

多特征不确定时间序列的关联趋势分析*

高亮¹, 孙卫²

(1. 空军工程大学 工程学院, 西安 710038; 2. 西安交通大学 系统工程研究所, 西安 710049)

摘要: 针对不确定信息的相似性度量方法无法充分反映信息之间的关联情况, 提出了直觉模糊集关联趋势分析法(RTIFS法)。利用直觉模糊集之间的距离表示不确定信息的差别, 通过区间数与直觉模糊集之间的等价关系, 利用区间数的距离计算直觉模糊集的关联度, 最后应用集对分析法对序列间的关联趋势进行分类。RTIFS法将关联度计算的范围推广到不确定信息环境下, 并给出多特征序列关联趋势的分类结果。实验结果表明, RTIFS法的分类准确率较高, 算法运行时间短。

关键词: 时间序列; 相关度; 直觉模糊集; 区间数; 集对分析

中图分类号: TP301.6 **文献标志码:** A **文章编号:** 1001-3695(2012)09-3255-04

doi:10.3969/j.issn.1001-3695.2012.09.014

Analysis approach of relevance trend of time series with multiple uncertain features

GAO Liang¹, SUN Wei²

(1. Institute of Engineering, Air Force Engineering University, Xi'an 710038, China; 2. Institute of Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: In view of the deficiency of means to directly calculate the relational degree between time series of uncertain information, this paper proposed the approach of the relevant trend analysis between series of intuitionistic fuzzy sets. The approach, firstly, quantified the difference between uncertain information by using the distance of intuitionistic fuzzy sets. Secondly, using equivalence relation between IFS and interval numbers, calculated the relevant degree of time series, on the basis of definition of distance between time series of interval numbers. Finally, classified the relevance trend of uncertain information series through the method of set pair analysis. The approach expanded the application range of relational degree from crisp numbers to uncertain environment expressed by intuitionistic fuzzy sets, and it classified the degree of relevance trend between time series with multiple uncertain features. Compared with the C-means algorithm and the simulation anneal algorithm, experiment results indicate, by using this method, the accurate rate of the algorithm is higher; and the false alarm rate and the missing alarm rate are both lower; furthermore, it reduces the running time effectively.

Key words: time series; relational degree; intuitionistic fuzzy set(IFS); interval number; set pair analysis

0 引言

时间序列的趋势分析^[1~3]作为一种数据驱动的故障诊断方法,通过关联度分析序列之间的相似、相近程度,定性地判断设备故障发生的时间和发展趋势。趋势分析法的物理意义直观,适用于大量序列的关联分析^[4]。但现有的序列关联分析方法无法直接计算不确定信息环境下序列之间的关联度。由于系统模型不完整、构成部分不确定,或者受环境噪声、传感器自身误差和操作误差的影响,工程实践中获得时序信息的连续性和完整性很难得到保证,这些数据信息包含不精确、不确定的成分,有时甚至可能是通过人员观察得到自然语言变量表示的信息^[5,6]。现有的时间序列关联趋势分析法只有在预处理消除了数据的不确定性后才能完成序列关联分析,而数据预处理过程造成了数据信息的损失^[7,8]。

Atanassov 增加非隶属度来描述“非此非彼”的模糊概念,建立直觉模糊集(intuitionistic fuzzy sets, IFS)以讨论不确定性

信息,所含信息量多于模糊集,表述更接近人类认知特点。本文利用 IFS 表述模糊概念和自然语言的优势,分析含有多个特征的序列间关联趋势,提出了直觉模糊集关联趋势分析法(RTIFS法)。RTIFS法将序列关联度计算从精确数范围推广至 IFS 范围,从单一特征序列推广至包含多个特征的序列,并考察自然语言表示的特征。最后给出实例,说明 RTIFS 方法能够计算多特征不确定序列的相关程度,在处理自然语言表示的特征时更具优势。

1 直觉模糊集和集对分析法

1.1 直觉模糊集

设 X 为给定论域, Zadeh 引入模糊集概念:

$$F = \{ \langle x, \mu_F(x) \rangle \mid x \in X \}$$

基本构成 $\mu_F(x)$ 称为隶属度,相应的非隶属度为 $1 - \mu_F(x)$ 。在现实生活中,当问及某人对某事表示的偏好程度时,总是会有不确定或者犹豫的成分,这部分内容在模糊集中无法得到表

收稿日期: 2012-02-21; 修回日期: 2012-03-22 基金项目: 国家自然科学基金资助项目(10971164)

作者简介: 高亮(1975-),男,新疆乌鲁木齐人,博士研究生,主要研究方向为不确定性分析和决策(ellenyuanzhi@163.com);孙卫(1966-),女,辽宁大连人,副教授,博士,主要研究方向为并行计算优化算法。

示。为了解决这个问题,Atanassov 增加了表示犹豫程度的量,将 Zadeh 的模糊集理论推广为直觉模糊集理论。

定义 1 直觉模糊集^[9]。设 X 为给定论域,则 X 上的一个直觉模糊集 A 为

$$\text{IFS}_A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle | x \in X \}$$

其中: $\mu_A(x)$ 表示隶属度, $\nu_A(x)$ 表示非隶属度,且有 $\mu_A: X \rightarrow [0, 1]$, $x \in X \rightarrow \mu_A(x) \in [0, 1]$, $\nu_A: X \rightarrow [0, 1]$, $x \in X \rightarrow \nu_A(x) \in [0, 1]$, 对所有的 $x \in X$, 满足 $\mu_A(x) + \nu_A(x) \leq 1$ 。

对 X 中每一个 IFS_A , 称 $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$ 为 x 对 A 犹豫程度的测度, 也称直觉指数 (intuitionistic index)。对于每一个 $x \in X$, $0 \leq \pi_A(x) \leq 1$, 如果 $\pi_A(x) = 0$, 直觉模糊集退化为模糊集。

1.2 区间数的距离

定义 2 区间数。称实数域 R 中的闭区间 $[x_{\text{low}}, x_{\text{up}}]$ 为区间数, 记做 I_x 。不失一般性, 令 $0 \leq x_{\text{low}} \leq x_{\text{up}}$ 。当 $x_{\text{low}} = x_{\text{up}}$ 时, I_x 退化为一个确定数。区间数还可以用二元组表示为 $I_x = \langle o_x, w_x \rangle$ 。其中: $o_x = (x_{\text{low}} + x_{\text{up}})/2$ 为区间数的中心, $w_x = (x_{\text{low}} - x_{\text{up}})/2$ 为区间数的半宽。当且仅当两个区间数的中心和半宽都相等时, 区间数相等。直觉模糊集可以用区间数表示, 即 $I_x = [\mu_A(x), 1 - \pi_A(x)]$ 。不失一般性, 记 $x_{\text{low}} = \mu_A(x)$ 和 $x_{\text{up}} = 1 - \pi_A(x)$ 。为方便表达, IFS 之间的计算用区间数计算表示。

区间数之间的距离没有直观的解释, 即使两个非 0 区间数的上、下限相等, 也不能得出这两个区间的距离为 0 的结论。本文采用距离作为反映区间差别的度量。

设区间数集内的任意三个区间数 I_x, I_y 和 I_z , 存在映射 $d: I_x \times I_y \rightarrow [0, 1]$, 如果实数 $d(I_x, I_y)$ 满足以下公理:

公理 1 $d(I_x, I_y) = 0 \Leftrightarrow I_x = I_y$;

公理 2 $d(I_x, I_y) = d(I_y, I_x)$;

公理 3 $d(I_x, I_z) \leq d(I_x, I_y) + d(I_y, I_z)$;

则称 $d(I_x, I_z)$ 为区间数 I_x 与 I_y 之间的距离度量 (简称为距离)。区间数中心反映区间位置, 半宽显示信息不确定程度, 它们在度量区间数间的分离程度上处于同等重要的地位。

定义 3 区间数距离。设 $I_x = \langle o_x, w_x \rangle$ 和 $I_y = \langle o_y, w_y \rangle$ 为区间数集内的任意两个区间数, 则区间数 I_x 与 I_y 之间的距离为

$$\text{dis}(I_x, I_y) = [|o_x - o_y|^n + |w_x - w_y|^n]^{1/n} \quad (1)$$

其中: $n \in \mathbf{N}$, 通常取 $n = 2$ 。

证明 只有满足距离度量的三条公理, 定义 2 才是有效的。公理 1 和 2 显然成立, 下面证明公理 3, 即需证明:

$$\begin{aligned} [|o_x - o_z|^n + |w_x - w_z|^n]^{\frac{1}{n}} &\leq [|o_x - o_y|^n + |w_x - w_y|^n]^{\frac{1}{n}} + \\ &[|o_y - o_z|^n + |w_y - w_z|^n]^{\frac{1}{n}} \quad n \in \mathbf{N} \end{aligned}$$

当 $n = 1$ 时, 由实数域内三角不等式知其显然成立; 当 $n > 1$ 时, 利用加减变量法构造下式:

$$\begin{aligned} |o_x - o_z|^n + |w_x - w_z|^n &= |o_x - o_y + o_y - o_z|^n + \\ |w_x - w_y + w_y - w_z|^n &= |o_x - o_y + o_y - o_z| |o_x - o_z|^{n-1} + \\ |w_x - w_y + w_y - w_z| |w_x - w_z|^{n-1} \end{aligned}$$

由三角不等式有

$$\begin{aligned} |o_x - o_y + o_y - o_z| |o_x - o_z|^{n-1} &\leq |o_x - o_y| |o_x - o_z|^{n-1} + \\ |o_y - o_z| |o_x - o_z|^{n-1} & \\ |w_x - w_y + w_y - w_z| |w_x - w_z|^{n-1} &\leq |w_x - w_y| |w_x - w_z|^{n-1} + \\ |w_y - w_z| |w_x - w_z|^{n-1} \end{aligned}$$

令 $q = \frac{n}{n-1}$, 由 Hölder 不等式有

$$\begin{aligned} |o_x - o_y| |o_x - o_z|^{n-1} + |w_x - w_y| |w_x - w_z|^{n-1} &\leq \\ (|o_x - o_y|^n + |w_x - w_y|^n)^{1/n} (|o_x - o_z|^{(n-1)q} + |w_x - w_z|^{(n-1)q})^{1/q} \end{aligned}$$

同理可得

$$\begin{aligned} |o_y - o_z| |o_x - o_z|^{n-1} + |w_y - w_z| |w_x - w_z|^{n-1} &\leq \\ (|o_y - o_z|^n + |w_y - w_z|^n)^{1/n} (|o_x - o_z|^{(n-1)q} + |w_x - w_z|^{(n-1)q})^{1/q} \end{aligned}$$

将 Hölder 不等式得到的结果带入证明开始时的不等式左边, 再在得到的不等式两边同时除以 $(|o_x - o_z|^{(n-1)q} + |w_x - w_z|^{(n-1)q})^{1/q}$, 即得证公理 3。所以定义 2 满足距离的公理, 证毕。

1.3 区间数的距离

定义 4 设 $I_x(t), I_y(t)$ 分别为时间序列 I_x 和 I_y 在时刻 t 的特征量, 即 $I_x(t) = \{ I_{x_1}(t), \dots, I_{x_i}(t), \dots, I_{x_m}(t) \}$, $I_y(t) = \{ I_{y_1}(t), \dots, I_{y_j}(t), \dots, I_{y_n}(t) \}$, $i, j, m, n \in \mathbf{N}$, 设 $1 \leq m \leq n$, m 和 n 分别称为序列 I_x 和 I_y 的特征数, 不妨取表示描述同一现象序列的特征数均为 m 。序列 $I_x = \{ I_x(1), \dots, I_x(l), \dots, I_x(p) \}$, $I_y = \{ I_y(1), \dots, I_y(s), \dots, I_y(q) \}$, $r, s, p, q \in \mathbf{N}$, 设 $1 \leq p \leq q$, 其中 p 和 q 分别称为 I_x 和 I_y 的时间长度。

为统一分析标准, 对序列各特征进行规范化处理, 其中, $I_{x_i}(t) = [x_{\text{low}}(t), x_{\text{up}}(t)]$, 有

$$I_{x_i}^*(t) = \left[\frac{x_{\text{low}}(t) - \min_{t=1, \dots, p} \{x_{\text{low}}(t)\}}{\max_{t=1, \dots, p} \{x_{\text{up}}(t)\} - \min_{t=1, \dots, p} \{x_{\text{low}}(t)\}}, \frac{x_{\text{up}}(t) - \min_{t=1, \dots, p} \{x_{\text{low}}(t)\}}{\max_{t=1, \dots, p} \{x_{\text{up}}(t)\} - \min_{t=1, \dots, p} \{x_{\text{low}}(t)\}} \right] \quad (2)$$

当序列特征一致时, 序列在时刻 t 的距离为

$$\text{dis}[I_x^*(t), I_y^*(t)] = \sum_{i=1}^m k_i \text{dis}[I_{x_i}^*(t), I_{y_i}^*(t)] \quad (3)$$

其中: k_i 表示序列第 i 个特征在关联度分析中所占的权重, 且 $\sum_{i=1}^m k_i = 1$, 序列之间的距离为

$$\text{dis}[I_x, I_y] = \sum_{t=1}^p \sum_{i=1}^m \text{dis}[I_{x_i}^*(t), I_{y_i}^*(t)] \quad (4)$$

称反映已知状态 (一般为设备正常工作状态) 的测量序列为基准序列, 记为 I_x 。称反映未知状态的测量序列为测试序列, 记为 I_y 。 I_y 和 I_x 的趋势比对可以反映两个序列之间的相似趋势, 进而解释 I_y 的状态。

1.4 集对分析法

集对分析法^[10]从同、异、反的角度对两种事物的确定和不确定关系进行了量化描述。

定义 5 给定两个集合 A 和 B , 并设这两个集合组成集对 $H = (A, B)$, 在某个具体问题的背景 W 下, H 有 n 个特性, 其中有 s 个特性为 A 和 B 所共有, p 个特性为 A 和 B 相对立, f 个特性为 A 和 B 既不共有也不对立, 则称 s/n 为 A 和 B 在问题 W 下的同一度 (以 a 表示), f/n 为问题 W 下的差异度 (以 b 表示), p/n 为问题 W 下的对立度 (以 c 表示)。 A 和 B 的联系度为

$$\mu_W(A, B) = \frac{s}{n} + \frac{f}{n}i + \frac{p}{n}j = a + bi + cj \quad (5)$$

其中: $a + b + c = 1$, $0 \leq a, b, c \leq 1$; j 为对立度的系数, 一般取 -1 ; $i \in [-1, 1]$ 为差异度的系数。当 i 取 -1 或 1 时, 差异度的不确定性完全消失; 当 i 趋向于 0 时, 差异度增加。在实际应用中, 确定的 μ_W 值会掩盖集合之间的同、异、反的趋势变化, 不改变 i 和 j 的值更有利于判断集合之间的趋势。

按联系度三个参数的大小关系可以将同、异、反关系分为13个等级。由序列与表示设备正常工作状态的序列相合、背离的趋势可以将设备状态分为三种。具体分类如图1所示。

正常	轻微故障	严重故障
同势 I 级 $a > b > c$	均势 I 级 $a = c > b$	均势 III 级 $b > a = c$
同势 II 级 $a > b = c$	均势 II 级 $a = b = c$	反势 III 级 $a < b < c$
同势 III 级 $a > c > b$	反势 V 级 $a < c < b$	反势 II 级 $a = b < c$
同势 IV 级 $a = b > c$	反势 IV 级 $a < b = c$	反势 I 级 $c > a > b$
同势 V 级 $b > a > c$		

图1 同、异、反关系分级及设备状态分类

1.5 序列关联度

用灰色关联度 r 度量序列之间的相关性^[11,12],如果两个序列变化的趋势(方向、大小和速度等)基本一致,则它们之间的关联度较大;反之,关联度较小。若关联度 $0 \leq r \leq 1$,1 表示序列之间变化趋势完全一致,0 表示无关联。

定义6 $I_x(t)$ 、 $I_y(t)$ 分别为 I_x 和 I_y 在时刻 t 的特征序列, $t=1, \dots, p$, 若 M_1 、 M_2 和 M_3 都存在,则 I_x 和 I_y 的关联度为

$$r = \left[1 - \cot \frac{\pi M_1}{2p} - \cot \frac{\pi M_2}{2(p-1)} - \cot \frac{\pi M_3}{2(p-2)} \right]^{-1} \quad (6)$$

其中:

$$M_1 = \sum_{t=1}^p \text{dis}[I_x^*(t), I_y^*(t)]$$

$$M_2 = \sum_{t=1}^{p-1} |\text{dis}[I_x^*(t+1), I_y^*(t+1)] - \text{dis}[I_x^*(t), I_y^*(t)]|$$

$$M_3 = \sum_{t=2}^{p-1} |\text{dis}[I_x^*(t+1), I_y^*(t+1)] - 2\text{dis}[I_x^*(t), I_y^*(t)] + \text{dis}[I_x^*(t-1), I_y^*(t-1)]|$$

定义6中应用余切函数克服了B型关联度不具备规范性的缺点^[13]。作为区间数序列的特例,精确数序列同样适用于定义6。

在故障诊断应用中,当 $\text{dis}[I_x^*(t), I_y^*(t)] = 1$ 时,表示序列 I_y 和 I_x 在时刻 t 的状态差距最大,即测试序列 I_y 所表示的设备在 t 时刻工作异常;当 $\text{dis}[I_x^*(t), I_y^*(t)] = 0$ 时,表示序列 I_y 和 I_x 在时刻 t 状态完全相同,设备工作正常; $\text{dis}[I_x^*(t), I_y^*(t)]$ 在 $0 \sim 1$ 内变化时,表示表示序列 I_y 和 I_x 在时刻 t 的状态由完全相同变化至完全不同。因此,构造序列 $\{1 - \text{dis}[I_x^*(t), I_y^*(t)]\}$ 来判断设备工作状态。

假设精确数最优序列为 $(\frac{p}{1}, \dots, \frac{p}{1})$, 则序列 $\{1 - \text{dis}[I_x^*(t), I_y^*(t)]\}$ 与这个最优序列的关联度 r^+ 为

$$\left[1 - \cot \frac{\pi \sum_{t=1}^p |1 - \text{dis}[I_x^*(t), I_y^*(t)]|}{2p} - \cot \frac{\pi \sum_{t=1}^{p-1} |\text{dis}[I_x^*(t+1), I_y^*(t+1)] - \text{dis}[I_x^*(t), I_y^*(t)]|}{2(p-1)} - \cot \frac{\pi \sum_{t=2}^{p-1} |\text{dis}[I_x^*(t+1), I_y^*(t+1)] - 2\text{dis}[I_x^*(t), I_y^*(t)] + \text{dis}[I_x^*(t-1), I_y^*(t-1)]|}{2(p-2)} \right]^{-1} \quad (7)$$

假设精确数最劣序列为 $(\frac{p}{0}, \dots, \frac{p}{0})$, 则序列 $\{1 - \text{dis}[I_x^*(t), I_y^*(t)]\}$ 与这个最劣序列的关联度 r^- 为

$$\left[1 - \cot \frac{\pi \sum_{t=1}^p |1 - \text{dis}[I_x^*(t), I_y^*(t)]|}{2p} - \cot \frac{\pi \sum_{t=1}^{p-1} |\text{dis}[I_x^*(t+1), I_y^*(t+1)] - \text{dis}[I_x^*(t), I_y^*(t)]|}{2(p-1)} - \cot \frac{\pi \sum_{t=2}^{p-1} |\text{dis}[I_x^*(t+1), I_y^*(t+1)] - 2\text{dis}[I_x^*(t), I_y^*(t)] + \text{dis}[I_x^*(t-1), I_y^*(t-1)]|}{2(p-2)} \right]^{-1} \quad (8)$$

因此,有

$$(r^+)^{-1} - (r^-)^{-1} = \cot \left[\frac{\pi}{2p} \sum_{t=1}^p |1 - \text{dis}[I_x^*(t), I_y^*(t)]| \right] -$$

$$\cot \left[\frac{\pi}{2p} \sum_{t=1}^p |\text{dis}[I_x^*(t), I_y^*(t)]| \right]$$

令 $(r^+)^{-1} - (r^-)^{-1} = \Phi$, 有 $d(r^+ + r^-)/dr^- = 1 + (\Phi r^- + 1)^{-2} > 0$, 因此知 $r^+ + r^-$ 在 $[0, 1]$ 区间上对于 r^- 是单调递增的, 即 $r^+ + r^- \leq 1 + (\Phi + 1)^{-1}$, 有

$$(\Phi + 1)(\Phi + 2)^{-1} r^+ + (\Phi + 1)(\Phi + 2)^{-1} r^- \leq 1 \quad (9)$$

其中: $(\Phi + 1)(\Phi + 2)^{-1} r^+$ 反映了 I_y 和 I_x 的接近程度, $(\Phi + 1)(\Phi + 2)^{-1} r^-$ 反映了 I_y 和 I_x 的偏离程度, 将它们分别定义为集对的一度 a 和对立度 c 。取 $a = (\Phi + 1)(\Phi + 2)^{-1} r^+$, $c = (\Phi + 1)(\Phi + 2)^{-1} r^-$, b 值由集对定义计算。

2 多特征不确定序列关联分析

对多特征不确定序列之间的关联趋势,建立直觉模糊集关联趋势分析法(RTIFS)。具体步骤为:

a) 用 IFS 描述包含多个特征的序列。将 10 等级自然语言变量对应的 IFS 定义如表 1 所示。

表1 自然语言对应的 IFS

语言变量	IFS
极好	EG [1.00, 0.00]
非常非常好	VVG [0.90, 0.10]
非常好	VG [0.80, 0.10]
好	G [0.70, 0.20]
微弱好	MG [0.60, 0.30]
中等	M [0.50, 0.40]
微弱差	MB [0.40, 0.50]
差	B [0.25, 0.60]
非常差	VB [0.10, 0.75]
非常非常差	VVB [0.10, 0.90]

b) 由式(2)对各特征变量进行规范化处理。

c) 由式(3)计算 t 时刻,测试序列 $I_y^*(t)$ 与最优、最劣序列特征之间的距离,特征权重可以采用多种方法确定。为方便算法讨论,实例采用指定权重。

d) 由式(7)(8)计算测试序列 I_y 与最优、最劣序列之间的距离,得到 r^+ 和 r^- 。

e) 计算同一度 a 和对立度 c , 分析 I_y 与 I_x 之间的关联趋势。

3 实验结果与分析

以某减速箱轴承振动的测量数据^[14]为例,共有 47 组数据,每组数据由 10 个离散值构成,数据对应的轴承状态都已由实验室观测确定。任意抽取 10 组正常状态的数据构建基准序列,剩余 37 组数据作为测试数据。为了分析方便,按实验室实测结果将 37 组数据编号,分别对应轴承正常(第 1~23 组)、存在轻微故障(第 24~35 组)和存在严重故障(第 36、37 组)。因为这些数据是从同一时间点开始记录,所以分析中不考虑时移和翻转转换。算法运行环境为 CPU 2.6 GHz、内存 2 GB 的 PC,采用 C++ 进行算法编程。

在 RTIFS 法中,取构建 I_x 的 10 组数据在时刻 t 的最大值(最小值)作为序列在该时刻的上(下)限。 I_y 由精确数序列本身作为中心、半宽为 0 的区间数序列构成。对序列添加一个自然语言表示特征,此特征由 37 组数据和 I_x 的位置关系确定,将 I_y 的 10 个离散值在 I_x 对应时刻上、下限内的总个数作为自然语言序数,即 10 个点都在 I_x 内,则 I_y 的第二个特征为 EG,9

个点为VVG。以此类推,具体对应表1的10种自然语言变量。 I_x 本身的第二特征记为EG。两个特征的权重取 k_1, k_2 。

C-均值法和模拟退火法对37组10维数据进行聚类分析,算法截止条件是分类数为3。特征权重 $k_1 = 1, k_2 = 0$ 时的RTIFS法、C-均值法和模拟退火法分析数据的结果如表2所示。

表2 测试数据的聚类结果

状态	数据组编号		
	RTIFS法	C-均值法	模拟退火法
正常	1,2,3,4,5,6,7,8,9,10, 11,12,13,14,15,16, 17,18,20,21,22,23	1,2,3,4,5,6,7,9,10, 11,12,13,14,15,16, 17,18,19,20,21,22,23,31	1,2,3,4,5,6,7,8,9,10, 11,12,13,14,15,16, 17,18,19,20,21,22,23
	19,24,25,26,27,28,29, 30,31,32,33,34,35	8,24,25,27,28,29, 30,32,33,34,35	24,25,27,28,29, 30,31,32,33,34,35,36
轻微故障			
严重故障	36,37	26,36,37	26,37

采用以下三个指标衡量算法性能:

分类正确率 = (分类正确数/总测试数据数) × %

虚警率 = [(正常被分为轻微故障数 + 正常被分为严重故障数 + 轻微故障被分为严重故障数)/总测试数据数] × %

漏警率 = [(严重故障被分为正常数 + 严重故障被分为轻微故障数 + 轻微故障被分为正常数)/总测试数据数] × %

三种算法性能比较如表3所示。

表3 三种算法性能比较

算法	分类正确率 /%	虚警率 /%	漏警率 /%	运行时间 /ms
RTIFS法	97.3	2.7	0	77.53
C-均值法	91.9	5.4	2.7	337.59
模拟退火法	94.6	2.7	2.7	213.23

由表2和3可以看出,RTIFS法分类正确率最高,发生了一个虚警,没有发生漏警的情况;C-均值法分类正确率最低,虚警和漏警均有发生;模拟退火法虽有较高的正确率,但虚警和漏警均有发生。由于RTIFS法不涉及最优值求解,算法运行时间比其他两种算法的运行时间分别减少了77%和63.4%。

从RTIFS法的计算结果中任意取出五组进行集对分析。各组计算同一度 a 和对立度 c ,结果如图2所示。

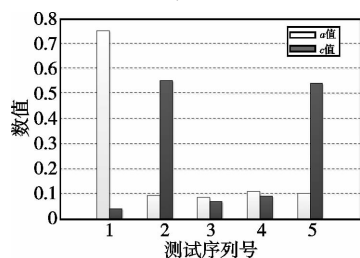


图2 各组测试序列 a 和 c 值比较

第2和5组的 c 明显大于 a ;第1组的 c 远小于 a ;第3、4组的 a 和 c 相差无几, a 略小于 c 。测试序列2、5相对于基准序列处于反势I级,测试序列3、4对于基准序列处于同势V级,而测试序列1相对基准序列处于同势I级。序列2、5偏离基准序列的趋势较强,这两段序列的趋势说明减速箱轴承存在故障的可能性最大,建议对该轴承和相关设备进行彻底检查。序列3、4表示轴承基本处于正常运行状态,这两个序列处于与基准序列很弱的同势,因此需要加大监测的力度,对该设备运行参数要作经常性的记录分析。序列1与设备正常状态趋势基本一致,建议对该设备进行间隔周期较长的阶段性监测。RTIFS法能够定量地分析序列间的相对发展趋势,这种作用是其他聚类算法所缺少的。

为研究具有不同权重特征对RTIFS方法的影响,分别令 (k_1, k_2) 为(0.8, 0.2)和(0.2, 0.8),分类结果如表4所示。

表4 特征取不同权重的RTIFS聚类结果

状态	数据组编号		
	$(k_1, k_2) = (0.8, 0.2)$	$(k_1, k_2) = (0.5, 0.5)$	$(k_1, k_2) = (0.2, 0.8)$
正常	1,2,3,4,5,6,7,8,9, 10,12,13,14,15,16,17,18,19,20,21,22,23	1,2,3,4,5,6,7,8,9, 10,11,12,13,14,15,16,10,11,12,13,14,15,16,17, 19,20,21,22,23	1,2,3,4,6,7,8,9, 10,11,12,13,14,15,16,17, 19,20,21,22,23
	11,24,25,26,27,28, 29,30,31,32,33,34,35	24,25,26,27,28,29, 30,32,33,34,35,36	5,18,24,25,27,28, 29,30,31,32,33,34,35,37
轻微故障			
严重故障	36,37	37	25,36

通过聚类结果发现,RTIFS方法随权重 k_1 减小,虚警和漏警数目开始增加,一方面说明自然语言特征比重越大,分类结果越不确定;另一方面也说明实例中第二特征的选择不是非常合理。虽然如此,RTIFS法却给出了自然语言特征进行定量分析的方法,为很多难以精确测量或无法测量却能被自然语言描述的特征的应用给出了一种有效的解决途径。

4 结束语

时序状态的定性分析是快速分析大量信息的有效方法,是定量分析方法的重要补充。序列分析中常常含有大量的不确定信息和包括自然语言在内的描述性变量,区间数、模糊集、直觉模糊集等表示方法已成为分析这些信息的有效手段。RTIFS法提出了IFS表示的序列关联度计算,对不确定信息之间的相似程度进行分析,将关联性计算由精确数范围推广至IFS范围,并进一步研究了包含多个特征的序列关联性。算例中对自然语言表示特征的有效分析,说明RTIFS法对不确定性故障诊断和预测具有理论价值。下一步研究将主要集中在不确定信息条件下的模式匹配和聚类分析问题。

参考文献:

- [1] KULIK I, HORNSBY K S, BISHOP I D. Modeling geospatial trend changes in vegetation monitoring data original research article computers[J]. *Environment and Urban Systems*, 2011, 35(1): 45-56.
- [2] MAURYA M R, PARITOSH P K, RENGASWAMY R, et al. A framework for on-line trend extraction and fault diagnosis[J]. *Engineering Applications of Artificial Intelligence*, 2010, 23(6): 950-960.
- [3] CHOI J, YI S Y, LEE K C. Analysis of keyword networks in MIS research and implications for predicting knowledge evolution[J]. *Information & Management*, 2011, 48(8): 371-381.
- [4] AGAMI N, SALEH M, SHISHINY H E. A fuzzy logic based trend impact analysis method[J]. *Technological Forecasting and Social Change*, 2010, 77(7): 1051-1060.
- [5] BRAGLIA M, CARMIGNANI G, FROSOLINI M, et al. Data classification and MTBF prediction with a multivariate analysis approach[J]. *Reliability Engineering & System Safety*, 2012, 97(1): 27-35.
- [6] HASSANI H, SOOFI A S, ZHINGLJAVSKY A A. Predicting daily exchange rate with singular spectrum analysis[J]. *Nonlinear Analysis: Real World Applications*, 2010, 11(3): 2023-2034.
- [7] ZHU Hong-ning, ZHANG Bin, GAO Xian-wen. A service selection method of non-critical task in parallel structure [C]//Proc of the 2nd International Workshop on Knowledge Discovery and Data Mining. Piscataway: IEEE Press, 2009: 846-850.

(上接第 3258 页)

- [8] WONG H, HU Bao-qing, IP W C, *et al.* Change-point analysis of hydrological time series using grey relational method [J]. **Journal of Hydrology**, 2006, 18(1): 323-338.
- [9] ATANASSOV K T. Intuitionistic fuzzy sets[J]. **Fuzzy sets and systems**, 1986, 20(1): 87-96.
- [10] LI Xiao-bai, SARKAR S. Against classification attacks: a decision tree pruning approach to privacy protection in data mining [J]. **Operations Research**, 2009, 57(6): 1496-1509.
- [11] JIN Min, ZHOU Xiang, ZHANG Z M, *et al.* Short-term power load

forecasting using grey correlation context modeling[J]. **Expert Systems with Applications**, 2012, 39(1): 773-779.

- [12] FU Ming-fu, XIE Ming-xiang, RAO Hong, *et al.* Fault diagnosis of suction fan based on the synthetically relational analysis [J]. **China Mechanical Engineering**, 2007, 18(20): 2403-2405.
- [13] TAKI Y, HASHIZUME H, SASSA Y, *et al.* Correlation among body height, intelligence, and brain gray matter volume in healthy children [J]. **NeuroImage**, 2012, 59(2): 1023-1027.
- [14] 温熙森. 模式识别与状态监控[M]. 北京: 科学出版社, 2007: 245-248.