

基于注意力机制的评论情感分析及情感词检测

李苑 李智星 滕磊 王化明 王国胤

计算智能重庆市重点实验室 重庆 400065

(yuandaidai1992@163.com)



摘要 评论情感分析是用户生成内容分析的一个研究热点。评论对象的多样性与评论者用语的随意性,导致评论情感分析成为一个非常具有挑战性的任务。现有方法主要通过预先构建情感词表来计算评论的情感极性,但这类方法无法处理同一个词语在不同语境下情感极性存在差异的问题。针对这一问题,文中提出了一种基于注意力的卷积-递归神经网络模型,对评论的情感极性和词语在不同语境下的情感极性进行了建模。通过结合词语在句子中的上下文语境,所提方法可以将注意力集中在主要情感词周围的一个小范围内,并以一种自适应的方式对情感词的情感极性进行计算,提高了词语情感极性判断的准确率,进而提高了短文本的情感极性准确率。与 CRNN, CNN 以及基于情感词典的方法相比,所提方法在中文数据集(美团评论、党建评论)和英文数据集(亚马逊商品评论数据集)上都达到了更好的效果。

关键词:情感分析;注意力机制;卷积-递归神经网络;多粒度

中图法分类号 TP391

Comment Sentiment Analysis and Sentiment Words Detection Based on Attention Mechanism

LI Yuan, LI Zhi-xing, TENG Lei, WANG Hua-ming and WANG Guo-yin

Chongqing Key Lab of Computation Intelligence, Chongqing 400065, China

Abstract Comment sentiment analysis is one of the research hotspots in user generated content field. Because of the diversity of comment objects and the casualness of commentators' language, comment sentiment analysis has become a challenging issue. The existing methods mainly calculate the emotional polarity of comments by pre-building the emotional vocabulary. However, these methods cannot adapt to the problem that the same words have different emotional polarities in different contexts. To overcome this problem, the attention based convolutional-recurrent neural network (A-CRNN) model was proposed to model the emotional polarity of comments and words in different contexts. By combining the context of words in sentences, the proposed method can focus attention on a small scale around the main emotional words. The A-CRNN model calculates the emotional polarity of the words through an adaptive method, which improves the accuracy of words' emotional polarity judgment and the accuracy of short texts' emotional polarity. Compared with CRNN, CNN and emotional dictionary methods, the proposed method achieves better results in Chinese dataset including Meituan Review, Party Building Review and English dataset including Amazon Product Review.

Keywords Emotional analysis, Attention mechanism, Convolutional-recurrent neural network, Multi-granularity

1 引言

随着信息技术的飞速发展,我国已进入“互联网+”时代,国民经济稳定增长,网络购物使用率提升至 63.8%。目前各大电子商务平台都提供了评论功能,越来越多的人愿意分享自己的体验,用户评论数量呈爆炸式增长。调查显示,75%以上的个人消费者以及 90%以上的企业对在线评论是否对购

买决策具有重要影响持肯定态度^[1]。互联网中的评论以其个性化、去权威性和匿名化的特点存在,它们具有独立性和非商业性,具有很大的参考价值,因此人们对于评论中的内容也抱有极大的信任。评论数据具有极高的潜在商业价值,因此商业界期望可以利用文本分析技术挖掘更多有价值的信息。然而,与规范文本如新闻、通告等相比,评论的篇幅长短不一,评论中网络新兴的流行语(不忘初心、柠檬精)丰富,还有缩略语

收到日期:2018-10-31 返修日期:2019-03-25 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家重点研发计划项目(2017YFB0802300);国家自然科学基金青年项目(61502066);重庆市基础与前沿研究计划项目(cstc2015jcyjA40018)

This work was supported by the National Key R&D Program of China (2017YFB0802300), National Natural Science Foundation of China (61502066) and Chongqing Basic and Frontier Research Project(cstc2015jcyjA40018).

通信作者:王国胤(wanggy.cq@hotmail.com)

(LZ:楼主)的出现,评论的形式也多种多样,没有统一的格式或指定要包含的内容,因此对评论的处理工作十分困难。以实际评论为例:“吃了过后我们都拉肚子,拉得虚脱了,估计是食物中毒,再也不吃这种自助餐了”,其中“中毒”表示食物不好,是贬义,而另一条评论“已经买了好几次了,特别好喝,这个酸奶有毒!”中的“有毒”则表达酸奶很好喝,是褒义。虽然同样是关于“毒”的信息,但是所表达的情感却大相径庭,如果仅从词语层面对情感倾向进行分析,则会失去很多细节上有价值的信息,导致人们对评论的理解有失偏颇,从而失去评论本身所具有的可借鉴性。因此,从不同粒度对评论信息进行分析是十分必要的。对于评论的情感倾向分析,由粗粒度到细粒度进行分析更加合理,更符合人类的认知过程,这种方式不仅可以得到评论的整体情感极性,还可以得到更加具体的词语级别的情感信息或语义信息。其中,情感倾向是在粗粒度角度对一个评论的判定,是人们对于一个评论的初步认知。而从词语层面上分析则是从细粒度角度出发,探究评论中表达情感的片段。相比粗粒度而言,细粒度分析可以更加具体地得到评论者所表达的语义。

目前,评论情感分析的研究方法主要分为两个方向:1)基于情感知识的方法(情感词典)。Jha等^[2]通过构建全自动印度语多领域情感词典来解决目标领域中未标记评论的情感分类问题,以减少人工标记的费用。Jha等统计词语之间的共现信息来计算词语间的相关度,并结合种子情感词典对未知词语的情感极性进行估计,进而对评论的情感极性进行分析。但这种方法得到的词语情感极性不可避免地会包含较多噪声。在传统词袋模型中,包含相同词语但词序不同的句子会转化为相同向量,Liu等^[3]通过增加依存句法特征来解决该问题,并利用K-means对词语聚类,进而构建情感词典对在线产品评论进行情感分类。2)基于机器学习的方法^[4-6],此类方法采用朴素贝叶斯、最大熵、支持向量机等若干已标注的评论中进行分类学习,然后对未标注情感的评论进行分类。这些方法能够从句子级别,也就是粗粒度上得到该评论的总体观点、情感,但无法从细粒度上对情感词等词汇层面进行具体分析。

情感词典通常由一系列词语及词语情感极性指数构成,例如知网的情感词典情感词典中除各类情感词外,还包含了情感评价词语等多种类型的词语。词语的情感极性一般是一个预定义的值,由词语本身确定,而一个评论的情感倾向则通常需要结合上下文语境来进行分析,如存在中文的贬词褒用等用法。因此,仅考虑词语在情感词典中的极性,而不考虑词语所在语境,不根据上下文调整词语情感极性是不妥的。与此同时,结合上下文语境对情感词进行分析更符合实际意义。另一方面,机器学习方法虽然可以综合评论中所有词语对评论的情感倾向进行判断,但很显然该方式无法定位到语句中表达评论者主观情感的词语片段位置。

针对以上问题,本文提出基于注意力的卷积-递归神经网络模型(A-CRNN),该模型能够从不同粒度对评论进行分析。通过该模型不仅能得到评论的情感倾向,还可以结合上下文

自适应地计算词语的情感极性,从而定位评论中表达情感的词语片段位置。A-CRNN结合注意力机制,将卷积神经网络与递归神经网络联合使用,即使用卷积神经网络进行特征自提取,学习细粒度信息,随后将卷积神经网络的输出作为递归神经网络的输入。评论文本本身是一种序列化的数据,而递归神经网络对于这类全局信息的把握程度更佳,因此使用递归神经网络从粗粒度上对评论内容进行分析,从而学习文本中有价值的信息。为了使模型自适应地发现情感词,我们引入注意力机制,自适应地计算每一个词语在不同上下文中的权重,对情感词自适应地进行重点关注。本文方法不但可以对评论情感倾向进行分类,还可以自适应地寻找评论中的情感词。该模型可以帮助产品生产商、政府部门等对评论进行自动分析,从而完成相关词库的自动构建与更新、舆情分析与监控等任务。

为了验证所提模型的有效性,本文分别在不同类型的数据集上进行对比实验。其中,美团评论、党建评论为中文数据,亚马逊商品评论为英文数据。在情感倾向的等级划分上,党建评论数据集、亚马逊商品评论数据集有2个等级,美团评论数据集有5个等级。本文的贡献主要有以下方面:

1)提出了一种基于注意力机制的卷积递归神经网络模型,可以对短文本进行有效分类;

2)提出了一种结合上下文的情感词极性计算方法,避免了预定义情感词典无法适应不同语境的问题;

3)实验表明,本文提出方法具有良好的适应性,在不同类型的数据集上都达到了比现有方法更好的效果。

2 相关工作

由于评论分析可帮助产品生产者、服务提供者有针对性地了解客户需求,也可以帮助政策决策者把握社会舆情动态,因此在现实应用中有极大需求,从而促进了理论研究的发展。目前,评论的情感分析有基于传统机器学习(如朴素贝叶斯、支持向量机等)的研究,也有专门针对情感词典自动化或半自动化构建的研究,还有一些学者使用深度学习对此问题进行研究。

基于机器学习的方法普遍面临的问题是需要构建以备模型使用的最优特征。Tripathi等^[7]尝试在电影评论数据上结合TFIDF,分别使用朴素贝叶斯和线性支持向量机对评论的情感进行分类。Bilal等^[8]尝试利用朴素贝叶斯、决策树(J48)和K近邻(k-Nearest Neighbor, KNN)算法对英文与罗马乌尔都语的公众意见极性进行分类。为了在有限的文本中得到更多有价值的信息,Liu等^[3]首先将依存句法分析所得的句法结构作为增加的特征,并与由Word2Vec所得的词向量相结合;然后通过K-means对语义相似的词聚类,从而得到每个词及其对应的类簇;最后用簇标签替换每个单词,并以此作为特征构建情感分类器。机器学习对于特征的选择较为依赖,可以得到评论的整体情感倾向,但无法对评论中词语的情感倾向进行细致分析,因此可将其看作一种粗粒度的分析,而基于情感词典的方法更偏重于评论的细粒度分析。Li等^[9]提

出了一种基于单词分布式表示的跨语言结构对应学习方法。该方法分别从源语言与目标语言中选择具有相似语义的词语作为中心特征。随后,它通过训练所得的词向量计算词语的相似度,并根据相似度建立一对多的映射,形成跨语言词典,从而实现跨语言情感分析。

目前,深度学习应用领域十分广泛,如机器翻译^[10]、问答系统^[11]、信息抽取^[12]等,在情感分析中也有深度学习的应用。目前大部分情感分析工作集中于句子级和篇章级的褒贬意分类任务,评论的情感分析在一定程度上可转化为短文本情感分类问题。对于短文本分类的研究,有部分学者考虑卷积神经网络能够自适应地学习特征,并从局部特征逼近整体特征,因此尝试使用卷积神经网络^[13-17]。他们大多首先将单词变为词向量,随后使用卷积、池化等操作对文本进行分类,虽然具体的操作有所差异,但是总体思想是类似的,他们期望利用卷积神经网络对文本信息进行自适应学习,从多角度挖掘文本特征。除单独使用卷积神经网络外,也有学者将卷积神经网络与其他方法结合使用。李杰等^[18]以有监督的方式训练 CNN,并对未标记的评论进行词性标注,随后使用训练好的 CNN 模型进行特征词提取和评论的情感分类;然后将得到的产品特征词聚类,结合特征词的情感标签与情感分类结果对产品评论进行进一步分析。徐莹莹^[19]提出基于修辞结构理论(RST Parsers)构建长短记忆网络的方法。该方法为 RST 解析树的每个结点增加各自的遗忘门,将隐藏层的状态作为分类器的输入,在 RST 解析树的每个结点上添加一个 softmax 分类器来预测情感分布。该方法在特征表达上的优劣依赖于 RST 解析树的结构,解析树的构建仍值得思考。根据处理文本粒度的不同,我们更加期待可以从词语级、短语级、句子级、篇章级等粒度层次上对文本进行情感分析。如何在有限的文本信息容量中挖掘更多有价值的信息是值得思考的。卷积神经网络擅长从局部特征逼近整体特征,LSTM 则擅长处理序列问题。本文所提模型结合了上述两种深度学习模型的优点并增加了注意力机制,通过卷积神经网络提取特征进行特征自学习,来获取文本中不同类型的特征,同时考虑了文本上下文内容并且通过注意力机制从细粒度上对评论语句进行分析,以自动检测到情感词进而得到评论的情感倾向。

3 方法

3.1 问题定义

本文中, $S = \{s_1, s_2, \dots, s_n\}$ 表示评论语句集合,其中 $s_i = t_1 t_2 \dots t_k \dots$ 是由词语 $t_1, t_2 \dots$ 组成的评论。 $Y = [y_1, y_2, \dots, y_j, \dots, y_n]$ 为 S 的真实标签。由于词语的情感极性和上下文紧密相连,比如“屏幕色彩锐利”中“锐利”表示色彩解析度高,是褒义词,但“边角锐利,割手”中的“锐利”表示做工粗糙,是贬义词。因此,对于每一个词语 t_i ,本文定义它在上下文 c_i 中的情感极性为 $\alpha(t_i, c_i)$,其中 $c_i = t_{i-w} \dots t_{i-1} t_{i+1} \dots t_{i+w}$ 为 t_i 前后文窗口长度为 $2w$ 范围内的词语。令所有上下文为 $C = \{c\}$,那么上下文敏感的词语情感可以定义为矩阵 $\mathbf{A} = \mathbf{R}^{|C| \times |T|}$,其中 $|C|$ 为上下文数量, $|T|$ 为词汇数量。这也是本

文方法和情感词典的主要区别。由于上下文的全集 C 实际上是不可计算的,本文利用注意力机制来拟合情感矩阵 \mathbf{A} 。总的来说,给定评论 $S = \{s_1, s_2, \dots, s_n\}$ 与标签 $Y = \{y_1, y_2, \dots, y_j, \dots, y_n\}$,问题的形式化定义为:

$$\hat{\mathbf{A}}(\hat{\theta}) = \arg \min_{\hat{\theta}} \sum |f(s_i) - y_i| \quad (1)$$

其中, f 为本文提出的 A-CRNN 模型,为情感极性矩阵,是 A-CRNN 的参数。

3.2 总体框架概述

A-CRNN 模型主要包含输入层、词嵌入层、卷积层、注意力层、池化层、长短记忆层和输出层。注意力层作为一个特殊层,帮助 A-CRNN 模型完成语句在不同粒度上的情感倾向的分类和情感词检测任务。

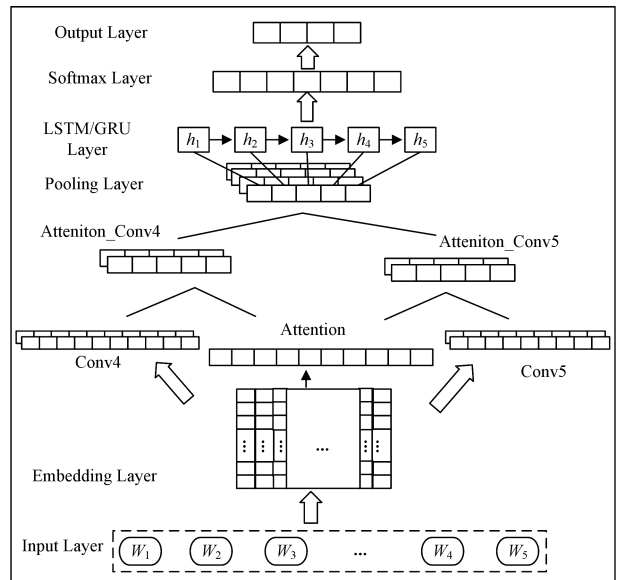


图1 A-CRNN 框架图

Fig. 1 Framework of A-CRNN

3.3 模型的具体结构

3.3.1 数据预处理

本文提出的 A-CRNN 模型不受语言种类限制,可对不同语种的评论进行分类,并分析检测情感词。中文语句中的词汇没有像英文等自然分隔,本文使用 jieba 分词¹⁾对中文分词,如“我喜欢吃中国菜”处理后为“我 喜 欢 吃 中 国 菜”,并利用 jieba 分词得到每个词的词性标注。英文词汇本身由空格分隔,因此不需做分词处理,同时通过 corenlp 得到英文词汇的词性标注。

3.3.2 词嵌入层

为方便计算,评论语句被转变为向量形式作为模型输入。将单词转变为向量(即 s_i 变为 $s_i = [v_1, v_2, \dots, v_l]$)的常用方式有 one-hot 编码和 word2vec。本文使用 word2vec 模型^[20]对单词进行向量化表示。由于评论集中词汇总量巨大,经 one-hot 编码所产生的向量矩阵过于稀疏,因此本文选取文本中的 TOP10 词性进行 one-hot 编码,并与对应词的 word2vec 向量直接拼接。

¹⁾ <https://github.com/fxsjy/jieba>

3.3.3 卷积层

卷积层由多个特征图(Feature Map)组成,每个特征图都由多个神经元构成,神经元通过卷积核与上一层连接,自动地提取特征。本文通过多卷积核(kernel)的方式,利用多卷积过滤器(filter)来帮助模型进行特征自学习,该方式可以尽可能多地学习到不同类型的特征。 $H_i \in \mathbf{R}^{d \times w}$ 为卷积层的权重矩阵, b_i 为偏置,其中 w 为卷积的滑动窗口大小,那么在特征图 $C \in \mathbf{R}^{l \times w+1}$ 中的第 i 元素可通过式(2)得到。第 i 列到第 $i+w$ 列的向量由 $V[* , i:i+w]$ 表示,将得到的卷积结果与注意力层相结合。

$$c_i = \sigma(\sum(V[* , i:i+w] \circ H_i) + b_i) \quad (2)$$

3.3.4 注意力层

注意力机制是一种人类特有的视觉处理方式,它对无用的信息进行抑制,对需要关注的目标区域投入更多的注意力资源,以获取目标的更多细节信息。不同任务注意力机制对注意力的资源分配方式是不同的,如上文所述,在评论的这种序列数据中,上下文之间往往存在联系,而这些具体的联系是什么或如何进行刻画是本文需要重点解决的一个问题。本文提出使用矩阵 $\mathbf{A} = \mathbf{R}^{|C| \times |T|}$ 刻画词语的情感极性,但由于上下文无法预先确定 $|C|$,且随着语料库的增大 $|C|$ 会急剧增加,因此无法直接对 \mathbf{A} 进行计算。在注意力层,本文使用注意力机制来对 \mathbf{A} 进行拟合。对于某一个句子中的词语 t_i ,其注意力分配的计算方式如下:

$$a_i = \sigma(\sum(V[i-w:i+w] \circ \mathbf{H}_a) + b_a) \quad (3)$$

其中, \mathbf{V} 是上下文词语 c_i 的向量表示, w 为窗口大小, $\mathbf{H}_a \in \mathbf{R}^{d \times w}$ 为注意力层的权重矩阵, b_a 为注意力层的偏置。本文所使用的注意力机制类似于滑动窗口的原理,当人们需要的上下文内容改变时,可以对 w 的大小进行改变。 w 变小时,减少与当前词有关的上下文内容; w 变大时,增加与当前词有关的上下文信息内容。随后,将注意力矩阵 \mathbf{A} 与特征图 C 结合得到新的特征图 \hat{C}_A ,将其中对应元素相乘 $\hat{C}_A = C \cdot \mathbf{A} = c_i = c_i \cdot a_i$,从而得到新的特征图 \hat{C}_A 。其中,改变卷积核 w 的大小,可以从不同粒度对评论句进行划分,通过这一方式可以实现从不同粒度对语句进行分析。该方式也可扩展到更粗的粒度上,例如在分析篇章时,可使用该方式为语句分配合理的注意力资源,也可以对更细粒度上的词语进行分析,进而从多粒度的角度出发对评论等文本加以分析。在注意力机制的帮助下,A-CRNN模型能够更好地进行特征自学习,注意力层也可以自适应地对情感词赋予区别于其他词的注意力值,使该模型在整体模型调整的过程中可得到更优的效果。

3.3.5 池化层与连接层

池化层的作用是特征二次提取,常用的池化函数有最大池化、平均池化等。本文采用最大池化,其对于周围的最大值比较敏感而对于精确位置并不敏感,这有利于选取注意力值相对较高的特征。该方式可以降低维数,防止过拟合。卷积层窗口大小 $h=4$ 时,得到 $pool_4 = \max\{\hat{c}_1, \hat{c}_2, \dots, \hat{c}_m\}$,卷积层窗口大小 $h=5$ 时,得到 $pool_5 = \max\{\hat{c}_1, \hat{c}_2, \dots, \hat{c}_m\}$ 。经池化得到结果 $X = \oplus(pool_4, pool_5)$ ($X \in \mathbf{R}$),将其连接所得的新矩阵 $\mathbf{X} = [x_1, x_2, \dots, x_n]$ 作为LSTM层输入。

3.3.6 LSTM层

递归神经网络常被用于处理序列任务。区别于传统递归神经网络,长短期记忆网络(LSTM)可长期记忆相关信息,它由一系列重复的时间模块组成,每个模块由遗忘门、输入门、输出门和一个记忆细胞结构 $c_t \in \mathbf{R}$ 构成。在时间片 t 时,输入是 x_t ,隐藏状态为 h_{t-1} ,记忆细胞为 c_{t-1} ,输出是 h_t 和 c_t 。LSTM层中的更新方式为:

$$i_t = \sigma(\mathbf{W}_i x_t + \mathbf{U}_i h_{t-1} + b_i) \quad (4)$$

$$f_t = \sigma(\mathbf{W}_f x_t + \mathbf{U}_f h_{t-1} + b_f) \quad (5)$$

$$o_t = \sigma(\mathbf{W}_o x_t + \mathbf{U}_o h_{t-1} + b_o) \quad (6)$$

$$g_t = \tanh(\mathbf{W}_g x_t + \mathbf{U}_g h_{t-1} + b_g) \quad (7)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (8)$$

$$h_t = o_t \odot \tanh(c_t) \quad (9)$$

其中, $\mathbf{W}_f, \mathbf{W}_g, \mathbf{W}_o, \mathbf{W}_i$ 为LSTM权重矩阵, b_f, b_g, b_o, b_i 为LSTM的偏置, σ 为激活函数sigmoid, \odot 为点乘操作。

3.3.7 全连接层

该层可整合卷积层和池化层中具有类别区分性的局部信息,也结合了LSTM层的全局信息。本文使用softmax函数进行处理,计算式如下:

$$\hat{y}_i = \frac{\exp(o_i)}{\sum_{j=1}^{\hat{C}_A} \exp(o_j)} \quad (10)$$

与此同时,本文还使用dropout技术来避免训练过拟合,A-CRNN模型在本层利用交叉熵 $H_y(y') = -\sum_i y_i \cdot \log(y_i')$ 分类,得到结果 $Y' = [y_1', y_2', \dots, y_j', \dots, y_n']$ 。

4 实验

4.1 实验数据

为了验证本文模型的有效性,分别在美团餐厅评论数据集(以下简称美团评论)、党建宣传网页与某邪教组织网页数据集(以下简称党建评论)、亚马逊食品评论数据集(以下简称亚马逊数据集)上进行实验。其中,美团评论数据集是一个中文多分类数据集,其数据来自于美团网上用户对一些餐厅的真实评论;党建评论是一个二分类中文数据集,党建评论为党建部门官网所发布的和谐积极评论,而某邪教组织网页评论内容多为反动言论;亚马逊食品评论是二分类英文数据集,按照人们对商品的满意程度进行划分的数据集。这些数据集从不同维度验证了该模型在不同语言、不同应用场景以及不同分类任务下均可检测到情感词并得到评论的情感倾向。数据集具体信息如表1所列(其中,类别比例一列中“1:2:3:4:5”表示美团评论中评分为1-5的5个类别)。

美团评论:该数据集是一个多分类数据集,它是在美团上自行爬取的用户对重庆地区的真实评论。其包含评论者ID、纯文本评论语句、评分。本文根据评分不同对1-5分所对应的评论进行分类(1表示消极,2表示稍消极,3表示中性,4表示稍积极,5表示积极)。数据集总计126000余条评论。

党建评论数据集:数据来源于外网中有关党政的负面评论以及中国共产党官网等相关的正确报道内容的数据集。该数据集是从党建部门官网和某邪教组织网页评论上自行爬取得到的,是一个二分类文本数据集,总计38000余条评论,我

们将党建信息、某邪教网站评论分别作为正负两类。

亚马逊数据集¹⁾:该数据集是一个二分类数据集^[21]。数据时间跨越 10 年以上,包括截至 2012 年 10 月的 568454 条评论,内容包括产品信息、用户信息、评分以及纯文本评论。该数据集由于包含的内容较多,使用的范围较广,也曾被用于他分析任务,例如修辞结构理论研究^[22]、主题模型研究^[23]等。本文使用该数据集中的评分与相应的纯文本评论内容。

表 1 数据集信息汇总
Table 1 Summary of datasets

数据集	类别	训练集 样本数	测试集 样本数	类别比例	词汇 数量
美团评论	5	115000	11388	1:2:3:4:5= 1:1:1.5:3:5	36413
党建评论	2	36000	2050	Negative:Positive=1:5	46821
亚马逊评论	2	50000	5047	Negative:Positive=1:4	114133

4.2 参数设置

词嵌入大小为 300 维,并额外增加 10 维词性的独热编码。卷积层为多卷积核,卷积核大小为 4 和 5 两种,两种卷积核数量均为 200。此外,模型使用批处理,批处理大小为 64,epoch 数目为 60。为了防止过拟合,还引入了 dropout 技术。

4.3 实验结果分析

由表 2 可知,在精确度指标 (ACC) 方面,A-CRNN 方法不仅在美团评论中文多分类数据集上明显优于 CRNN^[24] 和 CNN^[25],同时在中文二分类和英文二分类数据集上也优于其他两种方法。因此 A-CRNN 不受限于类别数量和语言种类。

表 2 不同算法的精确率比较

Table 2 Results of ACC of different algorithms

数据集	(单位:%)		
	A-CRNN	CRNN	CNN
美团评论	56.272	54.882	47.098
党建评论	98.195	97.920	89.038
亚马逊评论	92.867	92.339	92.278

除了比较深度学习的方法外,本文还与传统的机器学习方法中的朴素贝叶斯、KNN、Logistic Regression、随机森林、决策树、GBDT 等方法进行了比较。由表 3 可知,在精确度指标方面,A-CRNN 方法在 3 个数据集上的实验效果均明显优于传统的机器学习方法,并且具有较大的优势。

表 3 机器学习方法的精确率比较

Table 3 Results of ACC of machine learning algorithms

数据集	(单位:%)		
	美团评论	党建评论	亚马逊评论
A-CRNN	56.272	98.195	92.867
Naive Bayes	45.16	89.83	80.18
KNN	43.37	86.30	80.92
Logistic Regression	52.20	95.33	83.92
Random forest	41.88	90.80	82.06
Decision tree	37.70	87.28	76.56
GBDT	52.22	94.55	82.10

为有效说明类别内部分类的效果,本文除了使用精确度 (Accuracy) 指标,还使用了带权平均的准确率 (Precision)、召

回率、F1 值等评价指标来评价分类结果。通常,带权平均的评测指标更能反映模型在样本分布不平衡数据集上的真实效果。带权平均的准确率 (Precision)、召回率、F1 值的计算方法如下:

$$Macro_P = \sum_i \frac{n_i}{n_1 + n_2 + \dots + n_m} \times P_i \quad (11)$$

$$Macro_R = \sum_i \frac{n_i}{n_1 + n_2 + \dots + n_m} \times R_i \quad (12)$$

$$Macro_F = \frac{2 \times Macro_P \times Macro_R}{Macro_P + Macro_R} \quad (13)$$

其中, n_i 为第 i 类数据数量, P_i 为第 i 类的准确率, R_i 为第 i 类的召回率。

由表 3 可知,在 3 个数据集上 A-CRNN 模型的带权平均的准确率、召回率、F1 值都优于 CRNN 和 CNN 方法。这说明 A-CRNN 模型更加行之有效,也说明该方法不受限于语种和类别数量。

表 4 3 个数据集上带权平均的准确率、召回率、F1 值的结果

Table 4 Results of Macro_P, Macro_R, Macro_F on three datasets

		A-CRNN	CRNN	CNN
美团评论	Macro_P	0.5042	0.5021	0.3988
	Macro_R	0.5605	0.5488	0.4760
	Macro_F	0.5309	0.5244	0.4340
党建评论	Macro_P	0.9820172	0.9792	0.8920
	Macro_R	0.9819635	0.9793	0.8899
	Macro_F	0.9819903	0.9792	0.8909
亚马逊评论	Macro_P	0.8418	0.8144	0.8344
	Macro_R	0.8596	0.8177	0.8427
	Macro_F	0.8506	0.8161	0.8386

4.4 可视化结果分析

图 2、图 3 所示的可视化结果分别展示了 A-CRNN 模型在细粒度上情感可视化的分布情况。颜色越深,代表该词语在本条评论中的情感倾向越消极;颜色越浅,代表该词语在本条评论中的情感倾向越积极。相同或者类似的词语在不同评论中的颜色不同。在图 2 中,美团的有关食物的评价中“这个酸奶有毒!”,“有毒”实际是对酸奶的夸奖,是十分好喝的意思,而同样是类似于有毒的词语在“估计是食物中毒”的评论中则表达食物不新鲜的含义,此时“食物中毒”这个词的颜色较深。这说明 A-CRNN 模型在引入注意力机制后,能够更好地结合上下文内容给出相应的结果。从图 3 中可以看到,A-CRNN 模型同样可识别出亚马逊英文评论中的“love”“perfect”等表示喜欢的情感词语,以及“bitter”“weak”等表示消极含义的词语。从 A-CRNN 模型的可视化结果可以看出,该模型不但不会受限于语言种类,而且在不依赖任何情感词典的前提下能够有效识别出表达评论者情感的词语。本文所进行的实验在情感词检测部分区别于传统的情感词检测任务,因为它并没有使用预先得到的情感词词典,而是使用注意力机制结合上下文信息检测到情感词。相比较而言,传统的情感词检测是静态的过程,而本文所使用的方式是动态过程。例如,知网的情感词典²⁾ 提供了很多表示正面和负面的词,并且

¹⁾ <https://www.kaggle.com/snap/amazon-fine-food-reviews>

²⁾ http://www.keenage.com/html/c_bulletin_2007.htm

结合了当前的语境信息。如“现在的仿冒产品太逼真了,很难判断真伪。”中“逼真”所表达的并不是正面的含义。因为现实生活中所遇到的更多的是没有标签的情感词问题,人们在得到这些语句的情感倾向的同时,也渴望探知传递出这一情感的具体词汇是哪些。本文使用 A-CRNN 模型从一个新的角

度进行尝试,在具体语境中探寻情感词。A-CRNN 模型从不同程度上对评论情感加以分析并对情感词进行初步检测。在情感词检测方面可以通过改变当前词语的上下文信息的数量来进行动态的分析。对评论的情感倾向做到“知其然,知其所以然”,这更符合人类认知的过程。

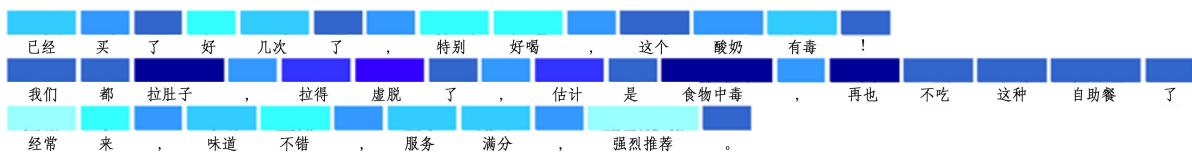


图2 美团评论细粒度情感可视化

Fig. 2 Visualization of fine-grained emotion on Meituan dataset



图3 亚马逊数据集细粒度情感可视化

Fig. 3 Visualization of fine-grained emotion on Amazon dataset

结束语 本文所提出的 A-CRNN 是一种基于注意力机制的卷积递归神经网络模型,该模型在卷积层进行特征自学习,并引入注意力机制协助模型刻画词语的情感极性,然后通过递归神经网络将评论中的上下文语境信息融合到模型中,最终得到评论的情感分类和评论中的情感词。为验证模型的有效性,本文分别在中文数据集美团评论、党建评论以及英文数据集亚马逊商品评论上进行验证。实验结果表明,A-CRNN 模型在情感分类问题上可以从不同粒度对评论的情感倾向加以分析,能够达到更高的准确率并且可以自适应地检测得到评论中的情感词。该模型具有良好的适应性,能够避免预定义情感词典无法适应不同语境的问题,与此同时也规避了不同语言的限制问题。在未来的工作中,我们可以将该模型应用于各大社交媒体的违法信息词语和人身攻击词语的过滤工作中。同时可以尝试改进模型,增加评论者的信息等以进一步提高评论情感分类的效果。

参考文献

[1] LIU C Y, RAN Q. Attitudinal Resources in Customers' Remarks in CSC and Their Impacts on Potential Customers' Purchase Decisions[J]. Journal of University of Science and Technology Beijing (Social Sciences Edition), 2017, 6(33): 1-7.

[2] JHA V, SAVITHA R, SHENOY P D, et al. A novel sentiment aware dictionary for multi-domain sentiment classification[J]. Computers & Electrical Engineering, 2018, 69: 585-597.

[3] LIU F, WEI F, YU K, et al. Sentiment Classification of Reviews on Automobile Websites by Combining Word2Vec and Dependency Parsing[C]//Proceedings of the International Conference on Smart Computing and Communication. Springer, Cham, 2017: 206-221.

[4] TRIPATHY A, RATH S K. Classification of sentiment of reviews using supervised machine learning techniques[J]. International Journal of Rough Sets and Data Analysis (IJRSDA),

2017, 4(1): 56-74.

[5] MANEK A S, SHENOY P D, MOHAN M C, et al. Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier[J]. World Wide Web, 2017, 20(2): 135-154.

[6] YORDANOVA S, KABAKCHIEVA D. Sentiment Classification of Hotel Reviews in Social Media with Decision Tree Learning [J]. International Journal of Computer Applications, 2017, 158(5): 0975-8887.

[7] TRIPATHI G, NAGANNA S. Feature selection and classification approach for sentiment analysis[J]. Machine Learning and Applications: An International Journal, 2015, 2(2): 1-16.

[8] BILAL M, ISRAR H, SHAHID M, et al. Sentiment classification of Roman-Urdu opinions using Naive Bayesian, Decision Tree and KNN classification techniques [J]. Journal of King Saud University-Computer and Information Sciences, 2016, 28(3): 330-344.

[9] LI N, ZHAI S, ZHANG Z, et al. Structural Correspondence Learning for Cross-Lingual Sentiment Classification with One-to-Many Mappings[C]//Thirty-First AAAI Conference on Artificial Intelligence. San Francisco, USA, 2017: 3490-3496.

[10] GEHRING J, AULI M, GRANGIER D, et al. Convolutional Sequence to Sequence Learning[C]//Proceedings of the International Conference on Machine Learning. Sydney, NSW, 2017: 1243-1252.

[11] HE H, GIMPEL K, LIN J. Multi-perspective sentence similarity modeling with convolutional neural networks[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar, 2015: 1576-1586.

[12] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks [C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar, 2015: 1753-1762.

- [13] KIM Y. Convolutional Neural Networks for Sentence Classification[C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar, 2014:1746-1751.
- [14] ZHANG Y, WALLACE B. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification[C] // Proceedings of the Eighth International Joint Conference on Natural Language Processing. Taipei, China, 2017, 1:253-263.
- [15] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification[C] // Processing of the 2015 Neural Information Processing Systems. San Diego, USA, Neural Information Processing Systems Foundation, 2015:649-657.
- [16] CONNEAU A, SCHWENK H, BARRAULT L, et al. Very deep convolutional networks for text classification[C] // Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Valencia, Spain; Association for Computational Linguistics, 2017:1107-1116.
- [17] GUO H. A Deep Network with Visual Text Composition Behavior[C] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, 2017, 2:372-377.
- [18] LI J, LI H. Research on Product Feature Extraction and Sentiment Classification of Short Online Review Based on Deep Learning[J]. Information Studies: Theory & Application, 2018, 2:26.
- [19] XU Y Y. Research of Sentence-Level Sentiment Classification for Text Based on Deep Neural Network[D]. Shenzhen: Shenzhen University, 2016.
- [20] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[C] // Proceedings of the International Conference on Learning Representations. Scottsdale, Arizona, 2013:1-12.
- [21] MCAULEY J J, LESKOVEC J. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews [C] // Proceedings of the 22nd International Conference on World Wide Web. Rio de Janeiro, Brazil; Association for Computing Machinery, 2013:897-908.
- [22] KRAUS M, FEUERRIEGEL S. Sentiment analysis based on rhetorical structure theory; Learning deep neural networks from discourse trees[J]. Expert Systems with Applications, 2019, 118:65-79.
- [23] JOSHI A, JAIN P, BHATTACHARYYA P, et al. Who would have thought of that!: A Hierarchical Topic Model for Extraction of Sarcasm-prevalent Topics and Sarcasm Detection[C] // Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM). 2016:1-10.
- [24] WANG X, JIANG W, LUO Z. Combination of convolutional and recurrent neural network for sentiment analysis of short texts [C] // Proceedings of the 26th International Conference on Computational Linguistics. Osaka, Japan; Association for Computational Linguistics, 2016:2428-2437.
- [25] KIM Y. Convolutional neural networks for sentence classification[C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar; Association for Computational Linguistics, 2014: 1746-1751.



LI Yuan, born in 1992, master. Her main research interests include deep learning and network security.



WANG Guo-yin, born in 1970, Ph. D., professor, Ph. D supervisor. His main research interests include rough set, granular computing, data mining and machine learning.