

# 结合语法结构和语义信息的情感三元组提取<sup>①</sup>



杨芳捷<sup>1</sup>, 冯广<sup>2</sup>, 唐业凯<sup>1</sup>

<sup>1</sup>(广东工业大学 计算机学院, 广州 510006)

<sup>2</sup>(广东工业大学 自动化学院, 广州 510006)

通信作者: 杨芳捷, E-mail: 2291959954@qq.com

**摘要:** 针对目前大多数方面情感三元组提取方法存在着没有充分考虑语法结构和语义相关性的问题. 本文提出一种结合语法结构和语义信息的方面情感三元组提取模型, 首先提出使用依赖解析器得到所有依赖弧的概率矩阵构建语法图, 提取丰富的语法结构信息. 其次利用自注意力机制构建语义图, 表示单词与单词之间的语义相关性, 从而减低噪声词的干扰. 最后设计了一个相互仿射变换层, 让模型可以更好地交换语法图和语义图之间的相关特征, 提升模型情感三元组提取的表现. 在多个公开数据集上进行验证. 实验表明, 与现有的情感三元组提取模型相比, 精确度 ( $P$ )、召回率 ( $R$ ) 和  $F1$  值整体都有提高, 验证了结合语法结构和语义信息在方面情感三元组提取的有效性.

**关键词:** 方面情感三元组提取; 语法结构; 语义信息; 图卷积网络; 自注意力机制

引用格式: 杨芳捷, 冯广, 唐业凯. 结合语法结构和语义信息的情感三元组提取. 计算机系统应用, 2024, 33(3): 255-263. <http://www.c-s-a.org.cn/1003-3254/9438.html>

## Sentiment Triple Extraction Combining Grammatical Structure and Semantic Information

YANG Fang-Jie<sup>1</sup>, FENG Guang<sup>2</sup>, TANG Ye-Kai<sup>1</sup>

<sup>1</sup>(School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China)

<sup>2</sup>(School of Automation, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** Most of the current aspect sentiment triplet extraction methods do not fully consider syntactic structure and semantic relevance. This study proposes an aspect sentiment triplet extraction model that combines syntactic structure and semantic information. First, the study proposes to construct a grammatical graph with a dependency parser to get the probability matrices of all dependency arcs, extracting rich information of syntactic structure. Second, it utilizes the self-attention mechanism to construct a semantic graph, which represents the semantic correlation between words, thus reducing the interference of noisy words. Finally, a mutual affine transformation layer is designed to allow the model to better exchange the relevant features between the syntactic graph and semantic graph to improve the performance of the model in sentiment triplet extraction. The model is validated on several public datasets. The experiments show that compared with the existing sentiment triplet extraction models, the precision ( $P$ ), recall ( $R$ ), and  $F1$  value are all improved. This validates the effectiveness of combining syntactic structure and semantic information in aspect sentiment triplet extraction.

**Key words:** aspect sentiment triplet extraction; grammatical structure; semantic information; graph convolutional network (GCN); self-attention mechanism

## 1 引言

随着互联网的迅猛发展, 社交媒体逐渐成为情感

表达和信息分享的主要平台, 越来越多的消费者在互联网平台上分享他们对某个产品所发表的观点和评价.

① 基金项目: 国家自然科学基金重点项目 (62237001); 广东省哲学社会科学青年项目 (GD23YJY08)

收稿时间: 2023-09-19; 修改时间: 2023-10-20; 采用时间: 2023-11-03; csa 在线出版时间: 2024-01-18

CNKI 网络首发时间: 2024-01-19

这些评论数据蕴含着用户的情绪信息,对这些用户的情绪信息进行挖掘不仅能够帮助平台商家更好地理解用户需求,而且在面对产品推广和改进时,了解用户对其产品或具体方面的情感反馈,可以用于产品推广和广告策略,从而强调产品的优点。目前情感分析在计算领域有着广泛的应用,方面级情感分析 (aspect-based sentiment analysis, ABSA) 又称细粒度情感分析。它不仅是识别文本的情感极性,还进一步提取关于不同方面的情感信息。而方面情感三元组提取 (aspect sentiment triplet extraction, ASTE) 是 ABSA 的一种新变体。ASTE 旨在从句子中提取方面术语和意见术语,并确定它们的情感极性。例如,图 1 中展示的是 ASTE 的例子,价格和餐厅环境两个方面的情绪极性分别为积极和消极。其中方面词和意见词分别用绿色和蓝色显示,情感极性用红色表示。黑色箭头表示这两个单词具有句法依赖性,即方面词“price”和意见词“reasonable”之间存在依赖关系。通过 ASTE 可以精确地识别用户对某个方面的态度,而不是简单地对整个句子分配情感极性,从而帮助理解人们对于不同方面的意见和情感倾向。

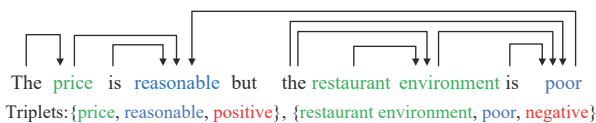


图 1 ASTE 的典型例子

目前大多数方面情感三元组提取方法存在以下两个问题。

(1) 根据依赖树中的依赖类型提取方面词和意见词时,没有充分考虑句子的语法信息,导致会出现依赖关系分析结果的不准确性的问题。

(2) 由于评论数据的复杂性和不规则性,方面词和意见词之间隐含的语义关系往往被忽略,导致无法全面准确地反映用户对特定方面复杂的情感表达。

为解决上述问题,本文提出了一种结合语法结构和语义信息的方面情感三元组提取模型。特别地,针对依赖关系分析结果的不准确性的问题,我们使用最先进的依赖解析模型 LAL-Parser 得到所有依赖弧的概率矩阵,其中包含了丰富的语法信息,可以有效增强方面词和意见词之间的对应关系,减少依赖关系分析错误。此外针对无法全面准确地获取用户对特定方面的情感表达,清楚用户需求,我们设计了一个自注意力机制模块。该模块利用自注意力机制构造语义图的邻接矩阵,

表示词之间的语义相关性,通过自注意力机制降低噪声信息的不利影响,从而提升模型对语义相关性的理解。最后,我们还提出相互仿射变换层,用于更好地捕获语法结构和文本中的语义信息。

## 2 相关研究

传统的情感分析任务是面向句子级或文档级的。相比之下,ABSA 是情感分析中更细粒度的任务。而 ASTE 是 ABSA 领域的最新子任务之一。最近, Peng 等人<sup>[1]</sup>首先提出了 ASTE 任务并采用了两阶段管道模型。第 1 阶段是预测所有方面词和它们相关的情感极性,以及预测能够描述方面词对应的意见词。第 2 阶段是将方面词与情感和意见词进行配对,但是这些方法会存在错误传播的问题,为了避免这种流水线方法的缺点, Xu 等人<sup>[2]</sup>提出了一种创新的位置感知标记方案,利用端到端的 JET 模型来提取三元组。与传统方法相比,我们首次设计了一套表达能力更强的新标签集,从而直接避免了原模型在特征提取的完整性方面存在的缺点。类似的, Wu 等人<sup>[3]</sup>提出了一种全新的标记方案,即网格标记方案 (GTS),它通过一个统一的网格标记任务一次性提取所有情感三元组。Span ASTE<sup>[4]</sup>是一种针对 ASTE 任务的方法,它特别注重了短语之间的相互作用。相比现有方法,它克服了提取不完全和情感冲突的问题,提供了更准确和一致的结果。Zhang 等人<sup>[5]</sup>为了进一步探索这项任务,他们成功实现了一个多任务学习框架,并结合启发式规则来生成最终的三元组。Mao 等人<sup>[6,7]</sup>将 ASTE 任务转换为机器阅读理解问题,并利用共享的 BERT 编码器在多个阶段解码以获得三元组。在这个转换过程中,他们将 ASTE 的输入文本和标签转化为适合机器阅读理解的格式,使得模型可以根据给定的问题和文本来提取相关的三元组信息。与以往基于标记方案的方法不同,还有研究引入了一种端到端无标记解决方案<sup>[8]</sup>,解决了先前基于标记的方法存在的不足之处。此外,ASTE 领域还有许多其他的建模范式被用于处理任务。在一项研究中<sup>[9]</sup>,作者将 ASTE 任务转换为文本生成任务,并设计了两种生成范式。另一项研究<sup>[10]</sup>中的作者将 ASTE 任务转换为无序三元组集识别问题,并将其建立在编码器-解码器架构上。

近期研究证明图卷积网络可以有效地处理 ASTE 任务。Chen 等人<sup>[11]</sup>使用图神经网络 (GNNs) 为 ASTE 任务设计了一种图-序列双重表示和建模范式,并且也

证明了可行性. Sun 等人<sup>[12]</sup>提出了 ABSA 任务的依存树 (CDT) 模型的卷积. 该模型引入了解析器解析的依存树信息, 帮助识别与方面词相关的意见词, 并结合 GCN 在依存树上学习的语法信息特征. 但这些方法都会出现依赖关系分析结果的不准确性的问题. 此外, Zhou 等人<sup>[13]</sup>提出了一种新的基于语法和知识的 GCN (SK-GCN) 模型. Hu 等人<sup>[14]</sup>提出了一个新的端到端的模型, 结合了语法成分解析树和常识知识图, 该模型通过图卷积网络 (GCN) 将它们整合到 ASTE 任务, 帮助更好地识别方面词和与之相关的意见词. Shi 等人<sup>[15]</sup>提出了一种创新的依赖图增强的交互式注意力网络, 该网络明确地考虑了词之间的语法关系. 最近, Chen 等

人<sup>[16]</sup>提出了一种新颖的解决方案, 该方案将句子中任意两个词之间的关系分为 10 种不同的类型, 并在建模过程中引入了语言特征. 然而, 目前现有的利用单词之间关系的方法存在一些局限性. 他们忽略了句子中语法结构和丰富的语义信息. 无法全面准确地获取用户对特定方面的情感表达, 清楚用户的需求. 因此仍然存在改进的空间.

### 3 结合语法结构和语义信息的情感三元组提取

本文提出的结合语法结构和语义信息的情感三元组提取模型框架图如图 2 所示, 模型主要由以下 7 个部分组成.

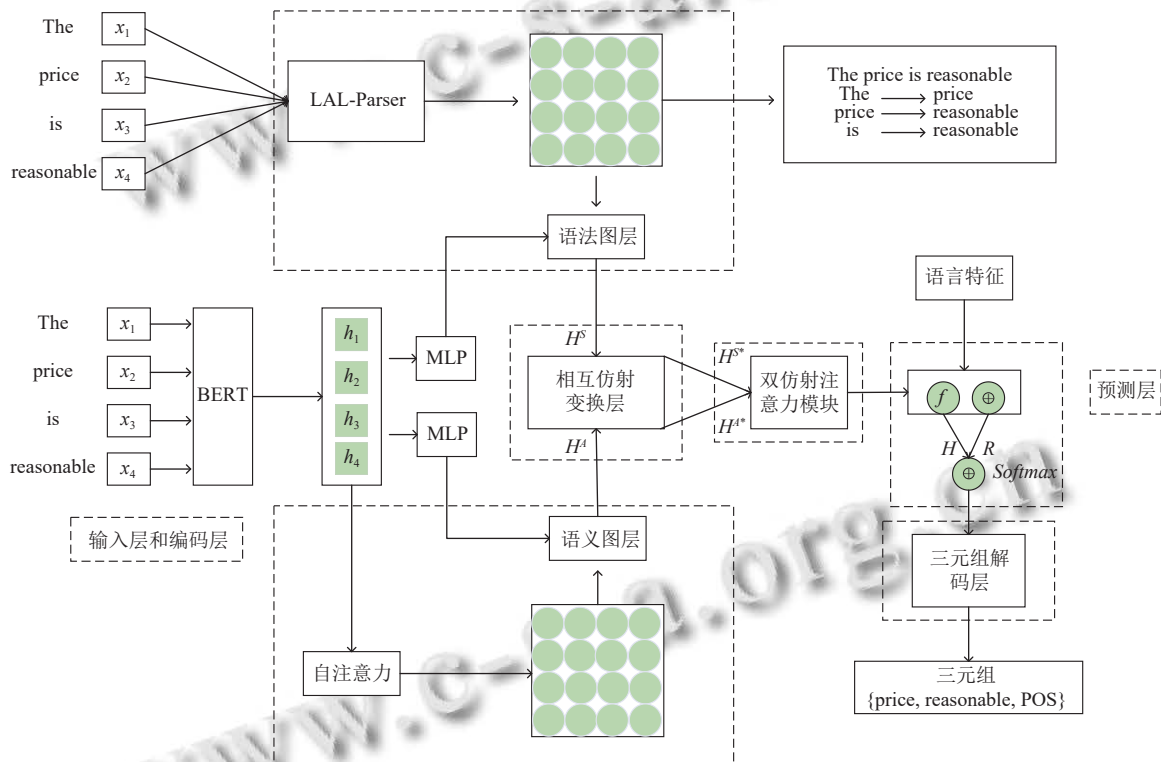


图 2 结合语法结构和语义信息的情感三元组提取模型图

(1) 输入层和编码层. 采用 BERT 模型对输入句子进行编码处理, 得到编码词序列.

(2) 语法图层. 利用依存解析模型生成所有依赖弧的概率矩阵, 获取句子中潜在的语法结构信息, 得到句子中词与词之间的语法关系.

(3) 语义图层. 使用注意力机制对编码词序列的语义信息进行动态地捕捉从而获取句子中单词与单词之间丰富的语义关系.

(4) 相互仿射变换层. 可以有效地传递和交换语法

图层和语义图层模块之间的相关特征.

(5) 双仿射注意力模块. 将交换语法图层和语义图层后的表示输入至双仿射注意力网络层进行词对关系分析, 得到多通道关系邻接张量.

(6) 预测层. 在依据多通道关系邻接张量计算出节点聚合矩阵后, 根据节点聚合矩阵、多通道关系邻接张量和预设语言特征张量计算词节点表示和边表示. 并采用分类器链根据词节点表示和边表示进行分类预测, 得到标签概率分布.

(7) 三元组解码层. 根据标签概率分布对目标文本句子进行三元组解码分析, 得到情感三元组, 情感三元组包括方面词、意见词和情感极性.

### 3.1 任务描述

给定一个输入句子  $X = \{x_1, x_2, \dots, x_n\}$ , 有  $n$  个单词, 我们模型的目标是从句子输出一组或多组三元组  $T = \{(a, o, s)_m\}$ , 例如, {"price", "reasonable", "positive"}, 其中  $a$  和  $o$  分别表示方面词和意见词,  $s$  属于情感标签集  $s = \{POS, NEU, NEG\}$ . 也就是说, 情绪标签集由 3 种情绪极性组成: 积极、中性和消极.  $m$  表示句子  $X$  有情感三元组的总数.

### 3.2 标记方案

我们为 ASTE 定义了句子中单词的 10 种关系. 具

体来说, 这 10 种关系 {B-A, I-A, A, B-O, I-O, O, POS, NEU, NEG,  $\perp$ }, 用于对每个词对  $(w_i, w_j)$  之间的关系进行标注. 具体而言, 其中 4 个关系 {B-A, I-A, B-O, I-O} 旨在提取方面术语和意见术语. B 和 I 分别表示术语的开头和内部, 而 -A 和 -O 子标签则用于确定术语是方面还是意见. A 和 O 关系用于检测由两个不同的词组成的词对是否分别属于同一个方面或意见词. 另外, 还有 3 个情感关系 {POS, NEU, NEG}, 用于检测一个词对是否匹配, 并判断方面意见对的情感极性. 采用这种标记方案, 我们能够准确地对词对之间的关系进行建模, 并提取出方面术语、意见术语以及它们之间的情感极性. ASTE 的标记方案如图 3 所示, 其中方面词和意见词分别以绿色和蓝色显示, 情感极性用红色表示.

	The	price	is	reasonable	but	the	restaurant	environment	is	poor
The	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$
price	$\perp$	B-A	$\perp$	POS	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$
is	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$
reasonable	$\perp$	POS	$\perp$	B-O	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$
but	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$
the	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$
restaurant	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	B-A	A	$\perp$	NEG
environment	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	A	I-A	$\perp$	NEG
is	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$
poor	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	NEG	NEG	$\perp$	B-O

图 3 ASTE 任务的标记方案

### 3.3 模型

#### 3.3.1 输入层和编码层

BERT 已经证明了它在各种任务中的有效性. 我们使用 BERT 作为句子编码器来提取隐藏的上下文表示. 给定一个输入句子  $X = \{x_1, x_2, \dots, x_n\}$  作为输入, 编码层输出隐藏的表示序列  $H = \{h_1, h_2, \dots, h_n\}$ , 其中,  $n$  表示句子长度.

#### 3.3.2 语法图层

我们使用 BERT 作为句子编码器, 用于提取隐藏

状态向量表示, 随后将经过多层感知机得到隐藏状态向量作为语法图的初始节点表示. 为了编码语法信息, 我们采用了最先进的依赖解析模型 LAL-Parser<sup>[17]</sup>, 用于生成所有依赖弧的概率矩阵. 这个依赖概率矩阵包含了句子中潜在的语法结构信息, 通过提供不同的依赖弧概率, 反映了句子中词之间的语法关系. 通过将 BERT 和 LAL-Parser 结合起来, 这种结合可以帮助我们更好地理解句子中的语法结构.

给定一个有  $n$  个节点的图, 我们使用依赖解析器获

得的依赖概率矩阵  $A^s \in R^{n \times n}$  的语法编码, 语法图表示  $H^S = \{h_1^S, h_2^S, \dots, h_n^S\}$  使用式 (1) 获得:

$$h_i^S = \sigma \left( \sum_{j=1}^n A_{ij}^s W h_j + b \right) \quad (1)$$

其中,  $W$  为权矩阵,  $b$  为偏置项,  $A^s$  是依赖解析器的最终离散输出表示,  $A^s$  中的元素  $A_{ij}^s$  表示第  $i$  个节点是否连接到第  $j$  个节点. 如果第  $i$  个节点与第  $j$  个节点相连,  $A_{ij}^s = 1$ , 否则  $A_{ij}^s = 0$ .  $\sigma$  为激活函数 (ReLU). 这里,  $h_j$  是第  $j$  个节点的隐藏表示.

### 3.3.3 语义图层

通过使用自注意机制获得注意力得分矩阵作为邻接矩阵, 我们可以构建一个更加灵活和适应性强的图结构, 用于表示句子中的单词之间的关系. 这种注意力机制允许模型在句子的表示中动态地捕捉重要的语义信息, 无需依赖预定义的语法结构. 因此, 自注意力能够更好地适应不同类型的句子和语言表达方式. 因为它能够更全面地捕捉句子的语义信息, 并且对于处理不规则性和复杂语法的句子也具有较好的适应性.

自注意力可以并行计算每对元素的注意分数. 在我们的模型中, 我们使用自注意层计算注意力得分矩阵  $A^a \in R^{n \times n}$ , 可以表示为:

$$A^a = \text{Softmax} \left( \frac{QW^Q \times (KW^K)^T}{\sqrt{d}} \right) \quad (2)$$

其中, 矩阵  $Q$  和  $K$  都是语义图模块前一层图表示, 而  $W^Q$  和  $W^K$  都是可学习的权重矩阵. 另外,  $d$  是输入节点特征的维数. 我们使用一个自注意头来获得一个句子的注意得分矩阵.

我们使用注意力得分矩阵邻接矩阵  $A^a \in R^{n \times n}$  的语法编码, 语义图表示  $H^A = \{h_1^A, h_2^A, \dots, h_n^A\}$  使用式 (3) 获得:

$$h_i^A = \sigma \left( \sum_{j=1}^n A_{ij}^a W h_j + b \right) \quad (3)$$

其中,  $W$  为权矩阵,  $b$  为偏置项,  $A^a$  是注意力得分矩阵邻接矩阵最终输出表示,  $A^a$  中的元素  $A_{ij}^a$  表示第  $i$  个节点是否连接到第  $j$  个节点. 如果第  $i$  个节点与第  $j$  个节点相连,  $A_{ij}^a = 1$ , 否则  $A_{ij}^a = 0$ .  $\sigma$  为激活函数 (ReLU). 这里,  $h_j$  是第  $j$  个节点的隐藏表示.

### 3.3.4 相互仿射变换层

为了有效地传递和交换语法图层和语义图层模块之间的相关特征, 我们引入了一种相互的仿射变换作为它们之间的桥梁. 这个相互的仿射变换可以在语法

结构和语义信息之间建立强大的联系, 并促进信息的有效传递.

$$H^{S^*} = \text{Softmax}(H^S (H^A)^T) H^A \quad (4)$$

$$H^{A^*} = \text{Softmax}(H^A (H^S)^T) H^S \quad (5)$$

其中,  $H^S, H^A$  分别为语法图和语义图表示,  $(H^S)^T, (H^A)^T$  分别为  $H^S, H^A$  的转置表示.  $\text{Softmax}$  函数生成概率分布.

### 3.3.5 双仿射注意力模块

已经有研究证明在语法依赖解析任务中, 双仿射注意力模块是一种有效的方法. 因此, 我们采用了双仿射注意力模块来捕获句子中每个词对之间的关系概率分布. 双仿射注意过程被表述为:

$$R = \text{Biaffine}(MLP(H), MLP(H^{S^*})) \quad (6)$$

$$R^* = \text{Biaffine}(MLP(H^{A^*}), MLP(H)) \quad (7)$$

$$R = R + R^* \quad (8)$$

其中,  $MLP, \text{Biaffine}$  分别是多层感知器和双仿注意力模块,  $H$  为 BERT 编码层输出的隐藏表示序列.  $H^{S^*}, H^{A^*}$  分别是式 (4) 和式 (5) 语法图表示和语义图表示.

### 3.3.6 预测层

首先, 为了建模单词之间的各种关系, 我们对最原始的图卷积网络 GCN 进行了扩展. 这个扩展版本包含双仿射注意模块, 用于构建多通道邻接张量  $R \in R^{n \times n \times m}$ . 每个通道都代表了对单词之间关系的不同建模方式. 然后, 我们使用 GCN 沿着每个通道对每个节点进行信息聚合. 这个过程可以用以下方式进行表述:

$$H_m = \sigma(R_m H W_m + b_m) \quad (9)$$

$$H = f(H_1, H_2, \dots, H_m) \quad (10)$$

其中,  $R_m \in R^{n \times n}$  表示  $R_m$  的第  $m$  个通道切片.  $m$  为标记方案中 10 种标签关系数量.  $W_m$  和  $b_m$  是可学习的权重和偏差.  $\sigma$  是一个激活函数 (ReLU).  $H$  为 BERT 编码层输出的隐藏表示序列. 平均池化函数  $f$  应用于所有通道的节点隐藏表示. 最后的  $H$  即为节点聚合矩阵.

同时为了增强我们的模型, 我们为每个词对引入了与 EMC-GCN<sup>[16]</sup> 中一致的 4 种类型的语言特征, 包括词性组合、语法依赖类型、基于树的距离和相对位置距离即  $R^{psc}, R^{dep}, R^{tbd}$  和  $R^{rpd}$ , 特别地, 这 4 个邻接张量是根据对应的语言特征随机初始化的. 使用这些邻接张量进行图卷积操作以获得节点表示  $H^{psc}, H^{dep}$ ,

$H^{ibd}$  和  $H^{rpd}$ . 最后, 分别将平均池化函数和连接操作应用于所有节点表示和所有边.

$$H = f(H, H^{psc}, H^{dep}, H^{ibd}, H^{rpd}) \quad (11)$$

$$R = R \oplus R^{psc} \oplus R^{dep} \oplus R^{ibd} \oplus R^{rpd} \quad (12)$$

其中,  $H = \{h_1, h_2, \dots, h_n\}$  表示词对的节点表示,  $R = \{r_{1,1}, r_{1,2}, \dots, r_{n,n}\}$  表示词对的边表示,  $\oplus$  表示连接.

其次, 为了获得用于标签预测的词对  $(w_i, w_j)$  的表示, 我们将它们的节点表示  $h_i$ 、 $h_j$  和它们的边表示  $r_{ij}$  连接起来. 此外, 受多标签分类任务中的分类器链方法<sup>[18]</sup> 的启发, 我们采用了一种细化过程, 通过将方面和意见提取的隐含结果融入到词对的表示中, 以提高对情感关系的判断准确性. 这样的策略可以使得模型更好地理解词对之间的语义关联. 具体来说, 假设有两个词,  $w_i$  属于方面词,  $w_j$  属于意见词. 那么这两个词组成的词对  $(w_i, w_j)$  更有可能被预测为情感关系, 即 POS、NEU 或 NEG. 否则, 它们不太可能匹配. 即:

$$s_{i,j} = h_i \oplus h_j \oplus r_{ij} \quad (13)$$

最后, 我们将词对表示  $s_{i,j}$  输入线性层, 然后使用 *Softmax* 函数生成标签概率分布  $p_{i,j}$ , 即:

$$p_{i,j} = \text{Softmax}(W s_{i,j} + b) \quad (14)$$

其中,  $W$  和  $b$  是可学习的权重和偏差.

### 3.3.7 三元组解码层

我们采用了一个上三角表来解码三元组. 首先, 使用基于主对角线的所有词对的预测关系  $(w_i, w_i)$ , 来提取方面术语和意见术语, 并将其作为方面术语和意见术语的提取结果. 如果预测关系中存在任何情感关系, 则认为方面术语和意见术语是配对的, 否则就不配对. 最后, 为了判断方面-意见对的情感极性, 将预测关系中出现次数最多的情感关系作为情感极性. 因此, 我们得到了一个三元组  $\{a, o, s\}$ .

为了精确捕捉单词之间的关系, 我们对经过双仿射注意力模块获得的相邻张量施加约束.

$$L_b = - \sum_i^n \sum_j^n \sum_{c \in C} F(y_{ij} == c) \log(r_{i,j|c}) \quad (15)$$

其中,  $F$  表示指示函数,  $y_{ij}$  是词对  $(w_i, w_j)$  的真值,  $n$  表示句子长度,  $C$  表示关系集,  $r_{i,j|c}$  是经过双仿射注意力模块后预测生成标签的概率值. 同样, 我们对语言特征产生的 4 个相邻张量施加关系约束, 约束成本表示为

$L^{psc}$ 、 $L^{dep}$ 、 $L^{ibd}$  和  $L^{rpd}$ .

标准交叉熵损失  $L_p$  用于 ASTE 任务, 即:

$$L_p = - \sum_i^n \sum_j^n \sum_{c \in C} F(y_{ij} == c) \log(p_{i,j|c}) \quad (16)$$

我们的目标是最小化目标函数:

$$L = L_p + \mu L_b + \nu (L^{psc} + L^{dep} + L^{ibd} + L^{rpd}) \quad (17)$$

其中, 参数  $\mu$  和  $\nu$  用于调整对应关系约束损失的影响.

## 4 实验

在本节中, 我们将介绍有关实验的信息, 包括数据集、评估指标、基线、实现细节、实验结果和消融研究.

### 4.1 数据集

我们在两个 ABSA 数据集上评估我们的方法. 第 1 个数据集 ASTE-Data-v1 由 Wu 等人<sup>[3]</sup> 注释. 第 2 个数据集 ASTE-Data-v2 由 Xu 等人<sup>[2]</sup> 注释. 这两个数据集都包含来自笔记本电脑领域和餐厅领域的用户评论数据. 它们全部分为 3 个部分: 训练集、验证集和测试集. 这些数据集的详细信息如表 1 所示.

表 1 实验数据集统计表

Dataset	Rest14		Lap14		Rest15		Rest16		
	#S	#T	#S	#T	#S	#T	#S	#T	
ASTE-Data-v1	train	1259	2356	899	1452	603	1038	863	1421
	dev	315	580	225	383	151	239	216	348
	test	493	1008	332	547	325	493	328	525
ASTE-Data-v2	train	1266	2338	906	1460	605	1013	847	1394
	dev	310	577	219	346	148	249	210	339
	test	492	994	328	543	322	485	326	514

注: #S和#T分别表示句子数和三元组数量.

### 4.2 评估指标

为了客观地评估不同方法在 ASTE 任务中的性能, 我们采用了广泛使用的评估指标, 包括精确度 ( $P$ )、召回率 ( $R$ ) 和  $F1$  值. 在这种评估中, 我们定义一个三元组在其所有元素与基准数据中的真实取值一致时, 该三元组为正确的. 为了确保结果的稳定性, 我们使用  $F1$  值选择最佳模型权重, 以确保选择最优模型进行测试. 同时, 我们使用 3 次不同的随机种子进行实验, 并取平均值作为最终的报告结果.

### 4.3 基线

我们将以下模型作为本方法的基线模型进行比较. 这些方法可以简单地分为管道方法、端到端方法和基

于阅读理解这3类方法。

(1) 管道方法: Peng-twostage+IOG 和 IMN+IOG.

(2) 端到端方法: OTE-MTL, JET, GTS-BERT, DGEIAN, GCN-EGTS (BERT), EMC-GCN.

(3) 基于MRC的方法: BMRC.

#### 4.4 实现细节

本文选用BERT预训练的词向量作为词嵌入. 为了编码语法信息, 我们利用依赖解析器得到所有依赖弧的概率矩阵来捕获丰富的语法结构信息, 语法图层和语义图层的dropout率设置为0.1, 层数设置为1, 它们的隐藏状态维数设置为300. AdamW优化器用于BERT微调的学习率为 $2 \times 10^{-5}$ , 其他可训练参数的学习率为 $10^{-3}$ , dropout设置为0.5. BERT和GCN的隐藏状态维数分别设置为768和300. 该模型在100个epoch中训练, 批量大小为16.

#### 4.5 实验结果

主要实验结果如表2和表3所示. 在F1指标下,

我们的模型在所有数据集上优于所有其他替代方案. 此外, 在大多数情况下, 我们的模型的性能超过其他方法在P和R的度量. 我们观察到, 端到端和基于MRC的方法比流水线方法取得了更显著的改进, 因为它们建立了这些子任务之间的相关性, 我们观察到端到端和基于MRC的方法比流水线方法取得了更显著的改进, 因为它们建立了这些子任务之间的相关性, 并通过联合训练多个子任务来缓解错误传播问题. OTE-MTL和GTS-BERT的标记方案类似于表填充. 与GTS-BERT相比, 我们的模型在ASTE-Data-v1和ASTE-Data-v2数据集上F1值显著超过了GTS-BERT的F1值平均3.26%和3.67%, 在ASTE-Data-v1数据集上, 与EMC-GCN相比我们的模型在Rest14, Lap14, Rest15和Rest16上分别实现了F1值的1.20%, 1.35%, 2.19%和1.90%的明显绝对增加. 实验结果表明在数据量比较小的Rest15和Rest16数据集上, 融合语法结构和丰富的语义信息可以更好地对情感三元组进行提取.

表2 ASTE-Data-v1上的实验结果(%)

Model	Rest14			Lap14			Rest15			Rest16		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Peng-unified-R+LOG	58.89	60.41	59.64	48.62	45.52	47.02	51.70	46.04	48.71	59.25	58.09	58.67
IMN+IOG	59.57	63.88	62.40	49.21	46.23	47.68	55.24	52.33	53.75	—	—	—
DGEIAN	71.03	62.63	66.55	<u>60.74</u>	45.56	51.72	<b>64.87</b>	52.75	57.11	<b>69.07</b>	65.64	67.30
GCN-EGTS (BERT)	70.14	68.07	69.20	54.54	52.27	53.64	59.23	58.15	58.84	66.89	65.86	66.28
GTS-BERT	70.92	69.49	70.20	57.52	51.92	54.58	59.29	58.07	58.67	<u>68.58</u>	66.60	67.58
EMC-GCN*	<u>71.57</u>	<u>72.22</u>	<u>71.89</u>	58.37	<b>58.16</b>	<u>58.27</u>	56.77	<u>62.06</u>	<u>59.30</u>	64.72	<u>71.50</u>	<u>67.94</u>
Ours	<b>72.41</b>	<b>73.79</b>	<b>73.09</b>	<b>61.64</b>	<u>57.79</u>	<b>59.65</b>	<u>60.44</u>	<b>62.57</b>	<b>61.49</b>	67.07	<b>72.84</b>	<b>69.84</b>

注: “\*”表示相同情况下原论文方法复现的结果, 其他结果是从已发表论文中得到的, 其中最佳效果用粗体标出, 次佳效果用下划线标出.

表3 ASTE-Data-v2上的实验结果(%)

Model	Rest14			Lap14			Rest15			Rest16		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
OTE-MTL	62.00	55.97	58.57	49.53	39.22	43.42	56.37	40.94	43.42	62.88	52.10	56.96
JET (M=6)	61.50	55.13	58.14	53.03	33.89	41.35	<u>64.37</u>	44.33	52.50	<u>70.94</u>	57.00	63.21
DGEIAN	71.68	61.62	66.26	60.15	43.44	51.14	61.84	50.99	55.89	69.40	60.15	64.37
GTS-BERT	68.09	<u>69.54</u>	<u>68.81</u>	59.40	<b>59.40</b>	55.42	59.28	<u>57.93</u>	58.60	68.32	<u>66.86</u>	<u>67.58</u>
BMRC	<b>75.61</b>	61.77	67.99	<b>70.55</b>	48.98	<u>57.82</u>	<b>68.51</b>	53.40	<u>60.02</u>	<b>71.20</b>	61.08	65.75
Ours	<u>71.71</u>	<b>74.69</b>	<b>73.16</b>	<u>60.51</u>	<u>58.04</u>	<b>59.24</b>	61.55	<b>65.08</b>	<b>63.25</b>	66.96	<b>72.12</b>	<b>69.44</b>

注: 模型结果是从已发表论文中得到的, 其中最佳效果用粗体标出, 次佳效果用下划线标出.

#### 4.6 消融研究

为了研究我们的模型中不同模块的有效性, 我们对ASTE-Data-v2进行了消融研究. 从表4的实验结果表明, 本文设计的模块对模型性能的提升均有影响, 语法图模块对模型有一定影响, 因为该模块可以利用依赖解析器中的概率矩阵来表示单词之间的依赖关系概

率, 从而捕捉丰富的语法信息. 此外, 语义图模块对模型也有一定影响, 因为该模块可以通过注意分数来精确捕捉词与词之间的语义相关性, 提升了模型对语义关联性的理解. 其中移除相互仿射变换模块对模型性能影响最大, 这充分说明了移除该模块后语法信息和语义信息就不能很好进行交互, 因此, 在数据集上性能

会大幅下降. 消融实验结果示于表4中. 我们的模型在 Rest14, Lap14, Rest15 和 Rest16 上分别实现了  $F1$  值的 1.51%, 1.47%, 1.72% 和 1.76% 的明显绝对增加.

表4 ASTE-Data-v2 上的消融实验结果 (%)

Model	Rest14	Lap14	Rest15	Rest16
EMC-GCN*	71.65	57.77	61.53	67.68
-语法图模块	72.20	58.46	62.13	68.32
-语义图模块	72.86	58.64	62.31	68.85
-相互仿射变换模块	71.89	57.96	62.11	67.95
Ours	<b>73.16</b>	<b>59.24</b>	<b>63.25</b>	<b>69.44</b>

注: “\*”表示相同情况下原论文方法复现的结果, “-”表示将该模块从模型中去除, 其他部分和我们的模型保持一致, 最佳结果用粗体表示.

## 5 结论

在本文中, 我们提出了一种新的端到端模型, 用于进行 ASTE 任务, 该模型充分地融合了句子的语法结构和丰富的语义信息, 可以更好地理解复杂评论中方面词和意见词之间隐含的语义关系. 首先, 为提高依赖关系分析结果的准确性, 模型利用依赖解析器中的概率矩阵来表示单词之间的依赖关系概率, 从而捕捉丰富的语法信息. 此外, 该模型还设计了一个自注意机制模块, 该模块利用自注意力网络构造语义图的邻接矩阵, 从而表示词之间的语义相关性. 另外, 为了更好地结合语法结构和语义信息来增强三元组的提取任务, 模型提出了一种相互仿射变换模块. 这个模块通过多次交互有效地捕捉了句子的语法结构和语义信息. 在多个公开数据集上进行了广泛的实验, 本文提出的模型在 ASTE 任务上始终优于所有基准模型. 这些实验证明了在 ASTE 任务中, 有效利用句子的语法结构和丰富的语义信息是一个有前途的方向.

## 参考文献

- Peng HY, Xu L, Bing LD, *et al.* Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 8600–8607. [doi: 10.1609/aaai.v34i05.6383]
- Xu L, Li H, Lu W, *et al.* Position-aware tagging for aspect sentiment triplet extraction. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. ACL, 2020. 2339–2349. [doi: 10.18653/v1/2020.emnlp-main.183]
- Wu Z, Ying CC, Zhao F, *et al.* Grid tagging scheme for

aspect-oriented fine-grained opinion extraction. Proceedings of the 2020 Findings of the Association for Computational Linguistics. ACL, 2020. 2576–2585. [doi: 10.18653/v1/2020.findings-emnlp.234]

- Xu L, Chia YK, Bing LD. Learning span-level interactions for aspect sentiment triplet extraction. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. ACL, 2021. 4755–4766. [doi: 10.18653/v1/2021.acl-long.367]
- Zhang C, Li QC, Song DW, *et al.* A multi-task learning framework for opinion triplet extraction. Proceedings of the 2020 Findings of the Association for Computational Linguistics. ACL, 2020. 819–828. [doi: 10.18653/v1/2020.findings-emnlp.72]
- Mao Y, Shen Y, Yu C, *et al.* A joint training dual-MRC framework for aspect based sentiment analysis. Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI, 2021. 13543–13551. [doi: 10.1609/aaai.v35i15.17597]
- Chen SW, Wang Y, Liu J, *et al.* Bidirectional machine reading comprehension for aspect sentiment triplet extraction. Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI, 2021. 12666–12674. [doi: 10.1609/aaai.v35i14.17500]
- Mukherjee R, Nayak T, Butala Y, *et al.* PASTE: A tagging-free decoding framework using pointer networks for aspect sentiment triplet extraction. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana: ACL, 2021. 9279–9291. [doi: 10.18653/v1/2021.emnlp-main.731]
- Zhang WX, Li X, Deng Y, *et al.* Towards generative aspect-based sentiment analysis. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. ACL, 2021. 504–510. [doi: 10.18653/v1/2021.acl-short.64]
- Fei H, Ren YF, Zhang Y, *et al.* Nonautoregressive encoder-decoder neural framework for end-to-end aspect-based sentiment triplet extraction. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(9): 5544–5556. [doi: 10.1109/TNNLS.2021.3129483]
- Chen ZX, Huang H, Liu B, *et al.* Semantic and syntactic enhanced aspect sentiment triplet extraction. Proceedings of the 2021 Findings of the Association for Computational Linguistics. ACL, 2021. 1474–1483. [doi: 10.18653/v1/2021.



- [findings-acl.128](#)]
- 12 Sun K, Zhang RC, Mensah S, *et al.* Aspect-level sentiment analysis via convolution over dependency tree. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong: ACL, 2019. 5679–5688. [doi: [10.18653/v1/D19-1569](https://doi.org/10.18653/v1/D19-1569)]
  - 13 Zhou J, Huang JX, Hu QV, *et al.* SK-GCN: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification. Knowledge-based Systems, 2020, 205: 106292. [doi: [10.1016/j.knosys.2020.106292](https://doi.org/10.1016/j.knosys.2020.106292)]
  - 14 Hu ZD, Wang ZX, Wang YL, *et al.* Aspect sentiment triplet extraction incorporating syntactic constituency parsing tree and commonsense knowledge graph. Cognitive Computation, 2023, 15(1): 337–347. [doi: [10.1007/s12559-022-10078-4](https://doi.org/10.1007/s12559-022-10078-4)]
  - 15 Shi LL, Han DH, Han JY, *et al.* Dependency graph enhanced interactive attention network for aspect sentiment triplet extraction. Neurocomputing, 2022, 507: 315–324. [doi: [10.1016/j.neucom.2022.07.067](https://doi.org/10.1016/j.neucom.2022.07.067)]
  - 16 Chen H, Zhai ZP, Feng FX, *et al.* Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin: ACL, 2022. 2974–2985. [doi: [10.18653/v1/2022.acl-long.212](https://doi.org/10.18653/v1/2022.acl-long.212)]
  - 17 Mrini K, Dernoncourt F, Tran QH, *et al.* Rethinking self-attention: Towards interpretability in neural parsing. Proceedings of the 2020 Findings of the Association for Computational Linguistics. ACL, 2020. 731–742. [doi: [10.18653/v1/2020.findings-emnlp.65](https://doi.org/10.18653/v1/2020.findings-emnlp.65)]
  - 18 Read J, Pfahringer B, Holmes G, *et al.* Classifier chains for multi-label classification. Machine Learning, 2011, 85(3): 333–359. [doi: [10.1007/s10994-011-5256-5](https://doi.org/10.1007/s10994-011-5256-5)]

(校对责编: 孙君艳)