Cash-out User Detection based on Attributed Heterogeneous Information Network with a Hierarchical Attention Mechanism

Binbin Hu[1], Zhiqiang Zhang[2], Chuan Shi[1], Jun Zhou[2], XiaoLong Li[2], Yuan Qi[2]
[1]Beijing University of Posts and Telecommunications, Beijing, China
[2]Ant Financial, Hangzhou, Zhejiang, China

CONTENTS

CONTENTS

Cash-out User Detection

- **Credit Payment Services :** such as offline credit card services in commercial banks and online credit payments in internet financial institutions.
- **Cash-out Fraud :** pursue cash gains with illegal or insincere means, e.g., through buying pre-paid cards or other goods then reselling them.
- **Cash-out User Detection :** predict whether a user will do cash-out transactions or not in the future.

Classical Methods

Conventional solutions

- First perform subtle feature engineering for each user
- then a classifier (e.g., tree based model or neural network) is trained based on these features.
- Seldom fully exploit the interaction relations → Limited prediction ability

Integrate more auxiliary information

- The fund transfer relation among users
- The login relation between users and devices
- The transaction relation between users and merchants
- Abundant attribute information

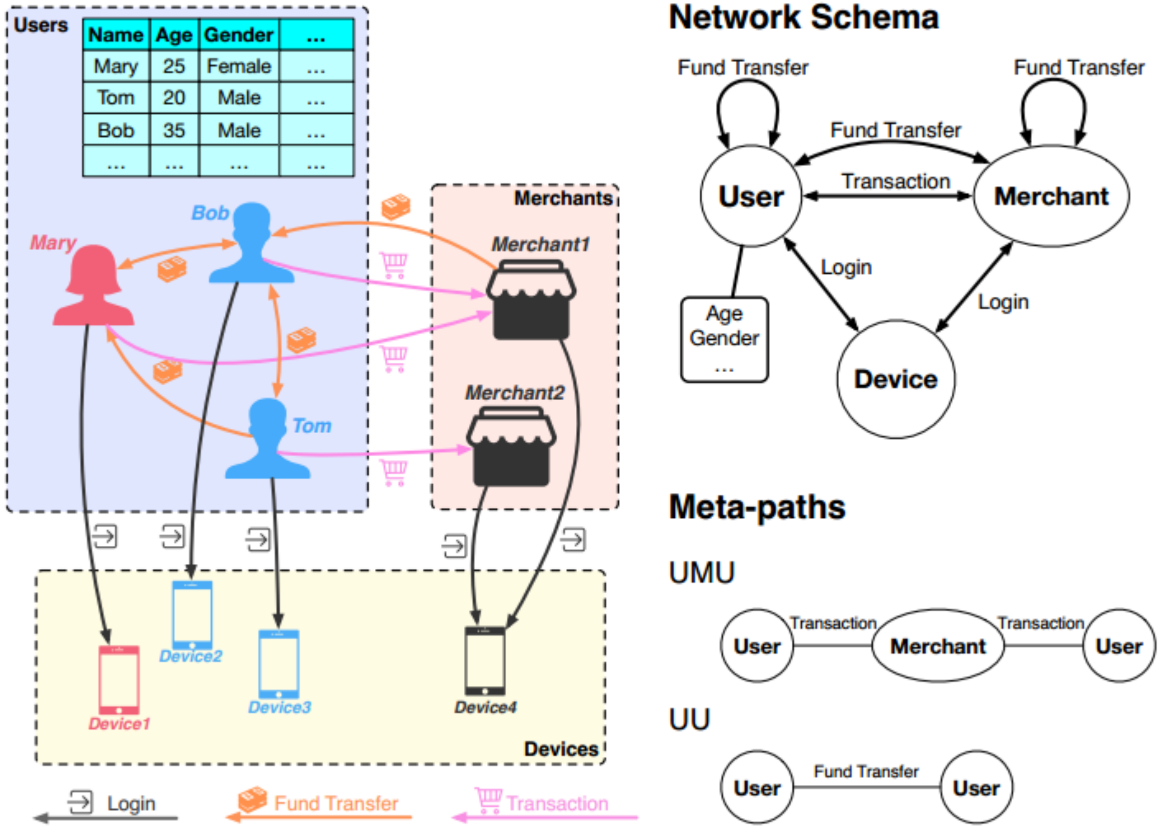
Attributed Heterogeneous information network is a promising way to integrate auxiliary data.

Attributed Heterogeneous Information Network (AHIN)

- Include multiple types of nodes or links and rich attribute information
- Flexibly characterize heterogeneous data
- Contain rich semantics

Meta-path

- A relation sequence connecting two objects in HIN
- Extract structural features
- Embody path semantics



(a) Scenario of credit payment service (b) Network schema and meta-path examples

Contributions

- First study the cash-out users detection problem, which is a very important and widely existing problem in financial fraud field
- Propose to model the cash-out user detection problem as a classification problem in AHIN which is constituted by different types of objects and their rich interactions in the scenario of credit payment service
- Propose a novel model HACUD to solve the problem with meta-path based neighbors and a hierarchical attention mechanism
- Extensive experiments on two real datasets illustrate the best performance of the proposed HACUD

CONTENTS

Observations in real data

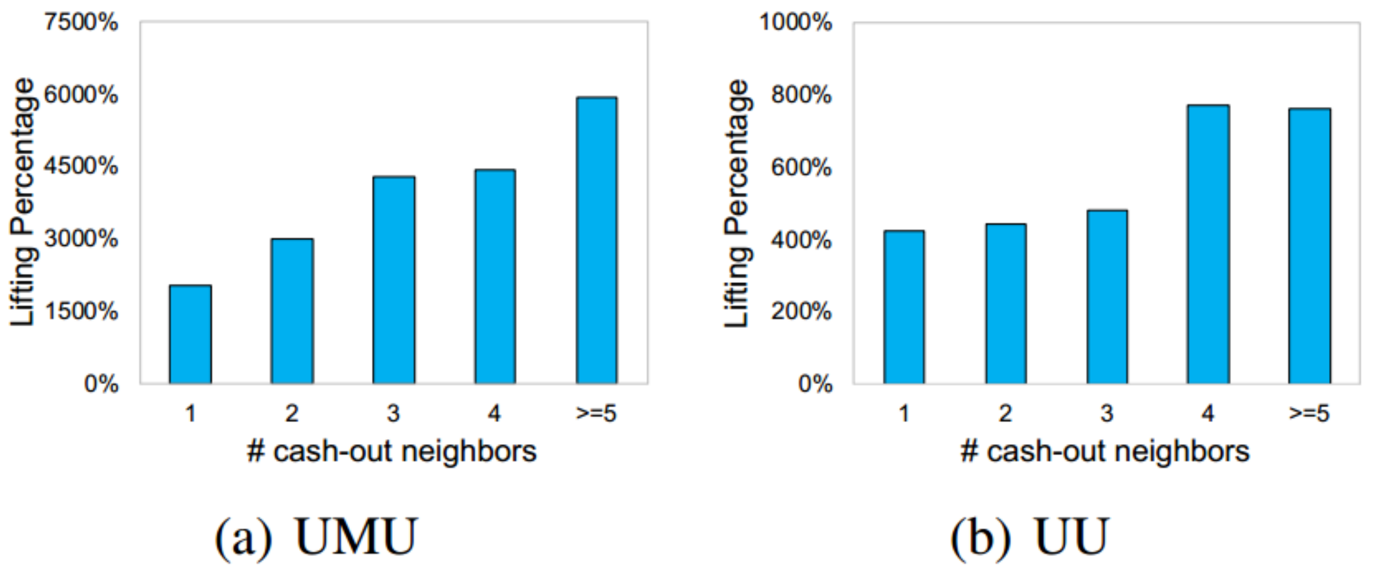
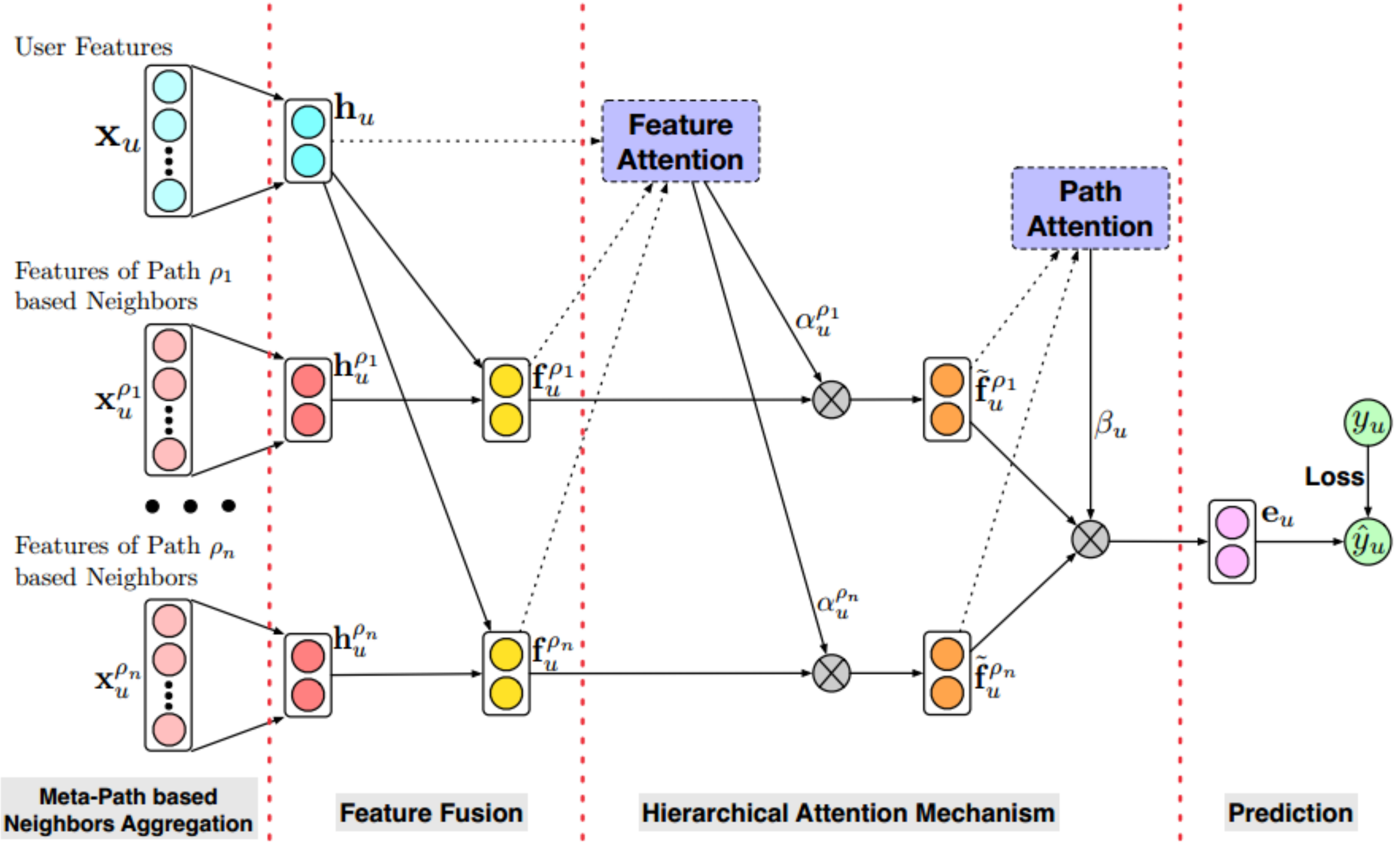


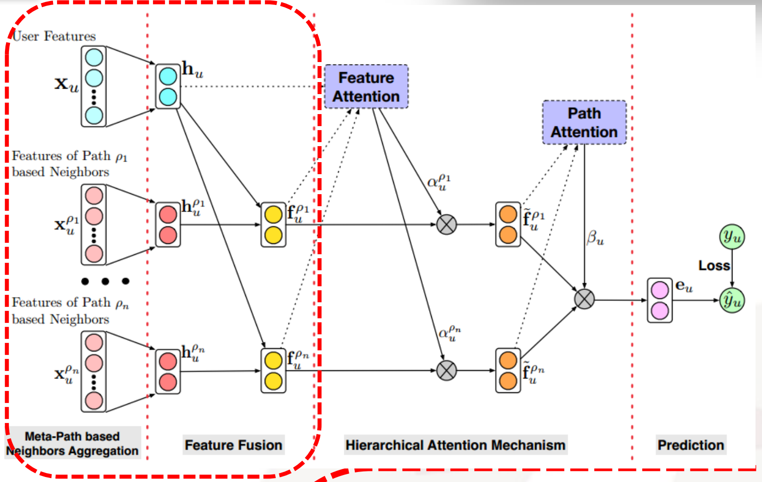
Figure 3: The lifting percentages of cash-out rate in users with different amount of cash-out neighbors against users without any cash-out neighbor in two meta-paths.

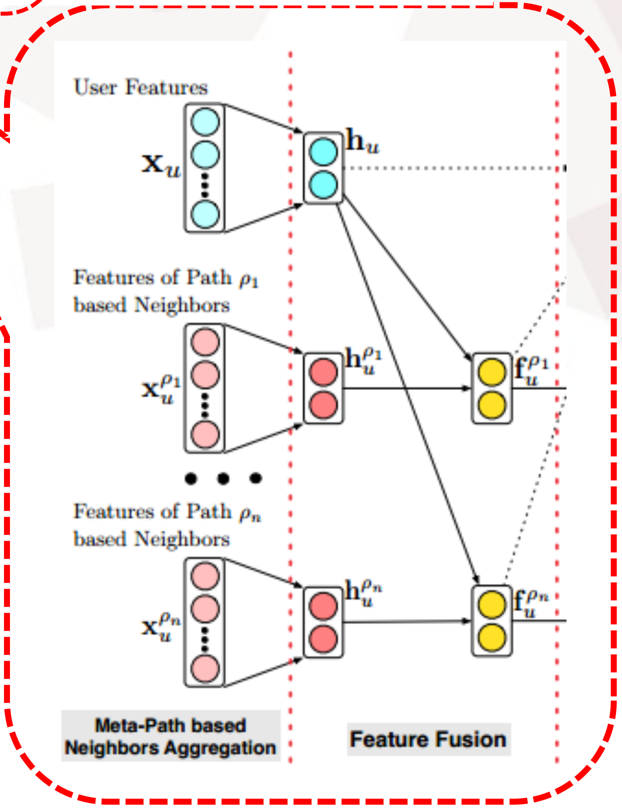
- Users with higher cash-out rate tend to have more cash-out neighbors.
- Different meta-path based neighbors have different impacts on users.

Overall Framework

Hierarchical Attention mechanism based Cash-out User Detection model (**HACUD**)







Meta-path based Neighbors

Giving a user u in an AHIN, the meta-path based neighbors is defined as the set of aggregate neighbors under the given meta-path for the user u in the AHIN.

Aggregation

$$\mathbf{x}_u^\rho = \sum_{j \in \mathcal{N}_u^\rho} w_{uj}^\rho * \mathbf{x}_j,$$

Feature Fusion

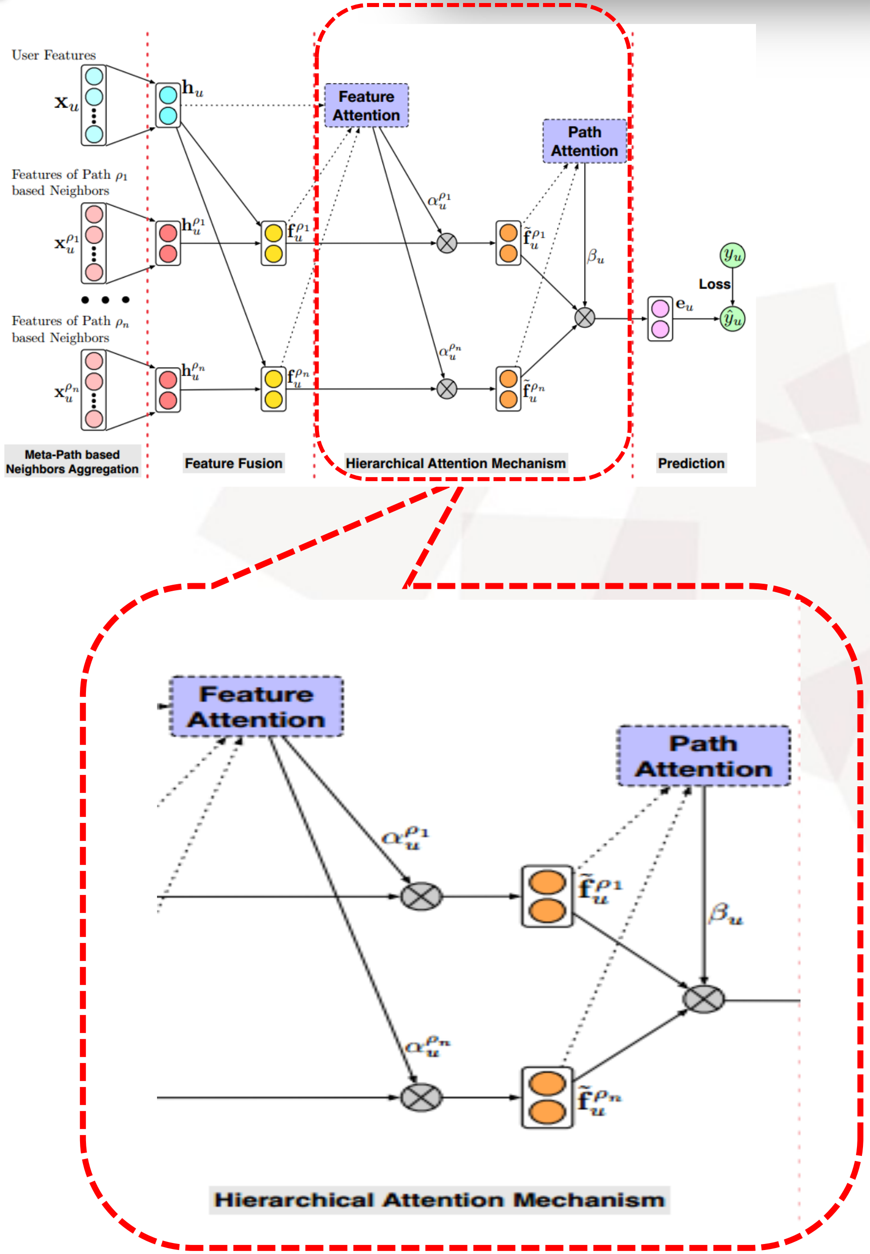
- Project the original sparse features

$$\mathbf{h}_u = \mathbf{W}\mathbf{x}_u + \mathbf{b}, \quad \mathbf{h}_u^\rho = \mathbf{W}^\rho \mathbf{x}_u^\rho + \mathbf{b}^\rho,$$

- Fuse the latent representations

$$\mathbf{f}_u^\rho = \mathrm{ReLU}(\mathbf{W}_F^\rho g(\mathbf{h}_u, \mathbf{h}_u^\rho) + \mathbf{b}_F^\rho).$$

2 HACUD



Hierarchical Attention Model

Different preferences over the features based on different meta-paths as well as attribute information

Feature Attention

Attention score

$$v_u^\rho = \mathrm{ReLU}(\mathbf{W}_f^1[\mathbf{h}_u;\mathbf{f}_u^\rho] + \mathbf{b}_f^1),$$

$$\alpha_u^\rho = \mathrm{ReLU}(\mathbf{W}_f^2 v_u^\rho + \mathbf{b}_f^2),$$

Softmax

$$\hat{\alpha}_{u,i}^\rho = \frac{\exp(\alpha_{u,i}^\rho)}{\sum_{j=1}^{K}\exp(\alpha_{u,j}^\rho)}.$$

Apply

$$\tilde{\mathbf{f}}_u^\rho = \hat{\alpha}_u^\rho \odot \mathbf{f}_u^\rho,$$

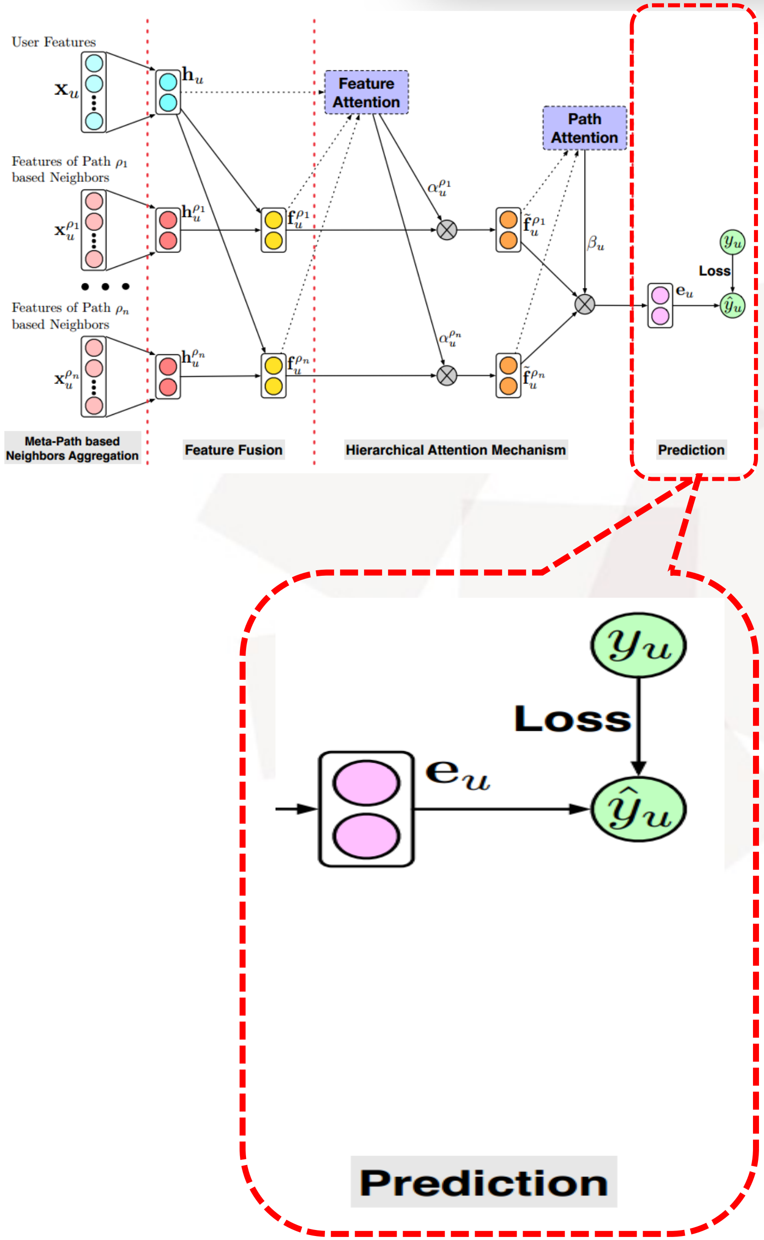
Path Attention

Attention score

$$\beta_{u,\rho} = \frac{\exp(\mathbf{z}^{\rho\,\mathrm{T}} \cdot \tilde{\mathbf{f}}_u^C)}{\sum_{\rho' \in \mathcal{P}} \exp(\mathbf{z}^{\rho'\,\mathrm{T}} \cdot \tilde{\mathbf{f}}_u^C)},$$

Apply

$$\mathbf{e}_u = \sum_{\rho \in \mathcal{P}} \beta_{u,\rho} * \tilde{\mathbf{f}}_u^\rho,$$



Model Learning

Multi-layer Perceptron

$$\mathbf{z}_u = \mathrm{ReLU}(\mathbf{W}_L \cdots \mathrm{ReLU}(\mathbf{W}_1 \mathbf{e}_u + \mathbf{b}_1) + \mathbf{b}_L),$$

Prediction

$$p_u = \mathrm{sigmoid}(\mathbf{w}_p^T \mathbf{z}_u + b_p).$$

Loss

$$\mathcal{L}(\Theta) = \sum_{\langle u, y_u \rangle \in \mathcal{D}} (y_u \log(p_u) + (1 - y_u) \log(1 - p_u)) + \lambda ||\Theta||_2^2,$$

Discussion

Flexibility

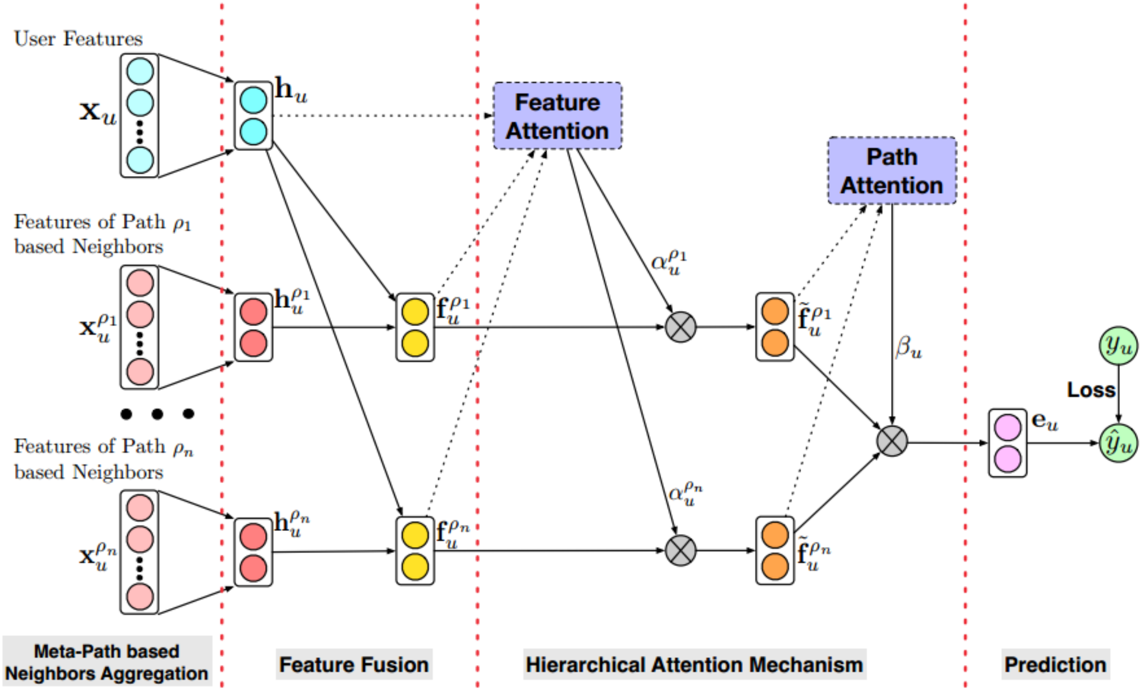
Flexible framework to leverage structure and attribute information through capturing multiple aspects in AHIN

Scalability

Compared to traditional network embedding methods (SDNE, Structure2vec), which represent nodes via their context (e.g., adjacency matrix), our method is more suitable for large-scale networks

Dynamics

For a new user which never appears in training set, the proposed model can also learn the representation through his/her meta-path based neighbors in networks.



CONTENTS

Datasets

Ten Days Dataset

Contains 1.88 million users ranging from 2018/03/21 to 2018/03/31 for training

One Month Dataset

Contains 5.16 million users ranging from 2018/03/01 to 2018/03/31 for training

For both datasets, we predict the cash-out probability of users in 2018/05/01 (around 0.17 million users).

Metrics

$$AUC = \frac{\sum_{u \in \mathcal{U}^+} rank_u - \frac{|\mathcal{U}^+| \times (|\mathcal{U}^+|+1)}{2}}{|\mathcal{U}^+| \times |\mathcal{U}^-|}.$$

AHIN

- **#users 56.75 millions**
- **#merchants 0.51 millions**
- **#fund transfer relations 77.40 millions**
- **#transaction relations 20.64 millions**
- **#attributes 123**

Table 1: Selected meta-paths and meta-path based neighbors statistics.

Meta-paths	#Neighbors (Min / Max / Avg.)
User-(transaction) -Merchant-(transaction)-User	1 / 16860 / 309
User-(fund transfer) -User	1 / 26235 / 150
User-(transaction) -Merchant	1 / 81 / 4

Methods to Compare

Attribute only or Structure only
- GBDT
- Node2vec
- Metapath2vec

Structure + Attribute
- Node2vec + Feature
- Metapath2vec + Feature

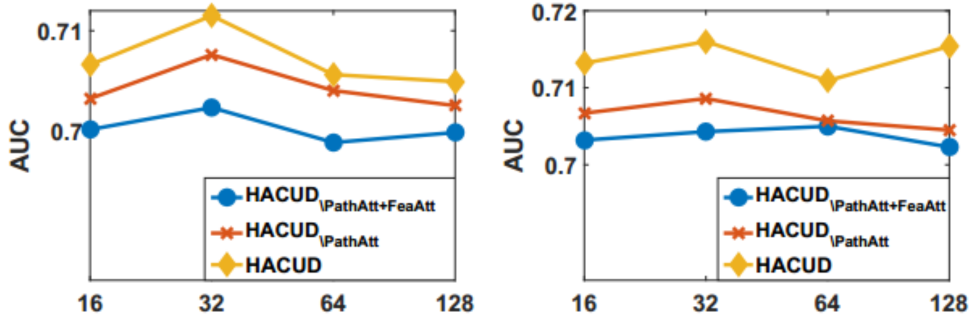
Structure + Attribute + Label
- Structure2vec
- $\text{GBDT}_{\text{Struct}}$

Table 2: Results of effectiveness experiments on two datasets *w.r.t.* the dimension of latent representation d. A larger value indicates a better performance.

Algorithm	AUC							
	Ten Days Dataset				One Month Dataset			
	$d = 16$	$d = 32$	$d = 64$	$d = 128$	$d = 16$	$d = 32$	$d = 64$	$d = 128$
Node2vec	0.5893	0.5913	0.5926	0.5930	0.5980	0.6963	0.6009	0.6021
Metapath2vec	0.5914	0.5903	0.5917	0.5920	0.6005	0.5976	0.5995	0.5983
Node2vec + Feature	0.6455	0.6464	0.6510	0.6447	0.6541	0.6561	0.6607	0.6518
Metapath2vec + Feature	0.6456	0.6429	0.6469	0.6485	0.4850	0.6552	0.6523	0.6545
Structure2vec	0.6537	0.6556	0.6598	0.6545	0.6641	0.6632	0.6657	0.6678
GBDT	0.6389	0.6389	0.6389	0.6389	0.6467	0.6467	0.6467	0.6467
GBDT_{Struct}	0.6948	0.6948	0.6948	0.6948	0.6968	0.6968	0.6968	0.6968
HACUD	**0.7066**	**0.7115**	**0.7056**	**0.7049**	**0.7132**	**0.7160**	**0.7109**	**0.7154**

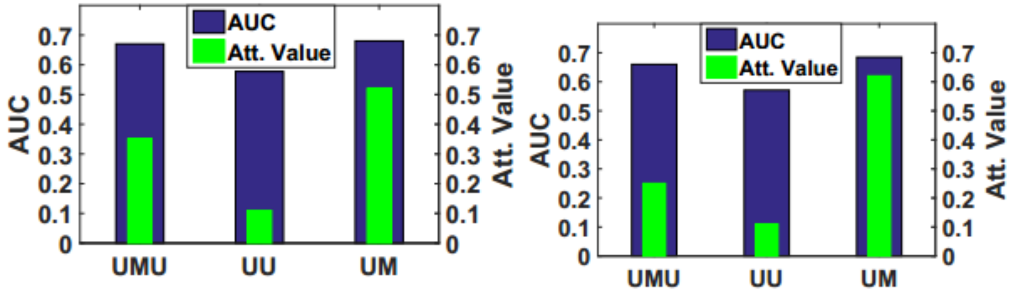
3 Experiments

Effects of Hierarchical Attention



(a) Ten Days Dataset (b) One Month Dataset

Figure 4: Performance comparison of hierarchical attention *w.r.t.* the dimension of latent representation d.



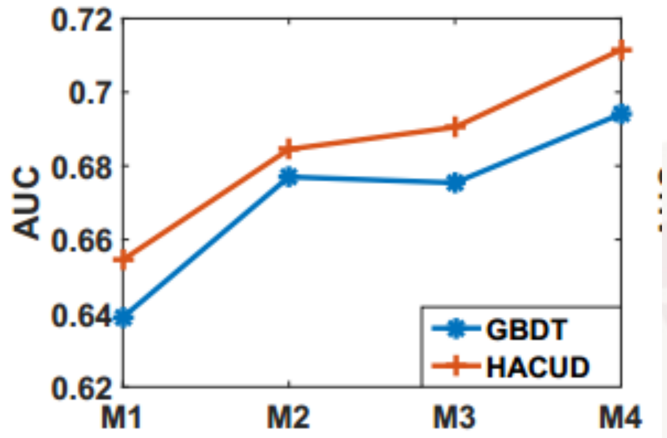
(a) Ten Days Dataset (b) One Month Dataset

Figure 5: Performances comparison on different meta-paths and corresponding attention values.

The hierarchical mechanism is able to better utilize the user feature and features generated by meta-paths

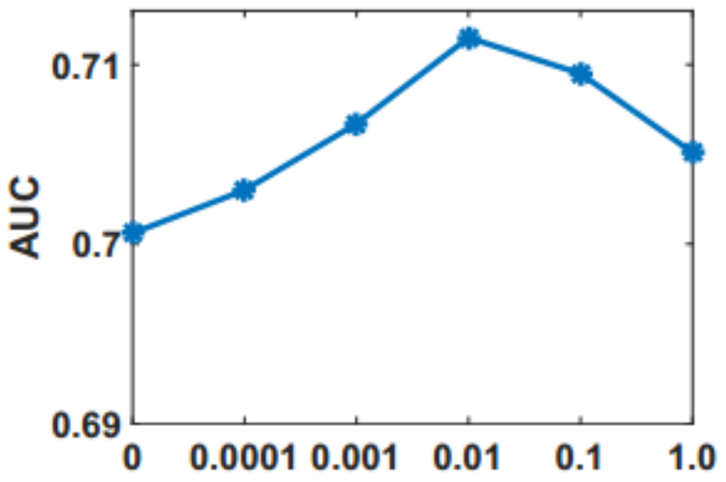
the corresponding attentions are positively correlated (i.e., important meta-paths tend to attract more attentions)

Impact of Different Meta-paths (Ten Days Dataset)



(1) M1 : user feature only;
(2) M2 : user feature + UMU;
(3) M3 : user feature + UMU + UU;
(3) M4 : user feature + UMU + UU + UM

Impact of parameter λ (Ten Days Dataset)



CONTENTS

4 Conclusions

- we first study the cash-out user detection problem under the AHIN framework and propose a novel HACUD model
- we design a hierarchical attention mechanism to model user preferences towards attributes and meta-paths
- extensive experiments for the cash-user detection task demonstrate the effectiveness of our model

Thanks
Q&A