

基于多种模式识别的切削液辨别研究^{*}

李波,汪永超,李强,吴爱强
(四川大学机械工程学院,成都 610065)

摘要: 为了实现同时辨别不同种类的切削液并对其质量进行分析,在数据预处理基础上,提出一种基于向量夹角余弦法和神经网络的切削液辨别分析方法。首先,鉴于切削液的组分复杂,对其组分数据进行标准化及主成分降维处理;其次,利用夹角余弦法与神经网络分析得到未知样品与标准指纹图谱之间识别相似度并对比辨别切削液种类,并采用阈值算法对两种方法测得的结果进行一致性验证;最后,结合实际应用验证了该方法的科学性和可靠性,从而为切削液辨别分析提供有效的理论参考。

关键词: 切削液; 模式识别; 向量夹角余弦法; 神经网络; 阈值算法

中图分类号: TH16; TG506

文献标识码: A

Research on Cutting Fluid Identification Based on Multiple Pattern Recognition

LI Bo, WANG Yong-chao, LI Qiang, WU Ai-qiang

(School of Mechanical Engineering, Sichuan University, Chengdu 610065, China)

Abstract: To identify different types of cutting fluids at the same time and analyze their quality, based on the data preprocessing, a cutting fluid identification analysis method based on vector angle cosine method and neural network is proposed. First, because the components of the cutting fluid are complex, the component data is standardized and the principal component dimension reduction processing is performed. Then use the angle cosine method and neural network analysis to obtain the similarity between the unknown sample and the standard fingerprint, compare and identify the type of cutting fluid, and use the threshold algorithm to verify the consistency of the results measured by the two methods. Finally, the scientificity and reliability of the method were verified in combination with practical applications, so as to provide an effective theoretical reference for discriminating analysis of cutting fluids.

Key words: cutting fluid; pattern recognition; vector angle cosine method; neural network; threshold algorithm

0 引言

在机械加工中,使用性能优良的切削液是提高工件加工效率和加工精度、延长刀具使用寿命、减少刀具与工件摩擦的必要途径之一^[1]。故保证机械加工液的质量就显得尤为重要^[2]。如果切削液质量不佳,不仅对机械加工质量和效率造成严重影响,更对机加工的安全性造成威胁,甚至会阻碍产品的正常加工生产。由于零件的加工精度要求越来越高,切削液质量控制也变得越发重要。

传统切削液辨别方法主要有靠经验推定^[3](目测和嗅觉等)、浓度测定^[4](折光仪法、酸解破乳法和海明滴定法等)、pH测定^[5]、仪器检测^[6](色谱仪等),虽然这些方法在实际检测过程中各有千秋,但对人员和检测仪器等外界条件要求较高,精度难以保证。除此之外,研究人员也通过应用数学方法对切削液质量进

行辨别,如黄建等提出了一种以模糊数学和神经网络BP算法为基础的切削液选择方法^[7]。陶桂宝等利用模糊综合评价方法进行评价,并提出了用于剃齿加工的切削液综合选择模型^[8]。但以上研究方法必定会受到主观因素的影响,并且没有综合考虑切削液内部的组成成分。

机加切削液种类繁多,成分复杂,为了实现识别不同种类的切削液,就必须研究适合切削液的辨别分析方法。基于此,在前人的研究基础知识上,综合利用神经网络和夹角余弦法等模式识别方法进行对比校核,探索一套快速、准确的切削液辨别分析方法。

1 研究原理及方法

1.1 数据提取及标准指纹图谱建立

1.1.1 数据提取

切削液要进行识别与质量分析,首先需得到切削

收稿日期: 2020-01-18

^{*} 基金项目: 国家自然科学基金资助项目(51875370)

作者简介: 李波(1994—),男,四川巴中人,四川大学硕士研究生,研究方向为模式识别与系统开发、计算机辅助设计等。(E-mail) lib0420@163.com。

液的相关数据,在数据完整无误的情况下,再通过一系列的识别分析手段,得到我们想要的信息。

切削液的仪器检测以及数据提取方面,现行的各类分析仪器如色谱仪、质谱仪、气相色谱-质谱联用仪等都可用于切削液的检测中,只需要保证标准指纹图谱的建立与样品识别处于同一条件下即可。本文切削液检测仪器为气相色谱-质谱联用仪,切削液样品相关数据由四川某润滑剂公司提供。

1.1.2 标准指纹图谱建立

标准指纹图谱的建立是为了与切削液检测样品形成参照,是切削液精细化识别与质量分析的基础。建立标准指纹图谱时,应选取适当个数的切削液样品作为标准样本录入。

标准指纹图谱的建立一定程度上影响到后续样品精细化识别与质量分析的有效性。随着切削液识别与分析的应用增加,标准指纹图谱的数量与质量也会随之增加,标准指纹图谱库越丰富,识别分析的精度就越高,这对于后续切削液精细化识别与质量分析有着重要的意义。

1.2 数据处理方法

1.2.1 数据预处理

由于切削液的种类良莠不齐,组分复杂,其组分不乏有很多杂质成分的存在。杂质成分处理不当会影响切削液的辨别精度,因此有必要对样品数据中的杂质数据进行剔除。本文采用的是马氏距离判别法,下面对该方法进行简要介绍:

马氏距离是 Peter M Roth 用于计算两个未知样本集相似度的方法,它表示数据的协方差距离^[9]。它独立于测量尺度,与组分数据的量纲无关,还考虑到各种特性之间的联系,相比欧式距离它还可以排除变量之间相关性的影响^[10]。其计算步骤如下:

设有 m 个切削液样品,每个样品有 n 个组分,则构成了一个 $m \times n$ 的矩阵;

$$R_{m \times n} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$$

Step1: 对切削液样品数据进行标准化转换,得到:

$$X_{ij} = \frac{r_{ij} - \bar{r}_j}{s_j} \quad j = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (1)$$

$$\text{其中 } \bar{r}_j = \frac{\sum_{i=1}^m r_{ij}}{m}, s_j^2 = \frac{\sum_{i=1}^m (r_{ij} - \bar{r}_j)^2}{m-1}$$

从而可以得到标准化矩阵:

$$X_{m \times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

Step2: 将标准化矩阵 X 与其转置矩阵相乘,得到它的协方差矩阵 D :

$$D_x = \frac{1}{n} X \cdot X^T \quad (2)$$

Step3: 将标准化矩阵 X 的每列求平均值,得到均值向量 \bar{X}_j :

$$\bar{X}_j = \frac{\sum_{i=1}^m x_{ij}}{m} \quad j = 1, 2, \dots, n \quad (3)$$

Step4: 计算组分与总体 G 之间的马氏距离 $C^2(x, G)$:

$$C^2(x, G) = \sqrt{(x - \bar{x})^T X^{-1} (x - \bar{x})} \quad (4)$$

对于某个组分 x ,假设有待测样品总体 G_1 与标准样品总体 G_2 ,当 $C^2(x, G_1)$ 远大于 $C^2(x, G_2)$,则说明 $x \notin G_1$,该组分为杂质成分,作剔除处理。

1.2.2 主成分分析法降维处理

主成分分析法也称为主分量分析法,是一种通过数学统计分析方法将多个指标转变成少数几个可以反映绝大部分原始指标信息的综合指标的数据处理方法^[11]。它不仅可以降低数据空间的维数,还可以消除切削液组分数据之间的相互影响,使得在最大化保证数据不失真的情况下数据得到简化。故本文选用主成分分析法对切削液组分数据进行降维,从而满足切削液的识别分析要求,下面对该方法计算步骤进行简要介绍:

Step1: 通过式(1)得到的标准化矩阵 X 可以算出相关系数矩阵 C ,公式如下:

$$C_x = \frac{1}{m-1} X \cdot X^T \quad (5)$$

Step2: 根据特征值分解公式:

$$\begin{cases} C_x = S \Lambda S^T \\ C_y = P C_x P^T = P S \Lambda S^T P^T = P P^{-1} \Lambda P P^{-1} = \Lambda \end{cases} \quad (6)$$

通过以上两步计算出特征值矩阵 Λ ,并按照降序顺序排列得到特征值 λ_i 及对应特征向量 e_i 。

Step3: 计算各切削液组分的贡献率及累计贡献率:

$$l = \frac{\lambda_i}{\sum_{i=1}^n \lambda_i} \quad i = 1, 2, \dots, n \quad (7)$$

前 k 个主成分累计贡献率:

$$l_k = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} \quad (8)$$

在一般情况下,当这前 k 个切削液主成分的累计贡献率可以达到要求时,就可以选取这 k 个切削液主成分代替原来的组分作为新的变量参与后续运算,从而达到降维的目的。

2 切削液识别分析

切削液的主要识别与分析方法由夹角余弦法和神经网络方法组成,这两种方法各自有着其优缺点。通过对这两种方法进行合理的结合应用,可以同时反映出切削液样品的线性与非线性特征,通过两种方法的对比校核也能够使得识别与分析结果更具有可靠性。

2.1 向量夹角余弦法

向量夹角余弦法是一种相似性分析法,它可以通

过简单的计算评估两个多变量矢量之间的相似程度。在切削液的识别分析中,通过向量夹角余弦法可以对检测数据进行相似性判断^[12]。向量夹角余弦方法能够直接根据各待评价对象加权标准化特征值与理想目标之间的相似度,快速地多目标决策提供准确的判断。

设切削液组分理想数据集表示为 $Y_0 = (y_{01}, \dots, y_{02}, \dots, y_{0n})$, 切削液待测样品组分标准化数据集为 $X_i = (x_{i1}, \dots, x_{ii}, \dots, x_{in})$, $(i = 1, \dots, m)$, 其中 m 表示切削液组分个数。则夹角余弦法公式表示二者的相似度如下:

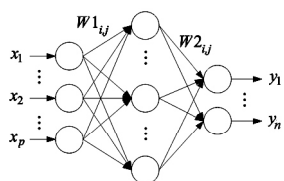
$$S_i = \cos \langle X_i, Y_0 \rangle = \frac{\sum_{j=1}^n x_{ij} y_{0j}}{\sqrt{\sum_{j=1}^n x_{ij}^2} \sqrt{\sum_{j=1}^n y_{0j}^2}} \quad (9)$$

S_i 越大, 相似度越好, 当 $S_i = 1$ 相似度最好。

2.2 神经网络算法

神经网络是一种模仿大脑神经网络信息处理方式的算法模型^[13]。神经网络模拟的是人脑的形象思维, 是一个由大量人工神经元构成的非线性系统。神经网络由多层神经元构成, 其中包括输入层、输出层以及隐含层, 每一层神经元的输出端连接下一层神经元的输入端, 经过每一层神经元的作用, 最后得到理想化的输出。

图 1 所示为典型的三层神经网络结构模型。



输入层节点 隐含层节点 输出层节点

图 1 人工神经网络结构模型

图中 X 为人工神经网络的输入, Y 为输出, $W1$ 、 $W2$ 为各层之间的连接权值。训练过程如图 2 所示。

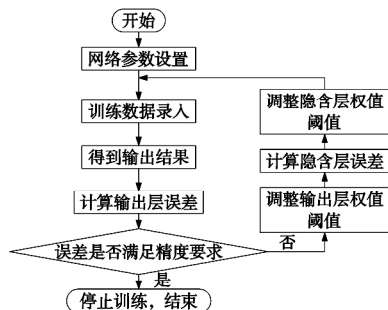


图 2 人工神经网络训练流程

BP 神经网络算法计算公式如下:

$$Z = \omega^T X + b \quad a = f(Z) \quad (10)$$

式中 ω 是权重 b 是偏置 X 为 n 维输入向量 Z 为神经元获得的信号加权 a 为神经元输出 f 为激活函数。以此得到目标函数为:

$$J(\omega, b) = \sum_{i=1}^N L(y^i, f(x^i | \omega, b)) + \frac{1}{2} \lambda \|W\|_F^2 \quad (11)$$

在最小化目标函数 $J(\omega, b)$ 的情况下, 采用梯度下降法, 得:

$$\omega_{ij}^{l-1} = \omega_{ij}^l - \alpha \cdot \frac{\partial J(\omega, b)}{\partial \omega_{ij}^l} \quad (12)$$

$$\frac{\partial L}{\partial \omega_{ij}^l} = \delta_i^l \cdot a_j^{l-1} \quad (13)$$

$$\delta_i^l = \frac{\partial a_i^l}{\partial Z_i^l} \cdot \frac{\partial Z_i^{l+1}}{\partial a_i^l} \cdot \frac{\partial L}{\partial Z_i^{l+1}} = \frac{\partial a_i^l}{\partial Z_i^l} \left(\sum_{j=1}^N \omega_{ji}^{l+1} \cdot \delta_j^{l+1} \right) \quad (14)$$

其中 l 为神经网络的层数 δ 是误差项, 通过以上公式求得每层神经元的误差调整量 δ , 继而逐层调整神经元权重及偏置向量, 完成训练。

由于识别样品众多, 且识别中容易产生系统误差干扰, 线性化的识别方法无法完全满足样品的识别要求, 应用神经网络正好弥补这个不足。

神经网络在切削液的具体数据分析中, 除了进行种类识别外, 对于样品的质量稳定性分析可通过样品数据的实际网络输出与理论输出之间的欧氏距离来判断。欧氏距离公式如下:

$$d = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (15)$$

对应的相似度公式表示为:

$$S = \frac{\sqrt{\sum_{k=1}^n y_k^2 - d}}{\sqrt{\sum_{k=1}^n y_k^2}} \quad (16)$$

式中 x 为实际网络输出 y 为理论网络输出, 其中理论网络输出的设定基于切削液样品的识别结果。

2.3 结果一致性验证

在实际情况下, 两种模式识别方法测得的相似度结果导向可能不完全一致, 故需要对两种方法测得的结果进行一致性验证。本文采用阈值算法, 其具体步骤如下:

(1) 提取同类型的切削液样本与该类型标准指纹图谱相似度的最大值 S_{\max} ;

(2) 计算同类型的切削液样本与该类型标准指纹图谱相似度的标准差 σ :

$$\sigma = \sqrt{\sum_{i=1}^k (S_i - \bar{S})^2 / k} \quad (17)$$

其中 S_i 为第 i 个切削液样本与标准指纹图谱的相似度值, \bar{S} 为各相似度的平均值, k 为切削液的样本数量。

(3) 计算相似度的阈值 S_t :

$$S_t = S_{\max} - 5\sigma \quad (18)$$

将待测切削液样本测试的相似度 S 与计算出的相似度阈值 S_t 进行比较, 若 $S > S_t$, 则说明该待测样本属于该类型切削液; 反之, 则不属于该类型切削液。如果两种算法的计算结果导向不一致, 则需要对其质量进行进一步的分析。

3 实例分析

3.1 数据处理

由于切削液的组分较多, 通过仪器测试所得的指

纹图谱数据维数也较大,而且所得数据相关性也很高,若直接进行识别分析,势必会造成识别分析效率低下,识别精度不高的问题,因此有必要对数据进行标准化和降维处理,使得数据变量较少,且数据相互独立。下面以切削液的检测数据为例,对其数据主成分降维过程进行说明。

实验室选用了 A、B、C 三种不同类型的切削液样品,每种类型的切削液分别再测试 10 个样品,通过气相色谱等仪器分析共得到 81 个组分数据并利用主成分降维分析得到前 10 个切削液主成分,前 10 个主成分的累计贡献率如表 1 所示。

表 1 切削液图谱数据主成分累计贡献率

主成分	累计贡献率%	主成分	累计贡献率%
P1	54.12	P6	87.61
P2	82.82	P7	88.25
P3	84.92	P8	89.11
P4	85.49	P9	89.54
P5	87.02	P10	90.15

由表 1 可知:前 10 个切削液主成分的累计贡献率就已经达到了 90% 以上,故这 10 个切削液主成分就包含了切削液数据中的绝大部分信息。因此,本文选择 10 个主成分对切削液进行识别分析。

3.2 识别分析

由第 2 小节可知,神经网络和夹角余弦法都可以计算得到待测样本与标准样本的类别相似度结果,从而可以判断待测样本所属类别,进而对切削液进行识别与质量分析。本次实验数据分析选用了 3 种不同类型的切削液,每种类型选取各 10 个检测样品进行测试,测试结果如表 2 所示。

表 2 神经网络与夹角余弦法的切削液对比分析结果

样品	各方法所得相似度		识别结果	样品	各方法所得相似度		识别结果
	神经网络	夹角余弦法			神经网络	夹角余弦法	
A 型样品							
A1	0.988	0.982	A 型	A2	0.993	0.992	A 型
A3	0.983	0.996	A 型	A4	0.988	0.991	A 型
A5	0.977	0.962	A 型	A6	0.989	0.984	A 型
A7	0.989	0.982	A 型	A8	0.987	0.992	A 型
A9	0.974	0.975	A 型	A10	0.984	0.982	A 型
B 型样品							
B1	0.968	0.967	B 型	B2	0.984	0.978	B 型
B3	0.984	0.991	B 型	B4	0.967	0.973	B 型
B5	0.975	0.982	B 型	B6	0.963	0.976	B 型
B7	0.971	0.964	B 型	B8	0.991	0.984	B 型
B9	0.964	0.975	B 型	B10	0.966	0.968	B 型
C 型样品							
C1	0.980	0.968	C 型	C2	0.981	0.965	C 型
C3	0.957	0.951	C 型	C4	0.968	0.958	C 型
C5	0.960	0.968	C 型	C6	0.975	0.959	C 型
C7	0.962	0.972	C 型	C8	0.982	0.957	C 型
C9	0.971	0.964	C 型	C10	0.974	0.941	C 型

合神经网络和夹角余弦法优缺点,得到的不同类型切削液的相似度分析结果如图 3、图 4 所示。

神经网络的主要优势有泛化能力强、具有自学习能力、非线性映射能力强等,其缺点在于计算量大、算法收敛速度慢、样本依赖性大等问题。在主成分分析模式下,神经网络算法计算量大和收敛速度慢的问题有很大改善。

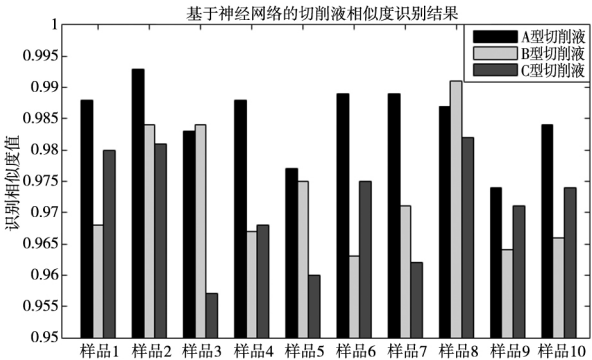


图 3 基于神经网络的切削液相似度识别结果

夹角余弦法的主要优势有计算量小、结果客观、对维度不敏感等,而其缺点是太线性化导致的小变量价值忽略。

神经网络法与夹角余弦法在样本识别与分析领域均有着广泛的应用,两者的组合应用可以同时反映出切削液样品的线性与非线性特征,通过两种方法的对比校核,能够使得识别结果更具有可靠性。

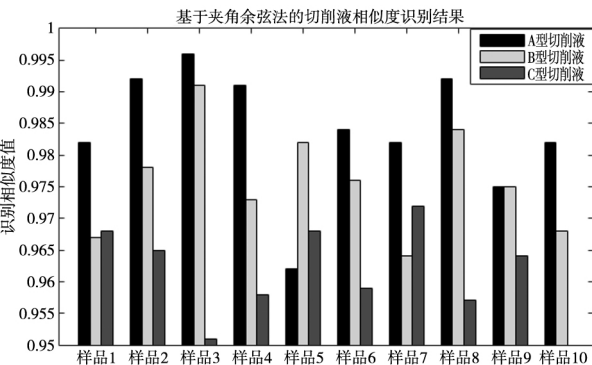


图 4 基于夹角余弦法的切削液相似度识别结果

通过不同模式识别方法,由图 3、图 4 可以判断切削液样本类型为 A 型的精确性和可靠性最高,样本类型为 C 型的最差。

3.3 一致性检验

以表 2 中 A 型切削液样本神经网络所测得的相似度结果为例,10 个样本中样本与标准指纹图谱之间相似度的最大值为 0.993,其相似度的标准差为 0.0055,从而得到相似度阈值:

$$S_t = S_{\max} - 5\sigma = 0.993 - 5 \times 0.0055 = 0.9652$$

由表 2 中可知,神经网络测试的 10 个检测样本所得相似度结果都超过了该相似度阈值,一致性检验通过。同理可得,该样品的夹角余弦法测试所得相似度结果也大于这个阈值。故通过二个算法得出的结果可以判断这 10 个切削液样本的类型为 A 型的精确性和可靠性都很高。(下转第 84 页)

- [J]. Mechanical Systems and Signal Processing, 2015, 60-61: 770-784.
- [8] Xu W J, Poon C C Y, Yam Y, et al. Motion compensated controller for a tendon-sheath-driven flexible endoscopic robot [J]. International Journal of Medical Robotics and Computer Assisted Surgery 2016, 13(1): 1-14.
- [9] Kaneko M, Yamashita T, Tanie K. Basic considerations on transmission characteristics for tendon drive robots [C]//International Conference on Advanced Robotics. California, USA, 1991: 827-832.
- [10] Palli G, Melchiorri C. Optimal control of tendon-sheath transmission systems [J]. IFAC Proceedings Volumes, 2006, 8(2): 185-191.
- [11] Palli G, Melchiorri C. Model and control of tendon-sheath transmission systems [C]//IEEE International Conference on

- Robotics & Automation, Florida, USA, 2006: 988-993.
- [12] Phee S J, Low S C, Huynh V A, et al. Design of a master and slave transluminal endoscopic robot for natural orifice transluminal endoscopic surgery [C]//International Conference of the IEEE Engineering in Medicine & Biology Society, Minneapolis, USA, 2009: 1192-1195.
- [13] Chen L, Wang X S, Tian F X. Tendon-sheath actuated robots and transmission system [C]//International Conference on Mechatronics & Automation. Changchun, China, 2009: 3173-3178.
- [14] Chen L, Wang X S, Lu J, et al. Design and preliminary experiments of a robot for searching interior of rubble [C]//IEEE International Conference on Systems, Nanjing, China, 2011: 1956-1961.

(编辑 李秀敏)

(上接第 79 页)

3.4 系统验证

通过对识别分析的流程进行分析,然后将本文模式识别算法模型嵌入识别分析系统,利用可视化编程技术与 Server 数据库技术,开发了一套切削液智能化识别分析系统并利用系统对识别方法进行了验证,证明了该方法的可行性。

系统界面如图 5~图 7 所示,该系统实现了基础数据的定义、标样的采集与训练、样本的识别分析、系统综合查询、系统管理维护等功能,可以准确的对切削液进行分类识别并自动生成识别分析报告。



图 5 系统主界面

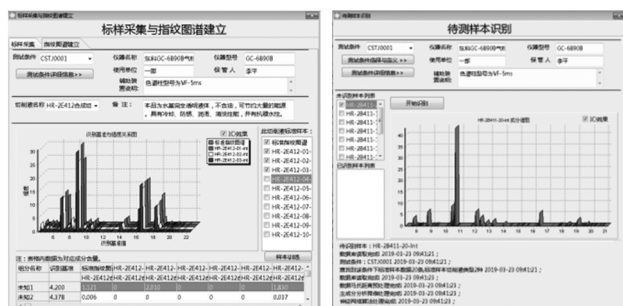


图 6 数据采集与图谱建立

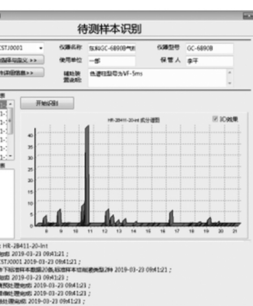


图 7 样本识别分析

4 结论

本文综合考虑了传统切削液辨别方法的缺点以及模式识别方法的应用,对切削液的辨别分析方法进行了深入的研究。在获取了切削液组分数据的基础上,对其数据进行了标准化处理,利用欧氏距离对杂质成分作了剔除处理,并利用主成分分析方法进行了降维

处理等一系列的数据预处理。对比夹角余弦法与神经网络法的分析结果,辨别出了切削液种类,并采用阈值算法对两种方法测得的结果进行了一致性验证,也通过实例验证了该方法的准确性和可行性。最终开发了一套切削液智能化识别分析系统并利用系统对识别方法进行了验证,解决了切削液的类别识别以及质量分析问题。

[参考文献]

- [1] 戴永红. 金属切削液的应用现状及发展趋势 [J]. 设备管理与维修, 2014(4): 66-67.
- [2] Zhong W W, Zhao D B, Wang X, et al. Adaptive Fuzzy Control of Cutting Temperature Based on Cutting Fluid in High-Speed Machining [J]. Advanced Materials Research, 2010, 97-101: 2381-2386.
- [3] 李金年. 机加工行业应合理有效使用切削液 [J]. 金属加工(冷加工), 2015(7): 28.
- [4] 孙元元, 肖岩. 水溶性切削液浓度的测定方法及应用 [J]. 辽宁化工, 2011, 40(8): 882-883, 886.
- [5] 张秀玲, 贾晓鸣. pH 值对合成切削液润滑性能的影响 [J]. 润滑与密封, 1998(1): 46-47.
- [6] 张璐熠, 朱明. 乳化切削液监测的过滤性检测方法研究 [J]. 润滑与密封, 2006(9): 183-186.
- [7] 黄建, 汪永超, 贾明刚. 基于模糊数学和神经网络 BP 算法的切削液选择 [J]. 组合机床与自动化加工技术, 2018(8): 160-163.
- [8] 陶桂宝, 王丽丹, 曹华军. 绿色制造制齿切削油液选择模型及应用 [J]. 重庆大学学报, 2015, 38(6): 75-83.
- [9] Roth P M, Hirzer M, Köstinger M, et al. Mahalanobis Distance Learning for Person Re-identification [M]. Person Re-identification, 2014.
- [10] 毕武, 袁小龙, 段新力, 等. 用马氏距离方法提取化探综合异常 [J]. 新疆有色金属, 2015, 38(2): 66-68.
- [11] Tipping M E, Bishop C M. Probabilistic Principal Component Analysis [J]. Journal of the Royal Statistical Society, 2010, 61(3): 611-622.
- [12] 景翔. 基于均方差和夹角余弦的水电机组振动诊断 [J]. 云南电力技术, 2013(1): 70-71.
- [13] 张景阳, 潘光友. 多元线性回归与 BP 神经网络预测模型对比与运用研究 [J]. 昆明理工大学学报(自然科学版), 2013, 38(6): 61-67.

(编辑 李秀敏)