基于内部非线性映射模型PLS算法的比较研究

宋斯男[1] 师佳[2]
(1.厦门大学信息科学与技术学院自动控制技术研究所; 2.厦门大学化学化工学院化学工程与生物工程系 厦门 361005)

摘 要:本文从线性偏最小二乘(PLS)算法出发对基于内部非线性映射模型的PLS算法的思想本质进行深入的剖析和探讨。结合拟合非线性过程的实际问题,从拟合精度、计算复杂度两方面对基于二次函数和神经网络作为内部模型的PLS方法以及基于误差反馈调整的PLS算法进行了比较,在此基础上对不同的非线性PLS算法的应用提出了若干指导性建议。

关键词:主成分分析 PCA 偏最小二乘 PLS 神经网络
中图分类号:0657.3 **文献标识码**:A **文章编号**:1674-098X(2010)03(a)-0004-02

偏最小二乘方法是集多种技术于一身的一种多元统计方法,这些技术包括最小二乘回归分析、主成分分析(PCA)、典型相关分析(ICA)[1],因此该方法具备了这些方法的诸多特性。与多元线性回归相比,偏最小二乘方法能够有效的克服自变量间的多重相关性对系统建模的不良影响,在多变量、小样本的建模问题上表现良好。但由于线性PLS算法并不能很好的应对实际系统普遍存在的非线性特性,因此近年来在线性PLS算法的基础上提出了很多非线性PLS算法。

本文所研究的是以非线性映射模型作为内部模型的一类非线性PLS算法,通过对这类算法的数学描述和算法流程的剖析。揭示了这类非线性PLS算法的实质,并将这类算法中三个比较有代表性的算法用于非线性过程的拟合,从计算精度、运算速度方面进行了比较,在比较结果的基础上对这些非线性PLS算法的应用提出了一些指导性建议。

1 基于非线性内部映射的非线性PLS算法

为了揭示内部基于非线性映射模型PLS算法的本质,首先回顾线性PLS算法的核心思想

1.1 线性PLS算法的核心思想

线性PLS算法建立的线性模型分为外部模型和内部模型,两部分如下描述:

$$\text{外部模型:}\begin{cases} X = TP^T + E & (1) \\ Y = U\,Q^T + F & (2) \end{cases}$$

$$\text{内部模型:}\ U = TB \qquad (3)$$

外部模型分别给出了输入变量 X 和输出变量 Y 的PCA模型,其中,T 和 U 分别是输入变量和输出变量的主元矩阵,P 和 Q 分别是载荷矩阵。内部模型则描述了输入变量和输出变量主元之间的线性回归关系。B 为回归的系数矩阵。E、F为残差矩阵。从上述描述可以看出PLS算法的特点在于通过提取输入输出变量的主元,使他们之间的协方差达到最大,让Y能够被尽量少的PLS成分线性的充分解释,这里,构成矩阵 T 的向量相互正交,消除了自变量间的线性相关性。

1.2 基于非线内部性映射的PLS算法

为了描述实际过程或系统中不可避免的非线性特点,必须在PLS算法中建立非线性模型,所谓基于内部模型非线性映射的基本思路是保持线性PLS算法框架不变,只是在建立输入和输出主元(t_n 和 u_n)之间的映射关系时,不采用模型回归而采用非线性映射,即 $U = f(T)$。

为提高非线性PLS的拟合精度,对以二次多项式和神经网络作为非线性内部映射的PLS算法进行改进,改进的主要思路是利用主元的预测误差来对自变量权值 w 进行校正,实现以内部模型的建模误差来动态地指导主元的提取,算法中最为核心的问题是如何根据拟合误差 e 来确定每一步 w 的修正量 Δw。该问题在数学上等价于如下优化问题:

$$\min_{\Delta w} J_k = u_k - f(X(w_k + \Delta w_k)) \qquad (4)$$

$$f(X(w_k + \Delta w_k)) \approx f(t_k) + e = f(Xw_k) + e \qquad (5)$$

为了求解上述优化问题,将非线性函数 $f(Xw)$ 在 w_k 处做Newton-Raphson线性化处理,有:

$$f(X(w_k + \Delta w_k)) = f(t_k) + \frac{\partial f}{\partial w}|_{w=w_k}\Delta w = f(Xw_k) + \frac{\partial f}{\partial w}|_{w=w_k}\Delta w \qquad (6)$$

比较(5)(6)式可得:$u_k - f(t_k) = \frac{\partial f}{\partial w}|_{w=w_k}\Delta w$

这里 u_k 是实际的输出主元,由前一步迭代确定,即:$u_k = F_{k-1}q$

映射函数 f 已知(二次多项式或神经网络),故 $f(Xw_k)$ 和 $\frac{\partial f}{\partial w}$ 也可计算,令

$$\frac{df}{dw}|_{w=w_k} = Z_k$$

$$e_k = u_k - f(Xw_k)$$

则优化问题(4)简化为确定 Δw 使得下式成立:$e_k \approx Z\Delta w$

利用二次线性回归技术可知每一步最优的调整量为 $\Delta w_k = (Z^TZ)^{-1}Z^Te_k$ (6)

至此,可以得到基于误差反馈校正的迭代非线性PLS算法流程如下:

Step1:将输入输出数据矩阵 X、Y 做标准化处理得到 E_0、F_0,令 $E = E_0$、$F = F_0$。

Step2:取 F 的一列作为输出主元u的初值。

Step3:计算提取输入主元的权重向量并将其单位化:

$$w^T = \frac{u^TE}{u^Tu},\ w = \frac{w}{\|w\|}$$

Step4:计算输入主元:$t = \frac{Ew}{w^Tw}$

Step5:用 u 和 t 拟合非线性映射 f,得到预测输出 $u = f(t)$

Step6:计算 Y 上的载荷向量,并将其单位化:$q^T = \frac{t^TF}{t^Tt}$, $q = \frac{q}{\|q\|}$

Step7:$u = \frac{Yq}{q^Tq}$

Step8:更新 w,并将其单位化,:

$$\Delta w = (Z^TZ)^{-1}Z^Te = (Z^TZ)^{-1}Z^T(Fq - f(t)),$$

$$w = w + \Delta w,\ w = \frac{w}{\|w\|}$$

Step9:计算新的主元:$t = \frac{Ew}{w^Tw}$

Step10:检验 t 是否收敛,如果不收敛则返回Step3,如果收敛则

表1 不同算法的计算时间与拟合误差统计数据

	NN-PLS	EB-Q-PLS	EB-NN-PLS	BP-NN
主元个数	4	2	1	–
均方误差	0.992	0.409	0.168	0.199
运算时间 (s)	1.326	0.035	4.128	2.677

向下进行

Step11:计算 X 上的载荷向量: $p^T = \frac{t^T E}{t^T t}$

Step12:计算输入输出残差矩阵: $E = E - t p^T$, $F = F - f(t) q^T$

Step13:如果精度不满足要求则重复Step2

2 非线性PLS算法的探讨

对于内部模型映射是神经网络的PLS算法来说,算法本质上构成了一种新型的神经网络,与一般的神经网络建模相比,该神经网络的输入层和输出层都利用了主成份分析技术对输入输出数据进行了压缩,从而减小了内部神经网络建模的复杂度。

非基于误差的算法将新加入的两个线性层与神经网络自身的割裂开来,使神经网络本身强大的非线性能力被外部不可变化的线性层所束缚,降低了非线性拟合能力。而基于误差反馈的PLS算法把PLS算法的诸多优秀特性与神经网络非线性能力充分结合起来的同时,将内部模型的建立和提取主元的过程联系起来,提高了内部模型与外部模型的匹配程度,从而提升了神经的非线性拟合能力。

以二次多项式和神经网络作为内部模型映射,先后得到两种改进的非线性PLS算法:分别为EB-Q-PLS和EB-NN-PLS。本文接下来将NN-PLS算法与以上两种算法以及基于BP算法的神经网络用于一个非线性过程的拟合问题,通过对拟合精度和计算复杂度的比较,分析不同的非线性PLS算法的效能,并由此给出一些应用这些算法的指导性建议。

3 非线性PLS算法的效能比较

3.1 拟合非线性差分方程

分别用NN-PLS、EB-Q-PLS、EB-NN-PLS以及BP神经网络四种算法对由以下的非线性差分方程描述的动态过程进行建模:

$$y(k) = e^{2x_1(k-2)\sin(p\,y(k-1))} + \sin(x_2(k-1)y(k-2)) + \cos(4x_1(k-1)x_2(k-2))$$

输入 x_1、x_2 分别为区间[-0.25,0.25]上的随机数,利用上述差分模型得到600组输入输出数据,其中500组用于建模,100组用于验证。仿真结果如表1。

由从表1的数据可知,尽管EB-Q-PLS和EB-NN-PLS分别只提取了2个和1个主元,但拟合误差却小于提取了4个主元的NN-PLS算法。尤其是EB-NN-PLS算法,拟合误差要远远小于同样采用神经网络作为内部映射的NN-PLS算法,因此,在拟合精度方面,采用误差反馈机制的算法要好于没有采用误差反馈机制的算法。

从表中还可以看出,EB-Q-PLS算法耗时最少,这主要是由于其内部非线性映射为二次多项式,其算法的复杂程度要远远小于神经网络,并且上面提到,EB-Q-PLS提取的主元数目比NN-PLS少,但拟合误差却比NN-PLS要小,因此,从运算效率、误差数据上来看,EB-Q-PLS都要好于NN-PLS算法。

BP神经网络算法具备很好的非线性拟合能力,从表中可以看出,该算法的拟合精度较好,因此,采用BP神经网络作为内部非线性模型映射的BP-NN-PLS算法在拟合精度的表现上要好于以二次函数作为内部映射的EB-Q-PLS,但对神经网络的训练却使算法大为耗时。

同样采用神经网络,NN-PLS每提取一个主元只需训练一个SISO网络,与之相比,EB-NN-PLS采用迭代的的方式提取主元,每迭代一次就要训练一个SISO网络并且每一次迭代的次数并不能控制,而BP-NN则要训练一个MIMO网络,因此,NN-PLS算法训练网络的方式是三种算法中最简单的,从表中的数据来看,耗时是三种算法中最少的。

3.2 建议

基于上述的比较结果,在应用何种非线性PLS算法来拟合非线性模型或过程的问题上给出如下建议:

(1)基于误差反馈的PLS算法在非线性拟合能力方面要好于没有基于误差反馈的PLS算法,因此,在使用PLS的解决非线性问题的时候建议首先考虑使用基于误差反馈的PLS算法。如EB-Q-PLS、EB-NN-PLS。

(2)EB-Q-PLS算法结构简单、实施方便,运算速度较快,但由于采用二次函数作为内部非线性映射,对非线性拟合能力有一定限制,建议在输入输出变量数目较多但非线性程度中等偏弱的系统建模问题中使用,以达到比NN-PLS各方面更好的效果。同时该算法也是进一步发展在线建模技术的首选算法。

(3)EB-NN-PLS算法既具备PLS算法的优点,又能将神经网络的非线性拟合能力强的特点充分发挥出来,与传统的BP-NN网络相比,可以去除变量间的相关性、并且需要训练的网络参数较少,在处理大量数据的时候更不容易产生过拟合的问题,是目前一种非常先进的PLS算法,因此,在更为普遍的多输入、多输出的高度非线性系统建模问题中,建议采用该方法,以建立更为精确、泛化能力更好的数据模型。

4 总结与展望

本文从线性PLS算法出发,对几种基于内部非线性映射模型的PLS算法的本质进行了剖析。同时结合非线性过程的拟合问题对几种算法的效能进行了比较,并在此基础上对非线性PLS算法的应用给出了指导性建议。

与其他基于数据驱动的建模算法一样,非线性PLS算法也属于后验方法,同样面临着算法如何与先验的机理模型相结合来有效提高建模精度的问题。这是非线性PLS的一个研究方向。另外,考虑到实际过程可能存在的参数时变性,用本文提到非线性PLS算法不适应这个问题,目前已经提出了一些能够根据系统的变化自适应地调整模型的在线PLS算法,如Qin等人提出的分块递推式PLS算法[9],但这些在线算法均是基于线性PLS算法,如何将本文提到的非线性建模思路与这些方法的自适应思想相结合也是一个值得研究的问题。

参考文献

[1] Johnson R.A.Wichern D.W 实用多元统计分析[M].2001.
[2] 王惠文.偏最小二乘回归方法及其应用[M].1999.
[3] Hoskuldsson A, PLS regression methods[J],Chemometrics, 1988.
[4] Wold, S.,Kettaneh-Wold,N.,& Skagerberg,B. Non-linear PLS modelling. Chemometrics and International Laboratory Systems 1989.
[5] Qin S.J,McAVoy T.J.Nonlinear PLS Modeling Using NeuralNetworks[J].Computer&Chemical Engineering,1992.
[6] G.Baffi.E.B. Martin. A.J. Morris Non-linear projection to latent structures revisited: the quadratic PLS algorithm, Computer&Chemical engineering 1999.
[7] G.Baffi.E.B. Martin. A.J. Morris Non-linear projection to latent structures revisited (the neural network PLS algorithm), Computer&Chemical engineering 1999.
[8] Liang J,Qian J X. Multivariate statistical process monitioring and control:Recent development s and applications to chamical industry, 2003.
[9] Recursive PLS algorithms for adaptive data modeling[J]. Computer&Chemical Engneering 1998.