HAKG: Hierarchy-Aware Knowledge Gated Network for Recommendation

Yuntao Du[1], Xinjun Zhu[2], Lu Chen[1], Baihua Zheng[3], Yunjun Gao[1]

[1]*College of Computer Science, Zhejiang University, China*

[2]*School of Software Technology, Zhejiang University, China*

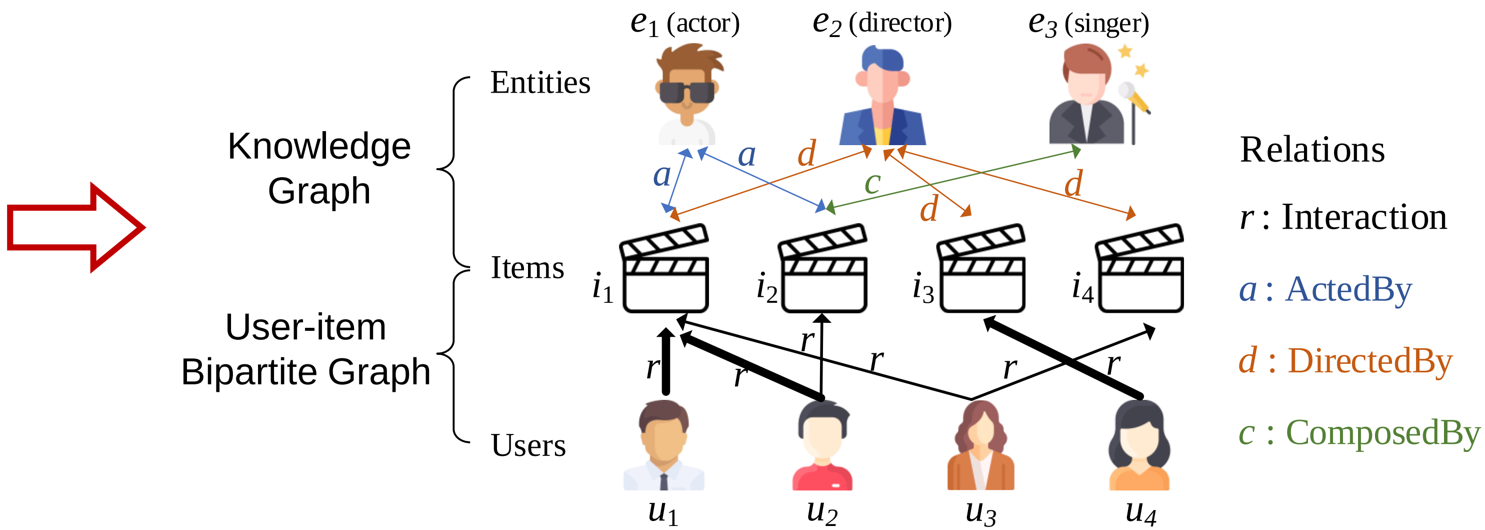
[3]*School of Computing and Information Systems, Singapore Management University, Singapore*

Outline

- **Motivation**
- Related Work
- Preliminaries
- Methods
- Experimental Evaluation
- Conclusions

Motivation

- **Data sparsity and cold-start issues**
  - KG provides rich semantic information about Items
- **Poor explainability of recommendation results**
  - KG improves the accuracy and interpretability



Knowledge-aware Recommendation!

Motivation

General steps of existing methods

a) Modeling both user-item interactions and KG in Euclidean space.

b) Considering the relations are of equal importance in KG.

c) Designing the KG-oriented item aggregation schemes.

Two important facts.

- Hierarchical structures and relations.
- High-order collaborative signals of items.

HAKG: Hierarchy-Aware Knowledge Gated Network for Recommendation

Outline

- Motivation
- **Related Work**
- Preliminaries
- Methods
- Experimental Evaluation
- Conclusions

Related Work

- **Embedding-based methods**
  - Directly embed entities and relations in KG to serve as item embedding in recommendation.
  - Drawbacks: fail to capture <u>high-order dependence</u> of user-item relations.

- **Path-based methods**
  - Define meta-paths in KG, and then connect items and users to discover long-range connectivity for recommendation.
  - Drawbacks: <u>time-consuming</u> and <u>poor generalization</u>.

- **Propagation-based methods**
  - Iteratively perform heterogeneous information aggregation mechanism from neighborhood nodes.

Outline

- Motivation
- Related Work
- **Preliminaries**
- Methods
- Experimental Evaluation
- Conclusions

Preliminaries

Problem Statement

- ***User-Item Bipartite Graph*** $\mathcal{G}_b = \{(u, i) \mid u \in \mathcal{U}, i \in \mathcal{I}\}$
  - $\mathcal{U}$ is the set of users, $\mathcal{I}$ is the set of items, (u, i) pair indicates that user u has interacted with item i.

- ***Knowledge Graph*** $\mathcal{G}_k = \{(h, r, t) \mid h, t \in \mathcal{E}, r \in \mathcal{R}\}$
  - Let $\mathcal{T}$ be the triplet set, $\mathcal{E}$ be a set of entities, and $\mathcal{R}$ be the relation set
  - Triplet $(h, r, t) \in \mathcal{T}$ means that there is a relation r between head entity h and tail entity t.
  - Assume all the items appear in KG as entities (i.e., $\mathcal{I} \subset \mathcal{E}$).

- ***Task Description***
  - Given a user-item graph $\mathcal{G}_b$ and a KG $\mathcal{G}_k$, our task of knowledge-aware recommendation is to predict how likely that a user would adopt an item that she has never engaged with.

Preliminaries

Hyperbolic geometry

- ***Poincaré Ball & Tangent Space***

$$\mathbb{B} = \left\{(x_1, \dots, x_n): x_1^2 + \dots + x_n^2 < \frac{1}{c}\right\}$$

- in $\mathbb{R}^n$. The tangent space $\mathcal{T}_{\mathbf{z}}\mathbb{B}$ at point z on $\mathbb{B}$ is a n-dimensional Euclidean space that best approximates $\mathbb{B}$ around z.

- ***Exponential Map & Logarithmic Map***

$$\exp_{\mathbf{z}}(\mathbf{x}) = \mathbf{z} \oplus \tanh\left(\frac{\| \mathbf{x} \|}{1 - \| \mathbf{z} \|^2}\right)\frac{\mathbf{x}}{\| \mathbf{x} \|}$$

$$\log_{\mathbf{z}}(\mathbf{y}) = (1 - \| \mathbf{z} \|^2) \cdot \tanh^{-1}(\| -\mathbf{z} \oplus \mathbf{y} \|)\frac{-\mathbf{z} \oplus \mathbf{y}}{\| -\mathbf{z} \oplus \mathbf{y} \|}$$

where $\oplus$ represents Möbius addition:

$$\mathbf{x} \oplus \mathbf{y} = \frac{(1 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \| \mathbf{y} \|^2)\mathbf{x} + (1 - \| \mathbf{x} \|^2)\mathbf{y}}{1 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \| \mathbf{x} \|^2 \| \mathbf{y} \|^2}$$

- the exponential map can map the tangent space $\mathcal{T}_{\mathbf{z}}\mathbb{B}$ to the hyperbolic space $\mathbb{B}$, and the logarithmic map maps $\mathbb{B}$ to $\mathcal{T}_z B$ conversely.

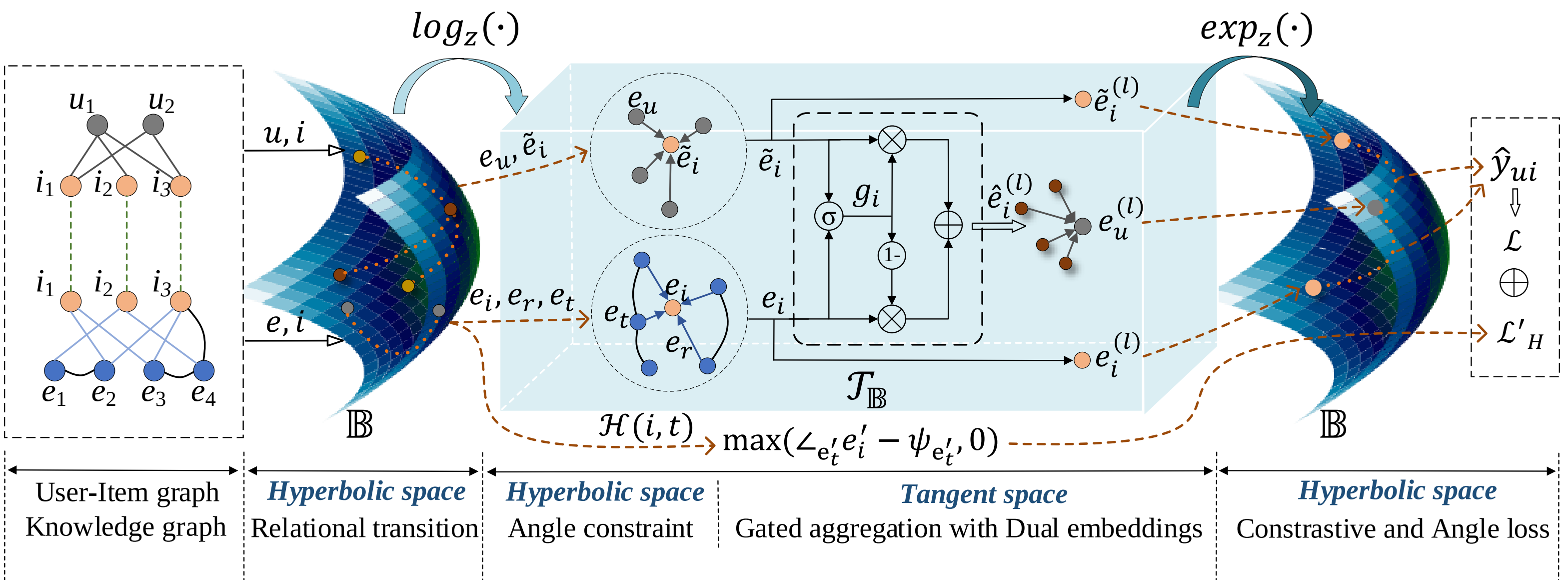
Outline

- Motivation
- Related Work
- Preliminaries
- **Methods**
- Experimental Evaluation
- Conclusions

Overview

HAKG framework

- **Component 1:** *Hierarchy-Aware Modeling*
  - Embed users and items as well as entities and relations in hyperbolic space.
  - Hyperbolic Relation-Transitive Aggregation.
  - Angle Constraint of Hierarchical triplets.
- **Component 2:** *Gated Aggregation with Dual embeddings*
  - Collaborative Aggregation for Items.
  - Information Gated Aggregation for Users.



Hierarchy-Aware Modeling

Hyperbolic Relation-Transitive Aggregation

- Integrate information from each connection (i, r, e) in hyperbolic space, and preserve the relation dependencies.

$$\mathbf{e}_i^{(l)} = \exp_{\mathbf{0}}\left(\frac{1}{|\mathcal{N}_i|}\sum_{(r,t)\in\mathcal{N}_i}\log_{\mathbf{e}_i^{(l-1)}}\left(\mathbf{e}_t^{(l-1)} \oplus \mathbf{e}_r\right)\right)$$

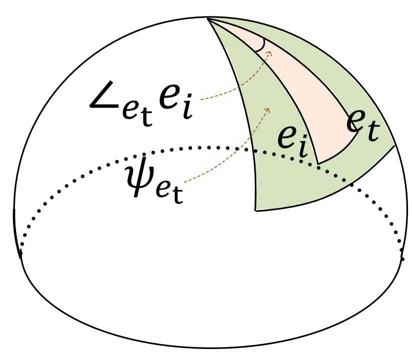
Angle Constraint of Hierarchical triplets

- Hyperbolic cone forms a nested structure in embedding space, and the width of the cone can indicate the "attribute" semantics of embeddings.

Cone width: $\psi_{\mathbf{x}} = \arcsin\left(K\frac{1-\|\mathbf{x}\|^2}{\|\mathbf{x}\|}\right)$

Angle: $\angle_{\mathbf{x}}\mathbf{y} = \arccos\left(\frac{\langle\mathbf{x},\mathbf{y}\rangle(1+\|\mathbf{x}\|^2)-\|\mathbf{x}\|^2(1+\|\mathbf{y}\|^2)}{\|\mathbf{x}\|\|\mathbf{x}-\mathbf{y}\|\sqrt{1+\|\mathbf{x}\|^2\|\mathbf{y}\|^2-2\langle\mathbf{x},\mathbf{y}\rangle}}\right)$

Angle loss: $\mathcal{L}'_{\mathrm{H}} = \sum_{(i,t)\in\mathcal{H}} max\left(\angle_{\mathbf{e}'_t}\mathbf{e}'_i - \psi_{\mathbf{e}'_t}, 0\right)$



Gated Aggregation with Dual embeddings

Collaborative Aggregation for Items

- Initialize new item representation to collaborative information.

$$\tilde{\mathbf{e}}_i^{(l)} = \exp_0\left(\frac{1}{|\tilde{\mathcal{N}}_i|}\sum_{(u)\in\tilde{\mathcal{N}}_i}\log_0\left(\mathbf{e}_u^{(l-1)}\right)\right)$$

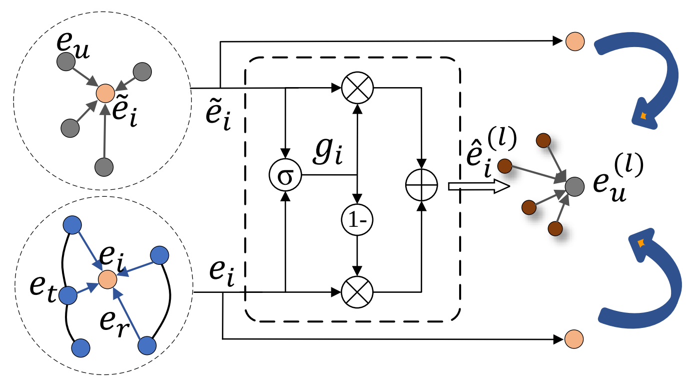
Information Gated Aggregation for Users

- A fusion gate to control the combination of two different types of semantic item representation.

$$\mathbf{g}_i^{(l)} = \sigma\left(W_1\log_0\left(\mathbf{e}_i^{(l)}\right) + W_2\log_0\left(\tilde{\mathbf{e}}_i^{(l)}\right)\right)$$

$$\hat{\mathbf{e}}_i^{(l)} = \exp_0\left(\mathbf{g}_i^{(l)}\cdot\mathbf{e}_i^{(l)} + \left(1-\mathbf{g}_i^{(l)}\right)\cdot\tilde{\mathbf{e}}_i^{(l)}\right)$$

$$\mathbf{e}_u^{(l)} = \exp_{\mathbf{0}}\left(\frac{1}{|\mathcal{N}_u|}\sum_{(i)\in\mathcal{N}_u}\log_{\mathbf{0}}\left(\hat{\mathbf{e}}_i^{(l-1)}\right)\right)$$



Outline

- Motivation
- Related Work
- Preliminaries
- Methods
- **Experimental Evaluation**
- Conclusions

Experimental Evaluation

Datasets

- Alibaba-iFashion, Yelp2018 and Last-FM are widely adopted in the state-of-the-art methods, and vary in terms of domain, size and sparsity.

		Alibaba-iFashion	Yelp2018	Last-FM
User-Item Interaction	#Users	114,737	45,919	23,566
	#Items	30,040	45,538	48,123
	#Interactions	1,781,093	1,185,068	3,034,796
Knowledge Graph	#Entities	59,156	90,961	58,266
	#Relations	51	42	9
	#Triplets	279,155	1,853,704	464,567

Competitors

- 9 state-of-the-art methods, including KG-free methods, embedding-based methods, propagation-based methods and hyperbolic-based methods.

Metrics

- Recall@20, NDCG@20.

Experimental Evaluation (Cont.)

Overall results of HAKG

	Alibaba-iFashion		Yelp2018		Last-FM	
	recall	ndcg	recall	ndcg	recall	ndcg
MF	0.1095	0.0670	0.0627	0.0413	0.0724	0.0617
CKE	0.1103	0.0676	0.0653	0.0423	0.0732	0.0630
UGRec	0.1006	0.0621	0.0651	0.0419	0.0730	0.0624
KGNN-LS	0.1039	0.0557	0.0671	0.0422	0.0880	0.0642
KGAT	0.1030	0.0627	<u>0.0705</u>	<u>0.0463</u>	0.0873	0.0744
CKAN	0.0970	0.0509	0.0646	0.0441	0.0812	0.0660
KGIN	<u>0.1147</u>	<u>0.0716</u>	0.0698	0.0451	<u>0.0978</u>	<u>0.0848</u>
Hyper-know	0.1057	0.0648	0.0685	0.0447	0.0948	0.0812
LKGR	0.1033	0.0612	0.0679	0.0438	0.0883	0.0675
HAKG	**0.1319***	**0.0848***	**0.0778***	**0.0501***	**0.1008***	**0.0931***
%Imp.	14.99%	15.43%	10.35%	8.21%	3.07%	9.79%

- HAKG consistently yields the best performance on all datasets.
- HAKG achieves significant improvement even over the strongest baselines *w.r.t.* ndcg@20 by 15.43%, 8.21%, and 9.79% in Alibaba-iFashion, Yelp2018, and Last-FM, respectively.

Experimental Evaluation (Cont.)

Ablation Study

Impact of angle loss and gated aggregation

	Alibaba-iFashion		Yelp2018		Last-FM	
	recall	ndcg	recall	ndcg	recall	ndcg
w/o A&G	0.1218	0.0799	0.0737	0.0458	0.0946	0.0872
w/o A	0.1272	0.0825	0.0763	0.0485	0.0963	0.0907
w/o G	0.1253	0.0817	0.0758	0.0471	0.0959	0.0894

- The absence of the angle constraint and gated aggregation dramatically degrades the performance.

Impact of hierarchical modeling

	Alibaba-iFashion		Yelp2018		Last-FM	
	recall	ndcg	recall	ndcg	recall	ndcg
Euclidean	0.1231	0.0798	0.0756	0.0484	0.0981	0.0916
PH-Relation	0.1317	0.0845	0.0772	0.0494	0.1001	0.0928
GH-Relation	0.1320	0.0845	0.0776	0.0498	0.1005	0.0929

- The performance degrades for all three datasets when we remove the hyperbolic geometry for HAKG.
- The hierarchical types of KG relations are not explicitly available.

Experimental Evaluation (Cont.)

Ablation Study

Impact of dual item embedding

	Alibaba-iFashion		Yelp2018		Last-FM	
	recall	ndcg	recall	ndcg	recall	ndcg
Single	0.1186	0.0755	0.0769	0.0492	0.0989	0.0904
Dual	0.1319	0.0847	0.0778	0.0501	0.1008	0.0931

- Discarding the collaborative item embeddings would consistently degrade the performance cross three datasets.

Impact of the number of layers L

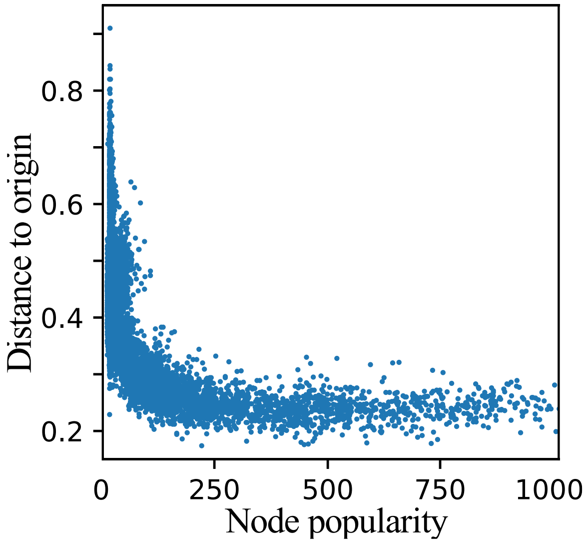
	Alibaba-iFashion		Yelp2018		Last-FM	
	recall	ndcg	recall	ndcg	recall	ndcg
HAKG-1	0.1313	0.0845	0.0766	0.0489	0.0972	0.0897
HAKG-2	0.1306	0.0831	0.0778	0.0501	0.0988	0.0913
HAKG-3	0.1319	0.0848	0.0774	0.0498	0.1008	0.0931

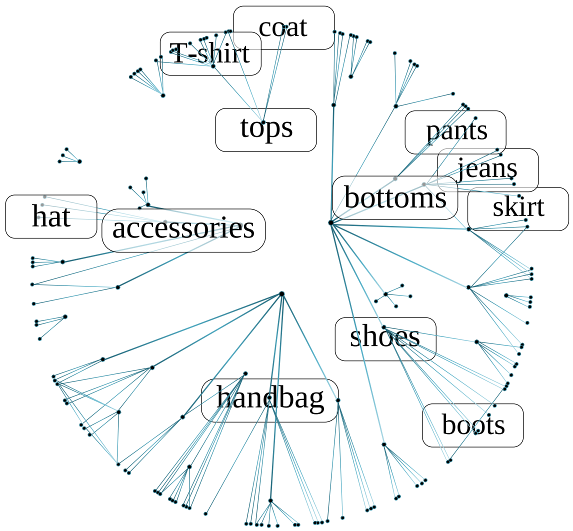
- HAKG is less sensitive to the model depth, compared with other propagation-based methods.

Experimental Evaluation (Cont.)

Hierarchies Visualization

- We first train HAKG with two-dimensional embeddings on the Alibaba-iFashion dataset, and separately analyze the hierarchies that exhibit in $\mathcal{G}_b$ and $\mathcal{G}_k$.





- The left figure indicates a clear exponential trend that distance to the origin increases exponentially for less popular items.

- The connectivities in the right figure show clear hierarchical relations between entities.

Outline

- Motivation
- Related Work
- Preliminaries
- Algorithms
- Experimental Evaluation
- **Conclusions**

Conclusions

- A new perspective. We present knowledge-aware recommendation by taking the hierarchy and high-order items' collaborative signals into consideration.
- Hierarchy-Aware. HAKG captures the underlying hierarchical structure of data in hyperbolic space, and characterize items with hierarchical relations in KG.
- Dual embeddings. HAKG employs dual item embeddings to separately encode items' collaborative signals and knowledge associations, and develops a gated mechanism to control discriminative signals towards the users' behavior patterns.
- Extensive experiments. Considerable experimental results demonstrate the superiority of HAKG.

Thank you !

Questions?

ytdu@zju.edu.cn

HAKG: Hierarchy-Aware Knowledge Gated Network for Recommendation

https://github.com/Scottdyt/HAKG

SIGIR 2022