

面向航天器测试的时序数据模式表示方法研究^{*}

周家杰^{1,2}, 余丹¹, 马世龙¹, 陈丽萍²

(1. 北京航空航天大学软件开发环境重点实验室, 北京 100191; 2. 中国人民解放军 91880 部队, 山东 青岛 266300)

摘要: 通过计算某一点与其左右两相邻点斜率的比值确定出变化点, 连接这些变化点, 就得到一种基于斜率变化阈值的时间序列分段线性 STC 表示算法。来自航天器不同分系统的真实数据集实验表明, 使用 STC 算法作为航天器测试时间序列的模式表示方法, 与其他分段线性表示算法相比, 该方法具有算法简单、拟合程度高和适应能力强的特点。

关键词: 航天器测试; 时序数据; 数据挖掘; 模式表示; 分段线性表示

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2011)01-0127-04

doi: 10.3969/j.issn.1001-3695.2011.01.035

Spacecraft test oriented pattern representation of time series

ZHOU Jia-jie^{1,2}, YU Dan¹, MA Shi-long¹, CHEN Li-ping²

(1. State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China; 2. 91880 PLA Troops, Qingdao Shandong 266300, China)

Abstract: This paper proposed a method of piecewise linear representation of time series based on slope change threshold (SCT). The algorithm calculated the two segments' slope ratio of the certain point connecting with the two adjacent points and then it combined the points with the line orderly. Detailed experiments on real datasets from various subsystems of spacecraft show that STC, as the selected method, is simpler, has a higher degree of fitting and stronger adaptability than other methods of piecewise linear representation.

Key words: spacecraft test; time series data; data mining; pattern representation; piecewise linear representation

航天器实验过程中会产生大量的时间序列测试数据, 这些测试数据是按照时间先后顺序排列的各个测试记录的有序集合, 通常具有如下特点: a) 非平稳性。它是指航天器在其测试过程中既有平稳的时间阶段, 同时在某一阶段或几个阶段上又有突变现象, 即有些变量的变化幅度在某一或某几个时间段中有突然增大的现象。例如, 在航天器测试过程中, 因为测试接口电路在阻抗上的不匹配就会造成信号幅度小, 相位不匹配就会造成信号异常、脉冲多或少、开门信号宽或窄、互相“串电”等, 以及在电测设备之间、系统之间可能会存在或多或少的相互干扰, 这些因素都可能会导致测试参数值的突变^[1]。b) 波动幅度随时间变化。一般来说, 对于要求标称值在某一范围内的航天器测试参数来说, 它的方差不是固定不变的, 而是随时间的变化而变化, 即该参数离其期望值的距离是随时间变化的。c) 海量、超高维数。在航天器测试过程中, 单个参数在每秒钟内会有多个数据下传入库。同时, 为了保证航天器测试的可靠性, 航天器测试中需要对航天器的多项参数进行详细反复的测试, 这样, 在航天器的持续测试过程中就会产生海量的测试数据, 理论估计多航天器并行测试中产生的年测试数据量不低於 2 TB^[2]。d) 存储形式多样。由于各航天器测试部门的信息化程度不同, 航天器测试数据有多种存储形式, 如数据库、数据文件、数据报文等, 而仅数据库这一类中就有 Oracle、SQL Server 等多种数据库管理系统软件, 同样, 数据文件也存在着

多种文件格式。

研究如何有效地从这些复杂的海量时间序列中挖掘潜在的有用知识, 一直是航天器测试数据分析人员关心的问题。鉴于航天器测试中得到的时序数据具有上述特点, 如果直接在其时间序列上进行数据挖掘, 不但在存储和计算上要花费高昂的代价, 而且可能会影响算法的准确性和可靠性。因此, 许多研究者提出了时间序列的模式表示方法。该方法就是将时间序列数据通过模式变换得到新的时间序列, 来刻画时间序列的主要形态而忽略微小细节, 这样不仅可以对时序数据进行压缩, 减少数据的存储和计算的代价, 而且可以保留原时间序列的主要形态, 忽略那些细小而不改变性质的变化, 提高数据挖掘的效率和准确性, 更好地为数据分析人员提供决策支持。

1 相关工作

目前, 常用的时间序列模式表示方法主要有频域表示法、符号表示法、奇异值表示法和分段线性表示法等。

频域表示法是将一条时间序列看做是时间域上的一个信号, 采用正交变换, 把信号变换到频率域上, 然后忽略那些幅度值很小的对时间序列形态很小的频率, 这样就得到时间序列一种新的表示方法。变换后的频域序列可以再通过逆变换到时间域上, 得到时间序列的一个近似表示。该方法主要分为两

收稿日期: 2010-07-05; **修回日期:** 2010-08-13 **基金项目:** 软件开发环境重点实验室探索性自主研究课题资助项目 (SKLSDE-2009ZX-13)

作者简介: 周家杰 (1979-), 男, 山东青岛人, 硕士研究生, 主要研究方向为时序数据挖掘、航天器自动化测试 (zjjclp201@126.com); 余丹 (1979-), 湖北武汉人, 博士, 主要研究方向为航天器自动化测试、安全苛刻性系统、可信软件; 马世龙 (1953-), 男, 北京人, 博导, 主要研究方向为网络计算、海量信息处理等; 陈丽萍 (1977-), 女, 甘肃天水人, 工程师, 主要研究方向为自动化测试。

种:a)通过离散傅里叶变换^[3,4],将时间序列从时域空间映射到频域空间;b)采用离散小波变换^[5]。

符号表示法是一种离散化方法,就是通过一些离散化方法将时间序列的实数值或者一段时间内的时间序列波形映射到有限的符号表上,将时间序列表示为有限符号的有序集合。Park 等人^[6,7]采用等宽离散化和最大熵离散化方法将时间序列的实数值映射到有限的符号,以后缀树为索引,提出了一种动态时间弯曲距离的下界距离,在提高查询效率的同时保证了查询的完备性。他们的工作使得基于符号化表示的时间序列相似性查询能够支持比欧式距离更鲁棒的动态时间弯曲距离。Lin 等人^[8]在 2003 年提出的新型的符号化表示方法——符号聚合近似 SAX 方法,比其他符号算法更简便、高效,在符号化过程中实现了减维降噪,它主要适合于遵循高斯分布且在有限方差范围内有较高分布密度的时序数据。钟清流等人^[9]为克服 SAX 方法对时序信息描述不完整的缺陷,提出基于统计特征的时序数据符号化算法。与 SAX 不同的是,该算法将时序符号看做矢量,而各时序子段的均值和方差则分别作为描述其平均值及发散程度的分量。

奇异值表示法是利用奇异值分解的方法对时序数据进行降维,已经成功用于图像和文本的索引^[10,11]。该方法通过分析所有时间序列,计算新的坐标系,使得第一条坐标轴对应最大的方差,第二条坐标轴对应次大的方差并与之前的坐标轴正交,依次得到所有的坐标轴,根据这些坐标轴将时间序列从原始空间变换到新的坐标空间。奇异值表示法在数据重构上误差最小^[12],这使得该方法在一些情况下能够取得很好的性能。但其时间复杂度为 $O(Mn^2)$ 。其中: M 是指时间序列数据库的大小, n 是指时间序列的平均长度。当插入或删除一条时间序列时,时间代价很高。

分段线性表示法是从时间序列中抽取一些特征点,将这些特征点依次相连构成的线段序列就是时间序列的分段线性表示。Prat 等人^[13]提出了基于重要点的分段方法。重要点被定义为在局部范围内的极值点,并且与端点的比值超过参数 R 。将重要点用线段连接,就得到了基于重要点的时间序列分段线性表示,通过选择不同的参数,可以获得精细粒度不同的分段线性表示。Keogh 等人^[14]和 Yi 等人^[15]分别独立提出了时间序列的分段聚集近似 PAA 表示方法,将时间序列等宽度划分,每个子段用时间序列在该子段上的平均值来表示。该方法简单直观,能够支持任意长度的相似性查询以及加权欧氏距离,而且能够用于索引以提高查询的效率。Xiao 等人^[16]提取时间序列的特征点作为时间序列的分段点,通过连接这些特征点,得到时间序列的分段线性表示,它的输入参数为 v ,表示特征区间大小。詹艳艳等人^[17]将线段的斜率与时间序列的特点结合起来,根据斜率的变化幅度选取一些时间序列的边缘点,将这些边缘点依次用线段连接,就得到了时间序列的一种分段线性表示,称为时间序列的 SEEP 表示。

通过上述文献分析表明,离散傅里叶变换适合那些自然发生的正弦信号,但不适合表示不连续间断的信号。在离散小波变换方法中,Haar 小波变换是最常用的,但由于基函数不光滑,采用阶梯状的结构近似地模拟信号,需要保留较多的小波系数的数目才能很好地近似连续函数。符号表示法因其离散化、非实数表示的特点得到了越来越多的关注。其不足之处在于如何对时间序列进行离散化,而符号表大小的选择也是个难点。奇异值表示法是一种依赖于数据内容的降维方法,通过计算给定数据集的特征值和特征向量将数据进行转换,使得大多

数信息集中在某些维上,取数据在这些维上的坐标作为原数据集的压缩。其主要缺陷是当数据改变时,特征向量需要重新计算,因此不适合动态变化的数据库。分段法主要通过将序列分段,取每段的特征(如极值点、变化趋势等)组成序列的特征表示。在时间序列的各种模式表示方法中,分段线性表示方法具有简单直观、支持时间序列的动态增量更新、支持快速相似性搜索及支持新的聚类、分类算法和奇异点检测等优点。

2 STC 算法

2.1 相关定义

定义 1 时间序列。它是由一系列元素组成的有序集合,这些元素本身由记录时刻和记录值构成,记为 $X = \langle x_1 = (t_1, v_1), x_2 = (t_2, v_2), \dots, x_n = (t_n, v_n) \rangle$ 。其中元素 $x_i = (t_i, v_i)$ 表示在 t_i 时刻取得记录值为 v_i ,这里记录时刻 t_i 是严格单调增加的,即 $i < j \Leftrightarrow t_i < t_j$ 。

通常时间序列的记录时间间隔 $\Delta t = t_{i+1} - t_i$ 是相等的,因此对于上述时间序列可以取 $t_1 = 0, \Delta t = 1$,那么时间序列 $X = \langle x_1 = (t_1, v_1), x_2 = (t_2, v_2), \dots, x_n = (t_n, v_n) \rangle$ 可以简记为 $X = \langle x_1, x_2, \dots, x_n \rangle$ 。 $|X|$ 称为时间序列 X 的模,表示时间序列的长度,即元素 x 的数量。对于广义时间序列,记录值 v_i 可以是离散符号、结构数据、多媒体数据等,但本文只采用狭义的时间序列,即 v_i 的取值为实数类型。

定义 2 时间序列的模式表示。设有时间序列 $X = \langle x_1, x_2, \dots, x_n \rangle$,时间序列的模式表示就是指通过提取时间序列的模式,将时间序列变换到模式空间,从而得到时间序列的模式表示。用符号定义如式(1)所示。

$$X(t) = f(w) + e(t) \tag{1}$$

其中: W 是时间序列的模式,它是指该时间序列的某种变化特征,这种特征可以是该时间序列在一段时间的均值或中值,也可以是时间序列离散化后的符号,甚至可以是时间序列的傅里叶变换系数; $f(w)$ 是时间序列的模式表示; $e(t)$ 是时间序列的模式和它的模式表示之间的误差。

2.2 基本思想

本算法借鉴 SEEP 算法中利用直线斜率的思想,首先计算连接同一点(除线段的两个端点外)左右两条线段的斜率;然后通过斜率的比值来确定该点是否是变化点;最后将这些变化点依次用线段相连接,就得到了时间序列的一种分段线性表示。该算法的关键是变化点的确定。其变化点的确定遵循以下原则:a)第一个点和最后一个点确定为变化点;b)当某点与其左边邻点连线的斜率为零时,该点与其右边邻点连线的斜率在 $(-d, +d)$ 外,就认为该点为变化点;c)当某点与其左边邻点连线的斜率不为零时,该点与其右边邻点连线的斜率和该点与其左边邻点连线的斜率的比值在 $(1-d, 1+d)$ 外,就认为该点为变化点。上述中的 d 是一个阈值参数。

如图 1 所示,当 A, B 两点确定以后,第三点落在平行于纵轴直线 l 上,输入阈值参数 d ,可以计算出 C' 点和 C'' 点。如果第三点落在线段 $C'C''$ 之间,那么认为该点不是变化点;如果第三点落在线段 $C'C''$ 以外直线 l 的其余点上,则认为该点是变化点。

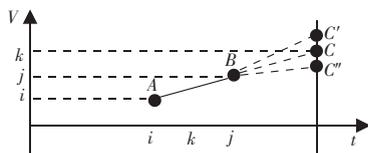


图 1 时间序列变化点的确定

2.3 算法表示

算法名称: STC 算法

输入: 时间序列 $X = \langle x_1 = (t_1, v_1), x_2 = (t_2, v_2), \dots, x_n = (t_n, v_n) \rangle$, 参数 n 为原始时间序列的长度, d 为阈值参数。

输出: 时间序列的 STC 表示。

$ii = 0; jj = 1; kk = 2;$

$xx = \{ (x_1, 1) \};$ // 第一个序列点是变化点

for ($i = 1; i < n - 1; i++$)

$l_1 = (x[jj] - x[ii]) / (jj - ii); l_2 = (x[kk] - x[jj]) / (kk - jj);$

if ($x[ii] = x[jj]$) // 前两点连线斜率为零

if ($l_2 > d$) || ($l_2 < -d$)

then $xx = xx + \{ (x_i, i) \};$ // 序列点 x_i 加入变化点集合 xx

$ii = i; jj = i + 1; kk = i + 2;$

else $jj = i + 1; kk = i + 2;$

else // 前两点连线斜率不为零

if ($(l_2/l_1) > (1 + d)$) || ($(l_2/l_1) < (1 - d)$)

then $xx = xx + \{ (x_i, i) \};$ // 序列点 x_i 加入变化点集合 xx

$ii = i; jj = i + 1; kk = i + 2;$

else $jj = i + 1; kk = i + 2;$

$xx = xx + \{ (x_n, n) \};$ // 最后一个序列点是变化点

output $L(X) = \{ L(x_{i_1}, x_{i_2}), L(x_{i_2}, x_{i_3}), \dots, L(x_{i_{k-1}}, x_{i_k}) \}$ ($x_{i_m},$

$m \in xx$);

3 实验

3.1 实验数据

在实验中采用来自航天器测试领域的实际数据集和人工合成的随机游走时间序列数据集来比较各分段线性表示的性能。这些实际数据集都采集自某航天器的各分系统。其中数据集 1 来自电源分系统, 数据集 2 来自测控分系统, 数据集 3 来自数管分系统, 数据集 4 和 5 来自姿态和轨道控制分系统。各数据集的相关信息如表 1 所示。

表 1 数据集描述

数据集名称	长度/行	大小/千字节
数据集 1	68 531	3 667
数据集 2	152 366	6 589
数据集 3	21 570	1 192
数据集 4	135 872	60 157
数据集 5	95 367	4 496
随机游走数据集	65 536	1 120

3.2 评价指标

本文使用压缩率和拟合误差来对面向航天器测试的时序数据模式表示进行评价。

1) 压缩率 设有时间序列 $X = \langle x_1, x_2, \dots, x_n \rangle$, 通过线性分段算法得到的时间序列 $X = \langle x_1', x_2', \dots, x_n' \rangle$ 。其中 $x_1' = x_1, x_n' = x_n$ 。那么该时间序列模式表示的压缩率用符号定义如式(2)所示。

$$\eta = (1 - \frac{n'}{n}) \times 100\% \quad (2)$$

2) 拟合误差 设有时间序列 $X = \langle x_1, x_2, \dots, x_n \rangle$, 通过线性分段算法得到时间序列的分段表示 $L(X) = \langle L(x_{i_1}, x_{i_2}), (x_{i_2}, x_{i_3}), \dots, (x_{i_{k-1}}, x_{i_k}) \rangle$ 。其中 $L(\cdot, \cdot)$ 表示连接两点的直线段。将 $L(X)$ 经过线性插值后得到的时间序列记为 $X^c = \langle x_1^c, x_2^c, \dots, x_n^c \rangle$, 那么时间序列分段线性表示与原始时间序列之间的拟合误差用符号定义如式(3)所示。

$$E = \sqrt{\sum_{i=1}^n (x_i - x_i^c)^2} \quad (3)$$

3.3 实验方法

Keogh 等人^[14]的实验表明, 分段线性表示方法与离散傅

里叶变换、离散小波变换以及奇异值分解相比, 具有简单、直观和高效等优点。因此, 结合航天器时序数据的特点, 以及上述时间序列模式表示方法的分析, 实验主要从以下两个方面对五种分段线性表示算法在航天器时序数据模式表示上进行相关性能的对比:

a) 在同一压缩率下, 比较运用各算法后得到的时序数据序列与原始时序数据序列的拟合误差;

b) 在不同压缩率下, 比较各算法表示在同一数据集上的拟合误差。

虽然数据集都是来自航天器测试领域, 但分属于不同的分系统, 序列值相差很大。为了便于对比, 首先对时间序列作规范化处理, 将序列值规范化到 $[0, 1]$ 。规范化公式用符号定义如式(4)所示。

$$\text{norm}(x_i) = \frac{x_i - \min(X)}{\max(X) - \min(X)} \quad (4)$$

其中: $\min(X)$ 和 $\max(X)$ 分别表示序列中的最大值和最小值。

3.4 实验结果与分析

1) 在同一压缩率下, 比较运用各算法后得到的时序数据序列与原始时序数据序列的拟合误差

PAA 算法的压缩率只能为 $(1 - \frac{1}{\omega}) \times 100\%$ 。其中参数 ω 是正整数, 这里取 $\omega = 4$, 即压缩率为 75%。实验结果如表 2 所示。其中算法 1 是指基于重要点的分段线性表示算法, 算法 2 是指分段聚集近似表示算法, 算法 3 是指基于特征点的分段线性表示算法, 算法 4 是指 SEEP 算法, 算法 5 是指 STC 算法。

表 2 压缩率为 75% 的拟合误差

数据集	算法				
	1	2	3	4	5
1	0.78	0.81	0.25	0.08	0.07
2	1.61	1.62	1.77	0.56	0.49
3	2.11	2.30	3.04	1.34	1.08
4	1.38	2.43	1.78	0.97	1.36
5	2.45	2.55	3.62	2.32	1.79

从表 2 中可以看出, 在时间序列集中的五条时间序列上, STC 算法在其中的四条时间序列上具有最小的拟合误差, 其余一条时间序列的拟合误差也较小。对于斜率变化比较集中的数据集 1, STC 算法的拟合误差与 SEEP 算法相当。但是从航天器的数据特点来看, 航天器测试得到的时间序列数据集不仅存在着斜率变化集中的序列, 而且也存在着大量斜率变化比较大的序列。因此, 在上述算法中, STC 算法比较适合航天器测试时间序列的模式表示, 而且对于来自航天器不同分系统的时间序列 STC 算法都有较好的拟合效果, 说明其适应性很强。

2) 在不同压缩率下, 比较各算法表示在同一数据集上的拟合误差

数据集采用随机游走序列 Random_Walk, 压缩率分别取值为 90%、85%、80%、75%、70% 和 65%, 实验数据和结果如图 2 所示。其中算法 1 是指基于重要点的分段线性表示算法, 算法 2 是指分段聚集近似表示算法, 算法 3 是指基于特征点的分段线性表示算法, 算法 4 是指 SEEP 算法, 算法 5 是指 STC 算法。

图 2 给出了五种算法在随机游走时间序列 Random_Walk 上的拟合误差的比较。从图中可以看出, STC 算法在所有压缩率设置下的拟合误差都是最小的, 并且随着压缩率的降低其拟合误差也逐渐减小。STC 算法实现简单, 可以直接用于时间序列的在线分段, 只需扫描序列一次就可以完成变化点的确定,

算法的时间复杂度为 $O(n)$, 其中 n 是时间序列的长度。从实验中可以看到, STC 算法拟合程度较高, 适应性较强。

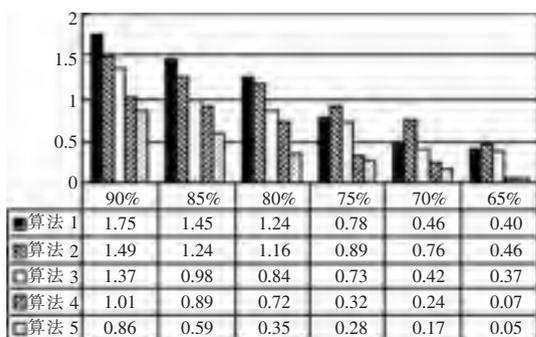


图 2 时间序列 Random_Walk 上的拟合误差

4 结束语

时序数据的模式表示是时序数据挖掘的基础, 本文结合航天器测试的时序数据资源的特点, 提出了一种基于斜率变化阈值的时间序列分段线性 STC 表示算法。实验表明, 该算法具有拟合程度高、适应能力强等特点, 是一种较优的航天器时序数据挖掘数据预处理方法。

参考文献:

[1] 王庆成. 航天器电测技术[M]. 北京: 中国科学技术出版社, 2007.
 [2] 陈蔚文. 面向航天器测试的通用数据资源管理平台的设计和实现[D]. 北京: 北京航空航天大学, 2007.
 [3] AGRWAL R, FALOUTSOS C, SWAMI A. Efficient similarity search in sequence databases[C]//Lecture Notes Computer Science. [S. l.]: Springer, 1993: 69-84.
 [4] RAFIEI D, MENDEIZON A. Similarity-based queries for time series data[C]//Proc of ACM SIGMOD Conference on Management of Data. Tucson: ACM Press, 1997.
 [5] CHAN K P, FU A W C. Efficient time series matching by wavelets[C]//Proc of the 15th International Conference on Data Engineering. Washington DC: IEEE Computer Society, 1999: 126.
 [6] PARK S, CHU W W, YOON J, et al. Efficient searches for similar subsequences of different lengths in sequence databases[C]//Proc of the

16th International Conference on Data Engineering. Washington DC: IEEE Computer Society, 2000: 23.
 [7] PARK S, CHU W W, YOON J, et al. Similarity search of time-warped subsequences via a suffix tree[J]. Information Systems, 2003, 28(7): 867-883.
 [8] LIN J, KEOGH E, LONARDI S, et al. A symbolic representation of time series, with implications for streaming algorithms[C]//Proc of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. New York: ACM Press, 2003: 2-11.
 [9] 钟清流, 蔡自兴. 基于统计特征的时序数据符号化算法[J]. 计算机学报, 2008, 31(11): 1857-1864.
 [10] WU D, SINGH A, AGRAWAL D, et al. Efficient retrieval for browsing large image databases[C]//Proc of the 5th International Conference on Information and Knowledge Management. New York: ACM Press, 1996: 11-18.
 [11] KANTH K V, AGRAWAL D, SINGH A. Dimensionality reduction for similarity searching in dynamic databases[J]. Computer Vision and Image Understanding, 1999, 75(1): 59-72.
 [12] RIPLEY B. Pattern recognition and neural networks[M]. London: Cambridge University Press, 1996.
 [13] PRAT K B, FINK E. Search for patterns in compressed time series[J]. International Journal of Image and Graphics, 2002, 2(1): 89-106.
 [14] KEOGH E, CHAKRABARTI K, PAZZANI M, et al. Dimensionality reduction for fast similarity search in large time series databases[J]. Journal of Knowledge and Information Systems, 2001, 3(3): 263-286.
 [15] BYOUNG-KEE Y, FALOUTSOS C. Fast time sequence indexing for arbitrary Lp norms[C]//Proc of the 26th International Conference on Very Large Data Bases. San Francisco: Morgan Kaufmann Publishers Inc, 2000: 385-394.
 [16] XIAO Hui, FENG Xiao-fei, HU Yun-fu. A new segmented time warping distance for data mining in time series database[C]//Proc of International Conference on Machine Learning and Cybernetics. 2004: 1277-1281.
 [17] 詹艳艳, 徐荣聪, 陈晓云. 基于斜率提取边缘点的时间序列分段线性表示方法[J]. 计算机科学, 2006, 31(11): 139-142, 161.

(上接第 126 页)地进行调整; 非线性参数是由训练样本和启发方法直接决定的, 而没有用优化算法来确定。虽然高斯宽度在学习时可以自适应地调整, 但学习规则却很简单。仿真结果表明, 由于使用了修剪技术的自适应系统, 使得 D-FNN 具有紧凑的系统结构、强大的泛化能力以及快速的学习速度。

参考文献:

[1] CHEN S, COWAN C F N, GRANT P M. Orthogonal least squares learning algorithm for radial basis function network[J]. IEEE Trans on Neural Networks, 1991, 20(2): 302-309.
 [2] LEONTARITIS I J, BILLINGS S A. Input-output parametric models for nonlinear systems, part 1: deterministic nonlinear systems[J]. International Journal of Control, 2006, 41(2): 303-344.
 [3] CHELLAPPA R, WILSON C L, SIROHEY S. Human and machine recognition of faces: a survey[J]. Proceedings of the IEEE, 1995, 83(5): 705-741.
 [4] BRUNELLI R, POGGIO T. Face recognition: features versus templates[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2003, 25(10): 1042-1053.
 [5] CHAO C T, CHEN Y J, TENG C C. Simplification of fuzzy-neural systems using similarity analysis[J]. IEEE Trans on System, Man,

and Cybernetics, Part B: Cybernetics, 1996, 26(2): 344-354.
 [6] PLATT J. A resource-allocating network for function interpolation[J]. Neural Computation, 2007, 19(2): 213-225.
 [7] LU Ying-wei, SUNDARARAJAN N, SARATCHANDRAN P. A sequential learning scheme for function approximation by using minimal radial basis function networks[J]. Neural Computation, 1997, 9(2): 461-478.
 [8] JANG J S R. ANFIS: adaptive-network-based fuzzy inference system[J]. IEEE Trans on System, Man, and Cybernetics, 1993, 23(3): 665-685.
 [9] KADIRKAMANATHAN V, NIRANJAN M. A function estimation approach to sequential learning with neural networks[J]. Neural Computation, 2004, 16(4): 954-975.
 [10] HONG Zi-quan. Algebraic feature extraction of image for recognition[J]. Pattern Recognition, 1991, 24(3): 211-219.
 [11] TURK M A, PENTLAND A P. Eigenfaces for recognition[J]. Cognitive Neuroscience, 2006, 3(1): 71-86.
 [12] VALENTIN D, ABDI H, O' TOOLE A J, et al. Connectionist models of face processing: a survey[J]. Pattern Recognition, 1994, 27(9): 1209-1230.