Market-Aware Models for Efficient Cross-Market Recommendation

Bhargav, S.; Aliannejadi, M.; Kanoulas, E.

DOI
10.1007/978-3-031-28244-7_9

Publication date
2023

Document Version
Final published version

Published in
Advances in Information Retrieval

License
Article 25fa Dutch Copyright Act (https://www.openaccess.nl/en/in-the-netherlands/you-share-we-take-care)

Citation for published version (APA):

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 426, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)
Market-Aware Models for Efficient Cross-Market Recommendation

Samarth Bhargav, Mohammad Aliannejadi, and Evangelos Kanoulas
University of Amsterdam, Amsterdam, The Netherlands
{s.bhargav,m.aliannejadi,e.kanoulas}@uva.nl

Abstract. We consider the cross-market recommendation (CMR) task, which involves recommendation in a low-resource target market using data from a richer, auxiliary source market. Prior work in CMR utilised meta-learning to improve recommendation performance in target markets; meta-learning however can be complex and resource intensive. In this paper, we propose market-aware (MA) models, which directly model a market via market embeddings instead of meta-learning across markets. These embeddings transform item representations into market-specific representations. Our experiments highlight the effectiveness and efficiency of MA models both in a pairwise setting with a single target-source market, as well as a global model trained on all markets in unison. In the former pairwise setting, MA models on average outperform market-unaware models in 85% of cases on nDCG@10, while being time-efficient—compared to meta-learning models, MA models require only 15% of the training time. In the global setting, MA models outperform market-unaware models consistently for some markets, while outperforming meta-learning-based methods for all but one market. We conclude that MA models are an efficient and effective alternative to meta-learning, especially in the global setting.

Keywords: Cross-market recommendation · Domain adaptation · Market adaptation

1 Introduction

Cross-market recommendation (CMR) involves improving recommendation performance in a target market using data from one or multiple auxiliary source markets. Data from source markets, which have rich- and numerous interactions, are leveraged to aid performance in a target market with fewer interactions. For instance, an e-commerce company well-established in Germany may want to start selling its products in Denmark. Using CMR methods, data from the German market can be utilised to augment recommender performance in the Danish market. This task is challenging since target market data can be scarce or otherwise unavailable, and user behaviours may differ across markets [2,7,24].

Research in CMR tackles multiple challenges. One challenge is to select the best source market, which is crucial since user behaviours across markets may
vary [2, 24], which may harm performance instead of bolstering it. Furthermore, effectively utilising data from multiple markets at the same time without harming performance can be challenging [2]. Another key obstacle is effectively modelling a market, in addition to users and items. Bonab et al. [2] treat recommendation in each market as a task in a multi-task learning (MTL) framework, using meta-learning to learn model parameters. This is followed by a fine-tuning step per market. These two steps enable models to learn both common behaviours across markets as well as market-specific behaviours. However, meta-learning can be resource intensive compared to other methods. In addition to this, utilising new data from source markets requires re-running the meta-learning step.

We propose market-aware (MA) models to address these limitations. We aim to explicitly model each market as an embedding, using which an item representation can be transformed and ‘customised’ for the given market. Compared to meta-learning models, we show that MA models are far more efficient to train. Furthermore, they are trained in one go, enabling easier model updates when new data is collected. MA models are built on the hypothesis that explicit modelling of markets allows better generalisation. In essence, an item representation is a product of (i) an across-market item embedding and (ii) a market embedding. The former is learnt from data across markets, and aims to capture an item representation applicable across markets; the latter enables market-specific behaviours to be captured.

In our experiments, we compare MA models with market-unaware baselines as well as meta-learning models. We do so in multiple settings, utilising data from several markets: the pairwise setting, which deals with a single target-source pair, and the global setting which trains one model for recommendation in all markets. In the pairwise setting, we show that MA models improve over market-unaware models for many markets, and match or beat meta-learning methods. This is significant since we show that training MA models require approximately the same time as market-unaware models and only 15% of the time required to train meta-learning models. We show that MA models especially excel in the global setting, outperforming meta-learning methods for nearly every market.

We examine the following research questions:

**RQ1.** Given a single source and target market, does explicitly modelling markets with embeddings lead to effective performance in the target market? We compare MA models against market-unaware as well as meta-learning models. We show MA models achieve the best performance for most markets, and when a single, best source is available they match or outperform baselines for all markets.

**RQ2.** How computationally expensive are MA models compared to market-unaware and meta-learning models? We show that MA models require similar training times as market-unaware models, and require fewer computational resources to train compared to meta-learning models while achieving similar or better performance.

---

1 https://github.com/samarthbhargav/efficient-xmrec.
RQ3. How do MA models compare against market-unaware models and meta-learning models when a global model is trained on all markets in unison? We show that MA models outperform or match market-unaware baselines, outperforming meta-learning models for all but one market.

2 Related Work

While both cross-domain recommendation (CDR) and CMR focus on improving recommender effectiveness using data from other domains (i.e., item categories) or markets, they present different challenges: CDR involves recommending items in a different domain for the same set of users, with the general assumption that the model learns from interactions of overlapping users. In CMR, items are instead shared across different markets, with each market having a different set of users. Interactions from auxiliary markets are leveraged to boost performance for users in the target market for a similar set of items.

Cross-domain Recommendation. CDR has been researched extensively [6, 12, 14, 17, 18, 20, 22, 23]. Prior approaches involve clustering-based algorithms [21] and weighing the influence of user preferences based on the domain [23]. Lu et al. [20] show that domain transfer may sometimes harm performance in the target domain. Neural approaches using similarity networks like DSSM [13] or transfer learning [6, 12] can be effective. DDCTR [18] utilizes iterative training across domains. Augmenting data with ‘virtual’ data [4, 22], as well as considering additional sources [27] have been shown to help. Other approaches leverage domain adaptation [9] for leveraging content for full cold-start [15], utilizing adversarial approaches [19, 25] or formulating it as an extreme classification problem [26]. Our approach is inspired by contextual invariants [17], which are behaviours that are consistent across domains, similar to our hypothesis that there are behaviours common across markets.

Cross-market Recommendation. CMR is relatively new and understudied compared to CDR. Ferwerda et al. [7] studied CMR from the perspective of country-based diversity. Roitero et al. [24] focus on CMR for music, investigating trade-offs between learning from local/single markets vs. a global model, proposing multiple training strategies. [2] release a new dataset for the Cross Market Product recommendation problem, which we utilise in our experiments. They design a meta-learning approach to transfer knowledge from a source market to a target market by freezing and forking specific layers in their models. The WSDM Cup 2022 challenge also dealt with this dataset, where most top teams utilized an ensemble of models based on different data pairs. Cao et al. [3] builds on the XMRec dataset and proposes multi-market recommendation, training a model to learn intra- and inter-market item similarities. In this work, we show that meta-learning methods are expensive to train. Instead, we show that market embeddings can encode and effectively transfer market knowledge,
beating or matching the performance of complex models while being much more efficient to train.

3 Methodology

We outline market-unaware models in Sect. 3.1, followed by market-aware models as well as meta-learning models in Sect. 3.2.

**Notation.** Given a set of markets \( \{M_0, M_1, \ldots, M_t\} \), such that market \( l \) has a items \( I_l \) and \( z_l \) users \( U_l = \{U^1_l \ldots U^{z_l}_l\} \). We assume the base market \( M_0 \) has \( I_0 \) s.t. \( I_0 \supset I_l \) for all \( 1 \leq l \leq m \). The task is to adapt a given market \( M_l \) using data from other markets \( M_{m \neq l} \) as well as data from the target market. We use \( p_u \) for the user embedding for user \( u \), \( q_i \) for the item embedding for item \( i \), and finally \( o_l \) for the market embedding for market \( l \). \( y_{ui} \) and \( \hat{y}_{ui} \) is the actual and predicted rating respectively. \( \odot \) denotes an element-wise product.

3.1 Market-Unaware Models

These models do not differentiate between users and items from different markets and are termed market-unaware since they do not explicitly model the market. We first outline three such models previously employed for CMR [2, 11]:

- **GMF**: The generalized matrix factorization (GMF) model computes the predicted rating \( \hat{y}_{ui} \) given \( p_u, q_i \) and parameters \( h \):
  \[
  \hat{y}_{ui} = \text{sigmoid}(h^T (p_u \odot q_i))
  \]

- **MLP**: An multi-layer perceptron (MLP) uses a \( L \) layer fully-connected network, such that:
  \[
  m_0 = \begin{bmatrix} p_u \\ q_i \end{bmatrix}
  \]
  \[
  m_{L-1} = \text{ReLU}(W_{L-1}^T \text{ReLU}(\ldots \text{ReLU}(W_1^T m_0 + b_1)) + b_{L-1})
  \]
  \[
  \hat{y}_{ui} = \text{sigmoid}(h^T m_{L-1})
  \]

- **NMF**: neural matrix factorization (NMF) combines both MLP and GMF. Given \( p^1_u, q^1_i \) for the MLP, and \( p^2_i, q^2_u \) for GMF, the NMF model computes the score as follows:
  \[
  m_0 = \begin{bmatrix} p^1_u \\ q^1_i \end{bmatrix}
  \]
  \[
  m_{MLP} = \text{ReLU}(W_{L}^T \text{ReLU}(\ldots \text{ReLU}(W_1^T m_0 + b_1))) + b_L)
  \]
  \[
  m_{GMF} = p^2_u \odot q^2_i
  \]
  \[
  \hat{y}_{ui} = \text{sigmoid}(h^T \begin{bmatrix} m_{GMF} \\ m_{MLP} \end{bmatrix})
  \]
For adapting to CMR, different sets of users from different markets are treated similarly, and training is performed on a combined item pool resulting in a single model. During inference for a user, however, only items from that market are ranked.

3.2 Market-Aware Models

We first discuss models proposed by Bonab et al. [2], followed by our proposed methods.

Meta-learning Baselines. Bonab et al. [2] propose using meta-learning in an MTL setting where each market is treated as a ‘task’. model-agnostic meta-learning (MAML) [8] is employed to train the base NMF model across markets. MAML employs two loops for training, an inner loop that optimises a particular market, and an outer loop that optimises across markets. This makes training expensive, as we will show in our experiments. Once a MAML model is trained, the FOREC model is obtained as follows for a given source/target market: (a) the MAML model weights are copied over to a new model, ‘forking’ it, (b) parts of the weights of the model are frozen and finally (c) the frozen model is fine-tuned on the given market.

Both MAML and FOREC are market aware but do not explicitly model the market i.e. a single item embedding is learned in MAML models for all markets, and while market adaptation is achieved through fine-tuning for FOREC, it requires maintaining separate sets of parameters, unlike the proposed MA models.

Market Aware Models. Markets here are explicitly modelled by learning embeddings for each of them, in addition to user and item embeddings. A market embedding adapts an item to the current market, which we argue is crucial for items that may be perceived differently in different markets. This aspect should be reflected in the latent representation of the item, motivating our approach. Both meta-learning and MA models learn item representations across markets, but MA models this explicitly via an element-wise product between a representation for an item and a market embedding. This produces item embeddings adapted to a given market. We augment the market-unaware baselines with market embeddings, producing MA models. We leave more complex methods, for instance—a neural network that models item/market interactions instead of an element-wise produce—for future work.

To obtain a market-adapted item embedding, we first (one-hot) encode a market \( l \), to obtain a market embedding \( o_l \); the dimensionality of \( o_l \) is the same as \( p_u \) and \( q_i \). The scores are computed as follows for the three proposed models:

- **MA-GMF**: For a user \( u \) in market \( l \), and item \( i \), we have embeddings \( p_u, o_l \) and \( q_i \):
  \[
  \hat{y}_{ui} = \text{sigmoid}(h^T(p_u \odot (o_l \odot q_i)))
  \]
- **MA-MLP**: This is the same as the MLP, with the initial embedding $m_0$ augmented with market information: $m_0 = \begin{bmatrix} p_u \\ q_i \odot o_l \end{bmatrix}$

- **MA-NMF**: The NMF model utilises both modifications listed above. That is:

\[
\begin{align*}
m_{GMF} &= p_u^2 \odot (o_l \odot q_i^2) \\
m_0 &= \begin{bmatrix} p_u^1 \\ q_i^1 \odot o_l \end{bmatrix}
\end{align*}
\]

These models are trained similarly to the market-unaware models, except the market is taken into consideration when making recommendations. Market embeddings are learned via backpropagation, similar to how user and item embeddings are learned, using a binary cross entropy loss [11].

Our proposed technique adds market awareness to all the models. Besides this, the proposed models are easier to update with new interactions compared to MAML/FOREC. While FOREC requires the expensive MAML pre-training followed by the fork and fine-tune step, MA models simply can be trained with new interaction data. In spite of this simplicity, MA models achieve similar performance compared to meta-learning models while requiring far lesser time to train, which we demonstrate in the following section.

### 4 Experimental Setup

We conduct two sets of experiments. The first set of experiments trains models with a single auxiliary source market for improving recommendation performance in a given target market. We term these pairwise experiments since one model is trained for a given source-target market pair. The second set of experiments deals with a global model trained on all markets in unison, with the goal of improving overall performance. We outline the dataset, evaluation, baselines, hyperparameters and training followed by a description of the experiments.

**Dataset.** We use the XMarket dataset [2] for all experiments. XMarket is an CMR dataset gathered from a leading e-commerce website with multiple markets. We utilise the largest subset, ‘Electronics’, considering the following markets (\# users, \# items, \# interactions): de (2373/ 2210/ 22247), jp (487/ 955/ 4485), in (239/ 470/ 2015), fr (2396/ 1911/ 22905), ca (5675/ 5772/ 55045), mx (1878/ 1645/ 17095), uk (4847/ 3302/ 44515), us (35916/ 31125/ 364339). We consider all markets except us as a target market, with all markets (including us) as possible source markets. Experiments are limited to XMarket as it is the only public dataset for research in CMR.

**Evaluation.** The data (per market) is split into a train/validation/test set, where one left-out item from the user history is used in the validation and test
set. This follows the leave-one-out strategy [5,10–12,16]. We extract 99 negatives per user for evaluating recommender performance in the validation/test set, following Bonab et al. [2]. In the pairwise experiments, the best-source market is picked based on the validation set performance. We report nDCG@10 on the test set in all results, with significance tests using a paired two-sided t-test with the Bonferroni correction. While we report only nDCG@10, we note that we observed similar trends for HR@10.

**Compared methods.** Market-aware models are denoted with an ‘MA-’ prefix, and are compared with the following models:

- Single-market models: These are models trained only on the target market data without any auxiliary source data, see Sect. 3.1. We train all three models GMF, NMF and MLP.
- Cross-market models: In addition to target market data, these models are trained with either one source market (for pairwise experiments), or all source markets (for global experiments). Models trained with at least one source market have a ‘++’ suffix e.g. GMF++ and MA-GMF++.
- Meta-learning models (see Sect. 3.2) similarly utilise data from one or more auxiliary markets:
  - MAML [2,8]: These are models trained using MAML, with weights initialised from a trained NMF++ model [2].
  - FOREC [2]: This model uses the trained MAML model to first freeze certain parts of the network, followed by a fine-tuning step on the target market.

**Model hyperparameters.** We set model parameters from [2]: the dimensionality of the user, item, and market embeddings are set to 8, with a 3-Layer network for MLP/NMF models. For MAML models, we set the fast learning rate $\beta = 0.1$ with 20 shots.

**Training.** All models are trained for 25 epochs using the Adam optimiser with a batch size of 1024. We use learning rates from [2], for GMF we use 0.005, for MLP and NMF we use 0.01. All models also utilise an L-2 regularisation loss with $\lambda = 1e - 7$. The NMF model is initialised with weights from trained GMF and MLP models. MAML models are trained on top of the resulting NMF models, and FOREC models utilise the trained MAML models for the fork-and-fine-tune procedure [2]. MA variants use the same hyperparameters as the market-unaware models. The objective function for all models is binary cross-entropy, given positive items and 4 sampled negatives [2,11]. For pairwise experiments, data from the source market is (randomly) down-sampled to the target market [2], which ensures that models are comparable across different-sized source markets. For global models, all data is concatenated together without any down-sampling.

---

2 [https://github.com/hamedrab/FOREC](https://github.com/hamedrab/FOREC).

3 We observed that this greatly improved performance for almost all markets.
Table 1. AVG results: Models are first trained on a single target-source pair and performance across sources are averaged. We report the nDCG@10 on the test set, with best performance in **bold**. Significance test \((p < \frac{0.05}{9})\) results are also reported comparing MA models with market-unaware (†), MAML (*) and FOREC (+).

<table>
<thead>
<tr>
<th>Method</th>
<th>de</th>
<th>jp</th>
<th>in</th>
<th>fr</th>
<th>ca</th>
<th>mx</th>
<th>uk</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMF++</td>
<td>0.2045</td>
<td>0.0916</td>
<td>0.1891</td>
<td>0.2026</td>
<td>0.1937</td>
<td>0.4204</td>
<td>0.3222</td>
</tr>
<tr>
<td>MA-GMF++</td>
<td>0.2148(^{†})</td>
<td>0.1079</td>
<td>0.2013</td>
<td>0.2022</td>
<td>0.2203(^{†})</td>
<td>0.4283(^{†})</td>
<td>0.3327(^{†})</td>
</tr>
<tr>
<td>MLP++</td>
<td>0.2836</td>
<td>0.1653</td>
<td>0.4376</td>
<td>0.2704</td>
<td>0.2905</td>
<td>0.5274</td>
<td>0.4346</td>
</tr>
<tr>
<td>MA-MLP++</td>
<td>0.2909(^{†+})</td>
<td>0.1741</td>
<td><strong>0.4502</strong></td>
<td>0.2805(^{†})</td>
<td><strong>0.3073(^{†+})</strong></td>
<td>0.5311</td>
<td>0.4349(^∗)</td>
</tr>
<tr>
<td>NMF++</td>
<td>0.2927</td>
<td><strong>0.1826</strong></td>
<td>0.4403</td>
<td>0.2844</td>
<td>0.2844</td>
<td>0.5367</td>
<td><strong>0.4379</strong></td>
</tr>
<tr>
<td>MA-NMF++</td>
<td><strong>0.3055(^{†+})</strong></td>
<td>0.1824</td>
<td>0.4471</td>
<td><strong>0.2893(^{†+})</strong></td>
<td>0.3002(^{†+})</td>
<td><strong>0.5387(^{†+})</strong></td>
<td>0.4370(^∗)</td>
</tr>
<tr>
<td>MAML</td>
<td>0.2808</td>
<td>0.1770</td>
<td>0.4320</td>
<td>0.2785</td>
<td>0.2794</td>
<td>0.5288</td>
<td>0.4296</td>
</tr>
<tr>
<td>FOREC</td>
<td>0.2835</td>
<td>0.1758</td>
<td>0.4345</td>
<td>0.2816</td>
<td>0.2772</td>
<td>0.5302</td>
<td>0.4330</td>
</tr>
</tbody>
</table>

**Pairwise Experiments.** The first set of experiments dealing with RQ1 and RQ2, which we call **pairwise** (Sect. 5.1), assumes a single auxiliary market is available for a given target market. Since there are multiple source markets, we report both the average performance in the target market **across source markets** — termed AVG — as well as performance in the target market using the **best source market**, termed BST. The two tables relay different results: the average performance indicates the expected performance of a method since the ‘best’ source market might be unknown, or only a single source may exist; whereas the best-source results are indicative of the maximum achievable performance if a good source market is already known (this is typically unknown [24]).

**Global Experiments.** The second set of experiments corresponding to RQ3 utilises data from multiple auxiliary markets at once to train a **global** recommender, with the goal to achieve good performance for all markets. We term these experiments **Global** (Sect. 5.2). We describe the results of the two sets of experiments in the following section.

5 Results and Discussion

5.1 Pairwise Experiments

Tables 1 and 2 report the results of the **pairwise** experiments, where the models only use one auxiliary market at a time. We report both AVG , the average performance of models using different auxiliary markets for the same target market (Table 1), as well as BST , the best auxiliary market (Table 2). The best auxiliary market is determined based on the validation set performance. Moreover, the results of the single-market baseline models are only reported in Table 2. We first examine RQ1, comparing the performance of MA models against baselines in both the AVG and BST settings. We end with discussion of RQ2, which compares training times across models.
Do MA models improve over market unaware models on average? Using Table 1, we first examine if MA models outperform market-unaware models in the AVG setting e.g. GMF+++ against MA-GMF+++. We see that the MA-GMF+++ outperforms GMF+++ for every market except fr. MA-MLP+++ outperforms MLP+++ for all markets, and MA-NMF+++ outperforms NMF+++ on all markets except jp and uk. For the de and ca markets, we see that MA models always outperform their non-MA variant. In addition, for the uk and mx markets, MA-GMF+++ significantly outperforms GMF+++; and for fr we see that MA-MLP+++ significantly outperforms MLP+++. Despite large improvements in some markets e.g. MA-MLP+++ improves nDCG@10 by 0.12 points over MLP+++ for in, we do not see a significant result, which may be due to the conservative Bonferroni correction, or fewer test users for in (requiring larger effect sizes). Overall, MA models outperform their market unaware equivalent in 18 of 21 settings. In summary, we can conclude that in the AVG setting, the proposed market-aware models outperform market-unaware baselines for nearly all markets. This demonstrates the robustness of MA models since these improvements are across multiple source markets.

How do MA models compare against meta-learning models in the AVG setting? We compare MA models against MAML and FOREC considering AVG, in Table 1. MA-GMF+++ never outperforms MAML/FOREC, but the differences in model sizes render this comparison unfair. A fairer comparison would be with MA-NMF+++; we see that it outperforms MAML for 5 of 7 markets: de, fr, ca, mx and uk. Additionally, FOREC is significantly outperformed by MA-NMF+++ for 4 of 7 markets: de, fr, ca and mx. We note, however, that at least one MA model outperforms both MAML/FOREC for all markets, and at least one MA model significantly outperforms MAML/FOREC for de (both), fr (both), ca (both), mx (MAML only) and uk (MAML only). Therefore, we can thus conclude that market-aware models either match or outperform meta-learning models for many markets in AVG setting.

Do MA models outperform market-unaware models when trained with the best available source? Viewing Table 2, we first note that MA models outperform all single market variants, highlighting the utility of selecting a good source market, consistent with prior research [2,24]. MA models significantly outperform single-market variants depending on the market and model, with more significant improvements seen for MA-GMF+++ (5 of 7 markets) than MA-MLP+++ (3 of 7) or MA-NMF+++ (3 of 7). Consistent improvements over the single-market models are surprisingly seen for some larger markets i.e. ca and de (but not for uk), showing larger markets can sometimes benefit from auxiliary market data. However, the results are less consistent when comparing the MA models with their augmented but market-unaware models, especially as model size increases. MA-GMF+++ improves over GMF+++ in 4 of 7 markets, MA-MLP+++ improves over MLP+++ in 3 of 7 markets, and finally, MA-NMF+++ improves over NMF+++ only in 2 markets. In fact, for in, fr, mx and uk, we see that NMF+++ outperforms MA-NMF+++. Furthermore, only MA-NMF+++ on ca
**Table 2. BST:** Models are trained on all source markets, the best source is selected based on validation set performance. We report nDCG@10 on the test set, along with significance test results ($p < \frac{0.05}{12}$) comparing MA models with single market (†), market unaware (‡), MAML (∗) and FOREC (+).

<table>
<thead>
<tr>
<th>Method</th>
<th>de</th>
<th>jp</th>
<th>in</th>
<th>fr</th>
<th>ca</th>
<th>mx</th>
<th>uk</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMF</td>
<td>0.2574</td>
<td>0.0823</td>
<td>0.0511</td>
<td>0.2502</td>
<td>0.2566</td>
<td>0.5066</td>
<td>0.4136</td>
</tr>
<tr>
<td>GMF++</td>
<td>0.2670</td>
<td>0.1093</td>
<td>0.2838</td>
<td>0.2708</td>
<td>0.2818</td>
<td>0.5338</td>
<td>0.4399</td>
</tr>
<tr>
<td>MA-GMF++</td>
<td>0.2831†</td>
<td>0.1453†</td>
<td>0.3338†</td>
<td>0.2654†</td>
<td>0.2907†</td>
<td>0.5145</td>
<td>0.4336†</td>
</tr>
<tr>
<td>MLP</td>
<td>0.2986</td>
<td>0.1340</td>
<td>0.4506</td>
<td>0.2869</td>
<td>0.2934</td>
<td>0.5367</td>
<td>0.4465</td>
</tr>
<tr>
<td>MLP++</td>
<td>0.3170</td>
<td>0.1865</td>
<td>0.4470</td>
<td>0.3016</td>
<td>0.3100</td>
<td>0.5455</td>
<td>0.4585</td>
</tr>
<tr>
<td>MA-MLP++</td>
<td>0.3167†</td>
<td>0.1806†</td>
<td>0.4584†</td>
<td>0.3026†</td>
<td>0.3105†++</td>
<td>0.5419</td>
<td>0.4544</td>
</tr>
<tr>
<td>NMF</td>
<td>0.3214</td>
<td>0.1717</td>
<td>0.4265</td>
<td>0.3014</td>
<td>0.2848</td>
<td>0.5430</td>
<td>0.4488</td>
</tr>
<tr>
<td>NMF++</td>
<td>0.3332</td>
<td>0.1921</td>
<td>0.4595</td>
<td>0.3271</td>
<td>0.3008</td>
<td>0.5590</td>
<td>0.4702</td>
</tr>
<tr>
<td>MA-NMF++</td>
<td><strong>0.3415†++</strong></td>
<td>0.1896</td>
<td>0.4433</td>
<td>0.3228†</td>
<td><strong>0.3158†++</strong></td>
<td>0.5573</td>
<td>0.4578</td>
</tr>
<tr>
<td>MAML</td>
<td>0.3168</td>
<td><strong>0.2083</strong></td>
<td>0.4491</td>
<td>0.3152</td>
<td>0.2989</td>
<td>0.5463</td>
<td>0.4671</td>
</tr>
<tr>
<td>FOREC</td>
<td>0.3040</td>
<td>0.1983</td>
<td>0.4458</td>
<td>0.3191</td>
<td>0.2927</td>
<td>0.5442</td>
<td>0.4683</td>
</tr>
</tbody>
</table>

Significantly outperforms NMF++. We can thus conclude that while MA models improve over market unaware models in some cases, selecting a source market remains an important factor for improving performance given a target market. While this conclusion holds, we note that in general, data from multiple source markets may be unavailable, or otherwise data from target markets might be unavailable—making best source selection unviable [24]. In such cases, results from the average-source experiments have to be considered.

**How Do MA models compare against meta-learning models when trained on the best source?** We now compare MA models against MAML/FOREC. We first note that at least one MA model beats MAML/FOREC for all markets but **jp** and **uk**: MA-NMF++, in particular, outperforms both MAML and FOREC for 4 of 7 markets. We see MA-NMF++ significantly outperforms both MAML/FOREC for **de** and **ca**. MAML achieves the best performance for **jp**, beating other models by a large margin. In conclusion, we observe similar performance of our MA models compared to meta-learning models, while outperforming them in some cases. This again indicates the effectiveness of our market embedding layer, especially when the training times are considered, which we discuss next.

**How do training times compare across models? Are MA models time-efficient?** We plot the time taken to train all models for a given target market (distributed across the seven different source markets) in Fig. 1, where the time taken is on a log scale. From this, we can see that the meta-learning models take far longer to train compared to MA models. We note that MA models
Fig. 1. Time taken to train a model for a target market across all source markets, where time is on a log scale. MA and market-unaware models require similar training times, while meta-learning models require significantly more.

require only 15% of the time taken to train meta-learning models, with MA models requiring about the same time to train as market-unaware models. This is unsurprising, since MAML requires an inner and outer loop, as well as requiring the expensive computation of second-order derivatives [1,8]. FOREC uses MAML in addition to fine-tuning the target market, so training FOREC takes up even more training time. In conclusion, MA models achieve better or similar performance to MAML/FOREC while requiring much less training time.

Discussion. Overall, we can conclude for AVG that MA models outperform both market-unaware baselines as well as meta-learning models, demonstrating the effectiveness of MA models across multiple sources. For BST i.e. when best-source selection is viable, the results are mixed: MA models always outperform single model variants; they outperform market-unaware models for many, but not all, markets; and an MA model either matches or outperforms meta-learning models for all markets.

A fair question to ask is whether an increase in performance of MA over market-unaware models can be attributed to the increase in the number of parameters from the market embeddings. However, this increase is minuscule compared to model sizes, especially for NMF and MLP i.e. for $t$ markets and $D$ dimensional user/item/market embeddings, the increase is just $tD$ parameters. In the pairwise experiments, this difference is just $16 (= 2 \times 8)$, much fewer than 19929, the number of parameters of a MLP model for the smallest target/source pair (in/jp).

While meta-learning models implicitly model the market during training, MA models show that this may be insufficient. We attribute the success of MA models to this explicit modelling of the markets: by adapting item representations depending on the market, the model may be better able to distinguish between recommendation in different markets more than market-unaware and meta-learning models. As we observe a better performance on AVG, we can conclude for RQ1, that market-aware models exhibit a more robust performance
Table 3. Global experiments: All markets are trained in unison. Best model for a market is in **bold**. Significance test ($p < 0.05$) results are also reported comparing MA models with market unaware (†), MAML (∗) and FOREC (+).

<table>
<thead>
<tr>
<th>Method</th>
<th>de</th>
<th>jp</th>
<th>in</th>
<th>fr</th>
<th>ca</th>
<th>mx</th>
<th>uk</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMF++</td>
<td>0.3166</td>
<td>0.1781</td>
<td>0.4535</td>
<td>0.2884</td>
<td>0.2921</td>
<td>0.5245</td>
<td>0.4481</td>
</tr>
<tr>
<td>MA-GMF++</td>
<td>0.3073</td>
<td>0.1817</td>
<td>0.4554</td>
<td>0.2836</td>
<td>0.3015</td>
<td>0.5262</td>
<td>0.4504</td>
</tr>
<tr>
<td>MLP++</td>
<td>0.3268</td>
<td>0.2127</td>
<td>0.4479</td>
<td>0.2953</td>
<td>0.3048</td>
<td>0.5376</td>
<td>0.4491</td>
</tr>
<tr>
<td>MA-MLP++</td>
<td>0.3158</td>
<td>0.2195</td>
<td>0.4398</td>
<td>0.2958</td>
<td><strong>0.3178</strong></td>
<td>0.5258</td>
<td>0.4535</td>
</tr>
<tr>
<td>NMF++</td>
<td>0.3262</td>
<td>0.1930</td>
<td><strong>0.4796</strong></td>
<td>0.3030</td>
<td>0.2851</td>
<td>0.5340</td>
<td>0.4476</td>
</tr>
<tr>
<td>MA-NMF++</td>
<td><strong>0.3442</strong></td>
<td><strong>0.2212</strong></td>
<td>0.4602</td>
<td><strong>0.3052</strong></td>
<td><strong>0.3112</strong></td>
<td><strong>0.5536</strong></td>
<td><strong>0.4604</strong></td>
</tr>
<tr>
<td>MAML</td>
<td>0.3281</td>
<td>0.1860</td>
<td>0.4736</td>
<td>0.3022</td>
<td>0.2836</td>
<td>0.5317</td>
<td>0.4474</td>
</tr>
<tr>
<td>FOREC</td>
<td>0.3249</td>
<td>0.1956</td>
<td>0.4778</td>
<td>0.3033</td>
<td>0.2947</td>
<td>0.5409</td>
<td>0.4474</td>
</tr>
</tbody>
</table>

compared to other models either matching or outperforming baselines in many settings. While this indicates that market-aware models are more effective models in general, in some cases meta-learning models seem to learn better from the most suitable market: in these cases, MA models achieve similar performance. However, it is critical to note that MA models achieve this while requiring far less computational power. Moreover, it is evident that MA models do not add much complexity to non-MA models, while empowering the model to capture the market’s attributes more effectively, resulting in an efficient and effective model.

5.2 Global Experiments

Table 3 reports the results of training one global recommendation model for all markets. We see that MA models outperform baselines in many cases, even beating meta-learning models for almost all markets.

How do MA models compare with market-unaware models? MA-variant models outperform market-unaware models in 15 of 21 settings, but results differ across models: MA-GMF++ (5 of 7), MA-MLP++ (4 of 7) and MA-NMF++ (6 of 7). MA-MLP++ significantly outperforms MLP++ for **ca** whereas MA-NMF++ significantly outperforms NMF++ for four markets. We also note that MA models for the largest markets, **uk** and **ca**, outperform both market-unaware and meta-learning models. We observe mixed results for smaller markets: for **jp**, MA consistently improves over market-unaware variants, but for **in**, only MA-GMF++ outperforms GMF++. Overall, we can conclude that MA models outperform market-aware models in several settings, especially for larger markets and models.

How do MA models compare with meta-learning-based models? We first note that an MA model (typically MA-NMF++) beats MAML/FOREC...
for all markets except \textit{in}. Indeed, MA-NMF++ beats both MAML and FOREC for all markets except \textit{in}. It \textit{significantly} outperforms MAML for \textit{ca}, \textit{mx} and \textit{uk} markets, and FOREC for \textit{de}, \textit{ca} and \textit{uk}—the larger markets. For \textit{ca}, we see all three MA models significantly outperform MAML, with MA-MLP++ and MA-GMF++ significantly outperforming FOREC. On the whole, we see that in a global setting, MA models outperform meta-learning methods in nearly all markets, and in particular the larger markets.

\textbf{Discussion.} We can conclude for RQ3 that MA models are more suitable than market unaware or meta-learning models if a global model is used for recommendation across all markets. This is critical for cases where various markets exist, empowering the model to take advantage of various user behaviours across different markets to improve recommendation in the target market. Moreover, it also leaves the problem of selecting the ‘best source’ to the model (i.e. the market embedding), as the model consumes the whole data and synthesises knowledge from multiple markets. MA models seem to have an advantage over market-unaware and meta-learning models, especially for larger markets. This is likely due to the market embedding, allowing markets to distinguish source- and target-market behaviours. As more data is collected, MA models, which perform better in the global setting for larger markets, are likely to have a clear advantage.

6 Conclusions and Future Work

In this work, we proposed simple yet effective MA models for the CMR task. In a \textit{pairwise} setting where models are trained with a single source market, MA models on average outperform baselines in most settings, showcasing their robustness. Considering the best source market, we showed that MA models match or outperform baselines for many markets. We showed that they require far less time to train compared to meta-learning models. Next, we trained a global model for all markets and showed that MA models match or outperform market-unaware models in nearly all settings, and outperform meta-learning models for all but one market. For future work, we plan to experiment with more complex MA models in a limited data setting. We also plan to investigate the utility of MA models in a \textit{zero-shot} setting, substituting the market-embedding of the new market with a similar market. In addition, we want to consider data selection techniques, since we speculate that not all data from a given source market will be useful for a given target market.

\textbf{Acknowledgements.} The authors would like to thank Hamed Bonab for help with code and data, and Ruben van Heusden for help with statistical testing. The authors also thank the reviewers for their valuable feedback. This research was supported by the NWO Innovational Research Incentives Scheme Vidi (016.Vidi.189.039), the NWO Smart Culture - Big Data/Digital Humanities (314-99-301), and the H2020-EU.3.4. - SOCIETAL CHALLENGES - Smart, Green And Integrated Transport (814961). All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.
References


