Temporal Heterogeneous Interaction Graph Embedding for Next-Item Recommendation

Yugang Ji[1], Mingyang Yin[2], Yuan Fang[3], Hongxia Yang[2], Xiangwei Wang[2], Tianrui Jia[1] , Chuan Shi[1,4]

[1]Beijing University of Posts and Telecommunications, Beijing, China

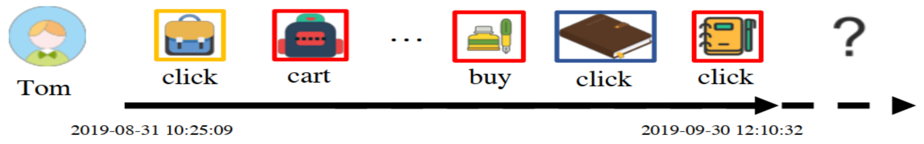
[2]Alibaba Group, Hangzhou, China

[3]Singapore Management University, Singapore

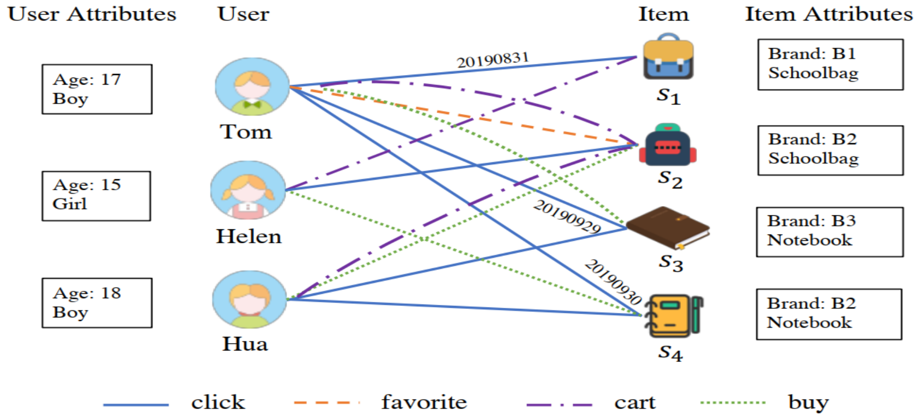
[4]Pengcheng Laboratory, Shenzhen, China

1 Overview

ECML PKDD 2020 GHENT, BELGIUM 14-18 SEPTEMBER



(a) The temporal evolution of interactions.



(b) An example of THIG.

C1: complex dynamic of historical habits and current demands

C2: temporal preference with heterogeneous interactions

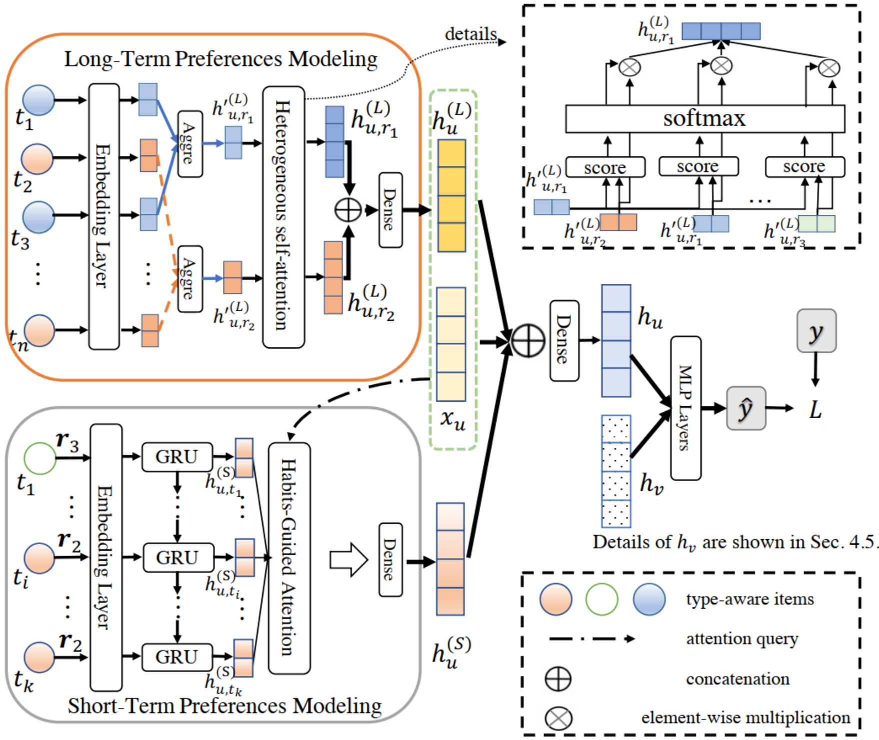
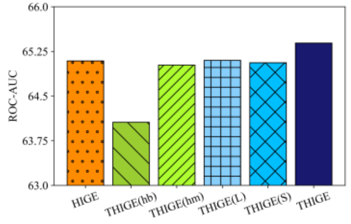
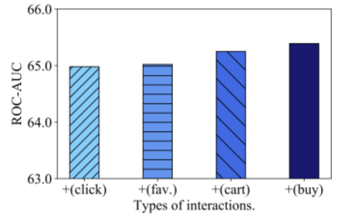


Table 2. Performance of next-item recommendation (with standard deviation). The best result is in bold while the second best is underlined. PR. denotes PR-AUC and ROC. denotes ROC-AUC.

Dataset	Metric	DIEN	STAMP	SHAN	M3R	MEIRec	GATNE	THIGE
Yelp	F1	39.52 (1.31)	40.37 (0.94)	40.17 (1.10)	33.49 (1.04)	<u>42.86</u> (0.44)	42.21 (0.96)	**43.77** (0.66)
	PR.	30.04 (0.37)	31.36 (1.23)	32.35 (1.10)	26.40 (0.92)	32.69 (0.54)	<u>33.39</u> (1.42)	**36.45** (1.66)
	ROC.	74.69 (0.57)	73.74 (1.15)	70.91 (1.14)	72.03 (1.33)	74.65 (0.23)	<u>76.15</u> (0.64)	**79.23** (0.80)
CT.	F1	25.70 (1.25)	21.42 (0.91)	26.25 (1.09)	<u>33.54</u> (1.67)	25.02 (0.98)	27.33 (0.50)	**37.17** (1.36)
	PR.	41.16 (0.22)	25.65 (0.44)	40.92 (1.09)	34.23 (0.95)	43.86 (0.42)	<u>44.74</u> (0.20)	**51.94** (0.43)
	ROC.	68.41 (0.34)	52.97 (0.52)	67.48 (1.06)	62.92 (0.89)	69.98 (0.35)	<u>71.22</u> (0.11)	**75.38** (0.33)
UB.	F1	<u>67.32</u> (3.45)	63.06 (1.51)	58.84 (7.83)	61.37 (2.20)	66.48 (1.16)	**67.81** (1.14)	67.19 (0.98)
	PR.	63.38 (0.19)	59.09 (0.22)	63.86 (4.76)	57.68 (0.03)	64.94 (0.15)	<u>65.42</u> (0.05)	**65.71** (0.09)
	ROC.	62.90 (0.23)	58.29 (0.40)	55.45 (3.98)	57.82 (0.09)	64.82 (0.16)	<u>65.06</u> (0.08)	**65.39** (0.06)





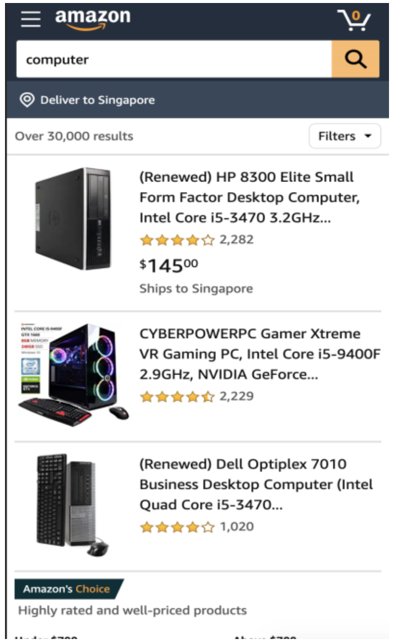
- The first use of THIGs for next-item recommendation.
- Propose THIGE to couple long- and short-term preferences of heterogeneous nature.
- Our THIGE performs better than SOTA, and the improvement of ROC-AUC is from **<u>0.51% to 5.84%.</u>**

CONTENTS

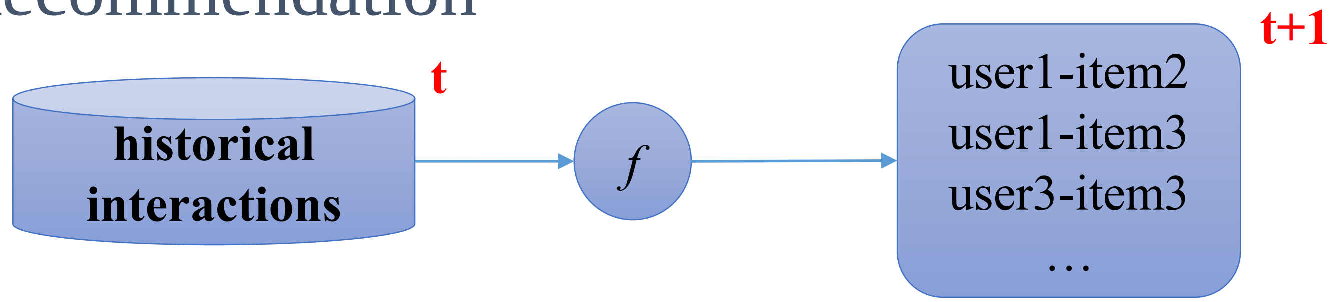
CONTENTS

1 Motivation

- E-commerce platforms have revolutionized our lifestyles.
- Recommender systems improve the shopping experience.

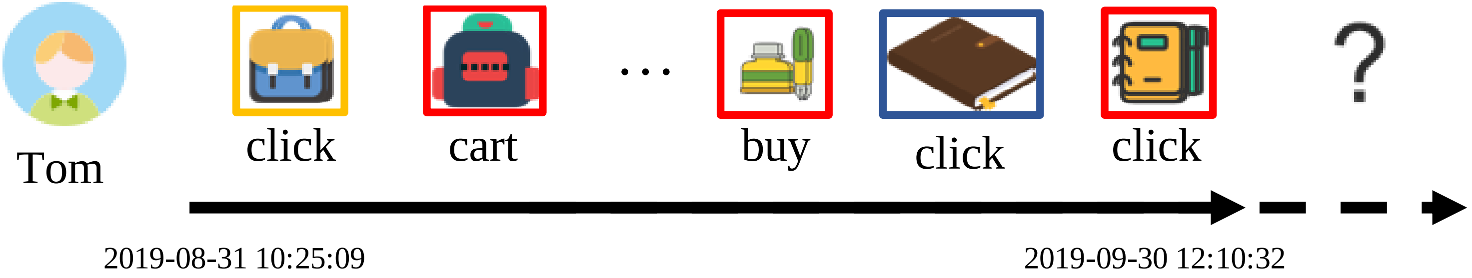


Next-Item Recommendation



Drawbacks of existing models (sequential)

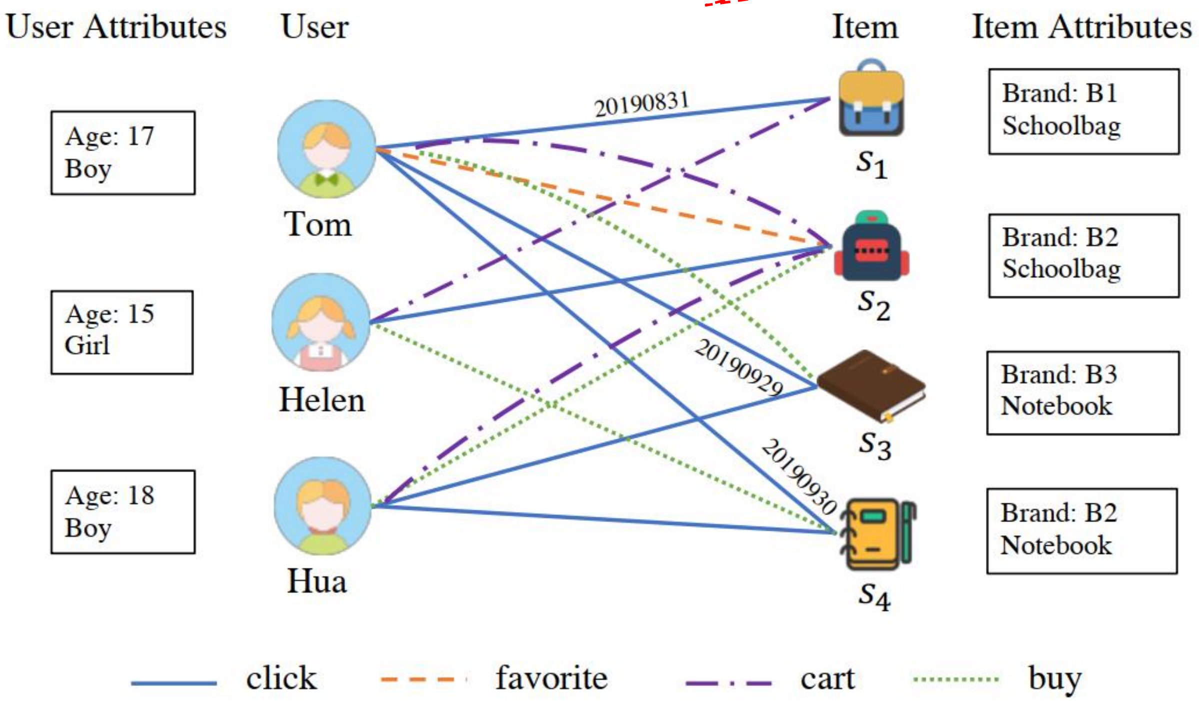
- unable to model historical habits
- ignoring the types of interactions
- cannot utilize structural information



Temporal heterogeneous interaction graph (THIG)

- multiple interactions with timestamps
- user attributes and item attributes

we focus on the problem of next-item recommendation on THIGs

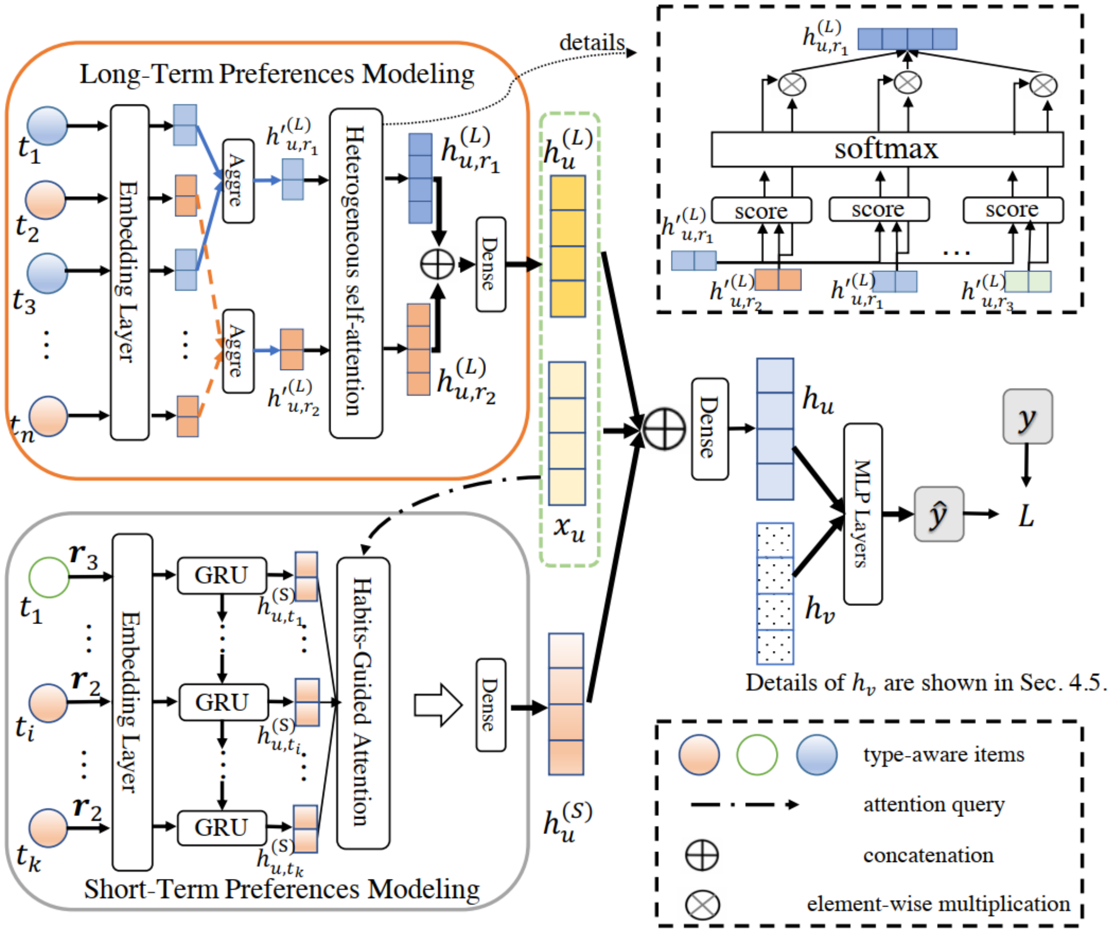


- *How to effectively model the complex dynamics, coupling both historical habits and evolving demands?*

- *How to make full use of the temporal heterogeneous interactions to model the preferences of different types?*

CONTENTS

Overall framework



- temporal embedding layer
- short-term preferences modeling
- long-term preferences modeling
- item preferences modeling
- optimization objects

Temporal embedding of an item *v* with timestamp *t*

- static component
- temporal component
- temporal embedding

projection matrix of type $\phi(v)$

$$\boldsymbol{x}_v = \boldsymbol{W}_{\phi(v)}\boldsymbol{a}_v$$

attributes

$$\oplus$$

projection matrix of buckets

timespan $\Delta t = T - t$

$$\boldsymbol{W}\xi(\Delta t)$$

mapping function to map Δt as a bucket id (one-hot feature)

$$\|$$

$$\boldsymbol{x}_{v,t}$$

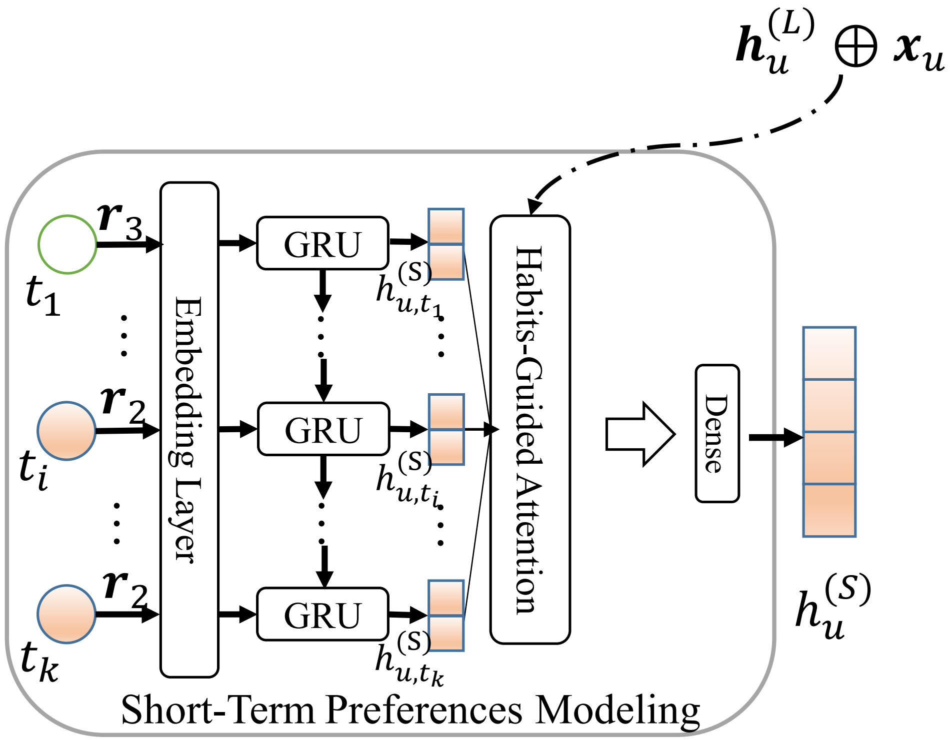
Evolving information based on GRU

$$\boldsymbol{x}_{v_i,t_i,r_i} = \boldsymbol{x}_{v_i,t_i} \oplus \boldsymbol{r}_i$$

$$\boldsymbol{h}_{u,t_i}^{(S)} = \mathrm{GRU}(\boldsymbol{x}_{v_i,t_i,r_i}, \boldsymbol{h}_{u,t_{i-1}}^{(S)}), \quad \forall 1 < i \leq k,$$

Habit-guided attention mechanism

$$\boldsymbol{h}_u^{(S)} = \sigma\left(W^{(S)} \cdot \sum_i a_{u,i} \boldsymbol{h}_{u,t_i}^{(S)} + b_s\right), \quad \forall 1 \leq i \leq k,$$



$$a_{u,i} = \frac{\exp\left([\boldsymbol{h}_u^{(L)} \oplus \boldsymbol{x}_u]^T \boldsymbol{W}_a \boldsymbol{h}_{u,t_i}^{(S)}\right)}{\sum_{j=1}^{k} \exp\left([\boldsymbol{h}_u^{(L)} \oplus \boldsymbol{x}_u]^T \boldsymbol{W}_a \boldsymbol{h}_{u,t_j}^{(S)}\right)},$$

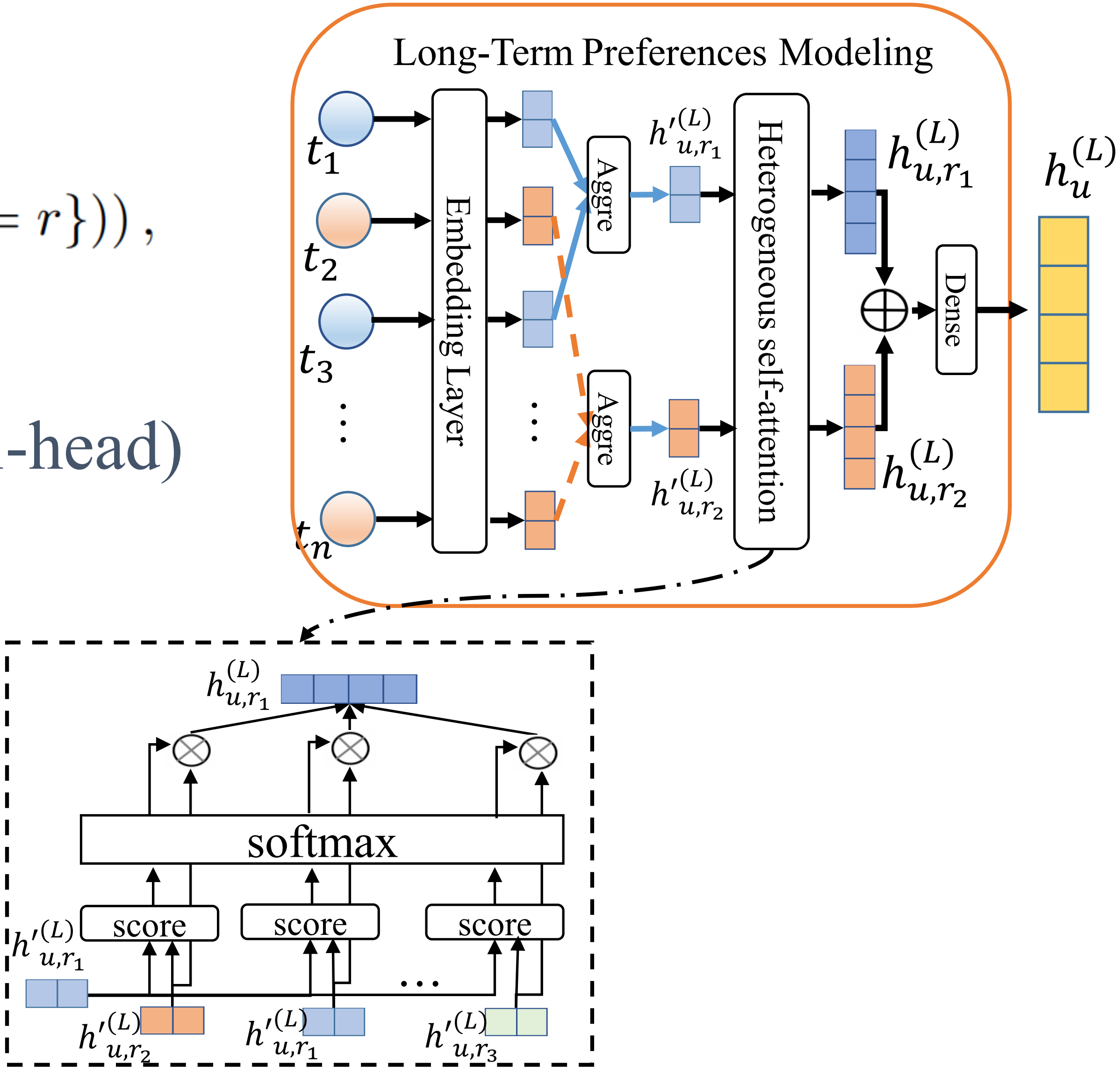
Heterogeneous aggregator

$$\boldsymbol{h}'^{(L)}_{u,r} = \sigma\left(\boldsymbol{W}_r \cdot \mathrm{aggre}(\{\boldsymbol{x}_{v_i,t_i} \mid 1 \leq i \leq n, r_i = r\})\right),$$

Heterogeneous self-attention (multi-head)

$$\boldsymbol{h}^{(L)}_{u,r} = \sum_{r' \in \mathcal{R}} \left(\frac{\exp\left(\boldsymbol{Q}^T_{u,r}\boldsymbol{K}_{u,r'}/\sqrt{d_a}\right)}{\sum_{r'' \in \mathcal{R}} \exp\left(\boldsymbol{Q}^T_{u,r}\boldsymbol{K}_{u,r''}/\sqrt{d_a}\right)} \boldsymbol{V}_{u,r'} \right),$$

$$\boldsymbol{h}^{(L)}_{u} = \sigma\left(\boldsymbol{W}^{(L)}(\oplus_{r \in \mathcal{R}} \boldsymbol{h}^{(L)}_{u,r}) + b_l\right),$$



Final embedding of users and items

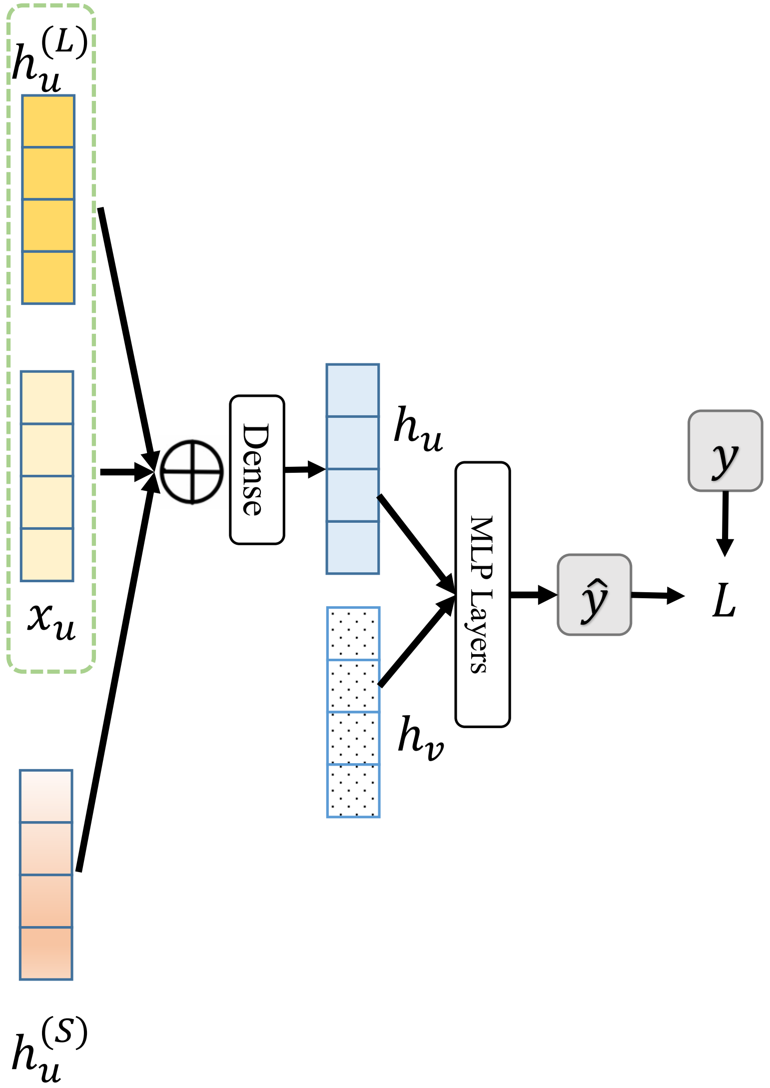
$$\boldsymbol{h}_u = \sigma(\boldsymbol{W_u}[\boldsymbol{x}_u \oplus \boldsymbol{h}_u^{(S)} \oplus \boldsymbol{h}_u^{(L)}] + b_u),$$

$$\boldsymbol{h}_v = \sigma(\boldsymbol{W}_v[\boldsymbol{x}_v \oplus \boldsymbol{h}_v^{(L)}] + b_v),$$

Optimization objective

$$\hat{y}_{u,v} = \text{sigmoid}(\text{MLP}(\boldsymbol{h}_u \oplus \boldsymbol{h}_v)),$$

$$L = -\sum_{\langle u,v \rangle} (1 - y_{u,v}) \log(1 - \hat{y}_{u,v}) + y_{u,v} \log(\hat{y}_{u,v}),$$



CONTENTS

Datasets

Table 1. Description of datasets.

Dataset	**Yelp**	**CloudTheme**	**UserBehavior**
# User	103,569	144,197	533,974
# Item/Business	133,502	272,334	4,152,242
# Interaction	1,889,132	1,143,567	122,451,055
# Interaction type	2	2	4
# Training instance	611,568	865,182	3,203,844
(Training time span)	5 years	2 weeks	1 weeks
# Test instance	108,408	216,295	800,961
(Test time span)	next one quarter	next day	next day

Baselines

- DIEN & STAMP
- SHAN & M3R
- MEIRec & GATNE

Tasks

- Comparison with baselines
- Comparison of model variants
- Analysis of key factors

Comparison with baselines

Table 2. Performance of next-item recommendation (with standard deviation). The best result is in bold while the second best is underlined. PR. denotes PR-AUC and ROC. denotes ROC-AUC.

Dataset	Metric	DIEN	STAMP	SHAN	M3R	MEIRec	GATNE	THIGE
Yelp	F1	39.52 (1.31)	40.37 (0.94)	40.17 (1.10)	33.49 (1.04)	<u>42.86</u> (0.44)	42.21 (0.96)	**43.77** (0.66)
	PR.	30.04 (0.37)	31.36 (1.23)	32.35 (1.10)	26.40 (0.92)	32.69 (0.54)	<u>33.39</u> (1.42)	**36.45** (1.66)
	ROC.	74.69 (0.57)	73.74 (1.15)	70.91 (1.14)	72.03 (1.33)	74.65 (0.23)	<u>76.15</u> (0.64)	**79.23** (0.80)
CT.	F1	25.70 (1.25)	21.42 (0.91)	26.25 (1.09)	<u>33.54</u> (1.67)	25.02 (0.98)	27.33 (0.50)	**37.17** (1.36)
	PR.	41.16 (0.22)	25.65 (0.44)	40.92 (1.09)	34.23 (0.95)	43.86 (0.42)	<u>44.74</u> (0.20)	**51.94** (0.43)
	ROC.	68.41 (0.34)	52.97 (0.52)	67.48 (1.06)	62.92 (0.89)	69.98 (0.35)	<u>71.22</u> (0.11)	**75.38** (0.33)
UB.	F1	<u>67.32</u> (3.45)	63.06 (1.51)	58.84 (7.83)	61.37 (2.20)	66.48 (1.16)	**67.81** (1.14)	67.19 (0.98)
	PR.	63.38 (0.19)	59.09 (0.22)	63.86 (4.76)	57.68 (0.03)	64.94 (0.15)	<u>65.42</u> (0.05)	**65.71** (0.09)
	ROC.	62.90 (0.23)	58.29 (0.40)	55.45 (3.98)	57.82 (0.09)	64.82 (0.16)	<u>65.06</u> (0.08)	**65.39** (0.06)

- Compared with the best competitors
  - 4.04% improvement on Yelp
  - 5.84% improvement on CouldTheme
  - 0.51% improvement on UserBehavior

(ROC-AUC)

- Perform better than sequential models

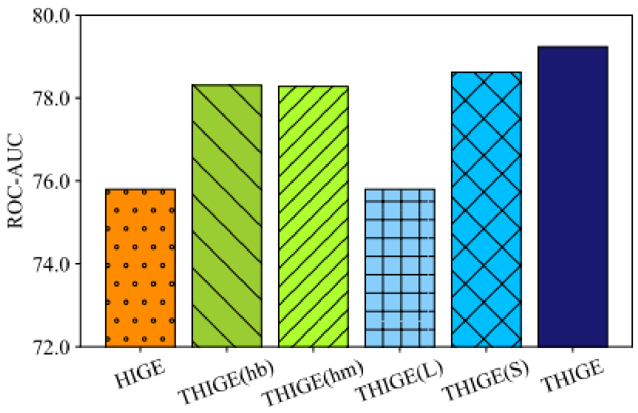
  effective overall preference modeling

  effective heterogeneity modeling
- Perform better than GNN-based models

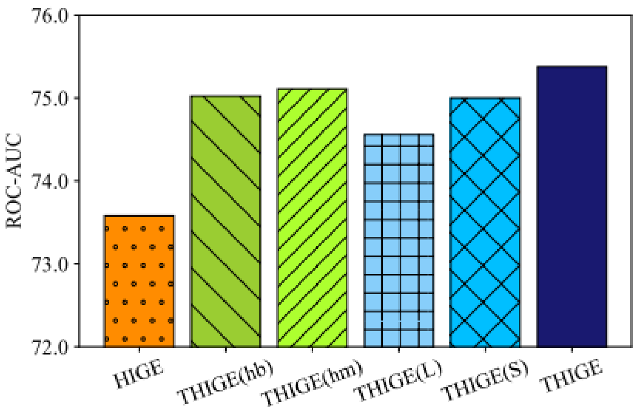
  effective temporal information modeling

Comparison of model variants

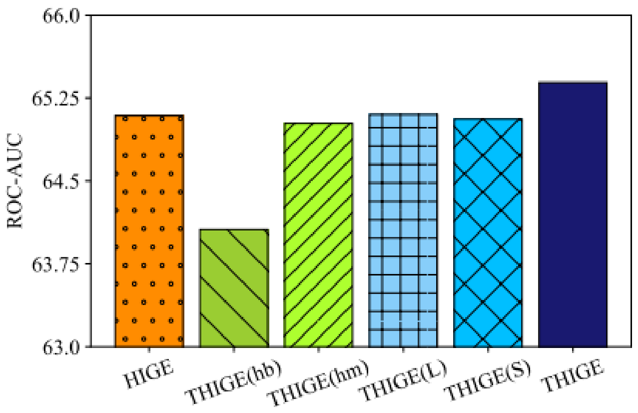
- Attention effect: THIGE(hm) & THIGE(hb) ✔
- Range of preferences: THIGE(L) & THIGE(S) ✔
- Temporal effect: HIGE ✔



(a) Yelp



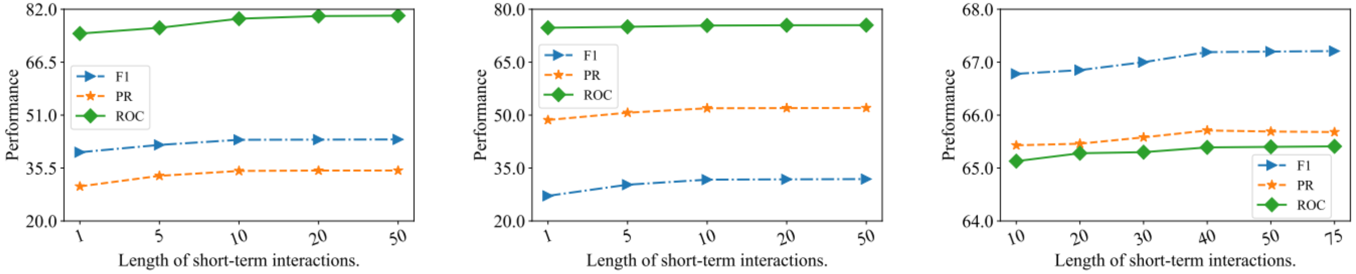
(b) CloudTheme



(c) UserBehavior

ECML PKDD 2020 GHENT, BELGIUM 14-18 SEPTEMBER

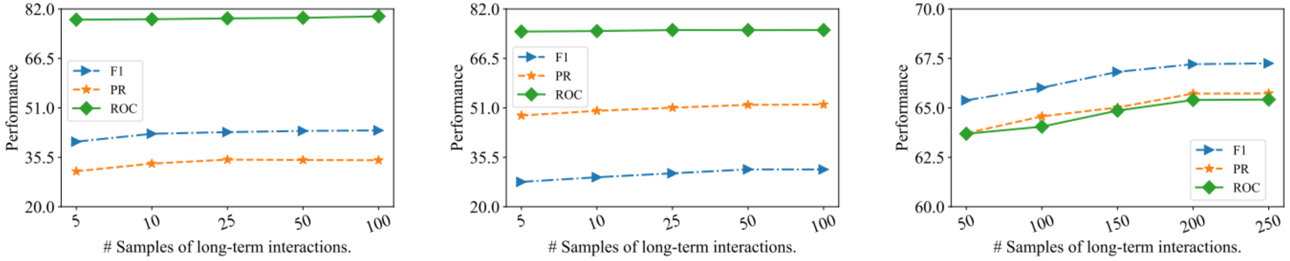
lengths of short-term interactions



(a) Yelp (b) CloudTheme (c) UserBehavior

Fig. 4. Analyzing the length of short-term interactions in THIGE.

sample of long-term interactions

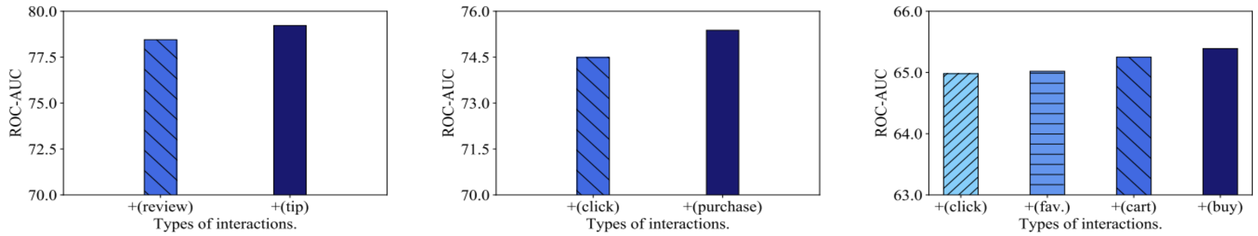


(a) Yelp (b) CloudTheme (c) UserBehavior

Fig. 5. Analyzing the samples of long-term interactions in THIGE.

- the performance of THIGE is continuously improved
- simply extending the length of the short term does not work
- too many interactions may be abnormal and introduce noise

different types of interactions

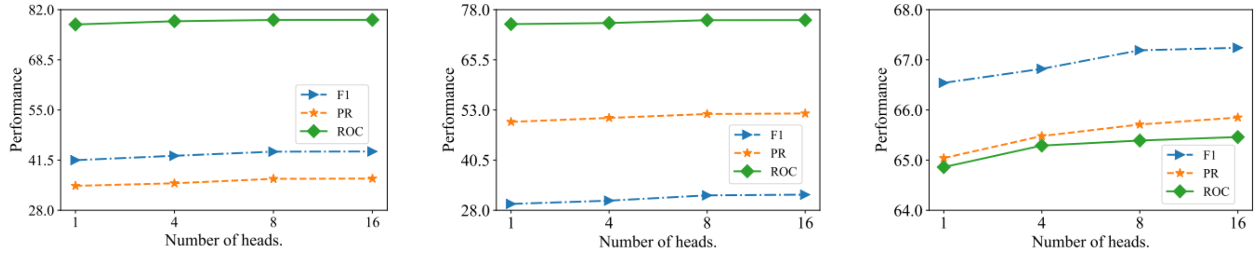


(a) Yelp (b) CloudTheme (c) UserBehavior

Fig. 6. Analyzing the types of interactions in THIGE.

- the performance gradually improves
- different types of interactions cannot be treated independently

number of heads



(a) Yelp (b) CloudTheme (c) UserBehavior

Fig. 7. Analyzing the number of heads in THIGE.

- h = 8 is a generally suitable and robust choice

CONTENTS

- we study the problem of representation learning on THIGs for next-item recommendation.
- To make full use of dynamic and heterogeneous information, we propose the THIGE to model short- and long-term preferences through habit-guided and heterogeneous self-attention mechanisms.
- Experimental results on three real-world datasets demonstrate the effectiveness of our proposed model.

Thank you !

Q&A