

通用词向量模型在文本多目标分类中的性能比较

王德志¹ 梁俊艳²

¹(华北科技学院计算机学院 河北 廊坊 065201)

²(华北科技学院图书馆 河北 廊坊 065201)

摘要 针对通用词向量模型在文本多目标分类中的不同性能评价比较问题,基于微博灾害数据集,设计四种多目标分类神经网络模型。通过实验,对比分析同一词向量模型在不同分类模型中的性能差异;分析不同词向量模型在分类模型中的性能特点;对模型训练时间和测试准确性进行分析。实验结果表明,Word2vec 模型在 CNN 和 LSTM 网络模型中针对微博灾害数据准确率最高。

关键词 词向量模型 情感分类 灾害数据集 自然语言处理

中图分类号 TP391

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2022.07.029

PERFORMANCE COMPARISON OF GENERAL WORD VECTOR MODEL IN TEXT MULTI-OBJECTIVE CLASSIFICATION

Wang Dezhi¹ Liang Junyan²

¹(Department of Computer Engineering, North China Institute of Science and Technology, Langfang 065201, Hebei, China)

²(Library, North China Institute of Science and Technology, Langfang 065201, Hebei, China)

Abstract Aiming at the problem of different performance evaluation and comparison of general word vector model in text multi-objective classification, we design four neural network models with multi-objective classification based on Weibo disaster data set. Through experiments, we compared and analyzed the performance differences of the same word vector model in different classification models. We analyzed the performance characteristics of different word vector models in classification models. And then the training time and testing accuracy of the model were analyzed. The experimental results show that Word2vec model has the highest accuracy for microblog disaster data in CNN and LSTM network models.

Keywords Word vector model Emotion classification Disaster data set Natural language processing

0 引言

文本数据多目标分类中,神经网络模型一般是基于经典的多层感知器(Multilayer Perceptron, MLP)、卷积神经网络(Convolutional Neural Networks, CNN)、循环神经网络(Recurrent Neural Network, RNN)和长短期记忆网络(Long Short-Term Memory, LSTM)^[1-2]。通过对这些网络进行修改和优化,构建各种不同应用领域的文本分析神经网络模型。所有的神经网络模型均需要对输入的文本数据进行预处理,把文本数据转

换为向量数据然后进行处理。而在文本数据转换为向量的研究中,主要集中在词向量模型上^[3]。谷歌公司基于词袋模型原理于 2013 年发布了 Word2vec 模型,它是将词表征为实数值向量的高效工具,利用深度学习的思想,通过大量数据训练,把对应文本中词处理简化为 K 维向量空间中的向量运算,而向量空间上的相似度可以用来表示文本语义上的相似度^[4-5]。2014 年斯坦福大学学者提出了基于全局矩阵分解方法(Matrix Factorization)和局部文本框捕捉方法的 GloVe 模型,该模型利用了全局词共现矩阵中的非零数据来训练,而不是只用了某词的局部窗口信息。Facebook

于 2016 年开源了基于 CBOW 模型原理的利用上下文来预测文本的类别的词向量与文本分类模型工具 fast-Text,其典型应用场景是带监督的文本分类问题。fast-Text 结合了自然语言处理和机器学习中最成功的理念,使用词袋、n-gram 袋表征语句和子字信息,并通过隐藏表征在类别间共享信息,从而提升自然语言情感分类的准确性和速度^[6]。ELMo 模型是 AllenNLP 在 2018 年 8 月发布的一个上下文无关模型,其是一种双层双向的 LSTM 结构^[7]。2018 年 10 月,谷歌公司发布了基于所有层中的左、右语境进行联合调整来预训练深层双向表征的新语言表示模型 BERT,其特点是可以更长的捕捉句子内词与词之间的关系,从而提升文本预测的准确性^[8-9]。这些模型都实现了文本分词的向量化,从不同的角度对词语进行了量化。但是,这些通用词向量模型在相同的神经网络结构中能否体现出相同的性能,尚缺少对比分析。

1 通用词向量模型

1.1 Word2vec 模型

Word2vec 词向量可以基于连续词袋模型(CBOW)和跳字模型(skip-gram)获得。其中 CBOW 模型是基于上下文来预测当前的词,而 skip-gram 是基于当前的词来预测上下文的词,适用于大型语料库中。在 skip-gram 模型中,通过给定的单词 w_i 来计算出上下文 $S_{wt} = (w_{i-k}, \dots, w_{i-1}, w_{i+1}, \dots, w_{i+k})$ 。其中 k 表示上下文中单词的数量, $k \leq n$, n 为文本的大小,则 c-skip-n-gram 模型为:

$$C = \{w_{i1}, w_{i2}, \dots, w_{kn} \mid \sum_{i=1}^{n-1} t_{i+1} - t_i \leq c + 1\} \quad (1)$$

式中: c 表示当前词所在上下文的长短; n 表示训练文本的数量。其训练目标函数采用最大平均对数函数,表示为:

$$P = \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c} \log p(w_{t+j} \mid w_t) \quad (2)$$

利用上述公式进行模型训练,当样本误差最小时,训练得到的每个单词的输入到隐藏层权值参数值就是该输入词的词向量。

1.2 GloVe 模型

GloVe 模型是一种基于词共现矩阵的词向量表示方法。在共现矩阵 X 中,坐标为 (i, j) 处的值 x_{ij} 表示目标词 w_i 与语料库中上下文中词 w_j 共同出现的次数,因此 X 为全局非零矩阵。GloVe 模型采用最小二乘法作为训练损失函数,其表示如下:

$$J = \sum_{i,j=1}^N P(x_{ij}) (\mathbf{w}'_i \mathbf{w}_j + b_i + b_j - \log x_{ij})^2 \quad (3)$$

$$P(x) = \begin{cases} \left(\frac{x}{x_{\max}}\right)^\beta & x \leq x_{\max} \\ 1 & x > x_{\max} \end{cases} \quad (4)$$

式中: N 为词典的大小, \mathbf{w}_i 和 \mathbf{w}_j 表示词 i 和 j 的向量。 b_i 和 b_j 为矩阵 X 中 i 行和 j 列的偏移值。 $P(x)$ 为加权函数,实现对训练中的低频词进行系数衰减操作,从而减少由于低频词造成的误差。在式(4)中 β 为常量,根据实验经验,一般为 0.75。

1.3 fastText 模型

fastText 模型采用类似 CBOW 方法,采用三层网络结构,即输入层、隐层和输出层。其中输入层输入的不是简单的单词数据,而是对单词进行深一步的分词结构。设分析文本为 S ,其由 n 个单词构成 $S = \{w_1, w_2, \dots, w_n\}$ 。则输入的第 t 个词 w_t 被进一步分解为 k 个子词 $w_t = \{c_1, c_2, \dots, c_k\}$,则最终输入的训练文本为:

$$S = \sum_{i=1}^n \sum_{j=1}^k c_{ij} \quad (5)$$

该模型输出为文本的标签。在做输出时模型采用分层 Softmax,使用哈夫曼编码对标签进行处理,从而极大地降低了运算的复杂度。

模型的目标函数如式(6)所示。式中: N 为语料库数量; y_n 为第 n 个输入文本对应的类别; $f(x)$ 为 softmax 损失函数; x_n 为文本的归一化处理值; A 和 B 为权重矩阵。最终训练得到输入层到隐藏的权重参数就是分词向量。

$$P = -\frac{1}{N} \sum_{n=1}^N y_n \log(f(\mathbf{B} \mathbf{A} x_n)) \quad (6)$$

1.4 ELMo 模型

ELMo 采用双向 LSTM 神经网络模型,其由一个向前和一个向后的 LSTM 模型构成,目标函数就是取这两个模型的最大似然数。其中,向前 LSTM 结构为:

$$p(t_1, t_2, \dots, t_N) = \prod_{k=1}^N P(t_k \mid t_1, t_2, \dots, t_{k-1}) \quad (7)$$

反向 LSTM 结构为:

$$p(t_1, t_2, \dots, t_N) = \prod_{k=1}^N P(t_k \mid t_{k+1}, t_{k+2}, \dots, t_N) \quad (8)$$

目标函数为:

$$P = \sum_{k=1}^N (\log p(t_k \mid t_1, t_2, \dots, t_{k-1}) + \log p(t_k \mid t_{k+1}, t_{k+2}, \dots, t_N)) \quad (9)$$

ELMo 首先通过训练此模型完成一定的任务,达到要求,然后就可以利用输入特定词,通过上述公式获得词向量,即把双向 LSTM 模型的每一个中间的权重进行求和,最终就得到词向量。

1.5 BERT 模型

BERT 模型采用类似 ELMo 结构,采用双向模型。但是它所使用不是 LSTM 模型,而是 Transformer 编码器模型,其目标函数为:

$$P = F(w_i | w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_n) \quad (10)$$

BERT 与传统的注意力模型的区别在于,它采用直接连接注意力机制。通过多层注意力结构,实现多头注意力能力。

2 微博灾害信息文本数据向量化

为实现文本多目标分类目的,本文采用微博灾害数据集为研究对象。微博灾害数据集来自于 CrisisNLP 网站。其提供了 2013 年到 2015 年的 2 万多条灾害相关的微博数据,并对微博数据进行了分类标注。为了实现利用各类通用词向量模型对相同灾害数据集处理结果输出格式的标准化,针对 5 种常用词向量模型,采用离线词向量数据集和词向量生成网络两种模式处理灾害数据集中的关键词。其中,针对 Word2vec、GloVe 和 fastText 三个模型采用官方提供的基于大量新闻信息训练模型生成的通用词向量数据集。通过查询常用词向量数据集的形式,在此数据集中查询提取出灾害数据集中对应的关键词的向量值,并进行存储。对于 ELMo 和 BERT 模型,由于没有常用词向量数据集,因此采用官方训练好的模型参数构建词向量神经网络模型,然后对灾害数据集中的关键词利用此神经网络生成对应的词向量值,并进行存储。为了实现标准化对比,5 类模型最终生成的每个关键词的词向量维度为(1,300)。如果词向量达不到此维度,则数据进行补零处理,从而保证所有模型处理完后每个关键词都具有(1,300)维度。

3 多目标分类神经网络模型构建

针对微博灾害数据集的词向量维度和多目标分类要求,基于四种经典模型(MLP、CNN、RNN 和 LSTM)分别设计四种多目标分类模型。

3.1 多层神经网络模型(MLP)

MLP 模型采用三层结构,输入层节点数目由词向量维度及博文中词个数构成,如式(11)所示。

$$\begin{aligned} N &= K \times M \\ K &= L_{word} \\ M &= \max(count(text_{word})) \end{aligned} \quad (11)$$

式中: K 代表词向量维度采用固定值,根据分词模型确

定, M 由博文中分词的数量确定。根据微博灾害数据词向量数据集的维度,MLP 输入层为(1,50×300)个神经元,中间层为(1,256)个神经元,输出层为(1,6)个神经元。

输出层函数采用 softmax 函数,如式(12)所示,隐层输出函数 f^2 采用 sigmoid 函数, w^1 和 w^2 分别表示输入层到隐层的全连接网络权重和隐层到输出层的权重。

$$f(X) = \text{softmax}(b^2 + w^2(f^2(w^1 X + b^1))) \quad (12)$$

$$f^2(a) = 1/(1 + e^{-a}) \quad (13)$$

为了提升模型的泛化能力,设计模型的 Dropout 正则化参数为 0.35,最终输出层为(1,6)。

3.2 卷积神经网络模型(CNN)

为了保障模型的运行效率,根据在灾害数据集的数量和词向量维度,CNN 模型采用具有 2 层卷积层的一维卷积模型结构,如图 1 所示。模型中卷积层为 2 层,卷积单元维度为(3,1)结构,输出结果为 Conv1D_1-out 和 Conv1D_2-out。最大池化层单元维度为(2,1)结构,输出结果为 MaxPooling_1-out 和 MaxPooling_2-out。在第 2 层最大池化层后是平铺层(Flatten-out)和隐藏层(Dense-out),最后为输出层(out-data)。

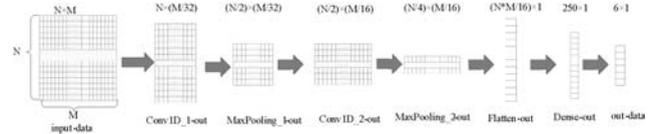


图 1 卷积神经网络

根据微博灾害数据词向量数据集的维度,CNN 输入层维度为(50,300)结构,两个卷积层激活函数采用“ReLU”函数,输出层激活函数采用“sigmoid”函数,优化器采用“Adam”。

3.3 递归神经网络模型(RNN)

循环神经网络采用序列式结构,当前单元的计算结果与前一个周期的单元结果建立关联关系,获得新的结果,同时为一个周期的单元结果提供输入。如此循环构成递归神经网络模型。RNN 采用 m 对 n 结构,如图 2 所示。

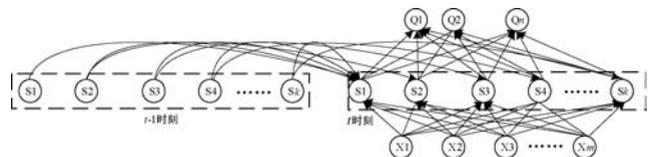


图 2 递归神经网络模型

根据微博灾害数据词向量数据,输入层维度为(1,6),隐藏层维度为(1,256)结构,输出层维度为(1,6)结构。隐藏层激活函数采用“ReLU”函数,输出层激活函数采用“sigmoid”函数,优化器采用“Adam”。为

保证模型的泛化性,避免过拟合发生,Dropout 正则化参数为 0.35。

3.4 长短期记忆网络模型(LSTM)

长短期记忆网络主要为解决 RNN 网络无长期记忆能力问题而提出。其主体结构与 RNN 网络相同,但是为实现对数据的长期记忆,在 LSTM 神经网络中,每一个神经元相当于一个记忆细胞,通过输入闸门、遗忘闸门和输出闸门实现控制记忆细胞状态的功能,如图 3 所示。

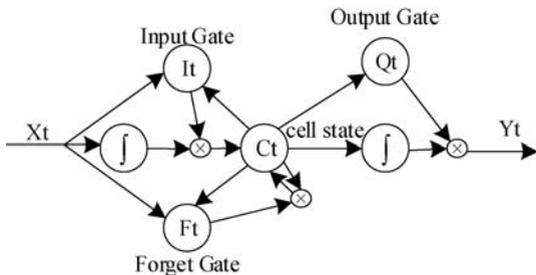


图 3 长短期记忆网络神经元结构图

根据微博灾害数据词向量数据,输入层维度为(1,32)。隐藏层维度为(1,256)结构。输出层维度为(1,6)结构。隐藏层激活函数采用“ReLU”函数。输出层激活函数采用“sigmoid”函数。优化器采用“Adam”。

4 实验与结果分析

本实验进行了 5 个通用词向量模型在 4 个多目标分类模型中的对比分析。实验中共使用 21 125 条微博数据,其中 70% 用来进行模型训练,30% 用来模型测试。

4.1 词向量模型在不同分类模型中结果分析

图 4 展示 Word2vec 模型在不同分类模型中的处理结果。可以看出,准确率最高为 LSTM,CNN 和 MLP 接近,最低为 RNN。RNN 出现了过拟合现象。CNN 准确率波动较大。RNN 在后半程的迭代周期,准确率提升明显。

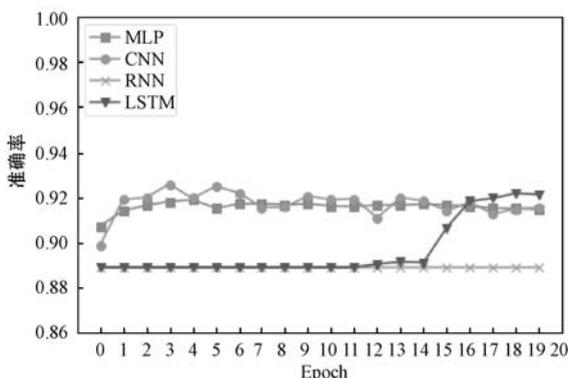


图 4 Word2vec 模型结果

图 5 展示 GloVe 模型在不同多目标分类模型中的处理结果。可以看出,GloVe 模型数据在 CNN 与 MLP 中性能接近,LSTM 次之,RNN 最低,RNN 模型波动较大。LSTM 在最后迭代周期准确率有提升。

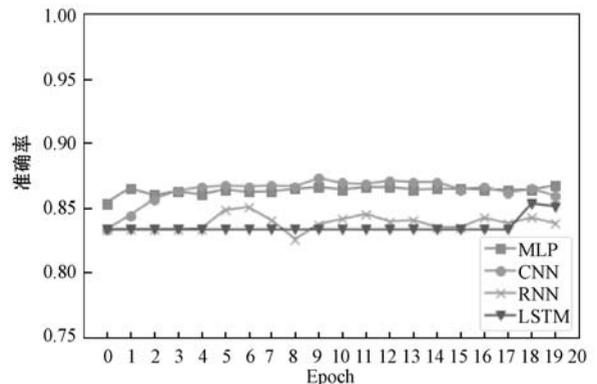


图 5 GloVe 模型结果

图 6 展示了 fastText 模型在不同多目标分类模型中的处理结果。可以看出,CNN 具有较高的测试准确率。其他三种模型准确率相似。MLP 和 LSTM 出现了过拟合现象,准确率变化不大。RNN 具有明显的波动性。

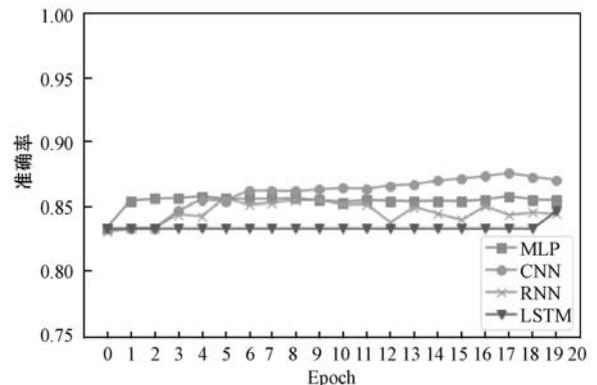


图 6 fastText 模型结果

图 7 展示了 BERT 模型在不同多目标分类模型中的处理结果。可以看出,BERT 模型数据在 RNN 和 LSTM 中没有出现过拟合现象。在测试准确性方面,RNN 的结果是最低的。CNN 据具有较好的线性增长和收敛性。

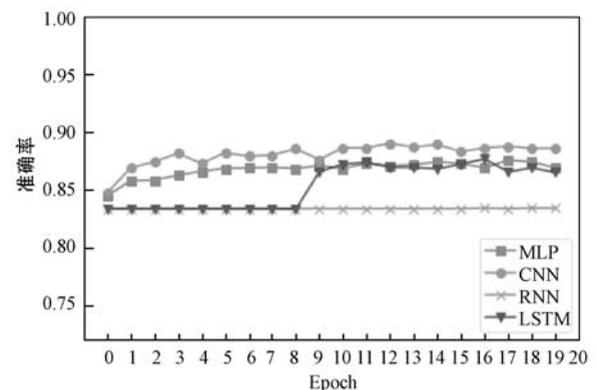


图 7 BERT 模型结果

图 8 展示了 ELMo 模型在不同多目标分类模型中的处理结果。可以看出,ELMo 模型数据在神经网络训练方面都出现过早收敛问题,测试数据的准确率基本相同,区分度不是很大。其中 LSTM 表现的数据变化不大,没有明显的线性变化过程。

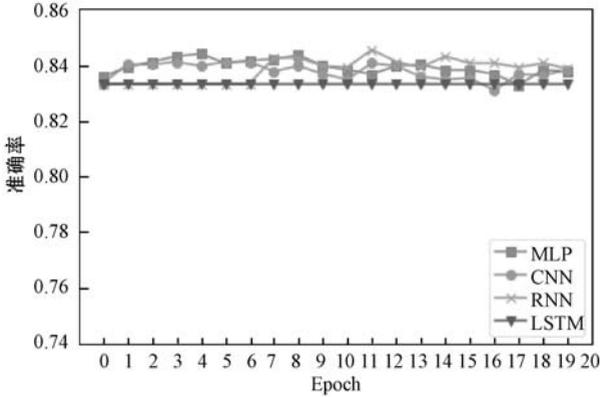


图 8 ELMo 模型结果

通过上述模型的分析,可以看出相同的通用词向量模型,在不同的多目标分类模型中的结果是不一样的,其中以 Word2vec 模型表现为最好,GloVe 和 fast-Text 模型具有相似的表现,BERT 模型在 LSTM 模型数据表现较理想。ELMo 模型在相同神经网络初始参数条件下出现了过早收敛现象,测试数据性能变化不大。

4.2 词向量模型在同一分类模型中结果分析

图 9 展示了 4 种分词模型在 MLP 模型中的训练结果比较。可以看出,除 ELMo 模型外,其他的训练数据较早地出现过拟合和现象。在测试数据方面,Word2vec 的准确率远远高于其他 4 种模型。而 ELMo 的测试准确率是最低的。

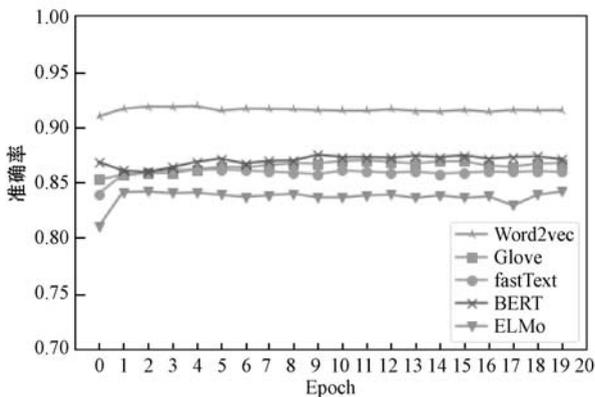


图 9 MLP 模型结果

图 10 展示了 4 种分词模型在 CNN 模型中的训练结果比较。可以看出,BERT 和 GloVe 都出现了过拟合现象。BERT、fastText 和 GloVe 的训练数据准确率保持了较好的线性增长。在测试准确率方面,Word2vec 具有最高的准确率,ELMo 准确率最低,BERT 和 GloVe

具有相似的准确率。

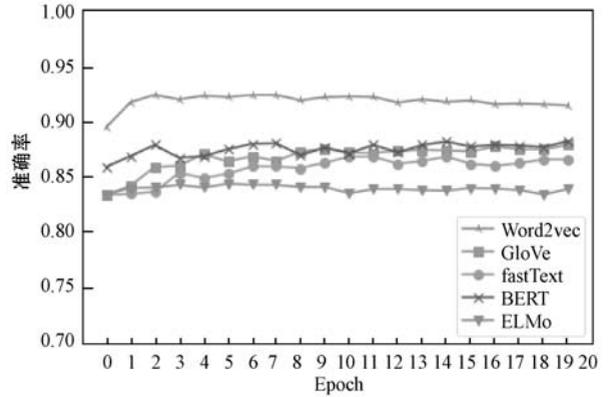


图 10 CNN 模型结果

图 11 展示了 4 种分词模型在 RNN 模型中的训练结果比较。除 GloVe 外,其他 4 种模型数据都没有出现明显的过拟合现象,尤其是 ELMo、BERT 和 fastText 模型具有相似的训练数据。GloVe 虽然训练数据有一定的线性变化,但是测试数据却出现了下降的现象,说明出现了一定的过拟合现象。

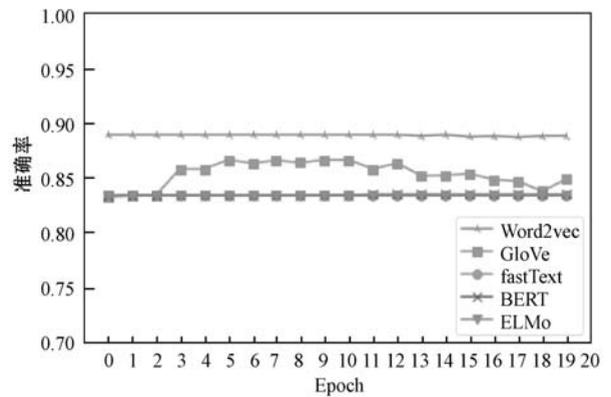


图 11 RNN 模型结果

图 12 展示了 4 种分词模型在 LSTM 模型中的训练结果比较。BERT 模型训练数据准确率有一定的线性增长,但是测试准确率却出现了下降,说明出现了过拟合现象。而其他 4 种模型训练数据准确率收敛较快,没有明显的线性增长。其中 fastText 与 ELMo 数据具有类似的训练结果数据。

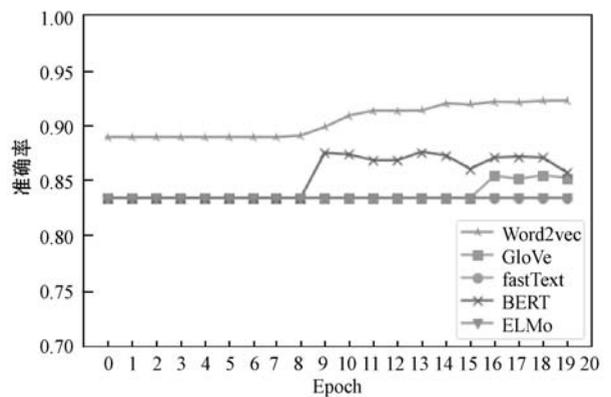


图 12 LSTM 模型结果

通过上述分析可以看出,在 MLP 和 CNN 模型中,不同的分词模型具有不同的性能参数,虽然有些模型出现了过拟合现象,但是还保持较高的测试准确率。而 RNN 和 LSTM 模型,随着没有出现过拟合现象,但是模型的准确率数据收敛较快,跟后期训练的迭代周期关系不大。而其中 fastText 和 ELMo 两种模型 RNN 和 LSTM 中具有相似的结果。

4.3 训练时间与准确率分析

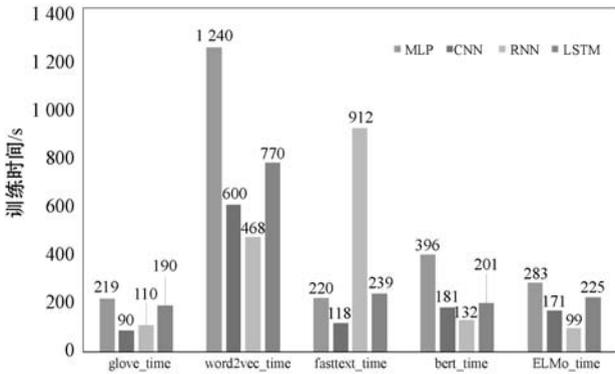


图 13 模型训练时间

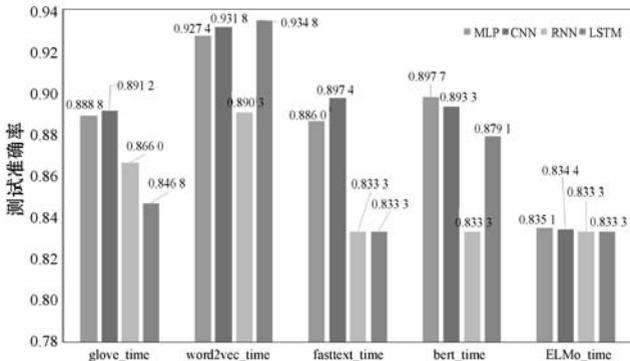


图 14 测试数据结果

图 13 和图 14 展示了不同自然语言分析模型的训练时间和最后测试数据的准确率。从训练时间上分析,Word2vec 用时最长,GloVe 模型用时最短。从测试数据结果分析,Word2vec 模型准确率最高,ELMo 模型准确率最低。Word2vec 模型用的训练时间最长,同时其测试数据准确率也是最高的。用时较少的是 ELMo 模型,但是其准确率在各个模型中变化不大,也是较小的。Word2vec 模型在 LSTM 模型中测试数据获得实验的最高准确率,达到了 0.934 8,而且模型训练时间也远远小于 MLP 时间,略高于 CNN 模型。RNN 模型虽然具有较短的训练时间,但是各个词分类模型在 RNN 上的表现不是很理想,一般都低于 LSTM 模型效果。

4.4 精确率、召回率和 F1 值比较分析

为验证模型在不同分类样本中的正确情况,对实

验数据除准确率外也进行了精确率、召回率和 F1 值计算。其中精确率和召回率采用加权平均法,如式(14)和式(15)所示。

$$PR = \sum_{i=1}^N \alpha_i \times Pre_i \quad (14)$$

$$Rec = \sum_{i=1}^N \alpha_i \times Rec_i \quad (15)$$

式中: α_i 为不同分类样本占总样本的比例; N 为分类总数,在本实验中共有 6 个分类。根据精确率和准确率计算了 F1 值,如式(16)所示,综合评价了模型的性能。实验结果如表 1 所示。

$$F1 = 2 \times (PR \times Rec) / (PR + Rec) \quad (16)$$

表 1 精确率、召回率和 F1 值比较

词向量	参数	MLP	CNN	RNN	LSTM
Word2vec	准确率	0.92	0.93	0.91	0.94
	精确率	0.91	0.92	0.39	0.89
	召回率	0.88	0.85	0.91	0.93
	F1 值	0.90	0.88	0.55	0.91
GloVe	准确率	0.86	0.88	0.83	0.83
	精确率	0.83	0.92	0.75	0.37
	召回率	0.88	0.87	0.83	0.83
	F1 值	0.86	0.89	0.79	0.51
fastText	准确率	0.88	0.89	0.84	0.91
	精确率	0.85	0.92	0.24	0.66
	召回率	0.88	0.86	0.84	0.91
	F1 值	0.87	0.89	0.37	0.76
ELMo	准确率	0.81	0.83	0.82	0.83
	精确率	0.36	0.74	0.61	0.24
	召回率	0.83	0.83	0.83	0.83
	F1 值	0.51	0.78	0.70	0.37
BERT	准确率	0.87	0.89	0.84	0.88
	精确率	0.72	0.88	0.73	0.81
	召回率	0.89	0.92	0.84	0.88
	F1 值	0.80	0.90	0.78	0.84

可以看出,各词向量模型在 CNN 模型上的分类结果精确率和准确率较高,差异不大,F1 值较高。而各模型在 RNN 上 F1 值较小,主要是由于精确率较低造成的。相同词向量模型在不同的分类模型中,一般 F1 值较低对应的准确率也较低,其中以 RNN 和 LSTM 表现较为明显,主要是由于精确率较低造成的,其中以

fastText 在 RNN 上 F1 值最低。

4.5 综合分析

为了进行对比分析,本实验过程也在斯坦福情感树银行数据集(Stanford-Sentiment-Treebank)和亚马逊评论数据库(Amazon Review Dataset)进行了验证,取得了相似的结果。但是准确率等指标相对微博灾害数据集较低,区分度不明显。可能原因是微博灾害数据集上各分类数据分布较均匀,同时在文本隐含信息中具有较高的相似性。而另外两个数据集通用性较强,文本隐含信息过多,分类结果不理想。

综上所述,从综合效果上看,Word2vec 词向量模型在 LSTM 网络模型中表现出了最佳的性能。这可能有三方面的原因。首先,在灾害数据集的预处理中,采用了基于词频的关键词提取。这导致降低了分词之间的上下文联系,从而导致基于上下文关系的词向量模型性能的下降,如 fastText 和 ELMo 模型。其次,由于 LSTM 神经网络模型具有长时记忆功能,通过迭代训练,能够发现训练数据集中的共同属性。因此,在特定领域数据集方面,对具有相同特点文本学习能够体现出更好的性能。最后,就是 Word2vec 词向量模型网络体现结构特点,Word2vec 没有过多依赖上下文的关系进行词分量,而是更多依靠多维度的信息描述方法。虽然过多的维度描述导致了词向量数据存储规模的增大,降低了处理速度,但是它具有更强的通用性,因此,在分词上下文关系较弱的自然文本分析中能够体现出更好的性能。

5 结 语

本文基于微博灾害数据集,对五种通用词向量模型在文本多目标分类中进行了比较分析。设计了 MLP、CNN、RNN 和 LSTM 多目标分类模型。通过实验分析了五种通用词向量模型在不同分类模型中的特点。实验结果表明,Word2vec 模型网络体系结构由于提供了更多的文本隐含信息,因此在微博灾害数据集情感分类中具有较高的准确性。下一步工作可以针对不同通用词向量模型设计特殊的神经网络模型,提升多目标分类的准确率。

参 考 文 献

[1] 任晓奎,郭娟,陶志勇. 嵌入常识的混合注意力 LSTM 用于主题情感分析[J]. 计算机应用与软件,2019,36(10): 151-155,161.

- [2] 缪裕青,汪俊宏,刘同来,等. 图文融合的微博情感分析方法[J]. 计算机工程与设计,2019,40(4):1099-1105.
- [3] 韩建胜,陈杰,陈鹏,等. 基于双向时间深度卷积网络的中文文本情感分类[J]. 计算机应用与软件,2019,36(12): 225-231.
- [4] Pan B, Yu C C. The improved model for Word2vec based on part of speech and word order[J]. Tien Tzu Hsueh Pao/Acta Electronica Sinica, 2018,46(8): 1976-1982.
- [5] Sung Y, Jang S, Jeong Y S, et al. Malware classification algorithm using advanced Word2vec-based Bi-LSTM for ground control stations[J]. Computer Communications, 2020, 153: 342-348.
- [6] Khatua A, Khatua A, Cambria E. A tale of two epidemics: Contextual Word2vec for classifying twitter streams during outbreaks[J]. Information Processing and Management, 2019, 56(1):247-257.
- [7] Wang Y, Wang J, Lin H, et al. Bidirectional long short-term memory with CRF for detecting biomedical event trigger in FastText semantic space[J]. BMC Bioinformatics, 2018, 19:507.
- [8] Gao Z, Feng A, Song X, et al. Target-dependent sentiment classification with BERT[J]. IEEE Access, 2019, 7:154290-154299.
- [9] Yu S, Su J. Improving BERT-based text classification with auxiliary sentence and domain knowledge[J]. IEEE Access, 2019, 7: 176600-176612.

(上接第 187 页)

- [11] Girshick R, Donahue J, Darrell T. et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014:580-587.
- [12] Shoeb A H. Application of machine learning to epileptic seizure onset detection and treatment[D]. Massachusetts Institute of Technology, 2009.
- [13] Muhammad U S, Shehzad K, Muhammad H A. Epileptic seizures prediction using deep learning techniques[J]. IEEE Access, 2020, 8:39998-40007.
- [14] Truong N D, Zhou L, Kavehei O. Semi-supervised seizure prediction with generative adversarial networks [C]//2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society(EMBC), 2019.
- [15] Fürbass F, Ossenblok P, Hartmann M, et al. Prospective multi-center study of an automatic online seizure detection system for epilepsy monitoring units[J]. Clinical Neurophysiology, 2015, 126(6):1124-1131.