

结合主题和分层注意混合网络的文本情感分析

饶冬章, 任淑霞, 赵宗现

天津工业大学计算机科学与技术学院, 天津

收稿日期: 2022年10月10日; 录用日期: 2022年11月8日; 发布日期: 2022年11月15日

摘要

文本情感分析一直以来都是自然语言处理的研究热点, 近几年, 深度神经网络在文本情感分析任务中取得了不错的效果。尽管取得了进展, 但提出的模型没有利用整个语料库的统计信息, 也没有将文档的体系结构的知识纳入到模型中。针对上述问题, 本文提出了一种结合主题和分层注意混合网络的文本情感分析模型。首先, 利用主题模型对数据集的主题进行提取, 并结合文本的词嵌入和句子嵌入来丰富特征空间, 以解决传统神经网络无法融入数据统计信息的问题; 然后, 采用卷积神经网络来降低特征空间的维度, 同时, 学习关键的主题信息; 最后, 使用带有主题感知的分层注意网络对模型进行训练, 来关注文本中更重要的单词和句子。实验结果表明, 提出的模型具有更好的分类性能, 能够更好地揭示文本的情感。

关键词

文本情感分析, 主题模型, 分层注意力机制, 混合神经网络

Text Sentiment Analysis of Combining Topic and Hierarchical Attention Hybrid Networks

Dongzhang Rao, Shuxia Ren, Zongxian Zhao

School of Computer Science and Technology, Tiangong University, Tianjin

Received: Oct. 10th, 2022; accepted: Nov. 8th, 2022; published: Nov. 15th, 2022

Abstract

Text sentiment analysis has long been a hot research topic in natural language processing, and in recent years, deep neural networks have achieved good results in text sentiment analysis tasks. Despite the progress made, the proposed models do not make use of statistical information from the whole corpus, nor do they incorporate knowledge of the architecture of the documents into the models. To address these issues, this paper proposed a model for text sentiment analysis that combines thematic and hierarchical attention hybrid networks. Firstly, a topic model is used to

extract the topics of the dataset and combine the word embeddings and sentence embeddings of the text to enrich the feature space to solve the problem that traditional neural networks cannot incorporate statistical information from the data; then, a convolutional neural network is used to reduce the dimensionality of the feature space while learning key topic information; finally, the model is trained using a hierarchical attention network with topic awareness to focus on more important words and sentences in the text. The experimental results show that the proposed model has better classification performance and can better reveal the sentiment of the text.

Keywords

Text Sentiment Analysis, Topic Models, Hierarchical Attention Mechanisms, Hybrid Neural Networks

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着网络上的社交媒体逐渐流行,人们很愿意在例如微博、推特等社交网络上发表自己对某个产品、事件等主体的情感,对这些有着情感倾向的数据进行分析,可以了解用户的想法和意见[1]。因此,文本情感分析一直以来都是自然语言处理的研究热点。

早期的大多数情感分析方法都是在手动提取的有效特征上训练浅层模型,这些模型通常采用传统机器学习的情感分析方法,包括支持向量机(SVM) [2]、主题模型(LDA) [3]、朴素贝叶斯[4]等等运用统计学的知识对文本的情感进行分析。这类方法存在不能充分利用上下文文本的语境信息的问题。近年来,深度学习方法被应用到情感分析任务中,其中,递归神经网络(RNN)和卷积神经网络(CNN)较为流行。基础的 RNN 不擅长处理较长的数据,此时数据会存在梯度消失和梯度爆炸[5]。长短期记忆网络 LSTM 和门控循环神经网络 GRU 的提出解决了这一问题,可以对长期依赖关系进行建模。Kim [6]提出通过一维卷积来获取句子中 n-gram 的特征表示,其对文本局部特征的抽取能力很强,在情感分析任务有着不错的表现。随着注意力机制在图像领域的兴起,Google 团队[7]实现了一种带有注意力机制的递归神经网络,可以帮助模型关注更重要的情感内容。注意力机制可以通过分配权重来突出上下文重要的信息,并有效地结合 RNN 和 CNN。Yang 等[8]利用 GRU 对单词和句子编码,结合分层的注意力,来捕捉重要的单词和句子。基于 CNN 和 RNN 的优点,人们也在探索这两个神经网络的结合。Behera 等[9]将单层 CNN 的输出输入到 LSTM 中,发现组合的性能要优于单独的 CNN 或 RNN 模型。Pasupa 等[10]以不同的方式融合了卷积神经网络和长短期记忆网络,发现最好的混合式深度学习是 BLSTM-CNN。Basiri 等[11]应用两个独立的双向 LSTM 和 GRU 层,同时考虑两个方向的上下文来预测情感极性。然而,传统的基于 LSTM 的模型具有一定的局限性,且没有考虑文档的结构层次,尤其是在对长文本数据进行情感分析的时候,过长的序列会存在丢失信息的问题。此外,它们没有充分利用所有的语料。具体地说,它们只利用了局部的上下文信息,没有利用整个文档的统计信息,这样得到的情感信息不够准确,影响模型得分。

为了解决上述问题,受到生成摘要中的 LDA 分布结合神经网络[12]的启发,本文提出了一种以分层注意网络为基础的新模型 THAHNN (Topic-Aware Hierarchical Attentional Hybrid Neural Networks)。THAHNN 主要思想是将预训练生成的 LDA 分布分配给文档词级和句子级的编码层,将数据集中全局共现的信息和编码层的局部内容窗口的语义特征相结合,使得模型既包含了局部的上下文信息,又融入了

整个文档的统计信息，充分利用了所有的语料。与此同时，使用带有主题感知的分层注意网络对模型进行训练，有效地将文档结构的知识融入到体系结构中，来关注文本中更重要的单词和句子。此外，引入卷积神经网络来降低特征空间的维度，解决模型参数数量过多的问题。为了验证框架的有效性，在两个数据集上进行了实验，分别是超党派新闻情感检测任务和亚马逊商品评论数据集。并将其情感分类的结果与几个流行的神经网络模型进行了比较，结果表明，该模型具有更高的精度。

2. 模型

结合主题的分层注意混合网络的总体结构如图 1 所示，为了解决传统神经网络无法融入文本统计信息的问题，本文将主题模型 LDA 生成的主题词分布和文档主题分布分别作为单词编码层的附加特征和句子编码层的附加特征，使得模型充分利用了所有的语料。单词编码层的附加特征和词嵌入技术 glove 相连接作为模型的输入，同时，联合 CNN 层来学习重要的主题信息，减少参数数量，提取更丰富的特征。最后，使用分层注意网络对单词和句子依次进行编码，有效地结合了文档的层次结构，将得到的文档表示与文档主题分布作为附加特征连接起来，以做出最终预测。

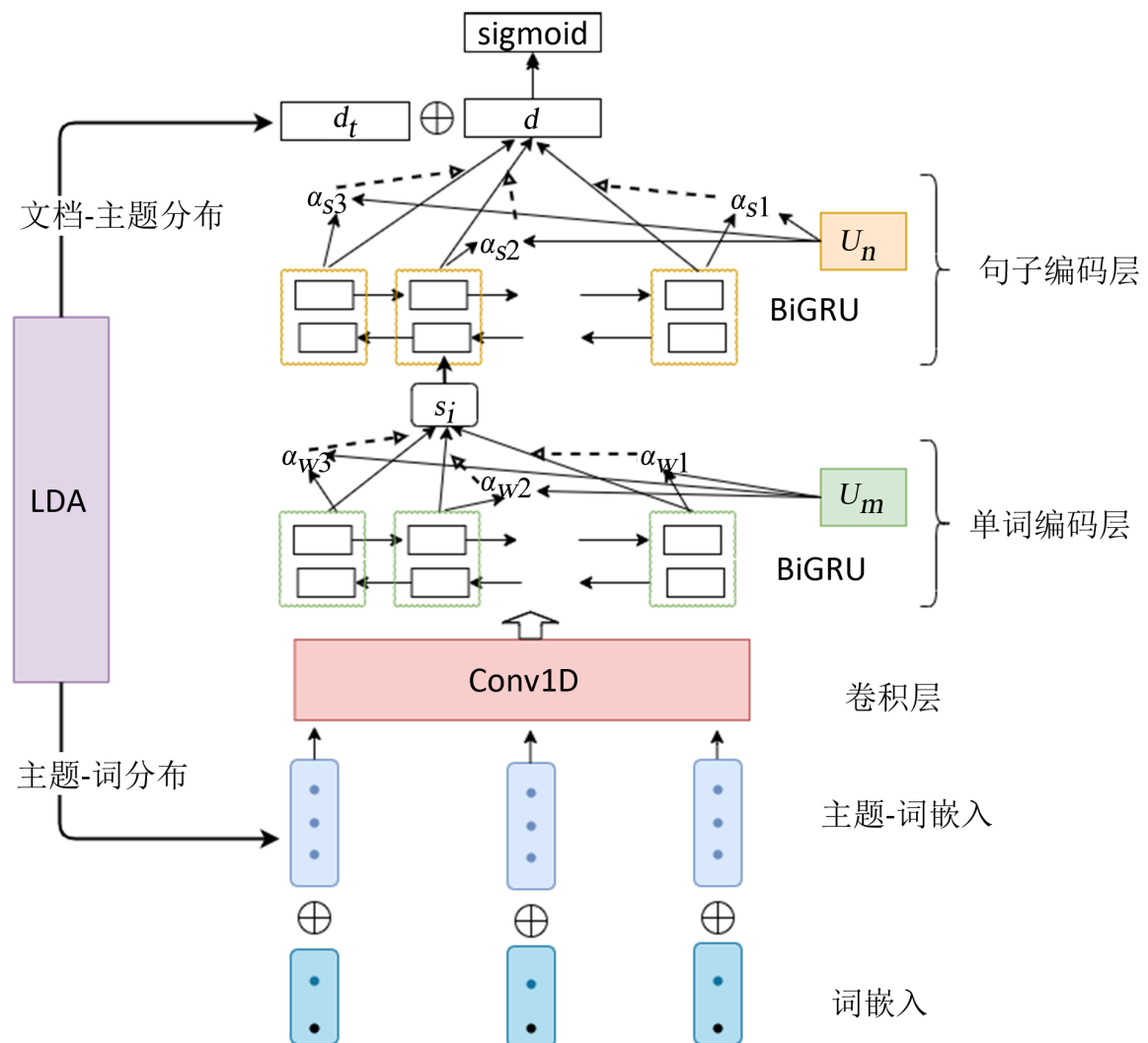


Figure 1. THAHNN model structure
图 1. THAHNN 模型结构

2.1. 神经主题模型

传统的 LDA 主题模型采用词袋的方法，将每一篇文档视为一个词频向量，通过分析文档数据，抽取它们的主题分布，之后根据主题分布进行文本分类。但是词袋方法没有考虑词与词之间的顺序，导致模型不能利用文本上下文语义。本文提出一种将主题与神经网络语言模型联合训练的机制，既保证了句子的序列语义，又加入了语义主题信息，充分利用了文档的所有语料。

对于主题信息，采用 LDA 模型计算出与文档相关性最高的主题词，之后预训练得到主题 - 词分布矩阵 $tw = (tw_1, tw_2, \dots, tw_n)$ 和文档 - 主题分布矩阵 $dt = (dt_1, dt_2, \dots, dt_m)$ ， tw_i 是维度为 T 的向量， T 是训练文本语料库的主题数，整个语料库由 m 个文档组成， dt_i 是一个 $m \times T$ 的矩阵，对应于每个文档的主题概率分布。

得到主题矩阵后，为了将 tw 和 dt 矩阵融入模型，在词嵌入阶段，把主题 - 词分布矩阵作为 embedding 层的权重得到主题嵌入向量 $(twe_1, twe_2, \dots, twe_n)$ ，同时，句子中的每个单词 (w_1, w_2, \dots, w_n) 匹配 glove 预训练的词向量 (g_1, g_2, \dots, g_n) ，最终代表每个句子的词嵌入 $we_{1:n}$ 表示为：

$$we_{1:n} = twe_{1:n} \oplus g_{1:n} \quad (1)$$

其中， \oplus 代表 Concatenate 操作，由此，每个输入语句 (w_1, w_2, \dots, w_n) 被表示为 $n \times k$ 的矩阵。同理，将 dt_i 与神经网络生成的文档嵌入进行 Concatenate 操作，之后输入到分类层，得到预测结果。神经主题模型将文档统计信息与神经网络联合起来，充分利用了文档的所有语料。

2.2. 卷积主题网络

由于引入 LDA 分布后模型的参数数量过多，为了更加准确地描述数据特征，本文采用卷积神经网络来降低特征的维度，同时，可以通过迭代卷积和池化操作学习关键的主题信息。

本文的卷积运算是在一维卷积中进行，将 e 个 h 字窗口的卷积滤波器 $W \in R_{h \times k}$ 重复应用于带有主题信息的词嵌入矩阵中产生新的特征。比如，在单词窗口 $we_{i:i+h-1}$ 我们可以得到新特征 M_i ：

$$M_i = f(W * we_{i:i+h-1} + b) \quad (2)$$

其中， f 代表 RELU 激活函数，模型使用激活函数用于输入矩阵， b 是偏置项。对于长度为 n 的句子每个过滤器取到的可能窗口为 $\{we_{1:h}, we_{2:h+1}, \dots, we_{n-h+1:n}\}$ ，由此产生的特征映射为 M ：

$$M = [M_1, M_2, \dots, M_{n-h+1}] \quad (3)$$

卷积层可以帮助模型提取文本的重要特征，明确主题信息。卷积层完成相应的操作后紧接着是池化层，池化的目的是对数据进行降维。本文采用最大池化，对卷积层产生的带有主题信息的特征地图应用最大值池化，来获得池中的最大元素 \hat{M} ：

$$\hat{M} = \text{Max}\{M\} \quad (4)$$

模型在卷积层应用不同大小的多层，来生成多个特征映射，最后池化层生成的元素表示为 M ：

$$M = \hat{M}_1 \oplus \hat{M}_2 \oplus \dots \oplus \hat{M}_m \quad (5)$$

其中， m 表示特征映射的数量。模型在引入卷积神经网络后，不仅减少了参数数量，也学习到了重要的主题信息。

2.3. 分层注意力网络

注意力机制可以通过分配权重来突出上下文重要的信息，新模型 THAHNN 采用分层的注意力网络

来更好地利用文档的结构层次特征，可以关注到对情感分析更有帮助的单词和句子，同时，联合文档 - 主题分布，来加强主题词的权重，使得输出的情感既与序列语义有关，又不偏离文本的主题。

本文根据文档的结构层次分为两个不同的部分，第一部分是词序编码器和词级注意力层，第二部分是句序编码器和句级注意力层。在单词编码器中，使用 BiGRU 来对单词进行编码，利用了前向和后向的上下文信息。对于给定的句子 s_i ，得到加上主题信息的词嵌入 $(we_1, we_2, \dots, we_n)$ ，通过双向的 GRU 层得到前向的 $\overline{h_{wi}}$ 和后向的 $\overleftarrow{h_{wi}}$ ：

$$\overline{h_{wi}} = \overline{GRU}(we_i), i \in [1, n] \quad (6)$$

$$\overleftarrow{h_{wi}} = \overleftarrow{GRU}(we_i), i \in [n, 1] \quad (7)$$

前向的隐藏状态 $\overline{h_{wi}}$ 读句子的顺序是从 we_1 读到 we_n ，而后向的 $\overleftarrow{h_{wi}}$ 读句子的顺序是从 we_n 读到 we_1 ，将前向隐藏状态 $\overline{h_{wi}}$ 和后向隐藏状态 $\overleftarrow{h_{wi}}$ 拼接起来得到 $h_{wi} = (\overline{h_{wi}} : \overleftarrow{h_{wi}})$ ，它包含了整个句子的信息。每个词对句子的重要性是不同的，模型使用注意力机制来量化单词对句子的重要性，具体地说，

$$u_{wi} = \tanh(W_m h_{wi} + b_m) \quad (8)$$

$$\alpha_{wi} = \frac{\exp(u_{wi}^\top u_m)}{\sum_i \exp(u_{wi}^\top u_m)} \quad (9)$$

$$s_h = \sum \alpha_{wi} h_{wi} \quad (10)$$

模型通过全连接层得到 h_{wi} 的隐藏表示 u_{wi} ，计算 u_{wi} 与词级上下文向量 u_m 的相似度来评估单词的重要程度。再通过 Softmax 层得到归一化的关注度权重 α ，之后通过加权得到句子表示 s_h 。单词上下文向量 u_m 在训练过程中被随机初始化并联合学习。

类似地，模型使用 BiGRU 来对句子进行编码，对于给定的句子向量 s_i 和文档表示 D ，得到的句子表示为：

$$\overline{h_{si}} = \overline{GRU}(s_i), i \in [1, h] \quad (11)$$

$$\overleftarrow{h_{si}} = \overleftarrow{GRU}(s_i), i \in [h, 1] \quad (12)$$

将前向隐藏状态 $\overline{h_{si}}$ 和后向隐藏状态 $\overleftarrow{h_{si}}$ 拼接起来得到 $h_{si} = (\overline{h_{si}} : \overleftarrow{h_{si}})$ ， h_{si} 概括了整个文档的上下文句子。然后使用注意力机制关注文档中重要的句子：

$$u_{si} = \tanh(W_n h_{si} + b_n) \quad (13)$$

$$\alpha_{si} = \frac{\exp(u_{si}^\top u_n)}{\sum_i \exp(u_{si}^\top u_n)} \quad (14)$$

$$d = \sum \alpha_{si} h_{si} \quad (15)$$

在上面的等式中， d 是通过注意力加权得到的文档表示向量。为了将主题信息融入神经网络，模型 THAHNN 将预训练的 LDA 主题模型得到的文档 - 主题分布 dt 与 d 相连接，输入到 sigmoid 层得到预测结果。

模型 THAHNN 的损失函数采用 binary_crossentropy，其中 \hat{y}_i 是真实标签， y_i 是输出 \hat{y}_i 标签的概率。

$$\text{loss} = -\sum_{i=1}^n \hat{y}_i \log y_i + (1 - \hat{y}_i) \log(1 - y_i) \quad (16)$$

3. 实验

3.1. 数据集

模型在两个公开的数据集进行实验，使用 80% 的数据进行训练，10% 的数据用于评估模型，10% 的数据用于测试。两个数据集的描述如下：

SemEval-2019 Task 4 [13]：这是一个检测新闻政治偏见的数据集，任务是确定新闻文章的情感倾向是否含有超党派偏见。它包括两个部分，出版社自动分类的语料库和手动标注的文章语料库。手动标注的语料库相对更可靠，本文只使用了文章语料库的数据。它包含 645 篇训练集，测试集的数据不对外开放，所以从中选择一些数据作为评估集和测试集。

Amazon Review [14]：该数据集出自亚马逊商品的好评和差评，它包含了 2065 个商品评论。

表 1 描述了上述数据的详细信息，图 2 展示了 SemEval-2019 Task 4 数据集的文档长度分布。

Table 1. Statistics for the dataset

表 1. 数据集的统计信息

Dataset	classes	documents	positive	negative
SemEval-2019 Task 4	2	645	407	238
Amazon Review	2	2065	1000	1065

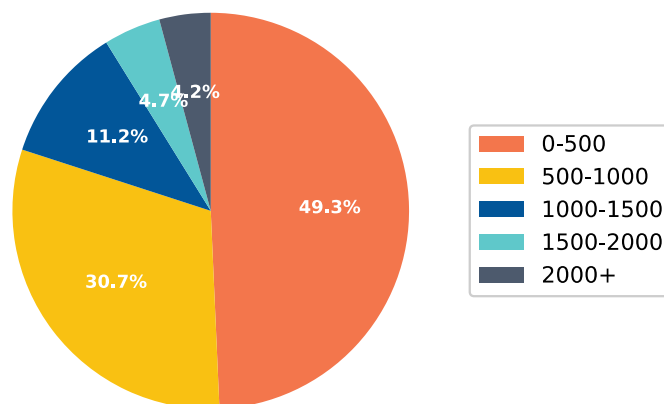


Figure 2. SemEval-2019 Task 4 dataset document length distribution

图 2. SemEval-2019 Task 4 数据集文档长度分布

3.2. 参数设置

由于数据集的文档大小差异较大，为了尽可能地多表达出句子描述的内容，每句话最多选 100 个单词，每篇文档最多 30 个句子。在单词嵌入层中，所有模型都使用公开的预先训练的 Glove 词向量作为权重，词向量的维度是 300 维。LDA 主题模型的主题数量采用一致性模型来确定，在 SemEval-2019 Task 4 和 Amazon Review 评论数据集中，得到最佳的主题数分别为 425 和 625。本文使用 Gensim LDA 模型生成主题词分布和文档主题分布。对于 CNN 层，使用了 32 个滤波器并且窗口大小分别为 4，5，6。对词和句子编码的 BiGRU 的存储维度设置为 50，分类器使用 sigmoid 激活函数。在训练过程中，所有数据集的 batch_size 设置为 64，Dropout 为 0.2，训练 8 个 epoch，采用学习率为 0.001 的 adam 优化器进行反向传播训练，损失函数为 binary_crossentropy，每次在评估集中验证模型的性能，并保存效果最好的模型。

3.3. 基线方法

本研究以下五种深度神经网络模型为基准进行情感极性分类，取得了良好的效果。五种深度神经网络模型如下所示：

CNN [6]：使用一维卷积来提取文本重要特征。

LSTM [15]：长短期记忆网络，可以学习到文本序列的长期依赖信息。

HAN [8]：利用文档的结构特征，分别对词和句子进行编码来施加注意力机制，可以关注到文档中对分类起着重要作用的单词和句子 s 。

Transformer [16]：是一个 `sequencetosequence` 的模型，使用自我注意力机制，同时采用位置嵌入来表示输入序列的顺序，使用多头自我注意力机制来理解词在上下文的含义，在很多 NLP 任务中有着不错的表现。

AC-BiLSTM [17]：使用卷积层从词向量中提取更高级别的短语表示，基于注意力的双向 LSTM 模型来同时考虑前向和后向的输入，从而能发现句子中重要的语义层面的信息。

4. 结果讨论

实验采用精确度(Pr)、召回率(Re)、F1 度量(F1)和准确度(ACC)等四个指标来评估模型，这四个指标被广泛用于情感分析任务中，它们的计算方法如下：

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (18)$$

$$\text{Rec} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (19)$$

$$\text{F1} = \frac{2 * \text{Pre} * \text{Rec}}{\text{Pre} + \text{Rec}} \quad (20)$$

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (21)$$

其中，TP、FP、TN 和 FN 分别代表真正类、假正类、真负类和假负类。文中将 THAHNN 与上述 5 种情感分析基线模型进行了比较。

Table 2. Comparison of results under the SemEval 2019 Task 4 dataset

表 2. SemEval 2019 Task 4 数据集下的结果比较

Method	SemEval 2019 Task 4				
	Class	Accuracy	Precision	Recall	F1
LSTM	Pos	67.69	0.7222	0.6192	0.6667
	Neg		0.6359	0.7528	0.6894
CNN	Pos	75.07	0.6549	0.7999	0.7201
	Neg		0.8146	0.7024	0.7543
Transformer	Pos	73.85	0.6285	0.8799	0.7333
	Neg		0.8527	0.6678	0.7490
HAN	Pos	78.96	0.8571	0.7199	0.7826
	Neg		0.7245	0.8762	0.7931
AC-BiLSTM	Pos	79.82	0.8083	0.7454	0.7755
	Neg		0.7968	0.8123	0.8044
THAHNN	Pos	81.53	0.9019	0.7729	0.8324
	Neg		0.7544	0.8825	0.8134

Table 3. Comparison of results under the Amazon Review dataset
表 3. Amazon Review 数据集下的结果比较

Method	Amazon review				
	Class	Accuracy	Precision	Recall	F1
LSTM	Pos	75.16	0.7279	0.8099	0.7667
	Neg		0.8245	0.7128	0.7645
CNN	Pos	79.22	0.8228	0.7806	0.8011
	Neg		0.7514	0.7925	0.7714
Transformer	Pos	79.23	0.7181	0.7899	0.7523
	Neg		0.8227	0.7678	0.7943
HAN	Pos	79.87	0.7747	0.8199	0.7966
	Neg		0.8314	0.7825	0.8062
AC-BiLSTM	Pos	80.87	0.7899	0.8099	0.7997
	Neg		0.8354	0.7915	0.8128
THAHNN	Pos	82.03	0.7962	0.8397	0.8173
	Neg		0.8412	0.8042	0.8222

首先, 在 SemEval 2019 Task 4 数据集下的实验结果如表 2 所示, 新模型 THAHNN 的准确率提高了 1.71%, 对于 F1 衡量标准, 正类的改进为 4.98%, 负类的改进为 0.9%, 正负类差距较大的原因主要是此数据集不平衡, 正类的数据更多。可以注意到, LSTM 和 Transformer 模型的效果相对较差, 由于该数据集的文本长度较大, LSTM 关注不到更重要的句子, 而 Transformer 缺少对序列建模所以表现的不好。基线模型中, AC-BiLSTM 的效果最好, 它同时考虑到了上述两个模型的缺陷。本文提出的 THAHNN 在准确率和 F1 衡量标准方面都要优于其他 5 个模型, 这些改进主要得益于主题模型结合分层结构丰富了特征空间, 同时加入了 CNN 来提取最重要的特征。

在 Amazon Review 数据集下的实验结果如表 3 所示, 新模型 THAHNN 的准确率提高了 1.16%, 对于 F1 衡量标准, 正类的改进为 1.62%, 负类的改进为 0.94%, 从结果可以看出, THAHNN 在准确率和 F1 衡量标准方面同样都要优于其他 5 个模型, 但是, 改进的数量相较于 SemEval 2019 Task 4 数据集要少。主要还是因为评论数据集的文本长度相对要小, 对于短文本来说, THAHNN 不会有显著的改进, 因为模型采用了分层的双向 GRU 层进行编码, 这个设计可以捕获长期的依赖。

5. 结论

本文提出了一种结合主题模型和深度神经网络的文档情感分析模型 THAHNN, 该方法将 LDA 分布和混合神经网络的分层模型结合起来, 以预测文本的情感。通过在词嵌入层和文本表达层都融入主题信息, 使得模型既考虑了文档的结构层次, 又充分利用了语料库的信息。除此之外, 使用卷积神经网络帮助模型学习关键的主题信息, 提取到更重要的局部特征。在文档长度差异较大的两个数据集中, 相较于对比模型, 本文的模型都取得了很好的结果。

基金项目

天津市自然科学基金(19JCYBJC18700)。

参考文献

- [1] Zhang, L., Wang, S. and Liu, B. (2018) Deep Learning for Sentiment Analysis: A Survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **8**, e1253. <https://doi.org/10.1002/widm.1253>
- [2] Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J. and Scholkopf, B. (1998) Support Vector Machines. *IEEE Intelligent Systems and Their Applications*, **13**, 18-28. <https://doi.org/10.1109/5254.708428>
- [3] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*, **3**, 993-1022.
- [4] McCallum, A. and Nigam, K. (1998) A Comparison of Event Models for Naive Bayes Text Classification. *AAAI-98 Workshop on Learning for Text Categorization*, **752**, 41-48.
- [5] 徐菲菲, 芦霄鹏. 结合卷积神经网络和最小门控单元注意力的文本情感分析[J]. 计算机应用与软件, 2020, 37(9): 75-80+125.
- [6] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 25-29 October 2014, 1746-1751. <https://doi.org/10.3115/v1/D14-1181>
- [7] Wen, S. and Li, J. (2018) Recurrent Convolutional Neural Network with Attention for Twitter and Yelp Sentiment Classification: ARC Model for Sentiment Classification. *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence*, Sanya, 21-23 December 2018, 1-7. <https://doi.org/10.1145/3302425.3302468>
- [8] Yang, Z., Yang, D., Dyer, C., et al. (2016) 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, 12-17 June 2016, 1480-1489. <https://doi.org/10.18653/v1/n16-1174>
- [9] Behera, R.K., Jena, M., Rath, S.K. and Misra, S. (2021) Co-LSTM: Convolutional LSTM Model for Sentiment Analysis in Social Big Data. *Information Processing & Management*, **58**, Article ID: 102435. <https://doi.org/10.1016/j.ipm.2020.102435>
- [10] Pasupa, K. and Seneewong Na Ayutthaya, T. (2022) Hybrid Deep Learning Models for Thai Sentiment Analysis. *Cognitive Computation*, **14**, 167-193. <https://doi.org/10.1007/s12559-020-09770-0>
- [11] Basiri, M.E., Nemati, S., Abdar, M., Cambria, E. and Acharya, U.R. (2021) ABCDM: An Attention-Based Bidirectional CNN-RNN Deep Model for Sentiment Analysis. *Future Generation Computer Systems*, **115**, 279-294. <https://doi.org/10.1016/j.future.2020.08.005>
- [12] Narayan, S., Cohen, S.B. and Lapata, M. (2018) Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, 31 October-4 November 2018, 1797-1807. <https://doi.org/10.18653/v1/D18-1206>
- [13] Kiesel, J., Mestre, M., Shukla, R., et al. (2019) SemEval-2019 Task 4: Hyperpartisan News Detection. *Proceedings of the 13th International Workshop on Semantic Evaluation*, Minneapolis, 6-7 June 2019, 829-839. <https://doi.org/10.18653/v1/S19-2145>
- [14] Amazon Review Dataset (2020). <https://gist.github.com/kunalj101/ad1d9c58d338e20d09ff26bcc06c4235>
- [15] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [16] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, 4-9 December 2017.
- [17] Liu, G. and Guo, J. (2019) Bidirectional LSTM with Attention Mechanism and Convolutional Layer for Text Classification. *Neurocomputing*, **337**, 325-338. <https://doi.org/10.1016/j.neucom.2019.01.078>