



基于 BERT 与注意力机制的方面级隐式情感分析模型

摘要

在方面级情感文本中存在部分不含情感词的评论句,对其情感的研究被称为方面级隐式情感分析.现有分析模型在预训练过程中可能会丢失与方面词相关的上下文信息,并且不能准确提取上下文中深层特征.本文首先构造了方面词感知 BERT 预训练模型,通过将方面词引入到基础 BERT 的输入嵌入结构中,生成与方面词信息相关的词向量;然后构造了语境感知注意力机制,对由编码层得到的深层隐藏向量,将其中的语义和句法信息引入到注意力权重计算过程,使注意力机制能更加准确地分配权重到与方面词相关的上下文.对比实验结果表明,本文模型的效果优于基线模型.

关键词

方面级情感分析;隐式情感分析;门控循环单元;注意力机制;BERT 模型

中图分类号 TP391

文献标志码 A

收稿日期 2022-09-14

资助项目 国家自然科学基金(61273229, 51705260);江苏省高校青蓝工程创新团队

作者简介

杨春霞,女,博士,教授,研究方向为自然语言处理.y.cx@163.com

韩煜(通信作者),硕士生,研究方向为自然语言处理.hanyu19970620@163.com

0 引言

方面级情感分析(Asspect-Based Sentiment Analysis, ABSA)属于情感分析中的细粒度任务,目的在于预测评论中某个具体方面的情感极性^[1].互联网的高速发展大大提升了舆论的产生和传播速度,评论中所包含的情感指向性和应用价值也成为关注的热点^[2].在常见的评论句中,存在着不少没有明确情感词的评论,却依旧可以清晰地表达情感.为了与带有明显情感倾向词的方面级显式情感区别开,将此类文本命名为方面级隐式情感.例如"Having USB3 is why I bought this Mini."句中方面词"USB3"并没有对应的情感词,但仍可以清楚地感受到对该方面的积极情感.如今互联网中存在大量类似的委婉评论,而开展方面级隐式情感分析研究可以有效提升对这些评论的理解,从中获得的数据可用于舆论分析、商品价格预测、民众心理分析等.

相比于显式情感,隐式情感的研究主要有两个难点:一是缺乏情感词导致情感语义特征不易识别;二是隐式情感与上下文中的主观表达联系更加紧密,导致对语义的理解难度加大.早期隐式情感分析的研究主要基于构建隐式情感数据集或者构建隐式情感词典.Russo 等^[3]提出在研究的语料中存在着隐含的情感极性,并提供了隐式情感数据集.Choi 等^[4]通过构造相应的 EffectWordNet 情感词典,检测到语料中的隐式情感.然而情感词典存在着更新代价大、效率不高等缺点,因此已经逐渐被更加高效的神经网络模型所取代.

为了方便提取句子中的语义信息,目前隐式情感研究通常会输入句进行词向量化处理.词向量模型主要分为静态和动态两种.静态词向量模型的主要功能是为每个词提供固定的词向量,如 Word2Vec^[5]和 Glove^[6]模型,但它们不适用于一词多义的文本,因此出现了动态词向量模型.动态词向量模型能够依据同一个词在不同语境中与其他词之间的联系而建模,并生成对应的词向量,如 ELMO^[7]模型.近年出现的预训练模型如 BERT(Bidirectional Encoder Representations from Transformers)^[8]也属于动态词向量模型, Li 等^[9]利用 BERT 预训练模型和外部语料库知识来增强模型识别情感语义特征的能力,将方面级文本中的显式与隐式情感标签相匹配,有效地捕获了上下文中的隐含情感知识,得到了更好的方面级隐式情感分类结果.但是该模型需要额外添加方面词遮掩任务才能感知方面词信息,这导致其在使用 BERT 进行预训练的过程中可能会丢失与方面词相

¹ 南京信息工程大学 自动化学院/江苏省大数据分析技术重点实验室/大气环境与装备技术协同创新中心,南京,210044

关的上下文信息,并且在后续的编码过程中很难弥补这种损失.

上下文感知模型与注意力(Attention)机制相结合的方法也在隐式情感分析中取得了不错的效果,其主要原理是先对句子中的上下文信息进行建模,再依据不同信息的权重去度量它们的重要程度.Klinger等^[10]提出一个共享任务,对一个自动标记的Twitter数据集,要求系统在不需捕获情感词特征的情况下能够识别其中的情感极性,参赛的前三名都采用BiLSTM与注意力机制相融合模型,而他们的区别在于:Balazs等^[11]采用基于深层的语境化(ELMo)表示编码词语信息;Chronopoulou等^[12]使用预训练模型的权重来初始化网络的特定层;Rozenal等^[13]在大规模文本的基础上训练语言模型并获取词向量作为输入.此外在方面级显式情感分析中,杨春霞等^[14]融合深度双向门控循环单元(Deep Bi-directional Gated Recurrent Unit, DBGRU)与注意力机制进行情感分类任务,通过实验说明DBGRU能更加高效地提取深层语义特征.受其启发,本文尝试将DBGRU与注意力机制相融合,用于提取方面级隐式情感中的深层特征.然而一般注意力机制存在无法有效利用深层特征中语境信息的缺点,导致其不能准确提取与方面词相关的深层情感特征.

本文提出一种融合深度双向门控循环单元及语境感知注意力机制的方面词感知BERT(Deep Bi-directional Gated Recurrent Unit Context-Aware Attention Aspect-Aware BERT, DCAB)模型,主要贡献如下:

1)本文改进了基础BERT的输入嵌入部分,构造了方面词感知BERT(Asspect-Aware BERT, AABERT),对句子中方面级隐式情感语料进行了预训练编码.AABERT将方面词作为输入句的相邻句进行预测,通过下一句预测任务,能够在预训练过程中有效捕获与方面词相关的上下文信息,最终获得与方面词相关的动态词向量表示.

2)本文构造了一种语境感知注意力(Context-Aware Attention, CAT)机制,用于关注深层情感特征中与方面词相关的重要信息.对于由DBGRU编码器得到的深层特征向量,CAT机制在一般注意力机制的权重参数中引入了深层语境信息,并添加了一个学习参数来明确每一个隐藏特征向量和深层语境信息对注意力权重大小的影响,有效地捕捉了深度上下文中与方面词相关的语义信息和句法信息.

1 DCAB模型

本文提出的DCAB模型的结构框架如图1所示,主要组成部分如下:

1)词嵌入层:将输入文本通过AABERT模型进行预训练,生成与方面词相关的动态词向量.

2)DBGRU词向量编码层:将AABERT词向量输入到DBGRU编码器中,提取包含上下文信息的深层特征向量,其中高层的隐藏特征包含上下文中的语义信息,低层的隐藏特征包含上下文中的句法信息.

3)CAT机制层:将DBGRU编码器输出的深层特征向量作为输入,使用CAT机制更加准确地关注上下文中与方面词相关的隐藏特征信息.

4)输出层:对CAT机制层输出的特征向量使用Softmax分类器进行不同情感类别概率的计算.

1.1 词嵌入层

BERT模型主要由输入嵌入、双向Transformer编码器^[15]和无监督任务三部分构成,其主要原理是通过额外添加的输出层结合语境进行语义微调,能够联合所有层级中的上下文单词训练双向语义表示^[16].其中:输入嵌入负责对输入句子进行标记处理;Transformer编码器是BERT的核心部分,由输入、自注意力和前馈神经网络组成,负责捕捉上下文中各个词之间的特征信息;无监督任务负责对Transformer编码器输出的隐藏向量进行预训练.而自注意力部分是Transformer中最重要的模块,它能够将在文本中的每个输入单词依次与其他单词进行建模,从而直接捕获隐藏层状态之间的依赖关系^[17].随后将得到的隐藏向量再传递到前馈神经网络部分,进行求和、归一化与前向反馈等流程后,最终得到输出词向量.

方面级隐式情感分析任务的目的是预测句子中给定方面所表达的情感,然而传统的BERT模型无论是输入嵌入、Transformer编码器还是无监督任务部分中都没有能够感知方面词的结构设计,这导致其在预训练过程中可能会丢失与方面词相关的上下文信息.为了解决这个问题,本文考虑从BERT的内部结构进行改进,使其获得方面词感知能力.但在实际情形中发现Transformer编码器和无监督任务部分的结构设计紧凑、逻辑严谨、修改难度较大,如果添加的参数过多很可能导致模型的复杂度大幅增加,出现过拟合现象,反而会降低原有的优秀建模能力;而输入嵌入部分结构则较为灵活,由于直接与输

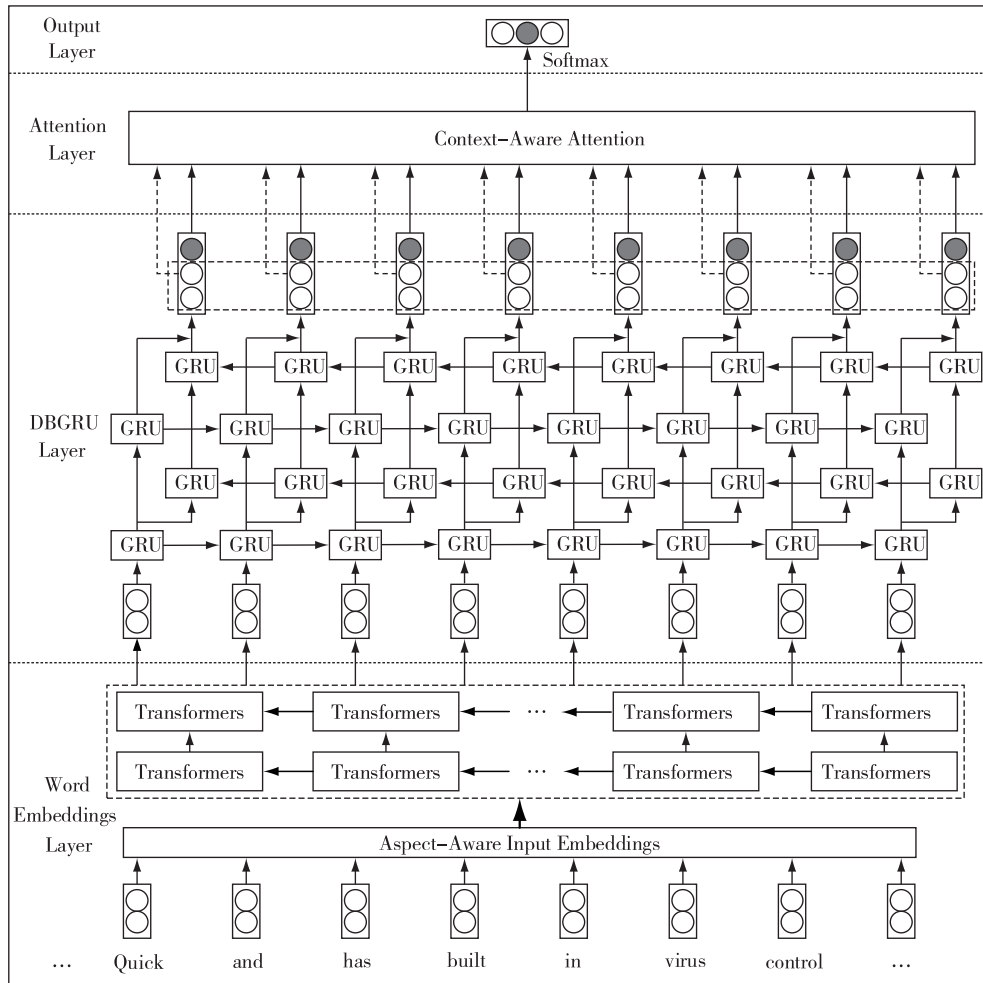


图 1 DCAB 模型结构

Fig. 1 Structure of DCAB model

入句子接触,添加方面词信息更加方便,且在修改的过程中不会影响到 BERT 原有的建模能力.因此为了使 BERT 在拥有方面词感知能力的同时保持其优秀的语言建模功能,本文通过修改输入嵌入部分,将 BERT 调整为能够感知上下文中与方面词相关信息的 AABERT.

AABERT 的改进主要分为三个步骤:首先,对于输入句子 $s = \{w_1, w_2, \dots, w_n\}$, 每次输入结束时的标点符号(比如句号‘.’)可以认为是一个弱分隔符,同时将给定的方面词插入到标点符号和最后的结束标记([SEP])之间,在标记嵌入(Token Embeddings)过程中对所有输入词汇(包括正常词汇和特殊符号)进行嵌入,就能生成对应的标点符号和方面词的位置标记,即提供了一个给定方面词的感知信号.其次在输入中再采用一个显式分隔符[SEP]来分隔输入的句子和方面词,进一步加深给定方面词的信号,它在位置嵌入(Position Embeddings)中的初始化标

记与句尾的[SEP]设置相同.最后将方面词的片段嵌入(Segment Embeddings)标记与前句的片段嵌入标记分别设置为 A、B,即在嵌入过程中将前句与方面词视为两个不同的句子.其输入部分结构如图 2 所示.

改进后,输入句的格式被设置为 $I_k = [\text{CLS}] + w_k + [\text{SEP}] + w_{\text{aspect}} + [\text{SEP}]$.在预训练过程中,先将输入句嵌入 I_k 输入到 Transformer 编码器中进行编码,其过程如式(1)所示:

$$\mathbf{x}_k, \dots, \mathbf{x}_{\text{aspect}} = \text{TransEnc}(I_k) \quad (1)$$

其中, \mathbf{x}_i 为输入句单词的隐藏输出向量, $\mathbf{x}_{\text{aspect}}$ 为方面词的隐藏输出向量.再将得到的隐藏向量送入掩码语言模型(Masked Language Model, MLM)和下一句预测(Next Sentence Prediction, NSP)两个无监督任务中来完成词向量的预训练^[8].MLM 会随机对单词遮掩(mask),让模型通过语境学习被遮掩的单词,它很好地利用了语境中的信息,但是不能直观地

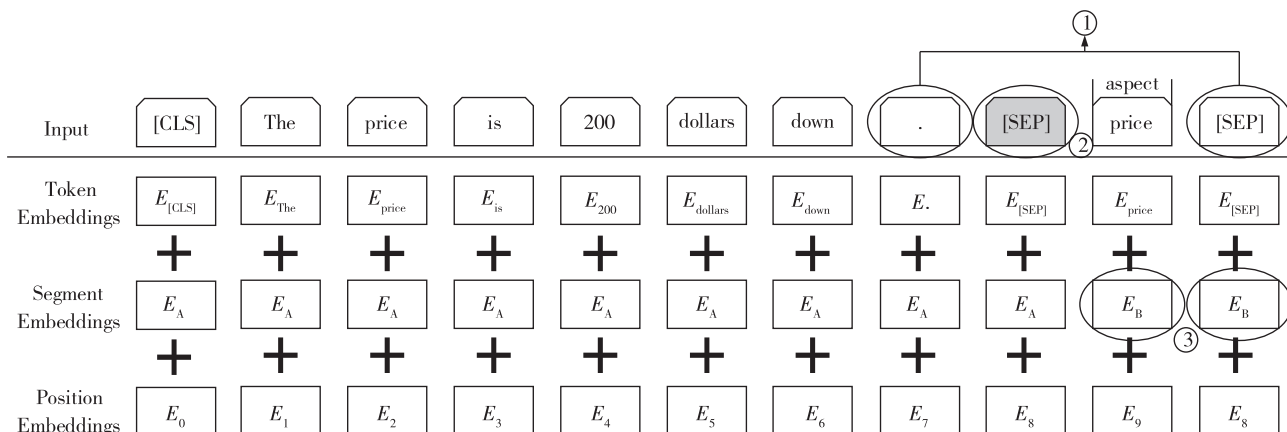


图2 AABERT输入序列

Fig. 2 AABERT input sequence

判断文本对之间的逻辑关系,而 NSP 解决了这个问题,它通过判断后句即方面词是否为前句的下一个句子,挖掘了句子之间的逻辑关系,最终捕获了与方面词相关的词向量表示 $\{e_1, e_2, \dots, e_n\}$.

1.2 深度双向门控循环单元编码层

门控循环单元 (Gated Recurrent Unit, GRU)^[14] 是由长短期记忆网络 (Long Short-Term Memory, LSTM) 改进而来的,它能够获取文本语境中的语义信息,并且相比于 LSTM 拥有更高的信息获取效率.但是 GRU 只能传递某一时刻中词向量序列的语义信息,忽略了全局文本中的语境所带来的影响,而 BiGRU 解决了这个问题,它由正反向 GRU 最后一个状态向量连接而成,可以提取正反向时间序列中的隐藏状态.DBGRU 由多层 BiGRU 叠加而来,其中每层 BiGRU 的输出将作为下一层的输入.本文采用由两层 BiGRU 组成的 DBGRU 来进一步挖掘方面级隐式情感语料中的深层语义信息.

GRU 的更新公式如下:

$$z_t = \sigma(\mathbf{W}_z \mathbf{e}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z), \quad (2)$$

$$r_t = \sigma(\mathbf{W}_r \mathbf{e}_t + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r), \quad (3)$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \mathbf{e}_t + \mathbf{U}_h (r_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h), \quad (4)$$

$$\mathbf{h}_t = z_t \odot \mathbf{h}_{t-1} + (1 - z_t) \odot \tilde{\mathbf{h}}_t, \quad (5)$$

其中: z_t, r_t 分别为 t 时刻的更新门和重置门的输出向量; σ 为 sigmoid 激活函数; $\mathbf{W}_z, \mathbf{W}_r, \mathbf{W}_h, \mathbf{U}_z, \mathbf{U}_r, \mathbf{U}_h$ 为权重矩阵; $\mathbf{b}_z, \mathbf{b}_r, \mathbf{b}_h$ 为偏置项; \mathbf{e}_t 为 t 时刻的输入向量; $\mathbf{h}_t, \mathbf{h}_{t-1}$ 分别为 t 时刻和 $t-1$ 时刻的细胞输出; $\tilde{\mathbf{h}}_t$ 为 t 时刻细胞的隐藏层输出; \odot 为 Hardamard 乘积运算.

BiGRU 的计算公式如下:

$$\overrightarrow{\mathbf{h}}_t = \overrightarrow{\text{GRU}}(\overrightarrow{\mathbf{h}}_{t-1}, \mathbf{e}_t), \quad (6)$$

$$\overleftarrow{\mathbf{h}}_t = \overleftarrow{\text{GRU}}(\overleftarrow{\mathbf{h}}_{t-1}, \mathbf{e}_t), \quad (7)$$

$$\mathbf{h}_t = [\overrightarrow{\mathbf{h}}_t, \overleftarrow{\mathbf{h}}_t], \quad (8)$$

DBGRU 每个单元如下:

$$\overrightarrow{\mathbf{h}}_t^n = \overrightarrow{\text{GRU}}(\overrightarrow{\mathbf{h}}_{t-1}^n, \overrightarrow{\mathbf{h}}_{t-1}^{n-1}), \quad (9)$$

$$\overleftarrow{\mathbf{h}}_t^n = \overleftarrow{\text{GRU}}(\overleftarrow{\mathbf{h}}_{t-1}^n, \overleftarrow{\mathbf{h}}_{t-1}^{n-1}), \quad (10)$$

$$\mathbf{h}_t^n = [\overrightarrow{\mathbf{h}}_t^n, \overleftarrow{\mathbf{h}}_t^n], \quad (11)$$

其中: $\overrightarrow{\mathbf{h}}_t^{n-1}$ 与 $\overleftarrow{\mathbf{h}}_t^{n-1}$ 分别是 t 时刻第 $n-1$ 层 $\overrightarrow{\text{GRU}}$ 与 $\overleftarrow{\text{GRU}}$ 的输出; $\overrightarrow{\mathbf{h}}_{t-1}^n$ 与 $\overleftarrow{\mathbf{h}}_{t-1}^n$ 分别是 $t-1$ 时刻第 n 层 $\overrightarrow{\text{GRU}}$ 与 $\overleftarrow{\text{GRU}}$ 的输出; $\overrightarrow{\mathbf{h}}_t^n$ 与 $\overleftarrow{\mathbf{h}}_t^n$ 分别是 t 时刻第 n 层 $\overrightarrow{\text{GRU}}$ 与 $\overleftarrow{\text{GRU}}$ 的输出; \mathbf{h}_t^n 为第 n 层 BiGRU 的最终输出向量.

1.3 语境感知注意力机制

在方面级语料中往往存在着不少具有情感倾向的词汇,它们对方面的情感极性有着不小的影响.为了提高情感分析的准确率,以往模型通常会采用一般注意力机制捕捉这些情感特征并分配其较大权重,其注意力权重向量 \mathbf{u}_h 计算过程如下:

$$\mathbf{u}_h = \tanh(\mathbf{W}_h \mathbf{h}_t^n + \mathbf{b}_h), \quad (12)$$

其中: \tanh 为激活函数; $\mathbf{W}_h \in \mathbf{R}^{n \times d}$ 为可训练的权重参数矩阵, d 为单词嵌入的维度; \mathbf{b}_h 为偏置项.本文在应用一般注意力机制后发现分析效果不太理想,原因可能是:1) 本文研究的方面级隐式情感语料与普通的方面级情感语料最大的不同在于句子中没有明显的情感词,这会导致一般注意力机制无法准确地捕获相关词向量;2) DBGRU 得到的特征向量为多层,其中含有不同类型的语义和句法信息,而一般注意力权重向量 \mathbf{u}_h 仅由一个可训练的参数矩阵 \mathbf{W}_h 控制,错失了有效利用这些深层语境信息的机会.由于

注意力机制中最重要的部分是计算权重系数,权重系数越高,此部分信息越重要.因此,为了有效突出方面级隐式情感文本中隐含的重要信息,本文对一般注意力机制进行了改进,将 DBGRU 中的多层语境信息引入到权重参数中,从而构造了 CAT 机制.CAT 的结构如图 3 所示.

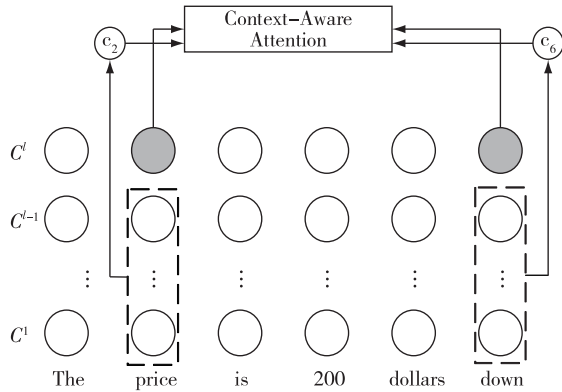


图3 语境感知注意力网络结构

Fig.3 Context-aware attention network structure

$$\hat{W}_h = (1 - \lambda_w) W_h + \lambda_w C U_w, \quad (13)$$

$$\lambda_w = \sigma(W_h V_h + C U_w V_c), \quad (14)$$

$$C = [C^1, \dots, C^{l-1}], \quad (15)$$

其中: $\hat{W}_h \in \mathbf{R}^{n \times d}$ 为语境感知注意力的权重参数矩阵; σ 表示 sigmoid 逻辑函数; $\lambda_w \in \mathbf{R}^{n \times 1}$ 为调节 \hat{W}_h 大小的学习参数; $U_w \in \mathbf{R}^{(l-1)d \times d}$, $V_h \in \mathbf{R}^{d \times 1}$ 和 $V_c \in \mathbf{R}^{d \times 1}$ 为可训练的参数矩阵; $C \in \mathbf{R}^{n \times (l-1)d}$ 为语境矩阵, l 为 BDGRU 编码器输入的特征向量 h_t 的层数,其中高层的隐藏状态可以提供语境的语义信息,低层的隐藏状态可以提供语境的句法信息.假设特征矩阵第 l 层 C^l 为当前输入层,注意力机制在分配权重时,通过将输入层下面第 1 到第 $l-1$ 层的语境特征 $[C^1, \dots, C^{l-1}]$ 引入到权重参数矩阵中,使其在感知输入层信息的同时,还能感知到与输入特征相关的深层语义和句法信息,从而更加精准地分配注意力权重到特征向量上.

此外,本文还在新的权重参数 \hat{W}_h 中分配了一个学习参数 λ_w ,它的作用是明确量化每个隐藏表示与语境矩阵 C 对注意力权重预测的贡献.参数 λ_w 自身也受到原参数矩阵 W_h 和语境矩阵 C 的影响,这样设计的好处在于不仅能对注意力权重的分配实现更加精准地控制,还能使注意力机制在面对不同上下文时的建模更加灵活.与此同时,本文将学习参数 λ_w 的值限定为 $[-1, 1]$,可以有效预防因 \hat{W}_h 量级过大

而导致模型梯度过小的问题.

由此可以得到新的语境感知注意力权重向量 \hat{u}_h :

$$\hat{u}_h = \tanh(\hat{W}_h h_t^n + \hat{b}_h). \quad (16)$$

继续对权重向量 \hat{u}_h 进行归一化处理,就可以得到关于特征 h_t 的注意力分数 α_t ,再将注意力分数与深层特征矩阵 $H(H = h_1^n, h_2^n, \dots, h_t^n)$ 中的子向量进行加权计算求和,可以得到文本特征向量 s_h :

$$\alpha_t = \frac{\exp(\hat{u}_{ht})}{\sum_{t=1}^n \exp(\hat{u}_{ht})}, \quad (17)$$

$$s_h = \sum_{t=1}^n \alpha_t h_t^n. \quad (18)$$

1.4 输出层

本文将 CAT 机制输出的特征向量 s_h 输入到 Softmax 分类层中,并计算其在不同情感极性中的概率,其计算过程如下:

$$g = \text{Softmax}(W s_h + b), \quad (19)$$

其中, W 为权重矩阵, b 为偏置项, g 为模型最终的输出向量.此外,选用了基于 L_2 正则化的交叉熵损失函数作为损失函数,其计算过程如下:

$$l = - \sum_{i \in D} \sum_{j \in E} g_i^j \ln(\hat{g}_i^j) + \lambda \| \varepsilon^2 \|^2, \quad (20)$$

其中 i 为数据样本的索引, j 为情感类别的索引, D 为训练集的容量大小, E 为数据集的种类数, g 为实际情感极性, \hat{g} 为模型预测的情感极性, λ 为 L_2 正则化参数, ε 为模型中需要训练的参数, $\lambda \| \varepsilon^2 \|^2$ 为交叉熵正则项.

2 实验

2.1 实验数据集与实验环境

本文选取的方面级隐式情感数据集来自 SemEval2014 task4 数据集^[18]中的隐式情感语料.其中: Restaurant 主题中的隐式数据占比达 27.47%,被称为餐厅隐式情感表达 (Restaurant Implicit Sentiment Expression, Rest_ISE); Laptop 主题中的隐式数据占比达 30.09%,被称为笔记本电脑隐式情感表达 (Laptop Implicit Sentiment Expression, Lap_ISE).这些数据说明 ABSA 数据集中广泛存在隐式情感语料,值得研究.此外,为了保证本文模型能够充分捕捉方面级隐式情感的特征,引入了数据集 Yelp 和 Amazon^[19]用于 AABERT 的预训练.以上数据集都将语料标记为积极、消极和中性三种情感极性.其具体数据

分布如表 1 所示.

表 1 数据集中的样本标签分布

Table 1 Sample label distribution in datasets

数据库	积极	消极	中性	总计
Rest_ISE-训练集	619	181	230	1 030
Rest_ISE-测试集	174	46	47	267
Rest_ISE	793	227	277	1 297
Lap_ISE-训练集	305	142	267	714
Lap_ISE-测试集	93	46	35	174
Lap_ISE	398	188	302	888
Yelp	117×10^4	39×10^4		156×10^4
Amazon	38×10^4	13×10^4		51×10^4

本文的实验环境如表 2 所示.

表 2 实验环境

Table 2 Experimental environment

实验环境	具体信息
操作系统	Microsoft Windows 10 (64 位)
显卡	NVIDIA GeForce RTX 3060
内存	16 GB
开发语言	Python 3.8
开发平台	Pytorch 1.6.0

2.2 实验参数与评价指标

模型提取方面级隐式情感特征的能力不仅与自身的设计结构相关,还会受到参数设置的影响,合理的参数能够有效提升模型的实验效果,因此参数调优是实验过程中不可或缺的一个步骤.本文在数据集 Rest_ISE 上以对实验结果影响较大的可调参数随机失活率(Dropout)与 DBGRU 隐藏层参数维度为例进行分析,使用控制变量法,在两种参数上分别选取了 10 个大小不同的指标进行对比,每个指标进行了 10 次反复实验并计算准确率的平均值作为结果.本文从对比实验中选取了具有代表性的 5 个指标,其实验结果如表 3 所示.

表 3 参数设置对准确率的影响

Table 3 Effect of parameter setting on accuracy

Dropout	准确率/%	DBGRU 隐藏维度	准确率/%
0.05	63.14	32	62.87
0.1	68.52	64	67.31
0.3	68.23	128	68.52
0.5	66.86	256	66.34
0.8	62.16	512	60.58

Dropout 在模型训练阶段的前向传播过程中,让

某些神经元的激活值以一定的概率停止工作,在一定程度上可以避免过拟合现象.由对比实验可知,DCAB 模型的 Dropout 取值为 0.1 时模型的准确率较高,其原因在于当 Dropout 值设置过小时,缓解过拟合现象的能力并不强,而当 Dropout 值设置过大时,又会使激活值失活较多从而丢失部分特征,导致准确率下降.DBGRU 的隐藏层参数维度即提取到的特征维度,由对比实验可知,当隐藏层参数维度设置为 128 时模型效果较好,其原因在于当隐藏层参数维度过小时,模型的特征提取能力较弱,而当维度过大时,可能会出现过拟合现象,并且模型的训练时间大大增加.

经过对比实验确定 DCAB 模型的参数后,在数据集 Rest_ISE 上对模型进行训练.由于本文模型使用了 12 层 Transformer,多头自注意力的个数为 12,因此 AABERT 的隐藏层参数维度为 768.当模型参数迭代至最优时需要大约 4 个 Epoch,并且使用了 Adam 优化器进行优化.模型整体参数如表 4 所示.

表 4 实验参数

Table 4 Experimental parameters

参数	值	参数	值
AABERT_维度	768	学习率	5E-5
Transformer_层数	12	Batch_size	16
DBGRU_维度	128	Epochs	4
Dropout	0.1	优化器	Adam

本文选取了准确率(Accuracy, A)和 F1 值作为模型的评价指标,具体公式如下:

$$P = \frac{TP}{TP+FP}, \quad (21)$$

$$R = \frac{TP}{TP+FN}, \quad (22)$$

$$A = \frac{TP+TN}{TP+FP+TN+FN}, \quad (23)$$

$$F1 = \frac{2PR}{P+R}. \quad (24)$$

2.3 对比模型

由于目前隐式情感分析研究较少,本文选用了一些常见的显式情感分析模型作为对比模型的补充.对比实验具体模型如下:

1) TD-LSTM^[20]:模型将目标信息合并到 LSTM 中,对句子中的目标及其上下文进行建模.

2) IIDYT^[11]:采用基于深层语境化(ELMo)的嵌入方法,将得到的词向量表示传递到 BiLSTM 中,

再结合注意力机制去提取相关的隐式语义信息。

3) RAM^[21]:该模型使用 GloVe 嵌入,将得到的词向量传递到 DBLSTM 中编码,再使用注意力机制捕获深层语义特征中的重要信息,最终识别评论句中方面词的情感极性。

4) GloVe+BiGRU+Attention(GBA)^[22]:一种融合了 BiGRU 和注意力机制的模型,能够在 GloVe 词嵌入过程中保留方面词向量的位置信息并消除其位置嵌入。

5) Amobee^[13]:使用 Transformer 编码器进行预训练,将得到的词向量传递到 BiLSTM 与注意力机制相结合的框架中分析隐式语料的情感极性。

6) BERT+BiGRU+Attention(BBA)^[23]:该模型将 BERT 模型作为输入层,BiGRU 模型作为隐藏层,再结合注意力机制提取其中的重要隐式情感特征,最后用 Softmax 作为输出层来进行情感分类。

2.4 实验结果与分析

对比实验的结果如表 5 所示。由表 5 可知,在数据集 Rest_ISE 和 Lap_ISE 中,本文提出的 DCAB 模型的准确率和 F1 值均高于其余基线模型,说明 DCAB 模型的性能要优于其他基线模型。TD-LSTM 仅对目标词及其上下文进行建模,IIIDYT 在 BiLSTM 的基础上加入了注意力机制,在提取上下文隐式情感特征的同时关注了其中与方面词相关的重要信息,因此性能要优于 TD-LSTM。RAM 在 IIIDYT 的基础上使用 DBLSTM 代替 BiLSTM 作为模型的编码器,它能够提取文本中更深层次的语义特征,所以效果较 IIIDYT 有所提升。而 GBA 使用了 BiGRU 对词向量进行建模,BiGRU 相比 BiLSTM 参数更少更容易收敛,因此 GBA 模型的精度在一定程度上要优于 IIIDYT 模型。另外,Amobee 和 BBA 模型都是基于 BERT 或 Transformer 结构的模型,Transformer 相比于 RNN 能直接并行计算序列数据,为不同的头分配不同的任务,从而更好地捕捉上下文中隐藏层状态之间的依赖关系,因此性能较其他仅使用 RNN 结构的模型有了较大提升,进而也说明本文使用 AABERT 作为预训练模型是合理的。

本文提出的 DCAB 模型在两个数据集上相比性能最优的基线模型 BBA 准确率分别提升 2.60 和 1.28 个百分点;F1 值分别提升 4.16 和 1.58 个百分点。其原因在于 DCAB 模型改进了提取与方面词相关的上下文信息的方法,即使用 AABERT 预训练模

表 5 对比模型分类结果

Table 5 Comparison of model classification results %

模型	Rest_ISE				Lap_ISE			
	P	R	F1	A	P	R	F1	A
TD-LSTM			54.64	53.71			52.92	37.68
IIIDYT	57.87	56.37	58.22	56.23	55.81	55.87	56.83	48.24
RAM			60.80	57.81			60.35	55.25
GBA	58.35	59.01	60.89	56.56	64.22	65.15	65.95	60.66
Amobee	62.77	62.74	63.21	62.54	69.59	70.13	71.14	73.43
BBA	66.53	64.29	67.33	65.92	77.37	73.65	78.89	75.36
DCAB	70.36	70.55	71.49	68.52	79.26	76.31	80.47	76.64

型生成与方面词相关的词向量,AABERT 在基础 BERT 的输入嵌入结构中引入了方面词,通过在标记嵌入、位置嵌入和片段嵌入中设置特殊标记格式从而加强方面词的信号,使模型获得方面词信息的感知能力,并且构造了语境感知注意力机制学习上下文中与方面词相关的深层特征,CAT 机制在分配注意力权重时能够感知上下文隐藏状态中的语义信息和句法信息,更加精准地关注与方面词相关的深层信息。综上所述,DCAB 模型是有效的,不仅能够在预训练过程中保留与方面词相关的上下文信息,还能准确提取上下文中与方面词相关的深层特征信息。

2.5 模型分析实验

2.5.1 不同词向量对模型性能的影响

为了验证 AABERT 预训练模型对文本向量化表示的能力,本文将 DCAB 中的 AABERT 词向量模型分别替换为 GloVe、ELMo、BERT,并在 Rest_ISE 数据集上进行对比实验,结果如图 4 所示。

由图 4 可知:GloVe 作为静态词向量模型,针对每个输入单词会生成固定的词向量,忽略了一词多义的情况,因此效果较差;ELMo 是一种动态词向量模型,它将输入句传递到 BiLSTM 中,生成的词向量包含了上下文语义信息,所以其性能要优于 GloVe;BERT 使用了双向 Transformer 语言模型,并结合 MLM 和 NSP 任务,从而生成动态词向量,能够有效解决一词多义的问题,因此效果要优于 ELMo;AABERT 在 BERT 的基础上对输入嵌入部分进行了改进,在引入方面词的同时加强了其信号,使模型获得了方面词相关信息的感知能力,因此性能要优于 BERT。综上,这组实验说明了 AABERT 模型的优越性。

2.5.2 不同注意力机制对模型性能的影响

为了验证语境感知注意力机制的有效性,本文

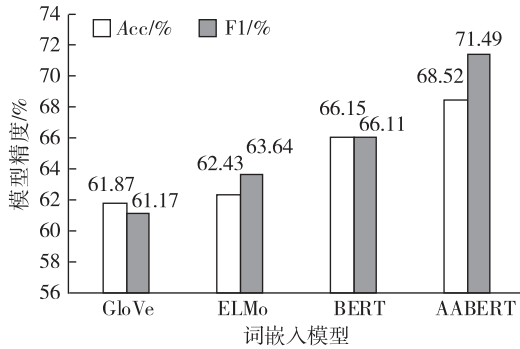


图4 词嵌入模型对比实验结果
Fig. 4 Comparative experimental results of word embedding models

移除注意力机制,建立了AD(AABERT+DBGRU)模型;使用传统的注意力机制代替CAT机制,建立了ADA(AABERT+DBGRU+Attention)模型.对比实验结果如图5所示.

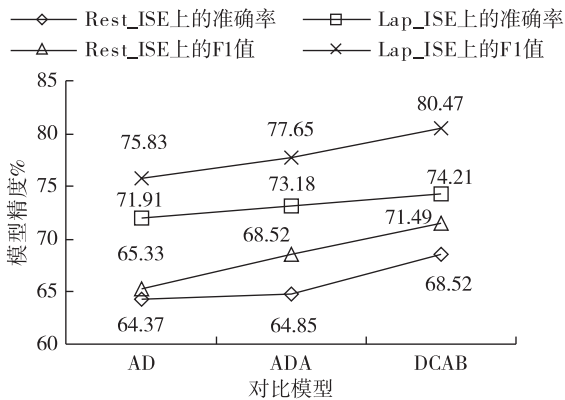


图5 注意力机制对比实验结果
Fig. 5 Comparative experimental results of attention mechanism

由图5可知,在两个数据集 Rest_ISE 和 Lap_ISE 中,DCAB 模型的准确率和 F1 值相比 AD 和 ADA 模型均有所提升,这说明 DCAB 模型的效果相较于 AD 和 ADA 模型要更加优秀.AD 模型仅提取向量中的深层上下文语义特征,效果较差;ADA 模

型在 AD 模型的基础上还关注了上下文语义特征中的重要隐式情感特征,效果次之;而 DCAB 模型在 AD 模型的基础上使用的 CAT 机制不仅能够提取上下文中的重要隐式情感特征,还能充分挖掘深层特征中所包含的语义信息与句法信息.这组实验结果说明了语境感知注意力机制的优越性.

2.5.3 实例分析

为了进一步分析 DCAB 模型捕获的深层隐式情感特征在方面级隐式情感分析中的有效性,本文选取数据集的几个例句,将本文的 DCAB 模型与 RAM 和 BBA 模型进行实例对比分析,结果如表6所示.句子中存在的方面词已用下划线标出,其中加粗的单词需要判断隐式情感极性.

从表6中可以看出,第1及第2句中只存在1个方面词,第4及第5句中存在2个方面词.由第1和第2句可知,对情感极性为积极或消极的方面级隐式情感句的判断准确率较低,难度较大,但是本文提出的 DCAB 模型相比 RAM 和 BBA 模型在积极或消极情感句上的性能最优.由第3句可知,3种模型在中性情感句上的判断都比较准确.由第4句可知, RAM 和 BBA 模型准确率较低,本文推测可能是由于存在着显式情感词"good"修饰方面词"battery life",影响了方面词"cord"隐式情感极性的判断;而 DCAB 的效果较好,其原因在于 CAT 机制能够关注与给定方面词相关的上下文特征信息,并忽略其他方面词的信息.综上,本文提出的 DCAB 模型在方面级隐式情感分析任务上有效提高了深层隐式情感特征提取的准确性.

虽然本文模型在总体效果上有一定的提升,但也引入了一些新的问题.由第5句可知,当句子中存在着两个隐式情感的方面词时,DCAB 模型会将方面词"online chat"的情感极性误判为消极,可能是受到了整个句子消极情感的影响,这是本文模型的局限所在.

表6 方面级隐式情感句对比

Table 6 Contrast of aspect-based implicit emotional sentences

方面级隐式情感句	实际情感	DCAB	RAM	BBA
1.I will <u>upgrade the ram</u> myself (because with this model you can you can do it) later on.	积极	积极	中性	中性
2.If you ask me,for this <u>price</u> it should be included.	消极	消极	中性	中性
3.I hope to edit this in the next few hours,I am going to try to install my own copy of <u>Windows 7</u> .	中性	中性	中性	中性
4.I charge it at night and skip taking the <u>cord</u> with me because of the good <u>battery life</u> .	中性	中性	积极	积极
5.I could not find a phone number anywhere to call an actual live person for <u>tech support</u> and had to result the their <u>online chat</u> .	中性	消极	中性	消极

3 结束语

为解决现有模型可能会在预训练过程中丢失方面词信息和提取深层特征信息不准确的问题,本文提出 DCAB 模型.首先通过 AABERT 生成与方面词相关的词向量,再通过 DBGRU 进行编码,接着使用 CAT 机制感知深层隐藏特征中的语义信息和句法信息,提取上下文中与给定方面词相关的深层特征信息.方面级隐式情感分析的对比实验结果表明,DCAB 模型性能优于基线模型.本文没有考虑到更大范围语境的影响,下一步将研究句子层面的信息是如何影响隐式情感的.

参考文献

References

- [1] 李胜旺,杨艺,许云峰,等.文本方面级情感分类方法综述[J].河北科技大学学报,2020,41(6):518-527
LI Shengwang, YANG Yi, XU Yunfeng, et al. A survey of text aspect-based sentiment classification[J]. Journal of Hebei University of Science and Technology, 2020, 41(6):518-527
- [2] 罗嘉,王乐豪,涂姗姗,等.基于 LSTM-BLS 的突发气象灾害事件中公众情感倾向分析[J].南京信息工程大学学报(自然科学版),2021,13(4):477-483
LUO Jia, WANG Lehao, TU Shanshan, et al. Analysis of public sentiment tendency in sudden meteorological disasters based on LSTM-BLS[J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2021, 13(4):477-483
- [3] Russo I, Caselli T, Strapparava C. SemEval-2015 task 9: CLIPeval implicit polarity of events[C]//Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). Denver, Colorado. Stroudsburg, PA, USA: Association for Computational Linguistics, 2015: 443-450
- [4] Choi Y, Wiebe J. +/-EffectWordNet: sense-level lexicon acquisition for opinion inference[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014:1181-1191
- [5] Rhanoui M, Mikram M, Yousfi S, et al. A CNN-BiLSTM model for document-level sentiment analysis[J]. Machine Learning and Knowledge Extraction, 2019, 1(3): 832-847
- [6] Pennington J, Socher R, Manning C. Glove: global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014: 1532-1543
- [7] Peters M, Neumann M, Iyyer M, et al. Deep contextualized word representations[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans, Louisiana. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018:2227-2237
- [8] Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[J]. arXiv e-print, 2018, arXiv:1810.04805
- [9] Li Z Y, Zou Y C, Zhang C, et al. Learning implicit sentiment in aspect-based sentiment analysis with supervised contrastive pre-training[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana, Dominican Republic. Stroudsburg, PA, USA: Association for Computational Linguistics, 2021
- [10] Klinger R, de Clercq O, Mohammad S M, et al. IEST: WASSA-2018 implicit emotions shared task[J]. arXiv e-print, 2018, arXiv:1809.01083
- [11] Balazs J, Marrese-Taylor E, Matsuo Y. IIDYDT at IEST 2018: implicit emotion classification with deep contextualized word representations[C]//Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Brussels, Belgium. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018:50-56
- [12] Chronopoulou A, Margatina A, Baziotis C, et al. NTUA-SLP at IEST 2018: ensemble of neural transfer methods for implicit emotion classification[C]//Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Brussels, Belgium. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018:57-64
- [13] Rozental A, Fleischer D, Kelrich Z. Amobee at IEST 2018: transfer learning from language models[C]//Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Brussels, Belgium. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018:43-49
- [14] 杨春霞,徐奔,陈启岗,等.融合深度 BiGRU 与全局图卷积的方面级情感分析模型[J].小型微型计算机系统,2023,44(1):132-139
YANG Chunxia, XU Ben, CHEN Qigang, et al. Aspect level sentiment analysis model based on deep BiGRU and global graph convolution [J]. Journal of Chinese Computer Systems, 2023, 44(1):132-139
- [15] Wolf T, Debut L, Sanh V, et al. Transformers: state-of-the-art natural language processing[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Stroudsburg, PA, USA: Association for Computational Linguistics, 2020:38-45
- [16] 林孟达,李书豪.融合 BERT 嵌入与注意力机制的方面情感分析[J].现代电子技术,2022,45(12):130-136
- [17] LIN Mengda, LI Shuhao. Aspect sentiment analysis integrating BERT embedding and attention mechanism[J]. Modern Electronics Technique, 2022, 45(12):130-136
- [17] 范昊,何灏.融合上下文特征和 BERT 词嵌入的新闻

标题分类研究[J].情报科学,2022,40(6):90-97

FAN Hao, HE Hao. News title classification based on contextual features and BERT word embedding [J]. Information Science, 2022, 40(6) : 90-97

[18] 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, Ireland. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014: 27-35

[19] He R N, McAuley J. Ups and downs: modeling the visual evolution of fashion trends with one-class collaborative filtering [C] // Proceedings of the 25th International Conference on World Wide Web. New York: ACM, 2016: 507-517

[20] Tang D, Qin B, Feng X, et al. Effective LSTMs for target-

dependent sentiment classification [J]. arXiv e-print, 2015, arXiv: 1512. 01100

[21] Chen P, Sun Z Q, Bing L D, et al. Recurrent attention network on memory for aspect sentiment analysis [C] // Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark. Stroudsburg, PA, USA: Association for Computational Linguistics, 2017

[22] Yadav R K, Jiao L, Goodwin M, et al. Positionless aspect based sentiment analysis using attention mechanism [J]. Knowledge-Based Systems, 2021, 226: 107136

[23] Liu Y, Lu J H, Yang J, et al. Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax [J]. Mathematical Biosciences and Engineering: MBE, 2020, 17(6): 7819-7837

Aspect-based implicit sentiment analysis model based on BERT and attention mechanism

YANG Chunxia¹ HAN Yu¹ CHEN Qigang¹ MA Wenwen¹

¹ School of Automation/Jiangsu Key Laboratory of Big Data Analysis Technology/Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science & Technology, Nanjing 210044, China

Abstract There are quite a few comment sentences without emotional words in aspect-level emotional texts, and the study of their emotions is called aspect-level implicit sentiment analysis. The existing models have the problems that the context information related to aspect words may be lost in the pre-training process, and the deep features in the context cannot be accurately extracted. Aiming at the first problem, this paper constructs an aspect-aware BERT pre-training model, and introduces aspect words into the input embedding structure of basic BERT to generate word vectors related to aspect words. Aiming at the second problem, this paper constructs a context-aware attention mechanism. For the deep hidden vectors obtained from the coding layer, the semantic and syntactic information is introduced into the attention weight calculation process, so that the attention mechanism can more accurately assign weights to the context related to aspect words. The results of comparative experiments show that the proposed model outperforms the baseline model.

Key words aspect-based sentiment analysis (ABSA) ; implicit sentiment analysis ; gated recurrent unit (GRU) ; attention mechanism ; BERT model