

基于 CNN 和 BiLSTM 网络特征融合的文本情感分析

李 洋, 董红斌*

(哈尔滨工程大学 计算机科学与技术学院 哈尔滨 150001)

(* 通信作者电子邮箱 852221609@qq.com)

摘 要: 卷积神经网络 (CNN) 和循环神经网络 (RNN) 在自然语言处理上得到广泛应用, 但由于自然语言在结构上存在着前后依赖关系, 仅依靠卷积神经网络实现文本分类将忽略词的上下文含义, 且传统的循环神经网络存在梯度消失或梯度爆炸问题, 限制了文本分类的准确率。为此, 提出一种卷积神经网络和双向长短时记忆 (BiLSTM) 特征融合模型, 利用卷积神经网络提取文本向量的局部特征, 利用 BiLSTM 提取与文本上下文相关的全局特征, 将两种互补模型提取的特征进行融合, 解决了单卷积神经网络模型忽略词在上下文语义和语法信息的问题, 也有效避免了传统循环神经网络梯度消失或梯度弥散问题。在两种数据集上进行对比实验, 实验结果表明, 所提特征融合模型有效提升了文本分类的准确率。

关键词: 词向量; 卷积神经网络; 双向长短时记忆; 特征融合; 文本情感分析

中图分类号: TP391 **文献标志码:** A

Text sentiment analysis based on feature fusion of convolution neural network and bidirectional long short-term memory network

LI Yang, DONG Hongbin*

(College of Computer Science and Technology, Harbin Engineering University, Harbin Heilongjiang 150001, China)

Abstract: Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) are widely used in natural language processing, but the natural language has a certain dependence on the structure, only relying on CNN for text classification will ignore the contextual meaning of words, and there is a problem of gradient disappearance or gradient dispersion in the traditional RNN, which limits the accuracy of text classification. A feature fusion model for CNN and Bidirectional Long Short-Term Memory (BiLSTM) was presented. Local features of text were extracted by CNN and global features related to text were extracted by BiLSTM network. The features extracted by the two complementary models were merged to solve the problem of ignoring the contextual semantic and grammatical information of words in a single CNN model, and the fusion model also effectively avoided the problem of gradient disappearance or gradient dispersion in traditional RNN. The experimental results on two kinds of datasets show that the proposed fusion feature model can effectively improve the accuracy of text classification.

Key words: word vector; Convolution Neural Network (CNN); Bidirectional Long Short-Term Memory (BiLSTM); feature fusion; text sentiment analysis

0 引言

近年来, 随着互联网的高速发展, 每天不断地产生大量文本、音频、图片、视频等数据, 其中文本信息的数据量最大, 但文本信息杂乱无章, 人工很难区分和整理, 因此, 对文本数据进行分类变得越来越重要。最初深度学习 (Deep Learning) 在图像和语音识别领域取得了优异成绩, 而近年来深度学习在文本分类中的应用也越来越广泛, 与传统的文本分类方法, 如朴素贝叶斯、K 最近邻 (K Nearest Neighbor, KNN) 和支持向量机 (Support Vector Machine, SVM) 等算法相比, 不需要人工设计特征^[1], 而是利用深度学习模型自动提取文本特征, 显著提高了文本分类的速度, 并取得比传统文本分类更好的分类效果。

深度学习是机器学习中的一条重要分支, 深度学习是利用多重非线性变换结构对数据进行高阶抽象的算法^[2]。最

近几年, 深度学习算法在自然语言处理领域取得了十分出色的成果, 其中卷积神经网络 (Convolutional Neural Network, CNN) 充分利用多层感知器的结构, 具备很好的学习复杂、高维和非线性映射关系的能力, 在图像识别任务和语音识别任务中得到广泛的应用^[3-4], 并取得很好的效果。Kalchbrenner 等^[5]提出把 CNN 应用于自然语言处理, 并设计了一个动态卷积神经网络 (Dynamic Convolution Neural Network, DCNN) 模型, 以处理不同长度的文本; Kim^[6]提出的英文文本分类的模型, 将经过预处理的词向量作为输入, 利用卷积神经网络实现句子级别的分类任务。虽然卷积神经网络在文本分类中取得了巨大的突破, 但是卷积神经网络更加关注局部特征而忽略词的上下文含义, 这对文本分类的准确率有一定的影响, 所以本文利用双向长短时记忆 (Bidirectional Long Short-Term Memory, BiLSTM) 网络解决卷积神经网络模型忽略词上下文

收稿日期: 2018-03-29; 修回日期: 2018-06-16; 录用日期: 2018-06-29。 基金项目: 国家自然科学基金资助项目 (61472095)。

作者简介: 李洋 (1991—), 男, 山东莱芜人, 硕士研究生, 主要研究方向: 机器学习、自然语言处理; 董红斌 (1963—), 男, 河北唐山人, 教授, 博士生导师, 博士, CCF 会员, 主要研究方向: 演化计算、计算智能、数据挖掘、多 Agent 系统、机器学习。

含义的问题。

神经网络在特征的自动学习和表达中发挥越来越重要的作用,对于序列化输入,循环神经网络(Recurrent Neural Network, RNN)能够把邻近位置信息进行有效整合^[7-8],处理自然语言处理的各项任务。RNN 的子类长短期记忆网络模型(Long Short-Term Memory, LSTM)^[9-10]对序列信号进行建模的神经网络模型,可以作为复杂的非线性单元用来构建大型的神经网络结构,同时能避免 RNN 的梯度消失问题,具有更强的“记忆能力”,能够很好地利用上下文特征信息和对非线性关系进行拟合的能力,保留文本的顺序信息。RNN 有多种变种循环神经网络模型,主要应用于文本分类的有双向循环神经网络(Bidirectional RNN)^[11],由于文本中词的语义信息不仅与词之前的信息有关,还与词之后的信息有关,两个 RNN 左右传播组合而成的双向循环神经网络能进一步提高文本分类的准确率。

本文的主要贡献如下:

- 1) 利用 BiLSTM 代替传统 RNN 和 LSTM, BiLSTM 解决传统 RNN 中梯度消失或梯度爆炸问题;同时一个词的语义与它之前信息和之后信息都有关,而 BiLSTM 充分考虑词在上下文的含义,克服了 LSTM 不能考虑词之后信息的弊端。
- 2) 将卷积神经网络和 BiLSTM 进行融合,既能利用卷积神经网络提取局部特征的优势,又能利用双向长短期记忆网络兼顾文本序列全局特征的优势,利用 BiLSTM 解决卷积神经网络在文本分类中忽略词的上下文含义的问题,提高了特征融合模型在文本分类的准确率。

1 词向量

1.1 词嵌入

深度学习方法进行文本分类的第一步是将文本向量化,利用词向量表示文本,作为卷积神经网络和 BiLSTM 网络模型的输入。传统的文本表示方法是基于向量空间模型或 one-hot 表示:向量空间模型中向量维度与词典中词的个数线性相关,随着词数增多容易引起维度灾难;而 one-hot 虽然简单但忽略了词之间的语义相关性。词向量解决了向量空间模型和 one-hot 的问题,将高维稀疏的特征向量映射为低维稠密的词向量,有效避免了维度灾难的发生,且可以直接计算词语之间的语义相关性。Bengio 等^[12]提出用神经网络概率语言模型(Neural network Probabilistic Language Model, NPLM)来处理文本信息。Mikolov 等^[13-14]基于 NNLM (Neural Network Language Model) 提出 Word2vec 模型,并给出了利用 CBOW (Continuous Bag-Of-Words) 和 Skip-gram 两种模型构建词向量。与 NNLM 不同的是 Word2vec 不局限于利用前 $n-1$ 个单词来预测第 n 个单词,而是以 n 为大小的窗口来计算窗口中心词出现的概率,实现了利用上下文预测。CBOW 和 Skip-gram 两个都是以 Huffman 树作为基础, Huffman 树中非叶子节点存储的中间向量的初始化值是零向量,而叶子节点对应词的词向量是随机初始化的。CBOW 是根据上下文预测一个词,训练过程中由三部分构成,如图 1 所示,分别为:输入层(input)、映射层(projection)和输出层(output);输入层为词 $W(t)$ 周围的 $n-1$ 个单词的词向量,如果 n 取 5,则词 $W(t)$ 的前两个词为 $W(t-2)$, $W(t-1)$, 后两个词为 $W(t+1)$, $W(t+2)$, 它们对应的向量记为 $V(W(t-2))$, $V(W(t-1))$, $V(W(t+1))$, $V(W(t+2))$, 从输入层到映射层将 4 个词的向量形式相

加,而从映射层到输出层需构造 Huffman 树,从根节点开始,映射层的值沿着 Huffman 树进行 logistic 分类,并不断修正各中间向量与词向量,得到词 $W(t)$ 所对应的词向量 $V(W(t))$ 。

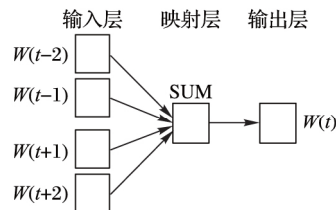


图 1 CBOW 模型

Fig. 1 CBOW model

Skip-gram 模型与 CBOW 刚好相反,如图 2 所示,同样由输入层、映射层和输出层构成。Skip-gram 输入是当前词 $W(t)$ 的向量形式,输出是周围词的向量形式,通过当前词来预测周围的词,如果上下文窗口大小设置为 4,已知中间词 $W(t)$ 所对应的向量形式为 $V(W(t))$,利用 $V(W(t))$ 预测出周围 4 个词所对应的向量形式, $Context(w) = \{V(W(t+2)), V(W(t+1)), V(W(t-1)), V(W(t-2))\}$, Skip-gram 模型计算周围词向量是利用中间词向量 $V(W(t))$ 的条件概率值来求解,公式如下:

$$P(V(W(i)) | V(W(t))) \quad (1)$$

其中 $V(W(i)) \in Context(w)$ 。

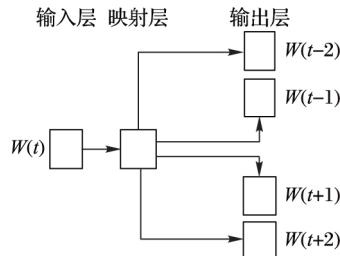


图 2 Skip-gram 模型

Fig. 2 Skip-gram model

1.2 词向量相似度

通过将词向量与向量空间模型和 one-hot 对比,发现词向量维度由上千维稀疏向量变为了低维稠密的向量形式,同时词向量中包含了自然语言中的语义和语法关系,通过 Skip-gram 模型训练得到的词向量可以更加容易地计算词语之间的语义相关性,利用词向量之间余弦距离表示词语之间的关系,余弦相似度值越大,词语间关系越大,余弦相似度值越小,词语间关系越小。如图 3 所示是利用如家酒店 3 000 条评论分词、去除停用词等预处理后计算与“房间”这个词最相关的 10 个词及余弦相似度值。

Enter word or sentence(EXIT to break): 房间	
Word: 房间 Position in vocabulary: 1137	
Word	Cosine distance
旧	0.98829
干净	0.98671
小	0.98663
舒服	0.98231
设施	0.98220
早餐	0.98.96
整洁	0.97443
热情	0.97326
一般	0.97198
态度	0.97049

图 3 词向量语义相似度

Fig. 3 Semantic similarity of word embedding

2 CNN 与 BiLSTM 特征融合模型

2.1 卷积神经网络模型

本文利用图 4 所示的卷积神经网络模型提取局部特征。卷积神经网络进行文本分类时,首先将词 $W(i)$ 利用 word2vec 转化为对应的词向量 $V(W(i))$,并将由词 $W(i)$ 组成的句子

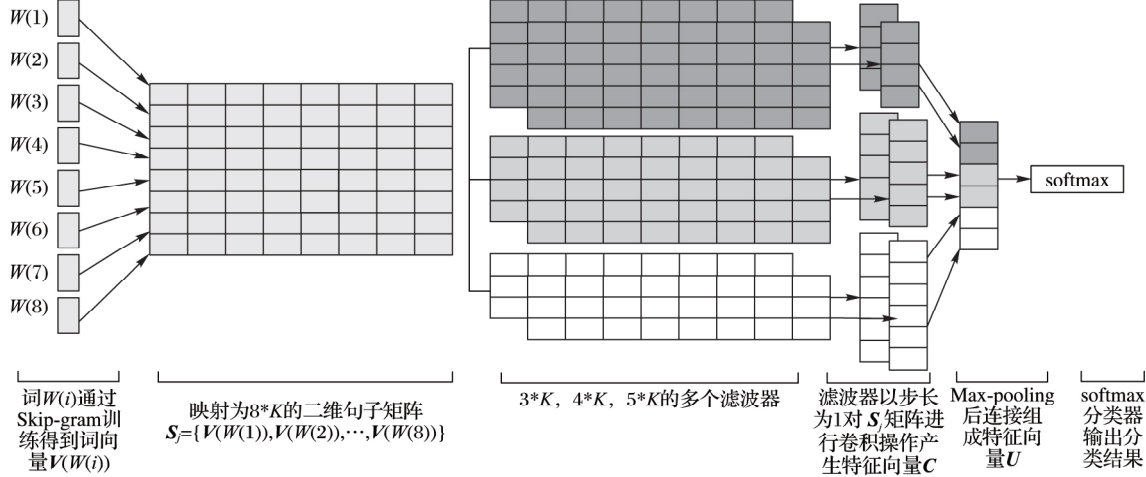


图 4 卷积神经网络模型示意图

Fig. 4 Schematic diagram of convolutional neural network model

卷积层用大小为 $r \times k$ 的滤波器对句子矩阵 S_j 执行卷积操作,提取 S_j 的局部特征;

$$c_i = f(F \cdot V(W(i:i+r-1)) + b) \quad (2)$$

其中: F 代表 $r \times k$ 的滤波器, b 代表偏置量; f 代表通过 RELU 进行非线性操作的函数; $V(W(i:i+r-1))$ 代表 S_j 中从 i 到 $i+r-1$ 共 r 行向量; c_i 代表通过卷积操作得到的局部特征。随着滤波器依靠步长为 1 从上往下进行滑动,走过整个 S_j ,最终得到局部特征向量集合 C :

$$C = \{c_1, c_2, \dots, c_{r-h+1}\} \quad (3)$$

对卷积操作得到的局部特征采用最大池化的方法提取值最大的特征代替整个局部特征,通过池化操作可以大幅降低特征向量的大小;

$$d_i = \max C \quad (4)$$

最后将所有池化后得到的特征在全连接层进行组合输出向量 U :

$$U = \{d_1, d_2, \dots, d_n\} \quad (5)$$

最后将全连接层输出的 U 输入 softmax 分类器中进行分类,模型利用实际分类中的标签,通过反向传播算法进行参数优化。

$$P(y|U, W, b) = \text{softmax}(F \cdot U + b) \quad (6)$$

2.2 BiLSTM 模型

由于 RNN 能学习任意时间长度序列的输入,但随着输入的增多,难以学习到连接之间的关系,产生长依赖问题,即对前面时间的一些节点的感知力下降,进而会发生梯度消失或者梯度爆炸现象。LSTM 能解决 RNN 以上的问题,其核心是利用记忆细胞记住长期的历史信息,并用门机制进行管理,门结构不提供信息,只是用来限制信息量,加入门其实是一种多层次的特征选择方式。如图 5 所示,门机制中各个门和记忆细胞的表达式如下:

LSTM 遗忘门表达式:

映射为句子矩阵 S_j 。

如图 4 所示,其中 $V(W(i)) \in \mathbf{R}^k$, 代表句子矩阵 S_j 中第 i 个词向量为 K 维词向量, $S_j \in \mathbf{R}^{m \times k}$, m 代表句子矩阵 S_j 中句子的个数,句子矩阵 S_j 作为卷积神经网络语言模型的嵌入层的向量矩阵。其中将句子矩阵表示为 $S_j = \{V(W(1)), V(W(2)), \dots, V(W(m))\}$ 。

$$Z^f = \text{sigmoid}(W_f \cdot [V(W(i)) \parallel h^{t-1}] + b_f) \quad (7)$$

LSTM 输入门表达式:

$$Z^i = \text{sigmoid}(W_i \cdot [V(W(i)) \parallel h^{t-1}] + b_i) \quad (8)$$

$$Z = \tanh(W_c \cdot [V(W(i)) \parallel h^{t-1}] + b_c) \quad (9)$$

LSTM 细胞更新表达式:

$$C^t = Z^f * C^{t-1} + Z^i * Z \quad (10)$$

$$Z^o = \text{sigmoid}(W_o \cdot [V(W(i)) \parallel h^{t-1}] + b_o) \quad (11)$$

LSTM 最后的输出表达式:

$$h^t = Z^o * \tanh(C^t) \quad (12)$$

其中: Z^f, Z^i, Z, Z^o 分别代表遗忘门、输入门、当前输入单元状态和输出门; h^{t-1}, h^t 分别代表前层隐层状态和当前隐层状态; W_f, W_i, W_c, W_o 分别代表遗忘门的权重矩阵、输入门的权重矩阵、当前输入单元转态权重矩阵和输出门的权重矩阵; b_f, b_i, b_c, b_o 分别代表遗忘门偏置项、输入门偏置项、当前输入单元偏置项和输出门偏置项。

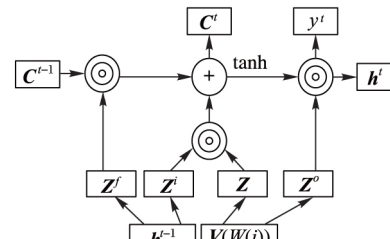


图 5 LSTM 单元结构图

Fig. 5 Structure diagram of LSTM cell

图 6 所示为本文所用的 BiLSTM 模型。虽然 LSTM 解决了 RNN 会发生梯度消失或者梯度爆炸的问题,但是 LSTM 只能学习当前词之前的信息,不能利用当前词之后的信息,由于一个词的语义不仅与之前的历史信息有关,还与当前词之后的信息也有着密切关系,所以本文利用 BiLSTM 代替 LSTM,既解决了梯度消失或者梯度爆炸的问题,又能充分考虑当前

词的上下文语义信息。利用 BiLSTM 对句子矩阵 $S_j = \{V(W(1)), V(W(2)), \dots, V(W(m))\}$ 学习,得到的文本特征具有全局性,充分考虑了词在文本中的上下文信息。

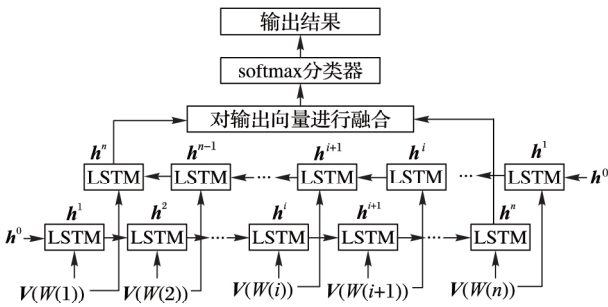


图6 BiLSTM模型

Fig. 6 BiLSTM model

2.3 本文特征融合模型

如图7所示,本文的特征融合模型由卷积神经网络和双向长短期记忆网络(BiLSTM)融合组成。卷积神经网络部分第一层是词嵌入层,将词嵌入层的句子矩阵作为输入,矩阵的列是词向量的维度,矩阵的行为 *sequence_length*;第二层是卷积层,进行卷积操作,提取局部特征,文献[15]对基准卷积神经网络的文本分类参数进行分析,当词向量100维时,滤波器为 3×100 , 4×100 , 5×100 会取得较好的分类效果,所以本文分别选用 3×100 , 4×100 , 5×100 大小滤波器各128个,步长 *stride* 大小设置为1, *padding* 为 VALID,进行卷积运算,通过卷积操作来提取句子的局部特征;第三层进行最大池化操作,提取关键特征,舍弃冗余特征,生成固定维度的特征向量,将三个池化操作输出的特征拼接起来,作为第一层全连接层输入特征的一部分。

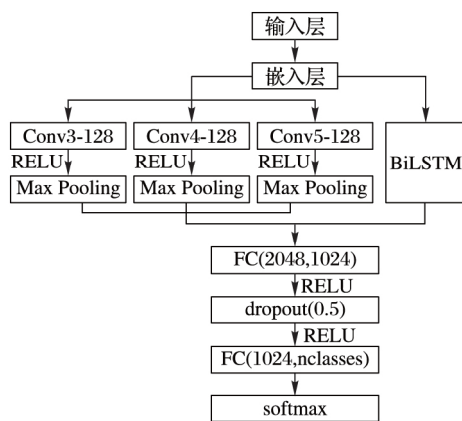


图7 CNN与BiLSTM特征融合模型

Fig. 7 Feature fusion model of CNN and BiLSTM

BiLSTM 部分第一层是词嵌入层,将嵌入层的句子矩阵作为输入,每一个词向量维度设置为100维;第二层、第三层均为隐藏层,隐藏层大小均为128,当前输入与前后序列都相关,将输入序列分别从两个方向输入模型,经过隐含层保存两个方向的历史信息和未来信息,最后将两个隐层输出部分拼接,得到最后BiLSTM的输出,代码如下:

```
output_blstm = rnn.static_bidirectional_rnn(fw_bw_inputs)
```

利用 BiLSTM 模型提取词的上下文语义信息,提取文本中词的全局特征。本文在第一个全连接层(Fully Connected layers, FC)前,使用 tensorflow 框架中的 `concat()` 方法对 CNN 和 BiLSTM 输出的特征进行融合,融合代码如下:

```
output = tf.concat([output_cnn, output_blstm], axis = 1)
```

将融合后的特征保存在 `output` 中,作为第一个全连接层的输入,在第一个全连接层与第二个全连接层之间引入 dropout 机制,每次迭代放弃部分训练好的参数,使权值更新不再依赖部分固有特征,防止过拟合,最后输入到 softmax 分类器输出分类结果,本文 softmax 回归中将 x 分类为类别 j 的概率为:

$$P(y^{(i)} = j | x^{(i)}; \theta) = \frac{\exp(\theta_j^T x^{(i)})}{\sum_{l=1}^k \exp(\theta_l^T x^{(i)})} \quad (13)$$

3 实验与分析

3.1 实验环境

本文实验环境如下:操作系统为 Ubuntu16.04, CPU 为 Intel Core i5-7500, GPU 为 GeForce GTX 1050Ti, 显卡驱动为 NVIDIA-SMI 384.111, 内存大小为 DDR3 8 GB, 开发环境为 Tensorflow 1.2.1, 开发工具使用的是 PyCharm。

3.2 实验数据

本文数据集分为两个部分:第一部分实验数据为搜狗实验室的全网新闻数据(SogouCA),来自多家新闻站点近20个栏目的分类新闻数据, SogouCA 数据为2012年6月—7月期间国内、国际、体育、社会、娱乐等18个频道的新闻数据,提供 URL 和正文信息,经过预处理后,数据大小为2.3 GB左右,将其利用 Skip-gram 模型训练得到100维的词向量;第二部分数据集是如家酒店的用户评论数据集,正负样本各3000条。其中90%用作训练集,10%用作测试集。由于 SogouCA 语料库规模足够大能训练出高质量的词向量,所以本文利用训练 SogouCA 得到的100维词向量去初始化第二部分的数据集,若第二部分数据集中的词在 SogouCA 数据中就将当前词利用 SogouCA 训练得到的词向量表示,若没有则采用随机初始化方式表示。

本文使用 word Embedding 将文本转化为低维、稠密的词向量。图8是利用 T-SNE (T-distributed Stochastic Neighbor Embedding) 对如家酒店的用户评论数据集中1000个词在低维空间中的可视化结果,低维空间可视化效果更直观,相关性越大的数据距离越近,越会聚集在一块。

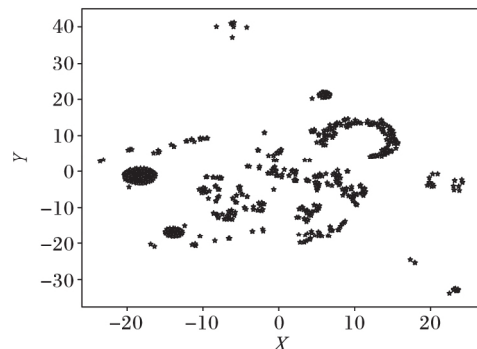


图8 词向量空间可视化

Fig. 8 Spatial visualization of word vectors

3.3 实验参数

实验参数的选取直接影响最后的实验结果,表1列出了本文融合模型的卷积部分和单 CNN 中的参数与对应的参数值。表2列出了本文融合模型的双向长短期记忆网络部分和

单 BiLSTM 中的参数与对应参数值。

通过固定参数的方法,分别比较 100 维、200 维的词向量,滑动窗口大小比较 3、4、5、7,滑动窗口数量分别取 40、80、128 进行比较,dropout 的比例对比了 0.3、0.5、0.6,L2 正则项 λ 比较了 3、5、7 对实验结果的影响,通过对比如上参数对模型准确率的影响,当取表 1 参数值时 CNN 模型取得了较好的分类效果。

表 1 卷积神经网络参数

Tab. 1 Parameters of convolution neural network

参数	值	参数	参数值
词向量维度	100	Pooling 方法	Max
滑动窗口大小	3 4 5	dropout rate	0.5
滑动窗口数量	128	L2	3
激活函数	RELU	Epoch	60

BiLSTM 的参数比较了 100 维、200 维的词向量,层数默认取 2 层,隐藏层大小对比了 128 和 256,最终发现词向量 100 维、隐藏层大小为 128 时模型分类准确率最高。Adam 通过计算梯度的一阶矩估计和二阶矩估计而为不同的参数设计独立的自适应性学习率,不断迭代更新网络参数,加快模型收敛,所以本文选取 Adam 作为优化函数。

3.4 实验结果及分析

为验证本文提出的 CNN 和 BiLSTM 特征融合模型的性能,分别将本文特征融合模型与单 CNN 模型、单 BiLSTM 模型、传统机器学习 SVM 及其他深度学习模型进行对比实验。

本文首先在第二部分数据集上将本文特征融合模型与单 CNN 模型和单模型 BiLSTM 进行了对比,且特征融合模型中 CNN 和 BiLSTM 的参数与单 CNN 模型、单 BiLSTM 模型中的参数相同,均为表 1、表 2 中的参数值,且学习率均设置为 0.001。

表 2 BiLSTM 模型参数

Tab. 2 Parameters of BiLSTM model

参数	值	参数	值
词向量维度	100	损失函数	交叉熵
层数	2	优化函数	Adam
学习率	0.001	Epoch	60
隐藏层大小	128		

图 9~10 分别给出了单 CNN 模型、单 BiLSTM 模型和本文模型的准确率和损失函数变化图。

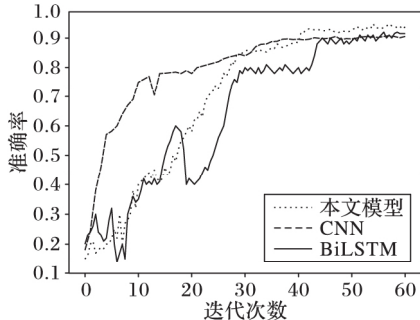


图 9 3 种模型准确率比较

Fig. 9 Comparison of accuracy of three models

由图 9 对比发现,融合模型在测试集上的收敛速度慢但

准确率均高于单 CNN、单 BiLSTM 模型。对比图 10 发现,单 CNN、单 BiLSTM 模型的 loss 值下降到稳定值的速度比融合模型下降到稳定值的速度快,但最终 loss 值都下降到了一个很低的稳定值,模型都取得较好收敛效果。

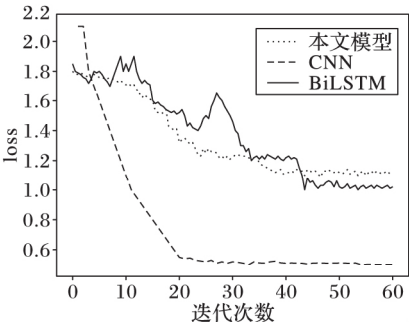


图 10 3 种模型损失比较

Fig. 10 Comparison of loss of three models

如表 3 所示,本文所提融合模型分类准确率比单 CNN^[15]模型分类准确率上提升了 4.27%,比单 BiLSTM 模型^[16]分类准确率提升 3.31%。利用 CNN 提取文本局部特征,BiLSTM 模型提取文本全局特征,将两种互补模型进行特征融合,取得了比单模型都好的分类准确率。

表 3 单模型和融合模型结果对比

Tab. 3 Comparison of single model and fusion model

模型	准确率/%	模型	准确率/%
CNN 模型 ^[15]	90.56	本文融合模型	94.83
BiLSTM 模型 ^[16]	91.52		

由于 Sogou 数据量过于庞大,本文从 Sogou 数据集中只选取了汽车和体育两个类别,且从每个类别随机抽取全部数据的一部分进行实验,本文利用 Sogou 新闻数据集中的汽车和体育两类数据集各 3 000 条,其中 90% 为训练集,10% 为测试集。

本文不仅与单模型进行了对比,还与传统的机器学习算法 SVM 及其他的深度学习模型进行对比,实验结果如表 4 所示。文献[15]将文本特征提取、文本特征表示和归一化后,输入到 SVM 中学习分类决策函数,基于结构风险最小化,转化为二次型寻优问题,求最优值,虽然取得了较好的分类效果,但本文融合模型分类效果明显优于 SVM;文献[18]中提出的 Attention Based LSTM 模型,通过引入 Attention model 计算历史节点对当前节点的影响权重,有效解决了信息丢失和信息冗余等长期依赖问题,探究了文本上下文对文本分类的影响,提高了文本分类的准确性,但本文所提融合模型在分类准确率上取得了更好分类结果;文献[19]中 Bag of Words 文本分类忽略了词序、语法和句法信息,每个词相互独立,而文本模型充分利用了卷积神经网络模型提取局部特征和 BiLSTM 兼顾文本序列全局特征取得了优于 BOW 文本分类的效果,并与文献[20]中所提 C-LSTM (CNN-LSTM) 的 CNN 和 LSTM 统一模型相比较,C-LSTM 首先通过使用 CNN 提取文本特征,然后将特征编码输入 LSTM 进行分类,发现本文所提融合模型有效提高了文本分类的准确率。

表4 融合模型与其他模型结果对比

Tab. 4 Result comparison of fusion model and other models

模型	准确率/%	模型	准确率/%
SVM ^[17]	84.02	C-LSTM ^[20]	94.40
Attention Based LSTM ^[18]	92.18	本文模型	94.58
Bag of Words ^[19]	92.85		

4 结语

本文提出了一种基于卷积神经网络和 BiLSTM 网络的特征融合模型用于文本分类研究,该模型既能利用卷积神经网络有效提取文本的局部特征,又可以利用 BiLSTM 兼顾文本的全局特征,充分考虑了词的上下文语义信息。将本文所提融合模型与单 CNN 模型、单 BiLSTM 模型进行了对比实验,本文所提融合模型分类准确率优于单 CNN、单 BiLSTM 模型。此外本文还与传统机器学习模型 SVM 及其他深度学习模型进行了比较,结果表明本文所提特征融合模型在分类准确率上优于对比的模型,本文融合模型有效地提升了文本分类的准确率。然而本文融合模型并未使用深度较深的卷积神经网络,未来将研究深度较深的卷积神经网络融合 BiLSTM 模型对文本分类准确率的影响。

参考文献 (References)

- [1] LI J, CAO Y, WANG Y, et al. Online learning algorithms for double-weighted least squares twin bounded support vector machines [J]. Neural Processing Letters, 2017, 45(1): 319–339.
- [2] 孙志军, 薛磊, 许阳明, 等. 深度学习研究综述[J]. 计算机应用研究, 2012, 29(8): 2806–2810. (SUN Z J, XUE L, XU Y M, et al. A review of deep learning research [J]. Application Research of Computers, 2012, 29(8): 2806–2810.)
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc., 2012: 1097–1105.
- [4] JOHNSON R, ZHANG T. Effective use of word order for text categorization with convolutional neural networks [J/OL]. arXiv Preprint, 2014, 2014: arXiv:1412.1058 (2014-12-01) [2015-03-26]. <https://arxiv.org/abs/1412.1058>.
- [5] KALCHBRENNER N, GREFFENSTETTE E, BLUNSON P. A convolutional neural network for modelling sentences [J/OL]. arXiv Preprint, 2014, 2014: arXiv:1404.2188 [2014-04-08]. <https://arxiv.org/abs/1404.2188>.
- [6] KIM Y. Convolutional neural networks for sentence classification [EB/OL]. [2014-09-03]. <http://emnlp2014.org/papers/pdf/EMNLP2014181.pdf>.
- [7] CHO K, van MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder decoder for statistical machine translation [EB/OL]. [2014-09-03]. <http://www.aclweb.org/anthology/D/D14/D14-1179.pdf>.
- [8] EBRAHIMI J, DOU D. Chain based RNN for relation classification [EB/OL]. [2014-09-03]. <http://www.aclweb.org/anthology/N/N15/N15-1133.pdf>.
- [9] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735–1780.
- [10] GRAVES A. Supervised Sequence Labelling with Recurrent Neural Networks [M]. Heidelberg: Springer, 2012: 35–42.
- [11] ANDRIY M, GEOFFREY H. A scalable hierarchical distributed language model [C]// Proceedings of the 21st International Conference on Neural Information Processing System. Lake Tahoe, Nevada: Curran Associates Inc., 2008: 1081–1088.
- [12] BENGIO Y, SCHWENK H, SENÉCAL J S, et al. A neural probabilistic language model [J]. Journal of Machine Learning Research, 2003, 3(6): 1137–1155.
- [13] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [J/OL]. arXiv Preprint, 2013, 2013: arXiv:1301.3781 (2013-01-16) [2013-09-07]. <http://arxiv.org/abs/1301.3781>.
- [14] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]// Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc., 2013: 3111–3119.
- [15] 冀胜利, 李波. 基于 SVM 的中文文本分类算法 [J]. 重庆工学院学报(自然科学版), 2008, 22(7): 84–87. (YI S L, LI B. Chinese text classification algorithm based on SVM [J]. Journal of Chongqing Institute of Technology (Natural Science Edition), 2008, 22(7): 84–87.)
- [16] 王树恒, 吐尔根江·依布拉音, 卡哈尔江·阿比的热西提, 等. 基于 BiLSTM 的维吾尔语文本情感分析 [J]. 计算机工程与设计, 2017, 38(10): 2879–2886. (WANG S H, IBRAHIM T, ABIDEREXITI K, et al. Sentiment classification of Uyghur text based on BiLSTM [J]. Computer Engineering and Design, 2017, 38(10): 2879–2886.)
- [17] 崔建明, 刘建明, 廖周宇. 基于 SVM 算法的文本分类技术研究 [J]. 计算机仿真, 2013, 30(2): 299–302. (CUI J M, LIU J M, LIAO Z Y. Research of text categorization based on support vector machine [J]. Computer Simulation, 2013, 30(2): 299–302.)
- [18] 张冲. 基于 Attention-Based LSTM 模型的文本分类技术的研究 [D]. 南京: 南京大学, 2016. (ZHANG C. Text classification based on Attention-Based LSTM model [D]. Nanjing: Nanjing University, 2016.)
- [19] ZHANG X, ZHAO J, LeCUN Y. Character-level convolutional networks for text classification [C]// Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2015: 649–657.
- [20] ZHOU C, SUN C, LIU Z, et al. A C-LSTM neural network for text classification [J/OL]. arXiv Preprint, 2015, 2015: arXiv. 1511.08630 (2015-11-27) [2018-04-01]. <https://arxiv.org/abs/1511.08630>.

This work is partially supported by the National Natural Science Foundation of China (61472095).

LI Yang, born in 1991, M. S. candidate. His research interests include machine learning, natural language processing.

DONG Hongbin, born in 1963. Ph. D., professor. His research interests include evolutionary computing, computational intelligence, data mining, multi-Agent system, machine learning.