计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.181102130

采用改进粒子群优化的SVM方法实现中文文本情感分类

王立志[1]　慕晓冬[1]　刘宏岚[2]
1 火箭军工程大学信息工程系　西安 710025
2 北京科技大学计算机科学与通信工程学院　北京 100083
(1546600487@qq.com)

摘　要　近年来,随着网络用户量的不断增加,用户评论数量也呈爆炸式增长,伴随而来的是大量可用于参考和深度挖掘的信息,文本情感分类应运而生。分类模型的预测精度和执行速度是衡量模型优劣的关键。使用传统的SVM进行文本情感分类,算法简单,易于实现,但其模型参数决定了分类准确率。针对这种情况,文中将改进粒子群优化算法与SVM分类方法相结合,采用了改进粒子群算法优化的SVM方法对影视剧评论的情感进行了研究分析。首先,通过网络爬虫获取豆瓣电影评论数据,将数据预处理后利用加权word2vec向量化文本信息,将其作为支持向量机可识别的输入;然后,使用自适应惯性递减策略并引入交叉算子来改进粒子群算法,并对SVM模型的损失函数、惩罚参数及核函数的参数进行优化;最后,实现文本的情感分类。在同一数据集上的实验结果表明,所提方法有效规避了传统的情感词典方法受词语顺序和不同语境影响的缺陷及使用卷积出现梯度消失或弥散的问题,同时也克服了粒子群算法易陷入局部最优的不足。相较于其他方法,所提分类模型的执行速度更快,有效地提高了分类准确率。

关键词:情感分析;网络爬虫;SVM分类;惯性递减;粒子群优化
中图法分类号　TP391

Using SVM Method Optimized by Improved Particle Swarm Optimization to Analyze Emotion of Chinese Text

WANG Li-zhi[1], MU Xiao-dong[1] and LIU Hong-lan[2]
1 Department of Information Engineering, Rocket Force University of Engineering, Xi'an 710025, China
2 School of Computer & Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

Abstract　In recent years, with the increasing number of network users, the number of user comments has also increased explosively, accompanied by a large number of information that can be used for reference and deep excavation. Text sentiment classification arises at this historic moment, the prediction accuracy and the execution speed of classification model are the keys to measure the quality of the model. Traditional algorithm by using SVM for text sentiment classification is simple and easy to implement, and its model parameters determine the classification accuracy. In this case, this paper combined the improved particle swarm optimization algorithm with the SVM classification method, used the SVM method optimized by improved particle swarm optimization to analyze the emotion of the movie and TV drama review. Firstly, Douban movie review data are obtained by internet crawler. Then the text information is vectorized by weighted word2vec after pre-processing, which becomes the recognizable input of support vector machine. Adaptive inertia decreasing strategy and crossover operator are used to improve particle swarm optimization algorithm. The loss function, penalty parameter and kernel parameter of SVM model are optimized by improved PSO. Finally, the text is classified by this model. Experimental results on the same data show that this method effectively avoids the shortcomings of traditional affective dictionary method affected by word order and different contexts, and solves the problem of gradient disappearance or dispersion caused by convolution. It also overcomes the possibility that PSO itself is easily trapped in local optimum. Compared with other methods, the proposed classification model performs faster and improves classification accuracy effectively.

Keywords　Sentiment analysis, Internet worm, SVM classification, Inertia diminishing, Particle swarm optimization

自然处理[1](Natural Language Processing, NLP),指以　计算机为工具,对人类独有的自然语言进行多种加工和处理

到稿日期:2018-11-09　返修日期:2019-05-01　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61702525)
This work was supported by the National Natural Science Foundation of China (61702525).
通信作者:慕晓冬(wascom4@sina.com)

的技术，是人工智能领域和计算机科学领域的一项重要内容。情感分析是自然语言处理的方向之一，也被称为情感提取或意见挖掘，是文本挖掘领域的研究热点[2]。文本的信息挖掘可以运用于用户反馈的评论信息、智能分析后的广告推荐、政府部门的舆情检测以及不文明不真实信息的处理等多个方面。

Dave 等[3]设计了最早的情感分析工具。Go 等[4]借助 Hashtag 设计的训练集数据，在以主题为基础的聚类上了多个分类。Joshi 等[5]设计了一套感情分析体系，依托微博的某类特点，如表情符，把微博评论分为正负两类情感。Gamon 等[6]利用聚类的功能获得用户观点，分析了对于汽车的情感评价倾向和强度。Li 等[7]使用 SVM、贝叶斯分类及 n 元语言方法，利用特征选择和提取完成了情感分析。李勇敢等[8]构建了无监督主题情感模型以实现情感分类。Jasson 等[9]将 CNN 应用于文本，借助文本数据的一维结构(即词序)进行了准确预测。Yoom Kim 使用卷积实现了英文文本句子级别的分类模型。Xue 等[10]提出了基于卷积神经网络和门控机制的分类模型。Parupalli 等[11]构建了一个有系统注释的语库，支持使用词语级注释来增强情感分析任务。Angelidis 等[12]提出了基于注意力的正负文本片段极性评分方法。Gui 等[13]在情感分类的基础上，提出通过对上下文信息进行进一步建模，来抽取文本情感原因。Yuan 等[14]运用全局解码器前馈网络实现对多语言文本的识别，为复杂文本分析奠定了基础。Bordoloi 等[15]设计了一种有效的情感分析模型，该模型基于图的关键词抽取方法对电子商务站点收集的手机评论进行高级分析。目前，卷积神经网络(Convolutional Neural Networks，CNN)在自然语言处理领域得到了广泛应用，但卷积神经网络包含多个卷积层和池化层，需要更多的参数，且参数优化调整代价较大。同时，卷积神经网络存在梯度消失或梯度爆炸问题，限制了文本分类的准确率。

利用支持向量机进行正负情感分类有较好的执行效率，但支持向量机模型的参数是决定预测精度的关键因素。为此，研究者引入粒子群算法对参数进行优化，以期获得较高的准确率。粒子群优化算法(Particle Swarm Optimization，PSO)是智能优化算法中的重要分支[16]，是 Kennedy[17]和 Eberhart 于 1995 年首先提出的。标准 PSO 算法在解决多种非线性优化问题上具有良好的性能，其借助个体最优和群体最优来控制整个迭代过程，收敛速度快，执行效率高。但是在迭代后期，种群中粒子个体的多样性较小，如果全局最优和局部最优的位置与粒子的位置在一定的迭代次数后相等，则算法可能陷入局部最优，致使全局性能较差。为提高粒子群算法的性能，1988 年 Shi 等[18]引入的惯性权重对 PSO 性能的改进起到了关键性作用。本文结合自适应惯性递减策略及交叉算子改善了粒子群算法的寻优质量，并借助改进的算法优化 SVM 模型的参数，进一步提高了预测精度。与其他情感分析模型相比，本文模型训练速度更快且具有良好的预测精度。

1 文本获取及预处理

在分类前，需要进行数据获取和预处理。

1.1 评论数据获取

数据的获取可以分为 URL 队列获得、相关网页解析、数据爬取、数据清洗及数据存储 5 个部分。

首先，从豆瓣影评中爬取影评数据。豆瓣影评是五星评价机制，一星到五星分别是：很差、较差、还行、推荐、力荐；本实验将一星、二星评论作为负向评论，四星、五星评论作为正向评论。然后，基于 Scrapy 框架实现数据抓取，生成相应的正负情感评论 csv 文件。最后，通过进一步筛选，获得正向评论21000 条，负向评论 9000 条，取其中的 2/3 作为训练数据，1/3 作为测试数据。实验数据分布如表 1 所列。

表 1 实验语料的数据分布

Table 1 Data distribution of experimental data

	评论数量	正面评价	负面评价
训练数据	20000	14000	6000
测试数据	10000	7000	3000

为获取足够规模的数据集，爬取过程中要实现模拟登录，突破网页对爬虫的限制。同时，某条评论可能在当前页面显示不完整，需要跳转到相应页面以获取全部评论。Scrapy 有默认去重机制，会判定第二次是重复爬取，因而需要解决 URL 重复问题。

1.2 文字预处理

文字预处理即对原始数据进行进一步处理，使数据成为下一步可以操作的对象。文字预处理分为文字去重、去噪、分词及去停用词 4 个部分。

(1)文字去重。预处理的文本中存在重复的文字，造成存储冗余，同时也增加了计算量，因此需要遍历去重。

(2)文字去噪。文本中出现一些干扰性文字或乱码，需要进行去噪，以降低计算复杂度，提高分析准确率。

(3)分词。将文本中的词语分割，结合词典赋予它们相应的词性。英文中每个单词都用空格隔开，处理起来比较容易，而中文分词对词典的全面准确性要求较高，因此本文采用 python 语言环境下的 jieba 分词。

(4)去停用词。停用词指文章中存在的语气助词、副词、介词、连词等，对实验结果没有帮助的词都可以归纳到停用词表中并去除。

预处理后的实验效果如表 2 所列。

表 2 文字处理效果

Table 2 Word processing effect

数据处理	执行过程	执行结果
未处理原文	不处理	【水军勿进】这部剧不仅好看，演员也都演技很好
分词处理	仅中文分词	【/w 水军/n 勿/d 进/v 】/w 这部/r 剧/n 不/d 仅/d 好看/a，/w 演员/n 也/d 都/d 演技/n 很/d 好/a
去词性标注、去停用词、去噪	利用正则化，使分词后输出的语句不带词性标注，加载停用词表和噪声词表去重去噪	【水军 勿 进】这部 剧 仅 好看 演员 演技

1.3 词向量化

文本是非结构化或半结构化的数据，SVM 分类器不能对其进行识别，因此需要将文本转化为向量形式，以便进一步的分析和处理。词向量化即将词语表示成向量形式，同时需要

保证处理的向量在语义相似度和相对相似度方面的相关性。词向量化能把词或短语映射为实数向量，把较高维度的向量空间特征降低到相对较低的维度空间。将词转化为实数性向量的模型有很多，如隐含狄利克雷分布(Latent Dirichlet Allocation，LDA)、隐含语义分析(Latent Semantic Analysis，LSA)。但以上模型的计算量会随着数据总量的增加而急剧增大，而 word2vec 较好地解决了这个问题，提高了效率。

word2vec 是 Google 于 2013 年开源的一个深度学习的多层神经网络结构，经过训练能够把对文本内容的处理简化成 K 维向量运算。其主体结构由输入层、多个隐藏层和输出层 3 部分构成。它将全部的特征词经过匹配转化为向量值，从而给予文本数据更深层的特征表示。基于此，本文使用 word2ve 实现词语向量化。假设预处理后的文本评论由 N 个词语组成，表示为 $d_N=\langle w_1,w_2,\cdots,w_N\rangle$。首先使用 word2vec 中默认的 Skip_gram 模型来训练搜狗新闻语料库；接着使用训练后的模型来计算 d_N 中每个词汇的 word2vec 向量，同时考虑到 word2vec 只解决了词间的语义关系，忽略了词语的重要程度，依据影评数据训练 TFIDF 模型，获得影评中每个词语的 TFIDF 权重，将其乘以相应的 word2vec 向量；最后获得支持向量机可识别的输入，每条影评的最终向量为 D。

$$D=(\sum_{i=1}^{N} word2vec(w_i)\times tfidf_{w_i})/N \tag{1}$$

2 SVM 分类

支持向量机能够构建一个超平面，使得正负两极在决策面之间的距离最大。其主要思想如图 1 所示。

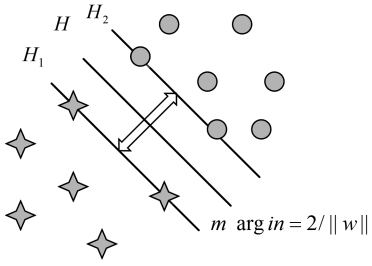

图 1 SVM 分类模型

Fig. 1 SVM classification model

图 1 中，四角星和圆形分别代表两类样本，SVM 使得这两类样本的直线距离最大。假设有 $X_1=\{x_i,i\in I_1\}$，$X_2=\{x_i,i\in I_2\}$ 两个训练集，X_1 中标签是+1，X_2 中标签是−1。

$$m=|I_1|,k=|I_2|,l=m+k,X=X_1\cup X_2 \tag{2}$$

当 X_1 和 X_2 线性可分时，SVM 通过构建一个线性分类超平面无差错地将这两类样本点分开，分类的超平面是：

$$H:\langle\omega\cdot\chi\rangle=\lambda,X_2=\{x_i,i\in I_2\} \tag{3}$$

SVM 是在线性可分的情况下提出最优平面分类，要求模型既能无差错区分数据，还能使得分类空隙达到最大。线性判别函数在多维空间一般表示为：

$$g(x)=wx+b \tag{4}$$

分类面的方程为：$wx+b=0$，归一化处理判别函数，使得这两类样点本到最优平面的距离大于等于 1。

$$y_i(w\cdot x_i+b)\geqslant 1,i=1,2,3,\cdots,n \tag{5}$$

此时离分类面最近的样本满足 $|g(x)|=1$，使得式(5)成立的样本即为支持向量。分类间隔的大小是 $2/\|w\|$，$\|w\|^2/2$ 最小即等价于间隔最大，满足 $\phi(w)=\|w\|^2/2$，同时满足式(5)的分类面即为最优分类面。但样本存在不能线性可分的情况，即 SVM 算法无法运行出可解的方案。对此，引入松弛变量的集合 $\xi=(\xi_1,\xi_2,\cdots,\xi_n)$：

$$\xi_i^*=\begin{cases}0, & |f(x)-y_i|<\varepsilon\\ |f(x)-y_i|-\varepsilon, & |f(x)-y_i|>\varepsilon\end{cases} \tag{6}$$

损失函数 ε、惩罚参数 C 和松弛变量 ξ 的引入，可将非线性问题转化为线性问题，实现样本分离的同时使错误率最小，具体公式为：

$$\min_{\omega,b,\xi}\Phi=\frac{1}{2}\omega^{\mathrm{T}}\cdot\omega+C\sum_{i=1}^{n}(\xi_i+\xi_i^*) \tag{7}$$

$$\text{s.t. } (w\cdot x_i+b)-y_i\leqslant\varepsilon+\xi_i$$

$$y_i-(w\cdot x_i+b)\leqslant\varepsilon+\xi_i^*$$

求解该问题的对偶问题，得到拉格朗日因子 a_i 和 a_i^*，回归方程的系数为：

$$w=\sum_{i=1}^{l}(a_i-a_i^*)x_i \tag{8}$$

使用核技巧，引用高斯径向基函数：

$$k(x_i,x)=N(x_i-x;0,\sigma^2 I) \tag{9}$$

其中，$N(x;\mu,\sum)$是标准正态分布。将点积替换为核估计，判别函数的表达式为：

$$f(x)=\sum_{i=1}^{l}(a_i-a_i^*)K(x_i,x)+b \tag{10}$$

损失函数 ε、惩罚参数 C 以及核函数的参数 σ 决定支持向量机的性能。损失函数 ε 是估计函数在误差上的期望，在一定程度上影响支持向量的个数；惩罚参数 C 过小会导致欠学习，过大会导致过学习；σ 是核函数的参数，反映了训练集的特性，决定了解的复杂程度。因此选择好的参数对分类效率和准确率的影响较大，本文使用改进粒子群算法实现了参数优化。

3 改进粒子群算法 PSO-W-GA

PSO 算法源于对鸟类觅食行为的研究[19]，此算法首先随机初始化一群粒子，每个粒子都是优化问题的一个可行解，并且根据目标函数确定适应值。粒子朝着当前最优粒子的方向运动，经逐代搜索获得最优解，每一代种群中都会有两个极值，一个是粒子本身找到的最优解 p_{best}，一个是全种群找到的最优解 g_{best}，每个粒子根据这两个极值不断更新，产生新一代群体。

找到这两个极值后，粒子依据下式更新自己的速度和位置：

$$v'_{id}=\omega V_{id}+c_1 rand()(P_{id}-X_{id})+c_2 rand()(P_{gd}-X_{id})id \tag{11}$$

$$X'_{id}=X_{id}+V'_{id} \tag{12}$$

其中，V_{id} 表示第 i 个粒子在第 d 维上的速度；ω 为惯性权重，是非负数；c_1 和 c_2 是非负常数。研究表明，加速度系数应该满足 $c_1+c_2<4(1+w)$[20]。Martniez 等[21]提出 $c_1=c_2$ 能使

二阶稳定域最大化，此外加速度系数相等能给所有最优值(全局和局部)赋予相同权重，以避免算法在开始阶段就陷入局部最优。P_{id} 和 P_{gd} 是相应维度的个体最优值和全局最优值。rand()生成(0,1)之间的随机数。

3.1 惯性权重的改进

惯性权重的调整主要分为4类：常数[22]、随机数[23]、时变及自适应惯性权重。初始化阶段，步长较大，需要较大的惯性权重 ω；中后期要求具有较强的局部开发能力，需要较小的 ω。为了提高寻优效率，避免陷入局部最优，本文建构了一种惯性递减策略，在迭代过程中逐步减小了 ω 的值，并引入了有 K 个元素的向量 $h(K$ 是常数)，得到如下算法：

$$h(k)=\max_{1\leqslant j\leqslant D}\{std(P_{gd}(t)-X_{id}(t))\},1\leqslant k\leqslant K$$
$$\omega(t)=\omega_{\text{start}}-\omega_{\text{end}}\frac{h(k)}{\max_{1\leqslant k\leqslant K}\{h(k)\}} \tag{13}$$

其中，k 是当前迭代次数 t 关于 K 的模；$\omega(t)$，ω_{start} 和 ω_{end} 分别为当前、起始和终止惯性权重。当粒子的最佳位置相互靠近时，$\omega(t)$就增加反向趋势，防止过早收敛，这样粒子就会进行更多的探索。每迭代 K 次，权重会逐渐减小，从而增强了粒子的局部寻优能力，最终有效地提高了寻优能力。

3.2 引入交叉算子

为解决算法迭代后期陷入局部最优的可能，引入交叉算子来加强粒子之间的信息交换。搜索过程由个体最优、群体最优以及个体的遗传操作共同控制，以弥补后期容易陷入局部最优的缺陷，从而使算法跳出局部最优，获得全局最优解。

3.3 仿真实验及分析

为测试 PSO-W-GA 算法的效果，选用 CEC2014 中的 Sphere，Schewefel，Rastrigin，Rosenbrock 4个基准函数来评价算法的性能，并将所提方法与以下改进方法进行比较：1)标准 PSO；2)APSO，其中 $\omega=\dfrac{1}{1+1.5\exp(-2.6f)}\in[0.4,0.5]$，$f$ 是利用粒子间距离计算的进化因子；3)AIWPSO，其中 $\omega=S(t)/N\in[0,1]$，N 是种群大小，$S(t)$是种群在 t 时间最好的位置。

依据经验，ω 取值范围为[0.4,0.9][24]，本文取 $\omega_{\text{start}}=0.9$，$\omega_{\text{end}}=0.4$，$c_1=c_2=1.5$，群体规模大小为30，最大迭代次数为1000，K 值为100，粒子维度为30。为评估算法的结果，将计算函数值的最大次数设置为 FE_S[35]：

$$FE_S=N\times T=N\times 10000\times\frac{D}{N} \tag{14}$$

各算法的运行结果如表3所列。

表3 各算法的运行结果($D=30$)

Table 3 Running results of each algorithm($D=30$)

f	指标	PSO	AIWPSO	APSO	W-G-PSO
f_1	Mean	1.3×10^{-1}	8.7×10^{-2}	2.4×10^{-2}	1.8×10^{-2}
	SD	9.1×10^{-2}	1.7×10^{-3}	3.2×10^{-3}	1.5×10^{-3}
f_2	Mean	9.4×10^{-2}	3.7×10^{-2}	2.8×10^{-2}	2.3×10^{-2}
	SD	6.3×10^{-2}	4.7×10^{-3}	4.4×10^{-3}	1.4×10^{-3}
f_3	Mean	7.5	3.8	6.7×10^{-1}	2.4×10^{-1}
	SD	4.7×10^{-1}	1.5×10^{-1}	4.3×10^{-2}	7.8×10^{-2}
f_4	Mean	7.3	6.8	4.4	3.1
	SD	4.9×10^{-1}	3.4×10^{-1}	2.3×10^{-1}	6.5×10^{-2}

表3中，f_1，f_2，f_3，f_4 分别是测试函数 Sphere，Schewefel，Rastrigin 及 Rosenbrock。Mean 和 SD 分别表示运行函数后获得的平均值和标准差。通过表3以及图2—图5可以发现，上述测试函数中，改进算法都能快速收敛到最优值，具有更优异的性能。

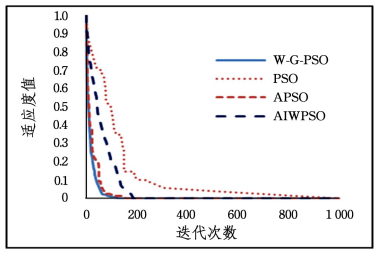

图2 单峰函数 Sphere 的测试结果

Fig. 2 Test results for single peak function Sphere

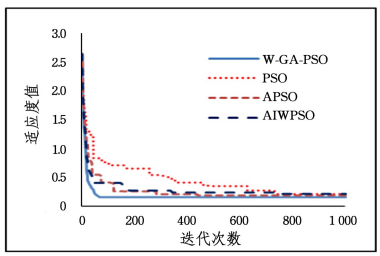

图3 单峰函数 Schewefel 的测试结果

Fig. 3 Test results for single peak function Schewefel

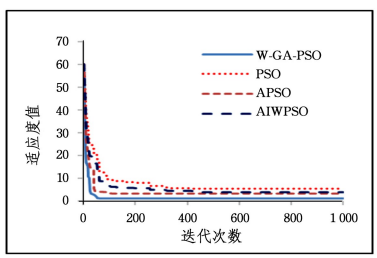

图4 多峰函数 Rastrigin 的测试结果

Fig. 4 Test results for multimodal function Rastrigin

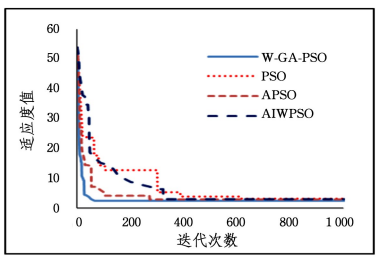

图5 多峰函数 Rosenbrock 测试结果

Fig. 5 Test results for multimodal function Rosenbrock

4 改进粒子群算法下的 SVM 分类

本文结合了粒子群算法全局寻优和 SVM 快速分类的特点，提出改进粒子群算法优化的 SVM 模型，以实现情感分类。具体的实现步骤如下。

(1)从网页中爬取评论信息，对文本进行预处理；

(2)将评论数据分成3个部分：训练集、验证集、测试集，

并进行正负标签设置；

(3)使用 word2vec 中默认的 Skip_gram 模型训练搜狗新闻语料库，再使用训练后的模型计算评论中每个词汇的 word2vec 向量 w_i；

(4)依据本实验的影评数据训练 TFIDF 模型，获得影评中每个词语的 TFIDF 权重 $tfidf_{w_i}$；

(5)通过式(1)获得每条影评的词向量，将其作为支持向量机的输入；

(6)初始化支持向量机的损失函数 ε、惩罚参数 C 及核函数的参数 σ，将向量 (ε, c, σ) 视为粒子，根据经验定义 $c_1 = c_2 = 2$，利用式(10)和式(12)生成 r_1，r_2，ω；

(7)定义适应度函数：

$$F_{\text{fitness}} = \frac{1}{m}\sum_{i=1}^{m}(f_i - y_i)^2 \tag{15}$$

其中，f_i 是预测值，y_i 为实际值，m 为样本个数；

(8)依据适应度函数计算每个粒子的适应值，如果当前适应值小于之前的适应值，则替换原来的 p_{best}，否则保持不变；

(9)取最小的 p_{best} 与 g_{best} 比较，若 p_{best} 小于 g_{best}，则用 p_{best} 取代 g_{best}，否则保留 g_{best}；

(10)引用 GA 算法的 OX 交叉算子产出新的粒子，计算适应值，重复步骤(8)和步骤(9)，更新 g_{best} 和 g_{best}；

(11)判断是否达到最大迭代次数，若达到则继续下一步，否则迭代次数加 1，跳转到步骤(6)；

(12)依据优化后的参数构建 SVM 模型，对加权 word2vec 处理后的特征向量进行分类，输出分类结果。

5 实验及结果分析

情感预测的效果有 4 个标准：准确率、精确率、召回率。预测结果可以分为 4 类：1)TP，将正向预测为正向的数量；2)FN，将正向预测为负向的数量；3)FP，将负向预测为正向的数量；4)TN，将负向预测为负向的数量。

正类准确率为 $P\_pos = TP/(TP+FP)$；正类召回率为 $R\_pos = TP/(TP+FN)$；正类 F_1 值为：

$$F_1 = \frac{2\times P\_pos \times R\_pos}{P\_pos + R\_pos} \tag{16}$$

实验以准确率和 F_1 值作为模型评价标准。选用正向评论 7000 条，负向评论 3000 条，绘制 roc 曲线，将曲线下的面积即 AUC(Area Under Curve)作为衡量 SVM 分类效果的标准。图 6 给出了改进粒子群优化的 SVM 分类的 AUC 值。

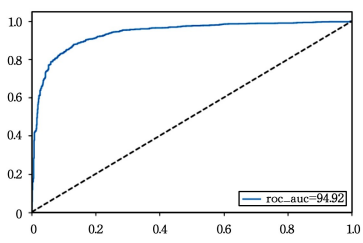

图 6 改进粒子群优化的 SVM 分类的 AUC 值

Fig. 6 AUC value of SVM classification improved by particle swarm optimization

最后，比较不同模型下的分类效果，结果如表 4 所列。实验证明，本文提出的方法有更高的预测准确率和相对较高的运行效率。

表 4 不同模型下的分类效果

Table 4 Classification effect under different models

模型	T/s	Auc/%	F1/%
情感词典	542	81.15	80.13
SVM	225	84.64	84.75
SVM+PSO	267	92.81	92.66
CNN	342	93.84	93.87
SVM+PSO-W-GA	283	94.92	94.82

结束语 本文利用改进的粒子群算法优化 SVM 模型参数进行中文文本情感分类。在 SVM 模型中使用核技巧学习非线性模型，降低了损耗。实验表明，该方法弥补了传统情感词典方法受词语顺序以及不同语境的影响的不足，也解决了粒子群算法易陷入局部最优的问题；同时避免了使用卷积出现梯度消失或弥散的问题，减小了参数优化调整的代价，有更高的运行效率和准确率，能很好地进行文本情感的预测。后续将进一步改进算法以提高预测精度，同时本文的实验数据都是从网上爬取的，因此增加数据规模，验证所提模型在大规模数据下的分类效果很有必要。本模型对于二值分类问题有较好的效果，因此我们将考虑推广模型以解决更复杂的分类问题。

参 考 文 献

[1] 冯志伟. 自然语言处理简明教程[M]. 上海：上海外语教育出版社，2012.

[2] KAUR H, MANGAT V, NIDHI. A survey of sentiment analysis techniques[C]// International Conference on I-Smac. IEEE, Palladam, India, 2017: 921-925.

[3] DAVE, KUSHAL, LAWRENCE, et al. Mining the peanut gallery: opinion extraction and semantic classification of product reviews[C]// Proceedings of the 12th International Conference on World Wide Web. NewYork: ACM, 2003.

[4] GO A, BHAYANI R, HUANG L. Twitter sentiment classification using distant supervision[J]. Processing, 2009, 150(12).

[5] JOSHI A, BALAMURALI A R, BHATTACHARYYA P, et al. C-Feel-It: A Sentiment Analyzer for Micro-blogs[C]// International Conference on Networked Computing & Advanced Information Management. IEEE Computer Society, 2008: 220-225.

[6] GAMON M, AUE A, CORSTON-OLIVER S, et al. Pulse: mining customer opinions from free text[C]// International Symposium on Intelligent Data Analysis. Berlin: Springer-Verlag, 2005: 121-132.

[7] LI S S, HUANG C R, ZHOU G D, et al. Employing personal/impersonal views in supervised and semi-supervised sentiment classification[C]// Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala: ACL, 2010.

[8] LI Y G, ZHOU X G, SUN Y, et al. Research and Implementation of Chinese Microblog Sentiment Classification[J]. Journal of Software, 2017, 28(12): 3183-3205.

[9] JOHNSON R, ZHANG T. Effective Use of Word Order for

Text Categorization with Convolutional Neural Networks[J]. arXiv:1412.1058.

[10] XUE W, LI T. Aspect Based Sentiment Analysis with Gated Convolutional Networks[C]// Association for Computational Linguistics. Melbourne, Australia, 2018:2514-2523.

[11] PARUPALLI S, RAO V A, MAMIDI R. BCSAT: A Benchmark Corpus for Sentiment Analysis in Telugu Using Word-level Annotations[C]// Association for Computational Linguistics. Melbourne, Australia, 2018:99-104.

[12] ANGELIDIS S, LAPATA M. Multiple Instance Learning Networks for Fine-Grained Sentiment Analysis[C]// TACL: Transactions of the Association for Computational Linguistics. Melbourne, Australia, 2018:17-31.

[13] GUI L, HU J, HE Y, et al. A Question Answering Approach to Emotion Cause Extraction[C]// Empirical Methods in Natural Language Processing. Copenhagen, Denmark, 2017:1593-1602.

[14] YUAN Z, JASON R, DANIEL G, et al. A Fast, Compact, Accurate Model for Language Identification of Codemixed Text [C]// EMNLP: Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018:328-337.

[15] BORDOLOI M, BISWAS S K. Graph-Based Sentiment Analysis Model for E-Commerce Websites' Data[C]// CISC: Cognitive Informatics and Soft Computing. Singapore: Springer, 2019:453-462.

[16] LI R Y, ZHANG W J, ZHOU Z Y. Improved PSO Algorithm and Its Load Distribution Optimization of Hot Strip Mills[J]. Computer Science, 2018, 45(7):214-218, 225.

[17] KENNEDY J. Particle Swarm Optimization[C]// Icnn95-international Conference on Neural Networks. IEEE, 2002.

[18] SHI Y, EBERHART R C. A modified particle swarm optimizer [C]// Proceedings IEEE Congress on Evolutionary Computation (CEC'98). Anchorage, 1998:69-73.

[19] KOU X L. Swarm Intelligence Algorithms and Their Application[D]. Xi'an: Xidian University, 2009.

[20] RAPAIĆ M R, KANOVIĆ Ž. Time-varying PSO-convergence analysis, convergence-related parameterization and new parameter adjustment schemes[J]. Information Processing Letters, 2009, 109(11):548-552.

[21] MARTÍNEZ J L F, GARCÍA E. The PSO family: deduction, stochastic analysis and comparison[J]. Swarm Intelligence, 2009, 3(4):245-273.

[22] SHI Y, EBERHART R C. A modified particle swarm optimizer [C]// Proceedings IEEE Congress on Evolutionary Computation (CEC'98). Anchorage, 1998:69-73.

[23] EBERHART R C, SHI Y. Tracking and optimizing dynamic systems with particle swarms[C]// Congress on Evolutionary Computation. IEEE, 2001.

[24] SHI Y, EBERHART R C. Empirical study of particle swarm optimization[C]// Congress on Evolutionary Computation. Washington: IEEE, 2002.

[25] LIANG J J, QU B Y, SUGANTHAN P N. Problem definitions and evaluation criteria for the CEC 2014 special session and competition on single objective real-parameter numerical optimization[R]. Technical Report 201311, 2013.

WANG Li-zhi, born in 1994, Ph.D. His main research interests include natural language processing and computer vision.

MU Xiao-dong, born in 1965, Ph.D supervisor. His main research interests include intelligent information processing and computer simulation.