

基于模糊模型支持向量机的混沌时间序列预测^{*}

崔万照 朱长纯 保文星 刘君华

(西安交通大学电子与信息工程学院, 西安 710049)

(2004 年 9 月 3 日收到 2004 年 11 月 10 日收到修改稿)

基于支持向量机强大的非线性映射能力和模糊逻辑易于将先验的系统知识结合到模糊规则的特性, 根据混沌动力系统的相空间重构理论, 提出了一种混沌时间序列的模糊模型的支持向量机预测模型, 并采用适用于大规模问题求解的最小二乘法来训练预测模型. 利用该模型分别对模型的整体预测性能与嵌入维数及延迟时间的关系进行了探讨. 最后利用 Mackey-Glass 时间序列和典型的 Lorenz 系统生成的时间序列对该模型进行了验证, 结果表明该预测模型不仅能够自动的从学习数据中获取知识产生模糊规则, 提取能够代表混沌时间序列内在规律的支持向量, 大大减少支持向量的数目, 精确地预测未来的混沌时间序列, 而且在混沌时间序列的嵌入维数未知和延迟时间不能合理选择的情况下, 也能取得比较好的预测效果. 这一结论预示着基于模糊模型的支持向量机是一种研究混沌时间序列的有效方法.

关键词: 模糊模型, 混沌时间序列, 支持向量机, 最小二乘法

PACC: 0545

1. 引 言

混沌现象是自然界普遍存在的一种不规则运动, 是一种来源于非线性动力系统内部的、本身固有的复杂的动态行为. 近年来, 随着混沌理论研究的不断深入及其在信号处理、自动控制和通信等领域中广泛的应用, 混沌时间序列的建模和预测已成为混沌信号处理研究领域的一个非常重要的研究方向^[1-5]. 混沌现象的一个独特之处是一个完全确定的、简单的非线性系统模型也会产生类似随机过程的复杂行为特征, 如果能够发现非线性系统的内在的规律并建立其模型, 则对其不确定性的未来的预测就有可能. Takens 的嵌入定理提供了预测混沌时间序列的理论依据, 但是如何根据相空间重构理论构造预测模型是混沌时间序列预测中的一个关键问题^[4,5], 为此人们提出了许多预测混沌时间序列的非线性预测方法. 这些方法大致可分为全局预测法^[6,7]、局域预测法^[2]、自适应非线性滤波预测法^[4]和基于神经网络的预测法^[8-11]. 其中基于相空间重构的基本思想和混沌时间序列产生的确定性非线性机理的前馈型神经网络非线性预测法, 对一些低维混沌序列的预测结果表明它能够进行精确的预测,

但是由于前馈型神经网络在训练中最小化样本点误差^[12,13], 因而不可避免地出现“过拟合”现象, 导致模型的泛化能力受到了限制, 而且还存在诸如隐层节点数的选择等问题, 从而大大限制了其进一步的应用. 而基于 Vapnik 等提出的统计学习理论^[14]的支持向量机(support vector machine, SVM)采用结构风险最小化准则, 在最小化样本点误差的同时缩小模型泛化误差的上界, 从而提高了模型的泛化能力. 这一优点在小样本学习中更为突出, 并且不依赖于系统的数学模型, 同时具有自学习自调整模型的特点, 能对各种混沌系统产生较好的预测效果^[15,16], 但是由于其训练后获得的支持向量数量与学习样本的数量一样多, 导致了当学习样本数目很大时, 得到的模型的支持向量数很大, 这样在预测未来的时间序列时的计算量也很大, 因而限制了其在实际中的广泛应用. 而模糊逻辑^[17,18]使人们易于将先验的系统知识结合到模糊规则中, 能够自动的从学习样本中提取典型的数据作为最终学习样本, 减少计算量, 已被成功地用于神经网络中并取得了良好的效果^[19]. 而基于模糊模型的支持向量机^[20]继承了支持向量机和模糊系统的优点, 使模型能够自动的提取混沌时间序列中具有内在规律的数据作为支持向量, 这样大大减少了模型的支持向量的数目, 因此在实际未知

^{*} 国家自然科学基金(批准号 60176020 和 60276037) 教育部博士点基金(批准号 20020698014) 资助的课题.

的混沌系统的研究中更具有很好的应用前景.

本文基于混沌动力系统的相空间重构理论和支持向量机思想的相似性,根据混沌系统固有的确定性和非线性,结合模糊逻辑推理系统,设计了一种基于模糊模型的支持向量机的混沌时间序列的预测模型,并利用适用于大规模问题求解的最小二乘方法对建立的模型进行训练,以具有时滞特性的 Mackey-Glass 时间序列和 Lorenz 系统产生的时间序列为例验证该模型的建模能力,并分别讨论了模型的预测性能与嵌入维数及延迟时间的关系.结果表明,基于模糊模型的支持向量机混沌时间序列的预测方法能够从学习样本数据中获取知识并提取代表混沌时间序列的内在规律的数据作为支持向量,不仅能够大大减少支持向量的数量,精确的预测未来的时间序列,而且在嵌入维数未知的情况下也能得到好的预测效果,同时发现延迟时间的选择对于该模型的预测性能影响并不是很大,因此基于模糊模型的支持向量机适合于实际中未知的具有混沌特性的时间序列的预测.

2. 系统模型结构

自 1965 年 Zadeh 创立模糊系统理论以来,模糊理论的研究和应用技术得到了蓬勃发展,并在 1974 年首次得以应用于控制领域. Wang^[18]等在 1992 年创造性地提出可将模糊系统作为通用逼近器,此后许多学者相继从数学和实际应用等不同的角度完善和发展了模糊系统理论,而且普遍认为现实世界太复杂以至于无法做到精确描述,所以为了得到一个合理的且可跟踪的模型就必须引入近似性(即模糊性)的概念.随着向信息时代的迈进,人类知识变得日益重要,人们需要一种理论能够系统地描述人类知识并将其同其他信息(如数学模型和感官测量)一起嵌入到工程系统中.模糊系统是一种基于知识或规则的系统,其核心是由所谓的 IF-THEN 规则组成的知识库.一个模糊的 IF-THEN 规则就是一个用连续隶属度函数对所描述的某些句子所做的 IF-THEN 规则的陈述.最常用的模糊规则推理系统是如下的一系列的语言规则:

$$R_j: \text{如果 } x_1 \text{ 为 } A_1^j, \dots, \text{ 且 } x_m \text{ 为 } A_m^j, \text{ 则 } y \text{ 为 } B^j, \\ j = 1, 2, \dots, N, \quad (1)$$

其中 $x = (x_1, x_2, \dots, x_m)^T \in U$ 和 $y \in V$ 分别是模糊系统的输入和输出(语言)变量, A_i^j 和 B^j 分别是

$U_i \subset R$ 和 $V \subset R$ 上的模糊集合.如果规则(1)中的模糊集 B^j 是标准模糊集,则采用模糊规则(1)乘积推理机和单值模糊器及中心平均解模糊器的模糊系统为

$$f(x) = \frac{\sum_{j=1}^N \bar{y}^j \left(\prod_{i=1}^m \mu_{A_i^j}(x_i) \right)}{\sum_{j=1}^N \left(\prod_{i=1}^m \mu_{A_i^j}(x_i) \right)}, \quad (2)$$

式中 $x \in U \subset R^m$ 是模糊系统的输入, $f(x) \in V \subset R$ 是模糊系统的输出, \bar{y}^j 为在输出空间上的采用高斯隶属度函数 μ_{B^j} 的所取得的最大值, $\mu_{A_i^j}$ 为隶属度函数,其最常用的形式为高斯函数.令模糊基函数 FBF(fuzzy basis function)为

$$p_j(x) = \frac{\prod_{i=1}^m \mu_{A_i^j}(x)}{\sum_{j=1}^N \prod_{i=1}^m \mu_{A_i^j}(x)}, \quad (3)$$

那么 $f(x)$ 可以表示为模糊基函数的线性组合,即(1)式的模糊推理系统等价于 FBF 的展开:

$$f(x) = \sum_{j=1}^N p_j(x) \bar{y}^j. \quad (4)$$

从(1)~(4)式可以看出模糊系统一方面是基于规则库的系统,是由一系列语言规则构造而成的;另一方面提供了一个把语言规则集合转变成非线性映射的系统化程序.但是如何选择一个合适的模糊规则数是模糊系统的一个比较困难的问题,为了充分发挥模糊系统能自动地从学习样本中获取知识的优点,结合支持向量机的基本思想,取模糊规则的数目等于 SVM 的支持向量数,这样不仅避免了模糊系统中模糊规则数的选择问题,而且继承了 SVM 通过控制学习机器的容量实现对推广能力的控制的优点.同时为了使模型能自动的从学习样本中提取最能代表混沌时间序列的发展趋势和内在本质的驱动力的数据,采用模糊曲线法^[19]对输入向量空间进行模糊划分.模糊曲线法是按照输入向量 X 的每一维来考核对其输出 Y 的影响,并在此基础上建立输入向量空间的模糊划分.对于具有 N 组 m 个输入单输出的学习样本,输入变量 x_i 的模糊隶属度函数为

$$\varphi_{ik}(x_i) = \exp \left\{ - \left[\frac{x_{ik} - x_i}{\sigma} \right]^2 \right\} \\ k = 1, \dots, N, i = 1, \dots, m, \quad (5)$$

其中 x_{ik} 是 X_k 的第 i 维分量, σ 为模糊隶属度函数的参数.这样每一个 φ_{ik} 和相应的 y_k 组成了一个模糊规则.

如果输入 x_i 属于 φ_{ik} , 则输出 y 为 y_k , (6)

当然 φ_{ik} 也可以为别的模糊隶属度函数, 包括梯形, 三角形等. 采用中心去模糊器、高斯模糊隶属度函数 (5) 和模糊规则 (6), 得到了每一个输入 x_i 的模糊曲线

$$C_i(x_i) = \frac{\sum_{k=1}^N \varphi_{ik}(x_i) \cdot y_k}{\sum_{k=1}^N \varphi_{ik}(x_i)} \quad i = 1, \dots, n, \quad (7)$$

从模糊曲线 C_i 中, 能够发现输入对输出的影响在不断地变化, 并能发现那些样本对于输出的影响比较大. 这样可以根据模糊曲线中局部最小点和最大点的数目来确定模糊规则数(支持向量数). 如果局部最大点和最小点相隔比较远, 或者这一段模糊曲线变化比较大, 则需要增加模糊规则. 这样可以根据每一个样本对于其对应的输出的影响, 从学习样本中选择对输出影响比较大的数据作为最终的学习样本, 这样这个最终的学习样本所对应的训练后模型的支持向量对于系统输出的影响相应的也比较大.

在模糊系统中, 模糊曲线函数中 σ 的取值是一个非常重要的参数, 比较难选择. 为了使模糊曲线能充分真实的反映出输入向量的每一维对其输出的影响, 结合 SVM 的机理, 取其与 SVM 核函数的超参数相等, 这样从根本上解决了 σ 的选择困难的问题. 模糊系统的这些优良特性为 SVM 从大量的混沌时间序列的实际数据中提取最能代表其内在驱动力的学习样本作为模型的支持向量, 大大减少有效支持向量的数目提供了可能.

混沌是非线性系统产生的一种类似随机且无规则运动的不确定性动力学行为, 由许多有序行为组合而成, 并且在正常情况下任何一个单独的行为都不在系统中占主导地位, 而且混沌条件下的系统对初始状态有敏感的“蝴蝶效应”, 因而一般认为基于混沌系统产生的时间序列具有不可长期预测性. 但是混沌理论的研究表明: 尽管从表面上看, 混沌是随机的、不可预测的, 事实上却是按照严格的、确定性的规则运动着, 即运动应该在一定的时间内还是可以预测的. 混沌时间序列预测的基础是 Takens 定理^[21], 其主要思想是系统任一分量的演化是由与之相互作用着的其他分量所决定的, 而且这些相关分量的信息就隐含在任一分量的发展过程中. 因此可以从某一分量的一批时间序列数据中提取和恢复出系统原来的规律, 这种规律是高维空间下的一种轨迹(混沌吸引子). 虽然这种轨迹在经过类似拉伸和

折叠后转化成与时间有关的序列时呈现出混乱的、复杂的特征, 但是由于混沌系统的策动因素是相互影响的, 因而在时间上先后产生的数据点也是相关的. 目前对序列动力学因素的分析广泛采用的是 Packard 等提出的延迟坐标状态空间重构法^[22]. Takens 同时也证明了先考虑一个分量, 然后将它在某些固定时间上的延迟点作为新维处理, 如果能够找到一个合适的嵌入维数, 即延迟坐标的维数 $m \geq 2d + 1$, d 是动力系统的关联维数, 在这个嵌入维空间里就可以把规律(吸引子)恢复出来, 亦即在重构的 R^m 空间中的轨迹与原动力系统保持微分同胚, 然后可以用这个规律来预测轨迹的走向, 这个由嵌入维决定的空间就是相空间. 而支持向量机是通过某种事先选择的定义在内积空间上的核函数, 自动的将输入空间的数据映射到一个高维(可能无限维)特征空间中

$$x_i \rightarrow \phi(x_i) = (\phi_1(x_i), \phi_2(x_i), \dots, \phi_n(x_i), \dots), \quad (8)$$

并在这个特征空间中自动的构造一个能够将所有训练样本正确分类, 而且使训练样本中离分类面最近的点到分类面的距离(定义为间隔)最大的最优分类超平面, 这样数据蕴涵的信息就自然的呈现在高维特征空间中, 因而能够容易的发现数据中内在的规律. 由上不难发现支持向量机的主要思想与混沌时间序列的相空间的重构理论有一些相似之处^[15]: 都把输入空间的数据通过某种方式扩展到高维空间中, 进而在高维空间中提取数据中蕴涵的信息和规律. 只不过相空间重构理论是将数据扩展到在事先选择的有限的嵌入维数的空间中提取其数据蕴涵的信息. 而支持向量机是通过非线性映射将输入向量自动的映射到高维(甚至无限维)空间中提取数据的信息. 而 Takens 定理仅仅说明了在已知嵌入维空间里可以把有规律的轨迹恢复出来, 这需要时间序列的先验知识. 对于实际中未知的复杂的时间序列, Takens 定理同样不再适用. 因此如果混沌时间序列能够根据相空间重构理论通过模糊逻辑选择代表时间序列内在规律的数据作为支持向量, 然后利用支持向量机的核函数将数据映射到高维特征空间中自动的构造超平面的能力, 实现数据的线性可分, 进而提取其蕴涵的信息, 则其混沌时间序列的模型能力将进一步提高.

在重构相空间中, 时间延迟 τ 和嵌入维数 m 的选择十分重要, 其选择直接决定着相空间重构后描

述奇异吸引子特征的不变量的准确度.关于 τ 和 m 的选取主要有两种观点,第一种认为 τ 和 m 是互不相关的,其选取可以独立进行(Takens 证明了对于无限长的、无噪声干扰的时间序列,其 τ 和 m 是相互独立的).常用的关于时间延迟和嵌入维数的选择有 3 种方法:1)序列相关法,如自相关法^[23]、互信息量法^[24]和高阶相关法^[25]等;2)相空间扩展法,如充填因子法^[26]、摆动量法^[27]、平均位移法^[28]、SVF 法^[29]等;3)复自相关法^[30,31].第二种观点则认为 τ 和 m 是相互关联的,因为现实中的时间序列都是有限长且不可避免地受到各种噪声的影响.大量实验表明, τ 和 m 的关系与重构相空间的时间窗 τ_w 密切相关($\tau_w = (m-1)\tau$)对于特定的时间序列,其 τ_w 相对固定, τ 和 m 的不恰当配对将直接影响重构后的相空间结构与原空间的等价关系,因此相应地产生了 τ 和 m 的联合算法,如时间窗口法^[32]、C-C 法^[33]等.人们普遍认为,第二种观点在实际应用中更为合理和有效.而基于统计学习理论的 SVM 由于其将数据映射到高维特征空间中提取数据的蕴涵其内在规律的信息,而使得 τ 和 m 的选取在大多数情况下,对于模型的预测性能影响不是很大,下面的仿真实验也证明了这一点.

利用模糊模型的 SVM 重构相空间的状态分量时,输入变量的个数应至少大于时间序列的嵌入维数与延迟时间的乘积.这样基于嵌入维 m , N 个混沌时间序列的模糊模型的 SVM 预测模型的输入矢量可表示为

$$X(t-1) = [x(t-\tau), x(t-2\tau), \dots, x(t-m\tau)] \\ t = 1, 2, \dots, N, \quad (9)$$

其中 $x(\cdot)$ 是混沌时间序列, τ 为延迟时间,通常 τ 取不小于 1 的整数, m 是嵌入维数.而利用模糊模型选择能代表混沌时间序列的发展趋势的数据作为 SVM 的学习样本的输入向量为

$$X_{\text{new}}(t-1) = [x_{\text{new}}(t-\tau), x_{\text{new}}(t-2\tau), \dots, \\ x_{\text{new}}(t-m\tau)] \\ t = 1, 2, \dots, N_{\text{new}}, \quad (10)$$

其中 N_{new} 是通过模糊曲线获得的最终的学习样本数,远远小于原来的学习样本数 N .利用 N_{new} 个学习样本训练 SVM 仅仅获得 N_{new} 个支持向量就能代表混沌时间序列发展趋势的内在关系,并能对未来的时间序列进行预报.因此基于 N 个混沌时间序列得到的 N_{new} 个支持向量的模糊模型的 SVM 的混沌时间序列的预测值为

$$\hat{y} = f(X(t-1)), \quad (11)$$

而构建的适于混沌时间序列的预测模型如图 1 所示.从图 1 可以看出,基于统计学习理论的支持向量学习机为模糊模型从有关时间序列的学习样本中提取代表其内在规律的支持向量,产生 IF-THEN 规则奠定了牢固的理论基础,因此基于模糊模型的 SVM 不仅具有支持向量机的优秀的泛化能力,而且克服了传统的神经网络(BP 神经网络)的“过拟合”问题,同时继承了模糊系统可以使人们易于将先验的系统知识结合到模糊规则中,较好的实现了 SVM 通过非线性变换将输入空间变换到一个高维特征空间,并在这个新空间中求取最优线性分类面,并提取输入向量蕴藏的系统信息的基本思想.若通过模糊模型选择后得到数据样本集 $\{x_k, y_k\}_{k=1}^{N_{\text{new}}}$, $x_k \in R^m$, $y_k \in R$,所要求的基于 SVM 的拟合函数形式为

$$y(x) = \omega \cdot \phi(x) + b, \quad (12)$$

其中 ω 为权向量, b 为偏差, $\phi(\cdot)$ 为把输入空间映射到一个高维特征空间中的非线性映射函数.由于 $\phi(\cdot)$ 可以用核函数来代替,而核函数可以用原空间中的函数来实现,无需知道 $\phi(\cdot)$ 的具体形式,因而根据统计学习理论^[14],基于支持向量机的混沌时间序列的拟合函数可表示为

$$y(x) = \sum_{k=1}^{N_{\text{new}}} \alpha_k k(x_k, x) + b, \quad (13)$$

其中 α_k 为支持向量; b 为偏差, $k(x_k, x_i)$ 为核函数,常用的核函数为高斯核函数:

$$k(x, x') = \exp(-\|x - x'\|^2 / \sigma^2), \quad (14)$$

其中 σ 为核函数的参数,可以通过 Cross-Validate 法选择一个合适的值.

利用模糊系统选择典型的能够代表混沌时间序列内在规律的数据,然后 SVM 根据构建的混沌时间序列的输入向量和输出向量进行学习,获得混沌时间序列模型的参数 α 和 b .预测模型的参数 N_{new} 个支持向量和偏差所蕴藏的关系就是混沌时间序列的各个向量过去的和将来的关系.至此基于模糊模型的支持向量机的预测模型已经建立.这样就可以利用该模型预测将来的混沌时间序列.

3. 学习算法

最小二乘法是一个比较古老的方法,由于在随机的情况下并不要求观测数据提供其概率统计方面的信息,而其估计结果却有很好的统计特性,因而常

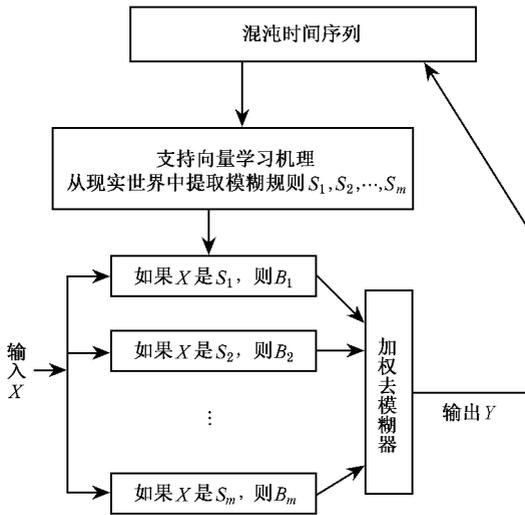


图1 混沌时间序列的基于模糊模型的支持向量机的推理模型

常用来拟合平面上的数值点. 同时由于它是一切平方和形式的目标函数的最优解的一个基本方法而广泛应用于科学实验和工程技术中的曲线拟合、函数逼近和数据处理中. 近年来最小二乘法在求解超定方程和处理最优化问题已有很大的发展, 新的算法不断产生. 如果采用最小二乘法拟合的数据中没有噪声, 则其解是一致的, 否则其解是不一致的, 但是最小二乘法在解不一致的情况下使获得的解的误差平方和最小. 基于最小二乘法的这些优点, 将支持向量机非线性映射到高维特征空间中的数据拟合成超平面并进行分类, 利用最小二乘法求取其最优解, 提取数据蕴涵的信息, 因此最小二乘支持向量机^[34, 35]是基于正则化理论对标准 SVM 的改进, 这个改进使 SVM 的解大大简化. 拟合函数中最优的权向量 ω , 和偏差 b 可以通过最小化由拟合误差的平方和及正则化项组成的目标函数得到

$$J(\omega, b) = \frac{1}{2} \|\omega\|^2 + \gamma \sum_{k=1}^{N_{\text{new}}} (y_k - \omega \cdot \varphi(x_k) - b)^2, \quad (15)$$

其中, 正实数 γ 是调节常数, 它能够使训练误差和模型复杂度之间取一个折衷, 以便使所求得的函数具有较好的泛化能力, 而且 γ 值越大, 模型的预测误差越小. 由于 ω 可能是无限维的, 因而通过最小化式(15)是不可能的, 但在对偶空间上求取 ω . 建立 Lagrangian 方程

$$\mathcal{L}(\omega, b, e; \alpha) = \min J(\omega, b, e) - \sum_{k=1}^{N_{\text{new}}} \alpha_k \{ \omega^T \varphi(x_k) + b + e_k - y_k \}, \quad (16)$$

其中 α_k 为 Lagrangian 因子(支持值). 最优的 α 和 b 可以通过 KKT(Karush-Kuhn-Tucker)^[34]条件获得

$$\begin{bmatrix} 0 & 1_V^T \\ 1_V & Z \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}, \quad (17)$$

其中 $Z = (K + \gamma^{-1}I)$, $K = \{k_{ki} = k(x_k, x_i)\}_{k,i=1}^{N_{\text{new}}}$, $1_V = [1; \dots; 1]$, $y = [y_1; \dots; y_{N_{\text{new}}}]$ 和 $\alpha = [\alpha_1; \dots; \alpha_{N_{\text{new}}}]$, $k(x_k, x_i)$ 为核函数. 从(17)式可以看出, 尽管采用模糊模型选取代表时间序列发展趋势的数据作为学习样本, 需要的计算机存储空间仍然比较大, 为了减少计算机的存储空间, 使算法适合于学习样本比较多的问题, (17)式可以变换为

$$\begin{bmatrix} 1_V^T Z^{-1} 1_V & 0 \\ 0 & Z \end{bmatrix} \begin{bmatrix} b \\ \alpha + bZ^{-1} \end{bmatrix} = \begin{bmatrix} 1_V^T Z^{-1} y \\ y \end{bmatrix}, \quad (18)$$

这样求取 α 和 b 的大规模问题的算法过程为

$$1) Z\eta = 1_V, \quad (19)$$

$$2) Zv = y, \quad (20)$$

$$3) s = 1_V^T Z^{-1} 1_V, \quad (21)$$

$$4) b = \eta^T y / s, \quad (22)$$

$$5) \alpha = v - \eta b. \quad (23)$$

这个算法的核心是利用 Z , y 和 1_V 解两个线性方程组, 因而可以充分利用 Z 是光滑的核矩阵 K 和比例单位矩阵 $\gamma^{-1}I$ 和的这一特性, 采用 N-S 形式的直接求解方程组^[36], 大大减小计算机的存储空间, 使模型可以应用于学习样本数目大的情况.

根据上述的分析, 对于建立的基于模糊模型的 SVM 预测模型, 采用最小二乘法训练的具体学习算法为

1) 初始化: 选择一个合适的嵌入维数 m 、选择核函数的超参数 σ 和调节常数 γ ;

2) 时间序列的生成: 根据选择的嵌入维数 m 生成模型所需的输入向量 X ($N \times m$ 维) 和输出向量 Y ($N \times 1$ 维);

3) 利用模糊曲线法选择能代表混沌时间序列发展趋势的支持向量的数据的输入向量 X_{new} ($N_{\text{new}} \times m$ 维) 和输出向量 Y_{new} ($N_{\text{new}} \times 1$ 维) 作为 SVM 的学习样本;

4) 模型求解: 利用大规模的问题的求解算法得到模型的参数 α 和 b .

至此基于模糊模型的支持向量机的预测模型已经建立, 这 N_{new} 个支持向量 α 和 b 所蕴藏的关系就是混沌时间序列的各个向量过去的和将来的关系, 利用获得的 α 和 b 就可以来预测将来的混沌时间

序列.

4. 预测实例

实践研究表明,将适用于大规模问题求解的最小二乘法与基于模糊模型的 SVM 理论结合起来的新预测方法在复杂非线性系统的建模中具有独特的优越性.由于混沌时间序列对初始状态极其敏感,给混沌时间序列的预测带来很大的困难.但实验表明,利用基于模糊模型的 SVM 在一定范围内,混沌时间序列的预测不仅是可行的,而且是精确的.本文将 Mackey-Glass 的混沌时间序列预测和 Lorenz 系统产生的混沌时间序列预测作为仿真实例,同时为衡量预测模型的精确性,采用相关系数

$$r = \left(\frac{1}{N_{\text{pred}}} \sum_{n=1}^{N_{\text{pred}}} (x(n, \text{true}) - \overline{x(n, \text{true})}) \right) \times (x(n, \text{pred.}) - \overline{x(n, \text{pred.})}) \times \sqrt{\left(\frac{1}{N_{\text{pred}}} \sum_{t=1}^{N_{\text{pred}}} (x(t, \text{true}) - \overline{x(t, \text{true})})^2 \right) \times \left(\frac{1}{N_{\text{pred}}} \sum_{t=1}^{N_{\text{pred}}} (x(t, \text{pred.}) - \overline{x(t, \text{pred.})})^2 \right)^{-1}} \quad (24)$$

作为评价模型整体预测效果的指标.其中 $x(n, \text{true})$ 和 $x(n, \text{pred.})$ 分别为第 n 个混沌时间序列的实际值和预测值, $\overline{x(n, \text{true})}$ 和 $\overline{x(n, \text{pred.})}$ 分别为混沌时间序列的实际值和预测值的平均值, N_{pred} 为预测的时间序列的数量.最后利用该模型讨论了模型的相关系数与嵌入维数及延迟时间的关系.

4.1. Mackey-Glass 的混沌时间序列预测

自 1977 年 Mackey 和 Glass 发现时滞系统中的混沌现象以来^[37],时滞混沌系统便引起了人们的广泛关注,并常常用其作为检验非线性系统模型性能的标准,其混沌时间序列由下面的时滞微分方程产生:

$$\frac{dx(n)}{dn} = \frac{0.2x(n-\Delta)}{1+x^{10}(n-\Delta)} - 0.1x(n), \quad (25)$$

其中 Δ 为时滞参数, Mackey-Glass 方程的行为特性已做过深入研究^[38]. $\Delta > 17$ 时呈现混沌性,其分形维数近似为 2.1,并且 Δ 值越大,混沌程度越高.图 2 为 $\Delta = 30$ 的混沌时间序列,从图 2 不难看出其时间序列具有极其复杂的非线性混沌特征,基于数学模型的方法直接建立其预测模型是很困难的.但是采用基于模糊模型的 SVM 时,利用 $N = 500$ 个数据点作为训练集,取嵌入维数 m 和延迟时间 τ 分别为

4 和 1,在 $\{\gamma, \sigma\} = \{100, 2\}$ 时进行训练建模,首先采用模糊曲线法选择模糊规则的数目(支持向量的数目),总共仅仅提取了 129 个典型的数据作为支持向量机的学习样本,比实际的学习样本(500 个)少了 371 个,大大减少了计算量.利用这 129 个学习样本训练 SVM 得到模型的 129 个支持向量和偏差 b 所蕴含的关系就是 Mackey-Glass 的各个变量的时间序列过去和未来的关系.为了验证模型的准确性,利用得到的 129 个支持向量对将来的 500 个数据点进行验证,图 3 表示了所建模型的预测值与系统的实际输出的比较曲线,其 MSE 误差为 0.0194,其相关系数 r 为 0.9985,说明该模型的预测值与系统的实际输出符合得很好.

通常模型的预测效果与延迟时间和嵌入维数有关,例如延迟时间的选择决定着重构相空间所包含的信息量.如果 τ 太小,输入向量的相邻延迟坐标元素间差别太小,即冗余度较大,重构相空间的样点所包含的关于原吸引子的信息偏小,表现在相空间形态上,即信号轨迹向相空间主方向轴压缩,相反若 τ 太大,相空间内延迟坐标元素间的相互信息丢失,即各元素不相关,信号轨迹可能出现折叠现象.此外,对于不同的嵌入维数,最佳的延迟时间也不一样.因此在建立混沌时间序列的模型时,选择合适的延迟时间是必需的.这里以 Mackey-Glass 时间序列为例,讨论嵌入维数 m 和延迟时间 τ 对模型预测未来混沌时间序列的影响.固定 $\{\gamma, \sigma\} = \{100, 2\}$,分别改变嵌入维数 m 从 1 到 18 和延迟时间 τ 从 4 到 100,模型的建模能力如图 4 所示,发现模型的建模能力并没有随着 τ 的增大而减少,说明 SVM 通过核函数把输入空间的向量映射到高维特征空间中,能够将其蕴涵的信息在这个特征空间中充分体现出来进行分类和提取而使得 τ 的选取对于模型的影响很小,模型的建模能力除 m 从 1 到 5,预测效果逐渐变好外,整体也没有随着嵌入维数 m 的增加而改善很多,甚至在高维时出现了一些预测效果比较差的情况.这可能是 SVM 刚开始(m 从 1 到 5)随着嵌入维数的增加,在相空间中通过核函数映射到高维空间中的信息增加而使得提取的信息增加导致了预测效果变好,而当嵌入维数越来越大,高维空间中信息越来越多,最小二乘法不能很好的对其分类导致了预测效果变差.总之外除极少数点外,绝大多数情况都大于 0.9800,但即使在最坏的情况下预测的有关未来的 500 个混沌时间序列的相关系数也都在 0.9000

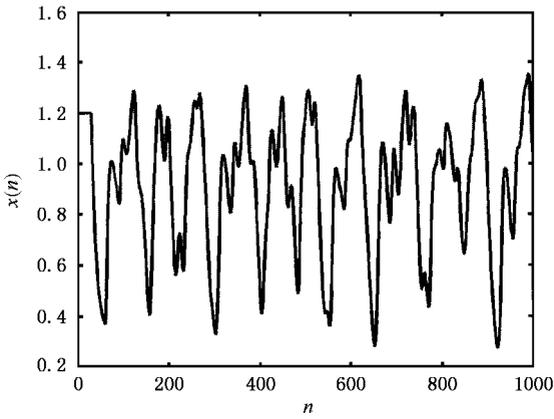


图2 $\tau = 30$ Mackey-Glass 混沌时间序列

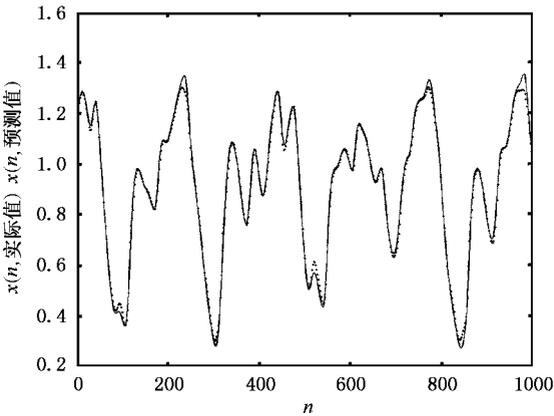


图3 Mackey-Glass 实际值(实线)和预测值(点线)比较曲线

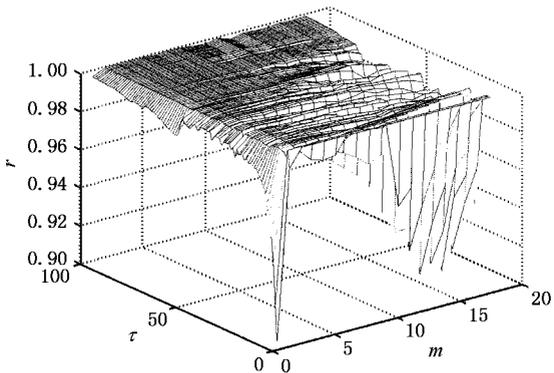


图4 Mackey-Glass 嵌入维数 m 和时间延迟 τ 与相关系数 r 的关系曲线

以上,说明基于模糊模型的 SVM 预测方法的延迟时间和嵌入维数选择范围比较大.因此基于模糊模型的 SVM 预测方法在实际混沌系统的研究中即使不能选取最佳的时间延迟 τ 值和嵌入维数 m 也能取

得好的预测效果.

4.2. Lorenz 混沌时间序列预测

自从 1963 年 Lorenz^[39]发现了 Lorenz 系统以来,至今人们已经对这个系统进行了大量而深入的研究,并对 Lorenz 系统的动力学行为的本质的有了进一步认识^[40-41].但是 Lorenz 系统作为一个最经典的三维二次多项式自治系统的混沌模型,对 Lorenz 系统的研究纵贯整个混沌科学的发展,几乎与所有混沌科学的重要发展都密切相关,因而研究 Lorenz 系统具有十分基本而重要的意义.下面以其产生的时间序列为例来验证模型的精确性.

$$\begin{aligned} \dot{x} &= \sigma(y - x), \\ \dot{y} &= \rho x - y - xz, \\ \dot{z} &= -\beta z + xy, \end{aligned} \tag{26}$$

其中 $\sigma = 10, \rho = 28, \beta = 8/3$. 假定初始条件 $x_0 = 5, y_0 = 5, z_0 = 15$, 利用四阶五级的 Runge-Kutta 算法求解 Lorenz 系统的数值解,采样间隔时间 $\Delta T = 0.05$ 产生关于 x 的混沌时间序列如图 5 所示.从图中不难看出其时间序列具有复杂的非线性特征,直接建立其预测模型存在很大的困难.对这样一个混沌时间序列,基于模糊模型的 SVM 是否能够基于过去的系统特性来预测其将来的轨迹?由利用前 $N = 600$ 个数据点作为训练集,取嵌入维数 m 和延迟时间 τ 分别为 6 和 1,在 $\{\gamma, \sigma\} = \{10, 5\}$ 时进行训练建立模型.首先利用模糊曲线法总共选择了 380 个模糊规则(支持向量),比实际的学习样本(600 个)少了近一半,这样减少了模型预测未来时间序列的计算量.利用 SVM 训练后得到的 380 个支持向量和偏差 b 所蕴涵的关系就是 Lorenz 的时间序列过去和未来的关系.为了验证模型的准确性,利用 380 个支持向

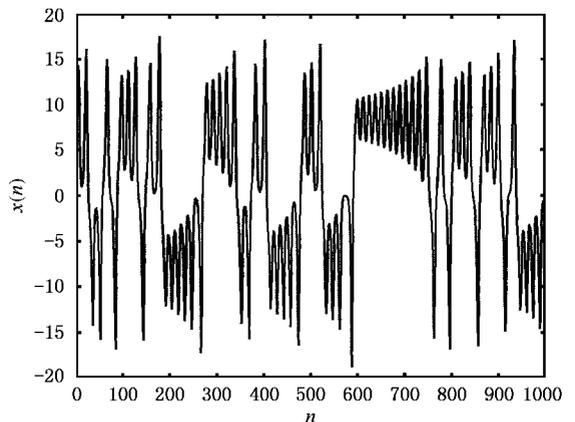


图5 Lorenz 混沌时间序列

量对将来的400个数据点进行了验证所建模型的预测值与系统的实际输出的比较曲线如图6所示,其相关系数 r 为 0.9991,从图 6 不难看出该模型的预测值与系统的实际值输出符合得比较好.

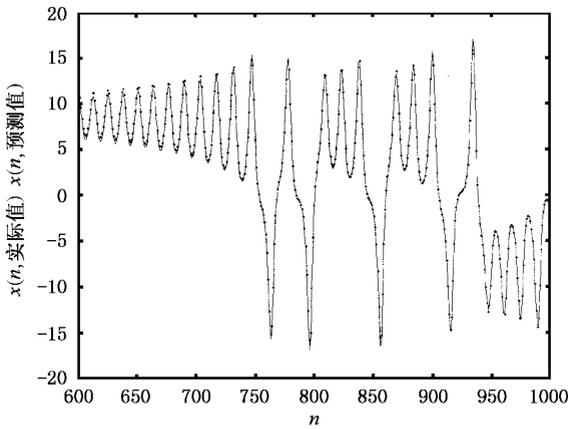


图 6 嵌入维数为 6 时,Lorenz 实际值(实线)和模型的预测值(虚线)比较曲线

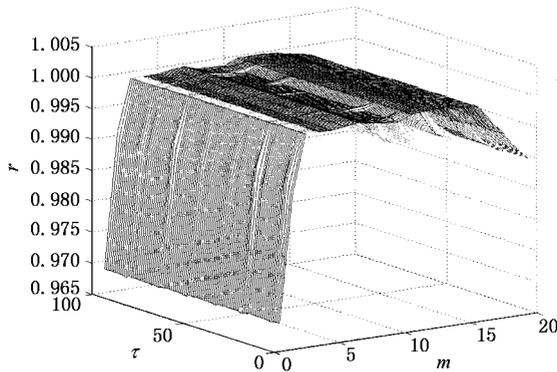


图 7 Lorenz 模型的嵌入维数 m 与相关系数 r 关系曲线

采用相空间重构法研究混沌时间序列时,要准确、全面的描述混沌时间序列中奇异吸引子的几何特性,必须准确地选择用于重构计算的嵌入维数 m 和延迟时间 τ ,而嵌入维数 m 和延迟时间 τ 又是密切相关的,因此选择最佳的嵌入维数 m 和延迟时间 τ 变得有些困难.以 Lorenz 时间序列为例,嵌入维数 m 和延迟时间 τ 对模型预测的影响到底有多大?固定 $\{\gamma, \sigma\} = \{10, 5\}$,分别改变嵌入维数 m 从 1 到 20 和延迟时间 τ 从 4 到 100,模型的建模能力如图 7 所示,发现模型的建模能力并没有随着 τ 的增大而减小,这再一次证明了 SVM 通过核函数把输入空间的向量映射到高维(可能无限维)特征空间中,能够在这个特征空间中提取其蕴含的系统信息而使得 τ

的选取对于模型的影响很小;而模型的建模能力除 m 从 1 到 3,预测效果逐渐变好外,整体随着嵌入维数 m 的增加有一些波动,最后随着嵌入维数 m 的增加反而预测效果降低,这可能是 SVM 刚开始(m 从 1 到 3)随着嵌入维数的增加,在相空间中通过核函数映射到高维特征空间中的信息量增加导致了预测效果变好,而当嵌入维数越来越大,映射到高维空间中信息太多以至于由于最小二乘法的局限性不能很好的对其分类导致了预测效果变差.分别以 m 为 2 和 20 为例,取 τ 为 5,所建模型的预测值与系统实际输出的比较曲线如图 8 和图 9 所示,其相关系

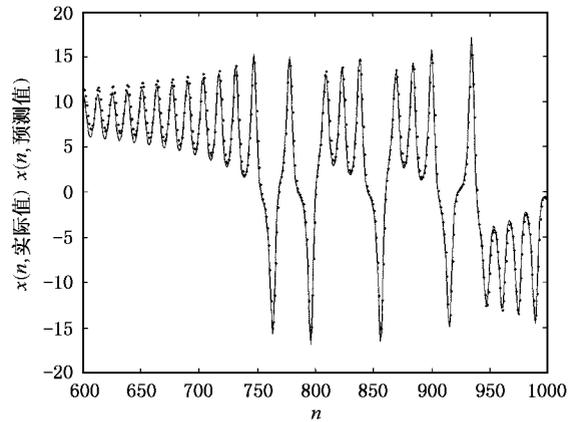


图 8 嵌入维数为 2 时,Lorenz 实际值(实线)和模型的预测值(虚线)比较曲线

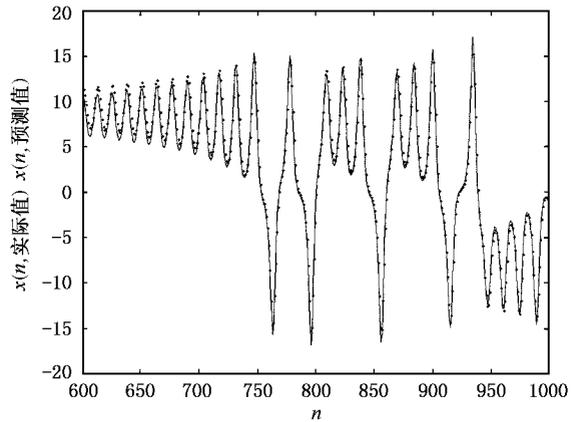


图 9 嵌入维数为 20 时,Lorenz 实际值(实线)和模型的预测值(虚线)比较曲线

数分别为 0.9897 和 0.9891.通过上面的分析发现对 Lorenz 系统产生的时间序列的预测,除嵌入维数小于 3 外,其余的相关系数都大于 0.9850,但即使最差

的情况其预测的未来的 400 个混沌时间序列的相关系数也都在 0.9700 以上,说明基于模糊模型的 SVM 预测方法的延迟时间和嵌入维数选择范围比较宽.因而基于模糊模型的 SVM 预测方法在实际混沌系统的研究中即使不能找到最理想的时间延迟 τ 值和嵌入维数 m 也能取得不错的预测效果.

5. 结 论

将基于统计学习理论的支持向量机和模糊逻辑应用到混沌时间序列预测的研究中,根据混沌时间序列固有的确定性和非线性,SVM 能够自动的把输入向量映射到一个高维特征空间中实现数据线性可分来提取信息的能力和模糊逻辑使人们易于将先验的系统知识结合到模糊规则中,利用混沌动力系统

的相空间重构理论,建立了混沌时间序列的基于模糊模型的支持向量机预测模型,并利用适用于大规模问题求解的最小二乘方法对建立的模型进行训练,该模型所蕴涵的函数映射关系就是混沌时间序列数据的内在的函数关系.最后以 Mackey-Glass 时间序列和典型的 Lorenz 系统产生的混沌时间序列为例,对模糊模型的 SVM 的建模能力进行了验证,并利用该模型对嵌入维数和模型整体预测效果及延迟时间关系进行了讨论.结果表明,混沌时间序列的基于模糊模型的支持向量机预测方法不仅可以大大减少计算量,而且在嵌入维数未知和延迟时间不能合理选择的情况下也能精确的预测未来的时间序列,因此基于模糊模型的支持向量机在混沌时间序列的研究中具有很强的实用价值.

- [1] Gershenfeld N *et al* 1999 *Nature* **397** 392
- [2] Farmer J D *et al* 1987 *Phys. Rev. Lett.* **59** 845
- [3] Gan J C *et al* 2003 *Acta Phys. Sin.* **52** 2995 (in Chinese) [甘建超等 2003 物理学报 **52** 2995]
- [4] Zhang J S *et al* 2000 *Acta Phys. Sin.* **49** 403 (in Chinese) [张家树等 2000 物理学报 **49** 403]
- [5] Zhang J S *et al* 2001 *Acta Phys. Sin.* **50** 1248 (in Chinese) [张家树等 2001 物理学报 **50** 1248]
- [6] Cao L *et al* 1995 *Physica D* **85** 225
- [7] Zhang J S *et al* 2000 *Chin. Phys. Lett.* **17** 88.
- [8] Casdal M 1989 *Physica D* **35** 335
- [9] Tan W *et al* 2003 *Acta Phys. Sin.* **52** 795 (in Chinese) [谭文等 2003 物理学报 **52** 795]
- [10] Gu W *et al* 1995 *Journal of Fudan University (Natural Science)* **34** 262 (in Chinese) [顾炜等 1995 复旦学报(自然科学版) **34** 262]
- [11] Gao L Y 1995 *Physica D* **85** 225
- [12] Cui W Z *et al* 2004 *Acta Phys. Sin.* **53** 1583 (in Chinese) [崔万照等 2004 物理学报 **53** 1583]
- [13] Cui W Z *et al* 2005 *Thin Solid Films* **473** 224
- [14] Vapnik V N 1995 *The Nature of Statistical Learning Theory* (New York : Springer-Verlag)
- [15] Cui W Z *et al* 2004 *Acta Phys. Sin.* **53** 3303 (in Chinese) [崔万照等 2004 物理学报 **53** 3303]
- [16] Cui W Z *et al* 2005 *Chin. Phys.* **14** 922
- [17] Wang L X 1992 *IEEE Trans. on Systems, Man and Cybern.* **22** 1412
- [18] Wang L X *et al* 1992 *IEEE Trans. on Neural Networks* **3** 807
- [19] Lin Y *et al* 1995 *IEEE Trans. on Fuzzy Systems* **5** 190
- [20] Chiang J H *et al* 2004 *IEEE Trans. on Fuzzy Systems* **12** 1
- [21] Takens F 1981 *Determining strange attractors in turbulence, Lecture notes in Mathematics* (Berlin : Springer-Verlag) 366
- [22] Packard N H *et al* 1980 *Phys. Rev. Lett.* **45** 712
- [23] Albano A M *et al* 1988 *Phys Rev A* **38** 3017
- [24] Fraser A M 1989 *IEEE Trans. on Information Theory* **35** 245
- [25] Albano A M 1991 *Physica D* **54** 85
- [26] Buzug PT 1992 *Phys. Rev. A* **45** 7073
- [27] Buzug PT 1992 *Physica D* **58** 127
- [28] Rossenstein MT *et al* 1994 *Physica D* **73** 82
- [29] Kember G *et al* 1993 *Phys. Rev. Lett.* **179** 72
- [30] Lin J Y *et al* 1999 *Signal Processing* **15** 220 (in Chinese) [林嘉宇等 1999 信号处理 **15** 220]
- [31] Ma H G *et al* 2004 *Journal of Xi 'an Jiaotong University* **38** 562 (in Chinese) [马红光等 2004 西安交通大学学报 **38** 335]
- [32] Kugiumtzis D *et al* 1996 *Physica D* **95** 13
- [33] Kim HS *et al* 1999 *Physica D* **127** 48
- [34] Suykens J A K *et al* 2002 *Least Squares Support Vector Machines* (Singapore : World Scientific)
- [35] Cui W Z *et al* 2004 *Journal of Xi 'an Jiaotong University* **38** 562 (in Chinese) [崔万照等 2004 西安交通大学学报 **38** 562]
- [36] Gines D L *et al* 1998 *Applied and Computational Harmonic Analysis* **5** 156
- [37] Mackey M C *et al* 1977 *Science* **197** 287
- [38] Farmer J D 1982 *Physica D* **3** 336
- [39] Lorenz E N 1963 *J. Atmos. Sci.* **20** 130
- [40] Mischalkow K *et al* 1998 *Mathematics of Computation* **67** 1023
- [41] Stewart I 2000 *Nature* **406** 948

Prediction of the chaotic time series using support vector machines for fuzzy rule-based modeling^{*}

Cui Wan-Zhao Zhu Chang-Chun Bao Wen-Xing Liu Jun-Hua

(*School of Electronics and Information Engineering , Xi 'an Jiaotong University , Xi 'an 710049 , China*)

(Received 3 September 2004 ; revised manuscript received 10 November 2004)

Abstract

Based on the powerful nonlinear mapping ability of support vector machines and the characteristics of fuzzy logic which can combine a prior knowledge into fuzzy rules , the forecasting model of the support vector machine for fuzzy rules-based model in combination with Takens ' delay coordinate phase reconstruction of chaotic time series has been established ; and the least squares method for large-scale problems is used to train this model . Moreover , based on this model , relationships among the prediction performances of this model , the embedding dimension and the delay time are discussed . Finally , the Mackey-Glass equation and the time series that Lorenz systems generate are applied to test this model , respectively , and the results show that the support vector machine for fuzzy rule-based modeling can not only acquire knowledge and generate fuzzy rules from the given data , reduce the number of support vectors greatly , but also predict chaotic time series accurately , and even if the embedding dimension is unknown and the delay time is appropriately selected , the predicted results are satisfactory . These results imply the support vector machine for fuzzy rule-based modeling is a good tool to study chaotic time series in practice .

Keywords : fuzzy logic , chaotic time series , support vector machine , least squares method

PACC : 0545

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 60176020 and 60276037) and the Doctoral Foundation of the National Education Administration of China (Grant No. 20020698014).