

基于改进灰狼算法优化多核支持向量回归机及其应用^①



王 颖, 朱刘涛, 童 勤, 张 强

(东北石油大学 计算机与信息技术学院, 大庆 163318)

通讯作者: 王 颖, E-mail: nepu_wy@163.com

摘 要: 为更好发现数据中的复杂规律, 避免核函数选择的盲目性和局部最优等非线性优化问题, 本文提出一种基于改进灰狼算法优化多核支持向量回归机算法. 首先, 基于全局核函数和局部核函数构建多核支持向量机采油速度预测模型; 其次, 利用基于云模型和二次插值算法改进灰狼优化算法对核函数权值和参数的选取进行优化; 最后, 应用灰色关联分析理论确定采油速度影响因素集, 并作为多核支持向量回归机预测模型的输入. 与 6 种采油速度预测方法进行对比, 所提方法具有较好的全局寻优能力和较高的预测率的优点.

关键词: 支持向量回归机; 多核学习; 灰狼优化算法; 云模型; 预测

引用格式: 王颖, 朱刘涛, 童勤, 张强. 基于改进灰狼算法优化多核支持向量回归机及其应用. 计算机系统应用, 2021, 30(1): 256-263. <http://www.c-s-a.org.cn/1003-3254/7774.html>

Optimization of Multi-Core Support Vector Regression Based on Improved Gray Wolf Algorithm and Its Application

WANG Ying, ZHU Liu-Tao, TONG Qin, ZHANG Qiang

(School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

Abstract: In order to find the complex rules in the data, avoid the blindness of kernel function selection and local optimal nonlinear optimization problems, this study proposes an improved gray wolf algorithm to optimize the multi-core support vector regression machine algorithm. Firstly, a multi-core SVM oil production speed prediction model is constructed based on the global kernel function and the local kernel function. Secondly, the gray wolf optimization algorithm is improved based on the cloud model and the quadratic interpolation algorithm to optimize the selection of the weights and parameters of the kernel function. Finally, the influencing factors set of oil production speed is determined by the grey correlation analysis theory and used as the multi-core SVM prediction model. Compared with 6 kinds of prediction methods of oil production rate, the proposed method has the advantages of better global optimization ability and higher prediction rate.

Key words: support vector regression; multi-core learning; grey wolf optimizer; cloud model; forecasting

采油速度是衡量油田开发速度和产能的关键指标, 其大小是油藏的年产量与地质储量的比值, 是油田开发

过程中必不可少的生产指标. 储层构造特征和物性特征、流体性质、井网部署情况、国家政策以及油价都会

① 基金项目: 国家自然科学基金(61702093); 黑龙江省自然科学基金(F2018003); 大庆市指导性科技项目(zd-2019-09); 东北石油大学青年科学基金(2020QNL-10)

Foundation item: National Natural Science Foundation of China (61702093); Natural Science Foundation of Heilongjiang Province (F2018003); Guiding Science and Technology Project of Daqing City (zd-2019-09); Young Scientists Fund of Northeast Petroleum University (2020QNL-10)

收稿时间: 2020-06-07; 修改时间: 2020-07-07, 2020-07-17; 采用时间: 2020-07-21; csa 在线出版时间: 2020-12-31

对采油速度产生影响^[1]。采油速度过低会增加生产成本,采油速度过高可能会导致含水快速上升、驱油效率下降,因此,确定油田合理的采油速度对油田高效开发意义重大。目前计算合理采油速度的常用方法主要有静态影响因素分析法、动态影响因素分析法、数值模拟法、类比法、线性回归法、多元逐步回归法和数理统计等^[2,3],这些预测方法多是以公式的形式进行表述。文献[4]利用含水率分流方程,建立了采油速度与油藏特征、含水率、含水饱和度及含水上升速度等影响因素的关系表达式,但预测效果受限于参与拟合数据的质量。支持向量回归机是建立在统计学习理论基础上的学习算法,与神经网络等算法相比具有更好的性能^[5],但实际应用中样本数据含有多源异构信息,采用单一核函数无法发现数据中隐含的复杂规律,进而准确地描述数据的内在特性,增强决策函数的可理解性^[6]。已有理论研究及应用表明,多个核函数的优化组合代替单个核函数能增强决策函数的可解释性,并能使支持向量回归机获取更优的性能^[7,8]。本文提出基于改进灰狼算法优化多核支持向量回归机模型,并将其应用到采油速度预测。首先应用灰色关联分析理论利用实际地质数据和开发数据对采油速度影响因素进行重要性分析,然后利用历史数据来训练构建的支持向量回归机模型来完成预测,使得预测模型既考虑动静数据,又能自适应历史数据,克服了历史拟合数据质量对已有采油速度预测算法的影响,具有很好泛化能力和推广前景。

1 多核支持向量回归机原理

支持向量回归机通过核函数的隐式映射将数据映射到高维特征空间,完成样本训练和数据预测。设 X 是输入空间, H 是更高维度空间,存在一个映射 $\varphi(x)$ 将数据从 X 映射到 H 。则回归函数为 $f(x) = \omega * \varphi(x) + b$, ω 为权值变量, b 为偏置。为求解标准 SVR 模型引入松弛变量 ξ 和 ζ , 因此新的目标函数就变成:

$$\min_{w,b,\xi,\zeta} L(w,b,\alpha^*,\alpha,\beta^*,\beta) = -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \varphi(x_i) \varphi(x_j) - \zeta \sum_{i=1}^l (\alpha_i^* + \alpha_i) + \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \quad (5)$$

将式(5)带入式(4)可得:

$$\begin{cases} \max_{\alpha^*,\alpha} -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \varphi(x_i) \varphi(x_j) - \zeta \sum_{i=1}^l (\alpha_i^* + \alpha_i) + \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ \text{s.t. } 0 \leq \alpha_i^*, \alpha_i \leq C, \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0, i = 1, 2, \dots, l \end{cases} \quad (6)$$

将式(6)转化为最小值优化可得:

$$\begin{aligned} \min_{w,b,\xi,\zeta} & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \zeta_i) \\ \text{s.t.} & \begin{cases} y_i - w * \varphi(x_i) - b \leq \zeta + \xi_i \\ w * \varphi(x_i) + b - y_i \leq \zeta + \zeta_i \\ \zeta_i, \xi_i \geq 0, i = 1, 2, \dots, l \end{cases} \end{aligned} \quad (1)$$

l 是样本点的个数, C 是惩罚因子, C 越大表示整个优化过程中对于总误差的关注度越高。采用拉格朗日对偶将式(1)转化为对应的拉格朗日目标函数为:

$$\begin{aligned} L(w,b,\zeta,\xi,\alpha^*,\alpha,\beta^*,\beta) &= \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \zeta_i) \\ &- \sum_{i=1}^l \alpha_i^* (\zeta + \xi_i - y_i + w * \varphi(x_i) + b) \\ &- \sum_{i=1}^l \alpha_i (\zeta + \zeta_i + y_i - w * \varphi(x_i) - b) - \sum_{i=1}^l \beta_i^* * \xi_i - \sum_{i=1}^l \beta_i * \zeta_i \\ \text{s.t. } & \alpha_i^*, \alpha_i, \beta_i^*, \beta_i \geq 0, i = 1, 2, \dots, l \end{aligned} \quad (2)$$

其中 $\alpha^*, \alpha, \beta^*, \beta$ 是拉格朗日乘子。由约束条件

$$\begin{cases} y_i - w * \varphi(x_i) - b \leq \zeta + \xi_i \\ w * \varphi(x_i) + b - y_i \leq \zeta + \zeta_i \end{cases} \text{ 可知:} \\ \alpha_i^*, \alpha_i, \beta_i^*, \beta_i \geq 0, i = 1, 2, \dots, l$$

$$\max_{\alpha^*,\alpha,\beta^*,\beta} L(w,b,\alpha^*,\alpha,\beta^*,\beta) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \zeta_i)$$

则式(1)可转化为:

$$\begin{cases} \min_{w,b,\xi,\zeta} \max_{\alpha^*,\alpha,\beta^*,\beta} L(w,b,\alpha^*,\alpha,\beta^*,\beta) \\ \text{s.t. } \alpha_i^*, \alpha_i, \beta_i^*, \beta_i \geq 0, i = 1, 2, \dots, l \end{cases} \quad (3)$$

式(3)经过拉格朗日对偶可得:

$$\begin{cases} \max_{\alpha^*,\alpha,\beta^*,\beta} \min_{w,b,\xi,\zeta} L(w,b,\alpha^*,\alpha,\beta^*,\beta) \\ \text{s.t. } \alpha_i^*, \alpha_i, \beta_i^*, \beta_i \geq 0, i = 1, 2, \dots, l \end{cases} \quad (4)$$

将式(2)对 w, b, ξ, ζ 求偏导, 将偏导结果带入式(2)可得:

$$\min_{\alpha^*, \alpha} \left\{ \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \varphi(x_i) \varphi(x_j) + \varsigma \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \right\}$$

$$\text{s.t.} \begin{cases} 0 \leq \alpha_i^*, \alpha_i \leq C \\ \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0, i = 1, 2, \dots, l \end{cases} \quad (7)$$

令核函数 $k(x_i, x_j) = \varphi(x_i) * \varphi(x_j)$, 将其带入式 (7) 可得:

$$\left\{ \begin{array}{l} \min_{\alpha^*, \alpha} \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(x_i, x_j) + \varsigma \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ \text{s.t. } 0 \leq \alpha_i^*, \alpha_i \leq C, \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0, i = 1, 2, \dots, l \end{array} \right. \quad (8)$$

预测函数为:

$$f(x) = \sum_{i=0}^l (\alpha_i^* - \alpha_i) k(x_i, x) + b \quad (9)$$

由对称半正定矩阵的性质可得:
 $K(x_i, x_j) = \sum_{z=1}^n k_z(x_i, x_j)$ 或者 $K(x_i, x_j) = \prod_{z=1}^n k_z(x_i, x_j)$,
 $K(x_i, x_j)$ 仍是核函数. 引入加权系数 λ , 可令:

$$K(x_i, x_j) = \lambda K_{\text{Poly}}(x_i, x_j) + (1 - \lambda) K_{\text{RBF}}(x_i, x_j) = \lambda (\gamma x_i^T x_j + r)^d + (1 - \lambda) \left(\exp(-\gamma \|x_i - x_j\|^2) \right) \quad (10)$$

因此多核支持向量机的目标函数为:

$$\left\{ \begin{array}{l} \min_{\alpha^*, \alpha} \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i, x_j) + \varsigma \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ \text{s.t. } 0 \leq \alpha_i^*, \alpha_i \leq C, \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0, i = 1, 2, \dots, l \end{array} \right. \quad (11)$$

多核支持向量机的预测函数为:

$$f(x) = \sum_{i=0}^l (\alpha_i^* - \alpha_i) K(x_i, x) + b \quad (12)$$

$$A = 2a \cdot r_1 - a \quad (15)$$

$$C = 2 \cdot r_2 \quad (16)$$

$$a = 2 - t * \frac{2}{G_{\max}} \quad (17)$$

2 改进灰狼优化算法原理

2.1 基本灰狼优化算法

灰狼优化算法^[9](Grey Wolf Optimizer, GWO) 的寻优原理取自于灰狼群体的等级机制和狩猎方式, 捕食行动由头狼领导, 其它灰狼个体进行围攻. 采用 GWO 求解函数优化问题时, α 代表适应值最高的灰狼个体, β 和 δ 分别代表两个次优灰狼个体, 其他灰狼个体被定义为 ω . 灰狼优化算法寻优时, 由 α 带领, β 和 δ 狼个体从旁协助, 其余 ω 完成包围、猎捕和攻击行为, 从而寻找全局最优解.

首先, 灰狼群体通过式 (13)–式 (17) 对猎物进行包围.

$$D = |C \cdot X_p(t) - X(t)| \quad (13)$$

$$X(t) = X_p(t) - A \cdot D \quad (14)$$

其中, t 表示当前迭代次数, A 和 C 是系数向量, X_p 是食物的位置向量, X 是青蛙的位置向量. G_{\max} 为最大迭代次数, a 为收敛因子, r_1 和 r_2 在 $[0, 1]$ 范围内随机取值.

其次, 由 α 、 β 和 δ 狼领导, 灰狼群体通过式 (18) 对猎物进行猎捕.

$$\begin{cases} D_\alpha = |C_1 \cdot X_\alpha(t) - X(t)|, X_1 = X_\alpha - A_1 \cdot D_\alpha \\ D_\beta = |C_2 \cdot X_\beta(t) - X(t)|, X_2 = X_\beta - A_2 \cdot D_\beta \\ D_\delta = |C_3 \cdot X_\delta(t) - X(t)|, X_3 = X_\delta - A_3 \cdot D_\delta \\ X(t+1) = \frac{X_1 + X_2 + X_3}{3} \end{cases} \quad (18)$$

最后, 灰狼群体完成攻击行为来寻获猎物. 攻击行为主要通过式 (15) 来实现. 当 $A \leq 1$ 时, 灰狼群体对猎物集中攻击实现局部搜索; 当 $A > 1$ 灰狼群体分散实现全局搜索.

2.2 改进灰狼优化算法

目前 GWO 在许多领域得到了广泛的应用,如车辆路径问题^[10]、多目标优化^[11]、函数优化^[12]和入侵检测^[13]等.然而,灰狼优化算法存在求解精度不高和收敛速度较慢等不足,对此,研究学者多在收敛因子改进和位置更新方面进行改进^[14-19].

(1) 基于云模型理论改进收敛因子

灰狼优化算法存在探索能力(全局搜索)和开发能力(局部搜索)两种操作.当系数 A 大于1时,灰狼扩大搜索猎物范围进行全局搜索,而系数 A 小于1时,灰狼会缩小包围圈进行局部搜索.根据公式(15)可知,参数 A 是由收敛因子 a 决定的,基本GWO中 a 由2线性减小到0,但是寻优过程并不是线性的, a 的线性递减策略不能完全体现出实际的优化搜索过程,本文采用云模型来确定收敛因子,摒弃了已有算法用一条精确曲线作为收敛因子的方式,更好的控制全局搜索与局部搜索之间的能力平衡,以便提高算法的收敛速度和稳定性.云模型采用期望 Ex ,熵 En 和超熵 He 这3个数字特征来表示.通过以下步骤计算收敛因子.

Step 1. 产生一个以 En 为期望值, He 为标准差的正态随机数 En' ;

Step 2. 取 x 为迭代次数,作为一个云滴;

Step 3. 利用公式 $a = 2 * \exp\left(\frac{-(x - Ex)^2}{2(En')^2}\right)$ 计算收敛因子.

本文期望 Ex 设置为0, En 为总迭代次数的1/3,收敛因子 a 呈非线性递减.由图1可知,GWO的收敛因子在迭代过程中以相同的速率减小,而本文所提方法在迭代初期递减速率较慢,使得计算的 A 可以较长时间

保持较大值,进而提升搜索效率和全局寻优;而到了迭代后期收敛因子以较快的速率减小,使 A 可以较长的保持较小值,进而加强局部搜索能力和提高精细搜索能力.通过本文的计算方法可以更加有效地平衡全局搜索和局部搜索能力.

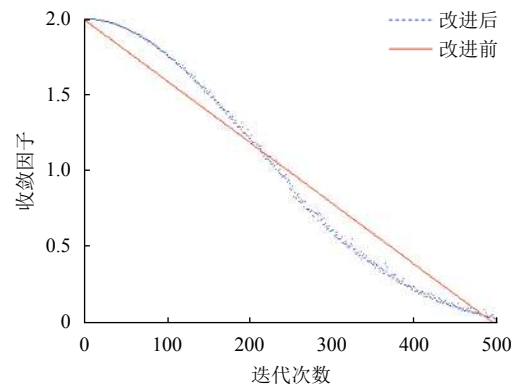


图1 收敛因子变化图

(2) 基于二次插值的头狼个体更新机制

灰狼优化算法在求解高维优化问题时,搜索空间规模较大造成后期收敛速度慢,寻优能力降低,其原因算法局部探索能力较差,如果对全局最优位置进行局部探索,则能够提高算法的收敛速度,依据灰狼进化原理,当前最优解也是通过三个位置的累加和确定头狼的移动方向,具有一定的随机性.本文对头狼的更新机制采用二次插值算法,提升算法的局部探索能力,如果通过二次插值得到的新解优于当前最优解则替换,否则不做任何操作.当前最优解 X^* 在二次插值中起到导向的作用,使搜索代理在其附近精细搜索,能强化搜索个体在当代最优解的精细搜索能力.

$$\begin{cases} X'_\alpha(t) = 0.5 \times \frac{(X_\beta^2(t) - X_\delta^2(t)) \times f(X_\alpha(t)) + (X_\delta^2(t) - X_\alpha^2(t)) \times f(X_\beta(t)) + (X_\alpha^2(t) - X_\beta^2(t)) \times f(X_\delta(t))}{(X_\beta(t) - X_\delta(t)) \times f(X_\alpha(t)) + (X_\delta(t) - X_\alpha(t)) \times f(X_\beta(t)) + (X_\alpha(t) - X_\beta(t)) \times f(X_\delta(t)) + \mu} \\ X_\alpha(t) = X'_\alpha(t), \quad \text{if } f(X'_\alpha(t)) < f(X_\alpha(t)) \end{cases} \quad (19)$$

同时为了避免迭代后期 $\alpha\beta\delta$ 个体适应度趋于一致性,因此在进行二次插值时先计算 $\beta\delta$ 个体适应度的距离,如果满足小于设定的阈值则随机在种群里选取灰狼个体替换 δ 个体依据式(7)进行二次插值获得新个体,进而保持种群的多样性.

(3) 基于动态权重策略的个体更新

在基本灰狼算法中 $\alpha\beta\delta$ 个体对其他个体的引领作用权重是一样的,已有文献通过引入适应度或是步长距离来计算3个狼在进化过程中的作用.本文提出一种综合适应度和步长距离的权重计算方法来动态调整权重,计算方法如式(20)所示,并将计算的3个权重带入式(18)来完成个体的更新.

$$\left\{ \begin{aligned} W_\alpha^1 &= \frac{f_\alpha + f_\beta + f_\delta}{f_\alpha}, W_\beta^1 = \frac{f_\alpha + f_\beta + f_\delta}{f_\beta}, W_\delta^1 = \frac{f_\alpha + f_\beta + f_\delta}{f_\delta} \\ W_\alpha^2 &= \frac{|X_1|}{|X_1| + |X_2| + |X_3|}, W_\beta^2 = \frac{|X_2|}{|X_1| + |X_2| + |X_3|}, W_\delta^2 = \frac{|X_3|}{|X_1| + |X_2| + |X_3|} \\ W_1 &= \frac{\sqrt{W_\alpha^1 * W_\alpha^2}}{\sqrt{W_\alpha^1 * W_\alpha^2} + \sqrt{W_\beta^1 * W_\beta^2} + \sqrt{W_\delta^1 * W_\delta^2}} \\ W_2 &= \frac{\sqrt{W_\beta^1 * W_\beta^2}}{\sqrt{W_\alpha^1 * W_\alpha^2} + \sqrt{W_\beta^1 * W_\beta^2} + \sqrt{W_\delta^1 * W_\delta^2}} \\ W_3 &= \frac{\sqrt{W_\delta^1 * W_\delta^2}}{\sqrt{W_\alpha^1 * W_\alpha^2} + \sqrt{W_\beta^1 * W_\beta^2} + \sqrt{W_\delta^1 * W_\delta^2}} \end{aligned} \right. \quad (20)$$

(4) 基于混沌策略的个体随机游走

为克服种群多样性衰减太快导致趋向局部最优而早熟的问题, 可以采用混沌映射在局部收敛区域外的空间范围内进行全局搜索. 为保持优化过程中的种群的多样性, 选取适应值最差个体 x_w 完成变异, 如果最差个体 x_w 根据公式 (18) 进化后得到的适应度得到改善则继续进入下一次迭代; 否则说明该灰狼个体需要被淘汰, 可以使用混沌理论产生新个体进入下一次迭代. 对于适应值最差的个体 x_w 采用偶数阶 Chebyshev 混沌映射进行更新.

$$x'_w = \cos(k \arccos x_w); \quad k = 2 \quad (21)$$

3 改进灰狼算法优化多核支持向量回归机流程

本文所构建的多核支持向量机采用多项式核函数和 RBF 核函数, 因此需要优化的参数分别为惩罚因子、RBF 和多项式核函数的核参数以及权重. 采用改进灰狼算法对于上述参数进行优化, 具体流程如下.

(1) 灰狼种群个体初始化. 将优化参数组合在一起作为灰狼个体进行编码, 并设定各参数的取值范围并初始化. 设定种群规模 m , 最大迭代次数 $maxiter$, 初始化 a, A 和 C .

(2) 将支持向量回归机的期望输出 y_i 与实际输出 o_i 的均方误差 MSE 作为灰狼个体的适应度, 具体计算公式如下.

$$MSE = \frac{1}{N} \sum_{i=1}^N (o_i - y_i)^2 \quad (22)$$

(3) 灰狼种群的所有个体根据式 (22) 计算适应度, 并将适应度最好的个体作为 α , 两个次优个体分别定义为 β 和 δ , 并基于二次插值算法利用式 (19) 的对 α 个体进行更新, 利用式 (20) 计算 3 个最优个体在个体更新过程中的权重来更新其他灰狼个体;

(4) 利用基于云模型理论算法计算收敛因子 A , 采用式 (16) 计算 C , 采用式 (13)、式 (14) 和式 (18) 完成个体的更新, 并判断种群中最差的个体是否有提高, 没有提高则用式 (21) 新产生一个新个体.

(5) 当迭代次数达到最大迭代次数 $maxiter$ 或适应度的值满足要求时停止迭代, 得到多核支持向量回归机优化参数, 建立采油速度预测模型.

4 实验仿真

4.1 改进灰狼优化算法寻优性能对比

为了验证 ICGWO 算法的寻优性能, 选取单峰、多峰和固定维度多峰 6 个基准测试函数进行测试, 6 个优化函数如表 1 所示. 同时选取基本 GWO 算法、文献 [14] IGWO 算法、文献 [15] AL-GWO 算法、文献 [16] 提出的 CGWO1 算法和文献 [17] 提出的 CGWO2 进行对比. 仿真实验的运行环境为 Intel Core i7 CPU, 主频 2.20 GHz, 内存 8 GB, Windows 10 操作系统, 实验仿真软件采用 Matlab R2015a. 所有对比算法种群个数统一设置为 30, 最大迭代次数 $maxiter$ 设置为 500 次, 优化维度除 F5 和 F6 之外都设置为 100 维. 实验独立运行 10 次, 对比各个算法的平均值和标准差结果如表 2 所示.

表1 优化函数

序号	测试函数	搜索空间	最优值
1	$f_1 = \sum_{i=1}^n x_i + \prod_{i=1}^n x_i $	$[-10, 10]^n$	0
2	$f_2 = \sum_{i=1}^n \left(\sum_{j=1}^i x_j \right)$	$[-100, 100]^n$	0
3	$f_3 = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos \left[\frac{x_i}{\sqrt{i}} \right] + 1$	$[-600, 600]^n$	0
4	$f_4 = 20 + \exp(1) - 20 \exp \left[-\frac{1}{5} \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \right] - \exp \left[\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i) \right]$	$[-32, 32]^n$	0
5	$f_5 = \left[\frac{1}{500} + \sum_{j=1}^{25} \frac{1}{j + \sum_{i=1}^2 (x_i - a_{ij})^6} \right]^{-1}$, $a_{ij} = \begin{pmatrix} -32, 16, 0, 16, 32, -32, 0, 16, 32 \\ -32, -32, -32, -32, -32, -16, 32, 32, 32 \end{pmatrix}$	$[-65.536, 65.536]^n$, $n = 2$	1
6	$f_6 = \sum_{i=1}^{11} \left[a_i - \frac{x_1(b_i^2 + b_i x_2)^2}{b_i^2 + b_i x_3 + x_4} \right]$	$[-5, 5]^n, n = 4$	0.0003075

表2 对比结果

函数	ICGWO		GWO		IGWO		AL-GWO		CGWO1		CGWO2	
	均值	标准差	均值	标准差	均值	标准差	均值	标准差	均值	标准差	均值	标准差
F1	0.00E+00	0.00E+00	3.94E-08	1.6E-08	4.1E-200	0.00E+00	6.81E-12	3.89E-12	8.32E-05	2.47E-05	8.19E-6	2.34E-04
F2	0.00E+00	0.00E+00	7.69E+02	7.27E+02	0.00E+00	0.00E+00	3.70E-02	6.89E-02	7.85E+03	2.61E+03	1.23E+06	1.55E+05
F3	2.31E-15	1.84E-15	1.07E-07	4.20E-08	4.09E-15	1.12E-15	7.64E-11	4.60E-11	1.74E-04	1.01E-04	2.00E-06	3.48E-06
F4	0.00E+00	0.00E+00	2.54E-03	8.03E-03	0.00E+00	0.00E+00	2.22E-17	4.68E-17	8.12E-03	2.57E-02	2.45E-03	1.91E-02
F5	9.98E+00	2.59E+00	2.87E+00	2.93E+00	1.04E+01	3.60E+00	4.72E+00	4.26E+00	2.95E+00	3.09E+00	1.19E+02	1.79E+02
F6	4.09E-04	2.55E-04	2.35E-03	6.33E-03	1.65E-03	1.24E-03	2.37E-03	6.32E-03	6.81E-04	4.45E-04	8.24E-02	1.23E-01

从表2的对比结果可知,本文所提算法在对仿真函数的优化过程中都取得了最好的结果,对于优化函数1、2和4都在迭代次数内寻找到最优解.IGWO寻优效果次之.分析其原因本文通过云模型改进收敛因子,基于二次插值的头狼个体更新机制和动态权重策略,可以很好地平衡全局探索和局部寻优,有利于加速寻优速度,这种改进在单峰函数中效果尤为明显,与GWO相比都找到了最优值.上述改进结合基于混沌策略的个体随机游走对于多峰函数的寻优也明显优于GWO,增加了跳出局部最优解的概率.虽然IGWO、CGWO1和CGWO2都对收敛因子采用非线性方法计算,但都属于采用一条精确曲线,而本文利用云模型定性概念与定量描述的不确定转换来完成收敛因子的计算,能更好地平衡全局搜索与局部搜索之间的寻优能力.IGWO和CGWO2采用权重的方式对灰狼个体的更新方式进行改进,通过实验结果可以得出IGWO的性能要优于CGWO2,但次于ICGWO,这是由于ICGWO不但考虑了位置向量对灰狼寻优位置的影响,同时也

考虑了3个最优狼本身适应度对群狼号召力的影响,CGWO1通过对随机扰动对头狼个体进行改进存在一定的随机性,而ICGWO则通过二次插值的方式通过当前3个较优解来加速局部寻优,具有一定的确定性,因此,通过以上改进原理的对比和实验结果可知,本文算法都优于其他对比算法,具有很好的搜索能力和良好的稳定性,且收敛精度较高.

4.2 基于多核支持向量回归机的采油速度预测

(1) 实际数据预处理

针对实际地质数据和开发数据,应用灰色关联分析理论的对采油速度影响因素的重要性进行分析.以采油速度为参考序列,含水率、地层压力、流压、有效厚度、地质储量、渗透率、束缚水饱和度、含水上升速度和压力恢复速度等多个影响因素为比较序列,利用均值法无量纲化,取分辨系数为0.5,得到不同分类油层的采油速度影响因素的关联度大小排序.最后选取含水、地层压力、流压、有效厚度、渗透率、束缚水饱和度6个属性用于确定

采油速度.

(2) 实例验证

首先选取油田某区块 2019 年 168 口油井每月的含水率、地层压力、流压、有效厚度、渗透率、束缚水饱和度和采油速度共 2016 条数据进行实例验证, 选用 2019 年 1 月到 9 月的数据作为训练数据集, 2019 年 10 月到 12 月的数据作为测试数据集. 分别采用静态影响因素分析法、动态影响因素分析法、多项式支持向

量回归机 (PLOY-SVR)、RBF 支持向量回归机 (RBF-SVR)、基于 GWO 优化的多核支持向量机 (GWO-MCSVR) 和基于 ICGWO 优化的多核支持向量回归机 (GWO-MCSVR) 进行对比. 采用均方误差 (Mean Squared Error, MSE) 和平均绝对误差 (Mean Absolute Error, MAE) 来对比各种算法的预测效果. 1 月到 9 月的训练数据集对比数据如表 3 所示, 10 月到 12 月的测试集对比结果如表 4 所示.

表 3 训练样本对比结果

月份	静态法		动态法		PLOY-SVR		RBF-SVR		GWO-MCSVR		IGWO-MCSVR		ICGWO-MCSVR	
	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
1	1.754	19.531	1.806	19.360	1.171	8.707	1.172	8.707	0.517	3.640	0.637	3.089	0.490	2.571
2	1.599	16.458	1.725	16.514	1.003	7.906	1.004	7.907	0.685	3.930	0.535	3.696	0.373	0.571
3	1.353	14.177	1.521	14.670	0.824	4.176	0.825	4.176	0.724	3.586	0.605	2.436	0.432	1.378
4	1.303	13.581	1.447	14.104	0.786	4.008	0.787	4.008	0.628	2.898	0.554	2.294	0.353	0.537
5	1.250	15.917	1.460	16.341	0.752	4.758	0.753	4.757	0.730	4.339	0.566	2.760	0.373	0.587
6	1.116	11.700	1.384	12.372	0.662	3.464	0.662	3.464	0.578	2.907	0.480	1.813	0.299	0.318
7	1.233	12.732	1.419	12.979	0.737	3.775	0.738	3.773	0.654	3.281	0.436	0.786	0.301	0.325
8	1.052	10.819	1.242	11.575	0.621	3.207	0.621	3.207	0.535	2.692	0.393	0.499	0.380	1.061
9	1.144	12.574	1.388	13.059	0.682	3.732	0.683	3.733	0.530	1.114	0.468	1.874	0.312	1.019

表 4 测试样本结果

月份	静态法		动态法		PLOY-SVR		RBF-SVR		GWO-MCSVR		IGWO-MCSVR		ICGWO-MCSVR	
	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
10	1.099	11.368	1.374	12.078	0.940	2.254	0.940	2.255	0.532	2.859	0.524	2.061	0.358	1.031
11	1.123	14.829	1.341	15.071	0.957	2.958	0.957	2.957	0.608	4.024	0.394	1.548	0.381	1.430
12	1.156	8.786	1.326	9.434	1.004	1.716	1.004	1.717	0.587	2.274	0.512	1.517	0.366	0.872

采油速度静态影响因素分析法计算方法如式 (23):

$$V_0 = \lg[(K * h / \mu)^a] + b * \eta^c - d \quad (23)$$

其中, V_0 为采油速度, μ 为地层原油粘度, h 为有效厚度, η 为注采井距, K 为油藏渗透率, a, b, c, d 为经验系数, 可以通过统计回归来得到.

动态影响因素分析法计算方法如式 (24):

$$V_0 = \frac{1}{b(1 - S_{wc})} \cdot \frac{1}{f_w(1 - f_w)} \cdot f_w' \quad (24)$$

其中, V_0 为采油速度, f_w 综合含水率, S_{wc} 束缚水饱和度, f_w' 为含水上升速度.

从表 3 可知, ICGWO 优化的多核支持向量机的拟合效果最好, 多核支持向量回归机的拟合效果要优于单核回归机, 静态影响因素分析法、动态影响因素分析法的拟合结果较差. 主要原因是这两种方法都是基于经验公式拟合, 且考虑影响因素较少, 拟合结果受数据的分布情况影响较大. 多个核函数的优化组合可以

弥补单一核函数选择盲目性的不足, 可以充分地提取数据的隐含信息. 从表 4 可知 ICGWO-MCSVR 对 10 月到 12 月的采油速度预测效果最好, 这是因为模型充分利用多核学习的优势, 以及利用 ICGWO 算法较好的寻优性能完成对核函数参数、惩罚因子和权值系数的最优选择.

5 结论

利用多核学习支持向量回归机模型预测采油速度, 针对模型中权值系数, 惩罚因子核和核参数的选择问题, 提出一种改进的灰狼优化算法, 算法利用云模型改进收敛因子的计算方式, 较好的平衡了算法全局和局部寻优性能, 利用二次插值方式改进了头狼的进化方式加速局部寻优. 实验结果表明, 经过优化的模型参数构造的多核支持向量回归机, 其预测精度相比较单核支持向量机来说有明显得提高.

参考文献

- 1 赵伦, 陈希, 陈礼, 等. 采油速度对不同黏度均质油藏水驱特征的影响. 石油勘探与开发, 2015, 42(3): 352–357.
- 2 吴文有, 王禄春, 张继风, 等. 特高含水期油田采油速度主要影响因素. 大庆石油地质与开发, 2012, 31(5): 51–55. [doi: 10.3969/J.ISSN.1000-3754.2012.05.010]
- 3 姜风光, 王小林. 采油速度影响因素及理论值研究. 西南石油大学学报(自然科学版), 2013, 35(6): 104–108.
- 4 陈军, 杨峻懿, 刘启国. 特高含水期油藏采油速度研究及影响因素分析. 油气藏评价与开发, 2018, 8(4): 17–21, 32. [doi: 10.3969/j.issn.2095-1426.2018.04.004]
- 5 吕达, 刘文婧, 陈肖洁. 规则的多核学习支持向量回归算法. 机械设计与制造, 2016, (12): 53–56. [doi: 10.3969/j.issn.1001-3997.2016.12.014]
- 6 邵朝, 李强. 基于特征加权的多核支持向量机. 西安邮电大学学报, 2017, 22(2): 84–88.
- 7 Li Y, Zhu ZC, Hou AL, *et al.* Pulmonary nodule recognition based on multiple kernel learning support vector machine-PSO. Computational and Mathematical Methods in Medicine, 2018, 2018: 1461470.
- 8 杜伟, 房立清, 齐子元, 等. 基于 LFOA 优化多核支持向量机的液压泵故障诊断. 机械设计与研究, 2018, 34(1): 137–140, 144.
- 9 Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Advances in Engineering Software, 2014, 69: 46–61.
- 10 黄戈文, 蔡延光, 戚远航, 等. 自适应遗传灰狼优化算法求解带容量约束的车辆路径问题. 电子学报, 2019, 47(12): 2602–2610.
- 11 鲍毅, 戴波, 汪志华, 等. 基于灰狼算法的多目标智能家居负荷控制算法. 系统仿真学报, 2019, 31(6): 1216–1222.
- 12 张新明, 王霞, 康强. 改进的灰狼优化算法及其高维函数和 FCM 优化. 控制与决策, 2019, 34(10): 2073–2084.
- 13 徐慧, 付迎春, 刘翔, 等. 融合杜鹃搜索的灰狼优化算法在网络入侵检测特征选择中的应用. 工程科学与技术, 2018, 50(5): 160–166.
- 14 郭振洲, 刘然, 拱长青, 等. 基于灰狼算法的改进研究. 计算机应用研究, 2017, 34(12): 3603–3606, 3610. [doi: 10.3969/j.issn.1001-3695.2017.12.019]
- 15 郭阳, 张涛, 胡玉蝶, 等. 基于自适应头狼的灰狼优化算法. 成都大学学报(自然科学版), 2020, 39(1): 60–63, 73.
- 16 谈发明, 赵俊杰, 王琪. 一种改进非线性收敛方式的灰狼优化算法研究. 微电子学与计算机, 2019, 36(5): 89–95.
- 17 晏福, 徐建中, 李奉书. 混沌灰狼优化算法训练多层感知器. 电子与信息学报, 2019, 41(4): 872–879. [doi: 10.11999/JEIT180519]
- 18 Luo KP, Zhao QH. A binary grey wolf optimizer for the multidimensional knapsack problem. Applied Soft Computing, 2019, 83: 105645. [doi: 10.1016/j.asoc.2019.105645]
- 19 Gupta S, Deep K. A novel random walk grey wolf optimizer. Swarm and Evolutionary Computation, 2019, 44: 101–112.