

基于局部上下文焦点机制和交谈注意力的方面级情感分析

林正超, 李弼程

引用本文

林正超, 李弼程. 基于局部上下文焦点机制和交谈注意力的方面级情感分析[J]. 计算机科学, 2023, 50(11A): 220900266-6.

LIN Zhengchao, LI Bicheng. Aspect-based Sentiment Analysis Based on Local Context Focus Mechanism and Talking-Head Attention [J]. Computer Science, 2023, 50(11A): 220900266-6.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[接诉即办智能派单业务调度算法研究](#)

Study on Scheduling Algorithm of Intelligent Order Dispatching

计算机科学, 2023, 50(11A): 230300029-7. <https://doi.org/10.11896/jsjcx.230300029>

[基于方面语义和门控过滤网络的方面级情感分析](#)

Aspect-based Sentiment Analysis Based on Aspect Semantic and Gated Filtering Network

计算机科学, 2023, 50(10): 193-202. <https://doi.org/10.11896/jsjcx.220900192>

[基于多粒度特征融合的新型图卷积网络用于方面级情感分析](#)

Novel Graph Convolutional Network Based on Multi-granularity Feature Fusion for Aspect-based Sentiment Analysis

计算机科学, 2023, 50(10): 80-87. <https://doi.org/10.11896/jsjcx.230600036>

[基于依赖模型的REST接口测试用例生成方法研究](#)

Study on REST API Test Case Generation Method Based on Dependency Model

计算机科学, 2023, 50(9): 101-107. <https://doi.org/10.11896/jsjcx.220800071>

[基于prompt和知识增强的方面级情感分析](#)

Aspect-based Sentiment Analysis Based on Prompt and Knowledge Enhancement

计算机科学, 2023, 50(6A): 220300279-7. <https://doi.org/10.11896/jsjcx.220300279>

基于局部上下文焦点机制和交谈注意力的方面级情感分析

林正超 李弼程

华侨大学计算机科学与技术学院 福建 厦门 361021

(1425721006@qq.com)

摘要 方面级情感分析是目前自然语言处理领域的一个重要研究方向,其目的是预测句子当中不同方面的情感极性。现有的方面级情感分析,通常忽略了情感极性与局部语境之间的关系,且在部分使用多头注意力机制的研究中,每个注意力头数的运算是相互独立的。为此,提出了一种基于局部上下文焦点机制和交谈注意力的方面级情感分析模型。首先,通过BERT预训练模型分别捕获局部上下文和全局上下文的初步特征。然后,在特征提取层,利用局部上下文焦点机制,通过上下文特征动态掩码层结合交谈注意力机制来进一步提取局部上下文特征;利用交谈注意力机制进一步提取全局上下文特征。最后,将局部和全局信息进行融合,输入非线性层获取情感分析结果。在3个公开数据集上进行了对比实验,实验结果表明,与现有的多个基线模型相比,所提模型的MF1值和准确率均有提升。

关键词: 方面级情感分析;BERT模型;局部上下文焦点机制;交谈注意力

中图法分类号 TP391

Aspect-based Sentiment Analysis Based on Local Context Focus Mechanism and Talking-Head Attention

LIN Zhengchao and LI Bicheng

College of Computer Science and Technology, Huaqiao University, Xiamen, Fujian 361021, China

Abstract Aspect-based sentiment analysis is an important research direction in the field of natural language processing, and its purpose is to predict the sentiment polarity of different aspects in sentences. The existing aspect-based sentiment analysis usually ignores the relationship between sentiment polarity and local context, and the operation of each attention head in the multi-head attention used is independent of each other. To this end, an aspect-based sentiment analysis model based on the local context focus mechanism and talking-head attention is proposed. First, preliminary features of local context and global context are captured by a BERT pretrained model. Then in the feature extraction layer, the local contextual focus mechanism is used, and the local contextual features are further extracted through the contextual feature dynamic mask layer combined with the talking-head attention, and talking-head attention is used to further extract global context features. Finally, the local and global information are fused and input to the nonlinear layer to obtain sentiment analysis results. Experiments are conducted on three public datasets. Experiments show that compared with multiple existing baseline models, the MF1 value and accuracy of the new model are improved.

Keywords Aspect-based sentiment analysis, BERT model, Local context focus mechanism, Talking-Head attention

1 引言

随着互联网的飞速发展,各种在线平台相继出现,从新闻、博客到论坛,互联网中用户的参与度越来越高,用户在网上浏览热点信息的方式,对发生的一些事件表达各自的观点和态度。与此同时,各类产品和娱乐方式也通过互联网的形式展现给用户,用户在购买和体验后,会对产品和服务发表大量表达自己观点的评论。这些带有用户观点的文本数据是非常重要的数据资源,分析这些带有用户观点的文本数据有着非常重要的意义。例如,商家可以通过分析这些数据获得用户对某种产品的喜好和产品存在的不足,从而找到更好的改进方向,提高产品的销量;突发事件发生后,通过分析人们对该事件所发表的评论数据,从而更好地把控舆论的走向;政府

出台新政策时,通过分析网民的观点判断提出的政策是否具有实效,从而进行调整。

随着社交媒体的不断进步,情感分析在自然语言处理领域具有重要的理论意义和应用价值^[1]。情感分析主要是对文本中带有不同情感表达进行情感分类。以往的大多数情感分析研究都是一种粗粒度分析,无法满足更加准确精细的分析需求。例如,对某一个产品来说,从不同的方面去分析该产品存在哪些优点和缺点等。方面级情感分析^[2]不同于以往的粗粒度研究,能够对句子中不同方面的情感极性进行分析,因此成为了情感分析领域中一个重要的研究方向^[3]。例如:“The price of this house is good, but the location is terrible”,这句针对房子的评论当中存在着两个方面词,分别是“price”和“location”,两者对应的观点词分别是“good”和“terrible”,

基金项目:装备预研教育部联合基金项目(8091B022150)

This work was supported by the Joint Fund Project of Ministry of Education for Equipment Pre-research(8091B022150).

通信作者:李弼程(lbclm@163.com)

其中“good”代表的是积极的情感，“terrible”代表的是消极的情感。在这种情况下，方面级情感分析能够更充分地捕获文中表达的不同方面的情感。由此可见，对于方面级情感分析的研究有着非常重要的研究意义和价值。

随着预训练模型的提出，以 BERT 为首的预训练模型受到了广泛关注，越来越多的研究人员将该模型用到方面级情感分析领域，这意味着 BERT 预训练模型对该领域的情感分析任务是可行的。以往大多数的方面级情感分析研究都是对句子中不同方面的情感极性进行预测，缺乏考虑情感极性和局部语境之间的关系。此外，大多数研究都是结合单个注意力或多头注意力机制进行的，但是多头注意力机制中的每个头的运算是相互独立的，为了能够在语言处理模型上取得更优的效果，我们将各个相互独立的头联系起来，得到了更强的注意力设计。为此，本文提出了基于局部上下文焦点机制和交谈注意力的方面级情感分析模型，简称 LCFTHA 模型。主要贡献如下：1) 利用局部上下文焦点机制，通过上下文特征动态掩码层(Context features Dynamic Mask, CDM)结合交谈注意力机制(Talking-Heads Attention, THA)来提取局部上下文特征。2) 考虑到全局上下文特征具有更加全面的方面词信息，使用交谈注意力机制提取全局上下文特征，并将局部和全局特征进行融合，从而进一步提升模型的具体表现。3) 结合 BERT 预训练模型，在 3 个方面级情感分析数据集上进行实验并跟目前的主流模型进行对比，结果表明本文提出的模型在性能上优于相关模型。

2 相关工作

情感分析又称为观点挖掘，根据 Liu^[4]的定义，情感分析的任务是提取 $\langle H, E, A, S, T \rangle$ (观点持有者, 目标实体, 目标方面, 情感, 发表时间) 观点元组。方面级情感分析是通过提取文本中的 $\langle A, S \rangle$ 元组，即提取目标方面和情感，常用于处理一些产品或服务的评价数据，从而分析用户从不同方面对产品或服务的评价。

早期的方面级情感分析主要通过构建情感词典来获取文档中情感词的情感值。构建情感词典的方法虽然简单，但是非常依赖情感词典的质量需求，因此存在情感词不够丰富的问题。并且情感词典大多都是通过人工进行构建，导致了工作量较大等问题。

随着相关研究的深入，基于深度学习的方面级情感分析^[5]得到了研究人员的广泛关注。Tang 等^[6]提出在方面级情感分析任务中引入长短期记忆网络(Long Short-Term Memory, LSTM)^[7]，提出了基于 LSTM 的 TD-LSTM 与 TC-LSTM 模型，虽然相较于原先的 LSTM 有着更好的效果，但是这类模型提取到的情感特征包含了非方面词的情感。为了更加充分地考虑方面信息的作用，Wang 等^[8]提出了一种基于注意力机制的 LSTM 模型，通过注意力机制^[9]来对句子中的重要部分进行关注，实验结果进一步表明方面词对情感判别的重要性；Ma 等^[10]提出了一种交互式注意力网络模型，以增强方面词和上下文信息之间的联系，首先对方面词和上下文进行建模，将方面词输入池化层得到方面信息的注意力，再利用注意力完成二者的交互操作，得到最后的情感特征；Fan 等^[11]提出了一种多粒度注意力网络 MGAN，因为以往的粗粒度注意力机制存在信息丢失的问题，所以 MGAN 通过

一种细粒度的注意力机制捕获方面词和上下文之间的交互信息，然后再与粗粒度注意力结合组成一种多粒度注意力网络框架来解决信息丢失的问题。

随着预训练模型的出现，许多研究人员提出使用预训练模型来进行方面级情感分析研究，通过微调，得到较好的情感分类结果，其中比较主流的预训练模型有 BERT, ALBERT 等。Sun 等^[12]通过构造辅助句的方法，构成“原句子和辅助句”的句子对，通过微调 BERT 预训练模型将方面级情感分析任务变成一种句子对的分类任务，取得了不错的情感分析结果；Song 等^[13]通过对 BERT 中间层的编码信息进行研究，弥补了以往大量的语义丢失，在内部的中间层都用注意力来进一步获取文本权重，提高了 BERT 在方面级情感分析任务中的有效性。现阶段的方面级情感分析的研究当中，大多数都是基于 BERT 预训练模型，原因是可以对 BERT 预训练模型进行微调，以此获得更好的情感分析结果。在接下来的研究当中，结合深度学习和预训练模型将成为方面级情感分析的主流方向。因此，本文充分考虑局部语境对情感极性的的重要性，利用局部上下文焦点机制结合交谈注意力提取局部信息，再与全局信息融合，结合 BERT 模型进一步提升模型效果。

3 LCFTHA 模型

本文提出的 LCFTHA 模型主要由 BERT 预训练层、特征提取层、特征学习层和输出层 4 个部分组成，如图 1 所示。

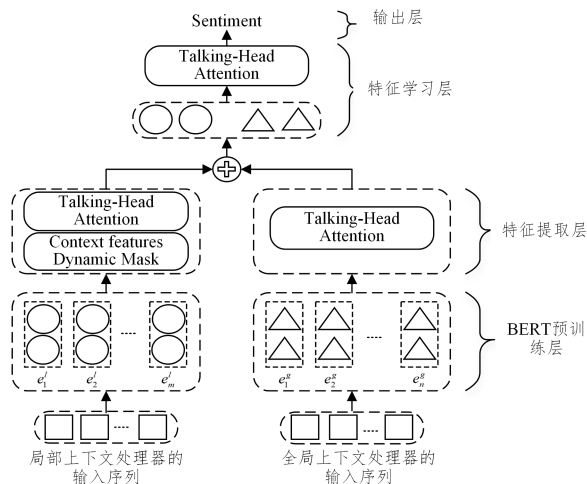


图 1 LCFTHA 模型

Fig. 1 LCFTHA model

3.1 BERT 预训练模型层

设计 BERT 预训练模型的目的是提升自然语言处理任务的性能。为了提升本文模型的性能，本文采用两个独立的 BERT 预训练模型作为下游任务来挖掘文本中的语义信息，分别对输入的文本中的词进行建模。BERT 预训练层是一个用于语言理解的预训练 Seq2Seq 模型，可以被视为嵌入层。BERT 预训练模型中有 [CLS] 与 [SEP] 两个特殊的标记，[CLS] 可以作为整个句子的语义表示。本文模型将每条样本预处理成“ $x=[CLS]+上下文+[SEP]$ ”和“ $x=[CLS]+上下文+[SEP]+方面词+[SEP]$ ”的形式，采用了两个独立的 BERT 预训练模型对输入的句子中的词进行建模，分别表示为 X^l 和 X^g ，得到初步的局部上下文特征和全局上下文特征的输出。

$$O_{\text{BERT}}^l = \text{BERT}^l(X^l) \quad (1)$$

$$O_{\text{BERT}}^g = \text{BERT}^g(X^g) \quad (2)$$

其中, O_{BERT}^l 和 O_{BERT}^g 分别是局部上下文和全局上下文处理器的输出表示, BERT^l 和 BERT^g 分别表示为局部上下文和全局上下文对应的 BERT 预训练模型建模。

3.2 特征提取层

在特征提取层中,主要采用了局部上下文焦点机制结合交谈注意力机制,来提取局部上下文特征。对于全局上下文特征,仅采用交谈注意力机制来提取。

3.2.1 局部上下文焦点机制

局部上下文焦点机制依赖于语义相对距离(Semantic Relative Distance, SRD),主要用来判断上下文是否属于目标方面的局部上下文,通过上下文特征动态掩码技术来帮助模型捕捉局部上下文特征。为了进一步提取局部上下文的深层语义特征,本文在上下文特征动态掩码层后结合使用交谈注意力机制。

1) 语义相对距离

语义相对距离是基于 Token-Aspect 对的概念,即句子中词的位置和方面词的位置,描述 Token 与 Aspect 之间的距离,可以理解成在二者之间相隔多少个词,即:

$$D_i = |i - F_a| - \left\lfloor \frac{n}{2} \right\rfloor \quad (3)$$

其中, i 表示词的位置, F_a 表示句子中方面词的位置, n 表示方面词的长度。因此 D_i 表示第 i 个词的位置和目标方面之间的距离。

2) 上下文特征动态掩码

除了局部上下文特征,上下文动态特征掩码层将屏蔽 BERT^l 层学习到的非局部上下文特征。虚线箭头指向的输出位置的特征将被屏蔽,实线箭头指向的输出位置的特征将被保留,如图 2 所示, POS 表示输出的位置。

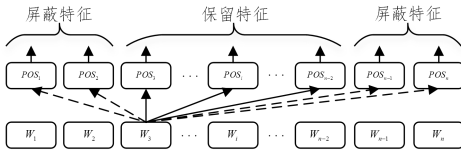


图 2 上下文特征动态掩码

Fig. 2 Context features dynamic mask

上下文特征动态掩码将非局部上下文的所有位置的特征设置为零向量。假设 O_{BERT}^l 是 BERT^l 的初始输出特征,可以得到如下的局部上下文特征 O_{CDM}^l :

$$V_i = \begin{cases} \mathbf{E}, & D_i \leq a \\ \mathbf{O}, & D_i > a \end{cases} \quad (4)$$

$$\mathbf{M} = [V_1, V_2, \dots, V_n] \quad (5)$$

$$O_{\text{CDM}}^l = O_{\text{BERT}}^l \cdot \mathbf{M} \quad (6)$$

其中, \mathbf{M} 是用于屏蔽非局部上下文特征的屏蔽矩阵, V_n 是输入序列中每个上下文词的掩码向量, a 是 SRD 阈值, n 是包括方面的输入序列的长度。其中与目标方面相关的 SRD 小于阈值 a 表示的是局部上下文, \mathbf{E} 是一个 1 向量, \mathbf{O} 是一个零向量(长度为 n)。

3.2.2 交谈注意力机制

在早期, Bahdanau 等^[14]通过研究,首次将注意力机制运用于机器翻译任务当中,这也代表了注意力机制在自然语言

处理领域取得了可观的效果。

2017 年,一些研究人员提出用注意力机制^[9]替代传统的循环神经网络来设计模型框架,并提出了一种多头注意力机制,通过多次缩放点积计算结果,并将其进行拼接,再通过一次线性变换就可以得到多头注意力的结果,即:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

$$Q^{(1)} = QW_Q^{(1)}, K^{(1)} = KW_K^{(1)}, V^{(1)} = VW_V^{(1)}, O^{(1)} = \text{Attention}(Q, K, V)$$

$$Q^{(2)} = QW_Q^{(2)}, K^{(2)} = KW_K^{(2)}, V^{(2)} = VW_V^{(2)}, O^{(2)} = \text{Attention}(Q, K, V)$$

⋮

$$Q^{(h)} = QW_Q^{(h)}, K^{(h)} = KW_K^{(h)}, V^{(h)} = VW_V^{(h)}, O^{(h)} = \text{Attention}(Q, K, V) \quad (8)$$

$$O_{\text{MHA}} = \text{concat}[\{O^{(1)}, O^{(2)}, \dots, O^{(h)}\} \cdot \mathbf{W}^{WH}] \quad (9)$$

其中, $W_Q^{(h)}, W_K^{(h)}, W_V^{(h)}$ 为参数矩阵, $W_Q \in R^{d_1 \times d_q}, W_K \in R^{d_1 \times d_k}, W_V \in R^{d_1 \times d_v}, \sqrt{d_k}$ 为缩放因子, h 为注意力头的个数, \mathbf{W}^{WH} 为一个权重矩阵。

但是在使用 Transformer 进行训练的过程中常常发现,多头注意力中注意力头的个数上升到一定程度其性能就不再上升,这是由于多头注意力存在低秩瓶颈(Low-Rank Bottleneck)^[15]的问题。并且当前多头注意力中的每个注意力头数的运算都是相互独立的,通过将各个独立的注意力头联系起来,就可以得到一种更强的注意力设计,进而提升模型的性能。通过在 softmax 操作前后引入各个注意力头之间的线性映射,来增强多个注意力机制之间的信息交流,即交谈注意力机制^[16],也就是将多个注意力头用一个参数矩阵重新进行融合,形成多个混合注意力,每一个新得到的混合注意力都融合了原先各个头的注意力,即:

$$\hat{J}^{(1)} = \frac{Q^{(1)} K^{(1)T}}{\sqrt{d_k}}, \hat{J}^{(2)} = \frac{Q^{(2)} K^{(2)T}}{\sqrt{d_k}}, \dots, \hat{J}^{(h)} = \frac{Q^{(h)} K^{(h)T}}{\sqrt{d_k}} \quad (10)$$

$$\begin{pmatrix} J^{(1)} \\ J^{(2)} \\ \vdots \\ J^{(h)} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1h} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{h1} & \lambda_{h2} & \dots & \lambda_{hh} \end{pmatrix} \begin{pmatrix} \hat{J}^{(1)} \\ \hat{J}^{(2)} \\ \vdots \\ \hat{J}^{(h)} \end{pmatrix} \quad (11)$$

$$P^{(1)} = \text{softmax}(J^{(1)}), P^{(2)} = \text{softmax}(J^{(2)}), \dots, P^{(h)} = \text{softmax}(J^{(h)}) \quad (12)$$

$$O^{(1)} = P^{(1)} V^{(1)}, O^{(2)} = P^{(2)} V^{(2)}, \dots, O^{(h)} = P^{(h)} V^{(h)} \quad (13)$$

$$O_{\text{THA}} = \text{concat}[\{O^{(1)}, O^{(2)}, \dots, O^{(h)}\} \cdot \mathbf{W}^{WH}] \quad (14)$$

其中, $J^{(h)}$ 表示在 softmax 操作前引入各个注意力头之间的线性映射, $O^{(h)}$ 表示 $\text{Attention}(Q^{(h)}, K^{(h)}, V^{(h)})$, $\sqrt{d_k}$ 为缩放因子, h 为注意力头个数, \mathbf{W}^{WH} 为一个权重矩阵。

为了避免 CDM 操作后特征分布的不均衡性,利用 THA 编码器学习并重新平衡被屏蔽的局部上下文特征。将 CDM 层学习到的局部特征放入 THA 层中,对于局部上下文焦点机制,就得到了局部上下文特征 O^l :

$$O^l = \text{THA}(O_{\text{CDM}}^l) \quad (15)$$

对于全局上下文特征 O^g ,直接通过 THA 编码器学习到的特征输出。

$$O^g = \text{THA}(O_{\text{BERT}}^g) \quad (16)$$

3.3 特征学习层

特征学习层主要用于交互式学习全局和局部上下文特征,通过将输出的局部上下文特征 O^l 和全局上下文特征 O^g 进行拼接得到 O^{lg} ,为了获得拼接向量的特征,对 O_{dense}^{lg} 进行THA编码,即:

$$O^{lg} = [O^l; O^g] \quad (17)$$

$$O_{dense}^{lg} = \mathbf{W}^{lg} \cdot O^{lg} + \mathbf{b}^{lg} \quad (18)$$

$$O_{FLL}^{lg} = THA(O_{dense}^{lg}) \quad (19)$$

其中, \mathbf{W}^{lg} 表示权重系数矩阵, $\mathbf{W}^{lg} \in R^{dn \times 2d_h}$, \mathbf{b}^{lg} 为偏置向量, $\mathbf{b}^{lg} \in R^{d_h}$, O_{dense}^{lg} 为一个全连接层,进行矩阵运算, O_{FLL}^{lg} 为特征学习层的输出。

3.4 输出层

本文模型采用softmax函数来预测最终方面词情感极性,即:

$$\hat{y} = \text{softmax}(\mathbf{W}_o O + \mathbf{b}_o) \quad (20)$$

其中, \mathbf{W}_o 为权重矩阵, \mathbf{b}_o 为偏置向量, \hat{y} 为预测的情感值。

损失函数采用带有正则化的交叉熵损失函数,引入正则化可以有效防止模型出现过拟合的情况,即:

$$L(\theta) = - \sum_1^C y_i \log y_i + \lambda \sum_{\theta \in \Theta} \theta^2 \quad (21)$$

其中, C 表示情感类别数, \hat{y}_i 为softmax函数预测得到的情感极性, y_i 为真实的情感值, λ 表示 L_2 正则化系数。

4 实验

4.1 数据集

在实验方面使用SemEval-2014Task4^[17]中的Restaurant14和Laptop14两个数据集,以及Twitter数据集^[18],来验证本文模型的有效性。在这3个公开数据集集中的评论信息当中,每个句子中的方面词都对应着3种不同的情感极性,数据细节如表1所列。

表1 数据集分布

Table 1 Data set distribution

数据集名称	积极(+1)	中性(0)	消极(-1)
Restaurant14 训练集	2164	637	807
Restaurant14 测试集	728	196	196
Laptop14 训练集	994	464	870
Laptop14 测试集	341	169	128
Twitter 训练集	1561	3127	1560
Twitter 测试集	173	346	173

4.2 基线模型

本文选择以下7个基线模型进行对比实验,来验证LCFTHA模型的有效性。

1)D-LSTM:该模型于2016年由Tang等^[6]提出,充分考虑对一个目标词的上下文信息,使用两个双向的LSTM,对目标词的信息建模,将两个LSTM得到的隐状态进行连接,输入非线性层进行情感分析。

2)RAM:该模型于2017年由Chen等^[19]提出,为了解决情感特征存在距离相隔较远的问题,采用多个注意力机制来对其进行捕获,预测情感极性,对无关信息具有更强的鲁棒性。

3)MGAN:该模型于2018年由Fan等^[11]提出,用一种细粒度的注意力机制捕获方面词和上下文之间的交互信息,再与粗粒度注意力相结合,组成一种多粒度注意力网络框架,用于最终的情感极性判断。

4)AEN-BERT:该模型于2019年由Song等^[13]提出,通过一个基于注意力的编码器对上下文和方面词进行建模,以此解决RNN无法并行化的问题,并在损失函数中引入了标签正则化,最后通过非线性层进行分类。

5)SPC-BERT:该模型于2019年由Song等^[13]提出,以句子对的形式输入到BERT预训练模型中,再通过非线性层来进行情感分类,取得了不错的效果。

6)LCF-BERT:该模型于2019年由Zeng等^[20]提出,其基本思想是通过关注方面词附近词的重要性,结合使用交谈注意力机制,再输入到非线性层来进行情感分类任务。

7)DREGCN-BERT:该模型于2020年由Liang等^[21]提出,采用了端到端ABSA多任务学习的依存句法知识增强交互体系结构,通过图卷积网络充分利用句法知识。

4.3 实验环境与参数配置

基于Pytorch深度学习框架搭建实验环境,具体配置如表2所列。

表2 实验环境

Table 2 Experimental environment

实验环境	环境配置
操作系统	Windows10 专业版 64位
CPU	Intel(R) Core(TM) i7-8700K
内存/GB	16
GPU	NVIDIA GeForce RTX 2080 Ti
开发工具	PyCharm

LCFTHA模型的部分超参数设置如表3所列。

表3 参数配置

Table 3 Parameter configuration

参数	值
优化器	Adam
BERT层学习率	0.00002
迭代次数	20
批量处理大小	32
句子最大长度	85
交谈注意力头数	12
L2正则化权值	0.01
BERT模型	bert-base-uncased

4.4 实验评价指标

本文采用了方面级情感分析任务中两个常用的评价指标,来验证模型的效果。第一个评价指标是准确率Accuracy,表示在任务中正确预测样本占所有样本的比例,预测正确的可能是正样本或负样本,即:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (22)$$

其中, TP 表示预测正确的正样本, TN 表示预测正确的负样本, FP 表示预测错误的正样本, FN 表示预测错误的负样本。

第二个评价指标是MF1,通常用来表示不同情感类别情况下模型的平均表现,即:

$$\begin{cases} P = \frac{TP}{TP + FP} \\ R = \frac{TP}{TP + FN} \\ MF1 = \frac{1}{|C|} \sum_i^C \frac{2P \times R}{P + R} \end{cases} \quad (23)$$

其中, P 表示情感类别中的精确率, R 表示情感类别中的召回率, C 表示情感类别的总数。

4.5 实验结果与分析

4.5.1 实验结果

通过在 Restaurant14, Laptop14 和 Twitter 3 个公开数据集上进行实验,对比 7 个基线模型,结果如表 4 所列。

表 4 模型准确率与宏平均

Table 4 Model accuracy and macro average
(单位:%)

模型	Restaurant14	Laptop14	Twitter
	Acc/MF1	Acc/MF1	Acc/MF1
TD-LSTM	77.77/64.93	71.78/68.43	70.98/68.50
RAM	80.23/70.80	74.49/71.35	69.36/67.30
MGAN	81.25/71.94	75.39/72.47	72.54/70.81
ANE-BERT	82.05/73.76	79.93/76.31	73.71/72.13
BERT-SPC	83.84/72.34	78.99/75.03	73.55/72.14
LCF-BERT-CDM	84.46/78.36	79.78/76.07	74.42/73.19
DREGCN-BERT	83.61/75.79	79.37/76.37	—
LCFTHA	86.52/80.69	81.19/77.57	75.87/74.65

观察表 4 中的实验结果可知,在 3 个公开数据集 Restaurant14, Laptop14 和 Twitter 上,本文提出的 LCFTHA 模型在准确率和 MF1 值上均优于基线模型,其准确率分别提升了 2.06%, 1.26%, 1.45%, MF1 分别提升了 2.33%, 1.2%, 1.46%。其中 ANE-BERT, BERT-SPC, LCF-BERT-CDM 3 个模型在相同情况下按照原论文方法复现模型。各模型在 Restaurant14 和 Laptop14 数据集上的效果均优于 Twitter 数据集,这是由于前两个数据集的评论者在撰写评论的过程中均使用了较为简单的语法,而 Twitter 数据集中一些方面词对应的实体大不相同,并且使用了一些网络流行用语及反讽语句,造成了分类的准确率较低。并且在前两个公开数据集中,积极的情感极性占比较大,因此在各个模型中都有较好的情感分类结果。

在上述 7 个基线模型中,TD-LSTM 模型的实验效果较差,这是由于 TD-LSTM 模型忽略了方面词与上下文词之间的语义关系,并且该模型未使用注意力机制,导致无法确认句子中词对方面的重要性。RAM 和 MGAN 两个模型的整体效果也不太理想,原因在于 GloVe 词嵌入只把一个词表示成一个词向量,无法区分一个词在不同语境下的不同含义,从而影响了模型最终的分类效果。与其他基于 BERT 预训练模型的方面级情感分析模型相比,本文提出的 LCFTHA 模型取得了更好的效果。BERT-SPC 模型以句子对的形式,将上下文词和句子中的方面词通过 BERT 预训练模型进行情感分类,ANE-BERT 对方面词与上下文的建模帮助模型更好地捕捉两者的内在联系,但二者均未考虑情感极性与局部语境之间的关系,因此本文模型利用局部上下文焦点机制,充分考虑情感极性与局部语境之间的关系。LCF-BERT 主要通过关注方面词附近词的重要性,再通过非线性层来进行情感分类任务,但模型中使用的多头注意力机制存在低秩瓶颈的问题,为此本文模型提出使用交谈注意力机制,具体地,将各个独立的注意力头联系起来,就可以得到一种更强的注意力设计,进而提升模型的性能。DREGCN-BERT 主要是通过从句法关系图上使用图神经网络,并使用 BERT 预训练模型来增强分类的性能,但得到的效果相较于本文模型仍然有待提高。实验结果进一步证明了本文提出的基于局部上下文焦点机制和交谈注意力的方面级情感分析模型在方面级情感分析任务中能够取得更好的效果。

4.5.2 参数选择实验

在模型 LCFTHA 中,语义相对距离是影响局部上下文

特征提取的重要因素,因此本文通过实验分析了在 3 个数据集上语义相对距离对准确率的影响。从图 3 和图 4 可以看出,当语义相对距离的阈值为 2 时,Restaurant14 数据集的准确率和 MF1 值的效果较好,准确率达到 86.52%, MF1 达到 80.69%。对于 Laptop14 数据集,当语义相对距离的阈值为 7 时,准确率和 MF1 值的效果均较好,准确率达到 81.19%, MF1 值达到 77.57%。对于 Twitter 数据集,语义相对距离的阈值与 Laptop14 数据集相同,当其阈值为 7 时,准确率和 MF1 值的效果均较好,准确率达到 75.87%, MF1 值达到 74.65%。

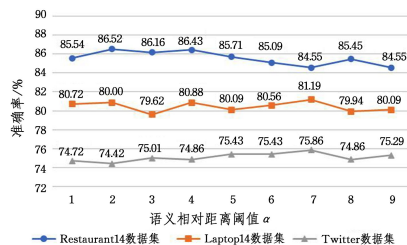


图 3 语义相对距离对准确率的影响

Fig. 3 Effect of semantic relative distance on accuracy

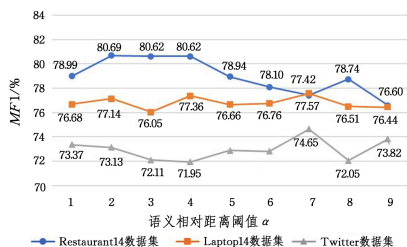


图 4 语义相对距离对 MF1 值的影响

Fig. 4 Effect of semantic relative distance on MF1

4.5.3 消融实验

本文在 3 个公开数据集上进行了消融实验,以此来验证 LCFTHA 模型中各个模块的重要性。其中 w/o 为 without 的缩写, FLL 为特征学习层的缩写,结果如表 5 所列。

表 5 消融实验结果

Table 5 Ablation experiment results
(单位:%)

模型组件	Restaurant14	Laptop14	Twitter
	Acc/MF1	Acc/MF1	Acc/MF1
仅全局上下文特征	85.27/77.70	79.47/76.11	74.57/73.00
仅局部上下文特征 (CDM+THA)	85.09/78.27	79.47/75.78	74.57/72.96
全局+局部 (CDM+THA) w/o FLL	85.71/79.25	79.62/76.65	75.58/74.24
LCFTHA	86.52/80.69	81.19/77.57	75.87/74.65

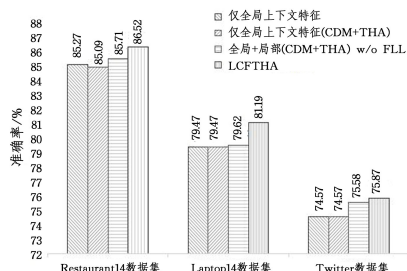


图 5 各部分模型组件对准确率的影响

Fig. 5 Effect of each model component on accuracy

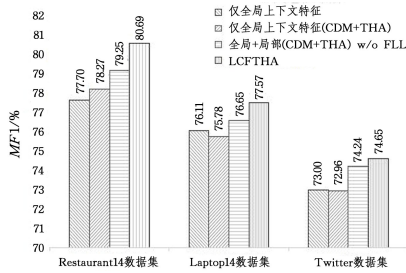


图6 各部分模型组件对 MF1 的影响

Fig. 6 Effect of each model component on MF1

观察表 5 可以看出,将全局和局部两个模块组件结合后,其效果比起单一的局部或全局模块更优,这充分说明二者之间的相互补充,可以提供方面词更加准确且完整的上下文表示。在此基础上,加上 FLL 可以明显看出模型的性能进一步提高,实验结果表明 FLL 层对该模型的设计具有重要意义。通过柱形图来更加直观地展示各个模型组件的重要性,如图 5 和图 6 所示。

结束语 本文通过研究方面级情感分析,提出了一种基于局部上下文焦点机制和交谈注意力机制的方面级情感分析模型 LCFTHA。首先,使用两个 BERT 预训练模型来分别提取局部和全局上下文特征;然后,将经过 BERT 提取到的初步局部特征再通过局部上下文焦点机制中的 CDM 层结合交谈注意力机制进一步捕获局部上下文特征,针对全局上下文特征部署一个交谈注意力编码器来学习全局上下文特征;最后,在特征学习层将局部和全局上下文特征进行融合后,输入到非线性层进行情感分析。在 Restaurant14, Laptop14 和 Twitter 3 个公开数据集上的实验证明,LCFTHA 模型的性能优于其他方面级情感分析任务中的基线模型,但是在一些使用讽刺语句或网络流行用语的情感表达场景下,效果有待提高。接下来将考虑如何对此进行进一步改进。

参考文献

- [1] ZHANG L, WANG S, LIU B. Deep learning for sentiment analysis: A survey[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8(4): e1253.
- [2] LV Y, WEI F, CAO L, et al. Aspect-level sentiment analysis using context and aspect memory network[J]. Neurocomputing, 2021, 428: 195-205.
- [3] ZHOU J, HUANG J X, CHEN Q, et al. Deep Learning for Aspect-Level Sentiment Classification: Survey, Vision and Challenges[J]. IEEE Access, 2019, 7: 78454-78483.
- [4] LIU B. Sentiment analysis: Mining opinions, sentiments, and emotions [J]. Computational Linguistics, 2016, 42(3): 1-4.
- [5] RUZ G A, HENRÍQUEZ P A, MASCARENO A. Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers[J]. Future Generation Computer Systems, 2020, 106: 92-104.
- [6] TANG D, QIN B, FENG X, et al. Effective lstms for target-dependent sentiment classification[C]// Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016.
- [7] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [8] WANG Y, HUANG M, ZHU X, et al. Attention-based LSTM

for aspect-level sentiment classification[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016: 606-615.

- [9] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems, 2017.
- [10] MA D, LI S, ZHANG X, et al. Interactive attention networks for aspect-level sentiment classification[C]// Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017.
- [11] FAN F, FENG Y, ZHAO D. Multi-grained attention network for aspect-level sentiment classification[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018: 3433-3442.
- [12] SUN C, HUANG L, QIU X. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence[J]. arXiv: 1903.09588, 2019.
- [13] SONG Y, WANG J, JIANG T, et al. Attentional encoder network for targeted sentiment classification [J]. arXiv: 1902.09314, 2019.
- [14] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv: 1409.0473, 2014.
- [15] BHOJANAPALLI S, YUN C, RAWAT A S, et al. Low-Rank Bottleneck in Multi-head Attention Models [J]. arXiv: 2002.07028, 2020.
- [16] SHAZEER N, LAN Z, CHENG Y, et al. Talking-Heads Attention[J]. arXiv: 2003.02436, 2020.
- [17] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2016 task 5: Aspect based sentiment analysis[C]// International Workshop on Semantic Evaluation, 2016: 19-30.
- [18] 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 (volume 2: Short papers). 2014: 49-54.
- [19] CHEN P, SUN Z, BING L, et al. Recurrent attention network on memory for aspect sentiment analysis[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017: 452-461.
- [20] ZENG B, YANG H, XU R, et al. Lcf: A local context focus mechanism for aspect-based sentiment classification[J]. Applied Sciences, 2019, 9(16): 3389.
- [21] LIANG Y, MENG F, ZHANG J, et al. A dependency syntactic knowledge augmented interactive architecture for end-to-end aspect-based sentiment analysis[J]. Neurocomputing, 2021, 454: 291-302.



LIN Zhengchao, born in 1997, postgraduate. His main research interests include natural language processing, intelligent data management and analysis, and deep learning.



LI Bicheng, born in 1970, Ph.D, professor, Ph.D supervisor. His main research interests include artificial intelligence, natural language processing, online public opinion monitoring and guidance.