

基于文本生成的多粒度评论情感分析

张佳威, 王中卿, 陈嘉沥

引用本文

张佳威, 王中卿, 陈嘉沥. 基于文本生成的多粒度评论情感分析[J]. 计算机科学, 2025, 52(10): 239-246.

ZHANG Jiawei, WANG Zhongqing, CHEN Jiali. [Multi-grained Sentiment Analysis of Comments Based on Text Generation](#) [J]. Computer Science, 2025, 52(10): 239-246.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于大型语言模型文本简化的细粒度情感分析](#)

Text Simplification for Aspect-based Sentiment Analysis Based on Large Language Model
计算机科学, 2025, 52(10): 258-265. <https://doi.org/10.11896/jsjcx.250100114>

[一种新的基于凸损失函数的离散扩散文本生成模型](#)

Novel Discrete Diffusion Text Generation Model with Convex Loss Function
计算机科学, 2025, 52(10): 231-238. <https://doi.org/10.11896/jsjcx.240800147>

[基于提示学习与超图的事件因果关系识别模型](#)

Event Causality Identification Model Based on Prompt Learning and Hypergraph
计算机科学, 2025, 52(9): 303-312. <https://doi.org/10.11896/jsjcx.240800121>

[多尺度多粒度解耦蒸馏模糊分类器及其在癫痫脑电信号检测中的应用](#)

Multi-scale Multi-granularity Decoupled Distillation Fuzzy Classifier and Its Application in Epileptic EEG Signal Detection
计算机科学, 2025, 52(9): 37-46. <https://doi.org/10.11896/jsjcx.250300096>

[结合预训练模型和数据增强的跨领域属性级情感分析研究](#)

Cross-domain Aspect-based Sentiment Analysis Based on Pre-training Model with Data Augmentation
计算机科学, 2025, 52(8): 300-307. <https://doi.org/10.11896/jsjcx.240900114>

基于文本生成的多粒度评论情感分析

张佳威 王中卿 陈嘉沥

苏州大学计算机科学与技术学院 江苏 苏州 215006

(20235227002@stu.suda.edu.cn)

摘要 随着社交媒体和在线评论平台的兴起,自动化的情感分析成为了理解公众情绪、消费者偏好及市场趋势的关键工具。传统的情感分析方法往往使用分类模型关注于提取文本的总体情绪倾向,忽视了评论中可能蕴含的复杂且多维度的情感信息。针对这一问题,提出了一种基于文本生成的多粒度评论情感分析模型,旨在细致地捕捉评论文本中方面级的情感和文档级的情感。同时,构建了一种结构化输出格式,其同时包含评论文本针对不同方面的情感标签和评论文本的总体情感标签。与传统的分类模型相比,所提模型通过不同的生成方式更全面地理解和反映了文本的情感结构,实现了对评论中多方面情感信息和总体情感的抽取和分类。实验结果表明,所提模型在总体情感和方面情感的识别中优于常规的分类方法,较 Bert+LSTM 模型 F1 值提升了 4.4%。

关键词: 自然语言处理; 文本生成; 结构化输出; 多粒度; 评论情感分析

中图分类号 TP391

Multi-grained Sentiment Analysis of Comments Based on Text Generation

ZHANG Jiawei, WANG Zhongqing and CHEN Jiali

School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

Abstract With the rise of social media and online review platforms, automated sentiment analysis has become a key tool for understanding public emotions, consumer preferences, and market trends. Traditional sentiment analysis methods often use classification models that focus on extracting the overall sentiment of the text, neglecting the complex and multidimensional emotional information that may be contained within the comments. Addressing this issue, this study proposes a multi-granularity text-based sentiment analysis model using generative models to intricately capture aspect-level and document-level emotions in review texts. Additionally, a structured output format is constructed that includes sentiment labels for different aspects of the review text as well as the overall sentiment label of the review text. Compared to traditional classification models, the proposed model more comprehensively understands and reflects the emotional structure of text, achieving extraction and classification of multifaceted emotional information and overall sentiment in comments. Experimental results show that the proposed model is better than conventional classification methods in the recognition of overall emotions and aspect emotions, and achieves a 4.4% higher F1-Score than the Bert+LSTM model.

Keywords Natural language processing, Text generation, Structured output, Multi-grained, Review sentiment analysis

在数字化时代背景下,互联网的普及和社交媒体平台的兴起带来了数据的爆炸性增长。这些数据,尤其是针对各类产品和服务的在线评论,对于消费者和商家具有重要作用。产品评论不仅反映了消费者的总体满意度,而且能够揭示消费者对产品不同方面的细致评价,如品质、性能、设计和用户体验等,从而帮助消费者做出有效的购买决策。通过分析这些评价,可以深入理解消费者的偏好和行为,从而对各行各业的产品开发和市场策略产生重要影响,帮助商家做出有效的商业决策^[1]。

在线评论情感分类是情感分类领域的一个重要任务。随着商品评论数量的不断增加,使用人工方式对海量的评论信

息进行挖掘是不切实际的,需要耗费大量的人力和物力,因此,使用计算机构建相应的情感分析方法对商品评论进行高效准确的情感分析具有重要意义^[2]。同时,一个产品往往拥有多方面的属性,产品的评论中也会包含产品的不同属性的信息以及对这些属性的相关评价,这些评价反映了用户对于产品各方面属性的情感极性。产品评论信息的多样性和复杂性,也为人们发掘产品评论中潜在的商业价值带来了巨大的挑战。

表 1 列出了一条包含产品多方面评论信息的评论示例,其中的文本包含了对酒类产品的外观、香气、口感、味道的评价,并给出了总体评分。例如,“略显浑浊的橙金色”是描述外

到稿日期:2024-08-05 返修日期:2024-11-07

基金项目:国家自然科学基金(62076175,61976146)

This work was supported by the National Natural Science Foundation of China(62076175,61976146).

通信作者:王中卿(wangzq@suda.edu.cn)

观的,评论中对各个方面都进行了描述,最后给出了总体评价描述“非常适口”。评价者也对该酒类的其他方面进行了打分,同时给出了总体评分。该产品各方面的评价综合构成了总体评价。综合各方面描述和总体描述,该产品的总体评价为积极的。

表1 评论示例
Table 1 Review example

属性	评价
商品评论	这是一种购买自 LCBO 的啤酒,保质期至 2011 年 4 月。啤酒倒出来呈现一种略显浑浊的橙金色。白色的泡沫很快消散,只留下一圈细小的泡沫和少量的挂杯。从香气上来说,主要是香蕉和橙子的味道,还带有一点丁香的气息,我很喜欢。味道上比较简单,典型的小麦啤酒风格,有面包味、香蕉和橙子的果味以及泡泡糖的甜味,最后还有一点丁香的味。唯一的不足是味道似乎有些淡,缺乏一些复杂性。口感很好,我喜欢有嚼劲的小麦啤酒。非常适口,一整天喝都没问题。
外观	3.5/5
香气	4/5
口感	4/5
味道	4/5
总体评价	4.5/5

国内外的研究者们对评论情感分析也有研究,并取得了一定的效果。但在以往的研究中,分类式模型(如 Bert)通常把评论文本看成输入,并生成一个特定的情感标签,这种粗粒度的情感分类难以同时挖掘文本中的细分方面的情感极性,也难以对一个产品的不同方面同时进行细致评价。对于生成式模型,已经有学者对文本进行细粒度的情感分析,表现为对文本中的具体方面进行情感极性的分析。实际上,产品的不同方面特性对评论的整体评价是有一定影响的,产品各个方面的评价越高,整体评价往往就越好。

因此,本文不再只对整个评论文本进行整体评价,而是在对产品评论的多个方面信息进行更细粒度评价的同时,生成对产品的整体评价。为了同时生成多方面评价和整体评价,本文提出了一种基于文本生成的多粒度情感分析模型(Multi-aspect Evaluation and Integrated Assessment T5, MEIA-T5)。具体来说,将商品评论输入到 T5 模型中。T5 模型是一个基于 Transformer 架构的文本到文本的生成模型,通过训练能够利用商品评论的输入得到特定格式的输出文本。对于这种输出,本文设计了一种统一的输出格式来同时表示方面级情感极性和整体情感极性,将每个方面属性的情感极性以特定的标签或短语形式表示,并将这些表示与整体情感极性的表示结合起来,形成一个完整的输出序列。这样,模型在生成输出时就能够同时考虑到不同方面属性的情感极性和整体情感极性,从而实现多粒度的情感分析,并提升评论整体情感分类的效果。通过设计相关实验,验证了同时生成产品多方面属性评价对于提升产品评论情感分类的效果具有明显的作用。

本文的主要贡献如下:

- 1)提出了一种基于文本生成的多粒度情感分析模型 MEIA-T5。相较于分类式的情感分析模型,MEIA-T5 同时考虑了产品评论中不同方面的情感极性和整体的情感极性。
- 2)对于方面级情感极性和整体情感极性,设计了一种统一的输出格式来进行多粒度情感极性的生成。
- 3)在改进的 Beer 数据集上对本文模型的有效性进行了

实验验证。实验结果表明,本文模型优于基准方法。

1 相关工作

情感分析是自然语言处理领域一个重要的研究方向。根据情感分析的细节和层次可分为粗粒度和细粒度的情感分析,粗粒度的情感分析主要识别文档级和句子级的总体情感,细粒度的情感分析则分析句子中具体短语或具体方面的情感。随着深度学习在各个领域取得巨大进展,利用深度学习进行情感分析已经逐步成为研究热点。

1.1 粗粒度的情感分析

粗粒度的情感分析分为文档级和句子级,主要分析篇章或者句子的总体情感倾向。Huang 等^[3]提出了一种基于卷积神经网络(CNN)的情感分析模型,对旅游目的地评论数据集进行了情感分析,取得了不错的效果。Bhatia 等^[4]提出了一种基于修辞结构理论(Rhetorical Structure Theory)的递归神经网络(RNN),其根据句子成分在修辞结构中的依赖表示进行重新加权,相较于基于分类的方法有显著的改进。Tang 等^[5]提出了一种基于门控递归神经网络的模型,利用门控递归神经网络对句子语义及其关系进行自适应编码,在来自 IMDB 和 Yelp 的 4 个大型数据集上进行了文档级情感分类。Baziotis 等^[6]提出的基于 LSTM 的模型,增强了注意机制,并在大量的推文中进行了预训练的词嵌入,在推文情感分析任务中取得了较好的效果。Rehman 等^[7]将长短期记忆网络(LSTM)和卷积神经网络结合,提出了一种 hybrid CNN-LSTM 的混合模型,该模型在 IMDB 电影评论数据集和亚马逊电影评论数据集上取得了一定的效果。Sasikala 等^[8]提出了一种深度学习改进的神经网络,其通过改进的神经模糊推理系统,对预测产品进行加权因子计算和分类,解决了在线产品情感分析准确率低的问题。Cao 等^[9]提出了一种基于改进的对称结构 Bert 模型的情感分析方法,以识别农产品评价中消费者的情感倾向。这些方法大都侧重于篇章或者文本的整体情感。

1.2 细粒度的情感分析

细粒度的情感分析可以分为短语级和方面级。相较于粗粒度的情感分析,细粒度的情感分析能够更深入地分析句子或篇章的具体部分的情感。Yin 等^[10]提出了 SentiBert,该模型将上下文表示与二元选区解析树相结合,以捕获语义组合,在短语级情感分析中取得了较好的性能。方面级的情感分析是近期的研究热点。Zhao 等^[11]从联合项和关系抽取的角度出发,提出了一种端到端的方法用于成对的方面项和意见项的提取任务。Cai 等^[12]提出了一个层次图卷积网络,其中低层次的图卷积网络对多个类别的内在关系进行建模,高层次的图卷积网络捕获方面类别和情感之间的相互关系,这种层次输出结构相较于已有的输出结构有显著的提升。Liu 等^[13]利用预训练的语言模型 Bert,将方面级情感分析的任务转换成自然语言生成任务,取得了较好的效果。随着方面级情感分析研究的深入,Cai 等^[14]引入了将评论中方面-类别-意见-情感进行九重提取的任务,以识别评论中隐含的方面和意见,通过在不同基准模型上的对比,证明了该任务在隐含方面和隐含意见上存在优势。

1.3 文档级多方面情感分类

文档级多方面情感分类旨在预测评论在给定方面的评级或者情感。它起源于在线评论系统,要求用户对产品和服务提供方面级评级。早期的工作大多先提取每个方面的特征,再使用回归技术来预测方面级的评分。Lu 等^[15]使用弱监督主题模型,先通过少量的先验知识进行主题与方面的对应,从而对句子进行标注,再使用总体评分和句子标注结果预测各方面评分。近年来,深度学习模型将文档级多方面情感分类表述为多任务分类问题。Li 等^[16]提出了一种层次用户方面评分网络,采用层次结构编码词、句子和文档级别的信息,结合用户偏好和整体评分进行方面评分。Zhang 等^[17]提出了一种基于注意力记忆网络的文档级多方面情感分析模型,采用基于注意力的记忆网络来构建方面和句子记忆,利用多跳注意力记忆网络将相邻的方面相关信息纳入最终的方面评级预测。

综上,已经有研究者对评论进行文档级多方面情感分类,但这些研究多将其纳入多任务情感分类框架进行实验,在预测方面级别的情感时,没有关注评论的总体情感。2019年,Google 提出的 T5(Text-to-Text Transfer Transformer)^[18]模型将所有 NLP 任务都视为文本到文本的任务,用相同的方式处理不同的任务,证明了生成式模型在生成结构化输出处理

各个任务方面的可行性。本文尝试使用基于 T5 模型的生成式多粒度情感分析方法同时捕获产品评论文本中的方面级信息和整体情感,在酒类评论数据集上的测试证实了,该模型的效果相比单独分析评论的整体情感效果有一定的提升。

2 基于 T5 的多粒度评论情感分析

2.1 生成式多粒度评论情感分析

生成式多粒度评论情感分析任务旨在从评论文本序列 X 中提取多方面级别情感,并将其与文本的总体情感极性结合起来,生成一个结构化的输出。该任务涉及方面级情感分析和文档级情感分析两部分。方面级情感分析检测所有元组 (a, ap) ,其中 a 表示评论中提到的产品的一个方面, ap 表示与 a 相关联的情感极性,如正面、负面或者中性。文档级情感分析指的是对文档的总体情感极性 p 进行分类,从而全面反映评论的整体情感基调。

为了使多粒度情感分析适用于生成式建模任务,设计了一种结构化的输出格式,以完整包含每个方面级情感以及整体评论的情感极性。图 1 给出了结构化输出的构建过程。通过在原有的评分数据集上根据设定的评分标准进行转换,得到评论各方面和整体的情感极性;再将方面标签和情感极性的组合进行串联,得到最后的结构化输出。

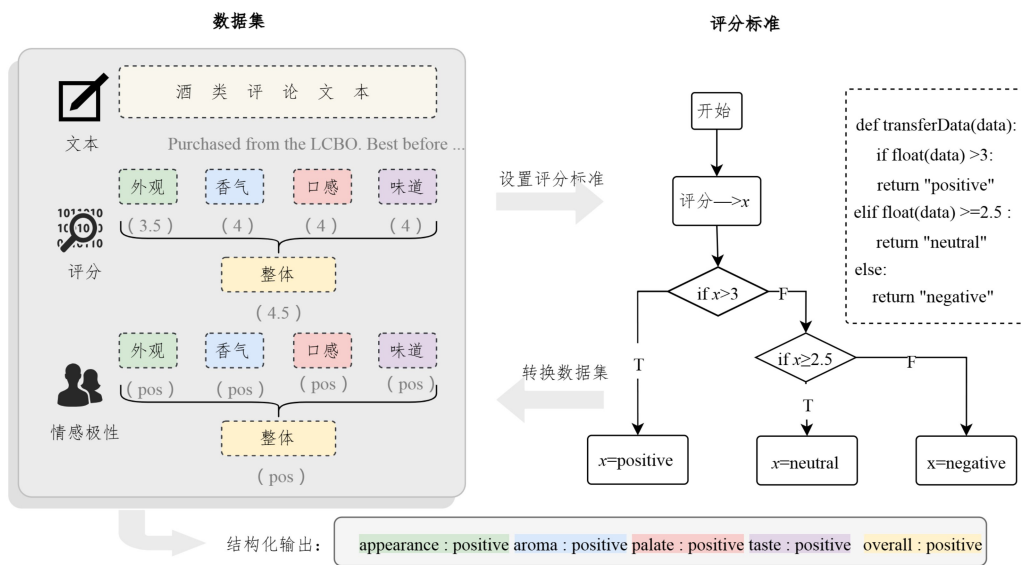


图 1 结构化输出的构建过程

Fig. 1 Process of building structured output

2.2 模型描述

本文模型是基于 T5 的多粒度评论情感分析模型 MEIA-T5。T5 模型是一个基于 Transformer^[19] 结构的新型神经网络模型,它在未标记文本和各种自然语言处理问题的标记数据上进行了训练,同时对各个任务进行了微调。作为一个编码器-解码器模型,它可以将所有 NLP 问题转换为文本到文本的形式,从而实现任务间的通用建模。它需要一个输入序列和一个输出序列,并结合注意力机制来进行训练。本文实验将评论文本作为输入序列 $X = \langle \langle s \rangle, x_1, x_2, \dots, x_{n-1}, x_n, \langle \backslash s \rangle \rangle$ 输入模型中,并使用 T5 模型的编码器和解码器训练生成期望格式的目标序列 $output =$

$\{o_1, o_2, o_3, \dots\}$ 。预测时,将评论文本输入训练好的模型中,生成构建好的结构化输出,这个结构化输出根据各个方面的评分来判定各个方面的情感倾向,并将其与方面标签相连。MEIA-T5 模型的整体结构如图 2 所示。

编码前对输入评论文本序列 $X = \langle \langle s \rangle, x_1, x_2, \dots, x_{n-1}, x_n, \langle \backslash s \rangle \rangle$ 进行向量化,得到向量化的评论文本,并将其输入到 T5 模型编码器中。T5 模型的编码器如图 3 的左半部分所示,输入向量中包含了多个方面的评论信息,如“非常适口”和“口感很好”具有一定的关联,使用多头自注意力机制同时关注到这些局部信息,包括商品评论中评论者对外观、香气、口感、味道的评论以及整体的评论。同时,在计算注意力时引入

相对位置编码,供所有层共享,使得模型对位置更加敏感。前馈神经网络对自注意力的输出进行进一步变换,从而增强模型的表达能力。残差连接把输入和网络的输出相连接,帮助梯度在模型中有效传播,并在每个组件的输入后都进行层归

一化,以稳定模型的训练过程。通过编码后得到隐向量 $\mathbf{H} = \{h_0, h_1, \dots, h_{n+1}\}$,其中包含了商品评论中的多方面信息。

$$\mathbf{H} = \text{T5Encoder}(\text{Embed}(X)) \quad (1)$$

其中, $\text{Embed}(X)$ 代表向量化后的序列。

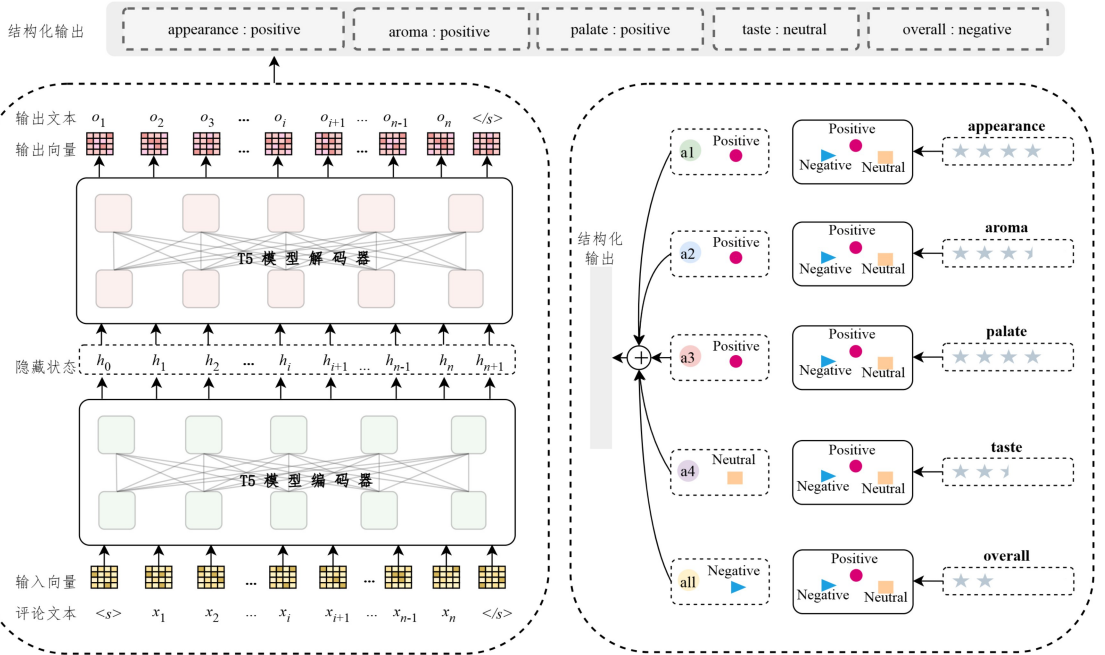


图2 MEIA-T5模型

Fig. 2 MEIA-T5 model

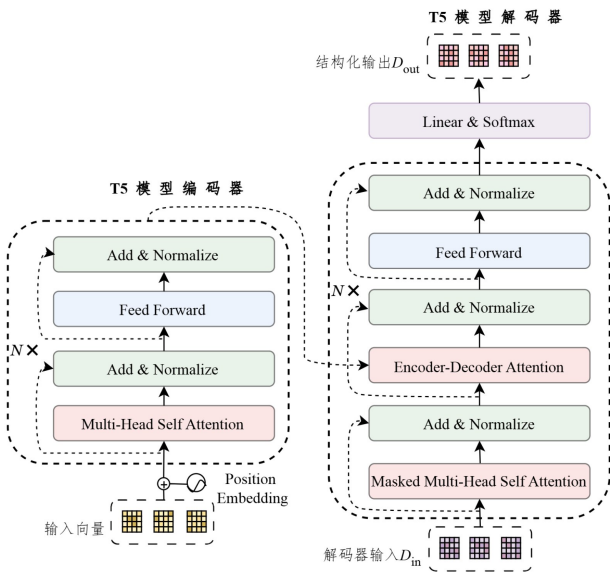


图3 T5模型编码器和解码器的结构

Fig. 3 Structure of T5 model encoder and decoder

解码器的结构与编码器类似,如图3的右半部分所示,但是在每个多头自注意力层后有一个编码器-解码器注意力层,该机制使每个解码器都能够访问编码器的输出 \mathbf{H} 。特定方面的情感极性除了依赖特定方面对应的评论之外,还与其他方面的情感极性具有一定的关联。以第一章中的例子进行说明,在生成整体评分对应的情感“整体:积极”时,编码器-解码器注意力层不仅能够关注到各个方面评分对应的情感“外观:积极 香气:积极 口感:积极 味道:积极”,还能够关注到隐向

量 \mathbf{H} 中代表“非常适口,一整天喝都没问题”的上下文信息。使用结构化表示 D_{in} 作为解码器的输入,通过掩码多头自注意力,编码器-解码器注意力和前馈神经网络训练解码器,生成解码器的输出 D_{out} 。

$$D_{out} = \text{T5Decoder}(D_{in}, \mathbf{H}) \quad (2)$$

具体来说,在生成 $D_{out} = \{y_1, y_2, y_3, \dots, y_n\}$ 的过程中,解码器利用隐藏向量 \mathbf{H} 逐步预测输出序列。在生成第 i 步时,解码器预测第 i 个词 y_i 。整个输出序列的条件概率由每个时间步的条件概率共同计算得到,计算过程如式(3)所示:

$$p(y|x) = \prod_{i=1}^{|y|} p(y_i | y_{<i}, x) \quad (3)$$

其中, $y_{<i} = y_1 \dots y_{i-1}$, $p(y_i | y_{<i}, x)$ 是在目标词汇上的概率。

对于结构化的解码器输出,根据每个方面对应的标签将方面情感提取出来,就能够评估本文提出的生成式情感分析模型的效果。

2.3 多头注意力机制

多头注意力机制是 Transformer 架构的核心组成部分,其基础是自注意力(Self-Attention)机制。MEIA-T5模型沿用了这一机制。与 Transformer 原始设计的在位置编码(Position Embedding)中直接将位置信息加到词嵌入向量上不同,T5采用相对位置编码,通过在每个注意力中共享位置编码,使得模型对位置更加敏感。多头注意力机制的结构如图4所示。

图4中, $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别代表查询矩阵、键矩阵和值矩阵,分别由对应的可训练的权重矩阵和上一层传入的输入向量相乘得到; R 代表相对位置信息。注意力层首先接收输入向量 \mathbf{X} ,

再通过线性变化生成 Q, K, V 。随后进行头分离,分离后分别进行注意力的计算并结合起来,通过在不同的表示子空间里并行处理信息,增强了模型的学习能力。在解码器的掩码多头自注意力机制中,每个时间步的位置嵌入通过可训练的权重矩阵分别生成 Q, K, V ,并结合掩码机制仅关注当前位置及其前序信息,从而捕捉各方面情感极性之间的上下文依赖。在编码器-解码器注意力中,使用解码器的前一状态的输入计算 Q 用于查询,使用编码器的输出 H 计算 K, V 作为键和值,从而通过编码器的整个输出来获取各个方面评论的上下文信息。多头自注意力的计算式如下:

$$\text{Atten}(Q, K, V) = \text{Softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}} + R\right) \times V \quad (4)$$

其中, d_k 代表每个注意力头的维度大小。

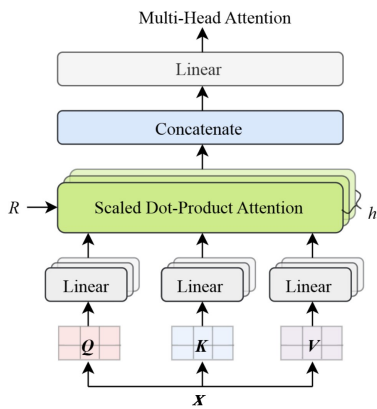


图4 多头注意力机制

Fig. 4 Multi-head attention

2.4 损失计算

为了得到最接近目标序列的结构化输出序列 X_O ,在给定评论文本 X_R 的基础上,最大化输出序列 X_O 的概率。因此使用交叉熵损失函数对预测分布与真实分布之间的差异进行建模,该损失函数如下所示:

$$L = -\frac{1}{|\tau|} \sum_{(X_R, X_O) \in \tau} \log p(X_O | X_R; \theta) \quad (5)$$

其中, θ 是模型参数, (X_R, X_O) 是训练集 τ 中的评论文本和结构化输出。

$$\log p(X_O | X_R; \theta) = \sum_{i=1}^n \log p(x_o^i | x_o^1, x_o^2, \dots, x_o^{i-1}, X_R; \theta) \quad (6)$$

其中, $\log p(x_o^i | x_o^1, x_o^2, \dots, x_o^{i-1}, X_R; \theta)$ 由 T5Decoder 计算。

3 实验

3.1 实验数据集

本文使用的数据集是来自 Beeradvocate 网站的酒类评分数据集,该数据集包含味道、外观等多个维度的评分。在实验中,先对数据进行清洗、归类、整理等一系列操作,从 1583422 组数据中随机抽取 10000 组不同酒的数据作为训练集,各抽取 2000 组与训练集不同酒的数据作为验证集和测试集来训练所提模型,数据集中包含了不同风格的酒品,种类丰富。其中每组数据包含对酒类的评论文本以及相关的评分数据,评分数据涵盖了酒品的 4 个评价维度,包括口感、香气、外观和

味道,综合反映了消费者对酒品的全面评价。

基于每组数据的评分,将每个评论对各个方面的情感极性分成 3 个类别:正面(positive)、中性(neutral)和负面(negative)。分类的依据是酒品的评分,对应评分标准如表 2 所列。

表2 酒类评分标准

Table 2 Wine classification standards

评分标准	情感极性
$X > 3$	positive
$2.5 \leq X \leq 3$	neutral
$X < 2.5$	negative

3.2 实验参数设置

使用 MEIA-T5 模型进行实验,其基于参数量为 220×10^6 的 T5-base 模型。所有的实验均在一张 Nvidia RTX 4090 显卡上进行,模型的参数如表 3 所列。模型的优化器选择 Adam,它基于随机梯度下降算法,能够根据梯度信息动态地调整学习率。max_len 受到数据集中评论文本长度的影响,设置为 256。batch_size 是每次训练抓取样本的数量,可以影响模型对文本特征的提取和梯度下降的方向,从而影响模型的训练效率。epochs 代表整个训练集在模型上的训练次数,如果训练后的验证损失在 3 次迭代中没有降低,则停止训练任务。

表3 模型参数

Table 3 Model parameter

参数	取值
optimizer	Adam
loss	CrossEntropy
learning_rate	0.0001
batch_size	16
max_len	256
epochs	20

3.3 评估准则

为了评估模型性能,主要采用准确率(Acc)和 F1 分数(F1)来衡量分类的效果。

1) 准确率:不论正负标签,预测正确的样本数占样本总数的比值。计算过程如式(7)所示:

$$\text{Acc} = \frac{\text{预测正确样本数}}{\text{样本总数}} \quad (7)$$

2) F1 分数:关注不同类别样本的分类效果,在计算时需要统计分类结果中的 TP(真正例)、TF(真负例)、FP(假正例)、FN(假负例),计算精确率 P 和召回率 R 。 P 和 R 的计算过程如式(8)、式(9)所示:

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$R = \frac{TP}{TP + FN} \quad (9)$$

据此计算出 F1 分数,如式(10)所示:

$$F1 = \frac{2PR}{P + R} \quad (10)$$

3.4 基准模型

随着人工智能领域的迅猛发展,计算机科学与技术不断取得进步,特别是在情感分析这一领域,已经诞生了众多卓越的算法模型。本研究将选取近年来表现较为出色的 7 种模

型,使用本文的数据集复现,进行一系列的比较和分析实验。

1)DPCNN:Johnson等^[20]提出的低复杂度的词级深度卷积神经网络架构,通过词级CNN的深化来有效地表示文本中的长距离关联,在情感分类上取得了不错的效果。

2)Bert:自Andrea等^[21]将Bert应用到Tweets的情感分析任务中并取得不错的效果后,研究者们纷纷将Bert分类模型用于各种情感分类任务。本实验采用Bert对酒类评论文本进行建模和评论情感分类,并将其作为本实验的基准模型。

3)Bert+LSTM:Bert模型在获取文本的深层语义特征上表现出了强大的能力,将其与LSTM结合,能够增强情感分类的准确性。

4)RoBERTa:Facebook于2019年提出的一种语言表示模型^[22],是对原始Bert的一种改进和优化。它使用了更大的数据集,训练了更长的时间,并去掉了Bert在预训练阶段的Next Sentence Prediction(NSP)任务以及动态调整掩码机制,在很多NLP任务中获得了非常不错的性能。

5)DeBERTa:微软于2020年提出的基于注意力解耦机制的解码增强型Bert^[23],同时采用新的虚拟对抗训练方法进行微调。

6)GPT3.5-Turbo:在原始的GPT3.5模型上进行了增强。其拥有庞大的参数数量,能够基于输入提示生成连贯且上下文相关的文本,且在使用一定的提示词后能够对评论的整体情感进行分析。

7)T5:由谷歌团队于2019年提出。该模型将所有的NLP任务转化成文本到文本的问题,并在情感分析任务上进行了微调。本文将T5-base模型进行微调,用于评论的文档级别的情感分析。

3.5 实验结果与分析

本文主要进行了两组实验:一组是将MEIA-T5模型与基准模型在文档级情感分类上进行对比;另一组是将方面级情感分析模型与MEIA-T5模型进行对比。

3.5.1 基准模型对比分析

为了验证生成式的评论情感分析方法的有效性以及MEIA-T5模型的效果,选择了在自然语言处理领域表现优异的BERT模型以及近年来在评论情感分析任务中表现优异的模型进行对比。实验结果如表4所列。

表4 总体评价对比

模型	Acc	F1
DPCNN	0.755	0.344
Bert	0.798	0.610
Bert+LSTM	0.816	0.626
RoBERTa	0.807	0.551
DeBERTa	0.802	0.568
GPT3.5-Turbo	0.732	0.622
T5	0.827	0.648
MEIA-T5(Ours)	0.835	0.670

从表4中可以得出以下结论:

1)在所有基线模型中,DPCNN的表现最差,F1值远远低于其他模型。这是因为DPCNN的模型结构较简单,获取长文本关联的效果不佳,且其对于数据中相对较少的中立和消

极情感难以起到很好的预测效果。当应用更复杂的模型时,DPCNN情感分类的效果会更好。预训练的Bert和RoBERTa模型都取得了不错的效果,证明了预训练模型的有效性。RoBERTa模型作为Bert模型的变体,在准确率上提升了0.9%,但是在F1值上却有着较大的差距,原因在于RoBERTa在预测时对消极情绪的预测能力较差。DeBERTa模型的准确率和RoBERTa相近,F1值提升了1.7%。此外,我们发现了模型的复杂性对于情感分析任务的作用,在Bert模型中加入LSTM模型后,模型的准确率提升了1.8%,F1值提升了1.6%。在使用GPT3.5-Turbo进行实验时,未进行训练微调,GPT3.5-Turbo的准确率比训练的模型低,但是F1值取得了与Bert+LSTM模型相当的水平,原因在于其不受实验数据的影响,且作为生成式模型,其对数据中相对较少的中立和消极的情感的预测效果较好。相比于最好的基线模型Bert+LSTM,使用T5模型生成评论总体评价的准确率提升了1.1%,F1值提升了2.2%,证明了生成式的T5模型在本文任务上的有效性。

2)通过T5单独生成评论整体情感极性和MEIA-T5的实验对比可以看出,MEIA-T5较使用T5模型单独生成整体评价,准确率提升了0.8%,F1值提升了0.9%;多粒度情感联合生成在生成时利用了评论的多方面评价,证明了MEIA-T5在生成时考虑方面级别评价后,整体评价的生成效果具备一定的优势。

3.5.2 模型方面信息生成效果

为了验证模型的多方面情感生成效果,将MEIA-T5模型与如下一些文档级多方面评论情感分析方法进行对比。

1)MHAN基于Yang等^[24]提出的一种具有层次结构的网络HAN(HAN利用注意力机制捕获单词和句子信息并应用于不同层次),将HAN与多任务学习框架相结合。

2)MGRUNAsp基于Li等^[16]提出的层次用户方面评级网络,在评估多方面情感时,忽略用户信息,以注意力的形式结合方面关键词信息对方面情感进行预测。

3)GAS基于Zhang等^[25]提出的一种生成式模型,为所有ABSA任务设计了统一的生成框架,并将其作为情感元素序列生成问题。

4)ME-T5情感分析模型是在MEIA-T5模型的基础上不考虑总体评价,只对多方面评价进行生成的模型。

实验结果如表5所列。

表5 方面评价对比

模型	Acc	F1
MHAN	0.785	0.507
MGRUNAsp	0.788	0.516
GAS	0.808	0.635
ME-T5	0.812	0.643
MEIA-T5(Ours)	0.815	0.641

从表5中可以得出以下结论:结合了多任务框架的MGRUNAsp模型由于结合了方面关键词,准确率比MHAN模型提升了0.3%,F1值提升了0.9%,但是还远不及本文提出的MEIA-T5模型。去除了总体评价预测的ME-T5模型在多方面评论情感分析中的生成效果与MEIA-T5模型相

近,在准确率降低 0.3% 的同时,F1 值提升了 0.2%。相较于同为生成式模型的 GAS,MEIA-T5 的准确率提升了 0.7%,F1 值提升了 0.6%。实验结果明了本文设计的结构化输出的有效性。

为了验证 MEIA-T5 较单独生成模型 UAE-T5(Unilateral Aspect Evaluation T5)的优势,考虑 MEIA-T5 在各个方面信息的生成效果,在模型相应参数保持不变的情况下,使用 UAE-T5 对酒类评价数据集中的各方面信息进行单方面情感极性的生成,并与本文提出的 MEIA-T5 进行对比实验。实验结果如表 6 所列。

表 6 单一方面评价对比

Table 6 Single aspect evaluation comparison

	UAE-T5	MEIA-T5(Ours)
Appearance	0.509	0.554
Aroma	0.635	0.653
Palate	0.603	0.641
Taste	0.683	0.680

表 6 中的实验结果展现了两个基于 T5 的生成模型 UAE-T5 和 MEIA-T5 对酒类数据集各个方面信息的分类的差异。MEIA-T5 在外观、香气、口感方面的 F1 值分别比 UAE-T5 高 4.5%,1.8%,3.8%,有明显提升;在味道方面的 F1 值与 UAE-T5 相近,仅低 0.3%。这表明,MEIA-T5 在处理多个方面的情感分析时,能利用完整的上下文信息,包括文本中各个方面之间的相互关系,从而在多个方面的生成上获得更好的效果。

3.5.3 各方面信息对总体评估的影响

MEIA-T5 模型同时考虑了外观、香气、口感和味道信息,表 7 列出了缺失某一方面的信息对多粒度情感分析总体情感生成效果的影响。

表 7 单一方面评价缺失的对比

Table 7 Comparison of missing evaluations in a single aspect

模型	Acc	F1
w/o appearance	0.825	0.686
w/o aroma	0.828	0.637
w/o palate	0.823	0.651
w/o taste	0.825	0.648
Ours	0.835	0.670

表 7 的实验结果表明,在去除香气、口感和味道方面的评价信息后,整体评价的 F1 值分别下降了 3.3%,1.9%,2.2%,证明了这些方面对于提升整体情感生成效果的有效性。但是,当去除外观方面的评价信息后,总体评价反而出现了 1.6% 的提升,证明了外观信息对于生成的总体评价的提升效果不佳。MEIA-T5 模型预测得到的结果也是如此,预测结果中外观的情感极性对总体情感的影响较大,但在人们对酒类进行评价的过程中,香气、口感和味道对于一款酒品的整体评价往往更关键。

结束语 本研究首先对比分析了基于 T5 的生成式模型和分类模型在总体情感和方面情感上的性能,然后探讨了 MEIA-T5 模型相比于使用 T5 模型单独生成单方面情感信息的效果优势,最后探究了各个方面的评价对于总体评价

生成的有效性。实验结果显示,MEIA-T5 在情感分类任务上的表现超越了所有分类式基线模型,并且在加入多方面评价后优于 T5 模型,体现了生成式的 T5 模型处理情感分析任务的强大能力以及加入方面评价对于总体评价生成的有效性。在方面评价生成上,MEIA-T5 优于传统的文档级多方面情感分析方法,但是去除总体评价后,方面级别的生成效果没有明显变化。

此外,使用 MEIA-T5 模型对各方面的信息联合生成的效果优于 UAE-T5 生成的单方面情感。这一结果表明,综合考虑评论的多方面信息不仅能够提高模型对评论整体情感倾向的预测准确性,还能够增强模型对评论各方面细节的理解能力。最后,本文通过去除单方面验证各方面对于总体评价的有效性,证明了外观对于生成的总体评价的提升效果最差。

通过多粒度情感联合生成方法,模型能够捕捉到文本中多方面的情感信号,并结合上下文信息更准确地反映文本的情感态度。经过反复实验并一一对比,所提 MEIA-T5 模型取得了显著的效果。

相较于现有的相关工作,本文利用生成的方式结合特定的结构化输出,更加全面地理解和反映了文本的情感结构。实验结果表明了 MEIA-T5 模型在总体情感和方面情感分类上的优越性,其为商品的多粒度评论情感分析提供了新的思路。同时,该方法不仅仅局限于商品评论,对酒店和旅游评论中各方面情感的识别也有一定的作用,具有一定的应用价值。然而,该模型也存在一定的局限性。在实际的应用中,评论中会存在图片等其他模态的数据,本文只针对酒类商品的评论文本进行多粒度的评论情感分析,忽视了不同模态和不同领域的的数据。未来,将进一步探索如何利用生成式模型在多领域和多模态数据集上进行多粒度的情感分析。

参考文献

- [1] WANG H W,ZHENG L J,YIN P,et al. A Review of Sentiment Polarity Classification of Online Reviews [J]. Journal of Information Science,2012,30(8):1263-1271,1276.
- [2] LI O. Sentiment Analysis and Application of Online Product Reviews [D]. Dalian:Dalian Maritime University,2017.
- [3] HUANG T. Research on Sentiment Classification of Tourist Destinations Based on Convolutional Neural Network[C]//2021 IEEE 3rd Eurasia Conference on IOT,Communication and Engineering(ECICE). IEEE,2021:358-361.
- [4] BHATIA P,JI Y,EISENSTEIN J. Better Document-level Sentiment Analysis from RST Discourse Parsing[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015:2212-2218.
- [5] TANG D,QIN B,LIU T. Document modeling with gated recurrent neural network for sentiment classification [C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015:1422-1432.
- [6] BAZIOTIS C,PELEKIS N,DOULKERIDIS C. DataStories at SemEval-2017 Task 4:Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis[C]//Proceedings of

- the 11th International Workshop on Semantic Evaluation (SemEval-2017). 2017:747-754.
- [7] REHMAN A U, MALIK A K, RAZA B, et al. A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis[J]. *Multimedia Tools and Applications*, 2019, 18: 26597-26613.
- [8] SASIKALA P L, Mary Immaculate Sheela. Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS [J]. *Journal of Big Data*, 2020, 7(1):33.
- [9] CAO Y, SUN Z X. A study of sentiment analysis algorithms for agricultural product reviews based on improved BERT model [J]. *Symmetry*, 2022, 14(8):1604.
- [10] YIN D, MENG T, AND CHANG K W. SentiBERT: A Transferable Transformer-Based Architecture for Compositional Sentiment Semantics[C]//*Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020: 3695-3706.
- [11] ZHAO H, HUANG L T, ZHANG R, et al. Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction[C]//*Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020: 3239-3248.
- [12] CAI H J, TU Y F, ZHOU X S, et al. Aspect-category based sentiment analysis with hierarchical graph convolutional network [C]//*Proceedings of the 28th International Conference on Computational Linguistics*. 2020:833-843.
- [13] LIU J, TENG Z Y, CUI L Y, et al. Solving Aspect Category Sentiment Analysis as a Text Generation Task[C]//*Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 2021:4406-4416.
- [14] CAI H J, XIA R, YU J F, et al. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions [C]//*Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. 2021: 340-350.
- [15] LU B, OTT M, CARDIE C, et al. Multi-aspect sentiment analysis with topic models[C]//*Proceedings of the 11th IEEE International Conference on Data Mining Workshops*. 2011:81-88.
- [16] LI J J, YANG H T, ZONG C Q. Document-level Multi-aspect Sentiment Classification by Jointly Modeling Users, Aspects, and Overall Ratings[C]//*Proceedings of the 27th International Conference on Computational Linguistics*. 2018:925-936.
- [17] ZHANG Q, SHI C. An Attentive Memory Network Integrated with Aspect Dependency for Document-Level Multi-Aspect Sentiment Classification[C]//*Asian Conference on Machine Learning*. PMLR. 2019:425-440.
- [18] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer [J]. *Journal of Machine Learning Research*, 2020, 21(140):1-67.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017: 6000-6010.
- [20] JOHNSON R, ZHANG T. Deep Pyramid Convolutional Neural Networks for Text Categorization[C]//*Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. 2017:562-570.
- [21] CHIORRINI A, DIAMANTINI C, MIRCOLI A, et al. Emotion and sentiment analysis of tweets using BERT [C]//*EDBT/ICDT Workshops*. 2021.
- [22] LIU Y, OTT M, GOYAL N, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach [J]. *arXiv*: 1907. 11692, 2019.
- [23] HE P, LIU X, GAO J, et al. DeBERTa: Decoding-enhanced BERT with Disentangled Attention [J]. *arXiv*: 2006. 03654, 2020.
- [24] YANG Z, YANG D, DYER C, et al. Hierarchical Attention Networks for Document Classification[C]//*Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2016:1480-1489.
- [25] ZHANG W, LI X, DENG Y, et al. Towards Generative Aspect-based Sentiment Analysis[C]//*Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. 2021:504-510.



ZHANG Jiawei, born in 2001, postgraduate. His main research interests include natural language processing and sentiment analysis.



WANG Zhongqing, born in 1987, Ph.D., associate professor. His main research interests include natural language processing and sentiment analysis.

(责任编辑:何杨)