

动态贝叶斯最小二乘支持向量机

潘宇雄, 任章, 李清东

(北京航空航天大学 a. 自动化科学与电气工程学院, b. 飞行器控制一体化技术国防科技重点实验室, 北京 100191)

摘要: 为了对涡扇发动机的运行参数变化进行实时高精度预测, 提出一种基于动态贝叶斯最小二乘支持向量机 (LS-SVM) 的时间序列预测算法. 该算法将贝叶斯证据框架理论用于推断 LS-SVM 的初始模型参数; 然后, 利用样本增减迭代学习算法实现 LS-SVM 的参数动态调整. 对某型涡扇发动机的摩擦力矩时间序列进行动态预测, 并与动态 LS-SVM 模型的预测结果进行比较. 结果显示, 动态贝叶斯 LS-SVM 具有较好的预测精度.

关键词: 动态预测; 时间序列预测; 最小二乘支持向量机; 贝叶斯证据框架

中图分类号: TP29

文献标志码: A

Dynamic Bayesian least squares support vector machine

PAN Yu-xiong, REN Zhang, LI Qing-dong

(a. School of Automation Science and Electrical Engineering, b. Science and Technology on Aircraft Control Laboratory, Beihang University, Beijing 100191, China. Correspondent: PAN Yu-xiong, E-mail: twhpyx@gmail.com)

Abstract: A time series prediction method based on the dynamic Bayesian least squares support vector machine (LS-SVM) is proposed to accurately predict operating parameters of the turbofan engine. By the Bayesian evidence framework theory, initial model parameters of the LS-SVM are inferred. Dynamic learning of the LS-SVM and dynamic prediction of time series are realized by the recursively incremental and decremental sample learning method. The method is applied to dynamically predict frictional moment time series of a certain type of the turbofan engine, and a comparison is made with the dynamic LS-SVM. The results show that the proposed method is effective in improving prediction accuracy.

Key words: dynamic prediction; time series prediction; LS-SVM; Bayesian evidence framework

0 引言

时间序列预测方法的基本思想是, 根据系统有限长度的观测数据, 建立能够比较精确地反映时间序列中所包含的动态依存关系的数学模型, 并以此对系统的未来行为进行预测^[1]. 目前, 关于时间序列预测的研究大都采用离线方法, 如传统统计学方法和各种人工智能方法等, 以批处理方式对时间序列数据进行建模, 然后利用模型进行预测. 然而, 在实际应用中所采集的时间序列将不断实时动态更新, 并且新采集的时间序列存在不确定性和数据量较大的情况, 利用现有的离线预测模型, 往往难以取得较好的精度, 并难以在短时间内实现算法和模型重建. 因此, 大多数时间序列预测问题, 如动态系统运行参数监测和经济运行指数等问题, 都需要以在线处理的方式予以解决.

针对时间序列动态预测问题, 学者们提出了滑

动时间窗模型^[2]、动态神经网络算法^[3]和动态 LS-SVM^[4] 等动态预测方法, 但绝大部分此类在线预测算法和模型并非真正意义上的在线工作方式, 而只是通过对于数据预处理或模型形式的改变实现在线预测. 其中动态 LS-SVM 模型能够实现 LS-SVM 模型某些参数的在线动态更新, 可展现出良好的发展潜力和应用价值, 因此日益受到研究者的重视. 但是, 对于数据特性变化大的时间序列, 其预测精度的控制问题需要进一步考虑.

针对上述问题, 并考虑涡扇发动机的高动态特性, 本文提出一种基于动态贝叶斯 LS-SVM 的时间序列预测方法. 贝叶斯 LS-SVM^[5] 具有较好的短期预测精度, 但随着时间推移, 误差积累, 中长期预测精度会逐渐下降, 特别是当时间序列变化较大时, 预测结果难以接受. 而且, 在规定时间内, 利用新加入的时

收稿日期: 2013-08-26; 修回日期: 2013-12-20.

基金项目: 国家自然科学基金项目 (61101004, 60874117); 高等学校学科创新引智计划项目 (B07009).

作者简介: 潘宇雄 (1983—), 男, 博士生, 从事可靠性、故障预测与健康管理的研究; 任章 (1957—), 男, 教授, 博士生导师, 从事可靠性、导航以及制导与控制等研究.

间序列样本进行贝叶斯 LS-SVM 的重建难于实现,因此需要对贝叶斯 LS-SVM 进行改进. 动态贝叶斯 LS-SVM 利用贝叶斯证据框架理论, 采用训练数据, 推断 LS-SVM 的初始参数, 建立初始贝叶斯 LS-SVM 预测模型. 随着新的时间序列样本的加入, 采用贝叶斯框架理论动态更新正则化参数, 并利用样本增减迭代学习方法和更新后的正则化参数, 实时调整模型参数, 从而实现时间序列的动态高精度预测.

1 最小二乘支持向量机

给定一个有 N 个数据的训练集合 $\{x_k, y_k\}, k=1, 2, \dots, N$. 其中: 输入数据 $x_k \in R^n$, 输出数据 $y_k \in R$.

函数拟合问题可表示为如下优化问题:

$$\min_{\mathbf{w}, b, e} J_1(\mathbf{w}, e) = \frac{\mu}{2} \mathbf{w}^T \mathbf{w} + \frac{\zeta}{2} \sum_{k=1}^N e_k^2 = \mu E_w + \zeta E_d,$$

约束条件为

$$y_k = \mathbf{w}^T \varphi(x_k) + b + e_k, k = 1, 2, \dots, N. \quad (1)$$

其中: $\varphi(\cdot) : R^n \rightarrow R^{n_h}$ 为将输入数据映射到高维特征空间的函数, 权值向量 $\mathbf{w} \in R^{n_h}$, 误差向量和偏置值满足 $e_k \in R$ 和 $b \in R$, μ 和 ζ 为超参数.

根据式(1)将模型变换到对偶空间加以解决, 得到如下 Lagrange 函数:

$$L(\mathbf{w}, b, e; \alpha) = J_1(\mathbf{w}, e) - \sum_{k=1}^N \alpha_k \{\mathbf{w}^T \varphi(x_k) + b + e_k - y_k\}, \quad (2)$$

其中 $\alpha_k \in R$ 为 Lagrange 乘子. 根据 KKT(Karush-Kuhn-Tucher) 最优条件, 对式(2)求偏导, 消去 e_k 和 \mathbf{w} 后整理可得

$$\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \mathbf{\Omega} + \gamma^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}. \quad (3)$$

其中: $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$; $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$; $\mathbf{1} = [1, 1, \dots, 1]^T$; $\mathbf{\Omega}$ 为核矩阵, 其第 i 列 j 行的元素为 $\Omega_{ij} = K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$; \mathbf{I} 为 $N \times N$ 的单位矩阵; $\gamma = \zeta/\mu$ 为正则化参数.

通过求解式(3)可以得到 $\boldsymbol{\alpha}$ 和 b , 则 LS-SVM 的拟合模型为

$$y(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b. \quad (4)$$

其中: $K(x, x_k) = \exp\{-\|x - x_k\|^2/\sigma^2\}$ 为满足 Mercer 定理的径向基核函数, 正则化参数 γ 和核函数系数 σ 通过贝叶斯证据框架选择.

2 基于贝叶斯证据框架的 LS-SVM

传统 LS-SVM 参数 (γ, σ) 的确定一般采用交叉验证方法, 即对每一对 (γ, σ) , 每次用训练集训练 LS-SVM 后, 对验证集进行预测. 在每对 (γ, σ) 预测不同

验证集之后, 选取预测误差值最小的 (γ, σ) 作为 LS-SVM 的模型参数. 采用交叉验证方法需花费大量时间, 特别不适用于大样本建模, 故本文采用贝叶斯证据框架理论来确定较优的 (γ, σ) , 以缩短建模时间.

贝叶斯证据框架的基本思想是最大化参数分布的后验概率, 而最优的参数值是在后验概率最大时获得的. 贝叶斯框架分为以下 3 层^[5].

第 1 层推断: 给定数据点 $D = \{x_k, y_k\}_{k=1}^N$ 和模型 H 中的正则化参数 γ , 根据贝叶斯准则, 模型参数 \mathbf{w} 和 b 可由下式推求:

$$p(\mathbf{w}, b|D, \lg \mu, \lg \zeta, H) = \frac{p(\mathbf{w}, b| \lg \mu, \lg \zeta, H)p(D|\mathbf{w}, b, \lg \mu, \lg \zeta, H)}{p(D| \lg \mu, \lg \zeta, H)}. \quad (5)$$

其中: $p(\mathbf{w}, b| \lg \mu, \lg \zeta, H)$ 为先验概率; 证据概率分布 $p(D| \lg \mu, \lg \zeta, H)$ 为一个标准化的常数, 对于所有的 \mathbf{w} 和 b 的积分值为 1; $p(D|\mathbf{w}, b, \lg \mu, \lg \zeta, H)$ 为相似度.

第 2 层推断: 应用贝叶斯定理可由数据 D 推导出正则化参数

$$\begin{aligned} \gamma &= \zeta/\mu, \\ p(\lg \mu, \lg \zeta|D, H) &= \frac{p(\lg \mu, \lg \zeta|H)p(D| \lg \mu, \lg \zeta, H)}{p(D|H)}, \end{aligned} \quad (6)$$

其中假设超参数满足均值分布.

第 3 层推断: 将贝叶斯准则应用于第 3 层, 通过选择不同的核函数, 可获得不同模型的后验概率为

$$p(H_j|D) \propto p(D|H_j)p(H_j) \propto p(D|H_j). \quad (7)$$

3 样本增减迭代算法

为了动态跟踪和预测系统的特征参数, 当系统产生新的数据时, 需要将此新信息加入系统的预测模型, 并相应地改变模型参数. 但是, 随着数据的增加, 核函数的维度也随之增加, 最终耗尽系统资源. 因此, LS-SVM 的动态训练和学习过程应包括增加样本和减少样本两个过程, 利用样本增减迭代方法进行.

设新增加一个样本 (x_{N+1}, y_{N+1}) , 并将其加入原训练样本集 $\{x_k, y_k\}_{k=1}^N$. 对应式(3), 样本为 N 的现有模型与样本为 $N+1$ 的新模型的增量关系为

$$\mathbf{A}_{N+1} \boldsymbol{\alpha}_{N+1} = \mathbf{Y}_{N+1}. \quad (8)$$

其中

$$\begin{aligned} \mathbf{A}_{N+1} &= \begin{bmatrix} \mathbf{A}_N & \boldsymbol{\beta} \\ \boldsymbol{\beta}^T & c \end{bmatrix}, \quad \mathbf{A}_N = \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \mathbf{\Omega} + \gamma^{-1} \mathbf{I} \end{bmatrix}, \\ \boldsymbol{\alpha}_{N+1} &= \begin{bmatrix} \boldsymbol{\alpha}_N \\ \alpha_{N+1} \end{bmatrix}, \quad \mathbf{Y}_{N+1} = \begin{bmatrix} \mathbf{Y}_N \\ y_{N+1} \end{bmatrix}, \quad \mathbf{Y}_N = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}, \\ \boldsymbol{\beta} &= [1, K(x_1, x_{N+1}), \dots, K(x_N, x_{N+1})], \\ c &= \gamma^{-1} + K(x_{N+1}, x_{N+1}). \end{aligned}$$

在线增量训练算法的目标是有效地更新逆矩阵 \mathbf{A}_{N+1}^{-1} , 根据文献 [6] 中的方法, 有

$$\mathbf{A}_{N+1}^{-1} = [c - \beta^T \mathbf{A}_N^{-1} \beta]^{-1} \begin{bmatrix} \mathbf{A}_N^{-1} \beta \\ -1 \end{bmatrix} + \begin{bmatrix} \beta^T \mathbf{A}_N^{-1} & -1 \\ \mathbf{0} & 0 \end{bmatrix}. \quad (9)$$

随着新样本的持续增加, 由式 (9) 计算出来的逆矩阵 \mathbf{A}^{-1} 的维数也不断增加, 导致储存资源耗尽, 影响工作效率. 为了避免这些问题的发生, 在增加样本之后, 必须进行相应的减小样本过程.

为了避免传统样本减少, 在此过程中重新计算矩阵的逆, 并将 \mathbf{A}_{N+1}^{-1} 更新为逆矩阵 $\overline{\mathbf{A}}_N^{-1}$, 可将下式作为更新规则来剪除 $N+1$ 个样本中的第 k 个样本^[7]:

$$\alpha_{ij} \leftarrow \alpha_{ij} - \alpha_{kk}^{-1} \alpha_{ik} \alpha_{kj}. \quad (10)$$

其中: $i, j = 1, 2, \dots, N$, 且 $i, j \neq k$; α_{ij} 表示 \mathbf{A}_{N+1}^{-1} 中第 i 行第 j 列的元素; k 表示第 k 个支持向量被移除. LS-SVM 的动态学习过程如图 1 所示.

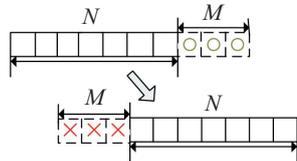


图 1 LS-SVM 动态学习框图

4 动态贝叶斯 LS-SVM 模型的建立

基于动态贝叶斯 LS-SVM 的时间序列预测模型的输入与输出之间的非线性关系由 LS-SVM 实现, 而模型参数和核函数系数根据贝叶斯证据框架推导得出, 最后利用样本增减迭代学习方法、反馈的超参数以及新增样本数据实现贝叶斯 LS-SVM 的动态预测, 如图 2 所示.

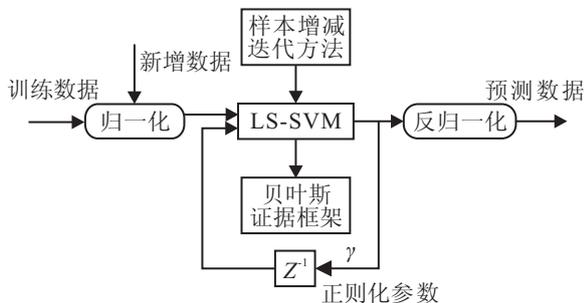


图 2 动态贝叶斯 LS-SVM 模型框架

动态贝叶斯 LS-SVM 模型的建模过程如下:

Step 1: 对训练数据进行归一化, 使数据的均值为 0, 方差为 1, γ 的初值取 10, σ 的初值取 $R/1.96$, R 为归一化后训练集的变化范围之和.

Step 2: 通过求解式 (3) (即第 1 层) 得到优化参数 \mathbf{w} 和 b ; 优化参数 μ 和 ζ 可由式 (6) 求得, 并得到正则化参数 γ .

Step 3: 通过式 (7) 可计算优化参数 μ 和 ζ 下的后验概率 $p(H_j|D)$.

Step 4: 返回 Step 2, 调整 σ 值, 重新计算相关参数和后验概率 $p(H_j|D)$.

Step 5: 当 $p(H_j|D)$ 值最大时, 所对应的参数 μ , ζ 和 σ 为 LS-SVM 在训练数据下的最优参数. 自此, 通过前 5 步的建模过程便得到离线条件下的贝叶斯 LS-SVM 模型.

Step 6: 当有新的样本数据输入时, 以样本增加前求得的参数 γ 为起始点, 由式 (7) 求得新样本下的 γ , 完成式 (3) 中正则化参数的更新.

Step 7: 将新样本值和更新后的正则化参数 γ 代入式 (9), 计算增加样本后的逆矩阵 \mathbf{A}_{N+1}^{-1} , 通过式 (10) 计算求得减少样本后的逆矩阵 $\overline{\mathbf{A}}_N^{-1}$, 以完成 LS-SVM 的增减学习, 并通过式 (3) 计算新的 α 和 b 值, 训练出一个新的 LS-SVM.

Step 8: 当新样本数据持续输入时, 重复 Step 6 和 Step 7, 动态调整 LS-SVM 的模型参数, 完成 LS-SVM 的动态训练和动态预测. 通过对预测数据进行反归一化, 得到预测结果.

5 实例分析

现今多数的喷气机引擎都是采用涡扇发动机作为动力来源, 涡扇发动机风扇的好坏直接影响到发动机的性能, 风扇的摩擦力矩过大将影响整个发动机的正常运行. 以某型发动机风扇为例, 对其摩擦力矩时间序列进行动态预测, 以便在摩擦力矩达到某一阈值时, 进行相应的维修.

采用某型风扇运行过程中一组摩擦力矩时间序列, 将获得的样本分为 200 组训练集和 160 组测试集. 设定参数优化范围, 采用贝叶斯理论推断最优参数, 得到 $(\gamma, \sigma) = (41.78, 18.24)$, 并对模型进行训练, 得到贝叶斯 LS-SVM 模型. 采用传统交叉验证方法优化 $(\gamma, \sigma) = (34.26, 19.89)$, 得到 LS-SVM 模型. 随着新序列样本的加入, 利用本文所提出的方法实时调整贝叶斯 LS-SVM 的模型参数, 建立动态贝叶斯 LS-SVM. 而对于 LS-SVM 模型, 利用传统交叉验证方法得到优化参数 (γ, σ) , 并采用传统的样本增减方法建立动态 LS-SVM. 图 3 为某型风扇摩擦力矩的预测结果, 图中描述了动态贝叶斯 LS-SVM 和动态 LS-SVM 的预测序列值与实际序列的对比情况.

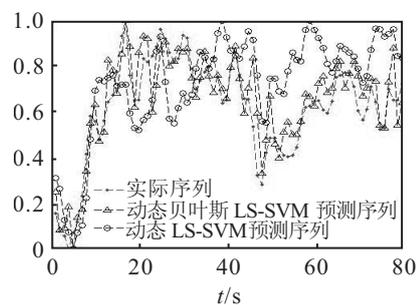


图 3 时间序列预测结果对比

由图3可以看出,动态贝叶斯LS-SVM时间序列预测模型能有效地预测摩擦力矩时间序列,而动态LS-SVM最开始能较好预测时间序列,但是随着时间的增加其预测误差将逐渐增大.为了全面比较算法的预测性能,采用两个定量的性能指标对预测结果进行评价,一是平均绝对误差(MAE),二是规范化均方根误差(NRMSE),分别为

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (11)$$

$$NRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (12)$$

其中: n 为预测集数据个数, y_i 为真实值, \hat{y}_i 为预测值, \bar{y} 为序列均值.

对摩擦力矩时间序列进行动态预测,进行5次实验的平均结果如表1所示.

表1 算法性能对比

序号	预测数据	算法	MAE	NRMSE
1	1 ~ 80	动态贝叶斯LS-SVM	0.049 5	0.241 4
		动态LS-SVM	0.087 0	0.547 2
2	81 ~ 160	动态贝叶斯LS-SVM	0.126 7	1.699 7
		动态LS-SVM	0.356 9	5.698 1

从表1可以看出,使用动态贝叶斯LS-SVM模型的时间序列预测方法,其平均绝对误差和规范化均方根误差均小于传统动态LS-SVM模型的误差值,并且随着时间的推移,动态贝叶斯LS-SVM相对于动态LS-SVM的预测精度越来越高.原因是随着新序列样本的加入,由于贝叶斯证据框架的推导,贝叶斯动态LS-SVM模型能够实时调整正则化参数 γ 的值,而传统动态LS-SVM模型的正则化参数 γ 是一个固定值,模型参数调整不完整,因此动态LS-SVM相对动态贝叶斯LS-SVM的预测精度越来越低.

6 结 论

为了高精度预测涡扇发动机的动态特性,本文提出了动态贝叶斯LS-SVM的时间序列预测模型.该

模型在贝叶斯LS-SVM的基础上,采用贝叶斯证据框架和样本增减迭代学习方法动态调整模型参数,提高LS-SVM的预测精度.实例分析结果表明,对于数据特性变化大的时间序列,与传统的动态LS-SVM相比较,动态贝叶斯LS-SVM预测精度较高,有着广阔的应用前景.

参考文献(References)

- [1] 韩敏.混沌时间序列预测理论与方法[M].北京:中国水利水电出版社,2007:7-8.
(Han M. Prediction theory and methods for chaotic time series[M]. Beijing: China Water Power Press, 2007: 7-8.)
- [2] 沈键,杨煜普.基于滚动时间窗的动态协同过滤推荐模型及算法[J].计算机科学,2013,40(2):206-209.
(Shen J, Yang Y P. Dynamic collaborative filtering recommender model based on rolling time window and its algorithm [J]. Computer Science, 2013, 40(2): 206-209.)
- [3] Vairappan C, Tamura H. Batch type local search-based adaptive neuro-fuzzy inference system with self-feedback for time-series prediction[J]. Neurocomputing, 2009, 72(7): 1870-1877.
- [4] 曹葵康,沈海斌.基于多样本的在线支持向量回归算法[J].浙江大学学报:理学版,2011,38(4):405-408.
(Cao K K, Shen H B. On-line support vector regression with multiple samples[J]. J of Zhejiang University: Science Edition, 2011, 38(4): 405-408.)
- [5] 陈磊,张士桥.基于贝叶斯最小二乘支持向量机的时用水量预测模型[J].天津大学学报,2006,39(9):1037-1043.
(Chen L, Zhang S Q. Hourly water demand forecast squares support model based on Bayesian least vector machine[J]. J of Tianjin University, 2006, 39(9): 1037-1043.)
- [6] Golub G, Van Loan C. Matrix computations[M]. Baltimore: Johns Hopkins University Press, 1996: 130-132.
- [7] Khaled B, Laurent B. Incremental and decremental multi-category classification by support vector machine[C]. Proc of the Int Conf on Machine Learning and Application. Miami: IEEE, 2009: 294-300.

(责任编辑:滕蓉)