计算机科学
COMPUTER SCIENCE

基于置信度引导提示学习的多模态方面级情感分析

李懋林, 林嘉杰, 杨振国

引用本文

李懋林, 林嘉杰, 杨振国. 基于置信度引导提示学习的多模态方面级情感分析[J]. 计算机科学, 2025, 52(7): 241-247.

LI Maolin, LIN Jiajie, YANG Zhenguo. Confidence-guided Prompt Learning for Multimodal Aspect-level Sentiment Analysis [J]. Computer Science, 2025, 52(7): 241-247.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240600126

基于置信度引导提示学习的多模态方面级情感分析

李懋林　林嘉杰　杨振国
广东工业大学计算机科学与技术学院　广州 510006
(gdutlml@gmail.com)

摘　要　面对日益增加的社交平台数据,多模态方面级情感分析对于理解用户的潜在情感至关重要。现有研究工作集中于通过跨模态融合图像和文本来完成情感分析任务,无法有效地捕获图像和文本中的隐含情感。此外,传统方法受限于模型具有的黑箱性质而缺乏可解释性。为应对上述问题,提出了基于置信度引导的提示学习(CPL)的多模态方面级情感分类模型。该模型由多模态特征处理模块(MF)、基于置信度的门控模块(CG)、提示构建模块(PC)和多模态分类模块(MC)组成。多模态特征提取模块用以提取多模态数据的特征;基于置信度的门控模块旨在通过自注意力网络的置信度评估样本的分类难度,对不同难易程度的样本进行自适应性处理;提示构建模块根据难易样本,采取不同的适应性模板提示,以引导 T5 大语言模型生成辅助情感线索;多模态分类模块用以预测结果。在公开数据集 Twitter-2015 和 Twitter-2017 的实验结果表明,与现有基线方法相比,所提出的多模态方面级情感分类模型具有显著性能优势,准确率分别提高了 0.48%和 1.06%。

关键词:多模态数据;大语言模型;情感分类;提示学习;分类置信度
中图分类号　TP311

Confidence-guided Prompt Learning for Multimodal Aspect-level Sentiment Analysis

LI Maolin,LIN Jiajie and YANG Zhenguo
School of Computer Science and Technology,Guangdong University of Technology,Guangzhou 510006,China

Abstract　With the increasing volume of data from social media platforms,multimodal aspect-level sentiment analysis is crucial for understanding the underlying emotions of users. Existing research primarily focuses on sentiment analysis tasks by fusing image and text modalities,but these methods fail to effectively capture the implicit emotions in both image and text. Furthermore,traditional approaches are often constrained by the black-box nature of the models,which lack interpretability. To address these issues,this paper proposes a confidence-guided prompt learning(CPL) based multimodal aspect-level sentiment analysis model,which consists of four key components:a multimodal feature processing module(MF),a confidence-based gating module(CG),a prompt construction module(PC),and a multimodal classification module(MC). The multimodal feature processing module is responsible for extracting features from multimodal data. The confidence-guided gating module evaluates the classification difficulty of samples using confidence assessment through a self-attention network and adaptively processes samples based on their difficulty. The prompt construction module generates adaptive prompt templates for different difficulty levels of samples to guide the T5 large language model in generating auxiliary sentiment cues. And the multimodal classification module is used for final sentiment prediction. Experimental results on the public datasets Twitter-2015 and Twitter-2017 show that,compared to existing baseline methods,the proposed multimodal aspect-level sentiment classification model achieves significant performance improvements,with accuracy increases of 0.48% and 1.06%,respectively.

Keywords　Multimodal data,Large language models,Sentiment classification,Prompt learning,Classification confidence

1　引言

随着互联网的高速发展,人们已不再满足于使用单一的文本,而更倾向于通过图像、文本等多种载体表达自我情感,同时社交媒体平台成为了人们发表自我观点和言论的主要阵地。由于图像能反映社交媒体用户的情感,联合图像和文本的多模态数据包含了更加丰富的情感信息,因此,对多模态数据的情感分类有助于进一步了解人们对某话题或热点的立场和态度,在监控舆情、民意调查、产品分析、推荐系统等方面存在巨大的应用价值。

在多模态情感分析的发展中,主流的方法是对提取的文本和图像特征进行融合后执行多分类任务,如基于神经网络的融合方法[1]、基于注意力的融合方法[2]、基于强化学习的方法[3]、基于对比学习的方法[4-5]等。而基于方面的多模态情感

到稿日期:2024-06-21　返修日期:2024-09-26
基金项目:广东省基础与应用基础研究基金(2024A1515010237)
This work was supported by the Guangdong Basic and Applied Basic Research Foundation(2024A1515010237).
通信作者:杨振国(yzg@gdut.edu.cn)

分类作为多模态情感分析的重要且具有前景的任务，受到越来越多的研究者的关注。如 Zhou 等[6]在语义上对齐图像和文本，并通过图卷积神经网络聚合情感信息；Yu 等[7]将每种模态的特定知识与一般知识表示结合，以促进不同模态之间的交互。

这类方法通常由模型大量的网络参数和复杂的网络结构组成，因此在训练过程中通过自动学习来建模复杂的非线性关系；同时，因为深度学习中网络的黑盒特性，其结果往往只能学习输入与输出之间的关联，而无法提供具体的解释或推理过程。此外，由于一些文本和图像数据中包含难以挖掘的隐式情感，多模态情感分析对于理解不同样本中的隐式情感的复杂程度不同，导致模型很难学习其中真实的关联性。

为解决该问题，本文提出了基于置信度引导大模型提示学习的多模态方面级情感分析模型（Confidence-Guided Prompt Learning for Multimodal Aspect-level Sentiment，CPL），通过引入大语言模型强大的推理能力来增强模型决策的可解释性，实现更加合理和准确的预测。具体来说，首先将图片数据转换为文本描述，然后采用基于置信度的门控模块（CG）确定样本分类的难易程度；根据样本分类难易的不同，在提示构建模块（PC）中设计了两种不同的提示，通过 T5 大语言模型获得推理的原因后将其再次送入 CG 模块中，以获得更加精准的预测。在两个基准数据集上的实验结果表明，本文方法有效提升了在多模态方面级情感分类中的有效性和可解释性。

本文的主要贡献总结如下：

1）提出基于置信度的门控模块（CG）来动态控制难易程度不同的样本的决策过程，提高了模型推理的自适应性；

2）提出提示构建模块（PC），根据样本分类的难易程度在提示模板中调整预测标签的选择范围，进而引导模型进行高质量推理；

3）在 Twitter-2015 和 Twitter-2017 两个通用数据集上进行实验，实验结果表明，与多个基线方法相比，CPL 在多模态方面级情感分类任务上具有优越的性能表现。

2 相关工作

2.1 多模态方面级情感分析

目前，多模态情感分析是一个热门的研究方向，而方面级情感分析则是细粒度情感分析任务，近年来受到越来越多的关注。Tang 等[8]将句子根据目标情感极性分为两个部分，并利用两个长短期记忆网络（Long Short-Term Memory）获得目标的左、右表示。随着社交媒体中多模态数据的不断丰富，文献[9-11]发现图像在方面术语提取中能提供大量的信息。Ling 等[12]基于 BART 构建了一个生成式多模态架构，用于视觉-语言预训练和下游的方面级情感分析任务。Yang 等[13]通过动态地控制视觉信息对不同方面的贡献，更改文本和图像影响结果的权重。Zhao 等[14]提出了一个知识增强框架来改进视觉注意力的能力，该框架利用外部知识来增强视觉处理中的理解和注意机制，通过利用预先存在的知识，如语义关系或上下文信息，模型能够更好地识别和关注相关的视觉区域。Zhao 等[15]提出了一个多粒度、多课程去噪框架，通过调整训练数据顺序实现去噪。Zhou 等[16]提出面向方面的方法检测与方面相关的语义和情感信息，设计了一个方面感知注意模块来同时选择与方面语义相关的文本标记和图像块。然而，上述方法侧重于对提取特征的进一步操作，由于模型固有的黑箱性质，其结果缺少可解释性，因此本文利用大语言模型的强大推理能力具现化推理过程，通过预设的提示模板获取模型推理的原因，从而增加模型预测的可解释性。

2.2 使用提示学习的情感分析

在自然语言处理领域的下游任务中，使用特定于任务的头部对预训练语言模型进行微调已成为主流范例，一些基于提示学习的方法相继被提出。Fei 等[17]设计了基于大规模语言模型的三跳思维链框架，以推理出隐式情感。Zhu 等[18]使用多提示学习解决不同训练数据不平衡的问题。Yang 等[19]设计了一种统一的多模态提示模板，以减少不同模态之间的差异，并将多模态示例动态地融入到多模态示例的上下文中。Liu 等[20]提出一种统一的目标导向多模态情感分类模型，利用描述性的提示重新构建目标导向的多模态情感分类。Bao 等[21]提出了一个多任务多提示模型，为不同任务引入不同模板。Shi 等[22]提出了一种基于软提示的跨域方面词联合学习的方法，结合外部语言特征减少方面词分布的差异。Li 等[23]利用情感知识增强提示，在统一框架内微调语言模型。虽然思维链方法的动态更改提示模板能够让语言模型获取更多的推理依据，但是更改提示中包含了过多的冗余信息，而固定的提示模板又缺乏灵活性。针对这个问题，本文设计了一个门控模块微调提示模板，其根据门控的选择对模板内容进行更改，在减少冗余信息的同时更具灵活性。

3 基于置信度引导的提示学习模型

本文提出基于置信度引导的提示学习模型，用于解决多模态方面级情感分类问题。如图 1 所示，模型主要由多模态特征处理模块、置信度门控模块、提示构建模块、多模态分类模块 4 部分组成。

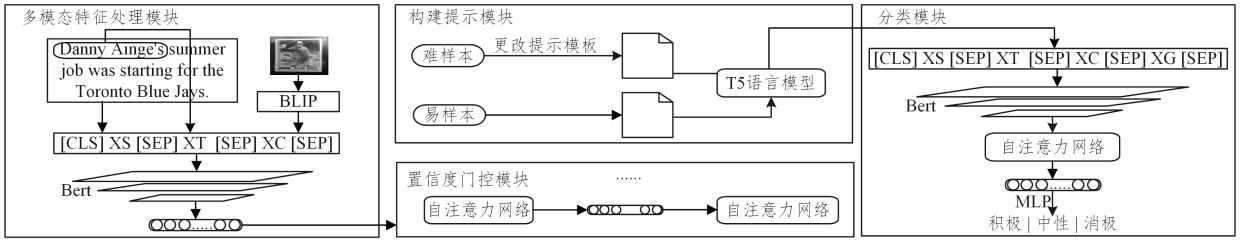


图 1 基于置信度引导提示学习的多模态方面级情感分析模型框架

Fig. 1 Framework of confidence-guided prompt learning for multimodal aspect-level sentiment analysis

3.1 多模态特征处理模块

对于视觉模态信息来说，多模态常使用深度残差神经网络(ResNet)和 VGG 等视觉预训练模型深入挖掘全局特征，但是由于 T5 大语言模型的输入不能接受文本之外的数据，因此本文使用预训练后的 BLIP 模型[24]将图片 V 转换为图像描述信息 X^C。

$$X^C = F_{\text{BLIP}}(V) \tag{1}$$

其中，$F_{\text{BLIP}}(\cdot)$代表预训练的多模态大语言模型。经过大规模语料库预训练的语言模型具有强大的文本特征抽取能力。给定句子信息X^S、方面术语X^T以及图像描述X^C，将其连接成一个统一的文本 X。本文使用预训练的 BERT 模型[25]进行文本特征提取。

$$X = [\text{CLS}]X^S[\text{SEP}]X^T[\text{SEP}]X^C[\text{SEP}] \tag{2}$$

$$\boldsymbol{U}_1 = F_{\text{Bert}}(X) \tag{3}$$

其中，[CLS]代表分类头标记，[SEP]代表句子分割标记，$\boldsymbol{U}_1 \in \mathbb{R}^{n\times d}$，$n$ 代表文本 X 的长度，d 代表词向量维度，$F_{\text{Bert}}(\cdot)$表示预训练的语言模型。

3.2 置信度门控模块

在多模态情感分析中，由于不同样本包含的潜在情感的复杂程度不同，因此模型较难学习其中的关联。如图 2 所示，本文通过引入 Vaswani 等[26]提出的自注意力机制来充分学习词与词之间的关联，进而挖掘其中的隐式情感。

$$\boldsymbol{U}_A = \boldsymbol{W}_1^{\mathrm{T}}\left(\text{softmax}\left(\frac{\boldsymbol{U}_1\boldsymbol{W}^Q(\boldsymbol{U}_1\boldsymbol{W}^K)^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{U}_1\boldsymbol{W}^V\right)+\boldsymbol{b}_1 \tag{4}$$

其中，$\boldsymbol{W}^Q$，$\boldsymbol{W}^K$，$\boldsymbol{W}^V$为可训练的参数矩阵，d_k表示 $\boldsymbol{Q}$，$\boldsymbol{K}$ 的向量维度，$\boldsymbol{W}_1$和 $\boldsymbol{b}_1$表示隐藏层可训练参数，softmax(·)表示激活函数。

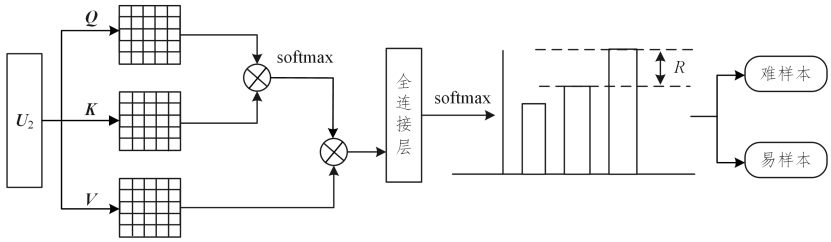


图 2 置信度门控模块结构

Fig. 2 Structure of confidence-based gating module

在自注意力模块充分学习到词与词之间的关系特征后，通过多层感知器获得不同结果的置信度表示。

$$U_2 = \text{softmax}(\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{U}_A + \boldsymbol{b}_2) \tag{5}$$

其中，$\boldsymbol{W}_2$和b_2代表全连接层可训练参数，softmax(·)表示激活函数，U_2代表每个样本类别的置信度。

在得到每个样本的置信度后，根据对不同样本进行分类的难易程度，将样本分为困难样本和简单样本。困难样本指的是U_2中的值非常接近，很难区分具体的类别，这意味着普通模型很难去准确分类此样本。对于不同分类难度的样本，本文利用大语言模型的强大推理能力进行进一步推理，通过门控模块对样本分类的难易程度进行区分，并对后续输送给大语言模型的提示模板进行动态调整。

首先，获取每个样本的置信度，得到置信度中的不同值$U_2 \in (0,1)$。通过定义门控的区分规则，对不同的样本难度进行区分。

$$k^i = \begin{cases} 1, & Z_g \in (Z^i_{\max} - Z^i_{\text{mid}} \leqslant \alpha) \\ 0, & Z_g \in \text{other} \end{cases} \tag{6}$$

其中，$Z^i_{\max}$，Z^i_{mid}分别表示U_2中置信水平的最大值和中间值，α表示置信度阈值。不同的k^i值归属于不同的难易程度，1 表示难样本，0 表示易样本。

3.3 提示构建模块

现有的利用大语言模型进行目标情感分类任务的研究通常通过思维链的方式辅助大语言模型推理，或者利用参数量更大的大语言模型进行更深层次的解读，这类方法通常使用固定的提示模板或者将复杂的问题分解为多个小问题进行推理，但是缺乏对推理过程中动态变化的适应。基于此，如图 3 所示，本文构建了一个提示模块，根据门控模块不同的输出值对提示模板进行动态调整，通过获取 T5 模型生成的辅助文本，增加对隐式情感的发掘，提升模型的可解释性。

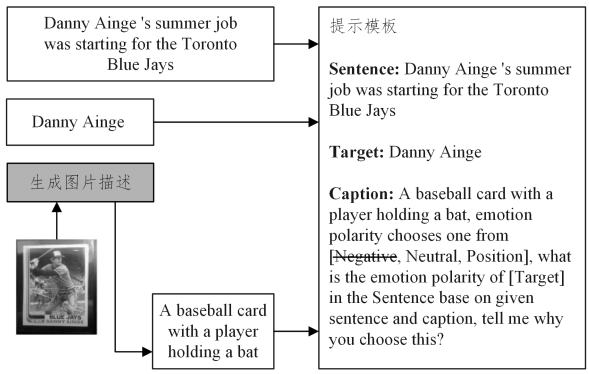


图 3 提示模板结构

Fig. 3 Structure of prompt construction module

从置信度门控模块获得输出值k^i：1)对于$k^i=0$的样本，使用本文预先设定好的提示模板；2)对于$k^i=1$的样本，通过获得U_2中的最小值对应的索引来屏蔽相应的分类标签。

$$L_i = S(U_2) \tag{7}$$

其中，$S(\cdot)$表示排序函数，L_i表示排序后的每个样本不同类别的置信度。对于$k^i=0$的样本，不对模板进行调整；对于$k^i=1$的样本，根据L_i得到的排序将最小值对应的类别从提示模板中掩盖掉。在获得提示模板 P 后，为了获取辅助推理文本 X^G，将其输送进大语言模型 T5 中。

$$X^G = F_{\text{T5}}(P) \tag{8}$$

其中，X^G表示 T5 大语言模型输出的文本，$F_{\text{T5}}(\cdot)$表示微调后的 T5 模型。

3.4 分类模块

为了提高目标分类的准确率，将辅助推理文本X^G与特征

提取预处理好的文本 X 连接，并输送到 BERT 模型获得全局特征。

$$\bar{X}=X[\mathrm{SEP}]X^{G}[\mathrm{SEP}] \tag{9}$$

$$\bar{\boldsymbol{U}}=F_{\mathrm{Bert}}(\bar{X}) \tag{10}$$

其中，$\bar{X}$表示连接后的样本；$\bar{\boldsymbol{U}}\in\mathbb{R}^{b\times d}$代表提取的特征，$b$ 代表文本$\bar{X}$的长度，d 代表向量维度；$F_{\mathrm{Bert}}(\cdot)$表示预训练的语言模型。特征 $\bar{\boldsymbol{U}}$ 中包含文本、图片的描述、目标词以及大模型给出的辅助文本的多重信息。之后，本文通过自注意力网络学习文本中更深层次的关联，以发掘隐式情感。

$$\bar{\boldsymbol{U}}_A=\boldsymbol{W}_U^{\mathrm{T}}\left(\mathrm{softmax}\left(\frac{\bar{\boldsymbol{U}}\boldsymbol{W}^{Q}(\bar{\boldsymbol{U}}\boldsymbol{W}^{K})^{\mathrm{T}}}{\sqrt{d_k}}\right)\bar{\boldsymbol{U}}\boldsymbol{W}^{V}\right)+\boldsymbol{b}_U \tag{11}$$

其中，$\boldsymbol{W}_U^{\mathrm{T}}$和$\boldsymbol{b}_U$代表可训练参数，softmax(·)表示激活函数。

本文通过多层感知器随机对结果进行预测。

$$\hat{\boldsymbol{U}}=\mathrm{softmax}(\boldsymbol{W}_q^{\mathrm{T}}\bar{\boldsymbol{U}}_A+\boldsymbol{b}_q) \tag{12}$$

其中，$\boldsymbol{W}_q^{\mathrm{T}}$和$\boldsymbol{b}_q$代表全连接层可训练参数，softmax(·)表示激活函数，$\hat{\boldsymbol{U}}$ 代表每个批次中对于每个样本不同类别的概率分布。

本文使用交叉熵损失作为损失函数。

$$\mathcal{L}=-\frac{1}{n}\sum_{i}^{n}\boldsymbol{U}_i\log(\hat{\boldsymbol{U}}_i)+(1-\boldsymbol{U}_i)\log(1-\hat{\boldsymbol{U}}_i) \tag{13}$$

其中，$\boldsymbol{U}_i$代表 1 个批次中第 i 个样本的真实标签，$\hat{\boldsymbol{U}}_i$代表 1 个批次中第 i 个样本的标签预测概率，n 代表每个训练批次的大小。

4 实验与分析

4.1 数据集

为验证所提模型的有效性，使用数据集 Twitter-2015 和 Twitter-2017 进行实验。Twitter-2015 和 Twitter-2017 是多模态数据集，其中包含文本内容、图片、方面信息及情感类别的信息。表 1 和表 2 列出了两个数据集的具体统计信息。

表 1 Twitter-2015 数据集统计信息

Table 1 Statistics on the Twitter-2015 dataset

数据集	情感分类			句子数量
	积极	中性	消极	
训练集	928	368	1883	3179
验证集	303	149	670	1122
测试集	317	113	607	1037

表 2 Twitter-2017 数据集统计信息

Table 2 Statistics on the Twitter-2017 dataset

数据集	情感分类			句子数量
	积极	中性	消极	
训练集	1508	1638	416	3562
验证集	515	517	144	1176
测试集	493	573	168	1234

4.2 实验设置

在训练本文模型的过程中，优化器使用 Adam，学习率设置为 0.0001，批次大小设置为 16，训练次数(epoch)设置为 50，自注意力为 4 层。在 Twitter-2015 中和 Twitter-2017 中，置信度阈值分别设置为 0.013 和 0.024。本文采用 PyTorch 框架，T5 模型采用 T5-base。

4.3 基线模型

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。

1)Res-RAM 和 Res-MGAN 采用 Hazarika 等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，之后采用 softmax 层将特征分类。

2)Res-RAM-TFN 和 Res-MGAN-TFN 采用 Zadeh 等[27]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。

3)MIMN[28]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。

4)EASFN[29]提出了一种面向实体的视觉机制来提取与目标实体相关的视觉块，并使用门控机制来消除视觉背景中的噪声。

5)UMAS[30]构建了一个多模态方面术语提取和方面级情感分类的统一框架，并引入词性信息提升性能。

6)SaliencyBERT[31]提出了一种循环注意力网络，以逐步优化目标敏感文本特征和视觉特征的对齐，增强模态之间的交互能力。

7)VilBERT[32]基于 BERT 架构扩展为一个多模态的双流模型，通过共同注意力分别处理视觉和文本的输入，并将不同模态进行信息传递。

8)BERT-Pair-QA[33]将方面级情感分析转换为句子对分类任务，并对预训练的 BERT 模型进行微调。

9)TomBert[25]利用 BERT 建立了基于方面术语和图像的跨通道交互表征，并利用自注意力机制对文本和目标词-图像的通道内交互进行建模。

10)FITE[34]引入图像中的面部表情作为情感线索，用于与方面术语对齐，增强融合模态的特征表示。

11)EF-CapTrBERT[35]利用 Transformer 生成图像的字幕作为辅助文本，丰富双流模型的文本信息。

4.4 实验结果与分析

4.4.1 对比实验

本文模型在多模态领域与现有方法在 Twitter-2015 上的性能对比如表 3 所列。一方面，CPL 在 Twitter-2015 数据集上的准确度超过了所有的基线模型。原因是其根据 CG 模块分类的不同样本设计了不同的提示模板，以更好地引导大语言模型生成高质量的辅助文本，从而提升了模型的推理能力。另一方面，CPL 的 F_1 分数比 UMAC 低 0.15%。这可能是 Twitter-2015 数据集中有很多无法识别的文本形成了干扰模型判断的噪声，导致模型判断错误。

本文模型在多模态领域与现有方法在 Twitter-2017 上的性能对比如表 4 所列。CPL 在 Twitter-2017 数据集上相较于其他基线方法，具有最高的准确率，高出 TomBert 1.06%。原因是 TomBert 无法充分挖掘隐式情感；而 CPL 利用大语言模型生成辅助文本，并通过自注意力网络捕捉文本之间的联系，能更加充分地发掘隐式情感地。

表 3 Twitter-2015 数据集上的性能对比

Table 3 Performance comparison on the Twitter-2015 dataset

(%)

模型	准确率	F_1 值
Res-RAM-TFN	69.91	61.49
Res-RAM	71.55	64.68
Res-MGAN-TFN	70.3	64.14
Res-MGAN	71.65	63.88
MIMN	71.84	65.69
ESAFN	73.38	67.37
UMAS	73.48	**73.34**
VilBERT	75.79	71.07
SaliencyBERT	77.03	72.36
TomBert	77.15	71.75
CPL	**77.63**	73.19

表 4 Twitter-2017 数据集上的性能对比

Table 4 Performance comparison on the Twitter-2017 dataset

(%)

模型	准确率	F_1 值
BERT-Pair-QA	63.12	59.66
Res-MGAN-TFN	64.10	59.13
Res-MGAN	66.37	63.04
MIMN	65.88	62.99
ESAFN	67.83	64.22
mPBERT	68.80	67.06
SaliencyBERT	69.69	67.19
EF-CapTriBERT	69.77	68.42
TomBert	70.34	68.03
FITE	70.90	68.70
CPL	**71.96**	**70.40**

4.4.2 消融实验

为验证模型中每个模块的有效性，设计了 3 种模型的变体，进行消融分析。

1)CPL_T：缺失 T5 大语言模型生成辅助文本。

2)CPL_G：不进行难易样本分类。

3)CPL_P：固定提示模板。

变体模型的性能如表 5 所列，分别通过缺失 T5 大语言模型生成辅助文本、不进行样本难易分类和固定提示模板验证了各个部分的作用。

表 5 消融实验结果

Table 5 Ablation experiment results

(%)

数据集	模型	准确率	精确率	召回率	F_1 值
Twitter-2015	CPL_T	75.80	70.28	72.03	71.09
	CPL_G	76.37	72.43	70.33	71.21
	CPL_P	76.86	71.14	71.04	71.09
	CPL	**77.63**	**73.32**	**73.87**	**73.19**
Twitter-2017	CPL_T	69.77	69.31	67.60	68.32
	CPL_G	70.02	68.62	68.36	68.48
	CPL_P	70.10	68.93	69.26	68.60
	CPL	**71.96**	**70.86**	**70.65**	**70.40**

相较于 CPL_P，CPL 取得了较好的性能，其原因是固定的提示模板限制了大语言模型的灵活性。而 CPL_T 在两个数据集的表现最差，其原因是大语言模型生成的辅助文本的缺失，导致模型缺乏对隐式情感的发掘。CPL_G 模块缺少了难易样本的分类，使得大语言模型的提示模板中混杂了偏差因素，引入了更多的噪声。从整体来看，基于置信度的门控模块(CG)和提示构建模块(PC)能够提高模型在情感分类上的表现，证明本文设计的两个模块是合理有效的。

4.4.3 模型表现

本文模型在处理不同类型情感上的表现如表 6 所列。在两个数据集中，CPL 在结果为中性的数据中准确率最高，而中性的情感在 Twitter-2015 中并不是最多的，这也可以说明 CPL 在训练过程中没有学习到数据集中结果分布不均匀导致的误差。此外，在 Twitter-2017 中，尽管训练集中消极的样本远远少于其他类别的样本，CPL 依旧能够对样本做出正确的预测，这也说明了本文模型的有效性。

表 6 CPL 处理不同类型情感的表现

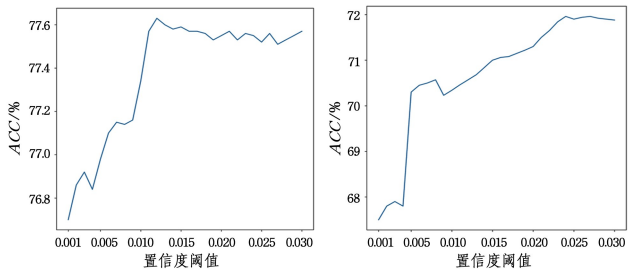
Table 6 Performance of CPL in processing different sentiment

(%)

	Twitter-2015			Twitter-2017		
	积极	中性	消极	积极	中性	消极
准确率	75.08	80.54	75.19	70.85	72.34	71.89
F1 值	67.28	82.16	73.14	73.99	70.07	71.38

4.4.4 置信度表现

不同置信度阈值下 CPL 预测的准确率如图 4 所示。从图 4 中可以看出，准确率随着置信度阈值的增加，在达到最高值后便趋于稳定。原因是阈值很小时，模型对情感的判断几乎是无差别的，以至于分类错误；在达到准确率最高值对应的阈值时，模型能充分利用大语言模型生成的辅助文本挖掘隐式情感，以进行精准的预测。



(a)Twitter-2015 (b)Twitter-2017

图 4 不同置信度阈值对应的准确率

Fig. 4 Accuracy of different confidence thresholds

4.5 样例分析

图 5 给出了本文模型在 Twitter-2015 数据集上的一些成功与失败的样例。前两行展示了本文模型的成功样例，后两行展示了本文模型的失败样例。从中可以看出，本文模型对于文本表述清晰的样本预测结果更加准确；前两行的例子相对来说，文本中不具备过多的干扰因素；而第 4 个例子的文本“Calories，今天不作数，全国冰淇淋日快乐”，前后情感均趋向积极，但对于“Calories”这个词来说，积极的情感并不对其造成影响，所以它是中性的，而这种形式的文本会引入噪声，影响模型对潜在情绪的理解，造成模型误判。以上结果也说明，大语言模型的强大推理能力虽然能处理大部分样本，但在一些文本中有太多噪声因素的情况下依旧会受到干扰。

图片	文本	目标词	图片描述	预测结果
	First snow fall of the year in **Summit County**, Colorado, USA :) Summit County	Summit County	a general view of the mountains	真实值：Neutral 预测值：Neutral ✓
	I always have a sweet spot for **Kate Hudson**! ! # MetGala	Kate Hudson	person in a gold gown	真实值：Positive 预测值：Positive ✓
	ENOUGH WITH ISRAEL **Israel** PM : Iran Nuke deal will make war in Middle East	Israel	a group of people sitting around a table	真实值：Neutral 预测值：Negative ✗
	RT @ Foodimentary : **Calories** don ' t count today ! Happy National Ice Cream Day !	Calories	a set of ice creams in different flavors	真实值：Neutral 预测值：Positive ✗

图 5 CPL 在 Twitter-2015 数据集上的成功和失败样例

Fig. 5 Success and failure examples of CPL on Twitter-2015 dataset

结束语 为了突破固定提示模板给大模型推理带来的局限性，并增强多模态方面级情感分类的可解释性，本文提出了基于置信度引导大模型提示学习的多模态方面级情感分析模型(CPL)。该模型由置信度门控模块(CG)和提示构建模块(PC)构成。CG 模块基于分类置信度确定不同样本的分类难易程度；PC 模块则根据不同的难度设计不同的模板，有效地增加了大语言模型提问的灵活性。本文模型在 Twitter-2015 和 Twitter-2017 数据集上，相比于其他基线模型，有着更优的表现。但是 CPL 在处理文本中含有大量噪声的情况时表现并不理想，其原因是，在多模态情感分析中，相比于其他模态，文本模态对结果的影响更大，其数据中的固有噪声导致难以挖掘隐藏在其中的情感。随着大模型技术的不断发展，未来将继续在大模型层面对多模态方面级情感分析模型进行优化，深入探索社交媒体数据中的情感本质特点，分析不同线索的可信程度，进一步提升多模态方面级情感分析模型的性能。

参 考 文 献

[1] AGARWALA, YADAV A, VISHWAKARMA D K. Multimodal sentiment analysis via RNN variants[C]// 2019 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering(BCD). IEEE, 2019: 19-23.

[2] TRUONG QT, LAUW H W. Vistanet: Visual aspect attention network for multimodal sentiment analysis[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 305-312.

[3] ZHANGD, LI S, ZHU Q, et al. Modeling the clause-level structure to multimodal sentiment analysis via reinforcement learning [C]// 2019 IEEE International Conference on Multimedia and Expo(ICME). IEEE, 2019: 730-735.

[4] YANG J, YU Y, NIU D, et al. ConFEDE: Contrastive Feature Decomposition for Multimodal Sentiment Analysis[C]// Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023: 7617-7630.

[5] MAF, ZHANG Y, SUN X. Multimodal Sentiment Analysis with Preferential Fusion and Distance-aware Contrastive Learning [C]// 2023 IEEE International Conference on Multimedia and Expo(ICME). IEEE, 2023: 1367-1372.

[6] ZHOU R, GUO W, LIU X, et al. AoM: Detecting aspect-oriented information for multimodal aspect-based sentiment analysis [J]. arXiv: 2306. 01004, 2023.

[7] YU Y, ZHAO M, QI S, et al. ConKI: Contrastive knowledge injection for multimodal sentiment analysis [J]. arXiv: 2306. 15796, 2023.

[8] TANG D, QIN B, FENG X, et al. Effective LSTMs for target-dependent sentiment classification[J]. arXiv: 1512. 01100, 2015.

[9] WUH, CHENG S, WANG J, et al. Multimodal aspect extraction with region-aware alignment network[C]// CCF International Conference on Natural Language Processing and Chinese Computing. Cham: Springer, 2020: 145-156.

[10] ZHANG Q, FU J, LIU X, et al. Adaptive co-attention network for named entity recognition in tweets[C]// Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. New Orleans: AAAI Press, 2018: 5674-5681.

[11] ASGARI-CHENAGHLU M, FEIZI-DERAKHSHI M R, FARZINVASH L, et al. CWI: A multimodal deep learning approach for named entity recognition from social media using character, word and image features[J]. Neural Computing and Applications, 2022, 34: 1905-1922.

[12] LING Y, YU J, XIA R. Vision-language pre-training for multi-

modal aspect-based sentiment analysis[J]. arXiv:2204.07955, 2022.

[13] YANGL, NA J C, YU J. Cross-modal multitask transformer for end-to-end multimodal aspect-based sentiment analysis[J]. Information Processing & Management, 2022, 59(5): 103038.

[14] ZHAOF, WU Z, LONG S, et al. Learning from adjective-noun pairs: A knowledge-enhanced framework for target-oriented multimodal sentiment classification[C]// Proceedings of the 29th International Conference on Computational Linguistics. 2022: 6784-6794.

[15] ZHAOF, LI C, WU Z, et al. M2DF: Multi-grained Multi-curriculum Denoising Framework for Multimodal Aspect-based Sentiment Analysis[J]. arXiv: 2310.14605, 2023.

[16] ZHOUR, GUO W, LIU X, et al. AoM: Detecting aspect-oriented information for multimodal aspect-based sentiment analysis[J]. arXiv: 2306.01004, 2023.

[17] FEI H, LI B, LIU Q, et al. Reasoning implicit sentiment with chain-of-thought prompting[J]. arXiv: 2305.11255, 2023.

[18] ZHUX, KUANG Z, ZHANG L. A prompt model with combined semantic refinement for aspect sentiment analysis[J]. Information Processing & Management, 2023, 60(5): 103462.

[19] YANGX, FENG S, WANG D, et al. Few-shot multimodal sentiment analysis based on multimodal probabilistic fusion prompts [C]// Proceedings of the 31st ACM International Conference on Multimedia. 2023: 6045-6053.

[20] LIUD, LI L, TAO X, et al. Descriptive Prompt Paraphrasing for Target-Oriented Multimodal Sentiment Classification[C]// The 2023 Conference on Empirical Methods in Natural Language Processing. 2023.

[21] BAO Y, LI X, REN F. 3M: Multi-Task Multi-Prompt Learning Model for Aspect Based Sentiment Analysis[C]// 2023 IEEE 9th International Conference on Cloud Computing and Intelligent Systems(CCIS). IEEE, 2023: 451-456.

[22] SHIJ, LI W, BAI Q, et al. Soft prompt enhanced joint learning for cross-domain aspect-based sentiment analysis[J]. Intelligent Systems with Applications, 2023, 20: 200292.

[23] LIC, GAO F, BU J, et al. Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis[J]. arXiv: 2109.08306, 2021.

[24] LIJ, LI D, XIONG C, et al. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation[C]// International Conference on Machine Learning. PMLR, 2022: 12888-12900.

[25] YU J, JIANG J. Adapting BERT for target-oriented multimodal sentiment classification. [C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. 2019: 5408-5414.

[26] VASWANIA, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2017: 6000-6010.

[27] ZADEH A, CHEN M, PORIA S, et al. Tensor fusion network for multimodal sentiment analysis[J]. arXiv: 1707.07250, 2017.

[28] XU N, MAO W, CHEN G. Multi-interactive memory network for aspect based multimodal sentiment analysis[C]// Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu, Hawaii: AAAI Press, 2019: 371-378.

[29] YUJ, JIANG J, XIA R. Entity-sensitive attention and fusion network for entity-level multimodal sentiment classification[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 28: 429-439.

[30] ZHOUR, ZHU H Z, GUO W Y, et al. A Unified Framework Based on Multimodal Aspect-Term Extraction and Aspect-Level Sentiment Classification [J]. Journal of Computer Research and Development, 2023, 60(12): 2877-2889.

[31] WANGJ, LIU Z, SHENG V, et al. SaliencyBERT: Recurrent attention network for target-oriented multimodal sentiment classification[C]// Chinese Conference on Pattern Recognition and Computer Vision(PRCV). Cham: Springer, 2021: 3-15.

[32] LUJ, BATRA D, PARIKH D, et al. VilBERT: pretraining task-agnostic visiolinguistic representations for vision-and-language tasks[J]. Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2019: 12-23.

[33] SUN C, HUANG L, QIU X. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence[J]. arXiv: 1903.09588, 2019.

[34] YANG H, ZHAO Y, QIN B. Face-sensitive image-to-emotional-text cross-modal translation for multimodal aspect-based sentiment analysis[C]// Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. 2022: 3324-3335.

[35] KHAN Z, FU Y. Exploiting BERT for multimodal target sentiment classification through input space translation[C]// Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 3034-3042.

LI Maolin, born in 2001, postgraduate, is a member of CCF(No. V1444G). His main research interests include deep learning and sentiment analysis.

YANG Zhenguo, born in 1988, Ph.D, associate professor, is a member of CCF (No. 77775M). His main research interests include online event detection and domain adaptation.

（责任编辑：柯颖）