

文章编号: 1002-0411(2002)02-102-04

数据流定性化方法研究

周 颢 邵晨曦 白方周

(中国科学技术大学计算机科学技术系 合肥 230027)

摘要: 数据流定性化处理,是在考虑噪声影响的情况下,对数据流进行处理,最终将数据流化分为单调变化的区段.这是定性仿真应用中的一个关键问题.本文在分析基于线性最小方差拟合的数据流定性化方法的基础上,改进其定性核心函数.提出了基于偏度系数的数据流定性化方法.并通过具体数据分析,证明了该方法的有效性.*

关键词: 数据流定性化;定性核心函数;偏度系数

中图分类号: TP13

文献标识码: B

RESEARCH OF FORMING THE QUALITATIVE TREND FROM THE MEASUREMENT STREAM

ZHOU Hao SHAO Chen-xi BAI Fang-zhou

(Department of Computer Science and Technology University of Science and Technology of China Hefei 230027)

Abstract: The qualitative trend is generated through a process which breaks the measurement stream into monotonic segments. It is important for the applications of qualitative simulation. After analyzed the method which using linear least-squared fit, we modified the qualitative kernel function and presented the method which using deflecting modulus. In the end of the article, we demonstrated the effectiveness of the method.

Keywords: qualitative trend, qualitative kernel function, deflecting modulus

1 引言(Introduction)

定性仿真^[1]思想由 Kuipers 于 1986 年提出,它的结果是系统的定性行为序列.在实际应用中,如系统模型精炼、仿真控制等^[2],往往需要将仿真结果与实际观测得到的数据流进行匹配和验证.为了实现这一目的,必须处理从实际观测得到的各个参数的数据流,从中得到该参数的定性状态.

在实际情况下,可以认为观测到的数据流是实际信号与一定噪声的复合.数据流定性化处理过程就是要在考虑到噪声影响的情况下,对数据流进行处理.最后结果是将数据流化分为若干单调变化的区段,每段有对应的符号(↑、↓、⊙)来表示其定性状态(上升、下降和顶值).在此基础上,可以通过基于神经网络的预测算法^[3,4]确定各个单调区段的上下界. Bernhard Rinner 等人,在其 SQUID 系统中,提出了使用基于线性最小方差算法来完成数据流定性

化处理^[5,6].但是由于该方法存在的边界效应问题,使得其对定性状态区段的边界定位上并不能获得令人满意的效果.本文提出基于偏度系数的数据流定性化处理方法,有效地解决了这些问题.

由于处理过程中必须考虑到观测得到的数据流中噪声的干扰.在本文讨论的情况中,假设存在以下一些前提:

- 观测得到的信号认为是纯信号和高斯噪声的复合.其中的高斯噪声是无意义的,并且具有固定的标准差.
- 采样频率足够高,可以反映纯信号的特性.但并不要求有固定的采样频率.
- 噪声的标准差已知,具体数值可能并不准确.

2 基于线性最小方差拟合的数据流定性化处理方法(Data stream processing method

* 收稿日期: 2001-11-14
基金项目: 国家自然科学基金(编号: 69974038)资助项目, 国家高性能计算基金(00215)支持

using linear least-squared fit)

该方法由 Bernhard Rinner 等人在 SQUID 系统中提出. 其核心思想是在数据流上移动一个固定长度的窗口, 利用基于线性最小方差拟合的定性核心函数(Qualitative Kernel Function)得到该窗口上的单调性(上升、下降或不确定). 再通过若干窗口之间的比较, 得出观测数据流的定性状态.

2.1 定性核心函数

定性核心函数是该处理的核心. 该函数处理由第 i 个数据点开始的窗口 W_i , 窗口的长度为 T_i , 函数 $k(W_i)$ 返回如下的结果:

- \uparrow w_i 中包含的是一个单调升的段;
- \downarrow w_i 中包含的是一个单调降的段;
- $*$ w_i 的单调性不确定.

该函数具体实现方法如下:

1) 对窗口中的数据点 (t_i, y_i) , 使用公式(1)计算窗口的坡度(slope).

$$\text{slope} = \frac{\sum_i (t_i - \bar{t})(y_i - \bar{y})}{\sum_i (t_i - \bar{t})^2} \quad (1)$$

2) 由于在观测数据流中存在噪声影响, 不能直接由 slope 的值来决定函数的返回值. 对每个窗口, 由公式(2)得到其标准差.

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

其中, σ_v 是包含高斯噪声干扰的观测数据流的标准差.

在 SQUID 系统中, 将 3.5σ 作为阈值. 将公式(1)得到的 slope 的值与阈值进行判断, 从而得到函数的返回值. 例如, $|\text{slope}| \leq 3.5\sigma$ 时, 函数返回* .

2.2 相邻窗口比较处理

由于存在噪声的影响, 不能直接利用定性核心函数的结果来进行单调性判断, 所以必须利用相邻窗口比较得到正确的结果.

在利用定性核心函数判断出各窗口的单调性之后, 如图 1 所示, 该方法利用相邻窗口的单调性进行比较, 从而将数据流进行划分, 完成数据流定性化处理.

- (1) 如果 $k(W_i) = k(W_{i+1}) = \uparrow$ 或 \downarrow , 则 $\text{Sign}(i+1) = k(W_{i+1})$;
- (2) 如果 $k(W_i)$ 与 $k(W_{i+1})$ 结果不同(\uparrow, \downarrow), 则在 w_{i+1} 中存在极大(小)值;
- (3) $k(W_{i+1}) = *$, 则无法判断;

(4) $k(W_i) = *$, $k(W_{i+1}) = \uparrow$ 或 \downarrow , 设 $j < i$ 是最后一个具有符号的数据点.

(a) $\text{Sign}(j) = k(W_{i+1})$, j 至 $i+1$ 具有与 $k(W_{i+1})$ 相同的符号;

(b) $\text{Sign}(j) \neq k(W_{i+1})$, 则 j 至 w_{i+1} 的末尾中存在一个极大(小)值.

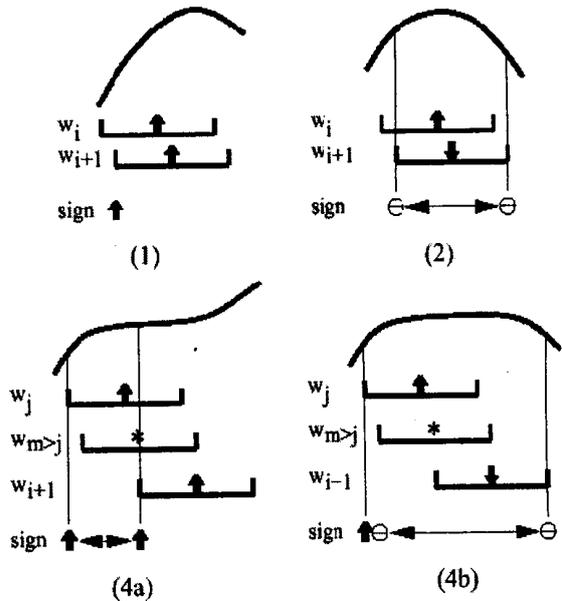


图 1 利用相邻窗口比较处理确定符号

Fig. 1 Determining signs from kernel values

3 基于偏度系数的数据流定性化处理 (Data stream processing method using deflecting modulus)

在我们的研究工作中, 提出了基于偏度系数的数据流定性化处理. 本方法借鉴了基于线性最小方差拟合的思想, 主要是修改了其定性核心函数, 由偏度系数来确定函数的返回值.

3.1 偏度系数

偏度系数 γ_1 的定义:

$$\gamma_1 = \frac{m(\xi - \mu)^3}{\sigma^3} \quad (3)$$

其中:

$$m(\xi - \mu)^3 = \sum_{i=1}^n (x_i - \mu)^k P_i \quad (4)$$

称为 3 阶中心矩.

偏度系数是对分布函数曲线对称程度的量度. 如图 2 所示, 由偏态系数的正负性可以得到分布函数曲线的单调性.

$\gamma_1 > 0$ 曲线下降; $\gamma_1 = 0$ 曲线对称; $\gamma_1 < 0$ 曲线上升.

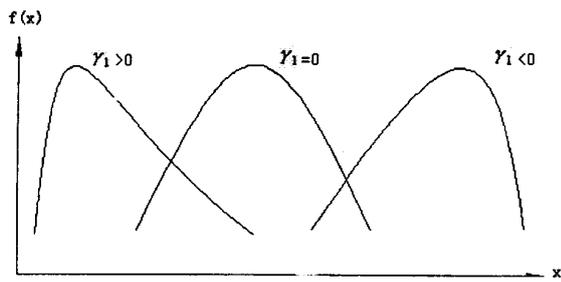


图 2 偏度系数

Fig. 2 Deflecting modulus

3.2 基于偏度系数的定性核心函数

本方法中,定性核心函数使用偏态系数来度量该窗口中数据的单调性.

在定性核心函数中,将 T_i 视为样本值, Y_i 视为相应样本值的分布. 由于在分布函数曲线中,样本值的分布应该为正值,所以在处理前,先对 Y_i 进行调整,使其全部为正值. 这样的调整,并不会改变曲线的单调性,对单调性的判断不会造成影响.

调整完成后,由公式(3)得到其偏态系数,然后与给定的阈值比较,即可得到函数的返回值(\uparrow 、 \downarrow 或 $*$).

定性核心函数完成后,相邻窗口间的比较处理与基于线性最小方差的方法基本类似.

4 分析(Analyze)

基于线性最小方差拟合与基于偏度系数的这两种数据流定性化处理方法,其差别主要表现在其定性核心函数上.

基于线性最小方差拟合的定性核心函数中使用的公式(1)的数学意义不明确,且具有明显的边界效应,对图 3 所示的数据窗口,虽然下降的段很少,但由于其处在边界上,所以该函数仍然返回 \downarrow . 具体数据见表 1.

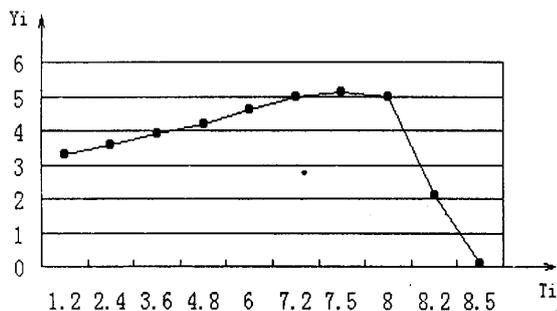


图 3 数据窗口

Fig. 3 Data windows

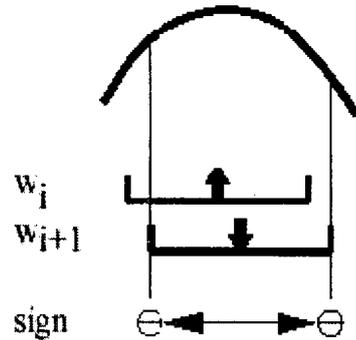
表 1 窗口数据

Tab. 1 Datas

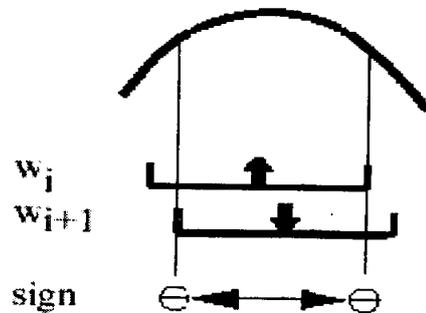
I	时间 T_i (Time)	值 Y_i (Value)
1	1.2	3.3
2	2.4	3.6
3	3.6	3.9
4	4.8	4.2
5	6.0	4.6
6	7.2	5.0
7	7.5	5.1
8	8.0	5.0
9	8.2	2.1
10	8.5	0.1

对表 1 的数据,运用上述定性核心函数得到 $\text{slope} = -8.296$; 返回结果为 \downarrow .

由于存在这样的问题,使得在进行相邻窗口比较处理时,不能很准确地定义顶点 \odot 的起止界.如图 1 中的情况(2),得到的起止界为 $(T_{i+1}, T_{i+1} + T_1)$. T_1 是窗口的长度,而不能是 $(T_{i+1}, T_i + T_1)$.



线性最小方差拟合
Linear least-squared fit



偏态系数
deflecting modulus

图 4 偏态系数方法可以得到更精确的界定
Fig. 4 Deflecting modulus method can give more precise boundary

在基于偏度系数的方法中, 使用了偏度系数. 该系数具有严格的数学意义. 并且杜绝了公式(1)中存在的边界效应. 例如在运用本方法处理表 1 的数据, 偏态系数 $\gamma_1 = -0.602$. 得到的结果为 \uparrow .

该方法可以对顶值的起止界给出更精确的界定. 如图 4 所示, 在相邻窗口间的比较处理的情况(2), 得到的起止界为 $(T_{i+1}, T_i + T_1)$. T_1 是窗口的长度.

从处理结果, 我们可以看出, 基于偏度系数的数据流定性化处理方法无论是从数学意义或是处理结果的准确性角度, 都优于运用基于线性最小方差的数据流定性化处理方法.

5 结束语(Conclusion)

数据流定性化处理是定性仿真应用中的一个关键性问题. 实际应用情况的需要, 要求数据流定性化处理方法能够在存在噪声干扰的情况下, 尽可能准确地处理观测得到的数据流, 抽取其定性特征, 为进一步进行模型精炼、仿真控制等应用作准备.

本文从数学上偏度定义出发, 改进了 Bernhard Rinner 等人的工作, 提出了基于偏度系数的数据流定性化处理方法. 并通过理论与实际结果分析, 证明了该方法的优异性.

参 考 文 献 (References)

- 1 Kuipers B. J. (1986). Qualitative simulation. *Artif. Intell.*, 29, 289~ 338
- 2 白方周, 张 雷. 定性仿真导论, 中国科学技术大学出版社, 1998, 218~ 244
- 3 Herbert Kay and Lyle Ungar. Estimating Monotonic Functions and Their Bounds using MSQUID. Technical Report TR A199-280, University of Texas at Austin, 1999
- 4 Herbert Kay and Lyle H. Ungar. Deriving Monotonic Function Envelopes from Observations. In Working Papers from the Seventh International Workshop on Qualitative Reasoning about Physical System (QR-93), pages 117-123, Oreas Island, Washington, 1993
- 5 H. Kay, B. Rinner and B. Kuipers. Semi-Quantitative System Identification. Technical Report TR A199-279, University of Texas at Austin, 1999
- 6 B. Rinner, B. Kuipers. Monitoring Piecewise Continuous Behaviors by Refining Semi-Quantitative Trackers, 16th IJCAI-99

作者简介

周 颢(1976-), 男, 博士研究生. 研究领域为定性仿真及其应用.

邵晨曦(1954-), 男, 副教授. 研究领域为系统仿真, 复杂系统行为和模型.

白方周(1933-), 男, 教授. 研究领域为定性建模与定性仿真.