

用于基于方面情感分析的 RCNN-BGRU-HN 网络模型

孙中锋 王 静

(南京工业大学计算机科学与技术学院 南京 211816)

摘 要 针对一般神经网络模型在处理基于方面情感分析任务中存在的句子间相互联系少以及单词之间的语义信息获取有限等问题,文中提出了一种新型结构的深度学习网络模型。该模型通过区域卷积神经网络(RCNN)可以很好地保留评论文本中句子的时序关系,同时结合双向门控循环单元(BGRU)可以大大降低模型训练的时间代价。此外,加入的高速公路网络(HN)使得该模型能够捕获更多单词间的语义信息;利用注意力机制来分配网络结构中特定方面的权重,可以有效获取特定方面在整个评论文本中的长距离依赖关系。该模型可以进行端到端的训练,在不同的数据集上取得了比现有网络模型更好的分类效果。

关键词 深度学习,基于方面情感分析,卷积神经网络,双向门控循环单元,高速公路网络,注意力机制

中图法分类号 TP183,TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.09.033

RCNN-BGRU-HN Network Model for Aspect-based Sentiment Analysis

SUN Zhong-feng WANG Jing

(School of Computer Science and Technology,Nanjing Tech University,Nanjing 211816,China)

Abstract The general neural network model has less inter-connectivity between sentences and cannot capture much more semantic information between words in the task of aspect-based sentiment analysis. To adress these problems, this paper proposed a deep learning network model with novel structure. The model can preserve the sequential relationship of sentences in the comment text through the regional convolutional neural network(RCNN). At the same time, the time cost of model training can be greatly reduced by combining bi-directional gated recurrent unit (BGRU). In addition, the introduction of highway network (HN) could enable the proposed model to capture much more semantic information between words. The attention mechanism is additionally utilized in an effort to assign weights of the concerned aspect in the network structure, which is able to effectively obtain the long-distance dependency of the concerned aspect in the whole review. The model can give end-to-end training and experiment on different datasets, achieving better performance than the existing network model.

Keywords Deep learning, Aspect-based sentiment analysis, Convolutional neural network, Bi-directional gated recurrent unit, Highway network, Attention mechanism

1 引言

近年来,随着社交网络的日益繁荣,特别是在线社区、博客、维基、微博平台和其他在线协作媒体上的社交媒体数据的快速发展,情感分析^[1]已成为自然语言处理(Natural Language Processing,NLP)领域的研究热点之一。作为情感计算研究的一个分支,情感分析旨在将文本极性分类为正面或负面(在某些情况下为中性)。早期关于情感分析的研究主要是判断文本单元的总体情绪^[2],即假设句子表达的是统一的情绪或极性。相比之下,基于方面情感分析扩展了情感分析的典型设置,更偏向于特定方面所对应的情感。例如,在句子“The design of the space is good but the service is horrible”中,“space”和“service”两个方面表达的情感是完全相反的。

因此,在同一个句子中,针对特定方面进行情感极性分析,有助于挖掘用户更深层次的情感表述。

自 Pang 等使用诸如朴素贝叶斯分类和最大熵等方法来研究句子中包含的情感以来^[3],机器学习技术已经被广泛应用于情感分析。传统的机器学习方法专注于设计手工特征,并使用支持向量机(Support Vector Machines,SVM)等情感分类器来完成任务。特征表示是许多机器学习系统的关键组件,这是因为其在很大程度上决定了系统的性能。这类方法在过去的研究中已经取得了很大的成功,但其通常需要依赖复杂的人工规则和特征工程。

深度学习方法能够在一定程度上自动捕捉来自文本的句法和语义特征,而不需要特征工程,近年来吸引了很多研究者的注意,并在 NLP 的许多领域取得了巨大的成功。同时,随

到稿日期:2018-07-18 返修日期:2018-10-02 本文受南京工业大学引进人才启动基金项目(39809110)资助。
孙中锋(1991—),男,硕士生,主要研究领域为机器学习与深度学习;王 静(1982—),女,副研究员,主要研究领域为无线传感器网络技术,
E-mail:wj_cec@126.com(通信作者)。

着注意力机制的有效应用,与神经网络结合的注意力机制在基于特定方面的任务中取得了比传统方式更有效的结果。自Mnih等^[4]将注意力机制成功应用于图像分类后,Bahdanau等^[5]将循环神经网络(Recurrent Neural Networks,RNN)与注意力机制相结合,在机器翻译任务中取得了显著的效果。Yin等^[6]提出了一种将卷积神经网络(Convolutional Neural Network,CNN)与注意力机制相结合的网络模型,并将其用于句子建模中;Wang等^[7]提出了一种基于注意力机制的LSTM网络方面情感分类模型,其取得了比以往研究更优越的性能。这些结合了注意力机制的神经网络模型可以更有效地针对特定方面进行模型训练,然后及时调整参数权重,以挖掘更深层的情感信息。但这些模型通常是以独立的句子作为输入,忽略了评论文本中句子之间的相互关系,且获取的单词之间的语义信息有限,很难有效判断一些复杂评论文本的情感极性。

针对这些问题,本文提出一种区域卷积神经网络与融入了高速公路网络的双向门控循环单元相结合的网络模型。该模型不但可以获取单词之间更多的语义信息,还可以更好地联系句子之间的情感极性,捕获特定情感在整个评论文本中的长距离依赖关系,从而更加有效地识别评论中不同方面的情感极性。

2 相关工作

2.1 方面情感分析

方面情感分析是针对文本中对特定方面所表达的情绪进行分析的任务,它面临的最大挑战是如何有效地表达整个句子的特定方面的情感信息。有关方面情感分析的早期工作主要依靠特征工程来表征句子^[8],最早的关于方面情感分析的是Hu等的研究^[9],该研究遵循一种基于频率的方法,基本思想是经常提到的名词更可能是要关注的方面。近年来,表示学习和深度学习模型已被有效地应用于不同级别的文本输入(如单词、短语、句子和文档)的语义表示学习中,它们在情感分析的任务中取得了显着的成果。同时,使用了深度学习技术的网络模型在基于方面情感分析任务中也取得了比传统机器学习方法更好的效果。例如,Poria等^[10]提出了一种深度卷积神经网络的学习方法,用于观点挖掘中的方面提取。该模型使用7层深度卷积神经网络将固定句子中的每个单词标记为方面单词或非方面单词,同时还为方面提取开发了一套语言模式,将它们与神经网络相结合,由此产生了集合分类器,并将该集合分类器与用于情感分析的单词嵌入模型相结合,从而可以提取出更加准确的方面。Tang等^[11]使用一种融入了注意力机制的深度记忆网络用于方面情感分析任务,该方法在推断方面情感极性时可以明确地捕获每个上下文单词的重要性,并使用了两种注意力机制使得模型可以更好地学习上下文权重并关注特定目标的情感信息。

2.2 卷积神经网络

CNN最初是由Lecun等于1998年提出,并用于计算机视觉的^[12]。它为计算机视觉任务提供了出色的结果,并且最近已被证明对于各种NLP任务也很有效,如POS标记、句子建模和句子分类等。CNN在潜在特征表现方面效果显著,其

主要由输入层、卷积层、池化层和全连接层组成。输入层对数据进行向量表示,卷积层利用不同的卷积核进行局部特征提取,池化层对特征向量图执行下采样操作,全连接层将池化层输出的矩阵进行分类。

2.3 BGRU

RNN是一种具有重复结构的神经网络结构,它可以不断地保存以前的信息。然而,在实践中,RNN的性能可能并不完美,它会失去保存和处理信息的能力,而且随着时间的推移,会出现梯度消失的情况。

为了解决这些问题,学者们提出了一些特殊的RNN替代方案,最有代表性的方法是长短时记忆网络(Long Short-Term Memory,LSTM)和门控循环单元(Gated Recurrent Unit,GRU)。它们改变了RNN的隐藏层,使其可以更好地捕捉深层连接,并能够有效解决RNN网络模型在训练过程中的梯度消失问题。LSTM^[13]由Hochreiter等提出,由输入门、遗忘门、输出门以及一个记忆核心(cell)组成。GRU^[14]由Cho等于2014年提出,在LSTM的基础上将遗忘门与输入门合成一个更新门,同时没有单独的存储单元。与LSTM相比,GRU结构更简单,且计算复杂度更低,因此本文选用GRU。

将基于GRU的网络用于训练和获取句子中每个输入词的语境语义特征。作为门控RNN,GRU通过重置门和更新门来控制信息流。在本文中,输入是 x_t ,其中 t 是当前时间步。 h_t 表示 t 时刻隐藏层的状态。

具体而言,GRU网络由两个关键部分组成,即重置门 r_t^j 和更新门 z_t^j 。对于第 j 个GRU隐藏单元,重置门 r_t^j 具有相应的权重矩阵 W_r 和 U_r 。当重置门关闭($r_t^j=0$)时,前一状态 h_{t-1} 将被忽略,当前状态仅由当前输入 x_t 定义,这意味着可以丢弃对当前隐藏状态无用的所有信息。更新门 z_t^j 的设计用于调节从先前状态 h_{t-1} 传输到当前隐藏状态 h_t 的信息的程度,其具有相对应的权重矩阵 W_z 和 U_z 。此外, $\tilde{h}_t^j$ 表示基于先前状态 h_{t-1} 和当前输入 x_t 的候选激活,其具有相应的权重矩阵 W_h 和 U_h 。隐藏单元 h_t 的最终输出包括通过乘以 $1-z_t^j$ 和 h_{t-1}^j 计算出的前一部分,以及通过乘以 z_t^j 和 $\tilde{h}_t^j$ 计算出的更新部分。

第 j 个 GRU 隐藏单元的全部数据流如下所示：

$$r_t^j = \sigma(W_r x_t + U_r h_{t-1})^j \tag{1}$$

$$z_t^j = \sigma(W_z x_t + U_z h_{t-1})^j \tag{2}$$

$$\tilde{h}_t^j = \tanh(W_h x_t + U_h (r_t^j \odot h_{t-1}))^j \tag{3}$$

$$h_t^j = (1 - z_t^j) h_{t-1}^j + z_t^j \tilde{h}_t^j \tag{4}$$

其中, σ 是激活函数sigmoid,在本文中表示为 $\sigma(x)=\frac{1}{1+e^{-x}}$; $\tanh$ 是双曲正切函数,可以表示为 $\tanh(x)=\frac{e^x-e^{-x}}{e^x+e^{-x}}$; $\odot$ 是点乘运算,它表示两个矩阵的相应元素相乘。

然而,上述GRU结构只能获取单向文本信息,为此本文设计了一种BGRU模型,以便从两个方向获取上下文信息,如图1所示。BGRU网络使用两个隐藏层来提取过去和未来的信息,这两个隐藏层连接到相同的输出层。该网络充分利用输入序列的上下文信息。左GRU层通过馈送 S 的正序列

来学习 x_t 的历史信息,右 GRU 通过馈送 S 的反向来获得 x_t 的未来信息。句子 S 的第 i 个单词的最终输出由式(5)表示:

$$h_i = [\vec{h}_i \oplus \vec{h}_i] \quad (5)$$

其中, $\vec{h}_i$ 和 $\vec{h}_i$ 分别表示通过左 GRU 层和右 GRU 层的第 i 个词的输出, $\oplus$ 表示左右两个矩阵相应元素之和。

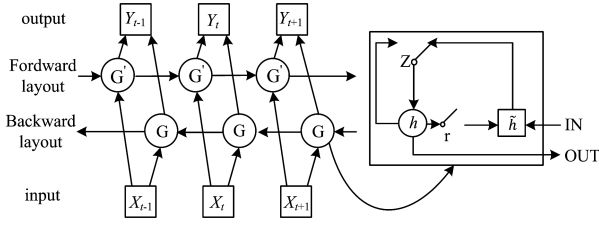


图 1 双向门控循环单元

Fig. 1 Bi-directional gated recurrent unit

3 RCNN-BGRU-HN 网络模型

本文提出的 RCNN-BGRU-HN 模型(见图 2)将一条用户评论按不同方面的目标词分割为固定长度的不同区域,利用一个区域 CNN 接收独立的区域作为网络的输入来提取整个评论中不同区域的局部特征信息;然后将其输出送入 BGRU 层,利用两个隐藏层来提取文本过去和未来的信息;之后经过高速公路网络层增强 BGRU 层获得的特征之间的相关性,并获得具有更深语义内涵的单词的向量表示;接着通过词语层注意力机制、句子层注意力机制来分配网络结构中特定方面的权重;最后通过输出层的 softmax 分类器输出分类结果。

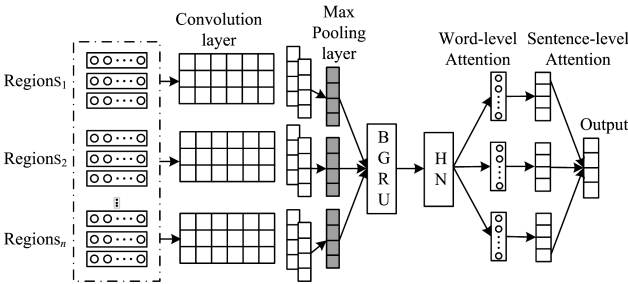


图 2 RCNN-BGRU-HN 网络模型

Fig. 2 RCNN-BGRU-HN network model

3.1 任务定义

对于由 n 个单词组成的给定句子 $S = \{c_1, c_2, \dots, t, \dots, c_n\}$, 其中 t 是句子 S 中特定方面的目标词, 每个句子都有 0 个或多个目标词, 分别对应 1 个或多个不同的方面。每个方面由实体和属性组成, 如 SERVICE # GRNERAL。本文要解决的任务是根据输入的句子判断句子中不同方面的情感极性。例如, 句子“The design of the space is good but the service is horrible”中有两个目标词“space”和“service”, 它们分别对应“SPACE # QUALITY”和“SERVICE # GRNERAL”这两个方面。在同一个句子中, 方面“SPACE # QUALITY”是积极情感, 而方面“SERVICE # GRNERAL”却是消极情感。本文将每一个方面映射为一个多维的连续值向量, 即 $a \in R^m$, 其中 a 为实体和属性的词向量平均值, m 为词向量的维度。

3.2 区域卷积神经网络

区域卷积神经网络通过一定的规则将用户的评论文本划

分为不同的区域, 然后按顺序通过卷积层和池化层, 这样既可以保留句子之间的时序关系, 也可以提取特定方面的情感特征^[15]。文献[15]中的模型使用一个句子作为一个区域, 将给定文本划分为 R 个区域; 与之不同, 本文根据句子中特定方面的目标词进行区域划分, 可以更好地判断同一个句子中不同方面的情感极性。例如, 给定句子 $S = \{c_1, c_2, \dots, t_1, t_2, \dots, c_n\}$, 对于 t_1 和 t_2 两个目标词, 我们将其划分为两个长度为 h 的独立区域, 即 $r_1 = \{c_i, c_{i+1}, \dots, t_1, \dots, c_{h+i-2}\}$, $r_2 = \{c_i, c_{i+1}, \dots, t_2, \dots, c_{h+i-2}\}$ 。将没有目标词的句子直接划分为一个长度为 h 的区域。

对于长度为 h 的独立区域 r_i , 本文将区域中的每个词语映射为一个 m 维的连续值向量, 得到每一个区域的矩阵表示 $r_i = [x_1, x_2, \dots, x_h]$, 其中 $x_i \in R^m$ 。在每个区域中, 我们使用长度为 L 的卷积核对其做卷积操作, 如式(6)所示:

$$y_i = f(W \cdot r_i + b) \quad (6)$$

其中, $W \in R^{m \times L}$ 和 b 分别表示权重矩阵和偏置。对每一个区域做卷积操作, 可以得到区域 r_i 的特征图 $y \in R^{h-L+1}$, 如式(7)所示:

$$y = [y_1, y_2, \dots, y_{h-L+1}] \quad (7)$$

池化层对卷积层的输出进行降采样处理, 对此本文采用 max-pooling 方法, 保留了最显著的特征。

3.3 高速公路网络

高速公路网络(HN)主要学习如何通过门控单元来控制网络中的信息流。高速公路网络中定义了两个非线性变换 T 和 C 。前者为具有相应权重矩阵 W_T 的变换门, 后者为具有相应权重矩阵 W_C 的进位门。因此, 输出 y 由下面的仿射变换计算:

$$y = (H(h, W_H)) \odot (T(h, W_T)) + h \odot (C(h, W_C)) \quad (8)$$

通常情况下, 为简单起见, C 被设置为 $1 - T$ 。因此, 高速公路网络层也可以按如下方式计算:

$$y = (H(h, W_H)) \odot (T(h, W_T)) + h \odot (1 - T(h, W_C)) \quad (9)$$

其中, h 和 y 必须有相同的维度, 从而使得 W_H 和 W_T 是方阵。 H 和 T 分别表示修正线性单元(ReLU)和 Sigmoid 函数的激活函数。这里, ReLU 函数被定义为 $\text{ReLU}(x) = \max\{0, x\}$, 其中 $x \in R$ 。

特别地, 从式(9)中我们可以观察到:

$$y = \begin{cases} h, & T(h, W_T) = 0 \\ H(h, W_H), & T(h, W_T) = 1 \end{cases} \quad (10)$$

通过对式(10)进行 Jacoby 变化, 我们可以得到:

$$d_y/d_h = \begin{cases} I, & T(h, W_T) = 0 \\ H(h, W_H)', & T(h, W_T) = 1 \end{cases} \quad (11)$$

在这里, 由于引入了门控单元, 高速公路网络能够平衡其在传统普通层和仅通过输入层之间的表现。相比于超深度神经网络, 高速公路网络的训练具有独特的优势, 即使网络深度数量增加到 100, 其优化也不会受到阻碍。

在 NLP 任务中, 高速公路网络可以加深上述神经网络提取的特征的内在相关性。一个典型的例子是结合 CNN 和高速公路网络来构建神经语言模型^[16]。与单一的 CNN 模型相比, 高速公路网络的增加可以提取 CNN 获得的人物特征之

间的内在联系,实验结果表明最终获得的人物表征更为全面和语义化。

这里使用高速公路网络来增强 BGRU 层获得的特征之间的相关性,并且获得具有更深语义内涵的单词的向量表示。 T 的偏差 b_T 一般被初始化为负数,使得进位函数 C 可以保留足够多的原始信息。本文中我们也采用这种方式来初始化 b_T 。

3.4 注意力机制

最近,源于人类视觉注意过程的神经注意力机制在众多应用中取得了巨大的成功,包括机器翻译、抽象语句概述、语音识别等。

本文利用了关系提取任务中的神经注意力机制。在图 2 中,注意力机制位于高速公路网络层后。很明显,每个实体对可能对应于语料库中的多个句子,每个句子由多个词组成。基于此,我们采用了两种注意力机制,即词语层注意力机制和句子层注意力机制。

3.4.1 词语层注意力机制

通常,如果通过 BGRU 和高速公路网络获得的所有单词向量在句子 S 中都被均等的表示,那么将会在一些不重要的单词上浪费计算时间。因此,我们给通过注意力词汇的输入序列的每个元素赋予权重,并根据句子中不同词语的重要性赋以不同的权重。权重表示当前输出对每个词的依赖性,其中“1”表示完全依赖,“0”表示完全独立。

这个注意力机制的关键问题是自动计算句子 S 中单词 C_i 的权重 α_i 。 α_i 通过如下公式计算得到:

$$u_i = \mathbf{W}^T y_i \tag{12}$$

$$\alpha_i = e^{u_i} (\sum_{i=1}^n e^{u_i})^{-1} \tag{13}$$

其中, y_i 由高速公路网络生成, $\mathbf{W}$ 是待训练的参数向量, $i \in \{1, 2, \dots, n\}$, n 是句子 S 的长度。

句子 S 的矢量表示 s 由 y_i 的加权和计算:

$$s = \sum_{i=1}^n \alpha_i y_i \tag{14}$$

最后,在 s 上应用一个非线性激活函数后,可以得到:

$$s^* = \tanh(s) \tag{15}$$

3.4.2 句子层注意力机制

本文利用句子层注意力机制来计算朝向指定实体对的所有句子的权重。设 $L = \{S_1, S_1, \dots, S_k\}$ 是包含特定实体对的所有句子的集合, k 是属于同一实体对的句子的数量。首先,我们计算实体对的句子 $S_j (j = 1, 2, \dots, k)$ 的权重 β_j :

$$v_j = (\mathbf{A} S_j^*)^T \mathbf{d} \tag{16}$$

$$\beta_j = e^{v_j} (\sum_{j=1}^k e^{v_j})^{-1} \tag{17}$$

其中, $\mathbf{A}$ 是加权对角矩阵, $\mathbf{d}$ 是与关系类有关的参数向量。

L 的向量表示 l 可以通过 S_j^* 的加权和来获得:

$$l = \sum_{j=1}^k \beta_j S_j^* \tag{18}$$

3.5 模型训练

本文使用 softmax 分类器接收句子层注意力机制的输出,然后得到待分类评论文本的输出结果,如式(19)所示:

$$y = \text{softmax}(\mathbf{W}_l l + \mathbf{b}_l) \tag{19}$$

其中, $\mathbf{W}_l$ 和 $\mathbf{b}_l$ 分别为相应的权重矩阵和偏置。

本文通过实际关系的交叉熵和预测关系来定义成本函数。此外,需要一个能够以最小的复杂度很好地解释训练数据的优化模型。因此,在经验风险背后增加了 L2-范数的正则化矩阵。正则化是选择经验风险和模型复杂性都尽可能小并且可以更好地避免过度拟合的模型的功能。成本函数通过式(20)计算得到:

$$\mathcal{J}(\theta) = -\frac{1}{m} \sum_{i=1}^m t_i \log(r_i) + \frac{\lambda}{2} \|\mathbf{w}\|^2 \tag{20}$$

其中,等号右边第一项是经验风险,第二项是正则项, $t \in R^m$, r_i 表示 softmax 分类器中第 i 个类的概率, m 表示关系类的数量。

4 实验结果和分析

本节将所提方法与几种基本基准方法和最先进的方法进行性能比较。

4.1 实验数据

本文使用的数据来自 2016 年语义评估研讨会任务 5: Aspect-Based Sentiment Analysis。SemEval-ABSA16 是 AB-SA 的多语言任务,涵盖了 8 种不同的语言(即阿拉伯语、英语、中文、荷兰语、法语、俄语、西班牙语和土耳其语)以及 7 个不同的领域(即餐厅、笔记本电脑、手机、数码相机、酒店、博物馆和电信)。我们选择了 restaurants 和 laptops 这两个领域的评论数据集来评估所提方法。数据集的相关统计信息如表 1 所列。实验中,我们去除了没有特定目标且句子属性标志为“OutOfScope”的样本,这些句子都没有注释情感极性。

表 1 实验数据统计
Table 1 Experimental data statistics

数据集	评论数	积极句子	消极句子	中性句子
REST-Train	350	1657	749	101
REST-Test	90	611	204	44
LAPT-Train	450	1637	1084	188
LAPT-Test	80	481	274	46

4.2 参数设置

对提出的模型进行训练并通过反向传播算法更新模型参数。通过 Glove 工具获得单词嵌入,并且将单词向量的维度设置为 300。该模型的整个超参数设置如表 2 所列。在 RC-NN 中,本文使用多窗口、多卷积核对划分区域进行卷积操作,以获得更丰富的特征信息。

表 2 参数设置
Table 2 Setting of parameters

参数名称	参数值
区域长度	30
卷积核窗口大小	2,3,4,5
每种卷积核数量	100
正则化强度 λ	1×10^{-4}
Mini-batch	35
Dropout	0.5
l_H	2
b_T	-2

对于高速公路网络层而言,所提模型中的重要超参数是公路网层数 l_H 和变换门的偏置 b_T 。这里,利用网格搜索来确定这两个超参数的最优值。我们人为地定义 l_H 的范围为

$\{1,2,\cdots,6\}$, b_T 的范围为 $\{-8.0,-7.0,\cdots,-1.0\}$ 。这两个超参数对模型平均精度 (Average Precision, AP) 的影响如图 3、图 4 所示。在 restaurants 领域评论数据集上进行测试,其中模型的其他超参数已被调整到最佳值。从图中可以看出 l_H 值为 2 且 b_T 值为 -2 时,模型的性能达到最优。其中,评价标准 AP 为积极样本、消极样本、中性样本 3 个样本精确度的平均值。

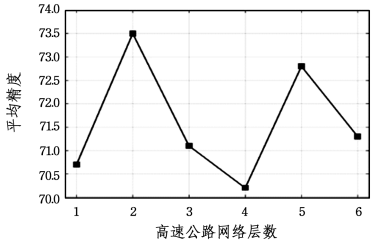


图 3 高速公路网络层数对平均精度的影响

Fig. 3 Impact of number of highway network layers on average precision

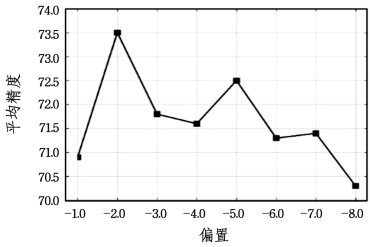


图 4 偏置对平均精度的影响

Fig. 4 Impact of bias on the average precision

4.3 对比实验

将本文提出的模型与典型的神经网络模型在 restaurants 和 laptops 这两个领域评论数据集上进行对比实验,以验证所提方法的有效性。

4.3.1 与其他网络模型的对比

本文使用 8 个网络模型在 restaurants 和 laptops 这两个领域评论数据集上进行三分类实验,结果如表 3 所列。

表 3 不同模型在 restaurants 和 laptops 领域数据集上的三分类结果
Table 3 Results of three classifications of different models on restaurants and laptops datasets
(单位: %)

模型	指标	数据集					
		EN-REST			EN-LAPT		
		Positive	Negative	Neutral	Positive	Negative	Neutral
CNN	P	91.64	65.45	39.68	82.54	66.16	33.18
	R	81.75	78.46	78.06	73.06	69.96	66.05
	F	86.41	71.37	52.61	77.51	68.01	44.17
LSTM	P	91.86	65.58	40.12	82.76	66.47	33.64
	R	82.02	78.89	78.13	73.28	70.86	66.68
	F	86.66	71.62	53.02	77.73	68.59	44.72
ATT-CNN	P	93.54	67.31	41.35	84.92	69.89	35.97
	R	83.14	80.76	80.43	76.03	74.53	68.99
	F	88.03	73.42	54.62	80.23	72.14	47.29
ATT-LSTM	P	92.89	70.18	47.93	84.96	72.29	42.12
	R	83.15	83.29	78.96	77.08	76.33	71.39
	F	87.75	76.18	59.65	80.83	74.26	52.98
RCNN-LSTM	P	93.08	69.04	49.65	86.04	70.18	43.68
	R	83.17	83.49	80.16	77.21	75.78	74.16
	F	87.85	75.58	61.32	81.39	72.87	54.98
HP-LSTM	P	92.98	70.05	51.12	85.27	71.15	43.89
	R	83.05	84.21	82.01	76.86	76.31	74.25
	F	87.83	76.48	62.98	80.85	73.64	55.17
RCNN-BGRU	P	93.31	69.53	54.60	85.63	71.22	45.60
	R	83.15	84.85	81.84	76.46	76.68	74.19
	F	87.94	76.43	65.50	80.79	73.85	56.48
RCNN-BGRU-HN	P	93.39	70.11	57.23	85.73	71.54	49.46
	R	83.60	85.36	82.03	77.21	76.71	78.29
	F	88.22	76.99	67.42	81.25	74.03	60.62

其中, P,R,F 分别代表精确率 (Precision)、召回率 (Recall) 和 F1 值。精确率是针对预测结果而言的,它表示预测为正的样本中有多少是真正的正样本。召回率是针对原样本而言的,它表示样本中的正例有多少被预测正确了。具体计算公式如下:

$$P=\frac{\text{True positives}}{\text{True positives}+\text{False positives}} \tag{21}$$

$$R=\frac{\text{True positives}}{\text{True positives}+\text{False negatives}} \tag{22}$$

$$F1=\frac{2 * P * R}{P + R} \tag{23}$$

从表 3 中可以看出,ATT-CNN 和 ATT-LSTM 网络模型中加入了特定方面的注意力机制,相比于没有加入注意力机制的 CNN 和 LSTM 网络模型具有更好的分类效果。例如,在 restaurants 领域数据集上,ATT-CNN 模型的积极样本的和消极样本的 F1 值相比于 CNN 模型分别提升了 1.62% 和 2.05%,从而验证了注意力机制在基于方面情感分析任务中的有效性。

RCNN-LSTM 和 HP-LSTM 网络模型中没有加入注意力机制,但是因为这两个模型考虑了评论文本中不同句子之间的相互联系,所以其结果优于以独立句子作为输入的

ATT-CNN 和 ATT-LSTM 网络模型。其中,RCNN-LSTM 网络模型在分类效果较好的 restaurants 领域数据集上 3 种样本的平均 $F1$ 值为 74.92%,相比于 ATT-CNN 和 ATT-LSTM 网络模型分别提升了 2.56%和 0.42%。这说明保留文本中句子的时序关系,考虑不同句子之间的相互联系,可以在训练过程中挖掘出句子间更深层次的依赖关系,从而在基于方面情感分析任务中表现更优。

从表 3 的实验结果还可以看出,本文提出的两个模型中,RCNN-BGRU-HN 模型相比于没有加入高速公路网络的 RCNN-BGRU 模型有更好的分类结果,在 restaurants 领域数据集上的平均 $F1$ 值提升了 0.92%。这也验证了高速公路网络层可以使网络模型在训练过程中有效地捕获单词之间更多的语义特征,从而可以在基于方面情感分析任务中取得更好的效果。

另外,RCNN-BGRU-HN 在所有任务中都优于其他模型,从而验证了本文所提模型的有效性。

4.3.2 网络训练时间的比较

本文分析对比了上述几种网络模型在 TensorFlow 框架下以相同的 CPU 和 GPU 在 restaurants 领域数据集上迭代一次所需的时间,结果如表 4 所列。

表 4 不同网络模型迭代一次所花的时间

Table 4 Time spent by iterating over different network models

网络模型	时间/s
CNN	6
LSTM	81
ATT-CNN	16
ATT-LSTM	126
RCNN-LSTM	40
HP-LSTM	72
RCNN-BGRU	36
RCNN-BGRU-HN	42

从表 4 中可以看出,与可以接收平行化输入的 CNN 模型相比,接收序列化输入的 LSTM 模型迭代一次需要花费更多的时间(5 倍之多),这说明 CNN 网络在训练时间上有更优的性能。加入了注意力机制的 ATT-CNN 网络迭代一次要花费 16 s 的时间,比 CNN 多 10 s,这是因为网络模型的加深增加了训练时间。而 ATT-LSTM 所消耗的时间最长,为 126 s,这说明以独立句子作为模型输入也会增加模型训练的时间。本文提出的 RCNN-BGRU,相比于 RCNN-LSTM,虽然多加入了注意力机制层,但因为 BGRU 结构更简单,且计算复杂度更低,所以其训练时间也减少了 4 s。本文所提另一种网络模型 RCNN-BGRU-HN 完成一次迭代需要 42 s,对比以往效果较好的 HP-LSTM 模型的 72 s,减少了 30 s,这说明结合了 CNN 和 BGRU 的网络模型能有效地缩短训练时间,从而验证了本文所提方法有着更优的训练性能。

结束语 本文针对基于方面情感分析任务提出的一种新型结构的深度学习网络模型是十分有效的。该模型通过区域 CNN 提取整个评论文本中重要的情感特征信息,且利用 BGRU 和高速公路网络使模型能捕获文本中更多的语义信息。此外,注意力机制的加入可以使得模型在训练过程中高度关注特定方面在整个评论文本中的长距离依赖关系。实验结果表明,本文模型在基于方面情感分析任务中表现得十分优异,可以实现端到端的训练,适应性较好,且有着更优的时间性能。

参 考 文 献

[1] CAMBRIA E,DAS D,BANDYOPADHYAY S,et al. A Practical Guide to Sentiment Analysis[M]. Berlin:Springer International Publishing,2017.

[2] DAS S R,CHEN M Y. Yahoo! For Amazon:Sentiment Extraction from Small Talk on the Web[J]. Management Science, 2007,53(9):1375-1388.

[3] PANG B,LEE L,VAITHYANATHAN S. Thumbs up?:sentiment classification using machine learning techniques[EB/OL]. <http://www.doc88.com/p-1416919772501.html>.

[4] MNIHV,HEESS N,GRAVES A. Recurrent models of visualattention[C]// Proceedings of Advances in Neural Information Processing Systems 27 (NIPS 2014). Cambridge, MA: MIT Press,2014:2204-2212.

[5] BAHDANAU D,CHO K,BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv:1409.0473,2014.

[6] YIN W,SCHÜTZ H,XIANG B,et al. Abcnn:Attention-based convolutional neural network for modeling sentence pairs[J]. Transactions of the Association for Computational Linguistics, 2016,4(11):259-272.

[7] WANG Y,HUANG M,ZHAO L,et al. Attention-based LSTM for aspect-level sentiment classification[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, USA,2016:606-615.

[8] KIRITCHENKO S,ZHU X,CHERRY C,et al. NRC-Canada-2014:Detecting Aspects and Sentiment in Customer Reviews[C]// International Workshop on Semantic Evaluation. Dublin: Ireland:Association for Computational Linguistics,2014:437-442.

[9] HU M,LIU B. Mining and summarizing customer reviews[C]// Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM,2004:168-177.

[10] PORIA S,CAMBRIA E,GELBUKH A. Aspect extraction for opinion mining with a deep convolutional neural network[M]. Amsterdam:Elsevier Science Publishers B. V.,2016:42-49.

[11] TANG D,QIN B,LIU T. Aspect Level Sentiment Classification with Deep Memory Network[J/OL]. <https://arxiv.org/abs/1605.08900>.

[12] LECUN Y,BOTTOU L,BENGIO Y,et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE,1998,86(11):2278-2324.

[13] HOCHREITER S,SCHMIDHUBER J. Long short-term memory[J]. Neural Computation,1997,9(8):1735-1780.

[14] CHUNG J,GULCEHRE C,CHO K H,et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. arXiv:1412.3555,2014.

[15] WANG J,YU L C,LAI K R,et al. Dimensional sentiment analysis using a regional cnn-lstm model[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin,Germany,2016:225-230.

[16] KIM Y,JERNITE Y,SONTAG D,et al. Character-Aware Neural Language Models[J/OL]. <https://arxiv.org/abs/1508.06615>.