

DOI: 10.3979/j.issn.1673-825X.202007280230

基于混合神经网络的序列推荐算法

刘纵横,汪海涛,姜 瑛,陈 星

(昆明理工大学 信息工程与自动化学院,昆明 650500)

摘 要: 为了提高推荐算法的推荐性能,在序列建模过程中,针对循环神经网络(recurrent neural network, RNN)无法并行运算导致建模速度与准确度较低,以及在偏好预测过程中对用户不同阶段偏好没有动态融合的问题,提出了一种基于混合神经网络的序列推荐算法。在算法模型的用户交互序列建模阶段,考虑到用户近期偏好变化频繁,对于时间效率与推荐准确度都有更高的要求,引入时间卷积网络(temporal convolutional network, TCN)对近期交互序列进行建模,解决了循环神经网络建模速度和准确度较低的问题;在用户偏好预测阶段,在考虑用户近期与长期偏好的基础上,基于注意力机制动态融合了用户近期与长期 2 个交互阶段的偏好,从而提高了推荐的性能。在公共数据集 MovieLens10M 与 LastFM 上进行了实验,结果证明了模型的有效性。

关键词: 序列推荐; 卷积神经网络; 长短期记忆网络; 注意力机制

中图分类号: TP391

文献标志码: A

文章编号: 1673-825X(2021)03-0466-09

Sequence recommendation algorithm based on a hybrid neural network

LIU Zongheng, WANG Haitao, JIANG Ying, CHEN Xing

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, P. R. China)

Abstract: In order to improve the recommendation performance of the recommendation algorithm, in the sequence modeling process, a hybrid neural network based sequence recommendation algorithm is proposed to solve the problem that the recurrent neural network (RNN) cannot operate in parallel, resulting in low modeling speed and accuracy, and there is no dynamic fusion of user preferences at different stages in the process of preference prediction. In the user interaction sequence modeling stage, considering the frequent changes in user preferences in recent times, there are higher requirements for time efficiency and recommendation accuracy, so the temporal convolutional network is introduced to model recent interaction sequences, which solves the problem of low speed and accuracy of recurrent neural network modeling. In the user preference prediction stage, considering the user's short-term and long-term preferences, based on the attention mechanism, the user's preferences in the recent and long-term interaction phases are dynamically integrated, thereby improving the performance of the recommendation. This paper conducted experiments on the public data sets MovieLens10M and LastFM, and the results proved the effectiveness of the model.

Keywords: sequence recommendation; convolutional neural network; long short-term memory networks; attention mechanism

收稿日期: 2020-07-28 修订日期: 2021-04-28 通讯作者: 汪海涛 kmwht@163.com

基金项目: 国家自然科学基金(61462049)

Foundation Item: The National Natural Science Foundation of China (61462049)

0 引言

推荐系统在互联网中有着广泛的应用,一个优秀的推荐系统可以帮助用户快速、准确地在海量的信息中寻找感兴趣的内容。它改变了人们获取信息的方式,给用户带来了极大的便利。推荐系统目前主要可以分为2类:常规推荐系统与序列推荐系统。常规推荐系统的目的是学习用户的静态偏好,静态偏好主要通过统计获得而且更新缓慢。基于因子分解的协同过滤系统^[1]就是这个领域中较有影响力的算法。但是常规推荐系统无法捕获到用户的兴趣变化,现阶段用户行为偏好具有内在的可变性,在一定时间段内会受到多种因素变化的影响,例如需求的变化、兴趣的变化等。相比于常规推荐系统,序列推荐系统可以建模用户的交互序列,捕获到用户兴趣的动态变化,已经成为推荐系统研究领域的热点方向。

在序列推荐系统的建模阶段,循环神经网络(recurrent neural network, RNN)及其变种网络占有主流地位^[2],但是RNN在对用户序列建模过程中依赖于过去的隐藏状态,无法对网络进行并行运算,所以建模速度是RNN网络的劣势。与RNN相比,时间卷积网络^[3](temporal convolutional network, TCN)不依赖于整个过去的隐藏状态,可以充分利用序列进行并行计算。因此,它在训练效率和预测性能上相较于RNN有一定的优势。受时间卷积网络在序列建模中有不错表现的启发,本文将时间卷积网络引入用户近期序列建模,用于提高基于RNN的推荐系统的训练效率与预测准确度。

用户的长期偏好反映的是用户较为固定的兴趣,如一直喜欢某种类型的电影,用户的近期偏好则反映的是最近的消费需求。尽管用户的喜好随时间动态变化^[4-5],但是仍会存在一些长期稳定的偏好影响用户的行为。现有方法倾向于将用户的长期偏好与近期偏好相结合^[6],但是2个方面简单的结合是不够的,因为过往的物品序列会隐含用户的下一项选择的倾向性,所以本文提出了一种基于注意力的融合方法来融合特定情境下用户的长期偏好和近期偏好。例如,下一个交互项目在上一个交互项目之后较短时间发生,则近期的用户交互可能在预测中起主要作用,否则,应该将更多的权重分配到用户的长期偏好。

基于以上表述,本文提出了一种基于混合神经

网络模型的序列推荐算法,命名为HNNRec。本文的贡献主要有以下2点。

1) 与先前大部分基于RNN的模型不同,本文在用户的近期偏好建模过程中,将自然语言处理中的时间卷积网络引入到推荐系统,对用户的近期偏好进行建模以提高推荐的效率与准确度。

2) 本文模型在考虑用户近期偏好与长期偏好的基础上进一步动态融合了用户2个阶段的偏好,提高了推荐的性能。

最后,本文在公共数据集MovieLens10M与LastFM上进行了实验,结果也证明了模型的有效性。

1 相关工作

在序列推荐的早期阶段,主要依靠马尔科夫链^[7]和基于特征的矩阵分解^[8-9]方法。与深度学习方法相比,马尔科夫链存在无法对复杂数据建模等缺陷。文献Caser^[10]认为,马尔科夫链方法无法对联合级序列模式进行建模,而且也无法捕获到用户项目序列跳跃行为对于推荐结果的影响。而基于因数分解的方法虽然通过序列项矢量的总和对序列建模,但是没有考虑到用户交互序列顺序对于推荐结果的影响。

与传统的推荐模型相比,深度学习模型在捕捉用户序列演变方面显示出了自己的优势,例如RNN这类网络模型在学习兴趣演变,构建基于会话的推荐系统,下一项推荐等领域显示出自己强大的功能。与基于RNN的推荐算法相比,基于卷积神经网络的推荐算法更具挑战性。Tuan等^[11]设计了一种3D CNN模型用于高维建模,证明了用户序列与元素复杂特征相关联时具有有效性;An等^[12]将交互序列编码为三维向量,并使用2D卷积滤波器学习局部的特征;Tang和Wang等^[10]将水平与垂直2种CNN应用于推荐系统,并且取得了不错的效果,但是它分开学习了用户与项目的嵌入,没有捕获到其中的交互信息,也没有加入上下文信息。

在神经网络中,注意力机制^[13]允许各部分在压缩成单个表达式时以不同方式对最终的预测做出贡献。在推荐算法中,注意力机制在算法模型中学习用户交互序列中项目的不同权重,并且根据用户的交互序列动态调整用户的偏好预测^[14]。Xiao等^[15]介绍了基于深度学习的注意力机制,并提出了注意力因子分解机(attentional factorization machines, AFM)模型。AFM的重要思想是建立一个注意力网

络,并依据此网络推导用户评分列表中每个交互的相对重要程度。最近的一些研究也调查了其在推荐任务中的能力,比如将注意力机制与神经网络结合,弥补了传统 ItemCF 预测准确度低的问题^[16]。Wu 等^[17]提出了一种基于情景的注意力机制,提高了推荐的性能与可解释性,但是并没有对用户进行显式建模;Li 等^[18]基于注意力机制建模了会话中的连续偏好信息,弥补了传统建模方式的不足。

2 模型算法

本节对模型算法进行详细介绍。第1部分是用户交互项目嵌入;第2部分是对用户交互序列建模,预测用户偏好。HNNRec 模型如图1。

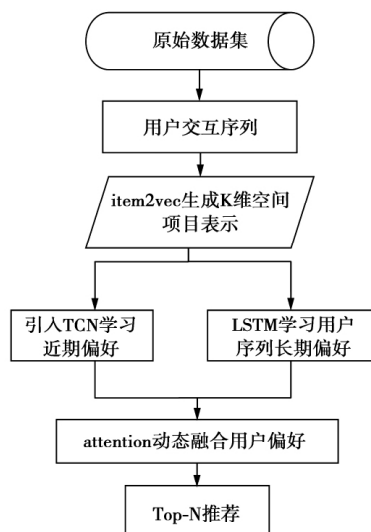


图1 算法整体流程图

Fig.1 Overall flow chart of the algorithm

设 $U=\{u_1, u_2, \dots, u_n\}$ 表示用户集合, $I=\{i_1, i_2, \dots, i_m\}$ 表示项目集合, 则一个用户的交互行为序列可以表示为 $H(u)=\{(i_1^u, t_1^u), (i_2^u, t_2^u), \dots, (i_n^u, t_n^u)\}$, 其中 (i_1^u, t_1^u) 表示用户 u 在 t_1 时间与项目 i_1 产生了交互。序列推荐算法通过捕获用户的交互序列的隐藏信息,最终生成向量 $p=i_{n+1}$ 来预测用户在下次最可能产生交互的项目。

2.1 项目嵌入

项目嵌入的目标是在项目序列中学习交互序列的相似度,生成项目的统一表示。推荐系统之前的项目嵌入是通过 one-hot 编码,然而当面对大量的高维数据时,使用 one-hot 进行编码会导致无法进行较好的优化,在模型的基础上再添加嵌入层又会对推荐模型的性能产生不小的损耗。针对上述问题,本

文引入 item2vec 来进行项目的嵌入。本文使用 $E=\{e_1, e_2, e_3, \dots, e_{n-1}, e_n\}$ 来表示用户交互序列 $G=\{I_1, I_2, I_3, \dots, I_n\}$ 的嵌入向量。

Item2vec 的目标是最大化以下函数

$$\operatorname{argmax} target = \frac{1}{L} \sum_{n=1}^L \sum_{m \neq n}^L \log p(x_m | x_n) \quad (1)$$

(1) 式中: L 表示用户交互序列的长度; x_n 表示序列中的项目; $p(x_m | x_n)$ 为 softmax 函数,此函数定义为

$$p(x_m | x_n) = \frac{\exp(w_n^T v_m)}{\sum \exp(w_n^T v_k)} \quad (2)$$

(2) 式中: w_n 与 v_n 对应 x_n 的目标和前后信息的潜在表示向量。(2) 式训练过程中存在复杂度较高的问题。为了降低梯度下降的复杂度,本文决定使用负采样来代替, (2) 式可转化为

$$p(x_m | x_n) = \sigma(w_n^T v_m) \prod_{k=1}^N \sigma(w_n^T v_k) \quad (3)$$

(3) 式中: $\sigma(x)$ 为 sigmoid 函数 $\sigma(x) = \frac{1}{1+e^x}$; N 表示为负样本数。通过以上运算, (1) 式则转换为

$$\operatorname{argmax} target = \frac{1}{L} \sum_{n=1}^L \sum_{m \neq 1}^L (\log \sigma(w_n^T v_m) + \prod_{k=1}^N \sigma(w_n^T v_k)) \quad (4)$$

之后,本文通过训练 item2vec 得到了目标项目的向量表示。在基于用户的交互序列上, item2vec 可以生成 K 维统一向量,并且为计算项目相似度提供了条件。

2.2 用户交互序列建模

本文将用户的交互序列分为2个阶段:①用户的近期交互序列;②用户的长期交互序列。

2.2.1 近期交互序列建模

受到时间卷积网络的启发,针对用户近期交互序列相较于长期交互序列变化频繁,更需要推荐准确度与时间优势的情况,本文将时间卷积网络引入到序列推荐系统之中进行用户近期偏好建模。卷积建模如图2。

面对序列问题,如果使用普通的 CNN 卷积模型会引起信息泄露,此过程借鉴因果卷积^[19]的思想,当预测项目 y_t 时,不允许卷积滤波器使用 y_t 的信息。在模型中,主要抽象为根据 $x_1 \dots x_t$ 和 $y_1 \dots y_{t-1}$ 去预测 y_t ,使得 y_t 接近于实际值,通过以上方式解决普通卷积模型引起的信息泄露的问题。

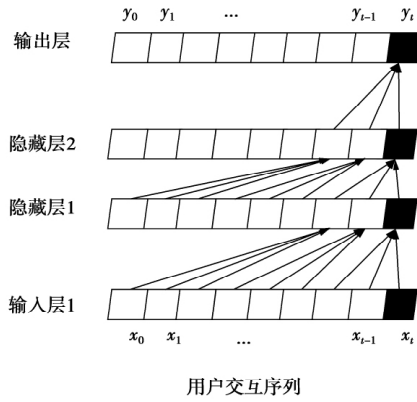


图2 卷积图

Fig.2 Convolution graph

为了获得用户序列项目之间更高级别的特征表示,常见的方法是增加模型中网络的深度,但是会引起梯度消失等问题,这样会使得模型的学习遇到困难。所以本文针对更深层的网络引入了残差学习^[20],将多个模型中卷积层堆叠在一起作为一个块,然后采用跳过连接方案,将前一层的特征信息传递到后一层。跳过连接方案允许显式拟合残差映射,从而维护了输入信息。

残差网络结构如图3,将所需的映射表示为 $H(E)$,本文让残差块适合 $F(E) = H(E) - E$ 的另一个映射。按元素方式将所需的映射转换为 $F(E) + E$ 。优化残差映射 $F(E)$ 比原始的未引用映射 $H(E)$ 容易得多。

残差块包含的分支通过 F 的转换,其输出作为(5)式的输入

$$o = \sigma(X + F(x)) \quad (5)$$

通过以上对于用户近期交互序列的分析,解决了近期偏好获取的问题。

2.2.2 长期交互序列建模

与近期交互序列相比,用户的长期交互序列都较为固定且更新缓慢,所以对时效性的要求较低。在所有基于RNN的模型中,推荐系统常使用长短期记忆网络(long short-term memory, LSTM)^[19]或门控制单元(gate recurrent unit, GRU)来进行用户建模。在本文的数据集中,观察到LSTM的表现略好于GRU,所以引入LSTM进行用户长期偏好建模。建模过程中序列更新过程为

$$h_k = g(x_k W + h_{k-1} U + b) \quad (6)$$

(6)式中: g 是激活函数; x_k 是最新的用户操作; h_{k-1} 是最后一个隐藏状态。(7)式—(11)式为用户序列建模所涉及的结构以及结构的运算过程。

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (7)$$

$$f_t = \sigma(W_f h_{t-1} + U_f h_{t-1} + b_f) \quad (8)$$

$$C_t = f_t c_{t-1} + i_t \tanh(W_c h_{t-1} + W_c x_t + b_c) \quad (9)$$

$$O_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (10)$$

$$h_t = o_t \tanh(c_t) \quad (11)$$

(7)式—(11)式中: W, U 为可训练参数; f_t, i_t, o_t 分别表示忘记、输入和输出门; c_t 表示细胞单元; x_t 表示第 t 个项目的输入嵌入向量; b 是偏置参数。通过以上的网络模型可以为用户的长期交互序列建模,捕捉用户的长期偏好,在这一阶段获得了用户长期交互序列的兴趣点 $p_u^{\text{normal}} = h_t$ 。

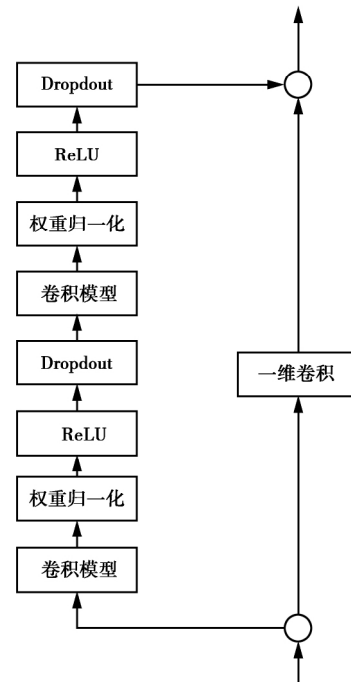


图3 残差结构

Fig.3 Residual structure

2.2.3 自适应融合阶段

单纯的设计用户的偏好为近期偏好与长期偏好相加并不是一个好的思路,因为用户的下一项推荐对于2种阶段偏好存在一定程度的倾向。本节设计了一种方法来融合用户的近期偏好与长期偏好。此方法中用户偏好是否获得更大的权重取决于当时的上下文。举这样一个例子,如果行为B在行为A不久之后发生,那么近期的偏好有大概率会提供更多的信息。基于注意力引出以下公式

$$f(i_h^u, e_j^u) = h^T \text{ReLU}(W(i_h^u e_j^u) + b) \quad (12)$$

(12)式中: $\text{ReLU}(\cdot)$ 是激活函数; e_j^u 代表在上2个阶段求得的用户偏好序列; i_h^u 表示注意力网络的候选项; $f(i_h^u, e_j^u)$ 形式化为前馈神经网络参与模型的训

练; W 与 b 分别代表矩阵的权重与偏置; h^T 代表从隐藏层到注意力权重的映射。

$$\alpha_{ji} = \frac{\exp(f(i_h^u, e_j^u))}{\sum_{j=1}^{|H_u|} \exp(f(i_h^u, e_j^u))} \quad (13)$$

$$\varphi = \sum_{l=1}^{|H_u|} \alpha_{ji} e_l^u \quad (14)$$

(14) 式中 φ 表示注意力网络的候选项与偏好序列中的相关性。当候选项与偏好序列相关程度较低甚至几乎没有相关性时, 则此节点为异常节点, 则应该降低其对推荐的影响。本文使用 softmax 函数将候选项目的权重转换为概率表达式, 使其具有清晰的概率解释。(15) 式表示出系统应该如何平衡近期序列偏好 p_u^{recent} 与长期序列偏好 p_u^{normal} 对于预测的影响。

$$p_u^{\text{final}} = \varphi p_u^{\text{recent}} + (1 - \varphi) p_u^{\text{normal}} \quad (15)$$

(15) 式中: p_u^{normal} 代表在用户长期序列阶段所求得的偏好; p_u^{recent} 代表用户的近期偏好。(14) 式求得的参数 φ 调节了不同阶段偏好在最终预测中所占权重。最后在此阶段获得了下一项推荐。

2.3 模型的训练和优化

本文将用户的交互序列的嵌入作为网络的输入, 通过网络运算输出 K 维的向量。在模型的整体学习阶段使用均方误差作为损失函数, 将此定义为

$$\text{loss} = \frac{1}{L} \sum_{G_u \in L} \frac{1}{|G_u| - l - 1} \sum_{l+1}^{|G_u|} (v_i - \hat{v}_i)^2 \quad (16)$$

(16) 式中: L 代表用户交互序列的大小; l 代表分界长度; v_i 表示用户的下一次交互的项目表示。在实验中使用 AdaGrad 对于损失函数进行优化。至此, 本文已经讨论了 HNNRec 的完整阶段。

2.4 算法描述

在算法模型中, 首先依据时间因素将数据集划分为 2 个子数据集, 再将处理之后的数据分别输入到不同阶段的网络中学习各自阶段的偏好, 之后基于注意力机制对于各阶段生成的预测进行动态融合, 最后输出推荐列表。

算法 基于混合神经网络的序列推荐算法

输入: 数据集 D 。

输出: Top-N 推荐列表。

步骤 1 将数据集 D 依据时间分为 2 部分子数据集 D_1 与 D_2 , D_1 代表长期用户序列数据集, D_2 代表近期用户序列数据集;

步骤 2 数据集 D 通过 item2vec 训练获得用户的全部项目

向量 $E = \{e_1, e_2, e_3, \dots, e_{n-1}, e_n\}$;

步骤 3 将用户的近期项目向量输入到时间卷积神经网络之中, 学习用户的近期偏好, 生成近期偏好预测;

步骤 4 将用户的长期序列向量输入到长短期记忆网络当中, 学习用户的长期偏好, 生成长期偏好预测;

步骤 5 引入注意力机制对于用户的长期偏好和近期偏好进行动态融合, 输出用户的最终偏好预测;

步骤 6 计算预测向量 p_u^{final} 与潜在空间中所有候选项的相似度, 输出推荐列表。

3 实验

本节将会通过 3 项实验来展示混合神经网络模型在下一项推荐中的优势: ①在下一项推荐中将混合神经网络模型的推荐效果与现有的先进算法进行比较, 以证明其在推荐领域有不错的推荐效果; ②对本文动态融合的参数进行分析, 证明了设置动态参数的有效性; ③对用户的交互序列长度进行了分析, 评估本模型与 LSTM, Caser, TransRec 算法在不同序列长度下的表现, 展现出混合神经网络推荐模型相较于其他深度模型在不同用户交互序列上都有着一定的优势。最后, 本文对算法的优点与缺点进行了分析。

3.1 数据集

本文采用 MovieLens10M 与 LastFM 公开数据集作为这次实验的数据集。MovieLens10M 包含了 72 000 个用户对 10 000 部电影的 1 000 万个评级和 100 000 个标签交互。平均交互长度在一百左右。LastFM 是在线音乐系统 last.fm 收集的用户信息, 包含了 951 名用户的 83 382 个交互, 平均交互序列在 90 左右。

为了确保实验的可靠性, 在实验之前, 对数据集进行了下列预处理: ①将交互数量小于 5 的项目过滤掉; ②对数据集采取随机划分的方法, 将数据集的 70% 作训练使用, 30% 作测试使用。本文将评分信息变为隐式反馈信息, 重点讨论隐式反馈。

本文的实验硬件为 CPU Intel Core i7, 16 GB 内存, Nvidia 1060 显卡, 软件环境为 Windows10 系统, 实验编程语言为 Python3, 使用 Keras 框架构建神经网络, 训练模型。

3.2 基线方法

本文将传统的推荐方法和基于深度学习模型的方法与本文提出的算法作对比, 下面是基线方法的描述。

1) Item-based k-NN: 此模型基于项目的相似度,

为用户推荐与历史项目相似的 Top-N 个项目;

2) Exp.Dec. Item-based k-NN^[21]: 此模型通过将指数衰减分量加入 item-KNN 模型, 来降低长时间间隔交互项目的权重;

3) Matrix Factorization (MF): 矩阵分解模型, 将大矩阵分解成小矩阵的乘积, 解决了近邻模型解决不了的问题;

4) Seq.Matrix Factorization^[22]: 在矩阵分解模型的基础上加入用户交互序列信息;

5) LSTM^[23]: 常规的 LSTM 模型, 通过对交互序列的嵌入捕捉用户偏好, 获得对于用户推荐的预测;

6) Caser^[10]: 通过卷积神经网络模拟用户的历史交互, 构建灵活的序列模式来提高用户个性化序列推荐的准确率;

7) TransRec^[24]: 该模型将嵌入思想应用于序列推荐之中, 输出的下一项推荐由用户向量与用户的最近交互向量决定。

3.3 参数设置

在模型嵌入阶段, item2vec 的训练过程中, 将上下窗口大小设置为 10, 嵌入维度设置为 300。在用

户序列建模过程中 Dropout 率设置为 0.1。

3.4 评估指标

本推荐模型推荐项目数量不大, 预测项目应处于候选列表的前列。所以选择以下 3 个指标设为评估指标。

1) Precision: 本次实验选择 Precision 作为评价本文算法的指标, Precision 指标衡量了推荐系统的准确率;

2) Recall: 在本次实验中将召回率作为衡量推荐系统性能的指标之一, 在本文中设定为是测试案例中前 20 个推荐项目中具有所需项目的比例;

3) MRR: 本次实验中将平均倒数排名作为实验的指标之一, MRR 指期望项目倒数排名的平均值。在本次实验中将 20 设置为贡献值。

3.5 实验结果及分析

3.5.1 推荐效果分析

对各算法的推荐性能做了对比。表 1 列出了所有方法的整体性能结果, 这些结果清楚地表明 HN-NRec 在序列推荐任务中获得了不错的准确性。

表 1 本文算法与基线算法性能对比

Tab.1 Performance comparison between this algorithm and the baseline algorithm

数据集	评估指标	Item-based k-NN	Exp.Dec. Item-based k-NN	MF	Seq.MF	LSTM	Caser	TransRec	HNNRec
ML-10m	Precision	0.103 65	0.104 01	0.076 58	0.102 31	0.131 36	0.162 13	0.168 91	0.182 12
	MRR	0.036 38	0.042 14	0.011 93	0.015 51	0.047 32	0.051 07	0.051 16	0.053 25
	Recall	0.121 41	0.128 54	0.077 43	0.107 29	0.157 79	0.182 62	0.185 21	0.211 29
LastFM	Precision	0.095 37	0.126 31	0.115 32	0.101 62	0.133 41	0.159 32	0.169 87	0.196 98
	MRR	0.044 37	0.055 95	0.032 87	0.027 68	0.143 72	0.182 51	0.189 17	0.204 31
	Recall	0.101 25	0.167 25	0.127 22	0.112 23	0.202 65	0.238 64	0.238 90	0.258 91

表 1 中将 HNNRec 算法与传统推荐算法、推荐领域领先的推荐算法进行了比较。从表 1 中可见, 加入时间卷积网络对用户的近期序列进行建模以及动态融合用户的偏好后, 算法的有效性较基线算法存在一定的提升, 由此可见, 本文算法在通过用户交互序列捕捉兴趣变化方面存在一定的优势。与基于向量嵌入翻译空间的 TransRec 模型、基于 CNN 的 Caser 模型和基于 RNN 的 LSTM 模型相比, 在 Precision, MRR 与 Recall 评价指标中都有着一定的提高。这归功于在用户的近期序列阶段引入时间卷积网络进行用户建模, 以及对用户不同交互序列阶段的动态融合。

3.5.2 动态融合的有效性

表 2 与表 3 分别展示了在不同数据集中参数 φ 的取值对推荐效果的影响。当 $\varphi=0$ 时, 表示仅考虑用户的长期交互序列; 当 $\varphi=1$ 时, 表示仅考虑用户的近期交互序列; φ 为固定值时, 表示本文依据以往经验为每个阶段搜索的较为合适的值; φ 为动态融合代表本文的实验思路。

从表 2 和表 3 中可以观察到, 在 2 个公共数据集上仅考虑近期交互序列总是比考虑长期偏好序列表现有优势, 某些时候用户 2 个阶段的简单组合甚至比只考虑近期交互序列表现还要逊色。因此, 证明了动态融合机制的有效性与必要性。

表 2 在 MovieLens10M 数据集上验证融合参数 φ 对推荐算法的影响

Tab.2 Verify the effect of fusion parameters φ on the recommendation algorithm on the MovieLens10M dataset

φ	Precision	MRR	Recall
0	0.131 36	0.047 31	0.157 78
1	0.142 53	0.050 12	0.171 82
固定大小	0.138 79	0.048 01	0.162 31
动态融合	0.182 12	0.053 25	0.211 29

表 3 在 LastFM 数据集上验证融合参数 φ 对推荐算法的影响

Tab.3 Verify the effect of fusion parameters φ on the recommendation algorithm on the LastFM dataset

φ	Precision	MRR	Recall
0	0.133 41	0.143 72	0.202 65
1	0.145 74	0.162 19	0.218 38
固定大小	0.139 71	0.158 01	0.209 31
动态融合	0.196 98	0.204 31	0.258 91

为了进一步分析与证明参数 φ 的有效性,且对 φ 有直观的效果展示,本文将用户距离上一次的行为时间间隔作为自变量,并根据时间间隔的不同划分自变量范围,将 φ 作为因变量。如图 4,时间间隔划分为 4 个级别,由短至长分别为一个小时、一天、一周、一个月。例如横坐标为一小时意味着模型做出的预测结果是在目标用户的最后交互行为最后一个小时内产生的。图 4 展示了此项实验的结果,可以看到有一个明显的趋势,即长期交互序列产生的偏好随时间间隔的增加而显得愈加重要,这与本文模型的假设相符。

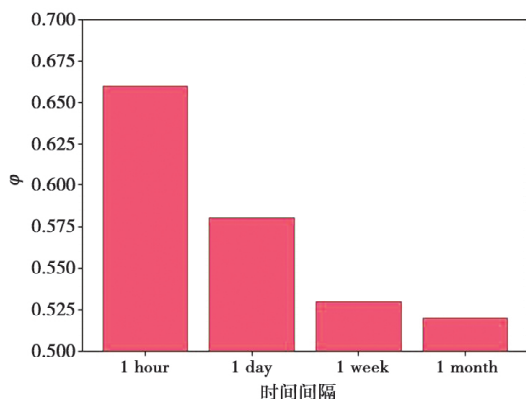


图 4 时间间隔对权重参数 φ 的影响

Fig.4 Effect of time interval on parameter φ

3.5.3 用户历史交互序列长度分析

本文将 HNNRec 模型与 LSTM, Caser 与 TransRec 作对比,进一步分析推荐效果。用到的数据集平均交互长度为 90—100,因此,将用到的数据划分为 3 类:交互序列长度小于 30, 30—60, 60 以上的整数。此目的是衡量复杂的用户历史交互序列长度对 HNNRec 和其他深度模型的影响。

基于表 1 的数据,可以观察到在数据集 MovieLens10M 与 LastFM 中,随着历史记录长度的增加,本文提出的 HNNRec 在 Recall 上有着不错的改进,不仅如此,还可以发现 HNNRec 与另外 3 种序列建模模型都随着交互序列长度的增加而有所改进。在其余的 3 种模型中, LSTM 随着交互序列的加长,性能的增幅逐渐减弱, Caser 与 TransRec 在 30—60 和 60 以上的性能提高较 HNNRec 有所减弱,显示出了 HNNRec 在不同交互长度下性能的鲁棒性。由此可得,用户交互的时间长度确实会对推荐系统的性能产生影响,也清楚地证明了在面对不同长度的交互序列时,通过动态融合用户的不同阶段偏好,可以提高算法的鲁棒性和推荐准确度,从而使得本模型在不同序列下都有着较好的表现。

图 5—图 7 直观展示了 4 种推荐算法在 MovieLens10M 上的推荐效果。

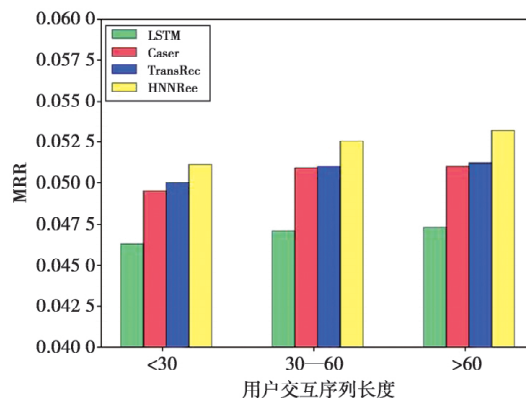


图 5 不同长度的用户交互序列在 MRR 的表现

Fig.5 Performance of user interaction sequences of different lengths in MRR

3.6 算法优缺点分析

本文提出一种基于混合神经网络的序列推荐算法。由上文的多项实验对比可知,本文算法具有较高的推荐准确性,面对不同长度的用户交互序列也有着较强的鲁棒性。较高的推荐准确性得益于时间卷积网络思想的合理引入,但是这也会导致在训练时需要更强的硬件条件。

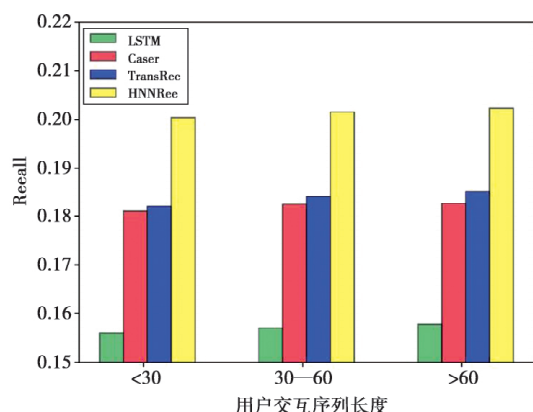


图6 不同长度的用户交互序列在 Recall 的表现

Fig.6 Performance of user interaction sequences of different lengths in Recall

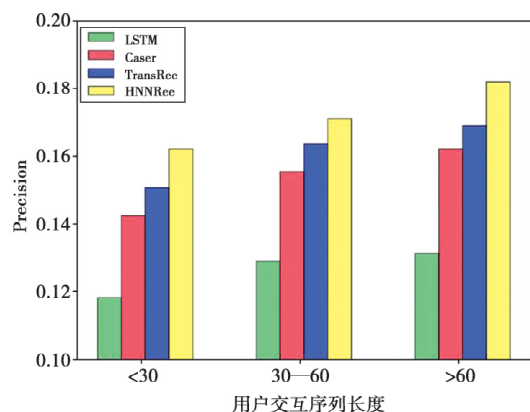


图7 不同长度的用户交互序列在 Precision 的表现

Fig.7 Performance of user interaction sequences of different lengths in Precision

4 结 论

本文提出了一种新颖的 HNNRec 模型。该模型基于用户的近期偏好与长期偏好建模,引入了时间卷积网络对用户项目序列进行建模,提高了推荐准确度,同时观察到用户的行为序列比其他领域序列遇到问题的复杂度更高,因此,进一步提出基于注意力的用户偏好动态融合方法,根据不同条件自适应地融合用户的长期偏好和近期偏好。本文在公共数据集上进行的实验证明了提出的模型具有有效性。在将来,模型的冷启动与模型效率的改进将是下一步探索的方向。

参考文献:

- [1] KOREN Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model [C]//Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2008: 426-434.
- [2] 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.
- [3] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [EB/OL]. (2018-04-19) [2020-07-11]. <https://arxiv.org/abs/1803.01271>.
- [4] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback [EB/OL]. (2012-05-09) [2020-07-08]. <https://arxiv.org/abs/1205.2618>.
- [5] 曾兆伟, 曹健. 数据分析服务流程模型推荐 [J]. 小型微型计算机系统, 2019, 40(7): 1374-1379.
- [6] ZENG Z W, CAO J. Data Analysis Service Process Model Recommendation [J]. Journal of Chinese Computer Systems, 2019, 40(7): 1374-1379.
- [7] VILLATEL K, SMIRNOVA E, MARY J, et al. Recurrent neural networks for long and short-term sequential recommendation [EB/OL]. (2018-07-23) [2020-07-15]. <https://arxiv.org/abs/1807.09142>.
- [8] 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: Association for Computing Machinery, 2010: 811-820.
- [9] 杨阳, 向阳, 熊磊. 基于矩阵分解与用户近邻模型的协同过滤推荐算法 [J]. 计算机应用, 2012, 32(2): 395-398.
- [10] YANG Y, XIANG Y, XIONG L. Collaborative filtering and recommendation algorithm based on matrix factorization and user nearest neighbor model [J]. Journal of Computer Applications, 2012, 32(2): 395-398.
- [11] 王建芳, 刘冉东, 刘永利. 一种带偏置的非负矩阵分解推荐算法 [J]. 小型微型计算机系统, 2018, 39(1): 69-73.
- [12] WANG Jian-fang, LIU Ran-dong, LIU Yong-li. Non-negative Matrix Factorization Algorithm with Bias in Recommender System. Journal of Chinese Computer Systems, 2018, 39(1): 69-73.
- [13] 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: Association for Computing Machinery, 2018: 565-573.
- [14] TUAN T X, PHUONG T M. 3D convolutional networks

- for session-based recommendation with content features [C]//Proceedings of the Eleventh ACM Conference on Recommender Systems. New York: ACM, 2017: 138-146.
- [12] YAN A, CHENG S, KANG W C, et al. CosRec: 2D convolutional neural networks for sequential recommendation [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York: ACM, 2019: 2173-2176.
- [13] RUSH A M, CHOPRA S, WESTON J. A neural attention model for abstractive sentence summarization [EB/OL]. (2015-09-03) [2020-07-08]. <https://arxiv.org/abs/1509.00685>.
- [14] 黄立威, 江碧涛, 吕守业, 等. 基于深度学习的推荐系统研究综述[J]. 计算机学报, 2018, 41(7): 1619-1647.
HUANG L W, JIANG B T, LV S Y, et al. Survey on Deep Learning Based Recommender Systems [J]. Chinese Journal of Computers, 2018, 41(7): 1619-1647
- [15] XIAO J, YE H, HE X N, et al. Attentional factorization machines: Learning the weight of feature interactions via attention networks [EB/OL]. (2017-08-15) [2020-06-08]. <https://arxiv.org/abs/1708.04617>.
- [16] HE X N, HE Z K, SONG J K, et al. Nais: Neural attentive item similarity model for recommendation [J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(12): 2354-2366.
- [17] WU J B, CAI R Q, WANG H N. Déjà vu: A Contextualized Temporal Attention Mechanism for Sequential Recommendation [C]//Proceedings of the Web Conference 2020. New York: Association for Computing Machinery, 2020: 2199-2209.
- [18] LI J, REN P J, CHEN Z M, et al. Neural attentive session-based recommendation [C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. New York: ACM, 2017: 1419-1428.
- [19] OORD A, DIELEMAN S, ZEN H, et al. Wavenet: A generative model for raw audio [EB/OL]. (2016-09-19) [2020-07-10]. <https://arxiv.org/abs/1609.03499>.
- [20] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. New York: IEEE, 2016: 770-778.
- [21] DING Y, LI X. Time weight collaborative filtering [C]//Proceedings of the 14th ACM international conference on Information and knowledge management. New York: ACM, 2005: 485-492.
- [22] KOREN Y. Collaborative filtering with temporal dynamics [C]//Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2009: 447-456.
- [23] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [24] HE R, KANG W C, MCAULEY J J. Translation-based Recommendation: A Scalable Method for Modeling Sequential Behavior [C]//IJCAI. International Joint Conferences on Artificial Intelligence. California, USA: [s.n.], 2018: 5264-5268.

作者简介:



刘纵横(1995),男,山东德州人,硕士研究生,主要研究方向为推荐系统、软件复用。
E-mail: 891726327@qq.com。



汪海涛(1967),女,云南昆明人,副教授,硕士研究生导师,主要研究方向为人工智能、软件复用。E-mail: kmwht@163.com。

(编辑: 王敏琦)