MS-HGAT: Memory-enhanced Sequential Hypergraph Attention Network for Information Diffusion Prediction

Ling Sun, Yuan Rao*, Xiangbo Zhang, Yuqian Lan, Shuanghe Yu
Xi'an Key Laboratory of Social Intelligence and Complexity Data Processing,
School of Software Engineering, Xi'an Jiaotong University, China
Shaanxi Joint Key Laboratory for Artifact Intelligence, China
{sunling, nick1001, Yuqian_Lan_xjtu, yushuanghe1997}@stu.xjtu.edu.cn, raoyuan@mail.xjtu.edu.cn

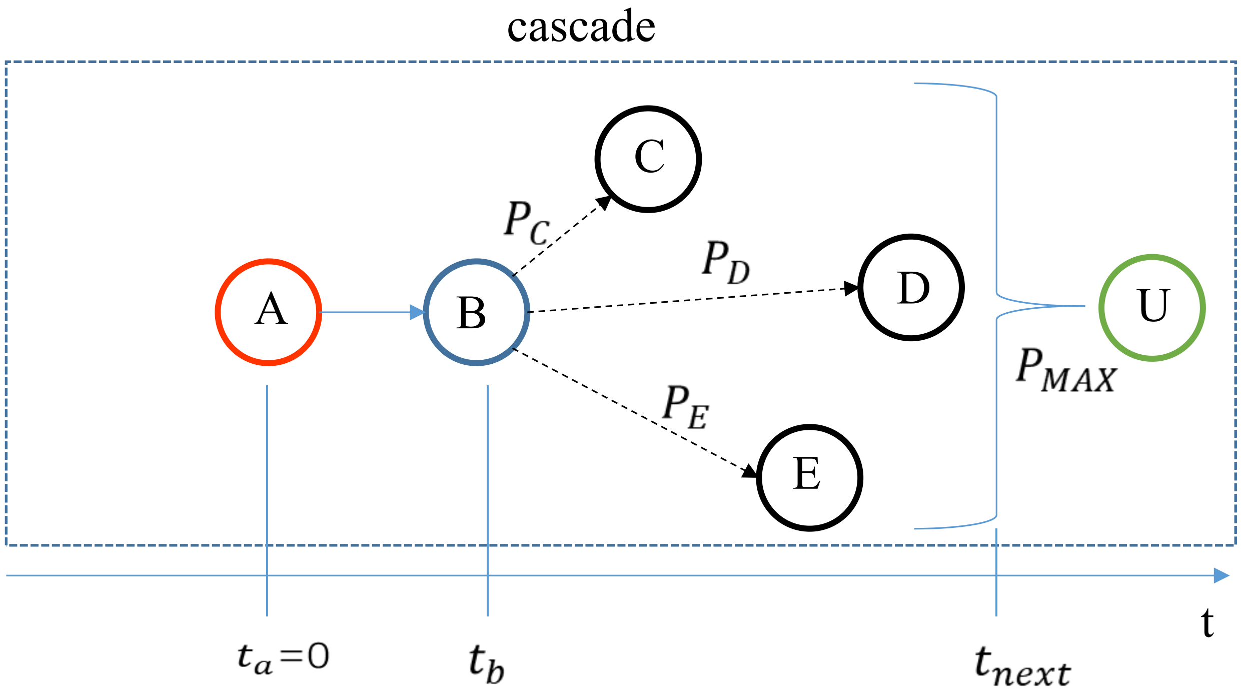
AAAI 2022
Code: https://github.com/slingling/MS-HGAT

Reported by Nengqiang Xiang

Introduction

- Using the friendship network is insufficient to describe users' interactive preferences.
  Ignored the dynamic connections between users and cascades

- Taking advantages of both the friendships and diffusion interactions of users.
  Proposing a sequential hypergraph attention network to learn the short-term interactions between users and cascades



Method

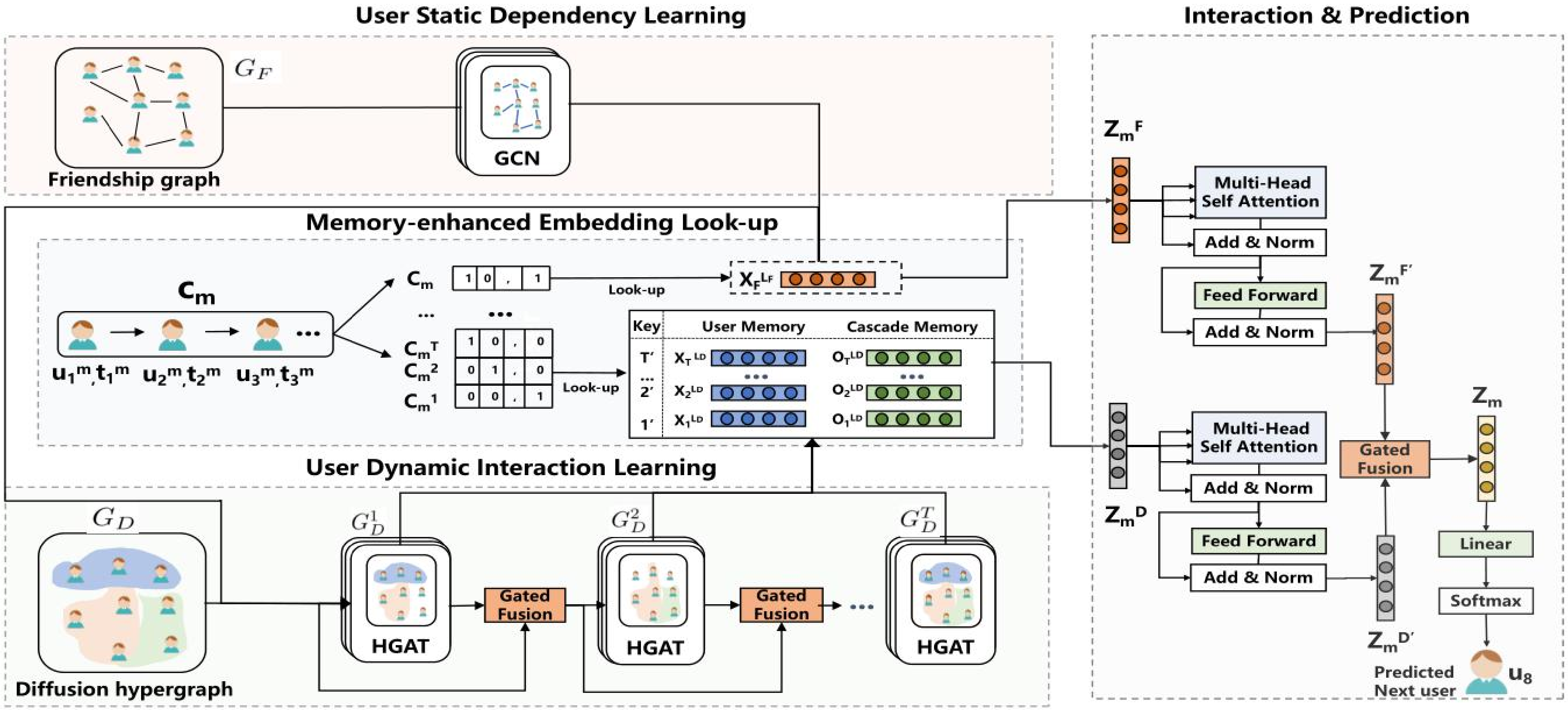


Figure 1: Four modules of MS-HGAT: 1) users' friendships are learned by GCN in user static dependency learning module; 2) user dynamic interaction learning module obtains interaction-based user and cascade embeddings through sequential HGATs; 3)memory-enhanced embedding look-up refers to finding the corresponding representation vectors in the static user representation and the dynamic memory block; 4)in interaction & prediction module, self-attention mechanisms are used to efficiently interact features in cascade, finally, probability of infection of candidates is calculated by Softmax function.

Method

Problem Formulation

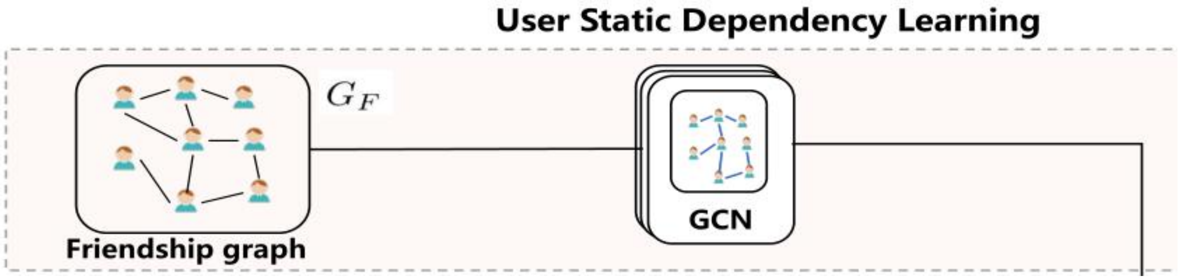
User set: U={$u_1, u_2, \ldots, u_n$}

friendship graph: $G_F=(U, E)$

historical diffusion cascades: $C = \{c_1, c_2, \cdots, c_M\}$ $|C| = M$

diffusion hypergraphs: $G_D=\{G_D^t \mid t = 1,2, \ldots, T\}$ $G_D^t=(U^t, \mathcal{E}^t)$

diffusion sequence: $c_m=\{(u_i^m, t_i^m) \mid u_i^m \in U\}$



User Static Dependency Learning

$$\mathbf{X}_F^{l+1} = \sigma\left(\tilde{\mathbf{D}}_\mathbf{F}^{-\frac{1}{2}} \tilde{\mathbf{A}}_\mathbf{F} \tilde{\mathbf{D}}_\mathbf{F}^{-\frac{1}{2}} \mathbf{X}_F^l \mathbf{W}_F\right) \tag{1}$$

$$\mathbf{X}_F^0 \in \mathbb{R}^{N \times d}$$

$$\tilde{A}_F = A_F + I$$

Method

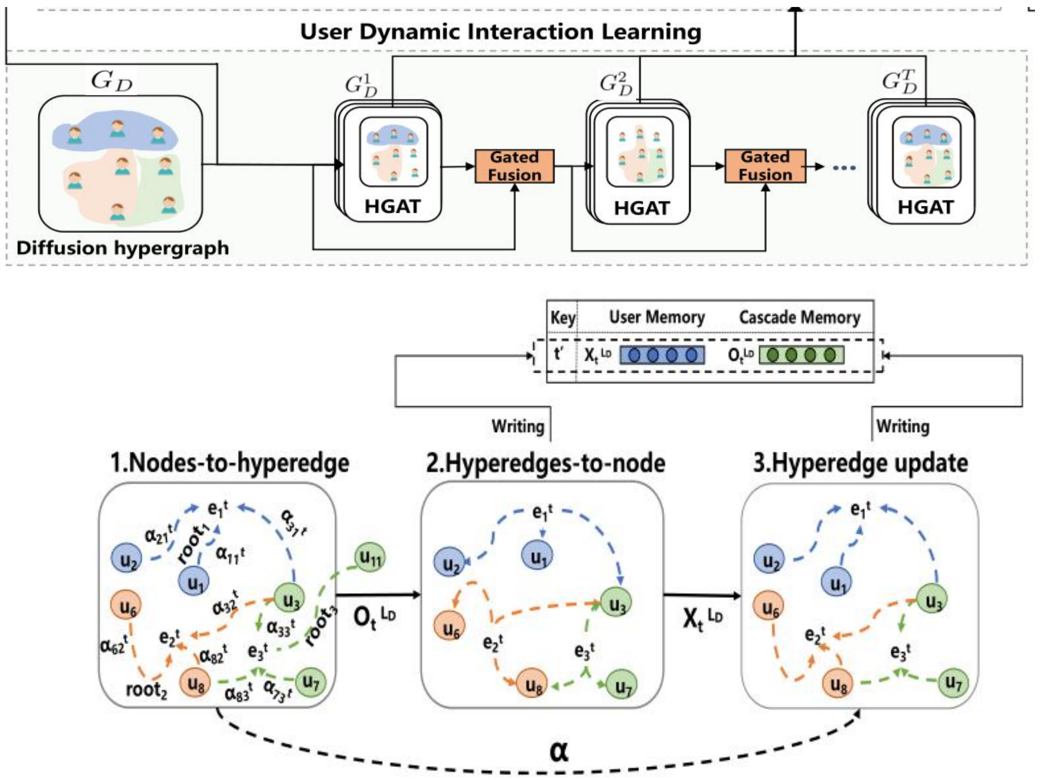


Figure 2: Stages of HGAT: (1) nodes-to-hyperedge aggregation, the root node of each cascade is determined in advance for attention calculation; (2) hyperedges-to-node aggregation, each hyperedge shares the same weight in this operation; (3) hyperedge update, the attention coefficient α is retained from first nodes-to-hyperedge aggregation. The learned user and the cascade representations are separately written to the memory module.

User Dynamic Interaction Learning:HGAT

Nodes-to-hyperedge aggregation:

$$\mathbf{o}_{j,t}^{l+1} = \sigma\Big(\sum_{u_i^t \in e_j^t} \alpha_{ij}^t \mathbf{W}_1 \mathbf{x}_{i,t}^l\Big) \tag{2}$$

$$\alpha_{ij}^t = \frac{\exp(-\operatorname{dis}(\mathbf{W}_1 \mathbf{x}_{i,t}^l, \mathbf{W}_1 \mathbf{r}_j^l))}{\sum_{u_p^t \in e_j^t} \exp(-\operatorname{dis}(\mathbf{W}_1 \mathbf{x}_{p,t}^l, \mathbf{W}_1 \mathbf{r}_j^l))} \tag{3}$$

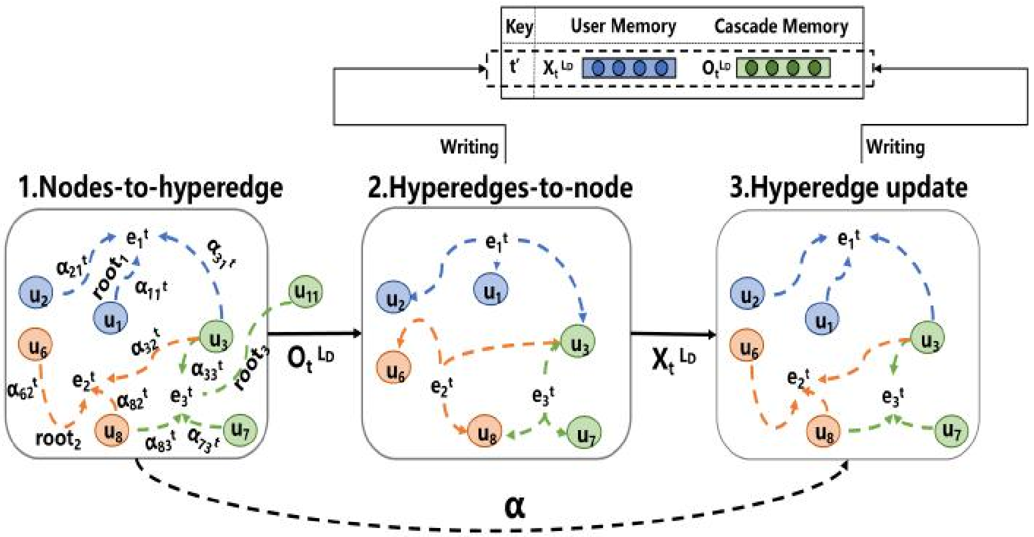
Hyperedges-to-node aggregation:

$$\mathbf{x}_{i,t}^{l+1} = \sigma\Big(\sum_{e_j^t \in \mathcal{E}_i^t} \mathbf{W}_2 \mathbf{o}_{j,t}^{l+1}\Big) \tag{4}$$

Update of hyperedges:

$$\mathbf{o}_{j,t}^{l+1'} = \sigma\Big(\sum_{u_i^t \in e_j^t} \alpha_{ij}^t \mathbf{W}_3 \mathbf{x}_{i,t}^{l+1}\Big) \tag{5}$$

Method



Sequential HGATs with Gated Fusion Strategy

$$
\mathbf{x}_{i,t+1}^{0} = g_{R_1}\mathbf{x}_{i,t}^{L_D} + (1 - g_{R_1})\mathbf{x}_{i,t}^{0}
$$

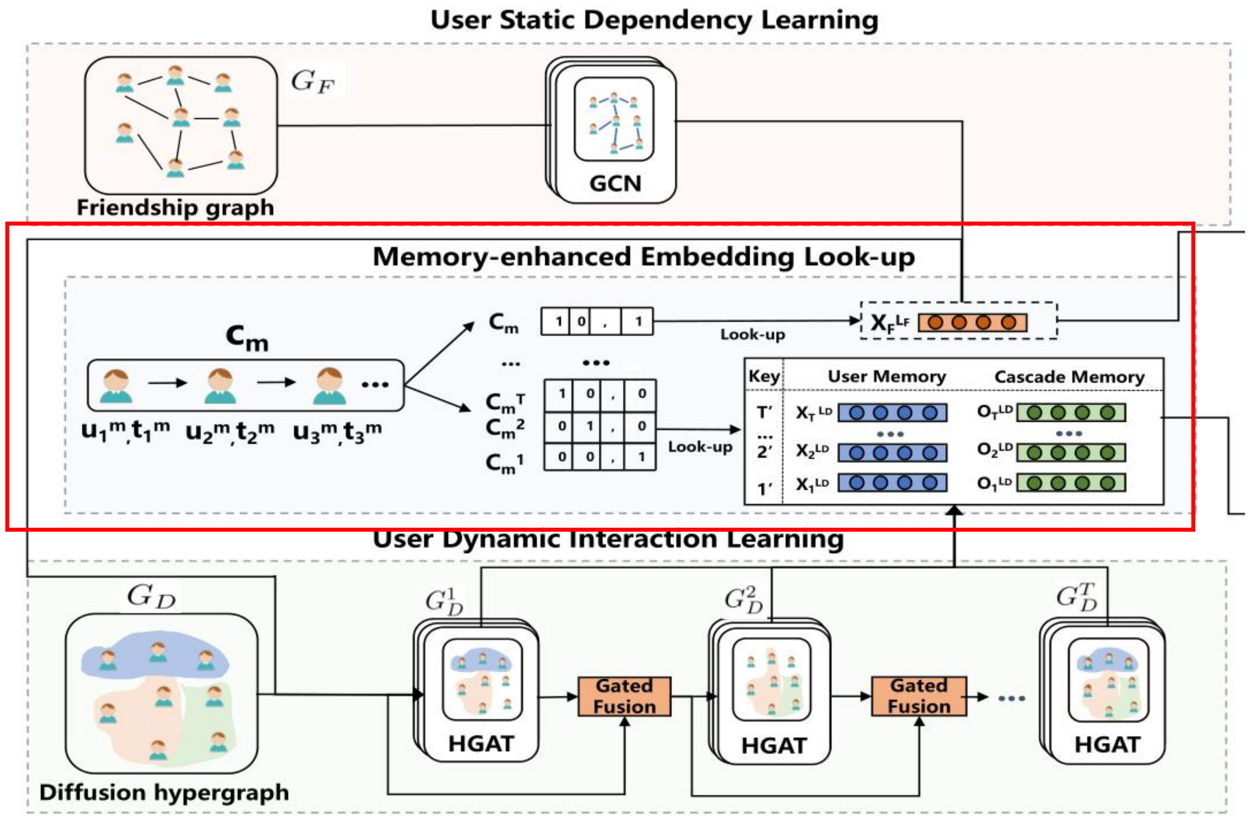
$$
g_{R_1} = \frac{\exp(\mathbf{W}_{Z_1}^{T}\sigma(\mathbf{W}_{R_1}\mathbf{x}_{i,t}^{L_D}))}{\exp(\mathbf{W}_{Z_1}^{T}\sigma(\mathbf{W}_{R_1}\mathbf{x}_{i,t}^{L_D}))+\exp(\mathbf{W}_{Z_1}^{T}\sigma(\mathbf{W}_{R_1}\mathbf{x}_{i,t}^{0}))} \quad (7)
$$

$$
\mathbf{X}_1^0 = \mathbf{X}_F^{L_F}
$$

Memory Module

$$
M_D = \left\{t' : (X_t^{L_D}, O_t^{L_D})\right\}, t = 1, 2, ..., T \qquad (6)
$$

Method



Memory-enhanced Embedding Look-up

Static representation look up:

Given the target cascade: $c_m=\{(u_i^m, t_i^m)|u_i^m \in U\}$

$$Z_m^F=[(x_i)] \in R^{|c_m|\times d} \quad i=0,1,\cdots,\mathcal{N}\text{-}1.$$

Static representation look up:

$$if\ t_i^m \ge t'\ and\ t_i^m < (t+1)' \rightarrow x_{i,t} \leftarrow X_t^{L_D}$$

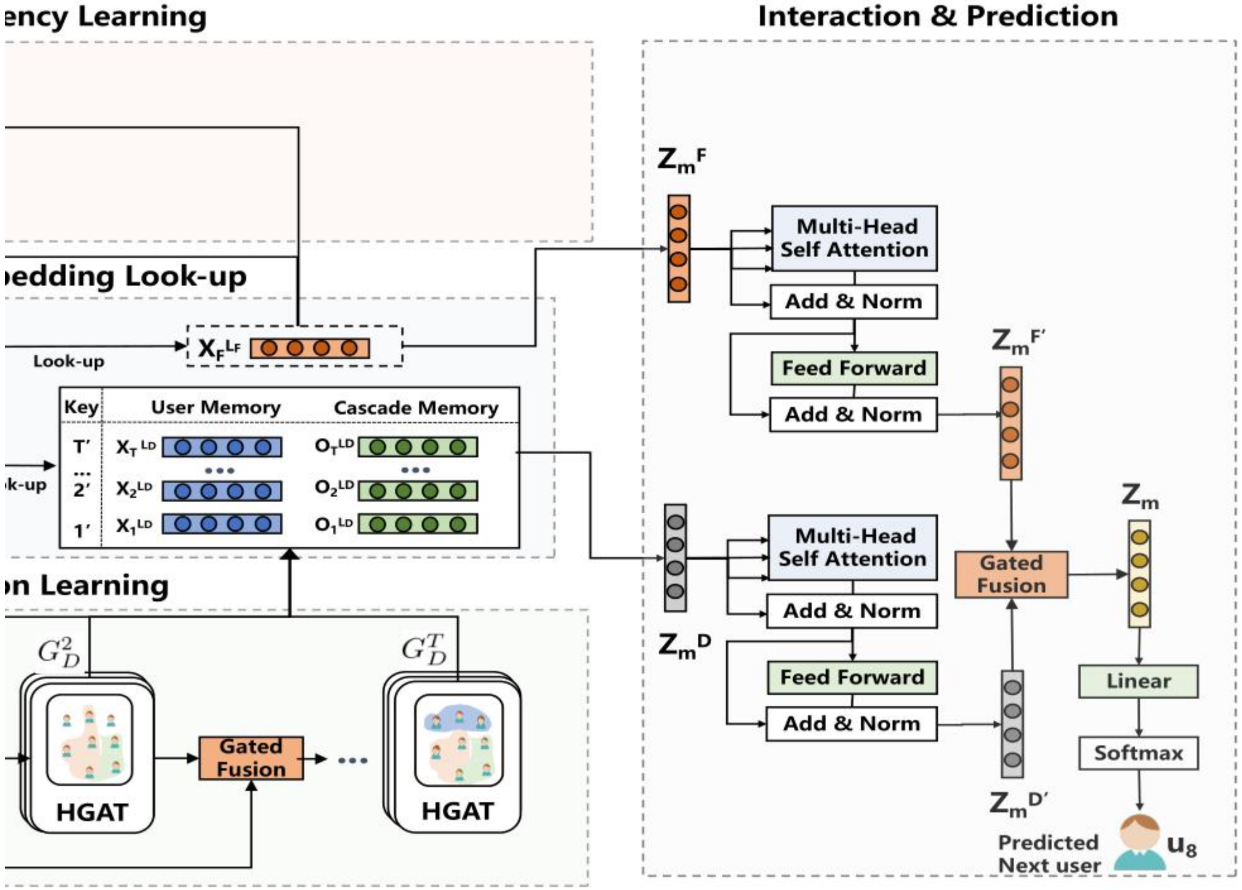
$$q_m^D=[(x_{i,t})] \in R^{|c_m|\times d} \quad i=0,1,\cdots,\mathcal{N}\text{-}1, t=1,2,\cdots,\mathcal{T}.$$

$$p_m^F=[(o_{m,t})] \in R^{|c_m|\times d} \quad t=1,2,\cdots,\mathcal{T}.$$

$$\mathbf{Z}_m^D = g_{R_2}\mathbf{p}_m^D + (1-g_{R_2})\mathbf{q}_m^D$$

$$g_{R_2} = \frac{\exp(\mathbf{W}_{Z_2}^T \sigma(\mathbf{W}_{R_2}\mathbf{p}_m^D))}{\exp(\mathbf{W}_{Z_2}^T \sigma(\mathbf{W}_{R_2}\mathbf{p}_m^D)) + \exp(\mathbf{W}_{Z_2}^T \sigma(\mathbf{W}_{R_2}\mathbf{q}_m^D))} \tag{8}$$

Method



Feature Interaction & Prediction

Self-attention:

$$\mathrm{Att}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d'}} + \mathbf{M}\right)\mathbf{V}$$

$$\mathbf{h}_{i,m}^F = \mathrm{Att}\left(\mathbf{Z}_m^F\mathbf{W}_i^Q, \mathbf{Z}_m^F\mathbf{W}_i^K, \mathbf{Z}_m^F\mathbf{W}_i^V\right) \qquad (9)$$

$$\mathbf{h}_m^F = \left[\mathbf{h}_{1,m}^F; \mathbf{h}_{2,m}^F; \ldots; \mathbf{h}_{H,m}^F\right]\mathbf{W}^O$$

$$M \in \mathbb{R}^{|c_m|\times|c_m|}$$

$$M_{i,j} = -\infty \text{ if } i>j \text{ else } M_{i,j} = 0$$

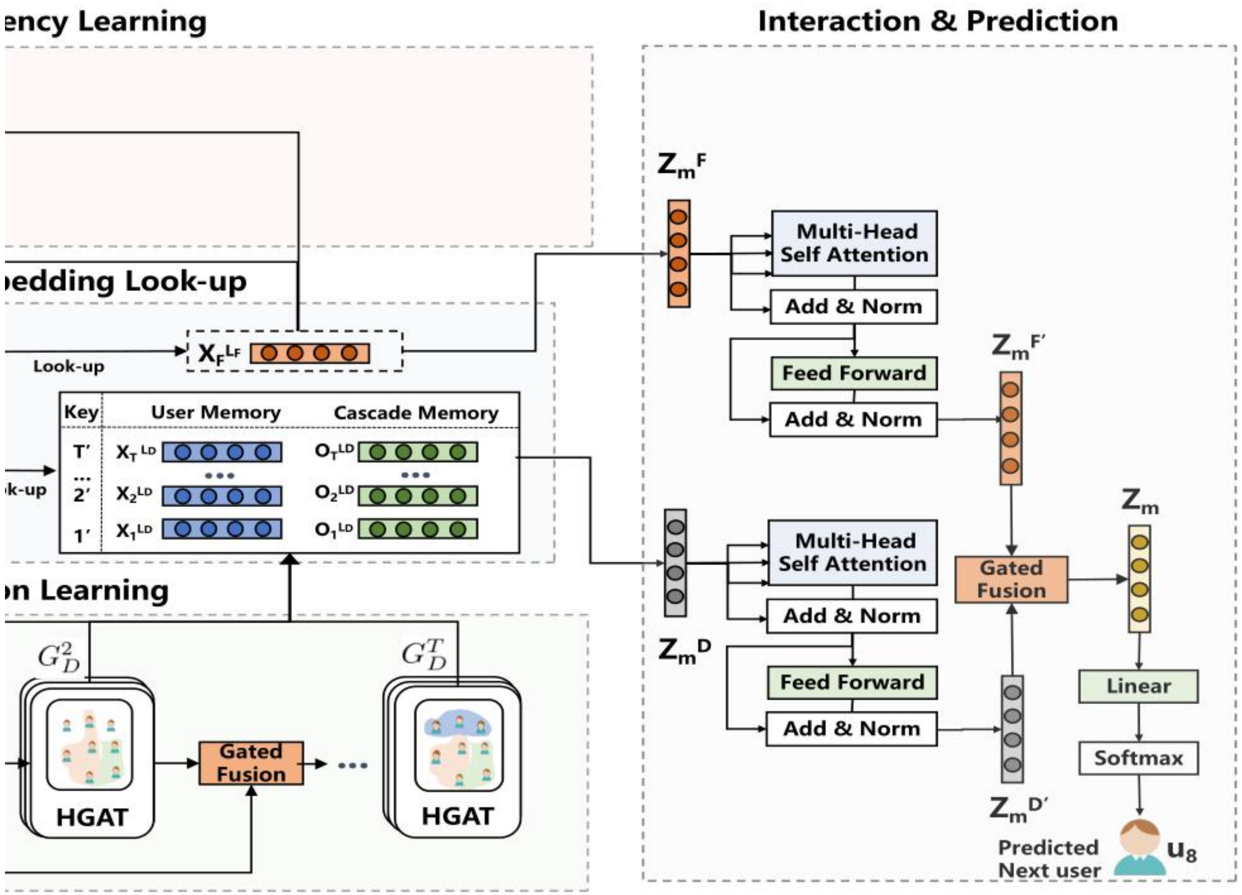
$$\mathbf{Z}_m^{F'} = \mathbf{ReLU}\left(\mathbf{h}_m^F\mathbf{W}_{A_1} + \mathbf{b}_1\right)\mathbf{W}_{A_2} + \mathbf{b}_2 \qquad (10)$$

Fusion Layer:

$$\mathbf{Z}_m = g_{R_3}\mathbf{Z}_m^{D'} + (1-g_{R_3})\mathbf{Z}_m^{F'}$$

$$g_{R_3} = \frac{\exp(\mathbf{W}_{Z_3}^T\sigma(\mathbf{W}_{R_3}\mathbf{Z}_m^{D'}))}{\exp(\mathbf{W}_{Z_3}^T\sigma(\mathbf{W}_{R_3}\mathbf{Z}_m^{D'}))+\exp(\mathbf{W}_{Z_3}^T\sigma(\mathbf{W}_{R_3}\mathbf{Z}_m^{F'}))} \qquad (11)$$

Method



Diffusion Prediction:

$$\hat{y} = \mathrm{softmax}(\mathbf{W}_p \mathbf{Z}_m + \mathbf{Mask}_m) \tag{12}$$

$$(\mathbf{Mask}_m)_{1,i}^{j,i} = 0$$

$$(\mathbf{Mask}_m)_{j+1,i}^{|c_m|,i} = -\infty$$

$$\mathcal{J}(\theta) = -\sum_{j=2}^{|c_m|} \sum_{i=1}^{|U|} \mathbf{y}_{ji} \log \left(\hat{\mathbf{y}}_{ji} \right) \tag{13}$$

in which θ represents all parameters that need to be learned in the model, if the user u_i participate in cascade c_m at the step j, $\mathbf{y}_{ji} = 1$, otherwise $\mathbf{y}_{ji} = 0$.

Experiments

Table 1: Statistics of datasets used in our experiments

Datasets	Twitter	Douban	Android	Christ.
# Users	12,627	12,232	9,958	2,897
		Frienship		
# Links	309,631	396,580	48,573	35,624
Density	24.52	30.21	4.87	12.30
		Interaction		
# Cascades	3,442	3,475	679	589
Avg. Length	32.60	21.76	33.3	22.9
Density	8.89	6.18	2.27	4.66

Experiments

Table 2: Experimental results on 4 dataset (%) (Hits@k scores for $K = 10, 50, 100$), scores are the higher the better.

Models	Twitter			Douban			Android			Christianity		
	@10	**@50**	**@100**	**@10**	**@50**	**@100**	**@10**	**@50**	**@100**	**@10**	**@50**	**@100**
DeepDiffuse	5.79	10.80	18.39	9.02	14.93	19.13	4.13	10.58	17.21	10.27	21.83	30.74
Topo-LSTM	8.45	15.80	25.42	8.57	16.53	21.47	4.56	12.63	16.53	12.28	22.63	31.52
NDM	15.21	28.23	32.30	10.00	21.13	30.14	4.85	14.24	18.97	15.41	31.36	45.86
SNIDSA	25.37	36.64	42.89	16.23	27.24	35.59	5.63	15.22	20.93	17.74	34.58	48.76
FOREST	28.67	42.07	49.75	19.50	32.03	39.08	9.68	17.73	24.08	24.85	42.01	51.28
Inf-VAE	14.85	32.72	45.72	8.94	22.02	35.72	5.98	14.70	20.91	18.38	38.50	51.05
DyHGCN	31.88	45.05	52.19	18.71	32.33	39.71	9.10	16.38	23.09	26.62	42.80	52.47
MS-HGAT (ours)	**33.50**	**49.59**	**58.91**	**21.33**	**35.25**	**42.75**	**10.41**	**20.31**	**27.55**	**28.80**	**47.14**	**55.62**
recurrence	33.38	49.49	**58.98**	**22.19**	**35.64**	**43.11**	**10.85**	20.23	**27.76**	28.00	**47.33**	**56.60**

Experiments

Table 3: Experimental results on 4 dataset (%) (MAP@k scores for $K = 10, 50, 100$), scores are the higher the better.

Models	Twitter			Douban			Android			Christianity		
	@10	@50	@100	@10	@50	@100	@10	@50	@100	@10	@50	@100
DeepDiffuse	5.87	6.80	6.39	6.02	6.93	7.13	2.30	2.53	2.56	7.27	7.83	7.84
Topo-LSTM	8.51	12.68	13.68	6.57	7.53	7.78	3.60	4.05	4.06	7.93	8.67	9.86
NDM	12.41	13.23	14.30	8.24	8.73	9.14	2.01	2.22	2.93	7.41	7.68	7.86
SNIDSA	15.34	16.64	16.89	10.02	11.24	11.59	2.98	3.24	3.97	8.69	8.94	9.72
FOREST	19.60	20.21	21.75	11.26	11.84	11.94	5.83	6.17	6.26	14.64	15.45	15.58
Inf-VAE	19.80	20.66	21.32	11.02	11.28	12.28	4.82	4.86	5.27	9.25	11.96	12.45
DyHGCN	20.87	21.48	21.58	10.61	11.26	11.36	6.09	6.40	6.50	15.64	16.30	16.44
MS-HGAT (ours)	**22.49**	**23.17**	**23.30**	**11.72**	**12.52**	**12.60**	**6.39**	**6.87**	**6.96**	**17.44**	**18.27**	**18.40**
recurrence	21.15	21.83	21.96	**12.09**	**12.72**	**12.83**	**6.51**	**6.97**	**7.07**	17.19	18.06	18.21

Experiments

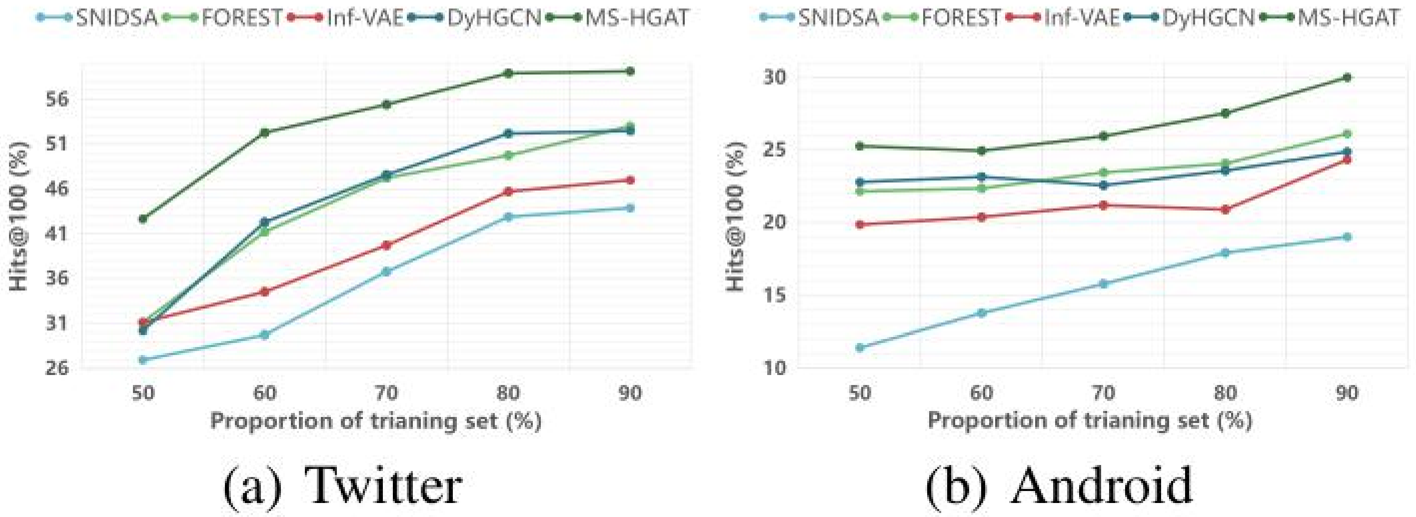
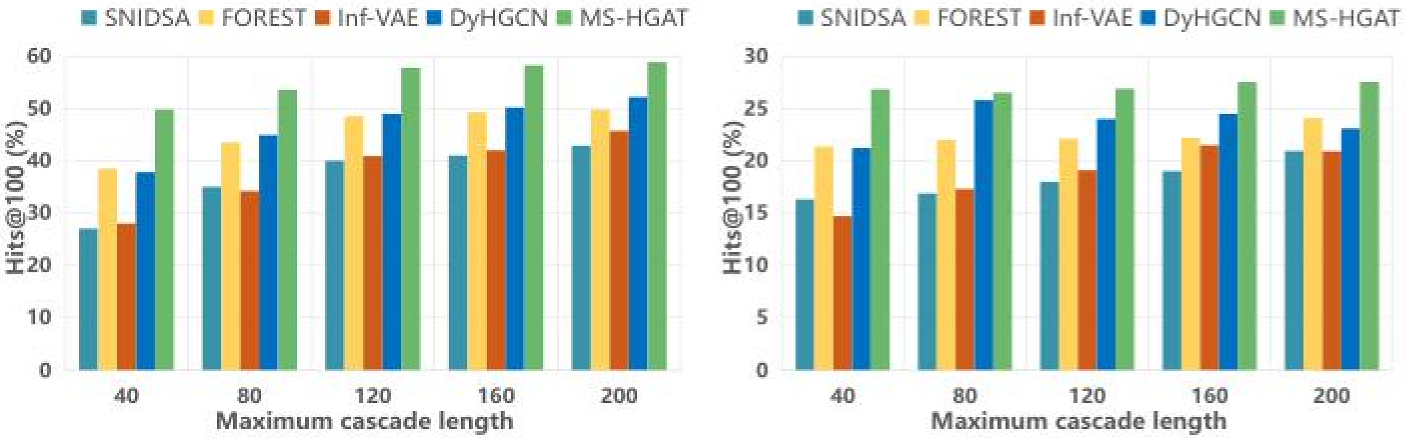


Figure 3: Impact of training proportion.

Experiments



(a) Twitter (b) Android

Figure 4: Impact of maximum cascade lengths.

Experiments

Table 4: Ablation study of MS-HGAT.

Models	Twitter		Android	
	Hits@100	**MAP@100**	**Hits@100**	**MAP@100**
MS-HGAT	58.91	**23.30**	**27.55**	**6.96**
w/o FG	57.20	21.38	26.32	6.86
w/o DH	57.41	22.24	26.74	6.78
w/o UM	58.63	22.74	26.40	6.83
w/o CM	58.32	21.96	27.09	6.77
w/o ATTH	**58.95**	22.76	27.03	6.75
w/o GF	57.93	22.19	27.26	6.89

Thanks