

基于预训练语言模型词向量融合的情感分析研究

魏上斐¹ 乔保军^{2*} 于俊洋¹ 姚相宇¹

¹(河南大学软件学院 河南 开封 475000)

²(河南大学计算机信息与工程学院 河南 开封 475000)

摘要 针对传统情感分类模型分类效果不足,无法准确地捕捉词语之间关系的问题,提出一种基于预训练语言模型词向量融合的 GE-BiLSTM(Glove-ELMO-BiLSTM)情感分析模型。通过预训练语言模型 ELMO 以语言模型为目的训练词向量,再与传统的 Glove 模型的训练结果进行运算融合,结合了全局信息以及局部上下文信息,增加了词向量矩阵的稠密度,词语之间的特征得到更好的表达,结合 BiLSTM 神经网络可以更好地捕捉上下文信息的关系。实验结果证明:GE-BiLSTM 情感分析模型可以达到更好的分类效果,准确率比传统模型提高了 2.3 个百分点,F1 值提升了 0.024。

关键词 Glove 预训练语言模型 ELMO 词向量融合 BiLSTM 情感分析

中图分类号 TP391 文献标志码 A DOI:10.3969/j.issn.1000-386x.2021.05.024

SENTIMENT ANALYSIS BASED ON PRE-TRAINED LANGUAGE MODEL WORD VECTOR FUSION

Wei Shangfei¹ Qiao Baojun^{2*} Yu Junyang¹ Yao Xiangyu¹

¹(Software College, Henan University, Kaifeng 475000, Henan, China)

²(School of Computer Information and Engineering, Henan University, Kaifeng 475000, Henan, China)

Abstract In order to solve the problem that the traditional sentiment classification model is not effective enough to capture the relationship between words accurately, this paper proposes a GE-BiLSTM(Glove-ELMO-BiLSTM) sentiment analysis model based on pre-trained language model word vector fusion. Through the pre-trained language model ELMO, the word vector was trained for the purpose of the language model. Then, it combined the global information and local context information with the traditional Glove model training results to increase the density of the word vector matrix, so that the characteristics between words could be better expressed. Combining with BiLSTM could better capture the relationship of context information. The experimental results show that the GE-BiLSTM sentiment analysis model can achieve better classification results. The accuracy rate is 2.3 percentage points higher than the traditional model, and the F1 value is increased by 0.17.

Keywords Glove Pre-trained language model ELMO Word vector fusion BiLSTM Sentiment analysis

0 引言

随着网络的盛行与网络技术的发展,用户可在网络中发表自己对事物的观点,这些观点中含有用户想要表达的情感因素。更加准确地发掘评论中的情感因

素已成为当前工业界和学术界的迫切要求。

情感分析又称为意见挖掘、主观因素挖掘等,是通过计算机分析并整理互联网上大量的评论信息,对带有情感因素的主观性文本进行分析、处理的过程^[1]。情感分析的主要工作是将文本所含有情感因素(如积极、消极、中立等)进行分类。目前,国内外学者提

出基于传统的机器学习方法,在语法层面上进行语义分析。周枫等^[2]提出了一种双向 GRU 与 CNN 相结合的文本分类模型,通过增大重要信息的权重,提高分类精度。现在流行的深度学习采用浅层神经网络的方法也可以进行情感分类。一些研究学者在文本的词向量表达上进行了改进,得到了更好的分类效果。Biswas 等^[3]利用特有领域的信息训练词向量,使情感分类模型有了更好的提升。Cotterell 等^[4]提出了一种新的形态学词向量,在词相似方面得到了很大的提升效果。Abdeddaïm 等^[5]提出了一种新的 MeSH-gram 模型,通过句向量捕捉语义联系,性能高于传统 skip-gram 模型。Yao 等^[6]提出了一种新方法 LSI,通过将外部知识转化为语义空间以增强词向量表达方式。文献[7-8]通过结合主题模型来训练词向量进行文本分类。文献[9-11]均使用深层次模型训练词向量,并应用在词表征任务上。

综上所述,现阶段的情感分类问题在于无法更好地抓取情感因素。词向量表示方法能够更好地对文本特征进行描述,在分类时的效果更好。本文提出 GE-BiLSTM 情感分析模型。通过预训练语言模型构建文本的特征词向量,然后将文本的特征词向量作为 BiLSTM 神经网络的输入,通过训练神经网络获取深度词向量特征,最后使用 softmax 层对文本进行情感分类。

1 相关介绍

1.1 Glove 模型

Glove(Global Vectors for Word Representation)模型由斯坦福大学的 Pennington 等^[12]提出,Glove 模型是基于共现矩阵分解生成词向量的模型。假设共现矩阵为 X , X_i 表示词 i , X_j 表示词 j 在目标词 i 的上下文中出现的次数。模型在训练时用到了全局信息,利用了词共现的信息状态,使用上下文信息建模,如果两个词在语料库中出现的上下文位置类似,这两个词的词向量就相似,对词语具有更好的描述。Glove 模型的损失函数为:

$$J = \sum_{i,j=1}^{|V|} f(X_{ij}) (\mathbf{W}_i^T \mathbf{W}_j + \mathbf{b}_i + \mathbf{b}_j - \log X_{ij})^2 \quad (1)$$

式中: $|V|$ 是词典的大小; \mathbf{W}_i 为词 i 的词向量; \mathbf{W}_j 为词 j 的词向量; \mathbf{b}_i 和 \mathbf{b}_j 分别为矩阵 X 行、列的偏移值。 J 的计算复杂度和共现矩阵矩阵 X 的非零元素的数目呈线性关系。设共现矩阵 X 中, X_{ij} 代表词 j 在词 i 上下文出现的次数。词 i 和词 j 的上下文是 k 时, k 与 i, j 存在共现关系。

$$P(k|i) = P_{ik} = \frac{X_{ik}}{X_i} \quad (2)$$

式中: $P(k|i)$ 为单词 i, k 共现概率。

$$P(k|j) = P_{jk} = \frac{X_{jk}}{X_j} \quad (3)$$

式中: $P(k|j)$ 为单词 j, k 共现概率。

$$F(w_i, w_j, w_k) = \frac{P_{ik}}{P_{jk}} \quad (4)$$

如果 $F(w_i, w_j, w_k)$ 的值较大,说明 k 和 i 的相关性强;反之,则说明 k 与 j 相关性比较大。Glove 模型的不足是上下文单词之间的联系不强,同一个词出现在不同场景的文本中,词汇的向量表征矩阵变化不大。

1.2 LSTM 网络结构

LSTM 网络结构是基于循环神经网络(RNN)改进而来的。循环神经网络最早由 Schuster 等^[14]提出,该网络结构更适合对序列化数据进行处理和预测,在处理文本信息时,可以使之之前的信息对后面的输入信息产生影响,输入信息为 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\}$, \mathbf{x}_t 的输入都会经过激活函数转化为输出值 \mathbf{o}_t 和上一个状态的记忆值 \mathbf{h}_t ,所以当前时刻 t 的输入除了来自当前输入层的 \mathbf{x}_t ,还有上一个状态的 \mathbf{h}_t 。LSTM 具体单元结构公式如下:

$$\mathbf{z}_t = \tanh(\mathbf{W}_z[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_z) \quad (5)$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \quad (6)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \quad (7)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o(\mathbf{h}_{t-1}, \mathbf{x}_t) + \mathbf{b}_o) \quad (8)$$

$$\mathbf{c}_t = \mathbf{f}_t \cdot \mathbf{c}_{t-1} + \mathbf{i}_t \cdot \mathbf{z}_t \quad (9)$$

$$\mathbf{h}_t = \mathbf{o}_t \cdot \tanh \mathbf{c}_t \quad (10)$$

式中: \mathbf{z} 为输入值; \mathbf{i} 为输入门; \mathbf{f} 为遗忘门; \mathbf{o} 为输出门; \mathbf{c}_t 为结合遗忘门选择的信息后输出的新状态; \mathbf{h}_t 为结合 \mathbf{c}_t 、输入和上一个状态所生成的进入下一层隐藏层的状态; σ 为 sigmoid 激活函数; \mathbf{b} 是训练时的偏置项。“门”机制可以更好地筛选有用的信息,虽然在训练时会产生更多的参数,但在数据量不大的数据集上具有更好的效果,所以本文选择了 LSTM 作为提取情感因素的网络结构。

2 GE-BiLSTM 情感分析模型

本文提出一种 GE-BiLSTM 情感分析模型,引入了预训练语言模型,针对于当前传统情感分类模型无法更好地抓取情感因素导致文本向量矩阵稀疏的问题进行了改进及处理。通过训练语言模型,以 BiLSTM 作为语言模型训练的网络结构在训练结束后可以得到输

入文本中每个词汇的 m 层向量矩阵,由于两个模型的训练机制不同,最终生成向量矩阵需进行运算重组,运算函数如下:

$$\mathbf{e}' = \sum_{i=1}^m \mathbf{e}_i/h \quad (11)$$

式中: \mathbf{e}' 为新的 ELMO 模型词向量; \mathbf{e} 为原始词向量; n 为词向量维度; h 为每个词汇的词向量层数。将预训练语言模型 ELMO 训练结果与 Glove 词向量通过运算融合为新的词向量矩阵。最终放入 BiLSTM 模型训练,进行情感分析。

2.1 预训练语言模型

本文引用了 Peters 等^[13]提出的 ELMO 模型(Embeddings from Language Models),以语言模型为目标在双向 LSTM 上训练模型,利用 LSTM 网络结构生成词语的表征。Word2Vec 等传统模型生成词向量时,不同句子中的相同词汇的词向量是一样的;而本文预训练双向语言模型的输入度量是字符而不是词汇,从而可以增强单词的联系,以语言模型为目的,更好地预测下一个单词的出现概率。该语言模型的前后网络结构如下。

前向 LSTM 结构:

$$p(t_1, t_2, \dots, t_N) = \prod_{k=1}^N p(t_k | t_1, t_2, \dots, t_{k-1}) \quad (12)$$

反向 LSTM 结构:

$$p(t_1, t_2, \dots, t_N) = \prod_{k=1}^N p(t_k | t_{k+1}, t_{k+2}, \dots, t_N) \quad (13)$$

联合前向和后向语言模型,并最大化前后向语言模型,得到该语言模型的联合似然函数如下:

$$\sum_{k=1}^N (\log p(t_k | t_1, \dots, t_{k-1}; \Theta_x, \vec{\Theta}_{\text{LSTM}} \Theta_s) + \log p(t_k | t_{k+1}, \dots, t_N; \Theta_x, \overleftarrow{\Theta}_{\text{LSTM}} \Theta_s)) \quad (14)$$

式中: θ_x 为输入的初始词向量参数; θ_s 为输出的 softmax 层参数。本文利用预训练语言模型的概念,使用双向 LSTM 训练语言模型,对该模型加以应用,每个词语经过语言模型的预训练后,情感词的权重通过最大条件概率随之增大,更加突出情感关键词的重要性,结合特定数据语境的情感词的条件概率可通过极大似然估计求得:

$$p(t_k | t_{k-1}) = \frac{C(t_{k-1} t_k)}{C(t_{k-1})} \quad (15)$$

通过结合全局信息和局部上下文信息,可以充分弥补传统情感分类模型中,只能通过前后句子判断关键情感词重要性的缺陷,情感词条件概率的增大,可以

通过不同场景,产生不同意义的词语表征。情感信息在融合新机制下得到了更好的保留,效果优于传统主题模型。

2.2 词向量的运算重组

本文所选择的传统词向量模型为 Glove 模型,该模型在 Word2Vec 模型的基础上加以改进,结合了全局信息,弥补了 Word2Vec 只能结合局部语义信息的缺点,在词表征任务上具有更好的效果。本文模型首先通过预训练语言模型训练词向量,在得到局部上下文信息后,再结合 Glove 模型训练得到的全局信息重组运算成新的矩阵。传统词向量模型 Glove 训练出来的词向量矩阵设为 \mathbf{G} ,文本分词之后每个词的 tokens 设为 g ,词向量矩阵维度过大容易造成训练参数增加、时间复杂度提升。经过实验证明,词向量维度在 200 ~ 300 维之间时词向量差别很小;当维度超过 300 维,会造成维度过长,训练时参数过多,提高时间复杂度。本文 Glove 模型词向量维度选择为 256 维,矩阵为文本词向量的集合 \mathbf{G} 。

$$\mathbf{G} = \begin{bmatrix} g_{1,1} & g_{1,2} & \dots & g_{1,256} \\ g_{2,1} & g_{2,2} & \dots & g_{2,256} \\ \vdots & \vdots & & \vdots \\ g_{76634,1} & g_{76634,2} & \dots & g_{76634,256} \end{bmatrix} \quad (16)$$

通过预训练语言模型训练出来的词向量矩阵设为 \mathbf{E} ,文本分词之后每个词的 tokens 设为 e ,维度与 Glove 模型输出保持一致,矩阵为文本词向量的集合 $\{e_{1,1}, e_{1,2}, \dots, e_{229902,256}\}$,训练出的每个词的词向量有 m 层,将 ELMO 模型每个词的词向量求平均值,转为新的词向量矩阵 \mathbf{E}' 。

$$\mathbf{E} = \begin{bmatrix} e_{1,1} & e_{1,2} & \dots & e_{1,256} \\ e_{2,1} & e_{2,2} & \dots & e_{2,256} \\ \vdots & \vdots & & \vdots \\ e_{229902,1} & e_{229902,2} & \dots & e_{229902,256} \end{bmatrix} \quad (17)$$

$$\mathbf{E}' = \begin{bmatrix} e'_{1,1} & e'_{1,2} & \dots & e'_{1,256} \\ e'_{2,1} & e'_{2,2} & \dots & e'_{2,256} \\ \vdots & \vdots & & \vdots \\ e'_{76634,1} & e'_{76634,2} & \dots & e'_{76634,256} \end{bmatrix}$$

将两个模型训练的词向量通过运算融合,即 $\mathbf{W} = \text{concat}(\mathbf{G}, \mathbf{E}')$,运算后的具体矩阵为:

$$\mathbf{W} = \begin{bmatrix} g_{1,1} & g_{1,2} & \dots & g_{1,256} & e'_{1,1} & e'_{1,2} & \dots & e'_{1,256} \\ g_{2,1} & g_{2,2} & \dots & g_{2,256} & e'_{2,1} & e'_{2,2} & \dots & e'_{2,256} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ g_{76634,1} & g_{76634,2} & \dots & g_{76634,256} & e'_{76634,1} & e'_{76634,2} & \dots & e'_{76634,256} \end{bmatrix} \quad (18)$$

融合后的词向量维度为 512 维。

2.3 GE-BiLSTM 模型结构

本文提出的 GE-BiLSTM 模型的关键部分在于通过预先训练语言模型,得到包含上下文信息的词向量,再结合传统的 Glove 所生成的全局词向量,提高了词语之间的联系,以语言模型为目的训练词向量可以更好地抓取情感因素,双向 LSTM 网络进一步提取文本特征,以提高分类精度。LSTM 网络结构是在 RNN (Recurrent Neural Network) 基础上改进得到的网络结构,其克服了 RNN 无法存储更多信息的缺点。但是在训练时,文本中每个词汇的理解不仅依赖于前面所存储的信息,也与后面输入的信息有关。LSTM 是单向的网络结构,如果要同时考虑上下文的信息,就需要同时考虑文本句子前后方向的信息。在经过对模型性能和训练复杂性之间进行权衡之后,本文选择双向的 LSTM 结构作为词向量融合之后训练的网络结构,以此可以充分地理解词汇的上下文信息。GE-BiLSTM 情感分析模型的流程如图 1 所示。

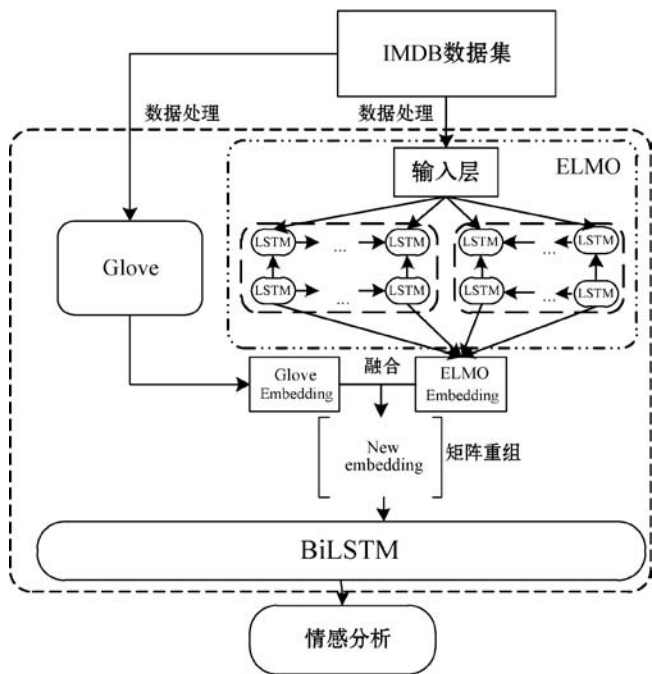


图 1 GE-BiLSTM 情感分析模型流程

在训练之前,先通过停用词表去除无用词并处理原数据,通过两个模型分别训练出词向量,经过维度调整和矩阵处理后,将预训练语言模型与 Glove 模型最终训练得到的矩阵运算融合成新的词向量矩阵。融合后分类交叉熵得到的目标函数为:

$$D = \sum_{i=1}^{|\mathcal{X}|} \sum_{n=1}^T P_i^{\tilde{t}}(W_n) \log(P_i(W_n)) \quad (19)$$

式中: $|\mathcal{X}|$ 为词向量矩阵大小; T 是模型训练结果矩阵大小; \tilde{t} 是训练中含有情感因素的词语; W 是预训练

语言模型分词过后的词语。具体融合细节图如图 2 所示。

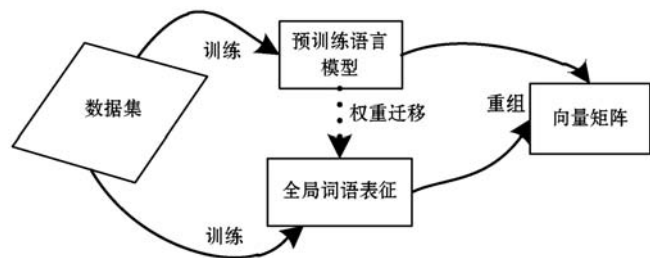


图 2 融合细节图

将融合之后的结果作为网络结构的输入放到输入层中,再次通过 BiLSTM 网络结构进一步抓取情感因素,将输出结果作为 softmax 层的输入,最终得到分类结果。

2.4 情感分析

最终生成的词向量矩阵通过 BiLSTM 网络训练,在得到情感特征向量后,利用向量矩阵做分类,softmax 函数是归一化的指数函数,其定义如下:

$$X = \frac{W^{Z_d}}{d} \sum_{i=1}^d W^{Z_i} \quad (20)$$

式中: X 是经过网络结构训练后的矩阵; d 是输入数据的向量维度, W 是输入层的向量矩阵。设 softmax 函数的输入数据为 d 维的向量 W , 最终 softmax 函数结果也是一个 d 维的向量,每个维度的值在 0 到 1 之间。利用 softmax 输出两个取值在 0 到 1 之间的数值,通过数值判断最终输出的概率,选择输出概率最大的作为文本分类的结果。GE-BiLSTM 算法流程如算法 1 所示。

算法 1 GE-BiLSTM

1. 通过对数据集处理成句子与标签对应形式
2. 通过式(4)和式(14)得到融合后的情感词向量矩阵
3. for i in $[1, N]$
4. 使用预训练语言模型与传统词向量模型训练词语表示
5. 使用式(15)的最大条件概率再次定义情感词权重
6. 运算重组训练后的词向量矩阵
7. end
8. 将 BiLSTM 再次训练之后的向量矩阵,通过式(19)更新权重参数
9. 将最终的词语表征矩阵输入到 softmax 分类器中,输出最终的结果

3 实验

本文实验数据集采用了 IMDB 大型电影评论集。该评论集具有良好的特征情感词,评论集中共有 50 000 条含有标签的评论,其中积极的文本为 25 000 条,消

极文本为 25 000 条。在训练时,采用 5:1 的比例划分训练集与测试集。训练集文本为 40 000 条,测试集文本为 10 000 条。通过统计数据集文本,每条文本评论在 500 词左右,本次实验对训练词汇设定统一长度,最大词汇数设定为 500,超过部分进行截断,不足的补 0。

3.1 实验数据集

本文所选数据集为 IMDB 大型电影评论集,该评论语料集中主要为积极评论与消极评论两种,表 1 所示为数据集中部分积极与消极评论。

表 1 实验数据集

情感分类	评论样本
积极	(1) "All the world's a stage and its people actors in it"—or something like that. Who the hell said that theatre stopped at the orchestra pit—or even at the theatre door? Why is not the audience participants in the theatrical experience, including the story itself? (2) I came in in the middle of this film so I had no idea about any credits or even its title till I looked it up here, where I see that it has received a mixed reception by your commentators.
消极	(1) I saw the capsule comment said "great acting." In my opinion, these are two great actors giving horrible performances, and with zero chemistry with one another, for a great director in his all-time worst effort. Robert De Niro has to be the most ingenious and insightful illiterate of all time. (2) This is the biggest insult to TMNT ever. Fortunately, officially Venus does not exist in canon TMNT. There will never be a female turtle, this took away from the tragic tale of 4 male unique mutants who will never have a family of their own, once gone no more.

3.2 实验参数

本文所提出的 GE-BiLSTM 模型训练参数如表 2 所示。该模型训练时使用的激活函数是 tanh,可以有效地降低在训练时梯度爆炸的概率。常用的 SGD 优化函数在优化过程中,容易陷入局部次优解,相比较下 Adam 优化函数可以直接并入梯度一阶矩估计中,其鲁棒性更强。

表 2 实验参数

参数名称	配置
深度学习框架 1	TensorFlow 1.9.0
深度学习框架 2	Keras 2.2.4
神经元激活函数(activation)	tanh
优化函数(optimizer)	Adam
Dropout	0.2

续表 2

参数名称	配置
网络迭代次数 (epochs)	6
词向量维度 (input_dimension)	512
批尺寸 (batch_size)	128

3.3 评价标准

本文选取查准率 P (precision)、召回率 R (recall) 和 F1 值 (F1-score) 作为情感分析的评价标准。查准率与召回率是一对矛盾量,为了达到一个较为平衡的评价标准,需要用到 F1 值来衡量。F1 公式如下:

$$F1 = \frac{2 \times P \times R}{P + R} \quad (21)$$

3.4 对比实验

为了验证 GE-BiLSTM 情感分析模型的有效性,本文将分别与 Glove-LSTM、ELMO-LSTM、Glove-BiLSTM、ELMO-BiLSTM 模型进行对比实验。

图 3 为 Glove 模型与双向 LSTM 训练后得到的 acc/loss 图,图 4 为 ELMO 预训练语言模型与双向 LSTM 训练后得到的 acc/loss 图,图 5 为本文模型的 acc/loss 图。可以看出,GE-BiLSTM 具有更好效果,训练集在第 6 轮左右,数据收敛趋于平缓。具体实验结果数据对比如表 3 所示。

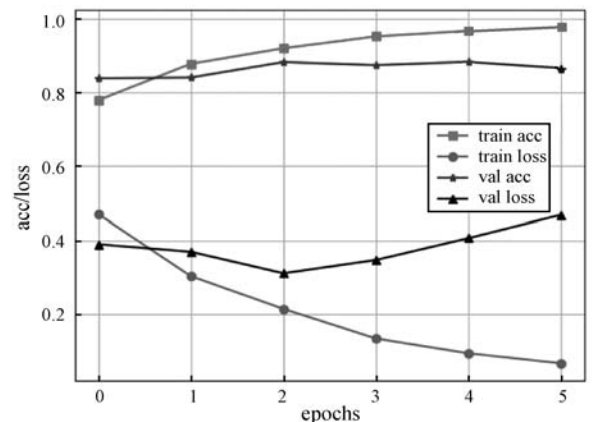


图 3 Glove-BiLSTM acc/loss 图

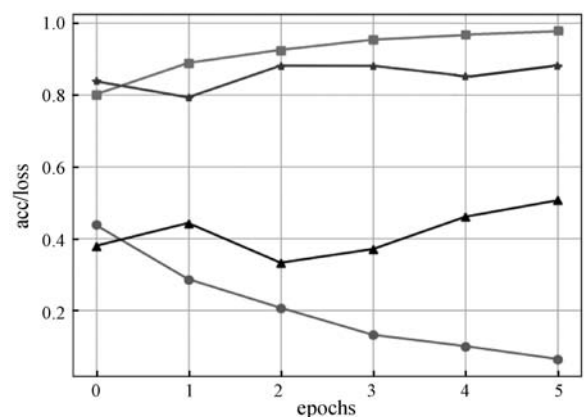


图 4 ELMO-BiLSTM acc/loss 图

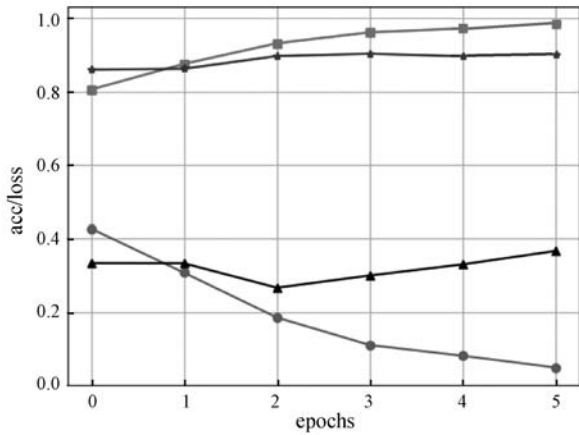


图5 GE-BiLSTM acc/loss 图

表3 实验结果对比

模型	Acc/%	P	R	F1
Glove-BiLSTM	87.5	0.893	0.869	0.881
ELMO-BiLSTM	87.9	0.895	0.826	0.859
Glove-LSTM	85.3	0.840	0.826	0.855
ELMO-LSTM	86.8	0.834	0.922	0.876
GE-BiLSTM(本文)	90.6	0.908	0.905	0.906
文献[16]方法	88.3	0.842	0.826	0.882

本文还与文献[16]方法做比较,通过结合 LDA 主题模型来捕捉单词之间的语义和情感相似性,分析语义成分,再结合上下文语义结构,通过无监督概率模型学习单词向量,最终将生成语义向量通过 SVM(支持向量机)进行情感分类。实验证明,本文提出的 GE-BiLSTM 情感分析模型具有更好的分类效果,F1 值比 ELMO-BiLSTM 方法提升 0.045,比 Glove-BiLSTM 方法提高 0.025。

4 结 语

本文提出了一种基于预训练语言模型 ELMO 与传统词向量模型融合的 GE-BiLSTM 情感分析模型。通过模型输出向量矩阵调整后,进行词向量矩阵运算融合,在文本向量化表示部分增加了文本的特征描述及上下文词汇之间的联系,减少了向量矩阵的稀疏度,再通过双向 LSTM 的训练可以更好地抓取情感因素。实验表明,GE-BiLSTM 情感分析模型具有更好的分类效果。下一步将对文本的特征描述做进一步研究。

参 考 文 献

[1] 陈龙,管子玉,何金红,等.情感分类研究进展[J].计算机研究与发展,2017,54(6):1150-1170.
 [2] 周枫,李荣雨.基于BGRU池的卷积神经网络文本分类模型[J].计算机科学,2018,45(6):235-240.

[3] Biswas E, Vijay-Shanker K, Pollock L. Exploring word embedding techniques to improve sentiment analysis of software engineering texts[C]//Proceedings of the 16th International Conference on Mining Software Repositories. IEEE, 2019: 68-78.
 [4] Cotterell R, Schütze H. Morphological word embeddings [EB]. arXiv:1907.02423,2019.
 [5] Abdeddaïm S, Vimard S, Soualmia L F. The MeSH-Gram neural network model: Extending word embedding vectors with MeSH concepts for semantic similarity[J]. Studies in HealthTechnology and Informatics,2019,264:5-9.
 [6] Yao S B, Yu D T, Xiao K L. Enhancing domain word embedding via latent semantic imputation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM,2019:557-565.
 [7] Ethayarajh K, Duvenaud D, Hirst G. Understanding undesirable word embedding associations [EB]. arXiv:1908.06361,2019.
 [8] Liu Y, Liu Z Y, Chua T S, et al. Topical word embeddings [C]//29th AAAI Conference on Artificial Intelligence. AAAI,2015:2418-2424.
 [9] Li D, Zhang J, Li P. TMSA: A mutual learning model for topic discovery and word embedding[C]//Proceedings of the 2019 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics,2019:684-692.
 [10] Nugaliyadde A, Wong K W, Sohel F, et al. Enhancing semantic word representations by embedding deeper word relationships[C]//Proceedings of the 2019 11th International Conference on Computer and Automation Engineering. ACM, 2019:82-87.
 [11] Kim Y. Convolutional neural networks for sentence classification[EB]. arXiv:1408.5882,2014.
 [12] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation [C]//The 2014 Conference on Empirical Methods in Natural Language Processing,2014:1532-1543.
 [13] 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. Association for Computational Linguistics,2018:2227-2237.
 [14] Schuster M, Paliwal K K. Bidirectional recurrent neural networks[J]. IEEE Transactions on Signal Processing,1997,45(11):2673-2681.
 [15] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation,1997,9(8):1735-1780.
 [16] Maas A L, Daly R E, Pham P T, et al. Learning word vectors for sentiment analysis[C]//Proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics,2011:142-150.