

融合用户和产品信息的多头注意力情感分类模型^①



蒋宗礼, 张 静

(北京工业大学 信息学部, 北京 100124)

通讯作者: 张 静, E-mail: zhangjing63519@163.com

摘 要: 针对传统情感分类方法提取文本信息单一的问题, 提出了一种融合用户信息和产品信息的分层多头注意力的情感分类模型. 首先, 采用分层的多头注意力代替单一注意力, 从多个视角获取有效信息. 然后在每个注意力中都融入用户信息和产品信息, 挖掘出用户和产品信息在多个子空间上的表现特征, 使模型在多个子空间上得到更全局的用户偏好和产品特点对情感评分的影响. 实验结果表明, 模型在 IMDB、Yelp2013、Yelp2014 数据集上的准确率较之前基于神经网络的情感分析模型均有所提高.

关键词: 文档级情感分类; 多头注意力机制; 层次化注意力网络; 用户产品信息

引用格式: 蒋宗礼, 张静. 融合用户和产品信息的多头注意力情感分类模型. 计算机系统应用, 2020, 29(7): 131-138. <http://www.c-s-a.org.cn/1003-3254/7447.html>

Multi-Head Attention Model with User and Product Information for Sentiment Classification

JIANG Zong-Li, ZHANG Jing

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

Abstract: Aiming at the single information problem of traditional sentiment classification method, a multi-head attention model with user and product information is proposed. Firstly, hierarchical multi-head attention is used to replace single-head attention and obtain effective information from multiple perspectives. Secondly, using multi-head attention with user information and product information, mining the performance characteristics of user and product information in multiple subspaces, the model can get a more global impact of user preferences and product characteristics on sentiment score in multiple subspaces. Experimental results on IMDB, Yelp2013, and Yelp2014 datasets show that the performance of the proposed model is better than the other advanced baselines.

Key words: document-level sentiment classification; multi-head attention mechanism; hierarchical attention network; user and product information

情感分类是 NLP 领域的研究热点之一, 目的是根据文本的含义和情感表达, 将文本分为两种或多种类型. 情感分类在评论网站中有很广泛的应用, 例如 IMDB、豆瓣电影、大众点评等. 本篇工作的研究任务是文档级别的情感分类, 它是情感分析和观点挖掘的基础任务, 它根据文本中单一的目标或产品来决定该文本的情感极性. 传统的情感分类方法^[1-3]是手动地从

文本中抽取、设计和选择一些高质量的特征, 然后训练各种分类器. 近几年来, 深度学习方法在 NLP 领域取得了重大突破, 越来越多的研究者关注如何利用深度学习技术提高情感分类的效果. 近来的大多数工作利用了神经网络去学习低维空间上的文本表征, 因为评论文本有很明显的序列特征, 所以处理序列信息的递归神经网络 (Recurrent Neural Network, RNN)^[4]能够

① 收稿时间: 2019-11-06; 修改时间: 2019-11-28; 采用时间: 2019-12-11; csa 在线出版时间: 2020-07-03

很好的学习到评论中的情感信息,尤其是它的变体长短时记忆网络(Long Short-Term Memory, LSTM)^[5],通过引入记忆单元和门控机制,缓解了传统循环神经网络出现的梯度爆炸或者是梯度消失的问题,在情感分类领域中取得了较好的效果.随着 Attention 机制^[6-9]在图像分类、机器翻译等领域的成功应用,越来越多的研究者开始将其引入到情感分类任务中,重点关注对情感分类有明显影响的单词或词语.但是单一的注意力机制难以提取到更丰富的情景语义信息,它很难对一些句法信息、指代信息等进行建模,并不能很好地判别句子的情感极性.

为了解决提取的情感信息单一的问题,本文提出了一种融合用户信息和产品信息的分层多头注意力的情感分类模型(Hierarchical User and Product Multi-head Attention model, HUPMA),在词语层和句子层中分别用多头注意力提取出词语和句子的情感信息,并且在每个注意力中融入用户信息和产品信息.本文工作的主要贡献如下:

(1) 提出了一种新的情感分类模型框架,在词语层和句子层分别使用多头注意力来学习更长的序列信息,捕捉到更丰富的上下文语境信息.

(2) 在多头注意力中引入用户和产品信息,在多个子空间上得到更全局的用户偏好和产品特点对情感评分的影响.

本文分别在公开数据集 IMDB、Yelp2013、Yelp2014 上验证了模型效果,实验结果表明,在相同运行环境下,我们的模型相比于基于单头注意力的分层网络模型,在准确率上有所提升.

本文第 1 节介绍基于文档级别的情感分类的相关工作.第 2 节介绍模型的基本原理.第 3 节是本文的对比实验以及对实验结果的详细描述和分析.最后是对本文工作的总结和未来展望.

1 相关工作

文档级别的情感分类作为情感分析的基础任务之一,一直是情感分析领域的研究热点.最早使用的方法是基于情感词典的方法.基于情感词典的方法,一般依靠情感词典的构建和规则的建立,文献[10]通过情感词典及词性求出文本中情感词的得分,并将其按特定规则组合计算,来判断文本的情感倾向性.李寿山等^[11]利用英文种子词典,借助机器翻译系统,构建了中文情感

词典.情感词典构建的好坏决定了分类的效果,并且情感词典的构建依赖于人工统计,耗费大量的人力物力的同时存在着误差.

另一种情感分类的方式是基于机器学习和深度学习的方法,利用深度学习技术可以自动提取文本的深层特征,使文本情感分类研究得到了进一步的发展.主流的深度学习网络包括卷积神经网络(Convolutional Neural Networks, CNN)^[12]和 RNN,其中卷积神经网络通过卷积操作实现对文本的局部信息的捕捉,通过池化操作实现对局部重要信息的提取. Kim 等人^[13]将经典的 CNN 结构应用在情感分类任务上. Xiang Zhang 等^[14]从字符角度来训练卷积神经网络,使网络不需要提前知道关于语法和语义的信息.但是 CNN 只能提取到文本的局部信息,丢失了词语的顺序信息,而词语的顺序信息在文本处理上是不能忽略的.相比于 CNN, RNN 的内在结构使得它在文本序列建模方面更有优势,能够很好地捕捉上下文信息.尤其是 RNN 的变体 LSTM 引入了门控机制,缓解了 RNN 中梯度消失和梯度爆炸的问题. Tang 等^[15]为了对句子之间的关系进行建模,引入自底向上的基于向量的篇章表示方法,首先用 LSTM 实现单句表示,再用 RNN 编码句子间的内在关系和语义联系,采用层次化 RNN 模型来对篇章级文本进行建模.该模型相比标准的循环神经网络模型实现很大的效果提升. Siwei Lai 等^[16]提出了 RCNN 模型,先使用双向循环神经网络得到上下文表示,再经过卷积、池化操作后输出分类结果.注意力机制(Attention Mechanism)使得模型能够在训练过程中高度关注特定的词语,更好地捕捉句子中的重要成分,提取出更多的隐藏特征. Wang 等^[17]采用方面嵌入的方式来生成注意力向量,关注句子的不同部分. Yang 等人^[18]提出了分层注意力网络(Hierarchical Attention Network, HAN)模拟文本的构成方式,对单词和句子进行分层编码,并根据其注意力权重融合成文本表示,不仅捕捉了句子中比较重要的词语,还给篇章中比较重要的句子赋予更高的权重.实验表明,结合注意力机制的神经网络模型在情感分类任务中取得了比传统方法更好的效果.在网络评论中,有些词语是表达用户情感的,有些词语则是对产品本身的评价.为了将这些用户和产品信息引入情感分类任务中,研究学者们在这方面做了很多工作,并且都获得了一定的提升.唐都钰等人^[19]在 CNN 网络的基础上添加了用户和产品的偏置矩阵和表示矩

阵来表示这些信息. 考虑到用户偏好矩阵训练比较困难, 陈慧敏等^[9]在层次网络中融入用户和产品的相关信息, 考虑了全局的用户偏好以及产品特点, 进一步提高了模型性能. 但是上述模型提取到的情感信息相对比较单一, 受 Transformer 模型^[20]的启发, 本文提出一种融合用户信息和产品信息的分层多头注意力的情感分类模型 HUPMA. 首先该模型利用分层多头注意力, 分别在词语层和句子层中采用不同的隐层向量权重参数, 形成指代信息更丰富的注意力权重分布; 其次, 在多头注意力中融入用户信息和产品信息, 同时采用不同的用户向量权重和产品信息权重参数, 形成多个子空间下不同的用户偏好和产品特点对情感评分的影响, 从而能够对指代信息丰富的上下文进行建模, 同时挖掘出用户信息和产品信息在多个维度上的深层特征.

2 模型

融合用户信息和产品信息基于分层多头注意力的情感分类模型 HUPMA 如图 1 所示.

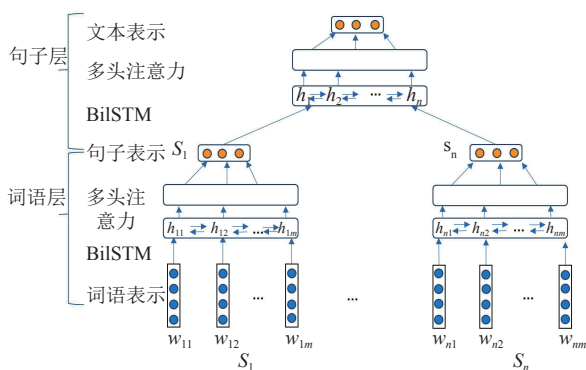


图 1 HUPMA 模型架构图

模型主要包含了 4 个部分: 词汇层编码、词汇层的多头注意力机制、句子层编码、句子层的多头注意力机制, 其中在词汇层的多头注意力机制和句子层的多头注意力机制中融入了用户和产品信息. 首先, 在模型的词语层, 利用双向 LSTM 计算单词的向量表示, 然后用多头注意力结合用户、产品信息来学习单词的注意力权重分布, 进而生成句子的向量表示; 在模型的句子层, 同样利用双向 LSTM 计算句子的向量表示, 用多头注意力结合用户、产品信息来学习句子的注意力权重分布, 得到文本的向量表示. 最后经过一个 Softmax 层, 完成评论文本的情感分类.

2.1 基于 LSTM 的分层网络结构

本模型采用“词-句子-文章”的自下而上的层次化结构来表示一篇文本, 获取更丰富的上下文情感特征, 既符合人类的阅读习惯, 也缓解了过长的序列导致的长句依赖的问题. LSTM 是一种带门控机制的循环神经网络, 通常被用于处理序列文本数据, 可以避免传统的循环神经网络的梯度消失或梯度爆炸的问题, 另外, LSTM 通过引入记忆单元和输入门、输出门、遗忘门 3 种门控机制来捕捉句子的依赖关系和上下文语境信息.

(1) 词语层 LSTM 网络

定义文档中的某个句子 $s_i = [x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}]$, x_{ij} 表示第 i 个句子的第 j 个词语, 首先将句子中的词语 x_{ij} 编码为向量表示 w_{ij} , 通过双向 LSTM 编码 w_{ij} 的上下文, 得到隐层表示 $h_i = [h_{i1}, h_{i2}, h_{i3}, \dots, h_{in}]$, 其中 $h_{ij} \in \mathbb{R}^d$, d 为 LSTM 单元输出向量的维度. 每一时刻的 LSTM 单元的输入由上一时刻隐藏层的输出和本时刻输入的词语向量组成. 具体计算方法如下:

$$\vec{h}_{ij} = \overrightarrow{LSTM}(w_{ij}) \quad (1)$$

$$\overleftarrow{h}_{ij} = \overleftarrow{LSTM}(w_{ij}) \quad (2)$$

$$h_{ij} = [\vec{h}_{ij}, \overleftarrow{h}_{ij}] \quad (3)$$

\overrightarrow{LSTM} 和 \overleftarrow{LSTM} 分别表示前向 LSTM 和后向 LSTM, \vec{h}_{ij} 和 \overleftarrow{h}_{ij} 表示前向 LSTM 生成的隐藏状态向量和后向 LSTM 生成的隐藏状态向量, 然后将 \vec{h}_{ij} 和 \overleftarrow{h}_{ij} 做连接操作, 生成最终的隐藏状态向量 h_{ij} .

(2) 句子层 LSTM 网络

因为同一评论中不同句子之间有一定的情感联系, 挖掘这种相互联系有助于整个评论文本的情感极性分析, 因此本文除了词语层的 LSTM 网络外, 还使用一个句子层的双向 LSTM 网络来进一步挖掘句子之间的依赖关系. 通过词语层之后就得到了句子向量 s_i , 然后句子层的 LSTM 网络将句子向量 s_i 编码为句子层的隐层向量 h_i .

$$\vec{h}_i = \overrightarrow{LSTM}(s_i) \quad (4)$$

$$\overleftarrow{h}_i = \overleftarrow{LSTM}(s_i) \quad (5)$$

$$h_i = [\vec{h}_i, \overleftarrow{h}_i] \quad (6)$$

\vec{h}_i 和 \overleftarrow{h}_i 分别表示前向 LSTM 生成的隐藏层句子向

量和后向 LSTM 生成的隐藏层句子向量, 同样拼接之后生成最终的句子层的隐藏向量 h_i .

2.2 融入用户和产品信息的多头注意力机制

(1) 多头注意力机制

注意力机制的主要目的是从众多信息中选择出对当前任务目标更关键的信息. 传统的单头注意力机制难以提取到更丰富的情景语义信息, 进而对评论文本的情感分类的效果也产生了影响, 为了从多个子空间上学习更加丰富的情感信息, 本文采用了多头注意力代替传统的单头注意力, 如图 2.

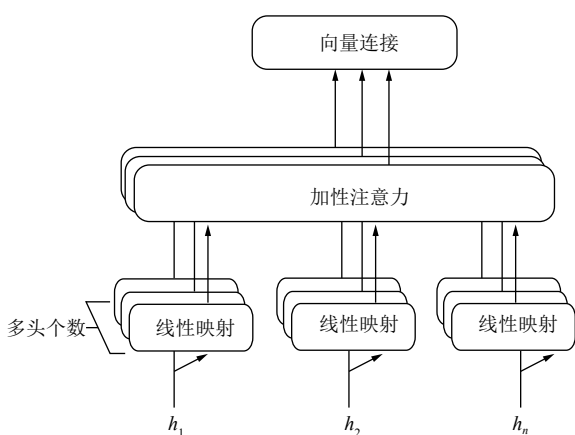


图 2 多头注意力机制

首先将 LSTM 层产生的隐层向量表示拆分为多个头, 线性变换后然后送入加性注意力层, 重复计算 $head_num$ 次, 其中 $head_num$ 为多头的个数, 最后拼接所有的 $attention$ 值作为下一模块的输入. 考虑到不同的用户、不同的产品会对情感分析造成一定的影响, 因此在多头注意力机制中引入了用户和产品信息, 其中每个 $head$ 的注意力计算如图 3.

(2) 词语层的多头注意力机制

不同的词语对句子的情感表达的重要度不同, 为了深度挖掘每个词语在多个子空间上的情感信息表达, 在词语层引入多头注意力机制, 同时不同的用户、不同的产品在评论中的情感用词也有差异, 因此, 将用户、产品信息融入到多头注意力的计算中, 计算方式如下:

$$s_i = \sum_j \alpha_{ij} h_{ij} \quad (7)$$

h_{ij} 是词语层的 LSTM 产生的隐层向量, α_{ij} 是在考虑用户信息和产品信息后, 句子中第 j 个词的权重, 度

量的是这个词语在整个句子 s_i 的重要程度. α_{ij} 由式 (8) 和式 (9) 计算而来:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_j \exp(e_{ij})} \quad (8)$$

$$e_{ij} = v_w^T \tanh(W_{wh}h_{ij} + W_{wu}u + W_{wp}p + b_w) \quad (9)$$

其中, e_{ij} 是向量 α_{ij} 未归一化时的权重表示, v_w^T 、 W_{wh} 、 W_{wu} 、 W_{wp} 和 b_w 分别表示计算 e_{ij} 时的前馈神经网络中对应的点积权重、隐层向量权重、用户向量权重、产品向量权重和偏置, 初始赋随机向量, 作为模型训练的参数.

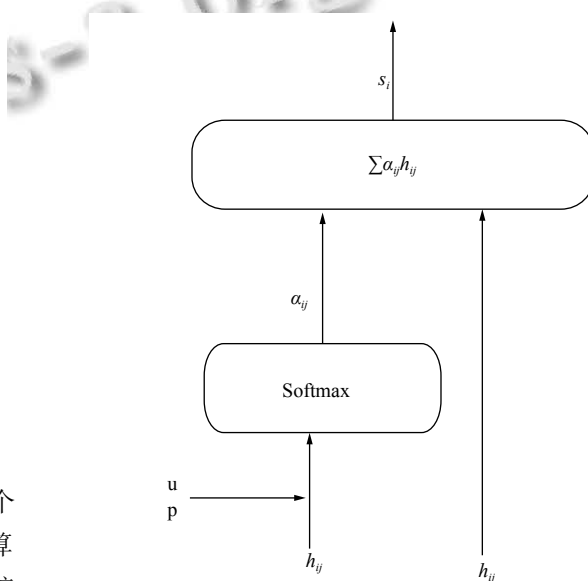


图 3 加性注意力机制

(3) 句子层的多头注意力机制

和词语层的多头注意力机制类似, 一条评论文本中的不同句子对最后的情感表达也有着不同的重要度, 因此在句子层引入多头注意力机制, 计算方式如下:

$$d = \sum_i \beta_i h_i \quad (10)$$

h_i 是经过句子层的 LSTM 产生的隐层向量, β_i 是在考虑用户信息和产品信息后的注意力权重, 度量的是某个句子在整个评论文本 d 中的重要程度. β_i 由式 (11) 和式 (12) 计算而来:

$$f_i = v_s^T \tanh(W_{sh}h_i + W_{su}u + W_{sp}p + b_s) \quad (11)$$

$$\beta_i = \frac{\exp(f_i)}{\sum_i \exp(f_i)} \quad (12)$$

其中, f_i 是向量 β_i 未归一化时的权重表示, v_s^T 、 W_{sh} 、 W_{su} 、 W_{sp} 和 b_s 分别表示计算 f_i 时的前馈神经网络中对应的点积权重、隐层向量权重、用户向量权重、产品向量权重和偏置, 初始赋随机向量, 作为模型训练的参数。

2.3 优化模型

本文的任务是对给定的评论文本, 判断其情感极性, 是一个情感分类的任务, 因此采用交叉熵损失函数作为代价函数, 通过最小化损失函数来训练模型, 找到最优参数。同时, 为了避免过拟合, 加入了 L2 正则化作为惩罚项, 损失函数公式如下:

$$loss = - \sum_i \sum_k y_i^k \log \hat{y}_i^k + \lambda \|\theta\|^2 \quad (13)$$

其中, y_i^k 表示第 i 篇评论文本的真实的情感类别, \hat{y}_i^k 表示模型对第 i 篇评论文本预测的情感类别, k 为情感类别的某一类, 以数字表示。 λ 表示 L2 正则化的惩罚系数, λ 越大, 惩罚力度越大。 θ 表示模型中的全部训练参数。

2.4 模型复杂度分析

在模型的词语层的多头注意力和句子层的多头注意力中, 对每个头的注意力都在维度上做压缩。首先将底层 LSTM 产生的隐层向量进行维度上的拆分, 拆分之后的维度为 $d_{head} = d_{lstm}/head_num$, 其中 d_{lstm} 为底层 LSTM 输出向量的维度, $head_num$ 是多头的个数。然后在维度为 d_{head} 的向量上做注意力函数运算, 所以在多头注意力中虽然需要计算多次, 但是相较于之前的单一注意力, 整体的计算成本并没有增加。同时, 多个

头之间的注意力计算是相互独立的, 因此可并行运算, 进一步提高模型的性能。

3 实验分析

3.1 实验设置

为了验证本文模型的有效性, 分别在 3 个情感分类任务的数据集上进行实验, IMDB、Yelp2013、Yelp2014 这 3 个数据集是 Tang 等^[19]构建的公开数据集。IMDB 是用户对电影评分数据集, Yelp2013 和 Yelp2014 是用户对商家的评分数据集。在数据预处理阶段, 本研究采用 Stanford CoreNLP^[21]实现评论文本中词语的切分和句子的切分, 同时提取数据集中的用户信息和产品信息, 预处理后的数据集统计如表 1。其中 IMDB 划分 10 个类别, Yelp2013 和 Yelp2014 划分 5 个类别, 级别越高代表评价越好。

将每个数据集中 80% 作为训练集, 10% 作为验证集, 剩余 10% 作为测试集, 用 Accuracy 和 RMSE 作为模型分类效果的评价指标。采用文献[9]中用 SkipGram 模型^[22]训练好的词向量作为初始化的词向量, 维度为 200, 此外, 用户和产品信息维度也设定为 200 维, 并随机初始化, 随着模型一起训练。LSTM 层的神经元维度设为 200 维, 用 Adam^[23]算法来优化超参数, 该算法通过计算梯度的一阶矩估计和二阶矩估计, 能够为不同的参数设计独立的自适应学习率。然后在验证集上选择表现最好的超参数作为模型参数, 最后用测试集测试模型在情感分类任务上的效果。

表 1 IMDB, Yelp2013 和 Yelp2014 数据集统计数据

数据集	类别	评论	用户	产品	评论/用户	评论/产品	句子/评论	词语/句子
IMDB	10	84 919	1310	1635	64.82	51.94	16.08	24.54
Yelp2013	5	78 966	1631	1633	48.42	48.36	10.89	17.38
Yelp2014	5	231 163	4818	4194	47.97	55.11	11.41	17.26

3.2 基准模型

本文模型将与以下基准模型进行实验对比。

(1) Majority 将训练集中出现次数最多的情感分类标签作为测试集的预测标签。

(2) Trigram 以评论文本的一元语法、二元语法和三元语法为特征, 以支持向量机 (SVM)^[24]为分类器, 训练模型。

(3) TextFeature^[25]将提取的词语和字符的 N 元语法、情感词典等作为特征, 训练 SVM 分类器。

(4) UPF^[26]从训练集中抽取 use-lenency 特征和相关产品特征, 作为 Trigram 方法和 TestFeature 方法的特征。

(5) AvgWordvec 用 Word2Vec 方法获得训练集和验证集的词语向量表示, 然后以词向量的平均向量为评论向量, 然后用这个向量为特征训练 SVM 分类器。

(6) NSC+UPA 在层次化的文本分类模型中引入了用户信息和产品信息来提高文档级别情感分类的效果。

3.3 模型性能对比

表2列出了HUPMA模型及基准模型的情感分类结果,我们将这些结果分为两组,第一组是仅仅考虑文本信息的模型,第二组是同时考虑文本信息以及用户和产品信息模型.为了验证分层多头注意力和用户产品信息对本文模型的性能影响,将用户信息和产品信息去掉,得到HUPMA(不含UP)的实验结果.HUPMA(含UP)一行是融合用户信息和产品信息的分层多头注意力的模型分类结果.

(1) 分层多头注意力的有效性

第1组的实验采用的是不考虑用户和产品信息分类方法,包括统计模型、传统的机器学习模型和神经网络模型,对比实验结果可看到, Majority 效果很差,因为它没有考虑任何文本信息、用户信息和产品信息,仅仅用一种统计思想寻找频率最高的情感分类标签.同样是基于机器学习的SVM分类器训练模型,用基于 N 元语法和情感词典的TextFeature比基于平均词向量的AvgWordvec在IMDB、Yelp2013和Yelp2014三个数据集上的情感分类准确率分别高出了9.8、3和4.2个百分点,说明采用了 N 元语法模型的词向量能够利用上下文信息丰富语义特征,比平均词向量效果更好.NSC模型和NSC+LA模型相比TextFeature在IMDB数据集上的情感分类准确率分别提高了4.1和8.5个百分点,在Yelp13数据集上的情感分类准确率分别提高了7.1、7.5个百分点,说明神经网络模型比传统的机器学习模型在情感分类任务上的表现更好,尤其是加入了注意力机制后,模型效果得到了进一步提升.不考虑用户信息和产品信息的情况下,本文提

出的HUPMA模型(不含UP)在3个数据集上的准确率比基准模型均提高了一到两个百分点,这也验证了引入分层多头注意力思想在提取文本上下文信息的有效性.

(2) 融合用户信息和产品信息的有效性

第2组的实验采用的是考虑用户和产品信息分类方法,结果表明,加入了用户信息和产品信息的模型相比于仅仅考虑评论文本信息的模型分类准确率更高,Trigram模型加入用户和产品信息后,在IMDB和Yelp2013数据集上的分类准确率分别得到了0.5、0.1个百分点的提升,NSC+UPA模型相比NSC+LA模型在3个数据集上的分类准确率分别得到了4.6、1.9、3.7个百分点的提升,通过对比可以发现,在神经网络结构中,用户和产品信息的引入对模型的提升效果更加明显.与不考虑用户和产品信息HUPMA模型相比,HUPMA(含UP)模型在IMDB数据集上的准确率提高了4.3个百分点,而在Yelp2013和Yelp2014两个数据集上的准确率仅提高了2.2个百分点,原因是IMDB数据集上的用户数据和产品数据相对比较丰富.这一结果验证了融入用户偏好和产品特征之后的模型在情感分类上的有效性.

(3) 融合用户和产品信息的多头注意力模型的有效性

同样是引入了用户和产品信息的分层的网络结构,本文模型HUPMA比NSC+UPA模型在3个数据集上的情感分类准确率分别提升了1.2、2、0.9个百分点.这一结果验证了融入用户偏好和产品特征之后的分层多头注意力模型在情感分类上的有效性.

表2 IMDB、Yelp2013和Yelp2014数据集上的情感分类结果

模型	IMDB		Yelp2013		Yelp2014		
	Acc	RMSE	Acc	RMSE	Acc	RMSE	
不考虑用户和产品信息模型	Majority	0.196	2.495	0.411	1.060	0.392	1.097
	Trigram	0.399	1.783	0.569	0.814	0.577	0.804
	TextFeature	0.402	1.793	0.556	0.845	0.572	0.800
	AvgWordvec+SVM	0.304	1.985	0.526	0.898	0.530	0.893
	NSC	0.443	1.465	0.627	0.701	0.637	0.686
	NSC+LA	0.487	1.381	0.631	0.706	0.630	0.715
	HUPMA(不含UP)	0.502	1.288	0.648	0.664	0.654	0.665
考虑用户和产品信息模型	Trigram+UPF	0.404	1.764	0.570	0.803	0.576	0.789
	TextFeature+UPF	0.402	1.774	0.561	1.822	0.579	0.791
	NSC+UPA	0.533	1.281	0.650	0.692	0.667	0.654
	HUPMA(含UP)	0.545	1.253	0.670	0.659	0.676	0.641

3.4 模型运行时间对比

为了验证本文模型 HUPMA 的运行时间,在同等运行环境下,在 Yelp2013 数据集上分别运行单头注意力模型和多头注意力模型,得到不同的模型单次迭代的运行时间,如表 3。从表中可以看出,多头注意力模型的单次迭代运行时间和单头注意力模型的单次迭代运行时间相差不大。

表 3 不同模型完成单次迭代的训练时间(单位: s)

模型	运行时间
单头注意力模型	1118.7
多头注意力模型	1147.2

4 结论与展望

本文提出了一种融合用户信息和产品信息的分层多头注意力的情感分类模型,首先,模型采用“词-句子-文章”的层次化结构来表示一条评论文本,为了从多个视角选择有用的信息,在词语层和句子层都引入多头注意力的思想。然后在每层注意力中融入了用户信息和产品信息,使模型在多个子空间上得到更全局的用户偏好和产品特点对情感评分的影响。实验结果表明,在同等条件下,本文模型在 IMDB 和 Yelp 数据集上的准确率均优于基准模型。在今后的工作中,我们试图将多头注意力引入基于 Aspect 的情感分类任务中,从更丰富的特征空间中识别评论文本中多个实体的情感极性。

参考文献

- Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. Philadelphia, PA, USA. 2002. 79–86. [doi: 10.3115/1118693.1118704]
- Pang B, Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Ann Arbor, MI, USA. 2005. 115–124.
- Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2008, 2(1–2): 1–135. [doi: 10.1561/1500000011]
- Elman JL. Finding structure in time. Cognitive Science, 1990, 14(2): 179–211. [doi: 10.1207/s15516709cog1402_1]
- Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: 10.1162/neco.1997.9.8.1735]
- Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv: 1409.0473, 2014.
- Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany. 2016. 207–212.
- Yin WP, Schutze H, Xiang B, et al. ABCNN: Attention-based convolutional neural network for modeling sentence pairs. Transactions of the Association for Computational Linguistics, 2016, 4: 259–272.
- Chen HM, Sun MS, Tu CC, et al. Neural sentiment classification with user and product attention. Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing. Austin, TX, USA. 2016. 1650–1659.
- Taboada M, Brooke J, Tofiloski M, et al. Lexicon-based methods for sentiment analysis. Computational Linguistics, 2011, 37(2): 267–307. [doi: 10.1162/COLI_a_00049]
- 李寿山, 李逸薇, 黄居仁, 等. 基于双语信息和标签传播算法的中文情感词典构建方法. 中文信息学报, 2013, 27(6): 75–81. [doi: 10.3969/j.issn.1003-0077.2013.06.011]
- LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278–2324, doi: 10.1109/5.726791.
- Kim Y. Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar. 2014. 1746–1751.
- Zhang X, Zhao JB, LeCun Y. Character-level convolutional networks for text classification. Advances in Neural Information Processing Systems. 2015. 649–657.
- Tang DY, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal. 2015. 1422–1432.
- Lai SW, Xu LH, Liu K, et al. Recurrent convolutional neural networks for text classification. Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin, TX, USA. 2015. 2267–2273.
- Wang YQ, Huang ML, Zhu XY, et al. Attention-based LSTM for aspect-level sentiment classification. Proceedings

- of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, TX, USA. 2016. 606–615.
- 18 Yang ZC, Yang DY, Dyer C, *et al.* Hierarchical attention networks for document classification. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, CA, USA. 2016. 1480–1489.
- 19 Tang DY, Qin B, Liu T. Learning semantic representations of users and products for document level sentiment classification. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China. 2015. 1014–1023.
- 20 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Advances in Neural Information Processing Systems. 2017. 5998–6008.
- 21 Manning CD, Surdeanu M, Bauer J, *et al.* The Stanford CoreNLP natural language processing toolkit. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Baltimore, MD, USA. 2014. 55–60.
- 22 Mikolov T, Chen K, Corrado GS, *et al.* Efficient estimation of word representations in vector space. Proceedings of 1st International Conference on Learning Representations. Scottsdale, AZ, USA. 2013.
- 23 Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv: 1412.6980, 2014.
- 24 Fan RE, Chang KW, Hsieh CJ, *et al.* LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research, 2008, 9: 1871–1874.
- 25 Kiritchenko S, Zhu X, Mohammad SM. Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research, 2014, 50: 723–762. [doi: [10.1613/jair.4272](https://doi.org/10.1613/jair.4272)]
- 26 Gao WL, Yoshinaga N, Kaji N, *et al.* Modeling user leniency and product popularity for sentiment classification. Proceedings of the 6th International Joint Conference on Natural Language Processing. Nagoya, Japan. 2013. 1107–1111.