

## 基于CNN特征空间的微博多标签情感分类

孙松涛<sup>1</sup>,何炎祥<sup>1,2</sup>

(1. 武汉大学 计算机学院, 湖北 武汉 430072; 2. 武汉大学 软件工程国家重点实验室, 湖北 武汉 430072)

**摘要:**面对微博情感评测任务中的多标签分类问题时,基于向量空间模型的传统文本特征表示方法难以提供有效的语义特征。基于深度学习的词向量表示技术,能够很好地体现词语的语法和语义关系,且可以依据语义合成原理有效地构建句子的特征表示向量。作者提出一个针对微博句子的多标签情感分类系统,首先从1个大规模的无标注微博文本数据集中学习中文词语的词向量表示,然后采用卷积神经网络(convolution neural network,CNN)模型进行有监督的多情感分类学习,利用学习到的CNN模型将微博句子中的词向量合成为句子向量,最后将这些句子向量作为特征训练多标签分类器,完成微博的多标签情感分类。2013年NLPCC(Natural Language Processing and Chinese Computing)会议的微博情感评测公开数据集中,相比最优评测结果的宽松指标和严格指标,本系统的最佳分类性能分别提升了19.16%和17.75%;采用Recursive Neural Tensor Network模型合成句子向量的方法,取得目前已知文献中的最佳分类性能,系统将2个指标分别提升了3.66%和2.89%。采用多种多标签分类器来对比不同的特征表示方法,发现基于CNN特征空间的句子向量具有最好的情感语义区分度;通过对CNN迭代训练过程的分析,体现了语义合成过程中的模式识别规律。进一步的工作包括引入更多合适的深度学习模型,并深入探索基于词向量的语义合成现象。

**关键词:**情感分类;多标签分类;词向量表示;卷积神经网络;语义合成

中图分类号:TP391

文献标志码:A

文章编号:2096-3246(2017)03-0162-08

### Multi-label Emotion Classification for Microblog Based on CNN Feature Space

SUN Songtao<sup>1</sup>, HE Yanxiang<sup>1,2</sup>

(1. School of Computer, Wuhan Univ., Wuhan 430072, China; 2. State Key Lab. of Software Eng., Wuhan Univ., Wuhan 430072, China)

**Abstract:** While the evaluation task of microblog emotion is a multi-label classification problem, the traditional text representing methods, which are usually based on vector space model, fail to provide more effective semantic features. Word embedding technology is based on deep learning, which can well capture the syntax and semantic relations between words, and build sentence representing effectively according to semantic compositionality. A multi-label emotion classification system was proposed. First, word embedding for Chinese words was learned from a large scale of unlabeled Chinese microblog text dataset. Second, the Convolution Neural Network (CNN) model was exploited to train a supervised multi-emotion classifier. Third, the learned CNN model was used to composite the feature vector for sentences from microblog. At last, these sentence vectors were treated as semantic features to train the multi-label classifier, which was used to finish the multi-label emotion classification for microblog. Based on the open dataset from microblog emotion evaluation task of NLPCC (Natural Language Processing and Chinese Computing) conference in 2013, the best performance of the proposed system achieved 19.16% and 17.75% improvement in the loose metric and the strict metric, respectively, comparing to the best performance of all the evaluation results. The state-of-art performance, which was achieved by the method of exploiting Recursive Neural Tensor Network model to composite the sentence vector, was also outperformed by the proposed system up to 3.66% and 2.89% on the two metrics. Several multi-label classifiers were employed to compare different feature representing methods, and the sentence vectors based CNN feature space were showed to have the most discriminative emotion semantic. The pattern recognition in the semantic composition procedure was showed by analyzing the training iteration of CNN model.

收稿日期:2016-08-07

基金项目:国家自然科学基金资助项目(61303115; 61373039; 61472290); 高等学校博士学科点专项科研基金资助项目(2013014111002 512)

作者简介:孙松涛(1986—),男,博士生。研究方向:自然语言处理、情感分析和机器学习。E-mail: stsun@whu.edu.cn

**Key words:** emotion classification; multi-label classification; word embedding; convolution neural network; semantic compositionality

近年来,微博已经成为人们沟通交流和情感表达的常用途径之一,产生的海量微博也为文本情感分析<sup>[1]</sup>提供了很好的数据基础<sup>[2]</sup>。相比新闻、电影和产品等评论文本,微博文本的数据量更大,包含的人类情感也更加复杂和多变。随着文本情感分析研究工作的发展,其主要研究对象已经从用户评论<sup>[3]</sup>等文本转移到微博文本上,主要研究任务也从主客观识别和正负情感分类逐步扩展为情感强度判别和多情感分类等更加复杂的问题。2013年NLPC的微博情感评测数据集中,标注了7种不同的情感,且部分微博还拥有主要和次要两种不同情感,成为了更加复杂的多标签分类问题<sup>[4-6]</sup>。

基于统计和机器学习的方法,是当前用于文本情感分析的4种主要技术<sup>[7]</sup>中最流行的一种。这些方法通常需要首先从文本空间 $T$ 中提取相关文本属性构建特征空间 $X$ ,然后利用分类器 $f$ 完成 $X$ 到标签空间 $Y$ 的映射,即 $f: X \mapsto Y$ ,其中 $X$ 又称为输入空间, $Y$ 又称为输出空间。

向量空间模型(vector space model, VSM)<sup>[8]</sup>为传统文本特征表示方法,最初在信息检索领域被提出,现已被广泛应用到文本分析的各个研究领域。文本情感分析中常用的文本属性包括 $n$ -gram、词性、情感词和TF-IDF(term frequency-inverse document frequency)等。VSM存在非常严重的数据稀疏性问题,使得分类器面临维度灾难,从而大大降低分类性能。此外,VSM采用的文本属性大多只是文本语义的表层形式,如果引入更多的语法和语义属性,就需要采用特征工程选择有效特征。

词向量表示(word embedding)<sup>[9]</sup>技术将词语表示为维度较小、数值稠密的实数向量,能够很好地捕捉词语的语法和语义关系。这些实数向量通常被称为词向量,具有很好的计算特性,即使简单的线性累加或平均后,仍然可以作为文本的有效特征。这样就能够在一定程度上缓解维度灾难。然而,根据计算语言学中的语义合成原理<sup>[10]</sup>,词向量的线性组合操作(由此构建的特征空间简称为EMB)很难有效地捕捉文本语义单元之间的递归性结构关系,并且难以区分词语之间的位置关系(例如“好/吃”与“吃/好”)。因此,一些具有语义合成能力的深度学习模型被提出来,用以解决利用词向量有效地合成更长文本段的向量表示这一难题,如MV-RNN(matrix-vector recursive neural network)<sup>[11]</sup>、RNTN(recursive neural tensor network)<sup>[12]</sup>、Tree-LSTM(tree-structured long short-term memory) network<sup>[13]</sup>和CNN<sup>[14-15]</sup>等。这些使用CNN模型的相关研究直接进行文本情感分析,特

别是针对微博的情感分类取得了很好的性能,表明其语义合成能力突出。然而,这些CNN情感分类模型仍然只是传统的单标签分类器,不能直接进行多标签分类。

对于传统的有监督分类问题,训练集为 $D = \{(\mathbf{x}_i, y_i) | 1 \leq i \leq m, \mathbf{x}_i \in X, y_i \in Y\}$ ,其中,当 $|Y| = 2$ 时为二分类问题,当 $|Y| > 2$ 时为多分类问题。对于多标签分类问题,训练集 $D^+ = \{(\mathbf{x}_i, Y_i) | 1 \leq i \leq m, \mathbf{x}_i \in X, Y_i \subset Y\}$ ,其中通常有 $|Y| > 2$ 且 $|Y_i| \geq 1$ 。可见,传统分类问题包含在多标签分类问题中,两者的特征空间都是一致的,只是标签空间有所不同。然而,一个特征向量对应多个不同标签,则表明其应当具有多种不同语义。相比传统分类方法,多标签分类方法需要具备更强的语义区分能力,或者使用语义区分度更好的特征空间。在现有传统机器学习方法的基础上,研究人员提出两种解决思路<sup>[16]</sup>:第1种是问题转换方法,即把多标签问题首先转换为单标签问题,然后利用现有的成熟方法求解,最后再将结果转换成多标签集合,代表方法有Binary Relevance(BR)<sup>[17]</sup>、Classifier Chains<sup>[18]</sup>、Random k-labelsets<sup>[19]</sup>和Calibrated Label Ranking<sup>[20]</sup>等;第2种是算法改编方法,即直接改编现有的常用机器学习方法以解决多标签问题,代表方法有Multi-Label KNN(ML-KNN)<sup>[21]</sup>、Collective Multi-Label Classification<sup>[22]</sup>、Rank-SVM<sup>[23]</sup>和Multi-Label Decision Tree<sup>[24]</sup>等。

从国内外文献来看,目前尚未发现利用CNN模型合成句子向量表示以进行微博多标签情感分类的研究和应用。作者围绕CNN模型构建一个多标签情感分类系统,充分利用CNN模型的语义合成能力从训练集中构建具有语义区分度的句子向量,并在此特征表示空间(简称为CNN)中使得多标签情感分类取得极大的性能提升。实验还选用了VSM和EMB两种特征空间作为对比,并从多标签分类方法的两种思路中分别选取BR和ML-KNN作为代表。其中,BR方法采用了NB(naïve bayes)、线性核SVM(support vector machine)和RBF(radial basis function)核SVM3种不同的基础分类器,分别记为BR-NB、BR-LIN和BR-RBF。这样,BR-NB和BR-LIN为线性分类器,BR-RBF和ML-KNN则为非线性分类器,从而可从线性可分性角度考察3个特征空间。

## 1 多标签情感分类系统

### 1.1 整体框架

所提出的多标签情感分类系统的整体框架如图1所示,其中包含3个主要流程,分别使用3种不同风格

的线型来区分:

第1个流程用实线表示,首先使用Word2Vec工具(<https://code.google.com/p/word2vec/>)中的Skip-gram模型<sup>[9]</sup>从一个大规模的无标注微博文本数据集 $U$ (详见2.2节)中学习中文词语的词向量表示,构建一个词向量查找表 $W^e$ 后,结合2013年NLPCCC的训练集进行CNN模型的有监督情感分类学习。

第2个流程用短划线表示,首先使用前一步充分训练的CNN模型,将训练集中的每个句子转化为向量表示,然后将这些句子向量和对应的情感标签输入到多标签分类器进行有监督的多标签情感分类学习。

第3个流程用点划线表示,首先利用CNN模型生成测试集的句子向量,然后输入训练好的多标签分类器进行情感分类,最终得到分类结果。

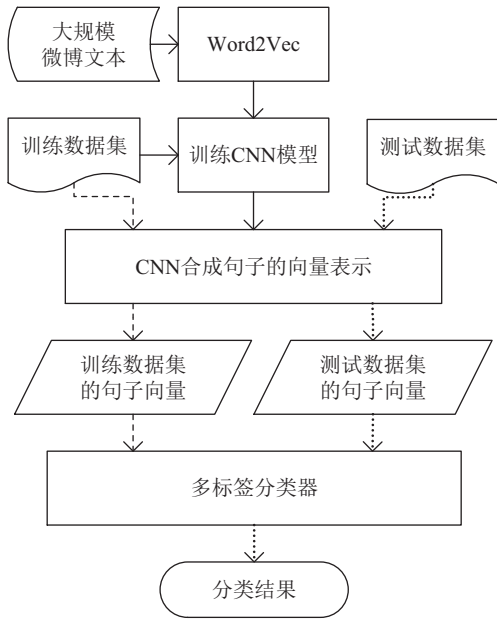


图1 整体框架

Fig.1 Framework

在这3个流程中,前面2个流程需要依次完成CNN模型和多标签分类器的训练,其中训练好的CNN模型将被进一步用来完成句子向量的合成,是整个系统的关键。由于CNN模型是传统的单标签分类器,因而需要将多标签训练集 $D^+$ 转换为单标签训练集 $D^* = \{(x_i, y_j) | \forall (x_i, Y_i) \in D^+, \forall y_j \in Y_i\}$ 。

## 1.2 CNN模型

采用一种改进的CNN模型<sup>[14]</sup>,具有多个不同宽度的卷积窗口,具体的模型结构如图2所示。

假设中文词典中词语个数为 $N$ ,单个词语的词向量维度为 $d$ , $W^e \in \mathbb{R}^{d \times N}$ 表示词向量查找表。对于微博句子 $S \in T$ ,统一表示为长度为 $k$ 的词语序列(不足时填充空白符),即有 $S = (w_1, w_2, \dots, w_k)$ ,则第 $j$ 个词语 $w_j$ 的词向量 $r_j$ 通过下面操作获得:

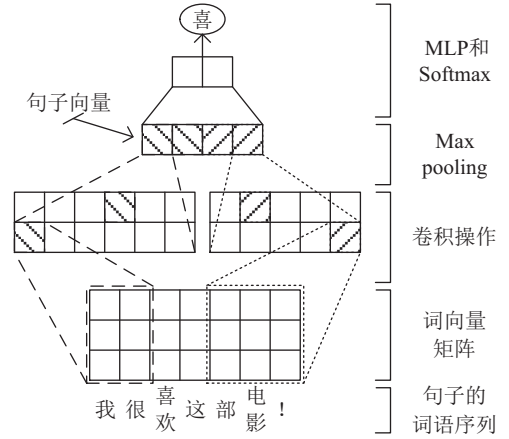


图2 CNN模型

Fig.2 CNN model

$$r_j = W^e v_j \quad (1)$$

式中, $v_j \in \mathbb{R}^N$ 为一个仅词语 $w_j$ 对应维度取值为1,其他维度全为0的 $N$ 维指示向量。

将句子中对应位置词语的词向量依次拼接起来,就得到这条句子的词向量表示矩阵:

$$R^s = r_1 \oplus r_2 \oplus \dots \oplus r_k \quad (2)$$

式中, $\oplus$ 表示行向量方向的拼接操作,则 $R^s \in \mathbb{R}^{d \times k}$ 对应图2中的词向量矩阵。

随后,对句子的词向量表示矩阵执行卷积操作。设卷积算子为 $W^f \in \mathbb{R}^{d \times d_{win}}$ , $d_{win}$ 表示卷积窗口的宽度,则:

$$\begin{aligned} r_{ij}^f &= g(W_i^f \otimes Z_j + b_i^f) \\ Z_j &= r_{j-d_{win}+1} \oplus \dots \oplus r_j \end{aligned} \quad (3)$$

式中, $r_{ij}^f$ 为第 $i$ 个卷积算子 $W_i^f$ (又称作过滤器)在词语 $w_j$ 处的卷积结果, $Z_j$ 为窗口 $d_{win}$ 范围内的词向量矩阵, $b_i^f$ 为第 $i$ 个卷积算子的偏置值, $\otimes$ 表示矩阵的内积运算, $g(\cdot)$ 为非线性变换函数Rectifier。

接着,对卷积结果执行最大化下采样操作,即获取每个卷积算子对整条句子执行操作后的最强激励信号。运算结果如下:

$$x_i = \max_{d_{win} \leq j \leq k} [r_{ij}^f], 1 \leq i \leq H \quad (4)$$

式中, $H$ 表示卷积算子的个数,最终得到 $x \in \mathbb{R}^H$ 即为句子向量。实验采用多种宽度的卷积算子,运算得到的多个句子向量直接拼接成整个句子向量。如图2所示,采用两种宽度的卷积算子,即 $d_{win}$ 分别为2和3;每种卷积算子都为2个,即2个 $H$ 都为2,因而句子向量 $x$ 的维度为4。

将句子向量 $x$ 输入到一个多层感知器MLP(multi-layer perceptron)之后,执行Softmax操作获得情感标签的概率分布,即

$$h = W^h x + b^h \quad (5)$$

$$P(y_j|S, \theta) = \frac{e^{h_j}}{\sum_{1 \leq i \leq |Y|} e^{h_i}}, 1 \leq j \leq |Y| \quad (6)$$

其中,  $\mathbf{W}^h \in \mathbb{R}^{|Y| \times H}$  和  $\mathbf{b}^h \in \mathbb{R}^H$  分别为MLP的参数矩阵和偏置量,  $\theta = \{\mathbf{W}^e, \mathbf{W}^f, \mathbf{W}^h, \mathbf{b}^f, \mathbf{b}^h\}$  为CNN模型的参数集合。

在训练集  $D^*$  中, 通过随机梯度下降算法最小化NLL(negative log-likelihood)学习最优的模型参数  $\theta$ :

$$\theta \mapsto \sum_{(S, y) \in D^*} -\log P(y|S, \theta) \quad (7)$$

采用AdaDelta优化技术, 对句子向量  $\mathbf{x}$  执行 dropout 以防止模型过拟合。

### 1.3 多标签分类器

BR方法采用问题转换的思路, 将多标签分类问题转换为  $|Y|$  个相互独立的二分类问题。为构建针对标签  $y_j \in Y$  的二元分类器, 需要将多标签训练集  $D^+$  转换为对应的二元相关训练集, 即

$$D_j = \{(\mathbf{x}_i, \phi(Y_i, y_j)) | \forall (\mathbf{x}_i, Y_i) \in D^+\} \\ \text{其中 } \phi(Y_i, y_j) = \begin{cases} +1, & y_j \in Y_i; \\ -1, & \text{其他} \end{cases} \quad (8)$$

这里的二元分类器  $\beta$  就是传统的单标签分类器, 即有  $g_j: X \mapsto \mathbb{R}$ , 即  $g_j \leftarrow \beta(D_j)$ 。实验采用微博情感分类效果最佳的NB和SVM两种分类器<sup>[25]</sup>。

接下来, 需要将  $|Y|$  个二元分类结果转换为多标签分类结果, 特别是本文实验的标签格式。假设对于测试样本  $\mathbf{x}_i$ , 预测得到的标签集合为  $Y_i$ 。对于SVM分类器, 其二元分类结果通常为  $\text{sign}(g_j(\mathbf{x}_i)) \in \{+1, -1\}$ , 其中  $g_j(\mathbf{x}_i)$  表示特征向量  $\mathbf{x}_i$  距离决策平面的距离, 在一定程度上反映出决策的置信度。最终, 实验选取  $g_j(\mathbf{x}_i)$  作为决策参考值, 并采用T-criterion策略<sup>[17]</sup>进行决策:

$$Y_i = \begin{cases} \text{top2}(Y'_i), & Y'_i \neq \emptyset; \\ Y_i^*, & \text{其他}; \end{cases} \quad (9) \\ Y'_i = \{y_j | g_j(\mathbf{x}_i) > 0, 1 \leq j \leq |Y|\}; \\ Y_i^* = \{y_j | j = \arg \max_{1 \leq j \leq |Y|} g_j(\mathbf{x}_i)\}$$

其中,  $\text{top2}(Y'_i)$  为从标签集合  $Y'_i$  中选取决策参考值最高的前两个标签或一个标签(当集合大小为1时)。当采用NB分类器时, 有  $g_j(\mathbf{x}_i) \equiv P(y_j | \mathbf{x}_i)$ , 可以直接作为决策参考值, 因而只需要修改公式中的  $Y'_i = \{y_j | g_j(\mathbf{x}_i) > 0.5, 1 \leq j \leq |Y|\}$  即可。

ML-KNN为一个采用算法改编思想的多标签分类方法, 不需要进行训练集的转换。ML-KNN算法结合惰性学习(lazy learning)和贝叶斯推理的优点: 1) 其决策边界能够自适应地根据每个新实例的近邻实例进行调整, 即具有非线性的决策边界; 2) 由于引入了不同标签的先验概率估计, 数据集不平衡问题可

以极大地得到缓解。同文献[4-6]一样, 本文在KNN算法中采用cosine作为距离度量函数<sup>[4]</sup>。

核函数在SVM中扮演重要的角色<sup>[26]</sup>, 已有研究表明文本数据应当采用线性核, 即VSM空间适用线性核。为了考察EMB空间和CNN空间的适用核函数, 引入另一个常用的RBF核函数。对于BR方法, 采用NB时记为BR-NB, 采用线性核SVM时记为BR-LIN, 采用RBF核SVM时记为BR-RBF。这样, BR-NB和BR-LIN具有线性决策面, BR-RBF和ML-KNN具有非线性决策面, 使得本文还可以从线性可分性考察3个特征空间的特性。

## 2 实验准备

### 2.1 标注数据集

采用2013年NLPCC情感评测任务的公开数据集组织实验, 其中包含一个训练集和一个测试集, 微博内容来自新浪微博。此标注数据集中共有7种不同的情感标签, 即happiness、like、sadness、disgust、anger、fear和surprise, 具有微博级和句子级的两种粒度的标签, 且每个实例有1个或者2个的情感标签。本文仅进行句子级的多标签情感分类实验, 统计信息参考文献<sup>[4]</sup>。

### 2.2 无标注数据集

为使用Word2Vec进行词向量的预学习, 通过新浪微博的API接口随机获取了2012年到2015年之间的大约  $3 \times 10^7$  条微博文本构建无标注数据集  $U$ 。经过简单的去噪和文本预处理(3.1节)之后, 共保留了  $3.12 \times 10^7$  条微博, 其中总共包含了  $1.08 \times 10^9$  个词语, 词典大小为347 886, 每条微博平均包含为34.6个词语。

### 2.3 评价指标

评测任务对包含2个情感标签的实例, 要求给出主次关系, 这点不同于一般的多标签分类问题。因而, 在常用的评价指标平均精度(average precision, AP)的基础上, NLPCC评测任务提出自定义的宽松指标(loose AP)和严格指标(strict AP)<sup>[4]</sup>。

对于多标签分类结果, 参照one\_error指标<sup>[16]</sup>定义one\_acc指标, 用以衡量预测的主情感标签出现在标准情感标签集合中的比例:

$$\text{one\_acc} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}[rk_1(\hat{Y}_i) \in Y_i] \quad (10)$$

为同时对比CNN模型与多标签分类器的单标签分类性能, 采用针对主情感标签的准确率(Accuracy, 记作acc)作为评价指标:

$$\text{acc} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}[rk_1(\hat{Y}_i) = rk_1(Y_i)] \quad (11)$$

### 3 实验过程

#### 3.1 文本预处理

使用NLPIR工具对微博文本进行分词,若词性属于集合{url,x,m,nr},则将对应的字符串作为基本语言单元。其中,词性url的对应字符串,转化为语言单元<URL>;以@开头且词性为nr的字符串,转化为语言单元<UNAME>;词性为m的字符串中的每个数字,统一转化为D表示;词性为x的字符串,如果以www.开头,则转化为<URL>,否则保持原样。

#### 3.2 超参数调节策略

使用Word2Vec工具中的Skip-gram模型从无标注数据集U中进行词向量的预学习,采用默认参数设置,训练得到维度为100、200、300和400的4种词向量查找表。

对于CNN模型,词向量维度 $d$ 的取值范围为(100,200,300,400), $d_{win}$ 的组合为(2,3,4)、(3,4,5)和(4,5,6), $H$ 在3种卷积算子上保持一致,取值范围为(50,100,150,200);从而构成了48种不同的参数组合,采用网格搜索(grid search)技术进行3倍交叉验证,最终选择使得准确率最优的一组,即 $d$ 为200, $d_{win}$ 组合为(3,4,5), $H$ 为100。

对于4种多标签分类器,每次实验前使用网格搜索技术在训练集上进行3倍交叉验证,从相应的超参数组合中选择最优的一组。4种多标签分类器的超参数取值组合为:

1)BR-NB: 仅有一个平滑因子 $\alpha$ ,取值范围(0,0.1,1.0,2.0,4.0,8.0)。

2)BR-LIN: 仅有一个正则化常数 $C$ ,取值范围(0.001,0.01,0.1,1.0,10)。

3)BR-RBF: 正则化常数 $C$ ,取值范围(0.1,1.0,10,100,1 000);核带宽的倒数 $\gamma$ 通常为特征向量维度之倒数的倍数,因而实验从(0.01,0.1,1.0,10,100)中取值后除以对应特征空间的维度(3.3节)。

4)ML-KNN: 仅有一个近邻个数 $k$ ,取值为区间[11,59]中的25个奇数。

#### 3.3 特征空间的信息

3.2节选定了CNN模型的超参数,采用3种不同窗口宽度的过滤器,每种个数都为100,因而最终的句子向量维度为300。为与CNN空间保持一致,EMBSpace同样选用维度为200的词向量,对句子中的词向量经过mean、max和min的线性操作并拼接后,得到的句子向量维度为600。

VSM空间的维度与选取的文本属性相关,对训练集进行文本预处理后,提取词语的unigram、bigram和trigram作为文本属性,然后过滤掉文档次数低于

3和文档频率高于0.95的文本属性,最终得到16 865个文本属性,即VSM空间的维度为16 865。构建句子向量时,采用文本属性的二元计数。

### 4 实验结果和分析

#### 4.1 评测结果

实验测试了3种不同句子特征表示方法在4个多标签分类器下的性能,最终的实验结果如表1所示。

表 1 实验结果

Tab.1 Experiment results

方法	looseAP	strictAP	one_acc
BR-NB+VSM	0.407 2*	0.372 9*	0.390 4*
BR-LIN+VSM	0.400 1	0.372 2	0.389 7
BR-RBF+VSM	0.398 0	0.370 7	0.388 3
ML-KNN+VSM	0.305 2	0.280 9	0.295 0
BR-LIN+EMB	0.496 5*	0.470 0*	0.488 6*
BR-RBF+EMB	0.467 9	0.448 9	0.467 2
ML-KNN+EMB	0.435 1	0.385 1	0.401 9
BR-LIN+CNN	0.533 9	0.501 8	0.513 6
BR-RBF+CNN	0.545 2	0.517 8	0.528 5
ML-KNN+CNN	0.556 6**	0.525 9**	0.534 9**

表1中,“\*”表示采用相同特征表示方法时的最优性能,“\*\*”表示整体的最优性能。由于BR-NB分类器以文本属性的计数作为特征取值,仅适用于VSM,因而没有对EMB和CNN进行实验。

可以明显看出,相比传统的VSM,使用EMB和CNN作为句子的特征表示方法,从整体上极大地提高了多标签分类性能,特别是CNN句子向量将3个性能指标提升了10%以上。对于多标签分类器,不同的特征表示方法具有不同的影响:BR-LIN相比BR-RBF,在简单的特征表示方法VSM和EMB中更加有效;BR-RBF在情感语义更加复杂的CNN句子向量中超过BR-LIN,表明这种特征表示方法的分类边界较为复杂;ML-KNN在3种特征空间中的分类性能差别巨大,表明这种惰性学习分类器相比其他两种分类器更加依赖于特征表示方法。进一步考察one\_acc还可以发现,VSM特征有利于主要情感的正确预测,CNN特征则同时有利于主次情感的正确预测。

#### 4.2 最优性能的对比分析

为更好地进行方法比较,摘录2013年NLPCC微博情感评测任务的最佳评测结果,以及在相同实验条件下(文献[6]额外使用了整条微博的情感标签)取得最优性能的文献[4]和[5]的结果,如表2所示。2013年NLPCC的2个最优评测结果,close表示仅使用了给定的训练集和测试集,open表示还使用了额外的数据集。本文的EMB和CNN特征表示方法,以及文

献[4]和[5]都使用了未标注的微博文本进行词向量的预学习。

表2 最优性能汇总

Tab.2 Collection of best performance

方法	<i>looseAP</i>	<i>strictAP</i>
NLPCC2013最佳评测结果(close)	0.343 9	0.330 5
NLPCC2013最佳评测结果(open)	0.365 0	0.348 4
文献[4]的最优性能	0.484 0	0.468 0
文献[5]的最优性能	0.520 0	0.497 0

对比close的最佳评测结果,在VSM特征空间中除了ML-KNN之外,其他3个BR分类器都明显超过此结果,最大的差距达到6.33%和4.24%。值得注意的是,这3个BR分类器的结果相差不大。这表明在VSM特征空间中,BR方法比较有效,且基础分类器的影响比较小;ML-KNN的性能低下,表明此空间中不同情感标签的样本点分布不规律(4.3节)。

文献[4]中的最优性能同样采用了词向量和ML-KNN,与本文不同之处有:1)以词语/词性为基本语言单元学习和使用词向量;2)使用词性选择句中的部分词语,采用线性加和运算合成句子向量时考虑否定词的作用;3)使用句子向量的随机线性加和构建新的训练样本,采取过采样策略构造平衡训练集。本文的EMB特征表示方法对句子中的所有词向量仅采用简单的线性平均及最大、最小采样的传统做法,却在使用BR-LIN分类器时取得超过文献[4]最优性能的分类结果。表明采用不同线性运算的词向量合成方法,构建出句子向量之间的差别不大,选择合适的多标签分类器也是提高分类性能的重要途径。当然,考察数据集不平衡性对于这些方法的影响,也值得进一步研究。

文献[5]采用RNTN模型合成句子向量,需要首先完成微博句子的句法分析,且假定句法树的内部结点具有与整个句子一样的情感标签。然而,微博文本很不规范,导致微博句子的句法树不够准确,这必将影响语义合成过程。再者,RNTN模型要求句法树内部的每个结点都具有情感标签,才能更好地指导不同结点处的语义合成,同时减弱反向传播时的误差消失问题。文献[5]并没有人工标注这些结点,而是采取简单的假定策略,将进一步降低RNTN模型的性能。本文采用的CNN模型,完全没有这两个方面的问题,能够直接对微博句子的词向量序列进行卷积操作以完成语义合成运算。此外,文献[5]的最优性能同样需要采用与文献[4]相同的过采样技术,以构建平衡训练集。本文采用CNN句子向量的3个分类器,在原始训练集上都取得超过文献[5]最优性能的分类结果,特别是ML-KNN分类性能分别超过了3.66%和

2.89%,体现出CNN模型相比RNTN模型在合成句子向量时具有明显的优势。最终,本文最优性能超过open的最佳评测结果达到19.16%和17.75%。

### 4.3 最优的超参数取值

除用于VSM空间的BR-NB分类器的最优超参数 $\alpha$ 取值为2.0之外,与表1实验结果对应的最优超参数取值如表3所示,其中BR-RBF的超参数 $\gamma$ 在表中的取值表示对应特征空间维度之倒数的倍数。

表3 最优超参数取值

Tab.3 Optimum values of hyper-parameters

分类器	BR-LIN	BR-RBF		ML-KNN
	$C$	$C$	$\gamma$	$k$
VSM	0.10	10.0	100.0X	45
EMB	0.01	1.0	10.0X	43
CNN	0.10	10.0	1.0X	19

BR-LIN的超参数 $C$ ,用以对SVM的结构风险和经验风险进行折中,其取值越小,决策边界越平滑,进而提高SVM的泛化能力。可以看到,空间维度相近的EMB和CNN,后者的超参数 $C$ 取值大于前者,表明CNN空间中不同类别样本点之间的区分度大于EMB空间,因而不需要通过此参数来提升。尽管VSM和CNN的超参数 $C$ 取值相同,然而前者的空间维度是后者的50倍以上,显然CNN空间的区分度更加优异。

BR-RBF中的超参数 $C$ 具有相同效用,且最优化取值也一致,这里不再赘述。BR-RBF特有的超参数 $\gamma$ ,反映出支持向量样本点的影响力范围大小,取值越小则影响范围越大。在VSM、EMB和CNN空间中, $\gamma$ 的最优取值依次减小,反映出同类别样本点的聚集度依次增高。

ML-KNN的超参数 $k$ ,用以控制决策时近邻参考样本点的个数,其取值越小,特征空间的局部判别性更强。在VSM、EMB和CNN空间中,超参数 $k$ 的最优取值依次减小,反映出局部判别性不断提高。这与前面超参数 $C$ 和 $\gamma$ 的分析结果一致:不同类别样本点之间的区分度越大,分类边界上不同类别样本点的出现几率越小;相同类别样本点之间的聚集度越高,类别内部同类别样本点的出现几率更大。

总的来看,最优的超参数取值进一步地揭示出3种特征空间的不同,体现了CNN空间的优秀特性。

### 4.4 迭代过程的性能分析

采用随机梯度下降法训练CNN模型,迭代过程中会产生一些候选模型。在这些候选模型上测试3个多标签分类器的性能表现,将有利于进一步观察CNN模型进行词向量语义合成的特点。

使用单标签评价指标 $acc$ 对比CNN模型和3个多标签分类器的不同,如图3所示。第1次迭代完成时,

尽管CNN模型仅仅遍历了约1/5的训练集,且准确率只有27.56%,但是其他3个分类器的准确率已在37%以上,反映此时CNN模型已经可以很好地合成句子向量。在第4次迭代完成时,CNN模型才将整个训练集遍历完毕,准确率也达到40%以上,其他3个分类器的性能仍然远远领先,且BR-LIN始终最好。第9次迭代完成时,4个分类器的准确率基本都为51%,之后的情况开始有所变化:CNN继续缓慢上升,直到最后一次迭代稍有下降;BR-LIN和BR-RBF开始基本保持不变,稍后前者开始逐步下降,而后者开始逐步提升,最后一次迭代大幅下降;ML-KNN在快速提升后始终保持最好,并同样在最后一有所下降。

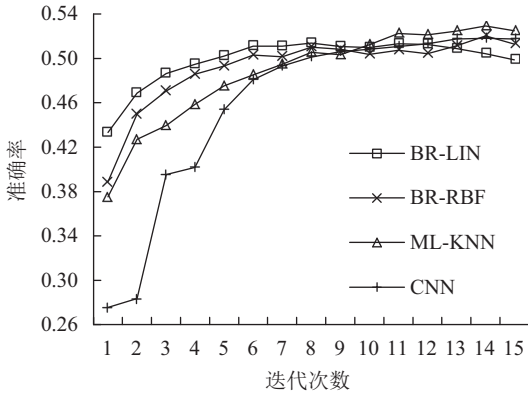


图 3 迭代过程的准确率对比

Fig.3 Accuracy of iterations

迭代过程中多标签分类指标 $looseAP$ 和 $strictAP$ 的变化如图4所示,前缀“lo”代表 $looseAP$ ,”st”代表 $strictAP$ 。可以看出,这两个指标的变化趋势与 $acc$ 类似。不同之处有:从整体看, $looseAP$ 相对 $strictAP$ 的曲线更平滑一些;对于BR-LIN, $looseAP$ 相比 $strictAP$ 在最后阶段的下降明显少许多;对于ML-KNN, $looseAP$ 相比 $strictAP$ 在中前期的提升明显多许多。

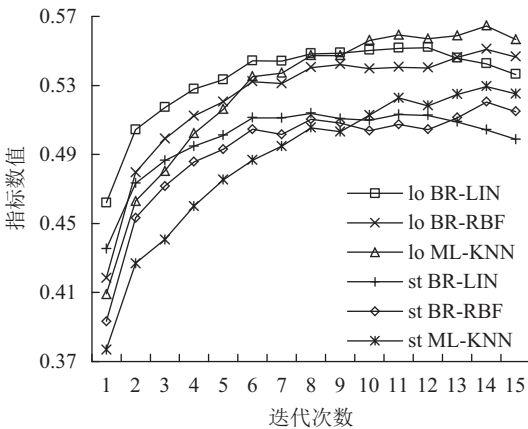


图 4 迭代过程的 $looseAP$ 和 $strictAP$ 对比

Fig.4 LooseAP and strictAP of iterations

总的来看,CNN模型在训练过程中不断发现不同情感标签的语义模式,使得合成的句子向量更加

具有区分度。第1次迭代后的句子向量就使得3个分类器取得不错的性能,表明CNN模型能够从小规模的训练样本中高效地发现具有普遍性的语义合成模式,对于分类边界简单的BR-LIN比较有利。随着更多的训练样本被加入,具有普遍意义的模式更加明确,因而3种分类器的性能都在不断提升。同时,具有局部区分度的模式开始不断增加,导致句子向量间的分类边界不断复杂,使得BR-LIN的线性边界很难适应而导致性能难以提升;对于具有非线性分类边界的BR-RBF和具有边界自适应能力的ML-KNN来说,还可以进一步提升。当这些局部模式积累到一定程度,就容易引起模型的过拟合而导致性能下降。复杂度较低的BR-LIN最先受到影响,性能持续下降。复杂度较高的BR-RBF和ML-KNN,则在最后一次迭代出现性能下降。

## 5 结论

为提高微博句子的多标签情感分类性能,围绕CNN模型构建了一个多标签情感分类系统。此系统利用CNN模型进行词向量的语义合成操作,由此构建的句子向量具有很好的情感语义区分度,能够明显提高多个多标签分类器的分类性能,文中实验的最好结果全部超过微博情感评测任务的最佳评测结果及已知文献的最优性能。通过对迭代训练过程中的CNN候选模型进行性能分析,揭示出语义合成过程中的模式识别规律,为进一步的研究工作指明方向。

本文提出的多标签情感分类系统具有通用性,可以考虑引入更多具有语义合成能力的深度学习模型,进一步探索词向量的语义合成现象。此后,即可进一步尝试将这些深度学习模型改进为多标签分类器,直接进行微博的多标签情感分类。

## 参考文献:

- [1] Zhao Yanyan, Qin Bing, Liu Ting. Sentiment analysis[J]. Journal of Software, 2010, 21(8): 1834-1848. [赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848.]
- [2] Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining[C]//Proceedings of the International Conference on Language Resources and Evaluation. Valletta, Malta: European Language Resources Association, 2010: 1320-1326.
- [3] Kim S, Hovy E. Automatic identification of pro and con reasons in online reviews[C]//Proceedings of the COLING/ACL on Main Conference Poster Sessions. Sydney: ACL, 2006: 483-490.
- [4] Chen T, Xu R, Lu Q, et al. A sentence vector based over-sampling method for imbalanced emotion classification[C]//

- Proceedings of 15th International Conference on Computational Linguistics and Intelligent Text Processing. Kathmandu: Springer, 2014: 62–72.
- [5] Xu R, Chen T, Xia Y, et al. Word embedding composition for data imbalances in sentiment and emotion classification[J]. *Cognitive Computation*, 2015, 7(2): 226–240.
- [6] Xu R, Wang Z, Xu J, et al. An iterative emotion classification approach for microblogs[C]//Proceedings of 16th International Conference on Computational Linguistics and Intelligent Text Processing. Cairo: Springer, 2015: 104–113.
- [7] Cambria E, Schuller B, Xia Y, et al. New avenues in opinion mining and sentiment analysis[J]. *IEEE Intelligent Systems*, 2013, 28(2): 15–21.
- [8] Salton G, Wong A, Yang C. A vector space model for automatic indexing[J]. *Communications of the ACM*, 1975, 18(11): 613–620.
- [9] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of Advances in Neural Information Processing Systems 26. Lake Tahoe: Curran Associates, Inc, 2013: 3111–3119.
- [10] Mitchell J, Lapata M. Composition in distributional models of semantics[J]. *Cognitive Science*, 2010, 34(8): 1388–1429.
- [11] Socher R, Huval B, Manning C D, et al. Semantic compositionality through recursive matrix-vector spaces[C]//Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island: Association for Computational Linguistics, 2012: 1201–1211.
- [12] Socher R, Perelygin A, Wu Y, et al. Recursive deep models for semantic compositionality over a sentiment treebank[C]//Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle: Association for Computational Linguistics, 2013: 1631–1642.
- [13] Tai K S, Socher R, Manning C D. Improved semantic representations from tree-structured long short-term memory networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing. Beijing: Association for Computational Linguistics, 2015: 1556–1566.
- [14] Kim Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: Association for Computational Linguistics, 2014: 1746–1751.
- [15] dos Santos C N, Gatti M. Deep convolutional neural networks for sentiment analysis of short texts[C]//Proceedings of the 25th International Conference on Computational Linguistics. Dublin: ACL, 2014: 69–78.
- [16] Zhang M, Zhou Z. A review on multi-label learning algorithms[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 26(8): 1819–1837.
- [17] Boutell M R, Luo J, Shen X, et al. Learning multi-label scene classification[J]. *Pattern Recognition*, 2004, 37(9): 1757–1771.
- [18] Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification[J]. *Machine Learning*, 2011, 85(3): 333–359.
- [19] Tsoumakas G, Vlahavas I. Random k-labelsets: An ensemble method for multilabel classification[C]//Proceedings of 18th European Conference on Machine Learning. Warsaw: Springer Berlin Heidelberg, 2007: 406–417.
- [20] Fürnkranz J, Hüllermeier E, Mencia E L, et al. Multilabel classification via calibrated label ranking[J]. *Machine Learning*, 2008, 73(2): 133–153.
- [21] Zhang M, Zhou Z. ML-KNN: A lazy learning approach to multi-label learning[J]. *Pattern Recognition*, 2007, 40(7): 2038–2048.
- [22] Ghamrawi N, McCallum A. Collective multi-label classification[C]//Proceedings of 14th ACM International Conference on Information and Knowledge Management. Bremen: ACM, 2005: 195–200.
- [23] Elisseeff A, Weston J. A kernel method for multi-labelled classification[C]//Proceedings of Advances in Neural Information Processing Systems 14. Vancouver: MIT Press, 2001: 681–687.
- [24] Clare A, King R D. Knowledge discovery in multi-label phenotype data[C]//Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery. Freiburg: Springer Berlin Heidelberg, 2001: 42–53.
- [25] Birmingham A, Smeaton A F. Classifying sentiment in microblogs: Is brevity an advantage?[C]//The 19th ACM International Conference on Information and Knowledge Management. Toronto: ACM, 2010: 1833–1836.
- [26] Burges C J A. A tutorial on support vector machines for pattern recognition[J]. *Data Mining and Knowledge Discovery*, 1998, 2(2): 121–167.

(编辑 李轶楠)

引用格式: Sun Songtao, He Yanxiang. Multi-label emotion classification for microblog based on CNN feature space[J]. *Advanced Engineering Sciences*, 2017, 49(3): 162–169. [孙松涛, 何炎祥. 基于CNN特征空间的微博多标签情感分类[J]. *工程科学与技术*, 2017, 49(3): 162–169.]