新闻推荐提示学习

Zizhuo Zhang 华中科技大学电子信息与通信学院，中国武汉 zhangzizhuo@hust.edu.cn

Bang Wang 华中科技大学电子信息与通信学院，中国武汉 wangbang@hust.edu.cn

arXiv:2304.05263v1 [cs.IR] 11 Apr 2023

摘要

一些近期的新闻推荐（NR）方法采用预训练语言模型（PLM）对新闻进行编码，以获得新闻表示，并遵循标准预训练与微调范式，同时精心设计针对推荐任务的神经网络和目标函数。由于任务目标与预训练语言模型的目标并不一致，我们认为，这类建模范式未能充分利用预训练过程中蕴含的丰富语义信息和语言知识。近期，一种名为提示学习的预训练、提示和预测范式在自然语言处理领域取得了诸多成功。本文首次尝试将这一新范式用于构建新闻推荐提示学习（Prompt4NR）框架，将预测用户是否会点击候选新闻的任务转化为完形填空式掩码预测任务。具体而言，我们设计了一系列提示模板，包括离散、连续和混合模板，并构建相应的答案空间，以检验所提出的Prompt4NR框架。此外，我们采用提示集成来整合多个提示模板的预测结果。在MIND数据集上进行的大量实验验证了Prompt4NR的有效性，并取得了一系列新的基准测试结果。

CCS 概念

• **信息系统** → **推荐系统**；语言模型；• **计算方法**。**自然语言处理**

关键词

提示学习，新闻推荐，预训练语言模型

ACM 参考格式：

Zizhuo Zhang 和 Bang Wang。2023 年。面向新闻推荐的提示学习。载于第 *46* 届 *ACM* 国际信息检索研究与发展 *SIGIR* 大会论文集（*SIGIR '23*），*2023*年*7*月*23*日至*27*日，台湾台北。ACM，纽约，纽约州，美国，共 11 页。https://doi.org/10.1145/3539618.3591752

允许免费制作本作品全部或部分内容的数字或纸质副本，供个人或课堂使用，但不得为营利或商业利益而复制或分发，且所有副本均须附上本声明及首页的完整引文。对于非作者拥有的本作品组成部分，必须尊重其版权。注明出处后，可以摘录。若要以其他方式复制、重新发表、发布到服务器或分发至列表，则须事先获得具体许可和/或支付费用。请通过 permissions@acm.org 申请许可。*SIGIR '23*，*2023*年*7*月*23*日至*27*日，台北，台湾© 2023 版权所有，归所有者/作者所有。出版权授权予 ACM。ACM ISBN 978-1-4503-9408-6/23/07…$ 15.00 https://doi.org/10.1145/3539618.3591752

1 引言

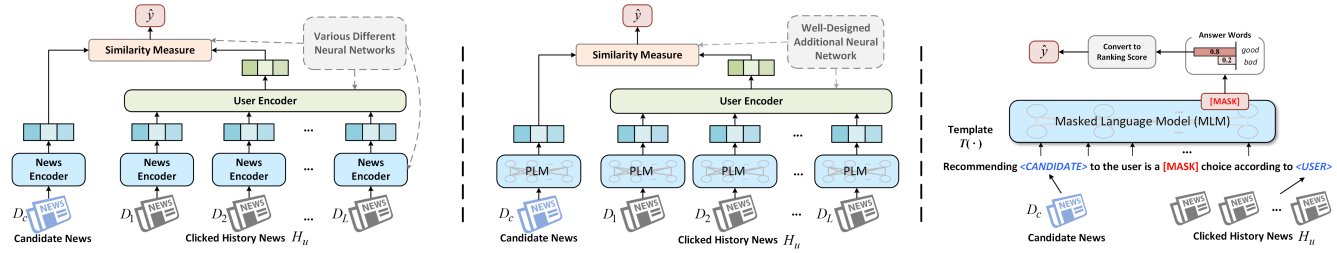
如今，Google 新闻等在线新闻平台已成为人们高效获取日常信息的重要门户 [4]。作为缓解信息过载问题的过滤工具，新闻推荐（NR）能够有效帮助用户从海量新闻中找到自己最感兴趣的新闻文章 [17, 24]，[25, 46]。

大多数现有的神经新闻推荐方法主要致力于设计各种精巧的神经网络，以编码新闻和用户的表示 [1, 11, 13, 30, 33, 37, 38, 42, 45, 56, 57, 60]。我们通过图 1(a) 总结了这些模型中的常见方法，其中核心模块包括新闻编码器、用户编码器和相似度度量。在相关文献中，这些模块通过不同的神经网络实现。例如，Wu 等人 [45] 采用多头自注意力网络作为新闻编码器和用户编码器。Wang 等人 [37] 提出通过堆叠空洞卷积学习一种分层多级新闻表示，以进行细粒度匹配。在这些神经模型中，静态词嵌入（例如 Word2Vec [23] 和 GloVe [27]）通常被用作模型训练的初始化参数，而训练主要关注从新闻推荐数据集中挖掘领域内信息，却忽略了来自现实世界大规模语料库的丰富语义和语言信息。

一些近期方法更进一步，引入预训练语言模型 （PLM）来学习新闻表示 [2, 9, 48, 54, 55, 58]。我们通过图 1(b) 总结了这些方法的共同做法：采用原始的预训练并微调 范式 [15] 来适配下游新闻推荐任务。在这一范式中，预训练语言模型仅用作新闻编码器，另行设计一个神经网络对用户进行编码。整个模型通过新闻推荐专属目标函数进行训练。尽管这些方法展现出令人瞩目的性能提升，但由于下游新闻推荐目标与预训练语言模型训练目标之间存在不一致，它们未能充分利用大规模预训练语言模型中丰富的百科全书式知识。

近期，一种名为提示学习的新颖预训练、提示与预测范式在自然语言处理（NLP）领域的诸多应用中取得了显著成功 [3, 53]。这一新范式的基本思想是通过设计与任务相关的提示模板和答案词空间，将下游任务重新表述为预训练语言模型的训练任务 [6, 15]。鉴于其展现出的巨大潜力，我们也希望考察提示学习范式在新闻推荐领域的适用性和有效性。据我们所知，本文首次进行了尝试，并提出了新闻推荐提示学习（Prompt4NR）框架。

图 1(c) 展示了我们的 Prompt4NR框架的设计理念：我们提出将一个新闻推荐任务转换为完形填空式的掩码预测任务。具体而言，给定用户的点击历史



(a) Common neural NR model framework

(b) Common PLM-based NR model framework with pre-train and fine-tune paradigm

(c) Idea of Prompt4NR framework with prompt learning paradigm

图1：神经新闻推荐模型、基于预训练语言模型的预训练与微调模型，以及我们的 Prompt4NR 框架之间的差异。

$H_u = \{D_1, D_2, \ldots\ldots D_L\}$和一条候选新闻 D_c，我们首先将 H_u 转换为一个句子，记作 <USER>，以便从历史记录中编码用户兴趣。我们还将 D_c 转换为一个句子，记作 <CANDIDATE>，用于表示一条候选新闻。随后，我们设计一个提示模板 T（<USER>，<CANDIDATE>），将这两个句子拼接成另一个包含 [MASK] 标记的句子。经过掩码语言模型（MLM）处理后，我们设计了一个与该模板对应的答案空间，用于预测 [MASK]，随后将其转换为排名分数，以确定是否应向用户推荐这条候选新闻。

本文围绕 Prompt4NR 的核心设计问题如何影响推荐性能，展开了一系列研究。这些问题包括：（1）模板：在新闻推荐领域，哪种模板更适合整合新闻数据和用户行为？（2）答案映射器：如何将推荐标签映射为用于 [MASK] 预测的答案词？（3）集成：能否整合不同模板的优势来提升性能？针对提示模板，我们从新闻推荐任务中 <USER> 与 <CANDIDATE> 之间最受关注的四个方面出发，设计了三类模板，包括离散模板、连续模板和混合模板，分别涉及语义相关性、用户情感、用户行为 和推荐效用。针对答案映射器，我们构建了一个包含两个含义相反的答案词的二元答案空间，并根据模板确定答案词，使其对应于用户是否点击候选新闻这一真实标签。针对集成，我们采用多提示集成，对不同模板的预测进行决策融合。我们在广泛使用的 MIND数据集 [52] 上设计了大量实验。一组新的基准测试结果验证了 Prompt4NR框架的有效性：其推荐性能优于最先进的竞争方法。

2 相关工作

本节回顾新闻推荐和提示学习领域现有的相关研究。

2.1 新闻推荐

人们广泛开发了各种神经网络，用于为新闻推荐任务[1, 7, 10, 11, 13, 16, 26, 29 33, 37 39, 41 45, 47, 49 51, 56, 57, 59, 60] – – – – 编码新闻和用户表示。例如，Wu 等人 [45] 提出了 NRMS，采用多头注意力网络作为新闻编码器和用户编码器。Wang 等人 [37] 提出了 FIM，通过堆叠扩张卷积实现细粒度匹配。Qi 等人 [33] 提出了 HieRec，根据新闻的主题和子主题，将每位用户表示为一棵层次化兴趣树。Li 等人 [13] 提出了 MINER，用于捕捉用户的多种兴趣表示。尽管这些浅层神经网络效果显著，但缺乏理解新闻文本深层语言和语义信息的能力。此外，它们仅从新闻推荐任务中的监督信号学习知识，忽略了现实世界大规模语料库可能带来的益处。

最近，随着预训练语言模型在自然语言处理领域取得巨大成功，一些方法将预训练语言模型引入新闻推荐任务，并取得了显著提升 [2, 9, 48, 54, 55, 58]。例如，Wu 等人 [48] 提出了一种基于预训练语言模型的直观新闻推荐框架，使用预训练语言模型直接实例化现有神经新闻推荐模型中的新闻编码器。Bi等人 [2] 采用多任务学习框架，增强了BERT 对新闻类别和实体等多字段信息的编码能力。上述方法都遵循针对新闻推荐任务特定目标的标准预训练与微调范式。尽管取得了一定的性能提升，但由于其目标与预训练语言模型的目标不一致 [15]，这种范式可能无法充分利用预训练过程中为新闻推荐任务学习到的知识。

2.2 提示学习

提示学习是一种基于预训练语言模型的新型学习策略，它通过在输入文本中添加模板，利用预训练语言模型将下游任务转换为 [MASK] 预测的形式。如何设计提示模板是提示学习的核心组成部分。通常，模板可分为离散模板、连续模板和混合模板 [15]。离散模板由现有的自然语言词语组成，非常依赖人工经验。例如，Petroni 等人 [28] 手动设计模板，以探测语言模型中的知识。Schick 等人 [34] 提出了一种名为 PET 的半监督训练流程，将特定任务重新表述为完形填空式任务。尽管离散模板已在许多任务中取得成功，但手工设计的模板可能耗费成本，且未必达到全局最优。为了解决这些问题，连续模板和混合模板引入可学习的提示词元，以自动搜索模板，例如 AutoPrompt [35]，、Prefix Tuning [14]，、P-Tuning [19]，、P-Tuningv2 [18] 等。本文聚焦于如何将提示学习应用于新闻推荐任务。我们提出并实验了一系列离散模板、连续模板和混合模板，以探究不同的设计考量。

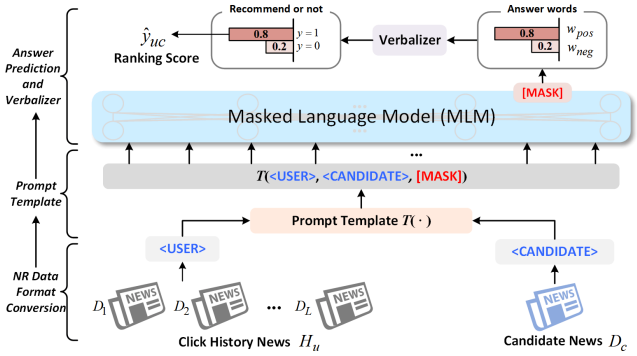


图 2：Prompt4NR框架示意图。

3 Prompt4NR框架

我们首先介绍新闻推荐（NR）的问题定义，接着介绍我们的 Prompt4NR 框架。然后，我们将解释训练策略和多提示集成。

3.1 新闻推荐任务与提示概述

分别将 U 和 D 表示为用户集和新闻集。每条新闻 $D \in$D 主要配有标题 $Title = \{w_1, w_2, \ldots\ldots w_M\}$，标题由一系列词语组成，其中 M表示词语数量。给定用户 u 的点击历史 $H_u = \{D_1, D_2, \ldots\ldots D_L\}$（包含 L 条已点击新闻）以及一条候选新闻 D_c，新闻推荐任务旨在预测排名分数 $\hat{y}_{uc}$，以估计用户 u 点击候选新闻 D_c 的概率。得分最高的候选新闻将推荐给该用户。

图 2 展示了我们的 Prompt4NR框架，其中包含三个主要模块：（1）NR 数据格式转换（第 3.2 节）；（2）提示模板（第 3.3 节）；（3）答案预测和答案映射器（第 3.4 节）。具体说明如下：

3.2 NRDataFormat 转换

给定点击历史 H_u和候选新闻 D_c，我们将其转换为自然语言句子，以适配后续的提示学习范式，分别表示为 <USER> 和 <CANDIDATE>。对于 <USER>，我们拼接用户历史记录 H_u中的新闻标题，并在每个标题的开头添加一个虚拟令牌 [NCLS]，以分隔每条已点击的新闻。对于 <CANDIDATE>，我们采用候选新闻 D_c的标题。我们将其形式化表示为：

$$\text{<USER>} \leftarrow \text{[NCLS]}\ Title_1 \ldots \text{[NCLS]}\ Title_L$$
$$\text{<CANDIDATE>} \leftarrow Title_c$$

其中$\{Title_1, , Title_L\}$ 对应于 $H_u = \{D_1, D_2, ..., D_L\}$中的新闻标题。直观地说，<USER>可以被视为对用户 u's兴趣领域的总结，而 <CANDIDATE> 则凝练了候选新闻的核心文本语义。二者都将作为后续提示模板的输入文本数据。

3.3 提示模板

作为 Prompt4NR 的核心组件，提示模板 $T(\cdot)$ 包装输入数据（<USER>，<CANDIDATE>），将新闻推荐任务转换为完形填空式任务，以预测 [MASK]：

$$x_{prompt} = T(\text{<USER>}, \text{<CANDIDATE>}, \text{[MASK]}) \quad (1)$$

作为首个将提示学习应用于新闻推荐任务的研究，我们希望了解哪种提示模板最有利于提升推荐性能。为此，我们从如何捕捉用户与候选新闻之间匹配信号的不同角度，设计了三种类型的模板，包括离散模板、连续模板和混合模板。表 1 总结了我们设计的提示模板。接下来，我们将介绍这些模板背后的设计思路。

3.3.1 离散模板。 离散模板是提示学习中最常见的一种模板工程形式，它通过人类可理解的自然语言构造输入数据，因此需要一定的先验经验知识。我们从四种不同的考量出发设计了四种离散模板，每种模板都对应一种衡量用户兴趣与候选新闻之间匹配信号的方法。也就是说，我们探索了哪种风格的 [MASK] 完形填空相似度度量适用于新闻推荐任务。

- **语义相关性：**这是为了考察相关新闻内容是否是用户阅读新闻的核心动机。换言之，用户是否对某些特定主题和内容抱有持续兴趣。为此，我们将新闻推荐任务转化为判断 <CANDIDATE> 与 <USER>之间的相关性，并将答案词设为“相关”和“不相关”。
- **用户情感：**这是为了探究用户对新闻的情感反应是否是影响最大的因素。换句话说，用户选择阅读某条新闻，仿佛这条新闻最能满足用户的情感需求。我们使用情感词“有趣”和“无聊”作为答案，以估计用户对 <候选>的情感反应。
- **用户行为：**这是为了研究掩码语言模型能否直接用作点击预测器。换句话说，兴趣引导行为，而行为反映兴趣。在告知掩码语言模型 <用户> 和<候选>后，我们让掩码语言模型直接预测用户是否会点击这条新闻，答案词为“是”和“否”。
- **推荐效用：**这是为了探究掩码语言模型能否自行判断推荐一条候选新闻的潜在优缺点，也就是进行此类推荐的效用。为此，我们向掩码语言模型提出一个关于效用的问题，并将“好”和“坏”作为推荐效用预测的答案词。

上述模板虽然在模板句子和答案词上看似只有少许差异，但预期能够通过预先定义的自然语言句子和预先选定的答案词，利用大型百科全书式预训练语言模型中蕴含的语义和语言学知识。一方面，我们注意到，这正是提示学习范式的核心理念：从预训练语言模型的词汇表中预测答案词的概率，仿佛将特定任务的输入句子插入大型语料库中，以训练预训练语言模型。另一方面，这类人工设计的模板虽然包含精心设计的自然语言句子，但显然无法涵盖所有可能的情况。因此，我们可以使用一些虚拟词元在模板中探索更多情况，即构建连续模板。

表 1：本文设计的提示模板，包括离散、连续和混合模板。

类型	观点	模板 T(<用户>, <候选项>, [掩码])	答案词汇
离散模板	相关性	<候选项> 是 [掩码] 对 <用户>	相关/不相关
	情感	用户感到 [掩码] 关于 <候选项> 根据其兴趣领域 <用户>	有趣/无聊
	行为	用户： <用户> [分隔符] 新闻： <候选项> [分隔符] 用户会点击这条新闻吗？ [掩码]	是/否
	效用	推荐 <候选项>给用户是一个 [掩码] 选择，依据 <用户>	好/坏
连续模板	相关性	$[Q_1]...[Q_{n_2}]$ <候选项> $[M_1]...[M_{n_3}]$ [掩码] $[P_1]...[P_{n_1}]$ <用户>	相关/不相关
	情感	$[M_1]...[M_{n_3}]$ [掩码] $[Q_1]...[Q_{n_2}]$ <候选> $[P_1]...[P_{n_1}]$ <用户>	有趣/无聊
	行为	$[P_1]...[P_{n_1}]$ <用户> [SEP] $[Q_1]...[Q_{n_2}]$ <候选项> [SEP] $[M_1]...[M_{n_3}]$ [MASK]	是/否
	效用	$[Q_1]...[Q_{n_2}]$ <候选项> $[M_1]...[M_{n_3}]$ [MASK] $[P_1]...[P_{n_1}]$ <用户>	好/坏
混合模板	相关性	$[P_1]...[P_{n_1}]$ <用户> [分隔符] $[Q_1]...[Q_{n_2}]$ <候选项> [分隔符] 这条新闻 [掩码] 符合用户感兴趣的领域	相关/不相关
	情感	$[P_1]...[P_{n_1}]$ <用户> [分隔符] $[Q_1]...[Q_{n_2}]$ <候选项> [分隔符] 用户对这条新闻的感受是 [掩码] 关于这条新闻	有趣/无聊
	行为	$[P_1]...[P_{n_1}]$ <用户> [分隔符] $[Q_1]...[Q_{n_2}]$ <候选项> [分隔符]用户会点击这条新闻吗？ [掩码]	是/否
	效用	$[P_1]...[P_{n_1}]$ <用户> [分隔符] $[Q_1]...[Q_{n_2}]$ <候选项> [分隔符] 向用户推荐这条新闻是一种 [掩码] 选择	好/坏

3.3.2 连续模板。 表 1 展示了我们设计的四种连续模板，每种模板都对应一种离散模板。我们分别在 <USER>、< CANDIDATE >和 [MASK]，前添加一些可学习的虚拟词元，记作 $[P_1]...[P_{n_1}]$、$[Q_1]...[Q_{n_2}]$ 和 $[M_1]...[M_{n_3}]$，其中 n_1、n_2、n_3 表示虚拟词元的数量。答案词和词元位置设置方面，我们参考了先前的离散模板。尽管连续模板赋予模型更大的自由度，但这些虚拟词元的嵌入是随机初始化的，这可能会引入歧义，导致预训练语言模型中的知识未能得到充分利用。我们进一步设计了一类混合模板，试图结合离散模板和连续模板各自的优点。

3.3.3 混合模板。在混合模板中，我们保留 <USER>和 < CANDIDATE>前面的虚拟词元 $[P_i]$和 $[Q_j]$，旨在自动搜索适当的格式，以便将这些信息呈现给预训练语言模型。我们将这些虚拟词元 $[M_k]$ 替换为自然语言，并使用用于答案预测的 [MASK] 词元。如表 1 所示，我们仍然设计了四个具有代表性的自然句子，分别对应我们的一个设计考虑因素。因此，一个混合模板由连续模板、 [SEP] 词元和一个自然句子组成。与这些连续模板相比，我们认为，这种混合模板既可以通过连续模板利用虚拟词元获得更多选择，又可以通过离散模板保留自然句子来引导答案方向。

3.4 答案预测和答案映射器

给定一个点击历史 H_u 和一条候选新闻 D_c，它们对应一个真实标签 $y \in \{0, 1\}$，用于反映用户是否点击了该候选新闻（$y = 1$）($y = 0$)。我们设计了一个 答案映射器 $v(\cdot)$，将这些标签映射到预训练语言模型词汇表 W 中的两个答案词，如下所示：

$$v(y) = \begin{cases} w_{pos}, & y = 1, \\ w_{neg}, & y = 0, \end{cases} \tag{2}$$

其中 $\mathrm{W}_a = \{w_{pos}, w_{neg}\} \subset \mathrm{W}$ 是答案词空间，可能因所使用的提示模板而异。新闻推荐任务被转换为完形填空式任务，由预训练的掩码语言模型 M（例如 BERT [5]）预测答案词为 [MASK] 的概率：

$$P(y|H_u, D_c) = P_{\mathcal{M}}\left([MASK] = v(y)|x_{prompt}\right), \tag{3}$$

其中， $P_{\mathcal{M}}\left([MASK] = v(y = 1)|x_{prompt}\right)$ 可以视为是否推荐当前候选新闻的置信度。我们将其作为排名分数，以形成推荐列表。需要指出的是，本文采用了一个由两个预训练语言模型词汇表词语构成的简单答案空间。未来，我们将研究由更多词汇表词语甚至虚拟答案构成的更复杂答案空间。

3.5 训练

与训练额外的特定任务神经模型所采用的预训练与微调范式相比，我们的 Prompt4NR 模型只需调整预训练语言模型的参数。我们采用交叉熵损失函数来训练模型：

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log P_i + (1 - y_i)\log(1 - P_i)\right], \tag{4}$$

其中， y_i 和 P_i 分别是第 i个训练实例的真实标签和预测概率。我们使用带有 L2 正则化的 AdamW [22] 优化器进行模型训练。

3.6 多提示集成

不同的模板可能各有优势，因为它们分别关注特定的设计考量，从而对预训练语言模型中的语言知识和语义知识产生不同的利用方式。多提示集成融合各个提示的预测结果，以提升最终决策 [12, 15]。由于我们事先不知道哪个模板更好，因此简单地将每个提示中肯定答案词的概率相加，作为最终的排名分数：

$$\hat{y} = \sum_{e \in \mathcal{E}} P_e, \tag{5}$$

其中， P_e 是模板 e 对 w_{pos} 的输出概率，E 是融合模板的集合。本文考虑两种多提示集成方式。一种是融合来自相同

表 2：数据集统计信息。

	# 用户	# 新闻	# 点击量	# 曝光
MIND	94,057	65,238	347,727	230,117

类型模板的预测，其中 $\mathcal{E}$ ={相关性、情感、行为、效用}。也就是说，离散集成是融合四个离散模板的预测。另一种是融合来自不同类型模板的预测，称为跨类型集成。

4 实验设置

本节介绍实验设置，包括数据集、参数设置、评估指标和基线。

4.1 数据集

我们在公开的真实世界新闻推荐基准数据集 MIND[1] [52]，上进行实验，该数据集通过曝光记录用户行为。曝光日志记录了特定时间展示给用户的已点击和未点击新闻，以及该次曝光之前用户的历史新闻点击行为。MIND 收集了 2019年10月12日至11月22日 Microsoft News 平台上的用户行为日志。按照以往工作 [2, 33], 的做法，我们使用前四周的用户数据构建用户历史记录，将倒数第二周的用户数据用作训练集，最后一周的用户数据用作测试集，并从训练集中提取 5% 的曝光，构成验证集。表 2 总结了数据集的统计信息。

4.2 参数设置

在实验中，我们使用 Hugging Face[2] transformers [40] 实现的基础版 BERT（12 层，bert-base-uncased） [5] 作为预训练语言模型，来测试我们的 Prompt4NR。我们采用 AdamW优化器，以学习率 lr=2e-5 训练模型。我们使用 8 块 NVIDIA RTX-A5000 GPU，以分布式数据并行方式运行模型，每块 GPU 的批量大小设为 16，相当于单块 GPU 上批量大小为 128。我们采用与其他基线相同的负采样比例 4。按照以往工作 [37, 45, 48], 的做法，我们将用户最近点击的 50 条新闻作为其点击历史。对于历史记录中的每条新闻，我们将标题最大长度设为 10 个词；对于每条候选新闻，我们将标题最大长度设为 20 个词。此外，我们采用所有展示的平均曲线下面积、平均倒数排名、归一化折损累计增益@5 和归一化折损累计增益@10 作为评估指标，这些指标在新闻推荐研究中广泛使用 [33, 37, 48]。所有超参数均在验证集上调整。我们已在 https://github.com/resistzzz/Prompt4NR 发布源代码。

4.3 基线

我们将所提出的方法与以下竞争方法进行比较，这些方法可分为神经新闻推荐模型和 BERT 增强模型：

神经新闻推荐模型： 针对新闻推荐任务，研究者专门设计了各种神经网络，包括：(1) *NPA* [42] 使用个性化注意力网络对词语和历史记录中的点击新闻进行建模，以进行新闻与用户表示学习。(2) *LSTUR* [1] 捕捉用户的短期和长期兴趣，分别通过 GRU 网络和用户嵌入实现。(3) *NRMS* [45] 使用多头自注意力学习新闻和用户表示。(4) *FIM* [37] 采用分层膨胀卷积网络，从多个粒度视角对新闻表示进行编码。

BERT增强模型： Wu等人 [48]在上述若干方法中采用BERT作为新闻编码器。我们在此复现了三个基线以供比较，分别记为*BERT-NPA*、*BERT-LSTUR*和*BERT-NRMS*，仅将它们的新闻编码器替换为BERT，而其他神经网络部分（例如用户编码器）保持不变。

[1]https://msnews.github.io/ [2] https://huggingface.co/

表 3：整体性能结果比较。

数据集		MIND			
基线		AUC	MRR	归一化折损累计增益@5	归一化折损累计增益@10
神经网络方法	NPA	64.65	30.01	33.14	39.47
	LSTUR	65.87	30.78	33.95	40.15
	NRMS	65.63	30.96	34.13	40.52
	FIM	65.34	30.64	33.61	40.16
基于预训练语言模型的方法	BERT-NPA	67.56	31.94	35.34	41.73
	BERT-LSTUR	68.28	32.58	35.99	42.32
	BERT-NRMS	68.60	32.97	36.55	42.78
Prompt4NR (BERT)		曲线下面积	平均倒数排名	归一化折损累计增益@5	归一化折损累计增益@10
离散模板	相关性	68.77	33.42	37.20	43.36
	情感	68.77	33.29	37.12	43.19
	行为	68.76	33.22	37.02	43.26
	效用	68.94	33.62	37.47	43.61
	集成	**69.34**	**33.76**	**37.71**	**43.80**
连续模板	相关性	69.25	33.72	37.75	43.79
	情感	68.76	33.51	37.39	43.47
	行为	68.58	33.37	37.17	43.30
	效用	69.10	33.96	37.91	43.92
	集成	**69.43**	**34.06**	**38.11**	**44.14**
混合模板	相关性	68.47	33.26	37.20	43.24
	情感	68.59	33.26	37.19	43.29
	行为	**69.37**	**34.02**	**37.96**	**44.00**
	效用	68.79	33.45	37.35	43.49
	集成zh	69.22	33.78	37.77	43.87
跨类型集成		**69.64**	**34.26**	**38.30**	**44.33**

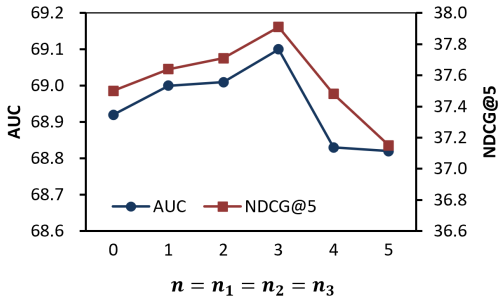
[1] 蓝色粗体表示整个表格中的最佳结果。在每种模板类型中，仅粗体表示最佳结果，第二佳结果以下划线标出。[2] 跨类型集成融合每种类型中表现最佳的结果进行决策，即$\mathcal{E}$ ={离散效用、连续效用、混合动作}。[3] 经学生 t 检验验证，这些改进具有显著性（$p < 0.01$）。

5 实验结果

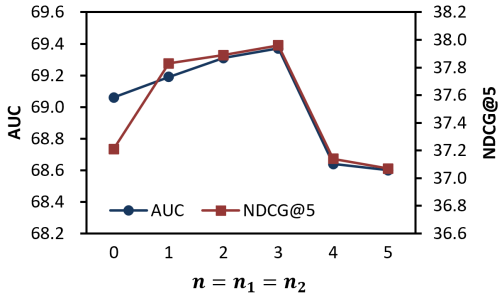
5.1 主要实验结果

表 3 总结了我们的 Prompt4NR 与这些最先进基线的比较结果。蓝色粗体表示整个表格中的最佳结果。在每种类型的模板中，粗体仅表示该类型中的最佳结果，而次佳结果则以带下划线的形式标出。从这些结果中，我们得出以下观察：

首先，使用 BERT 作为新闻编码器的模型始终优于使用浅层神经网络作为新闻编码器的神经模型，例如，BERT-NRMS 的表现优于 NRMS。这表明，预训练BERT 有助于新闻编码，而这



(a) 对 n_1, n_2, n_3 连续效用模板的影响。



(b) n_1, n_2 对混合动作 模板的影响。

图3：连续模板和混合模板中虚拟词元数量的影响。

表4：使用不同预训练语言模型的性能。

离散效用	AUC	MRR	归一化折损累计增益@5	归一化折损累计增益@10
BERT	68.94	33.62	37.47	43.61
RoBERTa	**69.20**	**34.00**	**37.77**	**43.96**
DeBERTa	69.01	33.98	37.67	43.74

归因于 BERT 在预训练过程 中学习到的丰富语义和语言知识。

其次，我们基于提示学习范式的Prompt4NR方法，无论采用离散、连续还是混合模板，几乎都优于那些基于预训练、微调范式的BERT增强模型。尤其是，平均倒数排名、归一化折损累计增益@5和归一化折损累计增益@10这几项指标分别均超过33.00、37.00和43.00，且某些模板上的曲线下面积结果超过69.00。这反映出提示学习范式在开发预训练BERT中嵌入的知识以支持新闻推荐任务方面的优越性。

第三，三种不同类型的模板取得了相近的表现，其最佳表现的排序为"离散效用 <连续效用 < 混合动作"，其中 混合动作 是所有单一类型模板中表现最好的。此外，四种离散模板之间的性能差距，相较于连续模板和混合模板之间的差距要小。这是因为连续模板和混合模板中的可学习虚拟词元带来了更多潜力，但也造成了更多变化。

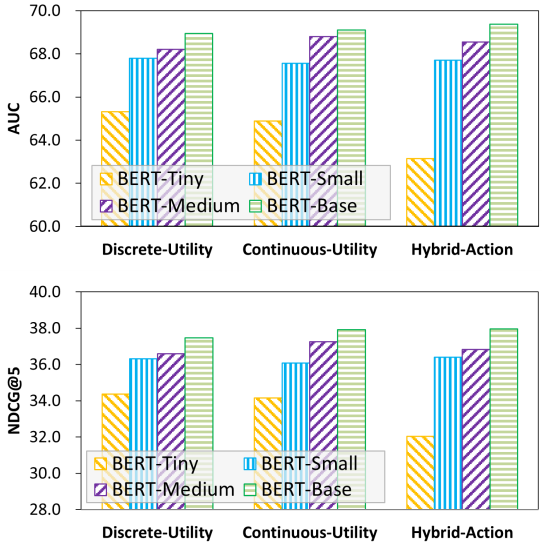


图4：BERT规模的影响。

最后，在离散模板和连续模板上，与仅使用单个提示相比，进行多提示集成显然能够提升推荐性能。这表明，提示集成能够融合多个提示，从而做出更好的决策。在混合模板上，集成结果优于相关性、情感和效用模板，但逊于行为模板。这说明，提示集成发挥作用的前提是各个被融合提示之间的性能差距不能太大。此外，跨类型集成融合了每种类型中表现最好的模板，即离散效用、连续效用 和混合动作，取得了最佳性能。这表明，提示集成不应受模板类型限制，在某些情况下，跨类型集成可能是不错的选择。

顺便一提，我们注意到，每种模板类型中排名前三的分别来自效用、效用 和行为 提示。在提示模板的四种考量因素中，行为 和效用提示与推荐任务直接相关；而相关性 和情感 提示则没有直接关联。相关性 和情感 提示分别通过推断新闻文本之间的语义相关性，以及预测用户的情绪反应，转变了兴趣匹配。我们认为，与其采用另一种底层推理方式，不如直接关联推荐属性的模板可能更适合新闻推荐任务。

5.2 超参数分析

我们的 Prompt4NR 的核心超参数是连续模板（即 n_1、n_2、n_3）和混合模板（即 n_1、n_2）中的虚拟词元数量。我们注意到，即使这些参数的取值范围很小，它们也有很多种不同的组合。我们采用粗粒度的超参数调优策略，即让 n_1、n_2、n_3 取相同的值进行调优，记为 $n = n_1 = n_2 = n_3$。

具体而言，我们在 {0、1、2、3、4、5} 中调整 n，其中我们注意到，不含任何词元的 $n = 0$ 称为空提示（NullPrompt）[21]。考虑到论文篇幅，我们仅展示表现最好的两种方法的调优结果，即 连续效用和混合动作，其他方法的趋势也类似。图 3 展示了不同 n取值下的推荐性能。我们发现，连续模板和混合模板的趋势相同：随着 n 增加，性能先提升后下降。这表明，提示只需少量虚拟词元即可：数量太少可能不足以引导预训练语言模型组织信息；而数量太多则可能导致更多歧义，因为随机初始化的虚拟词元缺乏预训练知识。此外，我们注意到，即使使用空提示（即 $n = 0$），曲线下面积也能达到 68.9，超过了这些 BERT 增强方法。这再次体现了提示学习范式在新闻推荐任务上的优势，它能够更好地利用预训练语言模型，捕捉新闻文本之间的潜在关系。

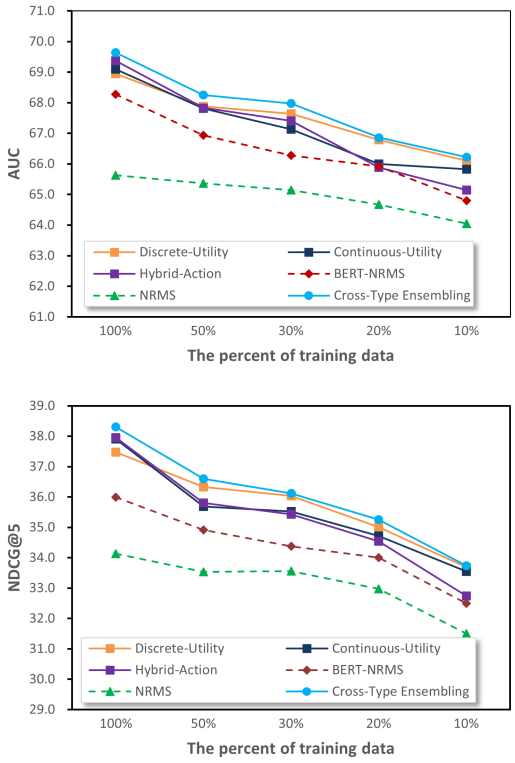
5.3 预训练语言模型的影响

5.3.1 预训练语言模型的规模影响。 我们研究了使用不同规模 BERT 的影响，即 BERT-Tiny（2层）、BERT-Small（4层）、BERT-Medium（8层）和 BERT-Base（12层）。我们在每种类型中表现最佳的模板上分别使用不同规模的 BERT。图4展示了曲线下面积和归一化折损累计增益@5的结果。我们观察到，使用参数更多、层数更深的大型预训练语言模型通常能够带来更好的推荐性能。这并不出乎意料，因为规模更大的预训练语言模型通常具有更强的能力来捕捉新闻文本深层次的语言和语义信息，而提示学习的性能也取决于预训练语言模型具备的这种能力。我们合理推测，使用更大的 BERT（例如具有24层的 BERT-Large）可以进一步提升性能。然而，我们认为应在推荐性能与模型规模之间取得平衡，因为过大的预训练语言模型可能不适合实际应用场景。

5.3.2 不同预训练语言模型的影响。 我们将BERT替换为另外两种流行的预训练语言模型，即 RoBERTa [20] 和 DeBERTa [8], ，以研究使用不同预训练语言模型的影响。RoBERTa 和 DeBERTa都是 BERT[3]的改进版本。RoBERTa移除了下一句预测，并在更大的语料库上以更大的批量大小进行预训练；DeBERTa引入了解耦注意力机制和增强的掩码解码器，以改进 BERT。表 4 展示了实验结果。我们观察到，RoBERTa和 DeBERTa的表现都优于 BERT，且RoBERTa 取得了最佳结果。这表明，下游新闻推荐任务能够受益于预训练过程中的改进。

5.4 少样本学习的性能

一些研究者报告称，在某些自然语言处理任务中，例如文本分类 [36], 隐式话语关系识别 [53]，提示学习范式在少样本场景下具有一定的鲁棒性。我们也希望考察 Prompt4NR 及其竞争方法在少样本学习场景下的表现。我们采用



Fig图 5：少样本学习的性能比较

学习。

下采样逐步缩减训练集，同时保持验证集和测试集不变。

图5总结了少样本学习的性能比较，其中虚线表示两个竞争者NRMS 和BERT-NRMS，实线表示我们的Prompt4NR 的三个单模板和一种跨类型集成。毫不意外的是，随着训练数据的减少，我们的 Prompt4NR 和竞争者的性能都会下降。我们观察到，NRMS的曲线始终处于最低位置；在大多数情况下，我们三个单模板的曲线都高于 BERT-NRMS 的曲线；交叉集成的曲线高于其他所有曲线。结果表明，与浅层神经网络相比，预训练语言模型具有更强的处理少样本场景的能力，即提示学习比标准预训练与微调范式更加稳健。提示集成策略可以进一步提升稳健性。此外，我们观察到，在减少训练集时，相较于离散效用 ，连续效用 和 混合动作更常处于优势位置。这可能是因为随着训练数据的减少，连续模板和混合模板中的虚拟词元没有得到充分训练。

5.5 Visualization

为了考察 BERT 在我们的 Prompt4NR框架中学到了什么，我们将图6中一个随机样例的 BERT 注意力权重可视化。由于 BERT 有 12 层，且每层有 12个注意力头，我们展示了第一层和

[3]BERT、RoBERTa 和 DeBERTa 均采用包含 12层 的基础版本。它们在 Hugging Face 上的模型名称为 {bert-base-uncased, roberta-base, deberta-base}。

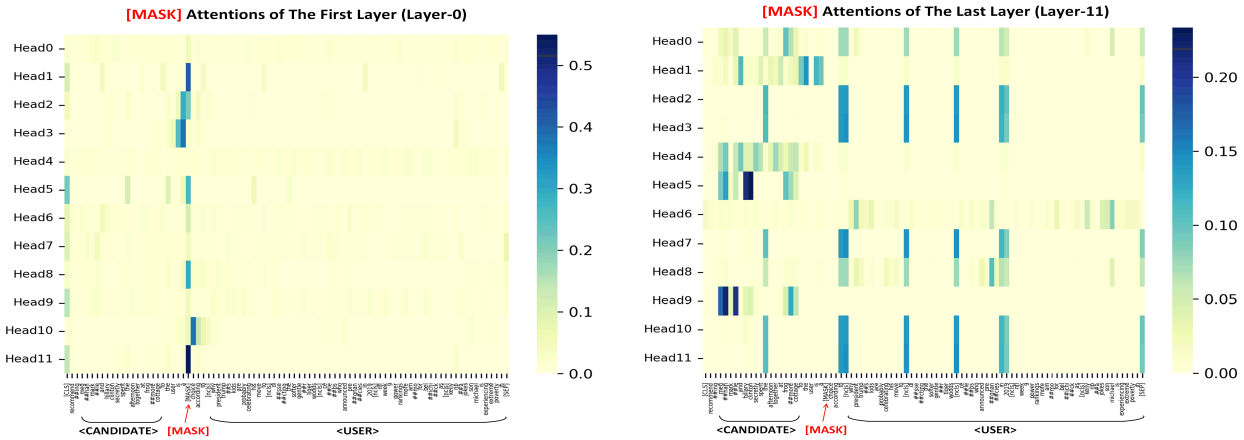


图6：某个抽样实例中，BERT 的 12个注意力头对 [MASK] 标记的注意力权重可视化：左侧为第一层的注意力权重，右侧为最后一层。*X*轴表示输入句子，*Y*轴表示 BERT 的 12 个注意力头；BERT 使用离散效用 模板进行训练。

最后一层的 [MASK] 注意力权重，以帮助我们理解学习过程。在第一层中，注意力主要集中在 [MASK] 周围区域的这些词元上。在最后一层中，我们观察到不同注意力头的关注点各不相同。一些注意力头仍然聚焦于 [MASK]，周围的区域，例如 Head1。一些注意力头更加关注 <CANDIDATE> 所在的区域，例如 Head4、Head5 和 Head9。还有一些注意力头更加关注 <USER> 所在的区域，其中 {Head6, Head8} 和 {Head2, Head3, Head7, Head10, Head11} 分别关注不同位置；Head6 和 Head8 广泛关注 <USER> 中的特定词元，我们称之为词元级注意力；而 Head2、Head3、Head7、Head10 和 Head11 则特别关注若干 [NCLS] 词元周围的区域，我们称之为新闻级注意力，因为每条历史新闻开头的虚拟词元 [NCLS] 可能能够表示其后续新闻。Head0 的注意力则分散在整个句子中。总之，与第一层相比，最后一层的注意力点更加明确，且每个注意力头都发挥了作用。这表明，经过在我们的 Prompt4NR框架中训练后，BERT 能够有选择地利用可用信息进行 [MASK] 预测和推荐。

5.6 案例研究

我们进行了案例研究，以进一步直观地阐明 Prompt4NR 的有效性。图 7 展示了某次曝光中，由四种单一的离散-{相关性、情感、行为、效用}模板以及一种离散集成构成的前 5 条推荐列表。在这次曝光中，包含用户*U56800* 点击过的 39 条历史新闻，以及用于排序的 60 条候选新闻[4]。为便于直观呈现，我们标注了新闻的类别标签。受篇幅所限，未详细展示全部 39 条历史新闻，但我们手动绘制了一个粗略的用户画像，其中包括类别分布（即饼图）和其历史新闻中的实体频率分布（即词云中，频率越高，字号越大）。通过这个案例，我们得出以下观察结果：

[4]案例研究中呈现的这些标识符可以直接与原始数据集匹配，例如 *Uxxxx*、*Nxxxx*。

该用户点击了涵盖10个类别的新闻。从该用户点击实体的词云中，诸如“梅根，萨塞克斯公爵夫人，哈里王子，萨塞克斯公爵，剑桥公爵夫人”之类的词语经常出现，我们可以直观地猜测，王室生活可能是该用户感兴趣的领域之一，但并非唯一兴趣。换言之，该用户的兴趣领域多种多样。此外，这些候选新闻来自多个领域，涵盖14个类别（因篇幅限制未绘制），并涉及众多不同的实体。这些观察表明，为该用户做出准确推荐可能并非易事。

虽然不容易看出，但我们发现，五个推荐列表都包含用户实际点击的真实新闻，简称真实新闻。我们注意到，每个模板都有各自的推荐列表风格。例如，离散相关性 命中的真实新闻更多，但多样性较低。离散情感 推荐了五条属于五个不同类别的新闻，其中甚至包括用户历史记录中从未出现过的“健康”新闻。离散行为 和离散效用 都命中了两条真实新闻：但离散行为 倾向于用户对皇室生活方式的兴趣，而离散效用 则发现了用户对“娱乐”新闻的另一处兴趣点。这些结果印证了我们的动机：尽管不同模板看起来只有少许差异，但它们可能关注预训练语言模型中蕴含的不同知识。此外，我们观察到，与每个单独模板相比，离散集成 实现了双赢。具体而言，离散集成同样命中了三条真实新闻，而且这些新闻的排序位置更有利：推荐的五条新闻涵盖四个类别，为用户提供了多样化的新闻。这表明，多提示集成能够融合每个独立提示的优势，从而提升推荐效果。

顺便提一下，我们注意到一个有趣的现象。第 20 条历史新闻 *N31748* 是关于电影的“预告片——《霹雳娇娃》”。在 离散行为推荐的前 5 条新闻中，有一条新闻*N58656* “《霹雳娇娃》主演：她们如今身在何处？”，与 *N31748*非常相关。尽管*N58656* 并非真正的新闻，我们仍认为这是一个不错的推荐，可以引导

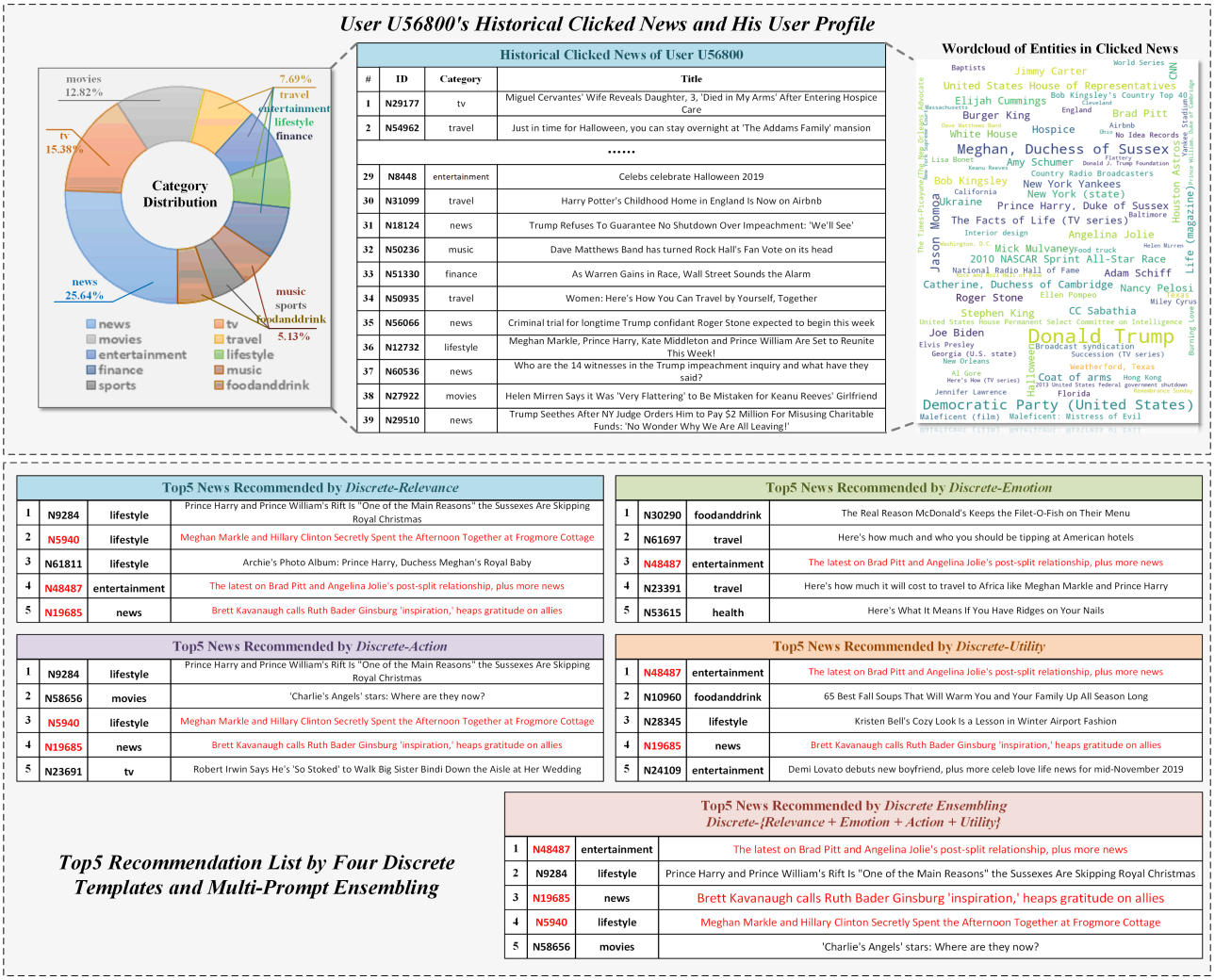


图7：四个Discrete-{相关性、情感、行为、效用}模板和离散集成所推荐的前5条新闻案例研究，展示了用户 *U56800* 的情况。用户实际点击的推荐新闻以红色突出显示。顶部是用户 *U56800* 历史记录中点击的新闻和手工制作的用户画像，包括其历史记录新闻的类别分布和实体词云。底部展示了五个推荐列表。

用户回顾自己的潜在兴趣并点击。这体现了 Prompt4NR 自行发现并串联新闻之间潜在关系的能力。

此外，我们观察到，这四个单模板并不能完美地充分挖掘用户的兴趣。例如，这里没有推荐体育新闻，但用户实际上点击了一条体育新闻*N40094*，其标题为"Baker Mayfield quick to condemn teammate Myles Garrett for brawl,attackon Mason Rudolph"。这表明我们的四个单模板不足以涵盖所有推荐考量，这促使我们在未来的工作中设计更加出色的模板。

6 结论

在本文中，我们提出了用于新闻推荐的 Prompt4NR 提示学习框架，该框架将将新闻推荐任务视为一种用于 [MASK] 预测任务的完形填空任务。作为首项探索性工作，我们针对一系列不同类型的提示模板开展了大量实验，包括离散、连续和混合模板。此外，我们采用了多提示集成，以结合不同提示的优势。实验结果验证了我们的 Prompt4NR 优于最先进的竞争方法。

7 致谢

本研究部分得到了国家自然科学基金（项目编号：62172167）的资助。计算工作在华中科技大学高性能计算平台上完成。

参考文献

[1] Mingxiao An、Fangzhao Wu、Chuhan Wu、Kun Zhang、Zheng Liu 和 Xing Xie。2019。基于用户长期与短期表征的神经新闻推荐。载于第57届计算语言学协会年会论文集。336–345。[2] Qiwei Bi、Jian Li、Lifeng Shang、Xin Jiang、Qun Liu 和 Hanfang Yang。2022。MTRec：基于 BERT 的新闻推荐多任务学习。载于计算语言学协会论文集：ACL 2022 发现。2663–2669。[3] Lu Dai、Bang Wang、Wei Xiang 和 Yijun Mo。2022。用于事件论元抽取的双向迭代提示调优。arXiv 预印本 arXiv:2210.15843（2022）。[4] Abhinandan S Das、Mayur Datar、Ashutosh Garg 和 Shyam Rajaram。2007。Google 新闻个性化：可扩展的在线协同过滤。载于第16届万维网国际会议论文集。271–280。[5] Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。2018。BERT：用于语言理解的深度双向变换器预训练。arXiv 预印本 arXiv:1810.04805（2018）。[6] Shijie Geng、Shuchang Liu、Zuohui Fu、Yingqiang Ge 和 Yongfeng Zhang。2022。将推荐视为语言处理（RLP）：统一的预训练、个性化提示与预测范式（P5）。载于第16届 ACM 推荐系统会议论文集。299–315。[7] Shansan Gong 和 Kenny Q. Zhu。2022。正面、负面与中性：基于会话的新闻推荐中的隐式反馈建模。载于第45届国际 ACM SIGIR 信息检索研究与发展会议论文集。1185–1195。[8] Pengcheng He、Xiaodong Liu、Jianfeng Gao 和 Weizhu Chen。2021。DeBERTa：采用解耦注意力的解码增强型 BERT。载于国际学习表征会议。https://openreview.net/forum? id=XPZIaotutsD[9] Qinglin Jia、Jingjie Li、Qi Zhang、Xiuqiang He 和 Jieming Zhu。2021。RMBERT：基于 BERT 的循环推理记忆网络新闻推荐。载于第44届国际 ACM SIGIR 信息检索研究与发展会议论文集。1773–1777。[10] Dhruv Khattar、Vaibhav Kumar、Vasudeva Varma 和 Manish Gupta。2018。Weave&rec：一种基于词嵌入的三维卷积网络，用于新闻推荐。载于第27届 ACM 信息与知识管理国际会议论文集。1855–1858。[11] Taeho Kim、Yungi Kim、Yeon-Chang Lee、Won-Yong Shin 和 Sang-Wook Kim。2022。只看标题就够了吗？利用正文文本丰富标题词，实现精准新闻推荐。载于第31届 ACM 信息与知识管理国际会议论文集。4138–4142。[12] Brian Lester、Rami Al-Rfou 和 Noah Constant。2021。参数高效提示调优的规模优势。载于2021年自然语言处理中的经验方法会议论文集。3045–3059。[13] Jian Li、Jieming Zhu、Qiwei Bi、Guohao Cai、Lifeng Shang、Zhenhua Dong、Xin Jiang 和 Qun Liu。2022。MINER：用于新闻推荐的多兴趣匹配网络。载于计算语言学协会论文集：ACL 2022 发现。343–352。[14] Xiang Lisa Li 和 Percy Liang。2021。前缀调优：优化用于生成的连续提示。载于第59届计算语言学协会年会暨第11届自然语言处理国际联合会议论文集（第1卷：长篇论文）。4582–4597。[15] Pengfei Liu、Weizhe Yuan、Jinlan Fu、Zhengbao Jiang、Hiroaki Hayashi 和 Graham Neubig。2021。预训练、提示与预测：自然语言处理提示方法的系统综述。arXiv 预印本 arXiv:2107.13586（2021）。[16] Rui Liu、Huilin Peng、Yong Chen 和 Dell Zhang。2020。HyperNews：通过双任务深度神经网络同时进行新闻推荐与活跃时间预测。载于 IJCAI。3487–3493。[17] Shenghao Liu、Bang Wang 和 Minghua Xu。2017。基于图随机游走和历史偏好重排序的事件推荐。载于第40届国际 ACM SIGIR 信息检索研究与发展会议论文集。861–864。[18] Xiao Liu、Kaixuan Ji、Yicheng Fu、Zhengxiao Du、Zhilin Yang 和 Jie Tang。2021。P-tuning v2：提示调优在不同规模和任务中普遍可与微调媲美。arXiv 预印本 arXiv:2110.07602（2021）。[19] Xiao Liu、Yanan Zheng、Zhengxiao Du、Ming Ding、Yujie Qian、Zhilin Yang 和 Jie Tang。2021。GPT 也能理解。arXiv 预印本 arXiv:2103.10385（2021）。[20] Yinhan Liu、Myle Ott、Naman Goyal、Jingfei Du、Mandar Joshi、Danqi Chen、Omer Levy、Mike Lewis、Luke Zettlemoyer 和 Veselin Stoyanov。2019。RoBERTa：一种鲁棒优化的 BERT 预训练方法。arXiv 预印本 arXiv:1907.11692 （ 2019 ）。[21] Robert Logan IV、Ivana Balazevic、Eric Wallace、Fabio Petroni、Sameer Singh 和 Sebastian Riedel。2022。精简提示与参数：使用语言模型进行简单的少样本学习。载于计算语言学协会论文集：ACL 2022 发现。2824–2835。[22] Ilya Loshchilov 和 Frank Hutter。2019。解耦权重衰减正则化。载于第七届国际学习表示会议（ICLR 2019），美国路易斯安那州新奥尔良，2019年5月6日至9日。OpenReview.net。[23] Tomas Mikolov、Kai Chen、Greg Corrado 和 Jeffrey Dean。2013。高效估计向量空间中的词表示。arXiv 预印本 arXiv:1301.3781（2013）。[24] Yijun Mo、Bixi Li、Bang Wang、Laurence T Yang 和 Minghua Xu。2018。基于反向随机游走和参与者规模控制的社交网络事件推荐。《未来一代计算机系统》79（2018），383–395。[25] Shumpei Okura、Yukihiro Tagami、Shingo Ono 和 Akira Tajima。2017。面向数百万用户的基于嵌入的新闻推荐。载于第23届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集。1933–1942。[26] Keunchan Park、Jisoo Lee 和 Jaeho Choi。2017。用于新闻推荐的深度神经网络。载于2017年 ACM 信息与知识管理会议论文集。2255–2258。[27] Jeffrey Pennington、Richard Socher 和 Christopher D Manning。2014。GloVe：词表示的全局向量。载于2014年自然语言处理中的经验方法会议（EMNLP）论文集。1532–1543。[28] Fabio Petroni、Tim Rocktäschel、Sebastian Riedel、Patrick Lewis、Anton Bakhtin、Yuxiang Wu 和 Alexander Miller。2019。语言模型能否作为知识库？载于2019年自然语言处理中的经验方法会议和第九届自然语言处理国际联合会议（EMNLP-IJCNLP）论文集。2463–2473。[29] Tao Qi、Fangzhao Wu、Chuhan Wu 和 Yongfeng Huang。2021。PP-Rec：结合个性化用户兴趣与时间感知新闻热度的新闻推荐。载于计算语言学协会第59届年会和第11届自然语言处理国际联合会议论文集。5457–5467。[30] Tao Qi、Fangzhao Wu、Chuhan Wu 和 Yongfeng Huang。2022。FUM：面向新闻推荐的细粒度快速用户建模。载于第45届国际 ACM SIGIR 信息检索研究与发展会议论文集。1974–1978。[31] Tao Qi、Fangzhao Wu、Chuhan Wu 和 Yongfeng Huang。2022。FUM：面向新闻推荐的细粒度快速用户建模。载于第45届国际 ACM SIGIR 信息检索研究与发展会议论文集。1974–1978。[32] Tao Qi、Fangzhao Wu、Chuhan Wu 和 Yongfeng Huang。2022。结合候选项感知用户建模的新闻推荐。载于第45届国际 ACM SIGIR 信息检索研究与发展会议论文集。1917–1921。[33] Tao Qi、Fangzhao Wu、Chuhan Wu、Peiru Yang、Yang Yu、Xing Xie 和 Yongfeng Huang。2021。HieRec：用于个性化新闻推荐的层次化用户兴趣建模。载于计算语言学协会第59届年会和第11届自然语言处理国际联合会议论文集。5446–5456。[34] Timo Schick 和 Hinrich Schütze。2021。利用完形填空问题进行少样本文本分类和自然语言推理。载于计算语言学协会欧洲分会第16届会议主会场论文集。255–269。[35] Taylor Shin、Yasaman Razeghi、Robert L. Logan IV、Eric Wallace 和 Sameer Singh。 [n. d.]。AutoPrompt：利用自动生成的提示从语言模型中引出知识。载于2020年自然语言处理中的经验方法会议（EMNLP）论文集。4222–4235。[36] Chengyu Wang、Jianing Wang、Minghui Qiu、Jun Huang 和 Ming Gao。2021。TransPrompt：面向少样本文本分类的自动化可迁移提示框架。载于2021年自然语言处理中的经验方法会议论文集。2792–2802。[37] Heyuan Wang、Fangzhao Wu、Zheng Liu 和 Xing Xie。2020。用于神经新闻推荐的细粒度兴趣匹配。载于计算语言学协会第58届年会论文集。836–845。[38] Hongwei Wang、Fuzheng Zhang、Xing Xie 和 Minyi Guo。2018。DKN：用于新闻推荐的深度知识感知网络。载于2018年万维网会议论文集。1835–1844。[39] Jingkun Wang、Yipu Chen、Zichun Wang 和 Wen Zhao。2021。通过多视角兴趣表示增强新闻推荐中的热度建模。载于第30届 ACM 信息与知识管理国际会议论文集。1949–1958。[40] Thomas Wolf、Lysandre Debut、Victor Sanh、Julien Chaumond、Clement Delangue、Anthony Moi、Pierric Cistac、Tim Rault、Rémi Louf、Morgan Funtowicz 等。2020。Transformers：最先进的自然语言处理技术。载于2020年自然语言处理中的经验方法会议系统演示论文集。38–45。[41] Chuhan Wu、Fangzhao Wu、Mingxiao An、Jianqiang Huang、Yongfeng Huang 和 Xing Xie。2019。基于注意力的多视角学习神经新闻推荐。载于第28届人工智能国际联合会议论文集

[42] Chuhan Wu、Fangzhao Wu、Mingxiao An、Jianqiang Huang、Yongfeng Huang 和 Xing Xie。2019。NPA：采用个性化注意力的神经新闻推荐。载于第 *25* 届 *ACM SIGKDD* 知识发现与数据挖掘国际会议论文集。2576–2584。 [43] Chuhan Wu、Fangzhao Wu、Mingxiao An、Yongfeng Huang 和 Xing Xie。2019。采用主题感知新闻表示的神经新闻推荐。载于第 *57* 届计算语言学协会年会论文集。1154–1159。[44] Chuhan Wu、Fangzhao Wu、Mingxiao An、Tao Qi、Jianqiang Huang、Yongfeng Huang 和 Xing Xie。2019。采用异构用户行为的神经新闻推荐。载于*2019* 年自然语言处理中的经验方法会议暨第 *9* 届自然语言处理国际联合会议（*EMNLP-IJCNLP*）论文集。4874–4883。 [45] Chuhan Wu、Fangzhao Wu、Suyu Ge、Tao Qi、Yongfeng Huang 和 Xing Xie。2019。采用多头自注意力的神经新闻推荐。载于*2019* 年自然语言处理中的经验方法会议暨第 *9* 届自然语言处理国际联合会议（*EMNLP-IJCNLP*）论文集。6389–6394。[46] Chuhan Wu、Fangzhao Wu、Yongfeng Huang 和 Xing Xie。2022。个性化新闻推荐：方法与挑战。*ACM*信息系统汇刊（*TOIS*）（2022）。 [47] Chuhan Wu、Fangzhao Wu、Tao Qi 和 Yongfeng Huang。2020。用于新闻推荐的点击偏好与阅读满意度用户建模。载于*IJCAI*。3023–3029。 [48] Chuhan Wu、Fangzhao Wu、Tao Qi 和 Yongfeng Huang。2021。利用预训练语言模型增强新闻推荐。载于第 *44* 届国际 *ACM SIGIR* 信息检索研究与开发会议论文集。1652–1656。[49] Chuhan Wu、Fangzhao Wu、Tao Qi 和 Yongfeng Huang。2022。一石二鸟：新闻推荐中兼顾召回与排序的统一模型学习。载于计算语言学协会论文集：*ACL 2022*。3474–3480。[50] Chuhan Wu、Fangzhao Wu、Tao Qi、Qi Liu、Xuan Tian、Jie Li、Wei He、Yongfeng Huang 和 Xing Xie。2022。Feedrec：结合多种用户反馈的新闻馈送推荐。载于*2022* 年 *ACM Web* 会议论文集。2088–2097。

[51] Chuhan Wu、Fangzhao Wu、Tao Qi、Chao Zhang、Yongfeng Huang 和 Tong Xu。2022。MM-Rec：视觉语言模型赋能的多模态新闻推荐。载于第 *45* 届国际 *ACM SIGIR* 信息检索研究与发展会议论文集。2560–2564。[52] Fangzhao Wu、Ying Qiao、Jiun-Hung Chen、Chuhan Wu、Tao Qi、Jianxun Lian、Danyang Liu、Xing Xie、Jianfeng Gao、Winnie Wu 等。2020。MIND：用于新闻推荐的大规模数据集。载于第 *58* 届计算语言学协会年会论文集。3597–3606。[53] Wei Xiang、Zhenglin Wang、Lu Dai 和 Bang Wang。2022。ConnPrompt：用于隐式话语关系识别的连接词填空提示学习。载于第 *29* 届国际计算语言学会议论文集。902–911。[54] Shitao Xiao、Zheng Liu、Yingxia Shao、Tao Di、Bhuvan Middha、Fangzhao Wu 和 Xing Xie。2022。利用预训练语言模型训练大规模新闻推荐系统。载于第 *28* 届 *ACM SIGKDD* 知识发现与数据挖掘会议论文集。4215–4225。[55] Yang Yu、Fangzhao Wu、Chuhan Wu、Jingwei Yi、Tao Qi 和 Qi Liu。2021。Tiny-NewsRec：高效且有效的基于预训练语言模型的新闻推荐。*arXiv* 预印本 *arXiv:2112.00944* （2021）。[56] Hui Zhang 、 Xu Chen 和 Shuai Ma。2019。利用层次注意力网络进行动态新闻推荐。载于*2019* 年 *IEEE* 国际数据挖掘会议（*ICDM*）。IEEE，1456–1461。[57] Qi Zhang、Qinglin Jia、Chuyuan Wang、Jingjie Li、Zhaowei Wang 和 Xiuqiang He。2021。Amm：用于新闻推荐的注意力多字段匹配。载于第 *44* 届国际 *ACM SIGIR* 信息检索研究与发展会议论文集。1588–1592。[58] Qi Zhang、Jingjie Li、Qinglin Jia、Chuyuan Wang、Jieming Zhu、Zhaowei Wang 和 Xiuqiang He。2021。UNBERT：用于新闻推荐的用户-新闻匹配 BERT。载于*IJCAI*。3356–3362。[59] Xuanyu Zhang、Qing Yang 和 Dongliang Xu。2021。结合显式实体图与隐式文本信息进行新闻推荐。载于*2021* 年万维网大会伴随论文集。412–416。[60] Qiannan Zhu、Xiaofei Zhou、Zeliang Song、Jianlong Tan 和 Li Guo。2019。Dan：用于新闻推荐的深度注意力神经网络。载于*AAAI* 人工智能大会论文集，第 33 卷。5973–5980。