Self-Supervised Interest Transfer Network via Prototypical Contrastive Learning for Recommendation

[1]College of Computer Science and Technology, Zhejiang University, China
[2]Alibaba Group, China
{sungq, wasdnsxchen, hyliu20, wuchunming}@zju.edu.cn,
{shilou.syb, zhousijin.zsj, chenyi.lcy, xianhui.wxh, mingyi.ff}@alibaba-inc.com

code:https://github.com/fanqieCoffee/SITN-Supplement.

AAAI 2023

Reported by Minqin Li

Introduction

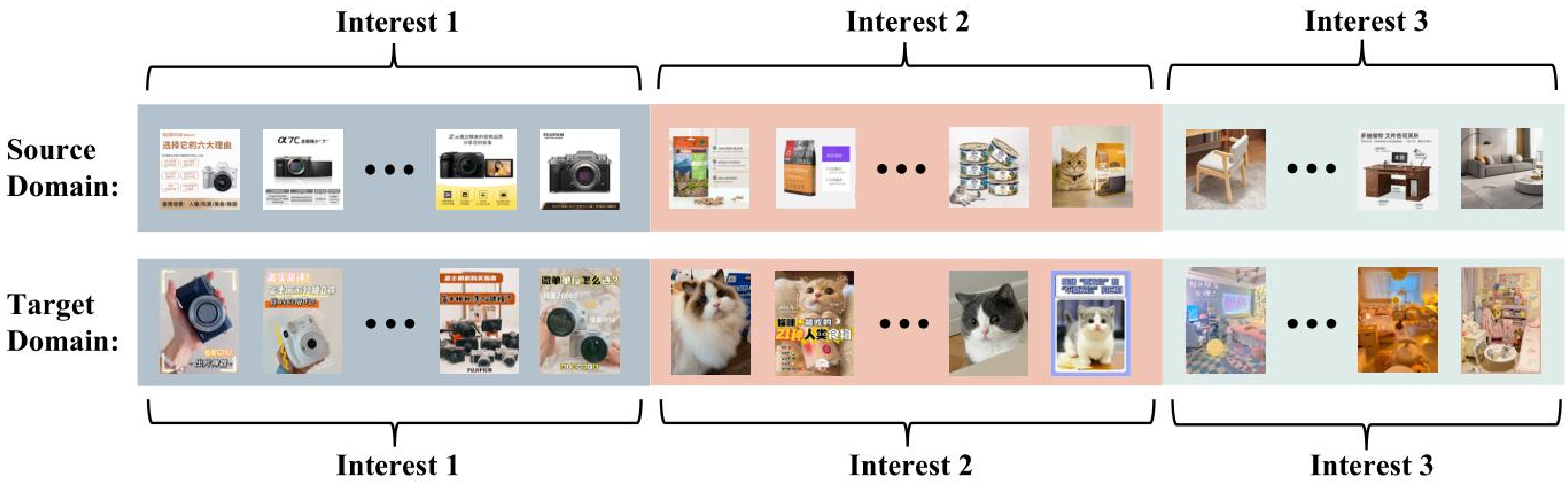


Figure 1: The user clicking sequences in a real-life scenario. The upper part is the user's clicking sequence in a product recommendation platform (source domain), and the lower part is the user's clicking sequence in a content/micro-video recommendation platform (target domain). The user has three interests in both domains and these interests correspond to each other.

Method

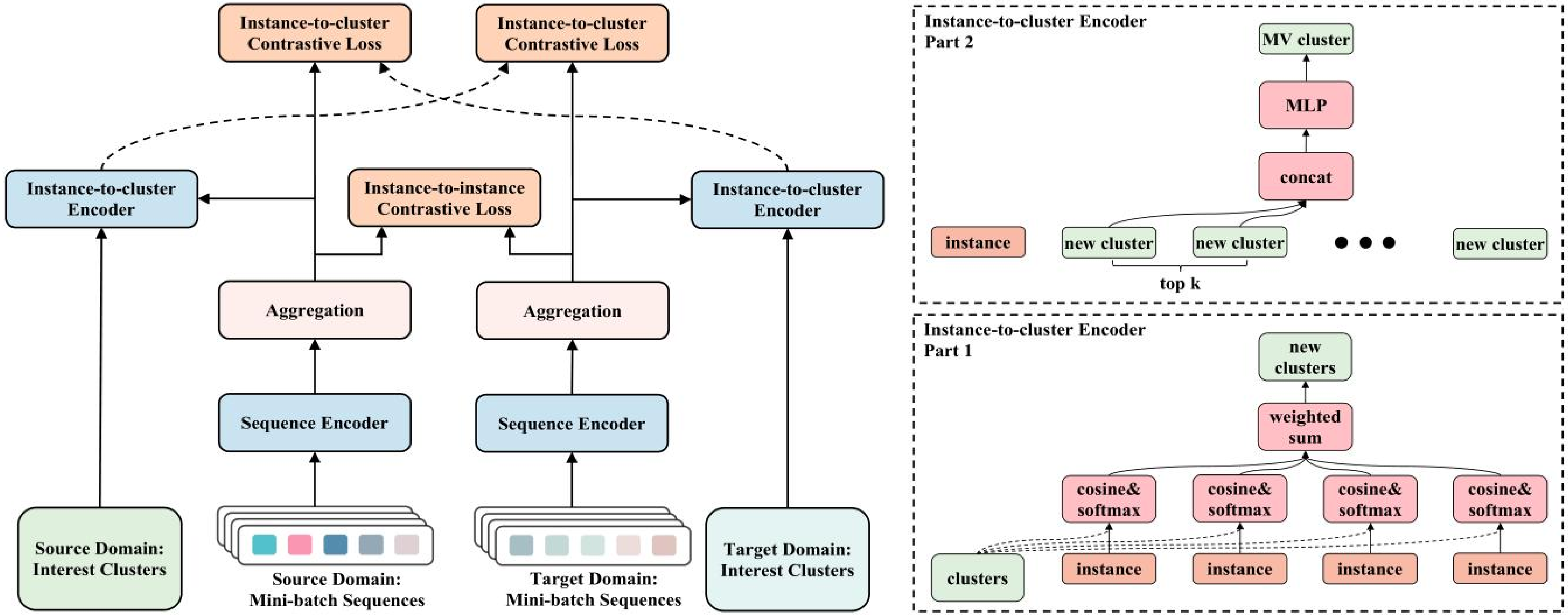
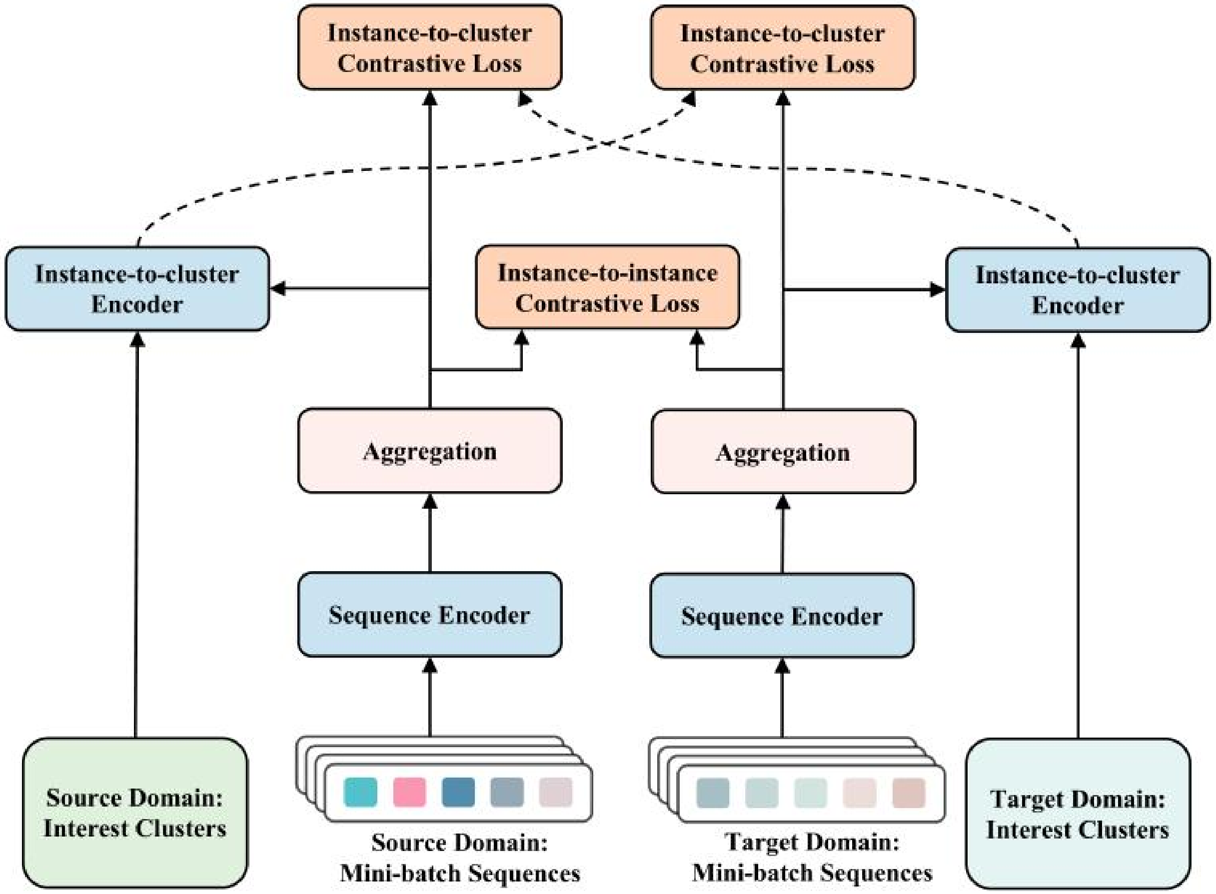


Figure 2: The overall framework of SITN's contrastive learning module. The left part shows the contrastive learning module at a high level. The right part shows the instance-to-cluster encoder. The role of the first part of instance-to-cluster encoder is to generate new clusters from instances and original clusters. The lower right part shows a example of clusters and four instances. The role of the second part of instance-to-cluster encoder is to generate MV clusters from instances and new clusters. The upper right part shows a example of new clusters and one instance, and the value of k is two.

Method



$$Attention(Q,K,V) = softmax(\frac{Q^T K}{\sqrt{d_k}})V \quad (1)$$

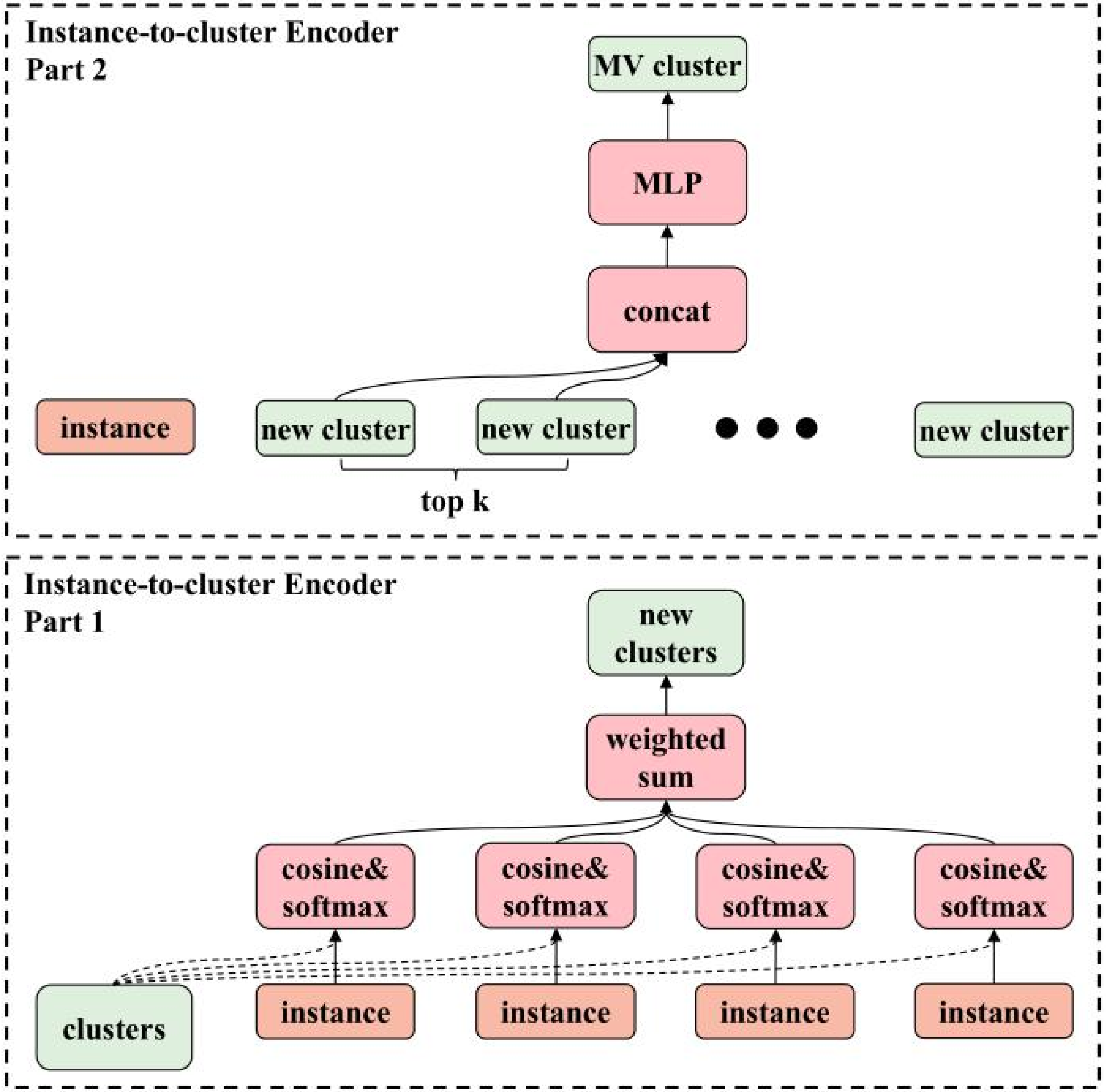
$$p_i^s = g(f_s(x_i^s)), \qquad p_i^t = g(f_t(x_i^t)) \quad (2)$$

$$\mathcal{L}_{i2i}^{s2t} = -\sum_{k=1}^{n} log \frac{exp(sim(p_k^s, p_k^t)/\tau)}{\sum_{i=1}^{n} exp(sim(p_k^s, p_i^t)/\tau)} \quad (3)$$

$$\mathcal{L}_{i2i}^{t2s} = -\sum_{k=1}^{n} log \frac{exp(sim(p_k^t, p_k^s)/\tau)}{\sum_{i=1}^{n} exp(sim(p_k^t, p_i^s)/\tau)} \quad (4)$$

$$\mathcal{L}_{i2i} = \mathcal{L}_{i2i}^{s2t} + \mathcal{L}_{i2i}^{t2s} \quad (5)$$

Method



$$u_k^s = \sum_{i=1}^{n} \pi_i^s(k) \cdot p_i^s,$$

$$\pi_i^s(k) = \frac{exp(sim(c_k^s, p_i^s))}{\sum_{k'=1}^{K} exp(sim(c_{k'}^s, p_i^s))} \tag{6}$$

$$q_i^s = MLP(Concat(u_{i1}^s, u_{i2}^s, \cdots, u_{ik}^s)) \tag{7}$$

$$q_i^t = MLP(Concat(u_{i1}^t, u_{i2}^t, \cdots, u_{ik}^t)) \tag{8}$$

$$e_{i2c}^{s2t} = -\sum_{k=1}^{n} log \frac{exp(sim(p_k^s, q_k^t)/\tau)}{exp(sim(p_k^s, q_k^t)/\tau) + \sum_{i=1}^{K} exp(sim(p_k^s, u_i^t)/\tau)},$$
$$e_{i2c}^{t2s} = -\sum_{k=1}^{n} log \frac{exp(sim(p_k^t, q_k^s)/\tau)}{exp(sim(p_k^t, q_k^s)/\tau) + \sum_{i=1}^{K} exp(sim(p_k^t, u_i^s)/\tau)}, \tag{9}$$
$$e_{i2c} = e_{i2c}^{s2t} + e_{i2c}^{t2s}.$$

$$\hat{y} = MLP(Concat(u_f^s, u_f^t, v)) \tag{10}$$

$$\mathcal{L}(y, \hat{y}) = -(ylog\hat{y} + (1-y)log(1-\hat{y})) \tag{11}$$

Experiments

Dataset	**Industrial**	**Amazon**
# Shared Users	23,706,610	87,896
# Items(Source)	54,829,040	673,826
# Items(Target)	13,149,185	100,164
# Instances	445,821,389	1,290,358
Avg. # clicked(Source)	94	39
Avg. # clicked(Target)	60	36

Table 1: Statistics of two utilized datasets. (avg. - average)

Experiments

Model		Industrial AUC	Industrial Logloss	Amazon AUC	Amazon Logloss
Single-domain	FM	0.6501	0.4486	0.6215	0.4624
	Wide&Deep	0.6939	0.3701	0.7154	0.3411
	DeepFM	0.7040	0.3672	0.7201	0.3358
	DIN	0.7173	0.3408	0.7370	0.3183
	DIEN	0.7185	0.3392	0.7397	0.3062
	CLSR	0.6935	0.3740	0.6752	0.4061
Cross-domain	MV-DNN	0.7169	0.3391	0.7293	0.3285
	CoNet	0.7188	0.3382	0.7275	0.3278
	MiNet	0.7218	0.3313	0.7305	0.3260
	SEMI	0.7304	0.3276	0.7439	0.3036
	SITN	**0.7351***	**0.3207***	**0.7525***	**0.2857***

Table 2: Test AUC and Logloss. * indicates the statistical significance for $p <= 0.01$ compared with the best baseline method based on the paired t-test.

Experiments

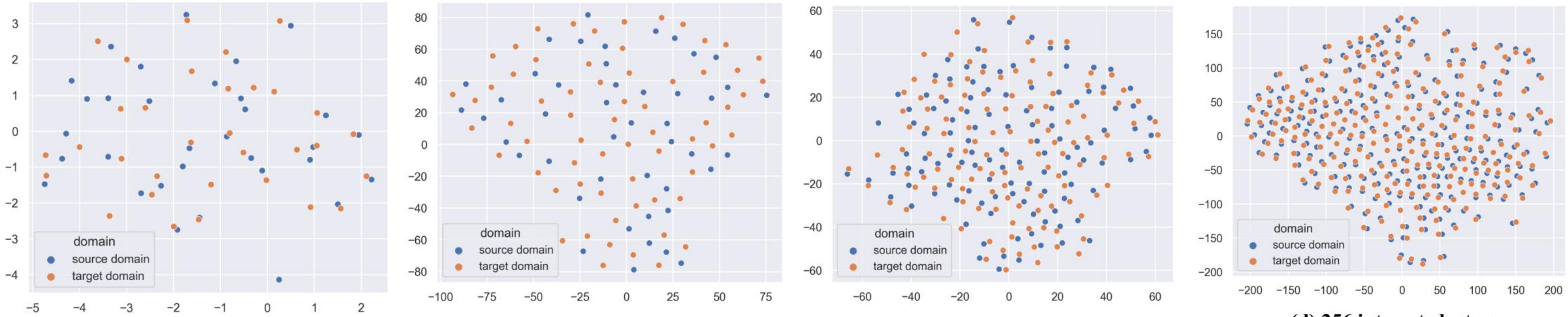
Method	Industrial AUC	Industrial Logloss	Amazon AUC	Amazon Logloss
w/o $\mathcal{L}_{i2c}$&$\mathcal{L}_{i2i}$	0.7203	0.3325	0.7328	0.3242
w/o $\mathcal{L}_{i2c}$	0.7295	0.3304	0.7456	0.3020
w/o $\mathcal{L}_{i2i}$	0.7324	0.3259	0.7489	0.2959
SITN	**0.7351**	**0.3207**	**0.7525**	**0.2857**

Table 3: Effect of instance-to-cluster contrastive learning.

Method	Industrial AUC	Industrial Logloss	Amazon AUC	Amazon Logloss
w/o MG&MV	0.7310	0.3343	0.7462	0.3013
w/o MG	0.7335	0.3236	0.7493	0.2953
w/o MV	0.7327	0.3305	0.7475	0.2965
SITN	**0.7351**	**0.3207**	**0.7525**	**0.2857**

Table 4: Effects of multi-granularity and multi-view modules.

Experiments



(a) 32 interest clusters. **(b) 64 interest clusters.** **(c) 128 interest clusters.** **(d) 256 interest clusters.**

Figure 3: Visualization of the interest clusters in the source domain and the target domain. For (a), (b), (c) and (d), the numbers of clusters in both domains are the same, 32, 64, 128 and 256, respectively.

Thanks