学校代码：**1006**

学　　号：

SOOCHOW UNIVERSITY

硕士学位论文

（学术学位）

TreeCRF-based High-Order Syntactic Parsing

研 究 生 姓 名 _____

指导教师姓名 _____

专 业 名 称 _____

研 究 方 向 _____

所 在 院 部 _____

论文提交日期 _____ 年　月

苏州大学学位论文独创性声明

本人郑重声明：所提交的学位论文是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含其他个人或集体已经发表或撰写过的研究成果，也不含为获得苏州大学或其它教育机构的学位证书而使用过的材料。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人承担本声明的法律责任。

论文作者签名：张宇　　日　期：2021.6.8

苏州大学学位论文使用授权声明

本人完全了解苏州大学关于收集、保存和使用学位论文的规定，即：学位论文著作权归属苏州大学。本学位论文电子文档的内容和纸质论文的内容相一致。苏州大学有权向国家图书馆、中国社科院文献信息情报中心、中国科学技术信息研究所（含万方数据电子出版社）、中国学术期刊（光盘版）电子杂志社送交本学位论文的复印件和电子文档，允许论文被查阅和借阅，可以采用影印、缩印或其他复制手段保存和汇编学位论文，可以将学位论文的全部或部分内容编入有关数据库进行检索。

涉密论文□

本学位论文属　　在_____年___月解密后适用本规定。

非涉密论文☑

论文作者签名：张宇　　日　期：2021.6.8

导师签名：李正华　　日　期：2021.6.8

摘 要

句法分析任务是句子理解的重要中间过程之一. 其中，概率估计一直是句法分析领域的一个核心问题. 然而，无论是神经网络方法还是深度学习时代以前的方法，采用基于全局概率模型的句法分析工作都非常少，主要的原因在于树形条件随机场（TreeCRF）推断的高复杂度. 在本文中，我们提出将 TreeCRF 应用到依存句法和成分句法这两个主要的句法分析任务. 为了解决 TreeCRF 的低效问题，关键的想法是批次化树结构的推断算法，并且用基于自动求导的反向传播代替 Outside 算法. 目前句法模型被不断简化，采用局部损失目标是当前句法分析方法的一个趋势，我们则进一步在一阶 TreeCRF 的基础上采用了高阶拓展. 高阶 TreeCRF 进一步增加了算法复杂度，为此，我们还提出利用基于平均场变分推断的近似推断算法代替精确推断的 TreeCRF 方法，从而增加了解析效率.

关键词：句法分析，依存句法分析，成分句法分析，树形条件随机场，变分推断

作　　者：张　宇

指导老师：李正华

TreeCRF-based High-Order Syntactic Parsing

Abstract

Syntactic parsing is one of the most important intermediate processes in sentence comprehension, and probability estimation has always been a core problem in the parsing field. However, in either deep learning (DL) era or pre-DL era, there exist very few works based on global probabilistic modeling, mainly due to the high complexity of tree-structure CRF (TreeCRF) inference. This thesis proposes to apply TreeCRF to both dependency parsing and constituency parsing. The key idea to solve the inefficiency issue is to batchify the inference algorithm for tree structures, and meanwhile avoid the complex Outside algorithm via back-propagation. Currently, parsing models are greatly simplified, and it's a trend to adopt local loss for syntactic parsing. In contrast, we propose a high-order extension to first-order models. While high-order modeling further increases the algorithm complexity, we also try to apply mean field variational inference (MFVI) as an alternative to exact inference of TreeCRF method, which greatly improves the parsing efficiency.

Key words: Syntactic Parsing, Dependency Parsing, Constituency Parsing, TreeCRF, Variational Inference

Written by Yu Zhang

Supervised by Zhenghua Li

目　　录

第一章　绪论

1.1　研究背景和意义

中文参考文献[?]

英文参考[?]

第二章　图文情感相关性语料构建与数据分析

现有图文关系定义仅关注文本方面与图像的显式对齐或全局相关性，忽略图文之间隐式情感关联。本章针对该问题从情感表达角度定义方面级跨模态情感相关性，构建数据语料并进行数据分析。同时设计了基于情感相关性感知的多模态方面情感分类模型，为后续研究提供了统一的基准与评估框架。

2.1　引言

多模态方面级情感分类是情感分析领域中一项关键的细粒度任务，其核心目标是判断文本-图像对中针对文本所提及特定方面的情感极性 (即积极、消极或中性)。例如在图2-1中，多模态方面级情感分类的任务是识别出针对“Trey Songz”这一特定方面的情感为积极。近年来，多模态方面级情感分类已取得显著进展[1–3]。随着社交媒体平台愈发多地采用多模态内容进行信息传播，精准的多模态方面级情感分类对提升信息理解能力至关重要，同时也能为观点挖掘[4]、推荐系统[5] 和虚假信息检测[6] 等任务提供有力支持。

与仅考虑单一文本模态的文本情感分析不同，多模态情感分析需要融合文本和视觉语义信息来判断情感极性。因此，在多模态场景中，文本与图像的关系对情感分类起着关键作用：相关图像能够补充文本信息，助力更全面地理解情感；而无关图像则可能引入噪声，误导模型做出错误的情感预测。基于上述考量，已有研究尝试利用图文跨模态相关性来提升情感分析性能。Ju 等人[7] 采用了 Vempala 和 Preoţiuc-Pietro[8] 提出的跨模态关系方案，并取得了性能提升。该方案从全局视角考虑文本整体与图像的关系：当图像能够直观描述或补充文本内容时，即认为其与文本整体相关。尽管该标注方案已被证实对多模态方面级情感分类有益，但在实际辅助多模态方面级情感分类任务时仍存在局限：多模态方面级情感分类的核心是关注文本中每个方面的情感，而在包含多个方面的文本中，单个方面与图像的相关性未必与全局图文相关性一致。仅依赖全局相关性可能引入无关视觉信息，最终导致特定方面的情感判断不准确。例如在图2-1-(a) 中，图像与文本整体相关，但“Oklahoma”这一方面与图像无关，此时依赖全局相关性会导致情感预测错误。为捕捉细粒度的方面级跨模态相关性，Yu 等人[9] 提出了一种方面-图像相关性数据集。该方案基于显式对齐来定义文本方面与

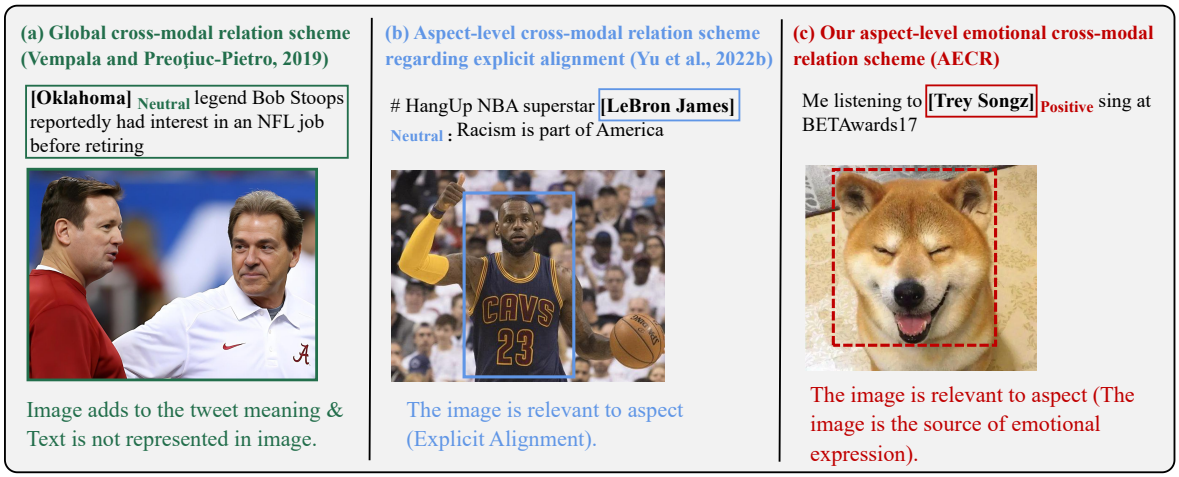


图 2-1　本章提出的跨模态相关性定义与 Vempala[8] 等人和 Yu[9] 等人所提出的相关性定义的对比示例。图中的实线框代表显式对齐关系，虚线框则代表隐性情感关联；相应地，文本中的实线框用于突出标注与图中实线框或虚线框内容相关联的文本信息。

图像的相关性，即当图像明确描绘或包含文本中提及的方面时，认为二者相关。如图2-1-(b) 中，“勒布朗　詹姆斯 (LeBron James)”这一方面在图像中清晰呈现，因此判定该方面与图像相关。尽管全局跨模态相关性和方面级跨模态相关性均能提升多模态方面级情感分类的性能，但这些方法仍忽略了隐式情感关联，即当方面未在图像中显式出现时，图像仍能增强或传递针对该方面的情感。例如在图2-1-(c) 中，“Trey Songz”未在图像中显式出现，但小狗欢快的表情传递了针对该方面的积极情感。

为解决上述局限，本章提出一种方面级情感跨模态相关性 (AECR)，用于捕捉文本的方面与图像间的显式对齐和隐式情感关联。如图2-1 所示，尽管“Trey Songz”未在图像中显式呈现，但用户通过图像中小狗的愉悦表情表达了对该方面的积极情感，因此在该跨模态关系方案中，图像与该方面存在隐式情感关联。具体而言，图像是整个图文对的情感来源，因此二者相关。基于该跨模态关系方案，本文构建了方面级情感跨模态相关性数据集 (AECR-Twitter)。为验证所提跨模态关系方案和新构建数据集在下游多模态方面级情感分类任务中的有效性，本章探索并对比了多种将跨模态相关性信息融入多模态方面级情感分类的方法，包括轻量级多模态模型和大型语言模型 (LLMs)，并在 Twitter-15 和 Twitter-17 两个广泛使用的数据集上进行了大量实验。

2.2　图文情感相关性语料构建

本节首先详细介绍图文情感相关性的标注规范以及提供相应的实例说明。紧接着，本节对方面级情感跨模态关系数据集的构建过程进行详细描述，包括标注数据的

Happy birthday **[Garrett Walker]**$_{\text{Positive}}$!

(a) The image is semantically related to the text but emotionally irrelevant

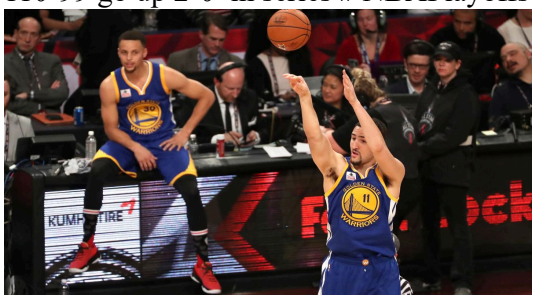
[SamHunt]$_{\text{Positive}}$ Performs at Stagecoach.

(b) The image is the source of emotional expression

[Justin Bieber]$_{\text{Negative}}$ broke history and global records but he only won 2 awards.

(c) The image enhances the emotional expression conveyed in the text

Warriors overwhelm **[TrailBlazers]**$_{\text{Negative}}$ 110-99 go up 2-0 in series # NBAPlayoffs



(d) The image is irrelevant to aspect

图 2-2　本章中四种相关性类型的示例

选取、标注工具的使用、标注团队的组建与培训以及标注的具体流程

2.2.1　标注规范

本章在综合前人跨模态关系定义[8,10–12] 后，并结合多模态方面级情感分类任务特点，提出一种情感语义感知的方面级图文关系标注体系。该方案基于图文对中显式表达的对象和隐式传递的情感语义信息，将图文关系划分为“相关”和“不相关”两类。基于此方案，我们制定了包含详细说明和示例的标注规范，旨在构建高质量的方面级情感跨模态关系数据，为下游方面级情感分析任务提供支持。具体而言，“相关”的图文关系包含以下三种类型：

- 语义相关但情感无关：图像明确包含文本中提到的方面，但未提供任何相关情感信息。图2-2-(a) 给出了一个示例：图像中包含“Garrett Walker”这一方面 (人物实体)，但未传递与推文相关的任何情感信息。
- 情感表达来源：图像中的情感语义信息是判断整体多模态情感的依据，仅依靠

文本无法完成这一判断。例如，在图2-2-(b) 中，图像中人物的表情补充了情感表达，明确了文本中对“SamHunt”这一方面的模糊情感倾向，因此该图像可作为该方面的情感来源。

- 增强情感表达：图像通过提供额外的情感语义信息，强化了方面的情感表达。例如，在图2-2-(c) 中，图像中女性的悲伤表情，强化了文本中对“Justin Bieber”仅获得两项奖项的遗憾情绪，直观体现了图像对情感表达的增强作用。

“不相关”的图文关系指图文对不属于上述任何一类情况，即图像既不包含文本中提到的方面，也不提供任何额外的情感相关信息。例如，在图2-2-(d) 中，方面“TrailBlazers”指NBA开拓者队，但图像展示的是NBA勇士队，因此该图像与“Trail-Blazers”这一方面显然不相关。在此情况下考虑图像内容只会引入无关信息。与已有的标注体系不同，以往方案要么仅将文本中的方面与图像中的对象显式对齐时定义为相关[9]，其余情况均视为不相关；要么聚焦于文本整体与图像的全局相关性[8,12]，而非方面级相关性。而本文提出的方案同时考虑了文本与图像之间的方面级显式对齐和隐式情感语义关系。这意味着，即使方面未显式出现在图像中，只要图文双方共同参与整体情感的表达，仍将其判定为相关。例如，在图2-2-(c) 中，文本中的“Justin Bieber”这一方面与图像中的对象并无显式对齐，但图像强化了对“Justin Bieber”的遗憾和悲伤情绪。识别这种隐式情感相关性，有助于我们更好地利用以往被忽视但有价值的图文关系，最终提升方面级情感分析的性能。

2.2.2　人工标注数据构建

本章采用一套科学的数据标注流程来对图文方面级情感分析数据集进行高质量的人工标注。具体地，为确保标注的准确性，本研究采取严格的双人独立标注方法，即两位标注员进行独立标注，当两位标注员给出不一致的结果时，将由一位第三方专家介入处理这种不一致，以确定最终的正确答案。通过该标注过程，成功构建了一个高质量的多模态方面级图文情感相关性数据集 (Aspect-level Emotional Cross-modal Relation dataset, AECR-Twitter)，该数据集包括3562个样本。接下来，本小节将从标注数据的选择、标注工具、标注人员的招聘与培训和标注质量的控制等方面详细介绍数据构建的具体细节。

1. 标注数据的选择

本文构建的方面级情感跨模态关系数据集以 Twitter-17 数据集[13] 为基础，遵循先前相关研究的做法[9,12]。Twitter-17 数据集包含从社交媒体平台 Twitter 收集的多模态帖子，涵盖文本 - 图像对，且每个文本 - 图像对均标注了文本方面及对应方面的情感极性。我们选取 Twitter-17 数据集中的训练集 (含 3562 个样本) 进行标注，标注内容包括：1. 单模态语境下各方面的情感标签；2. 方面级情感跨模态关系。其中，单模态语境下的方面情感标注将作为辅助信号，助力更精准地标注方面级情感跨模态关系类型。随后，我们将构建的 AECR-Twitter 数据集按 8:1:1 的比例随机划分为训练集、验证集和测试集，详细统计信息见表 3。值得注意的是，该数据集经过精心筛选，已剔除有害内容。

2. 标注工具

图中展示了标注工具的标注界面。对于每个需要标注的图文对，标注工具会显示该图文对以及目标方面。当根据我们的标注流程，在我们标注工需要经过三个步骤，首先是只显示文本和目标方面，选择目标方面对应的情感极性，接着第二个步骤中，标注工具只会显示图片和目标方面，需要根据图片内容去标注方面的情感，若图片与方面无关那么需要标注无情感。在最后的步骤中会展示完整的图文对和目标方面，并提供数据集中已有的金标情感极性，结合上述信息来分别从信息层面和情感层面来标注方面图文情感相关性。

3. 标注人员的招聘与培训

为了确保数据集的高质量标注，本研究聘请 4 名具备自然语言处理或多模态分析相关背景的研究生作为标注员，以及 1 名经验丰富的资深标注员作为专家仲裁。其中，专家的核心职责是协调并最终裁决标注过程中出现的分歧。所有标注人员的报酬是根据他们的标注质量和完成的工作量来决定的，以此激励标注员严谨、细致地执行标注任务。为确保标注判断的全面性与客观性，我们为每个推文图文对设计了“分阶段递进”的标注流程，避免标注员仅凭单一模态信息仓促下结论：

- 第一步 (文本单模态标注)：仅向标注员提供推文文本和对应的方面实体，要求其仅基于纯文本内容判断该方面的情感极性，旨在建立纯文本维度的情感基准，为后续跨模态关系判断提供参考。

- 第二步 (图像单模态标注)：仅向标注员提供图像和对应的方面实体，要求其仅基于视觉内容 (如图像中的人物表情、场景氛围等) 判断该方面的情感倾向，用于获取纯视觉维度的情感线索，与文本维度形成互补。
- 第三步 (方面级图文情感相关性判断)：向标注员完整呈现图文对，并提供 Twitter-17 数据集中已有的多模态方面级情感标签作为参考，要求标注员结合前两步的独立判断结果，综合评估并最终确定“方面级情感跨模态关系”的具体类型。这种“先分后合”的流程，能有效减少单一模态信息的误导，提升跨模态关系判断的准确性。

在正式启动标注工作前，所有标注员需参加专项培训：一方面，详细解读标注准则 (包括 4 种关系类型的定义、边界案例的判断标准等)，通过典型示例演示帮助标注员形成统一认知；另一方面，指导标注员熟练操作我们自主开发的标注工具，确保其能高效、规范地记录标注结果，避免因工具使用不熟练或对规则理解不一致导致的系统性误差。

4. 标注质量的控制

在标注过程中，本研究基于开发的标注工具进行严格的独立双重标注，以保证标注数据的质量，工作流程如图所示。具体来说，每个图文对被随机分配给两名不同的标注者，由二者独立完成方面级情感相关性标注；若两名标注者的标注结果一致，则直接将该结果作为最终标注；若结果不一致，由第三位专家标注者在分析双方标注后确定最终答案。

2.3　方面级情感跨模态相关性数据集分析

为了评估构建的 AECR-Twitter 数据集的质量，本节对其进行了标注一致性分析。此外，为了探索相关性标签和情感标签的相关关系，以及方面级图文情感相关性和前人标注体系的区别，本节从不同的角度对对标注的 AECR-Twitter 数据集进行了分析。

2.3.1　AECR 标注一致性分析

在标注过程中每项标注任务会分配给两名标注者进行独立标注。我们针对标注完成的跨模态关系计算了标注者间一致性比例，结果显示 84% 的任务中两名标注者得出了一致结果，另有 16% 的任务需由第三位专家标注者进行进一步审核。经调查

表 2-1　我们的 AECR-Twitter 数据集标签分布情况

Relevance	Type	Number	Proportion
Relevant	The image is semantically related to the text but emotionally irrelevant	868	24.37%
	The image enhances the emotional expression conveyed in the text	730	20.49%
	The image is the source of emotional expression	248	6.96%
Irrelevant	The image is irrelevant to aspect	1,716	48.18%

表 2-2　与前人工作的标注比较

	Relevant	Irrelevant
Yu 等[9]	39.1%	60.9%
Our annotation	45.3% (Explicit: 37.7% Implicit: 7.6%)	54.7%

发现，标注不一致主要原因在于标注者对图像和文本中情感表达的解读存在差异。这表明，执行严格的双重标注是保障数据质量的关键。

2.3.2　AECR 标签分布

表2-1展示了 AECR 数据集中标签的分布情况。由表可以得知，其中，“Irrelevant”(无关) 标签占比 48.18%。这一现象反映了社交媒体平台的内容特性，主要由两方面原因造成。首先，在同时包含多个方面 (aspect) 的推文中，图像通常只与其中一个方面相关，而与其他方面无关。其次，用户在发布推文时也可能附上与文本内容完全无关的图像。对于“Relevant”(相关) 标签，最常见的类型是：图像呈现了文本中的方面，但不包含任何情感信息，占比 24.37%。第二常见的类型是图像通过提供额外的情感线索来增强文本的情感表达。最少见的类型是推文的情感完全通过图像传达。这表明用户更倾向于使用文本作为表达情感的主要模态，而图像通常用于辅助文本；仅通过图像传递情感的情况相对较少。

2.3.3　与前人工作的标注比较

我们将数据集中相关与无关标签的分布与已有研究进行对比，并在表2-2中给出了显性与隐性相关性的比例。与 Yu 等[9] 相比，其仅将具有显性对齐关系的文本-图像对视为相关，而忽略了隐性的情感关联，因此我们的数据集由于纳入了隐性的跨模

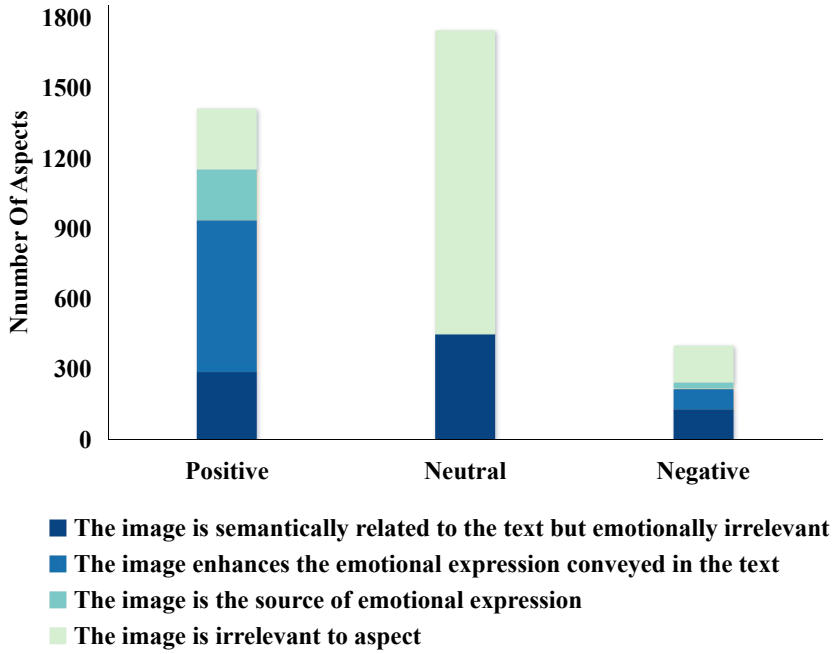


图 2-3　AECR-Twitter 数据集中不同相关性类型与情感类别之间的分布情况

态相关性，具有更高比例的相关标签。在我们的所标注的相关性类型中，显性相关与隐性相关分别占 37.7% 和 7.6%。这表明，在图文相关的场景中，大约六分之一的情感表达是通过隐含情感关联的图像 (如表情包等) 来传递的。与以往的跨模态关系标注方案相比，我们的方法能够更有效地捕捉图像与文本之间这些隐性的情感联系，从而更好地支持下游的多模态方面级情感分类 (MASC) 任务。

2.3.4　情感标签与相关性标签关系分析

我们分析不同相关性类型与情感类别之间的关系。如图2-3所示，当情感为中性时，占比最高的情况是该方面与图像无关，其次是图像与方面在语义上相关但不包含情感联系的情况。这表明，当用户对某一方面实体没有情感倾向时，图像往往与该方面无关，或仅作为描述性补充。在积极情感的场景中，占比最高的情况是图像能够增强方面的情感表达，说明用户倾向于通过文本与图像的共同作用来强化积极情绪。在消极情感中，图像无关、语义相关但情感无关，以及图像增强情感表达三类情况的比例较为接近。这说明负面情绪的表达方式更加多样化。

2.4　融合跨模态方面级情感相关性的多模态情感分析

本章提出的方法主要由三个部分构成。首先，我们给出了任务的形式化定义，明确界定了基于方面的跨模态关系推理任务、多模态方面级情感分类（MASC）任务，以及两者之间的内在联系。其次，我们总结了现有 MASC 模型采用的通用架构。最后，我们详细阐述了本文提出的模块，旨在引入方面级的情感跨模态相关性。

2.4.1　任务定义

为了降低无关图像的干扰并提升情感分类性能，我们将本文提出的“方面级情感跨模态相关性”引入到 MASC 任务中。具体而言，基于方面的跨模态关系推理任务旨在预测图像与文本中给定方面之间的相关性。随后，利用该相关性概率抑制无关图像的影响，从而改进多模态方面级情感分类任务的效果。

具体来说，数据集中的每个样本 $\mathbf{x}$ 包含一条推文，该推文由文本-图像对及其对应的方面组成。其中，x^t、x^i 和 x^a 分别表示文本、图像和方面。文本 x^t 是由 n 个单词组成的句子；而方面 x^a 是指文本中的命名实体（如人名、地名或组织名），它是 x^t 的一个包含 m 个单词的子序列。

基于方面的跨模态关系推理形式化为一个二分类问题。给定输入 $\mathbf{x} = (x^t, x^i, x^a)$，其目标是预测图像 x^i 与文本 x^t 针对给定方面 x^a 的关系 r，其中 $r \in \{relevant, irrelevant\}$。

多模态方面级情感分类旨在针对给定文本 x^t 和图像 x^i 对中的方面 x^a，预测其情感极性标签 $y \in \{positive, neutral, negative\}$。

2.4.2　现有 MASC 模型的通用架构

基础 MASC 模型的架构主要包含三个核心组件：文本/视觉编码器用于将文本和图像编码为特征表示；方面感知提取模块提取两个模态中聚焦于方面相关信息的特征表示；情感分类模块则融合文本和视觉特征以获得多模态表示，并将其输入 Softmax 层进行情感预测。

文本/视觉编码器。给定输入 $\mathbf{x} = (x^t, x^i, x^a)$（包含文本-图像对及一个方面），我们分别使用文本编码器和视觉编码器来提取文本和视觉模态的上下文特征表示。对于文本编码器，我们采用 RoBERTa[14] 或 BERT[15] 对文本 x^t 和方面 x^a 的拼接序列进行编码。对于视觉编码器，我们利用 Faster R-CNN[16] 或 ViT[17]。最终得到的文本和图

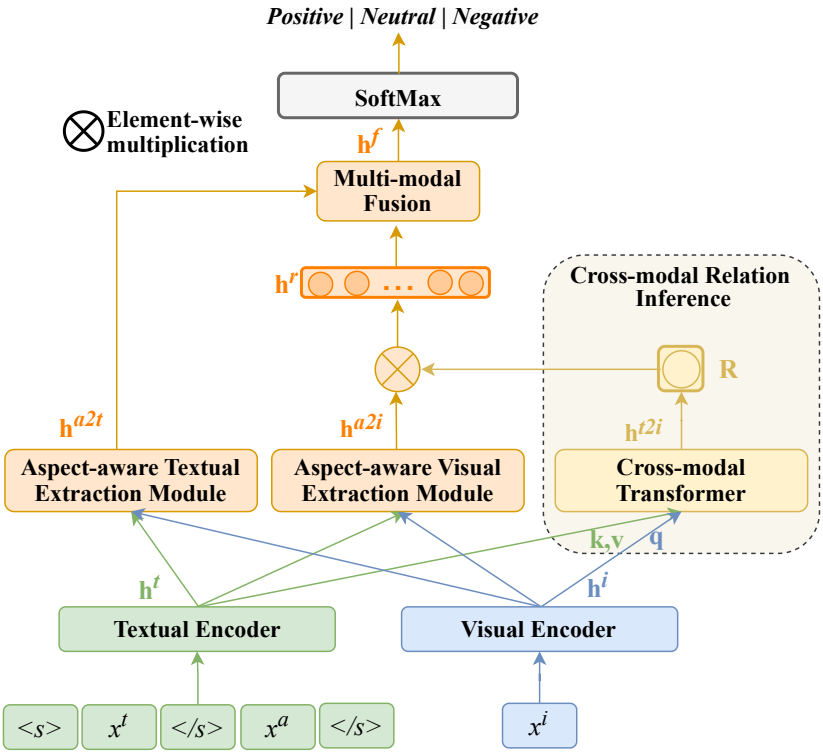


图 2-4　EmoCRel-MASC 架构概览

像表示分别记为 $\mathbf{h}^t$ 和 $\mathbf{h}^i$。

方面感知提取模块。在获得文本和图像表示后，我们将它们输入到方面感知提取模块中。该模块从文本和图像中提取并提炼与方面相关的表示。该模块可以通过多种方法实现，其中包括 ITM[9]、HIMT[18] 和 DPFN[19] 等框架中的模块。对于方面感知的文本表示，应用注意力机制或依存句法分析器等方法来获取方面感知的文本表示 $\mathbf{h}^{a2t}$。对于方面感知的视觉表示，则应用跨模态 Transformer (CMT)[20] 等方法来捕获方面感知的视觉表示 $\mathbf{h}^{a2i}$。

情感分类模块。我们通常采用 Wang 等[19] 中的拼接方法或 Yu 等[9,18] 中的 Transformer 方法来生成最终的多模态融合表示 $\mathbf{h}^f$。随后，我们将 $\mathbf{h}^f$ 输入 Softmax 层，以预测情感标签 $\mathbf{y}$ 的概率分布，从而进行情感分类：

$$p(\mathbf{y}) = \text{Softmax}(\mathbf{W}_s\mathbf{h}^f + \mathbf{b}_s) \tag{2.1}$$

在训练过程中，对于每个输入 $\mathbf{x}$，我们使用交叉熵损失最大化真实情感标签 y^* 的概率，目标函数如下：

$$\mathcal{L}_s = -\log P(y^*) \tag{2.2}$$

2.4.3　跨模态情感相关性感知的的情感分类增强

如图 2-4 所示，我们提出了一种跨模态情感相关性感知的多模态方面级情感分类方法（EmoCRel-MASC），旨在将预测的相关性有效地整合到 MASC 任务中。跨模态关系推理模块输出一个相关性概率，用于控制图像的贡献度；我们使用 AECR-Twitter 数据集作为监督信号来训练该模块。随后，关系感知的多模态融合模块利用预测的相关性概率来指导文本和视觉特征的融合，从而有效地抑制无关图像的影响。

方面级情感跨模态关系推理。为了将本文提出的方面级情感跨模态关系引入 MASC 任务，我们采用了一个跨模态关系推理模块来执行基于方面的跨模态关系推理，并获取相应的相关性概率。该概率用于控制图像的贡献度，并减少无关图像的干扰。具体而言，我们首先应用跨模态 Transformer（CMT）来捕获跨模态图像表示，其中图像表示 $\mathbf{h}^i$ 作为查询向量 **q**，文本表示 $\mathbf{h}^t$ 作为 Key **k** 和 Value **v**。接着，执行最大池化操作以获取用于跨模态关系推理的最显著跨模态特征 $\mathbf{h}^{t2i}$：

$$\mathbf{h}^{t2i} = \text{maxpool}(\text{CMT}(\mathbf{h}^i, \mathbf{h}^t, \mathbf{h}^t)) \tag{2.3}$$

最终的相关性概率计算如下。值得注意的是，我们将该模块预测为“相关”类别的概率作为相关性概率。

$$p(\mathbf{r}) = \text{Sigmoid}(\mathbf{W_r}\mathbf{h}^{t2i} + \mathbf{b_r}) \tag{2.4}$$

关系感知的跨模态融合。为了缓解 MASC 任务中无关图像的干扰，我们设计了一个关系感知的多模态融合模块，利用相关性概率引导模型关注相关图像，同时最小化无关图像的影响。具体而言，我们首先利用方面感知视觉提取模块来获取方面感知的视觉表示 $\mathbf{h}^{a2i}$。然后，我们将跨模态相关性概率 $p(r)$ 进行广播（broadcast）以匹配 $\mathbf{h}^{a2i}$ 的维度，得到矩阵 **R**。该矩阵 **R** 被用作权重，通过逐元素乘法获得关系感知的跨模态表示 $\mathbf{h}^r$：

$$\mathbf{h}^r = \mathbf{R} \odot \mathbf{h}^{a2i} \tag{2.5}$$

例如，当相关性概率 **R** 趋近于 0 时，表明图像与文本不相关，模型相应地在多模态融合过程中降低视觉特征的贡献。最后，我们将关系感知的跨模态表示 $\mathbf{h}^r$ 和方面感知

的文本表示 $\mathbf{h}^{a2t}$ 输入到上述多模态融合模块中，生成增强的多模态融合表示 $\mathbf{h}^{AECR-f}$。随后，我们将 $\mathbf{h}^{AECR-f}$ 作为最终的多模态特征输入 Softmax 层进行情感分类。此外，对于每个输入 $\mathbf{x}$，我们使用两个独立的交叉熵损失函数来分别最大化其真实跨模态关系标签 r^* 和真实情感标签 y^* 的概率，联合目标函数为：

$$\mathcal{L} = \mathcal{L}_s + \mathcal{L}_r = -\log P(y^*) - \log P(r^*) \tag{2.6}$$

在推理阶段，我们选择得分最高的标签作为预测的情感标签。

2.5　实验设置与结果分析

2.5.1　评价指标

本课题的图文多模态方面级情感分类问题是一个典型的分类任务，我们使用正确率 (Accuracy, Acc) 和宏平均 F1 值 (F1) 来评估性能。其中，F1 值是精确率 (Precision) 和召回率 (Recall) 的加权调和平均。在这些指标中，正确率既可以用于评估二分类系统的性能，也可以用于评估多分类系统的性能。而精确率、召回率和 F1 值则只能用于评估二分类系统的性能。本节将简要介绍以上评价指标在多分类系统中的定义，它们的计算公式如下：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2.7}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2.8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.9}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.10}$$

其中，TP表示该类被模型正确分类的样本数，FP表示其他类别被不正确地分类到该类别的样本数，FN表示该类被不正确地分类到其他类别的总样本数，TN表示其他类别被正确地分类到其他类别的总样本数。

2.5.2　实验参数设置

2.5.3　对比实验设置与结果分析

为了验证本章提出方法的有效性，本节在自建数据集 AECR-Twitter，Twitter-15 和 Twitter-17 数据集上，与以下几种相关基线模型进行了对比实验：

- ITM[9]：该方法提出了粗粒度与细粒度图像-aspect 对齐的多任务学习框架。由于 ITM 原本使用 Yu 等人提出的 aspect-image 相关性数据集来进行基于 aspect 的跨模态关系推理，我们在“ITM + Our AECR”中将其替换为我们的 AECR-Twitter 数据集，用于训练跨模态关系推理模块。
- HIMT[18]：该方法分层交互式多模态 Transformer 模型。在“HIMT + Our AECR”中，我们将跨模态关系推理模块集成到 HIMT 中，以获取图像与 aspect 之间的相关性权重。
- DPFN[19]：该方法双视角融合网络，用于同时建模特定 aspect 相关的全局与局部细粒度情感信息。在“DPFN + Our AECR”中，我们采用与“HIMT + Our AECR”相同的方式注入我们的相关性信息。
- A2II[21]：一种生成式模型，通过利用语言-视觉大模型 (LVLMs) 解决多模态融合的局限，并通过指令微调减轻无关图像噪声的影响。在“A2II + Our AECR”中，我们使用基于 AECR-Twitter 数据集微调后的 Qwen2-VL-7B 模型生成的关系感知知识，替换其原本由 InstructBlip-Flan-T5-xl[22] 模型生成的知识。
- Qwen2-VL[23]：一种广泛使用的 LVLM，表现出良好的性能。“Qwen2-VL + Our AECR”表示在指令中注入我们的关系感知知识。
- Qwen2.5-VL[24]：Qwen2-VL 的升级版本，具有更强的视觉-语言对齐能力与推理能力。“Qwen2.5-VL + Our AECR”沿用与“Qwen2-VL + Our AECR”相同的关系感知知识注入策略。
- AoM[2]：该方法面向 aspect 的方法，通过图卷积网络 (GCN) 在句法依存条件下融合语义与情感特征。“AoM + Our AECR”指采用与“A2II + Our AECR”相同的策略，将我们的关系感知知识注入 AoM 中。
- DQPSA[3]：该方法通过双查询提示实现视觉-文本对齐，并使用能量模型进行 span 边界配对。在“DQPSA + Our AECR”中，我们采用与“A2II + Our AECR”

相同的方式，将我们提出的关系感知知识注入 DQPSA。

表 2-3　对比实验结果

Methods	Twitter-15		Twitter-17		AECR-Twitter	
	ACC	F1	ACC	F1	ACC	F1
ITM	77.78	73.44	71.64	70.27	71.72	67.23
ITM+Our AECR	$78.36^{\dagger}_{\pm 0.20}$	$74.09^{\dagger}_{\pm 0.11}$	$72.93^{\dagger}_{\pm 0.91}$	$71.86^{\dagger}_{\pm 0.76}$	$73.31^{\dagger}_{\pm 0.39}$	$69.17^{\dagger}_{\pm 0.55}$
HIMT	76.01	71.46	67.77	65.34	68.31	62.05
HIMT+Our AECR	$76.37^{\diamond}_{\pm 0.43}$	$72.01^{\dagger}_{\pm 0.20}$	$68.82^{\dagger}_{\pm 0.12}$	$66.88^{\dagger}_{\pm 0.18}$	$69.94^{\dagger}_{\pm 1.18}$	$63.88^{\dagger}_{\pm 0.12}$
DPFN	76.75	71.72	70.06	68.68	66.29	65.96
DPFN+Our AECR	$76.53^{\diamond}_{\pm 0.31}$	$72.46^{\dagger}_{\pm 0.37}$	$71.12^{\dagger}_{\pm 0.40}$	$69.38^{\dagger}_{\pm 0.32}$	$71.06^{\dagger}_{\pm 1.20}$	$70.85^{\dagger}_{\pm 1.21}$
A2II	76.61	72.70	72.96	71.37	71.15	66.96
A2II+Our AECR	$77.77^{\dagger}_{\pm 0.07}$	$73.13^{\dagger}_{\pm 0.35}$	$74.11^{\dagger}_{\pm 0.38}$	$72.74^{\dagger}_{\pm 0.38}$	$72.19^{\dagger}_{\pm 0.48}$	$68.63^{\dagger}_{\pm 0.67}$
Qwen2-VL	77.91	73.41	72.69	71.07	71.91	67.12
Qwen2-VL+Our AECR	$79.36^{\dagger}_{\pm 0.72}$	$74.25^{\dagger}_{\pm 0.36}$	$73.39^{\dagger}_{\pm 0.12}$	$72.61^{\dagger}_{\pm 0.29}$	$73.68^{\dagger}_{\pm 1.13}$	$70.06^{\dagger}_{\pm 1.13}$
Qwen2.5-VL	77.62	73.73	74.31	73.95	72.75	68.30
Qwen2.5-VL+Our AECR	$78.73^{\dagger}_{\pm 0.07}$	$74.19^{\dagger}_{\pm 0.07}$	$\mathbf{74.55^{\dagger}_{\pm 0.11}}$	$\mathbf{74.40^{\dagger}_{\pm 0.36}}$	$73.87^{\dagger}_{\pm 0.28}$	$71.10^{\dagger}_{\pm 0.02}$
AoM	77.90	73.84	73.52	72.38	73.37	70.17
AoM+Our AECR	$78.52^{\dagger}_{\pm 0.23}$	$75.08^{\dagger}_{\pm 0.25}$	$74.02^{\dagger}_{\pm 0.09}$	$72.92^{\dagger}_{\pm 0.31}$	$\mathbf{74.15^{\dagger}_{\pm 0.24}}$	$\mathbf{72.05^{\dagger}_{\pm 0.40}}$
DQPSA	80.80	80.80	71.88	71.88	71.34	71.34
DQPSA+Our AECR	$\mathbf{81.65^{\dagger}_{\pm 0.20}}$	$\mathbf{81.65^{\dagger}_{\pm 0.20}}$	$72.89^{\dagger}_{\pm 0.29}$	$72.89^{\dagger}_{\pm 0.29}$	$72.28^{\dagger}_{\pm 0.42}$	$72.28^{\dagger}_{\pm 0.42}$

表 2-3展示了各模型在 Twitter-15、Twitter-17 以及 AECR-Twitter 数据集上的多模态方面级情感分类 (MASC) 主要实验结果。其中，ITM、HIMT、DPFN、A2II、AoM 和 DQPSA 的实验结果均由我们重新实现。符号 † 表示该模型与引入 Our AECR 后的对应模型之间的性能差异在显著性水平 $p<0.01$ 下具有统计显著性，而符号 ◊ 表示差异不显著 ($p>0.05$)。

实验结果表明，在引入我们提出的情感感知跨模态相关性后，所有模型均取得了稳定且一致的性能提升。尤其是在当前两种先进方法 AoM 和 DQPSA 上，同样观察到了显著的性能增益，充分验证了所提出相关性建模方法的有效性。我们将性能提升主要归因于跨模态关系推理模块与所提出相关性建模机制的协同作用：一方面，该关系推理模块能够有效帮助现有 MASC 方法抑制无关图像带来的干扰，从而提升情感分析性能；另一方面，我们提出的相关性不仅能够建模文本 aspect 与图像之间的显式对齐关系，还能够捕获二者之间隐式的情感关联，使得 aspect–image 相关性建模更加

全面。

此外，Qwen2.5-VL 与 DQPSA 相较于其他模型表现出更为优越的性能，这主要得益于语言–视觉大模型 (LVLMs) 所具备的强大多模态理解能力以及视觉-语言预训练任务的有效性。在此基础上，引入我们提出的跨模态相关性进一步提升了 Qwen2.5-VL 和 DQPSA 的性能，进一步验证了情感感知跨模态相关性在增强 MASC 性能方面的有效性。在 Twitter-15 数据集上，DQPSA 的性能优于 Qwen2.5-VL。我们认为，这一优势来源于其在预训练阶段采用了基于提示的图像理解机制，使模型能够更加精准地关注与给定 aspect 相关的图像区域，从而实现更准确的情感预测。相比之下，“Qwen2.5-VL + Our AECR”在 Twitter-17 数据集上取得了更优的结果，这主要归因于其更强的泛化能力。然而，在 AECR-Twitter 数据集上，Qwen2.5-VL 的表现仍略逊于 AoM 和 DQPSA，这可能是由于 Qwen2.5-VL 需要更多训练数据才能充分适配任务，而 AoM 与 DQPSA 在小规模数据集上更具优势。

总体而言，我们提出的情感感知跨模态相关性在所有模型架构上均带来了稳定一致的性能提升，充分体现了其在捕获隐式情感关联以及提升多模态方面级情感分析性能方面的鲁棒性与有效性。

表 2-4　消融实验结果

Methods	**AECR-Twitter**	
	ACC	F1
Complete Model	**73.31**	**69.17**
without “Irrelevant” type	63.34	57.69
without “Emotional expression source” type	72.46	68.44
without “Enhancing emotional expression” type	71.62	67.36
without “Semantically related but emotional irrelevant” type	71.20	67.05

2.5.4　消融实验分析

为了更好地理解每种关系类型的贡献，如表2-4所示，我们在 AECR-Twitter 数据集上进行了一项消融实验，每次省略一个跨模态关系类型，并对比实验结果，具体见表格。”Complete Model” 表示包含所有四种关系类型的 MASC 模型，而“without x type”则表示在执行 MASC 时忽略了关系类型 x。需要注意的是，在消融实验中，训练集的大小保持不变。当去除某个关系类型时，我们保留完整数据集用于训练 MASC

任务，同时在基于方面的跨模态关系推断任务中屏蔽该类型的监督信号。

表格2-4展示了消融实验的结果。我们首先观察到，去除任何一种关系类型都会导致相较于“Complete Model”的一致性性能下降，这表明所有关系类型都对 MASC 任务有积极贡献。特别地，去除“无关”类型会导致最显著的性能下降。我们将其归因于社交媒体中普遍存在的弱图像-文本对齐现象，排除这类样本会严重削弱模型过滤跨模态噪声的能力。此外，去除“增强情感表达”和“语义相关但情感无关”类型会导致显著下降。这突出了这两种类型的重要作用：前者通过视觉上的隐性情感线索强化了文本情感表达，而后者则帮助模型区分情感相关性与语义相关性，有效捕捉图像中的关键信息。去除“情感表达来源”类型对性能的影响较小。这可能是因为该类型在数据集中占比较低，导致模型未能充分学习该类型的模式。

表 2-5　将跨模态相关性加入到 LLMs 中在 Twitter-15 和 Twitter-17 上的结果

Methods	**Twitter-15**		**Twitter-17**	
	ACC	F1	ACC	F1
GPT-3.5-Turbo	48.72	49.72	51.95	48.31
GPT-3.5-Turbo+Our AECR	52.49	52.36	52.55	49.25
Qwen-Turbo	65.19	59.80	57.37	54.57
Qwen-Turbo+Our AECR	68.56	61.83	59.72	56.02
DeepSeek-V3	62.64	61.64	59.40	58.87
DeepSeek-V3+Our AECR	**66.12**	**63.67**	**62.53**	**61.29**

2.5.5　在 zero-shot 场景下将跨模态相关性加入到 LLMs 中的性能

我们进一步通过 zero-shot 场景下的提示，探讨了我们提出的方面级情感跨模态关系在增强各种 LLMs 的能力方面的有效性。图 2-5 显示了我们设计的提示模板。具体来说，任务定义概述了 MASC 任务，并解释了如何利用跨模态相关性进行 MASC。在提示中，我们提供了方面、情感选项和推文中的句子。图像描述是从 BLIP-Image 和 BLIP-VQA 中提取的，依据[25]，并包含图像标题、视觉实体描述和视觉情感描述。对于跨模态关系，我们在 AECR-Twitter 数据集上微调了 Qwen2-VL-7B 模型，以生成跨模态关系感知的知识。我们评估了多种开源和闭源 LLMs 的 MASC 性能，包括 DeepSeek-V3、GPT-3.5-Turbo 和 Qwen-Turbo。如表 2-5 所示，所有 LLMs 在 zero-shot 场景下融入跨模态关系感知知识后，均表现出一致的性能提升。这一观察表明，

LLM Prompt Template

Task Definition: Detect the sentiment polarity towards the aspect from the sentence. Strictly format output as [sentiment]. If cross-modal relation is relevant, integrate the information of the sentence and the image description to make predictions. if cross-modal relation is irrelevant, only use the sentence to make predictions.

Aspect: Trey Songz

Options: -positive -neutral -negative

Make sentiment prediction for the following sentence:
Sentence: Me listening to Trey Songz sing at BETAwards17.
Image Description : The image is about a dog smiling happily. The image implies happy sentiment. The entity of the dog is present in the image.
Cross-modal relation: relevant

图 2-5　Relevance-aware zero-shot prompt template.

我们的关系感知知识提高了 LLMs 在 MASC 任务中的可靠性。特别地，开源 LLMs DeepSeek-V3 和 Qwen-Turbo 在两个数据集上都优于闭源的 GPT-3.5-Turbo。

2.5.6　错误模式分析

为了更好地理解我们方法所取得的改进，我们分析了在[9] 框架下，不同对比模型在错误模式上的变化。一个错误模式 X, Y 表示黄金情感标签“X”被错误地预测为“Y”或反之。图 2-6 显示了 Twitter-17 数据集上的结果。我们观察到，在三种错误模式中，“Neu, Pos”和“Neu, Neg”的错误发生最为频繁，表明区分中性情感与积极或消极情感仍然是一个主要挑战。这些错误通常发生在图像中的情感线索模糊时，导致模型预测出更强的情感。通过引入我们的关系感知，模型能够更好地利用图像中的隐性情感线索，区分中性情感与其他两个情感标签。相比之下，“Neg, Pos”错误模式出现的频率最低，但通常涉及讽刺或冲突的多模态线索。例如，一条表达消极情感的推文可能伴随一张积极的图像，这可能会误导模型做出错误的预测。我们的方法通过对与方面无关的图像进行降权，减轻了这个问题，使模型能够更专注于讽刺性的文本线索。尽管我们的模型在处理这种类型的错误时比[9] 表现得更好，但仍略微不如基准模型。我们将在未来的工作中探讨合适的方法来解决这一类型的错误。

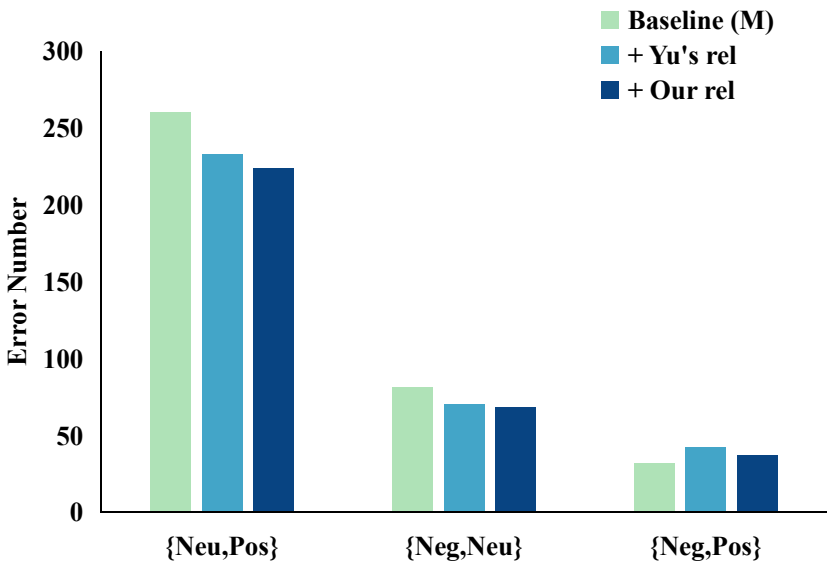


图 2-6　错误模式的统计分析。其中“Pos”、“Neu”和“Neg”分别代表正面、中性和负面

2.6　本章小结

本章定义了方面级情感相关性，并基于相关性的定义构建了一个方面级情感跨模态相关性数据集（AECR-Twitter），该数据集包含 3562 个样本。本章详细描述了 AECR-Twitter 的构建过程，涵盖了从标注规范到人工标注数据的全部流程，为确保数据集的高质量，本研究采用严格的独立双重标注。此外本章对 AECR-Twitter 数据集进行深入的分析，包括标注一致性分析、相关性标签分布、与前人定义相关性比较以及相关性标签与情感标签的关系分布。同时为验证所提跨模态关系方案和 AECR-Twitter 数据集在下游多模态方面级情感分类任务中的有效性，本章探索并对比了多种将跨模态相关性信息融入多模态方面级情感分类的方法，包括轻量级多模态模型的相关性插件方法和通过 prompt 注入相关性的大型语言模型方式，在 Twitter-15，Twitter-17 和 AECR-Twitter 数据集的实验结果表明加入我们的提出的方面级情感相关性能够给给现有方法提供一致的性能提升，验证了我们提出的方面级情感相关性的有效性。此外，本章将对方面级情感相关性进行了详细的分析，包括相关性标签类型的消融实验，错误模式的分析，深入探索了方面级情感图文相关关系的特点。

第三章　第三章 xx

现有的

3.1　引言

在第二章的工作中，我们定义了方面级情感相关性，基于相关性定义构建了情感相关性数据集。同时为验证提出的相关性的有效，我们提出了一种相关性感知的多模态方面级情感分类方法，通过相关性插件将相关性信息以图像权重比例的方式注入到模型之中。实验结果在多个模型和多个数据集得到了一致的提升表明了加入我们相关性的有效性。然而，现有方法在建模模态图文的情感线索时，主要通过全局的视觉特征或单一的文本语义 embedding 来代表模态情绪，但真实场景中的情绪线索往往分布在更细粒度的层面。例如，用户在文本中可能使用比喻、隐含情绪或未显式提及评价对象的表达方式；图像中的颜色、表情、场景构成等细节也可能蕴含间接的情绪信息。如果仅使用粗粒度的模态表示，模型难以准确捕捉这些细微却关键的情绪因素，从而导致情感识别偏差。另一方面。在上一章的工作主要通过计算图文是否相关的粗粒度权重系数来控制图像特征的使用，这种方式缺乏对细粒度情感相关性的深度挖掘。在复杂的实际场景中，图像中可能包含与文本情感无关的干扰背景，若仅进行粗粒度的权重调节，模型往往难以精确调控不同线索对最终情感决策的贡献，容易受到噪声干扰。面对上述问题，本章认为可以引入多模态大语言模型 (MLLM) 通过显式知识增强的方式提升特征质量。基于该思路，本章提出了一种基于 MLLM 知识增强与细粒度相关性控制的多模态方面级情感分类方法。该方法首先利用 MLLM 抽取图像与文本中的情感线索描述，使得隐含的模态情绪信息能够以自然语言形式显式表达，也通过文本化的方式弥合模态间的表征差异；其次，基于情感相关性构建差异化的输入指令，引导模型在关注有用线索的同时主动忽略无关信息的干扰；最后，设计了一种两路分支的后期融合机制，实现对文本与图像分支占比的灵活调节。总的来说，本章工作的主要贡献如下：- 显式情感知识增强机制：本章提出利用 MLLM 强大的推理能力，显式地从原始模态中抽取出深层的情感线索描述。通过将隐式的视觉特征转化为显式的文本语义线索，显著提升了情感特征的判别性与质量。- 基于相关性的精细化线索控制：本章设计了一套基于情感相关性的指令构建策略。通过在输入端引入不同类型的引导指令，使模型能够根据线索的贡献度进行动态调整，有效解

决了由于线索粒度过粗而导致的抗噪能力弱的问题。- 可控的多路融合架构：本章提出了一种两路分支的后期融合方式。根据情感相关性标签调节两条分支的贡献比例，从而有效抑制无关或冲突模态的影响，突出真正与 aspect 相关的模态情绪信号。

3.2　基于图文情感线索引导的多模态方面情感分类方法

本章提出了一种基于相关性控制与情感线索引导的多模态方面级情感分类方法(RC-ECG)，本章设计了一个包含离线数据增强与在线生成式分类的双阶段框架。具体来说，我们利用多模态大模型的强大能力来获得细粒度的方面级情感跨模态相关性，并从图像和文本中提取显式的、与方面相关的情感线索描述，为后续的分类任务提供参考信息。我们得到了包含情感线索描述和方面级情感跨模态相关性标签的扩充数据集。接着在相关性引导的情感线索训练阶段中，我们根据情感线索描述和相关性标签来动态构建文本分支与图像分支的提示输入，筛选出与目标方面相关的情感线索。最后，我们设计了一种两路分支的后期融合机制，根据情感相关性标签调节两条分支的贡献比例，从而有效抑制无关或冲突模态的影响，突出真正与目标方面相关的模态情绪信号。

3.2.1　方面级情感跨模态相关性与图文情感线索描述数据集构建

为满足本章提出的基于相关性控制与情感线索引导的多模态方面情感分类方法，我们为每个样本添加方面级情感跨模态相关性标签与情感线索描述。因此我们首先需要基于原有的图文对来构建情感线索描述和方面级情感跨模态相关性标签。具体来说，我们分别使用 Qwen2.5-VL-7B 模型和 Gemini 2.5 来进行数据集的扩充。整体的构建流程如图所示，接下来是数据构建的具体细节：

方面级情感跨模态相关性标签生成。对于数据集中的样本 x, 为了精准描述其中图像 x^i 与文本 x^t 中方面 x^a 之间的情感跨模态相关性，我们选用在视觉理解与推理任务上表现优异的 Qwen2.5-VL-7B 模型作为方面级情感跨模态相关性的分类器。具体来说，我们基于上一章构建的 AECR-Twitter 数据集，对 Qwen2.5-VL-7B 进行指令微调，然后在 Twitter-15 和 Twitter-17 数据集进行推理，使其能够根据图文内容输出三分类标签 $r \in \{0, 1, 2\}$。这些标签将作为后续生成式框架中引导情感线索组合使用的核心控制信号，用于筛选出与目标方面相关的情感线索。

- 无关 ($r = 0$)：图像内容与文本或目标方面存在矛盾或无关联，视为噪声。

- 语义相关 (r = 1)：图像在语义信息上与文本一致，但未包含明显的情感倾向。
- 情感相关 (r = 2)：图像表达或加强了针对目标方面的情感表达。

图文情感线索描述生成。为了弥补小参数模型在情感理解能力上的不足，我们利用 MLLM 的强大的推理能力提取显式的情感线索描述作为一种知识蒸馏的方式。我们通过构建 prompt 调用 MLLM 分别从图像 x^i 与文本 x^t 中提取方面 x^a 相关的情感线索描述。具体来说，对于图像情感线索描述 c^i，我们输入图片和目标方面词 x^a，要求模型输出与 x^a 相关的情感线索描述。而对于文本情感线索描述 c^t，我们输入文本 x^t 和目标方面词 x^a，要求模型输出与 x^a 相关的情感线索描述。具体的 prompt 如图所示。

3.2.2　相关性引导的生成式框架

在相关性引导的情感线索训练阶段中，我们根据上一节得到的情感线索描述和相关性标签来动态构建文本分支与图像分支的提示入喉输入到该生成式框架中。具体来说，该框架主要由相关性感知的提示构建、图文输入编码、以及两路分支融合三个模块组成。

3.2.3　相关性感知的提示构建

为了减轻无关情感线索的干扰，我们设计了一种方面级情感跨模态相关性感知的提示工程策略。我们构建了两个输入序列：文本分支提示, X_{txt} 和图像分支提示 X_{multi}。其构建逻辑严格受控于方面级情感跨模态相关性标签 r：文本分支提示 (X_{txt})，该分支专注于文本模态的情感线索描述。我们将输入文本 T、目标方面 x^a 以及提取的文本线索 C_{txt} 按照标准模板进行拼接：

$$X_{txt} = \text{“Aspect: } A\text{, Sentence: } T\text{, Clues: } C''_{txt} \tag{3.1}$$

图像分支提示，该分支旨在融入图像模态的情感线索描述，但通过相关性标签 r 进行控制是否引入。具体构建逻辑如下：

$$X_{multi} = \begin{cases} \mathcal{T}_{irrel}(T, A), & \text{if } r = 0 \\ \mathcal{T}_{rel}(T, A, C_{img}), & \text{if } r \in \{1, 2\} \end{cases} \tag{3.2}$$

具体而言，当 $r = 0$（无关）时，我们丢弃图像线索 C_{img}，并使用一条限制性指令 $\mathcal{T}_{irrel}$，强制模型忽略视觉输入。当 $r \geq 1$（相关）时，我们整合 C_{img} 并使用一条协同指令 $\mathcal{T}_{rel}$，结合图像情感线索 C_{img} 来引导模型关注与目标方面相关的情感线索。具体而言，$\mathcal{T}_{rel}$ 模板为：

$$\mathcal{T}_{rel}(T, A, C_{img}) = \text{“Aspect: } A\text{, Sentence: } T\text{, Clues: } C''_{img} \tag{3.3}$$

3.2.4　编码器输入

在本小节中，模型的输入层旨在将图文模态特征统一映射至预训练语言模型的语义空间。我们分别对视觉特征和文本提示进行处理，构建包含视觉前缀的序列化嵌入。具体来说，对于图像特征的编码，我们参考前人的工作利用强大的 MLLM 来获得图像中与方面相关的图像特征，首先对于图像 x^i，我们使用冻结的图像编码器 ViT-g/14 来提取图像特征 $\mathbf{h}^{img}$，然后我们创建一组可学习的查询嵌入 q，这些查询可以通过自注意力层与方面特征交互，也可以通过交叉注意力层与图像特征交互。正如 InstructBLIP 将图像特征和指令一起输入到 Q-Former 中，以使查询 Z 提取出指令所描述的任务信息更丰富的图像特征一样，我们也同时将图像特征 $\mathbf{h}^{img}$ 和方面 x^a 输入到 Q-Former 中，以获得与图像方面相关的细粒度特征。由于 Q-Former 已经过预先训练，可以通过查询提取包含语言信息的视觉表示，因此它可以有效地充当信息瓶颈，提供最有用的信息，同时过滤掉不相关的视觉信息。最后，我们得到最终的隐藏状态表示 h^f。值得注意的是，文本分支与图像分支共享该投影模块的参数，这有助于模型在统一的视觉语义空间下进行推理。对应的数学公式，vit 和 q-former 部分图像特征提取：

$$\mathbf{h}^{img} = \text{ViT-g/14}(\mathbf{x}^i) \tag{3.4}$$

方面特征提取：

$$\mathbf{h}^{a} = \text{Flan-T5}(\mathbf{x}^a) \tag{3.5}$$

Q-Former 编码：

$$\mathbf{h}^{f} = \text{Q-Former}(\mathbf{h}^{img}, \mathbf{h}^a) \tag{3.6}$$

接着基于前文构建的文本主导提示序列 X_{txt} 与多模态主导提示序列 X_{multi}，我们首先通过冻结的 T5 词嵌入层 Embed(·) 将其转换为离散的文本嵌入。随后，将生成的软视

觉前缀 $\mathbf{P}_{vis}$ 分别拼接至两个分支的文本嵌入序列前端，形成最终的编码器输入：

$$\begin{aligned}\mathbf{E}_{txt} &= [\mathbf{P}_{vis}; \mathrm{Embed}(X_{txt})] \\ \mathbf{E}_{multi} &= [\mathbf{P}_{vis}; \mathrm{Embed}(X_{multi})]\end{aligned} \tag{3.7}$$

其中 $[\cdot;\cdot]$ 表示在序列长度维度上的级联操作。通过这种前缀注入机制，双流分支在接收相同视觉信息的同时，受到不同指令与情感线索的引导，从而实现差异化的特征编码。

3.2.5 解码器输出和双流融合

构建好的输入嵌入 $\mathbf{E}_{txt}$ 与 $\mathbf{E}_{multi}$ 被并行输入至参数共享的 Flan-T5 生成式模型中。为了综合利用文本模态的稳定性与视觉模态的丰富性，我们在解码阶段采用逻辑层融合策略。在解码步骤 t，模型分别为两个分支计算当前时间步的隐状态，并映射至词表空间，得到预测分布：

$$\mathbf{z}_{txt}^{(t)} = \mathcal{F}_\theta(\mathbf{E}_{txt}, y_{<t}), \quad \mathbf{z}_{multi}^{(t)} = \mathcal{F}_\theta(\mathbf{E}_{multi}, y_{<t}) \tag{3.8}$$

其中 $\mathcal{F}_\theta$ 表示共享参数的模块，$y_{<t}$ 为 t 时刻之前生成的历史 Token 序列。同时我们通过引入一个可调节的平衡超参数 $\lambda \in [0, 1]$，对两个分支的 Logits 进行加权求和，得到最终的融合分布 $\mathbf{z}_{final}^{(t)}$：

$$\mathbf{z}_{final}^{(t)} = \lambda \cdot \mathbf{z}_{txt}^{(t)} + (1 - \lambda) \cdot \mathbf{z}_{multi}^{(t)} \tag{3.9}$$

该融合机制允许模型根据任务需求灵活调整对特定模态的依赖。在生成阶段，模型基于融合后的 Logits 采用贪婪解码策略选择概率最大的 Token 作为输出：

$$\hat{y}_t = \arg\max(\mathbf{z}_{final}^{(t)}) \tag{3.10}$$

3.2.6 损失函数

模型的训练采用端到端的监督学习方式。我们的目标是最小化预测序列与真实情感标签序列 $Y = \{y_1, y_2, \ldots, y_L\}$ 之间的差异。采用标准的交叉熵损失函数（Cross-Entropy Loss），基于融合后的 Logits 计算损失：

$$\mathcal{L} = -\frac{1}{L}\sum_{t=1}^{L}\log\left(\frac{\exp(\mathbf{z}_{final}^{(t)}[y_t])}{\sum_{v\in\mathcal{V}}\exp(\mathbf{z}_{final}^{(t)}[v])}\right) \tag{3.11}$$

其中，L 为目标序列长度，$\mathcal{V}$ 表示预训练模型的词表大小，$\mathbf{z}_{final}^{(t)}[y_t]$ 表示在时间步 t 真实标签对应的 Logit 值。通过最小化该损失函数，模型能够联合优化视觉投影层参数与解码策略，从而实现对多模态方面级情感的精准分类。

3.3　实验设置与结果分析

表 3-1　对比实验结果

Methods	**Twitter-15**		**Twitter-17**	
	ACC	F1	ACC	F1
ITM	78.36	74.09	72.93	71.86
HIMT	76.37	72.01	68.82	66.88
DPFN	76.53	72.46	71.12	69.38
Aom	77.9	73.84	73.52	72.38
Qwen2-VL-7B	**79.36**	74.25	73.39	72.61
Qwen2.5-VL-7B	77.62	73.73	74.31	73.95
Base Model	77.77	73.13	74.11	72.74
Promp Fuse Method	77.14	74.20	74.06	73.35
Our Method	77.91	**75.27**	**74.80**	**74.36**

3.3.1　对比实验设置与结果分析

如表 3-1 所示，我们将本文提出的方法与现有的主流基准模型进行了对比，实验结论分析如下：

(1) 相较于基准方法，本章提出的模型 (Our Method) 在两个数据集上均取得了极具竞争力的表现。特别是在 Twitter-17 数据集上，本章方法在 ACC 和 F1 值上均大幅超越了所有对比模型。在 Twitter-15 数据集上，尽管 Qwen2-VL-7B 在准确率上略有优势，但本章方法取得了最高的 F1 值。

(2) 通过与 Base Model 和 Prompt Fuse Method 的对比，进一步证实了细粒度相关性控制与情感线索引导机制的关键作用。具体来说，Base Model 仅使用二分类相关性来加权图像特征，缺乏明确的情感语义描述引导，导致模型难以充分挖掘图文对中的

深层情感交互，限制了性能上限。而 Prompt Fuse Method 简单地将图文情感线索拼接在输入端，这种粗粒度的融合方式不仅未能有效利用描述信息，反而可能引入语义噪声，导致模型无法聚焦于关键特征。相比之下，本章方法通过细粒度的相关性控制与显式的情感线索引导，有效解决了上述问题，显著提升了模型对 MASC 任务的适配能力。

(3) 虽然 Qwen 系列大模型拥有庞大的参数规模，并在 Twitter-15 的准确率上表现出色，但本章方法在 Twitter-17 的各项指标以及 Twitter-15 的 F1 值上均实现了反超。这表明，在面向 MASC 这类细粒度情感分析任务时，单纯依赖模型参数规模的堆叠并非最优解。通用大模型往往缺乏针对特定图文关系的精细化建模，而本文设计的相关性控制与情感引导机制，能够以更小的参数规模实现更精准的任务特征提取，展现了针对特定任务进行结构化设计的优越性。

表 3-2　消融实验结果

Model Name	Twitter-15		Twitter-17	
	ACC	F1	ACC	F1
Complete Model	77.91	**75.27**	74.79	**74.36**
w/o Relevance Control	76.66	71.82	71.31	70.62
w/o Emotional Clues	77.77	73.13	74.11	72.74
w/o Relevance and Emotional Clues	76.60	72.70	72.96	71.37
w/o Relevance and Text Emotional Clue	77.91	74.50	72.77	71.36
w/o Relevance and Image Emotional Clue	75.89	71.74	74.14	73.61

3.3.2　消融实验分析

为了验证本章提出的相关性控制模块与图文情感线索的有效性，我们在 Twitter-15 和 Twitter-17 数据集上设计了消融实验：

-w/o Relevance Control：保留图文情感线索，但移除相关性控制模块。此设置用于验证在缺乏相关性约束时，直接引入线索是否有效。

-w/o Emotional Clues：保留相关性控制模块，但移除所有情感线索。此设置用于验证情感线索是否为模型性能提供了重要线索。

-w/o Relevance and Emotional Clues：同时移除相关性控制和所有情感线索。

-w/o Relevance and Text/Image Emotional Clue：在移除相关性控制的基础上，分别仅保留图像情感线索或仅保留文本情感线索，用于探究不同模态线索对特定数据集的贡献差异。

实验结果如表3-2所示，可以得出以下结论：(1) 相较于完整模型 (Complete Model)，移除任意模块后的变体在 F1 指标上均出现了不同程度的下滑。这充分证明了细粒度相关性控制与多模态情感线索的结合能够显著增强模型对图文对的情感理解能力，对于提升方面级情感分类任务的判别准确性至关重要。

(2) 与移除情感线索引导相比，移除相关性控制模块的实验对模型性能的影响更大，如果仅引入情感线索而缺乏相关性控制，其性能甚至低于没有任何线索的基准模型。表明简单的图文线索堆叠会引入特征冲突或噪声。相关性控制模块在此处起到了关键作用，能够有效过滤无关信息的干扰，解决图文线索冲突问题，确保模型仅利用与目标方面强相关的情感描述。

(3) 在分别单独引入图像或文本线索的实验中，我们观察到了一个显著的现象：在 Twitter-15 数据集上，保留图像线索的性能显著优于保留文本线索。这是由于该数据集多包含单一方面目标，图像内容通常直接反映了核心情感，具有较高的信息增益。相反，在 Twitter-17 数据集上，保留文本线索的表现更佳。这是因为 Twitter-17 包含大量多方面的样本，复杂的图像背景往往成为噪声，难以对应到具体的特定方面，此时模型更依赖于文本描述中的精确语义线索。

表 3-3　相关性质量分析结果

Model Name	Twitter-15		Twitter-17	
	ACC	F1	ACC	F1
with Coarse-grained Relevance	77.04	73.76	**74.87**	73.98
with Fine-grained Relevance	**77.91**	**75.27**	74.80	**74.36**
Add 20% Relevance Noise	77.53	73.41	74.23	72.76
Add 50% Relevance Noise	76.47	73.22	72.77	72.20

3.3.3　相关性质量影响分析

为了深入探究相关性控制模块的质量对模型最终性能的具体影响，我们设计了针对相关性的对比实验。包含以下三种设置：

-粗粒度相关性 (Coarse-grained Relevance)：将相关性退化为二值控制，即仅区分图文是否相关，忽略了相关程度的强弱差异。

-细粒度相关性 (Fine-grained Relevance)：即本文提出的完整方法，将图文情感相关性分类为情感相关、语义相关和图文无关三种情况。

-相关性噪声干扰 (Relevance Noise)：为了验证模型对低质量相关性的敏感度，我们在细粒度相关性的基础上引入人工噪声。具体而言，我们随机选择 20% 和 50% 的测试样本，将其的相关性类型替换为随机类型，以此模拟相关性计算失效或对齐错误的情况。

实验结果如表 3-3 所示，可以得出以下结论：对比粗粒度与细粒度相关性的实验结果表明，精细化的相关性建模对于提升模型性能至关重要。在 F1 指标上，细粒度相关性设置在两个数据集中均取得了最优结果。我们认为二元相关性判定不足以捕捉复杂图文对中情感关联，而细粒度相关性能够识别情感相关的图文对保留对情感分类有贡献的关键模态线索。另外通过引入不同比例的噪声干扰，我们观察到模型性能呈现出明显的下降趋势。当引入 20% 的噪声时，模型在两个数据集上的 F1 值均出现了下滑；而当噪声比例增加至 50% 时，性能损失进一步扩大。说明了模型对相关性质量的依赖，进一步验证了本文所提出的相关性控制模块在保证图文特征对齐质量方面的核心作用。

3.3.4　情感线索质量影响分析

3.4　本章小结

第四章　第四章 xx

第三章提出的图文情感线索引导的多模态情感分析通过引入细粒度的方面级跨模态情感相关性标签和单模态情感线索描述，进一步提升了多模态方面情感分类的性能，然而对于方面级跨模态情感相关性的利用中，我们认为现有的相关性同时考虑了情感相关性和语义相关性，但是针对于图像特征的使用并没有细粒度进行控制区分，因此我们提出将方面级跨模态情感相关性进行解耦为两个角度情感相关性和语义相关性，来分别控制图像中的语义特征和图像中的情感特征的细粒度使用，进一步增强多模态方面级情感分类的性能。同时，我们提出了单模态情感监督增强的预热训练方式，来帮助模型能够更好地理解文本和图像本身的情感。

4.1　引言

第三章提出 RC-EmoCLue 方法通过引入细粒度的方面级跨模态情感相关性标签和单模态情感线索描述，有效的帮助模型理解图文之间的情感语义关联，进一步提升了多模态方面情感分类的性能。然而其对于方面级跨模态情感相关性的利用是将其分为三种类别：情感相关、语义相关、图文无关，我们认为提出方面级跨模态情感相关性同时考虑了情感角度和语义角度的图文相关关系，但是借助相关性对于图像特征的使用并从这两个角度进行细粒度地控制区分。本章考虑这个角度去进一步增强第三章的方法，通过挖掘更细分的方面级跨模态情感相关性，实现对于图像情感和语义特征的细粒度使用，从而提升模型对于图文内容的情感信息和语义信息的充分理解。基于该思路，我们将方面级跨模态情感相关性解耦为两个角度的相关性：情感相关性和语义相关性，来分别控制图像语义特征和图像情感特征的细粒度使用。对于图像情感特征和图像语义特征的提取，我们使用 Q-former 设置针对图像情感和图像语义的 prompt 来得到对应图像细粒度特征。同时，我们提出了两阶段的训练方式，首先是单模态情感监督增强的预热，通过将文本情感和图像情感作为多任务来帮助模型更好地理解文本和图像本身的情感。接着是多模态方面级情感分类和多角度相关性的多任务学习，同时训练两个任务，将多角度相关性输出的情感相关性和语义相关性作用于对应的图像特征，实现细粒度使用。

本章提出的多角度多任务框架在 twitter-15 和 twitter-17 数据集的实验结果表明

是有效的

4.2　xxxx 的多模态情感分析

4.3　实验设置与结果分析

4.3.1　对比实验设置与结果分析

4.3.2　主要实验结果

4.3.3　消融实验分析

4.3.4　样例分析

4.4　本章小结

第五章　总结与展望

参考文献

[1] Ling Y, Yu J, Xia R. Vision-language pre-training for multimodal aspect-based sentiment analysis [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022: 2149-2159.

[2] Zhou R, Guo W, Liu X, Yu S, Zhang Y, Yuan X. AoM: Detecting aspect-oriented information for multimodal aspect-based sentiment analysis [C]//Findings of the Association for Computational Linguistics. Association for Computational Linguistics, 2023: 8184-8196.

[3] Peng T, Li Z, Wang P, Zhang L, Zhao H. A novel energy based model mechanism for multi-modal aspect-based sentiment analysis [C]//Proceedings of the AAAI Conference on Artificial Intelligence: volume 38. 2024: 18869-18878.

[4] Huang W, Wang Y, Gan Z, Xi T, Xi J, Xu X, Liu T, Qi D, Liu W. An opinion leader mining method based on text contents and network features [J]. World Wide Web, 2025, 28(2): 21.

[5] Zhang X, Xie R, Lyu Y, Xin X, Ren P, Liang M, Zhang B, Kang Z, de Rijke M, Ren Z. Towards empathetic conversational recommender systems [C]//Proceedings of the 18th ACM Conference on Recommender Systems. 2024: 84-93.

[6] Luvembe A M, Li W, Li S, Liu F, Xu G. Dual emotion based fake news detection: A deep attention-weight update approach [J]. Information Processing & Management, 2023, 60(4): 103354.

[7] Ju X, Zhang D, Xiao R, Li J, Li S, Zhang M, Zhou G. Joint multi-modal aspect-sentiment analysis with auxiliary cross-modal relation detection [C]//Proceedings of the 2021 conference on empirical methods in natural language processing. 2021: 4395-4405.

[8] Vempala A, Preoţiuc-Pietro D. Categorizing and inferring the relationship between the text and image of twitter posts [C]//Proceedings of the 57th annual meeting of the Association for Computational Linguistics. 2019: 2830-2840.

[9] Yu J, Wang J, Xia R, Li J. Targeted multimodal sentiment classification based on coarse-

to-fine grained image-target matching [C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, 2022: 4482-4488.

[10] Chen T, Lu D, Kan M Y, Cui P. Understanding and classifying image tweets [C]// Proceedings of the 21st ACM international conference on Multimedia. 2013: 781-784.

[11] Xu N, Wang J, Tian Y, Zhang R, Mao W. Ananet: Association and alignment network for modeling implicit relevance in cross-modal correlation classification [J]. IEEE Transactions on Multimedia, 2022, 25: 7867-7880.

[12] Chen D, Su W, Wu P, Hua B. Joint multimodal sentiment analysis based on information relevance [J]. Information Processing & Management, 2023, 60(2): 103193.

[13] Yu J, Jiang J. Adapting bert for target-oriented multimodal sentiment classification [C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, 2019: 5408-5414.

[14] Liu Y. Roberta: A robustly optimized bert pretraining approach [J]. arXiv preprint arXiv:1907.11692, 2019.

[15] Devlin J, Chang M W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, 2019: 4171-4186.

[16] Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.

[17] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: Transformers for image recognition at scale [J]. arXiv preprint arXiv:2010.11929, 2020.

[18] Yu J, Chen K, Xia R. Hierarchical interactive multimodal transformer for aspect-based multimodal sentiment analysis [J]. IEEE Transactions on Affective Computing, 2023, 14(3): 1966-1978.

[19] Wang D, Tian C, Liang X, Zhao L, He L, Wang Q. Dual-perspective fusion network for aspect-based multimodal sentiment analysis [J]. IEEE Transactions on Multimedia, 2024, 26: 4028-4038.

[20] Tsai Y H H, Bai S, Liang P P, Kolter J Z, Morency L P, Salakhutdinov R. Multimodal transformer for unaligned multimodal language sequences [C]//Proceedings of the conference. Association for computational linguistics. Meeting: volume 2019. 2019: 6558.

[21] Feng J, Lin M, Shang L, Gao X. Autonomous aspect-image instruction a2ii: Q-former guided multimodal sentiment classification [C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation. 2024: 1996-2005.

[22] Dai W, Li J, LI D, Tiong A, Zhao J, Wang W, Li B, Fung P N, Hoi S. Instructblip: Towards general-purpose vision-language models with instruction tuning [C]//Advances in Neural Information Processing Systems: volume 36. Curran Associates, Inc., 2023: 49250-49267.

[23] Wang P, Bai S, Tan S, Wang S, Fan Z, Bai J, Chen K, Liu X, Wang J, Ge W, et al. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution [J]. arXiv preprint arXiv:2409.12191, 2024.

[24] Bai S, Chen K, Liu X, Wang J, Ge W, Song S, Dang K, Wang P, Wang S, Tang J, et al. Qwen2. 5-vl technical report [J]. arXiv preprint arXiv:2502.13923, 2025.

[25] Yang L, Wang Z, Li Z, Na J C, Yu J. An empirical study of multimodal entity-based sentiment analysis with chatgpt: Improving in-context learning via entity-aware contrastive learning [J]. Information Processing & Management, 2024, 61(4): 103724.

攻读学位期间的成果

· 论文

(1) **Yu Zhang**, Zhenghua Li, Min Zhang. 2020. Efficient Second-Order TreeCRF for Neural CRF Dependency Parsing. In Proceedings of ACL, pages 3295–3305, Online. (CCF-A 类会议)

(2) **Yu Zhang***, Houquan Zhou*, Min Zhang. 2020. Fast and Accurate Neural CRF Constituency Parsing. In Proceedings of IJCAI, pages 4046-4053, Online. (CCF-A 类会议)

(3) Houquan Zhou*, **Yu Zhang***, Zhenghua Li, Min Zhang. 2020. Is POS Tagging Necessary or Even Helpful for Neural Dependency Parsing?. In Proceedings of NLPCC. pages 179-191, Zhengzhou, China (CCF-C 类会议, ***Best Paper Award***)

(4) Wei Jiang, Zhenghua Li, **Yu Zhang**, Min Zhang. 2019. HLT@SUDA at SemEval 2019 Task 1: UCCA Graph Parsing as Constituent Tree Parsing. In Proceedings of SemEval, pages 11–15, Minneapolis, Minnesota, USA.

· 比赛

(1) 2020 语言与智能技术竞赛比赛，第六名.

(2) 2019 语义分析国际评测比赛，第一名.

· 实习

(1) 2020/8--2021/2. 杭州-阿里巴巴-达摩院.

致谢

养天地正气，法古今完人. 从本科到硕士，转眼间在美丽的苏州大学校园内度过了七年的求学时光. 在这段不算短的人生旅途中，无论是学识上还是生活阅历上我都成长良多.

首先，我要感谢我的导师李正华老师. 李老师永远以饱满的热情和专注的态度面对工作和生活，永远是我以后求学和工作的一个榜样.

感谢尊敬的张民老师，张老师以高标准要求每一个学生，营造了组内浓厚专一的科研氛围. 此外，张老师敏锐的思维、渊博的知识、平易近人的风格、深深的影响了我，平时的相处让我获益良多. 感谢陈文亮老师，陈老师开朗热情，在学业上给予了我很多指导. 同样感谢周国栋、朱巧明、李寿山、洪宇、段湘煜和李军辉等苏州大学自然语言处理实验室的所有老师，各位老师严谨的治学态度和进取的专业精神是我的榜样.

感谢周厚全师弟，厚全师弟涉猎广博，热爱阅读，富有好奇心，在平时的讨论中总是能给我很多启发. 在课题研究上我们有很多合作，也取得了很多成果，希望以后继续合作，互相促进.

感谢同组的夏庆荣师兄、龚晨师姐、李英师姐和张月师姐，各位师兄师姐总是十分热心的解决我生活和研究上遇到的困难. 感谢章波、黄德朋、江心舟师兄，彭雪师姐，在我还是萌新的时候对我的帮助，以及平时对我的关照. 感谢同届的蒋炜、陆凯华、吴锟、刘亚慧同学，大家在一起互相帮助，互相进步. 感谢沈嘉钰、李嘉诚、侯洋、李帅克、周仕林、刘泽洋、李扬师弟，还有周明月、杨浩萍师妹，十分珍惜与大家相处的美好时光.

此外，还要感谢实习期间相处的王涛师兄、蒋勇师兄，以及王新宇、胡泽川、蔡炯和马欣尹同学. 特别是感谢蒋勇师兄在我实习期间对我的关照，以及在课题研究上的悉心帮助和指导.

感谢我的父母还有家人们，你们总是我心灵上的港湾和寄托，无论何时都能给我最无私的帮助.

最后，我还要感谢各位评审老师，感谢各位老师们在百忙之中抽取时间对本文进行评审，并提出宝贵的修改意见.

学位论文答辩委员会决议

包括：1、对论文的评价，包括选题的理论价值和实践意义，论文理论、方法上的开拓与创新，论据的可靠充分与结论的正确性；论文所反映的作者学术视野（对本学科及相关领域研究动态的把握）、基础理论、专业知识、写作能力等；

2、对答辩的评价；

3、是否同意通过论文答辩，是否建议授予学位或是否建议在规定时间内修改论文后重新答辩一次的结论。

论文在依存句法分析和成分句法分析这两种句法分析任务上，探讨了基于树形条件随机场的高阶方法对句法分析器性能和效率的影响。选题具有很好的理论价值和实践意义，以及一定的创新性。 目前主流的句法分析方法大多基于神经网络方法，并采用了一个简化的学习目标，相对应地，传统方法中大多采用了结构化学习以及高阶建模。论文针对这一对比，提出了将当前的句法分析器与传统方法做一个联结。论文提出在神经网络模型中采用树形条件随机场来最大化树概率，并进一步提出采用高阶建模。为了改善带来的效率问题，论文分别尝试了批次化计算和变分推断近似方法来加速。结果表明结构化建模和高阶方法对于目前的句法分析器仍然是有益的。

作者比较全面地论述了相关研究领域国内外研究情况，所采用的研究方法和技术手段体现了作者良好的学术研究基础和能力。论文成果在研究方法等方面有所创新，其中的新方法、新思路具备了很好的应用价值。论文写作层次清晰，逻辑结构合理，文字流畅，符合学术规范。

论文质量优秀。答辩过程中陈述清楚，回答问题准确。

答辩委员会经讨论，认为该论文已达到硕士学位论文水平，一致同意其通过论文答辩，建议授予硕士学位。

答辩委员会主席：孔芳　　秘书：张雅静

委员：陈文亮、李培峰、钱龙华、朱晓旭

_____、_____、_____、_____

2021 年 5 月 22 日

注：本表内容（包括答辩名单）可手签或打印