

对多行为序列推荐中的工作 RIB 的复现与改进

陈小青

摘要

推荐系统是在线购物网站的重要组成部分，旨在向用户推荐其未来可能会购买的物品。随着电子商务的快速发展，在线购物网站记录了大量的用户与物品之间的交互行为。这些交互行为往往具有序列性，同时它们通常是异构的，即存在多种行为类型，如浏览、加入购物车和购买等。同时考虑行为序列性和异构性的推荐问题称为多行为序列推荐（MBSR）。由于 MBSR 更接近真实的推荐场景，但也面临着同时考虑序列性和异构性带来的新挑战。因此，MBSR 值得进一步研究。本文首先将对 MBSR 进行背景和相关工作进行介绍，接着对复现的工作 RIB 进行介绍，然后使用 PyTorch 进行复现，并在此基础上提出了在 GRU 模块中的两点改进。在基于数据集上的实验结果表明本文复现结果的准确性以及改进方法的有效性，从而实现了更高的推荐性能。

关键词：推荐系统；多行为序列推荐；

1 引言

现如今，随着科技发展，人们越来越通过互联网获取信息。然而，面对海量且复杂的网络信息，传统搜索引擎并不能很好地为每个用户过滤不重要或不感兴趣的信息，让人们面临着信息过载的问题。而能有效解决信息过载问题，并能够针对不同用户提供不同的个性化信息的推荐系统则具有重大意义。

推荐系统（Recommender Systems）^[1]是在线购物网站的重要组成部分，旨在向用户推荐其未来可能会购买的物品。推荐系统通常根据用户的历史行为数据来分析其对交互过和未交互过的物品的偏好，从而根据偏好值排序来挑选偏好预测值最高的一个或一部分物品，进而推荐给用户。

由于点赞或评分等显式反馈数据收集较为困难，因此，研究者更偏向于使用更容易获得的点击和加购等隐式反馈数据来作为用户的历史交互行为数据。基于某一种隐式反馈的推荐问题被称为单行为推荐，目前已有许多研究者进行了相关的研究^[2-4]。然而，在真实的场景中，人们和物品的交互行为往往是异构的，即行为有多种类型。例如，在电子商务平台中，人们可能会点击某个物品，加购某个物品，或是购买某个物品。同时，只采用一种行为类型较难直接表征用户的偏好和兴趣，而单一行为包含较少数据量也往往会造成冷启动问题^[5]。因此，研究者转而研究多行为推荐（Multi-behavior Recommendation, MBR）^[6-7]。

与 SBR 不同的是，单行为序列推荐（Single-behavior Sequential Recommendation, SBSR）在 SBR 问题的基础上考虑了用户行为所具有的序列性，如用户可能会先点击一个物品后点击或购买一个物品。目前，对于 SBSR，研究者已经提出了一些较为先进的方法来获得更佳的推荐性能^[8-10]。而在 SBSR 的基础上，考虑用户行为的异构性的推荐问题被称为多行为序列推荐（Multi-behavior Sequential Recommendation, MBSR）^[11-13]。MBSR 同时对用户的异构行为和行为的序列性进行建模，其更接近用户与物品的真实交互行为，从而可以为用户推荐其更感兴趣的物品。因此，设计一种用于 MBSR 的推荐算法具有重要的现实意义。

在推荐系统领域中，MBSR 是近几年来研究热点，目前已经应用于多个真实场景中，包含电子商务^[14-15]、视频推荐^[16-17]和新闻推荐^[18-19]等。为了更好地学习用户的真实偏好，研究人员还在 MBSR 的建模中纳入了停留时间和类别信息等。然而，MBSR 也面临一些困难和挑战。例如，如何同时对行为的异构性和序列性进行合适的建模，如何处理多种行为中的噪音和偏差，以及如何处理用户的长期偏好和短期偏好等问题。因此，在面临着这些挑战下，如何研究 MBSR 问题仍然值得进一步探讨。因此，本文主要对 MBSR 的经典工作 RIB^[12]进行复现和提出一些改进方法，以此来提高对 MBSR 的建模过程的认识和理解。

2 相关工作

目前，研究者已经提出了一些工作以解决 MBSR 问题。在这些工作中，所使用的方法在技术层面上可以分为基于矩阵分解的方法和基于深度学习的方法，这也是推荐系统领域中普遍使用的三大类方法的其中两种。由于推荐系统的另一类方法（基于邻域的方法）在 MBSR 问题中捕捉用户偏好的能力较差，效率也较低，因此研究人员往往不采用基于邻域的方法来解决 MBSR 问题，而是采用效率更高的基于矩阵分解的方法和基于深度学习的方法。

2.1 基于矩阵分解的方法

在 MBSR 中，由于深度学习在推荐系统的多个领域中往往表现出更好的性能，因此较少研究者使用基于矩阵分解的方法，目前较为经典的工作是 TransRec++^[20]。

TransRec++ 同时考虑了用户的点击和购买两种行为，并利用了矩阵分解的思想引入了行为转移向量来捕捉用户行为内和行为间的序列性，其中行为转移向量包含从点击到点击、从点击到购买、从购买到点击以及从购买到购买四种类型。TransRec++ 从行为转移的角度来建模 MBSR 问题，实现了较好的推荐性能。然而，在行为转移的建模过程中也可能存在噪音，如用户发生一个点击行为后的下一个点击行为可能是误点的情况。

2.2 基于深度学习的方法

基于深度学习的方法是推荐领域中最常用的建模方法。对于 MBSR 问题，目前已有一些采用深度学习进行建模的工作，主要可以分为基于 RNN 的方法、基于 GNN 的方法、基于 Transformer 的方法以及基于多种深度学习框架的方法。

2.2.1 基于 RNN 的方法

循环神经网络（Recurrent Neural Network, RNN）^[21]是一种经典的深度学习方法，在处理具有序列性的数据方法具有较好的性能。因此，研究人员使用了 RNN 的方法来对 MBSR 问题进行建模。

在这些相关的研究工作中，研究者使用的建模方法有所不同。例如，RLBL^[11]考虑了行为转移矩阵，以此来捕捉多种行为类型对用户偏好的影响；RIB^[12]将物品 embedding 和行为 embedding 进行拼接，然后将其送入 RNN 层和注意力层中，同时在建模过程中也考虑停留时间；BINN^[13]和 IARS^[22]分别通过改变 LSTM^[23]和 GRU^[24]的结构来提高推荐性能；DeepRec^[25]从局部和全局两种角度利用了多行为序列信息来进行跨服务器和客户端的联邦推荐；等等。

2.2.2 基于 GNN 的方法

图神经网络 (Graph Neural Network, GNN)^[26-27]也是深度学习中一种提取特征的方法。由于 GNN 可以充分利用节点之间的高阶邻居信息, 因此使用 GNN 的方法来对 MBSR 进行建模可以实现较好的推荐性能。

对于 MBSR 问题, 有研究者提出了 MGNN-SPred^[28], 在基于 GNN 的方法中考虑了行为转移信息, 从而取得了较好的推荐性能。在此之后, 也有研究者在这项工作的基础上, 提出了分别从全局图和个性化图中捕捉行为转移信息的 GPG4HSR^[29], 构建不同行为转移有向图的 BGNN^[30], 以及为不同行为序列构建有向图的 BA-GNN^[31]。在 MBSR 中, GNN 构建了用户-物品图, 增强了邻居节点的信息, 从而提高了推荐性能。

2.2.3 基于 Transformer 的方法

Transformer^[32]包含编码器和解码器两个部分, 并利用自注意力机制来提高训练速度。Transformer 在序列建模任务中表现优异, 因此也被应用于 MBSR 的相关研究工作中。

目前, 已有部分工作使用 Transformer 来解决 MBSR 问题。例如, DFN^[33]和 DUMN^[34]通过注意力网络去除用户行为中的噪音信息, FeedRec^[35]在 Transformer 中同时使用隐式反馈数据和显式反馈数据来预测用户的积极兴趣和消极兴趣, 以及 FLAG^[36]从全局和局部的角度使用 Transformer 来建模用户的全局偏好和个性化偏好以及用户个性化意图, 等等。

2.2.4 基于多种深度学习框架的方法

在 MBSR 中, 除了使用某一种深度学习方法之外, 有研究者也考虑了在建模中同时使用不同的深度学习方法, 如 GNN+RNN 和 GNN+Transformer 等, 从而融合了不同深度学习方法的优点, 也在一定程度上起到互补的作用。

在同时使用 RNN 和 GNN 的方法中, MKM-SR^[37]使用了 GGNN^[38]和 GRU 来构建多行为序列模型; MBGNN citeMLBDBI2021-MBGNN 利用了一个特定行为的物品序列和一个行为序列, 从全局的角度上进行建模。在同时使用 GNN 和 Transformer 的方法中, MBHT^[39]使用了 Transformer 和 GNN 分别建模用户的短期偏好和长期偏好; KHGT^[40]在同时使用 GNN 和 Transformer 的基础上考虑了物品间的关系来提高推荐效果。

3 本文方法

本文主要复现的工作是于 2018 年发表在 WSDM 会议上的一篇论文 “Micro Behaviors: A New Perspective in E-commerce Recommender Systems”。该论文提出了结合微观行为序列的推荐 (recommendation with sequences of micro behaviors, RIB)^[12]以多行为序列进行建模。RIB 在建模行为的异构性和序列性时还考虑了行为的停留时间。由于纯粹的 MBSR 问题一般只包含用户 ID、物品 ID、行为类型信息以及时间戳信息, 而不包含行为的停留时间。因此, 本文主要介绍 RIB 如何在纯粹的 MBSR 问题上进行建模。

3.1 本文方法概述

RIB 是基于 RNN 的学习方法来对 MBSR 进行建模，并使用交叉熵损失函数对模型进行训练，其示意图如 1 所示。RIB 一共包含输入层、GRU 层、注意力层和输出层。下面我们将一一对其进行介绍。

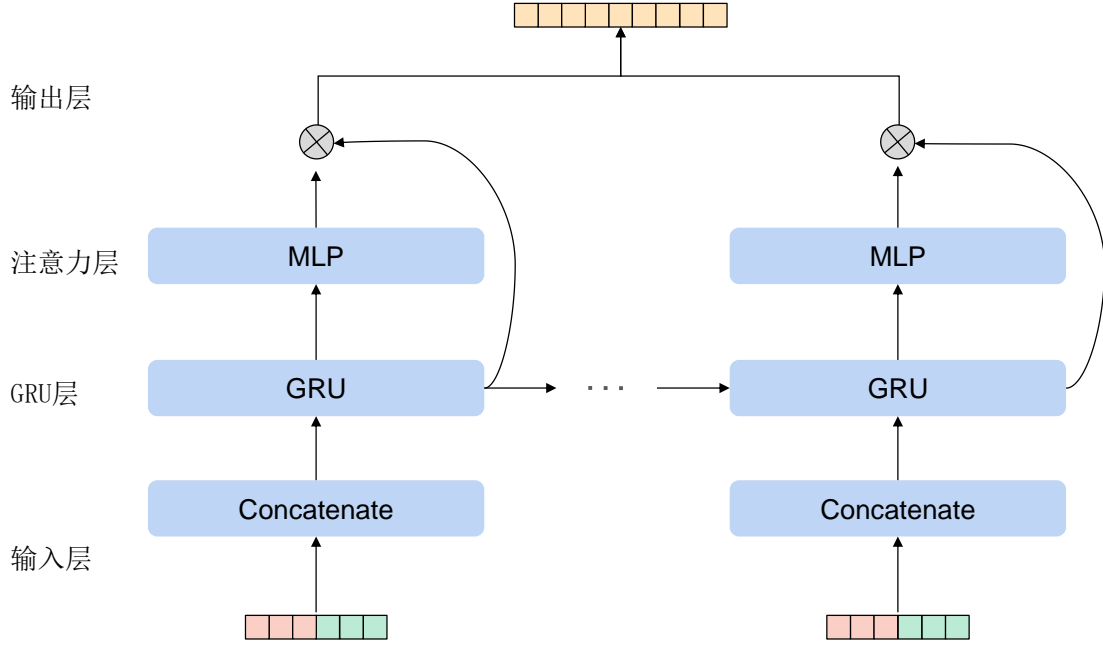


图 1: RIB 示意图

3.2 输入层

在输入层中，RIB 将物品表征 $V_{i_u^t}$ 和行为表征 $F_{f_u^t}$ 来表示用户 u 的交互序列 S_u 中的每个 (物品, 行为) 对在 t 时刻的物品 i_u^t 和行为 f_u^t 。接着，RIB 将物品表征和行为表征进行拼接，从而得到 t 时刻下的潜在表征：

$$\mathbf{e}_t = \text{concatenate}(V_{i_u^t}, F_{f_u^t}) \quad (1)$$

3.3 GRU 层

在 GRU 层中，输入层得到的潜在表征 \mathbf{e}_t 将作为 GRU 层的输入，以便通过对多行为序列建模来获得每个时间步的隐藏状态。在 t 时刻，GRU 内部的重置门 $\mathbf{r}_t \in \mathbb{R}^{d \times 1}$ 、更新门 $\mathbf{z}_t \in \mathbb{R}^{d \times 1}$ 、内部隐藏状态 $\mathbf{c}_t \in \mathbb{R}^{d \times 1}$ 和外部隐藏状态 $\mathbf{h}_t \in \mathbb{R}^{d \times 1}$ 的计算公式如下所示：

$$\mathbf{r}_t = \sigma(\mathbf{W}_{\text{er}}\mathbf{e}_t + \mathbf{W}_{\text{hr}}\mathbf{h}_{t-1}) \quad (2)$$

$$\mathbf{z}_t = \sigma(\mathbf{W}_{\text{ez}}\mathbf{e}_t + \mathbf{W}_{\text{hz}}\mathbf{h}_{t-1}) \quad (3)$$

$$\mathbf{c}_t = \tanh(\mathbf{W}_{\text{ec}}\mathbf{e}_t + \mathbf{W}_{\text{hc}}(\mathbf{r}_t \cdot \mathbf{h}_{t-1})) \quad (4)$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \mathbf{h}_{t-1} + \mathbf{z}_t \mathbf{c}_t \quad (5)$$

其中， $\mathbf{W}_{(\cdot)}$ 是待学习的模型参数， σ 是 Sigmoid 激活函数。

3.4 注意力层

注意力层将 GRU 层得到的隐藏状态 \mathbf{h}_t 作为输入，来获取每个时间步的注意力权重 α_t ：

$$\mathbf{M}_t = \tanh(\mathbf{W}_{\text{m}}\mathbf{h}_t + \mathbf{b}_{\text{m}}) \quad (6)$$

$$\alpha_t = \text{softmax}(\mathbf{W}_{\text{a}}\mathbf{M}_t + \mathbf{b}_{\text{a}}) \quad (7)$$

其中, $W_{(\cdot)}$ 、 $b_{(\cdot)}$ 和 M_t 都是待学习的模型参数。

3.5 输出层

最后, 在输出层, RIB 将获得的每个时间步的隐藏状态 h_t 与其对应的注意力分数 α_t 相乘后求和, 以此获得整条序列的整体潜在表征 h , 并将此输入到一个全连接层来获得用户对每个物品的预测偏好值 \hat{y} :

$$h = \sum_{t=1}^{|S_u|} \alpha_t h_t \quad (8)$$

$$\hat{y} = \text{Softmax}(Wh + b) \quad (9)$$

$$(10)$$

其中, W 和 b 分别是待学习的权重参数和偏置项。

4 复现细节

4.1 与已有开源代码对比

在代码可得性方面, RIB 的作者并没有在原论文中提供代码, 而之后有研究人员基于 TensorFlow 实现了该方法^[41]。为了有所不同, 本文将基于 PyTorch 框架来实现 RIB, 并采用与上述论文相同的数据集处理方式, 以更好地比对最终的复现结果。

除了使用不同的深度学习框架实现 RIB 之外, 本文还对 RIB 进行了一些改进。

第一个改进方法是在 RIB 中使用多个 GRU 的输出值的平均替代单个 GRU 的输出值。由于单个 GRU 在训练时可能会陷入局部最优解, 从而导致性能下降。通过使用多个独立的 GRU 并对它们的输出进行平均化, 可以降低这种风险, 从而提高模型的稳定性。同时, 通过集成多个模型的预测结果, 可以降低模型的预测误差, 从而提高模型的准确率。但需要注意的是, 使用多个 GRU 会增加计算和存储成本, 因此, 在本文的实验中, 采用了两个 GRU 的平均输出值。

第二个改进方法是在第一个改进方法的基础上, 使用多行为 GRU 单元 (Multi-behavior GRU units, MGRU) 替代传统 GRU, 即使用 2 个 MGRU 的平均输出值来替代 RIB 中 GRU 的输出值。MGRU 是 IARS 提出的一种方法, 在原论文中也没有提供源代码。MGRU 在 GRU 中引入了一个区分不同行为重要性的门, 并在不同行为之间共享同一组参数, 同时在输出隐藏状态前进行归一化。MGRU 的重置门 r_t 、更新门 z_t 和隐藏状态 h_t 的计算过程如下所示:

$$r_t = \sigma(W_r e_t + b_r) \quad (11)$$

$$z_t = \sigma(W_z e_t + b_z) \quad (12)$$

$$\tilde{h}_t = \tanh(W_h [r_t * h_{t-1}, V_{i_u}^t]) \quad (13)$$

$$\hat{h}_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (14)$$

$$h_t = \hat{h}_t / \|\hat{h}_t\| \quad (15)$$

其中 $W_{(\cdot)}$ and $b_{(\cdot)}$ 是待学习的模型参数, σ 和 \tanh 是激活函数。

4.2 实验

本实验基于 Windows 10 的操作系统和硬件平台，使用 Python 3.8.8 和 PyTorch 1.10.2 进行实现。以下将介绍所使用的数据集及其预处理步骤、评估指标、参数设置和本实验的创新点。

4.2.1 数据集及预处理

在本实验中，MovieLens 1M（ML1M，<https://grouplens.org/datasets/movielens/1m/>）是我们所使用的数据集。该数据集是一个电影评分数据集，包含了 6040 个用户对电影的将近 100 万个评分，其中评分的范围为 1-5。和前人研究^[42]一致，我们把评分为 5 的交互行为作为购买行为，1-4 分的交互行为作为点击。

在对 LM1M 数据集的预处理中，我们除去了后续出现的重复的（用户，物品，行为）三元组，然后删除购买次数少于 5 次的物品和用户。最后，我们将每个用户最后两次购买数据分别作为验证集数据和测试集数据，而剩余其他所有行为数据都作为训练集数据。在预处理后的数据集 ML1M 如表 1 所示。

表 1: ML1M 经预处理后的统计信息

数据集	用户	物品	点击	购买
ML1M	5645	2357	628892	223305

4.2.2 评估指标

本实验采用的评估推荐性能的指标是 HR@10 和 NDCG@10。其中，HR@10（Hit Rate at 10）是指在推荐列表中，用户实际发生交互（如点击、购买等）的物品中，排名前 10 的物品所占比例。而 NDCG@10（Normalized Discounted Cumulative Gain at 10）是指在推荐列表中，用户实际发生交互的物品按照实际发生交互的顺序进行加权，然后计算出推荐列表中前 10 个物品的加权累积分数，最后与理想情况下的加权累积分数进行比较得到一个标准化的分数。NDCG@10 和 HR@10 的取值范围都为 0 到 1，取值越高代表推荐系统的性能越好。

4.2.3 参数设置

与前人研究^[41]一致，我们将序列长度设置为 50，batch size 设置为 128，embedding size 设置为 64，dropout 率设置为 0.2，以及学习率设置为 0.001。

4.3 创新点

本实验的创新点在于，在基于 PyTorch 深度学习框架下复现了 RIB 中对多行为序列建模的部分，并对 RIB 的模型做了两点改进，分别为使用 2 个 GRU 的平均输出值替代了原来的单个 GRU 的输出值，以及使用了 2 个 MGRU 的平均输出值替代了原来的单个 GRU 的输出值。

5 实验结果分析

我们实现的三个模型在 ML1M 数据集上的实验结果如表 2 所示。在表 2 中，RIB (TensorFlow) 是指前人实现的 RIB^[41]，RIB (PyTorch) 是指本实验基于 PyTorch 实现的 RIB，RIB (Avg-GRU) 和 RIB (Avg-MGRU) 分别是改进的采用 2 个 GRU 和 MGRU 的平均输出值的 RIB。RIB (PyTorch) 和 RIB (TensorFlow) 的 HR@10 和 NDCG@10 结果相差 0.01 以内，表明了本实验使用 PyTorch 实现 RIB 的准确性，即复现基本成功。而在 HR@10 和 NDCG@10 上，RIB (Avg-GRU) 和 RIB (Avg-MGRU) 的值都比 RIB (PyTorch) 的值大，说明改进的两种方法都有一定的提升，其中 RIB (Avg-MGRU) 的提升更大，从而可以实现更优的推荐性能。

表 2: 在数据集 ML1M 上的实验结果

模型	HR@10	NDCG@10
RIB (TensorFlow)	0.1302	0.0646
RIB (PyTorch)	0.1371	0.0660
RIB (Avg-GRU)	0.1378	0.0667
RIB (Avg-MGRU)	0.1398	0.0676

6 总结与展望

本文首先对多行为序列推荐（MBSR）的背景和相关工作进行了介绍，然后对复现的 MBSR 的经典工作 RIB 进行了介绍，其中主要介绍了 RIB 对行为的异构性和序列性的建模方面，包含输入层、GRU 层、注意力层和输出层。接着，本文基于 PyTorch 实现了该方法，并提出了两点改进之处，即分别使用 2 个 GRU 和 2 个 MGRU 的输出值的平均值来代替使用单个 GRU 得到的输出值。最后，本文通过在 ML1M 数据集上进行了实验，实验结果表明了复现的准确性，以及改进方法的有效性。

然而，本文改进的两种方法的改进效果并不大，这意味着只是简单对 GRU 层进行改进不足以获得较好的推荐效果。因此，我们仍然需要在 RIB 的基础上做出更多改进，如加入物品类别信息等。由于 GRU 效率较低，以及存在梯度爆炸或梯度消失等问题，因此，将基于 GRU 的 RIB 与 GNN 等深层神经网络结合使用是一个可进一步研究的方向。

参考文献

[1] RICCI F, ROKACH L, SHAPIRA B. Recommender Systems Handbook (Second Edition)[M]. Springer, 2015.

[2] GAI L, LEI L. Dual Collaborative Topic Modeling from Implicit Feedbacks[C]//SPAC: Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics. 2014: 395-404.

[3] LI G, OU W. Pairwise Probabilistic Matrix Factorization for Implicit Feedback Collaborative Filtering [J]. Neurocomputing, 2016, 204: 17-25.

[4] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian Personalized Ranking from Implicit Feedback[C]//UAI'09: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. 2009: 452-461.

- [5] PAN W, ZHONG H, XU C, et al. Adaptive Bayesian Personalized Ranking for Heterogeneous Implicit Feedbacks[J]. Knowledge-Based Systems, 2015, 73: 173-180.
- [6] PAN W, LIU M, MING Z. Transfer Learning for Heterogeneous One-class Collaborative Filtering[J]. IEEE Intelligent Systems, 2016, 31(4): 43-49.
- [7] PAN W, YANG Q, CAI W, et al. Transfer to Rank for Heterogeneous One-class Collaborative Filtering [J]. ACM Transactions on Information Systems, 2019, 37(1): 1-20.
- [8] RENDLE S, FREUDENTHALER C, SCHMIDT-THIEME L. Factorizing Personalized Markov Chains for Next-basket Recommendation[C]// WWW'10: Proceedings of the 19th International Conference on World Wide Web. 2010: 811-820.
- [9] ZENG Z, LIN J, LI L, et al. Next-item Recommendation via Collaborative Filtering with Bidirectional Item Similarity[J]. ACM Transactions on Information Systems, 2019, 38(1): 1-22.
- [10] TANG J, WANG K. Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding[C]// WSDM'18: Proceedings of the 11th ACM International Conference on Web Search and Data Mining. 2018: 565-573.
- [11] LIU Q, WU S, WANG L. Multi-Behavioral Sequential Prediction with Recurrent Log-Bilinear Model [J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(6): 1254-1267.
- [12] ZHOU M, DING Z, TANG J, et al. Micro Behaviors: A New Perspective in E-commerce Recommender Systems[C]// WSDM'18: Proceedings of the 11th ACM International Conference on Web Search and Data Mining. 2018: 727-735.
- [13] LI Z, ZHAO H, LIU Q, et al. Learning from History and Present: Next-item Recommendation via Discriminatively Exploiting User Behaviors[C]// KDD'18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2018: 1734-1743.
- [14] SMITH B, LINDEN G. Two Decades of Recommender Systems at Amazon.com[J]. IEEE Internet Computing, 2017, 21(3): 12-18.
- [15] WANG J, HUANG P, ZHAO H, et al. Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2018: 839-848.
- [16] COVINGTON P, ADAMS J, SARGIN E. Deep Neural Networks for Youtube Recommendations[C]// Proceedings of the 10th ACM Conference on Recommender Systems. 2016: 191-198.
- [17] GOMEZ-URIBE C A, HUNT N. The Netflix Recommender System: Algorithms, Business Value, and Innovation[J]. ACM Transactions on Management Information Systems (TMIS), 2015, 6(4): 1-19.
- [18] LIU J, DOLAN P, PEDERSEN E R. Personalized News Recommendation Based on Click Behavior[C]// Proceedings of the 15th International Conference on Intelligent User Interfaces. 2010: 31-40.

- [19] WU F, QIAO Y, CHEN J H, et al. MIND: A Large-scale Dataset for News Recommendation[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 3597-3606.
- [20] ZHAN Z, HE M, PAN W, et al. TransRec++: Translation-based Sequential Recommendation with Heterogeneous Feedback[J]. Frontiers of Computer Science, 2022, 16(2): 162615.
- [21] MIKOLOV T, KARAFIÁT M, BURGET L, et al. Recurrent Neural Network Based Language Model. [C]// Interspeech: vol. 2: 3. 2010: 1045-1048.
- [22] XU Y, ZHU Y, YU J. Modeling Multiple Coexisting Category-Level Intentions for Next Item Recommendation[J]. ACM Transactions on Information Systems, 2021, 39(3): 23:1-23:24.
- [23] HOCHREITER S, SCHMIDHUBER J. Long Short-term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [24] CHUNG J, GULCEHRE C, CHO K, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. arXiv preprint arXiv:1412.3555, 2014.
- [25] HAN J, MA Y, MEI Q, et al. DeepRec: On-device Deep Learning for Privacy-Preserving Sequential Recommendation in Mobile Commerce[C]// WWW'21: Proceedings of The Web Conference 2021. 2021: 900-911.
- [26] SCARSELLI F, GORI M, TSOI A C, et al. The Graph Neural Network Model[J]. IEEE Transactions on Neural Networks, 2008, 20(1): 61-80.
- [27] WU Z, PAN S, CHEN F, et al. A Comprehensive Survey on Graph Neural Networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(1): 4-24.
- [28] WANG W, ZHANG W, LIU S, et al. Beyond Clicks: Modeling Multi-Relational Item Graph for Session-based Target Behavior Prediction[C]// WWW'20: Proceedings of the Web Conference 2020. 2020: 3056-3062.
- [29] CHEN W, HE M, NI Y, et al. Global and Personalized Graphs for Heterogeneous Sequential Recommendation by Learning Behavior Transitions and User Intentions[C]// RecSys'22: Proceedings of the 16th ACM Conference on Recommender Systems. 2022: 268-277.
- [30] LUO J, HE M, PAN W, et al. BGNN: Behavior-aware Graph Neural Network for Heterogeneous Session-based Recommendation[J]. Frontiers of Computer Science, 2022.
- [31] LIANG Y, SONG Q, ZHAO Z, et al. BA-GNN: Behavior-aware Graph Neural Network for Session-based Recommendation[J]. Frontiers of Computer Science, 2022, 0(0): 1-16.
- [32] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need[J]. Advances in Neural Information Processing Systems, 2017, 30.
- [33] XIE R, LING C, WANG Y, et al. Deep Feedback Network for Recommendation[C]// IJCAI'20: Proceedings of the 29th International Joint Conference on Artificial Intelligence. 2020: 2519-2525.

- [34] BIAN Z, ZHOU S, FU H, et al. Denoising User-aware Memory Network for Recommendation[C]// RecSys'21: Proceedings of the 15th ACM Conference on Recommender Systems. 2021: 400-410.
- [35] WU C, WU F, QI T, et al. FeedRec: News Feed Recommendation with Various User Feedbacks[C]// WWW'22: Proceedings of the ACM Web Conference 2022. 2022: 2088-2097.
- [36] HE M, LIN J, LUO J, et al. FLAG: A Feedback-aware Local and Global Model for Heterogeneous Sequential Recommendation[J]. ACM Transactions on Intelligent Systems and Technology, 2022.
- [37] MENG W, YANG D, XIAO Y. Incorporating User Micro-behaviors and Item Knowledge into Multi-task Learning for Session-based Recommendation[C]//SIGIR'20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020: 1091-1100.
- [38] WU S, TANG Y, ZHU Y, et al. Session-based Recommendation with Graph Neural Networks[C]// Proceedings of the AAAI Conference on Artificial Intelligence: vol. 33: 01. 2019: 346-353.
- [39] YANG Y, HUANG C, XIA L, et al. Multi-Behavior Hypergraph-Enhanced Transformer for Sequential Recommendation[C]//KDD'22: Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2022: 2263-2274.
- [40] XIA L, HUANG C, XU Y, et al. Knowledge-Enhanced Hierarchical Graph Transformer Network for Multi-Behavior Recommendation[C]//AAAI'21: Proceedings of the 35th AAAI Conference on Artificial Intelligence. 2021: 4486-4493.
- [41] HE M, PAN W, MING Z. BAR: Behavior-aware Recommendation for Sequential Heterogeneous One-class Collaborative Filtering[J]. Information Sciences, 2022, 608: 881-899.
- [42] PAN W, YANG Q, CAI W, et al. Transfer to Rank for Heterogeneous One-class Collaborative Filtering [J]. ACM Transactions on Information Systems, 2019, 37(1): 1-20.