

# 基于深度学习的微博用户情感分析方法研究与应用

邓智超\*

<sup>1)</sup>(天津大学, 智能与计算学部, 天津, 中国, 300354)

**摘要** 本文训练了一种结合 BERT 和 BiLSTM 的中文文本情感分类混合模型, 该模型使用 BERT-base-chinese 作为特征提取器, 将特征输入到 BiLSTM 中并通过全连接层将特征表示转化为最终的分类结果。实验结果表明, 该模型在中文情感分类任务中表现出色, 相较于单独使用 BERT-base-chinese 和 BiLSTM 的模型取得了更好的分类效果。基于上述模型, 本文设计并实现了一个网站应用平台, 该平台提供了丰富的可视化界面, 实现了微博数据的获取、中文文本情感分析等功能, 充分展示了模型在实际问题中的应用价值。所有代码已上传至[https://github.com/dzcgood/sentiment\\_analysis](https://github.com/dzcgood/sentiment_analysis)。

**关键词** 中文文本情感分析; BERT; BiLSTM; 网页应用

## Research and Application of Weibo User Sentiment Analysis Method Based on Deep Learning

Zhichao Deng

<sup>1)</sup>(College of Intelligence and Computing, Tianjin University, City Tianjin)

**Abstract** This paper introduces a Chinese text sentiment classification hybrid model combining BERT and BiLSTM. The model uses the pre-trained BERT-base-chinese model as a feature extractor, inputs the features into BiLSTM and converts the feature representation into final classification result. The experimental results show that the model performs well in the Chinese emotion classification task, and achieves better classification results than the model using BERT-base-chinese and BiLSTM alone. Based on the above model, this paper designs and implements a website application platform, which provides a rich visual interface, realizes functions such as microblog data acquisition and Chinese text sentiment analysis, and fully demonstrates the application value of the model in practical problems. Code is available in Github at [https://github.com/dzcgood/sentiment\\_analysis](https://github.com/dzcgood/sentiment_analysis).

**Keywords** Chinese text sentiment analysis; BERT; BiLSTM; web application

\* 在本文的实现过程中, 参考了一些开源的代码。其中, 数据预处理和模型构建基于fushengwuyu, BrownSweater, msft-vivi等人开源的代码改进; 微博爬虫相关的代码基于nghuyong开源的代码改进, 点击姓名可跳转至对应的代码仓库。

## 1 简介

最近发现微博的每个话题下面都会有对大众情绪的分析 and 统计, 将大众的情绪分为“平和”、“生气”、“惊讶”、“开心”、“喜欢”、“失望”六种。正好这学期学习了深度学习这门课程, 结合自己之前所学的 Web 开发相关知识, 我也想复刻一下这个功能。因此, 本文主要工作内容是结合目前有的模型, 对其做针对性的改进, 以微博文本为例, 训练了一个中文文本情感分类混合模型, 然后设计并实现了一个基于该模型的网站应用场景, 以展示模型在实际问题中的应用价值。本文的主要创新点包括:

- (1) 训练了一种基于 BERT-base-chinese<sup>[1]</sup> 和 BiLSTM<sup>[2]</sup> 的混合模型, 该模型能够有效地提取微博文本数据中的语义特征和情感倾向。
- (2) 构建了一个微博情感分析网站, 该网站能够实现直接文本情感分析、微博用户的动态数据的爬取和分析等功能, 并提供了饼图和词云等可视化结果。

## 2 相关工作

情感分析最早由 Nasukawa 等<sup>[3]</sup> 提出, 其任务是从文字中提取出人们的情感或观点。在三种主要的情感分析方法中, 基于情感词典的方法需要在进行情感计算前建立情感词典<sup>[4]</sup>; 基于机器学习的方法需要进行人工特征提取, 然后使用机器学习算法进行情感分类, 常见的机器学习算法有 SVM、NB、KNN 等。此外, 基于机器学习的方法可分为无监督的情感分析和有监督的情感分析两种方式。有监督的情感分析虽然精度较高, 但高度依赖于人工数据的标注; 无监督的情感分析则不需要人工标注数据, 但其精度一般较低。

基于深度学习的方法是日前研究热度最高的, 它不需要进行人工特征提取, 可以自动提取特征并进行情感分类。RNN 模型<sup>[5]</sup> 解决了 CNN 模型<sup>[6]</sup> 的输出只考虑前一个输入而不考虑其他时刻输入的问题。为解决 RNN 的梯度消失和梯度爆炸的问题, Jurgen Schmidhuber 等<sup>[7]</sup> 提出了 LSTM 模型。

Bahdanau 等<sup>[8]</sup> 首次将注意力机制应用到自然语言处理中, 通过给不同的位置分配不同的权重, 实现对重要信息的关注, 解决了 RNN 在处理长序列时忽略重要词汇的作用以及难以考虑到全部文本信息的问题。Vaswani 等<sup>[9]</sup> 提出的 Transformer 模型使用自注意力机制来学习文本表述, 可以考虑到每个词在全文中的作用。此外, 不少学者将不同的神经网络结合起来组成混合模型, 用于情感分析。预训练模型在众多 NLP 任务中取得了很好的结果, 可以将大量的数据训练好的模型保留下来, 出现一个新任务时, 只需要针对特定任务进行微调或者迁移学习。目前, 较新的预训练模型有 ELMo<sup>[10]</sup>、BERT 等。

目前主要的文本情感分析工作集中在英文数据集上, 中文和英文存在着语法、语义等方面的差异, 这使得中文的文本情感分析更加复杂。此外, 微博中经常使用的表情符号在现有的模型中也没有很好的体现, 大部分模型仍然使用二分类情感分析且对于多分类的情感分析效果不佳。若是基于传统的单一模型来训练, 在准确率上差强人意, 因此, 需要结合比较目前有的模型, 并对其做针对性的改进。

## 3 微博数据预处理

### 3.1 数据集介绍

本文使用了 SMP2020-EWECT<sup>[11]</sup> 比赛数据集中的通用微博数据集来训练模型, 这个数据集是由哈尔滨工业大学社会计算与信息检索研究中心标注的, 原始数据来自新浪微博, 由微热点大数据研究院提供。数据集分为训练集、验证集和测试集, 分别含有 27,768 条、2,000 条和 5,000 条微博。这些微博是随机采集的, 涉及各种不同的话题, 范围很广泛。每条微博都被归类为以下六种情绪之一: neutral (无情绪)、happy (积极)、angry (愤怒)、sad (悲伤)、fear (恐惧)、surprise (惊奇)。

### 3.2 数据统计与分析

为了更好地理解数据集的特点, 本文对数据集进行了数据分布特征的分析。具体而言, 本文从类

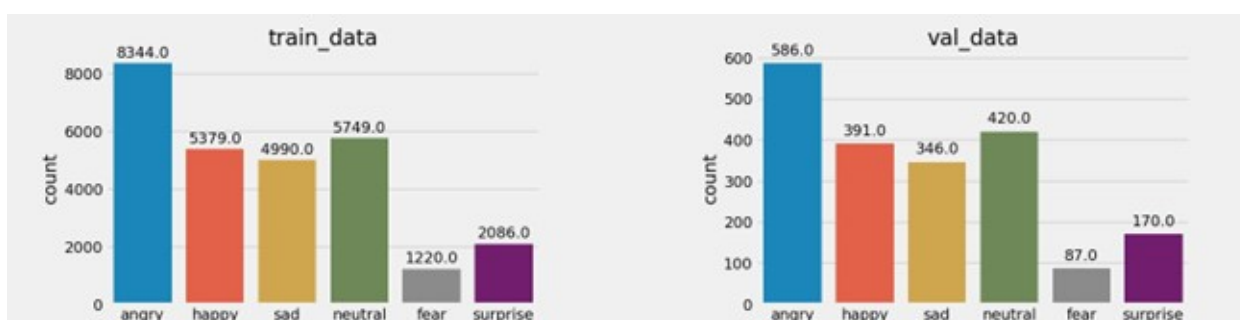


Fig. 1 train 和 val 数据集标签类别分布

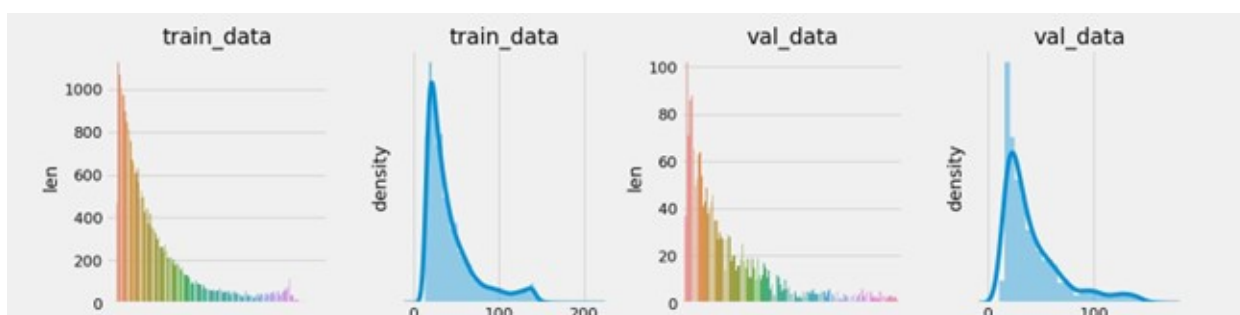


Fig. 2 train 和 val 数据集数据长度和频率统计

别占比和文本长度两个方面对数据集进行了分析。

图 1 是训练集和验证集中各类情绪标签数量的统计。可以看出数据集具有以下特点和问题：

- (1) 类别分布不平衡：在训练集中，生气文本数量最多，积极、伤心和无情绪文本占据大部分，而恐惧和惊奇的样本数较少，只占总数的 9.2%。这可能会导致模型在对恐惧和惊奇这两种情感的分类上表现较差，在学习 anrgy 标签的文本时可能会比其他情感文本更加具有倾向性，此外，这种不平衡也可能导致模型在测试集上表现不佳。
- (2) 数据总数较少：这个规模对于一些深度学习模型来说可能还不够大，可以考虑通过数据增强来提高模型的性能。
- (3) 训练集和验证集的各个情绪类别的分布比例大致一致，又因为数据是随机从微博采集的，所以也可以从一定程度上反映出目前微博平台上的文本情绪分布情况。

图 2 是对训练集和验证集中文本长度的统计。可以看出，文本长度有很大的差异，最长的超过一千字，最短的只有几个字。其中，文本长度在 0-50 字之间的占比最高，超过 150 字的占比极少。根据这个分布情况，我们可以在训练过程中设定文本输入模型的最大长度，以提高训练效率和准确性。

### 3.3 数据清洗与预处理

在进行微博文本情感分析之前，需要对原始数据进行清洗处理以确保数据的准确性和规范性。数据清洗的目的是消除原始文本数据中的无效信息和噪声，以便提高情感分析结果的精确度。微博文本情感分析中的数据清洗步骤包括：

- 去除文本中的 URL 链接
- 去除用户 ID 和转发信息
- 将繁体转换成简体、全角字符转换成半角字符
- 将表情包通过 emoji2text 转成文本

由于训练集数据量较少，只有 27,768 条。本文

对数据集进行了数据增强，以提高情感分析结果的准确性，并防止训练过程中出现过拟合。此外，微博文本中由于用户输入时的语音输入、输入法自动调整等原因，常常会出现较多的错别字和拼音错误，这些错误会影响到情感分析的准确性和可靠性。综合以上两点，本文选择的数据增强方法是同音字替换，在增加数据集的数量的同时模拟了真实微博文本的特点，丰富了文本数据的多样性以提高模型的泛化能力。

## 4 方法

### 4.1 模型结构

模型的整体结构如图 3 所示，它由三部分组成：基于中文文本预训练的 BERT-base-chinese、双向长短期记忆神经网络 (BiLSTM) 和全连接层。模型的输入是中文文本，经过 BERT 模型进行词向量表示后，将表示结果输入到 BiLSTM 中进行序列建模。在 BiLSTM 中，正向 LSTM 和反向 LSTM 对文本序列进行前向和后向建模，并将每个时刻的正向和反向输出拼接起来作为该时刻的输出。经过整个序列的建模后，将正向 LSTM 和反向 LSTM 最后一个时刻的输出连接在一起，将其作为输入送入全连接层进行分类。最终，模型将输出一个长度为 6 的向量，即文本属于各情感类别的概率。BERT-base-chinese + BiLSTM 的组合可以充分利用 BERT-base-chinese 的强大的预训练能力和 BiLSTM 的序列建模能力，实现微博文本情感分析的高效和精准。

### 4.2 模型训练

本文使用了一组特定的参数来训练模型，以提高其准确性和效率。具体来说，模型训练使用了 NVIDIA Tesla P100 显卡，并将 batch size 设置为 32，epoch 设置为 10。此外，还采用了 AMP 混合精度训练技术，以更好地平衡训练时间和计算资源消耗。对于 BERT 和 BiLSTM 两种不同的模型组成部分，分别设置学习率为  $1e-5$  和  $2e-5$ ，并在训练过程中使用了长度为 3 epoch 的 Warm up 学习率调整策略。基于 AdamW 在自然语言处理任务中

的优秀表现，本次研究选择其作为优化器，并将序列的最大长度设置为 140。

此外，由于 BERT-base-chinese 是预训练模型，具有更强的泛化能力，因此只需要进行小幅度的调整，相反，BiLSTM 需要参数的彻底更新，因此需要更大的学习率。因此，针对 BERT-base-chinese 和 BiLSTM 不同的特性和优化需求，采用了不同的学习率进行调整。对于 BERT 部分，设置了较小的学习率  $1e-5$ ，而对于 BiLSTM 部分，设置了较大的学习率  $2e-5$ ，以便更快地更新权重参数。

### 4.3 评价指标

模型训练中选择交叉熵损失函数，它能够度量真实概率分布与模型预测概率分布之间的差异，是连续可导的且在梯度下降过程中具有较好的稳定性，便于优化算法求解和模型收敛。式 (1) 是交叉熵损失函数的计算公式，其中， $M$  是指类别的数量；当样本  $i$  的真实类别等于  $c$  时， $y_{ic}$  的值取 1，否则取 0； $p_{ic}$  是指观测样本  $i$  属于类别  $c$  的预测概率。

$$\text{Loss} = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_c y_{ic} \log(p_{ic}) \quad (1)$$

Macro-F1，即宏平均 F1 值，是情感分类任务中常用的评价指标。宏平均 F1 值对每个类别的 F1 值取平均，能够平衡不同类别的数量和分布，更客观地评价模型的性能。具体计算方法如公式 (2) 和 (3) 所示：

$$P_e = \frac{TP_e}{TP_e + FP_e}, R_e = \frac{TP_e}{TP_e + FN_e}, F_e = \frac{2 \times P_e \times R_e}{P_e + R_e} \quad (2)$$

$$\text{Macro\_F} = \frac{1}{n} \sum_e F_e \quad (3)$$

其中， $e$  表示情感类别，Emotions 的取值是 happy, angry, sad, fear, surprise, neutral， $TP_e$  表示预测为情感  $e$  且正确分类的样本数， $FP_e$  表示预测为情感  $e$  但错误分类的样本数， $FN_e$  表示实际情感为  $e$  但被错误分类的样本数。

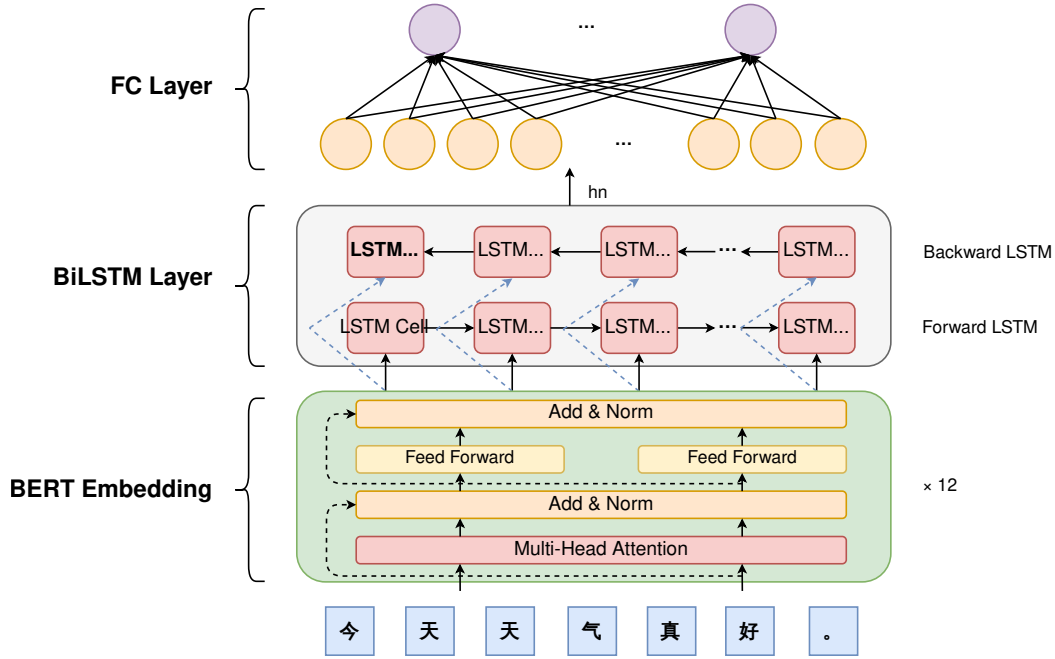


Fig. 3 模型结构图

## 5 实验

### 5.1 性能分析

图 4 展示了模型训练过程中准确率、宏平均 F1 值、损失函数的值的变化情况,可见准确率、宏平均 F1 值在前三个 epoch 逐渐提升并在第三个 epoch 达到最佳值,之后的 4 到 10 个 epoch 准确率和宏平均 F1 值总体呈下降趋势,这可能是由于数据量过少导致出现了过拟合现象。最终,模型在 val 集中准确率为 0.792、宏平均 F1 值为 0.762、损失函数数值为 0.615;在测试集中,准确率为 0.776、宏平均 F1 值为 0.745、损失函数数值为 0.628。

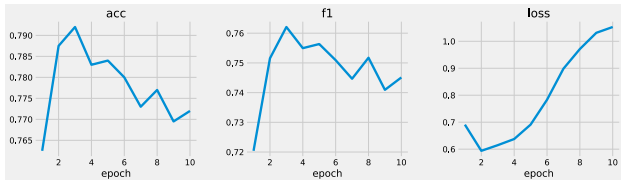


Fig. 4 val 数据集中测试结果

图 5 展示了模型在测试集中各类标签下的表现。可以看出,模型在“neutral”、“angry”和“happy”这三个类别下表现较好,准确率分别达到了 88.6%、83.5% 和 80.1%。然而,在“surprise”、“fear”

和“sad”这三个类别下的表现较差,准确率分别为 69.3%、65.8% 和 63.3%。造成这种现象的原因可能是训练集中“neutral”、“angry”、“happy”的数据数量多于“surprise”、“fear”、“angry”的数据数量。

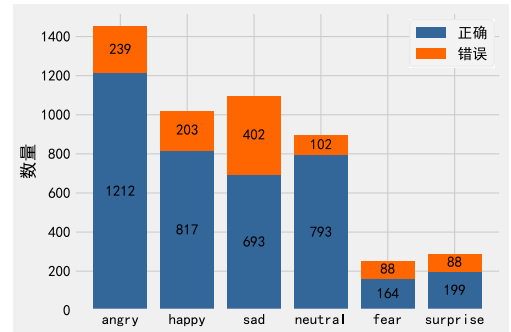


Fig. 5 模型在各标签中测试的表现

### 5.2 模型对比

本模型上述所提及数据集的 train 与 val 数据上进行训练,在 test 数据上进行测试,训练和测试的平台均为双核 Intel(R) Xeon(R) CPU @ 2.00GHz, NVIDIA Tesla P100。为了突显本模型相较于单个模型的提升效果,本文对比了使用相同平台下的单独 LSTM、BiLSTM 和 BERT 模型在测试集上的表现。表 1 展示了本模型与各模型的性能对比。

表 1 模型对比

模型	Acc	Loss	F1	模型大小	平均推理时间
LSTM + Classifier	0.644	1.178	0.611	744MB	8.541ms
BiLSTM + Classifier	0.671	1.080	0.639	907MB	15.716ms
BERT + Classifier	0.693	0.897	0.658	1.14GB	8.873ms
本模型	0.776	0.628	0.745	1.40GB	24.443ms

实验一，与单独使用 LSTM 对比：LSTM 模型的 hidden size 和 embedding size 设置为与本模型相同的 768 维，得到了 LSTM 模型在测试集上的表现，与本模型相比，LSTM 模型表现明显较弱。具体来说，本模型相比 LSTM 提升了 13.2% 的准确率、0.134 的 F1 值，并且损失值更小。这表明本模型在关键指标上的表现要优于 LSTM 模型。

实验二，与单独使用 BiLSTM 对比：在与 BiLSTM 的对比实验中，BiLSTM 模型的各项超参数设置与 LSTM 模型相同。相比 LSTM 模型，BiLSTM 模型在测试集上的准确率、F1 值、损失值均略有进步，但与本模型相比仍有较大差距，本模型的准确率相比 BiLSTM 提升了 10.5%，F1 值提升了 0.106。这再次表明，本模型在关键指标上的表现要优于单独使用 BiLSTM 模型。

实验三，与单独使用 BERT-base-chinese 对比：在此次实验中，设置 BERT-base-chinese 的各项参数都与本模型相同。可以看出 BERT-base-chinese 的表现虽优于 BiLSTM，但本模型仍具有优势，准确率提升了 8.3%，F1 值提升了 0.087。

## 6 微博情感分析网站涉及与实现

### 6.1 网站架构设计

图 6 是网站的架构图，网站采用前后端分离的架构，前端使用了 Vue3 和 Element Plus 组件库，网站一共包括四个主要界面，分别是导航栏界面、文本情感分析界面、微博情感分析界面、用户情感分析界面，页面的跳转和管理由 Vue Router 实现，使用 Axios 实现前后端交互；后端使用了 Python Flask 框架，将不同的功能模块分别封装为不同的接口，包

括进行情感分析、进行微博数据收集、生成词云等，使用 Scrapy 进行微博数据的收集、使用腾讯云提供的 API 进行图片的上传与下载。

后端共有四个模块，其中微博爬取模块负责根据系统需求，按照指定格式收集微博平台的数据并保存在本地；情感分析模块负责分析传入文本的情感倾向；词云生成模块负责将传入的文本分词后生成词云，并上传到腾讯云；接口响应模块负责响应前端的请求，根据不同的请求地址调用不同的功能模块并返回结果。

### 6.2 功能展示

文本情感分析界面的主要组件设计如附录图 1 所示，本页面提供了四个文本预设，用户可以选择一个文本预设，或者直接输入想要分析的文本，点击“开始分析”按钮，即可获取情感分析的结果。情感分析结果的展示由一张情绪图片和描述文字组成，分别是积极、愤怒、悲伤、恐惧、惊奇、无情绪，每种分类结果使用了对应的颜色来展示，附录图 2 是所有情绪图片和描述文字的展示，附录图 3 展示了进行文本情感分析之后，展示分析结果的界面。

微博情感分析界面包括三大组成部分，分别是输入部分，微博内容展示与分析部分以及历史记录部分。在输入部分，用户可以输入微博 ID 和 cookie，其中 cookie 是微博的登录信息，然后点击“爬虫一下”，在“微博内容展示与分析部分”就会展示爬取到的微博内容，包括用户头像、用户昵称、发表时间、来源、微博内容。点击“开始分析”按钮后网站会展示该微博对应的情感分析结果。附录图 4 展示了微博内容展示与分析部分的界面。

微博用户情感分析的界面由用户 ID 输入、微

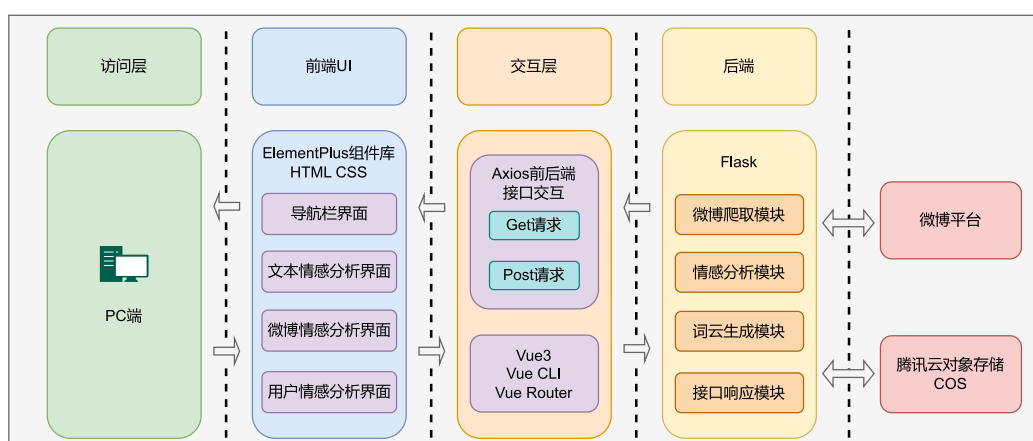


Fig. 6 网站架构图

博用户信息展示、饼图和词云展示、微博内容展示与分析四部分组成。附录图 5 和 6 展示了微博用户情感分析的结果，点击开始分析按钮，在“微博用户信息展示”部分会展示用户近十条微博的主情绪标签，即数量最多的情绪标签；在“饼图和词云展示”部分，会展示根据各情感标签的数量生成的饼图以及根据微博内容生成的词云图；在“微博内容展示与分析部分”会显示每一条微博的分析结果。

### 参考文献

- [1] LEE J, TOUTANOVA K. Pre-training of deep bidirectional transformers for language understanding: Vol. 3[A]. 2018: 8.
- [2] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks[J]. IEEE transactions on Signal Processing, 1997, 45(11): 2673-2681.
- [3] NASUKAWA T, YI J. Sentiment analysis: Capturing favorability using natural language processing[C]// Proceedings of the 2nd international conference on Knowledge capture. 2003: 70-77.
- [4] 王春东, 张卉, 莫秀良, 等. 微博情感分析综述.[J]. Computer Engineering & Science/Jisuanji Gongcheng yu Kexue, 2022, 44(1).
- [5] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization[A]. 2014.
- [6] FUKUSHIMA K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological cybernetics, 1980, 36(4): 193-202.
- [7] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [8] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[A]. 2014.
- [9] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [10] SARZYNSKA-WAWER J, WAWER A, PAWLAK A, et al. Detecting formal thought disorder by deep contextualized word representations[J]. Psychiatry Research, 2021, 304: 114135.
- [11] 中国中文信息学会社交媒体处理专业委员会. SMP2020-EWECT[EB/OL]. 2020[June 20, 2024]. <https://smp2020ewect.github.io/>.

## 附录

请输入一段需要分析的文本: [换一个示例](#)

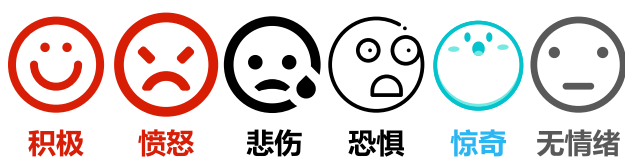
今天天气真好, 心情舒畅!

0 / 140

最多输入140字

开始分析

附录图 1 文本情感分析界面



附录图 2 情绪图片和描述文字

请输入一段需要分析的文本: [换一个示例](#)

听故事的时候,你总喜欢眼巴巴的问,后来呢,可显当后来,自己成为讲故事的人,才发现,后来故事和话语就在嘴边,后来眼泪几度泪疑于睫,是真的说不下去了。。

74 / 140

最多输入140字

开始分析

情感分析结果 悲伤

附录图 3 文本情感分析结果展示

DzcGood1  
2023-04-06 00:38:43 来自 iPhone客户端

苹果居然有这么逆天的统计.....

开始分析

情感分析结果 惊奇

附录图 4 微博情感分析结果界面





附录图 5 微博用户情感分析结果展示 - 1



附录图 6 微博用户情感分析结果展示 - 2