2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) | 978-1-7281-8470-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICMLA51294.2020.00059

TransCrossCF: Transition-based Cross-Domain Collaborative Filtering

Thanh-Nam Doan[§]
Center for Urban Informatics and Progress
University of Tennessee at Chattanooga
Chattanooga, TN 37403 USA
thanh-nam-doan@utc.edu

Shaghayegh Sahebi
Department of Computer Science
University at Albany, SUNY
Albany, NY 12222 USA
ssahebi@albany.edu

***Abstract*—The success of cross-domain recommender systems in capturing user interests across multiple domains has recently brought much attention to them. These recommender systems aim to improve the quality of suggestions and defy the cold-start problem by transferring information from one (or more) source domain(s) to a target domain. However, most cross-domain recommenders ignore the sequential information in user history. They only rely on an aggregate or snapshot of user feedback in the past. Most importantly, they do not explicitly model how users transition from one domain to another domain as users continue to interact with different item domains. In this paper, we argue that between-domain transitions in user sequences are useful in improving recommendation quality, dealing with the cold-start problem, and revealing interesting aspects of how user interests transform from one domain to another. We propose *TransCrossCF*, transition-based cross-domain collaborative filtering, that can capture both within and between domain transitions of user feedback sequences while understanding the relationship between different item types in different domains. Specifically, we model each purchase of a user as a transition from his/her previous item to the next one, under the effect of item domains and user preferences. Our intensive experiments demonstrate that *TransCrossCF* outperforms the state-of-the-art methods in recommendation task on three real-world datasets, both in the cold-start and hot-start scenarios. Moreover, according to our context analysis evaluations, the between-domain relations captured by *TransCrossCF* are interpretable and intuitive.**

***Index Terms*—recommendation system, cross-domain recommendation, collaborative filtering**

I. Introduction

With the rapid growth of online services, users are overwhelmed by the number of choices they can make (the "choice overload" phenomenon [1]). As a solution to alleviate this problem, recommender systems are increasingly used as an essential tool in such services. Recently, the abundance of user data collected from multiple systems and domains has led to an increased interest in cross-domain recommender systems. Originally, cross-domain recommender systems were introduced as a solution to the cold-start problem [2]. These recommender systems improve their suggestions by transferring information from one or more, typically dense, *source* or *auxiliary* domains to a *target* domain [3]. Particularly, recent researches have shown significant improvements by cross-domain recommender algorithms that transfer information from one item domain (e.g., books) to another item domain (e.g., music) [4]–[10]. Most of the current research in cross-domain recommenders focus on collaborative filtering cross-domain approaches. Many of these algorithms jointly model multiple domains by sharing common user's latent representations across them.

Similarly, sequential and time-based recommender systems have shown success in capturing the importance of succession in user interactions. Many of these recommenders are modeled as matrix or tensor factorizations [11], [12], translation-based models [13], or deep recurrent neural networks [14], [15].

As these cross-domain and sequential recommender systems have grown independently, how to represent user sequences in cross-domain recommendations has been less explored. The few recent works in this area have largely relied on modeling user sequences separately in each domain and using factorization-based modeling to transfer information [7], [10]. In this paper, we argue that items that appear consecutively in a user's sequence are tightly related to each other. Furthermore, there is an inherent relationship between items of different domains, especially for the ones that appear right after another in user sequences. Therefore, the sequence of user interactions is important to be modeled in cross-domain recommenders. Especially, user transition across domains can provide useful information in capturing how user interests map from one domain to another one. For example, a particular user is more likely to purchase Halloween-themed party supplies after purchasing a Halloween costume from the clothing domain.

In Figure 1, we illustrate another example of this domain transition in user purchase sequence. This user starts with reading the *Goblet of Fire* from *Harry Potter* book series. Then, the user continues to read the next *Harry Potter* books as they get published sequentially (within-domain sequence). In addition to that, the user watches movie adaptations of these books after reading them. Here is where the domain switches or between-domain transitions happen. In this paper, we bridge between cross-domain and sequential recommender systems literature by modeling both within-domain and between-domain transitions of user sequences. We demonstrate that modeling such transitions helps in improving the predicted recommendations, in both hot and cold-start settings.

[§]Most of the work was done when the first author was at University at Albany, SUNY

978-1-7281-8470-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ICMLA51294.2020.00059

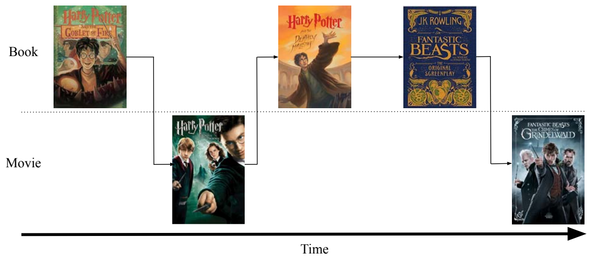

Fig. 1. Example of purchase sequence of a user.

We propose a new transition-based cross-domain recommendation model, TransCrossCF, which considers the sequence of user interactions as well as item domains. Specifically, our model assumes Markov property for user interactions with items. We model the user decision to select a new item as a transition from the previous selected item's latent space to the new item's latent space, under the operation of user preferences and item domains. To capture the information transfer across different domains, we model each domain with its own latent space and represent the transition with latent space mappings.

In summary, the contributions of our work is as follows

- We present TransCrossCF, our transition-based method, to capture the sequence of user interactions across different domains. Moreover, our model does not require the item and user embedding vectors to share the same latent feature space.
- Through our experiment, we demonstrate that our model outperforms other state of the art baselines in different experiments, both in hot and cold-start settings.
- Our in-depth analysis shows that TransCrossCF is able to capture meaningful information transfer patterns across different domains.

The structure of our paper is organized as follows. The next section will briefly review the related literature. The following two sections describe our proposed model and show its performance through extensive experiments and analyses. The last section concludes the paper and provides some future directions to extend our research further.

II. Related Work

In this section, we provide a brief overview of related works. We divide these works into the literature on cross-domain recommendations and sequential recommendation systems.

A. Cross-Domain Recommendations

Cross-domain recommendation is a branch of recommender systems, with the purpose of learning user preferences from data across multiple domains [16]. These recommender systems are increasingly gaining the researchers' attention, especially as a solution for the cold-start problem [7], [10], [17], [18]. There are two main types of cross-domain recommenders: collaborative filtering [4], [19], and content-based methods [20]. In this work, we focus on collaborative filtering cross domain recommendations. Similar to single-domain collaborative filtering, research works on cross domain recommendation usually use matrix factorization methods.

For example, Pan et al. [21] propose a cross domain recommendation system based on matrix factorization by using a coordinate system transfer method. Hu et al. [22] capture the triadic relation between user-item-domain as a tensor. Their algorithm to factorize the tensor can outperform the state-of-the-art baselines in rating and ranking prediction. Elkahky et al. [19] use a deep learning framework to improve the performance of cross domain recommendation and also provide a scalable method to handle large datasets.

Despite the fast-growing literature on cross-domain recommenders, most of the previous works do not take into account users' sequential purchases explicitly. In this paper, we bring the sequential cross-domain modeling to the field of cross-domain recommendation.

B. Sequential Recommendation Systems

Sequential recommendation systems have been researched for a long time. These models mainly rely on two main techniques: matrix factorization and recurrent neural network. Koren [11] proposed one of the first matrix factorization techniques to capture the sequence of users' interactions. Specifically, the author models the temporal dynamics of users and item over time to capture the long-term trends under factorization method. Other works [23], [24] achieve better performance than the work of Koren [11] due to integration of Markov condition. Eskandanian and Mobasher [25] proposed Hidden-Markov Model to capture the change point within the sequence of user activities which indicates the significant shift of user preferences.

Recently, because of the growth in deep neural networks, recurrent neural networks (RNN) have been used in recommendation systems [26], [27]. These works divide time into periods and model the temporal dynamics of users and items by recurrent neural network, then, calculate the rating prediction of users on items within the period. Tang and Wang [28] combine RNN with convolution neural networks to model the sequential activities of users and their evaluation shows that it yields a good performance.

From the success of transition-based models in knowledge discovery, He *et. al.* [13] proposed TransRec which integrates the idea of knowledge discovery in recommendation systems. Specifically, it assumes that the purchase of a user to an item is the transition to the latent feature of this item from the previous purchase of the user under his/her latent preference.

These works mostly focus on single domain recommendations, losing important information that can happen across domains. To address this, we propose a new cross-domain recommendation systems based on transition-based method that focuses on capturing sequential activities of users across different item domains.

III. Proposed Framework

In this section, we formulate our problem and describe our proposed framework in detail. For the ease of reading, Table I

TABLE I
TABLE OF NOTATIONS

Notation	Meaning
U/u	Set of users/a user
d	A particular domain
I_d/i_d	Set of items/an item in domain d
S_k^u	k-th interaction item of user u
γ_{i_d}	Latent feature vector of item i in domain d
t_u	Latent feature vector of user u
$W_{\to d}$	Linear mapping from user space to item space of domain d
$W_{d_1 \to d_2}$	Linear mapping from item space of domain d_1 to that of domain d_2
β_i	Latent bias of item i
$dist(\cdot,\cdot)$	Distance between two latent vectors
$p_{u,i,j}$	Probability that user u interacts with item i then item j

summarizes our notations used in this section. The parameter learning process is also discussed at the end of the section.

Problem Formulation. We would like to recommend the next interesting item to a user, given his/her past interaction history in different item domains. We denote the set of users as U and the set of items that belong to the domain d as I_d. The "next item" prediction task is defined as follows. For each user $u \in U$, the sequence of their interactions with items is represented as $S^u = (S_1^u, S_2^u, ..., S_{|S^u|}^u)$. Note that the k-th interaction item of user u, i.e. S_k^u, is a tuple of item index and the domain which this item belongs to. Given all users' sequential interactions, $S = \{S^{u_1}, S^{u_2}, ..., S^{u_{|U|}}\}$, we want to predict the next interaction item of each user. Notation I represents all items, i.e. $I = \bigcup_d I_d$, where d denotes the domain index.

TransCrossCF. We develop TransCrossCF according to the following main assumptions: i) items that appear right after each other in a user's sequence are tightly related to each other; ii) this relationship is built according to the specific user's interests; iii) user interests can be translated into item features; and iv) there is an inherent relationship between items of different domains, especially for the ones that appear consecutively in user sequences. TransCrossCF is designed to capture the transitions between items in the same and different domains in user sequences.

We model each item i_d of domain d with a latent feature vector $\gamma_{i_d} \in \Phi_d$ where Φ_d is the lower-dimensional item space for items of domain d. Similarly, we represent each user u with a latent feature vector $t_u \in \Omega$ where Ω is the lower-dimensional user space. Note that, unlike traditional collaborative filtering models, we allow the user and item spaces to be different. This provides the flexibility for items and users to not to have the exact same number and set of latent features. Instead, we provide a mechanism to map these two latent spaces, using a linear mapping $W_{\to d} : \Omega \to \Phi_d$ (assumption "iii" above).

To model the similarity between two consecutive items i and j in user u's interaction sequence, according to assumption "i", we expect γ_j to be similar to γ_i in space Φ regarding some distance metric such as $\mathcal{L}_2$. Inspired by [13], we model this similarity to be affected by user u's specific interests (assumption "ii"). Particularly, we expect the transition between γ_j and γ_i to be a function of t_u. In case i and j are both from the same domain d, using the mapping between user and item latent spaces, we have: $\gamma_j \approx \gamma_i + W_{\to d} t_u$.

Accordingly, in case u only interacts with items of one domain d, the probability that u transitions from previous item i to the next item j is given by:

$$\begin{aligned} p_{u,i,j} \propto \beta_j - dist(\gamma_i, \gamma_j | W_{\to d}, t_u) \\ \text{subject to } \gamma_i \in \Psi \subseteq \Phi_d, \text{ for all } i \end{aligned} \tag{1}$$

where Ψ is the subspace of Φ_d e.g. a unit ball. This constraint has been used in previous works to solve "curse of dimensionality" [29]–[31]. The notation $dist(\gamma_i, \gamma_j | W_{\to d}, t_u)$ denotes the distance between the two item embedding vectors γ_i and γ_j under the transition of user preference t_u and user-item latent space mapping $W_{\to d}$. For example, if $\mathcal{L}_2$ is employed, $dist(\gamma_i, \gamma_j | W_{\to d}, t_u) = \|\gamma_i + W_{\to d} t_u - \gamma_j\|^2$. Parameter β_j is a learned bias of item j, that can represent the item's popularity vector. The less the distance between i and j, the more likely it is for u to transition to j from i.

The above transition model assumes that the all items in user sequence belong to the same domain. To add a mechanism to represent multi-domain user sequences, we propose to capture the item transitions between different domains as a linear mapping between latent spaces of those domains (assumption "iv"). Suppose that in user u's interaction sequence, item j from domain d_j happens right after item i in domain d_i. When modeling the similarity between latent factor representations of these two items (γ_i and γ_j), two scenarios are possible:

- If $d_i = d_j$ (the two consecutive items are from the same domain), we approximate the item embedding of j by the transition from i under the latent feature of u i.e. $\gamma_i + W_{\to d_i} t_u \approx \gamma_j$. The distance between previous and current purchase of user under $\mathcal{L}_2$ distance is $dist(\gamma_i, \gamma_j | W_{\to d_i}, t_u) = \|\gamma_i + W_{\to d_i} t_u - \gamma_j\|^2$
- If $d_i \neq d_j$ (the two consecutive items are from different domains), we use a linear mapping $W_{d_i \to d_j} : \Phi_{d_i} \to \Phi_{d_j}$ to capture the information transfer from domain d_i to d_j. Therefore, we have the approximation $W_{d_i \to d_j}(\gamma_i + W_{\to d_i} t_u) \approx \gamma_j$ and the distance between two consecutive purchase of u is $dist(\gamma_i, \gamma_j | W_{\to d_i}, t_u) = \|W_{d_i \to d_j}(\gamma_i + W_{\to d_i} t_u) - \gamma_j\|^2$

We, then, plug our new definition of distance $dist(\gamma_i, \gamma_j | W_{\to d_i}, t_u)$ into the probability estimation in Eq. 1 to calculate the probability that user u transitions from item i to item j across various item domains.

Ranking Optimization. For each user and their sequential interaction history, we want to rank items such that the most likely ones happen earlier in our ranked list. Similar to Bayesian Personalized Ranking [32] framework, we choose sequential pairwise ranking to solve our optimization problem. Formally, we would like to maximize the following log likelihood function:

$$
\begin{aligned}
LLH &= \log \prod_{u \in U} \prod_{j \in S^u} \prod_{j' \in I \setminus S^u} P(j >_{u,i} j' | \Theta) P(\Theta) \\
&= \sum_{u \in U} \sum_{j \in S^u} \sum_{j' \in I \setminus S^u} \log \sigma(p_{u,i,j} - p_{u,i,j'}) - \Omega(\Theta)
\end{aligned} \tag{2}
$$

where i is the item preceding j in interaction sequence S^u; Θ is a set of all parameters. Particularly, we have $\Theta = \{t_u, \beta_j, \beta_{j'}, \gamma_i, \gamma_j, \gamma_{j'}, W_{\to d_i}, W_{d_i \to d_j},, W_{d_i \to d_{j'}}\}$; Ω is $\mathcal{L}_2$ regularization; and $\sigma(\cdot)$ denotes the Sigmoid function.

Learning the Parameters. First, we randomly initialize item and user embeddings as well as all parameters, setting the item embeddings to be unit vectors. We apply stochastic gradient ascent to maximize the objective function in Eq. 2. To do this, we randomly shuffle the users in training dataset. For each training data point, we randomly select the negative data point. Then, we calculate its gradient with respect to all parameters and update the parameters accordingly. In each iteration, we project the item embedding vectors into unit vectors. The whole process is repeated until convergence.

IV. Experiments

We evaluate TransCrossCF using three different domain pairs to answer the following four research questions:

- **RQ1.** How does TransCrossCF perform compared to other state-of-the-art single-domain and cross-domain recommender algorithms?
- **RQ2.** How does TransCrossCF perform in dealing with the cold-start problem?
- **RQ3.** Are the cross-domain relationships discovered by TransCrossCF meaningful?
- **RQ4.** How sensitive is TransCrossCF to different latent space dimensionalities in item domains?

A. Experiment Setup

Datasets. We use purchase sequences of Amazon users in three domains: Digital Music, Office Products, and Musical Instruments [33]. The datasets span from May 1996 to July 2014. We consider three cross-domain combinations of datasets: Digital Music + Musical Instrument (DM_MI), Digital Music + Office Products (DM_OP), and Musical Instrument + Office Product (MI_OP). Although our experiments are on domain-pairs, TransCrossCF can easily be applied to the combination of any number of domains. For each cross domain dataset, we select users whose purchase sequence is longer than five and items that have more than five users purchasing them (5-core processing). Table II describes the three datasets.

TABLE II
SUMMARY STATISTICS OF THE DATASET

Dataset	#users	#items	#ratings	sparsity
DM_MI	16,285	30,471	142,929	0.02%
DM_OP	22,778	38,719	202,419	0.023%
MI_OP	19,155	33,232	147,268	0.0231%

Train/Test Separation. We sort the purchase activity of each user by their timestamp. The purchase sequence of each user is divided into three test, validation, and train parts. The last purchase of each user is considered for test, second last item accounts for validation, and the rest is used for training.

Performance Measures. We employ Area Under the ROC Curve (AUC) and Mean reciprocal rank (MRR) to measure the performance of our model and the baselines. Below are the formula of these two metrics:

$$
\begin{aligned}
AUC &= \frac{1}{|U|} \sum_{u \in U} \frac{1}{|\mathcal{I} \setminus S^u|} \sum_{j' \in \mathcal{I} \setminus S^u} \mathbf{I}(R_{u,S^u_{|S^u|}} < R_{u,j'}) \\
MRR &= \frac{1}{|U|} \sum_{u \in U} \frac{1}{R_{u,S^u_{|S^u|}}}
\end{aligned} \tag{3}
$$

where $S^u_{|S^u|}$ is the latest purchased item of user u and $R_{u,i}$ indicates the predicted rank of item i for user u. Function $\mathbf{I}(\cdot)$ is an indicator function, which returns 1 if its input is true; otherwise, it returns 0. Note that the larger the value of these metrics, the better the model.

B. Baselines

For *rating prediction* and *cold-start* experiments, we compare TransCrossCF with the following baselines:

- Most Popular Item (MP): The simple method to predict items according to their popularity.
- Bayesian Personalized Ranking (BPR-MF) [32]: This model factorizes the interaction matrix of users and items based on ranking assumption, but it does not consider the sequence of users.
- Factorized Markov Chain (FMC) [24]: This method factorizes the item-to-item transition matrix of all users to find the global transition probabilities of item sequences.
- Factorized Personalized Markov Chain (FPMC) [24]: It combines factorization machines with Markov chain property to find the personalized transition probabilities of item sequences..
- Personalized Ranking Metric Embedding (PRME) [23]: This model models the Markov behaviors of users using sum of two Euclidean distances.
- Translation-based Recommendation (TransRec) [13]: The method combines user preferences and sequences with a sentence translation-based method. In this experiment, we only consider their $\mathcal{L}_2$ distance measure since it yields better results.
- Cross Domain Recommender (CDRec) [34]: The model is built upon Coupled Matrix Factorization algorithm to utilize both explicit and implicit similarities between datasets.
- Session-based Recommendation (GRU4Rec) [35]: The model was built upon neural networks. Specifically, it modified the gated recurrent unit [36] to capture the sequential property of purchase activities. In this work, we consider a purchase of each user as a session. The default values of its parameters are used in the experiment.
- Transition-based Cross-Domain Collaborative Filtering (TransCrossCF): Our method considers both sequence of

TABLE III
NEXT ITEM PREDICTION PERFORMANCE RESULTS. THE BEST PERFORMANCE IS HIGHLIGHTED. NOTATION * INDICATES A SIGNIFICANT IMPROVEMENT. THE VALUE OF MRR IS SHOWN IN PARENTHESIS.

AUC (MRR)	MP	BPRMF	FMC	FPMC	PRME	TransRec	CDRec	GRU4Rec	TransCrossCF
DM_MI	62.28%	65.81%	65.75%	66.15%	71.49%	73.89%	74.91%	73.1%	**80.04%***
	(2.78%)	(4.09%)	(4.22%)	(4.55%)	(4.77%)	(5.76%)	(5.82%)	(**6.78%**)	(6.57%)
DM_OP	64.94%	67.53%	67.76%	68.19%	72.55%	75.57%	74.95%	73.8%	**81.75%***
	(2.79%)	(4.02%)	(4.12%)	(4.67%)	(5.45%)	(6.87%)	(5.98%)	(7.02%)	(**7.25%***)
MI_OP	62.18%	65.1%	65.8%	66.32%	68.22%	70.41%	71.02%	79.03%	**80.01%***
	(2.77%)	(4.02%)	(4.97%)	(4.89%)	(7.12%)	(9.76%)	(9.88%)	(10.8%)	(**11.22%***)

users' activities and domains of items. In experiments, we use $\mathcal{L}_2$ distance.

In the first three experiments, we use the same number of latent features: 10. The learning rate for stochastic gradient ascent is set to 0.0001.

C. Next Item Prediction Experiments

In this set of experiments, we would like to evaluate the performance prediction of TransCrossCF compared to other baselines (**RQ1**). We particularly chose these baselines to cover various types of recommender algorithms: rank-based algorithms (BPR-MF), sequential algorithms (FMC, FPMC, PRME, TransRec), and cross-domain algorithms (CDRec). Among them, TransRec [13] has the most similar sequential model to our proposed model.

Table III shows the performance of TransCrossCF and all baselines. From the table, there are several observations. First, TransCrossCF outperforms all baselines in both metrics on the three datasets. For instance, its improvement is 6.8% in AUC and 12.8% in MRR compared the CDRec method on DM_MI dataset. This shows that modeling both cross-domain and sequential information produces higher-quality recommendations. Second, performance of CDRec and TransRec are consistently better than the other baselines. It indicates that considering cross domain information provides better recommendations to users. Third, the slight outperforming of CDRec over TransRec emphasizes the importance of modeling cross domain information over the order of user interactions. Fourth, GRU4Rec is deep-learning based method but its performance is lower than the one of TransCrossCF because GRU4Rec does not consider the latent features of users.

We further apply hypothesis testing to examine if the improvement of our model is actually significant over the baselines. Due to the large number of baselines, we only compare TransCrossCF with GRU4Rec method. Specifically, the *null hypothesis* is that the performances of TransCrossCF and the chosen baseline, across all users, are not statistically different and *alternative hypothesis* is that TransCrossCF is significantly better than the baseline. Paired t-test [37] is used to justify the above hypothesis. From the results in Table 3, we show that TransCrossCF is significantly better than the chosen baseline. The result is consistent in both metrics. Similar results can be obtained when we apply the same hypothesis testing for our method TransCrossCF and other baselines.

D. Cold-Start Experiments

In this section, we would like to evaluate if TransCrossCF performs well, compared to the baselines, in the cold-start setting (**RQ2**). To address this question, we analyze the algorithms' prediction performance for users whose purchase sequences are less than a specific threshold (n). We use various threshold values to explore the effect of user sequence length on recommendation performance and use the same parameter setting as rating prediction task.

Figure 2 shows the results of *TransCrossCF* and the baselines with different values of n. From the figure, we observe that *TransCrossCF* generally outperforms baselines in the three datasets under the two metrics. We also observe that TransCrossCF's performance is stable over different setting. Considering the good performance of TransCrossCF in the cold-start setting, we can conclude that TransCrossCF can capture the transfer of information between the two domains even when there is not much information on user sequences. Another observation is that in DM_MI dataset, the performance gap between TransCrossCF and CDRec is much clearer than the other two datasets. One potential reason for this is the sparsity of the DM_MI dataset, compared to the other two (see Table II). In this case, modeling the order of user actions enhances the prediction performance of cross domain recommendation systems in cold start settings. Another potential explanation is the relative closeness of the two domains in this dataset: Digital Music and Musical Instruments domains both include items that are related to music. We choose CDRec for comparison since it has the best performance among non-deep-learning methods after our method. For GRU4Rec, when we give more data i.e. increase the value of n, its performance also increases.

E. Cross-Domain Relation Analysis

In this section, we analyze if TransCrossCF can discover meaningful relations across domains, while modeling sequential transitions between them (**RQ3**). To do this, we focus on the DM_MI dataset. Similar finding could be found on the other two datasets but due to space limitation, we do not present in this section. For better interpretation of the results, we use the item *category* information that is provided for each item in Amazon dataset. Note that this information is not used in training our model and is only used to evaluate this analysis.

In this analysis, we focus on the linear between-domain mappings i.e. $W_{DM \to MI}$ and $W_{MI \to DM}$ to demonstrate the

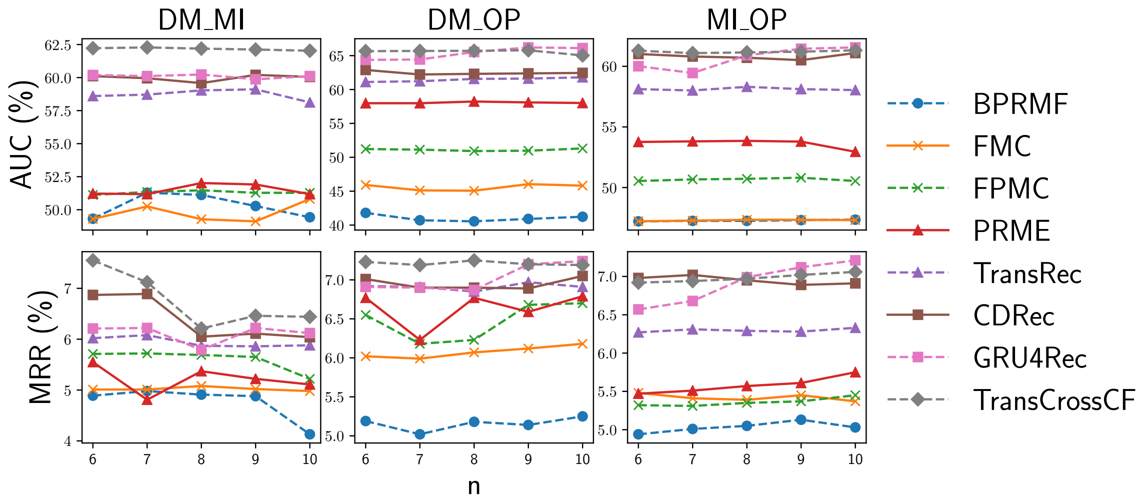

Fig. 2. Cold start performance for users whose purchase sequence length is less than n. The higher the value, the better the model is.

TABLE IV
TOPIC OF EACH ITEM LATENT FEATURE IN DM_MI DATASET. IT IS DERIVED FROM THE RESULT OF TRANSCROSSCF.

Topic	Digital Music	Musical Instrument
1	Latin Music	Studio Equipment
2	Hardcore Rock	Electronic devices
3	Alternative Rock	Accessories for music making
4	Folk + Pop	Instrument accessories
5	Southern Rock	*Unclear*
6	Rap + Hiphop	Sound device
7	Jazz	*Unclear*
8	R&B	Drum + Electronic Drum
9	Vocal Pop	*Unclear*
10	*Unclear*	Acoustic devices

relationships between latent factors of these two domains. To interpret the latent factors of each domain, we analyze item latent factors in the domains i.e. γ_i and item *category* information from the dataset. Specifically, for each latent factor f, we find its top-10 most representative items. These items (i) are those with highest values $\gamma_i[f]$. Due to ambiguity of item names, determining latent factor topics from item names is difficult. Therefore, we use the item *category* information for further analysis. This way, each latent factor's topic is determined by the most common categories of top 10 items in this latent factor. We use the topic "*Unclear*" for the cases that most common categories cannot be distinguished among the top 10 items and we found that there are four "*Unclear*" topics in DM_MI dataset.

Table IV shows the semantics of each latent factor in both domains. The left side of Figure 3 visualizes the heat-map of transfer matrix from Digital Music to Musical Instrument i.e. $W_{DM \to MI}$. The transpose of transfer matrix from Musical Instrument to Digital Music i.e. $W'_{MI \to DM}$ is also displayed on the right side of Figure 3. We use this transpose operation to group both matrices under one figure. Darker colors indicate higher values, and therefor more important transfer between topics. From the figure, we first observe that the two matrices are not symmetric. For example, transitioning from Digital Music to Musical Instruments, there is a high weight from Jazz to Acoustic Devices. But, in the reverse transition, from Acoustic Devices, Vocal Pop has a higher weight than Jazz. This shows that Acoustic Devices are highly relevant to both Jazz and Vocal Pop. However, the transfer between them is not always bi-directional. It indirectly infers that we need to explicitly model the direction of transfer from one domain to the other in cross-domain recommender systems, due to the difference in the nature of domains.

Second, from the value of transfer matrices, we notice some intuitive cases. For instance, in the transfer matrix from Digital Music to Musical Instrument, we see that there is a strong transfer from Alternative Rock to Electronic Devices which is intuitive. Also, a strong transfer from Jazz to Acoustic Devices is noticeable. For the transfer matrix from Musical Instrument to Digital Music, the strong transfers from Drum + Electronic Drums to Hardcore Rock and from Accessories for Music Making to Folk + Pop are easy to observe. On the other hand, we can see that Latin music and Electronic Devices or Alternative Rock and Acoustic Devices do not have a significant relationship in either of the two transitions.

F. Parameter Sensitivity Study

In this section, we study the impact of item latent space's dimensionality in both domains (**RQ4**). Specifically, we keep the number of user latent factors as 10 and measure the performance of TransCrossCF with different number of items' latent features in both domains. We only show the result on DM_MI dataset. Similar findings can be found in the other two datasets.

Table V shows the performance of TransCrossCF in DM_MI dataset, with various number of item latent features, in AUC

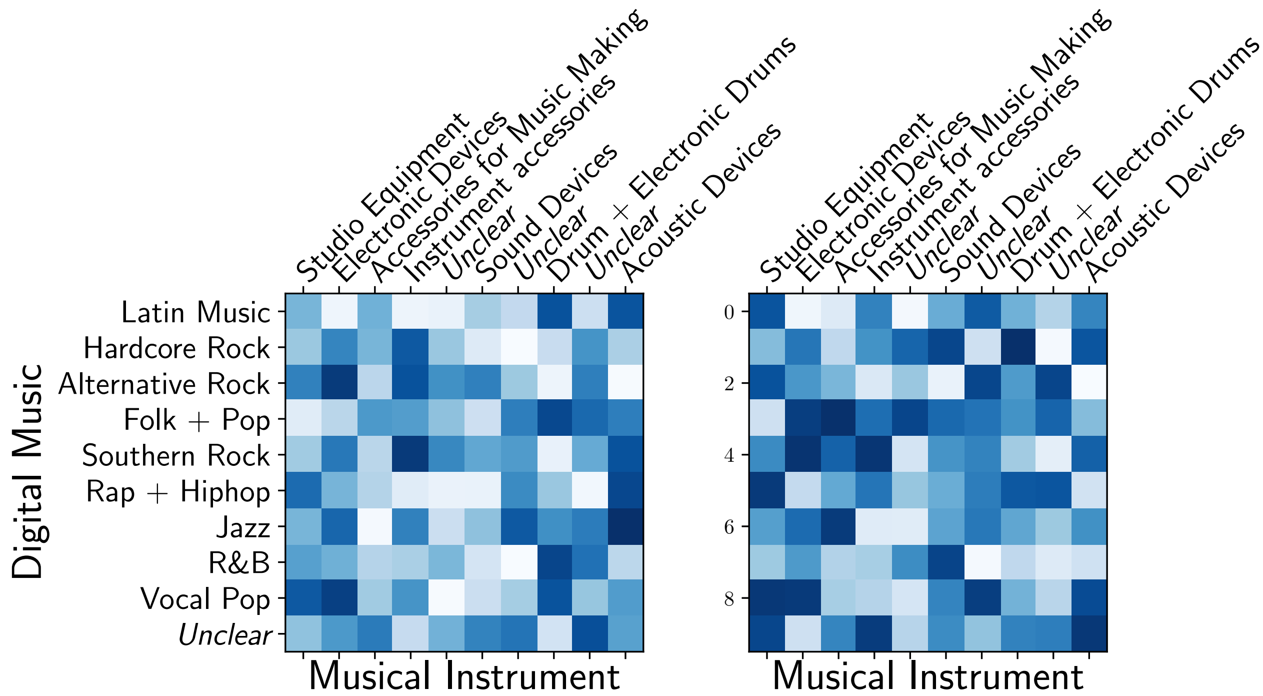

Fig. 3. Transfer matrix on DM_MI dataset. *Left:* The transfer matrix from Digital Music to Musical Instrument. *Right:* The transpose of transfer matrix from Musical Instrument to Digital Music. The darker the color, the higher the value is (i.e. the more correlated between the two corresponding topics).

TABLE V
THE PERFORMANCE OF TRANSCROSSCF IN DM_MI DATASET WITH DIFFERENT VALUES OF ITEM EMBEDDINGS IN EACH DOMAIN. THE BEST PERFORMANCE IS HIGHLIGHTED. THE VALUE OF MRR IS SHOWN IN PARENTHESIS.

AUC(MRR)	Digital Music			
Musical Instrument	5	10	15	20
5	75.82% (5.02%)	78.72% (6.15%)	79.56% (6.74%)	79.91% (6.85%)
10	77.25% (5.41%)	80.04% (6.57%)	80.86% (6.92%)	80.92% (6.96%)
15	78.12% (5.57%)	80.04% (6.62%)	80.88% (6.98%)	**81.06%** (7.02%)
20	78.35% (5.61%)	80.12% (6.58%)	80.97% (7.04%)	81.04% (**7.13%**)

and MRR metrics. From the table, we observe that increasing the number of latent feature improves the performance of TransCrossCF. For instance, when the number of item latent feature in Digital Music domain increases from 5 to 10, while keeping those for Musical Instrument as 5 AUC improves from 75.82% to 78.72%. Our second observation is that this performance improvement, that is achieved due to increase in number of item latent features, slows down as the difference in item latent dimensionalities in the two domains grows. For example, consider 5 as the number of item latent feature in Musical Instruments dataset. When we increase the corresponding value in Digital Music from 5 to 10, the AUC improvement is 3.82% but the improvement is only 0.44% when we increase this number from 15 to 20.

V. CONCLUSIONS

In this paper, we bridged between cross-domain and sequential recommender systems by proposing *TransCrossCF* a transition-based cross domain collaborative filtering model. We built this model assuming that user transitions between consecutive items, especially across different domains, reveal important information about the relationship between the items and the domains. We modeled this relationship under each user's specific interests. We designed and performed experiments on three datasets to answer four research questions. Through our experiments for next item prediction and cold-start settings, we showed that *TransCrossCF* outperforms the baselines with significant improvements. Moreover, our analysis of the transfer matrices showed the intuitive interpretation of *TransCrossCF*'s results. Finally, our parameter sensitivity study explained some connections between item embedding vectors across multiple domains.

There are several ways to extend *TransCrossCF* to enhance further. Reviews of users for items can provide more insightful information about user preference in addition to user interactions. For instance, analyzing cross-domain sequence of user reviews can potentially help in creating more accurate models. Another future direction is leveraging the social connections of users. As we know, user purchase decisions are under the influence of the choices of their friends. Therefore, incorporating users' social connections can reveal more information about user decisions [38].

VI. Acknowledgement

This paper is based upon work partially supported by the National Science Foundation under Grant No. 1755910.

References

[1] S. S. Iyengar and M. Lepper, "When choice is demotivating: Can one desire too much of a good thing?" *Journal of personality and social psychology*, vol. 79, pp. 995–1006, 01 2001.

[2] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," in *SIGIR*, 2002.

[3] I. Fernández-Tobías, I. Cantador, M. Kaminskas, and F. Ricci, "Cross-domain recommender systems: A survey of the state of the art," in *Spanish Conference on Information Retrieval*, 2012.

[4] B. Li, "Cross-domain collaborative filtering: A brief survey," in *Tools with Artificial Intelligence (ICTAI), 2011 23rd IEEE International Conference on*. IEEE, 2011, pp. 1085–1086.

[5] X. Zhang, J. Cheng, S. Qiu, Z. Zhu, and H. Lu, "When personalization meets conformity: Collective similarity based multi-domain recommendation," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2015, pp. 1019–1022.

[6] S. Sahebi and P. Brusilovsky, "It takes two to tango: An exploration of domain pairs for cross-domain collaborative filtering," in *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 2015, pp. 131–138.

[7] P. Dai, S.-S. Ho, and F. Rudzicz, "Sequential behavior prediction based on hybrid similarity and cross-user activity transfer," *Knowledge-Based Systems*, vol. 77, pp. 29–39, 2015.

[8] C. Yang, H. Yan, D. Yu, Y. Li, and D. M. Chiu, "Multi-site user behavior modeling and its application in video recommendation," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2017, pp. 175–184.

[9] A. Farseev, I. Samborskii, A. Filchenkov, and T.-S. Chua, "Cross-domain recommendation via clustering on multi-layer graphs," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2017, pp. 195–204.

[10] Y. Wang, C. Feng, C. Guo, Y. Chu, and J.-N. Hwang, "Solving the sparsity problem in recommendations via cross-domain item embedding based on co-clustering," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, ser. WSDM '19. New York, NY, USA: ACM, 2019, pp. 717–725. [Online]. Available: http://doi.acm.org/10.1145/3289600.3290973

[11] Y. Koren, "Collaborative filtering with temporal dynamics," in *KDD*, 2009.

[12] L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. G. Carbonell, "Temporal collaborative filtering with bayesian probabilistic tensor factorization," in *Proceedings of the 2010 SIAM international conference on data mining*, 2010.

[13] R. He, W.-C. Kang, and J. McAuley, "Translation-based recommendation," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, 2017.

[14] Y. Zhu, H. Li, Y. Liao, B. Wang, Z. Guan, H. Liu, and D. Cai, "What to do next: Modeling user behaviors by time-lstm." in *IJCAI*, 2017, pp. 3602–3608.

[15] A. Beutel, P. Covington, S. Jain, C. Xu, J. Li, V. Gatto, and E. H. Chi, "Latent cross: Making use of context in recurrent recommender systems," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, ser. WSDM '18, 2018.

[16] S. Sun, "A survey of multi-view machine learning," *Neural Computing and Applications*, vol. 23, no. 7-8, pp. 2031–2038, 2013.

[17] M. Enrich, M. Braunhofer, and F. Ricci, "Cold-start management with cross-domain collaborative filtering and tags," in *E-Commerce and Web Technologies*, 2013.

[18] F. Zhang, N. J. Yuan, D. Lian, and X. Xie, "Mining novelty-seeking trait across heterogeneous domains," in *Proceedings of the 23rd international conference on World wide web*. ACM, 2014, pp. 373–384.

[19] A. M. Elkahky, Y. Song, and X. He, "A multi-view deep learning approach for cross domain user modeling in recommendation systems," in *WWW*, 2015.

[20] S. Sahebi and T. Walker, "Content-based cross-domain recommendations using segmented models," in *Workshop on New Trends in Content-based Recommender Systems (CBRecsys)*. ACM, 2014, pp. 57–63.

[21] W. Pan, E. W. Xiang, N. N. Liu, and Q. Yang, "Transfer learning in collaborative filtering for sparsity reduction." in *AAAI*, vol. 10, 2010, pp. 230–235.

[22] L. Hu, J. Cao, G. Xu, L. Cao, Z. Gu, and C. Zhu, "Personalized recommendation via cross-domain triadic factorization," in *Proceedings of the 22Nd International Conference on World Wide Web*, ser. WWW '13, 2013.

[23] S. Feng, X. Li, Y. Zeng, G. Cong, Y. M. Chee, and Q. Yuan, "Personalized ranking metric embedding for next new poi recommendation," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[24] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, "Factorizing personalized markov chains for next-basket recommendation," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 811–820.

[25] F. Eskandanian and B. Mobasher, "Detecting changes in user preferences using hidden markov models for sequential recommendation tasks," *CoRR*, vol. abs/1810.00272, 2018. [Online]. Available: http://arxiv.org/abs/1810.00272

[26] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," *CoRR*, vol. abs/1511.06939, 2015.

[27] H. Jing and A. J. Smola, "Neural survival recommender," in *WSDM*, 2017.

[28] J. Tang and K. Wang, "Personalized top-n sequential recommendation via convolutional sequence embedding," *CoRR*, vol. abs/1809.07426, 2018.

[29] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Advances in neural information processing systems*, 2013, pp. 2787–2795.

[30] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Twenty-Eighth AAAI conference on artificial intelligence*, 2014.

[31] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[32] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "Bpr: Bayesian personalized ranking from implicit feedback," in *UAI*, 2009.

[33] R. He and J. McAuley, "Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering," in *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, 2016, pp. 507–517.

[34] Q. Do, W. Liu, and F. Chen, "Discovering both explicit and implicit similarities for cross-domain recommendation," in *PAKDD*, 2017.

[35] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," *arXiv preprint arXiv:1511.06939*, 2015.

[36] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

[37] H. Hsu and P. A. Lachenbruch, "Paired t test," *Wiley StatsRef: Statistics Reference Online*, 2014.

[38] T.-N. Doan and E.-P. Lim, "Pacela: A neural framework for user visitation in location-based social networks," in *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. ACM, 2018, pp. 13–21.