第 59 卷 第 2 期/2022 年 1 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于双层特征选择的空中目标分类算法研究

苏志刚[1,2*]，王雪萌[1]

[1]中国民航大学电子信息与自动化学院，天津 300300；

[2]中国民航大学中欧航空工程师学院，天津 300300

摘要 空中生物和非生物目标的分类是机场鸟击防治工作中的重要一环。基于轨迹信息的目标分类具有轨迹信息获得容易、部分特征区分度高的优势，但是不当的特征选择会导致近距离轨迹样本分类误差大。针对该问题，提出了一种基于双层特征选择的空中目标分类算法。首先对动态目标三维轨迹数据进行充分的特征提取，扩大特征选择范围；其次通过设计的双层特征选择算法选择特征子集，减少算法运算量，提高分类精细度；最后通过在线顺序极限学习机(OSELM)实现空中生物和非生物目标的实时分类。实验结果表明，所提算法兼顾了分类的精度与速度，分类精度达到了99.7%，平均分类时间仅为1.26 ms，满足了实时监测预警的需求。所提算法为机场条件下空中目标的实时分类提供了一种极具潜力的解决方案。

关键词 图像处理；在线顺序极限学习机；特征提取；特征选择；实时分类

中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP202259.0210018

Aerial Target Classification Algorithm Based on Double-Layer Feature Selection

Su Zhigang[1,2*], Wang Xuemeng[1]

[1]*College of Electronic Information and Automation, Civil Aviation University of China, Tianjin* 300300, *China*;

[2]*Sino-European Institute of Aviation Engineering, Civil Aviation University of China, Tianjin* 300300, *China*

Abstract The classification of biological and abiotic targets in the air is an important part of bird strike control in the airport. Target classification based on trajectory information has the advantages of easy access to trajectory information and high degree of discrimination of some features, but improper feature selection will result in large classification errors of close-range trajectory samples. Aiming at this problem, an aerial target classification algorithm based on double-layer feature selection is proposed. First, fully feature extraction is performed on the three-dimensional trajectory data of dynamic targets to expand the range of feature selection. Second, the feature subset is selected through the designed two-layer feature selection algorithm, which reduces the computational complexity of the algorithm and improves the classification precision. Finally, online sequential extreme learning machine (OSELM) is used to realize the real-time classification of aerial biological and abiotic targets. Experimental results show that the proposed algorithm takes into account the accuracy and speed of classification, the classification accuracy reaches 99.7%, and the average classification time is only 1.26 ms, which meets the needs of real-time monitoring and early warning. The proposed algorithm provides a potential solution for real-time classification of air targets under airport conditions.

Key words image processing; online sequential extreme learning machine; feature extraction; feature selection; real-time classification

收稿日期：2021-01-12；**修回日期**：2021-02-23；**录用日期**：2021-03-16

基金项目：军队后勤开放研究科研项目(BY119C009)、中国民航大学研究生科研创新项目(10502761)

通信作者：*ssrsu@vip.sina.com

1 引 言

机场周边鸟类活动严重威胁航空安全[1]，鸟击事故常常造成极大的人员及财产损失，因此无论军事机场还是民用机场均为消除鸟害影响在鸟类目标的探测识别及防治方面做了大量尝试。机场作为飞机的起降场所，机场周边的鸟击防治是一项保障飞行安全的重要工作，而对空中目标的实时分类是做好鸟情预警防治工作的重要前提。

雷达是国内外探测鸟情的主要手段，目前国内外研究学者提出了多种基于雷达探测的空中目标分类方法，这些方法大致可分为基于微多普勒特征的方法和基于目标运动的方法两类[2]。基于微多普勒特征的方法利用目标微动特征对鸟类和航空器进行区分[3]，特征通过鸟情监视雷达采集飞鸟翅膀扇动信息得到，翅膀扇动对回波产生的调制特性具有时变性和周期性，对微动信息建模并进行精细的特性认知，从而实现对鸟类和航空器的区分。研究人员通过提取目标的微多普勒特征来识别无人机和飞鸟。Molchanov 等[4]利用短时傅里叶变换获取微多普勒特征，并从微多普勒特征的相关矩阵中提取了特征对作为训练 3 个分类器（线性和非线性支持向量机分类器及朴素贝叶斯分类器）的输入特征，最高精度达到了 95%。然而基于微多普勒特征的方法的缺点是分类结果依赖于目标反射回波，但反射回波易受到地面杂波及降水等空域杂波的干扰，且剔除这些干扰的难度较大[5]，因此该方法存在一定的局限性。

基于运动信息的分类方法则是利用目标飞行运动的差异进行分类的。将轨迹信息作为空中目标分类的依据具有以下优势：相较于目标回波，轨迹信息是通过目标跟踪将每个真实目标在时间序列上检验并关联得到的[6]，目标跟踪能够消除大部分孤立的虚警。且鸟类和航空器的轨迹信息有明显的特征区分度[7]。通常情况下，生物目标飞行轨迹机动性较强，灵活多变[8]；而非生物目标飞行轨迹变化率较小，较为稳定，因而可以将轨迹作为分类的有效依据。同时随着激光雷达技术的进步[9]，主动探测技术和单光子探测技术能够实现精确测距，有效提高了作用距离[10]，较容易获得飞鸟这类典型“低小慢”目标的轨迹信息。Mohajerin 等[11]利用目标航迹的统计特征对航空器和鸟类进行分类，即从目标运动数据中手动提取有效特征，如速度、加速度和加速度的均值和方差，此外，还采用人工神经网络（MLP）对 4 个与目标雷达截面相关的特征进行分类。Chen 等[12]提出了一种基于计算模型出现概率的时域方差概率运动模型估计方法，该方法利用来自监视雷达的数据在航空器和鸟类之间进行分类，利用目标的运动方向、速度和位置信息建立它们的运动估计模型，以目标运动模型的转换频率为特征，同时提出了一种平滑算法来扩大鸟类和无人机的目标模型转换频率的估计之间的差距。上述大部分方法均需要有深厚的领域知识来构建特征库，并且特征库的好坏对分类结果有决定性影响。同时这些方法缺乏特征选择的步骤，尽管信息中存在对于分类结果贡献度较高的特征，但是目前的特征选择较为依靠人为经验，如何从较多特征中精确地进行特征选择是目前面临的主要问题。其次在机场条件下，针对空中目标的分类对实时性有较高要求，因此如何实现模型的快速训练并投入分类使用也是急需解决的重点。

针对上述问题，提出了一种基于双层特征选择的空中目标实时分类算法。首先使用基于可扩展假设检验的时序特征提取（TSFRESH）对轨迹数据进行特征提取；其次设计了一种双层特征选择算法进行特征子集选择，第 1 层基于假设检验进行初步特征筛选，第 2 层基于去冗余度的 relief 算法进行特征选择；最后将与分类结果相关度高的冗余度低的特征子集输入在线顺序极限学习机（OSELM）分类器，从而完成对空中目标的实时分类。所提算法可以在没有深厚的领域知识的条件下，实现自动提取和选择对于分类较为有效的特征，解决了特征选择困难的问题，在不增加硬件成本的基础上提高了分类精度，推进了机场环境下鸟情监测技术的发展。

2 问题描述

在机场净空条件下，空中的监视目标主要是航空器与飞鸟两类。根据监视系统的监视范围及数据重访率可以确定单条轨迹包含的最大点迹数 N，对监视系统采集的目标持续轨迹进行滑动截取或分段截取可获得 K 条轨迹。所有轨迹构成用于目标分类的数据集：

$$T=\left\{t_1,\ t_2,\ \cdots,\ t_K\right\}, \tag{1}$$

$$t_k=\left\{p_{k1},\ p_{k2},\ \cdots,\ p_{kN_k}\right\}, \tag{2}$$

式中：t_k 为数据集中第 k 条轨迹，$k=1,\ 2,\ \cdots,\ K$；p_{kn} 为第 k 条轨迹中的第 n 个位置点，$n=1,\ 2,\ \cdots,\ N_k$，

可由该位置点在站心坐标系中的坐标表示，即$p_{kn}=\left(x_{kn}, y_{kn}, z_{kn}\right)$；$N_k$为第$k$条轨迹中的点迹个数，$N_k \leqslant N$。

用于目标分类的数据集T中的轨迹既有航空器的轨迹，也有飞鸟的轨迹。由于监视鸟情的监视系统的监视范围通常较小，监视半径一般不超过15 km，该区域内航空器可被监视的时长约为1 min，导致航空器的轨迹长度较短，单条轨迹的点迹数目通常小于20。为保障分类算法输入的规范性，统一将轨迹的时长限定在1 min左右，这使得数据集T中所有轨迹的点迹数目小于20。以20作为滑动截取的步长，对目标的持续轨迹进行截取，一般将此处理叫作轨迹形成。因此，所讨论的分类问题是一个短时轨迹的分类问题。由于鸟类目标的机动性非常强，这使得所获得的轨迹t_k只是鸟类真实轨迹上的采样，与原始的轨迹存在着一定的差异，而且单纯对曲率等少量特征进行分类的效果不佳。为此，需要对获得的短时轨迹进行深度信息挖掘，从而获取关于轨迹的更充分信息。目前主流的特征提取方法包括3类：1）基于基本统计方法的特征提取方法提取数据的均值、方差、极值等统计数值作为特征；2）基于模型的特征提取方法用模型刻画时间序列，提取模型的系数作为特征；3）基于变换域的特征提取方法将时频参数和线性参数作为特征。TSFRESH是一种提取时间序列特征的工具[13]，包含多种基于统计和基于变换的时间序列提取方法。使用TSFRESH对轨迹数据进行全方位的特征提取，面对所提轨迹数据为时间序列，且分类结果为鸟类、航空器这样的二值分类问题，TSFRESH利用内置的计算时间序列的特征提取函数，对轨迹数据进行全方位的特征提取。提取结果为4200个特征。

基于短时轨迹的目标分类算法结构如图1所示。基于上述分析，目标分类的准备工作包括雷达测量的点迹数据获取，轨迹形成、TSFRESH轨迹特征提取。准备工作完成后，使用双层特征选择算法和OSELM算法输出目标分类结果。

3　双层特征选择与实时分类算法

3.1　设计的双层特征选择算法

空中目标分类的本质是模式识别问题。其中决定模式识别决策质量的关键环节是特征提取技术[14]。因此，如何从海量的轨迹数据集中有效地选择出规模小且最具辨识力的特征子集，已成为解决机场环境下实时鸟情监测问题的关键。

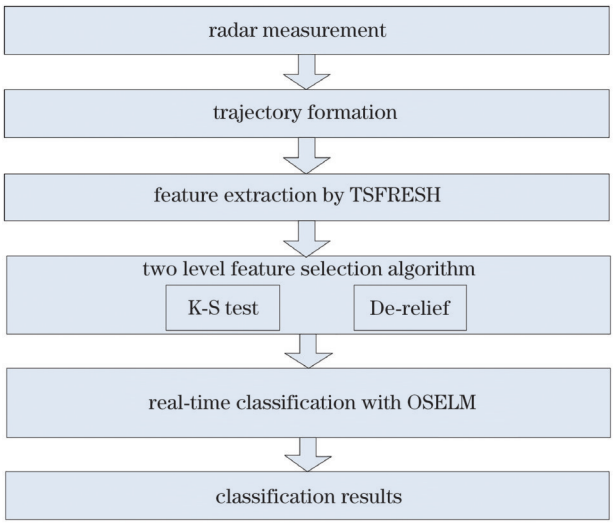

图1　空中目标特征选择与分类算法框架图

Fig. 1　Framework of air target feature selection and classification algorithm

现有的特征选择方法包括过滤式特征选择算法、封装式特征选择算法和嵌入式特征选择算法。封装式特征选择算法是特征选择和分类算法相结合的算法，需要多次调用分类器进行测试，时间复杂度较高，且特征子集的性能受特定分类算法影响，尤其在轨迹数据中包含大量近距离样本时，容易出现过拟合现象。嵌入式特征选择算法嵌入在分类算法中，分类算法训练结束就能得到特征子集，一般通过一阶正则化法实现特征选择，但该算法在处理非线性问题时效果不佳，不适合在此场景使用。过滤式特征选择算法的优点在于特征选择过程与分类过程是独立的，不必多次调用分类器进行测试，因此算法时间复杂度较低，特征评价结果适用于多种分类器。由于机场条件下空中目标分类实时性要求高，后续步骤使用在线分类器进行分类，需要适应度高的特征选择算法。且轨迹数据隐藏信息较多且较杂，特征选择步骤必不可少，因此借鉴过滤式特征选择算法的思想设计了双层特征选择算法。

双层特征选择算法的第1层使用假设检验方法[15]进行特征初步筛选，主要是进行特征数量的削减，通过判断特征与分类的相关程度进行特征的淘汰。对轨迹数据集$T=\left\{t_1, t_2, \cdots, t_K\right\}$进行特征提取后得到某一特征数据集$X$：

$$X=\{x_1, x_2, \cdots, x_K\}, \tag{3}$$

式中：x_k为数据集中第k条轨迹的某一特征X的值，$k=1, 2, \cdots, K$。使用K-S检验来判断特征X与分类结果$Y\in\{y_1, y_2\}$是否相关，y_1和y_2分别代表鸟类和航空器。$f_{X|Y=y_1}$和$f_{X|Y=y_2}$分别表示分类结果为鸟类和航空器时，特征X的经验分布。用$f_{X|Y=y_1}$和$f_{X|Y=y_2}$之间的差异程度衡量特征与分类标签的相关程度，检验问题可表示为

$$H_0: f_{X|Y=y_1}\equiv f_{X|Y=y_2} \leftrightarrow H_1: f_{X|Y=y_1}\neq f_{X|Y=y_2}, \tag{4}$$

式中：H_0和H_1分别表示鸟类与航空器在特征X上相似和存在明显差异。使用统计量Z检验假设问题：

$$Z=\max\left\{\left|f_{X|Y=y_1}-f_{X|Y=y_2}\right|\right\}。 \tag{5}$$

统计量Z对应的显著性水平由p表示，当统计量$Z\to 0, p\to 1$时，则接受H_0假设，认为特征X与分类结果Y无关，将此特征淘汰；反之，则拒绝H_0假设，将特征X筛选出来用作分类。通过K-S检验对所有特征进行判断，将与分类结果相关度较高的特征初步筛选出来。筛选出来的特征数量为原来特征的17%，为进一步的特征选择减小了运算量。尝试使用这些特征子集在时间序列上构造统计或机器学习模型[16]，在后续分类任务中使用[17]。

空中目标的轨迹数据特点是含有大量较为接近的近距离样本数据。对于此类样本，通过单一变量评分准则找到区分度较高的特征难度较大。因此选择多变量的特征选择算法relief[18]处理空中目标的轨迹数据，relief算法充分考虑了特征在同类近邻样本和异类近邻样本的差异，区分近距离样本能力强。算法对特征的选择标准与同类样本差距越小，与异类样本差距越大，则该算法特征区分能力越强[19]。Relief算法为每维特征赋予权重，以权重表征特征与类别的相关性。对于待解决的问题，首先给定经过第1层特征选择的训练特征样本集：

$$X'=\{x_1, x_2, \cdots, x_K\}。 \tag{6}$$

Relief算法随机从训练样本中选择一个特征样本x_k，当x_k对应的分类结果为鸟类时，先在鸟类样本中寻找x_k的最近邻$H(x_k)$，$H(x_k)$称为“同类近邻”；再从航空器样本中寻找x_k的近邻$M(x_k)$，$M(x_k)$称为“异类近邻”。此特征的权重v的表达式为

$$v=\sum_k\left[-\left|x_k-H(x_k)\right|+\left|x_k-M(x_k)\right|\right]。 \tag{7}$$

当一个特征对分类结果的贡献率高时，该特征到同类样本的距离较小，到非同类样本的距离较大，具有较大的权重。因此，对所有特征按照贡献率进行排序，利用预设的阈值，基于贡献率筛选出最能区分鸟类和航空器的特征子集。

双层特征选择算法的第2层使用改进的relief算法。Relief算法的劣势在于仅考虑了特征与分类标签之间的相关程度，而忽略了特征与特征之间的冗余程度。在特征选择的过程中，特征权重排名前列的特征组合并不一定能够达到最好的分类效果，这是由于特征之间存在冗余特征，冗余特征之间发生相互干扰导致分类性能的下降，且过多的冗余特征会大大增加训练模型的时间。

因此在relief算法中加入评估特征间关系的指标特征冗余度，提出了一种基于去冗余度的relief算法（De-relief）。De-relief算法在relief算法计算特征权重的基础上，加入了基于特征冗余度的特征筛选。轨迹信息是连续变量，令排序后特征的I元随机变量为$(\boldsymbol{F}_1, \boldsymbol{F}_2, \cdots, \boldsymbol{F}_I)$，relief算法计算出的特征权重向量$\boldsymbol{v}=[\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_I]$，对权重值进行排序后的权重向量为$\tilde{\boldsymbol{v}}=[\tilde{\boldsymbol{v}}_1, \tilde{\boldsymbol{v}}_2, \cdots, \tilde{\boldsymbol{v}}_I]$。为了尽可能选出冗余度低的特征子集，引入互信息指标。互信息是信息论中的一种信息度量，是用来衡量变量间相互依赖性的量度。任意两个随机变量之间的互信息的表达式为

$$I(\boldsymbol{F}_i, \boldsymbol{F}_j)=\iint_{f_i f_j} p(f_i, f_j)\log\frac{p(f_i, f_j)}{p(f_i)p(f_j)}, \tag{8}$$

式中：$p(f_i, f_j)$为特征随机变量$(\boldsymbol{F}_i, \boldsymbol{F}_j)$的联合分布；$p(f_i)$和$p(f_j)$分别为随机变量$\boldsymbol{F}_i$和$\boldsymbol{F}_j$的边缘分布。设轨迹数据的特征索引集合为$\boldsymbol{D}$，并初始化为包含权重最高的索引。令$\tilde{\boldsymbol{D}}=\{1, 2, \cdots, I\}-\boldsymbol{D}$，则某一特征$i\in\tilde{\boldsymbol{D}}$关于特征子集$\boldsymbol{D}$的冗余度的表达式为

$$R_{D,i}=\frac{\sum_{j\in D} I(\boldsymbol{F}_i, \boldsymbol{F}_j)}{|\boldsymbol{D}|}。 \tag{9}$$

特征冗余度计算完成后，综合特征权重的影响对特征进行重新打分。每个特征的打分公式为

$$g_i=\frac{\tilde{v}_i}{R_{D,i}}, i\in\tilde{\boldsymbol{D}}。 \tag{10}$$

然后将分数最高的特征从特征子集$\tilde{\boldsymbol{D}}$移入特

征子集$\boldsymbol{D}$中，以此循环。最后将最终生成的轨迹特征子集$\boldsymbol{D}$作为后续分类器的输入。

通过双层特征选择算法可以在低时间复杂度条件下选择出相关度高且冗余度低的特征子集，解决了特征选择困难的问题，保证了后续分类的精确性和实时性。

3.2　实时分类算法

机场条件下，鸟情防治工作对实时性有极高的要求。传统分类算法的分类速度无法满足实时分类场景的需求，因此引入实时分类算法解决实时性问题。

极限学习机(ELM)算法是一种单隐层神经网络算法[20]。ELM算法将传统神经网络的训练问题转化为求解线性问题的最小二乘解问题，收敛速度快，达到局部最优的时间较短。且该算法不需要更新隐层节点个数，经过单次学习就可以算出唯一的最优解，本质是通过随机初始化输入权重和偏置得到相应的输出权重，结构如图2所示。

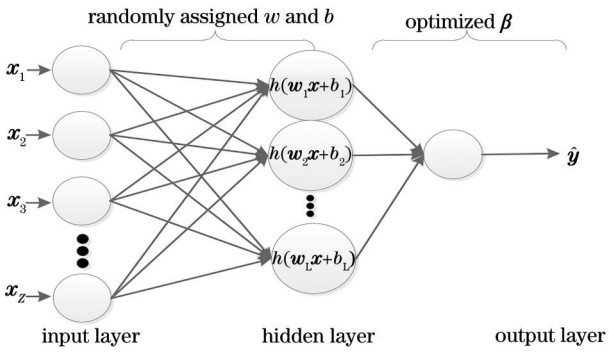

图2　ELM结构

Fig. 2　Structure of ELM

对于轨迹数据集中轨迹t_k的Z个不同特征，具有L个隐层节点，且激活函数为$h(\cdot)$的单隐层ELM算法，数学模型为

$$\hat{\boldsymbol{y}}_z=\sum_{l=1}^{L}\boldsymbol{\beta}_l h\left(\boldsymbol{w}_l\boldsymbol{x}_z+b_l\right)+\boldsymbol{\beta}_0,\ z=1,\cdots,Z,\tag{11}$$

式中：$\boldsymbol{x}_z\in\mathbf{R}^Z$和$\hat{\boldsymbol{y}}_l\in\mathbf{R}^m$为模型输入和输出，分别是特征子集和分类结果；$\boldsymbol{w}_l$和$b_l$为第$l$个隐藏层的输入权重和偏置参数；$\boldsymbol{\beta}_l$和$\boldsymbol{\beta}_0$是输出层的权重和偏置参数。

在ELM算法中，训练隐藏层相当于求出线性系统的最小二乘解。(11)式的Z个等式可以表示为

$$\boldsymbol{H}\boldsymbol{\beta}=\boldsymbol{Y},\tag{12}$$

$$\boldsymbol{H}=\begin{bmatrix}h\left(\boldsymbol{w}_1\boldsymbol{x}_1+b_1\right) & \cdots & h\left(\boldsymbol{w}_L\boldsymbol{x}_1+b_L\right)\\ \vdots & \cdots & \vdots\\ h\left(\boldsymbol{w}_1\boldsymbol{x}_Z+b_1\right) & \cdots & h\left(\boldsymbol{w}_L\boldsymbol{x}_Z+b_L\right)\end{bmatrix},\tag{13}$$

$$\begin{cases}\boldsymbol{\beta}=\begin{bmatrix}\boldsymbol{\beta}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{\beta}_L^{\mathrm{T}}\end{bmatrix}\\ \boldsymbol{Y}=\begin{bmatrix}\boldsymbol{y}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{y}_Z^{\mathrm{T}}\end{bmatrix}\end{cases},\tag{14}$$

式中：权值矩阵$\boldsymbol{\beta}$的维度为$L\times m$；输出数据矩阵$\boldsymbol{Y}$的维度为$Z\times m$。

$\boldsymbol{H}^{+}$为$\boldsymbol{H}$的广义逆矩阵。如果$\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}$非奇异，则(12)式可以改写为

$$\hat{\boldsymbol{\beta}}=\boldsymbol{H}^{+}\boldsymbol{Y}=\boldsymbol{P}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{Y},\tag{15}$$

$$\boldsymbol{P}=\left(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\right)^{-1},\tag{16}$$

式中：$\boldsymbol{P}$是隐藏层输出的协方差矩阵的逆。

传统ELM算法假定所有训练数据都可用于训练，整个学习过程需要一次全部完成。将鸟类与航空器轨迹数据全部用于模型训练的缺点有2个：1) 模型训练时间过长；2) 新的轨迹数据传来时，模型中参数值无法及时更新。为满足空中目标分类的实时性需求，对传统ELM进行改进来解决上述问题。

OSELM算法是一种增量式快速学习算法[21]，可以逐块学习新的数据块，实时进行模型参数的更新。OSELM算法可以分为2个阶段：初始化阶段和在线顺序学习阶段。初始化阶段沿用了ELM的求解过程，通过随机选取一部分的初始特征训练数据集$\{\boldsymbol{X}_0,\boldsymbol{Y}_0\}=\{\boldsymbol{x}_z,\boldsymbol{y}_z\}_{z=1}^{Z}$，对该数据集应用ELM算法，随机分配参数$\boldsymbol{w}_l$和$b_l$，求解$\boldsymbol{\beta}$矩阵，推算出隐藏层输出矩阵$\boldsymbol{H}_0$；在线顺序学习阶段中，每当获取到第$k+1$个新的训练数据块时，利用初始化阶段求解的结果，仅使用新数据计算$\boldsymbol{H}_{k+1}$，递推计算输出权值：

$$\boldsymbol{P}_{k+1}=\boldsymbol{P}_k-\boldsymbol{P}_k\boldsymbol{H}_{k+1}^{\mathrm{T}}\left(\boldsymbol{I}+\boldsymbol{H}_{k+1}\boldsymbol{P}_k\boldsymbol{H}_{k+1}^{\mathrm{T}}\right)^{-1}\boldsymbol{H}_{k+1}\boldsymbol{P}_k,\tag{17}$$

$$\hat{\boldsymbol{\beta}}_{k+1}=\hat{\boldsymbol{\beta}}_k+\boldsymbol{P}_{k+1}\boldsymbol{H}_{k+1}^{\mathrm{T}}\left(\boldsymbol{Y}_{k+1}-\boldsymbol{H}_{k+1}\hat{\boldsymbol{\beta}}_k\right)。\tag{18}$$

相比于其他流行的在线学习算法，OSELM是一种学习速度较快且泛化能力较好的在线学习算法，能够较好地应用在空中生物实时分类场景中。将双层特征选择算法与OSELM算法相结合，命名为Ret-OSELM算法。

4　实验结果

4.1　数据介绍及数据预处理

鸟类轨迹数据来自开源鸟类轨迹数据库[22]。该数据是通过高分辨率微型GPS设备获取的，GPS设备质量为16 g，占受试者质量的3%～4%。所有受试者最初都配备了橡皮泥假人砝码，砝码质量为16 g，与GPS记录器的尺寸和质量相同，用弹性安全带固定在背部，使受试者习惯于在负载下飞行和生活（设备的时间分辨率为0.2 s）。

数据预处理时，用高斯滤波器（高斯形的标准差$\sigma=0.4$ s）对飞行轨迹坐标进行平滑处理，并用3次β样条方法将曲线拟合到采样率为0.2 s的点上。

航空器数据来自开源网站Flightradar24。获取数据后，进行数据解析、航班归类等从而提取到需要的轨迹信息。同时，为了仿真激光雷达获取的目标轨迹特征，需要对真实轨迹数据进行处理，包括数据清洗、坐标转换、单位统一换算、基于激光雷达探测范围的轨迹选取4个步骤。由于轨迹数据获得过程中会面临GPS设备失联和数据重复采集的情况，轨迹数据中存在一些缺失值和重复值。数据清洗包括缺失值清洗和内容格式清洗，其中缺失值清洗是通过轨迹数据的缺失比例确定缺失值范围，对需要补齐的字段进行缺失值补齐；内容格式清洗是对轨迹数据进行去重处理和去除格式错误处理。坐标转换的目的是统一坐标系，主要是对飞机的坐标进行转换。最后对两类数据的单位进行统一换算，确定激光雷达的成像作用距离，通过分析激光雷达发射的单像素单脉冲接收光子数和工作距离的关系，将探测距离设置为15 km，并将重访率设置为5 s。

基于以上预处理步骤，数据集共有41835个坐标信息，以每15个坐标信息整合为一个数据块，共有2789个数据块。

4.2　实验初始参数的选取

相同的算法，相同的数据，但每次计算得到的结果都不同。原因是算法中存在导致最终的结果不稳定的随机因素。因此，为了比较随机算法的优劣或检验参数的最优解，需要多次实验，取平均值。在初始参数实验中，以8∶2的比例划分训练集和测试集。

图3为实验重复次数与分值均值函数关系图。从图中可以看出：起初，均值的波动幅度较大；随着重复次数的增长，均值也很快收敛到期望值附近；重复次数在200次以内时，曲线波动较大；当实验超过400次之后，均值几乎趋于稳定。因此，将实验的重复次数固定在400较为合理。

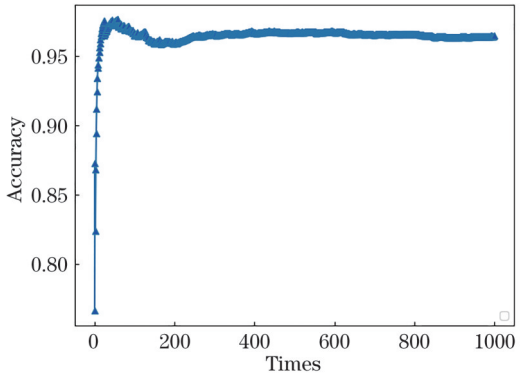

图3　实验重复次数与分值均值的关系

Fig. 3　Relationship between number of experimental repetitions and average score

相比于其他的神经网络算法，OSELM分类器需要确定隐层神经元个数和传递函数。只有确定了隐层神经元个数，OSELM分类器才能求出准确的输出权重；传递函数的作用是将权值结果转化成分类结果。神经网络常用的传递函数有Sigmoid、ReLU和Hadlim，对3种传递函数分别进行实验，分类的均方根误差（RSE）随隐层神经元个数变化的情况如图4所示。从图中可以看出：以Hadlim为传递函数时，误差较大；以ReLU和Sigmoid作为传递函数且隐层神经元个数较多时，两者误差都较小，但是在隐层神经元个数较少时，Sigmoid作为传递函数效果较好。因此，将模型的传递函数设置为Sigmoid。

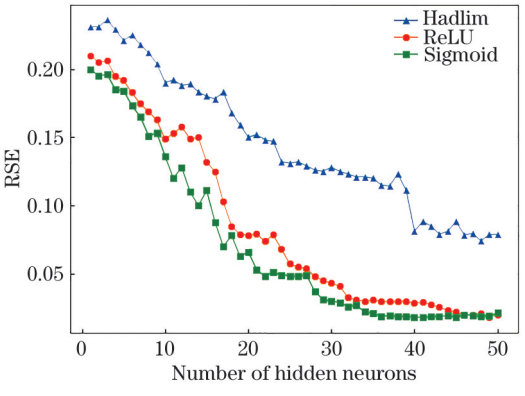

图4　不同传递函数实验效果对比

Fig. 4　Comparison of experimental effects of different transfer functions

接下来确定隐层神经元个数，初始训练数据个数取3000，初始化隐层神经元个数为1，不断增加隐层神经元个数，在Sigmoid函数下的效果如图5所示。从图中可以看出，训练误差及测试误差随神经元个数的增加而降低，当神经元个数增加为40时，

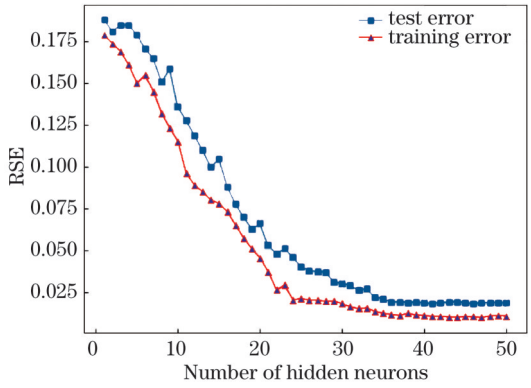

图5　训练误差、测试误差与隐层神经元个数关系

Fig. 5　Relationship between training error, test error and number of hidden layer neurons

RSE降为最低。由于随着神经元个数的增大，训练时间也随之变长。因此将神经元个数设为40。

4.3　对比实验

为了验证双层特征选择算法的有效性，与过滤式特征选择(FCBF)算法、封装式特征选择(EDA)算法、嵌入式特征选择(Lasso)算法进行了对比实验。FCBF算法是一种基于特征相关性和冗余性的特征选择框架，使用近似马尔科夫毯算法进行特征子集选取；EDA算法是一种近几年在进化领域新兴的优化算法，通过种群迭代更新特征权重；Lasso算法将特征选择过程与分类器训练过程融合，通过对系数添加惩罚项进行特征的稀疏选择。为了避免实验结果的不确定性，使用10-fold交叉验证对特征子集进行评价，将2789个轨迹数据块随机分为10等份，依次选取1份作为测试集，其他数据块作为测试集。

特征选择方法一般使用分类器的准确度评判特征子集的优劣，因此使用4种算法对鸟机轨迹数据分别进行特征选择，结合OSELM分类器进行分类，得到的平均结果如表1所示。从表中可以看出，双层特征选择算法在特征相关度标准下，挖掘出特征与类别及特征间的依赖关系，保留了与类别相关程度最高、最有区分度的特征子集，对比其他主流特征选择算法，双层特征选择算法选择的特征子集的分类精度较高。

表1　不同特征选择算法精度与标准差比较

Table 1　Comparison of accuracy and standard deviation of different feature selection algorithms　　unit: %

Algorithm	Precision	Standard deviation
K-S test+De-relief	99.78	±0.02
FCBF	97.75	±0.08
EDA	99.11	±0.12
Lasso	95.10	±0.06

同时改变选择特征的数量，各算法选择出的特征子集分类精度变化如图6所示。从图中可以看出，双层特征选择算法与其他算法在降维能力上的差异。在考虑了相关度的前提下，从冗余度的角度公平地删除部分冗余特征，双层特征选择算法能够选择出规模小且分类精度高的特征子集，不降低分类精度的同时增强了算法的降维能力。

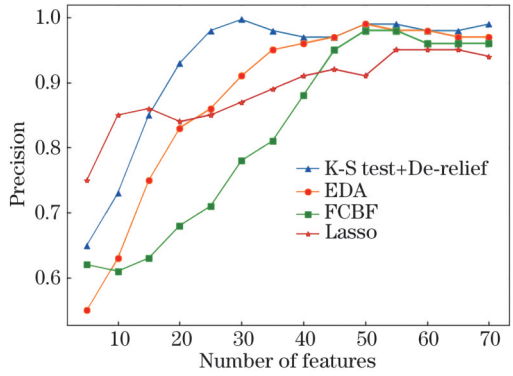

图6　各算法特征数量与精度对比

Fig. 6　Comparison of feature quantity and precision of each algorithm

对所提Ret-OSELM分类算法与LSTM神经网络、DNN神经网络、朴素贝叶斯和决策树4种在轨迹分类中较为常见的算法进行了对比，实验结果如表2所示，其中真正例率(TPR)代表预测为正例且真实情况为正例的占所有真实情况中正例的比例；反正例率(FPR)表示预测为正例但真实情况为反例的占所有真实情况中反例的比例。TPR越大，FPR越小，代表分类的结果越有可能是正确的。从表中可以看出：在精度指标上看，Ret-OSELM分类算法最高，其次是LSTM神经网络，再其次是DNN神经网络，最后是决策树和朴素贝叶斯；在TPR与FPR指标上看，Ret-OSELM分类算法的效果整体优于其他4种算法。原因是双层特征选择算法使用改进的relief算法，充分考虑了特征在同类近邻样本和异类近邻样本的差异，区分近距离样本能力较强。

图7为不同模型的训练时间对比图。从图中可

表 2　各分类算法的分类性能比较

Table 2　Comparison of classification performance of various classification algorithms

Algorithm	Precision /%	TPR	FPR
Ret-OSELM	99.7	0.996	0.016
LSTM	98.5	0.983	0.036
DNN	97.8	0.979	0.055
Naive Bayes	92.1	0.935	0.081
Decision tree	95.6	0.962	0.058

以看出，经过双层特征选择算法后模型的训练时间明显低于其他 3 种情况，在训练时间上有着明显的优势。原因在于双层特征选择算法选择的特征子集相关度高且规模较小，因此所需训练时间也大大减少。

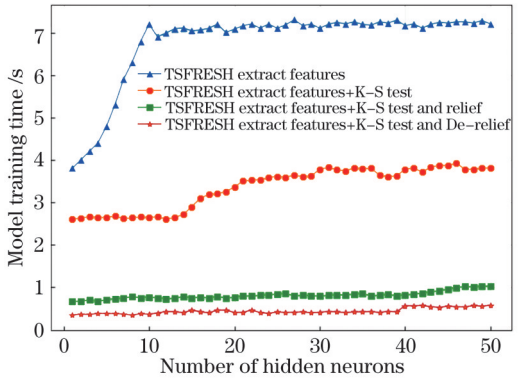

图 7　训练时间对比

Fig. 7　Comparison of training time

由于双层特征选择算法可以选出对于分类结果贡献较大且规模较小的特征子集，在后续数据的处理上可以直接计算这部分特征用于分类，而不需要再经过 TSFRESH 进行大量的特征计算，这大大减少了在特征提取方面的时间。同时不同特征提取算法的特征输入 OSELM 分类器进行测试的测试时间差异也很大，结果如表 3 所示。从表中可以看出，所提算法的测试时间大大缩短，这对于机场鸟类实时分辨并进行防治有着重要意义。

表 3　不同特征提取算法测试时间比较

Table 3　Comparison of test time of different feature extraction algorithms

Algorithm	Test time /ms
TSFRESH extract features	52.13
TSFRESH extract+K-S test	15.87
TSFRESH extracts+K-S test+relief	2.11
TSFRESH extracts+K-S test+De-relief	1.26

5　结　　论

为了满足在机场环境下空中生物目标和非生物目标分类实时性和准确性的需求，针对基于轨迹信息的目标分类特征选择困难的问题，提出了一种基于双层特征选择的空中目标实时分类算法。首先引入 TSFRESH 进行特征提取，用以提高特征可选范围；其次引入基于假设检验和基于改进的 relief 算法的双层特征选择算法，用以选择出高相关度、低冗余度的特征子集，减少分类算法运算量，提高算法分类精度；最后将筛选出的小规模且高贡献率的特征子集输入 OSELM 进行实时在线分类。实验结果表明，所提算法在准确性和实时性方面有着均衡的性能，能够满足机场条件下实时监测鸟情的需求。

参　考　文　献

[1] Li Y L, Shi X P. Investigation of the present status of research on bird impacting on commercial airplanes [J]. Acta Aeronautica et Astronautica Sinica, 2012, 33(2): 189-198.
李玉龙, 石霄鹏. 民用飞机鸟撞研究现状[J]. 航空学报, 2012, 33(2): 189-198.

[2] Samaras S, Diamantidou E, Ataloglou D, et al. Deep learning on multi sensor data for counter UAV applications: a systematic review[J]. Sensors, 2019, 19(22): E4837.

[3] Jahangir M, Baker C J, Oswald G A. Doppler characteristics of micro-drones with L-Band multibeam staring radar[C]//2017 IEEE Radar Conference (RadarConf), May 8-12, 2017, Seattle, WA, USA. New York: IEEE Press, 2017: 1052-1057.

[4] Molchanov P, Harmanny R I A, de Wit J J M, et al. Classification of small UAVs and birds by micro-Doppler signatures[J]. International Journal of Microwave and Wireless Technologies, 2014, 6(3/4): 435-444.

[5] Beason R C, Nohara T J, Weber P. Beware the boojum: caveats and strengths of avian radar[J]. Human-Wildlife Interactions, 2013, 7(1): 16-46.

[6] Chen W S, Liu J, Chen X L, et al. Non-cooperative UAV target recognition in low-altitude airspace based on motion model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(4): 687-694.
陈唯实, 刘佳, 陈小龙, 等. 基于运动模型的低空非合作无人机目标识别[J]. 北京航空航天大学学报, 2019, 45(4): 687-694.

[7] Torvik B, Olsen K E, Griffiths H. Classification of

birds and UAVs based on radar polarimetry[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(9): 1305-1309.

[8] Chen W S, Huang Y F, Chen X L, et al. Review on developments and applications of airport avian radar [J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(7): 1-22.
陈唯实, 黄毅峰, 陈小龙, 等. 机场探鸟雷达技术发展与应用综述[J]. 航空学报, 2021, 42(7): 1-22.

[9] Cai X Y, Sun J F, Lu Z Y, et al. Distortion compensation technology of coherent frequency modulation continuous wave lidar[J]. Chinese Journal of Lasers, 2020, 47(9): 0910003.
蔡新雨, 孙建锋, 卢智勇, 等. 相干调频连续波激光雷达畸变补偿技术研究[J]. 中国激光, 2020, 47(9): 0910003.

[10] Zu S, Hu P P, Pan Q. Extraction method of artificial landmark center based on lidar echo intensity[J]. Chinese Journal of Lasers, 2020, 47(8): 0810001.
祖爽, 胡攀攀, 潘奇. 基于激光雷达回波强度的人工路标中心提取方法[J]. 中国激光, 2020, 47(8): 0810001.

[11] Mohajerin N, Histon J, Dizaji R, et al. Feature extraction and radar track classification for detecting UAVs in civillian airspace[C]//2014 IEEE Radar Conference, May 19-23, 2014, Cincinnati, OH, USA. New York: IEEE Press, 2014: 0674-0679.

[12] Chen W S, Liu J, Li J. Classification of UAV and bird target in low-altitude airspace with surveillance radar data[J]. The Aeronautical Journal, 2019, 123(1260): 191-211.

[13] Christ M, Braun N, Neuffer J, et al. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh-A Python package)[J]. Neurocomputing, 2018, 307: 72-77.

[14] Zhang H, Zuo X L, Huang Y. Feature selection based on the correlation of sparse coefficient vectors with application to SAR target recognition[J]. Laser & Optoelectronics Progress, 2020, 57(14): 141029.
张虹, 左鑫兰, 黄瑶. 基于稀疏表示系数相关性的特征选择及SAR目标识别方法[J]. 激光与光电子学进展, 2020, 57(14): 141029.

[15] Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency[J]. The Annals of Statistics, 2001, 29(4): 1165-1188.

[16] Christ M, Kempa-Liehr A, Feindt M. Distributed and parallel time series feature extraction for industrial big data applications[EB/OL]. (2016-10-25)[2020-01-05]. https://arxiv.org/abs/1610.07717.

[17] Liu G, Li L L, Zhang L L, et al. Sensor faults classification for SHM systems using deep learning-based method with Tsfresh features[J]. Smart Materials and Structures, 2020, 29(7): 075005.

[18] Jin C, Kong X G, Chang J T, et al. Internal crack detection of castings: a study based on relief algorithm and Adaboost-SVM[J]. The International Journal of Advanced Manufacturing Technology, 2020, 108(9/10): 3313-3322.

[19] Robnik-Šikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF[J]. Machine Learning, 2003, 53(1/2): 23-69.

[20] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: a new learning scheme of feedforward neural networks[C]//2004 IEEE International Joint Conference on Neural Networks, July 25-29, 2004, Budapest, Hungary. New York: IEEE Press, 2004: 985-990.

[21] Liang N Y, Huang G B, Saratchandran P, et al. A fast and accurate online sequential learning algorithm for feedforward networks[J]. IEEE Transactions on Neural Networks, 2006, 17(6): 1411-1423.

[22] Nagy M, Akos Z, Biro D, et al. Hierarchical group dynamics in pigeon flocks[J]. Nature, 2010, 464(7290): 890-893.