

文章编号: 1003-0077(2023)03-0001-17

基于会话的推荐方法综述

陈晋鹏^{1,2}, 李海洋^{1,2}, 张帆^{1,2}, 李环³, 魏凯敏⁴

(1. 北京邮电大学 计算机学院(国家示范性软件学院), 北京 100876;

2. 北京邮电大学 可信分布式计算与服务教育部重点实验室, 北京 100876;

3. 浙江大学 计算机科学与技术学院, 浙江 杭州 310058; 4. 暨南大学 信息科学技术学院, 广东 广州 510632)

摘要: 近年来, 基于会话的推荐方法受到学术界的广泛关注。随着深度学习技术的不断发展, 不同的模型结构被应用于基于会话的推荐方法中, 如循环神经网络、注意力机制、图神经网络等。该文对这些基于会话的推荐模型进行了详细的分析、分类和对比, 阐明了这些方法各自解决的问题与存在的不足。具体而言, 该文首先通过调研, 将基于会话的推荐方法与传统推荐方法进行比较, 阐明基于会话的推荐方法的主要优缺点; 其次, 详细描述了现有的基于会话的推荐模型如何建模会话集中的复杂数据信息, 以及这些模型方法可解决的技术问题; 最后, 该文讨论并指出了在基于会话推荐的领域中存在的挑战和未来研究的方向。

关键词: 基于会话的推荐方法; 会话建模; 深度学习

中图分类号: TP391

文献标识码: A

Review on Session-based Recommendation Methods

CHEN Jinpeng^{1,2}, LI Haiyang^{1,2}, ZHANG Fan^{1,2}, LI Huan³, WEI Kaimin⁴

(1. School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. Key Laboratory of Trustworthy Distributed Computing and Service, Beijing University of Posts and Telecommunications, Beijing 100876, China;

3. College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang 310058, China;

4. College of Information Science and Technology, Jinan University, Guangzhou, Guangdong 510632, China)

Abstract: In recent years, session-based recommendation methods have attracted extensive attention from academics. With the continuous development of deep learning techniques, different model structures have been used in session-based recommendation methods, such as Recurrent Neural Networks, Attention Mechanism, and Graph Neural Networks. This paper conducts a detailed analysis, classification, and comparison over these models, and expounds on the target problems and shortcomings of these methods. In particular, this paper first compares the session-based recommendation methods with the traditional recommendation methods, and expounds the main advantages and disadvantages of the session-based recommendation methods through investigation. Subsequently, this paper details how complex data and information are modeled in session-based recommendation models, as well as the problems that these models can solve. Finally, this paper discusses and identifies the challenges and potential research directions in session-based recommendations.

Keywords: session-based recommendation method; session modeling; deep learning

收稿日期: 2022-05-22 定稿日期: 2022-10-04

基金项目: 国家自然科学基金(61702043, 61972178); 广东省自然科学基金(2019A1515011753, 2019B1515120010)

0 引言

随着大数据时代的到来,互联网的规模日益庞大,这为用户带来极其丰富且复杂的信息,使得用户能方便、快捷地获取信息。然而,面对海量的信息,用户往往不能迅速地获取自己需要的内容,导致对信息的使用效率有所下降,也就是所谓的信息超载(Information Overload)^[1]。解决信息超载问题,目前有两个常用的方法,一是搜索引擎,二是推荐系统。推荐系统根据用户的个人信息或历史行为来学习用户的兴趣偏好,并为用户生成其感兴趣的推荐序列。目前,推荐系统应用于许多领域^[2],例如,电子商务^[3]、新闻推荐^[4-5]、音乐推荐^[6]以及位置服务^[7]等。传统推荐方法主要包括两种:一是基于内容的推荐方法,二是基于协同过滤的推荐方法。

基于内容的推荐方法^[8]是根据用户(User)的兴趣简介与项目(Item)的特征描述实现推荐。其中,用户的兴趣简介也就是用户的兴趣偏好,它可以来源于用户主动输入,但往往是从用户与过去项目交互中获得的;项目的特征描述可以是项目的描述或用户对该项目的评论。基于内容的推荐方法在一定程度上可以解决冷启动(Cold-Start)问题^[9]。同时,基于内容的推荐方法具有很好的可解释性,但基于内容的推荐方法在项目的特征提取上存在一定难度,且难以挖掘出用户的潜在兴趣偏好。

基于协同过滤的推荐方法是基于用户或项目间的相关性进行推荐的方法,主要可以分为两类,即基于用户的协同过滤以及基于项目的协同过滤。基于用户的协同过滤,也可称为K-NN(K-Nearest Neighbor)协同过滤,其核心思想是根据用户的历史行为找到与当前用户行为偏好相似的用户,然后为该用户推荐这些相似用户感兴趣的项目。基于项目的协同过滤的方法关注的不是用户之间的相似性,而是项目之间的关联性,也就是说,若两个项目被相同的用户喜欢或不喜欢,那么这两个项目是相关的,系统为用户推荐相关的项目^[10]。基于协同过滤的推荐方法能够挖掘出用户的潜在兴趣,帮助用户发现新的兴趣,提高推荐的质量,但是其难以解决冷启动问题;同时,基于协同过滤的推荐方法受到数据稀疏性的影响,随着数据量的增大,数据的稀疏性也随之增大(如一个用户只与所有项目中的几个有交互),这就使得计算最邻近(最相似)用户或项目的准确率降低,影响最终推荐结果。

以上传统的推荐方法各有优缺点,随着互联网规模的增大,数据量也越来越大,传统推荐方法的不足也越来越明显。除了它们各自的缺点外,传统推荐方法存在的共同点是:只能够考虑用户长期静态偏好,而忽略了用户兴趣随时间的变化。因而出现了序列推荐(Sequential Recommendation)^[11]。这一类专注于行为序列性的推荐模型,通过对用户历史行为的序列化建模,学习用户的兴趣,进而对用户进行相关推荐。

随着序列推荐的深入研究,其弊端也开始出现,即:偏向于长期兴趣的学习,而忽略了用户短期内偏好的转移,使得用户在某一段时间内的兴趣变化被其历史交互行为所掩盖,从而生成不可靠的推荐结果。产生这个问题的原因是忽略了用户行为的事务性结构,例如用户之前长期购买牛仔裤,某一时段突然喜欢运动裤,但由于用户购买的牛仔裤历史记录远远多于运动裤,使推荐系统认为用户的偏好仍为牛仔裤。在此背景下,将用户行为分解成更小的粒度,考虑用户行为的事务性结构,捕捉用户短期偏好的转移是十分有必要的,因此基于会话的推荐系统(Session-Based Recommendation System, SBRS)得到了快速的研究和发展。

基于会话的推荐方法是将用户的全局交互行为分割成一个个更小粒度的事务单元,每个事务单元是由用户的部分交互行为组成的,这些事务单元被称为会话^[12-13]。会话可以在不同的场景中表现出不同的含义,例如,在电子商务领域,会话可以是用户一次购买的物品,或一小时内添加到购物车的商品;在旅游场景,会话可以是用户一年内游玩的景点;另外,会话也可以是用户一小时内浏览的网页、一天内看的电影等。基于会话的推荐算法通过学习这些会话内以及会话间项目的依赖关系,挖掘出用户的兴趣偏好,并为用户生成其感兴趣的推荐列表。如图1所示,用户的购物行为构成用户与项目的交互序列,其中, U 表示用户、 V 表示项目、 S 表示会话。在传统的推荐方法中往往通过全局的项目序列来挖掘用户兴趣,而在基于会话的推荐系统中,用户的购物行为会根据购物的次序将用户行为分割成粒度更小的会话,以这种方法保留用户行为的事务结构,获取用户行为中更多的转换信息和依赖关系;基于会话的推荐方法不仅仅关注用户最新的交互行为^①,同时包括当前会话中其他项目^②以及历史会话

① 图1中会话 S_3 中的项目 V_5

② 会话 S_3 中的项目 V_1, V_3

对当前推荐结果的影响,这种只需要根据会话中物品依赖关系进行推荐的特性,使其能够为匿名用户进行推荐。总的来说,基于会话的推荐方法以会话

作为基本单元,不仅能够捕捉更多项目间的转换信息和依赖关系,还能聚焦于当前会话,适时捕捉到用户兴趣的变化,从而改善推荐的效果。

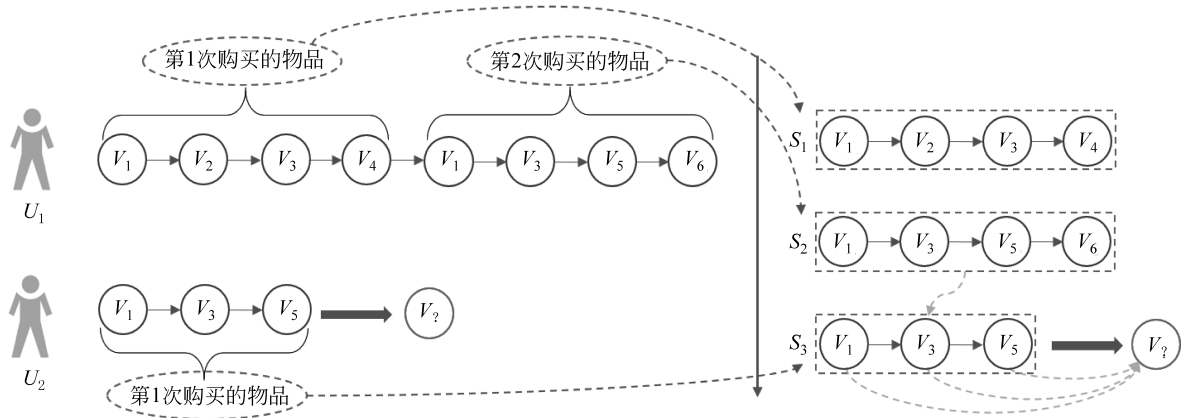


图 1 基于会话推荐方法的数据结构

基于会话推荐方法解决了传统推荐系统只关注用户长期静态偏好、无法及时发现用户兴趣变化的问题。但是,基于会话的推荐方法仍存在一些问題,例如,基于会话的推荐方法难以捕捉用户

长期静态偏好。表 1 在输入、核心思想、优点与缺点等方面,将传统的推荐方法与基于会话的推荐方法进行了对比,能直观地表现出各类推荐方法的优势与不足。

表 1 不同类型推荐方法的比较

推荐方法类型		输入	核心思想	优点	缺点
传统推荐方法	基于内容的推荐方法	用户简介、项目内容	计算用户兴趣嵌入与项目特征嵌入的相似性	可解决冷启动问题,可解释性强	特征提取较难,无法挖掘出用户的潜在兴趣偏好
	基于协同过滤的推荐方法	用户与项目的交互数据	建模用户与项目间的交互	广泛的适用性,能挖掘用户的潜在兴趣	无法解决冷启动问题、数据稀疏性问题
序列推荐方法		长序列数据	学习历史数据的时序关系	能够捕捉用户的兴趣偏好	难以发现用户短期的兴趣偏移
基于会话的推荐方法		会话集	建模会话及会话内项目间的复杂依赖关系	能够捕捉用户的短期兴趣偏好	挖掘用户的一般兴趣偏好较困难

针对基于会话推荐方法存在的问题,许多学者开展了深入的研究。近年来,深度学习的发展为基于会话的推荐方法带来了机遇。随着深度学习在自然语言处理、语音识别、图像识别等多个领域的突破性进展,循环神经网络(Recurrent Neural Network, RNN)、图神经网络(Graph Neural Networks, GNN)等深度学习模型被应用于基于会话的推荐算法中。目前,随着新技术的不断发展,国内外的基于会话的推荐领域的综述性文章^[12]较少,且这些文章并未对目前已提出的方法和已解决的问题进行详细的对比、分析和总结。因此,一篇详细归纳和总结在基于会话的推荐领域已取得的进展的综述文献十分必要。本文对基于会话的推荐方法进行了详细、全面的调研,对已有的研究成果进行梳理,以期对从事

推荐算法和系统的理论研究及实践开发人员提供参考和启发。

1 基础知识和相关定义

本节首先对下文中常用的符号进行说明,进而对与基于会话推荐方法相关的基础知识进行归纳和解释,并对基于会话的推荐方法进行数学形式的定义。

1.1 基础知识

基于会话的推荐是推荐系统领域的一个子任务,其本质是通过用户过去的行为记录来挖掘用户的兴趣偏好,并根据用户的兴趣偏好生成其感兴趣

的推荐列表。与传统的推荐方法关注用户长期偏好不同,基于会话的推荐方法聚焦于用户短期偏好的转移,这就意味着基于会话的推荐方法能够及时感知用户兴趣变化,且用户当前的意图不会被历史行为所淹没,从而使推荐的结果更加可靠。上述优点得益于会话这种基本单元的设置,它是通过将用户的全局行为分割成更小的粒度得到的。会话是一组项目的集合,如用户一次购买的项目、一小时内听的歌曲都可以构成一个会话。虽然会话也是由与用户交互的项目组成的,但其具有事务性。换句话说,会话能有效地保持数据的自然特性(更符合用户实际的行为习惯),避免了会话数据动态和局部信息的丢失。

基于会话的推荐方法不仅能充分分析会话内的项目分布,也能捕捉不同会话内项目的依赖关系以及不同会话间的转换信息。这些是传统推荐方法无法考虑的信息,这些信息可分别被称为会话内上下文以及会话间上下文。基于会话的推荐方法通过分析会话内上下文以及会话间上下文挖掘出用户行为模式,生成与用户偏好相匹配的个性化推荐结果。按推荐结果进行分类,基于会话的推荐方法可分为下一项推荐和下一会话推荐^[13]。下一项推荐是根据当前会话和历史会话对当前会话的下一个可能项目进行推荐;下一会话推荐是指根据邻近会话与历史会话推荐出下一个会话中一个或多个可能项目。目前大多数研究的是下一项推荐方法,而对于下一会话推荐方法的研究相对较少。

1.2 相关定义

基于会话的推荐方法的目标是通过分析用户行为,为用户推荐其感兴趣的项目。其中,系统中所有用户的集合可表示为 $U = \{u_1, u_2, u_3, \dots, u_m\}$,所有项目的集合可表示为 $V = \{v_1, v_2, v_3, \dots, v_n\}$ 。那么,用户 u 的用户行为 $V_u = \{v_{u,1}, v_{u,2}, v_{u,3}, \dots, v_{u,l}\}$,其中 l 是用户 u 的行为的基数。在实际数据中,用户行为是具有时间顺序的传统的推荐方法,根据全局用户行为分析用户偏好(通常是长期偏好)。基于会话的推荐方法为了考虑更加丰富的数据信息,捕捉用户的短期偏好,赋予用户行为事务特性,即将用户行为分解成粒度更小的会话,得到的会话集合可表示为 $S = \{s_1, s_2, s_3, \dots, s_k\}$ 。每一个会话 $s_i \in S$ 由同一用户 u 在一段时间内交互的多个项目组成,可表示为 $s_i = \{v_{i,1}, v_{i,2}, v_{i,3}, \dots, v_{i,j}\}$,其中 $v_{i,j} \in V_u$ 。在实际划分会话集时,往往会按照一定

的规则将用户行为进行分割,如间隔一小时、用户一次的购物行为。不同的规则会导致会话的平均长度不同(会话长度是指会话中项目的数量),同时,会话集的长度不同也可能影响模型学习用户偏好的准确性。后文将用到的符号如表2所示。

表2 符号表

符号	描述
$V = \{v_1, v_2, \dots, v_n\}$	项目集
$S = \{s_1, s_2, \dots, s_k\}$	会话集
$U = \{u_1, u_2, \dots, u_m\}$	用户集
v_i	项目集的某一项目
$s_i = \{v_{i,1}, v_{i,2}, \dots, v_{i,j}\}$	会话集的某一会话
u_i	用户集的某一用户
v_i	项目嵌入表示
s_i	会话嵌入表示
u_i	用户嵌入表示

已知用户的历史会话记录 $\{s_1, s_2, s_3, \dots, s_{k-1}\}$,其中 $s_i = \{v_{i,1}, v_{i,2}, v_{i,3}, \dots, v_{i,j}\}$ 。基于会话的推荐方法通过分析会话内项目(同一会话中的项目,如 $v_{i,1}, v_{i,2}, v_{i,3}$)、会话间项目(不同会话中的项目,如 $v_{i,1}, v_{i+1,1}$)的依赖关系,学习出用户的兴趣偏好。以下对下一项推荐和下一会话推荐进行正式定义:

- 下一项推荐指对于当前会话 $s_k = \{v_{k,1}, v_{k,2}, v_{k,3}, \dots, v_{k,j-1}\}$,用户在当前会话进行了 $j-1$ 次项目交互,通过分析当前会话 s_k 以及历史会话 $\{s_1, s_2, s_3, \dots, s_{k-1}\}$,下一项推荐的目的是预测用户交互的下一项 $v_{k,j}$;
- 下一会话推荐,即推荐系统通过分析用户的历史会话 $\{s_1, s_2, s_3, \dots, s_{k-1}\}$,学习用户的兴趣偏好,其推荐的目标是给出下一会话 s_k 的全部或部分项目。

2 基于会话推荐方法的分类

近年来,基于会话的推荐方法发展迅速,众多学者提出了大量的模型和方法,本节从技术角度对基于会话的推荐方法进行分类。首先介绍如协同过滤、矩阵分解、马尔科夫链等的传统会话推荐方法;进而介绍基于深度学习的会话推荐方法,包括循环神经网络、注意力机制、图神经网络等,最后介绍基

于强化学习的会话推荐方法。

2.1 传统的会话推荐方法

2.1.1 基于协同过滤的会话推荐方法

协同过滤^[14]是传统的推荐系统中常用的方法,其主要包括两种算法:基于用户的最邻近算法^[15]和基于项目的最邻近算法^[16]。

以基于用户的最邻近算法为例,其核心思想是根据用户的历史行为找到与当前用户相似的用户,然后为该用户推荐这些相似用户感兴趣的项目。该方法存在两个核心问题:一是如何寻找用户 u 的相似用户,二是如何计算 N 个相似用户对项目 v 的评分。

针对第一个问题,可以采用计算两个用户相似度的方法解决。计算相似度的方法有多种^[17],包括常见的余弦相似度,但余弦相似度适用于计算两个向量之间的相似关系,不太适合于此场景。在基于协同过滤的方法中,往往会考虑用户对项目的评分,因此可采用皮尔逊相关系数(Pearson Correlation Coefficient)来计算两用户之间的相似程度^[18],如式(1)所示。

$$\text{sim}(u_1, u_2) = \frac{\sum_{v \in V} (r_{u_1, v} - \bar{r}_{u_1})(r_{u_2, v} - \bar{r}_{u_2})}{\sqrt{\sum_{v \in V} (r_{u_1, v} - \bar{r}_{u_1})^2} \sqrt{\sum_{v \in V} (r_{u_2, v} - \bar{r}_{u_2})^2}} \quad (1)$$

其中, u_1, u_2 为两个用户, V 为用户 u_1, u_2 共同评分的项目集, $r_{u, v}$ 为用户 u 对项目 v 的评分, \bar{r}_u 是用户 u 的平均评分。皮尔逊相关系数的范围为 $(-1, +1)$, 值越接近 $+1$ 则认为两用户越相似(正相关), 反之越接近 -1 则认为两用户的兴趣相反(负相关), 值接近于 0 表明两用户之间几乎不存在相关性。

针对第二个问题,计算 N 个相似用户对项目 v 的评分, Resnick 等人^[18]提出可采用式(2)进行计算。

$$\text{red}(u, v) = \bar{r}_u + \frac{\sum_{u' \in N} \text{sim}(u, u') \times (r_{u', v} - \bar{r}_{u'})}{\sum_{u' \in N} |\text{sim}(u, u')|} \quad (2)$$

其中 N 为用户 u 的相似用户集合。由式(2)可知,该预测结果是从用户 u 的平均分出发,然后再根据每一个相似邻居对项目 v 的评分结果,对预测评分进行加权计算。

将协同过滤引入基于会话的推荐方法,其核心

思想不变。对于当前会话 s , 基于协同过滤的会话推荐算法首先利用式(1)的 $\text{sim}(s_i, s)$ 以得到 K 个相似会话, 构成相似会话集 N , 其中 s_i 属于历史会话集 S 。其次, 根据当前会话 s 计算相似会话 N 中每个可推荐项目 v 的相关性评分^[19], 计算方法如式(3)所示。

$$\text{score}(v, s) = \sum_{n \in N} \text{sim}(s, n) \times \mathbf{1}_n(v) \quad (3)$$

其中, 若 $v \in n$ 则 $\mathbf{1}_n(v) = 1$, 否则 $\mathbf{1}_n(v) = 0$, v 表示可推荐项目。通过计算每个可推荐项目与当前会话相关性的评分, 按照从高到低进行排序生成推荐列表。

上述基于协同过滤的会话推荐方法, 通过计算当前会话与每个项目之间的相关系数评分, 找出与当前会话相关的项目, 其本质上是通过相似会话寻找相关项目。当然, 也可以将基于用户的协同过滤方法引入到基于会话的推荐方法中, 即首先根据用户的历史行为(会话数据)计算相似用户; 然后, 在相似用户生成的会话中根据用户当前会话选择相关性较高的会话, 生成候选会话列表; 最后, 计算候选会话中项目与当前会话的相关性评分, 生成最终的推荐列表。

2.1.2 基于矩阵分解的会话推荐方法

推荐方法往往是通过分析用户的历史交互数据来挖掘用户偏好, 通常这些数据包含两种类型: 用户和项目。这些数据可以构成一个矩阵, 一维代表用户, 另一维代表用户感兴趣的项目。但由于大多数用户只与少量项目交互, 因此该矩阵是比较稀疏的, 单纯依赖这种稀疏矩阵去产生推荐的正确率是十分低下的。矩阵分解通过补充隐式反馈信息构造一个相对稠密的矩阵, 然后将用户和项目映射到同一潜在向量空间中, 使得每一个用户 u 都对应一个潜在的向量 q_u , 每一个项目 v 都对应一个潜在的向量 p_v , 此时, 用户 u 对项目 v 的评分 $r_{u, v}$ 可表示为式(4), 即用户的潜在向量 q_u 表示与项目的潜在向量表示 p_v 的点乘。

$$r_{u, v} = q_u^\top p_v \quad (4)$$

采用上述方法计算出用户与所有项目的评分, 即可实现推荐。因此, 矩阵分解的核心是如何获取用户与项目的潜在向量表示。Koren 等人^[20]指出在学习潜在向量表示 q_u, p_v 时, 可采用最小化已知评分集合上的正则化平方误差的方法, 其计算如式(5)所示。

$$\min_{q^*, p^*} \sum_{(u, v) \in K} (r_{u, v} - q_u^\top p_v)^2 + \lambda (\|q_u\|^2 + \|p_v\|^2) \quad (5)$$

其中, K 是训练集中所有用户和项目组成的二项对 (u, v) 集合; $r_{u,v}$ 是训练集中已知评分; λ 是一个常数, 用于防止过拟合现象的发生。

2.1.3 基于马尔科夫链的会话推荐方法

协同过滤为了寻找用户或项目的邻居, 往往过于依赖用户与系统的历史交互数据, 尤其是用户对项目的评级(评分)信息, 但是忽略了用户与项目交互的顺序性。例如, 用户在浏览商品时, 用户的点击浏览操作是有次序的(一个商品到另一个商品), 协同过滤无法对这种顺序行为进行建模, 从而无法挖掘出用户行为的序列模式。

用一阶马尔科夫链^[21]对会话内项目的转移进行建模, 其核心思想是计算一系列项目的转移概率, 然后, 将用户的购物序列与计算出的转移概率的序列进行匹配, 将概率较高的项目添加至候选列表。首先, 将会话集 S 构建成一个有向图, 图中的节点表示会话集中的项目, 其权重表示为该项目出现的频率; 边表示一个项目向另一个项目的转移, 其权重为两个项目共现的次数; 每一个会话 s 可以由图中的一条路径表示出来。那么, 项目 v_i 到项目 v_k 的转移概率可采用式(6)计算。

$$P(v_i \rightarrow v_k) = \frac{\text{freq}(v_i \rightarrow v_k)}{\sum_{v_i \in V} \text{freq}(v_i \rightarrow v_i)} \quad (6)$$

转移路径 $\{v_1 \rightarrow v_2 \rightarrow v_3\}$ 的预估概率可表示如式(7)所示。

$$P(v_1 \rightarrow v_2 \rightarrow v_3) = P(v_1) \times P(v_2 | v_1) \times P(v_3 | v_2) \quad (7)$$

通过以上方法, 给定一个项目序列作为马尔科夫链的输入, 系统将计算出以该项目序列为前缀的路径的概率, 选择概率高的路径, 将出现在这些路径中的项目添加至推荐列表, 完成推荐任务。

基于一阶马尔科夫链的会话推荐方法可以挖掘会话数据中项目的序列模式, 给定当前会话一个或多个项目, 该方法能够进行下一项或连续多项的可能推荐。但是, 针对推荐系统的数据往往具有稀疏性的特性, 采用一阶马尔科夫链生成的推荐序列的准确性会受到影响。

2.1.4 传统会话推荐方法的比较

本文介绍了三种常见的传统的会话推荐方法: 基于协同过滤的、矩阵分解的和马尔科夫链的会话推荐方法。这三种方法提出的时间较早, 应用也较为广泛, 都有其各自的优点, 但也存在着不足。三种推荐方法的对比如表3所示。

表3 传统推荐方法的比较

推荐方法	优点	缺点
协同过滤	适用广泛, 无需可推荐项目的知识	数据稀疏性限制协同过滤的性能, 无法解决冷启动问题
矩阵分解	能够添加隐式信息弥补显示反馈信息的不足	无法解决冷启动问题
一阶马尔科夫链	挖掘用户行为的序列模式	只能利用用户行为的显式信息

2.2 基于深度学习的会话推荐方法

随着算力的不断提升和深度学习的不断发展, 越来越多的科研人员将深度学习应用到基于会话的推荐方法中, 并取得了一系列的成果。基于深度学习的会话推荐方法不仅能够分析数据间的显式联系, 还能够学习到用户、项目等数据的隐式特征, 生成表达能力强大的嵌入表示, 进行更可靠的推荐。

不同的深度学习模型被应用于基于会话的推荐系统中来解决不同的问题。例如, 会话集常按照时间戳或用户实际的交互顺序来进行分割, 会话集中的会话往往具有序列关系, 同时, 会话中的项目也具有顺序关系, 因此对序列关系敏感的循环神经网络被应用于基于会话的推荐方法中。传统的推荐方法往往只能考虑到用户的长期偏好, 而基于会话的推荐方法将用户行为分割成更小粒度的会话集, 通常只能聚焦于用户的短期偏好。为了使基于会话的推荐方法既能考虑用户的长期偏好也能够关注用户的短期偏好, 注意力机制被应用其中; 图神经网络能够建模会话集中复杂的项目转移关系, 从而学习出表达能力极强的项目嵌入表示, 在基于会话的推荐方法中广泛应用。后文将根据模型类别对已有的推荐方法进行分类总结, 并对各类方法进行详细对比。

2.2.1 基于循环神经网络的会话推荐方法

传统的推荐方法在分析用户交互数据时, 每一个项目都是相互独立的, 无法捕获项目间的连续信息。然而, 用户的下一个交互项目, 往往会受到历史交互记录的影响(尤其是上一个交互项目), 因此关注项目间的序列信息是十分必要的。

RNN 最早被应用于自然语言处理领域^[22-24], 并取得了显著的成绩。RNN 的特性是对序列化数据敏感, 而在基于会话的推荐系统中, 每个会话是由一组连续的项目组成的, 可被看做是一个序列。因此, Hidasi 等人^[25]将 RNN 应用于基于会话的推荐

方法,提出了 GRU4Rec 模型,这是循环神经网络首次被应用于基于会话的推荐方法。在这篇文章中,作者采用门控循环单元(Gated Recurrent Unit, GRU)^[26]作为核心结构,GRU 的计算如式(8)所示。

$$\begin{aligned} r_t &= \sigma(W_r x_t + U_r h_{t-1}) \\ z_t &= \sigma(W_z x_t + U_z h_{t-1}) \\ \hat{h}_t &= \tanh(Wx_t + U(r_t \odot h_{t-1})) \\ h_t &= (1 - z_t)h_{t-1} + z_t \hat{h}_t \end{aligned} \quad (8)$$

GRU 是 RNN 的一个变体,比 RNN 更加复杂。但是,GRU 能够解决 RNN 梯度消失的问题,因此,基于循环神经网络的会话推荐方法常以 GRU 建模用户行为的序列模式,其模型结构如图 2 所示。首先,模型将会话中的项目进行独热编码,然后输入到嵌入层,得到每一个项目的向量表示;之后,将连续的项目表示输入到 GRU 层;最终,经过前馈层后输出,输出的结果是所有可能的下一个项目的得分。

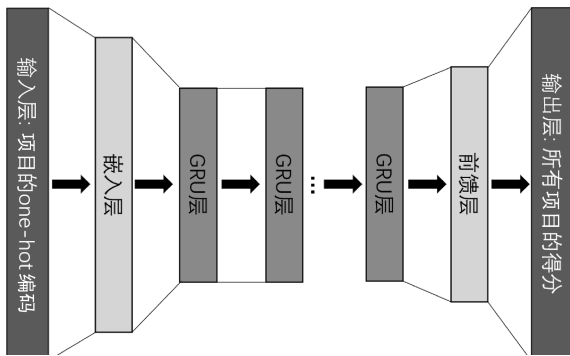


图 2 基于循环神经网络的会话推荐方法的总体架构

在此之前,RNN 常用于自然语言处理领域,将 RNN 应用到推荐领域中还存在一些问题。例如,如果项目数量过多,模型的维度就会增大,这就造成计算量大大增加。为减少计算量,GRU4Rec 采用 Mini-Batch 的处理方式,将部分会话第一个项目作为一个小批量,第二个小批量由会话的第二个项目构成,以此类推,若某一会话结束则可由其他可用会话继续替代,如图 3 所示。实验表明,这个思路虽然简单但效果明显,为后续的研究指引了方向。

GRU4Rec 在模型本身上并没有较大的创新,其意义在于首次将 RNN 引入推荐方法,并针对现实中存在的问题对模型进行调整,为后续深度学习在基于会话推荐方法的研究奠定了基础。

受到上述文章的启发,众多研究人员将循环神经网络应用于会话推荐中,并针对现有的问题提出改进策略。本文列举了部分典型的基于 RNN 的会

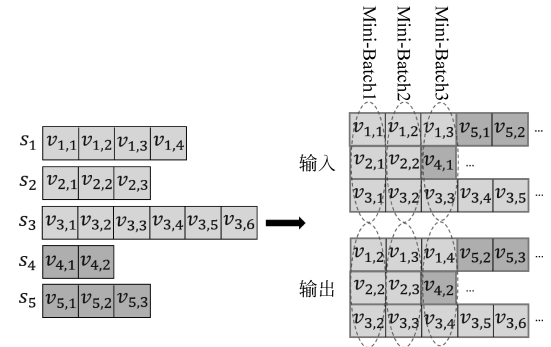


图 3 会话并行 Mini-Batch 创建

话推荐模型,并在数据特征、目标问题等方面进行了对比,如表 4 所示。Hidasi 等人^[27]针对 GRU4Rec 中仅对项目标识进行建模,无法充分利用项目丰富特征(图片、文字等)的不足,提出了一种面向丰富特征的基于循环神经网络的会话推荐模型——并行 RNN 模型(Parallel RNN, P-RNN)。P-RNN 由多个 RNN 构成,每个 RNN 用于学习项目的某一类型特征。最后将所有的特征表示融合得到项目的最终表示。P-RNN 采用多特征并行建模的方式,能够学习出表达能力更加强大的项目表示。

Bogina 等人^[28]提出将停留时间(用户与项目的交互时间间隔)应用于基于 RNN 的会话推荐方法中,并提出了 Dewell time based RNN,其将用户与项目的交互时间作为提取用户兴趣偏好的关键特征,该思想适用于当前流行的短视频推荐领域。针对同样的问题,Sun 等人^[29]提出了一种新的模型结构,基于带时间注意因子的循环神经网络的会话推荐模型——TA4Rec。TA4Rec 根据用户对项目的停留时间计算关注因子,并将其添加到 GRU 网络中,从而提高模型性能。

Ruocco 等人^[30]提出 II-RNN(Inter-Intra RNN)模型,将单个会话内的建模与历史会话间的建模相结合,以解决每个会话开始时,因无项目序列而无法获取用户兴趣偏好的问题。II-RNN 由两个 RNN 构成:会话内 RNN(Intra-session RNN)、会话间 RNN(Inter-session RNN),II-RNN 通过构建会话间的序列关系以及会话内项目序列关系,能够在一定程度上学习出用户的长期偏好和短期偏好,从而生成更加准确的推荐。

Quadrana 等人^[31]针对已有模型仅考虑项目标识及特性,忽略了用户标识的问题,提出了一个分层循环神经网络的个性化会话推荐方法 HRNN(Hierarchical Recurrent Neural Networks),在 GRU4Rec

表4 基于循环神经网络的会话推荐方法的比较

模型	数据特征	目标问题	工作机制	优点
GRU4Rec ^[25]	项目标识	关注项目间的序列信息	采用 GRU 建模会话中项目的序列	首次将 RNN 引入基于会话的推荐算法中,能够挖掘用户交互项目的顺序信息
P-RNN ^[27]	项目标识及特征	充分利用丰富的项目特征建模	采用并行 RNN 模型,分别建模项目不同特征	建模项目的丰富特征,生成表达能力强的项目表示
Dwell time based RNN ^[28]	项目标识,用户对项目的停留时间	考虑停留时间特征:用户对某一项目停留时间越长对其越感兴趣	对于长停留时间的项目,构建 Mini-Batch 时连续使用多次	学习到的用户兴趣偏好更准确
TA4Rec ^[29]	项目标识,用户对项目的停留时间		计算时间注意力因子,并将其添加到 GRU4Rec 模型	
II-RNN ^[30]	项目标识	解决每个会话开始时,因无项目序列而无法获取用户兴趣偏好的问题	构建会话内 RNN 以及会话间 RNN	构建会话间的序列关系以及会话内项目序列关系
HRNN ^[31]	项目标识、用户标识	上述模型仅考虑项目标识及特征,忽略了用户标识及特征	分层循环神经网络:在 GRU4Rec 的基础上,增加用户级 GRU	动态模拟用户兴趣变化,挖掘用户意图
AGRU ^[32]	项目标识、用户标识		将全局的用户嵌入和会话中的项目序列结合,同时输入到添加注意力机制 GRU 模型	能够同时结合用户、项目、上下文信息
ARNN ^[33]	项目标识、用户特征		采用基于乘积的神经网络(PNN)捕捉用户上下文与项目间的高阶交互来模拟用户的上下文偏好	能够增强大部分现有的基于 RNN 的会话推荐模型

的基础上,增加了用户级 GRU。用户级 GRU 能够跟踪用户在会话间的变化,即对用户与会话的交互进行建模,从而动态模拟用户兴趣变化,挖掘用户意图。Phuong 等人^[32]提出从用户与项目的交互数据中学习出用户和项目的嵌入表示,并将全局的用户嵌入和会话中的项目序列结合,通过 Attentive GRU(AGRU)型结构生成下一项推荐。HRNN 是根据用户与会话交互的隐式信息挖掘用户意图,并没有对用户的显式信息(如年龄、性别、位置等)进行建模。因此,Song 等人^[33]利用高阶用户上下文偏好扩充基于 RNN 的会话推荐方法,提出了增强 RNN(Augmented RNN,ARNN),ARNN 的核心结构是基于乘积的神经网络^[34](Product-Based Neural Network,PNN)能够捕捉用户上下文与项目间的高阶交互来模拟用户的上下文偏好,能够对已有的 RNN 模型效果进行增强。

循环神经网络是第一个被应用于基于会话推荐领域中的深度学习框架,主要原因在于基于会话的推荐系统的数据结构和数据特征符合循环神经网络的输入和特性:

(1) 基于会话的推荐方法本质上也是序列推荐^[11]的一种,区别在于会话是在一定的时间规则下对序列的一种分割。因此,在基于会话的推荐系统

中,数据的序列模式是比较明显的,RNN 具有记忆性、参数共享等特性,对于序列的非线性特征的学习具有一定的优势。

(2) 理论上讲 RNN 可以对无限长的序列进行建模,但是越长的序列越容易导致在训练过程中产生梯度爆炸的问题,但是在会话数据集中,组成会话的物品序列往往是较短的,RNN 通过对会话中项目序列的建模,挖掘项目间的顺序关系。

2.2.2 基于注意力机制的会话推荐方法

注意力机制最早被应用在 NLP 领域中的翻译模型^[35],其能够有效地抽取序列信息中的有用信息,忽略无效信息,随着对注意力机制的深入研究,其被应用于基于会话的推荐领域^[36-37]。注意力机制能够强调用户当前的意图,这意味着推荐方法能够忽略用户的一些错误交互行为(误点击),这是传统的基于 RNN 的会话推荐方法所欠缺的能力。

Li 等人^[38]首先对这个问题进行了研究,并提出了基于注意力机制的会话推荐方法——(Neural Attentive Recommendation Machine, NARM)。NARM 本质上是一个改进的编码-解码器(Encoder-Decoder)结构,在此之前,Encoder-Decoder 结构常用于自然语言处理领域^[39]。NARM 模型结构主要分为三个部分:编码器、会话特征生成器和解码器。

其中编码器用于分析用户行为,挖掘出用户当前的兴趣偏好和主要目的;会话特征生成器能够构建当前的会话表示;解码器的主要功能是根据生成的会话特征表示来生成推荐项目列表。NARM 的总体框架如图 4 所示。

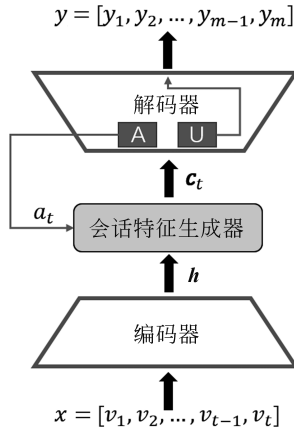


图 4 NARM 的总体框架结构

由图 4 可以看出,用户的点击序列 $x = [v_1, v_2, \dots, v_{t-1}, v_t]$ 输入到 NARM 的编码器中,得到一组高维的隐藏向量表示 $h = [h_1, h_2, \dots, h_{t-1}, h_t]$,并同当前时刻 t 的注意力信号 a_t 同时输入到会话特征生成器,以生成会话向量表示 c_t ,最终经过解码器生成推荐列表。 a_t 的目的是在时刻 t 强调或忽略 h 某些特征,是随时间动态变化的。

为了考虑用户行为的顺序关系,强调用户当前的意图,NARM 的编码器分为两部分:全局编码器和本地编码器。全局编码器用于模拟用户的顺序行为。全局编码器采用 GRU 作为循环单元,学习用户的顺序表示行为,GRU 的计算见式(8),并使用最终隐藏状态 h_t 作为用户序列行为特征的表示,如式(9)所示。

$$c_t^g = h_t \quad (9)$$

但是,全局编码器存在缺点,它不能准确地捕捉到用户当前会话的意图,因此 NARM 通过设计一个本地编码器来捕捉当前的主要目的。本地编码器的模型结构与全局编码器模型结构类似,其基本组件仍为 GRU,但是为了捕捉用户当前的主要意图,采用了注意力机制。注意力机制的作用是动态选择并线性组合输入序列的不同部分,即计算出当前输入序列的重要组成部分,计算方法如式(10)所示。

$$c_t^l = \sum_{j=1}^t a_{tj} h_j \quad (10)$$

其中, a_{tj} 为注意力因子,其目的是对输入序列

加权,以实现强调或忽略序列中的某一部分。注意力因子的计算方式如式(11)所示。

$$a_{tj} = q(h_t, h_j) \\ q(h_t, h_j) = v^T \sigma(A_1 h_t + A_2 h_j) \quad (11)$$

其中, A_1, A_2 的作用是将 h_t, h_j 转换到同一向量空间中, σ 为 sigmoid 函数。本地编码器利用注意力机制,捕捉用户当前明确的意图,这是以往基于 RNN 会话推荐方法(全局编码器)无法实现的。

NARM 的编码器部分由全局编码器和本地编码器一同组成,这意味着 NARM 在关注用户顺序行为(长期偏好)的同时,也能捕捉用户当前主要目的(短期偏好)。将用户的顺序行为和当前意图建模为统一的向量表示 c_t ,用于解码器的输入, c_t 的计算方式如式(12)所示。

$$c_t = [c_t^g; c_t^l] \quad (12)$$

NARM 提出了一个双线性解码方案,即采用双线性相似性函数计算当前会话(编码器)的输出与每个候选项目之间的相似度评分 S_i , S_i 的计算方法如式(13)所示。

$$S_i = \text{emb}_i B c_t \quad (13)$$

其中, B 为一个 $|D| \times |H|$ 的矩阵, $|D|$ 是每一个项目嵌入的维度,emb_{*i*} 指候选项目的嵌入表示。计算出每个候选项目与编码器输出 c_t 的相似度分数后,输入到 Softmax 层计算每个候选项目出现的概率,即可实现下一个项推荐的目标。解码器部分以及 NARM 的总体模型图如图 5 所示。

神经注意力推荐模型(NARM)是一个具有编码-解码器(Encoder-Decoder)结构的会话推荐模型。NARM 不仅能够关注用户行为的序列模式,还能捕捉用户当前的主要意图,从而生成更准确的推荐。

与 NARM 类似,Liu 等人^[40]提出了短期注意/记忆优先模型(Short-Term Attention/Memory Priority Model, STAMP),STAMP 的目的也是通过注意力机制强调用户的主要目的,但与 NARM 不同的是,STAMP 更强调最后一次点击对推荐系统学习用户兴趣的重要性,因此,STAMP 直接采用最后一次点击的项目表示作为用户的短期兴趣偏好,即 $s_t = v_t$ 。

为了计算用户一般兴趣,STAMP 设计了一个注意力网络(Attention Net)来明确用户的历史交互行为与最后一次交互行为的相关性,该注意力网络包括两个部分:

(1) 采用前馈神经网络(Feedforward Neural Network, FNN)为会话内每个项目生成注意力权重;

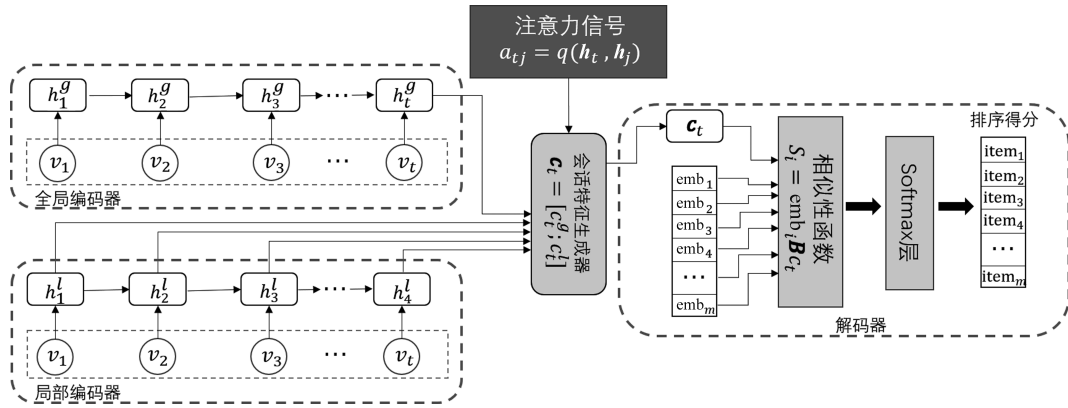


图 5 NARM 的模型结构图

(2) 合成注意力函数,生成用于基于注意力的一般兴趣 s_a 。

其计算方法如式(14)所示。

$$a_i = W_0 \sigma(W_1 v_i + W_2 v_t + W_3 s_t + b_a)$$

$$s_a = \sum_{i=1}^t a_i v_i \quad (14)$$

其中, v_t 指用户最后一次交互的项目的向量表示, s_t 为当前会话未添加注意力机制的一般兴趣表示,其计算方法如式(15)所示。

$$s_t = \sum_{i=1}^t \frac{1}{t} v_i \quad (15)$$

STAMP 提出的短期注意/记忆优先模型关注会话最后一次交互的重要性,其在基于会话的推荐领域效果明显,此后的基于深度学习的方法,常采用这种注意力机制来增强模型性能。

在基于会话的推荐系统中,由于误操作或偶然性操作,会话数据存在部分无效信息,这些信息与用户当前的偏好相关性较低,因此,在学习用户兴趣偏好时如何忽略这些信息是有必要的。在基于会话的推荐方法中,注意力机制本质上就是计算会话中各

个项目的注意力权重,并采用加权求和的方式生成最终的会话向量。它能够捕捉会话序列中用户感兴趣的信息,忽略无效信息,增强用户在当前会话中的主要意图。

2.2.3 基于图神经网络的会话推荐方法

以上基于深度学习的会话推荐方法主要聚焦于用户行为的序列模式,通过模拟连续项目之间的单项转换学习用户的偏好,忽略了会话中上下文之间的转换,即会话中其他的项目(包括同一会话中的项目以及不同会话中的项目)。图神经网络^[41-42]能够通过图节点之间的消息传递来捕捉节点间的依赖,随着图嵌入学习^[43]的发展,图神经网络也被广泛应用于基于会话的推荐方法中^[44-47]。

基于图神经网络的会话推荐(Session-based Recommendation with Graph Neural Networks, SR-GNN)^[48]首次将图神经网络应用到基于会话的推荐方法中。SR-GNN 将会话序列构建成只包含一种类型节点的图结构,并采用 GNN 捕捉项目间复杂的转换关系,生成表达能力更强的潜在项目表示向量。其模型结构如图 6 所示。

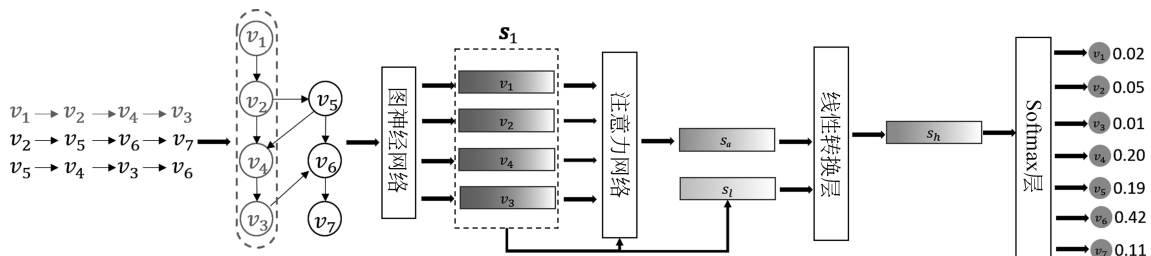


图 6 SR-GNN 的模型结构图

SR-GNN 将会话序列构建成为一个有向图 $G_s = (V_s, \epsilon_s)$, 图的每一个节点表示会话集中出现的项目 $v_i \in V_s$, 边 $(v_{i-1}, v_i) \in \epsilon_s$ 表示用户点击项目

v_{i-1} 后点击了项目 v_i 。构建会话图后,采用基于门控的图神经网络提取会话图特征,学习出项目的嵌入表示。基于门控的图神经网络的计算如式(16)

所示。

$$\begin{aligned} \mathbf{a}_{s,i}^t &= \mathbf{A}_{s,i} [\mathbf{v}_i^{t-1}, \dots, \mathbf{v}_n^{t-1}]^T \mathbf{H} + \mathbf{b} \\ \mathbf{q}(\mathbf{h}_i, \mathbf{h}_j) &= \mathbf{v}^T \sigma(\mathbf{A}_1 \mathbf{h}_i + \mathbf{A}_2 \mathbf{h}_j) \\ \mathbf{r}_{s,i}^t &= \sigma(\mathbf{W}_r \mathbf{a}_{s,i}^t + \mathbf{U}_r \mathbf{v}_i^{t-1}) \\ \tilde{\mathbf{v}}_i^t &= \tanh(\mathbf{W}_o \mathbf{a}_{s,i}^t + \mathbf{U}_o (\mathbf{r}_{s,i}^t \mathbf{v}_i^{t-1})) \end{aligned} \quad (16)$$

得到项目嵌入向量后,SR-GNN 采用 STAMP 中的注意力网络生成会话嵌入表示 \mathbf{s}_h , 计算方法如式(9)所示。最终,会话嵌入 \mathbf{s}_h 经过 Softmax 层后预测出所有候选项目的概率,选择 Top- k 个项目生成推荐列表。

SR-GNN 将图模型应用到会话推荐方法中,考虑了会话序列项目之间的复杂结构和转换关系,通过实验证明图模型能够显著提高基于会话推荐的效果,因此,众多学者应用图模型解决会话推荐中不同问题,并提出了一系列模型结构。针对会话中可能包含与当前会话相关的项目和无关的项目的转换的问题,Wang 等人^[49]提出了全局上下文增强图神经网络(GCE-GNN)用于基于会话的推荐。GCE-GNN 分别将局部会话序列和全局会话序列构建成局部会话图和全局会话图,采用图神经网络学习会话级项目嵌入以及全局项目嵌入,并通过注意力机制将会话级项目嵌入以及全局项目嵌入合并,最终生成会话序列的整体会话向量表示。Guo 等人^[50]提出了序列依赖增强图神经网络(SDE-GNN),采用长短时记忆网络(Long Short Term Memory Network,LSTM)和 GNN 两种模型结合的方式,设计了顺序依赖学习模块和项目转换学习模块,分别学习会话中的顺序依赖和项目转换。SDE-GNN 不仅能够解决 GNN 传播过程中节点信息存在的梯度消失和信息丢失问题,还能减少 GNN 中节点信息传播过程中无关节点(项)对目标节点的干扰。Yu 等人^[51]提出了目标注意图神经网络(Target Attentive Graph Neural Networks, TAGNN)模型,TAGNN 将会话序列构建成会话图,通过图神经网络学习项目嵌入,TAGNN 还设计了一个目标感知注意力机制,使模型捕捉项目间丰富的转换关系的同时,还能够学习到用户对不同项目兴趣的变化。

基于图神经网络的会话推荐方法,通过构建图结构(非欧氏空间),挖掘项目间的依赖关系。这些依赖关系主要体现在会话数据集中:一种是同一会话内项目间的顺序关系,另一种是不同会话内项目的依赖关系。图神经网络通过学习图中节点信息和网络拓扑结构(节点间的关系),将图中的节点(项目)表示为低维向量,在通过某种连接层(通常采用注意力网络或全连接网络)生成当前会话的嵌入表示。

2.2.4 其他基于深度学习的会话推荐方法

除了上文提到的基于循环神经网络、注意力机制、图神经网络的会话推荐方法外,近两年基于深度学习的会话推荐方法快速发展。

在实际的场景中,与用户交互的下一个项目往往是由之前多个项目共同作用触发的,即可能被推荐的下一个项目与之前的项目存在更高阶的关系,并非是简单的两两相互依赖(一阶依赖)。超图(Hypergraph)^[52-55]提供了一种方式去模拟项目间的高阶依赖,即一条边可以连接多个节点。为了捕捉项目间的高阶依赖,Xia 等人^[52]将会话集构建成超图,并采用双通道超图卷积网络(Dual-channel Hypergraph Convolutional Network, DHCN)。在 DHCN 中,每个会话 s 中的项目 $v_{s,i}$ 组成一个超边,这意味着在一个会话中任意两个项目都是相连的;构建成超图后,将每个会话建模为一个节点,不同会话之间通过共享项链接,生成该超图的线图(Line Graph),线图能够建模会话级的依赖。DHCN 构建超图捕捉项目之间的高阶依赖,能够学习出项目之间复杂的转换关系,从而提升推荐的准确度。

以上方法挖掘项目之间的相关性,往往只考虑项目之间的序列模式,而忽略了时间因素对这种相关性的影响。如图 7 所示,用户在浏览项目 v_1 后间隔 2s 浏览了项目 v_2, v_3 , 20s 后浏览了项目 v_4 ,那么在会话 $s = \{v_1, v_2, v_3, v_4\}$ 中,项目 v_1, v_2, v_3 之间相关性更强,而项目 v_4 与会话中其他项目相关性较低。由此可见,用户的兴趣会随时间的变化偏移,时间因素对于分析项目之间的相关性非常重要。

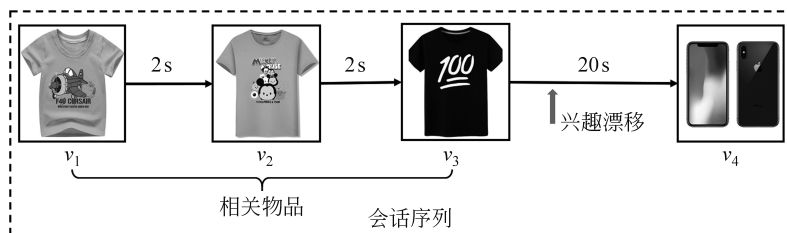


图 7 时间因素对项目相关性的影响

Guo 等人^[56]首次将时间间隔信息引入基于会话的推荐方法,提出了时间感知图神经网络模型(Time-Aware Graph Neural Network, TA-GNN)。TA-GNN 拥有一个时间感知生成器,它根据两个节点之间的时间间隔建立时间感知邻接矩阵,边的权重与时间间隔成反比。除此之外,Lv 等人^[57]提出的时间敏感的协同兴趣感知模型(Time-Sensitive Collaborative Interest Aware Model, TSCIA),Chen 等人^[58]提出的时效性增强注意力网络(Time-liness Enhanced Attention Network, TEAN),也将时间因素作为模型输入的重要特征,并显著提升了推荐效果。

在会话推荐中,大部分方法只能根据短期会话中产生的有限项目来完善相应的用户或会话表示,这会产生数据稀疏的问题。近年来,对比学习作为一类新兴的解决数据稀疏的方法,在会话推荐中逐

渐推广^[52,59-60]。Xia 等人^[59]提出了基于会话推荐的自监督图协同训练框架(Self-supervised Graph CO-Training Framework for Session-based Recommendation, COTREC)来增强基于会话的推荐。COTREC 首先利用会话-项目图来构建两个视图(项目视图和会话视图),展示会话的内部和外部连接。然后在这两个视图上建立两个不同的图编码器,并在共同训练的方案下进行训练。其中一个(主编码器)用于推荐,另一个作为辅助编码器来提升前者。这两个编码器递归地利用不同的连接信息生成可能的下一个项目,通过对比学习相互监督。

基于深度学习的会话推荐方法,能够建模会话集中丰富的数据信息,不同的深度学习模型都拥有其各自的优势,能够解决基于会话推荐中遇到的不同问题。表 5 列出了不同深度学习模型的工作机制和优缺点。

表 5 基于不同深度学习模型的会话推荐方法比较

模型	工作机制	优点	缺点
循环神经网络	建模项目的顺序关系,挖掘项目间的序列模式	对项目间的顺序关系敏感	只能模拟连续项目之间的单项转换
注意力机制	为每个会话中的项目生成注意力权重,合成注意力函数	能够忽略用户的一些错误交互行为,并强调用户当前的意图	学习出的长期偏好并非传统推荐方法中的一般兴趣偏好
图神经网络	建模会话上下文项目的转换	能够考虑项目之间的复杂结构和转换关系	只能考虑成对项目之间的依赖关系
超图卷积网络	将会话数据构建成超图,每个会话中的所有项目组成一个超边,模拟项目间的高阶依赖	能够考虑多个项目之间的高阶依赖	放弃了对项目间的序列行为建模
时间感知网络	将时间因素作为用户兴趣偏好学习的关键特征	解决用户的兴趣随时间的变化漂移的问题	针对不同场景时间窗口不同,时间因素对推荐效果的影响无法确定
对比学习	通过构造正负样例来学习特征	解决数据稀疏的问题	时间复杂度比较高

2.3 基于强化学习的会话推荐方法

会话推荐通常以自监督进行训练,给定会话序列,计算下一项推荐。这种训练模式可能会根据项目相关性找出最优的结果,但是对于用户来说有可能还有其他需求。近年来,推荐结果多样性和新颖性的重要程度逐渐提高,因为推荐多样化的项目更有可能符合用户的实际需求。

为了针对上述多个目标优化推荐系统,通常需要用可微函数进行优化,然而在目标只能以不可微的形式呈现的领域难以使用多目标优化(Multi-Objective Optimization, MOO)。为了解决这一问题,Stamenkovic 等人^[61]使用强化学习(Reinforcement Learning, RL)的策略,引入了标量化多目标强

化学习(Scalarized Multi-Objective Reinforcement Learning, SMORL)方法。SMORL 使用单个强化学习智能体(Agent)同时满足三个可能相互冲突的目标:提高点击率,使推荐多样化,以及引入新项目。该模型专注于选择的奖励,同时保持高相关性排名性能。

SMORL 在会话推荐中的应用如图 8 所示。首先使用生成模型 G 将用户-项目交互序列 $x_{1:t}$ 映射到潜在状态 s_t 。然后使用全连接层将 s_t 映射到 y_{t+1} 和三个一维 Q 值: Q_{acc} 、 Q_{div} 和 Q_{nov} 。传统的自监督学习通过交叉熵损失进行训练以进行排名。向量值 Q 值函数设置为: $Q = [Q_{acc}, Q_{div}, Q_{nov}]$ 。SDQL 损失是通过标量函数: $f_w(x) = wx$ 和 SDQL 算法得到的,与交叉熵损失一起用于训练模型,以改

进自监督学习的排名。

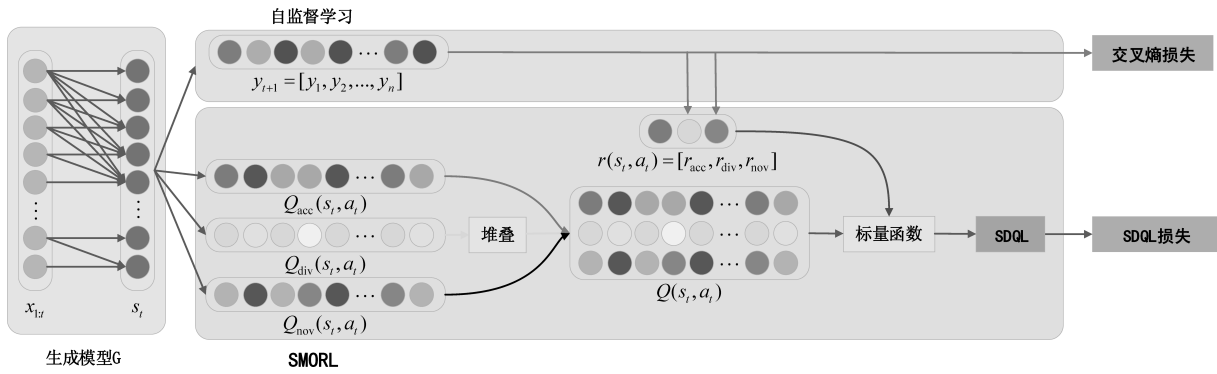


图 8 SMORL 在会话推荐的应用

3 数据优化及数据集

基于会话的推荐方法与传统推荐方法或序列推荐方法^[11]的不同点也体现在数据结构上,基于会话的推荐方法是以会话作为数据的基本单元。会话是用户行为被分解后形成的一个个有序项目组,在会话间不仅能够找到项目序的关系,也能够聚焦于用户行为的事务特性^①。

在深度学习被应用于基于会话的推荐领域初期,由于训练数据存在一些问题,导致最终的推荐效果并不是十分理想。因此,许多学者在数据预处理、排名损失函数^[62]等方面进行了优化。其中较为典型的是 Tan 等人^[63]提出的优化策略。

(1) 数据预处理

采用序列预处理方法^[64],即对于会话 $s = \{v_1, v_2, v_3, v_4\}$,在构建数据集时,将会话 s 分割成多个子序列及其标签: $\{v_1, T(v_2)\}$ 、 $\{v_1, v_2, T(v_3)\}$ 、

$\{v_1, v_2, v_3, T(v_4)\}$ 。该方法能有效增强训练,并减少过拟合现象的发生。

(2) 嵌入丢弃

嵌入丢弃(Embedding Dropout)是一种应用于输入序列的正则化形式^[65]。将其应用于基于会话的推荐方法中,用于随机删除会话中某一个项目。这种预处理的方法可以防止 RNN 陷入过拟合的情况。

(3) 特权信息

特权信息(Privileged Information)指未来的序列,如在会话 $s = \{v_1, v_2, \dots, v_r, v_{r+1}, \dots, v_n\}$,当根据 $\{v_1, v_2, \dots, v_{r-1}\}$ (原数据)预测 v_r 时, $\{v_n, \dots, v_{r+1}\}$ 即特权信息。Mikolov 等人^[66]提出采用广义蒸馏框架(Generalized Distillation Framework),分别训练一个 Teacher Model(利用特权信息)和一个 Student Model(利用原数据),这种方法对于小规模数据集帮助很大。

同时,本文也收集并整理了部分开源的数据集,详细信息如表 6 所示。

表 6 数据集

数据集	会话数量	项目数量	交互记录数量	平均会话长度	所属领域
Yoochoose ¹	5 917 745	29 618	8 326 407	5.71	电子商务
Diginetica ²	719 470	43 097	982 961	5.12	电子商务
Tmall ³	1 774 729	425 348	13 418 695	7.56	电子商务
Last.FM ⁴	169 576	449 037	2 887 349	17.03	音乐
NowPlaying ⁵	126 249	30 673	976 702	7.73	音乐
Globo ⁶	1 031 167	13 092	2 930 849	2.84	新闻
CLEF ⁷	1 644 442	742	5 540 486	3.37	新闻

注: 1. <https://2015.recsyschallenge.com/challenge.html>
 2. <http://cikm2016.cs.iupui.edu/cikm-cup>
 3. <https://tianchi.aliyun.com/dataset/dataDetail?dataId=42>
 4. <http://millionsongdataset.com/lastfm/>
 5. <https://www.kaggle.com/datasets/chelseapower/nowplaying-grs>

6. <https://www.koggle.com/datasets/gspmoreira/news-portal-user-interactions-by-grobocom>
 7. <https://www.news.rellchallenge.org/dataset/>

① 即关注会话的主要内容和目的

4 基于会话的推荐方法未来研究方向

近年来,基于会话的推荐方法快速发展,众多学者针对已存在的问题提出了不同的模型方法,如上文提到的利用循环神经网络挖掘会话中项目的序列模式、利用图神经网络建模项目间复杂的转换关系等。但目前对于基于会话的推荐方法的研究仍处于初级阶段,仍存在一些挑战,这些挑战是未来值得研究的方向。本节将针对基于会话的推荐方法可能面临的挑战进行阐述。

(1) 基于会话的推荐中的长期偏好

在推荐系统中,用户的偏好往往分为长期偏好(静态偏好)和短期偏好(动态偏好)。基于会话的推荐方法往往是通过学习出当前会话的嵌入向量去进行下一个项目的推荐,通常这只能捕捉用户的短期偏好(动态偏好),更多关注用户兴趣的变化。然而,在推荐系统的实际应用中,用户的长期偏好和短期偏好对于提升推荐方法的性能都是十分重要的。因此,为了提高基于会话的推荐方法的推荐效果,挖掘用户的长期偏好是有必要的。然而,会话具有事务属性,连续的会话相关性可能不强,使用连续的会话数据学习用户长期偏好的方法不可行。因此,如何

在会话数据集中构造长期相关会话(邻居会话)^[67]数据是学习长期偏好的关键。

(2) 用户多行为联合推荐

在实际场景中,用户的交互行为常常是多种类型操作同时进行的。例如,在购物场景中,用户的行为可能包含点击商品、添加购物车、购买商品等,而且这些类型的操作是无序的,即用户在进行点击操作后可能进行添加购物车操作或购买操作。而不同类型的交互行为对于下一项可能推荐的项目的影响是不同的。在以往研究所用的数据集中,每个会话往往是由同种交互类型的项目组成的,但是,下一个可能推荐的项目可能也会受到其他类型交互的影响。如图9所示,用户在会话 s_1 中进行了如下操作,连续点击项目 v_1, v_2, v_3 后进行购买项目 v_3 操作,其中项目 v_1, v_2, v_3 都属于T恤类别。用户在购买操作后,对当前类别商品的感兴趣程度降低,因此推荐方法应捕捉到用户当前的兴趣变化,为用户推荐其他类别的商品,如图9中的项目 v_4 。在会话 s_2 中,用户在进行点击操作后,执行了添加购物车行为,此时,用户对当前类别的商品仍十分感兴趣。由此可见,用户不同的交互行为会影响下一个项目的推荐。如何将用户多种行为结合,实现用户多行为的联合推荐是一个值得研究的方向。

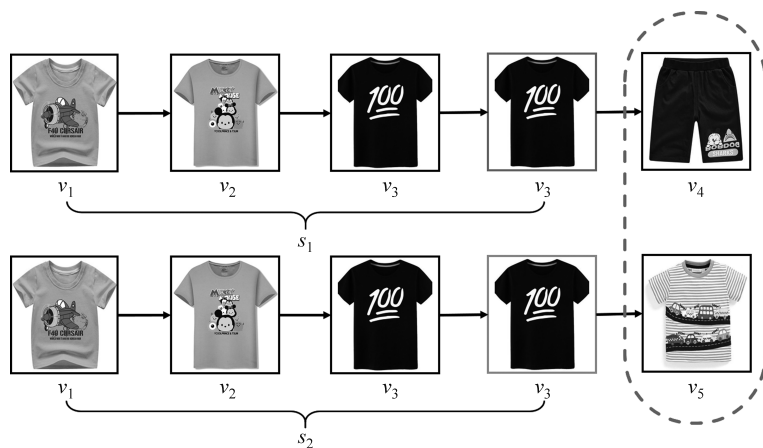


图9 用户多种类型交互行为构成的会话

其中,会话 s_1 和 s_2 的前三项表示用户的点击行为, s_1 的第二个 v_3 表示用户的购买行为, s_2 的第二个 v_3 表示用户的添加购物车行为。

(3) 下一个会话的推荐

目前已有的基于会话的推荐方法,大多数是根据当前会话的项目序列去推荐用户感兴趣的项目,推荐项目仍属于当前会话,这种推荐方式可以被称为当前会话的下一项推荐,但该方式存在一个问题,当一个新的会话产生时,基于会话的推荐方法无法

为用户推荐会话的第一个项目。虽然推荐方法可以根据用户的历史会话来挖掘用户偏好,但是若上一个会话距当前的时间间隔较长,则分析上一个会话的项目序列作为当前用户兴趣偏好的意义不大,因为推荐方法无法捕捉用户当前的兴趣偏好,从而无法发挥基于会话推荐方法的优势(关注用户的短期

偏好)。近年来,对于下一会话推荐^[68]的研究较少,但它仍是一个值得研究的方向。

(4) 跨域会话推荐

现有的会话推荐大多数是针对匿名用户的会话行为进行推荐,而在会话推荐中引入用户信息以提高推荐性能成为最近的研究热点。然而在实际使用中会出现用户信息稀疏的问题,此时可以考虑与跨域推荐相结合以缓解数据稀疏的问题^[69]。但是跨域推荐与会话推荐的结合面临着诸多的挑战。首先是不同域之间数据的差异性。由于不同的数据域包括不同种类的项目,如电影域和音乐域,将这两个域的数据进行结合时需要考虑如何进行跨域转换。其次是时间异步性,即用户在不同域上的相关行为可能时间跨度特别大。如用户多年前看过哈利波特小说,但可能最近才去看哈利波特的电影。因此对不同域的会话行为不能简单地按照发生的时间进行划分。

5 结论

本文对基于会话的推荐方法进行了详细地调研。首先分析了传统推荐方法的优劣势,针对其只能学习用户长期偏好的问题,阐明了基于会话的推荐方法捕捉用户短期偏好的重要意义,并详细介绍了基于会话推荐方法的工作原理和目标。其次,本文对当前已有的基于会话的推荐算法进行了分类总结,对比了基于不同模型下的推荐方法的优劣势,并详细介绍了这些模型方法的工作原理,如基于循环神经网络的会话推荐方法建模会话中项目的序列模式,基于注意力机制的会话推荐方法捕捉用户当前的主要意图,以及基于图神经网络的会话推荐方法建模会话上下文项目的转换等。最后,针对基于会话的推荐方法可能面临的挑战和有潜力的研究方向进行了说明。

参考文献

- [1] 许海玲, 吴潇, 李晓东, 等. 互联网推荐系统比较研究[J]. 软件学报, 2009, 20(02): 350-362.
- [2] 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(07): 1619-1647.
- [3] SCHAFFER J B, KONSTAN J A, RIEDL J. E-commerce recommendation applications[J]. *Data Mining and Knowledge Discovery*, 2001, 5(1): 115-153.
- [4] 孟祥武, 陈诚, 张玉洁. 移动新闻推荐技术及其应用研究综述[J]. 计算机学报, 2016, 39(04): 685-703.
- [5] GONG S, ZHU K Q. Positive, negative and neutral: Modeling implicit feedback in session-based news recommendation[C]//*Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2022: 1185-1195.
- [6] 李琳, 唐守廉. 基于多层注意力表示的音乐推荐模型[J]. 电子学报, 2020, 48(9): 1672-1679.
- [7] 刘树栋, 孟祥武. 一种基于移动用户位置的网络服务推荐方法[J]. 软件学报, 2014, 25(11): 2556-2574.
- [8] MOONEY R J, ROY L. Content-based book recommending using learning for text categorization[C]//*Proceedings of the 5th ACM Conference on Digital Libraries*, 2000: 195-204.
- [9] 于洪, 李俊华. 一种解决新项目冷启动问题的推荐算法[J]. 软件学报, 2015, 26(6): 1395-1408.
- [10] EKSTRAND M D, RIEDL J T, KONSTAN J A. Collaborative filtering recommender systems [J]. *Foundations and Trends® in Human-Computer Interaction*, 2011, 4(2): 81-173.
- [11] WANG S, HU L, WANG Y, et al. Sequential recommender systems: Challenges, progress and prospects[C]//*Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019: 6332-6338.
- [12] WANG S, CAO L, WANG Y. A survey on session-based recommender systems[J]. *arXiv preprint arXiv: 1902.04864*, 2019.
- [13] 肖楠. 基于会话的推荐算法研究综述[J]. 现代计算机, 2019(36): 33-36.
- [14] 任永功, 张云鹏, 张志鹏. 基于粗糙集规则提取的协同过滤推荐算法[J]. 通信学报, 2020, 41(01): 76-83.
- [15] 荣辉桂, 火生旭, 胡春华等. 基于用户相似度的协同过滤推荐算法[J]. 通信学报, 2014(35), 314(02): 16-24.
- [16] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9): 1621-1628.
- [17] JANNACH D, ZANKER M. Collaborative filtering: Matrix completion and session-based recommendation Tasks [J]. *Collaborative Recommendations- Algorithms, Practical Challenges and Applications*, 2018: 1-34.
- [18] RESNICK P, IACOVOU N, SUCHAK M, et al. Grouplens: An open architecture for collaborative filtering of netnews[C]//*Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 1994: 175-186.
- [19] BONNIN G, JANNACH D. Automated generation of music playlists: Survey and experiments [J]. *ACM Computing Surveys*, 2014, 47(2): 1-35.

- [20] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems [J]. *Computer*, 2009, 42(8): 30-37.
- [21] EIRINAKI M, VAZIRGIANNIS M, KAPOGIANIS D. Web path recommendations based on page ranking and markov models[C]//*Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management*, 2005: 2-9.
- [22] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [23] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[J]. *Advances in Neural Information Processing Systems*, 2014, 27: 185-196.
- [24] YIN W, SCHÜTZ H. Task-specific attentive pooling of phrase alignments contributes to sentence matching[C]//*Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 2017: 699-709.
- [25] HIDASI B, KARATZOGLOU A, BALTRUNAS L, et al. Session-based recommendations with recurrent neural networks [J]. *arXiv preprint arXiv: 1511.06939*, 2015.
- [26] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. *arXiv preprint arXiv: 1412.3555*, 2014.
- [27] 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.
- [28] BOGINA V, KUFLIK T. Incorporating dwell time in session-based recommendations with recurrent neural networks[C]//*Proceedings of the RecTemp@ RecSys*, 2017: 57-59.
- [29] SUN Y, ZHAO P, ZHANG H. Ta4rec: Recurrent neural networks with time attention factors for session-based recommendations[C]//*Proceedings of the International Joint Conference on Neural Networks*. IEEE, 2018: 1-7.
- [30] RUOCCO M, SKREDE O S L, LANGSETH H. Inter-session modeling for session-based recommendation[C]//*Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems*, 2017: 24-31.
- [31] 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.
- [32] PHUONG T M, THANH T C, BACH N X. Combining user-based and session-based recommendations with recurrent neural networks[C]//*Proceedings of the International Conference on Neural Information Processing*. Springer, Cham, 2018: 487-498.
- [33] SONG Y, LEE J G. Augmenting recurrent neural networks with high-order user-contextual preference for session-based recommendation[J]. *arXiv preprint arXiv: 1805.02983*, 2018.
- [34] QU Y, CAI H, REN K, et al. Product-based neural networks for user response prediction[C]//*Proceedings of the IEEE 16th International Conference on Data Mining*. 2016: 1149-1154.
- [35] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. *arXiv preprint arXiv: 1409.0473*, 2014.
- [36] CHEN C, GUO J, SONG B. Dual attention transfer in session-based recommendation with multi-dimensional integration[C]//*Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021: 869-878.
- [37] YUAN J, SONG Z, SUN M, et al. Dual sparse attention network for session-based recommendation [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(5): 4635-4643.
- [38] LI J, REN P, CHEN Z, et al. Neural attentive session-based recommendation[C]//*Proceedings of the ACM on Conference on Information and Knowledge Management*, 2017: 1419-1428.
- [39] SHANG L, LU Z, LI H. Neural responding machine for short-text conversation[J]. *arXiv preprint arXiv: 1503.02364*, 2015.
- [40] 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.
- [41] ZHOU J, CUI G, HU S, et al. Graph neural networks: A review of methods and applications[J]. *AI Open*, 2020, 1: 57-81.
- [42] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 32(1): 4-24.
- [43] GOYAL P, FERRARA E. Graph embedding techniques, applications, and performance: A survey[J]. *Knowledge Based Systems*, 2018: 78-94.
- [44] ZHOU H, TAN Q, HUANG X, et al. Temporal augmented graph neural networks for session-based recommendations[C]//*Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021: 1798-1802.

- [45] LAI S, MENG E, ZHANG F, et al. An attribute-driven mirror graph network for session-based recommendation[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022: 1674-1683.
- [46] GUO J, YANG Y, SONG X, et al. Learning multi-granularity consecutive user intent unit for session-based recommendation[C]//Proceedings of the 15th ACM International Conference on Web Search and Data Mining, 2022: 343-352.
- [47] PANG Y, WU L, Shen Q, et al. Heterogeneous global graph neural networks for personalized session-based recommendation[C]//Proceedings of the 15th ACM International Conference on Web Search and Data Mining, 2022: 775-783.
- [48] WU S, TANG Y, ZHU Y, et al. Session-based recommendation with graph neural network[C]//Proceedings of the National Conference on Artificial Intelligence, 2019, 33(01): 346-353.
- [49] 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.
- [50] GUO W, WANG S, LU W, et al. Sequential dependency enhanced graph neural networks for session-based recommendations [C]//Proceedings of the IEEE 8th International Conference on Data Science and Advanced Analytics, 2021: 1-10.
- [51] YU F, ZHU Y, LIU Q, et al. TAGNN: Target attentive graph neural networks for session-based recommendation[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020: 1921-1924.
- [52] XIA X, YIN H, YU J, et al. Self-supervised hypergraph convolutional networks for session-based recommendation[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(5): 4503-4511.
- [53] YADATI N, NIMISHAKAVI M, Yadav P, et al. Hypergn: A new method for training graph convolutional networks on hypergraphs[J]. Advances in Neural Information Processing Systems, 2019: 1509-1520.
- [54] LI Y, GAO C, LUO H, et al. Enhancing hypergraph neural networks with intent disentanglement for session-based recommendation[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022: 1997-2002.
- [55] ZHANG X, XU B, YANG L, et al. Price DOES matter! Modeling price and interest preferences in session-based recommendation [C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022: 1684-1693.
- [56] GUO Y, LING Y, CHEN H. A Time-aware graph neural network for session-based recommendation[J]. IEEE Access, 2020, 8: 167371-167382.
- [57] LV Y, ZHUANG L, LUO P, et al. Time-sensitive collaborative interest aware model for session-based recommendation[C]//Proceedings of the IEEE International Conference on Multimedia and Expo, 2020: 1-6.
- [58] CHEN D, ZHANG X, WANG H, et al. TEAN: Timeliness enhanced attention network for session-based recommendation[J]. Neurocomputing, 2020, 411: 229-238.
- [59] XIA X, YIN H, YU J, et al. Self-supervised graph co-training for session-based recommendation[C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management, 2021: 2180-2190.
- [60] WANG L, XU X, OUYANG K, et al. Self-supervised dual-channel attentive network for session-based social recommendation[C]//Proceedings of the IEEE 38th International Conference on Data Engineering. IEEE, 2022: 2034-2045.
- [61] STAMENKOVIC D, KARATZOGLOU A, ARAPAKIS I, et al. Choosing the best of both worlds: Diverse and novel recommendations through multi-objective reinforcement learning[C]//Proceedings of the 15th ACM International Conference on Web Search and Data Mining, 2022: 957-965.
- [62] 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.
- [63] 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.
- [64] DE BRÉBISSON A, SIMON É, AUVOLAT A, et al. Artificial neural networks applied to taxi destination prediction [J]. arXiv preprint arXiv: 1508.00021, 2015.
- [65] GAL Y, GHAHRAMANI Z. A theoretically grounded application of dropout in recurrent neural networks [J]. Advances in Neural Information Processing Systems, 2016: 1019-1027.

- [23] WIKIMEDIA. Wikipedia[EB/OL]. <http://wikipedia.org/>[2021-05-20].
- [24] KIM S, KANG I, KWAK N. Semantic sentence matching with densely-connected recurrent and co-attentive information [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(01): 6586-6593.
- [25] CHEN Q, ZHU X, LING Z H, et al. Enhanced LSTM for natural language inference [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017: 1657-1668.
- [26] HE H, LIN J. Pairwise word interaction modeling with deep neural networks for semantic similarity measurement [C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016: 937-948.
- [27] NIE Y, BANSAL M. Shortcut-stacked sentence encoders for multi-domain inference [C]//Proceedings of the 2nd Workshop on Evaluating Vector Space Representations for NLP, 2017: 41-45.



李志峰(1998—), 硕士, 主要研究领域为自动问答、常识问答、文本匹配。

E-mail: li_zaaachary@163.com



白岩(1978—), 硕士, 主要研究领域为文本匹配、自动问答。

E-mail: paladnix@outlook.com



洪宇(1978—), 通信作者, 博士, 教授, 主要研究领域为信息抽取, 篇章关系理解, 多模态机器翻译及智能问答。

E-mail: tianxianer@gmail.com

(上接第 17 页)

- [66] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [J]. Advances in Neural Information Processing Systems, 2013, 26.
- [67] HAN P, WANG N, LI K, et al. CASR: A collaborative attention model for session-based recommendation [C]//Proceedings of the IEEE 45th Annual Computers, Software, and Applications Conference, IEEE, 2021: 290-296.
- [68] WAN S, LAN Y, WANG P, et al. Next basket recommendation with neural networks [C]//Proceedings of the RecSys Posters, 2015.
- [69] WANG Y, GUO C, CHU Y, et al. A cross-domain hierarchical recurrent model for personalized session-based recommendations [J]. Neurocomputing, 2020, 380: 271-284.



陈晋鹏(1985—), 通信作者, 博士, 副教授, 博士生导师, 主要研究领域为社会媒体与数据挖掘、人工智能及应用。

E-mail: jpchen@bupt.edu.cn



李海洋(1996—), 硕士, 主要研究领域为机器学习、推荐系统。

E-mail: haiyangli@bupt.edu.cn



张帆(1997—), 博士, 主要研究领域为机器学习、推荐系统。

E-mail: zhang_fan@bupt.edu.cn