计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202220441
60(12): 2877−2889, 2023

基于多模态方面术语提取和方面级情感分类的统一框架

周　如　朱浩泽　郭文雅　于胜龙　张　莹
（南开大学计算机学院　天津　300350）
（zhouru@mail.nankai.edu.cn）

A Unified Framework Based on Multimodal Aspect-Term Extraction and Aspect-Level Sentiment Classification

Zhou Ru, Zhu Haoze, Guo Wenya, Yu Shenglong, and Zhang Ying
（*College of Computer Science*, *Nankai University*, *Tianjin* 300350）

Abstract　Aspect-term extraction (AE) and aspect-level sentiment classification (ALSC) extract aspect-sentiment pairs in the sentence, which helps social media platforms such as Twitter and Facebook to mine users' sentiments of different aspects and is of great significance to personalized recommendation. In the field of multimodality, the existing method uses two independent models to complete two subtasks respectively. Aspect-term extraction identifies goods, important people and other entities or entities' aspects in the sentence, and aspect-level sentiment classification predicts the user's sentiment orientation according to the given aspect terms. There are two problems in the above method: first, using two independent models loses the continuity of the underlying features between the two tasks, and cannot model the potential semantic association of sentences; second, aspect-level sentiment classification can only predict the sentiment of one aspect at a time, which does not match the throughput of aspect-term extraction model that extracts multiple aspects simultaneously, and the serial execution of the two models makes the efficiency of extracting aspect-sentiment pairs low. To solve the above problems, a unified framework based on multimodal aspect-term extraction and aspect-level sentiment classification, called UMAS, is proposed in this paper. Firstly, the shared feature module is built to realize the latent semantic association modeling between tasks, and to make the two subtasks only need to care about their upper network, which reduces the complexity of the model. Secondly, multiple aspects and their corresponding sentiment categories in the sentence are output at the same time by using sequence tagging, which improves the extraction efficiency of aspect-sentiment pairs. In addition, we introduce part of speech in two subtasks at the same time: using the grammatical information to improve the performance of aspect-term extraction, and the information of opinion words is obtained through part of speech to improve the performance of aspect-level sentiment classification. The experimental results show that the unified model has superior performance compared with multiple baseline models on the two benchmark datasets of Twitter2015 and Restaurant2014.

Key words　aspect-term extraction (AE); aspect-level sentiment classification (ALSC); unified framework; shared feature representation; sequence tagging

摘　要　通过方面术语提取和方面级情感分类任务提取句子中的方面-情感对，有助于Twitter，Facebook等社交媒体平台挖掘用户对不同方面的情感，对个性化推荐有重要的意义.在多模态领域，现有方法使用2

收稿日期：2022-05-28；**修回日期**：2023-02-20
基金项目：国家自然科学基金-联合基金项目（U1903128）
This work was supported by the National Natural Science Foundation of China-Joint Fund program（U1903128）.
通信作者：郭文雅（guowenya@dbis.nankai.edu.cn）

个独立的模型分别完成2个子任务，方面术语提取提取句子中包含的商品、重要人物等实体或实体的方面，方面级情感分类根据给定的方面术语预测用户的情感倾向. 上述方法存在2个问题：1）使用2个独立的模型丢失了2个任务之间在底层特征的延续性，无法建模句子潜在的语义关联；2）方面级情感分类1次预测1个方面的情感，与方面术语提取同时提取多个方面的吞吐量不匹配，且2个模型串行执行使得提取方面-情感对的效率低. 为解决这2个问题，提出基于多模态方面术语提取和方面级情感分类的统一框架UMAS. 首先，建立共享特征模块，实现任务间潜在语义关联建模，并且共享表示层使得2个子任务只需关心各自上层的网络，降低了模型的复杂性；其次，模型利用序列标注同时输出句子中包含的多个方面及其对应的情感类别，提高了方面-情感对的提取效率. 此外，在这2个子任务中同时引入词性：利用其中蕴含的语法信息提升方面术语提取的性能；通过词性获取观点词信息，提升方面级情感分类的性能. 实验结果表明，该统一框架在Twitter2015, Restaurant2014这2个基准数据集上相比于多个基线模型具有优越的性能.

关键词 方面术语提取（AE）；方面级情感分类（ALSC）；统一框架；共享特征表示；序列标注

中图法分类号 TP391.1

随着互联网的发展，社交媒体平台成为人们发表言论和观点的主要阵地，高效地识别用户对重要组织、重要人物、商品等实体及其方面①的情感对平台治理用户的不当言论、建模用户偏好以实现精准的个性化推荐有重要的实用意义. 同时也有助于监控消费者行为、评估产品质量、监控舆情、调研市场等.

不同于句子级情感分析任务为整个句子预测情感，方面术语提取和方面级情感分类(aspect-term extraction and aspect-level sentiment classification, AESC)任务的目标是抽取句子中的方面-情感对. 方面术语提取(aspect-term extraction, AE)提取句子中包含的方面术语，方面级情感分类(aspect-level sentiment classification, ALSC)预测用户对给定方面的情感. 比如来自Twitter的一条评论：“I love animals , so nice to see them getting along ! Here are our dogs , Greek and Salem , laying together”，提取出的方面-情感对为〈“Greek”，正面〉〈“Salem”，正面〉，即句子中包含方面“Greek”和“Salem”，表述者对它们的情感都是正面的.

在文本领域中，已有研究[1-3]实现了方面-情感对提取方法，并应用于商品评论数据的情感分析. 然而在Twitter, Instagram等社交媒体平台上，人们习惯发表短小且口语化的文字并配以图片，相关研究指出，文本单模态的模型在此类用户数据上表现并不好[4-6]. 考虑图片非仅仅依靠文本来分析用户发表的观点是时代的趋势，因此在多模态领域实现方面术语提取和方面级情感分类将具有一定的实用价值和现实意义.

在多模态领域，Zhang等人[7]和Yu等人[8]分别研究了方面术语抽取和方面级情感分类. 通过实体识别技术提取句子中包含的方面术语，接着将提取的方面术语和句子输入到方面级情感分类模型进行情感预测，可通过这种流水线方式实现方面-情感对的提取. 然而，目前的这种方法存在不足之处：首先，使用2个完全独立的模型分步实现方面-情感对的提取，使得建模特征的语义深度不同且不关联，忽略了2个任务之间潜在的语义关联，当句子中包含多个方面时，情感分类模型可能会混淆它们之间的上下文信息而造成预测失误；其次，方面术语提取模型一次提取句子中的多个方面术语，而情感分析模型一次只能预测一个方面的情感，前者的吞吐量大于后者，且情感分析必须在方面术语提取完成后进行，降低了方面-情感对的抽取效率.

针对以上问题，本文提出了一个同时进行方面术语提取和方面级情感分类的统一框架UMAS. 该统一框架包含3个模块：共享特征模块、方面术语提取模块、情感分类模块. 首先，该统一框架使用共享特征的方式表示方面术语提取和情感分类2个子任务的底层文本和图像特征，在学习的过程中建立2个子任务之间的语义联系. 相比于之前的方面术语提取模型和方面级情感分类模型使用不同的网络编码文本和图像的特征，本文所提出的特征共享的方法简化了模型. 其次，采用序列标注的方式，同时输出句子中包含的多个方面和对应的情感，方面术语提取模块和情感分类模块可并行执行，大大提升了方面-情感对提取的效率.

此外，既往多模态方面术语提取方法[7,9-10]未能充分利用文本的语法信息，而方面级情感分析方法[8,11]

① 方面指的是实体或实体的属性.

由于缺乏观点词的标注而未能通过观点信息更好地判断情感倾向. 为提升 2 个子任务的性能, 本文使用词性标注工具 spaCy[11] 获取单词的词性, 对 2 个子任务做如下改进: 在方面术语提取模块中, 使用多头自注意力机制获取词性特征, 融合视觉特征、文本特征、词性特征作为分类层的输入, 提升了方面术语提取的性能; 在情感分类模块, 为充分发挥观点词对情感分类的作用, 通过词性标注将动词、形容词、副词、介词标记为观点词, 在情感分类中增加对这些观点词的注意权重, 并将观点词特征融入到最后的分类层以提升情感分类的性能. 本文提出的方法与多个基线模型相比, 在方面术语提取、方面级情感分类、AESC 任务上的性能都有明显的提升.

本文的主要贡献有 3 个方面:

1) 在多模态领域提出方面术语提取和方面级情感分类的统一框架 UMAS (unified multi-modal aspect sentiment), 通过建模方面术语提取和方面级情感分类任务之间的语义关联, 同时提高了方面-情感对提取的性能和效率.

2) 本文通过引入词性特征提升了方面术语提取的性能; 通过词性标注获取观点词特征并结合位置信息, 提升了方面级情感分类的性能.

3) 该统一框架在 Twitter2015, Restaurant2014 这 2 个基准数据集上相比于多个基线模型在方面术语提取、方面级情感分类、AESC 任务上都具有优越的性能.

1 相关工作

目前, 文本领域的基于方面的情感分析研究发展的比较成熟, 现有研究[12-18] 在 Restaurant, Laptop, Twitter 等数据集上, 根据提供的方面术语预测情感类别; Ying 等人[19] 根据方面术语提取对应的观点并判断情感倾向; Oh 等人[20]、Chen 等人[21]、Xu 等人[22] 则使用多任务模型将方面术语提取、观点词提取、情感分类 3 个任务统一. 其中, Chen 等人[21] 详细阐述了 3 个任务之间的关系, 并在多层的网络模型 RACL 中通过关系传导机制促进子任务之间的协作, 最终以序列标注的方式分别输出 3 个任务的结果. RACL 将 3 个任务的关系总结如下: 方面术语和观点词存在对应关系(比如“美味”一词不适合描述地点), 方面术语和观点词的配对有助于预测情感, 观点词对情感预测有最直接的帮助, 方面术语是情感依托的对象. 文本领域的方面术语提取方法更关注文本的语法信息, Phan 等人[23] 和薛芳等人[24] 借助句法成分、依存关系提升方面术语提取的性能. 在情感分类中, Chen 等人[21]、He 等人[25] 利用观点词的信息提升了情感推断的准确性, He 等人[25] 还利用了位置信息使注意力集中在方面的上下文. 文本领域基于方面的情感分析的研究, 对多模态基于方面的情感分析的研究有重要的启发式意义.

在多模态领域, 可使用 Zhang 等人[7] 提出的方面术语抽取模型和 Yu 等人[8] 提出的方面级情感分类模型流水线式地抽取方面-情感对. 尽管流水线方法符合人们处理此类问题的直觉且有利于灵活变动 2 个模型, 但 Wang 等人[26] 指出该方法在方面术语提取中的错误将传播到情感预测阶段, 导致方面-情感对预测性能下降. 方面术语提取和方面级情感分类 2 个模型的独立无法像 RACL 一样建模 2 个任务之间的语义联系, 且串行执行使得模型效率低下. 多模态方面术语提取方法[7,9-10] 充分关注了图像对提取方面术语的帮助, 并且使用门控机制降低图像引入的噪音, 但忽视了文本中包含的语法信息. 在文本领域的方面级情感分类中, 多种方法[19-21] 利用观点词提取作为辅助任务提升情感分类的效果, 然而多模态方面级情感分类的数据集主要是 Twitter, 目前数据集中包含的信息包括句子、图片、方面、情感等的标准, 但是未有观点词的标注信息, 所以多模态领域中以观点词提取为辅助任务的方法不存在监督信息, 难以开展. 此外, 目前多模态方面级情感分类模型如 EASFN[8], ABAFN[12], 以句子、图像、方面术语为输入, 一次只能识别一个方面的情感, 而文本领域采用序列标注的方法可同时识别句子中所有方面的情感.

2 基于多模态方面术语提取和方面级情感分类的统一框架

本节主要介绍任务定义, 并详细阐述本文所提出的基于多模态方面术语提取和方面级情感分类的统一框架.

2.1 任务定义

给定长度为 n 的句子, 即 $S=\{w_1, w_2, \cdots, w_n\}$, 方面术语提取任务的目的是获取句子的方面术语标注序列 $Y^{\mathrm{A}}=\left\{y_1^{\mathrm{A}}, y_2^{\mathrm{A}}, \cdots, y_i^{\mathrm{A}}, \cdots, y_n^{\mathrm{A}}\right\}$, $y_i^{\mathrm{A}} \in \{\mathrm{B}, \mathrm{I}, \mathrm{O}\}$, 其中 B 表示方面术语的开始单词, I 表示方面术语的中间单词及结尾单词, O 表示不是方面术语. 而方面级情感分类任务的目的是获取句子的情感标注序列 $Y^{\mathrm{S}}=\left\{y_1^{\mathrm{S}}, y_2^{\mathrm{S}}, \cdots, y_i^{\mathrm{S}}, \cdots, y_n^{\mathrm{S}}\right\}$, $y_i^{\mathrm{S}} \in \{0, 1, 2, 3\}$, 其中 0 表示该单

词不是方面术语, 不被赋予情感, 1 表示情感为负面, 2 表示情感中立, 3 表示情感为正面. 方面术语提取和方面级情感分类的目的是抽取句子中包含的方面-情感对, 即$Y^{\mathrm{P}}=\{a_1^{\mathrm{s}}, a_1^{\mathrm{e}}, s_1, \cdots, a_i^{\mathrm{s}}, a_i^{\mathrm{e}}, s_i, \cdots, a_m^{\mathrm{s}}, a_m^{\mathrm{e}}, s_m\}$, 其中$a_i^{\mathrm{s}}, a_i^{\mathrm{e}}, s_i$分别为第 i 个方面术语的起始位置、终止位置和对应的情感类别.

2.2 模型概述

本文设计的方面术语提取和方面级情感分类的统一框架主要分为 3 个模块: 共享特征模块、方面术语提取模块和情感分类模块, 模型图如图 1 所示.

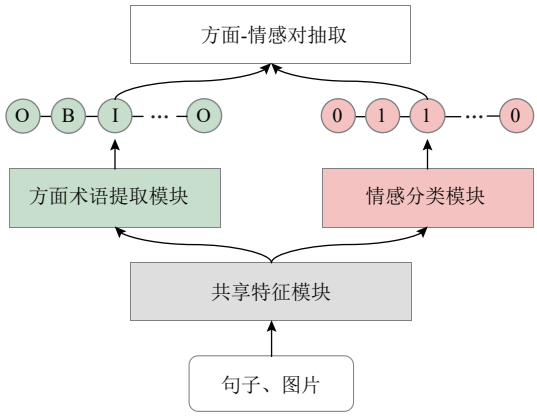


Fig. 1 Framework of our proposed model

图 1 本文模型框架

在共享特征模块, 使用 VGG-16 模型[27]获取图片特征表示, 通过双向长短期记忆网络(bi-long short-term memory, BiLSTM)获取单词和字符的联合特征表示, 通过多头自注意力机制[28]获取词性特征表示. 方面术语提取模块和情感分类模块以共享特征为输入, 编码出特定于各自任务的私有特征. 在方面术语提取模块, 通过文本和图像的交互注意力以及门控机制获取多模态表示, 并与文本及词性特征拼接, 作为方面术语提取模块最终的融合特征, 最后通过条件随机场(conditional random fields, CRF)层获取方面术语序列标注. 情感分类模块将共享特征和特有特征融合, 获取情感特征和观点词特征. 通过门控机制融合由情感特征引导的视觉注意特征和情感特征以获得多模态特征, 并通过情感文本注意、位置信息和词性获得观点词特征, 然后, 将多模态特征和情感特征以及观点词特征融合, 通过全连接层及 softmax 层获得情感序列标注. 在获得方面术语序列标签和情感序列标签后, 通过简单的代码提取方面-情感对, 实现 AESC 任务的目标. 图 2 是本文所提出的基于多模态方面术语提取和方面级情感分类的统一框架.

本文提出的方面术语提取和方面级情感分类的统一框架借鉴了多任务学习的思路, 即通过参数共享建模 2 个子任务的语义联系, 提升每个子任务的性能, 并使用子任务的加权损失作为模型的损失. 但多任务模型通常有多个主要目标, 而本文所提出的模型的主要目标只有 1 个, 即抽取方面-情感对.

2.3 共享特征模块

共享特征模块的图像特征、文本特征、词性特征分别由图像编码器、文本编码器、词性编码器生成.

2.3.1 图像编码器

裁剪图片为 224×224 像素, 作为 VGG-Net16[27]的输入, 图像编码器保留最后 1 层池化层输出结果作为图像特征(维度为 512×7×7). 其中, 7×7 代表图像的 49 个区域, 512 表示每个区域的特征维度. 所以图像特征可表示为$\tilde{v}_{\mathrm{I}}=\{\boldsymbol{v}_i|\boldsymbol{v}_i\in\mathbb{R}^{d_v}, i=1,2,\cdots,49\}$, $\boldsymbol{v}_i$代表图像区域 i 的具有 512 维度的特征向量.

2.3.2 文本编码器

字符级的嵌入式表示可以减轻罕见词和拼写错误的问题, 且能捕获前缀后缀的信息, 因此, 本文将字符级表示作为单词表示的一部分. 通过查找字符向量表, 可以获取第 t 个单词的字符表示$c_{t,\mathrm{W}}=\{\boldsymbol{c}_{t,1}, \boldsymbol{c}_{t,2}, \cdots, \boldsymbol{c}_{t,m}\}$, 其中$\boldsymbol{c}_{t,i}\in\mathbb{R}^{d_c}$为第 t 个单词第 i 个字母的向量表示, m 为单词的长度. k 个不同窗口大小的卷积核$[C_1, C_2, \cdots, C_k]$被应用在单词特征上, 每一次卷积后加一步最大池化操作, 最后将获得的 k 个特征$\boldsymbol{w}'_{t,1}$, $\boldsymbol{w}'_{t,2}$, $\cdots$, $\boldsymbol{w}'_{t,k}$拼接在一起作为单词的字符级表示, 即

$$\boldsymbol{w}'_t=[\boldsymbol{w}'_{t,1}\oplus\boldsymbol{w}'_{t,2}\oplus\cdots\oplus\boldsymbol{w}'_{t,k}]. \tag{1}$$

通过查询预训练的词向量矩阵, 可获得单词 t 的词嵌入式表示$\boldsymbol{w}''_t$, 将其与字符特征$\boldsymbol{w}'_t$拼接在一起作为单词 t 的联合表示, 即$\boldsymbol{w}_t=[\boldsymbol{w}'_t, \boldsymbol{w}''_t]$. 接着, 使用 BiLSTM 获取包含上下文信息的单词 t 的隐藏特征$\boldsymbol{h}_t$, 即

$$\boldsymbol{h}_t=[\overrightarrow{\boldsymbol{h}_t}, \overleftarrow{\boldsymbol{h}_t}], \tag{2}$$

$$\boldsymbol{H}=\left\{\boldsymbol{h}_j\middle|\boldsymbol{h}_j\in\mathbb{R}^d, j=1,2,\cdots,n\right\}, \tag{3}$$

其中 $\boldsymbol{H}$ 表示最终的共享文本特征, d 为隐藏特征的向量维度.

2.3.3 词性编码器

Phan 等人[23]使用句法成分信息提升了方面术语提取的准确率, 本文同样也使用 spaCy 工具获取单词的词性. 根据随机初始化的词性向量矩阵, 可获得句子的词性特征$\widetilde{\boldsymbol{P}}=(\widetilde{\boldsymbol{p}_1}, \widetilde{\boldsymbol{p}_2}, \cdots, \widetilde{\boldsymbol{p}_n})$($n$ 为句子长度). 然后, 本文使用文献 [27] 中的多头自注意力机制进一步获取深层次的词性嵌入式特征$\boldsymbol{P}$.

本文提出的模型中共有 2 个结构相同的文本编码器, 分别为共享文本编码器和情感模块的私有文

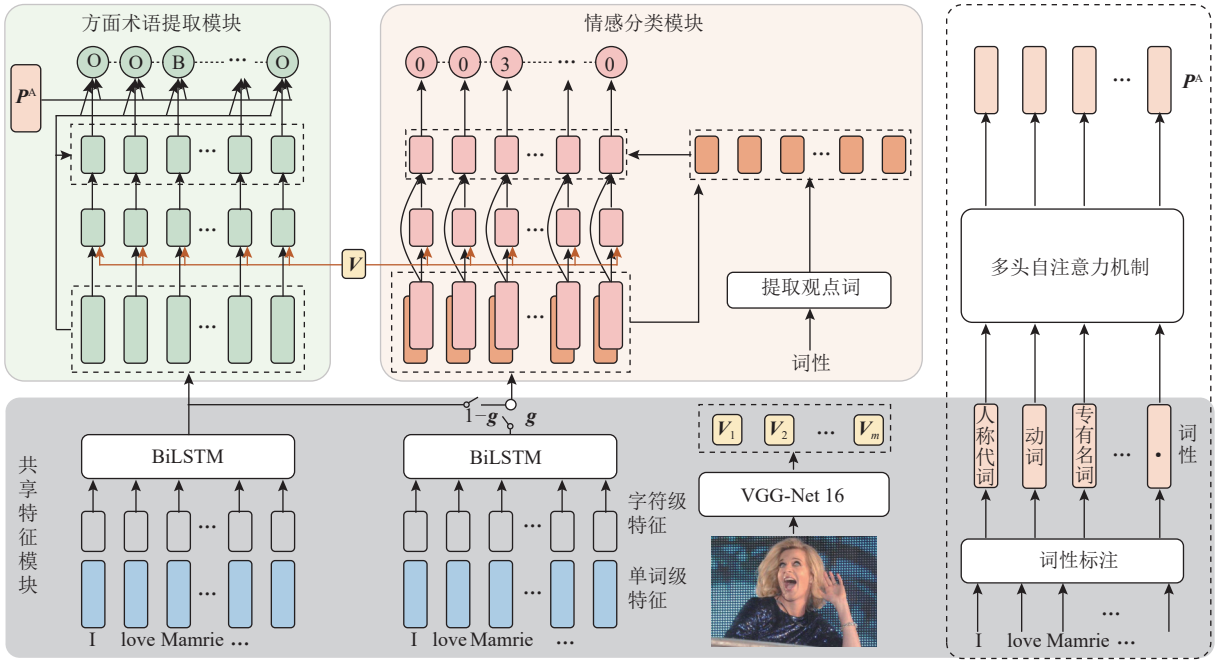


Fig. 2 Unified framework based on multimodal aspect term extraction and aspect-level sentiment classification

图 2 基于多模态方面术语提取和方面级情感分类的统一框架

本编码器. 方面术语提取模块和情感分类模块共享图像编码器、词性编码器、共享文本编码器的输出数据.

2.4 方面术语提取模块

方面术语提取模块通过文本注意和视觉注意建模不同模态之间的语义交互作用, 使用门控机制获取多模态融合特征, 并使用过滤门减少多模态引入的噪音, 最后将多模态融合特征、文本特征、词性特征拼接作为 CRF 解码器的输入, 获得方面术语标注序列.

首先, 使用线性层分别将图像特征映射到与文本同维度的空间, 将共享文本特征编码为方面术语提取模块的私有文本特征, 即

$$\boldsymbol{v}_{\mathrm{I}}^{\mathrm{A}}=\tanh(\boldsymbol{W}_{\mathrm{I}}^{\mathrm{A}}\tilde{\boldsymbol{v}}_{\mathrm{I}}+\boldsymbol{b}_{\mathrm{I}}^{\mathrm{A}}),\tag{4}$$

$$\boldsymbol{X}^{\mathrm{A}}=\tanh(\boldsymbol{W}_{\mathrm{H}}^{\mathrm{A}}\boldsymbol{H}+\boldsymbol{b}_{\mathrm{H}}^{\mathrm{A}}),\tag{5}$$

其中$\boldsymbol{W}_{\mathrm{I}}^{\mathrm{A}},\boldsymbol{W}_{\mathrm{H}}^{\mathrm{A}},\boldsymbol{b}_{\mathrm{I}}^{\mathrm{A}},\boldsymbol{b}_{\mathrm{H}}^{\mathrm{A}}$为可训练参数.

通常情况下, 句子中的单词只对应图像中的一小块区域, 为减小图像其他区域引入的噪音, 该模块使用文本引导的视觉注意来获取不同区域的权重, 图像区域与单词越相关, 它被赋予的权重越大. 给定一个单词的特征$\boldsymbol{x}_t^{\mathrm{A}}(\boldsymbol{x}_t^{\mathrm{A}}\in\boldsymbol{X}^{\mathrm{A}})$, 通过神经网络和 softmax 函数来生成单词 t 对应的图像权重分布$\boldsymbol{\alpha}_t$, 并通过加权和生成单词 t 对应的图像特征表示$\widehat{\boldsymbol{v}_t^{\mathrm{A}}}$, 即

$$\boldsymbol{z}_t^{\mathrm{M}}=\tanh(\boldsymbol{W}_{\boldsymbol{v}_{\mathrm{I}}^{\mathrm{A}}}\boldsymbol{v}_{\mathrm{I}}^{\mathrm{A}}\oplus(\boldsymbol{W}_{\boldsymbol{h}_t}\boldsymbol{x}_t^{\mathrm{A}}+\boldsymbol{b}_{\boldsymbol{h}_t})),\tag{6}$$

$$\boldsymbol{\alpha}_t=softmax(\boldsymbol{W}_{\boldsymbol{\alpha}_t}\boldsymbol{z}_t^{\mathrm{M}}+\boldsymbol{b}_{\boldsymbol{\alpha}_t}),\tag{7}$$

$$\widehat{\boldsymbol{v}_t^{\mathrm{A}}}=\sum_i\boldsymbol{\alpha}_{t,i}\boldsymbol{v}_i^{\mathrm{A}},\tag{8}$$

其中$\boldsymbol{x}_t^{\mathrm{A}}\in\mathbb{R}^d$, d 为单词和图像特征的维度, $\boldsymbol{v}_{\mathrm{I}}^{\mathrm{A}}\in\mathbb{R}^{d\times N}$表示 N 个图片区域的特征, $\boldsymbol{v}_i^{\mathrm{A}}\in\mathbb{R}^d$表示图片第 i 个区域的特征. $\boldsymbol{W}_{\boldsymbol{v}_{\mathrm{I}}^{\mathrm{A}}},\boldsymbol{W}_{\boldsymbol{h}_t},\boldsymbol{W}_{\boldsymbol{\alpha}_t},\boldsymbol{b}_{\boldsymbol{h}_t},\boldsymbol{b}_{\boldsymbol{\alpha}_t}$为可训练的参数. 符号$\oplus$表示 2 个特征的拼接, 当 2 个操作数分别为矩阵和向量时, 表示复制多个向量与矩阵的每一列进行拼接.

类似地, 上下文有助于丰富当前单词特征包含的信息, 且对上下文不同的单词应当有不同的关注程度, 所以本文通过视觉引导的文本注意力来获取单词 t 所需关注的上下文的权重$\boldsymbol{\beta}_t$, 通过对句子中单词的加权获得单词 t 的新的特征表示.

$$\boldsymbol{z}_t^{\mathrm{T}}=\tanh(\boldsymbol{W}_{\boldsymbol{X}^{\mathrm{A}}}\boldsymbol{X}^{\mathrm{A}}\oplus(\boldsymbol{W}_{\boldsymbol{X},\widehat{\boldsymbol{v}_t^{\mathrm{A}}}}\widehat{\boldsymbol{v}_t^{\mathrm{A}}}+\boldsymbol{b}_{\boldsymbol{X},\widehat{\boldsymbol{v}_t^{\mathrm{A}}}})),\tag{9}$$

$$\boldsymbol{\beta}_t=softmax(\boldsymbol{W}_{\boldsymbol{\beta}_t}\boldsymbol{z}_t^{\mathrm{T}}+\boldsymbol{b}_{\boldsymbol{\beta}_t}),\tag{10}$$

$$\widehat{\boldsymbol{x}_t^{\mathrm{A}}}=\sum_j\boldsymbol{\beta}_{t,j}\boldsymbol{x}_j^{\mathrm{A}},\tag{11}$$

其中$\boldsymbol{W}_{\boldsymbol{X}^{\mathrm{A}}},\boldsymbol{W}_{\boldsymbol{X},\widehat{\boldsymbol{v}_t^{\mathrm{A}}}},\boldsymbol{W}_{\boldsymbol{\beta}_t},\boldsymbol{b}_{\boldsymbol{X},\widehat{\boldsymbol{v}_t^{\mathrm{A}}}},\boldsymbol{b}_{\boldsymbol{\beta}_t}$为可训练的参数.

当句子中包含多个实体时, 往往并不是每个实体都存在与图像中的某个区域对应的关系, 可能图片中描述了一个实体, 而句子中有 3 个不同的实体. 为此, 在融合多模态特征时, 也需动态权衡视觉特征和文本特征的比例. 方面术语提取模块使用式(12)~(15)获取多模态融合特征$\boldsymbol{m}_t^{\mathrm{A}}$:

$$\boldsymbol{h}_{\widehat{\boldsymbol{v}_t^{\mathrm{A}}}}=\tanh(\boldsymbol{W}_{\widehat{\boldsymbol{v}_t^{\mathrm{A}}}}\widehat{\boldsymbol{v}_t^{\mathrm{A}}}+\boldsymbol{b}_{\widehat{\boldsymbol{v}_t^{\mathrm{A}}}}),\tag{12}$$

$$\boldsymbol{h}_{\widehat{\boldsymbol{x}_t^{\mathrm{A}}}}=\tanh(\boldsymbol{W}_{\widehat{\boldsymbol{x}_t^{\mathrm{A}}}}\widehat{\boldsymbol{x}_t^{\mathrm{A}}}+\boldsymbol{b}_{\widehat{\boldsymbol{x}_t^{\mathrm{A}}}}),\tag{13}$$

$$\boldsymbol{g}_t = \sigma\left(\boldsymbol{W}_{\boldsymbol{g}_t}\left(\boldsymbol{h}_{\widehat{v_t^{\mathrm{A}}}} \oplus \boldsymbol{h}_{\widehat{x_t^{\mathrm{A}}}}\right)\right), \tag{14}$$

$$\boldsymbol{m}_t^{\mathrm{A}} = \boldsymbol{g}_t \boldsymbol{h}_{\widehat{v_t^{\mathrm{A}}}} + (1-\boldsymbol{g}_t)\boldsymbol{h}_{\widehat{x_t^{\mathrm{A}}}}, \tag{15}$$

其中$\boldsymbol{W}_{\widehat{v_t^{\mathrm{A}}}}, \boldsymbol{W}_{\widehat{x_t^{\mathrm{A}}}}, \boldsymbol{W}_{\boldsymbol{g}_t}, \boldsymbol{b}_{\widehat{v_t^{\mathrm{A}}}}, \boldsymbol{b}_{\widehat{x_t^{\mathrm{A}}}}$为参数，$\boldsymbol{h}_{\widehat{v_t^{\mathrm{A}}}}, \boldsymbol{h}_{\widehat{x_t^{\mathrm{A}}}}$分别为使用线性层获取的单词 t 对应的新的视觉特征和文本特征，$\boldsymbol{g}_t$为通过 sigmoid 激活函数获取的视觉特征的权重.

尽管多模态融合特征考虑了文本和图像的权重，但方面术语提取所依赖的最重要的数据应该是文本，所以方面术语提取模块将初始的文本特征、多模态特征和词性特征拼接起来作为解码器的输入. 此外，当预测的单词是动词或副词时，加入图像特征会引起噪音，所以在拼接之前，对多模态特征进行过滤操作，具体公式为：

$$\boldsymbol{s}_t = \sigma(\boldsymbol{W}_{\boldsymbol{s}_t, \boldsymbol{x}_t^{\mathrm{A}}} \boldsymbol{x}_t^{\mathrm{A}} \oplus (\boldsymbol{W}_{\boldsymbol{m}_t^{\mathrm{A}}, \boldsymbol{s}_t} \boldsymbol{m}_t^{\mathrm{A}} + \boldsymbol{b}_{\boldsymbol{m}_t^{\mathrm{A}}, \boldsymbol{s}_t})), \tag{16}$$

$$\boldsymbol{u}_t^{\mathrm{A}} = \boldsymbol{s}_t(\tanh(\boldsymbol{W}_{\boldsymbol{m}_t^{\mathrm{A}}} \boldsymbol{m}_t^{\mathrm{A}} + \boldsymbol{b}_{\boldsymbol{m}_t^{\mathrm{A}}})), \tag{17}$$

$$\widehat{\boldsymbol{m}_t^{\mathrm{A}}} = \boldsymbol{W}_{\widehat{\boldsymbol{m}_t^{\mathrm{A}}}}(\boldsymbol{x}_t^{\mathrm{A}} \oplus \boldsymbol{u}_t^{\mathrm{A}} \oplus \boldsymbol{p}_t^{\mathrm{A}}), \tag{18}$$

$$\widehat{\boldsymbol{M}^{\mathrm{A}}} = \{\widehat{\boldsymbol{m}_1^{\mathrm{A}}}, \widehat{\boldsymbol{m}_2^{\mathrm{A}}}, \cdots, \widehat{\boldsymbol{m}_n^{\mathrm{A}}}\}, \tag{19}$$

其中$\boldsymbol{W}_{\boldsymbol{s}_t, \boldsymbol{x}_t^{\mathrm{A}}}, \boldsymbol{W}_{\boldsymbol{m}_t^{\mathrm{A}}, \boldsymbol{s}_t}, \boldsymbol{W}_{\boldsymbol{m}_t^{\mathrm{A}}}, \boldsymbol{W}_{\widehat{\boldsymbol{m}_t^{\mathrm{A}}}}, \boldsymbol{b}_{\boldsymbol{m}_t^{\mathrm{A}}, \boldsymbol{s}_t}, \boldsymbol{b}_{\boldsymbol{m}_t^{\mathrm{A}}}$为参数，$\boldsymbol{x}_t^{\mathrm{A}}, \boldsymbol{u}_t^{\mathrm{A}}, \boldsymbol{p}_t^{\mathrm{A}}$分别为单词 t 的文本特征、过滤后的多模态特征、词性特征，$\widehat{\boldsymbol{M}^{\mathrm{A}}}$为最终方面术语提取模块的句子表示.

最后，方面术语提取模块使用 CRF 作为解码器进行方面术语的序列标注. 以$\boldsymbol{X}=\{\boldsymbol{w}_0, \boldsymbol{w}_1, ..., \boldsymbol{w}_T\}$作为一般化的输入序列，其中$\boldsymbol{w}_i$表示第 i 个单词的特征向量，$Y=\{Y_0, y_1, ..., y_T\}$表示 $\boldsymbol{X}$ 对应的一种序列标签，Y 表示所有可能的序列标注集合. 对于给定的 $\boldsymbol{X}$，所有可能的 y 可以由式(20)计算得到：

$$p(y|\boldsymbol{X}) = \frac{\prod_{i=1}^{T} \Omega_i(y_{i-1}, y_i, \boldsymbol{X})}{\sum_{y' \in Y} \prod_{i=1}^{T} \Omega_i(y'_{i-1}, y'_i, \boldsymbol{X})}, \tag{20}$$

其中Ω表示可能性函数.

2.5 情感分类模块

情感分类模块可以分为 4 个部分：情感私有特征、多模态融合、观点词特征、情感分类.

2.5.1 情感私有特征

由于方面术语提取和情感分类的目标不一致，使用完全的共享特征机制会使训练效果不好，同时共享特征包含的信息有助于在底层更好地表现 2 个任务之间的语义联系，特别是方面作为情感的寄托者有助于情感的预测. 所以，在情感分类模块，存在一个私有的文本编码器以获取特有的情感特征. 接着，将共享表示层的文本特征和特有情感特征进行动态融合. 考虑使用动态融合是因为更关注共享特征中的方面而非其他单词. 该模块的情感私有特征表示$\boldsymbol{X}^{\mathrm{S}}$由式(21)~(25)获取：

$$\boldsymbol{H}^{\mathrm{S}} = f^{\mathrm{SC}}(\boldsymbol{S}), \tag{21}$$

$$\widehat{\boldsymbol{H}^{\mathrm{S}}} = \tanh(\boldsymbol{W}_{\boldsymbol{H}^{\mathrm{S}}} \boldsymbol{H}^{\mathrm{S}} + \boldsymbol{b}_{\boldsymbol{H}^{\mathrm{S}}}), \tag{22}$$

$$\widehat{\boldsymbol{H}} = \tanh(\boldsymbol{W}_{\boldsymbol{H}} \boldsymbol{H} + \boldsymbol{b}_{\boldsymbol{H}}), \tag{23}$$

$$\boldsymbol{g}^{\mathrm{S}} = \sigma\left(\boldsymbol{W}_{\boldsymbol{g}^{\mathrm{S}}}\left(\widehat{\boldsymbol{H}^{\mathrm{S}}} \oplus \widehat{\boldsymbol{H}}\right)\right), \tag{24}$$

$$\boldsymbol{X}^{\mathrm{S}} = \boldsymbol{g}^{\mathrm{S}} \widehat{\boldsymbol{H}^{\mathrm{S}}} + (1-\boldsymbol{g}^{\mathrm{S}})\widehat{\boldsymbol{H}}, \tag{25}$$

其中，f^{SC}表示表示情感模块私有文本编码器的函数，$\boldsymbol{S}$ 表示输入的句子，$\boldsymbol{W}_{\boldsymbol{H}^{\mathrm{S}}}, \boldsymbol{W}_{\boldsymbol{H}}, \boldsymbol{W}_{\boldsymbol{g}^{\mathrm{S}}}, \boldsymbol{b}_{\boldsymbol{H}^{\mathrm{S}}}, \boldsymbol{b}_{\boldsymbol{H}}$为参数.

2.5.2 多模态融合

用户在社交媒体发布的文字具有不完整、较短、口语化的特点，仅仅使用文本内容来推测情感是不充分的. 因此，情感分类模块使用图像信息来提升预测的准确性. 与方面术语提取模块一样，在判断单词 t(假设单词 t 为方面术语)的情感时，需要着重关注图像中该方面对应的区域，应尽量减少其他区域引起的干扰，所以使用相同的方法为不同的视觉区域分配不同的权重. 首先，将共享图像特征转换至与文本同一维度的空间内，然后使用情感引导的注意获取图像的权重分布$\boldsymbol{\gamma}_t$，最终加权获得单词 t 在情感分类模块对应的图像特征$\widehat{\boldsymbol{v}_t^{\mathrm{S}}}$，运算公式为：

$$\boldsymbol{v}_{\mathrm{I}}^{\mathrm{S}} = \tanh(\boldsymbol{W}_{\mathrm{I}}^{\mathrm{S}} \widetilde{\boldsymbol{v}_{\mathrm{I}}} + \boldsymbol{b}_{\mathrm{I}}^{\mathrm{S}}), \tag{26}$$

$$\boldsymbol{z}_t^{\mathrm{S}} = \tanh(\boldsymbol{W}_{\boldsymbol{v}_{\mathrm{I}}^{\mathrm{S}}} \boldsymbol{v}_{\mathrm{I}}^{\mathrm{S}} \oplus (\boldsymbol{W}_{\boldsymbol{h}_t} \boldsymbol{x}_t^{\mathrm{S}} + \boldsymbol{b}_{\boldsymbol{h}_t})), \tag{27}$$

$$\boldsymbol{\gamma}_t = softmax(\boldsymbol{W}_{\boldsymbol{\gamma}_t} \boldsymbol{z}_t^{\mathrm{S}} + \boldsymbol{b}_{\boldsymbol{\gamma}_t}), \tag{28}$$

$$\widehat{\boldsymbol{v}_t^{\mathrm{S}}} = \sum_i \boldsymbol{\gamma}_{t,i} \boldsymbol{v}_i^{\mathrm{S}}, \tag{29}$$

其中$\boldsymbol{W}_{\mathrm{I}}^{\mathrm{S}}, \boldsymbol{W}_{\boldsymbol{v}_{\mathrm{I}}^{\mathrm{S}}}, \boldsymbol{W}_{\boldsymbol{h}_t}, \boldsymbol{W}_{\boldsymbol{\gamma}_t}, \boldsymbol{b}_{\mathrm{I}}^{\mathrm{S}}, \boldsymbol{b}_{\boldsymbol{h}_t}, \boldsymbol{b}_{\boldsymbol{\gamma}_t}$为可训练的参数.

不同于方面术语提取模块对文本引入视觉注意的处理，在情感分类模块，为减少视觉特征引起的噪音，本文采用多头自注意的方式来获取单词 t 对上下文的关注，使某个位置的单词关注来自不同表示子空间的其他单词的特征. 该模块多头自注意力的查询矩阵、键矩阵、值矩阵都为情感特征矩阵. 最终多头自注意力输出的文本特征为$\widehat{\boldsymbol{X}^{\mathrm{S}}}$.

接着，同样通过门控机制获得情感特征和图像特征的多模态融合特征$\boldsymbol{m}_t^{\mathrm{S}}$.

$$\boldsymbol{h}_{\widehat{v_t^{\mathrm{S}}}} = \tanh(\boldsymbol{W}_{\widehat{v_t^{\mathrm{S}}}} \widehat{\boldsymbol{v}_t^{\mathrm{S}}} + \boldsymbol{b}_{\widehat{v_t^{\mathrm{S}}}}), \tag{30}$$

$$\boldsymbol{h}_{\widehat{x_t^{\mathrm{S}}}} = \tanh(\boldsymbol{W}_{\widehat{x_t^{\mathrm{S}}}} \widehat{\boldsymbol{x}_t^{\mathrm{S}}} + \boldsymbol{b}_{\widehat{x_t^{\mathrm{S}}}}), \tag{31}$$

$$\boldsymbol{g}_t^{\mathrm{S}} = \sigma\left(\boldsymbol{W}_{\boldsymbol{g}_t^{\mathrm{S}}}\left(\boldsymbol{h}_{\widehat{v_t^{\mathrm{S}}}} \oplus \boldsymbol{h}_{\widehat{x_t^{\mathrm{S}}}}\right)\right), \tag{32}$$

$$\boldsymbol{m}_t^{\mathrm{S}} = \boldsymbol{g}_t^{\mathrm{S}} \boldsymbol{h}_{\widehat{v_t^{\mathrm{S}}}} + (1-\boldsymbol{g}_t^{\mathrm{S}})\boldsymbol{h}_{\widehat{x_t^{\mathrm{S}}}}, \tag{33}$$

其中$\boldsymbol{W}_{\widehat{v_t^{\mathrm{S}}}}, \boldsymbol{W}_{\widehat{x_t^{\mathrm{S}}}}, \boldsymbol{W}_{\boldsymbol{g}_t^{\mathrm{S}}}, \boldsymbol{b}_{\widehat{v_t^{\mathrm{S}}}}, \boldsymbol{b}_{\widehat{x_t^{\mathrm{S}}}}$为参数.

2.5.3 观点词特征

由于人们表达情感是通过观点抒发的，即观点词有助于情感的判断，所以本文模块中使用词性标注识别的观点词信息帮助情感的预测，首先使用简单的神经网络编码得到观点词特征表示$\boldsymbol{X}^{\mathrm{O}}$.

$$\boldsymbol{X}^{\mathrm{O}}=\tanh(\boldsymbol{W}_{\mathrm{S}}^{\mathrm{O}}\boldsymbol{X}^{\mathrm{S}}+\boldsymbol{b}_{\mathrm{S}}^{\mathrm{O}}),\tag{34}$$

其中$\boldsymbol{W}_{\mathrm{S}}^{\mathrm{O}},\boldsymbol{b}_{\mathrm{S}}^{\mathrm{O}}$为参数.

观点描述的短语通常由动词、副词、形容词、介词构成，比如“agree with”“run fast”“beautiful”等. 在获取第t个单词的上下文时，应该给予这些单词更多的权重. 此外，通常情况下，观点词会出现在描述对象的附近，因此，位置关系也可以被考虑. 基于上述的分析，为获取单词t对应的观点信息，本文模块使用单词t(假设为方面术语)引导的注意，并考虑形容词、副词、动词和介词的权重以及位置权重，最终得到单词t对应的观点特征.

$$\boldsymbol{z}_t^{\mathrm{O}}=\tanh(\boldsymbol{W}_{\boldsymbol{X}^{\mathrm{O}}}\boldsymbol{X}^{\mathrm{O}}\oplus(\boldsymbol{W}_{\mathrm{O},\boldsymbol{x}_t^{\mathrm{S}}}\boldsymbol{x}_t^{\mathrm{S}}+\boldsymbol{b}_{\mathrm{O},\boldsymbol{x}_t^{\mathrm{S}}})),\tag{35}$$

$$\widehat{z_{t,i}^{\mathrm{O}}}=z_{t,i}^{\mathrm{O}}\cdot\left[\log\left(2+|i-t|\right)\right]^{-1}\cdot\varphi_i,\tag{36}$$

$$\varphi_i=\begin{cases}w, \text{ if } w_i \text{ is the opinion},\\ 0, \text{ if } w_i \text{ is not the opinion},\end{cases}\tag{37}$$

$$\boldsymbol{\kappa}_t=softmax(\boldsymbol{W}_{\boldsymbol{\kappa}_t}\widehat{\boldsymbol{z}_t^{\mathrm{O}}}+\boldsymbol{b}_{\boldsymbol{\kappa}_t}),\tag{38}$$

$$\widehat{\boldsymbol{x}_t^{\mathrm{O}}}=\sum_i\boldsymbol{\kappa}_{t,i}\boldsymbol{x}_i^{\mathrm{O}},\tag{39}$$

其中$\boldsymbol{W}_{\boldsymbol{X}^{\mathrm{O}}},\boldsymbol{W}_{\mathrm{O},\boldsymbol{x}_t^{\mathrm{S}}},\boldsymbol{W}_{\boldsymbol{\kappa}_t},\boldsymbol{b}_{\mathrm{O},\boldsymbol{x}_t^{\mathrm{S}}},\boldsymbol{b}_{\boldsymbol{\kappa}_t}$为参数. w为超参，表示观点词的权重. $\widehat{\boldsymbol{x}_t^{\mathrm{O}}}$为最终的观点词特征.

2.5.4 情感分类

将多模态融合特征、情感特征、观点特征融合，输入到分类层，得到最后的情感分类结果为：

$$\widehat{\boldsymbol{m}_t^{\mathrm{S}}}=\boldsymbol{m}_t^{\mathrm{S}}\oplus\boldsymbol{x}_t^{\mathrm{S}}\oplus\widehat{\boldsymbol{x}_t^{\mathrm{O}}},\tag{40}$$

$$\widehat{\boldsymbol{M}^{\mathrm{S}}}=\{\widehat{\boldsymbol{m}_1^{\mathrm{S}}},\widehat{\boldsymbol{m}_2^{\mathrm{S}}},...,\widehat{\boldsymbol{m}_n^{\mathrm{S}}}\},\tag{41}$$

$$p\left(y_t^{\mathrm{S}}\middle|\widehat{\boldsymbol{m}_t^{\mathrm{S}}}\right)=softmax(\boldsymbol{W}_s\widehat{\boldsymbol{m}_t^{\mathrm{S}}}+\boldsymbol{b}_s),\tag{42}$$

其中$\boldsymbol{W}_{\mathrm{s}},\boldsymbol{b}_{\mathrm{s}}$为可训练参数.

2.6 模型训练

AESC模块的损失函数是最小化交叉熵损失，实验的目标是最小化这2个模块的加权损失，即

$$\mathcal{L}^{\mathrm{A}}=-\frac{1}{N}\sum_i\log p\left(y_i^{\mathrm{A}}\middle|\widehat{\boldsymbol{M}^{\mathrm{A}}}\right),\tag{43}$$

$$\mathcal{L}^{\mathrm{S}}=-\frac{1}{N}\sum_i\log p\left(y_i^{\mathrm{S}}\middle|\widehat{\boldsymbol{M}^{\mathrm{S}}}\right),\tag{44}$$

$$\mathcal{L}=\alpha_1\mathcal{L}^{\mathrm{A}}+\alpha_2\mathcal{L}^{\mathrm{S}},\tag{45}$$

$$\alpha_1+\alpha_2=1,\tag{46}$$

其中α_1,α_2为超参，为2个模块损失函数的权重.

2.7 方面-情感对提取

通过AESC模块，可分别获取句子的方面术语和情感标注序列，即$Y^{\mathrm{A}}=\left\{y_1^{\mathrm{A}},y_2^{\mathrm{A}},\cdots,y_i^{\mathrm{A}},\cdots,y_n^{\mathrm{A}}\right\},y_i^{\mathrm{A}}\in\{\mathrm{B},\mathrm{I},\mathrm{O}\}$和$Y^{\mathrm{S}}=\{y_1^{\mathrm{S}},y_2^{\mathrm{S}},\cdots,y_i^{\mathrm{S}},\cdots,y_n^{\mathrm{S}}\},y_{\mathrm{i}}^{\mathrm{S}}\in\{0,1,2,3\}$. 为了实现AESC任务的目标，本文进行方面-情感对抽取，具体的算法如算法1所示.

算法1. 方面-情感对抽取.

输入：句子长度L，方面术语标注序列Y^{A}，情感标注序列Y^{S}；

输出：方面-情感对Y^{P}.

① 令$Y^{\mathrm{P}}=[],i=0$;
② while $i<L$ do
③ if $Y^{\mathrm{A}}[i]==\mathrm{B}$ then
④ 令 $start=i$, $end=i$;
⑤ i+=1;
⑥ while $i<L$ and $Y^{\mathrm{A}}[i]==\mathrm{I}$ do
⑦ $end=i$;
⑧ i+=1;
⑨ end while
⑩ $Y^{\mathrm{P}}.append((start,end,Y^{\mathrm{S}}[start]))$;
⑪ else
⑫ i+=1;
⑬ end if
⑭ end while

3 实 验

3.1 数据集

为验证本文所提出的模型的有效性，本文使用了数据集Twitter2015[8]和Restaurant2014[20]进行实验. Twitter2015[8]是一个多模态数据集，其包含文本内容、图片、方面信息以及情感类别信息. Restaurant2014[20]属于文本领域的方面级情感分类数据集，其不包含图片信息. 本文数据集的训练集、测试集以及验证集与来源保持一致. 表1和表2分别是这2个数据集的

Table 1 Statistics of Twitter2015 Dataset

表1 Twitter2015数据集统计信息

数据集	情感数量			句子数量	方面数量
	POS	NEG	Neutral		
训练集	928	368	1 883	2 101	3 179
验证集	303	149	670	727	1 122
测试集	317	113	607	674	1 037

Table 2 Statistics of Restaurant2014 Dataset

表 2 Restaurant2014 数据集统计信息

数据集	评分等级数量				句子数量	方面数量
	level 1	level 2	level 3	level 4		
训练集	1 747	645	520	73	2 436	2 985
验证集	417	162	117	18	608	714
测试集	728	196	196	14	800	1 134

注：level 是按评分等级划分的数据集.

统计信息.

3.2 实现细节

为了初始化模型中的词嵌入式表示，本文使用了 Zhang 等人[7]在 3 000 万条推特上预训练好的 GloVe[29]词嵌入式词典. 词嵌入式表示的维度为 200，不在词典内的单词被随机初始化，并服从−0.25~0.25 的均匀分布. 字符嵌入式表示、词性嵌入式表示的维度分别为 30 和 16，且随机初始化服从−0.25~0.25 的均匀分布. 句子和单词最大的长度都取数据集中的最大值，不满足最大值的单词或句子采用填充的方式使所有单词或句子等长. BiLSTM 输出的隐藏向量维度为 200，方面术语提取模块的私有特征维度为 200，情感分类模块私有特征的维度为 100. 方面术语提取和情感分类 2 个模块的损失权重分别为 0.5 和 0.5. 训练过程中，周期(epoch)为 50，批大小为 20，优化器为 Adam，学习率为 0.001.

3.3 基线模型

在实验中用作对比的模型主要包括文本领域和多模态领域的模型.

3.3.1 文本领域

CMLA+TCap 和 DECNN+TCap.CMLA[30]和 DECNN[31]是方面术语提取任务中经典的模型，TCap[32]是方面级情感分类领先的方法，本文分别将 2 个方面术语提取模型和 1 个情感分类模型进行整合，形成 2 个流水线模型.

1)MNN[26]. 该模型是使用联合标注方法的方面术语提取和情感分类统一的模型.

2)E2E-AESC[33]. 该模型是使用联合标注方法，并以观点词提取为辅助任务的方面术语提取和情感分类统一的模型.

3)DOER[34]. 该模型是联合训练方面术语提取和情感分类的多任务统一框架.

4)RACL[21]. 是将方面术语提取、观点词提取、情感分类统一的多任务模型，该模型使用多层叠加的框架.

5)UMAS-Text. 该模型是本文提出的方面术语提取和方面级情感分类的统一框架，它将模型中关于视觉特征处理的网络层去除，变成处理纯文本数据的模型.

3.3.2 多模态领域

1)VAM[9].VAM 使用视觉注意机制和门控机制的多模态方面术语提取模型.

2)ACN[7].ACN 使用文本注意机制、视觉注意机制和门控机制的多模态方面术语提取模型.

3)UMT[10].UMT 使用 Bert 预训练模型表征文本的多模态方面术语提取模型.

4)Res-RAM 和 Res-MGAN.它们是 2 个方面级情感分类模型. 采用 Hazarika 等人[35]提出的多模态融合方法将视觉特征和 RAM[36]或 MGAN[37]的文本特征融合，最后采用 softmax 层分类.

5)Res-RAM-TFN 和 Res-MGAN-TFN.它们是采用 Zadeh 等人[5]提出的多模态融合方法将视觉特征和 RAM 或 MGAN 的文本特征融合进行方面级情感分类的模型.

6)MIMN[38]. MIMN 是采用多跳记忆网络建模方面术语、文本和视觉之间交互关系的方面级情感分类模型，具有较高的性能.

7)EASFN[8]. EASFN 是目前多模态领域最新的方面级情感分类模型.

8)ACN-ESAFN. ACN-ESAFN 是使用 ACN[7]获取方面术语、ESAFN[8]获取方面级情感的流水线模型.

9)UMT-ESAFN. UMT-ESAFN 是使用 UMT[10]获取方面术语、ESAFN[8]获取方面级情感的流水线模型.

10)UMAS-AE. UMAS-AE 是将本文提出的模型中的共享特征模块和方面术语提取模块组合成单任务的方面术语提取模型.

11)UMAS-SC. UMAS-SC 是将本文提出的模型中的共享特征模块和情感分类模块组合成单任务的方面级情感分类模型.

12)UMAS-Pipeline. UMAS-Pipeline 是将独立的 UMAS-AE 和 UMAS-SC 模型使用流水线方式合并而成的模型.

13)UMAS: UMAS 是本文提出的多模态方面术语提取和方面级情感分类的统一框架，由 2 个模块共享浅层的特征表示.

3.4 评价指标

本文使用精确率(precision, P)、召回率(recall, R)、$F1$ 评价方面术语提取模型的性能，以下简记为 AE-P、AE-R、AE-$F1$；使用准确率(accuracy, ACC)、

*F*1 评价情感分类的性能，简记为 SC-*ACC*, SC-*F*1；使用 *F*1 评价方面-情感对提取的性能，简记为 AESC-*F*1，即当且仅当方面术语提取和情感预测同时正确时记为预测正确.

3.5 实验结果

3.5.1 与基线模型的对比

表 3 报告了本文所提出的模型 UMAS 在文本领域与现有方法的性能对比. 在文本数据集 Restaurant 2014 上，UMAS 的 *F*1 在方面术语提取、情感分类 2 个子任务上相较于第 2 优秀的模型 RACL-GloVe 的 *F*1 值分别提升了 0.21 个百分点和 1.9 个百分点，且方面-情感对的提取表现也是最好的. 说明 UMAS 在删除视觉处理的相关网络后，在文本领域也具有良好的表现.

Table 3 Performance Comparison of UMAS-Text and Existing Methods on Restaurant2014 Dataset

表 3 Restaurant2014 数据集上 UMAS-Text 与现有方法的性能对比 %

模型	AE -*F*1	SC-*F*1	AESC-*F*1
CMLA+TCap	81.91	71.32	65.68
DECNN+TCap	82.79	71.77	66.84
MNN	83.05	68.45	63.87
E2E-AESC	83.92	68.38	66.6
DOER	84.63	64.5	68.55
RACL	85.37	74.46	70.67
UMAS-Text	**85.58**	**76.36**	**70.70**

注：加粗数字表示最优结果.

表 4 和表 5 报告了 UMAS 在多模态领域与现有方法在方面术语提取和方面级情感分类 2 个子任务上的性能对比. 在多模态数据集 Twitter2015 上，UMAS 与当前 3 个方面术语提取模型相比，*F*1 值分别提升了 21.78 个百分点、4.25 个百分点、0.15 个百分点，比使用 BERT 预训练的方面术语提取模型 UMT 略有优势. 方面术语提取的 *P* 值比 ACN 高了 1.99 个百分点. 然而 *R* 值比 UMT 模型低了 2.22 个百分点. 这一定程度上体现了 UMAS 相对于 UMT 在识别方面时边界更加严格，提升了 *P* 值的同时损失了 *R* 值. 在情感分类任务中，UMAS 的性能超过了所有的基线模型，比当前最新的模型 ESAFN 的 *F*1 值提高了 5.97 个百分点、*ACC* 提高了 0.1 个百分点.

Table 4 Performance Comparison of AE on Twitter2015 Dataset

表 4 Twitter2015 数据集上 AE 性能对比 %

模型	AE -*P*	AE -*R*	AE -*F*1
VAM	58.10	56.70	57.39
ACN	79.10	71.17	74.92
UMT	78.50	**79.56**	79.02
UMAS（本文）	**81.09**	77.34	**79.17**

注：加粗数字表示最优结果.

Table 5 Performance Comparison of SC on Twitter2015 Dataset

表 5 Twitter2015 数据集上 SC 性能对比 %

模型	SC-*ACC*	SC-*F*1
Res-RAM	71.55	64.68
Res-RAM-TFN	69.91	61.49
Res-MGAN	71.65	63.88
Res-MGAN-TFN	70.3	64.14
MIMN	71.84	65.69
EASFN	73.38	67.37
UMAS（本文）	**73.48**	**73.34**

注：加粗数字表示最优结果.

表 6 报告了 UMAS 和当前多模态流水线方法的性能对比. UMAS 在多模态数据集上提取方面-情感对的 *F*1 值为 58.05%，分别高于现有流水线方法 2.49 个百分点和 1.16 个百分点，且时间效率是现有方法的 16.3 倍和 16 倍，体现了本文所提出的统一框架具有最优的性能.

表 7 报告了 UMAS 和单任务模型的性能对比. 结果表明，UMAS 相比于方面术语提取和情感分类

Table 6 Performance Comparison of AESC on Twitter2015 Dataset

表 6 Twitter2015 数据集上 AESC 性能对比

模型	AESC-*F*1/%	运行时间/s
ACN-ESAFN	55.56	163
UMT-ESAFN	56.89	160
UMAS（本文）	**58.05**	**10**

注：加粗数字表示最优结果.

Table 7 Comparison of Unified Model and Single-Task Model

表 7 统一框架和单任务模型的对比 %

模型	AE-*P*	AE-*R*	AE-*F*1	SC-*ACC*	SC-*F*1	AESC-*F*1
UMAS-AE	78.30	80.04	79.16			
UMAS-SC				71.26	70.79	
UMAS-Pipeline	78.30	**80.04**	79.16	71.26	70.79	56.76
UMAS（本文）	**81.09**	77.34	**79.17**	**73.48**	**73.34**	**58.05**

注：加粗数字表示最优结果.

单任务模型，性能都有一定的提升，*F*1 值分别提升了 0.01 个百分点和 2.55 个百分点，方面术语提取的 *ACC* 提升了 2.79 个百分点，情感分类的 *ACC* 提升了 2.22 个百分点. 然而，UMAS 中方面术语提取的 *R* 值相对于单任务下降了 2.7 个百分点，这可能是因为在 UMAS 中方面的特征表示受到了情感模块的影响. 此外，UMAS 的 AESC 性能与 2 个单任务串联的流水线模型对比，UMAS 对方面-情感对提取性能有 1.29 个百分点的提升. 结果表明了底层的特征共享对 2 个子任务的性能提升都有帮助，通过建立 2 个任务之间的语义联系有利于提高方面-情感对提取的准确率.

结合表 4、表 5、表 7，可以看出本文的方面术语提取单任务模型比 ACN 的性能高了 4.24 个百分点，验证了词性特征对方面术语提取的重要影响. 相比于其他方面级情感分类，本文的单任务情感分类模型也有较大的改善，说明观点词和位置信息对情感分类有一定的帮助.

3.5.2 消融实验

首先介绍 UMAS 的 7 个变体模型.

1）UMAS-no_visual. 删除视觉特征.

2）UMAS-no_POS_features. 删除词性特征.

3）UMAS-no_opinion. 删除情感分类模块中观点词特征.

4）UMAS-no_self_attention. 删除情感分类模块中情感特征的自注意机制.

5）UMAS-no_gate_fusion. 将情感分类模块中私有特征获取部分的门控融合机制改为直接拼接操作.

6）UMAS-special. 只保留情感模块中私有特征部分中的特有情感特征，删除共享文本特征.

7）UMAS-share. 只保留情感模块中私有特征部分中的共享文本特征，删除特有情感特征.

表 8 报告了变体模型的性能. 通过分别消除视觉特征、词性特征、观点特征、情感模块的自注意机制、情感模块私有特征的门控融合机制、情感模块的共享文本特征、情感模块的特有特征，验证了各个部分存在的作用. 由于 2 个模块之间存在参数的共享，所以一个模块的结构的变化不仅影响自身，而且影响另一个模块. 表 8 的第 1 行和最后 1 行的对比显示了视觉特征对方面术语提取和情感分类模块都有明显的性能提升，*F*1 值分别提升了 2.45 个百分点和 2.07 个百分点. 情感分类模块中的观点词特征将方面级情感分类的性能整体提升了 2.61 个百分点. 情感模块的自注意机制对该模块的性能有 2.83 个百分点的提升. 情感模块私有特征获取的门控融合机制，既考虑了方面对情感预测的影响，也考虑了情感特征本身的重要性，将情感分类的 *F*1 提升了 3.59 个百分点，AESC 性能提升了 2.35 个百分点. 根据表 8 最后 3 行的结果，可以看出在情感分类模块中的私有特征部分单独使用共享特征或特有特征的效果都不好，将这二者融合是最佳的选择.

Table 8 Results of Ablation Experiment

表 8 消融实验结果 %

模型	AE-*P*	AE-*R*	AE-*F*1	SC-*ACC*	SC-*F*1	AESC-*F*1
UMAS-no_visual	77.67	75.80	76.72	71.26	71.27	54.76
UMAS-no_POS_features	76.59	77.63	77.11	71.26	70.73	54.69
UMAS-no_opinion	75.16	79.36	77.20	73.00	72.28	55.44
UMAS-no_self_attention	75.87	**79.46**	77.63	71.36	70.51	55.77
UMAS-no_gate_fusion	75.30	78.78	77.02	71.26	69.75	55.70
UMAS-special	76.46	77.05	76.75	71.36	71.36	54.76
UMAS-share	75.44	78.78	77.08	68.27	67.91	52.55
UMAS（本文）	**81.09**	77.34	**79.17**	**73.48**	**73.34**	**58.05**

注：加粗数字表示最优结果.

3.5.3 补充实验

为了说明情感分类模块私有特征部分不同选择的不同效果，本节进行了相关的可视化分析. 首先，情感分类模块的私有特征可以有 3 种选择：情感模块私有文本编码器输出的特有情感表示、共享文本编码器输出的共享文本表示、特有情感表示和共享文本表示的融合特征. 为了方便说明，将这 3 种特征对应的模型记为 UMAS-special, UMAS-share, UMAS-combine. 表 9 说明了图 3、图 4 涉及的统计量的含义.

图 3 显示了不同情感私有特征表示的结果. 首先，在 AE 模块，UMAS-combine 预测正确且 UMAS-special 预测错误的数量为 128，而 UMAS-combine 预测错误且 UMAS-special 预测正确的数量为 99，说明 UMAS-combine 对 UMAS-special 的纠正能力要强于 UMAS-special 对 UMAS-combine 的纠正能力，即 UMAS-combine 模型的性能较优越. 通过图 3 中其他数据的对比分析，可以发现无论是对方面术语提取还是情感分类，UMAS-combine 的性能总是要强于 UMAS-special 和 UMAS-share. 其次，在情感分类模块，UMAS-special 预测正确而 UMAS-share 预测错误的数量为 83，而 UMAS-share 预测正确而 UMAS-special 预测错误的数量为 53，体现了特有情感特征和共享特征对情感模块性能的不同贡献. 图 4 展示了特有情感特征

Table 9 Instruction of Statistics

表 9 统计量说明

统计对象	统计量	说明
combine_true_special_wrong	UMAS-combine 预测正确而 UMAS-special 预测错误的数量.	体现了 UMAS-combine 对 UMAS-special 的纠正能力.
combine_wrong_special_true	UMAS-combine 预测错误而 UMAS-special 预测正确的数量.	
combine_true_share_wrong	UMAS-combine 预测正确而 UMAS-share 预测错误的数量.	
combine_wrong_share_true	UMAS-combine 预测错误而 UMAS-share 预测正确的数量.	
special_contribution	UMAS-share 预测错误而 UMAS-special 预测正确的数量.	体现了 UMAS-special 的特殊贡献.
share_contribution	UMAS-special 预测错误而 UMAS-share 预测正确的数量.	

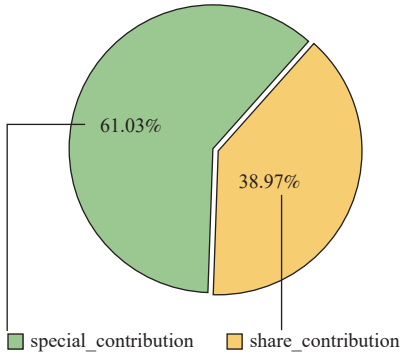


Fig. 4 Different representations contribute to sentiment classification

图 4 不同表示对情感分类的贡献

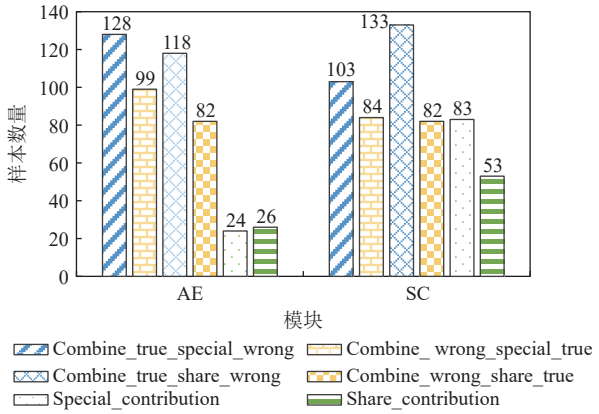


Fig. 3 Result comparison of different sentiment private features

图 3 不同情感私有特征的结果对比

和共享特征对情感模块的不同贡献程度，特有情感特征的贡献约为 60%，共享特征的贡献程度约为 40%.

综上体现了将特有情感特征和共享文本特征进行动态融合的必要性，且特有情感特征对方面级情感分类的贡献比较突出. 同时，也说明了方面术语提取和方面级情感分类 2 个任务之间既有联系又有区别，既要考虑 2 个任务之间的交互关系，又要充分考虑任务本身的特征.

4 总结与展望

为了解决目前 AESC 任务流水线方法的不足，本文提出了多模态方面术语提取和方面级情感分类的统一框架 UMAS. 该统一框架使用 3 个共享编码器，即文本、图像、词性编码器构建方面术语提取模块和情感分类模块底层的共享特征模块. 该共享特征模块不仅使模型在训练过程中学习到 2 个任务之间的语义联系，而且简化了模型. 同时，该统一框架能并行地执行 2 个子任务，同时输出句子中的多个方面及其对应的情感类别，解决了流水线方法效率低的问题. 此外，本文通过词性标注获取单词的词性，并使用多头自注意机制获取词性特征，将视觉特征、文本特征、词性特征融合作为方面术语提取模块解码器的输入，提升了方面术语提取的性能. 在情感分类模块，本文使用词性识别句子中的观点词，在情感分析中增加对这些观点词的注意权重并考虑位置信息以提升情感分类的性能. 本文所提出的统一框架在 Twitter2015和 Restaurant2014 这 2 个数据集上相比于其他基线模型都有良好的表现.

随着 transformer, BERT 等技术的不断发展，在未来的研究中可以考虑将预训练技术加入到本文模型中以获得更好的特征表示.

作者贡献声明：周如提出了算法思路和撰写论文；朱浩泽提出了实验方案并负责完成实验；郭文雅、于胜龙、张莹提出指导意见并修改论文.

参 考 文 献

[1] Ju Xincheng, Zhang Dong, Xiao Rong, et al. Joint multi-modal aspect-sentiment analysis with auxiliary cross-modal relation detection[C] // Proc of the 26th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2021: 4395−4405

[2] Li Z Y, Cheng Wei, Kshetramade R, et al. Recommend for a reason: Unlocking the power of unsupervised aspect-sentiment co-extraction[C]// Proc of the 26th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2021: 763−778

[3] Gong Chenggong, Yu Jianfei, Xia Rui. Unified feature and instance based domain adaptation for aspect-based sentiment analysis[C] // Proc of the 25th Conf on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2020: 7035−7045

[4] Cai Guoyong, Xia Binbin. Convolutional Neural Networks for Multimedia Sentiment Analysis[M] //Natural Language Processing and Chinese Computing. Cham: Springer, 2015: 159−167

[5] Zadeh A, Chen Minghai, Poria S, et al. Tensor fusion network for multimodal sentiment analysis[C] //Proc of the 22nd Conf on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2017: 1103−1114

[6] Mai Sijie, Hu Haifeng, Xing Songlong. Divide, conquer and combine: hierarchical feature fusion network with local and global perspectives for multimodal affective computing[C] //Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 481−492

[7] Zhang Qi, Fu Jinlan, Liu Xiaoyu, et al. Adaptive co-attention network for named entity recognition in tweets[C] //Proc of the 32nd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2018: 5674−5681.

[8] Yu Jianfei, Jiang Jing, Xia Rui. Entity-sensitive attention and fusion network for entity-level multimodal sentiment classification[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 429−439

[9] Lu Di, Neves L, Carvalho V, et al. Visual attention model for name tagging in multimodal social media[C] //Proc of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2018: 1990−1999

[10] Yu Jianfei, Jiang Jing, Yang Li, et al. Improving multimodal named entity recognition via entity span detection with unified multimodal transformer[C] //Proc of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2020: 3342−3352

[11] Honnibal M, Montani I, Van Landeghem S, et al. spaCy: Industrial-strength natural language processing in Python[EB/OL]. [2022-05-17].https://spacy.io

[12] Liu Lulu, Yang Yan, Wang Jie. ABAFN: Aspect-based sentiment analysis model for multimodal[J]. Computer Engineering and Applications, 2022, 58(10): 193−199 (in Chinese)
(刘路路, 杨燕, 王杰. ABAFN: 面向多模态的方面级情感分类模型[J]. 计算机工程与应用, 2022, 58(10): 193−199)

[13] Li Ruifan, Chen Hao, Feng Fangxiang, et al. Dual graph convolutional networks for aspect-based sentiment analysis[C] //Proc of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2021: 6319−6329

[14] Xiao Zeguan, Wu Jiarun, Chen Qingliang, et al. BERT4GCN: Using BERT intermediate layers to augment GCN for aspect-based sentiment classification[J]. arXiv preprint, arXiv: 2110.00171, 2021

[15] Qi Songzhe, Huang Xianying, Sun Haidong, et al. Aspect based sentiment analysis with progressive enhancement and graph convolution[J]. Application Research of Computers, 2022, 39(7): 2037−2042 (in Chinese)
(齐嵩喆, 黄贤英, 孙海栋, 等. 基于渐进增强与图卷积的方面级情感分类模型[J]. 计算机应用研究, 2022, 39(7): 2037−2042)

[16] Han Hu, Hao Jun, Zhang Qiankun, et al. Knowledge-enhanced interactive attention model for aspect-based sentiment analysis[J/OL]. Journal of Frontiers of Computer Science and Technology, 2022 [2021-12-31]. http://fcst.ceaj.org/CN/10.3778/j.issn.1673−9418.2108082 (in Chinese)
(韩虎, 郝俊, 张千锟, 等. 知识增强的交互注意力方面级情感分类模型[J]. 计算机科学与探索, 2022 [2021-12-31]. http://fcst.ceaj.org/CN/10.3778/j.issn.1673−9418.2108082)

[17] Mao Tengyue, Zheng Zhipeng, Zheng Lu. Aspect-level sentiment analysis based on improved self-attention mechanism[J]. Journal of South Central University for Nationalities: Natural Science Edition, 2022, 41(1): 94−100(in Chinese)
(毛腾跃, 郑志鹏, 郑禄. 基于改进自注意力机制的方面级情感分类[J]. 中南民族大学学报: 自然科学版, 2022, 41(1): 94−100)

[18] Sun Xiaowan, Wang Ying, Wang Xin, et al. Aspect-based sentiment analysis model based on dual-attention networks[J]. Journal of Computer Research and Development, 2019, 56(11): 2384−2395 (in Chinese)
(孙小婉, 王英, 王鑫, 等. 面向双注意力网络的特定方面情感分析模型[J]. 计算机研究与发展, 2019, 56(11): 2384−2395)

[19] Ying Chengcan, Wu Zhen, Dai Xinyu, et al. Opinion transmission network for jointly improving aspect-oriented opinion words extraction and sentiment classification[C] // Proc of the 9th CCF Int Conf on Natural Language Processing and Chinese Computing. Cham: Springer, 2020: 629−640

[20] Oh S, Lee D, Whang T et al. Deep context- and relation-aware learning for aspect-based sentiment analysis[C] //Proc of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2021: 495−503

[21] Chen Zhuang, Qian Tieyun. Relation-aware collaborative learning for unified aspect-based sentiment analysis[C] //Proc of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2020: 3685−3694

[22] Xu Lu, Li Hao, Lu Wei, et al. Position-aware tagging for aspect sentiment triplet extraction[C]//Proc of the 25th Conf on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2020: 2339−2349

[23] Phan M H, Ogunbona P O. Modelling context and syntactical features for aspect-based sentiment analysis[C] //Proc of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2020: 3211−3220

[24] Xue Fang, Guo Yi, Li Zhiqiang, et al. Aspect-level sentiment analysis based on double-layer part-of-speech-aware and multi-head interactive attention mechanism[J]. Application Research of Computers, 2022, 39(3): 704−710 (in Chinese)
(薛芳, 过弋, 李智强, 等. 基于双层词性感知和多头交互注意机制

的方面级情感分类[J]. 计算机应用研究, 2022, 39(3): 704−710)

[25] He Ruidan, Lee W S, Ng H T, et al. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis[C] //Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 504−515

[26] Wang Feixiang, Lan Man, Wang Wenting. Towards a one-stop solution to both aspect extraction and sentiment analysis tasks with neural multi-task learning[C/OL]// Proc of the 2018 Int Joint Conf on Neural Networks (IJCNN). Piscataway, NJ : IEEE, 2018 [2022-03-01].https://ieeexplore.ieee.org/abstract/document/8489042

[27] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint, arXiv: 1409.1556, 2014

[28] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C] // Proc of the 31st Conf on Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2017: 5998−6008

[29] Jeffrey P, Richard S, Christopher M. GloVe: Global vectors for word representation[C] //Proc of the 19th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2014: 1532−1543

[30] Wu Yuanbin, Zhang Qi, Huang Xuanjing, et al. Phrase dependency parsing for opinion mining[C] //Proc of the 14th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2009: 1533−1541

[31] Xu Hu, Liu Bing, Shu Lei, et al. Double embeddings and CNN-based sequence labeling for aspect extraction[C] //Proc of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2018: 592−598

[32] Chen Zhuang, Qian Tieyun. Transfer capsule network for aspect level sentiment classification[C] //Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 547−556

[33] Li Xin, Bing Lidong, Li Piji, et al. A unified model for opinion target extraction and target sentiment prediction[C] //Proc of the 33rd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2019: 6714−6721

[34] Luo Huaishao, Li Tianrui, Liu Bing, et al. DOER: Dual cross-shared RNN for aspect term-polarity co-extraction[C] //Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 591−601

[35] Hazarika D, Poria S, Zadeh A, et al. Conversational memory network for emotion recognition in dyadic dialogue videos[C] //Proc of the 16th Conf on Association for Computational Linguistics, North American Chapter. Stroudsburg, PA: ACL, 2018: 2122−2132

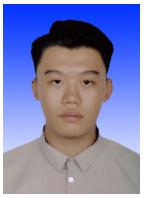
[36] Chen Peng, Sun Zhongqian, Bing Lidong, et al. Recurrent attention network on memory for aspect sentiment analysis[C]// Proc of the 22nd Conf on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2017: 452−461

[37] Fan Feifan, Feng Yansong, Zhao Dongyan. Multi-grained attention network for aspect-level sentiment classification[C] //Proc of the 23rd Conf on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2018: 3433−3442

[38] Xu Nan, Mao Wenji, Chen Guandan. Multi-interactive memory network for aspect based multimodal sentiment analysis[C] //Proc of the 33rd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2019: 371−378

Zhou Ru, born in 1999. Master candidate. Her main research interests include sentiment analysis and information retrieval.

周　如, 1999 年生. 硕士研究生. 主要研究方向为情感分析、信息检索.



Zhu Haoze, born in 2001. Undergraduate. His main research interests include sentiment analysis and information retrieval.

朱浩泽, 2001 年生. 本科生. 主要研究方向为情感分析、信息检索.

Guo Wenya, born in 1994. PhD. Her main research interests include sentiment analysis and data mining.

郭文雅, 1994 年生. 博士. 主要研究方向为情感分析、数据挖掘.

Yu Shenglong, born in 1998. Master candidate. His main research interests include data mining and reinforcement learning.

于胜龙, 1998 年生. 硕士研究生. 主要研究方向为数据挖掘、强化学习.

Zhang Ying, born in 1986. PhD, professor. Her main research interests include sentiment analysis, data mining, and information retrieval.

张　莹, 1986 年生. 博士, 教授. 主要研究方向为情感分析、数据挖掘、信息检索.