

# 哈爾濱工業大學

## 課程研究報告

課程名稱： 情感分析和計算

報告題目： 基於 Bi-LSTM 的微博評  
論情感三分類系統

所在院系： 計算學部

所在專業： 人工智能

學生姓名： 王藝丹

學生學號： 119020130

選課時間： 2022 年春季學期

評閱成績：

# 基于 Bi-LSTM 的微博评论情感三分类系统

王艺丹

1190201303

**摘要：** 随着社交媒体的飞速发展，微博等平台用户数量的快速增长，其中用户评论所携带的情感对社会影响愈发强大，甚至可能造成某种情绪的积累爆发，引发社会舆情，甚至影响社会稳定。因此，对微博评论进行情感分析，意义重大。微博评论情感分类的包括语料的预处理与情感分类方法等。常用的情感分类方法有基于情感词典的方法、基于机器学习的方法和基于深度学习的方法。随着深度学习从作为机器学习一个分支，逐渐发展至今，几乎成为一个独立的研究方向。其在 NLP 领域的使用也愈发广泛，本文对微博评论语料进行预处理，基于 word2vec 的文本特征表示方法，构建 Bi-LSTM 神经网络深度学习模型，对微博评论语料进行情感(positive, negative, neutral)三分类任务。

**关键词：** 微博评论；深度学习；Bi-LSTM；情感多分类

## 一、 研究目的和意义

文本情感分析又称意见挖掘，是指对带有情感色彩的主观性文本进行分析，挖掘其中蕴含的情感倾向，最终对其情感态度进行分类。

随着互联网的快速发展，大数据时代下，愈来愈多的用户在社交媒体上分享生活，发表对社会事件的看法，而微博、Twitter 等平台成为了火爆的公共舆论池，每天会发布传播上亿的信息。而这些海量文本信息中，很大一部分蕴含了用户的观点与情感倾向，是非常宝贵的资源。已经有许多研究者通过分析用户情感，解决了许多实际问题，如分析产品意见、股票收益、政治选举结果等。此外，社交平台上用户的情感可能会感染其他用户的情绪，使得某种情感在短时间内快速积累，造成社会舆论的爆发<sup>[1]</sup>。

因此，对各色平台上的不规则文本的情感态度进行分析计算，对于信息决策者，维持社会舆论良性发展等而言均有着巨大的价值意义。

## 二、 研究综述

### 2.1 国内外研究情况综述

在 20 世纪 90 年代末,文本情感分析在国外初步开始发展,发展至今,大致可分为基于情感字典的情感分析、基于机器学习的情感分析与基于深度学习的情感分析三类。

**基于情感字典的情感分析:**最初由 Riloff 和 Shepherd 文本数据的基础上进行了构建文本情感词典的相关研究;Hatzivassoglou 和 Mckeown 首次发现了连词对大规模的文本数据集中形容词语义表达的制约作用,研究发现在每个连词被认为是独立的情况下分类准确率达到 82%,启发了越来越多的研究者注重特征词与情感词的联系关系;Turney 等使用点互信息的方法进一步扩展了正负面情感词典,在分析文本情感时使用了极性语义算法。<sup>[2]</sup>

**基于机器学习的情感分类:**由于机器学习模型可以灵活地表示复杂关系,能够估计后验概率,执行时间较短,即使在有噪声数据的情况下,也具备良好的性能。因此,机器学习方法在情感分析中得到了广泛的应用。<sup>[3]</sup>其中基于机器学习的方法主要有:支持向量机(SVM)、K 最近邻(KNN)和最大熵(ME)。如 Pang 和 Le 等人对文本进行情感分类,分为正面和负面两种情感,并对比朴素贝叶斯、ME 和 SVM 这三种机器学习算法,对电影评论进行分类测试,最终发现 SVM 的分类效果最好。并且对于不同的输入特征,准确率大部分在 81%~83%;Boiy 和 Moens 基于机器学习的方法对三种语言(英文、荷兰语和法语)的博客、评论和论坛文本进行情感三分类。实验结果表明,三类语言分类的准确率依次为 83%、70%和 68%。国内则有唐慧丰和谭松波等人以 n-gram、名词、形容词和副词作为不同的文本表示特征,以互信息、信息增益、CHI 统计量和文档频率作为不同的特征选择方法,以中心向量法、KNN、Winnow、Naive Bayes 和 SVM 作为不同的文本分类方法,在不同的特征数量和不同规模的训练集情况下对情感分类结果进行了对比,发现采用 Bi-gram 特征表示方法、信息增益特征选择方法和 SVM 分类方法获得情感分类效果最好;夏火松和陶敏等人在文本预处理的过程中使用四种不同的停用词表,使用 TF-JIDF 权重计算方法进行特征选择,采用基于 RBF 核函数的 SVM 对携程网上的 4000 个酒店客户评论情感文本进行分类,结果发现不同的停用词表对情感分类的准确率不同。<sup>[4]</sup>

**基于深度学习的情感分类:**早期基于规则的分析方法要依靠人工设置的情感词典进行建模,不但有大量人力物力损耗,同时由于针对不同文本,使用的情感词典都有所不同,模型的通用性较差。而基于机器学习的分类方法虽不需要构建情感词典,但仍需要人工对文本特征进行筛选,且分析结果与人工筛选的特征准确性息息相关。而深度学习的特性之一就是可以从输入数据中自动学习所需的特征<sup>[5]</sup>,可以满足当下对海量文本进行情感分析的需求。其中主要应用的模型有 CNN、LSTM 等神经网络模型,并于注意力机制结合。Kim 等<sup>[6]</sup>使用 CNN 进行情感分析,将预训练的文本词向量作为输入,通过 CNN 进一步对文本特征进行提取,最终由全连接层送到分类器中,并取得了不错的效果。朱烨等<sup>[7]</sup>考虑到相似文本中的语义信息的影响,通过最邻近节点算法获得更准确的文本特征,之后使用双通道卷积神经网络模型对文本进行情感分析,也取得了不错的结果;Hochreiter 等<sup>[8]</sup>提出长短期记忆网络(LongShort-Term Memory, LSTM)模型,该模型不仅可以弥补 RNN 的不足,还能提取序列数据中长距离关联的语义特征。Liu 等<sup>[9]</sup>为了减少模型在处理文本数据、时间数据等序列信息的训练时间,Cho 等<sup>[10]</sup>提出的门控循环单元(Gated Recurrent Unit,GRU)模型在 LSTM 结构的基础上进行简化,能够有效的减少模型的训练时间;而注意力机制在自然语言处理方面的应用常常被看作是一种信息加权,通过分析每一时刻输出信息的重要程度来对输出

信息进行加权，可以帮助模型提取语义信息，这可以有效提高模型的学习效率。Yang 等<sup>[11]</sup>将注意力机制和神经网络结合来进行文本分析并取得了较好的实验效果。石磊等<sup>[12]</sup>将 Tree-LSTM 和 attention-Tree-LSTM 两种情感分析模型进行对比，结果表明引入注意力机制能有效的提高模型精度。

## 2. 2 本文研究方法选取原因

为避免传统机器学习算法时序预测中特征表达能力弱、准确率低且不能表示上下文语义关系与词序的缺陷。本文采取基于深度学习的方法，基于 word2vec 构建文本特征，采取双向多层 LSTM 神经网络训练，进行情感三分类任务。

# 三、 研究方法

## 3. 1 研究方法详叙

首先，分析待分类文本的特点，对数据进行清洗及预处理。考虑到表情、微博话题中蕴含的情感色彩，将其保留，去除微博用户名，url 等噪声词；对于用户转发的评论，保留用户自身的评论，若无则保留最近用户的评论，以此循环类推；进一步考虑否定词对情感色彩判断的影响，将一些具有代表性的否定词，如“不”等，与其紧紧跟随的情感色彩词语进行合并处理，作为一个词汇用于特征学习。基于 word2vec 文本表示，构建文本特征用于后续模型训练学习。

对于词向量的训练，分别采取对训练数据进行词典构建，训练词向量、使用知乎已经训练好的不同特征的词向量模型<sup>[13]</sup>，与后续各参数进行组合训练，选取测试集效果最好的参数组合作为最终模型参数。

此外，本文主要采用 Bi-LSTM、多层 LSTM 的方法，学习上下文关系，输入上文，输出下文，使用训练好的词向量作为双通道特征输入给双向长短期记忆网络进行学习，并将特征融合结果输入长短期记忆网络进一步学习，最后通过全连接层输出情感分类结果。

## 3. 1 补充说明

模型超参数的选取对性能有着很大的影响，因此，首先对预处理后的分词文本数据进行长度统计，选取合适的 max\_len 作为模型特征输入维度，统计结果如下图所示：

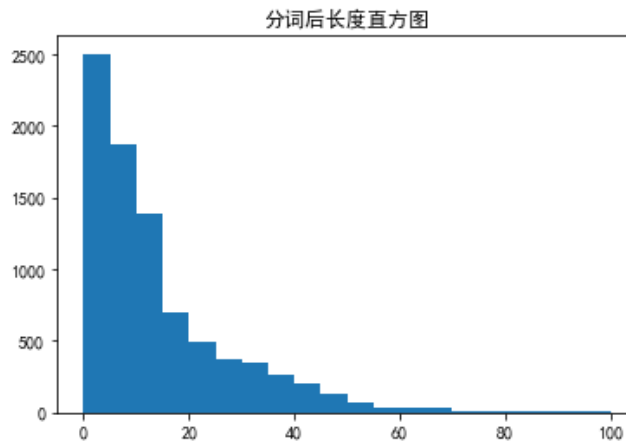


图 1 文本预处理长度分布统计直方图

结合既往经验考虑，最终各超参数备选值选取如下表所示：

词向量	训练集数据训练	word	word+bigrams		
max_len	40	60	70	80	100
batch_size	64	128	256		
随机丢弃概率	0.3	0.5			
优化目标	val_loss	val_acc			

表 1 超参数备选值

按上述超参数进行排列组合进行多次训练，以 macro\_f 作为模型评估指标，分别在测试集上进行验证，选取性能最佳的模型所对应的最优组合数值作为最终超参数。

## 四、实验结果及分析

### 4.1 结果展示

不同参数组合测试集表现如下（仅展示部分组合）：

词向量	max_len	batch_size	随机丢弃概率	优化目标	F1 值
word	100	128	0.5	val_loss	0.79
word	100	128	0.5	val_acc	0.77
word	100	128	0.3	val_loss	0.76
word	80	128	0.5	val_loss	0.76
word	100	256	0.5	val_loss	0.76
word	80	256	0.5	val_loss	0.76
word	100	256	0.3	val_loss	0.75
word+bigrams	100	128	0.5	val_loss	0.75
word+bigrams	100	256	0.5	val_loss	0.75

表 2 部分超参数组合及模型表现

最终确定最优参数组合如下：

词向量	max_len	batch_size	随机丢弃概率	优化目标
word	100	128	0.5	val_loss

表 3 最终最优超参数组合

模型在测试集最终表现结果如下，位列性能榜榜首：

F1(macro)	accuracy
0.811406	0.847667

表 4 模型最终效果（测试集）

### 4.2 结果分析

在超参数优化时发现 max\_len 与 batch\_size 改变时模型精度改变较大。

#### ➤ max\_len

最初认为选取 max\_len 为 80 已经可以保留足够多的文本特征信息，但数据

表明 `max_len` 为 100 时模型效果更佳。为进一步优化，也有选取 150、200 作为参数进行训练，但效果均无明显提升。经分析认为，长度越大的文本蕴含的情感更为复杂，可能是多重转折、否定、对比等复杂长语句，对上下文依赖较强，较短的 `max_len` 可能会造成部分关键词语的丢失从而影模型对整个句子情感的判断。

#### ➤ `batch_size`

在 `batch_size` 由 64 改为 128 时，模型精度明显提高。分析认为，当 `batch_size` 设为 64 时，结合丢弃率的设置，该配置对于该样本数据量而言过小，导致了模型的过拟合，造成精度损失。当 `batch_size` 设为 128 时，模型泛化能力较强，性能得到提升。而进一步选取 256 等数据时，精度稍有下降，且较大的 `batch_size` 可能虽然能够达到与较小 `batch_size` 相似的训练误差，但由经验研究来看，往往对测试数据的泛化效果较差。所以，最终选取 128 为 `batch_size` 大小。

## 五、 结论

本次实验基于简单的 Bi-LSTM 网络进行深度学习，完成了一个微博评论文本情感三分类系统。

首次实战深度学习模型与接触情感分类，搭建的神经网络结构与构造特征等较为简单，未来主要考虑改进方向有：

- 数据预处理：针对用户自身的评论与他人转发内容进行一定筛查，在保留用户自身评论的前提下尽可能去掉噪声词。
- 文本特征方面：未针对未登录词补充词向量；可以结合已有情感词典，对词向量赋予不同的权重，并以结合逆向词频等进行特征融合。
- 模型方面：参数调优，并考虑加入注意力机制，搭建更为复杂合理的网络模型进行训练。
- 整体展望：利用集成学习与其他模型进行融合，提高模型性能。

## 六、 学完本课程后的收获和体会

- ✧ 首次接触深度学习模型，并在实验过程中首次进行实战运用，对深度学习的理解与掌握进一步深入
- ✧ 对词向量的训练与表示方法有了更深入的学习理解
- ✧ 对于情感分析计算这一研究领域有了初步认知与进一步的了解学习
- ✧ 了解国内外前沿的情感分析计算方法

## 七、 参考文献

- [1] Chen Juan, Liu Yan-ping, Deng Sheng-li. An analysis on factors influencing the dissemination effect of rumor-refuting information[J]. Information Science, 2018, 36(1): 91-95. (in Chinese)
- [2] Hatzivassiloglou V, Mckeown K. Predicting the Semantic Orientation of Adjectives[J]. Proceedings of Acl, 1997: 174-181.
- [3] 李梦楠, 汪明艳. 基于机器学习的情感分析方法及应用研究综述[J]. 软件工程, 2021, 24(09): 21-23+8.
- [4] 徐琳宏, 林鸿飞. 基于语义特征和本体的语篇情感计算[J]. 计算机研究与发展, 2007(2): 356-360.
- [5] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations[J]. 2018: 2227-2237.
- [6] Kim Y. Convolutional Neural Networks for Sentence Classification[C]. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) 2014: 1746-1751.
- [7] 朱烨, 陈世平. 融合卷积神经网络和注意力的评论文本情感分析[J]. 小型微型计算机系统, 2020, 41(03): 551-557.
- [8] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [9] Liu P, Qiu X, Huang X. Recurrent Neural Network for Text Classification with Multi-Task Learning[C]. AAAI Press, 2016: 2873-2879.
- [10] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C]. Proceedings of the 2014 Conference on Empirical Methods in Natural Language. Doha: CoRR, 2014: 1724-1734.
- [11] Yang Z, Yang D, Dyer C, et al. Hierarchical attention Networks for document classification[C]. North American chapter of the Association for Computational Linguistics. San Diego, California: ACL, 2016: 1480-1489.
- [12] 石磊, 张鑫倩, 陶永才, 等. 结合自注意力机制和 Tree-LSTM 的情感分析模型[J]. 小型微型计算机系统, 2019, 40(7): 1486-1490.
- [13] Shen Li, Zhe Zhao, Renfen Hu, Wensi Li, Tao Liu, Xiaoyong Du, Analogical Reasoning on Chinese Morphological and Semantic Relations, ACL 2018