



研究与开发

数据增强的多模式时间感知序列推荐

李家乐, 王瑞琴, 于洋

(湖州师范学院信息工程学院, 浙江 湖州 313000)

摘要: 序列推荐系统以包含了显式信息的用户交互序列作为上下文, 推测用户的下一个可能动作。其中, 时间感知序列推荐挖掘了序列中的时间信息, 并考虑了时间信息对用户决策的影响。但是现有的时间感知序列推荐模型只运用到了原始时间信息, 原本的序列中还有很多额外信息没有被充分挖掘, 如用户评分、项目属性、项目流行度以及项目的标题和评论等文本信息。因此, 提出了DMTiSASRec模型, 它既可以以更高效的方式挖掘时间信息中的相关秩序, 还利用对比学习、多模态等技术对不同的额外信息进行挖掘。在5个不同领域、不同规模的公开数据集的实验数据表明, DMTiSASRec比现有模型更有效。

关键词: 序列推荐; 时间感知; 多模态; 对比学习

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2024234

Multi-pattern time-aware sequential recommendation with data augmentation

LI Jiale, WANG Ruiqin, YU Yang

School of Information Engineering, Huzhou University, Huzhou 313000, China

Abstract: In sequential recommendation systems, explicit user interaction sequences are used as context to infer the user's next possible action. Time-aware sequential recommendation models explore the temporal information within the sequence and consider the impact of time on user decisions. However, existing time-aware sequential recommendation models only utilize raw temporal information, while many additional pieces of information in the original sequence are not fully exploited, such as user ratings, item attributes, item popularity, and textual information like item titles and reviews. Therefore, the DMTiSASRec model was proposed, which not only efficiently extracted relevant orders beyond temporal information but also leveraged techniques like contrastive learning and multi-modal methods to mine different types of additional information. Experiments on five publicly available datasets across different domains and scales show that DMTiSASRec outperforms existing models in terms of effectiveness.

Key words: sequential recommendation, time-aware, multi-modal, contrastive learning

收稿日期: 2024-07-18; 修回日期: 2024-10-13

通信作者: 王瑞琴, wrq@zjhu.edu.cn

基金项目: 国家自然科学基金面上项目 (No.62277016)

Foundation Item: The National Natural Science Foundation of China (No.62277016)

0 引言

序列推荐技术已经应用于现实世界中的各种领域。它可以有效地解决信息过载,给予用户更好的体验。序列推荐的核心是将包含显示信息的用户序列作为输入,并且输出下一项的预测值。目前为止,序列推荐已经出现过许多优秀模型,从马尔可夫链^[1]到递归神经网络^[2],已成功用于捕获短期项目转换以进行推荐,但是在很大程度上忽略了历史交互中嵌入的其他有用的时间信息。之后受机器翻译(Transformer)^[3]的启发,SASRec^[4]模型将自注意力机制应用于序列推荐问题。图神经网络^[5-7]也利用了自注意力机制并且被运用到序列推荐中。TiSASRec^[8]进一步考虑了交互项目之间的时间间隔,认为更接近当前时间的交互对预测下一次交互有更大影响。整体时间模式(holistic temporal pattern, HTP)^[9]更是充分挖掘了时间戳中的信息,并且提出了一种新的基于整体时间模式的神经网络。除了对于位置和时间信息的挖掘,近年来,对比学习在机器学习和深度学习领域的应用越来越多。其核心思想是通过学习数据表示,使得相似数据点在表示空间中靠近,而不相似数据点远离,从而提高模型泛化能力。CLF4SRec^[10]采用多任务学习框架,结合对比学习和序列建模来优化用户表示。随着NLP和CV领域的发展,各种研究已经证明了从多模态中学习的有效性,越来越多的多模态技术已经被运用到了推荐系统中,例如,PMMRec^[11]仅依赖于项目的多模态内容,并学习足够通用的转换模式用以跨域和平台传输。前人对序列推荐的优化做了两个方向的努力。第一是对序列中已经包含的信息进行更充分的挖掘。如上文提到的递归神经网络、TiSASRec、HTP等体现了对序列中位置关系、时间信息的挖掘方式的发展以及效果的提高。第二是进行数据增强,以缓解可能存在的数据稀疏问题,并且在相同样本数量下提高

模型的精准度。如上文提到的PMMRec,输入模型中的数据不单单是原始序列,还有文本、图像等。而对比学习是一种特殊的数据增强手段,它额外提供了对象与对象之间相似性或者差异性的信息。基于此,本文提出了DMTiSASRec(multi-pattern time-aware sequential recommendation with data augmentation)模型。在数据增强方面,引入了预训练语言模型、用户维度的对比学习模块,并且引入了项目流行度指标。在改进对信息挖掘的方法方面,提出了一种对时间信息挖掘的新模式,以及对评分信息新的处理方法。实验结果表明,本文提出的模型在不同领域的数据集中都表现良好。验证了数据增强对于时间感知的序列推荐模型有增强效果。本文工作如下。

(1) 引入RoBERTa预训练语言模型,通过输入项目的标题和评论生成隐向量进行数据增强。通过对比学习,计算出用户之间差异度进行数据增强。还引入了项目流行度指标进行数据增强。

(2) 针对评分信息,本文设计了多通道嵌入层来改进对评分信息的挖掘。

(3) 针对时间信息,本文提出一种新的时间信息挖掘模式,不仅考虑了序列位置信息,还将时间信息分为相对时间模式和绝对时间模式。

1 相关工作

1.1 时间感知的序列推荐

考虑用户行为之间的时间间隔来学习用户兴趣爱好和项目特征已经出现在不久前的研究中。有学者通过为Time-LSTM^[12]配备专门设计的时间门来对用户交互之间时间间隔进行模拟,力求从短期、长期两个维度都可以抓取用户偏好特征,提高下一项的推荐准确率。TGSRec^[13]是一种近期的时间间隔感知顺序推荐方法,它通过协作注意力网络显式地对项目之间的协作信号和时间间隔进行建模,其中时间跨度表示为相应编码



时间嵌入的内核值。SASRec考虑了序列间位置信息,给访问序列相同的用户推荐相同的下一个项目,忽略序列的时间信息,但是项目被推荐的概率会随着时间的变化而变化,或者说项目综合评分会随着时间的推移而产生变化。TiSASRec基于SASRec引入了相对时间的概念,表明时间戳顺序更接近的项目对预测下一个项目有更大的影响。HTP将交互时间戳细分为交互之间的相对时间间隔和交互的绝对时间,然后分为3个种类,并且通过特定模块进行集成,形成最后的推测值。除此之外,为了集中在从用户侧探索时间间隔,研究用户连续偏好。TimelyRec^[14]通过处理分层时间粒度来建模用户行为中的异构时间模式,能够消除用户在某些时间间隔内可能不感兴趣的项目的影响。

1.2 序列推荐中的数据增强

在序列推荐中,数据增强是提升模型泛化能力和处理稀疏数据的重要手段,尤其是在用户交互行为序列较少或不完整的情况下。其中有不同的方法、策略来达到数据增强的目的。

HGNRec^[15]使用了序列截断策略,进行数据增强,并结合层次化网络来捕捉用户的长期与短期兴趣来提升模型性能。有学者通过提出一种基于行列式点过程(determinantal point process, DPP)的多样性采样方法^[16],实现序列推荐中的多样化,用以数据增强。还有学者提出了一种基于伪先验(Pseudo-prior Items)技术^[17]的序列推荐方法,通过对用户的历史交互行为进行序列扰动,生成额外的伪行为数据,帮助模型更好地捕捉用户兴趣变化。MMRec^[18]结合用户的交互历史和多模态信息来进行推荐引入了多模态融合机制,有效利用用户在文本、图像等多个模态下的兴趣点,提升推荐质量。特别是在用户交互数据稀疏或单一模态信息不足的情况下该框架通过跨模态信息融合,捕捉用户对不同模态的偏好,增强推荐结果的准确性。

1.3 对比学习

对比学习的核心在于通过对比相似和不相似样本来学习有用的数据表示。依赖于数据本身的结构来学习有意义的特征表示。基本思想是通过拉近相似样本的表示和拉远不相似样本的表示来训练模型。相似样本可以通过数据增强方法生成的不同视角的同一数据,不相似样本则是来自不同实例的数据。对比学习如今已经被运用到各种各样的领域^[19],并且确实取得了极大的进步。最后,Zhou等^[20]创造性地将推荐系统和对比学习模块进行结合,产生的模型有效减少了流行偏见并探究了对比学习目标与更具有难度的去偏见目标(如逆倾向得分)有理论相关性。不同于传统对比学习,Xie等^[21]提出了一种自监督对比学习框架,通过对用户行为序列进行不同形式的改变,如随机删除、替换等,生成正例和负例。这些改变的序列通过对比学习框架进行训练,增强模型对不同行为模式的鲁棒性。

2 模型设计

2.1 问题表述

令 U 、 I 、 T 、 G 分别为用户全域、项目全域和时间戳全域以及额外信息全域。在序列推荐的集合中,一个用户 $u \in U$ 与交互项目 S_u 相关联。 $S_u = [(s_1, t_1, g_1), \dots, (s_{|S_u|}, t_{|S_u|}, g_{|S_u|})]$,其中 $s_i \in I$, $t_i \in T$, $g_i \in G$ 并且 $t_i < t_{i+1}$ 。每个 (s_i, t_i, g_i) 对表示了用户 u 和项目 s_i 在时间戳为 t_i 时产生了交互并且附带了额外信息 g_i 。其中 $g_i = (\text{pos}_i, \text{pop}_i, \text{rate}_i, \text{title}_i, \text{rew}_i)$ 分别代表位置信息、项目流行度、评分信息以及经过预训练模型学习出来的项目标题和用户评论的隐向量表示。模型输入是 $(S_1^u, S_2^u, \dots, S_{|S^u|-1}^u)$,希望输出是每个时间段的下一项 $(S_2^u, S_3^u, \dots, S_{|S^u|}^u)$ 。其中项目流行度的计算是以全局视角,以某项目出现在某用户序列中的次数除以用户总数得到的

数值, 数值保留3位小数。并且在模型中设定了一个最大序列长度 L , 若一个序列长度大于预定义值 L , 就将其截断并选择 L 个最近项。否则, 将序列填充为长度 L 。最后, 在模型内设定一个常数 k 作为放缩后最大的时间间隔。

2.2 方法提出

尽管现在基于时间感知的序列推荐模型已经取得了很大进展, 但是仍有很大改进的空间。现有模型对时间信息没有进行模式区分, 没有更充分地对时间信息进行特征提取。对一些特定信息, 缺少高效的特征提取方法。并且缺少数据增强的方法, 可能导致模型在小型并且稀疏的数据集上表现不佳。

在考虑了上述几点后, 本文提出了一种对比学习与多模态增强的时间感知序列推荐模型。DMTiSASRec整体模型框架如图1所示。在数据预处理阶段, 模型会自动生成时间信息、用户评分信息、项目流行度、经过预训练语言模型产生项目的隐向量等。其中嵌入层接收除文本信息和评分信息外的所有信息, 并且集中生成嵌入表示。在时间感知层中, 模型会接收由嵌入层传来的关于时间与位置信息。其中对于时间信息, 将

其分成相对时间模式以及绝对时间模式, 并且结合位置信息进行综合考量, 最终输出两种模式的时间特征。多通道嵌入层则是对评分信息进行特殊处理。信息处理层则是对评分信息、项目流行度的嵌入表示以及隐向量进行综合处理。预测层利用最终得出的下一项的特征向量与项目嵌入进行点积得出预测得分。对比学习模块通过构造对和负对, 使不同用户之间差异更加明显。为了取得更好的效果, 模型采用二进制交叉熵损失以及对比学习模块用到的BPR损失作为模型目标函数。本模型与预测任务相结合, 实现较为精准的预测。

2.3 嵌入层

首先将项目和时间戳映射到两个不同的低维潜在空间。其中 $E \in \mathbf{R}^{N \times d}$ 表示由项目嵌入层产生项目的嵌入矩阵, 并且 $e_i = E_i$, e_i 为项目 i 的嵌入向量, 其中 N 是项目数量, d 是嵌入层大小。为了挖掘序列位置信息的作用, 本模型设置了一个可学习的位置信息嵌入层:

$$e_i^s = e_i + e_i^p \quad (1)$$

其中, $e_i^p \in \mathbf{R}^d$ 是一个序列中的位置信息的嵌入表示。 $M^p \in \mathbf{R}^{L \times d}$ 表示位置信息嵌入矩阵。并且设

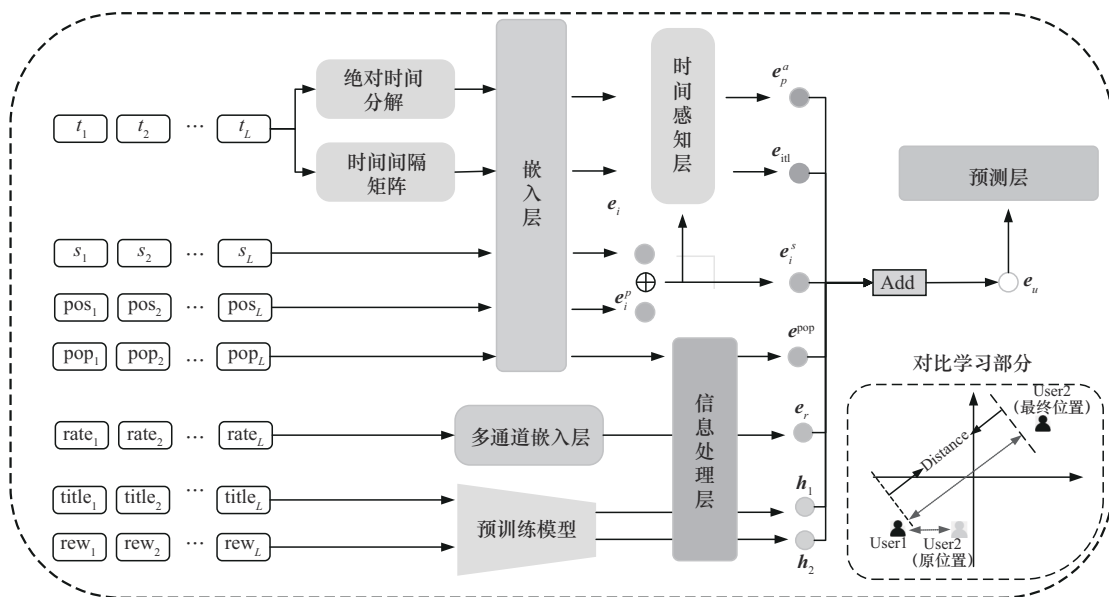


图1 DMtiSASRec整体模型框架



定相对时间间隔的嵌入矩阵 $\mathbf{M}_K^R \in \mathbf{R}^{k \times d}$, 用于生成键和值的嵌入表示, 最终得到嵌入矩阵 $\mathbf{E}_K^R \in \mathbf{R}^{N \times N \times d}$ 和 $\mathbf{E}_V^R \in \mathbf{R}^{N \times N \times d}$ 。

为了生成时间戳的信息嵌入。在数据预处理阶段, 将 t_i 分解成年、月、日、小时级别维度的绝对时间, $t_i = (t_i^y, t_i^m, t_i^d, t_i^h)$, 其中绝对时间戳嵌入表示为:

$$\mathbf{e}_{t_i} = \mathbf{e}_{t_i^y} + \mathbf{e}_{t_i^m} + \mathbf{e}_{t_i^d} + \mathbf{e}_{t_i^h} \quad (2)$$

其中, $\mathbf{e}_{t_i^y}$ 、 $\mathbf{e}_{t_i^m}$ 、 $\mathbf{e}_{t_i^d}$ 、 $\mathbf{e}_{t_i^h}$ 分别为以年、月、日、小时级别的时间嵌入表示, 它们分别由 $\mathbf{M}^y$ 、 $\mathbf{M}^m$ 、 $\mathbf{M}^d$ 、 $\mathbf{M}^h$ 嵌入层矩阵得到。与上述方法相同, 项目流行度的嵌入表示为 $\mathbf{e}_{\text{pop}}$, 其中, 对于用户评分的嵌入表示 $\mathbf{e}_{\text{rate}}$ 方法与上述方法不同。有了项目嵌入、时间戳嵌入以及其他信息嵌入, 可以将用户 u 的序列表示为 $\mathbf{S}_u = [(\mathbf{e}_i^s, \mathbf{e}_{t_i}, \mathbf{e}_{\text{pop}}, \mathbf{e}_{\text{rate}}, \mathbf{h}_1, \mathbf{h}_2)]$, 其中 $\mathbf{h}_1$ 、 $\mathbf{h}_2$ 表示经过预训练模型学习出来的项目标题和用户评论的隐向量表示。

2.4 时间感知层

该层考虑了相对时间以及不同维度的绝对时间对于用户兴趣以及物品特征建模的影响。直觉上来说, 相对时间 (某种意义上的时间间隔) 可以在某种意义上反映项目之间的关系, 例如, 如果一个用户在某时间看了系列电影的第一部, 那么他很有可能在短时间内看系列电影的第二部, 这说明了两部电影可能有一定相关性, 以及用户可能偏好这类电影。而绝对时间考虑了某种项目跟某个时间点用户交互行为的敏感相关性。例如, 西瓜更容易在夏天被购买, 羽绒服更容易在冬天被购买。

对于相对时间模式, 在数据预处理阶段构造了一个时间间隔矩阵, $\mathbf{M}^u \in \mathbf{R}^{L \times L}$, 对于给定属于 u 的固定长度的时间序列 $t = (t_1, t_2, \dots, t_L)$, 两个项目 i 和 j 的时间间隔表示为 $|t_i - t_j|$ 并且用户 u 的时间间隔集合 (除 0 之外) 表示为 R_u 。将用户的

最小时间间隔表示为 $r_{\min}^u = \min(R_u)$ 。比例缩放后的时间间隔表示为 $r_{ij}^u = \frac{|t_i - t_j|}{r_{\min}^u}$ 。考虑有些 r_{ij}^u 会非常大, 不利于计算, 因此设定了 k 为 r_{ij}^u 的最大值, 并将超过此最大值的 r_{ij}^u 替换为 k 。其中 r_{ij}^u 就是 $\mathbf{M}^u$ 的元素。

给定输入序列 $E = (e_1, e_2, \dots, e_N)$, 通过自注意力机制生成输出序列 $Z = (e_{\text{itl},1}, e_{\text{itl},2}, \dots, e_{\text{itl},N})$

$$e_{\text{itl},i} = \text{softmax} \left(\frac{e_i W^Q (e_j W^K + r_{ij}^k + e_j^p)^T}{\sqrt{d}} \right) e_j W^V, ij \quad (3)$$

其中, W^Q 、 W^K 、 $W^V \in \mathbf{R}^{d \times d}$ 分别是查询和键以及值的输入投影。式 (3) 考虑了项目的位置信息、相对时间间隔信息, 但是本质上还是线性组合, 所以在自注意力计算后, 引入了带有非线性激活函数的前馈层, 使模型拥有更强的拟合能力。

$$\text{FFN}(e_{\text{itl},i}) = \max(0, e_{\text{itl},i} W_1 + b_1) W_2 + b_2 \quad (4)$$

其中, $W_1 \in \mathbf{R}^{d \times d}$ 和 $W_2 \in \mathbf{R}^{d \times d}$ 是权重矩阵, $b_1 \in \mathbf{R}^d$ 和 $b_2 \in \mathbf{R}^d$ 是偏置向量。

为了解决过拟合、梯度消失问题, 本模型采用了层归一化、残差连接和 dropout 正则化技术。

$$e_{\text{itl},i} = e_{\text{itl},i} + \text{Dropout}(\text{FFN}(\text{LayerNorm}(e_{\text{itl},i}))) \quad (5)$$

对于不同维度的绝对时间, 以全局为视角, 创造一个项目的绝对时间模式, 令 T_i 是与来自所有用户的项目 i 相关联的交互时间戳的集合。项目 i 的绝对时间模式, 可以用式 (6) 表示为:

$$e_i^t = \sum_{t_j \in T_i} \alpha_{t_j} e_{t_j} \quad (6)$$

其中, $\alpha_{t_j} = \frac{n_{i,t_j}}{n_i}$, n_{i,t_j} 代表了所有用户在时间戳 t_j 时刻与项目 i 的交互数量, 而 n_i 代表了所有用户在任意时刻与项目 i 的交互数量。 α_{t_j} 越大表示项目 i 与某个时间节点之间的关系越密切, 例如, 圣诞树总是在圣诞节前后被购买的最多。而 e_i^t 经过

比例缩放, 汇总了所有绝对时间信息的综合绝对时间表示。

最后再通过缩放点积注意力机制, 产生一个假定的、某个用户很可能感兴趣的下一个选项的预测。这个预测表示为 $e_{p,i}^a \in \mathbf{R}^d$ 。其中 $e_{p,i}^a$ 的计算式为:

$$e_{p,i}^a = \text{Att}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (7)$$

$$e_{p,i}^a = \text{Att}(e_{t_{i+1}}, e_i^t, e_i^s) = \text{softmax}\left(\frac{e_{t_{i+1}}(e_i^t)^T}{\sqrt{d}}\right)e_i^s \quad (8)$$

其中, $\text{Att}(\cdot)$ 表示缩放点积注意力计算。 $e_{t_{i+1}}$ 表示 item $i+1$ 分解后的时间戳嵌入, 如式 (6) 所示, e_i^t 表示 item i 的绝对时间模式嵌入表示。如式 (2) 所示, e_i^s 为融合了位置信息与项目嵌入的综合表示, 如式 (1) 所示。其中绝对时间模式处理过程如图 2 所示。

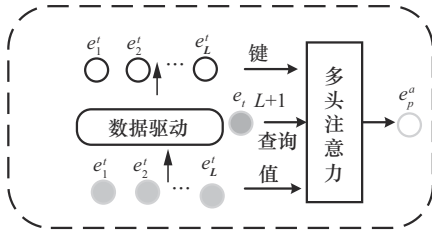


图2 绝对时间模式处理过程

2.5 多通道嵌入层

现有模型都没有很好地利用评分信息。对于评分信息, 一个用户对某个项目的评分, 包含了用户对它的喜好程度。考虑评分信息对模型的影响, 可以更好地对用户进行建模。但是有时候, 用户对项目的评分可能会包含一些偏见。例如, 某个用户很喜欢恐怖电影, 那么他很可能对恐怖电影打出高分, 但是他不喜欢爱情电影, 那么他很可能对爱情电影打出低分, 即使某些爱情电影很优秀。针对这个问题, 本模型提出了一种多通道嵌入层机制。评分信息通过衰减函数, 进入嵌入层, 得到最终结果。

假定整个评分数据集的最低分是 al 最高分是 ah。

$$\text{rate}_n = \text{decay}^n(\text{rate}) \quad (9)$$

$$\text{decay}(x) = \begin{cases} x-1, & x > \text{al} \\ \text{al}, & x \leq \text{al} \end{cases} \quad (10)$$

其中, $\text{decay}^n(\cdot)$ 表示函数 $\text{decay}(\cdot)$ 在原始输入上的 n 次迭代应用, 每次调用的输入为上一次调用的输出。这些迭代可以有效地削减用户在某些情况下的偏见。 $n \in \mathbf{Z}^+$ 且 $n \in [1, \text{ah} - \text{al}]$, Rating 为评分信息。

$$e_{r,0} = \text{embedding}(\text{rate}) \quad (11)$$

$$e_{r,i} = \text{embedding}(\text{rate}_i), i = 1, 2, \dots, \text{ah} - \text{al} \quad (12)$$

$$e_r = \text{MLP}(\text{concat}(e_{r,0}, e_{r,1}, e_{r,2}, \dots, e_{r,i})) \quad (13)$$

其中, $\text{embedding}(\cdot)$ 表示经过嵌入层, $\text{concat}(\cdot)$ 表示进行数据拼接。MLP($\cdot$) 表示经过全连接层竖立, $e_{r,0}$ 表示原始 Rating 数据的嵌入表示, $e_{r,i}$ 表示经过 decay 函数的 Rating 嵌入表示。多通道嵌入层机制如图 3 所示。

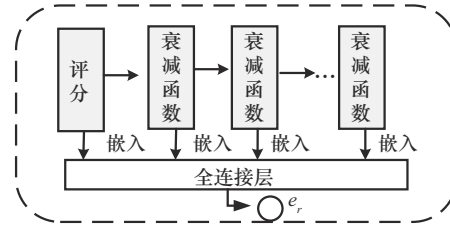


图3 多通道嵌入层机制

2.6 信息处理层

除了对于评分信息嵌入表示的处理。对于文本数据, 在某些短序列数据集中, 可以进行某种意义的额外信息增加, 提高模型准确率。很多研究^[18]都已经证明多模态技术运用到序列推荐中是有效果的。本模型采用了 RoBERTa 预训练语言模型, 学习数据集中, 项目名称和用户对项目评论进行某种意义上的数据增强。

$$h_1 = \text{RoBERTa}(\text{title}) \quad (14)$$

$$h_2 = \text{RoBERTa}(\text{Review}) \quad (15)$$



其中, RoBERTa($\cdot$)表示使用预训练语言模型的编码器。

信息处理层除了接受文本信息产生的隐向量,还处理项目流行度信息。对于项目流行度,本模型以全局为视角,计算出某个项目出现的频率。经过嵌入层得到项目流行度的嵌入表示 e_{pop} 。引入项目流行度可以弱化某些热门项目对用户决策产生的影响。

$$popularity_{s_i} = \text{round}\left(\frac{p_{s_i}}{|U|}\right) \quad (16)$$

其中, p_{s_i} 表示项目 s_i 在整个全局的序列数据中,与多少用户有过交互。 $|U|$ 表示全局用户个数。为了避免小数点后面位数过大导致嵌入层过大,采用 $\text{round}(\cdot)$ 函数保留小数点后4位有效数字。

2.7 对比学习模块

为了从用户交互序列中获得个性化的用户偏好向量,即使交互序列相似的用户向量也可以得到一定程度的区分并且得到某种意义的数据增强。本模块采用了一种对比学习机制。该机制让相似的建模结果尽可能相似,而不同的建模结果差距尽可能大。对于一个用户序列,让其通过两次自注意力层获得两个编码,即两个建模结果,组成一个正对。随机从其他用户中抽取一个序列,经过自注意力机制层,得到一个编码,与正对中的一个编码组成一对负对。对比损失函数定义如下。

$$L_{con} = - \sum_{u=1}^{|D|} \log \frac{e^{\frac{\sin(z_u^1, z_u^2)}{\tau}}}{e^{\frac{\sin(z_u^1, z_u^2)}{\tau}} + e^{\frac{\sin(z_u^1, z_n)}{\tau}}} \quad (17)$$

其中, z_u^1 和 z_u^2 表示对当前用户序列经过两次建模得到两个编码, z_n 表示随机挑选的用户序列经过建模后得到的编码, D 是训练数据集, $\sin(\cdot)$ 表示两个向量之间余弦相似度, τ 是温度参数,它控制分配给每个样本权重的尖锐程度。对比学习模块流程如图4所示。

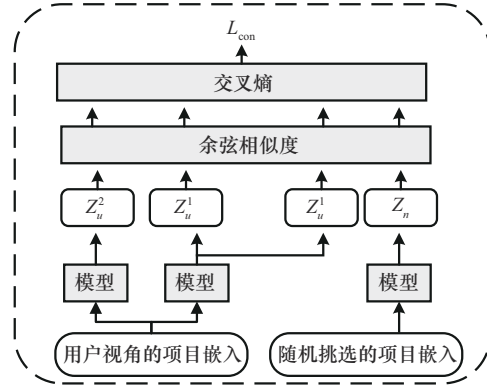


图4 对比学习模块流程

2.8 预测层

时间感知层以时间为切入点,以不同视角挖掘了用户感兴趣的下一个项目与时间的关系。信息处理层则是考虑了评分信息和项目流行度以及预训练语言模型学习出的隐向量的影响。对比学习模块,考虑了用户与用户之间的差异度对建模的影响。最后将各个要素组合到一起生成下一项的最终嵌入表示:

$$e_u = e_{itl} + e_p^a + e_r + e_{pop} + \alpha \cdot h_1 + \beta \cdot h_2 \quad (18)$$

其中, α 和 β 为可学习参数。最后通过使用 $r_o = e_u \cdot e_o$ 计算候选项嵌入 e_o 和 e_u 之间的相似度得分,并根据它们的相似度得分对所有候选项进行排序,从而生成Top- M 推荐列表。在给定的用户序列 S_u 中,出现在序列中的项目被视为正样本 $o \in S_u$,未出现在序列中的项被视为负样本 $o' \notin S_u$,生成一对训练样本 (o, o') ,使用二进制交叉熵作为损失函数:

$$L_{pred} = - \sum_{S_u \in S} \sum_{(o, o') \in D^u} [\log(\sigma(r_o)) + \log(1 - \sigma(r_{o'}))] \quad (19)$$

其中, D^u 表示对应于 u 的序列的训练样本集, $\sigma(\cdot)$ 表示sigmoid激活函数。最终损失函数如下。

$$\text{Loss} = L_{pred} + \gamma L_{con} + \lambda \|\theta\|_F^2 \quad (20)$$

其中, γ 为协调预测任务和对比学习的平衡系数,也是一个可学习参数, θ 为模型的可训练参数集合, λ 为正则化系数, $\|\cdot\|_F^2$ 表示Frobenius范数。

3 实验结果与分析

3.1 数据集

本文实验数据均选自亚马逊提供的公开数据集。为了提高实验合理性，选取了不同稀疏程度、不同数据大小、不同领域的数据集进行验证。数据集包含了用户 id、项目 id、交互时间和评分信息以及项目文本信息。为了进行预处理，实验遵循文献[4]中的过程。对于所有数据集，将评级或评论视为隐含的反馈，并根据时间戳对项目进行排序。只保留至少 5 个评级或评论的用户和至少 5 个评级或评论的项目。与文献[8-9]中的设置相同，实验使用序列中倒数第 2 个项目作为超参数调整的验证集，最后一个项目作为测试集。数据集统计数据见表 1。

表 1 数据集统计数据

数据集	#users	#items	#size
Software	21 663	375 147	0.659×10 ⁶
All_Beauty	32 586	324 038	0.371×10 ⁶
Digital_Music	456 992	840 372	1.584×10 ⁶
Prime-Pantry	10 814	247 659	0.572×10 ⁶
Luxury_Beauty	1 581	3 819	0.034×10 ⁶

3.2 评价指标

在前人研究的基础上，本次实验采用了 4 个常用的 Top- M 度量指标 Hit@10、NDCG@10、Hit@20、NDCG@20 来评估不同模型。Hit@10 (Hit@20) 统计基本事实项目在 10(20) 大推荐项目中所占百分比。NDCG@10、NDCG@20 是位置感知指标，为较高位置分配更大权重。为了避免所有用户项对的繁重计算，并加快评估过程，对于每个用户 u 随机抽取 100 个否定项，并将它们与基本事实项一起排序。将用户 u 在时间 t 之前未与之交互的项视为否定项。对每个模型进行了 5 次训练和评估，并记录了平均结果。

3.3 基准模型

(1) Caser^[22]：将一系列最近的物品嵌入时间和潜在空间中的“图像”中。该方法可以捕获高阶马尔可夫链，考虑了 L 个长度的序列的最新项。

(2) SASRec：利用自注意力来实现模型顺序性，而不需要任何额外时间信息。

(3) TiSASRec：使用自注意力机制并且结合了物品时间间隔信息。

(4) HTP：最近提出的时间感知自注意机制推荐模型，充分挖掘时间戳中包含的信息。

为了进行公平比较，在实验中，将模型维度都设为 50，迭代次数都设为 200，并且设定了相同的学习率。其他参数都为原模型的默认参数，特别是，本次实验将 Caser 模型的预测部分代码进行了重写，使其结果指标与其他模型保持一致。

3.4 对比实验

对比实验结果见表 2，显示了不同模型在所有数据集上的性能。粗体值表示所有模型中性能最好的，带下划线的值表示除最佳模型外的最佳性能。从实验结果可以看出，Caser 模型性能最差。SASRec 与 TiSASRec 有了显著的改进，DMTiSASRec 在选中的绝大部分数据集上获得了最佳结果。而 SASRec 只在 Prime_Pantry 数据集上的 Hit@20 指标上略高于 DMTiSASRec。说明 DMTiSASRec 可以在数据大小、领域、稀疏程度都不同的推荐任务上得到良好的推荐效果。说明时间感知层、对比学习模块、信息处理层以及多通道嵌入层作为一个整体对模型有提升作用。

3.5 消融实验

本文提出的模型由多个部分组成。在这一部分，通过消融实验，验证每个组件对模型性能的影响。消融实验结果如图 5 所示，其中 No_R 表示模型没有加入多通道嵌入层（评分信息只经过简单的嵌入层），No_TA 表示没有加入时间感知



表2 对比实验结果

数据集	Metric	Caser	SASRec	TiSASRec	HTP	DMTiSASRec
Software	Hit@10	0.331 2	<u>0.411 8</u>	0.409 8	0.358 2	0.457 9
	NDCG@10	0.199 3	0.241 7	<u>0.260 5</u>	0.223 6	0.279 3
	Hit@20	0.474 6	<u>0.582 8</u>	0.553 6	0.484 3	0.599 7
	NDCG@20	0.242 3	0.284 1	<u>0.295 2</u>	0.255 2	0.314 5
All_Beauty	Hit@10	0.925 7	0.940 3	<u>0.955 6</u>	0.950 5	0.969 3
	NDCG@10	0.789 9	0.804 5	<u>0.918</u>	0.906 3	0.941 7
	Hit@20	0.952 3	<u>0.971</u>	<u>0.971</u>	0.964 1	0.979 5
	NDCG@20	0.813 4	0.811 6	<u>0.921 2</u>	0.910 5	0.945 5
Digital_Music	Hit@10	0.315 5	0.567 3	0.243 9	0.514 6	0.568 4
	NDCG@10	0.292 2	<u>0.401 1</u>	0.251	0.315 3	0.402 1
	Hit@20	0.567 7	<u>0.742 7</u>	0.292 7	0.69	0.744 1
	NDCG@20	0.421 2	<u>0.443 4</u>	0.255 4	0.352 7	0.449 7
Prime_Pantry	Hit@10	0.454 2	0.470 6	0.471	<u>0.472 1</u>	0.476
	NDCG@10	0.265 5	<u>0.275 7</u>	0.274 8	0.273 3	0.290 6
	Hit@20	0.592 2	0.627 7	0.621 3	0.623 3	<u>0.624 2</u>
	NDCG@20	0.297 7	<u>0.313 9</u>	0.312 7	0.312	0.327
Luxury_Beauty	Hit@10	0.615 5	<u>0.650 2</u>	0.630 5	0.630 8	0.69
	NDCG@10	0.501 2	0.521 8	0.525 9	<u>0.530 5</u>	0.560 6
	Hit@20	0.638 8	<u>0.750 5</u>	0.719 9	0.724 8	0.778 1
	NDCG@20	0.514 2	0.546	<u>0.548 3</u>	0.545 5	0.582 1

层,只考虑序列位置信息。No_C表示没有加入对比学习模块, No_Bert表示没有考虑预训练语言模型的隐藏向量。Our表示完整的模型。实验结果表明,除了在极少部分的评分指标中,部分模块会抑制推荐性能,在绝大部分情况下,每个部分都能提高推荐性能,删除任何部分都会导致推荐性能不同程度的下降,其中时间感知模块对模型有显著影响。而预训练语言模型与对比学习模块对模型的影响相对较小,但在小而稀疏的数据集上效果比较明显。其原因是当数据集达到一定规模时,模型本身就可以通过序列学习相对精确的项目嵌入表示,而在如 All_Beauty、

Luxury_Beauty这两个小而稀疏的数据集中,没有足够数据来拟合精确的项目嵌入表示。而预训练语言模型以及对比学习模块对其进行了数据增强,缓解了数据稀疏的问题,使其可以学习相对更精确的项目嵌入表示。总体来说,对比学习模块、时间感知层、多通道嵌入层以及带有预训练语言模型的信息处理层均对提升模型推测效果有作用。

3.6 参数敏感性分析

模型的敏感性分析主要针对3个超参数,依次为正则化系数 λ 、最大时间间隔 k 和对比学习负对个数。保持其他参数不变,只改变待分析参数,测试4个评价指标。并且在3个数据集上进

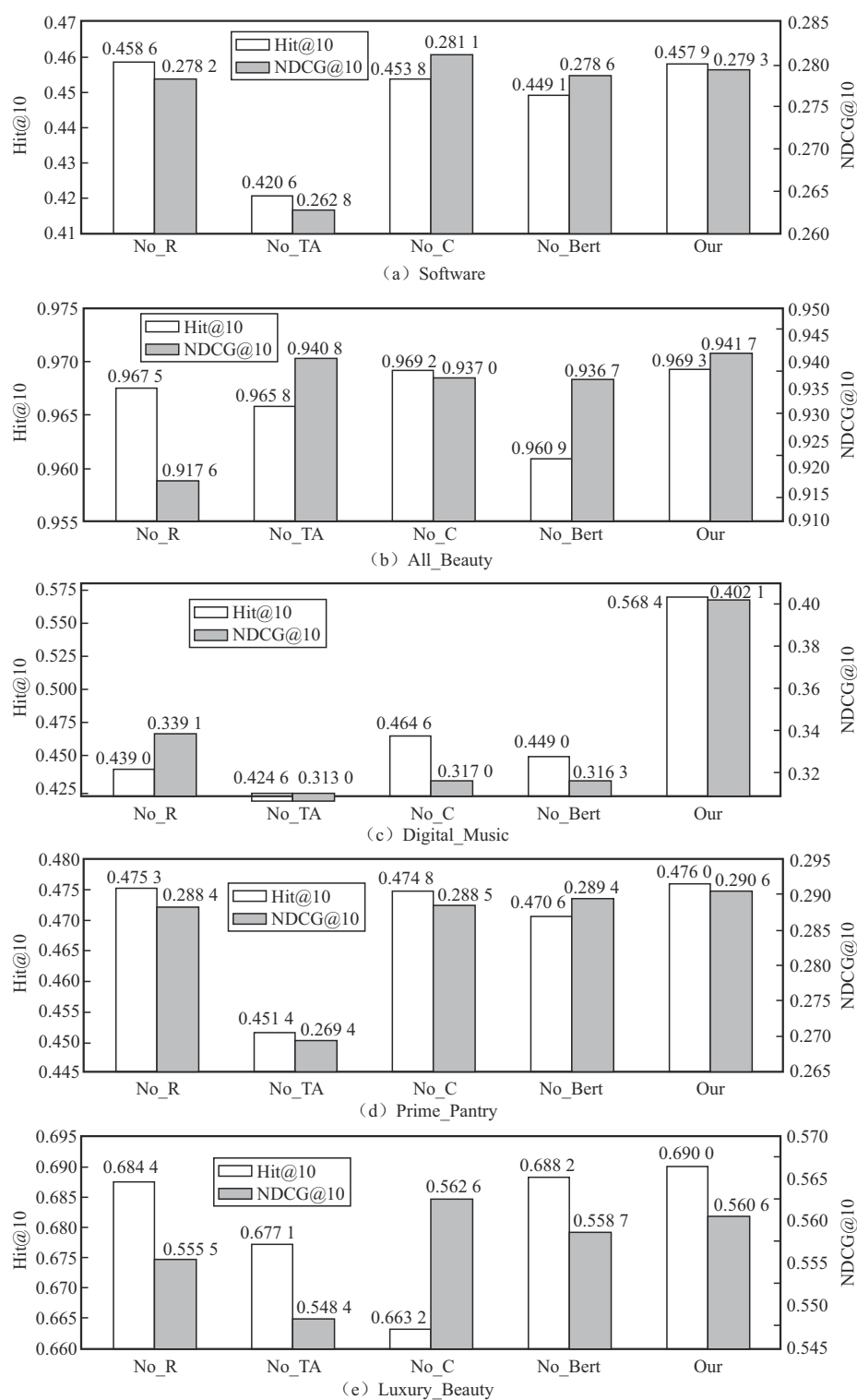


图5 消融实验结果

行验证。对于 λ ，在不同的数据集中取值不同，对效果的影响不稳定。对于 k ，在取值1 000时效果最佳。对于对比学习负对个数，在取值为4时

效果最佳。3个超参数的实验结果，正则化系数如图6所示，对比学习负对个数如图7所示，最大时间间隔 k 如图8所示。

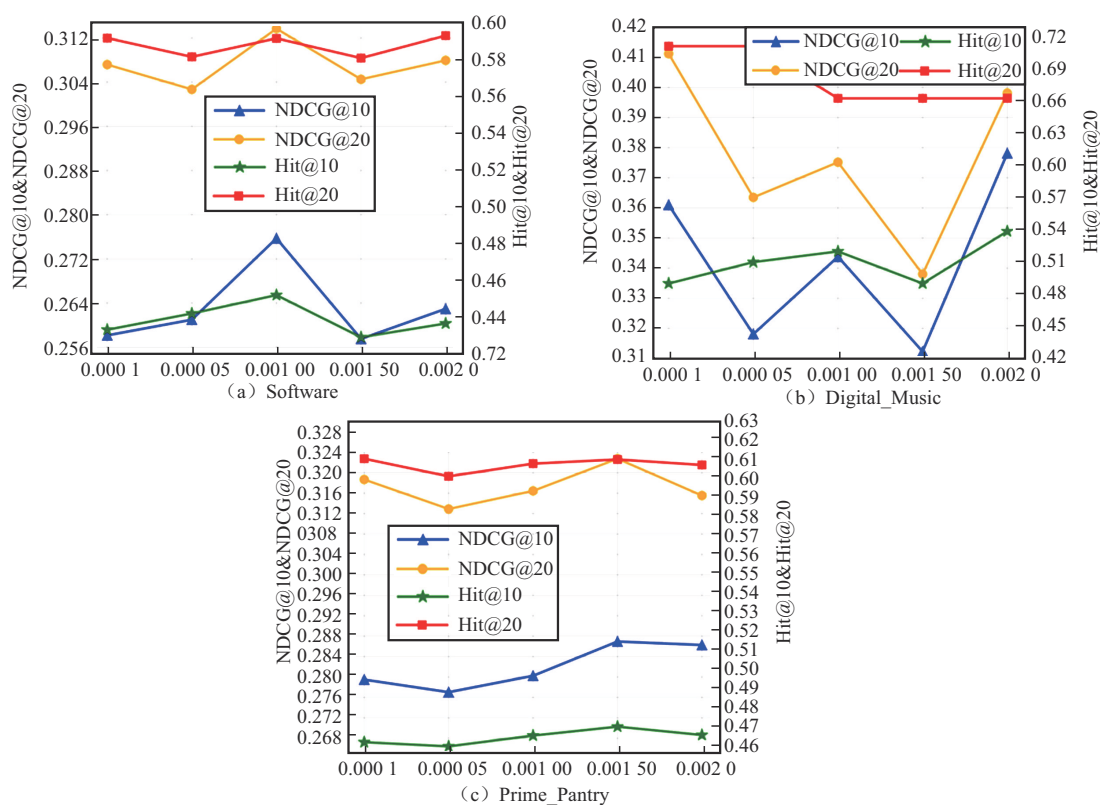


图6 正则化系数

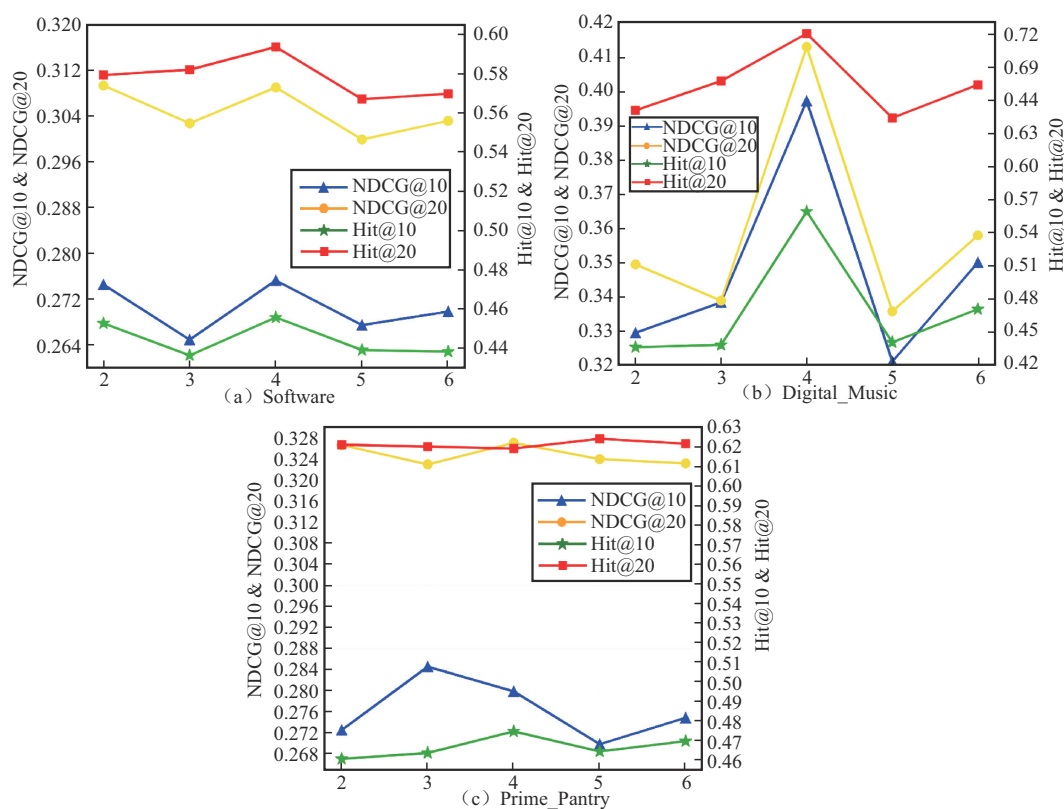
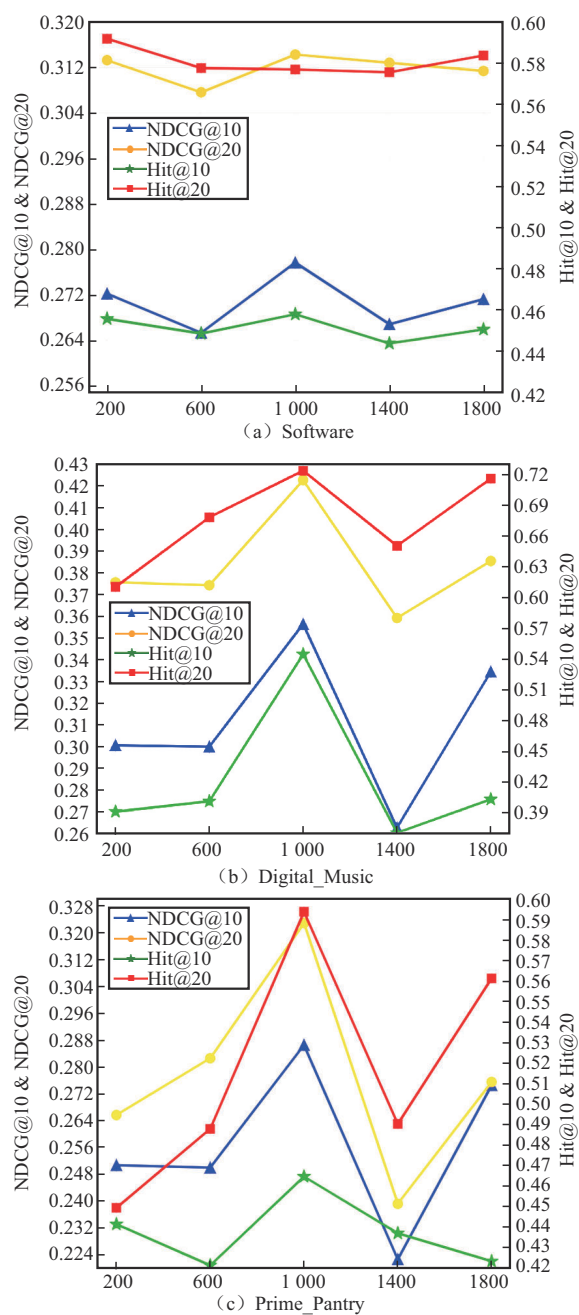


图7 对比学习负对个数

图8 最大时间间隔 k

4 结束语

本文提出了一个新型的序列推荐模型，DMTiSASRec模型。该模型在两方面有所改善。首先是对原有信息方法的处理改进。针对时间信息，提出了一种多模式的时间处理方式，使模型对时间信息的利用更加高效。针对评分信息，提

出了一种多通道的嵌入层来进行优化处理，可在一定程度上缓解数据中存在偏见的问题。其次是进行数据增强，分别引入了预训练语言模型、指标以及用户维度的对比学习。在一定程度上缓解了目前序列推荐模型遇到的数据稀疏的难题，提升了模型准确率。在5个不同类型、大小和稀疏程度的公开数据集上的大量实验表明，所提模型性能明显优于最新基准模型。进一步的消融实验证明了所提模型中各部分对于序列推荐任务的有效性。在未来的工作中，还需要借鉴多学科领域知识，尝试不同的更高效的推荐算法，进一步提升模型推荐性能。

参考文献:

- [1] RENDLE S, FREUDENTHALER C, SCHMIDT L. Factorizing personalized Markov chains for next-basket recommendation[C]//Proceedings of the 19th International Conference on World Wide Web. New York: ACM Press, 2010: 11-820.
- [2] HIDASI B, KARATZOGLOU A. Recurrent neural networks with top-k gains for session-based recommendations[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2018: 843-852.
- [3] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2017: 6000-6010.
- [4] KANG W, MCAULEY J. Self-attentive sequential recommendation[C]//Proceeding of the 18th IEEE International Conference on Data Mining. Singapore: IEEE Press, 2018: 197-206.
- [5] 黄熠昱, 纪其顺, 万超艺, 等. 注意力感知的边-节点交换图神经网络模型[J]. 电信科学, 2024, 40(1): 106-114.
HUANG Y M, JI Q S, WAN C Y, et al. Attention aware edge-node exchange graph neural network[J]. Telecommunications Science, 2024, 40(1): 106-114.
- [6] 杨帅, 王瑞琴, 马辉. 基于多通道的边学习图卷积网络[J]. 电信科学, 2022, 38(9): 95-104.
YANG S, WANG R Q, MA H. Multi-channel based edge-learning graph convolutional network[J]. Telecommunications Science, 2022, 38(9): 95-104.
- [7] 陶涛, 李珍, 王冀彬, 等. 基于图神经网络的权益推荐技术方案研究[J]. 电信科学, 2023, 39(8): 91-101.



- TAO T, LI Z, WANG J B, et al. Research on the graphical convolution neural network based benefits recommendation system strategy[J]. Telecommunications Science, 2023, 39(8): 91-101.
- [8] LI J, WANG Y, MCAULEY J. Time interval aware self-attention for sequential recommendation[C]//Proceedings of the 13th International Conference on Web Search and Data Mining. New York: ACM Press, 2020: 322-330.
- [9] CHEN R, LIANG G, MA C, et al. HTP: exploiting holistic temporal patterns for sequential recommendation[C]//Proceedings of the 2023 International Joint Conference on Neural Networks. Queensland, Piscataway: IEEE Press, 2023: 1-8.
- [10] ZHANG Y, YIN G, DONG Y, et al. Contrastive learning with frequency domain for sequential recommendation[J]. Applied Soft Computing, 2023, 144(C): 110481-110494.
- [11] LI Y, DU H, NI Y, et al. Multi-modality is all you need for transferable recommender systems[J]. arXiv preprint, 2023, arXiv: 2312.09602.
- [12] ZHU Y, LI H, LIAO Y, et al. What to do next: modeling user behaviors by time-LSTM[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia: AAAI Press, 2017: 3602 - 3608.
- [13] FAN Z, LIU Z, ZHANG J, et al. Continuous-time sequential recommendation with temporal graph collaborative[C]//Proceedings of the 30th ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2021: 433-442.
- [14] JUNSU C, DONGMIN H, SEONGKU K, et al. Learning heterogeneous temporal patterns of user preference for timely recommendation[C]//Proceedings of the 30th Web Conference. New York: ACM Press, 2021: 1274 - 1283.
- [15] CHEN M, PENG K, XUE L. Hierarchical gating networks for sequential recommendation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19). New York: ACM Press, 2019: 825 - 833.
- [16] SHI C Y, REN P J, FU D J, et al. Diversifying sequential recommendation with retrospective and prospective transformers[J]. ACM Trans. Inf. Syst. 2024, 42(5): 37-41.
- [17] LIU Z W, FAN Z W, WANG Y, et al. Augmenting sequential recommendation with pseudo-prior items via reversely pre-training transformer[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21). New York: ACM Press, 2021: 1608 - 1612.
- [18] WEI W, HUANG C, XIA L H, et al. Multi-modal self-supervised learning for recommendation[C]//Proceedings of the ACM Web Conference 2023 (WWW '23). New York: ACM Press, 2023: 790 - 800.
- [19] CHEN T, SIMON K, MOHAMMAD N, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of the 37th International Conference on Machine Learning. Online: JML R.org, 2020: 1597-1607.
- [20] ZHOU C, MA J, ZHANG J, et al. Contrastive learning for debiased candidate generation in large-scale recommender systems[C]// Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2021: 3985 - 3995.
- [21] XIE X, SUN F, LIU Z, et al. Contrastive pre-training for sequential recommendation[J]. arXiv preprint, 2020, arXiv: 2010.14395.
- [22] TANG J, WANG K. Personalized Top-N sequential recommendation via convolutional sequence embedding[C]//Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. Marina Del Rey: ACM Press, 2018: 565 - 573.

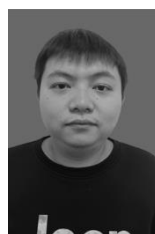
[作者简介]



李家乐 (1998-), 男, 湖州师范学院信息工程学院硕士生, 主要研究方向为数据挖掘、社交网络分析、个性化推荐。



王瑞琴 (1979-), 女, 博士, 湖州师范学院信息工程学院教授, 主要研究方向为自然语言处理、社交网络分析、个性化推荐。



于洋 (1999-), 男, 湖州师范学院信息工程学院硕士生, 主要研究方向为数据挖掘、自然语言处理、个性化推荐。