

基于多注意力卷积神经网络的特定目标情感分析

梁斌¹ 刘全^{1,2,3} 徐进¹ 周倩¹ 章鹏¹

¹(苏州大学计算机科学与技术学院 江苏苏州 215000)

²(软件新技术与产业化协同创新中心 南京 210000)

³(符号计算与知识工程教育部重点实验室(吉林大学) 长春 130012)

(bliang@stu.suda.edu.cn)

Aspect-Based Sentiment Analysis Based on Multi-Attention CNN

Liang Bin¹, Liu Quan^{1,2,3}, Xu Jin¹, Zhou Qian¹, and Zhang Peng¹

¹(College of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215000)

²(Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing 210000)

³(Key Laboratory of Symbolic Computation and Knowledge Engineering (Jilin University), Ministry of Education, Changchun 130012)

Abstract Unlike general sentiment analysis, aspect-based sentiment classification aims to infer the sentiment polarity of a sentence depending not only on the context but also on the aspect. For example, in sentence “The food was very good, but the service at that restaurant was dreadful”, for aspect “food”, the sentiment polarity is positive while the sentiment polarity of aspect “service” is negative. Even in the same sentence, sentiment polarity could be absolutely opposite when focusing on different aspects, so we need to infer the sentiment polarities of different aspects correctly. The attention mechanism is a good way for aspect-based sentiment classification. In current research, however, the attention mechanism is more combined with RNN or LSTM networks. Such neural network-based architectures generally rely on complex structures and cannot parallelize over the words of a sentence. To address the above problems, this paper proposes a multi-attention convolutional neural networks (MATT-CNN) for aspect-based sentiment classification. This approach can capture deeper level sentiment information and distinguish sentiment polarity of different aspects explicitly through a multi-attention mechanism without using any external parsing results. Experiments on the SemEval2014 and Automotive-domain datasets show that, our approach achieves better performance than traditional CNN, attention-based CNN and attention-based LSTM.

Key words attention mechanism; convolutional neural networks; aspect-based sentiment analysis; deep learning; natural language processing

收稿日期:2017-03-20;修回日期:2017-05-18

基金项目:国家自然科学基金项目(61272005,61303108,61373094,61472262,61502323,61502329);江苏省自然科学基金项目(BK2012616);江苏省高校自然科学研究项目(13KJB520020);吉林大学符号计算与知识工程教育部重点实验室基金项目(93K172014K04)

This work was supported by the National Natural Science Foundation of China (61272005, 61303108, 61373094, 61472262, 61502323, 61502329), the Natural Science Foundation of Jiangsu Province (BK2012616), the High School Natural Science Foundation of Jiangsu Province (13KJB520020), and the Foundation of the Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education (Jilin University) (93K172014K04).

通信作者:刘全(quanliu@suda.edu.cn)

摘要 特定目标情感分析作为情感分析一个重要的子任务,近年来得到越来越多研究人员的关注.针对在特定目标情感分析中,将注意力机制和 LSTM 等序列性输入网络相结合的网络模型训练时间长、且无法对文本进行平行化输入等问题,提出一种基于多注意力卷积神经网络(multi-attention convolution neural networks, MATT-CNN)的特定目标情感分析方法.相比基于注意力机制的 LSTM 网络,该方法可以接收平行化输入的文本信息,大大降低了网络模型的训练时间.同时,该方法通过结合多种注意力机制有效弥补了仅仅依赖内容层面注意力机制的不足,使模型在不需要例如依存句法分析等外部知识的情况下,获取更深层次的情感特征信息,有效识别不同目标的情感极性.最后在 SemEval2014 数据集和汽车领域数据集(automotive-domain data, ADD)进行实验,取得了比普通卷积神经网络、基于单注意力机制的卷积神经网络和基于注意力机制的 LSTM 网络更好的效果.

关键词 注意力机制;卷积神经网络;特定目标情感分析;深度学习;自然语言处理

中图分类号 TP391

随着社交网络的兴起,越来越多的人在网络上发表意见和表达情感,通过对这些数据的提取和分析可以帮助我们了解不同人对不同事物所发表的观点和情感.如何利用自然语言处理技术来分析社交网络短文本的情感倾向,已经成为研究人员关注的热点领域之一^[1-2].

近年来,随着深度学习在情感分析领域的研究和应用,越来越多的研究人员使用深度学习来对社交网络短文本进行情感分析.例如 Kalchbrenner 等人用卷积神经网络(convolution neural networks, CNN)解决 Twitter 的极性判断问题^[3];Zhou 等人使用 LSTM(long short-term memory)网络解决跨语言情感分析任务^[4];Yu 等人使用多种深度神经网络解决跨领域的情感分析任务^[5].这些使用深度神经网络的方法在情感分析任务中取得了比以往研究更好的分类效果.

特定目标情感分析(aspect-based sentiment analysis, ABSA)作为情感分析一个重要的子任务,是更深层次的情感分析^[6].和普通情感分析不同,特定目标情感极性的判别不仅依赖文本的上下文信息,同时还依赖特定目标的特征信息^[7].例如句子“The food was very good, but the service at that restaurant was dreadful”对于目标“food”是积极情感,而对于目标“service”则是消极情感.所以,就算是同一个句子,针对不同的目标可能会出现完全相反的情感极性.

注意力机制(attention mechanism)能很好解决同一句子中不同目标的情感极性判别问题.注意力机制最早是用在图像处理领域,目的是为了神经网络在处理数据时重点关注某些信息.Mnih 等人在 2014 年提出将注意力机制用在图像分类任务中^[8],

验证了注意力机制在图像处理领域的有效性,同时也使结合注意力机制的神经网络成为研究的热点.随后,Bahdanau 等人将注意力机制和循环神经网络(recurrent neural network, RNN)结合解决机器翻译任务^[9],使注意力机制成功融入自然语言处理领域.随着研究的推进,Yin 等人在 2015 年提出一种基于注意力机制的卷积神经网络,并将该网络用在句子建模任务中^[10];Wang 等人利用基于多层注意力机制的卷积神经网络进行句子关系分类^[11].这些方法的提出证明了注意力机制和 CNN 结合的有效性.

基于注意力机制的卷积神经网络可以通过一个注意力矩阵计算出句子相应的注意力特征,在模型训练时让神经网络高度关注某些信息.例如在特定目标情感分析任务中,注意力机制可以通过注意力特征图让模型在训练过程中关注目标词的特征信息,正确识别不同目标的情感极性.相比 LSTM 等序列性神经网络,卷积神经网络可以获取平行输入的信息,同时可以解决传统 LSTM 网络不能准确表示每个词在句子中的重要程度的缺点.为了更好地学习不同目标的情感信息和识别不同目标的情感极性,本文提出一种结合多注意力机制的卷积神经网络应用到特定目标的情感分析方法.该方法基于卷积神经网络,使模型能接收平行输入的文本信息.同时,该方法在以前研究的基础上提出了加入词性注意力机制,使模型能从多种特征信息关注句子的目标词和其他词语之间的联系,准确表示出每一个词在句子中的重要程度,获取更多的隐藏信息.

本文结合 3 种注意力机制来构造多注意力卷积神经网络(multi-attention convolution neural networks, MATT-CNN)模型:1)词向量注意力机制.将目标词的词向量和文本其他词的词向量进行运算,突出

目标词的重要性;2)词性注意力机制.将文本中的情感词和目标词重新进行词性标注,使网络充分利用文本的情感信息;3)位置注意力机制.以目标词为中心,计算其他词和目标词的位置关系,有效表示每个词在句子中的重要程度和对目标词的影响程度.最后,将本文提出的方法在 SemEval2014 和 ADD 两个数据集上进行实验,取得了比普通卷积神经网络、基于单注意力机制的卷积神经网络和基于注意力机制的 LSTM 网络更好的情感分类效果,验证了本文提出方法的有效性.

本文的主要贡献如下:

1) 提出一种多注意力卷积神经网络(MATT-CNN)用在特定目标情感分类任务中,和基于注意力机制的 LSTM 网络相比,MATT-CNN 可以接收句子的平行化输入,大大降低了模型的训练时间;

2) 本文提出的词性注意力机制可以让模型充分利用和学习输入文本的情感信息,同时可以弥补仅仅依赖内容层面注意力机制的不足;

3) 提出一个双向扫描算法,能有效标识不同词语在句子中的重要程度,使卷积神经网络可以充分利用文本每个词语的位置信息;

4) 多注意力卷积神经网络模型结构简单,无需使用例如依存句法分析和语义依存分析等外部知识,同时也不需要针对特定目标进行额外的向量化;

5) 在不同语言 and 不同领域的数据集上的实验结果表明,本文提出的基于多注意力卷积神经网络能有效识别不同目标的情感极性.

1 相关工作

1.1 特定目标情感分析

特定目标情感分析是通过学习文本上下文的信息来判别文本中特定目标的情感极性,是更深层次的情感分析,也是细粒度的文本分类任务,一直以来都得到很多学者的关注和研究.2004年,Hu 和 Liu 提出了基于不同商品的用户评论情感分析^[12-13]; Qiu 等人使用目标提取和情感词典扩充方法判别特定目标的情感极性^[14]; Kiritchenko 等人使用机器学习方法进一步研究特定商品的用户评论进行情感分析^[15].这些使用传统机器学习的方法通过构造情感特征,能很好地利用文本的情感特征信息,有效判别特定目标的情感极性.

近年来,深度学习技术在文本分类任务中取得了重大的突破,也有很多学者利用深度学习技术来

解决特定目标情感分析问题. Nguyen 和 Shirai 基于 RNN 和依存树提出一种特定目标情感分析模型^[16]; Dong 等人提出一种自适应的 RNN 用在特定目标的 Twitter 文本情感分类任务中^[17].这些方法使用深度学习技术来解决特定目标情感分析问题,可以在不需要情感词典的情况下通过深度神经网络更好地学习文本的情感特征信息,取得了比传统机器学习方法更好的分类效果.但是这些方法需要使用依存树、句法关系等外部知识,网络模型结构较为复杂. Ruder 等人提出一种分层的双向 LSTM 网络进行特定目标情感分析^[18],该方法利用一个双向的 LSTM 网络在不同层次对输入文本进行特征提取和学习,可以有效学习句子内部关系和句子之间的关系,从而可以判别特定目标的情感极性.但该方法更多的是针对句子中只包含一个特定目标的情况,对包含多个目标的句子情感极性判别较为困难.为了解决不同目标的情感极性判别问题, Wang 等人提出了一种结合注意力机制的 LSTM 网络^[19],通过对特定目标进行特定的向量化处理,将特定目标作为注意力机制输入 LSTM 网络,使网络模型在训练过程中可以高度关注特定的目标,从而可以有效识别不同目标的情感极性.从这些研究工作可以看出,基于 LSTM 网络的神经网络模型已经在特定目标情感分析任务中取得了很多的成功,但是将卷积神经网络应用到特定目标情感分析任务中的研究还很少.此外,基于 LSTM 等序列性输入网络在特定目标情感分析任务中无法接收平行化输入的文本,所以在训练过程中难以准确表示每一个词在句子中的重要程度.

本文提出的基于多注意力卷积神经网络的特定目标情感分析方法可以接收平行化输入的文本信息,通过不同的注意力机制可以准确表示出每一个词在句子中的重要程度.同时,该方法提出了一种可以充分利用文本情感特征信息的多注意力机制,使网络在训练过程中更好地学习不同目标的情感特征信息,有效识别不同目标的情感极性.

1.2 卷积神经网络

在自然语言处理任务中,卷积神经网络(CNN)无需对文本进行大量的预处理工作,有效缓解了特征工程的工作量.如图1所示,CNN主要由输入层、卷积层、池化层和全连接层组成.输入层是对输入数据的向量表示,对于给定的长度为 n 的句子,输入层矩阵可表示为

$$e \in \mathbb{R}^{n \times k}, \quad (1)$$

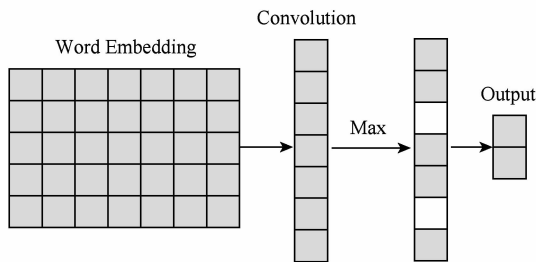


Fig. 1 The overall framework of convolution neural networks

图 1 卷积神经网络模型结构图

其中 k 为词向量维度. 卷积层使用不同的卷积核对输入矩阵进行卷积操作, 提取输入的局部特征, 得到卷积核特征向量图:

$$c = f(W \cdot x + b), \quad (2)$$

其中, x 为卷积核窗口词向量矩阵, W 为权重矩阵, b

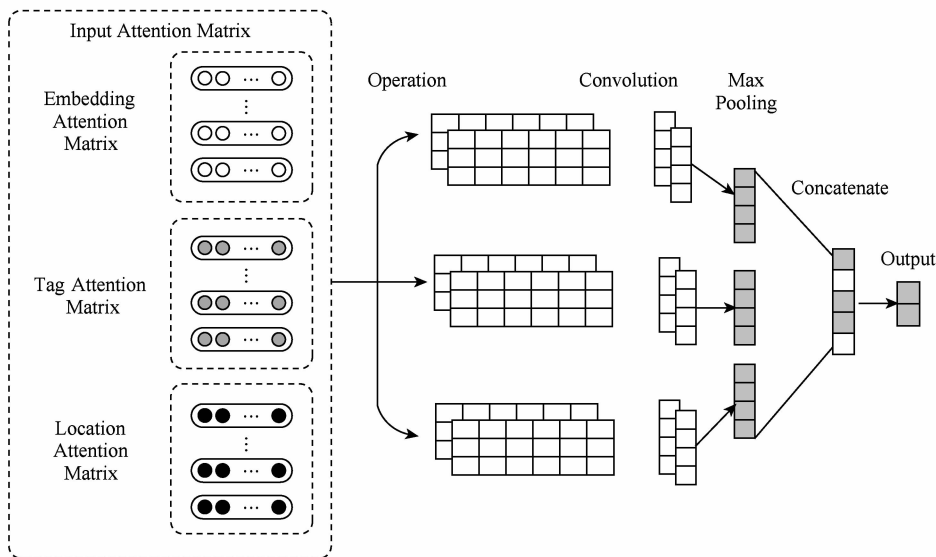


Fig. 2 The overall framework of multi-attention CNN

图 2 多注意力卷积神经网络框架图

1) 注意力输入矩阵. 存储不同注意力机制的注意力特征向量信息;

2) Operation 层. 对于 3 种不同的注意力机制, 本文使用不同的运算操作来获取输入文本的不同表示作为卷积神经网络的输入, 从而使模型在训练过程中从多方面信息来关注特定的目标;

3) 卷积层. 为了使网络模型简单化, 本文使用一层卷积层对不同的注意力机制输入信息进行特征提取, 通过多窗口多注意力机制的结合, 可以获取输入文本丰富的局部特征;

4) 池化层. 对于从不同注意力机制输入信息提取得到的局部特征, 分别进行池化操作, 可以获取不同注意力机制运算的最重要的情感特征信息;

为偏置, f 为激活函数. 池化层是卷积神经网络的重要网络层, 对于卷积层得到的特征向量图, 可以通过池化层来对特征向量图进行下采样操作, 提取重要的特征信息. 同时, 池化层可以输出一个固定大小的矩阵. 不同长度的句子输入和不同大小的卷积核可以通过池化层得到一个相同维度的输出, 并将输出传给全连接层对输入进行分类, 得到分类结果.

2 多注意力卷积神经网络

本文通过 3 种作用在卷积神经网络输入层的注意力机制介绍如何更加充分地学习文本的情感特征信息, 从而有效识别不同目标的情感极性. 如图 2 所示, 本文提出的多注意力卷积神经网络主要由 6 部分组成:

5) 合并层. 本文将从池化层提取得到的不同注意力机制的最重要信息通过合并层操作, 形成输入文本的特征表示, 并将特征表示作为全连接层的输入;

6) 输出层. 本文将通过 softmax 函数输出最终的分类结果.

2.1 任务定义

对于长度为 n 的句子 $s = \{\omega_1, \omega_2, \dots, a_i, \dots, a_j, \dots, \omega_n\}$, 其中 a_i 和 a_j 为 2 个不同的目标词, 本文的任务是针对不同的目标分析其情感极性. 例如句子 “The food was very good, but the service at that restaurant was dreadful”, 对于目标 “food” 来说, 是积极情感极性, 而对于目标 “service” 则是消极

的情感. 本文将句子以词为单位形成一个词序列, 然后将每一个词映射为一个多维的连续值词向量, 得到一个词向量矩阵 $E \in \mathbb{R}^{k \times |V|}$, 其中 k 为词向量维度, 即把每一个词映射为 k 维向量 $x_i \in \mathbb{R}^k$, $|V|$ 为词典的大小, 即数据集包含的所有词的数量. 对于长度为 n 的句子可以表示为式(3)所示的矩阵. 其中, \oplus 为拼接操作. 本文要通过句子中的词向量集合 $\{x_1, x_2, \dots, x_n\}$ 和目标集合 $\{t_i, t_j\}$ 之间的特征信息来判断目标集合 $\{t_i, t_j\}$ 中每一个目标的情感极性.

$$x_{1:n} = x_1 \oplus x_2 \oplus \dots \oplus t_i \oplus \dots \oplus t_j \oplus \dots \oplus x_n. \quad (3)$$

2.2 方法概述

为了更好地学习不同目标的特征信息以及识别不同目标的情感极性, 本文使用 3 种不同的注意力机制让模型以不同的方式注重学习某些需要关注的信息. 对于包含 t 个目标词的句子, 本文针对不同的目标将句子表示为 t 个分句, 并用一个特殊标识“ \langle ”标记目标词的位置, 对句子重新表示, 使得每一个分句只有一个目标词. 该方法能使模型在训练过程中获取目标词的位置, 同时也保留了标志的目标词和其他目标词之间的联系和相互影响.

如图 3 所示, 对于句子“The food was very good, but the service at that restaurant was dreadful”, 本文将特殊标识“ \langle ”标注为特殊的词性, 并映射为一个多维的向量, 可以在训练过程进行调整和学习.

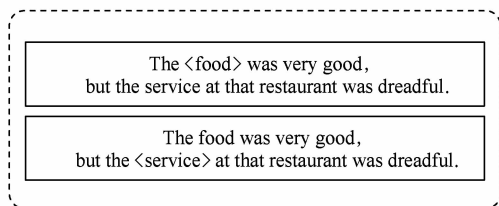


Fig. 3 Example of sentence representation

图 3 句子表示方法

对于长度为 n 的句子 $s = \{\omega_1, \omega_2, \dots, a_i, \dots, \omega_n\}$, 本文将句子中的目标词 a_i 的词性重新标注为“tar”, 特殊标识“ \langle ”的词性重新标注为“tag”, 积极情感词的词性重新标注为“pos”, 消极情感词的词性重新标注为“neg”. 本文通过词向量矩阵将每一个词表示为 k 维的向量, 提取出目标词 a_i 的词向量、词性和 a_i 在句子中的位置这 3 个特征来构造神经网络输入层的 3 种注意力机制:

1) 词向量注意力机制. 提取目标词的词向量作为目标词向量注意力矩阵, 将注意力矩阵和词向量矩阵进行运算操作获取注意力特征矩阵, 是关联文本内容的注意力机制.

2) 词性注意力机制. 是词向量注意力机制的补充, 是关联句子中词语的词性的注意力机制, 可以让模型用另一种方式关注句子的目标词, 从而学习到更多的隐藏信息.

3) 位置注意力机制. 本文使用的位置注意力有 2 种形式, 一种形式是作为参数加入词向量注意力机制和词性注意力机制的运算; 另一种形式是作为独立的注意力机制输入到网络中. 将位置注意力机制和其他注意力机制结合, 可以充分表示每一个词在句子中的重要程度.

通过结合这 3 种注意力机制作为卷积神经网络的输入, 可以让模型以多种形式学习目标词的特征信息, 达到更好的分类效果.

2.3 词向量注意力机制

注意力机制的目的是在训练过程中, 让模型知道输入数据中哪一部分信息是重要的, 从而使模型高度关注这些信息. 对于特定目标的情感分析, 内容层面的信息是最重要的, 可以通过分析文本的内容得出哪些词对句子中目标词的影响较大. 例如句子“The food was very good, but the service at that restaurant was dreadful”, 词语“good”是用来形容目标词“food”的, 而词语“dreadful”是用来形容目标词“service”的, 所以在该句子中情感词“good”比情感词“dreadful”对目标“food”的影响大得多. 同理, 情感词“dreadful”比情感词“good”对目标“service”的影响也大得多.

对于句子 $s = \{\omega_1, \omega_2, \dots, a_i, \dots, \omega_n\}$, 提取出目标词 a_i 的词向量作为注意力矩阵, 将注意力矩阵和句子 s 的词向量矩阵进行运算操作, 可以得到注意力特征矩阵 A^c , 如图 4 所示, 其中 A^c 为对角矩阵:

$$A_{i,i} = \text{innerproduct}(t_i, x_i), \quad (4)$$

$$A_{i,i}^c = \frac{\exp(A_{i,i})}{\sum_{j=1}^n \exp(A_{j,j})}, \quad (5)$$

此外, 矩阵 A^c 也可由

$$A_{i,i}^c = \alpha \times \frac{\exp(A_{i,i})}{\sum_{j=1}^n \exp(A_{j,j})} \quad (6)$$

得出, 其中, α 为可调参数, 用来控制不同词向量对目标词的影响程度, 表示出每一个词的重要程度. α 可以人工设定, 也可以由位置注意力机制计算给出, 通过 α 可以表示出每一个词的重要程度.

利用计算得到的注意力特征矩阵 A^c 和原来的词向量做运算可分别得到卷积神经网络的输入矩阵:

$$z_i^c = x_i \oplus A_{i,i}^c, \quad (7)$$

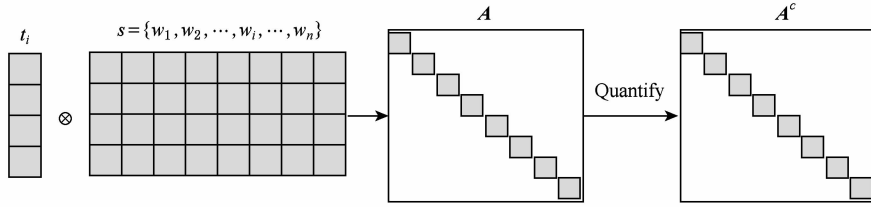


Fig. 4 The operation of attention mechanism

图4 注意力机制运算操作

$$z_i = x_i \cdot A_{i,i}, \quad (8)$$

这2种方法都可以用作输入矩阵的计算,本文实验中使用式(7)计算输入矩阵,即简单的向量拼接。

2.4 词性注意力机制

文本的内容信息是文本情感分析的关键,然而在分词错误以及情感词典中的情感词在数据集中的覆盖率较低的情况下,这种仅仅依赖文本内容信息对文本情感进行分类的准确率将会大大降低。针对该问题,本文提出一种基于词性的注意力机制,结合词向量注意力机制作为网络的输入。本文通过对特殊词的词性重新标注,可以让模型学习到目标词和情感词之间的联系信息。以句子“The food was very good, but the service at that restaurant was dreadful”为例,对句子中的特殊词词性重新标注如图5所示:

```
The/DT </TAG food/TAR> TAG was/VBD very/RB good/POS./,
but/CC the/DT service/TAR at/IN that/DT restaurant/NN was/
VBD Dreadful/NEG
```

Fig. 5 Part of speech re-tagging

图5 词性重新标注

和词向量一样,本文将每一个词性映射为一个多维的连续值向量,称为词性向量,即 $tag_i \in \mathbb{R}^l$,其中 l 为词性向量维度。对于长度为 n 的句子,可以将词性向量表示为如式(9)所示的向量矩阵,其中 tar 表示目标词的词性向量。

$$tag_{1,n} = tag_1 \oplus tag_2 \oplus \dots \oplus tar \oplus \dots \oplus tag_n. \quad (9)$$

本文提取目标词的词性向量作为词性注意力特征矩阵,即 $A' = tar$,然后计算网络的输入矩阵:

$$z'_i = \alpha \times \beta \times \frac{A'_i + tag_i}{2}, \quad (10)$$

其中 β 为权重系数。通过调整 β 的取值,可以利用句子的情感特征信息,本文在实验中对情感词取1.2,其他词取1.0。

2.5 位置注意力机制

在特定目标情感分析任务中,词语和目标词之

间的位置往往也隐含着很重要的信息,一种简单的距离计算方式如图6所示:

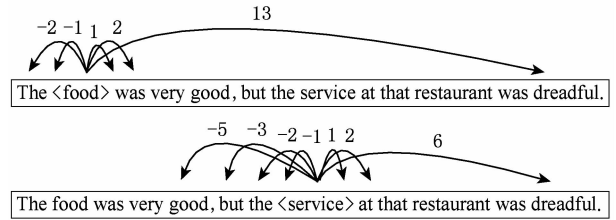


Fig. 6 Example of location calculation

图6 位置计算方式举例

我们普遍认为距离目标词更近的词对目标词的影响更大,例如句子中的情感词“good”比情感词“dreadful”距离目标词“food”更近,同时“good”也是目标词“food”的正确情感表达。但是对于目标词“service”来说,和情感词“good”的距离更近,但是情感词“dreadful”才是目标词“service”的正确情感表达。针对这个问题,本文提出一种双向扫描算法来确定词语和目标之间的位置取值,如算法1所示。

算法1. 双向扫描算法.

输入: 分词后的句子 s ;

输出: 每个词和目标词之间的位置取值集合 l .

1) 将目标词的位置取值置为0,其他词的位置取值置为 n ,其中 n 为句子的长度;

2) 以目标词为中心,设定2个工作指针分别往左右两边扫描(以下步骤以某一个指针为例子);

3) 记录工作指针和目标相对位置 i 的值 l_i ,如果该位置的词为标点符号,执行步骤4);如果该位置的词为目标词集合中的词,则执行步骤5);否则,执行步骤6);

4) 利用下式更新 l_i 的值,将位置取值 l_i 加入集合,继续扫描;

$$l_i = l_i + \min\left\{5, \frac{n-i}{2}\right\};$$

5) 利用式 $l_i = 1$ 更新 l_i 的值,将位置取值 i 加入集合,继续扫描;

6) 将位置取值 l_i 加入集合,继续扫描;

7) 当 2 个指针分别到达句子的起始和终止位置时, 将位置取值 l_i 加入集合, 停止扫描。

本文用矩阵 \mathbf{P} 来存储数据集所有句子的位置取值, 然后计算 α 的取值:

$$\alpha = 1 - \frac{\mathbf{P}_i + 1}{n + 1}, \quad (11)$$

将矩阵 \mathbf{P} 中存储的所有位置取值映射为一个多维向量, 即 $\mathbf{P}_i \in \mathbb{R}^k$, 然后计算网络的输入矩阵:

$$\mathbf{z}_i^p = \frac{\mathbf{P}_i + \mathbf{x}_i}{2}. \quad (12)$$

2.6 多注意力卷积神经网络模型

卷积神经网络可以通过多注意力机制使模型在训练过程中以不同的特征信息来关注目标, 可以提取和学习到目标更多的隐藏信息, 从而能更好地识别不同目标的情感极性。本文将针对多注意力机制介绍 3 种构造输入矩阵的方法:

1) AATT-CNN. 以词的内容为基础, 把词向量注意力、词性注意力和位置注意力作运算, 如式(13)所示。该方法以文本最重要的内容信息为基础, 结合文本的其他特征信息, 使网络更好地识别不同目标的情感信息。在式(13)中 λ_c, λ_l 和 λ_p 可以为系数, 也可以为系数矩阵。

$$\mathbf{z} = \lambda_c \mathbf{z}^c + \lambda_l \mathbf{z}^l + \lambda_p \mathbf{z}^p. \quad (13)$$

2) CATT-CNN. 将不同的注意力输入矩阵通过矩阵变换拼接形成新的输入矩阵, 式(14)给出一种拼接方式。其中 $\mathbf{G}, \mathbf{R}, \mathbf{U}$ 为权重矩阵, 可以在训练过程中学习和调整。

$$\mathbf{z} = \mathbf{z}^c \mathbf{G} \oplus \mathbf{z}^l \mathbf{R} \oplus \mathbf{z}^p \mathbf{U}. \quad (14)$$

3) SATT-CNN. 将不同的注意力输入堆叠形成 3 维的张量作为网络的输入。该方法的优点是可以形成不同的输入通道对输入矩阵进行卷积操作, 缺点是在输入层忽略了不同注意力矩阵的联系。

卷积神经网络可以通过卷积核同时对文本中的多个词进行卷积操作, 保留了词之间的联系。对于长度为 h 的卷积窗口, 对输入矩阵的卷积操作为

$$c_i = f(\mathbf{w} \cdot \mathbf{x}_{i:i+h-1} + b), \quad (15)$$

其中, $\mathbf{w} \in \mathbb{R}^{h \times k}$ 为卷积核权重, $b \in \mathbb{R}$ 为偏置, f 为激活函数, $\mathbf{x}_{i:i+h-1}$ 为一个卷积窗口的文本局部特征矩阵。对于长度为 n 的句子, 通过卷积操作可得到所示的特征图:

$$\mathbf{c} = [c_1, c_2, \dots, c_{n-h+1}], \quad (16)$$

即 $\mathbf{c} \in \mathbb{R}^{n-h+1}$ 。本文在池化层采用 max-over-time pooling 方法^[20]对特征图进行下采样, 提取最重要的特征 $\hat{\mathbf{c}} = \max\{\mathbf{c}\}$ 。对于有 m 个卷积核的窗口采样得到的特征信息如式(17)所示。下采样层输出的特征向量作为全连接层的输入, 本文通过一个 softmax 函数输出分类结果, 如式(18)所示。

$$\hat{\mathbf{c}} = [\hat{c}_1, \hat{c}_2, \dots, \hat{c}_m], \quad (17)$$

$$y = \text{softmax}(\mathbf{W}_f \mathbf{X}_p + \mathbf{B}_f), \quad (18)$$

其中, \mathbf{X}_p 为下采样层输出, $\mathbf{W}_f \in \mathbb{R}^{C \times |\mathbf{X}_p|}$ 为全连接层权重矩阵, $\mathbf{B}_f \in \mathbb{R}^C$ 为全连接层偏置。本文使用反向传播算法来训练模型, 通过最小化交叉熵来优化模型, 交叉熵代价函数为

$$\text{loss} = - \sum_{i=1}^D \sum_{j=1}^C \hat{y}_i^j \log y_i^j + \lambda \|\theta\|^2, \quad (19)$$

其中, D 为训练数据集大小, C 为类别数, y 为预测类别, \hat{y} 为实际类别, $\lambda \|\theta\|^2$ 为正项则。

3 实验

将本文提出的方法在 3 个不同领域数据的数据集上进行实验, 解决特定目标情感分析任务。通过和对比实验相比, 验证本文提出的方法在不同领域的数据集上都能取得较好的情感分类性能。在本文的实验中, 英文词向量采用 Pennington 等人^[21]提出的 Glove 词向量^①, 其中每个词向量为 300 维, 词典大小为 1.9 MB。中文词向量采用从汽车之家爬取的用户评论, 利用 word2vec 的 skip-gram 训练得到^[22], 每个词向量为 300 维。对于未登录词, 采用均匀分布 $U(-0.01, 0.01)$ 来随机初始化词向量。本文采用 ICTCLAS 分词工具^②对中文数据集进行分词和词性标注。对于单个词的目标, 本文使用词向量作为目标的向量表示, 例如“food”, “menu”, “staff”等。对于由多个词组成的目标, 本文使用多个词的词向量平均值作为目标的向量表示, 例如“dim sum”, “battery life”等。

3.1 实验数据

本文采用英文的 SemEval2014 数据集^③和中文的汽车领域数据集^④(automotive-domain data, ADD)进行实验。其中汽车领域数据集为 2016 年 CCF 大

① <http://nlp.stanford.edu/projects/glove/>

② <http://ictclas.nlpir.org/>

③ <http://alt.qcri.org/semeval2014/>

④ <http://www.wid.org.cn/data/science/player/competition/detail/description/237/>

数据与计算智能大赛(BDCI)中“基于视角的领域情感分析”赛题数据, SemEval2014 是语义评测比赛任务 4 的数据集, 包含 laptop 和 restaurant 两个领域的用户评论. 数据样本的情感极性分为积极、消极和中性, 表 1 给出本文实验使用数据的统计. 本文在 Laptop 领域数据集使用较少的训练数据, 用来分析不同模型在数据样本缺乏的数据集上情感分类性能.

Table 1 Statistic of The Datasets

表 1 实验使用数据统计

Data Set	Positive	Negative	Neutral
Laptop-train	834	709	452
Laptop-test	235	167	133
Restaurant-train	1 632	654	512
Restaurant-test	407	159	135
Automotive-train	2 686	1 089	1 657
Automotive-test	854	243	458

3.2 超参数

本文使用 3 种不同的注意力矩阵输入方法完成实验, 实验当中使用多种窗口卷积核对输入矩阵进行卷积操作. 卷积核函数为 rectified linear units, 训练过程采用 Zeiler^[23] 提出的 Adadelta 更新规则. 其他参数如表 2 所示:

Table 2 Hyper Parameters of Experiment

表 2 实验参数设置

Parameter	Parameter Description	Value
h	Windows Size	3, 4, 5
n	No. of Features Map	100
p	Dropout Rate	0.5
s	l_2 Constrain	3
b	Mini-batch Size	50

3.3 对比实验

将本文提出的 3 种方法 AATT-CNN, CATT-CNN, SATT-CNN 和以下的 8 种方法在 3 个不同领域的数据集上进行实验:

1) CNN. 基于 Kim^[24] 提出的卷积神经网络模型, 是最基础的卷积神经网络.

2) SVM. 文献[15]提出的基于特征的 SVM 分类模型, 该模型取得了比以往研究更好的分类效果.

3) ATT-CNN. 基于文献[11]提出的多层注意力机制卷积神经网络, 将词向量作为注意力机制构建网络的输入矩阵, 形成单注意力机制.

4) ATT-LSTM. 文献[19]提出的基于注意力

机制的 LSTM 网络, 该模型在 5 种特定目标的情感分类中取得了比传统 LSTM 网络更好的分类效果.

5) AATT-CNN-1. 本文提出的将多种注意力矩阵作运算作为卷积神经网络的输入模型, 实验中 λ_c, λ_i 和 λ_p 系数分别取 0.6, 0.3 和 0.1.

6) AATT-CNN-2. 本文提出的将多种注意力矩阵作运算作为卷积神经网络的输入模型, 实验中随机初始化 λ_c, λ_i 和 λ_p 系数矩阵, 并在实验中可调整.

7) CATT-CNN. 本文提出的将多种注意力机制拼接作为卷积神经网络的输入模型, 实验中 G, R, U 权重矩阵随机初始化, 并在实验中可调整.

8) SATT-CNN. 本文提出的将多种注意力机制堆叠形成 3 维张量作为卷积神经网络的输入模型.

3.4 实验结果与分析

本文将 8 组实验在 SemEval2014 和 ADD 数据集上进行实验, 分析特定目标的情感极性. 表 3 给出 8 组实验在 3 个不同领域的情感分类正确率.

Table 3 Accuracy on Aspect-based Sentiment Classification of Different Models

表 3 不同模型的特定目标情感分类正确率

Model	Laptop	Restaurant	Automotive
CNN	58.32	66.33	72.09
SVM	66.17	77.18	81.54
ATT-CNN	65.42	75.89	80.13
ATT-LSTM	68.22	77.32	83.09
AATT-CNN-1	65.79	76.89	82.06
AATT-CNN-2	67.48	77.89	82.70
CATT-CNN	67.10	79.17	84.63
SATT-CNN	68.04	78.46	83.53

从表 3 结果可以看出, 本文提出的 4 种方法在 3 个领域的数据集上都取得不错的情感分类效果, 其中 AATT-CNN-2, CATT-CNN 和 SATT-CNN 在 3 个领域的数据集中的情感分类效果都比使用传统的基于特征的 SVM 分类模型好. 没加入任何特征的 CNN 模型对特定目标的情感分析性能不是很理想, 在最好的汽车领域数据集的分类正确率也只有 72.09%. 分析实验结果可知, CNN 模型将很多句子中的不同目标判别为同样的情感极性, 普通的 CNN 模型没有针对特定的目标提取更多的特征信息, 所以模型无法准确识别同一个句子不同目标的情感极性. 基于注意力机制的 ATT-CNN 模型相比 CNN 模型分类正确率有明显的提高, 在最高

的 Restaurant 领域数据集提升了 9.56%，这说明注意力机制能使模型在训练过程高度关注特定的目标的特征信息，从而更好地识别特定目标的情感极性，验证了注意力机制在特定目标情感分类任务中的有效性。

对比本文提出的 4 种模型和使用单一的注意力机制的 ATT-CNN 模型可以看出，本文提出的 4 种模型在 3 个数据集的分类正确率都明显高于单一注意力机制的 ATT-CNN 模型，其中 Laptop 数据集提升最高为 2.62%，Restaurant 数据集提升最高为 3.28%，汽车领域数据集提升最高为 4.5%。实验结果表明对比单一注意力机制的 ATT-CNN 模型，结合多种注意力机制的卷积神经网络模型可以使网络在训练过程通过不同的注意力机制来关注和学习特定目标的情感信息，以及通过不同注意力机制之间的相互联系来提取挖掘更多的隐藏信息，有效弥补了单一注意力机制的不足，从而取得更好的分类效果。

对比本文提出的 4 种模型和文献[19]提出的 ATT-LSTM 模型可以看出，ATT-LSTM 模型在 Laptop 领域数据集取得了最好的分类效果(正确率为 68.22%)，而本文提出的 CATT-CNN 和 SATT-CNN 模型在 2 个训练数据较多的 Restaurant 领域数据集和汽车领域数据集的情感分类正确率均高于 ATT-LSTM 网络模型，其中表现最好的 CATT-CNN 模型在 2 个训练数据领域数据集上的分类正确率分别提升了 1.28% 和 1.54%。说明和结合注意力机制的 LSTM 网络相比，基于多注意力机制的卷积神经网络虽然在数据量较少的数据集上因为训练样本的不足容易出现过拟合现象，从而导致情感分类正确率较低，但是在训练集数据量较为充足的数据集中可以通过不同的注意力机制来关注输入文本，使模型能提取到更多的隐藏信息，从而有更好的情感分类效果。

对比 AATT-CNN-1 和 AATT-CNN-2 两组实验可以看出，AATT-CNN-2 模型在 3 个领域的数据集都好于 AATT-CNN-1 模型，其中分类正确率提升最高的是 Laptop 数据集，为 1.69%。这说明结合系数矩阵的注意力输入比单纯的固定系数注意力输入的情感分类效果更好。通过系数矩阵可以使网络在训练过程更细地调整特征向量的每一个分量，从而让网络在更低维度的调整中学习更多的隐藏信息。

对比 AATT-CNN-2, CATT-CNN 和 SATT-

CNN 这 3 组实验可以看出，采用通过矩阵变换的拼接方法虽然在训练数据较少的 Laptop 领域数据集的分类正确率最低，只有 67.10%，但是在 Restaurant 领域数据集和汽车领域数据集的情感分类正确率都明显高于其他 2 组实验。结合变换矩阵的 CATT-CNN 模型和 SATT-CNN 模型相比，CATT-CNN 模型可以在训练过程中通过调整变换矩阵使不同注意力机制在输入层就能很好地互相结合，而 SATT-CNN 模型只有在卷积层以后才会有不同注意力机制的信息交互。采用拼接操作的 CATT-CNN 模型和 AATT-CNN-2 模型相比，CATT-CNN 模型保留了不同注意力机制的原始输入信息，模型可以更好地调整注意力机制输入向量。

为了进一步比较本文提出的 AATT-CNN-2, CATT-CNN, SATT-CNN 这 3 个模型和 ATT-LSTM 模型在特定目标情感分析任务中的分类效果，本文只保留数据集中的积极和消极样本对 4 个模型进行对比实验。实验结果如表 4 所示：

Table 4 Accuracy on Aspect-based Sentiment Classification of Different Models Ignoring All Neutral Instances

表 4 不同模型在二分类实验中的情感分类正确率

Model	Laptop	Restaurant	Automotive
ATT-LSTM	86.57	89.22	92.16
AATT-CNN-2	84.83	88.52	91.98
CATT-CNN	85.57	89.75	93.98
SATT-CNN	85.82	90.28	92.34

从表 4 可以看出在去除中性极性数据之后，4 组实验的分类正确率都有了明显的提升，其中 ATT-LSTM 模型在 Laptop 领域数据的分类正确率最高，为 86.57%。说明 ATT-LSTM 模型在数据量较小的数据集中仍然可以取得比其他模型更好的分类效果。CATT-CNN 和 SATT-CNN 模型在 Restaurant 和汽车领域数据集的分类正确率都高于 ATT-LSTM 模型，说明在数据集数量较为充裕时，基于多注意力机制的卷积神经网络的情感分类效果要好于基于注意力机制的 LSTM 网络。AATT-CNN-2 模型在 3 个数据集的分类效果都是最差的，说明在输入层将注意力机制矩阵做运算会失去注意力机制原有的输入信息，在训练过程中网络只能通过调整系数矩阵和运算后的注意力矩阵来学习特征信息，导致不同的注意力机制没有得到充分的利用，降低了分类性能。

对比 CATT-CNN 模型和 SATT-CNN 模型可

可以看出,在 Laptop 领域和 Restaurant 领域数据集 SATT-CNN 模型的分类正确率高于 CATT-CNN 模型,分别提升 0.25% 和 0.53%,但是在汽车领域数据集中 CATT-CNN 模型要明显优于 SATT-CNN 模型,分类正确率提升了 1.64%。

为了进一步分析 CATT-CNN 模型和 SATT-CNN 模型在不同大小数据集上的分类效果,本文提取 Restaurant 领域 3 000 条数据和汽车领域 4 500 条数据进行 5 倍交叉验证实验,实验结果如图 7 所示。从图 7 结果可以看出随着数据集的样本数增加,CATT-CNN 模型和 SATT-CNN 模型都呈现出越来越好的分类效果。在数据集样本数较少的实验中,SATT-CNN 模型要好于 CATT-CNN 模型,而在数据集样本数较多的实验中,CATT-CNN 模型分类效果要明显优于 SATT-CNN 模型。因为 CATT-CNN 模型在训练过程中可以调整变换矩阵和注意力机制矩阵 2 个信息的变量,可以更好地拟合训练数据,所以在数据量较少的数据集中会出现过拟合的情况,而 SATT-CNN 模型只通过调整不同通道的注意力机制输入信息,过拟合现象远低于 CATT-CNN 模型,如图 7(a)所示。相反,随着数据集的样本数量的增加,CATT-CNN 模型能通过调整变换矩阵和注意力机制输入矩阵,使不同注意力机制更好地互相学习,达到更好的数据拟合效果,随着数据

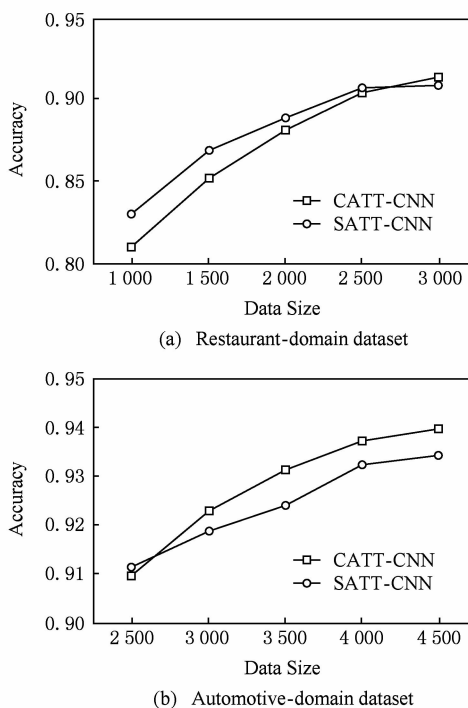


Fig. 7 Classification accuracy of different data size

图 7 不同大小数据集的分类正确率比较

量的增加也可以减轻过拟合的现象,所以随着数据量的增加,CATT-CNN 模型分类效果要比 SATT-CNN 模型更好,如图 7(b)所示。

本文在以前研究的基础上提出了加入词性注意力机制的方法,为了验证词性注意力机制的有效性,本文从 Restaurant 领域和汽车领域分别抽取 3 000 条数据在 AATT-CNN-2, CATT-CNN 和 SATT-CNN 这 3 个模型上使用不同维度的词性向量构造词性注意力机制进行 5 倍交叉验证实验,实验结果如图 8 所示,其中词性向量维度为 0 表示不使用词性注意力机制。从图 8 结果可以看出,加入词性注意力机制后,3 个模型在 2 个领域数据集上的分类效果都有显著提升。其中,Restaurant 领域分类正确率提升最高为 4.5%,汽车领域分类正确率提升最高为 3.83%。说明加入词性注意力机制可以让模型更加充分地学习文本的情感信息,达到更好的情感分类效果。此外,当词性向量维度大于 100 维时,AATT-CNN-2 模型和 CATT-CNN 模型在 2 个领域数据集上的分类正确率都出现了波动。分析实验结果可以看出,由于本文实验中使用的词向量维度为 300 维,AATT-CNN-2 模型和 CATT-CNN 模型

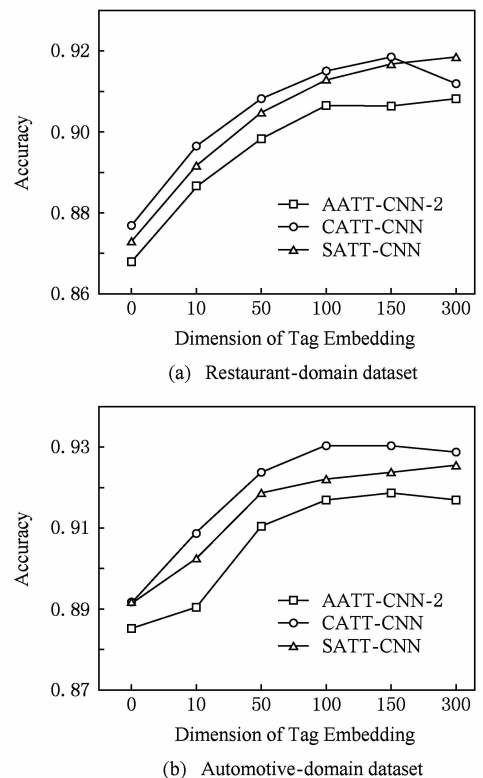


Fig. 8 Classification accuracy of different tag embedding dimension

图 8 不同维度词性向量的分类正确率比较

的输入矩阵是注意力机制有联系的输入,所以词向量注意力机制作为模型训练的主要特征,当词性向量的维度超过一定阈值时会影响模型对词向量注意力机制的参数调整,降低了模型在内容层面特征信息的学习效果.以独立注意力机制矩阵作为网络输入的 SATT-CNN 模型在训练过程中可以独立完成不同注意力的参数调整,所以当词性向量维度大于 100 时 SATT-CNN 模型的分正确率仍有提升,但提升幅度已经不明显.此外,随着向量维度的增加,模型的训练时间代价也会增加,所以本文在实验中选取 100 作为词性向量的维度.

3.5 训练时间分析

为了分析不同网络模型在相同条件下的时间性能,本文在相同的 CPU, GPU 和网络框架下完成所有神经网络实验,同时,所有实验使用相同的词向量矩阵和相同的数据集.表 5 给出不同网络模型在 Restaurant 领域数据集上完成一次迭代的训练时间对比结果.

Table 5 Runtime of Each Training Epoch

表 5 不同模型完成一次迭代的训练时间

Model	Time Cost
ATT-LSTM	500+
CNN	4
ATT-CNN	12
AATT-CNN-1	16
AATT-CNN-2	23
CATT-CNN	21
SATT-CNN	24

从表 5 结果可以看出 LSTM 网络的训练时间代价是非常高的,这主要因为 LSTM 网络训练的是序列性数据,每一个单元都需要相当复杂的运算操作.ATT-LSTM 模型完成一次迭代的训练时间超过 500 s,是本文提出方法的 20 多倍.此外,不加任何特征的 CNN 模型是最快的,完成一次迭代的训练时间只需要 4 s,加入注意力机制的 ATT-CNN 模型训练时间也只需 12 s,说明卷积神经网络模型训练的时间性能要远远好于 LSTM 网络.此外,本文提出的 4 种多注意力机制卷积神经网络模型的运行时间虽然高于 CNN 和 ATT-CNN 模型,但是远低于 ATT-LSTM 网络模型.这说明基于注意力机制的卷积神经网络可以接收句子的平行化输入,可以有效降低模型的训练时间.

4 总 结

在特定目标情感分析任务中,如今大部分研究都是将注意力机制和 LSTM 网络相结合的方法,这类方法基于循环网络模型,训练过程需要很高的时间代价.本文基于注意力机制和卷积神经网络结合的研究,提出结合多种注意力机制和卷积神经网络的特定目标情感分类方法,该模型大大降低了基于注意力机制网络模型的训练时间.同时,过去的研究主要是结合词向量注意力机制和普通的位置注意力机制,本文在以前的研究基础上提出了加入词性注意力机制的方法,并通过对比实验验证了加入词性注意力机制能有效提高特定目标情感分析的分类正确率.此外,本文使用一种双向扫描算法来计算句子的位置特征,使目标词的位置信息得到充分的利用.实验结果表明,本文提出的方法对比传统的卷积神经网络方法有很明显的提升,同时对比基于注意力机制的 LSTM 网络模型也有较好的分类效果.

从实验结果可以看出,本文提出的多注意力卷积神经网络方法在训练样本缺乏的数据集上容易出现拟合现象,导致分类效果不理想.所以本文接下来将针对这个问题对多注意力卷积神经网络进行改进.

参 考 文 献

- [1] Nasukawa T, Yi J. Sentiment analysis: Capturing favorability using natural language processing [C] //Proc of the 2nd Int Conf on Knowledge capture. New York: ACM, 2003: 70-77
- [2] Wang Zhongyuan, Cheng Jianpeng, Wang Haixun, et al. Short text understanding: A survey [J]. Journal of Computer Research and Development, 2016, 53(2): 262-269 (王仲远, 程健鹏, 王海勋, 等. 短文本理解研究[J]. 计算机研究与发展, 2016, 53(2): 262-269)
- [3] Kalchbrenner K 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
- [4] Zhou Xinjie, Wan Xiaojun, Xiao Jianguo. Attention-based LSTM network for cross-lingual sentiment classification [C] //Proc of the 2016 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2016: 247-256
- [5] Yu Jianfei, Jiang Jing. Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification [C] //Proc of the 2016 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2016: 236-246

- [6] Pang B, Lee L. Opinion mining and sentiment analysis [J]. *Foundations and Trends in Information Retrieval*, 2008, 2(1/2): 1-135
- [7] Pontiki M, Galanis D, Pavlopoulos J, et al. Semeval-2014 task 4: Aspect based sentiment analysis [C] //Proc of the 8th Int Workshop on Semantic Evaluation. Stroudsburg, PA: ACL, 2014: 27-35
- [8] Mnih V, Heess N, Graves A. Recurrent models of visual attention [C] //Proc of Advances in Neural Information Processing Systems 27 (NIPS 2014). Cambridge, MA: MIT Press, 2014: 2204-2212
- [9] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [C/OL] //Proc of ICLR 2015. [2017-03-15]. <https://arxiv.org/pdf/1409.0473.pdf>
- [10] Yin Wenpeng, Schütze H, Xiang Bing, et al. Abcnn: Attention-based convolutional neural network for modeling sentence pairs [J/OL]. arXiv preprint arXiv: 1212.05193, 2015. [2017-03-15]. <https://arxiv.org/pdf/1512.05193.pdf>
- [11] Wang Linlin, Cao Zhu, de Melo G, et al. Relation classification via multi-level attention cnns [C] //Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics. Stroudsburg, PA: ACL, 2016: 1298-1307
- [12] Hu Mingqing, 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
- [13] Hu Mingqing, Liu Bing. Mining opinion features in customer reviews [C] //Proc of AAAI 2004. Menlo Park: AAAI, 2004, 4(4): 755-760
- [14] Qiu Guang, Liu Bing, Bu Jiajun, et al. Opinion word expansion and target extraction through double propagation [J]. *Computational Linguistics*, 2011, 37(1): 9-27
- [15] Kiritchenko S, Zhu Xiaodan, Cherry C, et al. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews [C] //Proc of the 8th Int Workshop on Semantic Evaluation (SemEval 2014). Stroudsburg, PA: ACL, 2014: 437-442
- [16] Nguyen T H, Shirai K. PhraseRNN: Phrase recursive neural network for aspect-based sentiment analysis [C] //Proc of the 2015 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2015: 2509-2514
- [17] Dong Li, Wei Furu, Tan Chuangqi, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification [C] //Proc of the 52nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2014: 49-54
- [18] Ruder S, Ghaffari P, Breslin J G. 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
- [19] Wang Yequan, Huang Minlie, Zhao Li, et al. Attention-based LSTM for aspect-level sentiment classification [C] //Proc of the 2016 Conf on Empirical Methods in Natural

Language Processing. Stroudsburg, PA: ACL, 2016: 606-615

- [20] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch [J]. *Journal of Machine Learning Research*, 2011, 12(8): 2493-2537
- [21] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation [C] //Proc of the 2014 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2014: 1532-1543
- [22] Mikolov T, Chen Kai, Corrado G, et al. Efficient estimation of word representations in vector space [C/OL] //Proc of Workshop at ICLR 2013. [2017-03-15]. <https://arxiv.org/pdf/1301.3781.pdf>
- [23] Zeiler M D. ADADELTA: An adaptive learning rate method [J/OL]. arXiv preprint arXiv: 1212.5701, 2012. [2017-03-15]. <https://arxiv.org/pdf/1212.5701.pdf>
- [24] Kim Y. Convolutional neural networks for sentence classification [C] //Proc of the 2014 Conf on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2014: 1746-1751



Liang Bin, born in 1993. Master candidate. His main research interests include sentiment analysis, natural language processing, and deep learning.



Liu Quan, born in 1969. PhD, professor and PhD supervisor. His main research interests include intelligence information processing, automated reasoning and machine learning.



Xu Jin, born in 1992. Master candidate. His main research interests include reinforcement learning, deep learning and deep reinforcement learning.



Zhou Qian, born in 1992. Master candidate. Her main research interest is reinforcement learning.



Zhang Peng, born in 1992. Master candidate. His main research interest is continuous space reinforcement learning.