DOI:10.19651/j.cnki.emt.2107678

基于 GloVe 模型和注意力机制 Bi-LSTM 的文本分类方法*

周 燕
(华南农业大学数学与信息学院 广州 510642)

摘 要: 为了提高文本分类的准确性,扩展分类任务的多样性,提出一种结合一维卷积神经网络(1D-CNN)和双向长短期记忆网络(Bi-LSTM)的文本分类方法。首先,为了解决近义词、多义词的表征困难,采用 GloVe 模型表示词特征,充分利用全局信息和共现窗口的优势。然后,利用 1D-CNN 进行特征提取,以降低分类器或预测模型的输入特征维数。最后,对分类模块 Bi-LSTM 进行优化,其隐藏层由两个残差块组成,并引入注意力机制进一步改善预测的准确度。在多个公开数据集中进行二元分类和多元主题分类实验。实验结果表明,与其他优秀方法相比,所提方法在准确率、召回率和 F1 得分方面的性能更优,最高准确度达 92.5%,最高 F1 得分为 91.3%。

关键词: 文本分类;GloVe 模型;一维卷积神经网络;双向长短期记忆网络;注意力
中图分类号: TP391.1 **文献标识码:** A **国家标准学科分类代码:** 510.2060

Text classification method based on GloVe model and attention mechanism Bi-LSTM

Zhou Yan
(College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China)

Abstract: To improve the accuracy of text classification and expand different classification tasks, a text classification method combining one-dimensional convolutional neural network (1D-CNN) and bi-directional long short-term memory (Bi-LSTM) network is proposed. Firstly, in order to solve the difficulty of representing synonyms and polysemy, GloVe model is used to represent word features, making full use of the advantages of global information and co-occurrence window. Then, 1D-CNN is used for feature extraction to reduce the input feature dimension of classifier or prediction model. Finally, the classification module Bi-LSTM is optimized, which hidden layer is composed of two residual blocks, and the attention mechanism is introduced to further improve the accuracy of prediction. Binary classification and multiple topic classification experiments are carried out in multiple public data sets. The experimental results show that compared with other excellent methods, the proposed method has better performance in accuracy, recall and F1 score, with the highest accuracy of 92.5% and the highest F1 score of 91.3%.

Keywords: text classification; GloVe model; 1D-CNN; Bi-LSTM; attention

0 引 言

随着互联网使用的普及,社交媒体和网站不断产生海量的文本数据。由于文本数据大多是非结构化的,且包含一定的自然语言结构,因此,很难直接从中推导出有用的信息。传统的相关词文本分类方法[1]或字典分类方法[2]在自动化处理能力、稳定性和准确性方面都较为落后,已经明显不适用于现代文本分类。越来越多的研究利用深度学习进行基于自然语言的情感分类和自然语言推理[3]。

目前,有些研究者对传统方法进行改进,如文献[4]提出一种熵权约束稀疏表示的短文本分类方法。基于稀疏表示理论在过滤后的字典上,为目标函数设计一种熵权约束的稀疏表示方法,从而得到每个文本类的子空间。文献[5]在词频-逆文本词频(term frequency-inverse document frequency,TF-IDF)方法的基础上,提出了一种基于类内分散、类间分散和权重协调因子的改进信息增益文本分类方

收稿日期:2021-08-24
* 基金项目:2019 年广东省研究生教育创新计划项目(2019JGXM18)资助

法。这些改进方法针对特定文本分类取得了一些效果，但依然有较大局限。

长短期记忆(long short term memory，LSTM)网络提取高级文本信息的能力在文本分类中发挥着重要作用[6]。如文献[7]在 LSTM 的基础上，引入注意力机制，并将基于注意力机制的 LSTM 用于新闻文本分类。文献[8]提出一种基于可分离卷积层的文本分类方法，利用可分解卷积网络代替传统的卷积网络，一层词嵌入卷积层用来提取单词的词嵌入特征，另一层提取上下文特征，但该方法限定特定的数据库。为进一步提高非结构化文本的情感分类准确度，文献[9]提出了用于文本分类的一维卷积神经网络(convolutional neural network，CNN)方法。利用自注意力机制的 CNN 捕捉邻近词之间潜在的关联特征。但单纯使用 CNN 使得对文本特征的记忆能力有所缺失。文献[10]对 CNN、递归神经网络(recursive neural network，RNN)、LSTM 和门控递归单元(gated recurrent unit，GRU)方法进行了研究，并对词向量提取进行了分析。得出深度神经网络的结构和词向量的提取对文本分类的准确率和鲁棒性具有重要作用。

为了利用 LSTM 和 CNN 的各自优点并避免个体缺陷，本文提出了一种改进双向长短期记忆(Bi-directional long short term memory，Bi-LSTM)神经网络和一维 CNN 的文本分类方法，并利用注意力机制增强方法性能。其主要创新之处总结如下：

1)采用 GloVe 模型进行特征表示，优于传统的 Word2Vec 模型；并利用一维 CNN 从句子不同位置提取特征，由此降少了输入特征数量。

2)对 Bi-LSTM 进行改进，利用卷积层输出的特征进行上下文信息提取。其中，注意力机制在输入上使用偏差配准，向高度相关的输入组件分配权重，由此进一步减少训练阶段的参数数量。同时，也改善了不同文本长度序列的权重分布。

1　文本特征的表示

1.1　文本预处理

在将数据传入模型前，首先对文本进行预处理，其目的是过滤掉无关噪声，最大限度保留词特征，预处理操作包括移除空白符和无意义词，规范形式转换，以及移除重复词。预处理后的数据集提供了唯一且有意义的词序列，每个词具有唯一标识。嵌入层学习每个预处理后的输入 token 的分布式表征。token 的表征反映了可能出现在相同上下文的词之间的潜在关系。

1.2　基于 GloVe 的特征表示

传统的词特征表示方法采用 Word2Vec，但 Word2Vec 的缺点较为明显，词和向量都是 1 对 1 的关系，难以解决近义词，多义词，且需要词类比任务 gram 模型。而 GloVe 继承了 Word2Vec 的绝大部分优点，是一种无监督技术，使用了全局统计信息、全局先验信息，并融合了共现窗口的优势，使得在近义词、多义词的处理上更具有优势，能够更多地蕴含语义和语法信息。其基本模型结构如图 1 所示。

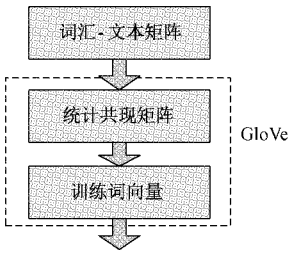


图 1　GloVe 模型

GloVe 的模型定义如下：

$$J=\sum_{i,j}^{N} f(\boldsymbol{M}_{i,j})[\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j+\boldsymbol{b}_i+\boldsymbol{b}_j-\ln(\boldsymbol{M}_{i,j})]^2 \tag{1}$$

式中：共现矩阵 $\boldsymbol{M}$ 的元素 $\boldsymbol{M}_{i,j}$ 表示词 i 和 j 共同出现在一个窗口中的次数；$\boldsymbol{M}$ 为 $N\times N$ 的矩阵；词 i 的向量为 $\boldsymbol{v}_i$，其偏差为 $\boldsymbol{b}_i$；词 j 的词向量 $\boldsymbol{v}_j$，其偏差项为 $\boldsymbol{b}_j$；通常设置窗口大小为 5～10；权重函数为 f，其定义如下：

$$f(l)=\begin{cases}(l/l_{\max})^{\alpha}, & l<l_{\max}\\ 1, & \text{其他}\end{cases} \tag{2}$$

其中，权重函数 f 满足：当 $l=0$ 时，$f(0)=0$。且 f 是非递减函数，即，词的共现次数增多，f 值不会下降。此外，当词出现的频率很高时，也不会过度加权。相关研究表明[11]：当 $l_{\max}=100$，$\alpha=0.75$ 时，GloVe 的特征表示最佳，故本文也采用该数值。

GloVe 模型比 Word2Vec 模型更佳，且不需要利用神经网络进行训练，可直接使用语料库对词向量进行计算，更容易并行化，由于使用全局信息，表征更准确。

2　文本特征的提取与分类

2.1　基于一维 CNN 的文本特征提取

二维 CNN 广泛用于图像处理，利用卷积层和子采样层(或最大池化层)，通过一连串的卷积和池化构建特征图。对于 1D-CNN，卷积层使用一维互相关操作，涉及到输入文本上从左到右的滑动卷积窗口，以及可变大小的卷积核[12]。由于一维互相关操作使用一维全局最大池化层的池化层，因此，减少了文本编码所需的特征数量。图 2 所示为本文 1D-CNN 的基本架构。

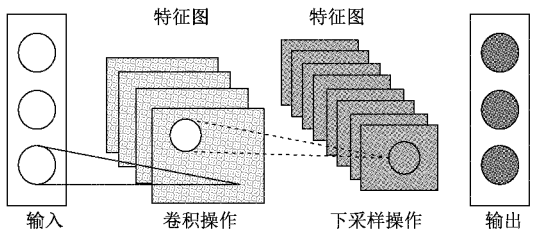


图 2　文本 1D-CNN 结构

本文方法中的卷积作用是从输入文本中提取特征，利用卷积层从原始文本中提取底层语义特征并减少维数；与之相比，Bi-LSTM 则将文本作为序列数据来处理。通过多个一维卷积核，在输入向量上执行卷积。式(3)定义了嵌入向量进行串联而得到的序列文本向量：

$$\boldsymbol{X}_{1:T}=[x_1,x_2,x_3,\cdots,x_T] \tag{3}$$

式中：T 为文本中的 token 数量。为了利用一维 CNN 捕捉文本的固有特征，将不同大小的卷积核应用到 $\boldsymbol{X}_{1:T}$，以捕捉文本的一元分词、二元分词和三元分词的特征。在第 t 次卷积过程中，取从 t 到 $t+d$ 的 d 个词的窗口作为输入，生成如下特征：

$$\boldsymbol{h}_{d,t}=\tanh(\boldsymbol{W}_d\boldsymbol{x}_{t:t+d-1}+\boldsymbol{b}_d) \tag{4}$$

式中：$\boldsymbol{x}_{t:t+d-1}$ 为窗口中词的嵌入向量，$\boldsymbol{W}_d$ 为可学习权重矩阵，$\boldsymbol{b}_d$ 为偏差向量。由于必须将每个滤波器应用到不同的文本区域，因此，卷积核大小为 d 的滤波器的特征图为：

$$\boldsymbol{h}_d=[\boldsymbol{h}_{d1},\boldsymbol{h}_{d2},\boldsymbol{h}_{d3},\cdots,\boldsymbol{x}_{T\ d+1}] \tag{5}$$

使用具有不同宽度卷积核的好处在于：可以捕捉到多个邻近词之间的隐藏关联，减少特征学习过程中可训练参数的数量[13]。通过大量卷积通道来处理输入，每个通道包含不同时间步长的数值。由此，在池化过程中，每个卷积通道的输出将为该通道中所有时间步长的最大值。

对于每个卷积通道，向卷积核大小为 d 的特征图应用 max-overtime 池化，可以得到：

$$\boldsymbol{p}_d=\mathrm{Max}^t(\boldsymbol{h}_{d1},\boldsymbol{h}_{d2},\boldsymbol{h}_{d3},\cdots,\boldsymbol{x}_{T\ d-1}) \tag{6}$$

为了得到窗口的最终特征图，对每个滤波器进行串联，提取一元、二元和三元分词隐藏特征为：

$$\boldsymbol{h}_d=[\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3] \tag{7}$$

使用 CNN 的好处是降低了分类器或自然推理预测模型的输入特征维数。且对于文本分类，一维 CNN 比 LSTM 更准确，特别是特征提取步骤中更是如此。

2.2　基于优化 Bi-LSTM 的分类

所提方法以基于注意力的 Bi-LSTM 为文本分类的基础。虽然上一节的一维 CNN 缩小了输入特征的维度，但对于所有输入词和句，其最终分类结果间的相关性并非相同。因此，本文充分利用一维 CNN 和 Bi-LSTM 的各自优势，完成对文本语义的相依性进行有效编码。本文方法的主要框架结构如图 3 所示。首先，利用 GloVe 进行文本特征表示，然后，利用一维 CNN 进行特征提取，通过优化的 Bi-LSTM 将提取后的文本特征进行分类，其中，注意力机制选择与最终分类结果高度相似的特征，并完成类别输出。

1)优化的 Bi-LSTM

LSTM 使用的各种门控机制有助于处理长期相依性，并解决梯度消失问题，特别是使用较长序列作为输入的情况。然而，LSTM 无法提取未来 token 的上下文信息，也无法提取局部上下文信息。此外，LSTM 不能识别一个文本的不同部分之间的关系。Bi-LSTM 神经网络由双向运行的 LSTM 单元组成，以结合过去和未来的上下文信息，能

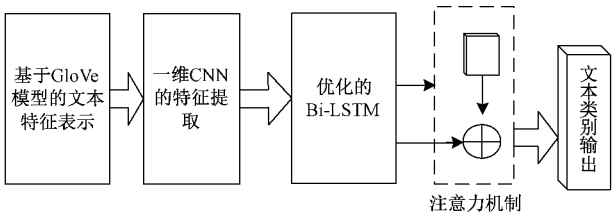


图 3　本文方法的整体框架图

够学习长期相依性，且无需保留重复的上下文信息。因此，在处理连续建模问题中表现出优异性能，并被广泛用于文本分类。不同于 LSTM 网络，Bi-LSTM 网络包含在两个方向上传播的两个并行层，通过正向和反向传递以捕捉两个上下文中的相依性[14]。本文在 Bi-LSTM 的基础上进一步优化，其结构如图 4 所示。其中，文本信息在水平和垂直方向流动，时间维度在垂直方向，空间维度在水平方向。与一般 Bi-LSTM 不同，本文方法的隐藏层由两个残差块组成，这两个残差块均由 2 个 Bi-LSTM 组成，优化后的 Bi-LSTM 可以融合累积文本特征，对叠加的文本信息进行获取。

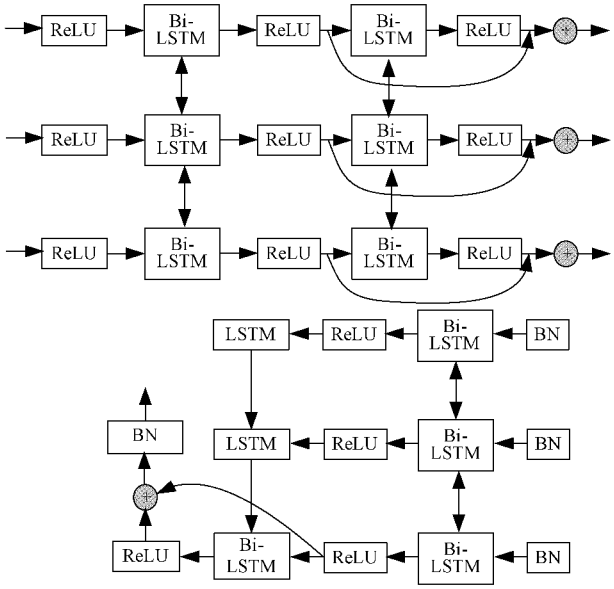


图 4　优化的 Bi-LSTM 模型

2)注意力机制

由于无法将所有信息压缩到一个固定大小的向量中，因此，存在信息丢失和梯度消失问题[15]。随着输入长度增加，准确度会下降。为此，在 Bi-LSTM 的基础上，增加注意力机制以改善预测准确度。其中，解码器在每个时间步对输出词进行预测，并再次参考来自编码器的整个输入句。但并不是对所有输入的句子进行等价参考。注意力机制向每个输入赋予初始权重，并根据每个输入与最终预测之间的相关性，在训练过程中对这些权重进行更新。注意力机制的公式如下：

$$\boldsymbol{Z}_t=\tanh(\boldsymbol{w}\boldsymbol{y}_t+\boldsymbol{b}_w) \tag{8}$$

式中：$\boldsymbol{w}$ 为注意力机制的权重矩阵，其矩阵维度为 128×320；$\boldsymbol{b}_w$ 为偏置向量，维度为 128×1 维列向量；$\boldsymbol{y}_t$ 的维度为 320×1。

注意力权重 $\boldsymbol{\alpha}_t$ 的定义为：

$$\boldsymbol{\alpha}_t = \frac{\exp(\boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{u}_w)}{\sum_{t=1}^{n}(\boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{u}_w)} \tag{9}$$

式中：$\boldsymbol{u}_w$ 为初始化的列向量，维度为 128×1，且 $\boldsymbol{\alpha}_t$ 是归一化操作的结果。

$$\boldsymbol{U} = \sum_{t-1}^{n}\boldsymbol{\alpha}_t\boldsymbol{y}_t \tag{10}$$

式(10)的作用是将计算出来的权重作为各个时刻的输出权重，对 Bi-LSTM 进行加权求和。

3　实验和结果

本文在配置了英特尔酷睿 i7-4500u CPU@ 1.8 GHz，内存 16 GB 的台式机器上进行仿真实验。采用 Python 3.5.0 语言编程，使用 Keras 提供的深度神经网络框架，使用 networkx、gensim 和 scikit-learn 软件包。在参数设置方面，嵌入大小为 500，所有数据集的批大小设为 128，dropout 率设为 0.2。代数设为 10，$l_{max}=100$，$\alpha=0.75$。

将仿真结果与文献[7]提出的注意力 LSTM 方法(attention-based long-short term memory，A-LSTM)、文献[9]提出的自注意力 CNN 方法(self-attention-based convolutional neural network，SA-CNN)、文献[10]中的 GRU 方法进行比较。

3.1　数据集

句子极性(sentence polarity)数据集包含 5 000 个正面电影评论和 5 000 个负面电影评论[16]。

主观性(subjectivity)数据集包含基于主观状态标记的评论，含有 5 000 个主观句和 5 000 个客观句[17]。

新闻(News)数据集是包含 32 602 个短文本文档的集合，这些文档是从网站 RSS 订阅源采集到的新闻，并根据其主题进行了分类。主题包括体育、商业、健康、科技、世界和娱乐等。文档由标题、描述、链接、ID、数据、源和新闻类别所组成。本文仅使用新闻的描述和类别[18]。

多领域情感(multi-domain sentiment)数据集包含从电商平台得到的 8 000 条产品评价，这些评论包括对书籍、DVD、电子产品和厨房用具的评论。对于这 4 类产品，每类均含有 1 000 条正面评论和 1 000 条负面评论。

20 类新闻组(20 Newsgroup)数据集包含被分入 20 个不同类别的 20 000 个新闻组文档[19]。

3.2　评估指标

实验中，使用 3 个评估度量进行模型评估：F1 得分、准确率和召回率。这些参数定义如下：

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

$$Recall = \frac{TP}{TP+FN} \tag{12}$$

$$F1 = \frac{2*(Precision*Recall)}{(Precision+Recall)} \tag{13}$$

式中：TP 表示正样本标签且被正确分入正样本类的数量；TN 表示负样本标签被正确分入负样本类的数量；FN 表示正样本标签被错误分入到负样本类的数量；FP 表示负样本标签被错误分入正样本类的数量；召回率(Recall)为被模型分入正样本类的文本数量在所有实际被标注为正样本数据中的占比；准确率(Precision)表示被模型分入正样本类的文本中实际包含正样本标签所占百分比；F1 得分(F1-score)为召回率和精度的调和平均值。

3.3　情感分类

情感分类是对文本的情感语义进行客观或主观、正面或负面的分类，是一种二元分类任务。在二元分类任务的几个数据集中，各方法的准确率、召回率和 F1 得分如表 1 所示。可以看出，所提方法具有更高的准确率、召回率和 F1 得分，其主要得益于较好的特征表征提取和文本语义分类，SA-CNN、A-LSTM 和 GRU 方法均采用 Word2Vec 模型，对于近似词和多义词的表征存在缺陷，只能采用 1 对 1 的方式表示，容易产生误分类。而本文采用 GloVe 的文本特征表示，低层语义表示更准确。此外，采用优化的 Bi-LSTM 配合注意力机制，对长距离词相依性进行有效编码，在语义特征分类方面具有优势。SA-CNN、A-LSTM 与所提方法相比，没有高效的独立文本特征提取模块，本文方法采用一维 CNN 网络进行文本特征提取。GRU 方法作为一种改进的循环神经网络方法，其性能表现最差，是因为 GRU 无法提取未来 token 的上下文信息，也无法提取局部上下文信息，其缺点与 LSTM 类似。

表 1　各方法在情感分类任务中的性能指标

数据集	度量	SA-CNN	A-LSTM	GRU	本文方法
句子极性	准确率	86.4	82.3	80.3	**91.1**
	召回率	86.7	83.3	85.2	**92.3**
	F1	87.9	83.7	80.8	**91.3**
主观性	准确率	89.1	88.6	85.1	**90.8**
	召回率	89.8	87.9	85.1	**90.7**
	F1	89.8	87.8	85.0	**90.7**
书籍	准确率	82.8	81.8	79.3	**86.1**
	召回率	82.6	81.7	79.1	**86.0**
	F1	82.6	81.7	79.1	**86.3**
DVD	准确率	81.6	80.9	80.9	**87.5**
	召回率	81.9	81.1	80.7	**87.2**
	F1	82.0	81.4	80.7	**87.7**
电子产品	准确率	81.0	82.4	80.3	**86.0**
	召回率	81.0	82.3	79.9	**85.9**
	F1	80.9	82.3	79.9	**85.8**
厨房用具	准确率	85.1	83.5	81.2	**88.2**
	召回率	84.9	83.3	81.2	**87.9**
	F1	84.9	83.3	81.1	**88.1**

3.4　多元主题分类

一般主题分类是一种多元分类任务，其数据集更为复杂，如 News 数据集包含的主题有体育、商业、健康、科技、世界和娱乐等，文档由标题、描述、链接、ID、数据、源和新闻类别所组成。对这些类别的分类需要对各个类别的语义特征进行提取分类，因此，比二元分类更为复杂。在 News 和 20NG 数据集中，各方法的准确率、召回率和 F1 得分如表 2 所示。其中，News 数据集上的结果通过十折交叉验证得到，而 20 NG 数据集则采用标准的训练/测试分割。由表 2 可知，与二元分类的结果(表 1)相比，各方法的性能指标都有所下降。这是因为语义特征更加细粒度化，甚至会有很多冗余特征的出现，所以分类的难度更大。但所提方法依然表现出较高的准确率、召回率和 F1 得分。这主要得益于所提方法在文本特征表示、提取和分类过程的改进，所提方法能够准确表达出相同上下文的词之间的潜在关系。

表 2　各方法在主题分类任务上的性能指标

数据集	度量	SA-CNN	A-LSTM	GRU	本文方法
20NG	准确率	76.9	76.4	71.7	**80.1**
	召回率	77.6	75.1	71.3	**79.2**
	F1	77.8	76.0	71.2	**79.3**
News	准确率	76.4	75.8	73.3	**78.3**
	召回率	77.4	74.9	72.2	**78.5**
	F1	77.3	75.8	72.3	**79.2**

3.5　代数与数据大小的实验分析

为了研究代数和数据大小对方法性能的影响，为了便于研究，以句子极性数据集为例，数据体量并不是太大，包含 5 000 个正面影评和 5 000 个负面影评。

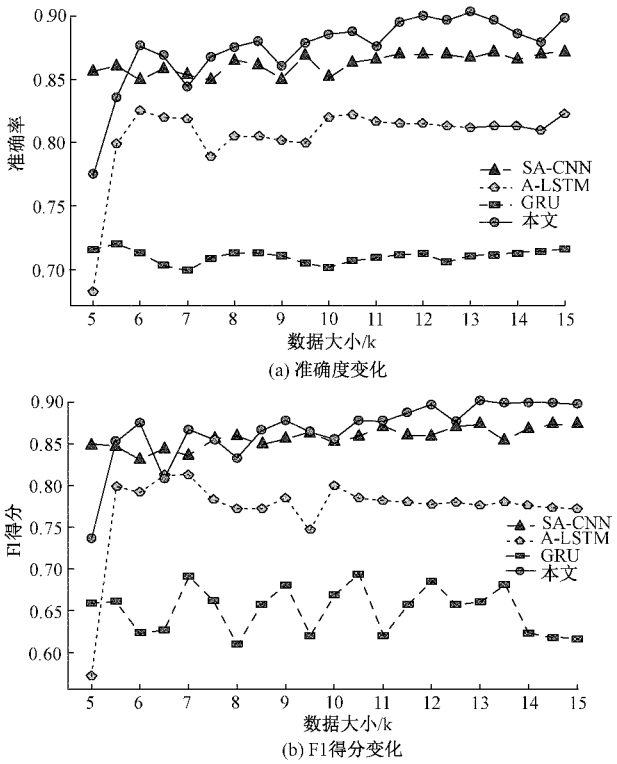
表 3 所示为随着代数的增加，各个方法的最高准确度、F1 得分、代数等值。可以看出，本文方法的最高准确度和平均准确度分别为 92.5%和 91.1%，均优于其他方法，且准确度会随着代数的增加而上升，解决了数据丢失和长期相依性问题。其后的准确度排名依次为 A-LSTM、SA-CNN 和 GRU。所提方法的最高 F1 得分为 91.3%，优于其他模型，其后依次为 SA-CNN、A-LSTM 和 GRU。各方法实现最优准确度的代数值分别为：SA-CNN 为 8 代，A-LSTM 为 15 代，GRU 为 16 代，本文方法为 13 代。SA-CNN 的代数最少说明该方法可以较快地达到收敛的状态，从 SA-CNN 框架看，自注意力 CNN 的主要计算量集中在 CNN 网络中，相比于其他循环神经网络的改进型，其效率较高。

图 5 所示为代数固定为 10 代的情况下，随着数据大小从 5 k 增加到 15 k 的各方法性能比较。其中，图 5(a)为准确度变化；图 5(b)是 F1 得分的变化。由图 5(a)可知，随着数据大小的增加，本文方法的准确度呈上升趋势，实现了更佳的性能，且准确度会随着数据量的增加而上升，解决了数

表 3　不同代数的实验结果　%

模型	SA-CNN	A-LSTM	GRU	本文模型
准确率(最大)	87.9	85.7	81.9	92.5
准确率(平均)	86.4	82.3	80.3	91.1
F1 得分	87.9	83.7	80.8	91.3
召回率	86.7	83.3	85.2	92.3
数据大小	20 k	20 k	20 k	20 k
代数	8	15	16	13

据丢失和长期相依性问题。在数据大小达到 11 k 之后，SA-CNN 方法的准确度渐趋稳定。A-LSTM 方法在数据大小增加到 10 k 后保持了稳定的准确度。对于 MLP 方法，其准确度均保持在 0.70～0.75 之间。在图 5(b)中，除 GRU 之外，其他方法均表现出与图 5(a)相似的趋势，而 GRU 方法则表现出不稳定数值，变化范围非常大。这是因为 GRU 的学习过程并不完整，用于分类的规则是不明确的。



(a) 准确度变化

(b) F1得分变化

图 5　数据大小与性能关系变化

4　结　　论

本文提出了用于文本分类的深度神经网络模型，基于一维 CNN 技术和改进注意力机制的 Bi-LSTM，且文本特征表示采用更为先进的 Glove 模型，实验结果验证了所提方法的性能优秀。此外，随着训练数据大小的增加和训练代数的增加，本文方法的准确度也随之上升。由此解决了

当前模型中存在的长期相依性问题，以及随着训练数据量增加而面临的数据丢失问题。

未来，本文将研究其他扩展应用，例如通过迁移学习等技术，将更充足的训练数据应用到模型中，还可增加用于多类预测的分类标签。

参考文献

[1] ANDRADE D, SDDAMASA K, TAMURA A, et al. Cross-lingual text classification using topic-dependent word probabilities[C]. the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015: 173-181.

[2] SIBUNRUANG C, POLPINIJ J. Concept-based text classification of Thai medicine recipes represented with ancient Isan language [M]. Springer International Publishing, Berlin, Germany, 2015.

[3] 杨睿，刘瑞军，师于茜，等. 面向智能交互的视觉问答研究综述[J]. 电子测量与仪器学报，2019，33(2)：117-124.

[4] 脱婷，马慧芳，李志欣，等. 熵权约束稀疏表示的短文本分类算法[J]. 电子学报，2020，48(11)：53-59.

[5] YE X M, MAO X M, XIA J C, et al. Improved approach to TF-IDF algorithm in text classification[J]. Computer Era, 2019, 45(1): 109-116.

[6] ZHANG L, WANG S, LIU B. Deep learning for sentiment analysis: A survey [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 34(Jan.): 1253-1267.

[7] 蓝雯飞，徐蔚，汪敦志，等. 基于 LSTM-Attention 的中文新闻文本分类[J]. 中南民族大学学报(自然科学版)，2018，37(3)：133-137.

[8] 严佩敏，唐婉琪. 基于可分离卷积神经网络的文本分类[J]. 电子测量技术，2020，43(13)：7-12.

[9] LU W, DUAN Y, SONG Y. Self-attention-based convolutional neural networks for sentence classification [C]. International Conference on Computer and Communications (ICCC), Chengdu, China: IEEE Press, 2020: 69-78.

[10] AYDOAN M, KARCI A. Improving the accuracy using pre-trained word embeddings on deep neural networks for Turkish text classification[J]. Physica A: Statistical Mechanics and its Applications, 2020, 541(C): 91-103.

[11] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]. Conference on Empirical Methods in Natural Language Processing, Doha, Qatar: IEEE Press, 2014: 1532-1543.

[12] KALCHBRENNER N, GREFENSTETTE E, BLUNSOM P. A convolutional neural network for modelling sentences[J]. Eprint Arxiv, 2014, 37(1): 2188-2196.

[13] 朱奕杰，蔺宏伟. 基于正则化一维卷积神经网络的网格模型显著性检测[J]. 计算机辅助设计与图形学学报，2020，32(2)：203-212.

[14] WANG S, HUANG M, DENG Z. Densely connected CNN with multi-scale feature attention for text classification[C]. International Joint Conferences on Artificial Intelligence, Stockholm, Sweden, 2018: 4468-4474.

[15] 王路，张璐，李寿山，等. 基于注意力机制的上下文相关的问答配对方法[J]. 中文信息学报，2019，33(1)：125-132.

[16] MOUMITA R. Empirical study of different classifiers for sentiment analysis[J]. Data Mining & Knowledge Engineering, 2014, 6(4): 187-195.

[17] 樊红梅. CNN 涉华报道中的主观性研究[D]. 哈尔滨：哈尔滨工程大学，2014.

[18] VITALE D, FERRAGINA P, SCAIELLA U. Classification of short texts by deploying topical annotations[C]. Advances in Information Retrieval, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012: 376-387.

[19] THELWALL M, BUCKLEY K, PALTOGLOU G, et al. Sentiment strength detection in short informal text [J]. Journal of Assoc. Information Science Technology, 2010, 61(12): 2544-2558.

作者简介

周燕，博士，主要研究方向为数据挖掘、深度学习、文本分类等中文信息处理。

E-mail：pigcbdlol@163.com