Artificial

User-Centric Conversational Recommendation with Multi-Aspect User Modeling

Shuokai Li[1,2,*], Ruobing Xie[3,*], Yongchun Zhu[1,2], Xiang Ao[1,2,†], Fuzhen Zhuang[4,5], Qing He[1,2,§]
[1]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China. [2]University of Chinese Academy of Sciences, Beijing 100049, China. [3]WeChat Search Application Department, Tencent, China. [4]Institute of Artificial Intelligence, Beihang Unversity, Beijing 100191, China. [5]SKLSDE, School of Computer Science, Beihang University, Beijing 100191, China.
{lishuokai18z, zhuyongchun18s, aoxiang, heqing}@ict.ac.cn, ruobingxie@tencent.com, zhuangfuzhen@buaa.edu.cn

Code :https://github.com/lisk123/UCCR

SIGIR 2022

2022.9.21 • ChongQing

Reported by Junhao Cao

1.Introduction

2.Method

3.Experiments

Introduction

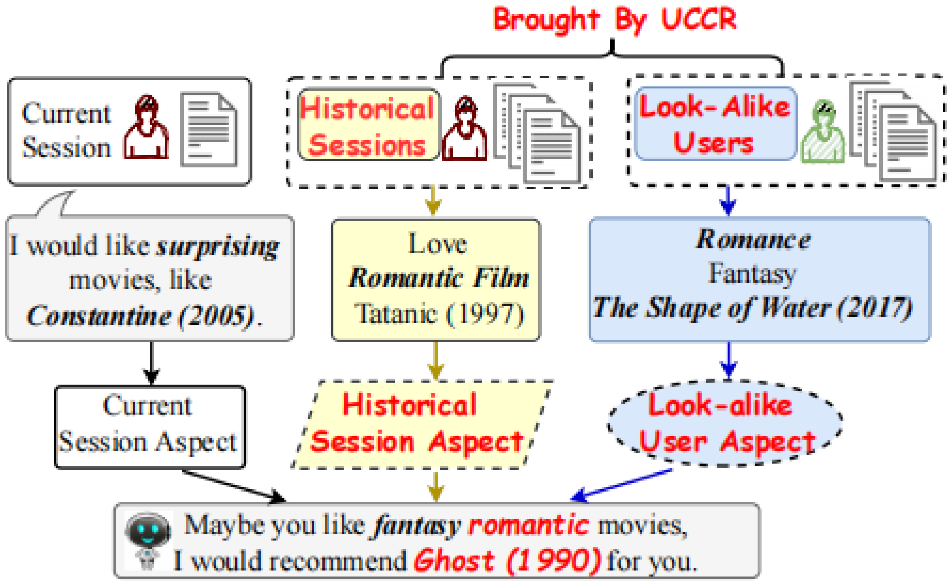


Figure 1: An example of the multi-aspect user information. UCCR introduces the historical dialogue sessions and look-alike users to CRS for user-centric preference learning.

most conventional CRS models mainly focus on the dialogue understanding of the current session, ignoring other rich multi-aspect information of the central subjects (i.e., users) in recommendation

In this work, we highlight that the user’s historical dialogue sessions and look-alike users are essential sources of user preferences besides the current dialogue session in CRS.

Introduction

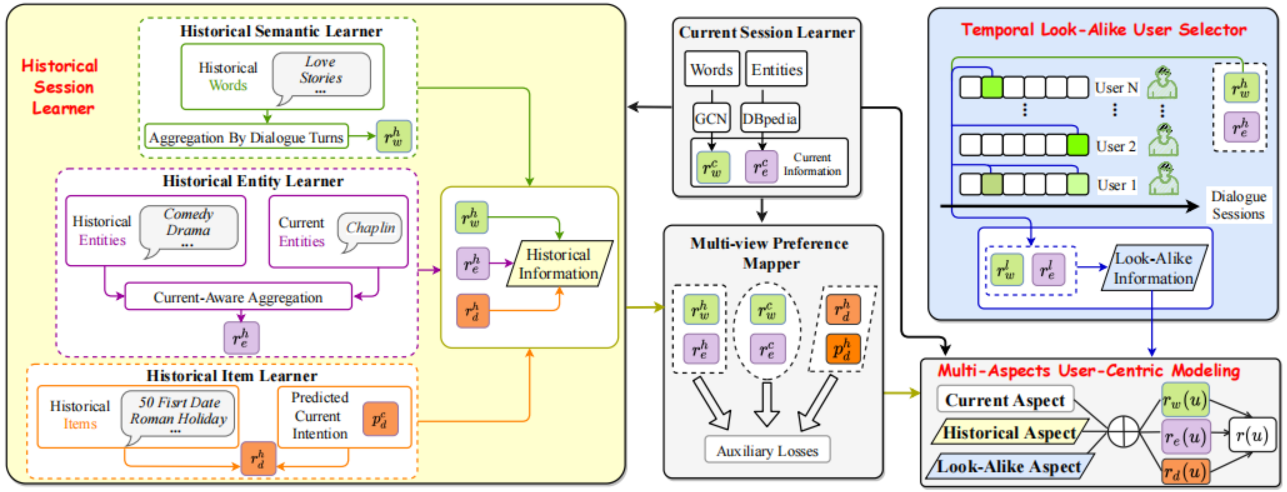


Figure 2: The overview of our model UCCR. First, the multiple views information is encoded by the historical and current session learners. Second, the multi-view preference mapper further explores the correlations between views. Next, the temporal look-alike users selector provides another aspect feature. Finally, these aspects are fused by the user-centric modeling module.

There are mainly three objects in CRS, namely user mentioned entities, words, and items. *User mentioned entities $e \in \mathcal{E}$* are entities in certain KG extracted from dialogues, which contain structural knowledge. In contrast, *words $w \in \mathcal{W}$* reflect semantic knowledge in dialogues. In conventional CRS, *items $d \in \mathcal{I}$* are recommended mainly via user preferences learned from user mentioned entities and words in the current dialogue sessions [39]. Note that in our

Method

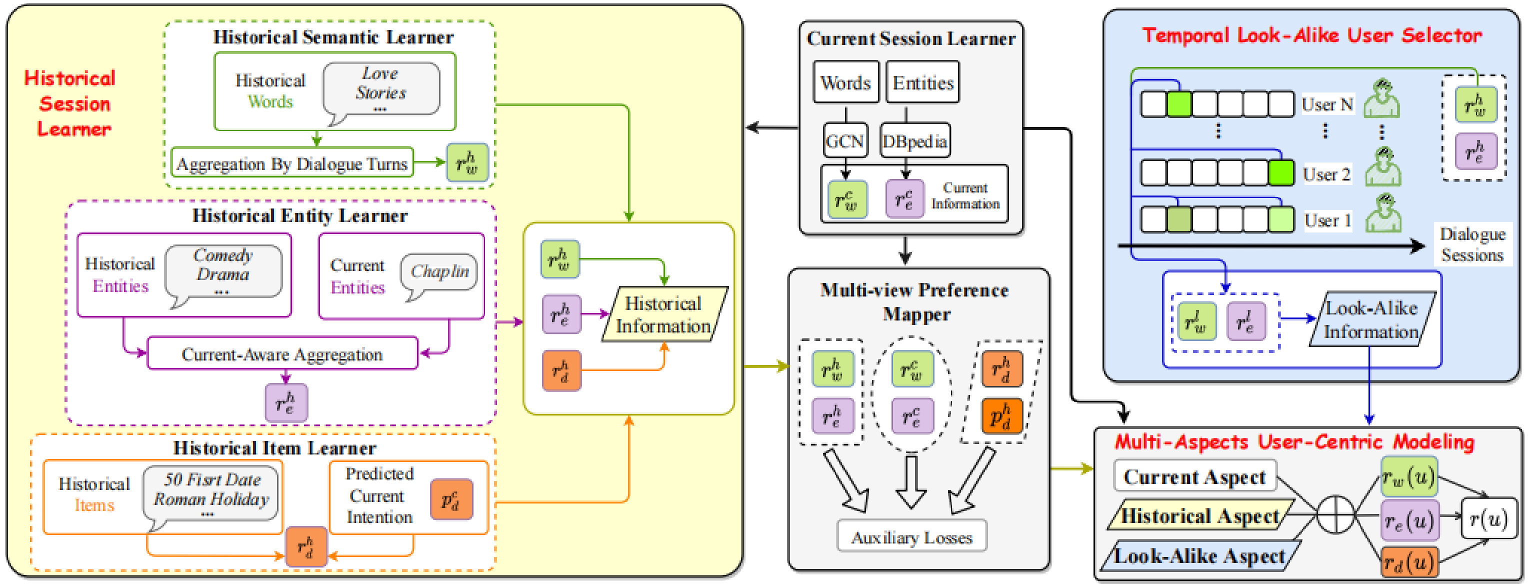


Figure 2: The overview of our model UCCR. First, the multiple views information is encoded by the historical and current session learners. Second, the multi-view preference mapper further explores the correlations between views. Next, the temporal look-alike users selector provides another aspect feature. Finally, these aspects are fused by the user-centric modeling module.

Current Session Learner

Current Entity Learner

$$C_e = \{e_1^T, \ldots, e_t^T\}$$

$$v_e^{l+1} = \sigma\Big(\sum_{r\in\mathcal{R}} \sum_{e'\in\mathcal{N}_e^r} \frac{1}{Z_{e,r}} W_r^l v_{e'}^l + W^l v_e^l\Big), \quad (1)$$

$$r_e^c = \text{R-GCN}(C_e) = \mathcal{F}(\mu_e (V_e)^{\mathrm{T}}), \quad (2)$$
$$\mu_e = \text{Softmax}(b_e \text{Tanh}(W_e V_e)),$$

Current Semantic Learner

$$C_w = \{w_1^T, \ldots, w_t^T\},$$

$$r_w^c = \text{GCN}(C_w)$$

Method

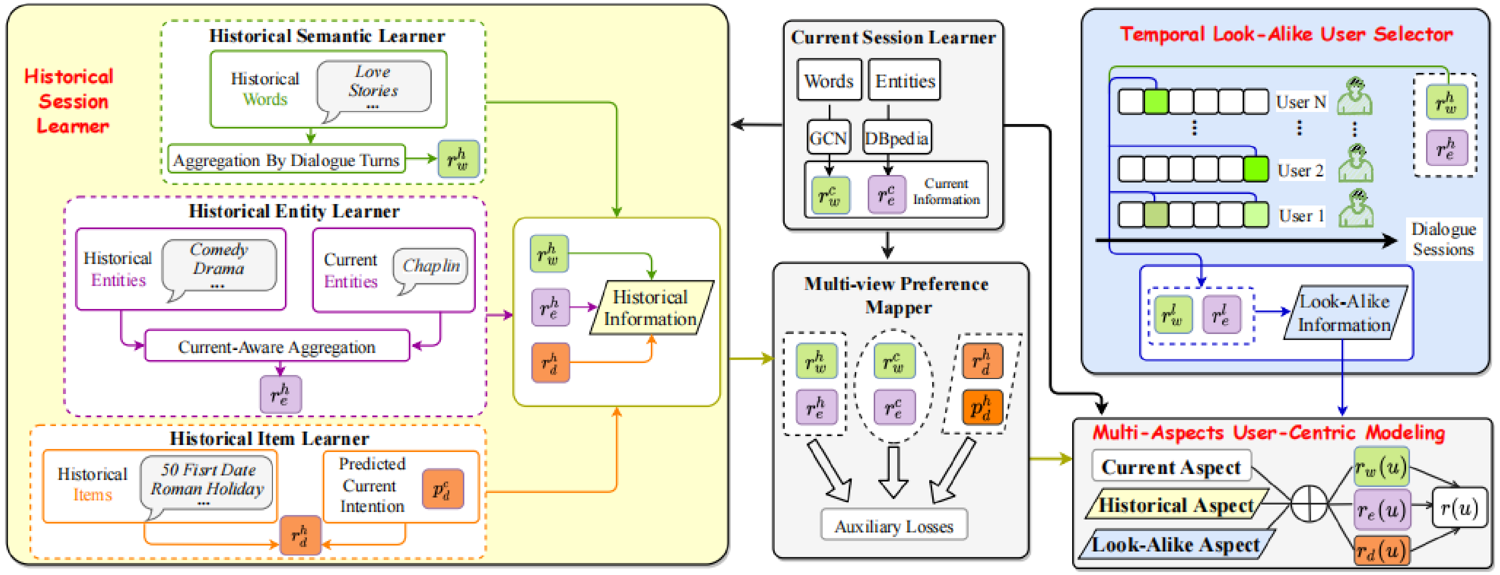


Figure 2: The overview of our model UCCR. First, the multiple views information is encoded by the historical and current session learners. Second, the multi-view preference mapper further explores the correlations between views. Next, the temporal look-alike users selector provides another aspect feature. Finally, these aspects are fused by the user-centric modeling module.

Historical Session Learner

Historical Entity Learner

$$\mathcal{H}_e = \{\mathcal{H}_e^1, ..., \mathcal{H}_e^{T-1}\}$$

$$\mathcal{H}_e^j = \{e_1^j, ..., e_t^j\}$$

$$h_e^j = \text{R-GCN}(\mathcal{H}_e^j)$$

$$r_e^h = \text{Agg}(r_e^c, h_e^1, ..., h_e^{T-1}) = \sum_{j=1}^{T-1} \varphi(h_e^j, r_e^c)\, h_e^j, \quad (3)$$

$$\varphi(h_e^j, r_e^c) = \text{Softmax}(h_e^j W_s r_e^c \,/\, \lambda_e),$$

Method

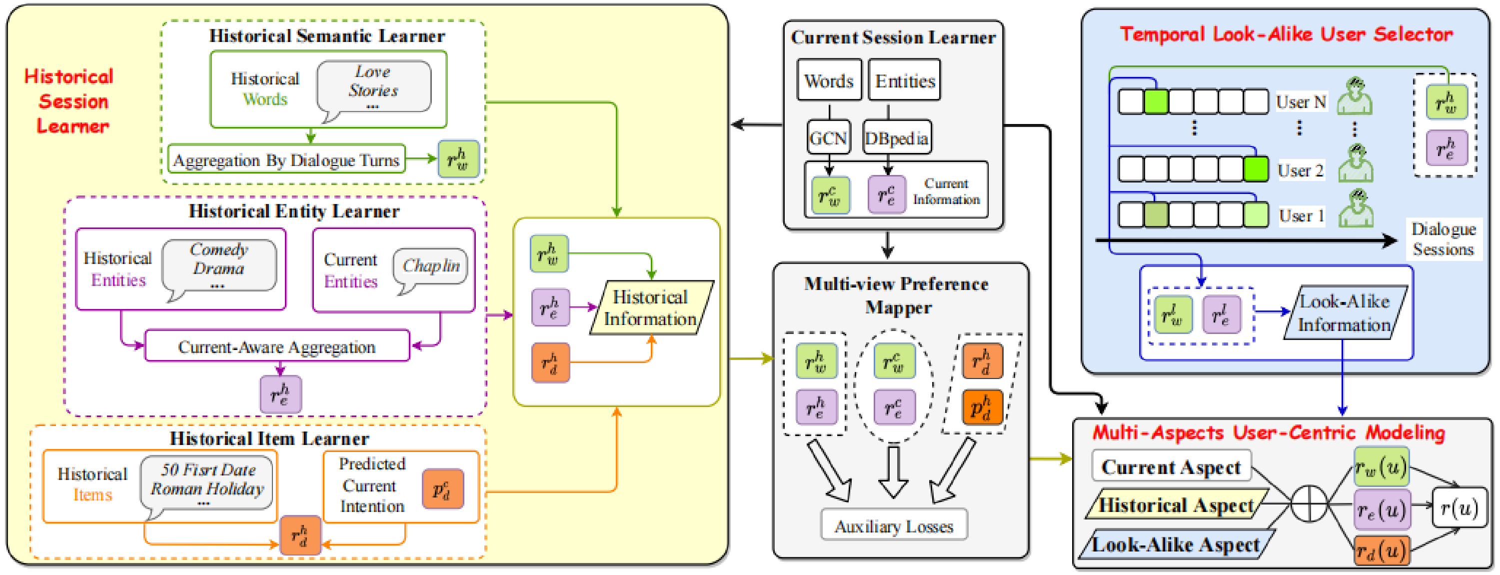


Figure 2: The overview of our model UCCR. First, the multiple views information is encoded by the historical and current session learners. Second, the multi-view preference mapper further explores the correlations between views. Next, the temporal look-alike users selector provides another aspect feature. Finally, these aspects are fused by the user-centric modeling module.

Historical Semantic Learner

$$\mathcal{H}_w^j = \{w_1^j, \ldots, w_t^j\}$$

$$V_w^j = \{v_{w_1^j}, \ldots, v_{w_t^j}\}$$

$$h_w^j = \mathcal{F}\left(\sum_{m=1}^{t} s(w_m^j)\, v_{w_m^j}\right), \quad (4)$$

where $s(w_m^j) = \mathrm{Softmax}(1, 2, \ldots, t)[m]$ is the importance of w_m^j according to the dialogue turn.

$$r_w^h = \sum_{j=1}^{T-1} s(h_w^j)\, h_w^j,$$

where j is the index of the dialogue sessions.

Method

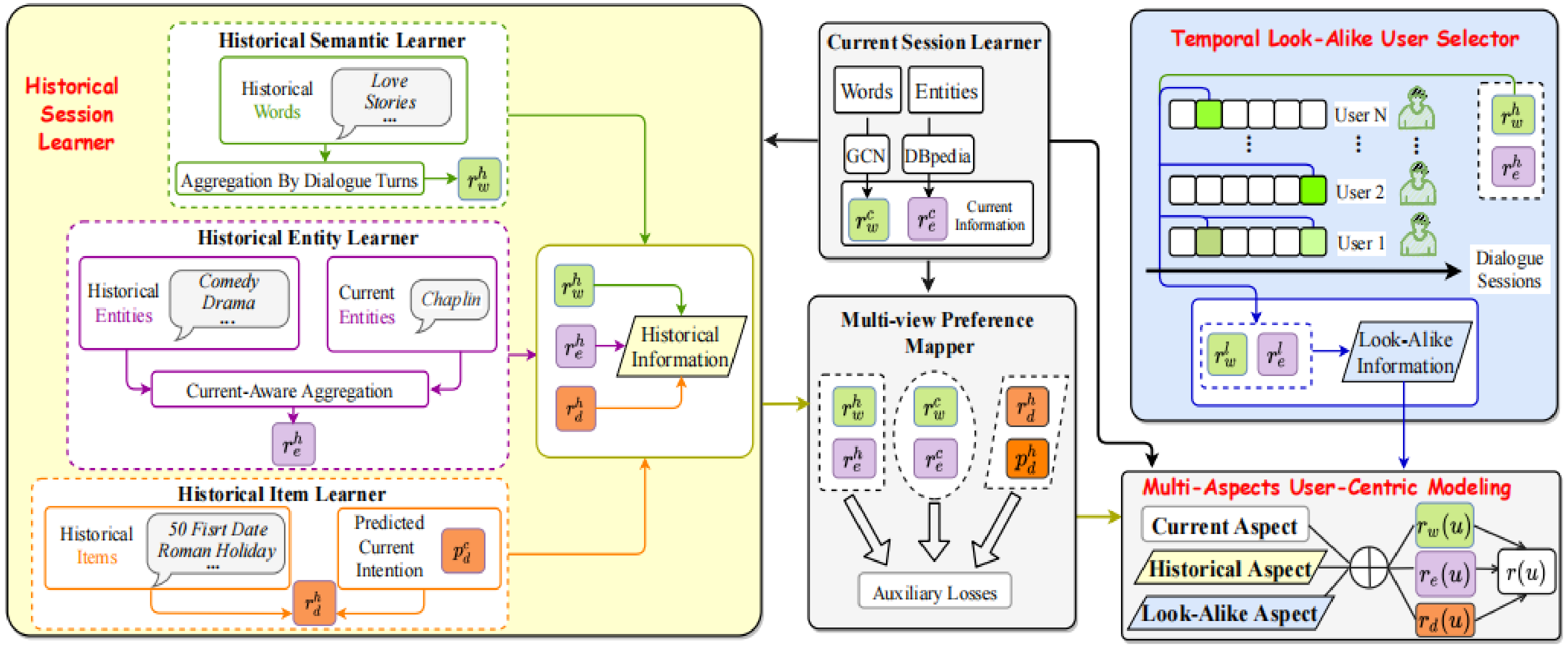


Figure 2: The overview of our model UCCR. First, the multiple views information is encoded by the historical and current session learners. Second, the multi-view preference mapper further explores the correlations between views. Next, the temporal look-alike users selector provides another aspect feature. Finally, these aspects are fused by the user-centric modeling module.

Historical Item Learner

$$\mathcal{H}_d^j = \{d_1^j, \dots d_t^j\}$$

$$\boldsymbol{h}_d^j = \text{R-GCN}(\mathcal{H}_d^j)$$

$$p_d^c = g(r_w^c, r_e^c) = \tau \cdot r_w^c + (1-\tau) \cdot r_e^c, \quad \tau = \sigma(W_g \text{Concat}(r_w^c, r_e^c)), \tag{5}$$

$$r_d^h = \text{Agg}(p_d^c, \boldsymbol{h}_d^1, \dots, \boldsymbol{h}_d^{T-1})$$

Method

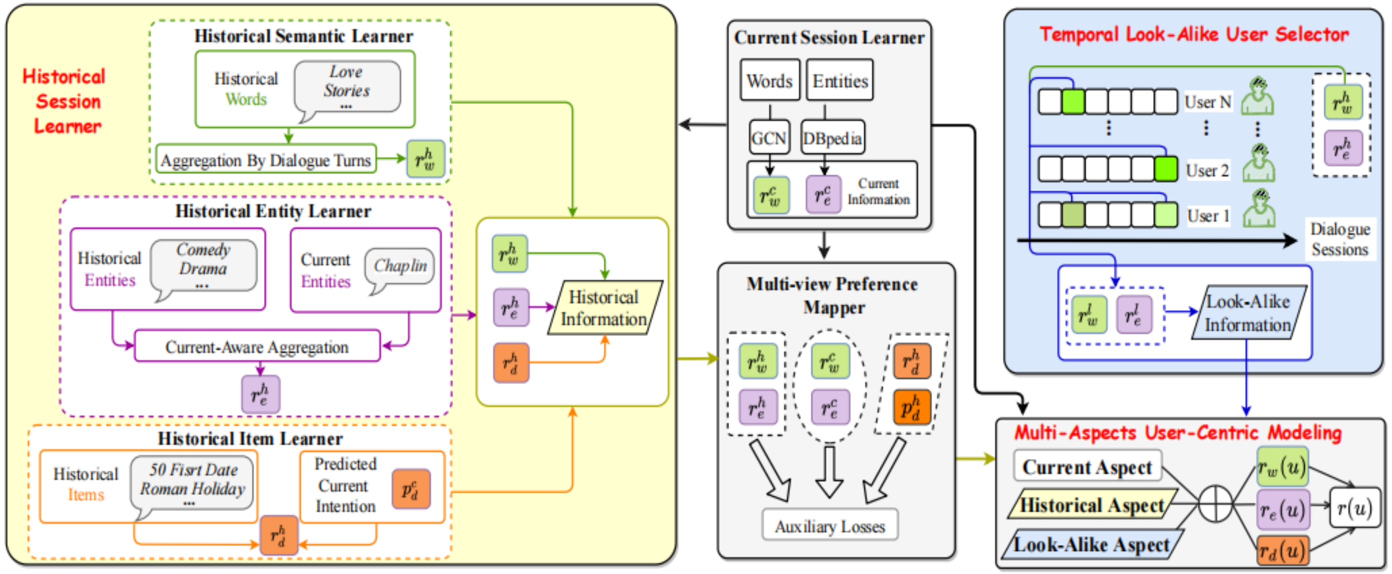


Figure 2: The overview of our model UCCR. First, the multiple views information is encoded by the historical and current session learners. Second, the multi-view preference mapper further explores the correlations between views. Next, the temporal look-alike users selector provides another aspect feature. Finally, these aspects are fused by the user-centric modeling module.

Multi-View Preference Mapper

$$\mathcal{L}_a(v_1, v_2) = \sum_{u\in\mathcal{B}}(1-\mathrm{sim}(v_1^u, v_2^u))^2 + \lambda_a \sum_{u,u'\in\mathcal{B}}(\mathrm{sim}(v_1^u, v_2^{u'}))^2, \tag{6}$$

where $\mathrm{sim}(\cdot,\cdot)$ is the cosine similarity function that measures the correlation between two views. Here, u' represents the negative users, which are all other users of batch $\mathcal{B}$ except for u.

Specifically, we have three alignment tasks: (1) $v_1 = r_w^c$, $v_2 = r_e^c$; (2) $v_1 = r_w^h$, $v_2 = r_e^h$; (3) $v_1 = r_d^h$, $v_2 = p_d^h$, p_d^h is the combination of historical words and entities (refers to Eq. (5)). Here we do not

Temporal Look-Alike User Selector

$$r_w^l(u, u') = \sum_{k=1}^{K} \delta(\mathrm{sim}(r_w^h(u), r_w^h(u'_k)))\, r_w^c(u'_k), \tag{7}$$

Suppose that a user u' has K recommendation turns totally (if there are T historical sessions and each session has t recommendation turns, we have $K = t * T$ turns in total). In each recommendation

where $\delta(x) = \max(0, x - \delta_w)$ is a clip function

Method

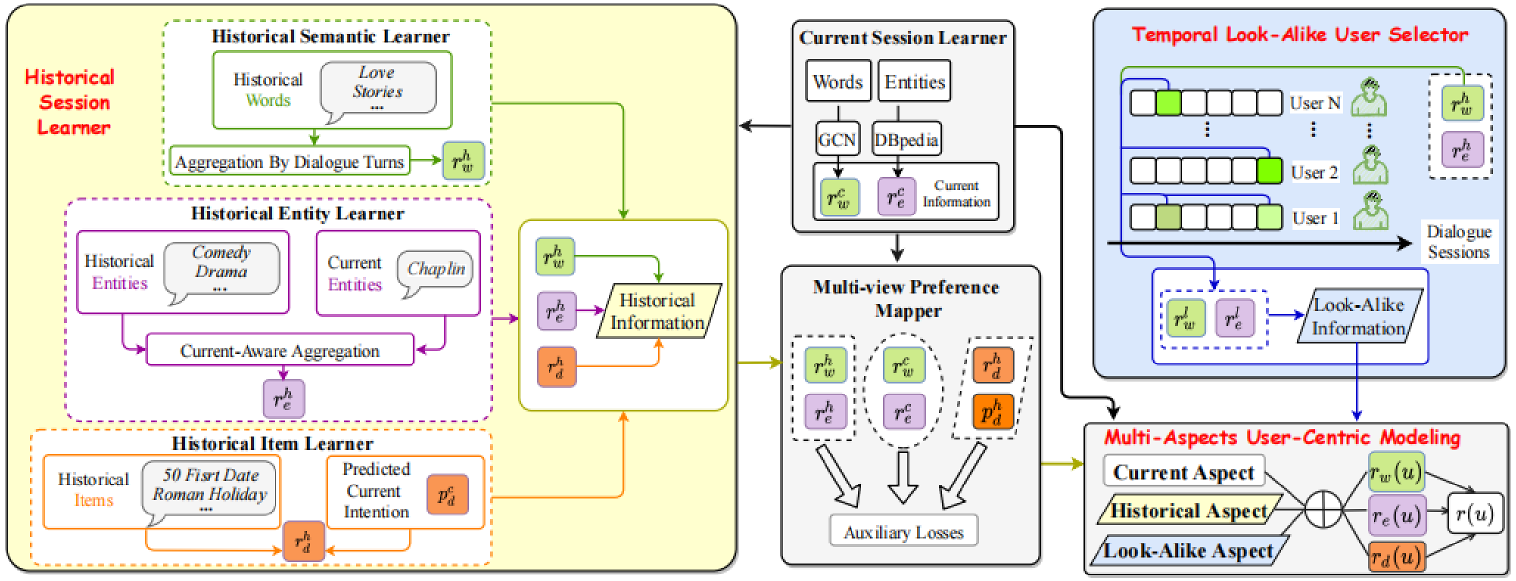


Figure 2: The overview of our model UCCR. First, the multiple views information is encoded by the historical and current session learners. Second, the multi-view preference mapper further explores the correlations between views. Next, the temporal look-alike users selector provides another aspect feature. Finally, these aspects are fused by the user-centric modeling module.

Multi-Aspect User-Centric Modeling

Multi-Aspect Entity View Modeling

$$\boldsymbol{r}_e(u) = \boldsymbol{r}_e^c(u) + \alpha_h \boldsymbol{r}_e^h(u) + \alpha_s \sum_{u' \in \mathcal{U}} \boldsymbol{r}_e^l(u, u'), \qquad (8)$$

$$\alpha_h = \mathcal{G}(\mathrm{Concat}(\boldsymbol{r}_e^c(u), \boldsymbol{r}_e^h(u))) \;/\; \tau_e, \qquad (9)$$

Multi-Aspect Word View Modeling

$$\boldsymbol{r}_w(u) = \boldsymbol{r}_w^c(u) + \beta_h \boldsymbol{r}_w^h(u) + \beta_s \sum_{u' \in \mathcal{U}} \boldsymbol{r}_w^l(u, u'), \qquad (10)$$

Multi-Aspect Item View Modeling

$$\boldsymbol{r}_d(u) = \gamma_h \boldsymbol{r}_d^h(u), \quad \gamma_h = \delta(\mathrm{sim}(\boldsymbol{p}_d^h,\ \boldsymbol{p}_d^c)), \qquad (11)$$

$$\boldsymbol{r}(u) = g(\boldsymbol{r}_w(u),\ \boldsymbol{r}_e(u)) + \boldsymbol{r}_d(u), \qquad (12)$$

Method

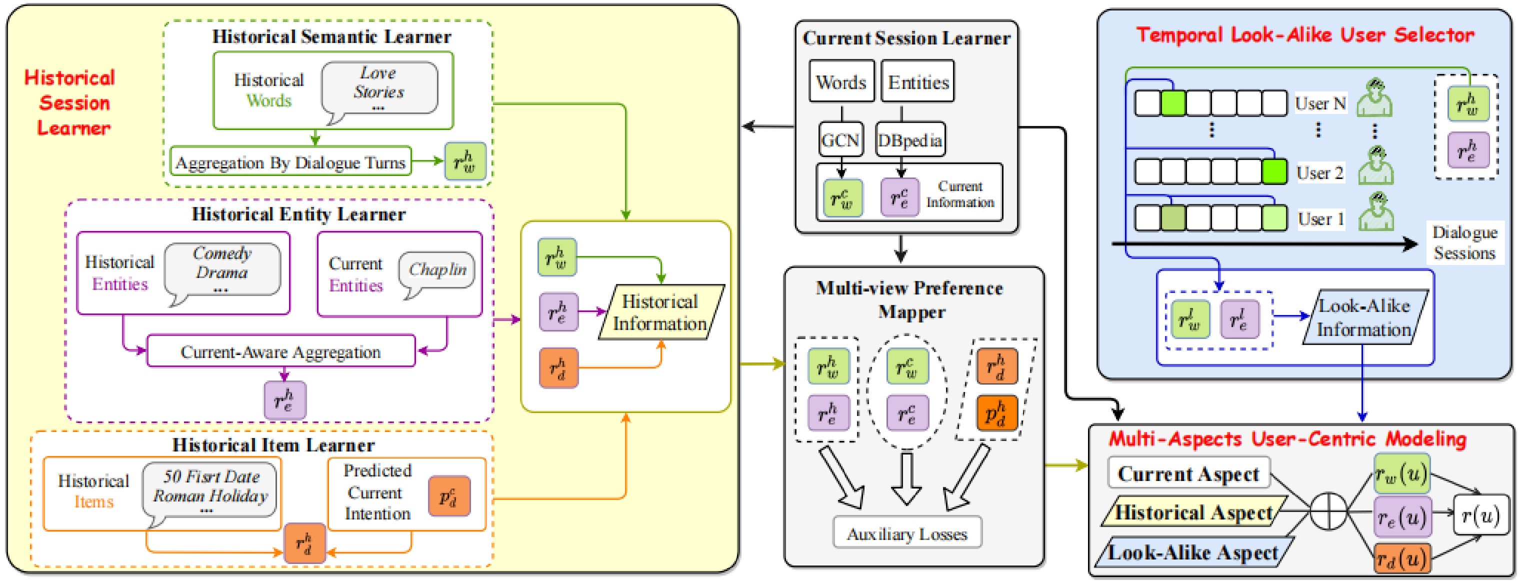


Figure 2: The overview of our model UCCR. First, the multiple views information is encoded by the historical and current session learners. Second, the multi-view preference mapper further explores the correlations between views. Next, the temporal look-alike users selector provides another aspect feature. Finally, these aspects are fused by the user-centric modeling module.

Optimization

Recommendation Objective

$$p_{rec}(u, d_i) = \text{Softmax}(\boldsymbol{r}(\boldsymbol{u})^\top \cdot \boldsymbol{d}_i), \tag{13}$$

$$\mathcal{L}_{rec} = -\sum_{u\in\mathcal{U}}\sum_{i=1}^{N_u} \log p_{rec}(u, d_i) + \lambda_{CL}\sum_{(v_1,v_2)} \mathcal{L}_a(v_1, v_2), \tag{14}$$

Dialogue Generation Objective

$$p_{dial}(y_t|y_1, ..., y_{t-1}) = \text{Softmax}(W^G \boldsymbol{q} + \mathcal{M}(\boldsymbol{r}(u))[y_t], \tag{15}$$

$$\mathcal{L}_{dial} = -\sum_{u\in\mathcal{U}}\sum_{t=2}^{N_t} \log(p_{dial}(y_t|y_1, ..., y_{t-1})). \tag{16}$$

Experiments

Table 1: The recommendation results. The marker * indicates that the improvement is statistically significant compared with the best baseline (t-test with p-value < 0.05).

Dataset	TG-ReDial						ReDial					
Method	HR@10	HR@50	MRR@10	MRR@50	NDCG@10	NDCG@50	HR@10	HR@50	MRR@10	MRR@50	NDCG@10	NDCG@50
SASRec	0.0048	0.0170	0.0011	0.0016	0.0019	0.0046	0.0418	0.1598	0.0385	0.0407	0.0473	0.0712
Text CNN	0.0052	0.0188	0.0015	0.0022	0.0029	0.0058	0.0733	0.1810	0.0438	0.0482	0.0576	0.0808
Bert	0.0098	0.0356	0.0027	0.0040	0.0051	0.0101	0.1499	0.2937	0.0683	0.0761	0.0813	0.1167
ReDial	0.0102	0.0370	0.0028	0.0041	0.0053	0.0107	0.1733	0.3359	0.0779	0.0841	0.0969	0.1351
KBRD	0.0141	0.0481	0.0045	0.0063	0.0072	0.0143	0.1827	0.3688	0.0784	0.0855	0.1004	0.1428
TG-ReDial	0.0168	0.0513	0.0061	0.0080	0.0088	0.0161	0.1893	0.3801	0.0801	0.0883	0.1032	0.1477
KGSF	0.0175	0.0543	0.0073	0.0088	0.0096	0.0175	0.2006	0.4034	0.0837	0.0932	0.1110	0.1556
KECRS	0.0113	0.0394	0.0033	0.0042	0.0057	0.0111	0.1772	0.3423	0.0780	0.0851	0.0983	0.1391
RevCore	0.0191	0.0581	0.0077	0.0093	0.0105	0.0189	0.2058	0.4088	0.0850	0.0946	0.1132	0.1583
UCCR w/o En	0.0167	0.0506	0.0071	0.0085	0.0092	0.0165	0.1976	0.3885	0.0812	0.0908	0.1084	0.1502
UCCR w/o Wo	0.0207	0.0592	0.0080	0.0095	0.0114	0.0196	0.2106	0.4196	0.0865	0.0959	0.1168	0.1613
UCCR w/o It	0.0211	0.0626	0.0082	0.0098	0.0116	0.0201	0.2146	0.4193	0.0865	0.0966	0.1173	0.1619
UCCR	$\mathbf{0.0232}^*$	$\mathbf{0.0664}^*$	$\mathbf{0.0088}^*$	$\mathbf{0.0107}^*$	$\mathbf{0.0122}^*$	$\mathbf{0.0214}^*$	$\mathbf{0.2161}^*$	$\mathbf{0.4258}^*$	$\mathbf{0.0883}^*$	$\mathbf{0.0981}^*$	$\mathbf{0.1182}^*$	$\mathbf{0.1642}^*$

Experiments

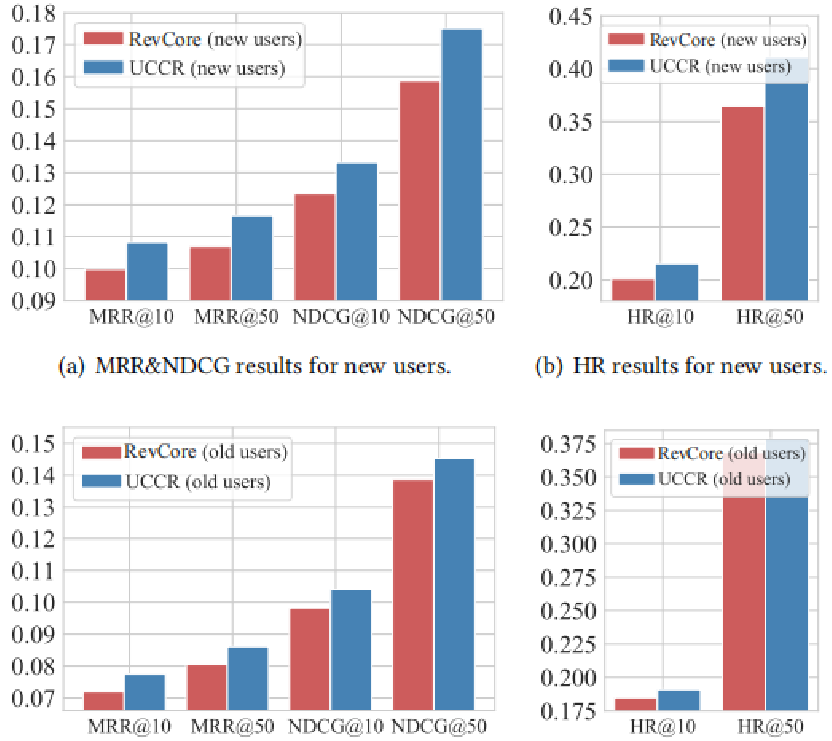
Table 2: Results on dialogue generation. Flu. and Inf. stand for Fluency and Informativeness, respectively. The marker * indicates that the improvement is statistically significant compared with the best baseline (t-test with p-value< 0.05).

Dataset	TG-ReDial								ReDial							
Method	Bleu-2	Bleu-3	Dist-2	Dist-3	Dist-4	PPL	Flu.	Inf.	Bleu-2	Bleu-3	Dist-2	Dist-3	Dist-4	PPL	Flu.	Inf.
ReDial	0.0409	0.0102	0.2672	0.5288	0.8012	55.71	0.71	0.75	0.0217	0.0078	0.0689	0.2697	0.4638	56.21	0.73	0.91
KBRD	0.0423	0.0119	0.3482	0.6911	0.9972	53.08	0.83	0.88	0.0238	0.0088	0.0712	0.2883	0.4893	54.89	0.82	1.00
KGSF	0.0461	0.0135	0.4447	1.0450	1.5792	51.27	1.01	1.09	0.0249	0.0091	0.0756	0.3024	0.5177	54.75	0.95	1.14
KECRS	0.0332	0.0078	0.1893	0.3799	0.6531	58.97	0.63	0.64	0.0133	0.0051	0.0473	0.1532	0.3188	59.35	0.59	0.71
RevCore	0.0467	0.0136	0.4513	1.0932	1.6631	51.03	1.06	1.11	0.0252	0.0098	0.0769	0.3065	0.5283	54.43	0.98	1.15
UCCR w/o En	0.0465	0.0138	0.4349	1.0289	1.5543	51.33	1.02	1.08	0.0245	0.0089	0.0729	0.3001	0.5082	54.95	0.96	1.12
UCCR w/o Wo	0.0478	0.0141	0.5093	1.2239	1.8583	50.68	1.07	1.14	0.0253	0.0097	0.0801	0.3195	0.5493	54.01	1.00	1.18
UCCR w/o It	0.0481	0.0142	0.5217	1.2589	1.9122	50.34	1.08	1.16	0.0255	0.0103	0.0815	0.3255	0.5561	53.56	1.03	1.18
UCCR	**0.0494***	**0.0145***	**0.5365***	**1.2783***	**1.9376***	**50.21***	**1.13***	**1.18***	**0.0257***	**0.0106***	**0.0818***	**0.3289***	**0.5635***	**53.24***	**1.06***	**1.22***

Experiments

Table 3: Results of cold-start scenarios on ReDial with different number of user's current entities.

#Entity	Method	H@10	H@50	M@10	M@50	N@10	N@50
0	RevCore	10.23	26.31	0.0317	0.0409	0.0483	0.0799
	UCCR	**11.61**	**28.36**	**0.0384**	**0.0471**	**0.0574**	**0.0906**
1	RevCore	23.88	41.76	0.1094	0.1186	0.1377	0.1764
	UCCR	**24.69**	**43.93**	**0.1153**	**0.1231**	**0.1409**	**0.1830**
2	RevCore	22.65	41.92	0.0939	0.1045	0.1271	0.1693
	UCCR	**23.44**	**42.12**	**0.0996**	**0.1084**	**0.1313**	**0.1725**
3	RevCore	23.15	44.69	0.0859	0.0967	0.1202	0.1684
	UCCR	**23.41**	**44.95**	**0.0886**	**0.0987**	**0.1214**	**0.1703**
≥ 6	RevCore	18.63	40.77	0.0789	0.0898	0.1048	0.1562
	UCCR	**19.28**	**41.64**	**0.0829**	**0.0942**	**0.1116**	**0.1617**



(a) MRR&NDCG results for new users. (b) HR results for new users.

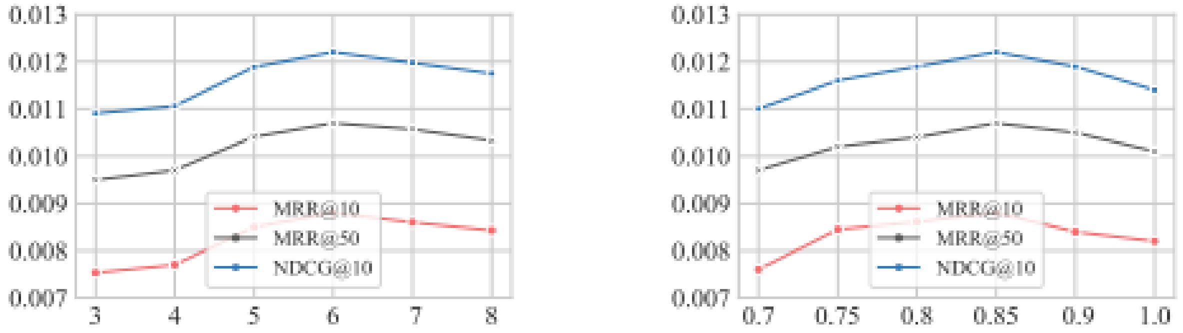
(c) MRR&NDCG results for old users. (d) HR results for old users.

Figure 3: The results for cold-start historical sessions scenario.

Experiments

Table 4: Ablation study for three aspects.

	H@10	H@50	M@10	M@50	N@10	N@50
UCCR w/o En	1.67	5.06	0.0071	0.0085	0.0092	0.0165
+ Current	1.95	6.12	0.0076	0.0093	0.0103	0.0192
+ Historical	2.14	6.33	0.0082	0.0101	0.0114	0.0204
+ Look-alike	2.32	6.64	0.0088	0.0107	0.0122	0.0214



(a) τ_e for historical aspect.

(b) δ_e for look-alike aspect.

Figure 4: Hyper-parameters sensitive analysis for historical entities and look-alike users on TG-Redial.

Thank you!