Intent Contrastive Learning for Sequential Recommendation (WWW_2022)

Yongjun Chen
Salesforce Research
Palo Alto, CA, USA
yongjun.chen@salesforce.com

Zhiwei Liu
Salesforce Research
Palo Alto, CA, USA
zhiweiliu@salesforce.com

Jia Li
Salesforce Research
Palo Alto, CA, USA
jia.li@salesforce.com

Julian McAuley
UC San Diego
La Jolla, CA, USA
jmcauley@eng.ucsd.edu

Caiming Xiong
Salesforce Research
Palo Alto, CA, USA
cxiong@salesforce.com

2022. 4. 17 • ChongQing

gesis Leibniz-Institut für Sozialwissenschaften

Reported by Lele Duan

Code & dataset:https://github.com/salesforce/ICLRec

1.Background

2.Method

3.Experiments

Background

- Users' interactions with items are driven by various intents:
  - E.g.: Preparing for holiday gifts, shopping for fishing equipment, etc.
- However, users' underlying intents are ofte unobserved/latent.
- To investigate the benefits of latent intents and leverage them effectively for recommendation.

Figure 1: Users' purchasing behaviors can be driven by underlying intents that are not observed.

➢ Over view

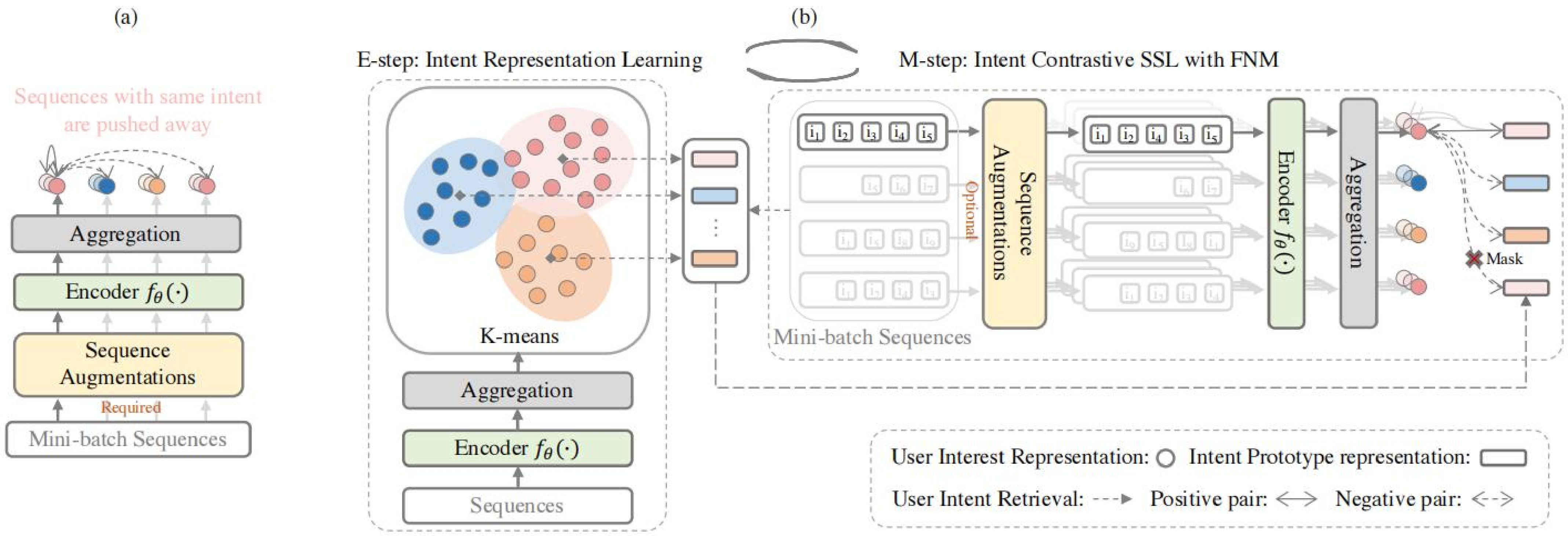

Figure 2: Overview of ICL. (a) An individual sequence level SSL for SR. (b) The proposed ICL for SR. It alternately performs intent representation learning and intent contrastive SSL with FNM within the generalized EM framework to maximizes mutual information (MIM) between a behavior sequence and its corresponding intent prototype.

Method

Deep SR Models for Next Item Prediction

$$\theta^* = \arg\max_{\theta} \sum_{u=1}^{N} \sum_{t=2}^{T} \ln P_\theta(s_t^u). \tag{1}$$

which is equivalent to minimizing the adapted binary cross-entropy loss as follows:

$$\mathcal{L}_{\text{NextItem}} = \sum_{u=1}^{N} \sum_{t=2}^{T} \mathcal{L}_{\text{NextItem}}(u, t), \tag{2}$$

$$\mathcal{L}_{\text{NextItem}}(u,t) = -\log(\sigma(\mathbf{h}_{t-1}^u \cdot \mathbf{s}_t^u)) - \sum_{neg} \log(1 - \sigma(\mathbf{h}_{t-1}^u \cdot \mathbf{s}_{neg}^u)), \tag{3}$$

where $\mathbf{s}_t^u$ and $\mathbf{s}_{neg}^u$ denote the embeddings of the target item s_t and all items not interacted by u. The sum operator in Eq. 3 is computationally expensive because $|V|$ is large. Thus we follow [3,

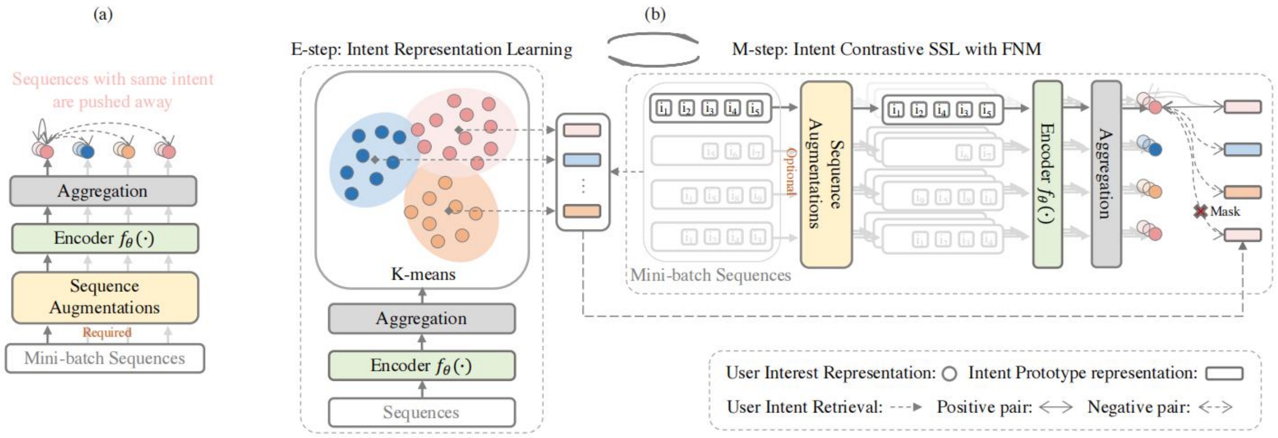

Figure 2: Overview of ICL. (a) An individual sequence level SSL for SR. (b) The proposed ICL for SR. It alternately performs intent representation learning and intent contrastive SSL with FNM within the generalized EM framework to maximizes mutual information (MIM) between a behavior sequence and its corresponding intent prototype.

Contrastive SSL in SR

Contrastive SSL in SR

$$\tilde{S}_1^u = g_1^u(S^u), \tilde{S}_2^u = g_2^u(S^u), \text{ s.t. } g_1^u, g_2^u \sim \mathcal{G}, \tag{4}$$

where g_1^u and g_2^u are transformation functions sampled from $\mathcal{G}$ to create a different view of sequence s_u. Commonly, views created

encoder $f_\theta(\cdot)$: $\tilde{\mathbf{H}}_1^u$ and $\tilde{\mathbf{H}}_2^u$

'Aggregation' layer : $\tilde{\mathbf{h}}_1^u$ and $\tilde{\mathbf{h}}_2^u$

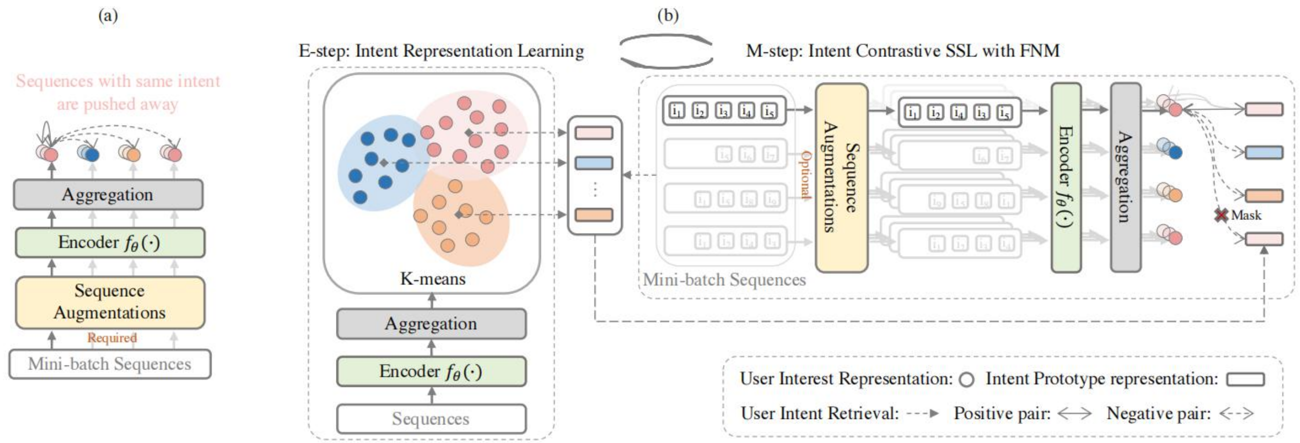

Figure 2: Overview of ICL. (a) An individual sequence level SSL for SR. (b) The proposed ICL for SR. It alternately performs intent representation learning and intent contrastive SSL with FNM within the generalized EM framework to maximizes mutual information (MIM) between a behavior sequence and its corresponding intent prototype.

optimize θ via InfoNCE loss:

$$\mathcal{L}_{\text{SeqCL}} = \mathcal{L}_{\text{SeqCL}}(\tilde{\mathbf{h}}_1^u, \tilde{\mathbf{h}}_2^u) + \mathcal{L}_{\text{SeqCL}}(\tilde{\mathbf{h}}_2^u, \tilde{\mathbf{h}}_1^u), \tag{5}$$

and

$$\mathcal{L}_{\text{SeqCL}}(\tilde{\mathbf{h}}_1^u, \tilde{\mathbf{h}}_2^u) = -\log \frac{\exp(\text{sim}(\tilde{\mathbf{h}}_1^u, \tilde{\mathbf{h}}_2^u))}{\sum_{neg} \exp(\text{sim}(\tilde{\mathbf{h}}_1^u, \tilde{\mathbf{h}}_{neg}))}, \tag{6}$$

where $sim(\cdot)$ is dot product and $\tilde{\mathbf{h}}_{neg}$ are negative views' representations of sequence S^u. Figure 2 (a) illustrates how SeqCL works.

Intent Contrastive Learning

K latent intent prototypes $\{c_i\}_{i=1}^{K}$ that affect users' decisions

$$\theta^* = \arg\max_{\theta} \boxed{\sum_{u=1}^{N}\sum_{t=2}^{T} \ln P_\theta(s_t^u).} \quad (1)$$

$$\theta^* = \arg\max_{\theta} \boxed{\sum_{u=1}^{N}\sum_{t=1}^{T} \ln \mathbb{E}_{(c)}\left[P_\theta(s_t^u, c_i)\right]}, \quad (8)$$

which is however hard to optimize. Instead, we construct a lower-bound function of Eq. (8) and maximize the lower-bound. Formally, assume intent c follows distribution $Q(c)$, where $\sum_c Q(c_i) = 1$ and

$Q(c_i) \geq 0$. Then we have

$$\begin{aligned}\sum_{u=1}^{N}\sum_{t=1}^{T} \ln \mathbb{E}_{(c)}\left[P_\theta(s_t^u, c_i)\right] &= \sum_{u=1}^{N}\sum_{t=1}^{T} \ln \sum_{i=1}^{K} P_\theta(s_t^u, c_i) \\ &= \sum_{u=1}^{N}\sum_{t=1}^{T} \ln \sum_{i=1}^{K} Q(c_i)\frac{P_\theta(s_t^u, c_i)}{Q(c_i)}.\end{aligned} \quad (9)$$

Based on the Jensen's inequality, the term in Eq. (9) is

$$\begin{aligned}&\geq \sum_{u=1}^{N}\sum_{t=1}^{T}\sum_{i=1}^{K} Q(c_i) \ln \frac{P_\theta(s_t^u, c_i)}{Q(c_i)} \\ &\propto \sum_{u=1}^{N}\sum_{t=1}^{T}\sum_{i=1}^{K} Q(c_i) \cdot \ln P_\theta(s_t^u, c_i),\end{aligned} \quad (10)$$

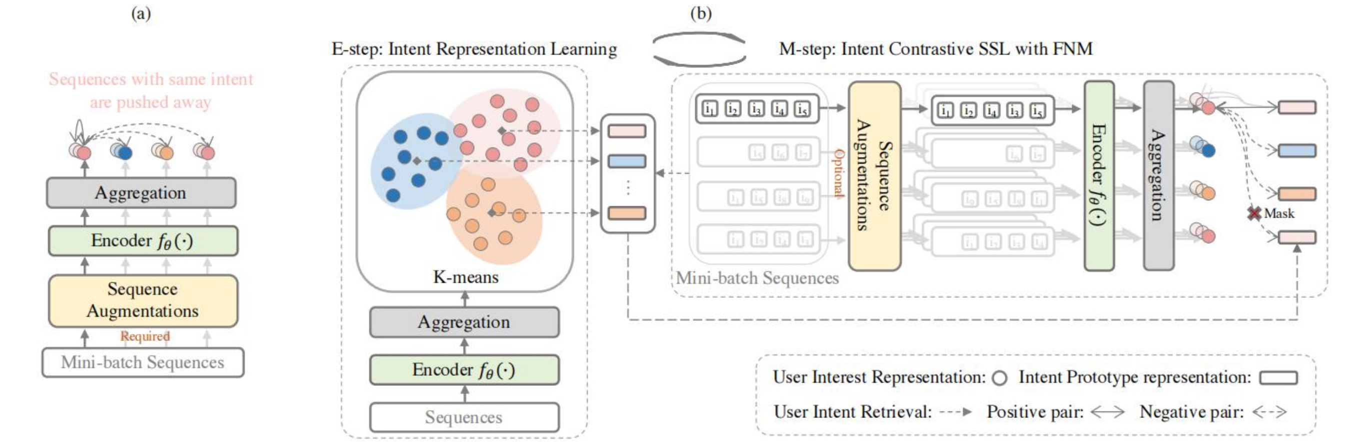

Figure 2: Overview of ICL. (a) An individual sequence level SSL for SR. (b) The proposed ICL for SR. It alternately performs intent representation learning and intent contrastive SSL with FNM within the generalized EM framework to maximizes mutual information (MIM) between a behavior sequence and its corresponding intent prototype.

where the $\propto$ stands for 'proportional to' (i.e. up to a multiplicative constant). The inequality will hold with equality when $Q(c_i) = P_\theta(c_i|s_t^u)$. For simplicity, we only focus on last positional step when optimize the lower-bound, which is defined as:

$$\boxed{\sum_{u=1}^{N}\sum_{i=1}^{K} Q(c_i) \cdot \ln P_\theta(S^u, c_i),} \quad (11)$$

where $Q(c_i) = P_\theta(c_i|S^u)$.

- ## Intent Representation Learning

K-means clustering over all sequence representations $\{\mathbf{h}^u\}_{u=1}^{|\mathcal{U}|}$

to obtain K clusters. After that, we can define the distribution function $Q(c_i)$ as follows:

$$Q(c_i) = P_\theta(c_i|S^u) = \begin{cases} 1 & \text{if } S^u \text{ in cluster } i \\ 0 & \text{else.} \end{cases} \tag{12}$$

We denote $\mathbf{c_i}$ as the vector representation of intent c_i, which is the centroid representation of the i^{th} cluster. In this paper, we use 'aggregation layer' to denote the the mean pooling operation over all position steps for simplicity. We leave other advanced aggregation methods such as attention-based methods for future work studies. Figure 2 (b) illustrates how the E-step works.

- ## Intent Contrastive SSL with FNM

$$\begin{aligned} P_\theta(S^u, c_i) &= P_\theta(c_i) P_\theta(S^u|c_i) = \frac{1}{K} \cdot P_\theta(S^u|c_i) \\ &\propto \frac{1}{K} \cdot \frac{\exp(-(\mathbf{h}^u - \mathbf{c}_i)^2)}{\sum_{j=1}^{K} \exp(-(\mathbf{h}_i^u - \mathbf{c}_j)^2)} \\ &\propto \frac{1}{K} \cdot \frac{\exp(\mathbf{h}^u \cdot \mathbf{c}_i)}{\sum_{j=1}^{K} \exp(\mathbf{h}^u \cdot \mathbf{c}_j)}, \end{aligned} \tag{13}$$

$$\sum_{u=1}^{N} \sum_{i=1}^{K} Q(c_i) \cdot \ln P_\theta(S^u, c_i), \tag{11}$$

where $Q(c_i) = P_\theta(c_i|S^u)$.

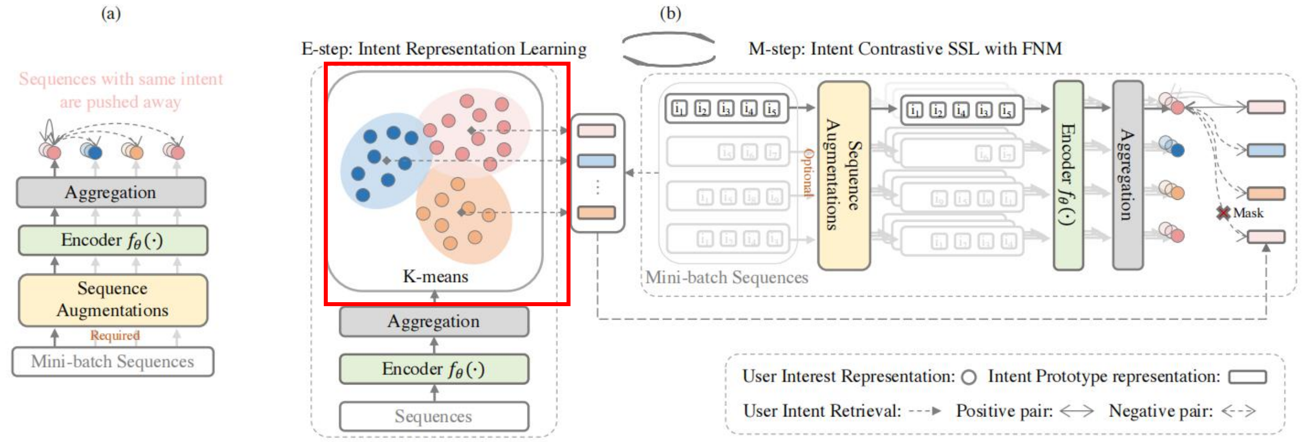

Figure 2: Overview of ICL. (a) An individual sequence level SSL for SR. (b) The proposed ICL for SR. It alternately performs intent representation learning and intent contrastive SSL with FNM within the generalized EM framework to maximizes mutual information (MIM) between a behavior sequence and its corresponding intent prototype.

where $\mathbf{h}^u$ and $\mathbf{c}_u$ are vector representations of S^u and c_i, respectively. Based on Eq. (11), (12), (13), maximizing Eq. (11) is equivalent to minimize the following loss function:

$$-\sum_{v=1}^{N} \log \frac{\exp(\text{sim}(\mathbf{h}^u, \mathbf{c}_i))}{\sum_{j=1}^{K} \exp(\text{sim}(\mathbf{h}^u, \mathbf{c}_j))}, \tag{14}$$

Intent Contrastive Learning

$$\mathcal{L}_{\mathrm{ICL}} = \mathcal{L}_{\mathrm{ICL}}(\tilde{\mathbf{h}}_1^u, \mathbf{c}_u) + \mathcal{L}_{\mathrm{ICL}}(\tilde{\mathbf{h}}_2^u, \mathbf{c}_u), \tag{15}$$

and

$$\mathcal{L}_{\mathrm{ICL}}(\tilde{\mathbf{h}}_1^u, \mathbf{c}_u) = -\log \frac{\exp(\mathrm{sim}(\tilde{\mathbf{h}}_1^u, \mathbf{c}_u))}{\sum_{neg} \exp(\mathrm{sim}(\tilde{\mathbf{h}}_1^u, \mathbf{c}_{neg}))}, \tag{16}$$

where c_{neg} are all the intents in the given batch. However, directly optimizing Eq. (16) can introduce false-negative samples since users in a batch can have same intent. To mitigate the effects of false-negatives, we propose a simple strategy to mitigate the effects by

not contrasting against them:

$$\mathcal{L}_{\mathrm{ICL}}(\tilde{\mathbf{h}}_1^u, \mathbf{c}_u) = -\log \frac{\exp(\mathrm{sim}(\tilde{\mathbf{h}}_1^u, \mathbf{c}_u))}{\sum_{v=1}^{N} \mathbb{1}_{v \notin \mathcal{F}} \exp(\mathrm{sim}(\tilde{\mathbf{h}}_1, \mathbf{c}_v))}, \tag{17}$$

where $\mathcal{F}$ is a set of users that have same intent as u in the mini-batch. We term this **False-Negative Mitigation (FNM)**. Figure 2 (b) illustrates how the M-step works.

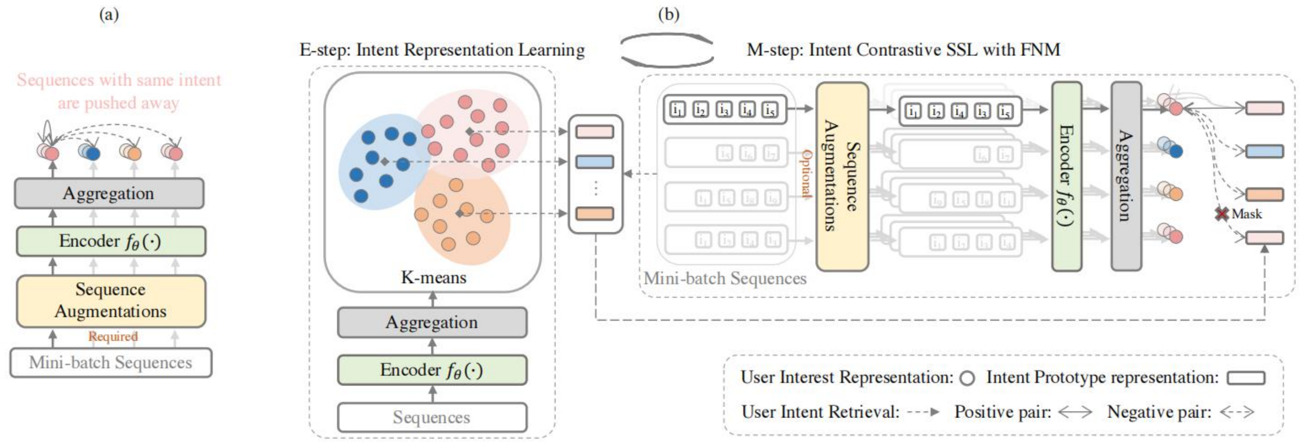

Figure 2: Overview of ICL. (a) An individual sequence level SSL for SR. (b) The proposed ICL for SR. It alternately performs intent representation learning and intent contrastive SSL with FNM within the generalized EM framework to maximizes mutual information (MIM) between a behavior sequence and its corresponding intent prototype.

Multi-Task Learning

$$\mathcal{L} = \mathcal{L}_{\mathrm{NextItem}} + \lambda \cdot \mathcal{L}_{\mathrm{ICL}} + \beta \cdot \mathcal{L}_{\mathrm{SeqCL}}, \tag{18}$$

where λ and β control the strengths of the ICL task and sequence level SSL tasks, respectively. Appendix A provides the pseudo-code of the entire learning pipeline. Specially, we build the learning paradigm on Transformer [13, 40] encoder to form the model ICLRec.

Experiment

Table 1: Performance comparisons of different methods. The best score is bolded in each row, and the second best is underlined. The last two columns are the relative improvements compared with the best baseline results.

Dataset	Metric	BPR	GRU4Rec	Caser	SASRec	DSSRec	BERT4Rec	S^3-Rec_{ISP}	CL4SRec	ICLRec	Improv.
Sports	HR@5	0.0141	0.0162	0.0154	0.0206	0.0214	0.0217	0.0121	0.0217±0.0021	**0.0283**±0.0006	30.48%
	HR@20	0.0323	0.0421	0.0399	0.0497	0.0495	0.0604	0.0344	0.0540±0.0024	**0.0638**±0.0023	18.15%
	NDCG@5	0.0091	0.0103	0.0114	0.0135	0.0142	0.0143	0.0084	0.0137±0.0013	**0.0182**±0.0001	33.33%
	NDCG@20	0.0142	0.0186	0.0178	0.0216	0.0220	0.0251	0.0146	0.0227±0.0016	**0.0284**±0.0008	24.89%
Beauty	HR@5	0.0212	0.0111	0.0251	0.0374	0.0410	0.0360	0.0189	0.0423±0.0031	**0.0493**±0.0013	16.43%
	HR@20	0.0589	0.0478	0.0643	0.0901	0.0914	0.0984	0.0487	0.0994±0.0028	**0.1076**±0.0001	8.30%
	NDCG@5	0.0130	0.0058	0.0145	0.0241	0.0261	0.0216	0.0115	0.0281±0.0018	**0.0324**±0.0017	15.51%
	NDCG@20	0.0236	0.0104	0.0298	0.0387	0.0403	0.0391	0.0198	0.0441±0.0018	**0.0489**±0.0013	10.90%
Toys	HR@5	0.0120	0.0097	0.0166	0.0463	0.0502	0.0274	0.0143	0.0526±0.0034	**0.0590**±0.0012	12.07%
	HR@20	0.0312	0.0301	0.0420	0.0941	0.0975	0.0688	0.0235	0.1038±0.0041	**0.1150**±0.0016	10.74%
	NDCG@5	0.0082	0.0059	0.0107	0.0306	0.0337	0.0174	0.0123	0.0362±0.0025	**0.0403**±0.0002	11.34%
	NDCG@20	0.0136	0.0116	0.0179	0.0441	0.0471	0.0291	0.0162	0.0506±0.0025	**0.0560**±0.0004	10.57%
Yelp	HR@5	0.0127	0.0152	0.0142	0.0160	0.0171	0.0196	0.0101	0.0229±0.0003	**0.0257**±0.0007	12.23%
	HR@20	0.0346	0.0371	0.0406	0.0443	0.0464	0.0564	0.0314	0.0630±0.0009	**0.0677**±0.0016	7.47%
	NDCG@5	0.0082	0.0091	0.008	0.0101	0.0112	0.0121	0.0068	0.0144±0.0001	**0.0162**±0.0003	12.50%
	NDCG@20	0.0143	0.0145	0.0156	0.0179	0.0193	0.0223	0.0127	0.0256±0.0003	**0.0279**±0.0006	8.98%

Experiment

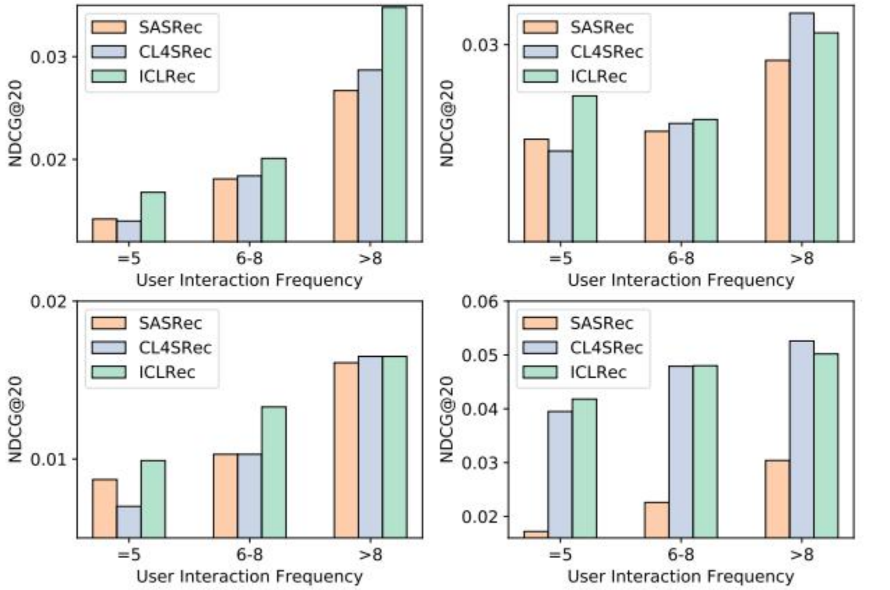

Figure 3: Performance comparison on different user groups among SASRec, CL4SRec and ICLRec (Upper left: Beauty, Upper right: Yelp, Lower left: Sports, Lower right: Toys).)

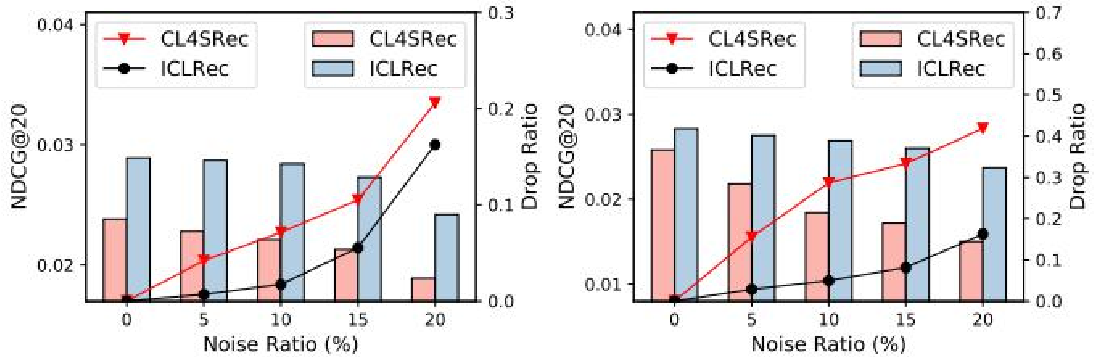

Figure 4: Performance comparison w.r.t. noise ratio on Sports and Yelp. The bar chart shows the performance in NDCG@5 and the line chart shows the corresponding drop rate.

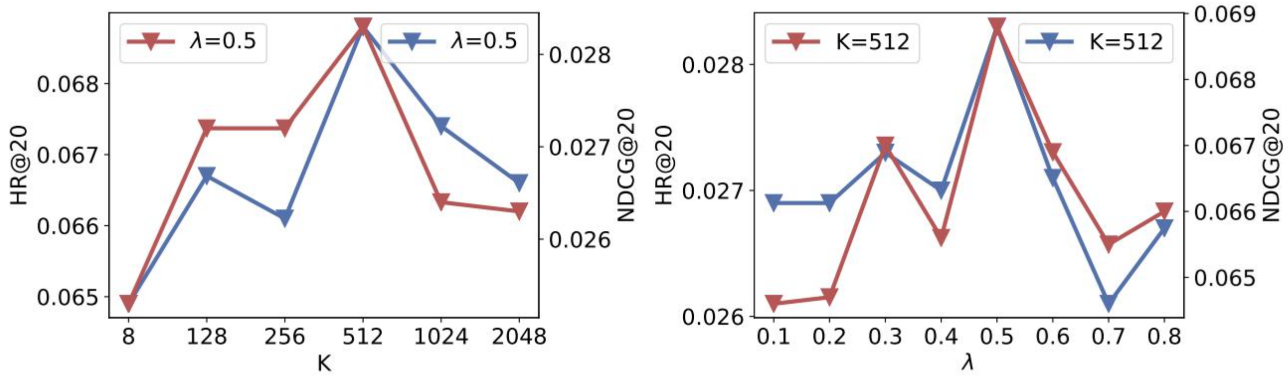

Figure 5: Impact of intent class numbers K and the intent contrastive learning strength λ on Yelp.

Experiment

Table 2: Ablation study of ICLRec (NDCG@20).

Model	Dataset			
	Sports	Beauty	Toys	Yelp
(A) ICLRec	**0.0287**	**0.0480**	**0.0554**	**0.0283**
(B) w/o FNM	0.0283	0.0465	0.0524	0.0266
(C) only ICL	0.0263	0.0429	0.0488	0.0267
(D) w/o ICL	0.0238	0.0428	0.0505	0.0258
(E), is (C) w/o seq. aug	0.0242	0.0414	0.0488	0.0213
(F) SASRec	0.0216	0.0387	0.0441	0.0179
(G) ICL + S^3-Rec_{ISP}	**0.0157**	**0.0264**	**0.0266**	**0.0205**
(H) S^3-Rec_{ISP}	0.0146	0.0198	0.0162	0.0127

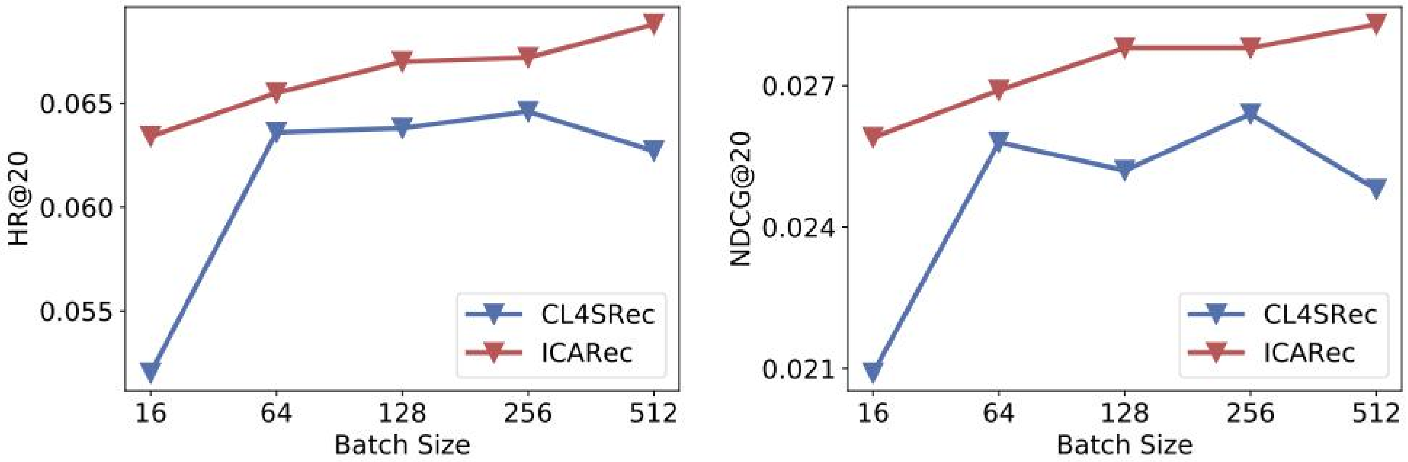

Figure 6: Performance comparison w.r.t. Batch Size.

Thank you!