

文章编号: 1001-0920(2006)05-0585-04

## 稀疏贝叶斯及其在时间序列预测中的应用

张旭东<sup>1,2</sup>, 陈 锋<sup>1</sup>, 高 隽<sup>2</sup>, 方廷健<sup>1</sup>

(1. 中国科技大学 自动化系, 合肥 230027; 2. 合肥工业大学 计算机与信息学院, 合肥 230009)

**摘 要:** 阐述了稀疏贝叶斯方法在时间序列预测中应用的理论基础, 将稀疏贝叶斯方法应用于 Logistic 方程产生的混沌时间序列和发动机油滑数据的预测, 并与支持向量机(SVM)和RBF神经网络时间序列预测进行了比较. 实验结果表明, 稀疏贝叶斯方法不仅具有 SVM 的性能, 而且比 SVM 使用更少的核函数, 取得了较好的预测效果.

**关键词:** 稀疏贝叶斯; 支持向量机; 非线性预测; RBF 神经网络

中图分类号: TP391

文献标识码: A

## Sparse Bayesian and Its Application to Time Series Forecasting

ZHANG Xu-dong<sup>1,2</sup>, CHEN Feng<sup>1</sup>, GAO Jun<sup>2</sup>, FANG Ting-jian<sup>1</sup>

(1. Department of Automation, University of Science and Technology of China, Hefei 230027, China; 2. Department of Computer and Information, Hefei University of Technology, Hefei 230009, China. Correspondent: ZHANG Xu-dong, E-mail: zxd163@mail.hf.ah.cn)

**Abstract:** The basic theoretic analysis of sparse Bayesian method in time series forecasting is introduced. Chaotic time series produced by Logistic equation and some type of engine lubrication time series are used for feasibility validation. In order to show its superiority, support vector machine (SVM) and RBF neural networks forecaster are also used during numerical simulations. Examples show that sparse Bayesian classification achieves comparable recognition accuracy to the SVM, and also requires substantially fewer kernel functions. Experimental results show the better performance in forecasting.

**Key words:** Sparse Bayesian classification; Support vector machine; Nonlinear forecasting; RBF neural network

### 1 引 言

时间序列预测在工业、交通、金融等许多领域都有广泛的应用, 国内外许多学者采用各种方法对时间序列进行了研究, 从传统的 ARMA 方法到目前采用较多的模糊逻辑和神经网络方法<sup>[1,2]</sup>. 各种神经网络预测方法以其良好的非线性、灵活且有效的学习方式在预测领域显示了极大的优势, 但这些方法都涉及到网络结构模型的选择和参数优化的问题<sup>[8]</sup>. 网络结构过于复杂或简单都会影响预测精度, 即神经网络方法难以解决系统中存在的复杂性和推广性(过学习)的矛盾.

为解决神经网络方法泛化能力低、最优网络结构难以确定, 从而影响预测精度的问题, 有些学者提

出了基于统计学习理论的支持向量机(SVM)预测方法<sup>[4]</sup>, 取得了较好的结果. 本文利用稀疏贝叶斯回归分类方法来进行时间序列的预测, 它不仅具有支持向量机的避免过学习的优点, 而且在达到与 SVM 相同性能的同时比 SVM 使用更少的核函数.

### 2 稀疏贝叶斯回归原理

人们对贝叶斯方法已研究了很多年, 但仅在近些年才广泛地应用于计算机视觉、信号处理、信息提取、数据挖掘以及数据分析等各个领域, 这主要是高速计算机的快速发展解决了边缘概率积分的复杂计算问题. 更进一步, MCMC 方法、EM 算法以及最近关于边缘概率计算的近似算法如变元推理等, 大大地扩展了贝叶斯方法的应用领域. 稀疏贝叶斯回归

收稿日期: 2005-03-15; 修回日期: 2005-05-10

基金项目: 国家自然科学基金项目(60175011, 60375011); 中国科技大学科学研究发展基金项目(030501F).

作者简介: 张旭东(1966—), 男, 合肥人, 副教授, 博士生, 从事模式识别与人工智能、图形图像处理等研究; 陈锋(1966—), 男, 副教授, 博士, 从事人工智能、智能交通和多传感器数据融合等研究.

就是一种基于贝叶斯框架<sup>[5,6]</sup>的核函数方法,它具有和SVM 相同的核函数形式

给定一系列输入向量和相应的输出, 监督学习的目的就是应用这些训练数据和先验知识来设计一个系统, 使设计出的系统对新的输入  $x$  预测出输出  $t$ . 假设有一组输入数据  $\{x_n\}_{n=1}^N$ , 以及相应的响应值  $t = \{t_n\}_{n=1}^N$ . 在这里假设目标值是标量, 并假设观察到的目标值是一个未知函数和一些噪声的结合, 即

$$t_n = y(x_n; w) + \epsilon_n, \quad (1)$$

其中:  $\epsilon_n$  是附加的噪声信号, 且是独立分布的;  $w$  是可调的参数或权值 函数族可由下式给出:

$$y(x_n; w) = \sum_{i=1}^M w_i \phi_i(x) = w^T \Phi(x), \quad (2)$$

它代表一组非线性基函数

$$\Phi(x) = (\phi_1(x), \phi_2(x), \dots, \phi_M(x))^T \quad (3)$$

的线性组合 由于式(1) 是模型采样加上噪声的结果, 假设噪声是均值为零、方差为  $\sigma^2$  的高斯噪声, 那么存在

$$p(t_n | x) = \mathbf{N}(t_n | y(x_n; w), \sigma^2). \quad (4)$$

基于独立性的假设, 全部数据的似然函数可写为

$$p(t | w, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} (t - \Phi w)^2\right\}$$

其中:  $\Phi = [\phi(x_1), \phi(x_2), \dots, \phi(x_N)]^T$  为设计矩阵, 大小为  $N \times (N + 1)$ ;

$$\phi(x_n) = [1, K(x_n, x_1), \dots, K(x_n, x_N)]^T;$$

$$t = (t_1, t_2, \dots, t_N)^T; w = (w_0, w_1, \dots, w_N)^T;$$

内积函数  $K(x_i, x)$  称为核函数

训练学习的目的就是获得参数  $w$ . 经典的贝叶斯方法是求  $w$  的点估计, 由于训练数据和参数很多, 如果按最大似然比函数法来估计参数  $w$ , 不可避免地会产生过学习现象 为避免这一情况, 一般的方法将权值系数  $w$  加上一个附加的约束 比如, 对似然函数加上复杂的惩罚函数或错误函数等 而稀疏贝叶斯方法直接给权值参数加上一个条件概率分布的限制, 给参数  $w$  定义一个均值为零的高斯概率分布或拉普拉斯条件概率分布, 用以限制模型的复杂度, 即

$$p(w | \alpha) = \prod_{i=0}^N \mathbf{N}(w_i | 0, \alpha^{-1}), \quad (5)$$

或

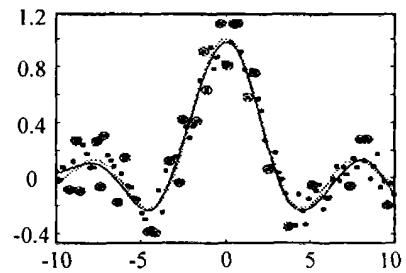
$$p(w | \alpha) = \exp\left\{-\sum_i \alpha_i |w_i|\right\} \exp\left\{-\sum_i \alpha w_i\right\}, \quad (6)$$

其中  $\alpha$  是假设参数向量

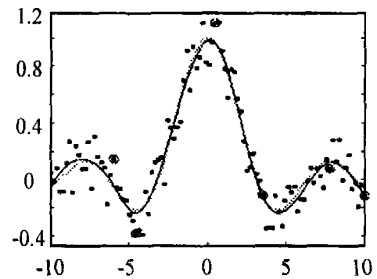
稀疏贝叶斯方法与给似然函数加上惩罚函数等方法的 不同之处在于: 稀疏贝叶斯方法在迭代优化

的过程中能使大部分权值  $w_i$  趋于零<sup>[7,8]</sup>, 其相应的核函数和训练样本被剔除, 只保留少数的权值  $w_i$  不为零, 其对应的保留下来的训练样本称为相关向量 (RV), 类似于支持向量机中的支持向量 (SV). 另外, 稀疏贝叶斯方法的核函数不受 Mercer 条件的限制, 而且还能给出回归和分类结果的后验概率, 其模型的稀疏性可根据训练数据自动调整; 而支持向量机的稀疏性则需要由参数  $C$  来控制

图 1 显示了稀疏贝叶斯和支持向量机对 100 个点的加上高斯噪声的  $\text{sinc}(x)$  函数的回归分析, 其中, 支持 / 相关向量用圆圈标出 训练后的稀疏贝叶斯的相关向量数目为 5, 而 SVM 的支持向量数目为 29; 稀疏贝叶斯回归的均方误差为 0.024 5, 而 SVM 的均方误差为 0.029



(a) 支持向量机



(b) 稀疏贝叶斯

图 1 支持向量机和稀疏贝叶斯对加噪  $\text{sinc}(x)$  函数的曲线拟合

### 3 基于稀疏贝叶斯回归的时间序列预测

#### 3.1 预测模型的确定

任何一个时间序列都可以看成是由一个非线性机制确定的输入输出系统 因此, 时间序列预测本质上就是利用历史数据序列寻求一个映射关系  $f: R^m$

$R$ , 使其逼近数据中的隐含非线性机制  $F$ . 对于给定的时间序列  $x(i), i = 1, 2, \dots, N$ , 假设已知  $x(t-1)$  预测  $x(t)$ , 则可建立映射  $f: R^m \rightarrow R$  使其满足

$$x(t) = f(x(t-1), x(t-2), \dots, x(t-m)). \quad (7)$$

映射  $f$  就是要求的逼近函数,  $m$  为预测系统的嵌入维数 用稀疏贝叶斯进行时间序列预测就是将  $x_t = [x(t-1), x(t-2), \dots, x(t-m)]$  作为  $m$  维输入向

量, 而将  $y_t = x(t)$  作为系统的期望输出, 来训练稀疏贝叶斯的权值, 从而得到逼近函数  $f = y(x_n; w)$ .

### 3.2 算法步骤

将时间序列分为两部分, 前一部分用于预测器的训练学习, 其余部分作为测试数据用于验证模型的有效性. 由于不能从公式中直接获得参数  $w$ , 故可采用 EM 迭代方法或 MCMC 方法获取参数  $w$ .

对于参数  $w$  按式 (6) 的拉普拉斯条件概率分布, 可采用下面的 EM 迭代步骤:

Step 1: 根据训练样本及所采用的核函数, 计算矩阵  $H$ , 其中  $n_{tr}$  为用于训练的时间序列数目:

$$x_i = [x_{i+1}, x_{i+2}, \dots, x_{i+m}],$$

$$i = 0, 1, \dots, n_{tr} - m; \quad (8)$$

$$h(x) = [1, K(x, x_1), \dots, K(x, x_n)]^T; \quad (9)$$

$$H = [h^T(x_1), \dots, h^T(x_n)]^T. \quad (10)$$

Step 2: 按下式对  $w$  进行初始化 ( $\epsilon = 10^{-6}$ ):

$$y = [x_{m+1}, x_{m+2}, \dots, x_{n_{tr}}]^T, \quad (11)$$

$$\hat{w}^{(0)} = (\epsilon I + H^T H)^{-1} H^T y. \quad (12)$$

Step 3: ( $E$  步) 根据当前  $\hat{w}^{(0)}$ , 按下式计算矩阵  $\Theta$ :

$$\Theta = \text{diag}(|\hat{w}_1^{(0)}|, \dots, |\hat{w}_m^{(0)}|). \quad (13)$$

Step 4: ( $M$  步) 根据下式获得新的估计值  $w(t+1)$ :

$$\hat{w}^{t+1} = \Theta(I + \Theta H^T H \Theta)^{-1} \Theta H^T y. \quad (14)$$

如果  $|\hat{w}^{(t+1)} - \hat{w}^{(t)}| / \hat{w}^{(t)} < \delta$ , 则迭代停止; 否则, 转向 Step 3.

### 3.3 评价标准

为了评价模型的预测性能, 需要计算预测的准确性. 这里采用均方根相对误差对模型的预测结果进行统计:

$$RMSE = \sqrt{\frac{1}{N_{te}} \sum_{n=1}^{N_{te}} ((y_n - t_n)/t_n)^2}, \quad (15)$$

其中:  $N_{te}$  表示测试样本数,  $y_n$  和  $t_n$  分别代表模型预测值和时间序列的实际值.

## 4 实验结果

### 4.1 模拟预测实验

为验证稀疏贝叶斯方法的预测效果, 设计了一个预测序列的模拟实验. 在混沌动力系统中, Logistic 比较简单而且研究得比较透彻, 所以用它来产生混沌序列. Logistic 方程常用的迭代形式为  $x_{n+1} = kx_n(1-x_n)$ , 参数  $k$  在  $3.57 < k < 4$  情况下, 系统呈现混沌状态. 实验中取  $k = 3.7, x_0 = 0.41$ , 嵌入维数  $m = 7$ , 时间序列  $n = 200$ , 前 100 个用于训练, 后 100 个用于测试. 预测的结果如表 1 所示.

表 1 不同预测步长下的根均方相对误差比较

预测步长	1	2	3	4	5
SVM	0.0036	0.0159	0.0193	0.0354	0.0430
稀疏贝叶斯	0.0050	0.0178	0.0217	0.0367	0.0471

表 1 数据表明, 对于 Logistic 混沌序列的预测, 稀疏贝叶斯方法可以达到和支持向量机相近的性能, 而且达到了较高的预测精度.

### 4.2 油滑数据金属含量预测实验

分别用 RBF 神经网络、支持向量机及本文方法对某发动机两次换油间油滑数据中的 Fe 金属含量 (数据来源见 <http://pw.matwav.com/skyhawk/main/gb.htm>) 进行预测. 支持向量机和稀疏贝叶斯均采用高斯径向基核函数, 参数  $\sigma = 0.5$ . 图 2 给出了 3 种方法在训练样本数为 60, 测试样本数为 34 的情况下多步预测的结果. 从图中可以看出, 3 种方法均取得了较好的结果.

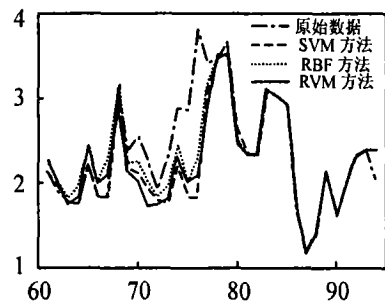


图 2 几种方法的预测结果比较

实验中通过减少训练样本数来分析几种方法进行预测时的泛化能力, 表 2 列出了在不同训练样本数的情况下, 3 种方法预测的均方根相对误差. 从中可以看出, 在训练样本数量较多的情况下, RBF 神经网络方法和其他两种方法的预测效果相近. 但随着训练样本数量的减少, RBF 神经网络方法预测的相对误差增加较快, 而支持向量机和稀疏贝叶斯方法预测的相对误差增加较为缓慢, 这种情况在多步预测时表现得更为明显. 这表明支持向量机和稀疏贝叶斯方法能较好地解决模型复杂性和推广性的矛盾, 具有较好的泛化能力, 而且在小样本情况下的预测效果要明显优于 RBF 网络.

另外还测试了在不同输入向量维数 (嵌入维数) 情况下, 支持向量机和稀疏贝叶斯单步预测时的相对均方根误差和支持 (相关) 向量数目, 表 3 列出了测试结果 (支持向量机参数:  $C = 800, \epsilon = 0.001$ , 训练样本数为 60). 从表 3 中可以看出, 随着向量维数的增加, 系统的测试误差减小, 预测精度提高; 但输入向量维数增加到一定程度时, 误差的减小

表2 几种方法在不同训练样本数和不同预测步长下(1步和5步)的相对均方根误差比较

训练样本数	60	50	40	30	20
RBF神经网络	0.0303(0.0556)	0.0999(0.2148)	0.3962(1.7976)	1.1319(4.2564)	1.3200(15.5277)
支持向量机	0.0325(0.0665)	0.1194(0.3094)	0.2284(0.4298)	0.4063(0.6241)	0.4103(0.4478)
稀疏贝叶斯	0.0310(0.0595)	0.1324(0.2632)	0.1704(0.2944)	0.3321(0.5350)	0.3383(0.3522)

表3 两种方法在不同输入向量维数(系统阶数)下的预测误差和支持(相关)向量数目的比较

输入向量维数(嵌入维数)		1	2	3	4	5	6	7	8	9
误差比较	SVM	0.2289	0.1844	0.1167	0.0598	0.0320	0.0320	0.0325	0.0325	0.0325
	稀疏贝叶斯	0.2304	0.2371	0.2517	0.0306	0.0299	0.0285	0.0304	0.0299	0.0310
支持(相关)向量数目比较	SVM	58	58	57	56	55	53	53	52	51
	稀疏贝叶斯	2	2	2	38	40	51	51	45	46

就不明显了。此外,在达到相同预测效果的情况下,稀疏贝叶斯方法使用了更少的相关向量,在输入向量维数较低的情况下,更为突出

## 5 结 语

本文采用稀疏贝叶斯方法对时间序列预测进行了研究。通过与支持向量机和RBF神经网络对时间序列进行预测的对比实验,得出以下结论:

1) 稀疏贝叶斯方法和神经网络、支持向量机等方法一样,可以用于时间序列的预测,并具有较高的预测精度

2) 与RBF神经网络相比,稀疏贝叶斯和支持向量机一样具有很强的推广能力,在样本数量较大、短期预测中,神经网络方法与稀疏贝叶斯和支持向量机相差不大;但在训练样本数较少、长期预测中,RBF神经网络效果相对较差

3) 在预测性能相近的情况下,稀疏贝叶斯与支持向量机相比,使用了较少的核函数,而且核函数的函数形式不象在SVM中受到限制,这意味着在实际应用中适应范围广且具有更快的计算速度和使用较少的内存

稀疏贝叶斯方法用于时间序列预测虽然具有较好的推广性能,但仍存在着需要离线训练的缺点,而且训练中需要进行大量的矩阵运算,使训练的计算复杂度提高。因此,下一步的工作目标是解决稀疏贝叶斯方法在训练过程中的计算复杂度问题,研究利用新的算法降低计算复杂度,提高收敛速率

## 参考文献(References)

[1] 韩敏,王晨,席剑辉. 基于改进RBF神经网络的非线性时间序列预测[J]. *仪器仪表学报*, 2003, 24(Z1): 574-575

(Han M, Wang C, Xi J H. Nonlinear Time Series Prediction Based on Improved RBF Network [J]. *Chinese J of Scientific Instrument*, 2003, 24(Z1): 574-575.)

[2] 梁艳春,王政,周春光. 模糊神经网络在时间序列预测中的应用[J]. *计算机研究与发展*, 1998, 35(7): 664-667.

(Liang Y C, Wang Z, Zhou C G. Application of Fuzzy Neural Networks to the Prediction of Time Series [J]. *Computer Research and Development*, 1998, 35(7): 664-667.)

[3] 张浩然,韩正之,李昌刚. 基于支持向量机的非线性系统辨识[J]. *系统仿真学报*, 2003, 15(1): 119-121.

(Zhang H R, Han Z Z, Li C G. Support Vector Machine Based Nonlinear Systems Identification [J]. *J of System Simulation*, 2003, 15(1): 119-121.)

[4] 朱家元,段宝君,张恒喜. 新型SVM对时间序列预测研究[J]. *计算机科学*, 2003, 30(8): 124-125

(Zhu J Y, Duan B J, Zhang H X. Prediction of Time Series Based Least Squares Support Vector Machines [J]. *Computer Science*, 2003, 30(8): 124-125.)

[5] MacKay D J C. Bayesian Interpolation [J]. *Neural Computation*, 1992, 4(3): 415-447.

[6] MacKay D J C. The Evidence Framework Applied to Classification Networks [J]. *Neural Computation*, 1992, 4(3): 720-736

[7] Tipping M. The Relevance Vector Machine [A]. *Advances in Neural Information Processing Systems* [C]. Cambridge: Cambridge MIT Press, 2000: 652-658

[8] Figueiredo M. Adaptive Sparseness for Supervised Learning [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2003, 25(9): 1150-1159