



## UvA-DARE (Digital Academic Repository)

### Cross-Market Product Recommendation

Bonab, H.; Aliannejadi, M.; Vardasbi, A.; Kanoulas, E.; Allan, J.

**DOI**

[10.1145/3459637.3482493](https://doi.org/10.1145/3459637.3482493)

**Publication date**

2021

**Document Version**

Final published version

**Published in**

CIKM '21

**License**

Article 25fa Dutch Copyright Act

[Link to publication](#)

**Citation for published version (APA):**

Bonab, H., Aliannejadi, M., Vardasbi, A., Kanoulas, E., & Allan, J. (2021). Cross-Market Product Recommendation. In *CIKM '21: proceedings of the 30th ACM International Conference on Information & Knowledge Management : November 1-5, 2021, virtual event, Australia* (pp. 110-119). The Association for Computing Machinery. <https://doi.org/10.1145/3459637.3482493>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Cross-Market Product Recommendation

Hamed Bonab  
University of Massachusetts Amherst  
bonab@cs.umass.edu

Mohammad Aliannejadi  
University of Amsterdam  
m.aliannejadi@uva.nl

Ali Vardasbi  
University of Amsterdam  
a.vardasbi@uva.nl

Evangelos Kanoulas  
University of Amsterdam  
e.kanoulas@uva.nl

James Allan  
University of Massachusetts Amherst  
allan@cs.umass.edu

## ABSTRACT

We study the problem of recommending relevant products to users in relatively resource-scarce markets by leveraging data from similar, richer in resource auxiliary markets. We hypothesize that data from one market can be used to improve performance in another. Only a few studies have been conducted in this area, partly due to the lack of publicly available experimental data. To this end, we collect and release XMarket, a large dataset covering 18 local markets on 16 different product categories, featuring 52.5 million user-item interactions.

We introduce and formalize the problem of cross-market product recommendation, i.e., *market adaptation*. We explore different market-adaptation techniques inspired by state-of-the-art domain-adaptation and meta-learning approaches and propose a novel neural approach for market adaptation, named *FOREC*. Our model follows a three-step procedure – pre-training, forking, and fine-tuning – in order to fully utilize the data from an auxiliary market as well as the target market. We conduct extensive experiments studying the impact of market adaptation on different pairs of markets. Our proposed approach demonstrates robust effectiveness, consistently improving the performance on target markets compared to competitive baselines selected for our analysis. In particular, FOREC improves on average 24% and up to 50% in terms of nDCG@10, compared to the NMF baseline. Our analysis and experiments suggest specific future directions in this research area. We release our data and code<sup>1</sup> for academic purposes.

## CCS CONCEPTS

• **Information systems** → **Recommender systems**: *Collaborative search*; • **Computing methodologies** → **Transfer learning**.

## KEYWORDS

Product Recommendation, Meta-Learning, Domain Adaptation, Market Adaptation, Cross-Market Recommendation

<sup>1</sup>Data and code: <https://xmrec.github.io>

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).

CIKM '21, November 1–5, 2021, Virtual Event, QLD, Australia

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8446-9/21/11...\$15.00

<https://doi.org/10.1145/3459637.3482493>

## ACM Reference Format:

Hamed Bonab, Mohammad Aliannejadi, Ali Vardasbi, Evangelos Kanoulas, and James Allan. 2021. Cross-Market Product Recommendation. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, QLD, Australia*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3459637.3482493>

## 1 INTRODUCTION

Nowadays online shopping in many countries is a part of people's daily lives. While online shopping brings several benefits and comfort to both users and vendors [23], it comes at the risk of overwhelming users with virtually unlimited options to choose from. Recommender systems are key in dealing with information overload, helping not only users finding interesting items, but also vendors finding the right customer for their products. E-commerce companies often operate across markets; for instance Amazon<sup>2</sup> has expanded their operations and sales to 18 markets around the globe.<sup>3</sup> This brings both opportunities and challenges.

While it is typical that several local e-commerce companies operate in every country, the presence of an international e-commerce company, like Amazon, eBay, and Etsy can benefit users even more if these companies can utilize the experience and data gathered across several markets. Using cross-market data however comes at a risk of assuming one-solution-fits-all and applying the same algorithms that are developed for and trained on large and data-rich markets, such as the U.S. [22], to small and data-scarce markets. The key challenge is that data, such as user interaction data with products (clicks, purchases, reviews), convey certain biases of the individual markets [7]. Algorithms that are optimized for a certain market learn various biases and distributions of the data [47]. Therefore, the algorithm trained on a source market, are not necessarily effective in a different target market [13], since utilizing the vast amount of data from a large market to improve the performance on low-resource markets comes at the risk of importing the wrong data distributions. For example, assume iPhone is the most popular smartphone in the U.S. (source market), while Samsung is the most popular smartphone brand in Germany (target market). Importing user preference from the U.S. market would yield to recommending iPhone in Germany more often than Samsung, which clearly is a wrong choice. Hence, even though there is a myriad of information to learn from a source market, careful adaptation of data is required.

The significance of Cross-Market Recommendation (CMR) has been pointed out in the literature [46]. However, small progress has been made in this area, mainly due to a lack of experimental

<sup>2</sup><https://www.amazon.com>

<sup>3</sup><https://sell.amazon.com/global-selling.html>

data. To this end, we construct a large-scale real-life product recommendation dataset, referred to as XMarket, drawn from 18 markets in 11 languages. To develop this dataset we crawled Amazon marketplaces around the globe, locating and including in the dataset the same products within different markets. Moreover, we analyze certain statistical properties and trends amongst multiple markets and product categories where we highlight the existence of crucial differences across different markets.

In this paper, we focus on the user-item interaction data through ratings and study the problem of recommending relevant products to users in relatively resource-scarce markets by leveraging data from similar, richer in resource auxiliary markets. Our hypothesis is that data from one market can be used to improve performance in another. For this purpose, we first introduce and formalize the problem of cross-market product recommendation, i.e., *market adaptation*. In order to solve the problem of CMR, we explore market-adaptation baselines inspired by domain-adaptation and meta-learning approaches. Then, we propose a novel neural approach, named *FOREC*, consisted of a three-step procedure – pre-training, forking, and fine-tuning – in order to fully utilize the data from an auxiliary market to boost the product recommendation performance in the target market. More specifically, FOREC learns a general recommendation system based on two markets (i.e., source and target) and employs a forking procedure by adding a *market specific* sub-network to the head of the model and freezing the bottom part in order to adapt the general internal representations to the target market. We conduct extensive experiments studying the impact of market adaptation on different pairs of markets. FOREC demonstrates robust effectiveness, consistently improving the performance compared to competitive baselines on 7 target markets we selected for our study. In particular, FOREC improves on average 24% and up to 50% in terms of nDCG@10, compared to the NMF baseline. Our analysis and experiments provide many insights on the CMR and suggest specific future directions in this research area. In summary, the main contributions of this paper constitute:

- Collecting a real-world pragmatic large-scale cross-market and cross-lingual product review dataset.
- Performing analysis of cross-market behavioral biases pivoting on the fact that our dataset includes the same items shared across different markets. In particular, we study how differently users in different markets interact with the same products.
- Proposing a novel neural architecture for market-adaptation and demonstrating the effectiveness of the model through extensive experiments. We adapt various Cross-Domain Recommendation (CDR) and meta-learning approaches for Cross-Market Recommendation (CMR).
- Analyzing the performance of the model, considering various setups and conditions to provide further insights.

## 2 RELATED WORK

This study is related to CMR and CDR, as well as meta-learning approaches. In this section, we briefly review the research done in these domains.

**Cross-domain and cross-market recommendation.** Research on CMR and CDR aims at improving the system’s effectiveness based on the external data that is available from other markets or

item categories. While having the same goal, the two tasks differ in various aspects, each bearing their own challenges. Particularly, in CDR the general assumption is that the model learns from interactions of overlapping users<sup>4</sup> in different domains (e.g. categories in product search) with the aim of improving the recommendation on the target domain items, using help from the source domain items. For CMR the situation is reversed: interactions of different set of users in the source market are leveraged to boost the recommendation for users in the target market. Here, we assume that the items are shared among different markets.

Im and Hars [22] conduct an experiment in two product domains, aiming to answer the question “does a one-size recommendation fit all?” where they observed that the performance of Collaborative Filtering (CF) is highly affected by the information-seeking mode of the users. Depending on the product domain, users adopt different strategies and therefore the system would not fit to all domains. Lu et al. [37] later argue that transferring all the knowledge from source domain into the target domain may harm the recommender due to some inconsistencies and propose a criterion for selecting the consistent part of knowledge to be transferred to the target domain. Elkahky et al. [12] apply domain adaptation using user behavior-based features for learning latent space. Rafailidis and Crestani [43] propose a collaborative ranking model with a weighting strategy that controls the influence of user preferences from auxiliary domains. Zhao et al. [55] use reviews to transfer user preference at aspect-level as a cross-domain recommendation framework. Krishnan et al. [31] propose using contextual invariances across domains to leverage data from a dense domain to improve learned representations in other sparse domains. Different from these works, in this paper we investigate the existence of different behavioral biases across different markets where unlike CDR, the items are the same across different markets but the users are different. Further research has aimed at mitigating these biases and transferring knowledge from one domain to another [20, 33]. Unsupervised domain adaptation [16] has also inspired various cross-domain recommender systems in recent years [26, 34, 50, 53]. CDR has been used specifically for mitigating the cold-start problem in a number of studies [15, 24, 25, 27, 40, 52, 55].

While there exists much work on domain adaptation, *market-adaptation* is relatively unstudied. CMR has attracted attention in music recommendation [13, 46] where Ferwerda et al. [13] analyze and study music diversity across countries and propose to use country-based diversity measurements for system evaluation. Roitero et al. [46] studied user behavior in 21 different markets on Spotify and highlight the need for market-specific algorithms, as opposed to a global algorithm. We take one step further in this direction by expanding our study to various item categories in e-commerce where users purchase items (rather than having a monthly subscription) and express their opinion and experience with item in the form of ratings and reviews.

**Meta-learning.** The goal of meta-learning is to train a model on multiple tasks, such that it can rapidly adapt to a new task after seeing a small number of new training samples [49]. In the context of recommendation systems, meta-learning has been used for several problems, including but not limited to recommender algorithm

<sup>4</sup>with a few exceptions like [31, 42] that assume non-overlapping users

**Table 1: General statistics of XMarket.**

# markets	18
# languages	11
# categories	16
# items	3,811,438
# users	9,562,260
# reviews	52,480,184
# reviewed items	1,000,829
# unique items	294,739

selection [10, 11, 38], cold-start problem [32, 48], and retraining the model [54]. For meta-learning of deep neural networks, a general and powerful technique, Model-Agnostic Meta-Learning (MAML), has been proposed that can be directly applied to any learning problem and model [14]. MAML framework is widely used in recommendation literature. For example, Lu et al. [36] use it on heterogeneous information networks to address the cold-start problem. Others used MAML to train a recommender which performs reasonably good enough both for cold- and warm-start users [5, 32].

Enabling the system to use beyond ratings across markets (such as text and image) is also interesting. Including textual data require cross-lingual information retrieval techniques [1, 6, 21]. In future, we plan to expand our study on these directions.

### 3 DATA COLLECTION & ANALYSIS

In this section, we first describe in detail the data collection process and provide statistics for the collected data. Then, we analyze the data and highlight important characteristics and similarities across different markets.

#### 3.1 Data Collection

We describe our new dataset, called XMarket, and provide details on how we generated it. We constructed the XMarket item and review collection on top of a large-scale publicly-available Amazon dataset [18, 39]<sup>5</sup>. The Amazon Product data [39] includes millions of item reviews collected from the Amazon U.S. marketplace in various categories. The dataset was collected in 2014 and later updated in 2018 [41]. We used this dataset as a seed to initiate our crawl. We located the same items that appear on the U.S. market in other markets, by matching the items' unique identifiers (aka. ASIN's) on all available Amazon markets. In particular we have crawled data from the following markets: Saudi Arabia (**sa**), Singapore (**sg**), Australia (**au**), United Arab Emirates (**ae**), Turkey (**tr**), Japan (**jp**), India (**in**), Spain (**es**), U.S. (**us**), China (**cn**), Germany (**de**), Netherlands (**nl**), France (**fr**), Brazil (**br**), Canada (**ca**), Mexico (**mx**), Italy (**it**), United Kingdom (**uk**). Our main criterion for including an item in the collection process was its popularity on the **us** market. Our decision was motivated by the idea of having a high-resource market (i.e., **us**) that would provide a wealth of data to other markets. Therefore, we discarded all items with less than 20 reviews in the past two years, as we did not consider them rich enough to be useful in other markets. In our preliminary analysis, we noticed that in most cases, if an item exists in another market its ASIN is the same. Therefore, we fed our crawler with the ASIN's that we collected

<sup>5</sup><https://jmcauley.ucsd.edu/data/amazon/>

**Table 2: Statistics of different markets for Electronics.**

	us	de	jp	in	fr	ca	mx	uk
<b>U</b>	2.7m	345k	117k	208k	230k	422k	100k	545k
<b>I</b>	35k	8k	4k	7k	6k	19k	9k	10k
<b>R</b>	4.1m	0.5m	174k	257k	392k	721k	169k	865k

from the U.S. market. In doing so, we collected cross-market meta-data information for over one million items and collected about 52 million multilingual reviews.

Among the existing multi-lingual review datasets, we find the Multilingual Amazon Reviews Corpus (MARC) [29] the most similar to XMarket. MARC consists of multi-lingual reviews extracted from different Amazon marketplaces, however, the scale of the dataset is much smaller. In particular, they do not cover all Amazon marketplaces and categories, whereas XMarket covers a wide range of categories in all 18 Amazon marketplaces. Moreover, XMarket includes rich item and review metadata (e.g., reviewer ID, item description, and related items) that can be utilized to pursue various research directions. We also found another similar dataset named as Amazon Customer Reviews Dataset<sup>6</sup> providing a collection of reviews from five marketplaces dated from 1995 to 2015. We notice that a vast majority of the provided data is only from United States and the provided meta-data is limited to only product title and reviews whereas our data covers more number of marketplaces with a full meta-data information (including product text, images, also bought, similar items). In addition, our product reviews are more recent. To the best of our knowledge, no other cross-market multi-lingual recommendation dataset with such a wide coverage of markets and categories exists in the community.

#### 3.2 Data Statistics

Table 1 summarizes some of the main characteristics of XMarket. It provides a cross-lingual e-commerce dataset of 16 shopping categories in 11 languages. We crawled data for ~300K unique items across all markets, which resulted in ~4M cross-market items. Also, Fig. 1a shows the distribution of items in each market. We see that the Canadian (**ca**) and Mexican (**mx**) markets have the most items in common with the U.S. market (**us**), which is expected due to the long-lasting presence of Amazon in these countries and their vicinity to **us**. Due to space considerations and similarity in results, in the remainder of the paper we analyze and discuss a subset of markets and categories. Our experiments and model evaluations are based on the Electronics category with statistics presented in Table 2. We see in the table the main characteristics of the studied marketplaces in terms of recommendation data, such as number of items, users, and ratings. We observe a relatively high number of users as well as items in the U.S. market, making it the most sparse market in our dataset.

#### 3.3 Cross-Market Analysis

Having in mind that the markets share the same set of products, we analyze how people in different regions interact with these items to uncover behavioral characteristics (similarities and differences) and biases.

<sup>6</sup><https://s3.amazonaws.com/amazon-reviews-pds/readme.html>

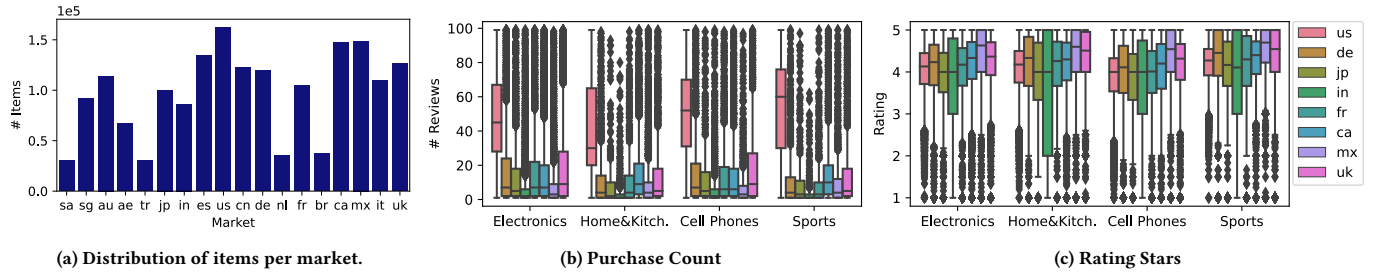


Figure 1: Distribution of items, purchase count, and rating stars, per category and per market.

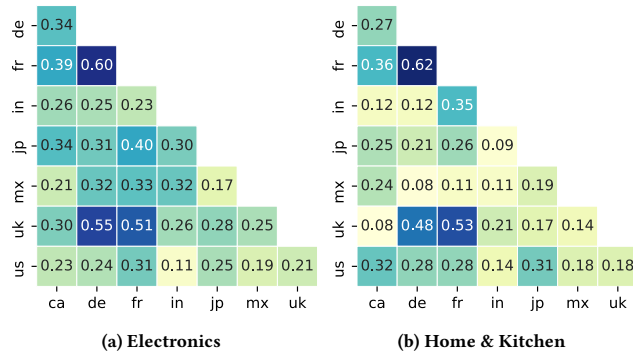


Figure 2: Market similarity based on the cosine distance.

**Distribution of ratings.** Several reasons may influence users to purchase a product such as users’ financial status, culture, and companies’ marketing strategies. Therefore, we study the difference in product rating (as a signal of product purchase) among different markets. We plot the distribution of product “purchase” in Fig. 1b. The dominance of **us** is obvious in this figure, with the highest median in all categories. We see that different markets exhibit different distributions across categories. In general we see that Electronics is most popular category among different markets. We also observe that the Home & Kitchen category shows a different trend compared to other categories, perhaps because such items are more regionally dependent.

**Distribution of rating stars.** We are interested in finding out if the same items are rated differently in each market. Also, if the differences happen across categories. We find significantly different behavior in giving rating stars among markets and categories, as shown in Figure 1c. We see a greater tendency of giving higher rating to items in **mx** market, whereas for **in** market, we see an opposite behavior. Interestingly, we observe a relatively similar trend in all categories where for example the median rating in **de** market is always higher than that of **us** and **fr** market, but slightly lower than the **uk** market. This clearly shows a general bias in user rating behavior, which should be taken into account when developing algorithms such as rating prediction.

**Market similarity.** Based on the observations that we had from Figure 1, we estimate the users’ purchase similarity between markets. For a pair of markets in the same category, we build item purchase count vectors and compute cosine similarity between the two vectors. In Figure 2 we plot the similarities of all market pairs

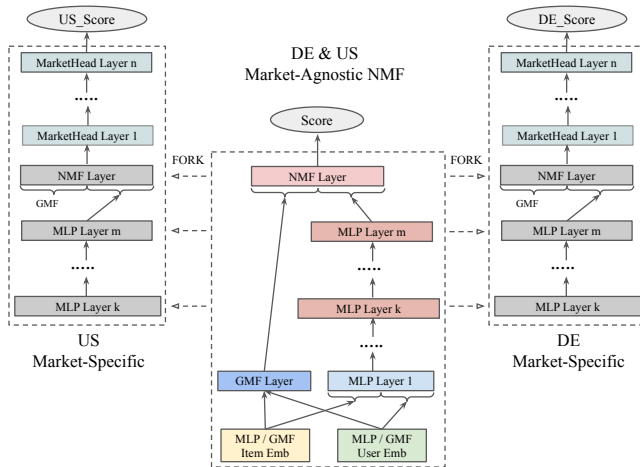
for two categories. Interestingly, we observe the highest rate of similarity between **de**, **fr**, and **uk**, highlighting the similarities in European markets. On the other hand, we see that the American countries do not share much in common. As we see **us**, **ca**, **mx** exhibit low similarity, which is surprising. Perhaps this is due to strong existence of local vendors in this category. In particular, we see the lowest similarity between **de** and **mx**.

**Remarks.** Overall, from our analyses it is evident that users in each market exhibit different behavior. These differences could be due to various reasons such as cultural biases or different marketing strategies adopted by companies. Another reason could simply be the popularity of Amazon as an e-commerce marketplace in different countries and how long it has been doing business in each country. For instance, we saw a very high similarity between **us**, **ca**, and **mx** in terms of common items that exist in the three markets, however, when it comes to product purchase data, we see very little similarity. This indicates that even though Amazon has a big influence on the e-commerce market in these countries, users act differently. With the existence of obvious differences and similarities, but at the same time having a mix of data-rich and scarce markets, learning from auxiliary markets is not trivial and requires careful development of market adaptation techniques.

## 4 PROBLEM STATEMENT

Assume we are given a set of parallel markets as  $M = \{\mathbb{M}_0, \dots, \mathbb{M}_t\}$ . Let  $\mathbb{M}_0$  be the base market with the set of items  $\mathcal{I}_0 = \{I_1, \dots, I_n\}$ . For the base market, one could assume the market with long-lasting existence offering the super-set of items and rich user-item interaction data. For example, with our XMarket settings, the **us** market can be thought of the base market and others such as **de** or **in** are considered the parallel target markets. We assume that  $\mathcal{I}_l \subset \mathcal{I}_0$  for  $1 \leq l \leq t$ . Depending on the provided parallel markets,  $M$ , the base market could change or there might be no base market. With any of these settings, a union set of items in all the parallel markets could be defined as  $\mathcal{I}_0$ , satisfying our assumption.

For a given target market,  $\mathbb{M}_l$ , let the set of market’s users as  $\mathcal{U}_l = \{U_l^1, \dots, U_l^z\}$ . Generally, a user can interact with different markets, but for simplicity, we assume that the set of users in each market are mutually disjoint with any other parallel market. The problem of market adaption is to use any of the parallel markets provided,  $\mathbb{M}_a \in M - \{\mathbb{M}_l\}$  as an auxiliary market to improve the quality of items recommended to users of the target market, i.e.  $\mathcal{U}_l$ . It is straightforward to use more than single auxiliary market. However,



**Figure 3: The general schema of our FOREC recommendation system. For a pair of markets, the middle part shows the market-agnostic model that we pre-train, and then fork and fine-tune for each market shown in the left and right.**

we focus on single auxiliary market and leave the other variations as the future work. For our experiments, we either augment with only **us** market or any of the parallel markets and report the results. Automatically selecting the most suitable parallel auxiliary market is another interesting problem that is out of the scope of this study and we plan to explore on that direction as our future work.

## 5 FOREC: A CMR SYSTEM

Here, we explain our proposed CMR system, named *FOREC*. A general schema of our model is presented in Fig. 3. We show an example model architecture for a pair of markets, **de** and **us**. The training phase for the FOREC system includes three ordered steps (i.e., pre-training, forking, and fine-tuning) that we explain in the following. Note that FOREC is capable of working with any desired number of target markets. However, for simplicity, we only experiment with pairs of markets for our experimental evaluation.

### 5.1 Pre-Training: Market-Agnostic NMF

In this step, we aim to train a recommendation model that is market-agnostic, in the sense that all the model parameters  $\theta$  are shared across markets and easily adaptable to every target market. This provides a generalized recommendation performance and a set of internal latent representations that are suitable for each individual market. Having such internal representations maximize the reusability of parameters translating into minimal effort on target market adaptation. To this end, we exploit the Model-Agnostic Meta-Learning (MAML) framework [14] from the few-shot learning literature.

The general neural architecture we use for our pre-training step is presented in middle part of Fig. 3. This architecture is first introduced by He et al. [19] and widely used in the literature. Here, we summarize the Neural Matrix Factorization (NMF) deep network before explaining our learning paradigm across markets. NMF model fuses two sub-networks namely Generalized Matrix Factorization (GMF) and Multi-Layer Perceptron (MLP). Both GMF and MLP sub-networks are trained with the data individually and then fused

---

### Algorithm 1: FOREC Training

---

**Require:** Market set:  $M = \{\mathbb{M}_0, \mathbb{M}_1, \dots, \mathbb{M}_t\}$   
**Require:** Step size parameters:  $\alpha, \beta$ ; Number of shots:  $K$

- 1 Initialize NMF model parameters  $\theta$
- 2 **while not done do**
- 3     **for all**  $\mathbb{M}_i \in M$  **do**
- 4         adapt\_batch = Sample  $K$  interactions from  $\mathbb{M}_i$
- 5         Evaluate:  $\nabla_{\theta} \mathcal{L}_{\mathbb{M}_i} \text{NMF}(\theta)$  using  $K$  adapt interactions
- 6         Compute:  $\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathbb{M}_i} \text{NMF}(\theta)$
- 7         eval\_batch = Sample another  $K$  interactions from  $\mathbb{M}_i$
- 8     **end**
- 9      $\theta = \theta - \beta \nabla_{\theta} \sum_{\mathbb{M}_i \in M} \mathcal{L}_{\mathbb{M}_i} \text{NMF}(\theta'_i)$  using eval batches
- 10 **end**
- 11 **for all**  $\mathbb{M}_i \in M$  **do**
- 12     NMF <sub>$i$</sub> : Fork a  $\mathbb{M}_i$  specific NMF model
- 13     Fine-tune NMF <sub>$i$</sub>  using only the training data of  $\mathbb{M}_i$
- 14 **end**

---

using the NMF model architecture. For each user and item, a one-hot vector is constructed and fed to the user and item embedding layers of GMF and MLP networks, respectively. Note that GMF and MLP sub-networks keep their own embeddings space (i.e., no embedding sharing). Let the user and item latent vectors be  $p_u$  and  $q_i$ , respectively. The GMF network simply perform an element-wise product between the vectors and uses a linear single-layer feed-forward network to calculate the prediction, i.e.  $h^T(p_u \odot q_i)$ . Here,  $h$  denotes the weights of the output layer. For individual training of GMF, the prediction  $\hat{y}_{\text{GMF}}$  is calculated using an activation function over the output layer's output. This implements a generalized matrix factorization network.

For the MLP sub-network, the user and item vectors are concatenated and fed to a deep feed-forward network for learning the interactions between user and items, i.e.  $z_1 = p_u \oplus q_i$ . Every layer of the deep network, takes the output of the previous layer  $z_k$  and calculates the output, i.e.  $W_k^T z_{k-1} + b_k$ .  $W$  and  $b$  denote the weight matrix and bias vector of each layer. Here, we use ReLU as activation function and calculate  $\hat{y}_{\text{MLP}}$  for individual training. The NMF model simply initialize the network with individual pre-trained parameters and for the output layer it concatenates the output layer of GMF and MLP with a hyper-parameter determining the trade-off. For the parameter initialization of our pre-training for market-agnostic NMF model, i.e.  $\theta$  initialization, we simply concatenate the data from all target markets (a pair of markets in our study) and train the model. For the loss function,  $\mathcal{L}$ , we use the binary cross entropy between the target and the model output.

Algorithm 1 lines 2-10 present our market-agnostic NMF pre-training. The general framework for meta-learning considers a probability distribution over tasks. Given the highly imbalanced training data that each market offers and our final goal to learn a set of generalized network parameters that works for each individual market, we consider equal task sampling across markets. To this end, we simply iterate over each market instead of sampling randomly (line 3). During the meta-training with  $K$ -shot setting, for each market the model is trained such that it can adapt itself with only  $K$  samples from the target market. This pushes the model parameters

such that they broadly become applicable to each individual market. For this purpose, we sample two  $K$  sized batches of the user-item interaction and perform adaptation and evaluation (lines 4 and 7).

Considering the NMF model with parameters  $\theta$ , with the adaptation step on market  $\mathbb{M}_i$  the model parameters become  $\theta'_i$ . With a single gradient update,  $\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathbb{M}_i} \text{NMF}(\theta)$  in which  $\alpha$  is the NMF model's learning rate (line 5-6). The meta-learning optimizes the NMF parameters across markets with the following meta-objective.

$$\min_{\theta} \sum_{\mathbb{M}_i \in \mathcal{M}} \mathcal{L}_{\mathbb{M}_i} \text{NMF}(\theta'_i)$$

The meta-optimization across markets are calculated using the  $K$  evaluation user-item interactions on each market with a meta-learning step size,  $\beta$ . This updates the original model parameters such that few gradient steps can tune the parameters to a specific target market (line 9).

## 5.2 Forking: Market-Specific Model Construction

After obtaining the general internal representations using our market-agnostic pre-training step, we conduct a series of operations for preparing a model that is market-specific for the desired target market. We call this step "forking" mainly due to the sharing of bottom parts and initialization of the middle parts of the market-specific model with the pre-trained model. The general schema of the forking operation is shown in Fig. 3 for our FOREC algorithm. As it can be seen, from the market-agnostic NMF model trained on **us** and **de**, we fork market-specific NMF models for each of the models, shown on the right and left sides of the figure.

Assuming the MLP part of the NMF model containing  $m$  layers, and one additional layer for the NMF model, our objective with forking is to maximize the reusability of the general parameters. To this end, Raghu et al. [44] studies the similarity between an adapted model's layers and the general model and suggests that the main body of the network barely changes and all the adaptation happens in the head layers of the network. Inspired by this finding, we freeze layers up to layer  $k$  of the MLP network ( $1 \leq k \leq m$ ), the only layer of the GMF network, as well as the user and item embeddings learned with each of the sub-networks. Given that freezing some part of the network limits the capacity of the network for learning market-specific parameters, we add  $n$  new market-specific layers on top of the original tower-style feed-forward network right after the NMF layer to increase the network's capacity. We call these new layers as MarketHead layers. We believe that our forking operation provides a network for balancing between the general market and target market-specific parameters after the final fine-tuning. The  $k$  and  $n$  values are experimentally explored on a few pairs and fixed for every market-specific forking in our experiments. The forking operation for each target market is shown in line 12 of Algorithm 1. Further experimental details are given in Section 6.1.

## 5.3 Fine-Tuning: Final Training on the Target Market

Over the forking step, we obtain a new market-specific NMF network for the desired target market that the bottom part of the network is frozen for any update for providing generalized internal

features, middle part initialized with the general market that could easily adapted, and the final part that randomly initialized and needs further training. One advantage of such a design is that it facilitates the maintenance of the entire network with the availability of new interactions on other market(s). Having the new market-specific model forked and initialized as described, we simply fine-tune the model using only the data from the target market. We keep the loss function the same for this part. However, one could easily change the loss function with this step to better adapt the market needs in the target market. The fine-tuning operation for each target market is shown in line 13 of Algorithm 1.

## 6 EXPERIMENTS

### 6.1 Experimental Setup

**Dataset.** We use XMarket dataset for our experimental evaluations. We specifically focus on Elec. category across 8 markets presented in Table 1. We prepare our data similar to single-market experimental setup in the literature [2-4, 45, 56]. For the ratings, we filtered items and users that there exist less than five transactions. We follow a long line of literature and use leave-one-out evaluation [9, 17, 19, 20, 27, 28, 33].

**Compared methods.** In order to show the effectiveness of our method, we employ the following models on each target market:

- **GMF, MLP, NMF:** The Generalized Matrix Factorization (GMF), Multi-Layer Perceptron (MLP) and Neural Matrix Factorization (NMF) models from [19] trained using only the target market.
- **GMF++, MLP++, NMF++:** The simplest way of leveraging the cross-market data is to train the model on the interactions inside both the source and target markets by sharing the item and user representations. We equally sample from both markets in the training phase—equal number of training interactions from each market is used. We experimentally observed that this training provides higher performance compared to simple concatenation of both markets.
- **DDTCDR:** CDR and CMR have similarities as discussed in Sec. 2. To test whether high performing CDR methods can be used to effectively solve the CMR problem, we include the algorithm proposed in [33], as one of the recent strong CDR methods in our comparison. As the assumption with CDR is that the set of users are shared across two domains, the original model connects the user features between a pair of MLP networks using an orthogonal transformation matrix. We adapt the model into CMR by connecting the item features between two market's MLP networks. We performed a similar modifications for the CoNet's network structure proposed by Hu et al. [20] for the CDR problem. We only report DDTCDR due to its consistent superiority compared to CoNet and the space limitations.
- **MAML:** Meta-learning is widely used in the recommendation literature for variety of problem settings—see Section 2. Here, we adapt the learning paradigm to the CMR by employing the MAML framework providing model-agnostic solution for meta-learning. Our MAML training phase is described in Algorithm 1 lines 1-10. Here, after the training phase, we perform a single pass with  $K$ -shots of sampled interactions from the validation split of the target market and fast adapt the pre-trained model parameters to the

**Table 3: Performance comparison of different CMR methods. Best performing method in each scenario is shown in bold fonts. Significance (Student’s t-test) with  $p < 0.05$  compared to MAML and NMF++ is indicated by \* and †, respectively.**

		nDCG@10							HR@10						
		de	jp	in	fr	ca	mx	uk	de	jp	in	fr	ca	mx	uk
Single	GMF	0.1109	0.2016	0.0324	0.4085	0.2960	0.5626	0.4685	0.4435	0.3033	0.0843	0.5246	0.4500	0.6452	0.5923
	MLP	0.2506	0.3331	0.5921	0.3980	0.2522	0.5368	0.4662	0.4642	0.4324	0.6867	0.5424	0.4444	0.6636	0.5978
	NMF	0.2486	0.3394	0.5970	0.3961	0.2917	0.5212	0.4853	0.4824	0.4447	0.6807	0.5451	0.4533	0.6648	0.6035
Best-Src	GMF++	0.2809	0.3253	0.5798	0.4074	0.2853	0.5598	0.4628	0.4605	0.4713	0.6807	0.5532	0.4411	0.6581	0.5945
	MLP++	0.2643	0.3268	0.5891	0.3991	0.2715	0.5557	0.4801	0.4753	0.4939	0.6807	0.5604	0.4389	0.6642	0.6053
	NMF++	0.3116	0.3890	0.6020	0.4165	0.3134	0.5511	0.5050	0.5083	0.5123	0.6988	0.5742	0.4692	0.6765	0.6232
	DDTCDR	0.2553	0.2960	0.4994	0.3702	0.2958	0.5182	0.4192	0.4245	0.3873	0.5542	0.5228	0.4717	0.6028	0.5855
	MAML	0.3402 <sup>†</sup>	<b>0.4258<sup>†</sup></b>	0.6076	0.4686 <sup>†</sup>	0.3512 <sup>†</sup>	0.5868 <sup>†</sup>	<b>0.5272<sup>†</sup></b>	0.5306 <sup>†</sup>	0.5635 <sup>†</sup>	0.6988	0.6046 <sup>†</sup>	0.5058 <sup>†</sup>	0.7066 <sup>†</sup>	<b>0.6478<sup>†</sup></b>
	NMF-FOREC	0.3515 <sup>†</sup>	0.3847	0.6249 <sup>†</sup>	0.4387 <sup>†</sup>	0.3454 <sup>†</sup>	0.5886 <sup>†</sup>	0.5102	0.5169	0.5082	0.7048	0.5748	0.4735	0.6832	0.6150
	FOREC	<b>0.3621<sup>*</sup></b>	0.4195 <sup>†</sup>	<b>0.6378<sup>*</sup></b>	<b>0.4755<sup>†</sup></b>	<b>0.3693<sup>†</sup></b>	<b>0.6160<sup>*</sup></b>	0.5252 <sup>†</sup>	<b>0.5480<sup>*</sup></b>	<b>0.5717<sup>†</sup></b>	<b>0.7169<sup>*</sup></b>	<b>0.6148<sup>†</sup></b>	<b>0.5159<sup>†</sup></b>	<b>0.7152<sup>†</sup></b>	0.6465 <sup>†</sup>
Ave-Src	GMF++	0.2707	0.3036	0.4986	0.4005	0.2827	0.5360	0.4587	0.4537	0.4198	0.5813	0.5412	0.4385	0.6442	0.5882
	MLP++	0.2573	0.3350	0.5667	0.3970	0.2619	0.5352	0.4688	0.4657	0.4588	0.6649	0.5495	0.4339	0.6552	0.5963
	NMF++	0.2876	0.3533	0.5809	0.4123	0.3044	0.5387	0.4899	0.4890	0.4685	0.6687	0.5595	0.4608	0.6652	0.6096
	DDTCDR	0.2155	0.2044	0.3095	0.3285	0.2551	0.4670	0.3923	0.3678	0.2945	0.4053	0.4560	0.4117	0.5675	0.5416
	MAML	0.3336 <sup>†</sup>	0.3756	0.5948	0.4555 <sup>†</sup>	0.3463 <sup>†</sup>	0.5793 <sup>†</sup>	0.5174 <sup>†</sup>	0.5111 <sup>†</sup>	0.4941	0.6764	0.5959 <sup>†</sup>	0.4939 <sup>†</sup>	0.6878 <sup>†</sup>	0.6335 <sup>†</sup>
	NMF-FOREC	0.3340 <sup>†</sup>	0.3742	0.5974	0.4374 <sup>†</sup>	0.3423 <sup>†</sup>	0.5747 <sup>†</sup>	0.5063	0.4968	0.4819	0.6713	0.5652	0.4673	0.6657	0.6079
	FOREC	<b>0.3523<sup>*</sup></b>	<b>0.4007<sup>*</sup></b>	<b>0.6140<sup>†</sup></b>	<b>0.4637<sup>*</sup></b>	<b>0.3616<sup>*</sup></b>	<b>0.5989<sup>*</sup></b>	<b>0.5200<sup>†</sup></b>	<b>0.5280<sup>*</sup></b>	<b>0.5187<sup>*</sup></b>	<b>0.6937<sup>†</sup></b>	<b>0.5994<sup>†</sup></b>	<b>0.5064<sup>*</sup></b>	<b>0.6981<sup>†</sup></b>	<b>0.6364<sup>†</sup></b>
Fix-Src (us)	GMF++	0.2571	0.3123	0.5649	0.3844	0.2825	0.5162	0.4628	0.4498	0.4201	0.6687	0.5382	0.4398	0.6311	0.5945
	MLP++	0.2566	0.3227	0.5728	0.3935	0.2773	0.5291	0.4693	0.4498	0.4344	0.6747	0.5547	0.4347	0.6519	0.6026
	NMF++	0.3008	0.3446 <sup>*</sup>	0.6020	0.4208	0.3101	0.5509	0.4994	0.4857	0.4508	0.6988	0.5598	0.4610	0.6630	0.6151
	DDTCDR	0.2376	0.2196	0.3763	0.3702	0.2958	0.3592	0.4192	0.3997	0.3299	0.4277	0.5228	0.4717	0.5433	0.5855
	MAML	0.3295 <sup>†</sup>	0.3154	0.5622	0.4403	0.3512 <sup>†</sup>	0.5970 <sup>†</sup>	0.5140	0.5065	0.4488	0.6506	0.5844 <sup>†</sup>	0.5058 <sup>†</sup>	0.7035 <sup>†</sup>	0.6315 <sup>†</sup>
	NMF-FOREC	0.3265 <sup>†</sup>	<b>0.3620<sup>*</sup></b>	<b>0.6249<sup>*</sup></b>	0.4417 <sup>†</sup>	0.3394 <sup>†</sup>	0.5671	0.5076	0.4879	0.4549	<b>0.7048</b>	0.5658	0.4635	0.6593	0.6109
	FOREC	<b>0.3306<sup>†</sup></b>	0.3563 <sup>*</sup>	0.6143 <sup>*</sup>	<b>0.4485<sup>†</sup></b>	<b>0.3693<sup>*</sup></b>	<b>0.6160<sup>†</sup></b>	<b>0.5252<sup>†</sup></b>	<b>0.5158<sup>†</sup></b>	<b>0.4877<sup>*</sup></b>	0.6928	<b>0.5886<sup>†</sup></b>	<b>0.5159<sup>†</sup></b>	<b>0.7152<sup>†</sup></b>	<b>0.6420<sup>†</sup></b>

target market. We observed no further improvements with more number of passes on the adaptations. This baseline provides the sole importance of our MAML adaptation to the CMR problem.

- **NMF-FOREC:** In order to show the impact of our MAML-based pre-training with the FOREC, we pre-train our market-agnostic model with only NMF++ method described above and perform forking and fine-tuning for each specific market. This baseline provides evidences in two ways; (1) The importance of MAML pre-training on the performance of the FOREC model, (2) The sole impact of forking and fine-tuning operations over a weak pre-training of internal features.

**Hyper-parameters.** For GMF, MLP and NMF we follow [19] and set all the network structure and the latent factor dimension as suggested, i.e. [16, 64, 32, 16, 8] with 8 as embedding dimensions. For the optimizer, we use Adam [30] and select the learning rate and  $l_2$ -regularization coefficient hyper-parameters using the validation data of a subset of our markets. For the learning rate we considered  $\{0.1, 0.05, 0.01, 0.005, 0.001, 0.0001\}$  and selected 0.01 for MLP and NMF and 0.005 for the GMF model. For regularization we observed that  $1e^{-7}$  is the best among our consideration set. We use 4 negative training samples for each user re-sampled with each epoch. ReLU is used as the activation function. For DDTCDR we use the same hyper-parameters provided in the original implementation of [33]. As the model uses a preset embeddings, we employ the GMF model for the initialization. For MAML training, we set the fast learning rate  $\beta = 0.1$  selected from  $\{0.5, 0.3, 0.1, 0.01, 0.001\}$

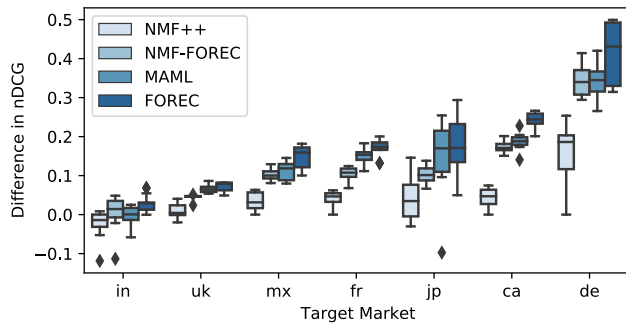
and the number of shots as 20 selected from  $\{5, 20, 50, 100, 200\}$ . For our FOREC’s market-agnostic part we use the same architecture as of NMF and employ the last three layers of the NMF network for forking and freeze the remaining bottom layers as well as the embeddings. For the market head layers, we considered three different layer sizes; (a) no new layers, (b) two 16 layers, and (c) 3 layers with [16, 32, 16], and evaluated experimentally using the validation data on a subset of markets. We selected (c) as our model’s market head layers for all experiments. In addition, we observed that setting higher  $l_2$ -regularization with the fine-tuning step helps the overall performance, especially with lower resourced markets—we set it to 0.001 for all fine-tuning steps.

**Evaluation Metrics.** We use Hit-Rate (HR) and nDCG as our evaluation metrics, commonly used in the literature. We report these metrics for a cut-off of 10. Similar to other works, we constructed the ground truth using the buying behavior by considering an item as relevant if the user gave a rating. In addition, we follow the literature and sample 99 negative items for each user in our evaluations.

## 6.2 Results and Discussion

We compare our FOREC model to several baseline techniques discussed in Sec. 6.1 in terms of recommendation performance. In theory, each target market can be paired with each auxiliary source market. However, for the 7 markets that we consider, all possible pairings of source and target markets leads to 49 different settings. For better readability and due to space limitation, we report our





**Figure 4: Impact of choosing different source markets on different target markets for different models.**

results in the following scenarios for every target market:

- **Best-Src:** Each of the parallel markets are once considered as source and the model is evaluated using the nDCG@10 on the validation set. For each method, the source market with the highest improvement on the validation set is selected. As such, for a single target market, the best source market for different CMR methods may be different.
- **Ave-Src:** In order to provide a rough indicator of the safer choice of the CMR method on each target market along with an overall insight on different source selections, we report the average performance of each model using different source markets.
- **Fix-Src:** We report the results when a fixed base market is available (i.e. **us** market) and all CMR methods can only use that market to improve the target market recommendation.

Table 3 summarizes the evaluation results with and without cross-market data. All three aforementioned source selection scenarios are reported respectively in the Table. The best performing method for each target market and each source selection scenario is shown by bold fonts.

We observe that FOREC is the winner in almost all target markets and across all the source selection scenarios, except for **jp** and **uk** in Best-Src scenario, and **jp** and **in** for Fix-Src. In addition, we observe that even our simple baselines are able to utilize to some degree the cross-market data provided and improve the target market’s recommendation performance. This suggests the importance of cross-market training, even with a fixed base market, for better recommendation systems across different local markets. It also suggests the importance of the model and source market selection for deployment purposes.

When the source market is selected by the validation performance (i.e. Best-Src scenario), FOREC and MAML noticeably outperform other CMR methods. For **uk**, MAML is slightly better than FOREC, though the difference is not substantial. For **jp**, MAML has a higher nDCG, but FOREC is better in terms of HR, suggesting that neither can be picked as the winner. For the other 5 markets, FOREC outperforms MAML both in terms of nDCG and HR.

Looking at the results in Ave-Src scenario, FOREC outperforms other powerful CMR methods on average, meaning that given a source and target market, FOREC is a safer choice for CMR. This is further illustrated in Table 4 where we report for each target market the number of source markets for which each of FOREC or MAML is the winner along with the relative percentage improvement of

**Table 4: nDCG Comparison of FOREC vs. MAML with different source markets. Values denote the percentage of relative improvement of FOREC compared with MAML. Positive values indicate FOREC superiority.**

	de	jp	in	fr	ca	mx	uk	avg	
Source Markets	de	-	-1.62	4.96	4.07	3.55	1.23	-0.74	1.64
	jp	7.33	-	2.87	1.05	0.40	3.39	0.06	2.16
	in	2.56	6.48	-	0.75	4.90	2.94	-0.40	2.46
	fr	14.34	-1.47	1.42	-	7.67	1.98	0.05	3.43
	ca	9.81	5.75	-0.08	0.32	-	3.44	0.62	2.84
	mx	4.51	6.24	3.48	2.86	5.28	-	0.79	6.18
	uk	3.87	24.51	2.39	3.99	4.86	3.62	-	6.18
	us	0.32	12.97	9.26	1.85	5.16	3.18	2.17	4.99
<b>avg</b>	5.34	6.61	3.04	1.86	3.98	2.47	0.32	-	

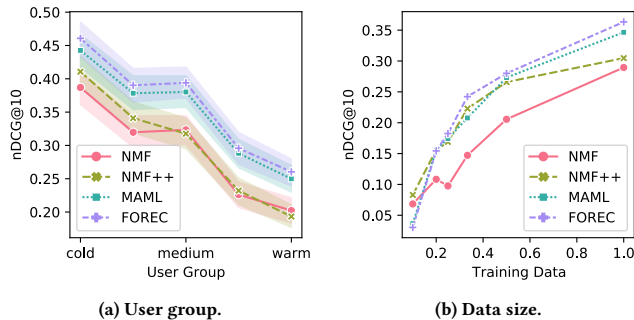
FOREC compared to MAML. As it can be seen, over all 49 possible combinations of source-target pairs, only in 5 of them MAML is slightly better than FOREC. As source markets, **mx** and **uk** are the ones with the highest average improvement of FOREC over MAML, while **de** market is the one where MAML performs closest to FOREC and even surpasses it in two target markets.

Comparing the Fix-Src results, suggests that even though **us** is not the best source market for all target markets in terms of performance boost (see Best-Src scenario), comparing with the single market results, it can be observed that **us** helps all of the tested target markets and the CMR comes to its highest performance when used with our FOREC model. In addition, we notice that NMF-FOREC is performing better in two markets compared to FOREC. We hypothesize that this might be due to the relative data size differences between the **us** and each of these markets (see Table 2), performing the pre-training using MAML provides a limited added value.

Comparing NMF-FOREC with FOREC, we observe that in most of the target markets, FOREC performs better. Based on the observations from Table 3, in most of the cases, MAML performs better than NMF++ and FOREC performs better than MAML. This shows the significance of the MAML-based pre-training as well as forking and fine-tuning, as proposed by our model.

Fig. 4 compares these four models more deeply. In this figure, the nDCG@10 improvement of these CMR models over the NMF on single market are shown for different target markets. For each method in each target market, the distribution of nDCG@10 improvements based on different source markets is given as a box plot. This figure provides a better illustration of the trend between these four models described earlier.

Finally, we observe that the adopted DDTCDR, as one of the state-of-the-art CDR methods, does not perform well for CMR. In order to adopt a CDR method to CMR scenario, as described before, the users and items should be interchanged: the shared users in CDR are analogous to the shared items in CMR. This change of perspective looks natural at first sight, but it introduces some issues. Here we discuss the issue. In the item recommendation problem, a number of past interactions of a user with items are used to predict her future interactions. This means that the evaluation is based on the accuracy of the predicted items for users. In CDR, the users are shared across domains and the interacted items from the source domain add to the per-user information of the target



**Figure 5: Performance comparison of FOREC on (a) different user groups; and (b) different training data size.**

domain. Differently, in CMR, the items are shared across markets and the users are separate. Changing the perspective of users/items during training but having a fixed evaluation based on the per-user predictions is the issue of naïvely adopting a CDR to CMR. An interesting future direction would be to analyze the impact of such adoption for user recommendation in CMR scenarios.

### 6.3 Impact of Cross-Market Training on Different Users

Here we study the impact of different cross-market training approaches on user groups in terms of their training data size. Following the work of Liu et al. [35] we split the users based on the interactions into five groups. The users with the least number of interactions are named *cold* (average of 5 interactions), and the ones with the highest number of interactions are called *warm* (average of 13.3 interactions). We create five equally-sized user splits and report the average nDCG@10 for each group in Fig. 5a. This figure contains the performance of four models on these five user groups in the *ca* market: single market NMF, NMF++, MAML and FOREC. Here, we used *de* as the source market.<sup>7</sup> We observe that MAML and FOREC almost uniformly improve the performance over the single market NMF in all five user groups. Comparing FOREC with MAML, again we see a consistent improvement in all five user groups with a slightly bigger gap toward the cold user group. This observation, together with other similar observations on our tests over other source-target pairs, provide experimental evidence that FOREC is suitable both for cold- and warm-start situations in the target market. NMF++, on the other hand, only helps the cold user groups, as can be seen in the figure.

### 6.4 Impact of the Target Market Data Size

In order to study the impact of target training data size, for a given market pair, we gradually decrease the number of training interactions for each user. Starting from the full target market user-item interactions, each time we halve the training data for the target market until only 10% of the target market’s training interactions remains for each user. We train four of our models (NMF, NMF++, MAML, FOREC) on each of these settings and test on the target market. Fig. 5b presents the resulting nDCG@10 performance on *ca* with *uk* as the source market—similar observations made for a many pairs. As it can be seen, among the cross-market methods

we see that FOREC and MAML are performing similarly especially when the target market’s size is extremely small. On the extremely small target market size (10% of the data), we see that single market NMF model as well as NMF++ are performing better. As more data becomes available, a significant boost among cross-market methods are observed, especially with FOREC—as opposed to NMF with lazy reaction to the new data availability. With further data availability, we see the full superiority of MAML as well as FOREC. This analysis suggests that some minimum amount of training data in the target market is essential for the cross-market models that we examined in order to be able to make use of the auxiliary source market. We hypothesize that when the target market provides limited amount of training data the pre-training through the MAML approach shifts parameters more toward the source market compared to the NMF++ training approach. However, the general observation from this analysis could be that FOREC and MAML are relatively resilient to the amount of the target training data. We note that this requires further analysis which is out of the scope of our study.

## 7 CONCLUSION & FUTURE WORK

We studied the problem of recommending relevant products to users in relatively resource-scarce markets by leveraging data from similar or richer-in-resource auxiliary markets. To this aim, we introduced a large-scale real-life dataset, named as XMarket, providing product information and reviews on 18 Amazon marketplaces featuring 52.5 million user-item interactions. We hypothesized and showed through extensive experiments on 7 target markets that data from one market can actually be used to improve the performance in another. Our model, named as FOREC, demonstrates robust effectiveness, consistently improving the performance on target markets compared to competitive baselines selected for our analysis. In particular, FOREC improves on average 24% and up to 50% in terms of nDCG@10, compared to the NMF baseline.

Our analysis and experiments suggest specific future directions in this research area. We show that models that are designed for CDR are not necessarily suitable for the market adaptation problem setting. One interesting extension of our study could be designing models that are domain and market agnostic in the sense that they can consume the data across different markets and domains and leverage that for the improved recommendation on a target market’s specific domain. In addition, we believe that data filtering or selection across markets could potentially be helpful for the models we discussed in our study. Moreover, using data augmentation techniques to generate synthetic ratings for target markets [8, 51] could be a potential solution for the extreme low-resource markets, i.e. cold-start markets. We believe that many potentially interesting problems are yet to be explored in the CMR area.

**Acknowledgments.** This work was supported in part by the Center for Intelligent Information Retrieval and in part by the NWO Innovational Research Incentives Scheme Vidi (016.Vidi.189.039), the NWO Smart Culture - Big Data / Digital Humanities (314-99-301), the Elsevier and NWO (612.001.551), and the H2020-EU.3.4. - SOCIETAL CHALLENGES - Smart, Green And Integrated Transport (814961). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsors.

<sup>7</sup>We repeated the experiments with different source and target markets and got similar results. Here we only show one case.

## REFERENCES

- [1] Aman Ahuja, Nikhil Rao, Sumeet Katariya, Karthik Subbian, and Chandan K Reddy. 2020. Language-Agnostic Representation Learning for Product Search on E-Commerce Platforms. In *WSDM*. 7–15.
- [2] Mohammad Aliannejadi and Fabio Crestani. 2018. Personalized Context-Aware Point of Interest Recommendation. *ACM Trans. Inf. Syst.* 36, 4 (2018), 45:1–45:28.
- [3] Mohammad Aliannejadi, Ida Mele, and Fabio Crestani. 2016. Venue Appropriateness Prediction for Contextual Suggestion. In *TREC*.
- [4] Mohammad Aliannejadi, Dimitrios Rafailidis, and Fabio Crestani. 2020. A Joint Two-Phase Time-Sensitive Regularized Collaborative Ranking Model for Point of Interest Recommendation. *IEEE Trans. Knowl. Data Eng.* 32, 6 (2020), 1050–1063.
- [5] Homanga Bharadhwaj. 2019. Meta-learning for user cold-start recommendation. In *IJCNN*. 1–8.
- [6] Hamed Bonab, Sheikh Muhammad Sarwar, and James Allan. 2020. Training Effective Neural CLIR by Bridging the Translation Gap. In *SIGIR*. ACM, 9–18.
- [7] Rocío Cañamares and Pablo Castells. 2018. Should I Follow the Crowd?: A Probabilistic Analysis of the Effectiveness of Popularity in Recommender Systems. In *SIGIR*. 415–424.
- [8] Dong-Kyu Chae, Jihoo Kim, Duen Horng Chau, and Sang-Wook Kim. 2020. ARCF: Augmenting Virtual Users and Items in Collaborative Filtering for Addressing Cold-Start Problems. In *SIGIR*. 1251–1260.
- [9] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Inspir, et al. 2016. Wide & Deep Learning for Recommender Systems. In *DLRS@RecSys*. 7–10.
- [10] Andrew Collins, Dominika Tkaczyk, and Joeran Beel. 2018. A Novel Approach to Recommendation Algorithm Selection using Meta-Learning. In *AICS*. 210–219.
- [11] Tiago Cunha, Carlos Soares, and André CPLF de Carvalho. 2018. Metalearning and Recommender Systems: A Literature Review and Empirical Study on the Algorithm Selection Problem for Collaborative Filtering. *Information Sciences* 423 (2018), 128–144.
- [12] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems. In *WWW*. 278–288.
- [13] Bruce Ferwerda, Andreu Vall, Marko Tkalcic, and Markus Schedl. 2016. Exploring Music Diversity Needs Across Countries. In *UMAP*. 287–288.
- [14] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *ICML*. 1126–1135.
- [15] Wenjing Fu, Zhaohui Peng, Senzhang Wang, Yang Xu, and Jin Li. 2019. Deeply Fusing Reviews and Contents for Cold Start Users in Cross-Domain Recommendation Systems. In *AAAI*. 94–101.
- [16] Yaroslav Ganin and Victor Lempitsky. 2015. Unsupervised Domain Adaptation by Backpropagation. In *ICML*. 1180–1189.
- [17] Yingqiang Ge, Shuyuan Xu, Shuchang Liu, Zuohui Fu, Fei Sun, and Yongfeng Zhang. 2020. Learning Personalized Risk Preferences for Recommendation. In *SIGIR*. 409–418.
- [18] Ruining He and Julian J. McAuley. 2016. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering. In *WWW*. 507–517.
- [19] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *WWW*. 173–182.
- [20] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. Conet: Collaborative Cross Networks for Cross-Domain Recommendation. In *CIKM*. 667–676.
- [21] Zhiqi Huang, Hamed Bonab, Sheikh Muhammad Sarwar, Razieh Rahimi, and James Allan. 2021. Mixed Attention Transformer for Leveraging Word-Level Knowledge to Neural Cross-Lingual Information Retrieval. In *CIKM*.
- [22] Il Im and Alexander Hars. 2007. Does a One-Size Recommendation System Fit All? The Effectiveness of Collaborative Filtering based Recommendation Systems across Different Domains and Search Modes. *ACM Trans. on Inf. Sys. (TOIS)* 26, 1 (2007), 4–es.
- [23] Ling Alice Jiang, Zhilin Yang, and Minjoon Jun. 2013. Measuring Consumer Perceptions of Online Shopping Convenience. *Journal of Service Management* 24, 2 (2013), 191–214.
- [24] Meng Jiang, Peng Cui, Xumin Chen, Fei Wang, Wenwu Zhu, and Shiqiang Yang. 2015. Social Recommendation with Cross-Domain Transferable Knowledge. *IEEE Trans. Knowl. Data Eng.* 27, 11 (2015), 3084–3097.
- [25] Yaru Jin, Shoubin Dong, Yong Cai, and Jinlong Hu. 2020. RACRec: Review Aware Cross-Domain Recommendation for Fully-Cold-Start User. *IEEE Access* 8 (2020), 55032–55041.
- [26] Heishiro Kanagawa, Hayato Kobayashi, Nobuyuki Shimizu, Yukihiro Tagami, and Taiji Suzuki. 2019. Cross-Domain Recommendation via Deep Domain Adaptation. In *ECIR*. 20–29.
- [27] SeongKu Kang, Junyoung Hwang, Dongha Lee, and Hwanjo Yu. 2019. Semi-supervised Learning for Cross-Domain Recommendation to Cold-Start Users. In *CIKM*. 1563–1572.
- [28] Wang-Cheng Kang and Julian McAuley. 2018. Self-Attentive Sequential Recommendation. In *ICDM*. 197–206.
- [29] Phillip Keung, Yichao Lu, György Szarvas, and Noah A. Smith. 2020. The Multilingual Amazon Reviews Corpus. In *EMNLP*. 4563–4568.
- [30] Diederik P Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [31] Adit Krishnan, Mahashweta Das, Mangesh Bendre, Hao Yang, and Hari Sundaram. 2020. Transfer Learning via Contextual Invariants for One-to-Many Cross-Domain Recommendation. In *SIGIR*. 1081–1090.
- [32] Hoyeop Lee, Jinbae Im, Seongwon Jang, Hyunsouk Cho, and Sehee Chung. 2019. MeLU: Meta-Learned User Preference Estimator for Cold-Start Recommendation. In *KDD*. 1073–1082.
- [33] Pan Li and Alexander Tuzhilin. 2020. DDTCCR: Deep Dual Transfer Cross Domain Recommendation. In *WSDM*. 331–339.
- [34] Ying Li, Jia-Jie Xu, Peng-Peng Zhao, Jun-Hua Fang, Wei Chen, and Lei Zhao. 2020. ATLRec: An Attentional Adversarial Transfer Learning Network for Cross-Domain Recommendation. *Journal of Computer Science and Technology* 35, 4 (2020), 794–808.
- [35] Yiding Liu, Tuan-Anh Pham, Gao Cong, and Quan Yuan. 2017. An Experimental Evaluation of Point-of-interest Recommendation in Location-based Social Networks. *Proc. VLDB Endow.* 10, 10 (2017), 1010–1021.
- [36] Yuanfu Lu, Yuan Fang, and Chuan Shi. 2020. Meta-Learning on Heterogeneous Information Networks for Cold-Start Recommendation. In *KDD*. 1563–1573.
- [37] Zhongqi Lu, Erheng Zhong, Lili Zhao, Evan Wei Xiang, Weike Pan, and Qiang Yang. 2013. Selective Transfer Learning for Cross Domain Recommendation. In *ICDM*. 641–649.
- [38] Mi Luo, Fei Chen, Pengxiang Cheng, Zhenhua Dong, Xiuqiang He, Jiashi Feng, and Zhenguo Li. 2020. MetaSelector: Meta-Learning for Recommendation with User-Level Adaptive Model Selection. In *TheWebConference*. 2507–2513.
- [39] Julian J. McAuley, Christopher Targett, Qinfeng Shi, and Anton van den Hengel. 2015. Image-Based Recommendations on Styles and Substitutes. In *SIGIR*. 43–52.
- [40] Nima Mirbakhsh and Charles X Ling. 2015. Improving Top-N Recommendation for Cold-Start Users via Cross-Domain Information. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 9, 4 (2015), 1–19.
- [41] Jianmo Ni, Jiacheng Li, and Julian J. McAuley. 2019. Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. In *EMNLP/IJCNLP*. 188–197.
- [42] Dilruk Perera and Roger Zimmermann. 2019. CnGAN: Generative Adversarial Networks for Cross-Network User Preference Generation for Non-Overlapped Users. In *WWW*. 3144–3150.
- [43] Dimitrios Rafailidis and Fabio Crestani. 2017. A Collaborative Ranking Model for Cross-Domain Recommendations. In *CIKM*. 2263–2266.
- [44] Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. 2020. Rapid Learning or Feature Reuse? Towards Understanding the Effectiveness of MAML. In *ICLR*.
- [45] Hossein A. Rahmani, Mohammad Aliannejadi, Mitra Baratchi, and Fabio Crestani. 2020. Joint Geographical and Temporal Modeling Based on Matrix Factorization for Point-of-Interest Recommendation. In *ECIR*. 205–219.
- [46] Kevin Roitero, Ben Carterre, Rishabh Mehrotra, and Mounia Lalmas. 2020. Leveraging Behavioral Heterogeneity Across Markets for Cross-Market Training of Recommender Systems. In *TheWebConference*. 694–702.
- [47] Wenlong Sun, Sami Khenissi, Olfa Nasraoui, and Patrick Shafto. 2019. Debiasing the Human-Recommender System Feedback Loop in Collaborative Filtering. In *WWW*. 645–651.
- [48] Manasi Vartak, Arvind Thiagarajan, Conrado Miranda, Jeshua Bratman, and Hugo Larochelle. 2017. A Meta-Learning Perspective on Cold-Start Recommendations for Items. In *NeurIPS*. 6904–6914.
- [49] Ricardo Vilalta and Youssef Drissi. 2002. A Perspective View and Survey of Meta-Learning. *Artificial intelligence review* 18, 2 (2002), 77–95.
- [50] Cheng Wang, Mathias Niepert, and Hui Li. 2019. RecSys-DAN: Discriminative Adversarial Networks for Cross-Domain Recommender Systems. *IEEE Trans. Neural Networks Learn. Syst.* 31, 8 (2019), 2731–2740.
- [51] Qinyong Wang, Hongzhi Yin, Hao Wang, Quoc Viet Hung Nguyen, Zi Huang, and Lizhen Cui. 2019. Enhancing Collaborative Filtering with Generative Augmentation. In *KDD*. 548–556.
- [52] Xinghua Wang, Zhaohui Peng, Senzhang Wang, S Yu Philip, Wenjing Fu, and Xiaoguang Hong. 2018. Cross-Domain Recommendation for Cold-Start Users via Neighborhood based Feature Mapping. In *DASEAA*. 158–165.
- [53] Feng Yuan, Lina Yao, and Boualem Benatallah. 2019. DARec: Deep Domain Adaptation for Cross-Domain Recommendation via Transferring Rating Patterns. In *IJCAI*. 4227–4233.
- [54] 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*. 1479–1488.
- [55] Cheng Zhao, Chenliang Li, Rong Xiao, Hongbo Deng, and Aixin Sun. 2020. CATN: Cross-Domain Recommendation for Cold-Start Users via Aspect Transfer Network. In *SIGIR*. 229–238.
- [56] Jie Zou, Yifan Chen, and Evangelos Kanoulas. 2020. Towards Question-based Recommender Systems. In *SIGIR*. 881–890.