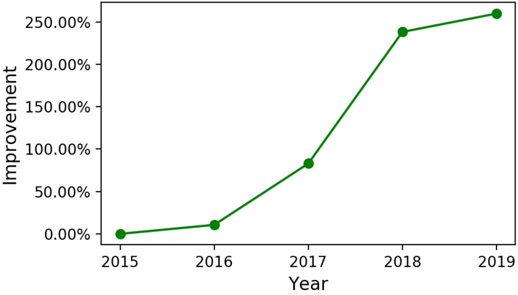


Figure 1: The Number of Insured Improvements w.r.t. 2015 on Jinguanjia App

Background

Users' behaviors in nonfinancial domain have influence on their behaviors in insurance domain. Users who have shopping experiences in nonfinancial domain are more willing to trust Jinguanjia, and more likely to buy insurance products.

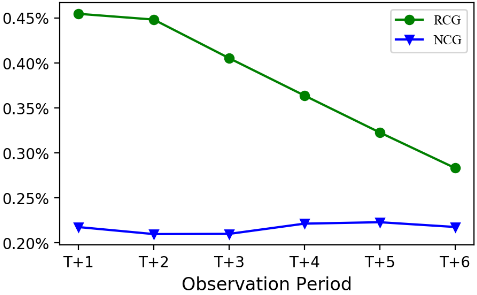


Figure 2: Group-buy-ratio of RCG and NCG.

Background

Different agents have different influence on users, a user assigned with the top 5% agent will be more likely to buy online insurance.

Table 1: Ask-buy-ratio of different agents.

communication frequency order	T+1	T+2	T+3
top 5%	4.9784%	4.9750%	4.7591%
top 10%	2.2700%	2.3214%	2.2698%
top 15%	1.0184%	1.0635%	1.0145%

Contributions

Our contributions in this paper are as follows:

- To the best of our knowledge, this is the first work to combine cross-domain mechanism and heterogeneous information network to give personalized recommendations for cold start users in insurance domain.
- For the complexity of insurance products, we construct a heterogeneous information network, which contains four types of nodes and six types of relations. And we employ three level aggregations over IHIN to learn more effective user and item representations in insurance domain.
- We conduct experiments on real-world recommendation scenarios, and the results prove the efficacy of HCDIR over several state-of-the-art baselines.

Framework

To provide recommendations to cold start users, we propose HCDIR. HCDIR contains three main parts: learning latent features of users in both insurance domain and nonfinancial domain, mapping of user latent features.

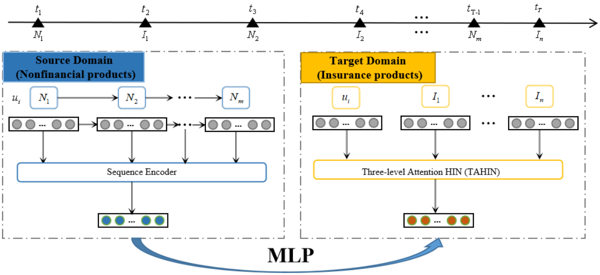
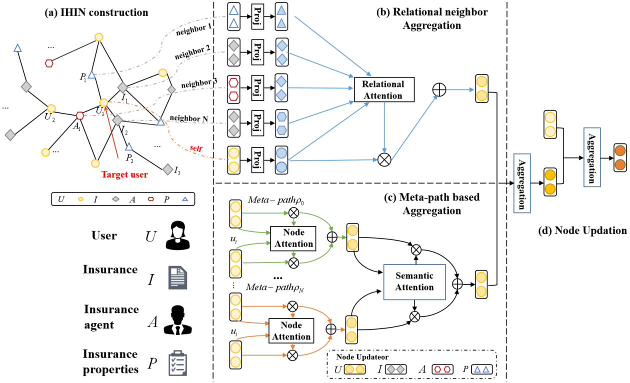


Figure 3: The Framework of HCDIR

Target Domain

- Aggregate one-hop heterogeneous neighbors
- Aggregate meta-paths based neighbors
- Aggregate meta-paths based neighbor sets



Target Domain

Attention Score:

$$\alpha_{ew} = \frac{\exp(f_r(\boldsymbol{h}_e^0, \boldsymbol{P}_r \boldsymbol{h}_w^0))}{\sum_{j \in \mathcal{N}_1(e)} \exp(f_r(\boldsymbol{h}_e^0, \boldsymbol{P}_r \boldsymbol{h}_j^0))}$$

Aggregation:

$$\boldsymbol{h}_e^1 = \sigma \left(\sum_{w \in \mathcal{N}_1(e)} \alpha_{ew} \boldsymbol{h}_w^0 \right),$$

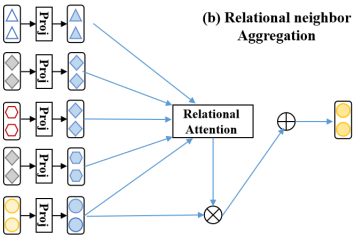


Figure 4: Relational Neighbor Aggregation

Target Domain

Meta-path based neighbors:

$$\beta_{ew}^{\rho} = \frac{\exp(f_\rho(\boldsymbol{h}_e^0, \boldsymbol{h}_w^0))}{\sum_{j \in \mathcal{N}_\rho(e)} \exp(f_\rho(\boldsymbol{h}_e^0, \boldsymbol{h}_j^0))}$$

$$\boldsymbol{h}_e^{\rho} = \sigma\left(\sum_{w \in \mathcal{N}_\rho(e)} \beta_{ew}^{\rho} \boldsymbol{h}_w^0\right)$$

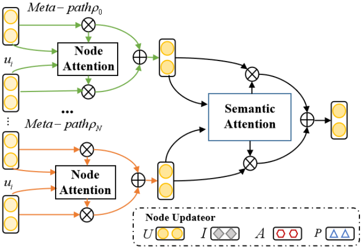


Figure 5: Meta-path based Aggregation

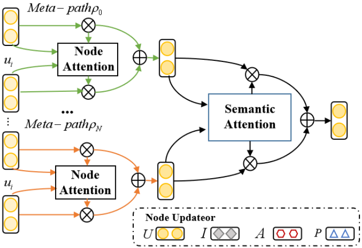
Target Domain

Meta-path based neighbor sets:

$$w_{\rho_j} = \frac{1}{|\mathcal{V}|} \sum_{e \in \mathcal{V}} \boldsymbol{q}^T \cdot \tanh(\boldsymbol{W} \boldsymbol{h}_e^{\rho_j} + b),$$

$$\gamma_{\rho_j} = \frac{\exp(w_{\rho_j})}{\sum_{j=1}^{N} \exp(w_{\rho_j})}.$$

$$\boldsymbol{h}_e^2 = \sum_{j=1}^{N} \gamma_{\rho_j} \cdot \boldsymbol{h}_e^{\rho_j}.$$



Source Domain

In Jinguanjia, each item in nonfinancial domain is associated with a description. In order to learn more effective latent features, we employ word2vec. Suppose there are n words in item's content. Then we utilize word2vec to obtain word vectors, which are represented as $\{\boldsymbol{w}_k^i\}_{k=1}^n$. Then we concatenate word vectors and apply a max pooling over it to get the final item embedding:

$$\boldsymbol{i} = \text{max-pooling}(\text{concat}(\{\boldsymbol{w}_k^i\}_{k=1}^n)).$$

- Background
- Contributions
- Framework
- Target Domain
- Source Domain

3 Results
- Dataset
- Performance

Dataset

Table 2: Statistics of Our dataset.

IS-domain (Target domain)		NF-domain (Source domain)	
#User Nodes	117,613	#Users	117,613
#Item Nodes	42	#Items	19,266
#Agent Nodes	90,377	#User-Item Interactions	1,995,168
#Insurance Property Nodes	35		
#User-Iten Relations	344,206		
#User-Agent Relations	97,343		
#Item-Property Relations	275		

Performance

Table 3: Performance comparison.

Jinguanjia dataset			Metrics			
η	Group	Method	NDCG	Rec@1	Rec@3	Rec@5
10%	Single-domain RS	BPR	0.0719	0.0213	0.0737	0.1248
		GRU4REC	0.0036	0.0017	0.0032	0.0057
	Cross-domain RS	EMCDR-BPR	0.0881	0.0324	0.0689	0.1543
		EMCDR-GRU	0.1013	0.0284	0.0961	0.2088
		HCDIR-RGCN	0.2468	0.0967	0.3448	0.3849
		HCDIR-HAN	0.3206	0.1236	0.3476	0.4828
		HCDIR	0.3674	0.1366	0.4002	0.5543
20%	Single-domain RS	BPR	0.0789	0.0241	0.0864	0.1348
		GRU4REC	0.0042	0.0022	0.0047	0.0061
	Cross-domain RS	EMCDR-BPR	0.0984	0.0347	0.0848	0.1611
		EMCDR-GRU	0.1112	0.0366	0.1308	0.2257
		HCDIR-RGCN	0.2579	0.1003	0.3516	0.4002
		HCDIR-HAN	0.3311	0.1273	0.3656	0.4927
		HCDIR	0.3769	0.1548	0.4189	0.5683

Performance

Table 4: Performance comparison (cont.)

Jinguanjia dataset			Metrics			
50%	Single-domain RS	BPR	0.0791	0.0274	0.1205	0.1735
		GRU4REC	0.0117	0.0027	0.0114	0.0213
	Cross-domain RS	EMCDR-BPR	0.1125	0.0402	0.1609	0.2281
		EMCDR-GRU	0.1289	0.0496	0.1594	0.2589
		HCDIR-RGCN	0.2701	0.1166	0.3611	0.4341
		HCDIR-HAN	0.3432	0.1341	0.3946	0.5372
		HCDIR	0.3895	0.1636	0.4354	0.5827
100%	Single-domain RS	BPR	0.1009	0.0354	0.1627	0.1809
		GRU4REC	0.0137	0.0054	0.0154	0.0221
	Cross-domain RS	EMCDR-BPR	0.1359	0.0511	0.2059	0.2556
		EMCDR-GRU	0.1498	0.0806	0.2124	0.2486
		HCDIR-RGCN	0.3067	0.1247	0.3739	0.4974
		HCDIR-HAN	0.3703	0.1357	0.4254	0.5627
		HCDIR	0.4109	0.1873	0.4654	0.6128

Performance

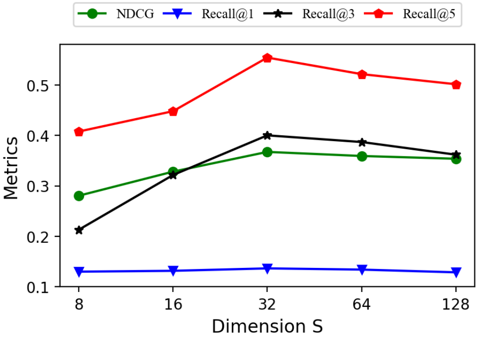


Figure 6: Study of the Final Embedding Dimension S.

Performance

Table 5: Performance of variants of HCDIR on Jinguanjia dataset at 10% sparsity level

inguanjia dataset at 10% sparsity level		Metrics			
		NDCG	Rec@1	Rec@3	Rec@5
Data Ablation	#HCDIR only with interactions	0.1013 **-72.43%**	0.0284 **-79.21%**	0.0961 **-75.99%**	0.2088 **-62.33%**
	#HCDIR without Agent	0.2157 **-41.29%**	0.0933 **-31.70%**	0.2287 **-42.85%**	0.3277 **-40.88%**
	#HCDIR without IP	0.2313 **-37.04%**	0.1073 **-21.45%**	0.2512 **-37.23%**	0.3665 **-33.88%**
Model Ablation	#HCDIR using RGCN	0.2468 **-32.83%**	0.0967 **-29.21%**	0.3448 **-13.84%**	0.3849 **-30.56%**
	#HCDIR using HAN	0.3206 **-12.74%**	0.1236 **-9.52%**	0.3476 **-13.14%**	0.4828 **-12.90%**
Full Model	#HCDIR	0.3674	0.1366	0.4002	0.5543

Performance

Metrics	G_HCDIR without agent		
	T+1 month	T+2 months	T+3 months
improvement percentage of UPCR vs G_Baseline	8.79%	12.87%	18.38%
improvement percentage of UPGR vs G_Baseline	10.97%	13.04%	15.31%
improvement percentage of runing time vs G_Baseline	-79.94%		

Performance

Metrics	G_HCDIR		
	T+1 month	T+2 months	T+3 months
improvement percentage of UPCR vs G_Baseline	12.94%	18.66%	23.20%
improvement percentage of UPGR vs G_Baseline	15.25%	20.41%	25.62%
improvement percentage of runing time vs G_Baseline	-76.59%		

Q&A

If you have any questions, please email to magicyebi@163.com

Acknowledgments

Thank you for your attention!