DOI:10.11918/j.issn.0367-6234.201709078

一种结合深度学习和集成学习的情感分析模型

金志刚, 韩 玥, 朱 琦

(天津大学 电气自动化与信息工程学院, 天津 300072)

摘 要: 随着社交媒体的不断发展,用户评价已成为网络决策的关键因素. 为了准确分析社交媒体用户评价的情感倾向性,更好地推进舆情分析、推荐算法等工作,本文通过对 Bi-LSTM 模型和 Bagging 算法的改进,提出了一种新的情感分析模型—Bi-LSTMM-B 模型. 该模型的特点在于将深度学习模型可提取抽象特征的优势和集成学习多分类器共同决策的思想相结合. 一方面在 Bi-LSTM 模型的基础上引入 Maxout 神经元,构建 Bi-LSTMM 模型,解决随机梯度下降算法中存在的梯度弥散问题,更好地优化训练过程. 另一方面,模型基于 Bagging 算法训练多个情感分类器,根据分类器性能优劣利用袋外数据为每个分类器分配指定类别的权重,并提出相应的改进投票策略,增强了模型的泛化能力. 实验结果表明:本文提出的 Bi-LSTMM-B 模型相比于传统的 LSTM 模型准确率提高 12.08%,其中 Maxout 神经元的引入对情感分析准确率有 8.28%的相对改善效果,改进后的投票策略对准确率有 4.06%的相对改善效果,并在召回率和 F 值两项指标上均优于其他对比模型. 由此证明,深度学习模型和集成学习思想相结合可提高情感分析的准确率,并具有一定的研究价值.

关键词: 情感倾向性分析;深度学习;集成学习;Bi-LSTM 模型;Maxout 神经元;Bagging 算法

中图分类号: TP391 **文献标志码**: A **文章编号**: 0367-6234(2018)11-0032-08

A sentiment analysis model with the combination of deep learning and ensemble learning

JIN Zhigang, HAN Yue, ZHU Qi

(School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: With the development of social media, users' evaluations have become a key factor in network decision-making. Owing to the necessity of making a more accurate analysis on the emotional tendency of social media users' evaluations as well as promoting public opinion analysis and recommendation algorithms, a sentiment analysis model called Bi-LSTMM-B (Bi-directional Long Short Term Memory Model with Maxout neurons in Bagging algorithm) is proposed. With the feature of combining deep learning model and the idea of ensemble learning, the model improves the Bi-LSTM model and the Bagging algorithm. On the one hand, the Bi-LSTMM model introduces the Maxout neural into the Bi-LSTM model to solve the vanishing gradient problem during the stochastic gradient descent training and optimize the training process. On the other hand, multiple emotional classifiers were trained at the foundation of the Bagging algorithm. The out of bag data assigns the weight for each classifier on specified category according to their performance. Hence the voting strategy is improved to enhance the generalization ability of the model. The experimental results indicate that the accuracy of the Bi-LSTMM-B model is improved by 12.08% compared to the traditional LSTM model. It is also superior to other contrast models in the recall rate and F value. Therein, the introduction of Maxout neurons has a relative improvement effect of 8.28% on the accuracy of sentiment analysis, while the improved voting strategy accounts for 4.06% on the accuracy. Thus, it proves that combining deep learning and ensemble learning contributes to the improvement of the accuracy of sentiment analysis, which shows some value in research.

Keywords: sentiment analysis; deep learning; ensemble learning; Bi-LSTM model; Maxout neural; Bagging algorithm

近年来,随着互联网技术的快速发展,各种各样的社交媒体不断涌现,越来越多的用户通过社交媒体平台表达自己的观点和想法. 作为一种重要的信息载体,用户评价已成为人们进行网络决策时考虑的重要因素. 情感倾向性分析,又称主客观分析,以挖掘文本信息中用户表达的情感极性为目标,其概念最早由 Nasukawa[1] 提出. 近年来,针对社交媒体用户评价的情感倾向性分析已成为自然语言处理领域研究的热点问题之一,具有重要的研究价值. 因此,本文的研究目的是提出一种文本情感倾向性分

收稿日期: 2017-09-14
基金项目: 国家自然科学基金项目(71502125)
作者简介: 金志刚(1972—),男,教授,博士生导师
通信作者: 朱 琦, zoomzq@163.com

析模型，为提高推荐算法、用户信誉评价、舆论分析等结果提供良好的模型基础.

随着深度学习的不断发展，以神经网络为代表的深度学习模型逐渐成为情感分析的主要模型. 但随着神经网络深度的增加，随机梯度下降算法在训练中会存在梯度弥散的问题. 为此，本文在Bi-LSTM模型基础上引入 Maxout 神经元，构建 Bi-LSTMM 模型，以解决该问题.

同时，将训练好的模型能具有一定适应性地应用到测试平台上，提高实际计算和预测的准确度，是模型具有实际效用的体现，因此泛化能力一直是机器学习所关注的另一个重要问题. 为了提高模型的泛化能力，本文充分利用随机采样类别样本的差异性，在传统 Bagging 算法的基础上对其投票策略进行改进.

通过以上两个部分的改进，本文构建了 Bi-LSTMM-B 情感分析模型，可得出该模型有助于提升情感分析准确率的结论.

1　相关工作

在情感倾向性分析任务中，通常有基于词典、基于机器学习和基于深度学习的三种方法. 基于情感词典的方法是一种典型的无监督学习. 早期 Kim 等[2]使用已有的情感词典，通过文本中情感词汇得分加和的方式来判断文本的情感极性. 该类方法对词典的要求很高，但互联网新词层出不穷，情感词典更新较慢，使分类性能大打折扣. 基于机器学习模型的有监督学习是情感倾向性分析的另一种重要方式. Pang 等[3]第一次提出使用有监督的机器学习模型进行情感分析，并使用朴素贝叶斯、支持向量机和最大熵三种分类算法进行尝试. Wan 等[4]提出基于 KNN 和支持向量机的文本分类方法，提高了分类的精确率，Wang 等[5]利用 N-gram 模型对文本进行建模，并采用朴素贝叶斯和支持向量机相结合的方法实现对文本的分类. 但有监督的机器学习方法需要大量人工标注的训练语料来提高模型的学习能力和泛化能力，这在实际应用中受到阻碍.

近年来兴起的深度学习方法较好地弥补了上述方法的缺陷，在图像和语音识别中取得了巨大的成功，并逐步应用到自然语言处理中[6]. Kim 等[7]使用 CNN 对短文本进行建模，完成句子级别的文本情感分类任务. ZHANG Xiang 等[8]提出基于字符级别的 CNN 模型，并用于文本情感分类.

但随着神经网络深度的增加，在模型训练过程中也出现了很多问题，如过拟合、梯度弥散等. 如何优化深度神经网络的训练过程，从而得到更高性能的网络模型，已成为深度学习研究的重点之一. 在 RNNs 基础上提出的 LSTM 模型利用其循环神经元的线性连接，解决了时间上的梯度弥散问题. Graves 等[9]基于 LSTM 的系统，完成了 2013 年电话号码的识别任务. Yao 等[10]进一步分析 Bi-LSTM 模型在中文分词中的作用. Chiu 等[11]将 Bi-LSTM 和 CNN 相结合，并证明该模型的良好表现. 但是，LSTM 模型在空间上仍存在梯度弥散的问题. 为解决这个问题，Goodfellow 等[12]将 Maxout 神经元引入到机器学习任务中. 该神经元通过选取一组线性单元中的最大值来实现它的非线性化，在语音识别、数据分类任务中均取得不错的效果. Cai[13]将 Maxout 神经元分别与 CNN 和 LSTM 模型相结合，构建 CMNN 和 RMNN 模型，并应用到语音识别领域.

同时，集成学习具有较强的泛化能力和先天的并行性，有利于提升分类的可靠性和效率. 在集成学习传统的投票策略中，认为每个基分类器都拥有相同的决定权，实际上每个基分类器的分类性能有一定差异，在为分类器赋权的过程中，如果分类器的权重相差悬殊，又会造成投票结果过分依赖分类性能好的分类器，与集成学习多个分类器共同决策的思想相违背. 因此在为基分类器赋权时，应遵循分类性能好的分类器权重高、各分类器权重差距不能悬殊的原则.

针对上述需求，首先将 Maxout 神经元引入到 Bi-LSTM模型中，构建 Bi-LSTMM 模型；其次将集成学习中经典的 Bagging 算法进行改进，结合提出的 Bi-LSTMM 模型，构建 Bi-LSTMM-B 情感倾向性分析模型；最后将该模型应用于豆瓣影评的情感倾向性分析中，为商品推荐、用户决策提供良好的理论基础，具有创新性和实用性.

2　Bi-LSTMM 模型

提出 Bi-LSTMM(Bi-directional Long Short Term Memory with Maxout neurons)模型是由传统的 LSTM 模型扩展而来. 本文借鉴两个独立 LSTM 模型联合的思想，引入 Maxout 神经元，构建基于 Bi-LSTMM 模型的文本分类模型.

2.1　Bi-LSTM 模型

LSTM 是一种引入记忆单元的特殊 RNN，其记忆单元结构见图 1. LSTM 模型可看作一种权值共享的深度神经网络，利用门信号的线性循环连接来解决时间层面的梯度弥散问题，进而学习长期的依赖关系.

除了 f_t 外，模型还有控制输入输出信息的输入门 i_t 和输出门 o_t. 如果令 x_t 和 y_t 分别是 t 时刻模型的输入输出信息，c_t 和 m_t 分别是神经单元的激活值

和输出单元的激活值,那么根据文献[14],LSTM 记忆单元算法如下:

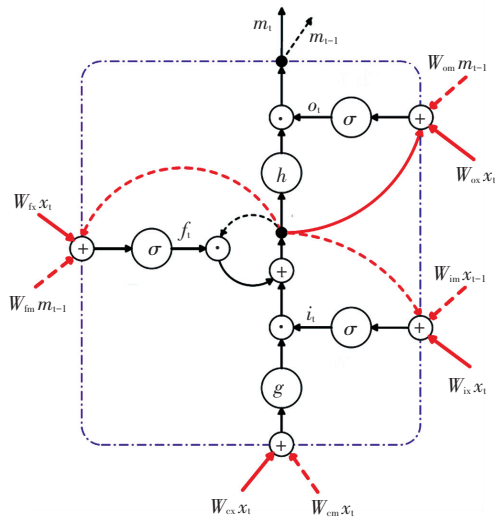

图 1 LSTM 记忆单元结构

Fig.1 Structure of LSTM memory blocks

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}C_{t-1} + b_i),$$
$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}C_{t-1} + b_f),$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cm}m_{t-1} + b_c),$$
$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}C_{t-1} + b_o),$$
$$m_t = o_t \odot h(c_t),$$
$$y_t = \varnothing(W_{ym}m_t + b_y).$$

式中:W_{ix},W_{fx},W_{ox} 和 W_{cx} 为连接输入层的权重;W_{im},W_{fm},W_{om} 和 W_{cm} 为连接激活单元的权重;b_i,b_f,b_c 和 b_o 为偏置项;W_{ic},W_{fc} 和 W_{oc} 为连接神经单元和门的对角窥视孔;W_{ym} 和 b_y 为输出层的权重和偏置;σ,g 和 h 为非线性函数,一般 σ 取 Sigmoid 函数,g 和 h 取 tanh 函数;$\odot$ 为向量对应元素的乘积;$\boldsymbol{\Phi}$ 为输出层的 Softmax 函数.

在此基础上,Yao 等[10]引入 2 个独立的 LSTM 模型,使输入信息分别正序和逆序地输入 2 个模型,然后拼接提取到的特征,构建所谓的双向神经网络—Bi-LSTM 模型. 在 t 时刻的输出向量 y_t 为

$$y_t = y_t^{\mathrm{f}} \| y_{n-t}^{\mathrm{r}}.$$

式中:y_t^{f} 和 y_{n-t}^{r} 分别为正序输入信息在 t 时刻的输出向量和逆序输入信息在 $n-t$ 时刻的输出向量;n 为输入信息的长度;$\|$ 为将 2 个输出向量进行拼接,提取最终的特征.

该模型的设计理念在于,t 时刻获取的特征同时拥有过去与将来的相关性信息,已验证其性能优于单个 LSTM 模型.

2.2 Maxout 神经元

虽然传统的 Sigmoid 函数在深度神经网络中应用广泛,但是在随机梯度下降训练中往往会出现梯度弥散的问题.

在梯度下降算法中,假设拟合函数为

$$h_\theta(X) = \theta_0 + \theta_1 X.$$

那么损失函数为

$$J(\theta_0,\theta_1) = \sum_i^n (h_\theta(x_i) - y_i),$$

式中:x_i 为样本特征 $\boldsymbol{X}$ 的第 i 个元素;y_i 为输出 $\boldsymbol{Y}$ 的第 i 个元素. 训练过程中不断更新参数,直到拟合

$$\theta_i = \theta_i - \alpha \frac{\partial J(\theta)}{\partial \theta_i}.$$

式中 α 为学习率.

传统的批量梯度下降算法使用整个样本集迭代计算参数 θ,每迭代一次复杂度为 $O(mn)$,式中 m 为样本个数,n 为每个样本特征 $\boldsymbol{X}$ 中的元素个数. 当样本 m 很大时,为减少计算复杂度,通常采用随机梯度下降算法. 即每读取一条样本,就对参数 θ 进行迭代更新,这样迭代一次的复杂度仅为 $O(n)$,大大减少了算法复杂度. 因此本文同文献[7,15]一样,选取随机梯度下降算法对模型进行训练.

Sigmoid 函数见图 2.

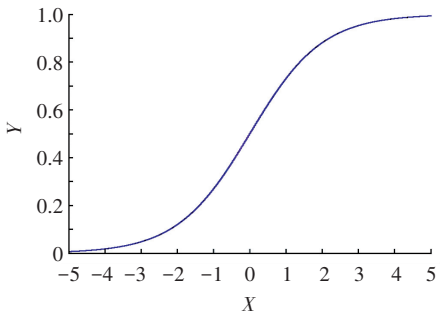

图 2 Sigmoid 函数图像

Fig.2 Sigmoid function image

由图 2 可知,若选用 Sigmoid 作为激活函数,那么在随机梯度下降算法计算导数时,随着网络深度的增加,梯度的幅值会急剧减小,导致最初几层的权重变化非常缓慢,以至于它们不能得到有效的学习,也就是所谓的梯度弥散问题. 不过,Maxout 神经元的出现恰好可解决这个问题,而且还有利于增加模型的深度[16],其结构见图 3.

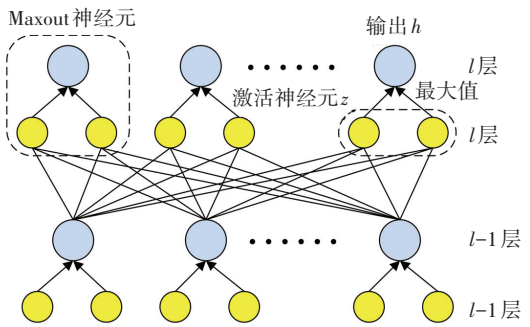

图 3 Maxout 神经元结构

Fig.3 Structure of the Maxout neurons

由图 3 可知，每一个 Maxout 神经元都是由多个不同的激活神经元构成，其输出是激活神经元中的最大值，为

$$h_l^i = \max_{j\in 1,\cdots,k} z_l^{ij}.$$

式中：h_l^i 为第 l 层第 i 个 Maxout 神经元的输出，图中用蓝色表示；k 为该 Maxout 神经元中激活神经元的个数；z_l^{ij} 为 l 层中第 i 个 Maxout 神经元的第 j 个激活神经元的激活值，图中用黄色表示. 这个激活神经元通过权重矩阵和偏置向量直接与前一层神经单元相连，为

$$z_l = \boldsymbol{W}_{l-1}^{\mathrm{T}} + b_l.$$

在随机梯度下降算法的训练过程中，Maxout 神经元的梯度为

$$\frac{\partial h_l^i}{\partial z_l^{ij}} = \begin{cases} 1, \text{if } z_l^{ij} > z_l^{is}, \forall s \in 1,\cdots,k; \\ 0, \qquad \text{otherwise.} \end{cases}$$

上式表明，Maxout 神经元的输出 h_l^i 对拥有最大激活值的激活单元 z_l^{ij} 的梯度为 1，其余为 0. 由此可见，Maxout 神经元在随机梯度下降的训练过程中，可通过产生连续的梯度来解决梯度弥散问题. 因此，Maxout 深度神经网络比传统的 Sigmoid 神经网络具有更好的训练效果.

2.3　基于 Bi-LSTMM 的文本分类模型

在此基础上提出基于 Bi-LSTMM 的文本分类模型，其结构见图 4.

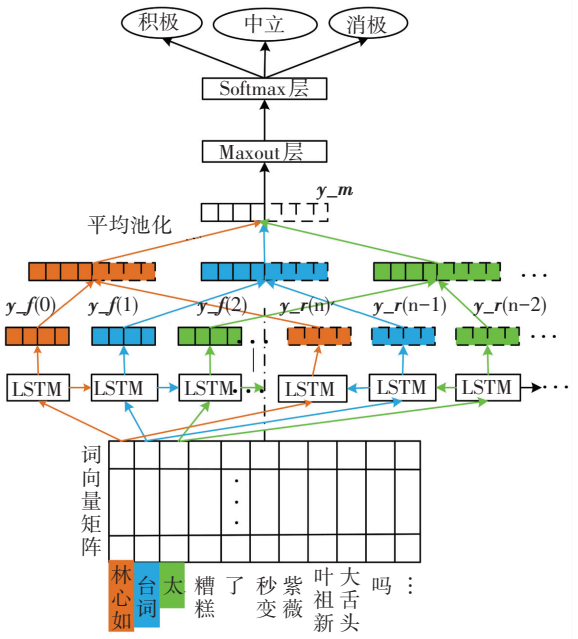

图 4　基于 Bi-LSTMM 的情感分类模型

Fig.4　Sentiment analysis model based on Bi-LSTMM

由图 4 可知，该模型在是实现文本分类时可分为以下几个关键步骤：

1）分词预处理. 利用结巴分词工具的精确模式将切分成以词为单位的形式，并把表情字符转为对应的含义.

2）构建词向量矩阵. Word2vec 是 Google 公司开源的一款将词表征为特征向量的工具，先利用该工具将中文分词处理后的词转为对应的词向量，然后构建文本的词向量矩阵. 该矩阵的行数为每条文本的词数，列数为每个词对应向量所指定的维数.

3）正反序提取词特征向量. 首先分别考虑每个词过去和未来的信息，把对应的词向量分别以正序和反序输入到两个 LSTM 模型，提取对应的正反特征. 然后把每个模型得到的词正向特征向量序列 $\boldsymbol{y_f}(0) - \boldsymbol{y_f}(\boldsymbol{n})$ 和反向特征向量序列 $\boldsymbol{y_r}(\boldsymbol{n}) - \boldsymbol{y_r}(0)$ 拼接起来，得到相应的词特征向量.

4）提取句子特征向量. 采用平均池化的方式生成句子的特征向量 $\boldsymbol{y_m}$.

5）情感分类. 将句子特征向量 $\boldsymbol{y_m}$ 输入到全连接层用 Maxout 神经元代替的 Softmax 层进行分类，实现文本积极、消极和中立的情感分类.

3　改进的 Bagging 算法

集成学习是通过构建并结合多个学习器来完成学习任务的，分为以 Bagging 和随机森林为代表的并行算法和以 Boosting 为代表的串行算法. 其中，Bagging 算法有助于训练复杂模型，防止出现过拟合问题，其原理见图 5. 该算法首先通过 T 次的随机采样得到 T 个采样集，并分别训练出 T 个独立的弱学习器（又称基分类器），再对这 T 个弱学习器通过投票策略来得到最终的强学习器.

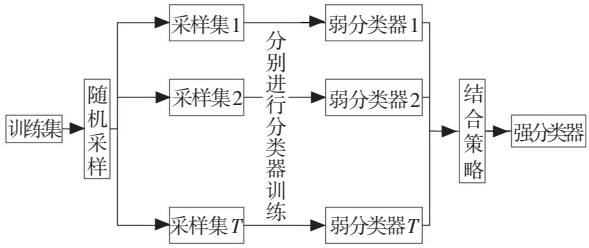

图 5　Bagging 算法原理

Fig.5　Illustration of Bagging algorithm

值得注意的是，算法中的随机采样一般采用的是自助采样法，即对于 m 个样本的原始训练集，有放回地随机采集 m 次，最终得到一个包含 m 个样本的采样集. 这样，每个基分类器的采样集不同于原始训练集和其他采样集，保证了集成学习各基分类器的差异性，有助于提升集成学习系统的泛化能力. 对于每个基分类器由自助采样法获得的训练集，理论上有 37% 的原始数据不会被选中，这些被忽略的数据称为袋外数据（Out of Bag，OOB），并可利用 OOB 数据对各基分类器的泛化能力进行预测.

在传统的 Bagging 算法中，多采用多数投票法

进行决策,忽略了基分类器的性能差异,使优劣分类器均具有相同的决策权. 传统的改进方式是为基分类器赋权,性能好的分类器权值高,性能差的分类器权值低,但往往对于每个基分类器仅赋一个权值. 由于各分类器训练集数据的代表性差异,对不同类别数据的分类能力是不同的,因此本文利用 OOB 数据的 F 值为每个基分类器的不同类别分别赋权,提出了一种基于 OOB 数据的改进投票策略.

首先,利用 OOB 数据计算各基分类器在不同类别上的可靠度 R, 计算如下:

$$R_{it}=\frac{2}{T}-\frac{1/\mathrm{OOB_}F_{it}}{\sum_{n=1}^{T}1/\mathrm{OOB_}F_{in}}.$$

式中: $i=\{-1,0,1\}$, 分别为情感类别消极、中立和积极; R_{it} 为第 t 个弱分类器在类别 i 上的可靠度; $\mathrm{OOB_}F_{it}$ 为第 t 个分类器的 OOB 数据对其在 i 类数据上的 F 值; T 为基分类器个数. R_{it} 具有以下两点特性:一是分类器在某类数据上预测性能越好,该可靠度值越大;二是满足 $\sum_{n=1}^{T}R_{in}=1$,保证在 i 类数据预测上各基分类器可靠度差异不至于过于悬殊, 避免过于依赖性能好的分类器,违背 Bagging 多分类器共同预测的初衷. 本文将该可靠度作为基分类器各类别上的权重.

随后,对于每个基分类器 t 都会对待测样本 s 产生一个情感分类 i^t,$i^t=-1,0,1$,将 t 个基分类器预测类别相同的对应权重分别累加,将权重和最大的类别作为预测的最终类别.

例如,现有 5 个基分类器,利用 OOB 数据计算的其各类别权重见表 1,表中“+”代表积极类;“-”代表消极类;“0”代表中立类.

表 1 权重样例

Tab. 1 Weight examples

基分类器		R_{it} / %	预测类别
	+	60	√
1	0	20	
	-	5	
	+	10	
2	0	25	
	-	30	√
	+	5	√
3	0	15	
	-	25	
4	+	15	
	0	30	√
	-	15	
	+	10	
5	0	10	
	-	25	√

现有一个预测样本 s. 经 5 个基分类器预测结果分别为:积极、消极、积极、中立、消极. 如采用传统的多数投票法,很难对其做出预测. 若采用本文提出的基于 OOB 数据的改进投票法,则积极类投票权重和为 0.6+0.05=0.65;消极类投票权重和为0.3+0.25=0.55;中立类投票权重和为 0.3. 因为 0.65>0.55>0.3,所以在积极类上投票可靠度(权重和)最大,可判断该样本 s 的情感分类为积极类.

4 Bi-LSTMM-B 情感分析模型

4.1 模型描述

已分析 Bi-LSTMM 模型和改进 Bagging 算法的具体内容,在此基础上提出了基于 Bi-LSTMM 和改进 Bagging 算法的 Bi-LSTMM-B 情感倾向性分析模型,其流程见图 6.

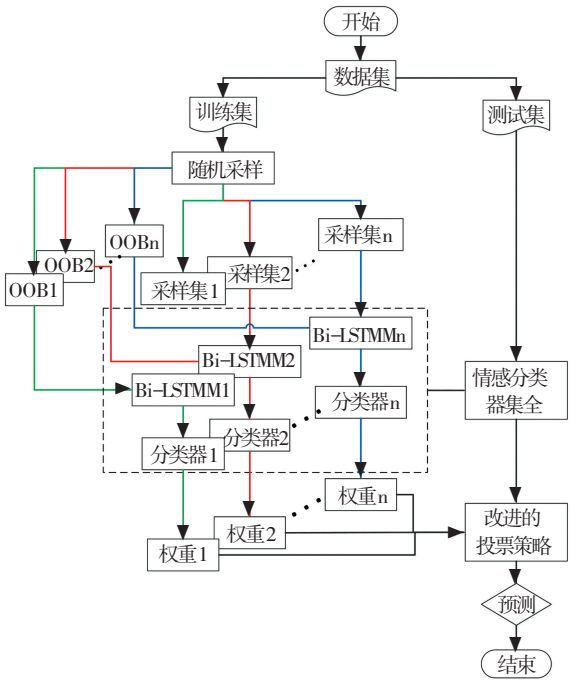

图 6 Bi-LSTMM-B 情感倾向性分析模型

Fig.6 Illustration of sentiment analysis model Bi-LSTMM-B

该模型的特点在于将集成学习的思想和深度学习的模型相结合,在传统 Bagging 算法的框架下,将引入 Maxout 神经元的 Bi-LSTMM 模型作为文本情感倾向性分析的分类器,并用 OOB 数据对分类器赋权,利用该权值提出新的投票策略,从而对预测文本的情感进行预测. 其具体实现流程如下:

1)对已标注情感标签的原始数据进行预处理,并将其分为训练集和预测集两部分.

2)Bagging 过程. 从训练集样本中随机有放回地抽取 T 份训练样本. 这样 T 个弱分类器训练过程相互独立,且保证了分类器之间存在一定的差异性.

3)在第 t 份训练样本的基础上,引入 Bi-

LSTMM 文本分类模型，训练出第 t 个 Softmax 分类器，作为第 t 个弱分类器.

4）收集第 t 份训练样本相对应的 OOB_t 数据.

5）利用步骤 3）中的训练过程和步骤 4）中的 OOB_t 数据，计算第 t 个弱分类器的在不同类别上可靠度 R_{it}.

6）重复步骤 3）~5），完成 T 个弱分类器及对应权重的训练.

7）将 T 个弱分类器和相应权重分布数据组合起来，利用改进的投票策略对预测集样本进行预测.

4.2　模型复杂度分析

神经网络通常采用矩阵运算，因而这里采用矩阵运算计算复杂度作为基础，按照运算顺序进行时间复杂度分析. Bi-LSTMM-B 模型与其余模型（模型说明见表 3）时间复杂度为

$$\begin{cases} T_{\text{Bi-LSTMM-B}} = \sum_{0}^{n} (T_1 + T_2' + T_3) + T_4; \\ T_{\text{Bi-LSTMM-MV}} = \sum_{0}^{10} (T_1 + T_2' + T_3) + T_4; \\ T_{\text{LSTM}} = T_{\text{Bi-LSTM}} = T_1 + T_2 + T_3; \\ T_{\text{RMNN}} = T_{\text{Bi-LSTMM}} = T_1 + T_2' + T_3; \\ T_1 = O(d_{\text{m}} * d * N); \\ T_2 = O(d_{\text{m}} * d * M); \\ T_2' = O(d_{\text{m}} * d * M^2); \\ T_3 = O(H * C); \\ T_4 = O(n) + O(C^2). \end{cases} \tag{1}$$

式中：T_1 为输入层 Word2vec 转化词向量操作；T_2 为利用没有引入 Maxout 层模型的训练操作；T_2' 为利用引入 Maxout 层模型的训练操作；T_3 为 Softmax 和情感分类操作；T_4 为投票操作；d 为每个词的维度；d_{m} 为最大微博长度；N 为训练语料词个数；M 为全连接层输入节点数；H 为循环神经元个数；C 为情感标签数；n 为分类器个数.

通过式（1）可发现，提出的 Bi-LSTMM-B 模型中引入的 Maxout 神经元虽然在模型训练过程中增加了模型复杂度，但远小于词向量转化过程的复杂度，因此总体上没有增加整个模型的复杂度. 同时，在 Bagging 算法中，提出的改进投票策略和相对多数投票法的时间复杂度一致，并且集成学习多分类器共同决策的思想并没有增加模型复杂度. 因此，本文提出的 Bi-LSTMM-B 模型相比于现有的经典模型，并没有增加模型的复杂度. 尽管由于需要训练多个并行分类器，在训练时间上会稍有增加，但 OOB 数据的引入可代替传统模型的十折交叉验证，因此在训练时间上并不会增加过于悬殊.

5　实验与分析

5.1　数据集和评价指标

实验数据来源于 Kaggle 公司提供的 Douban Movie Short Comments Dataset V2（DMSCD V2）. 该实验数据集包括了来自豆瓣电影网中 28 部电影的 200 多万条短评数据，每条数据由评论者 ID、电影名称、评论时间、影评内容、电影评分等组成. 从中随机选取50 000 条数据进行人工情感标注，并对情感一致性进行检测. 为保证分类准确性，这里舍弃情感不一致的数据，最终得到23 691 条积极评论数据，5 592 条消极评论数据，6 455 条中立评论数据. 最后选取积极、消极、中立评论各5 000 条作为实验数据集.

实验针对三类情感（积极、中立和消极）中的每一类 i 都建立了一个混肴矩阵. 每个混淆矩阵都包括 TP，FN，FP 和 TN. 以积极类为例，TP 代表实际是积极，预测也为积极；FN 代表实际是积极，预测不是积极（可能是中立或消极）；FP 代表实际不是积极，预测是积极；TN 代表实际不是积极，而预测也不是积极. 那么 i 为积极类的评价指标如下：

召回率：$r_{\text{Recall}i} = \dfrac{TP}{TP + FN}$.

准确率：$p_{\text{Percision } i} = \dfrac{TP}{TP + FP}$.

F 值：$F_i = 2 * \dfrac{r_{\text{Recall } i} * P_{\text{percision } i}}{r_{\text{Recall } i} + P_{\text{percision } i}}$.

用各类召回率和准确率的均值作为模型分类性能的评价指标，并用 OOB 数据的 F 值计算基分类器的权重.

5.2　实验设计

该实验数据共设计了两组对比实验，一是与传统深度学习模型的对比，二是与普通投票法的对比. 在与传统深度学习模型对比的实验中，为验证本文将 Maxout 神经元引入 Bi-LSTM 模型思想的优越性，选取经典的 LSTM 模型、Bi-LSTM 模型和文献[13]中的 RMNN 模型作为对比模型. 在普通投票法对比的实验中，为了验证本文提出的投票策略具有更高的准确性，按少数服从多数的原则，选取相对多数投票法作为对比方法.

基于此，本文对 6 个模型进行实验，模型设置见表 2.

在实验中，模型选用反向传播算法进行训练，采用随机梯度下降算法计算梯度. 实验代码语言为 Python，处理器为英特尔酷睿 i5，主频 2.7 G，操作系统为 64 bit Windows 7. 其实验参数设置见表 3. 其中

模型 1~4 采用十折交叉验证,模型 5、6 从数据集中随机选取2 000 条数据作为测试集,其余13 000 条数据作为训练集.

表 2 模型类别设置

Tab.2 Model type setting

类别	模型名称	基本模型	Maxout 层	投票策略
1	LSTM	LSTM	无	无
2	Bi-LSTM	双向 LSTM	无	无
3	RMNN	LSTM	有	无
4	本文提出的 Bi-LSTMM	双向 LSTM	有	无
5	Bi-LSTMM-MV	双向 LSTM	有	相对多数
6	本文提出的 Bi-LSTMM-B	双向 LSTM	有	基于 OOB 的改进投票策略

表 3 模型参数设置

Tab.3 Model parameter setting

参数	值	单位
文本保留最大长度	140	字符
词向量维度	100	维
Word2vec 训练窗口大小	5	—
训练批大小(batch-size)	32	—
Dropout 参数	0.5	—
单个分类器训练样本数	8 000	个
分类器个数	100	个

5.3 实验结果与分析

图 7 展示了所有模型在训练过程中测试集的分类准确率,图中纵坐标为分类的准确率,横坐标为训练次数.

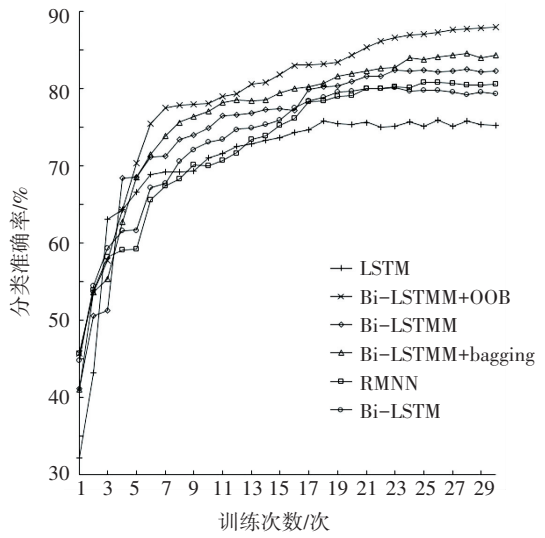

图 7 所有模型训练过程测试集分类准确率

Fig.7 The accuracy for all models on the training set

由图 7 可知,随着训练次数的增加,各模型的分类准确率逐渐提升,最终稳定在一个波动很小的区间. 最基本的 LSTM 模型在训练 20 次以后,其准确率停留在 75%左右,均低于其他模型. 在 LSTM 基础上引入 Maxout 神经元的 RMNN 模型准确率停留在 80%左右,相比 LSTM 模型提高了约 5%. 这充分说明了 Maxout 神经元在 LSTM 模型训练过程中起到了优化的作用,并有助于提升模型性能,提高分类准确率. 而 Bi-LSTM 模型是在 LSTM 模型基础上扩展而来,采用两个独立的 LSTM 正逆序拼接提取到的特征. 实验结果表明,该模型的准确率比单一 LSTM 模型提高约 4%,大约为 79%左右. 在此基础上,提出的 Bi-LSTMM 模型不仅拥有 Bi-LSTM 模型提取的特征同时具有过去与将来相关性信息的优势,而且还引入 Maxout 神经元,解决了训练过程中梯度弥散的问题,优化了训练过程. 实验表明,Bi-LSTMM 模型在准确率上明显高于其他模型,比最基本的 LSTM 模型提高了约 7%,稳定在 82%左右. 由此可见,该对比试验充分验证了本文提出的 Bi-LSTMM 模型的可行性和优越性.

同时,单一的 Bi-LSTMM 模型的准确率比集成学习下的 Bi-LSTMM-B 模型低约 5%,充分证明了集成学习多个分类器共同决策的思想可提高模型的分类性能. 同时,基于 OOB 数据的投票策略要优于传统的相对多数投票法 3%,说明了本文提出的改进 Bagging 算法可有效提高模型性能.

6 个模型训练所获得的最佳参数模型分类准确率、召回率、*F* 值见表 4.

表 4 最佳参数模型分类结果

Tab.4 Accuracy for models in optimal parameter %

模型	准确率	召回率	*F* 值
LSTM	75.85	79.60	77.88
Bi-LSTM	79.67	83.39	81.55
RMNN	80.07	84.07	82.27
Bi-LSTMM	82.13	85.74	83.73
Bi-LSTMM-MV	84.54	88.11	86.73
Bi-LSTMM-B	87.93	92.38	89.47

由表 4 可知,提出的 Bi-LSTMM-B 模型相比于 Bi-LSTMM-MV 模型准确率相对提高 4.06%,说明基于 OOB 数据的改进投票策略优于传统的投票策略. 同时,Bi-LSTMM 模型相比于 LSTM 模型准确率相对提高 8.28%,说明引入 Maxout 神经元的 Bi-LSTM 模型优于传统的 LSTM 模型. 总体上,Bi-LSTMM-B模型相比于传统的 LSTM 模型准确率提高 12.08%,在召回率和 *F* 值指标上也均优于其他模型,充分证明深度学习和集成学习相结合有助于提升情感分析的准确率.

6 结　论

为更加准确地分析社交媒体评论情感倾向性，本文基于双向循环神经网络和 Maxout 神经元提出了 Bi-LSTMM 模型，解决了随机梯度下降算法存在的梯度弥散问题，并证明该模型在 DMSCD V2 数据集上优于传统的 LSTM 等模型；其次，本文引入集成学习的思想，并利用 OOB 数据对投票策略进行了改进，增加了模型的泛化能力，并证明改进的投票策略优于传统的相对多数投票法. 本文提出的 Bi-LSTMM-B模型相比于传统的 LSTM 模型准确提高大约 12%，其原因可归为以下 4 点：1）借鉴双向循环神经网路，同时考虑过去与未来的信息，弥补单一 LSTM 模型的不足，提取特征更准确；2）引入 Maxout 神经元，替代传统的 Sigmoid 函数，提出 Bi-LSTMM 情感分类模型，有效解决梯度下降算法存在的梯度弥散问题，优化训练过程；3）借鉴集成学习多分类器共同决策的思想，提高模型泛化能力；4）提出改进的投票策略，利用 OOB 数据依据分类器在不同类别上的性能分配权重，提高情感预测的准确率. 因此，本文提出的情感倾向性分析模型不仅在情感分析准确率上有一定提升，有助于社交媒体评论的情感分析，更证明了深度学习和集成学习相结合的思想在未来自然语言处理领域的可行性和研究价值.

参考文献

[1] NASUKAWA T, YI J. Sentiment analysis: capturing favorability using natural language processing [C]// Proceedings of the 2nd International Conference on Knowledge Capture. New York, NY, USA: ACM, 2003: 70

[2] KIM S M, HOVY E. Extractingopinions, opinion holders, and topics expressed in online news media text [C]// Proceedings of the Workshop on Sentiment and Subjectivity in Text. Stroudsburg. PA, USA: Association for Computational Linguistics, 2006: 1

[3] PANG B, LEE L, VAITHYANATHAN S. Thumbs up: sentiment classification using machine learning techniques [C]// Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002: 79

[4] WAN C H, LEE L H, RAJKUMAR R, et al. A hybrid text classification approach with low dependency on parameter by integrating k-nearest neighbor and support vector machine [J]. Expert Systems with Applications, 2012, 39(15): 11880. DOI: 10.1016/j.eswa.2012.02.068.

[5] WANG S, MANNING C D. Baselines and bigrams: simple, good sentiment and topic classification [C]// Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. Stroudsburg, PA, USA: Association for Computational Linguistics, 2012: 90

[6] 曹宇慧. 基于深度学习的文本情感分析研究[D].深圳：哈尔滨工业大学(深圳)科学与技术学院,2016:2
CAO Yuhui. The research on text sentence analysis based on deep learning [D]. Shenzhen: Harbin Institute of Technology (Shenzhen). College of Computer Science and Technology, 2016: 2.

[7] KIM Y. Convolutional neural networks for sentence classification [J]. ArXiv Preprint ArXiv:1408.5882, 2014. DOI: 10.3115/v1/D14-1181.

[8] ZHANG Xiang, ZHAO Junbo, LECUN Y. Character-level convolutional networks for text classification [C]// Advances in Neural Information Processing Systems. Cambridge, MA, USA: MIT Press, 2015: 649

[9] GRAVES A, MOHAMED A, HINTON G. Speech recognition with deep recurrent neural networks [C]// Acoustics, Speech and Signal Processing, 2013 IEEE International Conference on. New York, NY, USA: IEEE, 2013: 6645

[10] YAO Yushi, HUANG Zheng. Bi-directional LSTM recurrent neural network for Chinese word segmentation [M]. Neural Information Processing. Berlin, Germany: Springer International Publishing, 2016: 345

[11] CHIU J P C, NICHOLS E. Named entity recognition with bidirectional LSTM-CNNs [J]. ArXiv Preprint ArXiv: 1511.08308, 2015.

[12] GOODFELLOW I J, WARDE-FARLEY D, MIRZA M, et al. Maxout networks [J]. Computer Science 2013, 28(3):1319

[13] CAI Meng, LIU Jia. Maxout neurons for deep convolutional and LSTM neural networks in speech recognition [M]. Speech Communication. B. V. Amsterdam, The Netherlands: Elsevier Science Publishers, 2016, 77: 53

[14] GRAVES A, MOHAMED A, HINTON G.Speech recognition with deep recurrent neural networks [C]// Acoustics, Speech and Signal Processing, 2013 IEEE International Conference on. New York, NY, USA: IEEE, 2013: 6645

[15] 何炎祥，孙松涛，牛菲菲，等. 用于微博情感分析的一种情感语义增强的深度学习模型 [J]. 计算机学报, 2017, 40(4): 773. DOI: 10.11897/SP.J.1016.2017.0073
HE Yanxiang, SUN Songtao, NIU Feifei, et al. A deep learning model enhanced with emotion semantics for microblog sentiment analysis [J]. Hinese Journal of Computer, 2017, 40(4):773. DOI: 10.11897/SP.J.1016.2017.00773.

[16] YU Dong, SEIDE F, LI Gang. Conversational speech transcription using context-dependent deep neural networks [C]// Proceedings of the 29th International Coference on International Conference on Machine Learning. USA: Omnipress, 2012: 1.

（编辑　苗秀芝）