

文章编号: 2095-2163(2023)08-0197-05

中图分类号: TP181

文献标志码: A

基于改进蜂群算法优化的支持向量机研究与应用

朱范炳, 陈泽, 张翔

(信阳学院 大数据与人工智能学院, 河南 信阳 464000)

摘要: 支持向量机是一种应用广泛的机器学习方法,其分类性能主要取决于相关模型参数的选择。本文提出了一种改进的人工蜂群算法来优化支持向量机的参数,并将其应用于人体活动数据识别。在标准数据集上测试,与基本人工群体算法、遗传算法和粒子群算法等优化算法相比,改进蜂群算法优化的支持向量机可以获得更高的分类准确率。验证了改进人工蜂群算法的有效性。利用改进人工蜂群算法优化的支持向量机对人体活动数据进行分类识别,结果显示该方法具有较高的分类准确率,说明本文所提方法具有实用性。

关键词: 支持向量机; 改进的人工蜂群算法; 参数优化; 人体活动识别

Research and application of Support Vector Machine based on improved Artificial Bee Colony algorithm

ZHU Fanbing, CHEN Ze, ZHANG Xiang

(School of Big Data and Artificial Intelligence, Xinyang College, Xinyang Henan 464000, China)

[Abstract] Support Vector Machine is the widely used machine learning method in which the classification performance mainly depends on the selection of relevant model parameters. In this paper, an improved Artificial Bee Colony algorithm is proposed to optimize the parameters of Support Vector Machine, and is applied to human activity data recognition. When tested on the standard data set, the SVM optimized by the improved Artificial Bee Colony algorithm can obtain higher classification accuracy compared with the basic artificial population algorithm, genetic algorithm and particle swarm optimization. The effectiveness of the improved Artificial Bee Colony algorithm is verified. The Support Vector Machine optimized by improved Artificial Bee Colony algorithm is used to classify and recognize human activity data. The results show that the proposed method has a high classification accuracy, which indicates that the proposed method is practical.

[Key words] Support Vector Machine; improved Artificial Bee Colony algorithm; parameter optimization; activity recognition

0 引言

支持向量机^[1](Support Vector Machine, SVM)是一种应用广泛的机器学习方法,其基础是Vapnik^[2]创立的统计学习理论。可以解决小样本和非线性的实际问题,现已广泛应用于智能故障诊断^[3]、图像识别^[4]、负荷预测^[5]等领域。在使用支持向量机进行分类时,模型参数的选择对分类准确率影响很大。如何选择模型参数以优化模型的性能是支持向量机研究领域的一个热点问题。目前常用的模型参数优化方法有遗传算法^[6]、高斯函数^[7]、模拟退火算法^[8]、粒子群优化算法^[9]等。在优化过程中,上述方法都不同程度地进入局部最优解,导致

分类模型的性能不能最大化。文献[10]提出了一种多维更新的人工蜂群算法,该方法很好地保持了挖掘搜索宽度和深度上的平衡,有效地避免了问题陷入局部最优解,但求解方法采用全维搜索策略会大大增加时间成本。

基于以上分析,本文提出一种进一步改进人工蜂群算法的方法,通过选择最优的搜索方向,在保证搜索精度的同时,能够降低算法的时间复杂度。基于RBF核函数,采用改进的人工蜂群算法对支持向量机模型参数进行优化。本文首先通过在标准数据集上的分类实验,验证改进人工蜂群算法优化支持向量机的有效性;然后通过人体活动数据分类识别,说明所提方法的实用性;最后得出结论。

作者简介: 朱范炳(1990-),男,硕士,讲师,主要研究方向:智能计算、复杂网络理论;陈泽(1993-),男,硕士,助教,主要研究方向:数字图像处理;张翔(1991-),男,硕士,讲师,主要研究方向:智能信息服务系统。

收稿日期: 2023-01-05

哈尔滨工业大学主办 ◆ 科技创新与应用

1 人工蜂群算法分析与改进

1.1 基本人工蜂群算法

人工蜂群算法^[11-12] (Artificial Bee Colony Algorithm, ABC) 是 Karaboga 等学者于 2005 年提出的一种仿生计算算法, 主要应用于函数极值优化问题。该算法模仿了蜂群的觅食行为, 并将蜜蜂分为 3 种不同的工作类型: 采蜜蜂、观察蜂和侦察蜂。采蜜蜂、观察蜂分别占蜜蜂总数的一半, 一个蜜源对应一只蜜蜂。蜜源优劣由适应度值评价, 一个蜜源有最大开采次数。算法通过迭代搜索蜜源, 进行目标问题解的寻优, 能够以较大的概率找到全局最优解。

人工蜂群算法的基本原理: 设有 N 个蜜源 $\{x_1, x_2, \dots, x_N\}$, 每个蜜源 $x_i (i = 1, 2, \dots, N)$ 有 D 个分量, 即待优化问题的解空间包含 N 个可行解, 每个可行解是 D 维向量。设定蜂群循环搜索的最大次数和每个蜜源的可重复开采次数, 同一蜜源开采超过可重复开采次数则放弃该蜜源。人工蜂群算法包括以下阶段:

(1) 蜂群的初始化阶段: 利用式(1)对任一解的任一分量进行全局随机初始化:

$$x_{id} = x_{idmin} + rand(0, 1)(x_{idmax} - x_{idmin}) \quad (1)$$

其中, x_{idmin} 和 x_{idmax} 分别表示可行解空间第 d 维分量的上下限, $rand(0, 1)$ 为 $[0, 1]$ 之间的随机数。

(2) 采蜜蜂搜索阶段: 采蜜蜂在初始蜜源邻域, 通过式(2)搜索产生一个新解, 作为候选蜜源进行开采:

$$v_{id} = x_{id} + rand(-1, 1)(x_{id} - x_{jd}) \quad (2)$$

其中, $j \in \{1, 2, \dots, N\}, j \neq i$ 表示在 N 个蜜源中随机选取一个不同于 x_i 的蜜源, $rand(-1, 1)$ 是在 $[-1, 1]$ 之间均匀分布的随机数, 决定采蜜蜂更新位置的扰动幅度。计算新解的适应度 fit_i 评价蜜源质量, 在 v_i 和 x_i 之中采用贪婪策略进行选择。

(3) 观察蜂跟随阶段: 所有采蜜蜂完成搜索后, 采蜜蜂会把解的信息及适应度分享给观察蜂。观察蜂通过选择概率 P_i 决定每只采蜜蜂被跟随的概率:

$$P_i = \frac{fit_i}{\sum_{k=1}^N fit_k} \quad (3)$$

或者

$$P_i = \frac{fit_i}{\max(fit_k)} \quad (4)$$

观察蜂通过轮盘赌策略选择采蜜蜂跟随。如果

采蜜蜂对应蜜源的选择概率值较大, 则会被更多的观察蜂跟随, 即适应度较大的蜜源附近会有更多的观察蜂进行搜索, 蜜源对应解的邻域搜索范围更广。若新解的适应度比之前的好, 观察蜂将会更新解; 反之, 观察蜂会将之前的解保留, 同时解的迭代搜索次数会加 1。

(4) 侦察蜂阶段: 所有观察蜂完成跟随搜索后, 如果某一蜜源在被搜索最大开采次数后仍没有被更新, 则认为该蜜源已被开采枯竭, 对应的解陷入局部最优。相应的采蜜蜂和观察蜂则会放弃该蜜源, 转换为侦察蜂模式, 按照式(1)进行全局随机搜索新蜜源。这是人工蜂群算法跳出局部最优的有效手段。然后返回到采蜜蜂的搜索阶段, 3 种蜜蜂依次工作, 重复循环搜索, 最终找到目标问题的最优解。

1.2 改进人工蜂群算法

基本人工蜂群算法中, 采蜜蜂和观察蜂随机选择蜜源的一个维度, 按照式(2)进行搜索。如果在这个维度上有更好的解, 但下一次迭代随机选择另外一个维度, 这可能会导致无法进一步开发当前维度更优的解, 后因达到最大开采次数而放弃这个蜜源。这会使算法错过很多寻找全局最优解的机会, 增加算法的收敛时间, 也会影响最终解的精度。

为此, 本文提出了一种改进的多维搜索策略。该方法在第一次迭代时搜索蜜源的每个维度进行贪婪选择。如果当前维度选择成功, 则在保留当前维度更新的基础上, 将该维度保存到外部文档 Dim 中, 然后进行下一个维度的选择。如果选择失败, 更新将不会被保留, 外部文档也不会存储该维度。结合上述想法, 本文将式(2)中随机更新一个维度替换为同时更新 Dim 中的维度, 即可得到新的更新公式如下:

$$\begin{cases} v_{i1} = x_{i1} + rand(-1, 1)(x_{i1} - x_{j1}) \\ v_{i2} = x_{i2} + rand(-1, 1)(x_{i2} - x_{j2}) \\ \dots\dots \\ v_{im} = x_{im} + rand(-1, 1)(x_{im} - x_{jm}) \end{cases} \quad (5)$$

其中, m 表示 Dim 中保存维度的个数。

每次搜索迭代搜索都会进行贪婪选择, 将每个蜜源中有价值的维度存储到 Dim 中。在下一次迭代中, 对 Dim 中存储的维度执行上述操作。多次迭代后, Dim 存储维度的数量将减少, 说明算法进化搜索的方向更加明确, 直至算法达到稳定收敛。

2 支持向量机参数选择

支持向量机在面对非线性分类问题时, 通过高

维空间变换将非线性分类问题转化为高维空间的线性分类问题,并引入拉格朗日函数来解决不等式约束下二次函数的极值问题。在使用支持向量机进行数据分类之前,应选择核函数类型、惩罚因子 C 和核函数参数^[13]。本文采用 RBF 径向基函数作为核函数,以 RBF 核函数中的宽度参数 g 和惩罚因子 C 为优化目标。为了说明惩罚因子 C 和参数 g 对模型性能的影响,表 1 和表 2 给出了不同参数下数据分类准确率。

表 1 $C=2$ 分类结果随 g 值的变化

Tab. 1 Change of classification results with g when $C=2$

g	分类准确率/%
0.1	51.43
1	35.24
10	17.62
100	17.62

表 2 $g=2$ 分类结果随 C 值的变化

Tab. 2 Change of classification results with C when $g=2$

C	分类准确率/%
0.1	17.62
1	26.67
10	27.14
100	27.14

惩罚因子 C 起到控制错误子样本惩罚的作用,从而实现错误子样本比例与算法复杂度之间的权衡。当 C 较小时,支持向量机对误差的容忍度较高,模型的精度会降低,但泛化能力增强。当 C 较大时,模型的精度提高,但复杂度增加,泛化能力也降低。参数 g 决定了数据样本在高维特征空间中的复杂度,称为空间维数。该维数决定在空间中可以构造的线性分类曲面的最大维数。维数越高,线性分类面越复杂,经验风险越小,置信范围越大,反之亦然。从表 1 和表 2 可以看出,惩罚因子 C 和参数 g 对分类准确率有显著影响。因此,为了得到性能较好的支持向量机,需要选择合适的宽度参数 g ,并在确定的特征空间中寻找合适的惩罚因子 C ,使模型的拟合能力和泛化能力得到最佳的结合。

3 改进人工蜂群算法优化支持向量机参数

利用智能计算方法优化支持向量机需要设置适应度函数和寻优解空间。由于使用算法对支持向量机进行优化的目的是为了获得更高的分类准确率,所以本文用分类准确率来设计算法的适应度函数;由前文的分析可知,惩罚因子 C 和宽度参数 g 会对

支持向量机的分类性能产生很大的影响。因此,设定惩罚因子 C 和宽度参数 g 为优化参数。

为了验证本文改进的人工蜂群算法的有效性和优越性,使用 UCI 标准数据集中 3 组数据集进行支持向量机的训练和测试验证。使用的数据集名称、样本规模、维度信息,见表 3。本文设计了不同改进方法的人工蜂群算法优化支持向量机的纵向比较实验和不同算法优化支持向量机的横向比较实验。

表 3 UCI 测试数据集信息

Tab. 3 The information of UCI test data sets

数据集	数据维度	训练样本数	测试样本数
Wine	13	120	58
BreastTissue	9	80	26
Glass	9	140	74

3.1 不同方法改进蜂群算法优化支持向量机参数

分别采用本文改进的人工蜂群算法 (IABC)、文献[10]提出的多维更新的人工蜂群算法 (MABC) 和基本人工蜂群算法 (ABC) 对支持向量机模型参数进行优化。实验结果由 10 次实验计算平均值所得。表 4 记录不同人工蜂群算法优化的支持向量机在不同数据集下的分类准确率,表 5 记录对应算法的运行时间。由结果数据可以看出,本文改进人工蜂群算法优化的支持向量机在保证分类准确率的同时,能够有效降低算法收敛时间。

表 4 改进蜂群算法优化支持向量机分类准确率

Tab. 4 Classification accuracy of SVM optimized by IABC %

算法	数据集		
	Wine	BreastTissue	Glass
ABC	94.8	67.7	64.2
MABC	96.5	73.1	66.8
IABC	95.9	72.5	67.6

表 5 改进蜂群算法优化支持向量机运行时间

Tab. 5 Running time of SVM optimized by IABC s

算法	数据集		
	Wine	BreastTissue	Glass
ABC	93.26	82.89	180.57
MABC	176.16	151.02	332.35
IABC	143.84	120.27	265.44

3.2 不同算法优化支持向量机参数

为了进一步验证改进算法的性能,利用 IABC 算法与网格搜索算法 (GS)、遗传算法 (GA) 和粒子群算法 (PSO) 优化支持向量机的参数。分别进行 10 次实验计算结果平均值。表 6 记录的是 4 种算法在不同数据集下的分类准确率。由结果数据可以

看出,对于不同样本规模和维度的数据集,IABC 优化的 SVM 模型都比其他算法具有更高的分类准确率。因此,采用 IABC 优化的 SVM 模型将局部最优解的搜索与全局最优解的搜索结合起来,具有较好的收敛性。

表 6 不同算法优化支持向量机分类准确率

Tab. 6 Classification accuracy of SVM optimized by different algorithms %

算法	数据集		
	Wine	BreastTissue	Class
IABC	95.9	72.5	67.6
GS	94.5	69.2	63.5
GA	93.7	63.5	62.7
PSO	93.1	65.3	63.1

4 实例分析

4.1 实验方案设计

应用改进蜂群算法优化的支持向量机对人体活动的实测数据进行分类,验证改进算法的实用性。数据采集过程中参与者需佩戴带有加速度、温度和高度传感器的智能手表。从加速度传感器采集的数据包括 X 、 Y 和 Z 方向的加速度。温度和高度传感器的数据包括参与者的体表温度和佩戴手表距地面的距离。采样频率设置为 52 Hz,采集 7 种不同的人体活动数据,数据的详细内容见表 7。

表 7 人体活动数据信息

Tab. 7 Activity data information

活动类别	具体描述
1	坐,使用电脑学习或办公
2	站立、行走、上下楼梯
3	静止站立
4	平面行走
5	上下楼梯
6	边走边与人交谈
7	静止站立与人交谈

4.2 实验与结果分析

在采集的人体活动数据中每类抽取 100 组样本,共计 700 组;在支持向量机训练和测试中,随机抽取 490 个样本作为训练集,其余 210 个样本作为测试集。本文主要运用 ABC 和 IABC 算法对 SVM 模型参数进行优化,从而进行人体活动识别。实验中设置蜂群算法的种群规模为 20,最大迭代次数为 500。ABC 和 IABC 算法优化的 SVM 模型分类准确率的迭代变化曲线分别如图 1 和图 2 所示。

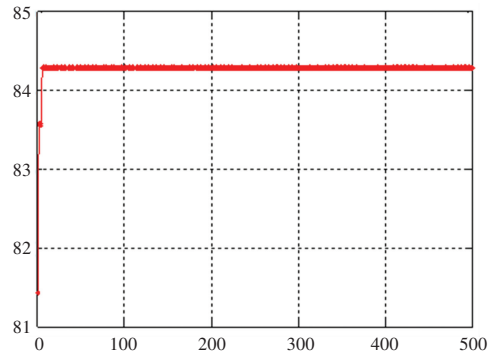


图 1 ABC-SVM 分类准确率变化曲线

Fig. 1 Classification accuracy curve of ABC-SVM

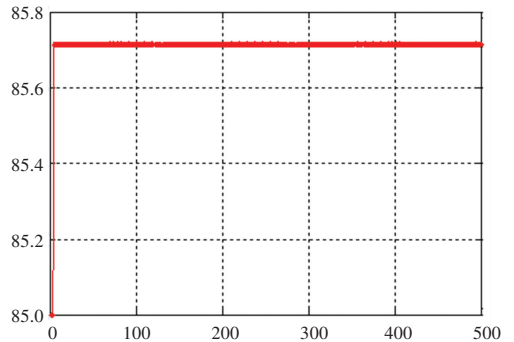


图 2 IABC-SVM 分类准确率变化曲线

Fig. 2 Classification accuracy curve of IABC-SVM

由图 1 和图 2 可见,在算法迭代次数均为 500 的情况下,ABC 算法优化的 SVM 模型(ABC-SVM)分类准确率收敛速度快,最高可达 84.31%;IABC 算法优化的 SVM 模型(IABC-SVM)分类准确率收敛速度较 ABC-SVM 慢,但分类准确率最高可达 85.73%,有显著提高。对上述实验数据分别采用 SVM、ABC-SVM、IABC-SVM 进行 10 次分类实验,计算分类准确率平均值记录于表 8 中。

表 8 不同算法模型分类准确率

Tab. 8 Classification accuracy under different algorithms

模型	分类准确率 /%
SVM	81.13
ABC-SVM	83.29
IABC-SVM	85.63

5 结束语

支持向量机模型参数是影响分类精度的重要因素。针对目前支持向量机训练算法计算量大、运行时间长、分类准确率待提高的问题,本文提出了一种改进的人工蜂群算法对支持向量机参数优化的方法。通过标准数据集的实验表明,改进的人工蜂群算法优化的支持向量机模型提高了分类准确率,降

(下转第 204 页)