

# 结合超图学习的多注意力机制新闻推荐方法

孟祥福 王琬淳 张雨萌 樊文懿

辽宁工程技术大学电子与信息工程学院 辽宁 葫芦岛 125105

(mengxiangfu@lntu.edu.cn)

**摘要** 在个性化新闻推荐中,图结构常被用来建立用户与新闻之间的交互关系。然而,普通图结构大多忽略了被点击新闻之间的高阶关联信息。此外,现有方法大多仅使用单一向量学习用户兴趣表示与候选新闻表示,导致建模不充分。针对上述问题,提出了结合超图学习的多注意力机制新闻推荐模型。首先,构建候选新闻超图,通过超图注意力网络的学习捕获候选新闻与其语义相似新闻的高阶相关性,丰富候选新闻语义;然后,构建新闻-主题超图用于建模用户兴趣,采用包含多种注意力机制的神经网络架构挖掘深层的用户细粒度兴趣特征;最后,通过引入激活单元,结合候选新闻特征进一步提取用户兴趣,从而提高推荐准确性。在 MIND-small 和 MIND-large 数据集上进行的大量实验,验证了所提方法的有效性。

**关键词:** 推荐系统;个性化新闻推荐;语义增强;用户兴趣;超图学习;超图注意力网络;注意力机制

**中图分类号** TP391

## Multiple Attention Mechanism News Recommendation Approach with Hypergraph Learning

MENG Xiangfu, WANG Wanchun, ZHANG Yumeng and FAN Wenyi

School of Electronic and Information Engineering, Liaoning Technical University, Huludao, Liaoning 125105, China

**Abstract** In personalized news recommendation, graph structures are often utilized to establish interaction relationships between users and news, however conventional graph structures mostly overlook the high-order association information among clicked news items. Furthermore, existing methods typically rely on a single vector to learn user interest representations and candidate news representations, leading to inadequate modeling. To address these issues, a multiple attention mechanism news recommendation model approach with hypergraph learning is proposed. Firstly, a candidate news hypergraph is constructed, leveraging a hypergraph attention network to capture high-order correlations between candidate news and their semantically similar news, thereby enriching the semantics of candidate news. Secondly, a news-topic hypergraph is built to model user interests, employing a neural network architecture with multiple attention mechanisms to explore deep, fine-grained user interest features. Lastly, an activation unit is introduced to further extract user interests from candidate news, enhancing recommendation accuracy. The experiments on the MIND-small and MIND-large datasets confirm the effectiveness of the proposed approach.

**Keywords** Recommendation system, Personalized news recommendation, Semantic augment, User interest, Hypergraph learning, Hypergraph attention network, Attention mechanism

## 1 引言

在线新闻发布平台上每天都有大量的新闻文章被创建和发布,用户不可能通过浏览所有新闻查找感兴趣的新闻。因此,个性化新闻推荐(Personalized News Recommendation, PNR)应运而生,旨在根据用户特定的兴趣为其推荐候选新闻,这是新闻平台帮助用户减轻信息过载和改善新闻阅读体验的关键技术。

近年来,图神经网络(Graph Neural Network, GNN)在图结构上的学习能力逐渐凸显,其因基于节点特征和图结构数据的强大表征能力受到广泛关注。一些研究尝试构造图结构来挖掘新闻间的交互信息以及深层的用户兴趣特征。例如,Ge等<sup>[1]</sup>将用户与新闻间的交互信息建模为图结构,并设计二跳图学习模块,采用图注意力网络聚合新闻及用户的邻居

嵌入,增强用户的表征能力。Hu等<sup>[2]</sup>在用户-新闻二分图的基础上,采用邻域路由机制对用户的潜在偏好因素解离合,提高了表示的表达性和可解释性。然而,在真实的新闻推荐场景中,被点击新闻之间的关系往往比成对关系更复杂,而普通图结构只能描述两点之间的关系,无法对高阶关系进行建模,导致高阶关联信息编码不足。此外,现有方法大多仅使用单一向量来学习用户兴趣表示与候选新闻表示,导致建模不够充分,忽略了用户对候选新闻的兴趣。

为了解决上述问题,提出了一种结合超图学习的多注意力机制新闻推荐模型(Multi-attention mechanism News Recommendation approach with Hypergraph Learning, MN-RHL)。该模型在候选新闻建模部分,构建候选新闻超图,并通过超图注意力网络的学习捕捉候选新闻与其语义相似新闻之间的高阶相关性,以丰富候选新闻语义。在用户兴趣建模

基金项目:国家自然科学基金(61772249)

This work was supported by the National Natural Science Foundation of China(61772249).

通信作者:王琬淳(13941658067@163.com)

部分,引入新闻的主题信息构建超图,再采用包含多种注意力机制的神经网络架构挖掘用户兴趣,以得到用户兴趣表示。首先,使用超图注意力网络探索新闻与主题之间复杂的多元关系。其次,使用加性注意力机制模拟用户对新闻的关注度,以准确建模用户兴趣。此外,考虑到用户对候选新闻的兴趣,引入激活单元,结合候选新闻特征进一步提取用户兴趣。最后,采用内积函数计算用户-新闻之间的关联分数,进而预测用户可能感兴趣的下一条新闻。

## 2 相关工作

### 2.1 个性化新闻推荐

个性化新闻推荐<sup>[3]</sup>是一种向用户提供满足其个性化阅读兴趣的新闻的重要技术,通常包含3个重要部分<sup>[4]</sup>,其框架如图1所示。首先,利用新闻编码器从新闻内容或其他特征中学习新闻表示。其次,利用用户编码器从用户的历史点击新闻中准确地学习用户表示。最后,根据候选新闻与用户兴趣之间的相关性对候选新闻进行排序。高质量的新闻推荐在很大程度上依赖于对新闻文章和用户兴趣的准确和及时表示。因此,新闻建模和用户建模在个性化新闻推荐中尤为重要<sup>[5]</sup>。

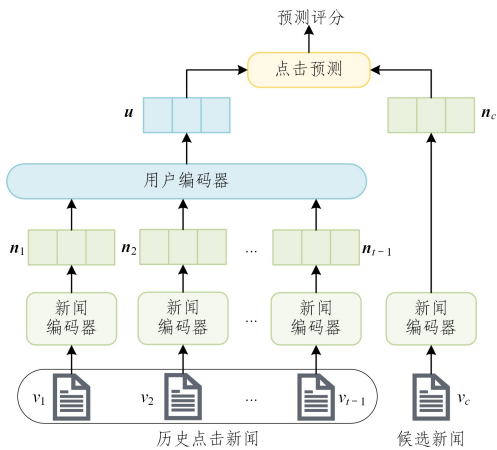


图1 个性化新闻推荐框架

Fig. 1 Framework of personalized news recommendation

新闻建模能够捕捉新闻文章的特征并理解其丰富的文本内容,是个性化新闻推荐方法中的关键步骤。基于深度学习的新闻建模方法旨在从原始输入中自动学习新闻嵌入表示。例如,Moreira等<sup>[6]</sup>采用卷积神经网络(Convolutional Neural Network, CNN)对新闻内容进行卷积,以生成新闻嵌入表示。但CNN网络难以捕捉长距离的文字交互,不适合长序列新闻推荐任务。因此,一些研究采用注意力机制(Attention Mechanism, AM)扩展神经网络,通过选择重要信息来学习新闻表示,提高文本特征提取的准确性。Wu等<sup>[7]</sup>提出基于多头自注意力机制的新闻推荐方法,通过捕捉远距离词之间的相互作用来增强新闻特征的表征能力;此外,还提出利用个性化注意力网络学习新闻标题的语义表示,利用自注意力机制学习新闻标题与正文中的词语语义表示,并采用交互式注意力网络来建模标题和正文间的关系。

用户建模旨在推断用户对新闻文章的偏好,是个性化新闻推荐系统中的关键步骤。例如,Wu等<sup>[7]</sup>采用注意力网络从点击新闻表示中学习用户表示。Zhang等<sup>[8]</sup>利用AM聚合点击新闻和候选新闻的不同信息以建模用户。然而,上述研

究虽然可以自动学习用户表示,但未充分考虑用户与新闻之间的高阶关系。因此,一些研究尝试构造图数据来挖掘更深层的用户兴趣特征,将用户与新闻交互数据作为主要信息源,以捕获用户和新闻的特征及高阶复杂关系,从而进行个性化新闻推荐。例如,Mao等<sup>[9]</sup>设计了新闻图与用户图,并在双图之间执行有效的特征交互,以更精确地学习新闻-用户匹配表示。Hu等<sup>[10]</sup>构造了一个用户-新闻-主题异构图来显式建模用户、新闻和潜在主题之间的交互,合并的潜在主题信息可以有效缓解数据的稀疏性,丰富新闻的语义表示。然而,上述基于图结构的研究方法通常将新闻关系建模为成对连接,而用户历史交互新闻的高阶关系信息并未得到充分挖掘。

### 2.2 超图学习

一般的图是由多个顶点及连接顶点的边组成,边连接了两个顶点,表示两个顶点之间的关系。超图<sup>[11]</sup>是在传统成对关系图的基础上进行扩展,相较于传统图关系更加泛化<sup>[12]</sup>。超图的边可表示多个顶点之间的关系,即一条边中可包含多个顶点。通常情况下,将研究对象赋予成对的关系,并用一般的图结构表示。然而,在许多现实问题中,对象之间的关系比成对关系更复杂。因此,一些研究选择使用超图结构来更好地建模高阶关系,捕捉实体(如新闻内容、新闻主题、用户等)之间更复杂的多元关系。例如,Feng等<sup>[13]</sup>提出超图神经网络(Hypergraph Neural Networks, HGNN),在超图结构中编码高阶数据相关性。Yadati等<sup>[14]</sup>提出了HyperGCN,将图卷积网络用于超图。

最近,一些研究将超图学习引入推荐系统的研究。例如,在下一项推荐(Next-item Recommendation)研究中,Wang等<sup>[15]</sup>使用超图表示短期项目相关性,生成包含项目之间短期相关性的动态项目嵌入,以捕获用户的动态偏好;在基于会话的推荐研究中,Xia等<sup>[16]</sup>将基于会话的数据建模为超图来捕捉项目间复杂的高阶信息,并提出了一种双通道超图卷积网络。Wang等<sup>[17]</sup>为每个会话构建超图,设计了超图注意力层聚合相关项目的上下文信息,以生成项目嵌入;在社交推荐研究中,Yu等<sup>[18]</sup>提出了一种多通道超图卷积网络,该网络利用高阶用户关系增强社交推荐;在新闻推荐研究中,Wang<sup>[19]</sup>利用超图模型对英文新闻文本之间的关系进行建模,并深入研究超图模型在英文新闻文本摘要领域的应用。受上述方法启发,本文将超图引入新闻推荐模型中,更好地捕捉了新闻推荐场景中的复杂交互关系,从而提升了推荐性能。

## 3 问题定义与模型结构

### 3.1 问题定义

在新闻推荐中, $H_u = \{n_1, n_2, \dots, n_{|H|}\}$ 表示用户 $u$ 的历史交互数据集,包含 $H$ 条用户点击新闻;候选新闻列表表示为 $N_c = \{c_1, c_2, \dots, c_{|M|}\}$ ,包含 $|M|$ 个新闻文档,每个新闻 $n$ 的文本内容为一个主题 $t$ 和一个长度为 $T$ 的单词序列组成,表示为 $T_n = \{\omega_1, \omega_2, \dots, \omega_{|T|}\}$ 。此外,由于每条新闻通常具有特定主题,令 $P_u = \{t_1, t_2, \dots, t_{|P|}\}$ 表示历史点击新闻的主题集。

针对用户点击数据,构造一个超图 $G_u = (V, E, W)$ ,其中 $V$ 是包含 $N$ 个节点的集合, $E$ 是包含 $M$ 个超边的集合。每条超边 $e \in E$ 都被分配有一个正权值,所有的权值都被存储在一

个对角矩阵  $\mathbf{W} \in \mathbb{R}^{M \times M}$  中。超图中使用关联矩阵  $\mathbf{H} \in \mathbb{R}^{N \times M}$  描述超图中的拓扑关系,使用关联函数  $h(v, e)$  定义节点  $v$  和超边  $e$  的关系:

$$h(v, e) = \begin{cases} 1, & v \in e \\ 0, & v \notin e \end{cases} \quad (1)$$

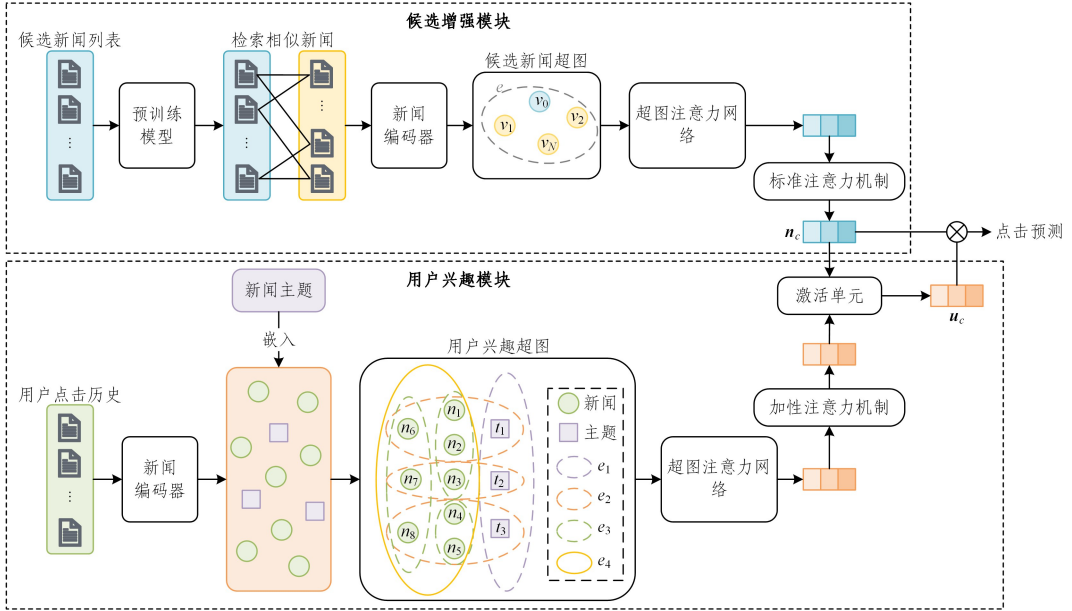


图2 MNRHL模型框架

Fig. 2 Framework of MNRHL model

### 3.2.1 新闻编码器

新闻编码器旨在从新闻文本内容  $T_n$  中提取语义表示  $\mathbf{h}$ 。首先,将新闻单词标记映射到新闻单词嵌入  $\mathbf{E}_n = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{|T|}]$ 。然后,采用多头自注意力网络<sup>[20]</sup>学习新闻的上下文表示  $\mathbf{C}_n \in \mathbb{R}^{|T| \times d}$  ( $d$  是特征维度),并利用注意力网络  $f_{att}(\cdot)$  来聚合新闻语义表示,得到  $\mathbf{h} \in \mathbb{R}^d$ :

$$\mathbf{C}_n = \text{MSA}(Q = \mathbf{E}_n, K = \mathbf{E}_n, V = \mathbf{E}_n) \quad (2)$$

$$\mathbf{h} = f_{att}(\text{ReLU}(\mathbf{C}_n)) \quad (3)$$

其中,注意力聚合函数  $f_{att}(\cdot)$  由前馈神经网络实现。

### 3.2.2 候选增强模块

本模块旨在探索候选新闻与其语义相似新闻之间的高阶相关性。

首先,由于新闻文本长度有限,对于单个候选新闻编码存在语义信息不足的问题,因此引入 Mao 等<sup>[9]</sup>的研究,通过搜索语义相似的新闻丰富候选新闻语义。给定候选新闻列表  $N_c$ ,使用一个预训练语言模型(Pretrained Language Model, PLM)  $\phi(\cdot)$  从新闻文本  $T_i$  和  $T_j$  中获取新闻表示,并计算新闻  $\mathbf{n}_i$  和  $\mathbf{n}_j$  的相似度分数  $s_{i,j}$ :

$$s_{i,j} = \text{sim}(\mathbf{n}_i, \mathbf{n}_j) = \text{cosine}(\phi(T_i), \phi(T_j)) \quad (4)$$

对于候选新闻  $c_i$ ,根据相似度分数对语义相似新闻列表排序,并选择得分最高的  $|N|$  个新闻,使用新闻编码器对其编码,用  $\{\mathbf{h}_{c,i}^{(0)}\}_{i=0}^{|N|} \in \mathbb{R}^{|N| \times d}$  表示。

其次,构建候选新闻超图  $G_n$ ,捕捉其中复杂的高阶关系。为了丰富候选新闻的语义信息,连接候选新闻与其语义相似新闻以构建超边,得到关联矩阵  $\mathbf{H}_c \in \mathbb{R}^{|M| \times (|N|+1) \times |M|}$ ,并使用超图注意力网络<sup>[21]</sup>学习其中包含的高阶关系。该处理包含两个阶段的传播过程:第一阶段是节点向量聚合到超边,获得超边的嵌入向量;第二阶段是超边向量聚合到节点,获得

## 3.2 模型结构

本文提出了一种结合超图学习的多注意力机制新闻推荐模型 MNRHL,其主要包含 3 个部分:候选增强模块、用户兴趣模块和点击概率预测。MNRHL 的具体结构如图 2 所示。

节点的嵌入向量。

节点到超边的信息传播过程:使用多头注意力机制将语义相关新闻嵌入  $\{\mathbf{h}_{c,i}^{(0)}\}_{i=0}^{|N|} \in \mathbb{R}^{|N| \times d}$  作为输入,并使用  $[\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_N]$  表示所有节点信息,将超边连接的节点信息聚合到  $e_j$ 。对于每条超边  $e_j$ ,计算其内部每一个节点  $\mathbf{v}_i$  的重要性权重:

$$\alpha_{i \rightarrow j}^k = \frac{\exp(\text{ReLU}(\mathbf{q}_{e_j}^k \mathbf{Q}_0^k \mathbf{v}_i))}{\sum_{n \in V_{e_j}} \exp(\text{ReLU}(\mathbf{q}_{e_j}^k \mathbf{Q}_0^k \mathbf{v}_n))} \quad (5)$$

其中,  $V_{e_j}$  是超边  $e_j$  的节点集,  $\mathbf{Q}_0^k$  是可学习参数,  $\mathbf{q}_{e_j}^k$  是超边  $e_j$  在第  $k$  个注意力头的可学习超图嵌入。

其次,将  $K$  个注意力头的输出进行拼接,得到最终的超边表示  $\mathbf{h}_{e_j}$ 。

$$\mathbf{h}_{e_j} = \parallel_{k=1}^K \mathbf{V}_0^k \left( \sum_{n \in V_{e_j}} \alpha_{i \rightarrow j}^k \cdot \mathbf{v}_n \right) \quad (6)$$

其中,  $\mathbf{V}_0^k$  是可学习参数。

超边到节点的信息传播过程:使用得到的超边信息更新节点嵌入,使其包含全局高阶关联信息。与第一阶段相同,使用注意力机制执行聚合,学习节点  $\mathbf{v}_i$  的下一层表示:

$$\beta_{j \rightarrow i}^k = \frac{\exp(\text{ReLU}(\mathbf{h}_{e_j}^k \mathbf{Q}_1^k \mathbf{v}_i))}{\sum_{e_n \in E_{v_i}} \exp(\text{ReLU}(\mathbf{h}_{e_n}^k \mathbf{Q}_1^k \mathbf{v}_i))} \quad (7)$$

$$\mathbf{h}_{v_i} = \parallel_{k=1}^K \mathbf{V}_1^k \left( \sum_{e_n \in E_{v_i}} \beta_{j \rightarrow i}^k \cdot \mathbf{h}_{e_n} \right) \quad (8)$$

其中,  $E_{v_i}$  是节点  $v_i$  的超边集,  $\mathbf{Q}_1^k$  和  $\mathbf{V}_1^k$  是可学习参数。经过  $L$  层后,得到语义相关新闻特征  $\{\mathbf{h}_{c,i}^{(L)}\}_{i=0}^{|N|}$ 。

最后,采用注意力机制聚合语义相似的新闻嵌入表示,以得到候选增强新闻表示  $\mathbf{n}_c$ :

$$\mathbf{n}_c = \text{Attn}(Q = \mathbf{h}_{c,0}^{(L)}, K = \{\mathbf{h}_{c,i}^{(L)}\}_{i=0}^{|N|}, V = \{\mathbf{h}_{c,i}^{(L)}\}_{i=0}^{|N|}) \quad (9)$$

其中,  $\text{Attn}(Q, K, V)$  表示标准注意力模块。当  $i=0$  时,  $\{\mathbf{h}_{c,i}^{(L)}\}_{i=0}^{|N|}$  表示候选新闻本身。

### 3.2.3 用户兴趣模块

本模块使用用户  $u$  的历史交互数据构建超图  $G_u$  对用户兴趣建模, 显式建模用户历史点击新闻和潜在主题之间的交互信息, 捕捉它们之间复杂的多元关系。然后采用包含多种注意力机制的神经网络架构挖掘用户兴趣, 以得到用户兴趣表示。

首先, 将用户的历史点击新闻  $H_u$  作为新闻级兴趣表示的一组新闻节点; 将被点击的新闻主题  $P_u$  作为主题级兴趣表示的一组主题节点。为了捕获新闻与主题之间的高阶交互信息, 用户兴趣超图中定义了 4 种类型的超边。1) 主题级超边  $e_1$ : 连接用户历史交互记录中的所有主题, 以捕获主题级别的用户兴趣变化。2) 主题-新闻超边  $e_2$ : 连接同一主题的所有新闻, 以捕获新闻与主题之间的隶属关系信息。3) 新闻-新闻超边  $e_3$ : 计算新闻嵌入之间的相似性, 连接语义相似的新闻, 以建模用户的细粒度语义偏好。4) 新闻级超边  $e_4$ : 连接用户历史交互记录中的所有新闻, 以捕获新闻级别的用户兴趣。然后, 聚合所有超边, 以生成用户超图关联矩阵表示  $\mathbf{H}_u \in \mathbb{R}^{(|H|+|P|) \times (|H|+|P|)}$ 。

与候选增强超图模块类似, 给定用户  $u$  的历史交互数据  $H_u$ , 使用新闻编码器学习用户历史点击新闻的嵌入表示  $[\mathbf{n}_1^{(0)}, \mathbf{n}_2^{(0)}, \dots, \mathbf{n}_{|H|}^{(0)}] \in \mathbb{R}^{|H| \times d}$ , 并将其与新闻主题节点嵌入  $[\mathbf{t}_1^{(0)}, \mathbf{t}_2^{(0)}, \dots, \mathbf{t}_{|P|}^{(0)}] \in \mathbb{R}^{|P| \times d}$  一起作为输入, 使用超图注意力网络来学习其中包含的高阶关系, 输出历史点击新闻节点特征  $\mathbf{h}_u^n = [\mathbf{n}_1^{(L)}, \mathbf{n}_2^{(L)}, \dots, \mathbf{n}_{|H|}^{(L)}] \in \mathbb{R}^{|H| \times d}$  和新闻主题节点特征  $\mathbf{h}_u^t = [\mathbf{t}_1^{(L)}, \mathbf{t}_2^{(L)}, \dots, \mathbf{t}_{|P|}^{(L)}] \in \mathbb{R}^{|P| \times d}$ 。然后, 将第  $i$  篇新闻表示与其相应主题表示连接起来, 用  $\mathbf{h}_i^n$  表示。

最后, 使用加性注意力机制<sup>[20]</sup>来模拟用户对不同新闻的不同关注度。用户对第  $i$  篇新闻的关注度为:

$$\alpha_i^n = \frac{\exp(\mathbf{q}_u^\top \tanh(\mathbf{V}_u \times \mathbf{h}_i^n + \mathbf{v}_u))}{\sum_{j=1}^{|H|} \exp(\mathbf{q}_u^\top \tanh(\mathbf{V}_u \times \mathbf{h}_j^n + \mathbf{v}_u))} \quad (10)$$

其中,  $\mathbf{q}_u^\top$ ,  $\mathbf{V}_u$  和  $\mathbf{v}_u$  是注意力网络中的参数。用户兴趣  $\mathbf{h}_u$  是对每篇新闻向量的加权求和。用户兴趣  $\mathbf{h}_u$  为:

$$\mathbf{h}_u = \sum_{i=1}^{|H|} \alpha_i^n \mathbf{h}_i^n \quad (11)$$

由于用户的点击行为依赖于用户对候选新闻的兴趣<sup>[22]</sup>, 因此, 基于候选新闻表示  $\mathbf{n}_c$  与用户兴趣表示  $\mathbf{h}_u$ , 通过一个激活单元自适应地学习用户对候选新闻的偏好, 生成融合用户对候选新闻兴趣的用户兴趣向量  $\mathbf{u}_c$ :

$$\alpha_c = \frac{\exp(\mathbf{n}_c \tanh(\mathbf{h}_u \times \mathbf{h}_c^n + \mathbf{h}_u))}{\sum_{j=1}^{|H|} \exp(\mathbf{n}_c \tanh(\mathbf{h}_u \times \mathbf{h}_j^n + \mathbf{h}_u))} \quad (12)$$

$$\mathbf{u}_c = \sum_{i=1}^{|H|} \alpha_c \mathbf{h}_i^n \quad (13)$$

### 3.2.4 点击概率预测模块

本模块旨在预测用户点击候选新闻的概率。以候选新闻表示  $\mathbf{n}_c$  和最终的用户兴趣表示  $\mathbf{u}_c$  作为输入, 通过内积函数计算用户与新闻之间的关联分数, 进行下一条新闻的预测。点击概率  $\hat{s}$  为:

$$\hat{s} = \mathbf{u}_c^\top \cdot \mathbf{n}_c \quad (14)$$

此外, 采用负采样策略<sup>[23]</sup>进行模型训练, 优化了训练数据集  $D$  上的 NCE 损失, 并使用对数似然函数作为损失函数。

$$\mathcal{L} = - \sum_{i=1}^{|D|} \log \frac{\exp(\hat{s}_i^+)}{\exp(\hat{s}_i^+) + \sum_{j=1}^{|D|} \exp(\hat{s}_{i,j}^-)} \quad (15)$$

其中,  $S$  是负样本的数量,  $\hat{s}_i^+$  表示第  $i$  个正样本的预测得分,  $\hat{s}_{i,j}^-$  表示第  $i$  个正样本匹配的负样本集中第  $j$  个样本的预测得分。

## 4 实验

### 4.1 数据集

为了验证 MNRHL 模型的有效性, 在两个大规模新闻数据集即 MIND-large 和 MIND-small<sup>[24]</sup>上进行大量实验。MIND 数据集由微软新闻网站 6 周内采集的 100 万匿名用户的真实新闻日志构成, 包括 MIND-large 和 MIND-small 两个版本。其中, MIND-large 包含 161 031 篇新闻、1 000 000 个用户和 24 155 470 条行为日志; MIND-small 则包含 93 698 篇新闻、50 000 个用户和 230 117 条行为日志。以 MIND-large 为例, 该数据集的统计信息如表 1 所列。

表 1 MIND-Large 数据集的统计信息

Table 1 Statistic information of MIND-Large dataset

统计信息	统计量
用户	1 000 000
新闻	161 013
印象日志	15 777 377
行为数据	24 155 470
新闻类别	20
标题平均长度	11.52
摘要平均长度	43.00
内容平均长度	585.05

### 4.2 实验设置与评价指标

本实验基于 PyTorch 框架实现模型。使用用户最近点击的 50 条新闻了解用户表征, 将新闻标题最大长度限制为 32 个单词。采用预训练 GloVe 嵌入<sup>[25]</sup>进行初始词嵌入, 超图的特征嵌入维度设置为 300, 多头注意力机制每个头的输出维度为 25。模型负采样率设置为 4, 新闻表示维度为 300, 采用 Adam<sup>[26]</sup>优化器以  $1 \times 10^{-4}$  的学习率来优化训练过程。

采用 4 种广泛评估推荐模型的评价指标, 分别为 AUC (计算 ROC 曲线下的面积)、MRR (平均倒数排名)、NDCG@5 和 NDCG@10 (归一化折损累计增益)。

AUC 是 ROC 曲线下面部分区域的面积, 即随机正样本排名高于随机负样本排名的概率, 取值范围为  $[0.5, 1]$ 。

$$AUC = \frac{|\{(i, j) | Rank(p_i) < Rank(n_j)\}|}{N_p N_n} \quad (16)$$

其中,  $N_p$  和  $N_n$  分别表示正、负样本的数量,  $p_i$  表示第  $i$  个正样本的预测得分,  $n_j$  为第  $j$  个负样本的得分。

MRR 强调位置关系与顺序性, 其反映推荐排序列表中用户感兴趣的项目是否在更靠前的位置。

$$MRR = \frac{1}{|U|} \sum_{u \in U} \frac{1}{rank_u} \quad (17)$$

其中,  $u \in U$  表示遍历所有用户,  $rank_u$  表示用户  $u$  推荐列表中第一个真正例所在的位置。

NDCG@ $k$  考虑推荐结果的排序顺序, 主要用于评价排序列表的准确性。

$$NDCG@k = \frac{1}{n} \sum_{i=1}^k \frac{2^{r_i} - 1}{\log_2(i+1)} \quad (18)$$

其中,  $r_i$  是第  $i$  级新闻的相关性得分, 若点击新闻,  $r_i$  值为 1, 否则  $r_i$  值为 0。

当算法给用户推荐一个新闻列表时,用户实际点击的新闻越靠前,表明推荐准确度越高,推荐效果越好,上述指标的数值也会越大。

### 4.3 实验效果分析

#### 4.3.1 对比实验

本文提出的结合超图学习的多注意力机制新闻推荐模型(即 MNRHL),分别与以下 8 种先进的推荐模型进行比较。为了保证实验的公平性,将在相同数据集上进行对比实验,并采用相同的评价指标对模型推荐效果进行评估。

1) LibFM<sup>[27]</sup>:一种基于因子分解机(Factorization Machine, FM)的模型,将每个特征表示为一个因子向量的线性组合,并通过对这些因子向量的乘积进行建模来捕捉特征之间的交互作用。

2) DeepFM<sup>[28]</sup>:结合了 FM 和深层神经网络,与 LibFM 特性相同。

3) NAML<sup>[7]</sup>:一种基于深度学习的新闻推荐方法,通过多视图角度来学习新闻信息和用户表示。

4) LSTUR<sup>[29]</sup>:一种结合用户长期和短期兴趣的新闻推荐方法,通过 GRU 网络来学习用户短期兴趣嵌入,并通过用户 ID 嵌入来建模用户长期兴趣。

5) GERL<sup>[1]</sup>:将用户与新闻间的交互信息建模为图结构,并设计二跳图学习模块,采用图注意力网络聚合新闻和用户的邻居嵌入。

6) GNewsRec<sup>[10]</sup>:构造一个用户-新闻-主题异构图来显式建模用户、新闻及潜在主题之间的交互。

7) GLORY<sup>[30]</sup>:一种结合全局图结构和局部个性化信息的个性化新闻推荐方法,以挖掘用户行为的隐含关联。

8) DIGAT<sup>[9]</sup>:一种由新闻图和用户图组成的双交互式新闻推荐模型,在新闻图和用户图之间执行有效的特征交互,能更精确地学习新闻-用户匹配表示。

#### 4.3.2 模型总体推荐效果

在 MIND-small 和 MIND-large 数据集上,采用 4 种评价指标(AUC, MRR, NDCG@5 以及 NDCG@10)进行对比实验,实验结果如表 2 所列。

表 2 总体比较

Table 2 Overall performance comparison

模型	MIND-small				MIND-large			
	AUC	MRR	NDCG@5	NDCG@10	AUC	MRR	NDCG@5	NDCG@10
LibFM	0.6001	0.2764	0.2992	0.3595	0.6116	0.2788	0.3006	0.3644
DeepFM	0.6213	0.2701	0.2997	0.3678	0.6365	0.2869	0.2937	0.3695
NAML	0.6588	0.3092	0.3411	0.4058	0.6765	0.3269	0.3623	0.4270
LSTUR	0.6611	0.3100	0.3419	0.4066	0.6752	0.3238	0.3610	0.4242
GERL	0.6527	0.3010	0.3293	0.3948	0.6810	0.3341	0.3634	0.4203
GNewsRec	0.6554	0.3027	0.3329	0.3980	0.6815	0.3345	0.3643	0.4210
GLORY	0.6604	0.3058	0.3302	0.4029	0.6900	0.3403	0.3719	0.4295
DIGAT	0.6877	0.3346	0.3714	0.4339	0.7008	0.3520	0.3846	0.4415
MNRHL	<b>0.6912</b>	<b>0.3387</b>	<b>0.3763</b>	<b>0.4365</b>	<b>0.7058</b>	<b>0.3572</b>	<b>0.3892</b>	<b>0.4456</b>

总体比较表 2 数据,可以得出以下结论。

传统的基于矩阵分解的 LibFM 模型的性能明显低于 DeepFM 模型,这表明在推荐算法中,深度模型在处理新闻和用户数据方面具有优越性。基于深度学习的新闻推荐方法(如 NAML, LSTUR 等)通常比传统的基于特征的推荐方法表现更好,这表明使用深度神经网络从原始数据中学习新闻内容和用户兴趣表示比手工提取的特征更有效。

大多情况下,基于图神经网络的方法(如 GERL, GNewsRec, GLORY, DIGAT 等)比基于传统深度学习技术的方法效果更好,这表明基于图神经网络的新闻推荐模型能够充分利用图结构中的节点和边信息学习节点之间的关系和信息传递规律,以增强模型的表征能力,在处理图结构数据和挖掘深层次交互信息方面表现更为优异。同时,利用图神经网络的信息聚合机制,聚合高阶邻域信息,挖掘更深层的用户兴趣特征,从而提升推荐性能。但传统图结构忽略了实体之间的高阶关系,使得用户建模不够充分。

本文提出的模型在两个数据集上的 4 种评价指标表现均优于其他对比方法,主要原因在于 MNRHL 模型不仅引入主题信息,通过构建超图并使用包含多种注意力机制的神经网络架构提取了更细粒度的用户兴趣,还结合用户对候选新闻的兴趣偏好,生成了更准确的用户兴趣特征表示。此外,该模型还构建了候选新闻超图,捕捉候选新闻与其语义相似新闻之间的高阶相关性,以丰富候选新闻语义,提高推荐的准确性。

#### 4.3.3 消融实验

由于在 MIND-large 数据集上的训练需要较高的 GPU 成本,因此实验只采用 MIND-small 数据集进行消融实验。为了评估 MNRHL 模型的有效性,设计了 3 个模型变体与所提模型进行对比。

1) MNRHL w/o HL:在候选增强模块与用户兴趣模块中,只使用普通图结构学习候选新闻和用户兴趣的表示,进行后续推荐。

2) MNRHL w/o CE:丢弃模型中的候选增强模块,仅学习单个候选新闻表示。

3) MNRHL w/o AU:在用户兴趣模块中,丢弃激活单元,即不考虑候选新闻对用户兴趣的影响。

从表 3 中可以发现,使用变体均会导致性能次优,表明各个组件在 MNRHL 中都是关键且有效的。与 MNRHL w/o HL 变体相比,原模型 MNRHL 具有明显优势。这是由于 MNRHL w/o HL 使用普通图结构,无法准确建模实体(新闻与主题、候选新闻与其语义相关新闻)之间的复杂关系。在超图结构中,每个节点(新闻、主题)都可以被表示为高维向量,其中包含丰富的特征信息,可进一步捕获丰富的新闻及用户特征表示。与 MNRHL w/o CE 变体相比, MNRHL 的性能有较大提升。这是因为 MNRHL 引入了语义相似新闻来丰富候选新闻的语义信息,从而为新闻推荐任务提高了匹配精度。MNRHL w/o AU 变体与原模型 MNRHL 相比,性能有所降低。这表明激活单元在提取用户历史点击兴趣的基础

上,可以进一步挖掘当前候选新闻对用户兴趣的影响,从而提升推荐模型的性能。

表 3 消融分析  
Table 3 Ablation analyses

模型	MIND-small			
	AUC	MRR	NDCG@5	NDCG@10
MNRHL	<b>0.6912</b>	<b>0.3387</b>	<b>0.3763</b>	<b>0.4365</b>
w/o HL	0.6802	0.3276	0.3647	0.4254
w/o CE	0.6807	0.3318	0.3679	0.4288
w/o AU	0.6876	0.3329	0.3701	0.4312

#### 4.3.4 超参数设置

超参数设置旨在寻找一组最优的参数,使模型达到最佳的推荐效果。为了验证部分超参数语义相似新闻数量  $N$  和超图注意力网络层数  $L$  对模型性能的影响,MNRHL 模型在 MIND-small 数据集上设置了如下两个超参数实验。

1) 探究语义相似新闻数量  $N$  对模型性能的影响。实验将  $N$  取值为  $\{1, 2, 3, 4, 5\}$ ,以 AUC 和 MRR 作为评价指标,观察其在 MIND-small 数据集上的表现。

实验结果如图 3 所示,模型在  $N=2$  时效果最佳。各评价指标先上升后下降,这是因为语义相似新闻虽然可以增强候选新闻的表示,但随着语义相似新闻数量的增加,使用 PLM 获取语义相似新闻时会增加噪声,同时还会削弱候选新闻的表征能力,从而影响模型性能。

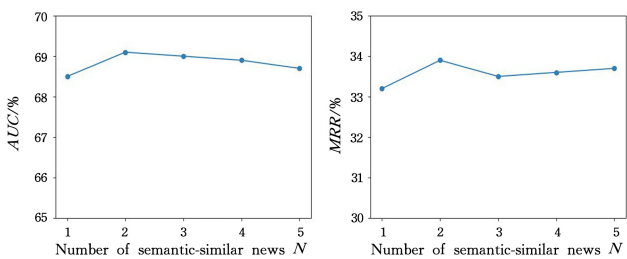


图 3 不同  $N$  值下 MNRHL 的性能

Fig. 3 Performance of MNRHL with different values of  $N$

2) 探究超图注意力网络层数  $L$  对模型性能的影响。实验将超图注意力网络层数  $L$  取值为  $\{1, 2, 3, 4\}$ ,以 AUC 和 MRR 作为评价指标,观察其在 MIND-small 数据集上的表现。

图 4 给出了 MNRHL 模型在不同的网络深度下的变化趋势。当  $L$  为 1 时,MNRHL 模型达到最佳性能。 $L$  过高会造成过拟合问题,降低模型的泛化能力,影响预测的准确性。这也说明,在信息传递过程中,超图结构可以使用较少的层数实现高阶信息传递。

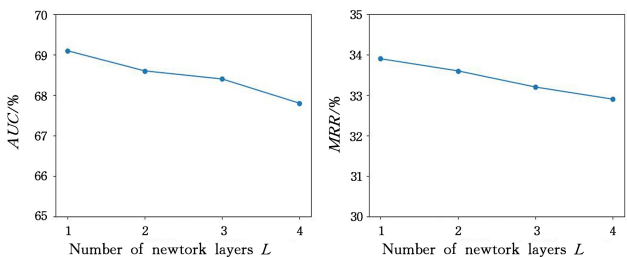


图 4 不同  $L$  值下 MNRHL 的性能

Fig. 4 Performance of MNRHL with different values of  $L$

**结束语** 本文提出了一种基于超图学习的多注意力机制新闻推荐模型(MNRHL 模型)。首先,构建候选新闻超图,以

捕获候选新闻与其语义相似新闻的高阶相关性,丰富候选新闻语义;其次,引入新闻主题信息构建用户兴趣超图,采用包含多种注意力机制的神经网络架构挖掘深层的用户细粒度兴趣特征。其中,为考虑候选新闻对用户兴趣的影响,引入激活单元,结合候选新闻特征进一步提取用户兴趣。在未来的研究中,考虑将更多的辅助信息(如社交关系、用户阅读的时间及位置等)融入用户历史交互数据中,充分挖掘用户的兴趣特征,构造更丰富有效的特征提取网络以进行推荐。此外,还将尝试引入对比学习任务,将其作为辅助任务与推荐预测任务进行联合优化,通过引入自监督信号来进一步缓解交互数据的稀疏性问题。

## 参考文献

- [1] GE S Y, WU C H, WU F Z, et al. Graph enhanced representation learning for news recommendation[C]// Proceedings of the Web Conference 2020. New York: ACM, 2020: 2863-2869.
- [2] HU L M, XU S Y, LI C, et al. Graph neural news recommendation with unsupervised preference disentanglement[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2020: 4255-4264.
- [3] WANG S Q, LI X X, SUN F Z, et al. Survey of research on personalized news recommendation techniques[J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(1): 18-29.
- [4] MENG X F, HUO H J, ZHANG X Y, et al. Survey of research on personalized news recommendation approaches[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(12): 2840-2860.
- [5] TIAN X, DING Q, LIAO Z H, et al. Survey on deep learning based news recommendation algorithm[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(6): 971-998.
- [6] MOREIRA G, JANNACH D, CUNHA A. Contextual hybrid session-based news recommendation with recurrent neural networks[J]. IEEE Access, 2019, 7: 169185-169203.
- [7] WU C H, WU F Z, AN M X, et al. Neural news recommendation with attentive multi-view learning[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 3863-3869.
- [8] ZHANG Q, JIA Q L, WANG C Y, et al. AMM: Attentive multi-field matching for news recommendation[C]// The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2021: 1588-1592.
- [9] MAO Z M, LI J, WANG H R, et al. DIGAT: Modeling news recommendation with dual-graph interaction[C]// Findings of the Association for Computational Linguistics. Stroudsburg: ACL, 2022: 6595-6607.
- [10] HU L M, LI C, SHI C, et al. Graph neural news recommendation with long-term and short-term interest modeling[J]. Information Processing and Management, 2020, 57(2): 102142.
- [11] SCHÖLKOPF B, PLATT J, HOFMANN T. Learning with Hypergraphs: Clustering, Classification, and Embedding[C]// Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference. Cambridge: MIT Press, 2007: 1601-1608.
- [12] HU B D, WANG X E, WANG X Y, et al. Survey on hypergraph learning: algorithm classification and application analysis[J].

- Journal of Software,2022,33(2):498-523.
- [13] FENG Y F, YOU H X, ZHANG Z Z, et al. Hypergraph neural networks[C]// Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'19/ IAAI'19/ EAAI'19). Palo Alto: AAAI Press, 2019:3558-3565.
- [14] YADATI N, NIMISHAKAVI M, YADAV P, et al. HyperGCN: a new method of training graph convolutional networks on hypergraphs[C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2019:1511-1522.
- [15] WANG J L, DING K Z, HONG L J, et al. Next-item recommendation with sequential hyper-graphs[C]// Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20). New York: ACM, 2020:1101-1110.
- [16] XIA X, YIN H Z, YU J L, et al. Self-supervised hypergraph convolutional networks for session-based recommendation [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 35. Palo Alto: AAAI Press, 2021:4503-4511.
- [17] WANG J L, DING K Z, ZHU Z W, et al. Session-based recommendation with hypergraph attention networks [C] // Proceedings of the 2021 SIAM International Conference on Data Mining (SDM). Philadelphia: SIAM, 2021:82-90.
- [18] YU J L, YIN H Z, LI J D, et al. Self-supervised multi-channel hypergraph convolutional network for social recommendation [C]// Proceedings of the Web Conference 2021. New York: ACM, 2021:413-424.
- [19] WANG Z H. English news text recommendation method based on hypergraph random walk label expansion[J]. Computational Intelligence and Neuroscience, 2022, 2022.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2017:6000-6010.
- [21] LIU W X, ZHANG Z Z, WANG B. Dual-view hypergraph attention network for news recommendation [J]. Engineering Applications of Artificial Intelligence, 2024, 133(C):108256.
- [22] FENG Y F, LYU F Y, SHEN W C, et al. Deep session interest network for click-through rate prediction [C] // Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI'19). Palo Alto: AAAI Press, 2019:2301-2307.
- [23] WU C H, WU F Z, GE S Y, et al. Neural news recommendation with multi-head self-attention [C] // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Stroudsburg: ACL, 2019: 6389-6394.
- [24] WU F Z, QIAO Y, CHEN J H, et al. MIND: A large-scale dataset for news recommendation [C] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2020:3597-3606.
- [25] PENNINGTON J, SOCHER R, MANNING C. GloVe: Global vectors for word representation [C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2014:1532-1543.
- [26] KINGMA D, BA J. Adam: a method for stochastic optimization [C] // Proceedings of the International Conference on Learning Representations. Washington: ICLR, 2015:1412-1427.
- [27] RENDLE S. Factorization machines with libFM[J]. ACM Transactions on Intelligent Systems and Technology, 2012, 3(3):1-22.
- [28] GUO H F, TANG R M, YE Y M, et al. DeepFM: A factorization-machine based neural network for CTR prediction [C] // Proceedings of the 26th International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2017:1725-1731.
- [29] AN M X, WU F Z, WU C H, et al. Neural news recommendation with long- and short-term user representations [C] // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2019:336-345.
- [30] YANG B M, LIU D R, SUZUMURA T, et al. Going beyond local: Global graph-enhanced personalized [C] // Proceedings of the 17th ACM Conference on Recommender Systems. New York: ACM, 2023:24-34.



**MENG Xiangfu**, born in 1981, Ph. D., professor, Ph. D supervisor. His main research interests include spatio-temporal big data analysis, recommendation system and artificial intelligence.



**WANG Wanchun**, born in 2000, master. Her main research interests include recommendation system and artificial intelligence.