Enhancing Domain-Level and User-Level Adaptivity in Diversified Recommendation

Yile Liang[1], Tieyun Qian[1,*], Qing Li[2], Hongzhi Yin[3,*]

[1]School of Computer Science, Wuhan University, Wuhan, China

[2]Department of Computing, Hong Kong Polytechnic University, Hong Kong, China

[3]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia

[1]{liangyile, qty}@whu.edu.cn, [2]qing-prof.li@polyu.edu.hk, [3]h.yin1@uq.edu.au

ABSTRACT

Recommender systems are playing a vital role in online platforms due to the ability of incorporating users' personal tastes. Beyond accuracy, diversity has been recognized as a key factor to broaden users' horizons as well as to promote enterprises' sales. However, the trade-off between accuracy and diversity remains to be a big challenge. More importantly, none of existing methods has explored the domain and user biases toward diversity.

In this paper, we focus on enhancing both domain-level and user-level adaptivity in diversified recommendation. Specifically, we first encode *domain-level diversity* into a generalized bi-lateral branch network with an adaptive balancing strategy. We further capture *user-level diversity* by developing a two-way adaptive metric learning backbone network inside each branch. We conduct extensive experiments on three real-world datasets. Results demonstrate that our proposed approach consistently outperforms the state-of-the-art baselines.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

recommender systems, diversified recommendation, metric learning, bilateral branch network

1 INTRODUCTION

Recommender systems have been widely deployed in many web services for addressing the information overload issues. The traditional recommendation approaches can be broadly classified into
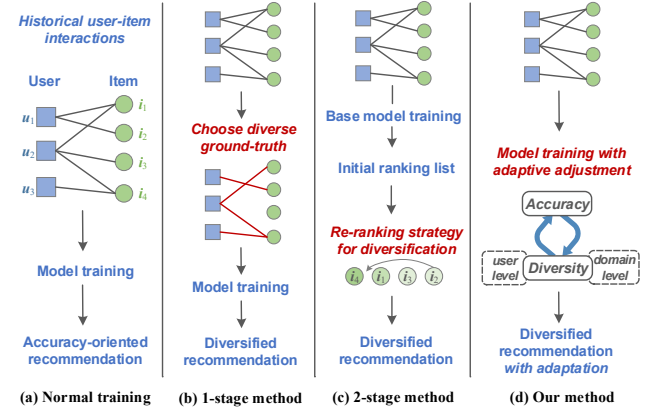
---

*Corresponding author.

Figure 1: The comparison among (a) normal training; (b) one-stage diversified recommendation, (c) two-stage; and (d) our method equipped with adaptation ability.

two main lines, i.e., collaborative filtering (CF) based and content based methods. CF based methods utilize historical interactions (e.g., clicks and purchases) to infer the user's personal preference and recommend items according to the matching score between the user and the target item. Content based methods [10, 32] recommend items by measuring content similarity (e.g., item attributes and text contents) between the user's historical items and the target item. Recent years have witnessed the success of deep learning-based methods like neural CF and neural matrix factorization [6, 16, 17].

While being one of the most successful applications in online systems, (neural) CF based recommendation favors popular items and may recommend items that are already known to the user [2], and content based recommendation may produce items matching the user's interests but covering a narrow scope of topics [33]. To tackle these problems, diversity-promoting (a.k.a., diversified) recommendation has been an active research topic in recent years as it helps to broaden users' horizons and increase e-commerce firms' sales [11, 36].

Diversified recommendation can be treated as a two-objective optimization problem, i.e., to maximize the overall relevance of a recommendation list and to minimize the similarity among the items in the list. Early work in this field often [2, 4, 33, 41] adopts a two-stage (post-processing) strategy that first generates a candidate set based on the accuracy metric and then selects a few items by maximizing the diversity metric. Some recent approaches [8, 14, 25, 37] apply determinantal point process (DPP) to modeling set diversity, and a few others [11, 34] improve recommendation

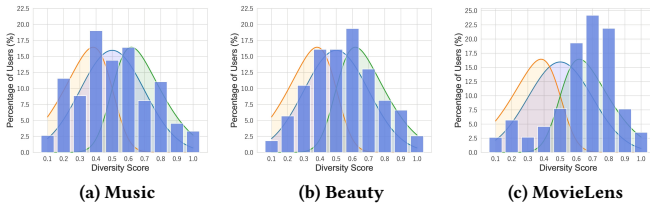**(a) Music** **(b) Beauty** **(c) MovieLens**

Figure 2: The distribution of users' diversity in three domains. The green, blue, and yellow curves denote the left-skewed, normal, and right-skewed distribution.

diversity in the manner of end-to-end (one-stage) supervised learning. Figure 1 (a-c) depicts the outline of the normal training, and that of one- and two-stage diversified recommendation methods. Despite their effectiveness, existing diversified recommendation approaches have the following inherent limitations.

- L1: None of existing methods has considered the domain and user biases. A diversity-promoting system may recommend a wide range of items to all users even if the target user has narrow and focused interests. Such a system will not change the recommendation strategy for different domains either.
- L2: Both the post-processing and DPP based approaches [2, 8, 25, 33, 37] require extra parameter tuning to balance accuracy and diversity. The end-to-end method DCF [11] also needs to select a proper trade-off parameter when generating the ground-truth for training.
- L3: Current studies only treat the recommendation as a task of finding items for the user, but the other side of the coin is that the products often have their own market orientations when they are designed and produced, which should not be disregarded.

Data-Driven Study. As an example, we plot the user's diversity distribution in three domains in Figure 2. The diversity score for each user is calculated by the number of interacted categories divided by number of interacted items. The higher score, the richer diversity. We can observe that the distribution is different in different domains. It is interesting that a majority of users have a great deal of variability when choosing the movies' category while many users have relatively fixed preferences for the music genre. We term this *the domain-level diversity*. We also find that each user has her own diversity preference even in the same domain. For instance, on MovieLens, a few users have a narrow diversity though most of them have a wide one. We term this *the user-level diversity*.

The domain-level and user-level diversity reflect the characteristics of the domain data and the individual user, respectively, and should be presented in the recommender systems. Equally treating all domains or equally treating all users will significantly decrease recommendation performance and affect user experience. Unfortunately, none of existing methods has taken these two types of diversity into account. As a result, they are unable to tackle the problem of domain and user biases.

In this paper, we propose a novel framework to focus on enhancing the domain-level and user-level adaptivity (EDUA), meanwhile

addressing the laborious parameter tuning and product orientation problems (Figure 1 (d)).

Our main idea is inspired by the bilateral branch network [43] in vision community. The architecture of two separate branches naturally equips the model with the ability of trading-off accuracy and diversity without extensive tuning of the trade-off parameter (corresponding to L2).

We move two steps further in that we first propose *an adaptive balancing strategy* which encodes the domain-level diversity by automatically determining the order of learning focus between bilateral branches (corresponding to L1). Moreover, we design *a two-way adaptive metric learning backbone network* inside each branch, which captures both the user's interests in specific items and the item's orientation towards target users (corresponding to L3), and then treats the user-level diversity as a special relation connecting the user and the item (corresponding to L1).

To summarize, the major contributions of this paper are:

- We highlight the importance of modeling both domain-level and user-level diversity in recommender systems. To the best of our knowledge, this is the first attempt to uncover and exploit these two types of diversity for recommendation.
- We generalize the bilateral branch network structure with an adaptive balancing strategy and a two-way adaptive metric learning backbone, which successfully overcome the limitations in existing methods.
- Extensive experimental results on three real-world datasets demonstrate that our proposed model yields significant improvements over the state-of-the-art baselines for the trade-off performance combining both accuracy and diversity.

2 RELATED WORK

2.1 Diversified Recommendation

Diversified recommendation can be viewed from both the aggregate and individual perspectives [38].

1) Aggregate diversity refers to the diversity of recommendations across all users [1]. A system with high aggregate diversity will recommend a majority number (if not all) of items out of the entire item pool. Since the classic Pareto principle reveals that the top 20% products often generate 80% of sales, current work on aggregate diversity pays more attention to long-tail items. For example, a few studies [1, 27] improve aggregate diversity by countering the effects of item popularity. Other studies [30, 40] propose clustering approaches and attempt to leverage long-tail items directly in a recommendation list. A recent work [20] constructs a pseudo ground-truth ranking order by clustering and relocating the tailed items.

2) Individual diversity refers to the diversity of recommendations for each target user [38]. The goal for individual diversity based recommendation is to achieve a balance between accuracy and diversity. Current research on individual diversity can be categorized into three groups.

The first group of methods adopts a post-processing heuristics. The classical MMR [4] approach represents relevance and diversity with independent metrics and maximizes the marginal relevance

with a trade-off parameter. Other heuristic methods [2, 33] maximize a submodular objective function to learn the diversified model.

The second group of methods is based on the determinantal point process (DPP) which measures set diversity by describing the probability for all subsets of the item set. DPP generates diverse lists through the maximum a posteriori (MAP) which is NP-hard and is computationally expensive even using the popular greedy algorithm. To address this issue, Chen et al. [8] develop a novel algorithm to accelerate the MAP inference for DPP. Based on the fast inference method [8], more recent studies [14, 25, 37] employ DPP to improve diversity in different recommendation tasks.

The third group of methods treats diversified recommendation as an end-to-end supervised learning task. Cheng et al. [11] propose the diversified collaborative filtering (DCF) to solve the coupled parameterized matrix factorization and structural learning problems. Sun et al. [34] apply bayesian graph convolutional neural networks (BGCF) to model the uncertainty between user-item and bring diversity into recommendation indirectly.

Compared with aggregate diversity, more research attention is paid on individual diversity, which is also the focus of our paper. Despite impressive progress, both post-processing and DPP methods require great efforts in balancing the trade-off between accuracy and diversity. Between two end-to-end methods, DCF [11] also needs to choose a trade-off parameter, and BGCF [34] increases diversity by adding the randomness into the data rather than explicitly modeling diversity.

2.2 Metric Learning for Recommendation

The goal of metric learning is to learn a distance metric that assigns smaller distances between similar pairs, and larger distances between dissimilar pairs. Metric learning has also been adopted in recommender systems [9, 13, 19] to address the issue that inner product similarity violates the triangle inequality.

The pioneering work CML [17] learns a joint user-item metric space for capturing fine-grained user preference. It minimizes the distance between each user-item interaction in Euclidean space, i.e., $d = \|u - v\|_2^2$, where u and v denotes a user and an item, respectively. To address the problem that the position of positive item may be close to the negative one, SML [24] proposes to measure the trilateral relationship from both user-centric and item-centric perspectives. PMLAM [26] is presented to generate adaptive margins for the training triples in the metric space.

Inspired by the relational metric learning in knowledge graph embedding, a number of methods [15, 29, 31, 39, 42] are proposed to measure the distance between user and item with $d = \|u + r - v\|_2^2$, where r is the connective relation. For example, TransCF [29] constructs the translation vectors by incorporating the neighborhood information of users and items to model the heterogeneity of user-item relationships. LRML [35] introduces the memory-based attention module to learn the relation vector.

Our framework adopts metric learning as the backbone network in each branch of BBN. Different from aforementioned translation based approaches that do not consider the diversity factor, our model aims to generate accurate and diverse recommendations. Furthermore, our model employs two-way ($u \rightarrow v$ and $v \rightarrow u$) translation while existing methods only perform $u \rightarrow v$ translation and ignore items' orientations towards users.

3 PROBLEM FORMULATION

Let $U = \{u_1, u_2, ..., u_M\}$ be a set of users and $V = \{v_1, v_2, ..., v_N\}$ be a set of items, where M and N denote the corresponding cardinalities. Let $R \in \mathbb{R}^{M \times N}$ be the user-item interaction matrix, which indicates whether the user purchased or clicked on the item. We focus on the recommendation task with implicit feedback [18], and the interaction matrix is defined as:

$$R_{uv} = \begin{cases} 1, \text{ if an interaction (user } u, \text{ item } v) \text{ is observed,} \\ 0, \text{ otherwise.} \end{cases}$$

The observed entries reflect users' interest in the item, while the unobserved entries are mixed with unknown data and negative views of the item. Given above interaction information, our main goal is to estimate users' interest for unobserved entries and rank the candidate items according to the predicted scores such that we can recommend top-K items for the traditional recommendation task.

Beyond that, we particularly focus on the *individual diversity in recommendation*, which aims to generate diversified recommendations for each target user. To be specific, on one hand, the recommendation quality is evaluated by a matching score between recommendation list and ground-truth list. On the other hand, the diversity is measured by category coverage and intra-list distance [28, 37, 44] of the list.

4 PROPOSED MODEL

4.1 Model Overview

The basic idea of Bilateral Branch network (BBN) is originated from visual recognition tasks [43]. It is designed for representation learning and classifier learning to improve the recognition performance of long-tailed data. In intuition, BBN structure enables the separation of the optimization process of two objectives, where one objective is accuracy and the other is diversity in our case. However, a direct application of BBN is inappropriate for our task.

- The fixed learning order prevents BBN from adaption for domain-level diversity. In BBN, representation learning is always the basis of classifier learning and should be trained first. In contrast, in our task, either diversity or accuracy may become the first priority depending on the domain bias.
- The conventional residual network backbone in BBN is not suitable for our task, because we need to measure the distance between the user and the item for recommendation, and we also wish to capture each user's diversity preference.

In light of this, we develop a new BBN paradigm for accurate and diversified recommendation which has the following two key properties to address the above two limitations: (1) We propose an adaptive learning strategy which encodes the domain-level diversity by automatically choosing the order of learning focus rather than the fixed learning order between bilateral branches. (2) We design a two-way adaptive metric learning backbone instead of the residual network for the recommendation task. The insight is to model the user-to-item and item-to-user relations and adaptively capture the user-level diversity simultaneously. The framework of the proposed EDUA model is shown in Figure 3.
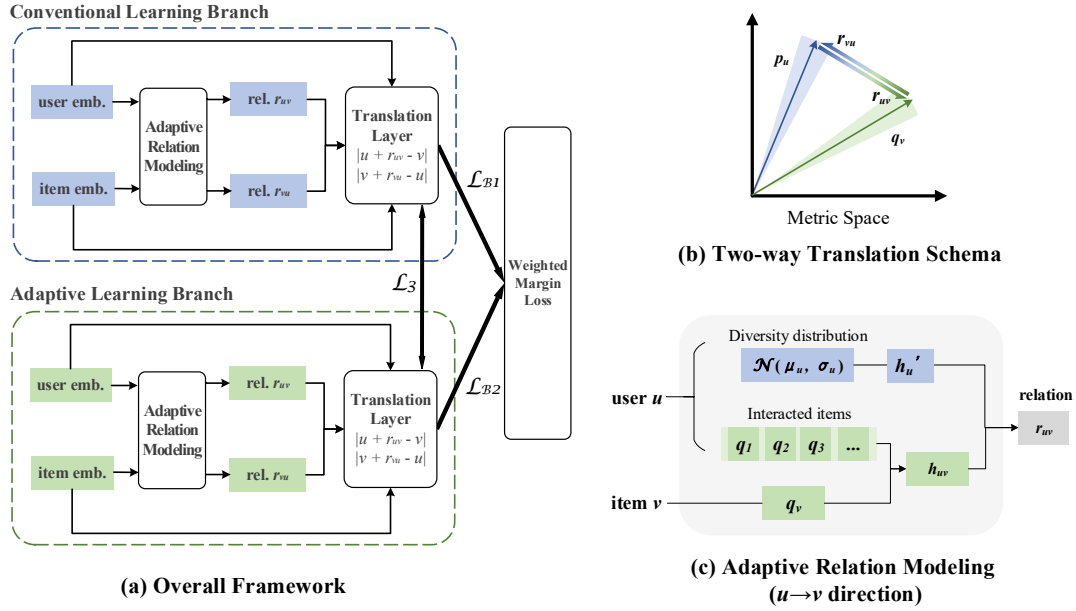
Figure 3: An overview of our EDUA model. (a) Overall framework. (b) Two-way translation scheme for modeling user-item relationship in metric space. (c) Detailed structure of adaptive relation modeling layer in the user-to-item direction. The similar process is applied to the item-to-user direction.

Concretely, we employ Bilateral Branch network as the main architecture, where one branch is used to improve accuracy in traditional recommendation and the other focuses on individual diversity as a subsidiary role, termed as "conventional learning branch" and "adaptive learning branch", respectively. We also present two-way adaptive metric learning with independent parameter space as the backbone for each branch.

4.2 Enhancing Domain-Level Adaptivity

Existing recommender systems favor popular items and produce items matching the user's interests in a narrow scope [33]. For example, if a user watched a romance film lately, the typical recommendation would be a series of redundant and similar movies. To tackle this problem, we employ BBN as the main architecture, where a conventional learning branch tries to maintain the ability to recommend relevant items, and an adaptive learning branch concentrates on the items neglected by the conventional learning. Moreover, to enhance the adaptivity for domain-level diversity, we first choose an order of learning focus for two branches in accordance with the domain bias, and then feed training samples into a corresponding branch using different sampling strategies.

4.2.1 Determining the order of learning focus in two branches. The original BBN [43] adopts the cumulative learning strategy where the order of learning focus is from representation learning to classifier learning branch by multiplying these two branches by a factor α and $1 - \alpha$. This is natural for long-tailed image classification tasks as discriminative representations are the foundation for learning robust classifiers.

However, such a fixed order of focus is inappropriate for our task since users in different domains may have various diversity

distributions. Essentially, the important branch should be trained later such that the network in this branch can be fine-tuned to get a better representation. For example, on MovieLens, most users have a wide range of diversity, and thus it is better to first train the conventional learning branch. If the adaptive learning branch for diversified recommendation is trained first on this dataset, the learned embeddings will deviate from earlier diversified representations at the later periods of conventional training. To address this issue, we propose to automatically determine the order of learning focus for balancing two branches by the domain-level diversity.

We first calculate the skewness score of the distribution X of users' diversity, which is defined as: $skewness(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$. A distribution with a zero skewness is a normal distribution, and a positive/negative distribution denotes a right/left skewed one. For a skewed/normal distribution, the model should place more emphasis on the adaptive/conventional learning branch. For this purpose, we associate the controlling factor α with the adaptive learning branch while $1 - \alpha$ with the conventional learning branch for a skewed distribution, and the order will be reversed for a normal distribution.

4.2.2 Determining the sampling strategy for each branch. In our model, the conventional learning branch is used to keep the accuracy of recommendation and adaptive learning branch for increasing the diversity. To this end, we apply uniform and reversed data sampler to the conventional and adaptive learning branch. Specifically, the uniform sampler samples each user-item pair with equal probability from all observed entries, and thus it can retain the characteristics of original distributions.

For the reversed sampler, we incorporate domain users' diversity preference by setting the sampling possibility of each user-item pair

proportional to the reciprocal of the samples in one category for the target user. Formally, suppose that $V_u^+ = \{v_1, v_2, ...\}$ is the item set interacted by a user u. Each item belongs to a category and C_u^+ is the category set of user u. The number of samples of category i in V_u^+ is N_i and the total number of samples in V_u^+ is $|V_u^+|$. We define the diversity score of the user u as $d_u = |C_u^+|/|V_u^+|$. The higher d_u, the more diverse interests the user has. We then construct the reversed sample set according to following steps: (1) Calculating the reversed and original sampling probability of category i as:

$$P_i^R = \frac{w_i}{\sum_{j=1}^{|C_u^+|} w_j}, \quad P_i^O = \frac{N_i}{|V_u^+|} \tag{1}$$

where $w_i = |V_u^+|/N_i$; (2) Sampling a category i according to the user diversity by comparing a noise $z \sim \mathcal{U}(0, 1)$ with d_u, if $z < d_u$ sampling a category using P_i^R and otherwise using P_i^O; (3) Uniformly selecting an item belonged to category i in V_u^+. In this way, we enforce the model to adaptively pay more or less attention to a broader or narrower range of categories of the specific user depending on the user's diversity preference.

4.3 Enhancing User-Level Adaptivity

Each user has her unique diversity preference which determines she is more likely to change the shopping habit or keep it as is, and recommender systems should have the ability to capture the user-level diversity. To achieve this, we propose a two-way adaptive metric learning backbone network in each branch of BBN. The basic idea of relational metric learning for recommendation is to measure the distance between the user and the item using the formula $d = \|u + r_{uv} - v\|_2^2$, where r_{uv} is the relevance relation between u and v. The two-way adaptive metric learning backbone in our framework has the following distinct properties.

Firstly, our method injects the user's/item's diversity preference into translation, i.e., r_{uv}/r_{vu} consists of a relevance relation and a diversity relation.

Secondly, our method performs both $u \rightarrow v$ and $v \rightarrow u$ translations instead of one-way ($u \rightarrow v$) translation such that the products' market orientations can also be taken into account.

4.3.1 Getting user and item embedding. We set up a lookup table to transform the one-hot representations of each user and item into low-dimensional vector in the metric space. After transformation, $p_u \in \mathbb{R}^{M \times D}, q_v \in \mathbb{R}^{N \times D}$ denotes the latent factor representation of the user u and the item v, respectively.

4.3.2 Adaptively measuring the relevance relation. We design a novel attentive matching method to adaptively measure the relevance relation between user u and item v in the metric space. Given the user-item pair (u, v), we learn the relevance relation vector h_{uv} and h_{vu} by selecting relevant pieces of historical behaviors. Suppose that $V_u^+ = \{v_1, v_2, ..., v_{L_u}\}$ are interacted items of user u, the relevance relation performs as a weighted sum to adaptively aggregate representation:

$$h_{uv} = f(p_u, q_{v_1}, ..., q_{v_{L_u}}) = \sum_{j \in V_u^+} a_{u,j} q_j. \tag{2}$$

Each element of the attention vector a_u can be defined as:

$$a_{u,j} = softmax(p_u^\top W_a q_j), \tag{3}$$

where $W_a \in \mathbb{R}^{D \times D}$ is the weight matrix to be learned, and the $softmax(.)$ is applied to the original attention weights for normalization. Similar operation can be conducted by extracting relevant information from interacted users of item v as:

$$h_{vu} = f(q_v, p_{u_1}, ..., p_{u_{L_v}}) = \sum_{j \in U_v^+} a_{v,j} p_j, \tag{4}$$

where $U_v^+ = \{u_1, u_2, ..., u_{L_v}\}$ is the user set interacted by item v.

4.3.3 Adaptively measuring the diversity relation. Apart from the relevance relation between user u and item v, we believe the user's preference to diversity would affect the translation process, e.g., users having broader interests are more likely to watch various types of movies. The same goes for the item. In view of this, we treat the user's/item's preference to diversity as an additional diversity relation h_u'/h_v' to connect u and v, and represent it as a Gaussian distribution for each user or item, i.e.,

$$h_u' \sim \mathcal{N}(\mu_u, \sigma_u), \quad h_v' \sim \mathcal{N}(\mu_v, \sigma_v), \tag{5}$$

where $\mu_u(\mu_v)$ and $\sigma_u(\sigma_v)$ are learned mean vector and standard deviation vector for user u (item v), respectively. To perform back-propagation from h', we adopt the reparameterization trick [22]. We sample $\epsilon \sim \mathcal{N}(0, I)$ and reparameterize $h' = \mu + \epsilon \odot \sigma$, where $\odot$ denotes element-wise product.

Note that, the parameters μ, σ are not item-specific or user-specific, because it is hard to learn the characteristics from interactions directly. Instead, we relate these parameters to coarse-grained aspects, which are more suitable for modeling the variance among clusters of items. Specifically, we first calculate the frequency vector from each user to $|C|$ categories from the interaction matrix. After applying PCA dimension reduction and min-max normalization to the user's frequency distribution on category, we obtain the user's attention distribution w_u over k aspects. We then construct the mean and standard deviation vector as:

$$\mu_u = T_\mu^\top w_u, \quad \sigma_u = T_\sigma^\top w_u, \tag{6}$$

where $T_\mu \in \mathbb{R}^{k \times D}$ and $T_\sigma \in \mathbb{R}^{k \times D}$ are the learnable weight matrix shared for all users. Similarly, μ_v and σ_v can be constructed in the similar way.

4.3.4 Performing two-way translation. So far, we have extracted two types of relations accounting for the relevance and diversity. We now incorporate these relations into the final connective relation as:

$$r_{uv} = h_{uv} + h_u', \quad r_{vu} = h_{vu} + h_v'. \tag{7}$$

Then we design the distance function for each user-item interaction. As we pointed out in previous section, classic one-way translation based metric learning has limited expression ability. Indeed, not only a user chooses preferable items according to her own interests, but also an item is oriented towards suitable consumers by its characteristic. Hence we perform a two-way translation in our model. Given the user-item pair (u, v), the distance function can be defined as:

$$\begin{aligned} d(u, v) &= \|p_u + r_{uv} - q_v\|_2^2, \\ d(v, u) &= \|q_v + r_{vu} - p_u\|_2^2, \end{aligned} \tag{8}$$

where $\|.\|_2$ is the L2 norm. The two-way translation improves the expression ability from two separate perspectives compared to the vanilla metric learning approaches.

4.4 Optimization and Learning

Under the BBN architecture, two types of training samples are fed into corresponding branches to calculate the loss. We define $\mathcal{L}_{B1}, \mathcal{L}_{B2}$ as the loss for conventional learning branch and adaptive learning branch, respectively. We extend the pairwise margin loss [17] for two-way adaptive metric learning in each branch, which ensures the distance between user u and positive item v is less than the distance between user u and negative item v^- by a fixed margin $m > 0$. The loss is defined as follows:

$$\mathcal{L}_{Bi} = \sum_{(u,v)\in\mathcal{I}} \sum_{(u,v^-)\in\mathcal{I}^-} \left[d(u,v) - d(u,v^-) + m\right]_+ \\ + \left[d(v,u) - d(v^-,u) + m\right]_+, \quad (9)$$

where $(u, v) \in \mathcal{I}$ indicates the positive sample, and $(u, v^-) \in \mathcal{I}^-$ denotes the negative sample and $[x]_+ = max(x, 0)$.

In addition, since the translations in two branches are conducted in different parameter spaces, there might be a small gap between them. We then define an additional consistency loss $\mathcal{L}_3$ to measure the distribution difference between two branches:

$$\mathcal{L}_3 = \mathcal{D}_{KL}\left(C_{fw}(x)||A_{fw}(x')\right) + \mathcal{D}_{KL}\left(C_{bw}(x)||A_{bw}(x')\right), \quad (10)$$

where $\mathcal{D}_{KL}$ is Kullback-Leibler divergence, $C(x)$ and $A(x')$ denote the probability distributions obtained by calculating the softmax function over the translations from conventional and adaptive learning branches, respectively. The subscript "fw" and "bw" denote user-to-item direction and item-to-user direction, respectively.

Finally, we define the overall loss function $\mathcal{L}$ as:

$$\mathcal{L} = \begin{cases} \alpha\mathcal{L}_{B1} + (1-\alpha)\mathcal{L}_{B2} + \mathcal{L}_3, & \text{if } |skewness| < \epsilon, \\ (1-\alpha)\mathcal{L}_{B1} + \alpha\mathcal{L}_{B2} + \mathcal{L}_3, & \text{if } |skewness| \geq \epsilon. \end{cases} \quad (11)$$

where α is the parameter to control the weights of two branches, which is calculated by $d_u * T/T_{max}$ to include the individual user's diversity preference, d_u is the diversity score of user u as defined in Section 4.2.2, and T and T_{max} are the current and total training epoch number. Also note that either α or $1 - \alpha$ can be associated with $\mathcal{L}_{B1}$ or $\mathcal{L}_{B2}$, depending on the domain has a skewed or normal diversity distribution. Practically, the real-world data might not strictly follow a normal distribution. As a rule of thumb [3], the distribution is approximately symmetric if skewness is between -1/2 and +1/2. Hence we set ϵ as 0.5 in our study.

5 EXPERIMENTS

In this section, we design experiments to verify the effectiveness and correctness of our model.

5.1 Experimental Setup

5.1.1 Datasets. We use three publicly available datasets from different domains. **Music** and **Beauty** are chosen from Amazon[1] which both contain metadata of diverse products. Following the evaluation settings in [5, 12], we take the 5-core version for experiments, where

Table 1: Statistics of the datasets.

Datasets	#Users	#Items	#Interactions	Sparsity	#Cate.
Music	5541	3568	64786	99.67%	60
Beauty	8159	5862	98566	99.79%	41
Movielens	6040	3706	1000209	95.53%	18

each user or item has at least five interactions. **MovieLens**[2] is a widely adopted dataset in the application domain of recommending movies to users. We employ the MovieLens-1M version.

In order to be consistent with the implicit feedback setting [6, 16], we transform the detailed rating into a value of 0 or 1, indicating whether a user has rated an item. The statistics of the datasets are shown in Table 1.

5.1.2 Evaluation Metrics. For all experiments, we evaluate the recommendation performance in terms of accuracy and diversity.

Accuracy. We evaluate the accuracy of ranking list using Recall [17] and Normalized Discounted Cumulative Gain (NDCG) [6, 16]. Recall considers whether the ground-truth is ranked amongst the top-K items (normally @5 or @10) while NDCG is a position-aware metric, which assigns higher score to the hits at higher positions.

Diversity. We evaluate the recommendation diversity by the intra-list distance (ILD) [28, 41] and category coverage (CC) [28, 37]. ILD measures the diversity of the recommended item set by the mean distance between all pairs of items, while CC measures the coverage of the user interests by counting the relevant categories in the set.

Trade-off. Following [11, 25], we employ F-score as a harmonic mean of conflicting accuracy and diversity for different methods, where *F-score* = $2 * accuracy * diversity/(accuracy + diversity)$.

5.1.3 Baselines. To demonstrate the effectiveness of our proposed EDUA model, we compare it with the following methods.

- **LFM** [23] is a classical matrix factorization method for collaborative filtering, which learns the latent factors by alternating least squares.
- **NCF** [16] presents a neural collaborative filtering method combining multi-layer perceptron with generalized matrix factorization to encode non-linearities.
- **CML** [17] is a metric learning method, which encodes the user and item into a joint metric space and measures the user-item pair by euclidean distance.
- **TransCF** [29] is a relational metric learning method which includes the neighborhood information of users and items to construct translation vectors.
- **ENMF** [6] proposes to learn neural recommendation model from the whole training data with a non-sampling strategy to enhance learning efficiency.
- **MMR** [4] is a canonical re-ranking method for diversified ranking problems by maximizing the marginal relevance.
- **DPP** [8] applies determinantal point process to optimize the trade-off between accuracy and diversity and generates recommendation list through the MAP inference.
- **PD-GAN** [37] proposes to learn users' personal preferences and item diversity through an adversarial learning process and adopts DPP model as generator to generate diverse results.

Table 2: The overall performance comparison on three datasets in terms of accuracy, diversity and trade-off evaluation, respectively. The best performance among all is in bold while the second best one is marked with an underline.

Datasets	Models	Acurracy				Diversity				Trade-off	
		Recall@5	Recall@10	NDCG@5	NDCG@10	ILD@5	ILD@10	CC@5	CC@10	F1@5	F1@10
Music	LFM	0.1264	0.1822	0.1273	0.1503	0.5904	0.6237	0.5762	0.6820	0.2082	0.2820
	NCF	0.1156	0.1668	0.1151	0.1364	0.6556	0.6809	0.5814	0.6897	0.1965	0.2680
	CML	0.1579	0.2204	0.1645	0.1879	0.5871	0.6366	0.5889	0.6892	0.2489	0.3274
	TransCF	0.1598	0.2242	0.1611	0.1871	0.5377	0.5801	0.5920	0.6809	0.2464	0.3234
	ENMF	0.1560	0.2179	0.1559	0.1796	0.5960	0.6418	0.5854	0.6917	0.2473	0.3253
	MMR	0.0690	0.1029	0.0708	0.0860	**0.7557**	**0.7822**	0.6085	**0.7273**	0.1265	0.1819
	DPP	0.0771	0.1528	0.0772	0.1102	0.6769	0.6800	0.5926	0.7136	0.1384	0.2495
	PD-GAN	0.1435	0.2068	0.1401	0.1658	0.6030	0.6376	0.5732	0.6806	0.2318	0.3123
	BGCF	0.1340	0.1934	0.1376	0.1621	0.6023	0.6246	0.5835	0.6870	0.2192	0.2953
	EDUA	**0.1685**	**0.2327**	**0.1718**	**0.1960**	0.6412	0.6893	**0.6135**	0.7168	**0.2669**	**0.3479**
Beauty	LFM	0.0505	0.0781	0.0626	0.0755	0.7452	0.7510	0.4820	0.6132	0.0946	0.1415
	NCF	0.0399	0.0654	0.0531	0.0657	0.7498	0.7719	0.4542	0.5864	0.0758	0.1206
	CML	0.0605	0.0977	0.0792	0.0949	0.7336	0.7513	0.4823	0.6133	0.1118	0.1729
	TransCF	0.0621	0.0970	0.0763	0.0927	0.6934	0.7130	0.4901	0.6125	0.1140	0.1708
	ENMF	0.0675	0.1037	0.0873	0.1042	0.7084	0.7257	0.4964	0.6155	0.1233	0.1815
	MMR	0.0424	0.0791	0.0557	0.0714	0.7450	0.7509	0.4704	0.5904	0.0802	0.1431
	DPP	0.0382	0.0800	0.0500	0.0696	**0.7785**	0.7854	0.4850	0.6350	0.0728	0.1452
	PD-GAN	0.0580	0.0899	0.0744	0.0881	0.7309	0.7487	0.4844	0.6056	0.1075	0.1605
	BGCF	0.0525	0.0867	0.0662	0.0824	0.7492	0.7524	0.4846	0.6181	0.0981	0.1555
	EDUA	**0.0721**	**0.1071**	**0.0926**	**0.1067**	0.7675	**0.7923**	**0.5009**	**0.6445**	**0.1318**	**0.1887**
Movielens	LFM	0.0835	0.1391	0.6106	0.6303	0.7928	0.8048	0.3670	0.5237	0.1511	0.2372
	NCF	0.0855	0.1431	0.6041	0.6260	0.7636	0.7798	0.3678	0.5178	0.1538	0.2418
	CML	0.0953	0.1578	0.6488	0.6634	0.7868	0.8012	0.3745	0.5309	0.1700	0.2637
	TransCF	0.0939	0.1562	0.6372	0.6555	0.7699	0.7869	0.3655	0.5180	0.1674	0.2607
	ENMF	0.0928	0.1546	0.6131	0.6355	0.7737	0.7863	0.3700	0.5231	0.1657	0.2584
	MMR	0.0435	0.0798	0.4125	0.4571	0.7950	0.8033	0.3521	0.5112	0.0825	0.1452
	DPP	0.0594	0.1071	0.4643	0.5096	0.8157	0.8131	0.3739	0.5126	0.1107	0.1893
	PD-GAN	0.0849	0.1443	0.5705	0.6006	0.7577	0.7750	0.3616	0.5092	0.1527	0.2433
	BGCF	0.0744	0.1257	0.5749	0.6004	0.8022	0.8117	0.3643	0.5243	0.1362	0.2177
	EDUA	**0.0975**	**0.1613**	**0.6509**	**0.6684**	**0.8672**	**0.8735**	**0.3901**	**0.5651**	**0.1753**	**0.2723**

- **BGCF** [34] models the uncertainty in the user-item bipartite graph by node copying and achieves accurate and diverse results using Bayesian graph convolutional neural networks.

We partition the baselines into two parts. The first part (from LFM to ENMF) includes the classical methods designed for improving the accuracy. The second part (from MMR to BGCF) includes typical diversity-promoting methods (post-processing, DPP, and end-to-end) developed for balancing accuracy and diversity.

5.1.4 Settings. For each user, we randomly select 80% of historical interactions as training set, and the remaining 20% data constitutes the testing set. During the test phrase, our evaluation protocol ranks all unobserved items in the training set for each user [7].

For the baselines, we follow the same hyper-parameter settings if they are reported by the authors and we fine-tune the hyper-parameters if they are not reported. We use WMF [18] to calculate the relevance score in first step for MMR and to construct kernel matrix for DPP. The trade-off parameter in MMR, DPP, and PD-GAN is set to 0.8, 0.9, and 0.9, respectively for getting the best trade-off performance. For BGCF, we set the number of sampled graphs to 5.

For our EDUA, we set the batch size = 128, the number of max epoch = 20, the initial learning rate = 0.0005, and we use Adam [21] as optimizer to self-adapt the learning rate. The margin m in margin loss = 1. The embedding dimension $D = 50$. The number of aspects $k = 20$. We sample $P = 20$ unobserved items as negative samples for each positive user-item pair [17].
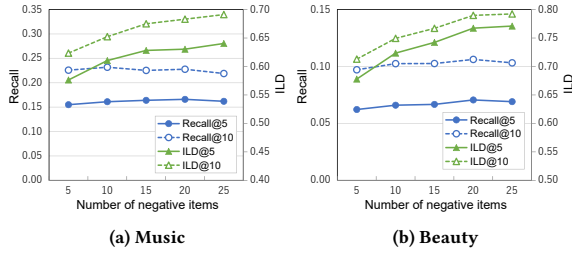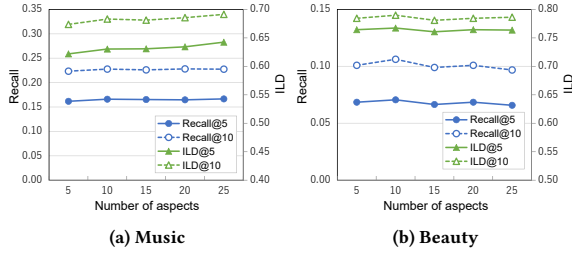
5.2 Performance Comparison

The performance comparison between our proposed EDUA model and the baselines on three datasets is reported in Table 2. From the results, we have the following important observations.

Firstly, it is clear that our EDUA significantly outperforms all the baselines in terms of accuracy and diversity trade-off (F1) on three datasets. The relative improvement of our model over the best baselines on three datasets are 5.58%, 4.87%, and 2.31% on F1@5, and 4.34%, 3.14%, and 2.35% on F1@10, respectively. In addition, EDUA achieves the best or second best recommendation accuracy (Recall, NDCG) and diversity (ILD, CC) in most of cases. It verifies the superiority of our proposed framework with the adapted BBN architecture and two-way adaptive metric learning backbone.

Secondly, compared with four accuracy-oriented baselines, our EDUA achieves the superior or competitive performance in terms of accuracy, and gains remarkable improvements on diversity. Among these baseline methods, LFM and NCF have inferior performance on three datasets. This indicates that the latent features obtained by matrix factorization are insufficient to capture user and item relationships and lead to limited performance. CML and ENMF perform better, owing to that CML overcomes the inherent limitation of the inner product and ENMF uses whole-data learning strategy. However, these methods focus on the recommendation accuracy and rely on the historical interactions to learn representation, resulting in redundant recommendation results.

Thirdly, compared with four diversity-promoting baselines, our EDUA gets significant improvements in accuracy while keeping a

Figure 4: Impacts of number of negative samples P.

Figure 5: Impacts of number of aspects k.

comparable or better diversity performance. In contrast, the post-processing methods MMR and DPP maintain the explicit trade-off between accuracy and diversity, and get impressive increase in diversity score. But, we also notice that they make great sacrifice of accuracy to generate a high diversity. The end-to-end methods PD-GAN and BGCF slightly enhance recommendation accuracy with a relatively low diversity. This reveals that our EDUA model provides a better paradigm by employing two branches to balance accuracy and diversity and injecting user-level diversity into internal structure.

5.3 Parameter Analysis

In this subsection, we analyze the impacts of hyper-parameters in EDUA, including the number of negative samples P and the number of aspects k. Figure 4 and Figure 5 show the results by tuning the negative sample number amongst {5, 10, 15, 20, 25}, and varying the number of aspects in the set of {5, 10, 15, 20, 25}, respectively. Due to the space limitation, we only present the Recall and ILD results on Music and Beauty datasets.

We first fix k to 20 and vary P. As shown in Figure 4, both recall and ILD gradually rise to different extent with the increase of the number of negative samples. When P is too small, e.g. 5, it is insufficient to achieve optimal performance especially on ILD. The reason is that the limited number of samples are unable to describe the border line between the positive and the negative. More negative instances are beneficial for pushing irrelevant and reduplicative items apart to improve diversity. While the number of negative samples reaches at 25, the accuracy performance degrades. It reveals that setting the sampling too large will hurt the performance.

We next fix P to 20 and vary k. We observe from Figure 5 that the number of aspects k does not have big impacts on the performance. Indeed, the curves are rather steady with the changing values of k. The reason might be that the aspect is a relatively coarse-grained form as category group or user cluster. In most cases, $k = 20$ can

generate good enough results. Hence we choose this as the setting for k on all datasets.

5.4 Ablation Studies

We conduct a series of ablation studies to assess the individual contributions of the components or strategies in our model, including the impacts of the order of learning focus, bilateral branches, and two-way adaptive metric learning. The ablation results on three datasets are shown in Table 3.

5.4.1 Impacts of the order of learning focus. As aforementioned in Section 4.2.1, the order of learning focus decides which branch is carefully trained at the later stage, and different datasets should have different orders to encode the domain-level diversity. We determine the order by the skewness of the diversity distribution, which is 0.2097, 0.1006, and -0.8569 for Music, Beauty, and Movielens, respectively. It indicates that Music and Beauty are approximately symmetric while Movielens is clearly left-skewed. Consequently, the training order for Music and Beauty is "adp → conv", and "conv → adp" for Movielens, where "adp" and "conv" denotes the adaptive and conventional learning branch, respectively.

In this experiment, we reverse the order to investigate its impact. Specifically, $\text{EDUA}_{\text{conv}\rightarrow\text{adp}}$ transfers the learning focus from conventional learning branch to adaptive one (i.e., $\alpha = d_u * T/T_{max}$) while $\text{EDUA}_{\text{adp}\rightarrow\text{conv}}$ is on the contrary (i.e., $\alpha = 1 - d_u * T/T_{max}$). It is clear that all the reversed order results in inferior performance. This is consistent with the domain biases, i.e., the users in Music and Beauty have focused interests, while a large number of users have broader interests in Movielens.

5.4.2 Impacts of two-way adaptive metric learning. Finally, we investigate the impacts of two-way adaptive metric learning. Specifically, $\text{EDUA}_{\text{w/o-attn}}$ and $\text{EDUA}_{\text{w/o-dist}}$ remove the attentive matching vector in relevance relation and the diversity relation, respectively from two types of connective relations, and $\text{EDUA}_{\text{w/o-twoway}}$ only reserves the commonly used user-to-item direction and removes the item-to-user direction.
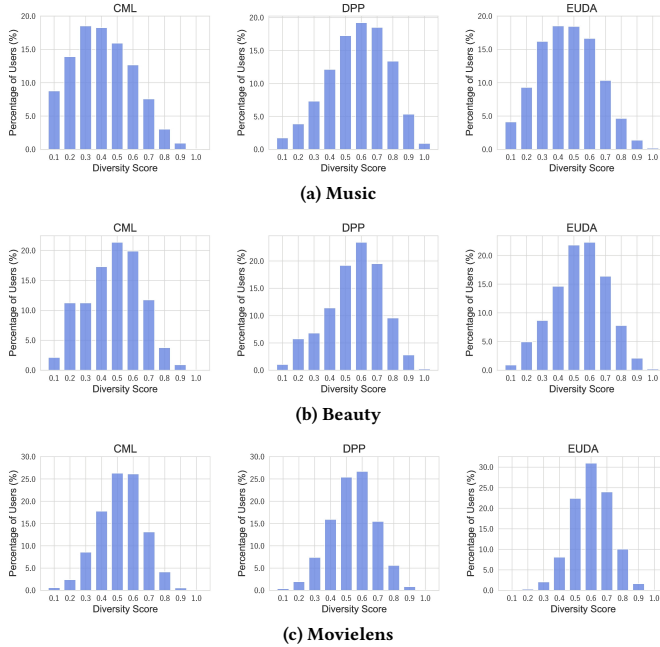
The accuracy performance of $\text{EDUA}_{\text{w/o-dist}}$ changes slightly but its diversity declines a lot compared to the standard EDUA. On the other hand, the downward trend of $\text{EDUA}_{\text{w/o-attn}}$ is more obvious on accuracy than diversity. This is consistent with our assumption that attention matching captures the exact relations between the historical behaviors and the target user while the diversity distribution describes the varying range of personal interests. Lastly, $\text{EDUA}_{\text{w/o-twoway}}$ drops sharply in terms of accuracy and diversity, proving that two-way translation is critical in modeling user-item relationship.

5.5 Analysis on Model's Adaptive Ability to Diversity

In this subsection, we analyze the adaptive ability of our proposed EDUA model on encoding diversity on different domains. Specifically, we define the predicted diversity score of each user as the number of recommended categories among top-K items divided by the total number of recommendation (K=10). In order to assess the adaptive ability, we plot the predicted user's diversity distribution conducted by our EDUA and that by other two typical baselines,

Table 3: Ablation results on three datasets

Datasets	Model	Acurracy				Diversity				Trade-off	
		Recall@5	Recall@10	NDCG@5	NDCG@10	ILD@5	ILD@10	CC@5	CC@10	F1@5	F1@10
Music	EDUA$_{conv \to adp}$	0.1636	0.2280	0.1671	0.1934	0.6305	0.6789	0.6014	0.7106	0.2598	0.3414
	EDUA$_{adp \to conv}$	**0.1685**	**0.2327**	**0.1718**	**0.1960**	**0.6412**	**0.6893**	**0.6135**	**0.7168**	**0.2669**	**0.3479**
	EDUA$_{w/o\text{-}dist}$	0.1633	0.2211	0.1707	0.1915	0.6022	0.6518	0.6046	0.7116	0.2569	0.3302
	EDUA$_{w/o\text{-}attn}$	0.1586	0.2160	0.1648	0.1901	0.6092	0.6622	0.5962	0.7010	0.2517	0.3257
	EDUA$_{w/o\text{-}twoway}$	0.1560	0.2147	0.1633	0.1865	0.6105	0.6655	0.5927	0.7082	0.2485	0.3247
Beauty	EDUA$_{conv \to adp}$	0.0659	0.1018	0.0866	0.1019	0.7572	0.7768	0.4876	0.6298	0.1212	0.1800
	EDUA$_{adp \to conv}$	**0.0721**	**0.1071**	**0.0926**	**0.1067**	**0.7675**	**0.7923**	**0.5009**	**0.6445**	**0.1318**	**0.1887**
	EDUA$_{w/o\text{-}dist}$	0.0678	0.1020	0.0865	0.1008	0.7450	0.7705	0.4908	0.6286	0.1243	0.1802
	EDUA$_{w/o\text{-}attn}$	0.0652	0.1002	0.0847	0.0995	0.7315	0.7583	0.4951	0.6267	0.1197	0.1770
	EDUA$_{w/o\text{-}twoway}$	0.0670	0.0976	0.0912	0.1037	0.7519	0.7786	0.4917	0.6318	0.1230	0.1735
Movielens	EDUA$_{conv \to adp}$	**0.0975**	**0.1613**	**0.6509**	**0.6684**	**0.8672**	**0.8735**	**0.3901**	**0.5651**	**0.1753**	**0.2723**
	EDUA$_{adp \to conv}$	0.0891	0.1468	0.6262	0.6425	0.8538	0.8606	0.3815	0.5591	0.1614	0.2508
	EDUA$_{w/o\text{-}dist}$	0.0951	0.1562	0.6272	0.6472	0.8422	0.8570	0.3779	0.5650	0.1709	0.2642
	EDUA$_{w/o\text{-}attn}$	0.0937	0.1531	0.6308	0.6468	0.8517	0.8620	0.3810	0.5551	0.1688	0.2600
	EDUA$_{w/o\text{-}twoway}$	0.0883	0.1447	0.6122	0.6324	0.8577	0.8624	0.3836	0.5618	0.1601	0.2478

Figure 6: The distribution of predicted user's diversity on three datasets.

namely CML (accuracy-oriented) and DPP (diversity-promoting) on three datasets.

As shown in Figure 6, on different domains, the distribution is presented as right-skewed for CML, and is left-skewed for DPP. The reason is that CML is based on CF and can only cover a small percentage of similar items, which results in its overall low diversity. On the other hand, DPP achieves high diversity using the global trading-off strategy but it ignores the individual difference. In contrast, our EDUA shows different skewed trends on different domains and yields diversity score distribution close to the original one in Figure 2. This proves that our model can adapt to different domain biases and generate user-specific recommendation lists.

Table 4: MSE performance comparison between the predicted user diversity and the original one.

Models / Datasets	CML	DPP	EDUA
Music	0.0024	0.0025	**0.0014**
Beauty	0.0012	0.0013	**0.0008**
Movielens	0.0111	0.0094	**0.0059**

We then further examine the difference between the predicted user diversity and the original one. We take the original diversity score defined in Section 4.2.2 as ground-truth and calculate the mean square error (MSE) for different methods. From the results in Table 4, it is clear that our EDUA achieves the best MSE. This suggests that our recommended results are mostly consistent with the user's individual diversity in original data and thus can improve the user experience.

6 CONCLUSION

In this paper, we present a novel EDUA model for diversified recommendation. It adopts the bilateral branch network as the main architecture for two-objective optimization, i.e., maintaining accuracy in conventional learning branch and increasing diversity in adaptive learning branch. We propose to adaptively balance the learning procedure between two branches so as to encode the domain-level diversity. We further present a two-way metric learning as the backbone for each branch, which captures the item's orientation towards target users and treats the user-level diversity as a special relation connecting the user and the item. Extensive experiments on three real-world datasets demonstrate the effectiveness of our proposed model as well as its distinct components or learning strategies.

REFERENCES

[1] Gediminas Adomavicius and YoungOk Kwon. 2012. Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques. *IEEE Trans. Knowl. Data Eng.* 24, 5 (2012), 896–911.

[2] Azin Ashkan, Branislav Kveton, Shlomo Berkovsky, and Zheng Wen. 2015. Optimal Greedy Diversity for Recommendation. In *IJCAI*. 1742–1748.

[3] Michael George Bulmer. 1979. *Principles of statistics.* Courier Corporation.

[4] Jaime G. Carbonell and Jade Goldstein. 1998. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. In *SIGIR*. 335–336.

[5] Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. 2018. Neural Attentional Rating Regression with Review-level Explanations. In *WWW*. 1583–1592.

[6] Chong Chen, Min Zhang, Yongfeng Zhang, Yiqun Liu, and Shaoping Ma. 2020. Efficient Neural Matrix Factorization without Sampling for Recommendation. *ACM Trans. Inf. Syst.* 38, 2 (2020), 14:1–14:28.

[7] Chong Chen, Min Zhang, Yongfeng Zhang, Weizhi Ma, Yiqun Liu, and Shaoping Ma. 2020. Efficient Heterogeneous Collaborative Filtering without Negative Sampling for Recommendation. In *AAAI*. 19–26.

[8] Laming Chen, Guoxin Zhang, and Eric Zhou. 2018. Fast Greedy MAP Inference for Determinantal Point Process to Improve Recommendation Diversity. In *NeuIPS*. 5627–5638.

[9] Shuo Chen, Joshua L Moore, Douglas Ross Turnbull, and Thorsten Joachims. 2012. Playlist prediction via metric embedding.. In *KDD*. 714ÍC–722.

[10] Tong Chen, Hongzhi Yin, Hongxu Chen, Rui Yan, Quoc Viet Hung Nguyen, and Xue Li. 2019. AIR: Attentional Intention-Aware Recommender Systems. In *ICDE*. 304–315.

[11] Peizhe Cheng, Shuaiqiang Wang, Jun Ma, Jiankai Sun, and Hui Xiong. 2017. Learning to Recommend Accurate and Diverse Items. In *WWW*. 183–192.

[12] Zhiyong Cheng, Ying Ding, Xiangnan He, Lei Zhu, Xuemeng Song, and Mohan S. Kankanhalli. 2018. Aˆ3NCF: An Adaptive Aspect Attention Model for Rating Prediction. In *IJCAI*. 3748–3754.

[13] Shanshan Feng, Xutao Li, Yifeng Zeng, Gao Cong, Yeow Meng Chee, and Quan Yuan. 2015. Personalized ranking metric embedding for next new POI recommendation. In *IJCAI*. 2069–2075.

[14] Lu Gan, Diana Nurbakova, Léa Laporte, and Sylvie Calabretto. 2020. Enhancing Recommendation Diversity using Determinantal Point Processes on Knowledge Graphs. In *SIGIR*. 2001–2004.

[15] Ruining He, Wang-Cheng Kang, and Julian J. McAuley. 2017. Translation-based Recommendation. In *RecSys*. 161–169.

[16] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *WWW*. 173–182.

[17] Cheng-Kang Hsieh, Longqi Yang, Yin Cui, Tsung-Yi Lin, Serge J. Belongie, and Deborah Estrin. 2017. Collaborative Metric Learning. In *WWW*. 193–201.

[18] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *ICDM*. 263–272.

[19] Mohammad Khoshneshin and W. Nick Street. 2010. Collaborative Filtering via Euclidean Embedding. In *RecSys*. 87–94.

[20] Yejin Kim, Kwangseob Kim, Chanyoung Park, and Hwanjo Yu. 2019. Sequential and Diverse Recommendation with Long Tail. In *IJCAI*. 2740–2746.

[21] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.

[22] Diederik P. Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. In *ICLR*.

[23] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *Computer* 42, 8 (2009), 30–37.

[24] Mingming Li, Shuai Zhang, Fuqing Zhu, Wanhui Qian, Liangjun Zang, Jizhong Han, and Songlin Hu. 2020. Symmetric Metric Learning with Adaptive Margin for Recommendation. In *AAAI*. 4634–4641.

[25] Yong Liu, Yingtai Xiao, Qiong Wu, Chunyan Miao, Juyong Zhang, Binqiang Zhao, and Haihong Tang. 2020. Diversified Interactive Recommendation with Implicit Feedback. In *AAAI*. 4932–4939.

[26] Chen Ma, Liheng Ma, Yingxue Zhang, Ruiming Tang, Xue Liu, and Mark Coates. 2020. Probabilistic Metric Learning with Adaptive Margin for Top-K Recommendation. In *KDD*. 1036–1044.

[27] Gal Oestreicher-Singer and Arun Sundararajan. 2012. Recommendation Networks and the Long Tail of Electronic Commerce. *MIS Q.* 36, 1 (2012), 65–83.

[28] Shameem Puthiya Parambath, Nicolas Usunier, and Yves Grandvalet. 2016. A Coverage-Based Approach to Recommendation Diversity On Similarity Graph. In *Recsys*. 15–22.

[29] Chanyoung Park, Donghyun Kim, Xing Xie, and Hwanjo Yu. 2018. Collaborative Translational Metric Learning. In *ICDM*. 367–376.

[30] Yoon-Joo Park. 2013. The Adaptive Clustering Method for the Long Tail Problem of Recommender Systems. *IEEE Trans. Knowl. Data Eng.* 25, 8 (2013), 1904–1915.

[31] Rajiv Pasricha and Julian J. McAuley. 2018. Translation-based factorization machines for sequential recommendation. In *RecSys*. 63–71.

[32] Tieyun Qian, Yile Liang, Qing Li, and Hui Xiong. 2020. Attribute Graph Neural Networks for Strict Cold Start Recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2020), 1–1. https://doi.org/10.1109/TKDE.2020.3038234

[33] Lijing Qin and Xiaoyan Zhu. 2013. Promoting Diversity in Recommendation by Entropy Regularizer. In *IJCAI*. 2698–2704.

[34] Jianing Sun, Wei Guo, Dengcheng Zhang, Yingxue Zhang, Florence Regol, Yaochen Hu, Huifeng Guo, Ruiming Tang, Han Yuan, Xiuqiang He, and Mark Coates. 2020. A Framework for Recommending Accurate and Diverse Items Using Bayesian Graph Convolutional Neural Networks. In *KDD*. 2030–2039.

[35] Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. Latent Relational Metric Learning via Memory-based Attention for Collaborative Ranking. In *WWW*. 729–739.

[36] Yuandong Wang, Hongzhi Yin, Hongxu Chen, Tianyu Wo, Jie Xu, and Kai Zheng. 2019. Origin-Destination Matrix Prediction via Graph Convolution: a New Perspective of Passenger Demand Modeling. In *KDD*. 1227–1235.

[37] Qiong Wu, Yong Liu, Chunyan Miao, Binqiang Zhao, Yin Zhao, and Lu Guan. 2019. PD-GAN: Adversarial Learning for Personalized Diversity-Promoting Recommendation. In *IJCAI*. 3870–3876.

[38] Qiong Wu, Yong Liu, Chunyan Miao, Yin Zhao, Lu Guan, and Haihong Tang. 2019. Recent Advances in Diversified Recommendation. *CoRR* abs/1905.06589 (2019).

[39] Xun Yang, Yunshan Ma, Lizi Liao, Meng Wang, and Tat-Seng Chua. 2019. TransNFCM: Translation-Based Neural Fashion Compatibility Modeling. In *AAAI*. 403–410.

[40] Hongzhi Yin, Bin Cui, Jing Li, Junjie Yao, and Chen Chen. 2012. Challenging the Long Tail Recommendation. *Proc. VLDB Endow.* 5, 9 (2012), 896–907.

[41] Mi Zhang and Neil Hurley. 2008. Avoiding monotony: improving the diversity of recommendation lists. In *Recsys*. 123–130.

[42] Yin Zhang, Yun He, Jianling Wang, and James Caverlee. 2020. Adaptive Hierarchical Translation-based Sequential Recommendation. In *WWW*. 2984–2990.

[43] Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen. 2020. BBN: Bilateral-Branch Network With Cumulative Learning for Long-Tailed Visual Recognition. In *CVPR*. 9716–9725.

[44] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *WWW*. 22–32.