Multimodal Meta-Learning for Cold-Start Sequential Recommendation

Xingyu Pan*
School of Information, Renmin University of China
Beijing, China
xy_pan@foxmail.com

Yushuo Chen*
Gaoling School of Artificial Intelligence, Renmin University of China
Beijing, China
chenyushuo@ruc.edu.cn

Changxin Tian*
School of Information, Renmin University of China
Beijing, China
tianchangxin@ruc.edu.cn

Zihan Lin*
School of Information, Renmin University of China
Beijing, China
zhlin@ruc.edu.cn

Jinpeng Wang
Meituan Group
Beijing, China
wangjinpeng04@meituan.com

He Hu ✉
School of Information, Renmin University of China
Beijing, China
hehu@ruc.edu.cn

Wayne Xin Zhao✉
Gaoling School of Artificial Intelligence, Renmin University of China
Beijing, China
batmanfly@gmail.com

ABSTRACT

In this paper, we study the task of cold-start sequential recommendation, where new users with very short interaction sequences come with time. We cast this problem as a few-shot learning problem and adopt a meta-learning approach to developing our solution. For our task, a major obstacle of effective knowledge transfer that is there exists significant characteristic divergence between old and new interaction sequences for meta-learning.

To address the above issues, we purpose a **M**ultimodal **M**eta-**L**earning (denoted as **MML**) approach that incorporates multimodal side information of items (*e.g.,* text and image) into the meta-learning process, to stabilize and improve the meta-learning process for cold-start sequential recommendation. In specific, we design a group of multimodal meta-learners corresponding to each kind of modality, where ID features are used to develop the *main meta-learner* and the rest text and image features are used to develop *auxiliary meta-learners*. Instead of simply combing the predictions from different meta-learners, we design an adaptive, learnable fusion layer to integrate the predictions based on different modalities. Meanwhile, we design a cold-start item embedding generator, which utilize multimodal side information to warm up the ID embeddings of new items. Extensive offline and online experiments demonstrate that MML can significantly improve the recommendation performance for cold-start users compared with baseline models. Our code is released at https://github.com/RUCAIBox/MML.

CCS CONCEPTS

• **Information systems** → **Recommender systems**.

KEYWORDS

Recommender Systems; Sequential Recommendation; Meta-Learning

ACM Reference Format:
Xingyu Pan, Yushuo Chen, Changxin Tian, Zihan Lin, Jinpeng Wang, He Hu ✉, and Wayne Xin Zhao✉. 2022. Multimodal Meta-Learning for Cold-Start Sequential Recommendation. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3511808.3557101

1 INTRODUCTION

In e-commerce platforms such as *Amazon* and *Meituan*, there is a large amount of sequential user behavior data, which contains important evidence to infer the underlying user preference. To better capture users' behavioral characteristics, the task of sequential recommendation [6, 16, 53] has been proposed and attracted significant research interest in recent years, which aims to model the user's preference according to user's historical interaction behavior and predict the items that a user is likely to interact with [6]. However, due to the emergence of new users, existing sequential recommender systems usually suffer from the cold-start issue [14, 35].

∗ This work was done during internship at Meituan.
✉ Corresponding authors. Wayne Xin Zhao is also with Beijing Key Laboratory of Big Data Management.

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557101

From the perspective of machine learning, cold-start recommendation can be considered as a *few-shot learning problem*, *i.e.,* predicting a user's preferences by only a few past interacted items. As a representative few-shot learning method, meta-learning has been introduced to recommendation scenarios for addressing the cold-start problem, showing promising results [4, 18] in recent years. The basic idea of meta-learning is to learn global transferable knowledge from existing tasks and then adapt such knowledge to the similar new tasks. For cold-start recommendation, most of exist meta-learning based methods [4, 18, 42] extend the model-agnostic meta-learning (MAML) [7] approach and regard the recommendation for each user as a task. These methods aim to learn global parameters based on the interaction data of *old users* and initialize the parameters of personalized recommender models for *new users*.

Specifically, for sequential recommendation, a major obstacle to achieve effective knowledge transfer is the characteristic divergence between the old and new interaction sequences, Firstly, compared with old sequences, new interaction sequences are usually very short, so the sequential characteristics (*e.g.,* periodicity and transition patterns) might be quite different, causing the *sequence-level characteristic divergence*. Secondly, as new user may interact with some new items that have not been observed by old users, causing the *item-level characteristic divergence*. Although existing methods transfer ID-based sequential characteristics [14] or cluster similar sequences based on item attributes [35] in meta-learning process, it still lacks a comprehensive consideration about the two types of characteristic divergence, thus leading to a limited or even negative knowledge transfer in MAML [4, 54].

To motivate our solution, we observe an important phenomenon in real online platforms: an item is usually associated with rich multimodal side information, such as textual descriptions and product photos. Our main idea is to utilize multimodal side information to explore intrinsic item correlations and capture more essential characteristics from interaction sequences. By conducting sequential learning from different perspectives (*i.e.,* modalities), we aim to devise a more stable meta-learning based approach to addressing the cold-start issue for sequential recommendation. Based on such an idea, the key lies in how to effectively reduce the characteristic divergences between old and new sequence with multimodal side information in meta-learning.

In this work, we present a **M**ultimodal **M**eta-**L**earning (denoted as **MML**) approach for cold-start sequential recommendation. The key point of our approach is to incorporate multimodal side information of items (*i.e.,* text and image) into the meta-learning process to alleviate the divergence between old and new tasks and improve the effectiveness of knowledge transferred to cold-start users. To reduce the aforementioned characteristic divergences, we design a group of multimodal meta-learners (Figure 1(a)) corresponding to each kind of modality, where ID features are used to develop the *main meta-learner* and the rest text and image features are used to develop *auxiliary meta-learners*. For the main meta-learner, we further enhance the sequence representations by proposing a feature-aware self-attention mechanism, which can inject attentional bias based on multimodal correlations among items. Instead of simply combing the predictions from different meta-learners, we design an adaptive, learnable fusion layer to integrate the predictions based on different modalities. Similar to the ensemble of meta-learners [28], our approach can stabilize the meta-learning process and enhance the original ID-only sequence modeling. Besides, we further design a cold-start ID generator (Figure 1(a)) to warm up the ID embeddings of new items, so as to improve the knowledge transfer to new items.

The main contribution of this work are threefold.

- Firstly, to the best of our knowledge, it is the first work that incorporates multimodal side information into a meta-learning framework for cold-start sequential recommendation, stabilizing and enhancing the meta-learning process.
- Secondly, we leverage multimodal side information of items to alleviate the characteristic divergence between old and new interaction sequences, which can improve the stability and effectiveness of knowledge transfer.
- Thirdly, we conduct both offline evaluations on large datasets and online A/B tests on the online platform Meituan to demonstrate the effectiveness of our approach.

2 RELATED WORK

In this section, we summarize the related work from three aspects, including sequential recommendation, cold-start recommendation and meta-learning for recommender system.

2.1 Sequential Recommendation

Early works on sequential recommendation mainly focus on modelling sequential patterns with the Markov Chain assumption. Based on the last interaction of the user, MC-based approaches [31] calculated an item-item transition probability matrix and used it to forecast the next item. In recent years, deep neural networks are introduced to model the sequential patterns. Hidasi *et al.* [10] firstly utilized gated recurrent units (GRU) to session-based recommendation, and a number of variants followed this approach [11–13, 29, 30]. Besides, some works utilize convolutional neural networks (CNN) [37] and self attention networks (SAN) [16, 36] to capture the sequential patterns. As graph neural networks (GNN) gain popularity recently, they are used to model complex item correlations [2, 43] by transforming sequential data into graph-structured data. Despite the success, existing approaches, aiming to improve overall performance by sequence representation learning, have limited prediction capability for cold-start users.

2.2 Cold-start Recommendation

Cold-start problem is one of the main challenges in recommender systems. The common solution to this issue can be categorized into two types, namely side information based and transfer learning based methods. The first type of methods aims to leverage additional data resources to enhance the recommendation performance. The traditional methods [26, 33, 38] mainly use user and item attributes to augment the data. For example, LCE [33] exploits items' properties and past user preferences by a local collective embedding learning method. Recent works introduce the binary hash codes [8] and cross & compress unit [39] to leverage the side information for cold-start scenarios. Another way to alleviate the cold-start problem is to transfer knowledge from other domains. These types of methods, such as cross-domain recommendation methods [46], transfer learning methods [34, 45], and meta-learning

methods [4, 18], regard the cold-start problem as a few-shot learning problem [41] and try to utilize the knowledge distilled from other domains.

2.3 Meta-learning for Recommender System

Meta-learning, also known as *learning to learn*, aims to adapt to new tasks quickly and effectively by leveraging prior knowledge gained from previous tasks [7, 24]. In recent years, the idea of meta-learning has been taken to solve the cold-start problems in recommendation scenario. Most of them adopt optimization-based meta-learning approach and choose model-agnostic meta-learning (MAML) [7] for model training and gain great success in cold-start problem for general recommendation model [4, 18, 21, 42, 47]. To utilize the side information of users and items, some works try to combine the information in heterogeneous information networks [23] or knowledge graph [5] with MAML. There are also some meta-learning methods purposed for cold-start problem in sequential recommendation [14, 35, 40, 47, 51]. For example, Mecos [51] aims to deal with the cold-start items in sequential recommendation, which applies a recurrent matching processor to match exist user with new items. MetaTL [40] designs the meta transitional learner to model the transition patterns of interaction sequences. metaCSR [14] proposes a meta-learning based cold-start sequential recommendation framework to solve the user's cold-start recommendation problem. CBML [35] designs a cluster-based meta-learning method to transfer shared knowledge across similar session. Although existing meta-learning methods achieve subtantial performance improvement, they haven't fully leverage rich side information to reduce the task divergence. Different from existing methods, in this work, we incorporate multimodal side information of items (*e.g.,* text and image) into the meta-learning process, in order to alleviate the the characteristic divergence and improve the meta-learning process for cold-start sequential recommendation.

3 PRELIMINARIES

In this section, we first formulate the task of cold-start user sequential recommendation and then introduce the meta-learning approach for this task.

Cold-start user sequential recommendation. Given the user set $\mathcal{U} = \{u\}$ and item set $\mathcal{I} = \{i\}$, an interaction record between user $u \in \mathcal{U}$ and item $i \in \mathcal{I}$ can be denoted as $r = \langle u, i^{id}, i^{te}, i^{im} \rangle$, where i^{id} is item ID, i^{text} and i^{img} are the text information and image information (*e.g.,* title and picture of item i), respectively. Generally, the user u has a chronologically-ordered interaction sequence $s_u = \{r_1, \cdots, r_n\}$, where n is the number of interactions and r_j is the j-th interaction record. The interaction sequences of all users constitute a sequence set $\mathcal{S} = \{s_u \mid u \in \mathcal{U}\}$. In real-world recommender systems, user interaction data aggregates over time. Suppose we have already collected a set of interaction data $D_{old} = \{\mathcal{U}_{old}, \mathcal{I}_{old}, \mathcal{S}_{old}\}$ before a timestamp T. For a new user $u_{new} \notin \mathcal{U}$ who comes after T, our goal is to predict the next item that u_{new} is likely to interact with at the $(n+1)$-th step based on his/her limited historical behaviors $s_{new} = \{r_1, \cdots, r_n\}$. Note that the s_{new} is usually very short and may contain some new items which are not in $\mathcal{I}_{old}$. We regard this kind of recommendation task as cold-start user sequential recommendation.

Meta-learning settings. In this work, we extend the classic Model-Agnostic Meta-Learning (MAML) [7] approach to cold-start sequential recommendation. The basic idea of MAML is to learn an initial parameters Θ from previous tasks. Then, based on Θ, the model can quickly adapt to a new task with limited training data. Following MAML, in our work, we consider the next-item prediction task on each sequence as *a single task*. Specifically, the task goal is to predict the last element (*i.e.,* an interaction item) in each sequence based on all the previous elements. As for the dataset, the dataset $D_{old} = \{D^S_{old}, D^Q_{old}\}$ collected before T will be used for meta-training and the dataset $D_{new} = \{D^S_{new}, D^Q_{new}\}$, which consists of new users' sequences coming in after T will be used for evaluation in meta-testing phase. During the meta-training phase, we use D^S_{old} for local update and D^Q_{old} for global update; while during the meta-testing process, we will firstly use D^S_{new} for warm-up training and then test the model performance on D^Q_{new}. For D^S_{old}/D^Q_{old} (D^S_{new}/D^Q_{new}), superscripts S and Q are used to denote support and query sets, respectively.

4 APPROACH

In the section, we introduce the proposed **M**ultimodal **M**eta-**L**earning (MML) method for cold-start sequential recommendation.

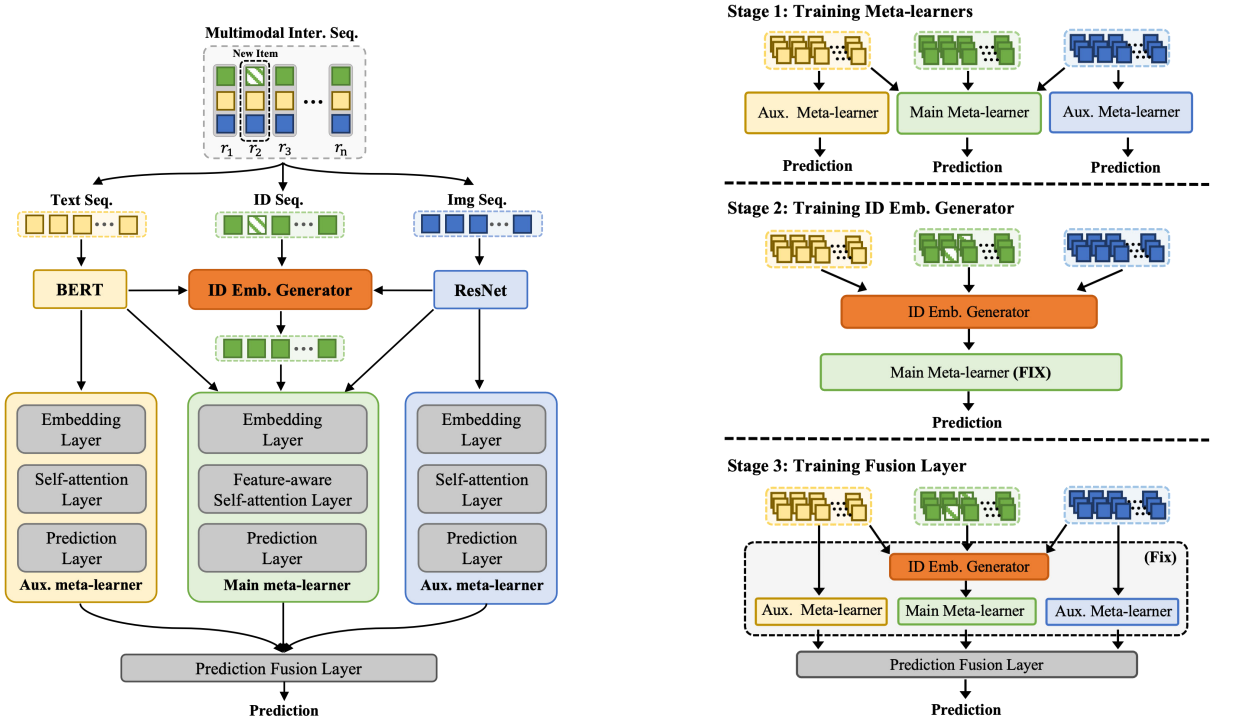
4.1 Overview

Under the MAML framework, our proposed MML incorporates the multimodal information (*i.e.,* the associated text and image data) into meta-learning process as a kind of auxiliary information to reduce the task divergence and improve the effectiveness of knowledge transfer across tasks. Specifically, we utilize the multimodal information of items in two aspects. Firstly, in order to minimize the divergence in sequential characteristics of old and new users, we elaborately design a group of multimodal meta learners corresponding to three different modalities (*i.e.,* ID, text and image), which can stabilize and improve the meta-training process by referring to each other's predictions. Secondly, considering the characteristic divergence of new items, we design a cold-start item embedding generator, which leverages the multimodal information to warm up the ID embedding for new items. The overall architecture of MML is illustrated in Figure 1(a).

In the rest of this section, we introduce our multimodal meta-learner group, which consists of three Transformer-based meta-learners in Section 4.2. Then in Section 4.3, we present the architecture of cold-start item embedding generator. Finally, we will introduce the training strategy of MML in Section 4.4.

4.2 Multimodal Meta-learner Group

To better model the sequential correlations among items in different modalities, based on the Transformer-based sequential model SASRec [16], we design a multimodal meta-learner group which consists of three meta-learners to model the interaction sequence in multiple aspects. Specifically, as shown in Figure 1(a), for an multimodal interaction sequence, we design a main meta-learner with feature-aware self-attention networks (SANs) for ID sequence and two auxiliary meta-learners with vanilla SANs for text and image sequences. And finally, we integrate the prediction result of



(a) The overall architecture of MML. **(b) The training strategy of MML.**

Figure 1: The overall architecture and training strategy of MML.

three meta-learners by an attention-based fusion layer. Next, we will discuss the implementation of each part in detail.

4.2.1 Main Meta-Learner. On the basis of SASRec, we adopt the main meta-learner to model the ID sequence, consisting of three layers: embedding layer, feature-aware self-attention block and prediction layer.

Embedding layer. Given a n-length item sequence, we maintain an item embedding matrix $\mathbf{M}_{id} \in \mathbb{R}^{|\mathcal{I}|\times d}$ and apply a look-up operation to obtain the ID embedding sequence $\mathbf{E}_{id} \in \mathbb{R}^{n\times d}$. Different from SASRec, the main meta-learner also utilizes text and image information of items to enhance sequence representations. For this purpose, we adopt two pretrained encoders (BERT [3] and ResNet [9]) to obtain the text representation $\mathbf{E}_{te} \in \mathbb{R}^{n\times d}$ and image representation $\mathbf{E}_{im} \in \mathbb{R}^{n\times d}$ of the input sequence.

Feature-aware self-attention block. In order to better model the sequential characteristics of user behaviors, we explore multimodal side information to capture intrinsic correlations among items. Intuitively, users might tend to interact with items having similar side information. However, the vanilla SANs in SASRec can't model the side information of items. To guide the learning of attention scores, we propose a *feature-aware self-attention* layer by incorporating text and image information of items through the attentional bias [19, 20]. Specifically, we implement the feature-aware self-attention layer with multi-head attention mechanisms as follows:

$$\mathbf{H}^l = [head_1, head_2, ..., head_h]\mathbf{W}^O, \tag{1}$$

$$head_i = \mathrm{FATT}(\mathbf{F}^l\mathbf{W}_i^Q, \mathbf{F}^l\mathbf{W}_i^K, \mathbf{F}^l\mathbf{W}_i^V, \mathbf{B}_f\mathbf{W}_i^B), \tag{2}$$

$$\mathbf{B}_f = (\mathbf{E}_{te}\mathbf{E}_{te}^\top) \odot \mathbf{W}_{te} + (\mathbf{E}_{im}\mathbf{E}_{im}^\top) \odot \mathbf{W}_{im}, \tag{3}$$

where the $\mathbf{F}^l$ is the input for the l-th layer (when $l = 0$, we set $\mathbf{F}^0 = \mathbf{E}_{id}$), $\mathbf{B}_f$ is the attention bias derived according to the multimodal side information, and the projection matrix $\mathbf{W}_i^Q \in \mathbb{R}^{d\times d/h}$, $\mathbf{W}_i^K \in \mathbb{R}^{d\times d/h}$, $\mathbf{W}_i^Q \in \mathbb{R}^{d\times d/h}$, $\mathbf{W}_i^B \in \mathbb{R}^{d\times d/h}$, $\mathbf{W}_{te} \in \mathbb{R}^{n\times n}$, $\mathbf{W}_{im} \in \mathbb{R}^{n\times n}$ and $\mathbf{W}^O \in \mathbb{R}^{d\times d}$ are the corresponding learnable parameters for each attention head. The feature-aware attention function is implemented by scaled dot-product operation:

$$\mathrm{FATT}(\mathbf{Q}, \mathbf{K}, \mathbf{V}, \mathbf{B}) = \mathrm{softmax}(\frac{\mathbf{Q}\mathbf{K}^\top + \mathbf{B}}{\sqrt{d/h}})\mathbf{V}, \tag{4}$$

where $\mathbf{Q}$, $\mathbf{K}$ and $\mathbf{V}$ respectively denote the queries, keys, and values of items in the sequence same as the vanilla SANs, and $\mathbf{B}$ is the attention bias (Eq. 3) and the $\sqrt{d/h}$ is the scale factor to avoid large values of the inner product. Here, we learn *multimodal feature-level correlations* for better capturing *item-level correlations*.

After performing multi-head self-attention, we further apply a point-wise feed-forward network on $\mathbf{H}^l$ to model interactions

between different latent dimensions as follows:

$$\mathbf{F}^{l+1} = [\mathrm{FFN}(\mathbf{H}_1^l)^\top; \cdots; \mathrm{FFN}(\mathbf{H}_n^l)^\top], \tag{5}$$

$$\mathrm{FFN}(x) = (\mathrm{ReLU}(x\mathbf{W}_1 + \mathbf{b}_1))\mathbf{W}_2 + \mathbf{b}_2, \tag{6}$$

where $\mathbf{W}_1 \in \mathbb{R}^{d\times d}$, $\mathbf{W}_2 \in \mathbb{R}^{d\times d}$, $\mathbf{b}_1 \in \mathbb{R}^d$ and $\mathbf{b}_2 \in \mathbb{R}^d$ are trainable parameters.

Prediction layer. After L-layer self-attention blocks, in the final layer, we estimate the user's preference score $\mathbf{P}^{id}(t+1)$ for c candidate items based on user's interaction sequence as follows:

$$\mathbf{p}^{id} = [p_1; \cdots; p_c], \tag{7}$$

$$p_i = \mathbf{e}_{id}^\top \mathbf{F}^L, \tag{8}$$

where $\mathbf{e}_{id}$ is the ID representation of item i from item embedding matrix $\mathbf{M}_I$, $\mathbf{F}^L$ is the output of the L-layer self-attention block and L is the number of self-attention blocks.

4.2.2 Auxiliary Meta-learner. In the above main meta-learner, we utilize multimodal data as auxiliary information to enhance the learning of sequential characteristics in terms of ID features. While, intuitively, the interaction sequence is likely to reflect some specific kind of correlation patterns from a single modality (either text or image). In order to better stabilize and improve the learning of main meta-learner, we further design two auxiliary meta-learners for the two modalities, respectively. Unlike main meta-learner, auxiliary meta-learners adopt *vanilla self-attention function* with the pretrained encoders (BERT for text and ResNet for image), and we can compute the preference scores $\mathbf{p}^{te}$ and $\mathbf{p}^{im}$ for c candidate items from two auxiliary meta-learners respectively.

4.2.3 Multimodal Prediction Fusion. Given the group of multimodal meta-learners, we design an attention-based fusion layer to integrate their predictions to generate more reliable predictions. Specifically, we firstly apply the layer normalization [1] on the preference scores of c candidate items and the output of the L-layer self-attention blocks from three meta-learners:

$$\mathbf{P}' = LN([\mathbf{p}^{id\top}; \mathbf{p}^{te\top}; \mathbf{p}^{im\top}]), \tag{9}$$

$$\mathbf{F} = LN([\mathbf{F}_{id}^{L\top}; \mathbf{F}_{te}^{L\top}; \mathbf{F}_{im}^{L\top}]), \tag{10}$$

where $\mathbf{P}' \in \mathbb{R}^{3\times c}$ and $\mathbf{F} \in \mathbb{R}^{3\times nd}$. And then, we apply the attention function which is implemented by a single-layer feed-forward neural network, parameterized by a weight matrix $\mathbf{W}_{att}$ and a bias vector $\mathbf{b}_{att}$, and the ReLU non-linearity. Besides, we apply the softmax normalization to make coefficients easily comparable across different modalities calculate the attention score $\mathbf{A} \in \mathbb{R}^{1\times 3}$ as follow:

$$\mathbf{A} = \mathrm{Softmax}(\mathrm{Att}(\mathbf{FW} + \mathbf{b})), \tag{11}$$

$$ATT(x) = \mathrm{ReLU}(x)\mathbf{W}_{att} + \mathbf{b}_{att}, \tag{12}$$

where $\mathbf{W}$ and $\mathbf{b}$ is the parameters of liner projection.

Finally, we can calculate the prediction scores $\mathbf{p} \in \mathbb{R}^c$ of all the candidate items given sequence $s = \{r_1, \cdots, r_t\}$ as:

$$\mathbf{p} = \mathbf{AP}'. \tag{13}$$

Recent studies have shown that meta-learning is likely to be unstable and produce unreliable predictions [44, 54], especially with limited training data. Similar to the meta-learner ensemble [28], we set up multimodal meta-learners and sufficiently fuse their predictions, which can help stabilize the learning of meta-learners. Instead of simply combining the prediction results, our approach lets the model learn how to select and fuse the predictions from different meta-learners. In the above approach, multimodal side information has been utilized in two ways: (1) the enhancement of sequence representations in the main meta-learner and (2) the learning of the auxiliary meta-learner.

4.3 Cold-start Item Embedding Generator

Besides sequential characteristics, there also exists divergence between old and new items, where new item IDs don't appear in cold user sequences, called *cold-start items*. Compared with the well-trained item embeddings in old sequences, the randomly generated ID embeddings of these cold-start items have negative impact on knowledge transfer to new sequences. Considering this issue, we further propose a cold-start item embedding generator to warm up item ID embeddings in new user' sequences by leveraging the multimodal information. Although there have been some studies on the cold-start embedding warm-up [25, 27], considering the model complexity, we design a simple yet effective attention network as the generator to address this problem.

As shown in Figure 1(a), given a ID sequence of new user, we firstly apply a look-up operation on $\mathbf{M}_{id}$ to identify the new items, then ignore the old items and generate the warm-up embedding for each new item only. Given a new item z and the representations of its text information $\mathbf{e}_z^{\mathrm{txt}}$ and image information $\mathbf{e}_z^{\mathrm{img}}$, we select the top K most similar old items $O = \{o\}$ according to the following similarity function:

$$q_{z,o} = \frac{\sum_m \mathrm{cosine}(\mathbf{e}_z^m, \mathbf{e}_o^m)}{|\mathcal{M}|}, \tag{14}$$

where $q_{z,o}$ is the similarity score of o, and $m \in \mathcal{M} = \{\mathrm{txt}, \mathrm{img}\}$ is the modality set. Then we apply the attention mechanism to calculate the attention weight for each old item $o \in O$ *w.r.t.* z:

$$a_o^m = \frac{\exp(\mathbf{W}_1^m[\mathbf{Y}^m\mathbf{e}_z^m || \mathbf{Y}^m\mathbf{e}_o^m])}{\sum_{j=1}^K \exp(\mathbf{W}_1^m[\mathbf{Y}^m\mathbf{e}_z^m || \mathbf{Y}^m\mathbf{e}_j^m])}, \tag{15}$$

where W_1^m and Y^m are the weight parameters of liner projection in modality m, e_o^m is the side information of o in modality m.

Furthermore, we generate the embedding from different modalities by the side information of $o \in O$ and z :

$$\mathbf{g}_z^m = \mathrm{ReLU}(\mathbf{Y}^m\mathbf{e}_z^m + \sum_{o=1}^K a_o^m\mathbf{Y}^m\mathbf{e}_o^m), \tag{16}$$

where g_z^m is the embedding generated by side information in modality m. Finally, we combine the generated embeddings in different modalities as:

$$\mathbf{e}_z^{id} = \frac{1}{|\mathcal{M}|}\sum \mathbf{g}_z^m, \tag{17}$$

4.4 The Training Algorithm of MML

To achieve fast adaption to cold-start users with insufficient data, we extend MAML [7] to our scenario and design a three-stage algorithm to train the meta-learners, the embedding generator and

the prediction fusion layer in turn as shown in Figure 1(b). Next, we will introduce our training algorithm in detail.

4.4.1 Data Preparation. As mentioned in Section 3, we regard the next-item prediction task on each sequence as a task and utilize the D_{old} and D_{new} as the training data and test data respectively. For D_{old}, we apply the data augmentation strategy on each sequence and split it into the query set and support set. Specifically, for each user u, we firstly employ data augmentation on his/her interaction sequence $s_u = \{r_1, \cdots, r_n\}$ following previous methods [22, 53] (*e.g.,* a sequence (r_1, r_2, r_3, r_4) is divided into three successive sequences: $\langle r_1, r_2\rangle$, $\langle r_1, r_2, r_3\rangle$, $\langle r_1, r_2, r_3, r_4\rangle$), then these sequences will be split into two sets. The first $n-1$ sequences form the support set D_u^S and the n-th sequence forms the query set D_u^Q.

4.4.2 Training Procedure. Our training procedure is descried in the following three stages:

Stage 1: training the meta-learners. Firstly, we train the three meta-learners on the D_{old} with the same task (*i.e.,* the next-item prediction task), which aims to predict the last item in each sequence based on all the previous items. For the sake of notation simplicity, we denote all the trainable parameters in embedding layer as Γ and the trainable parameters in self-attention blocks and prediction layer as Θ. Following the MAML [7], there are two steps in meta-training, *i.e.,* local update and global update. During local update, we update the Θ on each sequence b in the batch B by minimizing the BPR loss [32] on the support set D_b^S:

$$\Theta^b = \Theta - \alpha \nabla_{\Theta^b} Loss(D_b^S; \Theta^b), \tag{18}$$

where Θ is the global parameter, Θ^b is the local parameter learned via sequence b and α is the learning rate. Following [18], we don't update Γ in local update to ensure the stability of training process. During global update, we update the global parameter Γ and Θ on the query set D^Q by one gradient step on the sum of all the losses:

$$\Theta = \Theta - \beta \sum_{b \in B} \nabla_{\Theta} Loss(D_b^Q; \Theta^b), \tag{19}$$

$$\Gamma = \Gamma - \beta \sum_{b \in B} \nabla_{\Gamma} Loss(D_b^Q; \Theta^b), \tag{20}$$

where β is the learning rate. The three meta-learners follow the same meta-training process.

Stage 2: training the item embedding generator. Based on the trained main meta-learner in Stage 1, we train the cold-start item embedding generator with training data D_{old}. As mentioned in 4.3, the goal of the generator is to warm up the ID embeddings for cold-start items. Since all the items in training data are old items (their ID embeddings are trained in meta-learning process), to simulate the cold-start situation, for each sequence s^{id} in D_{old}, we will randomly "forget" some items' embedding in s^{id} according to a proportion of p_0, and regard them as the new items. And then, we fix all the parameters in main meta-learner and train the generator based on the recommendation loss on s^{id}. Here we choose the BPR loss [32] as the recommendation loss.

Stage 3: training the prediction fusion layer. Based on the trained meta-learners and the cold-start ID embedding generator, finally, we train the prediction fusion layer. Similar to Stage 2, we sample some of the items as new items to simulate the cold-start situation, and then fix the parameters of all the other modules and train the fusion layer only. The training task in Stage 3 remains the next-item prediction. Here we set the cross entropy loss for parameter update.

Table 1: Statistics of our datasets.

Dataset	Type	#User	#Item	#Inter	#Inter/User
Shanghai	Training set	18,055	81,765	463,335	25.6624
	Test set	1,448	81,765	14,616	10.0939
Hangzhou	Training set	10,861	50,777	279,865	25.7679
	Test set	972	50,777	10,408	10.7078
Changsha	Training set	9,209	40,519	258,902	28.1140
	Test set	943	40,519	9,530	10.1060

4.5 Time Complexity Analysis

For recommendation algorithms, inference time is more important to consider than training time. Next, we analyze the complexity of the inference time with our approach MML. Overall, for each request (*i.e.,* the recommendation to a user), the inference time cost of MML mainly comes from three parts, *i.e.,* the cold-start item embedding generator, three meta-learners and the prediction fusion layer. To predict the next item for a new user' interaction sequence, we firstly utilize the cold-start item embedding generator to warm up the ID embedding of new items, in which we use the Faiss [15] toolkit to retrieve the top-K nearest neighbors, and then we generate the embedding for cold-start items based on their neighbors. Considering the complexity of such a retrieval process in Faiss is sublinear, we can omit the cost of neighbors retrieval and the time complexity of this part is $O(nKd)$ (n-length sequences). As for the meta-learners, we will calculate the self-attention matrices for the sequence for three modalities in parallel, and the complexity of this part is $O(n^2dL)$ (L layers). Finally, we combine the prediction from three meta-learners via the prediction fusion layer. We need to calculate the attention scores for each modality and obtain the final prediction by the attentional sum, which leads to a time complexity of $O(ndh + c)$, where h is the hidden layer size and c is the number of candidate items.

As K and h are often small, thus, the total time complexity of the MML can be roughly written as $O(n^2dL + c)$. Such a process can be also further accelerated by parallel computing resources, which is acceptable in real-world online platform.

5 EXPERIMENT

We conduct experiments to answer the following questions:

- **RQ1:** How is the performance of MML compared with competitive baselines in offline evaluation?
- **RQ2:** How is the performance of MML in online deployment under real-world performance metrics?
- **RQ3:** What is the effect of different components of MML on its effectiveness?

Table 2: The overall performance. The best result is bolded and the runner-up is underlined. * indicates the statistical significance for $p < 0.01$ compared to the best baseline. The Impro. shows the improvement ratio of MML compared with the runner-up.

Dataset	Metric	Non-transfer based			Transfer based			Ours	
		SASRec	GRU4Rec	Full	SML	PT-FT	CBML	MML	Impro.
Shanghai	Recall@5	0.0925	0.0414	0.1174	0.1221	0.1360	0.2183	**0.2209**	+1.19%
	Recall@10	0.1153	0.0656	0.1533	0.1547	0.1547	0.2527	**0.2570**	+1.70%
	Recall@20	0.1478	0.0822	0.1899	0.2007	0.1637	0.2661	**0.2690**	+1.09%
	MRR@5	0.0678	0.0297	0.0820	0.0766	0.0790	0.0790	**0.0978**	+19.27%
	MRR@10	0.0708	0.0330	0.0867	0.0853	0.0815	0.0826	**0.1016**	+17.19%
	MRR@20	0.0729	0.0342	0.0893	0.0899	0.0822	0.0835	**0.1025**	+14.02%
	NDCG@5	0.0739	0.0327	0.0908	0.0977	0.0933	0.1138	**0.1287**	+13.09%
	NDCG@10	0.0811	0.0405	0.1024	0.1021	0.0993	0.1240	**0.1394**	+12.42%
	NDCG@20	0.0893	0.0447	0.1116	0.1088	0.1017	0.1274	**0.1423**	+11.70%
Hangzhou	Recall@5	0.0813	0.0134	0.0926	0.1223	0.1224	0.0918	**0.1910**	+56.05%
	Recall@10	0.1049	0.0226	0.1296	0.1558	0.1718	0.2833	**0.2960**	+4.48%
	Recall@20	0.1358	0.0453	0.1656	0.1766	0.1924	0.2890	**0.2986**	+3.32%
	MRR@5	0.0569	0.0076	0.0648	0.0599	0.0751	0.0874	**0.1135**	+29.86%
	MRR@10	0.0600	0.0089	0.0698	0.0768	0.0816	0.1111	**0.1264**	+13.77%
	MRR@20	0.0620	0.0104	0.0723	0.0785	0.0830	0.1115	**0.1266**	+13.54%
	NDCG@5	0.0629	0.0090	0.0717	0.0876	0.0868	0.0885	**0.1317**	+48.81%
	NDCG@10	0.0705	0.0121	0.0837	0.0988	0.1027	0.1487	**0.1646**	+10.69%
	NDCG@20	0.0782	0.0178	0.0928	0.0895	0.1079	0.1501	**0.1653**	+10.13%
Changsha	Recall@5	0.0954	0.0350	0.1060	0.1355	0.1029	0.1420	**0.1555**	+9.54%
	Recall@10	0.1145	0.0626	0.1495	0.1562	0.1548	0.1502	**0.1657**	+6.08%
	Recall@20	0.1410	0.0901	0.1845	0.1744	0.1888	0.1570	**0.1956**	+3.60%
	MRR@5	0.0622	0.0194	0.0715	0.0866	0.0666	0.1343	**0.1460**	+8.71%
	MRR@10	0.0646	0.0228	0.0771	0.0895	0.0738	0.1354	**0.1473**	+8.79%
	MRR@20	0.0664	0.0245	0.0795	0.0911	0.0761	0.1359	**0.1488**	+9.49%
	NDCG@5	0.0705	0.0232	0.0800	0.0988	0.0756	0.1362	**0.1483**	+8.88%
	NDCG@10	0.0765	0.0319	0.0939	0.1088	0.0926	0.1389	**0.1516**	+9.14%
	NDCG@20	0.0832	0.0386	0.1027	0.1158	0.1012	0.1406	**0.1557**	+10.74%

• **RQ4:** How is the performance of MML with limited training data?

In this section, we first present experimental settings, followed by results and analyses to answer each research question. We conduct our experiments based on the framework of RecBole [49, 50] and our code is released at https://github.com/RUCAIBox/MML.

5.1 Experimental Settings

5.1.1 Datasets. We conduct the offline experiments on three recommendation datasets collected from the large-scale and real-world click logs covering meals, hotel, tourism and other 6 businesses in Meituan[1] App from June to October 2021. In order to obtain the multimodal information, we crawl the textual title and product photo of each item in our dataset. The three datasets are collected from three different cities: Shanghai, Hangzhou and Changsha. We split the datasets into two parts: training set and test set. To simulate the cold-start scenario, we set a reference timestamp T as *September 10*, 2021. All the interactions before T will be selected into training set D_{old}, and users in D_{old} will be considered as old users. The rest of users will be considered as new users, and the interactions of these users are considered as test set D_{new}. The details of the data statistics are shown in Table 1.

5.1.2 Baselines. We compared the proposed MML method with two categories of recommendation methods: **non-transfer based methods** and **transfer based methods**, which either directly train the model based on entire training dataset or utilize some kind of knowledge transfer techniques. Here we select three non-transfer based methods as follows:

- **SASRec** [16]: It is the base recommendation modal of MML, which adapts the Transformer architecture for recommending the next item.
- **GRU4Rec** [10]: It utilizes the GRU network to model user click sequences for session-based recommendation. We represent the items using embedding vectors rather than one-hot vectors.
- **Full Retrain**: It is a model retraining baseline purposed in [47], the idea is to combine the training data of old and new users together to train the model, and then evaluate the performance only on the test data of new users.

[1] https://www.meituan.com

Besides, we select three transfer based methods as the baselines. Similar to MML, the transfer based methods will design specific knowledge transfer strategy on the training set and then help the model quickly adapt to test set.

- **SML** [47]: SML is a meta-learning based method for model retraining, which can also be used in our scenario. Specifically, we firstly train a recommender model on the data of old users and retrain this model on the new collected data of new users following the SML method.
- **Pre-train and Fine-tune (PT-FT)**: It is a widely used baseline for knowledge transfer [35, 47]. Different from meta-learning approach, this method trains the model in a multi-task learning manner. For fair comparison, we keep the same model structure with our MML and apply PT-FT method for model training. Specifically, we pretrain the model on the training set, save the best weights and then fine-tune on the support set of each cold-start user in test set, and evaluate the performance on the query set of each cold-start user.
- **CBML** [35]: Similar to our method, this method also focuses on the cold-start sequential recommendation and proposes a clustering-based meta-learning model, which clusters the similar sequences based on item attributes. In our dataset, we apply the item category as the item attributes to evaluate the performance of CBML.

5.1.3 Hyper-parameter Tuning. For fair comparison, we set the training batch size as 2048 and embedding size as 64 for all the comparison methods. For each non-transfer based method, we tune the learning rate in {1e-2, 7e-3, 5e-3, 3e-3, 2e-3, 1e-3} for the superior performance. For all the meta-learning based methods, we set the training epoch as 10. The number of local update in each training epoch is set as 5. We fix the local learning rate α as 1e-3 and tune the learning rate β of global update in {1e-4, 3e-4, 5e-4, 7e-4, 1e-3}. For our MML method, we set the hidden size as 128 in prediction fusion layer, and we select 5e-4, 1e-3 and 1e-3 as the learning rate β of global update for Shanghai, Hangzhou and Changsha respectively. All the other hyper-parameters of baselines are tuned following the suggestions in the original papers.

5.2 Offline Evaluation (RQ1)

To verify the effectiveness of our method, we evaluate the performance of MML with other baselines on the three collected datasets, *i.e.,* Shanghai, Hangzhou and Changsha datasets.

5.2.1 Evaluation Metrics. For evaluation metrics, we adopt three widely used metrics of top-K recommendation to evaluate the ranking list of recommendations: *Recall@K*, *Normalized Discounted Cumulative Gain (NDCG@K)* and *Mean Reciprocal Rank (MRR@K)* ($K \in \{5, 10, 20\}$). Since the sampled metrics may be unreliable [17, 48], we generate the ranking list for each user by considering all the items and calculate the metrics based on the ranking of all the candidate items. The listed results are averaged over all test users and all the metrics are calculated on the query set of cold-start users.

5.2.2 Experimental result. The experimental results on the three datasets are reported in Table 2, and we have the following important observations:

Table 3: The result of A/B test in online scenario. The improvement is calculated under the statistical significance for $p < 0.05$ compared to the baseline.

Method	CTR	CVR	New Pay
MML	+1.212%	+1.203%	+2.362%

1) As shown in all three datasets, the proposed model MML observably outperforms all the other baselines by large margin, including the strong meta-learning based baseline CBML, which shows the effectiveness of the proposed multimodal meta-learning approach. Compared with CBML, MML has two major merits in cold-start sequential recommendation. First, instead of simply leveraging the coarse-grain item attributes like item category, we design a more principled approach to leveraging multimodal side information, which can stabilized and improve the meta-learning process. Second, we have considered addressed the characteristic divergence at both item and sequence levels, thus leading to a better recommendation performance for cold-start users.

2) In all the datasets, simply training a deep recommender only on the support set of a cold-start user impairs the performance severely, since the support set only contains very limited interaction data. Among these baselines, the *pre-train and fine-tune* (PT-FT) method gains more improvement compared with no-transfer baselines, because PT-FT can leverage the knowledge learned from old users. However, we find the effective meta-learning strategy can further improve the recommendation performance, which is specially designed for fast adaption given limited data, indicating the importance of meta-learning algorithms in this test scenario.

5.3 Online A/B Tests (RQ2)

To further examine the performance of MML in real-world application scenarios, we conduct an online A/B test through the function labeled by "*Guess you like*" in Meituan App.

5.3.1 Evaluation Setup. Since the function "*Guess you like*" provides the recommendation service to all the Meituan users, we sample a traffic of nearly 17 million users for online evaluation. It is originally implemented based on a classic industrial implementation of multi-stage ranking systems, and we deploy the proposed MML approach as a new retrieval strategy in the recall module. For the sampled test users, following a standard setup for A/B test, we divide them into two groups: one with the original recall module and the other one with the improved recall module by our MML approach. For evaluation metrics, following [52], we adopt two widely used performance measurements CTR (Click-Through-Rate) and CVR (Click Conversion Rate). Besides, we also calculate the number of orders for new users, who have no order record in the past year, denoted as *New Pay*. For company privacy, we don't report the implementation details of the original recall module and the real performance. Instead, we report the performance gain ratio improved by our approach MML.

5.3.2 Experiment Result. The results of online A/B test are shown in Table 3. From the table, we can find that:

Table 4: The result of ablation study.

Method	Shanghai		Hangzhou	
	MRR@10	NDCG@10	MRR@10	NDCG@10
No-Gen	0.0935	0.1150	0.0877	0.1090
No-Fusion	0.0934	0.1130	0.0888	0.1097
MML_{ID}	0.0959	0.1169	0.0911	0.1126
MML	0.1016	0.1394	0.1264	0.1646

1) MML outperforms the online baseline on all the three metrics, which indicates that MML is capable of improving the effectiveness of cold-start recommendation on Meituan App. It is worth noting that the improvement ratios in Table 3 is significant after deploying the MML approach: one percent usually indicates a large improvement of the recommendation capacity in real-world application scenario, when tested on a large population of users.

2) Besides the click-based metrics of CTR and CVR, MML also leads to direct performance improvement based on the *New Pay* metric. This improvement demonstrates that the proposed MML can bring positive impact to the underlying recommender system, leading to more purchases from new users who have not paid before.

5.4 Ablation Study (RQ3)

Since our MML approach contains several important technical improvements, here we analyze their contribution to the recommendation performance by removing each one while keeping the rest. In particular, we consider the following two variants by ablating the ID embedding generator or the prediction fusion layer:

- **NO-Gen**: A variant of MML which replaces the item embedding generator by randomly generated embeddings for new items.
- **NO-Fusion**: A variant of MML which replaces the knowledge fusion layer by an average fusion for three meta-learners.

Besides, we also prepare an ID-only version for our MML approach in order to examine the effect of multimodal side information, denoted by MML_{ID}.

As shown in Table 4, we conduct the ablation experiments on the two datasets (*i.e.,* Shanghai and Hangzhou), and then evaluate the performance by MRR@10 and NDCG@10. From this table, we can observe that removing any one component from the MML approach will lead to a performance decrease. Such a finding shows that both components are important to improve the recommendation performance for cold-start users. Besides, we find that MML_{ID} also performs worse than the complete MML approach. It indicates that multimodal side information is indeed useful to boost the performance of our approach.

5.5 Influence Analysis of Training Data (RQ4)

A major advantage of meta-learning based methods is that they perform well under the few-shot setting [7]. Here, to examine the robustness of MML in few-shot scenarios, we conduct a specific experiment by varying the scale of the training data and evaluate the performance of MML on the test set, which will be fixed in this experiment. For comparison, we select the competitive baseline PT-FT for performance reference.

As shown in Figure 2, we report the performance of MML and PT-FT on the Hangzhou dataset with different sampling ratios. Here, we adopt Recall@10 (left subfigure) and Recall@20 (right subfigure) as the evaluation metrics. With the reduction of training data, the performance of both MML and PT-FT decrease accordingly. However, the performance of MML seems to be more stable than PT-FT and outperforms PT-FT in all the cases. The result shows our MML is more robust and can perform well even with very limited training data.

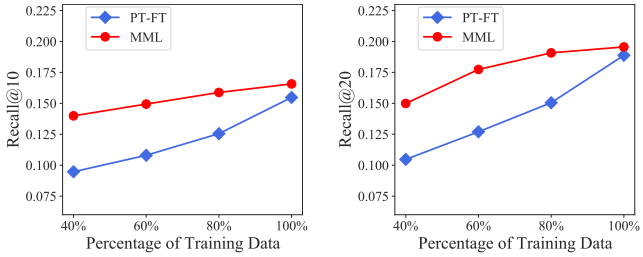


Figure 2: Performance of recommendation with different scale of training data.

6 CONCLUSION

In this paper, we presented the Multimodal Meta-Learning (denoted as MML) approach by leveraging multimodal side information of items for cold-start sequential recommendation. We designed modality-specific meta-learner to capture the sequential characteristics from different perspectives (*i.e.,* modalities), and adaptively integrated their predictions with a learnable fusion layer. These meta-learners can effectively reduce sequence-level divergence between old and new interaction sequences. Besides, we also designed a cold-start item embedding generator to warm up ID embeddings of new items. Our approach effectively reduced the characteristic divergence at both the item and sequence levels. Extensive experimental results show that the proposed MML outperforms several competitive baselines in both offline and online evaluation, leading to direct performance improvement after being deployed on Meituan App. To the best of our knowledge, it is the first time that multimodal side information of items have been leveraged to improve the meta-learning for cold-start sequential recommendation.

As future work, we will further consider extending our framework to other recommendation scenarios, such as cross-domain recommendation. Besides, we will explore the use of more modalities, such as audio and video, for enriching the side information.

ACKNOWLEDGEMENT

This work was partially supported by National Natural Science Foundation of China under Grant No. 61832017, Beijing Natural Science Foundation under Grant No. 4222027 and Beijing Outstanding Young Scientist Program under Grant No. BJJWZYJH012019100020098. This work is also partially supported by Beijing Academy of Artificial Intelligence (BAAI) and Meituan. We also sincerely thank non-author team members (Hongyu Wang and Chuyuan Wang) from Meituan for the guidance on online A/B tests. He Hu and Xin Zhao are the corresponding authors.

REFERENCES

[1] Lei Jimmy Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. 2016. Layer Normalization. *arXiv preprint arXiv:1607.06450* (2016).

[2] Jianxin Chang, Chen Gao, Yu Zheng, Yiqun Hui, Yanan Niu, Yang Song, Depeng Jin, and Yong Li. 2021. Sequential Recommendation with Graph Neural Networks. In *SIGIR*. ACM, 378–387.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT (1)*. Association for Computational Linguistics, 4171–4186.

[4] Manqing Dong, Feng Yuan, Lina Yao, Xiwei Xu, and Liming Zhu. 2020. MAMO: Memory-Augmented Meta-Optimization for Cold-start Recommendation. In *KDD*. ACM, 688–697.

[5] Yuntao Du, Xinjun Zhu, Lu Chen, Ziquan Fang, and Yunjun Gao. 2022. MetaKG: Meta-learning on Knowledge Graph for Cold-start Recommendation. *CoRR* abs/2202.03851 (2022).

[6] Hui Fang, Danning Zhang, Yiheng Shu, and Guibing Guo. 2020. Deep learning for sequential recommendation: Algorithms, influential factors, and evaluations. *ACM Transactions on Information Systems (TOIS)* 39, 1 (2020), 1–42.

[7] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *ICML (Proceedings of Machine Learning Research, Vol. 70)*. PMLR, 1126–1135.

[8] Casper Hansen, Christian Hansen, Jakob Grue Simonsen, Stephen Alstrup, and Christina Lioma. 2020. Content-aware Neural Hashing for Cold-start Recommendation. In *SIGIR*. ACM, 971–980.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *CVPR*. 770–778.

[10] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based Recommendations with Recurrent Neural Networks. In *ICLR*.

[11] Balázs Hidasi, Massimo Quadrana, Alexandros Karatzoglou, and Domonkos Tikk. 2016. Parallel Recurrent Neural Network Architectures for Feature-rich Session-based Recommendations. In *RecSys*. ACM, 241–248.

[12] Jin Huang, Zhaochun Ren, Wayne Xin Zhao, Gaole He, Ji-Rong Wen, and Daxiang Dong. 2019. Taxonomy-Aware Multi-Hop Reasoning Networks for Sequential Recommendation. In *WSDM*. ACM, 573–581.

[13] Jin Huang, Wayne Xin Zhao, Hongjian Dou, Ji-Rong Wen, and Edward Y. Chang. 2018. Improving Sequential Recommendation with Knowledge-Enhanced Memory Networks. In *SIGIR*. ACM, 505–514.

[14] Xiaowen Huang, Jitao Sang, Jian Yu, and Changsheng Xu. 2022. Learning to Learn a Cold-start Sequential Recommender. *ACM Trans. Inf. Syst.* 40, 2 (2022), 30:1–30:25.

[15] Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2019. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data* 7, 3 (2019), 535–547.

[16] Wang-Cheng Kang and Julian J. McAuley. 2018. Self-Attentive Sequential Recommendatiosn. In *ICDM*. 197–206.

[17] Walid Krichene and Steffen Rendle. 2020. On Sampled Metrics for Item Recommendation. In *KDD*. ACM, 1748–1757.

[18] Hoyeop Lee, Jinbae Im, Seongwon Jang, Hyunsouk Cho, and Sehee Chung. 2019. MeLU: Meta-Learned User Preference Estimator for Cold-Start Recommendation. In *KDD*. ACM, 1073–1082.

[19] Jiacheng Li, Yujie Wang, and Julian J. McAuley. 2020. Time Interval Aware Self-Attention for Sequential Recommendation. In *WSDM*. ACM, 322–330.

[20] Tianyang Lin, Yuxin Wang, Xiangyang Liu, and Xipeng Qiu. 2021. A Survey of Transformers. *arXiv preprint arXiv:2106.04554* (2021).

[21] Xixun Lin, Jia Wu, Chuan Zhou, Shirui Pan, Yanan Cao, and Bin Wang. 2021. Task-adaptive Neural Process for User Cold-Start Recommendation. In *WWW*. ACM / IW3C2, 1306–1316.

[22] Qiao Liu, Yifu Zeng, Refuoe Mokhosi, and Haibin Zhang. 2018. STAMP: Short-Term Attention/Memory Priority Model for Session-based Recommendation. In *KDD*. ACM, 1831–1839.

[23] Yuanfu Lu, Yuan Fang, and Chuan Shi. 2020. Meta-learning on Heterogeneous Information Networks for Cold-start Recommendation. In *KDD*. 1563–1573.

[24] Yao Ma, Shilin Zhao, Weixiao Wang, Yaoman Li, and Irwin King. 2022. Multimodality in Meta-Learning: A Comprehensive Survey. *Knowledge-Based Systems* 250 (2022), 108976. https://doi.org/10.1016/j.knosys.2022.108976

[25] Wentao Ouyang, Xiuwu Zhang, Shukui Ren, Li Li, Kun Zhang, Jinmei Luo, Zhaojie Liu, and Yanlong Du. 2021. Learning Graph Meta Embeddings for Cold-Start Ads in Click-Through Rate Prediction. In *SIGIR*. ACM, 1157–1166.

[26] Cosimo Palmisano, Alexander Tuzhilin, and Michele Gorgoglione. 2008. Using Context to Improve Predictive Modeling of Customers in Personalization Applications. *IEEE Trans. Knowl. Data Eng.* 20, 11 (2008), 1535–1549.

[27] Feiyang Pan, Shuokai Li, Xiang Ao, Pingzhong Tang, and Qing He. 2019. Warm Up Cold-start Advertisements: Improving CTR Predictions via Learning to Learn ID Embeddings. In *SIGIR*. ACM, 695–704.

[28] Minseop Park, Jungtaek Kim, Saehoon Kim, Yanbin Liu, and Seungjin Choi. 2019. MxML: Mixture of Meta-Learners for Few-Shot Classification. *arXiv preprint arXiv:1904.05658* (2019).

[29] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. 2017. Personalizing Session-based Recommendations with Hierarchical Recurrent Neural Networks. In *RecSys*. ACM, 130–137.

[30] Pengjie Ren, Zhumin Chen, Jing Li, Zhaochun Ren, Jun Ma, and Maarten de Rijke. 2019. RepeatNet: A Repeat Aware Neural Recommendation Machine for Session-Based Recommendation. In *AAAI*. AAAI Press, 4806–4813.

[31] Steffen Rendle. 2010. Factorization Machines. In *ICDM*. 995–1000.

[32] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI*. AUAI Press, 452–461.

[33] Martin Saveski and Amin Mantrach. 2014. Item cold-start recommendations: learning local collective embeddings. In *RecSys*. ACM, 89–96.

[34] Xiang-Rong Sheng, Liqin Zhao, Guorui Zhou, Xinyao Ding, Binding Dai, Qiang Luo, Siran Yang, Jingshan Lv, Chi Zhang, Hongbo Deng, and Xiaoqiang Zhu. 2021. One Model to Serve All: Star Topology Adaptive Recommender for Multi-Domain CTR Prediction. In *CIKM*. ACM, 4104–4113.

[35] Jiayu Song, Jiajie Xu, Rui Zhou, Lu Chen, Jianxin Li, and Chengfei Liu. 2021. CBML: A Cluster-based Meta-learning Model for Session-based Recommendation. In *CIKM*. ACM, 1713–1722.

[36] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. In *CIKM*. ACM, 1441–1450.

[37] Jiaxi Tang and Ke Wang. 2018. Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding. In *WSDM*. ACM, 565–573.

[38] Maksims Volkovs, Guang Wei Yu, and Tomi Poutanen. 2017. DropoutNet: Addressing Cold Start in Recommender Systems. In *NIPS*. 4957–4966.

[39] Hongwei Wang, Fuzheng Zhang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2019. Multi-Task Feature Learning for Knowledge Graph Enhanced Recommendation. In *WWW*. ACM, 2000–2010.

[40] Jianling Wang, Kaize Ding, and James Caverlee. 2021. Sequential Recommendation for Cold-start Users with Meta Transitional Learning. In *SIGIR*. 1783–1787.

[41] Yaqing Wang and Quanming Yao. 2019. Few-shot Learning: A Survey. *arXiv preprint arXiv:1904.05046* (2019).

[42] Tianxin Wei, Ziwei Wu, Ruirui Li, Ziniu Hu, Fuli Feng, Xiangnan He, Yizhou Sun, and Wei Wang. 2020. Fast Adaptation for Cold-start Collaborative Filtering with Meta-learning. In *ICDM*. IEEE, 661–670.

[43] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-Based Recommendation with Graph Neural Networks. In *AAAI*. AAAI Press, 346–353.

[44] Han-Jia Ye and Wei-Lun Chao. 2021. How to Train Your MAML to Excel in Few-Shot Classification. (2021).

[45] Fajie Yuan, Xiangnan He, Alexandros Karatzoglou, and Liguang Zhang. 2020. Parameter-Efficient Transfer from Sequential Behaviors for User Modeling and Recommendation. In *SIGIR*. ACM, 1469–1478.

[46] Tianzi Zang, Yanmin Zhu, Haobing Liu, Ruohan Zhang, and Jiadi Yu. 2021. A Survey on Cross-domain Recommendation: Taxonomies, Methods, and Future Directions. *arXiv preprint arXiv:2108.03357* (2021).

[47] Yang Zhang, Fuli Feng, Chenxu Wang, Xiangnan He, Meng Wang, Yan Li, and Yongdong Zhang. 2020. How to Retrain Recommender System?: A Sequential Meta-Learning Method. In *SIGIR*. ACM, 1479–1488.

[48] Wayne Xin Zhao, Junhua Chen, Pengfei Wang, Qi Gu, and Ji-Rong Wen. 2020. Revisiting Alternative Experimental Settings for Evaluating Top-N Item Recommendation Algorithms. In *CIKM*. ACM, 2329–2332.

[49] Wayne Xin Zhao, Yupeng Hou, Xingyu Pan, Chen Yang, Zeyu Zhang, Zihan Lin, Jingsen Zhang, Shuqing Bian, Jiakai Tang, Wenqi Sun, et al. 2022. RecBole 2.0: Towards a More Up-to-Date Recommendation Library. In *CIKM*. ACM.

[50] Wayne Xin Zhao, Shanlei Mu, Yupeng Hou, Zihan Lin, Yushuo Chen, Xingyu Pan, Kaiyuan Li, Yujie Lu, Hui Wang, Changxin Tian, Yingqian Min, Zhichao Feng, Xinyan Fan, Xu Chen, Pengfei Wang, Wendi Ji, Yaliang Li, Xiaoling Wang, and Ji-Rong Wen. 2021. RecBole: Towards a Unified, Comprehensive and Efficient Framework for Recommendation Algorithms. In *CIKM*. ACM, 4653–4664.

[51] Yujia Zheng, Siyi Liu, Zekun Li, and Shu Wu. 2021. Cold-start Sequential Recommendation via Meta Learner. In *AAAI*. AAAI Press, 4706–4713.

[52] Guorui Zhou, Xiaoqiang Zhu, Chengru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep Interest Network for Click-Through Rate Prediction. In *KDD*. ACM, 1059–1068.

[53] Kun Zhou, Hui Wang, Wayne Xin Zhao, Yutao Zhu, Sirui Wang, Fuzheng Zhang, Zhongyuan Wang, and Ji-Rong Wen. 2020. S3-Rec: Self-Supervised Learning for Sequential Recommendation with Mutual Information Maximization. In *CIKM*. ACM, 1893–1902.

[54] Pan Zhou, Yingtian Zou, Xiao-Tong Yuan, Jiashi Feng, Caiming Xiong, and Steven C. H. Hoi. 2021. Task similarity aware meta learning: theory-inspired improvement on MAML. In *UAI (Proceedings of Machine Learning Research, Vol. 161)*. AUAI Press, 23–33.