

# 方面级情感分析的研究与应用

蔡佳志, 冯 翱, 张 举

成都信息工程大学, 计算机学院, 四川 成都

收稿日期: 2022年7月20日; 录用日期: 2022年8月19日; 发布日期: 2022年8月24日

## 摘 要

在早期的细粒度情感分析任务上, 主要由人工进行特征选择, 人为的工作量大, 传统机器学习模型如支持向量机、朴素贝叶斯等被用于情感分类, 但随着近几年神经网络的兴起, 多种神经网络代替人为的工作, 实验效率大大地提升, 因为需要挖掘人们文本表达的观点的信息, 本文就多种模型应用于方面级情感分析进行研究和分析, 得出其中方面词和上下文之间的语义关系影响重大, 针对该方面对模型进行应用, 得到效果最好的模型。

## 关键词

细粒度情感分析, 特征选择, 情感分类, 神经网络

# Research and Application of Aspect-Based Sentiment Analysis

Jiazhi Cai, Ao Feng, Ju Zhang

School of Computer Science, Chengdu University of Information Technology, Chengdu Sichuan

Received: Jul. 20<sup>th</sup>, 2022; accepted: Aug. 19<sup>th</sup>, 2022; published: Aug. 24<sup>th</sup>, 2022

## Abstract

In the early tasks of Aspect-based Sentiment Analysis, the feature selection was mainly carried out manually with a high human workload, and traditional machine learning models such as support vector machines and Naive Bayes model were used for sentiment classification. However, with the rise of neural networks in recent years, a variety of neural networks have replaced human work, and experimental efficiency has been greatly improved. Because of the need to tap into information about the views expressed in people's texts, this paper studies and analyzes the application of various models in Aspect-based Sentiment Analysis, concluding that the semantic relationship between aspect words and context has a great influence, and that the models are applied to that

aspect to obtain the best results.

## Keywords

Aspect-Based Sentiment Analysis, Feature Selection, Sentiment Classification, Neural Networks

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

现在的互联网中，信息量大，文本信息丰富，如电商平台的评论文本，直播平台的评论文本，电影网站的评论等，都可以充分反映用户对所使用产品或所观看电影等的好感程度，我们可以利用某种手段挖掘出这些含有丰富信息的文本，进而根据用户的喜好对产品、服务等进行调整，让更多的用户使用该产品或服务，从而使商家的利益达到最大化。

情感分析[1]的含义是通过文本挖掘出用户对产品、服务等意见，获得对产品和服务等的好感度，情感分析根据研究对象的粒度的不同可分为三类：文档级、句子级和对象级，其中细粒度情感分析[2]为对象级情感分析，剩余的两项类别为粗粒度情感分析，细粒度情感分析是输入一段文本和需要分析出情感极性的对象，输出的是该对象的情感极性，因为粗粒度情感分析，不能更细致的分析出句子中对象的情感极性，无法从中获取更多有用的信息，无法提取出人们对特定对象的看法，只能笼统地得出整个句子的情感极性，目前研究的大多是细粒度情感分析。

比如“这里的景色很美，但这里的寿司却不好吃”。用户对目标“景色”的看法是正面的，而对目标“寿司”的看法却是负面的。所以一个句子可能多个方面词，而这多个方面词所带的情感极性是不同的，因为对应的对象不同。细粒度情感分析研究得更加细致，也更加地困难，在现有的方法中，一般是将方面信息加入到模型中来提取方面对应的情感极性，在早些时候，人们使用的传统机器学习模型，有支持向量机和朴素贝叶斯等，再往后可以运用的模型有 CNN [3]和 LSTM，但在 BERT 提出以后，这些模型就渐渐落后了，一般是直接用 BERT 或者将 BERT 和其他模型融合进行方面级情感分析，现有的情感分析对于方面词和上下文的语义关系处理不清楚，从而造成方面词情感极性的错误判断，我们在传统机器学习和神经网络方法上进行对比试验，通过门控机制、构造句子对任务和增强方面词的编码等方法，在 BERT 等模型的基础上进行实验，得到效果最好的模型。

## 2. 相关工作

### 2.1. 传统机器学习方法

情感分析的传统方法大致可以分为基于词典的方法和基于各种传统机器学习的方法[4]。

基于词典的情感分析的准确率极大地取决于词典的质量，词典主要包括电影、医疗等，并且词典需要人为的维护，词典的建立也需要人为的操作，基于词典的细粒度情感分析，要考虑句子中需要处理情感极性的对象，方便处理出对象的情感极性，进而进行情感分析，基于词典的情感分析，因为需要的人力巨大，现阶段该方法已经慢慢的被淘汰了。

基于各种机器学习的情感分析，主要使用监督学习方法，常用的分类方法主要有支持向量机(SVM)，朴素贝叶斯(NB)等，其主要困难为分类器的设计、数据的获取和一些未见过的短语的解释，基于机器学

习的情感需要人为的进行构建特征，不同的数据和不同的构造特征的方法造成的准确率的差别比较大，且基于机器学习的细粒度情感分析，需要考虑句子中的目标对象，可以生成依赖对象[5]的特征，也可以确定句子中对象的作用范围，确定其在那个范围内的情感极性，但是传统方法无法表示对象词语与其上下文的关系。

## 2.2. 深度学习方法

Word2vec [6]和 Glove [7]都是词嵌入模型，Word2vec 可以分为两个可实现的模型：CBOW 和 Skip-gram，CBOW 是通过上下文和周围词来预测当前值，比如“easyai 是()人工智能网站”，通过其他词来预测括号内的词，而 Skip-gram 是通过当前词来预测上下文的词，比如“()最好的()”，通过中心词预测其上下括号内的词。在 Word2vec 被提出不久，Glove 就被提出，作为 Skip-gram 模型的改良。早期的深度学习方法有 CNN、RNN 和 LSTM 等，随着 ELMO [8]和 BERT [9]的相继提出，BERT 在很多数据集上展现出了良好的性能，逐渐被越来越多的人使用，基于 BERT 的模型被越来越多的人所推广。

## 3. 模型

本文就传统方法和基于深度学习的方法应用于方面级情感分析，进行研究和对比。

### 3.1. 基于词典的方法、SVM 和 Bayes

基于词典和句法结构的方法是有一个个词典，词典里面含有服务等方面的词，输入一个句子，句子中的每个词分离开，按照词语的得分，对句子的情感度进行打分，SVM 和 Bayes 是分别直接用的 sklearn 里的 SVM 和 Multinomial NB 库。

### 3.2. E2E-ABSA

如下图 1，使用具有  $L$  个 Transformer 层的 BERT 组件来计算相应的上下文表示：

$$H_L = \{h_{-}(L_1), \dots, h_{-}(L_T)\} \quad (1)$$

其中  $H_L$  为相应的上下文表示。

将上下文的表示输入到任务特定层，预测标签序列：

$$y = \{y_1, \dots, y_T\} \quad (2)$$

其中  $y$  为预测的序列标签，相应的上下文表示经过 E2E-ABSA 层后，生成对应的预测的序列标签。

研究了几种不同的 E2E-ABSA [10]层的设计，即线性层、递归神经网络、自注意网络和条件随机场。

### 3.3. TC-LSTM 和 TD-LSTM

LSTM [11]是将整个句子输入到 LSTM 模型中，而 TD-LSTM [12]相比于 LSTM 的改动有，可以从目标词两边进行 softmax 大大提升了运行效率，左边的部分是左句子 + 方面词，右边的部分是方面词 + 句子，左边的部分输入到左边的 LSTM，右边的部分输入到右边的部分，更能体现目标词相关的语义信息，提升模型的准确率。TD-LSTM 的结构图如图 2 所示。

TC-LSTM 是将目标词的相关语义信息与上下文相结合，将  $v_{target}$  平均信息与每个词进行拼接。TC-LSTM 的结构图如图 3 所示。

### 3.4. BERT\_SPC 和 TD-BERT

BERT\_SPC [13]是将方面实体和上下文转换为句子对分类任务，在 BERT 的下游进行输入，方便进行

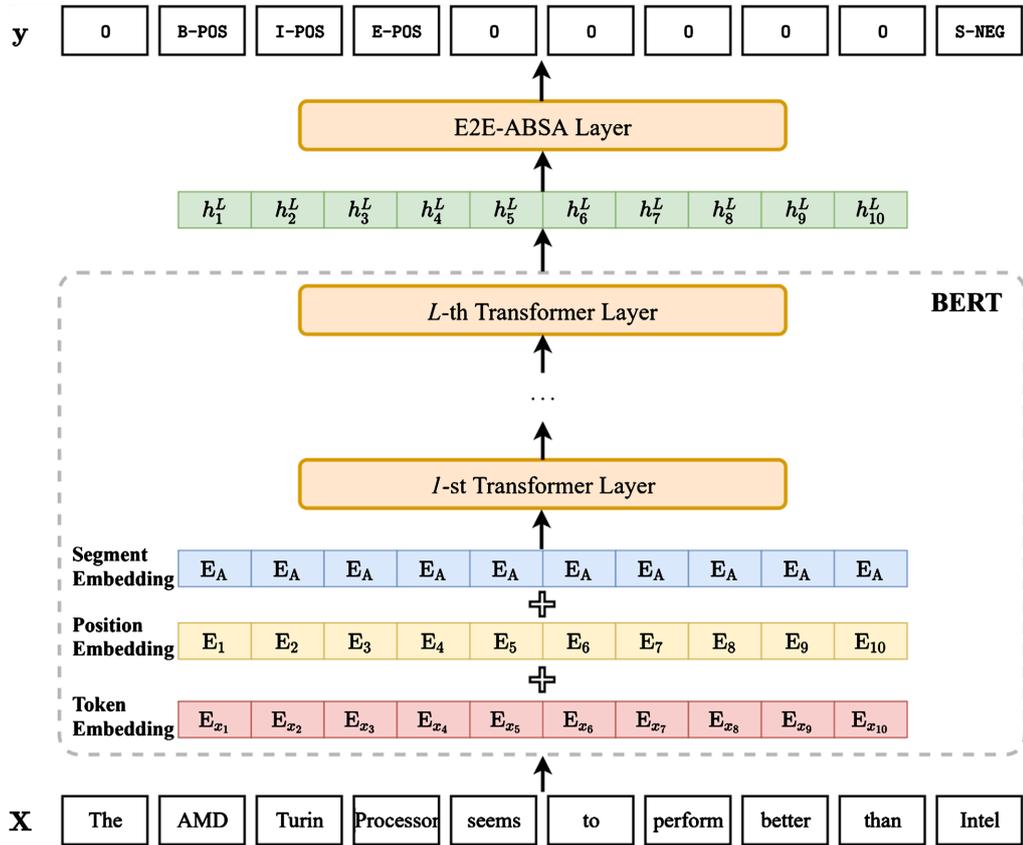


Figure 1. Overview of the designed model

图 1. 设计模型概述

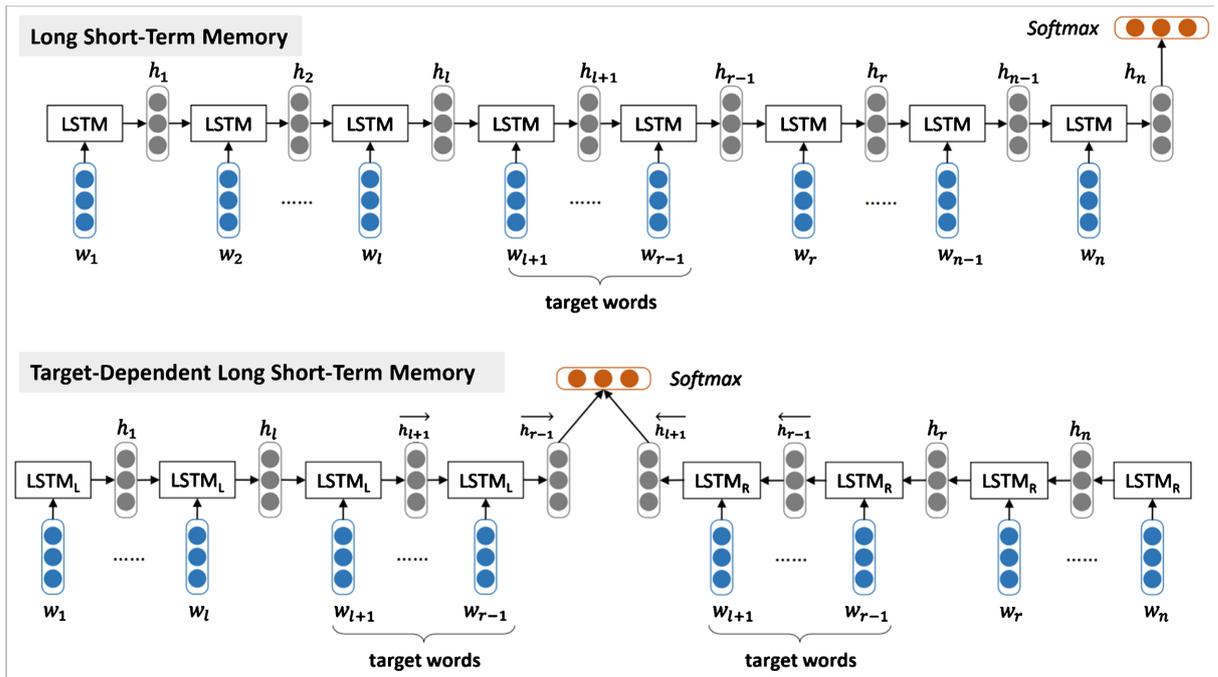


Figure 2. TD-LSTM

图 2. TD-LSTM

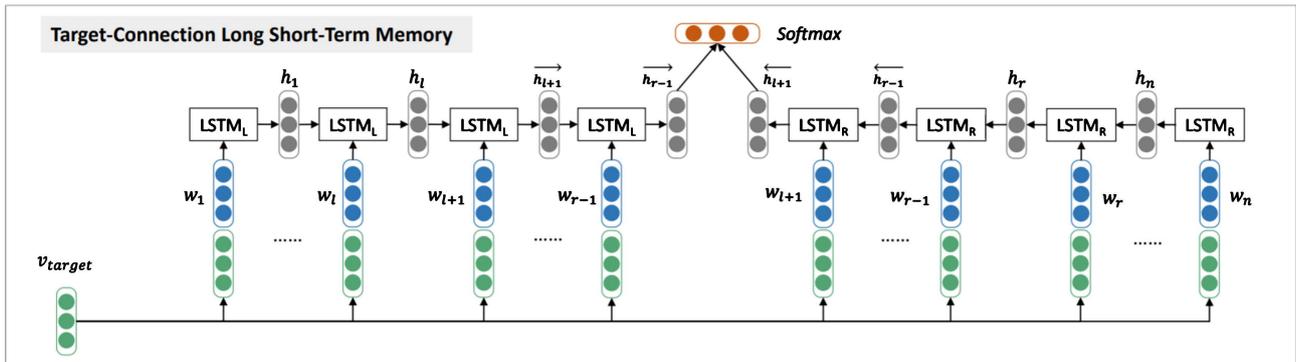


Figure 3. TC-LSTM

图 3. TC-LSTM

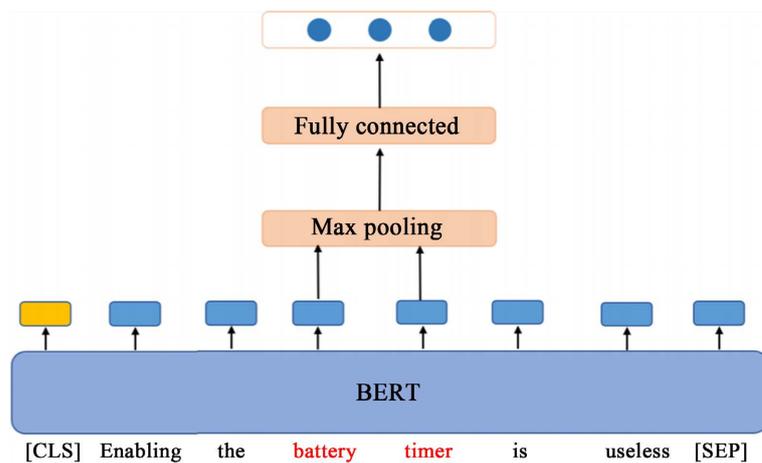


Figure 4. TD-BERT

图 4. TD-BERT

方面实体的情感极性的判断。

TD-BERT [14]将 CLS 输入顺序进行了改变，句子不再是从 CLS 开始输入，而是直接将方面词输入到 Max pooling 层，减少了相关工作量。TD-BERT 结构图如图 4 所示。

### 3.5. GCN

GCAE [15]在 ATSA 上的应用是在 CNN 的基础上加了一个门控机制，门控机制更方便地输出方面对象的情感极性，且是基于 CNN 的模型，可以并行化执行操作，对于 ATSA 上的应用的话，模型多出了一个包含多个方面词的卷积层，更能体现方面词的情感极性。GCN 应用在 ATSA 中的结构图如图 5 所示。

### 3.6. BERT-pair 和 BERT-single

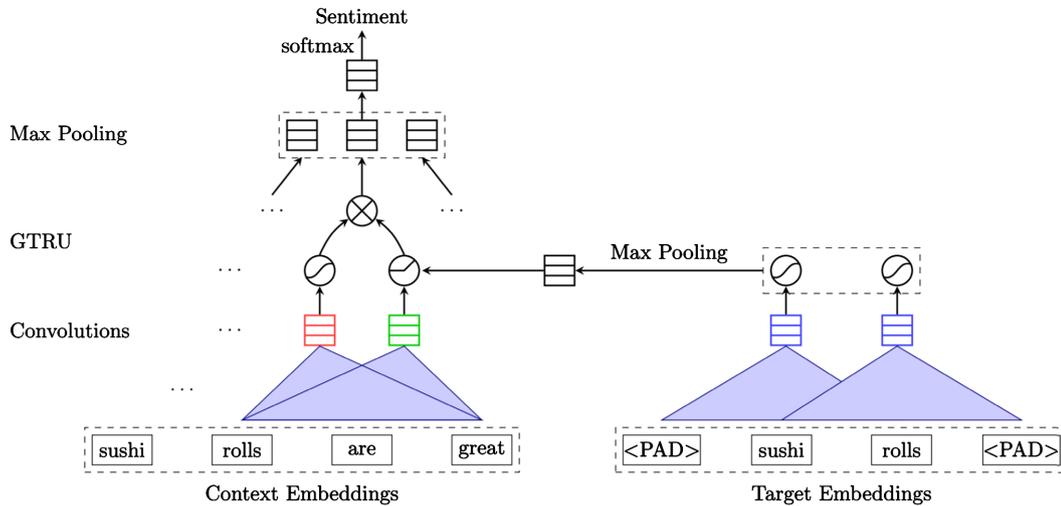
BERT-single 输入的是一个句子，对一个句子进行分类任务。

BERT-pair [16]是应用 4 个任务将 ABSA 任务转换为句子对分类任务，分别有如下 4 个句子对分类任务：

QA\_M: 目标方面对生成的句子是一个伪句。

QA\_B: 会添加标签信息，每个目标词方面对会产生三个句子，选择分数最高的句子。

NLI\_M: 创建的辅助句子是简单的伪句。



**Figure 5.** Application of model GCAE for ATSA task.  
**图 5.** GCAE 在 ATSA 任务的应用

NLI\_B: 辅助句子从疑问句变成了简单的伪句。

## 4. 实验分析

### 4.1. 数据集和评价指标

数据集为 SemEval 2014 Task4。

国际语义评测大会(SemEval)从 2014 年至今，每年均会发布相应的细粒度情感分类任务，并公布数据集。不过现目前大家使用得最多的还是 SemEval-2014 年任务 4 公布的数据集，这或许是为了更加方便进行实验对比。

在评价指标方面，均采用准确率(Accuracy)和 F1 作为衡量的指标。

$$Acc = (TP + TN) / (TP + TN + FP + FN) \tag{3}$$

$$P = TP / (TP + FP) \tag{4}$$

$$R = TP / (TP + FN) \tag{5}$$

$$F_1 = 2 \times P \times R / (P + R) \tag{6}$$

其中上述公式中  $TP$  表示正确的正样本， $FP$  表示错误的正样本， $TN$  表示正确的负样本， $FN$  表示错误的负样本， $Acc$  为模型准确率， $P$  为查准率， $R$  为查全率， $F_1$  为查准率和查全率的调和均值。

### 4.2. 实验结果

**Table 1.** Traditional machine learning methods  
**表 1.** 传统机器学习方法

Models	accuracy
SVM	85.6%
Bayes	85.8%

表 1 为传统机器学习的方法做的情感分析的准确率，SVM 是用的 sklearn 里的 svm 包，Bayes 是用的

sklearn 下 naive\_bayes 里的 MultinomialNB 包，可看出 SVM 和 Bayes 的准确率相差无几，而基于词典和句法结构<sup>1</sup>是根据互联网上的，准确率没有 SVM 和 Bayes 高。

**Table 2.** Deep learning methods

**表 2.** 深度学习方法

Models	Restaurant (Accuracy)	Laptop (Accuracy)
GCAE	74.51%	64.37%
BERT_SPC	85.27%	79.47%
LSTM	76.43%	66.61%
TD_LSTM	75.98%	67.87%
TC_LSTM	75.00%	65.20%
TD_BERT	86.16%	77.74%

表 2 中准确度最高为 TD\_BERT，最低为 GCAE，相关分析见实验分析。

**Table 3.** The results of BERT-pair

**表 3.** BERT 句子对分类任务结果

Models	F1	4 way (Accuracy)	3 way (Accuracy)	Binary (Accuracy)
BERT-single	91.06%	82.92%	86.12%	92.83%
BERT-pair QA_M	91.66%	84.19%	88.48%	93.74%
BERT-pair NLI_M	92.06%	84.39%	87.66%	93.85%
BERT-pair QA_B	92.35%	85.36%	89.41%	94.31%
BERT-pair NLI_B	91.56%	84.4%	88.0%	94.7%

表 3 中，F1 和准确率最高为 BERT-pair-QA\_B，最低为 BERT-single，相关分析见实验分析。

**Table 4.** The results of E2E BERT

**表 4.** E2E BERT 结果

Models	Precision	Recall	F1
BERT-linear	74.26%	73.92%	74.08%
BERT-tfm	71.11%	74.28%	72.65%
BERT-crf	71.58%	71.51%	71.54%
BERT-gru	71.10%	72.49%	71.78%
BERT-san	72.98%	74.55%	73.75%

表 4 中，F1 最高为 BERT-linear，最低为 BERT-crf，相关分析见实验分析。

### 4.3. 实验分析

实验对比表明 BERT 所做出的实验准确率远远高于其他预训练模型的准确率，BERT-SPC 和基于 BERT 的模型效果相差不多，说明改 BERT 框架时，只能起到少许作用。

<sup>1</sup>[https://github.com/hellonlp/sentiment\\_analysis\\_dict](https://github.com/hellonlp/sentiment_analysis_dict).

传统方法与基于神经网络的方法相比,传统方法所消耗的人力更多,有许多步骤需要人为的操作,而基于深度学习的方法,更加的方便,人为的操作更少且适合应用的场景更多,拥有更为强大的语言表征能力和特征提取能力[17]。

LSTM 只是简单地将整个句子作为其输入,无法更准确地表示方面词与其上文的联系,无法正确地表示方面词的情感极性,一个句子中可能有多个方面词,不同方面词可能有不同地情感极性,不同方面词与其上下文地联系可能不同,从而表达的语义信息不同,为了能够更准确地表示方面词地情感极性,双向 LSTM 应运而生,TC-LSTM 和 TD-LSTM 也取得了比 LSTM 高的准确率。

因为 BERT 拥有更为复杂的网络结构,计算操作更多,所以 BERT 的准确率要比 TD-LSTM 和 TD-LSTM 高,而且 BERT 可以实现并行计算操作。

对于 ATSA 上的应用的话,GCN 模型多出了一个包含多个方面词的卷积层,更能体现方面词的情感极性。因为是基于 CNN 的模型,所以可以并行计算,所以运行时间小于 TD-LSTM。

BERT-pair、BERT E2E-ABSA 和 TD-BERT 这三者比较可以看出在输入端做出改变的 BERT-pair 和 TD-BERT 比 BERT E2E-ABSA 的效果好,因为后者是在输出前加了一个 E2E 任务特定层进行预测,无法更能体现上下文与方面词之间的语义关系,其是没有微调 BERT,而是直接在 BERT 后添加了下游任务。

将 BERT 构造成句子对任务更能体现方面词与其上下文的语义关系,所以 BERT-SPC 和 BERT-pair 比基础的 BERT 效果要好,而 TD-BERT 改变了输入的形式,一般是从 CLS 开始输入,而 TD-BERT 跳过 CLS 直接将方面词输入到最大池化层,增强方面词的编码,更能体现方面词的相关语义信息,所以 TD-BERT 的效果最好,因此最能体现方面词和上下文语义之间的关系的模型效果是最好的,效果最好的为 BERT-pair 和 TD-BERT。

## 5. 结束语

细粒度情感分析与普通的情感分析相比,细粒度情感分析主要是添加了方面属性、方面类别和对于方面词的情感极性本文主要对在方面级情感分析上应用的几种模型进行了对比,发现传统的基于词典的方法和各种基于机器学习的方法所耗费的人力非常多,运用深度学习的方法可以大大减少人力的使用,在深度学习方法中,发现 BERT 展现出良好的准确率,基于 BERT 的模型也有较好的表现,但能否对这些模型进行改进还需要后续研究的推展,后期将会进一步探索如何在这些模型上进行融合改进[18]以获得更好的效果,在对特征进行不同的处理方面需要进行更多的研究,可以基于 BERT 并融合其它模型的优点进行创新,可以对 BERT 的输入和输出部分进行修改,对特征的提取和识别方面进行改良,对于句子中的构造成分很复杂的情况,怎么更能体现句子中多个方面词与上下文间的语义信息,成为我们要解决的问题。在数据集方面,大多用的是英文数据集,后期会在中文数据集上进行更多的实验。

## 参考文献

- [1] 王恒,唐孝国,郭俊亮. 基于电商评价的文本情感分析研究与应用[J]. 黑龙江科学, 2022, 13(12): 29-31.
- [2] 王海燕,陶皖,余玲艳,王鸣鹃. 文本细粒度情感分析综述[J]. 河南科技学院学报(自然科学版), 2021, 49(4): 67-76.
- [3] 马广兴,曲波,常琛,卞浩然. 基于 CNN 的供热管道泄漏识别方法研究[J/OL]. 电子测量技术: 1-9. <http://kns.cnki.net/kcms/detail/11.2175.TN.20220726.0942.006.html>, 2022-07-27.
- [4] 孙开华,代余杰,封晴. 基于机器学习的大数据分析和处理[J]. 中国新通信, 2021, 23(13): 65-66.
- [5] Casado-Mansilla, D., Moschos, I., Kamara-Esteban, O., Tsolakis, A.C., Borges, C.E., Krinidis, S., *et al.* (2018) A Human-Centric & Context-Aware IoT Framework for Enhancing Energy Efficiency in Buildings of Public Use. *IEEE Access*, 6, 31444-31456. <https://doi.org/10.1109/ACCESS.2018.2837141>

- 
- [6] Jatnika, D., Bijaksana, M.A. and Suryani, A.A. (2019) Word2Vec Model Analysis for Semantic Similarities in English Words. *Procedia Computer Science*, **157**, 160-167. <https://doi.org/10.1016/j.procs.2019.08.153>
- [7] Pennington, J., Socher, R. and Manning, C. (2014) Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, 25-29 October 2014, 1532-1543. <https://doi.org/10.3115/v1/D14-1162>
- [8] Liu, B. (2012) Sentiment Analysis and Opinion Mining. Vol. 5, Springer, Cham, 1-167. <https://doi.org/10.1007/978-3-031-02145-9>
- [9] Pontiki, M. and Pavlopoulos, J. (2014) SemEval-2014 Task 4: Aspect Based Sentiment Analysis. *Proceedings of International Workshop on Semantic Evaluation*, Dublin, 23-29 August 2014, 27-35. <https://doi.org/10.3115/v1/S14-2004>
- [10] Li, X., Bing, L., Zhang, W. and Lam, W. (2019) Exploiting BERT for End-to-End Aspect-based Sentiment Analysis. Association for Computational Linguistics. *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, Hong Kong (China), 4 November 2019, 34-41. <https://doi.org/10.18653/v1/D19-5505>
- [11] 韩雨恒. 基于 LSTM 神经网络的轨道预报算法研究[J]. 科学技术创新, 2022(21): 88-91.
- [12] Tang, D., Qin, B., Feng, X. and Liu, T. (2016) Effective LSTMs for Target-Dependent Sentiment Classification. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, 11-16 December 2016, 3298-3307.
- [13] Song, Y., Wang, J., Jiang, T., Liu, Z. and Rao, Y. (2019) Attentional Encoder Network for Targeted Sentiment Classification. arXiv:1902.09314.
- [14] Gao, Z., Feng, A., Song, X. and Wu, X. (2019) Target-Dependent Sentiment Classification with BERT. *IEEE Access*, **7**, 154290-154299. <https://doi.org/10.1109/ACCESS.2019.2946594>
- [15] Xue, W. and Li, T. (2018) Aspect Based Sentiment Analysis with Gated Convolutional Networks. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Vol. 1, Melbourne, 15-20 July 2018, 2514-2523. <https://doi.org/10.18653/v1/P18-1234>
- [16] Sun, C., Huang, L. and Qiu, X. (2019) Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. arXiv:1903.09588.
- [17] 张亚琪. 基于 BERT 预训练模型的功能肽预测问题特征提取算法研究[D]: [硕士学位论文]. 长春: 吉林大学, 2022. <https://doi.org/10.27162/d.cnki.gjlin.2022.002934>
- [18] 李铁飞, 生龙, 吴迪. BERT-TECNN 模型的文本分类方法研究[J]. 计算机工程与应用, 2021, 57(18): 186-193.