Advances in Applied Mathematics 应用数学进展, 2017, 6(9), 1073-1081
Published Online December 2017 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2017.69130

Hans汉斯

A Synchronous Optimization Algorithm for Increasing Accuracy of SVM Classification

Fan He, Changjing Lu

School of Mathematics and Physics, China University of Geosciences, Wuhan Hubei
Email: hefan@cug.edu.cn, luchangjing@cug.edu.cn

Received: Nov. 24th, 2017; accepted: Dec. 7th, 2017; published: Dec. 14th, 2017

Abstract

Support vector machines (SVM), which is a popular method for pattern classification, has been recently adopted in range of problems. In training procedure of SVM, feature selection and parameter optimization are two main factors that impact on classification accuracy. In order to improve the classification accuracy by optimizing parameter and choosing feature subset for SVM, a new algorithm is proposed through combining Bat Algorithm (BA) with SVM, termed BA + SVM. For assessing the performance of BA + SVM, 10 public data-sets are employed to test the classification accuracy rate. Compared with grid algorithm, conventional parameter optimization method, our study concludes that BA + SVM has a higher classification accuracy with fewer input features for support vector classification.

Keywords

SVM, Bat Algorithm, Feature Selection, Parameter Optimization, Classification

一种提高SVM分类能力的同步优化算法

何　凡，卢常景

中国地质大学数学与物理学院，湖北　武汉
Email: hefan@cug.edu.cn, luchangjing@cug.edu.cn

收稿日期：2017年11月24日；录用日期：2017年12月7日；发布日期：2017年12月14日

摘　要

近年来，支持向量机(SVM)理论广泛应用于模式分类，然而影响其分类准确率的两个主要因素特征选择和参数优化又是相互影响和制约的。文章提出一种BA + SVM算法，利用蝙蝠算法(BA)来同步完成SVM的参数优化和输入数据的特征属性选择，提高了SVM的分类能力。设计的三种实验方式在10个测试数据集

文章引用: 何凡, 卢常景. 一种提高 SVM 分类能力的同步优化算法[J]. 应用数学进展, 2017, 6(9): 1073-1081.
DOI: 10.12677/aam.2017.69130

上实验结果表明，BA + SVM同步优化算法与单一进行参数优化或单一进行特征选择算法相比，具有输入特征少分类准确率高的优势。

关键词

SVM，蝙蝠算法，特征选择，参数优化，分类

1. 引言

支持向量机(SVM) [1]是一种普遍、高效的模式分类方法，已被广泛应用于文本分类[2]、生物信息学和模式识别[3]等领域。提高 SVM 分类准确率面临的两个主要问题是：如何选择核函数的最佳参数，以及如何选择用于训练和测试的最佳特征子集。

提高分类准确率一般有三种方法。其一，优化 SVM 核函数的参数。具体途径是网格寻优或者使用交叉验证得到 SVM 的经验参数，并普遍应用于参数选择研究。其二，是对输入样本进行特征选择。因为选择不当的数据特征可能会导致较差的分类精度[4]。特征选择是选择输入样本的重要特征，不仅删除了一些冗余甚至无关的特征[5]，同时也可提高分类准确率。第三，SVM 的参数优化和特征选择同步进行。这种方法是近年来的研究热点。比如，Adriano L.I. Oliveira 等[6]提出的 GA + SVM 模型用于机器学习回归的特征选择和参数优化，Huang [7]和 Lin [8]采用的粒子群优化算法(PSO)来优化特征子集和 SVM 参数，都取得了有效的成果。最近，Li [9]采用遗传算法(GA)实现特征选择和优化 SVM 参数进行网络入侵检测。

蝙蝠算法(BA)是一种具有回波定位特征的元启发式优化算法[10]。为了得到优化结果，BA 采用频率调整方式，模仿蝙蝠的脉搏和响度的变化，使用自动缩放功能来平衡探索，简化搜索过程。

因此，这就意味着 BA 比其他算法效率更高。有结果表明基于支持向量机的 BA 参数优化算法取得了满意结果[11]。对这个结果做进一步的研究发现，这种元启发式算法还可以同步实现特征选择和 SVM 参数优化。文章提出的基于蝙蝠算法的特征选择与参数优化的向量机分类算法，称为 BA + SVM。这种方法通过蝙蝠的位置更新方式同步执行特征选择和参数优化，不但优化了 SVM 参数、选择了适当的样本特征，还提高了 SVM 的分类准确率。

2. 相关工作

2.1. SVM 分类器

给定一组具有标签的样本集(x_i, y_i)，$i = 1, 2, ..., m$，其中 m 是训练数据的个数，$x_i \in R^n$，$y_i \in \{-1, 1\}$，n 表示输入样本维数。对线性可分离数据点，SVM 可以通过选择最佳分离超平面来进行训练数据分类。

$$\langle \omega \cdot x_i \rangle + b = 0,\ i = 1, 2, \cdots, m \tag{1}$$

如果存在满足方程式的超平面(1)，通过求解以下的优化问题，可以找到最优线性分离超平面：

$$\begin{aligned} &\min_{1 \le i \le m} \frac{1}{2} \omega^{\mathrm{T}} \omega + C \sum_{i=1}^{m} \xi_i \\ &\text{subject to}: y_i \left(\langle \omega \cdot x_i \rangle + b \right) - 1 \ge 0 \end{aligned} \tag{2}$$

其中 C 是惩罚参数，ξ_i 是非负的松弛变量。引入拉格朗日乘数 α_i，原优化问题可以通过其对偶问题来解决。通过求解对偶问题得到 α_{*i}，最后确定最优超平面的参数 ω_* 和 b_*。其决策函数的定义如下：

$$f(x)=sign\left(\sum_{i=1}^{m}y_i\alpha_i^*\langle x_i\cdot x\rangle+b^*\right) \tag{3}$$

非线性 SVM 分类器可以表示为：

$$f\left(x,\alpha_i^*,b^*\right)=sign\left(\sum_{i=1}^{m}y_i\alpha_i^*k\left(x_i,x\right)+b^*\right) \tag{4}$$

2.2. 蝙蝠算法

蝙蝠是一种具有回声定位的哺乳动物[7]，当蝙蝠在飞行时通过分析返回的回波来导航并检测猎物[11]。此外，蝙蝠能够通过回声定位系统获得物体之外的距离，以防止碰撞[12]。基于蝙蝠的回声定位特征和觅食行为，Yang [10]提出了蝙蝠算法。蝙蝠算法的基于以下规则：

规则 1：蝙蝠都通过回声定位感知距离，并且能通过某种方式分辨猎物和障碍物。

规则 2：蝙蝠在位置 x_i 以速度 v_i 和频率 $f_{\min}$ 随机飞行，通过不断变化的响度 A_0 和频率来搜寻猎物。它们自动调整所发射脉冲频率，并依据距离来调整脉冲发射频率 $r\in[0,1]$。

规则 3：虽然响度变化范围可以不同，我们通常假设其从最大值 A_0 减小到最小值 $A_{\min}$。

基于以上规则，蝙蝠算法的基本步骤为：

步骤 1：初始化基本参数。包括响度衰减系数 α，脉冲发射强化系数 γ，最大固定频率 $f_{\max}$，最小频率 $f_{\min}$，迭代次数 $iter_{\max}$ 和种群大小 N；

步骤 2：初始化蝙蝠算子 $i(i=1,2,\cdots,n)$ 的参数：位置 x_i，速度 v_i，固定频率 f_i 和响度 A_i；

步骤 3：当 $t<iter_{\max}$ 时生成新的解决方案，通过公式(8)~(10)更新速度和解决方案；

步骤 4：如果 $rand>r_i$，从最佳解决方案中选择一个解，并在选择的最佳解决方案周围生成当局最优解，然后通过自由飞行生成一个新的解决方案；

步骤 5：如果 $rand<A_i$ 且 $f(x_i)<f(x_*)$，接收新的解决方案，让 r_i 增加并且 A_i 减少；

步骤 6：根据其适应度值对蝙蝠进行排名，并找出当前的最佳位置 x_*；

步骤 7：如果达到停止条件(即达到最大许可迭代次数或达到搜索精度)，转到步骤 8；否则转到步骤 3 并继续搜索；

步骤 8：输出最佳适应值和全局最优解。

假设 t 时刻蝙蝠 i 的位置为 x_i^t，速度为 v_i^t，那么，蝙蝠算法的更新公式为

$$f_i=f_{\min}+\left(f_{\max}-f_{\min}\right)\beta \tag{5}$$

$$v_i^{t+1}=v_i^t+\left(x_i^t-x_*\right)f_i \tag{6}$$

$$x_i^{t+1}=x_i^t+v_i^{t+1} \tag{7}$$

其中 β 表示[0,1]中的随机数，f_i 表示当前时刻蝙蝠 i 的变化频率。这里，x_* 是当前的全局最优解。

为了提高解决方案的灵活性，从蝙蝠种群中选择一个蝙蝠，并根据方程(8)，通过蝙蝠的自由飞行来创建一个新的当局解决方案。这种随机游走可以解释为本地搜索过程，用于创建新的解决方案。

$$x_{new}=x_{old}+\varepsilon A^t \tag{8}$$

其中，A^t 代表蝙蝠群体在 t 时刻的平均响度，$\varepsilon\in[-1,1]$ 是随机向量。x_{old} 表示从当前最优解中选择的随机解。

在 BA 算法的迭代过程中，使用以下方程来更新响度 A_i 和发射脉冲速率 r_i：

$$A_i^{t+1} = \alpha A_i^t \tag{9}$$

$$r_i^{t+1} = r_i^0\left[1-\exp(-\gamma t)\right] \tag{10}$$

其中 α 和 γ 是常数，在大部分应用程序中我们通常设置 $0<\alpha<1, \gamma>0$。因此，很容易得到，当 $t\to\infty$，$A_i^t \to 0$ 时 $r_i^t \to r_i^0$。

2.3. 特征选择

特征选择的策略可以分为两种不同的模型，分别称为包装模型和滤波模型的模型[13]。与包装模型相比，滤波器模型具有更快的速度。然而，包装模型可有利于找到最好的特征子集[14]。

3. 基于 BA 的 SVM 特征选择和参数优化

3.1. 蝙蝠位置的表示

使用 RBF 核函数的 SVM 来实现 BA + SVM 方法。蝙蝠位置向量由特征掩码(离散值)和参数掩码(连续值 C 和连续值 γ)两部分组成。蝙蝠位置向量的具体设计如表 1 所示。

其中，n 表示数据集的特征数量，输入特征掩码 $x_{i,1}\sim x_{i,n}$ 是布尔值，“1”表示选择了该特征，“0”表示未选择该特征。$x_{i,n+1}$ 表示参数 C 的值，$x_{i,n+2}$ 表示参数 γ 的值。

3.2. 蝙蝠位置的更新标准

在更新蝙蝠的位置和速度的过程中，速度公式(6)保持不变。对于特征掩码 1~n，新的蝙蝠位置的值由以下规则来计算：if $\text{rand}(.)<S(v_{i,d})$，then $x_{i,d}=1$；else $x_{i,d}=0$，其中 $S(v)=\dfrac{1}{1+\mathrm{e}^{-v}}$。$S(v)$ 是一个 sigmoid 函数，rand(.)是从[0,1]中选出的随机数。参数掩码的位置公式(7)不变。

3.3. 适应度函数

根据实验，更高的分类准确率和更少的特征数会产生更好的适应度值。因此，设计适应度函数如下：

$$fit_i = \omega_A \cdot acc_i + \omega_F \cdot \left(1-\frac{\sum_{j=1}^{n} f_j}{n}\right) \tag{11}$$

ω_A 是 SVM 的分类准确率权重，ω_F 是所选特征数量的权重，用户可根据需要进行适当调整。如果选择了特征 j，$f_j=1$，否则 $f_j=0$。acc_i 表示 SVM 分类准确率，由公式(12)给出。ac 和 uc 分别表示正确分类的样本数和不正确分类的样本数。

$$acc_i = \frac{cc}{cc+uc}\times 100\% \tag{12}$$

4. BA + SVM 参数优化和特征选择算法

BA + SVM 参数优化和特征选择的流程图如图 1 所示，详细的实验步骤如下：

步骤 1：预处理数据：对样本数据做归一化。

步骤 2：蝙蝠初始化及 SVM 参数初始化：即设置蝙蝠算法的参数，包括蝙蝠个数，迭代次数，噪音系数，脉冲强度系数，脉冲频率限制以及适应度函数的权重。以及 C 和 γ 的初始值。

步骤 3：建立 SVM 模型并计算准确率：选择用于训练 SVM 分类器的输入特征，并于计算基于 C 和 γ 的平均分类准确率。

Table 1. The composition of the location of the bat *i*
表 1. 蝙蝠 *i* 的位置组成表

	特征掩码				参数掩码	
形式	$x_{i,1}$	$x_{i,2}$	…	$x_{i,n}$	$x_{i,n+1}(C)$	$x_{i,n+2}(\gamma)$

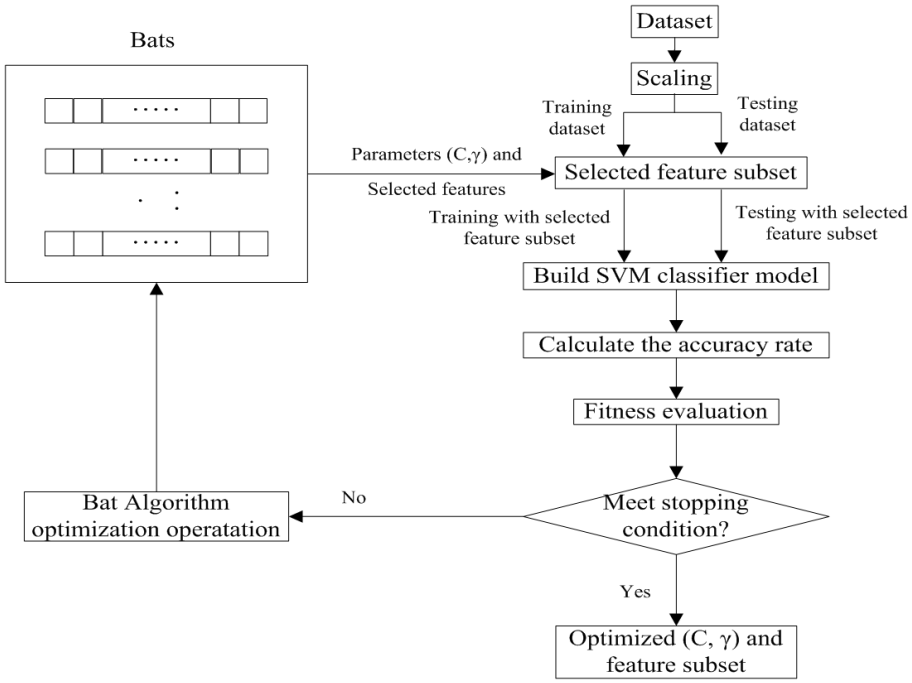

Figure 1. Flow chart: BA + SVM parameter optimization and feature selection
图 1. BA+SVM 参数优化和特征选择的流程图

步骤 4：更新全局最优解：该模型中每个蝙蝠的适应度值根据公式(11)进行计算。根据适合度值更新全局最优解。

步骤 5：判断停止条件：当满足停止条件时，输出最优特征子集和最优参数 C 和 γ；否则，进行下一步。

步骤 6：更新蝙蝠算法：根据蝙蝠算法更新每个蝙蝠的速度和位置并搜索更好的解决方案，然后转到步骤 3。

5. 实验结果

5.1. 平台和数据集

运行 BA + SVM 方法的平台是具有以下功能的 PC 端：Intel Pentium®双核 CPU，2.5GHz，2.00GB RAM，Windows 7 操作系统，运行环境为 MathWorks MATLAB 7.9 R2009b(Windows)，软件为 Libsvm [15]。

为了评估 BA + SVM 方法的分类能力，使用 UCI 机器学习库中的 10 个数据集进行测试，如表 2 所示。

5.2. 评估方法

停止标准是迭代次数达到 500，或适应度值在前 100 次迭代中没有改善。

多分类问题的准确率一般通过总体命中率进行评估，两分类问题的准确性评估一般需要三个指标进行详细说明，表 3 给出两分类问题的分类情况。

Table 2. UCI machine learning library data set
表 2. UCI 机器学习库的数据集

标号	数据集	分类数	实例数	特征数
1	Breast cancer	2	699	10
2	Car evaluation	4	1728	6
3	Glass	6	214	9
4	Ionosphere	2	351	34
5	Iris	3	150	4
6	Pima-Indian	2	768	8
7	Seeds	3	210	7
8	Sonar	2	208	60

Table 3. Classification of the two categories of issues
表 3. 两分类问题的分类情况

		预测	
		正样本	负样本
实际	正样本	TruePositive	FalseNegative
	负样本	FalsePositive	TrueNegative

TP 和 FN 分别表示正样本的正确分类率和正样本的不正确分类率，是两个重要的性能指标，计算公式表示如下：

$$\mathrm{TP}=\frac{\#\mathrm{TruePositive}}{\#\mathrm{FalseNegative}+\#\mathrm{TruePositive}} \tag{13}$$

$$\mathrm{TN}=\frac{\#\mathrm{TureNegative}}{\#\mathrm{TrueNegative}+\#\mathrm{FalsePositive}} \tag{14}$$

$$\text{Average accuracy}=\frac{\#\mathrm{TruePositive}+\#\mathrm{TruePositive}}{\#\mathrm{TestingSample}} \tag{15}$$

5.3. 实验结果

设计了三种实验方式，如表 4 所示。实验中，BA 群体大小为 20，响度衰减系数为 $\alpha=0.5$，脉冲发射系数为 $\gamma=0.5$，脉冲频率限制为 $f_i\in[0,2]$。SVM 中参数C的搜索范围为 0.01~10000，参数 γ 的搜索范围为 0.01~1000。适应度函数中，$\omega_A=0.8$ 和 $\omega_F=0.2$。

表 5 结果显示，BA + SVM 和 PSO + SVM 在不同数据集上产生比 SVM 更高的分类准确率，说明这两种组合算法确实可以获得更好的 SVM 参数和特征子集。此外，在没有特征选择的情况下，BA + SVM 在 6 个数据集中产生较高的分类准确率。进行特征选择的情况下，BA + SVM 在 7 个数据集中产生较高的分类准确率。一般来说，同步进行参数优化和特征选择的 BA + SVM 比 PSO + SVM 具有更好的性能。

网格算法是一种常规的参数优化方法。表 6 列出的是蝙蝠算法的参数寻优和网格算法的参数寻优比较结果。对于二分类数据集，BA 寻优的平均正命中率和平均负命中率均高于网格寻优算法。

图 2(a)到图 2(d)显示了 4 个数据集中三种优化方法的测试结果。显然，在选择的四个数据集上，同

Table 4. Experimental design
表 4. 实验方式设计

实验方式	特征掩码	参数掩码
1	开	关
2	关	开
3	开	开

Table 5. Comparison of classification results between BA + SVM and SVM and PSO + SVM
表 5. BA + SVM 与 SVM 和 PSO + SVM 的分类效果比较

数据集	SVM	特征选择(%)		无特征选择(%)	
		PSO + SVM	BA + SVM	PSO + SVM	BA + SVM
Breast cancer	88.97	91.11	94.38	92.27	95.96
Car evaluation	88.64	91.74	94.87	92.14	94.78
Glass	68.33	70.41	72.03	73.18	74.15
Ionosphere	92.14	94.78	92.76	95.22	94.22
Iris	79.28	90.36	90.09	90.33	94.69
Pima-Indian	77.87	80.78	80.04	85.5	82.11
Seeds	84.15	88.5	87.31	87.22	86.87
Sonar	85.14	93.41	92.1	96.26	95.24
Teaching	73.41	76.25	78.45	82.06	84.63
Wine	93.45	96.09	97.27	97.6	98.64

Table 6. Comparison of BA optimization and grid optimization without feature selection
表 6. 无特征选择情形下 BA 寻优和网格寻优对比

数据集	BA 寻优(%)			网格寻优(%)		
	TP	TN	准确率	TP	TN	准确率
Breast cancer	98.78	89.96	94.38	88.57	85.24	87.45
Ionosphere	94.63	90.24	92.76	92.84	90.21	90.50
Pima-Indian	82.48	79.19	80.04	79.15	74.02	75.24
Sonar	96.42	91.74	92.10	79.98	97.41	90.98

步进行参数优化和特征选择的 BA + SVM 具有最高的分类准确率，这表明同步优化的 BA + SVM 具有最优性能。

6. 结论

输入样本的特征选择和 SVM 核函数的参数设置是彼此相互影响的，提出的参数寻优和特征选择的同步算法就是为了解决这一博弈问题。仿真实验表明，无论是否进行特征选择，BA + SVM 都具有比 PSO + SVM

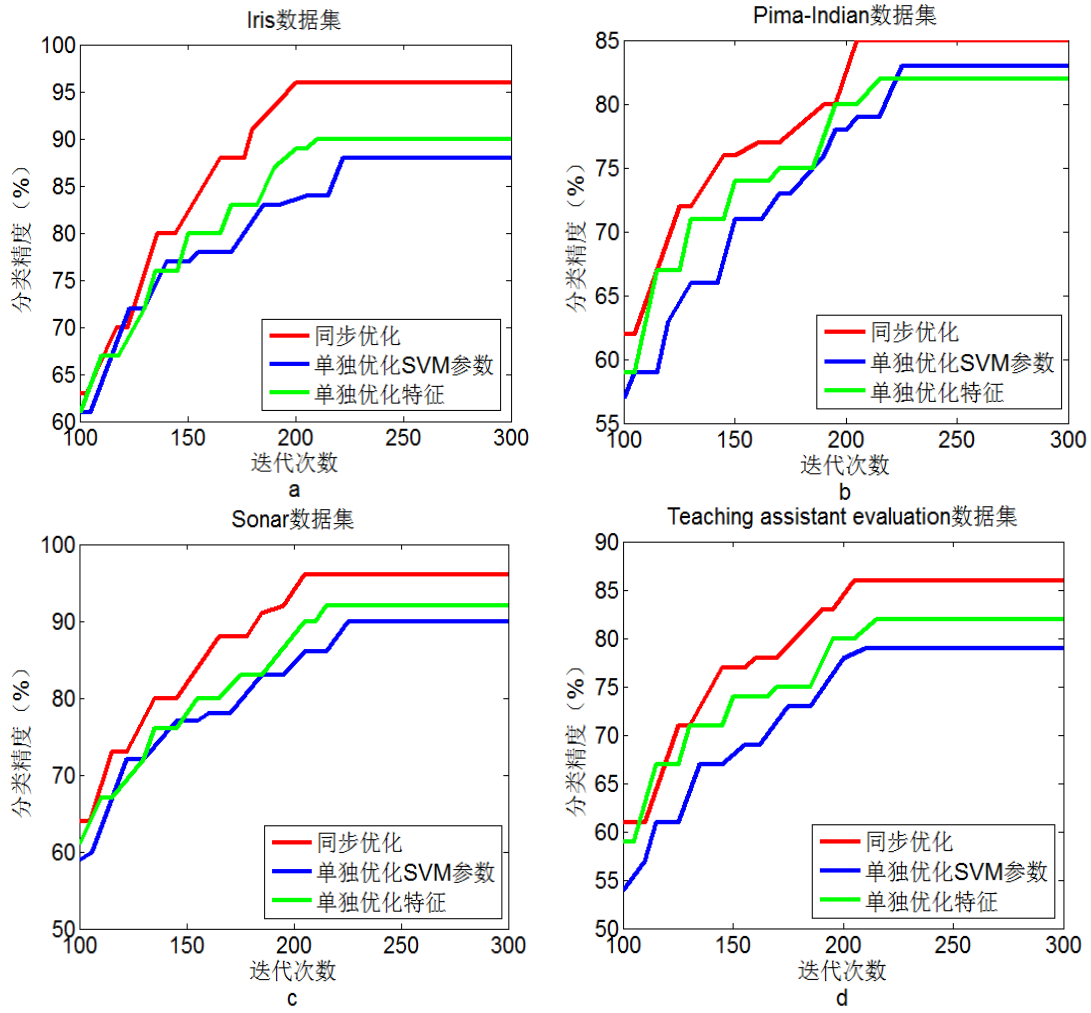

Figure 2. Line chart: Comparison of three experimental results on 4 data sets
图 2. 4 个数据集上三种实验结果比较

更高的分类准确率。就单纯参数优化而言，BA + SVM 有比网格寻优具有更好的 SVM 分类准确率。此外，4 个数据集上的仿真实验表明，同步进行参数优化和特征选择的 BA + SVM 比单独进行参数优化或特征选择的 BA + SVM 具有更好的分类效果。

基金项目

国家自然科学基金项目(11301492)；中国地质大学(武汉)基础研究基金项目(CUGL140420)。

参考文献 (References)

[1] Vapnik, V.N. (1995) The Nature of Statistical Learning Theory. Springer-Verlag, New York. https://doi.org/10.1007/978-1-4757-2440-0

[2] Joachims, T. (1998) Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *Proceedings of Support Vector Learning Machine Learning*: *ECML*-98, Volume 1398 of the Series Lecture Notes in Computer Science, 137-142. https://doi.org/10.1007/BFb0026683

[3] Yu, G.-X., Ostrouchov, G., Geist, A., *et al.* (2003) An SVM Based Algorithm for Identification of Photosynthesis-Specific Genome Features. *Proceedings of the* 2003 *IEEE Bioinformatics Conference on Computational Systems Bioinformatics*, California, 235-243. https://doi.org/10.1109/CSB.2003.1227323

[4] Keerthi, S.S. and Lin, C.-J. (2003) Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel. *Neural Computation*, **15**, 1667-1689. https://doi.org/10.1162/089976603321891855

[5] Xiao, H.J., Wei, Y.B., Yu, Z.L., *et al.* (2010) A New Active Learning Method for Instance Selection. *Journal of Information & Computational Science*, **7**, 2789-2795.

[6] Oliveira, A.L.I., Braga, P.L., Lima, R.M.F. and Cornélio, M.L. (2010) GA-Based Method for Feature Selection and Parameters Optimization for Machine Learning Regression Applied to Software Effort Estimation. *Information and Software Technology*, **52**, 1155-1166. https://doi.org/10.1016/j.infsof.2010.05.009

[7] Huang, C.-L. and Dun, J.-F. (2008) A Distributed PSO-SVM Hybrid System with Feature Selection and Parameter Optimization. *Applied Soft Computing*, **8**, 1381-1391. https://doi.org/10.1016/j.asoc.2007.10.007

[8] Lin, S.W., Ying, K.C., Chen, S.C., *et al.* (2008) Particle Swarm Optimization for Parameter Determination and Feature Selection of Support Vector Machines. *Expert Systems with Applications*, **35**, 1817-1824. https://doi.org/10.1016/j.eswa.2007.08.088

[9] Li, X.F. (2014) Network Intrusion Detection with Genetic Algorithm Synchronous Selecting Feature and SVM Parameters. *Computer Applications and Software*, **63**, S76.

[10] Griffin, D.R., Webster, F.A., Michael, C.R., *et al.* (1960) The Echolocation of Flying Insects by Bats. *Animal Behaviour*, **8**, 141-154. https://doi.org/10.1016/0003-3472(60)90022-1

[11] Tharwat, A., Hassanien, A.E. and Elnaghi, B.E. (2017) A BA-Based Algorithm for Parameter Optimization of Support Vector Machine. *Pattern Recognition Letters*, **93**, 13-22. https://doi.org/10.1016/j.patrec.2016.10.007

[12] Metzner, W. (1991) Echolocation Behaviour in Bats. *Science Progress Edinburgh*, **75**, 453-465.

[13] Liu, H. and Motoda, H. (1998) Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic, Boston.

[14] Chen, R.-C. and Hsieh, C.-H. (2006) Web Page Classification Based on a Support Vector Machine Using a Weighed Vote Schema. *Expert Systems with Applications*, **31**, 427-435. https://doi.org/10.1016/j.eswa.2005.09.079

[15] Chang, C.C. and Lin, C.J. (2001) LIBSVM: A Library for Support Vector Machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm