

文章编号: 1000-4750(2007)Sup.I-0036-05

模糊聚类分析法在抽风机故障诊断中应用

扶名福¹, *谢明祥¹, 饶 泓²

(1. 南昌大学机电工程学院, 江西 南昌 330031; 2. 南昌大学计算中心, 江西 南昌 330031)

摘 要: 介绍了模糊聚类分析的理论和方法, 讨论了数据规格化的三种方案, 指明极差值是数据规格化最简便的方案。并将该理论应用于抽风机故障诊断研究中, 在样本中引入正常状态和故障状态时的样本数据, 运用基于传递闭包的模糊等价矩阵进行聚类分析, 并通过待检测数据检验其可靠性。通过实例检验表明, 理论分析与现场检查结果相符, 证实了模糊聚类分析对提高抽风机故障诊断能力的有效性。

关键词: 抽风机; 故障诊断; 模糊聚类; 聚类分析; 传递闭包

中图分类号: TP391.5 **文献标识码:** A

APPLICATION OF FUZZY CLUSTER ANALYSIS TO THE FAULT DIAGNOSIS OF SUCTION FAN

FU Ming-fu¹, *XIE Ming-xiang¹, RAO Hong²

(1. Mechanical and Electronic School, Nanchang University, Nanchang, Jiangxi 330031, China;

2. Computer Center, Nanchang University, Nanchang, Jiangxi 330031, China)

Abstract: Fuzzy cluster analysis theory is introduced, three methods of data standardization for cluster analysis are discussed and the extremity standardization method is recommended. This method was applied to the fault diagnosis of suction fan. The fault data and normal data were input to compose the sample data, the fuzzy equivalent matrix based on the transitive closure was imposed to the cluster analysis, and the checking data were introduced to validate this method. A practical example shows that the result calculated by fuzzy cluster analysis is in good agreement with the actual inspections, indicating that the method is available for the fault diagnosis of suction fan.

Key words: suction fan; fault diagnosis; fuzzy cluster; cluster analysis; transitive closure

抽风机能否良好运行直接影响到工厂的生产, 因此保证抽风机的安全运行十分重要, 对抽风机的运行状态检测及故障诊断也引起了大家的重视。常规的抽风机故障诊断方法基本上是通过故障机理进行逻辑推理, 但由于抽风机在运转过程中, 受到外界和机器构件之间的相互作用, 不可避免的要产生机械振动, 使得抽风机的故障机理复杂, 故障特征提取困难, 故障的征兆与故障之间的对应关系往

往呈现出高度的非线性和耦合性, 故障的存在具有很强的模糊性、多层次性、相似性等特点, 很难进行精确的推理。本文引进模糊聚类分析法, 结合某公司车身厂涂装车间 1#抽风机组的实际情况, 先采集正常状态和故障状态作为样本信号, 再加入两组待检测信号与样本信号进行比较, 对其进行故障诊断。利用模糊聚类分析法, 能有效地将待诊断参数与运行状态样本参数归类, 对系统故障作出快速准

收稿日期: 2006-09-17; 修改日期: 2006-11-05

基金项目: 江西省科技厅攻关项目(2004100B100)

作者简介: 扶名福(1953), 男, 江西南康人, 教授, 硕士, 博导, 副校长, 主要从事力学及材料科学研究(E-mail: fmfu@ncu.edu.cn);

*谢明祥(1982), 男, 江西南康人, 硕士生, 主要从事智能故障诊断研究(E-mail: ncxmx57@163.com);

饶 泓(1973), 女, 江西高安人, 副教授, 硕士, 硕导, 主要从事智能故障诊断研究(E-mail: raohong111@163.com).

确的判断,实验表明该方法是可行的、有效的。

1 模糊聚类理论与分析

1.1 模糊聚类理论概述

聚类分析是多元统计分析的一种,也是非监督模式识别的一个重要分支。它把一个没有类别标记的样本集按某种准则划分成若干个子集(类),使相似的样本尽可能归为一类,而不相似的样本尽量划分到不同的类中^[1]。然而现实的分类问题往往伴有模糊性,即考虑的不是有无关系,而是关系深浅度,这就是具有模糊关系的分类问题。对这些问题,用模糊数学语言来表达更为自然和符合实际,从而产生了模糊聚类分析。因此,模糊聚类分析是根据事物特性指标的模糊性,应用模糊数学方法确定样本的亲疏程度而实现的分类方法。

常用的模糊聚类算法有基于相似性关系^[2]和模糊关系的方法^[3]、基于模糊等价关系的传递闭包^[4]、基于模糊图论的最大树法^[5]、基于目标函数法^[6]。本文采用的是基于模糊等价关系的传递闭包法,对故障进行聚类分析诊断。

1.2 模糊聚类算法

基于模糊等价关系的传递闭包聚类算法由以下几步组成:

第一步:数据规格化。

由于与识别有关的诸特征量的量纲不同,所以在计算其相应的系数之前,要将量纲的影响消除,即对数据进行规格化处理,以便分析和比较。将数据进行规格化处理的方法很多,常用的方法有标准差规格化、极大值规格化、极差值规格化、均值规格化、中心规格化和对数规格化等^[7]。

很显然,中心规格化和对数规格化不能达到无量纲的目的。但极大值规格化容易受个别极端值的影响,一般情况下不宜采用。许多研究者对数据的规格化采用标准差规格化,为了使规格化后的数据落在[0,1]区间内,需对数据进一步处理,通常的方法是再进行一次极差值规格化^[8],还有一些研究者采用的是均值规格化^[9,10]。

设一分类问题有 n 个待分类样本,有 m 个特性指标,则数据矩阵如下:

$$U' = \begin{bmatrix} u'_{11} & u'_{12} & \cdots & u'_{1m} \\ u'_{21} & u'_{22} & \cdots & u'_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ u'_{n1} & u'_{n2} & \cdots & u'_{nm} \end{bmatrix}$$

则标准差规格化、均值规格化和极差值规格化可以用如下公式表示。

(1) 极差规格化

$$u_{ij} = \frac{u'_{ij} - u'_{j\min}}{u'_{j\max} - u'_{j\min}}, \quad (i=1,2,\cdots,n, j=1,2,\cdots,m) \quad (1)$$

式中 $u'_{j\max} = \max\{u'_{1j}, u'_{2j}, \cdots, u'_{nj}\}$,

$$u'_{j\min} = \min\{u'_{1j}, u'_{2j}, \cdots, u'_{nj}\}。$$

(2) 标准差规格化

$$u_{ij} = \frac{u'_{ij} - \bar{u}'_j}{\sigma'_j}, \quad (i=1,2,\cdots,n, j=1,2,\cdots,m) \quad (2)$$

式中 $\bar{u}'_j = \frac{1}{n} \sum_{i=1}^n u'_{ij}$, $\sigma'_j = \sqrt{\frac{1}{n-1} \sum (u'_{ij} - \bar{u}'_j)^2}$ 。

(3) 均值规格化

$$u_{ij} = \frac{u'_{ij}}{\bar{u}'_j}, \quad (i=1,2,\cdots,n, j=1,2,\cdots,m) \quad (3)$$

式中 $\bar{u}'_j$ 意义同上。

第二步:标定。

为了确定各检测数据之间的亲疏关系,给检测数据中每两组数据都赋以一相似系数 r_{ij} 。相似系数可以按实际情况采用不同方法计算。计算相似系数的方法很多,常用的有数量积法、夹角余弦法、相关系数法、指数相似系数法、最大最小法、算术平均最小法、几何平均最小法、绝对值指数法等^[11]。如相关系数法的定义为:

$$r_{ij} = \frac{\sum_{k=1}^m |u_{ik} - \bar{u}_i| |u_{jk} - \bar{u}_j|}{\sqrt{\sum_{k=1}^m (u_{ik} - \bar{u}_i)^2} \sqrt{\sum_{k=1}^m (u_{jk} - \bar{u}_j)^2}}, \quad (i, j=1,2,\cdots,n) \quad (4)$$

其中 $\bar{u}_i = \frac{1}{m} \sum_{k=1}^m u_{ik}$, $\bar{u}_j = \frac{1}{m} \sum_{k=1}^m u_{jk}$

该值的意义为:当 $r_{ij} = 0$ 时,表示 u_i 与 u_j 截然不同,毫无相似之处;若 $r_{ij} = 1$,表明两者完全相似或等同;而 $i = j$ 表示自身相似程度,恒取 1。

由此可见,矩阵 $\mathbf{R} = [r_{ij}]_{n \times n}$ 必然满足自反性、对称性,为相似矩阵。

第三步:求传递闭包^[12]。

因为矩阵 $\mathbf{R}$ 为相似矩阵,一般具有自反性和对称性,不满足传递性,而模糊相似矩阵需具有传递性,故需对矩阵 $\mathbf{R}$ 进行改造。为此有以下引理:

引理 当 $|A| = n$ 时,如果 $\mathbf{R}$ 是 A 上的模糊相似

矩阵，则必然存在 $k \leq n$ ，使 R 的传递闭包 $t(R) = R^{2^k}$ 是模糊等价矩阵， $|A|$ 表示论域 A 中的元素个数。

对 k 值的确定可以按下述次序：

$$R \rightarrow R^2 \rightarrow R^4 \rightarrow \cdots \rightarrow R^{2^k} \rightarrow \cdots$$

经过有限次运算后，必有 $R^{2^k} = (R^{2^k})^2$ ，就可以得到一个模糊等价矩阵 $t(R) = R^{2^k}$ 。

第四步：根据传递闭包进行聚类分析。

对满足传递性的模糊分类关系的 $t(R)$ 进行聚类处理，给定不同置信水平的 λ ，求 $[t(R)]_\lambda$ 矩阵，找出 $t(R)$ 的 λ 显示，得到普通的分类关系。根据实例中模糊等价矩阵的各个元素及分类的要求，按不同的置信水平 λ 对数据进行模糊分类，将会得到不同的分类结果。

2 应用实例

选取某公司车身厂涂装车间 1#抽风机组(检测部位如图 1 所示)的振动信号作为样本数据，分别采集故障情况数据(1 组，2 组)和正常情况数据(3 组，4 组)作为运行状态样本，再采集两组数据(5 组，6 组)作为待检验数据。其数据见表 1。

第一步：对样本数据进行规格化处理。

本文采用的规格化方案有三种：第一种是极差值规格化；第二种是标准值规格化，采用这种方法

得出的 u_{ij} 范围为： $-2 \leq u_{ij} \leq 2$ ，为使其落在 $[0,1]$ 区间内，再进行一次极差值规格化；第三种是均值规格化，采用这种方法得出的 u_{ij} 范围为： $0 \leq u_{ij} \leq 1+2CV'_j$ (其中 CV'_j 表示变异系数)，为使其落在 $[0,1]$ 区间内，再进行一次极差值规格化。

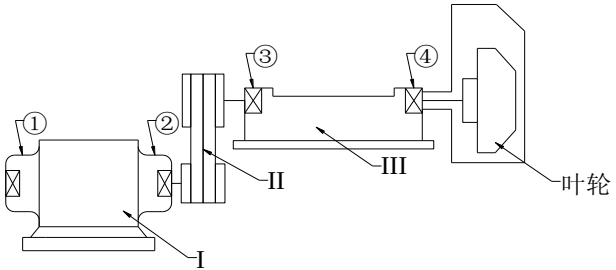


图 1 检测部位图
Fig.1 Inspected position chart

表 1 样本数据

	Table 1 Sample data							
	1H	1V	2H	2V	3H	3V	4H	4V
1	4.22	3.28	4.70	3.44	2.04	3.99	3.04	2.35
2	3.46	3.67	4.88	3.00	5.26	3.57	3.89	4.12
3	3.51	2.89	4.19	3.36	4.85	3.18	4.10	4.08
4	2.86	2.54	4.72	3.44	4.62	2.87	3.93	3.95
5	3.25	3.76	4.03	5.12	5.64	2.1	5.01	4.12
6	2.91	4.27	4.31	4.19	4.35	1.83	3.46	4.35

方案一、应用极差值规格化方法对样本数据进行规格化

利用式(1)可得规格化后的矩阵 U 为：

$$U = \begin{bmatrix} 1.0000 & 0.4277 & 0.7882 & 0.2075 & 0 & 1.0000 & 0 & 0 \\ 0.4412 & 0.6532 & 1.0000 & 0 & 0.8944 & 0.8056 & 0.4315 & 0.8850 \\ 0.4779 & 0.2023 & 0.1882 & 0.1698 & 0.7806 & 0.6250 & 0.5381 & 0.8650 \\ 0 & 0 & 0.8118 & 0.2075 & 0.7167 & 0.4815 & 0.4518 & 0.8000 \\ 0.2868 & 0.7052 & 0 & 1.0000 & 1.0000 & 0.1250 & 1.0000 & 0.8850 \\ 0.0368 & 1.0000 & 0.3294 & 0.5613 & 0.6417 & 0 & 0.2132 & 1.0000 \end{bmatrix}$$

方案二、先用标准差规格化方法对样本数据进行规格化，再用极差值规格化方法对数据进行后处理。

利用式(2)先将样本数据规格化，得矩阵 U 为：

$$U = \begin{bmatrix} 1.8762 & -0.2121 & 0.7337 & -0.4519 & -2.0856 & 1.3970 & -1.4265 & -2.2014 \\ 0.2019 & 0.4679 & 1.3121 & -1.0766 & 0.6894 & 0.8469 & -0.0247 & 0.4343 \\ 0.3121 & -0.8922 & -0.9051 & -0.5655 & 0.3361 & 0.3362 & 0.3216 & 0.3748 \\ -1.1199 & -1.5024 & 0.7980 & -0.4519 & 0.1379 & -0.0699 & 0.0412 & 0.1812 \\ -0.2607 & 0.6248 & -1.4192 & 1.9331 & -1.0169 & -1.0783 & 1.8223 & 0.4343 \\ -1.0097 & 1.5141 & -0.5195 & 0.6128 & -0.0948 & -1.4319 & -0.7339 & 0.7768 \end{bmatrix}$$

为使 u_{ij} 的值落在 $[0,1]$ 区间内，再用极差值规格化方法对数据进行后处理，由式(1)可得：

$$U = \begin{bmatrix} 1.0000 & 0.4277 & 0.7882 & 0.2075 & 0 & 1.0000 & 0 & 0 \\ 0.4412 & 0.6532 & 1.0000 & 0 & 0.8944 & 0.8056 & 0.4315 & 0.8850 \\ 0.4779 & 0.2023 & 0.1882 & 0.1698 & 0.7806 & 0.6250 & 0.5381 & 0.8650 \\ 0 & 0 & 0.8118 & 0.2075 & 0.7167 & 0.4815 & 0.4518 & 0.8000 \\ 0.2868 & 0.7052 & 0 & 1.0000 & 1.0000 & 0.1250 & 1.0000 & 0.8850 \\ 0.0368 & 1.0000 & 0.3294 & 0.5613 & 0.6417 & 0 & 0.2132 & 1.0000 \end{bmatrix}$$

方案三、先用均值规格化方法对样本数据进行规格化,再用极差值规格化方法对数据进行后处理。

利用式(3)先将样本数据规格化,得矩阵 U 为:

$$U = \begin{bmatrix} 1.2528 & 0.9642 & 1.0511 & 0.9153 & 0.4574 & 1.3649 & 0.7785 & 0.6138 \\ 1.0272 & 1.0789 & 1.0913 & 0.7982 & 1.1794 & 1.2212 & 0.9962 & 1.0762 \\ 1.0421 & 0.8496 & 0.9370 & 0.8940 & 1.0874 & 1.0878 & 1.0499 & 1.0657 \\ 0.8491 & 0.7467 & 1.0555 & 0.9153 & 1.0359 & 0.9818 & 1.0064 & 1.0318 \\ 0.9649 & 1.1053 & 0.9012 & 1.3623 & 1.2646 & 0.7184 & 1.2830 & 1.0762 \\ 0.8639 & 1.2553 & 0.9638 & 1.1149 & 0.9753 & 0.6260 & 0.8860 & 1.1363 \end{bmatrix}$$

为使 u_{ij} 的值落在 $[0,1]$ 区间内,再用极差值规格化方法对数据进行后处理,由式(1)可得:

$$U = \begin{bmatrix} 1.0000 & 0.4277 & 0.7882 & 0.2075 & 0 & 1.0000 & 0 & 0 \\ 0.4412 & 0.6532 & 1.0000 & 0 & 0.8944 & 0.8056 & 0.4315 & 0.8850 \\ 0.4779 & 0.2023 & 0.1882 & 0.1698 & 0.7806 & 0.6250 & 0.5381 & 0.8650 \\ 0 & 0 & 0.8118 & 0.2075 & 0.7167 & 0.4815 & 0.4518 & 0.8000 \\ 0.2868 & 0.7052 & 0 & 1.0000 & 1.0000 & 0.1250 & 1.0000 & 0.8850 \\ 0.0368 & 1.0000 & 0.3294 & 0.5613 & 0.6417 & 0 & 0.2132 & 1.0000 \end{bmatrix}$$

由以上可以知道,三种方案的结果是一样的,很明显,第一种极差值规格化方法是最简便的。

第二步:求初始相似矩阵和传递闭包。

本文采用的标定方法是相关系数法。利用式(4)计算,得初始矩阵 R ,经 3 次迭代后(即 $k=3$)矩阵具有传递性,得到模糊矩阵 $t(R)$:

$$R = \begin{bmatrix} 1.0000 & 0.7566 & 0.6776 & 0.7164 & 0.9275 & 0.8134 \\ 0.7566 & 1.0000 & 0.8120 & 0.7008 & 0.8677 & 0.5515 \\ 0.6776 & 0.8120 & 1.0000 & 0.8375 & 0.7781 & 0.7332 \\ 0.7164 & 0.7008 & 0.8375 & 1.0000 & 0.7402 & 0.8193 \\ 0.9275 & 0.8677 & 0.7781 & 0.7402 & 1.0000 & 0.7237 \\ 0.8134 & 0.5515 & 0.7332 & 0.8193 & 0.7237 & 1.0000 \end{bmatrix}$$

$$t(R) = \begin{bmatrix} 1.0000 & 0.8677 & 0.8134 & 0.8134 & 0.9275 & 0.8134 \\ 0.8677 & 1.0000 & 0.8134 & 0.8134 & 0.8677 & 0.8134 \\ 0.8134 & 0.8134 & 1.0000 & 0.8375 & 0.8134 & 0.8193 \\ 0.8134 & 0.8134 & 0.8375 & 1.0000 & 0.8134 & 0.8193 \\ 0.9275 & 0.8677 & 0.8134 & 0.8134 & 1.0000 & 0.8134 \\ 0.8134 & 0.8134 & 0.8193 & 0.8193 & 0.8134 & 1.0000 \end{bmatrix}$$

第三步:根据传递闭包进行聚类分析。

阈值 λ 的选取,首先应能确保正确区分正常状态样本与故障状态样本。由 $t(R)$ 可见, $\lambda = 0.8193$ 可将样本正确分类。若 λ 取值较大时,只能区分正

常与较严重故障;反之,若 λ 取值较小时,则能将较轻微的故障归类。现取阈值 $\lambda = 0.8193$,得到截矩阵 $[t(R)]_{\lambda=0.8193}$:

$$[t(\mathbf{R})]_{\lambda=0.8193} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

由以上 λ 截矩阵 $[t(\mathbf{R})]_{\lambda}$ 可明显地看出: 样本5属于样本1, 样本2所表示的故障状态; 样本6属于样本3, 样本4所表示的正常状态, 这和实际的情况相符。

3 结论

本文详细介绍了模糊聚类理论和模糊聚类算法的具体步骤, 重点讨论了三种数据规格化方案, 并将该方法应用于系统故障诊断中。在数据规格化三种方案比较中得出了用极差规格化方法是最简便的方案。

利用上述模糊聚类分析法, 可以在正常样本和故障样本保持不变的情况下, 加入新的待检测数据, 能有效地将待诊断参数与运行状态样本参数归类, 对系统故障作出快速准确的判断, 该方法是可行的、有效的。

参考文献:

- [1] 高新波. 模糊聚类分析及其应用[M]. 西安: 西安电子科技大学出版社, 2004.
Gao Xinbo. Fuzzy cluster analysis and its application [M]. Xi'an: Xidian University Press, 2004. (in Chinese)
- [2] Tamra S, Higuchi S, Tanaka K. Pattern classification based on fuzzy relations [J]. IEEE Transactions on SMC, 1971, 1(1): 217~242.
- [3] Backer E, Jain A K. A clustering performance measure based on fuzzy set decomposition [J]. IEEE Transactions on PAMI, 1981, 3(1): 66~74.
- [4] Le Zkim. Fuzzy relation compositions and pattern

- recognition [J]. Inf. Sci., 1996, 89: 107~130.
- [5] Wu Z, Leathy R. An optimal graph theoretic approach to data clustering: theory and its application to image segmentation [J]. IEEE Transactions on PAMI, 1993, 15(11): 1101~1113.
- [6] Bezdek J C. Pattern recognition with fuzzy objective function algorithms [M]. New York: Plenum Press, 1981.
- [7] 庄恒扬, 沈新平, 陆建飞, 黄丽芬. 模糊聚类计算方法的理论分析[J]. 江苏农学院学报, 1998, 19(3): 37~41.
Zhuang Hengyang, Shen Xinping, Lu Jianfei, Huang Lifan. Theoretical analyses of the computing method of fuzzy cluster [J]. Journal of Jiangsu Agricultural Collage, 1998, 19(3): 37~41. (in Chinese)
- [8] 吴福保, 李奇, 宋文忠. 模糊聚类分析的传递方法[J]. 东南大学学报, 1999, 29(2): 105~110.
Wu Fubao, Li Qi, Song Wenzhong. Transfer algorithm to fuzzy cluster analysis [J]. Journal of Southeast University, 1999, 29(2): 105~110. (in Chinese)
- [9] 罗曼莉, 蒋梁中, 翟敬梅. 基于模糊聚类的模糊推理系统及其在产品质量评价中的应用[J]. 机床与液压, 2003, (5): 289~290.
Luo Manli, Jiang Liangzhong, Zhai Jingmei. A fuzzy inference system based on fuzzy cluster and its application in product quality evaluation [J]. Machine Tool and Hydraulics, 2003, (5): 289~290. (in Chinese)
- [10] 马光志, 倪国元. 一种增量式模糊聚类算法[J]. 微计算机应用, 2005, 26(1): 5~7.
Ma Guangzhi, Ni Guoyuan. An increasable fuzzy clustering algorithm [J]. Microcomputer Applications, 2005, 26(1): 5~7. (in Chinese)
- [11] 温熙森, 胡芑庆, 邱静. 模式识别与状态监控[M]. 长沙: 国防科技大学出版社, 1997.
Wen Xisen, Hu Niaoqing, Qiu Jing. Pattern recognition and state detecting [M]. Changsha: National University of Defense Technology Press, 1997. (in Chinese)
- [12] 刘晋寅, 吴孟达. 模糊理论及其应用[M]. 长沙: 国防科技大学出版社, 1998.
Liu Jinyin, Wu Mengda. Fuzzy theory and application [M]. Changsha: National University of Defense Technology Press, 1998. (in Chinese)