计算机工程
Computer Engineering
ISSN 1000-3428,CN 31-1289/TP

《计算机工程》网络首发论文

题目：门控网络构建用户动态兴趣的序列推荐模型
作者：王燕，范林，赵妮妮
DOI：10.19678/j.issn.1000-3428.0062184
网络首发日期：2021-10-15
引用格式：王燕，范林，赵妮妮．门控网络构建用户动态兴趣的序列推荐模型[J/OL]．计算机工程. https://doi.org/10.19678/j.issn.1000-3428.0062184

门控网络构建用户动态兴趣的序列推荐模型

王燕，范林，赵妮妮

兰州理工大学 计算机与通信学院，甘肃 兰州 730050

摘　要：序列推荐是推荐系统领域非常重要的一部分，现有的方法将用户行为视为一个时间有序的序列进行用户兴趣建模，但用户兴趣的动态变化导致模型难以从用户行为序列中捕捉准确的用户兴趣信息，同时项目间成对的共现模式应该作为交互信息的补充。基于此，提出了门控网络构建用户动态兴趣的序列推荐模型 DCGN，首先使用门控线性网络对交互序列中的用户兴趣进行捕捉，并使用带有注意力权重的门控循环网络（GRU）学习用户的动态兴趣；然后对用户交互项目间的共现模式进行建模，与用户兴趣信息以及用户信息进行融合后输入深度神经网络（DNN），最后通过 DNN 给出推荐结果。在三个公开数据集上进行实验的结果验证了所提方法的有效性。

关键词：推荐算法；注意力机制；门控线性网络；项目共现模式；动态兴趣

开放科学（资源服务）标志码（OSID）：

Sequential recommendation method for building user's dynamic interest in gated network

WANG Yan, FAN Lin, ZHAO Ni-ni

College of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050,China

【Abstract】Sequence recommendation is a very important part in the field of recommender system. The existing meth-ods regard user behavior as a time ordered sequence for user interest modeling. However, the dynamic change of user interest makes it difficult for the model to capture accurate user interest information from user behavior sequence. At the same time, the paired co-occurrence pattern between items should be used as the supplement of interactive information. Based on this, a sequential recommendation model dcgn is pro-posed, in which the gating network is used to construct the user's dynamic interest. Firstly, the gating linear network is used to capture the user's interest in the interaction sequence, and the gating recurrent network (GRU) with attention weight is used to learn the user's dynamic interest; Then, the co-occurrence model of user interaction items is modeled, which is fused with user interest information and user information and then input to the depth neural network (DNN), and give the recommended results through DNN. Exper-imental results on three public datasets demonstrate the effectiveness of the proposed method.

【Key words】Recommendation algorithm; Attention mechanism; Gating linear network; Project co-occurrence mode; Dynamic interest

DOI:10.19678/j.issn.1000-3428.0062184

基金项目：基于大数据的“人肉搜索”建模、分析与控制，国家自然科学基金（61863025）;无人机关键技术研究，甘肃省重点研发计划-工业类（18YF1GA060）；

作者简介：王燕（1971-），女，教授，硕士生导师，主要研究方向为数据挖掘、推荐系统；范林（通信作者），硕士研究生；赵妮妮，硕士研究生。E-mail: figfavture@163.com

0 概述

互联网的飞速发展和移动设备的大量普及促使信息呈现爆发式增长，信息冗余致使用户无法从海量信息中准确、快速的筛选所需信息，从而导致用户体验不佳。推荐系统很好的解决了这一问题，并且在搜索引擎、电子商务、娱乐等许多网络应用中发挥着非常重要的作用，因此推荐系统的研究具有重要的实用价值及意义。

目前大多数推荐系统模型可分为基于协同过滤（Collaborative Filtering, CF）的推荐模型和混合推荐模型。传统的基于协同过滤推荐模型虽取得了不错的效果，但其基础的线性结构极大地限制了模型的表达能力。神经网络能够通过改变激活函数、选择和组合隐藏层单元来以任意精度近似任何连续函数特性，使得深度神经网络不但能处理复杂的用户交互模式，而且可以很好的拟合用户的偏好。凭借神经网络强大的数据泛化能力，结构复杂的混合推荐模型所表现出的性能远优于传统推荐模型，因此基于深度学习的混合推荐模型是本文的主要研究对象。

用户历史行为序列中包含丰富的用户兴趣信息，精确的捕捉用户兴趣是提升推荐系统推荐准确度的关键。最近有学者提出了对用户兴趣建模的方法，如文献[1]提出了一种基于贝叶斯图卷积神经网络的框架，用于对用户-物品交互的不确定性建模，以更好地描述用户和物品之间的关系，同时建模用户的偏好，但图模型的复杂度较高，提取用户兴趣偏好需要的计算量较大。文献[2]提出了一种新的长短期记忆网络建模用户兴趣的方法（LSTPM），将用户的交互记录看作是一个很长的序列，进行用户长短期兴趣的建模，但该方法在用户交互序列很长时，随着模型的复杂度增加，提取的用户兴趣变得不再准确，且模型将提取到的用户兴趣看作同等重要，这显然不符合现实逻辑。受自然语言处理（Natural Language Processing, NLP）领域新模型 Transformer 中注意力机制（self-attention）方法[3]的启发，研究者们将“注意力机制”应用到推荐系统中，解决了许多模型将用户兴趣项看作同等重要的问题。文献[4]提出了一种新的自适应分层注意力增强门控网络，该网络为了获取可区分的细粒度特征，引入注意力机制层来学习重要的语义信息特征以及这些特征之间的动态联系。文献[5]提出了一种新的基于自注意力的协同神经网络模型，将用户相似度和物品相似度结合起来，利用注意力机制从用户购买历史中多个方面计算物品的权重，进而捕捉用户兴趣信息。

尽管研究者们在捕捉用户兴趣信息方面已经有大量工作，但仍有如下问题需要解决：

1. 将用户历史行为视为按操作时间戳排序的操作序列，并未明确从特征层面捕捉用户兴趣，从特征层面捕捉细粒度的用户兴趣可有效的提升推荐性能；

2.捕捉项目之间的成对关系对用户兴趣的建模同样重要，这种项目间成对的共现模式在推荐系统中很常见，例如：购买电脑后再购买鼠标的可能性比购买鞋子的可能性大；

3.用户兴趣是动态变化的，如何建模用户的动态兴趣变化。

针对上述问题，本文提出了一种 DCGN（Dynamic construction of user interest in gated network）门控网络构建用户动态兴趣的序列推荐模型。本文的主要工作如下：

1. 从用户的历史行为数据中更加有效的捕捉用户兴趣，使用双层门控线性网络从特征的层面捕捉细粒度的兴趣信息，保留用户交互中重要的特征以及交互项目；

2.使用 GRU 对学习到的用户兴趣信息进行聚合，并使用注意力权重对 GRU 的隐藏状态进行动态更新，捕获用户动态变化的兴趣偏好；

3.使用一种双线性特征交叉的方法，对项目对之间的关系进行建模，因为密切相关的项目可能会在用户交互项目序列中接连出现；

4.使用多层全连接网络学习得到最终的输出结果，并在三个真实数据集上进行实验，表明本文的方法优于现有的模型。

1 相关工作

1.1 序列推荐模型

传统的协同过滤模型虽然取得了不错的效果，并且应用广泛，但它以静态的方式拟合用户与项目之间的交互模式，且其捕捉的用户兴趣对于用户来说是一种广义的模式，缺乏个性化。序列推荐模型不同于传统的协同过滤推荐模型，序列推荐模型的构建基于连续的用户-项目交互序列，用户-项目交互序列是一个动态变化的序列，且交互序列前后有着很强的关联性，模型通过拟合用户与物品之间的交互模式，捕捉用户的偏好，并利用用户交互序列中丰富的上下文信息描述用户的意图和消费趋势。

传统的基于马尔可夫链的序列推荐模型[6]，假设用户当前的交互项仅依赖于最近的几个交互项，但其忽略了用户较早的交互项对当前交互项的影响。基于因子分解机的序列推荐[7]将用户与项目的交互分解为用户和项目的隐因子向量，通过参数分解对特征之间的高阶交互进行建模，但模型易受数据稀疏性的影响使得推荐结果不够理想。深度学习技术的兴起，很好

的解决了传统序列推荐模型的不足，带来了令人惊喜的效果。

1.2 基于深度学习的序列推荐模型

利用深度学习技术构建序列推荐模型的常用方法时，首先将用户交互的项目序列转换为低维稠密的嵌入向量进行表达，然后通过一系列的方法得到用户的兴趣表示，这类方法一般使用加和、平均和取最大值的方法融合用户的历史交互信息得到用户的兴趣表达，但模型将用户交互序列中的交互项视为同等重要，结果中只包含了用户部分重要的兴趣，且无法建模用户兴趣的变化。

一些研究人员尝试利用更复杂的模型结构来构建用户兴趣，并提高推荐性能。基于 GRU 的序列推荐[8]通过建模给定的交互序列中的依赖来预测下一个可能的交互项，但其只使用了交互项信息，并未考虑到其他信息对结果的影响。文献[9]提出根据用户偏好自适应地选择项目中有吸引力的潜在特征，然后根据提取到的特征对用户兴趣进行建模。文献[10]基于 RNN 建模用户的交互序列，考虑了用户行为的特征信息，然后使用注意力机制计算每个交互项的权重，得到用户的兴趣向量表示。文献[11]基于 RNN 建模用户交互序列中的微观行为，利用 RNN 对用户行为的特征信息建模，在用户微观行为、商品等多个层次得到兴趣向量。基于 RNN 构建的序列推荐模型存在无法有效建模行为序列中多个行为间关联的问题，为解决此问题，研究者将 NLP 领域的 Transformer 模型应用到推荐系统中，文献[12]将用户的行为序列划分为多个会话，研究者发现每个会话区间内用户的兴趣往往是固定的，然后基于 Transformer 模型建模每个会话内用户的行为序列，再基于 RNN 聚合多个会话内的兴趣信息。为更好的捕捉用户兴趣，文献[13]尝试利用胶囊网络（胶囊网络是应用在图像领域的一种神经网络结构，可挖掘图像中丰富的空间信息）建模用户的行为序列，获得多个用户的兴趣向量，然后使用一个可控的多兴趣聚合模块平衡用户兴趣的多样性与准确性，这一模型的提出，也为序列推荐模型的研究提供了全新的思路。

针对现有序列推荐模型没有充分从特征层面考虑用户交互项之间的联系，以及用户兴趣动态变化和相似交互项连续出现的问题。本文提出一种门控网络构建用户动态兴趣的序列推荐模型，利用门控网络提取了用户兴趣，且通过注意力权重动态的调整信息聚合函数进而获得用户的动态兴趣，同时考虑了项目间成对的联系信息，提升了推荐的准确率。

2 DCGN 模型

这部分将详细介绍门控网络构建用户动态兴趣的模型。模型可以从用户交互序列中学习用户动态变化的兴趣表示，模型的目标是从特征的层面提取用户的兴趣，同时通过挖掘项目对之间的联系提升模型效果。模型结构如图 1 所示。首先，将高维稀疏的用户交互序列、项目特征及用户特征作为模型的输入；其次，通过嵌入层将所有的特征映射到一个低维的空间中，得到低维稠密的嵌入向量；然后，将用户交互序列嵌入向量输入到一个双层的门控网络中，利用门控线性网络 GLU（Gate Linear Unit, GLU）[14]从用户交互的特征层面对用户兴趣进行建模，该层实现了输入序列中项目特征的过滤及保留了那些对用户来说重要的交互项目，即用户兴趣；接着，使用双线性特征交叉方法，学习目标项目对之间的联系，捕捉两个项目之间的共现模式；之后，对门控网络提取到的用户兴趣进行聚合，这部分采用了 GRU 循环门控网络，结合目标项向量与提取到的兴趣向量使用注意力机制得到相应兴趣项的注意力权重，并利用该权重对 GRU 的隐状态进行更新，得到最终的用户动态兴趣表示；最后，通过堆叠多个全连接层来得到最终的预测结果。接下来，将详细介绍模型的细节。

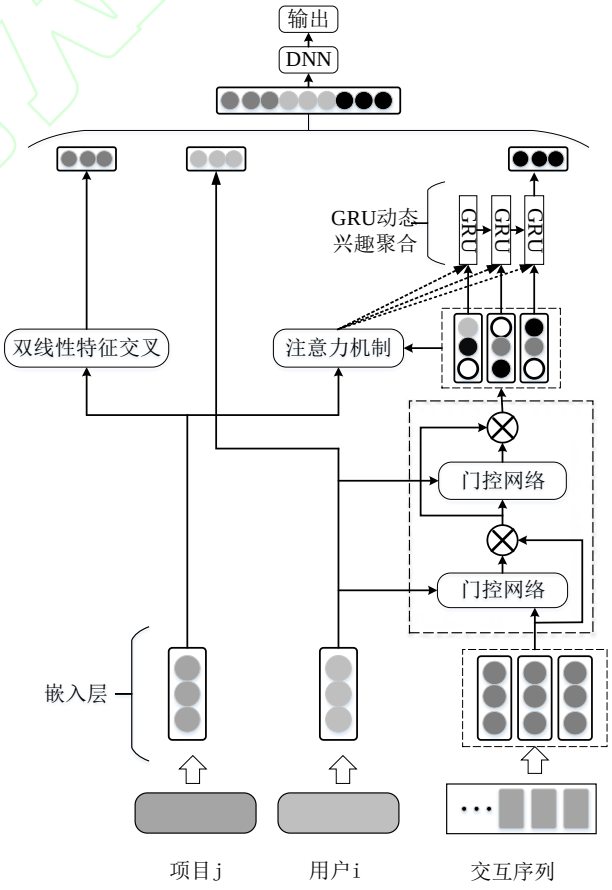


图 1 门控网络构建用户动态兴趣结构图

Fig.1 Construction of user dynamic interest structure diagram in gated network

2.1 嵌入层

为了对序列推荐任务建模，对于每个用户 i 的交互序列按时间顺序表示为 $S_i=\{s_i(1),\cdots,s_i(t),\cdots,s_i(l)\}$，其中 $S_i\in R^{d\times f\times l}$，$l\in R$ 为用户交互序列的长度，$f\in R$ 为交互项的特征数，$s_i(t)\in R^{d\times f}$ 表示用户交互序列中第 t 次购买/评价的项目。将该序列作为输入，解释为在用户-项目交互序列 S_i 中，给定 l 个连续项目，预测其他 N 个项目接下来将被交互的可能性。将交互序列输入到嵌入层转换为低维稠密的向量表示，用户第 t 个交互项的嵌入向量如式(1)所示：

$$s_i(t)=\left(e_1^i,e_2^i,\cdots,e_f^i\right) \tag{1}$$

其中 e_f^i 为交互项 $s_i(t)$ 中特征 f 的嵌入向量，将用户 i 的嵌入向量表示为 $u_i\in R^d$，项目 j 的嵌入向量表示为 $v_j\in R^d$，d 为嵌入向量的维度。

2.2 门控网络

Dauphin 等人提出的门控线性单元(Gate Linear Unit, GLU)[14]在语言建模任务中控制信息传递，用于下一个单词的预测， GLU 的公式为：

$$h_l(x)=(X*W+b)\otimes\sigma(X*V+c) \tag{2}$$

其中 X 为单词的嵌入向量，W,V 为卷积操作中的卷积核，b,c 为偏置参数，σ 是 sigmoid 函数，⊗ 是矩阵之间的元素乘积，$*$ 为卷积操作。公式中的后半部分，即有激活函数的卷积 $\sigma(X*V+c)$，就是所谓的门控机制，其控制 $(X*W+b)$ 中的哪些信息可以传入下一层。

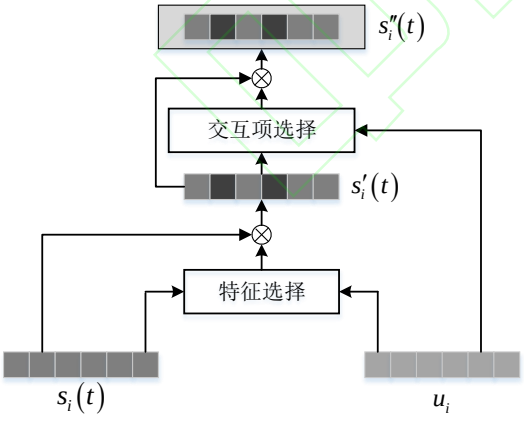


图 2 门控网络结构图

Fig.2 Structure diagram of gating network

将该方法用于用户的行为序列的兴趣提取，结构图如图 2 所示。由于卷积运算中卷积核的特性，导致提取到的特征只是一部分，使得数据中丰富的特征信息丢失，故将卷积运算替换为向量的内积，在运算中保留特征信息。使用两层线性门控网络提取交互序列中的重要特征和与未来交互有关联的项目的提取，即提取用户兴趣。先使用一层门控网络提取适合用户偏好的项目特征：

$$s_i'(t)=s_i(t)\otimes\delta\left(W_1\cdot s_i(t)+W_2\cdot u_i+b\right) \tag{3}$$

其中 $W_1,W_2\in\mathbb{R}^{d\times d}$，$b\in\mathbb{R}^d$ 为可学习的参数，$\otimes$ 为向量的元素积，σ 是 sigmoid 函数，u_i 为用户 i 的向量表示。然后此处采用了不同于使用注意力机制模型的做法，使用上一层门控网络提取到重要的交互特征作为输入，选取交互序列中与预测未来的交互项更相关的项目：

$$s_i''(t)=s_i'(t)\otimes\delta\left(W_3^T\cdot s_i'(t)+u_i^T\cdot W_4\right) \tag{4}$$

其中 $W_3\in\mathbb{R}^d$，$W_4\in\mathbb{R}^{d\times f}$ 为可学习参数，f 为特征数量，$s_i''(t)\in\mathbb{R}^{d\times f}$ 为门控网络最终提取出的用户兴趣，其中用户交互的主要特征和项目已被选择，无关的特征和项目被过滤。

2.3 注意力机制

从数学的角度来看，注意力机制只是对平均操作或加和操作进行改进，换成了加权平均或加权，但正是这种加权方式对于模型的效果有明显提升。从注意力权重计算的一种形式上来看，其与全连接网络的结构颇为相似，但不同的是，注意力权重参数是位置无关的，权重是根据输入的前后信息进行计算的，一旦输入的前后信息发生变化，其权重也会相应的发生变化。这种根据输入的前后信息来计算权重的方式，更好的刻画了输入序列中的重点信息，帮助模型对输入信息进行了区分。计算注意力分数的公式为：

$$g_t^i=avg\left(s_i''(t)\right) \tag{5}$$

$$a_t^i=\frac{\exp\left(g_t^iWv_j\right)}{\sum_{p=1}^{l}\left(g_p^iWv_j\right)} \tag{6}$$

其中 $g_t^i\in\mathbb{R}^d$ 为通过平均池化运算后的用户兴趣向量，v_j 为目标项目的向量，$W\in\mathbb{R}^{d\times d}$ 为可学习参数。

2.4 GRU 动态兴趣聚合

前面通过两层门控线性网络得到了用户兴趣的表示，本节使用 GRU 对提取到的用户兴趣进行聚合，结合注意力机制对用户的动态兴趣进行建模。注意力得分在 GRU 的每一步中都可以增强相关兴趣所起的作用，减弱无关兴趣对总体结果的影响，更好地建模用户对目标项的兴趣变化：

$$z_t^i=\delta\left(W_zg_t^i+p_zh_{t-1}^i+b_z\right) \tag{7}$$

$$r_t^i=\delta\left(W_rg_t^i+p_rh_{t-1}^i+b_r\right) \tag{8}$$

$$h_t^i=\tanh\left(W_hg_t^i+p_h\left(r_t\odot h_{t-1}^i\right)+b^h\right) \tag{9}$$

$$h_t^i=\left(1-z_t^i\right)\odot h_{t-1}^i+z_t^i\odot h_t^i \tag{10}$$

其中 $z_t^i \in \mathbb{R}^{d\times d}$ 与 $r_t^i \in \mathbb{R}^{d\times d}$ 分别为重置门和更新门，W_z, P_z 、W_r, P_r 和 W_h, P_h 为分别为重置门、更新门和输出门的可学习参数，$\delta(\cdot)$ 为 sigmoid 函数，$\odot$ 为点乘运算。隐藏状态 h_t^i 只捕捉了用户兴趣之间的相互依赖关系，并不能有效的表示用户动态变化的兴趣。通过使用注意力分数来控制 GRU 隐藏状态的更新来解决这个问题，保留原始重置门 z_t^i 的信息，注意力分数越小，对隐藏状态的影响越小，使用注意力分数更新 GRU 的隐藏状态：

$$z_t'^i = a_t^i \cdot z_t^i \tag{11}$$

$$h_t'^i = \left(1 - z_t^i\right) \odot h_{t-1}'^i + z_t'^i \odot h_t^i \tag{12}$$

其中 $z_t'^i$ 为注意力重置门，$h_t'^i$ 为最终输出的隐藏状态。

2.5 项目特征交叉与模型预测

项目之间成对关系的学习[15]对于推荐系统来说十分重要，在序列推荐问题中，密切相关的项目有很大概率会在将来的项目交互序列中连续出现。

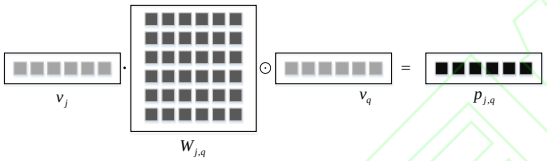


图 3 双线性特征交叉结构图

Fig.3 Bilinear feature crossing structure diagram

由于传统元素积的形式难以有效地对稀疏数据进行特征交叉建模，且模型的表达能力不强，为捕捉这种项目之间的共现模式，使用双线性交叉函数进行学习，其结构图如图 3 所示：

$$p_{j,q} = v_j \cdot W_{j,q} \odot v_q \tag{13}$$

其中 $W_{j,q} \in \mathbb{R}^{d\times d}$ 为参数矩阵，$\cdot$ 为内积运算，$\odot$ 为哈达玛积，$j,q \in (1,2,3,\cdots,m)$，m 为目标项目数量，对项目 j 与项目 q 特征交叉结果 $p_{j,q} \in \mathbb{R}^d$ 通过平均池化聚合为当前项目的向量表示 $p_j = avg\left(p_{j,q}\right)$，$p_j \in \mathbb{R}^d$。将 $x_i = Concat(p_j, h_t'^i, u_i)$ 向量输入到 MLP 中得到最终的预测结果：

$$a^{(l)} = \delta\left(W^{(l)} a^{(l-1)} + b^{(l)}\right) \tag{14}$$

其中 $a^{(0)}$ 为网络的输入 x_i，l 为网络的深度，δ 为 ReLU 激活函数，$W^{(l)}$，$b^{(l)}$ 为可学习参数，$a^{(l)}$ 为第 l 层的输出，最后一层 L 的输出输入到 sigmoid 函数得到最终的预测结果：

$$y = \delta\left(W^{L+1} a^L + b^{L+1}\right) \tag{15}$$

其中 δ 为 sigmoid 函数，a^L 为最后一层网络的输出，W^{L+1}, b^{L+1} 为可学习的参数。通过最小化下面的目标函数，进行模型的学习：

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log\left(y_i\right) + \left(1 - y_i\right)\log\left(1 - y_i\right)\right) \tag{16}$$

其中 y_i 为训练样本 i 的真实值，y_i 为预测值，N 为训练样本的数量。

3 实验

3.1 实验数据集

采用三个数据集即电影评分数据集（ML100K）[16]、亚马逊电子商务数据集（Amazon ecommerce dataset）1 和电子商务网站行为数据（Retailrocket）2 进行建模。Retailrocket 由一家个性化电子商务公司发布，包含六个月的用户浏览记录；ML100k 数据集包含了用户编号、电影编号、电影评分、时间戳及电影的相关信息，且该数据集已经过清洗；亚马逊电子商务数据集包含了多种类型（如：CD 销售数据、图书销售数据等）的数据，这里采用电子产品类的数据集(Amazon 5-Elect)，其中包含多个字段，本次实验仅使用商品编号、用户编号、产品评分和 Unix 格式时间戳字段。每个用户编号、商品编号都是唯一的，对每个用户交互的商品编号，根据数据字段中的时间戳进行排序，得到用户对应的交互序列。为了消除噪声数据对模型结果的影响，我们过滤掉出现次数少于 5 次的项目，然后删除数据集上少于 2 项的所有交互项，将评分大于 4 的设置为 1，小于 4 的设置为 0。数据集中的数据属性及数据特征如下表 1 所示：

表 1 Amazon 5–Elect 数据集数据属性

Table.1 Amazon 5-Elect dataset data properties

字段名	字段类型	字段描述
reviewerID	String	用户编号
asin	Int	商品编号
overall	Int	产品等级
unixReviewTime	Int	Unix 格式时间戳

所使用三个公开数据集的数据统计结果如表 2 所示。

表 2 实验数据统计

Table.2: Statistics of the datasets in experiments

数据集	用户数	项目数	用户交互数	密度
ML100k	943	1,682	100,000	6.3%
Amazon5-Elect	192,404	63,002	1,689,188	0.014
Retailrocket	1,407,580	36,968	2,756,101	0.0052%

[1] http://jmcauley.ucsd.edu/data/amazon/links.html

[2]http://www.kaggle.com/retailrocket/ecommerce-dataset

3.2 评价指标

本文使用精确率 Precision@K（precision rate）、归一化折损累积增益 NDCG@K（Normalized Discounted Cumulative Gain）和命中率 HR@K（Hit Ratio），K 为推荐列表的长度，设置 K=5，10 来评估模型的好坏。

精确率 PR 表示的是模型预测为正的样本中有多少是真正的正样本，计算公式为：

$$PR = \frac{TP}{TP + FP} \tag{17}$$

HR@K 指标衡量的是推荐结果的召回率，关注的是推荐模型的准确性，计算公式为：

$$HR@k = \frac{Hits@k}{|GT|} \tag{18}$$

NDCG@K 用于评价排序的准确性，即是评价推荐系统所给出的推荐列表的好坏。该指标关注的是预测到的项目是否放在推荐列表中更靠前的位置，即强调推荐系统的“顺序性”，计算式为：

$$NDCG@k = \frac{1}{k}\sum_{i}^{k}\frac{2^{r(i)}-1}{\log_2(i+1)} \tag{19}$$

其中 $r(i)$ 为推荐列表中项目 i 的相关性分数，如式(20)所示：

$$\begin{cases} 0, p_i \notin G \\ 1, p_i \in G \end{cases} \tag{20}$$

其中 p 为推荐列表，G 为实验测试集，p_i 表示推荐列表中的第 i 个项目。

3.3 对比模型及实验设置

为了评估本模型性能，选取几个序列推荐模型，在数据集 ML100k、Amazon 5-Elect 和 Retailrocket 上分别进行对比验证：

1） NARM[17]：该模型通过循环神经网络 RNN 最后一步的输出来捕捉用户的序列行为，利用 RNN 中每一步的输出通过注意力机制来捕捉用户的主要兴趣偏好。

2） GRU4Rec[18]：该模型首次利用 RNN 从用户的会话序列中获取用户交互序列中的顺序依赖关系，相比较于混合结构的序列模型，单一 RNN 的结构比较简单。

3） STAMP[19]：该模型是基于注意力机制的序列推荐模型，捕捉用户兴趣偏好，解决了传统的 RNN 结构不能很好的拟合用户兴趣问题。

4） NLR[20]：该模型将序列推荐看作一项认知任务，作者根据离散数学的理论设计了具有逻辑推理能力的网络用于推荐任务。

5） SASRec[21]：采用自注意力机制来对用户的历史行为信息进行建模，提取更有价值的信息。同时通过注意机制可以基于相对较少的用户交互进行预测。

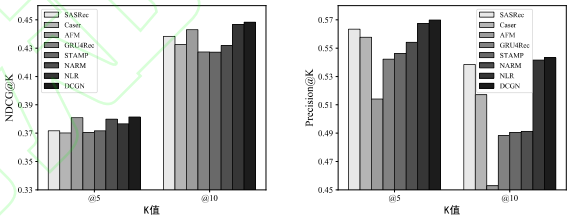
6） Caser[22]：采用卷积神经网络捕获用户行为序列，再利用全连接层将拼接的序列特征与用户偏好映射到用户在当前时间与每个物品交互的可能性。

7） AFM[23]：通过注意力机制来自动学习每个二阶交叉特征的重要性，过滤无用的交叉特征，提升模型的稳定性。

在此次实验中，学习率调整范围为[0.0001，0.001，0.005，0.01，0.1]，嵌入向量以及隐层维度调整范围为[8，16，32，64，128]，在 ML100k 数据集上进行参数调整。模型训练数据的批大小设置为 100，模型参数的初始化采用服从 N（0，0.1）高斯分布的随机数进行初始化。

3.4 实验结果比较

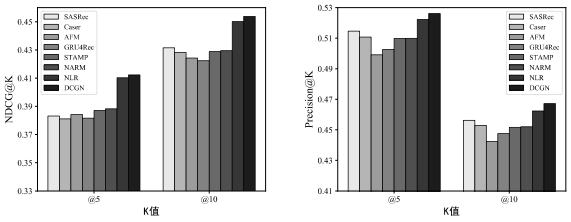
三个数据集 ML100K、Amazon 5-Elect 和 Retailrocket 上所有对比模型在两个评估指标 NDCG 和 Precision 的实验结果分别如图 4，5，6 所示：



(a) NDCG@K评价指标下的模型效果　(b) Precision@K评价指标下的模型效果

图 4 ML100K 数据集上模型效果对比

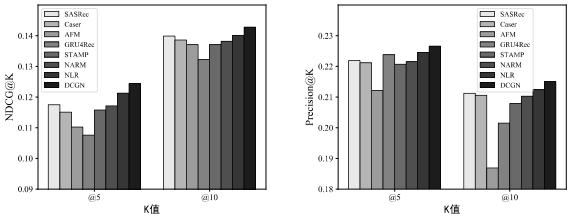
Fig.4 Comparison of models on ML100K dataset



(a) NDCG@K评价指标下的模型效果　(b) Precision@K评价指标下的模型效果

图 5 Amazon 5-Elect 数据集上模型效果对比

Fig.5 Comparison of Amazon 5-elect dataset models

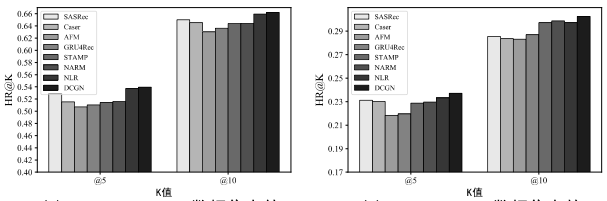


(a) NDCG@K评价指标下的模型效果　(b) Precision@K评价指标下的模型效果

图 6 Retailrocket 数据集上模型效果对比

Fig.6 Comparison of the effects of Retailrocket dataset models

在数据集 Amazon 5-Elect 和 Retailrocket 上所有对比模型在 HR 评估指标的实验结果如图 7 所示：



(a) Amazon 5-Elect数据集上的模型效果　(b) Retailrocket数据集上的模型效果

图 7 数据集 Amazon 5-Elect 和 Retailrocket 上模型效果对比

Fig.7 Performance comparison of models on Amazon 5-element and Retailrocket datasets

通过实验对比可以得出：

a) 在三个数据上，NARM 模型的效果始终优于 GRU4rec 模型，二者都是基于 RNN 的推荐模型，可见注意力机制的运用有效的提升了模型的性能；

b) 实验结果中 AFM 模型的性能总体较低，表明融入注意力机制的简单模型性能有限，模型性能的好坏还与模型的复杂程度有关；

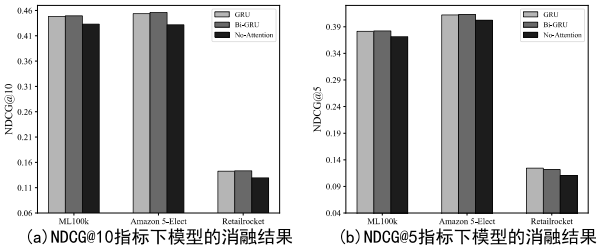
c) 实验中 STAMP 模型的效果次于 NARM 模型，二者都融入了注意力机制，表明基于 RNN 的模型相比基于 MLP 的模型的效果更佳，在推荐系统中仅仅考虑用户交互序列的特征交叉是不够的，还需要考虑用户交互行为前后的关系以捕获用户的顺序行为；

d) SASRec 模型的效果始终好于 Caser、GRU4Rec、STAMP、NARM 模型，说明叠加多个自注意力机制层能够学习更复杂的特征转换，可有效的提高模型的性能，相较于传统的基于 CNN、RNN 的模型展现出明显的优势；

e) DCGN 比 SASRec 和 NLR 模型具有更好的效果，可见利用门控网络和注意力机制从用户交互序列中学习用户动态兴趣有效的提升了推荐效果，采用传统结构的 DCGN 模型相较于采用复杂结构的 NLR 模型表现更好，意味着模型复杂的结构不是提升推荐性能的关键。

3.5 模型消融实验

在三个数据集 ML100K、Amazon 5-Elect 和 Retailrocket 上对模型进行了消融实验，验证了有无注意力机制对模型的影响，以及验证了用户兴趣聚合部分使用 GRU 和双向 GRU 给模型带来的差异。实验结果如图 8 所示，其中 No-Attention 代表无注意力机制，Bi-GRU 表示双向 GRU。



(a) NDCG@10指标下模型的消融结果　(b) NDCG@5指标下模型的消融结果

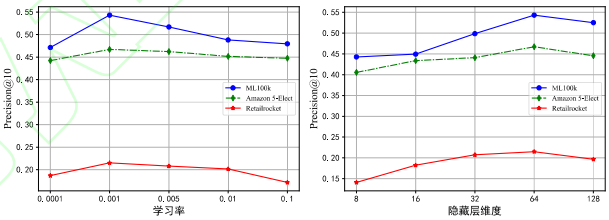
图 8 在三个数据集上模型消融结果对比

Fig.8 Comparison of model ablation results on three data sets

根据上面的实验结果可以观察到，首先，在无注意力机制时模型的效果下降明显，这反映出使用注意力机制可有效的捕获用户兴趣的动态变化，在不使用注意力机制进行 GRU 更新门的更新时，GRU 只能学习到用户静态的兴趣表示，进而模型的效果有所下降；其次，我们观察到，在使用了双向 GRU 后，并未给模型效果带来明显的提升，这是因为双向 GRU 较单向 GRU 更复杂，参数量的增多导致模型拟合不够充分；最后，通过在三个数据集上 NDCG 性能指标下的实验结果表明，模型在使用注意力机制后效果更好，为降低模型的训练难度，我们并未采用双向 GRU。

3.6 参数对模型的影响

学习率的大小和隐藏层的维度大小对实验结果产生一定的影响，尤其是隐藏层维度（即嵌入向量的维度）的大小限制了模型的表达能力，理论上嵌入向量的维度越大，所蕴含的信息越多，模型的效果越好。但实验表明，较高的维度会导致模型效果有所下降。在 ML100K 数据集上的实验结果如图 9 所示：



(a) 学习率对模型的影响　(b) 隐藏层维度对模型的影响

图 9 学习率及隐藏层维度对模型的影响

Fig.9 The influence of learning rate and hidden layer dimension on the model

由实验结果可知，隐藏层的维度较大时，模型的效果出现下降，在我们所选隐藏层的维度中，维度为 64 时效果最好，当隐藏层的维度下降时，在 ML100K 数据集上，模型效果下降最为明显。模型隐藏层的维度较大时，训练所需要的参数增多，模型训练所需的计算资源增多，训练周期变长，模型的过拟合风险增大。

4 结束语

在序列推荐问题中，目前现有的大多方法只使用了用户交互序列中的物品信息，而忽略了相似物品会在交互序列中相继出现这一重要信息，并且用户兴趣是动态变化的。针对以上问题本文提出了一种门控线性网络构建用户动态兴趣的推荐方法 DCGN，通过门控线性网络细粒度建模用户的兴趣表示，通过使用注意力机制动态构建用户的兴趣表示。通过双线性特征交叉方法，对项目之间的共现模式进行建模，模型

的泛化能力得到了明显的提升。

后续的工作将针对用户的长期兴趣和短期兴趣分别进行建模，以明确用户的长短期兴趣对于推荐效果的影响。此外，用户的长期兴趣和短期兴趣对于推荐用户未来将交互的项目时的影响是不同的，需要对二者的差异进行研究。

参考文献:

[1] Sun J, Guo W, Zhang D, et al. A framework for recommending accurate and diverse items using bayesian graph convolutional neural networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020: 2030-2039.

[2] Sun K , Qian T , Chen T , et al. Where to Go Next: Modeling Long- and Short-Term User Preferences for Point-of-Interest Recommendation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1):214-221.

[3] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information processing systems. 2017: 5998-6008.

[4] Liu D , Wu J , Li J , et al. Adaptive Hierarchical Attention-Enhanced Gated Network Integrating Reviews for Item Recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, PP(99):1-1.

[5] Ma S, Zhu J. Self-attention based collaborative neural network for recommendation[C]//International Conference on Wireless Algorithms, Systems, and Applications. Springer, Cham, 2019: 235-246.

[6] Cheng C, Yang H, Lyu M R, et al. Where you like to go next: Successive point-of-interest recommendation[C]//Twenty-Third international joint conference on Artificial Intelligence. 2013.

[7] Pasricha R, McAuley J. Translation-based factorization machines for sequential recommendation[C]//Proceedings of the 12th ACM Conference on Recommender Systems. 2018: 63-71.

[8] Liu L, Zhang P. A novel recommendation algorithm with knowledge graph[C]//Journal of Physics: Conference Series. IOP Publishing, 2021, 1812(1): 012035.

[9] Ma C, Kang P, Liu X. Hierarchical gating networks for sequential recommendation[C]//Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019: 825-833.

[10] Ni Y, Ou D, Liu S, et al. Perceive your users in depth: Learning universal user representations from multiple e-commerce tasks[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 596-605.

[11] Gu Y, Ding Z, Wang S, et al. Hierarchical User Profiling for E-commerce Recommender Systems[C]//Proceedings of the 13th International Conference on Web Search and Data Mining. 2020: 223-231.

[12] Feng Y, Lv F, Shen W, et al. Deep session interest network for click-through rate prediction[J]. arXiv preprint arXiv:1905.06482, 2019.

[13] Cen Y, Zhang J, Zou X, et al. Controllable multi-interest framework for recommendation[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020: 2942-2951.

[14] Dauphin Y N, Fan A, Auli M, et al. Language modeling with gated convolutional networks[C]//International conference on machine learning. PMLR, 2017: 933-941.

[15] Ning X , Desrosiers C , Karypis G . A Comprehensive Survey of Neighborhood-based Recommendation Methods[J]. Springer US. 2015: 37-76.

[16] Harper F M, Konstan J A. The movielens datasets: History and context[J]. Acm transactions on interactive intelligent systems (tiis), 2015, 5(4): 1-19.

[17] Li J, Ren P, Chen Z, et al. Neural attentive session-based recommendation[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017: 1419-1428.

[18] Hidasi B, Karatzoglou A, Baltrunas L, et al. Session-based recommendations with recurrent neural networks[J]. arXiv preprint arXiv:1511.06939, 2015.

[19] Liu Q, Zeng Y, Mokhosi R, et al. STAMP: short-term attention/memory priority model for session-based recommendation[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 1831-1839.

[20] Chen H, Shi S, Li Y, et al. Neural Collaborative Reasoning[C]//Proceedings of the Web Conference 2021. 2021: 1516-1527.

[21] Kang W C, McAuley J. Self-attentive sequential recommendation[C]//2018 IEEE International Conference on Data Mining (ICDM). IEEE, 2018: 197-206.

[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. 2018: 565-573.

[23] Xiao J, Ye H, He X, et al. Attentional factorization machines: Learning the weight of feature interactions via attention networks[J]. arXiv preprint arXiv:1708.04617, 2017.