

# 基于 XLNet-BiLSTM 模型的中文影评情感分析

张碧依, 陶宏才

(西南交通大学计算机与人工智能学院, 四川 成都 611756)

**摘要:**对影评进行情感分析可以为用户观看电影提供一定的参考意见。针对 Word2Vec 等静态词向量技术不能学习文本的深层信息、解决一词多义及 RNN 存在的长期依赖和上下文深层语义挖掘不充分的问题,提出一种基于 XLNet-BiLSTM 的中文影评情感分类模型。首先,使用 XLNet 预训练语言模型生成具有上下文依赖的词向量来对影评信息进行分布式表征;然后,将词向量输入到 BiLSTM 网络中,对评论的深层语义进行分析和计算;最后,使用 softmax 函数实现影评情感极性分类。通过爬取豆瓣电影上的评论对模型进行训练和测试,实验结果表明,模型的准确率为0.924,损失率为0.184,相比于相关的情感分析模型取得了更好的效果。

**关键词:**情感分析;词向量;XLNet;BiLSTM

**中图分类号:**TP391.12

**文献标志码:**A

**doi:**10.16836/j.cnki.jcuit.2021.03.004

## 0 引言

对影评进行情感分析,可以了解用户对某部电影褒贬倾向,同时也可以为其他用户选择电影提供一些参考意见。

CNN、RNN 和 LSTM 等深度学习算法被广泛地应用于情感分析。虽然和基于规则与机器学习<sup>[1]</sup>的方法相比,CNN 在情感分析领域的准确率有所提升,但是其学习文本上下文的能力还是有所不足。2013 年,Mikolov 等<sup>[2]</sup>使用 RNN 代替 CNN 对文本进行分类,解决了 CNN 存在的文本上下文学习不足的问题;但 RNN 又出现梯度弥散等新的问题,为了解这些问题,RNN 的众多变体被提出,如 LSTM、GRU 等。为了有效地学习文本双向上下文信息,本文将使用 BiLSTM 来构建基本模型。

研究表明,使用预训练技术可以学习到更多与下游任务相关的信息<sup>[3]</sup>。Word2Vec<sup>[2]</sup>词向量技术是一种典型的静态文本表示方法,这种方法只能学习文本的浅层表征<sup>[4]</sup>,且这种浅层表征在所有的语义环境中,对词的解释都一样,因此无法捕捉到一些深层次的信息<sup>[5]</sup>,对于下游任务的效果提升也非常有限。针对上述问题,使用 XLNet 这种动态的文本表示方法代替传统的静态文本表示方法来完成下游任务的初始化,使其能够同时学习到文本的浅层信息和深层信息。

## 1 相关工作

情感分析,也称为观点挖掘,是从一段文字中为某一特定主题提取情感或观点的过程。包括预测或分析

文本中存在的隐藏信息,这些隐藏信息有助于理解文本中的态度、观点和情绪<sup>[6]</sup>。由于通过各种网络和社交媒体表达意见的人所产生的数据基本上是非结构化的<sup>[7]</sup>,因此情感分析任务中最大的挑战是确定文本字面意义上陈述的真实性,并生成客观结果,以便根据正负极性对文本进行分类<sup>[8]</sup>。目前,情感分析的应用领域已经非常广泛,比如推荐系统、邮件过滤、各种网络评论情绪预测等<sup>[9]</sup>。

按照使用算法的类型分,情感分析主要有 3 种方法:基于规则的方法、基于机器学习的方法和基于深度学习的方法<sup>[10]</sup>。近年来,深度学习在情感分析中使用得越来越多。Kim 等<sup>[11]</sup>提出使用 CNN 对文本进行分类;刘龙飞等<sup>[12]</sup>使用 CNN 对微博语料进行情感分类。由于 CNN 没有充分学习上下文信息的能力,RNN 被更多应用于文本序列的处理。梁军等<sup>[13]</sup>和 Yangsen 等<sup>[14]</sup>用 RNN 对微博语料进行情感分析。针对 RNN 存在的梯度消失和梯度爆炸问题,Wang 等<sup>[15]</sup>使用 LSTM 对文本中词汇的时序关系进行学习;Zhu<sup>[16]</sup>使用 LSTM 将句子注释成单词序列进行情感分类;王业沛等<sup>[17]</sup>利用 LSTM 对判决书的判决倾向进行分析。

早期的预训练方法使用的是静态技术。2003 年,Bengio 等<sup>[18]</sup>使用神经网络实现了 NNLM 语言模型。2013 年,Tomas Mikolov 等<sup>[2]</sup>从 NNLM 得到启发,提出了 Word2vector 模型。次年,Jeffrey Pennington 等<sup>[19]</sup>提出了 Glove 词嵌入技术,Glove 在 Word2vector 的基础上,修改了目标函数以及权重函数<sup>[20]</sup>。随后,Word2Vec 和 Glove 在自然语言处理中得到了广泛的应用。但其本质是一种静态的预训练技术,不能解决在一词多义的问题,对下游任务的提升比较有限<sup>[21]</sup>。文献<sup>[22]</sup>提出了 ELMo 模型。ELMo 模型解决了静态文本存在一词多义

问题。因为 ELMo 在学习上下文时只是对节点左右信息进行了拼接,其两边的信息并没有交集,所以 ELMo 并不是严格意义上的双向语言模型。文献[23]提出了 GPT 预训练模型,首次用 Transformer 网络来捕获长距离语言结构。不久之后,Google 又提出了 BERT<sup>[24]</sup> 模型, BERT 同样使用了 Transformer 网络,在预训练时,通过屏蔽部分单词的方法来实现双向预测。但是, BERT 未考虑屏蔽词与未屏蔽词之间的联系<sup>[25]</sup>。2019 年, Google 团队在 BERT 的基础上提出了 XLNet 模型<sup>[25]</sup>, 其基于自回归 (autoregressive, AR)<sup>[26]</sup> 语言模型实现了新的双向编码,并且由于 AR 方法的缘故,解决了 BERT 未考虑单词之间关系的这一缺点。

## 2 情感分析模型

### 2.1 XLNet 模型

XLNet 模型在 BERT 的基础上做了一些改进。首

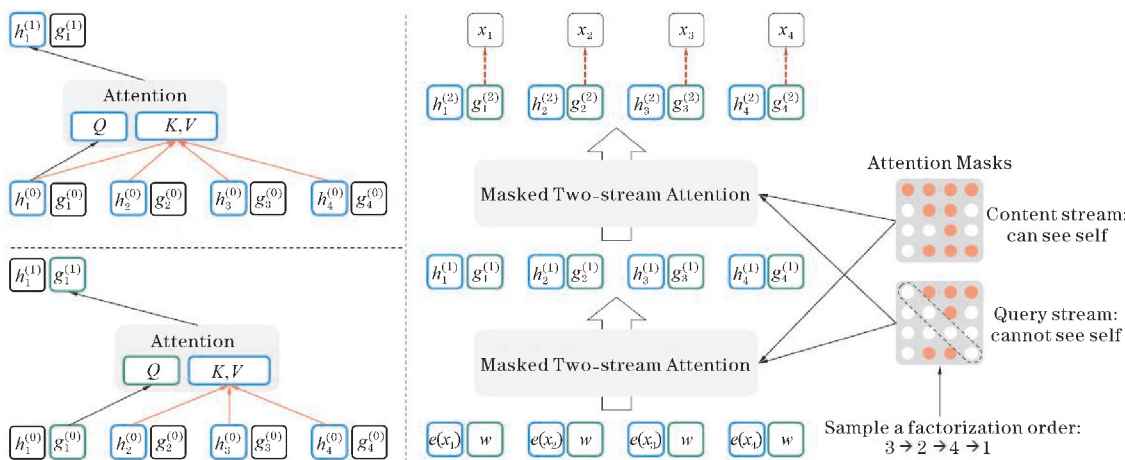


图1 双流自注意力列模型计算过程图

图1中,左上是Content流的注意力计算,假设预测第一个位置的词的概率,则  $KV=[h_1^{(0)}, h_2^{(0)}, h_3^{(0)}, h_4^{(0)}]$ ,  $Q=h_1^{(0)}$ ;左下是Query流的计算,  $KV=[h_2^{(0)}, h_3^{(0)}, h_4^{(0)}]$ ,  $Q=g_1^{(0)}$ 。图的右边是完整的计算过程,  $h$  和  $g$  分别被初始化为  $e(x_i)$  和  $W$ , 然后Content掩码和Query掩码依次计算每层的输出。最右边是掩码矩阵,假设排列后的顺序为  $[3, 2, 4, 1]$ , 则在Content掩码矩阵中,第一个词可以利用所有词的信息;第二个词可以利用第二个词和第三个词的信息,依此类推。因为Query掩码不能利用本身的信息,所以对角线为白点。

循环机制是XLNet的另一个重要部分,其借鉴了Transformer-XL的思想,即将上一个单词的隐藏层信息用于下一个单词的计算,从而获得更长距离的上下文信息<sup>[21]</sup>。在处理长文档的相关任务中,循环机制能够发挥很好的效果。

先, XLNet 不使用传统 AR 模型中固定的因式分解顺序,而是让所有因式分解顺序的期望对数似然值最大化。其次, XLNet 不依赖残缺数据,因此 XLNet 不存在 BERT 中训练-微调数据不一致的问题<sup>[27]</sup>。

XLNet 模型的核心思想是以排列组合的形式重构输入文本,把一部分下文的内容放到上文中,充分利用上下文信息实现双向预测的功能。排列语言模型的目标是让式(1)的似然概率最大:

$$\max_{\theta} E_{Z \sim Q_T} \left[ \sum_{i=1}^T \lg p_{\theta}(x_{z_i} | X_{Z_{<i}}) \right] \quad (1)$$

其中,

$$p_{\theta}(X_{z_i}=x | X_{Z_{<i}}) = \frac{\exp(e(x)^T g^{\theta}(X_{Z_{<i}}, z_i))}{\sum_{x'} \exp(e(x')^T g^{\theta}(X_{Z_{<i}}, z_i))} \quad (2)$$

因为在微调阶段无法排列组合原始的输入文本,所以在预训练时,只能在Transformer内部改变输入文本的顺序, XLNet 通过 Attention 掩码实现这一功能。在 XLNet 中,  $g_{\theta}(X_{Z_{<i}}, z_i)$  用双流自注意力机制实现,即 content 流  $h_{z_i}$  和 Query 流  $g_{z_i}$ , 其计算过程如图1所示。

### 2.2 BiLSTM 模型

LSTM 模型是 RNN 的改进版,在长序列训练任务上有更好的效果,其神经元结构如图2所示。LSTM 由遗忘门、输入门和输出门3个门控单元组成,计算公式如式(3)~(5)所示。遗忘门用来计算需要遗忘的信息;输入门用来计算需要更新的数据;输出门用来决定输出信息。对单元状态信息进行遗忘和更新的计算如式(6)、(7)所示。式(8)是计算最后的隐藏层输出。

$$f_t = \sigma(W_f \cdot (h_{t-1}, x_t) + b_f) \quad (3)$$

$$i_t = \sigma(W_i \cdot (h_{t-1}, x_t) + b_i) \quad (4)$$

$$o_t = \sigma(W_o \cdot (h_{t-1}, x_t) + b_o) \quad (5)$$

$$\tilde{c}_t = \tanh(W_c \cdot (h_{t-1}, x_t) + b_c) \quad (6)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (7)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (8)$$

其中,  $W_f, W_i, W_o, W_c$  为计算  $f_t, i_t, o_t$  和  $\tilde{c}_t$  时的权重。  
 $b_f, b_i, b_o, b_c$  为计算  $f_t, i_t, o_t$  和  $\tilde{c}_t$  时的偏移向量,  $\tilde{c}_t$  表示  $t$  时刻由  $\tanh$  新创建的向量,  $c_t$  表示  $t$  时刻更新后的状态信息。  $h_t$  表示 LSTM 最后隐藏层信息的输出。

BiLSTM 由两个方向相反的 LSTM 组成, 最后输出的结果由这两个方向相反的 LSTM 共同决定, 其模型结构如图 3 所示。

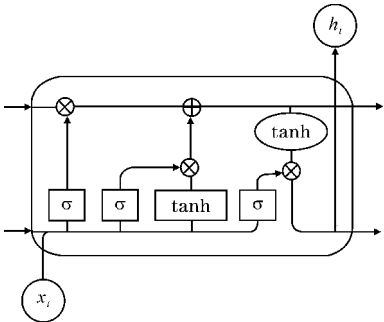


图 2 LSTM 神经元结构图

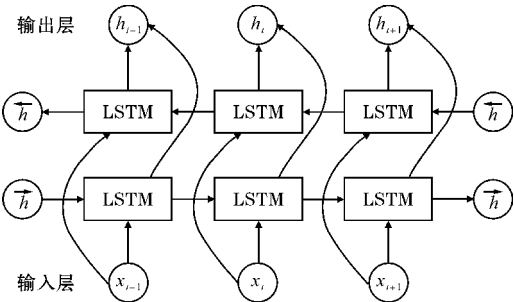


图 3 BiLSTM 结构图

图中,  $h_t$  为  $t$  时刻 BiLSTM 的输出结果,  $\vec{h}_t$  和  $\overleftarrow{h}_t$  分为  $t$  时刻 LSTM 的正向输出和反向输出。  $\vec{h}_t$  由  $t$  时刻的输入  $x_t$  和  $t-1$  时刻的正向输出  $\vec{h}_{t-1}$  计算得到,  $\overleftarrow{h}_t$  由  $t$  时刻的输入  $x_t$  和  $t-1$  时刻的反向输出  $\overleftarrow{h}_{t-1}$  计算得到,  $h_t$  由  $\vec{h}_t$  和  $\overleftarrow{h}_t$  计算得到。

2.3 XLNet-BiLSTM 模型设计

XLNet-BiLSTM 模型结构如图 4 所示, 分为输入层、嵌入层、隐藏层和输出层 4 个部分。

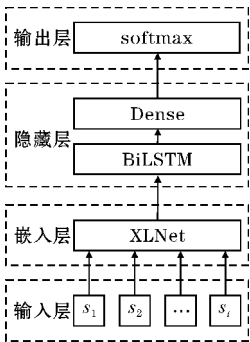


图 4 XLNet-BiLSTM 模型结构图

2.3.1 输入层

输入层的功能是把原始数据处理成模型需要的数据格式, 原始数据即豆瓣网爬取的评论数据。对原始数据的处理包括以下几个步骤:

步骤 1: 使用 python 爬虫技术获取豆瓣电影的网页信息, 并使用 BeautifulSoup 对网页进行解析, 得到用户对电影的评论信息, 包括评论内容、评分星级等。

步骤 2: 对评论信息进行去重、清洗, 包括使用正则表达式去除无用字符、将繁体转化为简体、删除评论长度过小的数据等。

步骤 3: 分别取评分为 1 分和评分为 5 分的两端评论数据, 将评论内容划分为负极和正极, 随机删除评论内容中汉字较少的评论数据, 使正负极数据量为 1 : 1。

步骤 4: 打乱评论数据, 并将其划分为训练集和测试集, 划分比例为 8 : 2。

2.3.2 嵌入层

嵌入层的功能是将句子向量化, XLNet 模型与传统的 Word2Vec 模型有所不同, 因此文本的输入格式也有所不同。 Word2Vec 模型需要事先将评论文本进行分词, 然后以数字序列的方式输入模型, 最后得到词向量, 其分词技术多使用 jieba 分词。而 XLNet 语言模型使用 SentencePiece<sup>[28]</sup> 技术来实现分词, 然后得到每条评论文本的数字序列。 SentencePiece 工具特别之处在于, 它不依赖于之前的训练, 而是从给定的训练集中学习, 并且不会因为语言不同而有不一样的表现, 因为它把所有字符串中的字符都当作 Unicode 字符处理。

由于谷歌官方并未发布中文版的 XLNet 预训练模型, 因此本文使用哈工大讯飞联合实验室发布的面向中文 XLNet 预训练模型的 base 版, 来生成词向量。 设样本  $S$  中有  $n$  个句子, 即  $S = \{s_1, s_2, \dots, s_n\}$ , 样本中的句子经 SentencePiece 分词后, 第  $j$  个句子表示为  $s_j = \{w_{j1}, w_{j2}, \dots, w_{ji}\}$ , 通过 XLNet 嵌入层预处理后, 得到句子的每个词向量, 词向量的维度为 768, 句向量通过对词向量求和取平均得到, 所以最后得到的句向量维数也是 768。 计算公式为

$$v_{jk} = \frac{1}{t} \sum_{i=1}^t c_{jik}$$
 (9)

其中,  $k$  表示词向量的第  $k$  维,  $c_{ji}$  表示第  $j$  个句子的第  $i$  个词向量。

2.3.3 隐藏层

BiLSTM 层的输入为经 XLNet 处理后的句向量, 主要目的是对输入的句向量做深层次的特征提取, 更加充分地学习上下文之间的信息。 在第  $i$  个时刻, 输入的第  $j$  个句向量为  $v_{ij}$ , 经 BiLSTM 语义编码后, 表示为  $h_{ij}$ , 具体公式为

$$h_{ij} = \text{BiLSTM}(v_{ij})$$
 (10)

2.3.4 输出层

输出层使用的激活函数是 softmax, 输出内容为评



论极性的预测分类结果,其计算公式为

$$y_j = \text{softmax}(\mathbf{W}h_{ij} + b) \tag{11}$$

其中,  $\mathbf{W}$  表示训练权重矩阵,  $b$  表示偏置,  $y_j$  为第  $j$  条评论的预测极性。

3 模型实验

实验环境为 Windows 10 操作系统, 编程语言为 Python, BiLSTM 和 XLNet 网络分别使用 Keras 库和 pytorch-transformers 来实现, 其中 Keras 库底层是通过 Tensorflow 框架来实现的。

表 1 数据集分布 单位:条

汉字长度 星级	>5	>10	>20	>30	>40	>50
1 星	52279	42511	28662	20915	16275	13108
5 星	178605	156185	120926	97735	81924	70049

表 2 评论样例

评论极性	样本内容
正极	1 全程搞笑,也没有什么低俗的硬梗,好看。
负极	0 不好看,作为音乐片,也没几首好听的歌。

3.2 模型参数

模型参数优化利用网格搜索超参数优化技术完成。首先,将 keras 模型封装到 KerasClassifier 类中。然后,创建超参数字典,构造 GridSearchCV 类对象并将超参数字典传给其 param\_grid 参数。最后,使用 GridSearchCV 的 fit( ) 函数返回优化结果。返回结果中有一个 best\_score\_成员,该成员记录了在参数优化时得到的最好评分,最佳结果的参数组合保存在 best\_params\_中。调优后的参数设置如表 3 所示。

表 3 XLNet-BiLSTM 模型参数设置

参数名	参数值
句向量维度	768
LSTM 隐藏层	128
Dense 隐藏层	128
Dropout	0.2
优化函数 optimizer	Adam
学习率 lr	0.0008
Loss 函数	categorical_crossentropy
batch_size	64
validation_split	0.25

3.3 对比实验与分析

实验分别从 3 个方面进行模型比较,一是使用了

3.1 实验数据

实验数据集选用豆瓣影评,搜集了 5000 部电影的前 220 条评论。经过数据清洗后得到 1019765 条评论。因为中间星级的评论数据多存在一些评论内容极性模糊的问题,所以分别取 1 星和 5 星两端评论数据作为正负极评论。该数据集中评论内容汉字长度分布如表 1 所示。由表 1 可以看出,5 星级的评论数量远多于 1 星级的评论。为平衡语料,分别从 1 星和 5 星评论中随机取 5 万条汉字长度大于 5 的数据用于实验,共计 10 万条。其中,正极用 1 表示,负极用 0 表示。

表 2 给出了评论的正极样例和负极样例。

预训练技术和未使用预训练技术模型的比较;二是 Word2Vec 词向量技术和 XLNet 词向量技术的比较;三是 CNN, LSTM 和 BiLSTM 的比较。本实验的评价指标为预测结果的准确率和损失率。损失率采用交叉熵损失函数计算。设总样本数为  $N$ , 预测正确的样本数为  $M$ , 则准确率为

$$\text{Accuracy} = \frac{M}{N} \tag{12}$$

由于 XLNet 和 Word2Vec 对输入数据的格式要求不同,因此在实现 Word2Vec 相关模型时,数据的处理方式也稍有不同。对于 Word2Vec 模型输入的数据处理主要包括 jieba 分词,去除不含情感词和分词后评论长度小于 5 的评论。模型输入的句子长度取 128, 不足的用 0 补充,长度大于 128 的采取截断的方式。字典长度设置为 5000,词向量维度为 300。

图 5 和图 6 为 10 次迭代过程中,模型训练的准确率和损失率动态变化情况。

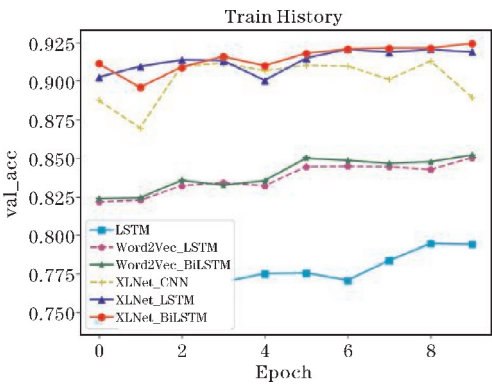


图 5 验证集准确率

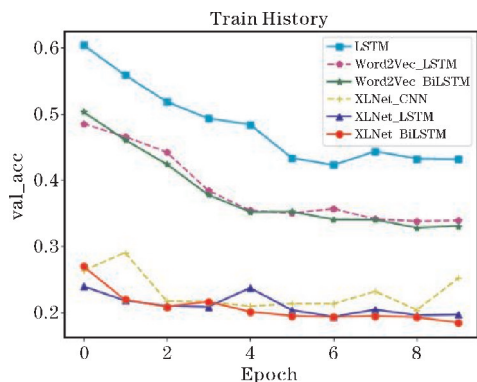


图6 验证集损失率

总体上看,各个模型的准确率都不断上升并趋于稳定。在4次迭代后,XLNet-BiLSTM模型的准确率一直高于其他模型。模型在测试集上的测试结果如表4所示。

表4 模型测试结果

模型	准确率	损失率
LSTM	0.759	0.483
Word2Vec-LSTM	0.831	0.391
Word2Vec-BiLSTM	0.835	0.387
XLNet-CNN	0.893	0.252
XLNet-LSTM	0.919	0.195
XLNet-BiLSTM	0.924	0.184

由表4可以看出,本文提出的XLNet-BiLSTM模型在准确率和损失率上均优于其他5种模型。6种模型中,未使用语言模型的LSTM模型效果最差,这也进一步证明了语言模型在自然语言处理中的重要性。在使用了预训练技术的模型中,BiLSTM稍优于LSTM,说明学习文本双向上下文信息能够更好地完成分类任务;而使用了XLNet预训练技术的模型又优于使用WordsVec预训练技术的模型,这说明相比于Word2Vec这种静态文本表示方法,XLNet这种动态文本表示方法对下游分类任务的效果提升更明显。

综上,在相关指标的测定下,本文提出的XLNet-BiLSTM模型在6个模型中效果最好。

4 结束语

提出一种基于XLNet-BiLSTM的中文影评情感分析模型,使用动态文本表征方法XLNet代替传统的静态文本表征方法Word2Vec来实现文本向量化,然后再使用BiLSTM来实现文本双向上下文信息深度提取,完成影评情感分类。将该模型与其他几种常见的情感分类模型在豆瓣影评数据上进行实验,结果表明,相比于其他使用Word2Vec的方法,使用XLNet的方法不但提高了准确率,同时也降低了损失率,而使用BiLSTM的结果又优于CNN。这一方面说明在影评情感分类模型中,XLNet文本表征方法对模型的效果提

升优于Word2Vec;另一方面也说明BiLSTM相比于CNN更适合于文本情感分类。

参考文献:

[1] 陈珂,梁斌,柯文德,等. 基于多通道卷积神经网络的中文微博情感分析[J]. 计算机研究与发展,2018,55(5):945-957.

[2] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrase and their compositionality[C]. Advances in Neural Information Processing Systems,2013:3111-3119.

[3] Kiros R, Zhu Y, Salakhutdinov R R, et al. Skip-thought vectors[C]. Advances in Neural Information Processing Systems,2015:3294-3302.

[4] Debg L, YU D. Deep learning: methods and applications[J]. Foundations and Trends in Signal Processing,2014,7(3/4):197-387.

[5] Yang X, Macdonald C, Ounis I. Using word embeddings in twitter election classification[J]. Information Retrieval Journal,2018,21(2/3):183-207.

[6] HarpreetKaur, VeenuMangat, Nidhi. A Survey of Sentiment Analysis techniques[C]. 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), IEEE, 2017:921-925.

[7] Zvarevashe K, Olugbara O O. A framework for sentiment analysis with opinion mining of hotel reviews[C]. 2018 Conference on Information Communications Technology and Society (ICTAS), IEEE, 2018:1-4.

[8] Kumar A, Sangwan S R, Arora A, et al. Sarcasm Detection Using Soft Attention-Based Bidirectional Long Short-Term Memory Model with Convolution Network[J]. IEEE Access,2019,7:23319-23328.

[9] Bandana R. Sentiment Analysis of Movie Reviews Using Heterogeneous Features[C]. 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), IEEE,2018:1-4.

[10] 李文江,陈诗琴. 基于深度学习的商品评论情感分类研究[J]. 知识管理论坛,2018,3(6):353-363.

[11] Kim Y. Convolutional Neural Networks for Sentence Classification[C]. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP),2014:1746-1751.

[12] 刘龙飞,杨亮,张绍武,等. 基于卷积神经网络的微博情感倾向性分析[J]. 中文信息学报,

- 2015,29(6):159-165.
- [13] 梁军,柴玉梅,原慧斌,等.基于深度学习的微博情感分析[J].中文信息学报,2014,28(5):155-161.
- [14] Yangsen J Y Z, Tong Y. Study of Sentiment Classification for Chinese Microblog Based on Recurrent Neural Network[J]. Chinese Journal of Electronics, 2016, 25(4):601-607.
- [15] Wang X, Liu Y, Sun C J, et al. Predicting polarities of tweets by composing word embeddings with long short-term memory[C]. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, 1:1343-1353.
- [16] Zhu X, Sobihani P, Guo H. Long short-term memory over recursive structures[C]. Proceedings of the International Conference on Machine Learning, 2015:1604-1612.
- [17] 王业沛,宋梦姣,王譞,等.基于深度学习的判决结果倾向性分析[J].计算机应用研究,2019,36(2):335-338.
- [18] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003(3):1137-1155.
- [19] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014:1532-1543.
- [20] 胡益淮.基于XLNET的抽取式多级语义融合模型[J].通信技术,2020,53(7):1630-1635.
- [21] 李舟军,范宇,吴贤杰.面向自然语言处理的预训练技术研究综述[J].计算机科学,2020,47(3):162-173.
- [22] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners[J]. Open AI Blog, 2019, 1(8):8-9.
- [23] Lundervold A S, Lundervold A. An overview of deep learning in medical imaging focusing on MRI[J]. Zeitschrift für Medizinische Physik, 2019, 29(2):102-127.
- [24] Young T, Hazarika D, Poria S, et al. Recent trends in deep learning based natural language processing[J]. IEEE Computational intelligence magazine, 2018, 13(3):55-75.
- [25] Yang Z, Dai Z, Yang Y, et al. Xlnet: Generalized autoregressive pretraining for language understanding[C]. Advances in neural information processing systems, 2019:5754-5764.
- [26] Gong X, Jin J, Zhang T. Sentiment Analysis Using Autoregressive Language Modeling and Broad Learning System[C]. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2019:1130-1134.
- [27] 姚贵斌,张起贵.基于XLnet语言模型的中文命名实体识别[J/OL].计算机工程与应用. <http://kns.cnki.net/kcms/detail/11.2127.TP.20201120.1251.012.html>, 2021-01-03.
- [28] Kudo T, Richardson J. SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing[C]. 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018:66-71.

## Sentiment Analysis of Chinese Film Review based on XLNet-BiLSTM Model

ZHANG Biyi, TAO Hongcai

(School of Computing & Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China)

**Abstract:** Sentiment analysis of film reviews can provide some reference for users to watch movies. The word vector technology such as Word2Vec can't learn the deep representation of text and solve the polysemy of the word, and the RNN can't fully exploit the deep semantics of context and has the characteristics of long-term dependence. In order to solve these problems, this paper proposed a Chinese film review sentiment classification model based on XLNet-BiLSTM neural network. Firstly, we use the XLNet model to generate a context-dependent word vector for distributed representation of the information, and then input the word vectors into the BiLSTM network to analyze and calculate the deep semantics of comments. Finally, we used the softmax function to classify the sentiment polarity. The model is trained and tested by reviews in Douban film. The experimental results show that the accuracy of the model is 0.924 and the loss rate is 0.184, which is better than the related sentiment analysis model.

**Keywords:** sentiment analysis; word vector; XLNet; BiLSTM