

主元分析-径向基函数神经网络 软测量方法研究

周平 冯少辉 钱锋

(华东理工大学自动化研究所, 上海 200237)

摘要:主元分析方法是基于多元统计分析的数据处理方法,径向基函数(RBF)网络能很好地用于非线性系统建模。针对丙烯腈聚合物水含量这一腈纶生产中的关键指标无法在现场进行在线检测的情况,运用主元分析方法对输入变量进行主元分析以达到降维的目的,再通过 RBF 神经网络建立非线性软测量模型。仿真结果表明,该模型具有较高的预测精度。

关键词:软测量;主元分析法;RBF 神经网络;丙烯腈聚合物;水含量

中图分类号:TP274

文献标识码:A

文章编号:0253-4320(2004)S2-0084-03

Study on soft measurement based on PCA-RBF neural network

ZHOU Ping, FENG Shao-hui, QIAN Feng

(Research Institute of Automation, East China University of Science and Technology, Shanghai 200237, China)

Abstract: Principal component analysis(PCA) is an effective multivariate statistical method for data processing, radial basis function (RBF) neural network is very useful in nonlinear system modeling. The water content of acrylonitrile polymer is one of the key quality target in the acrylonitrile production, however, it can't be measured on-line in plant presently. Considering this condition, the PCA method is applied to analyze the input variables, then the RBF neural network model is studied. The simulation results show this model is effective.

Key words: soft measurement; principal component analysis; RBF neural network; acrylonitrile polymer; water content

在过程控制中常有这样的情况:关键的被控变量无法在线测量,或者虽然能测量,但测量滞后很大而难以控制。软测量技术的基本思想是选择一组既与主导变量有密切联系又容易测量的变量,称为辅助变量,通过构造某种数学关系,实现对主导变量的估计。

主元分析(principal component analysis, PCA)^[1]是一种数据压缩和特征提取的多变量统计分析技术,它的最终目的是用少数几个不相关的变量获取反映原始变量所提供的绝大部分信息。前馈神经网络在选择恰当的网络层次与隐含层节点数的条件下,可以逼近任意精度的非线性函数,因而被广泛地应用于这些过程的建模和控制^[2],反向传播(back propagation, BP)网络和径向基函数(radial basis function, RBF)网络是目前工业中常采用的 2 种网络,其中 BP 网络常采用 S(Sigmoid)型函数 BP 算法, BP 网络的明显缺点是网络的结构难以确定,收敛速度慢,且参数估计必须基于非线性优化技术,在运用梯度下降法学习时容易陷入局部极小,其他改进的 BP 网络,例如基于遗传算法的 BP 神经网络、模拟退火法等,

有时能满足全局最优解,但计算量很大。为了克服局部极小,并提高算法的收敛速度,采用 RBF 网络能够达到较好的效果^[3]。

1 PCA-RBF 神经网络软测量模型

1.1 模型结构

PCA-RBF 神经网络模型由 PCA 和 RBF 神经网络 2 部分组成。其中 PCA 实施对初选的 n 个输入变量 $X = \{x_1, x_2, \dots, x_n\}$ 进行 PCA 主成分提取,得到 k 个线性独立变量 t_1, t_2, \dots, t_k ,在不过多损失输入变量信息的前提下,消除输入变量间的复共线性,减少神经网络的输入变量个数,简化模型结构。PCA-RBF 神经网络模型结构见图 1。

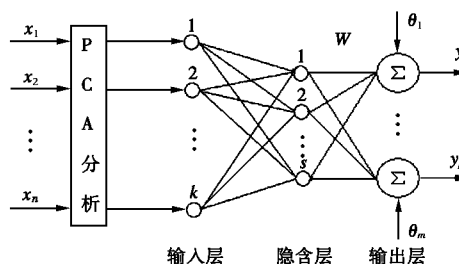


图 1 PCA-RBF 神经网络模型结构

RBF神经网络的产生具有很强的生物学背景。在人的大脑皮层区域中,局部调节及交叠的感受野(receptive field)是人脑反应的特点,基于感受野这一特性,Moody和Darken提出了一种神经网络结构^[4],即RBF网络。它是一种前向网络,隐含层的单元是感受野单元,每个感受野单元输出为:

$$\omega_i = R_i(\mathbf{X}) = R_i\left(\frac{\|\mathbf{X} - \mathbf{c}_i\|}{\sigma_i}\right), i = 1, 2, \dots, s \quad (1)$$

式(1)中, \mathbf{X} 是 N 维输入向量, \mathbf{c}_i 是与 \mathbf{X} 同维数的向量, σ_i 是第 i 个感知器变量(可以自由选择参数),它决定了该基函数的围绕中心的宽度, s 是感知单元的个数, $\|\mathbf{X} - \mathbf{c}_i\|$ 表示 \mathbf{X} 和 \mathbf{c}_i 之间的距离。 $R_i(\cdot)$ 具有局部感受的特点,例如 $R_i(\cdot)$ 取高斯函数,即 $R_i(\cdot) = \exp(-\|\mathbf{X} - \mathbf{c}_i\|^2/\sigma_i^2)$, $R_i(\cdot)$ 只有在 \mathbf{c}_i 周围的一部分区域内有较强的反应,这正体现了大脑皮质层的反应特点^[5]。输入层实现从 $\mathbf{x} \rightarrow R_i(\mathbf{X})$ 的非线性映射,输出层实现从 $R_i(\mathbf{X}) \rightarrow \mathbf{y}$ 的线性映射,即:

$$y_j = \sum_{i=1}^s W_{ij} R_i(\mathbf{X}) + \theta_j, j = 1, 2, \dots, m \quad (2)$$

式(2)中, W_{ij} 是第 i 个隐含层单元到第 j 个输出层单元的联接权值, θ_j 是第 j 个输出层节点的阈值。

1.2 RBF神经网络训练原理

RBF网络的训练过程分为径向基函数中心的选取和权值的求取。

1.2.1 径向基函数中心的选取

Moody和Darken于1989年建议在RBF中心值的选取问题中采用 k 均值聚类算法^[6],本文也采用该方法确定径向基函数中心。它是一种无监督学习算法,具体过程可描述如下:首先将隐含层单元激励函数中心 \mathbf{c}_i 的初值置为最初的 i 个训练样本,然后将所有的训练样本按最近的聚类中心分组,如果 x_i 与 \mathbf{c}_i 满足 $\min_i \|x_i - \mathbf{c}_i\|$,则 x_i 属于 \mathbf{c}_i 的集合 C ,最后计算聚类中心 $\mathbf{c}_i = \frac{1}{M_i} \sum_{x_i \in C} x_i$,其中 M_i 表示属于 C 的元素个数。

1.2.2 权值的求取

求解权值现在较常用的为正交最小二乘法(orthogonal least squares, OLS)算法和Givens最小二乘法。Givens最小二乘法运算的存储量大,运算速度慢,对实时运行有一定影响,OLS算法的优点是简单易行,运算速度快,故本文采用OLS算法。单输出RBF网络的OLS算法学习过程如下^[7],对于具有 N 对学习样本的RBF网络,其输入输出方程为:

$$\mathbf{d} = \mathbf{p}\Theta + \mathbf{E} \quad (3)$$

式中 $\mathbf{d} = [d(1), \dots, d(N)]^T$ 为期望输出, $\mathbf{p} = [p_1, \dots, p_N]$ 是隐含层单元输出阵, $\mathbf{p}_i = [p_i(1), \dots, p_i(N)]^T$, $\Theta = [\theta_1, \dots, \theta_N]^T$ 为输出权值矢量, $\mathbf{E} = [\epsilon(1), \dots, \epsilon(N)]^T$ 是误差矢量。OLS算法的过程是将 \mathbf{p}_i 集合变换成正交集,即 $\mathbf{p} = \mathbf{W}\mathbf{A}$, \mathbf{A} 是对角线为1的上三角阵, \mathbf{W} 是包含正交矢量 \mathbf{w}_i 的矩阵。因此有方程 $\mathbf{d} = \mathbf{W}\mathbf{g} + \mathbf{E}$, $\mathbf{A}\Theta = \mathbf{g}$,由于 \mathbf{A} 和 \mathbf{g} 可以在正交化的过程中得到,因此 Θ 也易求解出来。OLS算法在每步正交过程中,都要用“信息-贡献”准则进行正交优选,当OLS算法满足一定精度后,算法即终止,这时的正交矢量数目就是隐含层神经元数目。

2 干法腈纶生产工艺

丙烯腈生产工艺流程如图2所示。

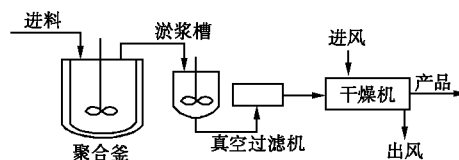


图2 丙烯腈工艺流程简图

腈纶装置聚合单元生产首先将共聚单体(丙烯酸甲酯、丙烯腈、回收单体和二氧化硫)、苯乙烯磺酸钠、氢氧化钠、催化剂、铁和脱盐水等物料按一定比例加入聚合釜,在 $55 \sim 60^\circ\text{C}$, pH值为 $2.5 \sim 3.0$ 的条件下发生聚合反应。反应生成的浆料经过转鼓式真空过滤机将聚合物分离出来,再制成固含量大约26%的浆料,送到淤浆混合系统。混合之后的浆料送入挤出机,从挤出机出来的聚合物下落到一个振动器上,振动器将这些聚合物在干燥机链板上铺成 $5 \sim 10\text{ cm}$ 厚的均匀层。干燥机共有4个区,空气在干燥机每个区的入口端进入并经蒸汽盘管加热,利用鼓风机使空气通过聚合物床层,再穿过有缝隙的链板循环,热空气带走聚合物中的水分。热空气再经过蒸汽盘管重新加热,经再加热后的热空气再通过聚合物床层对下一区进行干燥。这个过程沿干燥机的长度不断重复,空气流从聚合物入口处进入,沿干燥机的长度呈螺旋形前进,最终使其水含量由55%左右降低为 $0.1\% \sim 0.4\%$ 。

聚合物干燥过程是腈纶装置的核心之一,其操作情况稳定与否将直接影响最终产品的质量,因此在聚合物干燥生产中要求聚合物水含量稳定。聚合物水含量过高,对后续的一些工序如纺丝等有不利影响,而且聚合物风送系统也会被卡住;聚合物水含

量过低,一部分聚合物就会发生降解,使二甲基甲酰胺/聚合物溶液颜色变深,影响最终产品质量。但干燥过程中聚合物水含量这一关键指标无法在现场进行在线检测,只能借助于 4 小时一次的人工分析进行调控。干燥过程中聚合物水含量的波动已成为影响腈纶装置正常运行的主要“瓶颈”之一。因此及时、准确地测量聚合物水含量是提高丙烯腈聚合过程产品质量的关键技术。

3 PCA-RBF 神经网络模型仿真研究

3.1 输入变量初选

影响聚合物水含量的因素较多,经过对现场工艺的调查和对机理的定性分析,初步确定了作为软测量模型输入变量的参数分别是干燥机的 4 个干燥区热空气在链板上下的温差、待干燥聚合物的固含量以及相邻干燥区温差的变化值,即聚合物水含量与以上 8 个变量可以表示为如下非线性关系:

$$y_w = f(\Delta T_i, p, \delta t_j) \quad i = 1, 2, 3, 4; j = 1, 2, 3 \quad (4)$$

式(4)中 y_w 为聚合物水含量, ΔT_i 为干燥区的温差, p 为待干燥聚合物的固含量, δt_j 为相邻干燥区温差的变化值。

3.2 仿真实验

对包含以上 8 个变量的数据进行主元分析,取累计方差贡献率 $\eta_k \geq 85\%$ 的 5 个主元成分,即 $k = 5$,分别为 4 个区的温差和待干燥聚合物的固含量。

样本的选择过程为:根据样本分布情况,应用聚类分析方法对其进行分类,然后按照相同的比率从每一类样本中随机抽取训练样本和检验样本。这样样本就具有广泛的代表性,各区域内样本的映射关系都能被网络较好地学习,同时在检验过程中也能取得较好的效果。依据以上原理从现场 DCS 数据取 120 组样本,其中 80 组为训练样本,其余 40 组为检验样本。

将上述 PCA-RBF 神经网络模型应用于丙烯腈干燥过程水含量的仿真研究,本仿真实验隐含层节点数确定为 20 个。在网络训练之前,为了确保各变量在分析中的地位相同,对每一组输入输出数据进行如下归一化处理。

$$\text{输入: } x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (i = 1, 2, 3, 4, 5)$$

$$\text{输出: } y' = \frac{y - y_{\min}}{y_{\max} - y_{\min}}$$

式中 x'_i 表示归一化后的第 i 个输入样本数据, y' 表示归一化后的输出样本数据。训练结果如图 3

所示,检验结果如图 4 所示。可以看出,它们均取得了较好的拟合效果。表 1 给出了丙烯腈干燥过程水含量的训练样本和检验样本的误差值,检验样本最大绝对误差小于 0.02%,满足工艺要求,从而说明该模型能够较好地满足丙烯腈干燥过程水含量的软测量精度要求。

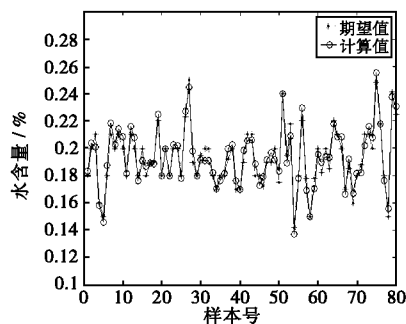


图 3 数据训练结果图

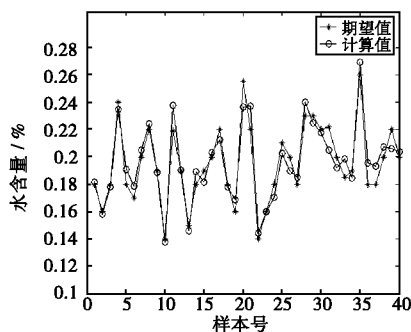


图 4 数据检验结果图

表 1 丙烯腈干燥过程水含量神经网络训练和检验误差值

	最大绝对 误差 / %	平均绝对 误差 / %	最大相对 误差 / %	平均相对 误差 / %
训练样本	0.0094	0.0049	4.9249	2.5422
检验样本	0.0185	0.0075	8.8129	3.7547

4 结语

以丙烯腈生产过程中聚合物水含量为研究对象,依据工艺机理和经验知识对过程变量进行初选,为了消除变量间的相关性、降低维数以及减少在 DCS 实现时的系统资源消耗,运用主元分析方法对输入变量进行主元分析,进一步压缩输入变量的个数,并消除变量间的复共线性,简化了软测量模型结构,再运用径向基函数神经网络,对提取出 PCA 主元后的变量与聚合物干燥过程水含量进行非线性建模。仿真结果表明:PCA-RBF 神经网络模型具有较高的精度和较强的泛化能力,且训练速度很快,可很

(下转第 88 页)

的基础,在欧美天然气管道输送行业充分发展的地区,这些数据甚至是决定一个企业能否长期生存的根本,将被直接用来进行每天甚至每小时的交易。在当今网上贸易盛行的时代,飞速发展的国内天然气行业也正在积极探索管理新思路以满足市场化的需求。显然,管理和工业监控应用的一体化是上至一个国家、地区的市场机制完善,下至一个企业的管理发展到一定阶段对信息技术应用提出的一项必然需求,是安全生产、实现利润最大化的基础。SCADA 系统不再是管道运营管理信息技术的全部,而只是整个应用的一个重要部分。

1.2 业务外延的拓展

华油天然气现有 100 名员工,在管理几十亿资产的同时,每年都在持续开展基础建设工作,目前正在进行陕甘宁至北京第二条管道(简称“陕京二线”)的建设。其业务在 4 省 2 市之间呈线性分布,大多数管理和技术人员都频繁地往返于 2 000 多 km 生产现场和施工现场之间,不同岗位角色的办公地点不再固定,客户端的环境和配置不再重要,一个信息孤岛中的局部资源已不足以支持业务的开展,要求跨平台、跨办公室、跨区间和移动办公。

1.3 应用集成的迫切性和必要性

不断的信息交流成为管理趋势,运营管理系统与 SCADA 系统之间不断方便、快捷而有效地进行数据交换成了迫切需要。每位员工也许都需要所有应用系统的某一部分数据,不同之处只是每个岗位需要掌握的信息重点不一样。一方面管理需要 SCADA 系统的数据,另一方面,作为原 SCADA 系统的主要使用者——调度管理人员,在直接指挥生产活动时,需要及时掌握其他业务系统的信息,以便及时准确地实施复杂管网的生产调配和安全控制。单靠监控 SCADA 数据,而没有管理信息系统的支持,已经很难进行现场作业的全面指挥。例如,当某台关键设施发生故障导致输气量变化时,调度管理人员可以通过与之关联的 GIS 系统查到现场的地理信息,同时通过 EAM 系统了解到设备的维护办法、周期及

负责人等情况,从而在最短时间内形成一个成熟而正确的调度解决方案,不再需要派遣维护人员立即赶赴现场进行现场勘查、调研和口头报告。

2 应用集成现状和存在的问题

2.1 应用现状

华油天然气原 SCADA 系统与管线于 1997 年同步建成,采用客户/服务器(C/S)模式,大量管理数据的获取或报表的生成通过开放数据库互联(ODBC)自动实现,一台位于 SCADA 局域网的个人计算机专门负责定时从 SCADA 数据库中提取每日所需的实时数据和历史数据,并自动生成 Excel 表格。

业务管理系统大都在近几年陆续建成,从维护和移动应用方便等角度考虑,普遍采用了浏览器/服务器(B/S)模式。多个业务应用系统之间,如 EAM(美国 MRO 公司的 Maximo 5.1)、OA(Lotus Notes)、企业信息门户(Microsoft)和移动短信平台之间已经通过标准开放的 API 接口或嵌套等方式实现共享。在与 SCADA 系统进行实时数据交换的过程中,例如, EAM 系统一方面提供开放的标准接口;另一方面采取了如下的临时解决方案获得主要的 SCADA 数据:利用浏览器获取前述个人计算机定时生成的 Excel 表格中的数据(这当然需要 SCADA 一端采用手工或自动定时的方式将数据发布出来),再将该表中的数据存放到 EAM 自身的数据库中,以便于 EAM 或其他信息系统查询检索或为报表生成器产生其他报表时用。

2.2 存在的问题

在强调安全生产的大前提下,SCADA 系统是生产控制与管理的基础,技术上应尽量避免潜在的安全漏洞和威胁。显然,在所述应用接口条件下,数据安全和方便高效是一对明显的矛盾,缺乏可靠保障。在没有足够的网络安全防范机制下,按照传统做法,一个企业往往会将 SCADA 与管理办公网进行物理隔离,一般不将 SCADA 所在的局域网直接对企业办公网开放。但是,如果不直接连接,系统应用之间将

(上接第 86 页)

好地应用于丙烯腈干燥过程聚合物水含量的软测量,取得了较好的效果。

参考文献

- [1] 张尧庭,方开泰.多元统计分析引论[M].北京:科学出版社,1982.
- [2] 丛爽.面向 MATLAB 工具箱的神经网络理论与应用[M].合肥:中国科学技术大学出版社,1998.

- [3] 薄翠梅,张湜,林锦国,等.[J].石油化工高等学校学报,2002,15(3):57-60.
- [4] Moody J, Darken C. Learning with localized receptive fields[A]. In: Touretzky D, Hinton G, Sejnowski T. Proc 1998 Connectionist Models Summer School[C]. San Francisco: Morgan Kaufmann Publishers, 1988.
- [5] 王旭东,邵惠鹤.[J].信息与控制,1997,26(4):272-284.
- [6] Moody J, Darken C.[J]. Neural Computation, 1989, (1):281-294.
- [7] Chen S, Cowan C F N, Grant P M.[J]. IEEE Trans on Neural Networks, 1991,2(2):302-309. ■