

黄山山. 基于 RoBERTa 模型的在线用户评论细粒度情感分析[J]. 智能计算机与应用, 2024, 14(7): 170-173. DOI: 10.20169/j.issn.2095-2163.240727

基于 RoBERTa 模型的在线用户评论细粒度情感分析

黄山山

(上海理工大学 管理学院, 上海 200093)

摘要: 本论文研究了基于 RoBERTa 的在线用户评论的细粒度情感分析模型。结合数据增强、模型微调、多任务学习和注意力机制等技术和方法,提升模型性能。本文使用了 Semeval2014 数据集,其中包含了各种情感类别的评论,如积极、消极和中性。实验结果显示,本文的方法在细粒度情感分析任务中,与传统方法相比取得了明显的性能提升,本文提出的模型能更好地识别评论中的情感极性。

关键词: 在线评论; 细粒度情感分析; NLP

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2024)07-0170-04

Fine-grained sentiment analysis of online user reviews based on the RoBERTa model

HUANG Shanshan

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: This paper studies the fine-grained sentiment analysis of online user reviews based on the RoBERTa model. By combining techniques and methods such as data augmentation, model fine-tuning, multi-task learning and attention mechanism, the model performance is improved. This paper uses the Semeval2014 dataset, which contains reviews of various fine-grained sentiment categories, such as positive, negative and neutral. The experimental results show that the proposed method achieves significant performance improvement compared with the traditional methods in the fine-grained sentiment analysis task. The proposed model can better identify the sentiment polarity in the reviews.

Key words: online reviews; fine-grained sentiment analysis; NLP

0 引言

在线评论通常定义为电子商务或者第三方网站上消费者发表的产品评价,主要表现为星级打分(1~5 颗星)和开放式评论文本^[1]。与传统的线下评论相比,在线评论具有速度快、范围广、信息量大、可存储、可测量性等特点,因此也成为了消费者购买决策过程中最具影响力的信息来源之一^[2]。随着互联网的日渐普及和各大电商平台的飞速发展,网络购物成交量逐年上升,电商平台上的产品评论数量也与日俱增。消费者在购物决策之时,除了关注商品的价格、销量外都会特别关注商品的评论,而商家也可以通过商品的评论来知晓消费者对该商品的感受,通过评论来强调目标商品的优点,改进商品的缺点。这些评论包含了各种情感,从积极的评

价到消极的抱怨,以及中性的观点。因此,情感分析成为了一个重要的研发课题,帮助企业了解用户对其产品和服务的感受,以及改善用户体验。细粒度情感分析进一步细化了情感分类,这就有利于更好地理解用户的情感。

在当下信息爆炸的网络环境中,用户和商家想要精准获取所需信息是十分困难的,很容易就迷失在海量的评论信息中,消费者及商家均已不再满足于对评论信息进行简单的好评或差评分类,而是需要了解目标商品的特征及情感。消费者渴望依据评论中对商品的某一特征的褒贬度选择中意的商品,而商家希望了解消费者对商品的各个特征的评价来做出后续精准的改进,这就需要对大量的评论进行整理分析,且在线评论中包含着大量的非结构化和半结构化的数据,所以如何高效且精准地对在线评

论进行情感分析成了当下研究热点。

1 相关工作

目前,国内外有很多在线评论相关的研究工作。这些研究获取社交媒体的评论文本数据,包括在线论坛数据、电商平台用户评论数据等,将从产品属性发现任务形式化为文本分类问题。但多是使用传统的粗粒度情感分析,如此就会影响其特定评价目标的情感判断,且整体情感并不能反映人们对意见目标的细粒度的情感表达,可能会在现实应用中传达出不准确信息。

情感分类是细粒度情感分析研究的另一重要方向。Munikaar 等学者^[3]使用了预先训练好的 BERT 模型,并将其用于 SST 数据集上的细粒度情感分类任务,该模型能够超越复杂的架构、如递归和卷积神经网络,有更高的准确性。Tan 等学者^[4]为了解决随着用户评论数量的增加而带来的情感分析方法高成本和高错误率的问题,提出了一种基于依赖树和图神经网络的细粒度情感分析方法。经数据验证,该模型优于大多数基线模型,能更好地编码语义关系信息,在捕获情感分析的重要语法结构方面的有效性得到显著提高。Balikas 等学者^[5]提出传统的情感分析方法。通过分别学习任务来解决三元和细粒度分类等问题,认为这种分类任务是相关的,并提出了一种基于递归神经网络的多任务方法(Bi-LSTM+Multitask),改进了在细粒度情感分类问题中的分类结果,从而为详尽分析多语言 and 不同情感粒度级别的长短不一的评论提供了可能。侯艳辉等学者^[6]将词频逆文档频率(TF-IDF)和 TextRank 算法用于中文影评的特征词提取上,并将 Bi-LSTM 和鲁契克多维度情感模型进行融合,实现了基于特征层面的情感强度的八分类。曹雪^[7]采用加权平均和门控单元机制对词语级、短语级和句子级三种粒度情感强度进行结合,进一步提升了情感分类的准确性。万岩等^[8]通过建立微博领域的情感词典及情感分析模型,提高了细粒度情感分类的准确率。Dong 等^[9]针对细粒度情感分类问题,根据语境和句法结构学习将词的情感传播给评价目标。Che 等^[10]利用文本的句法结构和语义特征训练一组分类器来完成细粒度情感分析任务。

2 数据集和预处理

实验主要使用 SemEval - 2014 Task4 中 Restaurant 和 Laptop 两个领域的评论数据集,每条

数据包含评论语句、评论语句中的评价对象以及评价对象的情感极性。这些数据集的标签有 3 个,分别是: Positive、Negative 和 Neutral。表 1 中是这 3 个数据集的训练集和测试集的一些统计情况。首先,进行数据预处理,包括分词、移除停用词和标点符号,并将评论转化为 RoBERTa 可接受的格式。

表 1 数据集统计情况

Table 1 The statistics of the data sets

Datasets	Positive		Negative		Neutral	
	Train	Test	Train	Test	Train	Test
Restaurant	2 164	728	807	196	637	196
Laptop	994	341	870	128	464	169

3 模型架构

RoBERTa 模型使用了动态掩码(Dynamic Masking),即每次向模型输入一个序列时都会生成新的掩码模式,而不是像原始的 BERT 那样在数据预处理期间执行一次掩码,得到一个静态掩码。使得模型可以适应不同的掩码策略,学习不同的语言表征。

RoBERTa 去掉了下一句预测(NSP)的任务,只保留了掩码语言模型(MLM)的任务。实验表明,在部分任务中 NSP 并不是必要的,而且可能会损害模型的性能。

RoBERTa 改变了文本编码方式,使用了字节级的 BPE(Byte Pair Encoding)词汇表,大小为 50 K,而不是原始的 BERT 使用的字符级的 BPE 词汇表,大小为 30 K。模型可以处理更多的词汇,而不需要对输入进行额外的预处理或分词。

RoBERTa 的模型架构是基于 Transformer 的,由多个层叠的自注意力层(Self-Attention Layers)和前馈神经网络(Feed-Forward Neural Networks)组成。在细粒度情感分析任务中,本文将使用 RoBERTa 的预训练权重,然后对其进行微调以适应特定任务。

RoBERTa 的自注意力层是核心组成部分,用于捕捉输入序列中不同位置的关联性,有助于理解文本的上下文信息。每个自注意力层包括以下组件。

(1) 自注意力机制(Self-Attention): 这是 RoBERTa 的关键部分,用于计算每个输入标记与其他标记之间的相关性权重。自注意力机制通过学习权重,使模型能够关注输入序列中不同位置的信息。同时允许模型在处理每个标记时考虑其他标记的上下文信息。自注意力机制的计算公式如下:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V} \quad (1)$$

其中, \mathbf{Q} 、 \mathbf{K} 和 \mathbf{V} 分别表示查询 (Query)、键 (Key) 和值 (Value) 的矩阵; d_k 表示查询和键的维度; Softmax 函数将计算结果归一化。

(2) 前馈神经网络 (Feed-Forward Neural Network): 在自注意力机制之后, 每个位置的表示会通过一个前馈神经网络进行变换和非线性映射。

(3) 残差连接 (Residual Connection) 和层归一化 (Layer Normalization): 这些组件有助于减轻梯度消失问题, 使模型更容易训练。

RoBERTa 模型在预训练阶段通过多层 Transformer 编码器堆叠, 学习文本的双向表示。然后, 在微调阶段, 模型通过添加一个分类层来适应特定任务。分类层的公式如下:

$$P(y | x) = \text{Softmax}(\mathbf{W}_o h_o) \quad (2)$$

其中, $P(y | x)$ 表示给定输入 x 预测标签 y 的概率分布; \mathbf{W}_o 表示分类层的权重矩阵; h_o 表示模型的输出表示。

3.1 模型微调

BERT 模型在微调阶段针对细粒度情感分析任务进行适应。微调的关键是调整分类层的参数, 以便将模型的输出映射到情感标签。微调的损失函数采用交叉熵损失函数, 公式如下:

$$L(\theta) = - \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(P(y_j | x_i)) \quad (3)$$

其中, $L(\theta)$ 表示损失函数; N 表示训练样本数量; C 表示情感类别数; y_{ij} 表示样本 i 的真实标签; $P(y_j | x_i)$ 表示模型的预测概率。

模型的参数 θ 将通过反向传播算法进行优化, 以最小化损失函数。采用梯度下降法进行参数更新。

3.2 数据增强

数据增强通过对原始数据进行变换或扩充来生成更多的训练样本, 以帮助模型更好地泛化到不同的情感表达方式。在本文细粒度情感分析任务中, 数据增强采用以下方法。

(1) 同义词替换: 对于输入评论中的一些关键词或短语, 可以将其替换为同义词或相似词汇。这有助于模型更好地理解不同方式表达相同情感的评论。

(2) 随机插入: 在评论中随机插入一些新词汇或短语, 以模拟用户的表达方式的多样性。这有助于模型更好地处理未见过的情感表达。

(3) 随机删除: 随机删除评论中的一些词汇或短语, 则可使模型更具鲁棒性, 能够处理信息不完整

的评论。

(4) 句法变换: 对评论进行句法结构的变换, 如句子重新排列、短语重组等, 以增加数据的多样性。

(5) 生成式方法: 使用生成式模型 (如 GAN 或 Seq2Seq 模型) 来生成新的评论样本, 以丰富数据集。

数据增强的目标是增加训练数据的多样性, 从而提高模型对不同情感表达方式的识别能力。

3.3 多任务学习

多任务学习是一种技术, 通过将细粒度情感分析任务与其他相关任务结合, 使模型能够更好地理解和处理文本。在这里, 将细粒度情感分析任务与情感极性分类任务和领域分类任务相结合。

(1) 情感极性分类任务: 情感极性分类是情感分析的一个常见子任务, 可将文本分类为积极、中性或消极。通过与细粒度情感分析任务结合, 模型可以更好地理解情感的整体情感趋势, 从而更准确地预测细粒度情感。

(2) 领域分类任务: 领域分类任务旨在将文本分类为不同的领域或主题。通过与细粒度情感分析任务结合, 模型可以更好地理解评论的背景信息, 从而更准确地识别情感。

多任务学习的关键是共享模型的层, 以便同时处理多个任务。在训练时, 可以使用多任务学习的损失函数, 同时考虑多个任务的损失, 就可优化共享模型的参数。这有助于模型在不同任务之间共享知识和特征, 提高了模型的性能。本文将细粒度情感分析任务与情感极性分类任务相结合, 以提高模型的性能。同时, 本文还引入了领域分类任务, 帮助模型更好地理解评论的背景信息。

总之, 数据增强和多任务学习是提高基于 RoBERTa 模型的细粒度情感分析性能的关键技术。可以增加数据多样性, 提高模型的泛化能力, 同时帮助模型更好地理解文本和情感的关系。结合这 2 个技术可以显著提升模型的性能。

4 评价指标

评价指标是判断训练出来的模型好坏的标准, 将测试数据集送入训练好的模型中进行测试, 按照评价指标高低确定模型参数的好坏。自然语言处理任务作为人工智能领域的一个重要研究方向, 如果采用单一衡量指标会存在片面的情况。因此, 本节主要介绍论文研究中所采用的评价指标, 即在测试模型性能时用到的评价指标准确率 (Precision) 和 F1 值 (F1-score), 通过表 2 的混淆矩阵进行计算。

表2 混淆矩阵
Table 2 Confusion matrix

类别	预测正例	预测负例
真正例	TP	TN
真实负例	FP	FN

(1) 准确度 (*Precision*): 准确度是最常用的模型性能评估指标, 表示模型在所有样本中正确分类的比例。准确度可以用于衡量模型的整体分类能力。具体计算公式为

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

(2) 召回率 (*Recall*): 用于衡量模型对正例样本的识别能力, 表示模型正确预测的正例样本数量占实际正例样本数量的比例。召回率较高, 表明模型能够较好地捕捉到真正例。具体计算公式为

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

(3) *F1* 分数 (*F1 - Score*): *F1* 分数是准确度和召回率的综合指标, 综合考虑了模型的预测精准性和召回率。*F1* 分数通常会被用作二分类任务中模型效果的重要评估指标。具体计算公式为

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (6)$$

5 实验结果

本文在 Semeval2014 数据集上进行了一系列实验, 并与传统方法进行了比较。实验对比结果见表3。实验结果表明, 本文的模型在细粒度情感分析任务中表现出色, 相比于传统方法, 准确率、召回率和 *F1* 值都有明显提升。

表3 实验结果对比

Table 3 Comparison of experimental results

Model	<i>Precision</i>	<i>Recall</i>	<i>F1</i>
RoBERTa	0.831	0.771	0.800
BERT	0.808	0.758	0.782
CNN	0.765	0.685	0.723
LSTM	0.749	0.699	0.723
TF-IDF	0.712	0.652	0.680
CRF	0.685	0.635	0.659

从表3中可以看出, RoBERTa 模型在细粒度情感分析任务上的表现最好, 其次是 BERT 模型, 而 CRF 模型的表现最差。这说明了基于预训练语言模型的方法在捕捉文本的语义和情感方面有着明显的优势, 而传统的特征提取和序列标注方法则相对较弱。

CNN 和 LSTM 对于 RoBERTa 模型表现较差, 其

准确度、召回率和 *F1* 分数都较低。这可能是因为受到了序列长度和结构的限制, 难以捕捉文本中的长距离依赖关系, 即使得识别评论中的复杂语义的效果较差。

6 结束语

本论文探讨了基于 RoBERTa 模型的在线用户评论细粒度情感分析, 并引入了数据增强、多任务学习和注意力机制等技术来提升模型性能。实验结果表明, 本文的方法在细粒度情感分析任务中表现出色, 有望在实际应用中提供更好的情感分析。

但在实际的在线评论中, 消费者对商家产品及相应服务的评论反馈不仅体现在在线文字评论上, 还包括商家评分、图片评论等。未来的研究可以继续探讨多模态综合分析, 以进一步提升细粒度情感分析任务的效果。

参考文献

- [1] 王安宁, 张强, 彭张林, 等. 在线评论的行为影响与价值应用研究综述[J]. 中国管理科学, 2021, 29(12): 191-202.
- [2] MUDAMBI S M, SCHUFF D. What makes a helpful online review? A study of customer reviews on Amazon. Com [J]. MIS Quarterly, 2010, 34(1): 185-200.
- [3] MUNIKAR M, SHAKYA S, SHRESTHA A. Fine-grained sentiment classification using BERT [C]//Proceedings of 2019 Artificial Intelligence for Transforming Business and Society (AITB). Kathmandu, Nepal: IEEE, 2019: 1-5.
- [4] TAN Yejin, GUO Wangshu, HE Jiawei, et al. A Fine-grained sentiment analysis method based on dependency tree and graph attention network [J]. Journal of Physics: Conference Series, 2020, 1651(1): 21-23.
- [5] BALIKAS G, MOURA S, AMINI M R. Multitask learning for fine-grained twitter sentiment analysis [C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Tokyo, Japan: ACM, 2017: 1005-1008.
- [6] 侯艳辉, 董慧芳, 郝敏, 等. 基于本体特征的影评细粒度情感分类[J]. 计算机应用, 2020, 40(4): 1074-1078.
- [7] 曹雪. 多粒度的汉语情感极性分类方法研究[D]. 哈尔滨: 黑龙江大学, 2020.
- [8] 万岩, 杜振中. 融合情感词典和语义规则的微博评论细粒度情感分析[J]. 情报探索, 2020(11): 34-41.
- [9] DONG Li, WEI Furu, TAN Chuanqi, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification [C]//Proceedings of the Annual Meeting of the Association for Computational Linguistics. Baltimore, USA: ACL, 2014: 49-54.
- [10] CHE Wanxiang, ZHAO Yanyan, GUO Honglei, et al. Sentence compression for aspect-based sentiment analysis [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(12): 2111-2124.