

[文章编号] 1003-4684(2023)04-0082-06

基于知识图谱的新闻推荐神经网络模型

程华松,熊才权,柯远志,吴歆韵

(湖北工业大学计算机学院,湖北 武汉 430068)

[摘 要] 新闻推荐的目标是根据用户的历史阅读习惯,为用户推送其感兴趣的新闻内容。现有的方法存在特征学习不足的问题,没有考虑到用户与浏览新闻之间的关系,以及不同候选新闻对用户的重要性。针对以上问题提出一种基于知识图谱的新闻推荐神经网络模型。首先使用 GloVe 模型(Global Vectors for Word Representation)和基于翻译的模型(TransX)分别从新闻语义和知识层面得到文本语义特征、实体特征和实体在知识图谱中的上下文特征。然后,使用 LSTM-CNN 网络挖掘新闻和用户深层次信息,得到用户的新闻偏好向量,同时引入注意力机制减少新闻无关题材的影响;最后通过点击求和计算用户对候选新闻的偏好值并生成最终的推荐结果。在真实新闻数据集 MIND 上的实验结果表明,本文所提出的模型相对现有的 DKN、DeepFM、DeepWide 模型,在 AUC、MRR、NDCG@k 指标上表现的更优异。

[关键词] 新闻推荐;神经网络;知识图谱;注意力机制

[中图分类号] TP18 **[文献标识码]** A

为了减轻信息过载的影响,向用户进行个性化新闻推荐成为新闻平台中的一个十分重要的功能^[1]。

由于新闻具有时效性特点,协同过滤方法难以做到实时更新,无法动态的捕获用户兴趣变化,在这种推荐方法基础上的推荐系统会面临严重的冷启动问题^[2]。基于内容的推荐则依赖于传统自然语言处理模型,这类模型在提取语义特征时往往忽视了词语顺序,不能对新闻文本语义进行准确刻画。Cheng 等^[3]提出 Wide&Deep 模型,通过联合训练线性模型和深度神经网络来解决稀疏输入的问题,但是线性模型的输入依赖于人工特征提取。Guo 等^[4]提出一种基于深度学习的模型(DeepFM),它结合了前馈神经网络和线性模型进行推荐,还新闻标题和个人资料嵌入作为特征,不需要人工特征工程。随着深度学习的不断发展,提取高质量的新闻特征和用户表示已经成为了新闻推荐系统中研究热点^[5]。人们在阅读新闻时通常会对多个特定的新闻类别感兴趣,如何根据用户对当前候选新闻的多样化阅读历史动态衡量用户的兴趣是新闻推荐的关键。Okura 等^[5]提出一种基于去噪自编码器和递归神经网络的个性化新闻推荐方法。新闻语言是高度浓缩的,由大量的知识实体和常识组成,传统的语义

模型^[6]和主题模型不能发现潜在的知识层面的联系,Wang 等^[7]提出将知识图谱引入到推荐模型中,来弥补仅使用词向量特征不足的缺陷。

针对上述问题,本文从新闻语义层面和知识层面得到两种不同的向量表示,通过多通道的卷积神经网络融合形成更精准的新闻特征表示,同时引入注意力机制使提取的新闻特征向量更能够代表上下文的语义信息。最后计算用户对候选新闻的偏好值并生成最终的推荐结果。通过在 MIND 新闻数据集对模型进行评估,获得了对于基线模型而言较好的推荐效果。

1 相关理论介绍

1.1 知识表示学习

2013 年,Mikolov 等^[8]在词向量空间中发现一种存在于词汇的语义关系和句子关系中的特殊平移规律,通过该现象提出了 TransE 模型^[9],并将该模型用于链接预测等任务。实验结果表明,TransE 比以往模型的性能有明显提升,经过改进,研究者依次提出了 TransR^[10],TransD^[11]等模型。

在当前的主流知识库中,知识实体以 (h, r, t) 三元组的形式存储,对于每个三元组 (h, r, t) ,其中 h 表示头实体, r 表示关系, t 表示尾实体,TransE 希

[收稿日期] 2022-04-14

[基金项目] 湖北省支持企业技术创新发展项目(2021BLB171);国家自然科学基金(61902116)

[第一作者] 程华松(1997-),男,湖北黄冈人,湖北工业大学硕士研究生,研究方向为人工智能。

[通信作者] 熊才权(1966-),男,湖北鄂州人,工学博士,湖北工业大学教授,研究方向为人工智能,辩论模型,智能决策。

望可以满足 $\boldsymbol{l}_h + \boldsymbol{l}_r \approx \boldsymbol{l}_t$ 。TransE 模型的损失函数定义如下：

$$f_r(h, t) = \| \boldsymbol{l}_h + \boldsymbol{l}_r - \boldsymbol{l}_t \|_{\frac{L_1}{L_2}} \tag{1}$$

TransE 用关系 r 的向量 \boldsymbol{l}_r 作为头实体向量 \boldsymbol{l}_h 和尾实体向量 \boldsymbol{l}_t 之间的平移,研究者将 TransE 称做翻译模型。由于 TransE 模型设计简单,计算的时间效率低,在高质量的知识图谱上提取知识库的复杂关系时有一定的局限性。为了解决这些问题,Ji 等^[11]提出了 TransD 模型,使用同一个投影矩阵可以在某种程度上限制模型的表达能力,来降低头实体和尾实体表示的差异。给定一个三元组 (h, r, t) ,TransD 模型设置了 2 个分别将头实体和尾实体投影到关系空间的投影矩阵 M_{rh} 和 M_{rt} ,定义如下：

$$M_{rh} = \boldsymbol{l}_{rp} \boldsymbol{l}_{hp} + \boldsymbol{I}^{d \times k} \tag{2}$$

$$M_{rt} = \boldsymbol{l}_{rp} \boldsymbol{l}_{tp} + \boldsymbol{I}^{d \times k} \tag{3}$$

其中 $\boldsymbol{l}_{hp}, \boldsymbol{l}_{tp} \in R^d, \boldsymbol{l}_{rp} \in R^d$,下标 p 代表该向量在向量空间中的投影向量,并且 M_{rh} 和 M_{rt} 和实体之间的关系都有联系,通过使用 2 个投影向量来构建投影矩阵,解决了 TransR 模型参数过多的问题,其损失函数定义如下：

$$f_r(h, t) = \| \boldsymbol{l}_h M_{rh} + \boldsymbol{l}_r - \boldsymbol{l}_t M_{rt} \|_{\frac{L_1}{L_2}} \tag{4}$$

1.2 LSTM 神经网络

RNN 是一种广泛用于序列任务的深度学习网络。RNN 不能很好地解决长时依赖问题,并且会产生梯度消失和梯度爆炸问题。针对 RNN 存在的问题,提出了一种带有门控机制的 LSTM 网络。LSTM 引入了门控机制,通过三个门控单元对输入的信息进行筛选和丢弃,有效避免了 RNN 存在的问题(图 1)。首先通过遗忘门,来对上一时刻输入的信息进行筛选,遗忘掉无关信息,保留有效信息。由 x_t 和 h_{t-1} 经过 sigmoid 函数得到一个 0—1 区间的值,公式如下：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

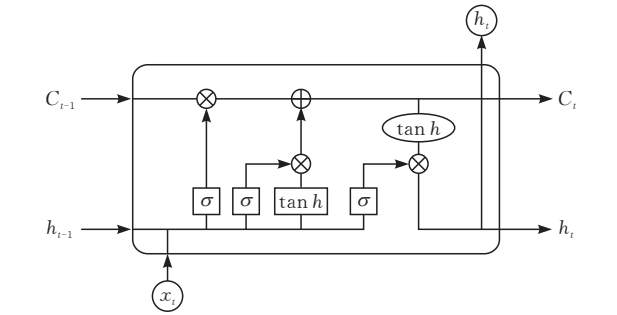


图 1 LSTM 内部结构图

接着通过输入门,对输入的信息进行选择性的保留,保留重要信息。由 x_t 和 h_{t-1} 经过 sigmoid 函数得到一个 0—1 区间的值,来决定当前细胞的更

新。将 x_t 和 h_{t-1} 经过 tanh 函数,得到候选值状态 \tilde{C}_t 。再由候选值状态 \tilde{C}_t 和 $t-1$ 时刻细胞状态来得到更新的细胞状态 C_t 。公式如下：

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{7}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{8}$$

最后通过输出门,得到 t 时刻的细胞状态 h_t ,传输给下一个时刻的 LSTM 细胞单元中,从而实现了获取距离的数据信息。由 h_{t-1} 和 x_t 经过 sigmoid 函数得到 o_t ,再由 t 时刻细胞状态 C_t 经过 tanh 函数与 o_t 相乘,得到最终输出细胞状态单元 h_t 。公式如下：

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = o_t * \tanh(C_t) \tag{10}$$

2 新闻推荐模型结构

提出一种基于知识图谱的新闻推荐神经网络模型(TransD,LSTM-CNN-Attention for News Recommendation,简称 TLCA-NR)。模型输入的数据为两种,分别是用户浏览过的新闻记录 and 候选新闻。该模型分为两部分,第一部分是通 过 TransD 和 GloVe 模型将原始新闻数据转化为特征向量,分别从语义层面和知识层面获得文本语义特征、实体特征和实体在知识图谱中的上下文特征。第二部分是推荐模型,使用 LSTM-CNN 挖掘更深层次的信息,实现知识层面和语义层面的信息融合,再通过注意力机制根据候选新闻不同动态计算用户偏好权重,将用户特征向量和候选新闻词向量通过点击求和运算做相似度计算,得到最终的推荐结果(图 2)。

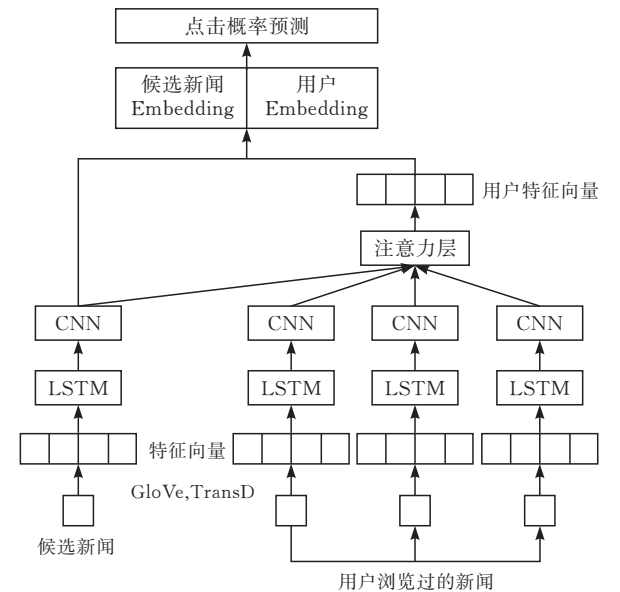


图 2 推荐模型结构

2.1 知识表示特征的获取

如今基于知识图谱的推荐系统有关研究受到广泛关注。现有的推荐系统技术,基于协同过滤的推荐算法存在数据稀疏和冷启动问题,其推荐效率不高。在知识图谱中存在着大量的知识实体和实体之间的关系,因此引入知识图谱对推荐效果的改善有很大的帮助。

实体嵌入信息表示的基本思路是将从新闻内容提取出的知识实体和来自 WikiData 知识图的知识图进行关联,通过 TransD 翻译模型从知识图谱中学习得到实体向量表示。在预处理过程中通过翻译模型得到新闻实体的向量,接着将文件中实体转化为对应的特征向量。由于通过翻译模型得到的特征向量,虽然能够体现新闻实体间的结构信息,但不能很好地体现效果,于是需要用知识图谱中的相邻实体来代表上下文实体,定义如下式:

$$context(e) = \{e_i \mid (e, r, e_i) \in Gor(e_i, r, e) \in G\}$$

(11)

G 是构建的知识图谱子图, r 是它们之间的关系,给定实体 e 的上下文,基于实体的上下文嵌入表示由上下文实体的平均值计算得到下式:

$$\bar{e} = \frac{1}{|context(e)|} \sum_{e_i \in context(e)} e_i$$

(12)

2.2 新闻语义特征向量的提取

以往的新闻推荐方法没有将新闻实体信息加入到推荐模型来,因此缺少新闻知识层面的特征,推荐的准确性很难达到理想效果。所以引入实体的知识层面信息,解决数据稀疏和冷启动问题,提高推荐精度。本方法提取了三种不同的特征:基于新闻的语义特征、实体特征和实体在知识图谱中的上下文特征。本文使用 GloVe 模型^[12]加载词向量,即可得到新闻语义信息的特征矩阵。GloVe 模型考虑了语料库中的全局信息,在机器理解、图像识别中的图片标注处理等强语义需求任务中有很好的效果。为了获得语义层面的特征,进一步提高推荐的准确性,在前文说明的来自 TransD 的实体向量的特征的基础上,本方法也使用了来自 GloVe 的语义词向量。

由于每个新闻标题长度的不一样,经过特征提取得到的特征矩阵长度也会不一样,所以在提取新闻语义特征之前,需要对新闻标题进行补齐或截断操作,如果文档中的单词数大于 k ,将文档截断为 k 个单词,如果文档中的单词数少于 k ,则将 0 填充到末尾,则所有新闻的标题长度确定为 k 。于是用 GloVe 模型得到新闻标题语义层面信息的向量,每个单词向量的维度是 n , n 是模型的超参数,再将每个单词的向量组合起来就可以得到 $n \times 10$ 的二维的

特征举证,即是新闻语义信息的特征矩阵。

2.3 基于 LSTM-CNN 的特征向量挖掘与融合编码

将上述提取的三种不同的矩阵进行统一,假设得到第一个基于语义的单词特征矩阵的格式是 $m_1 \times n$,另外两个基于实体和实体在知识图谱中的上下文特征的格式是 $m_2 \times n$, n 为新闻标题的长度是定值,不同的是 m_1 和 m_2 的值,因此将后面两个矩阵投影到第一个向量空间上,通过投影函数 f 如下所示:

$$Y = f(X)$$

(13)

其中 $X \in R^{m_2 \times n}$, $Y \in R^{m_1 \times n}$,让 X 左乘以矩阵 $M \in R^{m_1 \times m_2}$ 达到矩阵的统一,通过 ReLU 激活函数作非线性变换,得到其投影矩阵,最终得到统一的三种不同的特征矩阵。

本研究在将上述求得特征向量融合到一起,由于会出现以下几个问题,一是新闻语言具有非常强的逻辑性,语言前后顺序不同就会导致新闻表达的内容不同,这里通过循环神经网络(LSTM)去挖掘新闻标题中的逻辑性;二是经过知识图谱预处理的特征向量和经过 GloVe 提取的特征向量需要通过卷积神经网络(CNN)融合,于是使用 LSTM-CNN 网络做进一步的信息挖掘与融合。

新闻标题摘要中体现重要内容的一般是名词和动词来体现,比如一些时间、地点、人物和事件,其他介词、冠词的重要性可能没有那么多高,这里就通过 LSTM 遗忘门的机制,过滤掉新闻标题中的一些单词和实体,让这些词性的词对后续的推荐结果不产生什么影响。

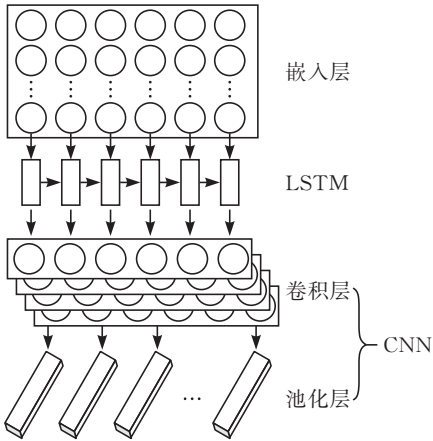


图 3 LSTM-CNN 网络结构

神经网络结构主要是两部分组成(图 3)。首先将嵌入层的输出结果送入到 LSTM 提取序列化的新闻特征向量,然后再通过卷积神经网络抽取空间结构信息,将多种信息融合就得到了用户的特征向量。下面将着重介绍具体操作方式,Embedding 的输入主要有两个,一个是经过 GloVe 预训练模型

处理的新闻标题 $w_{1,n}$ 得到的词向量, n 表示新闻标题的长度;另外是经过 TransD 模型得到的实体特征向量

$$E_{1,n} = [E_1, E_2, \dots, E_n]$$

其中 $E_1 \in P^{m \times 3}$, $E_{1,n} \in P^{n \times m \times 3}$, 数字 3 表示处理的词向量、实体的特征向量和实体对应的上下文特征向量三种向量, m 表示词向量的维度, n 表示新闻标题的长度。将经过处理的 $E_{1,n}$ 送入 LSTM 网络结构, 因为本文在 LSTM 网络模型中把输入维度和输出维度设置成相同, 经过 LSTM 的输出便得到了和词向量嵌入矩阵维度相同的 $E_{1,n}^{(LSTM)}$, 经过 LSTM 处理来提取新闻标题中的逻辑信息, 并将 $W_{1,n}^{(LSTM)}$ 送到 CNN 网络进行处理, 再通过 CNN 的卷积层抽取空间结构信息并融合用户的行为特征向量, 就可以得到最终用户的特征向量。这里的特征向量是通过 LSTM-CNN 神经网络处理用户的历史浏览新闻得到的, 也代表了该用户的综合偏好。

2.4 基于注意力机制的用户兴趣识别和预测

由于在线新闻平台中的题材和内容十分丰富, 用户在阅读中兴趣广泛, 用户感兴趣的主体可能不止一个, 但其浏览过的新闻会对推荐的结果产生一定的干扰。本文希望在面对不同的新闻场景时, 推荐模型会根据用户的历史浏览记录来动态分配不同的偏好权重, 所以通过注意力机制解决类似这样的问题。

本文通过注意力机制将当前候选新闻的特征向量和用户历史浏览的新闻特征向量逐条点积计算以获得两者之间的权重矩阵。对于用户的历史点击新闻 $\{t_1^i, t_2^i, \dots, t_{N_i}^i\}$, 其中 $t_j^i (j=1, \dots, N_i)$ 是用户 i 点击的第 j 条新闻的标题, N_i 是用户 i 点击的新闻总数。本文对每个候选新闻 t_j 分配不同的权重矩阵, 操作如下:

1) 通过上层神经网络得到了用户浏览新闻的词向量 $\{e(t_1^i), e(t_2^i), \dots, e(t_{N_i}^i)\}$ 和每条候选新闻的词向量 $e(t_j)$; 然后对每条新闻 $e(t_k^i)$ 做卷积运算得到新的带有权重的特征向量;

2) 将该带有权重的特征向量送入到 DNN 作为注意力网络, 得到针对当前候选新闻的加权矩阵, 再将加权矩阵送入到 softmax 层计算归一化, 得到加权矩阵

$$s_{tk,tj} = \frac{\exp(H(e(t_k^i), e(t_j)))}{\sum_{k=1}^{N_i} \exp(H(e(t_k^i), e(t_j)))} = \quad (14)$$

3) 注意力网络 DNN 接受两个新闻内容的嵌入作为输入并输出影响权重。所以用户 i 相对于候选新闻 t_j 的表达可以计算为用户点击新闻内容嵌入表

示的加权和, 即得到最终的用户点击新闻的词向量:

$$E(i) = \sum_{k=1}^{N_i} s_{tk,tj} e(t_k^i) \quad (15)$$

通过注意力机制得到了最终的用户特征向量, 接着将候选新闻词向量 $e(t_j)$ 和候选新闻的用户 i 的特征向量 $e(i)$ 做点积求和运算, 最终得到了两者之间的相似度

$$\rho_{i,tj} = G(e(i), e(t_j)) \quad (16)$$

其中 G 表示候选新闻词向量 $e(t_j)$ 和用户 i 的特征向量 $e(i)$ 点积运算, 再进行 reduce_sum 计算, $\rho_{i,tj}$ 表示用户 i 与候选新闻 t_j 的相似度分数, 计算出的分数越大代表用户对某一候选新闻兴趣更大, 最终分数会经过 sigmoid 函数的处理, 得到的值会在 0 到 1 之间, 该模型将计算出来的分数视作用户对候选新闻感兴趣的程度。最后使用 TOP-N 思想, 推送 N 个用户感兴趣的新闻内容来实现用户的个性化推荐。

3 实验验证与结果分析

3.1 数据集

数据集 (MIND)^[13] 是用于新闻推荐研究的大规模数据集, 作为新闻推荐的基准数据集, 促进新闻推荐和推荐系统领域的研究。该数据集集中的每篇新闻都有 ID、标题、摘要、类别和命名实体信息等详细信息。每个印象日志包含该用户之前的点击事件、未点击事件和历史新闻点击行为, 用 0 和 1 区分, 0 代表用户看到了新闻却没有点击进去浏览, 1 代表用户浏览过, 提供了划分好的训练集和测试集。

3.2 评价指标

推荐算法的好坏可以从 ROC 曲线下面积 (AUC)、平均倒数排名 (MRR) 和归一化折现累积收益 (NDCG) 三个性能指标进行评估。ROC 曲线下面积 (AUC) 的计算如下式:

$$AUC = \sum_{i \in Y} \frac{rank_i - \frac{P \times (P+1)}{2}}{M \times N}$$

P 表示预测得到的正样本在负样本之前的概率, Y 为样本数, M 为正样本数, N 为负样本数, $rank$ 是样本预测值的排名。

平均倒数排名 (MRR) 的计算如下式:

$$MRR = \frac{1}{Q} \sum_{i=1}^Q \frac{1}{rank_i}$$

MRR 通常是从推荐列表中正确检索结果值在检索结果中的排名, 从而来评估推荐系统的性能。

归一化折现累积收益 (NDCG) 的计算如下式:

$$NDCG@k = \frac{DCG_u@k}{IDCG_u}$$

NDCG 表示推荐系统某一用户返回的最好推

荐结果列表,即推荐结果的相关性越大,NDCG越大。

3.3 实验设置与结果分析

本实验使用深度学习框架中的 Tensorflow 框架进行开发,首先使用 GloVe 预训练模型提取新闻语义特征词向量,向量维度设置为 100,知识图谱实体嵌入向量是通过 TransD 方法从 WikiData 知识图里获取的,维度设为 100,最大标题词数为 16,对于用户历史记录文件,history_size 设置为 20;实验采用交叉熵作为损失函数来优化目标,并使用 Adam 优化器进行优化,学习率设置为 0.001,梯度裁剪参数为 0.5,每批设置为 128,迭代 50 次。

DKN^[7]是一个基于内容的深度推荐框架,用于点击率预测,它最大的特点是设计了一个多通道的知识感知卷积神经网络,融合了新闻的语义层面和知识层面的表示。

DeepFM^[4]是一个通用的深度推荐模型,它结合了解析机器的一个组件和共享输入的神经网络的一个组件。

DeepWide^[3]是一种基于深度学习的模型,它结合了前馈神经网络和线性模型进行推荐,还使用了连接新闻标题和个人资料嵌入作为特征。

LibFM^[14]是特征因子分解机模型的一个常用实现,该模型广泛用于点击率预测场景,将新闻标题和简介嵌入的串联作为输入特征。

为了验证本研究(TLCA-NR)在新闻推荐中的有效性,在相同的实验条件下,分别使用 DKN 模型、DeepFM 模型、DeepWide 模型、LibFM 模型和本文的模型进行实验对比。根据 AUC,MRR,NDCG 这些评估指标进行了以下比较,从表 1 可以看出,本文提出的模型对于 DKN 提升了 1.56%,对于 DeepFM 提升了 3.7%,对于 DeepWide 提升了 4.25%,对于 libFM 提升了 6.15%。

表 1 各模型在 MIND 上的结果对比

| 模型 | AUC | MRR | NDCG@5 | NDCG@10 |
|----------|--------|--------|--------|---------|
| TLCA-NR | 0.6608 | 0.3257 | 0.3463 | 0.3962 |
| DKN | 0.6452 | 0.3253 | 0.3382 | 0.3956 |
| DeepFM | 0.6238 | 0.2951 | 0.3157 | 0.3712 |
| DeepWide | 0.6183 | 0.2873 | 0.3096 | 0.3621 |
| LibFM | 0.5993 | 0.2823 | 0.3005 | 0.3585 |

本文还对基于 CNN 模型和基于 LSTM-CNN 的模型进行了对比,来证明 LSTM 能否通过遗忘门的机制来挖掘出新闻标题中的逻辑关系信息。图 4 可以看出,使用 LSTM-CNN 网络和单独使用 CNN 卷积神经网络在推荐指标上有更好的效果,体现出 LSTM 在抽取新闻特征方面起到了一定的作用。

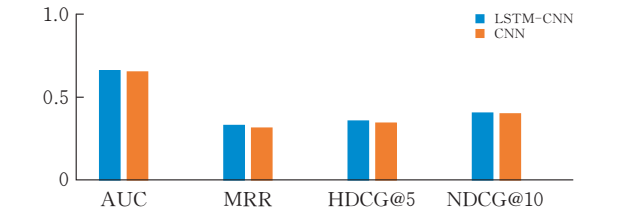


图 4 LSTM-CNN 和 CNN 模型比较

表 2 对知识图谱中实体对应的上下文特征是否引入对推荐模型的效果影响做出了对比:当引入知识图谱实体向量时,相对于仅使用词向量特征 AUC 指标提高了 1.14%;当引入实体向量和实体对应知识图谱中上下文特征时,相对于仅使用词向量特征 AUC 指标提高了 3.42%,从结果中可以看出使用了知识实体各项指标均优于不使用知识实体。

表 2 引入不同的特征向量对推荐结果影响

| 知识图谱的引入 | AUC | MRR | NDCG@5 | NDCG@10 |
|--------------|--------|--------|--------|---------|
| 仅使用词向量 | 0.6305 | 0.3083 | 0.3158 | 0.3592 |
| 使用词向量和实体 | 0.6419 | 0.3164 | 0.3229 | 0.3715 |
| 使用词向量、实体、上下文 | 0.6647 | 0.3257 | 0.3463 | 0.3962 |

4 结论

本文提出一种基于知识图谱的新闻推荐神经网络模型,将新闻知识层面信息融入到词向量中,使用 LSTM-CNN 网络挖掘新闻和用户深层次信息,再通过注意力机制根据候选新闻不同动态计算用户偏好权重,将用户特征向量和候选新闻词向量通过点击求和运算做相似度计算,得到最终的推荐结果。通过在真实数据集上的实验,证明了 TLCA-NR 模型的有效性。后续研究将进一步优化模型,可以考虑将新闻所属类别加入进来进一步提高新闻特征提取的精准性,从而来提高推荐的性能。

[参 考 文 献]

- [1] 孟祥武,胡勋,王立才,等.移动推荐系统及其应用[J].软件学报,2013, 24 (01): 91-108.
- [2] ÖZG ÖBEK Ö, GULLA J A, ERDUR R C. A survey on challenges and methods in news recommendation [C].WEBIST (2),2014: 278-285.
- [3] CHENG H-T, KOC L, HARMSSEN J, et al. Wide & deep learning for recommender systems[C].Proceedings of the 1st workshop on deep learning for recommender systems,2016: 7-10.
- [4] GUO H, TANG R, YE Y, et al. DeepFM: a factorization-machine based neural network for CTR predic-

tion[C]. Proceedings of the 26th International Joint Conference on Artificial Intelligence,2017: 1725-1731.

[5] 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: 1933-1942.

[6] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in neural information processing systems,2013, 26(04): 3111-3119.

[7] 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: 1835-1844.

[8] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [J]. Computer Science,2013,14(02):243-252.

[9] BORDES A, USUNIER N, GARCIA-DURÁN A, et al. Translating embeddings for modeling multi-relational data[C]. Proceedings of the 26th International Conference on Neural Information Processing Systems- Volume 2,2013: 2787-2795.

[10] LIN H, LIU Y, WANG W, et al. Learning entity and relation embeddings for knowledge resolution[J]. Procedia Computer Science,2017, 108(02): 345-354.

[11] JI G, HE S, XU L, et al. Knowledge graph embedding via dynamic mapping matrix[C]. Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers),2015: 687-696.

[12] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP),2014: 1532-1543.

[13] 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,2020: 3597-3606.

[14] RENDLE S. Factorization machines with libfm[J]. ACM Transactions on Intelligent Systems and Technology (TIST),2012, 3(03): 1-22.

News Recommendation Neural Network Model Based on Knowledge Graph

CHENG Huasong, XIONG Caiquan, KE Yuanzhi, WU Xinyun

(School of Computer Science , Hubei Univ. of Tech., Wuhan 430068, China)

Abstract: The goal of news recommendation is to push the news content of interest to users based on their historical reading habits. Existing methods suffer from insufficient feature learning and do not consider the relationship between users and browsing news, as well as the importance of different candidate news to users. Aiming at the above problems, this paper proposes a news recommendation neural network model based on knowledge graph. First, the Glove model (global log bilinear regression model) and the translation-based model (TransX) are used to obtain text semantic features, entity features, and context features of entities in the knowledge graph from news semantics and knowledge levels, respectively. Then, the LSTM-CNN network is used to mine the deep-level information of news and users to obtain the user's news preference vector, and at the same time, an attention mechanism is introduced to reduce the influence of irrelevant news topics; Finally, the user's preference value for candidate news is calculated by clicking the summation and the final news is generated recommended results. The experimental results on the real news dataset MIND show that the model proposed in this paper performs better in terms of AUC, MRR, and nDCG@k than the existing DKN, DeepFM, and DeepWide models.

Keywords: news recommendation; neural networks; knowledge graph; attention mechanism

[责任编辑：张岩芳]