

# 基于多任务学习注意力交互模型的方面级情感分析

叶洲铭, 王 瑛, 王 勇

广东工业大学, 计算机学院, 广东 广州

收稿日期: 2021年12月3日; 录用日期: 2021年12月30日; 发布日期: 2022年1月6日

---

## 摘 要

基于方面级的情感分析, 其现有算法大多是将方面词提取和情感极性分类分开独立处理, 或是只执行其中一个任务, 忽略两个子任务间潜在的关联; 或是使用流水线方法分为两个阶段分别处理, 导致系统误差传播以及复杂化。本文将两个任务统一定义为序列标注问题, 先使用Bi-LSTM和CNN提取特征, 其次使用基于多头注意力的特征交互学习机制对特征进一步筛选提取; 并且使用一个基于注意力交互机制, 获得任务间潜在关联特征。根据实验表明, 所提出的算法在3个公开数据集上, 在情感极性分类以及方面词提取上, 比当前代表模型的准确率有所提高。

## 关键词

方面级情感分析, 多头注意力机制, 多任务模型

---

# Aspect-Level Sentiment Analysis Based on Multi-Task Learning Attention Interaction Model

Zhouming Ye, Ying Wang, Yong Wang

School of Computer, Guangdong University of Technology, Guangzhou Guangdong

Received: Dec. 3<sup>rd</sup>, 2021; accepted: Dec. 30<sup>th</sup>, 2021; published: Jan. 6<sup>th</sup>, 2022

---

## Abstract

Based on aspect sentiment analysis, most of its existing algorithms separate aspect word extraction and sentiment polarity classification separately, or perform only one of the tasks, ignoring the

potential correlation between the two subtasks; or use pipeline method to analyze. They are processed separately for two stages, leading to system error propagation and complexity. This paper defines the two tasks as a sequence labeling problem. First, Bi-LSTM and CNN are used to extract features, and then the feature interactive learning mechanism based on multi-head attention is used to further filter and extract features; and an attention-based interaction mechanism is used to obtain the task potentially related features. According to experiments, the proposed algorithm is more accurate than the current representative model in terms of emotion polarity classification and aspect word extraction on three public data sets.

## Keywords

Aspect-Level Sentiment Analysis, Multi-Head Attention Mechanism, Multi-Task Model

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

方面级情感分析(Asspect Based Sentiment Analysis, ABSA), 主要是从带有主观情绪色彩的自然文本信息中, 分析、处理、归纳和研究用户对一个方面的情绪或者看法; 旨在对文本如用户评论进行细粒度的情感分析[1]。情感分析, 长期一直是自然语言处理的重点课题。随着研究深入, 情感分析的文本研究粒度可以细分为篇章级、句子级以及方面级, 其中方面级情感分析能够分析句子中所提及或者隐含的不同方面词的情感极性, 已经成为该领域的热门研究方向之一[2]。并且, 随着互联网大数据技术以及社交媒体的迅猛发展, 利用情感分析技术处理社交媒体以及网络平台的评论、博客、留言等文本, 获取其中的情感信息, 进而为不同实际应用场景, 诸如舆情管理[3]、商业决策、用户个性化推荐[4]等提供技术支持。

方面级情感分析, 主要包含两个子任务: 方面术语提取(Asspect term Extraction, AE)和方面级情感分类(Asspect-level Sentiment Classification, AS)。AE 任务主要目的是从给定文本中提取提及的方面术语; AS 任务则是对相关方面词进行情感极性判断。例如, “Coffee is a better deal than overpriced cosi sandwiches”, 其中 “Coffee” 和 “cosi sandwiches” 两个方面术语的情感极性分别为积极和消极。

## 2. 相关工作

传统的文本情感分析, 分别是基于情感词典和基于机器学习, 深度学习避免了大量人工标注成本以及复杂的特征工程, 从而在自然语言领域得到广泛应用。各种类型神经网络应用于方面级情感分析, 获得不错效果。研究者, 从不同方面提高深度学习在方面级情感分析的性能表现。比如: 常见的一词多义的问题, 随着 BERT [5]模型和 Elmo 模型等语言模型提出和应用, 有效解决该类问题, 因为其不同于 Glove, 可以在词嵌入过程捕捉上下文信息, 产生融合上下文信息的词嵌入。Song [6]等人提出 BERT-SPC 变体模型适用于细粒度情感分析, 获得高性能表现; Zeng [7]等人, 提出的 LCF-BERT 模型通过结合自注意力机制实现一种特征级的局部上下文聚焦机制, 在方面级情感分析获得较高性能提升。常用的神经网络提取单元, 多是长短期记忆网络(Long Short Term Memory, LSTM), 但是简单的 LSTM 容易忽略方面词的建构, Tang 等人[8]提出的 TD-LSTM 模型使用两个依赖方面词目标的 LSTM, 从而建构目标在句子上下文的语义表示。同时, 通过引入注意力机制, 提高针对句子中的方面词的聚焦能力, 从而提高情感分析的性能。彭祝亮等人[9]提出一种设置不同方面的注意力模块独立训练的 BLSTM-AAM 模型, 从而提高特征提取能

力，有效提高分类的精度。

目前研究，将两个任务分开独立进行处理，而且目前大多数的研究侧重于情感极性分类任务的研究，目前研究思路多是通过获取方面项与上下文的关系提高分类精度。如果忽略方面词提取任务的研究，缺乏相应足够的方面词就导致情感分类陷入无目标的困境。所以最近研究者尝试同时解决两个任务的方案，其中基于多任务模型的方法，有利于提高模型的泛化能力。本文模型主要基于多任务模型，构建端到端的方面级情感分析模型，这是因为当使用深度网络训练时，多任务的联合学习方法比集成方法更具有可行性，集成方法容易混淆不同子任务所学的特征。为了解决，不同任务间，因为相关性而存在的噪音，所导致的多任务负迁移问题，本文提出基于多头注意力机制的特征交互学习机制，通过筛选学习合适的特征。发现方面词与方面词的情感极性，是存在强联系的，而且二者的标注边界是一致的，所以使用所提出的基于注意力机制的交互机制，提高模型的精度。此外，使用 BERT 生成融合上下文信息的词嵌入表示。

### 3. 注意力交互多任务模型

本文提出的 Att-Inter-MT 模型主要由词嵌入层、特征提取层、语义交互提取层、信息交互层和输出层构成，模型总体架构如图 1 所示：

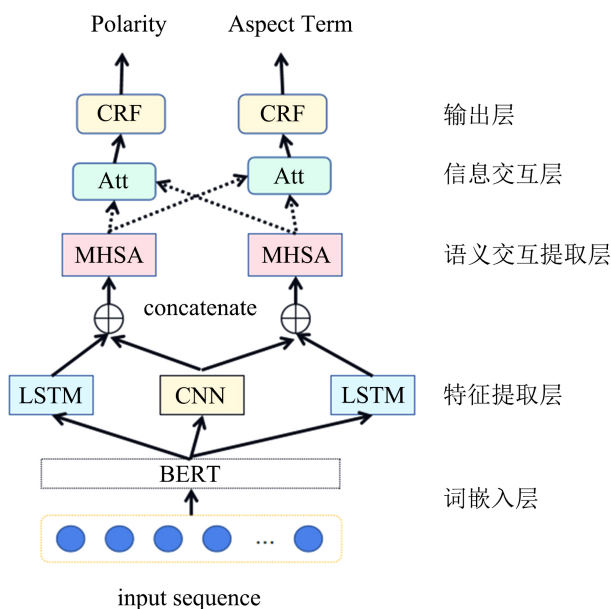


Figure 1. Overall architecture diagram of Att-Inter-MT model  
图 1. Att-Inter-MT 模型总体架构图

#### 3.1. 任务定义

将两个子任务定义为句子标注任务，AE 任务输出的标注标签为{B, I, O}，其中 B 标签指示方面词的首字，I 指示方面词的中间字段，O 指示的是其它字段；类似，AS 任务的输出标注标签为{PO, NT, NG, O} 分别表示情感极性中的积极、中性和消极，以及其它字段。

#### 3.2. 词嵌入层

BERT 模型，是一种基于双向 Transformer 构建的自编码预训练语言模型。常常通过微调实现与下游

任务模型的有效结合, 其能够获取上下文双向特征表示, 从而提取更为全面的特征, 适合处理高层语义信息提取的任务。本文使用 BERT 进行融合上下文语义的词嵌入预训练计算, BERT 词嵌入计算过程:

$$O_{\text{BERT}} = \text{BERT}(X) \quad (1)$$

### 3.3. 特征提取层

特征层使用两个多层双向 LSTM 分别构建两个子任务上下文依赖关系, LSTM 克服常规的 RNN 网络的长程遗忘问题, 而且缓解了 RNN 常见的梯度爆炸以及梯度消失等神经网络训练问题; 能够较好处理序列文本的较远距离间的依赖问题。AE 任务所得的特征表示为  $H^e = \{h_1^e, \dots, h_r^e\}$ ; AS 任务所得的特征表示为  $H^s = \{h_1^s, \dots, h_r^s\}$ 。同时, 使用一个多层不同卷积核大小的 CNN 网络, 获取不同尺寸大小的局部特征, 获得最佳的边界特征表示。因为极性分类以及方面词提取的标注输出的边界理论上是一致, 所以该特征作为多任务的共享特征空间。所以卷积层提取所得的特征表示为  $H^c = \{h_1^c, \dots, h_r^c\}$ 。

### 3.4. 语义交互提取层

该层主要针对特定特征空间和共享特征空间, 进行不同特征空间的融合计算, 获得对于子任务合适有效的语义表示。因为任务间的噪音存在, 这可能导致负迁移的现象, 从而限制模型的性能表现。有必要对公共特征空间进行筛选过滤, 使用注意力机制能够有效学得对应合适的特征表示。该层先是分别将子任务的 LSTM 和 CNN 所得的隐藏向量拼接; 然后, 通过全连接层进行维度调整。最后使用多头注意力机制; 其是基于放缩点积注意力机制(SDA), 进行注意力计算, 而且通过使用多头机制, 即是多次计算, 实现模型在不同多维度多层次的特征表示子空间计算, 能够学习到深层语义信息。其中 AS 任务的计算过程:

$$O^{sc} = [H^s; H^c] \quad (2)$$

$$O_{dense}^{sc} = w \times O^{sc} + b \quad (3)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) \times V \quad (4)$$

$$H_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (5)$$

$$\text{MHSA} = [H_1, \dots, H_h]W_h \quad (6)$$

$$O_{fil}^{sc} = \text{MHSA}(O_{dense}^{sc}) \quad (7)$$

其中  $H^s$  是 AS 任务的 LSTM 中间隐藏向量,  $H^c$  是 CNN 提取的公共特征空间的隐藏向量。  $d_k$  表示输入维度; 使用三个不同的参数矩阵  $W_i^Q$ 、 $W_i^K$ 、 $W_i^V$  进行线性变换, 其所得结果  $Q$ 、 $K$ 、 $V$  作为放缩点积注意力计算的输入; 进行  $h$  头计算,  $W_h$  参数矩阵对多头计算结果进行线性变换;  $O_{fil}^{sc}$  为语义交互提取的结果。

### 3.5. 信息交互层

两个任务间存在强关联, 不仅因为上下文融合方面项的语义表示, 能够有效提高情感极性判断的性能表现; 而且本文的任务定义为序列标注问题, 两个子任务的标注边界是要一致的。所以, 利用二者任务的中间隐藏向量, 进行交互计算, 以此获得相关联系。交互机制, 基于多头注意力机制, 分别将不同任务的中间隐藏向量作为输入, 进行不同特征空间的计算, 以求得二者输入相关性, 从而获得对应合适的语义表示。

AE, 方面词提取子任务, 将 AE 中间隐藏向量  $O_{fil}^{ec}$  作为注意力计算中的 Key 向量和 Value 向量, 将 AS 的中间隐藏向量  $O_{fil}^{sc}$  作为查询 Query 向量, 进行多头注意力计算:

$$O_{att}^e = \text{MHSA}(O_{fil}^{sc}, O_{fil}^{ec}, O_{fil}^{ec}) \quad (8)$$

AS, 方面级情感极性分类子任务, 将 AS 中间隐藏向量  $O_{fil}^{sc}$  作为注意力计算中的 Key 向量和 Value 向量, 将 AE 的中间隐藏向量  $O_{fil}^{ec}$  作为查询 Query 向量, 进行多头注意力计算:

$$O_{att}^s = \text{MHSA}(O_{fil}^{ec}, O_{fil}^{sc}, O_{fil}^{sc}) \quad (9)$$

### 3.6. 输出层

针对最终输出所生成的 AS 标签和 AE 标签, 既可以用全连接网络加 Softmax 函数求最大值得到标签, 也可以使用条件随机场(CRF)。使用线性链 CRF 作为输出的推理层, 其能够捕捉标签间的依赖, 进而使用依赖预测标签, 对标签序列进行建构, 能获得全局最优序列。比如: B 标签后续应该是标签 I, 而不是标签 O。本文模型以线性链 CRF 作为解码器, 两个子任务的输入分别为其信息交互层所得语义表示,  $O_{att}^s$  和  $O_{att}^e$ , CRF 通过其输入学习标签路径的概率分布。最后使用维特比算法生成 AS 标签和 AE 标签。CRF 的似然计算过程, 其中  $W$  为权重,  $b$  为偏置:

$$L(W, b) = \sum_i \log p(y|h; W, b) \quad (10)$$

## 4. 实验及结果分析

### 4.1. 实验环境及参数设置

本文实验环境如下: CPU 为 Intel Core i7 8700K, GPU 为 GeForce GTX 1080, 内存大小为 DDR4 16GB, 开发环境为 linux 64 位系统, pytorch 1.5.0。

实验参数设置: 词嵌入使用 BERT 预训练模型获得, 词嵌入向量维度 d<sub>bert</sub> 为 768; 特征提取层, LSTM 隐藏单元数为 200, 优化函数选择为 Adam, 实验学习率初始化为 0.001, CNN 模型使用多层堆叠卷积窗口, 卷积核尺寸大小设置为 2、3、4、5。实验引入 Dropout\_rate = 0.5, 提高模型泛化和减少过拟合风险。

### 4.2. 实验数据集及评价指标

数据集使用三个公共情感分析数据集, 分别是 SemEval 2014 任务的笔记本电脑、餐厅评论数据集, 和由 Dong 等人[10]收集的 ACL14 Twitter 数据集。该类数据集, 是主要应用于细粒度的情感分析任务中, 官方划分为训练集和测试集; 数据集含有多个方面词以及对应的情感极性(积极、中立和消极)。实验的数据集按照原来公开数据集的划分, 进行训练以及验证。表 1 所示为数据集中每个类别情感极性分布。

Table 1. Statistics of experimental data

表 1. 实验数据集的统计

Dataset	Positive		Neutral		Negative	
	Train	Test	Train	Test	Train	Test
Restaurant-14	2164	728	637	196	807	196
Laptop-14	994	341	464	169	870	128
Twitter	1561	173	3127	346	1560	173

本文, 采用分类任务中常用的准确率(ACC)和 F1 值作为实验的评价指标。

### 4.3. 对比实验及结果分析

为了综合分析模型特性以及表现, 与以下模型进行对比实验:

**BiLSTM [11]:** 基于双向 LSTM, 结合词嵌入进行方面提取。

**DTBCSNN [12]:** 是一种基于依赖树的堆叠卷积神经网络, 它使用线性链 CRF 进行方面提取。

**ATAE-LSTM [13]:** 在 AE-LSTM 的基础上, 将方面嵌入和词嵌入进行拼接, 共同组成模型的输入; 进行情感极性分类任务训练。

**PIPELINE-IMN:** 基于 IMN 模型使用流水线方法分别进行两个任务的训练。

**MNN:** 使用一种统一的序列标注模式定义训练任务, 同时进行方面抽取和情感分类。

**IMN [14]:** 使用交互架构, 实现端到端的 ABSA 多任务模型。

本文提出的 Att-Inter-MT 模型, 与上述基线模型在三个公开数据集上, 对比实验结果如表 2。

**Table 2.** Comparison of model accuracy and Macro-F1 score

**表 2.** 模型准确率及 F1 值对比

Model	Laptop			Restaurant			Twitter		
	F1AE %	AccAS %	F1AS %	F1AE %	AccAS %	F1AS %	F1AE %	AccAS %	F1AS %
BiLSTM	73.72	-	-	81.42	-	-	70.22	-	-
DTBCSNN	75.66	-	-	83.97	-	-	75.33	-	-
ATAE-LSTM	-	68.70	62.40	-	77.30	65.05	-	69.43	63.47
PIPELINE-IMN	78.38	72.29	68.12	83.94	79.56	69.59	73.54	74.02	66.33
MNN	76.94	70.40	65.98	83.05	77.17	68.50	72.05	71.05	63.87
IMN	77.96	75.36	72.02	83.33	<b>84.89</b>	<b>75.66</b>	75.63	75.45	67.05
<b>Inter-AE-MHSA</b>	<b>79.87</b>	<b>76.12</b>	<b>74.65</b>	<b>85.15</b>	83.96	74.89	<b>77.33</b>	<b>77.21</b>	<b>69.34</b>

从表 2 可以看出, 多任务模型 MNN、IMN 和本文模型在两类子任务上的性能表现, 显著优于单任务模型。这说明, 方面级情感分析的两类子任务间是存在强联系的, 通过多任务模型的特征共享, 可以利用该联系提升任务性能。而且, 本文模型在方面项抽取任务的表现, 明显优于其它多任务模型, 在 Laptop 数据集上比 MNN、IMN 等多任务模型效果提升了 2%~3%; 说明通过交互机制可以获得两子任务的边界信息等联系, 对方面词提取具有提升作用。

## 5. 总结

本文提出基于注意力机制的交互机制、基于多头注意力机制的特征交互融合机制和使用结合 Bi-LSTM 和 CNN 的特征提取方法的 Att-Inter-MT 多任务模型, 模型首先使用 BERT 预训练模型作为词嵌入生成工具, 获得融合上下文语义的词向量表示。然后, 使用结合 Bi-LSTM 和 CNN 的特征提取方法, 获取合适特征表示; 之后, 使用基于多头注意力机制的特征交互融合机制对提取特征进行多维度的交互学习, 获得合适的融合特征语义。最后通过基于注意力机制的交互机制获得最后的特征表示, 通过线性链获得输出的标签结果。经过比对多个模型, 本文模型在两个子任务的性能优于其它模型。说明多任务



模型在当前方面级情感分析数据有限的情景下, 其有效提高模型的泛化能力。

## 参考文献

- [1] Yadav, A. and Vishwakarma, D.K. (2020) Sentiment Analysis Using Deep Learning Architectures: A Review. *Artificial Intelligence Review*, **53**, 4335-4385. <https://doi.org/10.1007/s10462-019-09794-5>
- [2] Zhou, J., Huang, J.X., Chen, Q., et al. (2019) Deep Learning for Aspect-Level Sentiment Classification: Survey, Vision and Challenges. *IEEE Access*, **7**, 78454-78483. <https://doi.org/10.1109/ACCESS.2019.2920075>
- [3] 郝亚洲, 郑庆华, 陈艳平, 闫彩霞. 面向网络舆情数据的异常行为识别[J]. 计算机研究与发展, 2016, 53(3): 611-620.
- [4] 张宜浩, 朱小飞, 徐传运, 董世都. 基于用户评论的深度情感分析和多视图协同融合的混合推荐方法[J]. 计算机学报, 2019, 42(6): 1316-1333.
- [5] Devlin, J., Chang, M.W., Lee, K., et al. (2019) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceeding of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, 2-7 June 2019, 4171-4186.
- [6] Song, Y., Wang, J., Tao, J., et al. (2019) Attentional Encoder Network for Targeted Sentiment Classification.
- [7] Zeng, B., Yang, H., Xu, R., et al. (2019) LCF: A Local Context Focus Mechanism for Aspect-Based Sentiment Classification. *Applied Sciences*, **9**, 3389-3389. <https://doi.org/10.3390/app9163389>
- [8] Tang, D., Qin, B., Feng, X., et al. (2015) Effective LSTMs for Target-Dependent Sentiment Classification. *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers*, Osaka, 11-17 December 2016, 3298-3307.
- [9] 彭祝亮, 刘博文, 范程岸, 王杰, 肖明, 廖泽恩. 基于 BLSTM 与方面注意力模块的情感分类方法[J]. 计算机工程, 2020, 46(3): 60-65+72. <https://doi.org/10.19678/j.issn.1000-3428.0053929>
- [10] Dong, L., Wei, F., Tan, C., et al. (2014) Adaptive Recursive Neural Network for Target-Dependent Twitter Sentiment Classification. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Baltimore, 23-25 June 2014, 49-54. <https://doi.org/10.3115/v1/P14-2009>
- [11] Liu, P., Joty, S. and Meng, H. (2015) Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, 17-21 September 2015, 1433-1443. <https://doi.org/10.18653/v1/D15-1168>
- [12] Ye, H., Yan, Z., Luo, Z., et al. (2017) Dependency-Tree Based Convolutional Neural Networks for Aspect Term Extraction. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Jeju, 23-26 May 2017, 350-362. [https://doi.org/10.1007/978-3-319-57529-2\\_28](https://doi.org/10.1007/978-3-319-57529-2_28)
- [13] Wang, Y., Huang, M., Zhu, X., et al. (2016) Attention-Based LSTM for Aspect-Level Sentiment Classification. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, 1-5 November 2016, 606-615. <https://doi.org/10.18653/v1/D16-1058>
- [14] Wang, F., Lan, M. and Wang, W. (2018) Towards a One-Stop Solution to Both Aspect Extraction and Sentiment Analysis Tasks with Neural Multi-task Learning. *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, 8-13 July 2018, 1-8. <https://doi.org/10.1109/IJCNN.2018.8489042>