

文章编号: 1001-0920(2007)12-1377-04

一种基于特征提取方法的智能预测算法

李 鹏, 刘 民, 吴 澄
(清华大学 自动化系, 北京 100084)

摘 要: 针对色织生产调度过程中的一类整经轴数预测问题, 提出一种整经轴数智能预测算法. 首先基于线性特征提取方法 (PCA) 和非线性特征提取方法 (LLE) 对影响整经轴数的多维属性参数进行特征提取, 然后采用前向神经网络进行整经轴数预测. 数值计算结果表明, 所提出的方法能满足实际生产过程整经轴数预测的需要.

关键词: 特征提取; 人工神经网络; 预测; 纺织

中图分类号: TP273 **文献标识码:** A

An intelligent prediction algorithm based on feature extraction methods

LI Peng, LIU Min, WU Cheng

(Department of Automation, Tsinghua University, Beijing 100084, China. Correspondent: LIU Min, E-mail: lium@tsinghua.edu.cn)

Abstract: An intelligent algorithm is proposed to predict the trim-beam number in the colored weaving scheduling process. Linear and nonlinear feature extraction methods including PCA and LLE are used to extract features from high-dimensional properties which are relative to the trim-beam number. Then, neural networks are applied to predict the trim-beam number. Numerical computational results based on practical production data show that the proposed algorithm can satisfy the trim-beam number prediction requirements in the practical manufacturing process.

Key words: Feature extraction; Artificial neural network; Prediction; Textile

1 引 言

整经工序是纺织生产过程中的重要工序. 每个生产任务对应的整经轴数关系到该生产任务在整经工序的装配难度、机台占用数量、加工时间、占用的关键配件数量等, 是整经生产过程和调度过程中的重要参数. 在色织生产过程中, 该参数对生产调度过程的影响尤其明显. 纺织生产任务的整经轴数 (以下简称整经轴数) 要由专门的工艺人员完成复杂的整经工艺设计后确定, 经常滞后于生产调度过程, 影响了调度效率和调度效果, 因而需要对整经轴数进行快速准确的预测. 整经轴数预测问题的输入数据项是生产任务的属性参数, 数量多且关系复杂, 其输出数据项为整数的整经轴数.

采用常用的一些非线性回归方法直接解决整经轴数预测问题, 容易受到训练速度慢、输入项数量多

且耦合严重、数据噪声大等因素的制约, 而将整经轴数预测这类模式识别问题划分为特征提取和非线性回归两个阶段, 有助于提高训练速度和预测精度、降低噪声影响^[1-3]. 已有研究者使用 SVM (特征选择) + 前向神经网络 (非线性回归) + 改进的 BP 训练方法对整经轴预测问题进行研究, 并获得了一些成果^[4]. 但该研究只利用 SVM 这种特征选择方法, 而没有进行特征提取, 且能处理的样本数目和计算速度仍受到 SVM 的较大限制^[5].

本文研究整经轴数预测问题, 提出一种整经轴数智能预测算法. 它由离线与在线两部分构成, 集成了保持类平均向量分类信息的主成分分析 (PCA) 和局部线性嵌入 (LLE) 两种特征提取方法, 包含特征提取方法比较、非线性回归器训练等关键环节. 该方法能有效地控制计算量, 提高整经轴数预测精度.

收稿日期: 2006-08-23; 修回日期: 2006-10-23.

基金项目: 国家重点基础研究计划项目 (2002CB312200); 国家自然科学基金项目 (60443009, 60274045); 北京市科技计划重点项目 (D0305005040321); 国家 863 计划项目 (2006AA04Z163); 教育部新世纪优秀人才支持计划.

作者简介: 李鹏 (1975—), 男, 北京人, 博士生, 从事复杂生产过程建模与优化调度的研究; 刘民 (1965—), 男, 杭州人, 副教授, 博士, 从事复杂生产过程建模与优化调度的研究.

2 整经轴数预测问题描述

整经轴数预测问题可用图 1 所示模型描述。

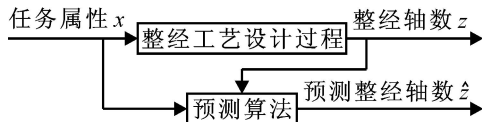


图 1 整经轴数预测问题模型

整经工艺设计过程的输入 $x(x \in R^D)$ 为生产任务的 D 维属性参数,包括布长、布宽、花型、纱色和生产工艺等。整经工艺设计过程的输出 $z(z \in Z^+)$ 为生产任务对应的整经轴数。整经工艺设计过程的输入输出二元组 (x, z) 构成预测模型的样本。设预测算法的输入输出函数关系为

$$\hat{z} = f(x, \theta) \quad R^+, \quad (1)$$

其中 θ 为广义参数。

整经轴数预测问题可描述为对样本集 $\{(x_i, z_i) \mid 1 \leq i \leq N\}$, 优化预测函数 f 及其广义参数 θ , 使得预测误差 $R(f, \theta)$ 最小, 即

$$R(f, \theta) = \sum_{i=1}^N (z_i - f(x_i, \theta))^2. \quad (2)$$

3 预测模型及算法

本文提出的预测模型如图 2 所示。

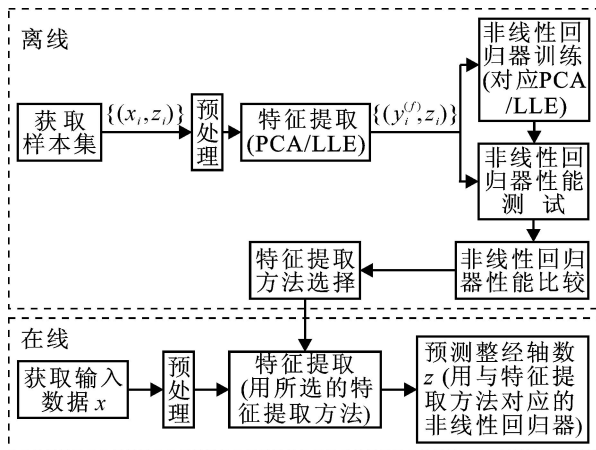


图 2 预测模型示意图

预测模型分为离线和在线两部分。离线部分定时运行,主要完成特征提取方法的比较和选择工作。在线部分利用离线部分选择得到的特征提取方法和非线性回归器,随时对新生产任务的整经轴数进行在线预测。预测模型通过离线、在线两部分的协同工作实现。

与文献[4]的研究不同的是,本预测算法使用了特征提取而非特征选择方法,这是基于如下考虑:特征提取方法能够保留更多与非线性回归有关的属性信息;特征提取方法所得到的低维特征可以是彼此独立的,能够去掉更多的冗余信息。

预测算法的实现步骤如下:

Step1: 对于时段 T 内已经完成整经工艺设计的生产任务,离线部分取其属性参数及其整经轴数作为样本集 $\{(x_i, z_i) \mid 1 \leq i \leq N\}$ 。通过预处理剔除样本集中的野值,同时规范所有属性值为实数。

Step2: 分别利用 PCA 和 LLE 特征提取方法,对样本集中的输入数据集 $\{x_i\}$ 进行特征提取,得到降维后的 d 维特征数据集 $\{y_i^{(f)}\}, f = \text{PCA/LLE}$ 。

Step3: 将每个 $\{y_i^{(f)}\}$ 分为训练集和测试集两部分。首先用训练集对非线性回归器进行训练,然后用测试集对训练完毕的非线性回归器进行测试,得到预测误差 $R^{(f)}$ 。比较两种特征提取方法对应的预测误差 $R^{(f)}$, 从中选择一个最好的特征提取方法。

Step4: 在线部分利用 Step3 得到的最优特征提取方法,对预处理后的新生产任务多维属性参数进行特征提取,进而用对应的(训练完毕的)非线性回归器进行整经轴数预测。

下面将介绍算法中 Step2 和 Step3 的详细内容。Step1 和 Step4 不再赘述。

3.1 特征提取方法

本算法集成了 PCA 和 LLE 两种特征提取方法:一种是线性降维方法,另一种是非线性降维方法。综合使用这两类特征提取方法是考虑到当整经轴数受主要特征量的线性影响较大时,这些特征量可以通过全局线性降维方法得到有效提取;而当企业生产品种的变化导致整经工艺制定原则和过程发生变化后,整经轴数特征空间可能变换为一个非线性空间,其流型结构需通过非线性特征提取方法来保存。

本算法使用的 PCA 方法和 LLE 方法都可通过线性代数的方法直接求解,与其他一些特征提取或特征选择方法(如 SVM)相比,计算量可得到严格控制,有利于缩短离线部分的运行周期,方便训练数据的更新和在线预测模块的升级。

3.1.1 保持类平均向量中分类信息的主成分分析方法

主成分分析方法(PCA)是最经典的线性降维方法之一,其基本思路是:通过 K-L 变换方法找到原问题的 K-L 坐标系,产生矩阵中的最大 d 个特征值对应的特征向量,将其作为变换矩阵对原高维数据进行变换,最终得到具有较好性质的低维数据。

本文使用的 PCA 方法是经典 PCA 方法的变型,将不同的整经轴数作为不同类别,其特征变换矩阵由总类内离散度矩阵 S_w 和经白化变换后的类间离散度矩阵 S_b 共同产生,降维过程中尽量保持类平均向量中的分类信息^[1]。具体步骤如下:

Step1: 原始数据集为 $\{x_i\}$, 整经轴数的类别总

量为 c . 总类内离散度矩阵 S_w 计算如下:

$$\begin{aligned} \mu_j &= E(x_i), \quad x_i \quad j, 1 \quad j \quad c; \\ j &= E[(x_i - \mu_j)(x_i - \mu_j)^T], \\ x_i & \quad j, 1 \quad j \quad c; \\ S_w &= \sum_{j=1}^c P(j) S_j. \end{aligned} \quad (3)$$

其中: j 为类别标签, P_j 为各类的分布概率.

类间离散度矩阵 S_b 计算如下:

$$S_b = \sum_{j=1}^c P(j) (\mu_j - \mu)(\mu_j - \mu)^T, \quad (4)$$

其中 $\mu = E(x_i)$.

Step2: 计算 S_w 的特征值矩阵 和特征向量矩阵 U , 并按特征值大小排序.

$$\begin{aligned} U^T S_w U &= \Lambda, \\ &= \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \lambda_D \end{bmatrix}, \\ U &= (U_1, U_2, \dots, U_D), \\ & \quad 1 \quad 2 \quad \dots \quad D. \end{aligned} \quad (5)$$

可计算前 k 个特征值的累计贡献度, 用于选取低维特征维数 d , 即

$$k = \sum_{i=1}^k \lambda_i / \sum_{i=1}^D \lambda_i, \quad k \leq D. \quad (6)$$

Step3: 对类间离散度矩阵 S_b 进行如下白化变换:

$$B = U^{-1/2}, \quad S_b = B^T S_b B. \quad (7)$$

Step4: 计算 S_b 的最大 d 个特征值对应的特征向量矩阵

$$V = (v_1, \dots, v_d). \quad (8)$$

Step5: 计算低维特征数据

$$y_i^{(PCA)} = (BV)^T x_i \quad R^d, \quad 1 \leq i \leq N. \quad (9)$$

3.1.2 局部线性嵌入方法

局部线性嵌入方法(LLE)是由 Roweis 和 Saul 于 2000 年提出的一种非线性属性降维方法^[6]. 作为一种有效的特征提取手段, 该方法在人脸识别等领域得到了应用^[7,8]. 在本文的研究中, LLE 构建了每个整经轴数样本点与其邻居样本点的重构关系, 在降维过程中尽量保持该重构关系不变, 保留了高维测量空间中的部分重要特征.

LLE 方法的基本步骤如下:

Step1: 原始数据为 $\{x_i\}$, 找到每个数据向量 x_i 的 K 个邻居数据点.

Step2: 计算通过邻居数据点重构 x_i 的最优权重 W_{ij} , 这需求解一个带约束的最小二乘问题

$$(W) = \min_j \left| x_i - \sum_j W_{ij} x_j \right|^2. \quad (10)$$

Step3: 基于 W_{ij} 计算最优的重构低维数据向量 y_i , 即令如下重构误差 (Y) 最小:

$$(Y) = \sum_i \left| y_i - \sum_j W_{ij} y_j \right|^2. \quad (11)$$

计算得到的 y_i 便构成了低维特征数据集 $\{y_i^{(LLE)}\}$.

3.2 神经网络非线性回归器的训练及特征提取方法的选择

本文使用多层前馈神经网络作为非线性回归器. 多层前馈神经网络能以任意精度逼近在紧密集上的任意连续实函数, 可以进行从高维到低维欧氏空间的非线性映射^[9], 因此被广泛应用于非线性回归.

设计两个多层前馈神经网络, 分别处理 3.1 节提出的两种特征提取方法得到的特征数据集. 两个神经网络结构相同: 输入层的单元个数为 d 个, 对应特征提取后的 d 维特征值; 隐层有 1 层, 单元个数为 $d+1$; 输出层的单元个数为 1 个, 输出值对应整经轴数.

按如下原则将每个特征数据集 $\{y_i^{(f)}\}$ 分为训练集和测试集:

1) 将原始样本点按不同的整经轴数进行分类后, 从每个类别中随机选取比例为 p 的样本点作为训练集, 其余样本点作为测试集.

2) 两个训练集所对应的原始样本点是一致的, 测试集亦然.

为了提高神经网络训练收敛速度, 使用 QP(QuickProp) 算法进行训练, 该算法是 Fahlman 于 1988 年提出的, 相关研究表明该方法无论是运行效率还是可推广性都优于标准 BP 算法^[10]. 在所有训练中均限定迭代次数为一个定值 L .

用训练完毕的两个神经网络对测试集进行测试, 针对测试集计算前文所定义的预测误差 $R^{(f)}$. 取 $R^{(f)}$ 最小的特征提取方法作为最优特征提取方法, 并保存其对应的神经网络训练权值, 供算法在线部分在实时预测整经轴数时使用.

4 数值计算和分析

通过数值计算比较本文算法中集成的两种特征提取方法的预测性能, 并与 ν -SVR(ν 支持向量回归) 方法进行比较. 后者使用参数 ν 控制支持向量的数量和误差, 具有较好的非线性回归能力^[11]. ν -SVR 中的参数 ν 按多次试算结果取为 0.5.

选定被考察的时间范围 T 为半年, 从实际生产和工艺历史数据中提取得到了规模 $N = 4138$ 的原始样本集, 其中属性信息 x 的维数 $D = 16$, 包括布长、布宽、花型、纱色、生产工艺等属性. 最小的整经轴数为 4, 最大的整经轴数为 21.

选取降维后的特征数量 $d = 6$, 可计算得到这个规模的特征量在 PCA 算法中特征累计贡献率大于 99.9%, 可保证特征提取过程的能量损失足够小, 且令神经网络的规模适中。

特征提取环节和神经网络训练环节的其他参数设置如下: LLE 的邻居数量 $K = 15$, 训练集占全部样本点的比例 $p = 75\%$, 神经网络 QP 训练算法的迭代次数 $I = 500$ 。为了减小神经网络的初始权值和 QP 算法中的随机参数带来的干扰, 对每个训练集的训练在随机重设神经网络权值后重复 10 遍; 然后选取输出误差最小的网络权值作为训练结果。

PCA/LLE/ ν -SVR 3 种预测方法所对应的预测误差 $R^{(f)}$ 如表 1 所示。使用 LLE 和 PCA 分别进行特征提取后, 进行神经网络训练的预测误差曲线如图 3 所示, 图中横轴 i 为训练代数。

表 1 特征提取方法比较

	ν -SVR	LLE	PCA
预测误差 $R^{(f)}$	1 225.61	1 166.25	668.02
与 PCA 性能比 / %	183.47	174.58	-
绝对误差均值	1.088	1.061	0.803
与 PCA 性能比 / %	135.50	132.13	-
相对误差均值	0.094	0.092	0.068
与 PCA 性能比 / %	138.24	135.29	-

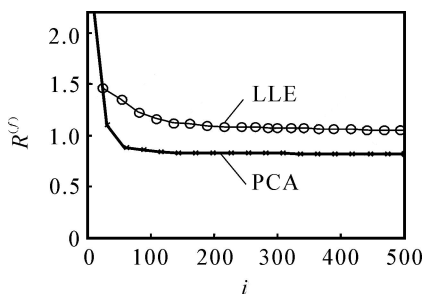


图 3 神经网络训练误差示意图

表 1 中“预测误差”的定义见第 2 节; “绝对误差”定义为神经网络输出的整经轴数预测值与真实整经轴数之差的绝对值; “相对误差”为测试误差与真实整经轴数之比。本预测算法无论使用 PCA/LLE 中的哪种特征提取方法 + 非线性回归器进行预测, 其性能都好于直接使用 ν -SVR 方法, 可见本算法机制有效地滤除了数据噪声, 提高了特征量的数据质量。在本数值实验考察的时间范围内, PCA 方法对应的预测误差比 LLE 方法小。从图 3 可见, PCA 对应的神经网络在训练过程中的收敛速度比 LLE 更快, 从而可知前者提取到的特征量质量更佳。PCA 方法占优说明了对于被考察的数据而言, 预测过程的输入数据与输出数据间的线性关联更为明显。但两种特征提取方法的绝对误差均值仅相差

30% 多, 说明 PCA 方法不占绝对优势, 如 3.1 节的分析, 最佳特征提取方法有可能随企业生产品种的变化而变化。

本文算法已作为某大型纺织企业的生产过程调度系统的组成部分之一, 应用于实际色织生产调度中。

5 结 语

本文提出了一种应用于色织生产调度的整经轴数智能预测模型及算法。该预测模型由离线与在线两部分构成, 集成了 PCA 和 LLE 两类特征提取方法, 包括特征提取方法比较、非线性回归器训练等关键环节, 能在有效控制计算量的前提下提高整经轴数预测精度。数值计算表明, 该方法能满足实际生产过程整经轴数预测需要, 可有效地为生产调度过程服务。

参考文献(References)

- [1] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2000.
(Bian Zhao-qi, Zhang Xue-gong. Pattern recognition [M]. Beijing: Tsinghua University Press, 2000.)
- [2] Sklansky J, Wassel G N. Pattern classifiers and trainable machines[M]. New York: Springer-Verlag, 1981.
- [3] Duda R O, Hart P E. Pattern classification and scene analysis[M]. New York: John Wiley & Sons, 1973.
- [4] 孙跃鹏, 刘民, 吴澄. 面向色织生产过程的整经轴数智能预测算法[J]. 控制工程, 2005, 12(6): 523-526.
(Sun Yue-peng, Liu Min, Wu Cheng. Intelligent prediction algorithm of trim beam number in colored textile manufacturing process[J]. Control Engineering of China, 2005, 12(6): 523-526.)
- [5] Vapnik V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995.
- [6] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290(5500): 2323-2326.
- [7] Hadid A, Kouropteva O, Pietikainen M. Unsupervised learning using locally linear embedding: Experiments with face pose analysis[C]. ICPR '02. Quebec, 2002, V1: 10111.
- [8] Saul L K, Roweis S T. Think globally, fit locally: Unsupervised learning of low dimensional manifolds[J]. J of Machine Learning Research, 2004, 4(2): 119-155.
- [9] Barron A R. Statistical properties of artificial neural networks[C]. Proc 28th Conf on Decision and Control. Tampa, 1989: 280-285.

(下转第 1389 页)

式中: $e_i = y_i - \hat{y}_i$, \hat{y}_i 和 y_i 分别为识别输出值和期望输出值。

BP 网络识别模型采用 5-13-1 的网络结构,学习因子取为 0.05,隐含层和输出层的传输函数均选取 Tansig 函数 ($f(x) = 2 / (1 + e^{-2x}) - 1$), 经典 SVM 模型也采用径向基核函数,识别结果见表 3。

表 3 识别结果比较

识别模型	RMSE	MAXPE	MAXNE	识别率 / %
BP 网络	0.267 9	1	- 0.997 1	90.6
经典 SVM	0.251 2	1	- 1	92.3
改进 LS-SVM	0.119 0	0.836 2	- 0.803 6	98.8

由表 3 可知,本文提出的改进 LS-SVM 识别模型在识别率等性能指标上明显优于 BP 模型和经典 SVM 模型,在识别精度、泛化能力上优于 BP 网络模型与 SVM 识别模型。此外,本文方法在油层识别率上未达到 100%,即有误判点,这与样本的选取不完备或构建的模型的泛化能力有限等因素有关。总之,本文提出的基于改进 PSO 的 LS-SVM 识别模型是切实可行且行之有效的,不但有效地避免了 BP 模型中存在的过学习、维数灾难、局部极小等问题,而且解决了 SVM 模型对于大规模样本所出现的计算复杂度高、计算速度慢等问题。

5 结 论

基于结构风险最小化 (SRM) 准则的 LS-SVM 是一种良好的分类方法。本文提出的基于改进 PSO 迭代优化的 LS-SVM 算法不仅避免了矩阵求逆计算,而且总能求得最优解,其训练速度和求解精度均得到提高。实际应用表明,本文提出的基于改进 PSO 的 LS-SVM 识别模型适用于解决油层识别等大规模样本问题,效果显著,具有很好的应用前景。

参考文献 (References)

[1] 《测井学》编写组. 测井学 [M]. 北京: 石油工业出版社, 1998.
(《Well logging》Compile Group. Well logging [M]. Beijing: Petroleum Industry Press, 1998.)

[2] Vladimir N Vapnik. The nature of statistical learning theory [M]. New York: Springer, 2000.

[3] Senf Alexander, Chen Xue-wen, Zhang Anne. Comparison of one-class SVM and two-class SVM for fold recognition [C]. Neural Information Processing — 13th Int Conf ICONIP 2006, Proc. Hong Kong, 2006: 140-149.

[4] Shih Frank Y, Zhang Kai. Support vector machine networks for multi-class classification [J]. Int J of Pattern Recognition and Artificial Intelligence, 2005, 19 (6): 775-786.

[5] Cortes C, Vapnik V. Support vector machine [J]. Machine Learning, 1995, 20(3): 273-297.

[6] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers [J]. Neural Processing Letters, 1999, 9(3): 293-300.

[7] Kennedy J, Eberhart R C. Particle swarm optimization [C]. Proc IEEE Int Conf Neural Networks. Piscataway: IEEE Press, 1995: 1942-1948.

[8] Shi Y, Eberhart R. A modified particle swarm optimizer [C]. IEEE World Congress on Computational Intelligence. Anchorage, 1998: 69-73.

[9] Clerc M. The swarm and the queen: Towards a deterministic and adaptive particle swarm optimization [C]. Proc of the Congress of Evolutionary Computation. Washington, 1999: 1951-1957.

[10] 焦荣昌. 线性方程组的正则化解法及其应用 [J]. 物探与化探, 1994, 18(3): 200-207.
(Jiao Rong-chang. The regularized solution of linear equations set and its application [J]. Geophysical and Geochemical Exploration, 1994, 18(3): 200-207.)

[11] 夏克文, 刘明霄, 张志伟, 等. 基于属性相似度的属性约简算法 [J]. 河北工业大学学报, 2005, 34(4): 20-23.
(Xia Ke-wen, Liu Ming-xiao, Zhang Zhi-wei, et al. An approach to attribute reduction based on attribute similarity [J]. J of Hebei University of Technology, 2005, 34(4): 20-23.)

(上接第 1380 页)

[10] Fahlman S E. An empirical study of learning speed in back-propagation networks [R]. CMU-CS-88-162, Carnegie-Mellon University, 1988.

[11] Scholkopf B, Smola A, Williamson R C, et al. New support vector algorithms [J]. Neural Computation, 2000, 12(5): 1207-1245.