Contrastive Scenario-Aware Meta Prompting for Multi-scenario Recommendation

Ang Li[1*], Jian Hu[2*], Ke Ding[1], Xiaolu Zhang[1], Jun Zhou[1]✉[**], and Yong He[1]

[1] Antgroup, Hangzhou, China
[2] Queen Mary University of London, London, UK
{liang268038, dingke.dk,yueyin.zxl,jun.zhoujun,heyong.h}@antgroup.com
jian.hu@qmul.ac.uk

Abstract. As the number of sub-scenarios increases, multi-scenario recommendation becomes essential for reducing operational costs while improving performance. Developing a unified model for diverse scenarios is challenging due to differences in information dimensions and semantic gaps. To address this, we propose integrating Large Language Models (LLMs) into multi-scenario recommendations by using text as a bridge to resolve semantic mismatches. However, two key challenges arise: (1) Manually crafting scenario-specific prompts for LLMs is time-consuming and inefficient; (2) Focusing solely on scenario-specific representations overlooks valuable knowledge from related scenarios, limiting performance. To address these issues, we introduce the Contrastive Scenario-Aware Meta Prompting (CSAMP) framework, comprising two modules: Scenario-Aware Meta Prompting (SAMP) and Semantic-Enhanced Contrastive Learning (SEC). SAMP uses meta-learning to automatically generate prompts, enabling knowledge transfer between similar scenarios. SEC allows ID representations to learn from text representations across scenarios, fostering a more integrated and effective system. Experiments on various datasets demonstrate the effectiveness of CSAMP.

1 Introduction

In practical recommendation systems, data characteristics can vary significantly across different scenarios. To achieve optimal performance, one approach is to train specialized models for each scenario. However, as the number of scenarios grows, this method becomes computationally expensive and resource-intensive. Additionally, managing and updating multiple models introduces complexity, underscoring the need for more adaptable and scalable models. This challenge has led to the rise of multi-scenario recommendation systems, which aim to create a unified model that leverages data from various scenarios, allowing it to serve multiple contexts simultaneously in an online environment [22].

Multi-scenario recommendation systems combine shared and specific parameters for diverse contexts [15, 19]. Item IDs vary across scenarios, hindering the use

[*] Equal contribution.
[**] ✉ Corresponding author.

Table 1: AUC of various prompts in multi-scenario recommendation.

Model	Content	Scenario_0	Scenario_1	Scenario_2	AVG
Prompt_0	User Profile: {}, item Details: {}, Predict: Will the user be interested in the item?	**0.7858**	**0.9313**	0.8207	0.8459
Prompt_1	Imagine a user and a product with the following characteristics: {}; {}. Narrate the likelihood of the user interacting with product.	0.7809	0.9311	**0.8219**	0.8446
Prompt_2	Data Analysis: Given a user with these attributes {}, and a product with these features {}, can you analyze and predict the potential for user-product interaction?	0.7850	0.9286	0.8160	0.8449

of implicit information. For example, purchasing basketball shoes and engaging with NBA content across scenarios suggest a consistent interest in basketball, but different item IDs obscure this link. Text, a universal feature, can bridge scenarios. Therefore, we suggest leverage LLMs [2, 16], which stand out in text processing, to unify these interests.

However, two issues exist. Firstly, the performance of LLMs in multi-scenario recommendations is heavily dependent on the prompts used. Each scenario requires carefully designed prompts to achieve optimal performance (as shown in Table 1), but manually crafting the most appropriate prompts for each scenario is a challenging and time-consuming task. Additionally, these prompts fail to capture the correlations between different scenarios, making it difficult to transfer useful knowledge across contexts. Secondly, LLMs require significant computational resources, making real-time predictions in large-scale online recommendation systems impractical. To address this, prior methods [13] have employed contrastive learning [3] to align text and ID representations, allowing the collaborative model to be used during online serving and mitigate the prediction delay. However, these approaches only align text and ID representations within the same sample, neglecting the alignment of similar samples across different scenarios. This limitation leads to suboptimal performance, as the non-interoperability of item IDs across scenarios hinders knowledge sharing and cross-scenario learning.

In this paper, we introduce a novel framework named Contrastive Scenario-Aware Meta Prompting (CSAMP) for multi-scenario recommendation. This innovative approach consists of two core components: Scenario-Aware Meta Prompting (SAMP) and Semantic-Enhanced Contrastive Learning (SEC). SAMP incorporates contextual scenario information into the auto-generation of prompts by meta-learning to gain optimal prompts automatically. It enables the prompts to discern and understand the intricate relationships between diverse scenarios dynamically. Consequently, such enriched prompts enhance the model's capability to transfer knowledge across various domains. SEC is designed to accomplish the alignment of representations across various scenarios. Different from the naive cross-modal contrastive learning, since we have similar samples from different scenarios, SEC aligns the ID representations with similar text representations across other scenarios. **Our contributions are as follows:**

1). We introduce **C**ontrastive **S**cenario-**A**ware **M**eta **P**ro-mpting (CSAMP) for multi-scenario recommendation. It leverages the generalization performance of LLMs to build bridges for recommendation systems across different scenarios.

2). To address the time-consuming and suboptimal issues caused by manually setting different prompts, we also introduce Scenario-Aware Meta Prompting (SAMP) module. It utilizes meta-learning to achieve automated prompt generation while integrating scenario characteristics.

3). We propose Semantic-Enhanced Contrastive Learning (SEC) to overcome the issue of inferior performance resulting from cross-modal alignment only within the same scenario. Specifically, SEC expands the scope and aligns the ID representations with similar text representations across various scenarios.

2 Related Works

Task-driven prompting aims to generate task-specific prompts to guide downstream tasks. Multimodal large models like CLIP [18] are exploited for their broad applicability, but their dependence on prompt nuances can significantly impact performance [24]. To mitigate this, studies such as [25, 20] have developed methods to tailor prompts to task features, enhancing model effectiveness. With the advent of billion-parameter models like the GPT series [6] and SAM [11], optimizing task-specific prompting has become crucial. Efforts like [7, 8] merge large models to generate context-aware prompts for personalized outcomes. Techniques like CoT [21] use logical reasoning to generate targeted prompts for language tasks. However, the extensive reasoning required can impede real-time application. Our CSAMP model addresses this by providing efficient, personalized prompts for improved performance in real-time scenarios.

Multi-Scenario Learning trains a unified model to handle multiple recommendation tasks across different scenarios, reducing model maintenance and enhancing effectiveness. The simplest approach, Full Shared Parameters, shares all parameters across scenarios but may overlook specific nuances, limiting performance. SharedBottom uses common bottom layers for all scenarios, capturing shared knowledge, while scenario-specific towers address unique requirements. MMoE [15] employs a Mixture-of-Experts framework with a gating network that optimizes task-specific performance by sharing expert models across scenarios. PLE [19] refines this by assigning unique expert layers to each scenario for precise modeling. Moreover, Meta Learning approaches in multi-scenario learning dynamically generate parameters for each scenario, offering adaptability and supporting a variable number of scenarios. APG [22] innovates by dynamically generating parameters for deep CTR models based on scenario context. MI-DPG [4] enhances parameter diversity by maximizing mutual information between the scenario-aware input and the scenario-conditioned dynamic weighting matrix. However, the aforementioned methods all overlook the utilization of textual information.

LLMs for Recommendation harness LLMs to understand user preferences through text, enhancing recommendation systems. CTR-Bert [13] integrates text and tabular data through twin-structured BERT-like encoders. However, the main drawback is its high resource demand during live operation. CTRL [17] addresses this by transferring knowledge from LLMs to collaborative networks via contrastive learning, enabling lighter network deployment during inference.

3 Preliminary

We present the foundational concept of multi-scenario recommendation. The dataset D is defined to include K different scenarios, represented as $\{D_1, D_2, ..., D_K\}$. Here, $D_k = \{x_i, y_i\}_{i=1}^{n_k}$ denotes the data corresponding to the k^{th} scenario, where x represents the input features and y represents the label. The input features, denoted by $x = \{(d_1, d_2, ..., d_S), d_{text}\}$, consist of two primary components: *ID information* and *Text information*. Here, *ID information* encompasses $(d_1, d_2, ..., d_S)$, which includes identifiers such as userID, itemID, and other relevant IDs. Alternatively, *Text information* pertains to d_{text}, exemplified by elements like the item title. The examples of these two componets are shown in Fig. 1. The label $y \in \{0, 1\}$ indicates whether a user clicked on an item.

For single-scenario recommendation, it is required to train a separate submodel $y = f_i(x)$ using data from each sub-scenario D_i, resulting in a set of K sub-models to serve K sub-scenarios simultaneously, which is cost-prohibitive. In contrast, multi-scenario recommendation utilizes data from K scenarios to train a single unified model $y = F(x)$, which serves all scenarios simultaneously, significantly reducing storage resources. However, ID information across different sub-scenarios is hard to interoperate. Textual information d_{text} is either overlooked or used directly as ID, lacking the capability of semantic understanding.

4 Methods

As shown in Fig. 1, CSAMP uses contrastive learning [3] and meta-learning [5] to transfer the LLM's strong text representation abilities to the collaborative model. CSMAP comprises three distinct modules: the Scenario-Aware Meta Prompting, the Collaborative Model, and the Semantic-Enhanced Contrastive Alignment.

4.1 Scenario-Aware Meta Prompting

Prompt learning is an innovative approach in natural language processing that has gained traction due to its ability to leverage pre-trained language models for various tasks without extensive task-specific fine-tuning. However, manually crafting prompts for various scenarios is neither the most effective nor the most efficient approach. To incorporate scenario-specific features into the prompts, we propose to adopt meta-learning techniques. This method enables us to infuse scenario information directly into the prompt creation process. Notably, we are pioneers in utilizing Soft Prompt techniques with LLMs to enhance their understanding of diverse scenarios.

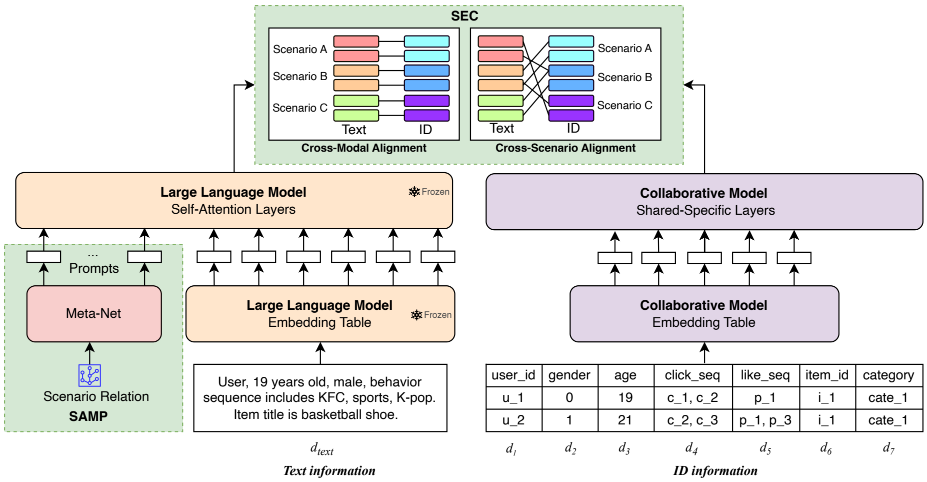

Fig. 1: Framework of our CSAMP. The prompts generated by Scenario-Aware Meta Prompting (SAMP) are concatenated with textual content and fed into the Large Language Model (LLM) to yield the Text representation. Concurrently, the raw features of the ID are processed by the collaborative model to produce the ID representation. Subsequently, these two representations are aligned using the Semantic-Enhanced Contrastive Learning (SEC).

Scenario Information Scenario information refers to certain contextual information when recommending items to users. We obtain the representation of scenarios from two perspectives: explicit and implicit. Explicit representation (ER) is measured using static features like the industry and category. Given the static features $S_k = \{s_1^k, s_2^k, \cdots, s_m^k\}$ of D_k, the ER^k is defined as:

$$ER^k = \text{concat}(e_{s_1}^k, e_{s_2}^k, \cdots, e_{s_m}^k), \tag{1}$$

where $e_{s_i}^k$ is the embedding of the static feature s_i, ER^k is the explicit representation of the k^{th} scenario.

Leveraging the Swing [23, 9] algorithm, implicit representation (IR) is measured by the overlap of users between two scenarios. The implicit similarity between scenario i and j is defined as:

$$Sim(i,j) = \textstyle\sum_{u_1} \sum_{u_2} \mathbf{1}\{u_1 \neq u_2\} w_{u_1} * w_{u_2} \frac{1}{\alpha + |S_{u_1} \cap S_{u_2}|}, \tag{2}$$

where $Sim(i,j)$ denotes the similarity between scenarios i and j, $u_1 \in U_i$ and $u_2 \in U_j$, U_i and U_j respectively represent the sets of users associated with scenarios i and j, $\mathbf{1}$ is the indicator function, S_{u_1}, S_{u_2} are the sets of scenarios visited by users u_1 and u_2, and $w_{u_1} = \frac{1}{\sqrt{|S_{u_1}|}}$ as well as $w_{u_2} = \frac{1}{\sqrt{|S_{u_2}|}}$ indicate the activity of the users, with more active users having lower weights. Highly active users receive lower weights to prevent their interactions from dominating and introducing bias. It ensures a balanced similarity measure by emphasizing the more specific and informative interactions of less active users. α is a hyperparameter used to avoid division by zero. The implicit representation is derived by averaging the explicit representations of the G scenarios most closely associated with k.

$$IR^k = \mathrm{f}_{mean}(\{ER_j \mid j \in \underset{G}{\operatorname{argmin}} \{Sim(k,j) \mid j \neq k\}\}), \tag{3}$$

where f_{mean} represents the Average Pooling operation. By concatenating the IR and ER, the final scenario representation $SR^k = \mathrm{concat}(IR^k, ER^k)$ is obtained.

Meta Network After acquiring representations of various scenarios, scenario representations (SR) are fed into a meta-learning network. The output generated by this network will serve as our soft prompt P_s.

$$P_s = W_{meta} SR, \tag{4}$$

where W_{meta} is the learnable parameters of the meta network. Subsequently, the prompt and the original text information d_{text} are jointly fed into the Large Language Model (LLM) following Prefix Tuning [12]. The resulting representations Rep_{text} obtained from the LLM are used for aligning with ID information.

$$Rep_{text} = LLM(P_s, d_{text}), \tag{5}$$

4.2 Collaborative Model

We then transform raw ID features $(d_1, d_2, ..., d_S)$ into embeddings and subsequently concatenate all embeddings to serve as the input for the collaborative module, which is shown as follows:

$$Input_{ID} = \mathrm{concat}(e_{d_1}, e_{d_2}, ..., e_{d_S}), \tag{6}$$

where e_{d_i} is the embedding of the ID d_i. We employ any multi-scenario recommendation (MSL) architecture to obtain intermediate representations (Rep_{id}) for each scenario. Specifically, the intermediate layers of each tower represent different scenarios. Their representation can be selected with the one-hot mask.

$$Rep_{id} = MSL(Input_{ID}), \tag{7}$$

4.3 Semantic-Enhanced Contrastive Alignment

During this phase, we use contrastive learning to align the ID representations Rep_{id} with the text representations Rep_{text}. The naive cross-modal alignment loss L_{ctr}^{id2t} within the same sample and scenario formula is as follows:

$$L_{ctr}^{id2t} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(sim(Rep_{id}^i, Rep_{text}^i)/\tau)}{\sum_{j=1}^{N} \exp(sim(Rep_{id}^i, Rep_{text}^j)/\tau)}, \tag{8}$$

where $sim(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$. The meaning of this formula is that for sample i, we try to minimize the distance between its ID representation Rep_{id}^{i} and text representation Rep_{text}^{i} while simultaneously maximizing the distance between its ID representation Rep_{id}^{i} and the text representations Rep_{text}^{j} of all other samples within the batch. However, this loss only affects the text and ID within the same example. In datasets with multiple scenarios where IDs differ, if samples from two scenarios have different IDs but show the same user interests, their text, and ID should also be linked by contrastive learning. To achieve this, we introduce the Semantic-Enhanced Contrastive Loss, which incorporates the cross-modal alignment loss L_{sec}^{id2t} within a sample and the cross-scenario alignment among similar samples across scenarios.

$$L_{sec}^{id2t} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\sum_{t \in T_i} \exp(sim(Rep_{id}^{i}, Rep_{text}^{t})/\tau)}{\sum_{j=1}^{N} \exp(sim(Rep_{id}^{i}, Rep_{text}^{j})/\tau)} \ , \tag{9}$$

where T_i denotes the set of the top k closest samples to the text representation of sample i. Specifically, T_i is derived from the computation of internal similarities using text representations Rep_{text} in Equation (10), and it includes samples from various scenarios.

$$T_i = \underset{k}{\operatorname{argmin}} \left\{ \operatorname{sim}(\operatorname{Rep}_{\text{text}}^{i}, \operatorname{Rep}_{\text{text}}^{j}) \mid 0 \leq j \leq N, j \neq i \right\}, \tag{10}$$

Similarly, the SEC loss for aligning text representations with ID representations L_{sec}^{t2id} is given by:

$$L_{sec}^{t2id} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\sum_{t \in T_i} \exp(sim(Rep_{text}^{i}, Rep_{id}^{t})/\tau)}{\sum_{j=1}^{N} \exp(sim(Rep_{text}^{i}, Rep_{id}^{j})/\tau)}, \tag{11}$$

The final alignment loss is the average of L_{sec}^{id2t} and L_{sec}^{t2id}.

$$L_{alignment} = \frac{1}{2}(L_{SEC}^{t2id} + L_{SEC}^{id2t}), \tag{12}$$

After aligning the representations of ID and text using the alignment loss, the collaborative model has assimilated the knowledge from LLMs. To further learn the knowledge of labels, we fine-tune the model using the binary cross-entropy loss. The formula is as follows:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]. \tag{13}$$

where N is the total number of samples, y_i is the true label for the i^{th} sample, which can be 0 or 1, p_i is the predicted probability that the i^{th} sample belongs to class 1. During the inference stage, only the collaborative model will be deployed in the online service. Consequently, there is no added latency in the scoring process in our method.

4.4 Datasets

We conduct experiments on both the publicly available **AntM^{2}C** [10] dataset and an industrial dataset **Search**. **AntM^{2}C** is a representative Multi-Scenario, Multi-Modal Click-Through Rate (CTR) dataset originating from the Alipay platform. It packages CTR data for various items, covering categories that include advertisements, vouchers, mini-programs, content, and videos. Beyond ID-based attributes, AntM^{2}C also encompasses raw textual descriptions. We randomly selected three distinct scenes based on the *scene* attribute for our experiments and utilized both text and ID features to evaluate our methods. **Search** is collected on content recommendations from two scenarios of Alipay search booths: Main Search and Vertical Search. Detailed description is in Table 2.

Table 2: Dataset statistics. We provide exposure, users, items, and click rate for each scenario of two datasets.

Dataset	Scenario	Exposure	Users	Items	Click Rate
AntM^{2}C	0	575,753	32,263	5,043	3.70%
	1	5,186,502	63,620	1,708	38.45%
	2	2,893,714	49,962	2,505	23.81%
Search	0	899,751	27,437	6,650	1.45%
	1	224,937	7,550	2,241	4.91%

5 Experiments

5.1 Experimental Setting

Baseline To validate the effectiveness of our CSAMP, we compare it with two categories of methods: **collaborative multi-scenario recommendation** approaches and **LLM-based methods**. Collaborative multi-scenario recommendation approaches include: Sharedbottom, MMoE [15], PLE [19], APG [22] and MI-DPG [4]. They utilize various network architectures that integrate shared and task-specific knowledge to optimize multi-task learning and enhance model generalization across multiple scenarios. LLM-based approaches contain CTR-Bert [17] and CTRL [13]. Both of them leverage the generalization ability of LLM to enhance the integration and knowledge alignment across different data modalities for improving performance across scenarios.

Implement Details The size of the embedding vectors for ID-based features is consistently configured to 16 for every model assessed. The meta network architecture comprises a two-layer MLP, with hidden units of sizes [64, 16] in each respective layer. For a fair evaluation, CSAMP and other methods utilize the same Sharedbottom network structure for the collaborative model. Within this network, the shared component consists of five hidden layers with dimensions of 512, 256, 128, 64, and 32. Each tower is parameterized with dimensions of [64, 16]. During the alignment process, we utilize a Projection Layer to map

Table 3: Experiments Results on *AntM^{2}C* and *Search* Datasets.

Model	Type	AntM^{2}C				Search		
		Scenario$_0$	Scenario$_1$	Scenario$_2$	AVG	Scenario$_0$	Scenario$_1$	AVG
SharedBottom	collaborative	0.7738 ± 0.0019	0.9237 ± 0.0012	0.8107 ± 0.0014	0.8361	0.6531 ± 0.0016	0.6767 ± 0.0010	0.6649
MMoE [15]	collaborative	0.7746 ± 0.0017	0.9254 ± 0.0015	0.8126 ± 0.0012	0.8375	0.6587 ± 0.0013	0.6825 ± 0.0007	0.6706
PLE [19]	collaborative	0.7761 ± 0.0010	0.9252 ± 0.0013	0.8136 ± 0.0011	0.8383	0.6593 ± 0.0011	0.6820 ± 0.0008	0.6707
APG [22]	collaborative	0.7743 ± 0.0010	0.9247 ± 0.0011	0.8119 ± 0.0012	0.8370	0.6605 ± 0.0013	0.6793 ± 0.0009	0.6699
MI-DPG [4]	collaborative	0.7774 ± 0.0009	0.9259 ± 0.0009	0.8135 ± 0.0010	0.8389	0.6611 ± 0.0010	0.6839 ± 0.0006	0.6725
CTR-Bert [13]	LLM-based	0.7811 ± 0.0007	0.9285 ± 0.0008	0.8173 ± 0.0006	0.8423	0.6641 ± 0.0008	0.6877 ± 0.0005	0.6759
CTRL [17]	LLM-based	0.7850 ± 0.0006	0.9342 ± 0.0006	0.8219 ± 0.0007	0.8470	0.6646 ± 0.0007	0.6887 ± 0.0005	0.6767
CSAMP (Ours)	LLM-based	**0.7927** ± 0.0003	**0.9381** ± 0.0004	**0.8272** ± 0.0002	**0.8527**	**0.6733** ± 0.0005	**0.6947** ± 0.0003	**0.6840**

Table 4: Component and LLM Model Size Analysis Results.

Component Analysis				LLM Model Size Analysis			
Model	Scenario$_0$	Scenario$_1$	*AUC*	Model	Scenario$_0$	Scenario$_1$	*AUC*
CSAMP	**0.6733** ± 0.0005	**0.6947** ± 0.0003	**0.6840**	*1.1b*	0.6709 ± 0.0007	0.6923 ± 0.0004	0.6816
CSAMP–Meta	0.6703 ± 0.0007	0.6909 ± 0.0005	0.6806	*1.7b*	0.6733 ± 0.0005	0.6947 ± 0.0003	0.6840
CSAMP–SAMP	0.6671 ± 0.0009	0.6876 ± 0.0006	0.6774	*3b*	0.6778 ± 0.0003	0.6955 ± 0.0003	0.6867
CSAMP–SEC	0.6722 ± 0.0006	0.6926 ± 0.0003	0.6824	*7b*	**0.6792** ± 0.0003	**0.6971** ± 0.0003	**0.6882**

the dimensions of the tower output to match that of the textual representation. Bloomz-1b7 [16] is used as the text backbone for all LLM-based methods.

5.2 Experimental Results and Analysis

Experiment Results The Area Under the Curve (AUC) is employed as an evaluative metric to evaluate the effectiveness of various approaches under the multi-scenario recommendation setting. An increase exceeding 0.3% is considered a significant improvement. For the AntM^{2}C dataset, we report the AUC of Scenario$_0$, Scenario$_1$, Scenario$_2$ and the average of these scenarios. Likewise, Scenario$_0$, Scenario$_1$, and AVG are used as the metrics for the Search dataset. All experiments were conducted five times, and the results were averaged and shown in Table 3. Among the collaborative methods, MI-DPG stands out by guaranteeing the diversity of generated parameters by maximizing mutual information, thus delivering optimal results. The two LLM-based approaches, CTR-Bert and CTRL, exhibit a notable improvement in performance due to information gains from textual data and LLM compared to collaborative models. However, a significant gap remains in comparison to our CSAMP strategy, which not only incorporates scenario information into the Prompt but also leverages SEC to align text with ID representations more effectively.

Ablation Study **Component Analysis.** In this section, we validate the impact of various components on an industrial dataset **Search**; the results are presented in the component analysis part in Table 4. *CSAMP* is our full solution with all features. *CSAMP-Meta* removes scenario information and meta-learning using a learnable prompt similar to Prefix Tuning. *CSAMP-SAMP* represents our model without the adaptive prompt module, relying only on fixed prompts. Finally, *CSAMP-SEC* excludes the SEC module, utilizing basic contrastive learning instead. This breakdown highlights the specific role of each component in the overall architecture. The comparison between the first and second rows shows the advantages of adding scenario information to the prompt. The differences between the first and third rows highlight the effectiveness of the SAMP module, revealing that using fixed Prompts significantly lowers performance compared to dynamic prompts with scenario information. Lastly, the comparison between the first and fourth rows demonstrates the SEC module's ability to align different samples across scenarios using textual data effectively.
LLM Model Size Analysis. In this section, we investigate whether the size of the LLM impacts the overall results, mainly focusing on whether a larger

LLM leads to improved performance in the collaborative model. Our experiments in the "LLM Model Size Analysis" part in Table 4, conducted with LLMs of various sizes on an industrial dataset *Search*, demonstrate that larger LLMs consistently enhance the model's performance, which is likely attributed to their more advanced text processing capabilities.

MSL Structure Analysis The structure of MSL can significantly influence the final results. In Fig.2, we conduct experiments to evaluate the metrics associated with different MSLs while maintaining the same LLM backbone. As illustrated in the figure, all MSL structures show notable improvements under the guidance of the LLM. MI-DPG [4] achieved the best performance, which can be attributed to its effective use of Meta Learning to model scenario information.

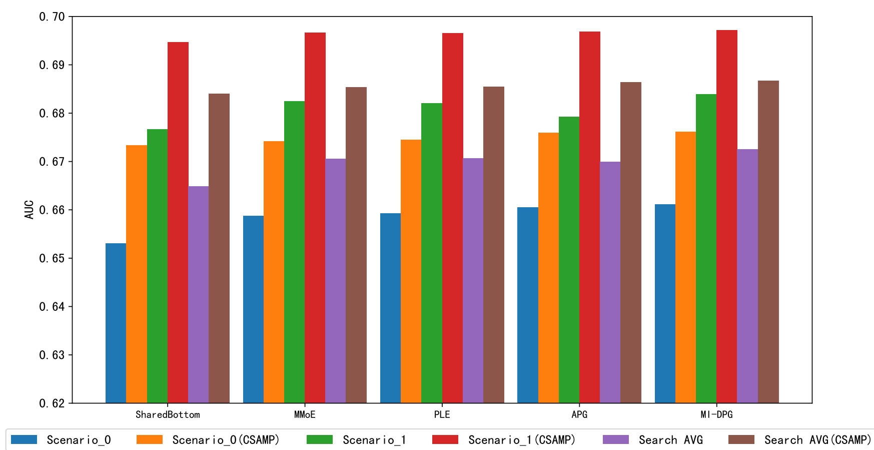

Fig. 2: Results of MSL Structure Analysis

Evaluation Latency Analysis As shown in Table, we further analyze the performance and inference time of CSAMP in comparison with other methods. It can be observed that CTR-Bert introduces a significant increase in inference time by 222 milliseconds (ms) due to the incorporation of the Language Model during the inference phase. However, both CTRL and CSAMP utilize contrastive learning to transfer the knowledge of LLMs to the collaborative model, resulting in no additional inference time compared to the collaborative model alone.

Table 5: Inference time of different models.

Model	Type	AUC	Inference time
SharedBottom	collaborative	0.6649	**13.5 ms**
CTR-Bert [13]	LLM-based	0.6759	235.5 ms
CTRL [17]	LLM-based	0.6767	**13.5 ms**
CSAMP	LLM-based	**0.6840**	**13.5 ms**

Online Deployment We deployed our CSAMP in our search systems through an eight-day A/B test. Due to limited inference resources, we only compared CSAMP with the collaborative model. The results show that CSAMP increased the online Page View Click-Through Rate (PVCTR) by 5.32% compared to the baseline, indicating the effectiveness of CSAMP in a real-world recommendation system.

6 Conclusion

In this paper, we introduces the Contrastive Scenario-Aware Meta Prompting (CSAMP) framework for multi-scenario recommendation, effectively addressing the challenges posed by differing information dimensions and semantic gaps across scenarios. By leveraging the generalization capabilities of Large Language Models (LLMs), CSAMP facilitates cross-scenario knowledge transfer and enhances recommendation systems' adaptability. Experiments on various datasets demonstrate the superiority of our CSAMP.

References

1. Bock, S., Weiß, M.G.: A proof of local convergence for the adam optimizer. 2019 International Joint Conference on Neural Networks (IJCNN) pp. 1–8 (2019), https://api.semanticscholar.org/CorpusID:203605988
2. Brown, T.B., et.al, B.M.: Language models are few-shot learners. ArXiv **abs/2005.14165** (2020), https://api.semanticscholar.org/CorpusID:218971783
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.E.: A simple framework for contrastive learning of visual representations. ArXiv **abs/2002.05709** (2020), https://api.semanticscholar.org/CorpusID:211096730
4. Cheng, W., Ding, K., Dong, X., He, Y., Zhang, L., Mo, L.: Mi-dpg: Decomposable parameter generation network based on mutual information for multi-scenario recommendation. Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (2023), https://api.semanticscholar.org/CorpusID:264350568
5. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning (2017), https://api.semanticscholar.org/CorpusID:6719686
6. Floridi, L., Chiriatti, M.: Gpt-3: Its nature, scope, limits, and consequences. Minds and Machines **30**, 681–694 (2020)
7. Hu, J., Lin, J., Gong, S., Cai, W.: Relax image-specific prompt requirement in sam: A single generic prompt for segmenting camouflaged objects. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 12511–12518 (2024)
8. Hu, J., Lin, J., Yan, J., Gong, S.: Leveraging hallucinations to reduce manual prompt dependency in promptable segmentation. arXiv preprint arXiv:2408.15205 (2024)
9. Huan, Z., et.al, A.L.: Samd: An industrial framework for heterogeneous multi-scenario recommendation. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (2023), https://api.semanticscholar.org/CorpusID:260500071

10. Huan, Z., et.al, K.D.: Antm2c: A large scale dataset for multi-scenario multi-modal ctr prediction. ArXiv **abs/2308.16437** (2023), https://api.semanticscholar.org/CorpusID:261395096
11. Kirillov, A., Mintun, E.e.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026 (2023)
12. Li, X.L., Liang, P.: Prefix-tuning: Optimizing continuous prompts for generation. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) **abs/2101.00190** (2021), https://api.semanticscholar.org/CorpusID:230433941
13. Li, X., Chen, B., Hou, L., Tang, R.: Ctrl: Connect tabular and language model for ctr prediction. ArXiv **abs/2306.02841** (2023), https://api.semanticscholar.org/CorpusID:263893102
14. Lian, J., et.al, X.Z.: xdeepfm: Combining explicit and implicit feature interactions for recommender systems. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (2018), https://api.semanticscholar.org/CorpusID:3930042
15. 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. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (2018), https://api.semanticscholar.org/CorpusID:50770252
16. Muennighoff, N., Wang, T., Sutawika, L., Roberts, A., Biderman, S., Scao, T.L., Bari, M.S., Shen, S., Yong, Z.X., Schoelkopf, H., et al.: Crosslingual generalization through multitask finetuning. arXiv preprint arXiv:2211.01786 (2022)
17. Muhamed, A., et.al, I.K.: Ctr-bert: Cost-effective knowledge distillation for billion-parameter teacher models (2021), https://api.semanticscholar.org/CorpusID:243844336
18. Radford, A., Kim, J.W.e.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021)
19. Tang, H., Liu, J., Zhao, M., Gong, X.: Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations. Proceedings of the 14th ACM Conference on Recommender Systems (2020), https://api.semanticscholar.org/CorpusID:221784966
20. Wang, Z., Zhang, Z.e.: Learning to prompt for continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 139–149 (2022)
21. Wei, J., Wang, X.e.: Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems **35**, 24824–24837 (2022)
22. Yan, B., Wang, P., Zhang, K., Li, F., Xu, J., Zheng, B.: Apg: Adaptive parameter generation network for click-through rate prediction. ArXiv **abs/2203.16218** (2022), https://api.semanticscholar.org/CorpusID:247793950
23. Yang, X., Zhu, Y., Zhang, Y., Wang, X., Yuan, Q.: Large scale product graph construction for recommendation in e-commerce. ArXiv **abs/2010.05525** (2020), https://api.semanticscholar.org/CorpusID:222291351
24. Yu, B.X., Chang, J.e.: Visual tuning. ACM Computing Surveys **56**(12), 1–38 (2024)
25. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Learning to prompt for vision-language models. International Journal of Computer Vision **130**(9), 2337–2348 (2022)