

# 用户行为序列个性化推荐研究综述

汪菁瑶, 吴国栋, 范维成, 涂立静, 李景霞

(安徽农业大学 信息与计算机学院, 合肥 230036)

E-mail:gdwul120@qq.com

**摘要:**作为缓解信息过载的一种重要方式,推荐系统可以帮助用户从海量信息中快速找到有价值的信息,其应用也越来越广泛.用户行为序列个性化推荐,又简称为序列推荐,主要根据用户与物品交互行为对用户特征进行建模,进而使用不同方法捕捉用户的长期偏好和短期偏好,向用户推荐其可能感兴趣的物品.本文从用户偏好会随时间变化的视角出发,分为用户长期偏好、短期偏好和长短期偏好3个方面,重点探讨了现有用户行为序列个性化推荐研究取得的主要进展;分析了当前用户行为序列个性化推荐研究中存在的冷启动、数据稀疏和噪声干扰等主要问题,并进一步展望了该领域未来主要研究方向.

**关键词:**用户行为序列;序列推荐;长期偏好;短期偏好

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2022)05-0921-15

## Research of Personalized Recommendation Based on User Behavior Sequence

WANG Jing-yao, WU Guo-dong, FAN Wei-cheng, TU Li-jing, LI Jing-xia

(School of Information and Computer, Anhui Agricultural University, Hefei 230036, China)

**Abstract:** As an important way to alleviate information overload, recommendation system can help users quickly find valuable information from massive information. Its application is becoming more and more widespread. User behavior sequence personalized recommendation, also referred to as sequence recommendation, mainly models user characteristics based on the sequential behavior of user interaction with items. Then uses different methods to capture long-term and short-term preferences of users and recommend items that may be of interest to users. From the perspective that user preferences change over time, this paper discusses the main progress of personalized recommendation research based on user behavior sequence from the perspective of user long-term preference, short-term preference and long and short-term preference. The main problems existing in the current research on personalized recommendation based on user behavior sequence, such as cold start, data sparseness and noise interference, are analyzed, and the main research directions in this field are prospected.

**Key words:** user behavior sequence; sequence recommendation; long-term preference; short-term preference

## 1 引言

在信息过载的数字经济时代,用户不得不从海量的信息中寻找自己感兴趣的物品,如何快速且精准地从繁杂的数据中获取有价值的信息已然成为当前社会所关注的热点话题.推荐系统的目的在于帮助用户从海量的数据中,挖掘出用户感兴趣的目标数据.目前,推荐系统在很多领域得到成功应用,包括电子商务(如Amazon、eBay、Netflix、阿里巴巴等)、信息检索(如iGoogle、MyYahoo、百度等)、社交网络(Facebook、Twitter、腾讯等)、位置服务(如Foursquare、Yelp、大众点评等)、新闻推送(如Google News、GroupLens、今日头条等)等各个领域,并取得了良好效果<sup>[1]</sup>.

在现实生活中,用户从进入到离开平台,通常会先后与多个物品产生交互,按照交互发生的先后顺序记录用户对每个物品的行为,这些行为所排列成的序列称为用户行为序列,现

实生活中有大量以序列形式存在着的事物<sup>[2]</sup>.传统的推荐系统主要是基于协同过滤算法和基于内容的推荐,他们大多根据用户和物品之间显性或隐性的交互来建模,得到用户的偏好,倾向于利用用户的历史交互来了解用户的静态偏好.然而,序列中前后不同的交互之间可能存在一定的依赖关系<sup>[3]</sup>,且用户的偏好是动态的,其下一个行为不仅取决于静态的长期偏好,而且在很大程度上还取决于用户的短期偏好.另一方面,序列是由一组按照时间排序的随机变量组成,反映用户偏好的变化趋势,现有研究大多忽略了用户交互之间的顺序性,会导致对用户偏好的建模不准确.与协同过滤和基于内容的传统推荐方法不同,用户行为序列推荐考虑交互过程中的时序信息,利用数据之间的关联信息,有效捕获用户的偏好,提高推荐效果.近年来,序列推荐在学术研究和实际应用中日渐流行起来,每年计算机领域的相关顶级会议(如AAAI, WWW, SIGIR, CIKM等)中都能看到与序列推荐的相

收稿日期:2021-11-09 收修改稿日期:2021-12-10 基金项目:国家自然科学基金项目(31671589)资助;安徽省自然科学基金项目(2108085MF209)资助;安徽省科技重大专项项目(202103b06020013)资助;嵌入式系统与服务计算教育部重点实验室开放基金项目(ESS-CKF2020-03)资助.作者简介:汪菁瑶,女,1997年生,硕士研究生,研究方向为深度学习、序列推荐等;吴国栋,男,1972年生,博士,副教授,CCF会员,研究方向为深度学习、推荐系统和智能商务;范维成,男,1997年生,硕士研究生,研究方向为深度学习、智能推荐等;涂立静,女,1980年生,博士,讲师,研究方向为人工智能、机器学习等;李景霞,女,1976年生,博士,讲师,CCF会员,研究方向为分布式计算、云计算等.

关研究,国内外很多公司(如阿里巴巴,YouTube,美团等)也对用户行为序列推荐系统开展了研究<sup>[4,5]</sup>.同样,序列推荐也应用于多种领域,如音乐推荐<sup>[6]</sup>,电影推荐<sup>[7]</sup>,新闻推荐<sup>[8]</sup>等等.

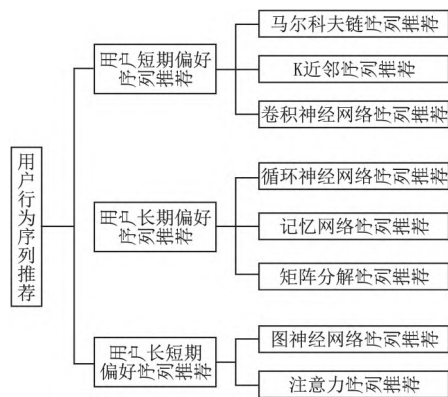


图1 用户行为序列推荐框架图

Fig. 1 User behavior sequence recommendation framework diagram

用户偏好往往随时间变化而变化,本文对现有的序列推荐研究进行较为全面地归纳,整体框架如图1所示.将序列推荐分成3种:用户短期偏好序列推荐、用户长期偏好序列推荐和用户长短期偏好序列推荐.

## 2 用户行为序列推荐概述

### 2.1 用户行为序列相关定义

#### 2.1.1 用户及用户属性

用户是推荐系统研究的对象,是产生行为以及接受推荐结果的主体.通常用 $u$ 表示一个用户,用户集表示为 $U = \{u_1, u_2, \dots, u_{|U|}\}$ , $|U|$ 表示用户的数量.当用户注册登录时,每个用户通常都有一个唯一的身份信息(Identity Document, ID),且每个用户都有一组与其相关的属性,如年龄、性别等,这些属性能够让用户得到更好的推荐.当然,在序列推荐中,用户也可以在匿名状态下进行操作.

#### 2.1.2 物品及物品属性

物品在推荐系统中作为被推荐的对象,是承受行为以及作为推荐结果的实体.通常用 $v$ 表示一个物品,物品集表示为 $V = \{v_1, v_2, \dots, v_{|V|}\}$ , $|V|$ 表示物品的数量.一般来说,不同的物品对应不同的属性,如对于音乐平台来说,物品对应的风格不同的乐曲,而在一些电商平台,物品对应的往往是一件商品.

#### 2.1.3 用户—物品交互及交互属性

用户和物品之间的交互是序列中最基本的构成,可以将用户—物品的交互表示为一个三元组 $\langle u, v, a \rangle$ , $a$ 表示交互行为类型,可以是浏览,点击,购买,评价等,当用户信息不可用或没有时,交互将变成匿名交互,即变成二元组 $\langle v, a \rangle$ .此外,序列中可能只有一种类型的交互,也可能会包含多种类型交互,当只有一种类型交互时,交互中将只包含物品,即 $\langle v \rangle$ .

### 2.2 用户行为序列推荐

用户行为序列推荐是以一组用户和物品的交互行为作为

输入,通常是有序且带有时间戳的,通过对交互序列中复杂的依赖关系建模,预测用户后续可能会产生的交互,输出形式一般是排名前 $n$ 的物品列表,排名可以通过概率、绝对值或者相对排序来确定,大多数情况采用softmax函数生成输出.

序列推荐可通过式(1)所示的公式表示,其中, $S_u$ 为用户行为序列, $f_{Rec}$ 表示用户行为序列的推荐算法, $y_u$ 输出表示推荐给用户 $u$ 的物品列表.

$$y_u = f_{Rec}(S_u) \quad (1)$$

用户行为序列推荐有两种代表性的任务:下个物品(Next-item)推荐和下一个篮子(Next-basket)推荐.下个物品推荐中,只有一个推荐物品,这个物品可以是一首歌、一部电影或一个地点;但下个篮子推荐中则包含多个物品.

### 2.3 长期偏好与短期偏好

用户的偏好可以分成两种,分别是长期偏好和短期偏好.序列推荐中既存在像传统推荐算法一样捕捉到用户的长期偏好,同时也有对用户的短期偏好进行建模,近年来,随着深度学习技术的发展,越来越多的新技术应用到序列推荐中,能够同时考虑到用户的短期偏好和长期偏好,提高了推荐的性能.

用户的长期偏好通常是不易改变的,代表着用户的长期兴趣,一般包含在整个用户行为序列中,通过建模用户对物品的交互历史序列,尽可能地考虑用户行为序列的全局信息,因而能够捕获到用户的长期偏好<sup>[9]</sup>.用户的短期偏好是随时间不断变化的,用户的短期偏好代表的是用户当前兴趣,通常由用户近期的交互行为所反映,通过建模包含当前交互的短序列或者会话,获得用户的短期偏好,这里会话是指一个时间段内发生的一组用户交互行为.

以线上购物场景为例,根据用户与物品的历史交互分析出用户可能是电子产品爱好者,因为他先后购买或点击了不同的电子产品,如手机、电脑等,可以推断该用户长期以来都对电子产品有着浓厚的兴趣,代表着用户的长期偏好.而若用户最近的交互是书籍,这代表着用户最近的兴趣可能是阅读,代表着用户的短期偏好.综合考虑用户的长期和短期偏好,可能会给用户推荐电子书阅读器,更符合该用户的兴趣.因此,结合两种类型的用户偏好为用户进行推荐,可以得到更好的推荐效果.

## 3 用户短期偏好序列推荐研究

### 3.1 马尔科夫链序列推荐

早期的序列推荐主要是依靠马尔科夫链(Markov Chain, MC),使用马尔科夫链对序列中的用户—物品交互数据进行建模,通过转移概率预测下一次交互.一阶马尔科夫链中,观测变量 $v$ 的取值仅依赖于其对应的状态变量 $V$ ,同时, $t$ 时刻的行为仅依赖于 $t-1$ 时刻的行为,与此前的 $t-2$ 个行为无关,即在推荐系统中的下一个交互仅由当前交互决定,不依赖于历史交互.

2005年Shani等人<sup>[10]</sup>首次将马尔科夫链应用在推荐系统当中,但是没有能充分利用用户的上下文信息.针对这个问题,Hosseinzadeh等人<sup>[11]</sup>将用户行为建模为隐马尔可夫链(Hidden Markov Model, HMM)以捕捉用户偏好的变化,并将用户的当前上下文信息建模为模型中的隐变量.同样,Le等

人<sup>[12]</sup>也是将用户行为建模为隐马尔科夫链,不同的是后者考虑了用户动态偏重排放 (Dynamic User-Biased Emissions) 等因素,例如不同用户在购买服装时都会考虑到季节,但是在每个季节所选择的服饰品牌是不同的。He 等人<sup>[13]</sup>和 Feng 等人<sup>[14]</sup>使用一阶马尔科夫链,只考虑最后一次交互行为,而 He 等人在文献[15]中采用高阶马尔科夫链,考虑了最后几次的用户交互行为。虽然在高阶马尔科夫链中考虑了更多的历史交互的依赖关系,但还是无法在一个相对较长的序列中捕捉更多的行为之间的依赖关系,即无法体现用户的长期偏好,除此之外,处理稀疏数据表现也不好<sup>[16]</sup>。

近年来,考虑到用户的偏好是多样的, Eskandarian 等人<sup>[17]</sup>提出了一个基于 HMM 的推荐框架,该框架关注用户行为序列中用户偏好开始发生变化的首次交互位置且考虑用户偏好的多样性,使用 HMM 检测用户偏好变更点,并根据变更点对用户行为序列进行分割,不同片段代表用户的不同偏好,通过矩阵分解将这种检测方法集成到推荐模型中。同样,Shao 等人<sup>[18]</sup>认为用户的历史行为序列中包含多种偏好,将分解用户历史行为序列看成马尔科夫决策过程,提出自适应分解用户偏好来构建序列推荐模型。

### 3.2 K 近邻序列推荐

K 近邻算法 (K-Nearest Neighbor, KNN)<sup>[19]</sup>简单、高效,可解释性强,被广泛应用在序列推荐当中<sup>[20,21]</sup>。从原理上看, K 近邻算法首先从序列数据中找出与当前交互或会话最相似的 K 个交互或会话。然后,根据相似度计算每个候选交互的得分,表示当前交互与历史交互的相似性,根据相似性的大小对用户进行推荐,相似度越高的物品,越有可能推荐给用户。

Linden 等人<sup>[22]</sup>在 2003 年提出了一种算法,通过计算序列中物品之间的相似度,找到与用户当前购买物品最近邻域中的物品。K 近邻相关算法会损失部分可用信息的重要部分,因此文献[21]融合了基于用户和基于物品的两种最近邻算法的优点,提出了统一最近邻协同过滤算法 (K Unified Nearest Neigh, KUNN)。但是该类算法<sup>[23]</sup>均只考虑序列中的最后一个交互,也就是只考虑到了用户的短期偏好。

相比之下,基于会话的 K 近邻推荐应运而生,计算当前会话序列与历史会话序列的相似度,得到 K 个邻居会话序列的集合。与基于物品的 K 近邻算法相比,基于会话的 K 近邻算法考虑的是整个会话序列,而不仅仅只是一个当前交互,因此可以获得更多的信息,进行更准确的推荐。基于会话的 KNN 推荐也衍生出很多变体,若仅仅只考虑最后一次交互物品的最近邻域,不考虑物品在序列中的位置和序列中的其他信息,忽视序列的历史交互,会影响推荐效果,为了解决这个问题,Garg 等人<sup>[24]</sup>提出了一种序列和时间感知领域 (Sequence and Time Aware Neighborhood, STAN) 算法,该算法添加了序列的顺序和时间信息,当前一个会话中,离当前交互越近交互权重应该越高,其中  $p(i, s)$  表示物品  $i$  在会话  $s$  中的位置,  $l(s)$  为会话长度,如公式(2)所示:

$$w_1(i, s) = \exp\left(\frac{p(i, s) - l(s)}{\lambda_1}\right) \quad (2)$$

不同会话中,越早产生的会话权重应该越低,公式见式(3):

$$w_2(s_j | s) = \exp\left(\frac{t(s) - t(s_j)}{\lambda_2}\right) \quad (3)$$

得到的邻域会话  $n$  中,  $i^*$  是同时存在邻域会话和当前会话中的物品,与当前会话距离越近的物品,权重越高,见式(4):

$$w_3(i | s, n) = \exp\left(-\frac{|p(i, n) - p(i^*, n)|}{\lambda_3}\right) I_n(i) \quad (4)$$

此外,现有的循环神经网络不能直接捕获推荐场景中用户购买每件物品的次数, Hu 等人<sup>[25]</sup>提出一种基于物品购买频率的 KNN 算法,充分利用率物品频率信息。RAC 等人<sup>[26]</sup>对简单的基于 KNN 的推荐进行扩展,添加了语义因子,通过语义层的属性更好的适应用户偏好的变化。

### 3.3 卷积神经网络序列推荐

卷积神经网络 (Convolutional Neural Networks, CNN) 通常用于处理时间序列数据,其中典型的结构由卷积层、池化层和前馈全连接层组成,它适合于捕捉序列中局部信息的依赖关系。在序列推荐中,基于卷积神经网络的模型可以很好地捕捉会话中的局部特征,并且还可以在输入层加入时间信息。

Tang 等人<sup>[27]</sup>提出一种卷积序列嵌入推荐模型 (Convolutional Sequence Embedding Recommendation Model, Caser), 结构如图 2 所示,模型由 3 个部分组成,分别是嵌入层、卷积层和全连接层,从序列中提取若干个连续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。卷积神经网络对当前序列的特征进行提取,模型没有直接从左到右单向结构来处理数据,而是整个将作为结构化的信息交给 CNN 来进行特征提取。

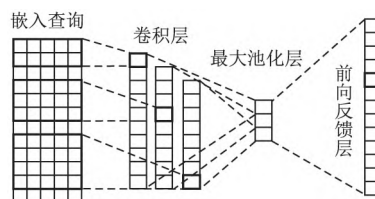


图2 Caser 结构图

Fig. 2 Architecture of caser

Yuan 等人<sup>[28]</sup>对 Caser 模型进行了改进, Caser 模型属于浅层网络,难以捕获复杂的关系以及对长期依赖建模,且在生成下一个物品时, Caser 模型只考虑了最后一个物品的条件概率分布。针对该问题, Hsu 等人<sup>[29]</sup>提出一种网络结构在用户嵌入层和水平卷积层与 Caser 模型非常相似,但没有加入垂直卷积层。文献[30]中进而设计三维卷积神经网络模型 (3 Dimensional Convolutional Neural Network, 3D-CNN), 用一个嵌入矩阵来串联物品身份信息、名称和类别,使用三维卷积神经网络针对序列推荐中的用户—物品交互数据以及内容特征更好地进行建模。2019 年,文献[31]提出一种 2D 卷积序列推荐 (2D Convolutional Network for Sequential Recommendation, CosRec), 将序列中的物品编码成一个三维张量,使用二维卷积滤波器学习序列的局部特征。Yuan 等人<sup>[32]</sup>继续进一步扩展,提出一个可以捕获物品高阶表示的模型,名为间隙填充推荐 (Gap-filling based Recommender, GRec), 主要结构是用层叠的空洞卷积 (Dilated Convolutional) 来扩大表示范围,并且使用剩余块结构优化深度网络, GRec 模型利用基于间隙填充的编码器-解码器框架和掩码卷积操作,同时考虑过去和未来的数据,而不存在 3D-CNN 中的数据泄漏问题,但是 GRec 模



型中使用的数据增强和自回归训练是训练序列数据常用的两种方法,这两种方法只考虑过去的用户行为,存在局限性。

表 1 归纳了本文中用户短期偏好序列推荐中的主要研究,从针对问题和主要优点方面进行了小结。

表 1 用户短期偏好的序列推荐主要研究

Table 1 Research on sequence recommendation based on user's short-term preference

类别	方法	针对问题	主要优点
马尔科夫链序列推荐	Shani 等人 <sup>[10]</sup>	以往的推荐系统都是在历史数据集上进行预测	首次将马尔科夫链应用在推荐上,考虑到当前交互
	Hosseinzadeh 等人 <sup>[11]</sup>	影响推荐效果的不仅仅只有序列本身	预测用户偏好时考虑了动态偏向用户和偏向上下文
	Le 等人 <sup>[12]</sup>	除序列顺序信息影响用户偏好外,还有其他重要因素	考虑到用户信息和上下文信息
	He 等人 <sup>[13]</sup>	传统方法构建针对 Web 查询的推荐较为困难	实时搜索查询引擎中的推荐效果很好
	Feng 等人 <sup>[14]</sup>	充分利用用户序列信息和构建个性化推荐模型较为困难	同时考虑序列信息和用户偏好信息
	He 等人 <sup>[15]</sup>	处理复杂的用户—物品之间的交互存在困难	能很好的处理大型且稀疏的数据集
	Eskandarian 等人 <sup>[17]</sup>	现有方法不能自动检测用户偏好变化点	能识别序列中偏好显著变化点
K 近邻算法序列推荐	Shao 等人 <sup>[18]</sup>	对于不同用户都利用统一的框架来预测,实际效果并不好	缓解数据稀疏,考虑用户的多样化偏好
	KUNN <sup>[21]</sup>	现有的最近邻相关算法会损失部分可用信息的重要部分	融合了基于用户和基于物品的两种最近邻算法的优点
	STAN <sup>[24]</sup>	传统的 K 近邻算法不考虑序列的顺序和时间信息	比起之前的基于 K 近邻算法的序列推荐效果有显著提升
	Hu 等人 <sup>[25]</sup>	现有的序列推荐大都不能直接捕获推荐场景中的物品频率信息	模型结构简单且缓解了循环神经网络算法的局限性
卷积神经网络序列推荐	RAC 等人 <sup>[26]</sup>	现有基于最近邻方法缺乏在单个会话期间检测兴趣突变点的能力	在稀疏数据集上表现好且很好适应兴趣变化
	Caser <sup>[27]</sup>	大部分推荐系统多根据用户的一般喜好推荐物品	网络的结构灵活,可以捕获序列中的重要特征
	Yuan 等人 <sup>[28]</sup>	基于循环神经网络的序列推荐模型的通常依赖整个过去的隐藏状态,训练速度受到限制	缓解了梯度消失的问题
	Hsu 等人 <sup>[29]</sup>	针对歌曲的基于神经网络的序列推荐模型相关研究较少	性能优于经典推荐系统
	3D-CNN <sup>[30]</sup>	现实推荐中,用户信息和历史行为没有充分利用	有效减少了模型的参数
	CosRec <sup>[31]</sup>	序列中的复杂关系难以捕获	学习局部特征且聚合高阶交互
	GRec <sup>[32]</sup>	建模序列时不一定要严格遵守序列中交互的顺序	可以在没有数据流的情况下工作

## 4 用户长期偏好序列推荐研究

### 4.1 循环神经网络序列推荐

循环神经网络 (Recurrent Neural Network, RNN) 是一类用于处理序列数据的神经网络,在建模序列依赖关系方面具有独特优势,使得它在序列推荐中占据了主导地位,大量的关于序列推荐的建模都是基于循环神经网络的。但是 RNN 需要在很长的时间序列的各个时刻重复应用相同操作来构建非常深的计算图,并且模型参数共享,在经过很多阶段传播以后会出现梯度消失或爆炸问题,使得推荐时难以捕获序列之间的长期依赖关系。因而引入了门控循环神经网络,包括基于长短期记忆 (Long Short-Term Memory, LSTM) 和基于门控循环单元 (Gated Recurrent Unit, GRU) 的网络,以更好地模拟长期序列数据。文献[33,34]中使用 RNN 框架从用户的历史行为序列中学习用户的长期兴趣。有实验证明,考虑用户的长期兴趣

对于个性化推荐是非常有价值的,例如,Quadrana 等人<sup>[35]</sup>提出的层次化循环神经网络 (Hierarchical RNN, HRNN) 模型在某些场景下,如用户明确表示的情况时,比文献[36]提出的模型效果高出 3.5%。

LSTM 算法在建模用户序列数据时有很好的效果<sup>[37]</sup>,GRU 算法是 LSTM 的改进算法,它的结构更加简单,后来研究人员多将 GRU 应用在序列推荐中。文献[36]提出了一种基于 GRU 的序列推荐模型 (GRU For Recommendation, GRU4Rec),是早期提出将 RNN 运用在序列推荐上的模型之一,与其他门控 RNN 的主要区别在于,单个门控单元同时控制遗忘因子和更新状态单元的决定,用于预测会话中的下一个用户动作,传统的推荐方法只能处理较短的序列数据,而不能处理较长的用户历史序列,很难获取用户的长期偏好,而 GRU4Rec 模型的在门控循环单元的帮助下对用户序列进行建模,图 3 显示了网络的总体框架,其输入是一组用户行为序

列,将序列中的物品经过独热编码,输入嵌入层进行降维,然后通过多个 GRU 层和前向传播层,输出层通过 softmax 等函数计算每个物品的点击概率。

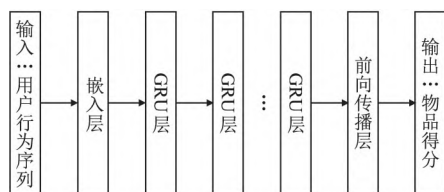


图3 GRU4Rec 结构图

Fig.3 Architecture of GRU4Rec

在 GRU4REC 的基础上,学者们提出了各种优化方案。Hidasi 等人<sup>[38]</sup>考虑了用户行为序列中的更多信息,诸如图像、文本,并设计了新的基于 RNN 的网络模型,称为并行循环神经网络(Parallel RNN, P-RNN)来利用这些信息,实验效果也有了明显提升。2016 年, Tan 等人<sup>[39]</sup>针对 GRU4REC 提出了模型改进方案,主要体现在序列预处理的数据增强,以及嵌入 Dropout 方法增强训练过程,减少过拟合。Hidasi 等人又在文献[40]中对 GRU4REC 算法中提到的采样和损失函数进行优化;而 Bogina 等人<sup>[41]</sup>将用户在序列中物品的停留时间长短考虑进去,停留时间越长,表示用户对该物品越感兴趣。Jannach 等人<sup>[42]</sup>提出将 KNN 算法与 GRU4REC 模型结合,以提高推荐的效果。而文献[35]的贡献在于提出一种层次化的 RNN 模型,捕获会话内和会话间的依赖关系,相比之前的工作,可以刻画会话序列中用户个人的兴趣变化,达到用户个性化的会话序列推荐。

除了对 GRU4REC 模型的优化以外,还有基于 GRU 算法的变体。文献[43]中提出一种分层时序卷积网络(Hierarchical Temporal Convolutional Networks, HierTCN),从跨会话数据中动态学习以预测用户将选择的下一个物品,模型有两个部分组成:高层模型使用 GRU 来捕获用户在不同会话中不断变化的长期兴趣,而低层模型使用时间卷积网络实现,利用会话之间的长期依赖和短期互动,以预测下一次互动。

就仅基于 RNN 的序列推荐而言,当用户序列数据较少或者没有的时候,每一层只有固定的输入矩阵和转移矩阵,不能感知序列中上下文信息的变化推荐效果会变差,针对这一局限性,可以通过建模用户的上下文信息来解决。Song 等人<sup>[44]</sup>提出一种增强循环神经网络(Augmented RNN, ARNN)模型,通过提取高阶的用户上下文偏好,以增强现有的 RNN 序列推荐模型。文献[45]中提出上下文感知循环神经网络模型(Context-aware RNN, CA-RNN),建模丰富的上下文信息,引入了上下文感知的输入矩阵和上下文感知的转移矩阵,使 RNN 每一层的矩阵参数随着输入上下文和转移上下文的不同而变化。类似的, Twardowski 等人<sup>[46]</sup>认为时间也是一种上下文信息,将时间信息与物品信息相结合进行推荐。文献[47]提出预测用户未来行为轨迹的循环推荐网络(Recurrent Recommender Networks, RRN),是通过 RNN 构建自回归模型来实现的,以适应用户的动态性。

Kumar 等人<sup>[48]</sup>提出了循环注意力深度语义结构模型(Recurrent Attention Deep Semantic Structured Model, RA-

DSSM),利用双向 LSTM 有效获取序列中包含的信息。与 RA-DSSM 模型类似,文献[49]提出名为 CHAMELEON 的模型,其同样依赖 RNN 作为模型的基本组成,但是与 RA-DSSM 不同的是,RA-DSSM 中不使用任何关于用户或物品的上下文信息,只依赖用户物品交互存在一定的局限性,限制模型在冷启动场景中推荐的准确性,CHAMELEON 中添加了多种类型的边信息,包括文本上下文信息,如最近流行度等;还有用户上下文信息,如时间、位置等以缓解冷启动问题,而且 CHAMELEON 支持在线学习,训练过程中模拟流场景(Streaming Scenario)。

而 Dai 等人<sup>[50]</sup>首次将时序点过程(Temporal Point Process)与 RNN 相结合,时序点过程有助于保留序列的时间信息和动态变化,不断地迭代更新用户和物品的嵌入表示。2018 年, Vassøy 等人<sup>[51]</sup>提出时序分层循环神经网络(Temporal Hierarchical RNN, THRNN),使用 RNN 分别对会话内和会话间关系建模,而利用时序点过程对交互时间预测。

#### 4.2 记忆网络序列推荐

用户行为序列推荐中的记忆网络(Memory Network)推荐通过引入外部存储器捕获序列中的交互与下一个交互之间的依赖关系,利用记忆组件保存场景信息,以实现长期记忆的功能。

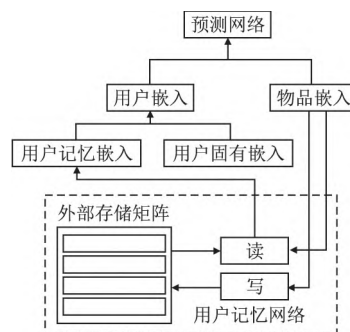


图4 MANN 结构图

Fig.4 Architecture of MANN

记忆网络比起 RNN 更适合建模长期序列,记忆向量的存储避免了直接存储长期行为序列,缓解存储压力,且记忆网络的读操作相较于 GRU 结合注意力的模型来说减小了计算压力。简单地将用户的历史序列信息嵌入到单个向量中,可能会丢失用户历史序列与长期偏好之间的相关信息。Chen 等人<sup>[52]</sup>提出了一种用于序列推荐的记忆增强神经网络(Memory-Augmented Neural Network, MANN),结构如图4所示,该模型通过外部存储矩阵(Memory Matrix)存储用户历史行为序列,并进行有选择的读取用户历史交互,生成用户表征,避免对所有的历史序列进行操作,根据用户历史行为序列对于当前物品的重要性来选择历史交互物品,即计算历史交互物品的权重,权重与历史交互物品嵌入表示相乘以表示用户嵌入表示。从记忆网络中得到的用户嵌入表示与用户固有的属性嵌入表示进行合并得到最终的用户嵌入表示,将用户嵌入表示和物品嵌入表示输入预测网络,以预测用户购买当前物品的概率。

Huang 等人<sup>[53]</sup>进一步扩展文献[52]的工作,单独的





表2 用户长期偏好的序列推荐主要研究  
Table 2 Research on sequence recommendation based on user's long-term preference

类别	方法	针对问题	主要优点
循环神经网络序列推荐	HRNN <sup>[35]</sup>	用户个性化的兴趣难以捕获	层次化的 RNN 模型,可以进行实时推荐
	GRU4REC <sup>[36]</sup>	克服传统序列推荐方法只考虑用户最后一次交互的局限性	结构简单,并行小批量处理加快训练速度
	P-RNN <sup>[38]</sup>	传统方法只考虑交互物品的基本信息而忽略其他信息	将文本和图像加入框架中
	Tan 等人 <sup>[39]</sup>	对 GRU4REC 模型进行优化	减轻训练时的过拟合,加快了训练速度
	Hidasi 等人 <sup>[40]</sup>	优化 GRU4REC 模型的采样环节和损失函数	克服随着负样本量的增加而导致的梯度消失现象
	Bogina 等人 <sup>[41]</sup>	传统方法忽略了交互序列的时间戳信息	考虑用户在物品上的停留时间长短
	Jannach 等人 <sup>[42]</sup>	传统基于 KNN 算法序列推荐无法捕获数据中交互之间的依赖关系	可以较好的捕获数据中的交互关系
	HierTCN <sup>[43]</sup>	CNN 难捕获序列之间的关联性	比传统的基于 RNN 的模型训练速度更快,使用内存减少
	ARNN <sup>[44]</sup>	现有的方法难以建模丰富的用户上下文信息	改进现有基于 RNN 模型,结合用户上下文信息进行个性化推荐
	CA-RNN <sup>[45]</sup>	在实际应用中,传统方法建模情景信息方面存在困难	可以建模复杂的现实环境,如时间、地点和天气
	Twardowski 等人 <sup>[46]</sup>	常用推荐系统假设用户身份是显式	无须直接使用用户信息
	RRN <sup>[47]</sup>	推荐系统通常认为用户信息和物品属性是静态的,不符合实际	可捕获动态的用户和物品信息
	Kumar 等人 <sup>[48]</sup>	新闻推荐中用户兴趣不同且不断变化	有效缓解用户冷启动和物品冷启动
	CHAMELEON <sup>[49]</sup>	以往的 RNN 推荐只使用点击序列作为输入时存在局限性	融入多种上下文信息,显著提高推荐准确性
记忆网络的序列推荐	Dai 等人 <sup>[50]</sup>	很多模型对用户变化过程和物品潜在特征有很强的参数化假设,与现实不符	能够捕获复杂的交互之间的影响以及捕获随时间变化的特征
	Vassøy 等人 <sup>[51]</sup>	尽管 RNN 本身能捕获序列中的时间依赖关系,但很大程度上着重序列中的顺序	分别对会话间和会话内的关系建模以及对用户返回时间预测
	MANN <sup>[52]</sup>	以往的方法会存在减弱相关度高的物品之间的信号	可通过记忆模块来动态更新并捕捉用户偏好的变化
	Huang 等人 <sup>[53]</sup>	大多数方法都忽略了用户关注物品的具体属性,可解释性低	提高准确率的同时还增强了推荐系统的可解释性
	HPMN <sup>[54]</sup>	对于处理用户终身兴趣,现有模型建模是有难度的	可以捕捉多尺度的序列特征
	CSR <sup>[55]</sup>	社区会话中的协作信息没有得到充分利用	两个并行记忆组件同时考虑当前会话和邻居会话
矩阵分解的序列推荐	DMAN <sup>[56]</sup>	传统方法处理长序列复杂度更高,更注重用户的短期偏好	充分利用交互中的信息,擅长追踪用户的长期偏好
	FPMC <sup>[58]</sup>	以往的序列推荐模型很难考虑用户的个性化	缓解数据稀疏性
	CoFactor <sup>[59]</sup>	当用户在平台上的行为较少时,推荐效果较差	面对用户冷启动时,有较好的表现
	FPMC-LR <sup>[60]</sup>	基于位置的社交网络推荐忽略了序列中的时间信息	计算成本降低,减少噪声信息,提高推荐效果
	FISM <sup>[63]</sup>	只描述共现物品的模型,两个物品没有同时被购买则相似度为零	有效地处理稀疏数据集
	FOSSIL <sup>[64]</sup>	现实数据普遍存在数据稀疏性的问题	用户冷启动和长尾问题得到缓解
	SSRM <sup>[65]</sup>	针对用户行为不确定性和会话数据特征	缓解数据量大以及高速的问题

和物品的嵌入表示. Wu 等人<sup>[67]</sup>首次将 GNN 使用在序列推荐中,提出基于会话的图神经网络推荐(Session-based Recommendation with GNN, SR-GNN),模型整体框架如图 6 所示,将会话序列建模成图结构数据,使用 GNN 提取物品的嵌入向量,在经过注意力层后将长期偏好和短期偏好组合,预测点击下一个物品的可能性.

与 SR-GNN 等模型相比,Wang 等人<sup>[68]</sup>提出一种全局上下文增强的图神经网络(Global Context Enhanced Graph Neu-

ral Networks, GCE-GNN),分别从当前会话图和全局图中学习两个层次的物品嵌入向量,考虑了全局信息的影响,利用近邻节点进行图构建,并在具体的建模时通过合理的设计来最大化利用相关物品的信息,降低噪声物品的影响.与 GCE-GNN 模型不同,Qiu 等人<sup>[69]</sup>提出全图神经网络(Full Graph Neural Network, FGNN)模型为每个会话序列构建一个会话图,并且提出多重加权图注意力网络来计算会话序列中物品之间的信息流,获得物品的嵌入表示.基于图神经网络的序列

推荐存在信息丢失问题,Chen 等人<sup>[70]</sup>提出的一种处理信息丢失(Lossless Edge-order Preserving Aggregation and Shortcut Graph Attention for Session-based Recommendation, LESSR)的模型针对这一问题进行了优化,首先,从会话序列到图的有损编码以及聚合消息传递期间不变的排列,会忽略有关会话转换的一些顺序信息;其次,由于层数的限制,无法捕捉会话中的长期依赖性.为了解决第一个问题,提出了一种无损编码方案和基于 GRU 的边缘顺序保留聚合层,该层专门设计用于处理无损编码图.为了解决第 2 个问题,提出了一个捷径图关注层,该层通过沿捷径连接传播信息来有效地捕获远程依赖关系.通过将两种类型的层组合在一起,能够缓解信息丢失问题并且在 3 个公共数据集上胜过最新模型.

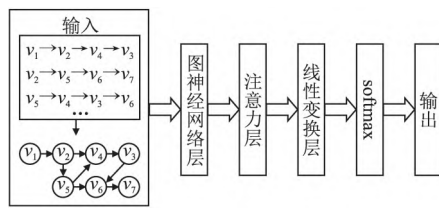


图 6 SR-GNN 结构图

Fig. 6 Architecture of SR-GNN

Song 等人<sup>[71]</sup>提出一种基于动态图神经网络的序列推荐,同时建模用户的动态兴趣和用户的社交关系,将两者结合起来,通过图注意力网络来捕捉社交关系对用户的动态影响,得到的向量和用户向量拼接得到最后的用户向量,最后得到下一个可能的物品的概率分布进行推荐.而动态图注意力神经网络(Dynamic Graph Attention Networks, DGAN)<sup>[72]</sup>构建了一个动态注意模块来捕获全局信息,并进一步提出了一个递归神经模块来学习局部动态表示,通过堆叠多个动态图注意力模块,可以更好地利用每个序列的丰富上下文信息. Kumar 等人<sup>[73]</sup>考虑到静态图的局限性,即无法建模用户节点或物品节点的变化,提出一种融合用户-物品动态嵌入表示的模型(JOINT DYNAMIC USER-ITEM EMBEDDING MODEL, JODIE),JODIE 将用户和物品的表示分为静态嵌入和动态嵌入,静态嵌入表示用户和物品的长期属性,而动态嵌入表示了用户随时间不断变化的属性,根据用户的交互序列动态更新嵌入表示,建模用户和物品的动态属性.

利用社交信息是提高推荐质量的一种重要方法,现有方法忽略了跨用户交互之间的时间关系,基于此 Li 等人<sup>[74]</sup>提出社交时间激励网络(Social Temporal Excitation Networks, STEN),引入时序点过程,通过利用时间信息对相关跨用户交互的影响进行建模,构建社交异质图从社交关系结构和交互中提取用户和物品的节点表示. Chen 等人<sup>[75]</sup>同样引入时序点过程用于缓解时间信息丢失问题,设计了一种动态表示学习模型序列推荐(Dynamic Representation Learning model for Sequential Recommendation, DRL-SRe),将整个时间轴分割成等长的时间片,对于每个时间片,构建用户-物品交互图以获得更好的动态用户和物品的嵌入表示.

## 5.2 注意力机制序列推荐

注意力机制最早起源于 Bahdanau 等人<sup>[76]</sup>的研究,并将其应用于机器翻译任务中,它可以捕获语言中的长期依赖,并

提供更流畅的翻译,强调输入的各部分对输出的影响程度不一样,即关注输入句子不同部分对输出词的重要性,在序列推荐系统中广泛使用的注意力机制已经也显示出巨大的成功.在此基础上,提出了将普通注意力(Vanilla Attention)作为循环神经网络的解码器,在序列推荐中得到了广泛的应用<sup>[77]</sup>.另一方面,机器翻译模型 Transformer<sup>[78]</sup>中的自注意力机制(Self-attention Mechanism)也被应用在序列推荐中,Transformer 模型由两部分组成,分别是多头自注意力(Multi-head Self-attention)和全链接前馈网络(Fully Connected Feed-forward Network)部分,与普通注意力机制相比,自注意力不包括循环神经网络结构,但在推荐中的效果好于普通注意力机制.

### 5.2.1 普通注意力机制序列推荐

Li 等人<sup>[79]</sup>提出的神经注意力推荐模型(Neural Attentive Recommendation Machine, NARM),原理是构建一个编码器-解码器框架,作者认为在之前的序列推荐中没有强调用户主要目的,所以将 RNN 与普通注意力结合,以捕获序列中的用户主要目的,NARM 模型的框架如图 7 所示,其中输入  $x = [x_1, x_2, \dots, x_{t-1}, x_t]$  为用户点击会话,  $x_i (1 \leq i \leq t)$  为  $m$  个物品中用户所点击物品项的索引,输出  $y = [y_1, y_2, \dots, y_{m-1}, y_m]$  为用户可能点击的下一个物品项排名列表,  $y_j (1 \leq j \leq m)$  对应的是物品  $j$  的得分,编码器将输入的点击序列  $x$  转化成高维隐藏表示  $h = [h_1, h_2, \dots, h_{t-1}, h_t]$ ,其与  $t$  时刻的注意力信号  $\alpha_t$  一起输入会话特征生成器中,注意力信号决定在  $t$  时刻应该强调或者忽略那个物品,然后构建当前会话表示  $c_t$ ,并在  $t$  时刻输入解码器,通过解码器中的矩阵,并使用激活函数得到输出排名.

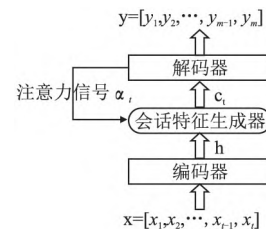


图 7 NARM 结构图

Fig. 7 Architecture of NARM

文献[80]对 NARM 模型进行了改进,调整了 NARM 模型的框架,增加了序列的上下文信息等优化方法,提高了模型的性能. Liu 等人<sup>[77]</sup>设计了短期注意力记忆优先(ShortTerm AttentionMemory Priority, STAMP)模型,提出一种新的注意力机制,根据会话序列的上下文计算注意力权重,同时捕捉用户的长期偏好和短期偏好. STAMP 模型中需要分别获取用户的长期偏好和短期偏好,  $m_s$  为用户点击序列的平均值,  $m_t$  为用户的最后一次点击,表示用户的当前兴趣,如公式(8)、公式(9)所示.

$$m_s = \frac{1}{t} \sum_{i=1}^t x_i \quad (8)$$

$$m_t = x_t \quad (9)$$

而用户的长期偏好是通过简单的前馈神经网络(Feed-forward Neural Network, FNN)得到的,向 FNN 中输入  $m_s$  和  $m_t$ ,得到每个物品的注意力,物品  $x_i$  的注意力系数  $\alpha_i$  计算公



式如式(10)所示,进而可以计算出用户的长期偏好  $m_a$ ,如公式(11)所示:

$$\alpha_i = W_0 \sigma(W_1 x_i + W_2 m_i + W_3 m_s + b_a) \quad (10)$$

$$m_a = \sum_{i=1}^t \alpha_i x_i \quad (11)$$

其中  $W_0$  为权重向量,  $W_1, W_2, W_3$  为权重矩阵,  $b_a$  为偏置向量,  $\sigma(\cdot)$  为 sigmoid 函数. 分别将  $m_a$  和  $m_i$  输入多层感知机(Multilayer Perceptron, MLP)中进行处理,以达到提取特征的目的,得到的隐表示向量分别为  $h_s, h_i$ . 对于每个候选物品,计算得分公式如公式(12)所示:

$$\hat{z}_i = \sigma(\langle h_s, h_i, x_i \rangle) \quad (12)$$

其中,3个向量  $h_s, h_i, x_i$  进行点积运算,  $\hat{z}_i$  表示了非归一化的余弦相似度,即物品  $x_i$  与用户一般兴趣和当前兴趣的相似度,最后利用 softmax 函数进行归一化处理,得到输出结果.

Wang 等人<sup>[81]</sup>提出基于注意力的事务上下文嵌入模型(Attention based Transaction Embedding Model, ATEM)中将注意力机制整合到浅层的网络中,构建没有严格序列关系的上下文表示,由于注意力机制的有效性,该模型能够关注更多相关的物品,而较少关注不相关的物品. NARM<sup>[79]</sup>、STAMP<sup>[80]</sup>和 ATEM<sup>[81]</sup>将普通的注意机制与 RNN 或多层神经网络结合起来,目的是在给定的会话中捕捉用户的主要目的.

除此之外, Ren 等人<sup>[82]</sup>考虑到重复消费的问题,提出了基于重复消费的网络模型(Repeat Consumption with Neural Networks, RepeatNet),将重复探索机制与注意力机制融入到 RNN 中. Yu 等人在文献[83]中提出一种多注意力排序模型,从多个角度将个体和联合级别的交互统一到偏好推理模型中,而 Ying 等人<sup>[84]</sup>提出了两层的注意力网络,第1层是根据历史购买的物品表示来学习用户的长期偏好,第2层通过耦合用户的长期和短期偏好来输出最终的用户表示. Wu 等人<sup>[85]</sup>提出基于会话的上下文感知注意力网络(Session-based Recommendation with Context Aware Attention Network, SR-CAAN),通过将序列预测与会话外部上下文感知方法相结合,增强了对用户偏好进行建模的能力,将外部知识与知识图谱(Knowledge Graph)相结合,通过使用注意力机制来获取会话的外部上下文,每个会话都作为会话外部上下文的组合呈现.

### 5.2.2 自注意力机制序列推荐

绝大部分的序列处理模型都采用编码器-解码器结构,其中编码器将输入序列  $(x_1, x_2, \dots, x_n)$  映射到连续表示  $\hat{z} = (z_1, z_2, \dots, z_n)$  中,然后解码器生成一个输出序列  $(y_1, y_2, \dots, y_m)$ ,每个时刻输出一个结果,Transformer 模型延续了这个模型,自注意力机制是 Transformer 模型的第1层,受到 Transformer 中的自注意力机制的启发,越来越多的学者将自注意力机制应用到序列推荐当中. 本质上,对于每个输入向量,自注意力机制产生一个向量,该向量在其邻近向量上加权求和,其中权重由物品之间的关系决定,自注意力机制在序列推荐中也获得了越来越多的关注. 自注意力网络的一个优点是计算一个序列中每对物品之间的注意权重来捕获长期依赖关系,很容易并行化并且捕捉长期依赖.

文献[86]中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. Zhang 等人<sup>[87]</sup>利用自注意力机制从用户的历史交互中推断出物品与物品之间的关系,有了自注意力机制,用户交互序列中每个物品的权重就能更好地代表用户的兴趣,模型最后使用度量学习框架进行训练,既考虑了短期特征又考虑了长期特征.

Li 等人<sup>[88]</sup>在 SASRec 模型的基础上,提出了感知时间间隔的自注意力序列推荐模型, (Time Interval Aware Self-attention based Sequential Recommendation, TiSASRec),不仅考虑物品的绝对位置,还考虑了序列中物品之间的时间间隔,结合绝对位置与相对时间间隔编码的优点,设计了一个具有时间间隔感知的自注意力机制去学习不同物品的绝对位置以及时间间隔的权重,以此进行未来物品的预测. 文献[89]同样也在 SASRec 模型的基础上提出改进,由于用户行为序列与文本序列存在相似性,作者使用了自然语言处理中的 BERT (Bidirectional Encoder Representation from Transformers) 模型,若将 BERT 模型直接使用在推荐上,会产生数据泄露问题,为解决此问题引入了填空任务(Cloze Task),对输入的序列进行随机掩码(Mask),根据上下文信息,预测掩码物品. BERT4Rec (BERT for Recommendation) 模型结构如图8所示,其中包含了 L 个 Transformer 层,对于物品  $v_i$ , Transformer 第1层输入为  $h_i^0$ ,由物品嵌入和位置嵌入之和表示,然后 Transformer 每一层迭代利用前一层输出的信息,经过 L 层的信息处理后,得到所有物品的最终表示,在第 t 步掩盖掉物品的  $v_t$ ,可以根据  $h_t^L$  对其进行预测. 序列中行为的顺序性对于实际生活中的用户行为是不合适的,基于 BERT 的推荐相较于基于 RNN 的推荐,其中的自注意力机制可以捕获任意位置的信息,相较于基于 CNN 的推荐模型,可以捕获全局的信息.

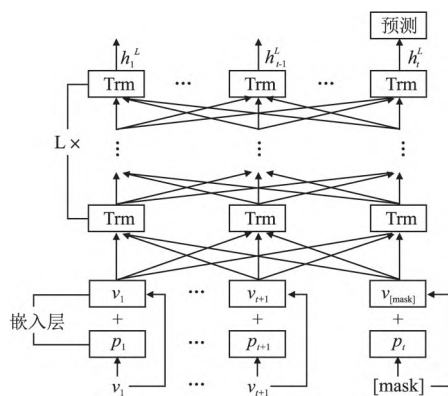


图8 BERT4Rec 结构图

Fig. 8 Architecture of BERT4Rec

文献[90]中提出一种序列深度匹配模型(Sequential Deep Matching, SDM),通过多头注意力机制在用户近期行为序列中捕获短期偏好;把用户长期行为序列中的所有物品对应的属性集合划分为不同的类,如 ID 集,物品店铺集,物品品种集,对于不同的集,都经过一层注意力机制进行建模,例如,每个用户对不同类别的书籍偏好程度不同,以此捕获用户的长期偏好. Wu 等人<sup>[91]</sup>认为很多序列推荐的处理时间信息上,基于 RNN, CNN, Transformer 的模型虽然效果不错,但是缺少个性化,为解决这种问题,提出了一种个性化的基于 Transformer 的序列推荐模型,考虑到用户嵌入在序列建模中的重

要性,并且使用的正则化技术是 SSE (Stochastic Shared Embedding), SSE 的主要思想是在 SGD 过程中随机地用另一个具有一定概率的嵌入来代替现在的嵌入,从而达到正则化嵌入层的效果,提高推荐结果的准确率. 文献[92]是阿里巴巴公司和北京大学联合提出的基于自注意力机制的用户异构行为建模框架,其命名为 ATRank (Attention-Based Rank),同时

考虑用户异构行为和时序,使用自注意力机制缓解 RNN、CNN 的局限性,提升推荐效果. Zhang 等人<sup>[93]</sup>提出一种基于特征集深自注意力网络 (Feature-level Deeper Self Attention Network, FDSA), 分别对物品级序列和特征级序列应用自注意力块,对物品转换模式和特征转换模式分别进行建模,将两个块的输出拼接起来进行预测推荐. 此外, Huang 等人<sup>[94]</sup>设

表 3 用户长短期偏好的序列推荐主要研究

Table 3 Research on sequence recommendation of users' long and short-term preferences

类别	方法	针对问题	主要优点
图神经网络序列推荐	SR-GNN <sup>[67]</sup>	以往的方法无法获得准确的用户向量,且忽略了物品之间的复杂关系	可以捕获复杂的物品关系的转换
	GCE-GNN <sup>[68]</sup>	现有的基于会话推荐模型大多只利用当前会话进行建模	考虑了全局信息的影响
	FGNN <sup>[69]</sup>	只依靠交互物品之间关系建模用户嵌入是不充足的	综合考虑了序列交互的顺序和交互的潜在关系
	LESSR <sup>[70]</sup>	基于 GNN 的序列推荐会出现信息丢失的问题	可以捕获任意长度序列的长期依赖和缓解信息丢失问题
	Song 等人 <sup>[71]</sup>	动态的用户兴趣和用户的社交关系的重要性被忽略	有效捕获用户不断变化的偏好
	DGAN <sup>[72]</sup>	忽略了不同用户序列之间动态协作信号细粒度的利用	实现了不同用户序列之间动态协作关系的编码
	JODIE <sup>[73]</sup>	主要缺陷在于只有用户发生变化时才会对其表示进行更新	能够学习用户在未来时间点的嵌入表示
	STEN <sup>[74]</sup>	以往方法忽略跨用户行为序列的时序关系	增强了时间信息的传播,融入用户的社交信息
注意力机制序列推荐	DRL-SRe <sup>[75]</sup>	现有大多研究只注重用户端信息而忽略物品端信息	能捕获细粒度的时间信息,缓解信息丢失问题
	NARM <sup>[79]</sup>	在基于会话的推荐中,一般只考虑物品本身的属性	考虑了序列中的时间信息,强调了用户的主要目的
	STAMP <sup>[80]</sup>	传统的循环神经网络结构不能同时捕获长期和短期兴趣	提出短期记忆优先机制,能更敏锐的感知用户偏好变化
	普通注意力机制 ATEM <sup>[81]</sup>	一个长序列经常包含很多余且与下次选择无关的交互	忽略了序列的顺序且能衡量序列中不同物品的权重
	RepeatNet <sup>[82]</sup>	以往的研究中没有强调用户的重复消费的模型	可以很好的捕获序列中的用户重复消费的意图
	Yu <sup>[83]</sup>	大多数现有工作只能捕获物品之间的关系	通过建模用户—物品交互行为,捕捉异构物品关系
	Ying <sup>[84]</sup>	用户长期偏好随时间变化的特征被忽略	能反映用户长期偏好的动态特征
	SR-CAAN <sup>[85]</sup>	忽略了物品与会话外部上下文的关系	增强了建模用户偏好能力
	SASRec <sup>[86]</sup>	较难在稀疏和密集数据集上同时效果好	在稀疏和密集数据集中的表现都优于以往的方法
	Zhang 等人 <sup>[87]</sup>	很多模型在稀疏数据上表现不好	能获取序列中物品的相对权重,在稀疏数据集表现良好
	TiSASRec <sup>[88]</sup>	序列中交互时间间隔信息没有充分利用	同时建模了物品在序列中的绝对位置和交互时间间隔
	自注意力机制 BERT4Rec <sup>[89]</sup>	对序列单向建模限制了序列隐表示能力	融合序列中物品可以接受来自双边的信息
	SDM <sup>[90]</sup>	现有方法难以捕获一个会话中的多个兴趣倾向	做到长期偏好和短期偏好的有效融合
	Wu 等人 <sup>[91]</sup>	SASRec 模型没做到个性化推荐	可解释性强,处理长序列有优势
	ATRANK <sup>[92]</sup>	下游应用建模用户特征时,会引入与行为不相关的噪声	考虑了用户的异构行为
	Zhang 等人 <sup>[93]</sup>	当前的方法通常只考虑物品交互间关系,忽略物品特征的关系	通过特征间的关系提高了推荐性能
	CSAN <sup>[94]</sup>	传统方法无法区分个体行为的重要性,效率低,灵活性差	动态建模上下文依赖关系,考虑用户异构行为和行为多义性

计的上下文自注意力网络 (Contextual Self-Attention Network, CSAN) 也考虑到用户异构行为,使用自注意力机制捕获多种

用户行为。

表3归纳了本文中用户长短期偏好序列推荐中的主要研究,从针对问题和主要优点方面进行了小结。

## 6 用户行为序列推荐研究存在的主要问题

### 6.1 冷启动和数据稀疏问题

冷启动问题即用户冷启动、物品冷启动和系统冷启动,而数据稀疏是缺少可用信息,这两个问题都是序列推荐长期要面对的问题,大多数现有的序列推荐算法<sup>[95-97]</sup>都忽略了这两个问题,其有效实现都依赖于对序列数据相对严格的要求,例如,丢弃小于最小阈值的序列等。用户的行为序列是将用户与物品之间的交互作为一组历史信息,交互中可用的用户和物品信息有限,Lv等人<sup>[90]</sup>利用序列中不同类型的边信息构建算法,如物品ID、品牌和店铺等,更好地模拟用户的长期偏好,这种方法也只能部分缓解这两个问题。

### 6.2 噪声信息干扰问题

一般来说,序列越长包含的信息量越多,同时与交互无关的噪声信息也越多,甚至在捕获用户长期偏好时,可能会被长序列中的噪声所淹没。如何对含噪声的序列进行建模,是序列推荐所面临的主要问题之一。目前的研究强调序列中的每个物品都与邻域的物品存在依赖关系,这种严格的规定,会影响推荐的效果,低估了依赖关系建模时噪声信息的影响,在现实生活中也是不现实的,因为用户的行为是不确定的。序列中会包含一些不相关的物品,这些物品会导致错误的依赖关系。受到注意力机制的启发,Ying等人<sup>[84]</sup>提出两层注意力网络,使用注意力为不同的物品不同的权重,更加强调相关物品,减少噪声物品的影响。另一个处理含噪声序列的方法是记忆网络,外部记忆网络有强大的存储能力,Chen等人<sup>[50]</sup>提出的记忆增强神经网络,使用一个内存矩阵存储之前状态,更好的模拟用户行为,提高推荐性能。

### 6.3 可解释性问题

大多数基于深度学习的序列推荐模型都缺乏可解释性,它们的决策过程是一个超出人类直接理解范畴的高维隐空间。如果缺乏推荐的原因,用户会对推荐结果产生怀疑,从而使推荐效果变差。另一方面,增强可解释性可以加强使用者对模型的理解<sup>[98]</sup>,如不同的超参数等因素是如何影响模型的输出,从而增强对整个推荐过程的控制。Peake等人<sup>[99]</sup>用矩阵分解在模型的输出上训练关联规则,以获得更好的可解释性。这类通过设计单独的方法来试图解释推荐产生的结果,没有直接将可解释性与边信息联系起来,而用户与物品的交互中,通常还包含一些辅助信息,如文本评论、社交关系等,这些信息都可以加强推荐的可解释性,但是却没有得到充分的利用。

### 6.4 嵌入方法过于简单

在以往的研究中,序列推荐大多采用自然语言处理的嵌入方法,通过嵌入方法以表示物品、用户或序列的信息,例如,Greenstein等人<sup>[100]</sup>使用词嵌入方法 GloVe<sup>[101]</sup>和 Word2Vec<sup>[102]</sup>得到物品嵌入表示。其中存在着一些问题,首先,利用物品之间的关系来学习物品的表示,然而交互物品之间的依赖关系是不断发生变化的,自然语言处理中的单词及其连接关系是相对固定的。其次,嵌入模型的预训练是相对复杂;而且序列推荐中的行为比单词要复杂的多,包含更多的信息,简单的嵌入

方法无法很好处理复杂用户行为序列。最后,对于通过嵌入方法学习到的用户表示而言,是通过相对静态的方式学习得到,难以体现出用户偏好的动态性。在这些情况下,序列推荐需要设计更先进、更特殊的嵌入方法。

### 6.5 未充分利用用户间协同信息

现有模型大都没有明确的利用不同用户之间的协同信息,也就是说,大多模型都侧重于编码每个用户自己的交互序列,训练和测试过程中对序列的编码都在一个序列内,忽略了不同用户序列之间的高阶连通性,然而在现实生活中,用户的偏好会受到好友的影响。同时,物品的语义信息也会随着一阶或者更高阶的相关性变化而变化,用户的兴趣也是动态的,这些模型也忽略了高阶协同信息在不同时刻的动态影响,推荐效果会受到影响。

## 7 用户行为序列推荐未来主要研究方向

### 7.1 上下文感知用户行为序列推荐

时间、天气、地点、季节、一天中的时间和地点等上下文因素可能会对用户的最终选择产生重大影响。上下文信息可以从序列中显式提取,也可以隐藏在用户的交互历史中。识别上下文是序列推荐中重要的参数之一,它在更好地帮助理解用户的需求和兴趣,在做出正确的预测方面发挥着重要作用。线上购物环境下,用户的购物目的或者当前心情等这些上下文因素是不能直接观察到的,必须从用户最近的行为中去分析,并最终基于用户和整个社区的行为模式<sup>[103,104]</sup>。对于新用户和匿名用户,考虑上下文信息尤其重要。总的来说,上下文感知用户行为序列推荐根据历史交互数据做出上下文适应的推荐,了解用户的兴趣和目的。

### 7.2 动态图神经网络用户行为序列推荐

受图神经网络在各种领域里成功应用的启发,研究人员也不断提出基于图神经网络的用户行为序列推荐<sup>[67,69]</sup>,将用户和物品关联起来构成一个图结构,通过 GNN 捕获复杂的交互关系,并相应生成物品的嵌入向量,这是传统序列推荐难以做到的。同时,时间在多方面影响着用户的兴趣偏好<sup>[105]</sup>,用户偏好会随时间而发生动态变化,随着交互物品的变化而变化,且会受到其他用户的影响,为克服这些挑战,可以将用户行为序列和动态协同信息建模到一个框架中,通过动态图结构将不同的用户序列连接起来,比起传统的静态图,动态图包含更多的时间和顺序信息,利用这些信息探索用户和物品的交互行为以提取用户偏好,将序列推荐的下一项预测任务转化为动态图中用户节点与物品节点之间的链接预测。

### 7.3 跨域用户行为序列推荐

通常用户购买的物品来自多个领域,而仅仅不是单一领域<sup>[106]</sup>。例如,用户阅读了《哈利波特》的书籍,他可能会对《哈利波特》的电影产生兴趣,以及用户可能因此购买哈利波特的周边产品,这些都是属于不同领域,且这些不同领域之间也存在着密切的联系,虽然不同领域物品之间的关系也要比单一领域要复杂很多,但是考虑多领域交互可以缓解单邻域数据稀疏的问题。跨领域推荐的主要问题就是如何利用源领域的的数据在目标领域产生理想的推荐,从序列数据中提取出属于不同领域的特征。然而,单独针对某一领域设计推荐模型的研究较少,大多数的算法都是假设其模型适用于所有领域,在



实际应用中却暴露出很多不足. 未来的研究可以考虑研究不同领域中序列数据的特征, 设计更好的序列推荐模型. 如, Zhuang 等人<sup>[107]</sup>利用用户来自不同领域的行为序列数据, 挖掘其特征以提高推荐性能. Ma 等人<sup>[108]</sup>通过识别同一账户下不同的用户行为, 将用户在两个域上的行为在每个时间戳上同步共享, 从不同领域获得相同时间段的用户行为数据, 提出同时生成两个域的推荐. 但是利用另一个域的用户行为信息同时也包含噪声数据, 因此跨域推荐中依旧存在着很多挑战.

#### 7.4 注意力机制多兴趣用户行为序列推荐

用户的兴趣越来越趋向于多样化且易变, 若仅仅只预测一类物品或一种偏好, 用户的选择会受到局限, 一些相对比较小众的物品就很难被发现. 序列中可能存在多个兴趣倾向, 基于深度学习的方法<sup>[109]</sup>通过对用户行为序列进行计算, 为每个用户生成一个低维度的嵌入表示, 但是单一的嵌入表示缺乏表现力, 并不能反映用户在一个时间段内的多重兴趣. 可以通过注意力机制对用户潜在的多重兴趣建模, 捕获用户的多种兴趣, 注意力机制还能有选择地保留和利用那些与下一次交互预测真正相关的交互信息, 以达到减少噪声的目的, 同时增加了模型的可解释性. 如 SDM<sup>[90]</sup>模型通过一个多头自注意力模块对用户行为序列进行编码, 以捕获多种类型的兴趣; Zhou 等人<sup>[5]</sup>提出用注意力机制使用户在不同的物品上有不同的表示, 以捕捉用户多样的兴趣, 但是这种方式需要重新计算每个物品的用户表示, 使得其只能适用于排序阶段. Pi 等人<sup>[110]</sup>通过记忆网络从长序列数据中捕获用户的多样兴趣, 使用记忆网络比起注意力机制更为复杂, 且不易捕获用户的长期兴趣.

## 8 结束语

用户行为序列在大多数情况下都是存在时间上的先后关系的, 且用户的偏好会随时间的变化而变化, 通过研究用户行为序列推荐领域内的进展, 本文阐述了序列推荐中的相关概念, 从时间维度将该领域内的主要工作划分为 3 个子领域, 分别为长期偏好序列推荐、短期偏好序列推荐和长短期偏好序列推荐, 深入分析了序列推荐中存在的冷启动问题、数据稀疏和可解释性差等问题, 最后展望了序列推荐未来的发展方向以及面临的挑战, 希望能为推进该领域的进步提供一定的帮助.

#### References:

- [1] Huang Li-wei, Jiang Bi-tao, Lv Shou-ye, et al. Survey on deep learning based recommender systems[J]. Chinese Journal of Computers, 2018, 41(7): 1619-1647.
- [2] Ren Shou-gang, Zhang Jing-xu, Gu Xing-jian, et al. Overview of feature extraction algorithms for time series[J]. Journal of Chinese Computer Systems, 2021, 42(2): 271-278.
- [3] Wang S, Hu L, Wang Y, et al. Sequential recommender systems: challenges, progress and prospects[C]//International Joint Conference on Artificial Intelligence, 2019: 6332-6338.
- [4] Cen Yu-ku, Zhang Jian-wei, Zou Xu, et al. Controllable multi-interest framework for recommendation[P]. Knowledge Discovery & Data Mining, 2020.
- [5] Zhou Guo-ru, Zhu Xiao-qiang, Song Chen-ru, et al. Deep interest network for click-through rate prediction[P]. Knowledge Discovery & Data Mining, 2018.
- [6] Bonnin G, Jannach D. Automated generation of music playlists: survey and experiments[J]. ACM Computing Surveys (CSUR), 2014, 47(2): 1-35.
- [7] Zhao W, Wang B, Yang M, et al. Leveraging long and short-term information in content-aware movie recommendation via adversarial training[J]. IEEE Transactions on Cybernetics, 2019, 50(11): 4680-4693.
- [8] Ma M, Na S, Wang H, et al. The graph-based behavior-aware recommendation for interactive news[J]. Applied Intelligence, 2021: 1-17.
- [9] Yuan Tao, Niu Shu-zi, Li Hui-yuan. A multi-scale temporal dynamic model for sequential recommendation with clockwork RNN[J]. Journal of Chinese Information Processing, 2020, 34(6): 97-105.
- [10] Shani G, Heckerman D, Brafman R I, et al. An MDP-based recommender system[J]. Journal of Machine Learning Research, 2005, 6(9): 1265-1295.
- [11] Hosseinzadeh Aghdam M, Hariri N, Mobasher B, et al. Adapting recommendations to contextual changes using hierarchical hidden Markov models[C]//Proceedings of the 9th ACM Conference on Recommender Systems, 2015: 241-244.
- [12] Le D T, Fang Y, Lauw H W. Modeling sequential preferences with dynamic user and context factors[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, Cham, 2016: 145-161.
- [13] He Q, Jiang D, Liao Z, et al. Web query recommendation via sequential query prediction[C]//IEEE 25th International Conference on Data Engineering, IEEE, 2009: 1443-1454.
- [14] Feng S, Li X, Zeng Y, et al. Personalized ranking metric embedding for next new poi recommendation[C]//24th International Joint Conference on Artificial Intelligence, 2015: 2069-2075.
- [15] He R, Fang C, Wang Z, et al. Vista: a visually, socially, and temporally-aware model for artistic recommendation[C]//Proceedings of the 10th ACM Conference on Recommender Systems, 2016: 309-316.
- [16] Florent Garcin, Christos Dimitrakakis, Boi Faltings. Personalized news recommendation with context trees[C]//Proceedings of the 7th ACM Conference on Recommender Systems, 2013: 105-112.
- [17] Eskandarian F, Mobasher B. Modeling the dynamics of user preferences for sequence-aware recommendation using hidden Markov models[C]//32nd International Flairs Conference, 2019.
- [18] Shao W, Chen X, Zhao J, et al. Sequential recommendation with adaptive preference disentanglement[EB/OL]. <https://arxiv.org/abs/2112.02812>, 2021.
- [19] Cover T, Hart P. Nearest neighbor pattern classification[J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.
- [20] Jannach D, Ludewig M. When recurrent neural networks meet the neighborhood for session-based recommendation[C]//Proceedings of the 11th ACM Conference on Recommender Systems, 2017: 306-310.
- [21] Verstrepen K, Goethals B. Unifying nearest neighbors collaborative filtering[C]//Proceedings of the 8th ACM Conference on Recommender Systems, 2014: 177-184.
- [22] Linden G, Smith B, York J. Amazon. com recommendations: item-to-item collaborative filtering[J]. IEEE Internet Computing, 2003,

- 7(1):76-80.
- [23] Davidson J, Liebal B, Liu J, et al. The YouTube video recommendation system [C]//Proceedings of the 4th ACM Conference on Recommender Systems, 2010:293-296.
- [24] Garg D, Gupta P, Malhotra P, et al. Sequence and time aware neighborhood for session-based recommendations; stan [C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019:1069-1072.
- [25] Hu H, He X, Gao J, et al. Modeling personalized item frequency information for next-basket recommendation [C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020:1071-1080.
- [26] Rac M, Kompan M, Bielikova M. Session-based k-NNs with semantic suggestions for next-item prediction [EB/OL]. <https://arxiv.org/abs/2012.08793>, 2020.
- [27] Tang J, Wang K. Personalized top-n sequential recommendation via convolutional sequence embedding [C]//Proceedings of the 11th ACM International Conference on Web Search and Data Mining, 2018:565-573.
- [28] Yuan F, Karatzoglou A, Arapakis I, et al. A simple convolutional generative network for next item recommendation [C]//Proceedings of the 12th ACM International Conference on Web Search and Data Mining, 2019:582-590.
- [29] Hsu K C, Chou S Y, Yang Y H, et al. Neural network based next-song recommendation [EB/OL]. <https://arxiv.org/abs/1606.07722>, 2016.
- [30] Tuan T X, Phuong T M. 3D convolutional networks for session-based recommendation with content features [C]//Proceedings of the 11th ACM Conference on Recommender Systems, 2017:138-146.
- [31] 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, 2019:2173-2176.
- [32] Yuan F, He X, Jiang H, et al. Future data helps training: Modeling future contexts for session-based recommendation [C]//Proceedings of the Web Conference, 2020:303-313.
- [33] Donkers T, Loepp B, Ziegler J. Sequential user-based recurrent neural network recommendations [C]//Proceedings of the 11th ACM Conference on Recommender Systems, 2017:152-160.
- [34] Li Z, Zhao H, Liu Q, et al. Learning from history and present: next-item recommendation via discriminatively exploiting user behaviors [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018:1734-1743.
- [35] Quadrana M, Karatzoglou A, Hidasi B, et al. Personalizing session-based recommendations with hierarchical recurrent neural networks [C]//Proceedings of the 11th ACM Conference on Recommender Systems, 2017:130-137.
- [36] Hidasi B, Karatzoglou A, Baltrunas L, et al. Session-based recommendations with recurrent neural networks [EB/OL]. <https://arxiv.org/abs/1511.06939>, 2016.
- [37] Bharadhwaj H, Joshi S. Explanations for temporal recommendations [J]. KI-Künstliche Intelligenz, 2018, 32(4):267-272.
- [38] Hidasi B, Quadrana M, Karatzoglou A, et al. Parallel recurrent neural network architectures for feature-rich session-based recommendations [C]//Proceedings of the 10th ACM Conference on Recommender Systems, 2016:241-248.
- [39] Tan Y K, Xu X, Liu Y. Improved recurrent neural networks for session-based recommendations [C]//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, 2016:17-22.
- [40] Hidasi B, Karatzoglou A. Recurrent neural networks with top-k gains for session-based recommendations [C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018:843-852.
- [41] Bogina V, Kuflik T. Incorporating dwell time in session-based recommendations with recurrent neural networks [C]//RecTemp@RecSys, 2017:57-59.
- [42] Jannach D, Ludewig M. When recurrent neural networks meet the neighborhood for session-based recommendation [C]//Proceedings of the 11th ACM Conference on Recommender Systems, 2017:306-310.
- [43] You J, Wang Y, Pal A, et al. Hierarchical temporal convolutional networks for dynamic recommender systems [C]//World Wide Web Conference, 2019:2236-2246.
- [44] Song Y, Lee J G. Augmenting recurrent neural networks with high-order user-contextual preference for session-based recommendation [EB/OL]. <https://arxiv.org/abs/1805.02983>, 2018.
- [45] Liu Q, Wu S, Wang D, et al. Context-aware sequential recommendation [C]//16th International Conference on Data Mining (ICDM), IEEE, 2016:1053-1058.
- [46] Twardowski B. Modelling contextual information in session-aware recommender systems with neural networks [C]//Proceedings of the 10th ACM Conference on Recommender Systems, 2016:273-276.
- [47] Wu C Y, Ahmed A, Beutel A, et al. Recurrent recommender networks [C]//Proceedings of the 10th ACM International Conference on Web Search and Data Mining, 2017:495-503.
- [48] Kumar V, Khattar D, Gupta S, et al. Deep neural architecture for news recommendation [C]//8th International Conference of the CLEF Initiative, Dublin, Ireland, CEUR Workshop Proceedings, 2017.
- [49] Gabriel De Souza P M, Jannach D, Da Cunha A M. Contextual hybrid session-based news recommendation with recurrent neural networks [J]. IEEE Access, 2019, 7:169185-169203, doi: 10.1109/ACCESS.2019.2954957.
- [50] Dai H, Wang Y, Trivedi R, et al. Deep coevolutionary network: embedding user and item features for recommendation [EB/OL]. <https://arxiv.org/abs/1609.03675>, 2016.
- [51] Vassøy B, Ruocco M, de Souza da Silva E, et al. Time is of the essence: a joint hierarchical rnn and point process model for time and item predictions [C]//Proceedings of the 12th ACM International Conference on Web Search and Data Mining, 2019:591-599.
- [52] Chen X, Xu H, Zhang Y, et al. Sequential recommendation with user memory networks [C]//Proceedings of the 11th ACM International Conference on Web Search and Data Mining, 2018:108-116.
- [53] Huang J, Zhao W X, Dou H, et al. Improving sequential recommendation with knowledge-enhanced memory networks [C]//41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018:505-514.

- [54] Ren K, Qin J, Fang Y, et al. Lifelong sequential modeling with personalized memorization for user response prediction [C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019:565-574.
- [55] Wang M, Ren P, Mei L, et al. A collaborative session-based recommendation approach with parallel memory modules [C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019:345-354.
- [56] Tan Q, Zhang J, Liu N, et al. Dynamic memory based attention network for sequential recommendation [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021:4384-4392.
- [57] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback [C]//Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, 2012:452-461.
- [58] 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, 2010:811-820.
- [59] Liang D, Altosaar J, Charlin L, et al. Factorization meets the item embedding: Regularizing matrix factorization with item co-occurrence [C]//Proceedings of the 10th ACM Conference on Recommender Systems, 2016:59-66.
- [60] Cheng C, Yang H, Lyu M R, et al. Where you like to go next: successive point-of-interest recommendation [C]//23rd International Joint Conference on Artificial Intelligence, 2013:2605-2611.
- [61] Lian D, Zheng V W, Xie X. Collaborative filtering meets next check-in location prediction [C]//Proceedings of the 22nd International Conference on World Wide Web, 2013:231-232.
- [62] Liu X, Liu Y, Aberer K, et al. Personalized point-of-interest recommendation by mining users' preference transition [C]//Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, 2013:733-738.
- [63] Kabbur S, Ning X, Karypis G. Fism: factored item similarity models for top-n recommender systems [C]//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013:659-667.
- [64] He R, McAuley J. Fusing similarity models with markov chains for sparse sequential recommendation [C]//IEEE 16th International Conference on Data Mining (ICDM), IEEE, 2016:191-200.
- [65] Guo L, Yin H, Wang Q, et al. Streaming session-based recommendation [C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019:1569-1577.
- [66] Zhou J, Cui G, Hu S, et al. Graph neural networks: a review of methods and applications [J]. AI Open, 2020, 1: 57-81, doi: 10.48550/arXiv.1812.08434.
- [67] Wu S, Tang Y, Zhu Y, et al. Session-based recommendation with graph neural networks [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1):346-353.
- [68] Wang Z, Wei W, Cong G, et al. Global context enhanced graph neural networks for session-based recommendation [C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020:169-178.
- [69] Qiu R, Li J, Huang Z, et al. Rethinking the item order in session-based recommendation with graph neural networks [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019:579-588.
- [70] Chen T, Wong R C W. Handling information loss of graph neural networks for session-based recommendation [C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020:1172-1180.
- [71] Song W, Xiao Z, Wang Y, et al. Session-based social recommendation via dynamic graph attention networks [C]//Proceedings of the 12th ACM International Conference on Web Search and Data Mining, 2019:555-563.
- [72] Zhang M, Wu S, Yu X, et al. Dynamic graph neural networks for sequential recommendation [EB/OL]. <https://arxiv.org/abs/2104.07368>, 2021.
- [73] Kumar S, Zhang X, Leskovec J. Predicting dynamic embedding trajectory in temporal interaction networks [C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019:1269-1278.
- [74] Li Y, Ding Y, Chen B, et al. Extracting attentive social temporal excitation for sequential recommendation [EB/OL]. <https://arxiv.org/abs/2109.13539>, 2021.
- [75] Chen Z, Zhang W, Yan J, et al. Learning dual dynamic representations on time-sliced user-item interaction graphs for sequential recommendation [C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management, 2021:231-240.
- [76] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [EB/OL]. <https://arxiv.org/abs/1409.0473>, 2015.
- [77] 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.
- [78] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Advances in Neural Information Processing Systems, 2017:5998-6008.
- [79] 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.
- [80] Santana M R O, Soares A. Hybrid model with time modeling for sequential recommender systems [EB/OL]. <https://arxiv.org/abs/2103.06138>, 2021.
- [81] Wang S, Hu L, Cao L, et al. Attention-based transactional context embedding for next-item recommendation [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1):2532-2539.
- [82] Ren P, Chen Z, Li J, et al. Repeatnet: a repeat aware neural recommendation machine for session-based recommendation [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1):4806-4813.
- [83] Yu L, Zhang C, Liang S, et al. Multi-order attentive ranking model for sequential recommendation [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1):5709-5716.
- [84] Ying H, Zhuang F, Zhang F, et al. Sequential recommender system based on hierarchical attention network [C]//International Joint



- Conference on Artificial Intelligence, 2018:3926-3932.
- [85] Wu J, Ou Z, Song M. Session-based recommendation with context-aware attention network[C]//Proceedings of the 2019 7th International Conference on Information Technology, IoT and Smart City, 2019:141-146.
- [86] Kang W C, McAuley J. Self-attentive sequential recommendation [C]//IEEE International Conference on Data Mining (ICDM), IEEE, 2018:197-206.
- [87] Zhang S, Tay Y, Yao L, et al. Next item recommendation with self-attentive metric learning[C]//33rd AAAI Conference on Artificial Intelligence, 2019, 9.
- [88] Li J, Wang Y, McAuley J. Time interval aware self-attention for sequential recommendation [C]//Proceedings of the 13th International Conference on Web Search and Data Mining, 2020:322-330.
- [89] 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, 2019:1441-1450.
- [90] Lv F, Jin T, Yu C, et al. SDM: sequential deep matching model for online large-scale recommender system [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019:2635-2643.
- [91] Wu L, Li S, Hsieh C J, et al. SSE-PT: sequential recommendation via personalized transformer[C]//14th ACM Conference on Recommender Systems, 2020:328-337.
- [92] Zhou C, Bai J, Song J, et al. Atrank: an attention-based user behavior modeling framework for recommendation [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1):4564-4571.
- [93] Zhang T, Zhao P, Liu Y, et al. Feature-level deeper self-attention network for sequential recommendation [C]//28th International Joint Conference on Artificial Intelligence, 2019:4320-4326.
- [94] Huang X, Qian S, Fang Q, et al. Csan: contextual self-attention network for user sequential recommendation [C]//Proceedings of the 26th ACM International Conference on Multimedia, 2018:447-455.
- [95] Smirnova E, Vasile F. Contextual sequence modeling for recommendation with recurrent neural networks[C]//Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems, 2017:2-9.
- [96] Du Z, Wang X, Yang H, et al. Sequential scenario-specific meta learner for online recommendation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019:2895-2904.
- [97] Sun Z, Guo Q, Yang J, et al. Research commentary on recommendations with side information: a survey and research directions[J]. Electronic Commerce Research and Applications, 2019, doi: 10.1016/j.elerap.2019.100879.
- [98] Meng W, Yang D, Xiao Y. Incorporating user micro-behaviors and item knowledge into multi-task learning for session-based recommendation[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020:1091-1100.
- [99] Peake G, Wang J. Explanation mining: post hoc interpretability of latent factor models for recommendation systems[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018:2060-2069.
- [100] Greenstein-Messica A, Rokach L, Friedman M. Session-based recommendations using item embedding [C]//Proceedings of the 22nd International Conference on Intelligent User Interfaces, 2017:629-633.
- [101] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[EB/OL]. <https://arxiv.org/abs/1301.3781>, 2013.
- [102] Pennington J, Socher R, Manning C D. Glove: global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014:1532-1543.
- [103] Hariri N, Mobasher B, Burke R. Context-aware music recommendation based on latent topic sequential patterns[C]//Proceedings of the 6th ACM Conference on Recommender Systems, 2012:131-138.
- [104] Natarajan N, Shin D, Dhillon I S. Which app will you use next? collaborative filtering with interactional context[C]//Proceedings of the 7th ACM Conference on Recommender Systems, 2013:201-208.
- [105] Yuan Fu-yong, Li Chen, Lei Yu, et al. POI dynamic recommendation algorithm based on fusion time series[J]. Journal of Chinese Computer Systems, 2020, 41(2):291-295.
- [106] Zhu F, Chen C, Wang Y, et al. Dtcdr: a framework for dual-target cross-domain recommendation [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019:1533-1542.
- [107] Zhuang F, Zhou Y, Zhang F, et al. Sequential transfer learning: cross-domain novelty seeking trait mining for recommendation [C]//Proceedings of the 26th International Conference on World Wide Web Companion, 2017:881-882.
- [108] Ma M, Ren P, Lin Y, et al.  $\pi$ -Net: a parallel information-sharing network for shared-account cross-domain sequential recommendations[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019:685-694.
- [109] Covington P, Adams J, Sargin E. Deep neural networks for youtube recommendations[C]//Proceedings of the 10th ACM Conference on Recommender Systems, 2016:191-198.
- [110] Pi Q, Bian W, Zhou G, et al. Practice on long sequential user behavior modeling for click-through rate prediction [C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019:2671-2679.

#### 附中文参考文献:

- [1] 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.
- [2] 任守刚, 张景旭, 顾兴健, 等. 时间序列特征提取方法研究综述[J]. 小型微型计算机系统, 2021, 42(2):271-278.
- [9] 袁涛, 牛树梓, 李会元. 一种基于 CW-RNN 的多时间尺度序列建模推荐算法[J]. 中文信息学报, 2020, 34(6):97-105.
- [105] 原福永, 李晨, 雷瑜, 等. 融合时间序列的 POI 动态推荐算法[J]. 小型微型计算机系统, 2020, 41(2):291-295.