

文章编号:1001-9081(2006)10-2395-03

基于小波分解和聚类模糊系统的时间序列预测

黄景春^{1,2}, 肖 建¹

(1. 西南交通大学 电气工程学院, 四川 成都 610031;
2. 兰州交通大学 机电与动力工程学院, 甘肃 兰州 730070)
(jdhjjcc@sina.com)

摘要:结合小波分析的多分辨特性和模糊规则的可解释性,提出了一种非平稳时间序列预测方法。首先将时间序列信号分解为各尺度上的细节信号和平滑信号,使用软、硬阈值折衷法消除噪声,并对各层重构信号单独使用聚类模糊系统进行预测,最后将各预测结果累加得到对整个时间序列的预测。仿真试验表明此方法是有效的。

关键词:小波分解; 模糊系统; 时间序列; 预测

中图分类号: TP311.13 文献标识码:A

Prediction of time series based on wavelet decomposition and clustering fuzzy systems

HUANG Jing-chun^{1,2}, XIAO Jian¹

(1. School of Electrical Engineering, Southwest Jiaotong University, Chengdu Sichuan 610031, China;
2. School of Mechatronics and Power Engineering, Lanzhou Jiaotong University, Lanzhou Ganshu 730070, China)

Abstract: A prediction method for non-stationary time series was proposed in association with multi-resolution of wavelet analysis and interpretability of fuzzy rules. The original time series were decomposed into the smooth and the detailed at different levels. After being denoised with the method of soft-hard threshold value, the smooth and the detailed at different levels were forecasted with the clustering fuzzy systems. Finally the sum of the forecasting results at different levels was the prediction of the original time series. Experiments show that the method is effective.

Key words: wavelet decomposition; fuzzy systems; time series; predict

0 引言

时间序列的预测在天气预报、经济学、信号处理、产品的生产和销售等诸多领域有着广泛的应用。对于平稳时间序列,现有的预测方法如移动平均模型(MA)、自回归模型(AR)、混合自回归移动平均模型(ARMA)、自回归积分移动平均模型(ARIMA)等都有很好的预测效果^[1]。但是很多实际系统往往是非平稳的,同时具有很强的非线性,上述这些预测方法对于非平稳的时间序列的预测效果不理想,因此尝试使用人工智能的方法比如:神经网络、灰色模型、遗传算法、支持向量机等来预测非线性、非平稳的时间序列成为研究重点^[2~5]。模糊系统是一种基于规则的非线性万能逼近器^[6],由于规则的可解释性,易于将人们的知识融入系统中,同时整个系统是依靠逻辑推理进行的,因此利用模糊系统预测时间序列具有独特的优势。小波分析由于具有时频局部化和多分辨率特性,非常适合处理非平稳信号^[7],利用小波变换将非平稳时间序列的各种趋势分解到不同层次上分别研究,可以减小各种趋势混杂对整个系统预测的影响。因此本文建立了一种基于小波分解和聚类模糊系统的时间序列预测方法,通过结合小波分析和模糊系统的各自优势来提高非平稳时间序列的预测效果。

1 小波分解和聚类模糊系统

1.1 时间序列的小波分解和消噪

对于任意函数 $f(t) \in L^2(R)$ 其连续小波变换定义为:

$$W_f(a, b) = \int_R |a|^{-\frac{1}{2}} f(t) \psi\left(\frac{t-b}{a}\right) d(t) \quad (1)$$

其中 a 为伸缩因子, b 为平移因子, $\psi(t)$ 为基小波, 基小波需满足如下容许性条件:

$$C_\psi = 2\pi \int_R \frac{|\hat{\psi}(\xi)|^2}{|\xi|} d\xi < \infty \quad (2)$$

为了计算方便和有利于在计算机上实现,通常将伸缩和平移因子离散化得到离散小波级数,函数 $f(t) \in L^2(R)$ 的小波级数展开为:

$$f(t) = \sum_{j,k \in Z} C_{j,k} \psi_{j,k}(t) \quad (3)$$

其中 $C_{j,k}$ 为 $f(t)$ 分解到小波 $\psi_{j,k}(t)$ 的系数:

$$C_{j,k} = 2^{\frac{j}{2}} \int_R f(t) \overline{\psi(2^j t - k)} d(t) \quad (4)$$

由于计算每个小波系数都要计算一个定积分,计算工作量也是很大的,因此 Mallat 在多分辨率分析的基础上提出了计算小波系数的快速递推算法。

Mallat 分解算法为:

$$c_{j,k} = \sum_n \bar{h}_{n-2k} c_{j+1,n} \quad (5)$$

收稿日期:2006-04-07; 修订日期:2006-06-20 基金项目:教育部博士点基金资助项目(20040613013)

作者简介:黄景春(1973-),男,吉林大安人,讲师,博士研究生,主要研究方向:模糊控制、机车粘着控制; 肖建(1950-),男,湖南衡阳人,教授,博士生导师,主要研究方向:计算机控制系统、鲁棒控制、电传动系统控制技术。

$$d_{j,k} = \sum_n \bar{g}_{n-2k} c_{j+1,n} \quad (6)$$

Mallat 重构算法为:

$$c_{j+1,k} = \sum_l h_{k-2l} c_{j,l} + \sum_l g_{k-2l} d_{j,l} \quad (7)$$

对于给定时间序列 $X = \{x_t, t = 1, 2, \dots, n\}$, 其中 x_t 是对应 t 时刻的数据值。利用 Mallat 塔式算法, 对时间序列 X 进行 N 层尺度分解, 得到 N 组细节信号 $w_i (i = 1, 2, \dots, N)$ 和一组平滑信号 v_N 。采用软、硬阈值折衷法来消除噪声, 其计算公式如下:

$$\hat{w}_i = \begin{cases} \text{sign}(w_i) \cdot (|w_i| - a\lambda) & |w_i| \geq \lambda, 0 \leq a \leq 1 \\ 0 & |w_i| < \lambda \end{cases} \quad (8)$$

式中 \hat{w}_i 为消除噪声后的细节信号, λ 为阈值, 其计算公式为:

$$\lambda = \sigma \sqrt{2 \lg(N)} \quad (9)$$

式中 N 为信号的长度, σ^2 为信号噪声的方差。对于平滑信号采用同样的方法得到消除噪声后的平滑信号 \hat{v}_N 。然后对每组细节信号 \hat{w}_i 和平滑 \hat{v}_N 单独利用 Mallat 塔式算法重构到原尺度上, 得到重构的细节信号 $\tilde{w}_i (i = 1, 2, \dots, N)$ 和平滑信号 \tilde{v}_N 。则消除噪声后的时间序列 X' 可以表示为:

$$X' = \sum_{i=1}^N \tilde{w}_i + \tilde{v}_N \quad (10)$$

式中 \tilde{w}_i 表示重构的第 i 层尺度上的细节信号, \tilde{v}_N 表示重构第 N 层尺度上的平滑信号^[8]。

1.2 聚类法设计的模糊系统

利用小波变换分解后得到的时间序列, 在每层上分别使用模糊系统进行预测, 模糊系统采用文献[9]提出的最近邻聚类法设计的模糊系统。原算法在模糊系统设计的过程中, 采用递推算法动态改变整个模糊系统, 实际上整个模糊系统只有在经过一定的训练后才能有满意的逼近效果。因此本文在训练模糊系统过程中, 不改变模糊系统结构以提高计算速度, 而在预测过程中采用动态改变模糊系统结构的方法, 具体算法如下:

1) 从第一对输入输出数据 $A_1(x_1^1, x_2^1, \dots, x_p^1, y_1)$ 开始, 将 $(x_1^1, x_2^1, \dots, x_p^1)$ 设为第一个聚类中心 X_c^1 , 同时记录 $y_c^1 = y_1$, $m_c^1 = 1$ 。

2) 假定考虑第 k 对数据 $A_k(x_1^k, x_2^k, \dots, x_p^k, y_k)$ 时, 已经有 M 个聚类中心 $X_c^1, X_c^2, \dots, X_c^M$, 计算 $(x_1^k, x_2^k, \dots, x_p^k)$ 与各聚类中心的距离, 假设与 X_c^t 类的距离最小并记为 r_{\min} 。

3) 设定半径 r , 如果 $r_{\min} \geq r$, 则产生一个新的聚类中心 X_c^{M+1} , 并令 $y_c^{M+1} = y_k, m_c^{M+1} = 1$; 如果 $r_{\min} < r$, 则数据 A_k 属于 X_c^t 类, 并令 $y_c^t = y_c^t + y_k, m_c^t = m_c^t + 1$ 。

4) 所有训练数据聚类完毕后, 假设共有 L 个聚类, 最后利用聚类方法得到的模糊系统计算公式为:

$$f(x) = \frac{\sum_{i=1}^L y_c^i \exp(-\frac{|x - x_{c1}^i|}{\sigma}) \exp(-\frac{|x - x_{c2}^i|}{\sigma}) \dots \exp(-\frac{|x - x_{cp}^i|}{\sigma})}{\sum_{i=1}^L m_c^i \exp(-\frac{|x - x_{c1}^i|}{\sigma}) \exp(-\frac{|x - x_{c2}^i|}{\sigma}) \dots \exp(-\frac{|x - x_{cp}^i|}{\sigma})} \quad (11)$$

1.3 基于分解和聚类的预测算法

小波分解和聚类模糊系统的预测算法如图 1 所示。首先利用小波变换对时间序列进行 N 层尺度分解, 使用软、硬阈值折衷法对分解后的信号消除噪声; 然后对每层分解信号单独

重构到原尺度上, 利用聚类模糊系统对各层信号单独预测; 最后累加各层预测结果得到最终的预测结果 $x(n+1)$ 。

$$x(n+1) = \sum_{i=1}^N \tilde{w}_i(n+1) + \tilde{v}_N(n+1) \quad (12)$$

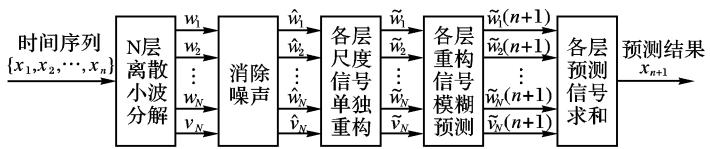


图 1 小波分解和聚类模糊系统的预测算法

2 仿真实验与结果分析

预测 Mackey-Glass 混沌时间序列, 其方程为:

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \quad (13)$$

令 $x(k) (k = 1, 2, 3, \dots)$ 为由式(13)在初值 $\tau = 30$, $x(0) = 0.8$ 产生的时间序列数据每隔 1s 抽样同时加入服从正态分布的均值为 0, 方差为 0.08 的随机噪声信号得到。图 2 给出了 400 个数据点。使用经过输入输出对训练的小波分解和聚类模糊系统模型, 对这 400 个点进行预测。图 3 为使用小波分解和聚类模糊系统模型的预测结果, 选用 Daubechies 小波, 分解尺度为 6, 聚类半径 $r = 0.15$, 参数 $\sigma = 0.2$ 。为了便于对比, 单独使用聚类模糊系统模型和 AR 模型进行同样的预测, AR 模型也选用 7 阶, 聚类模糊系统的聚类半径 $r = 0.15$, 参数 $\sigma = 0.2$ 。可以看出小波分解和聚类模糊系统模型的预测效果比其他两类预测方法的效果好, 主要是利用模糊系统对非线性、非平稳的系统有很好的逼近能力, 以及小波变换的消噪能力和多分辨率特性, 在不同尺度上分别使用聚类模糊系统预测从而提高了整个预测模型的精度。

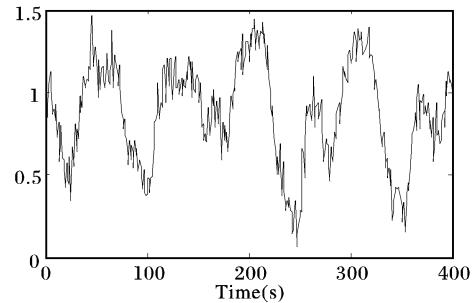


图 2 一段含有噪声的 Mackey-Glass 混沌时间序列

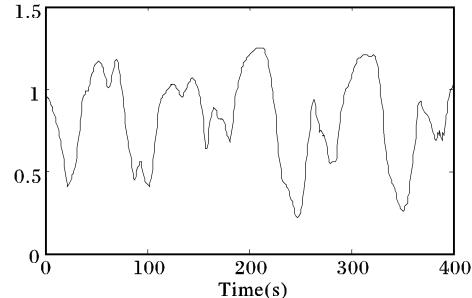


图 3 小波分解和聚类模糊系统模型预测结果

表 1 Mackey-Glass 混沌时间序列预测结果

预测方法	MSE
AR(7)	0.0039
聚类模糊系统	0.0037
小波分解和聚类模糊系统模型	0.0032

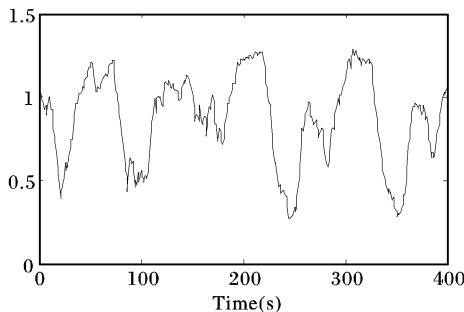


图 4 聚类模糊系统模型预测结果

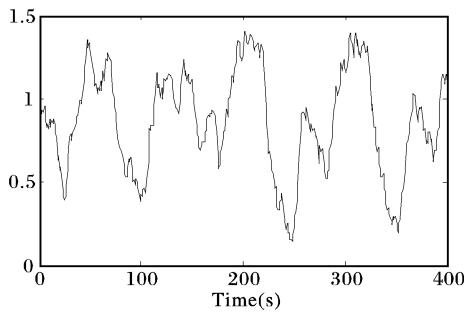


图 5 AR(7) 模型预测结果

3 结语

结合小波分析的时频局部化和多分辨特性,以及模糊系统的基于规则和推理的特性,本文提出了一种基于小波分解和聚类模糊系统的预测方法。仿真结果表明,此方法用来预测一类非平稳的时间序列是有效的。由于模糊系统是一种基于规则的逻辑推理系统,因此易于将人们的知识和经验融入

系统中,这对贫信息条件下的系统辨识和控制提供了一种有效的方法。进一步的研究工作主要在:1)设计基于最少规则且满足要求的模糊系统。最近邻聚类方法设计模糊系统,实际上聚的类对应于模糊系统的规则数。聚的类多,模糊系统复杂;聚的类少,模糊系统过于简单,精度差。2)针对一类强非线性、时变性、贫信息的系统(比如机车粘着控制系统),构建基于小波分解和聚类模糊系统的控制器,以提高其控制性能。

参考文献:

- [1] 王志贤. 最优状态估计与系统辨识[M]. 西安: 西北工业大学出版社, 2004.
- [2] KABOUDAN MA. Genetic programming prediction of stock prices [J]. Computational Economics, 2000, 16: 207 - 236.
- [3] HANSEN JV, NELSON RD. Neural networks and traditional time series methods: a synergistic combination in state economic forecasts [J]. IEEE Trans. Neural Networks, 1997, 8(4): 863 - 873.
- [4] 邓聚龙. 灰色系统理论教程[M]. 武汉: 华中理工大学出版社, 1990.
- [5] FLOYD S, WARMUTH M. Sample compression, learn ability and the Vapnik-Chervonenkis dimension [J]. Machine Learning, 1995, 21(3) : 269 - 304.
- [6] KOSKO B. Fuzzy systems as universal approximators [J]. IEEE Trans. Computers, 1994, 43(11): 1329 - 1333.
- [7] DAUBECHIES I. Ten Lectures on Wavelets[M]. The Society for industrial and applied mathematics, Philadelphia. 1992.
- [8] 冯象初, 甘小冰, 宋国乡. 数值泛函与小波理论[M]. 西安: 西安电子科技大学出版社, 2003.
- [9] 王立新. 模糊系统与模糊控制教程[M]. 王迎军, 译. 北京: 清华大学出版社, 2003: 150 - 164.

(上接第 2394 页)

比即为 74%, 有效减少了每个用户对应的网站数, 降低了代理服务器端 WUM 的复杂度。

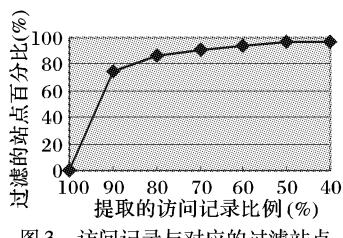
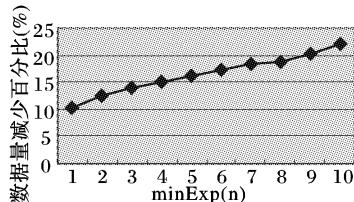


图 3 访问记录与对应的过滤站点

图 4 给出了分别将最小经验值 minExpV 设为 1 ~ 10, 减少的数据量百分比。

图 4 minExpV 与对应减少的数据量

另外,采用关联规则算法 Apriori 算法^[3] 分别在数据 1 和数据 2 上进行挖掘,发现频繁模式。对发现的频繁模式内容进行分析发现,经过基于兴趣特征的数据预处理后的数据 2,能够提升经验用户的访问模式的支持度和置信度,突出显著模式。将这些模式推荐给用户更能体现推荐系统的推荐价值。

3 结语

为了能够在尽量不丢失有价值信息的情况下降低代理服务器端 WUM 的数据规模,并从行为日志中发现更有推荐价值的访问模式,提出了基于用户兴趣特征的数据预处理方法。该方法根据用户兴趣的心理学特征,充分考虑用户兴趣的品质,将用户兴趣分为长期兴趣和短期兴趣。将那些偶然路过的短期兴趣用户的访问记录过滤掉,降低对长期经验用户访问模式的干扰,发现更有意义的行为模式。实验结果表明该方法能够过滤掉噪音数据,降低数据规模,从而减小代理端日志挖掘的复杂度,提高基于 WUM 进行个性化推荐的准确度。

参考文献:

- [1] NARASIMHAN S. An integrated approach to diagnosis of complex hybrid systems [A]. 15th Annual International Symposium on AeroSense[C]. Orlando, Florida, US: IEEE press, 2001. 309 - 322.
- [2] WANG W-H, LIU K-D, ZHOU D-H, et al. A fuzzy and rough sets integrated approach to fault diagnosis[A]. Proc. of the 15th World Congress of Int Federation of Automatic Control [C]. Barcelona, Spain: IFAC press, 2002. 1031 - 1037.
- [3] AGRAWAL R, SRIKANT R. Fast algorithms for mining association rules in large databases[A]. Proc. of the 20th International Conference on Very Large Databases[C]. Santiago, Chile: 1994. 487 - 499.
- [4] 高玉祥. 个性心理学[M]. 第 2 版. 北京: 北京师范大学出版社, 2005.
- [5] 郭岩, 白硕, 杨志峰, 等. 网络日志规模分析和用户兴趣挖掘[J]. 计算机学报, 2005, 28(9): 1483 - 1496.