

文章编号: 1001-0920(2002)06-952-04

# 定性辨识能力增强的数据流定性化方法

朱六璋, 陈宗海

(中国科学技术大学 自动化系, 安徽 合肥 230027)

**摘要:** 给出一种定性辨识能力增强的数据流定性化新方法。这一技术采用线性最小方差拟合和偏度测试构造定性核心函数序偶对在数据流上移动的固定长度观测窗中数据进行核心计算, 将数据流划分为由变量定性值、一阶导数和二阶导数构成的三元组确定的定性状态序列。这种数据定性化过程对定性知识获取和不确定系统描述具有信息完全和直观性、简单性的优点, 降低了辨识算法的复杂性和不确定性。

**关键词:** 系统辨识; 定性建模; 定性推理

**中图分类号:** O 231. 3      **文献标识码:** A

## Qualitative-identification-enhanced method of forming qualitative trends from the measurement stream

ZHU Liu-zhang, CHEN Zong-hai

(Department of Automation, University of Science and Technology of China, Hefei 230027, China)

**Abstract:** A new method is presented to obtain the qualitative trend delineated by the sequence of special qualitative states for qualitative system identification. The consequent qualitative states are defined qualitatively as the triplet of the qualitative value and its first and second derivatives, which did not vary in a time interval. Thus, the procedure transforms the noisy and quantitative information into a simple, intuitive and qualitative representation, which contains important features of the original stream and facilitates the identification algorithm.

**Key words:** system identification; qualitative modeling; qualitative reasoning

## 1 引言

定性系统辨识(QSD)是一种基于不精确系统模型空间和受噪声污染的观测数据流的定性建模方法,其策略的目的在于通过定性输入输出映射来辨识不确定复杂系统的参数和特性<sup>[1]</sup>。通过对受噪声污染的系统输入、输出观测数据流定性化处理,得到相应的定性状态序列(定性行为)作为辨识算法的输入,产生一个约束集作为定性模型输出,实现定性系

统辨识<sup>[2,3]</sup>。数据流定性化是获取辨识算法输入的关键预处理过程,对提高算法的准确性、稳定性和效率至关重要。

目前主要采用样条函数平滑数据流<sup>[4]</sup>和数理统计<sup>[5]</sup>两种数据流定性化方法。本文采用数理统计方法进行数据流定性化处理问题,将线性最小方差拟合<sup>[5]</sup>和偏度测试相结合,在沿测量数据流按一定步距滑动的固定宽度观测窗上构造两个定性核心函

收稿日期: 2001-07-23; 修回日期: 2001-09-05

作者简介: 朱六璋(1969—),男,安徽舒城人,博士生,从事人工智能、定性仿真与定性控制的研究;陈宗海(1963—),男,安徽桐城人,教授,博士生导师,从事人工智能、系统仿真等研究。

数来获得定性状态序列, 通过设计好的划分算法, 每个定性状态是单调的且具有相同的曲线形状(上凸、上凹或直线), 将定性状态定义成由变量的定性值、一阶导数和二阶导数构成的三元组  $v, dv, ddv$ ,  $v, dv, ddv \in \{+, 0, -\}$  且在一定时间区间内保持不变。这里假定观测信号为纯信号与固定标准差高斯噪声的复合且采样频率足够高。由于引入了高阶导数信息, 从而有效解决了定性推理的行为分支不确定性问题, 有利于提高辨识算法的特性。

## 2 数据流用定性状态序列定性表示








在定性推理的语义中, 最一般使用的是具有路标值  $\{+, 0, -\}$  的量空间<sup>[6]</sup>, 每个量根据它是大于零、等于零和小于零分成 3 个定性区间, 这种表示直观地反映量之间相互影响是正向还是反向的。基于此, 一个连续的定量变量的定性状态可以定义成由变量的定性值、一阶导数和二阶导数构成的三元组  $v, dv, ddv$ , 给出各定性值的定义<sup>[4]</sup> 如下

$$\begin{cases} v = \begin{cases} -, & x(t) < 0 \\ 0, & x(t) = 0 \\ +, & x(t) > 0 \end{cases} \\ dv = \begin{cases} -, & \dot{x}(t) < 0 \\ 0, & \dot{x}(t) = 0 \\ +, & \dot{x}(t) > 0 \end{cases} \\ ddv = \begin{cases} -, & \ddot{x}(t) < 0 \\ 0, & \ddot{x}(t) = 0 \\ +, & \ddot{x}(t) > 0 \end{cases} \end{cases} \quad (1)$$

三元组在一定时间区间上保持不变构成了一个定性状态, 若上述三元组中某元素发生变化, 则发生定性状态转移, 产生新的定性状态, 一个测量变量随时间变化的形状和趋势可用定性状态序列表示, 表明了变量的定性行为。

容易看出, 定性值  $v$  给出了变量  $x(t)$  的定性大小, 而  $dv$  和  $ddv$  的不同组合决定了  $x(t)$  的形状, 表 1 给出了所有可能的组合类型<sup>[4]</sup>。这样, 所有可能的定性状态可由表 1 的类型和定性大小(作为上标) 符号

表 1 定性状态的图形特性

type	A	B	C	D	E	F	G
$x$ 's shape							
$dv$	-	-	+	+	-	+	0
$ddv$	-	+	+	-	0	0	0

化表示, 如  $A^+$  或  $B^-$ 。通过这种处理, 含噪声的定量信息可由定性状态序列定性行为表示, 并保持了原来数据流的重要特性和本质。例如,  $x = \sin t, t \in [0, 2\pi]$  可由  $D^+, A^+, B^-, C^-$  定性表示。

## 3 扩展统计函数到数据流定性化过程

### 3.1 运用定性核心函数划分定性状态

定性状态划分将测量数据流分成单调的且具有相同图形特性的区间、极值或拐点的过程。每个区间具有表 1 所示的某一形态, 它提供了过程趋势定性描述的一个片段。定性状态划分过程是通过在沿测量数据流按一定步距滑动的固定宽度观测窗上构造两个定性核心函数进行统计计算实现的。用  $w_i$  表示第  $i$  个窗口, 两个定性核心函数定义  $S(w_i)$  和  $V(w_i)$ , 则  $S(w_i)$  返回  $+, -$  和  $*$  值之一, 分别表示  $w_i$  中包含的是一个单调升(+) 片段, 单调降(-) 片段或单调性不确定( $*$ )。

该函数具体实现方法<sup>[5]</sup> 如下:

1) 对窗口  $w_i$  中的数据点  $(t_i, y_i)$ , 使用式(2) 计算窗口的斜率(slope)

$$\text{slope} = \frac{(t_i - \bar{t})(y_i - \bar{y})}{(t_i - \bar{t})^2} \quad (2)$$

式中,  $\bar{t}, \bar{y}$  为观测窗相应数据的均值。

2) 由于观测数据流中存在噪声影响, 不能直接由 slope 值来决定  $S(w_i)$  的返回值。对于每个窗口, 由

$$\sigma = \frac{\alpha}{\sqrt{(t_i - \bar{t})^2}} \quad (3)$$

得到其标准差。式中  $\alpha$  为含高斯噪声的观测窗数据的标准差。

3) 这里将  $3.5\sigma$  作为阈值, 将式(2) 得到的 slope 与阈值进行比较, 得到  $S(w_i)$  的返回值。若  $|\text{slope}| > 3.5\sigma$  时, slope 不为零的可信度为 99.9%, 且大于零时函数返回  $+$ , 小于零时函数返回  $-$ , 否则, 函数返回  $*$ 。

$V(w_i)$  返回  $+, -$  和  $0$  值之一, 分别表示  $w_i$  包含的曲线是一个上凹(+) 片段, 上凸(-) 片段(或可看作是线性变化的) 或拐点区。

函数  $V(w_i)$  的实现如下: 对窗口  $w_i$  中的数据点  $(t_i, y_i)$ , 使用

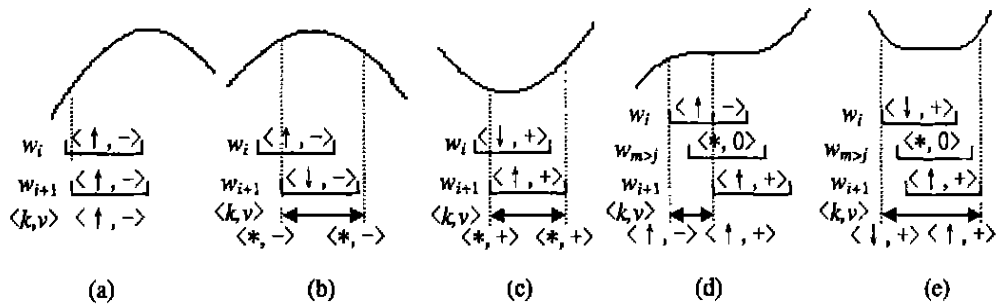


图1 由定性值决定符号

$$\text{skew} = \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^3}{\left[ \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2 \right]^{\frac{3}{2}}} \quad (4)$$

计算窗口的偏度(skewness)。

同样,由于噪声的影响,这里将 $2\sigma^2$ ( $\sigma^2$ 为估计的高斯噪声的方差)作为阈值,若 $\text{skew} > 2\sigma^2$ 则 $V(w_i)$ 返回+,若 $\text{skew} < -2\sigma^2$ 则 $V(w_i)$ 返回-,若 $|\text{skew}| \leq 2\sigma^2$ 则 $V(w_i)$ 返回0。

### 3.2 通过相邻窗口核心函数比较形成定性状态

$S(w_i)$ 和 $V(w_i)$ 构成的序偶表示第 $i$ 个窗口的测量数据特性,为得到定性状态序列,必须在相邻窗口间比较这些序偶。定性状态获取的策略是:设 $s(i), v(i)$ 为第 $i$ 次移动窗口所得定性核心值符号,并考虑相邻窗口 $w_i$ 和 $w_{i+1}$ ,有以下4种情况:

1) 若 $S(w_i) = S(w_{i+1}) = \pm$ 或 $0$ , $V(w_i) = V(w_{i+1}) = \pm$ 或 $0$ ,则 $s(i+1), v(i+1) = S(w_i), V(w_i)$ ,图1(a)给出这种情况中的1个示例;

2) 若 $S(w_i) = \pm, S(w_{i+1}) = \mp, V(w_i) = V(w_{i+1}) = \pm$ ,则在 $w_{i+1}$ 内存在极大值(见图1(b));

若 $S(w_i) = \pm, S(w_{i+1}) = \mp, V(w_i) = V(w_{i+1}) = \mp$ ,则在 $w_{i+1}$ 内存在极小值(见图1(c));

3) 若 $S(w_{i+1}) = *$ ,则在 $w_{i+1}$ 内无法确定极值;

4) 若 $S(w_i) = *$ 且 $V(w_i) = *, S(w_{i+1}) = \pm$ 或 $0$ 且 $V(w_{i+1}) = \pm$ 或 $0$ ,设 $j$ 是最后一个具有符号的数据点,有以下2种情形(分别见图1(d)和图1(e)):

若 $s(j) = S(w_{i+1})$ 且 $v(j) = V(w_{i+1})$ ,则 $j \sim i+1$ 具有相同单调性,因为 $s(j)$ 没有明显变化,但需要判断拐点;若 $j \sim i+1$ 存在一定长度测量数

据流且 $s(j), v(j) = *, 0$ ,则出现两个拐点,否则仅有一个拐点,可将规则2)中 $S(\bullet)$ 与 $V(\bullet)$ 互换进行判断;

若 $s(j) = S(w_{i+1})$ 且 $v(j) = V(w_{i+1})$ ,则当 $s(j) = -$ 时有极大值,当 $s(j) = +$ 时有极小值存在于 $j \sim w_{i+1}$ 的末尾中;若 $j \sim i+1$ 存在一定长度测量数据流且 $s(j), v(j) = *, 0$ ,则有一个极值区及两个拐点,拐点确定方法与中相同。

## 4 仿真实例

将上述方法用于含噪声的数据流定性化过程,在此给出仿真来显示该方法的可行性和正确性。考虑函数

$$y(x) = 4e^{-(x-3)^2} \cos(\pi(x-3)) + \eta$$

式中, $\eta$ 为高斯噪声(零均值、方差0.15),输入 $x$ 在区间 $[0, 6]$ 上均匀分布。选择采样步长为0.005,数据流共有1200个采样数据点,如图2所示。

仿真中选择具有240个数据点的观测窗口,以每次10个数据点的步距沿测量数据流向前移动,于是得到了窗口 $w_i (i = 1, 2, \dots, 97)$ 的斜率和偏度函数如图3所示,窗口定性核心函数序偶如图4所示。为方便显示,图3将skew的值放大3倍,图4中

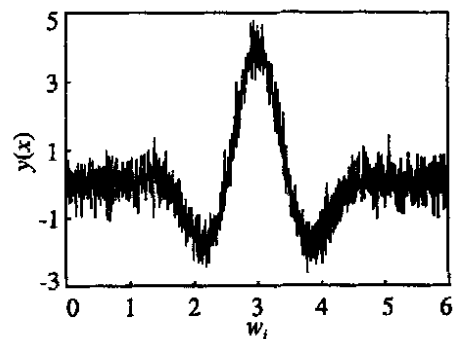


图2 函数采样数据曲线

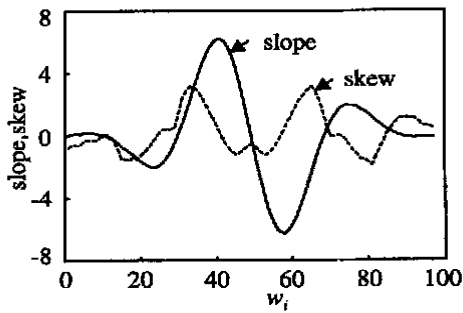


图 3 观测窗曲线斜率与数据偏差

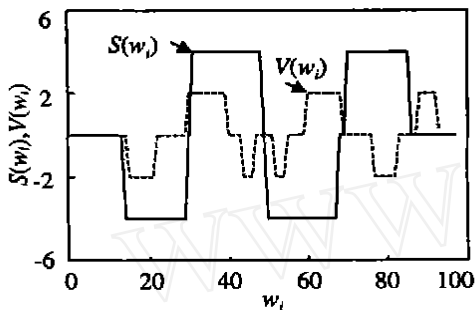


图 4 观测窗的核心函数值

$S(w_i) = 4(0, -4)$  表示  $S(w_i) = (*, -)$ ,  $V(w_i) = 2(0, -2)$  表示  $V(w_i) = +(0, -)$ 。

图 4 定性直观地反映了曲线变化的情况, 运用其数据  $S(w_i), V(w_i)$  和第 3 节所给出的规则定性计算出  $s(j), v(j)$ , 再集成定性状态及相应的区间。根据所编制的程序, 最后可得到定性状态序列为  $G^0 A^- E^- C^- F^+ D^+ G^0 A^+ E^+ F^- A^- B^+ G^0$  和相应的定性区间路标值以及可区分时间点, 这里的定性值(上标)0 表示近似为零。这一定性状态序列与所分析仿真函数的曲线特性基本一致。若该测量数据流用样条函数定性表示, 很难确定复杂的样条函数, 简化的样条函数使知识表示不完全。需说明, 观测窗宽度的选择对结果有影响, 一般尽量使窗口所含数据不失去趋势特征, 有关窗口宽度选择问题值得研究。

## 5 结 论

数据流定性化处理是定性系统辨识中关键性问题之一, 是定性推理和仿真实际应用的需要。本文引入斜率与偏度测试两种统计方法构造定性核心函数来判断数据的形状特性, 通过在数据流上移动固定长度观测窗, 并运用函数序偶进行核心计算, 利用规则比较相邻窗口核心函数值, 将数据流划分为定性状态序列, 并给出仿真过程。理论计算和仿真结果表明, 该数据流定性化方法对定性知识获取和不确定系统描述具有信息完全和直观性、简单性等优点, 降低了辨识算法复杂性和推理过程的行为分支不确定性问题。

### 参考文献(References):

- [1] Say A C C, Kuru S. Qualitative system identification: Deriving structure from behavior[J]. *Artificial Intelligence*, 1996, 83(1-2): 75-141.
- [2] Lichtenberg G, Lunze J. Identification of discrete event models for continuous-variable systems[A]. *UKACC Int Conf on Control 96*[C]. Edinburgh, 1996 1: 711-715.
- [3] Lo C H, Chow K M, Wong Y K, et al. Qualitative system identification with the use of on-line genetic algorithms[J]. *Simulation Practice and Theory*, 2001, 8(6-7): 415-431.
- [4] Schaich D, King R, Keller U, et al. Interval identification - A modelling and design technique for dynamic systems[A]. *13th Int Workshop on Qualitative Reasoning*[C]. Loch Awe, 1999 1-8.
- [5] Kay H, Rinner B, Kuipers B. Semi-quantitative system identification[J]. *Artificial Intelligence*, 2000, 119(1-2): 103-140.
- [6] De Kleer J, Brown J S. A qualitative physics based on confluences[J]. *Artificial Intelligence*, 1984, 24(1): 7-83.