doi: 10.13413/j.cnki.jdxblxb.2023224

基于用户长短期偏好的个性化推荐

叶 榕, 邵剑飞, 邵建龙
(昆明理工大学 信息工程与自动化学院, 昆明 650500)

摘要: 针对现有序列推荐模型忽略用户的长期偏好和短期偏好, 导致推荐模型不能充分发挥作用, 推荐效果不佳的问题, 提出一种基于用户长短期偏好的个性化推荐模型. 首先, 针对长期偏好序列长且不连续的特点, 采用 BERT(bidirectional encoder representations from transformers)对长期偏好建模; 针对短期偏好序列短且与用户交互的间隔时间较短, 具有易变性, 采用垂直水平卷积网络对短期偏好建模; 在得到用户的长期偏好和短期偏好后, 利用激活函数进行动态建模, 然后利用门控循环网络对长短期偏好进行平衡. 其次, 针对用户在日常交互中的误碰行为, 采用稀疏注意力网络进行建模, 在对长短期偏好建模前使用稀疏注意力网络进行用户行为序列处理; 用户特征偏好对推荐结果也会有影响, 使用带有偏置编码的多头注意力机制对用户特征进行提取. 最后, 将各部分得到的结果输入到全连接层得到最后的输出结果. 为验证本文模型的可行性, 在数据集 Yelp 和 MovieLens-1M 上进行实验, 实验结果表明该模型优于其他基线模型.

关键词: 序列推荐; 长期偏好; 短期偏好; 稀疏注意力网络; 垂直水平卷积网络
中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5489(2024)03-0615-14

Personalized Recommendations Based on Users' Long- and Short-Term Preferences

YE Rong, SHAO Jianfei, SHAO Jianlong
(*Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming* 650500, *China*)

Abstract: Aiming at the problem that the existing sequence recommendation model ignored the users' long-term preference and short-term preference, resulting in the recommendation model not being able to fully play its role and the recommendation effect being poor, we proposed a personalized recommendation model based on the users' long- and short-term preferences. Firstly, for the characteristics of long and discontinuous long-term preference sequences, BERT (bidirectional encoder representations from transformers) was used to model the long-term preference, for the short-term preference sequences and the short interval time between interaction with the user, which was volatile, vertical and horizontal convolutional networks were used to model the short-term preference, after obtaining the users' long-term preference and short-term preference, activation functions were used to model dynamically, and then a gated recurrent network was used to balance the long- and short-term preferences. Secondly, for the users' mis-touching behavior in daily interaction, sparse

收稿日期: 2023-06-02.
第一作者简介: 叶 榕(1998—), 女, 汉族, 硕士研究生, 从事推荐系统与自然语言处理的研究, E-mail: 771740131@qq.com.
通信作者简介: 邵剑飞(1970—), 男, 汉族, 硕士, 副教授, 从事数据挖掘与自然语言处理的研究, E-mail: 1515346516@qq.com.
基金项目: 国家自然科学基金(批准号: 61732005).

attention network was used for modeling, and sparse attention network was used to process the users' behavioral sequences before modeling the long- and short-term preferences. User feature preferences also had an impact on the recommendation results, and user features were extracted by using a multi-head attention mechanism with bias coding. Finally, the results obtained from each part were input into the fully connected layer to obtain the final output result. In order to verify the feasibility of the proposed model, experiments were conducted on Yelp and MovieLens-1M datasets, and the results show that the proposed model outperforms other baseline models.

Keywords: sequential recommendation; long-term preference; short-term preference; sparse attention network; vertical and horizontal convolutional network

推荐系统(recomm-ender systems, RS)能很好地帮助用户缓解信息过载带来的问题,已广泛应用于网络购物(如淘宝、天猫、京东等)[1]、电影(如猫眼、淘票票等)[2-3]、音乐(如QQ音乐、网易云等)、新闻阅读(如今日头条)[4]等领域. 目前使用较多的推荐系统主要分为两类:一般推荐系统和序列推荐系统. 一般推荐的目的是了解用户的长期偏好. 基于因式分解的协同过滤[5]是该领域应用最广的技术,其建模方式大多数是根据用户与物品之间的交互建模,而这种交互可能是显式的或隐式的,建模后得到的结果常趋于静态. Top-N 推荐[6-7]力求基于用户与物品之间的历史交互模拟用户对物品的偏好,在建模过程中只依赖于静态交互,而忽略了顺序依赖关系,将用户与项目的所有交互视为同等重要. 而用户的行为意图并非一成不变,在一段时间内用户的行为意图会受需求、环境等因素的影响. 为满足现实需求,近年来,序列推荐因其在捕获用户与物品间顺序关系方面的优势而得到广泛关注. 序列推荐弥补了一般推荐的不足.

循环神经网络(recurrent neural networks, RNN)[8-9]在自然语言处理(NLP)方面性能优异,该方法目前已成为顺序推荐的主流模型[10-13]. 这些方法具有短期记忆,因此会推荐与用户近期行为更相关的项目. 虽然上述模型对用户长短期偏好进行建模取得了较好的效果,但它们没有对长短期偏好分别有针对性的建模,对行为序列中的误点行为也未进行有效处理,从而影响了最后的推荐效果.

基于上述问题,本文提出一种融合稀疏网络与垂直水平卷积网络对用户长短期偏好建模的序列推荐方法,命名为FDSRec. 本文的贡献如下:

1) 提出了融合稀疏网络与垂直水平卷积网络对用户长短期偏好建模的序列推荐方法. 针对长期偏好,交互序列较长,普通循环神经网络在多次迭代后会出现梯度消失和梯度爆炸问题,因此无法处理长依赖问题,采用BERT(bidirectional encoder representations from transformers)对历史交互序列建模;针对短期偏好变化快,单一编码方式无法有效获取有效信息的问题,采用垂直水平卷积网络建模.

2) 针对用户的交互序列中会产生一些误点行为,导致一些干扰因素,采用稀疏网络进行建模;同时将 α-entmax 函数应用到其中,以减弱用户在误点时产生无关信息带来的影响.

3) 针对建模过程中的长期偏好与短期偏好平衡问题,采用门控循环单元解决该问题,将产生的长期偏好和短期偏好进行处理,给予不同的权重,进而产生精确的推荐结果.

4) 采用融合偏置编码的多头注意力机制进行特征提取. 位置编码的出发点在于关注了序列的位置顺序,偏置编码将其中的位置信息和与之有关的其他信息进行结合,构成一种新的表示输入模型,该模型具有学习能力. 基于此,本文采用融合偏置编码进行特征提取,融合偏置编码后的模型关注了序列之间的顺序关系,在得到序列的位置关系与顺序关系后,将二者融合能更好地进行特征提取.

5) 在两个真实的公开数据集上进行实验,实验结果表明FDSRec方法优于其他基线模型.

1 相关工作

传统的机器学习方法可用于序列推荐,基于Markov链(Markov chain, MC)的方法将用户的行为序列映射到Markov链中,根据用户的行为序列对用户的下一个行为进行预测[14]. Rendle等[15]提出

了 FPMC(factorizing personalized Markov chains)模型，通过一种基于一阶 Markov 链和矩阵分解进行结合，然后捕获序列模式和用户的长短期偏好，并据此进行推荐. 但基于 Markov 链的方法通常侧重于相邻序列之间的依赖关系，从而导致基于 Markov 链的推荐方法不能捕获长期偏好. 此外，这些方法不能有效模拟用户兴趣的动态变化.

由于深度神经网络的飞速发展，近年来，许多研究人员提出了许多基于 RNN 的方法对用户交互序列中的序列模型进行建模. 文献[8]提出了一种基于门控循环单元(gated recurrent unit，GRU)的序列推荐模型(GRU for recommendation，GRU4Rec)，该模型可通过单个门控单元同时控制遗忘因子和更新状态单元的决定，用于预测下一个用户的目的. 但该模型只能进行单向的信息提取，对信息的更新有一定的限制. Tang 等[16]提出了一种卷积序列嵌入推荐模型(convolutional sequence embedding recommendation model，Caser)，该模型从序列中提取若干个连续的项作为输入嵌入到神经网络中，使用水平卷积层和垂直卷积层捕获序列的局部特征，再通过全连接层得到更高级别的特征. 但卷积网络只对当前特征进行提取，忽视了之前特征对推荐结果的影响. Chen 等[17]和 Huang 等[18]采用记忆网络改进顺序推荐. STAMP(short-term attention/memory priority)利用多层感知器(MLP)网络捕捉用户的一般兴趣和当前兴趣[19].

注意力机制在建模序列数据中应用广泛，如机器翻译[20-21]和文本分类. 近期一些工作尝试采用注意力机制提高推荐性能和可解释性[22-23]. 如 Li 等[22]将注意力机制并入 GRU 以捕获用户的顺序行为和基于会话推荐中的主要目的. 文献[24]设计了一种基于自注意力的序列推荐模型(self-attention based sequential model，SASRec)，在每个时间步自适应地为之前的物品赋予权重，但该方法仍是一个单项的模型，使用一个偶然的注意掩模，依赖用户这一时刻之前的交互序列，并用其下一时刻作为标签对模型训练，可能会导致模型偏差. 相比之下，BERT4Rec 模型能实现双向编码，该模型针对短期序列有较好的效果.

本文根据不同长度的序列进行有针对性建模. 针对长期偏好序列，采用 BERT 对长期偏好建模. 因为 BERT 在文本和机器翻译领域都取得了较好效果，普洪飞等[25]将 BERT 应用在序列推荐中. BERT 在建模时使用双向编码的方式对用户行为序列进行编码，可利用双向编码的能力挖掘隐藏的信息. 此外，可以实现快速的并行方式，准确率也较传统模型有提高. 针对短期偏好，采用 Caser 进行建模，Caser 主要由两部分组成，能对单个目标项和后续目标都产生一定的作用，捕获用户最近活动的动态模式.

1.1　BERT 模型

BERT 模型[26]是一个基于预训练的模型，与传统模型的不同之处在于该模型采用了新方法，能生成一个深度的双向语言表征，且能充分利用上下文信息，在实际应用中也取得了较好效果. BERT 模型由多个 Transformer 层重叠而成，Transformer 的内部结构如图 1 所示.

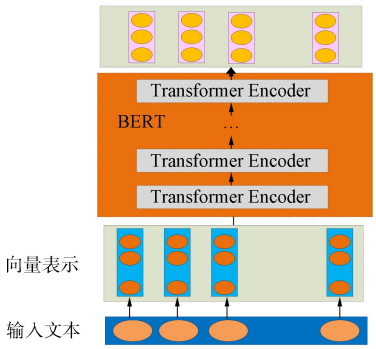

图 1　Transformer 内部结构

Fig. 1　Internal structure of Transformer

Transformer 层主要由 Encoder 和 Decoder 组成，在每层中还有其他小层. 在 Encoder 层中包含一个自注意力网络层和一个前馈神经网络；而 Decoder 层比 Encoder 层多一层注意力层. 其定义如下：在网络中的 query($\boldsymbol{Q}$)和 key-value($\boldsymbol{K}$)通过自注意力机制映射到某个输出的过程，经过该过程输出的向量即为根据 query 和 key 计算得到的权重作用于 value($\boldsymbol{V}$)的权重和. 多头注意力将信息进行融合，然后进行输出：

$$\mathrm{MultiHead}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Concat}(\mathrm{Head}_1,\mathrm{Head}_2,\cdots,\mathrm{Head}_h)\boldsymbol{W}^O, \tag{1}$$

$$\mathrm{Head}_i=\mathrm{Attention}(\boldsymbol{Q}\boldsymbol{W}_i^{\boldsymbol{Q}},\boldsymbol{K}\boldsymbol{W}_i^{\boldsymbol{K}},\boldsymbol{V}\boldsymbol{W}_i^{\boldsymbol{V}}). \tag{2}$$

1.2 门控循环网络

序列推荐中时间序列的时间步距离较大从而影响信息的捕捉. 而门控循环神经网络有特殊的结构，能通过门控制信息的流动进而解决该问题. 其中，GRU[27]是一种较常用的门控循环神经网络，其内部结构如图2所示.

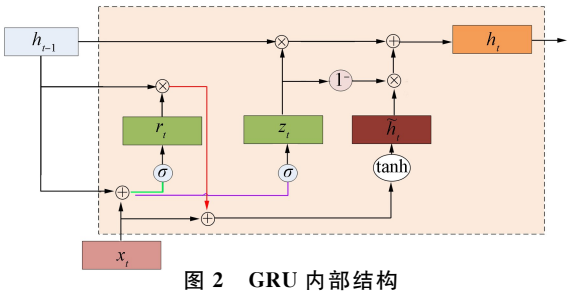

图2 GRU内部结构

Fig. 2 Internal structure of GRU

门控循环单元定义如下：

$$h_t = h_{t-1} + g(x_t, h_{t-1}; \theta), \tag{3}$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot g(x_t, h_{t-1}; \theta),\ z_t \in (0,1), \tag{4}$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z), \tag{5}$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h), \tag{6}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r), \tag{7}$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t, \tag{8}$$

其中 h_t 为网络的更新状态，z_t 为更新门，$\tilde{h}_t$ 为当前时刻的候选状态，r_t 为重置门，$W_z, U_z, W_h, U_h, W_r, U_r$ 为权重参数.

1.3 卷积神经网络

卷积神经网络(convolutional neural network，CNN)[28]是一个由卷积层和子采样层构成的特征抽取器，其主要由三部分组成：输入层(input)、特征提取层和输出层(output). 其中特征提取层又包含三层，分别是卷积层、池化层和全连接层. CNN的基本结构如图3所示.

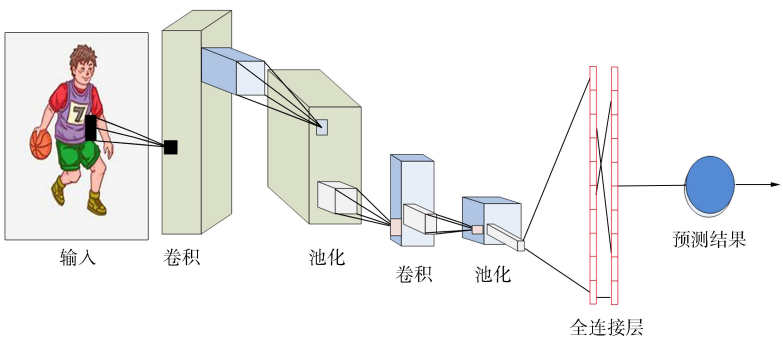

图3 CNN的基本结构

Fig. 3 Basic structure of CNN

卷积神经网络的作用较多：它能将音频之类的文件转化为图像，实现信息的可视化；还可以很好地处理各种目标，例如文本处理、视频处理等. 在现实生活中很多信息不仅是一维的简单信息，若要更好地获取信息就要采取不同的方式. CNN的卷积滤波器在捕获局部特征以进行图像识别[29-30]和自然语言处理[31]方面获得了成功. 由于CNN的特点扩大了其使用范围，Tang等[16]提出了一种Caser模型，该模型利用CNN的特点对RNN建模时对相邻但不相关的用户-物品交互而产生错误依赖的特点进行改进. CNN在序列推荐中提取特征的步骤如图4所示.

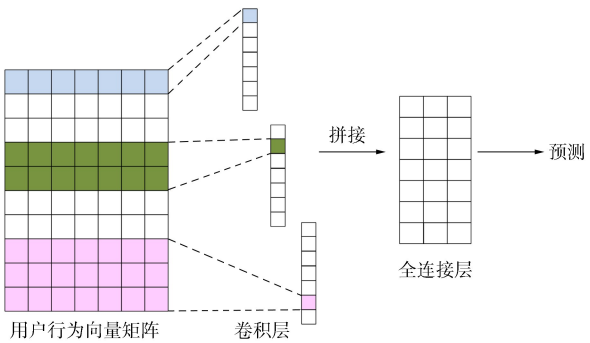

图4 CNN的特征提取

Fig. 4 Feature extraction of CNN

2 本文方法

2.1 相关问题概述

为便于描述，对用户集合和项目集合使用统一的符号表示，用 U 表示用户集合，用 I 表示项目集

合. 在用户与项目产生交互时，用 S^u 表示交互序列. 对每个用户 $u\in U$，$S^u\in I$，$S^u=\{S(1),S(2),\cdots,S(T)\}$，其中 T 显示了相互作用序列的长度和相互作用的数量. 此外，交互过程中产生的相互作用包含用户和项目本身的特征. 本文的目的是通过模拟用户与项目之间的互动序列预测用户后续感兴趣的内容.

2.2　模型描述

本文的总体模型如图 5 所示. 由图 5 可见，整个模型分为四部分，分别为稀疏网络层、长短期偏好建模层、用户项目特征提取层和输出层.

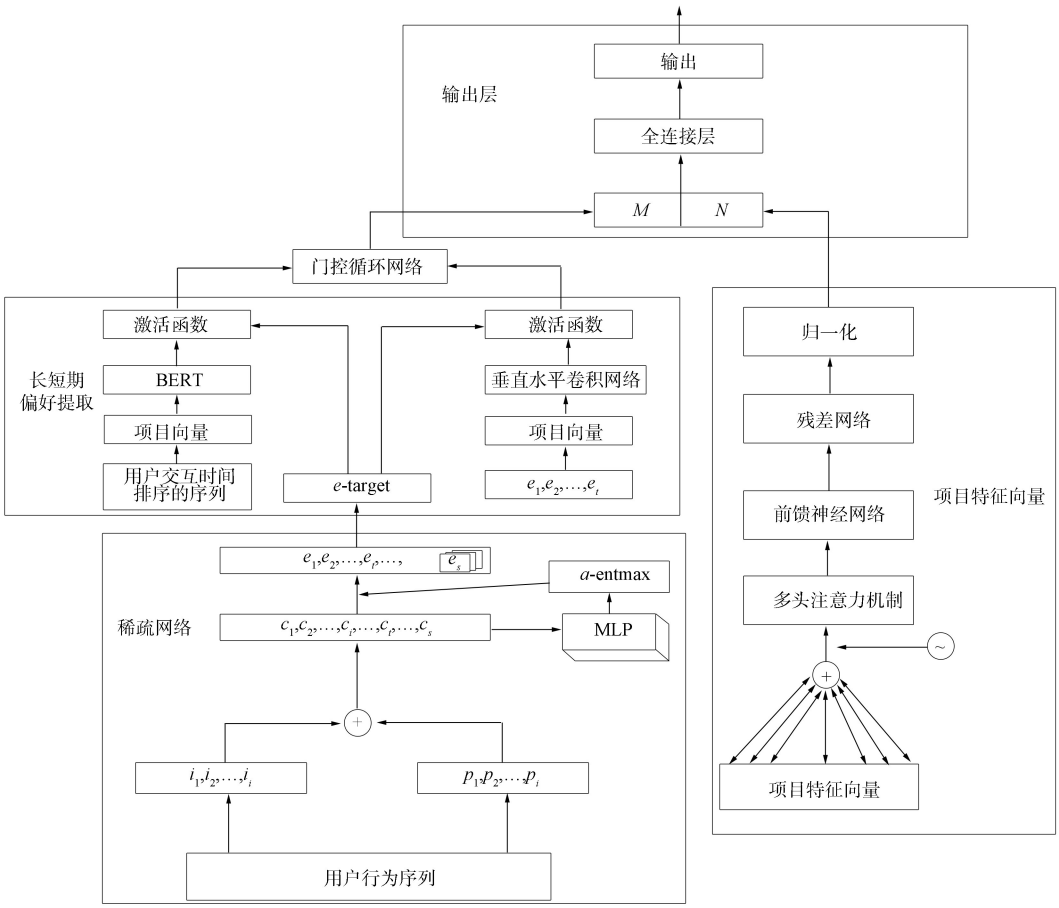

图 5　本文模型结构

Fig. 5　Structure of proposed model

2.3　稀疏网络层

将用户与项目在交互过程中产生的交互序列表示为 S^u. 在整个模型框架图中，稀疏注意力网络为最底层模型，即产生的交互序列先要输入到该层网络模型中进行处理. 在用户与项目的交互过程中会产生噪声，例如用户的误点行为，这些干扰项目会对最后的推荐结果产生影响. 在传统模型中，是将交互序列在建模过程中进行去噪，但效果不佳. 本文采取在进行用户长短期偏好建模前便进行交互序列的去噪处理. 针对交互序列中产生的噪声，本文采用稀疏注意力网络对交互序列进行一个过滤. 在稀疏网络有两个值得借鉴的点：一是在该网络中通过去噪处理后能得到一个目标项目可提高在后续兴趣提取中的准确性；二是在该网络中引入了 α-entmax 函数，与常用的 Softmax 函数相比，该函数具有选择性、更紧凑、注意力更集中的优点.

在稀疏注意力网络中的操作如下：

1）嵌入层. 将用户与项目交互产生的交互序列 S^u 转换为两个向量：

$$\boldsymbol{c}_i=\mathrm{concat}(\boldsymbol{x}_i,\boldsymbol{p}_i), \tag{9}$$

$$\widetilde{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_t,\boldsymbol{c}_s\}, \tag{10}$$

其中：$\boldsymbol{x}_i \in \mathbb{R}^d$ 为项目的嵌入；$\boldsymbol{p}_i \in \mathbb{R}^d$ 为位置的嵌入；$\boldsymbol{c}_i \in \mathbb{R}^{2d}$ 为项目和位置的串联嵌入；$\boldsymbol{c}_s$ 由 $\boldsymbol{x}_s$ 和 $\boldsymbol{p}_s$ 组成，能包含特殊信息的索引及待预测项目的位置，从而更准确地进行预测.

2) 目标嵌入学习. 在嵌入层处理后输出的结果为 $\widetilde{C}$，然后将输出结果输入到目标嵌入学习层，进行噪声过滤. 在预测项目之前需对无关信息进行处理，从而减少无关信息的干扰，提高预测结果.

先用带有稀疏变化的网络捕获交互序列中的依赖关系：

$$\widetilde{A} = \alpha\text{-entmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{2d}}\right)\boldsymbol{V}, \tag{11}$$

$$\alpha = \sigma(\boldsymbol{W}_a \boldsymbol{c}_s + b_a) + 1, \tag{12}$$

$$\boldsymbol{Q} = f(\widetilde{C}\boldsymbol{W}^Q + \boldsymbol{b}^Q), \tag{13}$$

$$\text{sparsemax}(\boldsymbol{x}) = \underset{\boldsymbol{p}}{\operatorname{argmin}} \| \boldsymbol{p} - \boldsymbol{x} \|^2, \quad \boldsymbol{p} \in \Delta^{d-1}, \tag{14}$$

$$\alpha\text{-entmax}(\boldsymbol{x}) = \underset{\boldsymbol{p}}{\operatorname{argmax}}\ \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{H}_\alpha^{\mathrm{T}}(\boldsymbol{p}), \quad \boldsymbol{p} \in \Delta^{d-1}, \tag{15}$$

$$\boldsymbol{H}_\alpha^{\mathrm{T}}(\boldsymbol{p}) = \frac{1}{\alpha(\alpha-1)}\sum_j (\boldsymbol{p}_j - \boldsymbol{p}_j^\alpha), \quad \alpha \neq 1, \tag{16}$$

$$\boldsymbol{H}_\alpha^{\mathrm{T}}(\boldsymbol{p}) = \boldsymbol{H}_s(\boldsymbol{p}), \quad \alpha = 1, \tag{17}$$

$$E = \text{SAN}(\widetilde{C}), \tag{18}$$

其中 $\boldsymbol{Q}$ 为查询的表示，$\boldsymbol{K}$ 为关键矩阵，$\boldsymbol{V}$ 为所参加项目的值矩阵，$\boldsymbol{W}_a \in \mathbb{R}^{2d\times 2d}$ 为加权矩阵，$b \in \mathbb{R}$ 为偏置值，σ 为激活函数 Sigmiod，α 为[0,1]之间的一个变量，$\boldsymbol{W}^Q \in \mathbb{R}^{2d\times 2d}$ 为加权矩阵，$\boldsymbol{b}^Q \in \mathbb{R}^{2d}$ 为偏置向量，$f(\cdot)$为激活函数 ReLU，E 为最终输出. 本文下面将其输出表示为 $\boldsymbol{e}_{\text{target}}$.

上述过程可简述为：将用户与项目交互过程中产生的交互序列 S^u 进行向量嵌入，分别为项目嵌入和位置嵌入，两者嵌入后将其相加输入到 MLP 中并使用 α-entmax 函数，最后输出目标项目的向量 $\boldsymbol{e}_{\text{target}}$.

2.4 长短期偏好建模

2.4.1 长期偏好建模

长期偏好是从用户长期的交互序列中学习到用户的一个相对稳定的、长期的偏好. 用户的长期偏好通常不容易改变，一般包含在用户与项目交互序列的整个过程中. 因此，长期偏好具有序列过长且不连续的特点，使得长期偏好的建模较困难. 普通的网络无法在长期序列中捕获交互序列之间的相互依赖关系，于是提出了基于神经网络的模型. 序列交互数据具有独特的特点，在处理序列数据方面循环神经网络有独特的优势，因此其被广泛应用于序列推荐建模. 但由于 RNN 本身的独特性，在建模过程中需要非常深的网络进行计算，而网络过深会产生梯度消失和梯度爆炸的问题. 基于长短时记忆的网络(long short-term memory，LSTM)和基于 GRU 的网络成为解决该问题的有效途径. 其中，LSTM 在序列建模时效果很好[32-33]，GRU 是基于 LSTM 的改进，不仅结构更简单，且效率也更高.

本文采用 BERT 对长期偏好建模[34]，如图 5 所示. 对用户与项目交互的时间排序的序列 S^u 进行编码后经过项目嵌入输出为项目向量 $\boldsymbol{e}$，将输出的项目向量 $\boldsymbol{e}$ 再输入到 BERT 模型中，得到输出后与稀疏注意力网络得到的目标向量一起输入激活函数模块[35]中，可以自适应地进行权重的分配，通过这种方法可减弱动态兴趣变化对推荐的影响，此时便能得到用户偏好的长期偏好 L. 激活函数定义为

$$a_i^L = \frac{\exp\{M_i^q \boldsymbol{W}_L \boldsymbol{e}_{\text{target}}\}}{\sum_{t=1}^{T}\exp\{M_t^q \boldsymbol{W}_L \boldsymbol{e}_{\text{target}}\}}, \tag{19}$$

$$L = \sum_{t=1}^{T} a_i^L M_t^q, \tag{20}$$

其中 T 为交互序列的长度，$\boldsymbol{W}_L$ 为 $d\times d$ 的权重参数，a_i^L 为权重，L 为长期偏好的输出.

2.4.2 短期偏好建模

用户的偏好不仅有长期偏好，还有短期偏好. 用户的长期偏好可通过交互的历史序列推断，但这不能表示用户近期的偏好. 用户的短期偏好主要描述了用户当前的偏好，想要全方位的预测用户的偏

好，就不能只单纯考虑用户的长期偏好，特别是在数据稀疏的情况下，用户的短期偏好对最终预测也十分重要. 因此，本文采用垂直水平卷积网络对短期偏好建模.

垂直水平卷积网络的原模型为 Caser 模型[16]，该模型将最近的项目序列嵌入在时空的图像中，并用卷积滤波器研究序列模型作为图像的局部性质[29-31]. 嵌入层、卷积层和全连接层为 Caser 模型的 3 个主要部分. 本文只使用该模型中的卷积层，它采用卷积神经网络捕获用户最近活动的动态模式. 该模型的卷积层主要由两部分组成：水平卷积网络和垂直卷积网络，它们分别用于发现联级和点级的序列模式，联级模式是指多个先前操作对后续目标的影响，点级模式是指历史序列中单个项目对目标项目的影响. 卷积层包含垂直卷积和水平卷积，两种不同的网络用不同的方式对信息进行提取.

垂直水平卷积网络是卷积序列嵌入推荐中的一个模型，这样在使用垂直水平卷积网络建模时就能对用户的兴趣偏好有一个更全面的建模. 操作过程如下.

1) 水平卷积网络. 水平卷积层中有 n 个水平卷积 $\boldsymbol{F}^k \in \mathbb{R}^{h\times d}$，其中 $1\leqslant k\leqslant n$，$h\in\{1,2,\cdots,L\}$ 为卷积的高度. 例如，当 $L=4$ 时，可以选择有 $n=8$ 个卷积核，两个卷积核对应 $\{1,2,3,4\}$ 中的每个 h. 卷积核从上到下滑动，并与第 i 项 $\boldsymbol{E}$ 的所有水平维度相互作用. 交互作用后即给定第 i 个卷积值：

$$c_i^k = \varphi_c(\boldsymbol{E}_{i:i+h-1} \odot \boldsymbol{F}^k), \tag{21}$$

其中：$\odot$ 表示内积算子；$\varphi_c(\cdot)$ 表示卷积层的激活函数，该值是 $\boldsymbol{F}^k$ 与由第 i 行到第 $(i-h+1)$ 行形成的子矩阵 $\boldsymbol{E}_{i:i-h+1}$ 之间的内积. $\boldsymbol{F}^k$ 的卷积即为一个向量：

$$\boldsymbol{c}^k = (c_1^k, c_2^k, c_3^k, \cdots, c_{L-h+1}^k). \tag{22}$$

得到向量后对 $\boldsymbol{c}^k$ 进行最大池化操作，从这个特定的卷积产生的值中获取一个最大值. 获取最大值的操作是卷积层提取的最重要特征. 因此，对该层的 n 个卷积的输出值为

$$o = \{\max(\boldsymbol{c}^1), \max(\boldsymbol{c}^2), \cdots, \max(\boldsymbol{c}^n)\}. \tag{23}$$

水平卷积通过嵌入 $\boldsymbol{e}$ 与每个连续的 h 项相互作用. 嵌入和卷积学习都是为最小化编码目标项预测误差的目标函数. 通过各种滑动卷积核能拾取重要的信号. 因此，可以训练水平卷积捕获具有多个联合大小的联合级别模式.

2) 垂直卷积网络. 垂直卷积网络与水平卷积网络的原理相同，但垂直卷积网络中卷积核的滑动方式是从左向右滑动，产生的垂直卷积结果为

$$\tilde{\boldsymbol{c}}^k = \sum_{l=1}^{L} \widetilde{\boldsymbol{F}}_l^k \cdot \boldsymbol{E}_l, \tag{24}$$

其中 $\boldsymbol{E}_l$ 是 $\boldsymbol{E}$ 的第 l 行. 因此，若要聚合项目的潜在表示，可用垂直卷积聚合先前项目的嵌入，再用垂直卷积进行加权和[34]得到点级序列模式.

上述过程可简述为：与长期偏好前期建模过程相同，将经过项目嵌入后得到的项目向量输入到垂直水平卷积网络中，得到输出后，将输出与目标嵌入向量 $\boldsymbol{e}_{\text{target}}$ 同时输入激活函数中得到短期偏好的输出 S_{t-1}，S_{t-1} 表示用户的短期行为集合，即用户最近(当前)的需求.

2.4.3　长短期偏好平衡

在现实生活中，针对不同用户偏好采用不同方法进行建模能得到更全面的预测推荐. 在得到二者的建模结果后如何采用一个合适的方式对长短期偏好进行平衡也是不可避免的问题. 一般的处理方式是进行两部分的线性连接或加权求和，但这些处理方式一般都是处于一种理想状态——所有用户的偏好都在历史序列中. 但现实生活中用户的意图受很多因素的影响，最重要的一点就是长短期的偏好对用户的影响不一样. 因此，本文选择门控循环网络平衡长短期的偏好，门控循环网络能控制具体信息的保留与丢弃，同时还能解决梯度消失和梯度爆炸的问题，本文门控循环网络自适应平衡长短期偏好的权重，定义如下：

$$R_t = \sigma(L_t W_{xr} + S_{t-1} W_{hr} + b_r), \tag{25}$$

$$Z_t = \sigma(L_t W_{xz} + S_{t-1} W_{hz} + b_z), \tag{26}$$

$$G = \tanh(L_t W_{xh} + R_t H_{t-1}) W_{hh} + b_h, \tag{27}$$

$$M = G * L_t + (1-G) S_{t-1}, \tag{28}$$

其中 * 为元素乘积，σ 为 Sigmoid 激活函数，W_{hr}，W_{hz}，W_{xh} 为权重参数，b_r，b_z，b_h 为偏置参数，R_t 为重置门，Z_t 为更新门，M 为平衡长短期偏好后的最终输出.

上述过程可简述为：在经过稀疏网络得到 $\boldsymbol{e}_{\text{target}}$ 后，与得到的长期偏好 L_t 和短期偏好 S_{t-1} 分别输入到激活函数中进行平衡；将平衡后的结果同时输入到门控循环网络中进行长短期偏好平衡，自适应地给予权重，能更好地对长短期偏好进行提取，得到平衡长短期偏好后的最终输出为 M.

2.5 用户项目特征提取

在实际生活中的选择会受多种因素影响，其中项目特征就是一个不可忽视的影响因子. 例如，对于商品的选择有不同的质量、价格等，不同的人会根据自己的需求进行不同的选择，有的人可能较重视价格，有的人可能较重视质量等. 看电影是日常生活中一种较常见的活动，在选择电影时，不同的人会根据电影的类型、电影的主演、上线的时间选择. 这些因素即为项目的特征，它对用户的选择影响也较大，所以也是一个不能忽视的问题.

本文采用带有偏置编码的多头注意力机制对项目特征进行提取. 在传统建模过程中采用位置编码，但如果想精确获取项目的特征偏好，还需要捕获交互的顺序关系及其中存在的偏差，故采用偏置编码的方式进行特征初步提取. 操作过程如下.

1) 偏置编码. 采用加入偏置编码的自注意力对项目特征向量进行提取：

$$\boldsymbol{BE}_{(k,t,c)}=\boldsymbol{W}_k^K+\boldsymbol{W}_t^T+\boldsymbol{W}_c^C,\qquad \boldsymbol{BE}\in\mathbb{R}^{K\times T\times d_{\text{model}}},\tag{29}$$

其中：$\boldsymbol{W}^K\in\mathbb{R}^K$ 为交互的偏置向量，k 为会话的索引；$\boldsymbol{W}^T\in\mathbb{R}^T$ 为会话中位置的偏置向量，t 为会话中行为的索引；$\boldsymbol{W}^C\in\mathbb{R}^{d_{\text{model}}}$ 为行为嵌入中单元位置的偏置向量，c 为行为嵌入中单元的索引. 在添加偏置编码后，用户行为会话 S 发生改变：

$$S^u=S^u+\boldsymbol{BE}.\tag{30}$$

对项目特征向量添加偏置编码后输入多头注意力机制进行特征提取，然后依次经过前馈神经网络、残差网络和归一化后输出项目特征 N.

2) 多头注意力机制：

$$\text{Head}_h=\text{attention}(\boldsymbol{S}_{kh}\boldsymbol{W}^{\boldsymbol{Q}},\boldsymbol{S}_{kh}\boldsymbol{W}^{\boldsymbol{K}},\boldsymbol{S}_{kh}\boldsymbol{W}^{\boldsymbol{V}}),\tag{31}$$

$$\text{Head}_h=\text{Sotfmax}\left(\frac{\boldsymbol{Q}_{kh}\boldsymbol{W}^S\boldsymbol{W}^{KT}\boldsymbol{S}_{kh}^T}{\sqrt{d_{\text{model}}}}\right)\boldsymbol{S}_{kh}\boldsymbol{W}^{\boldsymbol{V}},\tag{32}$$

其中 $\boldsymbol{W}^{\boldsymbol{Q}}$，$\boldsymbol{W}^{\boldsymbol{K}}$，$\boldsymbol{W}^{\boldsymbol{V}}$ 为线性矩阵，$\boldsymbol{S}_{kh}\in\mathbb{R}^{T\times dh}$ 是 $\boldsymbol{Q}_k$ 的第 h 个头.

3) 前馈神经网络：

$$N=\text{FNN}(\text{concat}(\text{Head}_1,\cdots,\text{Head}_H)\boldsymbol{W}^O),\tag{33}$$

$$N=\text{Avg}(N_K^S),\tag{34}$$

其中：$\boldsymbol{W}^O$ 为线性矩阵；FNN(·)为前馈神经网络；Avg(·)为平均池数，不同会话的自注意力机制共享权重；N 为用户第 k 次会话项目特征.

上述过程可简述为：项目特征向量与带有偏置编码的自注意力网络进行嵌入后输入到多头注意力机制中，此时的输出结果即为提取的项目特征；为得到更稳定精确的结果，将其输入到残差网络和归一化网络中，得到最后的项目特征 N.

2.6 输出层

对用户长短期偏好和项目特征偏好建模后，将建模得到的结果长短期偏好 M 和项目特征偏好 N 进行拼接，输入全连接层，最后输出对用户偏好的预测：

$$Y=\text{concat}(M,N).\tag{35}$$

3 实验及分析

3.1 实验设置

3.1.1 数据集

本文在两个公开的数据集上进行实验，这两个数据集分别是电影领域的 MovieLens[36] 和大众点评

的 Yelp. MovieLens 是一个被广泛使用的电影推荐基准数据集[37]，该数据集包含多个电影评分数据集. 本文实验采用 MovieLens-1M 版本. Yelp 是美国一个著名的商户点评网站，该网站有许多不同领域的商家，如餐馆、购物中心、酒店、旅游等. 在 Yelp 网站，用户除可以交流购物体验外，还可以对商家进行一个整体评价，以便其他用户在选择时有一个参考. 实验数据集信息列于表 1.

表 1　数据集信息

Table 1　Information of dataset

数据集	用户数量	物品数量	交互记录数量	平均长度	稀疏程度/%
MovieLens-1M	6 040	3 900	1 000 209	165.60	95.16
Yelp	23 695	27 927	284 104	12	99.95

3.1.2　实验环境及参数

本文实验采用的操作系统是 Windows11，显卡型号是 RTX 3090(24 GB)，14 vCPU Intel(R) Xeon(R) Gold 6330 CPU @ 2.00 GHz，Python 版本为 3.8，深度学习框架为 PyTorch 1.10.0，在 Pycharm 环境下进行实验. 在数据集 MovieLens-1M 下设置学习率为 0.000 01，数据集 Yelp 下设置学习率为 0.000 01. 实验参数设置如下：训练的最大轮数为 500，批训练大小为 258，嵌入向量维度为 64，垂直水平卷积层数为 16，MLP 层数为 1，注意力头数为 2，优化器选为 Adam，学习率设为 0.000 01，池化类型为 Mean，丢失率为 0.2.

3.1.3　评价指标

在推荐系统中，目前的评价指标多达十余种，本文采用其中几种评价指标进行模型评估，以捕获用户偏好. 实验中将数据集划分为训练集、验证集和测试集，其比例为 8∶1∶1[38-40]. 采用的评价指标为召回率(Recall@K)、平均倒数排名(MRR@K)和归一化折损累积增益(NDCG@K)[7,15,21,39]，其中 K 是每次推荐的项目数，本文将其取为 5,10.

1) 召回率(Recall)[34]表示正确预测出正样本占实际样本的概率：

$$\mathrm{Recall}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{36}$$

其中 TP,FN 是混淆矩阵中的元素. 在混淆矩阵中，每行表示待预测物品的实际值，每列表示待预测物品的预测值. TP 表示真正例，指正样本被判定为正例的数目；FN 表示假反例，指正样本被判定为负例的数目.

2) 平均倒数排名(MRR)是根据正确检索在所有检索结果中的排名评估检索系统的性能：

$$\mathrm{MRR}=\frac{1}{Q}\sum_{q=1}^{Q}\frac{1}{\mathrm{rank}_q}. \tag{37}$$

3) 归一化折损累积增益(NDCG)是将每个推荐结果相关性的得分值累加后作为整个推荐系统列表的得分. NDCG 在评价过程中考虑更多的是所推荐的项目是否出现在用户更容易关注到的位置，该评价指标更多强调是推荐列表的顺序性：

$$\mathrm{NDCG}=\sum_{k=1}^{K}\frac{2^{\mathrm{rel}_k}-1}{\log_2(k+1)}. \tag{38}$$

3.2　实验结果及分析

3.2.1　对比模型

为验证本文方法的有效性，将其与以下具有表示性的基线模型进行比较.

1) POP：该模型根据互动次数判断物品受欢迎的程度.

2) Caser[16]：该模型采用水平和垂直两种方式的卷积神经网络建模，进行顺序推荐.

3) BERT4Rec[34]：该模型是一个端到端的推荐模型，将用户的历史行为建模为一个序列.

4) GRU4Rec[9]：该模型使用基于 rank loss 的 GRU 对用户序列进行建模，实现基于会话的推荐.

5) SASRec[24]：该模型将 Transformer 的优势用到序列推荐以捕获用户的顺序行为.

6) NextItNet[31-32]：该模型由多个卷积层堆叠而成，可在不依赖池化操作的情况下有效增加感

受野，具有在历史序列中进行长范围依赖的建模能力.

3.2.2 结果分析

为证明本文模型的有效性，将本文方法与其他对比方法在 MovieLens-1M 和 Yelp 两个公共数据集基于评价指标 NDCG@5，NDCG@10，Recall@5，Recall@10 和 MRR@5，MRR@10 进行性能比较，实验结果列于表 2.

表 2 不同模型的实验结果

Table 2 Experimental results of different models

数据集	评价指标	模型						
		POP	Caser	BERT4Rec	GRU4Rec	SASRec	NextItNet	本文
MovieLens-1M	NDCG@5	0.254 2	0.384 0	0.493 8	0.450 6	0.559 4	0.514 8	0.585 6
	NDCG@10	0.273 7	0.426 6	0.526 1	0.493 7	0.598 3	0.551 9	0.621 4
	Recall@5	0.582 6	0.593 5	0.657 9	0.613 4	0.713 6	0.677 2	0.734 3
	Recall@10	0.555 6	0.604 6	0.770 5	0.747 8	0.821 5	0.791 9	0.832 9
	MRR@5	0.301 8	0.368 4	0.440 0	0.397 1	0.508 2	0.462 4	0.537 0
	MRR@10	0.332 7	0.352 7	0.449 9	0.414 9	0.528 4	0.477 4	0.555 1
Yelp	NDCG@5	0.285 7	0.475 4	0.253 7	0.503 7	0.615 9	0.567 7	0.616 8
	NDCG@10	0.334 8	0.515 5	0.294 3	0.536 2	0.644 3	0.602 5	0.645 5
	Recall@5	0.410 6	0.632 7	0.361 2	0.712 8	0.766 0	0.724 2	0.768 2
	Recall@10	0.562 6	0.756 0	0.487 3	0.834 6	0.856 2	0.831 7	0.858 5
	MRR@5	0.244 6	0.423 1	0.218 3	0.483 2	0.565 9	0.515 6	0.566 5
	MRR@10	0.264 8	0.439 8	0.235 0	0.497 2	0.577 4	0.530 1	0.578 4

由表 2 可见：

1) 在两个数据集上，基于顺序推荐的方法(如 GRU4Rec，SASRec 和本文模型)优于非序列推荐(如 POP)，表明顺序因素在推荐中具有重要作用.

2) 基于深度学习的方法(Caser，BERT4Rec，GRU4Rec，SASRec，NextItNet 和本文模型)通常优于传统的方法(POP). 因为深度学习方法强大的特征提取能力可捕获序列中复杂的关系，面对大量的数据，深度学习基于 GPU 的训练方法加快了训练时间，在数据集 Yelp 中交互为 30 万条，数据集 MovieLens-1M 交互为 100 多万条.

3) 基于双向编码的模型(BERT4Rec，本文模型)优于单项编码的模型(GRU4Rec，Caser)，在数据集 MovieLens-1M 上 BERT4Rec 均高于 Caser 和 GRU4Rec，但在数据集 Yelp 上优势则不明显甚至弱于 Caser 和 GRU4Rec，这可能是因为数据集 Yelp 相比于数据集 MovieLens-1M 有数据稀疏问题，而 BERT4Rec 因其双向架构的特点对长序列较有优势，而本文模型在两个数据集上均显示出优势，说明基于双向编码的模型优于单项编码的模型.

4) 特征信息的提取有助于推荐性能的提升，本文选取的对比模型中均无对特征信息的提取，与本文模型相比最后推荐精度在评价指标上存在一定差距，证明了本文进行特征提取的有效性.

5) 本文模型与 Caser，GRU4Rec，SASRec 相比性能更好，在这 3 个模型中采用不同的方式对短期偏好建模，在短期偏好建模中有一定优势，但对整体长短期的偏好建模效果却不理想，因此本文采用长期、短期偏好分别建模，实验结果证明了本文方法的可行性.

6) 本文模型与 BERT4Rec，NextItNet 相比，BERT4Rec 采用双向编码方式，能关注到长序列的偏好，NextItNet 模型使用自我注意力度量学习提取特征，网络结构由多个卷积层堆叠而成，可在不依赖于池化操作的情况下有效增加感受野，具备在历史序列中进行长范围依赖的建模能力，但却忽略了短期偏好的建模，使模型整体性能较低. 本文采用长期、短期分别建模，实验结果证明了本文方法的可行性.

实验结果表明，本文模型在数据集 MovieLens-1M 和 Yelp 上性能有很大提升，验证了本文模型的有效性，也证明长短期偏好建模的重要性.

3.2.3 消融实验

消融实验的目的是验证本文方法的有效性和创新性，本文在相同数据集下通过去除各种模块，对

实验结果进行对比验证，实验结果列于表 3.

表 3　去除各种模块后的消融实验结果

Table 3　Results of ablation experiments after removing various modules

模型	MovieLens-1M			Yelp		
	NDCG@10	Recall@10	MRR@10	NDCG@10	Recall@10	MRR@10
无稀疏网络	0.603 6	0.822 3	0.539 7	0.630 4	0.842 8	0.567 3
无 α-entmax 函数	0.610 4	0.825 7	0.542 4	0.635 6	0.848 0	0.569 7
无垂直水平卷积	0.602 3	0.814 6	0.537 5	0.623 9	0.837 6	0.553 5
无偏置编码的特征提取	0.620 4	0.827 5	0.552 0	0.643 1	0.853 2	0.574 3
本文	0.621 4	0.832 9	0.555 1	0.645 5	0.858 5	0.578 4

由表 3 可见：

1）在数据集 Yelp 上的实验结果效果略比数据集 MovieLens-1M 上的效果好.

2）去除稀疏网络模块时，在数据集 MovieLens-1M 上评价指标 NDCG@10，Recall@10，MRR@10 分别下降了 1.78，1.06，1.54 个百分点；在数据集 Yelp 上评价指标 NDCG@10，Recall@10，MRR@10 分别下降了 1.51，1.57，1.11 个百分点. 在两个数据集上评价指标均呈下降趋势，可见稀疏网络模块的重要性.

3）去除 α-entmax 函数时，在数据集 MovieLens-1M 上评价指标 NDCG@10，Recall@10，MRR@10 分别下降了 1.1，0.72，1.27 个百分点；在数据集 Yelp 上评价指标 NDCG@10，Recall@10，MRR@10 分别下降了 0.99，1.05，0.87 个百分点. 可见 α-entmax 函数在稀疏网络中的重要性，进而证明稀疏模块对本文模型的有效性.

4）去除垂直水平卷积模块时，在数据集 MovieLens-1M 上评价指标 NDCG@10，Recall@10，MRR@10 分别下降了 1.91，1.83，1.76 个百分点；在数据集 Yelp 上评价指标 NDCG@10，Recall@10，MRR@10 分别下降了 2.16，2.09，2.49 个百分点. 由实验数据可见，垂直水平卷积模块在整体模型中占有重要地位，无论是在数据集 MovieLens-1M 上还是在数据集 Yelp 上，去掉该模块后影响比其他模块更大，从而验证了在长短期偏好建模时将长期偏好与短期偏好分别建模的重要性，也验证了该模块的有效性.

5）去除偏置编码的多头注意力网络时，在数据集 MovieLens-1M 上评价指标 NDCG@10，Recall@10，MRR@10 分别下降了 0.1，0.54，0.31 个百分点；在数据集 Yelp 上评价指标 NDCG@10，Recall@10，MRR@10 分别下降了 0.24，0.53，0.41 个百分点. 由实验数据可见，该模块对整体模型的影响相对较小，但也起到了部分提升作用. 该模型中使用的偏置编码仅在位置编码上进行微小调整，并未牺牲速度作为代价换取效率的提升，验证了该模块的有效性.

3.2.4　参数对模型的影响

在实际应用中，推荐系统不可能是一个单独存在的个体，推荐系统常与人们所处的大环境有不可避免的交互. 此外，推荐系统本身也不是孤立的，所构成推荐系统的每一部分都会对其产生影响，例如用户因素、项目因素、数据因素、算法策略因素等. 这里主要分析参数在实验中的影响. 在实验中，会有很多因素影响模型最后的效率，但参数对模型的影响是本文研究的主要对象. 本文取序列最大长度分别为 5，10，20，30，50，70 作为主要研究对象，实验结果列于表 4. 图 6 为本文模型在不同交互长度下的对比实验结果.

由表 4 可见，在数据集 Yelp 上的整体效果比数据集 MovieLens-1M 上的效果稍好；在数据集 Yelp 上，当交互序列长度为 50 时，本文模型性能最优；在数据集 MovieLens-1M 上，当交互长度为 30 时，本文模型性能最优. 在两个数据集上，交互序列为 5 时本文模型性能最差，可见如果仅考虑短期偏好，则无法取得准确的推荐. 当交互序列为 70 时，本文模型在两个数据集上都有一个评价指标最高，但其他评价指标不是最优，原因是在过长的交互序列中会存在噪声干扰，进而影响推荐效果. 而本文考虑二者之间的平衡，能均衡长短期的偏好.

表4 不同长度序列对实验结果的影响

Table 4 Effect of different length sequences on experimental results

序列长度	MovieLens-1M			Yelp		
	NDCG@10	Recall@10	MRR@10	NDCG@10	Recall@10	MRR@10
5	0.620 5	0.824 3	0.556 8	0.622 0	0.843 9	0.552 2
10	0.624 6	0.829 6	0.559 8	0.631 3	0.848 9	0.562 6
20	0.626 5	0.830 1	0.562 1	0.635 5	0.854 8	0.565 8
30	0.628 4	0.834 8	0.563 3	0.643 4	0.858 1	0.575 4
50	0.621 4	0.832 9	0.555 1	0.645 5	0.858 5	0.578 4
70	0.625 2	0.835 3	0.559 3	0.645 1	0.860 9	0.576 7

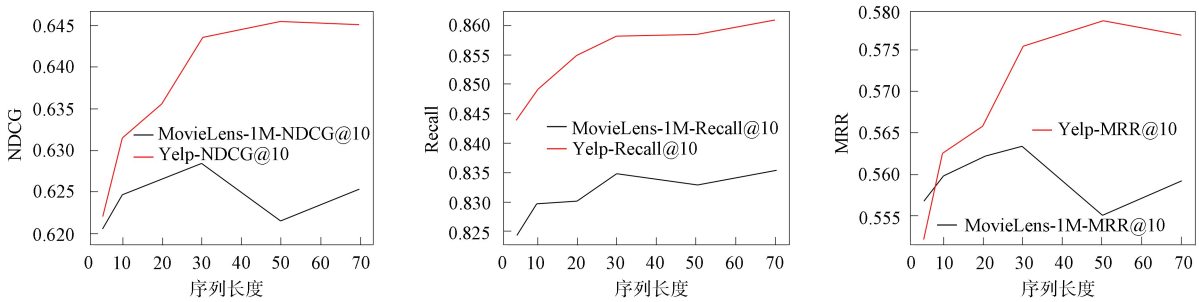

图6 本文模型在不同交互长度下的对比实验结果

Fig. 6 Comparative experimental results of proposed model under different interaction lengths

综上所述，针对现有序列推荐模型忽略用户的长期偏好和短期偏好，导致推荐模型不能充分发挥作用，推荐效果不佳的问题，本文提出了一种融合稀疏网络与垂直水平卷积网络对用户长短期偏好建模的序列推荐方法. 在本文模型中将 α-entmax 函数应用到对序列信息的筛选，去掉了无用信息对推荐结果的影响；将垂直水平卷积网络应用到长短期偏好建模，从不同维度进行偏好处理，能更全面地考虑建模的影响因素；将偏置编码应用到特征提取，重视用户特征偏好对推荐结果的影响. 最后，将本文模型与基线模型进行对比，证明其性能优异. 此外，进行消融实验验证了所用模块对性能提升的重要性. 实验结果表明本文模型优于其他对比模型.

参 考 文 献

[1] ZHOU G R, ZHU X Q, SONG C R, et al. Deep Interest Network for Click-through Rate Prediction [C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 1059-1068.

[2] COVINGTON P, ADAMS J, SARGIN E. Deep Neural Networks for Youtube Recommendations [C]// Proceedings of the 10th ACM Conference on Recommender Systems. New York: ACM, 2016: 191-198.

[3] LI Y Q, LIU M, YIN J H, et al. Routing Micro-videos via a Temporal Graph-Guided Recommendation System [C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM, 2019: 1464-1472.

[4] AN M X, WU F Z, WU C H, et al. Neural News Recommendation with Long- and Short-Term User Representations [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. [S. l.]: ACL, 2019: 336-345.

[5] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-Based Collaborative Filtering Recommendation Algorithms [C]//Proceedings of the 10th International Conference on World Wide Web. New York: ACM, 2001: 285-295.

[6] HU Y F, KOREN Y, VOLINSKY C. Collaborative Filtering for Implicit Feedback Datasets [C]//2008 Eighth IEEE International Conference on Data Mining. Piscataway, NJ: IEEE, 2008: 263-272.

[7] PAN R, ZHOU Y H, CAO B, et al. One-Class Collaborateve Filtering [C]//2008 Eighth IEEE International Conference on Data Mining. Piscataway, NJ: IEEE, 2008: 502-511.

[8] HIDASI B, KARATZOGLOU A, BALTRUNAS L, et al. Session-Based Recommendations with Recurrent

Neural Networks [EB/OL]. (2015-11-21)[2022-02-15]. https://arxiv.org/abs/1511.06939.

[9] JANNACH D, LUDEWIG M. When Recurrent Neural Networks Meet the Neighborhood for Session-Based Recommendation [C]//Proceedings of the Eleventh ACM Conference on Recommender Systems. New York: ACM, 2017: 306-310.

[10] WU C Y, AHMED A, BEUTEL A, et al. Recurrent Recommender Networks [C]//Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. New York: ACM, 2017: 495-503.

[11] YING H C, ZHUANG F Z, ZHANG F Z, et al. Sequential Recommender System Based on Hierarchical Attention Network [C]//International Joint Conference on Artificial Intelligence. New York: ACM, 2018: 3926-3932.

[12] ZHONG E H, LIU N, SHI Y, et al. Building Discriminateve User Profiles for Large-Scale Content Recommenddation [C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2015: 2277-2286.

[13] ZHOU G, MOU N, FAN Y, et al. Deep Interest Evolution Network for Click-through Rate Prediction [C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2019: 5941-5948.

[14] CHENG C, YANG H Q, LÜ M R, et al. Where You Like to Go Next: Successive Point-of-Interest Recommendation [C]//Twenty-Third International Joint Conference on Artificial Intelligence. New York: ACM, 2013: 2605-2611.

[15] RENDLE S, FREUDENTHALER C, SCHMIDT-THIEME L. Factorizing Personalized Markov Chains for Next-Basket Recommendation [C]//Proceedings of the 19th International Conference on World Wide Web. New York: ACM, 2010: 811-820.

[16] TANG J X, 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. New York: ACM, 2018: 565-573.

[17] CHEN X, XU H T, ZHANG Y F, et al. Sequential Recommendation with User Memory Networks [C]// Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. New York: ACM, 2018: 108-116.

[18] HUANG J, ZHAO W X, DOU H J, et al. Improving Sequential Recommendation with Knowledge-Enhanced Memory Networks [C]//Proceedings of SIGIR. New York: ACM, 2018: 505-514.

[19] YU J J, ZHU T Y. Combining Long-Term and Short-Term User Interest for Personalized Hashtag Recommendation [J]. Frontiers of Computer Science, 2015, 9(4): 608-622.

[20] DZMITRY B, KYUNGHYUN C, YOSHUA B. Neural Machine Translation by Jointly Learning to Align and Translate [EB/OL]. (2014-09-01)[2023-01-15]. https://arxiv.org/abs/1409.0473.

[21] WANG H, WANG N Y, YEUNG D Y. Collaborative Deep Learning for Recommender Systems [C]// Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2015: 1235-1244.

[22] LI J, REN P J, CHEN Z M, et al. Neural Attentive Session-Based Recommendation [C]//Proceedings of CIKM. New York: ACM, 2017: 1419-1428.

[23] LIU Q, ZENG Y F, MOKHOSI R, et al. STAMP: Short-Term Attention/Memory Priority Model for Session-Based Recommendation [C]//Proceedings of KDD. New York: ACM, 2018: 1831-1839.

[24] KANG W C, McAULEY J. Self-attentive Sequential Recommendation [C]//2018 IEEE International Conference on Data Mining (ICDM). Piscataway, NJ: IEEE, 2018: 197-206.

[25] 普洪飞，邵剑飞，张小为，等. 融合动态兴趣偏好与特征信息的序列推荐 [J]. 云南大学学报(自然科学版)，2022，44(4)：708-717. (PU H F, SHAO J F, ZHANG X W, et al. Sequential Recommendation by Fusing Dynamic Interest Preference and Feature Information [J]. Journal of Yunnan University (Natural Science Edition), 2022, 44(4): 708-717.)

[26] 胡胜利，林凯. 融合时间上下文与长短期偏好的序列推荐模型 [J]. 湖北民族大学学报(自然科学版)，2022，40(3)：328-335. (HU S L, LIN K. Sequential Recommendation Model Integrating Temporal Context and Long- and Short-Term Preferences [J]. Journal of Hubei University for Nationalities (Natural Science Edition), 2022,

40(3): 328-335.)

[27] HIDASI B, KARATZOGLOU A, BALTRUNAS L, et al. Session-Based Recommendations with Recurrent Neural Networks [EB/OL]. (2015-11-21)[2023-02-11]. https://arxiv.org/abs/1511.06939.

[28] ZHENG L, NOROOZI V, YU P S. Joint Deep Modeling of Users and Items Using Reviews for Recommendation [C]//Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. New York: ACM, 2017: 425-434.

[29] KARPATHY A, TODERICI G, SHETTY S, et al. Large-Scale Video Classification with Convolutional Neural Networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 1725-1732.

[30] SMIRNOV E A, TIMOSHENKO D M, ANDRIANOV S N. Comparison of Regularization Methods for Imagenet-Classification with Deep Convolutional Neural Networks [J]. Aasri Procedia, 2014, 6: 89-94.

[31] KIM Y. Convolutional Neural Networks for Sentence Classification [EB/OL]. (2014-08-25)[2023-02-10]. https://arxiv.org/abs/1408.5882.

[32] DE SOUZA P M G, FERREIRA F, DA CUNHA A M. News Session-Based Recommendations Using Deep Neural Networks [C]//Proceedings of the 3rd Workshop on Deep Learning for Recommender Systems. New York: ACM, 2018: 15-23.

[33] ZHAO C C, YOU J G, WEN X X, et al. Deep Bilstm Networks for Sequential Recommendation [J]. Entropy, 2020, 22(8): 870-1-870-14.

[34] SUN F, LIU J, WU J, et al. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York: ACM, 2019: 1441-1450.

[35] 孙淑娟, 过弋, 钱梦薇. 融合上下文信息的个性化序列推荐深度学习模型 [J]. 小型微型计算机系统, 2021, 42(6): 1121-1128. (SUN S J, GUO Y, QIAN M W. Deep Learning Model Based on Contextualized Personalized Sequence Recommendation [J]. Journal of Chinese Computer System, 2021, 42(6): 1121-1128.)

[36] HARPER F M, KONSTAN J A. The MovieLens Datasets [C]//ACM Transactions on Interactive Intelligent Systems. New York: ACM, 2015: 1-19.

[37] CHO E, MYERS S A, LESKOVEC J. Friendship and Mobility: User Movement in Location-Based Social Networks [C]//Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2011: 1082-1090.

[38] LIU D R, LAI C H, LEE W J. A Hybrid of Sequential Rules and Collaborative Filtering for Product Recommendation [J]. Information Sciences, 2009, 179(20): 3505-3519.

[39] YUAN Q, CONG G, SUN A X. Graph-Based Point-of-Interest Recommendation with Geographical and Temporal Influences [C]//Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. New York: ACM, 2014: 659-668.

[40] ZHAO S L, ZHAO T, YANG H Q, et al. STELLAR: Spatial-Temporal Latent Ranking for Successive Point-of-Interest Recommendation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2016: 315-321.

(责任编辑: 韩 啸)