

基于支持向量机的网络伪舆情识别研究*

刘 勘 朱怀萍 刘秀芹

(中南财经政法大学信息与安全工程学院 武汉 430073)

【摘要】针对网络伪舆情的识别问题,提出一种基于支持向量机的网络伪舆情识别方法。鉴于不同的舆情信息所反映出的舆情特征不同,而舆情特征的不同又可进一步辨别舆情的真假,因此首先构建针对网络舆情真伪的评价指标;基于支持向量机的分类机理,结合网络舆情的评价指标提出基于支持向量机的网络伪舆情识别模型,采用多项式核函数以及优化之后的径向基核函数产生的分类器。通过实验证明采用支持向量机构造舆情分类器所构建的识别算法能够对网络伪舆情进行有效识别。

【关键词】网络伪舆情 支持向量机 评价指标 核函数

【分类号】G202

Detection of Internet Deceptive Opinion Based on SVM

Liu Kan Zhu Huaiping Liu Xiuqin

(School of Information and Safty Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China)

【Abstract】This paper proposes a new approach to detect internet deceptive opinion based on Support Vector Machine (SVM). According to the different features in different opinions or sentiments, the authors firstly evaluate the effect or importance of various features. Then a deception detection model is developed with SVM. This model adopts polynomial kernel function and RBF kernel function after optimization to generate the classifier. The results of the experiment show that the proposed method is effective in identifying the deceptive opinion.

【Keywords】Internet deceptive opinion SVM Evaluation index Kernel function

1 引言

所谓网络伪舆情,也即网络虚假舆情,是指在网络这一特定环境下,一些有组织的公司或网民个体,通过论坛发帖、博文转载、QQ 消息群发等特定方式,散布“无中生有”或曲解原意的虚假信息和言论,在这些信息和言论的刺激下,媒体或网民出于蓄意制造或本能所表现出的具有明显倾向性、能够引导网络舆论走向、并可能造成强烈负面影响的所有认知、态度、情绪、意见和行为倾向的集合。

网民不仅是信息的接受者,更是信息的发布者和传播者。然而,少数别有用心者利用网络言论的强大影响力,制造网络谣言,推动网上网下群体性事件,甚至会导致社会动荡。因此,对网络上的虚假信息以及由此产生的网络伪舆情进行有效的识别、监控、预警及清除,已经成为当前迫切需要解决的问题。

2 相关研究

目前对网络舆情的研究主要集中于:网络热点话题提取^[1]、文本倾向性分类^[2]以及突发事件的检测与跟踪^[3]

收稿日期:2013-07-29

收修改稿日期:2013-10-08

* 本文系教育部人文社会科学基金项目“网络伪舆情的识别与防范机制研究”(项目编号:11YJAZH060)的研究成果之一。

等。网络伪舆情的研究是网络舆情分析的一个新的研究方向,“网络伪舆情”一词最早由宋常青^[4]提出,主要研究了网络伪舆情对政府决策的干扰;刘勘等^[5]对网络伪舆情的内涵和特征进行了研究;李兰玉^[6]则对网络伪舆情的成因和对策进行了分析。但是网络伪舆情的识别技术还在探索之中。

选取合适的评价指标,提取网络舆情的关键信息,对网络舆情(包括真实舆情和虚假舆情)进行特征抽取,形成可供机器学习的描述数据,是对网络伪舆情进行有效识别的基础。针对网络舆情的指标体系,张一文等^[7]给出了指标体系的概念界定和基本维度;王青等^[8]从舆情热度、强度、倾度和生长度等4个维度构建了舆情监测及预警体系;张玉亮^[9]通过对突发事件网络舆情发生周期的分析,提出了基于三个层次的21个风险评估指标体系;陈新杰等^[10]从传播方式、视听化程度、内容详略度等方面进行了指标的丰富。本文综合有关对舆情指标的研究,结合网络伪舆情的特点,构建了适于判别真伪舆情的指标体系。

支持向量机(SVM)是 Vapnik^[11]提出的一种基于统计学习理论的方法,SVM 通过学习自动寻找那些对分类有较好区别能力的支持向量,并将这些向量映射到一个更高维的空间,在这个空间建立一个最大间隔超平面,通过这个超平面把数据区分开。运用 SVM 方法得到的分类器,类和类之间的距离较大,推广性能和分类准确率较高。网络伪舆情的识别其本质可以看作是一个分类问题,网络舆情分类具有训练样本数量有限、维数较高等特点,从理论上说这些特点符合支持向量机的分类特征,因为 SVM 具有训练样本小、学习速度快、易于扩展等特点,并能处理高维和线性不可分问题。因此本文提出基于支持向量机的网络真伪舆情分类模型,能够对网络伪舆情进行有效识别。

3 基于 SVM 的网络伪舆情识别模型

3.1 基本思路

网络舆情包括网络真实舆情和网络伪舆情,网络伪舆情的识别其本质就是对网络真实舆情和网络伪舆情进行分类。基于支持向量机的网络伪舆情识别模型如图1所示。

首先,建立网络伪舆情的判别指标。网络舆情事件可以通过一些主要的指标数据进行描述,这些评价

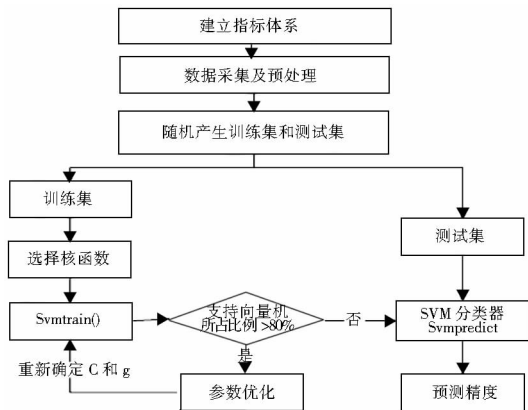


图1 网络伪舆情识别模型

指标之间存在着实体和属性的关系,合理、有效的评价指标的选取是对网络伪舆情进行有效识别的基础和前提。其次,根据舆情指标采集数据。收集的舆情数据既包括真实舆情,也包括虚假舆情,同时注重两类数据在数量上的平衡性。需要对收集到的数据进行预处理,主要是信息源数据权重的确定、来源于不同渠道的舆情信息直接信息系数的计算、缺失数据的补充以及不同量纲数据的归一化处理等。然后,将样本随机分为训练样本和测试样本,训练样本用于学习,以得到正确的分类器。最后,将测试样本输入到分类器中进行测试,分析测试结果,进而衡量分类器的精度。

具体实施过程中,主要涉及以下步骤:

- (1) 导入数据,并将数据进行归一化处理;
 - (2) 随机产生训练集和测试集;
 - (3) 核函数选择;
 - (4) 对训练样本进行 SVM 分类,学习产生分类器,确定支持向量的个数;
 - (5) 当支持向量的个数所占训练样本比例过高(>80%)时,采用参数优化方法选择最佳的惩罚因子 C 以及核函数中的参数 g;
 - (6) 运用新的参数惩罚因子 C 和参数 g 再次对训练样本进行分类,产生分类结果;
 - (7) 不断重复过程步骤(5)和步骤(6),直到支持向量的个数所占训练样本比例达到指定阈值(<80%),确定最终采用的惩罚因子 C 和参数 g;
 - (8) 对测试样本进行预测,统计预测准确率;
 - (9) 预测结果比较分析。
- 这里采用向量形式表示网络舆情信息。即针对每

个舆情事件,统一用向量 $X_i = (x_{i1}, x_{i2}, \dots, x_{i7}, y_i)$ 对其进行描述,其中, $x_{i1}, x_{i2}, \dots, x_{i7}$ 表示第 i 个样本的属性值,分别对应于每个舆情事件的信息源、总发文数、总报道数、直接信息系数、持续时间、观点总数和传播渠道数, $y_i \in (-1, 1)$, 表示两类训练样本的类别标识, $i = 1, 2, \dots, n, n$ 为样本总数。

3.2 网络伪舆情的识别指标

反映网络舆情的指标较多,但多数指标都难以用于判断网络舆情的真伪。根据网络伪舆情的主要特征^[5],本文选取了 7 个主要指标,包括:

(1)信息源:收集舆情事件相关信息时,追溯到其源头,判断信息源的类别,根据不同的类别赋予不同的权重。一般认为,当信息来源于官方网站和主流媒体时,可信度较大,赋予较高的权重;当信息来源于贴吧或个人微博时,可信度相对较小,赋予的权重也相应降低。

(2)总发文数:从第一篇文章开始计算,截止到统计时间点,在某一热点事件相关主题下发表的文章总数,包括论坛帖子数(原创和转帖)和博文数。

(3)总报道数:从第一篇文章开始计算,截止到统计时间点,在某一热点事件相关主题下所有网络媒体新闻报道的总数,包括消息和评论。

(4)直接信息系数:反映同一事件涉及的相关信息在不同信息来源上的分布状况的综合参考系数^[12]。计算方法分以下几个步骤:将所获得的同一事件的所有信息按其来源分类,并为不同的信息来源分配不同的权重;分别计算各类信息来源所包含的信息数与信息总数的比值,在此定义为信息显著度;计算各类信息来源信息显著度的加权平均数,即直接信息系数。

假设从 N 类信息源收集到有关某一事件的 M 条信息, S_k 为从第 K 类信息源中获得的信息数。 W_k 为第 K 类信息源的可靠性权重,直接信息系数为 I ,则具体计算公式如下:

$$\sum_{k=1}^N W_k = 1 \quad \sum_{k=1}^N S_k = M \quad I = \frac{\sum_{k=1}^N (W_k \frac{S_k}{M})}{N} \quad k \in (1, 2, \dots, N)$$

(5)持续时间:事件从发生到结束持续的天数,为使样本更加精确,也可以小时为单位进行度量。

(6)观点总数:将收集到的舆情信息加以整理,形成文本文档,再从文本文档中提取观点,进一步统计观点总数。

(7)传播渠道数:当信息的传播方式为串行传播

时,统一为一条传播渠道;当信息的传播方式为并行传播时,需要统计传播过程中涉及到的所有传播渠道数。

在这些指标中,有的能反映网络伪舆情的特征,如网络伪舆情总是来自于非官方的消息源、直接信息系数较低、被人为地加长持续时间、观点较为单一、传播渠道类似等。同时,也利用了发文总数、报道总数等指标,以表现其舆情特征。

3.3 支持向量机的核函数选取

核函数通过一个特征映射将输入空间(低维的)中线性不可分数据映射成高维特征空间中的线性可分数据,这样就可以对高维线性可分数据使用 SVM 方法进行识别。核函数本质上对应于高维空间的内积 $K(x_i, x_j) = \varphi(x_i) \varphi(x_j)$ 。本文主要采用以下几个核函数对模型进行测试:

(1)线性内积核函数^[13]:

$$K(x_i, x_j) = x_i \cdot x_j$$

(2)多项式核函数(Poly)^[13]:

$$K(x_i, x_j) = (g(x_i^T \cdot x_j) + r)^d \quad d > 0, d \text{ 阶多项式分类器}$$

(3)径向基内积核函数(RBF)^[14]:

$$K(x_i, x_j) = \exp(-g \|x_i - x_j\|^2) \quad g > 0$$

(4)S 型核函数(Sigmoid)^[15]:

$$K(x_i, x_j) = \tanh(g(x_i^T \cdot x_j) + r)$$

4 实验过程及结果

4.1 数据和实验结果

实验语料共收集了 79 个网络热点舆情数据,分为两类,包括真实舆情事件 41 个和伪舆情事件 38 个。具体实验过程中,将样本随机划分为训练集和测试集,其中训练数据 59 个,测试数据 20 个。针对评价指标中涉及的网络媒体,考虑到媒体的多样性和复杂性以及网民的参与程度,语料来源主要选取比较有代表性的商业网站和官方网站,包括新浪网、腾讯网、人民网、新华网、百度贴吧和天涯论坛。对所有网络媒体涉及到的报道、评论、帖子中所包含的观点进行人工提取。同时,针对每个舆情事件,都追溯其信息源,并根据信息源可信度的不同分别赋予不同的权重,权重的取值介于 0 和 1 之间。具体为:当事件来源于政府官方网站或主流媒体时将信息源可信度定为 0.8;当事件来源于地方报纸时将信息源可信度定位 0.5;当事件来源于论坛、贴吧、个人博客、手机短信或 QQ 群时将信息源可信度定为 0.2。实验数据的组成及结构如图 2 所示。

事件	总发文数	总报道数	信息总数	受众总人数	直接信息系数	传播渠道数	持续天数	信息来源可信度	总观点数	点击频率	回复频率	信息源
云南盈江地震	1917	113	2030	59923	0.029625042	8	13	0.85	16	53869	3889	云南省民政厅
雨润问题肉	175	92	267	17030	0.041847826	5	7	0.4	12	14151	2612	华商网
定鸡氮肥厂爆炸	1175	84	1259	3201	0.029371809	8	18	0.7	8	1615	365	西部网
台湾塑化剂污染食品	765572	136	765708	1022184	0.035428734	6	7	0.65	10	239989	16487	新京报
西安爆炸事件	48265	269	48534	410877	0.029147434	8	9	0.7	21	327436	22817	西部网
甘肃庆阳校车肇事	3696	900	4596	34987	0.029755253	8	8	0.7	22	23124	4048	中广网(兰州)

图2 部分实验数据示例图

整个实验在 Matlab 平台上完成。首先使用函数默认的参数值,分别选择不同的核函数进行分类,分类结果如表 1 所示:

表 1 不同核函数下的分类结果

核函数	C	g	d	r	支持向量数	支持向量所占比例	整体分类准确率	预测准确率
线性	1	-	-	-	4	6.78%	100%	100%
RBF	1	0.1	-	-	59	100%	100%	65%
多项式	1	0.1	3	0	5	8.47%	98.31%	95%
Sigmoid	1	0.1	-	0	54	91.53%	54.24%	45%

针对采用径向基核函数(RBF)时得到的准确率较高但支持向量所占训练样本比例过高的情况,分别采用基于 Grid Search(网格划分)的参数寻优方法^[16]、基于 GA(遗传算法)的参数寻优方法^[17]以及基于 PSO(粒子群)的参数寻优方法^[18]对参数进行了优化,进一步提高了分类器的预测准确率。优化之后的参数选择和实验结果如表 2 所示:

表 2 RBF 优化前后预测精度比较

优化方法	C	g	支持向量数	支持向量所占比例	整体分类准确率	预测准确率
未优化	1	0.1	59	100%	100%	65%
Grid Search	1	0.002	46	77.97%	100%	95%
PSO	0.95	0.001	44	74.58%	100%	90%
GA	8.645 5	0.085 8	59	100%	54.24%	45%

4.2 分类的准确率

实验中采用测试数据对识别模型的准确率进行客观评估,利用混淆矩阵(如表 3 所示)判断其准确率。

表 3 正、负样本的混淆矩阵

实际的类 \ 预测的类	C1	C2
	C1 (pos)	真正(t_pos)
C2 (neg)	假正(f_pos)	真负(t_neg)

本文使用三个度量指标对模型的准确率进行衡量。包括真正识别率 R (Recognition, 即正确识别的正样本在所有正样本中所占的百分比)、真负识别率 S (Specificity, 即正确识别的负样本在所有负样本中所占的百分比)和精度 P (Precision, 即正确识别出的正样本在所有预测为正样本的样本中的百分比)。具体计

算公式为:

$$R = \frac{t_pos}{pos} \quad S = \frac{t_neg}{neg} \quad P = \frac{t_pos}{t_pos + f_pos}$$

其中,t_pos 是正确分类的正样本数,pos 是正样本数;t_neg 是正确分类的负样本数,neg 是负样本数;而 f_pos 是将负样本错误判断为正样本的样本数。

根据上述分类准确率的度量方法,本次实验涉及的 20 个预测样本中,正样本数即 pos = 9,负样本数为 neg = 11,分别计算真正识别率(R)、真负识别率(S)和最终精度(P)。同时使用分类算法中常用的神经网络和朴素贝叶斯算法对实验数据进行分类,进而对比各个算法的分类准确性。计算结果如表 4 所示:

表 4 基于混淆矩阵的预测精度

分类方法	t_pos	t_neg	f_pos	R	S	P
线性	9	11	0	100%	100%	100%
RBF	9	4	7	100%	36.36%	56.25%
多项式	8	11	0	88.9%	100%	100%
Sigmoid	9	0	11	100%	0	45%
Grid Search	9	10	1	100%	90.9%	90%
PSO	9	10	1	100%	90.9%	90%
GA	9	4	7	100%	36.36%	56.25%
神经网络	8	5	6	88.9%	45.5%	57.14%
朴素贝叶斯	8	3	8	88.9%	27.27%	50%

4.3 结果及讨论

由表 1 可知,不同的核函数所对应的分类器的预测精度存在一定的区别,本实验中主要表现在:线性核函数和多项式核函数得到的预测精度最高,且支持向量的个数较为理想,但线性核函数产生的分类器可能存在“过学习问题”,也即分类器由于维数较高能够识别每个样本,但对样本之外的其他数据大多分类错误;径向基核函数得到的预测精度较高,但是产生的支持向量偏多,需要进行参数的优化;Sigmoid 核函数得到的预测精度很低,支持向量偏多。总体而言,针对网络伪舆情的识别,选择径向基核函数或多项式核函数较为合适,但径向基核函数的使用需要进行参数的优化。

对径向基核函数进行优化之后,结合表 2 所示结果可知:采用 Grid Search 方法和 PSO 方法进行优化效

果较好,它们不仅能够未优化的基础上提高预测的精度,还能减少支持向量的个数;而采用 GA 方法进行优化效果欠佳,不仅没有减少支持向量的个数反而降低了优化之前的预测精度。分析基于混淆矩阵的预测精度可知:多项式核函数产生的分类器预测精度较高,采用 Grid Search 方法和 PSO 方法进行优化之后的径向基核函数得到的分类器分类效果较好。该结果和表 2 所示的结果一致,说明多项式核函数比较适合于网络伪舆情识别模型,优化之后的径向基核函数所产生的分类器也具有良好的分类效果。

同时,通过实验结果的分析与比较发现:采用支持向量机方法、应用本文提出的识别模型对网络伪舆情进行识别较之其他分类方法的准确性得到了有效的提高。但是,运用 SVM 方法进行伪舆情的识别不仅要注意核函数的选择,更要考虑影响识别效果的舆情指标的选取。能够对网络舆情信息进行描述的评价指标很多,比如:报道数、点击率、浏览量、评论数、传播渠道数、观点倾向性等,但是对于网络伪舆情的识别而言,最为关键的是要选择那些能够使其明显区别于网络真实舆情的评价指标。基于对网络真实舆情和网络伪舆情特征的详细分析,结合实验结果发现:信息源、观点总数、持续时间是网络伪舆情区别于网络真实舆情的重要指标。具体表现是,网络伪舆情信息通常来自于论坛、贴吧的网民个人发帖或 QQ 消息群发,观点较为集中且持续时间相对较短;而网络真实舆情信息一般来源于官方网站或各大主流媒体,观点较为分散,不具有明显倾向性,并且事件发展持续时间较长。

5 结 语

支持向量机对于小样本数据具有良好的学习能力,已经被广泛应用于人脸识别、入侵检测以及机械故障诊断等许多方面。本文初步探讨了采用支持向量机方法对网络舆情中的网络伪舆情信息进行识别,并通过实验证明提出的基于 SVM 的识别模型能够将网络真实舆情和网络伪舆情有效区分开。核函数的使用是支持向量机的精髓,但是目前没有统一的选择标准,实验结果表明,利用线性的和多项式的核函数可以达到较好的效果,利用 RBF 核函数则需要进行网格划分或粒子群寻优才能得到好的结果。本文的识别方法具有算法简单、识别能力较强、分类精度较高的优点,为网

络虚假信息的甄别提供了一种新的途径。可以说分类器达到了预期的目的,但样本数量的局限性导致预测精度尚有改进的空间。

参考文献:

- [1] 周亚东,孙钦东,管晓宏,等. 流量内容词语相关度的网络热点话题提取[J]. 西安交通大学学报,2007, 41(10): 1142 - 1146. (Zhou Yadong, Sun Qindong, Guan Xiaohong, et al. Internet Popular Topics Extraction of Traffic Content Words Correlation[J]. *Journal of Xi'an Jiaotong University*, 2007, 41(10): 1142 - 1146.)
- [2] 廖祥文,曹冬林,方滨兴,等. 基于概率推理模型的博客倾向性检索研究[J]. 计算机研究与发展, 2009, 46(9): 1530 - 1536. (Liao Xiangwen, Cao Donglin, Fang Binxing, et al. Research on Blog Opinion Retrieval Based on Probabilistic Inference Model[J]. *Journal of Computer Research and Development*, 2009, 46(9): 1530 - 1536.)
- [3] 朱恒民,苏苏宁,张相斌. 基于链接网络图的互联网舆情话题跟踪方法[J]. 情报学报, 2011, 30(12): 1235 - 1241. (Zhu Hengmin, Su Xinning, Zhang Xiangbin. A Topic Tracking Method of Internet Public Opinion Based on Link Network Diagram[J]. *Journal of the China Society for Scientific and Technical Information*, 2011, 30(12): 1235 - 1241.)
- [4] 宋常青. “网络伪舆情”干扰政府决策[J]. 半月谈, 2010(3): 6 - 7. (Song Changqing. Raise Guard on Fake Internet Opinion Interfere with Government Decisions[J]. *China Comment*, 2010(3): 6 - 7.)
- [5] 刘勘,朱怀萍,胡航. 网络伪舆情的特征研究[J]. 情报杂志, 2011, 30(11): 57 - 60. (Liu Kan, Zhu Huaiping, Hu Hang. Study on Characteristics of Internet Fake Public Opinion[J]. *Journal of Intelligence*, 2011, 30(11): 57 - 60.)
- [6] 李兰玉. 网络伪舆情的成因及对策分析[J]. 人民论坛, 2013, 22(17): 42 - 43. (Li Lanyu. The Reasons and Strategies for Internet Fake Public Opinion[J]. *People's Tribune*, 2013, 22(17): 42 - 43.)
- [7] 张一文,齐佳音,方滨兴,等. 非常规突发事件网络舆情指标体系建立初探——概念界定与基本维度[J]. 北京邮电大学学报:社会科学版, 2010, 12(4): 6 - 14. (Zhang Yiwen, Qi Jiayin, Fang Binxing, et al. Research on the Index System of Public Opinion on Internet for Abnormal Emergency: Concept Definition and Basic Dimensionality [J]. *Journal of Beijing University of Posts and Telecommunications; Social Sciences Edition*, 2010, 12(4): 6 - 14.)
- [8] 王青,成颖,巢乃鹏. 网络舆情监测及预警指标体系构建研究[J]. 图书情报工作, 2011, 55(8): 54 - 57. (Wang Qing,

- Cheng Ying, Chao Naipeng. On the Construction of Internet Public Opinion Index System for Monitoring and Early Warning[J]. *Library and Information Service*, 2011, 55(8):54-57.)
- [9] 张玉亮. 基于发生周期的突发事件网络舆情风险评价指标体系[J]. *情报科学*, 2012, 30(7): 1034-1038. (Zhang Yuliang. Network Opinion Risk Evaluation Index System Based on the Cycle of Emergency[J]. *Information Science*, 2012, 30(7): 1034-1038.)
- [10] 陈新杰,呼雨,兰月新. 网络舆情监测指标体系构建研究[J]. *现代情报*, 2012, 32(5): 4-7. (Chen Xinjie, Hu Yu, Lan Yuexin. Index Systems of Network Public Sentiment for Monitoring [J]. *Modern Information*, 2012, 32(5):4-7.)
- [11] Vapnik V N. The Nature of Statistical Learning Theory [M]. New York, NY, USA: Springer-Verlag, 2000:138-156.
- [12] 冯维扬. 反竞争情报对策研究:虚假信息与信息可靠性定量分析[J]. *情报学报*, 2001, 20(6):728-732. (Feng Weiyang. Study on Strategy Against Counterintelligence: False Information and Quantitative Analysis of Reliability of Information[J]. *Journal of the China Society for Scientific and Technical Information*, 2001, 20(6):728-732.)
- [13] Cortes C, Vapnik V. Support - vector Networks [J]. *Machine Learning*, 1995,20(3):273-297.
- [14] Amari S, Wu S. Improving Support Vector Machine Classifiers by Modifying Kernel Functions [J]. *Neural Networks*, 1999, 12(6): 783-789.
- [15] Lin H T, Lin C J. A Study on Sigmoid Kernels for SVM and the Training of Non - PSD Kernels by SMO - type Methods[EB/OL]. [2013-10-05]. <http://www.csie.ntu.edu.tw/~cjlin/papers/tanh.pdf>.
- [16] Hsu C W, Chang C C, Lin C J. A Practical Guide to Support Vector Classification [EB/OL]. [2013-10-05]. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- [17] Huang C L, Wang C J. A GA - based Feature Selection and Parameters Optimization for Support Vector Machines [J]. *Expert Systems with Applications*, 2006, 31(2):231-240.
- [18] Paquet U, Engelbrecht A P. Training Support Vector Machines with Particle Swarms [C]. In: *Proceedings of the IEEE International Joint Conference on Neural Network*. New Jersey: IEEE Press, 2003: 1593-1598.

(作者 E-mail:liukan@znufe.edu.cn)

EBSCO 信息服务加入 ORCID

EBSCO 信息服务公司(简称 EBSCO)宣布成为研究人员及投稿人公开资料库(Open Researcher and Contributor ID, ORCID)的一员。ORCID 是一个开放的、非营利性的、社区驱动的组织,旨在创建并维护一个独一无二的持久的研究人员标识。ORCID 整个社区成员通力合作,将这一标识机制推广到整个科研工作流之中,并与科研人员的学术活动及学术产出进行关联。

EBSCO 执行副总裁兼订阅服务首席运营官 Robert Schoenvogel 认为,ORCID 将会产生深远的影响,“ORCID 目前正在做的工作会将大量的科研人员和机构联系在一起,整个科研学术圈都将受益无穷。通过将这一独特且持久的标识符整合到 EBSCO 的产品之中,特别是 EBSCO 即将推出的论文加工费管理服务,将有助于改善作者和出版商的体验,改善整个科学发现过程。”

ORCID 是 EBSCO 最近加入的一个支持开放信息交换的组织,EBSCO 同时还是开放获取学术出版协会(Open Access Scholarly Publishers Association, OASPA)的一员,该协会旨在维护开放获取期刊出版商的利益。

(编译自 :<http://www.prweb.com/releases/EBSCO/ORCID/prweb11254899.htm>)

(本刊讯)