

# MN-HDRM:长短兴趣多神经网络混合动态推荐模型

冯永<sup>1,2)</sup> 张备<sup>1,2)</sup> 强保华<sup>3),4)</sup> 张逸扬<sup>1,2)</sup> 尚家兴<sup>1,2)</sup>

<sup>1)</sup>(重庆大学计算机学院 重庆 400044)

<sup>2)</sup>(重庆大学信息物理社会可信服务计算教育部重点实验室 重庆 400044)

<sup>3)</sup>(桂林电子科技大学广西可信软件重点实验室 广西 桂林 541004)

<sup>4)</sup>(桂林电子科技大学广西云计算与大数据协同创新中心 广西 桂林 541004)

**摘要** 动态推荐系统通过学习动态变化的兴趣特征来考虑推荐系统中的动态因素,实现推荐任务随着时间变化而实时更新.该文提出一种携带历史元素的循环神经网络(Recurrent Neural Networks,简称RNN)推荐模型负责用户短期动态兴趣建模,而利用基于前馈神经网络(Feedforward Neural Networks,简称FNN)的推荐模型对用户长期兴趣建模.通过两种神经网络的融合,该文构建了一个兼顾用户短期动态兴趣和稳定长期兴趣的多神经网络混合动态推荐模型(Hybrid Dynamic Recommendation Model based on Multiple Neural Networks,简称MN-HDRM).实验结果表明相对于目前比较流行的多种动态推荐算法:TimeSVD++、基于HMM(Hidden Markov Model)的推荐模型、基于RNN(Recurrent Neural Networks)的推荐模型、基于LSTM(Long Short-Term Memory)的推荐模型和STG(Session-based Temporal Graph)推荐模型,MN-HDRM在精确率、召回率和平均倒数排名等多项评价指标上展现出更加优越的性能.

**关键词** 循环神经网络;前馈神经网络;动态推荐模型;长短期兴趣;时间因素

中图法分类号 TP18 DOI号 10.11897/SP.J.1016.2019.00016

## MN-HDRM: A Novel Hybrid Dynamic Recommendation Model Based on Long-Short-Term Interests Multiple Neural Networks

FENG Yong<sup>1,2)</sup> ZHANG Bei<sup>1,2)</sup> QIANG Bao-Hua<sup>3),4)</sup> ZHANG Yi-Yang<sup>1,2)</sup> SHANG Jia-Xing<sup>1,2)</sup>

<sup>1)</sup>(College of Computer Science, Chongqing University, Chongqing 400044)

<sup>2)</sup>(Key Laboratory of Dependable Service Computing in Cyber Physical Society, Ministry of Education, Chongqing University, Chongqing 400044)

<sup>3)</sup>(Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin, Guangxi 541004)

<sup>4)</sup>(Guangxi Cooperative Innovation Center of Cloud Computing and Big Data, Guilin University of Electronic Technology, Guilin, Guangxi 541004)

**Abstract** Recommendation system is a kind of automatic information retrieval technology to solve the problem of information overload in the Internet large data scene. However the traditional recommendation systems based on collaborative filtering model, such as the matrix decomposition obtain user hidden features by learning user history data to learn the user and recommend items on the basis of these hidden features to be in accordance with the needs of users. In today's mobile Internet environment, the traditional recommendation system has been difficult to cope with the rapid migration of user interests, so the dynamic recommendation system came into

收稿日期:2017-07-19;在线出版日期:2018-07-11. 本课题得到国家重点研发计划(2017YFB1402400)、国家自然科学基金(61762025)、重庆市基础与前沿研究计划(cstc2017jcyjAX0340)、广西可信软件重点实验室开放课题(kx201701)、广西云计算与大数据协同创新中心开放课题(YD16E01)、重庆市重点产业共性关键技术创新专项(cstc2017zdey-zdyxx0047)、重庆博士后科学基金(Xm2017125)、重庆市社会事业与民生保障科技创新专项(cstc2017shmsA20013)资助. 冯永,男,1977年生,博士,教授,主要研究领域为大数据分析 with 数据挖掘、人工智能与大数据处理、深度学习与大数据检索. E-mail: fengyong@cqu.edu.cn. 张备,男,1991年生,硕士,主要研究方向为智能推荐与神经网络. 强保华,男,1972年生,博士,教授,主要研究领域为大数据处理与信息检索. 张逸扬,男,1990年生,硕士,主要研究方向为深度学习与大数据检索. 尚家兴,男,1987年生,博士,讲师,主要研究方向为人工智能与大数据处理.

being. Dynamic recommendation system introduces dynamic factors into recommendation system by learning dynamic changed characteristics so that the recommended tasks can be updated in real time. The dynamic recommendation model mainly depends on the temporal data, that is, the user's historical behavior data with time stamp. Therefore, it is helpful to understand the dynamic features of users' interest in order to improve the accuracy of personalized recommendation. The recommendation system based on recurrent neural network can be used to find out the user's historical behavior patterns, so as to find out the rules of user interest transfer. In this paper, we researches the dynamic problems of the recommendation system and put forward the concept of user behavior term. Recurrent neural network will be introduced into the field of dynamic recommendation system; user behavior history data can be split into segmentation of user behavior term to learn model parameters about the short-term interest related learning neural network structure of the user. We make the further improvement for the recommendation system model based on recurrent neural network by add historical element composed of the behavior of user term data, used to store the historical term behavior, in order to improve traceability of historical information of recurrent neural network. At the same time, we put forward to increase the Embedding layer and Dropout layer to alleviate the overfitting problem of neural network training in recurrent neural network. We propose a novel recommendation model, Hybrid Dynamic Recommendation Model based on Multiple Neural Networks, referred to as MN-HDRM, which fuses two kinds of neural networks: RNN and FNN, into one neural network, to help recommender take into account the influence of the user's long-term interest. MN-HDRM recommendation system is combined with the characteristics of two kinds of neural network of recurrent neural network and feedforward neural network, integrated of short-term interest factors of user behavior under short-term condition and long-term interest factors under the users' global environment, to achieve dynamic balance between long-term and short-term preference in recommender systems. At the same time, in the paper we choose the Bayesian Personalized Ranking (BPR) as the objective function for best recommended items list, in order to achieve the best results in Top-N recommendation. The experiment results show that MN-HDRM has more superior performance on *precision*, *recall* and *MRR* compared to current popular dynamic recommendation algorithms, such as TimeSVD++, HMM (Hidden Markov Model) based model, RNN (Recurrent Neural Networks) based model, LSTM (Long Short-Term Memory) based model and STG (Session-based Temporal Graph) recommendation model.

**Keywords** recurrent neural network; feedforward neural network; dynamic recommendation model; long-short-term interest; time factor

## 1 引言

目前在推荐系统领域,个性化推荐系统一直受到更多的关注. 个性化推荐系统一般利用用户历史行为信息对用户的特征、兴趣等因素进行深入分析挖掘,并以此为用户匹配推荐符合用户需求的信息或服务. 它最重要的特点在于能够充分适应用户需求不明确和复杂的情况,并且能够利用用户的历史数据,构建合理的算法,捕获用户的兴趣. 比如,当用

户想寻找感兴趣的歌曲,然而面对海量的音乐时,存在选择困难的问题,此时个性化推荐系统会根据用户的历史行为记录挖掘用户偏好,为用户选择其感兴趣的物品. 因此,个性化推荐系统可降低筛选成本,减少用户选择的时间,从而提高用户的体验.

随着消费者与相关网站之间的交互变得更加频繁,人们对于推荐技术的依赖也越来越强. 因此,如何根据用户频繁的行为记录及时为用户提供精准的推荐服务成为实时推荐场景下亟待解决的问题. 但其实类似算法往往都会忽视用户兴趣漂移问题. 在

日常生活中,用户兴趣一般不会维持非常稳定的状态,尤其是在一些消费频繁的商品上,比如微博、音乐以及电子商务网站.其中,时序动态因子在推荐系统的实际应用中一直扮演着极其重要的角色,如 Amazon 的商品推荐、Netflix 的影视作品推荐以及 Google 的新闻和视频推荐等推荐系统.它们进行推荐的时候,对于用户动态兴趣的变化极其敏感,所以需要用户的动态信息进行更加深入的挖掘.因此,动态推荐系统在现阶段的推荐系统发展中占据及其重要的地位.

最近深度学习在多个领域取得了比较突出的贡献,深度学习在推荐系统的应用也相应出现,并获得了良好的效果.特别是在动态兴趣建模和实时推荐方面,循环神经网络(RNN)具有较好的优势.RNN 善于处理序列化数据,相比于纵向传播的神经网络,它更关注横向时间轴上信息的传播,传播过程中能够保留足量的历史信息,可以展示动态时序行为<sup>[1-6]</sup>.因此 RNN 在推荐过程中对于用户兴趣偏好的挖掘分析具有更加突出的辅助作用.

本文对影响动态推荐系统性能的相关因素进行深入研究,并重点探索循环神经网络模型在动态推荐系统方面的促进作用,同时结合相关研究提出合理的改进措施,进一步优化循环神经网络在推荐系统中对于历史因素信息的综合利用,并针对循环神经网络在推荐新项目问题上存在的不足进行改良,比如,只能获知最近的项目比以前的项目有更重要的影响,但无法获得全局范围的项目影响.本论文提出一重基于多神经网络的混合动态推荐模型(MN-HDRM)来解决这一问题,实现长短期兴趣的融合.MN-HDRM 主要利用神经网络框架融合两类神经网络模型:循环神经网络负责对动态数据建模,前向神经网络负责全局用户数据建模.混合神经网络推荐模型综合考虑用户动态性和用户间相关性等多重关联因素,融合长短期兴趣在推荐系统中发挥的影响.为了验证本文提出的混合神经网络推荐模型的有效性,在多个公开数据集上进行对比实验,并取得优异的效果.

## 2 相关工作

我们将分析不同时期在动态推荐系统模型构建过程中应用过的多种类型的时序预测和时间感知预测方法,其中主要包括:时间感知邻域模型(Time-aware Neighborhood Models)、时间感知因

子分解方法(Time-aware Factorization Methods)、基于马尔可夫链(Markov Chain)的方法和基于神经网络(Neural Network)的方法.

### 2.1 时间感知邻域模型

时间感知邻域模型<sup>[7-9]</sup>是用于时序信息建模的直观有效方法.这些方法一般通过使用基于邻域的推荐算法,给予距离当前时刻最近的观测项目更多关联性关注,而距离时间较远的项目相对较少的关注.因此,依赖这种方式虽然可以比较直接有效地获取时序影响的情况,但其方法本身依旧建立在协同过滤的基础上,并没有引入新的技术.

### 2.2 基于矩阵分解的方法

基于矩阵分解(Matrix Factorization)的方法已成为协同过滤模型中的先进方法.现在,基于 MF 的方法已经扩展到更一般和复杂的情况.其中,时间感知因子分解模型已被广泛研究.Tensor 因式分解(Tensor Factorization)将时间片视为另一个维度,并通过因子分解生成时间片的潜在向量,以捕捉行为历史中的潜在特性.在文献[10]中,作者将概率矩阵分解扩展为张量以分解时序关系数据.TimeSVD++<sup>[11]</sup>为不同时间片段中的用户和项目学习时间感知表示.在文献[11]中,作者考虑将个体用户和项目的极端影响封装在基准预测器内,这些基线预测因素倾向于捕获大部分观测信号,特别是数据中的大部分时间动态信息,因此可以通过这种方法实现增加了时序关系的矩阵分解.

### 2.3 基于马尔可夫链的方法

根据马尔可夫假设,基于马尔可夫链(MC)的方法是一种已经被广泛使用于序列预测任务的模型<sup>[12]</sup>.基于马尔可夫链的模型通过估计转移矩阵来预测用户的下一个行为,其中转移矩阵给出基于先前行为的动作概率.通过用户个性化概率转移矩阵的因式分解,批量式个性化马尔可夫链(Factorizing Personalized Markov Chain)<sup>[13]</sup>可以为每个序列提供更准确的预测.同时 FPMC 也通过使用用户组<sup>[14]</sup>或合并位置约束条件<sup>[15]</sup>的方式进行模型的扩展.最近,人类大脑科学的部分学说也被添加进入基于 MC 的方法中,包括兴趣遗忘曲线<sup>[16]</sup>和厌倦感觉动态变化模型<sup>[17]</sup>.除此之外,在文献[18]中,作者将矩阵分解模型和马尔可夫链结合起来,形成一个马尔可夫链上的个性化迁移图模型.其中,每个用户都单独拥有属于自己的迁移矩阵.除了基于马尔可夫链方法的推荐模型,研究人员也将隐马尔可夫模型(HMM)应用在动态推荐系统上<sup>[19]</sup>,将用户兴趣作

为隐藏状态,而商品作为可见状态,对用户兴趣动态变化过程进行建模,实现最终的基于用户兴趣迁移的动态推荐。

## 2.4 基于神经网络的方法

最近,基于神经网络结构形成了一些比较有效的预测模型,特别是语言模型<sup>[1]</sup>。在多样性语言模型中,循环神经网络(RNN)是模拟句子结构模型中最成功的一种<sup>[2-4]</sup>。它已成功应用于各种自然语言处理任务,如机器翻译<sup>[20-21]</sup>、对话机<sup>[22-23]</sup>和图像标记<sup>[24-25]</sup>。除此以外,基于RNN的模型也在其他领域取得了一系列的成果,比如在视频分析方面,RNN为行动识别带来满意的结果<sup>[5]</sup>。结合用户每次点击作为每个层的输入元素,RNN大幅提高了顺序点击预测的性能<sup>[6]</sup>。在空间位置预测方面,空间-时间周期性神经网络(ST-RNN)<sup>[26]</sup>在RNN框架中学习位置预测的地理距离特定过渡矩阵。

循环神经网络(RNN)作为一种比较优异的处理序列问题的方法,在推荐系统也得到应用<sup>[27]</sup>。动态回归模型(DREAM)<sup>[28]</sup>在每一层RNN中使用池化方法来聚合一个事务中的项目向量,它在下一个篮子推荐中实现了较好的性能。在上下文感知反复神经网络(CA-RNN)<sup>[29]</sup>中,RNN构建的推荐系统框架中同样利用包含多种上下文的信息处理行为序列问题。然而,当处理序列数据时,RNN也存在一些问题,假定时间依赖性随着序列中的位置单调地变化,这意味着序列中的一个元素通常比用于预测的先前元素具有更重要的效果。在文献<sup>[30]</sup>中,作者提出利用RNN和FNN模型进行共同推荐的思想,但未能充分考虑长短期因素的不同影响。

在上述采用RNN的方法中,我们只能获知最近的项目比以前的项目有更重要的影响,而且,这些RNN结构通常只在许多RNN的一部分并入更大的时标来捕获序列中的长期依赖性,从而导致RNN不能很好地模拟行为建模中的短期上下文。尽管在一些应用中它们确实比传统结构有更好的性能,但是它们仍然根据RNN结构中的连续顺序对输入元素建模,使得这些基于RNN的方法仍然很难在序列行为场景中对短期上下文进行良好建模。

## 3 主要工作

### 3.1 动态推荐中的行为周期

假设我们有一组用户和一组项目,分别表示为 $U = \{u_1, u_2, u_3, \dots\}$ 和 $V = \{v_1, v_2, v_3, \dots\}$ 。用户 $u$ 的

每个历史行为都与行为的时间戳标记相关联。然后,用户 $u$ 的时序行为历史包括项目 $V^u = \{v_1^u, v_2^u, v_3^u, \dots\}$ ,相应的行为时间戳 $T^u = \{t_1^u, t_2^u, t_3^u, \dots\}$ 。如果我们已经知道目标用户关于项目的行为历史记录,那么动态推荐问题可以表述为预测特定用户将在特定环境条件下,下一步会选择何种项目的问题。

下面我们通过电子商务的数据应用举例说明。在电商网站的交互过程中,用户存在点击、购买、加入购物车和确认购买等行为,所有这些行为都是用户历史记录的组成部分,而我们可以认为用户购买行为和他之前的历史行为存在着时间上的相关性,换言之用户历史行为暗示用户的购物倾向。因此,我们在这里定义一个行为周期 $s^u = \{v_{t_1}^u, v_{t_2}^u, v_{t_3}^u, \dots, v_{t_{r-1}}^u, v_{t_r}^u\}$ , $v_{t_i}^u \in V^u$ 。一个行为周期中 $t_1$ 时刻到 $t_{r-1}$ 时刻的商品为用户在一个行为周期历史记录中访问过的商品,而 $t_r$ 时刻的商品则是用户在当前行为周期内购入的商品。所以,行为周期可以理解为用户购买商品以及之前的一段时间的访问记录集合。根据定义,动态推荐模型在电商网站中的应用所解决的问题可以理解为根据用户的一个行为周期的访问商品的历史记录来预测用户在这个行为周期中可能购买的商品。上述描述是针对存在用户行为类型已进行分类标记的数据集而言的。如果用户历史记录行为不存在类型标记,则所有历史行为的类型统一识别为点击操作,动态推荐模型需要解决的问题就变成了如何根据用户一个行为周期的历史点击行为来预测用户下一次点击的商品。行为周期也就转变为点击行为以及之前一段时间的点击行为的集合。相似的定义可以使用在大多数拥有用户日志的在线网站的数据集上。

根据上文关于行为周期的定义,我们可知用户的整个历史行为记录其实由用户的一个个行为周期组成, $H_u = \{s_1^u, s_2^u, s_3^u, \dots, s_n^u\}$ ,其中 $H_u$ 表示用户 $u$ 的全部历史记录, $s_i^u$ 正如上文所示即用户的一个行为周期。由于定义的行为周期包括用于预测的商品序列,也包括该周期最终选择的商品,这样的数据组合不利于程序的输入输出运算。因此,我们将用于预测的行为周期信息和需要预测的行为周期结果分开,重新定义用户数据 $U_i = \langle S_i, R_i \rangle$ 二元组,作为算法的完整输入。 $U_i$ 表示第 $i$ 个用户的全部数据, $S_i$ 表示第 $i$ 个用户的所有行为周期的预测信息,即上文中的 $H_u$ 中的每个行为周期剔除最后一项组成的序列行为集合。而 $R_i$ 则表示每个行为周期最后一项结果组成的结果集合。

由上文的定义我们初步可知本文提到的动态推荐的基本流程和大致方式. 所以本文要解决的推荐问题可以转化为, 输入用户行为周期信息  $S_i$ , 输出用户行为周期结果  $R_i$ . 在这样的定义下, 动态推荐问题具体化为构建一个动态推荐模型负责用户信息的输入和推荐结果的输出.

### 3.2 携带历史元素的 RNN 推荐模型

然而, 当我们在使用序列数据进行建模的时候, 基本 RNN 模型假设时间依赖因素会随着序列中的位置变化而变化, 这意味着序列中某个成员相比之前的成员通常会产生更加显著的效果. 这通常适用于视频帧、自然语言或语音中, 这也是 RNN 应用的常见情况. 但它并不适应实际场景中用户行为的精准预测, 尤其是在一个用户行为历史过程中短周期突然添加新的元素. 而且一些最近的元素通常对用户的下一个选择有类似甚至更复杂的影响. 但基本 RNN 只能告诉我们最新的项目相比于更早之前的项目有更显著的影响. 因此, 我们可以说, 基本的 RNN 网络框架不能很好地对短期信息的行为进行建模. 此外, 在实际系统中, 我们没有足够的内存来维护所有可能的输入单元, 这种限制很有可能会影响神经网络模型的训练过程以及之后的预测过程.

为了解决上述提到的长序列的内存消耗和短周期内历史信息丢失的问题, 我们采用了一种折中方法来弥补这一缺陷. 虽然不能增加行为周期序列的最大长度, 但我们采用设置历史元素单元来收集历史行为周期信息的方式来补偿短周期序列历史信息的损失, 同时由于只增加了一个输入单元, 没有显著增加内存消耗, 可以使得推荐效率有效提升.

如图 1 所示, 我们安排的历史元素输入作为 RNN 模型第一个输入单元, 在用户行为周期中解决

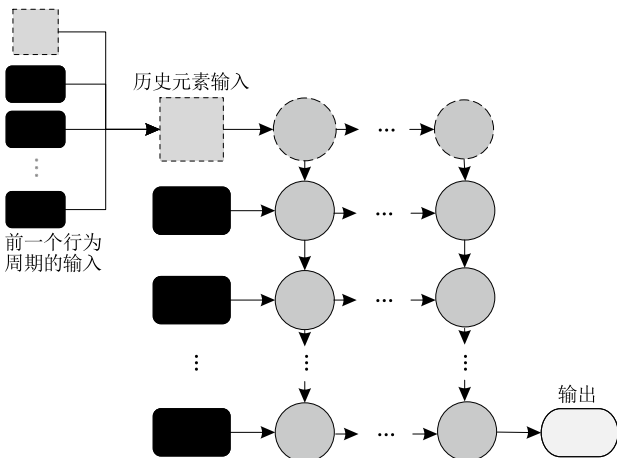


图 1 携带历史元素的 RNN 模型示例

历史信息丢失问题, 负责收集历史信息. 我们引入的历史元素, 融合所有的历史行为周期信息. 其中,  $[x_1, x_2, \dots, x_{m-1}, x_m]$  表示一个购物短周期内所有的项目集合. 然后, 我们将  $\mathbf{Xh}_s$  定义为下一个短周期历史信息的向量表示. 在下一个行为周期中, 我们将输入序列调整为  $[\mathbf{Xh}_s, x_1, x_2, \dots, x_{m-1}, x_m]$ . 在后续的行为周期内, 历史元素可以通过式(1)和(2)计算获得.

其中, 历史因素可以影响行为周期后续时刻的用户选择. 假设我们已经访问过第  $x$  个用户行为周期内容, 历史元素单元则会保留所有历史短周期的信息. 该单元还接受来自历史元素输入的递归关联信息. 对于历史元素输入单元, 我们会使用相同的更新函数导向其隐藏层组成神经单元.

$$\mathbf{Xs}_s^k = \begin{cases} \max(x_1^k, x_2^k, \dots, x_m^k) \\ \frac{1}{m} \sum_{i=1}^m x_i^k \end{cases} \quad (1)$$

$$\mathbf{Xh}_s = \alpha \mathbf{Xh}_{s-1} + (1-\alpha) \mathbf{Xs}_{s-1} \quad (2)$$

其中,  $\mathbf{Xh}_{s-1}$  表示前一个行为周期内的历史元素输入,  $\mathbf{Xs}_{s-1}$  表示前一个行为周期内用户访问过的所有项目向量加权组合向量特征, 下标  $s$  表示某一个用户的短周期标号索引. 如图所示, 当前行为周期内的历史元素输入  $\mathbf{Xh}_s$  的生成依赖  $\mathbf{Xh}_{s-1}$  和  $\mathbf{Xs}_{s-1}$  的加权组合, 其中  $\alpha$  表示历史元素的影响能力的衰减程度, 依据用户特定环境下的习惯设定  $\alpha$  越大, 表示用户受到历史行为周期的影响越大, 也表示用户历史行为周期之间存在的关联关系更加紧密.

如图 2 对于  $\mathbf{Xs}_{s-1}$  的向量融合计算方法, 我们尝试通过两种合并操作来处理这些项目向量, 来降低合并组合后的项目向量维度: 最大池化和平均池化操作. 最大池化操作, 即我们汇集一组向量, 选取在所有这些向量的每个维度分量的最大值作为生成向量该维度分量的值.  $\mathbf{Xs}_s^k$  表示为生成的新向量的每一个维度分量, 相关计算如式(1)的第一部分. 而平

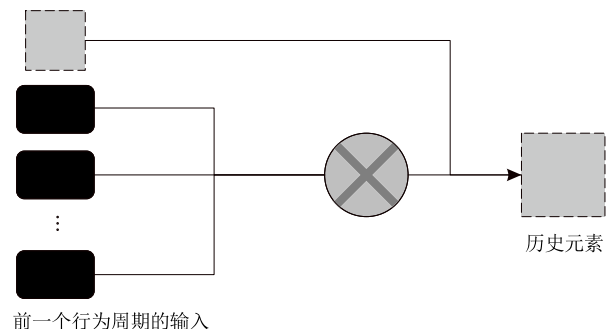


图 2 改进 RNN 结构中历史元素的产生过程

均池化操作类似于最大池化,只不过选取在所有这些向量的每个维度分量的平均值作为生成向量该维度分量的值。 $\mathbf{X}s_i^k$ 表示为生成的新向量的每一个维度分量,相关计算如公式的第二部分.在文献[28]中发现,在循环神经网络中,输入向量池化操作中最大池化操作能够取得更好的效果,这可能主要是因为推荐系统中用户倾向于选择自己最偏好的项目,而且对于排名靠前的项目的记忆远胜于其他,所以才导致这一情况出现.因此,在后续的实验,我们倾向于选择最大池化操作来进行向量融合计算.

### 3.3 基于前馈神经网络(FNN)的协同过滤模型

在一个较短的用户行为周期序列中,用户收到的动态影响比较多,所以推荐的时效性比较明显.从更深层次而言,在动态推荐系统模型中,我们假设短期兴趣可能处于主导地位,用户的兴趣以及行为模式在一定时间范围内(一个行为周期的跨度不长)是可能会发生变化的,所以上文中的基于RNN的动态推荐模型就是试图获取用户的短期兴趣变化或者行为模式.因此基于RNN的推荐系统模型对预测一个特定的用户未来的短期行为可能具有比较好的能力,但是往往会忽略用户全局的兴趣或者说长期兴趣.而这正是协同过滤方法的优势所在,它虽然对于动态因素并不敏感,但是对于用户全局的兴趣的把握还是比较准确的.CF的方法仍然可以从整个历史数据观察角度生成的一个很好的推荐.所以在本文中,我们建立了一个前馈神经网络(FNN)模拟CF方法融合到我们的神经网络模型中.

在文献[31-32]中,已经通过使用FNN构建了推荐系统模型,而且取得了较好的实验效果.在本文中,我们进一步简化了它的输入结构和学习过程.所以为了简单起见,所有用户在隐藏层中共享同一组神经元,对于所有用户适用这一标准框架,唯一区别在于每一层之间的连接权重不一样.接下来,我们将介绍具体的FNN推荐系统框架以取代CF推荐系统.

如图3所示,在FNN的输入层,我们设置 $U$ 个神经元,每一个用户对应一个神经元.其中每个神经元的输入都是一个one-hot向量,表示用户是否购买了商品或者购买商品的概率.同时为了简化推荐问题,我们假设所有的用户共用隐藏层神经元,这种方式目的是为了挖掘用户所共有的行为模式.

假设每个隐藏层有 $M$ 个神经元.第 $j$ 个神经元

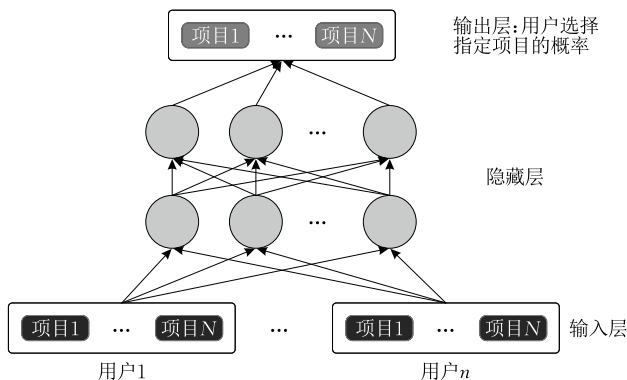


图3 基于FNN的协同过滤推荐示例图

在第 $i$ 个隐层,其值 $\alpha_j^i$ 可以通过如下公式更新:

$$\alpha_j^i = f\left(\sum_{x=0} w_j^{i-1} a_x^{i-1} + b_x^{i-1}\right) \quad (3)$$

式(3)中, $w_j^{i-1}$ 表示从第 $i-1$ 个隐藏层到当前隐藏层的连接权重,而 $b_x^{i-1}$ 则是对应位置的偏置权重.而FNN的输出层同样包含 $N$ 个神经元.而每个神经元代表用户对项目的评分或者选择项目的概率.

### 3.4 多神经网络的混合推荐模型框架

#### 3.4.1 混合动态推荐模型框架

在本节中,我们提出了一种基于多神经网络的混合动态推荐模型(Hybrid Dynamic Recommendation Model based on Multi-neural Networks,简称MN-HDRM)融合RNN和FNN两种神经网络,以吸收上述方法的优势,取长补短.

多模型神经网络框架提出主要是利用分治原则来解决复杂问题.在文献[33-34]中,作者主要利用它们解决分类问题,包括脸部图像分类问题和非平衡分类问题.它们主要将原始集分成多个集合,最后进行策略合并输出分类结果.本文提出引入的多神经网络目的有所不同,主要利用不同类型神经网络的特点弥补单一类型神经网络的不足.

MN-HDRM假设两种类型的神经网络可以应用于不同的处理场景.正如上文所述,两种神经网络:循环神经网络(RNN)和前馈神经网络(FNN)都可以用来预测推荐.但是它们主要解决的问题又存在一些区别,循环神经网络在推荐系统中的应用比较重视时效性,对于动态推荐比较合适,而前馈神经网络则并不具备这种能力,但是对于全局数据信息整体的把握能力较循环神经网络更强.动态推荐虽然更关注的是时间上的变化引起的兴趣变化,但是全局信息的加入并不会削弱它对于时间的敏感性,甚至可以作为动态推荐模型的一种补充,作为全局



信息对用户行为周期中出现的兴趣变化做出更加精确的判断。

因此, MN-HDRM 利用循环神经网络(RNN)保留时序信息, 利用前馈神经网络(FNN)保留全局信息. MN-HDRM 的神经网络框架如图 4 所示. 不同于使用 RNN 模型针对整个用户历史进行建模, 我们的模型融合短期用户兴趣和长期用户兴趣, 其中短期用户兴趣在 RNN 模型中随时间变化相对稳定, 而长期用户兴趣一般不会随着时间产生显著变化.

$$f(\mathbf{W}_1, \mathbf{W}_2) = \beta \mathbf{W}_1 + (1 - \beta) \mathbf{W}_2 \quad (4)$$

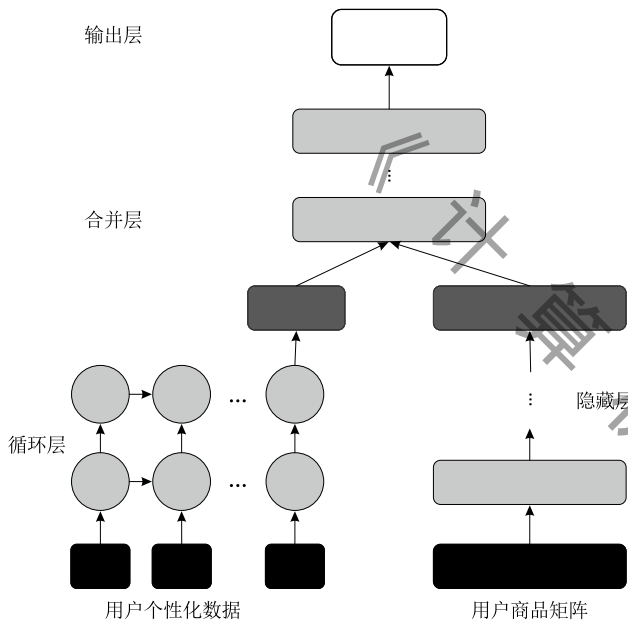


图 4 混合时序推荐模型的基本框架

在 MN-HDRM 网络框架中, 输入层来源分为两种类型: 个性化用户的数据作为短期兴趣获取来源, 整个用户商品矩阵数据作为长期兴趣获取来源. 具体而言, 该网络框架的输入由时序数据和全局数据组成. RNN 模型部分的时序数据是用户短周期数据, 而 FNN 模型部分则为全局数据用户商品矩阵.

两种神经网络模型的输出在图 4 的合并层(merge layer)进行融合, 形成混合商品向量输出, 合并层的具体融合方式如式(4)所示执行,  $\mathbf{W}_1$  和  $\mathbf{W}_2$  分别表示两种神经网络的中间输出向量, 代表混合推荐系统中长短期兴趣各自的影响力, 在原先的各自神经网络中它们表示的都是商品向量, 因此  $\mathbf{W}_1$  和  $\mathbf{W}_2$  维数一致. 它们按照长短兴趣的影响比例进行相加, 在式(4)中,  $\beta$  参数负责控制 RNN 和 FNN 模型对最终输出的影响效力,  $\beta$  值越大表示由 RNN 模型输出表示的短期兴趣影响效力越大, 反之亦然.

除此之外, 本文针对 one-hot 向量导致的低效率问题, 引入嵌入层用于优化神经网络. 如图 4 在输入层数据传递到下一层之前, 可以引入嵌入层, 使其映射到连续的低维空间. 嵌入化的短周期数据用于 RNN 层. 如果一个短周期结束, RNN 隐藏层单元会根据新的短周期进行重置. 同时, FNN 的全部输入数据也可以进行相应处理, 通过设置嵌入层降低输入数据的维数. 此外, 本文 Dropout 层也可以作为模型的合并层实现合并输出的正则化.

### 3.4.2 目标函数及训练方法

本文选择贝叶斯个性化排序(Bayesian Personalized Ranking, BPR)作为训练混合神经网络的目标函数. BPR 是目前关于隐含反馈数据效果最好的 pairwise 排名框架, 主要针对排名而提出. 在文献[35-36]中, 我们发现相比于一般的目标函数, pairwise 排名目标函数能够表现出更加优越的效果. 而 BPR 作为其中一种应用比较成熟的 pairwise 排名框架可以作为本文的神经网络的目标函数之一.

对于每一时刻  $t$ , MN-HDRM 的输出  $o_{u,t,v}$  即为用户项目  $v$  得分, 因而根据上文所述为了确保损失函数 BPR 函数的值最小化, 我们可以获得如下式(5), 用于计算我们选择的项目的分数大于负样本的概率:

$$p(u, t, v > v') = \sigma(o_{u,t,v} - o_{u,t,v'}) \quad (5)$$

其中,  $v'$  表示某一负样本项目, 我们的目标就是希望使正样本商品得分高于负样本的概率最大化.  $\sigma$  表示 sigmoid 函数, 计算分布概率值. 然后, 我们继续加上所有的似然对数结果和正则化项, 目标函数可以简化为式(6)所示:

$$J = \sum \ln(1 + e^{-(o_{u,t,v} - o_{u,t,v'})}) + \frac{\lambda}{2} \|\Theta\|^2 \quad (6)$$

其中,  $\Theta$  表示所有需要学习的模型参数,  $\lambda$  是一个参数控制的正则化参数.

本文中, 目标函数的梯度求导可以通过反向传播方法实现, 其中 RNN 模型的求导方法通过沿时间的反向传播方法(BPTT)实现. 反向传播方法分别重复计算  $J$  相对于相关参数的偏导数, 然后获取所有参数的梯度. 最后我们通过随机梯度下降算法(SGD)更新相关参数直到目标函数收敛.

对于目标函数的参数偏导的求导公式如下所示:

$$\begin{aligned} \frac{\partial J}{\partial \theta} &= \sum \frac{\partial}{\partial \theta} \ln(1 + e^{-(o_{u,t,v} - o_{u,t,v'})}) + \frac{\lambda \theta}{2} \frac{\partial}{\partial \theta} \|\Theta\|^2 \\ &\propto \sum \frac{e^{-(o_{u,t,v} - o_{u,t,v'})}}{1 + e^{-(o_{u,t,v} - o_{u,t,v'})}} \cdot \frac{\partial}{\partial \theta} (o_{u,t,v} - o_{u,t,v'}) + \lambda \theta \Theta \quad (7) \end{aligned}$$

如上文所说,我们采用随机梯度下降法最大化目标函数:

$$\Theta \leftarrow \Theta + \alpha \left( \sum \frac{e^{-(o_{u,t,v} - o_{u,t,v'})}}{1 + e^{-(o_{u,t,v} - o_{u,t,v'})}} \cdot \frac{\partial}{\partial \Theta} (o_{u,t,v} - o_{u,t,v'}) + \lambda_{\Theta} \Theta \right) \quad (8)$$

实际应用过程中,存在多种与梯度下降法类似的方法,如 Adam 等,可以择优选取。

### 3.5 Dropout

在推荐系统加入 Dropout 的动机是将其应用于行为周期访问序列,相当于随机删除部分访问结果的预处理步骤。简单而言,这种操作可以使得我们的模型降低对于噪音的敏感度。例如,用户有时候会不小心访问了某些并不喜欢的项目。也正因为如此,我们的模型也不会因为部分噪音数据而过拟合。某种程度上,Dropout 方法也可被视为一种数据增强方式,为模型训练而产生的行为周期序列数据可以被修剪得更加精细。

如图 5 所示,用户在自己的某一个行为周期序列总共访问过 4 个项目。问号标注的项目单元是与其相应训练序列对应的输出项目结果,训练过程中随机丢弃具有虚线轮廓的项目单元。

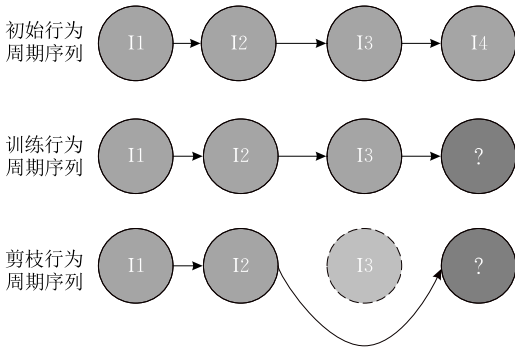


图 5 RNN 推荐模型中 Dropout 行为序列剪枝

## 4 实验及结果分析

### 4.1 数据集

为了比较相关推荐系统算法的精度,我们选择两组来源于真实业务场景的公开数据集进行实验验证。

(1) Tmall 数据集。这组数据集由天池大数据平台提供,主要收集了天猫电商平台上的部分用户历史行为数据。其中包含了 20 000 名用户的完整行为数据以及百万级的商品信息,主要记录了用户在移动端的历史行为数据。考虑 Tmall 数据集中商品集过于庞大,本文拣选其中访问比较频繁的部分商品

用于实验。

(2) Last.fm 数据集。Last.fm 是全球最大的网络音乐库,数据集包含用户的完整收听历史(到 2009 年 5 月 5 号为止,覆盖了近 1 000 名用户)。该数据集主要包含用户、收听时间戳、收听歌手、收听歌曲等成分。

### 4.2 模型相关环境和参数设置

本文选择 Keras 进行 MN-HDRM 框架的搭建,其中在本文中,MN-HDRM 框架 RNN 部分的隐藏层数设置为 1,FNN 部分的隐藏层数设置为 1,其中 RNN 的对于 Tmall 数据集的隐藏层神经元数为 3 000,而对于 Last.fm 数据集则为 1 000,dropout 值设置都为 0.5。而 FNN 部分的隐藏层数设置为 1,相应的神经元数都设置为 1 000。其中各个相关连接的激活函数都设置为 ReLU。其中  $\beta$  超参数的设置关系到长短期兴趣各自在 MN-HDRM 框架最终输出的影响,本文后续实验将讨论该参数的设置。

### 4.3 评价指标

为了评估本文所提出的基于多神经网络的混合动态推荐模型的准确性和有效性,本文采用了多种评价指标来对系统准确度进行测试。在模型返回给用户推荐的列表中,选择用户最感兴趣的  $K$  个项目作为最终的推荐结果。

$$\text{召回率}(Recall, R) = \frac{\text{系统推荐且符合用户需求的项目}}{\text{系统中用户喜欢的项目}}$$

$$\text{精确率}(Precision, P) = \frac{\text{系统推荐且符合用户需求的项目}}{\text{系统推荐的所有项目}}$$

当然,上述公式主要针对一个用户进行计算,如果要评价整个推荐系统的性能,我们一般选用单用户的评价指标的平均值来衡量。

这两个评价指标实际上都只是针对系统最终推荐项目正确总数的一种评价。其中,准确率是推荐系统给用户推荐的项目列表中,用户感兴趣的项目的比例。召回率是测试用户最终选择系统推荐项目与用户实际选择项目之间的关系。但是由于本文方法应用场景的特殊性,我们也需要引入关于排名的评价指标,对 Top-K 推荐项目的排序质量进行评判。本文也采用了平均倒数排名  $MRR$  (Mean Reciprocal Rank) 对推荐排名准确度进行评估。

具体而言,对于一个 Top-K 推荐列表,若第一个用户选择项目位于第  $n$  位,那么这个推荐列表得分为  $1/n$ 。所以, $MRR$  的计算公式如下所示:

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i} \quad (9)$$



其中,  $Q$  为推荐列表集合,  $|Q|$  表示推荐列表集合  $Q$  的长度, 而  $rank_i$  则表示在第  $i$  个推荐列表中用户第一个挑选的项目所在的位置.

为了衡量提出的推荐算法的效用性, 本文计算了各个算法的平均倒数排名, 推荐算法的效用性随着用户选择的项目在推荐列表中排名的上升而增强, 反之, 算法效用性随着排名下降而减弱. 比如, 对于目标用户来说, 如果他最终选择的物品在系统给出的推荐列表中排在最前面, 那么该系统对于用户的效用性最强.

#### 4.4 对比算法

(1) TimeSVD++<sup>[11]</sup>. TimeSVD++ 推荐算法, 是对于静态的基于矩阵分解技术的推荐方法的改进<sup>[37]</sup>. 该模型在已有的静态模型上提出项目动态效应评分和用户动态效应评分, 将矩阵分解过程中植入时序函数, 从而获得用户和项目的动态特征表示. 由于 TimeSVD++ 算法在推荐系统中表现出的优异性能, 所以将其作为本文的对比算法.

(2) 基于 HMM 的推荐算法<sup>[19]</sup>. 基于 HMM 的推荐算法运用隐马尔可夫模型对推荐系统进行建模. 该推荐模型认为用户的推荐过程中受到隐变量, 即用户兴趣的影响, 而且用户兴趣在不同时间存在差异, 会随着时间产生规律性变化. 该算法通过学习隐状态之间的迁移矩阵和隐状态到项目的发现矩阵, 从而来完成项目的动态推荐任务.

(3) 基于 RNN 和基于 LSTM(长短期记忆网络)的推荐算法<sup>[28,36]</sup>. 本文提出一种改进模型——携带历史元素的 RNN 推荐模型. 本文不再直接利用基本 RNN 模型进行实验, 直接选择改进的携带历史元素的 RNN 模型作为对比实验. 历史元素的加入相对基本 RNN 推荐模型考虑了更多的历史信息, 我们利用这组对比实验证明, MN-HDRM 推荐

模型兼顾长短期兴趣的改进措施相对于单独依赖 RNN 推荐模型实现了更高准确率的推荐. 同时, 还与基于 LSTM 的推荐模型进行实验比较.

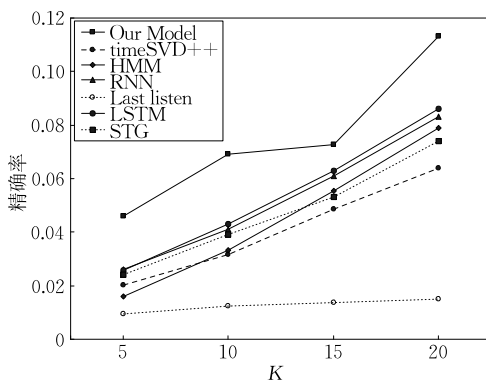
(4) STG 算法(Session-based Temporal Graph)<sup>[38]</sup>. STG 框架将 user-item 数据进一步才分为更短的周期行为数据集, 并将这些数据集和对应的用户进行一一对应, 实现用户兴趣在图模型中的最终交互.

在选择随机推荐对比结果时, 本文针对不同的数据集使用不同的方案. Last.fm 数据集中我们设置最近收听结果, 即 last listen 作为随机结果. 而在 T-mall 数据集中本文选择 random, 即随机推荐作为随机结果.

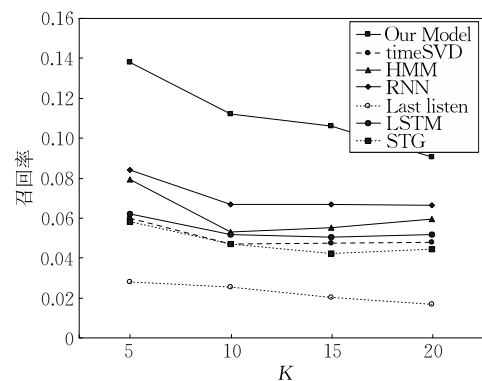
#### 4.5 实验结果分析

在实验中进行 Top-K 推荐时, 对推荐列表长度  $K$  取不同值分别进行实验, 并抽取  $K$  分别取 5、10、15 和 20 时的实验结果作为样本. 在准确率和召回率两个评价指标上的实验结果如图 6 和图 7 所示.

图 6 是在数据集 Last.fm 上进行实验所获得结果, 显示了本文提出的 MN-HDRM 推荐模型中的算法和选取的对比算法在  $K$  取不同值(5, 10, 15, 20)时准确率和召回率的对比及其变化趋势. 从图 6 中可以发现, 相比于最近收听算法, 本文提出的 MN-HDRM 推荐算法及其它对比算法在准确率和召回率上都有较大幅度的提升. 并且, MN-HDRM 推荐算法在召回率方面的表现明显要优于其他四种推荐算法. 除此之外, 我们的模型和基于 RNN 的推荐模型都属于基于循环神经网络的推荐算法, 它们在结果方面的表现也存在着明显优势, 这在某种程度上也表明基于 RNN 的推荐算法以及我们的改进算法在动态推荐领域拥有优势. 而在精确率方面, MN-HDRM 推荐模型依旧表现出比较明显的优势, 基于 RNN 的动态推荐算法则同样略高于 TimeSVD++ 和基



(a) Last.fm 精确率



(b) Last.fm 召回率

图 6 在 Last.fm 数据集上的精确率和召回率结果

于 HMM 的推荐算法. 图 7 是在 Tmall 数据集上的对比实验结果, 我们可以发现, 这两者中的大致趋势相似, 但是本文提出的 MN-HDRM 推荐算法没有表现出太明显的优势, 特别是和 RNN 推荐算法相比, 甚至在推荐列表较短的情况下表现不如 RNN 推荐算法. 我们认为可能是因为作为一个电商网站对于时效性的要求比较高. 在购物过程中, 个人的即时体验和感受对于用户的后续选择影响比较大, 特别是一个用户购买一件商品后, 可能不会再去购买

同一类型的商品. 所以在这一推荐过程中, 用户的全局信息对于推荐系统的重要性可能会有所降低, 不会像音乐影视网站中用户更加关注长期兴趣. 在图 6 和图 7 中可以发现随着推荐列表的长度  $K$  逐渐增大, 各算法的召回率逐渐下降而精确率逐渐上升. 另外, 从图 6 和图 7 中我们可以发现, 与各算法在推荐准确率的表现上都较低. 这主要是因为项目集合的数目比较庞大, 整体的稀疏性问题依旧比较严重, 从而导致精确率和召回率不高.

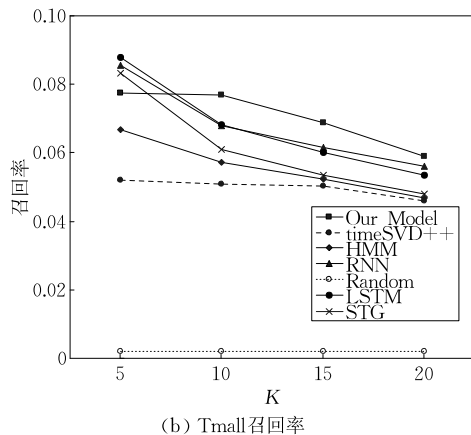
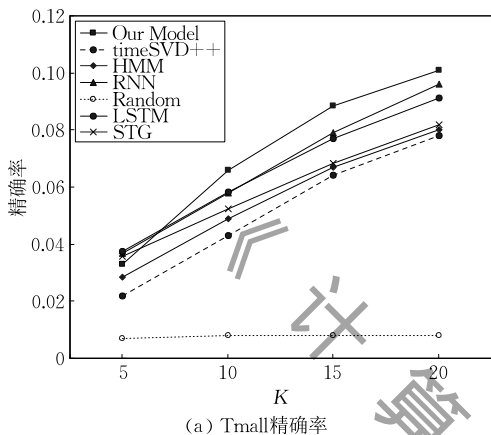


图 7 在 Tmall 数据集上的精确率和召回率结果

同时, 为了更准确地评估各个推荐算法的排名质量, 在实验过程中, 本文还对各个推荐算法的推荐平均倒数排名指数  $MRR$  进行计算并展示. 实验结果如图 8 所示, 算法的平均倒数排名指数越高, 表明用户最终所选择的项目在推荐列表中排名越靠前, 系统所输出的推荐列表越符合用户的需求. 在电子商务系统和音乐平台等真实的推荐应用中, 本文所提出的 MN-HDRM 推荐算法要优于 TimeSVD++、基于 HMM 的推荐算法、基于 RNN 推荐算法和基于 LSTM 的推荐算法等对比算法, 这也说明了充分考虑推荐系统的动态性因素对及时发现和满足用户需

求的重要性.

$\beta$  超参数主要的作用在于调节长短期兴趣对 MN-HDRM 最终输出结果的影响. 因此针对不同特点的数据集,  $\beta$  参数取值会影响到神经网络最优输出. 在图 9 中, 我们可以发现在不同的推荐结果长度下  $\beta$  超参数都是在 0.6 左右的时候使得 MN-HDRM 的推荐结果精确度取得最高. 这种结果某种程度上也证明了神经网络推荐模型中长短期兴趣, 即全局环境因素和动态因素对推荐结果有效性的不同影响, 实验结果表明在不同情况下它们能够在某一个值附近获得最佳输出.

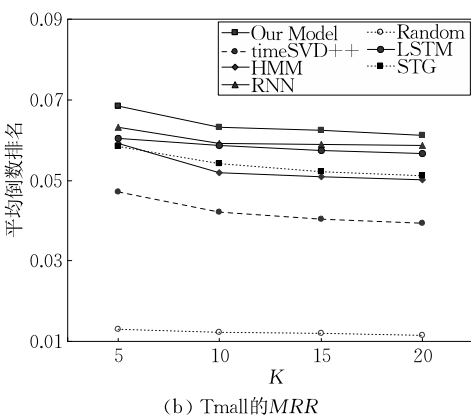
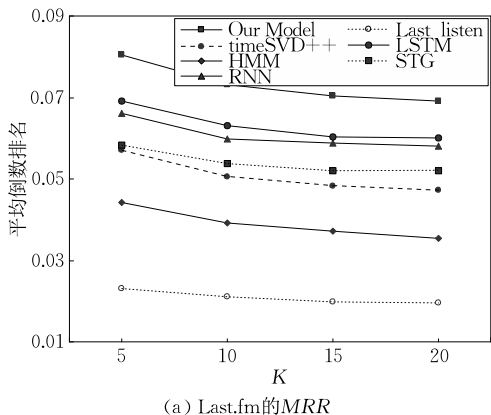
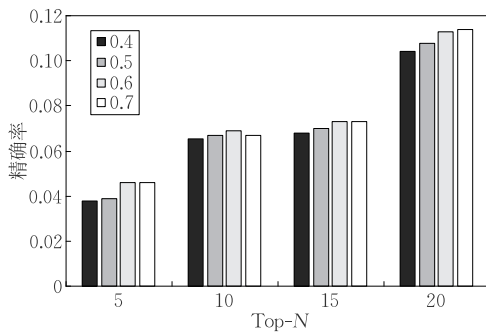
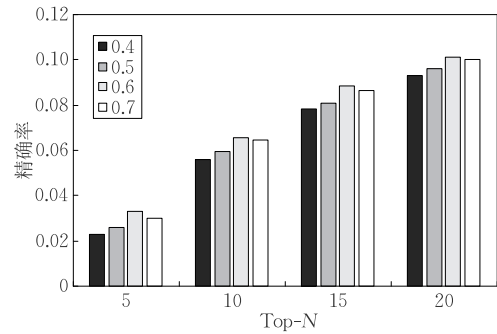


图 8 在数据集 Last.fm 和 Tmall 上的 MRR 结果



(a) Last.fm数据集下不同Beta参数精确率对比



(b) Tmall数据集下不同Beta参数精确率对比

图9 在数据集 Last.fm 和 Tmall 上的 MN-HDRM 关于  $\beta$  参数变化的精确度结果

## 5 结论与展望

本文提出引入历史元素单元用于存储本周期之前的历史信息,作为循环神经网络输入层的一部分,从而改进了传统循环神经网络模型(RNN).同时,基于改进的循环神经网络模型,本文提出一种多神经网络混合动态推荐系统模型(MN-HDRM).MN-HDRM 分别利用改进的 RNN 偏向关注短期行为兴趣的特点和 FNN 基础的 CF 模型关注长期整体行为兴趣的特点,将两种类型的神经网络进行融合,以便于推荐系统能够兼顾长短期兴趣.同时,本文使用 BPR 作为 MN-HDRM 推荐模型的训练算法,然后利用随机梯度下降法对模型进行训练.最终的实验结果表明本文提出的 MN-HDRM 推荐模型在召回率、精确率、MRR 等指标上都表现出比较优越的效果,相比于其它的几种动态推荐算法更为突出.

本文主要研究改进 RNN 模型在动态推荐系统领域的应用,同时提出利用改进 RNN 模型和 FNN 模型组合的 MN-HDRM 推荐模型来解决推荐预测问题.但是对于深度神经网络模型而言,过拟合问题对于模型的训练学习影响很大,所以如何通过优化神经网络来缓解这一问题是下一步研究工作之一.除此之外,虽然本文提出利用历史元素单元存储历史信息,而且方法比较直观且易于理解,但是当前行为周期对于历史信息的吸收过程还比较复杂,可能需要更加复杂的模型对这一过程进行解析和建模,这也是本文下一步的研究工作.

### 参 考 文 献

[1] Hu B, Lu Z, Li H, et al. Convolutional neural network architectures for matching natural language sentences.

Advances in Neural Information Processing Systems, 2015, 3: 2042-2050

- [2] Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model//Proceedings of the 11th Annual Conference of the International Speech Communication Association. Makuhari, Japan, 2010: 26-30
- [3] Mikolov T, Kombrink S, Burget L, et al. Extensions of recurrent neural network language model//Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Prague, Czech Republic, 2011: 5528-5531
- [4] Mikolov T, Kombrink S, Deoras A, et al. RNNLM — Recurrent neural network language modeling toolkit//Proceedings of the 2011 IEEE Workshop on Automatic Speech Recognition & Understanding. Waikoloa, USA, 2011: 196-201
- [5] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2014: 3104-3112
- [6] Serban I V, Sordani A, Bengio Y, et al. Building end-to-end dialogue systems using generative hierarchical neural network models//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA, 2016: 3776-3784
- [7] Lathia N, Hailes S, Capra L. Temporal collaborative filtering with adaptive neighbourhoods//Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval. Boston, USA, 2009: 796-797
- [8] Ding Y, Li X. Time weight collaborative filtering//Proceedings of the ACM CIKM International Conference on Information and Knowledge Management. Bremen, Germany, 2005: 485-492
- [9] Zimdars A, Chickering D M, Meek C. Using temporal data for making recommendations//Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence. Seattle, USA, 2001: 580-588
- [10] Deerwester S. Indexing by latent semantic analysis. Journal of the Association for Information Science and Technology, 1990, 41(6): 391-407

- [11] Rendle S, Gantner Z, Freudenthaler C, et al. Fast context-aware recommendations with factorization machines//Proceeding of the International ACM SIGIR Conference on Research and Development in Information Retrieval. Beijing, China, 2011; 635-644
- [12] Sharma S. Applied Multivariate Techniques. New York, USA; John Wiley & Sons, Inc. 1995
- [13] Rendle S, Freudenthaler C, Schmidt-Thieme L. Factorizing personalized Markov chains for next-basket recommendation//Proceedings of the 19th International Conference on World Wide Web. Raleigh, USA, 2010; 811-820
- [14] Cheng C, Yang H, Lyu M R, King I. Where you like to go next: Successive point-of-interest recommendation//Proceedings of the International Joint Conference on Artificial Intelligence. Beijing, China, 2013; 2605-2611
- [15] Natarajan N, Shin D, Dhillon I S. Which app will you use next?: Collaborative filtering with interactional context//Proceedings of the 7th ACM Conference on Recommender Systems. Hong Kong, China, 2013; 201-208
- [16] Chen J, Wang C, Wang J. A personalized interest-forgetting Markov model for recommendations//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA, 2015; 16-22
- [17] Kapoor K, Subbian K, Srivastava J, Schrater P. Just in time recommendations: Modeling the dynamics of boredom in activity streams//Proceedings of the 8th ACM International Conference on Web Search and Data Mining. Shanghai, China, 2015; 233-242
- [18] Wang P, Guo J, Lan Y, et al. Learning hierarchical representation model for next basket recommendation//Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. Santiago, Chile, 2015; 403-412
- [19] Sahoo N, Singh P V, Mukhopadhyay T. A hidden Markov model for collaborative filtering. MIS Quarterly, 2012, 36 (4): 1329-1356
- [20] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar, 2014; 1724-1734
- [21] Shen S H, Peng D, Frank R. Manifold learning for multivariate variable-length sequences with an application to similarity search. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(6): 1333-1344
- [22] Mandy K, James G. Spoken Language understanding for a nutrition dialogue system. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25 (7): 1450-1461
- [23] Shang L, Lu Z, Li H. Neural responding machine for short-text conversation//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing. Beijing, China, 2015; 1577-1586
- [24] Mao J, Xu J, Yang Y, et al. Deep captioning with multimodal recurrent neural networks (M-RNN)//Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA, 2015
- [25] Vinyals O, Toshev A, Bengio S, Erhan D. Show and tell: A neural image caption generator//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015; 3156-3164
- [26] Liu Q, Wu S, Wang L, Tan T. Predicting the next location: A recurrent model with spatial and temporal contexts//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA, 2016; 194-200
- [27] Twardowski B. Modelling contextual information in session-aware recommender systems with neural networks//Proceedings of the 10th ACM Conference on Recommender Systems. Boston, USA, 2016; 273-276
- [28] Yu F, Liu Q, Wu S, et al. A dynamic recurrent basket recommendation model//Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. Pisa, Italy, 2016; 729-732
- [29] Liu Q, Wu S, Wang D, et al. Context-aware sequential recommendation//Proceedings of the IEEE 16th International Conference on Data Mining. Barcelona, Spain, 2016; 1053-1058
- [30] Wu S, Ren W, Yu C, et al. Personal recommendation using deep recurrent neural networks in NetEase//Proceedings of the 32nd IEEE International Conference on Data Engineering. Helsinki, Finland, 2016; 1218-1229
- [31] Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann machines for collaborative filtering//Proceedings of the 30th International Conference on Machine Learning. Atlanta, USA, 2007; 791-798
- [32] Ouyang Y, Liu W, Rong W, et al. Autoencoder-based collaborative filtering//Proceedings of the Neural Information Processing-21st International Conference. Kuching, Malaysia, 2014; 284-291
- [33] Zhao Z Q. A novel modular neural network for imbalanced classification problems. Pattern Recognition Letters, 2009, 30(9): 783-788
- [34] Zhao Z Q, Huang D S, Sun B Y. Human face recognition based on multi-features using neural networks committee. Pattern Recognition Letters, 2004, 25(12): 1351-1358
- [35] Freudenthaler C, Gantner Z, Rendle S, et al. BPR: Bayesian personalized ranking from implicit feedback. Computing Research Repository, 2012, 1205(2618): 452-461
- [36] Hidasi B, Karatzoglou A, Baltrunas L, et al. Session-based recommendations with recurrent neural networks. arXiv preprint arXiv:1511.06939, 2015
- [37] Koren Y. Collaborative filtering with temporal dynamics. Communications of the ACM, 2009, 53(4): 89-97
- [38] Xiang L, Yuan Q, Zhao S, et al. Temporal recommendation on graphs via long- and short-term preference fusion//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, USA, 2010; 723-732



**FENG Yong**, born in 1977, Ph. D., professor. His current research interests include big data analysis and data mining, artificial intelligence and big data processing, deep learning and big data retrieval.

**ZHANG Bei**, born in 1991, M. S. His current research interests include intelligent recommendation and neural

networks.

**QIANG Bao-Hua**, born in 1972, Ph. D., professor. His current research interests include big data processing and information retrieval.

**ZHANG Yi-Yang**, born in 1990, M. S. His current research interests include deep learning and big data retrieval.

**SHANG Jia-Xing**, born in 1987, Ph. D., lecturer. His current research interests include artificial intelligence and big data processing.

## Background

This work is a part of the “Big Data Deep Understanding and Personalized Retrieval in Different Fields”, which is mainly supported by the National Key R&D Program of China under Grant No. 2017YFB1402400 and the National Natural Science Foundation of China under Grant No. 61762025. This project addresses four cutting-edge issues and difficulties in Big Data Personalized Retrieval, which involving extract efficient, accurate and complete label; establish orderly, dynamic and scalable multi-dimensional semantic label database; generate, store, represent and match massive retrieval template; recommend and quick rank diverse personalized retrieval results. This project aims to break through rich information dimension of big data, introduce deep learning, integrate ontology-based annotation, and propose effective

incomplete big data filling algorithm and high-speed training algorithm; establish dimensional semantic label cubes and scalable dynamic growth model of label database; integrate label cubes, match user’s retrieval behavior, and propose generation and matching algorithms of massive retrieval behavior templates to track retrieved content real-time and integrate retrieved content on-demand; understand different application fields in depth, establish multidimensional information drivers, integrate social circle-based recommendation model, and propose multidimensional information drivers and social behaviors-based rank and recommendation model to implement personalized, quick, accurate and diverse rank and recommendation.