



计算机科学

COMPUTER SCIENCE

增强列表信息和用户兴趣的个性化新闻推荐算法

蒲妍妍, 雷航, 李贞昊, 李晓瑜

引用本文

蒲妍妍, 雷航, 李贞昊, 李晓瑜. [增强列表信息和用户兴趣的个性化新闻推荐算法](#)[J]. 计算机科学, 2022, 49(6): 142-148.

PU Qian-qian, LEI Hang, LI Zhen-hao, LI Xiao-yu. [Personalized News Recommendation Algorithm with Enhanced List Information and User Interests](#)[J]. Computer Science, 2022, 49(6): 142-148.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于评论和物品描述的深度学习推荐算法](#)

Deep Learning Recommendation Algorithm Based on Reviews and Item Descriptions

计算机科学, 2022, 49(3): 99-104. <https://doi.org/10.11896/jsjcx.210200170>

[面向推荐应用的差分隐私方案综述](#)

Survey on Privacy Protection Solutions for Recommended Applications

计算机科学, 2021, 48(9): 21-35. <https://doi.org/10.11896/jsjcx.201100083>

[基于异质信息网络表示学习与注意力神经网络的推荐算法](#)

Recommendation Algorithm Based on Heterogeneous Information Network Embedding and Attention Neural Network

计算机科学, 2021, 48(8): 72-79. <https://doi.org/10.11896/jsjcx.200800226>

[时空图注意力网络在交叉口车辆轨迹预测的应用](#)

Application of Spatial-Temporal Graph Attention Networks in Trajectory Prediction for Vehicles at Intersections

计算机科学, 2021, 48(6A): 334-341. <https://doi.org/10.11896/jsjcx.200800066>

[基于二分图卷积表示的推荐算法](#)

Recommendation Algorithm Based on Bipartite Graph Convolution Representation

计算机科学, 2021, 48(4): 78-84. <https://doi.org/10.11896/jsjcx.200400023>

增强列表信息和用户兴趣的个性化新闻推荐算法

蒲妍妍 雷航 李贞昊 李晓瑜

电子科技大学信息与软件工程学院 成都 610054

(qianqianpu@std.uestc.edu.cn)

摘要 随着数据信息的不断扩增,点对点推荐模型作为深度学习中常用的推荐算法,能一定程度上解决信息过载的问题。但其仅通过单个用户与单个新闻预测得分,未利用新闻列表之间的交互信息。为了提升个性化推荐的质量,如何充分利用用户的历史信息、新闻的文本语义信息和列表信息等,以精准全面地表征用户和新闻是当前新闻推荐系统亟待解决的问题。对此,提出了一种增强列表信息和用户兴趣的新闻个性化推荐算法。利用用户历史点击新闻序列与新闻数据训练点对点模型进行特征构造,实现用户兴趣的个性化信息抽取,通过注意力网络处理用户与整个新闻列表特征来增强列表信息,实现直接对全局列表的推荐排序。实验结果表明,此种增强列表信息和用户兴趣的个性化推荐算法能够建模全局列表信息,与前沿的新闻推荐算法相比效果有显著提升。

关键词:推荐算法;列表排序算法;注意力网络;用户个性化;新闻推荐

中图法分类号 TP183

Personalized News Recommendation Algorithm with Enhanced List Information and User Interests

PU Qian-qian, LEI Hang, LI Zhen-hao and LI Xiao-yu

School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

Abstract With the continuous expansion of data and information, the point-to-point recommendation model, as a commonly used recommendation algorithm in deep learning, can deal with the problem of overloaded information to some extent. However, it predicts the recommendation score only by a single user and an isolated news, without using of the interactive information among relevant lists of news. To improve the quality of personalized recommendation, it is urgent for current news recommendation platforms to figure out how to accurately and comprehensively represent users and news by taking full advantage of users' browsing history, semantic meaning of news as well as list information. In view of this, this paper puts forward a personalized news recommendation algorithm with improved list information and user interest. Based on the historically browsed news sequence of the user and news data, the point-to-point recommendation model is trained for representation construction to realize the tailored information extraction catering to the users' interest, and the list information is enhanced by processing the characteristics of the user and news lists through the attention network, thus realizing the direct recommendation ranking of the lists as a whole. Experimental results show that this personalized recommendation algorithm with enhanced list information and user attraction can model global the comprehensive list information, presenting a significantly improved effect compared with cutting-edge news recommendation algorithms at present.

Keywords Recommendation algorithm, Listwise algorithm, Attention network, User personalization, News recommendation

1 引言

随着信息化与移动互联网的发展,人民群众对在线新闻阅读的需求大幅增加,今日头条、腾讯新闻等平台吸引了许多用户在线阅读新闻。这些在线平台的新闻不断扩增,随之而来的是信息过载问题。为解决该问题,新闻推荐系统^[1]应运而生,结合个性化技术^[2],通过分析用户过往的新闻浏览记录

和新闻的流行程度等信息,新闻推荐系统推断用户的偏好并为用户推荐他们感兴趣的新闻,使用户可以及时有效地获取新闻信息。

当用户打开新闻应用时,新闻以列表形式曝光给用户,新闻列表的排序对用户满意度和阅读体验有很大影响。为了优化排序的效果,人们提出了大量的排序算法。在这些新闻排序的算法中,大部分通常只考虑了单个用户与单个新闻的

到稿日期:2021-04-17 返修日期:2021-08-04

基金项目:四川省科技计划项目(2018GFW0198)

This work was supported by the Sichuan Science and Technology Program(2018GFW0198).

通信作者:雷航(hlei@uestc.edu.cn)

交互,而没有考虑列表中其他新闻的影响。另外的一些方法则使用了神经网络模型对新闻列表建模,但这类模型对新闻列表的交互建模能力有限,其长依赖问题使得模型中的信息含量随着列表长度的增加而降低。

本文根据现有的一些方法与模型,提出了一种增强列表信息和用户兴趣的个性化新闻推荐算法,通过结合个性化的用户嵌入与新闻特征嵌入,用基于注意力方法的模型来处理同一候选新闻列表的所有新闻信息,实现对整个新闻列表的整体打分,并根据分数进行最终结果的排序,从而实现个性化新闻推荐任务。

2 相关工作

个性化新闻推荐指通过采集用户和新闻的信息,使用一系列算法对将要曝光给用户的新闻进行筛选和排序。目前的个性化新闻推荐方法主要分为基于协同过滤的推荐算法^[3-4]和基于内容的推荐算法^[5-7]。协同过滤是最早的推荐模型,其主要思想是从历史用户对物品的评分中找到商品和用户的关系,通过评分矩阵来预测用户对未交互物品的评分以进行推荐。但协同过滤的推荐算法存在一些冷启动、个性化能力不强以及相似用户难以预测的问题,因此基于内容的推荐算法被提出^[8]。在基于内容的推荐模型中,每个用户的推荐结果由其历史行为决定,历史行为数据加强了推荐结果的个性化,且可以通过候选商品与用户的兴趣模型直接比较相似性计算推荐结果,一定程度上缓解了冷启动问题,但也存在模型与用户兴趣变化不匹配的不足。

近年来,学者们着力探索更有效的推荐技术,尤其是在深度学习方向上的探索^[9-10]。深度学习与海量的数据相结合,能更好地提取数据中的交叉特征信息。在新闻推荐方面,研究人员提出了一系列结合新闻数据特性的推荐算法,主要侧重于使用深度学习对用户和新闻进行向量表示^[11-12]。An等^[13]提出了一种能同时学习用户长期和短期兴趣的方法来推荐新闻,使用门控循环单元从用户最近浏览的新闻中学习短期用户兴趣。Wu等^[14]提出了一种多视角学习用户兴趣的新闻推荐方法,将新闻标题、新闻类别等不同类型的新闻数据视作不同的新闻视角,以提取高维信息量的数据。Wu等^[15]还提出了一种基于内容的神经新闻推荐算法,该算法使用多头自注意力机制来捕获单词之间的相关性,以学习新闻的嵌入表示。这些研究在个性化推荐方面都取得了一定的成功,但都是预测单个用户与单个新闻的得分,这种点到点进行预测的模式被称为 Pointwise,该模式没有考虑列表上下文的交互信息对推荐结果的影响。

3 增强列表信息和用户兴趣的推荐模型

3.1 问题定义

本文对本文算法涉及的新闻推荐问题进行如下定义:现有曝光场景集合 I_r ,在某一特定新闻曝光场景 i 下,已知其待曝光新闻集合 S_i 与该曝光场景下用户历史点击新闻集合 H_i ,根据历史新闻点击集合 H_i 与新闻的文本信息对待曝光新闻集合 S_i 进行排序,使得新闻排序结果中被用户点击的新闻尽可能靠前。表 1 列出了上述问题描述中和下文中常出现

的数学符号的定义。

表 1 本文所用符号及其描述

Table 1 Symbols used in this paper and their descriptions

符号	描述
I_r	全部曝光场景集合
S_i	第 i 个曝光场景中待曝光的新闻集合
H_i	第 i 个曝光场景中用户历史点击新闻集合
x	新闻特征嵌入向量
u	用户个性化特征嵌入向量
p	位置嵌入向量
$\theta, \hat{\theta}$	点击预估模型参数矩阵、排序算法参数矩阵
E	算法模型的输出矩阵
y_{ij}	第 i 个曝光场景中,第 j 条新闻是否被点击的标签
$P(y_{ij} \cdot)$	模型预测在第 i 个曝光场景中,第 j 条新闻是否被点击的概率

3.2 特征构造

本文的最终目标是对用户的一个新闻曝光场景中的新闻进行排序,对于推荐算法而言,初始的任务即为构造用户特征 u 与新闻的特征 x 。传统的方法主要使用用户与新闻的基本特征,例如用户性别、年龄以及新闻的类别等,本文则使用新闻的文本数据与用户历史点击数据通过深度神经网络模型构建输入特征。构造特征的网络模型使用 Pointwise 模式的点击预估模型,即输入用户与新闻的嵌入向量,计算点击概率,优化点击目标。训练完成后,训练好的用户与新闻嵌入向量即为实际送入算法模型的用户与新闻特征。Pointwise 模式的点击预估模型如图 1 所示。

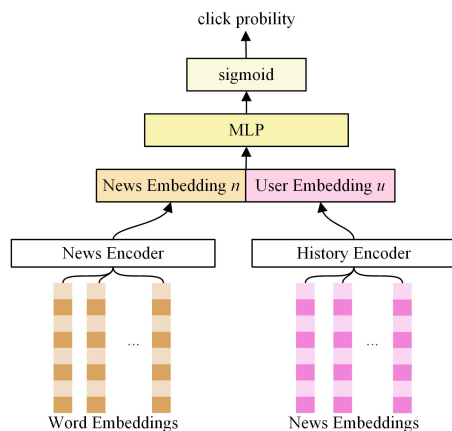


图 1 点击预估模型

Fig. 1 CTR model

3.2.1 新闻编码模块

新闻特征 x 的构造使用新闻的标题与正文的文本数据。首先对文本进行分词,对每个词语使用一个稠密词嵌入向量进行表示,将新闻转化为向量序列数据,接着对向量序列进行最大池化聚合,如式(1)所示:

$$x = \text{MaxPooling}(w_1, w_2, \dots, w_n) \quad (1)$$

其中, w_i 为词向量, n 为文本词语个数, x 为聚合形成的新闻特征向量,聚合方法可以根据点击预估模型来决定,常见的 RNN, Attention 等都是可选择的方法。

3.2.2 用户编码模块

对于用户特征,主要使用用户的历史点击新闻来构造。依靠用户的一些基本属性(如性别、年龄等)无法构造可靠的

用户特征,而历史点击数据更能反映用户的兴趣爱好与近期的关注点,因此本文选择历史点击新闻来构造用户特征。对于给定用户 u_i , 在一个曝光场景下有 m 个历史点击新闻, 对于每条新闻, 使用 3.2.1 节提出的新闻编码模块来将其编码为新闻向量, 获取历史点击新闻向量序列, 然后使用最大化聚合构造用户特征编码, 如式(2)所示:

$$\mathbf{u} = \text{MaxPooling}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) \quad (2)$$

其中, \mathbf{x} 为新闻特征向量, m 为历史新闻向量的个数, \mathbf{u}_i 为生成的用户特征向量, 聚合方式也可以根据点击预估模型进行修改。

3.2.3 点击预估模型

点击预估模型是一个用于预估用户是否点击新闻的二分类模型。此处以一个简单的点击预估模型为例进行算法说明。

在用户 i 的新闻曝光场景 I_r 中, 使用用户特征编码器将用户的历史点击新闻序列转为用户特征向量, 使用新闻特征编码器将待曝光的每一条新闻转为新闻特征向量。完成用户与新闻的特征构造后, 对于单个用户与新闻, 拼接用户特征 \mathbf{u}_i 和新闻特征 \mathbf{x}_{ij} , 将结果送入多层感知机模型以获取输出值, 输出值使用 Sigmoid 进行缩放, 该值即近似表示用户点击该新闻的概率, 计算过程如式(3)所示:

$$P(y_{ij} | \mathbf{x}_{ij}, \mathbf{u}_i; \boldsymbol{\theta}) = \text{Sigmoid}(\text{MLP}([\mathbf{x}_{ij}; \mathbf{u}_i])) \quad (3)$$

模型的优化使用交叉熵函数进行计算, 如式(4)所示:

$$L = - \sum_{i \in I_r} \sum_{j \in S_i} y_{ij} \log(P(y_{ij} | \mathbf{x}_{ij}, \mathbf{u}_i; \boldsymbol{\theta})) \quad (4)$$

实际上, 该模型可以被替换为任意常见的新闻点击预估模型, 如最前沿的 LSTUR, NRMS 和 NAML 模型, 在本文后续实验环节中也会对其进行对比。

3.3 个性化新闻算法模型

本文提出了一种增强列表的个性化新闻推荐算法, 后文简记为 PNR 模型 (Personalized News Recommendation), 模型的整体结构如图 2 所示。

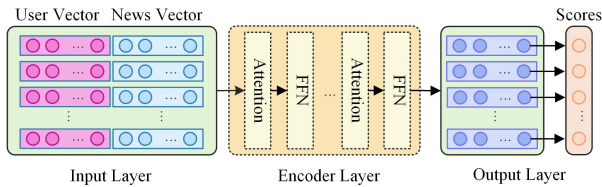


图 2 本文模型的整体结构

Fig. 2 Main structure of proposed model

3.3.1 输入层

输入层用于处理输入的特征, 将其转化为模型需要的形式。本模型的输入特征为用户特征以及曝光场景中的各个新闻特征。首先需要构造输入特征矩阵, 将用户特征向量复制 m 次, m 为列表中的新闻条数, 与新闻特征进行拼接, 如式(5)所示:

$$\mathbf{E}'' = [\mathbf{X}; \mathbf{U}] = \begin{bmatrix} \mathbf{x}_{i1}; \mathbf{u}_i \\ \mathbf{x}_{i2}; \mathbf{u}_i \\ \dots \\ \mathbf{x}_{im}; \mathbf{u}_i \end{bmatrix} \quad (5)$$

对于已有的训练好的用户与新闻特征, 其聚合了用户对

新闻的兴趣信息, 但是缺乏列表信息, 这对于本文的目标来说尚有欠缺, 因此需要补充新闻列表的列表信息。本文使用位置向量编码来补充这个信息。位置编码的计算式如式(6)所示:

$$P_{(pos, 2i)} = \sin(pos/1000^{2i/d}) \quad (6)$$

$$P_{(pos, 2i+1)} = \cos(pos/1000^{2i/d})$$

其中, pos 是新闻在初始列表中的排名, i 为位置嵌入向量的第 i 个位置, d 为位置嵌入向量的长度。

正如式(6)所示, 位置编码是表现新闻表项在列表中的顺序信息, 而这一顺序是来自于特征构造阶段的点击预估模型。点击预估模型通过计算用户与每条新闻的点击概率值由高到低对新闻进行顺序指定。最终构成的输入矩阵如式(7)所示:

$$\mathbf{E}' = \begin{bmatrix} \mathbf{x}_{i1}; \mathbf{u}_i \\ \mathbf{x}_{i2}; \mathbf{u}_i \\ \dots \\ \mathbf{x}_{im}; \mathbf{u}_i \end{bmatrix} + \begin{bmatrix} \mathbf{p}_{i1} \\ \mathbf{p}_{i2} \\ \dots \\ \mathbf{p}_{im} \end{bmatrix} \quad (7)$$

完成式(5)所示输入特征矩阵与位置嵌入的加法后, 使用线性变换的方法将特征嵌入矩阵 $\mathbf{E}' \in R^{m \times (d_x + d_u)}$ 转换为 $\mathbf{E} \in R^{m \times d}$, 其中 d 是编码层每个输入向量的维数。 \mathbf{E} 的表达式如式(8)所示:

$$\mathbf{E} = \mathbf{E}' \mathbf{W}_E + \mathbf{b}_E \quad (8)$$

其中, $\mathbf{W}_E \in R^{(d_x + d_u) \times d}$ 是线性变换矩阵, \mathbf{b}_E 是 d 维偏移向量。

3.3.2 编码层

编码层用于计算列表中不同新闻项之间的交互信息, 主要关注不同的新闻项之间的交互影响。在曝光场景下, 不同新闻项之间存在相互影响, 从而影响基于用户新闻点击预估模型的判断, 因此需要整体考虑针对列表的排序方法。本文模型使用多层多头自注意力^[16]方法来完成这一工作。

多头注意力层使用两个子层进行计算, 第一个子层为注意力子层, 计算式如式(9)所示:

$$\text{Attention}(\mathbf{E}\mathbf{W}_Q, \mathbf{E}\mathbf{W}_K, \mathbf{E}\mathbf{W}_V) = \text{Softmax}\left(\frac{\mathbf{E}\mathbf{W}_Q(\mathbf{E}\mathbf{W}_K)^T}{\sqrt{d}}\right)\mathbf{E}\mathbf{W}_V \quad (9)$$

其中, $\mathbf{E}\mathbf{W}_Q, \mathbf{E}\mathbf{W}_K, \mathbf{E}\mathbf{W}_V$ 分别表示查询、键和值; d 是 $\mathbf{E}\mathbf{W}_K$ 的维数, 除以 \sqrt{d} 可以避免较大的内积值; Softmax 用于将内积的值转换为对值向量 $\mathbf{E}\mathbf{W}_V$ 的相加权重, 用于表示两者的相关性。

为了建模更复杂的列表相互影响信息, 本文还使用了多头注意力机制, 将原始维度降低后执行多个注意力函数, 如式(10)所示:

$$\text{head}_i = \text{Attention}(\mathbf{E}\mathbf{W}_Q, \mathbf{E}\mathbf{W}_K, \mathbf{E}\mathbf{W}_V) \quad (10)$$

$$\mathbf{O} = \text{Concat}(\text{head}_1, \dots, \text{head}_h)\mathbf{W}_O$$

其中, $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in R^{d \times d}$, $\mathbf{W}_O \in R^{hd \times d_{\text{model}}}$ 是线性变换矩阵, h 是头的数量。

逐位置前馈网络 (Position-wise Feed-Forward Networks) 通过非线性的全连接层来处理输入向量不同维度之间的相互作用, 进而增强模型。该网络与注意力层之间使用 Dropout 和层归一化 (Layer Norm)^[17] 方法, 并进行残差连接。Dropout 对神经元进行少量随机丢弃, 可以提高模型的泛化性能, 层归一化能使每一层神经网络参数值呈现合理分布。层归一化方法如式(11)所示:

$$y = \frac{x - E[x]}{\sqrt{Var[x] + \epsilon}} * \gamma + \beta \quad (11)$$

其中, $E[x]$ 表示输入的期望, $Var[x]$ 表示输入的方差, γ, β 是两个可学习参数。

残差连接则对子层的输出与输入求和, 目的是避免子层计算丢掉了原始输入信息, 连接方式如式(12)所示:

$$y = SubLayer(x) + x \quad (12)$$

3.3.3 输出层

输出层的功能主要是为当前曝光场景 i 中的每条新闻 $j = j_1, j_2, \dots, j_m$ 生成一个分数 $Score(ij)$ 。输出层使用一层线性变换层并紧接一个 Softmax 函数, 用于将数值归一化为概率值。Softmax 函数的输出被视为当前用户可能点击该新闻的概率, 标记为 $P(y_{ij} | \mathbf{x}_{ij}, \mathbf{u}_i; \hat{\theta})$ 。本文使用 $P(y_{ij} | \mathbf{x}_{ij}, \mathbf{u}_i; \hat{\theta})$ 作为 $Score(ij)$ 对新闻进行重新排列。 $Score(ij)$ 的计算式为:

$$Score(ij) = P(y_{ij} | \mathbf{x}_{ij}, \mathbf{u}_i; \hat{\theta}) = \text{Softmax}(F^{(N \times)} \mathbf{W}_F + \mathbf{b}_F), i \in I_r, j \in S_i \quad (13)$$

其中, $F^{(N \times)}$ 是编码层的输出, \mathbf{W}_F 是可学习的线性变换矩阵, \mathbf{b}_F 是偏移向量。在训练过程中, 使用点击数据作为标签, 并最小化公式中的损失函数。

$$L = - \sum_{i \in I_r} \sum_{j \in S_i} y_{ij} \log(P(y_{ij} | \mathbf{x}_{ij}, \mathbf{u}_i; \hat{\theta})) \quad (14)$$

3.4 个性化新闻推荐算法的总流程

本文提出了一种增强列表信息和用户兴趣的个性化新闻推荐算法, 整体算法的流程如算法 1 所示。算法 1 首先基于点击预估模型获取用户向量、新闻向量以及预估点击概率, 根据预估点击概率对新闻曝光列表进行排序并建模, 以获取最终列表得分并将其输出。

算法 1 个性化新闻推荐算法

输入: 第 i 个曝光场景中待曝光的新闻集合 S_i , 用户历史点击集合 H_i

输出: 曝光场景 i 下候选列表新闻的得分 S

1. $his \leftarrow [1.. H_i.length]$
2. for $j \leftarrow 1$ to $H_i.length$ do
3. $his[j] \leftarrow \text{getNewsEncode}(H_i[j])$ // 获取历史新闻特征
4. end for
5. $\mathbf{u} \leftarrow \text{getUserEncode}(his)$ // 获取用户特征向量
6. $\mathbf{U} \leftarrow \text{repeat } \mathbf{u} \text{ } S_i.length \text{ times}$ // 复制 \mathbf{u} 向量多次
7. $\mathbf{X} \leftarrow [1.. S_i.length]$
8. $scores \leftarrow [1.. S_i.length]$
9. for $j \leftarrow 1$ to $S_i.length$ do
10. $\mathbf{X}[j] \leftarrow \text{getNewsEncode}(S_i[j])$ // 获取候选新闻特征向量
11. $scores[j] \leftarrow \text{getCTR}(\mathbf{u}, \mathbf{X}[j])$ // 获取点击概率
12. end for
13. $\mathbf{X} \leftarrow \text{sort } \mathbf{X} \text{ by } scores$ // 根据点击概率排序
14. $\mathbf{E} \leftarrow [\mathbf{X}; \mathbf{U}] + p$ // 添加位置编码
15. $\mathbf{O} \leftarrow \text{getAttentionEncode}(\mathbf{E})$ // 计算注意力结果
16. $\mathbf{S} \leftarrow \text{getOutput}(\mathbf{O})$ // 获取最终列表排序分数
17. return S

4 实验与分析

4.1 数据集

实验数据集选用微软 MIND 新闻数据集^[18], 该数据集总共

采集了从 2019 年 10 月 12 日至 11 月 22 日 6 周内用户曝光场景记录, 数据集中包含用户历史点击记录信息、用户曝光场景新闻信息以及新闻的文本信息。本实验从中随机挑选了 10 万名不同用户的曝光场景记录, 单个曝光场景包含用户、场景中的新闻 ID 以及用户的点击情况。为了模拟真实的新闻推荐过程, 实验针对训练集和测试集各自抽取了 5 万名用户, 并按照 4:1 的比例抽取曝光记录作为训练样本。本文所用新闻数据集中可用的数据统计如表 2 所列。

表 2 数据集统计

Table 2 Statistics of dataset

Dataset	User	News	Impressions
Train	50000	51283	502113
Test	50000	42417	125529

4.2 实验基准模型与评价指标

4.2.1 对比基准模型

为了验证本文提出的新闻推荐算法对推荐效果的改善, 本实验比较了现有主流的新闻推荐模型的结果, 主要比较的 Pointwise 基线模型包括:

(1) LSTUR^[13]。LSTUR 是一种具有长期和短期用户兴趣的神经新闻推荐方法, 它利用 GRU 从用户最近点击的新闻中塑造短期用户兴趣, 从整个点击历史中塑造长期用户兴趣。

(2) NAML^[14]。NAML 是一种专注多视角学习的神经网络新闻推荐方法, 将不同种类的新闻信息合并到新闻文章的表示中。

(3) NRMS^[15]。NRMS 是一种基于神经网络的新闻推荐方法, 利用多头自注意力从新闻文本中的单词中学习新闻表示, 从已点击的新闻文章中学习用户表示。

由于本文模型是在 Pointwise 模型之后进行列表排序, 因此本文也对比了同样进行列表排序的模型, 具体包括:

(1) GlobalRerank(GR)^[19]。GlobalRerank 是一种使用带有注意机制的 RNN 模型作为解编码器的列表排序算法。

(2) DLCM^[20]。DLCM 是一种使用 GRU 将本地上下文信息编码到全局向量中, 采用类似文本生成模型方式进行排序的排序算法。

4.2.2 度量标准

为了对比各模型的推荐效果, 本文选取了被广泛使用的两个评价指标: 平均倒数排名 (Mean Reciprocal Rank, MRR) 和归一化折损累计增益 (Normalized Discounted Cumulative Gain, NDCG)。本文的任务涉及对新闻列表进行排序, 因此主要的衡量指标应能够对列表进行判断。本文使用了在检索系统中常见的 MRR 和 NDCG 指标作为评估指标。

MRR 的计算式如式(15)所示。MRR 使用检索排名的倒数进行计算, 如果被点击的结果排名越高, 则分数越高, 可以直观地反映检索系统的性能。

$$MRR = \frac{1}{Q} \sum_{i=1}^Q \frac{1}{rank_i} \quad (15)$$

NDCG 的计算式如式(16)、式(17)所示。NDCG 主要采用了相关系数的计算得分, 相关系数越高且排名越高, 最终

得分才会越高,并对结果进行缩放,可以横向评价不同检索系统的性能。

$$DCG@K = \sum_{k=1}^K \frac{2^{rel_k} - 1}{\log_2(k+1)} \quad (16)$$

$$IDCG@K = \sum_{k=1}^{|REL|} \frac{2^{rel_k} - 1}{\log_2(k+1)} \quad (17)$$

$$NDCG@K = \frac{DCG@K}{IDCG@K} \quad (18)$$

4.3 模型效果对比实验

本实验对基线模型、使用不同排序算法结合的模型以及本文提出的新闻推荐算法模型进行了对比。本次实验采用10次实验取均值的方式进行,实验结果如表3所列。

表3 不同推荐算法的效果对比

Table 3 Result comparison of different recommendation algorithms

Models	MRR@5	MRR@10	NDCG@5	NDCG@10
LSTUR	0.3568	0.3665	0.3863	0.4447
NAML	0.4280	0.4324	0.4597	0.5144
NRMS	0.4328	0.4366	0.4653	0.5199
DLCM(LSTUR)	0.4030	0.4085	0.4306	0.4755
DLCM(NAML)	0.4268	0.4307	0.4576	0.5125
DLCM(NRMS)	0.4326	0.4358	0.4641	0.5189
GR(LSTUR)	0.4059	0.4116	0.4316	0.4778
GR(NAML)	0.4259	0.4302	0.4556	0.5121
GR(NRMS)	0.4139	0.4188	0.4444	0.5062
PNR(LSTUR)	0.4362	0.4406	0.4643	0.5158
PNR(NAML)	0.4515	0.4548	0.4802	0.5327
PNR(NRMS)	0.4570	0.4597	0.4863	0.5384

LSTUR模型结合不同排序算法的对比结果如图3所示。从实验结果来看,本文算法相比LSTUR基线模型有20.2%的MRR@10得分提升以及16%的NDCG@10得分提升。相比其他排序模型,如图3所示,PNR模型在训练过程中大幅领先其他排序模型,最终其NDCG@10相比DLCM模型有8.5%的提升,相比GlobalRerank有8%的提升。

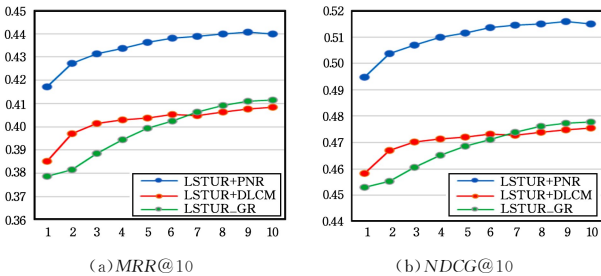
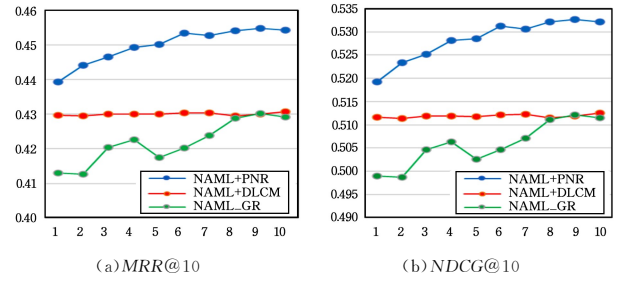


图3 LSTUR模型结合不同排序算法的对比

Fig. 3 Comparison of LSTUR model combined with different ranking algorithms

NAML模型结合不同排序算法的对比结果如图4所示。NAML模型是一种较复杂的新闻点击预估模型,其使用了注意力机制,因此结构更加多变,效果也很显著。PNR模型相比基线有3.6%的提升,相比其他排序模型,DLCM在MRR与NDCG指标上几乎都没有变化,GlobalRerank模型虽然表现出上升趋势,但是最终得分也与基线十分接近,说明对于这种复杂的点击预估模型,这些排序算法均不能很好地完成任务,而PNR模型则能够再次排序模型,使得效果更为出色。



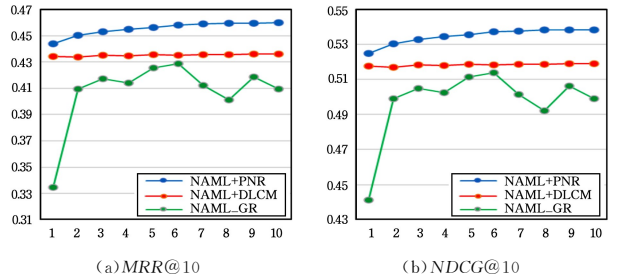
(a)MRR@10

(b)NDCG@10

图4 NAML模型结合不同排序算法的对比

Fig. 4 Comparison of NAML model combined with different ranking algorithms

NRMS模型的情况与NAML十分相似,其结合不同排序算法的对比结果如图5所示。DLCM和GlobalRerank模型均无法有效地提升模型表现,而PNR模型比NRMS模型有3.6%的提升,效果显著。



(a)MRR@10

(b)NDCG@10

图5 NRMS模型结合不同排序算法的对比

Fig. 5 Comparison of NRMS model combined with different ranking algorithms

这些实验的对比结果有效地论证了本文提出的PNR模型在新闻推荐问题上的有效性,PNR模型在其效果优于基线模型的同时还能正确学习复杂的点击预估模型生成的用户与新闻向量。

4.4 消融实验

4.4.1 自注意力头数、层数对推荐效果的影响

本节将探究本文模型中不同的自注意力头数目 H 以及层数 L 对模型推荐效果的影响。实验默认使用 $H=4$ 与 $L=3$ 在NRMS模型上进行,除了表格所示参数值外均使用默认参数。实验结果如表4所列,随着数目 H 的增大,MRR指标和NDCG指标的效果提升,但是 L 和 H 过大后效果会降低。分析式(10)可知,不同的自注意力头可以为将原始特征向量映射到不同的子空间中。如果使用更多的注意力头,可以在更多的向量空间中建模,使得编码层能够捕获更多的语义信息。

表4 自注意力头数和层数对MRR和NDCG的影响

Table 4 Effects of number of self-attention heads and layers on MRR and NDCG

Parameters	MRR@5	MRR@10	NDCG@5	NDCG@10
$H=4, L=3$	0.4570	0.4597	0.4863	0.5384
$H=1, L=3$	0.4504	0.4534	0.4794	0.5318
$H=2, L=3$	0.4525	0.4556	0.4815	0.5335
$H=8, L=3$	0.4538	0.4569	0.4829	0.5346
$H=2, L=1$	0.4485	0.4518	0.4779	0.5305
$H=2, L=2$	0.4531	0.4561	0.4812	0.5341
$H=2, L=4$	0.4526	0.4556	0.4812	0.5333

同时,为了使用多头注意力而不引入更多的参数,在使用时会对原始输入维度进行降维,使得总的参数量是一致的。当注意力头数变多时,则自身向量维度减少,虽然能从更多的向量空间中获取信息,但是由于自身维度减少导致自身所携带的信息减少,因此效果会产生回落。这一实验说明,可以通过适当增加头部来提升效果,但是需要小心引入较多的注意力头。

4.4.2 新闻列表长度 K 对推荐效果的影响

本节将对本文算法中不同的新闻列表长度 K 对评价指标 MRR 值和 $NDCCG$ 值的影响进行实验。实验结果如表 5 所列。

表 5 列表长度 K 对 MRR 和 $NDCG$ 的影响

Table 5 Effect of list length K on MRR and $NDCG$

K	$MRR@5$	$MRR@10$	$NDCG@5$	$NDCG@10$
5	0.4429	0.4466	0.4722	0.5252
10	0.4475	0.4510	0.4764	0.5285
15	0.4498	0.4530	0.4785	0.5309
20	0.4512	0.4545	0.4798	0.5323
25	0.4570	0.4597	0.4863	0.5384

从表 5 中可以看到,随着 K 值的增大,指标得分均有所提升。从结果分析可知,在新闻列表中,未考虑列表时一些用户可能点击兴趣大的新闻被排到了列表后面,但通过构建整个新闻列表的全局关系,对新闻特征和用户特征进行更丰富的建模之后,成功地将优质新闻排到了前面,或者虽然用户对某些新闻可能不是特别感兴趣,但当它在列表中与某些类别的新闻组合出现时,用户点击的可能性会加大。这反映了本文提出的个性化新闻推荐算法能满足用户在阅读新闻时的需求。

4.5 列表建模分析

本节将针对本文所实现的模型能否对列表建模进行分析。首先针对理论部分进行分析,在编码层中,其输入 $E = [e_1, e_2, \dots, e_n]^T$ 的每一个行向量 e_i 表示用户 u 和新闻列表第 i 个新闻特征的混合输入向量,其中 n 表示列表长度。在式(9)中, EW_Q, EW_K 和 EW_V 分别表示将输入 E 映射到 3 种不同向量空间后获得的向量表示, $\text{Softmax}((EW_Q(EW_K)^T)/\sqrt{d})$ 可以计算任意两个新闻项的点积,进而获得一个 $n \times n$ 的注意力得分矩阵,用于表示新闻与新闻之间的关联度,最终基于该注意力得分矩阵对 EW_V 进行加权求和来获取单个新闻表项的输出向量,最终第 i 个新闻表项的输出可以表示为:

$$\tilde{e}_i \times \alpha_i^1 + \tilde{e}_i \times \alpha_i^2 + \dots + \tilde{e}_i \times \alpha_i^n \quad (19)$$

其中, \tilde{e}_i 为 EW_V 中的第 i 个新闻表项的表示向量, α_i^j 为第 i 个新闻表项与第 j 个新闻表项之间的注意力分数。从计算结果来看,单个表项的输出结合了整个新闻列表的信息,可以认为对全局列表进行了建模。

对实验样本进行分析,在数据集中随机抽取一个曝光场景,并限定该场景中存在 10 条新闻,例如用户对新闻的点击情况为 $[0, 1, 0, 0, 0, 0, 0, 0, 0, 0]$,表示用户点击了列表之中的第二条新闻。图 6 给出了该场景在模型中的注意力权重,图 6 中第 i 行第 j 列表示当前第 i 条新闻受到了第 j 条新闻的影响,在本文中称之为注意力权重。从本质上讲,注意力值用于描述第 i 条新闻的分布式特征表示受到第 j 条新闻的分布式

特征表示影响的程度,颜色越深表示注意力值越大,影响也越大。

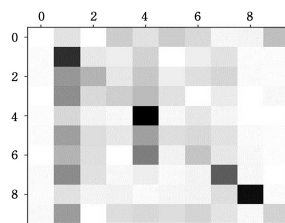


图 6 注意力可视化

Fig. 6 Attention visualization

从图 6 可以直观地看出,每条新闻除了自身的注意力的权重较大之外,还受到了其他新闻的影响,其中第二列整体的颜色都偏深,这与最终结果用户点击了第二条新闻相符,表明注意力机制可以学习到列表中不同新闻之间的相互影响与用户实际点击的关系。

4.6 实验小结

首先,对本文提出的新闻个性化推荐算法进行了对比实验,对比了本文模型与多个不同的基线、排序模型,展现了本文模型的有效性。其次,对算法的参数进行了消融实验,一方面证明了多头多层注意力可以有效提升效果,但头数 H 、层数 L 增加到一定数量后效果回落且带来了更高的计算消耗,另一方面对新闻列表长度 K 的实验也说明,在数据量允许的情况下提升新闻列表长度可以改善实验效果。最后进行了列表建模分析,再次说明了新闻列表上下文可以帮助对新闻列表整体的排序。

结束语 本文提出了一种增强列表信息和用户兴趣的用户个性化推荐模型,并且通过实验证明了这种基于整个新闻列表进行推荐排序的算法能够有效地提升新闻曝光场景中的表现效果和个性化展现。在未来的探索工作中,希望能够细化新闻的文本内容信息,引入文本模型,以研究其对推荐结果排名序列的影响。

参考文献

- [1] LI L, WANG D D, ZHU S Z, et al. Personalized news recommendation: a review and an experimental investigation[J]. Journal of Computer Science and Technology, 2011, 26(5): 754-766.
- [2] LIU J L, LI X G. Techniques for Recommendation System: A Survey[J]. Computer Science, 2020, 47(7): 47-55.
- [3] DAS A S, DATAR M, GARG A, et al. Google News Personalization: Scalable Online Collaborative Filtering[C]// Proceedings of the 16th International Conference on World Wide Web (WWW 2007). Banff, Alberta, Canada: ACM Press, 2007: 271-280.
- [4] KOREN Y. Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model[C]// Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008). New York, USA: Association for Computing Machinery, 2008: 426-434.
- [5] PHELAN O, MCCARTHY K, BENNETT M, et al. Terms of a Feather: Content-Based News Recommendation and Discovery

- Using Twitter [C] // Proceedings of the 33rd European Conference on Advances in Information Retrieval (ECIR 2011). Berlin; Springer, 2011; 448-459.
- [6] LI H, CAI F, LIAO Z. Content-based Filtering Recommendation Algorithm Using HMM [C] // 2012 Fourth International Conference on Computational and Information Sciences, Chongqing, China; IEEE, 2012; 275-277.
- [7] BANSAL T, DAS M, BHATTACHARYYA C. Content Driven User Profiling for Comment-Worthy Recommendations of News and Blog Articles [C] // Proceedings of the 9th ACM Conference on Recommender Systems (RecSys 2015). New York; Association for Computing Machinery, 2015; 195-202.
- [8] WANG S Q, LI X X, SUN F Z, et al. Survey of Research on Personalized News Recommendation Techniques [J]. Journal of Frontiers of Computer Science and Technology, 2020, 14(1): 18-29.
- [9] SHEN J, QIAO S J, HAN N, et al. Personalized Recommendation Model with Multiple Information Fusion [J]. Journal of Chongqing University of Technology (Natural Science), 2021, 35(3): 128-138.
- [10] HAO R F, ZHANG G M, CHENG Y Q. Socialized Matrix Factorization Recommendation Algorithm with User Rating Preference Confidence [J]. Journal of Chongqing University of Technology (Natural Science), 2020, 34(11): 138-146.
- [11] TIAN X, DING Q, LIAO Z H, et al. Survey on Deep Learning Based News Recommendation Algorithm [J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(6): 971-998.
- [12] TAO T Y, WANG Q Q, FU Y W, et al. Personalized Recommendation Algorithm for Financial News Based on Knowledge Graph [J]. Computer Engineering, 2021, 47(6): 98-103, 114.
- [13] AN M, WU F, WU C, et al. Neural News Recommendation with Long-and Short-Term User Representations [C] // The 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019). Florence, Italy; Association for Computational Linguistics, 2019; 336-345.
- [14] WU C, WU F, AN M, et al. Neural News Recommendation with Attentive Multi-View Learning [C] // Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI 2019). Macao, China; International Joint Conferences on Artificial Intelligence Organization, 2019; 3863-3869.
- [15] 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). Hong Kong, China; Association for Computational Linguistics, 2019; 6388-6393.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017). Red Hook, NY, USA; Curran Associates Inc., 2017; 6000-6010.
- [17] BA J L, KIROUS J R, HINTON G E. Layer Normalization [J]. arXiv:1607.06450, 2016.
- [18] WU F, QIAO Y, CHEN J H, et al. MIND: A Large-Scale Dataset for News Recommendation [C] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020). Association for Computational Linguistics, 2020; 3597-3606.
- [19] ZHUANG T, OU W, WANG Z. Globally Optimized Mutual Influence Aware Ranking in E-Commerce Search [C] // Proceedings of the 27th International Joint Conference on Artificial Intelligence. AAAI Press, 2018; 3725-3731.
- [20] AI Q, BI K, GUO J, et al. Learning a Deep Listwise Context Model for Ranking Refinement [C] // The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. New York, USA; Association for Computing Machinery, 2018; 135-144.



PU Qian-qian, born in 1998, postgraduate. Her main research interests include machine learning and recommendation systems.



LEI Hang, born in 1960, Ph.D, professor, Ph.D supervisor. His main research interests include embedded system and big data analysis.

(责任编辑:喻黎)