面向方面级情感分析的交互图注意力网络模型

韩 虎[①②] 吴渊航*[①] 秦晓雅[①]

[①](兰州交通大学电子与信息工程学院 兰州 730070)

[②](甘肃省人工智能与图形图像工程研究中心 兰州 730070)

摘 要：方面级情感分析目前主要采用注意力机制与传统神经网络相结合的方法对方面与上下文词进行建模。这类方法忽略了句子中方面与上下文词之间的句法依存信息及位置信息，从而导致注意力权重分配不合理。为此，该文提出一种面向方面级情感分析的交互图注意力网络模型(IGATs)。该模型首先使用双向长短期记忆网络(BiLSTM)学习句子的语义特征表示，并结合位置信息生成新的句子特征表示，然后在新生成的特征表示上构建图注意力网络以捕获句法依存信息，再通过交互注意力机制建模方面与上下文词之间的语义关系，最后利用softmax进行分类输出。在3个公开数据集上的实验结果表明，与其他现有模型相比，IGATs的准确率与宏平均F1值均得到显著提升。

关键词：自然语言处理；方面级情感分析；位置嵌入；图注意力网络；注意力机制

中图分类号：TN912 **文献标识码**：A **文章编号**：1009-5896(2021)11-3282-09

DOI: 10.11999/JEIT210036

An Interactive Graph Attention Networks Model for Aspect-level Sentiment Analysis

HAN Hu[①②] WU Yuanhang[①] QIN Xiaoya[①]

[①](*School of Electronics and Information Engineering*, *Lanzhou Jiaotong University*, *Lanzhou* 730070, *China*)

[②](*Gansu Artificial Intelligence and Graphic Image Engineering Research Center*, *Lanzhou* 730070, *China*)

Abstract: At present, aspect-level sentiment analysis uses mainly the method of combining attention mechanism and traditional neural network to model aspect and contextual words.These methods ignore the syntactic dependency information and position information between aspects and contextual words in sentences, which leads to unreasonable weight allocation of attention. Therefore, an Interactive Graph ATtention (IGATs) networks model for aspect-level sentiment analysis is proposed. Bidirectional Long Short-Term Memory (BiLSTM) network is firstly used to learn the semantic feature representation of sentences. And then the position information is combined to update the feature representation, a graph attention network is constructed on the newly generated feature representation to capture syntactic dependency information. Finally, interactive attention mechanism is used to model the semantic relations between the aspect and contextual words. Experimental results on three public datasets show that the accuracy and macro average F1 value of IGATs are significantly improved compared with other existing models.

Key words: Natural language processing; Aspect-level sentiment analysis; Position embedding; Graph attention network; Attention mechanism

1 引言

方面级情感分析，也称细粒度情感分析，是情感分析的一个子领域，旨在识别在线评论中明确给出的方面的情感极性[1]。例如：“The restaurant environment is very good, but the waiter attitude is so bad.”在这句话中，“restaurant environment”和“waiter attitude”是两个不同的方面，可以看出，“restaurant environment”的情感极性是积极的，而“waiter attitude”的情感极性是消极的。由于两个方面的情感极性是截然相反的，因此不能仅考虑句子层面的情感极性。

收稿日期：2021-01-11；改回日期：2021-09-27；网络出版：2021-10-09

*通信作者： 吴渊航 1903552800@qq.com

基金项目：国家自然科学基金(62166024)，国家社会科学基金(17BXW071)

Foundation Items: The National Natural Science Foundation of China (62166024), The National Social Science Foundation of China (17BXW071)

早期，情感分析问题主要采用知识工程与传统机器学习相结合的方法[2–4]。近年来，以神经网络为代表的深度学习在各个领域取得了显著进展[5–9]，由于其可以自动提取特征而不需要人工干预，一些学者采用神经网络解决情感分析问题[7–9]。然而，基于神经网络的方法不能有效地区分句子中每个单词的重要程度，而且句子的情感极性不仅由内容决定，还与所涉及的方面密切相关。为此，一些学者将注意力机制引入情感分类任务中[10–14]，以重点关注句子中的方面。文献[10]提出了一种基于注意力机制的长短期记忆网络模型(ATtention-based Long Short-Term Memory network with Aspect Embedding, ATAE-LSTM)，当不同的方面作为输入时，注意力机制可以关注句子中的不同部分。文献[11]将记忆网络(Memory Network, MemNet)引入情感分类任务中，并结合多层注意力机制生成方面与上下文的特征表示。文献[12]认为目标和上下文都需要特殊对待，需要使用交互注意力机制来学习各自的表示，因而提出了交互式注意网络(Interactive Attention Networks, IAN)。文献[13]指出IAN模型中的池化操作会忽略句子与目标之间的词对交互，需要以联合的方式对方面和句子进行建模，为此提出一种新的交互注意力模块(Attention-Over-Attention, AOA)。文献[14]提出了一种方面-上下文交互表示结构，该结构通过使用多头注意力机制来生成方面和上下文的特征表示。

虽然注意力机制与传统神经网络相结合的方法已经取得了一定效果，但是这类方法忽略了句子中方面与上下文词之间的句法依存信息以及位置信息，可能导致注意力机制难以学习到合适的注意力权重。具体来说，现有的注意力机制可能导致给定的方面错误地将句法上不相关的上下文词识别为情感词。为了解决这一问题，一些学者将图神经网络(Graph Neural Network, GNN)引入情感分类任务中[15,16]，使得句法依存信息得到了有效的利用。文献[15]使用图卷积网络(Graph Convolutional Network, GCN)从句子的依赖树中学习节点表示，并结合注意力机制将其用于情感分类任务。文献[16]提出了一种基于GCN的方面级情感分类模型，该模型可以有效地捕捉句子中多个方面之间的情感依赖关系。此外，方面与上下文词之间的位置信息对识别方面的情感极性来说也是至关重要的，句子中与方面相邻的词要比其他距离较远的词受到更多的关注。为了合理地利用位置信息，文献[17]提出了一种位置感知双向注意力网络模型(Position-aware Bidirectional Attention Network, PBAN)，该模型不仅关注方面的位置信息，还利用双向注意力机制相互建模方面和句子之间的关系。文献[18]提出一种基于依存树及距离注意力的句子属性情感分类模型(Dependency Tree and Distance Attention, DTDA)，该模型通过计算句子中方面与上下文词的相对距离与语法距离，构造包含语义信息和距离信息的句子特征表示。需要注意的是，上述方法大多对句法依存信息以及位置信息进行单独建模，依存信息以及位置信息并没有得到有效的利用，这使得注意力机制的作用没有得到充分的发挥。为此，本文提出一种面向方面级情感分析的交互图注意力网络模型(an Interactive Graph ATtention networks, IGATs)。本文的贡献如下：

(1)本文对句子中方面与上下文词之间的句法依存信息与位置信息进行了有效的建模，使得模型不仅能够学习到位置信息，还能够捕获到词与词之间的句法依存关系；

(2)模型使用交互注意力机制建模方面与上下文词之间的语义关系，并分别生成方面与上下文的表示，以判断方面的情感极性；

(3)本文在3个公开数据集上评估模型的性能，实验结果表明，该模型对不同数据集的准确率(Accuracy, Acc)与宏平均F1值(Macro average F1, MF1)均有所提升，并进一步证明了句法依存信息与位置信息对模型的重要性。

2　图注意力网络

2.1　句法依存树

本文借助Spacy文本处理库中的依存解析器生成句子的句法依存树。以“Its size is ideal but the weight is unacceptable.”为例，其依存树结构如图1所示。其中，黄色字体表示方面词，红色字体表示情感词，有向边表示依存关系，标签表示依存关系类型。为了使句法结构的作用得以充分发挥，本文使用图注意力网络(Graph Attention Network, GAT)编码句法依存树，以捕获词与词之间的句法依存关系。

2.2　图注意力网络

图注意力网络(GAT)[19]是GNN的一种变体，它通过使用多头注意力机制聚集邻居节点的表示来迭代更新每个节点的表示，实现了对不同邻居节点权重的自适应分配。给定一个有N个节点的依存图(即句法依存树)，设图中任意节点i在第l层GAT所对应的特征向量为$\boldsymbol{h}_i^{l+1}$, $\boldsymbol{h}_i^{l+1} \in R^{Kd^{l+1}}$, K表示图注意力的头数，d^{l+1}表示头数为1的情况下，图注意力网络输出的特征向量的维度。更新公式为

$$\boldsymbol{h}_i^{l+1} = \Big\|_{k=1}^{K} \sigma\left(\sum_{j\in N(i)} a_{ij}^{lk}\boldsymbol{W}^{lk}\boldsymbol{h}_j^{l}\right) \tag{1}$$

$$a_{ij}^{lk} = \frac{\exp(\text{LeakyReLU}(\boldsymbol{a}_{lk}^{\mathrm{T}}[\boldsymbol{W}^{lk}\boldsymbol{h}_i^{l} \,\|\, \boldsymbol{W}^{lk}\boldsymbol{h}_j^{l}]))}{\sum_{u\in N(i)}\exp(\text{LeakyReLU}(\boldsymbol{a}_{lk}^{\mathrm{T}}[\boldsymbol{W}^{lk}\boldsymbol{h}_i^{l} \,\|\, \boldsymbol{W}^{lk}\boldsymbol{h}_u^{l}]))} \tag{2}$$

其中，$\|$表示拼接操作，$N(i)$表示节点i的邻接节点，a_{ij}^{lk}表示第l层GAT的第k个注意力头计算出的权重系数，$\boldsymbol{W}^{lk}\in R^{d^{l+1}\times d^{l}}$与$\boldsymbol{a}_{lk}^{\mathrm{T}}\in R^{2d^{l+1}}$是权重参数，$\boldsymbol{h}_j^{l}\in R^{d^{l}}$表示邻接节点的隐藏状态向量，LeakyReLU表示激活函数。

3 交互图注意力网络模型IGATs

IGATs模型(如图2)主要由词嵌入层与LSTM层、位置嵌入层、GAT层、交互注意力层与输出层5部分组成。模型的组件在本节其余部分单独介绍。

3.1 词嵌入层与LSTM层

给定一个包含n个词的句子$(w_1,w_2,\cdots,w_{a+1},\cdots,w_{a+m},\cdots,w_{n-1},w_n)$，其中，$(w_{a+1},w_{a+2},\cdots,w_{a+m})$表示有$m$个词的方面(可以是方面词或方面短语)。本文首先使用嵌入矩阵$\boldsymbol{E}\in R^{|V|\times d_e}$将每个单词嵌入到低维实值向量空间中，其中，$|V|$表示词表的大小，$d_e$表示词嵌入的维数；然后，将初始化的单词向量输入到BiLSTM中，以生成句子的隐藏状态向量$\boldsymbol{H}=(\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_{a+1},\cdots,\boldsymbol{h}_{a+m},\cdots,\boldsymbol{h}_{n-1},\boldsymbol{h}_n)$；最后通过copy操作与mask操作分别生成上下文的隐藏状态向量$\boldsymbol{H}^{\mathrm{c}}=(\boldsymbol{h}_1^{\mathrm{c}},\boldsymbol{h}_2^{\mathrm{c}},\cdots,\boldsymbol{h}_{a+1}^{\mathrm{c}},\cdots,\boldsymbol{h}_{a+m}^{\mathrm{c}},\cdots,\boldsymbol{h}_{n-1}^{\mathrm{c}},\boldsymbol{h}_n^{\mathrm{c}})$与方面的隐藏状态向量$\boldsymbol{H}^{\mathrm{a}}=(0,\cdots,\boldsymbol{h}_{a+1}^{\mathrm{a}},\cdots,\boldsymbol{h}_{a+m}^{\mathrm{a}},\cdots,0)$，其中$\boldsymbol{h}_i\in R^{2d_h}$表示第$i$个词的隐藏状态向量，$d_h$为单向LSTM的输出维数。

3.2 位置嵌入层

位置嵌入层(Position Embedding, PE)采用方面与上下文词之间的距离来衡量二者的相关性，距

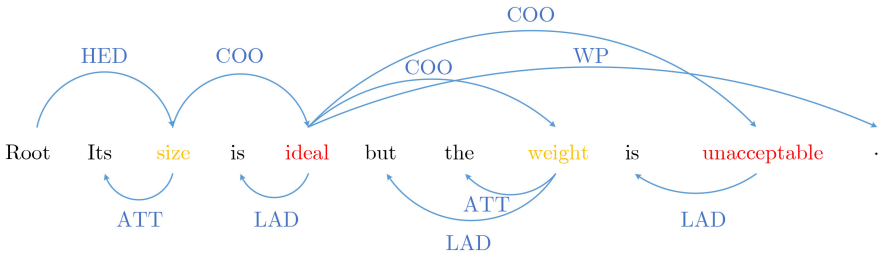

图 1 句法依存树

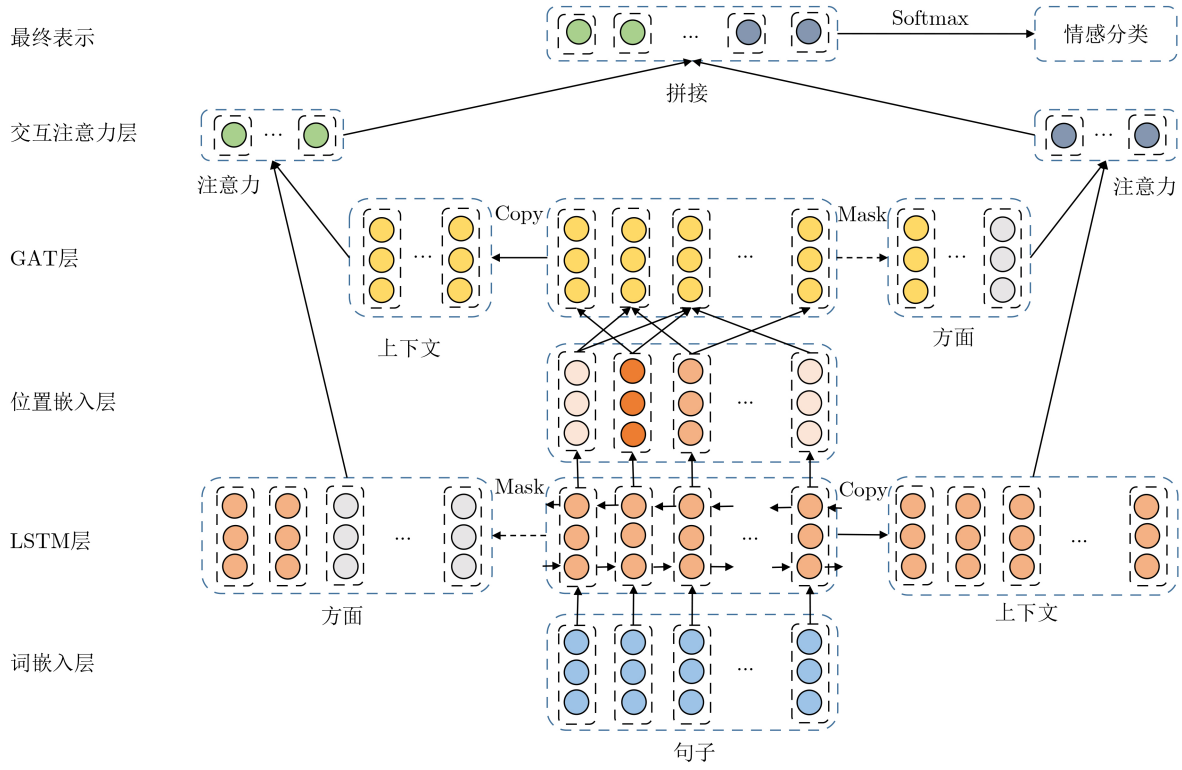

图 2 IGATs模型

离分为语法距离和相对距离。语法距离可由句法依存树得到，设方面词w_a与上下文词w_i之间的语法距离为d_i，文献[18]将d_i定义为二者在依存树中的路径长度，语法距离特征计算公式为

$$s_i=\begin{cases}1-\dfrac{d_i}{2d_{\max}}, 1\le i<a+1, a+m<i\le n\\ 0, \qquad\qquad a+1\le i\le a+m\end{cases} \tag{3}$$

通过计算式(3)可以得到句子中方面与上下文词之间的语法距离特征向量$\boldsymbol{S}=(s_1,s_2,\cdots,s_n)$。其中，$d_{\max}$表示路径长度的最大值，$s_i$表示方面与上下文词$w_i$之间的语法距离特征。若方面为短语，则以该短语中距离依存树根节点最近的一个单词作为方面词。相对距离可由方面与上下文词在句子中的路径长度得到，相对距离特征计算公式为

$$r_i=\begin{cases}1-\dfrac{a+1-i}{n}, & 1\le i<a+1\\ 0, & a+1\le i\le a+m\\ 1-\dfrac{i-a-m}{n}, & a+m<i\le n\end{cases} \tag{4}$$

通过计算式(4)可以得到句子中方面与上下文词之间的相对距离特征向量$\boldsymbol{R}=(r_1,r_2,\cdots,r_n)$，其中$r_i$表示方面与上下文词$w_i$之间的相对距离特征。然后，结合语法距离特征和相对距离特征生成新的距离特征表示

$$p_i=\frac{r_i+s_i}{2} \tag{5}$$

其中，p_i表示句子中方面与上下文词w_i之间的距离特征。最后，通过引入距离特征更新句子的隐藏状态，新生成的句子特征表示包含了单词的语义信息和位置信息。更新公式为

$$\boldsymbol{h}_i=p_i\boldsymbol{h}_i \tag{6}$$

3.3 GAT层

本文使用GAT捕获句子中方面与上下文词之间的句法依存关系，GAT层的初始输入为位置嵌入层的输出，为了使位置嵌入层的作用得以充分发挥，本文对每层GATs的输入都进行了位置嵌入，由前文可知，l层GATs的输出结果为$\boldsymbol{H}^l=\{\boldsymbol{h}_1^l,\boldsymbol{h}_2^l,\cdots,\boldsymbol{h}_{a+1}^l,\cdots,\boldsymbol{h}_{a+m}^l,\cdots,\boldsymbol{h}_{n-1}^l,\boldsymbol{h}_n^l\}$，$\boldsymbol{h}_i^l\in R^{2Kd_h}$。

3.4 交互注意力层

交互注意力层(Interactive Attention layer, IAtt)使用交互注意力机制建模方面与上下文之间的语义关系，并分别生成方面和上下文的表示。由于GAT层采用多头注意力机制来更新节点的隐藏状态，拼接操作会导致模型参数过多，为了防止过拟合，本文首先引入线性层对GAT层的输出进行降维

$$\boldsymbol{h}_i=\boldsymbol{W}\boldsymbol{h}_i^l+\boldsymbol{b} \tag{7}$$

其中，$\boldsymbol{W}\in R^{2d_h\times 2Kd_h}$和$\boldsymbol{b}\in R^{2d_h}$分别是权重参数和偏置项。然后，本文借助mask操作与copy操作分别生成面向方面的特征向量$\boldsymbol{A}=\{0,\cdots,\boldsymbol{a}_{a+1},\cdots,\boldsymbol{a}_{a+m},\cdots,0\}$与面向上下文的特征向量$\boldsymbol{C}=\{\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_{a+1},\cdots,\boldsymbol{c}_{a+m},\cdots,\boldsymbol{c}_{n-1},\boldsymbol{c}_n\}$，通过计算上下文的隐藏状态向量$\boldsymbol{h}_i^{\mathrm{c}}\in\boldsymbol{H}^{\mathrm{c}}$与面向方面的向量$\boldsymbol{a}_i\in\boldsymbol{A}$之间的语义相关性$\beta_t$，并为每个上下文词$\boldsymbol{h}_t^{\mathrm{c}}$分配相关的注意力权重$\alpha_t$

$$\beta_t=\sum_{i=1}^{n}\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{h}_i^{\mathrm{c}} \tag{8}$$

$$\alpha_t=\frac{\exp(\beta_t)}{\sum\limits_{i=1}^{n}\exp(\beta_i)} \tag{9}$$

$$\boldsymbol{c}=\sum_{t=1}^{n}\alpha_t\boldsymbol{h}_t^{\mathrm{c}} \tag{10}$$

通过计算方面的隐藏状态向量$\boldsymbol{h}_i^{\mathrm{a}}\in\boldsymbol{H}^{\mathrm{a}}$与面向上下文的向量$\boldsymbol{c}_i\in\boldsymbol{C}$之间的语义相关性$\phi_t$，为每个方面词$\boldsymbol{h}_t^{\mathrm{a}}$分配相关的注意力权重$\varphi_t$

$$\phi_t=\sum_{i=a+1}^{a+m}\boldsymbol{c}_i^{\mathrm{T}}\boldsymbol{h}_i^{\mathrm{a}} \tag{11}$$

$$\varphi_t=\frac{\exp(\phi_t)}{\sum\limits_{i=a+1}^{a+m}\exp(\phi_i)} \tag{12}$$

$$\boldsymbol{a}=\sum_{t=a+1}^{a+m}\varphi_t\boldsymbol{h}_t^{\mathrm{a}} \tag{13}$$

最后将生成的方面表示$\boldsymbol{a}$与上下文表示$\boldsymbol{c}$进行拼接得到最终的表示$\boldsymbol{r}$，其中$\boldsymbol{r}\in R^{4d_h}$。

3.5 输出层

本文将获得的最终表示$\boldsymbol{r}$输入到全连接层，最后通过softmax进行分类输出

$$p=\mathrm{softmax}(\boldsymbol{W}_p\boldsymbol{r}+\boldsymbol{b}_p) \tag{14}$$

其中，$p\in R^{d_p}$为极性决策空间，d_p表示情感标签的维数，$\boldsymbol{W}_p\in R^{d_p\times 4d_h}$和$\boldsymbol{b}_p\in R^{d_p}$分别是权重参数和偏置项。模型的损失函数采用交叉熵损失函数，使用标准的梯度下降法进行训练

$$\mathrm{loss}=-\sum_{i=1}^{d_p}y_i\lg p_i+\lambda\|\theta\|_2 \tag{15}$$

其中，$y\in R^{d_p}$为真实概率分布，i为概率分布y和p的第i个元素，λ是L2正则化系数。

4 实验

4.1 实验数据与实验平台

为了验证IGATs模型的有效性，本文在3个公开的数据集上进行了实验，分别是Twitter数据集[7]、Laptop数据集[1]、Restaurant数据集[1]。这些数据集包含3个标签(即$d_p=3$)，分别是积极(Positive)、中性(Neutral)、消极(Negative)。数据集统计结果如表1所示。

本文的实验平台如表2所示，实验主要采用Facebook的Pytorch-GPU深度学习框架与Spacy文本处理库，使用python3.7实现。

4.2 实验设置与评价指标

本文使用300维的GloVe向量[20]初始化单词嵌入，模型的所有权重参数均采用均匀分布进行初始化。GAT的层数设置为1，图注意力头数设置为1，此时模型的效果最好。模型的超参数设置如表3所示。

模型采用精准率(Precision, P)、召回率(Recall, R)、准确率(Acc)与宏平均F1值(MF1)等作为评价指标。对于单个类别，设TP为正确预测的样本，FP为其他类别被预测为当前类别的样本，FN为当前类别被预测为其他类别的样本，则P, R, F1值的计算公式为

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{16}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{17}$$

表 1 数据集统计

数据集	积极	中性	消极
Twitter-train	1561	3127	1560
Twitter-test	173	346	173
Laptop-train	994	464	870
Laptop-test	341	169	128
Restaurant-train	2164	637	807
Restaurant-test	728	196	196

表 2 实验平台

实验环境	具体信息
操作系统	Windows 10 教育版
CPU	Intel(R) Core(TM) i7-7700 CPU @ 3.60 GHz
内存	16.0 GB
显卡	GTX 1080
显存	8.0 GB

表 3 超参数设置

超参数	超参数值数量
词嵌入维度	300
隐藏状态向量维度	300
Batch size	16
训练迭代次数epoch	100
优化器Optimizer	Adam
学习率Learning rate	0.001
Dropout rate	0.3
L2正则化系数	0.00001

$$\mathrm{F1}=\frac{2\times P\times R}{P+R} \tag{18}$$

对于d_p个类别，准确率Acc与MF1的计算公式为

$$\mathrm{Acc}=\frac{\sum_{i=1}^{d_p}\mathrm{TP}_i}{\sum_{i=1}^{d_p}(\mathrm{TP}_i+\mathrm{FP}_i+\mathrm{FN}_i)} \tag{19}$$

$$\mathrm{MF1}=\frac{1}{d_p}\sum_{i=1}^{d_p}\mathrm{F1}_i \tag{20}$$

其中，MF1值是指所有类别的F1值的平均，模型重复运行3次，取Acc最高的一次作为最终结果。

4.3 对比模型

为了评估本文的模型，将它与一系列基准模型进行比较，各个模型简介如下所示：

SVM[4]：采用有监督的机器学习方法检测方面术语和方面类别，并检测客户评论中对方面术语和方面类别表达的情感。

LSTM[9]：一种目标相关的长短期记忆(LSTM)模型，该方法在生成句子表示时可以捕捉目标词和其上下文之间的联系，以用于目标相关的情感分类。

MemNet[11]：采用深度记忆网络与多层注意力机制相耦合的方法进行情感分类。

IAN[12]：使用两个注意力网络交互地建模目标与上下文之间的语义关系，并分别生成目标与上下文的表示，为判断方面的情感极性提供了足够的信息。

AOA[13]：对方面和句子进行联合建模，以捕捉方面和上下文之间的相互作用，共同学习方面和句子的表示，并自动地关注句子中的重要部分。

AOA-MultiACIA[14]：一种新的方面-上下文交互表示模型，它能够在其上下文序列建模过程中提取与特定方面相关的特征，并同时生成高质量的方面表示。

ASGCN[15]：通过在句子的句法依存树上构建多层GCN以捕获句法依存信息，并结合注意力机制进行情感分类。

GATs[19]：使用GAT模型进行情感分类，GAT模型描述如2.2节所示。

4.4 实验结果与分析

表4给出了各个模型在3个数据集上的性能对比。可以看出，虽然SVM模型在Laptop和Restaurant数据集上的性能优于LSTM模型，但是它依赖人工特征提取，费时费力且效率低下。基于注意力机制的模型(如MemNet, IAN, AOA, AOA-MultiACIA)性能均优于LSTM模型，这表明注意力机制在情感分析任务中是有效的，而且采用交互注意力机制的IAN与AOA模型性能要优于采用传统注意力机制的MemNet模型，甚至不亚于将句法依存信息考虑在内的ASGCN模型，这表明方面与上下文词都需要受到关注，而且二者之间的交互有利于情感信息的传播，并最终提升情感预测模型的性能。

然而，上述模型要么只考虑句法依存信息或位置信息，要么只关注于方面与上下文词之间的交互，难以实现二者之间的协调优化。为此，本文提出一个新的模型，即IGATs。可以看出，IGATs在3个数据集上的Acc与MF1值整体上优于其他模型。相较于仅包含多层注意力机制的AOA-MultiACIA模型，IGATs在Twitter数据集上的Acc与MF1分别提升了2.89%和4%，在Laptop数据集上的Acc与MF1分别提升了0.75%和1.81%，在Restaurant数据集上的MF1提升了1.86%，这表明句法依存信息与位置信息有利于识别方面的情感极性，而且Twitter数据集对句法信息与位置信息更加敏感。需要注意的是，IGATs在Restaurant数据集上的Acc稍低于AOA-MultiACIA模型，这是因为该模型采用多层交互注意力机制，它能够在上下文序列建模过程中提取与特定方面相关的特征，并生成高质量的方面表示。相较于将句法依存信息与相对位置信息考虑在内的ASGCN模型，IGATs在Twitter数据集上的Acc与MF1分别提升了3.14%和3%，在Laptop数据集上的Acc与MF1分别提升了0.47%和1%，在Restaurant数据集上的Acc与MF1分别提升了1.55%和1.97%，相较于将句法依存信息与位置信息考虑在内的GATs模型，IGATs在各个数据集上的Acc与MF1也均有所提升，这表明依存信息与位置信息需要结合交互注意力网络才能发挥更好的效果。

除此之外，文章统计了各个模型的可训练参数数量(1M=1Million)，如表5所示。由于SVM模型采用人工提取特征的方法进行情感分类，模型性能的好坏很大程度上取决于特征工程，其效率很低。MemNet模型的参数数量最少，其性能虽然优于LSTM，但是仍不及其他模型；IAN模型的参数规模与ASGCN相当，然而其在Laptop与Restaurant上的性能远不及ASGCN；AOA模型借鉴了IAN模型的交互注意力机制，其模型计算复杂度不亚于IAN，但是二者性能很接近；AOA-MultiACIA模型在AOA模型的基础上叠加了多层多头注意力机制，虽然其在各个数据集上的性能不亚于ASGCN，但是模型计算复杂度远高于AOA，更高于ASGCN；GATs模型的参数规模比ASGCN少，二者性能却十分接近，因此，模型效率高于ASGCN模型；IGATs模型的参数规模与GATs接近，但是其性能整体上优于其他模型，这表明IGATs模型的效率远高于其他模型。

4.5 消融实验

为了进一步检查IGATs的每个组件对性能的影响，本文对IGATs进行了消融研究，具体如下：

BiLSTM+IAtt：模型由LSTM层与IAtt层组

表 4　各个模型的性能对比(%)

模型	Twitter		Laptop		Restaurant	
	准确率(Acc)	宏平均F1	准确率(Acc)	宏平均F1	准确率(Acc)	宏平均F1
SVM	63.40	63.30	70.49	N/A	80.16	N/A
LSTM	69.56	67.70	69.28	63.09	78.13	67.47
MemNet	71.48	69.90	70.64	65.17	79.61	69.64
IAN	72.50	70.81	72.05	67.38	79.26	70.09
AOA	72.30	70.20	72.62	67.52	79.97	70.42
AOA-MultiACIA	72.40	69.40	75.27	70.24	**82.59**	72.13
ASGCN	72.15	70.40	75.55	71.05	80.77	72.02
GATs	73.12	71.25	74.61	70.51	80.63	70.41
IGATs	**75.29**	**73.40**	**76.02**	**72.05**	82.32	**73.99**

表 5 各个模型的可训练参数数量(M)

模型	可训练参数数量
SVM	–
LSTM	0.72
MemNet	0.36
IAN	2.17
AOA	2.10
ASGCN	2.17
GATs	1.81
IGATs	**1.81**

成，LSTM层用来学习句子的语义特征表示，IAtt层用来建立方面与上下文词之间的语义关系。

BiLSTM+GAT+IAtt：模型仅在LSTM层后新增了一个GAT层，用来捕获方面与上下文词之间的句法依存关系。

BiLSTM+PE+IAtt：模型仅在LSTM层后新增了一个PE层，用来将位置信息嵌入到句子的语义特征表示中。

IGATs：完整的模型，不仅考虑了位置信息，还考虑了句法依存信息。

各个模型的性能对比如表6所示。可以看出，BiLSTM+IAtt模型在3个数据集的Acc与MF1值整体上不及其他模型，这表明句子中方面与上下文词之间的句法依存信息与位置信息是不可忽略的。BiLSTM+GAT+IAtt模型的性能整体上优于BiLSTM+IAtt模型，但仍不及IGATs，这表明虽然句法依存信息有利于识别方面的情感极性，但是也不能忽略了位置信息的影响。BiLSTM+PE+IAtt模型的性能在Twitter数据集上比IGATs低，在Laptop数据集上要优于IGATs，在Restaurant数据集上的性能接近IGATs，这表明相较于句法依存信息，Laptop对位置信息更敏感。由此可见，PE层对IGATs模型的贡献最大，其次是GAT层，但是也不能忽视GAT层的影响。

4.6 其他因素

由于GAT层包含多层GAT，每层GAT又可能包含多个图注意力头，不能忽视这些因素对模型性能的影响。因此，本文研究了GAT头数K为1的情况下，GAT层数L对IGATs性能的影响。通过改变$L\in\{1,2,3,4,5,6,7,8,9,10\}$的值，测试IGATs在Restaurant数据集上的Acc和MF1值。实验结果如图3所示。可以看出，IGATs的准确率Acc随着L的增加而下降，MF1值在层数小于5的情况下比较稳定，在层数为5的情况下取得最大值，随后开始下降。

除此之外，本文也研究了GAT层数为1的情况下，图注意力头数K对IGATs最终性能的影响。通过改变$K\in\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$的值，测试IGATs在Restaurant数据集上的Acc和MF1值。实验结果如图4所示。可以看出，IGATs的准确率Acc随着K的增加整体呈现下降趋势，MF1值在图注意力头数为6时最大，随后开始下降。综合各方面因素，本文将GAT层数以及图注意力头数均设为1。

5 结束语

本文提出一种面向方面级情感分析的交互图注意力网络模型(IGATs)。首先，通过计算方面与上

表 6 消融研究(%)

模型	Twitter		Laptop		Restaurant	
	准确率(Acc)	宏平均F1	准确率(Acc)	宏平均F1	准确率(Acc)	宏平均F1
BiLSTM+IAtt	74.13	72.86	75.08	70.82	81.25	72.14
BiLSTM+GAT+IAtt	74.86	72.98	74.92	71.08	82.05	73.45
BiLSTM+PE+IAtt	74.42	72.35	**76.65**	**72.75**	82.23	**74.01**
IGATs	**75.29**	**73.40**	76.02	72.05	**82.32**	73.99

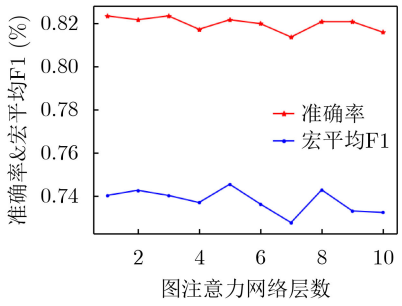

图 3 GAT层数L对模型性能的影响

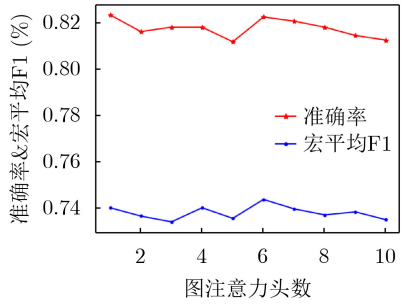

图 4 GAT头数K对模型性能的影响

下文词之间的语法距离与相对距离实现了位置信息的利用；其次，通过构建句法依存树和图注意力网络实现了句法依存信息的融合；最后，通过建立方面与上下文词之间的交互注意力模型，实现了两者之间的协调优化。实验结果表明IGATs能够有效捕获方面与上下文词之间的句法依存信息和位置信息，模型在多个数据集上的评价性能都得到了显著提升。同时，模型仍然存在一些有待研究的内容，例如，没有考虑方面与上下文词之间的依存关系标签以及词性对任务的影响，这些将在今后的工作中进行完善。

参 考 文 献

[1] PONTIKI M, GALANIS D, PAVLOPOULOS J, *et al*. Semeval-2014 task 4: Aspect based sentiment analysis[C]. The 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 2014: 27–35. doi: 10.3115/v1/S14-2004.

[2] DING Xiaowen, LIU Bing, and YU P S. A holistic lexicon-based approach to opinion mining[C]. 2008 International Conference on Web Search and Data Mining, Palo Alto, USA, 2008: 231–240. doi: 10.1145/1341531.1341561.

[3] JIANG Long, YU Mo, ZHOU Ming, *et al*. Target-dependent twitter sentiment classification[C]. The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, USA, 2011: 151–160.

[4] KIRITCHENKO S, ZHU Xiaodan, CHERRY C, *et al*. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews[C]. The 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 2014: 437–442. doi: 10.3115/v1/S14-2076.

[5] PAN S T, HUANG Zonghong, YUAN S S, *et al*. Application of hidden Markov models in speech command recognition[J]. *Journal of Mechanics Engineering and Automation*, 2020, 10(2): 41–45. doi: 10.17265/2159-5275/2020.02.001.

[6] WANG Lei. Application research of deep convolutional neural network in computer vision[J]. *Journal of Networking and Telecommunications*, 2020, 2(2): 23–29. doi: 10.18282/jnt.v2i2.886.

[7] DONG Li, WEI Furu, TAN Chuanqi, *et al*. Adaptive recursive neural network for target-dependent twitter sentiment classification[C]. The 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Baltimore, USA, 2014: 49–54. doi: 10.3115/v1/P14-2009.

[8] XUE Wei and LI Tao. Aspect based sentiment analysis with gated convolutional networks[C]. The 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 2018: 2514–2523. doi: 10.18653/v1/P18-1234.

[9] TANG Duyu, QIN Bing, FENG Xiaocheng, *et al*. Effective LSTMs for target-dependent sentiment classification[C]. COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 2016: 3298–3307.

[10] WANG Yequan, HUANG Minlie, ZHU Xiaoyan, *et al*. Attention-based LSTM for aspect-level sentiment classification[C]. 2016 Conference on Empirical Methods in Natural Language Processing, Austin, USA, 2016: 606–615. doi: 10.18653/v1/D16-1058.

[11] TANG Duyu, QIN Bing, and LIU Ting. Aspect level sentiment classification with deep memory network[C]. 2016 Conference on Empirical Methods in Natural Language Processing, Austin, USA, 2016: 214–224. doi: 10.18653/v1/D16-1021.

[12] MA Dehong, LI Sujian, ZHANG Xiaodong, *et al*. Interactive attention networks for aspect-level sentiment classification[C]. The 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017: 4068–4074. doi: 10.24963/ijcai.2017/568.

[13] HUANG Binxuan, OU Yanglan, and CARLEY K M. Aspect level sentiment classification with attention-over-attention neural networks[C]. The 11th International Conference on Social, Cultural, and Behavioral Modeling, Washington, USA, 2018: 197–206. doi: 10.1007/978-3-319-93372-6_22.

[14] WU Zhuojia, LI Yang, LIAO Jian, *et al*. Aspect-context interactive attention representation for aspect-level sentiment classification[J]. *IEEE Access*, 2020, 8: 29238–29248. doi: 10.1109/ACCESS.2020.2972697.

[15] ZHANG Chen, LI Qiuchi, and SONG Dawei. Aspect-based sentiment classification with aspect-specific graph convolutional networks[C]. The 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 2019: 4568–4578. doi: 10.18653/v1/D19-1464.

[16] ZHAO Pinlong, HOU Linlin, and WU Ou. Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification[J]. *Knowledge-Based Systems*, 2020, 193: 105443. doi: 10.1016/j.knosys.2019.105443.

[17] GU Shuqin, ZHANG Lipeng, HOU Yuexian, *et al*. A position-aware bidirectional attention network for aspect-

level sentiment[C]. The 27th International Conference on Computational Linguistics, Santa Fe, USA, 2018: 774–784.

[18] 苏锦钿, 欧阳志凡, 余珊珊. 基于依存树及距离注意力的句子属性情感分类[J]. 计算机研究与发展, 2019, 56(8): 1731–1745. doi: 10.7544/issn1000-1239.2019.20190102.
SU Jindian, OUYANG Zhifan, and YU Shanshan. Aspect-level sentiment classification for sentences based on dependency tree and distance attention[J]. *Journal of Computer Research and Development*, 2019, 56(8): 1731–1745. doi: 10.7544/issn1000-1239.2019.20190102.

[19] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, *et al*. Graph attention networks[C]. The 6th International Conference on Learning Representations, Vancouver, Canada, 2018: 1–12.

[20] PENNINGTON J, SOCHER R, and MANNING C. Glove: Global vectors for word representation[C]. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014: 1532–1543. doi: 10.3115/v1/D14-1162.

韩 虎：男，1977年生，教授，研究方向为神经网络与深度学习、数据挖掘与自然语言处理.

吴渊航：男，1997年生，硕士生，研究方向为深度学习与自然语言处理.

秦晓雅：女，1996年生，硕士生，研究方向为深度学习与自然语言处理.

责任编辑：余 蓉