

# 基于图神经网络和注意力机制提取兴趣的会话推荐算法

金 瑾

中国地质大学(武汉)数学与物理学院, 湖北 武汉

收稿日期: 2025年3月16日; 录用日期: 2025年4月9日; 发布日期: 2025年4月17日

## 摘 要

随着用户的隐私意识逐渐增强, 平台上存在着大量匿名用户的交互数据, 如何为匿名用户进行个性化推荐成为推荐系统备受关注的话题, 基于会话的推荐应运而生。会话推荐指的是基于用户当前会话中的交互历史进行推荐, 而不依赖用户长期历史行为数据或用户个人属性信息。本文针对现有会话推荐模型难以捕捉用户动态兴趣的问题, 提出了一种基于提示词的动态兴趣提取会话推荐模型(Dual-channel Prompt-based Interest Extraction, DPIE)。该模型使用图神经网络作为训练阶段, 引入兴趣提示词, 通过设计滑动注意力窗口机制, 对会话中的项目进行细粒度的表示学习, 自适应地捕捉用户兴趣的变化, 结合提示词增强项目的语义表示, 得到动态兴趣表示。同时, 引入位置编码, 结合长期兴趣提示词学习用户的长期兴趣表示。这种设计使得模型能够更好地建模用户兴趣的动态变化过程, 解耦动态兴趣和长期兴趣, 从而提供更精准的推荐, 在多个真实数据集上的实验结果验证了该方法的有效性。

## 关键词

会话推荐, 图神经网络, 提示学习, 注意力机制

# A Session-Based Recommendation Algorithm for Extracting Interest Based on Graph Neural Network and Attention Mechanism

Jin Jin

School of Mathematics and Physics, China University of Geosciences (Wuhan), Wuhan Hubei

Received: Mar. 16<sup>th</sup>, 2025; accepted: Apr. 9<sup>th</sup>, 2025; published: Apr. 17<sup>th</sup>, 2025

## Abstract

With heightened user privacy awareness, anonymous user interaction data is prevalent on platforms. Personalizing recommendations for anonymous users has become a focal point in recommender systems, leading to the rise of session-based recommendation. Session-based recommendation focuses on leveraging interaction history within the current session, rather than long-term user history or profile attributes. Addressing the challenge of existing session-based models in capturing dynamic user interests, this paper proposes a Dual-channel Prompt-based Interest Extraction (DPIE) model. DPIE employs graph neural networks in the training phase and introduces interest prompts. By designing a sliding attention window mechanism, it enables fine-grained representation learning of items within a session, adaptively capturing shifts in user interests, and enhances item semantic representations with prompts to derive dynamic interest representations. Furthermore, position encoding is incorporated, combined with long-term interest prompts to learn long-term user interest representations. This design allows the model to better model the dynamic evolution of user interests, decoupling dynamic and long-term interests for more precise recommendations. Experimental results on multiple real-world datasets validate the effectiveness of the proposed approach.

## Keywords

Session-Based Recommendation, Graph Neural Network, Prompt Learning, Attention Mechanism

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在数字化时代的今天，随着互联网和电子商务的快速发展，用户行为数据量也呈现爆炸式增长，推荐系统通过分析用户的历史行为和属性信息等，为用户提供个性化和精准的推荐服务[1]。然而，随着隐私保护意识的增强，越来越多的用户选择以游客(匿名)方式登录，这使得推荐系统无法访问用户的历史数据或个性化的属性信息，而只能依赖当前会话中的项目数据进行推荐。这种数据量少和匿名特性为推荐系统的设计和性能提升带来了新的挑战。会话推荐在这种背景下出现，通过分析用户在单个会话中较短的行为序列，预测用户可能感兴趣的项目。同时，与传统推荐系统相比，会话推荐面临多重挑战。首先，会话中的项目数据通常较少，会话的平均长度也较短。其次，用户需求通常是动态且模糊的，这对系统捕捉用户意图的能力提出了更高要求。此外，用户行为的多样性和复杂性加剧了建模难度，使得会话推荐模型需要在精准性和泛化性之间取得平衡[2]。如何在数据稀缺和兴趣动态变化的条件下实现高效的推荐仍是亟待解决的核心问题。

近年来，基于自然语言处理(Natural Language Processing, NLP)的提示学习(Prompt Learning)为解决类似问题提供了一种新的思路[3]。提示学习是一种通过设计合适的提示词(Prompt)，将下游任务转化为与预训练任务相似的技术，从而充分利用预训练模型的能力，在下游任务阶段只进行微调即可。这种方法不仅能有效降低任务设计的复杂度，还能在少样本或零样本学习场景中展现优异效果[4]。

最近，研究学者将提示学习从 NLP 领域的广泛应用扩展到图数据领域，图提示学习(Graph Prompt Learning)开始展现出强大的潜力。图提示学习通过为图数据(如节点、边、子图等)引入提示词，灵活适配

不同的任务需求[5]。这种方法不仅保持了图神经网络强大的表达能力，还赋予了模型更强的灵活性和泛化能力。

在会话推荐任务中，用户在不同时间步下所反映的兴趣会动态改变。基于图神经网络(Graph Neural Network, GNN)的方法虽然能够有效建模会话内的交互关系，但在动态兴趣提取方面存在两个主要局限性：首先，GNN 在聚合节点信息时采用统一的特征转换模式，无法根据不同时间步下的特点动态调整聚合策略。例如，当用户在浏览手机时突然转向查看配件，GNN 难以及时调整节点的重要性权重来反映这种兴趣转移。其次，GNN 将所有节点的嵌入表示视为固定特征，缺乏对时序信息的显式建模。这导致模型难以区分用户的长期稳定兴趣(如对某品牌的偏好)和短期动态兴趣(如临时的促销活动响应)。相比之下，提示学习通过引入可训练的提示词，能够根据会话数据动态调整特征转换方式，同时通过任务重构将时序信息融入到模型中，为解决上述问题提供了新的思路。

本文旨在将提示学习方法引入会话推荐领域，探索其在解决数据稀缺和动态兴趣建模挑战中的潜力。具体而言，本文提出了一种结合提示的会话推荐模型，通过用户在会话的历史数据中引入可优化的提示词，分别建模用户的动态兴趣与长期兴趣，增强其对用户兴趣的提取能力，进而提升推荐效果。本文的主要贡献如下：

1. 提出了 DPIE，一种基于提示学习的动态兴趣提取会话推荐模型(Dual-channel Prompt-based Interest Extraction for SBR, DPIE)，通过引入动态兴趣提示词和长期兴趣提示词，实现了对用户短期和长期兴趣的解耦建模，有效提升了推荐性能。

2. 设计了一种新颖的自注意力机制，能够动态评估会话中不同时间步数据的重要性，从而更准确地捕捉用户兴趣的实时变化。

3. 在多个真实数据集上进行了广泛的实验验证，实验结果表明 DPIE 的有效性。

## 2. 符号定义

为保持全文的一致性，本节首先详细介绍模型所使用的符号，具体符号说明见表 1。

**Table 1.** Symbol notation table

**表 1.** 符号说明表

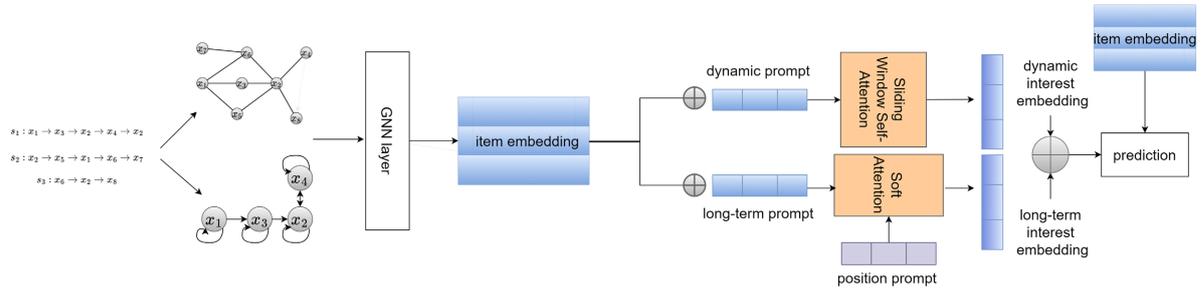
符号	定义
$I,  I /n$	会话中出现的项目的集合，所有项目的总数
$x_i$	第 $i$ 个被点击的项目
$s = [x_1, x_2, \dots, x_m]$	一个会话包含 $m$ 个项目
$s_i$	第 $i$ 个会话
$\mathcal{S}$	给定数据集中所有会话的集合
$e_i$	项目 $i$ 的嵌入向量

## 3. 基于提示词的动态兴趣提取会话推荐模型整体架构

用户在网站上的点击行为反映了其兴趣的动态变化，仅依赖长期行为建模而忽视这种动态性，难以精准捕捉用户兴趣的实时波动，从而影响推荐效果。为了解决这一问题，本文提出了一种“基于提示词的动态兴趣提取会话推荐模型”。该模型通过引入提示学习机制，设计动态兴趣提示词和长期兴趣提示词，并结合本文提出的自注意力机制，动态评估会话中不同时间步数据的重要性，从而融合用户的动态兴趣与长期兴趣，实现更加精准的推荐。

如图 1 所示,本文提出的基于提示词的动态兴趣提取会话推荐模型(Dual-channel Prompt-based Interest Extraction for SBR, DPIE)主要包含三个关键阶段:预训练阶段、提示阶段和预测阶段。

1. 预训练阶段:构建局部图和全局图,通过图神经网络对项目特性进行学习,从而生成项目的表示。
2. 提示阶段:预训练得到的项目特征被用于动态兴趣和长期兴趣建模。在动态兴趣通道中,引入动态兴趣提示词,通过拼接提示词与项目特征,并结合自注意力机制,提取用户的短期兴趣变化;在长期兴趣通道中,引入兴趣原型提示词,通过位置编码与软注意力融合机制,提取用户的长期兴趣特征。



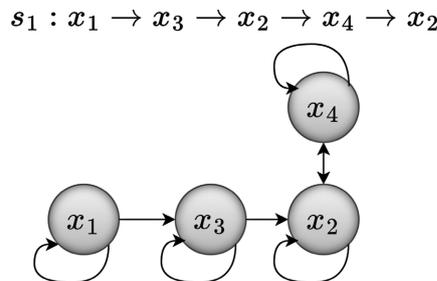
**Figure 1.** Dual-channel prompt-based interest extraction for SBR framework  
**图 1.** 基于提示词的动态兴趣提取会话推荐模型框架图

3. 预测阶段:结合动态兴趣表示与长期兴趣表示生成最终的用户兴趣表示,用于推荐下一个项目的预测。

### 3.1. 项目表示学习层

由于项目表示的质量直接影响模型的推荐效果,然而会话数据单个会话内的信息量有限,因此需要构建会话图来捕捉项目间的转换信息。本节首先将会话表示为局部会话图,作为 GNN 的输入来学习当前会话中项目的特征。此外,其他会话也蕴含着重要的信息,通过捕捉这些会话中项目与当前会话的联系,可以进一步丰富项目表示,因此,本文还构建了全局会话图来整合所有会话的信息。

对于给定的会话  $s = [x_1, x_2, \dots, x_m]$ , 设  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  为通过该会话构建的图,其中  $\mathcal{V}$  表示该会话中的项目集合,每个项目被定义为图中的一个节点,  $\mathcal{E}$  表示边的集合,会话中两个相邻节点为其连上一条边,同时,需要注意的是,为了区分项目的点击顺序,在构图时,本模型采用了有向边,会话序列  $s = [x_1, x_2, x_3, x_4]$  构建的会话图如图 2 所示。



**Figure 2.** Generate local graph  
**图 2.** 局部图构建

构建完局部会话图后,本文采用图注意力网络对项目节点进行特征学习。通过多层卷积,节点的特

征向量不仅整合了自身的信息，还聚合了邻居节点的信息。具体而言，在第  $l$  层的特征更新公式有如下步骤：

$$e_{ij} = \text{LeakyReLU}\left(a_{ij}^\top \left(v_i^{l-1} \odot v_j^{l-1}\right)\right)$$

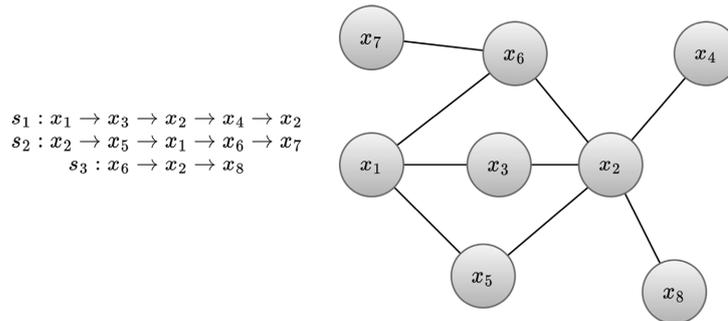
其中， $v_i$  表示节点  $x_i$  的初始嵌入， $a_{ij} \in R^d$  表示与边类型  $r_{ij}$  相关的权重向量，激活函数为 **LeakyReLU**。 $e_{ij}$  表示节点  $x_i$  和节点  $x_j$  之间的重要性得分。为区分不同的边类型的重要程度，模型训练四个权重向量，即  $a_{in}, a_{out}, a_{self}, a_{in-out}$ 。接下来，计算节点间的注意力权重  $\alpha_{ij}$ ，为了使不同节点之间的系数具有可比性，使用 **softmax** 函数进行归一化：

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{x_k \in N_i} \exp(e_{ik})}$$

其中  $N_i$  是节点  $x_i$  的一阶邻居。最后，局部会话图中的项目表示加权聚合了它的邻居特征：

$$e_i^s = \sum_{v_j \in N_i} \alpha_{ij} v_j$$

综上所述，局部会话图主要聚焦于当前会话内部的项目转换关系。然而，为了捕捉会话的全局上下文信息，本文进一步构建了一种全局图，具体而言，全局图是一个无向加权图，节点之间的边表示项目间的共现关系，对于任意两个项目  $x_i, x_j$ ，如果它们在会话中同时被点击，则在它们之间添加一条无向边，边的权重由这两个商品在所有会话中共同出现的频率来确定。全局图的构建如图 3 所示：



**Figure 3.** Generate global graph  
**图 3.** 全局图构建

构建完全局图后，接下来需要对全局图进行有效的表示学习。与局部会话图中的表示学习相似，首先输入项目的初始嵌入表示  $v_i$ ，聚合节点的一阶邻居表示：

$$e_i^g = \sum_{v_j \in N_{x_i}} \pi(x_i, x_j) v_j^{(l-1)}$$

其中， $\pi(x_i, x_j)$  衡量邻居节点  $x_j$  对于项目  $x_i$  的重要性，称为会话感知的注意力权重：

$$\pi(x_i, x_j) = q_i^\top \text{LeakyReLU}\left(W_1 \left[ s \odot v_i^{(l-1)} \parallel w_{ij} \right]\right)$$

$$\pi(x_i, x_j) = \frac{\exp(\pi(x_i, x_j))}{\sum_{v_k \in N_{x_i}} \exp(\pi(x_i, x_k))}$$

$\odot$  表示元素积， $\parallel$  表示拼接操作， $w_{ij} \in R^1$  是全局图中边的权重， $W_1 \in R^{d+1 \times d+1}$  和  $q_1 \in R^{d+1}$  是可学习的参数

矩阵,  $s$  是当前会话的特征表示, 由当前会话中所有的项目嵌入的平均值得到  $\left( s = \frac{1}{|S|} \sum_{x_i \in S} v_i \right)$ 。最后, 将全局项目嵌入  $e_i^g$  进行 Dropout 操作后与会话级项目嵌入  $e_i^s$  相加, 得到融合后的项目嵌入:

$$e_i = \text{dropout}(e_i^g) + e_i^s$$

### 3.2. 兴趣提取层

会话推荐中, 用户兴趣的动态变化是影响推荐效果的关键因素之一。用户的兴趣并非一成不变, 而是在每次交互中不断变化。传统的会话推荐通常依赖于用户的行为历史来进行兴趣建模, 然而, 这种方法忽视了用户在短时间内的兴趣变化。如何有效捕捉并建模这种变化仍是一个亟待解决的问题。

在动态兴趣通道中, 本文将项目的表示与一个可训练的动态兴趣提示词拼接在一起。该提示词旨在通过学习调整, 帮助模型更好地捕捉当前会话中的兴趣变化。具体而言, 可以通过以下公式来表示:

$$e_i^d = [e_i \parallel p_d]$$

其中,  $e_i$  表示项目  $x_i$  的嵌入表示,  $p_d$  为动态兴趣提示词,  $[\cdot \parallel \cdot]$  表示向量拼接操作。另一方面, 在长期兴趣通道中, 引入了长期兴趣提示词, 该向量代表用户的长期兴趣偏好, 通常不随会话的改变而快速波动。将项目嵌入与原始兴趣提示词拼接, 模型能够学习到用户的长期兴趣表示, 公式如下:

$$e_i^l = [e_i \parallel p_l]$$

其中,  $p_l$  表示长期兴趣提示词。通过在这两个通道中分别引入动态兴趣和长期兴趣提示词, 模型能够有效地解耦用户的动态兴趣和长期兴趣。然而, 单纯地引入提示词并不能直接实现对用户兴趣的提取。为了进一步提取动态兴趣和长期兴趣, DPIE 需要进行特定的聚合操作, 接下来介绍如何提取动态兴趣。

为了更好地提取用户的动态兴趣, 模型将动态兴趣表示输入到本文提出的滑动窗口注意力机制中。用户在不同时间步长中的点击行为能够反映出兴趣的变化, 结合不同时间步的兴趣可以实现更加精细的动态兴趣建模。传统的多头注意力机制对会话中所有商品计算注意力得分, 这种方式未能充分考虑用户兴趣的时变性。相比之下, 滑动窗口注意力机制为每个注意力头分配不同的感受野, 使不同注意力头关注会话中的不同时间窗口, 从而实现对兴趣的动态建模。具体而言, 首先, 对数据集中的会话长度进行统计, 生成长度的频率直方图和累积分布曲线, 如图 4 所示, 数据具有明显的长尾分布, 大量的会话长度集中在较短的范围内。

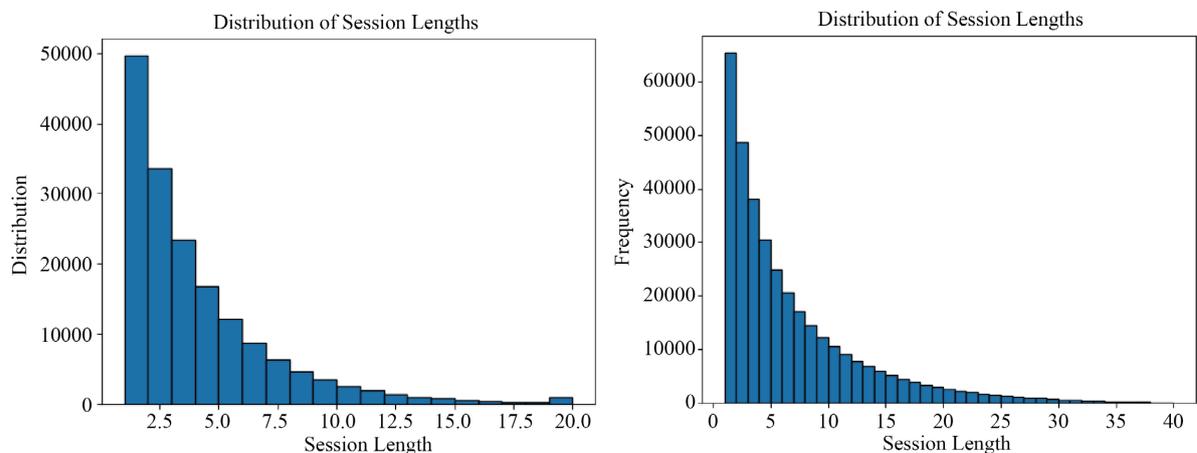
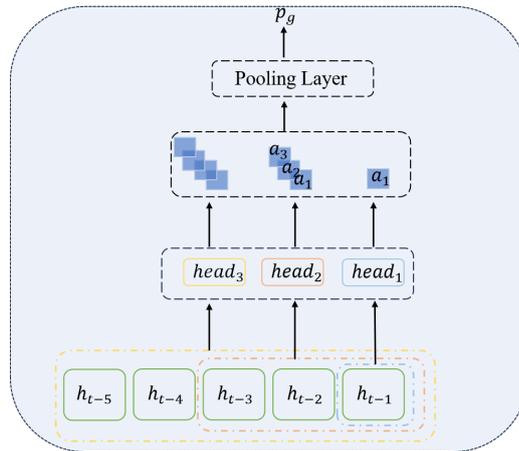


Figure 4. Session length distribution histograms for different datasets  
图 4. 不同数据集会话长度分布直方图

设定一个覆盖率阈值  $\epsilon$  (如 90%)，找到累计会话次数占总会话数  $\epsilon$  的长度临界值  $L_\epsilon$ 。因为会话数据的一个显著特点是，大部分会话较短，仅包含少量项目交互，只有少部分会话较长，针对这一特点，模型通过设置参数  $\epsilon$  选取覆盖大多数会话的长度范围，将注意力集中在短会话的动态变化上。避免关注长会话带来的噪声干扰。



**Figure 5.** Sliding window attention mechanism  
**图 5.** 滑动窗口注意力机制

接下来，如图 5 所示，针对每个滑动窗口，计算对应的注意力权重，并在窗口范围内完成加权求和，从而更新项目的表示。公式如下：

$$\beta_{ij}^k = \frac{\exp(q_i^{\top k_j} / \sqrt{d})}{\sum_{v \in W_k} \exp(q_i^{\top k_v} / \sqrt{d})}$$

其中， $W_k$  为滑动窗口的范围，表示在当前头下计算注意力时所考虑的邻居节点集合。 $q_i$  和  $k_j$  分别表示 query 向量和 key 向量，在滑动窗口注意力中，会有多个不同的头  $h=1,2,\dots,H$  (其中  $H$  是头的数量)，每个头都有不同的 query 和 key 向量，因此每个头会从不同的子空间学习信息。具体来说，对于第  $h$  个头，query 和 key 向量的计算为：

$$q_i^h = e_i^d W_q^h \in R^{d_k}$$

$$k_j^h = e_j^d W_k^h \in R^{d_k}$$

其中， $W_q^h \in R^{d \times d_k}$  和  $W_k^h \in R^{d \times d_k}$  是第  $h$  个头对应的参数矩阵， $d$  是项目嵌入的维度， $d_k = d/h$  是参数向量的维度。

在计算了注意力得分后，接下来会根据这些得分对项目向量进行加权聚合，更新项目的表示。具体而言，更新后的项目向量表示为：

$$h_i^k = \sum_{j \in W_k} \beta_{ij}^k e_j^d$$

其中， $e_j^d$  是项目  $x_j$  的嵌入向量， $\alpha_{ij}^k$  是项目  $x_j$  对  $x_i$  的注意力得分，最后，将所有头的输出拼接，使用全连接层对输出结果进行映射，得到最终的项目表示：

$$h_i = [h_i^1, h_i^2, \dots, h_i^H]$$

$$h_i^d = W_1(h_i + e_i^d) + b_1$$

其中,  $W_1$  和  $b_1$  为可训练的参数,  $h_i$  为经过包含了动态兴趣变化信息的项目嵌入。

接下来, 模型将学习用户的长期兴趣偏好, 通过位置编码使得模型能够区分会话中不同位置上项目的重要性, 结合长期兴趣提示词帮助模型捕捉用户的长期偏好。模型引入一个可学习的位置嵌入矩阵  $P = [p_1, p_2, \dots, p_m]$ , 其中每个位置向量  $p_i$  表示会话中第  $i$  个位置的编码,  $m$  是当前会话序列的长度。通过拼接原始兴趣提示词  $p_i$  和项目嵌入  $e_i$ , 再结合项目在会话中的位置信息进行非线性变换, 得到最终的长期兴趣表示。具体而言, 长期兴趣表示的计算过程如下:

$$h_i^l = \tanh(W_2[e_i \parallel p_{l-i+1}] + b_2)$$

其中,  $e_i^l$  是项目  $x_i$  的嵌入向量,  $p_{l-i+1}$  是项目在会话中的逆序位置嵌入向量,  $W_2$  和  $b_2$  是可学习的权重和偏置项。通过非线性激活函数  $\tanh$  进行映射, 最终得到长期兴趣表示  $h_i^l$ 。

提取了长期兴趣和动态兴趣表示后, 模型需要融合这两种兴趣, 以生成最终的会话表示。具体而言, 首先, 模型通过对会话中所有项目的表示进行加权求和来获得会话的信息表示  $\tilde{s} = \sum_{i=1}^m e_i$ 。对于每个项目  $x_i$ , 根据其长期兴趣表示  $h_i^l$  和动态兴趣表示  $h_i^d$ , 计算其在会话中的长期兴趣权重  $\gamma_i^l$  和动态兴趣权重  $\gamma_i^d$ :

$$\gamma_i^l = q^\top \sigma(W_3 e_i^l + W_4 \tilde{s} + b_3)$$

$$\gamma_i^d = q^\top \sigma(W_5 e_i^d + W_6 \tilde{s} + b_4)$$

其中,  $W_3, W_4 \in R^{d \times d}$ ,  $q, b_3 \in R^d$  为可训练的参数,  $\tilde{s}$  是会话中所有项目嵌入的平均表示,  $\sigma(\cdot)$  表示激活函数 sigmoid。

接下来, 将动态兴趣和长期兴趣的权重取平均, 得到最终的综合权重  $\gamma_i$ , 并使用这一权重对会话中所有项目的表示进行加权聚合:  $\gamma_i = \frac{1}{2}(\gamma_i^d + \gamma_i^l)$ , 最终, 所有项目的加权表示通过如下方式得到会话表示  $s$ :

$$s = \sum_{i=1}^m \gamma_i e_i$$

其中,  $e_i$  是项目  $x_i$  的更新表示,  $\gamma_i$  是根据长期兴趣和动态兴趣计算得到的加权系数。通过这种方式, 模型能够综合考虑长期兴趣和动态兴趣的权重, 并结合会话中各项目的顺序信息, 有效地生成当前会话的表示。

### 3.3. 预测和训练

预测过程基于生成的会话表示  $s$ , 结合候选项目的初始嵌入向量, 为每个候选项目计算其成为下一个点击项目的概率:

$$\hat{y}_i = \text{Softmax}(s^\top v_i)$$

其中,  $\hat{y}_i$  表示项目  $x_i$  为当前会话中下一个点击项目的预测概率,  $v_i$  是项目  $x_i$  的初始嵌入表示,  $\text{Softmax}$  函数确保了所有候选项目的预测概率之和为 1。

为了优化模型参数, 采用交叉熵损失函数对预测结果进行监督学习。损失函数定义如下:

$$\mathcal{L}(\hat{y}) = -\sum_{i=1}^m [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

其中,  $y_i$  是真实标签, 表示项目  $x_i$  是否为下一个点击项目(若为下一个点击项目则  $y_i = 1$ , 否则为  $y_i = 0$ ),

$\hat{y}_i$  是项目  $x_i$  的预测概率。通过最小化损失函数，模型能够学习到会话中项目与用户行为之间的相关性，从而提高推荐准确率。

## 4. 实验设置

本节主要介绍了 DPIE 模型的实验设置。

### 4.1. 数据集介绍

为了评估模型的有效性，本文在两个真实世界的数据集上进行了实验，如表 2 所示，分别是 Cosmetics 数据集和 Diginetica 数据集。这两个数据集被广泛用于评估基于会话的推荐算法的性能，并在领域、规模和稀疏性方面各不相同。Cosmetics 数据集来源于 Kaggle 比赛，记录了用户在化妆品电商平台上的行为数据，涵盖了用户的点击、浏览和购买等行为，同时包含了商品类别与价格信息，具有典型的长尾分布，数据集相对较小。Diginetica 数据集出自 2016 年的 CIKM Cup 竞赛的公开数据集，主要涉及电子商务领域，覆盖了多个商品类别，数据规模较大，数据集包含了用户在网站上的会话序列，每个会话代表了用户在一段时间内的一系列行为，如浏览、点击、购买等，本文仅使用其中的交易数据进行模型的训练和预测。

Table 2. Dataset statistics data

表 2. 数据集的统计数据

数据集	Diginetica	Cosmetics
候选项目数量	24,889	23,194
交互数	855,070	1,058,263
训练集	339,786	282,453
测试集	35,294	31,392
平均会话长度	4.56	6.74

### 4.2. 参数设置

本文中所有模型的嵌入维度均设置为 100，小批量大小固定为 100。对于 DPIE，使用 4 个注意力头，针对不同数据集设置在卷积层 {1, 2} 中寻找最优性能。所有参数均使用均值为 0，标准差为 0.1 的高斯分布进行初始化。模型优化采用 Adam 优化器，初始学习率为 0.001，每 3 个 epoch 后将衰减 0.1。通过调整超参数 epsilon 在 {0.3, 0.5, 0.7, 0.9} 中以及图卷积丢失率在 {0.02, 0.04, 0.06, 0.08, 0.1} 中寻找最优性能，此外，在全局图中，为加快训练速度和防止噪声，将邻居数量和相邻项目的距离分别设置为 12 和 3。

### 4.3. 基线模型

在本文中，为了验证提出的 CluHNN 模型的有效性，选取了会话推荐领域中的几种具有代表性的模型进行对比，选取的基线模型如下所述：

(1) POP 直接根据全局商品流行度进行推荐，认为在会话中，用户更倾向于选择热门商品。模型统计所有会话中商品的出现频率，并将其作为流行度排序依据，为所有用户推荐全局最流行的商品列表。

(2) GRU4REC [6] 首次将循环神经网络应用于会话推荐任务，认为用户的会话行为具有时序依赖性，可以通过 RNN 来捕捉这种序列模式。

(3) NARM [7] 是一种首次将注意力机制应用于会话推荐的模型，利用带有注意力机制的 RNN 来捕捉用户的主要意图。

(4) SR-GNN [8]将会话序列建模为图形结构化数据。

(5) SHARE [9]将超图注意力网络引入到会话推荐模型中，使用超图注意力网络在每个单独会话中学习上下文窗口内的信息。

(6) DHCN [10]提出了一种结合自监督学习的超图卷积网络，利用超图卷积网络来有效地建模会话数据中的项目转移关系。

(7) COTREC [11]提出了一种自监督会话模型，利用图神经网络和对比学习来缓解数据稀疏问题。

(8) SCL [12]模型提出了一种新的对比学习方法——自对比学习，此提出一种新的损失目标，直接将每个项目表示与所有其他项目表示区分开，从而促进项目表征空间内的均匀分布。

## 5. 实验结果与分析

### 5.1. 对比实验

为了验证 DPIE 模型的有效性，本节将其与多个基线模型在两个数据集上进行了对比实验。实验结果如表 3 所示，其中加粗值和下划线分别表示最佳和次优结果。

**Table 3.** Performance comparison of the model and baseline methods

**表 3.** 模型与基线方法的性能比较

model	cosmetics				Diginetica			
	P@10	M@10	P@20	M@20	P@10	M@10	P@20	M@20
POP	32.83	26.63	38.43	27.32	25.11	18.32	25.91	19.84
GRU4REC	19.41	30.4	47.63	30.80	45.68	20.24	55.76	21.1
NARM	42.59	35.05	46.09	35.30	40.68	20.87	50.78	21.57
SR-GNN	43.04	35.37	46.50	35.61	45.1	23.75	55.61	24.48
SHARE	45.76	36.37	49.80	36.65	46.55	23.94	58.08	24.74
DHCN	34.24	30.86	34.79	30.79	24.86	18.95	26.63	19.07
COTREC	39.28	34.65	41.30	34.80	44.85	23.81	53.68	24.57
SCL	<u>46.32</u>	<u>36.5</u>	<u>50.38</u>	<u>36.80</u>	<u>48.57</u>	<u>25</u>	<u>60.61</u>	<u>25.86</u>
DPIE	<b>46.35</b>	<b>36.52</b>	<b>50.57</b>	<b>36.8</b>	<b>48.95</b>	<b>25.13</b>	<b>60.89</b>	<b>25.94</b>

根据实验结果，可以发现：

(1) 基于循环神经网络的会话推荐模型，如 NARM、GRU4REC 在推荐效果上远远超过传统方法 POP，然而 GRU4REC 只关注内部序列信息，忽略了其他潜在关系，NARM 引入注意力机制后更能精准挖掘用户意图。

(2) 基于 GNN 的模型(如 SR-GNN、SHARE)普遍优于传统的 RNN 模型(NARM)，这说明图神经网络在捕获项目间复杂关系方面具有较大优势。此外 RNN 仅利用会话内的顺序信息，而会话数据中，过分强调顺序信息一定程度上限制了模型建模用户兴趣的能力，会话的拓扑结构则有效弥补了这一点。

(3) 对比学习能够有效提升模型的性能，DHCN、COTREC、SCL 模型均采用对比学习，其中 DHCN、COTREC 分别建立不同的项目视图，在两个通道中分别学习项目的表示，而 SCL 将复杂的对比目标简化，强调项目表示均匀性，其实验结果优于 DHCN、COTREC。

(4) DPIE 模型在两个数据集的所有评估指标上均取得了最佳性能，这证明了提示学习结合动态兴趣

提取的有效性。

## 5.2. 消融实验

为了验证 DPIE 模型中各个组件的有效性，本节进行了详细的消融实验，通过逐一去除 DPIE 中的不同模块并分析性能变化，实验结果见表 4。具体而言，设计了以下三组对照实验：

**Table 4.** Ablation experiment comparison

**表 4.** 消融实验对比

dataset	model	top10		top20	
		Prec	MRR	Prec	MRR
cosmetics	only attention	46.14	36.46	50.52	36.77
	prompt & pos	46.13	36.53	50.41	36.8
	promp & attention	41.94	34.86	44.55	35.04
	DPIE	46.35	36.52	50.57	36.8
digineticaBuy	only attention	48.49	25	60.74	25.83
	prompt & pos	48.59	25.09	60.61	25.92
	promp & attention	46.81	24.56	58.09	25.35
	DPIE	48.95	25.13	60.89	25.94

(1) 仅使用注意力机制的基础模型(w/o Prompt): 该模型移除了提示学习模块，经过图卷积学习项目特征后，分别通过滑动窗口注意力机制和位置编码聚合项目特征，得到动态兴趣表示和长期兴趣表示。

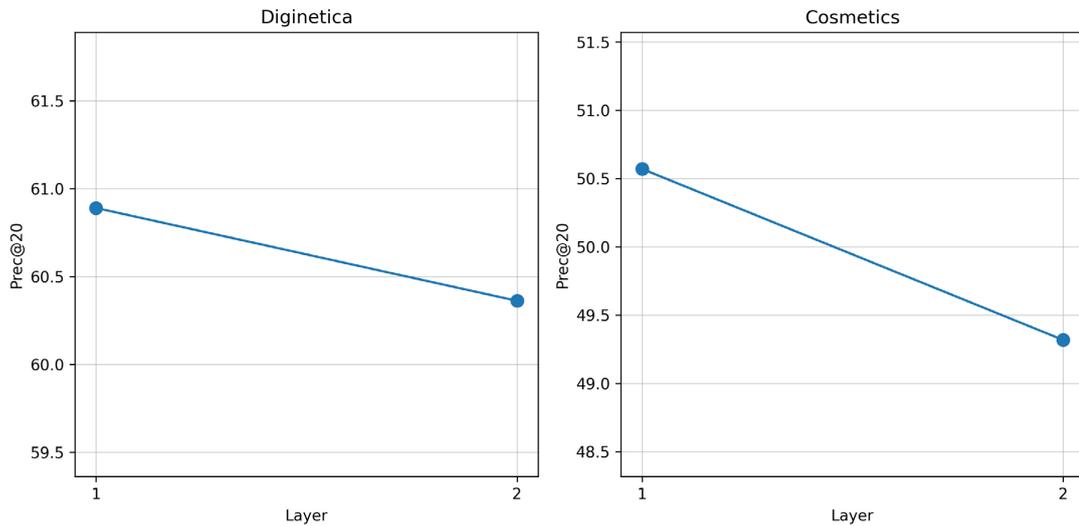
(2) 结合提示学习和位置编码的模型(w/o Attention): 该模型移除了滑动窗口注意力机制，保留提示学习模块，通过位置编码结合软注意力聚合项目表示。

(3) 结合提示学习和注意力机制的模型(w/o Position): 该模型移除了位置编码，保留提示学习模块，通过滑动窗口注意力机制聚合项目表示。

通过这三组消融实验，实验结果表明：无论是去除提示学习模块、滑动窗口注意力机制，还是位置编码，都显著影响了模型的预测性能。这表明 DPIE 中每个模块对最终结果的贡献不可或缺。单纯引入滑动窗口注意力模块(在实验 1 中)时，尽管模型在捕捉动态兴趣方面有一定的能力，但仅依赖项目特征的聚合无法充分挖掘用户兴趣的动态性和长期性。当提示学习结合滑动窗口注意力机制(在实验 2 中)或位置编码(在实验 3 中)被引入时，模型性能显著提升。这表明，加入提示学习模块外，能帮助捕捉动态兴趣或长期兴趣，但如果缺乏信息聚合机制，模型无法有效利用提示词进行微调。总体而言，在三者的协同作用下，DPIE 能够更好地捕捉用户的兴趣演化，表现出优异的性能。

## 5.3. 超参数分析

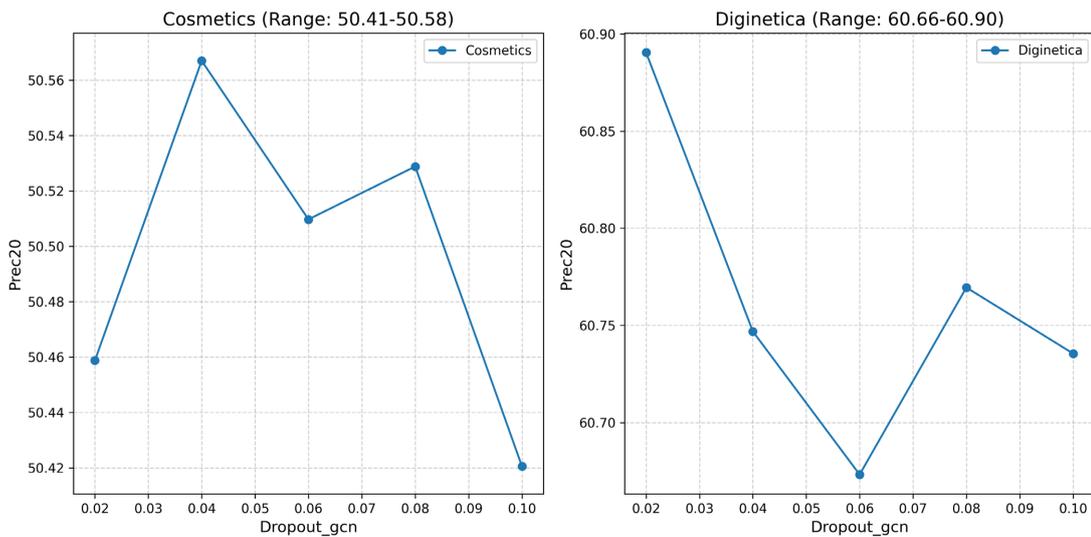
由于图卷积层数决定了模型对项目间关系建模的深度，但其设置需要权衡：层数过少可能导致模型无法充分捕捉项目间的高阶关联，而层数过多则可能引入噪声，导致过拟合或特征过度平滑。本节对图卷积层数进行分析，实验结果如图 6，图 7 所示，实验结果表明，当图卷积层数为 1 时，模型在 Cosmetics 和 Diginetica 数据集上均取得了最佳性能。



**Figure 6.** Performance impact of the number of graph convolutional layers

**图 6.** 图卷积层数影响性能图

此外, 本节对图卷积层的丢失率(dropout)进行了分析。丢失率是一个重要的正则化参数, 它通过在训练过程中随机“丢弃”一部分神经元, 来防止模型过拟合。在图神经网络中, 合适的丢失率设置尤为重要, 因为图结构数据往往存在较强的相关性, 容易导致模型过度依赖某些特定的连接模式。如果丢失率过高, 可能会导致重要的结构信息丢失; 如果丢失率过低, 则可能无法有效防止过拟合。实验结果见下图 7:



**Figure 7.** Performance impact of dropout rate parameter

**图 7.** 丢失率参数影响性能图

通过实验发现, 对于 Cosmetics 数据集, 当丢失率设置为 0.04 时, 模型取得较好的性能, Diginetica 数据集上的该值需设置在 0.02。这可能是由于 Cosmetics 数据集和 Diginetica 数据集的用户行为分布更加集中, 具有更高的相关性和较强的结构化特性。

## 6. 小结

本文提出了一种基于提示学习的动态兴趣提取会话推荐模型。通过引入动态兴趣提示词和长期兴趣提示词,结合滑动窗口注意力机制,实现了对用户短期兴趣变化和长期兴趣偏好的有效建模。在多个真实数据集上的实验结果表明,该模型在各项评估指标上均优于现有方法,验证了提示学习在提升会话推荐性能方面的有效性。未来的研究工作可以从以下几个方面展开:首先,可以探索更细粒度的兴趣分解方法,考虑用户的多维度兴趣特征;同时,在处理这些细粒度兴趣的过程中,也应关注如何应对会话中由于用户不经意点击而产生的噪声数据。此外,用户的兴趣通常存在多种复杂的相互关系,未来可以研究如何有效提取和融合不同兴趣之间的关系,以提升模型的推荐效果。

## 致谢

感谢审稿人对手稿的认真阅读以及有益的评价。

## 参考文献

- [1] 王喆. 深度学习推荐系统[M]. 北京: 电子工业出版社, 2020: 1-9.
- [2] Li, Z., Yang, C., Chen, Y., Wang, X., Chen, H., Xu, G., *et al.* (2024) Graph and Sequential Neural Networks in Session-Based Recommendation: A Survey. *ACM Computing Surveys*, **57**, 1-37. <https://doi.org/10.1145/3696413>
- [3] 程章桃, 黄浩然, 薛荷, 等. 基于提示学习与超图的事件因果关系识别模型[J/OL]. 计算机科学: 1-16. <https://link.cnki.net/urlid/50.1075.TP.20241211.1841.034>, 2025-04-14.
- [4] 崔金满, 李冬梅, 田萱, 等. 提示学习研究综述[J]. 计算机工程与应用, 2024, 60(23): 1-27.
- [5] Zhu, Y., Wang, Y., Shi, H., Zhang, Z., Jiao, D. and Tang, S. (2024) GraphControl: Adding Conditional Control to Universal Graph Pre-Trained Models for Graph Domain Transfer Learning. *Proceedings of the ACM Web Conference 2024*, Singapore, 13-17 May 2024, 539-550. <https://doi.org/10.1145/3589334.3645439>
- [6] Hidasi, B., Karatzoglou, A., Baltrunas, L. and Tikk, D. (2016) Session-Based Recommendations with Recurrent Neural Networks. *4th International Conference on Learning Representations, ICLR 2016*, San Juan, 2-4 May 2016.
- [7] Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T. and Ma, J. (2017). Neural Attentive Session-Based Recommendation. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Singapore Singapore, 6-10 November 2017, 1419-1428. <https://doi.org/10.1145/3132847.3132926>
- [8] Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X. and Tan, T. (2019) Session-Based Recommendation with Graph Neural Networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 346-353. <https://doi.org/10.1609/aaai.v33i01.3301346>
- [9] Wang, J., Ding, K., Zhu, Z. and Caverlee, J. (2021) Session-Based Recommendation with Hypergraph Attention Networks. In: Demeniconi, C., Davidson, I., Akoglu, L. and Terzi, E., Eds., *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, Society for Industrial and Applied Mathematics, 82-90. <https://doi.org/10.1137/1.9781611976700.10>
- [10] Xia, X., Yin, H., Yu, J., Wang, Q., Cui, L. and Zhang, X. (2021) Self-Supervised Hypergraph Convolutional Networks for Session-Based Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 4503-4511. <https://doi.org/10.1609/aaai.v35i5.16578>
- [11] Xia, X., Yin, H., Yu, J., Shao, Y. and Cui, L. (2021) Self-Supervised Graph Co-Training for Session-Based Recommendation. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 1-5 November 2021, 2108-2190. <https://doi.org/10.1145/3459637.3482388>
- [12] Shi, Z., Wang, X. and Lipani, A. (2024) Self Contrastive Learning for Session-Based Recommendation. In: Goharian, N., *et al.*, Eds., *Advances in Information Retrieval*, Springer, 3-20. [https://doi.org/10.1007/978-3-031-56027-9\\_1](https://doi.org/10.1007/978-3-031-56027-9_1)