On Cross-Domain Transfer in Venue Recommendation

Jarana Manotumruksa[1], Dimitrios Rafailidis[2]
Craig Macdonald[3], and Iadh Ounis[3]

[1]j.manotumruksa.1@research.gla.ac.uk,[2]dimitrios.rafailidis@maastrichtuniversity.nl
[3]{firstname.secondname}@glasgow.ac.uk
University of Glasgow, UK
Maastricht University, Netherlands

Abstract. Venue recommendation strategies are built upon Collaborative Filtering techniques that rely on Matrix Factorisation (MF), to model users' preferences. Various *cross-domain* strategies have been proposed to enhance the effectiveness of MF-based models on a target domain, by transferring knowledge from a source domain. Such cross-domain recommendation strategies often require user overlap, that is common users on the different domains. However, in practice, common users across different domains may not be available. To tackle this problem, recently, several cross-domains strategies without users' overlaps have been introduced. In this paper, we investigate the performance of state-of-the-art cross-domain recommendation that do not require overlap of users for the venue recommendation task on three large Location-based Social Networks (LBSN) datasets. Moreover, in the context of cross-domain recommendation we extend a state-of-the-art sequential-based deep learning model to boost the recommendation accuracy. Our experimental results demonstrate that state-of-the-art cross-domain recommendation does not clearly contribute to the improvements of venue recommendation systems, and, further we validate this result on the latest sequential deep learning-based venue recommendation approach. Finally, for reproduction purposes we make our implementations publicly available.

Keywords: Cross-domain recommendation · Venue suggestion · Transfer learning.

1 Introduction

Location-Based Social Networks (LBSN) such as Foursquare and Yelp have become popular platforms that allow users to find interesting venues to visit based on their preferences, share their location to their friends using checkins, as well as leave comments on venues they have visited. Matrix Factorisation (MF) [1] is a popular collaborative filtering technique that is widely used to predict users' ratings/checkins on venues by leveraging explicit/implicit feedback. The major challenge in MF-based venue recommendation systems is the sparsity problem, as users can visit a very limited number of venues. Consequently, the checkin user-venue matrix in LBSNs is extremely sparse. The sparsity problem hinders the effectiveness of MF-based models. To alleviate the sparsity problem, various MF-based models have been proposed to exploit additional sources of information such as friendships and textual content of comments [2–5]. Moreover, previous studies have shown that the sequential properties of user's interactions, that is the sequences of checkins or clicks, play an important role in alleviating sparsity for tasks such as movie recommendation and venue recommendation [6–10]. Indeed, sequential-based recommendation is more challenging than traditional recommendation (e.g. rating prediction problem) because a user's previous interactions may have a strong influence on their current preferences. For example, users are more likely to visit a bar directly after they have visited a restaurant.

Apart from these additional sources of information and the sequential properties of users' interactions, recently, Cross-Domain Collaborative Filtering (CDCF) models have been proposed to alleviate the sparsity problem in a target domain [11–15]. CDCF aims to improve the quality of recommendations in a target domain by leveraging the knowledge extracted from source domains. Thus, CDCF is a form of *transfer learning*, for which the existing approaches can be divided into two categories: namely *overlapping* and *non-overlapping* CDCF. The *overlapping* CDCF models, such as [14, 16, 15, 17], transfer knowledge from a source domain to a target one based on explicit links of users/items between the domains. For example, a user who has both Twitter and Foursquare accounts, *overlapping* CDCF aims to transfer user's preferences extracted from the user's tweets on Twitter (source domain) to improve the venue recommendation on Foursquare (target domain). *Non-overlapping* CDCF models like [11–13, 18] aim to transfer useful information from the source domain to the target one without any explicit links, which is the most challenging problem in cross-domain recommendation. An example of *non-overlapping* CDCF is a newly opened book e-commerce website that would like to build a recommender system. Due to the lack of user-book interactions at the outset, the effectiveness of MF-based models for the book e-commerce website would be low, due to the sparsity problem. Since the movie domain is related to the book domain in some aspects [19] – for example comedy movies correspond to humorous books – CDCF hypothesises that the rating matrix available from a popular movie rating website would alleviate the sparsity problem for the newly opened book e-commerce site, despite there being no users or items shared between the two domains.

Most of the *non-overlapping* CDCF models [12, 13, 18] are based on a Codebook Transfer technique (CBT) proposed by Li *et al.* [11]. CBT aims to transfer rating patterns from the source domain (e.g., the popular movie rating website) to the target domain (e.g., the newly opened book website). Although several studies have shown that CBT-based models can improve recommendation effectiveness, Cremonesi and Quadrana [20] demonstrated that CBT is not able to transfer knowledge between non-overlapping domains. This brings doubt into the usefulness of CBT. In this work, we are the first to explore the effectiveness of CBT-based *non-overlapping* CDCF in the context of cross-domain venue recommendation. Our assumption is that CDCF could enable knowledge from a source domain (e.g. Foursquare) to enhance the quality of recommendations for a target domain (e.g. Yelp). In particular, our contributions are summarised below:

- We investigate the performance of a of state-of-the-art *non-overlapping* CDCF framework, CrossFire [13] in the context of venue recommendation.
- Inspired by [13], we extend the state-of-the-art sequential-based deep learning venue recommendation framework of [10] using the CBT-based technique, in order to evaluate its effectiveness in the cross-domain sequential-based venue recommendation task. To the best of our knowledge, this is the first work that studies the effectiveness of cross-domain recommendation strategies by leveraging sequences of user-venue interactions.
- We conduct comprehensive experiments on three large-scale real-world datasets from Foursquare, Yelp and Brightkite to evaluate the performance of CBT-based *non-overlapping* CDCF. Our experimental results demonstrate that CrossFire is not able to transfer useful knowledge from the source domain to the target domain for the venue recommendation task, being consistent with the previous study of [20]. In particular, we find that the CBT-based technique does not clearly contribute to the improvements observed from CrossFire, compared to the traditional single-domain MF-based models. Indeed, through experiments conducted when equating the source and target domains, we show that such improvements may not be explained by transfer of knowledge between source and target domains. We postulate that, in fact, that the improvements are gained from the additional parameters introduced by CrossFire, which makes it more flexible than the traditional single-domain MF-based approaches. In addition, our experimental results on a state-of-the-art sequential-based deep learning venue recommendation framework of [10] further validate this result.

This paper is organised as follows. Section 2 provides the problem statement of cross-domain venue recommendation. Then, we describe single-domain MF-based approaches and *non-overlapping* Cross-Domain Collaborative Filtering approaches (CDCF). The experimental setup for our experiments is detailed in Section 3, while comprehensive experimental results comparing the effectiveness of *non-overlapping* CDCF approaches with various single-domain MF-based approaches are reported in Section 4. Concluding remarks follow in Section 5.

2 Cross-Domain Venue Recommendation Frameworks

In this section, we first formulate the problem statement of the cross-domain venue recommendation task, without overlaps between domains. Then, we briefly introduce state-of-the-art MF-based strategies for the single domain and cross-domain recommendation tasks. After presenting the original CBT model of [11], we detail the cross-domain recommendation framework of CrossFire, proposed by Shu *et al.* [13]. Note that this framework was not originally proposed for the cross-domain venue recommendation task but is sufficiently flexible to be applied to it. Next, we present the sequential-based deep learning model of [10] for venue suggestion in a single domain, and then our proposed extended model of Deep Recurrent Transfer Learning (DRTL) for the cross-domain recommendation task.

2.1 Problem Statement

The task of cross-domain venue recommendation is to exploit knowledge from a source domain to enhance the quality of venue recommendation in a target domain. The task of venue recommendation in the target domain is to generate a ranked list of venues that a user might visit given his/her historical feedback, that is the previously visited venues from the checkin data. Let $\mathcal{U}^s$, $\mathcal{V}^s$ and $\mathcal{U}^t$, $\mathcal{V}^t$ be the sets of users and venues in the source and target domains, respectively. Let $\mathcal{V}^s_u$ ($\mathcal{V}^t_u$) denote the list of venues the user u in the source (target) domain has previously visited, sorted by time. $\mathcal{S}^s_u$ ($\mathcal{S}^t_u$) denote the list of sequences of visited venues of user u in the source (target) domain, for example, if $\mathcal{V}^s_u = (v_1, v_2, v_3)$, then $\mathcal{S}^s_u = ((v_1), (v_1, v_2), (v_1, v_2, v_3))$. $s_t \in \mathcal{S}^s_u$ ($\mathcal{S}^t_u$) denotes the sequence of visited venues of user u at time t in the source (target) domain (e.g. $s_2 = (v_1, v_2)$). All checkins by all users in the source (target) domain are represented as a matrix $C_s \in \mathbb{R}^{m^s \times n^s}$ ($C_t \in \mathbb{R}^{m^t \times n^t}$) where m^s, n^s and m^t, n^t are the number of users and venues in the source and target domains, respectively. Let $c^s_{u,i} \in C_s$ ($c^t_{u,i} \in C_t$) denote a user $u \in \mathcal{U}^s$ who visited venue $i \in \mathcal{V}$ in the source (target) domain. Note that $c^s_{u,i} = 0$ means that user u has neither left a rating nor made a checkin at venue i. The goal of a cross-domain recommendation task is to generate a personalised list of venues in the target domain t, by exploiting both the checkin matrices C_s and C_t of the source and target domains, respectively.

2.2 Traditional MF-based Models

Matrix Factorisation (MF) is a collaborative filtering technique that assumes users who share similar preferences, like visiting similar venues, are likely to influence each other [1]. The goal of MF is to reconstruct the checkin matrix $C \in \mathbb{R}^{m \times n}$ where m and n are the number of users and venues by calculating the dot product of latent factors of users $U \in \mathbb{R}^{m \times d}$ and venues $V \in \mathbb{R}^{n \times d}$, where d is the number of latent dimensions. In particular, the MF model is trained by minimising a loss function, which consists of sum-of-squared-error terms between the actual and predicted checkins, as follows:

$$\underset{U,V}{min} \|C - UV^T\|^2_F \tag{1}$$

where $\|.\|^2_F$ denotes the Frobenius norm.

Collective Matrix Factorisation (CMF) is a MF-based model that leverages both user-venue interactions and social information [21]. CMF aims to approximate the checkin matrix C and social link matrix[1] S, simultaneously. Built upon the MF model, the authors introduced a latent factor of friends, $F \in \mathcal{R}^{m \times d}$,

[1] $S \in \{0, 1\}^{m \times m}$ is the adjacency matrix representing the relationship between users.

which is used to capture the social relationship between users. In particular, the loss function of CMF is defined as follows:

$$\min_{U,V,F} \|C - UV^T\|_F^2 + \|S - UF^T\|_F^2 \tag{2}$$

2.3 Cross-Domain Collaborative Filtering Models

Codebook Transfer (CBT) is a cross-domain collaborative filtering technique proposed by Li *et al.* [11], which assumes that user-venue interactions across different domains might share similar checkin/rating patterns. CBT consists of two steps: extracting the patterns/knowledge of user-venue interactions from a source domain and exploiting the extracted patterns/knowledge to enhance the quality of venue recommendation in a target domain t. To extract the patterns/knowledge from the source domain, based on the collaborative filtering technique, CBT aims to approximate the checkin matrix C_s of the source domain s by finding a decomposition of C_s, i.e. a dot product of the latent factors of users $U_s \in \mathcal{R}^{m\times d}$, venues $V_s \in \mathcal{R}^{n\times d}$ and the shared checkin patterns[2] $W \in \mathcal{R}^{d\times d}$. The loss function of CBT is defined as follows:

$$\min_{U_s,W,V_s} \|C_s - U_s W V_s^T\|_F^2 \tag{3}$$

Next, CBT approximates the checkin matrix C_t of the target domain t and exploits the shared patterns W to boost the venue recommendation accuracy in target domain t as follows:

$$\min_{U_t,V_t} \|C_t - U_t W V_t^T\|_F^2 \tag{4}$$

CBT only updates the latent factors of the target domain t, that is the latent factors U_t and V_t, while keeping the checkin patterns W fixed.

CrossFire, proposed by Shu *et al.* [13] is a cross-domain recommendation framework. In applying it to venue recommendation, it aims to transfer knowledge extracted from the user-venue interactions and users' social links of the source domain to improve the quality of venue recommendation in the target domain. In particular, the objective of the CrossFire framework is to jointly approximate the checkin matrices C_s and C_t and the social link matrix S_s and S_t of the source and target domains, respectively. Built upon the CBT-based technique, the loss function of CrossFire is defined as follows:

$$\begin{aligned} \min_{U_s,U_t,V_s,V_t,W,Q} & \left(\|C_s - U_s W V_s^T\|_F^2 + \|C_t - U_t W V_t^T\|_F^2\right) \\ & + \left(\|S_s - U_s Q U_s^T\|_F^2 + \|S_t - U_t Q U_t^T\|_F^2\right) \end{aligned} \tag{5}$$

where $W \in \mathcal{R}^{d\times d}$ is a matrix with the checkin patterns in the latent space, as the CBT method, and $Q \in \mathcal{R}^{d\times d}$ is the social patterns in the latent space shared across the source and target domains. Unlike the CBT-based technique, the CrossFire framework jointly updates the latent factors U_s, U_t, V_s, V_t, W, Q. More details on the optimisation strategy of CrossFire are available in [13].

[2] The shared patterns denote similarities between the latent factors of the domains.

2.4 Sequential-based Venue Recommendation Frameworks

Deep Recurrent Collaborative Filtering (DRCF) proposed by Manotumruksa *et al.* [10], is a state-of-the-art sequential framework, which leverages deep learning algorithms such as Muli-Level Perceptron (MLP) and Recurrent Neural Networks (RNN) to capture users' *dynamic* preferences from their sequences of checkins. The DRCF framework consists of three components: Generalised Recurrent Matrix Factorisation (GRMF), Multi-Level Recurrent Perceptron (MLRP) and Recurrent Matrix Factorisation (RMF). In particular, given the sequence of a user's checkins $S_{i,t}$, the predicted checkin $\hat{c}_{u,i}$ is estimated as:

$$\hat{c}_{u,i} = a_{out}(h(\phi^{GRMF} \oplus \phi^{MLRP} \oplus \phi^{RMF})) \tag{6}$$

where a_{out} is the activation function, h is the hidden layer, and $\oplus$ denotes the concatenation operation. ϕ^{GRMF}, ϕ^{MLRP} and ϕ^{RMF} denote the GRMF, MLRP and RMF models, which are defined as follows:

$$\phi^{GRMF} = \left[d_{u,t} \otimes p_u \otimes q_i\right] \tag{7}$$

$$\phi^{MLRP} = \left[a_L(h_L(...a_1(h_1(d_{u,t} \oplus p_u \oplus q_i))))\right] \tag{8}$$

$$\phi^{RMF} = (d_{u,t} + p_u) \circledcirc q_i^{Gd} \tag{9}$$

where $\otimes$ is the element-wise product operation, $\circledcirc$ is the dot-product operation, $d_{u,t}$ is the user's *dynamic* preferences of user u at time t that are projected from the RNN layer, p_u and q_i are the latent factors of user u and venue i and L is the number of layers. Next, instead of training the DRCF framework to minimise the *pointwise* loss between the predicted checkin $\hat{c}_{u,i}$ and observed checkin $c_{u,i}$, the DRCF framework follows the Bayesian Personalised Ranking (BPR) strategy of [22] to learn DRCF's parameters, as follows:

$$\underset{\theta_e,\theta_r,\theta_h}{min} \sum_{u \in \mathcal{U}} \sum_{s_t \in S_u} \sum_{j \in \mathcal{V} - s_t} \log(\sigma(\hat{c}_{u,i} - \hat{c}_{u,j}))$$

where i is the most recently visited venue in s_t and $\sigma(x)$ is the sigmoid function. θ_e, θ_r and θ_h are the parameter sets of latent factors, RNN layers and hidden layers, respectively.

Deep Recurrent Transfer Learning (DRTL) is our proposed extension of DRCF to perform cross-domain recommendation for the venue suggestion task, exploiting the user-venue interactions from the source domain based on the CrossFire framework. In particular, we extend the loss function of DRCF as:

$$\begin{aligned}\underset{\theta_e^s,\theta_r^s,\theta_h^s,\theta_e^t,\theta_r^t,\theta_h^t}{min} & \sum_{u \in \mathcal{U}^s} \sum_{s_t \in S_u} \sum_{j \in \mathcal{V}^s - s_t} \log(\sigma(\hat{c}^s_{u,i} - \hat{c}^s_{u,j})) \\ & + \sum_{u \in \mathcal{U}^t} \sum_{s_t \in S_u} \sum_{j \in \mathcal{V}^t - s_t} \log(\sigma(\hat{c}^t_{u,i} - \hat{c}^t_{u,j}))\end{aligned}$$

where θ_e^s, θ_r^s ,θ_h^s (θ_e^t, θ_r^t, θ_h^t) are the parameters of the latent factors, the RNN layers and the hidden layers for the source (target) domain. $\hat{c}_{u,i}^s$ ($\hat{c}_{u,i}^s$) is the predicted checkin for the source (target) domain, defined as follows:

$$\hat{c}_{u,i}^s = a_{out}(h^s(\phi_{GRMF}^s \oplus \phi_{MLRP}^s \oplus \phi_{RMF}^s)) \tag{10}$$

$$\hat{c}_{u,i}^t = a_{out}(h^t(\phi_{GRMF}^t \oplus \phi_{MLRP}^t \oplus \phi_{RMF}^t)) \tag{11}$$

Next, we extend the GRMF, MLRP and RMF models of the DRCF framework by adding the shared checkin pattern W, as follows:

$$\phi_{GRMF}^s = \left[d_{u,t}^s \otimes p_u^s \otimes q_i^s \otimes W_{GRMF}\right] \quad \phi_{GRMF}^t = \left[d_{u,t}^t \otimes p_u^t \otimes q_i^t \otimes W_{GRMF}\right] \tag{12}$$

$$\phi_{MLRP}^s = \left[a_L(h_L^s(...a_1(h_1^s(d_{u,t}^s \oplus p_u^s \oplus q_i^s \oplus W_{MLRP}))))\right] \tag{13}$$

$$\phi_{MLRP}^t = \left[a_L(h_L^t(...a_1(h_1^t(d_{u,t}^t \oplus p_u^t \oplus q_i^t \oplus W_{MLRP}))))\right] \tag{14}$$

$$\phi_{RMF}^s = (d_{u,t}^s + p_u^s) \circledcirc q_i^s \circledcirc W_{RMF} \quad \phi_{RMF}^t = (d_{u,t}^t + p_u^t) \circledcirc q_i^t \circledcirc W_{RMF} \tag{15}$$

where W_{GRMF}, W_{MLRP} and W_{RMF} are the checkin pattern parameters shared between the source and target domains. Inspired by CrossFire, we include these shared parameters into the GRMF, MLRP and RMF models, such that useful checkin patterns can be transferred from the source domain to the target domain e.g., ϕ_{RMF}^t and ϕ_{RMF}^s can share information via W_{RMF}.

3 Experimental Setup

In this section, we evaluate the effectiveness of the *non-overlapping* CDCF approach of CrossFire by comparing with single-domain MF-based approaches. In addition, we extend the Deep Recurrent Collaborative Filtering framework (DRCF), to perform cross-domain recommendation using the CBT-based technique, namely DRTL. Then, we compare the effectiveness of DRTL by comparing with the CRCF. In particular, we aim to address the following research questions:

RQ1 *Can a non-overlapping CDCF approach that relies on Codebook Transfer extract useful checkin patterns from a source domain that can enhance the quality of venue recommendation in a target domain?*

RQ2 *Can we enhance the effectiveness of a state-of-the-art sequential venue recommendation technique on a target domain by incorporating Codebook Transfer from a source domain?*

3.1 Datasets & Measures

We conduct experiments on publicly available large-scale LBSN datasets. In particular, we use two checkin datasets from Brightkite[3] and Foursquare[4], and a rating dataset from Yelp[5]. We follow the common practice from previous works [10,

[3] https://snap.stanford.edu/data/
[4] https://archive.org/details/201309_foursquare_dataset_umn
[5] https://www.yelp.com/dataset_challenge

Table 1: Statistics of the three evaluation datasets.

	Brightkite	Foursquare	Yelp
Number of users	14,374	10,766	38,945
Number of venues	5,050	10,695	34,245
Number of ratings or checkins	681,024	1,336,278	981,379
Number of social links	33,290	164,496	1,598,096
% density of User-Venue matrix	0.93	1.16	0.07

22, 23] to remove venues with less than 10 checkins/ratings. Table 1 summarises the statistics of the filtered datasets. To evaluate the effectiveness of cross-domain venue recommendation frameworks, following previous studies [24, 23, 22, 10], we adopt a *leave-one-out* evaluation methodology: for each user, we select her most recent checkin/rating as a ground truth and randomly select 100 venues that she has not visited before as the testing set, where the remaining checkins/ratings are used as the training and validation set. The venue recommendation task is thus to rank those 101 venues for each user, aiming to rank highest the recent, ground truth checkin/rating. Note that previous works [12, 13, 15, 20] on *non-overlapping* CDCF use Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to evaluate the quality of rating prediction. In contrast, we evaluate the quality of recommendation in terms of Hit Ratio (HR)[6] and Normalised Discounted Cumulative Gain (NDCG) on the ranked lists of venues – as applied in previous studies [10, 23, 24]. In particular, HR considers the ranking nature of the task, by taking into account the rank(s) of the venues that each user has previously visited/rated in the produced ranking, while NDCG goes further by considering the checkin frequency/rating value of the user as the graded relevance label. Lastly, significance tests use a paired t-test.

3.2 Source & Target Domains

In the subsequent experiments, we use all three LBSN datasets, separated into training and testing datasets as described above. We report results both without using any cross-domain transfer (i.e. using the MF and CMF baselines), and when conducting non-overlapping cross-domain transfer. In doing so, we set one LBSN dataset as the Source domain, and one as the Target domain. Finally, to determine if the cross-domain transfer from other domains brings new information, we set the source and target domains to be equal - but while retaining a fair train/test split.

3.3 Implementations & Parameter Setup

We implement all techniques using the Keras deep learning framework[7]. Following [10, 13, 24], we set equal the dimension of the latent factors d of MF-based approaches and cross-domain recommendation frameworks, $d = 10$, and the number of hidden layers $L = 3$ across the three datasets. Note that since the

[6] Hit Ratio (HR) is a simplification of Mean Reciprocal Rank (MRR), which has been commonly used in top-N evaluation for recommendation systems [24–26] when ground-truth data are extracted from the implicit feedback.

[7] `https://bitbucket.org/feay1234/transferlearning`

impact of the hidden layer's number L and dimension size d have been previously explored in [23, 24], we omit varying the size of the hidden layers and the dimension of the latent factors in this study. Following Manotumruksa *et al.* [10], we randomly initialise all hidden, latent factors, and RNN layers' parameters for the source and target domains, $\theta_r^s, \theta_e^s, \theta_h^s, \theta_r^t, \theta_e^t, \theta_h^t$, with a Gaussian distribution, setting the mean to 0 and the standard deviation to 0.01, and then we apply the mini-batch Adam optimiser [27] to optimise those parameters. In doing so, we achieve faster convergence than stochastic gradient descent and automatically adjust the learning rate for each iteration. We initially set the learning rate to 0.001[8] and set the batch size to 256.

4 Experimental Results

Table 2 reports the effectiveness of single-domain MF-based approaches and the CrossFile (*non-overlapping* CDCF) approach in terms of HR@10 and NDCG@10 on the three evaluation datasets. The Target Domain row indicates the dataset/domain for which we generate venue recommendations. The Source Domain row indicates the dataset that is used as auxiliary information for *non-overlapping* CDCF. Note that the single-domain MF-based approaches do not leverage the auxiliary information from the source domain.

On inspection of the results of CrossFire, we observe that it consistently and significantly outperforms MF and CMF, for both HR and NDCG, across both the Brightkite and Yelp datasets. This observation is consistent with the results reported in [13] when transferring knowledge between books and movie domains. In contrast, MF is more effective than CrossFire in terms of HR and NDCG on the Foursquare dataset. These results bring doubt that CrossFire generalises across different datasets. Next, when CrossFire uses Brightkite as both the source and target domains, then CrossFire is more effective than any other setup (i.e. Foursquare or Yelp as the source domain). Similarly, for CrossFire, we observe that when CrossFire uses Yelp as both the source and target domains, it outperforms CrossFire with Brightkite or Foursquare as the source domain.

These experimental results are counter intuitive, because we expect no improvement from CrossFire when the source and the target domains are identical – indeed, there should be no useful checkin patterns to be transfered from the source domain to the target one, as no new information has been obtained. Note that none of previous works [12, 13, 15, 20] on *non-overlapping* CDCF reports the effectiveness of their proposed approaches when setting the source domain equal to the target domain. In response to research question RQ1, our experimental results demonstrate that the CBT-based strategy of CrossFire does not clearly contribute to the improvements of the recommendation accuracy, compared to the traditional single-domain MF-based models. Indeed, we postulate that the observed improvements (for Brightkite and Yelp) are gained from the additional parameters introduced in CrossFire (namely, W and Q in Equation (5)), which make the CrossFire more flexible than the traditional single-domain MF-based

[8] The default learning rate setting of the Adam optimiser in Keras.

Table 2: Performance in terms of HR@10 and NDCG@10 of single-domain and cross-domain MF-based approaches. The best performing result in each row is highlighted in bold and $*$ indicates significant differences in terms of paired t-test with $p < 0.01$, comparing to the best performing result.

Target Domain	Brightkite				
Model	MF	CMF	CrossFire	CrossFire	CrossFire
Source Domain	-	-	Brightkite	Foursquare	Yelp
HR	0.5252*	0.5931*	**0.6140**	0.5906*	0.5668*
NDCG	0.3224*	0.3444*	**0.3670**	0.3546*	0.3421*

Target Domain	Foursquare				
Model	MF	CMF	CrossFire	CrossFire	CrossFire
Source Domain	-	-	Brightkite	Foursquare	Yelp
HR	**0.6897**	0.6750*	0.6737*	0.6483*	0.6722*
NDCG	**0.4279**	0.3692*	0.4159*	0.3997*	0.4189*

Target Domain	Yelp				
Model	MF	CMF	CrossFire	CrossFire	CrossFire
Source Domain	-	-	Brightkite	Foursquare	Yelp
HR	0.3458*	0.3472*	0.4364	0.4331*	**0.4399**
NDCG	0.1782*	0.1773*	**0.2332**	0.2275*	0.2331

approaches, and not by transferring knowledge from the source domain. This new evidence, for the venue recommendation task, supports the arguments of Cresmoni and Quadrana [20], namely that a CBT-based strategy cannot effectively transfer knowledge when the source and target domains do not overlap.

Next, Table 3 reports the effectiveness of single-domain and *non-overlapping* sequential-based venue recommendation frameworks: namely the Deep Recurrent Collaborative Filtering (DRCF) and the proposed extended CBT-based strategy of Deep Recurrent Transfer Learning (DRTL), respectively. In Table 3 we observe similar results as reported in Table 2. For example, DRTL consistently and significantly outperforms DRCF, for HR and NDCG, across the Brightkite and Yelp datasets, while DRCF is more effective than DRTL in terms of HR and NDCG on the Foursquare dataset.

In particular, when using Brightkite or Yelp as the target domain, we found that the performances of DRTL in terms of HR@10 and NDCG@10 with Foursquare as the source domain are more effective than other setups (i.e. Brightkite or Yelp as source domains). These results imply that DRTL may be able to transfer useful checkin patterns from the Foursquare dataset to enhance the quality of venue recommendation on the Brightkite and Yelp datasets. We note that the Foursquare dataset is larger than the Brightkite dataset, and hence it is possible that the checkin patterns extracted from Foursquare are reasonably useful for improving the effectiveness of recommendation system on the Brightkite dataset. Interestingly, the checkin patterns extracted from Foursquare are also useful for the Yelp dataset, perhaps due to the higher density of the checkins in the Foursquare dataset (see Table 1).

Table 3: Performance in terms of HR@10 and NDCG@10 of several sequential-based venue recommendation frameworks. The best performing result in each row is highlighted in bold and ∗ indicates significant differences in terms of paired t-test with $p < 0.01$, comparing to the best performing result.

Target Domain	Brightkite			
Model Source Domain	DRCF	DRTL Brightkite	DRTL Foursquare	DRTL Yelp
HR	0.5252*	0.6975*	**0.7083**	0.7036
NDCG	0.3224*	0.5244*	**0.5341**	0.5335*

Target Domain	Foursquare			
Model Source Domain	DRCF -	DRTL Brightkite	DRTL Foursquare	DRTL Yelp
HR	**0.8595**	0.8360*	0.8444*	0.8300*
NDCG	**0.7096**	0.6700*	0.6719*	0.6632*

Target Domain	Yelp			
Model Source Domain	DRCF -	DRTL Brightkite	DRTL Foursquare	DRTL Yelp
HR	0.5019*	0.5496*	**0.5577**	0.5350*
NDCG	0.2858*	0.3197	**0.3215**	0.3059*

On the other hand, as postulated above for CrossFire, a possible reason for the increased effectiveness of the DRTL model is the increased parameter space allowing more flexible learned models. To investigate this further, Figure 1 plots the number of parameters of each approach of the nine examined approaches (MF, CMF, 3x CrossFire, DRCF, 3x DRTL) versus the resulting effectiveness (HR & NDCG@10) on the Brightkite target domain. For instance, for MF, the number of parameters is defined by $m \times d + d \times n = 14374 \times 10 + 10 \times 5050 \approx 188,060$[9]. Examining the figure, some moderate correlation can easily be observed. We quantify this correlation using Spearman's ρ for each target domain and evaluation measure in Table 4. Here, we observe that in 3 out of the 6 settings, the observed correlations are significant, supporting our postulate that the increasing parameter space of the models – thereby allowing further flexibility – could explain the increasing effectiveness.

Overall, in response to research question RQ2, our experimental results demonstrate that the CBT technique of DRTL *appears to work* in the same settings that CrossFire works on. This may provide evidence that the checkin patterns extracted from the source domain that is larger than the target domain are useful for enhancing the quality of venue recommendation in sequential-based venue recommendation. However, we also provide some evidence that these improvements can be explained by the increased parameter space of the jointly-optimised transfer learning models used by the CBT technique.

[9] Recall that we remove sparse users and venues.

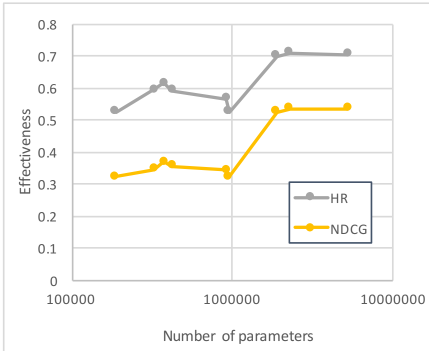

Fig. 1: Number of model parameters vs. effectiveness across all approaches, for the Brightkite dataset.

Target	HR	NDCG@10
Brightkite	0.62	0.66
Foursquare	0.53	0.67*
Yelp	0.85*	0.86*

Table 4: Spearman's ρ for $N = 9$ approaches: number of model parameters vs. effectiveness.

5 Conclusions

In this paper, we investigate the performance of a state-of-the-art *non-overlapping* cross-domain venue recommendation framework, CrossFire, that relies on the CodeBook Transfer (CBT) technique. Moreover, we extend the state-of-the-art sequential-based venue recommendation framework to perform cross-domain venue recommendation based on CrossFire. Our comprehensive experiments on three large-scale datasets from Brightkite, Foursquare and Yelp show that the CBT-based technique does not clearly contribute to the improvements of CrossFire, compared to the traditional single-domain MF-based approaches in the context of venue recommendation. In fact, such improvements may be due to the additional parameters introduced by the CBT-based technique. Regarding sequential-based recommendation, our experiments demonstrate that the CBT-based technique can enhance the effectiveness of a state-of-the-art sequential-based venue recommendation framework, namely DRCF. In particular, the results imply that the checkin patterns extracted from the source domain that is larger than the target domain can be useful for enhancing the effectiveness of DRCF. However, we also examined the parameter space of the resulting models, which showed at least moderate correlation (significant in 3 out of 6 cases) with the resulting effectiveness, suggesting that at least some of the benefit in CBT simply arises from the increased size of the parameter spaces.

As future work, we plan to investigate a *non-overlapping* cross-domain venue recommendation strategy that can effectively transfer knowledge across two domains. We will consider users' checkin behaviours in certain regions, instead of taking into account on how users checkin on platforms in general as state-of-the-art CDCF strategies do. It is the special characteristic of the cross-domain venue recommendation task that makes the CDCF approaches less stable, since users' checkin behaviours will highly depend on the regions that the users are located. For example, we will consider users' checkin behaviour at the center of a certain city in two different platforms like Yelp and Foursquare, and then weight the transfer learning accordingly in the cross-domain venue recommendation task.

References

1. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer (8) (2009)
2. Ma, H., Zhou, D., Liu, C., Lyu, M.R., King, I.: Recommender systems with social regularization. In: Proc. of WSDM. (2011)
3. Manotumruksa, J., Macdonald, C., Ounis, I.: Matrix factorisation with word embeddings for rating prediction on location-based social networks. In: Proc. of ECIR. (2017)
4. Manotumruksa, J., Macdonald, C., Ounis, I.: Regularising factorised models for venue recommendation using friends and their comments. In: Proc. of CIKM. (2016)
5. Yuan, F., Guo, G., Jose, J., Chen, L., Yu, H.: Joint geo-spatial preference and pairwise ranking for point-of-interest recommendation. In: Proc. of ICTAI. (2016)
6. Yu, F., Liu, Q., Wu, S., Wang, L., Tan, T.: A dynamic recurrent model for next basket recommendation. In: Proc. of SIGIR. (2016)
7. Tang, S., Wu, Z., Chen, K.: Movie recommendation via BLSTM. In: Proc. of ICMM. (2017)
8. Zhang, Y., Dai, H., Xu, C., Feng, J., Wang, T., Bian, J., Wang, B., Liu, T.Y.: Sequential click prediction for sponsored search with recurrent neural networks. In: Proc. of AAAI. (2014)
9. Cheng, C., Yang, H., Lyu, M.R., King, I.: Where you like to go next: Successive point-of-interest recommendation. In: Proc. of IJCAI. (2013)
10. Manotumruksa, J., Macdonald, C., Ounis, I.: A deep recurrent collaborative filtering framework for venue recommendation. In: Proc. of CIKM. (2017)
11. Li, B., Yang, Q., Xue, X.: Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction. In: Proc. of IJCAI. (2009)
12. Zang, Y., Hu, X.: LKT-FM: A novel rating pattern transfer model for improving non-overlapping cross-domain collaborative filtering. In: Proc. of ECMLPKDD, Springer (2017)
13. Shu, K., Wang, S., Tang, J., Wang, Y., Liu, H.: Crossfire: Cross media joint friend and item recommendations. In: Proc. of WSDM. (2018)
14. Farseev, A., Samborskii, I., Filchenkov, A., Chua, T.S.: Cross-domain recommendation via clustering on multi-layer graphs. In: Proc. of SIGIR. (2017)
15. Hu, L., Cao, J., Xu, G., Cao, L., Gu, Z., Zhu, C.: Personalized recommendation via cross-domain triadic factorization. In: Proc. of WWW. (2013)
16. Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., Yang, D.: Learning informative priors from heterogeneous domains to improve recommendation in cold-start user domains. (2016)
17. Mirbakhsh, N., Ling, C.X.: Improving top-n recommendation for cold-start users via cross-domain information. In: Proc. of TKDD. (2015)
18. He, M., Zhang, J., Yang, P., Yao, K.: Robust transfer learning for cross-domain collaborative filtering using multiple rating patterns approximation. In: Proc. of WSDM. (2018)
19. Coyle, M., Smyth, B.: (web search) shared: social aspects of a collaborative, community-based search network. In: Proc. of AH. (2008)
20. Cremonesi, P., Quadrana, M.: Cross-domain recommendations without overlapping data: myth or reality? In: Proc. of RecSys. (2014)
21. Singh, A.P., Gordon, G.J.: Relational learning via collective matrix factorization. In: Proc. of SIGKDD. (2008)

22. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: Bayesian personalized ranking from implicit feedback. In: Proc. of UAI. (2009)
23. He, X., Zhang, H., Kan, M.Y., Chua, T.S.: Fast matrix factorization for online recommendation with implicit feedback. In: Proc. of SIGIR. (2016)
24. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.S.: Neural collaborative filtering. In: Proc. of WWW. (2017)
25. Xiang, L., Yuan, Q., Zhao, S., Chen, L., Zhang, X., Yang, Q., Sun, J.: Temporal recommendation on graphs via long-and short-term preference fusion. In: Proc. of SIGKDD, ACM (2010) 723–732
26. Lee, S., Song, S.i., Kahng, M., Lee, D., Lee, S.g.: Random walk based entity ranking on graph for multidimensional recommendation. In: Proc. of RecSys, ACM (2011) 93–100
27. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)