

# 基于河马算法的贝叶斯分类技术研究<sup>\*</sup>

韩俊威<sup>1</sup> 王荣浩<sup>2</sup> 黄世沛<sup>1†</sup> 吴畏<sup>1</sup> 黄俊华<sup>3</sup> 骆茂森<sup>4</sup>

(1. 温州大学 电气数字化设计技术国家地方联合工程实验室, 温州 325035)

(2. 中国人民解放军陆军工程大学, 南京 210000)

(3. 国网浙江省电力有限公司淳安县供电公司, 杭州 311700)

(4. 国网浙江省电力有限公司建德市供电公司, 杭州 311600)

**摘要** 朴素贝叶斯分类器(NBC)凭借其坚实的概率理论基础,在处理包含不确定特征和噪声干扰的数据集时展现出了显著的分类优势.随着社会数据的复杂性日益攀升,以占比量来衡量先验概率的方法在一定程度上限制了朴素贝叶斯分类器的性能表现.先验概率的构造是贝叶斯分类研究中的重要问题,是决定朴素贝叶斯分类准确率的重要因素.如何有效估计和构造最优先验概率逐渐成为学者们关注的研究议题.为此,本文引入  $t$  分布变异和自适应权重对河马优化算法(HOA)的个体更新公式进行改进,并基于此提出了一种结合优化算法、训练样本和测试样本构造贝叶斯最优先验的优化方法,得到了较好的分类性能.具体流程为:采集系统数据并划分为训练集、验证集和测试集,以训练集得到的模型参数作为验证集的贝叶斯分类器初始输入,再以分类准确率为目标函数,采用改进河马优化算法搜寻验证集的贝叶斯最优先验,最后将寻优结果作为测试集的先验,得到分类准确率.通过切换电路系统仿真来测试所提出的方法,并与其他主流分类算法对比,结果显示所提方法表现出较高的分类准确性.

**关键词** 朴素贝叶斯分类, 河马优化算法, 最优先验概率, 数据分类, 改进策略

中图分类号: TP274.3

文献标志码: A

## Research on Bayesian Classification Technology Based on Hippopotamus Algorithm<sup>\*</sup>

Han Junwei<sup>1</sup> Wang Ronghao<sup>2</sup> Huang Shipai<sup>1†</sup> Wu Wei<sup>1</sup> Huang Junhua<sup>3</sup> Luo Maosen<sup>4</sup>

(1. National Local Joint Engineering Laboratory of Electrical Digital Design Technology, Wenzhou University, Wenzhou 325035, China)

(2. PLA Army Engineering University, Nanjing 210000, China)

(3. Chun'an City Power Supply Company, Zhejiang Electric Power Corporation, State Grid, Hangzhou 311700, China)

(4. Jiande City Power Supply Company, Zhejiang Electric Power Corporation, State Grid, Hangzhou 311600, China)

**Abstract** The Naive Bayes classifier(NBC), with its solid foundation in probability theory, exhibits significant classification advantages when dealing with datasets containing uncertain features and noise interference. With the increasing complexity of social data, the method of measuring prior probability based on proportion has to some extent limited the performance of Naive Bayes classifiers. The construction of prior probabilities is an important issue in Bayesian classification research and a crucial factor in determining the accuracy of Naive Bayes classification. How to effectively estimate and construct the highest priority probability has gradually become a research topic of concern for scholars. Therefore, this article introduces  $t$ -distribution variation and adaptive weights to improve the individual update for-

2024-09-03 收到第 1 稿, 2024-10-22 收到修改稿.

<sup>\*</sup> 国家自然科学基金资助项目(62173341), 江苏省自然科学基金(BK20231487), National Natural Science Foundation of China (62173341), Natural Science Foundation of Jiangsu Province (BK20231487).

<sup>†</sup> 通信作者 E-mail: hspei@sina.cn

mula of Hippopotamus Optimization Algorithm (HOA). Based on this improvement, an optimization approach is proposed, which integrates the optimized algorithm with the construction of Bayesian optimal priors using both training and testing samples. This method has achieved superior classification performance. The specific process is to collect system data and divide it into training set, validation set, and test set. The model parameters obtained from the training set are used as the initial input for the Bayesian classifier in the validation set. Then, with classification accuracy as the objective function, the improved Hippopotamus Optimization Algorithm is used to search for the Bayesian optimal prior in the validation set. Finally, the optimization result is used as the prior in the test set to obtain the classification accuracy. The proposed method was tested by switching circuit system simulations and compared with other mainstream classification algorithms. The results showed that the proposed method exhibited high classification accuracy.

**Key words** Naive Bayes classification, hippopotamus optimization algorithm, first priority verification probability, data classification, improvement strategy

## 引言

当今社会已进入大数据信息化时代,数据的规模和复杂性呈指数级增长,如何有效地处理和分析这些数据,挖掘其潜在价值,成为科研、商业乃至社会发展的重要课题. 分类技术作为数据挖掘领域中的一项关键技术,其重要性在学术研究中日益凸显. 它通过对数据进行系统、科学地分类,不仅能够揭示数据内在的规律和模式,还能为后续的统计分析、模型构建和预测分析提供坚实的基础. 在学术研究中,分类技术被广泛应用于各个领域,如生物信息学<sup>[1-3]</sup>、经济学<sup>[4-6]</sup>、社会学<sup>[7-9]</sup>等,成为推动学科发展和知识创新的重要工具.

当今主流的分类技术包括 K 近邻<sup>[10]</sup>、支持向量机<sup>[11]</sup>、决策树<sup>[12]</sup>、神经网络<sup>[13]</sup>和朴素贝叶斯分类器(Naive Bayes classifier, NBC)<sup>[14]</sup>等,其中朴素贝叶斯分类器基于概率理论具备坚实的数学基础,能够处理不确定性和噪声数据,适用于不同类型的数据集甚至是非常规复杂数据集,其强大的适用性和可解释性使其饱受学者青睐. 随着大数据时代的到来,数据的复杂性和多样性不断增加,对分类算法的要求也越来越高. 朴素贝叶斯分类器因其独特的优势,成为学者们研究的热点之一. 近年来,学者们对朴素贝叶斯分类器的研究不断深入,不仅在传统领域取得了丰硕的成果,还在新兴领域如机器学习<sup>[15]</sup>、文本处理<sup>[16]</sup>、模式识别<sup>[17]</sup>等领域中发掘出朴素贝叶斯分类器的巨大潜力. 例如, Tan 等<sup>[18]</sup>基于朴素贝叶斯分类器和概率神经网络提出

了一种新的气体泄漏监测传感器故障诊断方法,用于识别异常的安全监测数据并进行传感器故障分类;学者 Ruan 等人<sup>[19]</sup>使用统计特征加权技术,基于多项式朴素贝叶斯文本分类器提出了一种新的分类深度特征加权方法;Kim 等<sup>[20]</sup>考虑了标签之间的关系或依赖性,提出了一种新的多标签朴素贝叶斯分类器;Reddy 等<sup>[21]</sup>使用网络节点认证模块结合朴素贝叶斯分类模块,提出了一种分布式拒绝服务(distributed denial of service, DDoS)攻击严重性缓解解决方案,用于检测和隔离 DDoS 攻击流量模式;Zhang 等<sup>[22]</sup>提出了一种基于贝叶斯加性分类树的非线性分类方法,在和支持向量机(support vector machine, SVM)、随机森林(random forest, RF)、分类与回归树(classification and regression tree, CART)等方法的对比中有着较好的效果;Balaji 等<sup>[23]</sup>使用一种新的动态图割算法来分割皮肤病变,然后使用朴素贝叶斯分类器进行皮肤病分类,结果分别比全卷积神经网络(fully convolutional networks, FCN)和 SegNet 方法高出 6.5%和 8.7%. 鉴于朴素贝叶斯分类器在多个领域内的广泛应用,当今贝叶斯算法的重要性日益凸显.

随着朴素贝叶斯分类器在多个领域的广泛应用和深入研究,学者们发现,朴素贝叶斯分类器在面对特征间存在复杂分布关系的数据集时,其分类性能往往受限. 为了克服这一局限,研究者们开始探索先验概率优化的途径,旨在通过更加精细地估计先验分布,来提升朴素贝叶斯分类器的整体性能. 先验概率作为贝叶斯推断的基础,直接影响了

分类结果的准确性和可靠性。因此,对贝叶斯先验概率进行进一步的探索研究,成为推动朴素贝叶斯分类器性能提升的关键所在。近年来,学者们不仅关注先验概率的设定方法,还积极研究如何根据特定应用场景的数据特点去调整和优化先验概率,以提高分类器的性能。学者们做出了许多努力来开发模型选择的客观先验,比如,Llorente等<sup>[24]</sup>讨论了边际似然的先验敏感性问题及其在模型选择中的作用;Villa等<sup>[25]</sup>在模型嵌套情况下使用KL散度(Kullback-Leibler divergence)来评价每个模型的价值参数,从而确定针对不同模型的合适先验概率;Trang等<sup>[26]</sup>用模糊聚类方法(fuzzy-c means, FCM)将训练集中的种群信息与新的观测值结合起来确定分类问题的先验概率,从而提出了一种结合训练数据、新观测数据和FCM确定先验概率的算法;Boluki等<sup>[27]</sup>基于条件概率形式的约束的一般框架,提出一种新的先验概率构造法;Maruyama等<sup>[28]</sup>研究了具有球对称误差分布的线性模型中的贝叶斯变量选择问题以及 $g$ -先验混合的贝叶斯因子的拉普拉斯近似。

上述先验概率构造方法从模型本身出发,相比仅以占比量分配先验概率的经典方法,其先验概率的构造方法更加合理,朴素贝叶斯分类准确率也有一定的提升,但仍有继续提升的空间。为了尽可能提高分类的准确率,当今贝叶斯先验概率的研究领域不仅涉及概率论和统计学的知识,还融合了机器学习、优化算法等多个学科的理论。例如Tran等<sup>[29]</sup>基于Wasserstein距离的最小化,以函数先验推理的方式提出一种鲁棒性较强的框架来将函数先验与神经网络的函数先验进行匹配,从而调整神经网络的参数先验;Esfahani等<sup>[30]</sup>通过路径分析中的不完全先验信息,探索了优化先验概率构造的新途径,相比之下取得了较好的效果;Wong等<sup>[31]</sup>基于所有属性都遵循相同的广义狄利克雷先验的假设,使用朴素贝叶斯机制优化属性的筛选与排序,最后根据属性排序来搜索每个单一属性的最佳先验,取得了较好的效果。这些研究充分展示了融合机器学习与优化算法策略的先验概率优化方法能得到对数据集适应能力更强的先验参数与贝叶斯模型,且相比于朴素贝叶斯分类器和主流分类方法往往能在复杂数据集中展示更高的分类准确率。受此启发,本文提出了一种结合优化算法、训练样

本和测试样本确定特定数据集的贝叶斯最优先验概率的方法。

河马优化算法(hippopotamus optimization algorithm, HOA),是Amiri<sup>[32]</sup>于2023年提出的一种新型群智能优化算法,它模拟了自然界中河马寻找食物的行为过程,通过不断地迭代和搜索,找到问题的最优解。该算法收敛快,灵活性高,在力学设计、成本考虑、风电输出<sup>[32]</sup>等工程最优化问题上的表现良好。但是HOA在某些情况下也存在精度低、易陷入局部最优等缺点。本文的主要贡献如下:

(1)通过引入自适应线性权重与 $t$ 分布变异对HOA进行改进,提出一种改进的河马优化算法(Improved hippopotamus optimization algorithm, IHOA),改进后的算法在收敛速度与寻优精度方面的性能有所提升。

(2)基于改进河马优化算法并融合了数据的训练集、验证集和测试集,提出了一种确定数据集贝叶斯最优先验的方法。以二状态电路系统为例,以预测正确的正负样本数之和与测试集样本数的比率作为目标函数,通过迭代得到最优先验并用于测试集的分类,通过对比可知该方法分类准确率较高,为电路系统后续辨识提供了可靠的分类支持。

## 1 朴素贝叶斯模型

朴素贝叶斯分类器是一种基于贝叶斯定理的分类算法<sup>[33]</sup>。朴素贝叶斯分类器假设特征向量的各分量相对于决策变量是相对独立的。设一个训练样本集为 $D$ ,其样本个数为 $N$ ,类别个数为 $m$ ,即表示为

$$C = \{C_1, C_2, \dots, C_m\} \quad (1)$$

$N$ 个样本中第 $C_j$ 类样本个数为 $N_{C_j}$ ,其中 $j=1, 2, \dots, m$ 。若该样本包含属性数量为 $n$ ,则第 $r$ 个样本表示为

$$\mathbf{x}_r = (x_{r1}, x_{r2}, \dots, x_{rn}), \quad r=1, 2, \dots, N \quad (2)$$

其中 $x_{ri}$ 表示第 $r$ 个样本属性 $A_i$ 的取值, $i=1, 2, \dots, n$ 。

现给定一个待判样本 $\mathbf{y}$

$$\mathbf{y} = (y_1, y_2, \dots, y_n) \quad (3)$$

然后可以根据朴素贝叶斯分类原理计算出待判样本 $\mathbf{y}$ 属于类别 $C_j$ 的条件概率为

$$P(C_j | \mathbf{y}) = \frac{P(C_j)}{P(\mathbf{y})} P(\mathbf{y} | C_j) \quad (4)$$

由于 $p(\mathbf{y})$ 对所有类为常数,所以只需要 $p(\mathbf{y} |$

$C_j)p(C_j)$ 最大即可. 先验概率  $p(C_j)$  的估计值可利用样本集  $D$  得出, 即  $p(C_j) = N_{C_j}/N$ . 由于朴素贝叶斯分类器以类别条件独立性为假设前提, 因此有

$$P(y | C_j) = \prod_{k=1}^n P(y_k | C_j) \\ = P(y_1 | C_j)P(y_2 | C_j) \cdots P(y_n | C_j) \quad (5)$$

各条件概率  $P(y_1 | C_j), P(y_2 | C_j), \dots, P(y_n | C_j)$  可进行连续或离散情况下的讨论.

如果是连续值属性, 假设连续值属性服从高斯分布, 公式如下:

$$g(y, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad (6)$$

$$P(y_k | C_j) = g(y_k, \mu_{C_j}, \sigma_{C_j}) \quad (7)$$

其中  $\mu_{C_j}$  和  $\sigma_{C_j}$  为  $C_j$  类中特征属性的均值和标准差.

如果是离散值属性, 则通过相应属性中值为  $y_k$  且属于  $C_j$  类的个数来计算  $p(y_k | C_j)$ . 后验概率的数学公式描述如下:

$$P(C_j | y) = \frac{P(C_j)}{P(y)} \prod_{k=1}^n P(y_k | C_j) \quad (8)$$

测试样本会被分在最大化的后验概率所包含的类中, 综上所述, 可得到朴素贝叶斯分类器的模型为

$$V_{nb}(E) = \arg\max_C P(C) \prod_{k=1}^n P(y_k | C_j) \quad (9)$$

以上是传统贝叶斯分类算法的内容. 在朴素贝叶斯分类器的应用中, 先验概率的设定对于分类性能具有显著影响. 传统先验概率往往基于样本占比的假设来进行分配, 这在处理某些简单或特定类型的数据集时可能表现出良好的分类效果. 然而当面对复杂且非标准化的数据集, 特别是非常规数据集时, 这种基于均匀分布的先验概率设定方式往往无法准确反映数据的真实分布, 从而导致分类结果不理想. 为了提高分类结果的可靠性, 本文提出基于改进河马优化算法的先验概率构造方法, 旨在通过引入河马算法的寻优机制, 针对特定数据集对先验概率的分配进行优化, 使得分类器能够更好地适应复杂数据集的特性, 提高分类准确率. 下一节将简单介绍所采用的河马优化算法.

## 2 河马优化算法及其改进策略

本小节介绍了原始河马优化算法(HOA)原理, 提出了改进河马优化算法(IHOA)的改进策略, 并给出了 IHOA 的流程图介绍.

### 2.1 HOA 介绍

河马优化算法是一种新型的元启发式优化算法<sup>[32]</sup>. 通过模拟河马的社会行为来对粒子位置进行更新, 例如它们在河流或池塘中的位置更新, 对捕食者的防御策略以及逃避捕食者的方法. 河马优化算法能够自适应地调整搜索空间的分辨率和搜索速度, 快速而准确地找到最优解.

#### 2.1.1 初始化

设种群个体为  $N$ , 每个个体都是由  $m$  个决策变量构成的向量, 代表着粒子的搜索位置, 在数学上由一个  $N \times m$  的矩阵表示. 每只河马的位置代表了优化问题的候选解.

$$\mathbf{X} = \begin{bmatrix} \chi_1 \\ \vdots \\ \chi_i \\ \vdots \\ \chi_N \end{bmatrix}_{N \times m} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \end{bmatrix}_{N \times m} \quad (10)$$

与传统优化算法类似, 河马算法的搜索粒子在初始化阶段随机生成, 公式如下:

$$x_{ij} = lb_j + r \cdot (ub_j - lb_j), \quad \text{for } i = 1, 2, \dots, N, \text{ and } j = 1, 2, \dots, m \quad (11)$$

其中  $x_i$  表示第  $i$  个候选解的位置,  $r$  是 0 到 1 之间的随机数,  $x_j^u$  和  $x_j^l$  分别表示第  $j$  个决策变量的上限和下限.

#### 2.1.2 河马在河流或池塘中的位置更新

河马群由几只成年雌性河马、小河马、多只成年雄性河马和河马领袖组成. 初始个体位置生成后以适应度值为标准筛选出最优个体, 将此个体确定为河马种群领袖. 下式模拟了成年雄性河马被河马领袖赶出种群的社会行为, 雄性河马的位置更新如下所示

$$x_{ij}^{\text{Mhippo}} = x_{ij} + y_1 \cdot (D_{\text{hippo}} - I_1 x_{ij}) \\ \text{for } i = 1, 2, \dots, \left\lfloor \frac{N}{2} \right\rfloor, \text{ and } j = 1, 2, \dots, m \quad (12)$$

其中  $x_{ij}$  表示河马原始位置;  $x_{ij}^{\text{Mhippo}}$  表示被驱赶后的河马位置;  $D_{\text{hippo}}$  表示领袖河马位置, 即迭代以来适应度最佳的河马位置;  $y_1$  表示 0 到 1 之间的随机数;  $I_1, I_2$  表示一个值为 1 或 2 的随机变量.

而小河马往往会离开母河马, 在种群周围随机活动, 下式进行小河马的位置更新

$$x_{ij}^{\text{FHhippo}} = \begin{cases} x_{ij} + h_1 \cdot (D_{\text{hippo}} - I_2 \cdot x_i^{\text{MG}}), & T > 0.6 \\ \Xi, & \text{else} \end{cases} \quad (13)$$



$$\Xi = \begin{cases} x_{ij} + h_2 \cdot (x_i^{\text{MG}} - D_{\text{hippo}}), & r_6 > 0.5 \\ x_j^l + r_7 \cdot (x_j^u - x_j^l), & \text{else} \end{cases}$$

for  $i=1,2,\dots, \left\lfloor \frac{N}{2} \right\rfloor$ , and  $j=1,2,\dots,m$  (14)

其中,  $x_{ij}$  表示为更新前的位置,  $x_{ij}^{\text{FBhippo}}$  表示为更新后的小河马位置,  $h_1, h_2$  为相应随机数组中的元素,  $x_i^{\text{MG}}$  表示为随机选出的一组河马的位置平均值,  $r_6, r_7$  表示为 0 到 1 之间的随机数。

基于迭代数的非线性衰减因子为:

$$T = \exp\left(-\frac{t}{t_{\max}}\right) \quad (15)$$

$T$  表示为基于迭代数的非线性衰减因子,  $t$  为当前迭代数,  $t_{\max}$  为最大迭代数。衰减因子  $T$  大于 0.6 则表明小河马已经远离母亲, 否则转入下一个判断条件: 当  $r_6$  大于 0.5 时, 表明小河马远离母亲, 但是仍在群体周围活动; 否则小河马脱离了群体, 位置在搜索上限与下限之间随机生成。在该阶段, 如果雄性河马或小河马更新后的位置适应度值优于最优个体, 则该个体替换更新为最优个体。

### 2.1.3 河马防御捕食者

该阶段模拟了河马面对不同捕食者时所做的防御行为。首先在搜索上下限之间随机生成捕食者的位置, 公式如下:

$$x_j^{\text{Predator}} = x_j^l + r_8 \cdot (x_j^u - x_j^l) \quad (16)$$

$$d = |x_j^{\text{Predator}} - x_{ij}| \quad (17)$$

其中  $x_j^{\text{Predator}}$  为捕食者的位置,  $r_8$  为 0 到 1 之间的随机数,  $d$  为第  $i$  只河马到捕食者之间的欧氏距离。

河马面对捕食者时会表现出靠近捕食者以压迫其撤退的行为, 但是面对不同的捕食者会采取不同的措施。如果捕食者的适应度值低于河马, 河马会转向捕食者并朝其大步移动逼迫其撤退, 反之河马会出于安全考虑谨慎接近捕食者, 移动范围较小, 如下所示:

$$x_{ij}^{\text{HippoR}} = \begin{cases} \mathbf{R}_L \oplus x_j^{\text{Predator}} + \left(\frac{B}{c - e \times \cos(2\pi g)}\right) \left(\frac{1}{d}\right), & F_{\text{Predator}_j} < F_i \\ \mathbf{R}_L \oplus x_j^{\text{Predator}} + \left(\frac{B}{c - e \times \cos(2\pi g)}\right) \left(\frac{1}{2 \times d + r_9}\right), & F_{\text{Predator}_j} \geq F_i \end{cases} \quad (18)$$

其中  $x_{ij}^{\text{HippoR}}$  代表河马面对捕食者更新后的位置。  $\mathbf{R}_L$  是具有 Levy 分布的随机向量, 表示位置的突然变化, 增强全局搜索能力。  $B, c, e, g$  各代表区间  $[2,$

$4], [1, 2], [2, 3], [-2\pi, 2\pi]$  之间的随机变量。  $r_9$  表示为 0 到 1 之间的随机数。

在这个过程中, 如果更新后的位置适应度值优于当前最优适应度值, 则替换更新为最优个体。该阶段个体的全局搜索能力得到了显著增强, 第二阶段和第三阶段相辅相成, 有效提高了算法的寻优能力。

### 2.1.4 河马逃离捕食者

该阶段模仿河马无法击退捕食者, 就近进入池塘躲避敌人的情况。算法表现为在该个体周围随机生成新的搜索位置, 且搜索范围基于迭代数的增加而逐渐缩小, 相比之下这种策略增强了本地搜索的能力, 公式如下所示

$$x_{ij}^{\text{HippoE}} = x_{ij} + r_{10} \cdot \left[ \frac{x_j^l}{t} + s_1 \cdot \left( \frac{x_j^u}{t} - \frac{x_j^l}{t} \right) \right]$$

$$\text{for } i=1,2,\dots,N, \text{ and } j=1,2,\dots,m \quad (19)$$

其中  $x_{ij}^{\text{HippoE}}$  是局部搜索更新后的位置,  $t$  为迭代次数,  $r_{10}, s_1$  是区间  $[0, 1]$  的随机数。在这个过程中, 如果更新后的个体适应度值优于当前最优个体, 则替换更新。

最后, 在下次迭代开始前, 将经过前三个搜索阶段而产生的最优个体设为领袖, 保存其适应度值, 将其与这次迭代产生的最优个体进行比较, 以此不断更新最优个体, 直到达到最大迭代次数, 从而完成最优值的搜寻。

## 2.2 IHOA

通过上述分析可以看出, HOA 能够凭借其快速的收敛性和高度的灵活性, 在工程问题中实现高效迭代搜索, 从而找到最优解。但是该算法的种群个体更新依赖于大量的随机变量, 在某些复杂场景下收敛速度有限, 且容易陷入局部最优的困境。为了克服这些局限性, 本文提出了改进策略: 首先, 在该算法中引入了自适应权重因子改进第二阶段粒子的更新判断条件; 其次, 引入  $t$  分布变异参数对第三阶段的粒子进行变异更新, 避免其陷入局部最优, 以期增强算法的全局搜索能力和解的精确性。

### 2.2.1 自适应非线性权重

为了提高粒子前期的全局搜索能力以及后期的局部搜索能力, 避免搜索陷入局部最优, 设计一个非线性的权重因子  $r_G$  以更好地平衡算法的探索能力, 如下所示

$$r_G = a - \left[ b \times \left( \frac{t}{t_{\max}} \right) \right]^2 \quad (20)$$

在原始 HOA 的第二阶段中,雄性河马的更新权重为 0 到 1 之间的随机数,而小河马完全脱离种群后是在空间中随机生成新的位置,两者的位置更新有较高的随机性,在优化寻解过程中不能快速收敛至最优解,有一定的改进空间.

改进后的雄性河马个体位置的更新公式如下所示:

$$x_{ij}^{\text{Mhippo}} = x_{ij} + r_G \cdot y_1 \cdot (D_{\text{hippo}} - I_1 x_{ij}) \quad (21)$$

改进后的幼河马个体位置的更新公式如下所示:

$$x_{ij}^{\text{Fhippo}} = \begin{cases} x_{ij} + h_1 \cdot (D_{\text{hippo}} - I_2 \cdot x_i^{\text{MG}}), & T > 0.6 \\ \Xi, & \text{else} \end{cases}$$

$$\Xi = \begin{cases} x_{ij} + h_2 \cdot (x_i^{\text{MG}} - D_{\text{hippo}}), & r_6 > 0.25 \cdot r_G \\ x_j^l + r_7 \cdot (x_j^u - x_j^l), & \text{else} \end{cases} \quad (22)$$

通过引入该非线性权重因子  $r_G$ ,使改进后的雄性河马粒子在迭代前期搜索较激进,着重于全局搜索,在迭代后期搜索趋于保守,着重于局部搜索;小河马前期更侧重于脱离种群,进行全局搜索,后期则在种群周围搜索.改进之后算法在全局与局部的搜索能力得到了平衡.

### 2.2.2 自适应 $t$ 分布变异

$t$  分布<sup>[34]</sup>又称为学生分布,含有参数自由度  $n$ ,自由度  $n$  影响曲线的形态与特性, $n$  越大则曲线中间越高.当  $n$  为 1 时, $t$  分布接近于柯西分布;当  $n$  趋向于无穷时, $t$  分布接近高斯分布.

在原河马优化算法中,当河马面对适应度较差的捕食者时,个体会采取谨慎接近的策略,这可能会导致收敛速度低下甚至是陷入局部最优解.为了避免这种状况,我们在个体谨慎接近捕食者后引入自适应  $t$  分布变异,产生变异个体以增加种群个体的多样性,避免陷入局部最优.粒子更新公式如下

$$x_i^t = x_i + x_i \cdot T(t) \quad (23)$$

其中, $x_i^t$  表示变异后的河马个体位置, $x_i$  表示为第  $i$  个更新前的河马个体的位置, $T(t)$  表示为以算法迭代次数  $t$  为参数自由度的学生分布.

引入自适应  $t$  分布变异后,随着迭代次数的增加,算法将柯西变异与高斯变异的优点逐渐结合起来,在搜索初期具有良好的全局搜索性,而在后期具有较优的局部开发性.利用当前参数的不定性,使得个体能够跳出局部,促进收敛于全局最优,加快了收敛进程.

改进 HOA 框图如图 1 所示.

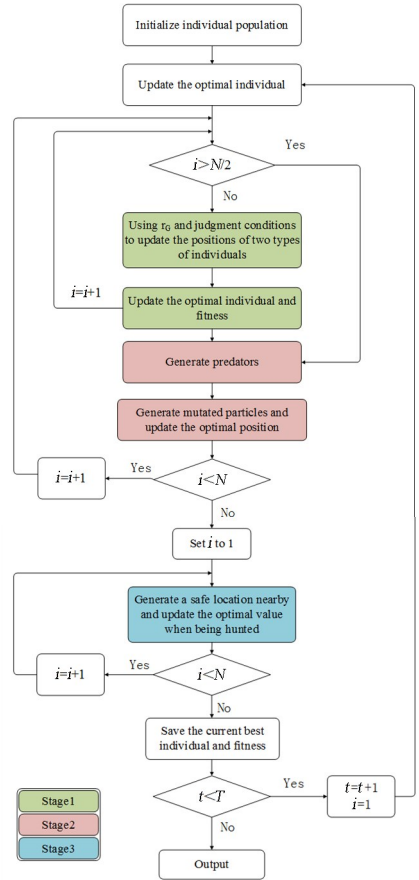


图 1 改进河马优化算法的流程框图  
Fig. 1 Process flowchart for improving hippopotamus optimization algorithm

我们已经深入探讨了朴素贝叶斯分类器(NBC)和河马优化算法(HOA)的原理及应用,并对 HOA 的寻优性能进行了一定改进,提出了一种改进河马优化算法(IHOA).传统的贝叶斯分类器在设定先验概率时往往基于均匀分布或经验选定,这在面对复杂数据集时可能导致分类性能不佳.为了克服这一局限性,我们利用 IHOA 来优化朴素贝叶斯分类器的先验概率.

### 3 基于 IHOA 的贝叶斯分类算法

将改进河马优化算法用于朴素朴素贝叶斯分类的先验概率优化,即利用 IHOA 的优化寻解能力为数据集确定最佳先验概率,从而提出了一种朴素贝叶斯先验概率优化方法(Naive Bayes classifiers-improving hippopotamus optimization algorithm, NBC-IHOA).

算法优化部分具体为,首先通过训练集数据及标签信息得到每个类别的均值  $\mu$  与标准差  $\sigma$ ;初始化解种群个体得到其中一个类别的先验概率  $p$ :

$$\mathbf{p} = [x_1, x_2, \dots, x_N]^T \quad (24)$$

其中  $x_i (1 \leq i \leq N)$  为各个体位置,在算法中  $x_i$  代表了先验概率。

已知其中一个类别的先验概率  $\mathbf{p}$ , 则该二分类数据的朴素贝叶斯分类先验概率为

$$\mathbf{P} = \begin{bmatrix} x_1 & 1 - x_1 \\ x_2 & 1 - x_2 \\ \vdots & \vdots \\ x_N & 1 - x_N \end{bmatrix} \quad (25)$$

将该先验概率组  $\mathbf{P}$  与训练参数  $\mu, \sigma$  作为验证集的朴素贝叶斯分类器输入参数, 得到  $P(y_k | C_j)$ , 即属于第  $C_j$  类的前提下待判样本  $\mathbf{y}$  在属性  $A_i$  取值为  $y_k$  的条件概率, 如下所示:

$$P(y_k | C_1) = g(y_k, \mu_{C_1}, \sigma_{C_1}) \quad (26)$$

$$P(y_k | C_2) = g(y_k, \mu_{C_2}, \sigma_{C_2}) \quad (27)$$

通过输入先验概率与条件概率, 对于每个测试样本  $y_k$  进行针对每个类别  $C_j$  的后验概率计算, 待判样本  $\mathbf{y}$  属于类别  $C_1$  与类别  $C_2$  的条件概率如下:

$$P(C_1 | \mathbf{y}) = \frac{x_i}{P(\mathbf{y})} \prod_{k=1}^N P(y_k | C_1) \quad (28)$$

$$P(C_2 | \mathbf{y}) = \frac{1 - x_i}{P(\mathbf{y})} \prod_{k=1}^N P(y_k | C_2) \quad (29)$$

此时利用 NBC 对向量  $\mathbf{y}$  进行分类, 后验概率的比较有

$$P(C_i | \mathbf{y}) > P(C_j | \mathbf{y}), \quad \text{for } i = 1, j = 2, \text{ or } j = 1, i = 2 \quad (30)$$

当  $P(C_i | \mathbf{y})$  最大时, 则认为  $\mathbf{y}$  的标号为  $C_i$ , 最终测试样本会被分在最大化的后验概率所包含的类中. 由此可得到预测标签并与验证集标签比对得到分类准确率 (accuracy)  $A(x_i)$  并保存最佳结果, 计算方法如下:

$$A(x_i) = \frac{TP + TN}{TP + TN + FP + FN} \quad (31)$$

其中,  $TP$  是被预测为正类的正类样本;  $FP$  是被预测为正类的负类样本;  $TN$  是被预测为负类的负类样本;  $FN$  是被预测为负类的正类样本. 算法优化过程中的适应度值 (fitness) 即为分类准确率 (accuracy). 将准确率作为 IHOA 的评价指标, 在验证集上对先验概率进行迭代寻优后得到最优分类准确率  $A_{\text{best}}$  和对应的最优先验概率  $\mathbf{p}_{\text{best}}$ . 最后将该最佳先验概率  $\mathbf{p}_{\text{best}}$  用于测试集分类。

NBC-IHOA 方法流程图如图 2 所示, 具体步骤描述如下:

Step 1 将系统所得数据样本进行预处理, 合理划分出训练集、验证集与测试集;

Step 2 初始化种群个体得到初始先验概率组  $\mathbf{p}$ , 结合由训练集样本得到的模型参数, 对验证集分类得到各个体的适应度;

Step 3 更新当前最优适应度值;

Step 4 一半种群个体作为雄性河马与幼河马使用公式 (21) 和公式 (22) 更新位置  $\mathbf{p}_{\text{new1}}$ , 结合模型参数对验证集进行分类, 得到各个体的适应度, 比较并更新较优位置与较优适应度;

Step 5 重复 Step 4 直到达到规定循环次数, 保存第一阶段的最优个体与适应度;

Step 6 使用公式 (16) 随机生成捕食者位置, 根据捕食者位置情况, 另一半种群个体根据公式 (18) 更新位置, 并通过公式 (23) 产生变异个体, 得到新位置  $\mathbf{p}_{\text{new2}}$ ;

Step 7 结合模型参数与  $\mathbf{p}_{\text{new2}}$ , 对验证集分类得到各个体的适应度, 比较并更新较优值;

Step 8 重复 Step 4 到 Step 7 之间步骤, 直到达到规定循环次数, 保存一、二阶段最优的位置与适应度;

Step 9 将  $i$  重设为 1;

Step 10 使用公式 (19) 更新个体位置  $\mathbf{p}_{\text{new3}}$ ;

Step 11 按照同样方式更新各个体的适应度, 比较并更新较优值;

Step 12 重复 Step 10、Step 11 的步骤, 直到达到规定循环次数, 保存第三阶段最优的位置与适应度;

Step 13 重复 Step 3 到 Step 12, 直到达到规定迭代次数, 输出最优先验概率  $\mathbf{p}_{\text{best}}$  与适应度  $A_{\text{best}}$ ;

Step 14 将由 IHOA 优化得到的先验概率  $\mathbf{p}_{\text{best}}$  输入测试集贝叶斯模型, 将测试集分为  $m$  类, 对比标签评估其分类准确率。

## 4 仿真算例

### 4.1 系统设置

本节采用一个以升压电路为核心的二分类系统来进行算例仿真. 系统为一个带有切换负载的 boost 电路, 拓扑结构如图 3 所示. 该仿真系统的输入量为恒定电压, 经过电路升压后输出. 输出负载串联一个受切换开关作用的负载, 因此切换开关可

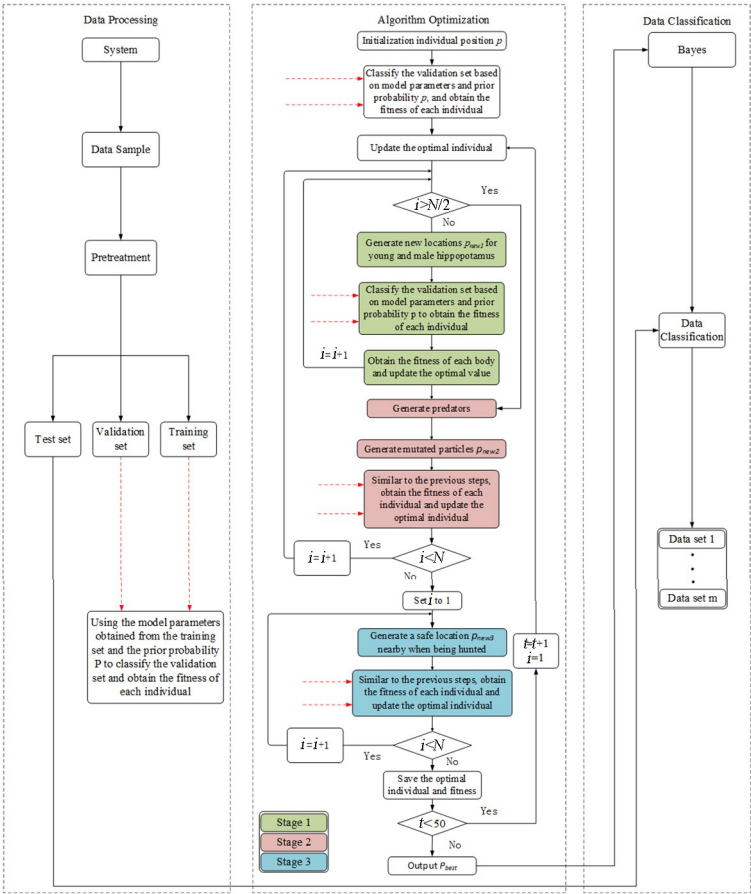


图2 NBC-IHOA流程图  
Fig. 2 NBC-IHOA process diagram

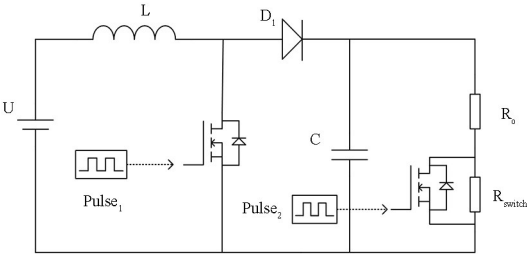


图3 仿真电路拓扑结构图  
Fig. 3 Simulation circuit topology diagram

以通过若干次切换调整电压输出。由于仿真系统会有不可避免的随机噪声干扰,在同样参数及结构的实际系统中,多次仿真的数据结果不尽相同,因此本节对输入电压与输出电压施加随机噪声以模拟实际系统,具体为每次采集数据时改变噪声的随机种子值,施加不同高斯噪声扰动以模仿实际系统的数据输出。采集多次完整数据并保留一次完整过程的数据为测试集,其余数据打乱重组,划分并设置为训练集与验证集。

系统参数设置如下:输入电压  $U$  为 40 V,电感值  $L$  为  $3e-3$  H,电容值  $C$  为  $1e-4$  F,开关频率  $f$  为 10 kHz,占空比  $D$  为 0.6,主负载  $R_o$  为 50  $\Omega$ ,

仿真时长为 0.3 s,输出结果受信噪比(SNR)为 5 dB 的高斯白噪声扰动。本文设置切换负载  $R_{switch}$  分别为 20  $\Omega$ 、10  $\Omega$  以及 2  $\Omega$  三种情况,通过切换信号  $Pulse_2$  来调整切换系统驻留在某个子系统上的时长,产生不同情况下的仿真数据,从而进行算法优化性与分类准确率的对比。每种情况下划分得到的训练集为 24 000 个样本点,验证集与测试集各为 6000 个样本点,训练集、验证集及测试集的数据比重为 4 : 1 : 1。

4.2 IHOA 寻优对比

本文设置不同切换信号来调整切换系统驻留在某个子系统上的时长,产生不同情况下的仿真数据,从而进行算法优化性与分类准确率的对比。IHOA 与模拟退火算法(SAA)<sup>[35]</sup>、粒子群算法(PSOA)<sup>[36]</sup>、HOA<sup>[32]</sup>、高尔 夫 优 化 算 法(GOA)<sup>[37]</sup>、霜冰优化算法(RIME)<sup>[38]</sup>在不同驻留时间比及不同切换负载共 9 种情况下分别运行 10 次取适应度曲线的平均值,对比结果如图 4 所示。图中横轴代表迭代次数,纵轴代表分类的准确率。



IHOA 及其他群优化算法种群规模  $n=20, t_{\max}=50$ , 搜索范围为  $[0,1]$ . 我们以驻留比  $0.6:0.4$ , 负载  $10\ \Omega$  的情况为例, 对原始输出信号与分类结果信号进行重合对比, 如图 5 及图 6 所示. 由图 5 可知, 原始数据存在阶跃波段, 考虑到后续的辨识工作, 该波段无法剔除从而对分类造成了干扰. 由图 6 可知经 NBC-IHOA 分类后, 各个子系统的分类

结果与原始数据基本相吻合.

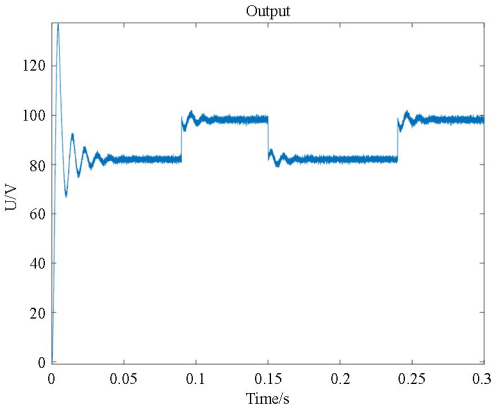


图 5 驻留比  $0.6:0.4$ , 负载  $10\ \Omega$  情况下的输出  
Fig. 5 Output under a residency ratio of  $0.6:0.4$  and a load of  $10\ \Omega$

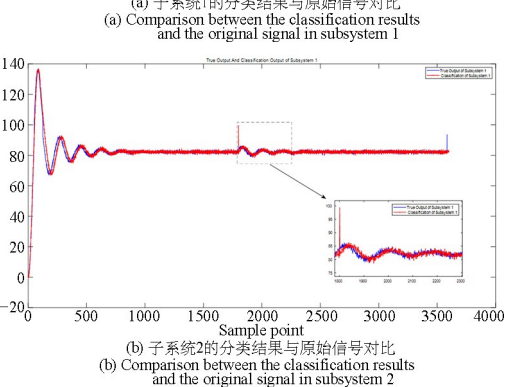
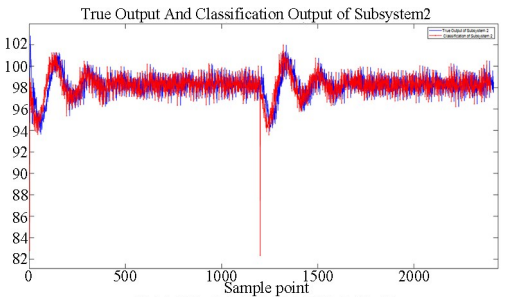


图 6 基于 NBC-IHOA 的分类结果与原始信号对比  
Fig. 6 Comparison between classification results and original signals based on NBC-IHOA

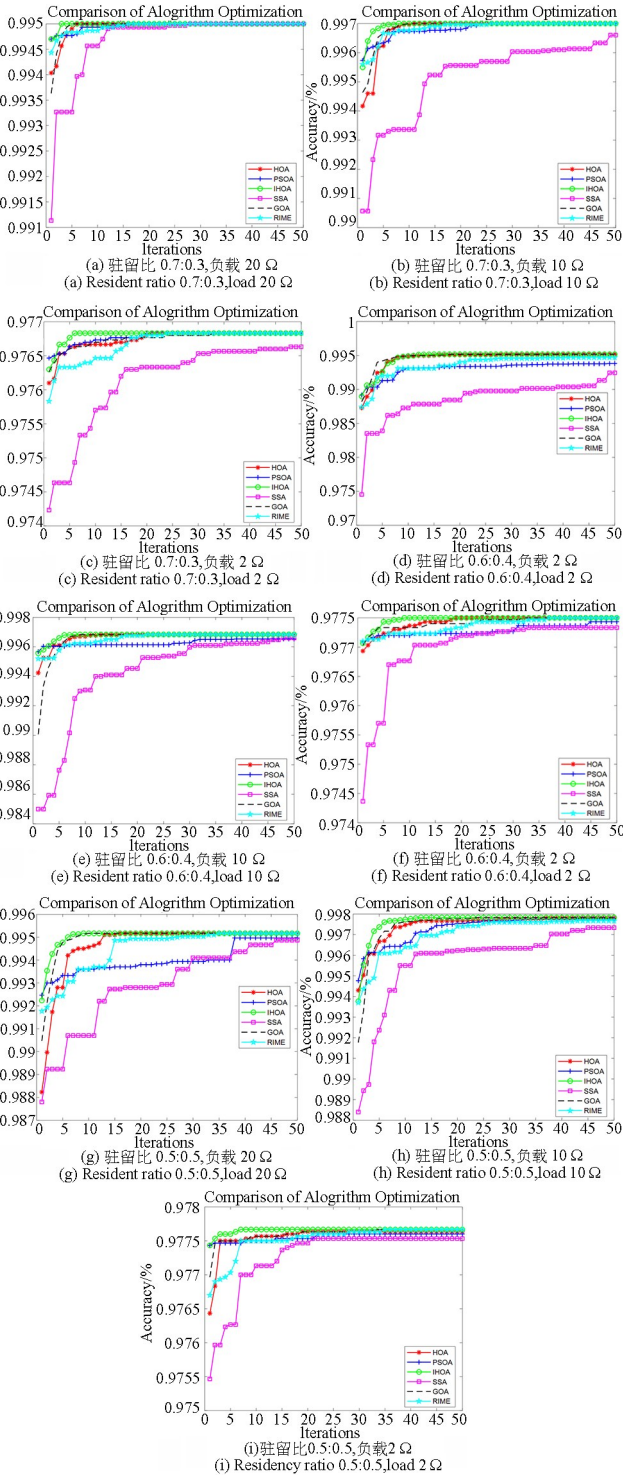


图 4 不同情况下 NBC-IHOA 与其他算法寻优对比结果  
Fig. 4 Comparison of optimization results between NBC-IHOA and other algorithms under different conditions

IHOA 与对比算法优化得到的先验概率  $p_{\text{best}}$  及适应度值 (即准确度), 如表 1 所示. 该表给出了不同驻留时间比及不同切换负载共 9 种情况下, 各优化算法所得的最优先验概率与最优准确度. 子系统数量为 2, 因此子系统的贝叶斯先验概率之比为  $p_{\text{best}}:(1-p_{\text{best}})$ . 分析表格结果得知相比于经典先验概率, 本文采用优化方法所得到的先验概率能得到更高分类准确率, 且 IHOA 的寻优性能相对其他优化算法更稳定.

由表 1 可知在不同情况下, 相比于准确率接近的 GOA 与 HOA, IHOA 在收敛速度上表现较好, 得到最优结果所需的平均迭代次数最少, 面对工程

运筹问题时能表现出快速稳定的寻解性能。

表 1 各算法寻优结果及对应分类准确率  
Table 1 Optimization results and corresponding classification accuracy of each algorithm

Resident ratio	Algorithm	R/Ω								
		20			10			2		
		$p_{best}$	Accuracy/%	$t$	$p_{best}$	Accuracy/%	$t$	$p_{best}$	Accuracy/%	$t$
0.7 : 0.3	NBC	0.7	98.93		0.7	98.38		0.7	96.78	
	PSOA	6.23e-02	99.50	17	1.06e-02	99.70	24	0.4794	97.68	20
	HOA	7.14e-02	99.50	6	1.18e-02	99.70	11	0.4576	97.68	21
	IHOA	8.95e-02	99.50	3	1.03e-02	99.70	8	0.4781	97.68	6
	SAA	9.48e-02	99.50	28	1.58e-02	99.66	/	0.4599	97.66	/
	RIME	9.70e-02	99.50	16	8.90e-03	99.70	17	0.4793	97.68	24
	GOA	6.24e-02	99.50	5	1.06e-02	99.70	10	0.4788	97.68	33
0.6 : 0.4	NBC	0.6	97.67		0.6	98.48		0.6	97.23	
	PSOA	1.70e-03	99.38	/	1.14e-02	99.65	/	0.3329	97.74	/
	HOA	1.16e-05	99.52	44	7.6e-03	99.68	18	0.3323	97.75	19
	IHOA	1.10e-05	99.52	15	1.9e-03	99.68	8	0.3320	97.75	9
	SAA	4.60e-03	99.24	/	1.38e-02	99.66	/	0.3117	97.72	/
	RIME	1.84e-04	99.48	/	3.9e-03	99.68	18	0.3317	97.75	37
	GOA	9.59e-06	99.52	19	7.2e-03	99.68	9	0.3328	97.75	32
0.5 : 0.5	NBC	0.5	98.47		0.5	98.55		0.5	97.63	
	PSOA	2.30e-03	99.50	/	8.95e-04	99.78	42	0.3977	97.76	/
	HOA	1.10e-03	99.52	16	8.82e-04	99.78	30	0.4069	97.76	/
	IHOA	6.49e-04	99.52	9	8.71e-04	99.78	13	0.4202	97.77	7
	SAA	4.20e-03	99.47	/	2.80e-03	99.73	/	0.3941	97.75	/
	RIME	5.62e-07	99.52	35	1.87e-04	99.77	/	0.4232	97.77	35
	GOA	8.88e-04	99.52	10	8.77e-04	99.78	24	0.4211	97.77	33

4.3 NBC-IHOA 方法对比

为了进一步验证 NBC-IHOA 的分类性能,将 9 种情况下使用 NBC-IHOA 方法得到的最优先验概率之比用于对应测试集的分类,并将经典朴素贝叶斯分类器(NBC)、高斯混合模型(GMM)、K 均值聚类(K-means)、支持向量机(SVM)、最小二乘支持向量机(LSSVM)以及极限学习机(ELM)等主流分类算法作为参照进行验证,结果如表 2 所示. 分析可知在以不能剔除异常波段为前提的系统数据集中,本文的 NBC-IHOA 具有较好的分类表现:当负载变化较大,两类子系统数据差异较明显时,NBC-IHOA 在分类准确率上相较于其他算法有一定的优势;而当切换负载量较小时,NBC-IHOA 的准确率与稳定性进一步凸显. 该提升得益于最优先验概率的选择,为后续辨识保障了数据集的准确性.

表 2 不同情况数据下各方法的分类准确率对比  
Table 2 Comparison of classification accuracy of various methods under different data conditions

Resident ratio	Algorithm	R/Ω		
		20	10	2
		Accuracy/%	Accuracy/%	Accuracy/%
0.7 : 0.3	NBC-IHOA	99.35	99.67	97.52
	NBC	98.92	98.52	96.65
	GMM	98.65	31.60	35.00
	K-means	98.50	97.87	69.45
	LSSVM	98.60	98.50	68.75
	SVM	98.48	97.80	94.53
	ELM	99.18	98.78	91.12
0.6 : 0.4	NBC-IHOA	99.35	99.70	97.62
	NBC	98.60	98.47	97.05
	GMM	98.70	96.57	44.87
	K-means	98.52	97.83	58.90
	LSSVM	98.57	98.38	58.00

续表

Resident ratio	Algorithm	R/ $\Omega$		
		20	10	2
		Accuracy/%	Accuracy/%	Accuracy/%
0.6 : 0.4	VM	98.52	97.80	95.33
	ELM	99.22	98.73	95.50
	NBC-IHOA	99.43	99.68	97.52
	NBC	98.48	98.30	97.20
	GMM	97.18	96.90	54.82
0.5 : 0.5	K-means	98.50	97.80	49.45
	LSSVM	98.53	97.98	94.85
	SVM	98.50	97.78	95.15
	ELM	99.18	98.70	95.65

5 结论

贝叶斯方法在复杂数据集中,尤其是在面对不确定性和小样本情况时,先验概率的构造尤为重要.针对这一问题我们提出了一种方法,首先通过数据集的训练样本得到各标签的均值、方差等模型参数,然后在验证集上以分类准确率为目标,采用改进的河马优化算法(IHOA)搜寻最佳分类准确率及对应先验,最后以此作为测试集的贝叶斯最佳先验并得到分类结果.为验证所提方法的有效性,以二状态电路系统数据集为例,在保留易干扰分类的阶跃波段情况下,利用 IHOA 来寻找特定数据集的最优先验概率,并将 IHOA 与其他群优化算法进行了寻优性能的比较,此外将该方法与其他主流分类算法进行了对比.仿真结果表明,IHOA 在寻优方面表现出了良好的寻优性能,而改进优化算法构造先验的朴素贝叶斯分类器(NBC-IHOA)相比其他分类方法展现出了较高的分类准确性.

基于上述研究我们了解到,结合 NBC-IHOA 能有效处理复杂数据集中的不确定性和非常规小样本情况.因此该算法能为动力学系统的模式识别和故障分类诊断提供有力的支持,在动力学与控制领域中具有相当程度的应用前景.随着动力学系统的日益复杂,业界对高精度分类算法的需求将更加迫切,我们期望本文所提的 NBC-IHOA 分类方法能够为业界提供有益的参考,助力相关领域的技术进步和应用创新.

参考文献

[1] PASHA A, LATHA P H. Bio-inspired dimensionality reduction for Parkinson's disease (PD) classification [J]. Health Information Science and Systems, 2020, 8(1): 13.

[2] 陈迪,陈云虹,叶青,等.融合 k-means 与多维特征分析的生物医学图像分类算法[J]. 信息技术, 2024, 48(7): 65—70.

CHEN D, CHEN Y H, YE Q, et al. Biomedical image classification algorithm combining k-means and multidimensional feature analysis [J]. Information Technology, 2024, 48(7): 65—70. (in Chinese)

[3] 张菡,舒静超,张晓静,等.鸭源马尿气球菌的鉴定及全基因组测序分析[J]. 中国预防兽医学报, 2024, 46(2): 128—139.

ZHANG H, SHU J C, ZHANG X J, et al. Identification and whole genome sequencing of a duck-derived *Aerococcus urinaequi* strain [J]. Chinese Journal of Preventive Veterinary Medicine, 2024, 46(2): 128—139. (in Chinese)

[4] NOFRIANI, BUDI KURNIAWAN N. Harnessing multi-label classification approaches for economic phenomena categorization [J]. ASEAN Journal on Science and Technology for Development, 2021, 38(2): 73—82.

[5] 江丽桃,曾晶.数字经济下的电商客户行为分析研究[J]. 商业观察, 2024(17): 93—95, 108.

JIANG L T, ZENG J. Analysis and research on E-commerce customer behavior under digital economy [J]. Business Observation, 2024(17): 93—95, 108. (in Chinese)

[6] 陈宇龙,孙广宇.中国股票市场操纵识别研究——基于机器学习分类算法[J]. 中央财经大学学报, 2023(3): 56—67.

CHEN Y L, SUN G Y. Manipulation recognition of China's stock market: based on machine learning classification algorithm [J]. Journal of Central University of Finance & Economics, 2023(3): 56—67. (in Chinese)

[7] KOROM P. The prestige elite in sociology: Toward a collective biography of the most cited scholars (1970-2010) [J]. The Sociological Quarterly, 2020, 61(1): 128—163.

[8] 周一夫,谭春辉,江婷,等.基于机器学习分类算

- 法的高质量专利成果筛选研究[J]. 现代情报, 2024, 44(2): 81—91.
- ZHOU Y F, TAN C H, JIANG T, et al. Research on the screening method of high-quality patent results based on machine learning classification algorithms [J]. Journal of Modern Information, 2024, 44(2): 81—91. (in Chinese)
- [9] 周刚, 张晓斌, 曲行达, 等. 基于机器学习和儿童神经心理行为检查量表鉴别孤独症谱系障碍和全面发育迟缓儿童的研究[J]. 中国当代儿科杂志, 2023, 25(10): 1028—1033.
- ZHOU G, ZHANG X B, QU X D, et al. Differential diagnosis of autism spectrum disorder and global developmental delay based on machine learning and Children Neuropsychological and Behavioral Scale [J]. Chinese Journal of Contemporary Pediatrics, 2023, 25(10): 1028—1033. (in Chinese)
- [10] AHMED M, SERAJ R, ISLAM S M S. The k-means algorithm: a comprehensive survey and performance evaluation [J]. Electronics, 2020, 9(8): 1295.
- [11] 赵志宏, 杨绍普. 基于 SVM 的混沌时间序列分析[J]. 动力学与控制学报, 2009, 7(1): 5—8.
- ZHAO Z H, YANG S P. Chaotic time series analysis based on support vector machine [J]. Journal of Dynamics and Control, 2009, 7(1): 5—8. (in Chinese)
- [12] COSTA V G, PEDREIRA C E. Recent advances in decision trees: an updated survey [J]. Artificial Intelligence Review, 2023, 56(5): 4765—4800.
- [13] 诸震宇, 王如彬. 基于核聚类的嗅觉神经网络对气味模式的识别[J]. 动力学与控制学报, 2020, 18(1): 93—101.
- ZHU Z Y, WANG R B. Odor pattern recognition of the olfactory neural network based on kernel clustering [J]. Journal of Dynamics and Control, 2020, 18(1): 93—101. (in Chinese)
- [14] AKER E, OTHMAN M L, VEERASAMY V, et al. Fault detection and classification of shunt compensated transmission line using discrete wavelet transform and Naive Bayes classifier [J]. Energies, 2020, 13(1): 243.
- [15] 许又文, 王尉又, 季日臣. 基于朴素贝叶斯机器学习的水电工程安全事故分类研究[J]. 科技通报, 2023, 39(9): 77—82.
- XU Y W, WANG W Y, JI R C. Research on engineering safety accident classification based on Naive Bayes machine [J]. Bulletin of Science and Technology, 2023, 39(9): 77—82. (in Chinese)
- [16] JEFRIYANTO J, AINUN N, AL ARDHA M A. Application of Naive Bayes classification to analyze performance using stopwords [J]. Journal of Information System, Technology and Engineering, 2023, 1(2): 49—53.
- [17] KERFA D. Moving objects detection in thermal scene videos using unsupervised Bayesian classifier with bootstrap Gaussian expectation maximization algorithm [J]. Multimedia Tools and Applications, 2024, 83(2): 6335—6350.
- [18] TAN Q, MU X W, FU M, et al. A new sensor fault diagnosis method for gas leakage monitoring based on the naive Bayes and probabilistic neural network classifier [J]. Measurement, 2022, 194: 111037.
- [19] RUAN S F, LI H W, LI C Q, et al. Class-specific deep feature weighting for Naive Bayes text classifiers [J]. IEEE Access, 2020, 8: 20151—20159.
- [20] KIM H C, PARK J H, KIM D W, et al. Multilabel naive Bayes classification considering label dependence [J]. Pattern Recognition Letters, 2020, 136: 279—285.
- [21] REDDY G K, THILAGAM S P. Naive Bayes classifier to mitigate the DDoS attacks severity in ad-hoc networks [J]. International Journal of Communication Networks and Information Security, 2020, 12(2): 221—226.
- [22] ZHANG J L, HÄRDLE W K. The Bayesian Additive Classification Tree applied to credit risk modeling [J]. Computational Statistics & Data Analysis, 2010, 54(5): 1197—1205.
- [23] BALAJI V R, SUGANTHI S T, RAJADEVI R, et al. Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier [J]. Measurement, 2020, 163: 107922.
- [24] LLORENTE F, MARTINO L, CURBELO E, et al. On the safe use of prior densities for Bayesian model selection [J]. Wiley Interdisciplinary Reviews: computational Statistics, 2023, 15(1): e1595.
- [25] VILLA C, WALKER S. An objective Bayesian criterion to determine model prior probabilities [J]. Scandinavian Journal of Statistics, 2015, 42(4): 947—



966.

[26] NGUYEN-TRANG T, VO-VAN T. A new approach for determining the prior probabilities in the classification problem by Bayesian method [J]. *Advances in Data Analysis and Classification*, 2017, 11(3): 629—643.

[27] BOLUKI S, ESFAHANI M S, QIAN X N, et al. Incorporating biological prior knowledge for Bayesian learning via maximal knowledge-driven information priors [J]. *BMC Bioinformatics*, 2017, 18: 552.

[28] MARUYAMA Y, STRAWDERMAN W E. Robust Bayesian variable selection with sub-harmonic priors [EB/OL]. (2020-11-25) [2024-09-03]. <https://arxiv.org/abs/1009.1926v4>.

[29] TRAN B H, ROSSI S, MILIOS D, et al. All you need is a good functional prior for Bayesian deep learning [J]. *Journal of Machine Learning Research*, 2022, 23(74): 1—56.

[30] ESFAHANI M S, DOUGHERTY E R. Incorporation of biological pathway knowledge in the construction of priors for optimal Bayesian classification [J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2014, 11(1): 202—218.

[31] WONG T T, CHANG L H. Individual attribute prior setting methods for naïve Bayesian classifiers [J]. *Pattern Recognition*, 2011, 44(5): 1041—1047.

[32] AMIRI M H, MEHRABI HASHJIN N, MONTAZERI M, et al. Hippopotamus optimization algorithm: a novel nature-inspired optimization algorithm [J]. *Scientific Reports*, 2024, 14(1): 5032.

[33] 徐光美, 杨炳儒, 秦奕青, 等. 基于互信息的多关系朴素贝叶斯分类器 [J]. *北京科技大学学报*, 2008, 30(8): 963—966.

XU G M, YANG B R, QIN Y Q, et al. Multi-relational Naïve Bayesian classifier based on mutual information [J]. *Journal of University of Science and Technology Beijing*, 2008, 30(8): 963—966. (in Chinese)

[34] DUONG V, HARIMA K, CHOY S, et al. GNSS best integer equivariant estimation using multivariant *t*-distribution: a case study for precise point positioning [J]. *Journal of Geodesy*, 2021, 95(1): 10.

[35] SHI K, WU Z T, JIANG B P, et al. Dynamic path planning of mobile robot based on improved simulated annealing algorithm [J]. *Journal of the Franklin Institute*, 2023, 360(6): 4378—4398.

[36] MIRMOHAMMADI S L, SAFDARI J, MALLAH M H, et al. The design and optimization of square cascades by PSO and GOA to provide fresh fuel for a nuclear power reactor [J]. *Separation Science and Technology*, 2023, 58(11): 2008—2026.

[37] MONTAZERI Z, NIKNAM T, AGHAEI J, et al. Golf optimization algorithm: a new game-based metaheuristic algorithm and its application to energy commitment problem considering resilience [J]. *Biomimetics*, 2023, 8(5): 386.

[38] SU H, ZHAO D, HEIDARI A A, et al. RIME: a physics-based optimization [J]. *Neurocomputing*, 2023, 532: 183—214.