



文章编号:1001-9081(2019)08-2186-06

DOI:10.11772/j.issn.1001-9081.2019010096

# 基于注意力机制的卷积-双向长短期记忆模型跨领域情感分类方法

龚琴\*, 雷曼, 王纪超, 王保群

(重庆邮电大学 通信与信息工程学院, 重庆 400065)

(\*通信作者电子邮箱 1772362617@qq.com)

**摘要:**针对现有跨领域情感分类方法中文本表示特征忽略了重要单词的情感信息,且在迁移过程中存在负迁移的问题,提出一种基于注意力机制的卷积-双向长短期记忆(AC-BiLSTM)模型的知识迁移方法。首先,利用低维稠密的词向量对文本进行向量表示;其次,采用卷积操作获取局部上下文特征之后,通过双向长短期记忆(BiLSTM)网络充分考虑特征之间的长期依赖关系;然后,通过引入注意力机制考虑不同词汇对文本的贡献程度,同时为了避免迁移过程中出现负迁移现象,在目标函数中引入正则项约束;最后,将在源领域产品评论训练得到的模型参数迁移到目标领域产品评论中,并在少量目标领域有标注数据上进行微调。实验结果表明,与 AE-SCL-SR 方法和对抗记忆网络(AMN)方法相比,AC-BiLSTM 方法的平均准确率分别提高了 6.5% 和 2.2%,AC-BiLSTM 方法可以有效地提高跨领域情感分类性能。

**关键词:**情感分类;跨领域;迁移学习;注意力机制;长短期记忆网络

**中图分类号:** TP181; TP389.1   **文献标志码:**A

## Cross-domain sentiment classification method of convolution-bi-directional long short-term memory based on attention mechanism

GONG Qin\*, LEI Man, WANG Jichao, WANG Baoqun

(School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

**Abstract:** Concerning the problems that the text representation features in the existing cross-domain sentiment classification method ignore the sentiment information of important words and there is negative transfer during transfer process, a Convolution-Bi-directional Long Short-Term Memory based on Attention mechanism (AC-BiLSTM) model was proposed to realize knowledge transfer. Firstly, the vector representation of text was obtained by low-dimensional dense word vectors. Secondly, after local context features being obtained by convolution operation, the long dependence relationship between the features was fully considered by Bi-directional Long Short-Term Memory (BiLSTM) network. Then, the contribution degrees of different words to the text were considered by introducing attention mechanism, and a regular term constraint was introduced into the objective function in order to avoid the negative transfer phenomenon in transfer process. Finally, the model parameters trained on source domain product reviews were transferred to target domain product reviews, and the labeled data in a small number of target domains were fine-tuned. Experimental results show that compared with AE-SCL-SR (AutoEncoder Structural Correspondence Learning with Similarity Regularization) method and Adversarial Memory Network (AMN) method, AC-BiLSTM method has average accuracy increased by 6.5% and 2.2% respectively, which demonstrates that AC-BiLSTM method can effectively improve cross-domain sentiment classification performance.

**Key words:** sentiment classification; cross-domain; transfer learning; attention mechanism; Long Short-Term Memory (LSTM) network

## 0 引言

随着互联网的蓬勃发展,互联网信息呈现爆炸式增长,京东、天猫、淘宝、亚马逊等大型电子商务网站迅速发展,网上购物也因此受到了越来越多人的青睐。为了提高用户满意度,网络商家通常允许用户对其所购买的商品进行评价,这导致商品评价数量与日俱增。挖掘并分析评论文本中的情感倾向,不仅可以优化用户的购买决策,而且能够帮助生产商和销售商通过反馈信息来改进产品,从而提高市场竞争力,因此,

情感分析(Sentiment Analysis, SA)<sup>[1]</sup>应运而生。

情感分析,又称为意见挖掘,它是利用自然语言处理(Natural Language Processing, NLP)、文本分析、机器学习等技术自动分析商品评论的文本内容,挖掘出消费者对该商品的褒贬态度和意见<sup>[2]</sup>。然而,当训练集和测试集不属于同一领域时,典型的有监督情感分类方法的效果就变得很差。这是由于训练集里有强烈情感倾向性的词语在测试集里呈现微弱的情感倾向性,由此产生了跨领域情感分类(Cross-Domain Sentiment Classification, CDSC)问题。随着评论文本的不断增

收稿日期:2019-01-14;修回日期:2019-03-19;录用日期:2019-03-19。   基金项目:长江学者和创新团队发展计划项目(IRT\_16R72)。

**作者简介:**龚琴(1993—),女,四川成都人,硕士研究生,主要研究方向:自然语言处理、深度学习;雷曼(1992—),女,重庆人,硕士研究生,主要研究方向:推荐系统、社交网络;王纪超(1993—),男,河南郑州人,硕士研究生,主要研究方向:机器学习、数据挖掘;王保群(1993—),男,山东聊城人,硕士研究生,主要研究方向:机器学习、数据挖掘。



加、新兴领域的不断涌现,需要大量的人力来标注训练集,这是费时费力的,因此研究跨领域的情感分类具有十分重要的研究价值。

跨领域情感分类的研究主要集中于从一个或多个源领域训练得到分类器并且应用在不同的目标领域进行训练。当前存在的技术主要包含两类:1)通过目标领域中少量标注数据来辅助训练;2)目标领域均是未标注数据。本文主要针对第一类进行研究。

近年来,迁移学习已经在学术界和工业界引起了广泛的关注和研究。其中,领域自适应(Domain Adaptation, DA)是迁移学习中的一种代表性方法,通过利用信息丰富的源领域样本来提升目标领域模型的性能。然而在领域自适应实际应用中常常存在概念漂移<sup>[3]</sup>和特征漂移<sup>[4]</sup>的问题。比如,不同领域特征分布差异太大将会导致情感分类模型在跨领域应用中分类效果下降,同时在不同领域下相同的特征可能代表相反的情感。

为了解决跨领域所存在的概念漂移(特征漂移)问题,本文选择亚马逊网站中4个领域产品评论作为实验数据,以情感分析作为研究任务,提出一种基于注意力机制的卷积-双向长短期记忆模型(Convolution-Bi-directional Long Short-Term Memory based on Attention mechanism, AC-BiLSTM)以实现不同领域环境下的知识迁移。

## 1 相关工作

情感分类是自然语言处理领域的一个研究热点,受到了学术界和工业界的高度重视。文献[5-6]总结了以往的情感分类研究技术及其所使用的数据集。然而这些研究主要集中在单一领域,随着数据量的快速增长,要想在跨领域环境下构建一个具有鲁棒性和普遍性的情感分类器变得十分困难,这激励研究者去分析跨领域数据的情感极性,即跨领域情感分类。

以往展开的大多数研究都是采用特征迁移<sup>[7]</sup>来实现,它致力于解决源领域与目标领域特征空间存在差异的问题,主要通过把各个领域的数据映射到同一空间下,使源领域与目标领域具有相同的分布,并利用源领域中的训练数据来解决目标领域的学习问题<sup>[8]</sup>。其中,Blitzer等<sup>[9]</sup>提出的结构对应学习(Structural Correspondence Learning, SCL)模型是经典的方法,该模型利用源领域和目标领域部分枢纽特征,构造枢纽特征与非枢纽特征的关联模型,在此基础上,构造基于枢纽特征和非枢纽特征的特征空间,用于情感分类任务;Pan等<sup>[10]</sup>提出谱特征对齐(Spectral Feature Alignment, SFA),首先构造领域枢纽特征与非枢纽特征的二分图,然后在二分图上利用谱聚类获取新的特征表达,最后在新的特征上构造分类器;文献[11]使用主成分分析法(Principal Component Analysis, PCA)进行特征选择,然后用朴素贝叶斯(Naive Bayes, NB)进行情感分类;文献[12]采用半监督方式,结合改进的最大熵(Maximum Entropy, ME)模型与二分图聚类模型,针对情感词分类取得了相对较好的准确率;文献[13]中提出一种基于短文本特征扩展的迁移学习模型——CATL-PCO(Correlation Analysis Transfer Learning-Probability Co-Occurrence)以解决古代诗歌这类短文本的情感分析问题。然而这些方法使用专家

设计的规则或者n-gram进行句子的特征提取,通常未考虑到上下文之间的关系和重要单词的情感信息,因此不能较好地表示跨领域大规模数据。

随着深度学习研究工作的开展,深度学习被广泛应用于图像分类、语音识别和自然语言处理等领域<sup>[14-15]</sup>。近年来,有研究者提出将深度学习用来学习跨领域场景下情感分类的公有特征和共享参数的方法,这些方法取得了显著的成就。Glorot等<sup>[16]</sup>利用堆叠降噪自编码器(Stacked Denoising Autoencoder, SDA)对多个领域的未标注数据进行预训练,结合源领域的标注数据和预训练模型来训练情感分类模型;Chen等<sup>[17]</sup>提出边缘降噪自动编码器(marginalized SDA, mSDA),它保留了强大的特征学习能力并且解决了高昂的计算成本和SDA的可扩展性问题;随后,许多SDA的扩展方法被提出以提高跨领域情感分类的性能<sup>[18-19]</sup>。然而这些方法缺乏可解释性,即无法证明网络是否充分学习到了枢纽特征。

为了提高深度学习的可解释能力,Ganin等<sup>[20]</sup>提出域对抗神经网络(Domain Adversarial Neural Network, DANN)来解决跨领域情感分类问题;文献[21]中提出AE-SCL-SR(AutoEncoder Structural Correspondence Learning with Similarity Regularization)算法,结合自编码和结构对应学习的优势来提高跨领域情感分类的性能;Li等<sup>[22]</sup>提出对抗记忆网络(Adversarial Memory Network, AMN)模型,利用注意力机制来自动捕获枢纽特征从而实现跨领域情感分类。

## 2 AC-BiLSTM模型跨领域情感分类方法

### 2.1 问题定义

为了解决跨领域情感分类问题,本文给出问题定义:给定已标记情感的源领域 $D_S = \{(X_{S_i}, \gamma_{S_i})\}_{i=1}^{n_S}$ 和无标记情感的目标领域 $D_T = \{(X_{T_j})\}_{j=1}^{n_T}$ ,其中, $n_S$ 和 $n_T$ 分别表示源领域 $D_S$ 和目标领域 $D_T$ 的评论数量, $X_{S_i}$ 和 $X_{T_j}$ 分别表示源领域 $D_S$ 和目标领域 $D_T$ 的第*i*和*j*条评论, $\gamma_{S_i}$ 为 $X_{S_i}$ 对应的情感标签, $\gamma_{S_i} \in \{+1, -1\}$ ,+1和-1分别为正向情感标签和负向情感标签。本文研究的目标是利用源领域 $D_S$ 的有标签评论 $X_{S_i}$ 训练一个分类器来预测目标领域 $D_T$ 的无标签评论 $X_{T_j}$ 的情感标签 $\gamma_{T_j}$ 。

### 2.2 相关概念

互联网上商品评论文本中的特征可分为两类:1)枢纽特征。源领域和目标领域中的公共特征,表示两个领域的一些公共知识。2)非枢纽特征。在某一领域多次出现而在另一个领域很少出现或不出现的特征。例如,“great”或“terrible”等特征在不同领域中表达的情感相似,在亚马逊网站的商品评论中,在源领域和目标领域都高频出现,这些特征被称为枢纽特征。表1给出了亚马逊网站商品评论中,Books和Kitchen两个领域的评论。

### 2.3 AC-BiLSTM模型

AC-BiLSTM模型如图1所示。该模型主要包含5个步骤:

- 步骤1 采用词向量进行文本表示;
- 步骤2 利用卷积操作获取局部特征;
- 步骤3 利用BiLSTM模型充分考虑到特征之间的长期



依赖关系；

步骤4 引入注意力机制表示不同特征的重要性；

步骤5 利用分类器进行情感分类。

表1 书籍和厨房用品两个领域的评论

Tab. 1 Reviews of books and kitchen domains

情感极性	书籍	厨房用品
+	This is <u>excellent</u> survey on deep learning methods.	10 dollars for a <b>sharp</b> knife like this, could not ask for a better <b>bargain</b> .
+	The story is <b>interesting</b> and <b>thrilling</b> . Could not put the book down.	The <u>excellent</u> quality of these <b>sharp</b> knives are well <b>worth</b> their price.
-	A <b>disappointing</b> end to a <b>boring</b> story. Utter <b>waste</b> of time.	Knives got <b>rusty</b> and <b>blunt</b> after normal usage for two weeks.

注:在这两个评论中,“+”表示评论呈正极性,“-”表示评论呈负极性;加粗字体的词汇记为非枢纽特征,带下划线的词汇记为枢纽特征。

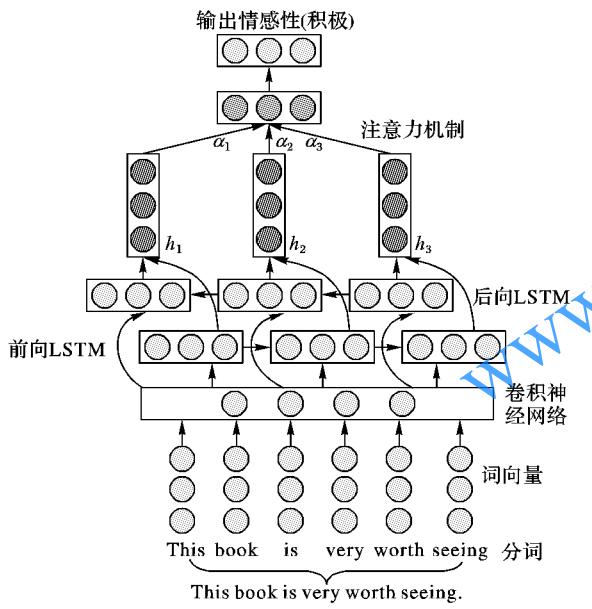


图1 AC-BiLSTM模型

Fig. 1 AC-BiLSTM model

输入层 作为词向量嵌入层,通过Word2Vec<sup>[23]</sup>预训练的低维稠密的词向量对词向量嵌入层的权重进行初始化,并采用均匀分布  $U(-0.01, 0.01)$  对未登录词进行向量初始化。嵌入层矩阵形式为  $\mathbf{x} \in \mathbb{R}^{n \times k}$ ,一条评论可以向量化为:

$$\mathbf{x}_{1:n} = \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \cdots \oplus \mathbf{x}_n \quad (1)$$

其中: $n$  为一条评论的长度, $k$  为词向量的维度, $\mathbf{x}_i$  为句子中第  $i$  个词的向量表示, $\oplus$  为连接运算符。

卷积层 卷积核的作用在于通过窗口滑动得到输入数据的局部特征,卷积滤波器  $\mathbf{m} \in \mathbb{R}^{h \times k}$  是指对窗口为  $h$  的  $k$  维词向量进行卷积操作。一条评论中第  $i$  个词的新的特征表示如下:

$$s_i = f(\mathbf{m}^T \cdot \mathbf{x}_{i:i+h-1} + \mathbf{b}) \quad (2)$$

其中: $\mathbf{m}^T$  和  $\mathbf{b}$  分别为权值和阈值,  $f$  为非线性激活函数ReLU(Rectified Linear Unit)。评论的特征表达为:

$$\mathbf{S} = [s_1, s_2, \dots, s_{n-h+1}]; \mathbf{S} \in \mathbb{R}^{n-h+1} \quad (3)$$

该层的输出  $\mathbf{y}_i \in \mathbb{R}^p$  由每个特征映射的第  $i$  维连接而成,如下所示:

$$\mathbf{y}_i = \mathbf{S}_1^i \oplus \mathbf{S}_2^i \oplus \cdots \oplus \mathbf{S}_p^i \quad (4)$$

LSTM 模型由三个门(输入门  $i_t$ 、遗忘门  $f_t$ 、输出门  $o_t$ )和一个记忆单元( $c_t$ )构成,通过这三个门对内部记忆进行选择性的输入、输出和遗忘操作,能够有效地克服梯度爆炸或梯度消失缺陷。LSTM 可表示为:

$$i_t = \sigma(\mathbf{W}_{iy} \mathbf{y}_t + \mathbf{W}_{ih} \mathbf{h}_{t-1} + \mathbf{b}_i) \quad (5)$$

$$f_t = \sigma(\mathbf{W}_{fy} \mathbf{y}_t + \mathbf{W}_{fh} \mathbf{h}_{t-1} + \mathbf{b}_f) \quad (6)$$

$$o_t = \sigma(\mathbf{W}_{oy} \mathbf{y}_t + \mathbf{W}_{oh} \mathbf{h}_{t-1} + \mathbf{b}_o) \quad (7)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(\mathbf{W}_{ch} \mathbf{y}_t + \mathbf{W}_{ch} \mathbf{h}_{t-1} + \mathbf{b}_c) \quad (8)$$

$$\mathbf{h}_t = o_t \odot \tanh(c_t) \quad (9)$$

其中: $\mathbf{h}_t$  为  $t$  时刻的隐藏状态; $\sigma$  为激活函数 Sigmoid; $\mathbf{y}_t$  为  $t$  时刻的输入词向量; $\mathbf{W}$  和  $\mathbf{b}$  分别为权值和阈值,均为模型训练参数; $\odot$  为逐点乘积。

虽然 LSTM 能够解决长期依赖的问题,但是它并没有利用文本的下文信息,因此本文采用 BiLSTM 模型同时考虑文本的上下文信息,其工作原理为:通过两个 LSTM 来得到时序相反的隐藏层状态并将其通过连接方式得到同一个输出,前向 LSTM 和后向 LSTM 分别获取输入序列的上文信息和下文信息,BiLSTM 模型能够有效地提高准确率。BiLSTM 在  $t$  时刻的隐藏状态  $\mathbf{H}_t$  包含前向的  $\mathbf{h}_t^{\text{forward}}$  和后向的  $\mathbf{h}_t^{\text{backward}}$ ,如下所示:

$$\mathbf{h}_t^{\text{forward}} = \text{LSTM}^{\text{forward}}(\mathbf{h}_{t-1}, \mathbf{x}_t, \mathbf{c}_{t-1}) \quad (10)$$

$$\mathbf{h}_t^{\text{backward}} = \text{LSTM}^{\text{backward}}(\mathbf{h}_{t-1}, \mathbf{x}_t, \mathbf{c}_{t-1}) \quad (11)$$

$$\mathbf{H}_t = [\mathbf{h}_t^{\text{forward}}, \mathbf{h}_t^{\text{backward}}] \quad (12)$$

将 BiLSTM 的输出  $\mathbf{H}_t$  当作文本的特征向量。

对于情感分类任务而言,句子中的情感词在整个句子的情感倾向性判别上具有十分关键的作用,因此,提出 AC-BiLSTM 模型,通过引入注意力机制<sup>[24]</sup>计算出文本中每个单词的注意力权重,使得情感词所在时刻的隐藏状态对情感分类有更大的贡献。

在 AC-BiLSTM 模型中,首先在每个时刻都将得到一个表示该时刻隐藏状态  $\mathbf{H}_t$  所占的注意力权重  $\alpha_t$ ,最后通过加权累加得到用于情感分类的隐藏状态  $\mathbf{v}$ ,可表示为:

$$\mathbf{u}_t = \tanh(\mathbf{W}^{\text{att}} \mathbf{H}_t + \mathbf{b}^{\text{att}}) \quad (13)$$

$$\alpha_t = \frac{\exp(\mathbf{u}_t^T \mathbf{u}_w)}{\sum_t \exp(\mathbf{u}_t^T \mathbf{u}_w)} \quad (14)$$

$$\mathbf{v} = \sum_t \alpha_t \hat{\mathbf{h}}_t \quad (15)$$

其中: $\mathbf{u}_t$  为  $\mathbf{H}_t$  的隐藏单元; $\mathbf{u}_w$  为上下文向量,被随机初始化并在训练中不断学习更新; $\mathbf{W}^{\text{att}}$  和  $\mathbf{b}^{\text{att}}$  均为注意机制参数; $\alpha_t \in [0, 1]$  为  $\mathbf{H}_t$  的权重且  $\sum_t \alpha_t = 1$ 。

最后,将注意力机制输出  $\mathbf{v}$  输入到 softmax 函数进行情感分类,分类结果为:

$$\hat{y} = \text{softmax}(\mathbf{w}_s \mathbf{v} + \mathbf{b}_s) \quad (16)$$

本文采用交叉熵作为损失函数。交叉熵损失函数公式如下:

$$\text{Loss} = - \sum_i y_i \log \hat{y}_i + \lambda \|\theta\|^2 \quad (17)$$

其中: $y_i$  为实际类别; $\hat{y}_i$  为预测类别; $\lambda \|\theta\|^2$  为正则项,防止



发生过拟合现象同时避免负迁移。

AC-BiLSTM 模型在源领域数据进行训练,训练好的模型表示为  $N_s$ 。为了实现迁移,本文还构造了另一个模型  $N_t$ ,这两个模型神经网络结构一致。首先将在  $N_s$  训练好的参数对  $N_t$  对应层进行初始化;然后冻结  $N_t$  的词向量层,  $N_t$  其余层的参数在目标领域含有少量标注的数据上进行微调;最后将训练好的  $N_t$  模型用来对目标领域进行预测。跨领域情感分类框架如图 2 所示。

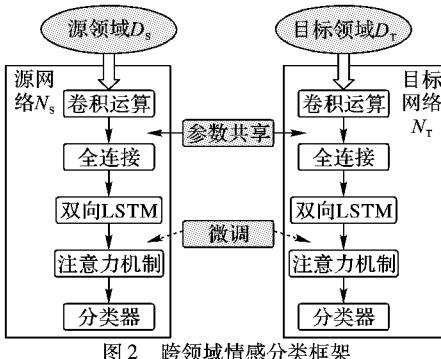


Fig. 2 Architecture of cross-domain sentiment classification

### 3 实验结果及分析

#### 3.1 数据集及参数设置

本文进行仿真实验所采用的数据集是宾夕法尼亚大学 Blitzer 等<sup>[9]</sup>收集的多领域情感数据集,具体包括:Books(B)、DVD(D)、Electronics(E)和Kitchen(K)4个领域。其中,每个领域包含正、负评论各 1 000 条,另外还包含大量的未标注数据。分别将各领域正负评论通过五折交叉验证划分为训练集和验证集。多领域情感数据集的统计信息如表 2 所示。

表 2 多域情感数据集统计信息

Tab. 2 Statistics of multi-domain sentiment datasets

领域	训练集	测试集	未标注数据集	消极评论占比/%
Books	1 600	400	6 000	50
DVD	1 600	400	34 741	50
Electronics	1 600	400	13 153	50
Kitchen	1 600	400	16 785	50

为了训练一个较好的模型,模型参数的设置十分重要,表 3 是本文方法的参数设置。

表 3 实验参数设置

Tab. 3 Experiment parameters setting

参数	值	参数	值
词向量维度	128	LSTM 单元状态维度	128
词向量训练模型	Skip-Gram	LSTM 隐藏状态维度	128
迭代次数	10	注意力权重维度	128
Batch Size	64	Dropout	0.5
滤波器数量	100	优化函数	Adam
卷积核大小	3	学习率	0.001

#### 3.2 评价指标

本文主要通过准确率(Accuracy)来评估情感分类的性能。计算公式如下所示:

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (18)$$

其中: $\hat{y}_i$  表示  $x_i$  的预测标签, $y_i$  则表示  $x_i$  的实际标签, $N$  表示测试集的大小。

#### 3.3 实验分析

将本文提出的 AC-BiLSTM 模型的跨领域情感分类方法与以下 8 种方法进行比较来验证所提方法的有效性。

1) NoTransf(No Transfer): 源领域训练的分类器直接用于目标领域。

2) SCL-MI (Structural Correspondence Learning Mutual Information)<sup>[7]</sup>: 首先采用互信息(Mutual Information, MI)构造出枢纽特征,最后分别计算源领域、目标领域枢纽特征与非枢纽特征之间的相关性。

3) SFA(Spectral Feature Alignment)<sup>[10]</sup>: 首先基于共现关系构建领域特定词汇和通用词汇的二分图,然后在二分图上进行谱聚类获取新的特征表示,最后在新的特征上构造分类器。

4) SS-FE(Feature Ensemble Sample Selection)<sup>[11]</sup>: 使用主成分分析法进行特征选择,然后用朴素贝叶斯算法进行情感分类。

5) DANN(Domain Adversarial Neural Network)<sup>[20]</sup>: 利用域对抗神经网络来实现跨领域情感分类。

6) DAmSDA(Domain Adversarial Stacked Denoising Autoencoders)<sup>[20]</sup>: 它是 DANN 的一种改进方法,从边缘堆叠去噪自编码中获得特征表达。

7) AE-SCL-SR(AutoEncoder SCL with Similarity Regularization)<sup>[21]</sup>: 结合自编码和结构对应学习的优势来提高跨领域情感分类的性能。

8) AMN(Adversarial Memory Network)<sup>[22]</sup>: 利用对抗记忆网络实现跨领域情感分类。

表 4 是上述各种方法的分类准确率对比结果。可以看出,本文提出的 AC-BiLSTM 模型在各个情感分析任务中的准确率均是最高的,平均准确率达到了 83.3%。与 NoTransf 方法相比,AC-BiLSTM 模型的平均分类准确率提升了 13.6%,原因在于 NoTransf 方法容易受到文本稀疏性的影响,所以平均分类准确率相对较低;与 SCL-MI、SFA 和 SS-FE 方法相比,AC-BiLSTM 模型的平均分类准确率分别提升了 12.1%、6% 和 7.3%,因为 SCL-MI、SFA 和 SS-FE 的性能取决于枢纽的特征选择方式,这将不能准确地捕获枢纽特征,因此它们的平均分类准确率相对很低;与 DANN、DAmSDA 和 AE-SCL-SR 方法相比,AC-BiLSTM 模型的平均准确率分别提高了 11.4%、9.3% 和 6.5%;与分类效果较佳的 AMN 方法相比,AC-BiLSTM 的平均分类准确率提升了 2.2%。

从表 4 的对比结果可以看出,在跨领域情感分类中,Kitchen 和 Electronics 领域之间的准确率相对其他领域来说较高,这说明 Kitchen 和 Electronics 特征分布最相似,领域自适应效果最佳。同时,从各个方法的平均分类准确率可以看出迁移学习的效果均优于没有进行迁移的情况,这说明迁移学习能够有效地提高分类准确率。

由表 4 还可以看出,与 NoTransf 方法相比,AC-BiLSTM 模型在 12 个跨领域情感分类任务中分别提升了 15.2%、16%、



15.6%、10.3%、12.5%、12.8%、17%、17.1%、6.4%、13.7%、16.7%、11.7%，特别是  $D \rightarrow E$  这个分类任务中，提升了 17.1%。

本文方法优于其他对比方法的原因在于： $N_s$  模型能够自动学习有助于情感分类任务的特征表示，并且源领域和目标领域可以共享该特征；将  $N_s$  中迁移过来的参数在少量目标领域数据上进行微调能够强化  $N_t$  模型去学习对目标领域分类任务有利的特征，从而提高预测精度；卷积操作能够有效捕

获局部上下文特征，BiLSTM 能够考虑特征之间的语义长短期依赖关系；通过引入注意力机制，充分考虑到不同词汇对文本的贡献程度，这将进一步提高分类精度。

图 3 表示准确率和交叉熵损失值随着迭代次数增加而变化的曲线。可以看出，本文提出的模型具有对领域自适应的充分性，同时在交叉熵损失值和分类准确率之间存在较强的相关性，随着迭代次数的增加，准确率整体呈现上升趋势，而交叉熵损失值整体呈现下降趋势。

表 4 不同方法的分类准确率对比结果

Tab. 4 Classification accuracy comparison results of different methods

领域	NoTransf	SCL-MI	SFA	SS-FE	DANN	DAmSDA	AE-SCL-SR	AMN	% %	
									AC-BiLSTM	AC-BiLSTM
$D \rightarrow B$	72.4	73.2	77.0	80.4	75.0	76.1	77.3	82.8	<b>83.4</b>	
$E \rightarrow B$	67.4	68.5	75.5	72.9	70.0	71.9	71.2	77.5	<b>78.2</b>	
$K \rightarrow B$	68.8	69.3	74.0	72.9	71.2	70.0	73.0	78.9	<b>79.5</b>	
$B \rightarrow D$	76.8	78.8	82.6	79.1	73.7	79.7	81.1	81.3	<b>84.7</b>	
$E \rightarrow D$	70.3	70.4	77.0	74.6	71.0	73.1	74.5	76.7	<b>79.1</b>	
$K \rightarrow D$	72.1	72.2	77.0	75.7	71.4	73.8	76.3	80.1	<b>81.3</b>	
$B \rightarrow E$	71.3	71.9	72.0	74.2	68.0	74.9	77.9	81.2	<b>83.4</b>	
$D \rightarrow E$	70.7	71.5	77.0	77.1	74.5	75.0	78.1	80.5	<b>82.8</b>	
$K \rightarrow E$	81.3	82.2	84.6	82.9	82.1	82.4	84.0	82.5	<b>86.5</b>	
$B \rightarrow K$	74.5	77.2	78.0	78.1	78.8	75.4	80.1	82.5	<b>84.7</b>	
$D \rightarrow K$	73.2	74.0	81.0	77.8	77.6	77.4	80.3	83.6	<b>85.4</b>	
$E \rightarrow K$	81.0	82.9	87.1	84.9	84.5	85.0	84.6	90.0	<b>90.5</b>	
平均	73.3	74.3	78.6	77.6	74.8	76.2	78.2	81.5	<b>83.3</b>	

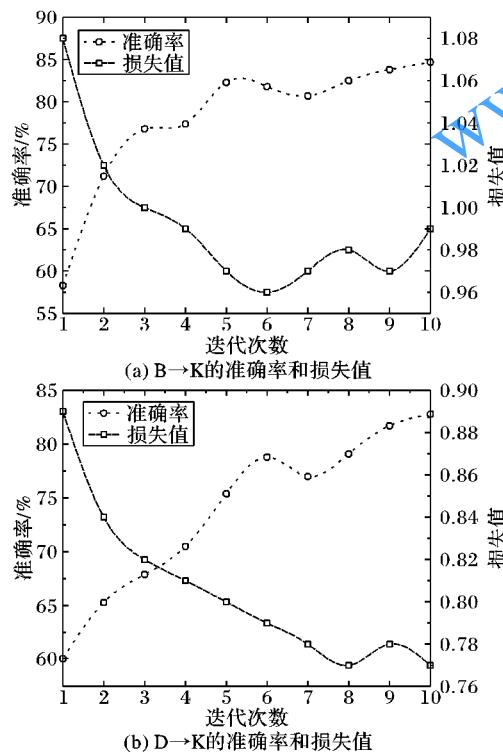


图 3 AC-BiLSTM 在 10 次迭代下的准确率和损失值

Fig. 3 Accuracy and loss values of AC-BiLSTM under 10 iterations

#### 4 结语

本文提出了一种 AC-BiLSTM 模型来进行跨领域情感分析研究。该模型通过引入注意力机制充分考虑到重要单词的情感信息，同时为了避免发生负迁移现象，在目标函数中加入

了正则项约束。在公开的大规模产品评价数据集上的实验结果表明，所提出的 AC-BiLSTM 模型可以有效地提高跨领域情感分类的性能。从整体上来看，本文方法在各个领域下的准确率均优于对比方法。

本文方法的局限性在于语料库的选择主要局限于英文。在后续研究中，将利用中文语料库来验证模型，并在领域跨度更大的环境（例如电子产品与餐饮类评论）中进一步验证模型的效果；同时，下一步还需要将该模型应用到实际工程中。

#### 参考文献 (References)

- [1] AGARWAL B, MITTAL N, BANSAL P, et al. Sentiment analysis using common-sense and context information [J]. Computational Intelligence and Neuroscience, 2015, 2015(78): Article No. 30.
- [2] 陈龙, 管子玉, 何金红, 等. 情感分类研究进展[J]. 计算机研究与发展, 2017, 54(6): 1150–1170. (CHEN L, GUAN Z Y, HE J H, et al. A survey on sentiment classification [J]. Journal of Computer Research and Development, 2017, 54(6): 1150–1170.)
- [3] GAMA J, ŽLIQOBIAITE I, BIFET A, et al. A survey on concept drift adaptation [J]. ACM Computing Surveys, 2014, 46(4): Article No. 44.
- [4] BARDDAL J P, GOMES H M, ENEMBRECK F, et al. A survey on feature drift adaptation: definition, benchmark, challenges and future directions [J]. Journal of Systems and Software, 2017, 127(52): 278–294.
- [5] MOHAMED HUSSEIN D M E D. A survey on sentiment analysis challenges [J]. Journal of King Saud University — Engineering Sciences, 2018, 30(4): 330–338.
- [6] RAVI K, RAVI V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications [J]. Knowledge-Based



- Systems, 2015, 89(17): 14–46.
- [7] BLIZER J, DREDZE M, PEREIRA F. Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification [C]// Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. Stroudsburg, PA: ACL, 2007: 440–447.
- [8] 庄福振, 罗平, 何清, 等. 迁移学习研究进展[J]. 软件学报, 2015, 26(1): 26–39. (ZHUANG F Z, LUO P, HE Q, et al. Survey on transfer learning research [J]. Journal of Software, 2015, 26(1): 26–39.)
- [9] BLITZER J, McDONALD R, PEREIRA F. Domain adaptation with structural correspondence learning [C]// Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2006: 120–128.
- [10] PAN S J, NI X, SUN J T, et al. Cross-domain sentiment classification via spectral feature alignment [C]// Proceedings of the 19th International Conference on World Wide Web. New York: ACM, 2010: 751–760.
- [11] XIA R, ZONG C, HU X, et al. Feature ensemble plus sample selection: domain adaptation for sentiment classification [J]. IEEE Intelligent Systems, 2013, 28(3): 10–18.
- [12] DESHMUKH J S, TRIPATHY A K. Entropy based classifier for cross-domain opinion mining [J]. Applied Computing and Informatics, 2018, 14(1): 55–64.
- [13] 吴斌, 吉佳, 孟琳, 等. 基于迁移学习的唐诗宋词情感分析[J]. 电子学报, 2016, 44(11): 2780–2787. (WU B, JI J, MENG L, et al. Transfer learning based sentiment analysis for poetry of the Tang dynasty and Song dynasty [J]. Acta Electronica Sinica, 2016, 44(11): 2780–2787.)
- [14] QIN B, LIU T, TANG D. Deep learning for sentiment analysis: successful approaches and future challenges [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2015, 5(6): 292–303.
- [15] SCHMIDHUBER J. Deep learning in neural networks: an overview [J]. Neural Networks, 2015, 61(27): 85–117.
- [16] GLOROT X, BORDES A, BENGIO Y. Domain adaptation for large-scale sentiment classification: a deep learning approach [C]// Proceedings of the 28th International Conference on Machine Learning. New York: ACM, 2011: 513–520.
- [17] CHEN M, XU Z, WEINBERGER K Q, et al. Marginalized denoising autoencoders for domain adaptation [C]// Proceedings of the 29th International Conference on Machine Learning. New York: ACM, 2012: 1627–1634.
- [18] CLINCHANT S, CSURKA G, CHIDLOVSKII B. A domain adaptation regularization for denoising autoencoders [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2016: 26–31.
- [19] PENG Y, WANG S, LU B L. Marginalized denoising autoencoder via graph regularization for domain adaptation [C]// Proceedings of the 2013 International Conference on Neural Information Processing, LNCS 8227. Berlin: Springer, 2013: 156–163.
- [20] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks [J]. Journal of Machine Learning Research, 2015, 17(1): 2096–2030.
- [21] ZISER Y, REICHART R. Neural structural correspondence learning for domain adaptation [C]// Proceedings of the 21st Conference on Computational Natural Language Learning. Stroudsburg, PA: ACL, 2017: 400–410.
- [22] LI Z, ZHANG Y, WEI Y, et al. End-to-end adversarial memory network for cross-domain sentiment classification [C]// Proceedings of the 2017 International Joint Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2017: 2237–2243.
- [23] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [C]// Proceedings of the 2013 International Conference on Learning Representations. Stroudsburg, PA: ACL, 2013: 1–12.
- [24] BAHDANA D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [C]// Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA: [s. n.], 2015: 1–15.

This work is partially supported by the Program for Changjiang Scholars and Innovative Research Teams in Universities (IRT\_16R72).

**GONG Qin**, born in 1993, M. S. candidate. Her research interests include natural language processing, deep learning.

**LEI Man**, born in 1992, M. S. candidate. Her research interests include recommendation system, social network.

**WANG Jichao**, born in 1993, M. S. candidate. His research interests include machine learning, data mining.

**WANG Baoqun**, born in 1993, M. S. candidate. His research interests include machine learning, data mining.