

第3章 基于滚磨光整加工数据库的工艺参数优选案例库构建

目前滚磨光整加工技术已广泛应用于各行各业的**零部件的表面加工中**，随着各企业生产的不断进行，已经积累了大量合格的加工实例，这些生产实例中包含着丰富的知识，为了使之能够在新工艺方案制定过程中发挥更大作用，本论文提出基于长期研发滚磨光整加工工艺实践和中试基地生产实例的数据积累，构建了面向全产业链应用的滚磨光整加工数据库。**然而面对光整加工零件多变、要求多样的实际，必须实现工艺方案优选的智能化**，为此提出在已经研发的滚磨光整加工数据库平台上，将案例推理技术（CBR）和专家推理（ER）有机结合，形成**融合推理**，实现滚磨光整加工工艺的快速、准确优选。即：在实际应用时，首先从大量的生产实例中挖掘信息，构建案例，形成案例库，同时将新问题（包括加工零件及加工要求）进行特征描述，然后采用加权案例推理技术（Weighted Case-based Reasoning, WCBR），在案例库中寻找与新问题匹配的原有案例，以便快速找到问题的解；如果没有匹配案例，则借助模糊专家系统（Fuzzy Expert System, FES），充分挖掘已有案例中的知识，构建区间值模糊规则，通过区间值模糊推理寻找新问题的相似案例。然后采用匹配案例或大于阈值的相似案例的工艺方案进行生产试验，满足加工要求后再投入批量生产。经过案例评价后，将合格的案例存入案例库。由此可知，基于滚磨光整加工数据库构建合理的案例库是实现加工工艺智能优选的前提和基础。

光整加工工艺案例库是**案例推理技术**的知识库，由于案例库中合格案例不断保存，案例库所包含的加工工艺知识越来越多，更有助于新零件加工方案的制订。但同时案例库中可能会出现大量相互非常相似的冗余案例，这样不仅需要增加案例存储空间，同时也会对案例推理过程中的检索效率产生影响，所以为了保证案例库中案例数量和分布的合理性，需要采取有效措施对案例库进行优化。本章采用**模糊 C 均值聚类算法（Fuzzy C-Means, FCM）**解决案例库中案例的冗余问题。由于在 FCM 聚类算法中聚类数和聚类中心的初始值选择，会直接影响到聚类过程的稳定性以及聚类结果的合理性，故提出首先采用**减法聚类算法（Subtractive Clustering Algorithm）**确定**FCM 聚类算法的初始聚类数和初始聚类中心**，提出**采用减法聚类算法的模糊 C 均值聚类方法（Subtractive Fuzzy C-Means, S-FCM）**对滚磨光整加工工艺案例库的进行优化，首先通过聚类找到案例库中的特殊案例，并保留；然后在各类中通过相似度计算，找到冗余案例并删除，从而更加有效地实现案例库优化。

3.1 滚磨光整加工工艺实例的数据分析

本论文第二章构建的滚磨光整加工工艺数据库平台是面向滚磨光整加工全产业链，

涉及众多企业和研发机构，包括国内外不同类型及规格的设备研制、固/液介质生产、整体工艺研发机构或企业以及针对不同产品不同需求应用滚磨光整技术的企业。数据库平台上拥有产业链中的不同企业的已有生产工艺数据和持续产生的新数据，其数据数量庞大、数据结构复杂。为了能够更好地利用案例推理和专家推理求解问题，必须构建合理的滚磨光整加工工艺优选案例库。只有依据工艺实例充分分析实例组成的数据特征，才能构造案例库模型并实现基于案例库的智能推理优选。

通过对大量的不同企业的生产实例报告和实验报告的分析，可以得到每次加工时涉及到的相关参数信息，主要有：加工设备参数（产地、参数）、零件加工前后的参数（零件本身的类型、材质、尺寸，以及加工前后的状态参数）、磨块参数（产地、参数）、磨液参数（产地、参数）、加工参数等。

通过对滚磨光整加工数据库平台大量的生产实例数据分析可知：

（1）零件尺寸数据方面：不同类型的零件其尺寸数据组成差异较大，既要考虑总体尺寸大小也要考虑局部结构参数，既要考虑重要尺寸的精度又要考虑加工受限部位尺寸。这些尺寸直接影响设备及磨块的选型。

（2）加工要求方面：以去毛刺、棱边倒圆、提高表面粗糙度等级、提高光亮度为主要加工要求。其中将加工前后表面粗糙度对比作为一个较为常见的加工要求，通常加工后表面粗糙度值普遍比加工前表面粗糙度值要低，并且加工效果以加工后表面粗糙度减少一个等级和接近一个等级的居多，加工后表面粗糙度等级提高两个等级以上的是少数。表面粗糙度的变化主要与所选择的磨块和加工时间有关。一般加工时间越长且选择与表面粗糙度等级匹配的磨块的案例加工效果较好。

（3）加工工艺参数：最简单的案例中包括设备、磨块、化学剂及加工时间。其中设备选型主要与零件结构尺寸相关；磨块的选择主要与加工之前零件的表面粗糙度和加工以后零件的表面粗糙度要求相关，化学剂选择主要与工件材质相关，需要说明的是这三者的选择也存在一定的交互相关性。大多数案例加工时间在 30min 以内，少数案例加工时间大于 30min 或因分步加工而有两个以上的加工时间。

3.2 基于滚磨光整加工数据库的工艺案例表征

目前滚磨光整加工生产现场已经积累了大量的滚磨光整加工的成功实例，这些实例是针对多变的被加工零件和多样的加工要求实施滚磨光整加工后形成的合格加工数据。它们大多是以传统纸质试验报告文件或者电子文档为数据保存手段，其中包含着许多丰富的光整加工专家经验，暗含着大量有关光整加工工艺参数之间相互关系的潜在规律。为了使这些合格的实例能够在新工艺方案制定过程中发挥更大作用，首先需要抽取待加工零件的特征和加工要求的主要特征及其对应的滚磨光整加工工艺参数，形成案例，并

存入案例库中，即完成生产实例的案例化表征，作为采用案例推理进行案例检索的知识库，为实施案例推理提供基础条件。

案例库构建是采用案例推理求解问题的第一步，也是最为关键的一步，把数据库中的“数据”，经过分析研究，找到影响加工工艺参数选择的“主要因素”，作为“案例特征”；然后对加工工艺的参数进行筛选，把主要的“加工工艺参数”作为案例的解，从而构成每个不同的“案例”；最后由大量的“案例”构成“案例库”，这些信息中的关键字段组成案例库结构。

案例库是由案例组成的，恰当的案例表示能够反映事物的本质特征，并且使案例检索系统能够迅速地从案例库中检索出所要的案例，从而提高案例检索效率^[91]。案例表示是指如何抽取待解决问题的特征以及特征之间的关系，并且存入案例库中，为推理提供基础条件。对于不同的问题，其表示也是各不相同，**一个合理的案例表示不仅会使问题变得容易求解，而且还有较高的求解效率**。在构建案例库时，由影响系统案例解的各种因素构成了案例的特征集合，筛选集合中的特征是决定系统性能的关键因素^[92]。

为了便于案例查询和存储，在表征案例时需要具有合理的组织结构。一般来说，一个典型的案例可以结构化为“案例的特征”和“案例的解”两部分，同时案例库中的案例应尽可能多地覆盖问题空间^[93,94]。为了更全面地表征案例，提高检索速度和检索准确度，本文采用实体联系模型（Entity-Relationship Model, E-R 模型）描述滚磨光整加工中的实体类型、属性及其联系。通过进一步分析加工工艺过程，根据上一节讨论的生产实际、加工实验报告以及文献手册中的相关数据，结合大量生产实例中的信息描述，选择影响滚磨光整加工工艺方案选择的因素主要有：被加工零件的特征、零件加工前特征以及用户给出的加工要求。图 3-1 为滚磨光整加工工艺方案优选的 **E-R 模型**。

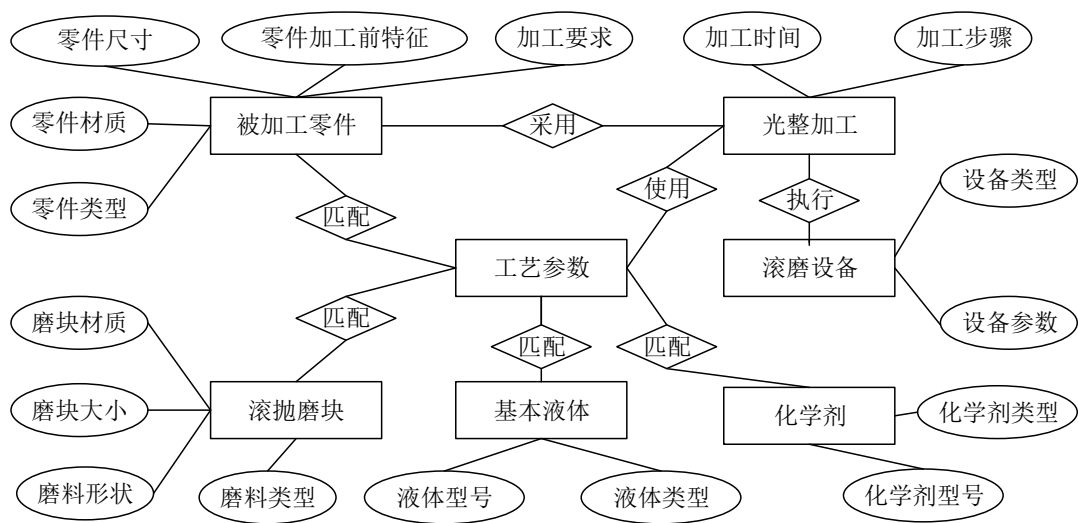


图 3-1 滚磨光整加工工艺方案优选的 E-R 模型

Figure 3-1 E-R model of the barrel finishing technology optimization

通过 E-R 图分析可知，滚磨光整加工工艺优选的案例包含：被加工零件的主要特征及其加工前状态、加工要求，以及相应的加工工艺参数。

(1) 被加工零件特征是指用以描述零件特征参数，主要有：零件的类型、零件的材质、零件的尺寸，等；

(2) 零件加工前特征是指零件未加工前的指标参数，主要有：零件表面粗糙度、毛刺、光亮度和硬度等，特殊情况时，还包括加工零件的前道工序；

(3) 用户加工要求是针对被加工零件加工后在某些特征上的具体指标要求，主要有：加工后零件的表面粗糙度、零件是否有毛刺、零件的光亮度、硬度、残余应力改善，等；

(4) 滚磨光整加工工艺参数主要有：滚磨光整加工的设备（设备的类型、设备的参数）、滚抛磨块（磨块材质、磨块大小、磨块形状、磨料类型）、磨液（包括：基本液体（液体型号、液体类型）和化学剂（化学剂类型、化学剂型号））、加工参数（主要有加工时间和加工步骤）等。

因此，滚磨光整加工工艺优选的案例特征主要包括被加工零件特征（包括零件描述、加工前特征）和用户提出的加工要求，而把滚磨光整加工工艺参数作为案例的解。图 3-2 为滚磨光整加工工艺优选案例的组成和结构。

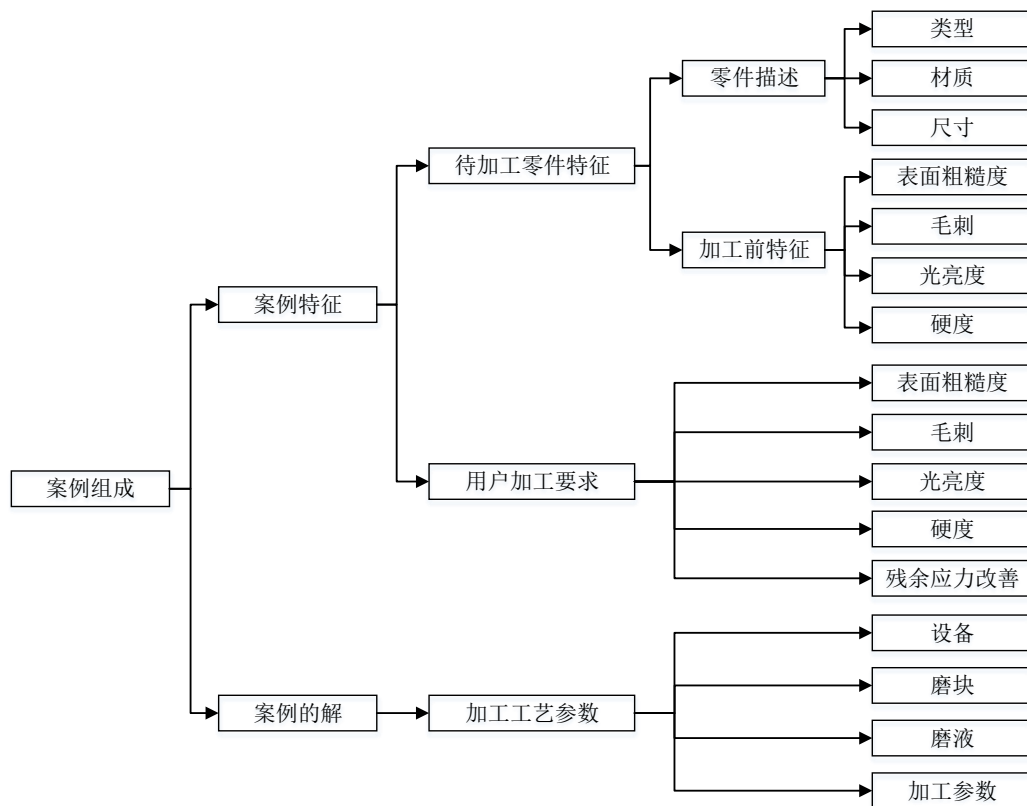


图 3-2 滚磨光整加工工艺优选案例的组成和结构

Figure 3-2 Case composition and structure of the barrel finishing technology optimization

一个合理的案例特征表示不仅会使问题变得容易求解，而且还有较高的求解效率。

由于不同类型零件的尺寸参数不同，且对于不同的加工问题，其案例特征的选择也各不相同，所以在案例构建时，需要按照零件类型建立案例，组成不同零件的案例库。

3.3 基于模糊 C 均值聚类算法的工艺案例库优化

随着滚磨光整加工生产的不断进行，越来越多的合格案例被保存到案例库中，一方面案例库中积累的知识和信息越来越全面，有助于采用案例推理方法更快更准地找到解决新问题的加工工艺方案；另一方面，如果不采取有效措施，不断处理更新案例，则案例库中可能会出现大量的相似度极高的案例，称为冗余案例，势必会降低案例检索的效率^[95]。因此，需要采取有效措施优化案例库，使案例库中案例的数目保持在合理的规模，以降低工艺方案的搜索空间，提高案例推理方法优选工艺的速度和准确性。

目前已有一些有关案例库优化方法，主要是通过精简案例库来减小案例库的容量，以提高案例检索的效率，并节省案例的存储空间。典型的有：（1）随机删除法^[96]：当案例库中的案例数超过规定的最大值后，随机删除一个案例。这是一种比较简单的案例删除方法；（2）案例效用度量法（Efficiency）^[97, 98]：根据计算案例使用的时间收益，来判断是否为无用案例，如果是，则删除该案例；（3）基于案例分类的删除策略^[99]：首先采用相似粗糙集进行案例属性约简，然后对案例库中的案例进行分类，最后确定删除策略；（4）基于孤立点的案例库优化^[100]：首先对案例库进行孤立点分析，找出不一致的案例，再进行删除；（5）模糊 C 均值（FCM）法^[101]：由于 FCM 算法的理论比较成熟完善，得到了广泛关注和应用。叶清等人^[102]采用主成分分析与 FCM 算法相结合，删除孤立数据，对入侵检测样本数据进行压缩，提高了入侵检测分类器的运算速率。徐瑾等人^[103]在建立径流预测模型时，为了提高径流预测精度，而减少模型的训练时间，采用 FCM 对案例库中的数据进行聚类，删除聚类错误的案例，精简了案例库，为小波神经网络选择了合理的训练样本。

3.3.1 基于减法聚类的模糊 C-均值聚类算法改进

在进行案例库优化时，需要选用合适的方法。由于在本文已经构建的滚磨光整加工工艺案例库中的案例都是经过检验和评价的合格案例，因此，不能采用随机删除法和孤立案例删除法，否则可能会导致被删除的案例是一个关键的案例，这样不仅会降低案例推理技术解决问题的能力，甚至可能使某个新问题永远无法得到解决。因此，本文提出的采用模糊 C 均值（FCM）聚类算法进行案例库优化方法的思路是：首先通过 FCM 聚类算法找到案例库中的孤立案例，即那些合格的特殊零件的案例数据，应该予以保留。然后将同类中的案例进行相似度计算，删除相似度很高的冗余案例，不仅可以节省案例库中案例的存储空间，并且可以提高案例检索效率。

(1) 基于 FCM 进行案例库中案例聚类的基本原理

根据案例数据之间的相似性可以把案例库中的案例聚类成为不同的类，其结果是相似性高的案例数据被聚在同一类中，而相似性低的案例数据分在不同的类中，这个过程称为聚类。由于加工工艺案例库中的案例数据具有一定的模糊性，无法使用传统的聚类算法，依据严格的划分界限将其归属于某一类，而应该通过隶属度进行划分。FCM 算法是一种基本的模糊聚类方法，由隶属度确定样本点的类属，因此，FCM 算法是一种对聚类对象的软划分。

基于 FCM 进行光整加工案例库中案例进行聚类的基本原理如下：

定义：案例库 $S=\{s_1, s_2, \dots, s_M\}$ 中有 M 个案例，其中 $s_i \in \mathfrak{R}^n$ 为滚磨光整加工工艺案例库中的第 i 个案例； i 代表案例序号， M 为案例的个数， $1 \leq i \leq M$ ； s_{ij} 为 s_i 的第 j 个特征， n 代表每个案例的特征个数， $1 \leq j \leq n$ ； c 为聚类数，即将案例库 S 划分为 c 类， $S=\{S_1, S_2, \dots, S_c\}$ ，其中 $2 \leq c \leq M$ 。假设聚类 S_i 非空，且满足：并集为全集 S ，交集为空集，即： $S=S_1 \cup S_2 \cup \dots \cup S_c$ ， $S_i \cap S_j = \emptyset$ ，其中 $i \neq j$ ； c 个聚类中心为 $v = \{v_1, v_2, \dots, v_c\}$ 。

FCM 的聚类准则是将目标函数最小化。目标函数的计算公式为：

$$J(S, R, v_1, v_2, \dots, v_k, \dots, v_c) = \sum_{k=1}^c \sum_{i=1}^M r_{ki}^{m_0} d_{ki}^2 \quad (3-1)$$

目标函数的约束条件为：

$$\begin{cases} \sum_{k=1}^c r_{ki} = 1, & 1 \leq i \leq M \\ r_{ki} \in [0, 1], & 1 \leq k \leq c, 1 \leq i \leq M \\ \sum_{i=1}^M r_{ki} \in [0, M], & 1 \leq k \leq c \end{cases} \quad (3-2)$$

式中， r_{ki} 为隶属度，即：第 i 个案例属于第 k 个类的程度； R 表示由 r_{ki} 构成的隶属度矩阵； $d_{ki} = \|v_k - s_i\|$ 表示第 k 个聚类中心 v_k 与第 i 个案例之间的欧氏距离； m_0 为模糊参数，代表模糊水平， $1 \leq m_0 < \infty$ 。当 $m_0=1$ 为硬分类， m_0 越大，模糊程度越大。通过大量的实验，当 $1.5 \leq m_0 \leq 3$ 有较好的分类效果^[104]。

通常采用拉格朗日法求解目标函数的最小值，其拉格朗日函数 F 为：

$$F = \sum_{k=1}^c r_{ki}^{m_0} d_{ki}^2 + \lambda \left(1 - \sum_{k=1}^c r_{ki} \right) \quad (3-3)$$

式中， λ 为拉格朗日乘子。

1) 根据目标函数最小值的条件，分别对 λ 和 r_{ki} 求一阶偏导：

$$\frac{\partial F}{\partial \lambda} = \left(1 - \sum_{k=1}^c r_{ki} \right) = 0 \quad (3-4)$$

$$\frac{\partial F}{\partial r_{ki}} = m_0 \times (r_{ki})^{m_0-1} \times d_{ki}^2 - \lambda = 0 \quad (3-5)$$

当 $d_{ki} = 0$ 时, $r_{ki} = 1$, $r_{ji} = 0 (k \neq j)$;

当 $d_{ki} \neq 0$ 时, 根据式 (3-5) 解得:

$$r_{ki} = \left[\frac{\lambda}{m_0 \times (d_{ki})^2} \right]^{\frac{1}{m_0-1}} \quad (3-6)$$

将式 (3-6) 代入式 (3-4), 得出:

$$\sum_{j=1}^c r_{ji} = \sum_{j=1}^c \left(\frac{\lambda}{m_0} \right)^{\frac{1}{m_0-1}} \left[\frac{1}{(d_{ji})^2} \right]^{\frac{1}{m_0-1}} = 1 \quad (3-7)$$

故:

$$\left(\frac{\lambda}{m_0} \right)^{\frac{1}{m_0-1}} = \left[\sum_{j=1}^c \left(\frac{1}{d_{ji}} \right)^{\frac{1}{m_0-1}} \right]^{-1} \quad (3-8)$$

将式 (3-8) 代入式 (3-6), 得出:

$$r_{ki} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{ki}}{d_{ji}} \right)^{\frac{2}{m_0-1}}} \quad (3-9)$$

2) 根据目标函数最小值的条件, 对 v_k 求一阶偏导, 得到:

$$v_k = \frac{\sum_{i=1}^M (r_{ki})^{m_0} s_i}{\sum_{i=1}^M (r_{ki})^{m_0}} \quad (3-10)$$

因此, FCM 聚类算法的步骤为:

Step 1: 设定: 初始的聚类数 c 、初始的聚类中心 V_0 、最大迭代次数为 T_{\max} (迭代次数 $t=1, 2, \dots, T_{\max}$)、收敛精度 ε ; 选取模糊参数 m_0 ;

Step 2: 通过式 (3-9) 得到初始的隶属度矩阵 R_0 、通过式 (3-1) 得到初始的目标函数 J_0 ;

Step 3: 根据式 (3-9) 对隶属度矩阵 R_t 进行更新;

Step 4: 根据式 (3-10) 对聚类中心 V_t 进行更新;

Step 5: 根据式 (3-1) 对目标函数 J_t 不断地更新, 如果 J_t 的变化量小于收敛精度 ε 或者迭代次数大于最大迭代次数 T_{\max} , 则聚类结束, 并输出隶属度矩阵 R 和聚类中心矩阵 V ; 否则, 转到 Step 3。

(2) 基于减法聚类算法的 FCM 参数初始值确定

根据上述的 FCM 算法的步骤可知: 在对加工工艺案例库中的案例数据集进行聚类时, 需要首先给出合适的初始聚类数和初始聚类中心, 才能获得正确的聚类结果。也就是说 FCM 算法对初始的聚类数和初始的聚类中心非常敏感, 其初始值的选择不仅会影响到聚类的质量, 而且还会影响到聚类的稳定性, 需要首先采用一定的方法来确定 FCM 算法的初始聚类数和初始聚类中心。

减法聚类 (Subtractive Clustering) 可以计算案例的密度指标, 来快速自动确定出聚类数和聚类中心^[105], 可以为 FCM 确定较为合理的初始值, 从而提高 FCM 的聚类质量和聚类稳定性。

通常, 采用减法聚类的步骤为:

Step 1: 计算案例库中每一个案例的密度指标, 可用公式 (3-11) 计算:

$$D_i^1 = \sum_{j=1}^M \exp \left[\frac{-\|s_i - s_j\|^2}{(0.5r_a)^2} \right], \quad i = 1, 2, \dots, M \quad (3-11)$$

式中, D_i^1 表示案例 s_i 的密度指标; r_a 表示邻域半径, 可用式 (3-12) 计算:

$$r_a = \frac{1}{2} \min_j \left\{ \max_i \left\{ \|s_i - s_j\| \right\} \right\} \quad (3-12)$$

计算得到各案例的密度指标后, 选择其中最大的密度指标, 即 $D_c^1 = \max_i \{D_i^1\}$, 其对应的案例 s_c^1 作为第一个聚类中心。

Step 2: 由于根据式 (3-11) 计算每个案例的密度指标时, 求出的最大密度指标的案例附近可能会有许多邻域案例数据, 虽然已经选择密度指标值最大的案例作为第一个聚类中心, 但是为了避免将第一聚类中心附近的案例被选择为另一个聚类中心, 需要从每个案例中减去与离第一聚类中心的距离成正比的密度指标。当案例指标降低时, 选择剩余密度指标最高的案例作为第二个聚类中心, 然后根据它们到第二个聚类中心的距离, 进一步降低每个案例的密度指标^[106]。因此, 在第 k 个聚类中心和对应的密度指标下, 为了削弱寻找出的聚类中心的影响, 修正寻找每个案例的案例密度指标的公式如 (3-13):

$$D_i^{k+1} = D_i^k - D_c^k \exp \left[\frac{-\|s_i - s_c^k\|^2}{(0.5r_b)^2} \right], \quad i = 1, 2, \dots, M \quad (3-13)$$

式中, s_c^k 表示第 k 个聚类中心; D_c^k 表示第 k 个聚类中心对应的密度指标; r_b 表示邻域半径, 通常可以取 $r_b = 1.2 r_a$ ^[107]。

Step 3: 将根据式 (3-13) 修正后的各案例的密度指标中, 取最大值所对应的案例作为第 $k+1$ 个聚类中心 s_c^{k+1} , D_c^{k+1} 为其对应的密度指标。

Step 4: 判断 $D_c^{k+1}/D_c^1 < \delta$ ($0 < \delta < 1$) 是否成立, 如果成立, 减法聚类的过程结束; 否则转到 Step 2。

(3) 基于减法聚类的 FCM 算法改进

由于 FCM 算法在应用的聚类效果对聚类数和聚类中心的初始值非常敏感, 故采用减法聚类来改进 FCM 算法, 通过减法聚类给出最大的聚类数和聚类中心的初始值, 然后计算聚类有效性函数最终确定最佳的聚类数, 提出了基于减法聚类改进的 FCM 算法, 其步骤如下:

Step 1: 首先设定最大迭代次数 T_{\max} 和收敛精度 ε 的值, 并设定模糊指数 $m_0=2$ 。

Step 2: 采用离差归一化方法对案例数据进行归一化, 可用式 (3-14) 计算:

$$s'_{ij} = \frac{s_{ij} - \min_j \{s_{ij}\}}{\max_j \{s_{ij}\} - \min_j \{s_{ij}\}} \quad (3-14)$$

式中, s_{ij} 表示第 i 个案例的第 j 个案例特征; $\max_j \{s_{ij}\}$ 和 $\min_j \{s_{ij}\}$ 分别表示第 j 个案例特征的最大值与最小值。

Step 3: 采用减法聚类算法对处理后的加工案例进行聚类, 得到最大聚类数 c_{\max} 和聚类中心 V^* , 并分别设定为 FCM 算法的初始聚类数 $c=c_{\max}$ 、初始聚类中心 $V_0=V^*$ 。

Step 4: 设定迭代次数 $t=1$ 。由式 (3-9) 和式 (3-1) 分别计算初始隶属矩阵 R_0 和初始目标函数 J_0 。

Step 5: 由式 (3-9) 和式 (3-10) 分别计算第 t 次迭代得到的隶属度矩阵 R_t 和聚类中心 V_t , 然后将 V_t 和 R_t 代入式 (3-1) 计算 FCM 的目标函数 J_t 。

Step 6: 判断 $|J_t - J_{t-1}| \leq \varepsilon$ 或 $t > T_{\max}$ 是否成立, 假如成立, 即得到隶属度矩阵 R 和聚类中心 V , 并转到 Step 7 执行; 假如不成立, 令 $t=t+1$, 转回到 Step 5。

Step 7: 在 Step 6 中得到隶属度矩阵 R 和聚类中心 V , 然后求出不同聚类数下的聚类有效性函数 $V_{xie}(R, V, c)$, 可用式 (3-15) 计算^[108]:

$$V_{xie}(R, V, c) = \frac{\frac{1}{M} \sum_{k=1}^c \sum_{i=1}^M r_{ki}^{m_0} \|v_k - s_i\|^2}{\min_{k \neq i} \|v_k - v_i\|^2} \quad (3-15)$$

由式 (3-15) 可知: $V_{xie}(R, V, c)$ 的值越小, 可以表明聚类的效果较好, 因为此时各个聚类集的内部比较紧凑, 且各聚类之间也比较独立。所以选择聚类有效性函数值最小值对应的聚类数作为最终的聚类数。

Step 8: 设 V^* 的前 c 个聚类中心作为 V_0 , 且 $c=c-1$, 判断 $c > 1$ 是否成立; 若成立,

则 $t=1$ ，转回到 Step 4；反之，跳转到 Step 9。

Step 9: 输出 $V_{xie}(R, V, c)$ 为最小时的聚类数 c 、聚类中心矩阵 V 和隶属度矩阵 R 。

图 3-3 为基于减法聚类改进的模糊 C 均值聚类算法（S-FCM）步骤框图。

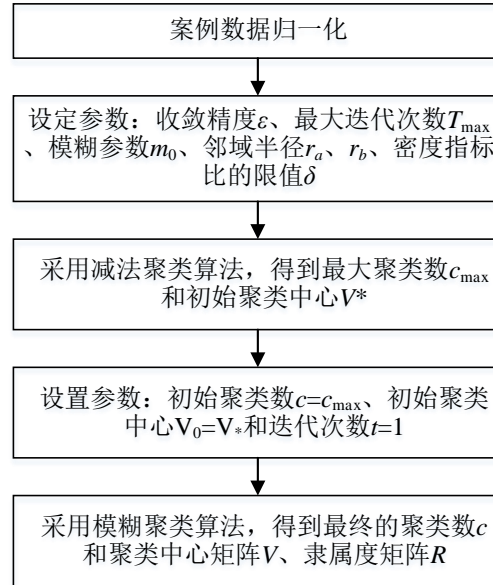


图 3-3 基于减法聚类改进的 FCM 聚类算法的步骤框图

Figure 3-3 The procedure diagram of improved FCM algorithm based on subtractive clustering

3.3.2 基于减法聚类的 FCM 的工艺案例库优化

实施案例推理优选滚磨光整加工工艺的前提是构建并且能够不断完善优化加工工艺案例库。为了解决案例不断增加后大量冗余案例的出现案例推理系统的检索效率的问题，本文提出采用减法聚类的 FCM（S-FCM）算法来解决案例库中案例的冗余问题，以提高案例检索的效率。

首先采用 S-FCM 算法对案例数据进行处理，计算得到案例库中每个案例属于不同聚类集的隶属度，则每个案例将被归属于隶属度最大的那个聚类集之中。在此引入隶属度阈值 u_{th} （表示某个案例属于不同聚类隶属度的下限），当某一案例属于每一类的隶属度都小于隶属度阈值时，将之定义为孤立案例，即特殊案例。由于本文所构建的滚磨光整加工工艺优选案例库中存储的案例都是满足实际加工要求的合格案例，所以应该保留这些特殊案例。然后将每类聚类集中的案例分别计算两两相似度。引入相似度阈值的概念，表示两个案例互为冗余案例，则可消除“案例对”中的一个案例。将其余的案例以及特殊案例组成新的案例库，完成案例库的优化过程。

图 3-4 为基于 S-FCM 的案例库优化的实施步骤，采用 S-FCM 进行滚磨光整加工工

艺案例库优化的步骤为：

Step 1：采用减法聚类算法确定 FCM 算法的聚类上限 c_{\max} 和初始聚类中心 $V_0=V^*$ ；

Step 2：采用 FCM 聚类算法对案例库进行聚类，计算不同聚类数下聚类有效性函数的值，最终聚类数即为最小值所对应的聚类数，并计算每个案例属于不同类别的隶属度。

Step 3：将属于每一类的隶属度都小于隶属度阈值 u_{th} 的案例组成特殊案例集 B_1 。

Step 4：将每一聚类中的案例计算其两两之间的相似度，将大于相似度阈值 S_{th}^* 的“案例对”中的一个案例，存入“冗余案例集”中，以备后续修正案例时再用；而将留下的案例组成案例集 B_2 。

Step 5：将 B_1 和 B_2 组成新的案例库，即为优化后的案例库。

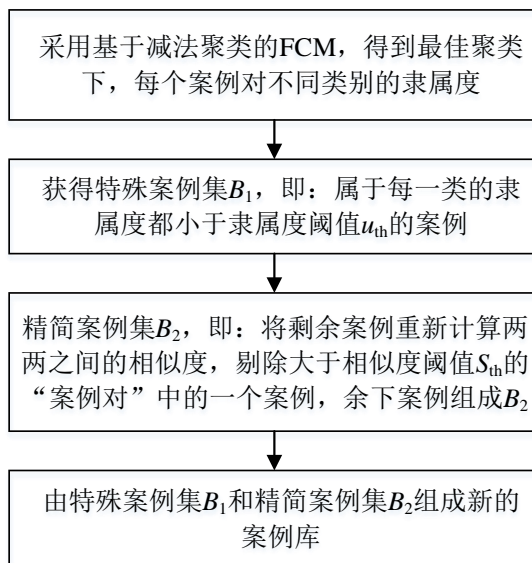


图 3-4 基于 S-FCM 的案例库优化实施步骤

Figure 3-4 The implementing procedure of the optimization of the case-base based on S-FCM

3.3.3 仿真研究

采用滚磨光整加工数据库平台上的大量成功案例进行深入的仿真研究，首先讨论减法聚类的结束条件 δ 的影响和选取，然后对比 S-FCM 和 FCM 的仿真结果，最后给出基于 S-FCM 的案例库优化结果。由于篇幅所限，下面仅给出案例库中现有的材质为钢的轴类零件的仿真结果。

（1）减法聚类算法中结束条件 δ 的影响和选取

由于参数 δ 的取值会影响减法聚类最终的聚类数，故首先对不同 δ 的值进行仿真研究。通过分析研究数据库中各案例的数据，选择案例特征有 11 个，包括：①轴类零件尺寸参数：长度和轴径；②零件加工前特征参数：零件表面粗糙度、毛刺、光亮度、硬度；③加工要求，即加工后的特征参数：零件表面粗糙度、毛刺、光亮度、硬度、残余应力

改善。

设定参数: $m_0=2$, $\varepsilon=10^{-5}$, $T_{\max}=1000$ 。由式(3-12)计算得到 $r_a=0.7923$, $r_b=1.2*r_a=0.9508$ 。

表 3.1 为减法聚类中的不同参数 δ 取值对应的聚类数。

表 3.1 减法聚类中的 δ 所对应的聚类数

Tab. 3.1 The corresponding clustering numbers to δ									
δ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
聚类数	13	7	5	5	3	3	3	3	2

FCM 算法中的初始聚类数和聚类中心选用由减法聚类得到的不同的聚类数和对应的聚类中心, 然后对案例库中的数据进行聚类分析。表 3.2 为计算得到的聚类数与对应的聚类有效性函数值, 图 3-5 为聚类数与聚类有效性函数值的关系曲线。

表 3.2 聚类数与聚类有效性函数值

Tab. 3.2 Clustering number and the values of clustering validity function						
聚类数	2	3	4	5	6	7
$V_{xie}(R, V_c)$	0.5268	0.2680	0.1806	1.9561	1.8766	1.6166
聚类数	8	9	10	11	12	13
$V_{xie}(R, V_c)$	3.0747	1.8225	1.2380	0.8659	0.8329	21537604.2634

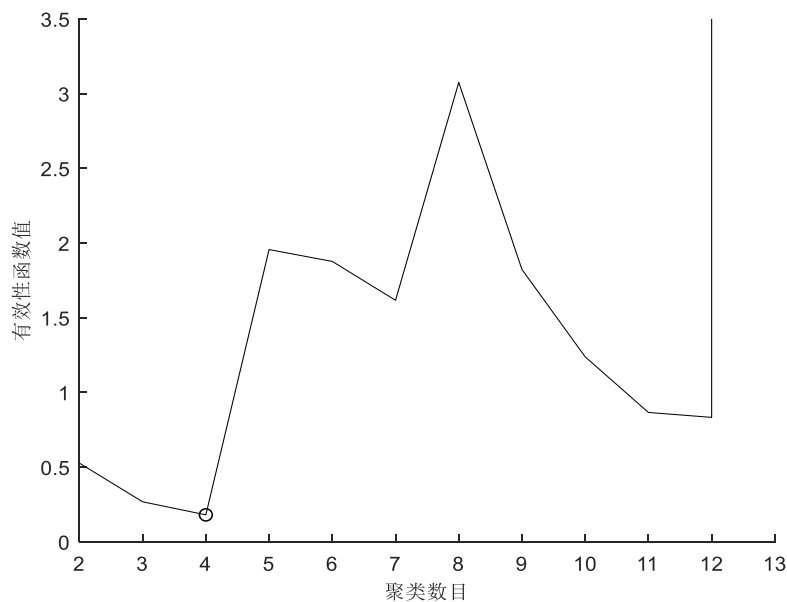


图 3-5 聚类数与聚类有效性函数值的关系曲线

Figure 3-5 The relation between the clustering number and the values of validity function

根据表 3.1 和表 3.2 可知, 当 $\delta=0.1$ 时, 对应的最大聚类数为 13, 此时聚类数较多, 且对应的聚类有效性函数值极大, 不能达到很好的聚类效果; 随着参数 δ 增大, 得到的最大聚类数减少; 当 δ 增大到一定范围内, 最大聚类数基本保持不变; 当参数 δ 太大时, 最大聚类数太小, 从而无法确定 FCM 算法的最佳聚类数。

由图 3-5 可知, 当聚类数为 13 时, 由于聚类数目过多, 导致原本属于一类的案例被分成两类甚至更多, 此时某两个聚类中心接近于重合, 即聚类有效性函数值 $V_{xie}(R, V, c)$ 公式 (3-15) 的分母极小, 因此此时的聚类有效性函数值极大; 当 FCM 聚类算法的聚类数为 4 时, 聚类有效性函数 $V_{xie}(R, V, c)$ 的值最小, 即案例聚为 4 类最为合理。通常, 将参数 δ 设置为 $\delta \in [0.2, 0.4]$, 可以得到最为合理的聚类数目。

当取 $\delta=0.2$ 时, 聚类数 $c_{\max}=7$, 其聚类中心如表 3.3 所示。

表 3.3 减法聚类中心
Tab. 3.3 The subtractive clustering centers

类号	对应的聚类中心
1	(0.0844, 0.0427, 0.0141, 0.5029, 1.0, 0.1471, 0.0072, 1.0, 1.0, 0.2941, 0.4307)
2	(0.2403, 0.0569, 0.0317, 0.4277, 0.0, 0.0588, 0.0195, 1.0, 0.5, 0.2941, 0.4307)
3	(0.0557, 0.1090, 0.0008, 0.5318, 1.0, 0.3529, 0.0006, 0.0, 1.0, 0.4706, 0.4901)
4	(0.1331, 0.0758, 0.0767, 0.4393, 0.0, 0.3824, 0.0275, 0.0, 0.5, 0.3824, 0.6089)
5	(0.1017, 0.1090, 0.0376, 0.5029, 0.0, 0.2059, 0.0341, 1.0, 1.0, 0.3235, 0.3762)
6	(0.3193, 0.1706, 0.0163, 0.5064, 1.0, 0.0882, 0.0070, 0.0, 1.0, 0.2059, 0.3564)
7	(0.5087, 0.1706, 0.0390, 0.4277, 1.0, 0.0588, 0.0221, 1.0, 1.0, 0.2647, 0.1188)

(2) S-FCM 和 FCM 的比较

当 FCM 聚类算法的聚类数设置为 4 时, 设置 $m_0=2$, $\varepsilon=10^{-5}$, $T_{\max}=1000$, 图 3-6 为通过减法聚类优化后的 FCM 聚类算法 (S-FCM) 和 FCM 的目标函数值的变化。

由图 3-6 可知, 采用 S-FCM 方法由于确定出初始聚类中心, 收敛速度较快, 在第 6 次迭代时, 已达到聚类结束条件 “目标函数变化值 < 0.00001 ”; 而采用 FCM 聚类算法具有较大振幅, 收敛速度较慢, 在第 18 次迭代时达到聚类结束条件, 收敛迭代次数更多, 因此, S-FCM 方法的收敛性更好。

(3) 精简案例库

当聚类数为 4 时, 表 3.4 为由 S-FCM 聚类算法获得 4 类最终的聚类中心, 表 3.5 为部分案例属于每一类的隶属度。

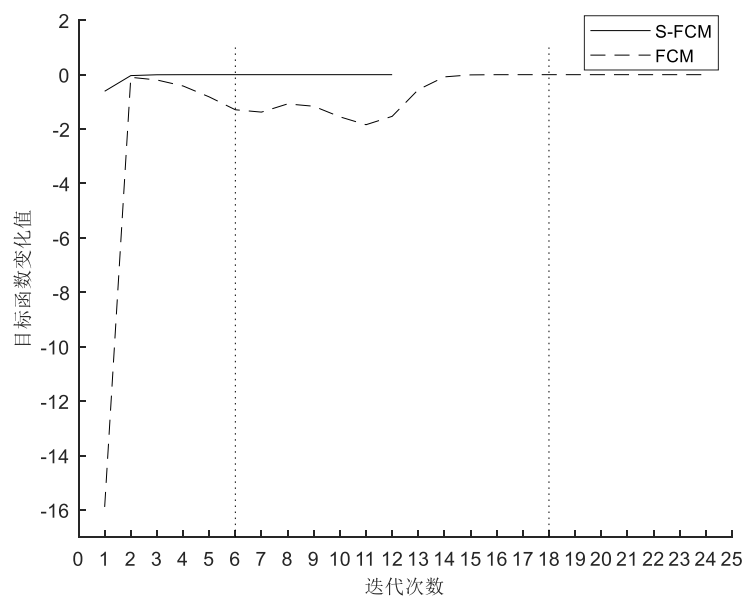


图 3-6 目标函数值变化

Figure 3-6 Objective function changing

表 3.4 S-FCM 的聚类中心

Tab. 3.4 The clustering centers from S-FCM algorithm

类号	对应的聚类中心
1	(0.2023, 0.1100, 0.0494, 0.4653, 0.9601, 0.1659, 0.0321, 0.9700, 0.9791, 0.3138, 0.4074)
2	(0.2561, 0.0995, 0.0491, 0.4579, 0.0253, 0.1904, 0.0292, 0.9705, 0.6367, 0.3343, 0.3850)
3	(0.2290, 0.0979, 0.0325, 0.4588, 0.9494, 0.2852, 0.0173, 0.0501, 0.9325, 0.3878, 0.4450)
4	(0.2612, 0.1143, 0.0519, 0.4676, 0.0500, 0.3569, 0.0254, 0.0748, 0.6278, 0.4185, 0.4597)

表 3.5 部分案例的隶属度

Tab. 3.5 Membership of some cases

案例编号	类别 1 隶属度	类别 2 隶属度	类别 3 隶属度	类别 4 隶属度
1	0.1638	0.0878	0.5992	0.1492
2	0.1316	0.6917	0.0662	0.1104
3	0.0708	0.1444	0.1239	0.6609
4	0.0444	0.0934	0.0824	0.7798
5	0.0729	0.1512	0.127	0.6489
6	0.0663	0.1346	0.1175	0.6816
7	0.0659	0.1375	0.1162	0.6804
8	0.0778	0.1535	0.1364	0.6323
9	0.674	0.1163	0.1376	0.0721
10	0.6949	0.1089	0.1288	0.0674

表 3.6 为每个聚类集中案例的数量，表中， D_1 、 D_2 、 D_3 、 D_4 分别表示 4 个聚类集。

表 3.6 聚类集中案例的数
Tab. 3.6 The case number of each cluster

D_1	D_2	D_3	D_4
80	97	66	58

在精简案例库时，首先根据隶属度阈值 u_{th} 找出特殊案例，然后在剩余案例中，重新计算两两案例之间的相似度，剔除大于某个相似度阈值 S_{th}^* 的案例的“案例对”中的一个案例，最后将剩余案例与特殊案例一起构成新的案例库。

表 3.7 为不同的隶属度阈值 u_{th} 及相似度阈值 S_{th}^* 的案例库优化结果。

表 3.7 案例库优化结果
Tab. 3.7 Case base optimization results

相似度阈值	精简前	精简后		
		$u_{th}=0.4$	$u_{th}=0.6$	$u_{th}=0.8$
$S_{th}^*=0.8$	301	82	155	277
$S_{th}^*=0.85$	301	136	195	285
$S_{th}^*=0.9$	301	198	246	290
$S_{th}^*=0.95$	301	260	272	293

由表 3.7 可知，案例库的优化结果与隶属度阈值 u_{th} 及相似度阈值 S_{th}^* 的取值大小有关：

- ① u_{th} 取值越大，则特殊案例的数量就越多，即被保护的案例就越多，使得案例库优化后仍会有更多的案例；反之，优化后的案例库中案例的数量就会减少。
- ② S_{th}^* 取值越大，则超过阈值的案例就越少，即被删减的案例就越少，使得优化后的案例库中的案例就越多；反之，被删减的案例就越多，则优化后案例库中的案例就越少。

因此，需要设定合适的隶属度阈值和合适的相似度阈值。经过对大量滚磨光整加工的生产实例的仿真研究，本文设定隶属度阈值为 0.6、相似度阈值为 0.9，可以保证在精简冗余案例的同时，特殊案例不被删除。

对案例库优化后，计算删除案例与对应聚类集中案例的相似度，其计算公式如下^[109]：

$$SIM(T, S) = \sum_{i=1}^n sim(T_i, S_i) \times w_i / \sum_{i=1}^n w_i \quad (3-16)$$

式中， $SIM(T, S)$ 为删除的案例 T 与对应聚类集中的案例 S 之间的总相似度； n 为案例特

征的个数； $sim(T_i, S_i)$ 为删除案例的第 i 个案例特征的值 T_i 与聚类集中对应案例特征的值 S_i 之间的相似度； w_i 为采用层次分析法确定的第 i 个案例特征的权重。

分别计算每一个被删除的案例与同一聚类集中所有剩余案例之间的相似度，表 3.8 为仿真结果。

表 3.8 删除案例的最大/最小相似度

Tab. 3.8 The maximum/minimum similarity of the deleted cases

隶属度阈值	相似度阈值	聚类数据集	最大相似度 S_{\max}	最小相似度 S_{\min}
$u_{th}=0.6$	$S_{th}^*=0.9$	D_1	1	0.305
		D_2	0.996	0.201
		D_3	1	0.115
		D_4	0.998	0.263

由表 3.8 可知，删除案例的最大相似度为 0.996 到 1 之间，最低相似度为 0.115 到 0.305 之间，表明该方法能够删除案例库中相似度很高的冗余案例，同时能够保留案例库中比较特殊的案例，从而在保证案例优选准确率的前提下，提高案例检索的效率。

在论文第四章进一步分别采用未优化前的案例库和优化后的案例库进行加权案例推理优选加工工艺参数，其结果更进一步佐证了本章提出的基于 S-FCM 的滚磨光整加工案例库优化算法的可行性和有效性。

此外，在课题组已经开发的滚磨光整加工数据库^[110]的基础上，将 S-FCM 应用于滚磨光整加工工艺案例库，对工艺案例库定期进行优化。在用户设置参数，选择零件类型和零件材质后，进行案例库优化后，可以在线查看优化过程参数和优化结果，实现了案例库的动态优化。

3.4 本章小结

为了使滚磨光整加工数据库中的大量生产实例能够在新工艺方案制定过程中发挥更大作用，首先深入剖析挖掘生产实例数据信息，抽取待加工零件的主要特征和加工要求的主要特征及其对应的滚磨光整加工工艺参数，形成案例，并存入案例库中，完成生产实例的案例化表征，为案例推理的应用提供知识和基础条件。

随着滚磨光整加工生产的不断进行，案例库中的案例将越来越多，如果不做定期的优化，可能会导致案例库中会出现大量的冗余案例，进而影响到案例推理系统的检索效率。为了提高滚磨光整加工工艺优选的效率，本章提出了采用减法聚类算法的模糊 C 均值聚类方法（S-FCM），对滚磨光整加工工艺优选案例库进行优化，研究结果表明：

- （1）采用减法聚类算法可以获得合适的 FCM 算法的初始聚类数和初始聚类中心，

然后计算聚类有效性函数，为 FCM 算法选择最终的聚类数，从而提高 FCM 的聚类质量。

（2）采用已研发的滚磨光整加工数据库平台上的合格案例进行大量的仿真研究，结果表明：所提出的方法可以合理有效地删除滚磨光整加工工艺案例库中的冗余案例，节省了案例的存储空间，提高了案例检索的效率，保证了知识库对生产线时效性需求的及时反馈，可以满足对生产现场的实时调控。该方法原理简单、步骤清晰，为实现滚磨光整加工工艺制定与批量生产过程中工艺参数的快速、准确优选提供保障。

