

基于粗糙集理论的螺旋桨设计规则提取

胡 豹¹, 高永卫¹, 魏斌斌²

(1. 西北工业大学 航空学院, 西安 710072;

2. 西安交通大学 航天航空学院, 西安 710049)

摘 要: 不同于由专家通过理论推导或经验得到设计规则的传统方法, 提出了数据驱动的螺旋桨设计规则提取方法。使用粗糙集理论对螺旋桨优化设计过程中产生的大量方案数据进行分析, 发展了适当的属性表示方法和属性离散方法。结果表明: 在对两类不同参数螺旋桨型号的研究中, 均提取出了“当各剖面诱导速度的差异越小时, 螺旋桨效率越高”的设计规则, 这一规则与 Betz 条件相当, 验证了方法的有效性。基于数据驱动的螺旋桨设计规则提取方法可为后续三维设计规则的发现和优化设计研究提供技术支持, 该思想也可供其他设计领域参考。

关 键 词: 规则提取; 螺旋桨设计; 数据驱动; 粗糙集理论; Betz 条件

中图分类号: V211.44

文献标志码: A

Propeller design rule extraction based on rough set theory

HU Bao¹, GAO Yongwei¹, WEI Binbin²

(1. School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China;

2. School of Aerospace Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Different from the traditional method of obtaining design rules by experts through theoretical derivation or experience, a data-driven propeller design rule extraction method was proposed. The rough set theory was used to analyze a large number of scheme data generated in the process of propeller optimization design. An appropriate attribute representation method and an attribute discrete method were developed. Result showed that, in the study of two types of propeller models with different parameters, the design rule that the smaller the difference of the induced velocity of each section of the propeller, the higher the propeller efficiency was extracted. This rule was equivalent to the Betz condition, which verified the effectiveness of the method. The data-driven propeller design rule extraction method can provide technical support for the subsequent discovery of three dimensional design rules and optimal design research. This idea can also be used as a reference for other design fields.

Keywords: rule extraction; propeller design; data driven; rough set theory; Betz condition

螺旋桨的气动效率对飞机性能有着至关重要的影响。为了设计出高效率的螺旋桨, 需要研究完善设计规则。

目前已有的螺旋桨设计规则主要由空气动力学领域的专家在结合物理定律和一些合理假设的

情况下, 通过数学推导得到, 如动量理论、片条理论、Betz 条件、涡流理论等^[1]。其本质是一种典型的专家系统, 这种设计规则发现方法依赖于专家的设计经验和专业知识。但由于螺旋桨三维设计的复杂性和设计要求的不断提高, 如今从理论

收稿日期: 2021-11-09

作者简介: 胡豹(1995—), 男, 博士生, 主要从事螺旋桨智能设计研究。E-mail: hu_bao@mail.nwpu.edu.cn

通信作者: 高永卫(1968—), 男, 教授、博士生导师, 博士, 主要从事设计空气动力学、实验空气动力学和气动噪声研究。

E-mail: gyw630@nwpu.edu.cn

引用格式: 胡豹, 高永卫, 魏斌斌. 基于粗糙集理论的螺旋桨设计规则提取[J]. 航空动力学报, 2023, 38(8): 1901-1908. HU Bao, GAO Yongwei, WEI Binbin. Propeller design rule extraction based on rough set theory[J]. Journal of Aerospace Power, 2023, 38(8): 1901-1908.

出发,通过数学推导得到更多的螺旋桨设计规则已经变得殊为困难。

随着计算科学的发展,通过数据挖掘得到新知识逐渐成为一种新的研究范式^[2-3]。数据挖掘是在大量数据中发现隐含的、先前未知的并具有潜在价值的知识的过程,它主要基于人工智能、机器学习、模式识别、统计学、数据库、以及可视化等技术。由于数据挖掘技术可以对数据中隐含的知识/规则进行分析,得到复杂参数之间的关联关系及影响规律,因而在多个领域中受到重视。Simpson 等^[4]指出数据挖掘可以用于研究设计变量与设计目标之间的影响机制,实现主要因素聚类,为类似的设计提供信息,提高设计效率。Jeong 等^[5-6]使用自组织映射(self-organizing map, SOM)和方差分析(analysis of variance, ANOVA)方法对翼型设计知识进行了挖掘,通过消除对设计问题影响很小的设计变量来简化设计空间,得到翼型设计的敏感变量,确定了最优解具有良好性能的原因。Chiba 等^[7-8]使用 ANOVA、SOM 和粗糙集方法对机翼设计进行了数据挖掘研究提取了设计空间的关键特征,发现了多学科设计优化过程中丢失的信息。汪伟等^[9]用 K-Means 聚类对仿真数据进行离散化处理,使用粗糙集属性重要度算法进行属性约简,采用决策树算法构建叶片优化设计知识决策树,挖掘了叶片优化设计变量弯扭、周向积迭线的控制点偏移量与目标函数总压损失系数之间隐含的设计规则。郭振东等^[10]基于显著变量识别、ANOVA 及 SOM 这 3 种方法对叶栅气动性能的影响机制进行分析,探索最优设计气动性能提高的本质原因。刘卓等^[11]将粗糙集理论中,基于正域重要度和基于条件信息熵的两种权重确定方法相结合,提出在机翼方案设计阶段进行翼型选择时,两断点方法是比较合适的属性离散化方法。刘深深等^[12]使用 ANOVA、等度量映射、决策树以及 SOM 这 4 种算法对飞行器气动布局优化设计过程中产生的数据进行分析,凝练出了适用于高升阻比滑翔飞行器的设计知识。庞梦洋等^[13]基于决策树和粗糙集进行航空发动机故障诊断,增强了提取规则的可解释性,提升了诊断的准确率。目前,数据挖掘在翼型、叶片、机翼、飞行器布局选型、航空发动机故障诊断中均得到了初步应用,针对螺旋桨的特殊性开展设计规则/知识发现的研究还鲜见公开报道。

本文针对两型无人机的使用要求,分别随机

生成大量设计方案进行分析,使用粗糙集理论挖掘数据中隐藏的螺旋桨设计规则。针对螺旋桨方案数据的特点,使用人工离散方法对决策属性进行离散化,使用动态层次聚类方法对条件属性进行离散化。经过属性约简,得到了与 Betz 条件相当的设计准则,验证了粗糙集理论在螺旋设计规则挖掘中的有效性,证明了基于数据驱动进行螺旋桨设计规则提取这一思路的可行性。

1 粗糙集理论简介

20 世纪 80 年代初,波兰数学家 Pawlak^[14-15]提出了粗糙集(rough set)理论,用于分析模糊和不确定问题,其基本思想是通过案例库的分类归纳出规则,是一种典型的数据挖掘手段。粗糙集理论的优势在于处理数据的过程中不需要先验知识,仅对现有数据样本进行分析,不掺杂人为主观因素,与模糊理论等其他需要依靠专家系统的方法相比,粗糙集理论具有高度的客观性,特别适用于处理缺乏系统理论支持的问题。

本节简要介绍粗糙集的基本概念、理论及相关算法,主要有决策信息表、属性约简、连续属性离散化、确定性因子及确定性规则^[16]。

1.1 决策信息表

要利用粗糙集方法,首先需建立决策信息表。设 (U, A, V, f) 是一个信息系统,其中 U 为对象集,也称论域, A 为属性集, V 是各属性对应取值的集合,信息函数 f 为 U 与 A 的关系集,是 $U \times A \rightarrow V$ 的映射,它为每个对象的所有属性进行赋值。当属性集 A 由条件属性集 C 和决策属性集 D 组成时,则称此信息系统为决策信息表。

1.2 属性约简

信息系统中各个属性并不是同等重要的,属性约简就是为了消除信息系统中的冗余属性。一般通过启发式算法实现属性约简,即先求决策信息表的相对核,然后根据属性重要度逐渐添加属性,最终得到相对约简。

定义(相对核) 给定信息系统 $K = (U, S)$ 和其中的两个等价关系族 $P, Q \subseteq S$, 对于任意的 $R \in P$, 若 R 满足

$$Z_{\text{pos,IND}}(P - \{R\})(I_{\text{IND}}(Q)) \neq Z_{\text{pos,IND}}(P)(I_{\text{IND}}(Q)) \quad (1)$$

则称 R 为 P 中 Q 必要的,将 P 中所有 Q 必要的知识组成的集合定义为 P 的 Q 核。式(1)中 $Z_{\text{pos}}(\cdot)$ 表示求正域, $I_{\text{IND}}(\cdot)$ 表示求不可分辨(IND)关系。

定义(相对约简) 给定信息系统 $K = (U, S)$ 和其中的两个等价关系族 $P, Q \subseteq S$, 对于任意的 $G \subseteq P$, 当 G 符合如下条件:

- 1) G 相对 Q 独立
- 2) $Z_{\text{pos}, G}(Q) = Z_{\text{pos}, P}(Q)$

则称 G 为 P 相对于 Q 的一个约简。

1.3 连续属性离散化

动态层次聚类的连续属性离散化方法是对条件属性进行离散化的常用方法, 在此对其原理进行简要介绍。

定义(不相容度) 令 $T = (U, C \cup D, V, f)$ 为一个决策信息表, 其中 U 指论域, $C = \{c_1, c_2, \dots, c_m\}$ 为条件属性集。设第 i 个条件属性 c_i 的不相容度为 α_i , 则按式(2)对 α_i 进行定义:

$$\alpha_i = \frac{N_{\text{card}}(\bar{C}_i)}{N_{\text{card}}(U)} \quad (2)$$

式中 $\bar{C}_i$ 表示仅考虑条件属性 c_i 时, 决策信息表中不相容对象的集合, $N_{\text{card}}(\cdot)$ 表示集合中的元素个数。

因为决策信息表中, 各属性之间出现不相容的情况在统计上基本相互独立, 所以整个决策表的不相容度 α_T 可以记作 $\alpha_T = \prod_{i=1}^m \alpha_i$ 。如果进一步把 α_i 看作近似相等, 记为 α , 那么 $\alpha_T = \alpha^m$, 于是 α 的估算可以记作 $\hat{\alpha} \approx \sqrt[m]{\alpha_T}$ 。

原始的连续决策信息表是相容的, 因此在完成属性离散化后, 离散的决策信息表理论上也应该是相容的, 即 $\alpha_T = 0$ 。但在实际应用中, 当数据量较大时, 一般很难实现 $\alpha_T = 0$, 故取 α_T 为一个足够小的数即可, 如 $\alpha_T = 0.0001$ 。

此外, 在决策表中不同属性之间的相容度也不尽相同。因此在实际应用中, 保证所有属性的不相容度 α_i 能够在 $\hat{\alpha}$ 的某一范围内即可。即保证式(3)成立:

$$|\alpha_i - \hat{\alpha}| \leq \beta \quad (3)$$

其中 β 为设定的误差范围。

离散化的具体流程为

- 1) 设定 α_T 、 β 和 δ 的值。

- 2) 求解 $\hat{\alpha}$ 。

- 3) 对于 $i=1, 2, \dots, m$, 重复以下步骤:

(1) 按照类间距 δ_i , 通过层次聚类方法得到条件属性 c_i 关于论域 U 的离散结果。

- (2) 求属性 c_i 的不相容度 α_i 。

- (3) 判断 $|\alpha_i - \hat{\alpha}| \leq \beta$ 是否成立, 如果成立,

$i \leftarrow i + 1$; 否则, 转到步骤(4)。

- (4) 若 $\alpha_i > \hat{\alpha} + \beta$, 则 $\delta_i \leftarrow (1 - \Delta)\delta_i$; 若 $\alpha_i < \hat{\alpha} - \beta$, 则 $\delta_i \leftarrow (1 + \Delta)\delta_i$, 并转到步骤(1)。

- 4) 将划分后符合误差要求的决策信息表输出。

流程中的 Δ 表示各个属性的类间距每次调整的步长。

1.4 确定性因子与确定性规则^[17]

属性离散化完成后, 最后一步是归纳决策规则。本文使用基于确定性因子的规则权重确定方法, 提取样本中的决策规则。令 $(U, C \cup D, V, f)$ 为一个决策信息表, 其中 C 、 D 分别是条件属性、决策属性, 用 C_i 和 D_j 分别代表 U/C 和 U/D 中的各个等价类, $E(C_i)$ 、 $E(D_j)$ 分别表示对等价类 C_i 、 D_j 的描述, 即等价类 C_i 、 D_j 对于各条件属性的特定取值。

决策规则用式(4)表示:

$$r_{ij}: E(C_i) \rightarrow E(D_j) \quad (4)$$

其中 $D_j \cap C_i \neq \emptyset$, 决策规则的确定性因子定义为

$$\mu(C_i, D_j) = \frac{|C_i \cap D_j|}{|C_i|}, 0 < \mu(C_i, D_j) \leq 1.$$

当 $\mu(C_i, D_j) = 1$ 时, 决策规则 r_{ij} 是确定的; 当 $0 < \mu(C_i, D_j) < 1$ 时, 决策规则 r_{ij} 是不确定的。

一般而言, 规则权重的确定多由专家根据经验知识确定, 如专家打分法, 这种确定方式主观影响大。而在粗糙集理论中, 基于确定性因子的规则权重确定方法不需要任何先验知识, 具有高度的客观性。

2 螺旋桨设计规则挖掘

如图1所示, 传统的专家系统是依靠该领域专家在结合物理定律和一些合理假设的情况下, 通过数学推导得到的设计规则进行初始方案设计, 然后根据专家的经验 and 优化设计理论进行迭代优化, 最终选出优选方案。这个过程中会产生大量的方案, 但专家系统往往不对这些样本进行分析, 事实上, 这些样本中隐藏了设计规则。本文的知识发现系统则以设计过程中产生的大量方案作为素材, 以数据挖掘技术中的粗糙集理论作为工具, 通过属性约简, 提取设计规则, 并希望在以后的研究中, 利用发现的设计规则, 通过反设计得到最终方案。

为了验证以上思路的可行性, 本节使用粗糙集理论对螺旋桨二维设计过程中产生的大量方案基于片条理论进行二维分析, 研究适用于螺旋桨数据的属性表示方法, 属性离散化方法, 确定影

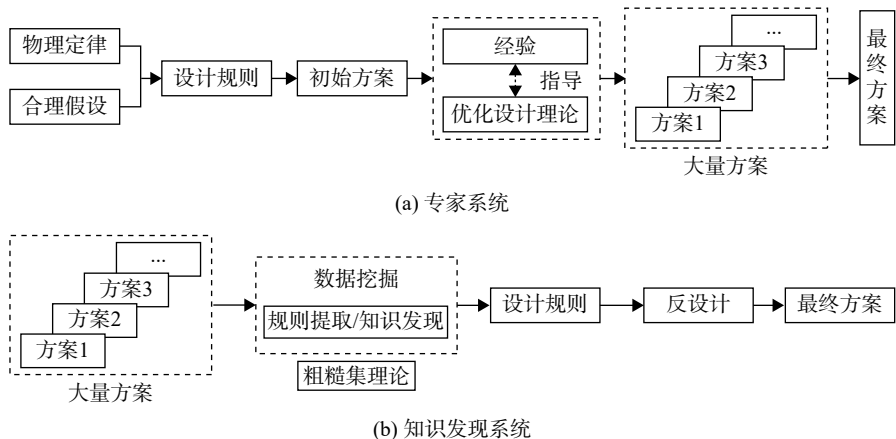


图 1 研究思路

Fig. 1 Research ideas

响螺旋桨效率的属性,进行设计规则提取。若能通过本文方法从中挖掘出螺旋桨设计规则——Betz 条件,则可以验证粗糙集理论在螺旋设计规则挖掘中的有效性,证明基于数据驱动进行螺旋桨设计规则提取这一思路的可行性。

Betz 条件是螺旋桨设计中,基于片条理论假设条件下的重要设计规则。由 Betz 在研究无摩擦阻力条件下,如何使螺旋桨效率最高的问题时,根据茹科夫斯基定理推导得出:当尾涡面像一个向下游移动的刚性螺旋面,即各半径处的轴向诱导速度相等时,螺旋桨效率最高^[1]。

由于 Betz 条件是螺旋桨二维设计中一个重要的设计准则,本文的目的又是证明可以通过适当的数据处理技术挖掘出该规则,即若能在大量外形随机变化的螺旋桨方案中,总结出高效率螺旋桨各剖面处的轴向诱导速度相等或接近相等,便证明了 Betz 条件的存在,也就证明了可以通过本文的方法挖掘出螺旋桨的设计规则。

2.1 数据准备

本文选取笔者设计的两个螺旋桨型号作为研究对象,表 1 列出了它们的基本参数,下面首先针对桨 I 的研究过程进行介绍。

表 1 螺旋桨主要参数
Table 1 Main parameters of propellers

参数	桨 I	桨 II
叶数	4	2
直径/mm	1 200	2 600
转速/(r/min)	2 263	450
速度/(m/s)	55.5	18.5
高度/km	10	12

在初始方案的基础上,对螺旋桨 0.55 倍、0.75 倍、0.9 倍半径处 3 个剖面的弦长和安装角分别进行调整,弦长变化幅度范围为 $\pm 3\%$,安装角变化幅度范围为 $\pm 1.5^\circ$,加上初始值一同排列组合,得到 729 个方案。

在螺旋桨的设计方案中,各剖面诱导速度的分布、推力系数和剖面升阻比这 3 个因素都会影响螺旋桨的效率。因此同样的效率变化,可能是其中某一个因素造成的,也可能是其中某两个,或者全部 3 个因素综合作用引起的。即相同的效率变化,可能对应着多种影响因素的变化情形。本文采用固定推力系数以及剖面升阻比影响的办法,来单独研究各剖面诱导速度的分布对于螺旋桨效率的影响,具体做法如下。

螺旋桨动量理论指出,不同推力系数下,螺旋桨的理想效率不同,故不同推力系数的方案之间比较效率容易造成混淆。为了固定推力系数对螺旋桨效率的影响,本文选择方案中推力系数接近的方案进行研究。以设计推力要求为基准,挑选与设计推力正偏差小于 5% 的方案作为研究样本。

各剖面升阻比的变化也是影响螺旋桨效率的重要因素,为了固定剖面升阻比的影响,本文使用加权升阻比的概念进行研究。定义螺旋桨方案的加权升阻比为

$$k = \sum_{i=1}^n \frac{w_i}{\sum_{i=1}^n w_i} \cdot k_i \quad (5)$$

其中 k_i 是剖面升阻比, w_i 为各剖面升阻比权重函数。二维理论指出 0.75 倍半径处剖面的性能最

能表征桨的特性, 经对比分析, 本文提出, 按式(6)确定剖面升阻比权重函数的值

$$w_i = \frac{1}{(r_i - 0.75)^2 + 1} \quad (6)$$

式中 r_i 是剖面的相对半径位置, $w_i / \sum_{i=1}^n w_i$ 是各剖面的权重系数。

利用式(5)求得方案的加权升阻比后, 选取加权升阻比与基准方案偏差在 $\pm 1\%$ 以内的方案进行分析。

经过以上两步筛选, 留下 26 个方案, 即选出的 26 个方案, 其推力系数和剖面升阻比对效率的影响相当。

2.2 属性表示方法

表征变量值之间差异程度的指标有很多, 统计分析中常用极差 J 、平均差 M 、标准差 T 、变异系数 B ^[18]。其定义式分别如下:

$$J = x_{\max} - x_{\min} \quad (7)$$

$$M = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad (8)$$

$$T = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (9)$$

$$B = \frac{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}}{\bar{x}} \quad (10)$$

式中 x 为样本, $\bar{x}$ 是算术平均值, n 为样本总数。

这几个指标中, 极差指明了样本的最大离散范围, 但未能利用全部测量值的信息, 且易受极端值的影响; 标准差对于不同的样本缺乏可比性; 变异系数在样本平均值接近零时, 对于微小的扰动过于敏感。因此为了综合评价各剖面处诱导速度的差异大小, 本文在求得螺旋桨各剖面诱导速度的极差、平均差、标准差、变异系数后, 进行归一化并求平均值, 作为诱导速度的分散程度属性 F (以下简称分散程度)

$$F = 0.25 \times (J_0 + M_0 + T_0 + B_0) \quad (11)$$

式中下标 0 指对应变量归一化后的值。

选择分散程度作为条件属性, 效率 η 作为决策属性, 组成的决策信息表如表 2 所示。

表 2 原始决策信息表

Table 2 Original decision information table

论域编号	分散程度(条件属性)	效率(决策属性)
1	0.108	0.792 2
2	0.150	0.790 4
3	0.123	0.790 8
4	0.154	0.789 1
5	0.289	0.775 6
6	0.341	0.773 4
7	0.306	0.771 2
8	0.205	0.787 7
9	0.280	0.775 0
10	0.329	0.772 8
11	0.412	0.768 5
12	0.404	0.767 3
13	0.093	0.787 0
14	0.098	0.785 8
15	0.131	0.784 1
16	0.279	0.769 8
17	0.156	0.784 4
18	0.155	0.782 8
19	0.260	0.769 4
20	0.307	0.767 3
21	0.353	0.763 8
22	0.063	0.776 8
23	0.082	0.780 9
24	0.113	0.779 7
25	0.122	0.778 2
26	0.252	0.763 9

2.3 属性离散方法

原始的连续决策信息表在进行离散化后才能进行属性约简。基于动态层次聚类的离散方法是对决策表中的属性进行离散化的常用方法^[19], 但对于本文的数据, 采用动态层次聚类方法容易导致离散后的决策表不相容。

图 2 中直线 a 是动态层次聚类方法对效率的 2 水平离散结果, 可以看到这一离散结果导致在分散程度属性划分为 2 水平后, 决策表出现了如表 3 所示的不相容, 同样的条件属性值(分散程度为 2 水平), 决策属性却有多水平值(效率为 1 或 2 水平), 为了保证相容性, 分散程度属性需要离散成更多水平, 但在分散程度方向上, 造成决不相容的点距离很近, 在使用动态层次聚类离散时, 首先会将其他点离散成不同水平, 因而最终 26 个方案离散成了 24 个水平(如图 3 所示), 才

实现决策表相容。这样的离散结果虽然保证了决策表的相容性,但分散程度属性离散水平过多,会导致最终规则的泛化能力不强,算法实用性很低。

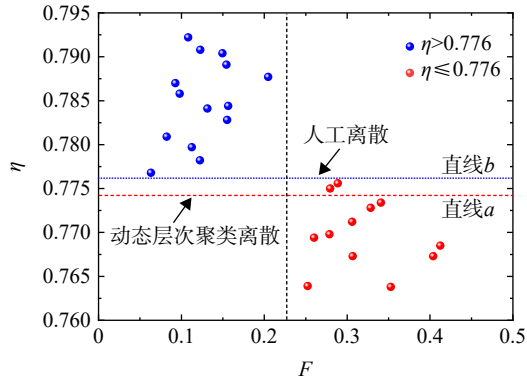


图 2 动态层次聚类离散与人工离散方法对比

Fig. 2 Comparison of dynamic hierarchical clustering discrete and artificial discrete methods

表 3 离散结果中的不相容

Table 3 Incompatibility in discrete results

分散程度水平值	效率水平值
1	2
2	1
2	2

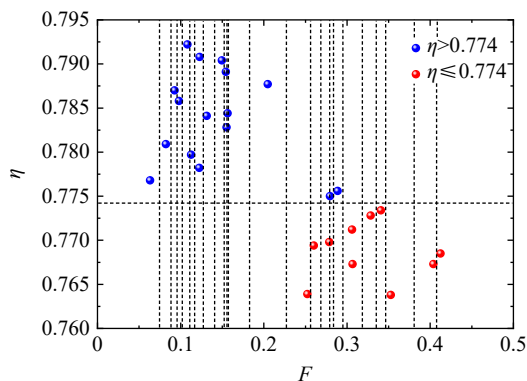


图 3 动态层次聚类离散结果

Fig. 3 Dynamic hierarchical clustering discrete results

如果效率按直线 b 离散为 2 水平,则分散程度属性只需要离散为 2 水平决策表即可相容。因此本文对决策属性(效率)的离散采用人工离散(如何自动进行离散仍需进一步研究),对条件属性采用动态层次聚类方法。

2.4 规则提取

在分散程度作为条件属性的基础上,为检验本文方法能否将无关属性剔除,在决策信息表中

增加一列条件属性作为干扰,属性值采用程序生成的一系列伪随机数。采用上述的方法离散后,得到的决策信息表如表 4 所示。

表 4 离散后的决策信息表

Table 4 Discrete decision information table

论域编号	离散后的条件属性		决策属性
	分散程度水平值	随机数	效率水平值
1	1	20	2
2	1	21	2
3	1	3	2
4	1	26	2
5	2	14	1
6	2	16	1
7	2	10	1
8	1	12	2
9	2	6	1
10	2	18	1
11	2	4	1
12	2	9	1
13	1	24	2
14	1	2	2
15	1	11	2
16	2	7	1
17	1	22	2
18	1	5	2
19	2	1	1
20	2	15	1
21	2	19	1
22	1	8	2
23	1	23	2
24	1	25	2
25	1	17	2
26	2	13	1

利用文中第 1.2 节中介绍的方法进行属性约简,求出相对约简为分散程度,而随机数被作为干扰属性约简掉。然后通过文中第 2.4 节中确定性规则求解方法,得到一条确定性规则

$$r_1: (c_1, 1) \rightarrow (d, 2) \quad (12)$$

即如果分散程度为低水平,那么效率为高水平。该规则的物理含义为螺旋桨各剖面诱导速度的差异越小,螺旋桨效率就越高。事实上,这便是专家系统推导的二维条件下的 Betz 条件^[1]。Betz 条件在二维条件下的表述为螺旋桨径向各剖

面处的诱导速度保持一致。

可见, 本文使用粗糙集理论对螺旋桨设计中产生的大量中间数据进行分析, 可以将无关属性约简, 留下与决策属性(本文为气动效率)相关的条件属性——分散程度, 并最终挖掘出了专家系统推导的螺旋桨设计中的一条重要准则——Betz 条件。

采用本文方法对螺旋桨Ⅱ的设计方案样本进行分析。通过与螺旋桨Ⅰ相同的方法得到 729 个方案, 在进行推力、方案加权升阻比的筛选后, 留下了 81 个方案。对分散程度和效率进行离散, 结果如图 4 所示。采用层次聚类离散方法离散效率时, 效率离散成 3 水平, 离散结果出现了与螺旋桨Ⅰ类似的不相容情况, 直线 a 造成了分散程度为 1 水平时, 效率为 2 或 3 水平。因此采用人工离散方法, 按照图中直线 b 离散效率的 2、3 水平, 保证决策表相容, 分散程度仍然采用层次聚类离散方法。

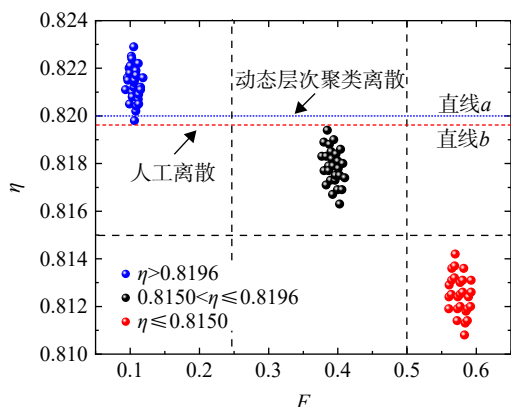


图 4 螺旋桨Ⅱ的离散结果

Fig. 4 Discrete results of propeller II

通过属性约简, 去掉干扰属性。基于确定性因子的规则权重确定方法提取出了确定性规则

$$r_1: (c_1, 1) \rightarrow (d, 3) \quad (13)$$

即当分散程度为低水平时, 效率为高水平, 也即二维条件下的 Betz 条件。因此在第二个螺旋桨型号的研究中, 采用本文方法也提取出了二维条件下的 Betz 条件, 进一步验证了本文方法的适用性。

综上所述, 本节采用粗糙集理论, 对两个螺旋桨型号进行了规则挖掘。首先进行方案挑选, 固定了螺旋桨推力和升阻比的影响; 然后在求得螺旋桨各剖面诱导速度的极差、平均差、标准差和变异系数后, 进行归一化并求平均值, 用于评价

诱导速度的径向差异; 此后在属性离散时, 对决策属性采用人工离散, 对条件属性采用动态层次聚类方法, 最后进行规则提取, 得到了与 Betz 条件相当的规则, 验证了本文方法的可行性。

3 结 论

本文提出了一种螺旋桨设计准则的知识发现系统, 该系统以粗糙集理论为内核, 可对隐藏在设计方案中的螺旋桨设计规则进行提取。①通过设定约束和确定权重的方法, 排除了螺旋桨方案推力的不同以及剖面升阻比的不同对螺旋桨效率分析的影响。②提出对螺旋桨各剖面诱导速度的极差、平均差、标准差和变异系数进行归一化并求平均值的方法, 用于评价诱导速度的径向差异。③针对螺旋桨设计中的数据特点, 对决策属性(螺旋桨效率)采用人工离散, 条件属性(分散程度)采用动态层次聚类离散, 在保证决策表相容性的同时, 避免了属性离散水平过多, 提高了算法的实用性。④通过分别对两种差异比较大的螺旋桨型号生成的数据库进行规则提取研究, 均得到了 Betz 条件, 验证了本文方法的可行性。

本文的知识发现系统区别于一般的专家系统, 以数据为驱动, 无需任何先验的专家知识, 更加客观, 其思想可用于后续三维设计准则的发现和其他设计领域。

在未来的工作中, 一方面需要进一步解决螺旋桨数据的属性离散化问题, 研究能够在保证离散结果相容性的同时, 提高离散自动化程度的属性离散方法。另一方面将研究能够在多因素综合影响效率时, 提取螺旋桨设计规则的方法。

参考文献:

- [1] 周盛, 顾高辉, 潘杰元. 航空螺旋桨与桨扇[M]. 北京: 国防工业出版社, 1994.
- [2] 张伟伟, 寇家庆, 刘溢浪. 智能赋能流体力学展望[J]. 航空学报, 2021, 42(4): 26-71.
ZHANG Weiwei, KOU Jiaqing, LIU Yilang. Prospect of artificial intelligence empowered fluid mechanics[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(4): 26-71. (in Chinese)
- [3] BRUNTON S L, NOACK B R, KOUMOUTSAKOS P. Machine learning for fluid mechanics[J]. Annual Review of Fluid Mechanics, 2020, 52: 477-508.
- [4] SIMPSON T, TOROPOV V, BALABANOV V, et al. Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come-or not[R]. AIAA-2008-5802, 2008.
- [5] JEONG S, SHIMOYAMA K. Review of data mining for multi-disciplinary design optimization[J]. Proceedings of the Institution of

- Mechanical Engineers: Part G Journal of Aerospace Engineering, 2011, 225(5): 469-479.
- [6] JEONG S, CHIBA K, OBAYASHI S. Data mining for aerodynamic design space[J]. *Journal of Aerospace Computing, Information, and Communication*, 2005, 2(11): 452-469.
- [7] CHIBA K, OYAMA A, OBAYASHI S, et al. Multidisciplinary design optimization and data mining for transonic regional-jet wing[J]. *Journal of Aircraft*, 2007, 44(4): 1100-1112.
- [8] CHIBA K, JEONG S, OBAYASHI S, et al. Knowledge discovery in aerodynamic design space for fly back-booster wing using data mining[R]. Canberra, Australia: the 14th AIAA/AHI Space Planes and Hypersonic Systems and Technologies Conference, 2006.
- [9] 汪伟, 莫蓉, 张岩. 叶片气动优化仿真数据的数据挖掘应用研究[J]. *计算机工程与应用*, 2013, 49(12): 11-15.
WANG Wei, MO Rong, ZHANG Yan. Applied research on simulation data of blade optimization designing based on data mining[J]. *Computer Engineering and Applications*, 2013, 49(12): 11-15. (in Chinese)
- [10] 郭振东, 宋立明, 李军, 等. 基于子元模型的全局优化与设计空间知识挖掘方法[J]. *推进技术*, 2015, 36(2): 207-216.
GUO Zhendong, SONG Liming, LI Jun, et al. Meta model-based global design optimization and exploration method[J]. *Journal of Propulsion Technology*, 2015, 36(2): 207-216. (in Chinese)
- [11] 刘卓, 李自启, 梁斌, 等. 基于粗糙集理论的翼型参数优选方法研究[J]. *航空工程进展*, 2016, 7(1): 87-93, 111.
LIU Zhuo, LI Ziqi, LIANG Bin, et al. Method of airfoil selection in wing design using rough set theory[J]. *Advances in Aeronautical Science and Engineering*, 2016, 7(1): 87-93, 111. (in Chinese)
- [12] 刘深深, 陈江涛, 桂业伟, 等. 基于数据挖掘的飞行器气动布局设计知识提取[J]. *航空学报*, 2021, 42(4): 350-364.
LIU Shenshen, CHEN Jiangtao, GUI Yewei, et al. Knowledge discovery for vehicle aerodynamic configuration design using data mining[J]. *Acta Aeronautica et Astronautica Sinica*, 2021, 42(4): 350-364. (in Chinese)
- [13] 庞梦洋, 索中英, 郑万泽, 等. 基于RS-CART决策树的航空发动机小样本故障诊断[J]. *航空动力学报*, 2020, 35(7): 1559-1568.
PANG Mengyang, SUO Zhongying, ZHENG Wanze, et al. Small sample fault diagnosis of aeroengine based on RS-CART decision tree[J]. *Journal of Aerospace Power*, 2020, 35(7): 1559-1568. (in Chinese)
- [14] PAWLAK Z. Rough set theory and its applications to data analysis[J]. *Cybernetics and Systems*, 1998, 29(7): 661-688.
- [15] PAWLAK Z. Rough sets[J]. *International Journal of Computer & Information Sciences*, 1982, 11(5): 341-356.
- [16] 苗夺谦, 李道国. 粗糙集理论、算法与应用[M]. 北京: 清华大学出版社, 2008.
- [17] 周炜, 周创明, 史朝辉, 等. 粗糙集理论及应用[M]. 北京: 清华大学出版社, 2015.
- [18] HAN Jiawei, KAMBER M, PEI Jian, 等. 数据挖掘: 概念与技术[M]. 范明, 孟小峰, 译. 机械工业出版社, 2012.
- [19] 郝丽, 莫蓉, 魏斌斌, 等. 粗糙集理论在关键功能零件识别中的应用[J]. *哈尔滨工业大学学报*, 2021, 53(2): 61-70.
HAO Li, MO Rong, WEI Binbin, et al. Application of rough set theory in identification of key functional parts[J]. *Journal of Harbin Institute of Technology*, 2021, 53(2): 61-70. (in Chinese)

(编辑: 王碧珺)