◆大数据技术◆

基于情感信息预处理和Bi-GRU的虚假评论识别模型*

张玉莹[1,2],朱广丽[1,2**],张友强[1,2],孙争艳[1,2],张顺香[1,2]

(1.安徽理工大学计算机科学与工程学院,安徽淮南 232001;2.合肥综合性国家科学中心人工智能研究院,安徽合肥 230088)

摘要:虚假评论识别在电子商务、社交媒体等领域具有重要的应用价值。尽管现有虚假评论识别模型融合了文本的情感信息,但在预训练过程中忽视了对情感信息的提取,导致准确率不高。针对此问题,本文提出一种基于情感信息预处理和双向门控循环单元(Bidirectional Gated Recurrent Unit,Bi-GRU)的虚假评论识别模型(FR-SG),用于提高虚假评论识别的准确率。首先,通过Albert模型获取文本的语义向量;然后,使用词频-逆文本频率(Term Frequency-Inverse Document Frequency,TF-IDF)和K-均值(K-means++)聚类的方法从评论中挖掘情感种子词,基于种子词对文本中的属性词和情感词进行掩码(mask);接着,使用面向情感的目标优化函数,将情感信息嵌入到语义表示中,生成情感向量;最后,将这两组向量的拼接结果输入虚假评论识别网络中,得到文本的分类结果。实验结果表明,相较于Bi-GRU+Attention模型,FR-SG提高了虚假评论识别的准确率。

关键词:虚假评论识别;情感信息预处理;情感种子词;Bi-GRU;目标优化函数

中图分类号:TP391 **文献标识码**:A **文章编号**:1005-9164(2023)01-0169-08

DOI:10.13656/j.cnki.gxkx.20230308.018

随着电商平台的发展,在线评论数据呈爆发式增长,评论文本[1]中包含用户大量的情感色彩。商家为了增加销量,往往会通过雇佣特定写手的方式,鼓吹自己店铺或诋毁同行店铺的产品,误导用户的消费决策,因此对虚假评论的识别非常有必要。在深度学习技术尚未成熟之前,虚假评论识别任务大多由机器学习技术来完成,包括无监督矩阵运算算法和监督学习的神经网络方法[2,3]等。虚假评论数据集一般来自公开的网站评论数据集或由研究者爬取的网站评论数据集。表1列举了一个公开的虚假评论数据集中的4个实例。从表1可以看出,虚假评论中往往会存在一些强烈的情感倾向性单词(粗体的单词),因此情感信息是识别虚假评论的一项重要特征。

根据调查研究[4],在虚假评论识别任务中,预处理阶段主要对文本进行分词、过滤停用词等处理,而忽略了对情感信息的提取。根据现阶段虚假评论领

收稿日期:2022-09-22 修回日期:2022-09-26

*国家自然科学基金面上项目(62076006)和安徽省高校协同创新项目(GXXT-2021-008)资助。

【第一作者简介】

张玉莹(1997-),女,在读硕士研究生,主要从事文本分类与情感分析研究。

【**通信作者】

朱广丽(1971-),女,副教授,主要从事文本挖掘与情感计算研究,E-mail:glzhu@aust.edu.cn。

【引用本文】

张玉莹,朱广丽,张友强,等.基于情感信息预处理和Bi-GRU的虚假评论识别模型[J].广西科学,2023,30(1):169-176.

ZHANG Y Y,ZHU G L,ZHANG Y Q,et al. Fake Review Detection Model Based on Pre-training of Sentiment Information and Bi-GRU [J]. Guangxi Sciences,2023,30(1):169-176.

表 1　虚假评论和真实评论实例

Table 1　Examples of fake reviews and real reviews

标签 Label	评论文本 Review text
Fake review	**Wow**! Mike Ditka′s in Chicago is one of my **favorite** restaurants,it′s seriously amazing…
Real review	I would return to Mike Ditka′s though if I had a bigger budget…
Fake review	I was **extremely happy** with my first visit to Weber Grill…
Real review	We choose the Weber Grill due to it having a great outdoor eating area…

Note:bold words have strong sentiment orientation

域的研究成果,需要考虑以下两点:①如何在预处理阶段提取情感信息;②如何更加充分地提取虚假评论相关特征[5]。基于上述考虑,为了提高虚假评论识别的准确率,本文提出一种基于情感信息预处理和双向门控循环单元(Bidirectional Gated Recurrent Unit,Bi-GRU)的虚假评论识别模型。模型框架主要分为以下 3 层:①预处理层。将输入文本内容表示成序列化的文本信息,输入到语义编码模块和情感编码模块中,生成语义向量和情感向量并进行拼接融合。②特征提取层。将拼接后的向量输入 Bi-GRU,结合注意力(Attention)机制进一步提取文本的上下文特征信息。③输出层。将特征提取层输出的向量输入到平均池化层,对数据进行降维。最后由 Softmax 激活函数判定是否为虚假评论。

1　相关工作

自 Jindal 等[6]提出虚假评论识别的问题以来,研究者们对这一问题展开了深入的研究,深度学习模型[7]的应用也越来越广泛。对虚假评论的识别按研究对象可分为文本内容识别[8]、虚假评论者识别和虚假评论群组识别。

1.1　文本内容识别

在虚假评论识别的研究中,评论文本内容成为识别虚假评论的主要途径。陆杉等[9]提出将多种不同粒度的文本特征进行融合,丰富语义信息,解决了少量数据情况下模型性能不佳的问题。陶晶晶[10]提出采用并联方式将神经网络识别模型进行混合,实验结果表明该方法识别准确率达到 90.3%。曾致远等[11]根据评论首尾情感极性更加强烈的特点,给予评论文本的头部和尾部更高的权重。李春雨[12]将深度学习的 Bi-GRU 方法与注意力机制相结合,构建垃圾评论的识别模型。Tian 等[13]首次提出在预处理阶段提取情感信息的 Sentiment Knowledge Enhanced Pre-training (SKEP)模型,采用无监督方式获取情感知识,从而将情感信息嵌入语义表示之中。在本文中,SKEP 模型作为对比基线之一。

1.2　虚假评论者识别

在对评论文本内容进行研究的基础上,研究者们发现用户或者商家的特征对虚假评论识别的准确率有影响。Gao 等[14]在英文数据集上结合文本情感特征、情感强度、文本相似度和基于特征加权模型的极端评价指标来识别虚假信息发布者。Barbado 等[15]通过自适应提升算法识别虚假评论者,研究表明此算法模型在 Yelp 数据集上的 $F1$ 值为 82%。孟园等[16]通过迭代修正方法对用户-评论-商户偏差进行改善,增强文本的虚假度。Ruan 等[17]通过集成学习方法将用户基础信息和地理位置序列特征相结合,提高识别虚假评论者的准确率。

1.3　虚假评论群组识别

虚假评论群组指以团队合作的形式,对某个商品批量发布虚假评论。张琪等[18]提出一种带权评论图的水军群组检测及特征分析方法,可以检测到高活跃的“水军群组”。Wang 等[19]在基础特征上增加评论内容与产品信息紧密性的特征,提出图稀疏隐狄利克雷分布(Latent Dirichlet Allocation,LDA)。韩忠明等[20]构建加权用户关系图模型,用于识别大规模的虚假评论群组,实验结果表明,该模型的识别准确率高于非加权用户关系图。

基于以上研究分析,现有模型在预处理阶段大多采用无监督方式挖掘情感词,不能保证情感词的代表性,且没有进一步提取评论文本的重要特征,导致识别准确率不高。本文提出的模型在预处理阶段将情感信息与原文本信息进行融合,并在特征提取阶段提取虚假评论其他特征,有利于提高识别的准确率。

2　文本特征向量的获取

通过语义编码模块和情感编码模块获取文本的语义向量和情感向量,并使用 concat()方法将两组向量进行拼接,生成文本特征词向量。

2.1　语义编码模块

Albert 是 BERT 的优化模型,在保证 BERT 性能的基础上,减少了实验的参数,数据吞吐量更高,拥有更强的语义提取能力,更加适用于文本虚假评论识别任务。

与 BERT 模型相同,Albert 同样需要获取序列

中每个字符向量 E_i、位置向量 P_i,E_i 通过词嵌入的方式获取。因为评论文本均为单句,所以不考虑句子分割信息。

$$P_i(\mathrm{pos},2i)=\sin\left(\frac{\mathrm{pos}}{10000^{2i/d_{\mathrm{model}}}}\right), \tag{1}$$

$$P_i(\mathrm{pos},2i+1)=\sin\left(\frac{\mathrm{pos}}{10000^{2i+1/d_{\mathrm{model}}}}\right), \tag{2}$$

其中,pos 是词的位置索引,当词的向量维度 $d_{\mathrm{model}}=512$ 时,$i=0,1,\cdots,255$。输入序列中每个词通过绝对位置信息的编码方式,获取位置向量 P_i。

$$Z_i=E_i+P_i, \tag{3}$$

其中,Z_i 为字符向量 E_i 与位置向量 P_i 的结合,将 Z_i 输入到 Transformer 中进行编码。

$$A_i=\mathrm{Albert}(Z_i\mid\theta_A), \tag{4}$$

其中,A_i 为语义编码模块输出的词向量,θ_A 为 Albert 预处理模型的参数。本文采用公开参数初始化模型。

2.2 情感编码模块

构建情感编码模块,用于获取文本情感上下文信息。基于 SKEP 模型,在预处理阶段挖掘数据集中代表性情感词,构建情感种子词集。情感编码模块如图 1 所示,图中笑脸代表被掩码(mask)的情感词为积极。模块具有以下 3 个特点。

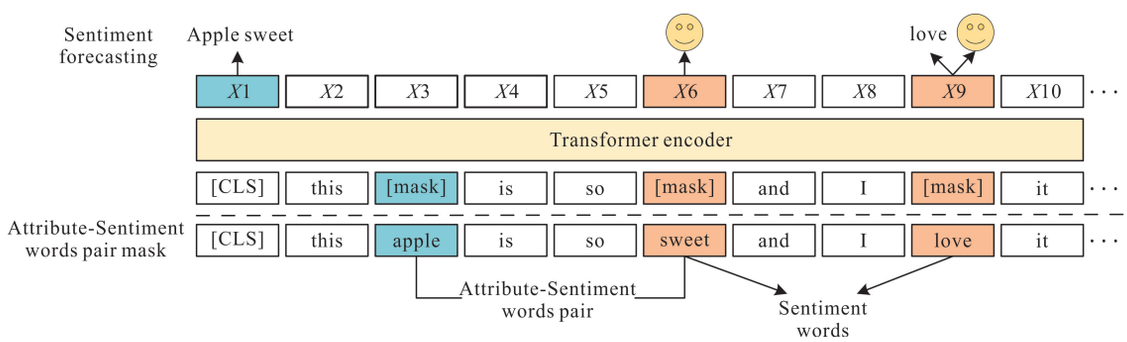

图 1 情感编码模块

Fig. 1 Sentiment coding module

一是采用词频-逆文本频率(Term Frequency-Inverse Document Frequency,TF-IDF)和 K-均值(K-means++)方法获取情感种子词集。首先采用 TF-IDF 方法挖掘数据集中的关键词,选取其中的情感词作为候选词,然后利用 K-means++方法确定最终的情感种子词。

$$\mathrm{TF-IDF}=\mathrm{TF}\times\mathrm{IDF}, \tag{5}$$

其中,TF 为词频,表示词条在文本中出现的频率;IDF 为逆文档词频,用于度量词条在文档中的重要性。

为了确保选取的种子词具有代表性,本文利用 K-means++方法对候选词进行聚类,如算法 1 所示。

算法 1:K-means++词聚类算法

输入:数据集 $D=\{w_1,w_2,\cdots,w_m\}$,聚类数 K,迭代数 N

输出:种子词

① 预设 K 个质心

② 随机选择一个词作为第一个聚类中心 u_1

③ if 词 w_i 在词集 D 中,计算 w_i 到 u_1 语义距离 $d(w_i)$

④ 选取 $d(w_i)$ 较大的词作为新的聚类中心

⑤ end if

⑥ for $(n=1,2,3,\cdots,N)$

⑦ 划分 C 初始化为 $C_t=\varphi,t=1,2,\cdots,k$

⑧ 计算词语 w_i 到各个质心的 $u_j=(j=1,2,3,\cdots,k)$ 的语义距离

⑨ 计算新的质心,直到质心距离不变为止

⑩ end for

⑪ 输出划分 $C=\{C_1,C_2,\cdots,C_k\}$

⑫ end

算法 1 中,步骤①-步骤⑤用于确定初始的 k 个质心,步骤⑥-步骤⑫用于确定最终的情感种子词集。情感种子词集如表 2 所示。

表 2 情感种子词集

Table 2 Sentiment seed word set

积极单词 Positive words	消极单词 Negative words
good, like, well, love, best, better, want, recommend, worth, great, right, fun…	problem, hard, waste, boring, less, junk, poor, least, worst, wrong…

二是基于情感词和属性-情感词 mask。由于成对的词之间并不相互排斥,所以挖掘文本属性-情感词对可以捕捉属性词和情感词之间的关系。根据就近原则(属性词和情感词的距离不超过 3)找出匹配的属性词,与情感词共同组成属性-情感词对,例如<apple,sweet>。基于情感词和匹配的属性-情感词进行 mask,需要保证被 mask 的令牌(token)个数不能超过当前句子 token 总数的 10%;若不足 10%,则 mask 句子中剩余情感词,保证比例补充到 10%。

三是定义了 3 个情感预处理的目标优化函数。将它们用于恢复 Transformer 编码器中被 mask 的情感信息。

$$L = L_{sw} + L_{wp} + L_{ap}, \tag{6}$$

其中,L 为目标优化函数,L_{sw} 为情感词目标优化函数,L_{wp} 为情感极性目标优化函数,L_{ap} 为属性-情感词对目标优化函数。

$$L_{sw} = -\sum_{i=1}^{n} m_i \times y_i^{sw} \log \hat{y}_i^{sw}, \tag{7}$$

$$\hat{y}_i^{sw} = \mathrm{Softmax}(\mathrm{Sent}(Z_i) \cdot W^{sw} + b^{sw}), \tag{8}$$

其中,Z_i 指输入的序列,m_i 用于表示是否为情感词,y_i^{sw} 是原始情感词 x_i 基于整个词汇表的独热编码(one-hot)表示,$\hat{y}_i^{sw}$ 为 Z_i 经过输出层后,由 Softmax 得到的概率分布,W^{sw} 和 b^{sw} 是情感词目标优化函数中输出层的可训练参数。如果第 i 个位置是输入序列 mask 的词,则 $m_i = 1$,否则就为 0。

$$L_{wp} = -\sum_{i=1}^{n} m_i \times y_i^{wp} \log \hat{y}_i^{wp}, \tag{9}$$

$$\hat{y}_i^{wp} = \mathrm{Softmax}(\mathrm{Sent}(Z_i) \cdot W^{wp} + b^{wp}), \tag{10}$$

其中,Z_i 依然指的是输入的序列,m_i 用于表示是否为情感词,y_i^{wp} 为被 mask 的情感词的极性,$\hat{y}_i^{wp}$ 为 y_i^{wp} 的概率估计值,W^{wp} 和 b^{wp} 是情感极性目标优化函数中输出层的可训练参数。可以理解为计算的是另外一种形式的 token 的损失,区别在于它总共分为两类:积极和消极。

$$L_{ap} = -\sum_{i=1}^{A} y_a \log \hat{y}_i^{ap}, \tag{11}$$

$$\hat{y}_i^{ap} = \mathrm{Softmax}(\mathrm{Sent}(Z_i) \cdot W^{ap} + b^{ap}), \tag{12}$$

其中,A 是输入序列被 mask 的属性-情感词对的数量,y_a 是目标属性-情感词对的稀疏表示,$\hat{y}_i^{ap}$ 为 y_a 的概率估计值,W^{ap} 和 b^{ap} 是属性-情感词对目标优化函数中输出层的可训练参数。

以上是情感编码模块内容,将字符向量 E_i、位置向量 P_i 拼接后的序列 Z_i 输入情感编码模块中进行预训练,生成融合上下文情感信息的向量 S_i。

$$S_i = \mathrm{Sent}(Z_i \mid \theta_s), \tag{13}$$

其中,θ_s 为情感编码模块的参数。

2.3 文本特征信息拼接

本文提出的模型在预处理阶段提取评论文本的情感向量 S_i,并使用 concat()将语义向量 A_i 和 S_i 直接拼接起来。拼接后的向量 V_i 不仅获取了文本丰富的情感信息,还保留了原文本的语义信息,有利于特征提取层进一步提取虚假评论其他相关特征。

$$V_i = \mathrm{concat}(\{A_i; S_i\}). \tag{14}$$

3 基于情感信息预处理和 Bi-GRU 的虚假评论识别模型

本文提出的虚假评论识别模型如图 2 所示,将情感向量和语义向量进行融合,以丰富语义信息,充分提取虚假评论重要特征,提高识别准确率。

3.1 预处理层

预处理层是对海量数据进行训练,得到含有更加丰富的语义信息的词向量 Embedding (E)。在评论文本的开头和结尾分别加上特殊字符[CLS]和[SEP],[CLS]用来表示整句话的语义,应用于下游的分类任务;[SEP]为句与句之间的分隔符,代表一句话结束。

首先,将评论文本序列通过语义编码模块生成对应的语义向量;然后,通过情感编码模块生成情感向量;最后,以拼接的方式将语义向量和情感向量融合为 Embedding (E)。

3.2 特征提取层

双向门控循环单元(Bi-GRU)由前向 GRU_l 和后向 GRU_r 组成,通过 Bi-GRU 层进一步提取评论文本特征,输出对应位置上的上下文信息。GRU_l 按顺序(从 V_1 到 V_n)读取序列信息,GRU_r 按逆序(从 V_n 到 V_1)读取序列信息。

$$\overrightarrow{h}_t = \sigma(W_{x\overrightarrow{h}} x_t + W_{\overrightarrow{h}\overrightarrow{h}} \overrightarrow{h}_{t-1} + b_{\overrightarrow{h}}), \tag{15}$$

$$\overleftarrow{h}_t = \sigma(W_{x\overleftarrow{h}} x_t + W_{\overleftarrow{h}\overleftarrow{h}} \overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}), \tag{16}$$

其中,x_t 为融合后的词向量,$\overrightarrow{h}_{t-1}$ 和 $\overleftarrow{h}_{t-1}$ 分别为 $t-1$ 时刻 Bi-GRU 正向和反向输出的序列,$W_{x\overrightarrow{h}}$ 表示 t 时刻输入 Bi-GRU 中 x_t 的权重,$W_{\overrightarrow{h}\overrightarrow{h}}$ 表示 t 时刻输入 Bi-GRU 中 $\overrightarrow{h}_{t-1}$ 或 $\overleftarrow{h}_{t-1}$ 的权重,$b_{\overrightarrow{h}}$ 为 t 时刻所对应的偏置向量,σ 为激活函数,$\overrightarrow{h}_t$ 和 $\overleftarrow{h}_t$ 分别为 t 时刻 Bi-

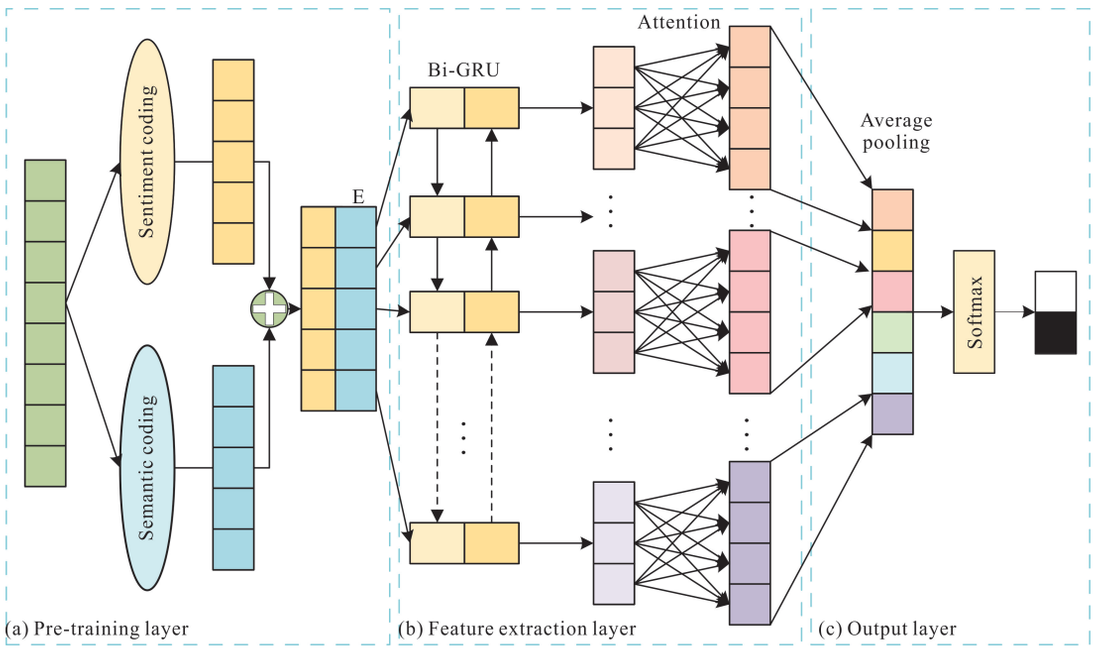

图 2　基于情感信息预处理和 Bi-GRU 的虚假评论识别模型

Fig. 2　Fake review detection model based on the pre-training of sentiment information and Bi-GRU

GRU 正向和反向输出的序列,拼接后得到文本特征向量矩阵,完整的序列中包含每个词的双向信息。

注意力(Attention)机制是一种权重参数的分配机制,通过给序列层中各隐藏节点分配对应的注意力权重,协助模型捕捉词汇、句法等其他重要特征。具体操作为给定一组<key,value>,以及一个目标向量 query。通过计算 query 与每一组 key 的相似性,求出每个 key 的权重系数,再通过对 value 加权求和,得到最终数值。

$$e_t = a(h_t)\,, \tag{17}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T}\exp(e_k)}\,, \tag{18}$$

$$S = \sum_{t=1}^{T}\alpha_t h_t\,, \tag{19}$$

其中,a 是学习函数,隐藏状态序列 h_t 经过学习函数得到其权重 e_t,e_t 做归一化处理,得到注意力的权重分布 α_t,根据注意力权值对所有单词向量做加权线性组合,得到最终的文本特征向量 S。

3.3　输出层

输出层主要由平均池化层和 Softmax 激活函数共同组成。平均池化层能保留整体数据的特征,在保证网络平移不变性的基础上,提升网络的泛化能力。Softmax 又称归一化指数,主要分为两步:①将模型预测的结果以指数函数的形式表示,保证概率的非负性;②将表示后的结果除以所有结果的总和,即占总体的百分比。

$$Q = \mathrm{softmax}(w_T T + b_T)\,, \tag{20}$$

其中,w_T 为平均池化层到输出层的权重系数矩阵,b_T 为相对应的偏置。将平均池化层的结果输入到 Softmax 激活函数中,得到概率值 Q,代表真实或虚假的概率,取概率相对较大的为分类结果。

4　实验与结果分析

4.1　数据集

为验证本文提出模型的效果,选取 Rayana 等[21]从 Yelp.com 构建评论的公开数据集 YelpZIP 和 Ott 等[22]有关酒店的评论数据集进行实验,数据集的信息如表 3 所示。

表 3　数据集信息

Table 3　Dataset information

数据集 Dataset	数据类型 Data type	来源 Source	数量 Count
Ott	Fake review	Crowdsourcing	500
	Real review	TripAdvisor	500
YelpZIP	Fake review	Yelp.com	10 466
	Real review	Yelp.com	19 305

Ott 等[22]通过众包的方式,有偿雇佣有关人员写了 500 条正面评论,属于虚假评论。真实评论为 500 条,来自 TripAdvisor 对芝加哥相同酒店的 6 977 条

点评。在本文中,数据集 Ott 用于评估虚假评论的模型的有效性,YelpZIP 用于整体实验分析。

原始数据集进行清洗和格式转换等处理后,得到初始实验数据集,并将数据分为训练集、验证集和测试集,比例为 8∶1∶1。在实验过程中,采用十折交叉验证(10-fold cross validation)的方式对模型进行调优,将 10 次结果的平均值作为最终结果。

4.2 评价指标

本文采用的评价指标为准确率(Precision)、召回率(Recall)和 $F1$ 值。

4.3 实验环境设置

实验主要使用 Pytorch 深度学习框架编写程序。实验配置环境:CPU 为 AMD Ryzen 7 5800H with Radeon Graphics;GPU 为 NVIDIA GeForce RTX 3060 16 GB;编程语言为 Python 3.7;深度学习框架为 Pytorch 1.11.0。

4.4 参数设置

在实验的预处理模块,本文模型的大部分参数采用默认的配置,模型参数如表 4 所示。为了防止模型的过拟合,采用 Dropout 来降低模型的密度。若超过 1 000 个 batch 后模型学习效果未提升,则提前结束训练。

表 4 模型参数设置

Table 4 Model parameter setting

参数 Parameter	值 Value
Embedding size	128
Hidden size	768
Batch size	16
Epoch	20
Dropout	0.6
Learning rate	0.3

4.5 实验方法

为验证本文提出的 FR-SG 的有效性,首先获取数据集的文本词向量,然后训练虚假评论识别网络并确定模型最优参数,最后使用测试集对模型进行验证。具体实验步骤如下。

步骤 1:加载数据集,将数据集划分为训练集、验证集和测试集。

步骤 2:依次对数据集的样本进行预处理,计算相对位置编码以及字符编码。

步骤 3:采用 TF-IDF 和 K-means++方法获取数据集中的情感词,并根据就近匹配原则找到属性-情感词对。基于情感知识生成融合情感信息的情感向量,并与语义编码模块生成的语义向量进行融合。

步骤 4:通过上下文信息训练虚假评论识别网络,并使用损失函数计算模型的损失,从而优化参数。

步骤 5:保存模型的最优参数,在测试集上对模型的准确率进行测试。

4.6 对比实验

分别采用以下 4 个模型在数据集上进行对比评估,具体实验结果如表 5 所示。

①Bi-GRU:FR-SG 的基础模型,用于解决长期记忆和反向传播中的梯度等问题。

②SKEP[13]:百度研究团队提出的基于情感知识增强的情感预训练算法,主要对情感词和属性词进行 mask,作为情感编码模块的基础。

③联合预训练模型[23]:将 SKEP 和 Roberta 进行融合,得出虚假评论分类结果。

④Bi-GRU + Attention[12]:本文识别网络的核心,用于提取虚假评论特征和上下文语义信息。

表 5 训练对比实验结果

Table 5 Comparative experimental results of training

模型 Model	准确率(%) Accuracy (%)	召回率(%) Recall (%)	$F1$ 值(%) $F1$ value (%)
Bi-GRU	86.9	87.8	86.3
SKEP	90.6	89.2	91.0
Joint pre-training model	91.3	91.3	92.3
Bi-GRU + Attention	92.9	**94.0**	93.2
FR-SG (Ours)	**93.8**	93.6	**93.7**

Note:bold words are the highest percentage of the experimental results in comparison of multiple models

由表 5 可见,FR-SG 的准确率、召回率和 $F1$ 值分别为 93.8%、93.6%和 93.7%。相较于其他模型中效果最好的 Bi-GRU + Attention,FR-SG 的准确率提高了 0.9%,$F1$ 值提高了 0.5%,且相对于 SKEP[13]和联合预训练模型[23],FR-SG 均有效提升了模型性能。

5 结论

本文提出了一种基于情感信息预处理和 Bi-GRU 的虚假评论识别模型,主要有以下贡献:

①采用 TF-IDF 和 K-means++方法构建专属情感种子词集,解决了种子词集与数据集不适配的问题。模型在预处理阶段将评论文本的情感信息与原

语义信息进行融合,解决了在提取情感信息时丢失原语义信息的问题,且直接拼接的方法可以最大程度地保留文本的重要特征。

②特征提取阶段采用 Bi-GRU 和 Attention 机制结合的方法,充分提取虚假评论词汇、句法等其他重要特征,解决了其他模型只考虑评论情感信息而忽视其他特征的问题,提高了虚假评论识别的准确率。

基于本文提出的模型,未来将考虑将更多的虚假评论相关特征,进一步优化模型,推进对虚假评论识别模型的研究;努力提升模型的泛化能力,使模型更好地应用于不同领域的文本虚假评论识别任务中。

参考文献

[1] 汤凌燕,熊聪聪,王媛,等.基于深度学习的短文本情感倾向分析综述[J].计算机科学与探索,2021,15(5):794-811.

[2] MANDHULA T,PABBOJU S,GUGULOTU N. Predicting the customer's opinion on amazon products using selective memory architecture-based convolutional neural network [J]. The Journal of Supercomputing,2020,76(8):5923-5947.

[3] 张国标,李洁.融合多模态内容语义一致性的社交媒体虚假新闻检测[J].数据分析与知识发现,2020,53(5):21-29.

[4] 李菲菲,吴璠,王中卿.基于生成式对抗网络和评论专业类型的情感分类研究[J].数据分析与知识发现,2021,52(4):72-79.

[5] 邱宁佳,杨长庚,王鹏,等.改进卷积神经网络的文本主题识别算法研究[J].计算机工程与应用,2022,58(2):161-168.

[6] JINDAL N,LIU B. Opinion spam and analysis [C]// Proceedings of the 2008 International Conference on Web Search and Data Mining. New York,USA:Association for Computing Machinery,2008:219-230.

[7] HAJEK P,BARUSHKA A,MUNK M. Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining [J]. Neural Computing and Applications,2020,32(23):17259-17274.

[8] 袁禄,朱郑州,任庭玉.虚假评论识别研究综述[J].计算机科学,2021,48(1):111-118.

[9] 陆杉,毛存礼,余正涛,等.融合多粒度特征的低资源语言词性标记和依存分析联合模型[C]//第二十届中国计算语言学大会论文集.呼和浩特,内蒙古:中国中文信息学会计算语言学专业委员会,2021:747-757.

[10] 陶晶晶.基于深度学习的商品虚假评论识别[D].成都:电子科技大学,2020.

[11] 曾致远,卢晓勇,徐盛剑,等.基于多层注意力机制深度学习模型的虚假评论检测[J].计算机应用与软件,2020,37(5):177-182.

[12] 李春雨.基于 Bi-GRU 的垃圾评论识别方法研究[D].上海:上海师范大学,2020.

[13] TIAN H,GAO C,XIAO X Y,et al. SKEP:sentiment knowledge enhanced pre-training for sentiment analysis [C/OL]//JURAFSKY D,CHAI J,SCHLUTER N,et al. Proceeding of the 58th Annual Meeting of the Association for Computational Linguistics. Online:Association for Computational Linguistics,2020:4067-4076. https://aclanthdogy.org/2020.acl-main.374.pdf.

[14] GAO X Y,LI S,ZHU Y Y,et al. Identification of deceptive reviews by sentimental analysis and characteristics of reviewers [J]. Journal of Engineering Science and Technology Review,2019,12(1):196-202.

[15] BARBADO R,ARAQUE O,IGLESIAS C A. A framework for fake review detection in online consumer electronics retailers [J]. Information Processing & Management,2019,56(4):1234-1244.

[16] 孟园,王悦.基于用户-评论-商户关系的虚假用户识别研究:用户偏差分析的视角[J].数据分析与知识发现,2022,66(6):55-70.

[17] RUAN N,DENG R Y,SU C H. GADM:manual fake review detection for O2O commercial platforms [J]. Computers & Security,2020,88:101657.

[18] 张琪,纪淑娟,傅强,等.基于带权评论图的水军群组检测及特征分析[J].计算机应用,2019,39(6):1595-1600.

[19] WANG Z,GU S M,XU X W. GSLDA:LDA-based group spamming detection in product reviews [J]. Applied Intelligence,2018,48(9):3094-3107.

[20] 韩忠明,杨珂,谭旭升.利用加权用户关系图的谱分析探测大规模电子商务水军团体[J].计算机学报,2017,40(4):939-954.

[21] RAYANA S,AKOGLU L. Collective opinion spam detection:bridging review networks and metadata [C]// Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York,USA:Association for Computing Machinery,2015:985-994.

[22] OTT M,CARDIE C,HANCOCK J. Estimating the prevalence of deception in online review communities [C]//Proceedings of the 21st International Conference on World Wide Web. New York,USA:Association for Computing Machinery,2012:201-210.

[23] 张东杰,黄龙涛,张荣,等.基于主题与情感联合预训练的虚假评论检测方法[J].计算机研究与发展,2021,58(7):1385-1394.

Fake Review Detection Model Based on Pre-training of Sentiment Information and Bi-GRU

ZHANG Yuying[1,2],ZHU Guangli[1,2**],ZHANG Youqiang[1,2],SUN Zhengyan[1,2],
ZHANG Shunxiang[1,2]

(1. School of Computer Science and Engineering,Anhui University of Science and Technology,Huainan,Anhui,232001,China;
2. Institute of Artificial Intelligence Research,Hefei Comprehensive National Science Center,Hefei,Anhui,230088,China)

Abstract:Fake review detection has important application value in e-commerce,social media,and other fields. Although existing review detection models integrate the sentiment information of the text,in the process of pre-training,the extraction of emotional information is ignored,resulting in low accuracy. Aiming at this problem,a Fake Review detection model (FR-SG) based on the pre-training of sentiment information and Bi-directional Gated Recurrent Unit (Bi-GRU) is proposed to improve the accuracy of fake review detection in this article. Firstly,the semantic vector of the text is obtained by Albert model. Then,Term Frequency-Inverse Document Frequency (TF-IDF) and K-means++ clustering methods are used to mine the sentiment seed words from reviews. Based on the seed words,the attribute words and sentiment words in the text are masked. Then,using the sentiment-oriented objective optimization function,the sentiment information is embedded into the semantic representation of the text to generate the sentiment vector. Finally,the joining results of these two groups of vectors are input into the fake review detection network to obtain the classification results of the text. Experimental results show that FR-SG improves the accuracy of fake review detection compared with the Bi-GRU+Attention model.

Key words:fake review detection;pre-training of sentiment information;sentiment seed words;Bi-GRU;objective optimization

责任编辑:梁　晓