

Yue-yang WANG, Wei-hao JIANG, Shi-liang PU, Yue-ting ZHUANG, 2020.
Learning embeddings of a heterogeneous behavior network for potential
behavior prediction. *Frontiers of Information Technology & Electronic
Engineering*, 21(3):422-435. <https://doi.org/10.1631/FITEE.1800493>

Learning embeddings of a heterogeneous behavior network for potential behavior prediction

Key words: Network embedding; Representation learning; Human
behavior; Social networks; Heterogeneous information network;
Attribute

Corresponding author: Yue-yang WANG

E-mail: yueyangw@zju.edu.cn



ORCID: <https://orcid.org/0000-0003-3210-0930>

Motivation

1. Exploring and analyzing the human behavior of specific groups can assist some organizations in making decisions.
2. Human behavior is implicated in social actions that individuals have already performed, but these actions are difficult to observe in their entirety.
3. Based on the insufficient information, predicting individuals who may take specific actions and further categorizing them into specific groups, which we call potential behavior prediction, is a challenging task.

Main idea

1. We propose a novel network embedding model that captures structural semantics, attribute semantics, and the interrelationships between them in the heterogeneous behavior network (HBN) for potential behavior prediction.
2. We design an attribute encoder network to learn the mapping of heterogeneous attributes into a common space.
3. Experiments on two real-world datasets show that this approach outperforms other similar methods on various heterogeneous information network mining tasks for potential behavior prediction.

Method

1. a⁴HNE jointly embeds human beings and actions associated with attributes in the HBN into a common space.
2. Three important points are considered, i.e., structure proximity which suggests that the embeddings of explicitly connected nodes are similar to each other, attribute resemblance which indicates that nodes with similar attributes tend to have similar embeddings, and heterogeneity fusion by which the heterogeneous structure and the attributes are fused.

Major results

- Results of link prediction on the AMiner and HIK datasets

Method	Results of link prediction on AMiner			
	$\alpha=30\%$	$\alpha=50\%$	$\alpha=70\%$	$\alpha=90\%$
Doc2Vec	0.6211	0.6199	0.6185	0.6112
DeepWalk	0.8966	0.9484	0.9660	0.9732
node2vec	0.8948	0.9465	0.9664	0.9736
($p=2, q=0.5$)				
node2vec	0.8947	0.9479	0.9665	0.9746
($p=0.5, q=2$)				
LINE	0.8976	0.9492	0.9620	0.9668
CANE	0.9065	0.9345	0.9439	0.9446
PTE	0.7429	0.8917	0.9444	0.9614
Metapath2vec++	0.8133	0.8610	0.8813	0.8870
a ⁴ HNE-NH	0.9357	0.9598	0.9684	0.9718
a ⁴ HNE-NC	0.9008	0.9580	0.9705	0.9741
a ⁴ HNE	0.9383	0.9659	0.9730	0.9747

Method	Results of link prediction on HIK			
	$\alpha=30\%$	$\alpha=50\%$	$\alpha=70\%$	$\alpha=90\%$
DeepWalk	0.6004	0.6581	0.7336	0.7690
node2vec	0.5995	0.6786	0.7363	0.7740
($p=2, q=0.5$)				
node2vec	0.5963	0.6709	0.7220	0.7575
($p=0.5, q=2$)				
LINE	0.5952	0.6640	0.7157	0.7583
PTE	0.5885	0.6572	0.7148	0.7592
Metapath2vec++	0.5043	0.5401	0.5391	0.5459
a ⁴ HNE-NC	0.6843	0.7089	0.7220	0.7960
a ⁴ HNE	0.8048	0.8335	0.8465	0.8645

Major results (Cont'd)

- Results of classification on the HIK and AMiner datasets

Dataset	Method	Results of classification			
		30%	50%	70%	90%
HIK (user)	DeepWalk	0.6026	0.6075	0.6085	0.6119
	node2vec	0.6146	0.6281	0.6321	0.6194
	($p=2, q=0.5$)				
	node2vec	0.5858	0.5979	0.5982	0.6071
	($p=0.5, q=2$)				
	LINE	0.5849	0.5971	0.5915	0.5952
	PTE	0.6015	0.6069	0.6097	0.6125
	Metapath2vec++	0.6063	0.6246	0.6347	0.6310
	a ⁴ HNE-NH	0.6331	0.6392	0.6397	0.6406
	a ⁴ HNE-NC	0.6071	0.6082	0.6252	0.6440
	a ⁴ HNE	0.6534	0.6654	0.6715	0.6833

AMiner (author)	Doc2Vec	0.6507	0.6674	0.6787	0.6785
	DeepWalk	0.7251	0.7316	0.7338	0.7378
	node2vec	0.7243	0.7312	0.7312	0.7307
	($p=2, q=0.5$)				
	node2vec	0.7237	0.7320	0.7334	0.7344
	($p=0.5, q=2$)				
	LINE	0.7285	0.7314	0.7347	0.7371
	CANE	0.7871	0.7920	0.7933	0.7971
	PTE	0.7670	0.7712	0.7749	0.7799
	Metapath2vec++	0.7589	0.7653	0.7676	0.7655
AMiner (paper)	a ⁴ HNE-NH	0.7674	0.7736	0.7746	0.7802
	a ⁴ HNE-NC	0.7777	0.7846	0.7876	0.7868
	a ⁴ HNE	0.7958	0.7968	0.8027	0.7996
	Doc2Vec	0.6652	0.6811	0.6938	0.6939
	DeepWalk	0.8956	0.9003	0.9016	0.9015
	node2vec	0.8986	0.9013	0.9043	0.9089
	($p=2, q=0.5$)				
	node2vec	0.8981	0.9023	0.9034	0.9061
	($p=0.5, q=2$)				
	LINE	0.8826	0.8869	0.8901	0.8914
	CANE	0.9206	0.9250	0.9249	0.9268
	PTE	0.9342	0.9372	0.9390	0.9411
	Metapath2vec++	0.9349	0.9386	0.9402	0.9389
	a ⁴ HNE-NH	0.9147	0.9185	0.9206	0.9242
	a ⁴ HNE-NC	0.9155	0.9201	0.9231	0.9276
	a ⁴ HNE	0.9434	0.9462	0.9469	0.9502

Major results (Cont'd)

- Results of clustering on the HIK and AMiner datasets

Method	NMI value for clustering		
	HIK (user)	AMiner (author)	AMiner (paper)
Doc2Vec	—	0.2977	0.2251
DeepWalk	0.0192	0.3791	0.6001
node2vec ($p=2, q=0.5$)	0.0159	0.3852	0.6238
node2vec ($p=0.5, q=2$)	0.0184	0.3783	0.6245
LINE	0.0006	0.3792	0.5275
CANE	—	0.4726	0.6684
PTE	0.0194	0.4358	0.6769
Metapath2vec++	0.0301	0.4514	0.7014
a ⁴ HNE-NH	0.0352	0.4564	0.6493
a ⁴ HNE-NC	0.0312	0.4641	0.6921
a ⁴ HNE	0.0379	0.4992	0.7203

Major results (Cont'd)

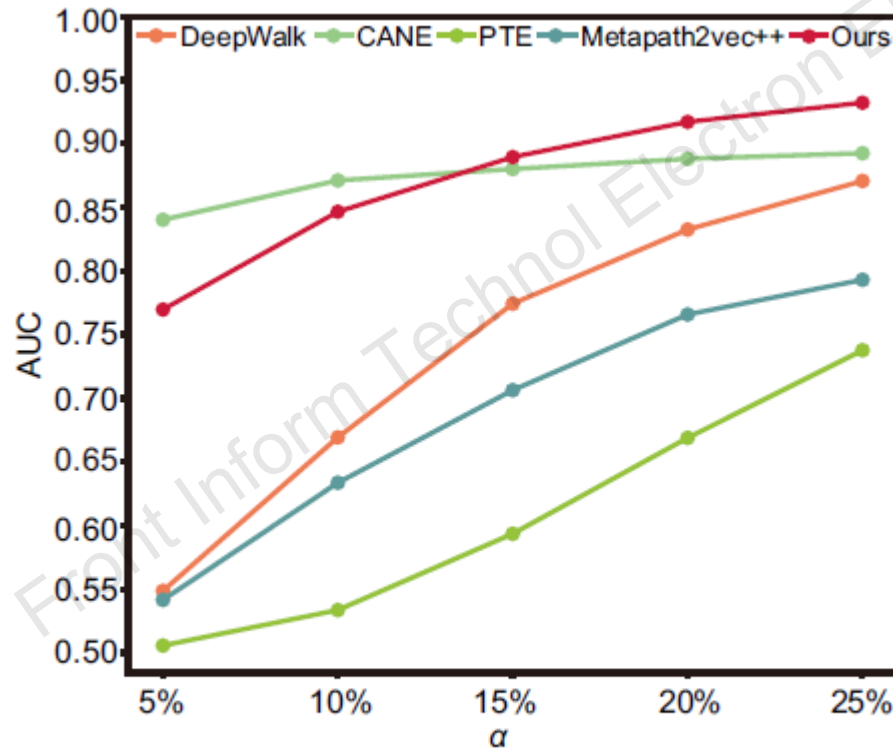


Fig. 3 Results of cold-start prediction on the AMiner dataset

Major results (Cont'd)

- Computing systems
- Theoretical computer science
- Computer networks & wireless communication
- Computer graphics
- Human-computer interaction
- Computational linguistics
- Computer vision & pattern recognition
- Databases & information systems

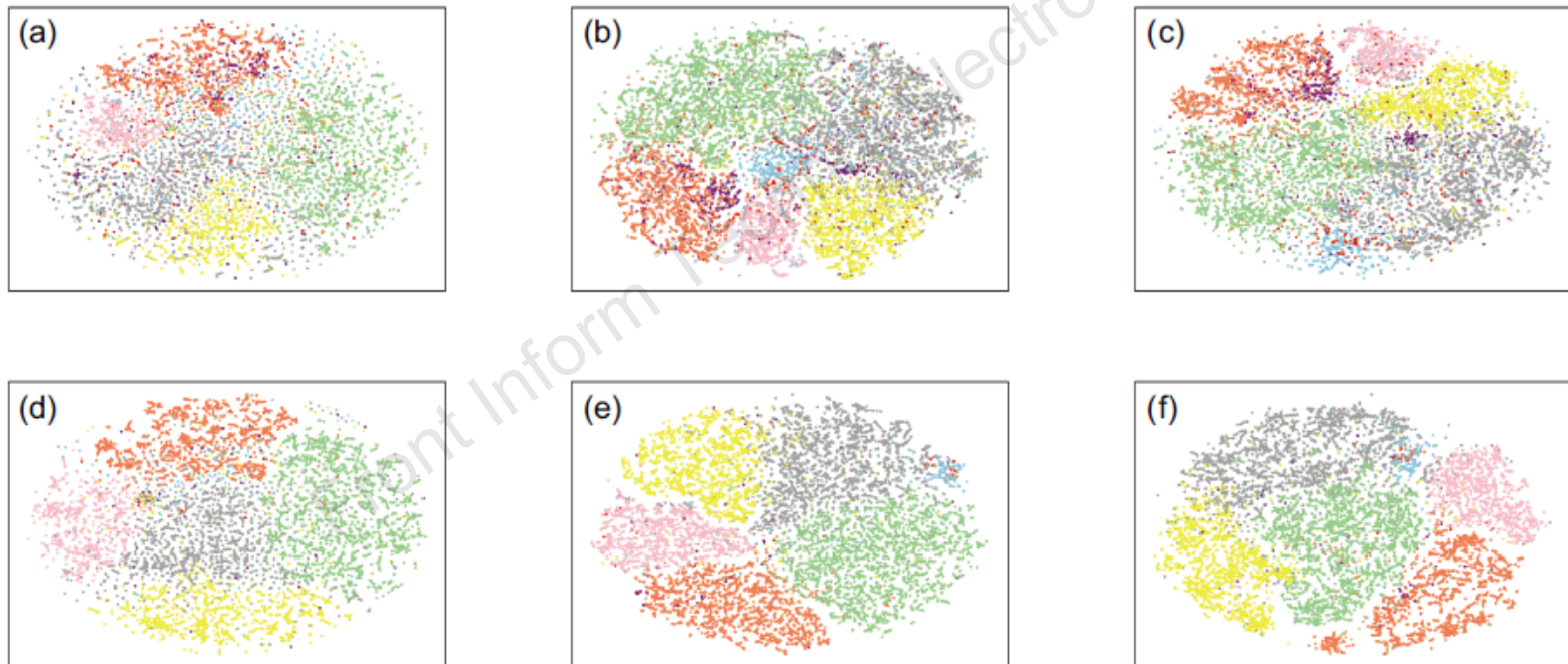


Fig. 4 Visualization results on the AMiner dataset. PTE (a), CANE (b), and a⁴HNE (c) are results for authors; PTE (d), CANE (e), and a⁴HNE (f) are results for papers. References to color refer to the online version of this figure

Major results (Cont'd)

- Hyper-parameter analysis on the HIK dataset (TR=50%)

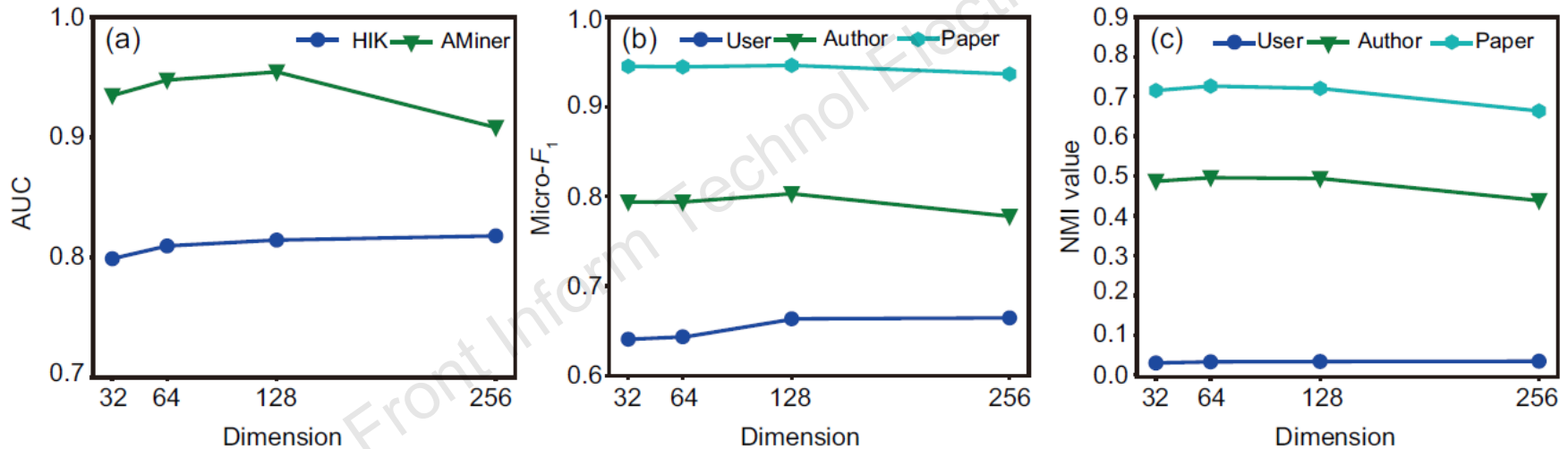


Fig. 5 Parameter sensitivity study: (a) link prediction, $\alpha = 50\%$; (b) classification, $T_R=50\%$; (c) clustering

Major results (Cont'd)

Table 5 Hyper-parameter analysis on the HIK dataset ($T_R = 50\%$)

Weight	Results of hyper-parameter analysis			
	$\beta_1=0.01$	$\beta_1=0.1$	$\beta_1=1$	$\beta_1=10$
$\beta_0=0.01$	0.6354	0.6490	0.6420	0.6375
$\beta_0=0.1$	0.6470	0.6404	0.6654	0.6553
$\beta_0=1$	0.6172	0.6473	0.6504	0.6561
$\beta_0=10$	0.6175	0.6257	0.6383	0.6471

The best results are in bold

Conclusions

1. We have constructed the HBN from mass humans and actions information.
2. We have introduced a novel network embedding method, human-action-attribute-aware heterogeneous network embedding (a^4HNE), which handles the HBN and embeds nodes in the HBN into a common space.
3. Extensive experiments demonstrated that a^4HNE is effective and robust for various heterogeneous information network mining tasks for human potential behavior prediction.