Collaborative Filtering with Attribution Alignment for Review-based Non-overlapped Cross Domain Recommendation

Weiming Liu, Xiaolin Zheng, Mengling Hu, Chaochao Chen∗
College of Computer Science and Technology, Zhejiang University, China
{21831010, xzheng, humengling, zjucccl}@zju.edu.cn

ABSTRACT
Cross-Domain Recommendation (CDR) has been popularly studied to utilize different domain knowledge to solve the data sparsity and cold-start problem in recommender systems. In this paper, we focus on the Review-based Non-overlapped Recommendation (RNCDR) problem. The problem is commonly-existent and challenging due to two main aspects, i.e., there are only positive user-item ratings on the target domain and there is no overlapped user across different domains. Most previous CDR approaches cannot solve the RNCDR problem well, since (1) they cannot effectively combine review with other information (e.g., ID or ratings) to obtain expressive user or item embedding, (2) they cannot reduce the domain discrepancy on users and items. To fill this gap, we propose Collaborative Filtering with Attribution Alignment model (CFAA), a cross-domain recommendation framework for the RNCDR problem. CFAA includes two main modules, i.e., rating prediction module and embedding attribution alignment module. The former aims to jointly mine review, one-hot ID, and multi-hot historical ratings to generate expressive user and item embeddings. The latter includes vertical attribution alignment and horizontal attribution alignment, tending to reduce the discrepancy based on multiple perspectives. Our empirical study on Douban and Amazon datasets demonstrates that CFAA significantly outperforms the state-of-the-art models under the RNCDR setting.

CCS CONCEPTS
• Information systems → Recommender systems; • Collaborative filtering;

KEYWORDS
Recommendation, Domain Adaptation, Transfer Learning

1 INTRODUCTION
With the advent of digital era, more and more users participate in multiple domains (platforms) for different purposes, e.g., buying books on Amazon and reading news on Flipboard [12]. How to comprehensively utilize the cross domain information to improve the performance of recommendation systems has become a hot topic. Therefore, Cross Domain Recommendation (CDR) becomes more and more attractive for establishing highly accurate recommendation systems [5, 25, 43, 54]. Most existing CDR models assume that both the source and target domains share the same set of users, which makes it easier to transfer useful knowledge across domains based on these overlapped users. Moreover, most existing CDR models assume the existence of sufficient negative knowledge, which limits their applications.

In this paper, we focus on a general problem in CDR where the target domain only has a small proportion of positive ratings but no negative ones. Meanwhile users in the source and target domains are totally non-overlapped. Besides user-item rating information, we also assume the existence of user-item reviews which are commonly used as side-information for alleviating the data sparsity problem [44, 48]. Specifically, we term this problem as Review-based Non-overlapped Cross Domain Recommendation (RNCDR), where we aim to transfer knowledge from a relative dense source domain to a sparse target domain to enhance the prediction performance, as shown in Fig. 1. The RNCDR problem is widely existed since the target domain is always suffering from the data sparsity problem. Therefore, the target domain only has a small proportion of observed positive samples without negative ones. We summarize the main challenges as follows. (1) Users/items always have diverse characteristics and preferences which should be well-exploited across domains. (2) Different domains tend to have varied behavioral patterns [15, 49], causing the existence of latent embedding attribution bias and discrepancy [22].

Although there have been previous studies on the RNCDR problem [48], they cannot solve it well. The current state-of-the-art model on RNCDR is TDAR [48] which adopts deep Domain Adversarial Neural Network (DANN) [16] to align the source and target user/item features. On the one hand, it fails to integrate the review information with one-hot ID and multi-hot historical rating [9, 34], leading to less expressive user and item embeddings [17, 24]. On the other hand, a deep adversarial network with domain discriminator is unstable and hard to train in practice [38]. Therefore, it cannot solve both challenges and leads to poor model performance. These two negative aspects make this approach difficult to achieve the RNCDR task.

To address the aforementioned issues, in this paper, we propose CFAA, a cross-domain recommendation framework for the RNCDR problem. In order to better model user/item embeddings and align
the latent embedding attributions across domains for high-quality rating predictions, we utilize two modules in CFAA, i.e., rating prediction module and embedding attribution alignment module. The rating prediction module aims to tackle the first challenge, capturing useful information and generating expressive user and item embeddings. For this purpose, we fuse the one-hot ID, multi-hot historical ratings, and text-based review information to better capture user and item collaborative preferences. The embedding attribution alignment module tends to tackle the second challenge, properly reducing the attribution discrepancy and transferring consistent useful knowledge across the source and target domains. Specifically, we propose dual perspectives on reducing the discrepancy, i.e., vertical attribution alignment with typical-sample optimal transport and horizontal attribution alignment with the adaptation on attribution graph. Vertical attribution alignment confines the probability distribution among the source and target domains for the corresponding attribution and horizontal attribution alignment tends to exploit the consistent relevant relationship between different attributes across domains. These two methods cooperate with each other to complete the alignment task and reduce the domain discrepancy.

We summarize our main contributions as follows: (1) We propose a novel framework, i.e., CFAA, for the RNCDR problem, which contains rating prediction module and embedding attribution alignment module. (2) The proposed rating prediction module can efficiently combine the text-based review information with one-hot ID and multi-hot historical rating embedding, and the embedding attribution alignment module equipped with vertical and horizontal alignment can better reduce the domain discrepancy. (3) Extensive empirical studies on Douban and Amazon datasets demonstrate that CFAA significantly improves the state-of-the-art models under the RNCDR setting.

2 RELATED WORK

Traditional Cross Domain Recommendation. Traditional Cross Domain Recommendation (CDR) emerges as a technique to alleviate the long-standing data sparsity problem in recommendation by assuming the same or partial proportion of overlapped user/item set across domains [20, 47]. Man et al. [28] first proposed to learn the mapping function on the overlapped users across domains. Zhao et al. [51] extended the research on the attribution aspects transfer.

Some recent method [21] even exploited bidirectional latent relations between users and items to obtain more reliable knowledge. Besides, Hu et al. [17] further proposed multi-task learning strategy which stitches both source and target information together. It is noticeable that the majority of traditional CDR frameworks always assume the existence of overlapped users as the bridge for knowledge sharing across different domains [54]. However, the source and target domains may not share overlapped users in the real-world applications which limits their application scenario.

Review-based Non-overlapped Cross Domain Recommendation. There are some CDR models that can solve the situations where users/items are non-overlapped. Yang et al. [20] was the first to propose the widely used codebook transfer approach for non-overlapped CDR. However, it cannot be extended to implicit feedback, which strongly limits its generalization capability. Furthermore, without overlapped users or items act as bridges for knowledge transfer, the recommendation accuracy is still not satisfying [11]. Recently, some researchers [52] have propose to utilize review text of users and items to enhance the model performance. This inspires us to take advantage of these review information as the transferring bridge across different domains. For example, Wang et al. [44] used text-enhanced information for solving both overlapped and non-overlapped CDR problems with deep adversarial learning strategy. Yu et al. [48] considered a more general situation where the target domain only has positive ratings, and proposed dual adversarial alignment on both user and item embeddings. However, these models cannot combine the one-hot ID with multi-hot historical rating into the review information, leading to limited model performance. Moreover, some studies [22] have shown the existence of embedding bias and discrepancy on CDR problem since two domains may always have different expressions. This enlightens us to transfer useful review knowledge by reducing the embedding bias for better results.

Domain Adaptation. Domain adaptation technique was proposed to transfer knowledge from a well-labeled source domain to a target domain without or with less labels [42]. The most classic method is Maximum Mean Discrepancy (MMD) [4, 35] which is based on the mean statistic of samples. Moreover, some scholars proposed Correlation Alignment (CORAL) [40, 41] to align the second-order covariance statistics. More recently, ESAM [8] extended CORAL with attribution correlation congruence for solving the long-tailed
item recommendation problem. However, performing statistic on mean or covariance may hard to capture the complex and high dimension data characteristics. For this, Ganin et al. [16] proposed Domain Adversarial Neural Network (DANN) which integrated a domain discriminator with adversarial training to align the embeddings across domains. Notably that previous review-based CDR models, e.g., TDAR [48] and Rec-GAN [44], also adopted adversarial training for aligning the source and target domains. However, latest researches [38] pointed out that the origin adversarial training may be unstable under some circumstances which will hurdle the performance. In this paper, we propose to enhance statistic-based embedding attribution alignment method by attribution distribution and attribution relevance alignment to exploit more expressive information on complex feature space.

3 MODELING FOR CFAA

First, we describe notations. We assume there are two domains, i.e., a source domain \( S \) and a target domain \( T \). There are \( N_S \) and \( N_T \) users in source and target domains respectively. There are \( N_S \) and \( N_T \) items in source and target domains respectively. Let \( \Omega^S \) and \( \Omega^T \) be the set of user reviews. Similarly, let \( \Omega^S \) and \( \Omega^T \) be the set of item reviews. Unlike the traditional cross domain recommendation which has the assumption that both source and target domains share the same set of users, in RNCDR, the source and target users are totally non-overlapped.

Then, we introduce the overview of our proposed CFAA framework, as is illustrated in Fig. 2. CFAA model mainly has two modules, i.e., rating prediction module and embedding attribution alignment module. The rating prediction module aims to combine the review information with one-hot ID and multi-hot historical rating and item rating matrices in \( S \) and \( T \) respectively. Moreover, let \( \Omega^S \) and \( \Omega^T \) be the set of user reviews. Similarly, let \( \Omega^S \) and \( \Omega^T \) be the set of item reviews. Unlike the traditional cross domain recommendation which has the assumption that both source and target domains share the same set of users, in RNCDR, the source and target users are totally non-overlapped.

Then, we introduce the overview of our proposed CFAA framework, as is illustrated in Fig. 2. CFAA model mainly has two modules, i.e., rating prediction module and embedding attribution alignment module. The rating prediction module aims to combine the review information with one-hot ID and multi-hot historical rating and item rating matrices in \( S \) and \( T \) respectively. Moreover, let \( \Omega^S \) and \( \Omega^T \) be the set of user reviews. Similarly, let \( \Omega^S \) and \( \Omega^T \) be the set of item reviews. Unlike the traditional cross domain recommendation which has the assumption that both source and target domains share the same set of users, in RNCDR, the source and target users are totally non-overlapped.

3.1 Rating Prediction Module

Firstly, we provide the details of the rating prediction module. For convenience, we use the notations and calculation process in the source domain as an example. For the \( i \)-th user and the \( j \)-th item, we define their corresponding one-hot ID vectors as \( X^U_i^S \) and \( X^V_j^S \), respectively. Let \( R^S_i \) and \( R^S_j \) denote the historical rating for the \( i \)-th user and the \( j \)-th item. For the \( i \)-th user, we use \( \Omega^S_i \) to denote the corresponding review which includes \( \Omega^S_i = (S^1_i, S^2_i, \ldots, S^k_i) \) sentences. Likewise, we use \( \Omega^S_j \) to denote the corresponding review for the \( j \)-th item which includes \( \Omega^S_j = (S^1_j, S^2_j, \ldots, S^h_j) \) sentences. Notably that we adopt the sentence segmentation component (Sentenciizer\(^1\)) to split the origin document into several individual sentences in \( \Omega^S_i \) and \( \Omega^S_j \).

We adopt a trainable lookup table to exploit the user and item one-hot ID embedding as \( \text{LookUp}(X^U_i^S) = E^U_i \) and \( \text{LookUp}(X^V_j^S) = E^V_j \). We utilize the fully connected layers \( F_S \) and \( F_T \) to obtain user and item behavior embeddings as \( F_S(R^S_i) = C^U_i \) and \( F_T(R^S_j) = C^V_j \), respectively. Meanwhile we utilize the pre-trained BERT model’s penultimate encoder layer to obtain the contextualized word embeddings for each sentence [30]. Then we average every sentence’s word embeddings to generate the sentence embedding. We then average the sentence embeddings of the \( i \)-th user or \( j \)-th item as their corresponding review-based embedding \( H^U_i \) and \( H^V_j \). Finally, we utilize fully connected layers \( G_S \) and \( G_T \) to obtain the user and item general embedding as \( G_S(E^U_i \oplus C^U_i \oplus H^U_i) = U^S \in \mathbb{R}^{N_S \times D} \) and \( G_T(E^V_j \oplus C^V_j \oplus H^V_j) = V^S \in \mathbb{R}^{N_T \times D} \). Where \( D \) denotes batch size, \( D \) denotes the dimension of the latent embedding, and \( \oplus \) denotes the concatenation operation. Likewise, we can also obtain the user and item general embeddings on the target domain as \( U^T \in \mathbb{R}^{N_T \times D} \) and \( V^T \in \mathbb{R}^{N_T \times D} \). After that, we adopt the fully connected layer \( F \) to predict user-item ratings as \( G_U(U^S, V^S) = \hat{R}^S \) and \( G_T(U^T, V^T) = \hat{R}^T \). We further use cross entropy loss \( L_C \) to minimize the prediction ratings and ground-truth ones as below:

\[
L_C = \frac{1}{N} \sum_{i=1}^{N} (\hat{R}_i^S \log \hat{R}_i^S + (1 - \hat{R}_i^S) \log (1 - \hat{R}_i^S) + \hat{R}_i^T \log \hat{R}_i^T).
\]

It is noticeable that we only have positive ratings in the target domain. Through this basic loss function, the network parameters can be adjusted to the training data efficiently.

3.2 Embedding Attribution Alignment Module

Although the rating prediction module can provide us a simple and good baseline, it still cannot depict user and item characteristics in the source and target domains well. The main reason lies in the data sparsity across domains which causes the embedding attribution bias and discrepancy. Specifically, we denote each dimension of an embedding as a certain kind of distribution. The attribution discrepancy includes two aspects, i.e., vertical probability discrepancy and horizontal relevance discrepancy on each attribution, as shown in Fig. 3 and Fig. 4. For example, Attr3 in both source and target domains has different probability distribution as shown in Fig. 3(b). Meanwhile, Attr2 and Attr3 have different relevance strength across domains, indicating the existence of the relevance discrepancy, as shown in Fig. 4. These biases deteriorate the knowledge transfer across domains and may even lead to the negative transfer phenomenon [46]. Therefore, it is essential to reduce the embedding attribution discrepancy for better knowledge transfer. The embedding attribution discrepancy module consists of two main algorithms, i.e., Vertical Attribution Alignment and Horizontal Attribution Alignment. These two algorithms proceed from different perspectives, and they complement each other to solve the embedding attribution discrepancy problem.

3.2.1 Vertical Attribution Alignment

We first introduce the attribution distribution alignment. It is reasonable to assume that each dimension of user (item) embeddings has certain meaningful information like occupation, income, hobby (style, theme, brand), and etc. Although we cannot directly decipher what is the exact meaning of each dimension, deep analysis of these latent embedding

\(^1\)https://spacy.io/api/sentencizer
In order to better model the attribution distribution between the source and target domains, we adopt the optimal transport technique [14]. Optimal transport is serious to handle the knowledge transfer across domains, which not only raises the training difficulty but also leads to the scattered target space with domain discrepancy. Vertical attribution discrepancy seriously hinders the knowledge transfer across domains, which can obtain the stable solution of \( \mathbf{M} \).

**Typical Sample Selection.** We first introduce the typical sample selection algorithm, which aims to find \( K \) typical samples \( \mathbf{M}_Z^S \in \mathbb{R}^{K \times D} \) in both source and target domains where the \( q \)-th column \( \mathbf{M}_Z^S \in \mathbb{R}^{K \times 1} \) represents the typical samples on the \( q \)-th attribution. Let \( X = \{S, T\} \) denote the domain index, and \( Z = \{U, V\} \) denote the user or item set. Inspired by [2, 32], we formulate the typical sample selection problem as:

\[
\begin{align*}
\min_q & \quad \sum_{i=1}^N \sum_{j=1}^K (\Psi_{ij}^X) \left( Z_{iq}^X - (\mathbf{M}_Z^S)_{ij} \right)^2 + \alpha \cdot \mathcal{R}((\Psi_{ij}^X)_{ij}) \\
\text{s.t.} & \quad (\Psi_{ij}^X)_{ij} = 1, \mathcal{R}((\Psi_{ij}^X)_{ij}) = (\Psi_{ij}^X)_{ij} \log(\Psi_{ij}^X), (\Psi_{ij}^X)_{ij} > 0,
\end{align*}
\]

where \( \Psi_{ij}^X \in \mathbb{R}^{N \times K} \) be the similarity matrix between the data samples and the typical proxies on the \( q \)-th attribution. The non-negative entropy norm term \( \mathcal{R}((\Psi_{ij}^X)_{ij}) = (\Psi_{ij}^X)_{ij} \log(\Psi_{ij}^X) \) is set to avoid trivial solution with \( \alpha \) denoting the regularization strength [2]. We will provide the optimization details on the typical sample selection algorithm in Appendix A. In short, alternatively updating \( \mathbf{M}_Z^S \) and \( \Psi_{ij}^X \) can solve Equation (1) efficiently as:

\[
(\Psi_{ij}^X)_{ij} = \frac{\exp(-\xi_{ij}/\alpha)}{\sum_{k=1}^K \exp(-\xi_{ik}/\alpha)}, \quad (\mathbf{M}_Z^S)_{ij} = \frac{\sum_{i=1}^N (\Psi_{ij}^X)_{ij} Z_{ik}}{\sum_{i=1}^N (\Psi_{ij}^X)_{ij}},
\]

where \( \xi_{ij} = (Z_{ik}^X - (\mathbf{M}_Z^S)_{ij})^2 \). Since this problem is convex, we can obtain the stable solution of \( \Psi_{ij}^X \) and \( \mathbf{M}_Z^S \) through iterations.

**Optimal Transport Alignment.** In order to better model the attribution distribution between the source and target domains, we adopt the optimal transport technique [14]. Optimal transport is based on Kantorovich problem [1], seeking for a general coupling

\[
\pi_q^Z \in X(S_q, T_q) \text{ between } S_q \text{ and } T_q \text{ on the } q \text{-th attribution:}
\]

\[
\pi_q^Z = \arg \min \int M((\mathbf{M}_Z^S)_{q}, (\mathbf{M}_Z^T)_{q}) d\pi_q^Z((\mathbf{M}_Z^S)_{q}, (\mathbf{M}_Z^T)_{q}).
\]

The cost function matrix \( M((\mathbf{M}_Z^S)_{q}, (\mathbf{M}_Z^T)_{q}) \) denotes the cost to move probability mass from \( (\mathbf{M}_Z^S)_{q} \) to \( (\mathbf{M}_Z^T)_{q} \) where \( (\mathbf{M}_Z^S)_{q} \) and \( (\mathbf{M}_Z^T)_{q} \) denote the typical selected samples on the \( q \)-th attribution for source and target domain respectively. The discrete optimal transport formulation can be expressed as:

\[
\hat{\pi}_q^Z = \arg \min \left\{ \langle \pi_q^Z, M \rangle + \epsilon \mathcal{R}(\pi_q^Z) \right\},
\]

where \( \hat{\pi}_q^Z \in \mathbb{R}^{K \times K} \) is the ideal coupling matrix between the source typical samples \( (\mathbf{M}_Z^S)_{q} \) and the target typical samples \( (\mathbf{M}_Z^T)_{q} \). The matching matrix follows the constraint of \( 1 = \sum_q \hat{\pi}_q^Z \). Therefore, we can calculate the optimal transport distance \( d_0^Z \) as:

\[
d_0((\mathbf{M}_Z^S)_{q}, (\mathbf{M}_Z^T)_{q}) = \frac{1}{K^2} \sum_{i=1}^K \sum_{j=1}^K \left| \hat{\pi}_q^Z(i, j) \right|^2 ((\mathbf{M}_Z^S)_{q} - (\mathbf{M}_Z^T)_{q})_{ij}^2.
\]

In summary, we propose the attribution distribution alignment loss as below:

\[
L_O = \frac{1}{D} \sum_{q=1}^D \left[ d_0((\mathbf{M}_Z^S)_{q}, (\mathbf{M}_Z^T)_{q}) + d_0((\mathbf{M}_V^S)_{q}, (\mathbf{M}_V^T)_{q}) \right].
\]

(3)

Take Fig. 3 for example, we first collect the typical samples of Attr3, e.g., the stars marked with numbers in Fig. 3(c), in both source and target domains. Then we adopt the Optimal Transport (OT) to align these typical samples across domains. Afterward, we minimize the distance between the matched typical samples. Finally, we can align the attribution probability distribution in Fig. 3(d).

### 3.2.2 Horizontal Attribution Alignment

Then we introduce the horizontal attribution alignment algorithm. Previous researches have pointed out that aligning the corresponding relevant attribution relationship, e.g., adopting the correlation alignment with covariance matrix in ESAM [8], can better enhance the model performance. However, covariance is hard to capture the complex and nonlinear hidden relationships between different attributions under the RNC problem [6]. Therefore, we propose horizontal attribution alignment with attribution subspace modelling and attribution graph alignment methods. We first use attribution subspace modelling to build the attribution graph, exploiting the hierarchical...
and topological structure between different attributions. Then we propose attribution graph alignment to align attribution graphs in source and target domains with Wasserstein distance metric.

**Attribution Subspace Modelling.** SLIM (Sparse Linear Methods) [10, 33] have been widely adopted in recommendation systems due to its good performance. It computes the item-item relations with statistical learning using the corresponding coefficient matrix. Meanwhile, one can even adopt it to measure the embedding attribution-attribute relations according to the following optimization problem:

$$\min_{B^X_Z} \frac{1}{2} \left\| X^X - X^X \cdot B^X_Z \right\|_2^2 + \nu \left\| B^X_Z \right\|_F^2 \quad \text{s.t.} \quad \text{diag} \left( B^X_Z \right) = 0,$$

(6)

where \( X = (S,T) \) denotes the domain index, \( Z = (U, V) \) denotes the user and item set, \( \text{diag}(B^X_Z) = 0 \) is a constant that avoids the trivial solution, \( \nu \) is the balance hyper parameter, and \( \| \cdot \|_F \), is the nuclear-norm which can make the matrix to become low rank [10]. The low rank constraint can also enhance the robustness and generalization of \( B^X_Z \). Considering that \( \left\| B^X_Z \right\|_F = \text{Tr} \left( \sqrt{(B^X_Z)^T B^X_Z} \right) \), the original optimization problem Equation (6) can be rewritten as:

$$\min_{B^X_Z, \Phi} \frac{1}{2} \left\| X^X - X^X \cdot B^X_Z \right\|_2^2 + \nu \text{Tr} \left( (B^X_Z)^T \Phi B^X_Z \right) \quad \text{s.t.} \quad \text{diag} \left( B^X_Z \right) = 0, \Phi = \left( B^X_Z (B^X_Z)^T \right)^{-\frac{1}{2}}.$$

(7)

Alternatively updating \( B^X_Z \) and \( \Phi \) can solve Equation (7) efficiently. For \( B^X_Z \), it has the following closed-form solution:

$$B^X_Z \left| \begin{array}{c} i = j \\ i \neq j \\ \end{array} \right. \left| \begin{array}{c} 0 \\ \text{Tr} \left( \frac{\theta_{ij}}{\theta_{jj}} \right) \\ \end{array} \right.$$  

(8)

where \( \Theta = ((Z^X)^T \cdot X^X + \nu (\Phi + \Phi^T))^{-1} \). After we have updated \( B^X_Z \), we fix it as a constant and update \( \Phi \) through the equality constraint \( \Phi = \left( B^X_Z (B^X_Z)^T \right)^{-\frac{1}{2}} \). By utilizing the iterative updating method until it converges, we can obtain the results of \( B^X_Z \) and \( \Phi \). The optimization details will be provided in Appendix B. Notably that \( B^X_Z \) may be asymmetric and contain negative values, we build up the attribution graph through the average value of \( |B^X_Z| \) and \( |(B^X_Z)^T| \) which can be depicted as \( A^X_Z = (|B^X_Z| + |(B^X_Z)^T|) / 2 \), where \( (A^X_Z)_{ij} \) depicts the attribution similarity on the \( i \)-th and \( j \)-th attribution [23]. Meanwhile \( A^X_Z \) can be viewed as the graph adjacent matrix where each node denotes the corresponding attribution. Therefore, we can establish the attribution graph to represent the topology structure among these attributions, as shown in Fig. 4(a) and Fig. 4(c) in the source and target domains.

**Attribution Graph Alignment.** After we obtain the attribution graph through the adjacent matrix \( A^X_Z \), we can match the source and target attribution graph across domains in Fig. 4(b). To start with, we first calculate the corresponding Laplacian matrix \( L^S_Z = D^S_Z - A^S_Z \) and \( L^T_Z = D^T_Z - A^T_Z \) through the origin adjacent matrix where \( D^S_Z \) and \( D^T_Z \) denote the degree matrix on graph.

According to previous research [37], a graph can be interpreted as a multivariate normal distribution. Specifically, the mean of the multivariate normal distribution is zero and the covariance is the inverse Laplacian matrix \( (L^S_Z)^+ \) and \( (L^T_Z)^+ \), as shown below:

$$P(G^S_Z) = N(0, (L^S_Z)^+), \quad P(G^T_Z) = N(0, (L^T_Z)^+).$$

(9)

The above formulation has been further used in many graph inference algorithms to represent the graph characteristics. Therefore, we can calculate the attribution graph distance and reduce the discrepancy based on the formulation. We adopt the Wasserstein distance [27] to measure the distance between different Gaussians distributions, which is shown as:

$$d_W \left( N(0, (L^S_Z)^+), N(0, (L^T_Z)^+) \right) = \text{Tr} \left( (L^S_Z)^+ + (L^T_Z)^+ - 2((L^S_Z)^+ (L^T_Z)^+)^{\frac{1}{2}} \right).$$

(10)

Meanwhile the Wasserstein distance on graph is more sensitive than traditional Frobenius norm distance which can detect more subtle modification [29]. Therefore we propose the attribution map distance alignment loss as below:

$$L_A = d_W \left( P(G^S_Z), P(G^T_Z) \right) + d_W \left( P(G^S_Z), P(G^T_Z) \right).$$

(11)

The relevance between different attributions will finally meet the consensus through attribution graph alignment, as shown in Fig. 4(b). In summary, vertical and horizontal attribution alignment are both indispensable, as shown in Fig. 5. With the vertical attribution alignment, one can just align the marginal distribution across domains, as shown in Fig 5(b). Otherwise, the horizontal attribution alignment can only align the attribution relationship but cannot reduce the distribution discrepancy as shown in Fig 5(c). Therefore, two methods can complement each other and work together to complete the task, as shown in Fig. 5(d).
3.3 Putting Together
The total loss of CFAA could be obtained by combining the losses of the rating prediction module and the embedding attribution alignment module. That is, the loss of CFAA is given as:

\[ L_{\text{CFAA}} = L_C + \lambda_O L_O + \lambda_A L_A, \quad (12) \]

where \( \lambda_O \) and \( \lambda_A \) are hyper-parameters to balance different type of losses. By doing this, CFAA can not only model the source user-item interactions, but also reduce the embedding bias and discrepancy on user and item across domains.

4 EMPIRICAL STUDY
In this section, we conduct experiments on several real-world datasets to answer the following questions: (1) \textbf{RQ1}: How does our approach perform compared with the state-of-the-art CDR methods? (2) \textbf{RQ2}: How do the vertical and horizontal attribution alignments contribute to performance improvement? (3) \textbf{RQ3}: How does the performance of CFAA vary with different values of the hyper-parameters?

4.1 Datasets and Tasks
We conduct extensive experiments on two popularly used real-world datasets, i.e., Douban and Amazon. First, the Douban dataset [53, 55] has three domains, i.e., Book, Music, and Movie. Second, the Amazon dataset [31, 51] has five domains, i.e., Movies and TV (Movie), Books (Book), CDs and Vinyl (Music), Instant Videos (Video) and Clothes (Clothes). Both datasets have user-item ratings and reviews. The detailed statistics of these datasets after pre-process are shown in Appendix C. We select the relative large datasets (e.g., Amazon Movie, Douban Book) as the source domains and the rest as the target domains. We remove the users and items less than 30 records to increase the density following existing research in the source domain [48]. Meanwhile we also delete some part of interactions in the target datasets to make them more sparse. For each datasets, we binarize the ratings to 0 and 1. Specifically, we take the ratings higher or equal to 4 as 1 and others as 0. Therefore we conduct several tasks that transferring the useful knowledge from the source to the target domains. Notably that it includes both easy and hard tasks, e.g., the transfer task between Amazon Movie \( \rightarrow \) Amazon Video is easy since they are rather similar, while Amazon Movie \( \rightarrow \) Amazon Clothes is hard because they are different.

4.2 Experiment Settings
We randomly divide the observed source and target data into training, validation, and test sets with a ratio of 8:1:1. Users and items are selected to be both non-overlapped across domains. We set batch size \( N = 256 \) for both the source and target domains. The latent embedding dimension is set to \( D = 300 \). For the English reviews on Amazon, we apply the pretrained transformer based on RoBERTa-Large for semantic textual similarity task [26]. For the Chinese reviews on Douban, we apply the pretrained Chinese BERT [13]. We set \( K = \frac{N}{2} \) for typical sample selection in attribution distribution alignment. For CFAA model, we set the balance hyper-parameters as \( \lambda_O = 0.5 \) and \( \lambda_A = 0.8 \). We set the hyper-parameters \( \alpha = 0.1 \) and \( \nu = 0.1 \) for solving the typical sample selection method and attribution subspace modelling respectively. For all the experiments, we perform five random experiments and report the average results. We choose Adam [18] as optimizer, and adopt Hit Rate@\( k \) (HR@\( k \)), Recall@\( k \), and NDCG@\( k \) [45] as the ranking evaluation metrics with \( k = 10 \).

4.3 Baseline
We compare our proposed CFAA with the following state-of-the-art recommendation models. (1) \textbf{DeepCoNN} [52] Deep Cooperative Neural Networks (DeepCoNN) is the first deep collaborative model to leverage both user and item textual features from reviews for recommendation. (2) \textbf{NARRE} [7] Neural Attentional Rating Regression with Explanations (NARRE) utilizes two parallel CNNs with attention mechanism to extract review-level information for recommendation. (3) \textbf{ESCOFILT} [36] Extractive Summarization-Based Collaborative Filtering (ESCOFILT) is a BERT-based state-of-the-art collaborative filtering model based on user and item review information. (4) \textbf{Rec-DAN} [44] Discriminative Adversarial Networks for Recommender System (Rec-DAN) adopts adversarial training strategy to align the joint user-item textual features to transfer useful knowledge. (5) \textbf{TDAR} [48] Text-enhanced Domain Adaptation (TDAR) is the state-of-the-art reviewed-based non-overlapped CDR model which adopts adversarial training strategy to user and item embeddings respectively with text memory network. (6) \textbf{ESAM} [8] Entire Space Adaptation Model (ESAM) adopts attribute correlation alignment to improve long-tail recommendation performance by suppressing inconsistent distribution between displayed and non-displayed items. (7) \textbf{DARRec} [50] Deep Domain Adaptation for Cross-Domain Recommendation via Transferring Rating Patterns (DARRec) adopts adversarial training strategy to extract and transfer knowledge patterns for shared users across domains. Note that the original ESAM and DARRec models cannot be directly applied to RNCDF tasks, and thus we adopt the same rating prediction module as CFAA for them. Besides, for a fair comparison, all the models use the same types of data and pre-processing methods during experiments.

4.4 Recommendation Performance (RQ1)
Results and discussion. The comparison results on Douban and Amazon datasets are shown in Table 1. Note that CFAA-Base represents the model that only adopts the rating prediction module for collaborative filtering without embedding attribution alignment. From them, we can find that: (1) Only adopting the single target domain information (e.g., DeepCoNN) cannot obtain satisfying results under the RNCDF settings due to the data sparsity problem. (2) CFAA-Base outperforms the previous reviewed-based recommendation model (e.g., ESCOFILT), indicating that adopting BERT-based review embedding with one-hot ID and multi-hot historical rating information can efficiently enhance the model performance. (3) Although ESAM achieves better performance than CFAA-Base due to its attribution correlation congruence, it still fails to capture the nonlinear and complex topology structure among different attributions. (4) Although DARRec with gradient reverse layer can obtain good performance, the unstable adversarial training may hinder the model to obtain more accurate results [38]. As a result, it will eventually cause coarsely matching across domains.
### Table 1: Experimental results on Douban and Amazon datasets.

| Source Domain        | Target Domain        | Model          | HR Recall | NDCG | HR Recall | NDCG | HR Recall | NDCG | HR Recall | NDCG |
|----------------------|----------------------|----------------|-----------|------|-----------|------|-----------|------|-----------|------|
| (Amazon) Movie → Video | (Amazon) Movie → Music | DeepCoNN      | .1459     | .1338 | .0944     | .1476 | .1588     | .0625 | .1039     | .1135 |
|                      |                      | NARRE         | .1538     | .1413 | .1032     | .1573 | .1707     | .0948 | .1346     | .1270 |
|                      |                      | Rec-GAN       | .1601     | .1692 | .1065     | .1787 | .1876     | .1193 | .1560     | .1402 |
|                      |                      | TDAR          | .1782     | .1786 | .1197     | .1915 | .2043     | .1485 | .1674     | .1461 |
|                      |                      | ESCOFILT      | .1896     | .1977 | .1304     | .2024 | .2115     | .1567 | .1751     | .1504 |
|                      |                      | DARec         | .2044     | .2220 | .1651     | .2462 | .2417     | .1709 | .1846     | .1708 |
|                      |                      | ESAM          | .2067     | .2275 | .1696     | .2408 | .2391     | .1680 | .1870     | .1715 |
| (Amazon) Movie → Clothes | (Amazon) Book → Video | CFAA-Base     | .2013     | .2168 | .1569     | .2316 | .2301     | .1624 | .1832     | .1686 |
|                      |                      | CFAA-V        | .2083     | .2263 | .1683     | .2440 | .2514     | .1736 | .1865     | .1713 |
|                      |                      | CFAA-H        | .2105     | .2205 | .1722     | .2391 | .2468     | .1705 | .1891     | .1735 |
|                      |                      | CFAA          | .2190     | .2414 | .1808     | .2552 | .2637     | .1901 | .1944     | .1798 |
|                      |                      | (Amazon) Book → Music | DeepCoNN | .1169 | .1624 | .0706 | .0939 | .1085 | .0439 | .1868 | .1540 | .0787 |
|                      |                      | NARRE         | .1582     | .1696 | .0933     | .1148 | .1137     | .0504 | .2037     | .1669 | .0944 | .1576 | .1580 | .0962 |
|                      |                      | Rec-GAN       | .1599     | .1817 | .1228     | .1261 | .1206     | .0613 | .2204     | .1782 | .1193 | .1798 | .1636 | .0984 |
|                      |                      | TDAR          | .1708     | .2032 | .1414     | .1465 | .1291     | .0688 | .2365     | .2016 | .1328 | .1913 | .1792 | .1109 |
|                      |                      | ESCOFILT      | .1784     | .2108 | .1570     | .1587 | .1369     | .0745 | .2396     | .2103 | .1399 | .2057 | .1904 | .1250 |
|                      |                      | DARec         | .1897     | .2295 | .1840     | .1694 | .1501     | .0873 | .2562     | .2415 | .1638 | .2203 | .2101 | .1405 |
|                      |                      | ESAM          | .1915     | .2321 | .1909     | .1712 | .1530     | .0897 | .2589     | .2366 | .1544 | .2219 | .2133 | .1393 |
|                      |                      | CFAA-Base     | .1860     | .2265 | .1761     | .1670 | .1488     | .0854 | .2530     | .2327 | .1502 | .2194 | .2075 | .1368 |
|                      |                      | CFAA-V        | .1924     | .2340 | .1912     | .1699 | .1522     | .0910 | .2628     | .2410 | .1713 | .2250 | .2146 | .1421 |
|                      |                      | CFAA-H        | .1913     | .2319 | .1876     | .1725 | .1543     | .0941 | .2605     | .2424 | .1690 | .2236 | .2137 | .1434 |
|                      |                      | CFAA          | .2021     | .2493 | .2015     | .1782 | .1595     | .1016 | .2715     | .2548 | .1821 | .2321 | .2210 | .1496 |

Figure 6: The t-SNE visualization of user latent embeddings on Douban Movie → Douban Music (first row) and item latent embeddings (second row). The user (item) latent embeddings in the source domain are shown with red dots and that in the target domain are shown with blue dots.

Meanwhile although DARec and TDAR both utilizes adversarial training, DARec equipped with more expressive user and item embeddings can obtain more delightful results. (5) CFAA consistently achieves the best performance, which proves that embedding
Figure 7: (a)-(b) show the effect of hyper-parameters $\lambda_O$ and $\lambda_A$ on model performance.

Visualization. To better show the user embeddings across domains, we visualize the t-SNE embeddings [19] for CFAA-Base, DARec, ESAM, and CFAA. The results of Douban Movie $\rightarrow$ Douban Music are shown in Fig. 6(a)-(d). The first and second row denote user and item embeddings, respectively. From it, we can see that (1) CFAA-Base cannot reduce the embedding bias and discrepancy on both users and items across domains, leading to insufficient knowledge transfer, as shown in Fig. 6(a). (2) ESAM and DARec can marginally align user and item embedding attributions to a certain extent, but there still exists domain discrepancy which causes negative transfer, as shown in Fig. 6(b)-(c). (3) CFAA with attribution distribution and relevance alignment can better match users and items across domains, as shown in Fig. 6(d). The visualization results illustrates the validity of our model.

4.5 Analysis (RQ2 and RQ3)

Ablation. To study how does each module of CFAA contribute on the final performance, we compare CFAA with its several variants, including CFAA-V and CFAA-H. CFAA-V only adapts the vertical attribution (i.e., attribution distribution) alignment while CFAA-H only adopts the horizontal attribution (i.e., attribution relevance) alignment. The comparison results are shown in Table 1. From it, we can observe that (1) CFAA-V and CFAA-H always get more accurate output predictions than CFAA-Base, which indicates that reducing embedding attribution alignment is essential. (2) However, CFAA-V and CFAA-H still cannot achieve the best results against CFAA. Simply aligning vertical embedding probability distribution on CFAA-V will sometimes neglect the attribution relevance and cause negative transfer. Likewise, only concentrating on the horizontal attribution alignment will sometimes ignore the attribution discrepancy across domains. Overall, the above ablation study demonstrates that our proposed embedding alignment module is effective in solving the RNCDD problem.

Distribution Discrepancy. The domain adaptation theory [3] suggests proxy $\mathcal{A}$-distance as a measure of cross-domain discrepancy. We adopt $d_{\mathcal{A}}(\mathcal{S}, \mathcal{T}) = 2(1 - e(h))$ to analysis the distance between two domains, where $e(h)$ is the generalization error of a linear classifier $h$ that discriminates the source domain $\mathcal{S}$ and the target domain $\mathcal{T}$ [3]. Table 2 demonstrates the domain discrepancy on Amazon Movie $\rightarrow$ Amazon Music and Douban Movie $\rightarrow$ Douban Music tasks using CFAA-Base, CFAA-V, CFAA-H, CFAA, and several baseline methods. From it, we can conclude that: (1) The large number of $d_{\mathcal{A}}$ on CFAA-Base indicates the existence of embedding attribution discrepancy between the source and target domains. (2) Most of the current baselines (e.g., ESAM) can reduce $d_{\mathcal{A}}$ but they still cannot obtain lower $d_{\mathcal{A}}$ than CFAA, indicating their limitations on domain adaptation. (3) By adopting the embedding attribution alignment methods, CFAA-V and CFAA-H can reduce the embedding attribution discrepancy according to the $d_{\mathcal{A}}$. Furthermore, we can observe that CFAA achieves the lowest $d_{\mathcal{A}}$ distance, because it considers both vertical and horizontal embedding attribution alignment. Lower $d_{\mathcal{A}}$ can always obtained better results since more useful knowledge can be transferred from source to target domains, which is consistent with the previous ablation study.

Effect of hyper-parameters. We finally study the effects of hyper-parameters on model performance. For CFAA, we vary $\lambda_O$ and $\lambda_A$ in $\{0, 0.1, 0.2, 0.5, 0.8, 1, 2, 5, 8\}$ and report the results in Fig. 7(a)-(b) on Amazon Movie $\rightarrow$ Amazon Music. Fig. 7(a)-(b) show the bell-shaped curves, indicating that choosing the proper hyper-parameters to balance the rating prediction loss and embedding alignment loss can effectively improve the model performance. Empirically, we choose $\lambda_O = 0.5$ and $\lambda_A = 0.8$.

5 CONCLUSION

In this paper, we propose Collaborative Filtering with Attribution Alignment model for solving review-based non-overlapped cross domain recommendation (CFAA), which includes the rating prediction module and the embedding attribution alignment module. We innovatively adopt horizontal and vertical attribution alignment to better reduce the embedding discrepancy from different perspectives. Vertical distribution alignment utilizes the typical sample selection with optimal transport to make them consistent across domains. Horizontal relevance alignment applies the subspace modelling with attribution graph alignment to reduce the discrepancy. We also conduct extensive experiments to demonstrate the superior performance of our proposed CFAA on several datasets and tasks.

ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (No.72192823 and No.62172362).
A TYPICAL SAMPLE SELECTION METHOD

As mentioned in Section 3.2.1, the typical sample selection algorithm is given by:

\[
\min t_\Omega = \sum_{i=1}^{N} \sum_{j=1}^{K} (\Psi_{q}^{X})_{ij} (Z_{iq}^X - (M_{Z}^{X})_{jq})^2 + \alpha \cdot \mathcal{R}((\Psi_{q}^{X})_{ij})
\]

s.t. \((\Psi_{q}^{X})_{1} = 1, \mathcal{R}((\Psi_{q}^{X})_{ij}) = (\Psi_{q}^{X})_{ij} \log((\Psi_{q}^{X})_{ij}), (\Psi_{q}^{X})_{ij} > 0.\) 

(13)

We now provide the optimization details on the typical-proxies algorithm. Alternatively updating \(M_{Z}^{X}\) and \(\Psi_{q}^{X}\) can solve Equation (13) efficiently.

**Update \(\Psi_{q}^{X}\).** We first fix \(M_{Z}^{X}\) and update \(\Psi_{q}^{X}\). By using Lagrangian multiplier to minimize the objective function, we have:

\[
\ell_{Q} = \sum_{i=1}^{N} \sum_{j=1}^{K} (\Psi_{q}^{X})_{ij} \zeta_{ij} + \alpha \cdot \mathcal{R}((\Psi_{q}^{X})_{ij}) + \chi \sum_{i=1}^{N} \sum_{j=1}^{K} (\Psi_{q}^{X})_{ij} - 1. 
\]

where \(\zeta_{ij} = (Z_{iq}^{X} - (M_{Z}^{X})_{jq})^2\) and \(\chi\) is the Lagrangian multiplier.

Taking the differentiation of Equation (14) w.r.t. \((\Psi_{q}^{X})_{ij}\) and setting it to 0, we obtain:

\[
\frac{\partial \ell_{Q}}{\partial (\Psi_{q}^{X})_{ij}} = \zeta_{ij} + \alpha (\log((\Psi_{q}^{X})_{ij}) + 1) + \chi = 0.
\]

(15)

By solving and simplifying Equation (15), we have:

\[
(\Psi_{q}^{X})_{ij} = \exp\left(-\frac{\alpha + \chi}{\alpha} \exp\left(-\frac{\zeta_{ij}}{\alpha}\right)\right). 
\]

(16)

Meanwhile, taking \(\sum_{j=1}^{K} (\Psi_{q}^{X})_{ij} = 1\) into Equation (16), we have:

\[
\sum_{j=1}^{K} \exp\left(-\frac{\alpha + \chi}{\alpha} \exp\left(-\frac{\zeta_{ij}}{\alpha}\right)\right) = \exp\left(-\frac{\alpha + \chi}{\alpha} \sum_{j=1}^{K} \exp\left(-\frac{\zeta_{ij}}{\alpha}\right)\right) = 1. 
\]

(17)

That is,

\[
\exp\left(-\frac{\alpha + \chi}{\alpha} \right) = \frac{1}{\sum_{j=1}^{K} \exp\left(-\frac{\zeta_{ij}}{\alpha}\right)}. 
\]

(18)

Thus, the final solution of \((\Psi_{q}^{X})_{ij}\) is given by:

\[
(\Psi_{q}^{X})_{ij} = \exp\left(-\frac{-\zeta_{ij}/\alpha}{\sum_{k=1}^{K} \exp(-\zeta_{ik}/\alpha)}\right). 
\]

(19)

**Update \(M_{Z}^{X}\).** After we have updated \(\Psi_{q}^{X}\), we fix it as a constant and update \(M_{Z}^{X}\). Thus, Equation (14) becomes

\[
\min_{M_{Z}^{X}} \sum_{i=1}^{N} \sum_{j=1}^{K} (\Psi_{q}^{X})_{ij} ||Z_{iq}^{X} - (M_{Z}^{X})_{jq}\|^2. 
\]

(20)

Taking the differentiation of Equation (20) w.r.t. \(M_{Z}^{X}\) and setting it to 0, we can update \(M_{Z}^{X}\) as:

\[
(M_{Z}^{X})_{jq} = \frac{\sum_{i=1}^{N} (\Psi_{q}^{X})_{ij} Z_{iq}^{X}}{\sum_{i=1}^{N} (\Psi_{q}^{X})_{ij}}. 
\]

(21)

We can obtain the stable solution of \(\Psi_{q}^{X}\) and \(M_{Z}^{X}\) through several iterations.

B ATTRIBUTION SUBSPACE MODELLING

As mentioned in Section 3.2.2, the attribution subspace modelling algorithm is given by:

\[
\min_{B_{Y}^{X}} \frac{1}{2} ||Z_{iq}^{X} - Z^{X} B_{Y}^{X}||^2_2 + \nu \mathbf{Tr}\left((B_{Y}^{X})^T \Phi B_{Y}^{X}\right) 
\]

s.t. \(\mathbf{diag}(B_{Y}^{X}) = 0, \Phi = (B_{Y}^{X})^T (B_{Y}^{X})^{-1}.\) 

(22)

Alternatively updating \(B_{Y}^{X}\) and \(\Phi\) can solve Equation (22) efficiently.

We first fix \(\Phi\) and update \(B_{Y}^{X}\). By using Lagrangian multiplier to minimize the objective function, we have:

\[
\ell_{B} = \frac{1}{2} ||Z_{iq}^{X} - Z^{X} B_{Y}^{X}||^2_2 + \nu \mathbf{Tr}\left((B_{Y}^{X})^T \Phi B_{Y}^{X}\right) + \psi \mathbf{diag}(B_{Y}^{X}). 
\]

(23)

Taking the differentiation of Equation (23) w.r.t. \(B_{Y}^{X}\) and setting it to 0, we obtain:

\[
\frac{\partial \ell_{B}}{\partial B_{Y}^{X}} = (Z_{iq}^{X} - Z^{X} B_{Y}^{X})^T (Z^{X} B_{Y}^{X} - Z^{X}) + \nu B_{Y}^{X} + \psi \mathbf{diag}(B_{Y}^{X}) = 0, 
\]

(24)

where \(\Phi + \Phi^T = \Xi\), \(\mathbf{diag}(\cdot)\) denotes the diagonal matrix. By solving and simplifying Equation (24), we have:

\[
B_{Y}^{X} = (Z^{X})^T (Z^{X} B_{Y}^{X} - Z^{X}) + \nu \Xi B_{Y}^{X} + \psi \mathbf{diag}(\Xi). 
\]

(25)

Here, we suppose that \(\nu\) is always sufficient large and \((Z^{X})^T (Z^{X} + \nu \Xi)^{-1}\) is invertible [39]. We define \(\Theta = ((Z^{X})^T (Z^{X} + \nu \Xi)^{-1}\) and substitute it into Equation (24):

\[
B_{Y}^{X} = I - \Theta \cdot \mathbf{diag}((\nu + \psi)). 
\]

(26)

The unknown value of \(\gamma\) can be solved by the diagonal constraint \(\mathbf{diag}(B_{Y}^{X}) = 0\). We can obtain that \(\gamma = \mathbf{diag}(\Xi)\), where \(\Theta\) denotes the elementwise division. Thus, the final solution of \(B_{Y}^{X}\) is given by:

\[
(B_{Y}^{X})_{ij} = \begin{cases} 0, & i = j \\ \Theta_{ij}, & \text{Others}. \end{cases} 
\]

(27)

After we have updated \(B_{Y}^{X}\), we fix it as a constant and update \(\Phi\) through the equality constraint \(\Phi = (B_{Y}^{X})^T (B_{Y}^{X})^{-1}.\) Utilizing this iteration method until it converges, we can obtain the results of \(B_{Y}^{X}\) and \(\Phi\).

C DATASETS AND TASKS

We conduct extensive experiments on two popularly used real-world datasets, i.e., Douban and Amazon. The Douban dataset includes Book, Music, and Movie and the Amazon dataset has five domains, i.e., Movies and TV (Movie), Books (Book), CDs and Vinyl (Music), Instant Videos (Video) and Clothes (Clothes). The detailed statistics of these datasets after pre-process are shown in Table 3.

---

**Table 3: Statistics on Douban and Amazon datasets.**

| Dataset       | Items | Users | Interactions | Density |
|---------------|-------|-------|--------------|---------|
| Amazon Movie  | 50,052| 123,960| 1,697,532    | 0.020%  |
| Amazon Book   | 43,168| 95,643 | 1,083,019    | 0.025%  |
| Douban Movie  | 34,893| 151,258| 1,278,401    | 0.024%  |
| Douban Book   | 38,776| 111,270| 965,041      | 0.022%  |
| Amazon Music  | 7,710 | 11,053 | 106,188      | 0.124%  |
| Amazon Video  | 1,580 | 4,555  | 11,137       | 0.155%  |
| Amazon Clothes| 21,554| 35,669 | 89,176       | 0.012%  |
| Douban Music  | 3,562 | 11,278 | 2,1451       | 0.053%  |