

基于 SSA-TSVR 的飞机状态预测方法研究

赵 晗, 樊智勇, 刘 涛

(中国民航大学 电子信息与自动化学院, 天津 300300)

摘要: 为了构建地面飞行安全态势监测系统, 针对飞机状态数据向地面传输过程中出现数据传输异常情况而导致无法对飞机状态进行实时监控的问题, 提出一种基于 SSA-TSVR 的飞机状态预测方法, 使用随机森林算法对真实飞行数据进行特征重要度分析, 筛选与待预测飞机状态参数关系密切的重要参数, 获得待预测参数与飞行数据间重要度关系; 通过孪生支持向量回归算法建立预测模型, 对缺失的关键飞行状态参数进行预测; 并运用飞鼠搜索算法对孪生支持向量回归模型进行优化, 根据不同预测对象选择对应的最优核函数, 提高了模型预测精度; 以飞行高度、速度为预测对象进行实验验证, 预测模型实现了利用不完整飞行数据对飞机状态进行准确预测, 对飞机飞行状态监测有着重要意义。

关键词: 飞行状态预测; 孪生支持向量回归; 飞鼠搜索算法; 随机森林; QAR 数据

Research on Aircraft State Prediction Method Based on SSA-TSVR

ZHAO Han, FAN Zhiyong, LIU Tao

(School of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In order to build ground flight safety situation monitoring systems, solve the problem that abnormal data transmission occurs during the transmission process of aircraft status data to the ground, resulting in the inability to the real-time monitoring of aircraft status data, an aircraft status prediction method based on squirrel search algorithm and twin support vector regression (SSA-TSVR) was proposed. The random forest method was used to analyze the feature importance of real flight data, screen and predict the important parameters closely related to the aircraft's state, and obtain the importance relationship between the parameters to be predicted and the flight data. The twin support vector regression algorithm was used to build the prediction model, and predict the missing key flight state parameters. The twin support vector regression model was optimized by using the SSA, and the optimal kernel function was selected according to different prediction objects, improving the prediction accuracy of the model. With flight altitude and speed as prediction objects, the prediction model realizes the accurate prediction of aircraft state by using incomplete flight data, which is of great significance for aircraft flight state monitoring.

Keywords: aircraft state prediction; TSVR; SSA; random forest; QAR data

0 引言

当今的航空飞机越来越复杂、越来越智能, 维护和保障成本越来越高, 能否保证航空飞机安全地工作直接影响着民航业的整体发展^[1-2]。

在全世界范围内, 尽管民用航空运输与其他交通方式相比一直保持着比较低的事故率以及死亡率, 并且已经被普遍认为是安全的出行方式, 但是近些年来发生的航空事故在一定程度上影响了人民群众对民用航空安全的信心^[3-4]。为了保证飞机在飞行全过程中的安全, 各国民用航空监督管理机构和运输航空公司一直都将航空器的运行监控作为一项重点工作来进行^[5-6]。为了对飞机状态进行实时监控, 常利用实时传输的飞行数据在地面端对飞机状态进行监控。但是当飞机在实际飞行过程中有可能发生意外状况, 无法确保所有数据都能完整地传输至地面端, 因此当

传输数据出现缺失时, 便无法对飞机进行监控。针对飞机在飞行中出现传输数据缺失而导致无法对飞机状态进行实时监控这一问题, 需要一种仅利用部分飞行数据便能够对飞机状态进行预测的方法。

对飞行状态的预测作为气动力建模的一个分支, 传统方法较复杂, 成本高, 样本数据需求量大, 且存在提取时间关联信息能力弱、性能表现差等弊端^[7-8]。随着人工智能技术的兴起, 智能方法如神经网络模型、支持向量机模型等陆续应用到飞行状态预测, 这些方法规避了对复杂流体的机理研究, 属于黑箱非线性系统建模, 预测精度较高。

在飞机状态预测研究中, Kou J^[9]利用长短时记忆网络(LSTM, long short-term memory)进行飞机失速预测建模; 王经纬^[10]等利用 Transformer 方法推测飞机的飞行状态并进行分类; 李桂毅^[11]等利用 RBF 神经网络飞机运行态势预测进行了研究; 赵元棣^[12]等将隐马尔科夫模型

收稿日期: 2024-03-08; 修回日期: 2024-04-03。

基金项目: 国家重点研发计划项目子课题(2023YFB4302901)。

作者简介: 赵 晗(1998-), 男, 硕士研究生。

樊智勇(1979-), 男, 硕士, 教授。

引用格式: 赵 晗, 樊智勇, 刘 涛. 基于 SSA-TSVR 的飞机状态预测方法研究[J]. 计算机测量与控制, 2024, 32(9): 125-132, 156.

(HMM, hidden markov model) 与反向传播 (BP, back propagation) 神经网络相结合, 建立了 HMM-BP 混合模型对飞行状态进行预测; 何磊^[13]等使用支持向量机方法预测了气动数据。

支持向量机 (SVM, support vector machine) 是一种能够从大规模数据中自动提取特征并进行高级别抽象的机器学习技术^[14-16]。和人工神经网络相比, SVM 的泛化能力更强。目前, SVM 已在求解分类和回归两方面问题中得到成功应用, 其中 SVM 常用于解决分类问题, 支持向量回归 (SVR, support vector regression) 常用于解决回归问题。

在对飞机状态的预测过程中, 需要对大量数据进行运算从而得到待预测飞机状态参数。在实际工程应用中, 当飞机出现传输数据缺失这一现象后, 为了尽可能保证飞机的安全, 必须要以最快速度通过算法模型对缺失数据进行预测, 因此这一过程在实际应用中对实时性要求较高。

孪生支持向量机 (TWSVM, twin support vector machines) 是在 SVM 基础上进行改进的一种算法, 由 Javadeva 等人在 2007 年提出^[17]。该算法在训练速度上明显优于传统的 SVM 算法。和 SVM 不同, TWSVM 要寻找的是两个不平行的分类超平面, SVM 需要求解一个大型二次规划问题, 而 TWSVM 将其化简为求解两个较小的二次规划问题, 从理论上来说, 其训练效率是 SVM 的 4 倍^[18]。

与 SVM 推广到 SVR, 将求解分类问题推广至解决回归问题类似, 孪生支持向量回归机 (TSVR, twin support vector regression) 是在 TWSVM 基础上为了解决回归问题提出的一种机器学习方法, 由 Peng 等人在 2010 年提出^[19-21]。和 SVR 相比, TSVR 的训练速度也得到了很大的提高。根据实际需求中, 飞机状态预测过程实时性要求较高这一特点, 采用 TSVR 算法建立预测模型。

虽然通过 TSVR 模型已经能够对飞机状态进行初步预测, 但是在模型的应用中仍然有一定缺陷, 在模型训练过程中有许多超参数需要进行手动调参从而导致模型的预测精度无法确保满足要求。模型中参数选择对模型的学习效果和泛化性能有直接的影响, 因此合适的参数是实现算法优势的前提。得到的 TSVR 预测模型在建模过程中需要一种参数寻优算法对模型参数进行调整, 保证寻优后的相关参数计算得到最优超平面, 从而使得训练数据集尽可能地落入最优回归超平面附近的间隔带内。

传统的参数优化方法主要有交叉验证技术、网格搜索法等, 但这些方法都存在极大的计算量, 非常耗时, 所以并不符合实际的应用场景。飞鼠搜索算法 (SSA, squirrel search algorithm,) 是 2018 年 Mohit Jain 等提出的一种新颖的群智能进化自然启发式算法^[22-23]。与其他传统参数优化方法相比, 其计算量更小并且优化性能和跳出局部最优的能力都更强。因此利用 SSA 算法对 TSVR 模型进行优化, 解决超参数寻优问题, 并提高模型的预测精度。

在建立飞机状态预测模型之前, 需要确定预测模型的

输入与输出关系。对于模型的输出, 即待预测的飞机状态参数; 在实际飞机状态监测中, 往往更关注飞机的整体飞行状态, 因此选取飞行高度、飞行速度这两个关键飞行状态参数作为待预测对象。

对于模型的输入, 即根据哪些参数对飞机状态进行预测; 虽然可以利用所有历史数据对模型进行训练、预测, 但这种方式让模型训练时间过长, 无法满足监控系统的实时性要求, 并会产生参数冗余问题, 使得建模过程中噪点数据过多, 且还有可能引起模型过度拟合。因此需要对真实飞行数据进行特征重要性分析, 使用随机森林算法对输入参数与输出参数之间的相关性进行评估, 从而确定各参数对预测模型的相对重要性, 挖掘出对飞机飞行状态影响较大的因素, 对输入空间进行降维。

基于上述分析, 提出基于 SSA-TSVR 算法的飞机状态预测方法的基本框架。首先根据待预测飞行状态参数, 即飞行高度、飞行速度进行特征重要性分析, 提取飞行数据中与待预测参数关系最密切的重要参数作为预测模型输入; 为满足飞行状态监控中的实时性要求, 凭借 TSVR 算法时间复杂度较低的优势, 建立 TSVR 预测模型; 针对 TSVR 模型中的预测精度不足以及超参数寻优问题, 利用 SSA 算法对其进行优化得到 SSA-TSVR 模型; 通过真实飞行数据进行实验, SSA-TSVR 的飞机状态预测模型能够根据不同预测对象选择对应最优核函数, 提高了模型预测精度, 实现了利用不完整飞行数据对飞行状态进行预测的功能, 可以满足实际应用需求。

1 特征重要性分析

特征重要性是对输入参数与输出参数之间的相关性进行评估, 确定飞行数据中各参数对待预测对象的相对重要性, 挖掘出对飞机飞行状态影响较大的因素, 对输入空间进行降维。根据需求, 以 QAR 数据作为真实数据, 利用随机森林算法进行特征重要性分析, 从中选取对飞行高度、速度影响较大的参数, 并利用这些参数作为后续预测模型的输入数据进行预测。

随机森林算法通过有放回的随机采样, 从原始数据集中抽取数据样本放入采样集中, 然后利用决策树算法对采样集数据进行学习得到回归树。对数据进行随机采集的过程中, 根据计算有 36.8% 的数据未被采集, 这些数据被称为袋外数据。对参数重要度进行评估时, 主要使用袋外数据进行计算。

给定训练数据集, 每个输入数据样本包含有 m 个特征。使用随机森林算法进行特征重要性评估的过程如下所示:

$$errOOB_i = \frac{1}{2} \sum_{i=1}^N [y_i - f(x_i)]^2, (x_i, y_i) \in B_i$$

1) 输入训练数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 特征 $X = \{x_{i1}, x_{i2}, \dots, x_{im}\}$;

2) 过程:

①从原始训练数据集中有放回的随机抽取 n 个输入数

据样本放入采样集中, 用同样的方法抽取数据获得采样集, 直至得到 k 个采样集;

②利用每个采样集训练得到一颗回归树, 记为 $\{T_1, T_2, \dots, T_K\}$ 。与采样集对应的产生 k 组未被采样的袋外数据集, 记为 $\{B_1, B_2, \dots, B_K\}$;

③训练得到回归树后, 用与采样集对应的袋外数据集计算袋外误差, 计算公式为:

$$errOOB_i = \frac{1}{2} \sum_{i=1}^N [y_i - f(x_i)]^2, (x_i, y_i) \in B_i$$

其中: $f(x)$ 为学习得到的回归树;

④对袋外数据中某一参数进行噪点数据干扰, 对于干扰后的袋外数据计算袋外误差,

⑤重复过程③、过程④, 直至完成所有袋外数据的袋外误差计算;

⑥通过公式计算每个输入样本空间中参数的特征重要度得分, 计算公式为:

$$Score = \sum_{i=1}^K \frac{1}{K} |errOOB_i - errOOB'_i|$$

3) 输出样本空间中每个参数的重要性得分。

根据随机森林算法流程可以对 QAR 数据进行特征重要性进行分析, 但是在没有一定的评价标准的情况下, 输入参数的个数 k 的确定缺乏说服力, 因此为了找到合适的输入参数个数, 按照输入参数对飞行高度影响重要性的排序, 测试不同的 k 值对模型的精度的影响。随着输入参数的个数不断增加, 模型的精度值不断增加, 模型的精度值变化范围不大, 直至输入 7 个参数时模型的精度值达到最高, 输入参数值超过 7 个后, 模型的精度值下降, 说明输入 7 个参数训练模型时, 模型的效果最好。

由于飞机在不同飞行阶段中影响飞行状态的重要参数并不相同, 因此对 QAR 数据进行飞行阶段划分, 将其划分为爬升、巡航、降落 3 个阶段, 并对每个阶段分别进行特征重要性分析来保证分析得到重要参数的准确性和说服力。

2 TSVR 预测模型的建立

2.1 孪生支持向量回归机原理

TSVR 通过在训数据点两侧产生一对不平行的函数, 分别确定回归函数的 ϵ 不敏感上、下界。通过训练数据的 ϵ_1 不敏感下界:

$$f_1(x) = w_1^T x + b_1 \tag{1}$$

与 ϵ_2 不敏感上界:

$$f_2(x) = w_2^T x + b_2 \tag{2}$$

确定最终的回归函数, 而这对函数可以通过求解下面的一对二次规划问题得到。

$$\begin{aligned} \min & \frac{1}{2} \|Y - \epsilon\epsilon_1 - (Av_1 + eb_1)\|^2 + C_1 e^T \xi \\ \text{s. t. } & Y - (Av_1 + eb_1) \geq \epsilon\epsilon_1 - \xi, \xi \geq 0 \\ \min & \frac{1}{2} \|Y + \epsilon\epsilon_2 - (Aw_2 + eb_2)\|^2 + C_2 e^T \eta \end{aligned} \tag{3}$$

$$\text{s. t. } Y - (Aw_1 + eb_1) \geq \epsilon\epsilon_1 - \xi, \xi \geq 0 \tag{4}$$

式中, $C_1, C_2 > 0, \epsilon_1, \epsilon_2 > 0$ 为常数, ξ, η 为松弛变量, e 为 1×1 维的单位列向量。引入拉格朗日乘子 α 和 γ , 并结合 KKT 条件, 可以得到式 (3) 和式 (4) 的对偶优化问题为:

$$\begin{aligned} \max & -\frac{1}{2} \alpha^T G (G^T G)^{-1} G^T + f^T G (G^T G)^{-1} G^T \alpha - f^T \alpha \\ \text{s. t. } & 0 \leq \alpha \leq C_1 e \end{aligned} \tag{5}$$

$$\begin{aligned} \max & -\frac{1}{2} \gamma^T G (G^T G)^{-1} G^T \gamma - h^T G (G^T G)^{-1} G^T \gamma + h^T \gamma \\ \text{s. t. } & 0 \leq \gamma \leq C_2 e \end{aligned} \tag{6}$$

其中: $G = [A \quad e], f = Y - \epsilon_1$ 和 $h = Y + \epsilon_2 e$, 优化之后得到下面的目标回归函数:

$$\begin{aligned} f(x) &= \frac{1}{2} (f_1(x) + f_2(x)) = \\ &= \frac{1}{2} (w_1 + w_2)^T x + \frac{1}{2} (b_1 + b_2) \end{aligned} \tag{7}$$

式中, $[w_1 \quad b_1]^T = (G^T G)^{-1} G^T (f - \alpha)$, $[w_2 \quad b_2]^T = (G^T G)^{-1} G^T (h + \gamma)$

2.2 核函数

模型在学习非线性的样本空间时, 映射获得特征空间的好坏对模型的学习有着极大的影响, 而核函数隐式的表达了特征空间。常见的核函数如下。

1) 线性 (linear) 核函数:

$$K(x_i, x_j) = x_i^T x_j \tag{8}$$

2) 多项式 (poly) 核函数:

$$K(x_i, x_j) = x_i^T x_j^d \tag{9}$$

其中: d 为多项式的次数。

3) 径向基 (rbf) 核函数:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{10}$$

其中: σ 为高斯核的带宽。

4) Sigmoid 核函数:

$$K(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta) \tag{11}$$

飞机飞行数据中包含许多具有时变性、非线性特点的数据, 使用 TSVR 算法学习模型时, 需要借助核函数将原始数据集向高维空间进行映射, 使得原始数据集变得线性可分。不同的核函数映射获得的高维空间不同, 从而得到的超平面也不同, 回归函数的解答式也会不同, 因此核函数的选择很大程度上决定了 TSVR 模型的性能。因此在对飞机不同飞行阶段进行建模时, 需要对不同核函数的性能进行评估, 选择合适的核函数用于建立预测模型。

2.3 TSVR 预测模型建模流程

在使用 TSVR 算法建立预测模型时: 首先, 将数据进行划分, 分为训练集数据和测试集数据, 其中训练集数据用于 TSVR 模型学习, 测试集数据用于测试 TSVR 模型的预测准确度; 然后使用不同的核函数进行测试, 选择性能最优的核函数训练获得的模型。其基本建模框图如图 1 所示。

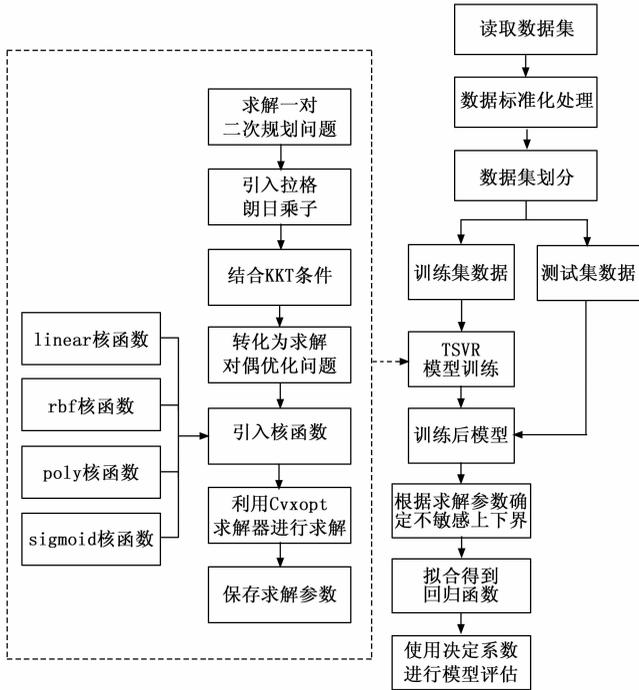


图 1 TSVR 预测模型建框图

3 SSA 优化 TSVR 重要参数

3.1 SSA 原理

飞鼠搜索算法是一种收敛速度快、跳出局部最优能力强的新型算法，其基本原理及搜索寻优步骤如下。

1) 随机初始化种群：

首先需要设置飞鼠的种群数量，将其设置为 FS ，在这个种群中，每一个飞鼠个体转化为一个向量，并将所有飞鼠的位置转化为如式 (12) 的矩阵，为了对飞鼠位置进行随机分布，用式 (13) 初始化所有飞鼠的位置。

$$FS = \begin{bmatrix} FS_{1,1} & FS_{1,2} & \cdots & FS_{1,d} \\ FS_{2,1} & FS_{2,2} & \cdots & FS_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ FS_{n,1} & FS_{n,2} & \cdots & FS_{n,d} \end{bmatrix} \quad (12)$$

$$FS_{i,j} = FS_L + U(0,1) \times (FS_U - FS_L) \quad (13)$$

式中， $FS_{i,j}$ 为森林中第 i 个飞鼠在第 j 维度上的位置；设置 FS_U 和 FS_L 为飞鼠活动范围的上下界； $U(0,1)$ 是在 $0 \sim 1$ 之间的随机数。

2) 适应度评估：

然后对飞鼠适应度进行评估，主要是通过计算得出每一只飞鼠在不同位置的食物资源优劣，从而反映出飞鼠的生存概率，式 (14) 为飞鼠适应度函数。

$$f = \begin{bmatrix} f_1 & ([FS_{1,1} & FS_{1,2} & \cdots & FS_{1,d}]) \\ f_2 & ([FS_{2,1} & FS_{2,2} & \cdots & FS_{2,d}]) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_n & ([FS_{n,1} & FS_{n,2} & \cdots & FS_{n,d}]) \end{bmatrix} \quad (14)$$

3) 适应度排序及分类：

将适应度结果升序排列，适应度最小的飞鼠位于山核

桃树上，次之的飞鼠位于橡树上，其余的飞鼠在普通树上。

4) 位置更新：

飞鼠在进行觅食活动时实现位置更新，根据捕食者出现的概率来影响飞鼠滑翔的方向。飞鼠进行位置更新的情况分为 3 种。

第一种情况是获得次优级食物资源即橡树的飞鼠为了得到更好的食物资源飞往最优级食物资源山核桃树。位置更新公式如式 (15) 所示：

$$\begin{cases} FS_{at}^t + d_g \times G_c \times (FS_{ht}^t - FS_{at}^t) \text{rand} \geq P_{dp} \\ \text{Random location otherwise} \end{cases} \quad (15)$$

第二种情况是在普通树上的飞鼠不能满足自己的食物需求，需要寻找更好的食物资源，因此飞鼠可以选择飞往橡树。位置更新公式如式 (16) 所示：

$$FS_{mt}^{t+1} = \begin{cases} FS_{mt}^t + d_g \times G_c \times (FS_{st}^t - FS_{mt}^t) \text{rand} \geq P_{dp} \\ \text{Random location otherwise} \end{cases} \quad (16)$$

第三种情况是普通树上的飞鼠还可以选择飞往山核桃树位置更新公式如式 (17) 所示：

$$\begin{cases} FS_{st}^t + d_g \times G_c \times (FS_{ht}^t - FS_{st}^t) \text{rand} \geq P_{dp} \\ \text{Random location otherwise} \end{cases} \quad (17)$$

式中， FS_{at} 、 FS_{ht} 、 FS_{st} 分别为飞鼠在橡树上的位置、在山核桃树上的位置、在普通树上的位置。 d_g 为滑步行步长； G_c 为滑动常数；rand 是在 $0 \sim 1$ 之间的随机数； t 为迭代次数。

5) 季节判别及算法停止条件：

为了增强算法跳出局部最优的能力，加入了季节判别因素，当处于冬季时，被捕食者捕获的概率增加，从而导致飞鼠变得不再活跃，当冬季结束时，飞鼠再次进行活跃的觅食活动。

设置季节常数 S_c^t 与季节常数最小值进行比较，从而判断是否处于冬季。如式 (18)、(19) 所示：

$$S_c^t = \sqrt{\sum_{k=1}^d (FS_{at,k}^t - FS_{ht,k}^t)^2} \quad (18)$$

$$S_{\min} = \frac{10E^{-6}}{(365)^{(2.5t/T_s)}} \quad (19)$$

当冬季结束时，普通树上未获得食物的飞鼠可根据式 (20) 进行位置更新：

$$FS_{mt}^{t+1} = FS_L + Levy \times (FS_U - FS_L) \quad (20)$$

式中， $Levy$ 是列维飞行获得步长方式。

当算法达到所设定的最大迭代次数时，算法停止。

3.2 SSA-TSVR 模型

SSA 算法是为了对 TSVR 模型中的超参数进行优化，实现自动寻优并提高 TSVR 模型预测精度。在 TSVR 模型中，主要需要调节的参数包括惩罚因子 C 、不敏感参数 ϵ ，以及核函数参数 γ 。

其中惩罚因子 C 决定了支持向量与最优分类超平面的距离，也就是决定的错误分类的容忍程度。 ϵ 可以改变两个不平行超平面之间的距离，从而影响使模型的泛化性能。

核函数参数 γ 的取值大小与 TSVR 的分类精细程度密切相关。

因此, 为 TSVR 模型选择合适的核函数参数 γ 、惩罚因子 C 值、不敏感参数 ϵ 能够于提高 TSVR 的预测精度及计算速度。

利用 SSA 对 TSVR 的核函数参数 γ 、不敏感参数 ϵ 和惩罚因子 C 进行寻优的流程框图如图 2 所示, 其过程如下:

1) 首先需要初始化 SSA 算法的各类参数, 主要包括飞鼠种群规模、迭代次数、滑行步长、捕食概率等。将核函数参数 γ 、不敏感参数 ϵ 和惩罚因子 C 作为飞鼠的初始位置及优化目标。其中, 核函数参数 γ 取 $0 \sim 100$, 不敏感参数 ϵ 取 $0 \sim 1$, 惩罚因子 C 取 $0 \sim 100$ 。

2) 根据适应度函数计算每只飞鼠的适应度, 对得出的适应度值进行排序, 进而确定出飞鼠在山核桃树和橡树的位置, 分别作为全局最优解与局部最优解。

3) 对于仍在捕食的飞鼠, 不断更新其位置信息以靠近山核桃树和橡树。

4) 加入季节性影响, 根据季节常数进行判断是否处于冬季, 若处于冬季则飞鼠的觅食活动将不再活跃, 等冬季结束时飞鼠再重新变得活跃起来, 利用 Levy 飞行对位置进行更新从而防止算法陷入局部最优的情况。

5) 对算法迭代次数进行判断, 是否达到最大迭代次数, 若不满足条件, 则需要重新对飞鼠适应度进行计算, 直到达到最大迭代次数。将山核桃树上飞鼠位置当作全局最优解输出, 即得到核函数参数 γ 、不敏感参数 ϵ 和惩罚因子 C 的 3 个待寻优参数的最优值。

TUDE (1 013.25 mB) 为预测目标, 剔除 QAR 数据中所有与飞行高度直接相关的数据, 然后利用其他数据进行特征重要性分析。在爬升、巡航、下降阶段的分析结果分别如表 1~3 所示。

表 1 爬升阶段飞行高度特征重要性分析结果

特征	重要度
TOTAL PRESS	0.236 97
STATIC PRESS	0.183 12
MAX BUFFET AIRSPEED	0.166 59
GNSS LONGITUDE	0.152 38
LOCAL P0 SENSOR	0.148 51
MACH	0.080 11
DISTANCE TO GO	0.032 29

表 2 巡航阶段飞行高度特征重要性分析结果

特征	重要度
GNSS LATITUDE	0.259 67
GNSS LONGITUDE	0.249 76
PRES POSN LAT -FMC	0.229 80
PRES POSN LONG -FMC	0.160 22
AVERAGED STATIC PRESS	0.070 06
FMC AIRSPEED SELECTED	0.030 26
CENTER MAIN FUEL OTY	0.002 21

表 3 下降阶段飞行高度特征重要性分析结果

特征	重要度
SUBFRAME COUNTER	0.197 76
STATIC PRESS AVG(CORR)	0.171 89
PRES POSN LAT - FMC	0.170 93
MAX BUFFET AIRSPEED	0.155 08
CABIN DIFF PRESS CPC2	0.133 34
LOCAL P0 SENSOR	0.087 21
GROUND SPEED	0.083 76

以飞行速度为预测对象时, 选取 QAR 数据中的 AIR-SPEED 为预测目标, 剔除 QAR 数据中所有与飞行速度直接相关的数据, 然后利用其他数据进行特征重要性分析, 在爬升、巡航、下降阶段的分析结果分别如表 4~6 所示。

表 4 爬升阶段飞行速度特征重要性分析结果

特征	重要度
STATIC PRESS AVG(CORR)	0.206 43
TE FLAP SKEW 3 POSN	0.191 91
EEC MAX CLIMB N1 RATING	0.151 73
CABIN DIFF PRESS CPC2	0.135 83
BARO COR ALT	0.133 83
ELEV PITCH CMD	0.124 18
TOTAL FUEL	0.056 06

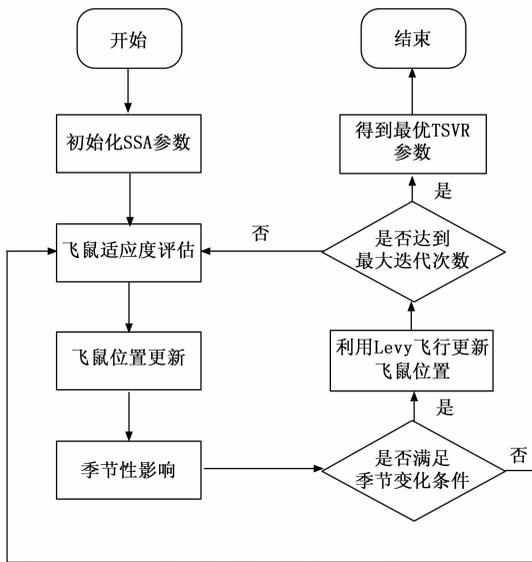


图 2 SSA 优化 TSVR 参数流程框图

4 实验与分析

4.1 特征重要性分析结果

以飞行高度为预测对象时, 选取 QAR 数据中的 ALTI-

表 5 巡航阶段飞行速度特征重要性分析结果

特征	重要度
TOTAL PRESS	0.175 63
INERTIAL ALTITUDE	0.168 71
SEL PT2	0.157 24
PRES POSN LAT	0.141 17
CENTER TANK VOLUME	0.140 43
AVERAGED STATIC PRESS	0.123 19
MAX ALLOW AIRSPEED	0.093 60

表 6 下降阶段飞行速度特征重要性分析结果

特征	重要度
PRES POSN LAT	0.418 06
INERTIAL ALTITUDE	0.308 75
AVERAGED STATIC PRESS	0.245 73
TOTAL PRESS	0.012 94
SEL PT2	0.008 92
MAX ALLOW AIRSPEED	0.005 56
CENTER TANK VOLUME	0.000 48

4.2 TSVR 预测模型评估

以飞行高度以及飞行速度为预测对象，对预测模型进行验证，根据飞行高度特征重要性分析结果，使用这些重要参数对模型进行训练，在使用不同核函数的情况下得到 TSVR 预测模型的预测精度。

需要对模型的性能进行评估，确保建立的模型具有应用价值。对于回归模型，模型的评估指标主要有均方根误差 (RMSE)、平均绝对误差 (MAE)、决定系数 (R^2)。

RMSE 对训练数据中的噪点数据最为敏感，将噪点数据的误差进行了指数倍的放大，这就对噪点数据赋予了更高的优先级，均方根误差越小，说明模型对噪点值的抗干扰能力越强。RMSE 的计算公式为：

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (21)$$

同均方根误差类似，平均绝对误差同样是对预测值与真实值之间的差值进行评估，不同于均方根误差的是，平均绝对误差将真实值与预测值的差值的绝对值进行相加，平均绝对误差越小，说明训练得到的模型的精度越高，模型越接近实际情况。MAE 的计算公式为：

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (22)$$

R^2 评估的对象是输出值，其反应了模型拟合的准确程度，结果在 0~1 之间，结果越接近 1，模型对输出值的描述越准确。 R^2 的计算公式为。

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (23)$$

其中： y 和 y_i 是实际值和预测值， i 是数据记录编号， \bar{y}

是实际值的平均值，是数据总数。

利用前面进行的特征重要性分析结果，选取相应数据对 TSVR 模型进行训练，以高度为预测对象，在爬升阶段的各核函数预测值与真实值对比结果如图 3 所示。在此阶段中，poly 核函数的预测性能最强， R^2 为 0.919 0，RMSE 为 0.082 9，MAE 为 0.063 5。

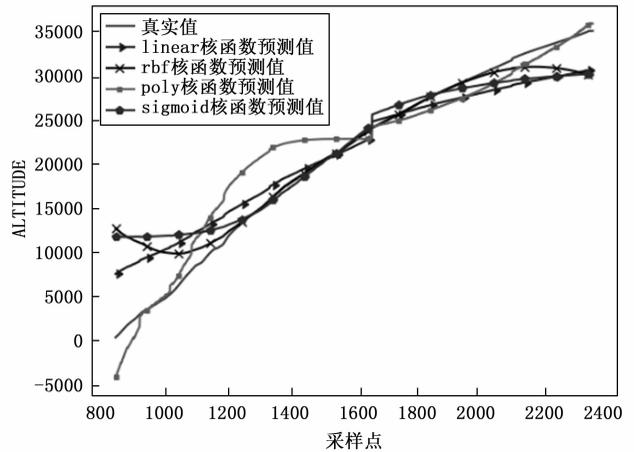


图 3 爬升阶段飞行高度 TSVR 预测对比

以飞行高度为预测对象，在巡航阶段的各核函数预测值与真实值对比结果如图 4 所示。在此阶段中，linear 核函数的预测性能最强， R^2 为 0.940 3，RMSE 为 0.087 7，MAE 为 0.080 5。

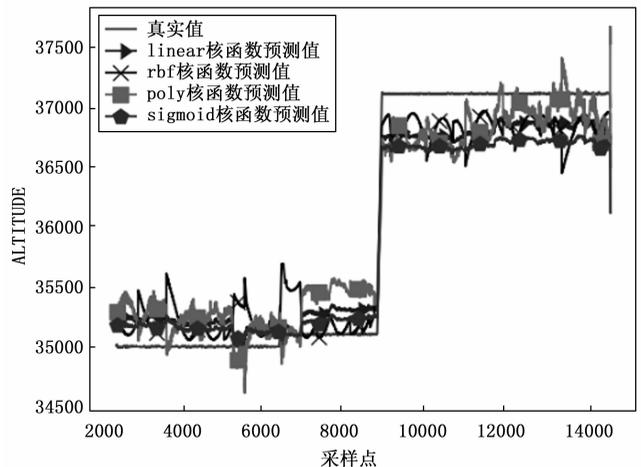


图 4 巡航阶段飞行高度 TSVR 预测对比

以飞行高度为预测对象，在下降阶段的各核函数预测值与真实值对比结果如图 5 所示。在此阶段中，rbf 核函数的预测性能最强， R^2 为 0.926 3，RMSE 为 0.081 5，MAE 为 0.046 8。

以飞行速度为预测对象，在爬升阶段的各核函数预测值与真实值对比结果如图 6 所示。在此阶段中，linear 核函数的预测性能最强， R^2 为 0.939 6，RMSE 为 0.061 0，MAE 为 0.039 3。

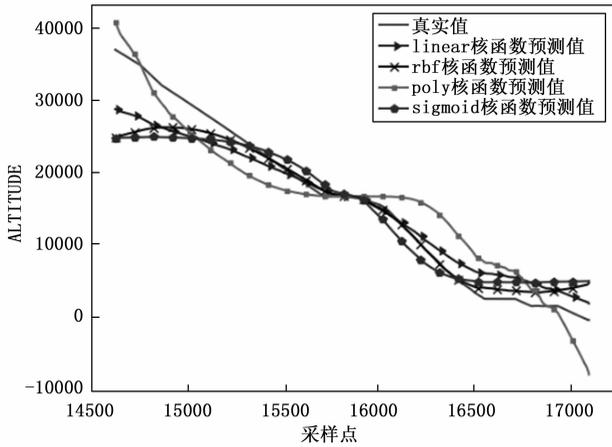


图 5 下降阶段飞行高度 TSVR 预测对比

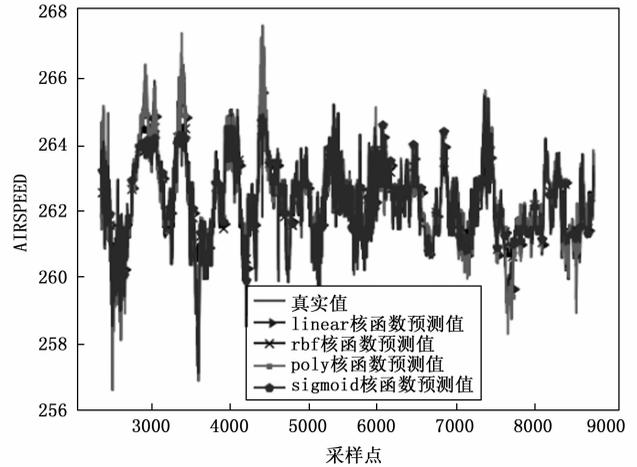


图 7 巡航阶段飞行速度 TSVR 预测对比

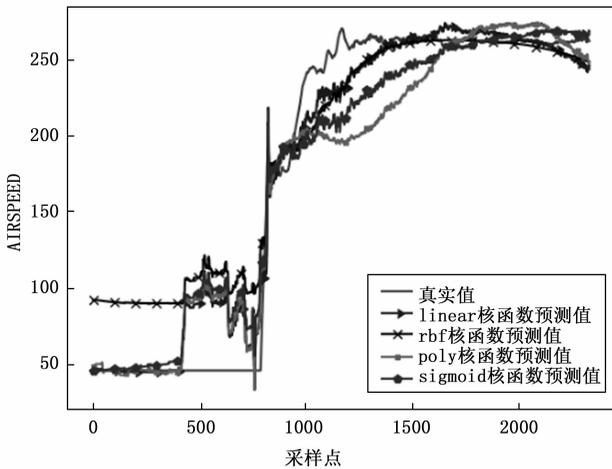


图 6 爬升阶段飞行速度 TSVR 预测对比

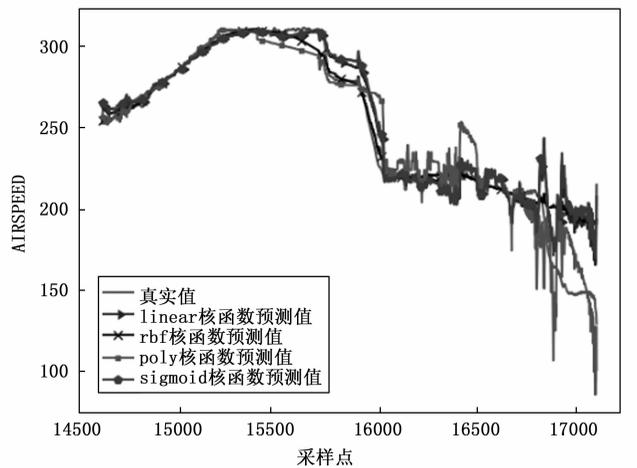


图 8 下降阶段飞行速度 TSVR 预测对比

以飞行速度为预测对象, 在巡航阶段的各核函数预测值与真实值对比结果如图 7 所示。在此阶段中, linear 核函数的预测性能最强, R^2 为 0.965 2, RMSE 为 0.027 6, MAE 为 0.020 3。

以飞行速度为预测对象, 在下降阶段的各核函数预测值与真实值对比结果如图 8 所示。在此阶段中, poly 核函数的预测性能最强, R^2 为 0.925 3, RMSE 为 0.071 9, MAE 为 0.045 7。

对上述指标及图像进行分析, 在以飞行高度为预测对象时, 在 3 个飞行阶段中预测精度最高的核函数分别为 poly、linear、rbf 核函数。在不同的飞行阶段中, 模型的预测精度变化较大, 例如在爬升以及下降阶段中, 不同核函数之间的 R^2 最大相差 0.061 8、0.067 8, 但在巡航阶段, 不同核函数之间最大差值仅有 0.026 2, 这是由于在爬升、下降阶段中, 飞行高度数值变化范围较大、变化更加频繁, 导致模型预测精度较低, 并且不同核函数之间预测精度也相差较大。以飞行速度为预测对象时, 在 3 个飞行阶段中预测精度最高的核函数分别为 linear、linear、poly 核函数。

综上所述, linear 核函数与 poly 核函数的预测性能比较好, 而 sigmoid 核函数的预测性能较差, 在六次实验当中均不是最优核函数。

4.3 SSA-TSVR 预测模型评估

利用 SSA 算法对 TSVR 模型中的超参数进行寻优, 并根据寻优得到的核函数参数 gamma、不敏感参数和惩罚因子 C 对模型进行训练。以飞行高度为预测对象时, 在爬升、巡航、下降阶段的优化后结果分别如表 7~9 所示。

表 7 爬升阶段飞行高度 SSA-TSVR 优化对比

	核函数	R^2	RMSE	MAE
优化前	poly	0.919 0	0.082 9	0.063 5
优化后	linear	0.971 4	0.049 2	0.023 4

表 8 巡航阶段飞行高度 SSA-TSVR 优化对比

	核函数	R^2	RMSE	MAE
优化前	linear	0.940 3	0.087 7	0.080 5
优化后	sigmoid	0.989 0	0.036 5	0.017 0

表 9 下降阶段飞行高度 SSA-TSVR 优化对比

	核函数	R^2	RMSE	MAE
优化前	rbf	0.926 3	0.081 5	0.046 8
优化后	rbf	0.973 6	0.045 5	0.032 7

以飞行速度为预测对象时，在爬升、巡航、下降阶段的优化后结果分别如表 10~12 所示。

表 10 爬升阶段飞行速度 SSA-TSVR 优化对比

	核函数	R^2	RMSE	MAE
优化前	linear	0.939 6	0.061 0	0.039 3
优化后	linear	0.974 8	0.036 5	0.019 7

表 11 爬升阶段飞行速度 SSA-TSVR 优化对比

	核函数	R^2	RMSE	MAE
优化前	linear	0.965 2	0.027 6	0.020 3
优化后	linear	0.982 8	0.024 5	0.018 2

表 12 爬升阶段飞行速度 SSA-TSVR 优化对比

	核函数	R^2	RMSE	MAE
优化前	poly	0.925 3	0.071 9	0.045 7
优化后	poly	0.976 0	0.020 8	0.019 9

为了更直观地展示优化前后的效果对比，选择以飞行高度为预测对象，将 TSVR 模型最优核函数预测值与 SSA-TSVR 最优核函数预测值进行比较，在爬升阶段与下降阶段的对比分别如图 9、图 10 所示。

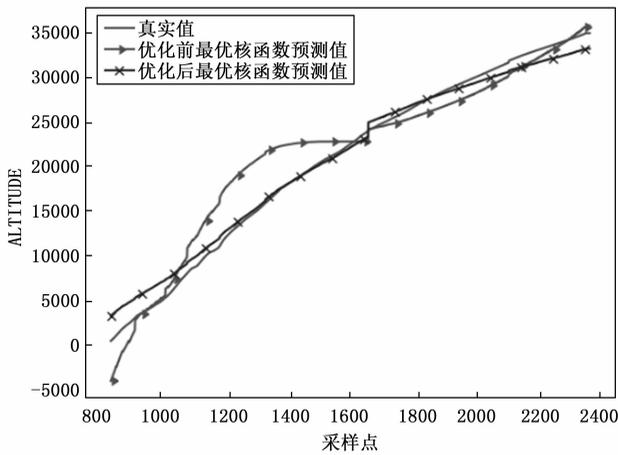


图 9 爬升阶段飞行高度优化前后对比

通过与优化前的 TSVR 模型进行对比，以飞行高度为预测对象时，不同飞行阶段中的最优核函数分别为 linear、sigmoid、rbf 核函数，在爬升阶段中，SSA-TSVR 模型 R^2 提高了 5.70%，RMSE 降低了 40%，MAE 降低了 63%；在巡航阶段中， R^2 提高了 5.17%，RMSE 降低了 58%，MAE 降低了 78%；在下降阶段中， R^2 提高了 5.11%，RMSE 降低了 55%，MAE 降低了 63%。

以飞行速度为预测对象时，不同飞行阶段中的最优核

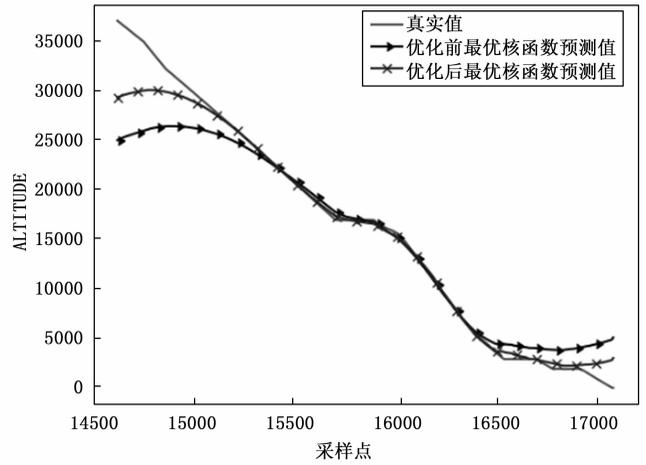


图 10 下降阶段飞行高度优化前后对比

函数分别为 linear、sigmoid、rbf 核函数，在爬升阶段中，SSA-TSVR 模型 R^2 提高了 3.75%，RMSE 降低了 40%，MAE 降低了 49%；在巡航阶段中， R^2 提高了 1.82%，RMSE 降低了 11%，MAE 降低了 10%；在下降阶段中， R^2 提高了 5.48%，RMSE 降低了 49%，MAE 降低了 62%。

通过对 SSA-TSVR 模型与 TSVR 模型的对比分析，优化后的预测模型在 R^2 、RMSE、MAE 等评估指标上均有一定程度上的提升。除此之外，优化前预测性能较差的核函数经过 SSA 算法的超参数寻优，预测性能有着很大程度上的提升。以 sigmoid 核函数为例，在 TSVR 模型中的表现较差，在六次实验中均为作为最优核函数，但在 SSA-TSVR 模型中，有两次作为最优核函数进行输出。

综上所述，建立的 SSA-TSVR 数据驱动预测模型不仅能够部分数据缺失的情况下对飞机状态参数进行预测，且在预测精度方面有着较好表现。解决了关键飞行状态丢失时无法进行飞行状态实时监控的问题。

5 结束语

安全是民航永恒的主题，没有安全就没有发展，没有安全就没有效益，只有加强应急管理体系，才能不断提高民航抵御各种风险和应对突发事件的能力。

针对飞机在飞行过程中出现意外状况从而导致的关键飞行参数丢失问题，提出了一种基于 TSVR 的飞机状态预测模型，此模型可以使用不同核函数对待预测对象进行预测，并且能够凭借 TSVR 在模型训练过程中时间复杂度较低的优势，来增强预测模型的实时性。

针对 TSVR 模型在实际使用过程中需要手动对模型超参数进行调参的问题并且为了提高模型的预测精度，提出一种 SSA-TSVR 预测模型，旨在结合 SSA 优化精度高、收敛速度快、跳出局部最优能力较强的优势对模型进行优化。

在模型验证过程中，利用 QAR 数据模拟部分关键飞行参数丢失的情况，并使用随机森林算法对数据进行特征重要性分析，选取与待预测对象关系最密切的参数，并利用这

(下转第 156 页)