Commonsense Knowledge Graph towards Super APP and Its Applications in Alipay

Xiaoling Zhang*, Binbin Hu*, Jun Chu, Zhiqiang Zhang, Gangnan Zhang,
Jun Zhou, Wenliang Zhong
Ant Group, Hangzhou, China

1 Background

Why Commonsense Knowledge Graph towards Super APP?

Alipay

A wide variety of services and contents provided through mini-programs

Transport

Shopping and Catering

Finance content

City service

Entertainment

Super App

- Open ecology
- Numerous scenarios
- Heterogeneous data

Challenges

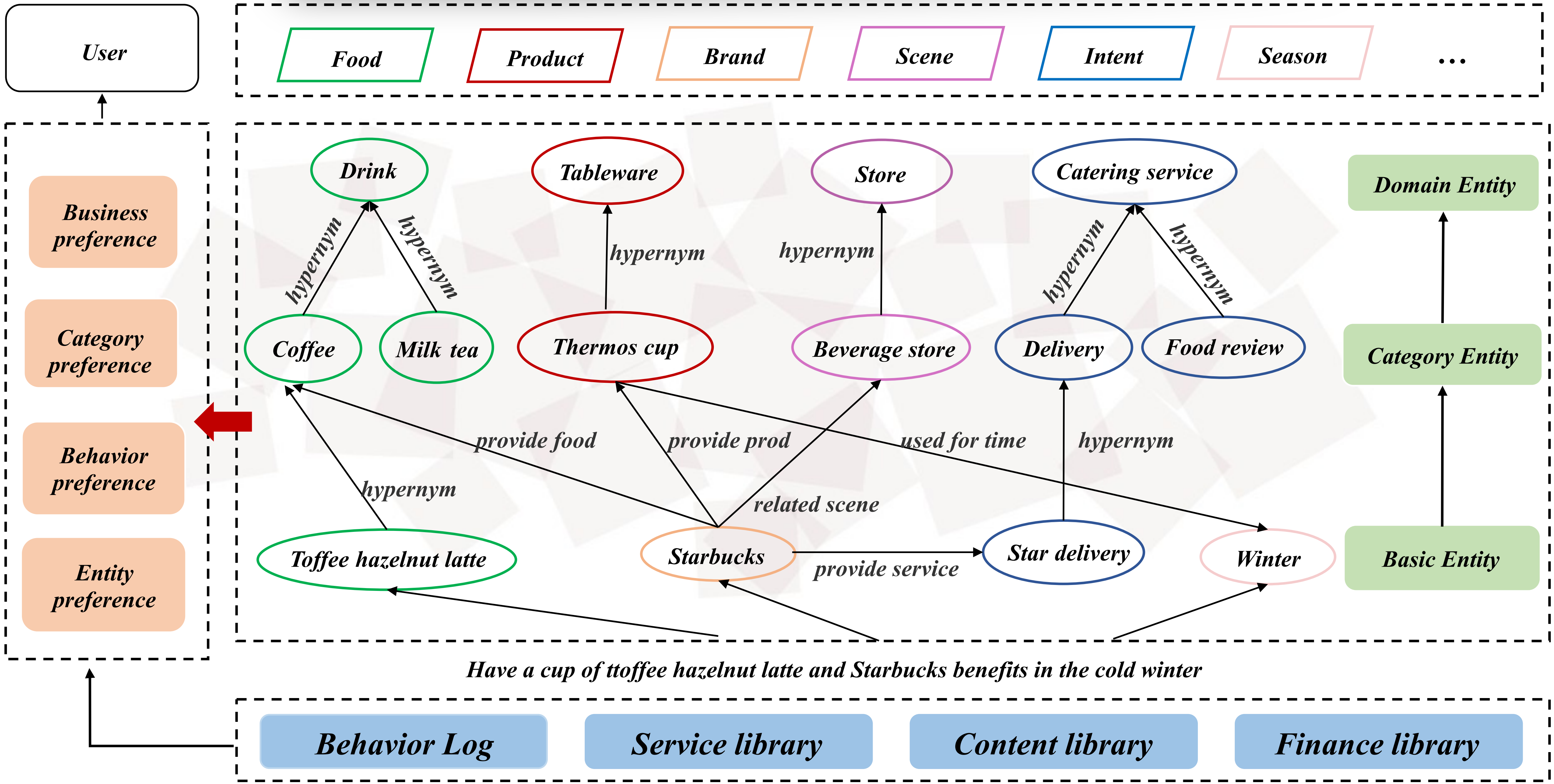
- Huge gap between scenarios
- Difficulty for effective information transfer and information

Solution A identical "Bridge"

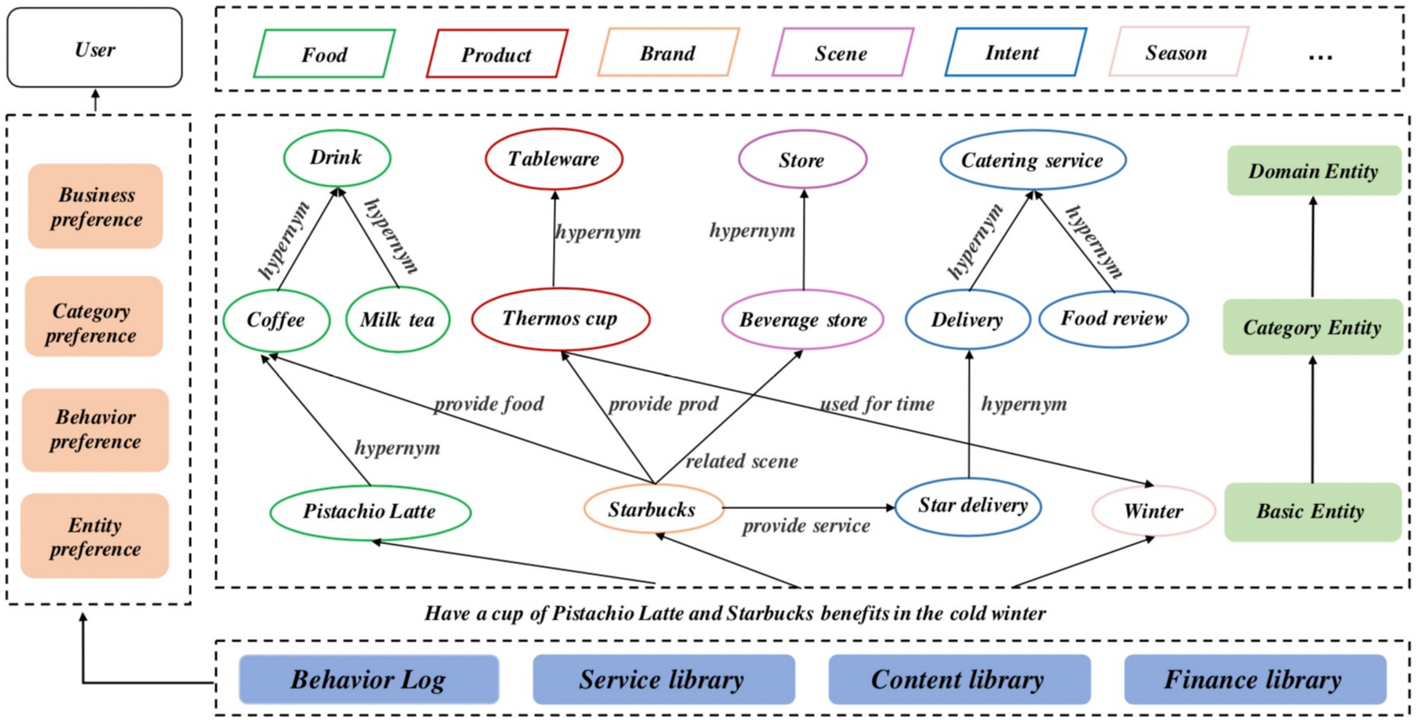
SupKG

2 SuperKG

An Overview



Construction



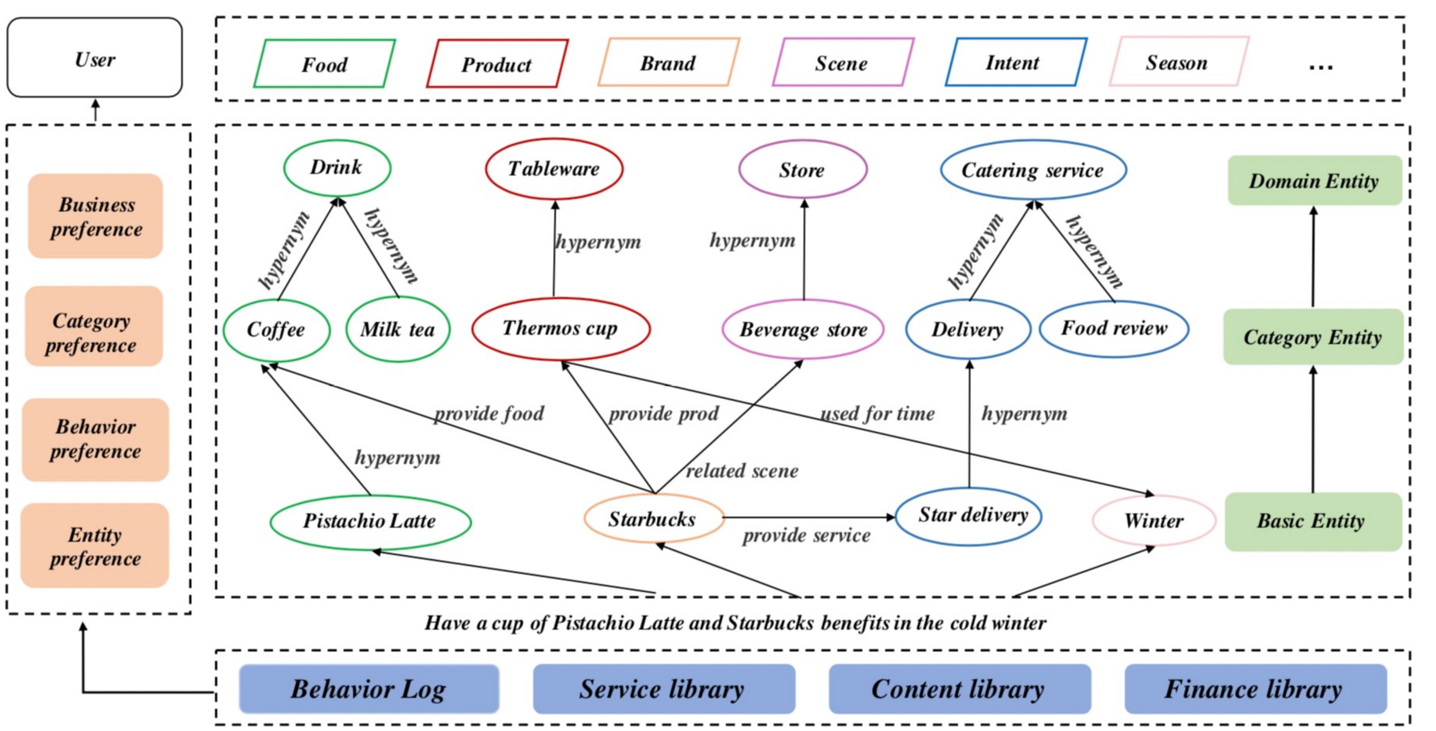
Entity Extraction

Fields with strong expertise	Crawlers and external public data	**Knowledge-related intent**	Directly introduced
Existing structural knowledge	Knowledge system with domain experts	**knowledge from user behaviors and services**	Entity extraction with sequence labeling

3

SuperKG

Construction

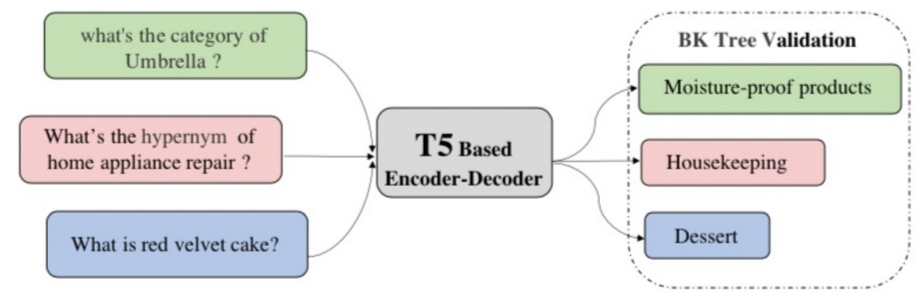


Entity Extraction

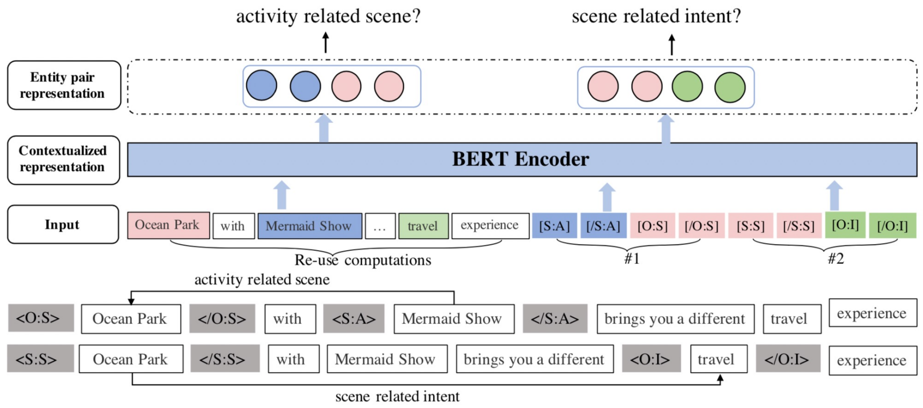
Fields with strong expertise	Crawlers and external public data
Knowledge-related intent	Directly introduced
Existing structural knowledge	Knowledge system with domain experts
knowledge from user behaviors and services	Entity extraction with sequence labeling

Relation Construction

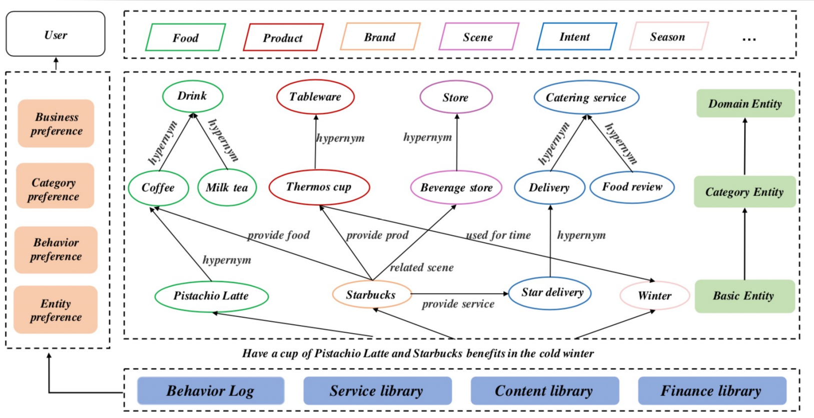
a) Hierarchical Relation



a) Spatio-temporal Relation



Challenge



Representation Learning

Challenges

- **Textual and structural information complement each other in KGE**

  Both structural match and semantic match matters

  $(<Chinese\_gold, brand\_provide\_service, Jewelry>)$

- **SupKG exhibit the skewed data distribution**

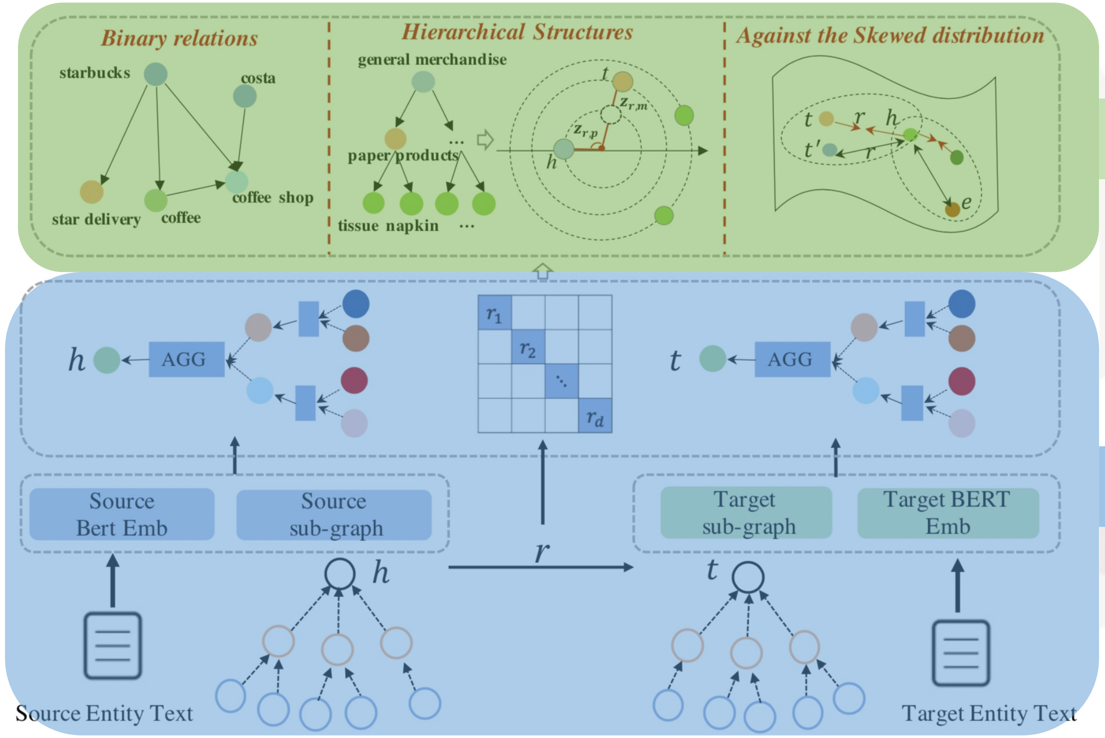
  More than 80% of entities with less than 5 degrees;

  KGE methods are easily misled towards remaining high-degree entities

- **Hierarchical structures are ubiquitous in SupKG**

  $(Pistachio\_Latte \rightarrow Coffee \rightarrow Drink)$

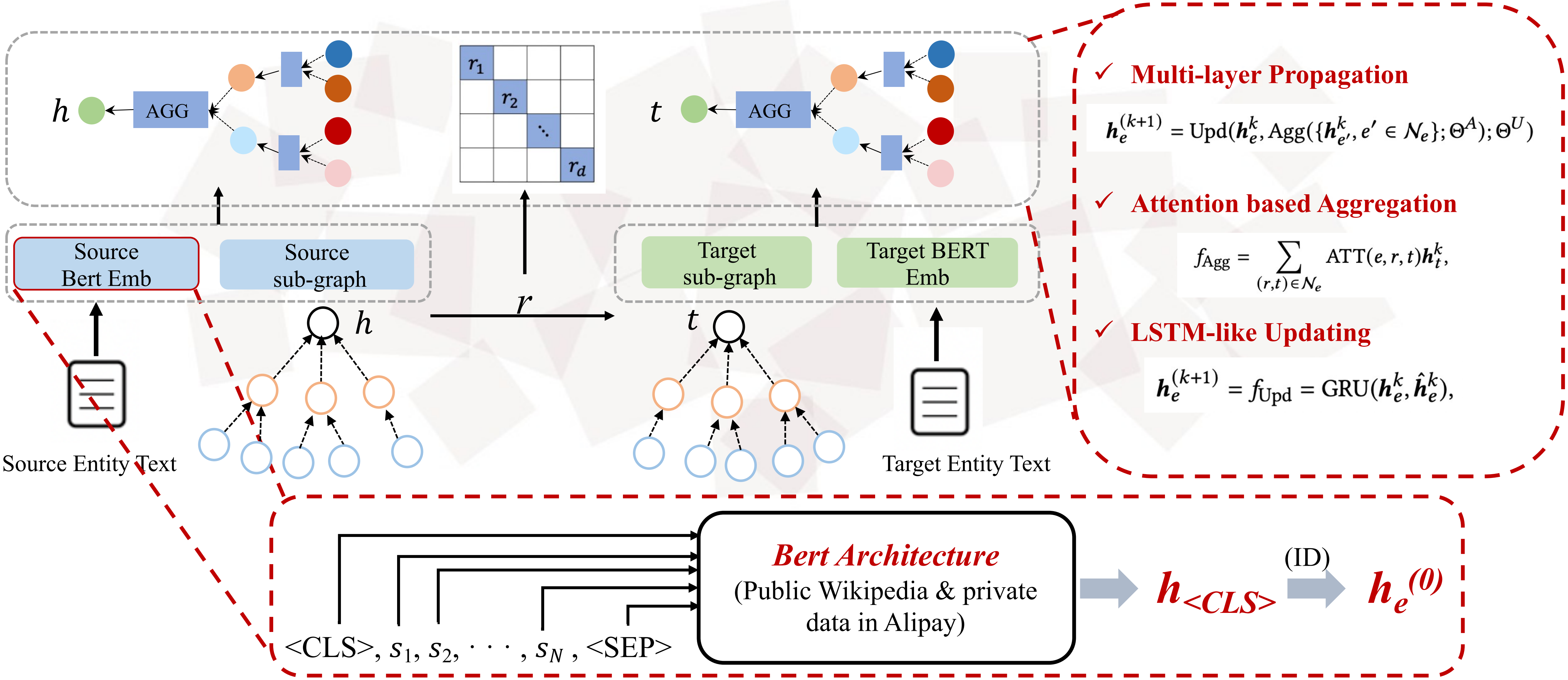
Representation Learning of SupKG



Fine-grained Relation Pattern Preservation in SupKG with a Multi-task Component

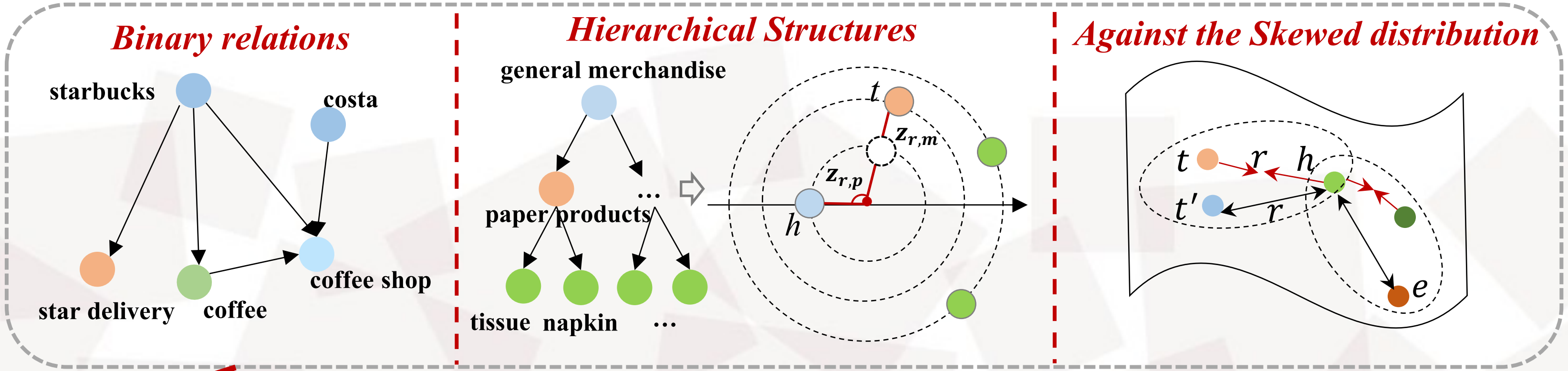
Bridging Language Representations with Knowledge Structure in SupKG

Bridging Language Representations with Knowledge Structure in SupKG



Representation Learning of Super KG

Fine-grained Relation Pattern Preservation in SupKG with a Multi-task Component



- ✓ **DistMult based interaction**

$$\boldsymbol{m}_{h,r,t} = \boldsymbol{z}_h^{\mathcal{D}^T} \mathbf{M}_r \odot \boldsymbol{z}_t^{\mathcal{D}},$$

- ✓ **MLP based scoring**

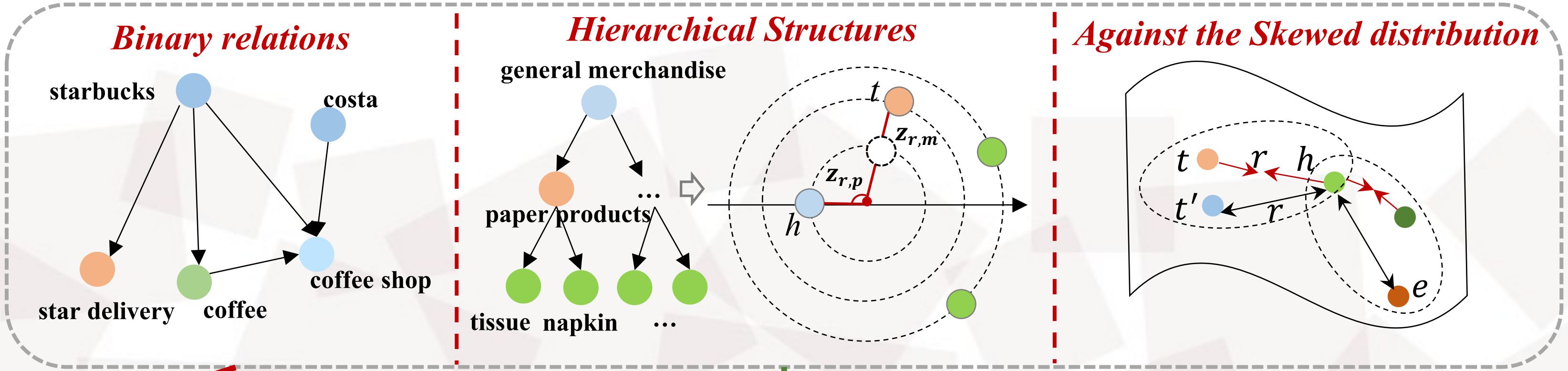
$$s^{\mathcal{D}}_{h,r,t} = f^K(\mathbf{W}_K \cdots f^1(\mathbf{W}_1 \boldsymbol{m}_{h,r,t} + \boldsymbol{b}_1) + \boldsymbol{b}_L),$$

- ✓ **CR loss with negatives**

$$\mathcal{L}^{\mathcal{D}} = - \sum_{<h,r,t>\sim\mathcal{G}, <h',r,t'>\sim\mathcal{G}'} \log(s^{\mathcal{D}}_{h,r,t}) + \log(1 - s^{\mathcal{D}}_{h',r,t'})$$

Representation Learning of Super KG

Fine-grained Relation Pattern Preservation in SupKG with a Multi-task Component



- ✓ **DistMult based interaction**

$$m_{h,r,t} = {z_h^{\mathcal{D}}}^T \mathbf{M}_r \odot z_t^{\mathcal{D}},$$

- ✓ **MLP based scoring**

$$s^{\mathcal{D}}_{h,r,t} = f^K(\mathbf{W}_K \cdots f^1(\mathbf{W}_1 m_{h,r,t} + b_1) + b_L),$$

- ✓ **CR loss with negatives**

$$\mathcal{L}^{\mathcal{D}} = -\sum_{<h,r,t>\sim\mathcal{G},<h',r,t'>\sim\mathcal{G}'} \log(s^{\mathcal{D}}_{h,r,t}) + \log(1 - s^{\mathcal{D}}_{h',r,t'})$$

- ✓ **Mapping as modulus and phase**

$$h_e^{(L)} \longrightarrow [z^{\mathcal{H}}_{e,m}; z^{\mathcal{H}}_{e,p}]$$

- ✓ **Distance in the polar coordinate system**

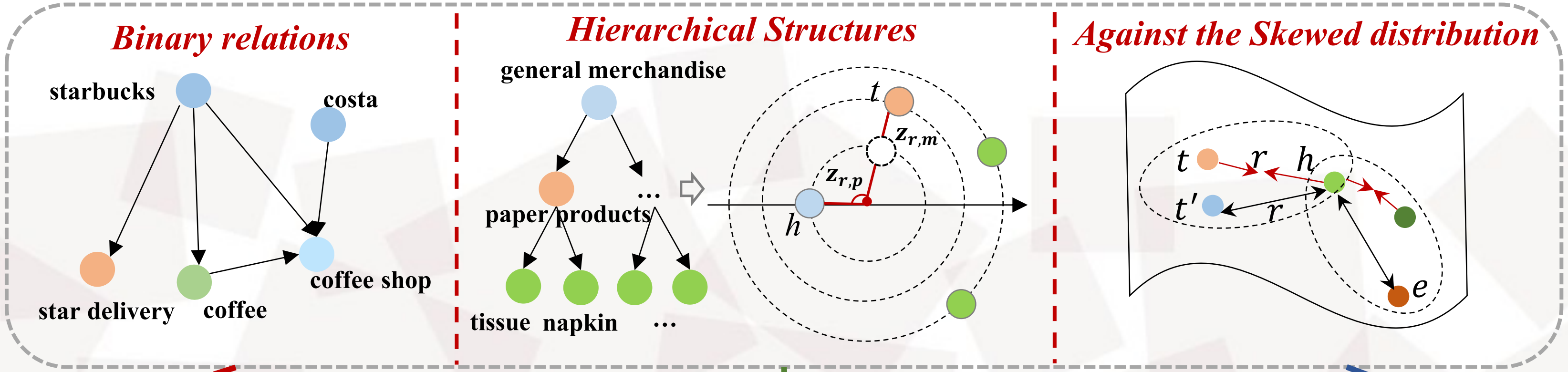
$$s^{\mathcal{H}}_{h,r,t} = ||z^{\mathcal{H}}_{h,m} \odot z^{\mathcal{H}}_{r,m} - z^{\mathcal{H}}_{t,m}||_2 + \lambda || \sin(z^{\mathcal{H}}_{h,p} + z^{\mathcal{H}}_{r,p} - z^{\mathcal{H}}_{t,p})||_1$$

- ✓ **Margin based loss**

$$\mathcal{L}^{\mathcal{H}} = - \sum_{<h,r,t>\sim\mathcal{G},<h',r,t'>\sim\mathcal{G}'} \log(\sigma(\gamma - s^{\mathcal{H}}_{h,r,t}) + \log(\sigma(s^{\mathcal{H}}_{h',r,t'} - \gamma)).$$

Representation Learning of Super KG

Fine-grained Relation Pattern Preservation in SupKG with a Multi-task Component



- ✓ **DistMult based interaction**

$$m_{h,r,t} = {z_h^{\mathcal{D}}}^T \mathbf{M}_r \odot z_t^{\mathcal{D}},$$

- ✓ **MLP based scoring**

$$s^{\mathcal{D}}_{h,r,t} = f^K(\mathbf{W}_K \cdots f^1(\mathbf{W}_1 m_{h,r,t} + b_1) + b_L),$$

- ✓ **CR loss with negatives**

$$\mathcal{L}^{\mathcal{D}} = -\sum_{\langle h,r,t\rangle\sim\mathcal{G},\langle h',r,t'\rangle\sim\mathcal{G}'} \log(s^{\mathcal{D}}_{h,r,t}) + \log(1 - s^{\mathcal{D}}_{h',r,t'})$$

- ✓ **Mapping as modulus and phase**

$$h_e^{(L)} \longrightarrow [z^{\mathcal{H}}_{e,m}; z^{\mathcal{H}}_{e,p}]$$

- ✓ **Distance in the polar coordinate system**

$$s^{\mathcal{H}}_{h,r,t} = ||z^{\mathcal{H}}_{h,m} \odot z^{\mathcal{H}}_{r,m} - z^{\mathcal{H}}_{t,m}||_2 + \lambda||\sin(z^{\mathcal{H}}_{h,p} + z^{\mathcal{H}}_{r,p} - z^{\mathcal{H}}_{t,p})||_1$$

- ✓ **Margin based loss**

$$\mathcal{L}^{\mathcal{H}} = -\sum_{\langle h,r,t\rangle\sim\mathcal{G},\langle h',r,t'\rangle\sim\mathcal{G}'} \log(\sigma(\gamma - s^{\mathcal{H}}_{h,r,t}) + \log(\sigma(s^{\mathcal{H}}_{h',r,t'} - \gamma)).$$

- ✓ **Entity-level contrast**

$$\mathcal{L}^C_e = -\sum_{\langle h,t\rangle\sim\mathcal{G}} \left(\log \frac{\exp(s(z_h^{(1)}, z_h^{(2)})/\tau)}{\sum_{e\in\mathcal{E}} \exp(s(z_h^{(1)}, z_e^{(2)})/\tau)} + \log \frac{\exp(s(z_t^{(1)}, z_t^{(2)})/\tau)}{\sum_{e\in\mathcal{E}} \exp(s(z_t^{(1)}, z_e^{(2)})/\tau)} \right),$$

- ✓ **Margin based loss**

$$\mathcal{L}^C_t = -\sum_{\langle h,r,t\rangle\sim\mathcal{G}} \log \frac{\exp(\phi(h,r,t)/\tau)}{\sum_{\langle h,r,t'\rangle\in\mathcal{N}} \exp(\phi(h,r,t')/\tau)},$$

Compared to AliCoCos

✓ Heterogeneous and unstructured data source

	AliCoCo	AliCoCo2	SupKG
# Entity	57, 125	163, 460	17, 343, 492
# Relation type	2	91	88
# Relation instance	131, 968	813, 315	103, 526, 390

SupKG has to deal with a lot of vanilla text of rather multiplex and heterogeneous behaviors, covering city service, traveling, entertainment, health care, and so on

Compared to AliCoCos

✓ **Heterogeneous and unstructured data source**

	AliCoCo	AliCoCo2	SupKG
# Entity	57, 125	163, 460	17, 343, 492
# Relation type	2	91	88
# Relation instance	131, 968	813, 315	103, 526, 390

SupKG has to deal with a lot of vanilla text of rather multiplex and heterogeneous behaviors, covering city service, traveling, entertainment, health care, and so on

✓ **Distinct emphasis from e-commerce for relation extraction**

(SupKG aims at answering “which service is needed at what time and where”)

Compared to AliCoCos

KDD2023

✓ Heterogeneous and unstructured data source

	AliCoCo	AliCoCo2	SupKG
# Entity	57, 125	163, 460	17, 343, 492
# Relation type	2	91	88
# Relation instance	131, 968	813, 315	103, 526, 390

SupKG has to deal with a lot of vanilla text of rather multiplex and heterogeneous behaviors, covering city service, traveling, entertainment, health care, and so on

✓ Distinct emphasis from e-commerce for relation extraction

(SupKG aims at answering "which service is needed at what time and where")

✓ More powerful representation capability

Incorporating language representations in the information propagation process is a more reasonable way for complementing textual and structural information.

CL empowers the representation

	Hit@5	Hit@10	Hit@15	Hit@20	MRR
AliCoCo2	0.2298	0.2844	0.3009	0.3128	0.1693
OURS	**0.2514**	**0.2958**	**0.3179**	**0.3403**	**0.1984**

Table 7: Performance comparison with Alicoco2 on triplets $< h, r, t >$ with semantic distance $\cos(\boldsymbol{h}, \boldsymbol{t}) \geq \delta$. Here $\boldsymbol{h}, \boldsymbol{t}$ are semantic embedding from BERT for head and tail entities, respectively.

More optimization-friendly in learning hierarchy

	Hit@5	Hit@10	Hit@15	Hit@20	MRR
Ours-HyL	0.3339	0.4406	0.4900	0.5246	0.2487
OUS	**0.3557**	**0.4446**	**0.4954**	**0.5310**	**0.2571**

Table 8: Hyperbolic loss (*i.e.,* "Ours-HyL") versus polar coordinate system (*i.e.,* "OUS") in the proposed framework.

Performance Comparison

➢ Overall performance

Methods	Hit@K K = 5	K = 10	K = 15	K = 20	MRR
TransE [4]	0.2346	0.3145	0.3652	0.4019	0.1555
TransR [19]	0.1751	0.2247	0.2622	0.2916	0.1410
TransD [17]	0.2483	0.3068	0.3407	0.3638	0.1834
TransH [29]	0.2488	0.3071	0.3419	0.3667	0.1828
ConvE [10]	0.1658	0.2229	0.2718	0.3133	0.1234
RESCAL [22]	0.2825	0.3238	0.3487	0.3681	0.2206
BLP [9]	0.2299	0.3115	0.3613	0.3981	0.1515
HAKE [33]	0.2169	0.2541	0.2732	0.2871	0.1669
RGCN [24]	0.0675	0.0962	0.1189	0.1414	0.0526
KGNN [15]	0.1477	0.2187	0.2699	0.3128	0.1053
AliCoCo2 [20]	0.3402	0.4395	0.4705	0.4926	0.2433
OURS	**0.3557**	**0.4446**	**0.4954**	**0.5310**	**0.2571**

Table 3: Quantitative comparison of different methods.

➢ Ablation study

TI	GS	HL	CL	Hit@K K = 5	K = 10	K = 15	K = 20	MRR
✓		✓	✓	0.2605	0.3032	0.3319	0.3548	0.2095
	✓	✓	✓	0.2576	0.3076	0.3408	0.3664	0.2014
✓	✓			0.2102	0.2897	0.3417	0.3822	0.1485
✓	✓	✓		0.3341	0.4236	0.4786	0.5186	0.2462
✓	✓		✓	0.2359	0.3254	0.3826	0.4251	0.1651
✓	✓	✓	✓	**0.3557**	**0.4446**	**0.4954**	**0.5310**	**0.2571**

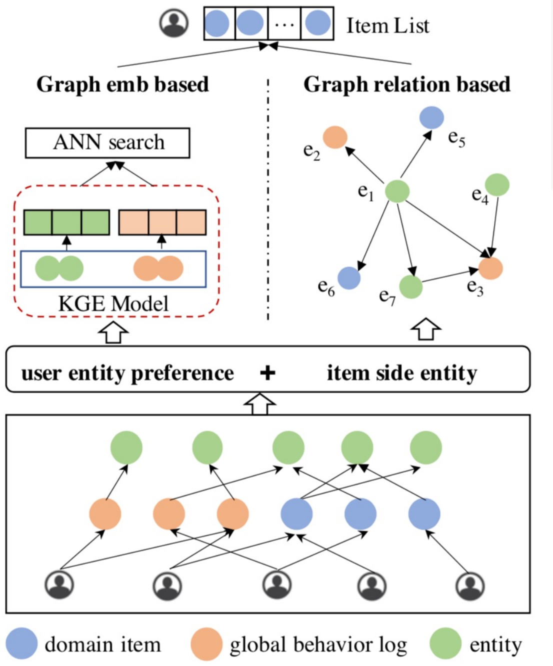
Table 4: Ablation studies of our proposal. "TI" means textual information; "GS" means graph structure; "HL" means hierarchy aware learning module; "CL" means contrastive learning module.

- ✓ Our complete representation framework consistently and significantly surpasses all the baselines across all metrics in the industrial knowledge graph
- ✓ In the ablation study, we could observe an apparent performance decrease once the corresponding component is removed, indicating all signals considered play critical roles in high-quality learning

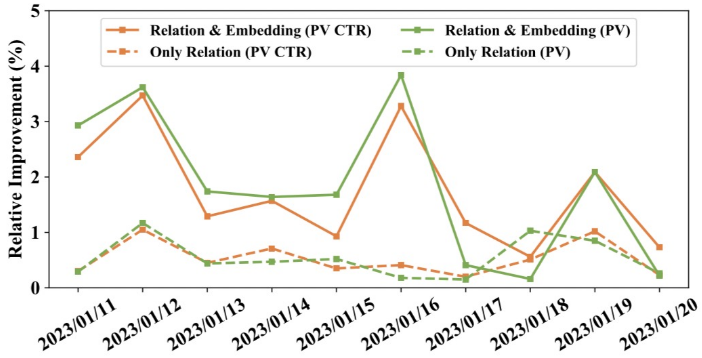
10 Application

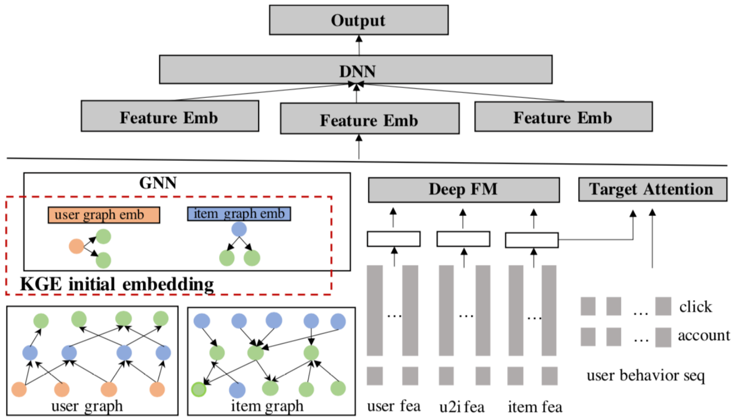
Source entity	Relation	Target entities retrieved
Musical instrument shop (乐器行)	scene_related_prod	Ukulele (乌克里里), Folk drum (民族鼓), Violin (小提琴), old records (老唱片) Percussion instrument (敲打乐器)
CBA	activity_need_prod	Basketball shoes (篮球鞋), Basketball (篮球), Sneaker (球鞋), Jersey (球衣)
Anti-inflammatory (消炎药)	prod_in_scene	Drugstore (药店), Fair-price drugstore (平价药店), TCM pharmacy (中药坊), Children's hospital (儿童医院), Community hospital (社区医院)
Bartending (调酒)	intent_related_food	Cocktail (鸡尾酒), Blueberry wine(蓝莓酒),Plum wine(青梅酒), Foreign wine(洋酒)
Family trip (亲子游)	intent_related_scene	Parent-child park(亲子乐园), Adventure park (探险乐园), wild animal park(野生动物园)

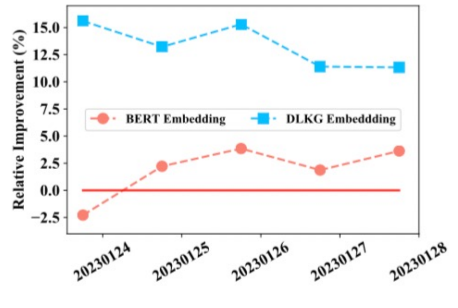
Supplementing 5 million potential knowledge through semantic and structural matching



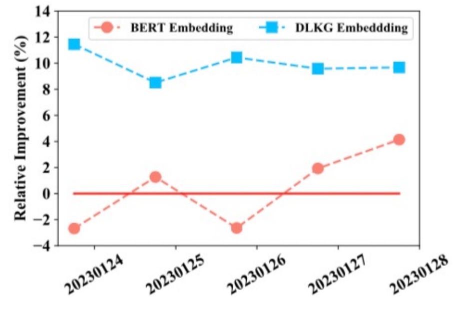
Integrating global behaviors with SupKG for matching







(a) PV Click



(b) PV CTR

Enhancing item and user representation in graph-based ranking

11 Conclusion

- ✓ We propose SupKG, a commonsense knowledge graph toward Super APP to help comprehensively characterize user behaviors across different business scenarios in a more fine-grained manner.
- ✓ We devise a novel representation learning framework, enabling various applications to draw support from effective representations of entities and relations from SupKG.
- ✓ We perform a series of offline/online to demonstrate that i) the proposed representation learning framework could substantially help supplement potential knowledge for SupKG; ii) the learned embedding and SupKG could well warm up various downstream by provide high-quality SupKG knowledge.

Thanks ! More details will be published in our paper

Commonsense Knowledge Graph towards Super APP and Its Applications in Alipay

ANT GROUP