

Neural News Recommendation with Attentive Multi-View Learning

Chuhan Wu , Fangzhao Wu , Mingxiao An , Jianqiang Huang , Yongfeng Huang , Xing Xie

摘要

个性化新闻推荐对于在线新闻平台来说至关重要，能够帮助用户快速找到自己感兴趣的新闻信息、缓解信息过载同时改善用户体验。新闻推荐的关键在于两个方面：新闻表示的学习和用户表示的学习。本文提出的方法的核心在于新闻编码器和用户编码器。新闻编码器使用了一个注意力多视角学习模型来从新闻的标题、正文和主题类别中学习统一的新闻信息表示 [5]，同时把正文、标题和主题类别这些信息看作是新闻的不同视角。此外，文中还在新闻编码器中使用了词级和视角级的注意力机制来选择新闻当中重要的单词和视角。在用户编码器中，基于用户浏览过的新闻并使用注意力机制来挑选信息丰富的新闻来表示用户。在真实数据集中的大量实验表明，本文中提出的方法可以有效提高新闻推荐系统的性能。

关键词：个性化新闻推荐；注意力机制

1 引言

近年来，深度学习发展迅猛，并在许多领域都取得了不错的效果。在这篇论文之前，许多基于深度学习的新闻推荐方法已经提出并取得了一定的效果，比如基于 embedding 的新闻推荐系统 [2] 以及知识感知的 CNN 新闻推荐系统 [4]。但是，这些方法都存在一定的缺陷，它们往往只基于单一的新闻信息来学习新闻表示。要么是仅仅通过新闻正文来学习新闻表示，要么是仅仅基于新闻标题来学习新闻表示。对于准确地学习新闻表示和用户表示来说，从单一的新闻信息中学习可能是不够的，这在一定程度上限制了新闻推荐系统的性能。在新闻推荐系统中，最重要的是学习用户表示和新闻表示。用户表示对用户兴趣和行为进行建模，试图捕捉用户的长期和短期兴趣，以便更加精准地推荐用户可能感兴趣地新闻。新闻表示对新闻内容进行建模，试图捕捉新闻中的关键信息，比如主题、关键词以及情感等，以便更加精准地匹配用户兴趣。

基于以上观察，我们发现新闻中不同种类的新闻信息对于准确地学习新闻表示是非常重要的，因此在学习新闻表示和用户表示时需要综合考虑新闻的标题、正文和主题类别信息。此外，不同种类的新闻信息具有不同的特点，比如标题通常从非常简短而正文通常非常详细，这就需要不同程度地提取各个部分重要的信息。同时，新闻中不同的单词包含的信息量也是不同的，有的单词能够更加准确地表示指定的新闻，因此需要对新闻中重要的单词进行提取。

因此，本文提出了注意力多视角学习的新闻推荐方法以学习更有信息量的新闻表示和用户表示。能够有效解决仅仅提取单一的新闻信息存在的问题，并取得不错的效果。这篇论文吸取了传统的使用深度学习的新闻推荐系统的经验，综合考虑不同的新闻信息，具有较大的复现意义。

2 相关工作

新闻推荐系统的关键在于学习准确的新闻表示和用户表示。许多现有的新闻推荐方法对于新闻表示和用户表示的学习采用手动特征工程。比如说 Liu 提出的使用主题类别和贝叶斯模型产生的兴趣特征作为新闻表示 [1]，Son 提出的从维基百科的页面中提取主题和位置特征作为新闻表示 [3]。这种使用特征工程学习新闻表示的方法往往需要大量的领域知识并且需要很大的工作量，而且这种方法无法获取新闻中上下文的信息以及单词间的顺序。

还有一些近年提出的新闻推荐方法使用深度学习进行新闻表示和用户表示的学习。比如 Okura 提出的使用自编码器从新闻正文中学习新闻表示 [2]，并且使用循环神经网络从用户浏览过的新闻中学习用户表示；Wang 提出的使用卷积神经网络从新闻标题中学习新闻表示 [4]。这些使用深度学习学习出的新闻表示和用户表示虽然学习了新闻中的上下文特征，却没有考虑到新闻中各个部分的作用，仅仅从单一的新闻信息中进行学习。

3 本文方法

3.1 本文方法概述

本文提出的方法采用注意力机制，将新闻中的标题、正文和主题类别当作新闻的不同视角，并综合考虑进行新闻表示和用户表示的学习。该方法主要包括三个模块，分别是基于注意力多视角学习的新闻编码器、使用注意力机制从用户浏览过的新闻中学习的用户编码器和预测用户点击候选新闻的点击预测器。论文的整体框架如图 1 所示：

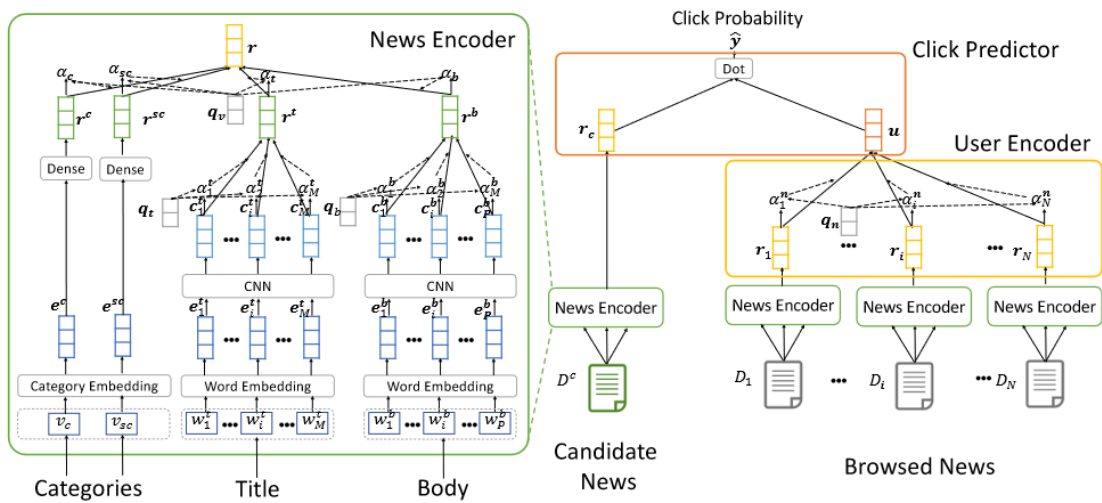


图 1. 注意力多视角学习的新闻推荐系统框架

3.2 新闻编码器

在本文方法中，新闻编码器用来从不同种类的新闻信息中学习新闻表示，将新闻的标题、正文和主题类别所包含的信息融合在一起。该编码器使用注意力多视角学习框架来学习标题、正文和主题类别的统一新闻表示，主要由四个部分组成。即标题编码器、正文编码器、类别编码器和注意力层。

标题编码器这个部分主要是从新闻标题中学习新闻表示，提取出新闻标题中的重要信息。这个编码器由三层组成，分别是词嵌入层、卷积神经网络层和词级的注意力网络。词嵌入层用于将新闻标题的词向量序列转化为低维的语义向量。卷积神经网络用来学习新闻标题单词中的局部上下文信息。同一个新闻标题中的不同新闻单词所包含的信息量是不同的，因此使用词级的注意力网络从新闻标题中挑选出更有信息量的单词 [6]。

正文编码器从新闻的正文中学习新闻表示，提取出新闻正文中的重要信息，和标题编码器一样由词嵌入层、卷积神经网络层和注意力层组成。主题类别编码器用于从新闻的主题类别中提取出重要信息。该编码器由嵌入层和致密层组成。新闻编码器的最后一个部分是注意力网络组成，新闻的标题、正文和主题类别对于这个新闻表示含有的信息量是不同的，使用注意力网络从各个部分中选取对于表示这个新闻包含信息量最多的信息。标题、正文、类别和子类别的注意力权重分别用 $\alpha_t, \alpha_b, \alpha_s, \alpha_{sc}$ 表示，经过编码器得到的标题、正文、主题类别和子类别的表示分别为 r^t, r^b, r^c, r^{sc} 那么最终通过新闻编码器得到的新闻表示为：

$$r = \alpha_c r^c + \alpha_{sc} r^{sc} + \alpha_t r^t + \alpha_b r^b$$

3.3 用户编码器

本文中的用户编码器用来从用户浏览过的新闻中学习用户表示。同一个用户浏览的不同新闻所包含的表示这个用户特征的信息量不同，用户浏览的小众新闻往往更能够反映出用户的喜好。用户编码器采用注意力网络，给用户浏览过的新闻赋予相应的权重，最后权值相加得到用户表示。

3.4 点击预测器

点击预测器根据用户表示预测用户浏览候选新闻的概率。通过用户表示向量和候选新闻的新闻表示向量做点积得到用户点击候选新闻的概率。将候选新闻 D^c 表示为 r_c ，用户表示为 u ，那么点击预测率表示如下：

$$\hat{y} = u^T r_c$$

3.5 损失函数定义

本文采用了负采样法进行模型训练。对于每个用户浏览的新闻（正样本），随机选择同一会话中未被点击的新闻作为负样本。采用的损失函数是所有正样本的负对数似然，衡量模型预测点击概率和实际概率之间的差异。通过最小化该损失函数，模型能够更好区分用户可能感兴趣的新闻和不感兴趣的新闻。损失函数定义如下

$$L = - \sum_{i \in S} \log(p_i)$$

4 复现细节

4.1 与已有开源代码对比

复现本论文中的代码，本人参考了微软官方 github 上的 recommenders 库的代码，并在该代码的基础上修改。该库中包含了多种推荐系统的模型，包括本论文中的新闻推荐系统模型。微软官方的代码库地址为 <https://github.com/recommenders-team/recommenders>。但是，我对按照论文的方法编写的代码作了一定的修改，将从新闻的标题、正文和主题类别学习新闻表示的新闻编码器改为了仅从新闻的标题和主题类别学习新闻表示，并取得了一定的效果。

4.2 实验环境搭建

本次复现论文使用的环境配置如表 1：

表 1. 实验环境配置

操作系统	Ubuntu
平台	colab
GPU	Nvidia Tesla V100 24GB
框架	pytorch

4.3 训练数据集

本次论文复现，我采用的数据集是微软官方提供的新闻数据集。该数据集包括新闻数据集和用户点击过的新闻数据集，新闻数据集包括新闻的 ID、正文、主题类别、标题等信息，用户点击过的数据集记录了用户点击的历史记录和映像日志。但是该数据集并不能够直接作为神经网络训练的数据，需要对数据进行预处理。首先，需要将新闻数据集和用户点击过的数据集进行合并，将用户点击过的新闻和用户 ID 对应起来。然后，需要将新闻数据集中的新闻正文、标题和主题类别进行分词，将每个新闻的正文、标题和主题类别分别转化为一个词向量序列。最后，将新闻数据集和用户点击过的数据集分为训练集和测试集。新闻数据集的一个示例如表 2：

表 2. 新闻数据集示例

新闻 ID	N37378
类别	运动
子类别	高尔夫球
标题	PGA 巡回赛冠军
摘要	PGA 巡回赛新晋冠军图库。
代码	https://www.msn.com/en-us/sports/golf/pga-tour-winners/ss-AAjnQjj?ocid=chopendata

4.4 代码实现

我编写代码参考了微软官方提供的代码，并对其进行一定的修改。代码结构如图 2所示：

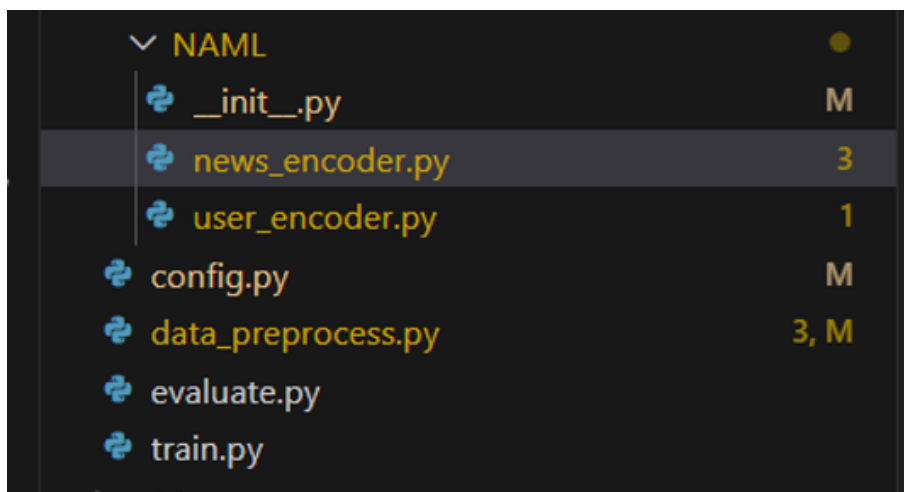


图 2. 代码结构

4.5 创新点

阅读论文并运行按照论文的方法编写的代码，从新闻标题、正文和主题类别中学习新闻表示，我发现对于大型的新闻数据集，训练的过程较慢。于是我思考，新闻的正文部分比较长占用了大量的内存开销并且很多信息是冗余的，对于较好的新闻仅仅通过新闻标题和主题类别就足以获取足够的新闻信息。因此在复现过程中，对于新闻编码器部分，我取消了对于新闻正文部分的编码，发现训练速度加快了，最终 AUC 还获得了一定程度的提升，其他的评价指标也取得了和基准模型差不多的效果。

5 实验结果分析

本次实验将以往提出的基于单一的新闻信息训练出的新闻推荐系统作为基准模型，并进行对比，得到了比较好的结果。结果如表 3所示：

表 3. 实验结果

Methods	AUC	MRR	nDCG@5	nDCG@10
LibFM	0.5880	0.3054	0.3202	0.4090
CNN	0.5909	0.3068	0.3221	0.4109
Wide&Deep	0.5846	0.3009	0.3167	0.4062
DeepFM	0.5896	0.3066	0.3221	0.4117
DFM	0.5996	0.3133	0.3288	0.4165
DKN	0.5966	0.3113	0.3286	0.4165
NAML	0.6434	0.3411	0.3670	0.4501
NAML*	0.6611	0.3115	0.3457	0.4086

其中我复现代码时进行了一定的修改，原文是对新闻的标题、正文和主题类别都进行编码以得到一个新闻表示，而我复现的代码取消了对正文的编码。最后一行是我复现的代码各种评价指标取得的结果，倒数第二行加粗的部分是原文取得的结果。可以看到，对比其他基准模型，经过修改之后，AUC 指标取得了一定的提升，其他指标取得了和其他基准模型基本差不多的效果。

5.1 评价指标

- AUC

推荐系统常用的模型评价指标，衡量推荐系统排序的准确性。AUC 的计算方式是随机给定一个正样本和负样本，分类器输出该正样本为正的的概率值比分类器输出该负样本为正的的概率值要大的概率。AUC 的取值范围是 0.5 到 1，AUC 越大，说明模型的排序效果越好。

- MRR

MRR 也是用于衡量推荐系统排序能力的评价指标，即平均倒数排名。MRR 的计算方式是对于每个用户，将推荐列表按照推荐分数从高到低进行排序，计算该用户第一个点击的新闻在排序后的列表中的位置，然后取倒数作为该用户的得分，最后对所有用户的得分求平均值。

- nDCG

全称是归一化折损累计增益。它的计算方式是将当前的折损累计增益除以最理想的 DCG，从而得到一个介于 0 到 1 之间的分数，表示当前的排序结果与最优的排序结果的相似度。

6 总结与展望

深度学习技术的不断发展为新闻推荐系统提供了新的思路，并促进了其推荐效果的提升。我这次复现的论文就是一个典型的例子，总结了前人的经验并结合当时比较火的注意力机制，提出了新的新闻推荐系统模型。该论文与传统的推荐系统模型主要的不同就是，将新闻的正文、标题和主题类别当成一篇新闻不同的视角，并使用注意力机制选取比较有代表性的新闻信息用统一的形式表示一篇新闻。对于用户表示同样地使用注意力机制提取更能彰显用户兴趣的新闻对用户进行表示。最后使用点击预测器预测用户点击候选新闻的概率。该论文的创新点在于使用了注意力机制，将新闻的不同信息当成不同的视角，综合考虑新闻的不同信息，学习出更好的新闻表示和用户表示。

然而，我对该论文的复现工作存在一定的不足，我在复现的过程中，仅仅修改了编码器的部分。没有对损失函数进行修改或者尝试用 L2 正则化替代原文采用的 dropout 正则化。因此，下一步我计划对原文中损失函数改为交叉熵损失函数，并尝试使用 L2 正则化替代 dropout 正则化，看看是否能够取得更好的效果。

参考文献

- [1] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. Personalized news recommendation based on click behavior. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 31–40, 2010.
- [2] Shumpei Okura, Yukihiro Tagami, Shingo Ono, and Akira Tajima. Embedding-based news recommendation for millions of users. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1933–1942, 2017.
- [3] Jeong-Woo Son, A-Yeong Kim, and Seong-Bae Park. A location-based news article recommendation with explicit localized semantic analysis. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 293–302, 2013.
- [4] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. Dkn: Deep knowledge-aware network for news recommendation. In *Proceedings of the 2018 world wide web conference*, pages 1835–1844, 2018.
- [5] Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. Neural news recommendation with attentive multi-view learning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 3863–3869. International Joint Conferences on Artificial Intelligence Organization, 7 2019.
- [6] Chuhan Wu, Fangzhao Wu, Mingxiao An, Yongfeng Huang, and Xing Xie. Neural news recommendation with topic-aware news representation. In *Proceedings of the 57th Annual meeting of the association for computational linguistics*, pages 1154–1159, 2019.