When Multi-scenario Meets Multi-attribute: Scenario and Attribute-aware Recommendation with Contrastive Learning

Jin Huang(✉), Yu Qian, Zixu Yang, and Zhijun Sun

Dewu, Shanghai, China
{huangjin01,qianyu01,yangzixu,zhangliang1}@shizhuang-inc.com

Abstract. Recently, multi-scenario learning (MSL) has achieved flourishing development in recommendation systems of E-commerce platforms. Current numerous models have been proposed that attempt to use a unified model to serve multiple scenarios. In these works, user's sequential behavior modeling is infrastructure for capturing the dynamic nature of the user profiles and ensuring their predictive performance. However, recent multi-scenario learning methods usually ignore the utilization of item attributes when modeling behavior, which contain crucial prior information for deciding which items to recommend. Specifically, considering item IDs only in the model will lead to insufficient characterization of interests especially in highly sparse recommender settings. Meanwhile, the lack of abundant attribute information is not conducive to model the significant interests discrepancy among multiple scenarios. In this paper, we address these limitations by proposing a **S**cenario and **A**ttribute-aware **C**ontrastive **N**etwork (**SACN**) for multi-scenario learning. First, SACN employs Item-level Preference Extracting (IPE) and Attribute-level Preference Extracting (APE) modules capturing user's coarse-grained and fine-grained preferences dynamically in terms of scenario context and item attributes. Then, by introducing self-supervised learning in Scenario Contrastive Module (SCM), we can distinguish users' interests in different scenarios sufficiently and achieve superior performance. Offline experiments on real-world datasets and online A/B testing both show that the proposed model significantly outperforms all state-of-the-art models in the task of multi-scenario recommendation (MSR).

Keywords: Click-through Rate Prediction · Multi-scenario Learning · User Interest Modeling.

1 Introduction

With rapid development of the E-commerce platforms, recommendation systems play an increasingly critical role in boosting business revenue and improving users' online experience. Naturally, multiple shopping scenarios are rapidly developed to meet the diversified needs of users. At large commercial companies like

Taobao and Dewu[1], there are diverse business scenarios (e.g., home page feed, banner feed) [22]. For users, a user may visit some of the scenarios and click or buy different items. For items, an item may be recommended to different users in different scenarios. Clearly, the user behaviors can vary greatly across different scenarios (e.g., users have diverse preferences for price or category of items) [24,10]. Therefore, it is important to model distinctions in multi-scenarios learning. Moreover, item attributes plays an essential role in sequential recommendation tasks (e.g., CTR prediction) [28,13]. These additional attributes proved indispensable in getting high-quality recommendations in highly sparse settings with rich item attributes such as in online fashion stores [4] and unique item recommendation settings like online auctions [13].

From the perspective of multi-scenario learning (MSL), various types of strategies have been proposed to tackle the multi-scenario problem: (1) Training a separate model for each scenario[6,2,3,26]. The commonalities among multiple scenarios is neglected in this method. It is challenging for new and minor scenarios with limited data to train models sufficiently. Besides, developing a separate model for each scenario will require more complex computation and maintenance cost. (2) Training a unified model with whole scenario samples and treating each scenario as a task, which is inspired by multi-task learning (MTL) methods [14,27,7,5,15,9,25]. This strategy requires specific network for each scenario, which adds excessive parameters with the increase of scenarios. (3) Training a unified model with scenario related features as independent input and employing dynamic weighting operations on the bottom-level embedding and top-level network hidden units to model inter-scenario distinctions [20,22,21,1]. This strategy has become mainstream to solve multi-scenario problem as it is more flexible and parameter-efficient than MTL based approaches.

In the above multi-scenario learning (MSL) methods, effective user's behavior modeling is fundamental to ensure good recommendation performance. However, existing approaches use item IDs only when modeling behaviors and do not consider the impact of item's multiple attributes. On the one hand, multiple attributes are indispensable to generate rich latent interest representations and overcome highly sparse settings achieving superior performance. Sparsity is very significant in multi-scenario recommendation problems. For example, in some small and new scenarios, users generate very few historical behaviors and the amount of sample data is very small. On the other hand, users in different scenarios have significant interests discrepancy, which can be explicitly reflected in item's attributes. For instance, users in home page of an E-commerce APP prefer browsing a variety of categories and brands of items as they have more divergent preference in this scenario. Conversely, users in clicked items' detailed pages will have plainer preference and they like to view recommended items which are more related to the clicked ones (e.g., having the same category, brand or price range). In short, the inseparable relationship between the item's attributes and scenarios urgently needs to be modeled.

[1] `https://www.poizon.com/`

From the perspective of attribute-aware sequential modeling, current state-of-the-art approaches use self attention mechanism to fuse attribute information. we categorize them into three types, i.e., early fusion, late fusion and hybrid fusion. Early fusion methods (e.g., $\mathrm{SASRec_F}$ [28], CARCA [12]) first combine item IDs and attributes, then feed them into the self attention block to generate outputs. In late fusion [23], the networks for processing item IDs and attributes are independent. Early fusion cannot resist noise interference and may result in information invasion, while late fusion lacks effective interaction between item IDs and attributes. Hybrid fusion (e.g., NOVA [8], DIF-SR [18] and ASIF [17]) addresses the above limitations, allowing item IDs and attributes to interact in the middle layer. As far as we know, none of the existing attribute-aware models utilize target scenario and item context, which may result in a loss of information restricting the expressive power of the network.

To address these issues, we propose a novel method called SACN, an scenario and attribute-aware model for MSR, which involves both the scenario and attribute information to capture more comprehensive preferences of user in different scenarios and learn discrepancy among scenarios explicitly. Concretely, we leverage multi-head self attention mechanism as the backbone to model user's coarse-grained (item-level) and fine-grained (attribute-level) preferences dynamically in terms of target scenario context and item's attributes. Subsequently, To model the significant interests discrepancy among multiple scenarios, we utilize the Scenario Contrastive Module (SCM) which explicitly distinguishes the differences via contrastive learning [11]. We regard item-level and attribute-level preference representations of current target scenario as positive contrast samples, item-level preference representations of other scenarios as negative samples. We perform contrastive learning between the positive and negative samples, i.e., enhancing the similarity score between the two positive samples and weakening the similarity score between the negative sample and the two positive samples. To summarize, in this paper we make the following contributions:

(1) To the best of our knowledge, We are the first to model the multi-scenario learning problem from the perspective of fusing multi-attributes information of items, which significantly improves the effectiveness of multi-scenario recommendation tasks.

(2) We design several useful modules for SACN, which model user's multi-grained preferences in terms of scenario information and item attributes, and learn the significant interests discrepancy among multiple scenarios explicitly.

(3) We conducted extensive offline experiments on the industrial dataset and online A/B testing. Experimental results on the task of multi-scenario CTR prediction show that our proposed approach outperforms existing state-of-the-art approaches remarkably. SACN has already been deployed in our recommendation system.

2 The Approach

The overall framework of SACN is shown in Figure 1, and the details will be introduced next.

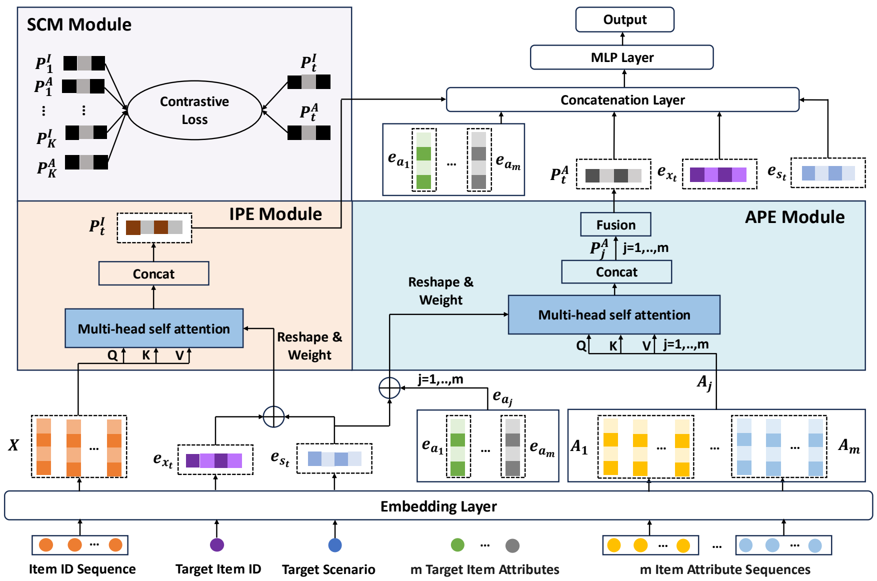


Fig. 1. Overall framework of the SACN model.

2.1 Problem Formulation

Let $\mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{A}_j$, $\mathcal{S}$ denote the sets of users, items, item IDs, the j-th type of attributes and scenarios, respectively. $\mathcal{B}^u = \{v_u^1, \dots, v_u^N\}$ represents the historical sequence of interactions in chronological order for user $u \in \mathcal{U}$, where $v_u^i \in \mathcal{V}$ is the i-th item in the user interaction sequence and N is the maximum length of the sequence. Suppose we have m types of attributes, then $v_u^i = \{x_u^i, a_{1,u}^i, \dots, a_{m,u}^i\}$, where $x_u^i \in \mathcal{X}$ is the item ID of the i-th interaction, and $a_{j,u}^i \in \mathcal{A}_j$ represents the j-th type of the attributes of the i-th interaction. Given target item $v_t \in \mathcal{V}$ and $v_t = \{x^t, a_1^t, \dots, a_m^t\}$, target scenario $s_t \in \mathcal{S}$ and the interaction history $\mathcal{B}^u$, the goal of multi-scenario recommendation is to predict the probability that the user u may be interested in target item v_t at the scenario s_t. Considering CTR prediction task, it can be formalized as: $Pr(click = 1|s_t, v_t, \mathcal{B}^u)$.

2.2 Item-level Preference Extracting

To capture user's coarse-grained (item-level) preference, Item-level Preference Extracting (**IPE**) module takes scenario-aware multi-head self-attention (MHSA) mechanism to process user history sequence of item IDs. We emphasize the importance of considering target scenario and item related information. Let $\boldsymbol{X} \in \mathbb{R}^{N\times d}$ represents the embedding matrices of item IDs, The target scenario and target item ID features are denoted as $\boldsymbol{e_{s_t}}$ and $\boldsymbol{e_{x_t}}$. The final output of scenario-aware MHSA $\hat{\boldsymbol{X}}$ is calculated as:

$$\begin{aligned} \hat{\boldsymbol{X}} &= MultiHead\left(\boldsymbol{X}\right) = concat\left(head_1, \cdots, head_h\right)\boldsymbol{W^O}, \\ head_i &= Attention\left(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}\right) = Softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_k}}\right)\boldsymbol{V} \end{aligned} \tag{1}$$

To fully interact the target item and historical behavior items, and utilize the target scenario to guide the behavior encoding, the $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}$ integrate the target scenario and item ID embeddings into self-attention parameters. The process is:

$$\begin{aligned} \boldsymbol{Q} &= \boldsymbol{X}\left(\boldsymbol{W_i^Q} \otimes Reshape\left(\boldsymbol{e_{s_t}}, \boldsymbol{e_{x_t}}\right)\right) \\ \boldsymbol{K} &= \boldsymbol{X}\left(\boldsymbol{W_i^K} \otimes Reshape\left(\boldsymbol{e_{s_t}}, \boldsymbol{e_{x_t}}\right)\right) \\ \boldsymbol{V} &= \boldsymbol{X}\left(\boldsymbol{W_i^V} \otimes Reshape\left(\boldsymbol{e_{s_t}}, \boldsymbol{e_{x_t}}\right)\right) \end{aligned} \tag{2}$$

where $\boldsymbol{W_i^Q}, \boldsymbol{W_i^K}, \boldsymbol{W_i^V} \in \mathbb{R}^{d\times d_k}$ are transformation matrix. $\otimes$ denotes element-wise product. The $Reshape\left(\boldsymbol{e_{s_t}}, \boldsymbol{e_{x_t}}\right)$ is reshaped from the embedding of target scenario and item ID, which has the same dimensions as above matrix. In this way the information from target scenario and item ID can get involved into the coarse-grained preference extracting process element-wisely and thoroughly. The final preference representation $\boldsymbol{P^I} \in \mathbb{R}^{Nd}$ of IPE is formulated as follow:

$$\boldsymbol{P_t^I} = concat\left(\hat{\boldsymbol{X}}_1, \cdots, \hat{\boldsymbol{X}}_N\right) \tag{3}$$

Similarly, preference representation of other scenarios can be calculated as $\boldsymbol{P_k^I}, k \in [1, \cdots, K]_{k\neq t}$ with replacing $\boldsymbol{e_{s_t}}$ in Equation (2) with $\boldsymbol{e_{s_k}}$, where K denotes the number of scenarios

2.3 Attribute-level Preference Extracting

Item's attributes are indispensable to capture more comprehensive preferences of users in different scenarios. However, as far as we know, none of the existing attribute-aware models utilize target scenario and item information, which may result in a loss of information restricting the expressive power of the network. Let $\boldsymbol{A_j} \in \mathbb{R}^{N\times d}$ represents the embedding matrices of the j-th item attribute of behavior, the target scenario and j-th target item attribute features are denoted as $\boldsymbol{e_{s_t}}$ and $\boldsymbol{e_{a_j}}$. Symmetrically, we replace the $\boldsymbol{X}$ in Equation (1) and $\boldsymbol{e_{x_t}}$ in Equation (2) with $\boldsymbol{A_j}$ and $\boldsymbol{e_{a_j}}$, respectively. Then Attribute-level Preference Extracting (**APE**) module can obtain m attribute-level preference representations corresponding to m types of attributes respectively. $\boldsymbol{P_j^A} \in \mathbb{R}^{Nd}$ refers to the j-th representation. To capture the user's varying appetite toward attributes (e.g., categories or brands), we fuse m attribute-level preference representations in target scenario with vanilla attention network, which is defined as follows:

$$\begin{aligned}\boldsymbol{\alpha} &= softmax\left(concat\left(\boldsymbol{P_1^A}, \cdots, \boldsymbol{P_m^A}\right)\boldsymbol{W_P} + \boldsymbol{b_P}\right) \in \mathbb{R}^m, \\ \boldsymbol{P_t^A} &= \sum_{j=1}^{m}\left(\boldsymbol{\alpha}_j \times \boldsymbol{P_j^A}\right)\end{aligned} \tag{4}$$

Similarly, we can obtain attribute-level preference representation of other scenarios as $\boldsymbol{P_k^A}$.

2.4 Scenario Contrastive Module

As mentioned before, users in different scenarios have significant interests discrepancy. Current supervised modeling approach lacks explicit supervision signals for fully distinguishing interests from various scenarios since there does not exist any annotated label of user interests. So we leverage self-supervised learning to achieve it. Different from existing methods [19,16] incorporating contrastive learning that tend to focus on complex data augmentation techniques, we analyse specific domain issues and design contrastive strategies based on raw data. Specifically, we regard item-level preference representation $\boldsymbol{P_t^I}$ and attribute-level preference representation $\boldsymbol{P_t^A}$ of current target scenario as positive contrast samples, and corresponding representations $\boldsymbol{P_k^I}, \boldsymbol{P_k^A}$ of other scenarios as negative samples. We leverage a contrastive learning loss [11] to teach the model to enhance the similarity score between the two positive samples, weaken the similarity score between the negative samples and the two positive samples. Formally, there are two contrastive tasks as follows:

$$L_{ssl1} = -log\frac{exp\left(\phi\left(\boldsymbol{P_t^I}, \boldsymbol{P_t^A}\right)/\tau\right)}{\sum_{k=1}^{K} exp\left(\phi\left(\boldsymbol{P_t^I}, \boldsymbol{P_k^I}\right)/\tau\right)} \tag{5}$$

$$L_{ssl2} = -log\frac{exp\left(\phi\left(\boldsymbol{P_t^I}, \boldsymbol{P_t^A}\right)/\tau\right)}{\sum_{k=1}^{K} exp\left(\phi\left(\boldsymbol{P_t^I}, \boldsymbol{P_k^A}\right)/\tau\right)} \tag{6}$$

where $\phi(\cdot)$ denotes the similarity function to compute cosine distance between two instances. τ is the temperature parameter. Scenario Contrastive Module (**SCM**) finally supervises the distinguishing interests among scenarios and enhances the model's proficiency in discerning distinctions across scenarios.

2.5 Prediction and Optimization

We concatenate outputs of IPE and APE, scenario features and target item features, then feed them into the multi-layer DNN tower:

$$p_t = Sigmoid\left(MLP\left(cocnat\left(\boldsymbol{P_t^I}, \boldsymbol{P_t^A}, \boldsymbol{e_{x_t}}, \boldsymbol{e_{a_j}}, \boldsymbol{e_{s_t}}\right)\right)\right) \tag{7}$$

p_t is the probability that a user will interact with target items. We used the widely used cross entropy loss as the objective function:

$$L_{main} = -\frac{1}{M}\sum_{i=1}^{M}\left(y_i log\left(p_i\right) + \left(1-y_i\right) log\left(1-p_i\right)\right) \tag{8}$$

where y_i is the label of a sample. M is the number of samples. The joint loss function with a hyper-parameter γ to balance the supervised and self-supervised objectives, can be formulated as follows:

$$L = L_{main} + \gamma\left(L_{ssl1} + L_{ssl2}\right) \tag{9}$$

3 Experiments

3.1 Experimental Setup

Datasets. We collect and sample online service logs from eight recommendation scenarios (#D1-#D8) in Dewu app between 2024/05/24 and 2024/05/31 as our experimental datasets. The item attributes in this dataset are the price, fine-grained categories, and the item's brand. All of the items' attributes are discrete and categorical. The data between 2024/05/24 and 2024/05/30 is used for training while the data in 2024/05/31 is collected for testing. Table 1 summarizes the detailed statistics of our datasets.

Table 1. Statistics of industrial dataset. (M-million)

Scenario	#D1	#D2	#D3	#D4	#D5	#D6	#D7	#D8
Users	0.76M	0.62M	0.56M	0.51M	0.43M	0.31M	0.27M	0.14M
Items	2.43M	2.31M	1.97M	1.88M	1.76M	1.62M	1.43M	1.31M
Samples	102.03M	92.40M	87.15M	82.37M	74.46M	48.81M	36.62M	29.74M

Evaluation Metrics. For the offline comparison, we use Area Under ROC Curve (AUC) as the evaluation metric, which are widely adopted in industrial recommender systems.

Competitors. To verify the effectiveness of the proposed SACN, we compare the performance of our model with a series of state-of-the-art multi-scenario learning (MSL) methods, i:e:, MMoE [9], PLE[14], M2M [22], PEPNet [1] and MARIA [15].

Implementation Details. All the models are implemented in distributed Tensorflow 1.4 and trained with 6 parameter servers and 30 workers. Embedding size for categorical features is set to 32. Adagrad optimizer with a learning rate of 0.01 and a mini batch size of 1024 is used for training. We report the results of each method under its empirically optimal hyper-parameters settings.

3.2 Overall Results

We repeat each model three times and report the averaged results. The offline comparison results are presented in Table 2 and the major observations can be summarized as follows:

Table 2. Comparison of different methods on industrial dataset.

	#D1	#D2	#D3	#D4	#D5	#D6	#D7	#D8	Overall
MMoE	0.7764	0.7802	0.7769	0.7452	0.7368	0.7386	0.7331	0.6983	0.7627
PLE	0.7778	0.7823	0.7784	0.7468	0.7376	0.7419	0.7343	0.7011	0.7649
M2M	0.7806	0.7844	0.7821	0.7492	0.7394	0.7452	0.7373	0.7035	0.7681
PEPNet	0.7817	0.7853	0.7831	0.7504	0.7408	0.7463	0.7388	0.7049	0.7694
MARIA	0.7839	0.7879	0.7857	0.7532	0.7443	0.7488	0.7425	0.7087	0.7726
SACN	**0.7862**	**0.7904**	**0.7889**	**0.7569**	**0.7486**	**0.7515**	**0.7489**	**0.7143**	**0.7763**

- MMOE is chosen to be the base model as its representativeness in MSL. Compared with MMOE, PLE achieves better performance as it introduces greater stability across scenarios by segmenting the expert network into two distinct groups and extracts differences and commonalities among scenarios more effectively.
- Both MMOE and PLE introduce scenario-specific DNN towers at the top of models and outputs multiple scores for different scenarios. Nevertheless, they ignore the bottom level (e.g., embedding layer) optimization of models, which will seriously degenerate the performance of multi-scenario recommendation. PEPNet uses scenario-aware gated units to adjust embedding layer and hidden layers adaptively. M2M introduce a novel meta unit that incorporates rich scenario knowledge to learn explicit inter-scenario correlations and enhance the capability of capturing the representation of scenario-specific features. They both perform better than MMoE and PLE.
- The MARIA outshines others by optimizing both the lower and upper structures of the model. However, All of these methods use item IDs only when modeling behaviors and do not consider the impact of item's multiple attributes, which are indispensable to generate rich latent interest representations and reflect user's interests discrepancy in different scenarios. Our SACN model yields the best performance compared with others in all scenarios due to the leveraging of item attributes information. Specially, the second best model MARIA has worse performance than SACN, especially for scenarios with sparse behaviors. For example, scenario #D7 and scenario #D8 have very large portion of users without sufficient behaviors (i.e., over 37.42% in #D7 and 41.23% in #D8 users have interactions with items less than 5 times in recent 180 days, respectively). SACN achieves a larger AUC improvement in these two scenarios than others.

3.3 Ablation Study

To investigate the effectiveness of each component in the proposed SACN, especially the attributed-related module, we conduct several ablation experiments:

- **w/o APE** removes the APE from SACN model, which represents that item attributes information are not introduced.
- **w/o APE (w DIF-SR)** replaces the APE module with DIF-SR [18], which is a self-attention based attribute-aware behavior modeling method without considering target scenario and item information.
- **w/o APE (w ASIF)** replaces the APE module with ASIF [17], which is also an attribute-aware behavior modeling approach with aligning item IDs and attributes.
- **w/o SCM** simply removes the SCM module from SACN, which means that the interests discrepancy is not distinguished explicitly.

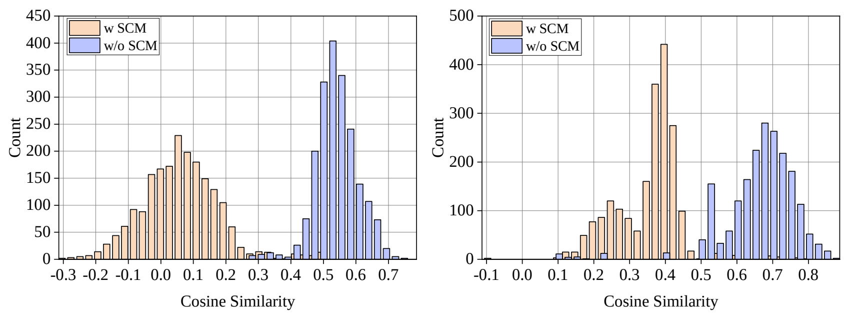
Table 3. Comparison of different SACN variants on industrial dataset.

	#D1	#D2	#D3	#D4	#D5	#D6	#D7	#D8
SACN	**0.7862**	**0.7904**	**0.7889**	**0.7569**	**0.7486**	**0.7515**	**0.7489**	**0.7143**
w/o APE	0.7796	0.7839	0.7815	0.7486	0.7391	0.7453	0.7351	0.7007
w/o APE (DIF-SR)	0.7851	0.7889	0.7874	0.7545	0.7461	0.7499	0.7456	0.7116
w/o APE (ASIF)	0.7854	0.7893	0.7875	0.7549	0.7464	0.7503	0.7457	0.7119
w/o SCM	0.7829	0.7871	0.7861	0.7526	0.7431	0.7479	0.7412	0.7074

As shown in Table 3, with APE removed, w/o APE suffers a significant performance degradation and is even worse than PEPNet. This observation proves the correctness and necessity of leveraging items attributes when modeling multiple scenarios. Additionally, removing APE has stronger impact on sparse scenarios, which indicates that item attributes are able to overcome highly sparse settings achieving superior performance. When replacing APE with other attribute-aware behavior modeling methods without using target scenario and item knowledge, a degradation of AUC is also observed, implying that these information are beneficial to the learning of fine-grained preference representation. Furthermore, without the SCM, model performance declines in all scenarios significantly. This demonstrates that self-supervised learning injecting unsupervised signals in terms of item IDs and attributes information is very helpful in distinguishing interests among scenarios.

3.4 In-depth Analysis

We conduct experiments to further explore whether the module understands the interests difference of different scenarios. We calculate the cosine similarity of scenario #D1 and scenario #D3, similarity of scenario #D6 and scenario #D8, based on the representations (i.e., $\boldsymbol{P_t^I}$ and $\boldsymbol{P_k^I}$) given by the module IPE with and without introducing the contrastive loss, meanwhile, plot the distribution of similarity scores in Figure 2. From the results, we can observe that embeddings learned by SACN have smaller similarity scores than those learned without SCM. This phenomenon indicates that SCM enables the model to distinguish scenario-wise interests effectively.



(a) Scenario #D1 and scenario #D3 (b) Scenario #D6 and scenario #D8

Fig. 2. The histogram of similarities between the item-level preference embeddings of different scenarios, with and without the SCM

4 Online A/B Test

To further demonstrate the effectiveness of the proposed SACN, we deploy it on our platform for A/B test. Due to industrial constraints, it was not feasible to compare all baseline models in the online system. Therefore, we select PEPNet as the baseline model for comparison. The online evaluation metric is pCTR, i.e., the number of clicks over the number of impressions. After conducting two-week online A/B test, we find the proposed SACN achieves consistent improvement over PEPNet model, i.e., achieving an average improvement of 4.17% pCTR overall. In a nutshell, the online A/B test results again demonstrate the effectiveness and practicability of our SACN model in the industrial setting. Now, SACN has been deployed on our platform and is serving real traffics.

5 Conclusion

In this paper, we present a novel method SACN for multi-scenario learning with introducing item attributes information when modeling user behaviors. SACN can capture user's coarse-grained and fine-grained preferences by utilizing item IDs and attributes. The preferences extraction process also considers the importance of using target scenario and item prior knowledge to achieve better performances. With the help of self-supervision, SACN distinguishes the user's preferences discrepancy in different scenarios in terms of user's item-level and attribute-level interest representations. Extensive experiments on our industrial dataset and online A/B testing demonstrate that SACN consistently outperforms state-of-the-art baselines.

References

1. Chang, J., Zhang, C., Hui, Y., Leng, D., Niu, Y., Song, Y., Gai, K.: Pepnet: Parameter and embedding personalized network for infusing with personalized prior information. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 3795–3804 (2023)
2. Cheng, H.T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., et al.: Wide & deep learning for recommender systems. In: Proceedings of the 1st workshop on deep learning for recommender systems. pp. 7–10 (2016)
3. Guo, H., Tang, R., Ye, Y., Li, Z., He, X.: Deepfm: a factorization-machine based neural network for ctr prediction. arXiv preprint arXiv:1703.04247 (2017)
4. He, R., McAuley, J.: Vbpr: visual bayesian personalized ranking from implicit feedback. In: Proceedings of the AAAI conference on artificial intelligence. vol. 30 (2016)
5. Jacobs, R.A., Jordan, M.I., Nowlan, S.J., Hinton, G.E.: Adaptive mixtures of local experts. Neural computation **3**(1), 79–87 (1991)
6. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009)
7. Li, P., Li, R., Da, Q., Zeng, A.X., Zhang, L.: Improving multi-scenario learning to rank in e-commerce by exploiting task relationships in the label space. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. pp. 2605–2612 (2020)
8. Liu, C., Li, X., Cai, G., Dong, Z., Zhu, H., Shang, L.: Noninvasive self-attention for side information fusion in sequential recommendation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 4249–4256 (2021)
9. Ma, J., Zhao, Z., Yi, X., Chen, J., Hong, L., Chi, E.H.: Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 1930–1939 (2018)
10. Min, E., Luo, D., Lin, K., Huang, C., Liu, Y.: Scenario-adaptive feature interaction for click-through rate prediction. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 4661–4672 (2023)
11. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
12. Rashed, A., Elsayed, S., Schmidt-Thieme, L.: Context and attribute-aware sequential recommendation via cross-attention. In: Proceedings of the 16th ACM Conference on Recommender Systems. pp. 71–80 (2022)
13. Rashed, A., Jawed, S., Schmidt-Thieme, L., Hintsches, A.: Multirec: A multi-relational approach for unique item recommendation in auction systems. In: Proceedings of the 14th ACM Conference on Recommender Systems. pp. 230–239 (2020)
14. Tang, H., Liu, J., Zhao, M., Gong, X.: Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations. In: Proceedings of the 14th ACM Conference on Recommender Systems. pp. 269–278 (2020)

15. Tian, Y., Li, B., Chen, S., Li, X., Deng, H., Xu, J., Zheng, B., Wang, Q., Li, C.: Multi-scenario ranking with adaptive feature learning. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 517–526 (2023)
16. Wang, C., Ma, W., Chen, C., Zhang, M., Liu, Y., Ma, S.: Sequential recommendation with multiple contrast signals. ACM Transactions on Information Systems **41**(1), 1–27 (2023)
17. Wang, S., Shen, B., Min, X., He, Y., Zhang, X., Zhang, L., Zhou, J., Mo, L.: Aligned side information fusion method for sequential recommendation. In: Companion Proceedings of the ACM on Web Conference 2024. pp. 112–120 (2024)
18. Xie, Y., Zhou, P., Kim, S.: Decoupled side information fusion for sequential recommendation. In: Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval. pp. 1611–1621 (2022)
19. Xu, J., Wang, C., Wu, C., Song, Y., Zheng, K., Wang, X., Wang, C., Zhou, G., Gai, K.: Multi-behavior self-supervised learning for recommendation. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 496–505 (2023)
20. Yan, B., Wang, P., Zhang, K., Li, F., Deng, H., Xu, J., Zheng, B.: Apg: Adaptive parameter generation network for click-through rate prediction. Advances in Neural Information Processing Systems **35**, 24740–24752 (2022)
21. Yang, X., Peng, X., Wei, P., Liu, S., Wang, L., Zheng, B.: Adasparse: Learning adaptively sparse structures for multi-domain click-through rate prediction. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 4635–4639 (2022)
22. Zhang, Q., Liao, X., Liu, Q., Xu, J., Zheng, B.: Leaving no one behind: A multi-scenario multi-task meta learning approach for advertiser modeling. In: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. pp. 1368–1376 (2022)
23. Zhang, T., Zhao, P., Liu, Y., Sheng, V.S., Xu, J., Wang, D., Liu, G., Zhou, X., et al.: Feature-level deeper self-attention network for sequential recommendation. In: IJCAI. pp. 4320–4326 (2019)
24. Zhang, Y., Wang, X., Hu, J., Gao, K., Lei, C., Fang, F.: Scenario-adaptive and self-supervised model for multi-scenario personalized recommendation. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 3674–3683 (2022)
25. Zhao, P., Gao, X., Xu, C., Chen, L.: M5: Multi-modal multi-interest multi-scenario matching for over-the-top recommendation. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 5650–5659 (2023)
26. Zhou, G., Zhu, X., Song, C., Fan, Y., Zhu, H., Ma, X., Yan, Y., Jin, J., Li, H., Gai, K.: Deep interest network for click-through rate prediction. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 1059–1068 (2018)
27. Zhou, J., Cao, X., Li, W., Bo, L., Zhang, K., Luo, C., Yu, Q.: Hinet: Novel multi-scenario & multi-task learning with hierarchical information extraction. In: 2023 IEEE 39th International Conference on Data Engineering (ICDE). pp. 2969–2975. IEEE (2023)
28. Zhou, K., Wang, H., Zhao, W.X., Zhu, Y., Wang, S., Zhang, F., Wang, Z., Wen, J.R.: S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization. In: Proceedings of the 29th ACM international conference on information & knowledge management. pp. 1893–1902 (2020)