

DOI:10.11918/202309032

雾化性能预测驱动的航空发动机燃油喷嘴选配

程祥辉¹, 胡志星², 张亚辉³, 胡小锋^{1,4}, 刘跃雄²

(1. 上海交通大学 机械与动力工程学院, 上海 200240; 2. 中国航发南方工业有限公司, 湖南 株洲 412002; 3. 上海交通大学 海洋装备研究院, 上海 200240; 4. 上海市网络化制造与信息化重点实验室(上海交通大学), 上海 200240)

摘要: 为减少航空发动机燃油喷嘴装配中的反复拆卸重装, 提高一次装配成功率, 提出了一种基于雾化性能预判的关键零部件选配方法。首先, 基于喷嘴历史装配数据构建喷嘴几何精度-雾化性能实例库; 然后, 考虑到样本空间大小和喷嘴几何精度波动较大、一致性差的影响, 采用自适应综合过采样方法对样本空间进行扩充, 同时利用改进的 *K*-means 聚类算法对连续属性离散化处理; 最后, 通过关联规则挖掘算法建立几何精度与雾化性能之间的关联关系, 并利用规则适应度评价方法量化每条规则的准确性, 基于这些关联规则集构建喷嘴雾化性能预判模型, 用于指导喷嘴装配。研究结果表明, 利用某双油路离心喷嘴的旋流器和副喷口的装配数据进行验证, 与决策树、支持向量机和人工神经网络等方法进行比较, 本方法提出的喷嘴雾化性能预判模型的预测效果最好, 预测精度高达 98.33%。可以对不同零件组合后的喷嘴雾化性能进行有效预判, 进而减少无效装配, 提高喷嘴的装配效率。

关键词: 燃油喷嘴; 选配; 关联规则挖掘; 样本扩充; 连续属性离散化

中图分类号: TH164

文献标志码: A

文章编号: 0367-6234(2024)12-0105-11

Selective assembly for aero-engine fuel nozzle driven by atomization performance prediction

CHENG Xianghui¹, HU Zhixing², ZHANG Yahui³, HU Xiaofeng^{1,4}, LIU Yuexiong²

(1. School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; 2. Aecc South Industry Company Limited, Zhuzhou 412002, Hunan, China; 3. Institute of Marine Equipment, Shanghai Jiao Tong University, Shanghai 200240, China; 4. Shanghai Key Laboratory of Advanced Manufacturing Environment (Shanghai Jiao Tong University), Shanghai 200240, China)

Abstract: In order to reduce the repeated disassembly and reassembly in aero-engine fuel nozzle assembly and improve the success rate of one assembly, a key part selective assembly method based on atomization performance prediction was proposed. First, based on the historical assembly data of nozzle, the nozzle geometric precision-atomization performance case library was constructed. Next, considering the impact of large fluctuations in sample space size and nozzle geometric accuracy, as well as poor consistency, the sample space was expanded by adaptive comprehensive oversampling method, and simultaneously the continuous attribute was discretized by improved *K*-means clustering algorithm. Finally, the correlation between geometric accuracy and atomization performance was established by association rule mining algorithm, and the accuracy of each rule was quantified by rule fitness evaluation method. Based on these association rule sets, the nozzle atomization performance prediction model was constructed to guide nozzle assembly. The research results show that the nozzle atomization performance prediction model proposed in this paper has the best prediction effect, with a prediction accuracy of up to 98.33%, compared with methods such as decision tree, support vector machine, and artificial neural network, which can effectively predict the atomization performance of different parts combination, thus reducing invalid assembly and improving the assembly efficiency of nozzle.

Keywords: fuel nozzle; selective assembly; association rules mining; sample expansion; continuous property discretization

复杂机械产品结构复杂、价格昂贵, 通常采用小批量生产模式, 其零部件往往根据需求定制加工, 一

般没有库存备件, 但此类产品装配精度要求高, 易出错, 选配能够在现有的加工条件下, 对待装零件的组

收稿日期: 2023-09-12; 录用日期: 2023-10-08; 网络首发日期: 2024-09-10

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20240909.1237.002>

基金项目: 国防基础科研项目(JCKY2021110B048)

作者简介: 程祥辉(1999—), 男, 硕士研究生; 胡小锋(1977—), 男, 研究员, 博士生导师

通信作者: 胡小锋, wshxf@sjtu.edu.cn

合进行优化,在获得较高装配精度的同时减小装配成本,提高装配效率^[1-2]。

目前,国内外许多学者在复杂机械产品选配上进行了研究并取得一定成果。Kannan 等^[3]将装配公差作为优化目标,采用粒子群优化算法对待装配零件进行选配,在降低产品的配合公差的同时还提高了零件的利用率。Asha 等^[4]研究了高精度组件的选配问题,提出了以最小化装配间隙和剩余零件为目标的选配方法,通过轴承的装配验证了选配方法的有效性。宿彪等^[5]面向工程机械再制造存在的问题,以零件装配精度和再制造资源利用率为优化目标,构建了基于多元质量损失函数和装配偏差度的选配模型。段黎明等^[6]为提高选配成功率和装配精度,考虑零件尺寸链与尺寸的关联关系,提出了基于密度的多目标进化算法进行求解,并利用某机械产品的选配验证了方法的可行性与有效性。上述研究虽然从不同优化目标出发提高了装配质量,但均是对装配尺寸链进行分析,通过控制零件偏差在装配尺寸链上的传递和积累来保证装配质量。航空发动机燃油喷嘴作为燃烧室的重要元件,其雾化性能决定了燃烧室燃烧效率,直接影响航空发动机的寿命、可靠性及经济性,往往具有明确性能指标。航空发动机燃油喷嘴由若干个关键零部件组成,不同的零部件组合对喷嘴性能有显著影响。即使各零部件几何精度均在公差范围内波动,但组装后的产品仍会出现不满足性能要求的现象,因此,基于装配尺寸链分析的选配方法难以适用于燃油喷嘴零部件之间的装配。

在燃油喷嘴实际生产中,工人往往通过随机挑选喷嘴关键零件进行组装,喷嘴性能存在显著差异且一次装配成功率低,需要反复拆卸重装并进行性能调试,甚至会尝试更换某个零件继续装配。如果能够提前预判出不同零件组成的喷嘴雾化性能是否合格,就可以减少无效装配,不仅能够提高生产效率,还可降低生产成本。

基于上述分析,通过构建喷嘴雾化性能预判模型,能够在实际装配前基于不同零件组合的几何精度预测喷嘴雾化性能,指导工人进行选配,可以有效提高喷嘴装配效率。喷嘴雾化性能预判模型的构建关键在于建立喷嘴几何精度与雾化性能之间的映射关系,当前相关研究可分为基于理论推导和数据驱动建模的两种方法。基于理论推导的方法主要结合实验方法和理论分析,得到雾化特性的半经验表达

式。Rizk 等^[7]研究了单油路离心喷嘴尺寸和操作条件对喷雾锥角和速度系数的影响,通过理论方法确定液膜厚度计算公式。Xiao 等^[8]通过实验研究了喷嘴几何形状对索特尔平均直径(Sauter mean diameter, SMD)的影响,得到 SMD 与以液膜厚度为特征的韦伯数之间的经验相关性,并建立了一个半经验预测模型来预测 SMD。肖为等^[9]针对双油路离心喷嘴开展理论和试验验证的研究,推导得到喷嘴液膜破碎长度和 SMD 的半经验预测公式。目前燃油雾化理论主要停留在简单雾化条件下的单一雾化机理揭示,在复杂气动条件下,喷嘴雾化特性与几何参数、实际工况以及液滴与空气之间的摩擦等多个因素有关,因此理论推导得到的雾化特性半经验表达式预测误差较大^[10]。数据驱动的建模方法从数据层面出发,通过相应的数据处理算法,建立数据之间的关联关系,可以克服由于误差等因素导致物理模型不准确的问题。盛鑫等^[11]基于响应面法建立了喷嘴雾化特性预测模型,根据得到的喷嘴结构参数和雾化性能之间的显式表达式。游云霞等^[12]以气液同轴式离心喷嘴为研究对象,采用基于改进鲸鱼算法(whale optimization algorithm, WOA)优化的 Kriging 模型建立喷嘴关键几何参数和雾化性能的映射关系。邱贵霞等^[13]利用回归分析,对喷嘴雾化特性与喷口几何结构参数进行多项式拟合,得到几何结构参数对雾化性能的影响规律。

航空发动机燃油喷嘴属于小批量生产件,不仅结构复杂且加工工艺多变,难以通过大批量试验获取足够样本;同一批次零件几何精度波动较大,通常不服从特定分布,随机挑选零件的组合往往难以满足性能要求,因此,为减少燃油喷嘴生产过程中的无效装配,本文提出了一种基于雾化性能预判的选配方法用于指导不同零件之间的选配。其中,考虑到样本数量不足对模型的影响,采用自适应综合过采样(adaptive synthetic sampling, ADASYN)方法进行样本扩充;考虑喷嘴的几何精度波动大、无特定分布,采用改进的 K-means 聚类算法对连续特征离散化处理,构建关联规则挖掘数据集。最后,基于 Apriori 算法建立喷嘴几何精度和雾化性能之间的关联关系,用于构建喷嘴雾化性能预判模型,通过规则适应度评价方法对模型中所有规则的准确性进行量化,从而有效指导零部件选配,即在喷嘴实际装配前,将零件几何参数输入预判模型,根据雾化性能预测结果来决定是否装配。

1 喷嘴几何精度 – 雾化性能实例库构建

本文研究的双油路离心喷嘴结构示意图如图 1 所示,主要由主喷口、副喷口和旋流器这 3 个关键零件组成,其中副喷口用于主油路雾化,旋流器用于副油路雾化。不同零件的组合决定着喷嘴的雾化性能,而实际生产现场往往通过人工随机挑选零件进行组装,待零件装配完成需要进行性能测试,根据测试结果来判断是否装配成功。当性能测试结果不合格时,装配人员将其拆卸并尝试更换某个零件继续装配,该过程不仅对装配人员的装配技术具有较高要求,且一次装配成功率不高,装配效率低下,因此,本文针对实际工程问题,为了减少喷嘴生产中的无效装配,提高一次装配成功率,提出了一种喷嘴选配方法,能够在零件装配前判断喷嘴雾化性能是否合格,从而决定是否装配。

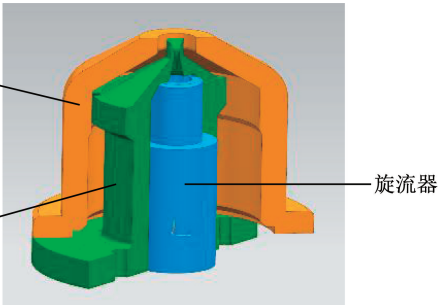


图 1 双油路离心喷嘴结构示意图

Fig. 1 Schematic diagram of dual oil circuit centrifugal nozzle structure

与主油路相比,副油路工况更加复杂,且影响副油路的几何参数更多,几何参数对副油路雾化性能的影响比主油路更加繁多多变,因此,本文选择某一工况下的副油路为研究对象,研究不同旋流器和副喷口组合后的雾化性能是否满足要求,用以验证本文所提选配方法的可行性和有效性。图 2 为副油路结构简图,由副喷口和旋流器组成。旋流器结构如图 3 所示,旋流器有 2 个旋流槽(记为旋流槽 1 和旋流槽 2),燃油经旋流槽沿螺旋线流入旋流室,最后经副喷口喷射。选择等直段直径 d_0 ,副喷口外径 d_1 ,旋流槽 1 的深度 xd_1 和宽度 xw_1 ,旋流槽 2 的深度 xd_2 和宽度 xw_2 ,旋流槽偏心距 l ,副喷口配合面直径 D_1 和旋流器配合面直径 D_2 等几何特征以及副油路流量和雾化锥角这 2 个表征雾化性能的参数,并将流量和雾化锥角与性能要求比较,分为雾化性能合格与不合格 2 类(合格记为 1,不合格记为 0),构建喷嘴几何精度和雾化性能一一对应的实例库,为喷嘴雾化性能预判模型提供数据支持。

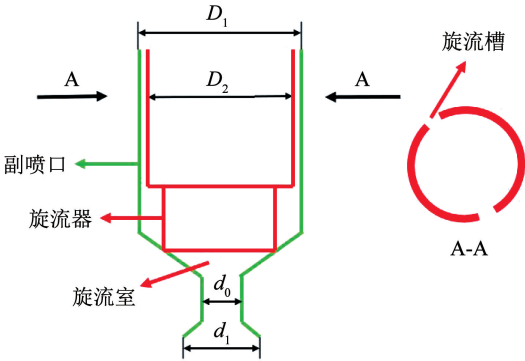


图 2 副油路结构简图

Fig. 2 Sketch of secondary oil circuit structure

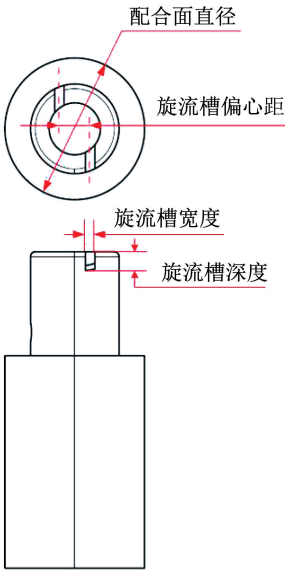


图 3 旋流器结构

Fig. 3 Structure of cyclone

2 关联规则挖掘与 Apriori 算法概述

2.1 关联规则挖掘

关联规则^[14]是一种挖掘和描述数据项或数据项集之间统计关系的一种算法,可以发现隐藏在数据集中有意义的关联关系,形成知识和规则。由于关联规则挖掘结果可解释性好,在许多领域都得到广泛应用,如故障诊断^[15-16]、交通^[17]、旅游^[18]等。

令 $I = \{i_1, i_2, \dots, i_n\}$ 表示所有项目的集合,而 $T = \{t_1, t_2, \dots, t_N\}$ 是所有事务的集合,每个事务 t_i 包含的项集都是 I 的子集。关联规则是形如 $X \rightarrow Y$ 的蕴含表达式, X 称为规则前项, Y 称为规则后项,其中 X 和 Y 是不相交的项集,即 $X \cap Y = \emptyset$ 。关联规则通过支持度 (support) 和置信度 (confidence) 来度量关联强度,其中支持度可以衡量规则出现的频繁程度,置信度则衡量 Y 在包含 X 的事务中出现的频率,支持度和置信度计算公式如下:

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$$

(1)

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (2)$$

式中： $\sigma(\cdot)$ 为计数函数， N 为事务总个数， s 为支持度， c 为置信度。

给定事务集合 T ，关联规则挖掘是找出满足支持度大于或等于最小支持度 (minsup) 且置信度大于或等于最小置信度 (mincon) 的所有规则，其中 minsup 和 mincon 是对应的支持度和置信度阈值。大多数关联规则挖掘算法通常采用的策略是将挖掘任务分解成 2 个子任务：生成频繁集和产生规则。频繁集 (frequent itemsets, FI) 是满足最小支持度阈值的所有项集，当规则 $Z \in \text{FI}$ ，并且置信度大于等于 mincon，此规则称为强关联规则。在实际应用中，需要根据实际情况设置 minsup 和 mincon。

2.2 Apriori 算法

作为最经典的关联规则挖掘算法之一—Apriori 算法^[19]，通过逐层搜索迭代，从候选项集中生成频繁项集，算法流程见图 4。

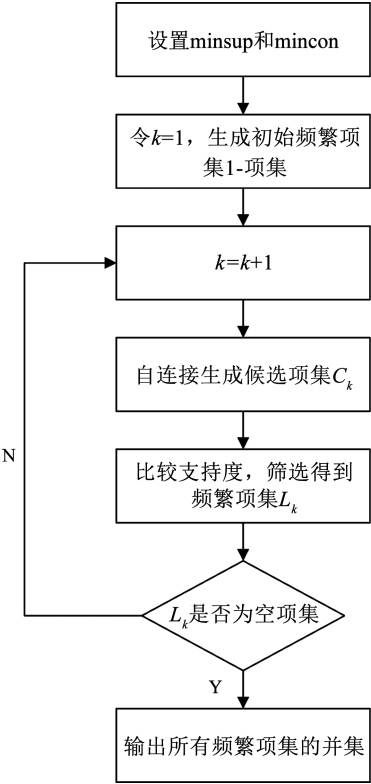


图 4 Apriori 算法流程

Fig. 4 Apriori algorithm flow

3 喷嘴雾化性能预判模型构建

3.1 数据预处理

3.1.1 样本扩充

虽然关联规则挖掘对样本数量没有最小要求，但是与小样本相比，关联规则挖掘能够从更大的样本中挖掘出隐藏在数据间的关联关系，获取更多有

用的关联规则。双油路离心喷嘴由于结构复杂，具有体积小、结构紧凑的特点，对几何精度和形位精度有极高的要求，属于难加工作件，只能小批量生产，难以收集充足样本，因此，本文采用自适应综合过采样方法^[20] (ADASYN) 对样本进行扩充，解决样本数量不足导致结果偏差大和模型泛化能力差等问题。

ADASYN 是一种 SMOTE (synthetic minority over-sampling technique) 变体算法，能够根据样本学习难度自适应确定样本生成数量，即针对特征空间中少数类(相对)样本密度低的区域生成更多样本，在密度较高的区域生成更少样本，样本扩充过程见图 5。

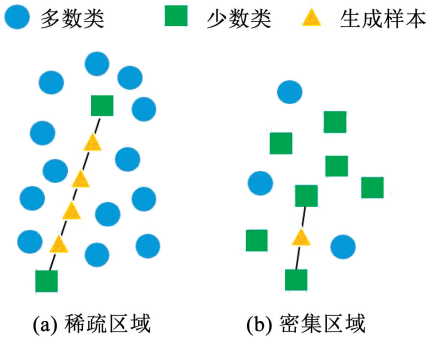


图 5 ADASYN 扩充样本

Fig. 5 ADASYN expand sample

算法具体步骤如下：

输入 设样本数据集包含 m 个样本， $\{x_i, y_i\}$ ($i=1, 2, \dots, m$)，其中 x_i 为特征数据， $y_i \in y = \{0, 1\}$ ，是与 x_i 对应的标签数据。 $y_i=0$ 表示雾化性能不合格，本文定义不合格样本为负类，合格样本为正类；以负类样本的扩充为例，定义 m_s 和 m_l 分别为负类和正类的样本数量。

Step1 确定样本生成数量 N 。

Step2 对于负类中样本 x_i ，计算 K 近邻中正类样本占比：

$$r_i = \frac{\Delta_i}{K} \quad (3)$$

Step3 对 r_i 进行标准化处理：

$$\hat{r}_i = \frac{r_i}{\sum_{i=1}^{m_s} r_i} \quad (4)$$

Step4 计算负类中每个样本 x_i 需要生成的样本数量 N_i 为

$$N_i = \hat{r}_i \times N \quad (5)$$

Step5 通过 SMOTE 算法生成样本：

$$s_i = x_i + (x_{zi} - x_i) \times \tau \quad (6)$$

式中： s_i 为生成的虚拟样本， x_{zi} 为 x_i 的 K 近邻中随机抽取的一个负类样本， τ 为一个随机数， $\tau \in [0, 1]$ 。

3.1.2 连续属性离散化

由于关联规则挖掘只能处理离散型数据,而喷嘴的几何精度均为连续型数据,需要对其离散化处理。喷嘴几何精度离散化结果的好坏会直接影响关联规则挖掘结果的准确性。等宽法^[21]和等频法^[22](或等深法)是连续属性离散化中最常用的 2 种方法,等宽法将属性的值域划分成具有相同宽度的区间,而等频法指离散化后每个区间所包含的样本个数相同。等宽法和等频法虽然原理简单,使用方便,但对数据质量要求也较高。等宽法没有考虑数据的分布特征,仅依赖数域宽度,易导致不均匀的离散化结果。对异常点敏感,当存在离群点时,可能会导致离散化区间之间存在极大偏差,出现某个区间包含大量样本,而另一区间没有样本的情形;且离散化区间个数由人为指定,存在较大主观性,离散化效果的好坏和研究人员的专业水平有直接联系。等频法虽然能够保证每个区间都存在样本,但为了保证区间样本个数相同,可能会将两个数值相同的样本划分到两个不同的区间,对后续模型的建立和算法的影响造成严重影响。基于聚类的连续属性离散化方法

可以充分挖掘属性值中固有的聚合特性,有效降低离散过程中由于随机性和主观性对离散效果的影响,不仅能够得到合理离散区间,还可以提高数据离散化效率^[23-24],因此,本文采用基于聚类的方法对喷嘴几何精度离散化。

采用 K -means 聚类算法对单个连续属性进行离散化,通过肘部法确定最佳聚类个数,即离散区间个数。肘部法是根据聚类结果误差平方和(sum of the squared errors, SSE)随聚类个数变化关系来确定最佳聚类个数 k , SSE 随聚类个数 k 增加而减小,在某个 k 处, SSE 下降速度会突然减小,形成一个拐点,对应最佳聚类个数。然而,当数据集中存在噪声或者离群点时, SSE 随 k 的变化曲线会发生变化,可能表现为无拐点或多个拐点,从而无法确定最佳聚类个数,见图 6。

$$S_{SE} = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (7)$$

式中: S_{SE} 为聚类结果误差平方和, C_i 为第 i 个簇, p 为 C_i 中的样本点, m_i 为 C_i 质心。

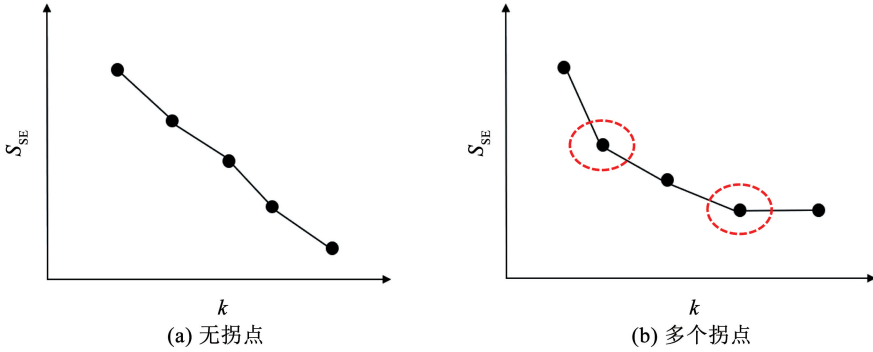


图 6 误差平方和变化曲线

Fig. 6 Sum of squared errors variation curve

针对肘部法失效问题,引入轮廓系数^[25](silhouette coefficient, SC)和戴维森堡丁指数^[26](Davies-Bouldin index, DBI)来辅助确定最佳聚类个数。轮廓系数通过计算每个对象距所在区间的凝聚度和相邻区间的分离度来评价每个对象的离散化效果,戴维森堡丁指数通过计算区间内与区间间距离来评价离散化。轮廓系数越大,表示聚类效果越好;戴维森堡丁指数越小,聚类效果越佳。如果肘部法能够判断最佳聚类个数时,以肘部法结果为主;当时肘部法无法确定最佳聚类个数时,通过轮廓系数和戴维森堡丁指数变化曲线来确定最佳聚类个数。

轮廓系数和戴维森堡丁指数计算公式如下:

$$C_s = \frac{b - a}{\max(a, b)} \quad (8)$$

式中: C_s 为某个对象 x_i 的轮廓系数, a 为 x_i 与同簇其他样本的平均距离,称为凝聚度; b 为 x_i 与最近簇中

所有样本的平均距离,称为分离度。本文通过计算所有对象的平均轮廓系数来衡量聚类效果。

$$I_{DB} = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \frac{W_i + W_j}{C_{ij}} \quad (9)$$

式中: I_{DB} 为戴维森堡丁指数, k 为聚类个数, W_i 为簇 C_i 中所有样本到其聚类中心的平均距离, W_j 为类 C_j 中所有样本到其聚类中心的平均距离, C_{ij} 为类 C_i 和类 C_j 中心之间的距离。

3.2 关联规则生成及后处理

根据最小支持度 minsup 和最小置信度 mincon, 采用 Apriori 算法生成所有频繁项集;再与设定阈值比较,得到所有强关联规则。将规则后项设为喷嘴雾化性能,如果规则后向没有指定唯一的类别,可认为是无效规则,进行剔除。根据筛选得到所有有效关联规则,构建喷嘴雾化性能预判模型。

通过喷嘴雾化性能预判模型来预测喷嘴雾化性

能的原理是将喷嘴的几何精度数据与模型所有规则前项进行匹配,被激活的规则结果即为喷嘴的雾化性能预测结果。当出现一个喷嘴激活多条规则的情形时,此时往往取置信度最大的那条规则结果作为最终预测结果。但是仅通过置信度来评估规则的准确性有一定局限,当数据集中只存在一个样本与某条规则匹配,此时该规则置信度为 100%,但可能只是特例,不具有泛化性,因此,仅根据置信度无法保证规则的预测精度。基于上述分析,本文使用适应度函数^[27] (Fitness) 来评价每条规则的准确性,计算模型中每一条规则的适应度大小。当一条数据同时激活多条规则时,依据它们的适应度大小,选择具有最高适应度的规则对喷嘴雾化性能进行预测。

适应度值基于 3 个独立参数计算到,分别是可靠度 (Confidence)、完备度 (Completeness) 和简洁度 (Simplicity)。可靠度和完备度计算依赖混淆矩阵,用于衡量规则的预测精度,表 1 为混淆矩阵。

表 1 混淆矩阵

Tab. 1 Confusion matrix

预测结果	真实结果	
	正类	负类
正类	T_P	F_P
负类	F_N	T_N

可靠度和完备度计算公式如下:

$$V_{\text{con}} = \frac{T_P}{T_P + F_P}$$

(10)

$$V_{\text{com}} = \frac{T_P}{T_P + F_N}$$

(11)

式中: V_{con} 为可靠度, V_{com} 为完备度, T_P 为训练集中满足规则前项的样本数量, F_P 为满足规则前项但不满足规则后项的样本数量, F_N 为不满足规则后项但满足规则前项的样本数量, T_N 为既不满足规则前项也不满足规则后项的样本数量。

简洁度用来表示规则的可理解性,由关联规则前项的特征个数 N 计算得到:

$$V_{\text{sim}} = \frac{1}{N}$$

(12)

式中 V_{sim} 为简洁度。

综合考虑可靠度、完备度和简洁度,规则适应度计算公式如下:

$$V_{\text{fit}} = w_1 \times (V_{\text{con}} \times V_{\text{com}}) + w_2 \times V_{\text{sim}}$$

(13)

式中: V_{fit} 为规则适应度, w_1 、 w_2 分别为用于控制规则可靠度、完备度和简洁度的权重,本文令 $w_1 = 0.7$, $w_2 = 0.3$ 。

3.3 基于喷嘴雾化性能预判模型的喷嘴选配流程

基于喷嘴雾化性能预判模型的副喷口和旋流器选配流程如图 7 所示,根据喷嘴历史装配数据,构建喷嘴几何精度-雾化性能实例库,对实例库中样本进行数据预处理,包括样本扩充和连续属性离散化。基于 Apriori 算法生成关联规则,计算每条规则适应度,构建喷嘴雾化性能预判模型。在旋流器和副喷口装配前,将两者几何精度数据作为模型输入,对其装配后的雾化性能进行预测,如果预测结果为雾化性能合格,则根据该组合进行装配;否则认定为无效选配组合,更换旋流器或副喷口继续选配,直至雾化性能合格。

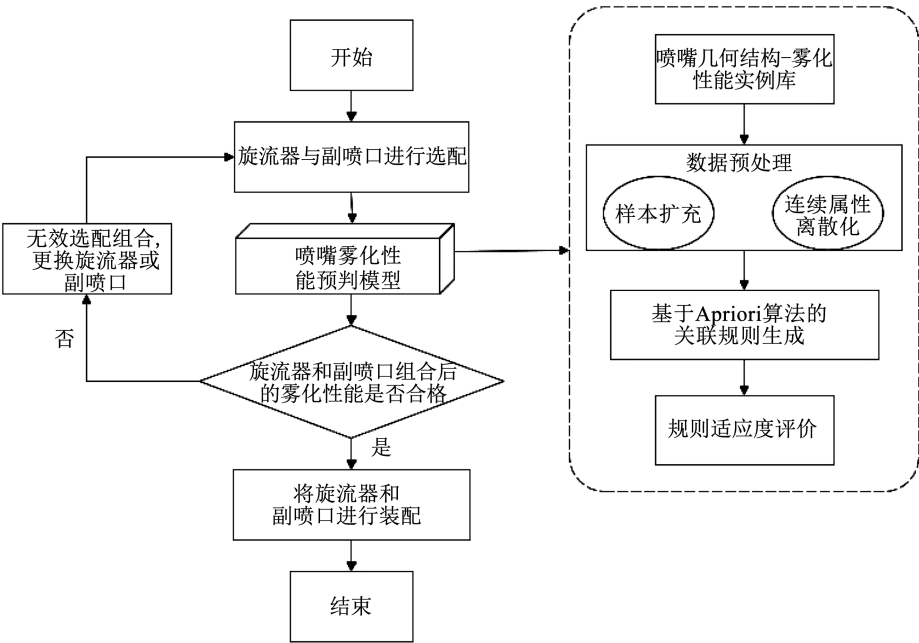


图 7 基于喷嘴雾化性能预判模型的旋流器和副喷口选配流程

Fig. 7 Cyclone and sub-nozzle selective assembly process based on nozzle atomization performance prediction model

4 结果与分析

本文以某双油路离心喷嘴为研究对象,共 20 组副油路压力为 0.3 MPa 下的雾化性能及几何精度参数。副油路流量设计要求为 29 ~ 32 g/min,雾化锥角的变化设计要求为 90° ~ 110°,部分喷嘴雾化性能分类结果见表 2。表 2 中喷嘴 1 和喷嘴 2 虽然锥角合格,但流量不合格,因此这两个喷嘴雾化性能不合格。综合喷嘴雾化性能分类结果,构建喷嘴几何精度-雾化性能实例库,表 3 为实例库部分数据,共有 11 个样本雾化性能不合格,9 个样本雾化性能合格。

表 2 部分喷嘴雾化性能分类结果

Tab. 2 Partial nozzle atomization performance classification results

喷嘴序号	流量/(g·min ⁻¹)	锥角/(°)	雾化性能
1	33.6	90.21	0
2	32.3	103.11	0
3	31.4	91.09	1
4	31.7	92.30	1
5	33.6	101.51	1
6	30.5	92.33	1

表 3 部分喷嘴几何精度-雾化性能实例库

Tab. 3 Partial nozzle geometry accuracy-atomization performance example library

喷嘴序号	旋流槽 1 深度/mm	旋流槽 2 深度/mm	旋流槽 1 宽度/mm	旋流槽 2 宽度/mm	旋流器配合面直径/mm	旋流槽偏心距/mm	等直段直径/mm	副喷口外径/mm	副喷口配合面直径/mm	雾化性能
1	0.188 9	0.192 7	0.179 8	0.176 1	2.336 7	0.258 1	0.252 7	0.758 8	2.473 1	0
2	0.192 2	0.178 1	0.177 6	0.190 5	2.378 1	0.285 5	0.252 7	0.766 7	2.458 8	0
3	0.188 9	0.202 6	0.181 7	0.186 9	2.373 7	0.243 7	0.259 1	0.767 2	2.454 5	1
4	0.206 4	0.210 1	0.180 7	0.187 5	2.353 2	0.238 8	0.268 7	0.764 0	2.465 0	1
5	0.210 4	0.190 6	0.179 3	0.185 1	2.371 7	0.250 8	0.257 4	0.731 0	2.490 9	1
6	0.177 8	0.199 0	0.191 8	0.188 9	2.361 7	0.261 4	0.272 7	0.764 1	2.462 1	1

4.1 数据预处理结果

关联规则挖掘结果的可靠性往往与数据集大小成正比,当样本数量越大时,喷嘴雾化性能预判模型的预测精度越高。本研究的样本数量显然不够,因此通过 ADASYN 生成一定数量样本,对实例库进行扩充。在分类问题中,基于机器学习的数据模型预测结果通常向多数类样本偏倚,如果各类样本数量相当,预测结果则更准确。

考虑到样本数据集的大小对预测精度具有重要影响,本文通过比较每一类样本扩充数量的大小对预测精度的影响来确定最佳扩充数量,试验结果如图 8 所示,预测精度随样本扩充数量增加呈先上升后下降的趋势,当样本数量扩充至 150 时,模型此时预测精度最高。这是因为当样本扩充数量不够时,对原始样本空间的信息挖掘不够充分,而样本扩充数量过多时,扩充样本的空间则会覆盖原始样本空间,导致原始样本难以被学习,预测精度下降。综上所述,将正类和负类样本数量均扩充至 150,生成新的几何精度-雾化性能实例库。

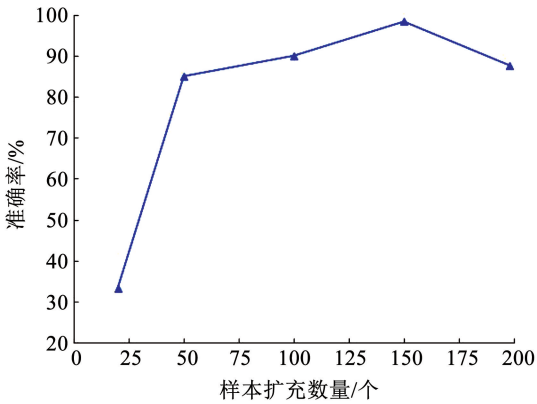


图 8 不同样本扩充数量对预测精度的影响

Fig. 8 Impact of different sample expansion quantities on prediction accuracy

基于样本扩充后的实例库,对连续属性离散化处理。图 9(a)、(b)分别为旋流槽 1 深度 xd_1 和等直段直径 d_0 离散化过程,由于评价指标 SSE 和 SC、DBI 数量级相差较大,需要对其作归一化处理才能与 SC 和 DBI 的变化曲线一起比较。由图 9(a)可知拐点在 $k=2$ 处,且该点的轮廓系数最大,戴维森堡丁指数最小, xd_1 离散效果最好。在图 9(b)中,误差平方和随聚类个数的变化曲线比较平滑,根据肘部

法原则,无法通过找出拐点位置来确定最佳聚类个数,此时可以比较 SC 和 DBI 变化曲线来辅助确定最佳聚类个数。可以发现 SC 随聚类个数变化呈上

升趋势,DBI 随聚类个数变化呈下降趋势,两者均在 $k=3$ 处变化最大;随着 k 继续增大,SC 和 DBI 波动幅度很小,因此,当 $k=3$ 时, d_0 离散化效果最佳。

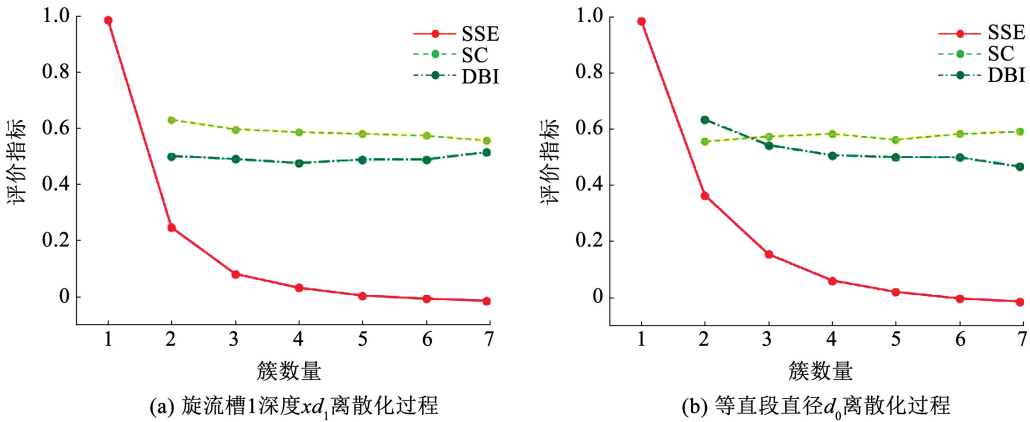


图 9 连续属性离散化
Fig. 9 Continuous property discretization

表 4 为喷嘴几何精度离散化结果,对于每个特征,将各个区间根据聚类中心大小顺序排列,取相邻区间的上界和下界的平均值作为离散化边界。

表 4 喷嘴几何精度离散化结果
Tab. 4 Discretization results of nozzle geometric accuracy

几何结构	旋流槽 1 深度/mm	旋流槽 2 深度/mm	旋流槽 1 宽度/mm	旋流槽 2 宽度/mm	旋流器配合 面直径/mm	旋流槽 偏心距/mm	等直段 直径/mm	副喷口 外径/mm	副喷口配合 面直径/mm
区间 1	[0.170 3, 0.186 4)	[0.174 8, 0.191 5)	[0.170 3, 0.183 8)	[0.174 9, 0.181 3)	[2.336 7, 2.364 8)	[0.235 5, 0.274 7)	[0.191 5, 0.240 4)	[0.731 0, 0.749 0)	[2.439 6, 2.469 9)
区间 2	[0.186 4, 0.197 5)	[0.191 5, 0.210 1]	[0.183 8, 0.191 8]	[0.181 3, 0.186 6)	[2.364 8, 2.383 8]	[0.274 7, 0.332 0]	[0.240 4, 0.269 1)	[0.749 0, 0.769 2]	[2.469 9, 2.497 9)
区间 3	[0.197 5, 0.211 9]			[0.186 6, 0.193 8]			[0.269 1, 0.293 5]		[2.497 9, 2.526 3]

4.2 喷嘴雾化性能预判模型

将数据预处理之后的数据集随机划分 80% 为训练集,20% 为测试集。基于 Apriori 算法对训练集进行关联规则挖掘,建立喷嘴几何精度和雾化性能之间的映射关系。在关联规则挖掘过程中,最小支持度 minsup 和最小置信度 mincon 的选择对关联规则的生成结果好坏至关重要。如果 minsup 和 mincon 过小,则会生成许多无用的规则;如果 minsup 和 mincon 过大,一些有意义的关联规则可能会被删除,降低预判模型的性能预测能力,因此,本文通过网格搜索法,比较不同支持度和置信度下的模型预测精度,确定最佳 minsup 和最佳 mincon。本次实验设置 minsup 搜索范围为[0.05,0.20],步长为 0.05; mincon 搜索范围为[0.8,1.0],步长为 0.1。

表 5 为网格搜索优化结果,当 mincon 取 1.0, minsup 取 0.05 或 0.10 时,此时喷嘴雾化性能预判模型的预测精度最高,为 0.98。虽然 minsup 取不同

值时,模型预测精度相等。但当 minsup = 0.05 和 minsup = 0.10 时,分别挖掘到 1 059 条和 98 条关联规则;minsup 越小,挖掘到的冗余规则就越多,大大增加计算负担和成本。因此,本文[minsup, mincon]的最佳组合为[0.1,1.0]。

表 5 网格搜索优化结果
Tab. 5 Grid search optimization results

mincon	minsup			
	0.05	0.10	0.15	0.20
0.8	0.65	0.65	0.67	0.67
0.9	0.87	0.90	0.88	0.57
1.0	0.98	0.98	0.60	0.27

在确定最优 minsup 和最优 mincon 后,最终得到由 98 条关联规则构成的喷嘴雾化性能预判模型。根据式(10)~(13)计算模型中每条关联规则的适应度值,结果见表 6,表 6 为模型部分关联规则,雾化性能记为 C。

表 6 模型部分关联规则

Tab.6 Model part association rules		
规则序号	适应度	关联规则
1	0.440 0	$d_0 = 1 \rightarrow C = 0$
2	0.284 4	$xw_1 = 2, xw_2 = 1 \rightarrow C = 0$
3	0.150 0	$d_0 = 3, xd_2 = 2 \rightarrow C = 1$
4	0.100 0	$d_0 = 3, xd_1 = 1, d_1 = 1 \rightarrow C = 1$
5	0.279 2	$d_1 = 1, d_0 = 2, xd_1 = 2 \rightarrow C = 0$

4.3 实验评价指标

为更加全面、有效的评估模型的预测性能,选取基于混淆矩阵的分类评价指标体系,将查准率 f_1 、查全率 f_2 及马修斯相关系数(Matthews correlation coefficient, Mcc)作为模型评价指标。查准率 f_1 是衡量模型误判程度的指标,查准率越高意味着模型的可靠度越高;查全率 f_2 是衡量模型漏判程度的指标,查全率越高表示模型的识别灵敏度越高;马修斯相关系数 c_{Mc} 则是一个较为平衡的评价指标,综合考虑了查准率和查全率。马修斯系数在 $[-1,1]$ 之间变化, $c_{Mc} = 1$ 表示完美预测, $c_{Mc} = 0$ 意味着比随机预测结果差, $c_{Mc} = -1$ 表示预测结果与真实值完全相反。

表 7 不同样本扩充方法结果对比

Tab.7 Comparison of different sample expansion methods results					
扩充方法	准确率/%	f_1 /%	f_2 /%	c_{Mc} /%	规则数量
未扩充	75.00	50.00	37.50	0	509
bootstrap	78.33	66.67	53.34	68.44	367
SMOTE	90.00	65.38	60.00	82.61	215
Borderline-SMOTE	90.00	65.48	60.00	82.53	276
ADASYN	98.33	66.67	65.71	96.71	98

将本文提出的离散化方法与等宽法、等频法进行对比,结果见表 8。采用等宽法和等频法连续属性离散化时,可以发现属性离散区间个数越多,生成的规则越少,模型的预测精度下降,在各项指标上得分均次于本文方法,表明本文方法离散化效果最佳,

表 8 不同离散化方法结果比较

Tab.8 Comparison of different sample expansion methods results						
离散化方法	区间个数	准确率/%	f_1 /%	f_2 /%	c_{Mc} /%	规则数量
等宽法	2	81.67	66.67	53.90	71.82	105
	3	73.33	66.67	48.00	62.75	52
等频法	2	95.00	64.20	63.81	90.41	217
	3	83.33	61.90	56.76	72.44	35
本文方法		98.33	66.67	65.71	96.71	98

计算公式如下:

$$f_1 = \frac{n_{TP}}{n_{TP} + n_{FP}}$$

(14)

$$f_2 = \frac{n_{TP}}{n_{TP} + n_{FN}}$$

(15)

$$c_{Mc} = \frac{n_{TP} \cdot n_{TN} - n_{FP} \cdot n_{FN}}{\sqrt{(n_{TP} + n_{FP})(n_{TP} + n_{FN})(n_{TN} + n_{FP})(n_{TN} + n_{FN})}}$$

(16)

式中: c_{Mc} 为马修斯相关系数, n_{TP} 为划分正确的正类样本个数, n_{FN} 为假的负类样本个数, n_{FP} 为假的正类样本个数, n_{TN} 为划分为正确的负类样本个数。

4.4 实例验证与比较

为验证本文样本扩充方法的有效性,分别与原始数据集、SMOTE、Borderline-SMOTE 以及 bootstrap 等几种样本生成方法进行比较,结果见表 7。与原始数据集相比,采用不同样本扩充方法后,喷嘴雾化性能预判模型预测精度均有所提高。通过 ADASYN 样本扩充后的数据集质量最好,生成的规则数量远小于其他方法,有效减小冗余规则的数量;在各个指标上得分均最高,准确率和马修斯相关系数显著高于其他方法。

表 9 测试样本参数

Tab. 9 Test sample parameters

喷嘴 序号	旋流槽 1 深度/mm	旋流槽 2 深度/mm	旋流槽 1 宽度/mm	旋流槽 2 宽度/mm	旋流器配合 面直径/mm	旋流槽 偏心距/mm	等直段 直径/mm	副喷口 外径/mm	副喷口配合 面直径/mm	雾化 性能
1	2	2	1	1	1	1	2	2	2	0
2	3	1	1	2	2	1	2	1	2	1

根据给定样本几何精度参数,从喷嘴雾化性能
预判模型中筛选出所有被激活的规则,选择规则适
应度值最大的规则为最优规则预测喷嘴性能。
表 10为喷嘴雾化性能预测结果,1 号喷嘴与预判模
型中 3 条规则匹配,最优规则为:当副喷口配合面直
径属于区间 2,旋流器过油槽 1 深度属于区间 2,等
直段直径属于区间 2 时,雾化性能为 0,规则适应度
为 0.279 2;2 号喷嘴与预判模型中 5 条规则匹配,
最优规则为:当副喷口配合面直径属于区间 2,旋流
器过油槽 1 深度属于区间 3,旋流器过油槽偏心距
属于区间 1 时,雾化性能为 1,规则适应度为 0.100 0,
基于喷嘴雾化性能预判模型的预测结果与实际结果
一致。表 11 为喷嘴雾化性能预判模型对所有测试
数据集的预测结果,只有一个样本预测错误,表明本
文提出的方法可以有效预判旋流器和副喷口组合后
的雾化性能,减少无效装配,降低装配成本。

为进一步验证本文提出的喷嘴雾化性能预判模
型的有效性,与决策树,支持向量机及人工神经网络
这几种经典分类模型进行比较,对比结果如图 10 所
示,本文方法在测试数据集上的表现最好,预测准确
率达 98.33%,马修斯系数为 96.71%,模型预测精
度高,泛化能力强。

表 10 喷嘴雾化性能预判模型预测结果

Tab. 10 Prediction results of nozzle atomization performance
prediction model

喷嘴 序号	激活规 则数量	最优规则	适应度
1	3	$d_1 = 2, xd_1 = 2, d_0 = 2 \rightarrow C = 0$	0.279 2
2	5	$d_1 = 2, xd_1 = 3, l = 1 \rightarrow C = 1$	0.100 0

表 11 雾化性能预测结果混淆矩阵

Tab. 11 Confusion matrix of atomization performance prediction
results

预测结果	真实结果	
	0	1
0	34	0
1	1	25

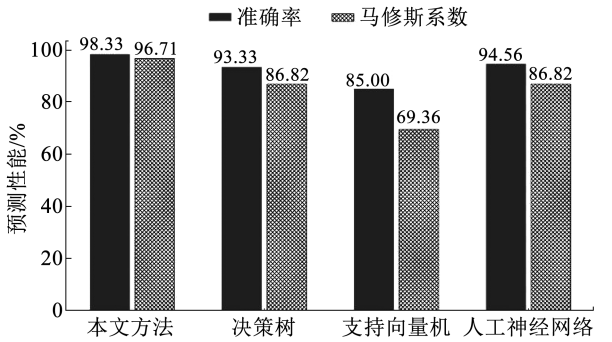


图 10 不同方法对比结果

Fig. 10 Comparison results of different methods

5 结 论

1) 利用自适应综合过采样方法(ADASYN)可以
对样本空间进行有效扩充,改善了样本数量不足对
预测模型的影响,使预测结果更加准确。本文所提
的改进 K-means 聚类算法与等宽法、等频法相比,充
分考虑数据整体分布,能够更好地挖掘属性值中固
有的聚合特性,离散化效果更好。

2) 通过 Apriori 算法建立喷嘴几何精度和雾化
性能之间的关联关系,并采用适应度评价方法对关
联规则准确性进行量化,在此基础上构建的喷嘴雾
化性能预判模型,可以在触发多条规则时,根据适
应度大小选择最准确的规则进行预测,提高预测精
度。

3) 与其他方法比较,本文提出的喷嘴雾化性能
预判模型预测效果最好,预测精度达到 98.33%,
马修斯系数为 96.71%,可以有效判断喷嘴不同零
件组合后的雾化性能是否满足要求,进而指导实际
生产现场装配,提高喷嘴装配效率。

参考文献

[1]曹杰,高智勇,高建民,等. 基于制造公差
的复杂机械产品精准
选配方法[J]. 计算机集成制造系统, 2020, 26(7): 1729
CAO Jie, GAO Zhiyong, GAO Jianmin, et al. Precise selective
assembly method for complex mechanical products based on
manufacturing tolerance [J]. Computer Integrated Manufacturing
Systems, 2020, 26(7): 1729. DOI: 10.13196/j. cims. 2020. 07.
001
[2]任水平,刘检华,何永熹,等. 机械产品多质量要求下的选择装
配方法[J]. 计算机集成制造系统, 2014, 20(9): 2117
REN Shuiping, LIU Jianhua, HE Yongxi, et al. Selective assembly

- method for mechanical product with multi-objective assembly function [J]. Computer Integrated Manufacturing Systems, 2014, 20(9): 2117. DOI: 10.13196/j.cims.2014.09.006
- [3] KANNAN S M, SIVASUBRAMANIAN R, JAYABALAN V. Particle swarm optimization for minimizing assembly variation in selective assembly[J]. The International Journal of Advanced Manufacturing Technology, 2009, 42(7/8): 793. DOI: 10.1007/s00170-008-1638-7
- [4] ASHA A, BABU J R. Comparison of clearance variation using selective assembly and metaheuristic approach [J]. International Journal of Latest Trends in Engineering and Technology, 2017, 8(3): 148. DOI: 10.21172/1.83.020
- [5] 宿彪, 黄向明, 任莹晖, 等. 基于蚁群算法的工程机械再制造优化选配方法研究[J]. 机械工程学报, 2017, 53(5): 60
- SU Biao, HUANG Xiangming, REN Yinghui et al. Research on selective assembly method optimization for construction machinery remanufacturing based on ant colony algorithm [J]. Journal of Mechanical Engineering, 2017, 53(5): 60. DOI: 10.3901/JME.2017.05.060
- [6] 段黎明, 涂玉林, 李中明, 等. 基于密度的进化算法的机械产品选配方法[J]. 计算机集成制造系统, 2020, 26(2): 312
- DUAN Liming, TU Yulin, LI Zhongming, et al. Selective assembly of mechanical product based on density-based evolutionary algorithm [J]. Computer Integrated Manufacturing Systems, 2020, 26(2): 312. DOI: 10.13196/j.cims.2020.02.004
- [7] RIZK N K, LEFEBVRE A H. Prediction of velocity coefficient and spray cone angle for simplex swirl atomizers[J]. International Journal of Turbo and Jet Engines, 1987, 4(1/2): 65. DOI: 10.1515/TJJ.1987.4.1-2.65
- [8] XIAO Wei, HUANG Yong. Semi-empirical correlation to predict the sauter mean diameter of the pressure-swirl atomizer [C]//ASME International Mechanical Engineering Congress and Exposition. San Diego: American Society of Mechanical Engineers, 2013: 1. DOI: 10.1115/IMECE2013-62907
- [9] 肖为, 罗莲军, 马柱, 等. 双油路离心喷嘴雾化特性的半经验预测方法[J]. 航空动力学报, 2018, 33(2): 463
- XIAO Wei, LUO Lianjun, MA Zhu, et al. Semi-empirical prediction on spray characteristics of dual-orifice atomizers [J]. Journal of Aerospace Power, 2018, 33(2): 463. DOI: 10.13224/j.cnki.jasp.2018.02.025
- [10] 严红, 陈福振. 航空发动机燃油雾化特性研究进展[J]. 推进技术, 2020, 41(9): 2038
- YAN Hong, CHEN Fuzhen. Review on fuel atomization in aeroengine [J]. Journal of Propulsion Technology, 2020, 41(9): 2038. DOI: 10.13675/j.cnki.tjjs.200333
- [11] 盛鑫, 侯力, 游云霞, 等. 离心式喷嘴结构参数的多目标优化设计[J]. 机械设计与制造, 2023(8): 60
- SHENG Xin, HOU Li, YOU Yunxia, et al. Multi-objective optimization design of pressure swirl nozzle geometric parameters [J]. Mechanical Design and Manufacturing, 2023(8): 60. DOI: 10.19356/j.cnki.1001-3997.20230310.011
- [12] 游云霞, 侯力, 易宗礼, 等. 气液同轴离心式喷嘴雾化性能及优化设计研究[J]. 机械工程学报, 2022, 58(1): 201
- YOU Yunxia, HOU Li, YI Zongli, et al. Study on spray performance and optimization design of gas-liquid coaxial swirl nozzle [J]. Journal of Mechanical Engineering, 2022, 58(1): 201. DOI: 10.3901/JME.2022.01.201
- [13] 邱贵霞, 侯力, 易宗礼, 等. 离心喷嘴出口结构参数对雾化性能的影响分析[J]. 推进技术, 2020, 41(12): 2782
- QIU Guixia, HOU Li, YI Zongli, et al. Effects of structure parameters of centrifugal nozzle on atomization performance [J]. Journal of Propulsion Technology, 2020, 41(12): 2782. DOI: 10.13675/j.cnki.tjjs.190640
- [14] AGRAWAL R, IMIELIŃSKI T, SWAMI A. Mining association rules between sets of items in large databases [C]//Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. Washington: ACM, 1993: 207. DOI: 10.1145/170035.170072
- [15] DOOSTAN M, CHOWDHURY B H. Power distribution system fault cause analysis by using association rule mining [J]. Electric Power Systems Research, 2017, 152: 140. DOI: 10.1016/j.epsr.2017.07.005
- [16] 周家玉, 侯慧娟, 盛戈峰, 等. 状态参量关联规则挖掘及深度学习融合的变压器故障诊断算法[J]. 高压电器, 2023, 59(3): 108
- ZHOU Jiayu, HOU Huijuan, SHENG Gehao, et al. Transformer fault diagnosis algorithm based on association rules mining of state parameters and deep learning [J]. High Voltage Apparatus, 2023, 59(3): 108. DOI: 10.13296/j.1001-1609.hva.2023.03.015
- [17] ZHENG Xiaofeng, WANG Shu. Study on the method of road transport management information data mining based on pruning Eclat algorithm and MapReduce [J]. Procedia-Social and Behavioral Sciences, 2014, 138: 757. DOI: 10.1016/j.sbspro.2014.07.254
- [18] VERSICHELE M, DE GROOTE L, BOUUAERT M C, et al. Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium [J]. Tourism Management, 2014, 44: 67. DOI: 10.1016/j.tourman.2014.02.009
- [19] AGRAWAL R, SRIKANT R. Fast algorithms for mining association rules [C]//Proceedings of the 20th Very Large Data Bases Conference. Santiago: VLDB, 1994: 487
- [20] HE Haibo, BAI Yang, GARCIA E A, et al. ADASYN: adaptive synthetic sampling approach for imbalanced learning [C]//2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). Hong Kong: IEEE, 2008: 1322. DOI: 10.1109/IJCNN.2008.4633969
- [21] HONG Jialin, QIN Mengzhao. Multisymplecticity of the centred box discretization for Hamiltonian PDEs with $m \geq 2$ space dimensions [J]. Applied Mathematics Letters, 2002, 15(8): 1005. DOI: 10.1016/S0893-9659(02)00077-0
- [22] FUKUDA T, MORIMOTO Y, MORISHITA S, et al. Mining optimized association rules for numeric attributes [C]//Proceedings of the fifteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems. Montreal: ACM, 1996: 182. DOI: 10.1145/237661.237708
- [23] 焦磊, 刘晓军, 刘庭煜, 等. 一种面向生产调度规则挖掘的数据离散化方法[J]. 计算机集成制造系统, 2016, 22(1): 257
- JIAO Lei, LIU Xiaojun, LIU Tingyu, et al. Data discretization method for rules discovery of production scheduling [J]. Computer Integrated Manufacturing Systems, 2016, 22(1): 257. DOI: 10.13196/j.cims.2016.01.025
- [24] GUPTA A, MEHROTRA K G, MOHAN C. A clustering-based discretization for supervised learning [J]. Statistics & Probability Letters, 2010, 80(9/10): 816. DOI: 10.1016/j.spl.2010.01.015
- [25] ROUSSEEUW P J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis [J]. Journal of Computational and Applied Mathematics, 1987, 20: 53. DOI: 10.1016/0377-0427(87)90125-7
- [26] DAVIES D L, BOULDIN D W. A cluster separation measure [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1979. PAMI-1(2): 224. DOI: 10.1109/TPAMI.1979.4766909
- [27] FREITAS A A. A survey of evolutionary algorithms for data mining and knowledge discovery [J]. Advances in Evolutionary Computing, 2003: 819. DOI: 10.1007/0-387-25465-X_20