文章编号:1001-9081(2010)06-1539-04

基于动态特征提取和神经网络的数据流分类研究

汪成亮^{1,2}, 庞 栩¹, 陆志坚¹, 罗昌银¹

(1. 重庆大学 计算机学院, 重庆 400044; 2. 重庆大学 电气工程学院, 重庆 400044)

(ltpx2007@sina.com)

摘要:为提高数据流分类的精确性和适应性,提出了一种新的数据流分类方法。该方法基于总体最小二乘法对数据流进行分段拟合,并将传统曲线分析算法——滑动窗口(SW)和在线数据分割(OSD)进行结合、改进,以可变滑动窗口算法实现对数据流的合理分割,提高趋势分析精度。在此基础上,对数据流进行动态特征提取和判断,并以神经网络对数据流特征进行模式识别,精确分类,进而对监控对象提供早期预警、状态评估和决策支持。实验结果表明,该方法能对数据流进行有效的动态特征描述,分类效果明显。

关键词:分类;神经网络;数据流;可变滑动窗口;趋势分析

中图分类号: TP391 **文献标志码:** A

Classification of data stream based on dynamic feature extraction and neural network

WANG Cheng-liang^{1,2}, PANG Xu¹, LU Zhi-jian¹, LUO Chang-yin¹

(1. College of Computer Science, Chongqing University, Chongqing 400044, China;

2. College of Electrical Engineering, Chongqing University, Chongqing 400044, China)

Abstract: To improve the accuracy and adaptability of the classification of data stream, this paper presented a new method of classification. This method used total least squares to fit the segmentation of data stream, and presented a variable sliding window algorithm to achieve a reasonable segmentation and improve the accuracy of trend analysis by combining Sliding Window (SW) algorithm with extrapolation for On-line Segmentation of Data (OSD) algorithm. By extracting the dynamic feature of data stream, neural network was used for the pattern recognition of data stream and classification so it can provide early warnings, severity assessments of monitored subjects and information for decision support. The test results show that this method can describe the dynamic characteristic of data stream effectively, and the effect of classification is evident.

Key words: classification; neural network; data stream; changeable sliding window; trend analysis

0 引言

近几年,数据流在股票交易、电话通信记录、网络流量监控、传感器网络等许多领域得到了广泛应用。数据流中蕴涵大量信息,可以用来作为智能决策的依据。动态数据流的特征提取及描述目的就是提取变化信息,对数据流信号进行分类识别以达到决策支持^[1],如股市曲线分析、音频波形分析、心电图波形识别及工业生产曲线识别等。

但由于应用领域的不同,动态数据流自身曲线特征千差万别,且大多存在于非线性、多变量耦合、时变和大时滞的某过程体系中。因此新算法要求数据流趋势分析方法适应性强、分析精度高,并对此建立统一的数据模型,能够应用于各个领域的动态数据流分析、预测。

传统的曲线趋势分析算法有滑动窗口(Sliding Window, SW)算法、外推式在线数据分割(On-line Segmentation of Data, OSD)算法,但它们有明显的不足。

1) SW算法的滑动窗口最大长度没有限制,当检测点阈值比较大时,窗口的长度可能过长造成拟合误差变得很大。且到达的数据均需进行一次完整的回归建模和数据段分割点

判定,随着数据的增多,计算效率极为低下^[2-4]。

2) OSD算法在SW算法基础上设定了最小滑动窗口,且对于新到达的数据,只将其代入已建立的模型,分析外推累积误差。虽然提高了效率,但滑动窗口内的突变点无法检测^[5]。

3) SW和OSD算法对滑动窗口最大长度都没有限制,当检测点阈值比较大时,窗口的长度可能很长使得趋势分析的误差变大。

4) SW和OSD均采用常规最小二乘法进行曲线拟合。与常规最小二乘法相比,总体最小二乘法具有更高的直线拟合精度^[6]。

因此,本文对传统数据流趋势分析算法进行了改进,以提高趋势分析精度,使数据流特征表现更为精确^[7]。并在此基础上,利用神经网络所具有的非线性映射能力强、能够学习和适应不确定性系统的动态特性,对特征参数进行建模以达到识别、分类各曲线的目的。

1 数据流问题描述

数据流实时趋势分析的基础是按某一统计特性指标(比

收稿日期:2009-12-25;修回日期:2010-03-29。 基金项目:国家863计划项目(2007AA12Z306);中国博士后科学基金资助项目(20080430750);重庆市自然科学基金资助项目(CSTC2007BB6118)。

作者简介:汪成亮(1975-),男,重庆人,副教授,博士,主要研究方向:人工智能、软件系统设计; 庞栩(1981-),男,重庆人,硕士研究生,主要研究方向:数据仓库、数据挖掘、神经网络; 陆志坚(1985-),男(布依族),贵州贵阳人,硕士研究生,主要研究方向:人工智能; 罗昌银(1983-),男,重庆人,硕士研究生,主要研究方向:数据库。

如均方差、累计误差、广义似然比统计量)实时地分割,使所分割的数据段内数据服从同一统计模型,而相邻分段服从不同的统计模型。

为了研究需要,我们定义:不断到达的一维时序数据流为^[8]:

$$Y = \{v_{t_1}, \dots, v_{t_i}, \dots, v_{t_c}, \dots\} \quad (1)$$

其中 t_c 为当前时刻。

数据流分割(本文采用均方差比较)将 Y 分割成一系列连续的非空数据段(即滑动窗口):

$$\{Y_1, \dots, Y_j, \dots, Y_s, \dots\} \quad (2)$$

其中第 j 数据段为:

$$Y_j = \{v_{t_{j,1}}, \dots, v_{t_{j,\lambda}}, \dots, v_{t_{j,n_j}}\} \quad (3)$$

对应的数据到达时间:

$$t_{j,\lambda} \in \{t_1, \dots, t_i, \dots, t_c, \dots \mid j \in \mathbf{N}, 1 \leq j \leq s; \lambda \in \mathbf{N}, 1 \leq \lambda \leq n_j\} \quad (4)$$

在式(3)~(4)中,用 n_j 表示数据段 Y_j 的长度,即滑动窗口 Y_j 的长度,而且令 $t_{1,1} = t_1$ 。记数据段 Y_i 包含当前数据 v_{t_c} 的数据段。

设 Y_j 中的数据可用线性回归模型进行拟合,即

$$v(t) = (a_j + \delta_j(t))t + (b_j + \varepsilon_j(t)); t \in \{t_{j,1}, \dots, t_{j,n_j}\} \quad (5)$$

其中: a_j 和 b_j 为模型参数,参数 a_j 称为数据段 Y_j 的趋势特征值; $\delta_j(t)$ 为 a_j 的误差扰动; $\varepsilon_j(t)$ 为独立同分布零均值白噪声。

为了算法描述的需要,设数据段 Y_{j+1} 的第1个数据元素 $v_{t_{j+1,1}}$ 为 Y_j 的分割点。另外,令 β 为模型参数向量

$$\beta = [a_j \quad b_j]^T \quad (6)$$

数据流实时趋势分析的基本任务是:在当前已接收数据序列 $Y_{s,n} = \{v_{t_{s,1}}, \dots, v_{t_{s,n}}\}$ 的基础上,对新到达的数据流元素 v_{t_c} 完成以下计算:1) 分割点检测(比如检测 v_{t_c} 是否作为数据段 $Y_{s,n}$ 的分割点);2) 对当前数据段建立回归模型,计算出当前拟合的模型参数值 a_j 和 b_j 。

2 数据流趋势分析方法

2.1 总体最小二乘法回归建模

总体最小二乘法是一种较为先进的最小二乘法结构,总体最小二乘法认为回归矩阵存在干扰,在计算最小二乘解时考虑了这个因素,而在一般最小二乘法时没有考虑该因素的影响。

通过总体最小二乘法进行曲线趋势拟合(优化),能够提高曲线趋势分析的精度。

设当前数据序列 $Y_{s,n} = \{v_{t_{s,1}}, v_{t_{s,2}}, \dots, v_{t_{s,n}}\}$ 构造的线性回归模型^[9]为:

$$v(t) = (a_{s,n} + \delta)t + b_{s,n} + \varepsilon \quad (7)$$

若该模型函数通过点 $(t_{s,i}, v_{t_{s,i}})$, 则式(7)可转化为:

$$v - v_{t_{s,i}} = a_{s,n}(t - t_{s,i}) \quad (8)$$

设 $(\bar{t}, \bar{v})$ 为当前数据段 $Y_{s,n}$ 的几何中心点。其中 $\bar{t}, \bar{v}$ 分别是:

$$\bar{t} = \frac{1}{n} \sum_{i=1}^n t_{s,i}$$

$$\bar{v} = \frac{1}{n} \sum_{i=1}^n v_{t_{s,i}}$$

此时回归模型方程为:

$$v - \bar{v} = a_{s,n}(t - \bar{t}) \quad (9)$$

令中心化的数据向量为:

$$\mathbf{T} = [t_{s,1} - \bar{t}, t_{s,2} - \bar{t}, \dots, t_{s,n} - \bar{t}]^T$$

$$\mathbf{V} = [v_{t_{s,1}} - \bar{v}, v_{t_{s,2}} - \bar{v}, \dots, v_{t_{s,n}} - \bar{v}]^T$$

常规最小二乘法极小化 $\varepsilon = \mathbf{V} - \mathbf{T}a_{s,n}$ 的范数 $\xi_{LS} = \|\varepsilon\|^2$, 这等价于在条件 $\mathbf{T} \cdot a_{s,n} = \mathbf{V} - \varepsilon$ 的约束下, $\|\varepsilon\|^2$ 达到最小化。也就是说,使校正项 ε 尽可能小,同时通过强令 $\mathbf{T} \cdot a_{s,n} = \mathbf{V} - \varepsilon$ 补偿于数据流纵坐标 Y 中的噪声。 $\mathbf{T}$ 和 $\mathbf{V}$ 分别是数据点的横坐标向量和纵坐标向量,它们有对等的地位。常规最小二乘法认为纵坐标中有噪声而横坐标中无噪声显然是不合理的。而总体最小二乘法同时考虑 $\mathbf{T}$ 和 $\mathbf{V}$ 中的噪声,令 φ 代表横坐标 $\mathbf{T}$ 向量中的误差,则有 $(\mathbf{T} - \varphi) \cdot a_{s,n} = \mathbf{V} - \varepsilon$ 。在总体最小二乘法中, $a_{s,n}$ 的选取应使得总体误差 $\mathbf{F} = [\Delta\varepsilon]$ 的 Frobenius 范数最小,即 $\min \|\mathbf{F}\|_F^2$, 其中:

$$\|\mathbf{F}\|_F^2 = \text{tr}[\mathbf{F}^T \cdot \mathbf{F}] \quad (10)$$

其中 $\text{tr}[\cdot]$ 代表矩阵的迹。

基于总体最小二乘法的直线拟合可以归结为一个简单的二阶矩阵的特征向量求解,运算量很小。该算法同时考虑了数据点在横坐标方向和纵坐标方向的不确定性,因而较常规的最小二乘法直线拟合具有更高的拟合精度。

因此,本文的数据流趋势分析算法决定采用总体最小二乘法对各个数据分段进行直线拟合,以提高趋势分析的精度。

2.2 可变滑动窗口算法

SW 算法的滑动窗口最大长度没有限制,当检测点阈值比较大时,窗口的长度可能很长使得趋势分析的误差变大;而 OSD 算法限定了最小滑动窗口长度,使得在最小滑动窗口内的突变点无法检测。

针对 SW 和 OSD 算法中滑动窗口存在的缺陷,本文提出一种动态改变设定窗口长度的可变滑动窗口算法,以合理地数据序列进行分割。该算法首先设置数据段基准窗口长度和最长数据窗口长度,从当前数据段的起始点开始,对每一新到达的数据流元素都重新建立回归模型,提高精度。在当前数据段小于基准窗口长度时,用模型的拟合均方差与噪声函数 G 的返回值比较,检测基准窗口内是否存在异常点;当前数据段长度若大于等于基准窗口长度,而且模型的拟合均方差大于预先设置的标准分割点阈值时,则认为新到达的数据为当前数据段的分割点。如果当前数据段长度大于最长数据窗口长度,则回溯,从当前数据段的起点开始,从中寻找一个拟合均方差与标准分割点阈值最接近的数据点作为当前数据段的分割点。

可变滑动窗口算法解决 SW 算法和 OSD 算法的固定窗口问题,实现了对数据流的合理分割,因此趋势分析的精度得到了提高。

具体算法的伪代码如下。

基准窗口长度 L : 一般根据所应用领域分析的数据流的波形变化特点决定。

最长数据窗口长度 $MaxLength_Window$: 一般根据所分析的数据流波形变化的特点决定。

标准分割点阈值 θ : 一般为所应用领域专家经验决定。

噪声函数 $G(\cdot)$: 一个接口函数,在当前数据短小于基准

窗口时调用。其功能是返回分割点阈值,该值要根据所应用领域的具体条件决定。

设模型的拟合均方差为 $\mu_{s,n}$, 总体最小二乘法为 $f(\cdot)$ 。

```

Begin
  While (Y不为空)  // 由式(1)可知,Y为所分析的数据流
  {
     $\mu_{s,n} = f(Y_{s,n})$ 
    // 由前面的分析可知, $Y_{s,n}$ 为当前已接收数据序列,
    //  $n$ 为该序列的长度
    If ( $n < L$ )
    {
      If ( $\mu_{s,n} > G(\cdot)$ )
      {
         $v_{tc}$ 是数据段  $Y_{s,n}$  的分割点, $Y_{s,n}$  的趋势特征值为  $\beta = [a_{s,n}, b_{s,n}]^T$ , 并将后续到达的数据归入新的当前数据段, 启动新的趋势分析过程
      }
      Else
      {
        继续分析下一个到达的数据
      }
    }
    If ( $n \geq L$ )
    {
      If ( $\mu_{s,n} > \theta$ )
      {
         $v_{tc}$ 是数据段  $Y_{s,n}$  的分割点, $Y_{s,n}$  的趋势特征值为  $\beta = [a_{s,n}, b_{s,n}]^T$ , 并将后续到达的数据归入新的当前数据段, 启动新的趋势分析过程
      }
      Else if ( $n < \text{MaxLength\_Window}$ )
      {
        继续分析下一个到达的数据
      }
      Else
      {
        While ( $Y_{s,n}$  不为空)
        {
          搜索拟合均方差与  $\theta$  最为接近的数据点作为数据段  $Y_{s,n}$  的分割点, 并将后续到达的数据归入新的当前数据段, 启动新的趋势分析过程
        }
      }
    }
  }
End

```

3 BP神经网络的建立

3.1 神经网络原理

神经网络基于从输入到输出的映射,对非线性函数的优化进行求解运算。在整个网络学习中,从输入输出数据中提取规律,将其保存于网络权值中并应用于一般情形,具有极强的自学习能力和推广、概括能力。

由于神经网络自身特点,其广泛运用于社会各领域之中。它由输入层、隐含层和输出层组成,能实现信号从输入空间到输出空间的变换,其信息处理能力来自于简单非线性函数的多次复合,通常比较适合预测、模式识别以及非线性函数的逼近。实际上,BP网络的学习过程就是不断调整网络的权值和偏差,使得实际输出和期望输出不断接近的过程^[10]。理论上,一个具有至少一层隐含层和足够数量隐藏节点的多层前向网络能够逼近任何测量函数。

3.2 数据流动态特征的提取

动态数据流趋势分析得到了表述该曲线特征的分割点,本文利用各分割点、分割点间数据段参数提取该数据流的曲线特征,并将各特征数据作为BP神经网络的输入参数。

经过不断实验调整,本文所提出的曲线特征提取参数如下。

- 1) 压缩比数 Num : $Num = \text{数据流数据量} / \text{分割点数}$ 。
- 2) 平均上升斜率为:

$$avg_A_inc = \left(\sum_{i=1}^k A_inc_i \right) / Num$$

其中: A_inc_i 为斜率为正的数据段斜率; $i = 1, 2, \dots, k; k < Num$ 。

- 3) 平均下降斜率为:

$$avg_A_dec = \left(\sum_{i=1}^t A_dec_i \right) / Num$$

其中: A_dec_i 为斜率为负的数据段斜率; $i = 1, 2, \dots, t; t < Num$ 。

- 4) 极值点数 $extremun_num$ (若分割点前后数据段斜率符号相反,该分割点为极值点)。

- 5) 数据流平均值 $\bar{Y}$ 为:

$$\bar{Y} = \left(\sum_{i=1}^N Y_i \right) / N$$

- 6) 数据流的均方差 θ 为:

$$\theta = \sum_{i=1}^N (Y_i - \bar{Y})^2 / N^2$$

其中: $\bar{Y}$ 为数据流均值, N 为数据流数据量。

3.3 输入层、隐藏层、输出层的确定

输入层、隐藏层、输出层节点数完全根据具体要求来设计。本文的研究中,研究对象是某段数据流所代表的状况。因此,输出层即为曲线状况,输入参数为曲线特征提取的6种参数。BP模型可表示为:

$y = \text{net}(Num, avg_A_inc, avg_A_dec, extremun_num, \bar{Y}, \theta)$
神经网络的传递函数采用 Logsig 算法,学习函数采用 Levenberg-Marquardt 优化算法 (trainlm 函数)。

隐含层单元数目的问题并没有一个明确、统一的解决方案,一般通过逐步实验,一直达到理想输出结果为止。对应三层神经网络而言,经验上可得到隐含层单元数 h 的几个公式^[11]:

$$h = \sqrt{n + m} + a$$

$$h = \sqrt{n \cdot m}$$

其中 a 为 0 ~ 10 的常数。本文经过实验取隐藏层节点数 13,因此 BP 模型为 $6 \times 13 \times 1$ 结构。

4 实验与分析

铝电解槽电压是一种平稳的数据流,它的有效趋势提取可提供槽况恶化的早期预警,评估槽况、工况的状态,对氧化铝浓度的判断提供支持信息。

4.1 数据流趋势分析实验

论文基于总体最小二乘法建立回归模型以拟合各个分段的数据,采用可变滑动窗口算法进行分割点检测,对铝电解过程的重要参数(槽电压)进行实时趋势分析。

实验分别采用本文算法、SW 算法和 OSD 算法对同一数据流进行趋势分析,并用不同的分割点阈值分别检验各算法以下性能指标。

- 1) 分隔点点数 N_1 。
- 2) 压缩比 λ 为:
 $\lambda = \text{数据流点数} / N_1$
- 3) 拟合均方误差 E_{MS} 为:

$$E_{MS} = \frac{1}{n} \sum_{i=1}^n [(at_i + b) - v_i]$$

其中; n 表示数据段的长度,即电压采样点的点数; a 为该数据段的拟合的模型参数值; v_i 为实际电压, t_i 为横坐标时间。

在满足上述限定下,对各算法的参数尽可能优化设置。综合考虑铝电解情况^[12],我们设置:基准窗口长度 $L = T/4$ (T 为电压曲线波形的周期,该周期随运行工况可变,初始值为 110 min),最长数据窗口长度 $MaxLength_Window = Lk$ (k 为可变参数, $1.5 \leq k \leq 2$, 初始值为 1.5),标准分割点阈值 $\theta = r$ ($r \in \mathbf{N}, 2 \leq r \leq 10$)。

图 1 为 3 种算法对部分数据流的趋势分析,该数据流共 836 个采样点,每个采样点时间间隔为 10 s (为便于观察将 3 种算法得出结果加上某固定值)。它们基于相同分割点阈值,而本算法涉及的噪声函数 $G(\cdot)$ 需要根据计算的工况、槽况自适应调整(在此不过多阐述),以便检测基准窗口内的突变值。

表 1 列出了在不同分割点检测阈值下 3 种算法的分割点个数、压缩比及拟合均方误差具体数据。图 2~3 显示了它们

的压缩比和拟合均方误差比较。

图 1 3 种算法的数据流趋势分析

由图 1 可知:与 SW 算法和 OSD 算法相比,本文的算法可以更好地逼近原始槽电压数据,SW 算法的精确度在数据长期趋于平稳后增大,OSD 算法缺乏对基准窗口内的突变点检测。再由表 1、图 3 可见,文中算法比 SW 算法、OSD 算法具有更高的拟合精度,特别是当分割点检测的阈值较大的时候更为明显。但本算法的分隔点数和压缩比性能略差于 SW 和 OSD,但差别很小,可忽略不计。

表 1 3 种算法的分割点数、压缩比、拟合均方误差比较

分割点阈值	N_1			λ			E_{MS}		
	本文算法	SW 算法	OSD 算法	本文算法	SW 算法	OSD 算法	本文算法	SW 算法	OSD 算法
2	213.00	210.00	206.00	29.90	30.00	30.80	3.00	3.21	3.32
4	205.00	187.00	125.00	31.00	32.27	33.99	5.13	5.26	5.37
6	172.00	164.00	161.00	37.06	38.66	39.38	5.92	6.26	7.06
8	161.00	157.00	153.00	39.59	40.42	41.62	6.15	7.85	9.19
10	158.00	150.00	145.00	40.34	42.42	43.88	6.35	11.03	12.35

图 2 3 种算法的压缩比比较

图 3 3 种算法的拟合均方误差比较

图 4 3 种算法下的神经网络分类识别精度结合

结合图 1、4 可得出,神经网络模型分类识别精度与方差拟合精度、分割点阈值相关。参考基于 SW 和 OSD 算法的分类模型,分类精度基本和拟合精度正相关,和分割点阈值负相关。而采用本文算法的分类模型时,随拟合精度、分割点阈值的变化,分类精度先大后小,且总体无太大变化。

结果表明:由于 SW 和 OSD 算法自身缺陷,趋势分析精度随分割点增大而减小,且逐渐失去对数据流的动态特征描述,即使神经网络有很强的映射性,也无法准确对数据流进行识别。OSD 算法其在分割阈值较大时和 SW 算法无差别。但 OSD 设定了最小窗口,因此滤掉了部分分割点,导致数据流动态特征提取误差更大,分类模型识别精度随阈值增大而下降明显。

本文所提的分类模型则很好地解决了上述问题。虽然拟合精度也和分割点阈值负相关,但变化不大;且根据实际情况,算法蕴涵有某分割点阈值,能对数据流进行最佳动态特征描述,再利用神经网络的特点,数据流信号分类效果较好。

5 结语

本文将改进后的数据流趋势分析方法和神经网络相结合 (下转第 1567 页)

4.2 神经网络分类实验

根据铝电解工艺专家,将数据流信号对应的槽况分为 7 种状态:健康、基本健康、亚健康、临界健康、轻微病态、中度病态、严重病态。

实验同样采用不同的分割点阈值,以本文算法、SW 算法和 OSD 算法对数据流进行趋势分析,并基于本文的特征提取方法对数据流进行动态特征描述,进而建立神经网络分类识别模型。

实验采集了 700 个样本数据,训练样本数为 500,测试样本数为 200,其中单样本由 800 个采样点组成。神经网络的训练逼近精度均达到 0.1%,各状态下的分类识别精度见图 4。

数及算法结果见表1。

图1 基于模板分形的陶瓷图案样例

表1 基于模板分形算法用例参数及结果

序号	模板 规格 样图	迭代次数	模板空间	图案
1	5×5	3	$(5 \times 5)^3$	图1(e)
2	6×6	3	$(6 \times 6)^3$	图1(a)、(b)
3	3×3	4	$(3 \times 3)^4$	图1(f)
4	3×3	3	$(3 \times 3)^3$	图1(c)
5	3×3	4	$(3 \times 3)^4$	图1(d)

表1列出了算法使用的3×3模板2个,5×5模板、6×6模板各1个,应用本算法迭代表1中迭代代数列出的次数后,生成的模板经色调处理后生成的陶瓷图案如图1所列。由表1所列的5组数据可知,当算法指定的模板为 $m \times n$ 规格的模板时,如算法总共迭代 k 次,则迭代完成后算法最终生成的

图案模板的内在空间规模为 $(m \times n)^k$,表明随迭代次数的增加,算法生成的模板空间以 $m \times n$ 倍规模迅速增加。纵观本文所列的5次实验用例迭代生成的最终图案及迭代次数,可以发现应用模板分形生成图案的迭代次数通常为3~5次,绝大多数情况下迭代3次即可取得较满意的结果。表1所列的第4~5次应用可以发现,生成的最终图案(见图1(c)~(d))差别不大,甚至图1(c)拥有比图1(d)更丰富的图案细节。

4 结语

本文重点讨论了在建筑陶瓷图案设计过程中图案创作的步骤,分析了建筑陶瓷行业的应用特点,对模板与分形相结合的图案生成技术作了深入研究,在模板、分形技术基础上的提出并实现了一种基于模板分形的建筑陶瓷图案生成算法。算法生成的图案取决于模板的选择,图案内容充分体现的分形图形的自相似特征,最终图案足以满足建筑陶瓷行业应用需求,而计算机生成图案的自动、实时、多变的特点又有助于建筑陶瓷业增加产品的花色,满足用户的不同需求。该算法已在陶瓷图案CAD系统中成功应用。

参考文献:

- [1] 王小铭. 分形图案的构图艺术及其计算机实现[J]. 计算机辅助设计与图形学学报, 2001, 13(1): 83-86.
- [2] 孙天凯, 邵晓根, 王兴元. 扩展的分形L-系统与自然景观的动态模拟[J]. 计算机工程与应用, 2009, 45(2): 182-185.
- [3] 王万根, 余小清, 等. 基于几何参数的植物真实感模拟[J]. 计算机应用, 2009, 29(1): 97-100.
- [4] 王小铭, 李福清. 基于分形矩阵的编织物图案设计与生成[J]. 纺织学报, 2003, 8(24): 317-319.
- [5] BARB A. On a class of fractal matrices (I): Excess matrices and their self-similar properties[J]. International Journal of Bifurcation and Chaos, 1992, 2(4): 841-860.
- [6] 魏宝刚, 庞向斌, 朱文浩, 等. 基于纹理渲染的分形图案设计[J]. 中国图象图形学报, 2006, 11(5): 689-694.
- [7] GRÜNBAUM B, SHEPHARD G C. Tilings and patterns [M]. New York: Freeman, 1987.

(上接第1542页)

合,提出了一种新的数据流分类识别方法。该算法在基于总体最小二乘法上,可灵活设置数据基准窗口、最大窗口、分割点阈值,并采用可变滑动窗口算法。因此其可根据具体应用领域状况,对数据流进行相应的动态特征提取,提高了适应性及趋势分析精度。在此基础上,利用神经网络强大的学习能力和映射能力,对数据流进行分类识别,以对状态进行诊断或提供决策支持。该方法在某铝厂电解槽上进行的实时测试结果表明,该算法能准确分析槽电压实时趋势变化及电解槽工作状态。

参考文献:

- [1] MELEK W W, LU Z, KAPPS A, et al. Comparison of trend detection algorithms in the analysis of physiological time-series data [J]. IEEE Transactions on Biomed Engineering, 2005, 52(4): 639-651.
- [2] KOSKI A, JUHOLA M, MERISTE M. Syntactic recognition of ECG signals by attributed finite automata[J]. Pattern Recognition, 1995, 28(12): 1927-1940.
- [3] SHATKAY H, ZDONIK S. Approximate queries and representations for large data sequences [C]// Proceedings of 12th IEEE International Conference on Data Engineering. Washington: IEEE Computer Society, 1996: 546-553.
- [4] KEOGH E, CHU S, HART D, et al. Segmenting time series: A

survey and novel approach [C]// Proceedings of IEEE International Conference on Data Mining. Los Jose: IEEE Computer Society, 2001: 289-296.

- [5] SYLVIE C, CARLOS G B, CATHERINE C, et al. Trends extraction and analysis for complex system monitoring and decision support [J]. Engineering Applications of Artificial Intelligence, 2005, 18(1): 21-36.
- [6] 张贤达. 矩阵分析与应用[M]. 北京: 清华大学出版社, 2004.
- [7] 汪成亮, 陆志坚, 庞栩. 一种数据流趋势分析方法的研究与应用[J]. 计算机系统与应用, 2010, 19(1): 152-156.
- [8] 周黔, 吴铁军. 一种动态数据流的实时趋势分析算法[J]. 控制与决策, 2008, 23(10): 1182-1185.
- [9] 杨云, 孙群, 朱长青. 曲线数据压缩的总体最小二乘算法[J]. 西安电子科技大学学报: 自然科学版, 2008, 35(5): 946-950.
- [10] 周志华, 曹存根. 神经网络及其应用[M]. 北京: 清华大学出版社, 2004.
- [11] LEE S C, HEINBUCH D V. Training a neural-network based intrusion detector to recognize novel attacks [J]. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 2001, 31(4): 294-299.
- [12] 汪成亮, 罗昌银, 庞栩, 等. 基于工况槽况的电解生产控制研究及实现[J]. 轻金属, 2009(4): 24-28.