

# A Fast Algorithm for Waveform Pattern Matching 波形模式匹配的一种加速算法\*

李德华等\*\*

(华中工学院)

〔提要〕 本文提出了用于某些波形识别模板匹配的一种模式识别与阈值加速算法。这种方法的基本思想是将波形的幅度期望值和波形的上升、下降、水平的波动特征构成筛选阈值和特征向量，对已知的波形集合进行筛选，仅对满足阈值要求并且与特征向量完全相同的波形才进行模板匹配的相关计算。对于有一定特征的波形集合信息压缩率达到 90~98%，提高计算速度 10~30 倍。

## 一 问题的提出

设  $f(t)$  为输入波形(图1)，根据采样定理可将其离散化，得到序列：

$$\{f(t_j)\} = \{f(t_0), f(t_1), \dots, f(t_{p-1})\}$$

$$j = 0, 1, 2, \dots, p-1 \quad (1)$$

设  $F = \{F_i(t)\} i = 1, 2, \dots, N$  为某已知基准波形集合，其中  $F_i(t)$  可离散化为：

$$\{F_i(t_j)\} = \{F_i(t_0), F_i(t_1) \dots F_i(t_{p-1})\}$$

$$j = 0, 1, 2, \dots, p-1 \quad (2)$$

识别输入波形  $f(t)$  实质上是将其作为一个模板对  $F$  集合中之元素一一进行匹配计算，找出与  $f(t)$  相关程度最高者，即可认为  $f(t)$  与之同属

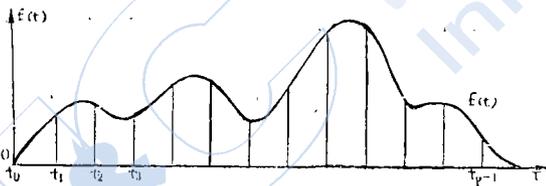


图 1 输入波形  $f(t)$

一类(或全同)。一般可以采用互相关算法，平均绝对差(MAD)算法和平均方差算法(MSD)互相关算法公式为：

$$G_i = \frac{K_{f(t), F_i(t)}}{\sigma_{[f(t)]} \cdot \sigma_{[F_i(t)]}}$$

$$= \frac{\frac{1}{p} \sum_{j=0}^{p-1} [f(t_j) - \overline{f(t)}]}{\sqrt{\frac{1}{p-1} \sum_{j=0}^{p-1} [f(t_j) - \overline{f(t)}]^2}} \cdot \frac{[F_i(t_j) - \overline{F_i(t)}]}{\sqrt{\frac{1}{p-1} \sum_{j=0}^{p-1} [F_i(t_j) - \overline{F_i(t)}]^2}} \quad (3)$$

其中  $K_{f(t), F_i(t)}$  是  $f(t)$  与  $F_i(t)$  的协方差；

$\sigma_{[f(t)]}$  是  $f(t)$  的均方差；

$\sigma_{[F_i(t)]}$  是  $F_i(t)$  的均方差；

$\overline{f(t)}$ 、 $\overline{F_i(t)}$  分别是  $f(t)$  与  $F_i(t)$  的平均值。平均绝对差的公式(MAD)为

$$G_i' = \frac{1}{p} \sum_{j=0}^{p-1} |f(t_j) - F_i(t_j)| \quad (4)$$

平均方差的公式(MSD)为

$$G_i'' = \frac{1}{p} \sum_{j=0}^{p-1} [f(t_j) - F_i(t_j)]^2 \quad (5)$$

判断准则是：

如果  $G_i = \max[G_j], j \in I$

则  $f(t) \triangleq F_i(t)$

如果  $G_i' = \min[G'_j], j \in I$

则  $f(t) \triangleq F_i(t)$

如果  $G_i'' = \min[G''_j], j \in I$

则  $f(t) \triangleq F_i(t)$

\* 本文曾在 1981 年中国自动化学会模式识别与机器智能委员会学术会上宣读。

\*\* 本文的作者还有华中工学院的胡昌赤、余苏宁、钱进、方放四同志。

这里  $I = \{1, 2, \dots, N\}$ ,  $\triangle$  表示同类。

很显然当  $N$  充分大时要识别  $f(t)$  必须对  $F$  集合进行全局搜索, 计算量大, 效率很低, 必须设法进行信息压缩, 提高效率。

## 二 波形的筛选特征

如果两波形完全类同, 那么它们的采样时间对应点的上升、下降、水平的形态模式特征应该是相近或一样的。因此我们可以求出  $f(t)$  的形态模式特征对  $F$  集合中的  $N$  个波形进行识别预选, 凡是符合筛选特征要求的才用(3)、(4)、(5)进行相关识别计算, 最后决定其类属。否则抛弃, 这样可以达到信息压缩的目的。

在没有噪声的情况下, 我们可以求出:

$$\text{SGN}[f(t_{j+b}) - f(t_j)] = \begin{cases} 1 & \text{如果 } t_j \text{ 到 } t_{j+b} \text{ 两点} \\ & \text{间曲线幅值上升} \\ 0 & \text{如果 } t_j \text{ 到 } t_{j+b} \text{ 两点} \\ & \text{间曲线幅值相同} \\ -1 & \text{如果 } t_j \text{ 到 } t_{j+b} \text{ 两点} \\ & \text{间曲线下降} \end{cases} \quad (6)$$

$$b = 1, 2, \dots, \beta < P$$

$$j = 0, 1, \dots, p - b - 1$$

(6) 式的值可以表示出间隔为  $b$  的两点波形幅值的上升、下降和水平的变化。

$$\text{令 } h_{bj} = \text{SGN}[f(t_{j+b}) - f(t_j)]$$

$$j = 1, 2, \dots, p - b - 1$$

实际上只需选出若干个  $h_{bj}$  就可构成筛选特征向量, 例如选  $r$  个, 有:

$$\vec{h} = (h_1, h_2, \dots, h_r) \quad (7)$$

在有噪声存在的情况下, 就有可能因噪声的叠加改变了原有波形的  $b$  间隔点间波动形态 (如原为上升的变为下降, 原为下降的变为上升), 如果把这样的值也作为筛选特征向量的元素, 就会出现误识别现象。因此就必须修改式(6), 把即使叠加了噪声, 但上升下降趋势仍然不改变的点找出来, 构造  $\vec{h}$ 。

如果噪声是方差为  $\sigma_n$ , 均值为 0 的正态随机变量, 根据三倍均方差原则, 叠加噪声值

达  $|3\sigma_n|$  的概率仅为 0.003; 而  $f(t_{j+b})$  与  $f(t_j)$  叠加相反方向噪声的概率一般为 0.5; 所以这两点叠加方向相反, 并且大于等于  $|3\sigma_n|$  的噪声的概率  $\ll \frac{1}{1000}$  (约为 0.0000045) 即

$$P(|f(t_{j+b}) - f(t_j)| > 6\sigma_n) \ll \frac{1}{10000} \quad (8)$$

所以我们可以用  $6\sigma_n$  作为一个门限值

$$\text{如果 } |f(t_{j+b}) - f(t_j)| > 6\sigma_n \quad (9)$$

$$\text{则 } h_{bj} = \text{SGN}[f(t_{j+b}) - f(t_j)] \quad (10)$$

在研究工作中, 我们常遇到有的波形当  $b = 1$  时, 找不到满足条件(9)的点, 而在  $b = 2, 3$  情况下可找到许多满足(9)的点。根据我们的经验常常当  $b = 3$  时满足(9)的点比  $b = 1$  多。令  $\sigma_{z_b}$  为间隔为  $b$  点的粗糙度:

$$\sigma_{z_b} = E\{[f(t_{j+b}) - f(t_j)]^2\} \quad (11)$$

如果  $b$  不太大 (例如 3) 且  $b > b_1 > b_2$ , 则常有

$$\sigma_{z_b} \geq \sigma_{z_{b_1}} \geq \sigma_{z_{b_2}}$$

因此我们先算  $b \gg 1$  的情况, 如若找不到满足(9)的点, 然后, 再往  $b$  小的方向进行搜索。一般总能找出一批点构造  $\vec{h}$ 。从  $\{h_{bj}\}$  中选取  $r$  个构成

$$\vec{h} = (h_1, h_2, \dots, h_r) \quad (12)$$

筛选向量。

为叙述的方便我们定义算子  $H^*$ , 它具有上述(9)(10)(12)的功能, 即:

$$H^*[f(t)] = \vec{h} = (h_1, h_2, \dots, h_N) \quad (13)$$

$$H^*[F_i(t)] = \vec{h}_i = (h_{i1}, h_{i2}, \dots, h_{iN}) \quad (14)$$

## 三 波形的筛选阈值

如前述, 待识别波形  $f(t)$  根据所述准则分类判决, 可归属于  $F$  中的某个类别。

设  $f(t)$  判决为  $F_i$  类 ( $F_i \in F$ ), 则  $f(t)$  与  $F_i$  之间的一些差异, 往往比它与  $F$  中其余元素之间的差异小。

$f(t)$  与  $F_i$  的平均波幅差即是其一。显然, 如果两个波形完全类同, 或差异甚小, 则其平均波幅差应为 0 或甚小。

因此, 可以利用平均波幅差作为阈值, 在

进行全集合内的匹配计算时,凡落在  $f(t) \pm$  阈值  $H_0$  范围之外的基准波形类别均可抛弃,从而缩小搜索范围,达到信息压缩之目的(见图 2)。

阈值可由统计计算,计算机模拟等方法确定。下面以计算机模拟为例,说明阈值的计算选取方法。设  $f(t)$  是  $F$  中的某类波形,按一定误差模型迭加而生成(例如  $f(t)$  测取时的各种误差因素及干扰等),而这些误差模型是已知的,因之,  $f(t)$  是  $F$  中的某个类别。

这样,就可以由基准波形集合  $F$ , 通过一系列计算,得到符合误差模型的与之成为映射关系的待识别波形集合  $R$ 。

设集合  $F$  为

$$F = (F_1, F_2, \dots, F_i, \dots, F_n)$$

经过对  $F_i$  的  $K$  种误差迭加后,得到一一对应的“拟待识别”波形矩阵  $R$ :

$$R = (f_1, f_2, \dots, f_i, \dots, f_n)$$

$$ER = (Ef_1, Ef_2, \dots, Ef_i, \dots, Ef_n)$$

$$EF = (EF_1, EF_2, \dots, EF_i, \dots, EF_n)$$

式中,  $F_i$  —— 前述之式 (2)

$f_i$  —— 由  $F_i$  按已知误差模型生成的“拟待识别”波形

$$EF_i = \frac{1}{P} \sum_{j=1}^P F_i(t_j) \text{—— 已知波形波幅均值}$$

$$Ef_i = \frac{1}{P} \sum_{j=1}^P f_i(t_j) \text{—— “拟待识别” 波形波幅均值}$$

幅均值

$$E = EF - ER = (E_1, E_2, \dots, E_i, \dots, E_n)$$

式中:  $E_i$  为  $F_i$  与  $I_i$  平均波幅差。

令  $H_0 = \max(E_i)$

则  $H_0$  为我们所选之阈值。

#### 四 两级信息压缩加速匹配计算

首先用算子  $H^*$  作用于  $F$  集合求出筛选特征向量

$$H^*[F_i(t)] = \vec{h}_i = (h_{i1}, h_{i2}, \dots, h_{ir})$$

$$i = 1, \dots, r$$

并计算出  $H_0$ 。

第一级信息压缩:以筛选阈值  $H_0$ ,进行预处理  
若  $|\overline{f(t)} - \overline{F_i(t)}| > H_0$  则抛掉波形  $F_i(t)$ ,  
因为两平均幅值相差太大;

若  $|\overline{f(t)} - \overline{F_i(t)}| < H_0$  则转入第二级信息压缩;

第二级信息压缩:用筛选特征向量  $\vec{h}$  进行预处理

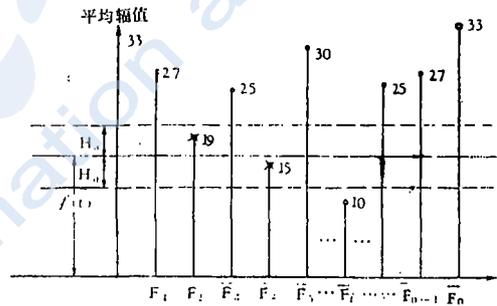
若  $\vec{h} = \vec{h}_i$  则进行(3)(4)(5)式的相关计算

若  $\vec{h} \neq \vec{h}_i$  则抛弃波形  $F_i(t)$ ,继续对  $F_{i+1}$

( $t$ ) 重复上述筛选过程。

我们也可以通过计算机模拟和对  $F$ , 对噪声分析研究,相当准确地确定  $H_0$  阈值,使筛选效率大为提高。可满足实时性的要求。

我们采用上述模式识别和阈值加速方法对波形数为 2500~3000 的集合进行模板匹配计算,在选定  $b=1, 2, 3$ ;  $r=5$  的情况下,筛选率高达 90~98% (甚至 99.7%) 大大压缩了信息处理量,效率提高 10~30 倍,而且由于运用此法除掉了一些有可能用相关法会混淆的波形,因而降低了误识别率。



· —— 为可抛弃波形, × —— 需进行相关运算之波形;  
 $\overline{F_i}$  —— 已知波形平均幅值;  $H_0$  —— 筛选阈值  
 $\overline{f(t)}$  —— 待识别波形平均幅值

图 2

(上接18页)

#### 附录 2

设对应底界为  $\omega_m$  的侧界线座标为

$$E_1 = rE_{1m} = rK_1K_2T^2M_m \quad (2.1)$$

令(25)中  $e_1 = E_1$ , 从而有底界座标

$$\omega_m = \sqrt{2r} K_1 T M_m \quad (2.2)$$

令  $E_1 = e_1(0)$ ,  $\omega_m = \omega(0)$ , 代入(22), 得:

$$M(0) = (r - 3\sqrt{2r}/2)M_m \quad (2.3)$$

$$M(T) = (-r + \sqrt{2r}/2)M_m \quad (2.4)$$

要求  $M(0) \geq -M_m$ , 则由(2.3)解得:

$$r \leq 1/2 \quad (2.5)$$

由此可见当底界为  $\omega_m$  时, 侧界为  $1/2E_{1m}$ 。

## **Frame Structure and Categories in Compatible Telecontrol System**

*Yu Ning, Tan Wenshu*

Some problems about compatible telecontrol systems are discussed in this paper. Frame structures and frame categories suitable for compatible telecontrol system are proposed. In addition, a 8080 microprocessor-based compatible telecontrol system has been designed. (pp.31—36)

## **A Fast Algorithm for Waveform Pattern Matching**

*Li Dehua*

An algorithm for pattern recognition and the calculation of the threshold value has been proposed to speed up the pattern matching of certain waveforms.

The main idea is that from the expected amplitude and the vertical and horizontal fluctuation of the waveform a characteristic vector is derived and the threshold value calculated. For certain a set of waveforms the matching operation is carried out only for those waveforms that have the same vector as the pattern, and are within the threshold value. For some waveforms, the information compression ratio reaches 90~98%, the speed of the computation is increased by a factor of 10—30. (pp.37—39)

## **Discussion on "The Law of Requisite Variety"**

*Wang Dingchang*

This paper comments on Sahal's and Aulin-Ahmarara's inference and some points at issue about "The Law of Requisite Variety". Recent applications are reviewed with emphasis on its importance in social control systems. (pp.40—45)

## **Automatic Control Device for Feeding Raw Material to Blast Furnace**

*Liu Guorong, Yao Gexin*

This paper discusses the working principle of a raw material feeding device used in blast furnace No.1 and No.2 at the Wuhan Iron and Steel company. The methods to improve its reliability and the effect of its use are also discussed. The device is noted for using common components and simple electric scheme. The circuitry of the device is quite different from the conventional one in putting electronic and power circuits together to achieve improved reliability in operation. (pp.57—60)

## **Cascade Control of Air Supply to Oil Boilers**

*Shen Botao*

An automatic cascade control system for air supply to oil boiler has been designed around a wing air flowmeter. The experiences in the field and methods of adjustment are described in detail. (pp.61—62)