

DOI: 10.11992/tis.201905025

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.tp.20190719.1413.002.html>

基于人工鱼群算法的孪生支持向量机

李景灿^{1,2}, 丁世飞^{1,2}

(1. 中国矿业大学 计算机科学与技术学院 江苏 徐州 221116; 2. 矿山数字化教育部工程研究中心, 江苏 徐州 221116)

摘 要: 孪生支持向量机 (twin support vector machine, TWSVM) 是在支持向量机的基础上产生的机器学习算法, 具有训练速度快、分类性能优越等优点。但是孪生支持向量机无法很好地处理参数选择问题, 不合适的参数会降低分类能力。人工鱼群算法 (artificial fish swarm algorithm, AFSA) 是一种群智能优化算法, 具有较强的全局寻优能力和并行处理能力。本文将孪生支持向量机与人工鱼群算法结合, 来解决孪生支持向量机的参数选择问题。首先将孪生支持向量机的参数作为人工鱼的位置信息, 同时将分类准确率作为目标函数, 然后通过人工鱼的觅食、聚群、追尾和随机行为来更新位置和最优解, 最后迭代结束时得到最优参数和最优分类准确率。该算法在训练过程中自动确定孪生支持向量机的参数, 避免了参数选择的盲目性, 提高了孪生支持向量机的分类性能。
关键词: 孪生支持向量机; 人工鱼群算法; 模式分类; 参数优化; 准确率; 群体智能; 二次规划; 并行处理; 全局优化
中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2019)06-1121-06

中文引用格式: 李景灿, 丁世飞. 基于人工鱼群算法的孪生支持向量机 [J]. 智能系统学报, 2019, 14(6): 1121-1126.

英文引用格式: LI Jingcan, DING Shifei. Twin support vector machine based on artificial fish swarm algorithm[J]. CAAI transactions on intelligent systems, 2019, 14(6): 1121-1126.

Twin support vector machine based on artificial fish swarm algorithm

LI Jingcan^{1,2}, DING Shifei^{1,2}

(1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China; 2. Mine Digitization Engineering Research Center of Ministry of Education of the People's Republic of China, Xuzhou 221116, China)

Abstract: Twin support vector machine (TWSVM) is a machine learning algorithm based on the support vector machine. It has the advantages of fast training speed and superior classification performance. However, the algorithm cannot handle the parameter selection problem effectively, and the inappropriate parameters will reduce the classification ability. The artificial fish swarm algorithm (AFSA) is a group intelligent optimization algorithm with a strong global optimization ability and parallel processing capability. In this paper, TWSVM and AFSA are combined to solve the parameter selection problem of the TWSVM. First, the parameters of the support vector machine are taken as the position information of the artificial fish, and the classification accuracy is taken as the objective function. Then, the position and the optimal solution are updated by the artificial fish's preying, swarming, following, and random behavior. At the end of the iterations, the optimal parameters and the optimal classification accuracy are obtained. The algorithm automatically determines the parameters of the TWSVM in the training process, avoiding the blindness of parameter selection, and improves the classification performance of the TWSVM

Keywords: twin support vector machine; artificial fish swarm algorithm; pattern classification; parameter optimization; accuracy; swarm intelligence; quadratic programming; parallel processing; global optimization

于1995年提出。其在解决小样本、非线性和高维模式识别问题中有许多优势,目前已被应用到很多领域,例如文本分类、时间序列分析、模式识别和图像处理等^[3-7]。为了提高支持向量机的性能,学者们提出了许多改进算法。例如最小二乘支持向量机^[8]、近似支持向量机^[9]和基于广义特征值的近似支持向量机^[10]等。2007年,基于广义特征值近似支持向量机的思想, Jayadeva 等^[11]提出了孪生支持向量机。和 SVM 不同, TWSVM 是生成两个非平行的分类超平面,使得其中一个超平面尽量接近一类样本,并且尽可能的远离另一类样本。TWSVM^[12-13]将单个大型二次规划问题转化为两个较小的二次规划问题,明显缩短了运行时间。虽然孪生支持向量机的发展时间较短,但由于其坚实的理论基础和出色的性能表现,已经成为机器学习领域的一个研究热点。许多学者对 TWSVM 的研究做出了贡献,并取得了很多成果。例如光滑孪生支持向量机^[14]、最小二乘孪生支持向量机^[15-17]、模糊孪生支持向量机^[18]和投影孪生支持向量机^[19-20]等。

尽管近年来对 TWSVM 的研究在算法改进及其应用方面取得了很大的进展,但是仍有不足之处,其中之一就是 TWSVM 无法很好地处理参数选择问题。参数对分类结果有很大的影响,一个好的参数将直接提升 TWSVM 的性能。网格搜索^[21]是最普遍的参数选择方法,但是这种方法很耗时。Ding 等^[22-23]采用粒子群算法和量子粒子群算法对 TWSVM 的参数进行优化。Zhai 等^[24]将粒子群算法与局部搜索算法结合,提出了一种混合粒子群算法来优化小波孪生支持向量机的参数。随后, Sartakhti 等^[25]将最小二乘孪生支持向量机与模拟退火算法结合来提高分类准确率。

人工鱼群算法 (artificial fish swarm algorithm, AFSA)^[26]是一种基于动物行为的群体智能优化算法。该算法具有收敛速度快、灵活性强、容错性强、准确率高等优点。本文将人工鱼群算法与孪生支持向量机进行结合,以分类准确率作为适应度值,来对惩罚参数和核参数进行优化,提高算法的有效性。

1 相关理论

1.1 孪生支持向量机

假设在 n 维实空间 R^n 中有 m 个训练样本,其中属于+1类的有 m_1 个,属于-1类的有 m_2 个。令

$A \in R^{m_1 \times n}$ 表示+1类的训练样本, $B \in R^{m_2 \times n}$ 表示-1类的训练样本。TWSVM 的训练过程是要在 R^n 中找到两个非平行的超平面:

$$K(x^T, C^T)w_1 + b_1 = 0, K(x^T, C^T)w_2 + b_2 = 0 \quad (1)$$

其中 K 为核函数, $C=[A;B]$ 代表所有的训练样本, $w_i (i=1,2)$ 是分类超平面的法向量, $b_i (i=1,2)$ 为偏置。

通过求解下面两个二次规划问题可以得到这两个超平面:

$$\min_{w_1, b_1, \xi_1} \frac{1}{2} \|K(A, C^T)w_1 + e_1 b_1\|^2 + c_1 e_2^T \xi_2 \quad (2)$$

$$\text{s.t. } -(K(B, C^T)w_1 + e_2 b_1) \geq e_2 - \xi_2, \xi_2 \geq 0$$

$$\min_{w_2, b_2, \xi_2} \frac{1}{2} \|K(B, C^T)w_2 + e_2 b_2\|^2 + c_2 e_1^T \xi_1 \quad (3)$$

$$\text{s.t. } (K(A, C^T)w_2 + e_1 b_2) \geq e_1 - \xi_1, \xi_1 \geq 0$$

其中 c_1 和 c_2 是惩罚参数, e_1 和 e_2 是两个全1列向量, ξ_1 和 ξ_2 是松弛变量。

测试样本离哪个超平面近,就属于哪一类别,若

$$K(x^T, C^T)w_r + b_r = \min_{i=1,2} |K(x^T, C^T)w_i + b_i| \quad (4)$$

则 x 属于第 r 类,其中 $r \in \{1,2\}$ 。

1.2 人工鱼群算法

人工鱼群算法是一种基于群体智能的优化算法,其灵感来自于鱼群行为。在 AFSA 中,每条人工鱼 (artificial fish, AF) 根据其当前状态和周围环境状态调整其行为。在每次迭代过程中,人工鱼通过觅食、聚群、追尾和随机这4种行为来更新自己。

假设 $X_i = (x_1, x_2, \dots, x_n)$ 是人工鱼 AF_i 的当前状态; $Y_i = f(X_i)$ 是 X_i 的适应度函数; Visual 是人工鱼的视野范围; try_number 是觅食行为的最大尝试次数; δ 是拥挤度因子; Step 表示人工鱼移动的步长; n_f 表示视野内的人工鱼数目。对于人工鱼 AF_i , 其视野内的一个随机状态 X_j 可以用式 (5) 来表示,其中 Rand() 是介于0和1之前的随机数。当满足更新条件时, AF_i 用式 (6) 更新其状态。

$$X_j = X_i + \text{Visual} \cdot \text{Rand}() \quad (5)$$

$$X_i^{t+1} = X_i^t + \frac{X_j - X_i^t}{\|X_j - X_i^t\|} \text{Step} \cdot \text{Rand}() \quad (6)$$

人工鱼的4种行为描述如下:

1) 觅食行为: AF_i 用式 (5) 在其视野范围内随机选择一个状态 X_j 。如果在求极大值问题中, $Y_i < Y_j$ (在求极小值时为 $Y_i > Y_j$, 它们之间可以转换), 则根据式 (6) 向 X_j 移动一步。否则, 再重新

随机选择状态 X_j , 判断是否满足要求 $Y_i < Y_j$ 。如果反复尝试 try_number 次后, 仍不满足要求则执行随机行为。

2) 聚群行为: 假设 X_c 是视野范围内的中心位置, 如果 $Y_c > Y_i$ 且 $Y_c/n_f > \delta \cdot Y_i$, 表明伙伴中心有更多食物且不拥挤, 则根据式 (6) 向 X_c 移动一步, 否则执行觅食行为。

3) 追尾行为: 假设 X_b 是视野范围内找到的最好位置, 如果 $Y_b/n_f > \delta \cdot Y_i$, 表明位置 X_b 有更多食物且不拥挤, 则根据式 (6) 向 X_b 的方向前进一步, 否则执行觅食行为。

4) 随机行为: AF_i 在其视野范围内随机选择一个位置, 然后向该方向移动一步, 它是觅食行为的一个缺省行为。

以上 4 种行为在不同条件下会相互转换, 人工鱼会选择合适的行为以找到更优解的位置。

2 AFSA-TWSVM 算法

AFSA-TWSVM 的核心思想是通过 AFSA 找到 TWSVM 的最优参数。人工鱼 AF_i 的位置 $X_i = (x_1, x_2, \dots, x_n)$ 对应 TWSVM 的一组参数, 该位置是一个向量, 其维数代表参数的个数。AFSA-TWSVM 的目标函数是 TWSVM 的分类准确率, AFSA 找到的最佳位置就是 TWSVM 的最佳参数。

AFSA-TWSVM 的算法步骤如下:

1) 初始化设置, 包括人工鱼群规模 N 、最大迭代次数 K 、每个人工鱼的初始位置、步长 Step 、视野 Visual 、尝试次数 try_number 、拥挤度因子 δ 和 TWSVM 的参数上下限。

2) 以人工鱼的位置为参数计算孪生支持向量机的分类准确率, 将分类准确率作为目标函数进行寻优。得到每个人工鱼的适应度值, 并记录全局最优的人工鱼的位置。

3) 每个人工鱼都执行聚群行为和追尾行为, 并判断个体是否得到改善。如果得到改善, 则选择一个更优行为执行, 否则执行觅食行为。

4) 执行人工鱼选择的行为, 更新每条人工鱼的位置。

5) 更新全局最优人工鱼的状态。

6) 判断是否达到最大迭代次数, 如果是则输出最优解和对应的参数组合, 否则迭代次数加 1, 并跳转到 2)。

AFSA-TWSVM 的流程图如图 1 所示, 通过

这个流程图可以直观地看出本文所提出算法的过程。

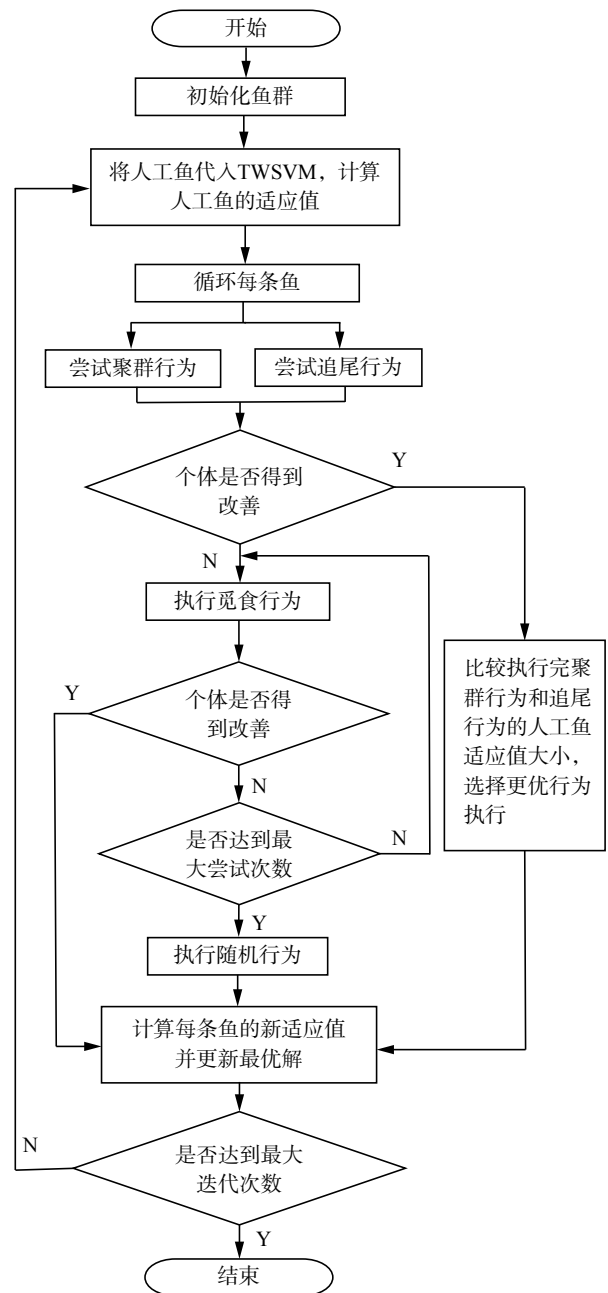


图 1 AFSA-TWSVM 流程图

Fig. 1 The flowchart of AFSA-TWSVM

3 实验结果及分析

为了验证基于人工鱼群算法的孪生支持向量机的有效性, 选取 UCI 机器学习数据库中的 8 个常用数据集来进行测试, 数据集的基本特征如表 1 所示。为了得到更可靠的测试结果, 本文使用十折交叉验证方法。实验在 Matlab 环境下实现, 硬件配置为 Windows 10 操作系统, 4 GB 内存, 250 GB 硬盘, i5-2400 和 3.1 GHz 主频 CPU 的计算机。

表1 实验数据集的特征
Table 1 Data feature of the datasets

数据集	样本数	特征数
Sonar	208	60
Haberman	306	3
Heart-c	303	13
WPBC	198	34
Statlog	270	13
Votes	435	16
Ionosphere	351	34
Liver	345	6

在实验中,选取未优化参数的孪生支持向量、基于网格搜索的孪生支持向量机(Grid-TWSVM)和基于粒子群算法的孪生支持向量机(PSO-TWSVM)作为对比实验,与AFSA-TWSVM进行比较。由于Gauss径向基函数具有较好的适应性,故选取该函数作为核函数进行实验。

孪生支持向量机的惩罚参数 c_1 、 c_2 和高斯核函数相关参数 $R(R=1/\sigma^2)$ 的取值范围为 $(0,10]$ 。人工鱼群算法的最大尝试次数try_number为5,视野Visual为0.5,步长Step为0.1,拥挤度因子 δ 为0.618。粒子群算法中速度上界为1,速度下界为-1,学习因子 η_1 和 η_2 都为2。AFSA和PSO的种群规模 N 为10,最大迭代次数为50。

表2给出了4种算法的十折交叉验证的平均分类准确率和标准差,表3给出了在不同数据集上的运行时间,图2是分类准确率效果图。为了更好地体现算法的优越性,用加粗数字代表特定数据集下的最高分类准确率。同时,Mean acc给出了所有算法的平均准确率,“W-T-L”(win-tie-loss)比较了AFSA-TWSVM与其他算法性能。从中可以看到,本文所提出的AFSA-TWSVM在大多数数据集中都获得了最佳分类准确率,并且平均分类准确率也是最好的。因此可以得出结论,与传统的分类算法相比,AFSA-TWSVM具有更好的分类性能。这是因为AFSA具有较好的全局搜索能力,能避免自己过早地陷入局部最优,从而能更精确地找到参数。由于有更好的参数选择,AFSA-TWSVM提高了TWSVM的分类准确率。

表2 各算法在不同数据集上的分类准确率
Table 2 Classification Accuracy of Different Algorithms on Different Datasets

数据集	准确率±标准差/%			
	AFSA-TWSVM	PSO-TWSVM	Gird-TWSVM	TWSVM
Sonar	94.73 ±3.34	91.83±3.72	88.42±7.80	86.64±7.31
Haberman	80.70 ±3.72	80.34±6.14	79.40±4.03	73.62±4.61
Heart-c	85.09 ±7.38	84.46±5.80	84.09±7.77	79.82±4.90
WPBC	84.86 ±3.11	82.27±6.54	82.39±7.05	78.65±8.81
Statlog	83.70±5.79	84.81 ±5.60	82.96±6.67	79.26±6.02
Votes	95.62 ±2.18	94.72±3.94	94.23±3.47	93.54±3.73
Ionosphere	96.88 ±2.32	96.01±3.43	96.54±3.80	94.31±4.21
Liver	78.00±5.42	78.25 ±5.29	77.92±9.46	73.90±6.24
Mean acc	87.45	86.59	85.75	82.47
W-T-L		6-0-2	8-0-0	8-0-0

表3 不同数据集上的运行时间

Table 3 Running time on different data sets min

数据集	AFSA-TWSVM	PSO-TWSVM	Gird-TWSVM
Sonar	34.01	15.13	20.97
Haberman	40.03	18.76	27.83
Heart-c	73.47	21.01	26.10
WPBC	34.07	19.98	23.07
Statlog	30.17	18.47	33.86
Votes	105.05	46.03	45.69
Ionosphere	79.21	30.91	35.84
Liver	82.42	25.27	30.99

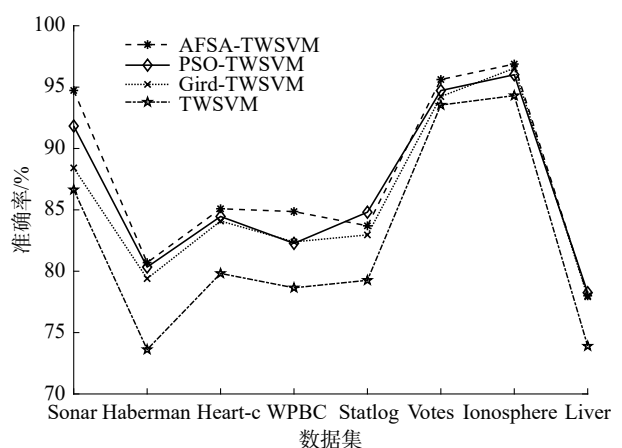


图2 实验结果

Fig. 2 The effect diagram of experimental results

4 结束语

与传统的分类算法相比, TWSVM 有许多特有的优势, 但也存在着一些不足。本文针对 TWSVM 参数难以设置的问题, 提出了一种基于人工鱼群算法的孪生支持向量机 (AFSA-TWSVM), 利用 AFSA 优化参数, 避免了参数选择的盲目性。通过选取 UCI 数据集进行实验, 与 TWSVM、Grid-TWSVM 和 PSO-TWSVM 相比, AFSA-TWSVM 有更好的分类性能。但是不足之处在于, 人工鱼群算法寻优速度较慢, 如何进一步提高算法的寻优速度是下一步需要研究的问题。

参考文献:

- [1] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述[J]. *电子科技大学学报*, 2011, 40(1): 1-10.
DING Shifei, QI Bingjuan, TAN Hongyan. An overview on theory and algorithm of support vector machines[J]. *Journal of University of Electronic Science and Technology of China*, 2011, 40(1): 1-10.
- [2] CORTES C, VAPNIK V. Support-vector networks[J]. *Machine learning*, 1995, 20(3): 273-297.
- [3] CHEN Zhensong, QI Zhiqian, WANG Bo, et al. Learning with label proportions based on nonparallel support vector machines[J]. *Knowledge-based systems*, 2017, 119: 126-141.
- [4] CHEN Zhenyu, FAN Zhiping. Distributed customer behavior prediction using multiplex data: a collaborative MK-SVM approach[J]. *Knowledge-based systems*, 2012, 35: 111-119.
- [5] MORAES R, VALIATI J F, NETO W P G. Document-level sentiment classification: an empirical comparison between SVM and ANN[J]. *Expert systems with applications*, 2013, 40(2): 621-633.
- [6] ZHANG Xiekai, DING Shifei, XUE Yu. An improved multiple birth support vector machine for pattern classification[J]. *Neurocomputing*, 2017, 225: 119-128.
- [7] DING Shifei, ZHANG Xiekai, AN Yuexuan, et al. Weighted linear loss multiple birth support vector machine based on information granulation for multi-class classification[J]. *Pattern recognition*, 2017, 67: 32-46.
- [8] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers[J]. *Neural processing letters*, 1999, 9(3): 293-300.
- [9] FUNG G, MANGASARIAN O L. Proximal support vector machine classifiers[C]//Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, 2001: 77-86.
- [10] MANGASARIAN O L, WILD E W. Multisurface proximal support vector machine classification via generalized eigenvalues[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2006, 28(1): 69-74.
- [11] JAYADEVA, KHEMCHANDANI R, CHANDRA S. Twin support vector machines for pattern classification[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2007, 29(5): 905-910.
- [12] HUANG Huajuan, WEI Xiuxi, ZHOU Yongquan. Twin support vector machines: a survey[J]. *Neurocomputing*, 2018, 300: 34-43.
- [13] 丁世飞. 孪生支持向量机: 理论、算法与拓展[M]. 北京: 科学出版社, 2017: 16-31.
DING Shifei. Twin support vector machine: theory, algorithm and extension[M]. Beijing: Science Press, 2017: 16-31.
- [14] KUMAR M A, GOPAL M. Application of smoothing technique on twin support vector machines[J]. *Pattern recognition letters*, 2008, 29(13): 1842-1848.
- [15] KUMAR M A, GOPAL M. Least squares twin support vector machines for pattern classification[J]. *Expert systems with applications*, 2009, 36(4): 7535-7543.
- [16] KUMAR M A, KHEMCHANDANI R, Gopal M, et al. Knowledge based Least Squares Twin support vector machines[J]. *Information sciences*, 2010, 180(23): 4606-4618.
- [17] MIR A, NASIRI J A. KNN-based least squares twin support vector machine for pattern classification[J]. *Applied intelligence*, 2018, 48(12): 4551-4564.
- [18] CHEN Sugan, WU Xiaojun. A new fuzzy twin support vector machine for pattern classification[J]. *International journal of machine learning and cybernetics*, 2018, 9(9): 1553-1564.
- [19] CHEN Xiaobo, YANG Jian, YE Qiaolin, et al. Recursive projection twin support vector machine via within-class variance minimization[J]. *Pattern recognition*, 2011, 44(10/11): 2643-2655.
- [20] XIE Xiaomin. Improvement on projection twin support vector machine[J]. *Neural computing and applications*, 2018, 30(2): 371-387.
- [21] XU Zongben, DAI Mingwei, MENG Deyu. Fast and efficient strategies for model selection of Gaussian support vector machine[J]. *IEEE transactions on systems, man, and cybernetics, part B (cybernetics)*, 2009, 39(5): 1292-1307.
- [22] DING Shifei, YU Junzhao, HUANG Huajuan, et al. Twin support vector machines based on particle swarm optimization[J]. *Journal of computers*, 2013, 8(9): 2296-2303.

- [23] DING Shifei, WU Fulin, NIE Ru, et al. Twin support vector machines based on quantum particle swarm optimization[J]. *Journal of software*, 2013, 8(7): 1743–1750.
- [24] ZHAI Shijun, PAN Juan, LUO Hongwei, et al. A new sense-through-foliage target recognition method based on hybrid particle swarm optimization-based wavelet twin support vector machine[J]. *Measurement*, 2016, 80: 58–70.
- [25] SARTAKHTI J S, AFRABANDPEY H, SARAEE M. Simulated annealing least squares twin support vector machine (SA-LSTSVM) for pattern classification[J]. *Soft computing*, 2017, 21(15): 4361–4373.
- [26] 李晓磊, 邵之江, 钱积新. 一种基于动物自治体的寻优模式: 鱼群算法 [J]. *系统工程理论与实践*, 2002, 22(11): 32–38.
- LI Xiaolei, SHAO Zhijiang, QIAN Jixin. An optimizing method based on autonomous animats: fish-swarm algo-

rithm[J]. *Systems engineering-theory & practice*, 2002, 22(11): 32–38.

作者简介:



李景灿, 男, 1995 年生, 硕士研究生, 主要研究方向为支持向量机和机器学习。



丁世飞, 男, 1963 年生, 教授, 博士生导师, 主要研究方向为人工智能、机器学习、模式识别、数据挖掘。主持国家重点基础研究计划 (973 计划) 课题 1 项、国家自然科学基金面上项目 2 项。出版专著 4 部, 发表学术论文 200 余篇。