

# 融合渐近性的灰狼优化支持向量机模型

武玉坤 肖杰 李伟 楼吉林

浙江工业大学计算机科学与技术学院 杭州 310023

(051092296wyk@163.com)



**摘要** 大数据的发展对数据分类领域的分类准确性有了更高的要求;支持向量机(Support Vector Machine, SVM)的广泛应用需要一种高效的方法来构造一个分类能力强的 SVM 分类器;SVM 的核函数参数与惩罚因子以及特征子集对预测模型的复杂度和预测精度有着重要影响。为提高 SVM 的分类性能,文中将 SVM 的渐近性融合到灰狼优化(Grey Wolf Optimization, GWO)算法中,提出了新的 SVM 分类器模型,该模型对 SVM 的参数与数据的特征子集同时进行优化,融合 SVM 渐近性的新灰狼个体将灰狼优化算法的搜索空间导向超参数空间中的最佳区域,能够更快地获得最优解;此外,将获得的分类准确率、所选特征个数和支持向量个数相结合,提出了一种新的适应度函数,新的适应度函数与融合渐近性的灰狼优化算法将搜索引向最优解。采用 UCI 中的多个经典数据集对所提模型进行验证,将其与网格搜索算法、未融合渐近性的灰狼优化算法以及其他文献中的方法进行对比,其分类准确率在不同数据集上均有不同程度的提升。实验结果表明,所提算法能找到 SVM 的最优参数与最小特征子集,具有更高的分类准确率和更短的平均处理时间。

**关键词:** 灰狼优化算法;参数优化;特征选择;渐近性;支持向量机

中图法分类号 TP391

## Support Vector Machine Model Based on Grey Wolf Optimization Fused Asymptotic

WU Yu-kun, XIAO Jie, Wei William LEE and LOU Ji-lin

College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310023, China

**Abstract** The development of big data requires higher accuracy of data classification. The wide application of support vector machine (SVM) requires an efficient method to construct an SVM classifier with strong classification ability. The kernel parameter, penalty parameter and feature subsets of dataset have an important impact on the complexity and prediction accuracy of the model. In order to improve the classification performance of SVM, the asymptotic of SVM was integrated into the gray wolf optimization (GWO) algorithm, and a new SVM classifier model was proposed. The model optimizes feature selection and parameter optimization of SVM at the same time. The new grey wolf individual integrating the asymptotic property of SVM directs the search space of grey wolf optimization algorithm to the optimal region in super-parameter space, and can obtain the optimal solution faster. In addition, a new fitness function, which combines the classification accuracy obtained from the method, the number of chosen features and the number of support vectors, was proposed. The new fitness function and GWO fused asymptotic lead the search to the optimal solution. This paper used several classical datasets on UCI to verify the proposed model. Compared with the grid search algorithm, the gray wolf optimization algorithm without asymptotic convergence and other methods in the literature, the classification accuracy of the proposed algorithm has different degrees of improvement on different data sets. The experimental results show that the proposed algorithm can find the optimal parameters and the smallest feature subset of SVM, with higher classification accuracy and less average processing time.

**Keywords** Gray wolf optimization algorithm, Parameters optimization, Feature selection, Asymptotic, Support vector machines

## 1 引言

支持向量机(Support Vector Machine, SVM)是一种被广泛使用的机器学习分类与回归算法,由 Vapnik 于 20 世纪的最后 10 年提出<sup>[1]</sup>。SVM 已被应用于许多领域,如生物识

别<sup>[2]</sup>、生物信息学<sup>[3]</sup>和化学信息学<sup>[4]</sup>。在 SVM 分类中,训练数据用于构建分类模型,分类器对未知样本进行分类。在使用 SVM 时,有两个问题需要解决,即如何选择最优的特征子集和最优的参数。特征子集与 SVM 的惩罚参数和内核参数对 SVM 的性能有着重要影响。数据集集中的有些特征是冗余

到稿日期:2019-01-12 返修日期:2019-06-10 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(61502422);浙江省自然科学基金(LY18F020028)

This work supported by the National Natural Science Foundation of China (61502422) and Natural Science Foundation of Zhejiang Province (LY18F020028).

通信作者:李伟(lewei@zjut.edu.cn)

和不相关的,丢弃这些无用的特征可以提高计算效率;惩罚参数可以权衡最小化训练误差和最大分类间隔,内核参数定义了输入特征空间到一个高维特征空间的非线性映射。同时,特征子集的选择将会影响 SVM 参数的选择,反之亦然。因此,求解最优特征子集和 SVM 参数必须同时进行。

群智能算法在解决优化问题方面具有强大的竞争力<sup>[5-6]</sup>,灰狼优化(GWO)算法<sup>[7]</sup>是 Mirjalili 于 2014 年提出的一种新型群智能优化算法。GWO 具有算法结构简单、需调参数少、收敛速度快等优点;更重要的是,该算法表现出了非常强的探索与利用能力<sup>[8]</sup>。许多研究者已经把该算法应用于很多优化问题,如训练多层传感器、经济调度指派<sup>[9]</sup>、车间调度<sup>[10]</sup>、约束优化<sup>[11]</sup>等。

许多研究根据经验尝试有限数量的值来选择 SVM 参数,以获取最小的误分类率。此过程需要在搜索空间上进行详尽搜索才能找到可行的解决方案,这是 SVM 的一大挑战。Friendrichs 等使用网格搜索 SVM 参数空间的最佳参数<sup>[12]</sup>,在此解决方案中,参数在参数空间内以固定的步长变化,使用不同的性能度量来评估每个参数组合。但是,该参数优化解决方案仅适用于极少数参数的调整,而且复杂、耗时。鉴于此,后续有研究者使用启发式进化算法来搜索 SVM 参数的最佳值以构建具有高预测精度和稳定性的分类模型。在以往的研究中,很多工作都是围绕特征子集的选择或者 SVM 参数优化中的某一方面来进行的。Huang 等利用 GA(遗传算法)对 SVM 参数进行优化,并依据精确率与权重惩罚因子提出了一种新的适应度评估函数<sup>[13]</sup>;Wu 等提出了一种混合遗传算法(GA)用于 SVM 参数优化,以预测最大值电力日负荷<sup>[14]</sup>;Zhang 等用 ACO 搜索最优 SVM 参数来优化支持向量机<sup>[6]</sup>,提升了性能且耗时短;Chen 等使用 PSO 优化 SVM 参数,而且为了降低运算复杂度和提高搜索速度,采用了类内平均距离与类外中心距离之差作为适应度函数<sup>[15]</sup>;Guo 等使用免疫算法优化 SVM 参数和特征权重,将优化模型应用于 EEG 信号中的意识任务分类<sup>[16]</sup>;Tharwat 使用蝙蝠算法来优化 SVM 参数,并且验证了 RBF 核函数具有比多项式函数更好的分类性能<sup>[17]</sup>。优化 SVM 参数的其他启发式算法还有鲸鱼优化算法<sup>[18]</sup>、萤火虫优化算法<sup>[19]</sup>、布谷鸟搜索算法<sup>[20]</sup>和引力搜索算法<sup>[21]</sup>。

文献[22-25]的研究主要集中于特征子集的选择,未考虑 SVM 中参数的优化。近年来的研究趋势是通过在选择特征的同时优化参数来提高准确率。Huang 等<sup>[13]</sup>首先使用遗传算法进行特征选择与参数优化,相比网格优化取得了更高的准确率与更小的特征子集;随后,Lin 等利用 PSO 代替 GA,取得了更有竞争力的实验结果<sup>[26]</sup>;Zhou 等提出了用遗传算法优化 SVM 的医院住院费用的预测模型<sup>[27]</sup>,同时进行参数优化和特征选择,取得了较高的准确率;文献[28]提出的 MVO+SVM 优化采用了两种体系结构来比较多分类任务;文献[29-32]分别利用蚱蜢优化算法、蜂群优化算法、改进的鱼群优化算法和改进的烟花算法来同时进行参数优化与特征选择。但上述文献均未考虑支持向量机的渐近行为,只有文献[33-34]考虑了支持向量机的渐近行为,采用网格优化与遗传算法提高了分类准确率,同时缩短了处理时间。实验验证了遗传算法优于网格搜索。

本文的主要目标是采用灰狼优化算法融合支持向量机的

渐近行为来进行特征子集的选择与 SVM 参数的选择,以提高其分类准确率。首先介绍 SVM 特征子集选择与参数优化的相关工作;其次描述 SVM 的基本思想及其渐近行为和灰狼优化算法;然后提出融合渐近行为的灰狼优化特征子集选择与 SVM 参数优化方法;最后进行实验验证与讨论。

## 2 相关方法

### 2.1 SVM 的基本思想

支持向量机是一种基于结构风险最小化的被广泛使用的机器学习算法,能够解决小样本和非线性分类问题。SVM 的基本原理是寻找最优分割超平面来使问题变成线性可分。假设有  $n$  个训练样本集  $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 。其中,  $x_i$  是训练数据;  $y_i \in \{-1, 1\}$  是样本类别, +1 代表正例, -1 代表负例。若样本集线性可分,则存在超平面  $\mathbf{W}^T X + b = 0$  能对样本数据进行正确分类,其中  $\mathbf{W}$  是样本的权重向量,  $b$  是偏置,  $X$  是训练样本。若超平面已知,则样本空间中任意点  $x_i$  到超平面的几何距离可记为:

$$r = \frac{|\mathbf{W}^T X + b|}{\|\mathbf{W}\|} \quad (1)$$

这个距离  $r$  为最大类间隔。若超平面方程在最大类间隔的情况下可以正确地分离样本,则将最优分离平面的求解转换为以下目标函数和约束:

$$\min \frac{1}{2} \|\mathbf{W}\|^2 \quad (2)$$

$$\text{s. t. } y_i (\mathbf{W}^T X_i + b) - 1 \geq 0, \forall i = 1, 2, \dots, N$$

针对线性不可分数据,即存在离群点的情况下,为保证数据线性可分,引入松弛因子  $\xi_i \geq 0 (i = 1, 2, \dots, n)$ , 优化问题的目标函数被重写为:

$$\min \frac{1}{2} \|\mathbf{W}\|^2 + C \sum_{i=1}^N \xi_i \quad (3)$$

$$\text{s. t. } y_i (\mathbf{W}^T X_i + b) - 1 + \xi_i \geq 0, \forall i = 1, 2, \dots, N$$

其中,  $C$  代表惩罚因子,其可以平衡最大间隔与松弛变量。

上文讨论的都是线性数据。若数据是非线性可分的, SVM 使用核函数把数据映射到高维空间,在高维空间中可以找到一个超平面,从而使数据线性可分,其目标函数为:

$$\min \frac{1}{2} \|\mathbf{W}\|^2 + C \sum_{i=1}^N \xi_i \quad (4)$$

$$\text{s. t. } y_i (\mathbf{W}^T \phi(X_i) + b) - 1 + \xi_i \geq 0,$$

$$\forall i = 1, 2, \dots, N$$

其中,  $\phi(X_i)$  为非线性函数。

核函数被定义为非线性函数的点积,这样可以解决由高维投影导致的维数灾难问题。目前主要有 3 种常用的核函数: 1) 线性核函数:  $K(X_i, X_j) = \langle X_i, X_j \rangle$ ; 2) 径向基函数:  $K(X_i, X_j) = \exp(-\|X_i - X_j\|^2 / 2\sigma^2)$ ; 3) 多项式核函数:  $K(X_i, X_j) = (\langle X_i, X_j \rangle + c)^d$ 。

由于径向基核函数在近似任何非线性函数方面有着显著性能,而且有更少的参数进行优化,因此本文采用径向基核函数。

### 2.2 支持向量机的渐近行为

Keerthi 等于 2003 年提出了支持向量机的渐近行为<sup>[35]</sup>,对具有高斯核的支持向量机的渐近行为进行了理论分析与数学推理;把  $\log C$  和  $\log \sigma^2$  作为超参数空间的参数,在参数空间的渐近区域  $(\log C, \log \sigma^2)$  内存在着泛化错误的一条等高线,此条等高线把参数空间划分为过拟合区域、欠拟合区域和

好区(见图1),好区内的参数设置具有好的泛化错误性能。对于每一个固定的 $\tilde{C}$ ,用式(5)来描述等高线,图1中的虚线对应着 $\tilde{C}$ ,即线性支持向量机的惩罚因子,也对应了线性向量机最小的泛化错误。

$$\log \sigma^2 = \log C - \log \tilde{C} \quad (5)$$

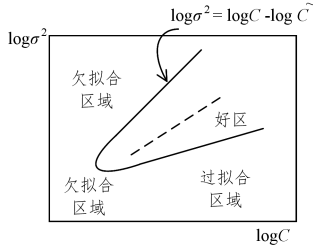


图1 超参数空间

Fig.1 Hyperparameter space

若能找到图1中的好区,指导相关算法在好区中寻找最优参数组合,将会降低时间复杂度,同时提高算法的准确率。

### 2.3 灰狼优化算法的主要思想

GWO算法模拟灰狼的狩猎机制。灰狼生活在一个严格等级的群体中:最重要是负责决策的 $\alpha$ 狼,其次是 $\beta$ 狼和 $\delta$ 狼,其余称为 $\omega$ 狼<sup>[7]</sup>。该算法中,猎物位置是最优解,灰狼代表搜索空间中的潜在解。更接近猎物的狼是 $\alpha$ 狼,它们是目前为止的最优解;从等级上讲, $\beta$ 狼是次优解;而 $\delta$ 狼是第三最优解。它们的位置在搜索空间中表示为 $X_\alpha, X_\beta$ 和 $X_\delta$ , $\omega$ 狼根据 $\alpha$ 狼、 $\beta$ 狼和 $\delta$ 狼的相对位置来更新它们在搜索空间中的位置。图2给出了灰狼和猎物的定位,以及用于更新灰狼在搜索空间中的位置等式中涉及的参数。狩猎猎物的具体步骤为包围猎物、狩猎、攻击和探索<sup>[36]</sup>。

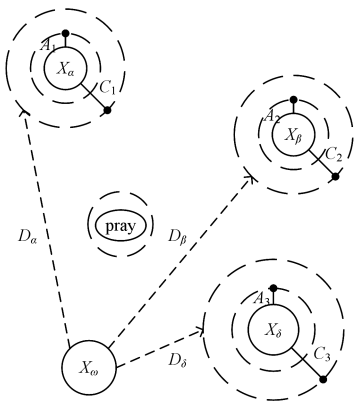


图2 搜索空间中灰狼与猎物的定位

Fig.2 Location of grey wolves and prey in the search space

#### (1)包围猎物

根据猎物的位置更新搜索空间中各个灰狼的位置,更新方程如下:

$$X(t+1) = X_p(t) + A * D \quad (6)$$

其中, $t$ 是迭代次数, $X_p$ 是猎物位置, $X$ 是灰狼位置, $D$ 表示猎物与灰狼之间的距离,定义如下:

$$D = |C * X_p(t) - X(t)| \quad (7)$$

$A$ 和 $C$ 为向量系数,其计算公式为:

$$A = 2ar_1 - a \quad (8)$$

$$C = 2r_2 \quad (9)$$

其中, $a$ 随着迭代的进行从2线性递减到0, $r_1$ 和 $r_2$ 为 $[0,1]$ 间的随机数。

#### (2)狩猎阶段

根据 $\alpha$ 狼、 $\beta$ 狼和 $\delta$ 狼的信息更新狼群中各个灰狼个体的位置,更新公式为:

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (10)$$

其中, $X_1, X_2$ 和 $X_3$ 的定义为:

$$X_1 = |X_\alpha - A_1 * D_\alpha| \quad (11)$$

$$X_2 = |X_\beta - A_2 * D_\beta| \quad (12)$$

$$X_3 = |X_\delta - A_3 * D_\delta| \quad (13)$$

其中, $X_\alpha, X_\beta$ 和 $X_\delta$ 是第 $t$ 次迭代中最优的3个解; $D_\alpha, D_\beta$ 和 $D_\delta$ 分别被定义为:

$$D_\alpha = |C_1 * X_\alpha - X| \quad (14)$$

$$D_\beta = |C_2 * X_\beta - X| \quad (15)$$

$$D_\delta = |C_3 * X_\delta - X| \quad (16)$$

#### (3)攻击阶段

攻击猎物是狩猎过程的最后阶段。接近猎物可以通过减小 $a$ 的值来进行数学建模。注意, $A$ 的变化可以通过 $a$ 的变化来实现,即 $A$ 是区间 $[-2a, 2a]$ 中的任意值,整个迭代过程中 $a$ 的区间为 $[0, 2]$ 。当 $|A| < 1$ 时,灰狼可以移动到它的当前位置与猎物之间的任何位置。

#### (4)探索阶段

探索阶段允许背离狼群去寻找新空间,以发现更优的猎物,在算法中通过设置 $|A| > 1$ 来实现<sup>[7]</sup>。

灰狼优化算法的步骤如下:

Step1 初始化灰狼数量 $N$ 、迭代次数 $t$ 、维度 $D$ 等参数以及灰狼群体 $X = (X_1, X_2, \dots, X_N)$ ,随机初始化灰狼的位置 $X_i = (X_{i1}, X_{i2}, \dots, X_{iD})^T (i = 1, 2, \dots, N)$ ;

Step2 计算各个灰狼的适应度值 $f_i$ ,找出适应度值最好的前3个灰狼个体 $X_\alpha, X_\beta, X_\delta$ ,并将适应度最好的 $X_\alpha$ 设置为最优解;

Step3 开始迭代,遍历灰狼种群,按照式(10)更新当前灰狼和猎物的位置;

Step4 更新参数 $a, A$ 和 $C$ ,根据式(23)评估各个灰狼的适应度,更新 $\alpha, \beta$ 和 $\delta$ ;

Step5 迭代结束,返回最优解 $X_\alpha$ 。

## 3 融合 SVM 渐近性的 GWO 的参数优化与特征选择

为了解决 SVM 参数设定与特征子集选择的问题,将灰狼优化算法与支持向量机相结合,融合支持向量机的渐近特性产生新的灰狼个体,到更好的搜索空间寻找最优解,构建新的 SVM 优化模型。

### 3.1 灰狼个体编码的设计

本文用灰狼优化算法进行 SVM 参数优化和特征子集选择。由于优化算法的最优解是灰狼个体 $X_\alpha$ ,因此灰狼个体的维度包含两个参数( $C, \sigma$ )和数据集的特征数量 $n$ 两部分,设计方案如图3所示。

1	2	...	$n$	$n+1$	$n+2$
$\alpha_1$	$\alpha_2$	...	$\alpha_n$	$C$	$\sigma$

图3 灰狼个体编码

Fig.3 Individual code of gray wolf

参数  $C$  和  $\sigma$  都是实值型数据。 $n$  个特征变量被初始化为  $[0, 1]$  之间的随机数,若值高于或等于 0.5,则取值为 1,表示该特征被选择;若值低于 0.5,则取值为 0,表示该特征未被选择。

### 3.2 融合渐近性产生新灰狼个体

为了在算法中融合渐近行为,必须首先获得图 1 中虚线对应的  $\tilde{C}$ 。本文选择灰狼优化算法中适应度值最好的  $X_a$ ,  $X_\beta$  和  $X_\delta$  等前  $r$  个灰狼个体作为  $\tilde{C}$  值产生的依据。为计算方便,把式(5)转化为:

$$\tilde{C} = \frac{C}{\sigma^2} \quad (17)$$

为方便叙述,令  $\tilde{C} = K$ ,  $\gamma = \frac{1}{\sigma^2}$ , 即:

$$K = C\gamma \quad (18)$$

为了使图 1 中的虚线更加接近好区中的参数最优组合,本文引入一个修正量  $\delta$ ,把式(18)变换为:

$$\delta K = C\gamma, 0 < \delta < 1 \quad (19)$$

把初始化参数  $C$  变量值的搜索空间离散化为  $d$  个值 ( $C_1, C_2, \dots, C_d$ ),然后根据式(19)计算出  $d$  个对应的  $\gamma_{ij}$ :

$$\gamma_{ij} = \frac{\delta \tilde{C}_i}{C_j}, i=1, 2, \dots, r; j=1, 2, \dots, d \quad (20)$$

最后,根据产生的新的参数组合与对应的特征重新生成灰狼个体。新的灰狼个体如表 1 所列,新灰狼个体中的  $f_1, f_2, f_3$  为  $X_a, X_\beta$  和  $X_\delta$  分别对应的特征子集。

表 1 由适应度最高的前  $r$  个灰狼个体产生的新灰狼个体

Table 1 New gray wolf individual produced by the top  $r$  gray wolf individual of highest fitness

$\alpha$ 狼新个体	$\beta$ 狼新个体	$\delta$ 狼新个体	第 $r$ 个灰狼产生的新个体
$(C_1, \gamma_{11}, f_1)$	$(C_1, \gamma_{21}, f_2)$	$(C_1, \gamma_{31}, f_3)$	$(C_1, \gamma_{r1}, f_r)$
$(C_2, \gamma_{12}, f_1)$	$(C_2, \gamma_{22}, f_2)$	$(C_2, \gamma_{32}, f_3)$	$(C_2, \gamma_{r2}, f_r)$
...	...	...	...
$(C_d, \gamma_{1d}, f_1)$	$(C_d, \gamma_{2d}, f_2)$	$(C_d, \gamma_{3d}, f_3)$	$(C_d, \gamma_{rd}, f_r)$

新的灰狼个体选择了合适的  $\tilde{C}$ , 蕴含了支持向量机的渐近特性。新生成的灰狼个体加强了灰狼优化算法的搜索力度,提高了支持向量机的分类准确率。

### 3.3 融合渐近性的灰狼优化 SVM 模型

基于融合渐近性的灰狼优化算法进行支持向量机的特征选择与参数优化的模型如图 4 所示。

首先对参数初始化、数据预处理、特征子集选取和适应度评估进行详细阐述,然后给出所提模型的主要算法步骤。

#### (1) 参数初始化

首先,初始化灰狼优化算法的基本参数,包括灰狼数量、每个个体的维度、初始位置以及迭代次数等。由于在提出的优化 SVM 模型中需要寻找最优的  $(C, \sigma)$  组合和最优的特征子集,因此搜索空间是  $2+n$  维的,并且灰狼的每个位置就是一个  $(C, \sigma)$  与特征子集的组合。根据初始化的灰狼种群,随机选择 3 个灰狼个体作为  $X_a, X_\beta$  和  $X_\delta$ 。

#### (2) 数据预处理

数据归一化处理是数据挖掘的一项基础工作。不同的评价指标往往具有不同的量纲,数据间的差别可能很大,不进行处理会影响到数据分析的结果。为了消除指标之间量纲和取

值范围差异的影响,本文对数据进行标准化处理,以获得更高的分类精度。本文采用最小-最大规范化,即对原始数据进行线性变换,使每个特征可以缩放到  $[0, 1]$  范围,转换公式如下:

$$x^* = \frac{x - \min}{\max - \min} \quad (21)$$

其中,  $x$  为样本属性的原始值,  $\min$  为样本数据属性的最小值,  $\max$  为样本数据属性的最大值,  $x^*$  是变换后的值。

#### (3) 特征子集的选取

根据最优灰狼个体  $X_a$  的前  $n$  个值进行特征提取,保留相关特征,丢弃不相关的特征,用训练特征子集和测试特征子集分别进行 SVM 训练和测试训练好的模型的泛化能力。

#### (4) 适应度函数与评估

适应度函数是灰狼优化算法的基础组成部分,用于评估灰狼个体的优劣。以往的研究人员将准确性和特征数量作为两个重要参数,并评估每个参数的特征子集<sup>[13]</sup>。由文献[36]可知,支持向量的个数对模型的泛化能力也有重要影响,本文在式(22)的基础上增加了模型支持向量个数项,提出了新的适应度函数,如式(23)所示,此适应度函数的最优值可以获得最高的准确率、最小的特征子集和最少的支持向量。

$$fitness = W_A \times SVM\_accuracy + W_F \times \left( \sum_{i=1}^{n_f} C_i \times F_i \right)^{-1} \quad (22)$$

$$fitness = W_A \times SVM\_accuracy + W_F \times \left( \sum_{i=1}^{n_f} C_i \times F_i \right)^{-1} + W_V \times \left( \sum_{i=1}^l V_i \right)^{-1} \quad (23)$$

其中,  $SVM\_accuracy$  是分类准确率;  $W_A$  是其权重;  $W_F$  是特征数量的权重;  $F_i$  是第  $i$  个特征,被选中取 1,未被选中取 0;  $C_i$  是第  $i$  个特征的权重;  $n_f$  是数据集的总特征数;  $V_i$  表示第  $i$  个样本,若是支持向量则取值为 1,否则取 0;  $l$  是总的样本数;  $W_V$  是支持向量的权重,一般取 0.05。

由于每个灰狼个体表示  $(C, \sigma)$  组合和选取的特征子集,当用测试集获取准确率后,每个灰狼个体就可式(23)进行评估,从而选择出最优的灰狼个体  $X_a, X_\beta$  和  $X_\delta$ 。

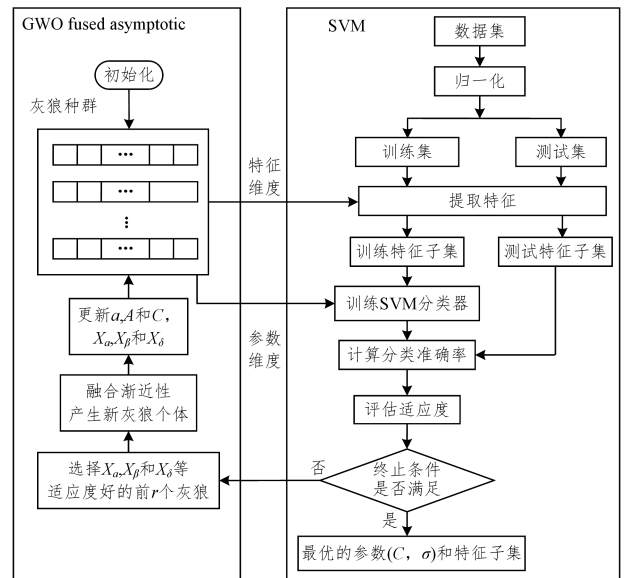


图 4 融合渐近性的灰狼优化 SVM 的流程

Fig. 4 Flowchart of GWO SVM fused asymptotic

所提模型的处理过程如下:

Step1 数据处理。数据集中的特征按照式(21)进行归一化处理,然后划分为训练集与测试集。

Step2 初始化。初始化灰狼数量  $N$ 、迭代次数  $t$ 、维度  $D$  以及参数  $\alpha$ 、 $A$  和  $C$ 。

Step3 选取特征子集。依据初始灰狼种群中特征的维度,选择训练子集与测试子集。

Step4 训练 SVM。根据训练子集与灰狼个体中的参数维度  $(C, \sigma)$  进行训练,获取支持向量机中的权重向量  $W$  和偏置  $b$ 。

Step5 计算分类准确率。使用测试子集和 Step4 获得的相关变量计算分类准确率。

Step6 评估适应度。根据式(23)评价每个灰狼个体的适应度,并且找出适应度最高的前  $r$  个灰狼个体。

Step7 产生新灰狼种群。依据式(17)一式(20),由适应度最高的前  $r$  个灰狼个体融合渐近性产生新的灰狼个体,继而产生新的灰狼种群。

Step8 更新参数。据式(8)、式(9)、式(11)和式(12)分别更新参数  $A$ 、 $C$  和灰狼个体  $X_\alpha$ 、 $X_\beta$  和  $X_\delta$ ,依据新的灰狼种群训练 SVM。

Step9 终止条件。若最大迭代次数未达到,则转 Step3,否则转 Step10。

Step10 获取最优解。训练过程终止,得到最优参数与最优特征子集。

## 4 实验

### 4.1 实验环境

本文实验平台是 Intel Core CPU i7-7700HQ, 8 GB RAM, Windows 10 操作系统;开发环境是 Matlab2017a, Libsvm (3.1)<sup>[37]</sup>,设置参数的搜索范围为  $(0.01, 35000)$ ,  $\sigma$  的搜索范围为  $(0.0001, 32)$ 。

### 4.2 实验数据集

从 UCI 机器学习库中选取了 10 个数据集,来验证本文方法在不同数据集中的有效性。这 10 个数据集分别为: Australian, Breast cancer, German, Heart disease, Sonar, Liver, Vehicle, Vowel, Iris and Wine。

表 2 列出了这些数据集的类别数、样本数以及初始特征数等相关属性。实验中,采用  $k$  折交叉验证。交叉验证是指将数据集  $D$  划分为  $k$  个大小相似的互斥子集,每个子集都尽

可能保持数据分布的一致性,子集从数据集  $D$  中分层采样得到;然后,每次用  $k-1$  个子集的并集作为训练集,将余下的子集作为测试集,于是可实现  $k$  次训练和测试;最后得到的是这  $k$  次测试结果的均值,本文实验中  $k$  置为 10。

表 2 数据集描述

Table 2 Dataset description

序号	数据集	类别数	样本数	特征数
1	Australian	2	653	15
2	Breast cancer	2	683	10
3	German	2	1000	30
4	Heart disease	2	270	13
5	Sonar	2	208	60
6	Liver	2	345	6
7	Vehicle	4	846	18
8	Vowel	11	528	10
9	Iris	3	150	4
10	Wine	3	175	13

### 4.3 实验结果与分析

(1)表 3 列出了本文方法与文献[34]中的方法以及网格搜索算法的结果对比,分类准确率采用平均数±标准差的形式表示。可以看出,本文算法在 10 个数据集上的分类准确率都高于文献[34]中方法的准确率。

表 3 本文算法与文献[34]、网格搜索分类算法的准确率比较

Table 3 Accuracy comparison of proposed algorithm and literature [34], grid search algorithm

(单位:%)

数据集	本文算法	文献[34]方法	网格搜索
Australian	92.68±2.91	91.59±2.14	84.74±4.52
Breast cancer	99.69±1.21	99.00±1.66	95.43±2.50
German	86.73±1.91	86.10±1.97	78.90±1.73
Heart disease	95.93±2.63	95.56±2.34	88.15±5.18
Sonar	99.30±1.58	99.00±2.11	90.50±8.32
Liver	85.72±5.38	83.14±7.19	72.89±5.60
Vehicle	88.27±1.44	88.24±1.47	83.94±2.74
Vowel	99.70±0.82	99.60±0.71	95.36±2.24
Iris	100.00±0.00	100.00±0.00	94.09±4.77
Wine	100.00±0.00	100.00±0.00	97.16±2.88

(2)为比较本文算法和没有融合渐近性的灰狼优化算法的差异,在上述 10 个 UCI 数据集上采用 10 折交叉验证计算其分类准确率,并用非参数的 Wilconxon 秩和检验进行检验。如表 4 所列,除 Iris, Vowel 和 Wine 的  $P$  值大于规定的显著性水平 0.005 外,其他的  $P$  值均小于 0.005。因此,与没有融合渐近特性的灰狼优化支持向量机算法相比,本文算法具有更高的准确率和较少的特征子集。

表 4 本文算法与未融合渐近性的 GWO 算法的对比

Table 4 Comparison between proposed algorithm and GWO algorithm without asymptotic

数据集	特征数	本文算法		GWO+SVM		$P$ 值
		准确率/%	选择的特征数	准确率/%	选择的特征数	
Australian	15	92.68±2.91	7.6±2.01	89.04±1.45	8.0±2.41	<0.005
Breast cancer	10	99.69±1.21	2.9±1.20	95.69±2.01	3.1±1.46	<0.005
German	30	86.73±1.91	12.6±2.21	80.73±2.81	13.4±1.89	<0.005
Heart disease	13	95.93±2.63	6.3±1.86	91.12±1.91	6.8±1.75	<0.005
Sonar	60	99.30±1.58	23.4±2.21	94.56±2.41	29.7±4.61	<0.005
Liver	6	85.72±5.38	3.6±1.18	77.43±7.29	3.9±1.45	<0.005
Vehicle	18	88.27±1.44	10.23±1.85	84.59±3.09	10.34±2.61	<0.005
Vowel	10	99.70±0.82	6.05±1.36	98.97±1.51	6.9±1.64	0.3980
Iris	4	100.00±0.00	1.4±0.64	99.36±2.11	1.5±0.61	0.3681
Wine	13	100.00±0.00	3.88±0.56	99.50±1.81	4.2±0.91	0.3681

(3)为了评估本文方法的有效性,表5列出了本文算法和其他算法在UCI数据集Australian上的准确率。由表5可知,本文算法具有较强的竞争力。

表5 本文算法与其他算法在Australian数据集上的分类准确率

Table 5 Classification accuracy of proposed algorithm and other algorithms on Australian dataset

文献	方法	准确率/%
本文算法	融合渐近性的灰狼算法	92.68
文献[34]算法	带特征染色体的遗传算法	91.59
文献[29]算法	GOA+SVM	85.07
文献[28]算法	MVO+SVM	86.81
文献[38]算法	GA+feature weigh+SVM	88.55
文献[36]算法	CGPGA+SVM	90.13
文献[27]算法	PSO+SVM	91.03
文献[13]算法	GA+SVM	88.10
—	决策树(C4.5)	85.22
—	随机森林	86.81
—	贝叶斯	72.46
—	Boosting	86.23

表6 本文算法和未融合渐近性的GWO算法以及网格搜索算法的性能对比

Table 6 Performance comparison of proposed algorithm and GWO algorithm without asymptotic, grid search algorithm

数据集	特征数	准确率/%			选择的特征数		平均处理时间/s		
		GWO-FA	GWO-NFA	网格搜索	GWO-FA	GWO-NFA	GWO-FA	GWO-NFA	网格搜索
Breast cancer	10	99.69	95.69	95.43	2.9	3.1	11.59	13.02	15.90
Heart disease	13	95.93	91.12	88.15	6.3	6.8	9.26	9.68	12.08
Ionosphere	60	99.30	94.56	90.50	23.4	29.7	10.08	11.16	16.32

从实验结果可以看出,GWO-FA获得了99.69%的准确率,所需的平均处理时间为11.59s,得到的特征数更少,而GWO-NFA和网格搜索算法分别需要13.02s和15.90s,其他数据集的比较类似,从而验证了上述时间复杂度分析的正确性。本文算法能获得更高的准确率、更小的特征子集和更少的处理时间,具有更好的性能。

**结束语** 本文使用灰狼优化算法来进行支持向量机的特征选择和参数优化,同时融合了支持向量机的渐进性,从而将灰狼优化算法的搜索导向超参数空间的好区中的最佳泛化误差直线上。其本质是将支持向量机的性能特征融入算法,使其不仅具备灰狼优化算法的搜索功能,而且具有融合渐近性灰狼个体的搜索功能,因此新算法具备更强的搜索力度,且分类准确率得到显著提高。

适用度函数采用支持向量机分类准确率与被选择的特征数、支持向量数作为标准,同时进行参数优化与特征选择。实验结果表明,与网格搜索算法、不带融合渐近性的算法以及其他相关文献中的分类方法相比,本文算法在分类准确率、特征子集的选择和时间复杂度方面都表现出了良好的特性,是一种有效的算法。下一步将在更多的数据集上对本文算法进行验证,同时考虑融合数据集特征权重来验证分类准确率。

## 参考文献

[1] DRUCKER H, WU H, VAPNIK V N. Support vector machines for spam categorization[J]. IEEE Transactions on Neural Networks, 1999, 10(5): 1048-1054.

[2] VATSA M, SINGH R, NOORE A. Improving biometric recognition accuracy and robustness using dwt and svm water-

(4)对融合渐近性的灰狼优化算法(GWO-FA)和不融合渐近性的灰狼优化算法(GWO-NFA)进行时间复杂度的对比分析。在这两种算法中,适应度评价是最耗费时间的。在GWO-NFA中,若种群规模为 $N$ ,每迭代一次就评价一次适应度,时间复杂度为 $O(N)$ ,当迭代次数是 $t_1$ 时,耗时为 $t_1 \cdot O(N)$ 。而在GWO-FA中,计算代价还包括对新生成的个体进行一次适应度评价,若新个体规模为 $n_f$ ,则时间复杂度为 $O(N+n_f)$ ;当迭代次数为 $t_2$ 时,耗时为 $t_2 \cdot O(N+n_f)$ ,忽略 $N$ 与 $n_f$ 的差异,总耗时记为 $t_2 \cdot O(N)$ 。

实验表明,融合渐近性的GWO达到预定的分类准确率所需的种群规模小,迭代次数少,因此GWO-FA的时间复杂度比GWO-NFA小。

本文采用UCI库中的Breast cancer, Heart disease, Ionosphere这3个数据集对上述两种算法以及网格搜索算法进行性能对比实验。实验中,采用10折交叉验证来获得分类准确率、特征数、平均处理时间,结果如表6所列。

marking[J]. IEICE Electronics Express, 2005, 2(12): 362-367.

[3] BYVATOV E, SCHNEIDER G. Support vector machine applications in bioinformatics[J]. Applied bioinformatics, 2002, 2(2): 67-77.

[4] DOUCET J P, BARBAULT F, XIA H, et al. Nonlinear svm approaches to qspr/qsar studies and drug design[J]. Current Computer-Aided Drug Design, 2007, 3(4): 263-289.

[5] LIN S W, YING K C, CHEN S C, et al. Particle swarm optimization for parameter determination and feature selection of support vector machines[J]. Expert Systems with Applications, 2008, 35(4): 1817-1824.

[6] ZHANG X L, CHEN X F, HE Z J. An aco-based algorithm for parameter optimization of support vector machines[J]. Expert Systems with Applications, 2010, 37(9): 6618-6628.

[7] MIRJALILI S, MIRJALILI S M, LEWIS A. Grey wolf optimizer[J]. Advances in Engineering Software, 2014, 69(5): 46-61.

[8] JALILIS M R. How effective is the grey wolf optimizer in training multilayer perceptrons[J]. Applied Intelligence, 2015, 42(4): 608-619.

[9] SONG H M, SULAIMAN M H, MOHAMED M R. An application of grey wolf optimizer for solving combined economic emission dispatch problems[J]. International Review on Modelling and Simulations, 2014, 7(5): 838-844.

[10] SULAIMAN M H, MUSTAFFA Z, MOHAMED M R, et al. Using the grey wolf optimizer for solving optimal reactive power dispatch problem[J]. Applied Soft Computing, 2015, 32: 286-292.

[11] LONG W, ZHAO D Q, XU S J. Improved grey wolf optimization algorithm for constrained optimization problem[J]. Journal of

- Computer Applications,2015,35(9):2590-2595.
- [12] FRIEDRICHSF,IGEL C. Evolutionary tuning of multiple SVM parameters[J]. Neurocomputing,2005,64(2):107-117.
- [13] HUANG C L, WANG C J. A GA-based feature selection and parameters optimization for support vector machines[J]. Expert Systems with Applications,2006,31:231-240.
- [14] WU C H, TZENG G H, LIN R H. A novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression[J]. Expert Systems with Applications,2009,36(3):4725-4735.
- [15] CHEN J Y, XIONG H, ZHENG H B. Parameters Optimization for SVM Based on Particle Swarm Algorithm [J]. Computer Science,2018,45(6):197-203.
- [16] GUO L, WU Y X, ZHAO L, et al. Classification of Mental Task From EEG Signals Using Immune Feature Weighted Support Vector Machines [J]. IEEE Transactions on Magnetics, 2011,47 (5):866-869.
- [17] THARWAT A. A BA-based algorithm for parameter optimization of Support Vector Machine [J]. Pattern Recognition Letters,2017,93:13-22.
- [18] ALA' M A, HOSSAM F, JA' FAR A. Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingua contexts [J]. Knowledge-based systems,2018,8(153):91-104.
- [19] CHAO C F, HORNG M H. The construction of support vector machine classifier using the firefly algorithm [J]. Computational Intelligence and Neuroscience,2015,1:1-8.
- [20] PRABUKUMAR M, AGILANDEESWARI L, GANESAN K. An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier [J]. Journal of Ambient Intelligence and Humanized Computing,2019,1(10):267-293.
- [21] XUE H X, BAI Y P, HU H P. A Novel Hybrid Model Based on TVIW-PSO-GSA Algorithm and Support Vector Machine for Classification Problems [J]. IEEE Access, 2019 (7): 27789-27801.
- [22] MAO K Z. Feature subset selection for support vector machines through discriminative function pruning analysis [J]. IEEE Transactions on Systems, Man, and Cybernetics, 2004, 34 (1): 60-67.
- [23] RAYMER M L, PUNCH W F, GOODMAN E D, et al. Dimensionality reduction using genetic algorithms [J]. IEEE Transactions on Evolutionary Computation,2000(2):164-171.
- [24] YANG J, HONAVAR V. Feature subset selection using a genetic algorithm [J]. IEEE Intelligent Systems,1998,13(2):44-49.
- [25] GUYON I, WESTON J, BARNHILL S, et al. Gene selection for cancer classi? cation using support vector machines [J]. Machine Learning,2002,46(1/2/3):389-422.
- [26] LIN S W, YING K C, Chen S C, et al. Particle swarm optimization for parameter determination and feature selection of support vector machines [J]. Expert Systems with Applications,2008,35(4):1817-1824.
- [27] ZHOU T, LU H L, WANG W W, et al. GA-SVM based feature selection and parameter optimization in hospitalization expense modeling [J]. Applied Soft Computing,2019,75 (2):323-332.
- [28] FARIS H, HASSONAH M A, AL-ZOUBI, et al. A multi-verse optimizer approach for feature selection and optimizing SVM parameters based on a robust system architecture [J]. Neural Computing and Applications,2017,30(8):2355-2369.
- [29] ALJARAHA I. Simultaneous Feature Selection and Support Vector Machine Optimization Using the Grasshopper Optimization Algorithm [J]. Cognitive Computation,2018,10(3):478-495.
- [30] CHEN Y, MA H W. Feature selection and parameter optimization of support vector machine based on the bees algorithm [J]. Modular Machine Tool&Automatic Manufacturing Technique, 2013 (11):41-43.
- [31] LIN K C, CHEN S Y, JASON C, et al. Feature Selection and Parameter Optimization of Support Vector Machines Based on Modified Artificial Fish Swarm Algorithms [J]. Mathematical Problems in Engineering,2015, Article ID 604108.
- [32] SHEN Y L, SONG J, WAN Z C. Improved Fireworks Algorithm for Support Vector Machine Feature Selection and Parameters Optimization [J]. Microelectronics&Computer,2018,35(1):21-25.
- [33] IKEDA K, AOISHI T. An asymptotic statistical analysis of support vector machines with soft margins [J]. Neural Networks, 2005,18(3):251-259.
- [34] ZHAO M Y. Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes [J]. Expert Systems with Applications,2011,38(5):5197-5204.
- [35] KEERTHI S S, LIN C J. Asymptotic behaviors of support vector machines with Gaussian kernel [J]. Neural Computation, 2003,15:1667-1689.
- [36] CHEN Z. A Parallel Genetic Algorithm Based Feature Selection and Parameter Optimization for Support Vector Machine [J]. Scientific Programming,2016:1-10.
- [37] CHANG C C, LIN C J. LIBSVM: a library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology,2011,2(3).
- [38] PHAN A V, NGUYEN M L, BUI L T. Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems [J]. Applied Intelligence,2016,46(2):455-469.



**WU Yu-kun**, born in 1980, doctoral student, is member of China Computer Federation (CCF). His main research interests include machine learning and big data.



**Wei William LEE**, born in 1958, Ph.D , professor. His main research interests include big data, block chain, IOT and smart city development.