

# 深度学习框架下微博文本情感细粒度研究<sup>①</sup>



王 儒, 王嘉梅, 王伟全, 符 飞

(云南民族大学 电气信息工程学院, 昆明 650504)

(云南民族大学 云南省高校少数民族语言文字信息化处理工程研究中心, 昆明 650504)

通讯作者: 王嘉梅, E-mail: [wangj\\_2004@163.com](mailto:wangj_2004@163.com)

**摘 要:** 情感细粒度分析是情感分析的分支, 随着社交网络规模的扩大, 简单的划分积极或消极的粗粒度情感分析不能满足实际应用的需要, 基于评价对象及其属性的细粒度情感分析得到了重视. 近几年深度学习在自然语言处理领域的成功应用给情感细粒度分析提供了新的思路. 以 NLPCC2013 任务二微博数据集为研究对象, 探究微博短文本在不同神经网络结构中的情感细粒度分类结果并加入词向量进行优化, 最后分析与总结了神经网络微博短文本细粒度情感分析的影响因素及发展方向.

**关键词:** 细粒度情感分析; 神经网络; 微博短文本

引用格式: 王儒, 王嘉梅, 王伟全, 符飞. 深度学习框架下微博文本情感细粒度研究. 计算机系统应用, 2020, 29(5): 19-28. <http://www.c-s-a.org.cn/1003-3254/7371.html>

## Fine-Grained Analysis and Research of Emotion in Microtext under Framework of Deep Learning

WANG Ru, WANG Jia-Mei, WANG Wei-Quan, FU Fei

(School of Electric and Informative Engineering, Yunnan Minzu University, Kunming 650504, China)

(Engineering Research Center of Yunnan Higher Education for Information Processing of Minority Languages and Characters, Yunnan Minzu University, Kunming 650504, China)

**Abstract:** Fine-grained analysis of emotion is a branch of sentiment analysis, with the expansion of social network, the division of simple positive or negative coarse-grained sentiment analysis cannot satisfy the need of practical application. Thus the fine-grained emotional analysis based on evaluation objects and their attributes has received attention in recent years. The successful application of deep learning in the field of natural language processing in recent years provides a new idea for the fine-grained analysis of emotion. Take NLPCC2013 task 2 Weibo data set as the research object, explore the classification results of microtext in different neural network structures and add word vectors for optimization. Finally, the influencing factors and development direction of finer-grained emotion analysis of neural network micro-blog essay are analyzed and summarized.

**Key words:** fine-grained analysis of emotions; neural network; microtext

据第 43 次《中国互联网络发展状况统计报告》<sup>[1]</sup>, 中国网民规模达 8.29 亿, 2018 年网页长度 19 061 579 332 918 字节数, 相比 2017 年增长 11.4%, 对规模越发庞大的网络文本内容进行有效的挖掘处理, 对促进网络的发展和网络环境的净化有着重要意义. 随着网络

的发展, 形如微博、邮件、短信、弹幕、新闻评论、微信、聊天记录의 短文本内容逐渐增多, 网络成为人们表达观点与抒发情感的重要平台, 与传统通过书信报刊形式表达观点有所不同, 上述形式的短文本有以下几个特点:

① 收稿时间: 2019-09-18; 修改时间: 2019-10-15; 采用时间: 2019-10-30; csa 在线出版时间: 2020-05-07

- 1) 文本内容长度一般都较短.
- 2) 内容形式不局限于纯文本, 也会结合表情图像.
- 3) 持有观点的群体和观点的内容具有多样性.
- 4) 表达的观点或内容具有一定的倾向性和重复性.
- 5) 一般都缺少上下文信息.

本文以 NLPCC2013 任务二微博数据集为研究对象, 探究微博短文本在不同神经网络模型中的情感细粒度分类结果, 论文其余部分是这样组织的, 第 1 节介绍了情感细粒度分析, 第 2 节介绍词向量及论文实验所用到的几个不同模型及结构, 第 3 节介绍实验所用数据、环境、参数及评价标准, 第 4 节对实验结果进行了分析与评价, 最后对未来的工作进行了总结和展望.

## 1 情感细粒度分析

情感细粒度分析是情感分析的分支, 从文本颗粒度层次上来说, 情感分析可以分为篇章级、句子级、属性级 3 个层次<sup>[2]</sup>, 按能否获取观点评价对象的不同可分为粗粒度情感分析(篇章级、句子级)与细粒度情感分析(属性级), 篇章级和句子级的情感分析只能得到整篇文档或整条句子的情感倾向, 无法确切得知观点评价对象的喜好, 属性级的情感分析基于评价对象及其属性上的观点信息, 可以得到观点评价对象对某一事物的具体情感. 传统的情感分析方法往往是将文本划分为积极和消极的二分类, 或加入中性的三分类, 随着社交网络规模的扩大, 简单的二分类或三分类无法囊括丰富的文本情感, 如“我早上很难过, 一个人郁郁寡欢的, 小王可能察觉到了我的异样, 跑来带我出去逛街玩了好久, 有人关心的感觉真好啊!”. 这句话前半部分表达了今天心情的颓丧, 后半部分表达了有人关心心情因此得到缓解而愉悦的转变, 这句话无法用简单的二分类或三分类进行囊括, 也因此细粒度情感分析在实际应用中有很强需求<sup>[3,4]</sup>. 与其他文本相比, 微博文本具有篇幅较短(一般在 140 字以内), 噪声大, 上下文的语义信息稀疏, 符号化(表情、图像)和口语化(缩写)严重等特点<sup>[5,6]</sup>, 个性化的表达非常适合进行细粒度情感分析的研究.

张云秋等<sup>[7]</sup>利用情感词典的方法在大连理工大学情感词汇本体 DUTIR<sup>[8]</sup>的 7 大类情感词汇基础上增加了表示“疑”这一类情感的词汇, 共 52 个词, 根据情感

强弱不同分为 4 个等级, 并在自行构造的药物微博评论数据集上使用 PMI 方法进行了“乐、好、怒、哀、惧、恶、惊、疑”8 种情感的细粒度分析, 这些评价结果基于当前情感类别, 没有给出整体的宏平均分析结果.

刘丽<sup>[9]</sup>利用条件随机场 (CRF) 结合语法树剪枝的方法对产品评论进行了细粒度情感分析, 借助 Map-Reduce 中并行化协同训练的方法对语料进行了半自动标注, 并进行了可视化表示, 在识别情感要素和情感评价单元两个领域均取得了 89% 左右的综合准确率.

李阳辉等<sup>[10]</sup>利用深度学习中降噪自编码器的方法在多个数据集上进行了实验, 对文本进行了句子级的情绪二分类(消极、积极), 没有更进一步对文本的情感细粒度进行分析.

张谦等<sup>[11]</sup>利用深度学习中词向量模型结合传统机器学习 TF-IDF 模型进行词汇加权对 15 种微博主题文本进行了主题分类, 证明了合并后的模型结果比单一的词向量加权模型和 TF-IDF 模型分类结果更好.

还有许多工作<sup>[12-15]</sup>结合深度学习的知识对文本进行情感分析以后取得了很好的效果, 这里不一一列举.

## 2 词向量与神经网络模型

### 2.1 文本-数值转化

自然文本输入计算机以后需要进行文本-数值的转化, 传统方法是通过向量空间模型 (Vector Space Model, VSM)<sup>[16]</sup>进行文本-数值之间的转换将文本转换成向量以空间上的相似度表示文本间的相似度, 通过计算空间相似度来表示文本相似度计算简单方便, 但存在文本稀疏、一词多义、易忽略文本上下文信息等问题难以解决. 机器学习中传统的向量空间模型如 TF-IDF 模型、One-Hot 向量, 都是通过将文本中的词频信息转化为向量衡量该词语所占的权重.

Word2Vec<sup>[17,18]</sup>自 2013 年提出以来迅速取代向量空间模型成为自然语言处理领域新的基础, 与 VSM 相比, Word2Vec 是一种分布式的神经概率语言模型<sup>[19]</sup>(图 1), Word2Vec 层数并不深, 是一种浅而双层的神经网络, 词向量的每一维都具有一定的语义和语法特征, 能够捕捉更多有关上下文的信息, 同 VSM 一样, 词向量也很难解决一词多义的问题, 这是由于训练出的词向量矩阵是静态固定不变的. 词向量模型有两种,

CBOW 模型与 Skip-gram 模型. 其中 CBOW 模型利用当前词上下文的若干个词语预测当前词的概率, 如图 1 通过当前词前后 2 个词预测当前词语, 而 Skip-gram 模型则相反, 通过当前词的概率预测上下文的若干个词语.

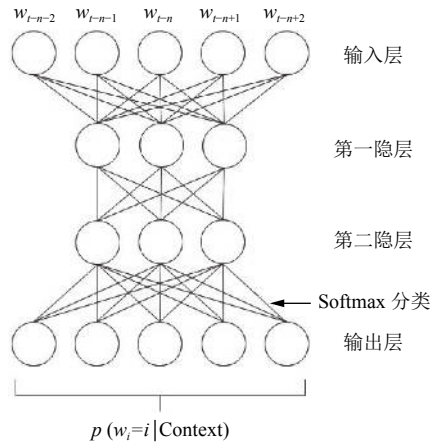


图 1 神经网络语言模型

图 1 中  $(w_{t-n-2}, \dots, w_{t-n+2})$  可看做按时间序列排序的词语集, 通过神经网络结合上下文信息 (context) 计算语言模型中的参数来预测当前词, 依据窗口大小为 2 的上下文预测当前词, 最后由 Softmax 层输出预测的当前词. 原理和 2-gram 语言模型相似.

## 2.2 CNN 和 TextCNN 模型

CNN (Convolutional Neural Network) 也称卷积神经网络, 1962 年生物学家 Hubel 和 Wiesel<sup>[20]</sup>发现猫的视觉神经细胞中存在一种被称为感受野的细胞结构, 这种细胞结构能作用于视觉输入空间, 对视觉形成的图像局部特征的提取能力很强, 1980 年, Fukushima<sup>[21]</sup>根据上述结构提出了神经感知机, 被认为是卷积神经网络的雏形, 到 1998 年 LeCun<sup>[22]</sup>提出 LeNet-5, 宣告了现代的卷积神经网络结构正式出现. 卷积神经网络由输入层、卷积层、池化层、全连接层、输出层等 5 部分构成, 典型的卷积神经网络 LeNet-5 结构如图 2.

CNN 的卷积原理可用图 3 表示.

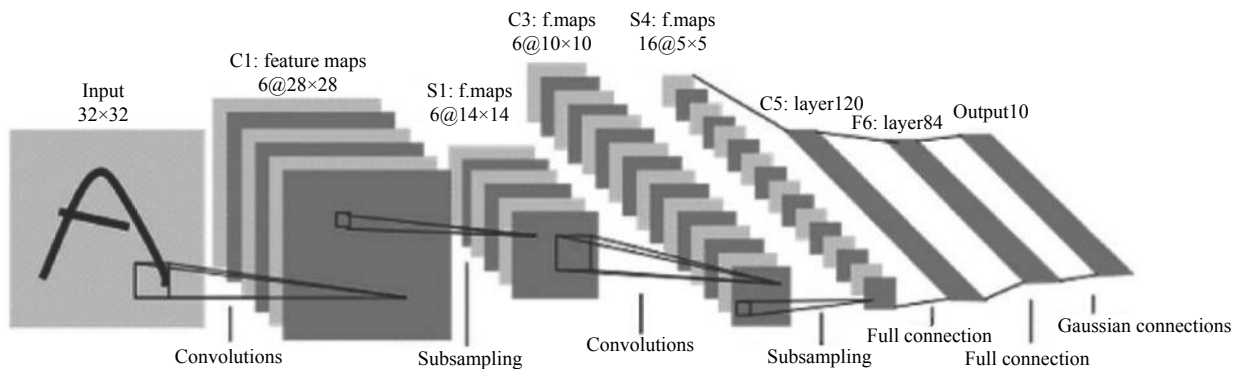


图 2 LeNet-5<sup>[22]</sup>

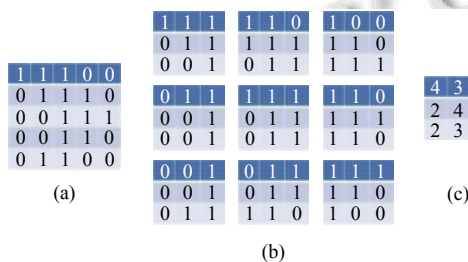


图 3 卷积核滑动原理

假定输入是一个  $5 \times 5$  的矩阵图 3(a), 通过  $3 \times 3$  窗口大小的卷积核滑动分别提取特征图如图 3(b), 图 3(c) 为卷积后得到的特征图. 此时得到的矩阵很好计算, 但如果是一幅长宽为 256 位的图片或一条长度

为 200 的句子此时得到的特征图仍然十分臃肿, 需要通过池化层进一步降维, 常用的池化方法有最大值池化和平均池化两种. Nal Kalchbrenner 等<sup>[23]</sup>提出了  $k$ -Pooling 方法, 将所有特征值得分在 Top-K 的值及这些特征值原始词语的先后顺序保留, 这种方法对一条前半句与后半句表达情感相反或有所区别的句子效果很好. 以上简单的介绍了 CNN 工作原理, 更多内容也可阅读文献<sup>[24,25]</sup>.

TextCNN<sup>[26]</sup>是 CNN 的一个变种, 由 Kim Y 在 2014 年提出, 是卷积神经网络在自然语言处理领域的首次应用, 在当时取得了 state-of-the-art 的效果, 结构



如图4. TextCNN 工作原理与 CNN 相差不大, 通过卷积核滑动提取文本中的特征, 但图像是二维数据(长、宽), 文本是一维数据, 需要通过构造不同的 Filter(卷积核) 的窗口大小形成维度上的差异完成文本上的二维表示, 常用的卷积核窗口大小为 3、4、5, 这与  $N$ -

gram 语言模型在某些方面不谋而合,  $N$ -gram 语言模型通过计算当前词前后  $N$  个词语的概率得到当前词语与上下文之间的关系. Zhang Y<sup>[27]</sup>测试了不同参数设置如词向量、激活函数、Feature Map 大小、Pooling 的方法等对 Text-CNN 模型分类效果的影响及最佳参数.

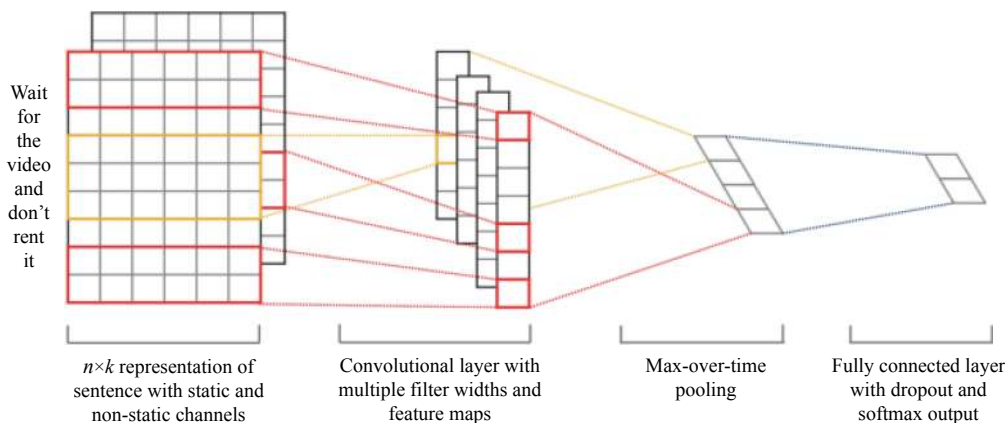


图4 TextCNN<sup>[26]</sup>

Text-CNN 在中文文本分类(词级别)中工作原理如图5.

用于处理如语音、文字、视频图像等形式的时间序列数据. 典型的 RNN 如图6.

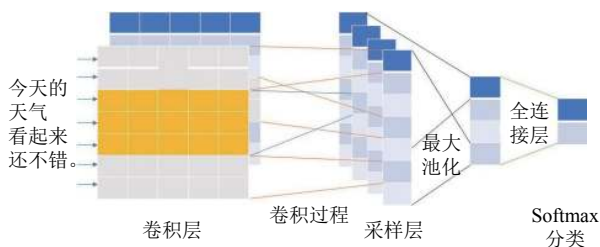


图5 Text-CNN 中文词级别分类原理

从图5可以看出 TextCNN 模型要先经过中文分词, 再按卷积神经网络的原理对文本转化的数值进行处理, 卷积神经网络相比循环神经网络更多关注局部的信息, 因而中文处理领域影响较大的停用词处理对卷积神经网络而言就非必须, 分词时可根据需要决定是否去除停用词.

### 2.3 LSTM 和 GRU 模型

RNN 也称循环神经网络 (Recurrent Neural Network) 或递归神经网络 (Recursive Neural Network), 本文中 RNN 指循环神经网络, 循环神经网络既有前馈通路, 又有反馈通路, 可以将上一时刻的信息传入下一时刻, 使模型捕捉到序列数据中有关上下文的信息, 故而常

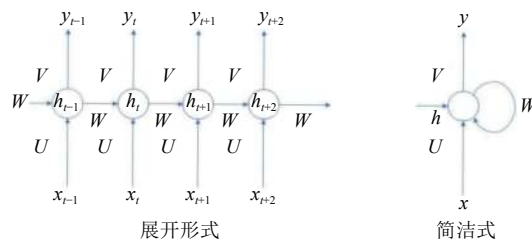


图6 RNN 结构图

RNN 公式, 如式 (1) 及式 (2):

$$h_t = \tanh(Ux_t + Ws_{t-1}) \tag{1}$$

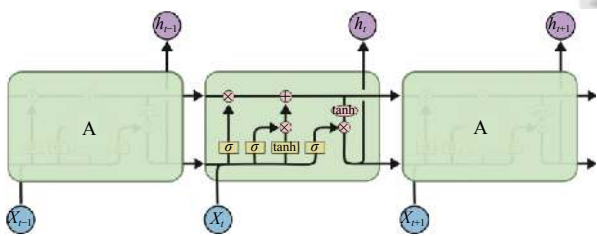
$$\hat{y}_t = \text{Softmax}(Vh_t) \tag{2}$$

式中,  $x_t$  为输入当前时间步的输入信息,  $y_t$  为当前时间步信息经隐藏层  $h_t$  后得到的输出信息, 同时也是下一时间步的输入信息,  $s_t$  是隐藏层的值,  $U$  是输入层到隐藏层的权重矩阵,  $V$  是隐藏层到输出层的权重矩阵,  $W$  则是隐藏层上一次的值作为这一次的输入的权重. RNN 训练时的梯度变化由于模型的时序性存在不能简单的通过 BP(反向传播) 算法进行, 需加入基于时间的计算过程, 也就是 BPTT 算法.

RNN 可以捕捉文本的上下文信息, 但当模型时间

步过长时,过去时间步的历史信息不停累积容易使模型出现梯度爆炸或消失的情况, LSTM<sup>[28]</sup>和 GRU<sup>[29]</sup>均是在此基础上进行的改进. RNN 有很多变种结构, Jozefowicz R 等<sup>[30]</sup>测试过一万多种 RNN 及其变种结构的效果.

LSTM (Long and Short Term Memory) 的主要结构由输入门 $i_t$ (input gate)、遗忘门 $f_t$ (forget gate)、输出门 $o_t$ (output gate)、以及细胞态 $\tilde{c}_t$ (memory cell) 组成, 为了避免出现梯度爆炸或消失的情况, LSTM 通过遗忘门先选择性的遗忘掉上一时间步的部分历史信息, 再通过细胞态对当前时间步的模型状态进行更新. 典型的 LSTM 如图 7.

图 7 LSTM 结构图<sup>[27]</sup>

LSTM 公式如式 (3)~式 (8):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (6)$$

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

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

式中,  $W$  和  $b$  分别表示相应状态下的权重系数矩阵和偏置项,  $\sigma$  和  $\tanh$  表示 Sigmoid 和双曲正切激活函数, LSTM 在训练时先按式 (3)~式 (8) 计算当前时间步各个状态值, 再通过 BPTT 算法反向计算网络的整体误差, 通过误差对模型梯度进行优化并更新网络的权重.

GRU (Gated Recurrent Unit) 在 LSTM 的基础上进一步简化了模型结构, 将 LSTM 的输入门 $i_t$ 、遗忘门 $f_t$ 、输出门 $o_t$ 更改为更新门 $z_t$ (update gate) 与重置门 $r_t$ (reset gate), 并将细胞态 $\tilde{c}_t$ 与隐层状态 $h_t$ 进行了合并为 $\tilde{h}_t$ . GRU 结构如图 8.

GRU 公式如式 (9)~式 (12):

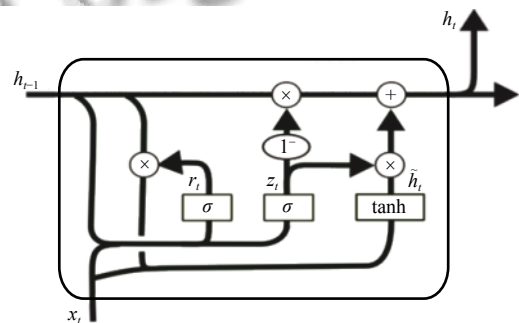
$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (9)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (10)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad (11)$$

$$h_t = (1 - z_t)h_{t-1} + z_t * \tilde{h}_t \quad (12)$$

式中,  $W$  表示相应状态下的权重矩阵,  $h_t$  表示隐藏层状态信息,  $\sigma$  和  $\tanh$  表示 Sigmoid 和双曲正切激活函数, 比较 GRU 与 LSTM 的结构会发现, GRU 的输入输出结构与 LSTM 相似, 区别在于 GRU 根据式 (12) 使用  $z_t$  即可完成对输入信息的遗忘与选择, 使用参数更少, 结构更加简单.

图 8 GRU 结构图<sup>[28]</sup>

## 2.4 双向结构

在类时间序列数据的处理中, 只有前馈输入还不够, 还需要有反馈通路以利用上一时间步的信息, 双向 (Bi-directional) 结构很好的解决了这个问题. 缺点是相比单层结构, 双向结构需要更多的计算资源, 程序运行时间也更长.

双向结构如图 9 所示.

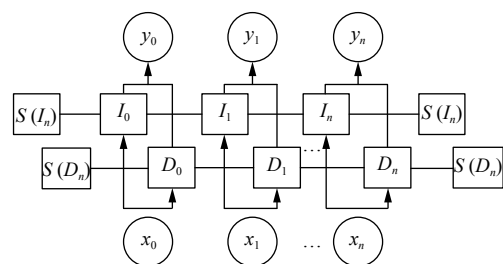


图 9 双向结构

## 2.5 CLSTM 和 CNN-LSTM 模型

LSTM 神经网络可以捕捉序列中有关上下文的信息, CNN 神经网络可以获取全文的局部特征, Ghosh S<sup>[31]</sup>结合两个模型提出了结合上下文信息的混合神经

网络模型 CLSTM(也称时空网络)在接续语句预测上相比当时其他模型提高了 20%, Zhou CT<sup>[32]</sup>等将这一模型应用到文本分类、情感分析问题上也取得了不错的效果. 孙晓等<sup>[33]</sup>将 LSTM 与 CNN 的输出结果进行直接拼接, 在讽刺语识别数据集上进行了测试, 效果比词袋模型或单一 RNN、CNN 模型要好. 本文在 CNN 和 RNN 的组合模型与拼接模型的基础上进行了改进, 加入了双向结构以利用上下文信息.

C\_BiLSTM 用双向 LSTM 连接 CNN 中的卷积层与池化层, 通过 CNN 卷积层提取序列数据中的重要特征, BiLSTM 层对提取后的特征进行处理得到序列数据中有关上下文的信息, 模型如图 10.

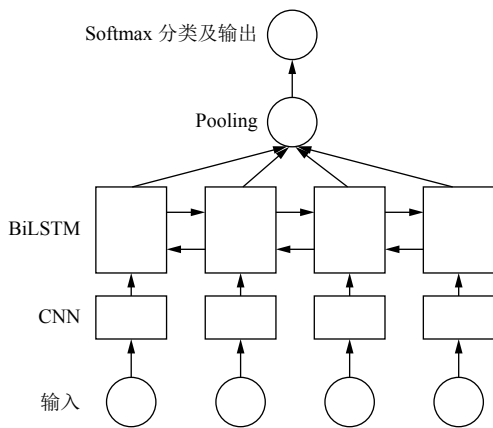


图 10 C\_BiLSTM 模型示意图

CNN-BiLSTM 模型将 CNN、BiLSTM 两个神经网络模型的输出结果经 Keras 的 Concatenate 层进行简单拼接, 最后经 Softmax 进行分类, 模型如图 11.

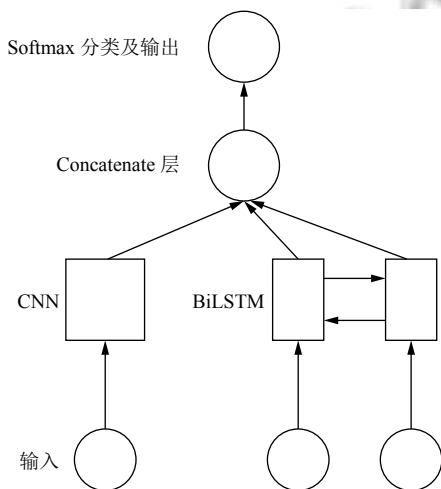


图 11 CNN-BiLSTM 模型示意图

### 3 实验准备工作

#### 3.1 实验数据

本文所用数据为 NLPCC2013 任务二微博数据集, 数据已经标注. 如图 12, 共 8 类情感标签 none, sadness, like, anger, happiness, disgust, fear, surprise 其中各类情感标签数目分别为 6753, 1144, 2120, 640, 1476, 1394, 139, 334 共 14 000 条, 包含 4000 条训练集与 10 000 条测试集, 在实验过程中我们将两个数据集合并成一个数据集, 借助 Python 的机器学习包 SKlearn 中的数据处理函数 Train\_Test\_Split 按训练集与测试集 8:2 的比例进行重新划分.

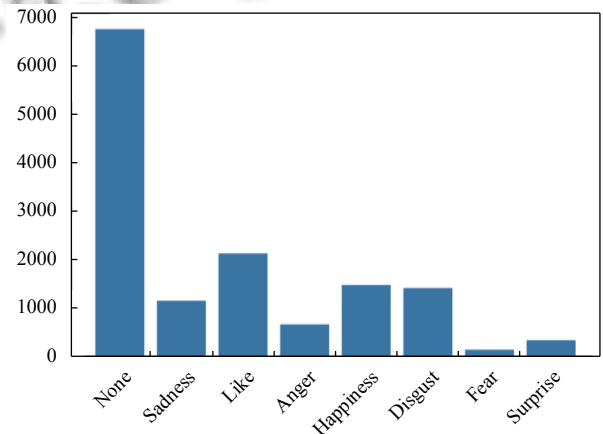


图 12 微博情感标签分布图

#### 3.2 实验环境

本文实验环境配置见表 1.

表 1 实验环境配置

实验环境	实验配置
硬件	小米笔记本电脑 Pro 15.6''64 位
操作系统	Windows 10 家庭中文版 17134.885
内存	16 GB (15.9 GB 可用)
CPU	Intel(R)Core(TM)i7-8550U CPU@1.8 GHZ 2.0 GHZ
编程语言	Python 3.7.3
深度学习框架	Keras

#### 3.3 词向量

词向量是深度学习用于自然语言处理的关键一环, 论文采用了两种词向量的训练方式, 一种是通过 Keras 中 Embedding 层的动态训练生成的词向量, 一种是预训练词向量模型.

关于预训练词向量, 原始数据中包含 5 万条原始微博语料, 但这部分语料规模较小, 训练出的词向量规

模也较小,本文选用了北京师范大学中文信息处理研究所与中国人民大学 DBIIR 实验室合作的工作<sup>[34]</sup>,这份工作的内容包括一百多种中文领域的词向量,选择其中的微博词向量 531 MB (557 675 674 字节, 300 维) 作为本文所用预训练词向量. 从这份词向量中随机抽取了部分词语,如图 13,可以看到不同词语间的语义相关性.

### 3.4 神经网络参数

神经网络模型中参数对模型最终的结果意义极大,表 2 为论文所用模型参数,分别为输入长度、层数、词向量维度、神经元个数、卷积核、dropout、optimizer、epoch,为了便于比较论文部分模型设置了相同的参数.

表 2 中 dropout 一行括号内参数为循环时上一时间步到当前时间步循环 dropout 的比率,在统计层数时,去掉了模型的输入层与输出层.

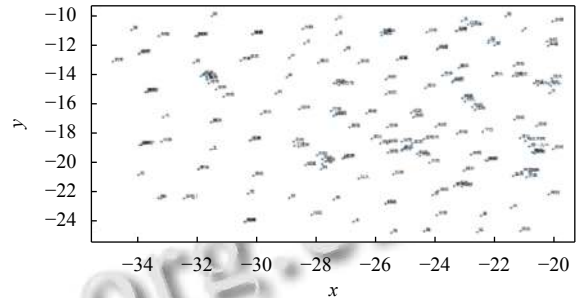


图 13 微博词向量可视化

表 2 神经网络参数

参数	CNN	TextCNN	LSTM	GRU	BiLSTM	BiGRU	C_Bi LSTM	CNN-BiLSTM
输入长度	200	200	200	200	200	200	200	200
层数	3	1	1	1	2	2	3	1
词向量维度	300	300	300	300	300	300	300	300
神经元	128	256	128	128	128	128	128	128
卷积核	3 4 5	3 4 5					3	3
Dropout	0.2	0.2	0.2(0.1)	0.2(0.1)	0.2(0.1)	0.2(0.1)	0.2(0.1)	0.2(0.1)
optimizer	adam	adam	rmsprop	rmsprop	adam	adam	adam	adam
epoch	10	10	10	10	10	10	10	10

### 3.5 实验评价指标

在进行实验结果分析之前,需要明确判别模型性能的标准,本文采用两种评价指标进行评价.一种是机器学习中常用的准确率,  $F1$  值,一种是混淆矩阵,准确率和  $F1$  值可以得到整个模型的整体分类效果,但不能得到某一情感类别文本的细粒度分类效果,混淆矩阵可以比较文本预测类别与文本真实类别的差异看出某一模型具体的细粒度分类效果.

准确率的定义是以关注的类为正类,其他类为负类,分类器在测试数据集上的预测或正确或不正确,4 种情况出现的总数分别记作:

$TP$ : 将正类预测为正类数.

$FN$ : 将正类预测为负类数.

$FP$ : 将负类预测为正类数.

$TN$ : 将负类预测为负类数

准确率的定义为:

$$P = \frac{TP}{TP+FP} \quad (13)$$

$F1$  值的定义为:

$$F1 = \frac{2TP}{2TP+FP+FN} \quad (14)$$

混淆矩阵的定义是矩阵的每一列代表文本的预测情感标签,每一列的总数表示预测为该类别情感标签的文本的数目.每一行代表数据的真实归属情感标签,每一行的数据总数表示该类别情感标签的文本实际的数目.每一列中的数值表示此实际情感标签文本被预测为该类别情感标签的数目.

## 4 实验分析

由于 dropout 机制的存在,每次训练结果会出现一定的偏差,本文比较了 8 种模型在 10 次训练中整体准确率效果最好的一次结果,还比较了加入词向量优化后的模型训练 10 次训练中整体准确率最好的一次结果,如表 3,取小数点后 4 位.

通过表 3 可以看出,在最好结果的对比上, CNN 模型比 RNN 模型整体准确率更好,加入双向结构以



后 LSTM、GRU 整体准确率变低, CNN-BiLSTM 模型的拼接模型整体准确率比使用双向结构的 RNN 模型更好, C\_BiLSTM 模型整体准确率最低, 整体准确率最佳的 LSTM 模型 53.71% 与最低的 BiGRU 模型整体准确率 50.36% 相差 3.35%。同加入词向量以后的准确率作比较发现 CNN 和 LSTM 模型整体准确率分别下降 4.89%、0.85%, 其余 6 个模型的最佳整体准确率保持不变或有所提高, 加入词向量以后最佳整体准确率也由 LSTM 模型的 53.71% 提高到 CNN-BiLSTM 模型的 55.07%, 说明加入词向量对提升整体准确率有一定效果, 值得注意的是加入词向量以后除 TextCNN 模型与 C\_BiLSTM 模型, 其余模型的  $F1$  值均有所下降, 说明模型准确率有提高但分类性能有所下降, 词向量覆盖掉了原始文本的部分语义, 本文通过将混淆矩阵进一步分析模型的细粒度分类效果。

表3 不同模型准确率及  $F$  值 (%)

模型	准确率	$F1$ 值	准确率 (+词向量)	$F1$ 值 (+词向量)
CNN	53.43	47.05	48.54	43.40
Text CNN	52.39	44.73	53.14	45.91
LSTM	53.71	49.84	52.86	43.89
GRU	51.29	47.93	51.29	43.35
BiLSTM	50.64	49.89	53.64	46.03
BiGRU	50.36	49.46	54.50	47.05
C_BiLSTM	43.18	42.79	54.68	42.82
CNN-BiLSTM	52.54	48.35	55.07	44.17

图 14 中, 图 14(a) 表示模型未加入预训练词向量模型的混淆矩阵, 图 14(b) 表示加入预训练词向量模型的混淆矩阵。由混淆矩阵的表现可以看出, 在不加入词向量的情况下, 对 happiness, like, none, sadness 等情感类别文本预测效果较好。Anger, disgust 等情感类别文本预测效果一般。Fear 和 surprise 情感标签文本预测效果极差, 甚至出现类别缺失的情况, 为此查阅了原始的微博文本, 发现部分句子标签为 fear 的微博文本很难表征强烈的 fear 情感, 如“和总部来的领导们在珠江新城吃饭, 完后她们想去沙面逛逛, 我正好开了车所以当然由我送去。然后, 我就开始不停问同事沙面怎么走, 然后同事说先往 2 沙岛方向走, 然后我开始问 2 沙岛怎么走, 再然后很巧的我的 GPS 就死机了一直搜不到卫星信号[衰]你们说, 领导会不会觉得我是故意的?! [泪]”。“去微博广场看了十几页, 全这句话: 我喜欢周韵汤唯全智贤。嘛意思啊? 2 亿微博用户不会全体中毒

吧?真可怕!!这样宣传比买 CCTV 的 8 点广告好值啊!!!”。相比强烈的 fear 情感, 这两句话更像是在吐槽个人的想法并抒发个人的情绪, 第二句中“好值啊”和前半部分表达的情感有区别, 句子的整体情感也因此产生变化, 也一定程度影响模型预测的结果。

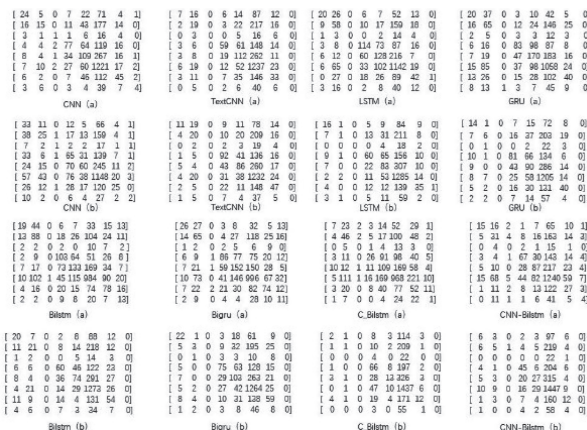


图 14 混淆矩阵

结合混淆矩阵的结果, 加入词向量以后模型的整体准确率有所提高, 但具体到某一情感类别, 特别是情感标签数较少的类别, 模型的预测结果不甚理想, 出现在了预测时情感类别缺失的状况, 这种情况以数据量较小的情感标签为主。与未加入词向量时的结果相比, 情感类别缺失更多, 说明词向量覆盖了部分情感词汇的文本信息, 不能很好的表征微博文本的细粒度情感, 可以尝试构建更大规模的基于微博文本的词向量, 同时, 中文自然语言处理流程中停用词的去除, 也对口语化、符号化、图形化严重的微博文本产生了一些影响, 可以尝试构建基于微博文本的情感词汇库或微博分词系统。

与欧阳纯萍等<sup>[35]</sup>基于多策略融合的细粒度情感分析方法相比, 基于神经网络模型的方法最佳准确率高了大约 25%, 需要注意的是欧阳纯萍等的工作所用数据集剔除了情感标签为 None 的文本。与传统的基于机器学习方法的文本细粒度情感分类的效果相比, 基于神经网络模型的微博文本细粒度分类仍大有可为。

### 5 总结与展望

论文结合深度学习方法对微博文本进行了细粒度情感分析, 但对文本情感的探究较少, 试验了深度学习



中几种模型在情感细分类中的效果, 下一步除了构造新的模型, 还应当结合一定的情感知识; 分词时简单地用结巴分词进行了分词, 而微博文本新词、缩略词、表情符号较多, 这种分词不甚准确, 影响文本情感的表征, 也在一定程度上影响了情感细分类结果, 下一步可以考虑构建微博的情感词典; 同时, 不同情感标签的数目差距较大, 情感标签数目最多的 None 与最少的 fear 比例近乎 50:1, 对分类结果也有极大的影响; 模型对文本的上下文信息利用较少, 可考虑使用 Elmo、Bert 等基于上下文信息的预训练语言模型, 或加入 Attention 机制进行优化. 与微博文本领域主题分类的相关工作相比, 微博文本的情感细粒度分类任务仍任重而道远.

### 参考文献

- 1 CNNIC 中国互联网络信息中心. 第 43 次中国互联网络发展状况统计报告. 北京: 中国互联网络信息中心, 2019.
- 2 唐晓波, 刘广超. 细粒度情感分析研究综述. 图书情报工作, 2017, 61(5): 132-140.
- 3 魏韡, 向阳, 陈千. 中文文本情感分析综述. 计算机应用, 2011, 31(12): 3321-3323.
- 4 陆文星, 王燕飞. 中文文本情感分析研究综述. 计算机应用研究, 2012, 29(6): 2014-2017. [doi: 10.3969/j.issn.1001-3695.2012.06.003]
- 5 丁兆云, 贾焰, 周斌. 微博数据挖掘研究综述. 计算机研究与发展, 2014, 51(4): 691-706. [doi: 10.7544/issn1000-1239.2014.20130079]
- 6 张剑峰, 夏云庆, 姚建民. 微博文本处理研究综述. 中文信息学报, 2012, 26(4): 21-27, 42. [doi: 10.3969/j.issn.1003-0077.2012.04.003]
- 7 敦欣卉, 张云秋, 杨铠西. 基于微博的细粒度情感分析. 数据分析与知识发现, 2017, 1(7): 61-72.
- 8 陈建美. 中文情感词汇本体的构建及其应用[硕士学位论文]. 大连: 大连理工大学, 2009.
- 9 刘丽, 王永恒, 韦航. 面向产品评论的细粒度情感分析. 计算机应用, 2015, 35(12): 3481-3486, 3505. [doi: 10.11772/j.issn.1001-9081.2015.12.3481]
- 10 李阳辉, 谢明, 易阳. 基于深度学习的社交网络平台细粒度情感分析. 计算机应用研究, 2017, 34(3): 743-747. [doi: 10.3969/j.issn.1001-3695.2017.03.024]
- 11 张谦, 高章敏, 刘嘉勇. 基于 Word2vec 的微博短文本分类研究. 信息安全, 2017, (1): 57-62. [doi: 10.3969/j.issn.1671-1122.2017.01.009]
- 12 梁军, 柴玉梅, 原慧斌, 等. 基于极性转移和 LSTM 递归网络的情感分析. 中文信息学报, 2015, 29(5): 152-159. [doi: 10.3969/j.issn.1003-0077.2015.05.020]
- 13 刘龙飞, 杨亮, 张绍武, 等. 基于卷积神经网络的微博情感倾向性分析. 中文信息学报, 2015, 29(6): 159-165. [doi: 10.3969/j.issn.1003-0077.2015.06.021]
- 14 李骁, 黄征. 基于 GRU 网络的互联网信息挖掘. 信息技术, 2018, (3): 1-5, 9.
- 15 吉祥飞, 李明东, 陶卫国, 等. 基于多通道 LSTM-CNN 模型的 Twitter 情感分析. 廊坊师范学院学报(自然科学版), 2019, 19(2): 21-24.
- 16 Salton G, Wong A, Yang CS. A vector space model for automatic indexing. Communications of the ACM, 1975, 18(11): 613-620. [doi: 10.1145/361219.361220]
- 17 Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. arXiv: 1301.3781, 2013.
- 18 Le QV, Mikolov T. Distributed representations of sentences and documents. arXiv: 1405.4053, 2014.
- 19 Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model. Journal of Machine Learning Research, 2003, (3): 1137-1155.
- 20 Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology, 1962, 160(1): 106-154. [doi: 10.1113/jphysiol.1962.sp006837]
- 21 Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 1980, 36(4): 193-202. [doi: 10.1007/BF00344251]
- 22 LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324. [doi: 10.1109/5.726791]
- 23 Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. arXiv: 1404.2188v1, 2014.
- 24 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述. 计算机学报, 2017, 40(6): 1229-1251. [doi: 10.11897/SP.J.1016.2017.01229]
- 25 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述. 计算机应用, 2016, 36(9): 2508-2515, 2565. [doi: 10.11772/j.issn.1001-9081.2016.09.2508]

- 26 Kim Y. Convolutional neural networks for sentence classification. arXiv: 1408.5882, 2014.
- 27 Zhang Y, Wallace B. A sensitivity analysis of (and Practitioners' Guide to) convolutional neural networks for sentence classification. arXiv: 1510.03820v1, 2016.
- 28 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735–1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)]
- 29 Cho K, van Merriënboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv: 1406.1078v3, 2014.
- 30 Jozefowicz R, Zaremba W, Sutskever I. An empirical exploration of recurrent network architectures. Proceedings of the 32nd International Conference on Machine Learning. Lille, France. 2015. 2342–2350.
- 31 Ghosh S, Vinyals O, Strope B, *et al.* Contextual LSTM (CLSTM) models for large scale NLP tasks. arXiv: 1602.06291v2, 2016.
- 32 Zhou CT, Sun CL, Liu ZY, *et al.* A C-LSTM neural network for text classification. arXiv: 1511.08630v2, 2015.
- 33 孙晓, 何家劲, 任福继. 基于多特征融合的混合神经网络模型讽刺语用判别. *中文信息学报*, 2016, 30(6): 215–223.
- 34 Li S, Zhao Z, Hu RF, *et al.* Analogical reasoning on Chinese morphological and semantic relations. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia. 2018. 138–143.
- 35 欧阳纯萍, 阳小华, 雷龙艳, 等. 多策略中文微博细粒度情绪分析研究. *北京大学学报 (自然科学版)*, 2014, 50(1): 68–67.