Attention Over Self-Attention: Intention-Aware Re-Ranking With Dynamic Transformer Encoders for Recommendation

Zhuoyi Lin\textsuperscript{a}, Sheng Zang, Rundong Wang\textsuperscript{a}, Zhu Sun\textsuperscript{a}, J. Senthilnath\textsuperscript{a}, Senior Member, IEEE, Chi Xu, and Chee Keong Kwoh\textsuperscript{a}, Senior Member, IEEE

Abstract—Re-ranking models refine item recommendation lists generated by the prior global ranking model, which have demonstrated their effectiveness in improving the recommendation quality. However, most existing re-ranking solutions only learn from implicit feedback with a shared prediction model, which regrettably ignore inter-item relationships under diverse user intentions. In this paper, we propose a novel Intention-aware Re-ranking Model with Dynamic Transformer Encoder (RAISE), aiming to perform user-specific prediction for each individual user based on her intentions. Specifically, we first propose to mine latent user intentions from text reviews with an intention discovering module (IDM). By differentiating the importance of review information with a co-attention network, the latent user intention can be explicitly modeled for each user-item pair. We then introduce a dynamic transformer encoder (DTE) to capture user-specific inter-item relationships among item candidates by seamlessly accommodating the learned latent user intentions via IDM. As such, one can not only achieve more personalized recommendations but also obtain corresponding explanations by constructing RAISE upon existing recommendation engines. Empirical study on four public datasets shows the superiority of our proposed RAISE, with up to 13.95\%, 9.60\%, and 13.03\% relative improvements evaluated by Precision@5, MAP@5, and NDCG@5 respectively.

Index Terms—Item re-ranking, user-specific prediction, user intention modeling, dynamic transformer

1 INTRODUCTION

In the era of Big Data, recommender systems are widely adopted by the online platforms (e.g., Amazon and Youtube), so as to alleviate the problem of information overload [1], [2]. Accordingly, latent factor models, e.g., matrix factorization [3], [4], and deep learning models, e.g., NeuMF [5], have demonstrated their effectiveness to achieve personalized recommendations by learning user and item representations. Despite the great success, one fundamental assumption of the above solutions is that a global ranking model is designed to optimize the overall performance of item recommendations. This could be sub-optimal for individual users because it ignores the local item distributions for each user [6], [7].

To remedy the above issue, a shared re-ranking model is employed to refine the initial recommendation lists provided by the prior global ranking model for each user. Specifically, by taking into account the relationships among item candidates in a user’s initial recommendation list, the inter-item relationships for the target user can be explicitly modeled [6], [7], [8]. As such, re-ranking models are able to re-rank the item candidates and generate more personalized recommendation list for each target user. However, we argue that the existing manner of modeling inter-item patterns is insufficient, as such item relationships could vary a lot under different user intentions.\textsuperscript{1} Intuitively, users have multiple intentions to adopt certain items; meanwhile, different user intentions could drive different user behaviors.

\textsuperscript{1}Following Wang et al. [9], user intention is a high-level concept, which could be the reasons of a user to purchase a certain item (e.g., timing, price, user interest and shopping for others). Note that we focus on refining the item recommendation lists with learnt latent user intentions from text reviews rather than predicting the next item with instant intention for the target user (e.g., sequential recommendation models such as BERT4Rec [10] and MF\textsuperscript{2}OU [11]).
We conduct extensive experiments and ablation studies on four public datasets to verify the effectiveness and interpretability of our proposed RAISE.

and item relationships [9], [12]. Taking Fig. 1 as an example, the representations of two cheap items could be similar for a price-sensitive user, while the representations of the two aforementioned items could be different for a user who often buys gifts for her friends, as she may be more concerned with whether items are well packaged. Besides, although the existing re-ranking models can capture the inter-item relationships among item candidates, they are shared by all users in a dataset, which may not be sufficiently ‘personalized’.

Consequently, we are seeking to investigate the re-ranking task through modeling user-specific inter-item relationships based on user intentions. However, this is not trivial because of two main challenges: (1) Basically, user intentions are diverse and complex, which may vary greatly when confronted with different items. How to accurately capture user intentions is of crucial importance to deliver a performance-enhanced re-ranking model; (2) To provide sufficient personalization, a shared re-ranking model is not feasible. However, it is impractical to assign each user a prediction model. Hence, a tailored solution need to be designed in order to achieve both effectiveness and efficiency.

To tackle these challenges, we propose a novel Intention-aware Re-ranking Model with Dynamic Transformer Encoder (RAISE). In particular, we first devise an intention discovering module (IDM) to mine latent user intentions from text reviews. As illustrated in Fig. 1, such auxiliary information contains users’ preferences and item properties, which could be useful for modeling user intentions and item relationships [13], [14]. Given a user-item pair, IDM applies a co-attention network to estimate review-to-review matching scores and differentiate users’ diverse intentions from text reviews. This enables intention-aware representations to be generated by weighting the text representations with learnt attention signals. We then design a dynamic transformer encoder (DTE) to perform user-specific predictions by seamlessly accommodating the learnt latent user intentions. Under the hood, the dynamic self-attention mechanism captures the user-specific inter-item relationships and provides the driving force: an individual attention network is applied over the self-attention layer, to contextualize the item representations based on the learnt user intentions. Our proposed DTE advances the classic transformer encoder by learning specialized transformations of input item sequences, which increases the representational capability with limited extra computational cost and keeps efficient inference. To summarize, this paper makes the following contributions.

- We emphasize the importance of modeling diverse user intentions for the re-ranking task, whereby an IDM is devised to help extract user intentions from text reviews.
- We design a DTE to explicitly capture the user-specific inter-item patterns based on the learnt user intentions via IDM. By applying an individual attention network over the self-attention layer, DTE enables our proposed RAISE to perform user-specific predictions in an efficient manner.
- We conduct extensive experiments and ablation studies on four public datasets to verify the effectiveness and interpretability of our proposed RAISE.

2 RELATED WORKS

This section first briefly reviews existing re-ranking studies. Since our proposed RAISE aims to perform user-specific prediction based on the input item sequences and user intentions, we then present existing review-aware recommendation methods and input-dependent recommendation models.

Re-Ranking Models for Recommendation. Re-ranking methods have demonstrated their effectiveness to improve user satisfaction and been widely applied in industrial applications. For example, diversity-aware re-ranking methods focus on presenting relevant but diverse results at the top of a ranked list [15], [16], while fairness-aware re-ranking methods aim to provide each item a fair proportion of exposure [17], [18], [19]. In this work, we aim to capture the mutual relationships between items in the initial recommendation list for each user in order to improve model accuracy. For example, DLCM [6] uses gated recurrent unit (GRU) to sequentially encode the information of top candidates into document representations, and SetRank [20] uses multi-head self-attention mechanisms and an attention rank loss function to re-rank documents for the document retrieval task. In addition, Seq2Slate [21] exploits mutual influences between items with pointer networks, which treats the recommendation task as a sequence generation problem. Meanwhile, a graph neural network (GNN) based re-ranking method IRGPR [8] has been proposed to fuse the information from multi-hop neighbors and item relationships. It essentially models the global user-item and item-item relationships rather than the local item-item relationships among the initial recommendation lists. A more recent work [22] propose to combine graph attention networks (GAT) and transformer model to capture the sequential signals underlying users’ behavior sequences for complementary product recommendation. Then it uses a hinge loss to perform re-ranking for sequential recommendation tasks without a specific model architecture. The most related work to ours is PRM [7]. By employing transformer encoders to encode the input items, the mutual influences between item candidates can be captured to refine the initial recommendations. However, the transformer encoders are
shared by all users in PRM, which ignores user intentions and could lead to sub-optimal performance.

To sum up, existing works follow the same paradigm of modeling the inter-item relationships to re-rank the initial recommendation and regrettably ignore diverse user intentions. By contrast, we focus on exploiting latent user intentions from text reviews to capture the inter-item relationships specifically for each user in an effective manner. This enables RAISE to provide user-specific prediction and achieve more personalized recommendation.

Review-Aware Recommendation Methods. Our work benefits from review-aware recommendation models which aim to exploit richer semantic information from the text reviews. Early studies such as HFT and TopicMF [13], [14] demonstrate that better rating prediction accuracy can be achieved by modeling text reviews with topic models. Empowered by the powerful representational capabilities of deep neural networks, deep learning-based recommendation models such as DeepCoNN, TransNets, and MPCNN [23], [24], [25] are proposed to predict missing ratings from text reviews with convolutional neural networks (CNN) [26] and attention networks [27]. As for ranking tasks, TAFA [28] provides recommendations by jointly learning representations from user reviews and implicit feedback. Recently, TIM [29] models users and items within the topic space which is learned from the review data. Another direction of this field is to provide explanations for item recommendations based on text reviews [30], [31], [32]. Different from the above works which mainly focus on to tackle rating prediction tasks or ranking tasks with text reviews, we focus on distinguishing user intentions from reviews so as to construct user-specific re-ranking model.

Input-Dependent Models. Recently, input-dependent models have shown effectiveness in various domains, such as language modeling [33], [34] and computer vision [35], [36]. In recommendation, IFM [37] and DIFM [38] are presented to re-weight the representations of features and weights for different input instances before performing feature interactions. Inspired by these studies, we design a dynamic transformer encoder which performs an individual attention network on the self-attention layer and enables modeling user-specific inter-item patterns unveiled by the user intentions. We demonstrate that the proposed dynamic transformer encoder is computationally efficient with superior performance in this study.

3 PRELIMINARIES

Notations and Problem Formulation. Given a set of users $U = \{u_1, u_2, \ldots\}$, a set of items $I = \{i_1, i_2, \ldots\}$, and observed interaction scores $y_{ui}$, we adopt GMF [5] as the prior global ranking model, which predicts the missing interaction score $\hat{y}_{ui}$ between the target user $u$ and target item $i$ with their learnt representations denoted as $p_u$ and $q_i$, respectively. In this way, an initial recommendation list $S_u = \{i_1, i_2, \ldots, i_n\}$ can be generated by ranking the estimated interaction scores for user $u$, where $n$ is the length of $S_u$. Given $S_u$, $p_u$ and $\{q_i | i \in S_u\}$ learnt from GMF, the goal of our study is essentially how to effectively capture the inter-item relationships under different user intentions from text reviews and efficiently perform user-specific refinement for $S_u$. Hence, we further introduce two review sequences $R_u = \{r_u^{(1)}, r_u^{(2)}, \ldots, r_u^{(n)}\}$ and $R_i = \{r_i^{(1)}, r_i^{(2)}, \ldots, r_i^{(n)}\}$ for user $u$ and item $i$, which contains the reviews written by user $u$ and reviews received by item $i$, respectively. Note that $l_u$ and $l_i$ are the maximum number of reviews of $u$ and $i$. As such, each user-item pair in the training set can be denoted as a 6-tuple $(u, i, p_u, q_i, R_u, R_i)$, and RAISE is trained to re-rank the item candidates in $S_u$ for user $u$.

Self-Attention Mechanism. Our work benefits from the self-attention mechanism due to its effectiveness to capture the inter-item patterns among item candidates [27]. As shown in Fig. 2a. After linearly projecting the input sequence $S$ to obtain queries ($Q$) and keys ($K$) of dimension $d_k$, and values ($V$) of dimension $d_v$, the attention function produces outputs of dimension $d_v$ which is defined as follows:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where $Q = SW^Q$, $K = SW^K$, and $V = SW^V$, (1)

where $W^Q$, $W^K$, and $W^V$ are transform matrices.

In addition, the self-attention mechanism can be multi-headed by employing a number of heads $h$ [27], [34]. As shown in Fig. 2b, each head learns separate attention weights from different representation subspaces

$$\text{Multi-head}(Q, K, V) = \text{Concat}(head_1, \ldots, head_h)W^Q$$

where $head_i = \text{Attention}(QW^Q_i, KW^K_i, VW^V_i)$, (2)

where $W^Q_i \in \mathbb{R}^{h_d \times d}, W^K_i \in \mathbb{R}^{d \times d_h}, W^V_i \in \mathbb{R}^{d \times d_v}$ are transform matrices used in the multi-head self-attention.
Normally, $d_b$ and $d_c$ are set to $d/h$. The extra computational cost of multi-head self-attention is limited compared with that of normal self-attention function with full dimensionality (see Eq. (1)). We will provide a more detailed discussion in Dynamic Self-attention Analysis.

4 THE PROPOSED METHOD

In this section, we first briefly introduce the architecture of our proposed re-ranking method RAISE, then we present each component of RAISE in detail.

4.1 Overall of RAISE

The overall architecture of RAISE is presented in Fig. 3a, which consists of intention discovering module (IDM), dynamic transformer encoders (DTE), and prediction layer. Given an initial list $S_u$ and learnt representations $p$ and $q$ from GMF, the goal of RAISE is to refine $S_u$ by modeling the local inter-item relationships based on user intentions. To this end, IDM first discovers user intentions from text review with a co-attention module. The learnt intention-aware representations from text review are then fed into the DTE to capture user-specific item relationships. Accordingly, a new predicted score for each item candidate in $S_u$ is generated via the prediction layer. Finally, a re-ranked item list can be achieved by ranking the predicted scores. We list main mathematical notations used in RAISE in Table 1.

4.2 Intention Discovering Module (IDM)

4.2.1 Intention-Aware Review Representation

Learning informative user intentions is of crucial importance for RAISE to perform more personalized re-ranking. Intuitively, text reviews written by users are usually semantic and encode users’ diverse intentions, which could lead to better modeling of user behavior when confronted with an item [23], [25].

We hence employ a co-attention module to generate two intention-aware representations for each user-item pair with regard to the diverse intentions behind their reviews, as illustrated in Fig. 3c. Specifically, given two input review sequences $R_u$ and $R_i$, we first obtain the representation of each single text review by simply adding its constituent word representations. As such, two embedding sequences $R_u^v = \{r_1^v, r_2^v, \ldots, r_n^v\}$ and $R_i^v = \{r_1^i, r_2^i, \ldots, r_n^i\}$ can be achieved. To distinguish the importance of reviews according to the intentions behind them, we then compute the review-level matching scores between every user-item review pair with a co-attention network, given by,

$$c_{kj} = f(r_k^v)^T M f(r_j^i),$$

(3)

where the matching matrix $C = [c_{kj}]$ indicates how well the intentions behind a user review and an item review matches each other; $M \in \mathbb{R}^{d \times d}$ is a trainable transform matrix; $r_k^v, r_j^i \in \mathbb{R}^d$ denote the representations of the $k$th review of user $u$ and the $j$th review of item $i$, respectively. Meanwhile, we employ two MLPs to encode the review representation...
denoted as $f(\cdot)$ in Eq. (3). Since each row (column) of $C$ indicates how much the main intention of a user (item) review matches those of all item (user) reviews, the refined representations of a text review regarding users and items can be weighted by the strength of its main intention,

$$r_{k}^{(u)} = \frac{1}{|R_u|} \sum_{k} c_{kj} \ast r_{k}^{u};$$

$$r_{j}^{(i)} = \frac{1}{|R_i|} \sum_{j} c_{kj} \ast r_{j}^{i}.$$  

(4)

Then we add up all the reviews of user $u$ and item $i$ respectively in order to obtain the intention-aware representations $r_{k}^{(u)}$ and $r_{j}^{(i)}$. Despite the simplicity, it is an effective way to encode the quantity of each intention behind a review sequence into the final intention-aware representations,

$$r_{k}^{(u)} = \sum_{k} r_{k}^{(u)}; \quad r_{j}^{(i)} = \sum_{j} r_{j}^{(i)}.$$  

(5)

Note that for each user-item pair, their intention-aware representations are unique and will be different w.r.t. other users or items. In other words, the intention-aware representations are contextually learnt according to the input review sequences. This enables RAISE to capture inter-item patterns based on the matched user intentions between the review sequences of the target user $u$ and those of item candidates in $S_u$ leading to more personalized recommendations.

4.2.2 Intention-Aware Sequential Representation

Before introducing the DTE, a meaningful input sequential representation $S$ should be generated in advance, to encode sequential item candidates in the initial list $S_u$. We notice that most of the existing re-ranking models capture inter-item relationships from the implicit feedback data [7], [21], [39]. However, implicit feedback only indicate if a user has interacted with an item, thus it could be hard to learn effective item representations from implicit feedback alone. Besides, the learnt item representations from implicit feedback are shared by all users, which is unreasonable as item representations may vary greatly under different user intentions. Hence, we propose to represent item candidates with both the implicit feedback data and text review data. To this end, the user-specific item representations $s_i$ can be achieved with latent intention representations learnt by GMF and intention-aware representations obtained in the IDM

$$s_i = W^S[\text{Concat}(s_{i}^{(im)}, s_{i}^{(re)})] + o,$$

$$s_{i}^{(im)} = f(\text{Concat}(p_{u}, q_{i}))$$

$$s_{i}^{(re)} = f(\text{Concat}(r_{i}^{(u)}, r_{i}^{(i)})),$$  

(6)

where $W^S$ is a projection matrix; $s_{i}^{(im)}$ and $s_{i}^{(re)}$ denote item representations learnt from the implicit feedback data via GMF and text review information via IDM, respectively; inspired by PRM [7], we also encode the initial positions of item candidates in $S_u$ denoted as $o$. At last, the representation of $S_u$ can be obtained by concatenating the representation ($s_i$) of each item in $S_u$

$$S = \text{Concat}(s_1, s_2, \ldots, s_n)^T,$$  

(7)

which is then fed into DTE for further refinement.

4.3 Dynamic Transformer Encoder (DTE)

4.3.1 User-Specific Transformer Encoders

A fundamental assumption of the transformer architecture is that the transform matrices used in the self-attention layer are shared by all input samples. However, user intentions could be diverse and complex in recommendation scenarios, thus the prediction model should be different for each user to achieve maximal personalization. Having realized the vital role of user-specific recommendation engines, we hence propose the DTE as described in Fig. 3b. By performing attention network over the self-attention layer in the transformer encoder, the capacity of the transformer encoder is improved without increasing its depth (blocks of transformer encoders) or width (number of heads). An illustration of how we obtain user-specific queries $(Q_u)$, user-specific keys $(K_u)$, and user-specific values $(V_u)$ is presented in Fig. 2c. Specifically, given the input embedding $S$ obtained from the IDM, the output of dynamic self-attention $S$ can be described as follows:

$$S = \text{Attention}(Q_u, K_u, V_u) = \text{Softmax}(Q_u K_u^T) V_u$$

(8)

where $Q_u = SW^Q_u$, $K_u = SW^K_u$, and $V_u = SW^V_u$. $S \in \mathbb{R}^{d \times d}$ and $\hat{S} \in \mathbb{R}^{d \times d}$ are the input and output of DTE respectively; $W^Q_u$, $W^K_u$, and $W^V_u$ are learned user-specific transform matrices. By comparing Eqs. (8) with (1), it showcases an important merit of our dynamic self-attention module. By assembling the transform matrices before scaled dot-product, it achieves stronger representation power while keeping computational efficiency. We will analyze the efficiency of DTE in Model Complexity and Scalability. In what follows, we present how to learn the three user-specific transform matrices $W^Q_u$, $W^K_u$, and $W^V_u$.

4.3.2 Attention Over Self-Attention

Inspired by recent efforts [34], [40] which improve multi-head transformer architecture with mixture of experts and attention mechanisms, in this paper, we propose to learn three specialized transform matrices for each user by attentively aggregating multiple transform matrices before the scaled dot-product process

$$W^Q_u = a_1 W^Q_1 + a_2 W^Q_2 + \cdots + a_t W^Q_t$$

$$W^K_u = a_1 W^K_1 + a_2 W^K_2 + \cdots + a_t W^K_t$$

$$W^V_u = a_1 W^V_1 + a_2 W^V_2 + \cdots + a_t W^V_t,$$  

(9)

where $a_i$ is the shared attention weight across the three types of transform matrices, and $t$ is the number of transform matrices for each $Q_u$, $K_u$, and $V_u$.

In order to learn a meaningful attention weight $a = [a_1, a_2, \ldots, a_t]$ for each user, we employ an attention network over the self-attention layer in the DTE

$$a = \text{Softmax}(W^E e + b^E)$$

$$e = \text{ReLU}(W^E \bar{p}_{S_u} \odot \bar{q}_{S_u} + b^E),$$  

(10)
where \( W^A, W^E \) are learnable projection matrices; and \( b^A, b^E \) are bias vectors. In order to encode all information in the initial list \( S_u \) and learn effective attention weights for each user, we obtain \( p_{S_u} \) and \( q_{S_u} \) by accumulating all latent representations \( p_{i} \) and \( q_{j} \) learnt via GMF and intention-aware representations \( r_i^{(u)} \) and \( r_j^{(v)} \) learnt via IDM,

\[
\begin{align*}
 p_{S_u} &= \frac{1}{n} \sum_{i \in S_u} (p_i + r_i^{(u)}); \\
 q_{S_u} &= \frac{1}{n} \sum_{i \in S_u} (q_i + r_i^{(v)}),
\end{align*}
\]

(11)

where \( n \) is the length of \( S_u \). By estimating \( a \) with information from both implicit feedback data and text review data, the specialized transform matrices can be obtained by Eq. (9), thus leading to user-specific predictions.

### 4.4 Prediction Layer

In this layer, a new ranking score \( \hat{y}_{ui} \) is estimated for each item candidate in the initial list \( S_u \). This enables us to generate a final re-ranked list for the target user \( u \) by sorting the item candidates according to their new scores from highest:

\[
\hat{y}_{ui} = \text{Softmax}(F^{(b)}W^p + b^p), i \in S_u,
\]

(12)

where \( F^{(b)} \) is the output of \( b \) blocks of the DTE. Meanwhile, \( W^p \) and \( b^p \) are transform matrix and bias vector for the prediction layer. We employ a negative log likelihood loss to train our RAISE, as suggested by [7],

\[
\mathcal{L} = -\sum_{u \in \mathcal{U}_t} \sum_{i \in S_u} y_{ui} \log \hat{y}_{ui},
\]

(13)

where \( \mathcal{U}_t \) refers to the set of users in the training set.

### 4.5 Dynamic Self-Attention Analysis

We investigate the properties of our dynamic self-attention in this section, which is the key of RAISE. Generally, the self-attention mechanism is usually applied in order to compute the mutual relationships among item candidates in \( S_u \), whose computational complexity is \( O(n^2d) \). Recent studies [7, 20] propose to capture the mutual relationships from different sub-spaces, multi-head self-attention mechanism can be employed, and its computational complexity is \( O(n^2d + nd^2) \), i.e., the extra computational cost is \( nd^2 \). In this study, we propose DTE to aggregate transform matrices with attention weights learnt from an attention network based on user intentions. The key insight is that within reasonable cost of model size, DTE provides user-specific predictions and improves representation capability in an efficient way (low extra computational cost). Specifically, the extra computational cost of DTE is \( O(td^2) \) compared with the self-attention mechanism. As such, DTE is more computationally efficient than multi-head self-attention on the premise that \( t < n \) which usually holds (the optimal \( t \) is no larger than 4, while \( n \) is 50 in this study). Note that DTE takes only one scaled dot-product operation. This means that one can increase the capacity of the prediction model by increasing \( t \) with only a small increase in inference cost because each additional parameter requires only 1 additional multiply-add.

As noted in Table 2, our proposed dynamic transformer encoder advances the classic multi-head transformer encoder in following aspects. (1) Compared to multi-head transformer encoder, DTE provides a more efficient solution to boost the representational capability without increasing the depth (number of DTE blocks \( b \) and width (number of heads \( h \)). (2) By parameterizing the transform matrices in a dynamic self-attention layer as a linear combination of \( t \) experts, DTE is able to perform user-specific (input-dependent) prediction in an efficient manner.

**Algorithm 1. The RAISE Algorithm**

1: for each user \( u \in \mathcal{U} \) do
2: for each item candidate \( i \in S_u \) do
3: Obtain \( r_i^{(u)} \) and \( r_j^{(v)} \) with Eqs. (3), (4), and (5);
4: Compute \( S \) for the initial list \( S_u \) with Eqs. (6) and (7);
5: Compute \( y_{ui} \) with Eq. (12);
6: Calculate \( \mathcal{L} \) based on Eq. (13);
7: Update \( \Theta \) to minimize \( \mathcal{L} \), using \( \nabla_u \mathcal{L} \);

### 5 EXPERIMENTS

In this section, we conduct extensive experiments to answer the following research questions.

- **RQ1:** Is RAISE able to perform competitively to baseline re-ranking solutions?
- **RQ2:** How does the proposed IDM and DTE affect the performance of RAISE?
- **RQ3:** Is RAISE able to generate meaningful explanations for item recommendation?
- **RQ4:** How do the key hyper-parameters affect the performance of RAISE?
- **RQ5:** How does the running cost of RAISE compare with baselines?
- **RQ6:** How do the text reviews affect the performance of re-ranking methods?
- **RQ7:** Can we achieve better results with different co-attention functions and aggregation functions in the IDM module?

#### 5.1 Experimental Setup

**Datasets.** We conduct experiments on four public datasets from Amazon\(^2\) as listed in Table 3, including ‘Sports and
Outdoors, ‘Health and Personal Care,’ ‘Clothing, Shoes and Jewelry’ and ‘Video Games’. The selected datasets are widely used in recommendation studies and come from different domains. After obtaining the initial recommendation lists generated by GMF, we randomly select 80% users to construct the training set, use 10% users as the validation set, and leave the remaining 10% as the testing set for each dataset [6], [7]. In addition, we follow previous works [5], [7] to convert explicit rating data to binary implicit feedback. In particular, the label is 1 if the target user has rated the target item; otherwise 0.

Comparing Methods. We compare with the following state-of-the-art counterparts.

- **GMF** [5] generalizes the matrix factorization model in a non-linear manner, which is a widely adopted baseline for recommendation tasks. Note that GMF is the prior global ranking model of all re-ranking baselines in this paper.
- **DLCM** [6] is a classic re-ranking model, which encodes item candidates in the initial list sequentially with GRU.
- **PRM** [7] is constructed based on the transformer architecture. It employs a pre-trained model to generate personalized vectors for candidate items, which are then fed into the transformer network to refine initial item lists together with the latent representations learnt from GMF.
- **SetRank** [20] re-ranks items with the multi-head self-attention mechanism and an attentive loss function. We implement both the SetRank with and without positional embeddings, which are denoted as SetRank\(_w/o\) PE and SetRank\(_w/o\) PE, respectively.

Note that the source code of Seq2Slate [21] is not released and our re-implemented version performs unsatisfactorily. Meanwhile, we also find that the performance of IRGPR [8] is poor in our experimental setting, although we have carefully tuned its parameters. Therefore, we omit the comparison with Seq2Slate and IRGPR in this paper.

Training Details. For a fair comparison, we set \(d = 32\) and \(n = 50\) for all re-ranking methods. We follow the configuration presented in [5], [6], [7], [20], and all the baselines are trained until convergence. For our proposed RAISE, we tune the number of hidden layers from 1 to 4 for MLP structures. The maximum number of reviews \(l_u\) and \(l_i\) are set to 20. Instead of constructing and fine-tuning for an end-to-end NLP model to obtain the word representations for text information, we obtain pre-trained language model word representations from BERT’s pre-trained model in this work since many NLP tasks are benefit from BERT to get the SOTA [41]. The learning rate is selected from \(\{e^{−1}, e^{−2}, e^{−3}, e^{−4}\}\); batch size is chosen from \(\{256, 512, 1024\}\) and dropout rate varies in the range of \([0.1, 0.5]\) stepped by 0.1. Moreover, the number of transform matrices \(t\) and the number of DTE blocks \(b\) are searched from \(\{1, 2, 4, 8, 10\}\), and \(\{1, 2, 3, 5, 8, 10\}\). Our model is implemented with Pytorch, optimized with Adam [42] and trained on one Nvidia TITAN Xp GPU with 12 GB memory associated with Intel Exon CPU E5-2630 v4@2.20 GHz.

Evaluation Metrics. We adopt the same evaluation metrics with PRM [7] to evaluate the performance of all methods: Precision (Pre@\(k\)) and Mean Average Precision (MAP@\(k\)), where \(k\) is the length of the recommendation list. Precision evaluates the fraction of correct recommendations in recommendation lists for all users, and MAP computes the mean average precision of all ranked lists cut off by \(k\). In addition, we also evaluated the recommendation performance by normalized discounted cumulative gain (NDCG@\(k\)) which takes the position of correct recommendations into account [4], [5]. Note that higher metric values indicate a better recommendation performance.

### 5.2 Experimental Results

#### 5.2.1 Performance Comparison (RQ1)

Table 4 presents the overall performance of the proposed RAISE and comparing methods on four Amazon datasets. As a whole, all re-ranking methods are able to refine the initial recommendation list generated by the global ranking model GMF. This verifies the effectiveness of modeling inter-item patterns for re-ranking tasks. Moreover, we can observe that RAISE performs better than the other three state-of-the-art re-ranking models: DLCM, SetRank and PRM. In particular, RAISE achieves up to 13.95% relative improvement w.r.t. Pre@\(5\), 12.30% relative improvement w.r.t. MAP@\(5\), and 9.60% relative improvement w.r.t. NDCG@\(5\) on ‘Health and Personal Care’ dataset, compared to the second best re-ranking algorithm. We also observe that RAISE gains slightly less improvement on the ‘Clothing, Shoes and Jewelry’ dataset. A possible explanation could be the difficulty of model training due to the extremely sparse data, especially for RAISE which refines recommendations with review information, as we can see the review density of this dataset is lower than 0.05% as shown in Table 3. Another interesting finding is that considering the positional embeddings for SetRank leads to better results on the four datasets.

#### 5.2.2 Effects of IDM and DTE (RQ2)

To investigate the effectiveness of two key components of RAISE, we compare RAISE with its three variants listed as follows: (1) RAISE\(_w/o\) IDM: removing the intention-aware representations learnt via IDM in Eq. (6); (2) RAISE\(_w/o\) DTE: replacing DTEs with normal transformer encoders as shown in Eq. (1); (3) RAISE\(_w/o\) Both: removing both IDM and DTEs
| Datasets                      | Models                        | Pre@1 | Pre@5 | Pre@10 | MAP@5 | MAP@10 | NDCG@5 | NDCG@10 |
|------------------------------|-------------------------------|--------|-------|--------|-------|--------|--------|---------|
|                              | GMF                           | 32.14  | 34.76 | 26.73  | 46.76 | 48.66  | 42.26  | 51.56   |
|                              | DLCM                          | 55.12  | 41.44 | 29.48  | 63.56 | 60.08  | 52.77  | 59.99   |
|                              | PRM                           | 79.44  | 54.83 | 35.39  | 81.53 | 76.98  | 70.46  | 74.36   |
|                              | RAISE                         | 85.45* | 60.63 | 37.77  | 86.96*| 83.48*| 77.84*| 80.30*  |
|                              | Improvement                   | 7.57%  | 10.58%| 6.73%  | 6.66% | 8.44%  | 10.47% | 7.99%   |
|                              | GMF                           | 30.65  | 33.52 | 26.07  | 44.66 | 46.56  | 40.72  | 49.65   |
|                              | DLCM                          | 49.23  | 37.73 | 27.52  | 57.70 | 55.41  | 48.48  | 56.17   |
|                              | PRM                           | 73.96  | 52.17 | 34.56  | 78.04 | 74.15  | 67.61  | 72.01   |
|                              | RAISE                         | 83.08* | 59.48*| 36.73  | 85.53*| 82.21*| 76.42*| 78.99*  |
|                              | Improvement                   | 12.33% | 13.95%| 6.28%  | 9.60% | 10.87% | 13.03% | 9.69%   |
|                              | GMF                           | 42.83  | 46.52 | 33.24  | 59.69 | 62.42  | 57.96  | 67.19   |
|                              | DLCM                          | 68.06  | 50.79 | 34.65  | 73.40 | 71.10  | 67.02  | 74.32   |
|                              | PRM                           | 79.08  | 54.60 | 35.73  | 81.95 | 77.10  | 65.26  | 62.00   |
|                              | RAISE                         | 83.37* | 59.08*| 38.28  | 84.71 | 80.66  | 72.22  | 82.18   |
|                              | Improvement                   | 3.52%  | 6.75% | 3.25%  | 3.44% | 4.68%  | 5.35%  | 3.41%   |
|                              | GMF                           | 40.80  | 43.66 | 34.23  | 57.23 | 57.86  | 50.39  | 58.63   |
|                              | DLCM                          | 58.82  | 47.09 | 35.33  | 66.77 | 64.00  | 56.68  | 63.86   |
|                              | PRM                           | 69.47  | 50.16 | 36.19  | 76.06 | 70.55  | 61.93  | 66.76   |
|                              | RAISE                         | 71.93  | 54.11 | 38.53  | 77.37 | 72.78  | 65.90  | 70.57   |
|                              | Improvement                   | 8.98%  | 6.58% | 4.23%  | 5.54% | 5.58%  | 7.48%  | 5.75%   |
|                              | GMF                           | 60.63  | 37.77 | 86.96  | 83.48 | 77.84  | 80.30  |
|                              | DLCM                          | 59.70  | 37.44 | 86.88  | 82.76 | 76.85  | 79.40  |
|                              | PRM                           | 59.18  | 37.22 | 86.14  | 82.03 | 76.23  | 78.94  |
|                              | RAISE                         | 58.50  | 36.78 | 86.00  | 81.83 | 75.55  | 78.28  |
|                              | Improvement                   | 45.8%  | 42.3% | 5.54%  | 5.58% | 7.48%  | 5.75%  |

'Improvement' indicates how much RAISE has improved over the second best method which is underlined. We use '*' to denote statistically significant improvements (paired t-test with p-value < 0.05). Note that "MAP@5" and "NDCG@5" is equal to the "Pre@1," so they are omitted in this table. We run all methods for 5 times and report their average value in this table.

| Dataset                                | Model                        | Pre@1 | Pre@5 | Pre@10 | MAP@5 | MAP@10 | NDCG@5 | NDCG@10 |
|----------------------------------------|------------------------------|-------|-------|--------|-------|--------|--------|---------|
| Sports and Outdoors                    | RAISE                        | 60.63 | 37.77 | 86.96  | 83.48 | 77.84  | 80.30  |
|                                       | RAISE_{w/o IDM}              | 59.70 | 37.44 | 86.88  | 82.76 | 76.85  | 79.40  |
|                                       | RAISE_{w/o DTE}              | 59.18 | 37.22 | 86.14  | 82.03 | 76.23  | 78.94  |
|                                       | RAISE_{w/o Both}             | 58.50 | 36.78 | 86.00  | 81.83 | 75.55  | 78.28  |
| Health and Personal Care               | RAISE                        | 59.48 | 36.73 | 85.53  | 82.21 | 76.42  | 78.99  |
|                                       | RAISE_{w/o IDM}              | 58.34 | 36.12 | 84.08  | 81.02 | 75.06  | 77.92  |
|                                       | RAISE_{w/o DTE}              | 58.54 | 36.57 | 84.78  | 81.24 | 75.38  | 78.26  |
|                                       | RAISE_{w/o Both}             | 53.10 | 34.38 | 79.93  | 76.03 | 68.95  | 73.14  |
| Clothing, Shoes and Jewelry            | RAISE                        | 70.39 | 41.66 | 95.32  | 93.25 | 91.33  | 92.38  |
|                                       | RAISE_{w/o IDM}              | 69.10 | 40.86 | 94.81  | 92.59 | 90.19  | 91.53  |
|                                       | RAISE_{w/o DTE}              | 70.12 | 41.56 | 95.26  | 93.12 | 91.07  | 92.37  |
|                                       | RAISE_{w/o Both}             | 67.68 | 40.58 | 93.36  | 91.06 | 88.50  | 90.51  |
| Video Games                            | RAISE                        | 59.43 | 40.68 | 83.99  | 79.33 | 73.16  | 76.30  |
|                                       | RAISE_{w/o IDM}              | 57.43 | 39.65 | 82.06  | 77.51 | 70.66  | 74.08  |
|                                       | RAISE_{w/o DTE}              | 58.77 | 40.39 | 83.52  | 78.79 | 72.42  | 75.49  |
|                                       | RAISE_{w/o Both}             | 56.39 | 39.39 | 81.14  | 76.24 | 69.43  | 73.30  |
from RAISE simultaneously. The results are reported in Table 5. By analyzing the performance comparison, we have the following observations. (1) Being equipped with IDM, our proposed RAISE is able to effectively learn useful patterns from review information and further enrich the representations of users and items. (2) By employing an attention network over the self-attention layer, the proposed DTE performs better than the normal transformer encoder under our setting, showing its stronger representational capability without increasing the depth or width. (3) The combination of IDM and DTE brings the largest improvements, which showcases that RAISE is able to generate more personalized recommendations by seamlessly accommodating the learnt latent user intentions extracted from the text reviews.

One may wonder why performance gaps exist among RAISE w/o Both, SetRank, and PRM methods since they are all based on the transformer architectures. This could come from the following three reasons. First, the input data is different, which results in the different performance. In particular, RAISE improves the personalization of the recommendation engine by empowering it to be input-dependent, while the PRM model employing a pre-trained model to generate personalized vectors for candidate items, which are learned from implicit feedback. As mentioned in their paper, such personalized vectors encode users’ generic preferences and could be beneficial to the predictions. They are then fed into the transformer architecture together with item latent representations \( q_i \). As such, the input of PRM is not the same as ours, because RAISE w/o Both is trained with both user representations \( p_u \) and item representations \( q_i \) in an end-to-end manner, rather than encoding user information via a pre-trained model. In summary, SetRank and RAISE w/o Both take the latent representation of users \( p_u \) as input, whereas PRM does not. Second, the range of hyper-parameters are different. As mentioned in the Section 5.1, in order to reproduce the results of comparing methods and make a fair comparison, we follow the same configurations presented in their paper, including the search spaces of hyper-parameters, which could make a considerable impact on the model performance. Third, the choice of loss functions could be influential to model performance. To be specific, SetRank adopts an attentive loss function, which is different from the negative log likelihood loss used in PRM and RAISE.

5.2.3 Interpretability of RAISE (RQ3)

We now discuss another important property of our proposed RAISE: the interpretability. To this end, we randomly pick eight user-item review pairs with high matching scores among the testing set of the four datasets and highlight the similar intention(s) between user and item reviews in Fig. 4. Taking Fig. 4a as an example, user 719 wrote about “I feel so well equipped when I have this multi-tool with me...” and then the co-attention mechanism assigns higher matching score to the item review that mentioned “This one has all the tools I was looking for.” We find that, even there are very few words that occur in common between two reviews, the selected review pairs are identified to be consistent with regard to some high-level concepts. This indicates it is a practical way to find meaningful item reviews for the target user. By distinguishing the importance of reviews between two input sequences \( R_u \) and \( R_i \), RAISE is capable of boosting the interpretability of recommendation engines.

5.2.4 Hyper-Parameter Sensitivity Analysis (RQ4)

In this section, we examine how the number of transform matrices \( t \), the length of the initial recommendation list \( n \) and the number of DTE blocks \( b \) affect the performance of RAISE. The results are depicted in Figs. 5, 6, and 7, and

Fig. 4. Selected user-item review pairs with high matching scores from the four datasets.
similar trends can be observed with \( k = \{10, 20\} \) as well as the rest two datasets.

From Fig. 5, we can see that a small number of transform matrices \( (t \leq 4) \) is sufficient to distinguish the main intentions of users. This indicates our DTE does not actually add much space complexity and verifies our assumption that user behavior is driven by multiple intentions with different contributions. From Fig. 6, we notice that the performance of RAISE initially improves with the increase of \( n \) \((n \leq 50)\), and the performance begins to decrease after a certain point \((n > 50)\) on both datasets. Intuitively, a longer initial list contains more item candidates and richer inter-item patterns, which is beneficial to re-ranking models. However, since RAISE is empowered by the transformer architecture, it may not be a good choice to blindly increase the length of initial list, as the transformer architecture is quadratic to \( n \) (see Model Complexity and Scalability). Besides, a long initial list could be hard for the transformer architecture to model effective mutual relationships. Furthermore, the results on the number of DTE blocks are plotted in Fig. 7, indicating that only one or two DTE blocks are usually sufficient for RAISE to achieve its optimal performance.

5.2.5 Running Time Comparison (RQ5)

Since we only model limited items for targets users in the re-ranking process (50 items for each user), it could be more efficient to model text information in the re-ranking process rather than modeling text information by a global ranking model as most review-aware ranking models do. However, modeling text reviews may result in significantly higher costs for RAISE compared to approaches that do not take reviews into account. To investigate the computational costs in practical situations, we hence compare the running time of all methods as shown in Table 6.

We can observe that DLCM takes the most running time. Meanwhile, RAISE takes much more time than PRM and SetRank due to the review modeling process happened in the IDM. Although modeling text review takes unexpected computational costs, it benefits the re-ranking model in threefold.

- First, IDM distinguishes the importance of reviews according to the intentions behind them, enabling RAISE to capture user-specific inter-item relationships and hence perform user-specific predictions.
- Second, modeling text information brings additional performance gains (see Section 5.2.2).
- Third, RAISE is able to provide meaningful explanations for target users by modeling text information with the co-attention mechanism (see Section 5.2.3).

In summary, with more time spent, RAISE provides better personalization and interpretability than comparing methods by modeling text information. In order to alleviate the above problem and decrease the training time, future works will introduce efficient attention mechanisms for the computations of co-attention and dynamic self-attention. For example, localizing the attention span and using memory-compressed attention are simple yet effective methods to decrease computational costs [43].

5.2.6 Effects of Text Review (RQ6)

To investigate the effect of user reviews and item reviews, we compare RAISE with its two variants listed as follows: (1) RAISE\textsubscript{w/o urvw}: removing the user review information from RAISE; (2) RAISE\textsubscript{w/o irvw}: removing the item review information from RAISE. As shown in the Table 7, we can see the default setting of RAISE (i.e., considering both user and item reviews) performs the best on three out of four presented datasets. Surprisingly, the model performance on the video games dataset indicates that modeling item reviews only (i.e., RAISE\textsubscript{w/o urvw}) is able to achieve better results than RAISE which models both user and item reviews. This indicates the importance of modeling item reviews for the Video Games dataset.

Furthermore, we feed both text information and latent representation learned from implicit feedback into all re-

![Fig. 6. Impact of \( n \) at Pre@5(%) and MAP@5(%).](image)

**Fig. 6. Impact of \( n \) at Pre@5(%) and MAP@5(%).**

![Fig. 7. Impact of \( b \) at Pre@5(%) and MAP@5(%).](image)

**Fig. 7. Impact of \( b \) at Pre@5(%) and MAP@5(%).**

---

**Table 6**

| Dataset                        | Model      |
|-------------------------------|------------|
|                               | DLCM       | PRM       | SetRank   | RAISE     | RAISE\textsubscript{w/o IDM} |
| Sports and Outdoors           | 00:59:51   | 00:10:54  | 00:26:09  | 04:22:58  | 00:24:08             |
| Health and Personal Care      | 05:35:59   | 00:09:13  | 00:26:48  | 05:01:59  | 00:24:59             |
| Clothing, Shoes and Jewelry   | 05:36:45   | 00:11:57  | 00:27:29  | 05:14:56  | 00:19:14             |
| Video Games                   | 04:59:51   | 00:06:14  | 00:18:04  | 03:07:19  | 00:11:13             |

*We follow the default setting of DLCM and train it for 10,000 iterations. As for PRM, SetRank RAISE and RAISE\textsubscript{w/o IDM}, they are trained for 100 epoch respectively.*
ranking methods, to further investigate how much the text information can improve the model performance. Specifically, given an embedding review sequence \( \mathcal{R}_i = \{r^{(i)}_1, r^{(i)}_2, \ldots, r^{(i)}_l\} \), we first obtain the representations of text information for each item by summing up its corresponding representation of reviews:

\[
r_i = \sum_j r^{(j)}_j.
\]

Then we concatenate the representations of text information \( r_i \) and item latent representation \( q_i \) to replace the original item representations \( q_i \) in the baseline re-ranking models. In addition, we implement the IDM for the PRM model to further evaluate the effectiveness of IDM since the PRM model achieves the best results on four datasets among all baseline methods, which is denoted as \( \text{PRM}_{\text{IDM}} \). From Table 8, we can see the model performance of review-aware baselines do not improve their performance as expected. In particular, considering text information significantly degrades the performance of DLCM on all datasets. This showcases that the model architectures of baselines should be modified accordingly in order to leverage reviews, and simply introducing textual information as input can lead to worse results. Besides, the IDM-enhanced PRM (i.e., \( \text{PRM}_{\text{IDM}} \)) consistently outperforms \( \text{PRM}_{\text{review}} \). Compared with \( \text{PRM}_{\text{review}} \), it generates more meaningful review representations for the prediction model by computing matching scores between every user-item review pair with a co-attention network. This helps confirm the effectiveness of the IDM module.

### Table 7
The Effects of Text Reviews of RAISE (%)

| Dataset                  | Model | Pre  | MAP  | NDCG |
|--------------------------|-------|------|------|------|
|                          |       | @5   | @10  |      | @5   | @10  |      |       |
| Sports and Outdoors      | RAISE | 60.63| 37.77| 86.96| 83.48| 77.84| 80.30|
|                          | RAISE \(_{w/o \text{rvw}}\) | 59.94| 37.36| 86.30| 82.85| 77.00| 79.10|
|                          | RAISE \(_{w/o i_{\text{rvw}}}\) | 59.04| 37.25| 85.69| 81.71| 76.02| 78.56|
| Health and Personal Care | RAISE | 59.48| 36.73| 85.53| 82.21| 76.42| 78.99|
|                          | RAISE \(_{w/o \text{rvw}}\) | 57.86| 36.10| 83.84| 80.85| 75.03| 77.77|
|                          | RAISE \(_{w/o i_{\text{rvw}}}\) | 57.87| 36.13| 84.49| 80.96| 75.17| 77.83|
| Clothing, Shoes and Jewelry | RAISE | 70.39| 41.66| 95.32| 93.25| 91.33| 92.38|
|                          | RAISE \(_{w/o \text{rvw}}\) | 69.13| 41.07| 94.79| 92.50| 89.98| 91.44|
|                          | RAISE \(_{w/o i_{\text{rvw}}}\) | 69.24| 41.07| 94.69| 92.40| 90.01| 91.39|
| Video Games              | RAISE | 59.43| 40.68| 83.99| 79.33| 73.16| 76.30|
|                          | RAISE \(_{w/o \text{rvw}}\) | 59.81| 40.51| 84.09| 79.91| 73.61| 76.55|
|                          | RAISE \(_{w/o i_{\text{rvw}}}\) | 58.39| 40.17| 82.58| 78.09| 71.85| 75.24|

### Table 8
The Performance Comparison Between RAISE and Baselines With Review Information as Input (%)

| Dataset                  | Model   | Pre  | MAP  | NDCG |
|--------------------------|---------|------|------|------|
|                          |         | @5   | @10  |      | @5   | @10  |      |       |
| Sports and Outdoors      | DLCM\(_{\text{review}}\) | 4.80 | 4.86 | 9.09 | 9.71 | 5.23 | 6.57 |
|                          | SetRank\(_{\text{review}}\) | 49.48| 33.89| 76.39| 71.04| 63.49| 68.95|
|                          | PRM\(_{\text{review}}\) | 50.60| 34.31| 76.15| 71.17| 64.28| 69.53|
|                          | PRM\(_{\text{IDM}}\) | 55.08| 36.15| 82.55| 77.71| 71.20| 75.17|
|                          | RAISE   | 60.63| 37.77| 86.96| 83.48| 77.84| 80.30|
| Health and Personal Care | DLCM\(_{\text{review}}\) | 5.58 | 5.39 | 11.82| 12.18| 6.43 | 7.82 |
|                          | SetRank\(_{\text{review}}\) | 46.68| 32.07| 71.56| 67.41| 59.63| 65.50|
|                          | PRM\(_{\text{review}}\) | 49.38| 33.53| 75.34| 70.65| 63.28| 68.80|
|                          | PRM\(_{\text{IDM}}\) | 53.76| 35.03| 80.19| 76.07| 69.85| 73.97|
|                          | RAISE   | 59.48| 36.15| 85.33| 82.21| 76.42| 78.99|
| Clothing, Shoes and Jewelry | DLCM\(_{\text{review}}\) | 4.54 | 4.21 | 8.76 | 9.10 | 5.38 | 6.47 |
|                          | SetRank\(_{\text{review}}\) | 58.22| 37.91| 93.98| 79.66| 76.16| 81.21|
|                          | PRM\(_{\text{review}}\) | 61.64| 39.16| 86.72| 83.01| 80.33| 84.63|
|                          | PRM\(_{\text{IDM}}\) | 66.41| 40.60| 92.17| 89.26| 86.69| 89.30|
|                          | RAISE   | 70.39| 41.66| 95.32| 93.25| 91.33| 92.38|
| Video Games              | DLCM\(_{\text{review}}\) | 20.25| 19.84| 32.50| 32.91| 21.91| 28.09|
|                          | SetRank\(_{\text{review}}\) | 53.42| 38.15| 77.02| 72.15| 65.08| 69.50|
|                          | PRM\(_{\text{review}}\) | 51.79| 39.51| 74.35| 69.90| 62.52| 67.52|
|                          | PRM\(_{\text{IDM}}\) | 55.85| 39.52| 79.75| 74.76| 68.22| 72.61|
|                          | RAISE   | 59.43| 40.68| 83.99| 79.33| 73.16| 76.30|
### Table 9
Ablation Analysis for IDM (%)

| Dataset                | Model          | Pre  | MAP  | NDCG |
|------------------------|----------------|------|------|------|
|                        |               | @5   | @10  | @5   | @10  |
| Sports and Outdoors    | RAISE         | 60.63| 37.77| 86.96| 83.48| 77.84| 80.30|
|                        | Co-ATT (Soft) | 59.51| 37.09| 85.85| 82.28| 76.44| 78.56|
|                        | Co-ATT (MLP)  | 58.69| 37.01| 85.26| 81.39| 75.52| 78.07|
|                        | Aggr (Mean)   | 59.57| 37.37| 86.15| 82.36| 76.57| 78.88|
| Health and Personal Care | RAISE        | 59.48| 36.73| 85.53| 82.21| 76.42| 78.99|
|                        | Co-ATT (Soft) | 59.00| 36.76| 85.34| 82.27| 76.46| 78.97|
|                        | Co-ATT (MLP)  | 58.10| 36.38| 84.39| 81.26| 75.40| 78.15|
|                        | Aggr (Mean)   | 58.55| 36.40| 84.76| 81.50| 75.83| 78.32|
| Clothing, Shoes and Jewelry | RAISE      | 70.39| 41.66| 95.32| 93.25| 91.33| 92.38|
|                        | Co-ATT (Soft) | 70.18| 41.77| 95.58| 93.21| 91.06| 92.40|
|                        | Co-ATT (MLP)  | 70.83| 41.74| 95.89| 93.68| 91.71| 92.61|
|                        | Aggr (Mean)   | 70.86| 41.73| 95.86| 93.74| 91.71| 92.64|
| Video Games            | RAISE         | 59.43| 40.68| 83.99| 79.33| 73.16| 76.30|
|                        | Co-ATT (Soft) | 58.67| 40.14| 82.99| 78.51| 72.17| 75.44|
|                        | Co-ATT (MLP)  | 59.84| 40.46| 83.90| 79.72| 73.52| 76.34|
|                        | Aggr (Mean)   | 59.72| 40.43| 84.23| 79.75| 73.52| 76.32|

"Co-ATT" denotes the co-attention function used in the Eq. (3), and “Aggr” is short for aggregation function in the Eq. (5).

One may wonder whether the number of reviews affects the model complexity. In our work, we set the number of reviews $l_u$ and $l_i$ to a constant (i.e., 20), and the text reviews are pre-processed before training. In this way, RAISE loads user and item text representations for the training process, and the model complexity only grows with the number of users and items, and without regard to the number of reviews. In summary, additional memory costs are needed for training on larger datasets because RAISE needs to train and save the representations of users and items, which is very common for modern deep learning recommendation models.

#### 5.2.7 Ablation Analysis for IDM (RQ7)

In the IDM, we adopt a bilinear co-attention function (Eq. (3)) to compute the matching scores, which enables RAISE to capture intention-aware information and to provide meaningful explanations. We have tried two more co-attention function to compute the review-level matching scores in this section. The first one omits the transform matrix $M$ and the latter one utilizes only one MLP to compute matching scores [25].

\[
\begin{align*}
    c_{kj} &= f(r_i^{(u)})^T f(r_j^{(i)}) \
    c_{kj} &= f(\text{Concat}(r_i^{(u)}, r_j^{(i)})).
\end{align*}
\]

The above co-attention functions are denoted as Co-ATT (Soft) and Co-ATT (MLP), respectively. We present the performance of Co-ATT (Soft) and Co-ATT (MLP) in the Table 9. In addition, we change the aggregation function from the sum pooling to the mean pooling in order to discover the effect of different aggregation functions in the Eq. (5), which is denoted as “Aggr (Mean)” in the Table 9. We observe the relative ranking of all three variants (bilinear, soft and MLP) are always interchanging across different datasets. On the other hand, changing the aggregation function could marginally improve performance on the later two datasets.

In summary, the default setting can help achieve the best performance on some datasets. For the other datasets, although the default setting do not perform the best, there is not huge performance difference between different settings.

As such, there is no universal settings for all datasets, and it may vary case by case.

### 6 Conclusion and Future Works

In this paper, we propose a novel re-ranking method RAISE to refine the recommendation list. Equipped with the intention discover module (IDM) and dynamic transformer encoder (DTE), our proposed RAISE performs user-specific re-ranking by exploiting user intentions with the help of text reviews. By constructing RAISE upon prior global ranking models, one can easily achieve personalization, efficiency, and interpretability without modifying their current recommendation engines. Such scalability enables RAISE to refine recommendation lists generated by existing ranking models in an efficient manner. Empirical study verifies the additional gains brought by the devised IDM and DTE. In future work, we will investigate how to mine user intentions from other auxiliary information such as social networks and knowledge graphs for further performance-enhanced re-ranking approaches.

### References

[1] Z. Sun et al., “Research commentary on recommendations with side information: A survey and research directions,” Electron. Commerce Res. Appl., vol. 37, 2019, Art. no. 100879.

[2] Z. Lin, L. Feng, R. Yin, C. Xu, and C. K. Kwoh, “GLIM: Global and local item graphs for top-n recommender systems,” Inf. Sci., vol. 580, pp. 1–14, 2021.

[3] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” Computer, vol. 42, no. 8, pp. 30–37, 2009.
Zhuoyi Lin received the BS degree in electronic engineering from Ming Chuan University, Taiwan, in 2017. He is currently working toward the PhD degree with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He is a research scientist with the Institute for Infocomm Research (I²R), Agency for Science, Technology and Research (A∗STAR), Singapore. His research interests include collaborative filtering, deep learning and their applications.

Sheng Zang received the BS degree in computer science and technology from Nankai University, China. His research interests mainly include deep learning, reinforcement learning and their applications.

Rundong Wang is currently working toward the PhD degree with the School of Computer Science and Engineering, Nanyang Technological University. His research interest are (Multi-agent) Reinforcement Learning and application in Game, Finance, and so on. He has published several papers in the top AI conference such as ICML, ICLR, AAAI, IJCAI, etc.
Zhu Sun received the PhD degree from Nanyang Technological University, Singapore, in 2018. Her research interests include machine learning, data mining and recommender systems. Her research has been published in leading conferences and journals in related domains (e.g., IJCAI, AAAI, CIKM, ACM RecSys, IEEE Transactions on Knowledge and Data Engineering and IEEE Transactions on Neural Networks and Learning Systems). Currently, she is with Institute of High Performance Computing and Centre for Frontier AI Research, A*STAR, Singapore.

J. Senthilnath (Senior Member, IEEE) received the PhD degree in aerospace engineering from the Indian Institute of Science (IISc), India. He is a scientist with the Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore. His current research interests include artificial intelligence, multi-agent systems, generative models, online learning, reinforcement learning and optimization. He has published more than 100 high-quality papers and won five best paper awards. He is a member of Artificial Intelligence, Analytics And Informatics (AI3), A*STAR. He has been serving as guest editor/organizing chair/co-chair/PC members in leading AI and data analytics journals and conferences.

Chi Xu received the bachelor's and PhD (Honors) degrees from Nanyang Technological University (NTU), Singapore, in 2003 and 2010 respectively. After joined SIMTech A*Star, in 2010, he has been working in the areas of information management and operations analytics for enterprise and supply chain applications. He has developed various technologies which have been licensed to companies in different industries such as aerospace, construction, transportation&logistics, and fast moving consumer goods (FMCG). Currently, he is deputy group manager of planning and operations management group (POM) of SIMTech. He is responsible to drive the group research direction and capabilities development for operation analytics and system optimization. He is an adjunct assistant professor of the School of Computer Science and Engineering (SCSE), Nanyang Technological University (NTU).

Chee Keong Kwoh (Senior Member, IEEE) received the bachelor’s degree (Hons.) in electrical engineering and the master’s degree in industrial system engineering from the National University of Singapore, Singapore, in 1987 and 1991, respectively, and the PhD degree from the Imperial College of Science, Technology and Medicine, University of London, in 1995. He has been with the School of Computer Engineering, Nanyang Technological University (NTU), since 1993. His research interests include data mining, soft computing and graph-based inference; applications areas include bioinformatics and biomedical engineering. He has done significant research work in his research areas and has published many quality international conferences and journal articles. He has often been invited as an organizing member or referee and a reviewer for a number of premier conferences and journals, including GIW, IEEE, BIBM, RECOMB, PRIB, BIBE, ICDM, and iCBBE. He is also a member of the Association for Medical and Bioinformatics, Imperial College Alumni Association of Singapore. He has provided many services to professional bodies in Singapore and was conferred the Public Service Medal by the president of Singapore, in 2008. His research interests include data mining, soft computing and graph-based inference; applications areas include bioinformatics and biomedical engineering. He is an editorial board member of the International Journal of Data Mining and Bioinformatics, the Scientific World Journal, Network Modeling and Analysis in Health Informatics and Bioinformatics, Theoretical Biology Insights, and Bioinformation. He has been a guest editor of many journals, such as the Journal of Mechanics in Medicine and Biology, the International Journal on Biomedical and Pharmaceutical Engineering, and others.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.