

# Sequential Recommendation for Cold-start Users with Meta Transitional Learning

Jianling Wang  
Texas A&M University  
jllwang@tamu.edu

Kaize Ding  
Arizona State University  
kaize.ding@asu.edu

James Caverlee  
Texas A&M University  
caverlee@tamu.edu

## ABSTRACT

A fundamental challenge for sequential recommenders is to capture the sequential patterns of users toward modeling how users transit among items. In many practical scenarios, however, there are a great number of cold-start users with only minimal logged interactions. As a result, existing sequential recommendation models will lose their predictive power due to the difficulties in learning sequential patterns over users with only limited interactions. In this work, we aim to improve sequential recommendation for cold-start users with a novel framework named MetaTL, which learns to model the transition patterns of users through meta-learning. Specifically, the proposed MetaTL: (i) formulates sequential recommendation for cold-start users as a few-shot learning problem; (ii) extracts the dynamic transition patterns among users with a translation-based architecture; and (iii) adopts meta transitional learning to enable fast learning for cold-start users with only limited interactions, leading to accurate inference of sequential interactions.

## CCS CONCEPTS

• **Information systems** → **Recommender systems**;

## KEYWORDS

Recommendation Systems, Cold-start, Meta-learning

## ACM Reference Format:

Jianling Wang, Kaize Ding, and James Caverlee. 2021. Sequential Recommendation for Cold-start Users with Meta Transitional Learning. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*, July 11–15, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3404835.3463089>

## 1 INTRODUCTION

Recommendation systems play an essential role in connecting users with content they enjoy on many large-scale online platforms. Different from traditional recommendation systems [8, 20, 32] that model the general preferences of users, sequential recommendation systems [9, 10, 12, 25, 30, 36] aim to infer the dynamic preferences of users from their behavioral sequences.

And yet, in many real-world scenarios, sequential recommenders may face difficulty in dealing with new users who have only limited interactions with the system, leading to inherently long-tailed interaction data [35]. Ideally, an effective recommendation system should be able to recommend items to new users who have only a few interactions with items. However, most existing sequential recommenders are not designed to handle such cold-start users due to the difficulty of characterizing user preferences with limited data. Since new users may migrate away from the platform if they receive poor recommendations initially, *how to capture the preference of these cold-start users* becomes a critical question for building a satisfactory recommendation service.

Though quite a few cold-start recommendation methods have been proposed, most require side information [14, 15, 37] or knowledge from other domains [1, 11, 17] during training, and commonly treat the user-item interactions in a static way. In contrast, cold-start sequential recommendation targets a setting where no additional auxiliary knowledge can be accessed due to privacy issues, and more importantly, the user-item interactions are sequentially dependent. A user's preferences and tastes may change over time and such dynamics are of great significance in sequential recommendation. Hence, it is necessary to develop a new sequential recommendation framework that can distill short-range item transitional dynamics, and make fast adaptation to those cold-start users with limited user-item interactions.

In this work, we propose a new meta-learning framework called *MetaTL* for tackling the problem of cold-start sequential recommendation. In order to improve the model generalization capability with only a few user-item interactions, we reformulate the task of cold-start sequential recommendation as a few-shot learning problem. Different from existing methods that directly learn on the data-rich users, MetaTL constructs a pool of few-shot user preference transition tasks that mimic the targeted cold-start scenarios, and progressively learns the user preferences in a meta-learning fashion. Moreover, we build the proposed framework on top of a translation-based architecture, which allows the recommendation model to effectively capture the short-range transitional dynamics. This way the meta-learned sequential recommendation model can extract valuable transitional knowledge from those data-rich users and make fast adaptation to cold-start users to provide high-quality recommendations. The contributions of this work can be summarized as follows:

- We explore the challenging problem of sequential recommendation for cold-start users without relying on auxiliary information and propose to formulate it as a few-shot learning problem.
- We develop a novel meta-learning paradigm – *MetaTL* to model the transition patterns of users, which can make fast adaption for cold-start users in inferring their sequential interactions.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8037-9/21/07...\$15.00

<https://doi.org/10.1145/3404835.3463089>

- With extensive experiments on three real-world datasets, we verify the effectiveness of the proposed MetaTL in cold-start sequential recommendation and shows that it can bring in 11.7% and 5.5% improvement compared with the state-of-the-art in sequential recommendation and cold-start user recommendation.

## 2 RELATED WORK

**Sequential Recommendation.** One of the first approaches for sequential recommendation is the use of Markov Chains to model the transitions of users among items [21]. More recently, TransRec [7] embeds items in a “transition space” and learns a translation vector for each user. With the advance in neural networks, many different neural structures including Recurrent Neural Networks [9, 10], Convolutional Neural Networks [25, 36], Transformers [12, 24] and Graph Neural Networks [31, 34], have been adopted to model the dynamic preferences of users over their behavior sequences. While these methods aim to improve the overall performance via representation learning for sequences, they suffer from weak prediction power for cold-start users with short behavior sequences.

**Meta-learning.** This line of research aims to learn a model which can adapt and generalize to new tasks and new environments with a few training samples. To achieve the goal of “learning-to-learn”, there are three types of different approaches. Metric-based methods are based on a similar idea to the nearest neighbors algorithm with a well-designed metric or distance function [28], prototypical networks [3, 23] or Siamese Neural Network [13]. Model-based methods usually perform a rapid parameter update with an internal architecture or are controlled by another meta-learner model [22]. As for the optimization-based approaches, by adjusting the optimization algorithm, the models can be efficiently updated with a few examples [4, 5, 18]. In this work, we explore how to design an effective approach to handle the cold-start sequential recommendation for short sequences.

**Cold-start Recommendation.** Under the complete cold-start setting with no historic interaction available for new users or items, previous works usually learn a transformation between the auxiliary information with the well-trained latent factors [6, 15, 26]. Under the incomplete cold-start setting, Dropoutnet utilizes Dropout layer to simulate the data missing problem [29]. Meanwhile, meta-learning has been applied to train a model tailored for cold-start cases. To solve the user cold-start problem, MetaRec [27] proposes a meta-learning strategy to learn user-specific logistic regression. There are also methods including MetaCF [33], Warm-up [19] and MeLU [14], adopting Model-Agnostic Meta-Learning (MAML) methods to learn a model to achieve fast adaptation for cold-start users. However, none of them are designed with the dynamics of user preferences in mind (as is the case in sequential recommendation).

## 3 THE PROPOSED MODEL

In this section, we introduce the details of the proposed MetaTL. In essence, the design of MetaTL aims to answer the following research questions: **RQ1:** How to enable the model to transfer knowledge from data-rich users to cold-start users? **RQ2:** How do we capture the short-range transition dynamics in user-item interaction sequences? and **RQ3:** How to optimize the meta-learner for making accurate recommendations for cold-start users?

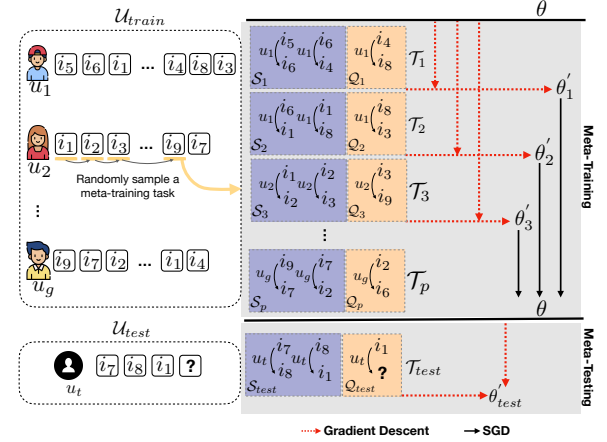


Figure 1: Overview of the proposed MetaTL.

### 3.1 Problem Setup

Let  $\mathcal{I} = \{i_1, i_2, \dots, i_p\}$  and  $\mathcal{U} = \{u_1, u_2, \dots, u_G\}$  represent the item set and user set in the platform respectively. Each item is mapped to a trainable embedding associated with its ID. There is no auxiliary information for users or items. In sequential recommendation, given the sequence of items  $Seq_u = (i_{u,1}, i_{u,2}, \dots, i_{u,n})$  that user  $u$  has interacted with in chronological order, the model aims to infer the next interesting item  $i_{u,n+1}$ . That is to say, we need to predict the preference score for each candidate item based on  $Seq_u$  and thus recommend the top-N items with the highest scores.

In our task, we train the model on  $\mathcal{U}_{train}$ , which contains users with various numbers of logged interactions. Then given  $u$  in a separate test set  $\mathcal{U}_{test}$ ,  $\mathcal{U}_{train} \cap \mathcal{U}_{test} = \emptyset$ , the model can quickly learn user transition patterns according to the  $K$  initial interactions and thus infer the sequential interactions. Note that the size of a user’s initial interactions (i.e.,  $K$ ) is assumed to be a small number (e.g., 2, 3 or 4) considering the cold-start scenario.

### 3.2 Meta Transitional Learning

**Few-shot Sequential Recommendation (RQ1).** Meta-learning aims to learn a model which can adapt to new tasks (i.e., new users) with a few training samples. To enable meta-learning in sequential recommendation for cold-start users, we formulate training a sequential recommender as solving a new few-shot learning problem (i.e., meta-testing task) by training on many sampled similar tasks (i.e., the meta-training tasks). Each task includes a *support* set  $\mathcal{S}$  and a *query* set  $\mathcal{Q}$ , which can be regarded as the “training” set and “testing” set of the task. For example, while constructing a task  $\mathcal{T}_n$ , given user  $u_j$  with initial interactions in sequence (e.g.,  $i_A \xrightarrow{u_j} i_B \xrightarrow{u_j} i_C$ ), we will have the a set of transition pairs  $\{i_A \xrightarrow{u_j} i_B, i_B \xrightarrow{u_j} i_C\}$  as support and predict for the query  $i_C \xrightarrow{u_j} ?$ .

How can we generate a pool of meta-training tasks from data-rich users to mimic the targeted cold-start scenarios? Assume that we are focusing on predicting for cold-start users with  $K$  initial interactions and want to predict for their  $K + 1^{th}$  interactions. To construct a meta-training task, firstly, we will randomly select a user  $u_j$  from  $\mathcal{U}_{train}$  and also randomly sample  $K + 1$  interactions from the user’s logged interactions. With the  $K + 1$  interactions ordered

chronologically  $(i_1, i_2, \dots, i_K, i_{K+1})$ , we will have  $i_1 \xrightarrow{u_j} i_2, i_2 \xrightarrow{u_j} i_3, \dots, i_{K-1} \xrightarrow{u_j} i_K$  in the support set and  $i_K \xrightarrow{u_j} i_{K+1}$  for the query.

**Meta transitional Learner (RQ2).** Our goal is to gain the capability of learning the transition patterns to help predict the next interactions for a new user with only a few initial interactions.

Firstly, for a task  $\mathcal{T}_n$  sampled from user  $u_j$ , we want to retrieve the transitional dynamics with the transition pairs in support set  $\mathcal{S}_n$ . Let  $i_h \xrightarrow{u_j} i_t$  denote a transition pair and  $\mathbf{i}_h$  and  $\mathbf{i}_t$  as the item embedding. Following [2], then we can use a simple MLP network to represent the transitional information with item  $i_h$  and item  $i_t$ ,

$$\mathbf{t}_{h,t} = \sigma(\mathbf{W}(\mathbf{i}_h \parallel \mathbf{i}_t) + \mathbf{b}) \quad (1)$$

in which  $\mathbf{W}$  and  $\mathbf{b}$  is the trainable transform matrix and bias vector correspondingly.  $\sigma(\cdot)$  denotes the Sigmoid activation function.

Since there are multiple transition pairs in a support set  $\mathcal{S}_n$  for the current task  $\mathcal{T}_n$ , we need to aggregate the transitional information from all the pairs to generate the final representation  $\mathbf{tr}_n$ . As a straightforward solution, we take the average of all the transitional information from each transition pair in the support set  $\mathcal{S}_n$ :

$$\mathbf{tr}_n = \frac{1}{|\mathcal{S}_n|} \sum_{(i_h \xrightarrow{u_j} i_t) \in \mathcal{S}_n} \mathbf{t}_{h,t}, \quad (2)$$

in which  $|\mathcal{S}_n|$  denotes the size of  $\mathcal{S}_n$ . After generating  $\mathbf{tr}_n$ , we need a scoring function to evaluate its effectiveness in characterizing the transition pattern of the user. Following the idea in translation-based recommendation [7], a user can be viewed as the translation (or transition) between two consecutive items in the interaction sequence. Thus if  $\mathbf{tr}_n$  is effective in characterizing the transition pattern of a transition pair in  $\mathcal{S}_n$  (i.e.,  $i_h \xrightarrow{u_j} i_t$ ), the translation  $\mathbf{i}_h + \mathbf{tr}_n$  should be close to  $\mathbf{i}_t$ . That is to say, the score  $s(i_h \xrightarrow{u_j} i_t) = \|\mathbf{i}_h + \mathbf{tr}_n - \mathbf{i}_t\|^2$  will have a small value. Then we calculate the marginal loss based on all the transition pairs in support  $\mathcal{S}_n$  as:

$$\mathcal{L}_{\mathcal{S}_n} = \sum_{(i_h \xrightarrow{u_j} i_t) \in \mathcal{S}_n} [\gamma + s(i_h \xrightarrow{u_j} i_t) - s(i_h \xrightarrow{u_j} i'_t)]_+ \quad (3)$$

in which  $[\cdot]_+ \triangleq \max(0, \cdot)$  and  $\gamma$  denotes the margin. Here  $i'_t$  is a negative items without interaction from  $u_j$ .

**Optimization and Fast Adaptation (RQ3).** Next, we explain the procedure to optimize the model so that it can learn the transition pattern for new users with just a few interactions. We denote the meta model as a parameterized function  $f_\theta$  with parameters  $\theta$ . Since we hope that the model can obtain a small value in  $\mathcal{L}_{\mathcal{S}_n}$ , we update the model  $f_\theta$  by minimizing the  $\mathcal{L}_{\mathcal{S}_n}$  with one gradient step:

$$\theta'_n = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{S}_n}(f_\theta), \quad (4)$$

in which  $\alpha$  is the task-learning rate. Note that we can extend this to perform multiple-step gradient updates based on Eq. (4).

After updating  $\theta$  to be  $\theta'_n$  with the support set  $\mathcal{S}_n$ , we can generate the updated  $\mathbf{tr}_n$  with  $f_{\theta'_n}$ . We evaluate its performance on the query set  $\mathcal{Q}_n$  with loss  $\mathcal{L}_{\mathcal{Q}_n}$ , which can be calculated following Eq. (3) with transition pair in  $\mathcal{Q}_n$ . Our goal is to determine the  $\theta$  that can work as a good initialization for each task. That is, it can

	# Users	# Items	Avg. Length of Sequences	Splitting Timestamp
<b>Electronics</b>	22,685	20,712	15.26	1/1/2014
<b>Movie</b>	26,933	18,855	28.97	1/1/2014
<b>Book</b>	90,892	58,250	27.81	1/1/2017

**Table 1: Dataset Statistics.**

minimize the loss on query sets across multiple meta-training tasks,

$$\theta = \min_{\theta} \sum_{\mathcal{T}_n \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{Q}_n}(f_{\theta'_n}), \quad (5)$$

in which  $p(\mathcal{T})$  is the distribution of meta-training tasks and can be obtained by randomly sampling meta-training tasks. To solve this equation, we can perform optimization via stochastic gradient descent (SGD), such that:

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_n \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{Q}_n}(f_{\theta'_n}), \quad (6)$$

where  $\beta$  is the meta step size.

When **testing** on a new user  $u_{test}$ , we will firstly construct the support set  $\mathcal{S}_{test}$  based on the user's initial interactions. With Eq. (4), the model  $f_\theta$  is fine-tuned with all the transition pairs in  $\mathcal{S}_{test}$  and updated to  $f_{\theta'_{test}}$ , which can be used to generate the updated  $\mathbf{tr}_{test}$ . Given the test query  $i_o \xrightarrow{u_{test}} ?$ , the preference score for item  $i_p$  (as the next interaction) is calculated as  $-\|\mathbf{i}_o + \mathbf{tr}_{test} - \mathbf{i}_p\|^2$ .

## 4 EXPERIMENTS

### 4.1 Experimental Setup

**Datasets.** We adopt three public real-world datasets: **Electronics** is adopted from the public Amazon review dataset [16], which includes reviews ranging from May 1996 to July 2014 on Amazon products belonging to the "Electronics" category. **Movie** is similarly drawn from the "Movie" category of the same Amazon review dataset. For both, we treat a user review as an interaction. **Book** is scraped from Goodreads, a book platform for users to tag, rate and review books. We treat all these interactions on items equally.

For all of the datasets, we filter out items with fewer than 10 interactions. We split each dataset with a corresponding cutting timestamp  $T$ , such that we construct  $\mathcal{U}_{train}$  with users who have interactions before  $T$  and construct  $\mathcal{U}_{test}$  with users who start their first interactions after  $T$ . Summary statistics of the datasets can be found in Table 1. When evaluating few-shot sequential recommendation for a choice of  $K$  (i.e., the number of initial interactions), we keep  $K$  interactions as initialization for each user in  $\mathcal{U}_{test}$  and predict for the user's next interactions.

**Baselines.** Aside from a standard matrix factorization method (BPR-MF) [20], we compare the proposed MetaTL model with two categories of widely used recommendation models for Top-N recommendation: (i) *Sequential recommendation baselines* utilize different methods to capture the sequential patterns in the interaction sequences of users. **TransRec** [7] embeds items into a "transition space" and learns a translation vector for each user. **GRU4Rec** [10], **TCN** [36] and **SASRec** [12] rely on Gated Recurrent Units, the simple convolutional generative network, and the self-attention layers to learn sequential user behaviors, respectively. **BERT4Rec**

	Electronics		Movie		Book	
	Hit@1	MRR	Hit@1	MRR	Hit@1	MRR
<b>BPR-MF</b>	0.066	0.123	0.025	0.083	0.043	0.098
<b>TransRec</b>	0.183	0.296	0.208	0.321	0.335	0.454
<b>GRU4Rec</b>	0.185	0.301	0.189	0.309	0.330	0.466
<b>TCN</b>	0.182	0.303	0.186	0.314	0.349	0.489
<b>SASRec</b>	0.193	0.318	0.211	0.345	0.347	0.488
<b>BERT4Rec</b>	0.200	0.323	0.220	0.351	0.369	0.513
<b>MeLU</b>	0.172	0.265	0.168	0.289	0.318	0.423
<b>MetaCF</b>	0.210 <sup>†</sup>	0.330 <sup>†</sup>	0.234 <sup>†</sup>	0.365 <sup>†</sup>	0.398 <sup>†</sup>	0.528 <sup>†</sup>
<b>MetaTL</b>	<b>0.224</b>	<b>0.352</b>	<b>0.258</b>	<b>0.380</b>	<b>0.420</b>	<b>0.555</b>

**Table 2: Comparison of Different Models under  $K = 3$ . The improvement of MetaTL is statistically significant compared with the next-best model with  $p < 0.05$**

[24] adopts the bi-directional transformer to extract the sequential patterns and it is the state-of-the-art for sequential recommendation; (ii) *Cold-start baselines* include methods focusing on providing accurate recommendations for cold-start users with limited information. **MeLU** [14] learns a user preference estimator model based on Model-Agnostic Meta-Learning (MAML), which can be rapidly adapted for cold-start users. We modify MeLU as in [33] to fit for the case without auxiliary information. **MetaCF** [33] learns a collaborative filtering (CF) model which can quickly adapt to new users. We adopt the version on top of NGCF [32] for better performance.

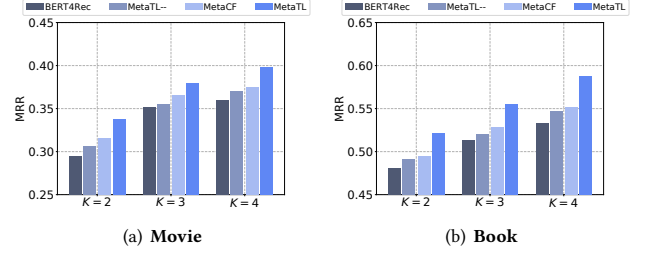
**Evaluation Metrics.** In the experiment, each user only has one ground-truth item for testing. With the predicted scores, as in [12, 33], we rank the ground-truth positive item with 100 randomly sampled negative items. Mean Reciprocal Rank (MRR) indicates the rankings of the ground-truth items. We also evaluate the Hit Rate (Hit) for the top-1 prediction.  $Hit@1 = 1$  if the ground-truth item is ranked top-1, otherwise  $Hit@1 = 0$ . Also note that  $Hit@1$  is equal to the recall or NDCG for top-1 prediction.

## 4.2 Overall Performance

We compare the performance of MetaTL and the baseline models under  $K = 3$  and report the results in Table 2. The best performing method in each column is boldfaced, and the second best method is marked with <sup>†</sup>. MetaTL achieves the best performance under different evaluation metrics in all of the datasets, illustrating the effectiveness of MetaTL in providing accurate sequential recommendation for cold-start users with limited interactions.

Starting from the simplest collaborative filtering (BPR-MF), we find that it performs weakly since it ignores the dynamic patterns in the user interaction sequences and fails to learn effective embeddings for cold-start users. TransRec also becomes ineffective in learning translation embeddings for cold-start users since there is insufficient data to update the embeddings. Then we compare the performance of various neural models for sequential recommendation. We can see that GRU4Rec and TCN perform the worst. SASRec and BERT4Rec (utilizing transformers to extract the sequential patterns) work better since they are able to aggregate the items with attention scores and thus obtain more informative representations for users with limited interactions.

MeLU and MetaCF are both meta-learning based methods for providing cold-start recommendations. Since MeLU requires side



**Figure 2: Comparison for different model variants w.r.t. K.**

information for both users and items, we treat their historic interactions as the side information as in [33]. Alas, MeLU is unable to obtain satisfying results since it designed for scenarios with abundant auxiliary user/item information which is absent in this case. Meanwhile we find that MetaCF can achieve the second-best performance for sequential recommendation, illustrating the importance of fast adaption in cold-start scenario. It still falls behind the proposed MetaTL since it is unable to learn the transition patterns for cold-start sequential recommendation.

## 4.3 Ablation Analysis

To further evaluate the effectiveness of the proposed MetaTL model, we compare it with its variants and some of the baselines under different K values (i.e., the number of initial interactions) in Figure 2. Note that *MetaTL-* is the simplified version of MetaTL by removing the MAML optimization step. BERT4Rec is the state-of-the-art sequential recommendation method and MetaCF is the strongest cold-start baseline from our original experiments (and illustrates the performance of CF with meta-learning).

In both datasets, BERT4Rec loses the prediction power on sequences consisting of only a few items and thus performs weakly in the cold-start sequential recommendation task. Even without the fast adaptation learning module, MetaTL- can outperform BERT4Rec since it is carefully designed to learn the transition patterns on extremely short sequences. However, it still falls behind MetaCF, which demonstrates the necessity of training a model with fast adaptation in cold-start scenarios. With the well-designed optimization steps and meta transitional learner, the proposed MetaTL can further improve MetaTL- and outperform both the state-of-the-art methods in sequential recommendation and cold-start user recommendation with different numbers of initial interactions.

## 5 CONCLUSION

We propose a novel framework MetaTL to improve sequential recommendation for cold-start users. To enable the fast adaptation to cold-start users, we reformulate our task as a few-shot learning problem and adopt meta-learning for solving the problem. Powered by a translation-based architecture, the model is able to capture the transition patterns from the transition pairs. Meanwhile, given a pool of few-shot user preference transition tasks that mimic the targeted cold-start scenarios, MetaTL learns a model which can be adapted to new users with just a few interactions in a meta-learning fashion. With experiments on three real-world datasets, the proposed MetaTL can bring in significant improvement compared with the state-of-the-art methods.

## REFERENCES

- [1] Ye Bi, Liqiang Song, Mengqiu Yao, Zhenyu Wu, Jianming Wang, and Jing Xiao. 2020. DCDIR: A deep cross-domain recommendation system for cold start users in insurance domain. In *SIGIR*.
- [2] Mingyang Chen, Wen Zhang, Wei Zhang, Qiang Chen, and Huajun Chen. 2019. Meta relational learning for few-shot link prediction in knowledge graphs. In *EMNLP*.
- [3] Kaize Ding, Jianling Wang, Jundong Li, Kai Shu, Chenghao Liu, and Huan Liu. 2020. Graph prototypical networks for few-shot learning on attributed networks. In *CIKM*.
- [4] Kaize Ding, Qinghai Zhou, Hanghang Tong, and Huan Liu. 2021. Few-shot Network Anomaly Detection via Cross-network Meta-learning. In *The Web Conference*.
- [5] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*.
- [6] Zeno Gantner, Lucas Drumond, Christoph Freudenthaler, Steffen Rendle, and Lars Schmidt-Thieme. 2010. Learning attribute-to-feature mappings for cold-start recommendations. In *ICDM*.
- [7] Ruining He, Wang-Cheng Kang, and Julian McAuley. 2017. Translation-based recommendation. In *RecSys*.
- [8] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *WWW*.
- [9] Balázs Hidasi and Alexandros Karatzoglou. 2018. Recurrent neural networks with top-k gains for session-based recommendations. In *CIKM*.
- [10] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939* (2015).
- [11] SeongKu Kang, Junyoung Hwang, Dongha Lee, and Hwanjo Yu. 2019. Semi-supervised learning for cross-domain recommendation to cold-start users. In *CIKM*.
- [12] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *ICDM*.
- [13] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. 2015. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*.
- [14] Hoyeop Lee, Jinbae Im, Seongwon Jang, Hyunsouk Cho, and Sehee Chung. 2019. MeLU: Meta-Learned User Preference Estimator for Cold-Start Recommendation. In *KDD*.
- [15] Jingjing Li, Mengmeng Jing, Ke Lu, Lei Zhu, Yang Yang, and Zi Huang. 2019. From zero-shot learning to cold-start recommendation. In *AAAI*.
- [16] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *SIGIR*.
- [17] Nima Mirbakhsh and Charles X Ling. 2015. Improving top-n recommendation for cold-start users via cross-domain information. In *TKDD*.
- [18] Alex Nichol, Joshua Achiam, and John Schulman. 2018. On first-order meta-learning algorithms. In *arXiv preprint arXiv:1803.02999*.
- [19] 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*.
- [20] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *UAI*.
- [21] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing personalized markov chains for next-basket recommendation. In *WWW*.
- [22] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. 2016. Meta-learning with memory-augmented neural networks. In *ICML*.
- [23] Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. In *NeurIPS*.
- [24] 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. *CIKM*.
- [25] Jiayi Tang and Ke Wang. 2018. Personalized top-n sequential recommendation via convolutional sequence embedding. In *WSDM*.
- [26] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. In *NeurIPS*.
- [27] Manasi Vartak, Arvind Thiagarajan, Conrado Miranda, Jeshua Bratman, and Hugo Larochelle. 2017. A meta-learning perspective on cold-start recommendations for items. In *NeurIPS*.
- [28] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. 2016. Matching networks for one shot learning. In *NeurIPS*.
- [29] Maksims Volkovs, Guangwei Yu, and Tomi Poutanen. 2017. Dropoutnet: Addressing cold start in recommender systems. In *NeurIPS*.
- [30] Jianling Wang, Kaize Ding, Liangjie Hong, Huan Liu, and James Caverlee. 2020. Next-item recommendation with sequential hypergraphs. In *SIGIR*.
- [31] Jianling Wang, Kaize Ding, Ziwei Zhu, and James Caverlee. 2021. Session-based Recommendation with Hypergraph Attention Networks. In *SDM*.
- [32] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *SIGIR*.
- [33] 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. *ICDM*.
- [34] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based recommendation with graph neural networks. In *AAAI*.
- [35] Jianwen Yin, Chenghao Liu, Weiqing Wang, Jianling Sun, and Steven CH Hoi. 2020. Learning Transferrable Parameters for Long-tailed Sequential User Behavior Modeling. In *KDD*.
- [36] Fajie Yuan, Alexandros Karatzoglou, Ioannis Arapakis, Joemon M Jose, and Xiangnan He. 2019. A Simple Convolutional Generative Network for Next Item Recommendation. In *WSDM*.
- [37] Ziwei Zhu, Shahin Sefati, Parsa Saadatpanah, and James Caverlee. 2020. Recommendation for New Users and New Items via Randomized Training and Mixture-of-Experts Transformation. In *SIGIR*.