

# 一种用于基于方面情感分析的深度分层网络模型

刘全<sup>1),2),3)+</sup> 梁斌<sup>1)</sup> 徐进<sup>1)</sup> 周倩<sup>1)</sup>

<sup>1)</sup>(苏州大学计算机科学与技术学院 江苏 苏州 215006)

<sup>2)</sup>(软件新技术与产业化协同创新中心 南京 210000)

<sup>3)</sup>(吉林大学符号计算与知识工程教育部重点实验室 长春 130012)

**摘要** 近年来,基于方面情感分析已成为自然语言处理领域的研究热点之一。结合注意力机制的深度网络模型在基于方面情感分析任务中取得了令人瞩目的成功,针对以独立句子作为网络模型输入的方法无法获取句子间相互关系,以及仅使用词语层注意力机制难以充分获取同一评论中句子间的相互联系等问题,提出一种结合区域卷积神经网络和分层长短记忆网络(regional convolutional neural network-hierarchical long short-term memory, RCNN-HLSTM)的深度分层网络模型用在基于方面情感分析任务中。该模型通过区域 CNN 既可以保留不同句子在评论中的时序关系也可以大大降低仅使用 LSTM 网络的时间代价。此外,该模型利用一个分层 LSTM 网络来获取待分类句子内部词语之间的相互联系,以及待分类句子和评论中其他句子之间的情感特征信息。通过词语层和句子层注意力机制能有效获取特定方面在句子中的局部特征和整个评论中的长距离依赖关系,弥补了仅使用词语层注意力机制的不足。最后在多种语言的不同领域数据集上进行实验,取得了比传统的深度网络模型、结合注意力机制的深度网络模型以及考虑句子间关系的双向分层 LSTM 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; 分层模型; 循环神经网络; 卷积神经网络; 注意力机制

中图法分类号 TP18

## A Deep Hierarchical Neural Network Model for Aspect-based Sentiment Analysis

LIU Quan<sup>1),2),3)</sup> LIANG Bin<sup>1)</sup> XU Jin<sup>1)</sup> ZHOU Qian<sup>1)</sup>

<sup>1)</sup>(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu, 215006)

<sup>2)</sup>(Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing, 210000)

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

**Abstract** Aspect-based sentiment analysis has become one of the research hotspots in the field of natural language processing (NLP) in recent years. Different from ordinary sentiment analysis, aspect-based sentiment classification is a fine-grained task of sentiment analysis in the field of NLP, which need to infer different sentiment polarity of different aspects in the same sentence, because there will be more than one aspect in the same sentence usually. Previous studies, in general, usually consider the independent sentence as the input of neural networks and only focus on the given aspect in each sentence in the process of training. These approaches, however, will ignore the long-distance dependency of the given aspect in the entire long text and cannot take full

本课题得到国家自然科学基金项目(61272005, 61303108, 61373094, 61472262, 61502323, 61502329, 61772355)、江苏省自然科学基金(BK2012616)、江苏省高校自然科学基金项目(13KJB520020, 16KJB520041)、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)、苏州市应用基础研究计划工业部分(SYG201422, SYG201308)资助。刘全(通讯作者),男,1969年生,博士,教授,博士生导师,中国计算机协会(CCF)高级会员,主要研究领域为智能信息助理、自动推理和机器学习。E-mail: quanliu@suda.edu.cn。梁斌,男,1993年生,硕士研究生,主要研究方向为情感分析、自然语言处理和深度学习。E-mail: bliang@stu.suda.edu.cn。徐进,男,1992年生,硕士研究生,主要研究方向为强化学习、深度学习和深度强化学习。E-mail: 20154227016@stu.suda.edu.cn。周倩,女,1992年生,硕士研究生,主要研究方向为强化学习。E-mail: 20154227029@stu.suda.edu.cn。

advantage of the context relations of different sentences across the same review, which is bad for those ambiguous sentences and short sentences of a review in the training process. To address these problems, this paper proposes a hierarchical model of combining regional convolutional neural network and hierarchical long short-term memory (HRCNN-LSTM) for the task of aspect-based sentiment classification on long text customer review. This approach can extract more feature information of independent sentence and the relations of different sentences in the whole review via combining regional CNN and hierarchical LSTM, and is able to infer the sentiment polarity of different aspects discriminatively without any external information such as semantic dependency parsing. We divide a long text review into several regions based on different targets of aspect in the sentences, and then we utilize a regional CNN to receive different independent regions of the review to extract the information across the entire review. This regional CNN is able to capture the long-distance dependency of the concerned aspect across the whole review and keep the order of different regions, as well as, save the training time of using LSTM network only. Meanwhile, dividing regions based on different targets can assist our model to discriminate different sentiment polarity of different aspects in the same sentence better. On the other hand, we present a hierarchical LSTM combined with regional CNN to concentrate on both word-level and sentence-level information through a hierarchical attention mechanism. Aspect embedding is considered as a word-level attention and combined with word embedding as a sequential input and fed into a word-level LSTM to focus on the given aspect in the process of training and generate a sentence-level attention. The final output of word-level LSTM is combined with the extracted feature information from regional CNN and fed into a sentence-level LSTM. Such hierarchical word-level and sentence-level attention between regional CNN and hierarchical LSTM can capture more in-depth information from sentence and the entire review as well as consider both intra-sentence and inter-sentence relations in the prediction process, which provide a good ability to discriminate the sentiment polarity of short and ambiguous sentences. Finally, experimental results on multi-domain datasets of two languages from SemEval2016 show that, our approach yields better performance than several competitor models on aspect-based sentiment classification with word vectors only.

**Keywords** deep learning; aspect-based sentiment analysis; hierarchical model; recurrent neural network; convolutional neural network; attention mechanism

## 1 引言

随着社交网络的日益繁荣,如何从社交网络的文本信息中挖掘用户的情感倾向已得到越来越多研究人员的关注<sup>[1]</sup>。基于方面情感分析作为更深层次的情感分析,可以针对不同的目标挖掘用户更细腻的情感表述,近年来已成为自然语言处理(natural language processing, NLP)领域的研究热点之一<sup>[2-3]</sup>。和普通情感分析不同,基于方面情感分析需要正确判断句子中不同方面的情感极性,这不仅依赖于文本的上下文信息,同时也要考虑不同方面的情感信息<sup>[4]</sup>,例如句子“Good food but dreadful service at that restaurant”,在同一个句子中对于方面“food”是积极情感,而对于方面“service”则是消极情感。所以在同一个句子中针对不同的方面分析其情感极性可以更加有效地帮助人们了解用户对不同方面的情感表达。

在过去的研究中,很多学者使用传统的机器学习方法来解决文本分类问题。这类方法通过人工标注一部分数据作为训练集,然后对训练集上的数据进行特征提取和学习来构建文本分类模型,最后使用该模型来对未标注的数据进行预测,自动实现文本的分类<sup>[5]</sup>。基于机器学习的方法在过去的研究中已取得了很多成果,但这类方法通常需要依赖复杂的人工规则和特征工程。

近年来,深度学习已在NLP领域取得了令人瞩目的成功,也被广泛应用于自然语言处理任务中<sup>[6]</sup>。同时,结合注意力机制的神经网络模型在基于目标的任务中取得了比传统方法更好的效果。例如基于特定方面的关系分类<sup>[7]</sup>,基于特定目标的句子对建模<sup>[8]</sup>,以及基于特定目标的机器翻译<sup>[9]</sup>。这类结合注意力机制的神经网络模型可以在训练过程中高度关注特定目标的特征信息,并且能有效针对不同的目标调整神经网络的参数信息,挖掘更多的

隐藏特征信息，从而可以取得比传统方法更好的效果。但这类方法通常是以独立的句子作为输入，在长文本的评论中忽略了句子之间的关系，且在训练过程中只能高度关注特定目标本身，无法有效识别情感极性模糊、句子长度较短、含有隐含情感等句子的情感极性。

在同一个用户评论中不同句子往往有着相似的情感，如图 1 所示。在同一个评论中，3 个独立的句子针对不同的方面情感极性都为积极情感，说明同一用户针对不同方面的评论信息会有相似的情感极性。句子之间的相互联系有助于分析特定方面的情感极性，但以独立句子作为输入的深度网络模型忽略了同一评论中不同句子之间的联系。此外，用户评论中往往存在一些句子长度较短或者情感极性表达不清晰的句子，这类句子的情感特征信息较为有限，只依赖独立的句子输入和单一的注意力机制并不能很好地判别这类句子的情感极性。针对这些问题，本文提出一种结合区域卷积神经网络和分层 LSTM 网络的深度分层模型来解决基于特定方面的情感极性分析问题。该模型通过挖掘评论中的上下文句子关系，有效弥补了以独立句子作为输入的不足，使模型可以挖掘特定方面在整个评论的长距离依赖关系，从而更充分地利用整个评论中的特征信息，有效识别情感极性表达不清晰等句子的情感极性。同时，该模型通过词语层和句子层的分层注意力机制，可以更加有效地识别句子中不同方面的情感极性。

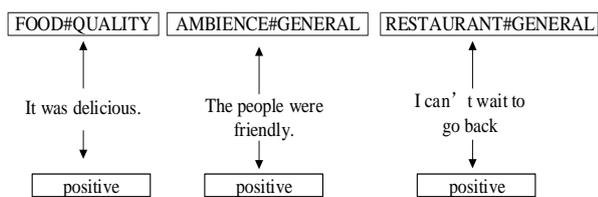


图 1 评论中不同句子相互联系举例

本文主要贡献如下：

1) 提出了一种区域 CNN 和 LSTM 网络结合的模式。该模型既保留了整个评论中的内容信息和句子间的时序关系，同时也有效降低仅使用 LSTM 网络的训练时间。此外，该模型在无需额外的依存关系分析的情况下可以有效获取长文本评论中句子内部和句子之间的相互联系，同时也可以获取特定方面在句子中和在整个评论中的依赖关系。有效解决了情感极性表达不清晰、句子长度较短等难句的情感极性判断。

2) 引入了一种基于特定方面的目标词来划分区域的方法。该方法可以将同一个句子基于不同的方面进行展开，使每一个独立区域都含有不同方面的重要特征信息。从而使网络模型在训练过程中可以针对不同方面所在的区域调整参数信息，有效保留特定方面的重要情感信息。

3) 介绍了一种注意力机制和分层 LSTM 网络结合的方法。该方法可以使模型在训练过程中通过词语层和句子层注意力机制来获取句子中不同方面在句子内部和句子之间的相互联系，有效解决了短句和情感表达不清晰等句子情感极性判断困难的问题。在分类正确率最高的 restaurants 领域数据集上的二分类实验的平均 F1 值为 84.37%，相比以往研究中表现最好的 HP-LSTM 模型提升了 1.74%。

本文第 2 节介绍基于方面情感分析的相关工作以及卷积神经网络 (convolutional neural network, CNN)、长短记忆网络 (long short-term memory, LSTM) 和注意力机制 (attention mechanism) 的基本原理；第 3 节重点描述本文提出的 RCNN-HLSTM 模型；第 4 节是本文的对比实验以及对实验结果的详细描述和分析；最后是本文工作的总结和未来展望。

## 2 相关工作

### 2.1 基于方面情感分析

在过去的一些研究中，传统的机器学习方法在关联目标的情感分析任务中取得了成功<sup>[10-11]</sup>，但这类方法通常需要对输入文本进行大量的预处理和复杂的特征工程，以及如依存关系分析等外部知识。近年来，深度学习在文本分类任务中取得了重大的突破。使用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分类效果。同时，基于神经网络的方法大大缓解了模型对特征工程的依赖，使模型用更低的代价取得更好的分类效果。在基于深度神经网络的方法中，Dong 等人<sup>[12]</sup>提出一种自适应递归神经网络 AdaRNN 模型，并加入语义结构和上下文信息来学习目标和其他词语的情感关系。该模型通过自适应的深度网络来高度关注特定目标的情感特征信息，以及挖掘特定目标更丰富的语义信息和情感特征，并通过依存句法分析的结果来获取其他词语和特定目标的依赖关系。Nguyen 和 Shirai 结合递归神经网络 (recursive neural network, RNN) 和依存树

提出一种 Phrasernn 模型<sup>[13]</sup>。该模型可以针对不同方面构建二叉短语依存树来表示特定方面的特征信息,挖掘更丰富的情感信息,从而可以有效识别句子中不同方面的情感极性。这类基于深度神经网络的模型与传统的机器学习方法相比,大大降低预处理和特征工程的工作量,并且不需要使用人工规则和情感词典。但这类方法仍需要结合一些如依存句法分析、依存关系树等外部知识。

基于注意力机制的深度神经网络模型在基于方面情感分析任务中取得了比上述结合依存句法分析等外部知识的网络模型更好的分类性能。这类基于注意力机制的方法可以使模型在训练过程中高度关注特定方面的特征信息,有效识别不同方面和句子中词语的相互关系,挖掘更多的隐藏特征,取得了更好的情感分类效果。如 Wang 等人<sup>[14]</sup>对不同的方面进行特定的向量化处理,使用基于注意力机制的 LSTM 网络对特定方面进行情感分析。该方法使用一个作用在 LSTM 隐藏层的注意力机制可以有效地将特定方面的向量信息结合到每一个 LSTM 网络单元,从而可以使网络在训练过程中针对特定的方面来学习和调整网络模型的参数信息,取得了比以往方法更好的情感分类效果。此外, Tang 等人<sup>[15]</sup>使用一种基于注意力机制的深度记忆网络用在基于方面的情感分析任务中。该方法使用内容层面注意力机制和特定方面在句子中的位置注意力机制,使网络模型在训练过程中可以通过多方面信息关注输入文本的特定目标的情感特征信息,有效弥补了单一注意力机制的不足。这些结合注意力的深度神经网络模型在无需额外的语义分析等外部知识的情况下,在基于方面情感分析任务中取得了比传统神经网络更好的分类效果。但这类方法通常是以独立的句子作为网络的输入忽略了同一个评论中不同句子之间的相互联系。

在长文本评论的基于方面情感分析任务中, Ruder 等人<sup>[16]</sup>提出一种分层双向 LSTM 网络模型,该模型可以充分考虑不同句子在评论中的相互联系,同时也保留了特定方面在句子中和词语之间的联系,有效提取句子内部和句子之间的特征信息。受文献[16]启发,本文提出一种将区域 CNN 和分层 LSTM 网络结合的深度网络模型用在基于方面的情感分析任务中。相比文献[16],本文利用句子层的 LSTM 网络和区域 CNN 结合既保留了不同句子的时序关系,同时也可以通过对区域 CNN 挖掘特定方面在整个评论中的长距离依赖关系,有效提高情感

分类的性能。

## 2.2 卷积神经网络

卷积神经网络 (CNN) 作为深度学习的重要网络之一,近年来已被广泛应用于众多领域,并取得了令人瞩目的成功<sup>[6]</sup>。CNN 主要通过卷积层和池化层来学习输入的局部特征和提取重要的特征信息<sup>[17-19]</sup>。在自然语言处理任务中, CNN 无需对文本进行大量的预处理工作,有效减轻了特征工程的工作量。如图 2 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:

$$E \in \mathbf{R}^{n \times m} \quad (1)$$

其中  $m$  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,如式 (2) 所示:

$$c = f(\mathbf{W} \cdot \mathbf{x} + b) \quad (2)$$

其中  $\mathbf{x}$  为卷积核窗口词向量矩阵,  $\mathbf{W}$  为权重矩阵,  $b$  为偏置,  $f$  为卷积核激活函数。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,对于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。

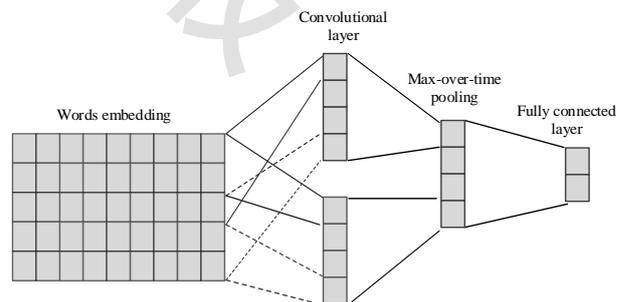


图 2 卷积神经网络模型结构图

## 2.3 长短记忆网络

长短记忆网络 (LSTM) 是对循环神经网络 (recurrent neural network, RNN) 的改进,可以有效解决 RNN 网络模型在训练过程中的梯度消失问

题。在 NLP 任务中, LSTM 网络可以学习词语在句子中的长期依赖关系<sup>[20-22]</sup>。

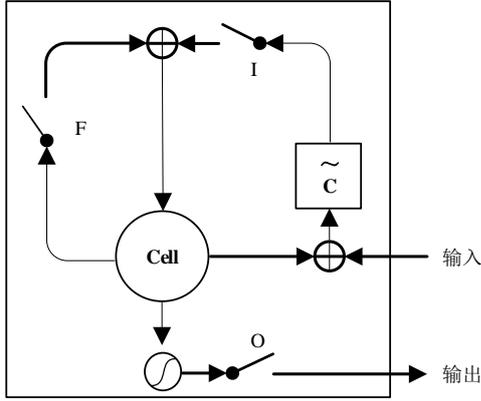


图 3 LSTM 网络单元结构

如图 3 所示, LSTM 网络通过一个记忆单元来记忆存储输入句子的重要特征信息, 同时也可以遗忘不重要的信息。每一个 LSTM 网络神经元包含核心元素 Cell 和 3 个门单元。其中 F 为遗忘门, I 为输入门, O 为输出门。

LSTM 单元对输入信息的计算如式 (3) - (8) 所示:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (3)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (4)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1}) \quad (6)$$

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \quad (7)$$

$$h_t = o_t \tanh(c_t) \quad (8)$$

其中  $f_t$  为 LSTM 单元  $t$  时刻遗忘门的输出,  $i_t$  为输入门的输出,  $o_t$  为输出门的输出。 $W_f$ 、 $W_i$ 、 $W_o$ 、 $b_f$ 、 $b_i$  和  $b_o$  分别为遗忘门、输入门和输出门的权重矩阵和偏置项。LSTM 单元可以用式 (6) 计算得到新的上下文记忆信息  $\tilde{c}_t$ , 并通过式 (7) 更新 LSTM 网络单元的记忆内容  $c_t$ 。 $h_{t-1}$  为  $t-1$  时刻的隐藏层输出,  $t$  时刻的隐藏层输出可由式 (8) 计算得到。

## 2.4 注意力机制

结合注意力机制的深度神经网络最早是在图像处理领域中提出, 目的是为了网络模型在训练过程中高度关注特定目标的特征信息<sup>[23]</sup>。在自然语言处理 (NLP) 领域中, Bahdanau 等人<sup>[9]</sup>将注意力机制和 RNN 结合用在机器翻译任务中, 证明了注

意力机制在 NLP 任务中的有效性。Wang 等人<sup>[24]</sup>将多层注意力机制和卷积神经网络 (CNN) 结合用在句子关系分类任务中, 该模型利用作用在输入层和池化层的多层注意力机制来针对不同的实体目标进行特征提取, 使网络可以有效区分不同实体目标的特征信息, 更有效地学习输入的句子信息, 最终在多个数据集上的实验结果证明使用注意力机制的模型相比传统网络有更好的关系分类效果。在基于方面的情感分析任务中, 网络模型需要在训练过程中针对不同方面来提取特征信息, 所以网络模型需要高度关注句子中特定的方面。结合注意力机制的深度网络模型能很好地解决句子中不同方面的情感极性判别问题。文献[14]提出了一种基于注意力机制的 LSTM 网络模型, 该模型将特定方面的向量和上一神经元的隐藏层输出结合作为当前神经元的输入, 使模型中每一个神经元都能学习特定方面的情感信息, 从而在训练过程中可以高度关注该方面的特征信息, 有效针对该方面来学习输入句子的情感特征。最终在不同领域的数据集上的实验结果也证明了结合注意力机制的深度网络能取得更好的情感分类效果。文献[15]将 2 种注意力机制加入深度记忆网络来构建网络模型, 该模型在基于方面的情感分类任务中能通过不同的注意力机制来关注特定方面的情感信息, 针对不同方面来挖掘输入文本的特征信息。该方法除了使用常用的内容层面注意力机制之外, 还加入了词在句子中的位置注意力机制, 通过在不同数据集上进行实验验证了使用内容层面和位置注意力机制的有效性。该方法进一步证明了注意力机制在基于方面情感分析任务中的有效性。

## 3 一种结合区域 CNN 和分层 LSTM 的深度分层网络模型

本文提出的 RCNN-HLSTM 网络模型是基于文献[16]提出的分层网络模型的改进, 如图 4 所示, 一个待分类句子在网络中的训练框架主要由以下 3 部分组成:

1) 区域 CNN: 本文将一个用户评论按不同方

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;

2) 词语层 LSTM 网络: 将特定方面的向量和隐藏层输出结合作为词语层 LSTM 网络的序列化输入, 使网络在训练过程中高度关注特定方面的情感特征信息;

3) 句子层 LSTM 网络: 区域 CNN 和词语层 LSTM 网络的输出结合形成句子层 LSTM 网络的序列化输入, 获取特定方面在整个评论中的依赖关系。

本文提出的 RCNN-HLSTM 模型是以整个评论作为网络的输入, 既可以获取同一评论中不同句子间的相互联系, 也能降低模型的训练时间。对于当前时刻的待分类句子, 词语层的 LSTM 可以针对该句的特定方面挖掘词语间的特征联系。因为在整个评论中, 特定方面所在句子的词语往往包含对特定方面最重要的特征信息, 词语层 LSTM 可以有效挖掘特定方面在句子中的局部特征。此外, 该模型利用区域 CNN 来提取整个评论中不同区域的局部特

征信息, 并将这些特征信息按顺序输入到句子层 LSTM 网络中, 可以保留整个评论的时序关系。区域 CNN 可以接收句子的平行话输入, 同时也可以在整个评论的句子中共享特征信息, 通过这种方法, 可以进一步降低网络的训练时间。句子层 LSTM 网络可以利用整个评论的特征信息来获取特定方面在整个评论中的长距离依赖关系。在实际的长文本用户评论中, 某些句子的情感极性表达不清晰或句子长度较短, 仅利用词语层 LSTM 网络难以获取足够的特征信息来对特定方面进行情感极性判断。而同一评论中的不同句子之间往往有密切的情感联系, 所以本文将句子层 LSTM 提取得到的特征信息和词语层 LSTM 网络结合起来对特定方面进行情感极性的判断, 通过句子层 LSTM 网络可以挖掘整个评论的特征信息, 从而可以挖掘特定方面更深层的情感特征信息, 在词语层 LSTM 无法获取足够特征信息时也能协助网络模型对当前句子进行情感分类, 有效判断特定方面的情感极性。

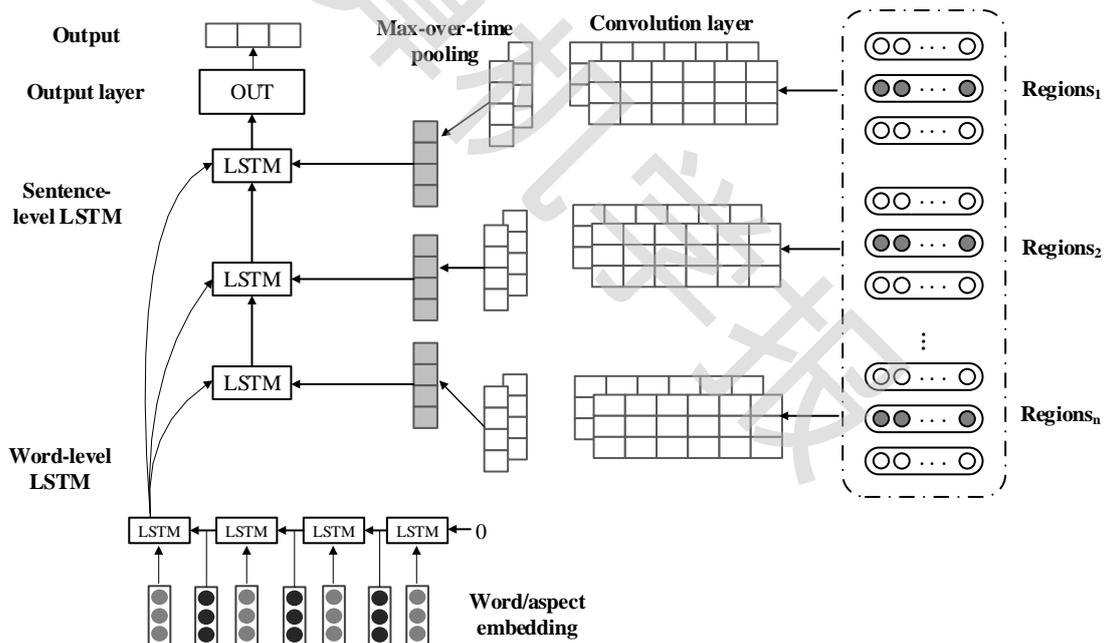


图4 结合区域 CNN 和分层 LSTM 的深度分层网络模型

### 3.1 任务定义

给定一个句子  $s = \{w_1, w_2, \dots, t, \dots, w_n\}$ , 即每个句子由一系列的词语  $w_i$  组成, 其中  $t$  是句子  $s$  中特定方面的目标词, 每个句子有 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 FOOD#QUALITY。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“Good food but dreadful service at that restaurant”, 该句子有 2 个目标词“food”和“service”, 分别对应 2 个方面“FOOD#QUALITY”和“SERVICE#GENERAL”, 在同一个句子中, 对于方面“FOOD#QUALITY”是积极情感, 而对于方面

“SERVICE#GENERAL”则是消极情感。本文将每一个方面映射为一个多维的连续值向量，即  $\mathbf{a} \in \mathbf{R}^m$ ，其中  $\mathbf{a}$  为实体和属性的词向量平均值， $m$  为词向量维度。

### 3.2 区域卷积神经网络

区域卷积神经网络通过把一个包含多个句子的长文本用户评论分割为若干个独立的区域，既可以保留文本中不同句子的时序关系，同时也可以获取特定方面在整个文本中的长距离依赖关系<sup>[25]</sup>。文献[25]是以句子作为独立的区域，和文献[25]不同，本文提出一种基于句子中不同方面的目标词来划分区域的方法。对于每一个句子，本文将重点关注句子中不同方面的目标词和目标词周围的词语，并基于不同的目标词对句子进行区域划分。例如句子  $s = \{w_1, w_2, \dots, t_1, \dots, t_2, \dots, w_n\}$ ，有  $t_1$  和  $t_2$  两个目标词，本文将该句子按照两个不同的目标词划分成 2 个长度为  $h$  的独立区域  $r_1 = \{w_i, w_{i+1}, \dots, t_1, \dots, w_{h+i-2}\}$  和  $r_2 = \{w_j, w_{j+1}, \dots, t_2, \dots, w_{h+j-2}\}$ 。对于没有目标词的句子，直接将该句子划分为一个长度为  $h$  的区域。基于不同方面的目标词来划分区域可以有效根据不同的目标词来区分不同方面在句子中的情感特征信息，使模型在训练过程中可以有效识别同一句子中不同方面的情感极性。

对于长度为  $h$  的独立区域  $r_i$ ，本文将区域中每一个词语映射为一个  $m$  维的连续值向量，得到每一个区域的矩阵表示  $r_i = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_h]$ ，其中  $\mathbf{x}_i \in \mathbf{R}^m$ 。将一个评论中的每一个独立区域作为卷积神经网络的输入矩阵，然后使用长度为  $l$  的卷积核对区域做卷积操作，如式 (9) 所示：

$$c_i = f(\mathbf{w} \cdot r_i + b) \quad (9)$$

其中  $\mathbf{w} \in \mathbf{R}^{m \times l}$  为卷积核权重， $b \in \mathbf{R}$  为偏置项。对每一个区域通过卷积操作可以得到区域  $r_i$  的特征图  $\mathbf{c} \in \mathbf{R}^{h-l+1}$ ，如式 (10) 所示：

$$\mathbf{c} = [c_1, c_2, \dots, c_{h-l+1}] \quad (10)$$

对于每一个区域，本文采用 max-over-time pooling 方法进行局部特征的下采样，提取每一个特征图  $\mathbf{c}$  最重要的特征信息，即  $\hat{\mathbf{c}} = \max\{\mathbf{c}\}$ 。对于有  $k$  个卷积核的窗口可以得到如式 (11) 所示的下采样

特征信息图。

$$\hat{\mathbf{c}} = [\hat{c}_1, \hat{c}_2, \dots, \hat{c}_k] \quad (11)$$

### 3.3 分层 LSTM 网络

为了获取更丰富的上下文情感特征信息，本文使用一个分层 LSTM 网络分别接收词语层和句子层的序列化输入，通过词语层和句子层的注意力机制使网络在训练过程中既可以关注特定方面在句子内部的相互关系，也可以获取特定方面在整个评论中的长距离依赖关系。

#### 3.3.1 词语层 LSTM 网络

基于方面情感分析需要网络模型在训练过程中高度关注不同方面的特征信息，词语层 LSTM 接收一个以词语为单位的句子作为网络的序列化输入，每一个 LSTM 单元的输入由上一个单元隐藏层的输出和本次输入的词语向量组成。词语层 LSTM 网络可以针对特定方面提取其在句子中的局部特征信息。同时，通过特定方面的注意力机制可以让网络在训练过程中高度关注特定方面在句子中的情感信息。如图 5 所示，是本文提出的词语层 LSTM 网络结构图。

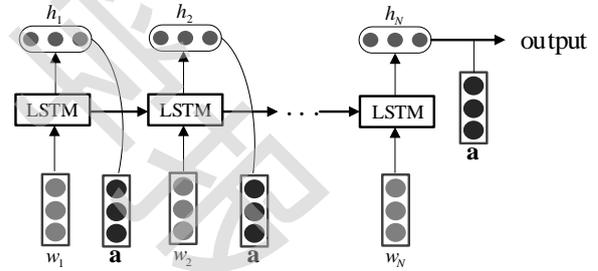


图 5 词语层 LSTM 结构图

对于长度为  $N$  的句子，可以得到隐藏层输出矩阵  $H = [h_1, h_2, \dots, h_N]$ ，即  $H \in \mathbf{R}^{d \times N}$ ，其中  $h_i \in \mathbf{R}^d$ ， $d$  为 LSTM 单元输出向量的维度。为了使网络在训练过程中可以高度关注待分类句子特定方面的特征信息，本文将特定方面向量  $\mathbf{a}$  和网络的上一个神经单元的隐藏层输出  $h_i$  结合作为词语层 LSTM 单元的输入，如式 (12) 所示：

$$\mathbf{e} = W_h \cdot h_i + W_a \cdot \mathbf{a} \quad (12)$$

其中  $W_h \in \mathbf{R}^{m \times d}$  为隐藏层输出  $h_i$  的权重矩阵， $W_a \in \mathbf{R}^{m \times m}$  为特定方面向量  $\mathbf{a}$  的权重矩阵。网络在训练过程中可以通过调整每一个输入向量来调整特

定方面的参数信息,从而可以使网络在训练过程中高度关注特定方面的特征信息。

### 3.3.2 句子层 LSTM 网络

词语层 LSTM 网络可以有效获取特定方面和句子内部词语之间的相互关系,但是对于情感极性模糊和句子长度较短的句子,其所含的特征信息非常有限,仅仅使用词语层的 LSTM 网络难以正确识别这类句子的情感极性。此外,同一评论中不同句子往往有相似的情感极性。为了挖掘待分类句子和评论中其他句子的相互联系,本文使用一个句子层 LSTM 网络来进一步挖掘句子之间的依赖关系。将区域 CNN 的输出按句子的先后顺序形成序列化矩阵,和词语层 LSTM 网络的最终输出结合,可以得到句子层 LSTM 网络每一个单元的输入。计算方法如式 (13) 所示:

$$s = h_N \oplus \hat{c}_i \quad (13)$$

其中  $h_N$  为词语层 LSTM 网络最后一层 LSTM 单元的输出,  $\hat{c}_i$  为区域卷积神经网络第  $i$  个下采样特征向量,  $\oplus$  为简单拼接操作。

### 3.4 模型训练

本文使用一个 *soft max* 函数接收句子层 LSTM 网络的输出,得到待分类句子情感极性的结果输出,如式 (14) 所示:

$$y = \text{soft max}(Wh_s + b) \quad (14)$$

其中  $W$  为权重矩阵,  $b$  为偏置。本文使用反向传播算法来训练网络模型,训练过程中,通过全连接层的每一次输出最小化交叉熵来优化模型,交叉熵公式由式 (15) 给出:

$$\text{loss} = -\sum_{i \in D} \sum_{j \in C} \hat{y}_i^j \log y_i^j + \lambda \|\theta\|^2 \quad (15)$$

其中  $D$  为训练集数据量,即训练集合大小,  $C$  为数据的类别数,  $y$  为待分类句子的预测类别,  $\hat{y}$  为实际类别,  $\lambda \|\theta\|^2$  为交叉熵正则项。

## 4 实验及结果分析

本文提出一种结合区域卷积神经网络和分层 LSTM 网络的深度分层网络模型用在基于方面情感分析任务中,在 2 种语言(英文和中文)和 4 个领域(restaurants、laptops、cameras 和 phones)的数据集上进行实验,验证本文提出方法的有效性。

### 4.1 实验数据

本文采用 SemEval2016 任务 5 数据集<sup>1</sup>中的 restaurants、laptops、cameras 和 phones, 4 个领域数据集进行实验<sup>2</sup>。每个领域中的数据由若干个用户评论组成,每个用户评论包含若干个句子,每个句子中含有 1 个或多个不同的方面,不同方面对应 0 个或多个目标词。基于方面情感分析任务需要根据句子的特征信息,在同一句子中针对特定的方面判断其情感极性。其中,restaurants 和 laptops 领域中的数据有积极、消极和中性 3 种情感极性。cameras 和 phones 领域数据有积极和消极 2 种情感极性。实验数据统计如表 1 所示。

表 1 实验数据统计

语言	数据集	评论数	积极句子	消极句子	中性句子
EN	REST-Train	350	1657	749	101
	REST-Test	90	611	204	44
	LAPT-Train	450	1637	1084	188
	LAPT-Test	80	481	274	46
CH	PHNS-Train	140	758	575	0
	PHNS-Test	60	310	219	0
	CAME-Train	140	809	450	0
	CAME-Test	60	344	137	0

### 4.2 实验参数设置

在本文的实验中,英文数据集采用 Penning 等人<sup>[26]</sup>提出的 Glove<sup>2</sup>词向量来初始化实验数据的词向量,中文词向量采用 Leipzig Corpora Collection<sup>3</sup>词向量进行初始化。其中每个词向量为 300 维的连续值,采用均匀分布  $U(-0.1,0.1)$  对未登录词进行词向量初始化。本文使用 ICTCLAS 分词工具<sup>4</sup>对中文数据集进行分词处理。在区域 CNN 中,为了获取丰富的特征信息,本文使用多窗口,多卷积核对独立区域进行卷积操作,其中窗口大小分别为 2、3、4、5,每种窗口的个数均为 100。每个区域的固定长度设置为 30,即长度超过 30 的句子,本文基于特定方面的目标词对句子进行切割,对于长度小于

<sup>1</sup> 本文使用 restaurants、laptops、cameras 和 phones 领域的 4 个数据集进行实验,在实验中本文去除没有特定目标的样本和句子属性标志为“OutOfScope”的样本。

<sup>2</sup> <http://nlp.stanford.edu/projects/glove/>

<sup>3</sup> <http://corpora2.informatik.uni-leipzig.de/download.html/>

<sup>4</sup> <http://ictclas.nlpir.org/>

50 的句子，本文以 0 向量填补句子的输入矩阵。此外，为了防止过拟合，本文在实验中使用了 dropout 机制和权重的正则化限制。由于实验的训练数据集不大，所以本文在实验中使用了较小的 mini-batch。区域 CNN 在实验中使用的详细参数如表 2 所示。

表 2 参数设置

超参数	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
正则项限制	3
Mini-batch	32
Dropout	0.5

### 4.3 对比实验

将本文提出的深度分层网络模型和典型的神经网络模型在 4 个领域的数据集上进行对比实验，验证本文提出方法的有效性：

1) CNN-HLSTM：本文提出的深度分层网络模型。使用分层 LSTM 网络可以获取特定方面在句子内部的依赖关系，同时也可以获取不同句子之间的相互联系。但该模型没有基于方面的目标词对句子进行区域划分，仅仅以独立的句子作为 CNN 的输入，无法有效针对不同方面的目标词在同一句子中进行展开。

2) RCNN-HLSTM：本文提出方法的完整模型，基于特定方面的目标词对输入句子进行区域划分，使模型在训练过程中可以针对不同方面的目标词调整和优化模型参数，有效识别不同方面在句子内部的情感特征信息和在整个评论中句子之间的长距离依赖关系。

3) CNN：基于文献[18]提出的卷积神经网络模型，使用独立句子作为网络模型的输入，没有结合特定方面的注意力机制，无法针对特定方面对模型进行训练和优化，也无法获取特定方面在句子内部和句子之间的依赖关系，是最基础的卷积神经网络模型。

4) LSTM：基于文献[22]提出的 LSTM 网络模型，以独立句子作为网络的输入，是最基础的 LSTM 网络模型。该模型可以保留句子中词语的时序关

系，获取词语间的依赖关系。但在训练过程中无法高度关注特定方面的情感特征，不能有效判别同一句子中不同方面的情感极性。

5) LSTM-R：基于文献[22]提出的 LSTM 网络模型，该模型以整个用户评论作为网络的输入，考虑了同一用户评论中不同句子之间的相互联系。但该模型没有针对特定方面加入注意力机制，网络在训练过程中不能针对特定方面进行学习和调参。

6) ATT-CNN：基于文献[8]提出的基于注意力机制的卷积神经网络模型，该模型在句子对建模任务中取得了比以往研究更好的效果。本文使用一种作用在卷积层的特定方面注意力机制，可以使模型在训练过程中高度关注特定方面的情感特征信息，从而有效区别同一个句子中不同方面的情感极性。但该模型以独立句子作为网络的输入，无法获取句子之间的相互联系。

7) ATT-LSTM：基于文献[14]提出的结合注意力机制的 LSTM 网络模型，该模型在以独立句子作为输入的基于方面情感分析任务中取得了比以往研究更好的效果。该模型可以在训练过程中高度关注特定的方面，有效识别不同方面的情感极性。但该模型以独立句子作为网络的输入，忽略了同一评论不同句子之间的相互联系。同时该模型的注意力机制需要复杂的矩阵运算，模型的训练时间较长。

8) RCNN-LSTM：基于文献[25]提出的区域卷积神经网络和 LSTM 网络结合来获取文本长距离依赖关系的网络模型，该模型以独立句子作为不同的独立区域，并把不同的独立区域作为卷积神经网络的输入，通过把卷积神经网络的输出按顺序输送到 LSTM 网络可以保留句子之间的时序关系，也可以获取整个评论的长距离依赖关系。但该模型没有针对特定方面加入注意力机制，网络模型在训练过程中不能针对特定方面进行学习和调参。

9) HP-LSTM：基于文献[16]提出的分层 LSTM 网络模型，该模型可以有效获取特定方面在句子内部的依赖关系，同时也可以获取同一评论中不同句子之间的相互联系，在长文本用户评论的基于方面情感分析任务中取得了突破性的效果。但该模型无法有效针对特定方面来获取整个评论中的情感特征信息。

### 4.4 实验结果与分析

本文使用 9 个网络模型在 phones、cameras 领

域的所有数据样本和 restaurants、laptops 领域中的积极和消极情感数据对比本文提出的 2 组模型和 7

组对比实验, 进行二分类对比实验, 验证本文提出方法的有效性。实验结果如表 3 所示。

表 3 不同模型在 4 个领域数据集上的二分类结果, 其中 P、R、F 分别代表精确率、查全率和 F1 值

模型	指标	数据集							
		EN-REST		EN-LAPT		CH-PHNE		CH-CAME	
		Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
CNN	P	92.32	60.45	85.00	66.00	76.87	62.10	88.37	56.67
	R	82.65	79.41	77.75	75.91	69.68	70.32	77.33	74.45
	F	87.20	68.64	81.21	70.61	73.10	65.95	82.48	64.35
LSTM	P	92.98	60.58	85.85	65.74	76.92	62.96	88.96	57.14
	R	82.32	81.37	76.92	77.74	70.97	69.86	77.33	75.91
	F	87.33	69.45	81.14	71.24	73.83	66.23	82.74	65.20
LSTM-R	P	93.92	62.73	86.11	66.25	77.82	65.25	89.97	58.79
	R	83.47	83.33	77.34	78.10	73.55	70.32	78.20	78.10
	F	88.39	71.58	81.49	71.69	75.62	67.69	83.67	67.08
ATT-CNN	P	94.28	63.37	86.90	67.81	77.89	63.93	89.67	58.56
	R	83.63	84.80	78.59	79.20	71.61	71.23	78.20	77.37
	F	88.64	72.54	82.54	73.06	74.62	67.38	83.54	66.66
ATT-LSTM	P	94.89	65.92	88.10	69.81	78.89	65.83	90.03	59.44
	R	85.11	86.27	80.04	81.02	73.55	72.15	78.78	78.10
	F	89.73	74.73	83.88	75.00	76.13	68.85	84.03	67.50
RCNN-LSTM	P	95.06	66.04	88.40	69.14	78.77	66.24	90.30	59.34
	R	85.11	87.25	79.21	81.75	74.19	71.69	78.49	78.83
	F	89.81	75.18	83.55	74.92	76.41	68.86	83.98	67.71
HP-LSTM	P	94.92	66.67	88.51	70.00	79.18	66.95	90.60	59.56
	R	85.60	86.27	80.04	81.75	74.84	72.15	78.49	79.56
	F	90.02	75.23	84.06	75.42	76.95	69.45	84.11	68.12
CNN-HLSTM	P	95.28	67.42	88.79	70.75	80.07	67.65	90.73	60.89
	R	85.92	87.25	80.67	82.12	75.16	73.52	79.65	79.56
	F	90.36	76.06	84.53	76.01	77.54	70.46	84.83	68.98
RCNN-HLSTM	P	95.84	69.08	88.94	72.12	80.68	68.20	91.09	61.80
	R	86.74	88.72	81.91	82.12	75.48	74.43	80.23	80.29
	F	<b>91.06</b>	<b>77.68</b>	<b>85.28</b>	<b>76.80</b>	<b>78.00</b>	<b>71.18</b>	<b>85.32</b>	<b>69.84</b>

从表 3 所示的实验结果可以看出, 本文提出的 CNN-HLSTM 和 RCNN-HLSTM 模型在 4 个领域的数据集上都取得了比其他网络模型更好的分类效果, 其中 RCNN-HLSTM 模型在 4 个领域的数据集上都取得了最好的分类效果, 在分类效果最好的 restaurants 领域数据集的 F1 值分别为 91.06% 和 77.68%, 比 HP-LSTM 模型提升了 1.04% 和 2.45%, 比 RCNN-LSTM 模型提升了 1.25% 和 2.50%, 验证

了本文提出方法的有效性。

没有加入特定方面注意力机制的基础模型 CNN 和 LSTM 的分类效果都不理想, 在分类效果最好的 restaurants 领域数据集上的积极和消极样本平均 F1 值也只有 77.92% 和 78.39%, 而加入注意力机制的 ATT-CNN 和 ATT-LSTM 模型在 restaurants 领域数据集上的平均 F1 值相比 CNN 和 LSTM 模型分别提升了 2.67% 和 3.84%。表 4 给出这两个基础

模型和结合注意力机制的网络模型在包含多个方面的典型句子分类结果。

从表 4 结果可以看出, 句子 1 中同一个句子的不同方面有相同的情感极性, 4 种模型都能正确识别不同目标的情感极性。但是基础的 CNN 和 LSTM 网络模型无法在训练过程中高度关注特定方面的特征信息, 所以在基于方面情感分析任务中将同一个句子中的不同方面都判定为相同的情感极性。而同一句子中针对不同的方面往往存在不同的情感极性, 如句子 2 和句子 3, 基础的 CNN 和 LSTM

网络模型把不同的方面都判定为相同的情感极性, 得到了错误的情感分类结果。结合注意力机制的 ATT-CNN 和 ATT-LSTM 模型可以在训练过程中通过特定方面的注意力机制高度关注特定方面的情感特征信息, 有效区分同一句子不同方面的情感极性。所以对于句子 2 和句子 3, ATT-CNN 和 ATT-LSTM 网络都可以通过特定方面的注意力机制来判别不同方面的情感极性, 得到正确的情感分类结果, 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。

表 4 包含多目标的典型句子举例

序号	句子	方面	情感	CNN	LSTM	ATT-CNN	ATT-LSTM
1	Service here was great, food was	SERVICE#GENERAL	Pos.	Pos.	Pos.	Pos.	Pos.
	fantastic	FOOD#QUALITY.	Pos.	Pos.	Pos.	Pos.	Pos.
2	Service was slow, but the people	SERVICE#GENERAL	Neg.	Pos.	Neg.	Neg.	Neg.
	were friendly	SERVICE#GENERAL	Pos.	Pos.	Neg.	Pos.	Pos.
3	Sometimes I get good food and	SERVICE#GENERAL	Neg.	Pos.	Pos.	Neg.	Neg.
	bad service	FOOD#QUALITY	Pos.	Pos.	Pos.	Pos.	Pos.

此外, 从表 3 结果也可以看出, 在 2 个基础的 LSTM 模型中, 以整个评论作为网络输入的 LSTM-R 模型比以独立句子作为网络输入的 LSTM 模型有更好的分类效果, 其中在效果最好的 restaurants 领域数据集上 LSTM-R 模型的 F1 值相比 LSTM 模型提升了 1.06% 和 2.13%, 说明考虑整个评论中不同句子的相互联系能有效提升模型的性能。在加入注意力机制的模型中, 相比以独立句子作为输入的 ATT-CNN 和 ATT-LSTM 网络模型, 考虑评论中句子之间相互关系的 RCNN-LSTM、HP-LSTM、CNN-HLSTM 和 RCNN-HLSTM 模型在 4 个领域数据集上都有更好的分类效果。其中, 在 phones 领域数据集上本文提出的 RCNN-HLSTM 相比 ATT-LSTM 模型的 F1 值提升最高, 为 1.87% 和 2.33%。

为了进一步分析考虑同一评论中不同句子之间相互联系在长文本的基于方面情感分析任务中比以独立句子作为输入有更好的分类效果, 本文加入 restaurants 和 laptops 领域数据集的中性情感样本在 restaurants 和 laptops 领域数据集上进行三分类对比实验, 对比实验结果如表 5 所示。

从表 5 实验结果可以看出, 在三分类的基于方面情感分析任务中, 考虑不同句子之间相互联系的 RCNN-LSTM、HP-LSTM、CNN-HLSTM 和 RCNN-HLSTM 模型在 restaurants 和 laptops 领域数据集上的分类效果都好于以独立句子作为输入的

ATT-CNN 和 ATT-LSTM 模型, 其中本文提出的 RCNN-HLSTM 模型在 restaurants 领域数据集上的分类效果最好, 在 3 种极性数据集中的平均 F1 值为 77.45%, 比 ATT-CNN 模型的 72.32% 提升了 5.13%, 比 ATT-LSTM 模型的 74.65% 提升了 2.8%。没有加入注意力机制的 RCNN-LSTM 模型在 restaurants 领域数据集上的平均 F1 值为 74.70%, 相比 ATT-CNN 模型提升了 2.38%, 相比 ATT-LSTM 模型提升了 0.05%。说明与以独立句子作为输入的网络模型相比, 考虑句子之间相互关系的模型可以在训练过程中挖掘整个评论中更深层的句子间依赖关系, 从而可以取得更好的情感分类效果, 即使在没加入注意力机制的模型中也有不错的分类效果, 进一步验证了考虑评论中句子间相互联系的有效性。

此外, 从表 5 的实验结果也可以看出, 模型对于数据集中简单句子都有很好的识别能力, 例如以往研究的 4 组模型在 restaurants 领域数据集消极样本中的查全率都在 80% 以上, 其中效果最差的 ATT-CNN 的查全率也有 80.39%。但是, 分析结果可以发现, 效果最好的 RCNN-LSTM 和 HP-LSTM 模型的查全率都为 83.82%, 分类效果无法取得进一步的提升, 说明数据集中存在一些以往研究提出的网络模型无法识别的样本。而对于本文提出的 CNN-HLSTM 和 RCNN-HLSTM 模型, 在 restaurants 领域中消极样本的分类查全率为 84.80% 和 85.29%,

查全率相比 RCNN-LSTM 和 HP-LSTM 模型提升了 0.98% 和 1.47%，说明本文提出的方法在基于方面情

感分析任务中能对其他模型难以判断的句子进行正确的极性判断，从而提升模型的分类效果。

表 5 不同模型在 restaurants 和 laptops 领域数据集上的三分类结果

模型	指标	数据集					
		EN-REST			EN-LAPT		
		Positive	Negative	Neutral	Positive	Negative	Neutral
ATT-CNN	P	93.85	68.62	42.17	84.94	70.24	36.78
	R	82.49	80.39	79.55	75.05	74.09	69.57
	F	87.80	74.04	55.13	79.69	72.11	48.12
ATT-LSTM	P	92.86	70.12	48.61	85.55	71.33	41.77
	R	82.98	82.84	79.55	76.30	76.28	71.74
	F	87.64	75.95	60.35	80.66	73.72	52.80
RCNN-LSTM	P	93.33	68.95	49.30	85.31	71.09	43.59
	R	82.49	83.82	79.55	76.09	76.28	73.91
	F	87.58	75.66	60.87	80.44	73.59	54.84
HP-LSTM	P	93.21	70.37	50.70	85.38	71.33	44.16
	R	83.14	83.82	81.82	76.51	76.28	73.91
	F	87.89	76.51	62.61	80.70	73.72	55.29
CNN-HLSTM	P	93.38	69.48	54.55	85.35	71.19	45.45
	R	83.14	84.80	81.82	76.30	76.64	73.91
	F	87.96	76.38	65.46	80.57	73.81	56.29
RCNN-HLSTM	P	93.42	69.88	57.14	85.68	71.53	49.32
	R	83.63	85.29	81.82	77.13	76.64	78.26
	F	<b>88.25</b>	<b>76.82</b>	<b>67.29</b>	<b>81.18</b>	<b>74.00</b>	<b>60.51</b>

本文在以往的研究基础上提出了一种结合区域 CNN 和分层 LSTM 网络的深度分层网络模型用于基于方面情感分析任务中。为了进一步验证该模型在基于方面情感分析任务中比以往研究有更好的情感分类效果，对比本文提出的 CNN-HLSTM、RCNN-HLSTM 模型和以前研究中取得最好效果的 HP-LSTM 模型在二分类和三分类的情感分类结果进行对比分析。对比结果如图 6 和图 7 所示。

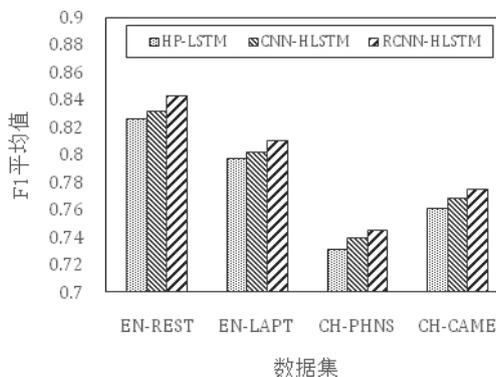


图 6 不同模型的二分类实验结果对比

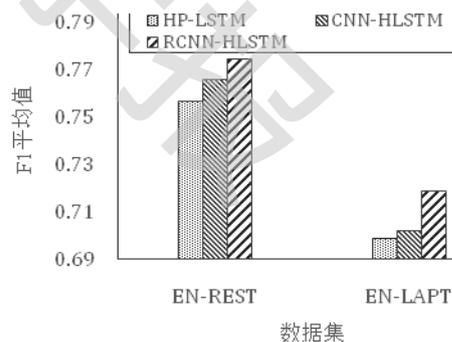


图 7 不同模型的三分类实验结果对比

综合图 6 和图 7 对比结果可以看出，本文提出的 CNN-HLSTM 和 RCNN-HLSTM 模型在所有的实验中都取得了比 HP-LSTM 模型更好的分类效果。其中，在二分类实验中提升最高的 restaurants 领域数据集上 RCNN-HLSTM 模型的 F1 平均值为 84.37%，相比 HP-LSTM 模型的 82.63% 提升了 1.74%。没有基于句子中不同方面的目标词来划分

独立区域的 CNN-HLSTM 模型的 F1 平均值为 83.21%，也比 HP-LSTM 模型提升了 0.58%。同时，RCNN-HLSTM 模型在三分类实验中的 cameras 领域数据集上的平均 F1 值为 71.90%，比 HP-LSTM 模型的 69.90% 提升了 2.0%。分析结果可知，本文提出的结合区域 CNN 和分层 LSTM 的深度分层网络模型不仅可以获取句子内部特定方面和词语之间的依赖关系，也可以获取同一评论中句子之间的相互关系。同时，结合区域 CNN 和分层 LSTM 的深度分层网络模型可以在训练过程中通过分层 LSTM 网络使模型从不同层面来关注特定方面的情感特征信息，有效挖掘特定方面的隐含情感特征，弥补了仅使用词语层面注意力机制的不足。同时，结合区域 CNN 和分层 LSTM 网络的模型可以通过区域 CNN 获取整个用户评论的特征信息，保留了评论中不同句子之间的时序关系，使网络模型在训练过程中可以挖掘更丰富的句子间的情感特征。所以本文提出的深度分层网络模型在所有领域的数据集上都取得了比 HP-LSTM 模型更好的分类效果，验证了本文提出的网络模型在基于方面情感分析任务中的有效性。

综合表 3、表 5 实验结果和图 6、图 7 的对比结果可以看出，本文提出的 CNN-HLSTM 和 RCNN-HLSTM 模型，在所有实验中都取得了比其他模型更好的分类效果。同时，基于句子中不同方面的目标词来划分独立区域的 RCNN-HLSTM 模型在所有的实验中都比以独立句子作为区域的 CNN-HLSTM 模型表现更好，说明本文提出的基于不同方面的目标词来划分不同区域的方法在基于方面情感分析任务中有更好的分类效果。为了进一步验证基于不同方面的目标词来划分区域方法的有效性，本文对句子进行不同长度的区域划分，在 laptops 和 restaurants 领域数据集上对 CNN-HLSTM 和 RCNN-HLSTM 模型进行二分类和三分类对比实验，如图 8-图 11 所示，其中正确率为 accuracy。

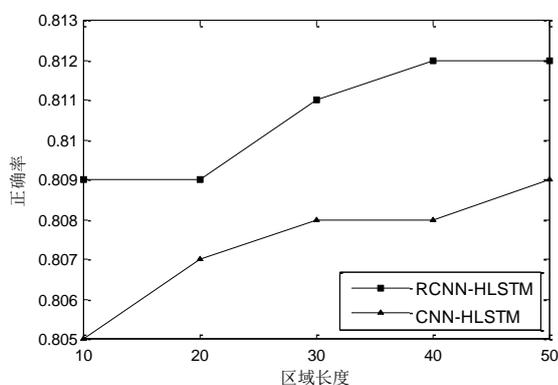


图 8 Laptops 领域数据集上不同区域长度的二分类结果

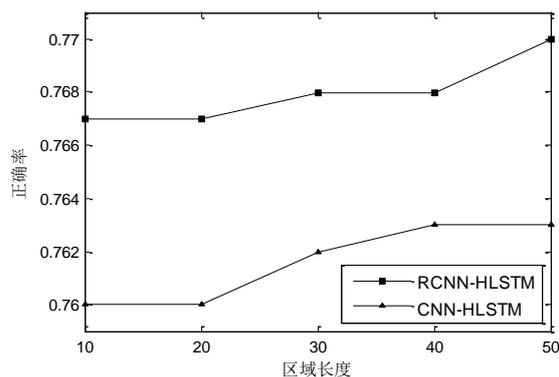


图 9 Laptops 领域数据集上不同区域长度的三分类结果

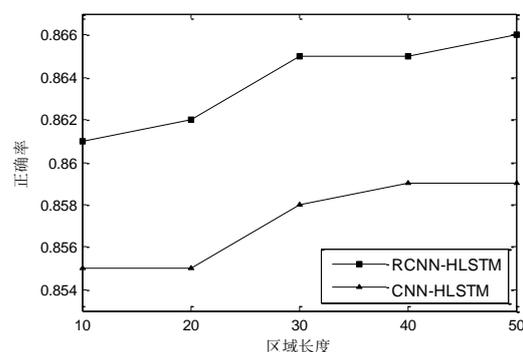


图 10 Restaurants 领域数据集上不同区域长度的二分类结果

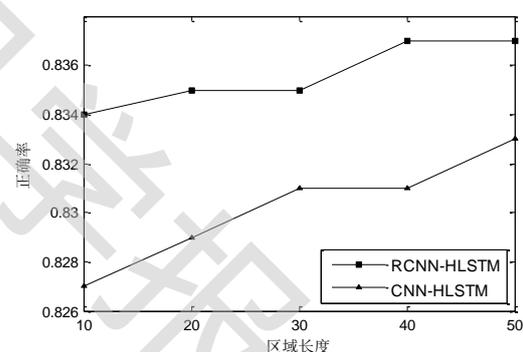


图 11 Restaurants 领域数据集上不同区域长度的三分类结果

综合图 8-图 11 的实验结果可以看出，本文提出的 RCNN-HLSTM 模型对比 CNN-HLSTM 模型在 restaurants 和 laptops 2 个数据集上的二分类和三分类实验中都有更好的分类效果。同时也可以看出，CNN-HLSTM 模型在随着区域长度的变化，分类正确率的波动比 RCNN-HLSTM 模型更大。说明以独立句子直接划分区域的方法在区域长度较小时在训练过程中会丢失对特定方面有重要影响的词语。而 RCNN-HLSTM 模型是基于特定方面的目标词来对句子进行区域划分，当区域长度减小时，也可以保留和特定方面有密切关系的词语，从而取得比直接划分区域的 CNN-HLSTM 模型更好的分类效果，

验证了基于不同目标词划分区域的有效性。

#### 4.5 典型难句分析

为了进一步分析本文提出的 RCNN-HLSTM 深度分层网络模型在解决情感极性表达不清晰、句子

长度较短等难句相比以往的研究有更好的效果，本文从数据集中抽取一些典型的难句进行分类结果对比分析，如表 6 所示。

表 6 数据集中的典型难句举例

Id	句子	关联句子	方面	情感	ATT-LSTM	HP-LSTM	RCNN-HLSTM
1	Can be used for gaming as well.	Awesome laptop for college!	LAPTOP#MISCELLANEOUS	Pos.	Neu.	Pos.	Pos.
2	Leave room for dessert	Great dining experience	FOOD#QUALITY	Pos.	Neu.	Neg.	Pos.
3	3.7 的屏	相当不错!!	DISPLAY#DESIGN_FEATURES	Pos.	Neg.	Pos.	Pos.
4	DONOT GO!	I have been to this place, folks and it is BAD.	RESTAURANT#GENERAL	Neg.	Neu.	Neu.	Neg.
5	We never had to wait more than 5 minutes.	This is a great place to get a delicious meal.	SERVICE#GENERAL	Pos.	Neg.	Pos.	Pos.
6	I waited for 10-15 minutes for service ordered a beer	Save yourself the waste of time	SERVICE#GENERAL	Neg.	Neg.	Neu.	Neg.
7	还有谁是弧形屏幕啊?	秒杀其他所有手机外形。	DISPLAY#DESIGN_FEATURES	Pos.	Neg.	Pos.	Pos.
8	The only positive thing about Mioposto is the nice location.	Poor customer service, poor pizza.	LOCATION#GENERAL RESTAURANT#GENERAL	Pos. Neg.	Pos. Pos.	Pos. Pos.	Pos. Neg.

从表 6 结果可以看出，句子 1 和句子 2 中没有明显的情感词，属于情感表达不清晰的句子。使用独立句子作为输入的 ATT-LSTM 模型无法通过句子中有限的情感特征来对特定方面进行正确的极性判断，对于这类无明显情感的句子都是偏向中性的极性判断，所以句 1 和句 2 都得到错误的结果。对于考虑了同一评论中不同句子之间相互联系的 HP-LSTM 模型，该模型虽然借助了不同句子之间的相互联系来帮助模型对这类句子进行情感分类，但是仅使用一个分层网络，难以挖掘充分的特征信息，所以该模型能正确识别句 1 的情感极性，但是对句 2 是错误的分类结果。对于本文提出的 RCNN-HLSTM 模型，该模型利用句子层 LSTM 来挖掘整个评论的情感特征信息。例如对于句 1，句子层 LSTM 网络可以通过挖掘评论中有相同方面的关联句子“Awesome laptop for college!”的特征信息和情感极性来对句 1 进行情感极性的判断，从而可以得到正确的分类结果。句 3 和句 4 属于句子长度较短的句子，这类句子由于长度的限制，所含的情感信息有限，仅靠从句子本身提取的特征信息，难以有效地对句子的

特定方面进行正确的情感极性判断。所以 ATT-LSTM 模型对句 3 和句 4 都得到错误的分类结果。对于 HP-LSTM 模型，虽然该模型通过获取评论中句子之间的相互联系，能有效利用其他句子的特征信息来协助对短句的情感极性判断，例如句 3 的分类结果是正确的。但是对于句 4，由于和句 4 有相同方面且极性相同的关联句子“I have been to this place, folks and it is BAD.”在评论中和句 4 的距离较远，HP-LSTM 模型在对句 4 进行分类的过程中难以利用到关联句子的特征信息，所以得到了错误的分类结果。对于本文提出的 RCNN-HLSTM 模型，该模型通过词语层 LSTM 和区域 CNN 的每一个输出进行组合操作，可以有效获取待分类句子和每个区域的相互联系，从而即使在关联句子距离句 4 较远的情况下，也能提取到句 4 和关联句子的长距离依赖关系，从而得到正确的情感分类结果。句 5 和句 6 是含有时间的句子，这类句子中最重要的信息是时间信息，但是对于不同方面或不同评论中，时间对应的情感极性可能是完全不一样的，例如句 5 是积极情感，而句 6 是消极情感。对于 ATT-LSTM 模型，

因为该模型无法获取句子之间的相互联系，模型通过对独立句子的特征信息挖掘，会将这类句子都判别为相同的情感极性，因此无法正确识别句 5 的情感极性。对于 HP-LSTM 模型，该模型因为考虑了评论中句子之间的相互联系，所以能正确识别句 5 的情感极性。但是因为句 6 本身含有很强烈的中性情感信息，而 HP-LSTM 模型没有针对待分类句子获取其和评论中每个句子的相互联系，因此难以正确识别句 6 的情感极性。而对于 RCNN-HLSTM 模型，该模型通过区域 CNN 可以有效提取和待分类句子有紧密关联的局部特征，通过分层 LSTM 网络，可以将局部特征和待分类句子的特征信息进行结合，从而得到正确的情感分类结果。句 7 属于带有疑问性质的句子，这类句子仅通过单一的句子信息，难以正确判断句子的情感极性，所以 ATT-LSTM 模型得到了错误的分类结果。而对于 HP-LSTM 和 RCNN-HLSTM 这 2 个模型，可以通过获取评论中的上下文信息来协助句子的情感极性判断，从而可以得到正确的分类结果。句 8 属于句子结构复杂，且表达情感的词语较多的句子，这类句子往往是对多个方面提出情感表达，同时情感极性的表达相对混乱。例如句 8 中有“positive”和“nice”这 2 个情感强烈的词语，所以 ATT-LSTM 和 HP-LSTM 将句子中不同的方面都判断为积极情感，所以对于方面“RESTAURANT#GENERAL”得到了错误的分类结果。对于本文提出的 RCNN-HLSTM 模型，模型的区域 CNN 可以针对不同方面提取句子中的情感特征信息，能有效获取某一方面在不同句子总的局部特征。同时，模型通过分层 LSTM 网络可以有效挖掘某一方面在整个评论中的情感特征信息。对于句 8 这类句子，RCNN-HLSTM 模型通过词语层 LSTM 网络来提取句子中不同方面的特征信息，然后将词语层 LSTM 的输出跟评论中不同句子的局部特征信息进行结合来学习某一方面在整个评论中的长距离依赖关系，判断其情感极性。例如方面“RESTAURANT#GENERAL”虽然在句子中有积极情感的表现，但是在整个评论中，该方面是消极的，所以模型通过整个评论的情感信息，能得到该方面的正确情感判断，最终得到正确的情感分类结果。

#### 4.6 网络训练时间分析

本文在相同的网络框架，相同的 CPU、GPU 下分析不同网络的训练时间，并对比分析不同网

络在 restaurants 领域数据集上完成一次迭代的训练时间，对比结果如表 7 所示。

表 7 不同模型完成一次迭代的训练时间

模型	时间 (s)
CNN	6
LSTM	82
LSTM-R	35
ATT-CNN	17
ATT-LSTM	128
RCNN-LSTM	42
HP-LSTM	73
CNN-HLSTM	60
RCNN-HLSTM	54

从表 7 结果可以看出，在相同的环境下 LSTM 网络的训练时间远远高于 CNN 的训练时间，在 restaurants 领域数据集上，LSTM 模型完成一次迭代的训练时间为 82 秒，几乎是 CNN 模型（一次迭代时间为 6 秒）的 14 倍。这主要是因为 LSTM 网络接收的是序列化输入，所以训练时间高于接收平行化输入的 CNN 模型。相比以独立句子作为输入的 LSTM 模型，LSTM-R 模型是以整个评论作为网络的输入，网络模型能同时处理整个评论中的句子，所以训练时间远低于以独立句子作为输入的 LSTM 模型。此外，从表中结果也可以看出，使用独立句子作为网络输入并且加入注意力机制的 ATT-LSTM 模型完成一次迭代所用的时间最多，为 128 秒，而 ATT-CNN 模型完成一次迭代的时间仅为 17 秒，再一次证明 CNN 网络的训练时间要远低于 LSTM 网络。本文提出的 CNN-HLSTM 和 RCNN-HLSTM 模型完成一次迭代所需时间为 60 秒和 54 秒，对比以往研究中取得最好效果的 HP-LSTM 模型的 73 秒分别降低了 13 秒和 19 秒，说明结合 CNN 的网络模型能有效降低模型的训练时间。而 ATT-LSTM 模型完成一次迭代所需时间是本文提出的 RCNN-HLSTM 模型的 2 倍多，从而验证了本文提出的方法有更优的训练性能。

## 5 总结

本文提出了一种结合区域卷积神经网络和分层长短记忆网络的 RCNN-HLSTM 模型用在基于方面情感分析任务中。通过和传统的深度神经网络以及目前在文本分类任务中取得突破性成果的

网络模型进行对比实验,验证了本文提出方法的有效性。在基于方面情感分类任务中,用户评论往往存在一些情感极性表达不清晰、句子长度较短、句子情感表达混乱等句子。目前的研究方法大多是以独立句子作为网络模型的输入,忽略了句子之间的相互联系,所以难以准确判断这类句子的情感极性。本文提出的 RCNN-HLSTM 模型通过区域 CNN 能有效提取整个评论中的重要情感特征信息,这些特征信息可以用于整个评论中所有句子的情感极性判断,使得模型可以有效判断情感极性表达不清晰等句子的情感极性。此外,该模型通过一个分层 LSTM 网络来接收词语层和句子层的注意力机制,可以高度关注特定方面在待分类句子内部的依赖关系,以及待分类句子在整个评论中的依赖关系,取得比对比模型更好的分类效果。实验结果表明,本文提出的 RCNN-HLSTM 模型在所有领域的数据集上都取得了最佳的分类效果,在分类效果最好的 restaurants 领域数据集上的二分类实验的平均 F1 值为 84.37%,相比以往研究中表现最好的 HP-LSTM 模型提升了 1.74%,验证了本文提出的 RCNN-HLSTM 模型在基于方面情感分析任务中的有效性。此外,本文提出的方法结合了可以接收平行化输入的 CNN,降低了网络模型的训练时间,对比以往提出的模型有更优的时间性能。

此外,从实验结果也可以看出,本文提出的 RCNN-HLSTM 模型对区域的划分有一定的依赖性,在区域长度减小时分类正确率有一定的下降。同时,本文的区域划分方法仅基于特定方面的目标词,对于没有目标词的句子仍然使用直接划分区域的方法,这对分类效果有一定的影响。所以本文接下来的工作将针对这个问题对网络模型进行改进,以达到更好的情感分类效果。

### 参考文献

- [1] Pang B, Lee L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2008, 2(1-2): 1-135
- [2] Pontiki M, Galanis D, Papageorgiou H, et al. Orphé De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeny Kotelnikov, Nuria Bel, Salud Maria Jiménez-Zafra, and Gülsen Eryigit. 2016. SemEval-2016 task 5: Aspect based sentiment analysis//Proceedings of the 10th International Workshop on Semantic Evaluation. San Diego, USA, 2016: 19-30
- [3] Wang Zhong-Yuan, Cheng Jian-Peng, Wang Hai-Xun, Wen Ji-Rong. Short text understanding: A survey. *Journal of Computer Research and Development*, 2016, 53(2): 262-269 (in Chinese)  
(王仲远,程健鹏,王海勋,文继荣.短文本理解研究.计算机研究与发展,2016,53(2):262-269)
- [4] Pontiki M, Galanis D, Pavlopoulos J, et al. Semeval-2014 task 4: Aspect based sentiment analysis//Proceedings of the 8th International Workshop on Semantic Evaluation. Dublin, Ireland, 2014: 27-35
- [5] Boiy E, Moens M F. A machine learning approach to sentiment analysis in multilingual Web texts. *Information Retrieval*, 2009, 12(5): 526-558
- [6] Yu Kai, Jia Lei, Chen Yu-Qiang, Xu Wei. Deep learning: yesterday, today, and tomorrow. *Journal of Computer Research and Development*, 2013, 50(9): 1799-1804 (in Chinese)  
(余凯,贾磊,陈雨强,徐伟.深度学习的昨天、今天和明天.计算机研究与发展,2013,50(9):1799-1804)
- [7] Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016: 207-212
- [8] Yin W, Schütze H, Xiang B, et al. Abcnn: Attention-based convolutional neural network for modeling sentence pairs. *Transactions of the Association for Computational Linguistics*, 2016, 4(11): 259-272
- [9] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014
- [10] Kiritchenko S, Zhu X, Cherry C, et al. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews//Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland, 2014: 437-442
- [11] Zhang Z, Lan M. Ecnu: Extracting effective features from multiple sequential sentences for target-dependent sentiment analysis in reviews//Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). Denver, USA, 2015: 736-741
- [12] Dong L, Wei F, Tan C, et al. Adaptive recursive neural network for target-dependent Twitter sentiment classification//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Baltimore, USA, 2014: 49-54
- [13] Nguyen T H, Shirai K. PhraseRNN: Phrase recursive neural

- network for aspect-based sentiment analysis//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal, 2015: 2509-2514
- [14] Wang Y, Huang M, Zhao L, et al. Attention-based LSTM for aspect-level sentiment classification//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, USA, 2016: 606-615
- [15] Tang D, Qin B, Liu T. Aspect level sentiment classification with deep memory network//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, USA, 2016: 214-224
- [16] Ruder S, Ghaffari P, Breslin J G. A hierarchical model of reviews for aspect-based sentiment analysis//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, USA, 2016: 999-1005
- [17] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324
- [18] Kim Y. Convolutional neural networks for sentence classification//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar, 2014: 1746-1751
- [19] He Yan-Xiang, Sun Song-Tao, Niu Fei-Fei, Li Fei. A deep learning model enhanced with emotion semantics for microblog sentiment analysis. Chinese Journal of Computers, 2016, 40(4): 773-790 (in Chinese)  
(何炎祥, 孙松涛, 牛菲菲, 李飞. 用于微博情感分析的一种情感语义增强的深度学习模型. 计算机学报, 2016, 40(4): 773-790)
- [20] Zhou X, Wan X, Xiao J. Attention-based LSTM network for cross-lingual sentiment classification//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, USA, 2016: 247-256
- [21] Liu P, Qiu X, Chen J, et al. Deep fusion LSTMs for text semantic matching//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016: 1034-1043
- [22] Wang X, Liu Y, Sun C, et al. Predicting polarities of Tweets by composing word embeddings with long short-term memory//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Beijing, China, 2015: 1343-1353
- [23] Mnih V, Heess N, Graves A. Recurrent models of visual attention//Proceedings of the Advances in Neural Information Processing Systems 27 (NIPS 2014). Montreal, Canada, 2014:2204-2212
- [24] Wang Linlin, Cao Zhu, de Melo G, et al. Relation classification via multi-level attention cnns//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016: 1298-1307
- [25] Wang J, Yu L C, Lai K R, et al. Dimensional sentiment analysis using a regional cnn-lstm model//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016: 225-230
- [26] Pennington J, Socher R, Manning C D. Glove: Global Vectors for Word Representation//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar, 2014, 14: 1532-1543

**LIU Quan**, born in 1969, Ph.D. , professor, Ph.D. supervisor. His main research interests include intelligence information processing, automated reasoning and machine learning.



**LIANG Bin**, born in 1993. Master candidate. His main research interests

include sentiment analysis, natural language processing, and deep 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.

## Background

Sentiment analysis from customer reviews has become flourishing in these years. And deep learning has achieved much attention in the field of NLP and the approaches based on

neural networks have become pervasive in sentiment analysis recently. For the task of aspect-based sentiment classification, attention mechanism can be well combined with neural

network to obtain better results. However, such approaches are only in a position to receive an independent sentence and focus on the aspect itself in the process of training, which ignoring the impact of the words around the target on sentence, and cannot extract the important information of the aspect across the whole review. This paper proposes a regional CNN-hierarchical LSTM, which is able to consider both word-level and sentence-level attention of the aspect in the process of training. Meanwhile, this approach can leverage the relations both of intra-sentence and inter-sentence. The sole regional CNN aims to capture inter-sentence relations by means of receiving several regions of a review and extract the long-distance dependency of aspect across the whole review. The hierarchical LSTM architecture consists of word-level and sentence-level LSTMs. Word-level LSTM receive a word-level

sequential input and an aspect attention that aims to concentrate on different aspects in the sentence we are considering. The sentence-level LSTM is able to capture sentence-level attention by receiving a sequential input of the review from regional CNN, which can capture the important relations of the sentence on the whole review explicitly.

This paper is partially supported by National Natural Science Foundation of China (61272005, 61303108, 61373094, 61472262, 61502323, 61502329, 61772355), Natural Science Foundation of Jiangsu (BK2012616), High School Natural Foundation of Jiangsu (13KJB520020, 16KJB520041), Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University (93K172014K04), Suzhou Industrial application of basic research program part (SYG201422, SYG201308).