

文章编号: 1003-0077(2021)02-0116-09

融合多特征的分段卷积神经网络对象级情感分类方法

周 武¹, 曾碧卿², 徐如阳¹, 杨 恒¹, 韩旭丽¹, 程良伦³

(1. 华南师范大学 计算机学院, 广东 广州 510631;

2. 华南师范大学 软件学院, 广东 佛山 528225;

3. 广东省信息物理融合系统重点实验室, 广东 广州 510006)

摘 要: 对象级情感分类旨在判断句子中特定对象的情感极性类别。在现有基于卷积神经网络的研究中,常在模型的池化层采用最大池化操作提取文本特征作为句子表示,该操作未考虑由对象所划分的上下文,因此无法得到更细粒度的对象上下文特征。针对该问题,该文提出一种融合多特征的分段卷积神经网络(multi-feature piecewise convolution neural network, MP-CNN)模型,根据对象将句子划分为两个部分作为上下文,并在池化层采用分段最大池化操作提取上下文特征。此外,该模型还将有助于情感分类的多个辅助特征融入其中,如词的相对位置、词性以及词在情感词典中的情感得分,并通过卷积操作计算词的注意力得分,有效判断对象的情感极性类别。最后在 SemEval 2014 数据集和 Twitter 数据集的实验中,取得了较基于传统机器学习、基于循环神经网络以及基于单一最大池化的卷积神经网络分类模型更好的分类效果。

关键词: 多特征;分段;卷积神经网络;对象级情感分类

中图分类号: TP391

文献标识码: A

Multi-Feature Piecewise Convolution Neural Network for Aspect-Based Sentiment Classification

ZHOU Wu¹, ZENG Biqing², XU Ruyang¹, YANG Heng¹, HAN Xuli¹, CHENG Lianglun³

(1. School of Computer Science, South China Normal University, Guangzhou, Guangdong 510631, China;

2. School of Software, South China Normal University, Foshan, Guangdong 528225, China;

3. Guangdong Provincial Key Laboratory of Cyber-Physical System, Guangzhou, Guangdong 510006, China)

Abstract: Aspect-based sentiment classification aims at judging the sentiment polarity of a particular aspect in a sentence. In the existing research on convolution-based neural networks, the maximum pooling operation is often used to extract text features as sentence representation in the pooling layer of the model. This operation does not consider the context divided by the aspect and fails to get finer-grained aspect context features. To solve this problem, this paper proposes a multi-feature piecewise convolution neural network (MP-CNN) model. According to the aspect, the sentence is divided into two parts of context, and in the pooling layer, the maximum pooling operation is used to extract the context features. In addition, this paper also integrates several auxiliary features into the model, such as relative position of words, part of speech and sentiment score of words in sentiment lexicon, and calculates the attention score of words through convolution operation. The experiments of SemEval 2014 and Twitter datasets confirm the best performance among the baselines.

Keywords: multi-feature; piecewise; convolutional neural network; aspect-based sentiment classification

收稿日期: 2019-06-19 定稿日期: 2019-08-07

基金项目: 国家自然科学基金(61876067);广东省普通高校人工智能重点领域专项(2019KZDZX1033);广东省信息物理融合系统重点实验室建设专项(2020B1212060069)

0 引言

情感分类是自然语言处理领域中一个常见的任务,旨在挖掘人们在文本中所表达观点情感倾向的类别^[1]。主要研究集中在对篇章级别、句子级别和对象级别的情感分类任务上^[2]。其中,对象级情感分类是一种相对比较细粒度的分类问题,旨在判断句子中对某个特定对象的情感类别。例如,“这台笔记本电脑的功能很强大,但价格太高了!”,该任务需要针对对象“功能”判断正向的情感类别,以及针对对象“价格”判断负向的情感类别。

基于深度学习的方法能够自动提取文本的语义特征,避免了人工提取特征繁琐的过程,因此该方法广泛应用于自然语言处理的各个研究领域,如词性标注^[3]、问答系统^[4]、机器翻译^[5]等。在对象级情感分类问题的研究中,以往的方法大多利用长短时记忆网络(long short-term memory network, LSTM)具有捕捉序列特征的能力或者在 LSTM 的基础上融入注意力机制(attention mechanism)来对分类任务进行建模。其中, Tang 等人^[6]提出 TD-LSTM 模型,该模型使用预训练词向量表示单词,根据对象在句子中的位置,利用 LSTM 提取对象的上下文特征作为最终的句子表示。Wang 等人^[7]提出 ATAE-LSTM 模型,其认为不同的词将会对分类的结果产生不同程度的影响,作者在 LSTM 的基础上引入注意力机制计算每一个词的注意力权重,为句子中对情感极性分类起到重要作用的形容词、词组等分配较大的权重。

有研究指出,卷积神经网络(convolutional neural networks, CNN)和 LSTM 在文本分类任务上效果基本接近^[8],但 LSTM 在进行序列建模时具有时间依赖性,需要训练较长的时间,而 CNN 结构相对简单且可并行化,因此在 CNN 的基础上,梁等人^[9]提出一种基于词向量、词性和位置的多注意力机制卷积神经网络模型 CATT-CNN 来识别不同对象的情感极性。Xue 等人^[10]提出一种基于门控机制的卷积神经网络模型 GCAE,该门控机制可独立工作且能根据给定的对象有选择地输出情感特征,因此模型结构更加简单且训练快速。

就人理解的角度而言,在判断某个对象的情感极性时,对象的上下文会对判断产生很大的影响。但在现有以 CNN 为基础的研究方法中,大多并未考虑到由对象所划分出的上下文对分类效果的影

响,在池化层采用最大池化(max-pooling)操作提取句子特征,因此无法得到更细粒度的上下文特征。针对该问题,本文提出一种融合多特征的分段卷积神经网络模型,根据对象将句子划分为包含上下文的两个部分,并在池化层分别对这两个部分进行最大池化操作,即通过分段的方式提取包含对象上下文的特征。此外,本文还充分考虑到有利于情感分类的多个辅助特征,将词的相对位置、词性以及词在情感词典中的情感得分融入模型,并通过卷积操作为每个词计算注意力得分,在 SemEval 2014 数据集和 Twitter 数据集的实验中,结果表明,本文提出的模型能够有效地判别句子中对象的情感极性。

本文的主要贡献总结如下:

(1) 提出了一种融合多特征的分段卷积神经网络模型,模型根据对象将句子划分为包含上下文的两个部分,并在模型的池化层采用分段最大池化操作,有效地提取包含对象上下文信息的细粒度特征;

(2) 利用词的相对位置、词性以及词在情感词典中的情感得分作为辅助特征嵌入到模型中,使模型能够学习到更加丰富且有利于情感分类的特征,同时使用卷积操作为每个词计算注意力得分,有效地为那些有利于分类的词分配更大的注意力权重;

(3) 在 SemEval 2014 数据集和 Twitter 数据集的实验中,对比基于传统机器学习、基于循环神经网络以及基于单一最大池化机制的卷积神经网络分类模型,本文提出的方法取得了更好的分类效果。

1 相关工作

1.1 对象级情感分类相关研究

近年来,在深度学习领域,学者们利用 LSTM 和 CNN 具有自动提取文本的语义特征的能力,避免了传统机器学习算法人工提取特征的繁琐过程,对分类任务进行有效的建模。其中, Tang 等人^[6]利用 LSTM 提取对象的上下文特征作为句子表示,提出对象依赖的分类模型 TD-LSTM。Wang 等人^[7]提出一种基于注意力机制的 LSTM 模型 ATAE-LSTM,该模型利用注意力机制为句子中的每个词计算注意力权重,对词向量加权求和以表示句子,取得了较 TD-LSTM 更好的分类效果。在 CNN 的基础上,梁等人^[9]根据词向量、词性和位置信息,提出一种基于多注意力的卷积神经网络模型 CATT-

CNN 来识别不同对象的情感极性。Xue 等人^[10]提出 GCAE 模型,该模型在 CNN 上引入一种能够独立工作的门控机制,其能够根据给定的对象有选择地输出情感特征,使分类更加快速和准确。

1.2 辅助特征

Matsumoto 等人^[11]在带情感标注的日语和英语语料库上,提出了一种估计句子中情感表达的方法,该方法充分利用词性和词之间的位置特征,能更好地识别出情感词或者短语。陈等人^[12]利用外部情感词典资源,在句子中提取和情感表达相关的序列片段,凭借词语本身具有的属性特征来构建词向量以作为 CNN 的输入之一,在其研究领域数据集上的情感分类效果优于未引入词典资源的 CNN 模型。显然,词性、词的位置信息以及外部情感词典的运用,在一定程度上对情感分类问题发挥了积极影响。

1.3 分段卷积神经网络

在知识图谱关系抽取领域的研究中,要对两个给定实体之间的关系进行分类。Zeng 等人^[13]认为句子本质上可由给定的两个实体划分成 3 个部分,为了获取两个实体之间更加丰富的语义特征,作者设计了一个如图 1 所示的分段卷积神经网络模型。其中,两个实体在 Word Embedding 层用深黑色标识,Convolution 层使用 3 个不同的过滤器提取局部 n-gram 特征,Pooling 层采用分段最大池化操作,返回每个分段中的最大值,而不是单个句子的最大值,

与传统最大池化方法相比,该方法取得了更好的分类效果。类似地,在对象级情感分类任务中,句子可由对象划分成上下两个部分,其中包含了对象的上下文信息,若使用分段卷积神经网络对上下文进行建模,采取分段最大池化策略,将得到更加细粒度的对象上下文特征。

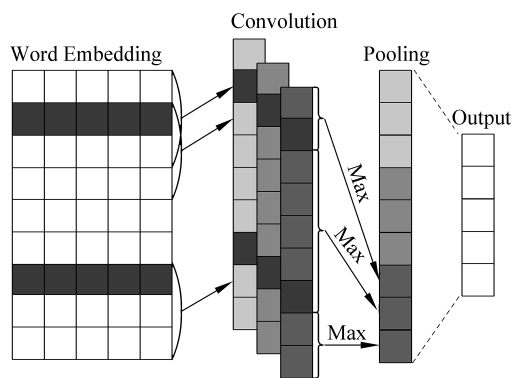


图 1 分段卷积神经网络^[13]

2 融合多特征的分段卷积神经网络模型

考虑到由对象所划分上下文的重要性,本文使用分段卷积神经网络对上下文进行建模。图 2 为本文提出的 MP-CNN 模型的网络结构,模型主要分为五个部分:词的多特征向量表示、卷积层、分段池化层、特征连接层以及 softmax 层。以下将对这些部分以及注意力计算方式进行详细说明。

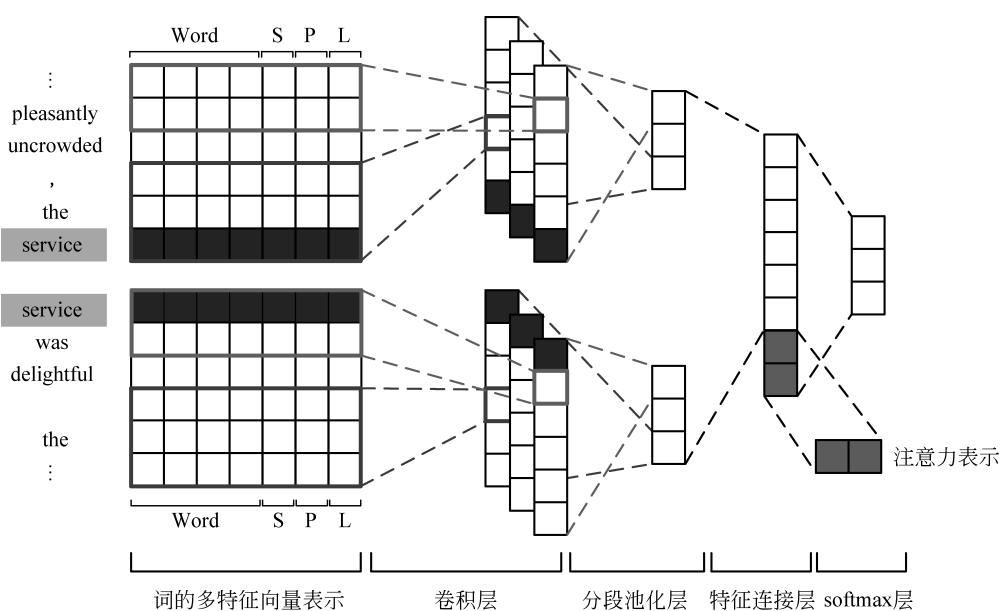


图 2 MP-CNN 网络结构图

2.1 词的多特征向量表示

为将词和多个辅助特征嵌入模型,需要将其进行向量化表示。假设包含对象的任意句子中包含 n 个词, $i \in [1, n]$, $e^i \in \mathbb{R}^{|V|}$ 为 0 和 1 组成的 one-hot 向量,向量中只有表示该词所在词典中位置处的元素值为 1,其余元素值为 0, V 表示数据集中词的集合, $|V|$ 为词表的大小,则词和辅助特征的嵌入表示描述如下:

词嵌入 指的是在“上下文相似的词,其语义也相似”假设的基础上,将词嵌入到一个维度较低且固定的空间当中,以得到一种稠密且包含语义的分布式词向量。目前训练词向量的方法主要包括 Word2Vec^[14-15]、GloVe^[16] 以及 FastText^[17]。假设 $x_i \in \mathbb{R}^d$ 为句子中某个词的词向量, $0 \leq d \leq 300$ 词嵌入矩阵表示为 $W_e \in \mathbb{R}^{d \times |V|}$,则获取词向量的查找操作 look-up 可表示如式(1)所示。

$$x_i = W_e e_i \quad (1)$$

情感得分特征嵌入 词的情感得分指反映其情感倾向程度的值。值越大代表该词为正向情感词的概率越大,反之则越小。设该特征嵌入的维度为 m_1 , (s_1, s_2, \dots, s_n) 为情感得分特征向量,其中 s_i 为句子中第 i 个词对应的情感得分特征值,抽取自情感词典 NRC Hashtag Sentiment Lexicon,该词典由 Mohammad 等人^[18]用 777 310 条包含大量情感词的 Twitter 推文所构建。 $S \in \mathbb{R}^{n \times m_1}$ 为句子的情感特征嵌入矩阵,则特征值 s_i 对应的情感得分特征嵌入向量 $S_i \in S$ 可表示如式(2)所示。

$$S_i = W_s e_i \quad (2)$$

其中, $W_s \in \mathbb{R}^{m_1 \times |V|}$ 为可训练矩阵,在模型训练过程中不断学习和更新,对于未登录词以及包含对象的词,其情感特征值设置为 0。

词性特征嵌入 词性信息亦是一种有利于情感极性判断的特征,因为若一个词为形容词,那么该词很有可能用以表达观点。类似地,设该特征嵌入的维度为 m_2 ,词性特征向量为 (p_1, p_2, \dots, p_n) ,其中 p_i 为句子中第 i 个词对应的词性特征值,由自然语言处理工具 NLTK^① 提取得到。 $P \in \mathbb{R}^{n \times m_2}$ 为句子的词性特征嵌入矩阵,则特征值 p_i 对应的词性特征嵌入向量 $P_i \in P$ 可表示如式(3)所示。

$$P_i = W_p e_i \quad (3)$$

同理, $W_p \in \mathbb{R}^{m_2 \times |V|}$ 为可训练矩阵,在训练中得到学习和更新,对于未登录词以及包含对象的词,其词性特征设置为字符 '0'。

位置特征嵌入 在对象级情感分类任务中,距离对象越近的词越有可能为描述对象的词,其所在的位置也就越重要。如图 3 所示,一般取词与对象的相对位置作为位置特征,但是该方法使情感词“horrible”相对于对象的距离较非情感词“was”和“so”要远。

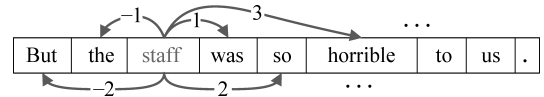


图3 无位置重排

针对该问题,本文提出一种位置重排算法对位置进行重新排序,算法表示如下:

输入: 位置特征向量 (l_1, l_2, \dots, l_n) ;

输出: 重排后的特征向量 $(l'_1, l'_2, \dots, l'_n)$ 。

(1) 根据词的相对位置及其情感得分计算位置得分向量 $score$,如式(4)所示。

$$score_i = \frac{\exp(l_i)}{\sum_{j=0}^n \exp(l_j)} + \frac{\exp(s_i)}{\sum_{j=0}^n \exp(s_j)} \quad (4)$$

其中, $score_i$ 为第 i 个词对应的位置得分,由 softmax 函数对该词的原位置值以及情感得分进行归一化,两部分相加得到。该操作相当于在考虑原位置的基础上引入了情感得分因素,使情感词能够分配到更大的位置得分。

(2) 根据位置得分计算新的位置值,如式(5)所示。

$$l'_i = f(l_i) \text{top_k}(score_i, score),$$

$$f(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases} \quad (5)$$

其中, top_k 为输出 $score_i$ 在所有位置得分当中排在第几位的函数,函数 f 值取 -1 表示词出现在对象的左边,取 1 表示词出现在对象的右边,取 0 则表示对象本身。

如图 4 所示,经过计算,“horrible”的位置特征值由 3 变为 1,意味着“horrible”处在更重要的位置上。设该特征嵌入的维度为 m_3 , $L \in \mathbb{R}^{n \times m_3}$ 为句子的位置特征嵌入矩阵,则重排后的位置特征值 l'_i 对应的位置特征嵌入向量 $L_i \in L$ 可表示如式(6)所示。

$$L_i = W_L e_i \quad (6)$$

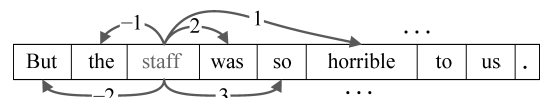


图4 有位置重排

① <http://www.nltk.org/>

同上, $\mathbf{W}_L \in \mathbb{R}^{m_3 \times |V|}$ 可训练且会在训练中学习和更新, 对于包含对象的词, 其位置特征设置为 0。

2.2 卷积层

根据对象在句子中的位置将句子划分为 up 和 down 两个部分。假设 k 为对象在句子中的位置, $\mathbf{X}_{\text{up}} = \{x_1, x_2, \dots, x_k\}$, $\mathbf{X}_{\text{down}} = \{x_k, x_{k+1}, \dots, x_n\}$ 为句子上下文词嵌入矩阵, 类似的 \mathbf{S}_{up} 、 \mathbf{S}_{down} 、 \mathbf{P}_{up} 、 \mathbf{P}_{down} 、 \mathbf{L}_{up} 、 \mathbf{L}_{down} 分别对应由对象划分出的辅助特征嵌入矩阵。为融合多个特征, 模型将词嵌入矩阵与多个辅助特征嵌入矩阵拼接到一起, 得到矩阵 \mathbf{U} 和矩阵 \mathbf{D} , 拼接之后矩阵中词向量的维度为 m 。在卷积操作过程中, up 和 down 两个部分分别由 3 种不同长度的卷积过滤器以提取多种不同的局部特征, 将其表示为 $\mathbf{W}_{\text{up}}, \mathbf{W}_{\text{down}} \in \mathbb{R}^{w \times m \times n_{\text{conv}}}$ 。卷积操作可表示如式(7)~式(10)所示。

$$\mathbf{U} = [\mathbf{X}_{\text{up}} \oplus \mathbf{S}_{\text{up}} \oplus \mathbf{P}_{\text{up}} \oplus \mathbf{L}_{\text{up}}]^T \quad (7)$$

$$\begin{aligned} C_{\text{up}}(i, j) &= f(\mathbf{U} * \mathbf{W}_{\text{up}}(:, :, j) + \mathbf{b}_1(j)), \\ i &\in [1, (k-1) - w + 1], \\ j &\in [1, n_{\text{conv}}] \end{aligned} \quad (8)$$

$$\begin{aligned} \mathbf{D} &= [\mathbf{X}_{\text{down}} \oplus \mathbf{S}_{\text{down}} \oplus \mathbf{P}_{\text{down}} \oplus \mathbf{L}_{\text{down}}]^T \quad (9) \\ C_{\text{down}}(i, j) &= f(\mathbf{D} * \mathbf{W}_{\text{down}}(:, :, j) + \mathbf{b}_2(j)), \\ i &\in [1, (n-k-1) - w + 1], \\ j &\in [1, n_{\text{conv}}] \end{aligned} \quad (10)$$

其中, $*$ 为两个矩阵元素对应相乘相加操作, f 表示非线性激活函数 ReLU, w 为卷积过滤器的长度, 其值表示每次提取局部特征对应词的个数, “ \oplus ”为矩阵拼接操作, $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^{n_{\text{conv}}}$ 为偏置向量, n_{conv} 为使用到的过滤器的数量, 经过卷积操作后输出结果分别为 \mathbf{C}_{up} 和 \mathbf{C}_{down} 。

2.3 分段池化层

为了得到包含对象上下文更加细粒度的特征, 本文将 up 和 down 两个部分卷积操作的结果分别做最大池化操作, 即分段池化以得到包含上下文特征的结果 \mathbf{z}_{up} 和 \mathbf{z}_{down} , 该过程表示如式(11)、式(12)所示。

$$\mathbf{z}_{\text{up}}(j) = \max(\mathbf{C}_{\text{up}}(:, j)), j \in [1, n_{\text{conv}}] \quad (11)$$

$$\mathbf{z}_{\text{down}}(j) = \max(\mathbf{C}_{\text{down}}(:, j)), j \in [1, n_{\text{conv}}] \quad (12)$$

2.4 注意力计算

Seo 等人^[19]对一条电影评论每次取 5 个词做卷积操作, 输出的结果作为中心词的注意力得分, 可将此得分作为该词在句子中重要程度的依据。考虑到描述对象的词会集中在句子中的某个部分, 则该部

分所包含的词应赋予更大的注意力权重, 而且使用卷积操作计算注意力的方式可并行, 不影响整个模型的训练, 因此, 本文采用该注意力计算方式。设 x_i 为中心词, w_a 为卷积窗口长度, 在一个窗口内使用参数矩阵 $\mathbf{W}_{\text{att}}^1 \in \mathbb{R}^{w_a \times d}$ 和偏置项 b_{att}^1 计算每个词的注意力得分, 计算过程如式(13)、式(14)所示。

$$\mathbf{X}_{\text{att}, i} = (x_{i+\frac{-w_a+1}{2}}, x_{i+\frac{-w_a+3}{2}}, \dots, x_i, \dots, x_{i+\frac{w_a-1}{2}})^T \quad (13)$$

$$s_i = g(\mathbf{X}_{\text{att}, i} * \mathbf{W}_{\text{att}}^1) + b_{\text{att}}^1 \quad (14)$$

其中, $g(\cdot)$ 为 sigmoid 函数, s_i 即为句中第 i 个词的注意力得分, 使用 s_i 对词向量进行加权得到新的词向量表示 x'_i , 并对加权后的向量采取求和操作, 如式(15)、式(16)所示。

$$x'_i = s_i x_i, i \in [1, n] \quad (15)$$

$$y = \sum_{i=1}^n x'_i \quad (16)$$

最后由参数矩阵 $\mathbf{W}_{\text{att}}^2 \in \mathbb{R}^{d \times n_a}$ 和偏置项 $b_{\text{att}}^2 \in \mathbb{R}^{n_a}$ 计算得到注意力表示 \mathbf{z}_{att} , 如式(17)所示。

$$\begin{aligned} \mathbf{z}_{\text{att}}(i) &= g(y * \mathbf{W}_{\text{att}}^2(:, i) + \mathbf{b}_{\text{att}}^2(i)), \\ i &\in [1, n_a] \end{aligned} \quad (17)$$

其中, $*$ 为两个矩阵元素对应相乘相加操作, n_a 为过滤器的数量, $g(\cdot)$ 为 tanh 函数。

2.5 softmax 层

设 $\#n_w$ 为模型所用卷积过滤器的类型, 其值指示一个特定长度的卷积过滤器, 若模型中使用到 3 个不同长度的过滤器, 则 $\#n_w$ 的取值为 $[1, 2, 3]$ 。将池化操作以及上下两个部分单词注意力计算结果 $\mathbf{z}_{\text{att_up}}$ 、 $\mathbf{z}_{\text{att_down}}$ 进行串联, 以丰富句子的特征表示, 得到结果 \mathbf{z}_{out} , 将其交给 softmax 分类器输出情感分类的概率 P , 过程表示如式(18)、式(19)所示。

$$\begin{aligned} \mathbf{z}_{\text{out}} &= [\mathbf{z}_{\text{up}}^1 \oplus \mathbf{z}_{\text{up}}^2 \oplus \dots \oplus \mathbf{z}_{\text{up}}^{\#n_w} \\ &\oplus [\mathbf{z}_{\text{down}}^1 \oplus \mathbf{z}_{\text{down}}^2 \oplus \dots \oplus \mathbf{z}_{\text{down}}^{\#n_w} \\ &\oplus \mathbf{z}_{\text{att_up}} \oplus \mathbf{z}_{\text{att_down}}] \end{aligned} \quad (18)$$

$$P = \text{softmax}(\mathbf{W}_p \mathbf{z}_{\text{out}} + \mathbf{b}_p) \quad (19)$$

其中, “ \oplus ”为向量连接操作, \mathbf{W}_p 和 \mathbf{b}_p 为权重参数和偏置参数。

2.6 损失函数

在模型反向传播过程中, 本文采用基于 Adam 的 mini-batch 梯度下降法, 通过最小化交叉熵来优化模型, 交叉熵损失函数如式(20)所示。

$$\text{Loss} = - \sum_{i=1}^D \sum_{j=1}^C \hat{P}_i^j \log P_i^j + \lambda \|\theta\|^2 \quad (20)$$

其中, D 为训练数据集大小, C 为情感类别数量, 其值为 3, \hat{P} 为数据集中样本的实际类别分布, P 为模型输出的预测类别分布, $\lambda \|\theta\|^2$ 为 L_2 正则项。

3 实验

本文将在 3 个不同领域的对象级情感分类数据集上进行实验, 并将实验结果与现有模型进行比较, 验证了本文提出的模型具备良好的分类性能。其中, 实验在如表 1 所示的环境配置下完成。

表 1 实验环境配置

实验环境	环境配置
操作系统	Windows 10
CPU	i5-8300H
GPU	GTX-1050Ti
内存	16 GB
深度学习框架	TensorFlow 1.8.0

3.1 数据集

本文使用的数据集包括 SemEval 2014 提供的 Restaurant 和 Laptop 两个领域的用户评论, 以及 Twitter 提供的用户推文数据集。数据中包含对象词或对象词组, 对象的情感极性被分为消极、中性以及积极。数据集被划分为训练集和测试集, 3 个情感类别在各数据集上的数量统计结果如表 2 所示。

表 2 实验数据集情感类别统计

数据集	消极	中性	积极
Restaurant-train	805	633	2164
Restaurant-test	196	196	728
Laptop-train	866	460	987
Laptop-test	128	169	341
Twitter-train	1 411	2 826	1 411
Twitter-test	173	346	173

3.2 参数设置

在模型的输入层, 本文使用 GloVe 预训练英文词向量作为模型的输入之一, 维度 d 设置为 300, 对于未登录词, 其词向量用 0 向量初始化, 三个可训练

特征矩阵使用均匀分布 $U(-1, 1)$ 随机初始化; 在多特征嵌入当中, 经过多次实验发现, 嵌入维度对模型的影响不大, 考虑到时间性能因素, 本文将 3 个特征嵌入的维度 m_1, m_2, m_3 统一设置为较低的 50 维; 在卷积层和注意力计算过程中, 模型均采用了多个卷积过滤器进行卷积计算; 为防止模型过拟合, 模型在池化层和优化损失函数过程中分别采用了 dropout 和 L_2 正则化策略, 其中模型的参数设置如表 3 所示。

表 3 参数设置

参数	参数描述	参数值
w	卷积层窗口长度	[3, 4, 5]
n_{conv}	特征图数量	100
w_a	注意力卷积窗口长度	5
n_a	注意力特征图数量	400
λ	L_2 正则化系数	0.5
d_r	dropout 比率	0.5

3.3 对比实验

实验结果将和以下 10 种不同类型的方法进行对比。

- (1) **SVM**^[20]: 一种基于可用特征的传统机器学习算法支持向量机 (support vector machine, SVM) 分类模型;
- (2) **SVM+lexicons**: 在上述 SVM 的基础上引入了多个情感词典优化的模型;
- (3) **TD-LSTM**^[6]: 根据对象将句子划分成两个部分, 利用 LSTM 提取上下文特征进行情感分类;
- (4) **ATAE-LSTM**^[7]: 将对象和其上下文表示连接到一起, 在 LSTM 的基础上引入注意力机制计算每一个词的注意力权重而提出的情感分类模型;
- (5) **IAN**^[21]: 利用两个 LSTM 对上下文和对象进行建模, 并用注意力机制交互地学习上下文和对象中的关键特征表示, 最后将得到的表示连接到一起进行情感分类;
- (6) **TextCNN**^[22]: 基于 CNN 提出的最基本的文本分类模型;
- (7) **Att-CNN**^[23]: 在 CNN 基础上添加多层注意力的文本分类模型;
- (8) **CATT-CNN**^[9]: 结合词向量、词性、位置信息计算注意力得分, 提出的基于多注意力 CNN 对象级情感分类模型;

(9) **GCAE**^[10]: 一种基于 CNN 和门控机制对象级情感分类模型。其中, 门控机制 \tanh -ReLU 单元可以根据给定的对象选择性地输出情感特征;

(10) **Mul-AT-CNN**^[24]: 一种利用多层卷积神经网络对上下文进行多次建模, 并使用注意力机制显式学习情感中间表示, 最后将所有特征集成到一起进行情感分类。

本文提出的模型如下:

(1) **MP-CNN**: 融合多特征的分段卷积神经网络模型, 该模型能够根据对象在句子中的位置划分句子, 在模型的池化层采用分段最大池化操作, 提取包含对象上下文信息的细粒度特征。

(2) **MP-CNN-2**: 受 CNN 在图像处理中应用的启发, 本文认为更深层的 CNN 能够学习到更多有利于分类的特征, 实验发现 2 层 CNN 效果最优, 即 MP-CNN-2。

3.4 实验结果

本文提出的两个模型 MP-CNN 和 MP-CNN-2 在 3 个数据集上都取得了不错的分类准确率, 实验结果如表 4 所示。

表 4 不同模型分类准确率(%)

模型	Restaurant	Laptop	Twitter
SVM	77.13	63.61	—
SVM+lexicons	80.16	70.49	—
TD-LSTM	75.63	68.13	66.73
ATAE-LSTM	77.20	68.70	—
IAN	78.60	72.10	—
TextCNN	66.33	58.32	—
Att-CNN	75.89	65.42	—
CATT-CNN	79.17	67.10	—
GCAE	79.67	69.14	—
Mul-AT-CNN	79.46	75.39	71.25
MP-CNN	80.42	72.92	71.72
MP-CNN-2	81.25	74.56	72.81

首先, 对比基于传统机器学习算法的两个模型 SVM 和 SVM+lexicons, 引入情感词典的 SVM+lexicons 模型较 SVM 模型在 Laptop 数据集上效果提升了 6.88%, 这进一步说明了引入辅助特征的有效性。因为深度学习模型具有强大自动提取特征的能力, 并且又添加了更多的辅助特征, 所以与需要手

工设计特征且只引入了情感词典的 SVM+lexicons 模型相比, 本文表现较好的模型 MP-CNN-2 取得了更好的实验结果, 在 Restaurant 和 Laptop 数据集上效果分别提升了 1.09% 和 4.07%。

其次, 对比基于 LSTM 的模型。其中, 添加注意力机制的 ATAE-LSTM 和 IAN 模型与只考虑对象上下文信息的 TD-LSTM 模型相比, 效果都有一定程度的提升, 这说明注意力机制在对象级情感分类当中起到一定程度上的改进作用, 但这 3 种方法都没有考虑额外的辅助特征, 而引入 3 个辅助特征的 MP-CNN-2 模型具有更好的特征学习能力。因此, 分类效果与基于 LSTM 的模型中表现较好的 IAN 模型相比, 在 Restaurant 和 Laptop 数据集上的分类准确率分别提升了 2.65% 和 2.46%。

最后, 对比基于 CNN 的分类模型。较最基本的模型 textCNN 来说, 因未考虑对象的上下文且未引入任何辅助特征, 捕获特征的能力不足, 所以分类结果较差, 这也说明了本文分段提取对象上下文特征的有效性。Att-CNN、CATT-CNN 以及本文提出的两个模型都分别在 CNN 模型的基础上使用了注意力机制, 较未使用注意力机制的 TextCNN 模型分类效果都有所提高, 所以将注意力机制应用于 CNN 上亦可改进分类效果, 但这两个模型在池化层都采用最大池化操作, 并未考虑更细粒度的上下文特征, 所以本文提出的两个模型表现均优于这两个模型。和 GCAE 模型相比, MP-CNN-2 在 Restaurant 和 Laptop 数据集上分别取得了 1.58% 和 5.42% 的提升, 说明上下文特征结合多个辅助特征要比单一的门控机制表现力强; Mul-AT-CNN 模型和本文提出的两个模型均考虑了对象的上下文信息, 但建模方式不同, 由于本文考虑了多个辅助特征, 使模型更具特征表现力, 所以在 Restaurant 和 Twitter 数据集上表示更为出色, 分别提高了 1.79% 和 1.56%。而 Mul-AT-CNN 在 Laptop 数据集上领先 0.83%, 这反映了该模型的上下文建模方式在 Laptop 数据集上有着更好的效果, 但却不能泛化到其他数据集上。

3.5 实验分析

为验证本文方法的有效性, 下面将针对位置重排算法、注意力机制以及辅助特征设计分析实验, 描述如下。

位置重排算法验证实验 在进行位置嵌入时, 将模型设置为有位置重排和无位置重排进行对比实

验,实验结果如图 5(a)和图 5(b)所示。其中 with 表示有位置重排,without 表示无位置重排,从图的对比中可以看出,本文提出的两个模型 MP-CNN 和 MP-CNN-2,使用位置重排模型的分类效果在所有

数据集上均优于未使用该算法的模型,这是因为重排后的位置使得情感词处在更重要的位置上,因此模型学习到了更有利于情感分类的位置特征,所以位置重排算法在一定程度上能够改进模型的分类效果。

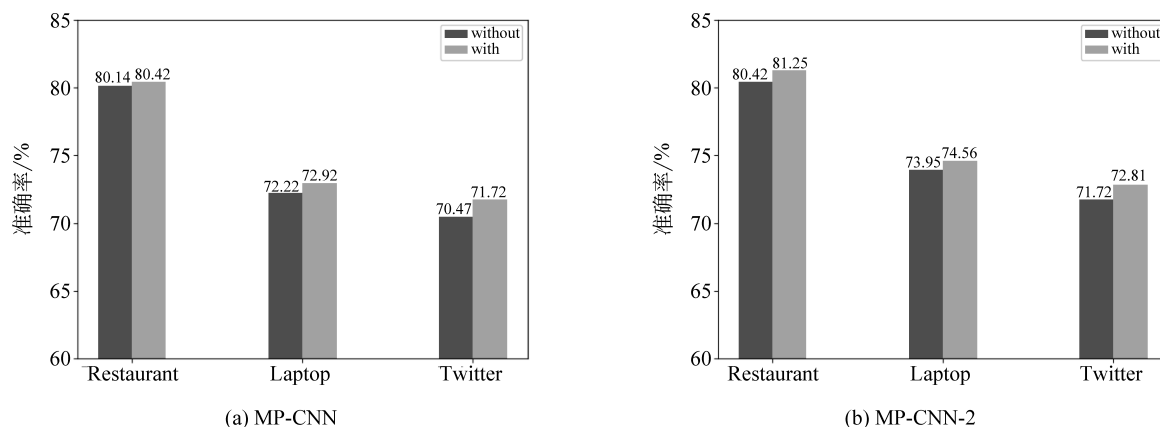


图 5 辅助特征验证实验结果

注意力值可视化实验 为验证注意力机制能够关注到与描述对象有关的词,我们在数据集上分别随机挑选两个句子进行测试,输出各个词的注意力得分,并对其进行可视化,结果如图 6 所示,对于第

一个句子,描述对象“bread”的单词“is”“top”“notch”以及第二个句子描述对象“hardware”的单词“was”“slow”“locked”等都被赋予了较大的权重,这说明注意力机制在一定程度上发挥积极作用。

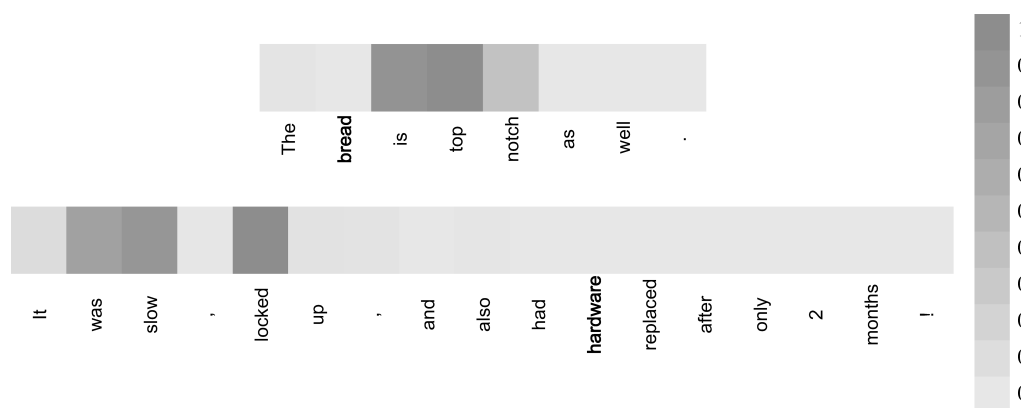


图 6 注意力可视化

辅助特征验证实验 为验证在本文提出的模型上添加辅助特征的有效性,实验首先将所有辅助特征从模型中剔除,然后再逐个添加,实验结果如图 7 所示。其中,“0”表示剔除所有辅助特征;“1”表示只包含一个辅助特征,选取实验中 3 个辅助特征得到的最大准确率作为该实验的结果;“2”表示包含两个特征,3 个辅助特征共有 3 种不同的组合方式,亦选 3 种组合中取得的最大准确率作为该实验的结果;“3”为添加所有辅助特征。可见,随着特征的添加,分类的准确率在 3 个领域的数据集上都呈现上升趋势,这是因为随着辅助特征的增多,使模型

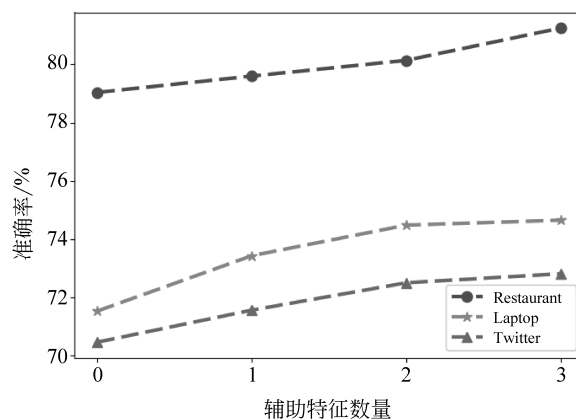


图 7 辅助特征验证实验结果

能够提取到更多有利于情感分类的特征信息,因此特征表现力就越强,所以引入辅助特征能够改进模型的效果。

4 总结和展望

本文提出了融合多特征的分段卷积神经网络模型 MP-CNN 和 MP-CNN-2,考虑到由对象划分出句子上下文的重要性,选择分段最大池化策略提取更加细粒度的对象上下文特征。此外,根据情感分类的特点,引入多个辅助特征,将词的位置、词性以及词在情感词典中的情感得分特征融合进模型,增强了模型的特征表现力。最后模型采用卷积的方式计算词的注意力得分,帮助改善模型的分类效果。

虽然本文提出的两个模型在 3 个不同领域的对象级情感分类数据集上均取得了较好的分类结果,但多个辅助特征+注意力的句子表示方式是否可以泛化到其他更细粒度的情感分类领域仍然未知,下一步的工作重点将是验证该组合方式在更细粒度的分类任务上的有效性。

参考文献

- [1] 李然,林政,林海伦,等. 文本情绪分析综述[J]. 计算机研究与发展, 2018, 55(1): 30-52.
- [2] Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8(4): e1253.
- [3] Plank B, Søgaard A, Goldberg Y. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016: 412.
- [4] Yu A W, Dohan D, Luong M T, et al. QANet: Combining local convolution with global self-attention for reading comprehension[J]. arXiv preprint arXiv: 1804.09541, 2018.
- [5] Gehring J, Auli M, Grangier D, et al. Convolutional sequence to sequence learning[C]//Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017: 1243-1252.
- [6] Tang D, Qin B, Feng X, et al. Effective LSTMs for target-dependent sentiment classification [C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics; Technical Papers, 2016: 3298-3307.
- [7] Wang Y, Huang M, Zhao L. Attention-based LSTM for aspect-level sentiment classification[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016: 606-615.
- [8] Yin W, Kann K, Yu M, et al. Comparative study of CNN and RNN for natural language processing[J]. arXiv preprint arXiv: 1702.01923, 2017.
- [9] 梁斌,刘全,徐进,等. 基于多注意力卷积神经网络的特定目标情感分析[J]. 计算机研究与发展, 2017, 54(8): 1724-1735.
- [10] Xue W, Li T. Aspect based sentiment analysis with gated convolutional networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 2514-2523.
- [11] Matsumoto K, Ren F. Estimation of word emotions based on part of speech and positional information[J]. Computers in Human Behavior, 2011, 27(5): 1553-1564.
- [12] 陈钊,徐睿峰,桂林,等. 结合卷积神经网络和词语情感序列特征的中文情感分析[J]. 中文信息学报, 2015, 29(6): 172-178.
- [13] 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, 2015: 1753-1762.
- [14] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv: 1301.3781, 2013.
- [15] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality [C]//Proceedings of the Advances in Neural Information Processing Systems, 2013: 3111-3119.
- [16] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014: 1532-1543.
- [17] Bojanowski P, Grave E, Joulin A, et al. Enriching word vectors with subword information[J]. Transactions of the Association for Computational Linguistics, 2017, 5: 135-146.
- [18] Mohammad S M, Kiritchenko S, Zhu X. NRC-Canada: Building the state-of-the-art in sentiment analysis of Tweets [J]. arXiv preprint arXiv: 1308. 6242, 2013.
- [19] Seo S, Huang J, Yang H, et al. Representation learning of users and items for review rating prediction using attention-based convolutional neural network [C]//Proceedings of the 3rd International Workshop on Machine Learning Methods for Recommender Systems (MLRec)(SDM'17), 2017.

(下转第 132 页)

- [17] Xu W, Tan Y. Semi-supervised rarget-level sentiment analysis via variational autoencoder [J/OL]. arXiv preprint arXiv: 1810.10437, 2018.
- [18] 周健, 田莹, 崔晓晖. 基于改进 Sequence-to-Sequence 模型的文本摘要生成方法[J]. 计算机工程与应用, 2019, 55(1): 128-134.
- [19] 谢志峰, 吴佳萍, 马利庄. 基于卷积神经网络的中文财经新闻分类方法[J]. 山东大学学报(工学版), 2018, 48(03): 38-43, 70.
- [20] Sennrich R, Haddow B, Birch A. Neural machine translation of rare words with subword units[J/OL]. arXiv preprint arXiv: 1508.07909, 2015.
- [21] Kim Y. Convolutional neural networks for sentence classification [C]//Proceedings of EMNLP 2014 Conference on Empirical Methods in Natural Language Processing, 2014: 1746-1751.
- [22] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Proceeding of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA. 2019: 4171-4186.



杜朋(1996—), 硕士研究生, 主要研究领域为自然语言处理。

E-mail: 18810911898@163.com



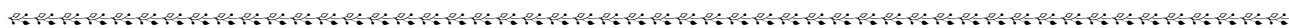
卢益清(1980—), 通信作者, 博士, 副教授, 主要研究领域为商务计算、供应链管理。

E-mail: luyiqing@126.com



韩长风(1996—), 硕士研究生, 主要研究领域为供应链和博弈论。

E-mail: 15137092320@163.com



(上接第 124 页)

- [20] Kiritchenko S, Zhu X, Cherry C, et al. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews [C]//Proceedings of the 8th International Workshop on Semantic Evaluation, 2014: 437-442.
- [21] Ma D, Li S, Zhang X, et al. Interactive attention networks for aspect-level sentiment classification [C]//Proceedings of the International Joint Conference on Artificial Intelligence, 2017: 4068-4074.
- [22] Kim Y. Convolutional neural networks for sentence classification [J]. arXiv preprint arXiv: 1408.5882, 2014.
- [23] Wang L, Cao Z, de Melo G, et al. Relation classification via multi-level attention CNNs [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, 1: 1298-1307.
- [24] Zhang S, Xu X, Pang Y, et al. Multi-layer attention based CNN for target-dependent sentiment classification [J]. Neural Processing Letters, 2019: 1-15.



周武(1995—), 硕士研究生, 主要研究领域为自然语言处理、情感分析。

E-mail: zwersir@gmail.com



曾碧卿(1969—), 通信作者, 博士, 教授, 硕士生导师, 主要研究领域为人工智能与自然语言处理。

E-mail: zengbinqing0528@163.com



徐如阳(1995—), 硕士研究生, 主要研究领域为自然语言处理、自动文本摘要。

E-mail: taeyanglucky@qq.com