

Relief 和 SVMRFE 在高超声速 进气道不起动预测中的应用

刘欢¹, 黄俊¹, 张勇², 刘志勤¹, 王耀彬¹

(1. 西南科技大学 计算机科学与技术学院, 四川 绵阳 621010;

2. 中国空气动力研究与发展中心 吸气式高超声速技术研究中心, 四川 绵阳 621010)

摘要: 高超声速进气道不起动预测研究中主要包括确定压力传感器位置和建立起/不起动分类面, 属于机器学习特征选择问题和分类问题, 而常用特征选择算法(基于支持向量机的递归特征消除 SVM-RFE)单一并且耗时较长; 为解决该问题寻找较优的特征选择算法, 建立一个高超声速二元进气道/隔离段模型, 通过数值模拟获得内流道上表面压力数据样本; 利用 Relief 和 SVMRFE 组合式算法 Relief-Corre 方法, Relief-SVMRFE 方法, Relief-PSO-SVMRFE 方法进行特征选择; 支持向量机 SVM 训练分类面; 最后得出 Relief-SVMRFE 方法性能最优, 运行效率比 SVMRFE 提高了约 3 倍, 准确率比其他基于 Relief 组合方法高; 获得最优特征的分类面具有较高的泛化性与鲁棒性, 证明该分类面的有效性。

关键词: 高超声速进气道; 数值模拟; 特征选择; Relief; 支持向量机的递归特征消除

Application of Relief and SVMRFE in Predicting Unstart of Hypersonic Inlet

Liu Huan¹, Huang Jun¹, Zhang Yong², Liu Zhiqin¹, Wang Yaobin¹

(1. College of Computer Science and Technology, Southwest University of Science and Technology,

Mianyang 621000, China; 2. Air-breathing Hypersonic Technology Research Center,

China Aerodynamics Research and Development Center, Mianyang 621000, China)

Abstract: The predicting unstart of hypersonic inlet consists of calculating positions of pressure sensors and establishing classification of inlet start/unstart, the feature selection algorithm and the classification algorithm was taken to solve this two problems, the common algorithm (SVMRFE) is time intensive on computation. A 2D hypersonic inlet / isolator model were simulated to generate wall static pressures data set. Hybrid feature selection algorithm based on SVMRFE algorithm and Relief algorithm were used to select optimal pressure points, which were called as Relief-Corre algorithm, Relief-SVMRFE algorithm, Relief-PSO-SVMRFE algorithm. The support vector machine (SVM) algorithm was used to train the classification plane. Finally, the performance of Relief-SVMRFE algorithm is proved best, since its operation efficiency is three times higher than SVMRFE and it has higher accuracy than other algorithms based on Relief. The hyperplane with strong robust performance and generalization performance conforms to the actual physical law, so the result shows that the criterion is valid.

Keywords: hypersonic inlet; numerical simulation; feature selection; Relief; SVMRFE

0 引言

高超声速进气道作为飞行器发动机关键气动部件^[1], 若发生进气道不起动现象直接影响着推进系统整体效能的发挥, 发动机不能产生推力。若能对进气道工作状态进行及时的判断与预测, 对进气道的再起控制 and 发动机安全运行起关键作用, 因此在进气道的研究控制中, 进气道起/不起动判断与预警

已经成为一个重要的研究工作^[2-3]。通常使用模式分类来解决进气道起/不起动判断与预警^[4], 数值计算产生进气道起/不起动样本数据作为训练集, 采用计算结果中多个压力点和压力值作为特征属性和特征属性值。但在实际应用中, 对压力点的要求是最少最灵敏的压力点, 因此在对进气道起/不起动判断预警中, 特征选择算法尤为重要, 一般要求选择出 2 个最优压力点, 再在二维空间中训练分类准则, 对进气道起/不起动现象进行准确判别。目前可用的特征选择算法仅有 SVMRFE, 分类算法有 FLD 和 SVM^[4-5], 而 SVMRFE 算法只能每次删除一个特征, 这对于维度较高的数据集, 大大增加了计算负担。

特征选择算法在机器学习中常用到的一类算法, 可以分为 wrapper 和 filter 两大类^[6-7], wrapper 类算法是利用后续学习算法的训练准确率来评估特征子集, 因此评估结果和后续算法性能偏差较小, 但由于评估每个特征子集时都要运行学习算法, 计算量很大, 大的数据集耗时会很长; filter 类算法和后续学习算法无关, 一般直接利用所有训练数据的统计性能评

收稿日期: 2017-08-04; 修回日期: 2017-09-15。

基金项目: 国家自然科学基金面上项目(61672438); 西南科技大学研究生创新基金(16ycx048)。

作者简介: 刘欢(1992-), 女, 硕士研究生, 主要从事机器学习、数值模拟方向的研究。

张勇(1969-), 男, 博士, 副研究员, 主要从事吸气式高超声速推进技术方向的研究。

刘志勤(1962-), 女, 硕士, 教授, 主要从事高性能计算、数值模拟、计算机网络方向的研究。

估特征，如相关性、一致性、信息增益等，因此 filter 类算法具有速度快的优点，但评估结果和后续算法的性能有较大的偏差。而这两类算法中最具代表性的算法是 SVMRFE 和 Relief，SVMRFE 算法准确性能有一定的保证，但是运行时间较长，而 Relief 耗时少，但是 Relief 的缺点是不能去掉冗余特征。

目前基于 Relief 的改进算法也较为常见，使用 Relief 算法快速去掉一大部分无关特征，再进一步有效地去掉冗余特征^[8-10]。本文在数值计算结果中，内流道上表面会多达几百个压力点，数据维度较高，所以 SVMRFE 算法选择特征耗时较长，本文结合 Relief 算法和 SVMRFE 算法进行特征选择，寻找新的处理进气道压力数据的方法，提高运行效率，并对结果准确率进行分析。

1 数据准备

本文的数据集是通过数值计算而获得的，包括进气道启动/不起动两类样本，压力点为属性，对应压力值为属性值。

1.1 数值计算

1.1.1 物理模型和计算网格

给出高超声速进气道/隔离段模型见图 1，是参考文献[11]进行建模。图中点 A 为鼻锥前缘，点 B 为二级楔面起点，点 C 为进气道外罩唇口，点 D 为进气道喉道部位，点 E 为隔离段出口位置。 L_1, L_2, L_3 为点 A 到点 B, C, D 的位置， L_4 为隔离段长度， H_1, H_2 分别为喉道和出口的高度。图中具体数据如表 1 所示。

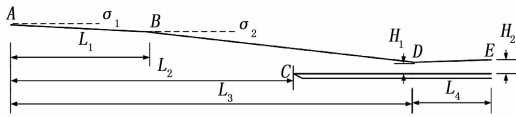


图 1 进气道/隔离段几何模型

表 1 几何模型数据

L_1/m	L_2/m	L_3/m	L_4/m
0.6300	1.3015	1.8475	0.3620
H_1/m	H_2/m	$\sigma_1/(^\circ)$	$\sigma_2/(^\circ)$
0.044	0.0600	2.7	6.2

计算网格如图 2，网格划分软件为 Gridgen，网格类型为结构网格，其中壁面向第一层网格间距取值为 0.0005 mm， $y^+ < 1$ 。二维网格总数约为 12.7 万。

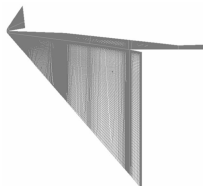


图 2 计算网格

1.1.2 数值计算方法

采用 fluent14.5 软件^[12]进行数值模拟，其中使用的控制方程为 RANS 方程，本文采用的湍流模型为 SST $k-\omega$ ，无粘通量离散格式为 AUSM，热完全气体模型，采用定常模拟方法，数值计算的收敛准则为连续方程的标准化残差小于 10^{-3} 。

参考飞行试验中 X-51A 飞行器飞行条件，设置边界条件如表 2，设置了 3 组不同的动压条件，每组动压条件下分别对应的马赫数分别为 4.5, 5.0, 5.5，攻角大小分别为 $[0, 2^\circ, 4^\circ, 6^\circ, 8^\circ]$ 。计算自由流压力条件下的流场，此时喉道质量加权平均压力为 P ，背压比定义为隔离段出口条件中的压力 P_0 与压力 P 之比，即背压比 $r = P_0/P$ 。按 $[2, 3, 4, 5 \dots]$ 逐步增大背压比，计算流场直到不起动，随后在相邻的启动/不起动背压比之间，采用二分取值法细分背压比来计算流场。

表 2 自由来流条件

动压 Q/atm	马赫 数 M	高度 H/m	静压 P/Pa	静温 T/K	总温 T_∞/K	攻角 A/deg
0.5	4.5	22711	3574	219	1108	0,2,4,6,8
	5.0	24068	2895	221	1324	0,2,4,6,8
	5.5	25302	2393	222	1565	0,2,4,6,8
0.75	4.5	20124	5361	217	1095	0,2,4,6,8
	5.0	21465	4343	218	1309	0,2,4,6,8
	5.5	22685	3589	219	1546	0,2,4,6,8
1.0	4.5	18301	7148	217	1094	0,2,4,6,8
	5.0	19636	5790	217	1300	0,2,4,6,8
	5.5	20846	4785	218	1533	0,2,4,6,8

1.1.3 数值计算结果数据分析

变化的背压引起进气道是否启动主要体现在隔离段激波串前沿位置的变化，变化的背压，使得激波串几乎完整地进行简单滑动和平移。如果背压过大，会出现激波串前沿越过喉道位置处的情况，本文将此现象定义为进气道不起动。

在激波串移动的过程中，造成隔离段内上表面压力增大，当进气道不起动时，隔离段内整个上表面压力值增大。由于楔面转角处出现膨胀波，导致压力下降，所以图 3 中进气道启动状态下，喉道处压力会突然减小。

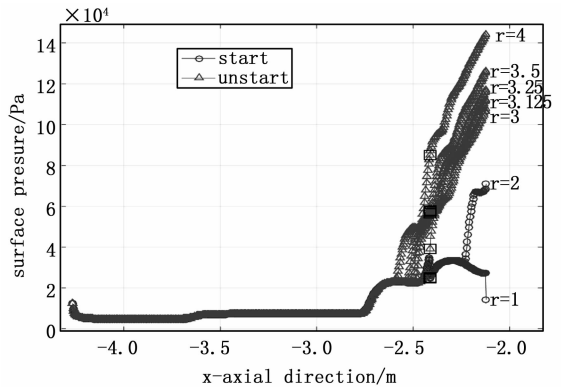


图 3 条件 $Q = 0.5, M = 4.5, A = 0$ 下上表面压力分布

由此，本文通过上表面压力值变化趋势来判别进气道启动/不起动。最终从数值计算的结果中共获得 321 条样本数据，进气道启动样本数据共 161 条、不起动样本数据 160 条，每条数据共包含 576 个压力数据值。

1.2 数据集描述

定义该数据集为某二元进气道/隔离段上表面压力数据集，数据集中包括 321 条进气道启动/不起动样本，576 个属性（即指 CFD 网格点上 576 个压力分布点），样本类别为启动/不起动，属性值为数值形式。每个样本表示成一个序列向量 X_i ， i 的取值为 1~321。

$$\mathbf{X}_i = (\text{label}: x_1 x_2 \cdots x_j \cdots x_{576})$$

其中: label 是样本类别标签, $\text{label} \in \{+1, -1\}$, +1 表示进气道起动机样本, -1 表示进气道不起动机样本。

2 特征选择算法

2.1 基本算法简介

Relief 算法^[13]设计了一个“相关统计量”来度量特征的重要性, 主要是通过特征间的距离来计算相关统计量, 认为特征的分类能力越强, 越能聚集同类样本, 疏远异类样本。有样本集 $\{(x_1, y_1), (x_1, y_1), \dots, (x_n, y_n)\}$, 对于一个样本 x_i , 在同类中最邻近的样本 $x_{i,nb}$, 异类中的邻近样本 $x_{i,mb}$, 相关统计量对应于属性 j 的分类为:

$$\sigma^2 = \sum_i |x_i^j - x_{i,nb}^j|^2 + |x_i^j - x_{i,mb}^j|^2 \quad (1)$$

多次随机抽取样本 x_i , 对基于这些样本得到的估计结果进行平均, 就得到各个特征的相关统计分量, 分量值越大, 则对应特征越好。

SVMREF 算法^[13-15]使用递归特征去除 (Recursive Feature Elimination, REF) 算法来保证在特征排序的过程中优先保留优化特征子集。在对特征进行排序时, 一般可以使用支持向量机 (SVM) 判别函数中的权值来计算特征排序准则。每次迭代消去排序准则最小的特征, 直到所有特征被删除, 越是重要的特征越是晚被删除。

SVM 分类器的权值 w 计算方法为公式 (1), m 是支持向量的个数, α_i 是非零拉格朗日乘子, x_i 为训练样本, y_i 为对应类别标签。

$$w = \sum_{i=1}^m \alpha_i y_i x_i \quad (2)$$

常见的排序准则^[16]有如下两种, 公式 (2) 为线性核函数 SVM 分类器的排序准则, w_h 为消去第 h 个特征后的权重; 非线性核函数 SVM 分类器对应的排序为公式 (3), 其中 α 是所有非零拉格朗日乘子, \mathbf{H} 是一个矩阵, 其元素为 $y_i y_j K(x_i, x_j)$, $\mathbf{H}(-h)$ 为消去第 h 个特征后的矩阵, K 为核函数。

$$DJ(h) = (w_h)^2 \quad (3)$$

$$DJ(h) = \frac{1}{2} \alpha^T H \alpha - \frac{1}{2} \alpha^T H(-h) \alpha \quad (4)$$

粒子群优化算法 (PSO: Particle Swarm Optimization) 是通过群体中个体之间的协作和信息共享来寻找最优解。初始化一群随机粒子 (随机解), 然后再每一次的迭代中, 粒子通过跟踪两个极值 (个体最优和种群最优) 来更新自己。粒子速度与位置更新公式^[17]如下:

$$v_i(t+1) = w \cdot v_i(t) + c_1 \text{rand}() \cdot (p_j(t) - x_i(t)) + c_2 \text{rand}() \cdot (p_g(t) - x_i(t)) \quad (5)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (6)$$

其中: v_i 是粒子速度, t 是迭代次数, w 是惯性权重, $\text{rand}()$ 是 (0, 1) 间的随机数, P_j 是个体最优, P_g 是种群最优, c_1 和 c_2 是加速系数。

2.2 基于 Relief 组合式特征选择算法

2.2.1 Relief-Corre 方法

Relief-Corre 方法首先运行 Relief 算法, 得到每个特征的权重 w_i , 将取值小于设定无关阈值的特征删除, 在采用相关系数 (Correlation coefficient) 评估两两特征间的冗余度。若相关度小于设定的阈值, 就删除权重较小的一个属性。具体算法流程如下^[8]:

Step1: 输入数据集 D , 特征集 $S = \{s_i, i=1, 2, \dots, n\}$, Relief 无关阈值 σ , 相关度阈值 α 。

Step2: 运行 Relief, 得到降序排列的特征权重 $W = \{w_i, i=1, 2, \dots, n\}$ 。

Step3: 删除 $w_i > \sigma$ 的特征, 得到特征集 $S = \{s_i, i=1, 2, \dots, m\}$ 和数据集 D 。

Step4: for $i=1: m-1$

For $j=i+1: m$

计算两个特征间的相关度 $\text{CORR}(i, j)$

If $(\text{CORR}(i, j) \geq \alpha)$, 删除特征 s_j

Step5: 得到最优的特征子集。

2.2.2 Relief-SVMRFE 方法

Relief-SVMRFE 方法仍然在第一步采用 Relief 算法去掉无关特征, 在第二步中就采用序列后向搜索 (Sequential Backward Search, SBS) 方法 SVMRFE 算法去掉冗余特征。具体流程如下:

Step1: 输入数据集 D , 特征集 $S = \{s_i, i=1, 2, \dots, n\}$, relief 无关阈值 σ 。

Step2: 运行 Relief, 得到降序排列的特征权重 $W = \{w_i, i=1, 2, \dots, n\}$, 删除 $w_i > \sigma$ 的特征, 得到特征集 $S = \{s_i, i=1, 2, \dots, m\}$ 和数据集 D 。

Step3: 运行 SVMRFE 算法。

Step4: 得到最优的特征子集。

2.2.3 特征选择算法 Relief-PSO-SVMRFE

第一阶段使用 Relief 过滤掉权重小于无关阈值的特征, 在第二阶段一分类器的准确率作为自己的评估标准, 采用 PSO 逐步寻找最优特征子集, 删除冗余特征。算法描述如下:

Step1: 输入数据集 D , 特征集 $S = \{s_i, i=1, 2, \dots, n\}$, relief 无关阈值 σ 。

Step2: Relief 剔除无关特征。

Step3: 随机产生 N 个长度为 P 的初始粒子种群, 即特征子集。

Step4: 利用支持向量机的精度来评估当前粒子的优劣性。

Step5: 更新局部最优和全局最优, 更新每个粒子的位置。

Step6: 转到 Step4, 直到达到最大迭代数。

Step7: 输出最优特征子集。

3 实验与分析

3.1 算法结果分析

基于 MATLAB 环境下调用 `fitsvm` 函数完成上述原始 SVMRFE 算法, 其中核函数选择线性核函数, 惩罚系数 $C = \infty$, 优化方法选择序列最小优化 SMO, 并运行 SVMRFE 算法及第二节中基于 Relief 算法进行特征选择, 并对分类准确率和耗时记录分析。

在 Relief 阶段每次随机取值 300 次, 共计算 3 次, 统计权重 W 的平均值, 无关阈值设置为 0.5, 该阶段能删除约 69.5% 的特征。Relief-Corre 方法的相关阈值设置为 0.8 和 0.9, Relief-PSO-SVMRFE 中 $c_1 = c_2 = 2$, $w = 0.9$, 种群大小为 50, 迭代次数为 200。首先在整个数据集上测试, 重复 5 次统计最优特征子集和平均耗时 (由于 PSO 随机产生种群的原因, 5 次结果中出现 3 个最优特征子集)。对特征子集进行 5 次 5 折交叉验证, 给出 SVM 和朴素贝叶斯的分类错误率。

表 3 各算法特征选择结果耗时与 SVM 和 NBC 分类误差

序号	算法	最优特征子集	平均耗时, s	SVM 平均错误率	NBC 平均错误率
1	Relief-Corre(0.8)	[1,506]	2.6846	0.0031	0.16142
2	Relief-Corre(0.9)	[1,413,506]	2.6458	0.0031	0.17326
3	Relief-SVMRFE	[406, 505,506]	4.0782	0.00168	0.17078
4	Relief-PSO-SVMRFE	[405,506]	130.7324	0.00062	0.17816
		[406,506]		0.00124	0.17704
		[408,506]		0.00062	0.18056
5	SVMRFE	[406,505,506]	16.775	0.00168	0.17078

表 3 中显示, 所有算法都能选择出 2 个或 3 个特征, Relief-Corre 方法耗时最少, Relief-PSO-SVMRFE 耗时相当长, 而 Relief-SVMRFE 方法对于 SVMRFE 在效率上提高了约 3.1 倍。朴素贝叶斯分类时, 每个特征组对应的误差并不大, 而在 SVM 分类时, Relief-Corre 方法对应分类误差明显较大, 其他 3 个算法对应分类误差区别不明显, 但 Relief-PSO-SVMRFE 方法不能一次就选择出唯一准确的结果。通过特征选择过程耗时与特征子集分类误差判断, Relief-SVMRFE 方法的性能明显优于其他方法。

3.2 结果应用分析

在进气道起动预测中, 一般会选择 2 个最优、最敏感的压力传感器放置位置, 即选择两个特征。但 3.1 节中通过 Relief-SVMRFE 选择出的 [406, 505, 506] 特征并不能直接运用于实际应用中, 因此对 [406, 505] 和 [406, 506] 两组特征进行讨论, 删除冗余特征; 由于实际应用中压力传感器测到的数据存在误差, 再对得到的分类面的鲁棒性进行分析。

通过五次五折对 [406, 505] 和 [406, 506] 特征 SVM 分类面判别误差进行计算, 分别为 0.0031 和 0.00124, 特征 505, 506 对应 Relief 权重分别为 16.1077 和 15.9996, 所以 506 特征是优于 505 特征的, 因此删除 505 特征。令第 406 与 506 个压力测试点分别为 P_1, P_2 , 其中 P_1 是进气道内某个位置上的压力测试点, P_2 为靠近喉道位置某个压力测试点。基于 P_1 和 P_2 两点训练分类面, 分类面为 $p_2 - 4.7058 * p_1 + 3388.6156 = 0$, 两条边界分别为 $p_2 - 4.7058 * p_1 + 9763.7793 = 0$ (起动) 和 $p_2 - 4.7058 * p_1 + 2986.5481 = 0$ (不起动)。数据分类情况如图 4 所示。

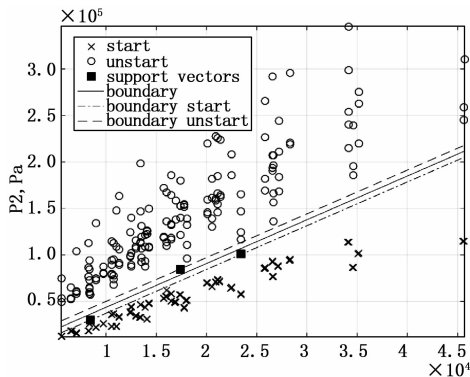


图 4 SVM 训练的分类结果

由于压力测试点位置的进度已经达到 0.1 mm, 在实际应用中, 不能保证压力传感器测试到的压力值就是 P_1, P_2 点的值, 因此应检验分类面的鲁棒性能。结合点的位置坐标与 Re-

lief 权重因素, 选择相似的点进行考察, 这些点对应的压力值代表在实际应用中压力传感器可能会测到的压力值, 获得两组考察压力点 405~407 和 503~509, Relief 权重和压力点坐标位置如表 4。一共可以获得 21 个组合, 用已得到的分类面对这些数据进行分类, 表 5 中所有的分类误差均不大于 0.001。尽管压力传感器没有测到 P_1 和 P_2 点的压力, 实际测到的压力值具有一定的偏差, 该分类面依然可以较准确的将进气道起动/不起动分类开。

表 4 两组压力点 Relief 权重和坐标位置

特征点	405	406	407	503	504
Relief 权重	0.7617	0.9319	1.1060	15.5074	15.6610
坐标位置	-2.8134	-2.8078	-2.8019	-2.4124	-2.4123
特征点	505	506	507	508	509
Relief 权重	15.9351	16.0507	15.9692	15.6776	15.2243
坐标位置	-2.4122	-2.4120	-2.4118	-2.4117	-2.4114

表 5 特征子集对应分类误差

特征组合	[405,503]	[405,504]	[405,505]	[405,506]
分类误差	0.0062	0.0031	0.0031	0.0031
特征组合	[406,503]	[406,504]	[406,505]	[406,506]
分类误差	0.0062	0.0031	0.0031	0
特征组合	[407,503]	[407,504]	[407,505]	[407,506]
分类误差	0.0093	0.0031	0.0031	0
特征组合	[405,507]	[405,508]	[405,509]	
分类误差	0.0031	0.0062	0.0093	
特征组合	[406,507]	[406,508]	[406,509]	
分类误差	0.0031	0.0062	0.0062	
特征组合	[407,507]	[407,508]	[407,509]	
分类误差	0	0.0031	0.0062	

4 结果

本文主要解决高超声速进气道问题, 使用 Relief 算法和 SVMRFE 算法来保证特征选择的效率与准确性。试验中, 特征选择使用到了 Relief-Corre 方法、Relief-SVMRFE 方法、算法 Relief-PSO-SVMRFE 和 SVMRFE 算法。Relief-SVMRFE 方法和常用的分析进气道起动方法 (SVMRFE 算法) 在获得同样结果的基础上, 效率上有了很大的提高, 而其他的基于 relief 算法在该方面应用并没有明显的优势; 基于文中最优特征子集 [406, 506] 得到分类面具有较高分类性能和鲁棒性能, 有实际应用的意义。最后特征选择结果中特征数量是否可以进一步减少, 将在接下来的研究中继续展开讨论。

参考文献:

- [1] 吴颖川, 贺元元, 贺伟, 等. 吸气式高超声速飞行器机体推进一体化技术研究进展 [J]. 航空学报, 2015, 36 (1): 245-260.
- [2] Heiser W, Pratt D, Daley D, et al. Hypersonic airbreathing propulsion [M]. American Institute of Aeronautics and Astronautics, 2010.
- [3] Cox C, Lewis C, Pap R, et al. Prediction of unstart phenomena in hypersonic aircraft [M]. International Aerospace Planes and Hypersonics Technologies, 1995.
- [4] Yu D, Chang J, Bao W, et al. Optimal classification criterions of hypersonic inlet start/unstart [J]. Journal of Propulsion & Power, 2007, 23 (2): 310-316.

(下转第 190 页)