

基于注意力机制的多通道 CNN 和 BiGRU 的文本情感倾向性分析

程艳¹ 尧磊波¹ 张光河¹ 唐天伟² 项国雄³ 陈豪迈⁴ 冯悦¹ 蔡壮¹

¹(江西师范大学计算机信息工程学院 南昌 330022)

²(江西师范大学管理决策评价研究中心 南昌 330022)

³(江西师范大学新闻与传播学院 南昌 330022)

⁴(豫章师范学院数学与计算机学院 南昌 330022)

(chyan88888@jxnu.edu.cn)

Text Sentiment Orientation Analysis of Multi-Channels CNN and BiGRU Based on Attention Mechanism

Cheng Yan¹, Yao Leibo¹, Zhang Guanghe¹, Tang Tianwei², Xiang Guoxiong³, Chen Haomai⁴, Feng Yue¹, and Cai Zhuang¹

¹(School of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022)

²(Center of Management Decision Evaluation Research, Jiangxi Normal University, Nanchang 330022)

³(School of Journalism and Communication, Jiangxi Normal University, Nanchang 330022)

⁴(School of Mathematics and Computer, Yuzhang Normal University, Nanchang 330022)

Abstract CNN(convolutional neural network) and RNN(recurrent neural network) have been widely used in the field of text sentiment analysis and have achieved good results in recent years. However, there is a problem of contextual dependency between texts, although CNN can extract local features between consecutive words of a sentence, it ignores the contextual semantic information between words. BiGRU(bidirectional gated recurrent unit) network can not only solve the problem of gradient disappearance or gradient explosion in traditional RNN model, but also make up for the shortcomings that CNN can't effectively extract contextual semantic information of long text, while it can't extract local features as well as CNN. Therefore, this paper proposes a MC-AttCNN-AttBiGRU (multi-channels CNN and BiGRU network based on attention mechanism) model. The model can notice the important words for sentiment classification in the sentence. It combines the advantages of CNN to extract local features of text and BiGRU network to extract contextual semantic information of long text, which improves the text feature extraction ability of the model. The experimental results on the Tan Songbo Hotel Review dataset and IMDB dataset show that the proposed model can extract richer text features than other baseline models, and can achieve better classification results than other baseline models.

收稿日期:2019-12-17;修回日期:2020-09-21

基金项目:国家自然科学基金项目(61967011);江西省自然科学基金项目(20202BABL202033);江西省重点研发计划项目(20161BBE50086);

江西省教育厅科技重点项目(GJJ150299);教育厅人文社科重点(重大)项目(JD19056)

This work was supported by the National Natural Science Foundation of China (61967011), the Natural Science Foundation Project of Jiangxi Province (20202BABL202033), the Primary Research and Development Program of Jiangxi Province (20161BBE50086), the Science and Technology Key Project of Education Department of Jiangxi Province (GJJ150299), and the Humanities and Social Sciences Key (Major) Project of the Education Department (JD19056).

Key words CNN (convolutional neural network); text sentiment orientation analysis; BiGRU (bidirectional gated recurrent unit); attention mechanism; multi-channels

摘要 近年来,卷积神经网络(convolutional neural network, CNN)和循环神经网络(recurrent neural network, RNN)已在文本情感分析领域得到广泛应用,并取得了不错的效果。然而,文本之间存在上下文依赖问题,虽然CNN能提取到句子连续词间的局部信息,但是会忽略词语之间上下文语义信息;双向门控循环单元(bidirectional gated recurrent unit, BiGRU)网络不仅能够解决传统RNN模型存在的梯度消失或梯度爆炸问题,而且还能很好地弥补CNN不能有效提取长文本的上下文语义信息的缺陷,但却无法像CNN那样很好地提取句子局部特征。因此提出一种基于注意力机制的多通道CNN和双向门控循环单元(MC-AttCNN-AttBiGRU)的神经网络模型。该模型不仅能够通过注意力机制关注到句子中对情感极性分类重要的词语,而且结合了CNN提取文本局部特征和BiGRU网络提取长文本上下文语义信息的优势,提高了模型的文本特征提取能力。在谭松波酒店评论数据集和IMDB数据集上的实验结果表明:提出的模型相较于其他几种基线模型可以提取到更丰富的文本特征,可以取得比其他基线模型更好的分类效果。

关键词 卷积神经网络;文本情感倾向性分析;双向门控循环单元;注意力机制;多通道

中图法分类号 TP391

情感分析,也称为观点挖掘,是指人们对服务、产品、组织、个人、问题、事件、话题及其属性的情感、观点、评价、态度和情绪^[1]。在自然语言处理领域,情感倾向性分析是一项基础性任务,近年来已经吸引了国内外学者的广泛关注。随着互联网和移动网络的快速发展,用户在网上发表的观点和评论也越来越多,从这些内容中挖掘出重要的情感信息也值得探索。情感分析主要分为主客观情感识别和情感分类这2类,主客观情感识别是将给定文本分成主观性或客观性文本,情感分类则是把主观性文本分类为正面、负面或者中性^[2]。情感分析方法主要有基于机器学习和基于情感词典的这2类方法,随着神经网络研究的再次复苏,基于深度学习的方法被广泛应用到文本情感分析任务中。基于情感词典的方法最先被用于情感分析领域,它首先需要构建一个情感词典,然后对情感词典进行手工极性和强度标注,最后实现对文本的情感分类。这种方法虽然能够对文本进行情感分类,但是分类的效率并不高,因为它需要人工构建情感词典并进行手工标注。20世纪90年代,机器学习方法开始在文本情感分析中崭露头角^[3-4]。机器学习方法模型的函数虽然简单,但通常需要复杂的特征工程,而特征的好坏能够直接影响后面的分类效果,此外,机器学习方法的泛化能力比较低。近年来兴起的深度学习方法很好地弥补了基于机器学习和基于情感词典2种方法的缺陷,并被成功地应用于自然语言处理任务中,尤其是被应用

在情感分析任务中,如卷积神经网络^[5-7](convolution neural network, CNN)、长短时记忆(long-short term memory, LSTM)网络^[8-9]、循环神经网络(recurrent neural network, RNN)和CNN是目前深度学习领域比较热门的2种模型。但这2种模型也都存在一定的缺陷,CNN不能考虑到上下文语义信息,从而忽视了句子之间的相互联系;RNN模型缺乏对文本特征提取的能力,不能很好地提取句子的局部特征。针对2种模型存在的缺陷问题,本文利用CNN、双向门控循环单元(bidirectional gated recurrent unit, BiGRU)网络模型和注意力机制,结合word2vec词向量^[10]表示模型,提出一种基于注意力机制的多通道CNN和BiGRU网络的模型来解决文本情感分类问题。模型使用3个CNN进行信息抽取,利用不同大小的卷积窗口进行粒度不同的情感特征提取,从而可以提取到更丰富的文本情感特征。BiGRU网络能够考虑文本上下文信息和保持句子间单词时序,因此能够更加充分地利用整个文本中的特征信息,有效识别文本情感极性。同时在CNN通道和BiGRU网络通道都引入注意力机制,能够更加关注句子中对分类重要的单词信息,以此来进一步提升分类准确率。

本文的主要贡献有3个方面:

- 1) 利用双向GRU(BiGRU)代替单向GRU和传统的RNN模型,使模型既能够考虑到文本上下文语义信息,又能够解决梯度消失问题。

2) 提出一种结合 CNN 和 BiGRU 的网络模型。该模型既能够利用不同大小的卷积核在 CNN 不同通道内提取文本中粒度不同的特征信息,又能够使用 BiGRU 网络充分考虑到句子上下文语义信息,弥补了各自的缺陷,增强了模型的文本特征提取能力,提高了分类准确率。

3) 将注意力机制引入到本文模型中,它能够提取影响情感极性分类的关键文本特征,使模型能够关注到句子中对文本分类贡献大的词语,以此来提升 CNN 和 BiGRU 网络的特征提取效果。

1 相关工作

情感分析是研究人们对实体的看法、态度和情感的。情感倾向性分析一直都是自然语言处理领域的研究热点。早些年,主要使用基于情感词典和传统的机器学习的方法进行情感倾向性分析。基于情感词典的方法是使用情感词典中已经人工标注好的词条和句子中分好词的每个词语进行匹配,再用相关公式对每个句子进行计算并打分,最后整个句子得分大于 0 的分为正类,小于 0 的分为负类,以此进行文本情感倾向性分析。对于英文词典,Kamps 等人^[11]使用 WordNet 英文情感词典判断英文文本的情感倾向性。与英文情感词典相比,规范的中文情感词典相对缺乏,最早也是最普遍使用的是由董振东等人^[12]标注的大型情感分析词语库 HowNet。基于情感词典进行分类的方法虽然操作简单,但是情感词典的构建耗时耗力,且扩充范围有限,具有领域局限性。而传统的基于机器学习的情感分类方法通常被看作是一类有监督学习问题。Pang 等人^[13]在 2004 年利用最大熵、朴素贝叶斯和支持向量机(support vector machine, SVM)在文本情感分析中进行了实验比较,发现使用 SVM 进行文本情感分类能够达到最优效果。Lee 等人^[14]将最大熵分类应用于给定电子产品评论的极性评估。这些机器学习方法虽然能够进行文本情感分类,但是在分类的过程中需要复杂的特征选择,这个过程同样需要人工设计,同样会导致其扩展性差,很难适应不同的领域和应用需求。

近些年来,深度学习方法逐渐成为了情感分析的主流方法。它不需要人工进行干预,是一种端到端的方法,可以实现对特征进行自动选择,因此效率会比基于词典和机器学习的方法更高,且领域适用性更强。Collobert 等人^[15]于 2011 年首先提出使用 CNN

解决词性标注等自然语言处理(natural language processing, NLP)领域的问题。在 2014 年, Kim^[6]提出将 CNN 应用于英文文本情感分类任务中,并取得了当时不错的分类结果。Kalchbrenner 等人^[5]提出了一种宽卷积模型,并且用 k -max 池化代替了传统 CNN 的最大池化,这样做可以保留更多的特征信息。这种模型不需要任何的先验信息输入,也不需要构造非常复杂的人工特征。Yin 等人^[16]使用多通道不同尺寸的卷积神经网络进行句子分类。陈珂等人^[17]提出一种多通道卷积神经网络模型,使用多个卷积神经网络提取句子的多方面特征,在中文微博情感分析任务上取得了不错的效果。然而,基于 CNN 的文本情感分类存在不能考虑句子上下文语义信息问题。

相较于 CNN, RNN 引入了记忆单元使得网络具备了一定的记忆性,能够捕捉到文本间的长距离依赖关系。但是它也存在一定的弊端,即在训练的过程中会出现梯度消失问题。而 LSTM 和门控循环单元(gated recurrent unit, GRU)网络在传统的 RNN 基础上引入了门机制,较好地克服了 RNN 中梯度消失的弊端。基于 LSTM 和 GRU 模型,很多人在它们基础上进行改进,并取得了不错的分类效果。Tang 等人^[18]首先使用 CNN 或 LSTM 实现单句表示,然后再用 gated RNN 编码句子间的内在关系和语义联系,最后再实现篇章级文本的编码,通过此方法可以较好地捕获句子间的语义信息。Lai 等人^[19]提出了一种 RCNN 模型,首先利用双向循环网络模型得到单词的上下文信息,然后再通过 CNN 的卷积池化过程进行文本分类,在 ZONews, Fudan, ACL, SST 这 4 个测试集上取得了最好的效果。Zhou 等人^[20]提出了一种 C-LSTM 模型,首先利用卷积神经网络提取文本特征,然后再用 LSTM 网络代替最大池化层得到最终分类结果。Ruder 等人^[21]使用一种层次化的双向 LSTM 模型进行方面级别的文本情感分类。Rao 等人^[22]使用 2 层 LSTM 进行文档级别的情感分类。Sachin 等人^[23]使用 LSTM、GRU、双向 GRU 和双向 LSTM 方法进行亚马逊评论情感分析,并取得了不错的效果。虽然 LSTM 和 GRU 模型能够提取到上下文语义信息和考虑到单词间时序问题,但是它们并不能像 CNN 那样很好地提取句子的局部特征。

为了能够充分发挥 CNN 和 RNN 的各自优势,越来越多的研究者将 CNN 与 RNN 进行结合以此来进行文本情感倾向性分析。Wang 等人^[24]使用

一层 CNN 和一层 RNN 构成一个融合模型, 进行短文本情感分析, 实验证明效果优于单纯使用 CNN 和 RNN 模型. Zhou 等人^[25]提出一种结合双向 LSTM 和 CNN 的模型, 并在池化层使用 2 维最大池化代替传统的最大池化得到最终分类结果. Zhang 等人^[26]提出一种 CNN-LSTM 模型, 用于预测 Twitter 文本的情感强度. Zhang 等人^[27]提出一种多通道 CNN-LSTM 模型, 用于 Twitter 文本情感分类任务. Alayba 等人^[28]提出一种结合 CNN 和 LSTM 模型, 用于阿拉伯语情感分析, 并取得了不错的分类效果. Zhang 等人^[29]使用基于 Convolution-GRU 模型对 Twitter 仇恨评论文本进行情感极性判别.

近几年来, 注意力机制被广泛应用于文本分类任务中, 因为注意力机制能够让模型从大量信息中有选择性地筛选出重要信息, 同时忽略一些不重要的信息, 能够让模型注意到句子中对情感分类重要的词语, 例如 Yang 等人^[30]结合了双向 RNN 与注意力机制, 构建了注意力模型应用在篇章级文本分类任务中, 取得了当时文本分类任务的最佳结果; Wang 等人^[31]将多层注意力机制和 CNN 结合用于句子关系分类任务中, 模型在多个数据集上的实验结果表明, 使用注意力机制的模型比未使用的模型有更高的分类精度. Yang 等人^[32]提出 2 种不同的基于注意力机制的双向 LSTM 模型用于目标相关的情感分类, 在当时达到了最好的分类效果. Wu 等人^[33]使用基于注意力机制的 CNN-LSTM 网络模型进行 Twitter 文本情感强度分析, 并取得了不错的效果. 程艳等人^[34]提出一种融合卷积神经网络与层次化注意力网络的模型, 先利用 CNN 学习词语特征, 然后再将其输入到基于注意力机制的双向 GRU 网络中得到句子的情感倾向. 吴小华等人^[35]提出一种结合双向 LSTM 和自注意力机制的模型, 先使用双向 LSTM 得到文本的上下文语义特征, 然后通过自注意力机制自动调整特征权重, 在短文本情感分类任务上取得了不错的效果. 袁和金等人^[36]提出一种提出了一种基于多通道卷积与双向 GRU 网络的情感分析模型, 并在双向 GRU 网络上引入注意力机制来自动关注对情感极性影响较强的特征. 鉴于融合注意力机制的神经网络模型的出色表现, 本文在对文本情感倾向性分析任务中也引入了注意力机制, 使网络模型能够更多地注意到对文本情感极性贡献大的词语.

基于 CNN, RNN 和注意力机制各自的优点, 本文提出一种基于注意力机制的多通道 CNN 和

BiGRU 网络模型, 该模型利用 3 通道 CNN 提取文本不同粒度的局部特征信息, 并且使用 BiGRU 网络模型代替传统的 RNN 模型, 提取文本上下文语义信息, 然后在每一个通道上都引入注意力机制, 最后将它们进行拼接, 得到一种新的基于注意力机制的多通道神经网络模型. 本文综合利用 CNN, BiGRU 网络和注意力机制这 3 种深度学习方法的优点, 既考虑了文本的局部信息又考虑到了上下文语义信息, 避免了文本信息的丢失, 还能通过注意力机制关注到句子中对情感分类贡献较大的词语, 提高了模型对文本情感分类性能.

2 模型

2.1 卷积神经网络

CNN 最早被应用在计算机视觉领域中, 近些年被广泛应用到自然语言处理任务中, 并取得了不错的效果^[6,37]. 常见的 CNN 模型主要由卷积层、池化层和全连接层构成, 卷积层是用来提取输入数据的特征, 而池化层是对卷积层提取的特征进行选择和过滤, 全连接层等价于传统前馈神经网络的隐藏层, 一般都是在最后和输出层进行连接, 实现最后的输出. 在本文模型中, 我们将输入句子用一个矩阵 $S \in \mathbb{R}^{n \times d}$ 进行表示, 其中 n 表示的是句子中词的个数, d 表示的是每个词的词嵌入向量维度. 假设卷积核 $W \in \mathbb{R}^{h \times d}$, 其中 d 表示的是卷积核的长度, 它的大小和词向量的维度相同, h 表示的是卷积核的宽度. 对于输入矩阵 $S \in \mathbb{R}^{n \times d}$, 通过重复应用卷积核 W 进行卷积操作得到特征映射向量 $\mathbf{O} = (o_0, o_1, \dots, o_{n-h}) \in \mathbb{R}^{n-h+1}$, 其中特征向量 \mathbf{O} 中元素的计算公式如式(1)所示:

$$o_i = W \cdot S_{i:i+h-1}, \quad (1)$$

其中, $i = 0, 1, 2, \dots, n-h$, (\cdot) 表示矩阵的点乘操作, $S_{i:j}$ 表示的是矩阵 S 从 i 行到 j 行的子矩阵, 即第 i 个单词到第 j 个单词的词向量矩阵. 进行完卷积操作之后, 得到的每个特征映射向量 \mathbf{O} 都被送到池化层以生成潜在的特征和进行特征的过滤. 常见的池化策略是最大池化(max-pooling), 其作用是捕获卷积之后的最重要特征 v , 这种策略能够很自然地把一个变长句子处理为固定长度: $v = \max_{0 \leq i \leq n-h} \{o_i\}$.

我们在 2.1 节已经详细描述了用一个过滤器提取一个特征的过程. 图 1 中模型利用了几个不同大小的过滤器(具有不同窗口大小)来提取不同的特征, 接下来把每一个过滤器提取的特征进行 max-

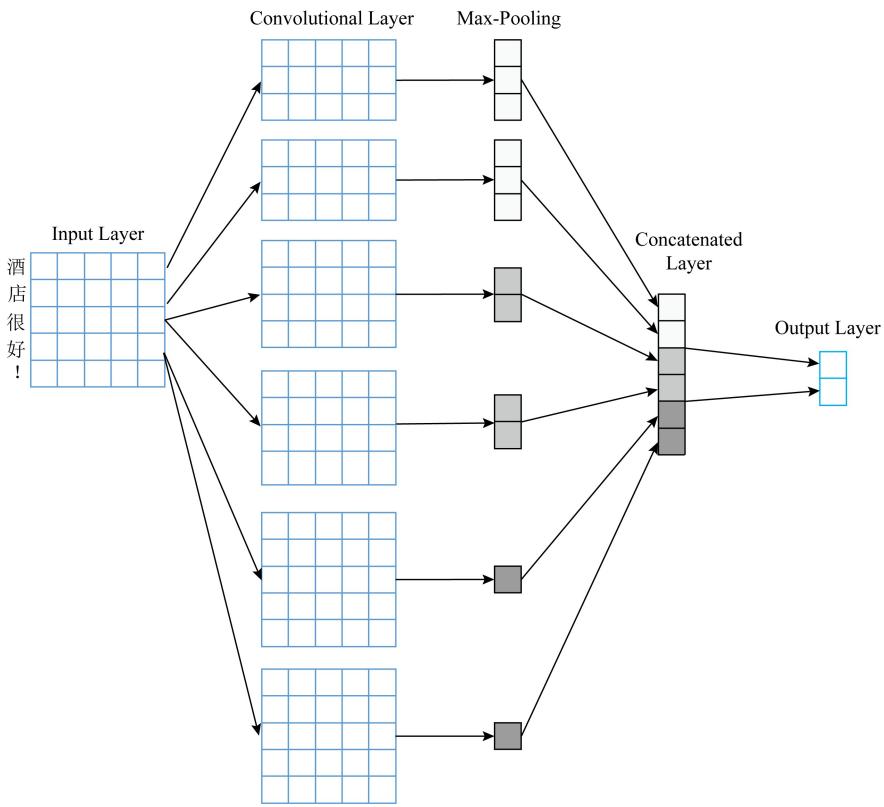


Fig. 1 The structure of the CNN

图 1 CNN 结构

pooling 操作,以此来得到多个不同的最重要特征,然后把这些特征进行拼接,最后再传递给全连接的 Softmax 层,其输出是标签上的概率分布.

2.2 门控循环单元络和双向门控循环单元网络

Cho 等人^[38]在 2014 年最初提出了一种 GRU 网络模型,以使每个循环单元能够自适应地捕获不同时间尺度的依赖性,有关 GRU 的结构说明如图 2 所示:

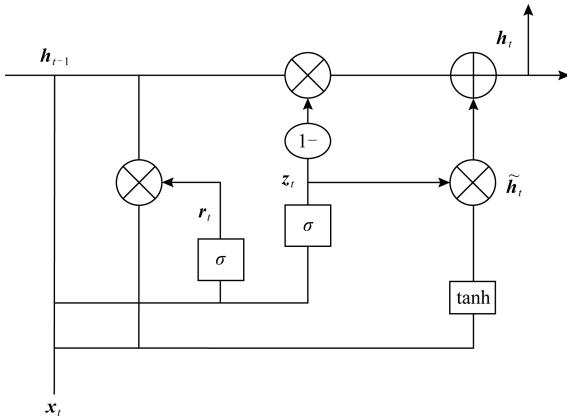


Fig. 2 The structure of the GRU

图 2 GRU 结构

相比于 LSTM,GRU 的模型更为简单,仅由更新门 z 和重置门 r 构成,比 LSTM 少了一个门,所以在训练的时候参数更少,收敛时间更快.可以通过式(2)~(5)更新参数:

$$r_t = \sigma(W_r x_t + U_r h_{t-1}), \quad (2)$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1}), \quad (3)$$

$$\tilde{h}_t = \tanh(Wx_t + Uh_{t-1}), \quad (4)$$

$$h_t = (1 - z_t)h_{t-1} + z_t \tilde{h}_t, \quad (5)$$

其中, W_r, W_z, W, U_r, U_z, U 都是 GRU 的权值矩阵, σ 表示逻辑 sigmoid 函数, \odot 表示元素相乘操作, z_t 表示更新门,它能够决定 GRU 单元激活值的更新程度,是由当前输入状态和上一隐藏层的状态共同决定, r_t 表示的是重置门,它的更新过程和 z_t 过程类似, \tilde{h}_t 表示候选隐藏层, h_t 表示隐藏层.GRU 其实就是 LSTM 的一种变体,它主要是把 LSTM 中的输入门和遗忘门合并成了一个单一的更新门,减少了模型的训练参数与模型收敛时间,降低了训练复杂度,是当前比较流行的循环神经网络模型之一.

单向 GRU 的状态是从前往后单向传输的,也就是忽略了后文单词对整体状态的影响;BiGRU 是

GRU 的一个变体,它的输出取决于前向状态和后向状态的双重影响,解决了单向 GRU 存在的问题,使

得最后的输出结果更加准确.BiGRU 的模型结构如图 3 所示:

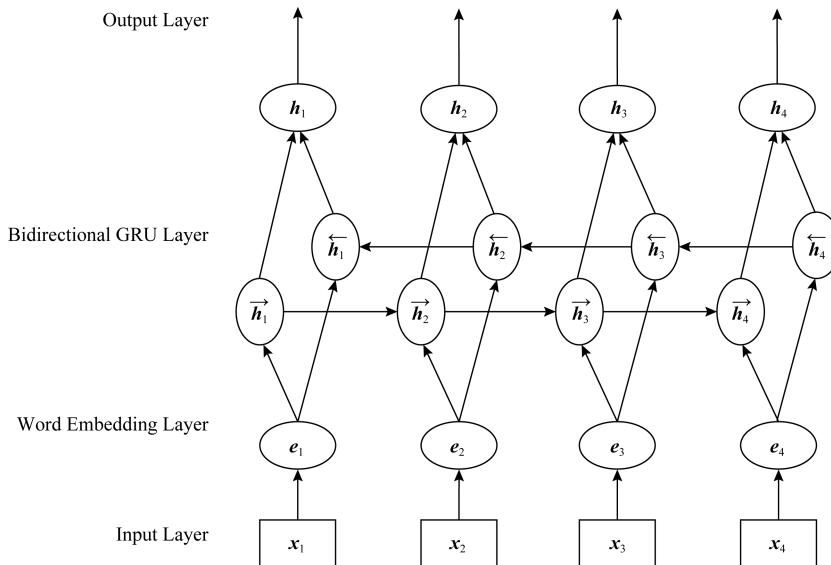


Fig. 3 The structure of the BiGRU

图 3 双向 GRU 结构

在图 3 中, x_i 表示输入的句子单词, e_i 表示词向量, \vec{h}_i 表示 e_i 经过 GRU 层输出的正向的隐藏状态, \overleftarrow{h}_i 表示 e_i 经过 GRU 层输出的反向的隐藏状态, h_i 表示正向 \vec{h}_i 和反向 \overleftarrow{h}_i 拼接后的隐藏状态.

2.3 注意力机制

注意力机制早在 2014 年便被应用在机器翻译任务上^[39],并取得了当时最好的效果. 2016 年, Yang 等人^[30]把注意力机制应用于文本篇章级别的情感分类任务中,取得了不错的分类效果. 因为对于每个句子而言,每个词的重要程度不同,所以本文引入注意力机制来提取句子中重要单词的语义信息. 注意力机制可以抽象为由 Encoder 和 Decoder 这 2 个模块组成. Encoder 一般为编码器,对输入数据

做一定的变换得到语义向量; Decoder 一般为解码器,同样经过一定的变换后得到输出数据. 注意力机制结构如图 4 所示.

主要公式如式(6)~(8)所示:

$$\mathbf{u}_i = \tanh(\mathbf{W}_i \mathbf{h}_i + \mathbf{b}_i), \quad (6)$$

$$\alpha_i = \frac{\exp(\mathbf{u}_i^T \mathbf{u}_w)}{\sum_{i=0}^L \exp(\mathbf{u}_i^T \mathbf{u}_w)}, \quad (7)$$

$$\mathbf{h}_i = \sum_{i=0}^L \alpha_i \mathbf{h}_i, \quad (8)$$

其中, \mathbf{u}_w 为随机初始化的上下文向量, 在训练的过程中更新; \mathbf{u}_i 为隐藏层向量 \mathbf{h}_i 经过一个全连接运算得到的结果; \mathbf{W}_i 和 \mathbf{b}_i 分别为注意力计算的权值矩阵和偏置项; α_i 为句子中第 i 个单词的注意力分数, L 为句子的长度.

2.4 模型整体结构

本文在上述基础模型上提出一种基于注意力机制的多通道 CNN 和 BiGRU 网络的文本情感分析模型. 该模型由卷积神经网络、双向门控循环网络和注意力机制组成, 模型结构如图 5 所示.

本文的总体模型是由多个通道组成, 主体是 3 个拼接的 CNN 模型通道和一个 BiGRU 模型通道, 3 个 CNN 通道主要是用来提取句子间词语的不同的局部特征, 而 BiGRU 通道是用来得到句子上下文语义信息.CNN 通道的输入是每个句子经过分词

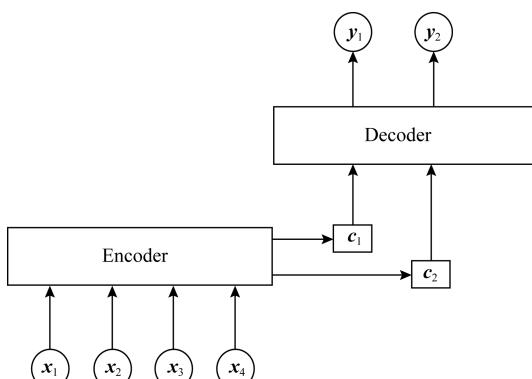


Fig. 4 The structure of attention mechanism

图 4 注意力机制结构

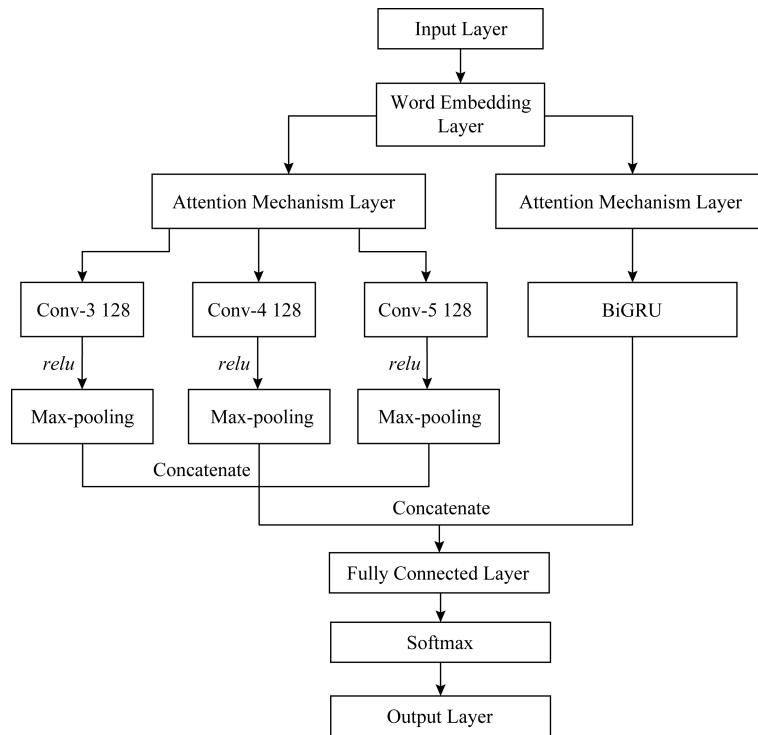


Fig. 5 The overall structure of model

图 5 模型整体结构

操作得到的词语。第 1 层是嵌入层,作用是把输入的每个词语映射成一个向量表示;第 2 层是注意力机制层,主要是为了提取句子间重要单词信息;第 3 层是卷积层,主要用来提取词语间的局部特征,本文词向量维度选取 300 维,选择的 3 种过滤器大小分别为 $3 \times 300, 4 \times 300, 5 \times 300$,每一种过滤器各 128 个,步长 $stride=1, padding=valid$ (不需要进行补零操作),经过卷积操作之后就能得到句子的局部特征;第 4 层为池化层,主要是把卷积层得到的局部特征进行最大池化操作,提取句子间最重要的特征,丢弃部分不相关和无用特征,以此生成固定维度的特征向量,再将 3 个经过最大池化操作输出的特征进行拼接,作为全连接层输入的一部分,经过 3 种不同过滤器得到的 3 种不同的特征经过拼接后可以为后面句子情感分类提供更丰富的特征信息。

BiGRU 通道的第 1 层亦为词嵌入层,词向量的维度也设置为 300 维,第 2 层同样是一个注意力机制层,和用于 CNN 模型的注意力机制一样,均用来提取句子中重要的单词信息;BiGRU 的第 3 层和第 4 层分别为前向 GRU 和后向 GRU 结构,它们的隐藏层大小都设置为 128,因为当前时刻的输入单词和前后序列单词都有一定的关系,所以将输入序列分别从前后 2 个方向输入 GRU 模型中,如图 3 所

示,然后通过隐藏层来保存前后方向文本的信息,最后将这 2 个隐藏层的输出进行拼接,得到 BiGRU 最后输出,BiGRU 输出代码如下:
 $BiGru_output = keras.layers.Bidirectional(gru(hidden_dim, return_sequences=True))(inputs)$.

利用 BiGRU 模型主要是用来提取句子中词的上下文语义信息,以此来提取文本中词语的全局特征,然后我们将 CNN 通道中经过池化操作拼接起来的特征和 BiGRU 的输出进行融合,融合的代码为:
 $output = keras.layers.merge.concatenate([cnn_output, BiGru_output])$.

将融合的特征作为全连接层的输入,全连接层中的激活函数都是使用 relu,在全连接层之后加入 dropout 机制,每次训练的时候都让一部分神经元不工作,这样做的目的是为了防止过拟合,最后再把它输入到 Softmax 分类器中得到最后的分类结果,本文 Softmax 分类器将 x 分类为类别 j 的概率如式(9)所示:

$$P(y^{(i)} = j | x^{(i)}; \theta) = \frac{\exp(\theta_j^T x^{(i)})}{\sum_{n=1}^k \exp(\theta_n^T x^{(i)})}, \quad (9)$$

其中, θ 为训练中的所有参数, $k=2$, 本文均为二分类任务。

3 实验分析

3.1 实验环境

本文实验环境如下:操作系统为 Windows10, CPU 是 Intel Core i7-6700U, GPU 为 GeForce GTX 1080, 内存大小为 DDR3 8 GB, 开发环境为 Keras 2.1.0, 开发工具使用的是 PyCharm.

3.2 实验数据集

本文实验部分在 2 个公开数据集上进行:

1) 谭松波酒店评论数据集. 用户在国内不同酒店上的评论数据集, 然后经过谭松波教授的整理, 用于情感倾向性分析.

2) IMDB^[40] 数据集. 国外用户在网站上对电影

的评论数据, 评论包括正向和负向的情感, 用于情感倾向性分析.

数据集中的每条评论已经被人为地设置了情感标签, 这 2 个数据集具体信息如表 1 所示, 部分评论数据如表 2 所示, 为了减少训练过程中带来的随机性影响, 我们分别在这 2 个数据集上进行 10 折交叉验证实验. 本文句子的长度我们用词语个数来衡量.

Table1 Specific Information of Two Datasets

表 1 2 个数据集具体信息

Datasets	Category	Number	Average Length	Class
Tan Songbo Hotel Review	Train	3 000	45	2
	Test	3 000		
IMDB	Train	25 000	240	2
	Test	25 000		

Table 2 Part of Review Data

表 2 部分评论数据

Datasets	Positive	Negative
Tan Songbo Hotel Review	很满意的一家酒店, 装修不错, 那天住下晚上房间的网络出了问题, 负责网络的人又下班了, 服务员很热心的带我到楼下商务中心免费上网. 送机服务也很好, 总之很满意, 以后去昆明就住经贸了!	住过这么多酒店, 从来没有住过这么小旧脏老臭的“大床房”! 一进门就被一股臭味熏了出来, 开了一天的窗户也不见减少多少. 洗手间又小又脏, 恶心死了! 我和老公从进门开始, 就只能一个在床上, 一个才有站在地上的空间
IMDB	Bromwell High is a cartoon comedy. It ran at the same time as some other programs about school life, such as “Teachers”...	Story of a man who has unnatural feelings for a pig. Starts out with a opening scene that is a terrific example of absurd comedy...

3.3 实验预处理与模型超参数设置

词向量在自然语言处理中起着至关重要的作用, 结合词向量进行预训练有助于提高模型的准确率^[41]. 对于中文数据集, 本文使用大规模中文 Wikipedia 百科数据集来预训练 CBOW 模型, 以此来得到中文词向量, 经过训练之后, 每个语义相近的中文词语被映射到向量空间中相近的位置上, 在训练过程中, 我们设置词向量的维度大小为 300. 首先用 Jieba 分词工具对输入的句子进行分词操作, 之后去除停用词, 然后使用词向量训练工具把这些词语转换为向量形式后作为神经网络的输入. 对于英文数据集, 本文使用大规模英文 Wikipedia 数据集来训练 CBOW 模型, 以此来得到英文词向量, 在训练过程中我们同样设置词向量维度为 300. 然后通过空格把每个句子分成一个个单词, 再去除停用词, 本文中我们把句子最大长度设置为 200, 小于该长度时, 对句子进行补零操作, 大于该值时进行截断操作. 本文实验基于 Keras 深度学习框架, 模型的优化函数采用 Adam^[42] 函数, 因为它能为不同的参数设计独立的自适应性学习率, 并且能够加快网络收敛速度.

在本文实验中, 为了得到更加丰富的情感特征信息, 将多通道 CNN 所选取的卷积核窗口大小分别为 3, 4, 5, 每个卷积核的数量都为 128. 为了防止过拟合现象, 我们使用了 L2 正则化和 dropout 机制. 本文模型的详细超参数设置如表 3 所示:

Table 3 Hyper Parameters of Experiment

表 3 模型超参数设置

Hyper Parameters Description	Value
Size of Word Vector	300
Window Size	(3, 4, 5)
Number of Features Map	128
Hidden Size of BiGRU	128
Mini-batch Size	100
Epochs	50
Learning Rate of Adam	0.001
Size of L2 Regularization	0.001
Dropout Rate	0.5

3.4 实验对比

为了验证本文提出的模型的分类性能, 我们使用准确率作为评估指标, 将本文基于注意力机制的

多通道 CNN 和 BiGRU 模型(MC-AttCNN-AttBiGRU)分别与 13 种模型方法在 2 个数据集上进行实验对比。

1) SVM.相较于其他机器学习方法,SVM 在情感分类上面的效果较好^[13],本文先把句子中的每个单词用词向量表示,然后再把这些单词向量进行加权平均作为 SVM 的输入进行分类。

2) Fasttext^[43].Fasttext 是 Facebook 开源的文本分类工具.在本实验中,将模型学习率设为 0.01,词向量的维度设置为 300.

3) CNN^[6].普通的单通道 CNN 分类,把句子中每个词语经过词嵌入之后得到的词向量作为 CNN 的输入,然后经过卷积层、池化层、全连接层和最后的 Softmax 输出层进行分类。

4) MC-CNN.这是卷积核窗口大小分别为 3,4,5 的 3 个卷积层通道的拼接,即 3 通道 CNN 情感分类模型,过程和单个 CNN 分类一样,只是最后把 3 个经过池化层得到的最大特征进行了拼接,以此来获得更丰富、不同粒度的特征信息。

5) RCNN^[19].这是一个单通道的 CNN-RNN 模型,该模型使用 CNN 和 RNN 进行文本分类,在训练的过程中对输入词向量进行微调。

6) C-LSTM^[20].这是一个单通道 CNN-LSTM 模型,首先把句子中每个词语经过词嵌入之后得到的词向量作为 CNN 的输入,然后把抽取到的特征作为 LSTM 的输入,最后再经过全连接和 Softmax 层得到最终分类结果。

7) Convolution-GRU^[29].这是一种结合 CNN 和 GRU 的复合模型,首先把句子中每个词语经过词嵌入之后得到的词向量作为 CNN 的输入,然后把抽取到的特征作为 GRU 的输入,最后再经过全连接和 Softmax 层得到最终分类结果。

8) MC-CNN-LSTM^[27].这是一种多通道的 CNN-LSTM 模型,该模型由 2 部分组成:一部分是多通道 CNN 模型;另一部分是 LSTM 模型.首先使用多通道 CNN 抽取不同的 n-gram 特征,然后把得到的特征作为 LSTM 的输入,最后再经过全连接和 Softmax 层得到最终分类结果。

9) BiGRU.使用经过词嵌入之后的单词向量作为输入经过 BiGRU 网络模型进行句子情感分类。

10) ATT-MCNN-BGRUM^[36].该模型是一种单通道的 CNN-BiGRU 模型,首先使用多通道的 CNN 模型提取到文本的不同的 n-gram 特征,然后将其输入到基于注意力机制的 BiGRU 模型中,最

后使用 Maxout 神经元得到最后的分类结果。

11) MC-CNN-BiGRU.这是 3 个 CNN 通道和 BiGRU 通道的拼接,3 个 CNN 通道和 BiGRU 通道的输入都是经过词嵌入之后单词的词向量,但 3 个 CNN 通道和 BiGRU 通道都不加入注意力机制。

12) MC-AttCNN-BiGRU.这是 3 个 CNN 通道和 BiGRU 通道的拼接,并且在 3 个 CNN 通道上都加入注意力机制,而 BiGRU 通道不加入注意力机制,其他参数设置和 MC-CNN-BiGRU 一致。

13) MC-CNN-AttBiGRU.这是 3 个 CNN 通道和 BiGRU 通道的拼接,并且在 BiGRU 通道上加入注意力机制,而 3 个 CNN 通道不加入注意机制,其他参数设置和 MC-CNN-BiGRU 一致。

14) MC-AttCNN-AttBiGRU.这是本文提出的最终模型,3 个 CNN 通道和 BiGRU 通道的拼接,并且在 BiGRU 通道和 3 个 CNN 通道上都加入注意力机制,以经过词嵌入之后得到的词向量作为输入,分别通过 CNN 通道和 BiGRU 通道提取特征,最后将提取到的 2 种特征融合进行最后的情感分类。

3.5 实验结果分析

本文用所提出的 MC-AttCNN-AttBiGRU 与 3.4 节其他 13 种神经网络模型方法在谭松波酒店评论数据集和 IMDB 上进行实验对比,实验结果如表 4 所示:

Table 4 Accuracy Comparison of the Proposed Model and Other Baseline Models on Two Public Datasets

表 4 本文模型和其他基线模型在 2 种公开数据集上的准确率对比

Models	Tan Songbo Hotel Review Dataset	IMDB Dataset	%
SVM	80.16	79.86	
Fasttext	83.96	81.69	
CNN	89.88	88.84	
MC-CNN	90.38	89.38	
RCNN	90.09	89.20	
C-LSTM	90.29	89.13	
Convolution-GRU	90.42	89.25	
MC-CNN-LSTM	90.78	89.74	
BiGRU	89.50	89.00	
ATT-MCNN-BGRUM	91.22	90.80	
MC-CNN-BiGRU	91.01	90.24	
MC-AttCNN-BiGRU	91.58	91.09	
MC-CNN-AttBiGRU	91.42	90.82	
MC-AttCNN-AttBiGRU (Proposed Model)	92.28	91.70	

从表 4 可以看出,本文提出的 MC-AttCNN-AttBiGRU 模型方法在 2 个数据集上都取得了比其他模型更好的分类效果,在谭松波酒店评论数据集上准确率达到了 92.28%,在 IMDB 数据集上准确率达到了 91.70%,说明了本文所提模型的优越性。

由表 4 可知,基于神经网络模型方法(CNN, RCNN 等)在分类效果上还是显著优于传统机器学习模型方法 SVM 的,因为 SVM 模型只是简单地对句子中所有单词向量进行加权平均,并没有考虑到句子间的上下文语义和一些更深层次的信息,所以分类效果不好,基于神经网络模型方法准确率大概比基于传统机器学习方法 SVM 高 10% 左右,说明了深度学习方法在文本情感分类任务上更有效果;此外,Fasttext 模型分类准确率(83.96%,81.69%)优于 SVM 模型(80.16%,79.86%),证明 Fasttext 模型在文本情感分析中的优良性能,因为其模型简单,运算速度快,因此本文把它作为基线模型;在谭松波酒店评论数据集上,CNN 和 BiGRU 的分类准确率相近,因为它们都能提取到一定的语义信息,但 CNN 的准确率要比 BiGRU 的稍微高一点,这是因为谭松波酒店评论数据集较小并且评论文本较短,CNN 可以很好地提取句子整体特征,而 BiGRU 却无法在短文本上充分发挥其优势;在 IMDB 数据集上,BiGRU 的分类准确率要比 CNN 准确率要稍微高一点,这是因为 IMDB 数据集的每条评论都很长,BiGRU 可以充分发挥其长距离记忆的优势;通过 MC-CNN 和单通道 CNN, C-LSTM 和 MC-CNN-LSTM, ATT-MCNN-BGRUM 和 MC-AttCNN-AttBiGRU 这 3 组对比实验可以看出,多通道模型相对于单通道模型能够提取到更丰富的情感特征信息,有助于提高情感分类性能;通过对比 C-LSTM 和 Convolution-GRU 可以看出,Convolution-GRU 模型在本文数据集上的分类效果要优于 C-LSTM 模型;通过对比 MC-CNN-BiGRU, MC-CNN-LSTM 和 RCNN, C-LSTM, Convolution-GRU 可以看出多通道的 CNN-RNN 模型还是优于单通道 CNN-RNN 模型;通过 MC-CNN-BiGRU 和 CNN, BiGRU 进行对比可以看出融合了 CNN 和 BiGRU 的多通道模型还是比单独的 CNN, BiGRU 分类效果要好得多,这是因为 MC-CNN-BiGRU 结合了 CNN 和 BiGRU 各自的优点,既能提取到句子间连续单词的局部信息又能够获得句子上下文语义信息,从而得到更高的分类准确率。通过 MC-AttCNN-BiGRU, MC-CNN-AttBiGRU 和 MC-CNN-BiGRU 对比发

现,加入注意力机制的多通道模型可以更好地提高分类准确率,因为它能让模型更加注意到句子中那些对情感分类贡献较大的单词;通过 MC-AttCNN-AttBiGRU 和 MC-AttCNN-BiGRU, MC-CNN-AttBiGRU 对比发现,在 CNN 通道和 BiGRU 通道同时引入注意力机制还能再进一步提升文本分类准确率。

本文提出的 MC-AttCNN-AttBiGRU 模型方法不仅能够提取到句子中连续单词的局部特征信息,还能够捕获句子上下文语义信息,充分发挥了 CNN 和 BiGRU 各自的优势,并且引入注意力机制能够让模型在特征提取的过程中更加关注句子中比较重要的情感词语信息,减少那些对分类不重要的词语的影响,所以本文提出的模型相较于其他几种基线模型能够获得最好的效果。

3.6 注意力可视化

为了更加直观有效地展示模型效果,本文在实验部分使用 seaborn 库和 matplotlib 库进行句子中单词注意力权重分配的可视化展示,本文选取正面和负面评论各一段进行可视化展示实验。

文本 1: 总体感觉还不错.房间很干净、简洁、舒适.

文本 2: 酒店前台小女孩子服务太次,态度很差.

对于文本 1 和文本 2,我们首先使用 Jieba 分词工具对句子进行分词,然后再去除标点符号和停用词,得到“总体 感觉 还 不错 房间 很 干净 简洁 舒适”和“酒店 前台 小 女孩子 服务 太 次 态度 很差”这些单词,然后再分别绘制单词级别的注意力权重热力图如图 6、图 7 所示。图中颜色越深即灰度值越大,

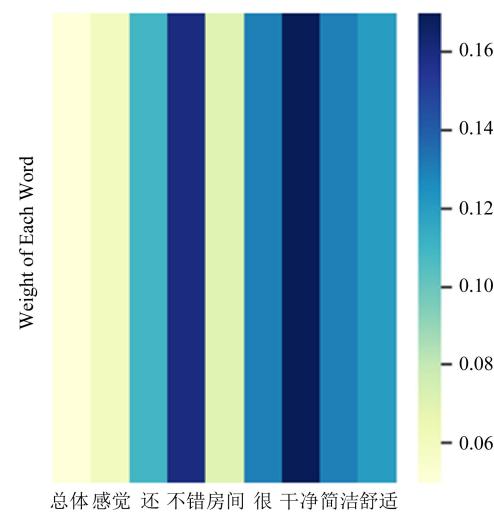


Fig. 6 Text one word attention weight heatmap

图 6 文本 1 单词注意力权重热力图

表示分配给单词的权重越高,说明该单词对后续情感分类的重要性更高。

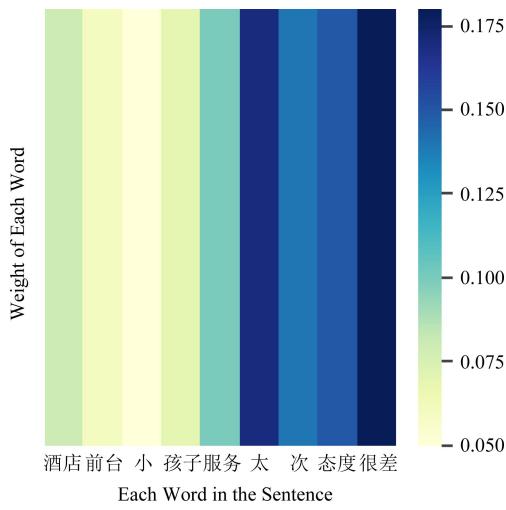


Fig. 7 Text two word attention weight heatmap

图 7 文本 2 单词注意力权重热力图

文本 1 中经过 Jieba 分词和去除停用词和标点之后包含 9 个单词,从图 6 中可以看出模型对“还”、“不错”、“很”、“干净”、“简洁”、“舒适”这几个单词分配了较高的权重,这些都是有关正面评论的词语;文本 2 经过 Jieba 分词和去除停用词之后也包含 9 个单词,从图 7 可以看出模型对“太”、“次”、“很差”这几个单词分配了较高的权重,这些都是有关负面评论的词语。

通过在这 2 段正负评论文本上进行注意力权重可视化实验,结果表明:本文使用注意力机制可以很好地找出句子中对情感分析结果影响较大的单词,证明了注意力机制确实能够关注句子中比较重要的情感词语。

4 总结和展望

本文提出了一种基于注意力机制的多通道 CNN 和 BiGRU 的网络模型用于文本情感倾向性分析,该模型既能够利用 CNN 有效地提取到文本间相邻词语的局部特征,又能够利用 BiGRU 捕获句子上下文全局语义信息,而且引入注意力机制还能更好的注意到句子中对情感分类贡献大的词语。在中英文数据集上多个实验结果充分证明了本文提出模型的有效性,这种多通道混合模型比单一的 CNN 和 BiGRU 网络提取的特征效果更好,且本文提出的模型较其他几种基线模型在 2 个公认的数据集上都取得了最好的分类效果。

虽然本文模型可以通过 CNN 提取局部特征和 BiGRU 网络提取一定的上下文语义结构信息,但是这对于文本情感分类任务还是不够的。传统的文本情感分类方法一般加入了一些句法结构特征,而本文提出的模型没有使用任何句法结构特征,因此下一步工作我将探讨如何让深度学习方法与传统的方法结合,并考虑引入更多文本特征进一步来提高分类准确率。

参 考 文 献

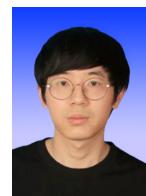
- [1] Liu Bing. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions [M]. Cambridge, UK: Cambridge University Press, 2015
- [2] Pang Bo, Lee L. Opinion mining and sentiment analysis [J]. Foundations and Trends in Information Retrieval, 2008, 2(1/2): 1-135
- [3] Pang Bo, Lee L, Vaithyanathan S. Thumbs up?: Sentiment classification using machine learning techniques [C] //Proc of the Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2002: 79-86
- [4] Hu Minqing, Liu Bing. Mining and summarizing customer reviews [C] //Proc of the 10th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2004: 168-177
- [5] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences [C] //Proc of the 52nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2014: 655-665
- [6] Kim Y. Convolutional neural networks for sentence classification [C] //Proc of the 2014 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2014: 1746-1751
- [7] Zhang Ye, Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification [J]. arXiv preprint, arXiv: 1510.03820, 2015
- [8] Wang Xin, Liu Yuanchao, Sun Chengjie, et al. Predicting polarities of tweets by composing word embeddings with long short-term memory [C] //Proc of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2015: 1343-1353
- [9] Liu Pengfei, Joty S, Meng H. Fine-grained opinion mining with recurrent neural networks and word embeddings [C] //Proc of the 2015 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2015: 1433-1443
- [10] Mikolov T, Chen Kai, Corrado G, et al. Efficient estimation of word representations in vector space [J]. arXiv preprint, arXiv:1301.3781, 2013

- [11] Kamps J, Marx M. Words with attitude [C] //Proc of the Belgian-Netherlands Conf on Artificial Intelligence. Enschede: University of Twente, 2002: 332–341
- [12] Dong Zhendong, Dong Qiang, Hao Changling. HowNet and the Computation of Meaning [M]. Hackensack, NJ: World Scientific, 2006
- [13] Pang Bo, Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts [C] //Proc of the 42nd Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA: ACL, 2004: 271–278
- [14] Lee H, Bhd C. Chinese sentiment analysis using maximum entropy [C] //Proc of the Workshop on Sentiment Analysis Where AI Meets Psychology. New York: ACM, 2011: 89–93
- [15] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch [J]. Journal of Machine Learning Research, 2011, 12: 2493–2537
- [16] Yin Wepeng, Schütze H. Multichannel variable-size convolution for sentence classification [C] //Proc of the 19th Conf on Computational Natural Language Learning. Stroudsburg, PA: ACL, 2015: 204–214
- [17] Chen Ke, Liang Bin, Ke Wende, et al. Chinese micro-blog sentiment analysis based on multi-channels convolutional neural networks [J]. Journal of Computer Research and Development, 2018, 55(5): 945–957 (in Chinese)
(陈珂, 梁斌, 柯文德, 等. 基于多通道卷积神经网络的中文微博情感分析[J]. 计算机研究与发展, 2018, 55(5): 945–957)
- [18] Tang Duyu, Qin Bing, Liu Ting. Document modeling with gated recurrent neural network for sentiment classification [C] //Proc of the 2015 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2015: 1422–1432
- [19] Lai Siwei, Xu Liheng, Liu Kang, et al. Recurrent convolutional neural networks for text classification [C] //Proc of the 29th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2015: 2267–2273
- [20] Zhou Chunting, Sun Chonglin, Liu Zhiyuan, et al. A C-LSTM neural network for text classification [J]. arXiv preprint, arXiv:1511.08630, 2015
- [21] Ruder S, Ghaffari P, John G B. A hierarchical model of reviews for aspect-based sentiment analysis [C] //Proc of the 2016 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2016: 999–1005
- [22] Rao Guozheng, Huang Weihang, Feng Zhiyong, et al. LSTM with sentence representations for document-level sentiment classification [J]. Neurocomputing, 2018, 308 (12): 49–57
- [23] Sachin S, Tripathi A, Mahajan N, et al. Sentiment analysis using gated recurrent neural networks [J]. SN Computer Science, 2020, 1(2): 1–13
- [24] Wang Xingyou, Jiang Weijie, Luo Zhiyong. Combination of convolutional and recurrent neural network for sentiment analysis of short texts [C] //Proc of the 26th Int Conf on Computational Linguistics. New York: ACM, 2016: 2428–2437
- [25] Zhou Peng, Qi Zhenyu, Zheng Suncong, et al. Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling [J]. arXiv preprint, arXiv:1611.06639, 2016
- [26] Zhang You, Yuan Hong, Wang Jin, et al. YNU-HPCC at EmoInt-2017: Using a CNN-LSTM model for sentiment intensity prediction [C] //Proc of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Stroudsburg, PA: ACL, 2017: 200–204
- [27] Zhang Haowei, Wang Jin, Zhang Jixian, et al. Ynu-hpcc at semeval 2017 task 4: Using a multi-channel CNN-LSTM model for sentiment classification [C] //Proc of the 11th Int Workshop on Semantic Evaluation. Stroudsburg, PA: ACL, 2017: 796–801
- [28] Alayba A M, Palade V, England M, et al. A combined CNN and LSTM model for Arabic sentiment analysis [C] //Proc of the Int Cross Domain Conf for Machine Learning and Knowledge Extraction. Berlin: Springer, 2018: 179–191
- [29] Zhang Ziqi, Robinson D, Tepper J. Detecting hate speech on twitter using a Convolution-GRU based deep neural network [C] //Proc of European Semantic Web Conf. Berlin: Springer, 2018: 745–760
- [30] Yang Zichao, Yang Diyi, Dyer C, et al. Hierarchical attention networks for document classification [C] //Proc of the 2016 Conf of the North American Chapter of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2016: 1480–1489
- [31] Wang Linlin, Cao Zhu, Melo G D, et al. Relation classification via multi-level attention CNNs [C] //Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2016: 1298–1307
- [32] Yang Min, Tu Wenting, Wang Jingxuan, et al. Attention based LSTM for target dependent sentiment classification [C] //Proc of the 31st AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2017: 5013–5014
- [33] Wu Chuhan, Wu Fangzhao, Liu Junxin, et al. Thu_ngn at semeval-2018 task 1: Fine-grained tweet sentiment intensity analysis with attention CNN-LSTM [C] //Proc of the 12th Int Workshop on Semantic Evaluation. Stroudsburg, PA: ACL, 2018: 186–192
- [34] Cheng Yan, Ye Ziming, Wang Mingwen, et al. Chinese text sentiment orientation analysis based on convolution neural network and hierarchical attention network [J]. Journal of Chinese Information Processing, 2019, 33(1): 133–142 (in Chinese)
(程艳, 叶子铭, 王明文, 等. 融合卷积神经网络与层次化注意力网络的中文文本情感倾向性分析[J]. 中文信息学报, 2019, 33(1): 133–142)

- [35] Wu Xiaohua, Chen Li, Wei Tiantian, et al. Sentiment analysis of Chinese short text based on self-attention and Bi-LSTM [J]. Journal of Chinese Information Processing, 2019, 33(6): 100–107 (in Chinese)
(吴小华, 陈莉, 魏甜甜, 等. 基于 Self-Attention 和 Bi-LSTM 的中文短文本情感分析[J]. 中文信息学报, 2019, 33(6): 100–107)
- [36] Yuan Hejin, Zhang Xu, Niu Weihua, et al. Sentiment analysis based on multi-channel convolution and bi-directional GRU with attention mechanism [J]. Journal of Chinese Information Processing, 2019, 33(10): 109–118 (in Chinese)
(袁和金, 张旭, 牛为华, 等. 融合注意力机制的多通道卷积与双向 GRU 模型的文本情感分析研究[J]. 中文信息学报, 2019, 33(10): 109–118)
- [37] Liu Longfei, Yang Liang, Zhang Shaowu, et al. Convolutional neural networks for Chinese micro-blog sentiment analysis [J]. Journal of Chinese Information Processing, 2015, 29(6): 159–165 (in Chinese)
(刘龙飞, 杨亮, 张绍武, 等. 基于卷积神经网络的微博情感倾向性分析[J]. 中文信息学报, 2015, 29(6): 159–165)
- [38] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. arXiv preprint, arXiv: 1406.1078, 2014
- [39] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J]. arXiv preprint, arXiv: 1409.0473, 2014
- [40] Maas A L, Daly R E, Pham P T, et al. Learning word vectors for sentiment analysis [C] // Proc of the 49th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2011: 142–150
- [41] Zhang Ye, Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification [J]. arXiv preprint, arXiv: 1510.03820, 2015
- [42] Kingma D P, Ba J. Adam: A method for stochastic optimization [J]. arXiv preprint, arXiv: 1412.6980, 2014
- [43] Joulin A, Grave E, Bojanowski P, et al. Bag of tricks for efficient text classification [J]. arXiv preprint, arXiv: 1607.01759, 2016



Cheng Yan, born in 1976. PhD, professor. Her main research interests include artificial intelligence, intelligent information processing, deep learning and sentiment analysis.



Yao Leibo, born in 1996. Master candidate. His main research interests include deep learning and sentiment analysis. (yaoleibo_jxnu@163.com)



Zhang Guanghe, born in 1979. PhD, lecturer. His main research interests include artificial intelligence and sentiment analysis. (guanghe_zhang@163.com)



Tang Tianwei, born in 1966. PhD, professor. His main research interests include machine translation and sentiment analysis. (1149930364@qq.com)



Xiang Guoxiong, born in 1965. Bachelor, professor. His main research interests include machine learning and deep learning. (1529159315@qq.com)



Chen Haomai, born in 2000. Undergraduate. His main research interests include big data, virtual reality technology. (1152437102@qq.com)



Feng Yue, born in 1984. PhD, lecturer. Her main research interests include deep learning, face recognition. (53689828@qq.com)



Cai Zhuang, born in 1996. Master candidate. His main research interests include deep learning and image processing. (1844791294@qq.com)