



## 融合Transformer和交互注意力网络的方面级情感分类模型

程艳, 胡建生, 赵松华, 罗品, 邹海锋, 詹勇鑫, 富雁, 刘春雷

引用本文:

程艳, 胡建生, 赵松华, 罗品, 邹海锋, 詹勇鑫, 富雁, 刘春雷. 融合Transformer和交互注意力网络的方面级情感分类模型[J]. 智能系统学报, 2024, 19(3): 728-737.

CHENG Yan, HU Jiansheng, ZHAO Songhua, et al. Aspect-level sentiment classification model combining Transformer and interactive attention network[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(3): 728-737.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202303016>

## 您可能感兴趣的其他文章

### 基于双特征嵌套注意力的方面词情感分析算法

An algorithm for aspect-based sentiment analysis based on dual features attention-over-attention  
智能系统学报. 2021, 16(1): 142-151 <https://dx.doi.org/10.11992/tis.202012024>

### 混合神经网络和条件随机场相结合的文本情感分析

Text sentiment analysis combining hybrid neural network and conditional random field  
智能系统学报. 2021, 16(2): 202-209 <https://dx.doi.org/10.11992/tis.201907041>

### 面向推荐系统的分期序列自注意力网络

Recommendation system with long-term and short-term sequential self-attention network  
智能系统学报. 2021, 16(2): 353-361 <https://dx.doi.org/10.11992/tis.202005028>

### 层次化双注意力神经网络模型的情感分析研究

Hierarchical double-attention neural networks for sentiment classification  
智能系统学报. 2020, 15(3): 460-467 <https://dx.doi.org/10.11992/tis.201812017>

### 基于注意力融合的图片描述生成方法

An image caption generation method based on attention fusion  
智能系统学报. 2020, 15(4): 740-749 <https://dx.doi.org/10.11992/tis.201910039>

### 加入自注意力机制的BERT命名实体识别模型

BERT named entity recognition model with self-attention mechanism  
智能系统学报. 2020, 15(4): 772-779 <https://dx.doi.org/10.11992/tis.202003003>

DOI: 10.11992/tis.202303016

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20231110.0952.002>

# 融合 Transformer 和交互注意力网络的方面级情感分类模型

程艳<sup>1,2</sup>, 胡建生<sup>1</sup>, 赵松华<sup>1</sup>, 罗品<sup>1</sup>, 邹海锋<sup>1</sup>, 詹勇鑫<sup>1</sup>, 富雁<sup>3</sup>, 刘春雷<sup>4</sup>

(1. 江西师范大学计算机信息工程学院, 江西南昌 330022; 2. 智能信息处理与情感计算江西省重点实验室, 江西南昌 330022; 3. 江西软云科技股份有限公司, 江西南昌 330200; 4. 江西和壹科技有限公司, 江西南昌 330200)

**摘要:** 现有的大多数研究者使用循环神经网络与注意力机制相结合的方法进行方面级情感分类任务。然而, 循环神经网络不能并行计算, 并且模型在训练过程中会出现截断的反向传播、梯度消失和梯度爆炸等问题, 传统的注意力机制可能会给句子中重要情感词分配较低的注意力权重。针对上述问题, 该文提出了一种融合 Transformer 和交互注意力网络的方面级情感分类模型。首先利用 BERT(bidirectional encoder representation from Transformers) 预训练模型来构造词嵌入向量, 然后使用 Transformer 编码器对输入的句子进行并行编码, 接着使用上下文动态掩码和上下文动态权重机制来关注与特定方面词有重要语义关系的局部上下文信息。最后在 5 个英文数据集和 4 个中文评论数据集上的实验结果表明, 该文所提模型在准确率和  $F_1$  上均表现最优。

**关键词:** 方面词; 情感分类; 循环神经网络; Transformer; 交互注意力网络; BERT; 局部特征; 深度学习

**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1673-4785(2024)03-0728-10

中文引用格式: 程艳, 胡建生, 赵松华, 等. 融合 Transformer 和交互注意力网络的方面级情感分类模型 [J]. 智能系统学报, 2024, 19(3): 728-737.

英文引用格式: CHENG Yan, HU Jiansheng, ZHAO Songhua, et al. Aspect-level sentiment classification model combining Transformer and interactive attention network[J]. CAAI transactions on intelligent systems, 2024, 19(3): 728-737.

## Aspect-level sentiment classification model combining Transformer and interactive attention network

CHENG Yan<sup>1,2</sup>, HU Jiansheng<sup>1</sup>, ZHAO Songhua<sup>1</sup>, LUO Pin<sup>1</sup>, ZOU Haifeng<sup>1</sup>,  
ZHAN Yongxin<sup>1</sup>, FU Yan<sup>3</sup>, LIU Chunlei<sup>4</sup>

(1. School of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022, China; 2. Key Laboratory of Intelligent Information Processing and Emotional Computing in Jiangxi Province, Nanchang 330022, China; 3. Jiangxi Ruanyun Technology Corporation Limited, Nanchang 330200, China; 4. Jiangxi Heyi Technology Co., Ltd., Nanchang 330200, China)

**Abstract:** At present, most researchers use a combination of recurrent neural networks and attention mechanisms for aspect-level sentiment classification tasks. However, the recurrent neural network cannot be computed in parallel, and the models encounter problems, such as truncated backpropagation, gradient vanishing, and gradient explosion, in the training process. Traditional attention mechanisms may assign reduced attention weights to important sentiment words in sentences. An aspect-level sentiment classification model combining Transformer and interactive attention network is proposed to solve these problems. In this approach, the pretrained model, which considers bidirectional encoder representation from Transformers (BERT), is initially used to construct word embedding vectors. Then, Transformer encoders are used to perform parallel encoding for input sentences. Subsequently, the contextual dynamic mask-off code and the contextual dynamic weighting mechanisms are applied to focus on local context information semantically relevant to specific aspect words. Finally, the model is tested on five English datasets and four Chinese review datasets. Experimental results demonstrate that the proposed model outperforms others in terms of accuracy and  $F_1$ .

**Keywords:** aspect term; sentiment classification; recurrent neural network; transformer; interactive attention network; BERT; local feature; deep learning

收稿日期: 2023-03-08. 网络出版日期: 2023-11-13.

**基金项目:** 国家自然科学基金项目 (62167006, 61967011); 江西省科技创新基地——智能信息处理与情感计算江西省重点实验室 (原江西省智能教育重点实验室) 项目 (20212BCD42001); 江西省 03 专项及 5G 项目 (20212ABC03A22); 江西省主要学科学术和技术带头人培养计划——领军人才项目 (20213BCJL22047); 江西省自然科学基金项目 (20212BAB202017).

**通信作者:** 程艳. E-mail: [chyan8888@jxnu.edu.cn](mailto:chyan8888@jxnu.edu.cn).

随着互联网的快速发展, 越来越多的用户在网上购物、学习、社交等, 并发表和分享个人言论。当用户对某一实体发表主观意见时, 往往会以不同的情感极性来评价某一实体的不同方面。方面级情感分析是一项细粒度情感分析任务, 包

括方面词提取和方面级情感分类<sup>[1]</sup>。本文研究的是方面级情感分类任务,旨在判断评论语句 $S$ 中某一特定方面词 $A$ 的情感极性(如积极、中性和消极)<sup>[2]</sup>。例如,以评论文本“lots of extra space but the keyboard is ridiculously small”为例,space 和 keyboard 为两个方面词,对应的情感类别为积极和消极。

方面级情感分类任务的研究基于情感词典和机器学习方法开展。Deng 等<sup>[3]</sup>构建了情感词汇,通过层次化监督主题模型来判断句子的情感极性。Pang 等<sup>[4]</sup>使用支持向量机(support vector machines, SVM)来确定情感分类任务中的文本表示。Jiang 等<sup>[5]</sup>提出一种结合特征工程来确定情感分类的统计方法。上述方法需要依赖大量的人工干预,语义理解能力较弱,很难突破性能瓶颈。

近年来,基于神经网络的方法,如循环神经网络(recurrent neural network, RNN)、长短期记忆网络(long short-term memory, LSTM)和门控循环单元(gated recurrent unit, GRU)等深度学习技术在方面级情感分类任务中均表现出优越性。Tang 等<sup>[6]</sup>设计了 TD-LSTM 和 TC-LSTM 两种经过改进的模型,使用两个 LSTM 分别提取前上下文和后上下文情感语义特征。Ma 等<sup>[7]</sup>提出了一种扩展的 LSTM 模型,使用两阶段注意力机制来构建方面词与上下文之间的关联程度。Gu 等<sup>[8]</sup>提出位置感知的双向门控循环单元(bi-directional gated recurrent unit, Bi-GRU)进行方面级情感分类。但是 RNN 对每个句子进行编码时,无法建模长序列关系,难以进行并行计算,需要花费较长的训练时间。Tang 等<sup>[9]</sup>使用深度记忆网络(memory networks, MemNet)来关注上下文对特定方面词的重要程度。Xue 等<sup>[10]</sup>利用卷积神经网络(convolution neural network, CNN)来解决并行计算问题。虽然 CNN 可以有效地减少模型的训练时间,但它无法捕获句子中的长距离依赖关系。Vaswani 等<sup>[11]</sup>提出了一种基于注意力机制的 Transformer 模型,该模型能够对输入序列进行并行编码,在提高分类性能的同时减少了模型训练时间。

注意力机制能够对句子中每个单词分配不同的注意力权重。Chen 等<sup>[12]</sup>提出了一种基于记忆机制的递归注意力(recurrent attention on memory, RAM)模型,使用 GRU 实现对注意力的多层抽象。Wang 等<sup>[13]</sup>提出了一种 ATAE-LSTM 网络,使用 LSTM 来表示上下文和方面词的隐藏状态,然后通过注意力机制来关注句子中重要的情感词。Ma 等<sup>[14]</sup>提出了一种交互注意力网络(inter-

active attention networks, IAN)对方面词与上下文进行建模。Song 等<sup>[15]</sup>使用注意力编码器网络(attentional encoder network, AEN)来表示目标词与上下文之间的隐藏状态和语义交互。Xu 等<sup>[16]</sup>使用多注意力网络进行方面级情感分类任务。肖宇晗等<sup>[17]</sup>提出了一种基于双特征嵌套注意力的方面词情感分析算法,有效地捕捉了方面词与上下文之间的语义关联。曾义夫等<sup>[18]</sup>提出了一种双记忆注意力网络,该网络首先使用 GRU 来表示句子的隐层状态,然后通过注意力机制从一个复杂的句子中提取出某一特定方面词的情感极性。虽然上述模型在方面级情感分类任务中取得不错效果,但是它们在捕获一个句子中方面词与上下文语义交互时忽略了语法依赖关系。例如,在评论“the cuisine from what i have gathered is without a doubt the most delicious”中,“cuisine”为方面词,模型会重点关注“without a doubt”短语,因为它与方面词“cuisine”的距离较近,而上下文“delicious”会被分配较低的注意力权重,因为它与方面词“cuisine”的距离较远,从而得到情感极性为 negative,使得模型预测错误。

针对上述问题,本文提出了一种融合 Transformer 和交互注意力网络的方面级情感分类模型(aspect-level sentiment classification model combining Transformer and interactive attention network, TIAN)。首先使用 BERT 预训练模型来初始化词嵌入向量,利用 Transformer 编码器获取输入向量的隐藏状态;然后为了加强方面词与上下文之间的语义交互,使用依存句法树中两个单词之间的最短路径作为语法相对距离,根据语法相对距离使用上下文动态掩码(context dynamic mask, CDM)和上下文动态权重(context dynamic weighting, CDW)来关注与特定方面词有重要语义关系的局部上下文信息。本文使用的 CDM 和 CDW 机制与文献[19]不同,文献[19]简单地将两个单词之间的单词计数作为语义相对距离(syntactic relative distance, SRD),而忽略了单词之间句法关系。本文结合了依赖解析树中两个单词之间的最短距离和句子序列中单词距离作为最终的 SRD,通过 SRD 掩码和权重衰减与方面词相关度较小的上下文,从而得到局部上下文表示。为丰富方面词信息,将局部和全局特征进行融合操作,通过多头自注意力机制和池化操作映射到全连接层和 softmax 分类器;最后输出方面级情感分类结果。

## 1 本文模型

方面级情感分类任务定义为: 给定一个长度为  $n$  的句子  $S = \{w_1, w_2, \dots, w_i, \dots, w_{i+m-1}, \dots, w_n\}$  和长度

为  $m$  的方面词  $A = \{w_i, w_{i+1}, \dots, w_{i+m-1}\}$ , 方面级情感分类主要目的是判断方面词  $A$  在句子  $S$  中的情感极性。TIAN 模型框架包括以下 4 个部分(图 1): 词嵌入层、Transformer 编码层、交互注意力层和输出层。

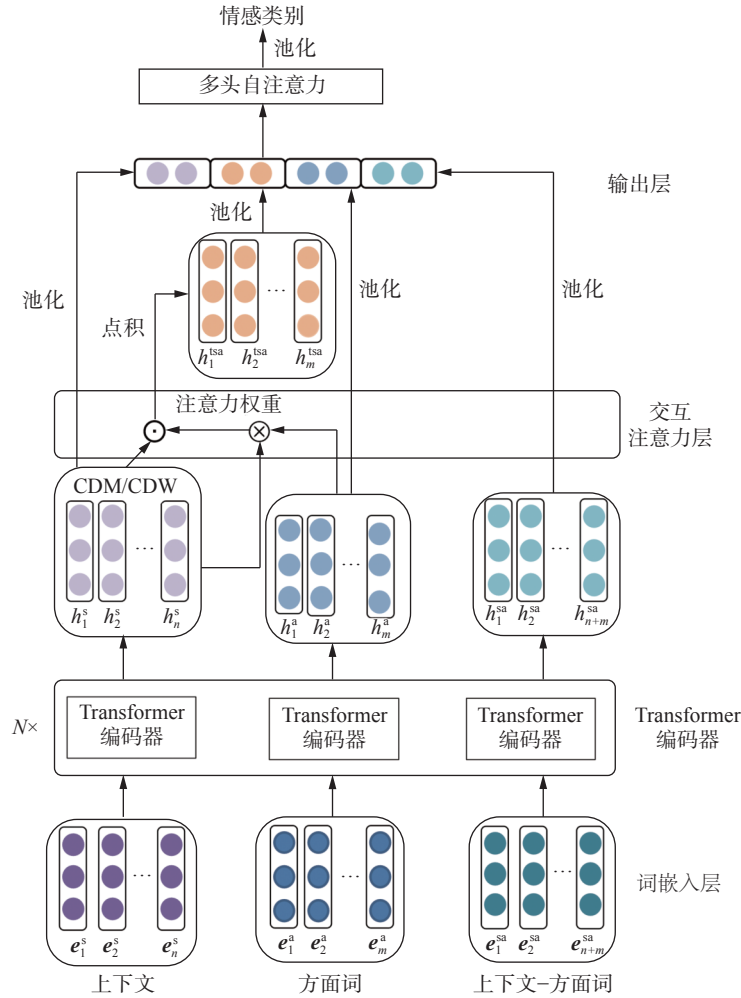


图 1 TIAN 模型结构

Fig. 1 Structure of TIAN model

### 1.1 词嵌入层

词嵌入是将句子中的每个单词  $w_i$  映射到一个低维、连续、稠密的向量。BERT 不仅能克服一词多义的关键性问题, 而且还能充分考虑单词上下文语境。因此, 本文使用 BERT 预训练模型来初始化单词的向量表示。为了更好地使 BERT 进行训练和微调, 本文将句子转换成以下形式作为模型的输入, 即“[CLS]+上下文+[SEP]”“[CLS]+方面词+[SEP]”和“[CLS]+上下文+[SEP]+方面词+[SEP]”结构, 其中 [CLS] 表示句子输入的开始, [SEP] 用于分隔上下文与方面词。经过词嵌入层处理后, 分别得到上下文词向量  $E^s = \{e_1^s, e_2^s, \dots, e_n^s\} \in \mathbf{R}^{n \times d_w}$ , 方面词向量  $E^a = \{e_1^a, e_2^a, \dots, e_m^a\} \in \mathbf{R}^{m \times d_w}$  和全局特征向量  $E^{sa} = \{e_1^{sa}, e_2^{sa}, \dots, e_{n+m}^{sa}\} \in \mathbf{R}^{(n+m) \times d_w}$ , 其中  $d_w$  表示词向量的维度,  $n$  表示上下文长度,

$m$  表示方面词长度。

### 1.2 Transformer 编码层

Transformer 编码器主要由两部分组成: 多头注意力机制和前馈网络。每一部分之后连接残差网络和层归一化模块, 图 2 为 Transformer 编码器结构<sup>[11]</sup>。

多头注意力机制是通过多个注意力机制进行堆叠操作来得到句子中每个单词的注意力权重。注意力机制输入是 3 个维度相同的矩阵  $Q = [q_1 \ q_2 \ \dots \ q_n]$ 、 $K = [k_1 \ k_2 \ \dots \ k_n]$  和  $V = [v_1 \ v_2 \ \dots \ v_n]$ , 其中  $q_i \in \mathbf{R}^{d_h}$ ,  $k_i \in \mathbf{R}^{d_h}$ ,  $v_i \in \mathbf{R}^{d_h}$ ,  $d_h$  表示隐藏层的维度。通过查询序列  $Q$ 、键序列  $K$  和值序列  $V$ , 可以得到注意力值为

$$A(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

其中 $\sqrt{d_k}$ 表示缩放因子, 主要目的是为了防止反向传播过程中梯度消失。

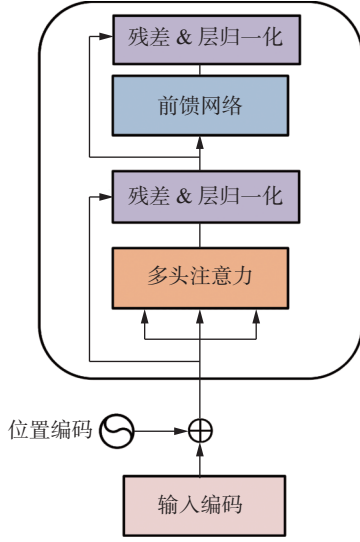


图 2 Transformer 编码器结构

Fig. 2 Structure of Transformer encoder

多头注意力机制是从不同的表示子空间中获  
得文本的语义特征向量, 然后将多个注意力头得  
到的结果拼接并线性变换:

$$h_i = A(W_i^Q Q, W_i^K K, W_i^V V) \quad (2)$$

$$M = C(h_1, h_2, \dots, h_h) W^O \quad (3)$$

其中:  $W_i^Q \in \mathbf{R}^{d_q \times d_h}$ ,  $W_i^K \in \mathbf{R}^{d_k \times d_h}$ ,  $W_i^V \in \mathbf{R}^{d_v \times d_h}$  和  $W^O \in \mathbf{R}^{d_o \times d_h}$   
为权重矩阵。

输入向量  $E^s$ 、 $E^a$ 、 $E^{sa}$  与式 (3) 进行残差网络和  
层归一化运算, 以  $E^s$  为例, 计算如下:

$$X^s = L(E^s + M(E^s)) \quad (4)$$

前馈网络对多头注意力机制获得的隐层表示

进行转换, 假设输入序列为  $H$ , 其计算为

$$F(H) = \max(0, HW_1 + b_1) W_2 + b_2 \quad (5)$$

其中:  $W_1$  和  $W_2$  表示权重矩阵,  $b_1$  和  $b_2$  表示偏置值。  
最后通过残差网络和层归一化计算得到  $H^s$ 、 $H^a$  和  
 $H^{sa}$  隐层表示, 以上下文隐层表示  $H^s$  为例, 计算  
式为

$$H^s = L(X^s + F(X^s)) \quad (6)$$

### 1.3 交互注意力层

上下文特征向量  $H^s$  未考虑方面信息, 为了加  
强方面词与上下文之间的情感语义关系, 使用交  
互注意力网络进行学习。本文首先利用依赖解析  
树来计算上下文与方面词之间 SRD, 然后使用 CDM  
和 CDW 两种机制分别掩盖和削弱与方面词无关  
的上下文权重。本节使用变量  $S$  来表示上下文  
与方面词之间 SRD, 使用  $S_{seq}()$  和  $S_{dep}()$  来计算  
SRD 的序列距离和依存距离。序列距离是指一个  
句子中两个单词之间的实际距离, 如图 3 所示,  
 $S_{seq}(\text{the service, slow}) = 2$ 。依存距离是利用依赖解  
析树来计算单词对应节点之间的最短距离, 当方  
面词是短语时, 则上下文与方面词之间的 SRD 值  
等于方面词中的每个单词与上下文单词之间的平  
均距离, 如  $S_{dep}(\text{the, slow}) = 3$ ,  $S_{dep}(\text{service, slow}) = 2$ ,  
方面词 the service 与上下文 slow 之间的依存距离  
为  $S_{dep}(\text{theservice, slow}) = 2.5$ 。因此, 本文提出了一种新的 SRD 计算方法, 即将 SRD 的序列距离和依  
存距离进行综合考虑, 计算公式为  $S = \min(S_{seq}, S_{dep})$ 。  
由此, 方面词 the service 与上下文 slow 之间的最  
终 SRD 计算结果为  $S = \min(S_{seq}, S_{dep}) = 2$ 。

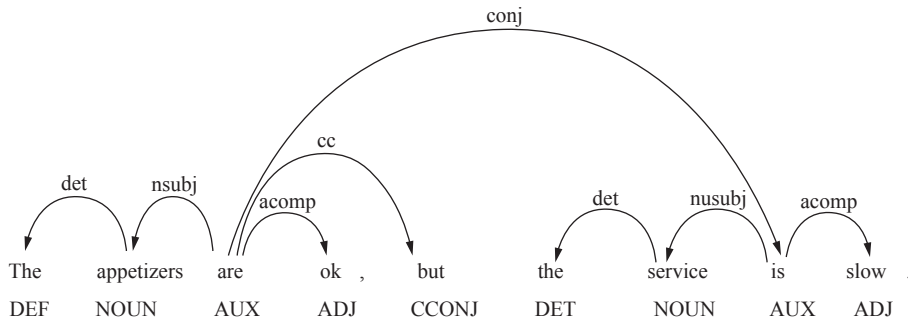


图 3 Restaurant 训练数据集中一个例句的依存句法树实例

Fig. 3 Example diagram of a dependency syntax tree for an example sentence in the Restaurant training dataset

CDM 表示掩盖了方面词的 SRD 大于预定义  
阈值  $\alpha$  的低语义上下文特征。根据上一层得到的  
上下文隐藏状态  $H^s$ , 基于 SRD 和阈值  $\alpha$  得到掩码  
向量  $V_i^m$  和新的上下文隐藏状态  $H^{sm}$ :

$$v_i^m = \begin{cases} O, & S_i > \alpha \\ E, & S_i \leq \alpha \end{cases} \quad (7)$$

$$M = C(v_1^m, v_2^m, \dots, v_n^m) \quad (8)$$

$$H^{sm} = H^s \odot M \quad (9)$$

CDW 表示完全保留小于等于阈值  $\alpha$  的上下  
文特征, 通过权重衰减来削弱大于阈值  $\alpha$  的上下  
文特征:

$$v_i^w = \begin{cases} \left(1 - \frac{S_i - \alpha}{N}\right) \cdot E, & S_i > \alpha \\ E, & S_i \leq \alpha \end{cases} \quad (10)$$

$$\mathbf{W} = \text{Concate}(\mathbf{v}_1^w, \mathbf{v}_2^w, \dots, \mathbf{v}_n^w) \quad (11)$$

$$\mathbf{H}^{\text{sw}} = \mathbf{H}^s \odot \mathbf{W} \quad (12)$$

其中:  $\mathbf{O}$  和  $\mathbf{E}$  分别表示 0 和 1 向量,  $\mathbf{O} \in \mathbf{R}^h$ ,  $\mathbf{E} \in \mathbf{R}^h$ ;  $h$  表示上下文隐藏向量  $\mathbf{h}_i^s$  的维度;  $\odot$  表示矩阵逐元素乘法运算;  $\mathbf{H}^{\text{sw}} = \{\mathbf{h}_1^{\text{sw}}, \mathbf{h}_2^{\text{sw}}, \dots, \mathbf{h}_n^{\text{sw}}\}$ 。

在得到上下文隐层表示  $\mathbf{H}^{\text{sm}}$  和  $\mathbf{H}^{\text{sw}}$  后, 使用注意力机制来捕获上下文与方面词之间的交互情感特征。当使用 CDM 机制来获取局部上下文隐层表示时, 如下式所示:

$$f(\mathbf{h}_j^{\text{sm}}, \mathbf{h}_k^a) = \tanh\left(\left(\mathbf{h}_j^{\text{sm}}\right)^T \cdot \mathbf{W} \cdot \mathbf{h}_k^a + \mathbf{b}\right) \quad (13)$$

$$I_{kj} = \frac{\exp\left(f(\mathbf{h}_j^{\text{sm}}, \mathbf{h}_k^a)\right)}{\sum_{i=1}^n \exp\left(f(\mathbf{h}_i^{\text{sm}}, \mathbf{h}_k^a)\right)} \quad (14)$$

$$\mathbf{h}_k^{\text{tsa}} = \sum_{j=1}^n I_{kj} \cdot \mathbf{h}_j^{\text{sm}} \quad (15)$$

同理, 当使用 CDW 机制获取局部上下文隐层表示时, 如下式所示:

$$f(\mathbf{h}_j^{\text{sw}}, \mathbf{h}_k^a) = \tanh\left(\left(\mathbf{h}_j^{\text{sw}}\right)^T \cdot \mathbf{W} \cdot \mathbf{h}_k^a + \mathbf{b}\right) \quad (16)$$

$$I_{kj} = \frac{\exp\left(f(\mathbf{h}_j^{\text{sw}}, \mathbf{h}_k^a)\right)}{\sum_{i=1}^n \exp\left(f(\mathbf{h}_i^{\text{sw}}, \mathbf{h}_k^a)\right)} \quad (17)$$

$$\mathbf{h}_k^{\text{tsa}} = \sum_{j=1}^n I_{kj} \cdot \mathbf{h}_j^{\text{sw}} \quad (18)$$

其中:  $I_{kj} \in \mathbf{R}^{m \times n}$  是由方面词隐藏状态  $\mathbf{h}_k^a$  与第  $j$  个上下文隐藏状态  $\mathbf{h}_j^{\text{sm}}$  计算而得到的注意力权重得分,  $\mathbf{h}_k^{\text{tsa}} \in \mathbf{R}^{m \times d_h}$  是隐藏表示的加权和,  $\tanh()$  是一个非线性激活函数,  $\mathbf{W} \in \mathbf{R}^{d_h \times d_h}$  和  $\mathbf{b} \in \mathbf{R}^{d_h}$  分别表示权重矩阵和偏置。

#### 1.4 输出层

通过平均池化得到新的隐藏表示, 然后再融合局部特征和全局特征。当使用 CDM 机制时, 其计算方法为

$$\mathbf{H}_{\text{avg}}^{\text{sm}} = \frac{1}{n} \sum_{j=1}^n \mathbf{h}_j^{\text{sm}} \quad (19)$$

$$\mathbf{H}_{\text{avg}}^{\text{tsa}} = \frac{1}{m} \sum_{i=1}^m \mathbf{h}_i^{\text{tsa}} \quad (20)$$

$$\mathbf{H}_{\text{avg}}^a = \frac{1}{m} \sum_{k=1}^m \mathbf{h}_k^a \quad (21)$$

$$\mathbf{H}_{\text{avg}}^{\text{sa}} = \frac{1}{n+m} \sum_{p=1}^{n+m} \mathbf{h}_p^{\text{sa}} \quad (22)$$

$$\mathbf{O}^{\text{sa}} = \mathbf{W}^{\text{sa}} \cdot \left[ \mathbf{H}_{\text{avg}}^{\text{sm}} \quad \mathbf{H}_{\text{avg}}^{\text{tsa}} \quad \mathbf{H}_{\text{avg}}^a \quad \mathbf{H}_{\text{avg}}^{\text{sa}} \right] + \mathbf{b}^{\text{sa}} \quad (23)$$

同理, 当使用 CDW 机制时, 只需将式 (19) 中  $\mathbf{H}_{\text{avg}}^{\text{sm}}$  和  $\mathbf{h}_j^{\text{sm}}$  分别变换为  $\mathbf{H}_{\text{avg}}^{\text{sw}}$  和  $\mathbf{h}_j^{\text{sw}}$ , 其他公式保持不变。其中,  $\mathbf{W}^{\text{sa}} \in \mathbf{R}^{d_h \times 4d_h}$  和  $\mathbf{b}^{\text{sa}} \in \mathbf{R}^{d_h}$  分别表示权重矩阵和偏置值。通过多头自注意力机制得到融合后的特征:

$$\mathbf{O} = \mathbf{M}(\mathbf{O}^{\text{sa}}, \mathbf{O}^{\text{sa}}, \mathbf{O}^{\text{sa}}) \quad (24)$$

然后, 对  $\mathbf{O}$  进行池化操作:

$$\mathbf{x} = \text{POOL}(\mathbf{O}) \quad (25)$$

将  $\mathbf{x}$  通过全连接层映射到维度为  $C$  的向量空间中:

$$\mathbf{x}' = \mathbf{W} \cdot \mathbf{x} + \mathbf{b} \quad (26)$$

其中  $\mathbf{W} \in \mathbf{R}^{C \times d_h}$ 、 $\mathbf{b} \in \mathbf{R}^C$  分别表示权值矩阵和偏置值。通过 softmax 归一化计算, 最后得到方面级情感极性概率  $\mathbf{y} \in \mathbf{R}^C$ :

$$\mathbf{y} = \text{softmax}(\mathbf{x}') = \frac{\exp(\mathbf{x}')}{\sum_{i=1}^C \exp(\mathbf{x}')} \quad (27)$$

其中  $C$  表示情感类别数。

#### 1.5 损失函数

本文所提出的模型在有监督的数据上以端到端的学习方式进行训练, 训练的目的是优化所有参数, 从而尽可能最小化损失函数。损失函数由交叉熵损失函数和  $L_2$  正则化项两部分组成:

$$L(\theta) = - \sum_{i=1}^C \mathbf{y}_i \log(\hat{\mathbf{y}}_i) + \lambda \sum_{\theta \in \Theta} \theta^2 \quad (28)$$

其中:  $\hat{\mathbf{y}}_i$  表示某个方面词模型预测得到的情感极性;  $\mathbf{y}_i \in \mathbf{R}^C$  是某个方面词真实的情感极性, 它用 one-hot 向量来进行表示;  $\Theta$  表示模型参数集合,  $\lambda$  表示  $L_2$  正则化系数。

## 2 实验及结果分析

### 2.1 实验数据集和参数设置

为了验证模型的有效性, 本文首先在 5 个主流的英文数据集上进行实验, 即 SemEval 2014 的任务 4 (restaurant 和 laptop)、SemEval 2015 的任务 12 (restaurant)、SemEval 2016 的任务 5 (restaurant) 和 Twitter 数据集。数据集中每一条评论文本包括句子、方面词和情感极性, 情感类别为积极、中性和消极。SemEval 2014 的任务 4 中 Laptop 和 Restaurant 包括积极、中性、消极和冲突 4 个情感极性, 由于冲突类别的评论语句较少, 因此本文借鉴了前人的数据预处理的方法, 删除了冲突类别的评论文本, 5 个英文数据集的详细信息如表 1 所示。

表 1 英文数据集信息  
Table 1 Information of English datasets

数据集	积极		中性		消极		总计	
	训练集	测试集	训练集	测试集	训练集	测试集	训练集	测试集
Restaurant2014	2 164	728	637	196	807	196	3 608	1 120
Laptop2014	994	341	464	169	870	128	2 330	638
Restaurant2015	1 178	439	50	35	382	328	1 610	802
Restaurant2016	1 620	597	88	38	709	190	2 417	825
Twitter	1 561	173	3 127	346	1 560	173	6 248	692

为了评估 TIAN 模型对中文信息的处理能力,本文在 4 个中文评论数据集上也进行了实验,分别为 Phone、Car、Notebook 和 Camera 数据集。中文评论数据集由 Peng 等<sup>[20]</sup>在文献[21]的基础

上处理得到,评论文本的情感极性为二分类,分别是积极和消极。数据集中存在个别评论语句方面词缺失和情感极性标记缺失等问题,通过处理后,得到表 2 所示的 4 个中文评论数据集信息。

表 2 中文评论数据集信息  
Table 2 Information of Chinese review datasets

数据集	积极		消极		总计	
	训练集	测试集	训练集	测试集	训练集	测试集
Phone	1 319	341	668	156	1 987	497
Car	708	164	213	66	921	230
Notebook	328	88	168	35	496	123
Camera	1 197	322	546	113	1 743	435

在实验过程中,基于 BERT 模型使用 BERT<sub>base</sub> 预训练模型来初始化单词的嵌入向量,在基于非 BERT 的基线模型中,英文数据集使用 GloVe 来获得词嵌入,中文数据集使用中文预训练字向量 (Chinese word vectors)<sup>[22]</sup>来得到词嵌入。基于 BERT 模型词嵌入的维度设置为 768,学习率为  $2 \times 10^{-5}$ ; GloVe 预训练词向量的维度设置为 300,学习率为 0.001。所有权重矩阵使用均匀分布  $U(-\varepsilon, \varepsilon)$  进行随机初始化,其中  $\varepsilon$  大小设置为 0.01,所有偏置设置为 0,英文和中文数据集的句子最大长度分别设置为 85 和 50,模型训练过程中使用 Adam 优化器来优化所有参数, batch\_size 为 16,使用 dropout(dropout=0.1)和  $L_2(1 \times 10^{-5})$  正则化项来防止模型过拟合。本文使用准确率 (Accuracy, Acc) 和  $F_1$  指标来评估模型的性能。

## 2.2 对比模型

为了评估 TIAN 性能,选取了方面级情感分类任务中近几年具有代表性的模型进行了实验对比。本文对比的基线模型分为两组:基于非 BERT 模型和基于 BERT 模型。本文对比模型如下:

基于非 BERT 的基线模型:

1)ATAE-LSTM<sup>[13]</sup>:在输入层中充分考虑了方面嵌入,并使用注意力机制的 LSTM 来得到情感

分类结果。

2)IAN<sup>[14]</sup>:使用注意力机制学习上下文与方面词之间的语义交互。

3)MemNet<sup>[9]</sup>:使用记忆网络将上下文作为外部记忆,并利用多层注意力机制得到特定方面词对上下文的注意力权重。

4)RAM<sup>[12]</sup>:采用双向 LSTM 和循环注意力机制来获得文本特征,同时构建了一种位置加权记忆模块来捕获远距离情感信息。

5)MGAN<sup>[23]</sup>:基于 IAN 和 AOA 两种模型,在损失函数上增加了一种方面对齐损失,有效地提高了具有相同上下文中各方面词之间的相互影响。

6)TNet-LF<sup>[24]</sup>:由 CNN、LSTM 和 CPT 模块组成,CPT 模块包含特定方面词表示和上下文保存机制。

7)MAN<sup>[16]</sup>:该模型通过全局和局部注意力机制来捕获上下文和方面词之间的语义交互。

基于 BERT 的基线模型:

1)BERT-SPC<sup>[15]</sup>:将序列“[CLS]+上下文+[SEP]+方面词+[SEP]”输入到 BERT 预训练模型中,用于句子对分类任务。

2)AEN-BERT<sup>[15]</sup>:使用注意力编码网络来建模上下文和方面词之间的语义交互,利用标签平

滑正则化项来解决数据集中标签不可靠性问题。

3) LCF<sup>[19]</sup>: 该模型使用句子序列中单词距离作为语义相对距离, 通过语义相对距离来丢弃局部上下文中与方面词无关的情感词。

4) MGMD<sup>[25]</sup>: 该模型是一种基于 BERT 的多源数据融合的网络, 从多个资源库中学习情感词倾向, 同时还设计了一种特殊的架构集成来自多层次情感词典的数据。

5) RMN-BERT<sup>[26]</sup>: 该模型使用图卷积神经网络来生成方面表示, 并利用双注意力机制捕获方面词与上下文之间的语义关系。

6) TIAN\_M/TIAN\_W: 本文提出的融合 Trans-

former 和交互注意力网络的方面级情感分类模型, TIAN\_M 和 TIAN\_W 分别表示 TIAN 使用 CDM 和 CDW 机制。

### 2.3 实验结果与分析

本文提出的 TIAN 模型在 5 个英文基准数据集上与上述 13 个基线模型进行实验对比, 实验结果如表 3 所示。表 3 标有\*的基线模型表示直接引用了原论文中的实验结果, “—”表示原论文在该数据集上未取得实验结果。表 3 中其他基线模型的实验结果, 均根据原论文作者发布的代码进行了实验复现, 表中加粗的字体表示最好的实验结果。

表 3 各模型在英文数据集上的实验结果

Table 3 Experimental results of different models on English datasets

模型		Restaurant2014		Laptop2014		Restaurant2015		Restaurant2016		Twitter	
		Acc	$F_1$	Acc	$F_1$	Acc	$F_1$	Acc	$F_1$	Acc	$F_1$
Non BERT	ATAE-LSTM <sup>*</sup>	77.20	—	68.70	—	—	—	—	—	—	—
	IAN <sup>*</sup>	78.60	—	72.10	—	—	—	—	—	—	—
	MemNet <sup>*</sup>	80.95	—	72.71	—	—	—	—	—	—	—
	RAM	80.61	69.79	73.86	70.93	80.78	68.89	84.37	70.50	69.39	67.35
	MGAN	81.23	71.90	75.39	72.47	80.78	68.94	85.28	72.53	72.56	70.78
	TNet-LF	80.75	70.80	76.21	71.49	81.47	68.79	85.03	71.49	74.55	73.31
	MAN <sup>*</sup>	84.38	71.31	78.13	73.20	82.65	69.10	85.87	73.28	76.56	72.19
BERT	BERT-SPC	84.36	76.91	78.56	75.23	83.29	68.98	86.47	74.58	73.55	72.14
	AEN-BERT	83.46	74.20	79.82	76.27	83.80	69.19	88.63	76.75	74.68	73.11
	LCF-CDM	85.71	78.82	79.96	76.15	84.34	71.08	90.01	77.86	75.66	74.21
	LCF-CDW	85.94	79.28	80.23	76.34	84.91	69.23	90.66	78.29	<b>76.73</b>	74.72
	MGMD <sup>*</sup>	86.16	78.83	79.18	74.86	—	—	—	—	—	—
	RMN-BERT <sup>*</sup>	84.56	79.05	77.95	70.83	82.94	66.95	89.38	71.88	—	—
	TIAN_M	86.81	80.71	80.57	76.83	85.79	<b>72.57</b>	91.40	78.46	75.72	74.38
	TIAN_W	<b>87.23</b>	<b>81.93</b>	<b>80.92</b>	<b>77.21</b>	<b>86.15</b>	72.13	<b>92.21</b>	<b>79.39</b>	76.68	<b>74.75</b>

由表 3 可知, TIAN\_M 和 TIAN\_W 在 5 个英文数据集上取得最优结果。TIAN\_W 比 TIAN\_M 优越, 原因是 CDW 把与方面词较远上下文隐藏表示的权重只是被降低了, 而 CDM 则是把与方面词较远的上下文的权重被完全掩盖为 0。在基于 RNN 的模型中, ATAE-LSTM、IAN、MemNet、RAM 都使用了注意力机制, 在各个数据集上的实验结果都有所提升, 表明 RNN 与注意力机制结合, 对于优化模型性能具有重要意义。MGAN 和 MAN 取得较优的分类效果, 验证了交互注意力机制的有效性。BERT-SPC 和 AEN-BERT 在 5 个英文数据集上均优于使用非 BERT 的基线模型, 表明使用 BERT 预训练模型能够更好地对方面词与上下文进行向量表示。TIAN\_M 和 TIAN\_W 在除 Twitter 数据集上的准确率外, 在其他数据集

上均优于 LCF-CDM 和 LCF-CDW, 表明了语义相对距离对掩码和衰减与方面词相关度较小的上下文具有重要的影响。RMN-BERT 对上下文进行位置编码时简单的将两个单词之间的单词计数作为语义相对距离, 相比 RMN-BERT, 本文提出的方法在 4 个英文数据集上的准确率提升约 3%, 表明了通过依赖解析树中两个单词之间的最短路径作为语法相对距离能更好地捕获与方面词相关的上下文特征, 验证了本文模型的有效性。

为了验证 TIAN 对中文信息的处理能力, 本文在 4 个中文评论数据集上进行实验, 实验结果如表 4 所示。从表 4 可以看出, TIAN 性能最优。TIAN\_M 和 TIAN\_W 在大多数中文数据集上的实验效果基本持平, 主要原因是中文数据集上的评

论文本长度较短,并且这些评论语句大多数都较为简单,此时当 CDM 和 CDW 通过 SRD 阈值  $\alpha$  来掩盖和削弱与方面词无关的上下文权重时,句子中的大多数上下文与方面词之间的距离都在 SRD 阈值  $\alpha$  范围内,无论使用 CDW 还是 CDM,上下文被分配较低和完全掩盖为 0 的隐藏表示权重的

情况较少,大部分上下文隐藏表示的权重都是为 1 的向量。TIAN 通过交互注意力机制不仅能捕获方面与上下文之间的语义关系,而且还能着重关注方面词信息,使用 CDM 和 CDW 机制能够根据句法相对距离来减少远离目标方面词的不相关上下文的负面影响。

表 4 各模型在中文数据集上的实验结果

Table 4 Experimental results of different models on Chinese datasets

%

模型		Phone		Car		Notebook		Camera	
		Acc	$F_1$	Acc	$F_1$	Acc	$F_1$	Acc	$F_1$
Non BERT	ATAE-LSTM	86.44	82.63	81.40	71.50	83.36	82.15	85.29	81.83
	IAN	89.74	88.14	85.96	81.58	84.52	79.08	88.75	85.87
	MemNet	89.89	88.39	84.91	79.54	86.45	82.72	84.95	81.24
	RAM	90.51	88.83	86.67	81.27	86.57	83.22	89.10	86.85
	MGAN	91.35	89.90	85.61	81.37	87.10	83.91	89.62	87.13
	TNet-LF	91.88	90.69	87.37	82.79	87.24	83.97	90.83	88.53
BERT	BERT-SPC	96.02	95.34	96.49	95.44	94.84	93.82	95.67	94.87
	AEN-BERT	96.78	96.29	96.14	95.13	92.90	91.44	95.42	94.56
	TIAN_M	96.97	96.45	<b>98.60</b>	<b>98.22</b>	94.84	93.89	96.54	95.83
	TIAN_W	<b>97.12</b>	<b>96.64</b>	98.25	97.78	<b>95.48</b>	<b>94.56</b>	<b>96.89</b>	<b>96.29</b>

## 2.4 模型训练时间分析

为了更好地比较本文所提模型与基于 RNN 模型的训练时间, TIAN 这里使用 Glove 预训练模型来初始化单词嵌入向量。同时, 本文在相同的 GPU 和深度学习框架等环境下分析不同模型在 Laptop2014 数据集上完成一次迭代的训练时间, 实验结果如表 5 所示。

表 5 不同模型在 Laptop2014 数据集上一次迭代的训练时间

Table 5 Training time for different models to complete one epoch on the Laptop2014 dataset

模型	训练时间
ATAE-LSTM	284.35
RAM	257.64
MGAN	273.59
TNet-LF	361.08
MemNet	53.82
MAN	46.71
TIAN_M	44.52
TIAN_W	<b>43.64</b>

由表 5 可知, TIAN\_M 和 TIAN\_W 模型的训练时间最短。相反, 基于 LSTM 的基线模型, 模型训练时间较长。这主要是因为 Transformer 编码器中的注意力机制的时间复杂度为  $O(l_m^2 d_w)$ , 而 LSTM 的时间复杂度为  $O(l_m d_w^2)$ ,  $l_m$  表示输入序列的最大长度,  $d_w$  表示词向量的维度<sup>[27]</sup>, 通常情况下,  $l_m$  比  $d_w$  更小。本文在实验过程中,  $l_m$  设置为 85,

$d_w$  设置为 300。Transformer 模型在训练文本时是并行处理的, 而 RNN 模型对当前的输入需要建立在上一时刻的输出上, 模型训练时非常缓慢, 其循环结构增加了模型的训练难度。

## 2.5 消融实验

为了验证 TIAN 模型中各个模块的合理性, 本节在 5 个英文数据集上进行消融实验, 实验结果如表 6 所示, “w/o”表示“未使用”。TIAN(M) 和 TIAN(F) 分别表示 Transformer 编码层中仅使用多头注意力机制和仅使用前馈神经网络。TIAN(L) 表示使用双向 LSTM 来代替 Transformer 模型。

由表 6 可知, 除了在某些特殊情况外, TIAN\_M 和 TIAN\_W 在 5 个英文数据集上均取得了最好的分类效果。TIAN w/o MW 实验结果略差, 表明 CDM 和 CDW 机制对 TIAN 模型的贡献程度较大, 这也验证了本文使用 CDM 和 CDW 机制的有效性。TIAN w/o Local 在 5 个数据集上的准确率下降 3.11%~4.59%, 表明局部上下文特征能够有效地提高模型识别方面的情感修饰词的能力。TIAN\_M w/o Global 和 TIAN\_W w/o Global 在 5 个数据集上的准确率分别下降 1.96%~4.38% 和 2.25%~5.14%, 表明 TIAN 只关注局部上下文特征会丢失句子中的重要信息, 全局上下文特征对方面情感预测也是不可或缺的。TIAN(M)\_W 在 Twitter 数据集上取得不错的分类效果, 这可能是因为 Twitter 数据集中的评论语句表达方式更适

合 TIAN(M)\_W 学习文本的语义特征。虽然 TIAN(L)\_W 在 Restaurant2014 数据集上的准确率

达到最佳,但是 TIAN(L)\_W 不能进行并行计算,模型需要花费较长的训练时间。

表 6 消融实验结果  
Table 6 Ablation experiment results

%

模型	Restaurant2014		Laptop2014		Restaurant2015		Restaurant2016		Twitter	
	Acc	$F_1$	Acc	$F_1$	Acc	$F_1$	Acc	$F_1$	Acc	$F_1$
TIAN(M)_M	85.46	78.13	78.83	74.47	84.21	68.35	88.39	75.06	74.09	72.67
TIAN(M)_W	85.65	78.74	78.98	74.82	84.76	68.54	88.42	75.41	<b>76.73</b>	74.67
TIAN(F)_M	85.62	78.80	77.75	74.55	84.37	67.37	87.92	74.39	73.81	72.94
TIAN(F)_W	86.30	78.92	78.19	74.71	84.50	68.46	89.20	76.23	73.95	73.12
TIAN(L)_M	86.91	79.68	79.47	75.84	83.95	68.15	88.57	75.56	73.37	72.53
TIAN(L)_W	<b>87.34</b>	79.89	79.94	76.23	85.02	72.04	88.96	76.18	73.51	72.61
TIAN w/o Local	83.67	76.75	77.81	74.04	82.49	68.20	87.62	74.21	72.94	71.93
TIAN_M w/o Global	84.85	77.23	78.22	74.20	82.81	68.27	87.02	73.98	73.25	72.22
TIAN_W w/o Global	84.98	77.42	78.35	74.48	83.02	68.60	87.07	74.06	73.32	72.29
TIAN w/o MW	85.11	77.59	78.56	74.80	83.21	68.83	88.04	74.86	74.68	73.31
TIAN_M	86.81	80.71	80.57	76.83	85.79	<b>72.57</b>	91.40	78.46	75.72	74.38
TIAN_W	87.23	<b>81.93</b>	<b>80.92</b>	<b>77.21</b>	<b>86.15</b>	72.13	<b>92.21</b>	<b>79.39</b>	76.68	<b>74.75</b>

### 3 结束语

针对方面级情感分类任务中存在模型训练时间较长、未有效地建立方面词与上下文交互信息的问题,本文提出了一种融合 Transformer 和交互注意力网络的方面级情感分类模型(TIAN)。该模型使用 BERT 预训练模型来初始化词嵌入,这种词嵌入方法比常用的 Word2vec 和 Glove 等方法效果更好。TIAN 使用 Transformer 编码器来表示文本的隐藏状态,在提高性能的同时,模型的训练时间明显减少,然后使用 CDM 和 CDW 机制来关注与特定方面词有重要语义关系的局部上下文信息。为了验证所提模型的有效性,在 5 个英文数据集和 4 个中文评论数据集上进行了广泛实验,TIAN 的实验结果优于目前主流的方面级情感分类模型。今后将考虑对 Transformer 编码器内部结构进行改进,如在编码器中嵌入 RNN、CNN、动态路由等多种结构。此外,我们也会考虑使用更大文本语料库的 RoBERTa 和 XLNet 等预训练模型来初始化方面词和上下文的嵌入向量,通过上下文与方面词之间的交互,学习到更广泛和更深入的语言知识,提供更丰富的文本表示能力,提高模型性能。

### 参考文献:

- [1] PONTIKI M, GALANIS D, PAVLOPOULOS J, et al. SemEval-2014 task 4: aspect based sentiment analysis[C]// Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin: Association for Computational Linguistics, 2014: 27–35.
- [2] LIU Bing. Sentiment analysis and opinion mining[J]. *Synthesis lectures on human language technologies*, 2012, 5(1): 1–167.
- [3] DENG Dong, JING Liping, YU Jian, et al. Sentiment lexicon construction with hierarchical supervision topic model[J]. *IEEE/ACM transactions on audio, speech, and language processing*, 2019, 27(4): 704–718.
- [4] PANG Bo, LEE L, VAITHYANATHAN S. Thumbs up?: sentiment classification using machine learning techniques[C]//Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - EMNLP '02. Morristown: Association for Computational Linguistics, 2002: 79–86.
- [5] JIANG Long, YU Mo, ZHOU Ming, et al. Target-dependent Twitter sentiment classification[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1. Portland: ACM, 2011: 151–160.
- [6] TANG D, QIN B, FENG X, et al. Effective LSTMs for Target-Dependent Sentiment Classification[C]//Proceedings of the 26th International Conference on Computational Linguistics. Stroudsburg: ACL, 2016: 3298–3307.
- [7] MA Yukun, PENG Haiyun, KHAN T, et al. Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis[J]. *Cognitive computation*, 2018, 10(4): 639–650.
- [8] GU S, ZHANG L, HOU Y, et al. A position-aware bidirectional attention network for aspect-level sentiment analysis[C]//Proceedings of the 27th International Conference on Computational Linguistics. Stroudsburg: ACL, 2018: 774–784.
- [9] TANG Duyu, QIN Bing, LIU Ting. Aspect level sentiment classification with deep memory network[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Associ-

- ation for Computational Linguistics, 2016: 214–224.
- [10] XUE Wei, LI Tao. Aspect based sentiment analysis with gated convolutional networks[EB/OL]. (2018–05–18) [2023–03–08]. <http://arxiv.org/abs/1805.07043>.
- [11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017: 6000–6010.
- [12] CHEN Peng, SUN Zhongqian, BING Lidong, et al. Recurrent attention network on memory for aspect sentiment analysis[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2017: 452–461.
- [13] WANG Yequan, HUANG Minlie, ZHU Xiaoyan, et al. Attention-based LSTM for aspect-level sentiment classification[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2016: 606–615.
- [14] MA Dehong, LI Sujian, ZHANG Xiaodong, et al. Interactive attention networks for aspect-level sentiment classification[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: ACM, 2017: 4068–4074.
- [15] SONG Youwei, WANG Jiahai, JIANG Tao, et al. Attentional encoder network for targeted sentiment classification[EB/OL]. (2019–02–09)[2023–03–08]. <http://arxiv.org/abs/1902.09314>.
- [16] XU Qiannan, ZHU Li, DAI Tao, et al. Aspect-based sentiment classification with multi-attention network[J]. *Neurocomputing*, 2020, 388(C): 135–143.
- [17] 肖宇晗, 林慧苹, 汪权彬, 等. 基于双特征嵌套注意力的方面词情感分析算法 [J]. *智能系统学报*, 2021, 16(1): 142–151.
- XIAO Yuhan, LIN Huiping, WANG Quanbin, et al. An algorithm for aspect-based sentiment analysis based on dual features attention-over-attention[J]. *CAAI transactions on intelligent systems*, 2021, 16(1): 142–151.
- [18] 曾义夫, 蓝天, 吴祖峰, 等. 基于双记忆注意力的方面级别情感分类模型 [J]. *计算机学报*, 2019, 42(8): 1845–1857.
- ZENG Yifu, LAN Tian, WU Zufeng, et al. Bi-memory based attention model for aspect level sentiment classification[J]. *Chinese journal of computers*, 2019, 42(8): 1845–1857.
- [19] ZENG Biqing, YANG Heng, XU Ruyang, et al. LCF: a local context focus mechanism for aspect-based sentiment classification[J]. *Applied sciences*, 2019, 9(16): 3389.
- [20] PENG Haiyun, MA Yukun, LI Yang, et al. Learning multi-grained aspect target sequence for Chinese sentiment analysis[J]. *Knowledge based systems*, 2018, 148: 167–176.
- [21] ZHAO Yanyan, QIN Bing, LIU Ting. Creating a fine-grained corpus for Chinese sentiment analysis[J]. *IEEE intelligent systems*, 2015, 30(1): 36–43.
- [22] LI Shen, ZHAO Zhe, HU Renfen, et al. Analogical reasoning on Chinese morphological and semantic relations[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Stroudsburg: Association for Computational Linguistics, 2018: 138–143.
- [23] FAN Feifan, FENG Yansong, ZHAO Dongyan. Multi-grained attention network for aspect-level sentiment classification[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2018: 3433–3442.
- [24] LI Xin, BING Lidong, LAM W, et al. Transformation networks for target-oriented sentiment classification[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2018: 946–956.
- [25] CHEN F, YUAN Z, HUANG Y. Multi-source data fusion for aspect-level sentiment classification[J]. *Knowledge-based systems*, 2020, 187: 104831.
- [26] ZENG Jiandian, LIU Tianyi, JIA Weijia, et al. Relation construction for aspect-level sentiment classification[J]. *Information sciences*, 2022, 586: 209–223.
- [27] VASWANI A, BENGIO S, BREVDO E, et al. Tensor2Tensor for neural machine translation[C]//Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track). Boston: Association for Machine Translation in the Americas, 2018: 193–199.

## 作者简介:



程艳, 教授, 博士生导师, 主要研究方向为人工智能、智能信息处理、情感分析。发表学术论文 70 余篇, 出版学术专著 2 部。E-mail: [chyan88888@jxnu.edu.cn](mailto:chyan88888@jxnu.edu.cn)。



胡建生, 硕士研究生, 主要研究方向为情感分析、深度学习。E-mail: [1875560437@qq.com](mailto:1875560437@qq.com)。



赵松华, 硕士研究生, 主要研究方向为知识追踪、教育数据挖掘。E-mail: [3140851935@qq.com](mailto:3140851935@qq.com)。