

# 基于深度学习的新闻推荐算法研究综述\*

田 萱<sup>1,2+</sup>, 丁 琪<sup>1,2</sup>, 廖子慧<sup>1,2</sup>, 孙国栋<sup>1,2</sup>

1. 北京林业大学 信息学院, 北京 100083

2. 国家林业草原林业智能信息处理工程技术研究中心, 北京 100083

+ 通信作者 E-mail: tianxuan@bjfu.edu.cn

**摘 要:** 新闻推荐可以有效缓解新闻信息过载, 是当今人们获取新闻资讯的重要方式, 而深度学习成为近年来促进新闻推荐发展的主流技术, 使新闻推荐的效果得到显著提升, 受到研究者的广泛关注。主要对基于深度学习的新闻推荐方法研究现状进行分类梳理和分析归纳。根据对新闻推荐的核心对象——用户和新闻的建模思路不同, 将基于深度学习的新闻推荐方法分为“两段式”方法、“融合式”方法和“协同式”方法三类。在每类方法中, 根据建模过程中的具体子任务或基于的数据组织结构进行更进一步细分, 对代表性模型进行分析介绍, 评价其优点和局限性等, 并详细总结每类方法的特点和优缺点。另外还介绍了新闻推荐中常用数据集、基线算法和性能评价指标, 最后文末分析展望了该领域未来可能的研究方向及发展趋势。

**关键词:** 新闻推荐; 深度学习; 用户兴趣建模; 新闻建模

**文献标志码:** A **中图分类号:** TP311

## Survey on Deep Learning Based News Recommendation Algorithm\*

TIAN Xuan<sup>1,2+</sup>, DING Qi<sup>1,2</sup>, LIAO Zihui<sup>1,2</sup>, SUN Guodong<sup>1,2</sup>

1.School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China

2.Engineering Research Center for Forestry-oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China

**Abstract:** News recommendation (NR) can effectively alleviate the overload of news information, and it is an important way to obtain news information for users. Deep learning (DL) has become a mainstream technology to promote the development of NR in recent years, and the effect of news recommendation has been significantly improved. In this paper, the methods of deep learning-based news recommendation (DNR) are classified, analyzed and summarized. In the research of NR, modeling users or news are two key tasks. According to the different strategies

**基金项目:** 国家重点研发计划 (2018YFC1603302, 2018YFC1603305)。

This work was supported by the National Key Research and Development Program of China (2018YFC1603302, 2018YFC1603305).

of modeling users or news, the news recommendation methods based on deep learning are proposed to be divided into three types: "two-stage" method, "fusion" method and "collaboration" method. Each type of method is further subdivided in terms of sub-tasks or model structures. The representative models of each method are introduced and analyzed, and its advantages and limitations are evaluated. The characteristics, advantages and disadvantages of each type of method are also summarized in detail. Furthermore, the commonly used datasets, baseline and performance evaluation indicators are introduced. Finally, the possible future research directions and development trends in this field are analyzed and predicted.

**Key words:** news recommendation; deep learning; user interest modeling; news modeling

新闻推荐 (News Recommendation, 简称 NR) 是一种缓解新闻信息过载的有效方式, 能帮助用户从众多的新闻资源中筛选并推送其可能感兴趣的新闻<sup>[1, 2]</sup>。目前, 各大新闻 APP 如今日头条和新闻网站如 MSN News 等都在其商业系统中嵌入了推荐算法, 以提供千人千面的新闻推荐, 提升用户体验。例如今日头条从新闻内容特征、用户个人特征与所处环境特征三个维度构建新闻推荐算法, 并成功应用于其平台<sup>[3, 4]</sup>。相比于商品、电影、音乐等其他推荐, NR 具有新闻时效性强、语义信息丰富、用户兴趣变化多样等特点, 在用户整体偏好建模、用户时序兴趣建模、新闻建模等子任务方面面临着挑战。近年来 NR 成为国内外众多科研机构研究的热点, 获得了信息检索、数据挖掘和人工智能等领域国内外学术会议的高度关注<sup>[5, 6]</sup>。

深度学习 (Deep Learning, 简称 DL) 是一种表征学习 (Representation Learning) 方法<sup>[7]</sup>, 已广泛应用于自然语言处理、图像处理和语音识别等领域。基于深度学习的新闻推荐 (Deep learning based News Recommendation, 简称 DNR) 研究与日俱增, 成为当前推荐系统研究领域热点之一。目前已经有多篇相关综述性论文涉及调研 DNR 技术, 如文献[8, 9]从深度学习技术角度剖析了推荐算法, 多处提及到前沿的 DNR 技术; 文献[10, 11]从个性化新闻推荐框架角度介绍了几种基于深度学习的新闻推荐模

型等。然而, 这些文献都缺乏对 DNR 的系统性、全面性分析, 未能表达出深度学习模型针对新闻推荐任务的设计特色和方法特征。本文的主要贡献在于从深度学习技术角度全面、系统论述 DNR 研究进展, 尤其着重从新闻推荐的核心对象——用户和新闻出发剖析深度学习技术建模特色和处理特征。

按照建模思路的不同, 我们将 DNR 方法分为“两段式”方法、“融合式”方法和“协同式”方法三类分别进行介绍, 其中, “两段式”方法强调按照“先新闻、后兴趣”的思路, 分步建立推荐模型; “融合式”方法是将新闻和用户信息混合在一起, 同时学习新闻和用户特征; “协同式”方法则是不止利用了用户自己的新闻点击行为, 还考虑了相邻用户与新闻的交互。在每一类方法里, 或根据建模过程中的具体子任务不同, 亦或根据其基于的数据组织结构不同, 从“建模新闻”、“建模用户整体偏好”、“建模用户时序兴趣”、“用户-新闻二分图”等多个角度分别剖析不同 DNR 技术的建模特点和设计特色。其分类如图 1 所示。

全文内容安排如下: 第 1 节介绍 NR 早期研究和 DNR 的相关背景; 第 2 节重点分析讨论三类 DNR 方法, 并根据 DNR 应用任务等不同进一步介绍每类方法的基本思想、设计机制等; 第 3 节介绍常用公共数据集、基线算法和算法性能评价指标等; 第 4 节总结讨论 DNR 面临的挑战和未来发展趋势。

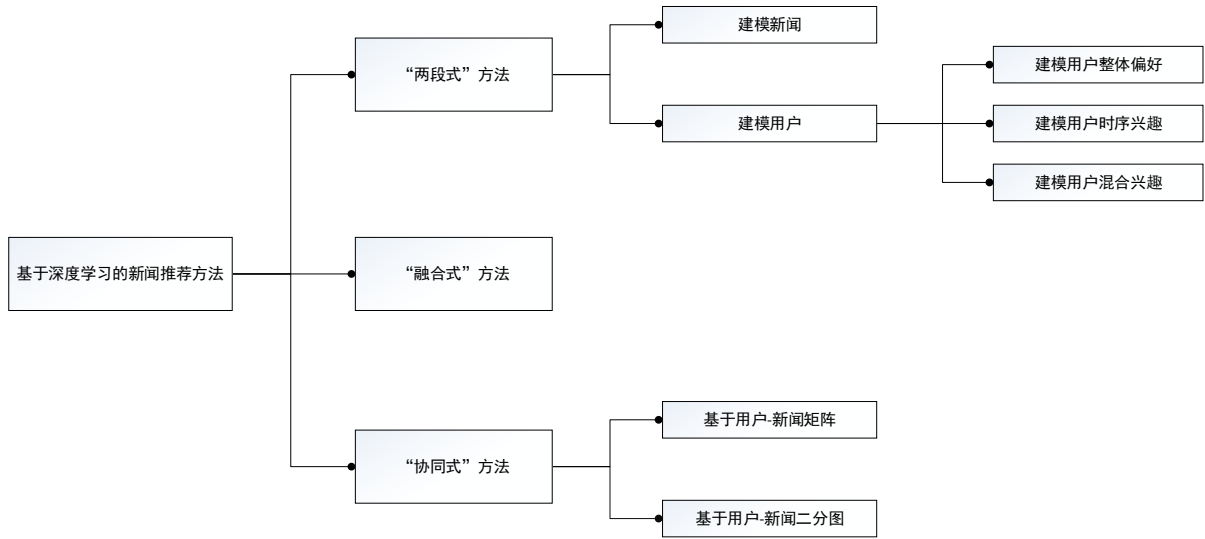


Fig.1 A taxonomy of DNR

图 1 基于深度学习的新闻推荐方法分类

## 1 早期研究和 DNR 相关背景介绍

如今，网络新闻成为获取实时资讯最流行的方式，用户可以通过专业的新闻网站阅读新闻，也可以通过其他的社会化网络服务渠道（如 Twitter、微博等）浏览新闻资讯。NR 的任务是从众多新闻资讯中为用户推送出其可能感兴趣的新闻资源，有效过滤无关新闻，尽可能满足不同用户对新闻资讯的需求，提高用户阅读质量。当前，网络传播具有传播广泛、即时性强、信息量大和交互灵活等特点，以网络传播为主的新闻推荐面临着新闻数量爆炸增长、新闻时效性强、语义信息丰富<sup>[12]</sup>和用户兴趣动态变化等挑战。早期新闻推荐方法等往往采用手工构造特征，依赖于贝叶斯模型<sup>[13]</sup>等浅层模型进行预测，无法提取新闻及用户的深层次特征，限制了 NR 的有效性可扩展性。

DL 是机器学习中一种基于数据表征学习的方法，在自然语言处理、语音识别、图像处理等领域广泛使用。新闻推荐中，DL 技术能够通过学习非线性网络结构，表征新闻和用户相关的海量数据<sup>[9]</sup>，

提取新闻内容和用户兴趣的深层次特征<sup>[14]</sup>，可应对 NR 面临的新闻数目巨大和用户兴趣动态变化<sup>[15]</sup>等挑战，有效提高推荐的准确率。此外，DL 技术能够融合蕴含丰富用户兴趣的多源异构辅助性信息，学习不同数据的统一表示<sup>[16]</sup>，缓解新闻推荐中冷启动、数据稀疏等问题。主流的 DL 模型有卷积神经网络（Convolutional Neural Networks，简称 CNN）、循环神经网络（Recurrent Neural Networks，简称 RNN）、注意力机制模型（Attention Model，简称 AM）等。CNN 因其有效捕获局部和全局特征的能力，能够在 NR 领域更好地提取新闻文本特征和用户特征，受到诸多研究者青睐。RNN 因其递归处理历史信息 and 建模历史记忆的特点，特别适合处理序列数据，是建模新闻推荐中随时间变化用户兴趣的有效手段。AM 能够学习不同部分的相对重要程度<sup>[17]</sup>，可以在建模过程中对新闻和用户的特征进行动态调整，在 DNR 领域研究中逐渐成为焦点。还有一些 DL 技术也逐渐在 NR 领域得到应用，如图卷积神经网络（Graph Convolutional neural Network，简称 GCN）、自动编码器（Auto-Encoder，简称 AE）

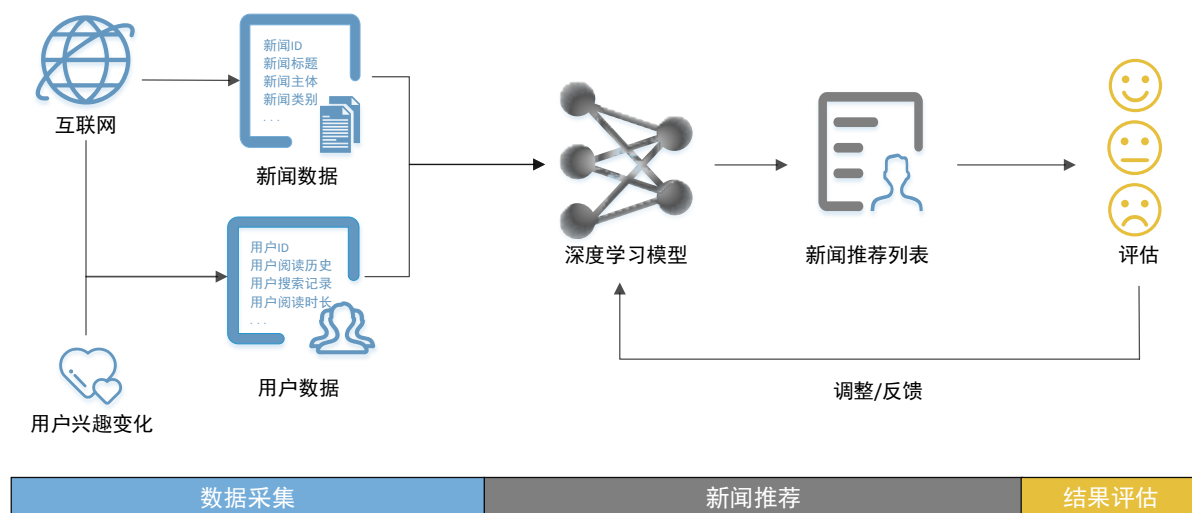


Fig.2 Work flow of DNR

图 2 基于深度学习的新闻推荐的一般流程

以及多层感知机 (Multi-Layer Perceptron, 简称 MLP) 等。目前 DNR 研究已经成为 NR 领域的研究热点之一, 许多国内外大学和研究机构如美国的麻省理工大学、微软亚研院、清华大学、北京邮电大学、上海交通大学、中国科学技术大学等对 DNR 开展了广泛深入研究。近几年的信息检索、数据挖掘、人工智能等相关国际会议 (如 SIGIR<sup>1</sup>、SIGKDD<sup>2</sup>、WWW<sup>3</sup>、CIKM<sup>4</sup>、AAAI<sup>5</sup>、IJCAI<sup>6</sup>等) 和推荐系统大会 (RecSys<sup>7</sup>) 中, DNR 相关论文数量逐年提升。

总结梳理基于深度学习的新闻推荐一般流程, 可将其概括为数据采集、新闻推荐和结果评估三个步骤, 如图 2 所示。其中, 深度学习模型是 DNR 的核心。事实上, 目前研究中, 少数 DNR 方法使用单一深度学习技术, 大部分则混合运用多种深度学习技术设计 NR, 帮助有效提升推荐效果。例如 DADM (Dynamic Attention Deep Model) 模型<sup>[18]</sup>结

合了 CNN 和 AM, 既利用 CNN 提取新闻内容特征和种类特征, 获取新闻的高质量表示, 又利用 AM 调整不同时刻下不同新闻的重要性权重, 综合考虑新闻被选择的概率; DeepJoNN (Deep Joint Neural Networks) 模型<sup>[19]</sup>则结合了 CNN 和 RNN, 构造了基于字符的多维输入 CNN, 建模了带有读者兴趣意图的新闻内容特征, 提升了推荐系统的推荐精准性。

## 2 基于深度学习的新闻推荐方法

DNR 方法通常将 NR 算法与 DL 相结合, 利用 DL 技术对新闻数据和用户数据进行特征提取。为便于剖析介绍, 按照图 1 所示, 本文将 DNR 方法划分为 3 类: “两段式”方法、“融合式”方法和“协同式”方法, 在每类方法中, 将根据建模过程中的具体子任务或其基于的数据组织结构进行更细致划分。表 1 对这 3 类方法从方法特点、优点和缺点等几个方面进行了分析与比较。下面将对每类方法进行详细分析介绍。

<sup>1</sup> International Conference on Research on Development in Information Retrieval

<sup>2</sup> ACM Knowledge Discovery and Data Mining

<sup>3</sup> International World Wide Web Conferences

<sup>4</sup> ACM International Conference on Information and Knowledge Management

<sup>5</sup> AAAI Conference on Artificial Intelligence

<sup>6</sup> International Joint Conference on Artificial Intelligence

<sup>7</sup> ACM Recommender Systems Conference

Table 1 Comparison of three DNR methods

表 1 3 种 DNR 方法对比

类别	方法特点	优点	缺点
“两段式”方法	“先新闻、后兴趣”分步建模，一般有独立的新闻表示向量与用户兴趣表示向量	对新闻特征和用户兴趣特征显式建模，可解释性更强	1.绝大多数方法采用文章级匹配方式，可能会丢失隐藏在更细粒度新闻片段中的语义特征与兴趣特征 2.倾向于推荐与用户点击过的新闻内容相似的新闻，推荐结果缺乏新颖性
“融合式”方法	将新闻和用户信息混合在一起建模，一般没有独立的新闻表示向量	在更细的数据粒度（如新闻类别、内容单词）上同时学习新闻和用户特征，兴趣描述更加细化	1.没有得到明确的新闻表示与用户表示，可解释性弱 2.推荐结果缺乏新颖性
“协同式”方法	不仅使用了用户自身与新闻的交互行为，还利用了相邻用户与新闻的交互	1.考虑相邻用户（或新闻）之间的关联性，新闻和用户兴趣特征更加丰富 2.推荐结果具有新颖性	缺少用户点击历史的顺序信息利用，无法建模兴趣的动态变化

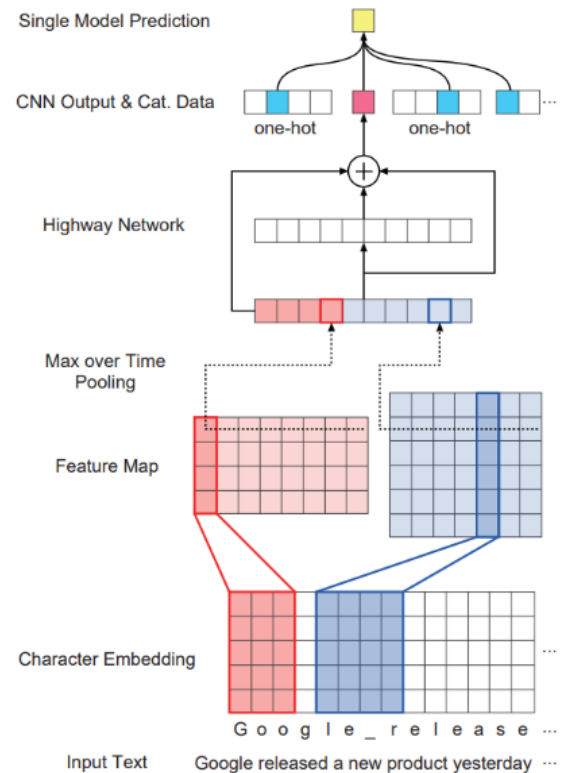
## 2.1 DNR 中的“两段式”方法

在基于深度学习的新闻推荐方法中，多数方法是先提取每篇新闻的特征，再在此基础上学习用户兴趣表示，其中，大部分的工作将每篇新闻建模为一个独立统一的向量。这种“先新闻，后用户”的分步建模思路，本文将其总结为“两段式”新闻推荐方法。根据其建模子任务不同，本文从建模新闻、建模用户整体偏好、建模用户时序兴趣和建模用户混合兴趣来分别介绍。其中，“建模用户整体偏好”和“建模用户时序兴趣”为两种常见的 DNR 应用任务，其区别在于，前者在利用用户新闻点击历史等信息建模用户兴趣时不强调其时间顺序，而后者更注重通过用户点击历史的时序信息捕捉用户兴趣随时间的变化。

### 2.1.1 建模新闻

精准地建模新闻是 NR 中的一项重要任务。一些工作采用单一的深度学习技术对新闻建模，获得新闻的语义表示。CNN 是一种从新闻文本中提取特征的常见方法，例如，文献[20]提出的新闻推荐深度学习元架构 CHAMELEON，其内容表征模块基于 CNN 从单词级别对新闻文本内容进行卷积计算，生成新闻内容嵌入表示；文献[21]在其提出的模型 DAINN（Dynamic Attention-Integrated Neural Network）中，同样利用 CNN 从新闻文本内容的词级表示中提取语义信息。为丰富语义信息，一些工作不止选用了一种新闻数据来建模新闻。文献[18]为

新闻编辑从业者设计出一个专业的新闻筛选推荐系统，针对新闻编辑在筛选新闻时较多依赖新闻文字质量而较少注重关键词、主题等元数据造成的新闻筛选标准不明确的问题，提出利用新闻文本内容

Fig.3 CNN architecture for news recommending prediction<sup>[18]</sup>图 3 用于新闻推荐预测的 CNN 架构<sup>[18]</sup>

新闻类别 2 种数据来建模新闻，预测新闻的筛选标准，其框架如图 3 所示。在该框架中，一方面采用



含有 1 个卷积层和总共 1050 个卷积核的 CNN 模型捕获单词序列中的潜在语义模式来表示文本内容, 另一方面使用 one-hot 向量表示新闻类别等元数据信息, 最后将两者拼接起来预测新闻筛选的概率。此外, 该模型将 CNN 构建于字符之上, 提高了对不同语言的泛化能力。但由于字符中语义信息不足, 会使得字符级 CNN 提取的新闻语义特征不够丰富, 且输入序列的增长可能导致计算成本的增加。

文献[22]在考虑新闻标题的同时, 还将比新闻标题信息量更丰富的新闻概要 (profile, 是从新闻中提取的实体及实体类型) 纳入新闻数据信息, 提出了一种深度注意力神经网络 DAN (Deep Attention Neural Network)。为了学习新闻特征表示, DAN 中设计了由两个并行卷积神经网络组成的 PCNN (Parallel CNN) 组件, 结构如图 4 所示。PCNN 分别以单词级别的新闻标题和概要作为输入, 学习新闻的标题级和概要级 (profile-level) 表示, 并拼接在一起作为最终的新闻特征表示。因有更多数据特征支撑, 带有 PCNN 组件的模型比单纯依靠新闻标题的模型更具竞争力。从以上工作可以看出, CNN 在建模新闻的任务中被广泛应用, 但由于 CNN 固定感受野的限制, 不利于较长新闻词语序列的建模。

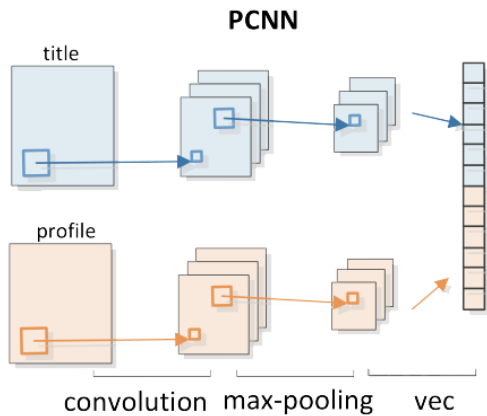


Fig.4 The architecture of PCNN<sup>[22]</sup>

图 4 PCNN 结构<sup>[22]</sup>

除了 CNN 外, 自动编码器 (Auto-Encoder, 简称 AE) 也常被用来学习新闻或用户表示。AE 是一种无监督的神经网络模型, 具有典型的输入-隐含-

输出三层神经网络, 旨在通过一个编码和一个解码过程来重构输入数据, 学习数据的深层表示。在 DNR 领域, AE 在应用过程中往往被加入高斯噪声, 升级为降噪自动编码器 (Denoising Auto-Encoder, 简称 DAE), 在重构输入数据过程中通过去除噪声来学习对输入数据更泛化的表达能力。文献[23]通过改进的降噪自编码器建立新闻的分布式表示。具体来说, 该文献对 DAE 的损失函数进行优化, 加入了弱监督信息, 提高了模型对差距的表达能力, 解决了基于词的方法中“无法识别相似词”的问题, 使得模型能够更为精确学习新闻间的差异与相似性。但由于 DAE 不考虑新闻中词语的顺序信息, 可能会限制该方法获得准确语义表示的能力。

还有一些工作探索多种深度学习技术的混合使用来提高准确建模新闻的能力。文献[24]提出一种带有个性化注意力机制的新闻推荐模型 NPA (neural News recommendation with Personalized Attention), 在其新闻编码器中, 先利用 CNN 处理新闻标题单词序列, 通过捕获新闻标题中单词的局部上下文学习单词的语义表示; 再基于不同的单词在表示新闻时具有不同信息量这一思想, 设计出个性化的 AM 来为不同的用户识别重要的单词, 其模块结构如图 5 所示。图中, 首先用户 ID 经过全连接层生成用户偏好查询向量 (Preference Query), 其次, 根据偏好查询向量和单词表示向量  $[c_1, c_2, \dots, c_M]$  计算出每个单词的注意力权重  $\alpha$ , 将新闻标题单词表示向量依据注意力权重加权处理得到新闻特征  $r$ , 最后, 新闻特征  $r$  作为该模块的输出, 用于最后点击预测计算的输入。该模型根据用户特征动态地关注新闻中重要的单词, 增强了新闻的表示学习能力。该模型可能存在的问题在于, 仅利用了新闻标题这一种新闻信息, 丢失了如新闻主体中丰富的语义信息, 导致新闻语义表示不够全面。

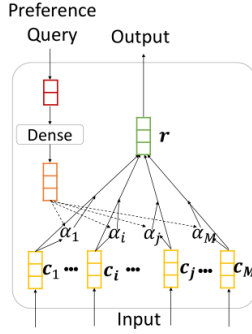


Fig.5 The architecture of the personalized attention module in NPA approach<sup>[24]</sup>

图 5 NPA 方法中的个性化注意力模块结构<sup>[24]</sup>

文献[25]提出的 LSTUR (both Long- and Short-Term User Representations) 模型, 同样采用 CNN 与 AM 相结合的方式提取新闻标题特征, 不同的是, LSTUR 增加了新闻主题和子主题信息, 从二者的 ID 嵌入中学习表征, 并将标题、主题和子主题的表达作拼接得到最终的新闻表示。其结构如图 6 所示。文献[26]提出一种基于 AM 的多视角学习模型 NAML (neural News recommendation with Attentive Multi-view Learning), 除了在建模新闻标题和主体时采用 NPA 模型的 CNN 与 AM 结合的方法外, 还将标题、主体、类别等不同的新闻数据视为不同的新闻视角, 利用 AM 学习不同视角在建模不同新闻时所具有的不同重要程度 (即注意力权重), 得到最终的新闻表示, 有效提升了个性化新闻推荐的性能。另外, 以上 2 个模型的局限性可能在于, 简单的显式主题或类别信息或许不足以细致地描述新闻主题, 特别是当新闻同属于两个或多个不同主题时, 因此, 可以通过挖掘潜在的主题信息以提高新闻表示的准确性。

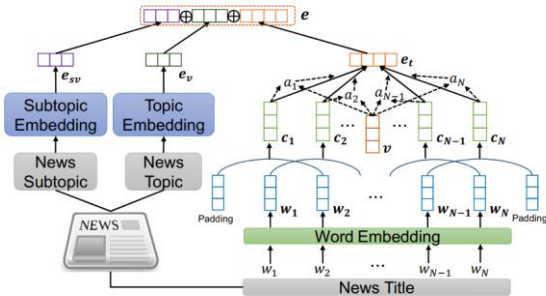


Fig.6 The architecture of the news encoder in LSTUR<sup>[25]</sup>

图 6 模型 LSTUR<sup>[25]</sup>中的新闻编码器架构

CNN 擅长提取局部特征, 却无法捕捉长距离的单词交互, 并且同一新闻中一个单词可能和多个单词之间存在交互, 多头自注意力机制 (Multi-Head Self-Attention, 简称 MHSA) 能够更好地建模上述情况。文献[27]提出一种基于多头自注意力机制的神经新闻推荐方法 NRMS (Neural News Recommendation with Multi-Head Self-Attention), 在其新闻编码器中, 使用 MHSA 从新闻标题中建模单词之间的交互来学习单词的上下文表示, 并使用 AM 选择重要的词语来学习更具有信息量的新闻表示。文献[28]提出的 CPRS (Click Preference and Reading Satisfaction for news recommendation) 模型中的文本编码器, 同样采用了 MHSA 与 AM 相结合的方式获得更准确的文本表示, 结构如图 7 所示。不同之处在于, CPRS 不仅使用了新闻标题还使用了新闻主体内容, 并且对二者分别进行编码。MHSA 不仅能够捕捉新闻中远距离的词与词之间的关联, 还能建模一个词与多个词之间的语义交互, 增强了新闻特征表达能力。

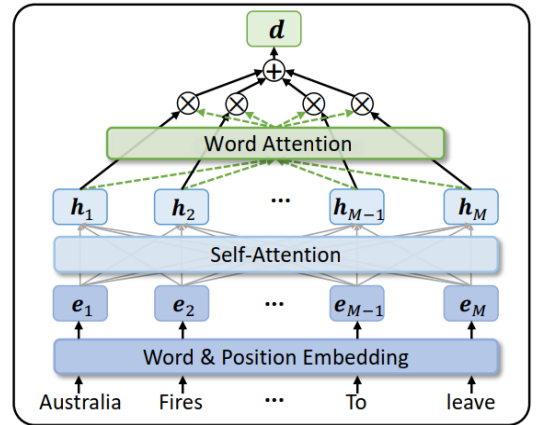


Fig.7 The architecture of the text encoder in CPRS<sup>[28]</sup>

图 7 模型 CPRS<sup>[28]</sup>中的文本编码器架构

新闻语言具有高度浓缩的特点, 包含大量的知识与实体, 仅从语义层面学习新闻表示, 无法充分发现新闻之间潜在的知识层面的联系。近年来, 知识图谱 (Knowledge Graph, 简称 KG) 被广泛应用于机器阅读<sup>[29]</sup>, 文本分类<sup>[30]</sup>和单词表示<sup>[31]</sup>等领域。受 KG 的启发, 文献[32]提出一种基于深度知识感知的新闻推荐系统 DKN (Deep Knowledge-aware

Network), 基于 CNN 利用知识图谱实现新闻推荐。DKN 的关键组件是一个多通道且单词与实体对齐的知识感知卷积神经网络 KCNN(Knowledge-aware Convolutional Neural Network), 实现过程中, 它融合语义层面和知识层面的表示, 将单词、实体、实体的上下文作为 CNN 的多个通道输入, 并在卷积期间维持其对齐关系, 其结构如图 8 所示。KCNN 的实体和单词的对齐机制融合了异构的信息源, 可以更好捕捉新闻之间的隐含关系, 合理扩展推荐结果, 提高推荐性能, 相较于仅使用单词的单通道 CNN 结构, 其 F1 和 AUC 分别提高了 2.7% 和 2.2%。不足的是, 该方法仅利用了新闻标题, 没有充分利用新闻主体等中包含丰富语义的实体。此外, 新闻中更新迅速的人名等未登录词, 在知识图谱中可能缺乏对应的知识, 也会限制模型的推荐效果。

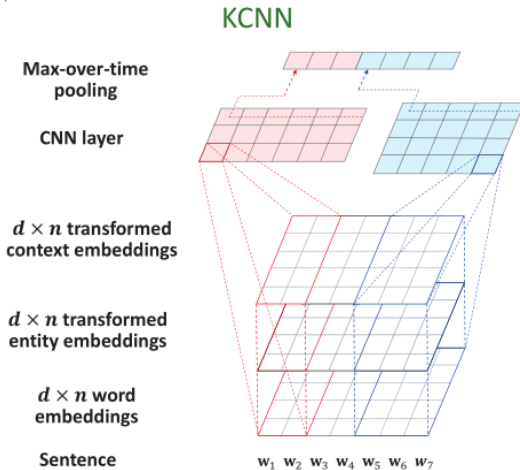


Fig.8 The architecture of KCNN<sup>[32]</sup>

图 8 KCNN 结构<sup>[32]</sup>

在 DKN 的基础上, 文献[33]利用自注意力机制构建语义级和知识级的新闻表示, 其架构中含有词级自注意力模块和项目级自注意力模块。其中, 词级自注意力模块将新闻标签、外部知识图谱中的实体及其上下文作为输入来计算新闻的语义级和知识级表示。其具体操作是, 根据标签和新闻的相关性, 选择  $n$  个单词作为每篇新闻的关键词, 通过这些关键词从知识图谱中选择出关键实体和关键上下文扩充关键词; 然后将这些关键词输入到该词级自注意力模块, 聚合新闻中其他单词, 获得词级新闻表示。

作者采用同样的方法获得了实体级新闻表示和上下文级新闻表示。项目级自注意力模块将以上三个层级(单词级、实体级以及上下文级)的表示进行加权融合, 以获得更准确的新闻表示。

图卷积神经网络(Graph Convolutional neural Network, 简称 GCN)是一种从图数据结构中提取特征的深度学习方法, 因其强大的特征表达能力受到广泛关注, 并在图像识别、自然语言处理等领域得到应用, 它可以聚合每个中心节点及其相邻节点的信息来增强节点的特征表达<sup>[34]</sup>。文献[35]提出了一种上下文感知图架构 CAGE(Context-Aware Graph Embedding for session-based news recommendation), 不仅利用 CNN 从新闻中提取文本级新闻特征、利用知识图谱提取语义级新闻特征丰富新闻语义, 而且提出构建文章级的图, 以利用文章之间的邻域结构信息丰富新闻特征。具体而言, 将文章作为图的节点, 将文章之间的两两相似度值作为边的权值, 移除相似度较小的边以得到一个稀疏图, 并利用多层的 GCN 进一步提炼新闻表示。文献[36]同样采用了知识图谱与图神经网络(Graph Neural Network, 简称 GNN)方法, 提出了一个基于主题的知识图推荐系统 TEKGR(Topic-Enriched Knowledge Graph Recommendation system)。TEKGR 设计了 3 个编码器从语义和知识 2 个角度来处理新闻标题获得新闻表示, 结构如图 9 所示。其中,

词级新闻编码器依次通过词嵌入层、双向门控循环单元(Bidirectional Gated Recurrent Units, 简称 Bi-GRU)层和注意力层来关注对于表征新闻重要的单词, 并学习新闻表示; 知识编码器从新闻标题中的实体概念出发, 利用知识图谱中的关系事实来提取新闻标题的主题向量; 知识图谱级新闻编码器通过在实体之间添加来自知识编码器的主题关系向量构造子图, 并应用 GNN 得到新闻嵌入向量。该模型挖掘了实体之间的主题关系与实体的上下文特征, 提高了新闻建模的准确度。该模型局限性在于没有充分利用除新闻标题外的其他新闻信息, 忽视了可以为标题实体提供上下文语境的新闻主体内容等, 新闻特征表达能力不够丰富。



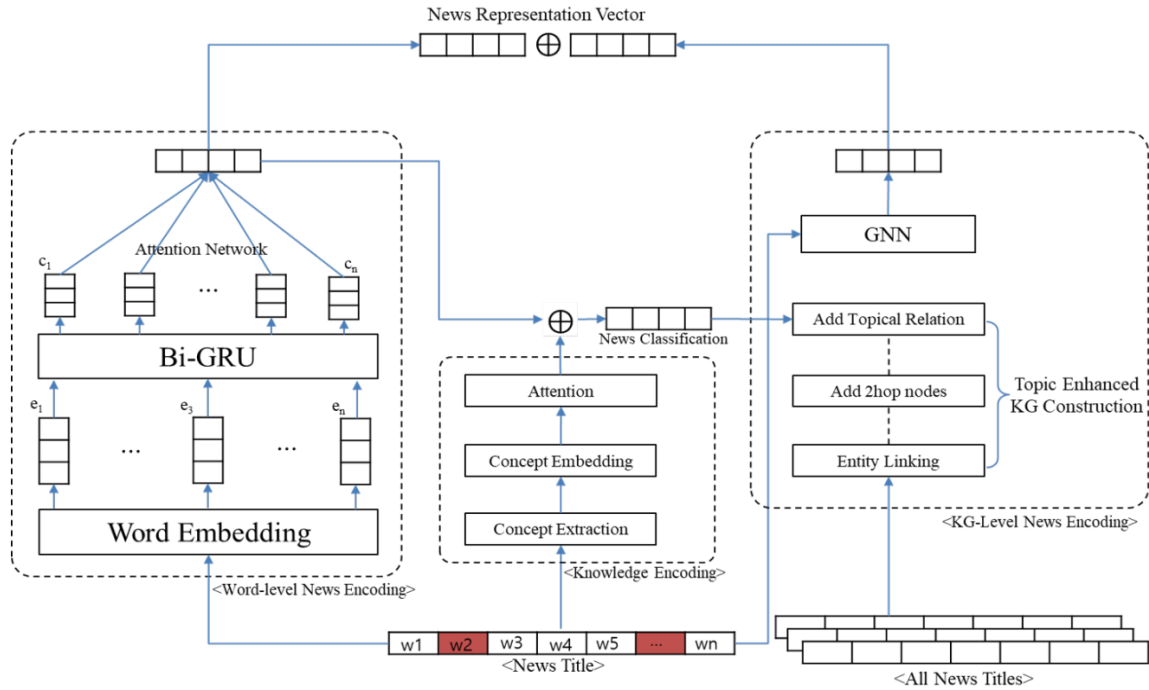


Fig.9 The architecture of KG-based news modeling layer in the TEKGR<sup>[36]</sup>

图 9 TEKGR 模型中的基于知识图谱的新闻建模层结构<sup>[36]</sup>

### 小结

在“两段式”新闻推荐方法的建模新闻方面，各研究工作以获得精确的新闻表示为目标采取了多样的建模方式。一部分工作借助 CNN 强大的局部特征提取能力，有效捕获新闻标题、概要等单词序列的语义特征，提升了模型在 AUC 和 F1 上的性能，且由于 CNN 网络的可并行学习性使得模型有较快的运算速度，但 CNN 固定的感受野可能限制了新闻词语序列的长度。一部分工作利用 AM 可学习不同数据重要性程度的能力，关注新闻中更具有信息量的单词，提升了新闻表示的重点性或个性化程度，其推荐结果在 AUC 或 MRR 上有明显提升。一部分

工作使用 MHSA 建模长距离的依赖，增强了对单词之间长距离的交互的学习能力。一部分工作引入知识图谱来获取新闻的外部知识特征，提高了新闻表示的丰富性，增加了模型在 AUC 或 NDCG 上的性能，但新闻中快速产生的新实体可能无法及时被知识图谱所包含。还有部分工作进一步采用 GCN 技术引入邻域新闻信息，增强了新闻特征表达。表 2 总结了在建模新闻方面不同模型的关键技术、选取的数据种类、数据粒度、方法特点、优点和局限性等。（一些工作只介绍其使用文本型数据而未明确指出所用数据种类，本文将其归类为文本内容。）

Table 2 Comparison of typical models on news modeling in “two-stage” DNR

表 2 “两段式”新闻推荐方法中各代表模型在建模新闻方面的对比

类别	代表模型	发表会议	关键技术	数据种类	数据粒度	方法特点	优点	局限性
单一深度 学习 技术	CHAMELEON <sup>[20]</sup>	RecSys'18	CNN	文本内容	词级	通过捕获单词的上下文，学习文本的局部特征	有效突出新闻局部特征	CNN 固定感受野的限制，不利于较长新闻词语序列的建模
	DAINN <sup>[21]</sup>	ML'19	CNN	文本内容	词级			
	DAN <sup>[22]</sup>	AAAI'19	CNN	新闻标题 新闻概要	词级	增加新闻概要数据	丰富新闻特征	

混合 深度 学习 技术	DADM <sup>[18]</sup>	SIGKDD'17	CNN	文本内容 新闻类别	字符级	增加新闻类别数据； 采用字符级 CNN	丰富新闻特 征；提高对 不同语言的 泛化能力	字符中缺乏 语义信息； 长字符序列 增加计算成 本
	文献[23]	SIGKDD'17	DAE	文本内容	词级	增加弱监督信息	提高对相似 词的识别能 力	DAE 不考虑 新闻中词语 的顺序信息 ，限制新闻 表示准确性
	NPA <sup>[24]</sup>	SIGKDD'19	CNN; AM	新闻标题	词级	针对不同用户来个 性化地关注单词的 重要程度	建立个性化的 新闻特征	数据种类单 一，新闻语 义表示不够 全面
	LSTUR <sup>[25]</sup>	ACL'19	CNN; AM	新闻标题 新闻主题 新闻子主题	词级	关注标题中重要的 单词；增加主题及子 主题信息	抓住新闻重 点信息；丰 富新闻特征	仅使用简单 的显式主题/ 类别，未挖 掘潜在的题 目信息
	NAML <sup>[26]</sup>	IJCAI'19	CNN; AM	新闻标题 新闻主体 新闻类别	词级	学习不同数据种类 在建模不同新闻时 所具有的不同重要 程度	提高新闻特 征的个性化 程度	
	NRMS <sup>[27]</sup>	ACL'19	MHSA; AM	新闻标题	词级	捕捉长距离的单词 交互； 关注标题中重要的 单词	提高长距离 交互的特征 学习能力	数据种类单 一，新闻语 义表示不够 全面
	CPRS <sup>[28]</sup>	IJCAI'20	MHSA; AM	新闻标题 新闻主体	词级	捕捉长距离的单词 交互； 关注标题中重要的 单词	提高长距离 交互的特征 学习能力； 丰富新闻语 义特征	MHSA 层数 较少，加深 层数或许能 增强新闻特 征表达能力
	DKN <sup>[32]</sup>	WWW'18	CNN; KG	新闻标题	词级	通过 KG 引入外部知 识	引入知识丰 富语义；能 够捕捉新闻 间的隐含关 系	仅利用标题/ 关键词等单 一数据没有 充分利用新 闻主体等中
	文献[33]	WI'18	SA; KG	新闻标签 (即关键 词)	词级	通过 KG 引入外部知 识；使用 SA 关注重 要的词级特征	引入知识丰 富语义；抓 住新闻重点 信息	包含丰富语 义的实体； 新闻中更新 迅速的人名 等未登录词 在知识图谱 中可能缺乏
	CAGE <sup>[35]</sup>	RecSys'20	CNN; KG; GCN	文本内容	词级	通过 KG 引入外部知 识；通过 GCN 利用 文章间的邻域结构 信息	引入知识丰 富语义；利 用邻域新闻 丰富新闻特 征	

TEKGR <sup>[36]</sup>	CIKM'20	Bi-GRU; KG; GNN	新闻标题	词级	利用实体外部知识 和新闻的主题关 联性来获得准确的 新闻表示	提高新闻表 示的准确性	对应的知识
-----------------------	---------	--------------------	------	----	---	----------------	-------

### 2.1.2 建模用户

NR 中的用户兴趣含有多样性与变化性等特点,使得准确的用户建模具有一定挑战。本文将“两段式”新闻推荐方法的建模用户部分分为“建模用户整体偏好”、“建模用户时序兴趣”和“建模用户混合兴趣”3类进行介绍。其中,“建模用户整体偏好”将用户的新闻点击历史当作集合,建模用户兴趣时不强调其时间顺序;“建模用户时序兴趣”将用户的点击历史当作时间序列,更注重应用时序信息捕捉用户兴趣随时间的变化;“建模用户混合兴趣”则兼顾以上两种建模兴趣角度。

#### 建模用户整体偏好

用户整体偏好是用户过往一段时间的整体兴趣,强调兴趣的整体性、重点性,往往是对用户长期兴趣的建模。通常来讲,聚合用户的新闻点击历史可以得到用户的整体偏好,然而并非每一篇点击新闻对建模用户兴趣都具有相同的重要性,因此,很多工作采用 AM 机制以学习用户不同点击新闻的重要性权重。DKN<sup>[32]</sup>模型、TEKGR<sup>[36]</sup>模型和文献[33]所提模型均使用 AM 来表征推荐候选新闻对用户点击历史新闻的影响。以 DKN 为例,具体过程如图 10 所示,在获得每条输入新闻的表示向量的基础上,通过 AM 机制计算候选新闻向量与每条点击新

闻向量之间的注意力权重,最后使用该权重动态聚合(按照权重加权)用户的点击历史计算出用户兴趣表示作为推荐依据。NPA<sup>[24]</sup>模型同样使用了 AM 机制建模用户整体偏好,但与 DKN 不同的是,其 AM 的查询向量(Query Vector)并非是候选新闻表示而是用户的 ID 嵌入。其作者认为同一篇新闻在建模不同用户时具有不同信息量,故设计了一个个性化的注意力网络,基于用户 ID 嵌入来对该用户所点击的新闻赋予不同权重,以此获得用户兴趣的最终表示。该类模型利用 AM 加权聚合用户的点击历史,使得偏好的重点性得以突出,提高了用户兴趣建模的准确性。不过该方法没有对点击新闻之间的潜在联系进行建模,其推荐效果还有提升的潜力。

除了关注同一用户不同点击新闻的重要性差异外,同一用户的历史点击新闻之间也可能存在关联性,并且一篇新闻可能和多篇新闻相关,关注新闻之间的关联性同样有助于挖掘用户兴趣。NRMS 模型<sup>[27]</sup>在用户编码器中使用多头自注意力机制捕捉新闻之间的长距离依赖关系,以增强用户的特征表达,其模型结构如图 11 所示。每篇新闻的表示是由多个独立自注意头输出的表示拼接而成,针对不同的新闻在建模用户时可能具有不同的信息量,采用注意力机制选择重要的新闻来学习用户表示。相较

于 DKN 模型, 该模型在 AUC 上增加了 4.03%, 提高了推荐模型的推荐结果排序能力。但该方法还缺乏对多种新闻信息的有效整合, 没有充分挖掘各种新闻信息的语义表示。

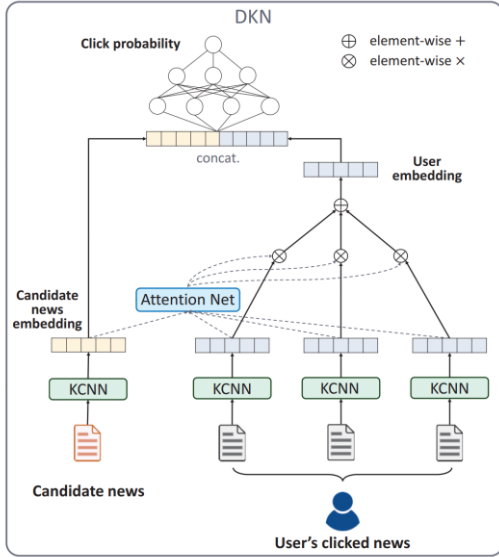


Fig.10 The architecture of DKN approach<sup>[32]</sup>

图 10 DKN 方法的架构<sup>[32]</sup>

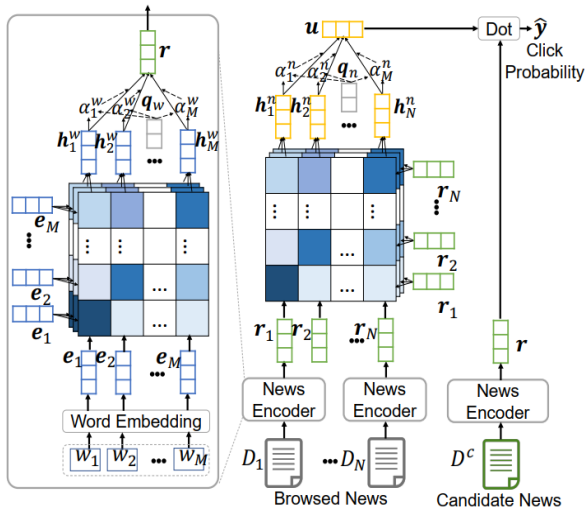


Fig.11 The architecture of NRMS approach<sup>[27]</sup>

图 11 NRMS 方法的架构<sup>[27]</sup>

以上工作均基于用户对新闻的点击行为建模用户兴趣, 而 CPRS<sup>[28]</sup>提出从用户对新闻标题的点击

行为和对新闻内容的阅读行为两方面建模用户整体偏好, 以避免“用户被标题吸引而点击但对内容并不感兴趣”的情况。具体而言, 提出了一个个性化的阅读速度度量标准来衡量用户对新闻内容的满意程度, 该度量标准是基于用户的阅读停留时间、新闻内容长度和历史平均新闻阅读速度来计算的。CPRS 从用户阅读过的新闻内容和对新闻的满意度中学习用户表示, 以此来建模用户对新闻内容的兴趣, 其模型结构如图 12 所示。模型构建了内容注意力网络挑选重要的新闻内容来学习新闻内容表示, 同时构建了满意度注意力网络识别用户满意的新闻内容, 其查询向量为个性化的阅读速度; 此外, 还构建标题注意力网络从用户点击过的新闻标题中建模用户对新闻标题的偏好, 并构建行为注意力网络将这两种用户表示结合成统一的用户兴趣表示用于新闻推荐。阅读行为的引入和 AM 的灵活运用使得对用户兴趣的建模更加全面而准确, 有助于向用户推荐内容满意的新闻。然而, 公开的数据集中几乎没有用户阅读停留时间此类相关数据, 限制了模型的可复现性。

在工业实践中, 商业化的新闻推荐系统通常从多个渠道/服务 (新闻阅读、网页浏览、网页搜索) 中收集数据以丰富用户特征, 不同用户在不同渠道中有不同的活跃度分布, 不同渠道存在差异需要根据内容特征来选择。多层感知机 (Multi-Layer Perception, 简称 MLP) 也是一种提取用户或新闻特征的常用方法, 它是一种基础的深度神经网络, 近年来被广泛应用于推荐领域<sup>[37][38]</sup>, 相比于 CNN、GCN、DAE 等网络在特征表示能力上的强大, MLP 的特征表示更为简单高效。文献[39]提出的深度融合模型 DFM (Deep Fusion Model) 使用了 MLP 和 AM 两种技术提高建模用户兴趣的能力。DFM 利用多个不同深度的 MLP 网络来并行学习不同渠道中的用户表示, 并向最终的激活层提供不同层次的组合特征, 以此提高对用户的表示学习能力。AM 模块主要负责将从不同渠道中得到的用户表示按照学习到的注意力权重进行加权以获得最终的用户表

示, 其 AM 的查询向量为用户的所处的位置、时间等上下文信息。该模型通过 AM 机制融合借鉴了其他渠道信息, 在 Bing News 数据集上的 AUC 高达 0.8386, 不仅可帮助提高推荐准确率, 还有助于解

决 NRS 中的用户冷启动问题。由于需要大量用户个人信息与时空环境信息, 该模型适用于需要用户登录的场景或移动场景。不过该模型在很大程度上依赖于手工特征工程, 需要大量的领域知识。

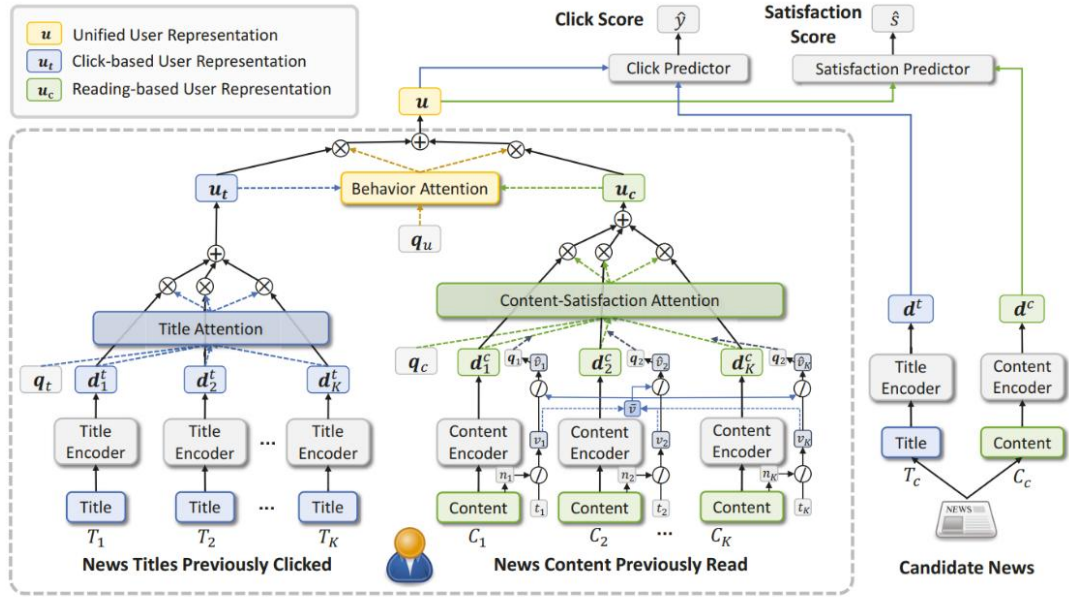


Fig.12 The architecture of CPRS approach<sup>[28]</sup>

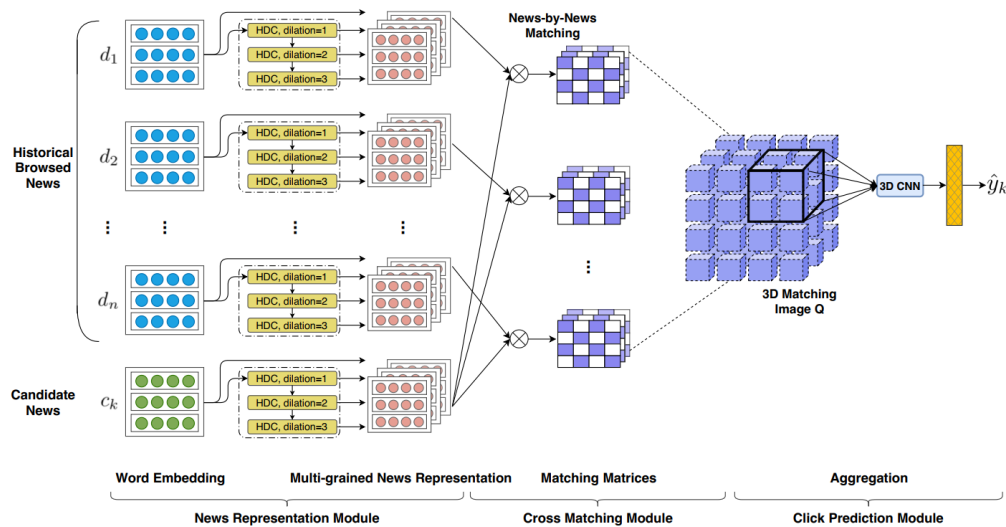
图 12 CPRS 方法的架构<sup>[28]</sup>

从小本节以上工作可以看出, 现有建模用户整体偏好的工作一般将每个用户表示为单一向量, 然后匹配候选新闻向量, 但这在推荐时可能会丢失细粒度信息, 因为一些重要语义特征往往隐含在不同粒度的新闻文本片段中。文献[40]提出一种基于细粒度兴趣匹配的新闻推荐方法 FIM (a Fine-grained Interest Matching network), 没有将用户点击过的所有历史新闻聚合成一个统一的向量, 而是通过层级扩张卷积 (Hierarchical Dilated Convolution, 简称 HDC) 构造新闻的多级表示, 然后在每个语义层次上对每个浏览新闻和候选新闻进行细粒度匹配, 结构如图 13 所示。HDC 逐步从词汇、短语、句子等不同粒度上的局部相关性和长期依赖性中获得新闻的语义特征; 再基于多层级的新闻表示为每个历史-

候选对构建不同粒度的文本段 (特征) 相似度矩阵, 得到多通道的匹配度矩阵; 最后使用 3D-CNN 提取高阶显著特征来预测用户对候选新闻的点击概率。这种细粒度的兴趣匹配方法细化了用户兴趣特征, 提高了推荐结果的准确性。另外, FIM 虽然没有将新闻或用户建模为一个单一的向量, 但依然符合“先新闻, 后用户”的“两段式”建模方法。

此类“建模用户整体偏好”的推荐方法, 能够很好地建模兴趣的整体性与重点性, 并依赖于用户整体兴趣偏好对候选新闻进行推荐, 但忽略了用户点击新闻的顺序信息的影响。事实上, 用户点击历史的顺序信息能够更好地反映用户在一段时间内兴趣的变化和多样性, 有助于更精确地对用户偏好建模。



Fig.13 The architecture of FIM approach<sup>[40]</sup>图 13 FIM 方法的架构<sup>[40]</sup>

### 建模用户时序兴趣

用户时序兴趣是指随时间动态变化的用户兴趣，对其建模的关键在于捕捉用户新闻点击历史中的时序信息，抓住用户兴趣随时间变化的特征。循环神经网络的网络单元中存在内部隐藏状态，可以记录和保存历史信息，特别适用于对序列数据建模。但普通的 RNN 结构普遍存在梯度消失问题，难以解决数据长期依赖（Long-Term Dependency）<sup>[41]</sup>关系的学习问题，因此 RNN 的变种，如长短期记忆网络（Long Short-Term Memory，简称 LSTM）和门控循环单元（Gated Recurrent Unit，简称 GRU）等应用广泛。文献[42]将用户点击历史按照 30 分钟无活动间隔划分为多个会话（Session），尝试从两个角度建立 LSTM 来捕捉动态变化的用户兴趣过滤候选新闻。其中一种是基于会话的 LSTM，将用户当前会话中的点击历史作为输入，捕获用户的即时兴趣，根据其过滤新发布的新闻；另一种是基于历史的 LSTM，将用户过去一段的点击历史作为输入，捕获用户的短期兴趣，根据其过滤新发布的新闻。最后，针对这些通过 LSTM 获得的候选新闻，依据其新闻类别与用户长期偏好新闻类别间的相似性进行重新排序得到推荐结果。文献[23]和文献[35]的 CAGE 模型均采用 GRU 建模用户时序兴趣，利用

GRU 捕捉用户和新闻之间的顺序交互能够更好地建模用户动态变化的兴趣，并且不会发生 LSTM 可能出现的梯度爆炸问题。以文献[23]为例，GRU 的输入为用户浏览文章的序列，输出为当前用户兴趣状态，其中，当前用户兴趣状态 $u_t$ 是该用户上一状态 $u_{t-1}$ 和浏览状态 $a_t^u$ 决定的，结构如图 14 所示。另外，该模型在考虑大用户群背景下利用内积操作将用户表示和新闻表示进行匹配完成推荐，在真实的百万级用户新闻推荐上获得了较好效果，在工业级应用方面具有一定的借鉴意义。该模型的局限性在于，由于 GRU 比较费时，可能会限制获得用户兴趣表示的速度。

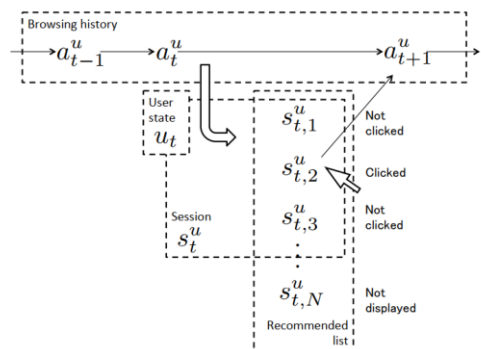
Fig.14 A user's browsing and clicking history of literature<sup>[23]</sup>

图 14 文献[23]中的一个用户的浏览和点击交互历史

为区分历史新闻序列中新闻的不同重要性，文献[43]在利用 LSTM 建模用户的基础上加入了 AM，提出了一种使用双向 LSTM 模型的混合循环注意力 HRAM (Hybrid Recurrent Attention Machine)。HRAM 使用双向 LSTM (Bidirectional LSTM, 简称 Bi-LSTM) 构建用户历史组件，在学习了每篇新闻的分布式表示基础上，以顺序方式将用户的点击历史数据提供给基于注意力的递归层，其结构如图 15 所示。这种 Bi-LSTM 和 AM 相结合的方法有效捕获了用户点击顺序中隐藏的大量兴趣信息，并据此区分新闻的重要程度，从而理解和适应用户不断变化的兴趣，提高了推荐精度。

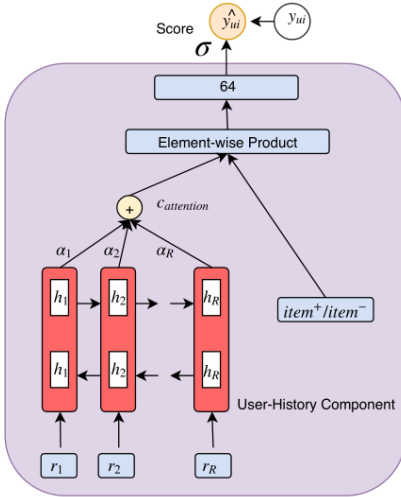


Fig.15 User-history component using attention-based recurrent network<sup>[43]</sup>

图 15 使用基于注意力机制的循环神经网络的用户历史组件<sup>[43]</sup>

### 建模用户混合兴趣

以上两种建模兴趣方式只学习用户兴趣的单一表现形式，可能无法充分捕获用户特征。“建模用户混合兴趣”的用户建模方式中同时包含对用户整体偏好与时序兴趣的建模，既考虑用户的长期兴趣，又考虑用户兴趣的动态变化。文献[22]提出的 DAN 模型不仅引入了一个 ANN (Attention Neural Network) 组件基于点击历史集合来学习用户兴趣嵌入 (即整体偏好)，还设计了一个 ARNN (Attention-based RNN) 组件来捕捉用户点击历史中的潜在顺

序特征，学习用户历史序列嵌入 (即时序兴趣)。基于用户每次点击都受其之前新闻选择的影响这一假设，作者将 ARNN 组件设计为一种基于 AM 的 RNN，在 LSTM 的每个状态上都加入 AM，以获得不同点击时间下丰富的序列特征，最终这些来自不同时间的特性被集成为用户点击历史的顺序特性表示。最后，将用户兴趣嵌入与用户历史序列嵌入作拼接并经过全连接网络得到用户嵌入，即最终的用户兴趣表示。该模型在 Adressa-1week 和 Adressa-10week 两个公共数据集上的结果表明，同时考虑用户整体偏好与时序兴趣的模型相较于只考虑整体偏好的模型其 F1 分别提高 0.89% 与 3.27%，其 AUC 分别提高 3.77% 与 0.80%。

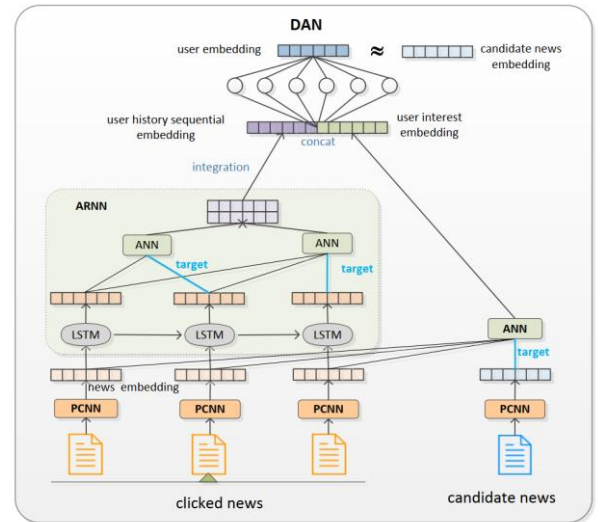


Fig.16 The architecture of DAN approach<sup>[22]</sup>

图 16 DAN 方法的架构<sup>[22]</sup>

LSTUR 模型<sup>[25]</sup>也是一种同时学习用户整体偏好表示和时序兴趣表示的新闻推荐模型。该方法基于 GRU 设计出两种融合用户整体偏好和时序兴趣的推荐框架：一种是使用用户整体偏好表示 (由用户 ID 嵌入得到) 初始化 GRU 网络隐藏层状态，并将 GRU 网络用于学习用户时序兴趣作为最终用户表示进行推荐，如图 17(a)所示；另一种是将用户整体偏好表示和 GRU 生成的时序兴趣表示拼接成最终用户表示进行推荐，如图 17(b)所示。两种方法中，前者推荐结果更加准确而后者性能更加稳定。相较

于基于 AM 或 LSTM 的推荐框架,基于序列的 GRU 因其能够捕捉用户的新闻阅读模式、具有更少的参数、更小的过拟合风险使得推荐性能得以提升,两种方法的 AUC、MRR 和 NDCG 相较于其基线算法均至少提高 1%。不过,该模型仅使用用户 ID 的嵌入向量来建模用户整体偏好,缺乏对用户的长期历史等信息的利用,在一定程度上限制了其对用户整体偏好建模的能力。

不同于以上利用 RNN 系列神经网络来建模用户混合兴趣的方法, SASKR<sup>[44]</sup>设计了一个带多头自注意力机制 (HMSA) 的编码-解码结构,还利用了蕴含丰富信息知识图谱。其结构如图 18 所示,在建模用户时序兴趣方面,利用 HMSA 对用户的点击

序列进行编码,再利用 HMSA 解码计算候选新闻基于用户时序兴趣的推荐得分。作者还建模了用户整体偏好的扩展,利用知识图谱,将点击历史新闻中的实体在知识图谱上进行一跳扩展,形成上下文实体矩阵,再利用注意力机制计算候选新闻基于知识的推荐得分。并将二者加权融合得到最终得分。该文献基于 HMSA 有效提取了时序兴趣特征,并利用知识图谱挖掘新闻之间的深层联系,对用户整体偏好进行扩展,提高了推荐的合理多样性和可扩展性。但在实际应用过程中,由于新闻的高时效性产生大量未登录词,很多新实体无法及时更新在知识图谱中,或者在实体链接等过程中出现差错,可能导致推荐效果的下降。

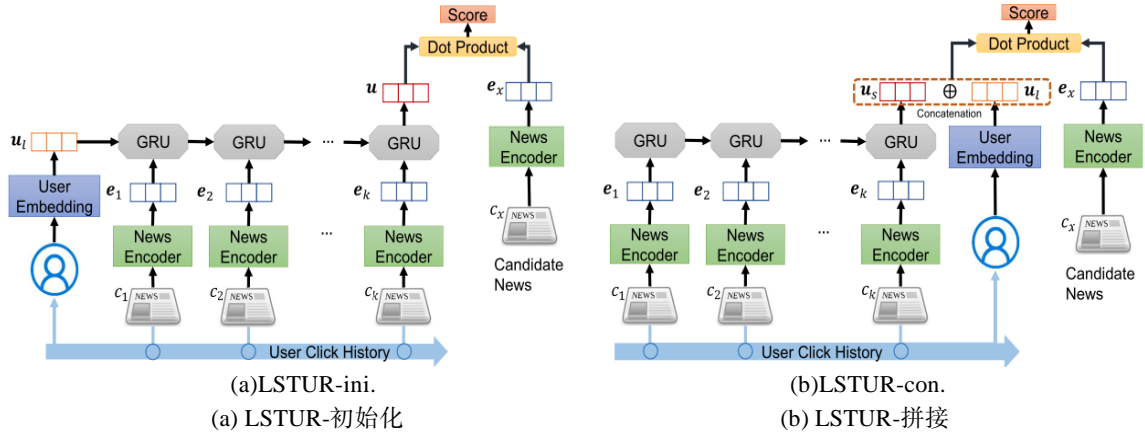


Fig.17 The two frameworks of our LSTUR approach<sup>[25]</sup>

图 17 LSTUR 方法的两种架构<sup>[25]</sup>

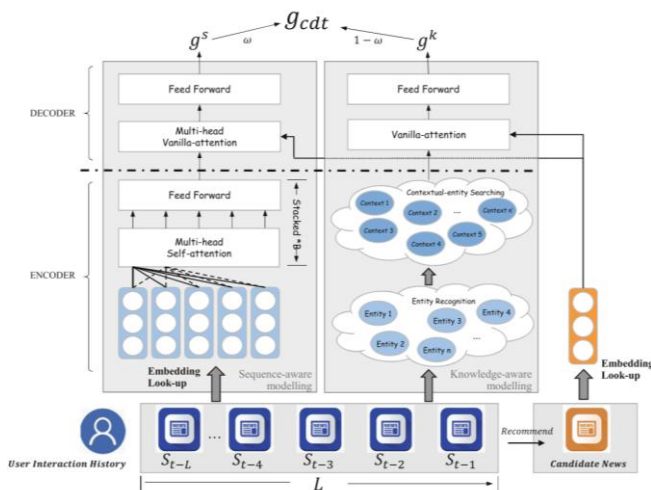


Fig.18 The architecture of SASKR approach<sup>[44]</sup>

图 18 SASKR 方法的架构<sup>[44]</sup>

和其它模型不同的是, DAINN 模型<sup>[21]</sup>除了融合用户整体偏好、用户时序兴趣外,还引入大众兴趣共同建模用户兴趣表示。大众兴趣由同时间段内近邻用户的点击新闻做平均池化 (mean pooling) 来提取,以减少由用户误点历史记录带来的错误兴趣,同时缓解用户冷启动问题。在获取用户时序兴趣方面, DAINN 模型与之前的模型不同在于,引入了用户所处的时空因素来建模。具体而言,以周 (week) 为时间单位将用户在  $t$  时刻的点击行为分解为某天 (Day of Week)、某小时 (Hour of Day)、地点 (Location) 三个方面的时空特征,将这三种特征的融合嵌入表示与用户历史点击新闻的嵌入表示拼

接起来, 利用 DNN 网络计算注意力影响权重, 再依照注意力权重对用户历史点击新闻做加权计算, 最终形成用户在  $t$  时刻的时序兴趣表示。而在用户整体偏好建模方面, DAINN 利用主题模型学习用户的兴趣主题分布, 聚合用户点击行为的主题分布作为用户的兴趣主题, 计算用户的兴趣主题与点击新闻的相似度得到联合向量作为用户整体兴趣。最后将大众兴趣、用户时序兴趣与用户整体偏好做拼接, 并输入 GRU 网络来动态描述用户兴趣的发展。DAINN 将大众兴趣引进模型, 纠正了建模存在的兴趣误差; 同时将用户所处的时空环境因素融入模型中, 使得模型对用户的兴趣变化更加敏感。此外, 该模型适合于强调时空环境变化的移动推荐场景。

### 小结

在“两段式”新闻推荐方法的建模用户方面, 各研究工作以获得准确的用户兴趣表示为目标采用

了多种建模角度与建模技术。一部分工作建模用户整体偏好, 其模型中多含有 AM, 它可以根据用户的需求特点动态关注对其重要的新闻数据, 从而适应不同用户的偏好, 有效提高用户表示学习的准确性, 相较于无 AM 的模型而言, 引入 AM 能使推荐结果在 AUC 或 F1 指标上有明显提升。一部分工作建模用户时序兴趣, 多利用 RNN 系列的深度学习技术适合处理时间序列的特性, 学习用户新闻点击顺序中隐藏的大量兴趣信息, 捕获用户兴趣动态变化, 但由于 RNN 的不可并行性导致其运行速度较 CNN 更慢。一部分工作建模用户混合兴趣, 综合考虑用户整体偏好和时序兴趣, 或加入如大众兴趣等其它因素做调整, 其 AUC 指标值会有所增加。表 3 对比总结了在建模用户方面不同模型的关键技术和特点优势等。

Table 3 Comparison of typical models on user modeling in “two-stage” DNR

表 3 “两段式”新闻推荐方法中各代表模型在建模用户方面的对比

类别	代表模型	发表期刊/会议	关键技术	方法特点	优点	局限性
建模用户整体偏好	DKN <sup>[32]</sup>	WWW'18	AM	捕捉不同候选新闻下用户点击新闻的不同重要程度, 动态聚合用户历史	用户特征更具有信息量	忽略用户点击新闻的顺序信息, 缺乏对用户兴趣动态变化的建模
	TEKGR <sup>[36]</sup>	CIKM'20	AM			
	文献[33]	WI'18	AM			
	NPA <sup>[24]</sup>	SIGKDD'19	AM	利用个性化注意力网络捕捉建模不同用户时其点击新闻的不同影响, 动态聚合用户历史	用户特征更加个性化	
	NRMS <sup>[27]</sup>	ACL'19	MHSA; AM	挖掘同一用户的历史点击新闻之间可能存在的关联性	挖掘深层用户兴趣特征	
	CPRS <sup>[28]</sup>	IJCAI'20	AM	从内容兴趣与标题偏好两方面考虑, 提出个性化阅读速度度量来建立 AM 以选择重要的新闻内容	避免用户被标题吸引而点击但对内容并不感兴趣的情况	
	DFM <sup>[39]</sup>	IJCAI'18	MLP; AM	融合多个渠道信息	丰富用户兴趣特征	
	FIM <sup>[40]</sup>	ACL'20	HDC; 3D-CNN	构造新闻的多级表示, 在每个语义层次上对每个点击新闻和候选新闻进行细粒度匹配	细化用户兴趣特征	

建模用户时序兴趣	文献[42]	CIKM'17	LSTM	通过捕捉用户的即时兴趣与短期兴趣来建模用户时序兴趣	适应用户兴趣变化	缺乏对用户长期偏好的建模
	文献[23]	SIGKDD'17	GRU	建模用户与新闻的顺序交互中的时序兴趣	适应用户兴趣变化；避免 LSTM 可能出现的梯度爆炸问题	
	CAGE <sup>[35]</sup>	RecSys'20	GRU			
	HRAM <sup>[43]</sup>	CIKM'18	Bi-LSTM; AM	捕获了用户阅读顺序中时序兴趣信息，并区分了新闻的重要程度	适应用户兴趣变化；提高用户特征质量	
建模用户混合兴趣	DAN <sup>[22]</sup>	AAAI'19	整体偏好: AM	在 LSTM 的每个状态上都加入 AM，以获得不同点击时间下丰富的序列特征	兼顾用户兴趣的长期性与变化性	能有效降低上述 2 种方法的局限性
			时序兴趣: LSTM; AM			
	LSTUR <sup>[25]</sup>	ACL'19	整体偏好: 用户 ID 嵌入	尝试了初始化和拼接两种方式融合整体偏好与时序兴趣		
			时序兴趣: GRU			
	SASKR <sup>[44]</sup>	CCL'19	整体偏好扩展: KG; AM	利用 KG 合理扩展用户整体偏好	兼顾用户兴趣的长期性与变化性；对用户整体偏好合理扩展	
			时序兴趣: MHSA			
	DAINN <sup>[21]</sup>	ML'19	大众兴趣: 平均池化	将用户的细粒度时空环境纳入模型，并引入大众兴趣调整错误兴趣	提高用户兴趣变化的敏感度；纠正兴趣错误	
			整体偏好: 主题模型			
时序兴趣: AM						

总体而言，DNR 中的“两段式”方法具有“先新闻，后兴趣”的分步建模的特点，一般可得到独立的新闻表示向量与用户兴趣表示向量，是对新闻特征与用户兴趣特征的显式建模，因而可解释性更强。但是，大多数方法采用文章级匹配方式，可能会丢失隐藏在更细粒度新闻片段中的语义特征与兴趣特征，且往往仅基于用户自己的点击新闻，倾向于推荐与用户点击过的新闻内容相似的新闻，推荐结果缺乏新颖性。

## 2.2 DNR 中的“融合式”方法

在基于深度学习的新闻推荐中，一部分方法并不先为每一篇建立单独的新闻表征，而是将新闻和用户信息融合在一起，同时学习新闻和用户特征。本文将之总结为“融合式”新闻推荐方法。

2018 年，文献[19]提出的 DeepJoNN 模型，将新闻和用户相关的多维数据（包括新闻类别、关键字和实体、新闻 ID 和用户 ID 等）融入到同一矩阵中同时学习特征。具体而言，该模型把新闻类别、关键字、实体、用户 ID 等信息分别编码为向量，将多个向量垂直堆叠形成字符级嵌入矩阵，再将该矩阵作为 CNN 的输入对新闻和用户联合建模；模型还将 CNN 和 LSTM 通过上下分层的形式相耦合，以同时学习新闻上下文特征和点击流中的时序模式，并预测用户的下一次点击行为，结构如图 19 所示。其评价指标 R 与 MRR 在新闻数据集 Adressa 和音乐数据集 Last.fm 上均有提升，证明该模型具有一定的普适性。此外，该模型仅使用了新闻类别、关键字、实体等粗粒度的信息，缺乏对更加全面细致的新闻语义信息的挖掘利用。



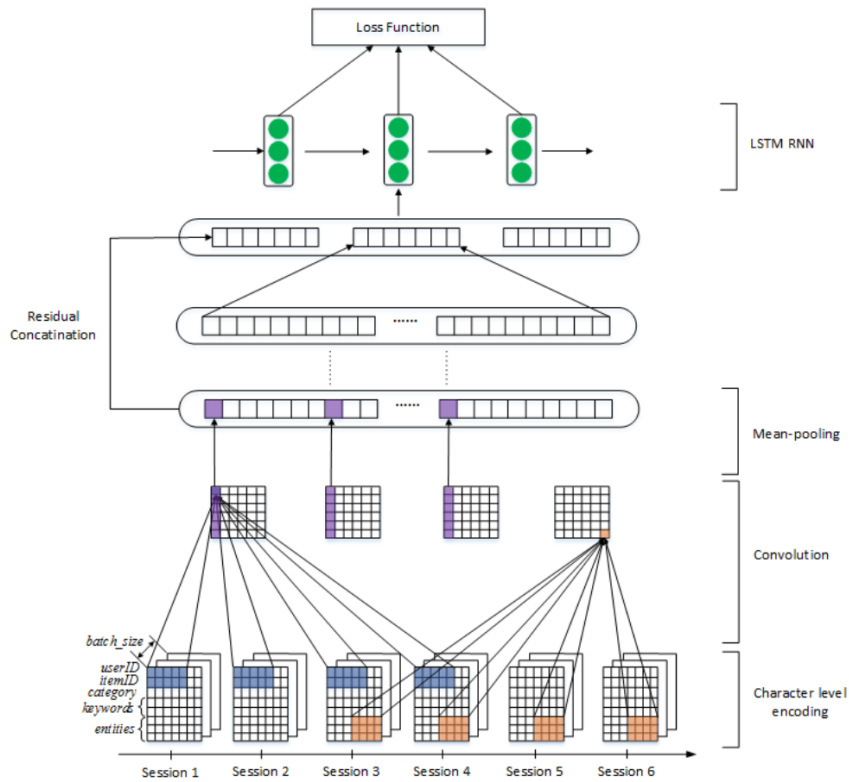


Fig.19 The architecture of DeepJoNN approach<sup>[19]</sup>

图 19 DeepJoNN 方法的架构<sup>[19]</sup>

文献[45]将用于视频动作识别领域的 3-D CNN (3-D Convolutional Neural Network) 模型<sup>[46]</sup>引入新闻推荐方法中, 将用户和新闻信息融合到一个 3-D CNN 中, 建模用户时序兴趣, 其结构如图 20 所示。作者并没有先对每篇历史新闻建立单独特征向量, 而是借助 3-D CNN 可通过 3D 卷积操作在时间和空间两个维度上提取特征的特点, 将同一用户的点击历史与候选新闻融合在一个 3-D CNN 网络中进行特征分析, 通过计算用户点击历史中每篇文章的每

个单词与候选新闻中每个单词的相似度, 得到一个三维相似度张量, 基于该相似度张量进行 3D 卷积以提取用户随时间变化的阅读兴趣, 其中捕获用户兴趣历史的时间窗口大小可由卷积核大小决定。总体来看, 该模型使用单词之间的语义相似度作为 3-D CNN 的输入, 建模用户动态变化的时序兴趣, 推荐效果在命中率 (Hit Ratio, 简称 HR) 和归一化折扣累计效益 (Normalized Discounted Cumulative Gain, 简称 NDCG) 上有所提升。

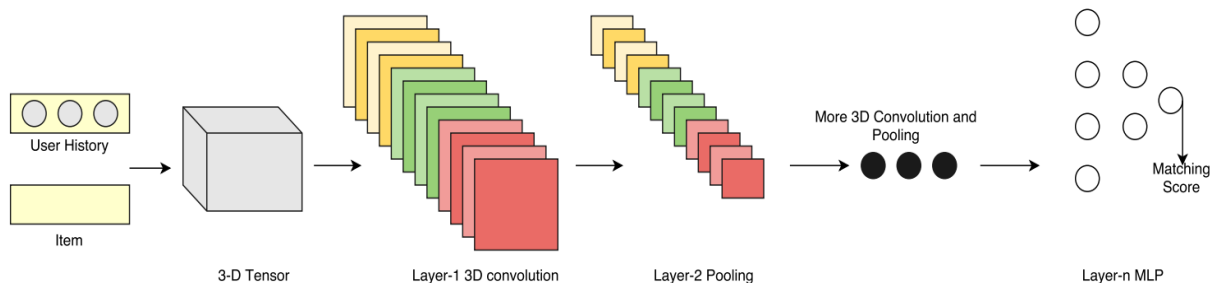


Fig.20 The architecture of 3-D CNN in literature[45]

图 20 文献[45]中的 3-D CNN 结构

文献[47]在文献[45]的基础上，增加了一个 2-D CNN（2-Dimensional Convolutional Neural Network），提出一种 Weave&Rec 框架来进行新闻推荐，将用户点击历史与候选新闻分开建模，其结构如图 21 所示。与上文相同，作者并没有先对每篇历史新闻建立单独特征向量，而是直接将用户所有点击历史作为 3-D CNN 的输入，并经过 3D 卷积与池化的交替处理，其中 3D 矩阵由多个 2D 矩阵堆叠而成，每个 2D 矩阵为一篇点击历史新闻的前 50 个单词 word2vec 向量表示组成；候选新闻则被输入 2-D CNN，并经过 2D 卷积与池化的交替处理。最后两部分结果经过内积运算和一个全连接层的计算输出预测评分。该模型中 3-D CNN 可以同时学习空间信息（新闻特征）和时序信息（用户点击历史中的顺序特征），捕捉用户兴趣的动态变化，最终得到较高质量的推荐效果。

小结

总的来说，在“融合式”新闻推荐方法中，各

研究工作将新闻和用户信息混合在一起建模，并不先将每篇新闻建模为一个整体的向量表征，故一般没有独立的新闻表示向量。“融合式”方法直接在更细的数据粒度（如新闻类别、内容单词）上同时学习新闻和用户特征，使得兴趣更加细化，但相比于“两段式”方法，其没有得到明确的新闻表示与用户表示，因此可解释性更弱，且同样具有推荐结果缺乏新颖性的局限性。一些工作将新闻和用户相关数据编码为多个向量并堆叠形成字符级 CNN，其实验结果中 R 和 MRR 在多个数据集上均有提升；一些工作使用用户历史新闻单词和候选新闻单词之间的语义相似度作为 3-D CNN 的输入，其推荐效果在 HR 和 NDCG 上有所提升。表 4 对比总结了“融合式”新闻推荐方法中不同模型的关键技术、数据种类、数据粒度、方法特点、优点和局限性等。

Table 4 Comparison of typical models in “fusion” DNR

表 4 “融合式”新闻推荐方法中各代表模型对比

类别	代表模型	发表期刊/会议	关键技术	数据种类	数据粒度	方法特点	优点	局限性
“融合式”新闻推荐方法	Deep-JoNN <sup>[19]</sup>	HT'18	CNN; LSTM	新闻类别 新闻关键字 新闻实体 新闻 ID 用户 ID	字符级	分层耦合两种不同的神经网络对新闻和用户联合建模	能够同时学习新闻上下文特征和点击流中的时序兴趣特征	缺乏对全面细致的新闻语义信息的挖掘利用
	3-D CNN <sup>[45]</sup>	ICDMW'17	3-D CNN	文本内容	词级	使用 3-D CNN 学习新闻特征与用户点击历史中的时序兴趣特征		捕捉时序兴趣的时间窗口大小有一定局限性
	Weave&Rec <sup>[47]</sup>	CIKM'18	3-D CNN; 2-D CNN	文本内容	词级			

2.3 DNR 中的“协同式”方法

在基于深度学习的新闻推荐方法中，一些方法不仅从用户自身的点击行为挖掘兴趣，还将其他用户与新闻之间的交互纳入模型，即考虑了相邻用户（或新闻）之间的关联性。本文将此类基于深度学习的新闻推荐总结为“协同式”新闻推荐方法。根据其基于的数据组织结构不同，本节将按照基于用

户-新闻矩阵和基于用户-新闻二分图（bipartite graph）两种类型分别介绍。

基于用户-新闻矩阵

用户-新闻矩阵是描述用户与新闻之间的交互矩阵，用户对新闻的点击等行为代表用户与该新闻存在交互，反之无交互，用 0 填充。由于用户只能与有限数量的新闻进行交互，因此原始的用户-新闻

矩阵往往非常稀疏。文献[48]基于用户-新闻矩阵利用栈式降噪自动编码器（Stacked Denoising Auto-Encoder, 简称 SDAE）实现了新闻推荐。SDAE 由多层稀疏式 AE 组成, 具有较强的特征提取能力。模型使用 SDAE 从原始稀疏的用户-项目矩阵中提取有用的低维特征对用户建模, 之后将提取的特征采用余弦相似度方法计算用户之间的相似性得到用户 top-N 推荐, 实验结果中准确率（Precision, 简称 P）有显著提升。该方法对于新闻与用户特征提取上展示出更强的能力, 提取的特征更具有鲁棒性, 能在一定程度上避免噪声影响, 有助于提高新闻推荐准确率。但该方法仅基于用户-新闻矩阵, 忽略了对新闻的文本内容等信息的利用, 而由于新闻具有高时效性, 新闻条目经常被替换, 所以该方法无法解决冷启动问题。

### 基于用户-新闻二分图

用户-新闻二分图是一种描述不同用户与新闻之间交互行为的图结构, 如图 21 左图所示, 在新闻推荐中, 用户是一类节点, 新闻是一类节点, 利用由用户节点和新闻节点组成的二分图, 可以通过相邻用户为目标用户推荐可能感兴趣的新闻。将用户-新闻二分图展开后可得到用户-新闻高阶连接图(如图 22 右图所示), 利用高阶连接关系能够得到更多的潜在在用户和新闻的交互中的协作信号 (collaborative signal)。

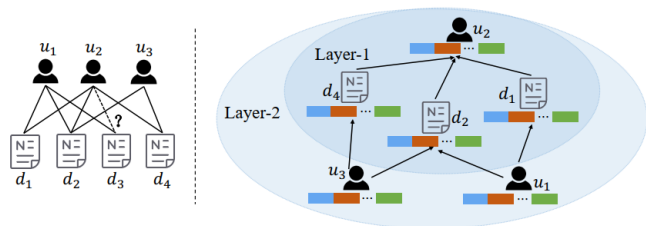


Fig.21 An illustration of user-news interaction graph and high-order connectivity<sup>[49]</sup>

图 21 用户-新闻交互二分图和高阶连接图示<sup>[49]</sup>

文献[49]为揭示导致一个用户点击不同新闻的潜在偏好因素, 在用户-新闻二分图上提出一种带有无监督偏好解离的图神经新闻推荐方法 GNUD

(Graph neural News recommendation with Unsupervised preference Disentanglement)。GNUD 结构如图 22 所示, 通过沿图的信息传播将高阶连接关系编码到用户和新闻的表示中, 利用邻域路由机制 (neighborhood routing mechanism) 对表示中的潜在偏好因素进行解离合<sup>[50]</sup> (即识别和解离隐藏在数据中的不同潜在解释因素), 并设计了一个偏好正则器来帮助提高新闻和用户解离合表示的质量。GNUD 不仅可动态识别可能导致用户点击某新闻的潜在偏好因素, 而且可相应地将新闻分配到一个可卷积特定该因素特征的子空间, 有利于提高特征表示的表达性和可解释性。另外, 该模型沿用了 DAN 中的 PCNN 方法基于新闻标题和新闻概要学习新闻语义表示, 在考虑用户-新闻之间的协作信息的同时还兼顾了新闻内容信息。

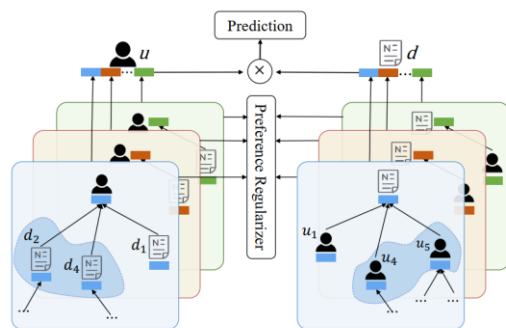


Fig.22 The architecture of GNUD approach<sup>[49]</sup>

图 22 GNUD 方法的结构<sup>[49]</sup>

和文献[49]类似, 文献[51]同样通过引入用户与新闻之间的高阶连接性来增强用户和新闻的学习表示, 提出了一个用户新闻推荐的图增强表示学习方法 GERL (Graph Enhanced Representation Learning for news recommendation)。GERL 包括一个单跳交互学习模块和一个两跳图学习模块, 如图 23 所示。在单跳交互学习模块中, 通过 transformer 架构形成新闻语义表示, 其中的多头自注意力网络可以对标题中的长距与短距单词依赖进行编码; 它还通过聚合用户点击的新闻来学习用户的表示。在两跳图学习模块中, 通过图注意力网络 (Graph Attention

Network, 简称 GAN) 聚合用户和新闻的邻居嵌入, 增强了用户和新闻的特征的表达能力, 提高了推荐的准确性。不过, 用户的兴趣往往随着时间产生变化, GERL 和 GUND 均没有考虑用户-新闻交互的顺序信息, 其推荐的性能还有待提升。

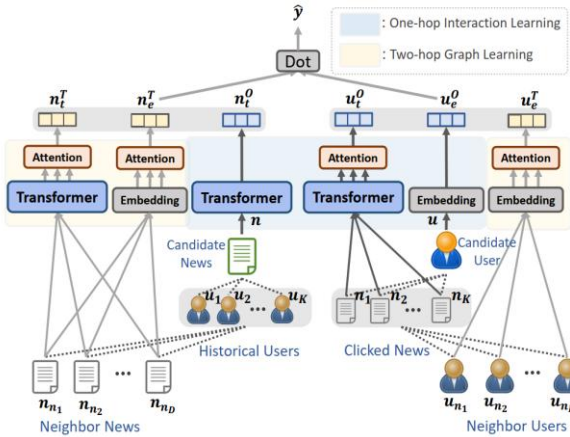


Fig.23 The architecture of GERL approach<sup>[51]</sup>  
图 23 GERL 方法的结构<sup>[51]</sup>

小结

总的来说,“协同式”新闻推荐方法主要利用相邻用户与新闻之间的交互, 对用户和新闻特征进行增强。或基于用户-新闻矩阵, 使用深度学习技术来增强其特征的鲁棒性, 其推荐结果在 P 和 R 指标上有显著提升; 或基于用户-新闻二分图, 编码用户与新闻交互背后的高阶关系, 以丰富用户和新闻特征表示, 提升了 AUC 的数值结果。“协同式”方法因其纳入了其他用户与新闻的交互关系, 使得推荐结果相较于以上 2 种方法更具有新颖性, 但是, 目前大部分的“协同式”方法都缺少用户点击历史中顺序信息的利用, 难以建模用户兴趣的动态变化。表 5 对比总结了“协同式”新闻推荐方法中不同模型的关键技术、数据种类、数据粒度、方法特点、优点和局限性等。

Table 5 Comparison of typical models in “collaboration” DNR  
表 5 “协同式”新闻推荐方法中各代表模型对比

类别	代表模型	发表期刊/会议	关键技术	数据种类	数据粒度	方法特点	优点	局限性
基于用户-新闻矩阵	文献[48]	ICIS'17	SDAE	用户-新闻矩阵	无	使用 SDAE 从原始稀疏矩阵中提取有用的低维特征	新闻与用户特征更具鲁棒性	未利用新闻文本语义信息; 无法解决冷启动问题
基于用户-新闻二分图	GNUM <sup>[49]</sup>	ACL'20	GCN	新闻标题 新闻概要	词级	编码用户-新闻交互的高阶关系到新闻和用户表示中; 解离一个用户点击不同新闻的潜在偏好因素	提高特征表示的表达性和可解释性	未考虑用户-新闻交互的顺序性, 未挖掘用户兴趣变化
	GERL <sup>[51]</sup>	WWW'20	MHSA; AM; GAN	新闻标题	词级	既利用了新闻文本的丰富的语义信息, 又利用了用户-新闻交互的高阶关系	同时丰富新闻特征与用户兴趣特征	

3 基于深度学习的新闻推荐的数据集、基线算法及评价指标

进行新闻推荐实验时要获得公平、客观的评价, 其充分条件在于使用合理的数据集和统一的评价指

标。本章主要介绍一些在 DNR 实验中常用的数据集、基线算法以及衡量算法性能的指标。

### 3.1 新闻数据集

DNR 整个过程由收集数据开始,需要采集的数据主要包括新闻相关数据(新闻的标题、类别、摘要、正文等)和用户相关数据(用户的点击历史、用户的时间信息、位置信息等)。研究者们通常会根据自身的实验条件及实验环境选择一些公共数据集进行实验。我们在表 6 中整理统计了近几年 DNR 研究中出现过的公开数据集。

在新闻推荐研究中经常使用的公共数据集主要包括:

(1) Adressa<sup>[52]</sup>。由挪威 Adresseavisen 新闻出版社和挪威科技大学共同发布,采集自

www.adresseavisen.no 新闻网站。该数据集包含 Adressa 1G 和 Adressa 16G 两个版本,其中,Adressa1G 包含 2017 年 1 月 1 日-1 月 7 日共 1 周内的 11 207 篇新闻、561 733 个用户和 2 286 835 条阅读记录,Adressa 16G 包含 2017 年 1 月 1 日-3 月 31 日共 3 个月内的 48 486 篇新闻、3 083 438 个用户和 27 223 576 条阅读记录。

(2) Plista<sup>[53]</sup>。发布于 RecSys2013 新闻推荐挑战赛,该数据集包含德国 13 个新闻门户网站在 2013 年 6 月 1 日-30 日内的 1 095 323 篇新闻、14 897 978 个用户和 84 210 795 条阅读记录。

Table 6 Statistics of common public datasets of NR

表 6 常用于新闻推荐的公开数据集统计信息

数据集	项目数量及类型	用户数量	交互记录数量	下载链接
Adressa 1G	11 207 新闻	561 733	2 286 835	<a href="http://reclab.idi.ntnu.no/dataset/">http://reclab.idi.ntnu.no/dataset/</a>
Adressa 16G	48 486 新闻	3 083 438	27 223 576	<a href="http://reclab.idi.ntnu.no/dataset/">http://reclab.idi.ntnu.no/dataset/</a>
Plista	1 095 323 新闻	14 897 978	84 210 795	<a href="http://www.dai-labor.de/en/irml/datasets/plistane_wsrecommendationdataset/">http://www.dai-labor.de/en/irml/datasets/plistane_wsrecommendationdataset/</a>
Last.fm	17 632 艺术家	1 892	92 834	<a href="https://grouplens.org/datasets/hetrec-2011/">https://grouplens.org/datasets/hetrec-2011/</a>
MovieLens 1M	3 952 电影	6 040	1 000 209	<a href="https://grouplens.org/datasets/movielens/1m/">https://grouplens.org/datasets/movielens/1m/</a>
MovieLens 10M	10 681 电影	71 567	10 000 054	<a href="https://grouplens.org/datasets/movielens/10m/">https://grouplens.org/datasets/movielens/10m/</a>
MovieLens 20M	27 278 电影	138 493	20 000 263	<a href="https://grouplens.org/datasets/movielens/20m/">https://grouplens.org/datasets/movielens/20m/</a>
MIND	161 013 新闻	1 000 000	24 155 470	<a href="https://msnews.github.io">https://msnews.github.io</a>
MIND-sm all	93 698 新闻	50 000	230 117	<a href="https://msnews.github.io">https://msnews.github.io</a>

(3) Last.fm<sup>[54]</sup>。由西班牙马德里自治大学的信息检索小组创建,发布于第二届推荐系统信息异构与融合国际研讨会(The 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems)。Last.fm 数据集采集自同名社交音乐平台,主要用于音乐推荐,在新闻推荐研究

中也常被研究者用于评估新闻推荐方法的通用性。数据集中包含 1 892 个用户、17 632 个艺术家和 92 834 条收听记录。此外,该数据集还含有用户社交网络信息。

(4) MovieLens<sup>[55]</sup>。由美国明尼苏达大学的 GroupLens 研究小组发布的电影评分数据集,常被



用来评估新闻推荐方法的通用性。该数据集常用的版本为 MovieLens 1M、MovieLens 10M 和 MovieLens 20M 三个版本,其中,MovieLens 1M 包含 3 952 部电影、6 040 个用户和 1 000 209 个评分(用户对其看过的电影进行 1-5 分评价);MovieLens 10M 包含 10 681 部电影、71 567 个用户和 10 000 054 个评分;MovieLens 20M 包含 27 278 部电影、138 493 个用户和 20 000 263 个评分。

(5) MIND<sup>[24-26, 56]</sup>。采集自微软新闻网站的匿名行为日志的大型新闻推荐数据集。它包含 MIND 和 MIND-small 两个版本,其中,MIND 含有 161 031 篇新闻、1 000 000 个用户和 24 155 470 条行为日志,每篇新闻中包含标题、摘要、正文、类别和实体,每个行为日志中点击事件、未点击事件和该行为日志前的历史点击新闻;轻量级的 MIND-small 则包含 93 698 篇新闻、50 000 个用户和 230 117 条行为日志。

最近几年也有一些新闻推荐研究团队提供了数据集用于自己团队内部评测,包括:

(1) Bing News<sup>[32, 39]</sup>。采集自 Bing News 在线新闻网站,每条阅读记录主要包含时间戳,用户 ID,新闻 URL,新闻标题和点击次数等属性。该数据集一般为微软亚洲研究院研究团队所使用。

(2) NAVER News<sup>[42]</sup>。采集自韩国最大的搜索引擎和门户网站 NAVER,包括 8 620 854 个用户和 5 759 377 篇新闻,其中每条阅读记录包括其用户加密登录 ID、会话 ID、时间戳、新闻文章 ID 等属性。该数据集一般为韩国 NAVER 公司研究团队所使用。

(3) Yahoo! JAPAN<sup>[23]</sup>。采集自 Yahoo! JAPAN 的移动端主页,包括约 2 000 000 篇新闻、约 12 000 000 个用户和约 100 000 000 个阅读记录。该数据集一般为日本雅虎公司研究团队所使用。

### 3.2 基线算法

基线(Baseline)算法是研究论文中用于实验结果对比的前人算法。我们将 DNR 领域常用的基线分为两类,如表 7 所示,一类是传统的基于统计机器学习的新闻推荐方法,如 POP(Popular-based method)、Item KNN、BPR(Bayesian Personalized Ranking)、LibFM 等。另一类是基于深度学习的新闻推荐方法如 DSSM(Deep Structured Semantic Models)、Wide&Deep、DeepFM(a Factorization-Machine based neural network)、YouTubeNet、RA-DSSM(Recurrent Attention Deep Structured Semantic Models)、HRNN、NeuMF(Neural Matrix Factorization)等。

Table 7 Common baselines for DNR

表 7 DNR 中常用的基线算法

引用论文	基线算法	说明	发表会议/期刊	代码下载链接
[19, 21, 33, 42, 43, 45]	POP	一种基于用户-项目最大交互次数的推荐方法	N/A <sup>8</sup>	N/A
[19, 21, 42, 45]	Item KNN <sup>[57]</sup>	K 最近邻分类算法,利用余弦距离来计算文章之间相似性	WWW'01	N/A
[19, 21, 43, 47]	BPR <sup>[58]</sup>	一种基于用户隐式反馈的贝叶斯个性化排序算法	UAI'09	<a href="https://github.com/gamboviol/bpr">https://github.com/gamboviol/bpr</a>
[22, 24-28, 32, 36, 40, 49, 51]	LibFM <sup>[59]</sup>	基于特征矩阵分解的推荐方法	TIST'12	<a href="https://github.com/srendle/libfm">https://github.com/srendle/libfm</a>
[22, 24-28, 40, 49, 51]	DSSM <sup>[60]</sup>	一种用于文档排序的深度结构化语义模型	CIKM'13	<a href="https://github.com/InsaneLife/dssm">https://github.com/InsaneLife/dssm</a>

<sup>8</sup> N/A 表示相关论文未提及、无法确定该信息

[18, 22, 24-28, 32, 33, 40, 49, 51, 61]	Wide&Deep <sup>[62]</sup>	一种被广泛使用的结合了宽度线性通道和深度非线性通道的深度推荐方法	DLRS'16	<a href="https://github.com/tensorflow/tensorflow/blob/r1.4/tensorflow/examples/learn/wide_n_deep_tutorial.py">https://github.com/tensorflow/tensorflow/blob/r1.4/tensorflow/examples/learn/wide_n_deep_tutorial.py</a>
[22, 25-28, 36, 40, 49]	DeepFM <sup>[63]</sup>	一种被广泛使用的结合因子分解机和深度神经网络的推荐方法	IJCAI'17	<a href="https://github.com/princewen/tensorflow_practice/tree/master/recommendation/Basic-DeepFM-model">https://github.com/princewen/tensorflow_practice/tree/master/recommendation/Basic-DeepFM-model</a>
[32, 39]	YouTubeNet <sup>[37]</sup>	使用深度候选生成网络和深度排名网络对 YouTube 大规模候选集进行视频推荐的方法	RecSys'16	N/A
[44, 47]	RA-DSSM <sup>[64]</sup>	一种使用基于注意力的双向 RNN 解决用户兴趣变化的问题并采用 DSSM 捕获用户与物品之间的相似度的推荐方法	CLEF'17	N/A
[19, 21]	HRNN <sup>[65]</sup>	一种具有跨会话信息无缝传输的个性化 RNN 模型	RecSys'17	<a href="https://github.com/mquad/hgru4rec">https://github.com/mquad/hgru4rec</a>
[43, 47]	NeuMF <sup>[66]</sup>	一种神经矩阵分解模型。它将隐式反馈作为二分类问题，建立推荐模型	WWW'17	<a href="https://github.com/hexiangnan/neural_collaborative_filtering">https://github.com/hexiangnan/neural_collaborative_filtering</a>

### 3.3 评价指标

随着推荐技术的发展和数据处理能力的提高，当前 NR 领域中评价指标包含准确度、多样性、新颖性和满意度等多种。其中，准确度用来衡量推荐算法在多大程度上能够准确预测用户对新闻的感兴趣程度，是衡量推荐算法最基础的指标<sup>[67]</sup>。事实上，由于多样性、新颖度的计算公式较为简单粗糙，而满意度往往需要通过用户在线调查获得，故而目前大多数研究方法都以提升算法准确度为研究目标。目前 NR 领域常用的实验性能评价指标包括：准确率（Precision，简称 P）、召回率（Recall，简称 R）、准确率和召回率调和平均值 F1、ROC（Receiver Operating Characteristic）曲线下的面积（Area Under the ROC curve，简称 AUC）、命中率（Hit Ratio，简称 HR）、平均倒数排名（Mean Reciprocal Rank，简称 MRR）和归一化折扣累计效益（Normalized Discounted Cumulative Gain，简称 NDCG）等。其中，准确率 P 表示用户对推荐结果的点击概率；召回率 R 表示用户感兴趣的新闻被推荐的概率，往往与准确率 P 呈负相关性；F1 能够综合推荐结果的准确率 P 和召回率 R 提供更全面的评估。AUC 表示一个推荐系统能够在多大程度上将用户感兴趣的新闻

与不感兴趣的新闻区分出来<sup>[68]</sup>，可综合衡量算法的整体表现，不仅适合有明确“喜欢/不喜欢”评价机制的推荐系统，也适合没有明确喜好阈值的推荐系统（如 5 分制评价）。HR@n 用于直观衡量测试新闻是否存在于推荐列表的前 n 名中。MRR 和 NDCG 均为衡量推荐列表中排序准确度的指标，目的都是通过考察推荐结果的排序位置来检验用户的体验程度，区别在于两者的计算方式不同——MRR 是通过累加相关结果的排序位置倒数而得，而 NDCG 则是按照 log 调和级数形式进行计算。在实际研究中，为了能更公正有效地分析推荐效果，通常采用两种或两种以上的评价指标综合分析推荐算法的表现，例如将 R 和 MRR 组合评价<sup>[19, 42]</sup>，将 AUC、MRR 和 NDCG 组合评价<sup>[23]</sup>等。

## 4 总结和展望

如今，深度学习技术已经广泛应用到新闻推荐领域。本文主要对基于深度学习的新闻推荐方法研究现状进行了分类、梳理和总结。根据对新闻和用户的建模思路不同，本文将基于深度学习的新闻推荐方法（DNR）分为“两段式”、“融合式”和“协同式”3类，对每类方法的代表性算法进行了研究、

分析和对比,并概括总结了每类方法的技术特点和优缺点。在现有研究成果的基础上,我们总结 DNR 研究领域的重点问题和发展趋势,认为该领域还存在以下一些具有挑战性的研究方向:

#### (1) 精准的用户建模

精准的用户建模一直是 DNR 领域的挑战,探索如何对用户兴趣建模并实现更为准确的推荐一直是一个富有挑战的研究焦点。已有的新闻推荐方法往往仅依赖用户的新闻点击历史对用户建模,但这种单一类型的用户表示难以较为全面地提取用户兴趣特征。除新闻点击历史外,用户的搜索引擎搜索历史和网页浏览历史等多种信息中也包含了丰富的用户兴趣,可以作为用户兴趣建模的辅助信息,如文献[39]利用 AM,对从新闻阅读、网页浏览等多个渠道收集的用户数据进行融合,丰富用户特征。因此,如何综合多种用户行为数据进行用户兴趣建模是一个具有挑战性的问题。此外,用户的兴趣具有动态变化的特点,用户不仅对一些方面具有长期偏好,还对另一些方面具有短期兴趣,这种短期兴趣往往和用户所处的时间空间等上下文因素相关,特别是移动新闻推荐中, NR 过程与用户所在的地理位置以及用户的运动轨迹有重要联系<sup>[69]</sup>,文献[21]利用 AM 将用户阅读新闻时的时间、地点融入用户特征,获得用户兴趣表示。因此,如何对用户的长期偏好和短期兴趣精准刻画以及如何协同二者进行新闻推荐同样是值得深入研究方向。

#### (2) 基于图结构的新闻推荐

图结构是一种非线性的复杂数据结构,在 NR 领域,图结构通常被用来建立多个用户与多个新闻之间的交互关系,图结构中所反映的用户与新闻之间的高阶连接性蕴含着丰富的特征信息。近来,图神经网络在图结构上的学习能力逐渐凸显,它能够自然地将节点信息和拓扑结构集成起来,有效建模节点之间的依存关系。在 DNR 中,图神经网络在建模用户与新闻之间的高阶连接性上具有强大能力。文献[51]设计了两跳图学习模块,通过图注意力网络聚合用户和新闻的邻居嵌入,增强用户和新

闻的表示。文献[49]在用户-新闻二分图的基础上,对用户的潜在偏好因素解耦合,提高了表示的表达性和可解释性。事实上,图结构能够融合推荐系统中的多源异构数据,特别有助于个性化推荐。目前有关基于图结构的 NR 研究还比较少,如何基于图结构结合 DL 技术并将其有效应用于 NR 也是目前研究人员亟需解决的问题。

#### (3) 新闻推荐中的虚假新闻传播

由于在线新闻传播相较于传统新闻成本较低,所以一些虚假新闻或低质量新闻大量涌入了读者视野。假新闻的传播会对个人和社会产生严重的负面影响<sup>[70]</sup>,破坏新闻生态系统。假新闻的存在也严重影响了 NR 的推荐效果,一方面,假新闻增加了新闻文章数量,使得原有真实新闻又多了一些“假”的版本,增加了 NR 的计算成本。另一方面,假新闻含有比较明显的错误信息,若推荐此类新闻,将极大降低读者对 NR 的推荐满意度。当前,已有一些研究人员开展虚假新闻检测研究和新闻质量识别研究,文献[71]提出一个基于 RNN 的虚假新闻检测模型,能够自动化预测虚假新闻。文献[72]基于 GRU 研究了假新闻文章、假新闻创作者和假新闻主题之间的关系,实现了一个虚假新闻检测模型。文献[73]基于用户的阅读率和停留时间等行为自动识别新闻的质量。由于检测假新闻所需的信息通常不可用或不充分,假新闻往往被在线用户广泛分享,检测假新闻仍然是新闻推荐中一个重要但又具有挑战性的问题。

#### (4) 新闻推荐中的隐私和安全

互联网是一个复杂的环境,在对用户进行新闻推荐时候应为用户考虑两点,一是要确保 NR 向用户推荐的新闻或者新闻链接是安全的,二是要确保用户的个人隐私,并且需要在尽可能少利用用户隐私数据的情况下做出准确、合理的推荐。NR 需要建立相应的隐私保护机制,为用户提供用户个人信息安全保障以及保密措施,提高用户对推荐系统的信任度与好感度。文献[74]指出针对敏感数据信息的保护贯穿于网络中每个层次,可以通过修改或隐

藏原始信息的局部或全局敏感数据来保护隐私, 还可以通过加密技术对信息进行保护。目前由于网络发展迅速, 用户接受信息的途径日益增多, 如短信、链接、网页弹窗等, 如何在新闻推荐中全面考虑隐私安全保护机制并建立异常情况解决方案, 也是研究者需要努力的方向。

#### (5) 新闻推荐中的可解释性

推荐的可解释性是指向用户或系统设计者提供解释, 使之知道为什么推荐这些项目, 这有助于提高推荐系统的有效性、说服力和用户满意度<sup>[75]</sup>, 也有助于帮助研究者探索模型的内部机理。深度学习以其强大的表征作用广泛应用于新闻推荐系统并切实提高了推荐准确性, 然而深度学习中的深层神经网络被普遍认为是高度不可解释的, 因此, 可解释的新闻推荐似乎是一项艰巨的任务。近年来受到广泛关注的 AM 在一定程度上为该问题提供了解决方法, 其中的解释性主要来自于注意力权重分配。如文献[76]从商品评级和用户评论中共同学习商品和用户信息, 将评论词的注意力权重作为推荐的解释。新闻推荐模型 NPA<sup>[24]</sup>中, 对词语和新闻级别的注意力权重做了高亮的可视化展示, 可在未来进一步形成推荐的解释。另外, 知识库中包含了丰富的用户和项目信息, 可以帮助推荐生成更直观、更有针对性的解释。如文献[77]提出的 Ripple 网络, 是一种将知识图谱整合到推荐系统中的端到端架构, 其解释可以通过在知识图谱上查找用户和推荐项的路径来提供。在 DNR 领域, 关于对推荐结果做出解释的工作还非常少, 可以依赖已经能够使模型对重要因素进行突出强调的 AM 机制和包含丰富外部知识的知识图谱, 使新闻推荐达到可解释水平。

#### (6) 更统一的评估方法

对 NR 进行有效评估是一个重要课题, 一个统一的评估方法能够更加客观而合理地对推荐结果进行评价。然而, 现有 NR 研究中数据集的来源较为散乱, 缺少一套广泛接受的标准化评估数据集的来源较为散乱, 缺少一套广泛接受的标准化评估数据集, 导致不同研究的评估差别较大。另外, 大部分评价指标只重视准确度, 忽略了推荐结果的惊喜度、

覆盖率等特性, 能否将这些指标结合提出一个综合性的评价指标, 也是 NR 中需要继续深入研究的问题。

## References

- [1] Prawesh S, Padmanabhan B. The "top N" news recommender: count distortion and manipulation resistance[C]. Proceedings of the fifth ACM conference on Recommender systems. ACM, 2011.
- [2] Gulla J A, Yu B, özgöbek Ö, et al. 3rd international workshop on news recommendation and analytics (INRA 2015)[C]. Proceedings of the 9th ACM Conference on Recommender Systems. 2015.
- [3] Zhao X, Wang C, Chen M, et al. AutoEmb: Automated Embedding Dimensionality Search in Streaming Recommendations[J]. arXiv preprint arXiv:2002.11252. 2020.
- [4] Xu J, He X, Li H. Deep learning for matching in search and recommendation[C]. The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 2018.
- [5] Kazai G, Yusof I, Clarke D. Personalised news and blog recommendations based on user location, Facebook and Twitter user profiling[C]. Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. 2016.
- [6] Karkali M, Pontikis D, Vazirgiannis M. Match the news: A firefox extension for real-time news recommendation[C]. Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. 2013.
- [7] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. nature. 2015, 521(7553): 436-444.
- [8] Zhang S, Yao L, Sun A, et al. Deep learning based recommender system: A survey and new perspectives[J]. ACM Computing Surveys (CSUR). 2019, 52(1): 1-38.
- [9] Huang L, Jiang B, Lv S, et al. Survey on Deep Learning Based Recommender Systems[J]. Chinese Journal of Computers. 2018, 41(07): 1619-1647.
- [10] Wang S, Li X, Sun 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.
- [11] Li M, Wang L. A Survey on Personalized News Recom-

- mendation Technology[J]. IEEE Access. 2019, 7: 145861-145879.
- [12] Liang S, Zhang C, Cao L, et al. Collaborative Joint Embedding Based on Personalized News Recommendation[J]. Journal of Chinese Information Processing. 2018, 32(11): 72-78.
- [13] Liu J, Dolan P, Pedersen E R. Personalized news recommendation based on click behavior[C]. Proceedings of the 15th international conference on Intelligent user interfaces. 2010.
- [14] Kelly D, Teevan J. Implicit feedback for inferring user preference: a bibliography[C]. Acn Sigir Forum. ACM New York, NY, USA, 2003.
- [15] Wu X, Huang B, Fang Z, et al. Application of sequence generative adversarial network in recommendation system[J]. Computer Engineering and Applications. 2020, 56(23): 175-179.
- [16] Peng Y, Zhu W, Zhao Y, et al. Cross-media analysis and reasoning: advances and directions[J]. Frontiers of Information Technology & Electronic Engineering. 2017, 18(1): 44-57.
- [17] Chaudhari S, Polatkan G, Ramanath R, et al. An attentive survey of attention models[J]. arXiv preprint arXiv:1904.02874. 2019.
- [18] Wang X, Yu L, Ren K, et al. Dynamic attention deep model for article recommendation by learning human editors' demonstration[C]. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017.
- [19] Zhang L, Liu P, Gulla J A. A deep joint network for session-based news recommendations with contextual augmentation[M]. Proceedings of the 29th on Hypertext and Social Media, 2018, 201-209.
- [20] de Souza Pereira Moreira G. CHAMELEON: a deep learning meta-architecture for news recommender systems[C]. Proceedings of the 12th ACM Conference on Recommender Systems. 2018.
- [21] Zhang L, Liu P, Gulla J A. Dynamic attention-integrated neural network for session-based news recommendation[J]. Machine Learning. 2019, 108(10): 1851-1875.
- [22] Zhu Q, Zhou X, Song Z, et al. Dan: Deep attention neural network for news recommendation[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2019.
- [23] Okura S, Tagami Y, Ono S, et al. Embedding-based news recommendation for millions of users[C]. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017.
- [24] Wu C, Wu F, An M, et al. Npa: Neural news recommendation with personalized attention[C]. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019.
- [25] An M, Wu F, Wu C, 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. 2019.
- [26] Wu C, Wu F, An M, et al. Neural news recommendation with attentive multi-view learning[J]. arXiv preprint arXiv:1907.05576. 2019.
- [27] Wu C, Wu F, Ge S, 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). 2019.
- [28] Wu C, Wu F, Qi T, et al. User Modeling with Click Preference and Reading Satisfaction for News Recommendation[C]. Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence (IJCAI-PRICAI-20) . 2020.
- [29] Yang B, Mitchell T. Leveraging knowledge bases in lstms for improving machine reading[J]. arXiv preprint arXiv:1902.09091. 2019.
- [30] Wang J, Wang Z, Zhang D, et al. Combining Knowledge with Deep Convolutional Neural Networks for Short Text Classification.[C]. IJCAI. 2017.
- [31] Xu C, Bai Y, Bian J, et al. Rc-net: A general framework for incorporating knowledge into word representations[C]. Proceedings of the 23rd ACM international conference on conference on information and knowledge management. 2014.
- [32] Wang H, Zhang F, Xie X, et al. DKN: Deep knowledge-aware network for news recommendation[C]. Proceedings of the 2018 world wide web conference. 2018.
- [33] Gao J, Xin X, Liu J, et al. Fine-Grained Deep Knowledge-Aware Network for News Recommendation with Self-Attention[C]. 2018 IEEE/WIC/ACM International



- Conference on Web Intelligence (WI). IEEE, 2018.
- [34] Xu B, Cen K, Huang J, et al. A Survey on Graph Convolutional Neural Network[J]. Chinese Journal of Computers. 2020, 5(43): 755-780.
- [35] Sheu H, Li S. Context-aware Graph Embedding for Session-based News Recommendation[C]. Fourteenth ACM Conference on Recommender Systems. 2020.
- [36] Lee D, Oh B, Seo S, et al. News Recommendation with Topic-Enriched Knowledge Graphs[C]. Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020.
- [37] Covington P, Adams J, Sargin E. Deep neural networks for youtube recommendations[C]. Proceedings of the 10th ACM conference on recommender systems. 2016.
- [38] Wang X, He X, Nie L, et al. Item silk road: Recommending items from information domains to social users[C]. Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval. 2017.
- [39] Lian J, Zhang F, Xie X, et al. Towards Better Representation Learning for Personalized News Recommendation: a Multi-Channel Deep Fusion Approach.[C]. IJCAI. 2018.
- [40] Wang H, Wu F, Liu Z, et al. Fine-grained Interest Matching for Neural News Recommendation[C]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.
- [41] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. MIT press, 2016.
- [42] Park K, Lee J, Choi J. Deep neural networks for news recommendations[C]. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017.
- [43] Khattar D, Kumar V, Varma V, et al. HRAM: A hybrid recurrent attention machine for news recommendation[C]. Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018.
- [44] Chu Q, Liu G, Sun H, et al. Next News Recommendation via Knowledge-Aware Sequential Model[C]. China National Conference on Chinese Computational Linguistics. Springer, 2019.
- [45] Kumar V, Khattar D, Gupta S, et al. Word semantics based 3-d convolutional neural networks for news recommendation[C]. 2017 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, 2017.
- [46] Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE transactions on pattern analysis and machine intelligence. 2012, 35(1): 221-231.
- [47] Khattar D, Kumar V, Varma V, et al. Weave&rec: A word embedding based 3-d convolutional network for news recommendation[C]. Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018.
- [48] Cao S, Yang N, Liu Z. Online news recommender based on stacked auto-encoder[C]. 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). IEEE, 2017.
- [49] Hu L, Xu S, 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. 2020.
- [50] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives[J]. IEEE transactions on pattern analysis and machine intelligence. 2013, 35(8): 1798-1828.
- [51] Ge S, Wu C, Wu F, et al. Graph Enhanced Representation Learning for News Recommendation[C]. Proceedings of The Web Conference 2020. 2020.
- [52] Gulla J A, Zhang L, Liu P, et al. The Adressa dataset for news recommendation[C]. Proceedings of the international conference on web intelligence. 2017.
- [53] Kille B, Hopfgartner F, Brodt T, et al. The plista dataset[C]. Proceedings of the 2013 International News Recommender Systems Workshop and Challenge. 2013.
- [54] Cantador I, Brusilovsky P, Kuflik T. Second workshop on information heterogeneity and fusion in recommender systems (HetRec2011)[C]. Proceedings of the fifth ACM conference on Recommender systems. 2011.
- [55] Harper F M, Konstan J A. The movielens datasets: History and context[J]. Acm transactions on interactive intelligent systems (tiis). 2015, 5(4): 1-19.
- [56] Wu F, Qiao Y, Chen J, et al. Mind: A large-scale dataset for news recommendation[C]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.
- [57] Sarwar B, Karypis G, Konstan J, et al. Item-based collabor-

- ative filtering recommendation algorithms[C]. Proceedings of the 10th international conference on World Wide Web. 2001.
- [58] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback[J]. arXiv preprint arXiv:1205.2618. 2012.
- [59] Rendle S. Factorization machines with libfm[J]. ACM Transactions on Intelligent Systems and Technology (TIST). 2012, 3(3): 1-22.
- [60] Huang P, He X, Gao J, et al. Learning deep structured semantic models for web search using clickthrough data[C]. Proceedings of the 22nd ACM international conference on Information & Knowledge Management. 2013.
- [61] Zheng G, Zhang F, Zheng Z, et al. DRN: A deep reinforcement learning framework for news recommendation[C]. Proceedings of the 2018 World Wide Web Conference. 2018.
- [62] Cheng H, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems[C]. Proceedings of the 1st workshop on deep learning for recommender systems. 2016.
- [63] Guo H, Tang R, Ye Y, et al. DeepFM: a factorization-machine based neural network for CTR prediction[J]. arXiv preprint arXiv:1703.04247. 2017.
- [64] Kumar V, Khattar D, Gupta S, et al. Deep Neural Architecture for News Recommendation.[C]. CLEF (Working Notes). 2017.
- [65] Quadrana M, Karatzoglou A, Hidasi B, et al. Personalizing session-based recommendations with hierarchical recurrent neural networks[C]. Proceedings of the Eleventh ACM Conference on Recommender Systems. 2017.
- [66] He X, Liao L, Zhang H, et al. Neural collaborative filtering[C]. Proceedings of the 26th international conference on world wide web. 2017.
- [67] Zhu Y, Lv L. Evaluation Metrics for Recommender Systems [J]. Journal of University of Electronic Science and Technology of China M. 2012, 41(02): 163-175.
- [68] Hanley J A, Mcneil B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve.[J]. Radiology. 1982, 143(1): 29-36.
- [69] Meng X, Chen C, Zhang Y. A Survey of Mobile News Recommend Techniques and Applications[J]. Chinese Journal of Computers. 2016, 39(04): 685-703.
- [70] Qin X. Causes and Countermeasures of false news: Taking the top ten fake news in 2017 as an example[J]. New Media Research. 2018, 4(13): 19-20.
- [71] Ruchansky N, Seo S, Liu Y. Csi: A hybrid deep model for fake news detection[C]. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017.
- [72] Zhang J, Cui L, Fu Y, et al. Fake news detection with deep diffusive network model[J]. arXiv preprint arXiv:1805.08751. 2018.
- [73] Lu H, Zhang M, Ma W, et al. Quality effects on user preferences and behaviors in mobile news streaming[C]. The World Wide Web Conference. 2019.
- [74] Zhang H, Han W, Lai X, et al. A Survey on Cyberspace Security[J]. Scientia Sinica(Informationis). 2016, 46(02): 125-164.
- [75] Zhang Y, Chen X. Explainable recommendation: A survey and new perspectives[J]. arXiv preprint arXiv:1804.11192. 2018.
- [76] Lu Y, Dong R, Smyth B. Coevolutionary recommendation model: Mutual learning between ratings and reviews[C]. Proceedings of the 2018 World Wide Web Conference. 2018.
- [77] Wang H, Zhang F, Wang J, et al. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems[C]. Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018.

## 附中文参考文献:

- [9] 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.
- [10] 王绍卿, 李鑫鑫, 孙福振, 等. 个性化新闻推荐技术研究综述[J]. 计算机科学与探索. 2020, 14(1): 18-29.
- [12] 梁仕威, 张晨蕊, 曹雷, 等. 基于协同表示学习的个性化新闻推荐[J]. 中文信息学报. 2018, 32(11): 72-78.
- [15] 伍鑫, 黄渤, 方志军, 等. 序列生成对抗网络在推荐系统中的应用[J]. 计算机工程与应用. 2020, 56(23): 175-179.
- [34] 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述[J]. 计算机学报. 2020, 43(05): 755-780.

[67] 朱郁筱, 吕琳媛. 推荐系统评价指标综述[J]. 电子科技大学学报. 2012, 41(02): 163-175.

[69] 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.

[70] 秦希屏. 虚假新闻的成因及其应对策略——以2017年十大假新闻为例[J]. 新媒体研究. 2018, 4(13): 19-20.

[74] 张焕国, 韩文报, 来学嘉, 等. 网络空间安全综述[J]. 中国科学:信息科学. 2016, 46(02): 125-164.



田萱 (1976-), 女, 山东济宁人, 2008 年于中国人民大学获得博士学位, 现为北京林业大学副教授, CCF 高级会员, 主要研究领域为智能信息处理, 文本挖掘等。

TIAN Xuan was born in 1976. She received the Ph.D. degree from Renmin University of China in 2008. She is an associate professor at Beijing Forestry University, and the senior member of CCF. Her research interests include intelligent information processing, text mining, etc.



丁琪 (1996-), 女, 山东邹城人, 北京林业大学硕士研究生, 主要研究领域为智能信息处理, 个性化推荐等。

DING Qi was born in 1996. She is an M.S. candidate at Beijing Forestry University. Her research interests include intelligent information processing, personalized recommendation, etc.



廖子慧 (1994-), 女, 重庆忠县人, 北京林业大学硕士研究生, 主要研究领域为智能信息处理, 推荐系统等。

LIAO Zihui was born in 1994. She is an M.S. candidate at Beijing Forestry University. Her research interests include intelligent information processing, recommendation system, etc.



孙国栋 (1981-), 男, 黑龙江哈尔滨人, 2009 年于哈尔滨工业大学获得博士学位, 现为北京林业大学副教授, 主要研究领域为无线传感器网络, 移动计算, 数据挖掘等。

SUN Guodong was born in 1981. He received the Ph.D. degree from Harbin Institute of Technology in 2009. He is an associate professor at Beijing Forestry University. His research interests include wireless sensor network, mobile computing, data mining, etc.