

思维链推理的方面级情感分析模型

黄俊光¹, 缪裕青¹, 刘同来², 张万桢², 蔡国永¹

¹(桂林电子科技大学 计算机与信息安全学院, 广西 桂林 541004)

²(仲恺农业工程学院 信息科学与技术学院, 广州 510225)

E-mail: kystring@163.com

摘要: 现有的方面级情感分析方法, 大多是通过分析句子中意见项和句子语义信息得到方面项的情感极性, 较少考虑意见项出现的背后可能原因和句子本身所处的语境等与方面相关的潜在外部信息, 而这些潜在外部信息有助于方面级的情感分析。本文提出一种基于思维链推理的方面级情感分析模型 4-T-4-T 思维链模型在推理过程中, 由浅入深地不断给出提示, 使语言模型通过推理, 不断挖掘方面相关的潜在外部信息; 利用推理出的潜在外部信息作为辅助, 增强模型对方面情感的理解。同时, 为提高推理的逻辑正确性和优化训练效率, 设计思维一致性约束和特征裁剪。实验结果表明, 所提模型的效果优于多个对比模型, 进一步提高了方面级情感分析的准确率。

关键词: 方面级情感分析; 思维链; 思维一致性约束; 推理提示; 特征裁剪

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2026)02-0403-10

Aspect-level Sentiment Analysis Model Based on Chain-of-thought Reasoning

HUANG Junguang¹, MIAO Yuqing¹, LIU Tonglai², ZHANG Wanzen², CAI Guoyong¹

¹(School of Computer Science & Information Security, Guilin University of Electronic Technology, Guilin 541004, China)

²(College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China)

Abstract: Most existing aspect-level sentiment analysis methods focus on analyzing opinion targets and sentence semantics to determine sentiment polarity, with little consideration for the underlying reasons behind the appearance of opinion targets or the contextual information in which the sentence is situated. These latent external factors are useful for aspect-level sentiment analysis. This paper proposes a 4-T model for aspect-level sentiment analysis based on chain-of-thought reasoning. The 4-T reasoning model provides incremental hints during the reasoning process, enabling the language model to dig deeper into latent external information related to aspects. This external information is then used to enhance the model's understanding of aspect sentiment. To improve logical consistency and optimize training efficiency, we design reasoning consistency constraints and feature pruning. Experimental results show that the proposed model outperforms several baseline models, improving aspect-level sentiment analysis accuracy.

Keywords: aspect-level sentiment analysis; chains of thought; consistency constraints; reasoning prompts; feature pruning

0 引言

方面级情感分析 (Aspect-Based Sentiment Analysis, AB-SA) 是一种细粒度的情感分析任务, 用于识别给定句子中特定方面术语的情感极性^[1]。例如句子 “The cafeteria ambiance is nice, but the fish doesn't feel fresh”, 句中存在两个方面, “cafeteria ambiance” 和 “fish” 分别对应的情感倾向为积极和消极^[2]。ABSA 任务旨在识别这两个方面在句子中包含的情感: 积极、消极还是中性^[3]。

对于方面级情感分析的研究, 注意力机制的方法在 AB-SA 任务中已经取得了一些不错的成果^[4]。考虑到注意力机制的局限性, 一些学者通过构建语法树匹配方面项相关的意见信息^[5], 利用语法树的语法结构信息分析方面的情感。该类方法通常在句法依赖关系上使用图神经网络, 利用图神经

网络确定上下文方面相关特征。比如: Chen 等人^[6]提出了一种类型感知图卷积网络, 显式地利用依赖类型增强方面的依赖信息提取。Tian 等人^[7]引入了组合式范畴语法, 从句子的整体语义出发, 提取方面的语义信息。

不论是基于注意力机制的方法, 还是构建语法树的方法, 大都是都是从句子本身出发, 通过分析句子中意见项和句子语义信息得到方面项情感极性。已有的研究^[8]表明, 很多情况下方面的情感理解除了从句子本身出发之外, 还需要考虑句子的潜在外部信息, 如: 意见项出现的背后可能原因和句子本身所处的语境等。但是, 现有方法较少对意见项和句子本身之外潜在语义信息进行挖掘, 缺少对句子潜在外部信息的利用。或者简单地引入外部知识库作为辅助信息分析方面项情感, 容易导致外部信息和句子语义信息不匹配而损害模型性能。

收稿日期: 2025-01-02 收修改稿日期: 2025-03-03 基金项目: 广东省研究生教育创新计划项目 (2022JGXM115) 资助; 教育部产学研合作协同育人项目 (202102191071, 202102211021) 资助; 国家自然科学基金项目 (62366010, 62366011) 资助; 桂林电子科技大学研究生教育创新计划项目 (2024YCX068) 资助。作者简介: 黄俊光, 男, 1998 年生, 硕士研究生, 研究方向为方面级情感分析; 缪裕青, 女, 1966 年生, 博士, 教授, 研究方向为数据挖掘、大数据、情感分析; 刘同来 (通信作者), 男, 1982 年生, 博士, 副教授, 研究方向为数据挖掘、区块链和智能计算; 张万桢, 女, 1981 年生, 硕士, 讲师, 研究方向为数据挖掘; 蔡国永, 男, 1971 年生, 博士, 教授, 博士生导师, 研究方向为安全可信网络软件、社会计算。

为了解决上述问题,本文设计了一种用于方面级情感分析的思维链推理模型 4-T。通过思维链推理挖掘意见项背后的可能原因,同时从整体出发挖掘句子的语境信息;并在推理过程中依托语言模型大量的参数辅助知识的推理。此外,为减轻思维链推理过程中产生的错误推理而导致推理信息与目标方面项不匹配的干扰,设计了思维一致性约束方法,以提高推理过程的逻辑正确性。

思维链的推理过程类似于人类思考的过程,通过由浅入深的分析问题,能在推理的过程中挖掘一些潜在的信息,从而提高对问题的理解。例如:图 1 中的例子,语言模型对于输入的句子“The sauces used are also not that exciting”,能推断出对应于方面“sauces”的意见项为“not that exciting”;并进一步地推理出“not that exciting”的原因可能是“sauces might be bland, lacking in depth and complexity of flavors that make them stand out”并理解句子的语境为“customer review”,最终根据上述推理信息可得出方面“sauces”的情感极性为消极。

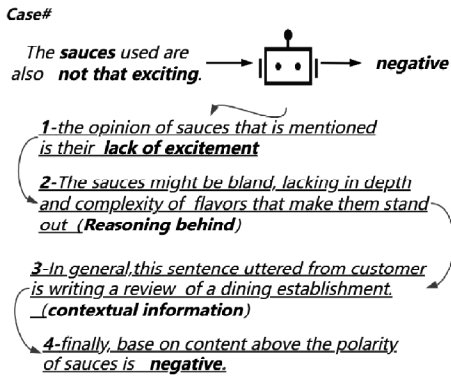


图 1 语言模型思维链推理

Fig. 1 Language model reasoning chain

通过这一例子可看出思维链的推理过程可帮助方面级情感分析。本文的主要工作和贡献:

- 设计了一种用于方面级情感分析的思维链推理模型 4-T; 4-T 模型由浅入深地挖掘句子的潜在外部信息;利用推理出的潜在外部信息作为辅助,增强模型对方面情感的理解。
- 设计了思维一致性约束方法和特征裁剪策略,前者使得推理出的外部信息易于与目标方面项匹配,以提高推理逻辑的正确性;后者使得每一步的特征尽量简洁以提升训练效率。
- 在 3 个公开数据集上进行了充分实验,相比已有一些模型取得了更好的性能,验证了所提模型的有效性。

1 相关工作

方面级情感分析 (ABSA) 在过去 10 年引起了广泛的关注。早期的研究侧重于使用简单的预训练模型结合下游特定任务进行情感分析。例如:Devlin 等人^[9]使用通用的预训练 BERT 模型和一个适应 ABSA 任务的分类层来进行情感分析。

一些学者在依赖树上应用 GNN 来挖掘语法信息。这些模型通常将句子中的单词视为图中的节点,并通过聚集相邻

节点的信息来学习节点表示。Tang 等人^[10]提出一种依赖图增强的双变换结构,并在图中融合了序列表示和图表示。黄俊等人^[11]使用语序图卷积网络结合语序图卷积网络学习上下文相关情感信息。巫浩盛等人^[12]提出距离与图卷积网络,用以提高卷积神经网络和循环神经网络处理图结构数据的能力。Wang 等人^[13]调整方面依赖树结构,并使用新型 GAT (Graph Attention Networks) 关系对依赖树进行建模。Tian 等人^[14]显式地利用具有类型感知的 GCNs (Graph Convolutional Networks) 的 ABSA 依赖类型,针对这两种依赖结构设计相应的解码器提取情感特征。

学者们注意到依赖树在 ABSA 任务中的局限性而使用改良的依赖树或者新型的语法树结构。如 Chen 等人^[15]提出了一种特定方面和语言无关的离散潜在树模型,作为依赖树的替代结构。Ma 等人^[16]提出新型的 ARM 语法树,通过路径聚合方法捕获方面相关的情感信息。

上述分析方法,大多只从句子本身进行分析,较少考虑句子中潜在的外部信息,潜在的外部信息有助于方面项的情感分析。随着大语言模型的兴起,思维链被广泛应用于大模型对复杂逻辑问题的推理。思维链 Chain-of-Thought (CoT) 是一系列推理步骤的有序集合^[17,18]。思维链通过要求模型在输出最终答案之前,显式输出中间的推理步骤。一个完整的思维链推理链条应该由指令 (Instruction), 逻辑依据 (Rationale), 示例 (Exemplars) 3 部分^[19,21]组成。

- 指令:用于描述问题并告知大模型输出的格式。
- 逻辑依据:推理过程的提示,可包含任何对推理有帮助的外部知识或者问题解决方法。
- 示例:大模型输入输出的一种格式。

按照是否包含示例,可以将思维链分为两类:单指令思维链 (Zero-Shot-CoT) 和多指令思维链 (Few-Shot-CoT)。单指令思维链:不包含示例仅仅添加一行指令“Let’s think step by step”;多指令思维链:包含多个指令和示例。

鉴于思维链在大模型中对复杂逻辑问题的强劲推理能力,结合方面级情感分析,将思维链用于情感分析推理方面项的情感极性成为可能。

2 思维链设计

情感分析思维链应遵循以下原则:

- 推理链条长度合适;
- 当前推理提示需要基于前置推理结果进行推理;
- 每个推理提示应只包含一个核心问题。

为了探究合适长度的思维链,在有限的实验环境中设置了 6 组不同长度的情感分析思维链。在数据集 Restaurants、MAMS 和 Laptops 使用思维链结合 flan-t5 (large) 语言模型进行推理情感。从图 2 中可看出随着思维链长度的增加,F1 值呈缓慢增长,而推理时间增长迅速。而在长度为 4 的情况下,能较好地兼顾推理时间和准确率。

此外,已有的研究^[21,22]表明,当前推理提示应当基于前置结果进行推理,如果没有利用前置结果有可能导致中间推理过程中有效信息的缺失。同时,思维链中的推理提示应只包含一个核心问题,否则容易产生推理幻觉(即生成看似合理

但实际上不准确的信息).

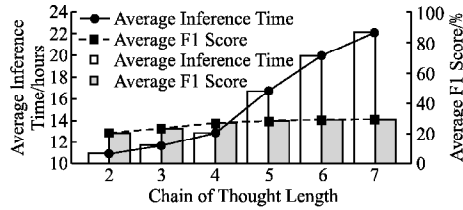


图2 不同思维链长度效率和时间对比

Fig. 2 Efficiency and time by thought chain length

根据 CoT-style 构建方法^[17,21,22],推理提示应分为两部分.第1部分为已知的条件或者上一步推理的结果,第2部分为当前轮次需要推理的核心问题.

Prompt: condition/result + problem?

在实际的推理过程中,模型需要推理外部信息.需要模型输出中间推理分析过程,为下一步推理挖掘有效信息.因此,4-T模型在第1和第4轮次采用了 CoT-style 方法构建推理提示,而在第2和第3轮次,则在 CoT-style 方法基础上加入了对推理原因的分析.如下面例子所示:

Prompt: condition/result + problem + and why?

3 基于思维链的情感分析模型

3.1 任务定义

在 ABSA 任务中,给定句子 $S = \{w_1, w_2, \dots, w_n\}$ 和一个特定的方面项 $A = \{a_1, a_2, \dots, a_m\}$,任务的目的是分析方面项在句子中的情感极性:积极、消极和中性,其中 w_i 表示子中的单词, a_i 表示组成方面项的单词.

3.2 模型概述

本文提出的方面级情感分析 4-T 思维链推理模型结构图如图3所示.模型主要由思维链推理、特征裁剪、思维一致性约束和思维链推理提示生成4部分组成.

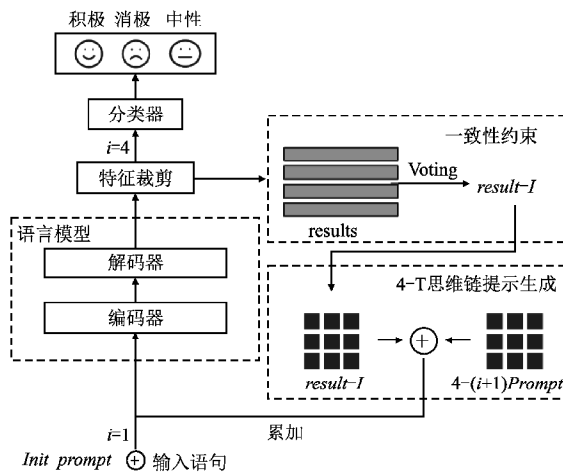


图3 4-T思维链推理模型结构图

Fig. 3 4-T chain-of-thought reasoning model structure diagram

句子在输入之前,会附加上初始提示,此时 $i=1$,经过语言模型中的编码器和解码器生成结果 results,然后这些结果经过特征裁剪之后再送入思维一致性约束,从候选的意见项

中选出一致性高的意见项 $result-i$.将该意见项输入到思维链推理提示生成模块中,在该模块中将该意见项和下一轮的推理提示拼接,得到的结果再作为下一轮的输入.

依次进行第2轮推理,直到 $i=4$,将第4轮的特征裁剪结果作为分类器的输入,通过分类器最终得到情感极性.

3.3 4-T思维链推理

4-T思维链的推理过程分为4轮,这4轮推理模拟人类对问题的一般思考推理方式,即问题描述的客观主体是什么?针对主体的限定条件是什么?根据限定条件能推理出的中间结论是什么?根据中间结论继续推理的最终结果是什么?通过这种拟人的思维推理模式,4-T思维链推理模型将在推理过程中推理意见项出现的背后可能原因和句子本身所处的语境.下面,将详细介绍4-T思维链推理中的每轮推理.

在初始输入的提示中,设计针对意见项的提示,在提示中加入方面项以确定分析的主体.通过初始提示,可使语言模型大致确定针对目标方面项的意见描述,得出意见项对方面项的基本情感倾向,提示形式如下:

1-T: $In [S], \text{what opinion is a description regarding aspect}[a]?$

其中, $[S]$ 表示输入模型的句子, $[a]$ 表示目标的方面项 a ,整个1-T可以用条件概率形式化地表示为 $O = \max P(o | S, a)$, O 表示模型推理出的意见项, o 表示句子中可能出现的所有意见项, a 为目标方面项, S 为输入句子.

第2轮推理中,对意见项在句子中出现的背后原因进行推理.在第2轮的推理开始之前,将第1轮推理产生的结果 $result-1$ 和1-T拼接,将该结果用以构造2-T,提示形式如下:

2-T: $shift([1-T] | [result-1]), \text{what are the possible reasons behind the emergence of this opinion, and why?}$

其中, $shift(M|m)$ 表示:将疑问句 M 在已知结果为 m 的情况下将句子 M 从疑问句转换为陈述句.2-T用条件概率可形式化地表示为 $R = \max P(r | S, a, O)$, R 表示模型推理出的意见项原因, r 表示所有可能的推理原因, S 为输入句子, a 为目标方面项, O 为1-T中推理的意见项.

第3轮推理中,对语境信息进行推理.将第2轮推理中产生的推理结果 $result-2$ 拼接上2-T,将该结果作为整体用以构造3-T.提示形式如下:

3-T: $shift([2-T] | [result-2]), \text{In general, what circumstances is this sentence uttered, and why?}$

3-T用条件概率可形式化地表示为 $C = \max P(c | S, a, O, R)$, C 表示模型推理出的语境, c 表示所有可能的语境, S 为输入句子, a 为目标方面项, O 表示1-T中推理的意见项, R 为2-T中推理的意见项出现原因.

第4轮推理中,对方面项的情感极性进行推理.将第3轮中产生的推理结果 $result-3$ 拼接上3-T,该结果加入到4-T的提示中,提示形式如下:

4-T: $shift([3-T] | [result-3]), \text{Based on the above analysis and opinion, what is the sentiment polarity of aspect}[A]?$

4-T用条件概率可以形式化地表示为 $Q = \max P(q | S, a, O, R, C)$,其中 Q 表示推理的方面项的情感极, q 表示所有

可能的情感极性(积极、消极和中性), S 为输入句子, a 为目标方面项, O 为 $1-T$ 中推理的意见项, R 为 $2-T$ 中推理的意见项出现原因, C 为 $3-T$ 中推理的语境信息。

以上的推理步骤, 每一轮推理都基于上一轮推理的结果进行推理, 这样每一轮推理都不断挖掘更加深层次的外部信息。这些信息将辅助模型更好地理解方面项情感。

3.4 特征裁剪

在多轮的推理中, 当前轮次的输入由上轮的结果拼接而成, 输入模型的 *token* 长度会随着轮次的迭代不断增加。这会导致模型的推理开销不断增加, 为了减小开销并提升推理效率设计了特征裁剪。一般情况下, 模型推理产生的结果有效大部分在第 1 段, 也就是第 1 个 $\langle \text{eos} \rangle$ 结束符之前的内容。所以本模块将保留第一个结束符 $\langle \text{eos} \rangle$ 前半部分的内容。对于后半部分的内容设计一个可学习的特征裁剪网络对特征进行裁剪。公式如下:

$$X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{d \times n} \quad (1)$$

$$z_i = \sigma(x_i) = \frac{1}{1 + e^{-f(x_i)}} \quad (2)$$

$$f(x_i) = W \cdot x_i + b \quad (3)$$

$$y'_i = z_i \cdot x_i \quad (4)$$

$$Y' = [y'_1, y'_2, \dots, y'_n] \in \mathbb{R}^{d \times n} \quad (5)$$

$$X' = \text{grid_sample}(Y', \text{range}) \quad (6)$$

$$X' = [x'_1, x'_2, \dots, x'_m] \in \mathbb{R}^{d \times m} \quad (7)$$

其中, X 为第 1 个 $\langle \text{eos} \rangle$ 结束符之后的结果矩阵, x_i 为每个 *token* 的 d 维向量表示, W 是可学习的权重矩阵, b 可学习的是偏置项, σ 是 Sigmoid 激活函数, x'_i 为裁剪之后的特征, $\text{range} \in [p, q] \in [0, 1]$ 为采样网格范围, p 和 q 为超参数, grid_sample 为双线性插值函数。剪裁过程中为每个 *token* 特征建立索引, 便于后续的相似度计算。同时, 希望最小化相似度的差异, 确保裁剪后的词向量与原始词向量尽可能相似; 因此, 使用余弦相似度计算损失。

$$L_{fp} = \sum_{k=1}^s \sum_{l=1}^4 \sum_{i=1}^m \left(1 - \frac{x_i \cdot x'_i}{\|x_i\| \|x'_i\|} \right) \quad (8)$$

其中, m 为裁剪之后 *token* 特征数量, 4 为推理轮次, s 为样本数量。

3.5 思维一致性约束

思维一致性方法的目的是提高每一步推理逻辑正确性从而使推理出的外部信息易于与目标方面项匹配。思维一致性方法通过采样多个不同的推理结果, 选择最一致的推理结果作为下一步的输入。

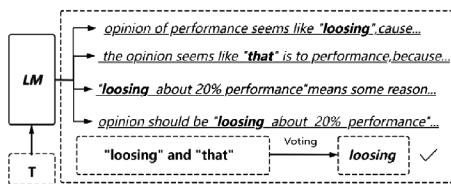


图 4 一致性约束

Fig. 4 Structure diagram of consistency constraint

对于每一轮推理, 语言模型解码器生成多个不同的结果, 从多个结果中, 采用投票机制选取一致性高的答案作为下一步的输入。实际情况下, 对于出现次数多的答案, 往往表示着

模型的推理较多偏向于这种结果, 意味着该结果一致性高; 而出现次数少的结果, 往往预示着模型的推理较少偏向这种结果, 其一致性就低。

以图 4 为例, 给定的句子 $T = \text{System is loosing about 20\% of performance because of that}$, 语言模型 (LM) 生成了多个推理, 其中有 3 个包含指向 *loosing* 的答案, 则通过投票选择出占据多数推理答案的“*loosing*”作为结果。

3.6 推理提示生成

推理提示生成模块的输入为一致性约束模块的输出结果。先将一致性约束模块的输出结果与本轮的推理输入 $i-T$ 做语句转换, 再拼接下一轮的推理提示; 拼接结果作为下一轮推理中语言模型的输入。相关计算公式如下:

$$(i+1) - T = \text{Gen}((P|Q), R)$$

$$P = \text{shift}[i - T]$$

$$Q = [\text{result} - i]$$

$$R = 4 - (i+1) \text{Prompt} \quad (9)$$

其中, $\text{result} - i$ 表示当前轮次推理出的结果, $i - T$ 表示本轮推理的推理输入, $4 - (i+1) \text{Prompt}$ 表示下一轮的推理提示, $\text{shift}(M | m)$ 表示: 将疑问句 M 在已知结果为 m 的情况下将句子 M 从疑问句转换为陈述句, $\text{Gen}(a, b)$ 表示 $a \oplus b$ 。

3.7 情感分析和损失计算

最后一轮推理, 将特征裁剪结果作为分类器的输入, 通过分类器最终得到情感极性。损失函数由交叉熵损失函数和裁剪损失函数组成, 计算公式如下:

$$\text{loss} = \text{cross_entropy}(\text{pre}, \text{gold}) + L_p \quad (10)$$

其中, pre 为预测值, gold 为标签值。

4 实验设计与结果分析

4.1 实验数据集

本文在公开发布的数据集 Restaurants^[23]、MAMS^[24] 和 Laptops^[25] 上进行实验。以上 3 个数据集标注了句子中的方面项和情感极性: 积极、消极和中性。

表 1 实验数据集划分

| Table 1 Division of experimental datasets | | | |
|---|-------------|---------|-------|
| 数据 | Restaurants | Laptops | MAMS |
| 训练 | 2375 | 1447 | 7680 |
| 验证 | 150 | 150 | 150 |
| 测试 | 1083 | 685 | 3356 |
| 合计 | 3608 | 2282 | 11186 |

上述 3 个数据集经过筛选和预处理后, 按照训练集、验证集和测试集进行划分, 统计如表 1 所示。

4.2 实验设置及评价指标

通过深度学习框架 PyTorch 搭建模型, 使用 GPU (NVIDIA GeForce RTX 4090) 加速。经过预处理后, 每个句子加上对应的初始提示 $1-T$ 。预处理好的数据输入 $4-T$ 思维链推理模型经过 4 轮推理得到情感分析结果。

实验的最优结果涉及多个参数, 具体设置如表 2 所示。保持表 2 中超参数的设置进行实验。

3 个数据集, 均使用准确率 (Accuracy) 和宏平均 F1 值 (macro average F1, MF1) 作为评价指标。

$$Acc = \frac{\sum_i^C TP_i}{\sum_i^C (TP_i + FN_i)} \quad (11)$$

$$MF1 = \frac{1}{C} \sum_i^C F1_i \quad (12)$$

$$F1_i = 2 \times \frac{P_i \times R_i}{P_i + R_i} \quad (13)$$

$$P_i = \frac{TP_i}{TP_i + FP_i} \quad (14)$$

$$R_i = \frac{TP_i}{TP_i + FN_i} \quad (15)$$

表2 模型参数设置

Table 2 Setting of model parameters

| 超参数 | Restaurants | Laptops | MAMS |
|------------|-------------|-----------|-----------|
| 词嵌入维度 | 100 | 100 | 100 |
| 批处理大小 | 32 | 16 | 8 |
| 学习率 | 1e-4 | 1e-4 | 1e-4 |
| 投影维度 | 300 | 300 | 300 |
| 梯度累加步 | 1 | 2 | 1 |
| 采样网格范围 p,q | 0.21,0.85 | 0.22,0.85 | 0.24,0.87 |
| 梯度剪裁 | 2 | 1 | 1 |
| Patience | 100 | 100 | 70 |
| seed | 42 | 42 | 42 |
| max_length | 400 | 300 | 600 |

其中,情感类别的数量 $C=3$, TP_i (true positive) 表示正确预测当前类别的样本数量; FP_i (false positive) 表示将其他类别错误预测为当前类别的样本数量; FN_i (false negative) 表示将当前类别错误预测为其他类别的样本数量。

4.3 对比模型

为验证所提模型的有效性,与现有一些模型进行比较,说明如下:

BERT^[9]:使用通用的预训练 BERT 模型和一个适应 ABSA 任务的分类层作为 ABSA 任务的核心模型。

DGEDT^[10]:基于依赖图增强的双变换结构,融合了序列表示和图表示。

R-GAT^[13]:使用新型 GAT (Graph Attention Networks) 关系对依赖树进行建模。

T-GCN^[14]:显式地利用具有类型感知的 GCNs 的 ABSA 依赖类型,通过这两种依赖类型构造依赖树,再针对这两种依赖结构设计相应的解码器。

dotGCN^[15]:使用一种基于特定方面和语言无关的离散潜在树作为依赖树的替代结构。

APARN^[16]:使用一种新的 ARM 语法结构来构建语法树,并通过路径聚合网络实现 ARM 语法信息的提取。

TextTG^[26]:使用双视图 Transformer,并在此基础上设计了一种新算法来在图中实现边特征。

其中,DGEDT、R-GAT、T-GCN、dotGCN、TextTG 基本架构都是图结构和预训练模型结合的。

4.4 实验结果及分析

4.4.1 模型有效性实验

为了和现有模型进行公平比较,在同一实验环境下复现了已有的一些模型,实验结果如表 3 所示。

BERT 的 MLM 预训练目标涉及随机掩盖输入序列中的部分词汇,并要求模型基于上下文同时预测这些被掩盖的词。这种双向上下文建模虽然增强了对局部语义的理解,但在生成任务中却导致了未来信息泄露的问题,即模型在预测某个位置的词时能够访问该位置之后的信息,这违背了自回归生成过程的本质要求。此外,BERT 的非自回归特性意味着它一次性预测多个位置的词汇,而非逐词生成,这与生成任务所需的顺序解码不兼容。因此,要使纯 BERT 实现思维链推理,需要设计额外的解码器形成编-解码器的结构,适配下游的推理任务。所以在实验中直接使用具有编-解码器结构的 flan-t5 语言预训练模型作为核心的语言模型。

表3 模型对比实验结果

Table 3 Experimental results of model comparison

| 模型 | Restaurants | | Laptops | | Accuracy | |
|--------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | Accuracy | Macro -F1 | Accuracy | Macro -F1 | Accuracy | Macro -F1 |
| BERT | 85.02 | 77.97 | 77.54 | 72.38 | 80.24 | 81.66 |
| DGEDT | 86.30 | 80.00 | 79.80 | 75.60 | 83.01 | 80.54 |
| R-GAT | 86.61 | 81.35 | 77.93 | 73.88 | 81.63 | 80.23 |
| T-GCN | 86.16 | 79.95 | 81.03 | 77.13 | 82.36 | 81.57 |
| dotGCN | 86.16 | 80.49 | 81.03 | 78.10 | 84.95 | 84.44 |
| APARN | 87.32 | 82.30 | 80.96 | 79.10 | 84.89 | 85.21 |
| TextTG | 84.15 | 81.64 | 81.59 | 78.84 | 86.65 | 86.43 |
| T5 (base) + 4 - T | 86.13 | 79.84 | 80.29 | 74.81 | 84.38 | 84.36 |
| T5 (large) + 4 - T | 88.13 | 82.32 | 81.92 | 79.25 | 85.38 | 86.54 |

表中有两组加粗的数据 T5 (base) + 4 - T 和 T5 (large) + 4 - T, T5 (base) + 4 - T 是使用了 T5 - base 作为对应的语言模型,其包含的参数是 250M 的版本, T5 (large) + 4 - T 使用的语言模型是 T5 - large,对应的参数数量为 780M。其中, T5 - large 取得的提升最显著,在 Restaurants 数据集上的 F1 值 T5 - large 比 T5 - base 提升了 2.48%; 在 Laptops 上 F1 值 T5 - large 比 T5 - base 提升了 4.44%; 在 MAMS 上 F1 值 T5 - large 比 T5 - base 提升了 2.18%。这说明思维链对模型的参数大小有着较强的依赖性,语言模型的参数量越大越能发挥 4 - T 思维链的能力。不同语言模型参数分析将在 4.4.5 中进行详细讨论。

T5 - large + 4 - T 与其他模型相比, T5 - large + 4 - T 在 3 个数据集上的各项指标均取得显著提升,说明挖掘出的外部信息对情感分析起到正向作用,使得模型理解方面情感更加全面。 T5 - large + 4 - T 与 DGEDT 相比,所提模型在 3 个数据集上性能表现更优,说明思维一致性使得推理出的外部信息易于与目标方面项匹配,使模型对方面的情感理解更加准确。

与近期方法相比,所提模型的性能也具有竞争力。与 dotGCN 相比,所提模型在 3 个数据集上均取得较好的结果,但在 Restaurants 和 MAMS 上的大部分指标性能更好,一部分原因可能是这两个数据集的规模比 Laptops 更大,语言模型本身的模型损失在大量的数据上受到样本分布的波动更小。

与 APARN 相比,所提模型在 3 个数据集上均取得更好的性能。 APARN 模型引入了 ARM 语法树结构,对句子整体的语义信息能充分利用。而 4 - T 思维链模型考虑了意见项的原因和句子外部的一些语境信息。这些信息能帮助模型从

句子本身之外理解方面情感,让模型能更好分析方面情感。

通过上述分析可看到,与已有一些模型相比,所提模型在3个数据集的大部分指标上均取得更优的性能,验证了所提模型的有效性。

4.4.2 4-T思维链模型消融实验

为了验证4-T各个推理提示以及思维一致性约束的有效性,分别在Restaurants、MAMS和Laptops3个数据集上进行消融实验,实验说明如下:

T5(large)+4-T(1-T):移除4-T中的第1个思维链提示,即移除关于意见项的推理提示。

T5(large)+4-T(2-T):移除4-T中的第2个思维链提示,即移除意见项出现背后原因的推理提示。

T5(large)+4-T(3-T):移除4-T中的第3个思维链提示,即移除句子整体语境的推理提示。

T5(large)+4-T(CC):移除4-T中的思维一致性约束,只留下思维链推理。

T5(large)+4-T:保持原有所有模块。

消融实验采用Accuracy和Macro-F1作为评价指标,如表4所示。由表4可知,移除或替换所提模型的一部分均会降低模型的性能。

表4 模型消融实验结果

Table 4 Experimental results of model ablation

| 序号 | 模型 | Restaurants | | Laptops | | Accuracy | |
|----|--------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | Accuracy | Macro-F1 | Accuracy | Macro-F1 | Accuracy | Macro-F1 |
| 1 | T5(large)+4-T(1-T) | 84.39 | 76.32 | 79.68 | 74.31 | 83.21 | 82.11 |
| 2 | T5(large)+4-T(2-T) | 87.10 | 78.88 | 81.60 | 75.93 | 83.12 | 83.65 |
| 3 | T5(large)+4-T(3-T) | 86.09 | 78.93 | 81.64 | 76.51 | 84.55 | 83.77 |
| 4 | T5(large)+4-T(CC) | 86.43 | 79.33 | 80.21 | 74.79 | 84.54 | 83.88 |
| 5 | T5(large)+4-T | 88.13 | 82.32 | 81.92 | 79.25 | 85.38 | 86.54 |

对比实验1和实验5,F1值在Restaurants、Laptop和MAMS的准确率分别降低了6.0%、4.9%和4.4%;对比实验2和实验5,F1值在Restaurants、Laptop和MAMS数据集上的准确率分别降低了3.4%、3.3%和2.8%;说明去除意见项的推理提示比去除意见项背后原因推理提示对模型性能的伤害更大。这说明意见项对推理情感的过程影响是显著的。实验分析发现,意见项大多集中在短语或句子的特定部分,表现出较强的局部性特征。这种集中分布使得T5中的encoder能够高效地捕捉这些信息,并通过多头注意力机制(Multi-head Attention Mechanism)进一步强化对关键信息的关注。另一方面,在解码过程中,T5根据编码器生成的特征表示来预测输出。意见项的存在提供了直接的情感线索。这为后续的推理产生直接影响。

对比实验3和实验5,F1值在3个数据集上分别下降了3.3%、2.7%、2.7%。相比实验2和实验5,平均F1值下降要小,这说明对于意见项背后原因的推理比句子语境信息的推理更加重要。分析发现,通过推理出意见项背后的原因,实际上是让模型构建了一个更深层次的理解框架。意见项背后的

原因提供了明确的因果关系,因果关系属于高阶特征,同时,这种高阶特征本身具有强情感指示性和深层次的理解。递进式的推理能很好地保留这种高阶特征到下一轮推理。由于高阶特征的强情感指示性,在最后的推理中吸引编码器分配更多的注意力权重,使其成为模型学习的重点。例如,在面对类似“虽然环境很好,但整体感觉不佳”这样的模糊表达时,模型可以根据已知的因果关系(如服务质量问题)进行推理,从而得出更为可靠的结论。

相比之下,句子语境信息涵盖了更广泛的内容,包括但不限于主语、谓语、宾语等语法成分,以及上下文中的其它信息。这些特征特点是比较分散和模糊,在推理中往往作为辅助信息使用。

对比实验4、实验5,F1值下降了2.9%、4.4%、2.6%,可看到移除思维一致性约束性能明显下降,而下降的原因可能是推理过程中产生的错误推理导致了情感不匹配而损害性能。除此之外,在实验过程中发现,尽管思维一致性约束在推理过程中对于大部分情况都适用;但存在一些少数高置信情况可能产生错误推理。

为探究少数高置信情况对一致性约束的影响,对实验数据集进行深入分析发现,影响模型性能的3个主要情况是:隐式的意见项(上下文中没有意见项)、低频意见项、以及反讽。对于隐式的情况,模型在1-T推理时可通过直接推理生成意见项。这种情况下通过上下文的推理分析,2-T根据1-T推理的意见项进行推理得出的多个结果,高置信的结果占据多数,这是因为1-T根据上下文推理的意见项大部分是常见意见项,使得思维一致性约束大部分情况下仍是有效的。对于低频意见项,在训练初期会导致特征学习不足,出现少数高置信推理无法被正确选择情况,随着训练的进行,这种错误会不断收敛。对于反讽的情况,特殊语义特征使得一致性推理容易产生少数高置信推理无法被正确选择的情况。为了量化这3种情况对思维一致性约束的影响,分别对3个数据集进行特殊情况统计。其中,隐式可直接统计,对于低频情况(专业领域专业知识等)使用维基百科特定领域的知识数据集DBpedia^[27]作为知识库进行检测统计,对于反讽使用反讽检测模型VADER^[28]进行检测统计。结果如图5所示。

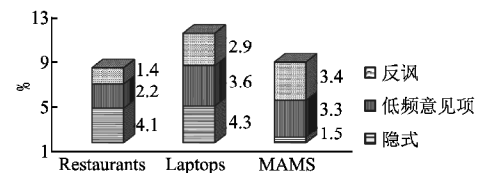


图5 少数高置信情况统计

Fig. 5 Statistics of minority high-confidence cases

发现3个数据集上3种特殊情况所占整个数据集的比例平均在8%左右。实验过程中发现这3种特殊的情况出现得并不多,思维一致性约束在大多数的情况下仍是有效的。

对比实验1~实验3,实验1的性能下降最明显,其次是实验2,最后是实验3;这说明了在推理过程中意见项对于模型推理方面的情感影响显著性最高,其次是意见项背后可能原因,最后是句子语境。

除此之外,在表4中对于Restaurants、Laptops和MAMS3

个数据集, T5 (large) + 4 - T 模型的性能表现也各不相同. 具体来说, 在 Restaurants 和 MAMS 数据集上模型的性能表现比 Laptops 数据集要好, 这可能因为 3 个数据集不同的数据规模量导致的. 从表 1 上可看出 Restaurants 和 MAMS 数据集的数据量比 Laptops 数据集的数据量更大, 而 4 - T 思维链推理模型在推理时需要学习大量的知识, Restaurants 和 MAMS 数据集数据量大意味着模型能够学习到更多的外部知识, 更有利于方面情感的理解.

通过上述分析可看到, 所提模型各部分均是有效的, 并且推理逻辑由浅入深.

其中第 1 轮关于意见项描述的提示对模型的影响最大, 其次是第 2 轮、最后第 3 轮; 各个部分对方面项的情感分析都起着积极的作用.

4.4.3 特征裁剪对模型作用分析

为探究特征裁剪对模型训练效率的影响, 在 T5-large、T5-base、T5-small 3 个语言模型上进行特征裁剪的对比实验, 结果如图 6 所示.

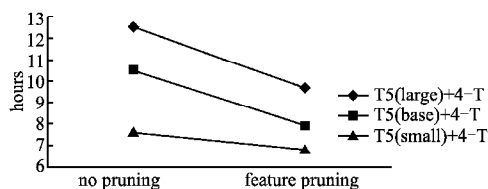


图 6 特征裁剪对比实验

Fig. 6 Feature pruning comparison experiment

从图 6 中可看出 T5-large、T5-base、T5-small 3 个语言模型进行特征裁剪之后训练时间明显降低. 说明特征裁剪能有效提升训练的效率.

为探究特征裁剪之后对信息丢失影响, 使用重建误差来衡量句子特征裁剪后的信息丢失. 使用采样网格范围 p 、 q 的最优参数, 得出 3 个数据集上每轮推理的平均信息丢失情况. 结果如图 7 所示.

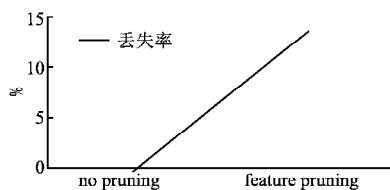


图 7 特征裁剪平均丢失率统计

Fig. 7 Feature pruning average dropout rate statistics

深入分析发现, 每轮裁剪后的信息丢失率维持在较低水平, 主要归因于所采用的可学习特征网络的独特设计. 该网络能够动态调整裁剪权重, 为与推理结果紧密相关的特征分配较低的裁剪权重, 使得大部分与核心推理相关的信息在每次推理过程中得以完整保留. 对于与核心问题关联度较低的虚词、连接词等一般性语言元素, 则分配较高的裁剪权重, 从而有效减少冗余信息, 提高计算效率和模型专注度.

但是, 不得不承认长文本通常具有复杂的结构和逻辑关系, 裁剪操作可能破坏这种连贯性, 使得剩余部分难以理解或解释. 但是实际上大部分情况下被裁剪掉的信息与核心的推

理内容的相关度不高, 通过舍弃这些信息而提高效率仍然是可接受的.

4.4.4 单指令思维链和多指令思维链对比实验

为了探究单指令思维链和多指令思维链 4 - T 对语言模型的不同影响, 对 4 - T 思维链推理和单指令思维链推理做了对比实验, 结果如表 5 所示.

实验说明如下:

T5 (base) + Prompt; T5 (base) + 单指令思维链.

T5 (large) + Prompt; T5 (large) + 单指令思维链.

GPT3 + Prompt; GPT3 + 单指令思维链.

T5 (base) + 4 - T; T5 (base) + 4 - T 思维链.

T5 (large) + 4 - T; T5 (large) + 4 - T 思维链.

GPT3 + 4 - T; GPT3 + 4 - T 思维链.

其中单指令思维链采用的方式是 Zero-Shot-CoT. 即在指令中仅仅加入“Let's think step by step”作为提示. 对于 GPT3 在 3 个数据集上各选取了 100 条数据进行实验. 实验采用 Accuracy、Macro-F1 作为评价指标, 实验结果如表 5 所示.

由表 5 可知, 单指令思维链整体比 4 - T 思维链 F1 准确率更低, 对比实验 1、实验 4 和实验 2、实验 5, 单指令 F1 准确率比 4 - T 思维链 F1 准确率在 T5 (base) 上平均低 5.3% 左右, 在 T5 (large) 上平均低 3.0% 左右, 说明 4 - T 思维链的推理过程比单思维链的推理方法能更加有效地引导模型挖掘句子的潜在外部信息, 从而帮助模型理解方面情感.

表 5 单指令思维链和多指令思维链对比实验

Table 5 Comparative analysis experiment between thought chains and single instruction prompts

| 序号 | 模型 | Restaurants | | Laptops | | Accuracy | |
|----|-----------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | Accuracy | Macro-F1 | Accuracy | Macro-F1 | Accuracy | Macro-F1 |
| 1 | T5 (base) + Prompt | 81.93 | 73.21 | 70.89 | 69.44 | 83.11 | 80.31 |
| 2 | T5 (large) + Prompt | 86.46 | 79.11 | 79.58 | 74.41 | 84.08 | 83.76 |
| 3 | GPT3 + Prompt (Unsup) | 64.13 | 57.29 | 50.77 | 44.01 | 66.31 | 59.97 |
| 4 | T5 (base) + 4 - T | 86.13 | 79.84 | 80.29 | 74.81 | 84.38 | 84.36 |
| 5 | T5 (large) + 4 - T | 88.13 | 82.32 | 81.92 | 79.25 | 85.38 | 86.54 |
| 6 | GPT3 + 4 - T (Unsup) | 68.83 | 64.71 | 54.01 | 45.85 | 67.58 | 61.98 |

一方面, 单指令思维链仅仅只在指令中加入“Let's think step by step”作为提示, 这可能导致后续的推理产生错误的结果而损害性能; 另一方面, 单指令没有对外部信息进行匹配, 而导致外部信息没有对情感的推理起到正向作用.

对比实验 1 ~ 实验 3 和实验 4 ~ 实验 6 可以看出, GPT3 的 F1 值和准确率相比 T5 (base) 和 T5 (large) 明显降低. 这是由于, GPT 的参数是不释放的, 所以无法通过微调的方法去调整参数, 只能以 API 的方式对其进行调用. 导致 GPT 在实际实验时只能以简单的问答的形式进行, 无法进行推理, 产生了大量错误的回答.

4.4.5 语言模型参数分析

为了探究语言模型的参数大小对 4 - T 模型的影响, 设

置了3组对照实验,采用 Google Flan - T5 的不同参数版本,其中 Flan - T5 - small 参数量 60M、Flan - T5 - base 参数量为 250M,另 Flan - T5 - large,对应的参数量为 780M. 分别在 3 个数据集上对 3 个不同参数的模型进行实验,并用柱状图表示 F1 值的变化,实验结果如图 8 所示.

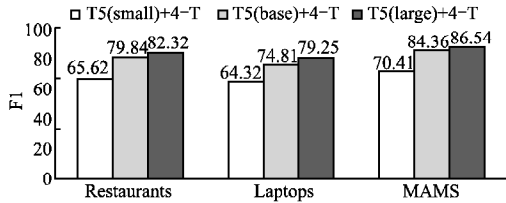


图 8 不同参数大小下的模型 F1 值
Fig. 8 F1 scores by model size

除此之外,设置两组不同语言模型的对照实验,其中 T5 使用的是 large 版本,BART 也使用 large 版本,其参数量为 409M,实验结果如图 9 所示.

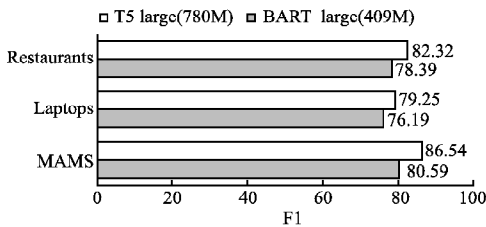


图 9 不同语言模型 F1 值
Fig. 9 F1 scores of different language models

由图 8 和图 9 中可知,模型随着模型参数数量的增大,4 - T 思维链推理的 F1 值提升.说明 4 - T 模型在参数量越大的语言模型上发挥的效果更加显著.这符合直觉,模型的参数越大,语言模型本身的外部知识量越丰富,对于模型推理外部信息起正向作用.

4.4.6 不同提问方式分析

为评估不同问题提示方式对推理过程的潜在影响,设计对比实验,选取多种提问方式进行分析.

不同语言习惯决定了多种提问方式的出现.考虑到全面列举各种提问方式的不切实际性,本文将重点探讨几种典型的提问方法.以 3 - T 为例,设置了 4 组不同的提问方式.在 encoder 编码之后取出句子 token 特征进行可视化分析.其中编号 1 为 3 - T,其余为设置的对照提示.实验结果如图 10 所示.

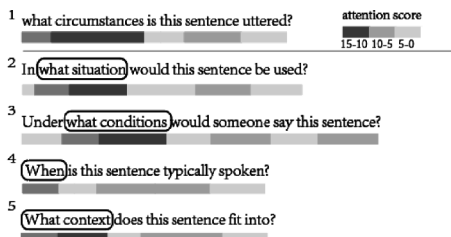


图 10 不同提问方式注意力得分
Fig. 10 Attention scores for different question types

可看出注意力得分较高的为疑问词和其之后的实词.这

说明,提问中的实词和疑问词是推理的关键.从句子语义角度分析;1 和 2 表达的意思比较接近,侧重于句子场景;3 侧重于说出这句话的具体因素或前提条件;4 侧重于句子的时间性,5 侧重于整体背景或上下文.为进一步探究提示语义和关键词对推理的影响,对这 5 组提示设置对照实验,将 2,3,4,5 不同提示分别与 3 - T 进行替换,进行推理;同时对关键词设置掩盖和不掩盖两组进行对照.分别在 3 个数据集上进行实验,得出平均推理错误率,实验结果如图 11 所示.

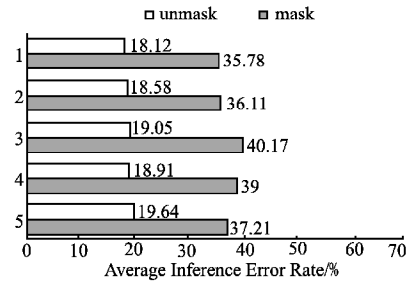


图 11 不同提问方式以及遮掩错误率对比

Fig. 11 Error rate:questioning methods & masking

从图 11 中看出,未进行关键词遮掩时,1 到 5 组提示的平均推理错误率相差不大.这说明不同的提问方式尽管形式不同,但是表达的意思相近,对推理的影响不大.进行关键词遮掩之后 1 ~ 5 组的平均推理错误率明显增大.这说明,提问关键词在整个推理中起主导作用.

深入分析 3 - T 推理过程可以发现,挖掘句子可能出现的场景对于推理至关重要,而这些信息主要由提示中的关键词决定.尽管提问方式在形式上有所不同,编码器在执行注意力操作时更关注的是关键词而非提问形式本身.因此,在句子语义相似的情况下,关键词对推理结果的影响更为显著.

4.4.7 案例分析

Case#: "The crust is thin, the ingredients are fresh and the staff is friendly"

aspect: ingredients sentiment: positive

1 - T: In "The crust is thin, the ingredients are fresh and the staff is friendly", what opinion is a description regarding aspect ingredients.

T5 - large: In the sentence "The crust is thin, the ingredients are fresh and the staff is friendly," the opinion regarding the aspect of ingredients is "the ingredients are fresh." This statement describes the quality of the ingredients used.

2 - T: In the sentence "The crust is thin, the ingredients are fresh and the staff is friendly," the opinion regarding the aspect of ingredients is "the ingredients are fresh." what are the possible reasons behind the emergence of this opinion, and why?

T5 - large: the Reason behind maybe Fresh ingredients often have a superior taste and flavor compared to those that are not fresh.

3 - T: In the sentence "The crust is thin, the ingredients are fresh and the staff is friendly," the opinion regarding the aspect of ingredients is "the ingredients are fresh." the possible reasons behind the emergence of this opinion is Fresh ingredients

often have a superior taste and flavor compared to those that are not fresh. In general, in what circumstances is this sentence uttered, and why?

T5 – large: A customer writes a review or testimonial online (e.g., Yelp, Google Reviews) or in a feedback form.

4 – T: In the sentence "The crust is thin, the ingredients are fresh and the staff is friendly," the opinion regarding the aspect of ingredients is "the ingredients are fresh." the possible reasons behind the emergence of this opinion is Fresh ingredients often have a superior taste and flavor compared to those that are not fresh. In general, the circumstances of this sentence uttered is A customer writes a review or testimonial online (e.g., Yelp, Google Reviews) or in a feedback form. Based on the above analysis and opinion, what is the sentiment polarity of aspect ingredients?

T5 – large: Based on the analysis provided, the sentiment polarity of the aspect "ingredients" in the sentence "The crust is thin, the ingredients are fresh and the staff is friendly" is positive.

每一轮推理中, T5 (large) 可能会产生多个不同的答案, 使用思维一致性约束, 选取一致性最高的结果, 将结果用作下一步的推理。例如: 对于 2 – T 中意见项原因的推理模型给出的答案有多个, 本案例中选择了一致性最高的结果 "Fresh ingredients often have a superior taste and flavor compared to those that are not fresh", 将该结果用来构成 3 – T 进行下一步推理。除此之外, 可以看到 4 – T 思维链推理由浅入深; 并能有效的引导模型从句子潜在外部语义出发理解句子中方面项的情感信息。

5 结束语

为解决现有模型较少考虑句子外部潜在信息和外部潜在信息与句子语义信息不匹配的问题, 本文提出 4 – T 思维链模型。4 个推理轮次中, 不断给出提示, 对句子的外部信息挖掘, 以利用外部信息辅助情感理解。每轮推理中, 为提高句子语义和外部潜在信息的匹配度, 设计思维一致性约束。为提高模型的推理效率, 设计特征裁剪。最后, 经过第 4 个轮次推理得出情感分析结果。在 3 个公开数据集上的实验结果表明, 所提模型是有效的, 通过 4 – T 思维链模型可进一步提高方面级情感分析的准确率。

尽管思维链推理取得了较好的性能, 但仍然存在不足。如: 特征裁剪采用可学习的特征裁剪网络来优化推理过程, 不可避免在裁剪过程中丢失一些推理信息。未来考虑采取更精细的裁剪策略在保持模型性能的同时进一步提高推理效率。

References:

- [1] ZHANG T Z, ZHOU G, LIU H B, et al. Text-image gated fusion mechanism for multimodal aspect-based sentiment analysis [J]. Computer Science, 2024, 51 (9) : 242-249.
- [2] Seoh R, Birlle I, Tak M, et al. Open aspect target sentiment classification with natural language prompts [C] // Conference on Empirical Methods in Natural Language Processing, 2021 : 6311-6322.
- [3] Birjali M, Kasri M, Beni Hssane A. A comprehensive survey on sentiment analysis: approaches, challenges and trends [J]. Knowledge-Based Systems, 2021, 226 : 107134, doi: 10. 1016/j. knosys. 2021. 107134.
- [4] Liang Y, Meng F, Zhang J, et al. A novel aspect-guided deep transition model for aspect based sentiment analysis [C] // Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019 : 5569-5580.
- [5] Xiao Z, Wu J, Chen Q, et al. Bert4gcn: using bert intermediate layers to augment gcn for aspect-based sentiment classification [C] // Conference on Empirical Methods in Natural Language Processing, 2021 : 9193-9200.
- [6] Chen G, Tian Y, Song Y. Joint aspect extraction and sentiment analysis with directional graph convolutional networks [C] // 28th International Conference on Computational Linguistics, 2020 : 272-279.
- [7] Tian Y, Chen W, Hu B, et al. End-to-end aspect-based sentiment analysis with combinatory categorial grammar [C] // Findings of the Association for Computational Linguistics, 2023 : 13597-13609.
- [8] Zhang Z, Zhou Z, Wang Y. Ssegcn: syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis [C] // Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022 : 4916-4925.
- [9] Devlin J, Chang M W, Lee K, et al. Bert: pre-training of deep bidirectional transformers for language understanding [C] // Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019 : 4171-4186.
- [10] Tang H, Ji D, Li C, et al. Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification [C] // 58th Annual Meeting of the Association for Computational Linguistics, 2020 : 6578-6588.
- [11] HUANG J, LIU Y, WANG Q F, et al. Dual-channel graph convolutional network with word-order knowledge for aspect-based sentiment analysis [J]. Application Research of Computers, 2024, 41 (3) : 779-785.
- [12] ZHANG X Y, YAN Z. Three-level joint prompt-tuning implicit sentiment analysis method incorporating LLMs [J]. Application Research of Computers, 2024, 41 (10) : 2900-2905.
- [13] Wang K, Shen W, Yang Y, et al. Relational graph attention network for aspect-based sentiment analysis [C] // 58th Annual Meeting of the Association for Computational Linguistics, 2020 : 3229-3238.
- [14] Tian Y, Chen G, Song Y. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble [C] // Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021 : 2910-2922.
- [15] Chen C, Teng Z, Wang Z, et al. Discrete opinion tree induction for aspect-based sentiment analysis [C] // 60th Annual Meeting of the Association for Computational Linguistics, 2022 : 2051-2064.
- [16] Ma F, Hu X, Liu A, et al. AMR-based network for aspect-based sentiment analysis [C] // 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023 : 322-337.
- [17] Fei H, Li B, Liu Q, et al. Reasoning implicit sentiment with chain-of-thought prompting [C] // 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers),

- 2023;1171-1182.
- [18] Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 24824-24837, doi: 10.48550/arXiv.2201.11903.
- [19] Besta M, Blach N, Kubicek A, et al. Graph of thoughts; solving elaborate problems with large language models[C]//AAAI Conference on Artificial Intelligence, 2024; 17682-17690.
- [20] Yang L, Yu Z, Zhang T, et al. Buffer of thoughts; thought-augmented reasoning with large language models[EB/OL]. <https://arxiv.org/abs/2406.04271>, 2024.
- [21] Fungwacharakorn W, Thanh N H, Zin M M, et al. Layer-of-thoughts prompting (lot); leveraging llm-based retrieval with constraint hierarchies[J]. *arXiv preprint arXiv:2410.12153*, 2024.
- [22] Yuan J, Du D, Zhang H, et al. Reversal of thought; enhancing large language models with preference-guided reverse reasoning warm-up[J]. *arXiv preprint arXiv:2410.12323*, 2024.
- [23] Dong L, Wei F, Tan C, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014; 49-54.
- [24] Jiang Q, Chen L, Xu R, et al. A challenge dataset and effective models for aspect-based sentiment analysis[C]//Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019; 6280-6285.
- [25] Pontiki M, Galanis D, Papageorgiou H, et al. Semeval-2016 task 5: aspect based sentiment analysis[C]//Workshop on Semantic Evaluation, 2016; 19-30.
- [26] Yin S, Zhong G. TextGT: a double-view graph transformer on text for aspect-based sentiment analysis[C]//AAAI Conference on Artificial Intelligence, 2024; 19404-19412.
- [27] Auer S, Bizer C, Kobilarov G, et al. Dbpedia: a nucleus for a web of open data[C]//International Semantic Web Conference, 2007; 722-735.
- [28] Hutto C, Gilbert E. Vader: a parsimonious rule-based model for sentiment analysis of social media text[C]//International AAAI Conference on Web and Social Media, 2014; 216-225.

附中文参考文献:

- [1] 张添植, 周刚, 刘洪波, 等. 基于文本和图像门控融合机制的多模态方面级情感分析[J]. *计算机科学*, 2024, 51(9): 242-249.
- [11] 黄俊, 刘洋, 王庆凤, 等. 基于语序知识的双通道图卷积网络方面级情感分析[J]. *计算机应用研究*, 2024, 41(3): 779-785.
- [12] 张小艳, 闫壮. 融合大语言模型的三级联合提示隐式情感分析方法[J]. *计算机应用研究*, 2024, 41(10): 2900-2905.