

基于 PLS 的 Elman 神经网络算法研究

丁世飞^{1,2}, 贾伟宽¹, 许新征¹, 苏春阳¹

(1. 中国矿业大学计算机科学与技术学院, 江苏徐州 221116;

2. 中国科学院计算技术研究所智能信息处理重点实验室, 北京 100080)

摘要: 针对特征变量多的小样本, 结合偏最小二乘(Partial Least Squares, PLS)法则原理与 Elman 神经网络结构性质, 提出基于 PLS 的 Elman 神经网络算法(PLS-Elman). 新算法通过 PLS 对高维小样本进行特征降维时, 顾及了与因变量的相关程度, 所得到的数据进行网络训练和仿真, 明显的简化了网络结构, 且可得较精确的网络模型. 通过实例分析, 结果表明新算法提高了网络的收敛速度、预测的精准率, 证明新算法提高网络处理问题的效率. 同时为便于验证新算法的有效性, 与基于主成分分析(Principal Component Analysis, PCA)的 Elman 神经网络算法(PCA-Elman)进行了比较, PLS-Elman 算法有明显的优越性.

关键词: Elman 神经网络; 偏最小二乘法; PLS-Elman 算法; 主成分分析

中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112 (2010) 2A-071-05

Elman Neural Network Algorithm Based on PLS

DING Shi-fei^{1,2}, JIA Wei-kuan¹, XU Xin-zheng¹, SU Chun-yang¹

(1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China;

2. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: As to small size samples which have many characteristic variables, when Partial Least Squares (PLS) principle and structural properties of Elman neural network are taken into account, PLS-Elman is put forward. The new algorithm, when carrying feature reduction on high-dimensional and small size sample, takes its relativity to dependent variable into account. Obtained data carries on network training and simulation, clearly simplifies network structure and can get more precise network models. According to case analysis, the result shows that new algorithm improves convergence rate of the network, the predicting precision and proves that new algorithm improves the efficiency of the dealing with problems of the network. In the meantime, in order to test the effectiveness of new algorithm, it is compared with Elman neural network algorithm based on Principal Component Analysis (PCA-Elman) and it is observed that PLS-Elman algorithm has more advantages.

Key words: elman neural network; PLS; PLS-elman algorithm; PCA

1 引言

人工神经网络(Artificial Neural Network ANN)^[1]是以计算机网络系统模拟生物神经网络的智能计算为基础, 在处理非线性问题和大规模计算时显现了强大的功能. Elman 神经网络^[2]是一种反馈型神经网络, 在前馈网络的隐含层中增加一个承接层, 作为延时算子, 达到记忆目的, 从而使系统具有适应时变特性的能力.

用神经网络处理高维样本时, 由于太多的特征变量输入, 会不利于网络的设计, 占用大量的存储空间和计算时间, 引起训练过程耗时费工, 妨碍训练网络的收敛, 甚至最终影响网络的识别精度. 就要从大量的特征变量

中分析、提取最有用的特征, 排除相关或重复因素的影响, 在不影响问题解决的前提下尽可能降低特征的维数. 旨在简化网络结构, 提高网络的训练速度、收敛性和泛化能力. 数据降维与神经网络有机融合, 是目前神经网络算法改进的热点之一, 也取得了可喜成果. C Melchiorre 将聚类分析与 ANN 结合进行滑坡敏感性分析, 引入距离测度对样本进行分类, 进而排除相关因素影响选择有意义的类, 以提高神经网络的预测能力^[3]; E Lewis 等用主成分分析(Principal Component Analysis, PCA)与神经网络组合分析复杂的光学光谱和时间分辨信号, 测定食品质量^[4]; K R Janes 运用因子分析与神经网络组合对分析农场气味建立了预测模型, 找出气味成分和发

生过程^[5]; E S Gopi 用独立成分分析提取伪造图像的特征, 通过 NN 对伪造的图像进行鉴别^[6]等算法改进在处理大样本时都取得了很好的效果.

然而当待处理样本为特征变量多、样本数目少的样本时, 经典的特征压缩理论就显得无能为力, 且处理效果很不理想. 基于偏最小二乘 (Partial Least Squares, PLS) 特征压缩算法^[7,8], 利用 PLS 进行特征压缩时顾及与因变量的相关程度, 对高维小样本压缩可以取得理想结果. 运用 PLS 对神经网络算法对改进已取得一定的进展, Li R H 等用 ANN-PLS 组合模型对发电机的绝缘定子的使用寿命做了预测^[9]; 骆中华等提出一种基于 PLS 快速剪枝法的 RBF 神经网络软测量建模方法, 并应用苯二甲酸晶体平均粒径测量, 取得了较好的效果^[10]; 尹建川等提出一种利用分级偏最小二乘回归方法构造径向基函数神经网络的方法, 并通过船舶航向跟踪预测控制仿真, 验证了该算法的可行性和有效性^[11]; 刘波平等用 PLS 与 Elman 神经网络组合, 对近红外光谱技术进行多组分定量分析验证了其有效性^[12].

以上实例大都取得了不错的效果, 本文通过 PLS 对高维小样本特征进行降维, 将低维数据对 Elman 神经网络进行训练, 以期探索基于 PLS 的 Elman 神经网络算法 (PLS-Elman). 并通过实例分析, 以验证该新理论的实用性.

2 PLS-Elman 组合神经网络算法

2.1 PLS 降维基本原理

PLS 是最小二乘法 (Ordinary Least Squares, OLS) 的进一步拓展, 其基本思想是对特征变量矩阵 X 进行压缩的同时, 兼顾与因变量矩阵 Y 的相关性. 假设有 n 个特征变量 x_1, x_2, \dots, x_n , p 个因变量 y_1, y_2, \dots, y_r , 对 X, Y 进行预处理后矩阵分解

$$X = TP^T + E \quad (1)$$

式中 T 得分矩阵, P 是载荷矩阵, E 是残差矩阵. 矩阵乘积 TP^T 可以表示为得分向量 t_i (T 的第 i 列) 与负荷向量 p_i (P 的第 i 列) 的乘积之和, 则式(1)可写为

$$X = \sum_{i=1}^n t_i p_i^T + E, \quad i = 1, 2, \dots, n \quad (2)$$

同理, Y 矩阵可分解为:

$$Y = UQ^T + F \quad (3)$$

式中 U 得分矩阵, Q 是载荷矩阵, F 是残差矩阵. 矩阵乘积 UQ^T 可以表示为得分向量 u_j (U 的第 j 列) 与负荷向量 q_j (Q 的第 j 列) 的乘积之和, 则式(3)可写为

$$Y = \sum_{j=1}^r u_j q_j^T + F, \quad j = 1, 2, \dots, r \quad (4)$$

PLS 分析就是分别从 X 和 Y 中提取各自的得分向量 t 和 u , 它们分别为特征变量与因变量的线性组合.

且二者满足最大程度地承载特征变量和因变量的变异信息; 二者之间的协方差最大. 建立回归方程

$$u_j = b_k t_i \quad (5)$$

其中 b_k 为回归系数, 用矩阵形式可表达为

$$Y = BX \quad (6)$$

其中 B 系数矩阵,

$$B = W(P^T W)^{-1} Q^T \quad (7)$$

W 为 PLS 的权重矩阵.

PLS 对每一维度互相利用对方的信息进行迭代计算, 每一次迭代不断根据 X, Y 的剩余信息 (即其残差矩阵) 调整 t_i, u_j 进行第 2 轮的成分提取, 直到残余矩阵中的元素绝对值近似为零, 精度满足要求, 则算法停止. 迭代过程中, t_i, u_j 能同时最大限度地表达 X 和 Y 的方差.

PLS 回归并不需要选用存在的所有的成分来建立回归方程, 只选取前 l 个成分 ($0 \leq l \leq n$), 即可得到较好的回归方程. 一般采用交叉验证法, 计算预报残差平方和, 确定 PLS 所抽取的成分数, 达到降维的目的.

2.2 Elman 神经网络基本原理

Elman 回归神经网络拓扑结构如图 1 所示, 一般分为 4 层: 输入层、隐含层、承接层和输出层, 承接层用来记忆隐含层单元前一时刻的输出值, 可以看作是一个一步延时算子. 它是在反向传播 (Back-Propagation, BP) 神经网络基本结构的基础上, 隐含层的输出通过承接层的延迟与存储, 自联到隐含层的输入, 这种自联方式使其对历史状态的数据具有敏感性, 内部反馈网络的加入增加了网络本身处理动态信息的能力. 通过存储内部状态使其具备映射动态性的功能, 从而使系统具有适应时变特性的能力.

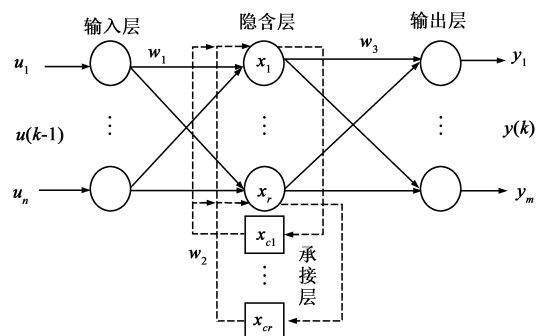


图1 Elman神经网络拓扑结构图

设有 n 个输入, m 个输出, 隐含层和承接层有 r 个神经元, 输入层到隐含层的权值为 w_1 , 承接层到隐含层的权值为 w_2 , 隐含层到输出层的权值为 w_3 ; $u(k-1)$ 表示神经网络的输入, 用 $x(k)$ 表示隐含层的输出, $x_c(k)$ 表示承接层的输出, $y(k)$ 表示神经网络的输出. 则

$$x(k) = f(w_2 x_c(k) + w_1 (u(k-1))) \quad (8)$$

$$x_c(k) = x(k-1) \quad (9)$$

$$y(k) = g(w_3 x(k)) \quad (10)$$

其中 f 为隐含层的传递函数,常用 S 型函数,即

$$f(x) = (1 + e^{-x})^{-1} \quad (11)$$

g 为输出层的传递函数,常为线性函数。

Elman 网络采用 BP 算法进行权值修正,网络的误差为:

$$E = \sum_{k=1}^m (t_k - y_k)^2 \quad (12)$$

其中 t_k 为目标输出向量。

2.3 基于 PLS 的 Elman 神经网络算法

利用 Elman 神经网络处理特征变量多、样本数目少的样本时,如果过多的网络输入,对隐含层的神经元个数确定是个难题,同时也会影响网络的训练;而相对较少的样本数又影响网络的训练精度,进而影响网络的识别精度。此时对自变量特征降维应顾及因变量的相关性,求精确的降维数据,如先将原始数据通过 PLS 进行特征降维,然后作为网络的输入。建立基于 PLS 的 Elman 神经网络算法(PLS-Elman 算法),算法流程如图 2 所示,这样便于隐层设计、提高训练速度,以提高网络的处理效率。

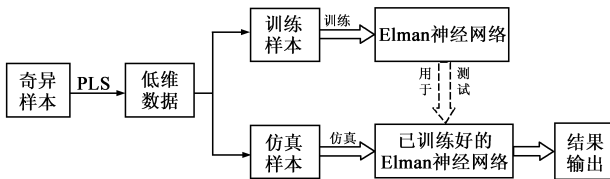


图2 PLS-Elman算法流程图

PLS-Elman 算法的基本步骤:

Step 1 将原始数据标准化,得特征变量矩阵 X 和因变量矩阵 Y ;

Step 2 分别在 X 、 Y 中提取第一对成分 t_1 、 u_1 ,并使之相关性达最大,分别建立 X 对 t_1 的回归以及 Y 对 t_1 的回归;

Step 3 用残差阵 E 和 F 代替 X 和 Y 重复 Step 2,直到残余矩阵中的元素绝对值近似为零;

Step 4 用交叉验证法,通过计算残差平方和确定抽取成分的个数;

Step 5 从信息特征压缩的角度,同时得到 X 和 Y 的压缩矩阵,看作新的样本;

Step 6 将新样本根据问题需要划分为训练样本和仿真样本;

Step 7 根据所抽取的成分个数和实际问题的输出设计 Elman 神经网络结构;

Step 8 利用训练样本对网络进行训练,确定 Elman 神经网络的隐层神经元个数,以及所有连接权值、阈值;

Step 9 用训练好的 Elman 神经网络模型对仿真样本进行预测。

根据 PLS-Elman 算法步骤,类似建立基于 PCA 的 Elman 神经网络算法(PCA-Elman)^[13]。

3 实例分析

已知数据^[14],利用气象因子预测小麦吸浆虫的发生程度。这里选择 1986-2000 年的数据资料 15 个样本作为研究对象,用 $x_1 - x_{14}$ 表示待处理原始数据的 14 个特征变量(气象因子), Y 表示当年小麦吸浆虫的发生程度(14 个特征变量,只有 15 个样本,用神经网络处理可以看作是低维小样本)。采用标准化方法对原始数据进行标准化处理(处理后的数据仍记为 X)。害虫发生系统的预测,在本质上都可以看作一个输入输出系统,

表 1 通过 PLS 与 PCA 分别降维得到的低维数据

年份	PLS 降维得低维数据						PCA 降维得低维数据					
	t1	t2	t3	t4	t5	t6	P1	P2	P3	P4	P5	P6
1986	-0.3399	2.1033	0.0324	-0.2966	-1.5854	-0.0422	-0.9923	-1.1104	-2.4299	-0.4116	1.5889	1.1063
1987	-0.3386	2.0559	0.2919	-1.2709	0.9692	1.4484	-1.3250	-1.4261	0.2452	-0.2305	0.6969	-0.7976
1988	-1.2719	0.7677	-0.5190	-0.0138	1.6462	-0.4957	0.5071	-1.4875	3.1671	-0.0898	0.6239	0.8705
1989	-3.8970	-1.1010	1.0484	-1.0285	0.1411	-0.5403	5.4469	0.0875	0.1535	1.4501	-0.3608	-0.3430
1990	-2.3352	-0.3105	0.3735	0.3927	-0.0044	0.3268	2.3190	0.3451	-0.2026	-0.5756	0.6745	-0.0961
1991	0.6033	0.4230	-1.1310	0.7038	0.5788	-0.2400	-1.3246	-0.9176	1.2502	-0.1244	-0.0051	-0.3802
1992	0.4783	-1.1772	-1.2116	-0.6397	0.0938	-0.9670	-0.1920	-1.5841	1.2374	-1.4616	-1.5315	-0.1984
1993	-0.2084	-1.5397	-1.4361	-0.3456	-0.6862	0.8983	1.1375	-1.3245	-1.6489	-0.3652	-0.9855	-0.3969
1994	-0.7313	0.5216	-1.7178	2.3387	-0.3698	-0.0425	-0.0388	-0.4420	-0.1982	0.8822	1.5901	-0.2613
1995	2.2686	0.1553	0.8227	-1.1167	-0.5840	-1.2815	-1.8865	0.3162	-0.4479	-0.0377	-1.1573	0.7443
1996	2.0200	-0.4237	-0.5989	-0.3853	0.9485	-0.6370	-1.7236	-0.1622	1.1836	1.7771	-0.4898	0.7158
1997	-0.9612	-0.4790	0.7784	0.2538	-1.1804	-0.1070	1.5592	1.0104	-0.8627	-1.8521	0.0057	0.4113
1998	1.8223	0.1781	-0.6179	-0.8766	-0.6402	0.8414	-1.7306	-0.8734	-2.2116	1.1485	-0.7525	-0.8944
1999	1.1180	0.7023	2.3765	1.5546	0.2085	-0.5170	-1.5223	3.5265	1.0909	-0.4766	0.6585	-1.4953
2000	1.7730	-1.8759	1.5086	0.7302	0.4643	1.3552	-0.2340	4.0420	-0.3262	0.3672	-0.5559	1.0149

转换关系包括数值拟和、模糊转换和逻辑推理三个方面,这些都可以用神经网络来表达。

为了更好地说明问题这里分别采用 Elman 神经网络、PCA-Elman 算法、PLS-Elman 算法对该问题进行测试,对比试验结果.试验中都选择 1986-1995 年的 10 样本作为训练样本,用 1996-2000 年 5 个样本做为仿真预测样本。

将样本数据通过 PCA 分析,该实例析取出 6 个主成分(列于表 1);再对实例中原始数据通过 PLS 特征压缩可提取 6 个特征(同样列于表 1)。

利用 MATLAB 工具箱,先建立输入层 14 个神经元,输出层 1 个神经元的 Elman 神经网络,对问题进行模拟,将处理结果列于表 2.再利用 PCA-Elman 算法进行

表 2 Elman 算法、PCA-Elman 算法、PLS-Elman 算法预测结果比较

网络结构	1996	1997	1998	1999	2000	训练步数
实际值	1	2	2	2	1	—
Elman 算法	0.8146	2.2433	1.7688	1.8216	1.1740	532
PCA-Elman 算法	0.7893	1.8905	2.2330	1.8470	0.8796	178
PLS-Elman 算法	1.0811	2.0671	1.8480	2.1125	1.2419	115

4 结果与讨论

本文针对特征变量多而样本量少的情况下,在 PLS 算法的基础上,结合 Elman 神经网络算法提出了 PLS-Elman 组合神经网络算法.当特征变量集合内存在较程度的相关时,通过特征降维,既减少数据冗余,又排除相关、重复数据的影响.PLS 能利用对系统中的数据进行分解和筛选,提取对预测变量解释性最强的综合变量,且对特征变量进行压缩的同时,顾及了与因变量的相关程度,其压缩结果将比 PCA 具有更实际的意义.而神经网络又具有较强的非线性处理问题的能力,Elman 神经网络是在 BP 神经网络基础上增加了反馈功能.将 PLS 与 Elman 神经网络二者优点有机结合的新算法,能够较好地拟合非线性预测问题.通过实例分析的结果表明,PLS-Elman 算法明显优于 PCA-Elman 组合算法,比单纯的神经网络预测有较大程度的改进,神经网络的自学能力增强、收敛速度加快、节省了运行时间,进而提高了网络的效率。

用 Elman 神经网络处理特征变量多、样本量少的样本时,训练得到的网络往往不是很理想,而 PLS 处理这类问题却有独特的优越性,原始数据经 PLS 处理后,再进行网络训练,得到的结果令人满意,表明算法值得推广。

参考文献:

[1] Mcclloch S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. Bulletin of Mathematical Biophysics,

预测,将样本数据通过 PCA 分析,得到的低维数据作为网络的输入,建立新的 Elman 网络,得出预测结果(列于表 2).最后利用 PLS-Elman 算法进行预测,利用 MATLAB 按新建立的算法,实例中原始数据通过 PLS 特征压缩,将得到的低维数据其作为新的样本,进行设计仿真,结果列于表 2。

表 2 显示,改进后的算法,对原始数据经 PCA 或 PLS 进行降维,虽有一定的信息损失,两种新算法的预测精度并没有降低,而收敛步数却相应减少,说明原始数据经降维后,减少了网络的输入,易于网络设计,简化网络结构,提高网络的训练速度,表明经改进后的算法优于传统的算法.而 PLS-Elman 组合算法的预测精度明显优于 PCA-Elman 组合算法。

1943, 10(5):115 - 133.

- [2] Elman J L. Finding structure in time[J]. Cognitive Science, 1990, 14(2):179 - 211.
- [3] Melchiorre C, Matteucci M, Azzoni A, et al. Artificial neural networks and cluster analysis in landslide susceptibility zonation [J]. Geomorphology, 2008, 94(3):379 - 400.
- [4] Lewis E, Sheridan C, Farrell M O, et al. Principal component analysis and artificial neural network based approach to analysing optical fibre sensors signals[J]. Sensors and Actuators A: Physical, 2007, 136(1):28 - 38.
- [5] Kevin R J, Simon X Y, Roger R H. Pork farm odour modelling using multiple-component multiple-factor analysis and neural networks[J]. Applied Soft Computing, 2005, 6(1):53 - 61.
- [6] Gopi E S. Digital image forgery detection using artificial neural network and independent component analysis [J]. Applied Mathematics and Computation, 2007, 194(2):540 - 543.
- [7] 王惠文. 偏最小二乘回归方法及其应用[M]. 北京:国防工业出版社, 2000.
Wang Hui-wen. Partial Least Squares Regression and Applications[M]. Beijing: National Defence Industry Press, 2000.
- [8] 丁世飞, 靳奉祥, 史忠植. 基于 PLS 的信息特征压缩算法[J]. 计算机辅助设计与图形学学报, 2005, 17(2):368 - 371.
Ding Shi-fei, Jin Feng-xiang, Shi Zhong-zhi. Information feature compression based on partial least squares[J]. Journal of Computer Aided Design & Computer Graphics, 2005, 17(2):368 - 371.
- [9] Li Rui-hua, Meng Guo-xiang, Gao Nai-kui, et al. Combined use of partial least-squares regression and neural network for resid-

- ual life estimation of large generator stator insulation[J]. Measurement Science and Technology, 2007, 18(7):2074 - 2082.
- [10] 骆中华, 刘瑞兰, 苏宏业, 等. 基于 PLS 快速剪枝法的 RBF 神经网络软测量模型建模方法和应用[J]. 化工自动化及仪表, 2005, 32(3):19 - 21.
- Luo Zhong-hua, Liu Rui-lan, Su Hong-ye, et al. RBF neural network soft sensor modeling algorithm based on fast-PLS-pruning and its application[J]. Control and Instruments in Chemical Industry, 2005, 32(3):19 - 21.
- [11] 尹建川, 侯建军, 等. 基于分级偏最小二乘回归的径向基函数神经网络[J]. 大连海事大学学报, 2007, 33(3):113 - 116.
- Yin Jian-chuan, Hou Jian-jun, et al. Two-stage partial least squares regression for constructing radial basis function networks[J]. Journal of Dalian Maritime University, 2007, 33(3):113 - 116.
- [12] 刘波平, 秦华俊, 罗香, 等. 基于 Elman 网络的近红外光谱技术多组分定量分析研究[J]. 光谱学与光谱分析, 2007, 27(12):2456 - 2459.
- Liu Bo-ping, Qin Hua-jun, Luo Xiang, et al. Study on the multicomponent quantitative analysis using near infrared spectroscopy based on building Elman model[J]. Spectroscopy and Spectral Analysis, 2007, 27(12):2456 - 2459.
- [13] Ding Shi-fei, Jia Wei-kuan, Su Chun-yang, et al. PCA-based elman neural network algorithm[A]. Kang Lishan, Cai Zhihua, Liu Yong (eds.) Advances in Computation and Intelligence, ISICA2008, LNCS, Springer, Heidelberg [C]. Heidelberg: Springer, 2008. 5370. 315 - 321.

- [14] 张映梅. 人工神经网络在小麦吸浆虫预测中的应用[D]. 西安:西北农林科技大学, 2003.
- Zhang Ying-mei. The Application of Artificial Neural Network in the Forecasting of Wheat Midge[D]. Xi'an: Northwest A&F University. 2003.

作者简介:



丁世飞 男, 1963 年生于山东省青岛市, 工学博士, 博士后. 现任中国矿业大学教授, 博士生导师. 主要研究方向为人工智能、机器学习、智能信息处理等.

E-mail: dingsf@cumt.edu.cn



贾伟宽 男, 1982 年生于山东省菏泽市, 中国矿业大学计算机应用技术专业研究生, 主要研究方向为智能信息处理、神经网络等.

许新征 男, 1980 年生于安徽淮北市, 博士生. 现任中国矿业大学讲师, 主要研究方向为图像识别、神经网络等.

苏春阳 男, 1984 年生于江苏连云港市, 中国矿业大学计算机应用技术专业研究生, 主要研究方向为遗传算法、智能信息处理等.