会话推荐方向总结整理

1会话推荐的含义

会话推荐（Session-based recommendation，SBR）是指仅仅利用用户在最近一段时间内的交互物品序列，以预测用户下一次可能交互的物品。由于长期用户属性通常由于隐私政策问题而不可用，因此在物品转换过程中捕捉用户的意图变得尤为重要。在会话推荐中，会话长度相对较短，而未登录用户的用户画像完全未知。数据总是在很短的时间段内产生，可能具有暂时的依赖性，因此SBR主要关注建模用户的近期实时兴趣，可被视为序列推荐的子领域。会话推荐方法可以在不侵犯用户隐私的情况下提供一定质量的推荐，这是一个值得研究的重要方向。在工业界的主流推荐场景中，通常包括召回、粗排、精排和重排四个阶段。基于会话推荐的相关工作主要应用于在线上的召回阶段。

2 传统的会话推荐方法

传统的SBRS方法利用传统的数据挖掘或机器学习技术，为基于会话的建议捕获嵌入会话数据中的依赖项。一般来说，SBRS的传统方法有三类：基于模式/规则挖掘的方法、基于K近邻的方法和基于马尔可夫链的方法。

2.1基于模式/规则挖掘的方法

一般来说，SBR有两种基于模式/规则挖掘的方法：基于频繁模式/关联规则挖掘的方法在与未登录会话的不同交互中挖掘关联规则；基于序列模式挖掘的方法在不同交互中挖掘序列模式与有序会话的交互顺序。这类方法只能处理数据集中所有操作都相同的单类型操作会话，因此，给定会话中的每个交互都被简化为一个项目。

2.2 基于最近邻模型的算法

首先寻找和目标物品或会话相近的物品和会话，然后通过计算候选集物品与当前会话的相似度来推荐。根据是否实际计算了项目或会话之间的相似度，基于KNN的SBRS方法可分为item-KNN和session-KNN。给定当前会话的上下文，基于item-KNN的SBR推荐与先前会话序列中物品相似的物品,相似度是2 个物品向量的余弦相似度。相比于item-KNN，session-KNN考虑整个会话上下文，而不仅仅是会话上下文中的当前项目，因此可以捕获更多的内容。

2.3 基于马尔可夫链的方法

基于马尔可夫链的SBRS采用马尔可夫链对会话内或会话间的交互转换进行建模，以预测给定会话上下文的下一次交互或会话。根据跃迁概率是基于显式观测还是基于潜空间计算，基于马尔可夫链的方法大致可分为基于基本马尔可夫链的方法和基于潜在马尔可夫嵌入的方法。基于马尔可夫链的方法擅长建模短期和低阶的顺序依赖，但通常忽略长期和高阶的依赖，其依赖的强假设：下一状态只能由当前状态决定，在时间序列中之前的行为均与之无关，导致在实际场景中运用受限。

这些传统的SBR 算法在物品间依赖关系比较简单的数据集上效果可以和基于深度学习的SBR算法相媲美，但是其忽略了用户的动态行为，无法建模会话内和会话间物品之间复杂的高阶依赖关系。

1. 基于深度学习的方法

**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,会话中的项目具有顺序关系，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。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强的项目嵌入表示,在基于会话的推荐方法中广泛应用…接下来将根据模型类别对已有的推荐方法进行分类总结。

从大体上来看，会话推荐主要有两种分支，一种是把会话看作一个序列，利用诸如处理流数据的RNN模型得到隐层向量，并基于此搭建预测模型。另一种是基于图神经网络GNN的方法，通过构建有向图，利用GNN的方法学习表达向量，基于此进行预测。

* 1. 基于序列的方法

基于物品的协同过滤和基于马尔可夫的决策忽略整个序列，GRU4Rec[1]用GRU对会话序列进行编码，获得会话向量以表征用户行为意图。采用小批量并行处理和采样两种策略来简化训练代价，提高训练效率，作为首个利用RNN对会话序列建模的工作，取得了很好的效果。NARM[2]发现会话内用户的兴趣分布大多数感兴趣的物品多集中在会话尾部，但并非所有物品都与下一个电机的物品相关。现阶段模型只注重会话本身，而没有挖掘用户的意图。因此基于Encoder-Decoder结构设计了两路多层GRU编码器并引入注意力机制，分别建模用户的序列行为和用户的主要意图，将注意力机制应用于对会话的顺序行为及主要意图分别建模。STAMP[3]利用多层感知机和注意力网络对会话内长短期兴趣分别表征，解决了传统的RNN结构设计不足以对通常包含由意外点击引起的用户兴趣漂移的长时间会话进行建模的问题，提出了一种新颖的短期注意力捕捉目标物品和用户兴趣的长/短期记忆的相关性。最近的工作MMSBR[4]指出现有的会话推荐方法大多集中于挖掘由商品ID暴露的有限的商品共现模式，而忽略了真正吸引用户购买商品的是页面上显示的丰富的多模态信息。基于这一动机，多模态会话推荐方法MMSBR对包括描述型信息（图片和文本）和数值型信息（价格）在内的多模态信息进行建模来捕获用户偏好，设计层次枢纽注意力机制以融合异构描述型信息全面刻画用户偏好。

3.2 基于图神经网络的方法

图神经网络因其在各种应用中令人信服的性能而越来越受欢迎。许多先前的研究也尝试将GNN应用于基于会话的推荐并获得了可喜的结果。SR-GNN[5] 提出了一个通过图来建模会话序列的全新框架。针对会话中用户行为数量有限时难以捕获用户的行为表示、RNN和马尔可夫方法只对相邻物品的单项转移进行建模，忽略了会话中未直接相连物品之间复杂的转变缺乏整体大局观的问题。它通过图神经网络来捕捉复杂的物品转移关系来建模不同的会话序列，不仅考虑了复杂的物品转移信息，也设计了一个同时考虑了局部兴趣和全局兴趣的策略，以此挖掘序列内物品之间复杂的过渡模式。

GC-SAN[6]同样关注到现有方法通常对连续项目之间的单向转换进行建模忽略了整个会话序列之间的复杂转换的问题，通过加入自注意力机制进一步扩展了SR-GNN[5]模型，在GNN获取的物品表示之上使用多个自注意力层来捕获物品间远距离依赖关系。图神经网络用于捕捉局部的依赖关系，而自注意力网络更倾向于获取全局依赖关系。除了使用注意力机制来融合序列信息之外，序列表示也被显式地加入到融合过程中来，利用自注意力网络和图神经网络的互补性来提高推荐性能。

尽管以上的两种初探索取得了不错的性能，基于图神经网络的方法仍主要存在以下三个问题：

3.2.1 图结构

目前数据集中用户的行为序列长度普遍偏短，如果只是在序列的物品之间添加节点来构建图将会限制图中的信息并难以利用GNN在图数据学习上的优势。最近的工作针对这一问题有两种主流的做法：

利用额外的序列丰富物品序列长度。GCE-GNN[7]利用了所有会话中的序列来帮助学习当前序列的行为模式，从两个维度考虑物品的转换关系，包括全局维度和会话维度。以更细微的方式利用所有会话上的物品转换关系，抽取出全局会话中与目标会话的相关信息来辅助目标会话更好的进行物品表征，以更好地推断当前会话的用户偏好。

另外一种主流做法是对当前序列的图结构进行调整。考虑到基于GNN的会话推荐方法存在两个信息丢失问题，在将会话转换为图时，会丢失序列信息，根据以前的构图方法没法还原到原始的会话。由于GNN的层数不是越深越好，所以导致GNN-based算法很难捕捉到长序依赖。LESSR[8]从信息损失和长序列这两个问题出发，构建了两种图结构。基于GRU的边缘顺序保留聚合层确保网络可以重构回会话，快捷图注意层通过沿着快捷连接传播信息来有效的捕获远程依赖关系。通过结合这两种层，能够构建一个没有信息丢失问题的模型。

3.2.2数据稀疏

COTREC[9]通过自监督学习图协同训练来对稀疏的序列进行增强，从而提升模型的性能。构建两类图分别从会话内和会话间两个角度来对关系进行挖掘。商品视角构建有向图，会话的角度构建无向图。这两个图能够相互提供互补信息，同时保持独立并在一定程度上表现出不同。在两个图上构建了两个不同的图编码器，利用不同的连接信息来生成伪标签，通过对比学习相互监督，并利用KL散度对两个编码器进行约束，即让两个编码器在一定程度上有所不同，避免两个编码器变得相似并对相同的会话生成相同的标签。

以往的会话推荐模型没有利用额外的有价值的知识，因此深受数据稀疏问题的困扰。MGS[10]提出一种用于会话推荐的镜像图增强模型，以利用项目嵌入向量上的项目属性信息来进行更准确的偏好估计。具体来说，MGS 利用两种图来学习项目表示，一种是从用户交互序列生成的会话图，描述了基于转换模式的用户偏好，另一种是由属性感知模块构建的镜像图，整合项目的属性信息为每个会话项目选择最具属性代表性的信息。应用一种迭代的双重细化机制来在会话和镜像图之间传播信息。引入对比学习策略随机采样属性相同的邻居来比较为同一会话生成的两个镜像图，实验证实性能超过了许多最先进的模型。

3.2.3高阶关系

基于序列的方法将基于会话的数据建模为单向的序列，但实际上在会话推荐中，用户的这些行为并没有严格的时间先后顺序。严格且唯一的对项目的相对顺序建模可能导致推荐模型易于过拟合。基于GNN的会话推荐方法将基于会话的数据建模为一个有向子图、将项目转换视作成对关系，这种做法忽略了项目间复杂的高阶信息。受到超图在超对关系建模的优势启发，超图最近被用来捕获物品之间的高阶关系以及跨会话信息。

为解决上述问题，DHCN[11]将每个会话视为一个超边，然后将所有会话整合起来变为一个超图。提出双通道超图卷积网络，它能够借助超图建模来捕获项目间的超对关系和跨会话信息。为解决数据稀疏问题，提高超图建模的质量，两个通道学习到的会话表示经过自监督学习使得两种会话表示间的互信息最大化，两个通道能够从彼此获取新信息，从而分别改进它们自己在项目特征提取和会话特征提取中的表现。

CoHHN[12]将价格因素引入到会话推荐任务当中，提出互导异质超图网络来从丰富的异质信息当中挖掘用户的价格偏好及兴趣偏好。在异质超图中设计双通道累积机制用来学习节点表示。基于学习到的节点表示采用注意力层来获取用户的原始价格偏好及兴趣偏好。随后设计了一个互导学习机制，建模用户价格偏好与兴趣偏好在决定用户购买行为时的复杂关系。

考虑到个人物品只能提供有限的证据来揭示用户意图，不同的用户对同一个物品的购买意图是不一样的。SHARE[13] 将每一个会话构建一个超图, 并在会话序列上应用多个滑动窗口来捕获上下文信息。首先构建每个会话的超图，同时建模各个上下文窗口中项目的相关性，引入超图注意力网络灵活聚合相关项目的上下文信息生成项目嵌入，有效生成信息丰富的动态项嵌入，推断通用和当前需求，并解码以预测下一个项目提供更准确的建议。

另一个有趣的工作MSGIFSR[14]指出，不同粒度的连续物品可能反映出不同的用户兴趣，因此希望建模这些不同粒度反映出的不同兴趣。相较于关注序列中的单个物品，关注如何利用GNN从连续片段中捕获用户偏好。作者提出了多粒度意图异构会话图，它捕获了不同粒度意图单元之间的交互并减轻了长依赖的负担。此外，作者提出了意图融合排名模块充分考虑不同意图下的会话表示，综合表达用户的偏好信息从而进行更好的推荐。

4 总结

传统推荐方法只能学习用户长期偏好，因此侧重于捕捉用户短期偏好的会话推荐方法具有重要的研究意义。对当前已有的基于会话的推荐算法按技术类别进行了分类总结，梳理了经典代表模型方法的核心思想和工作原理，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，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，基于超图的会话推荐方法被用来捕获物品之间的高阶关系以及跨会话信息。下一步研究计划结合超图的会话推荐这一方向进一步展开。

References

[1] The Session-based recommendations with recurrent neural networks ICLR2016

[2] Neural Attentive Session-based Recommendation CIKM2017

[3] Short-Term Attention/Memory Priority Model for session-based recommendation KDD 2018

[4] Beyond Co-occurrence: Multi-modal Session-based Recommendation TKDE2023

[5] Session-based recommendation with graph neural networks AAAI2019

[6] Graph Contextualized Self-Attention Network for Session-based Recommendation IJCAI2019

[7] Global Context Enhanced Graph Neural Networks for Session-based Recommendation SIGIR 2020

[8] Handling Information Loss of Graph Neural Networks for Session-based Recommendation KDD2020

[9] Self-supervised graph co-training for Session based Recommendation CIKM2021

[10] An Attribute-Driven Mirroring Graph Network for Session-based Recommendation SIGIR2022

[11] Self-supervised Hypergraph Convolutional Networks for Session-based Recommendation AAAI2021

[12] Price DOES Matter! Modeling Price and Interest Preferences in Session-based Recommendation SIGIR2022

[13] Session-based hypergraph attention network for recommendation SDM2021

[14] Learning Multi-granularity Consecutive User Intent Unit for Session-based Recommendation WSDM2022

[15] A survey on Session-based Recommender Systems ACM Comput. Surv.2022