Heterogeneous Information Network Embedding with Incomplete Multi-view Fusion

Susu ZHENG, Donghai GUAN, Weiwei YUAN

Frontiers of Computer Science, DOI: 10.1007/s11704-021-1057-6

Problems & Ideas

- Problems of HIN embedding with incomplete multi-view fusion.
  - Multi-view based HIN embedding can preserve rich semantic information of multiple views in HIN.
  - Nodes and edges are incomplete in single-view due to privacy issues or other reasons.

- The key idea is to reconstruct the incomplete view by aggregating neighbors in other views.
  - Single-view embedding introduces the multi-head self-attention mechanism to learn node embedding.
  - Single-view completion aggregates the embeddings of reconstructed neighbors to learn the embedding of missing nodes.
  - Multi-view fusion uses multi-view CCA method to learn the unified multi-view embedding.

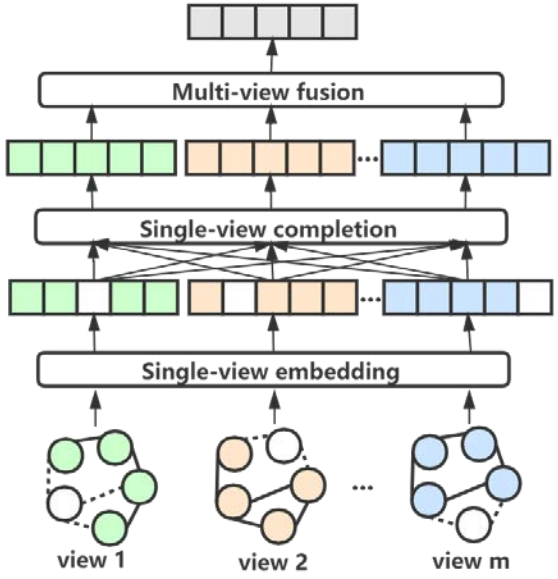

Fig1. The architecture of IMHE framework. Circles and squares represent nodes and node embeddings, and white circles and squares represent missing nodes and node embeddings.

Main Contributions

Table1: Micro-F1 scores of Node classification task

Methods	Training ratio					
	10%	15%	20%	30%	40%	50%
Line	0.497	0.522	0.536	0.566	0.576	0.572
M2V	0.526	0.553	0.595	0.637	0.643	0.645
Hin2vec	0.649	0.643	0.679	0.668	0.689	0.692
MNE	0.584	0.587	0.591	0.596	0.604	0.613
AHE	0.633	0.646	0.648	0.658	0.675	0.690
IMHE	**0.675**	**0.683**	**0.695**	**0.712**	**0.724**	**0.732**

Table2: Macro-F1 scores of Node classification task

Methods	Training ratio					
	10%	15%	20%	30%	40%	50%
Line	0.497	0.530	0.532	0.554	0.562	0.567
M2V	0.530	0.558	0.582	0.581	0.618	0.632
Hin2vec	0.642	0.646	0.651	0.662	0.674	0.686
MNE	0.575	0.573	0.580	0.585	0.594	0.611
AHE	0.648	0.657	0.658	0.675	0.682	0.679
IMHE	**0.677**	**0.679**	**0.689**	**0.700**	**0.708**	**0.716**

The results of node classification on the DBLP dataset show that the embedding performances of IMHE are better than the baselines, especially when the number of training samples is limited.