

信息不确定条件下时间序列的关联分析法

高亮¹, 孙卫^{2,3}, 朱荣昌¹

(1. 空军工程大学工程学院, 710038, 西安; 2. 浙江工业大学之江学院, 310014, 杭州;
3. 西安交通大学系统工程研究所, 710049, 西安)

摘要: 针对不确定信息环境中无法直接计算序列关联度的问题, 提出了区间数序列关联趋势分析法(RTASI法). 首先利用区间数距离表示不确定信息的差别, 经过时移和翻转转换对序列进行匹配后, 通过区间数序列之间的距离计算序列关联度, 最后应用集对分析法对序列间的关联趋势进行分类. RTASI法将关联度计算的范围推广到不确定信息环境下, 并给出序列关联趋势的分类结果. 实验结果表明, RTASI法的分类准确率较高, 其虚警率和漏警率均低于C均值法和模拟退火法, 运行时间分别比C均值法和模拟退火法减少了77%和63.4%.

关键词: 时间序列; 关联度; 区间数; 集对分析

中图分类号: TP301.6 **文献标志码:** A **文章编号:** 0253-987X(2010)06-0067-05

A Relational Analysis Approach of Time Series with Uncertain Information

GAO Liang¹, SUN Wei^{2,3}, ZHU Rongchang¹

(1. Engineering Institute, Air Force Engineering University, Xi'an 710038, China; 2. Zhijiang College, Zhejiang University of Technology, Hangzhou 310014, China; 3. Institute of Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: In view of the deficiency to directly calculate the relational degree between time series of uncertain information, an approach to analyze the relational trend between series of interval numbers is proposed. The approach quantifies the difference between uncertain information using the distance of interval numbers, and matches the series through time shifting and inversion transformation. Then the relational degree of time series is calculated based on the distance between time series of interval numbers, and the relational trends of uncertain information series are classified through the method of set pair analysis. The approach extends the application range of relational degree from crisp numbers to uncertain environment expressed by interval numbers, and classifies the degree of relational trend between time series. Comparisons with both the C-means algorithm and the simulation anneal algorithm show that the accuracy of the proposed algorithm is higher; the false alarm rate and the missing alarm rate are both lower; and the running time is reduced by 77% and 63.4% respectively.

Keywords: time series; relational degree; interval number; set pair analysis

时间序列的趋势分析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. 时间序列趋势分析方法的物理意义直观, 适用于大量序列的关联分析^[1-6], 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. 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,数据预处理过程造成了数据信息的损失^[7-8].

本文从区间数表示的不确定信息出发,将序列关联度计算从精确数范围推广至区间数范围,结合集对分析法,提出了区间数序列关联趋势分析法(RTASI法).RTASI法可以直接对区间数表示的不确定序列进行关联性分析,与常用的聚类算法比较,具有分类准确率高,虚警率和漏警率低,算法运行时间短的特点.

1 区间数和集对分析法

1.1 区间数的距离

在信息分析中常常将含有不确定性,分布规律无法确切得知的数据用区间数表示^[9].

定义 1 称实数域 $\mathbf{R}$ 中的闭区间 $[x_{\text{low}}, x_{\text{up}}]$ 为区间数,记作 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 I_x 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, $w_x = (x_{\text{up}} - x_{\text{low}})/2$ 为区间数的半宽.当且仅当 2 个区间数的中心和半宽都相等时,区间数相等.

区间数之间的距离没有直观的解释,即使 2 个非 0 区间数的上下限相等,也不能得出这 2 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 3 个区间数为 I_x, I_y 和 I_z ,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 $d(I_x, I_y)$ 为区间数 I_x 和 I_y 之间的距离度量(简称为距离).

区间数的中心反映了区间的位置,半宽显示了信息的不确定程度,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 2 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 2 个区间数,则区间数 I_x 和 I_y 之间的距离为

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

式中: $n \in \mathbf{N}$, 通常取 $n=2$.

证明 只有满足距离度量的 3 条公理,定义 2 才是有效的.公理 1 和公理 2 显然成立,下面证明公理

3,即需证明

$$[|o_x - o_z|^n + |w_x - w_z|^n]^{1/n} \leq [|o_x - o_y|^n + |w_x - w_y|^n]^{1/n} + [|o_y - o_z|^n + |w_y - w_z|^n]^{1/n}, n \in \mathbf{N}$$

当 $n=1$ 时,由实数域内三角不等式知其显然成立.

当 n 大于 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 $q=n/(n-1)$,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 2 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.

1.2 序列的时移和翻转

定义 3 设 $I_x(t), I_y(t)$ 分别为时间序列 I_x 和 I_y 在时刻 t 的区间值,即时间序列为 $I_x = \{I_x(0), \dots, I_x(i), \dots, I_x(n)\}$, $I_y = \{I_y(0), \dots, I_y(j), \dots, I_y(m)\}$, $i, j, n, m \in \mathbf{N}$,设 $0 \leq m \leq n$,其中 n 和 m 分别称为序列 I_x 和 I_y 的时间长度,则序列之间的距

$$\text{离 } d(I_x, I_y) = \sum_{t=0}^m d[I_x(t), I_y(t)].$$

将反映已知系统状态(一般为设备正常工作状态)的测量序列称为基准序列,反映未知系统状态的测量序列称为测试序列.通过测试序列与基准序列的趋势比过来解释测试序列代表的状态之前,需要对测试序列进行时移或翻转转换,从而在基准序列中找到与测试序列最相似的时间段.

假设 I_x 为基准序列, I_y 为测试序列,对序列 I_y 进行时移或翻转转换,使转换后的序列与 I_x 距离最小,将 I_x 中与 I_y 状态最接近的这段序列起点作为

序列关联分析的起始时刻. 寻找 I_x 中最接近 I_y 序列起始时刻的过程如下:

- (1) 计算序列 I_x 的翻转序列 I_x^- 在时间点 t 的区间数 $[-x_{\text{low}}(t) + \max_{i=1, \dots, n}(x_{\text{up}}(i)), -x_{\text{up}}(t) + \max_{i=1, \dots, n}(x_{\text{up}}(i))]$;
- (2) 使序列 I_y 依次由 0 到 $n-m$ 时刻依次滑动, 得到 $n-m+1$ 组序列 $I_y^j = \{I_y(t+j) = I_y(t), t=0, \dots, m; j=0, \dots, n-m\}$, 计算 $d(I_x, I_y^j)$ 以及 $d(I_x^-, I_y^j)$;
- (3) 取 $\min_{j=0, \dots, n-m} \{d(I_x, I_y^j), d(I_x^-, I_y^j)\}$ 时的 j 作为基准序列的起始时刻.

经过上面的遍历搜索过程, 可以找到序列 I_x 中和 I_y 最接近的序列段.

1.3 区间距离的规范化

在 t 时刻, 序列 I_x 和 I_y 的距离规范化公式为 (时间长度为 m)

$$d^*[I_x(t), I_y(t)] = \frac{d[I_x(t), I_y(t)] - \min_{i=1, \dots, m} \{d[I_x(i), I_y(i)]\}}{M}$$

式中

$$M = \max_{i=1, \dots, m} \{d[I_x(i), I_y(i)]\} - \min_{i=1, \dots, m} \{d[I_x(i), I_y(i)]\}$$

当 $d[I_x(t), I_y(t)]$ 恒定时, 令 $d^*[I_x(t), I_y(t)] = 1$.

1.4 集对分析法

集对分析法^[10]从同、异、反的角度对 2 种事物的确定和不确定关系进行了量化描述.

定义 4 给定 2 个集合 A 和 B , 并设这 2 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 A 和 B 在问题 W 下的差异度 (以 b 表示), p/n 为 A 和 B 在问题 W 下的对立度 (以 c 表示). A 和 B 的联系度为

$$\mu_w(A, B) = \frac{s}{n} + \frac{f}{n}i + \frac{p}{n}j = a + bi + cj \quad (2)$$

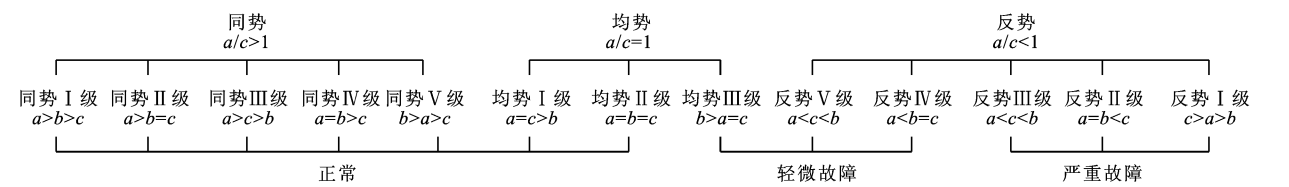


图 1 同、异、反关系分级及设备状态分类

式中: $a+b+c=1, 0 \leq a, b, c \leq 1; j$ 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 3 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 3 种. 具体分类如图 1 所示.

2 RTASI 法

用灰色关联度 r 度量序列之间的相关性^[11-12], 如果 2 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 $0 \leq r \leq 1$, 则 1 表示序列之间变化趋势完全一致, 0 表示无关联.

定义 5 设 $I_x(t), I_y(t)$ 分别为时间序列 I_x 和 I_y 在时刻 t 的区间值, $t=1, \dots, m$, 若 M_1, M_2 和 M_3 都存在, 则 I_x 和 I_y 的关联度为

$$r = \left[1 - \cot \frac{\pi M_1}{2m} - \cot \frac{\pi M_2}{2(m-1)} - \cot \frac{\pi M_3}{2(m-2)} \right]^{-1} \quad (3)$$

式中

$$M_1 = \sum_{t=1}^m d^*[I_x(t), I_y(t)]$$
$$M_2 = \sum_{t=1}^{m-1} |d^*[I_x(t+1), I_y(t+1)] - d^*[I_x(t), I_y(t)]|$$
$$M_3 = \sum_{t=2}^{m-1} |d^*[I_x(t+1), I_y(t+1)] - 2d^*[I_x(t), I_y(t)] + d^*[I_x(t-1), I_y(t-1)]|$$

定义 5 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 5.

在故障诊断应用中, $d^*[I_x(t), I_y(t)]$ 等于 1 时, 表示序列 I_x 和 I_y 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; $d^*[I_x(t), I_y(t)]$ 为 0 时, 表示序列 I_x 和 I_y 在时刻 t 的状态完全相同, 设备工作正常; $d^*[I_x(t), I_y(t)]$ 在 0 到 1 内变化时, 表示序列 I_x 和 I_y 在时刻 t 的状态由完全相同至完全不同变化. 因此, 构造序列 $\{1 - d^*[I_x(t), I_y(t)]\}$ 来判断设备工作状态.

假设精确数最优序列为 $(\overbrace{1, \dots, 1}^m)$, 则序列 $\{1 - d^*[I_x(t), I_y(t)]\}$ 与这个最优序列的关联度为

$$r^+ = \left\{ 1 - \cot\left[\pi \sum_{t=1}^m |d^*[I_x(t), I_y(t)]| / 2m\right] - \cot\left[\pi \sum_{t=1}^{m-1} |d^*[I_x(t+1), I_y(t+1)] - d^*[I_x(t), I_y(t)]| / 2(m-1)\right] - \cot\left[\pi \sum_{t=2}^{m-1} |d^*[I_x(t+1), I_y(t+1)] - 2d^*[I_x(t), I_y(t)] + d^*[I_x(t-1), I_y(t-1)]| / 2(m-2)\right] \right\}^{-1}$$

假设精确数最劣序列为 $(\overbrace{0, \dots, 0}^m)$, 则序列 $\{1 - d^*[I_x(t), I_y(t)]\}$ 与这个最劣序列的关联度为

$$r^- = \left\{ 1 - \cot\left[\pi \sum_{t=1}^m |1 - d^*[I_x(t), I_y(t)]| / 2m\right] - \cot\left[\pi \sum_{t=1}^{m-1} |d^*[I_x(t+1), I_y(t+1)] - d^*[I_x(t), I_y(t)]| / 2(m-1)\right] - \cot\left[\pi \sum_{t=2}^{m-1} |d^*[I_x(t+1), I_y(t+1)] - 2d^*[I_x(t), I_y(t)] + d^*[I_x(t-1), I_y(t-1)]| / 2(m-2)\right] \right\}^{-1}$$

因此, 有

$$(r^+)^{-1} - (r^-)^{-1} =$$

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

令 $(r^+)^{-1} - (r^-)^{-1} = \Phi$, 则 $\frac{d(r^+ + r^-)}{dr^-} = 1 + (\Phi r^- + 1)^{-2} > 0$, 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

$$(\Phi + 1)(\Phi + 2)^{-1}r^+ + (\Phi + 1)(\Phi + 2)^{-1}r^- \leq 1 \tag{4}$$

式中: $(\Phi + 1)(\Phi + 2)^{-1}r^+$ 反映了测试序列与基准序列的接近程度; $(\Phi + 1)(\Phi + 2)^{-1}r^-$ 反映了测试序列与基准序列的偏离程度, 将它们分别定义为集对的同一度 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.

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 至 23 组)、存在轻微故障(第 24 至 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的, 所以分析中不考虑时移和翻转转换. 算法运行环境为: CPU 2.4 GHz, 内存 2 GB 的 PC 中采用 C++ 进行算法编程.

在 RTASI 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. 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. RTASI 法、C 均值法和模拟退火法分析数据的结果如表 1 所示.

表 1 测试数据的聚类结果

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

采用以下 3 个指标衡量算法性能:

(1)分类正确率=(分类正确数/总测试数据数);

(2)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];

(3)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].

3 种算法的性能比较见表 2.

表 2 3 种算法的性能比较

算法	分类正确率/%	虚警率/%	漏警率/%	运行时间/ms
RTASI 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 1 和 2 可以看出,RTASI 法分类正确率最高,发生了 1 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 RTASI 法不涉及最优值求解,算法运行时间比其他 2 种算法的运行时间分别减少了 77%和 63.4%.

从 RTASI 法的计算结果中任意取出 5 组进行集对分析. 各组计算的同一度 a 和对立度 c 如图 2 所示.

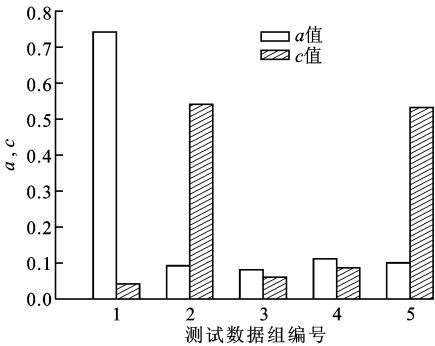


图 2 各组测试序列的 a 和 c 值的比较

由图 2 可以看出:第 2 和 5 组的 c 明显大于 a ; 第 1 组的 a 远大于 c ;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 2 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 2 个序列处于和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做经常性的记录分析. 序列 1 和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. RTASI 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缺少的.

4 结 论

时序状态的定性分析是快速分析大量传感器信息的有效方法,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、模糊语言等表示方法已成为分析这些信息的有效手段. 本文对不确定信息条件下的离散时间序列关联趋势进行了研究,实例分析说明 RTASI 法对故障诊断和预测具有理论价值,结果证明了该算法的有效性. 本文方法在进行序列关联分析时还缺少定量指标划分的界限,下一步研究将主要集中在分析关联度阈值,以及阈值和同、异、反关系界限之间的联系,并进一步研究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] MAURYA M R, RENGASWAM Y R, VENKATA-SUBRAMANIAN V. Fault diagnosis using dynamic trend analysis; a review and recent developments [J]. Engineering Applications of Artificial Intelligence, 2007, 20(2):133-146.

[2] BALASKO B, BANKO Z, ABONYI J. Analyzing trends by symbolic episode representation and sequence alignment [C] // Proc of 15th Mediterranean Conference on Control and Automation, Piscataway, NJ, USA; IEEE Press, 2007; 605-610.

[3] DERELI T, DURMUSOGLU A. A trend-based patent alert system for technology watch [J]. Journal of Scientific & Industrial Research, 2009, 68(8): 674-679.

[4] MAURYA M R, RENGASWAMY R, VENKATA-SUBRAMANIAN V. A signed directed graph and qualitative trend analysis-based framework for incipient fault diagnosis [J]. Chemical Engineering Research & Design, 2007, 85(10): 1407-1422.

[5] VILLEZ K, PELLETIER G, ROSEN C, et al. Comparison of two wavelet-based tools for data mining of urban water networks time series [J]. Water Science and Technology, 2007, 56(6): 57-64.

[6] CHARBONNIER S, GARCIA-BELTAN C, CADET

- C. Trends extraction and analysis for complex system monitoring and decision support [J]. *Engineering Applications of Artificial Intelligence*, 2005, 18(1): 21-36.
- [7] ZHU Hongning, ZHANG Bin, GAO Xianwen. A service selection method of non-critical task in parallel structure [C]//*Proc of 2nd International Workshop on Knowledge Discovery and Data Mining*. Piscataway, NJ, USA: IEEE Press, 2009: 846-850.
- [8] WONG Heung, HU Baoqing, 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] ZHU Jianjun, LIU Sifeng, LI Lihong. Research on four kinds of uncertain preference information aggregation approach in group decision making [J]. *Kybernetes*, 2009, 38(10): 1862-1869.
- [10] LI Xiaobai, 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] LI Zhijun, ZHANG Lei, CHEN Mianyun, et al. Research and application of data mining method based on gray trend relational analysis [C]//*Proc of 2008 Pacific-Asia Workshop on Computational Intelligence and Industrial Application*. Piscataway, NJ, USA: IEEE Press, 2008: 98-101.
- [12] FU Mingfu, XIE Mingxiang, 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] SUDKAMP T. Examples, counterexamples, and measuring fuzzy associations [J]. *Fuzzy Sets and Systems*, 2005, 149(1): 57-71.
- [14] 温熙森. 模式识别与状态监控[M]. 北京: 科学出版社, 2007: 245-248.

(编辑 刘杨)