

一种基于因子分析改进的 RBF 神经网络算法

杨 华,王 珂

(中国矿业大学 计算机科学与技术学院,江苏 徐州 221116)

摘 要: 针对 RBF 神经网络在处理大规模多维数据时,网络结构不易设计、收敛时间较长、训练次数较多等不足,在原有的 RBF 神经网络模型的基础上,结合因子分析算法可以对大规模数据进行降维处理的优点,提出一种 FA-RBF 神经网络算法.利用该算法首先可以先对输入的数据进行降维处理,将处理后的数据作为神经网络的输入进行网络的训练和仿真.FA-RBF 算法可以有效地简化网络结构、提高收敛速度、节省训练时间.将该算法用于私家车保有量的预测中,预测结果显示 FA-RBF 算法较之于 PCA-RBF 和 RBF 神经网络算法,其预测精度有所提高,训练时间及误差平方和都明显降低.

关键词: 人工神经网络;径向基函数;因子分析;FA-RBF 算法;私家车保有量

中图分类号: TP183

文献标识码: A

文章编号: 1000-7180(2011)10-0105-04

An Improve-Based RBF Neural Network Algorithm Based on Factor Analysis

YANG Hua, WANG Ke

(Department of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: FA-RBF neural network algorithm is proposed by combining factor analysis algorithm, which has advantage of reducing the data's dimensions, with RBF neural network. Using the new algorithm, the original data can first reduce its dimension, then, the processed data can be the input of neural network to network training and simulation. FA-RBF algorithm can effectively simplify the network structure, improve the convergence speed and save training time. Apply the algorithm for the prediction of private car ownership, the prediction results show that compared to PCA-RBF and RBF neural network algorithm FA-RBF algorithm significantly improved prediction accuracy, reduced training time and the sum of squared errors.

Key words: artificial neural networks; radial basis function; factor analysis; FA-RBF algorithm; private car ownership

1 引言

RBF^[1] (Radical Basic Function, RBF) 人工神经网络是一种结构简单、具有最佳逼近性能和全局最优特征的神经网络.对于 RBF 神经网络来说,虽然大规模的输入数据可以提供需要的信息,但同时也增加了其处理数据的难度.如果我们将不经过处理的大规模的数据直接作为 RBF 神经网络的输入,则会出现网络结构设计困难、占用过多存储空间、计

算时间过长等问题,而且多因素的输入数据和重复的训练样本,会给整个网络带来训练次数增加、收敛性降低等缺点,从而影响网络的预测精度.所以,对原始数据进行预处理并确定大规模数据中的有用影响因素、排除冗余影响因素、降低输入数据的维数是十分必要的.将经过降维处理的数据作为神经网络的输入数据,并确定其结构,可以有效避免上述相关问题.

文献[2-4]中分别提出了基于聚类分析(Clus-

收稿日期: 2011-04-23; 修回日期: 2011-06-25

基金项目: 国家大学生创新性实验计划项目(S2TP-2010-23)

ter Analysis, CA)、主成分分析法(Principal Components Analysis, PCA)和核主成分分析(Key Principal Components Analysis, KPCA)改进的 RBF 神经网络方法,其通过层次聚类将特征分散的样本分成不同的子层次,然后利用分层降维后的数据进行网络的构建和网络的训练,收到了一定的效果.文献[5]中利用独立分量分析(Independent Component Analysis, ICA)方法对输入数据进行降维处理并构建了新的神经网络模型.此外,文献[6-7]提出了一种利用因子分析方法(Factor Analysis, FA)改进的 BP 神经网络模型和基于因子分析和 RBF 神经网络的快速估算模型,在一定程度上提高了网络的预测精度,缩短了网络的训练时间.

在借鉴以上研究成果的基础上,现提出一种利用因子分析法改进的 RBF 神经网络算法,该算法将因子分析和 RBF 神经网络相结合,利用因子分析对原始数据进行降维处理,并利用处理后的数据进行网络结构的设计和训练,可在一定程度上降低训练时间,提高预测精度.

2 FA-RBF 神经网络算法

2.1 因子分析

因子分析是一种把变量从高维到低维作降维处理,从高维空间到低维空间的映射算法,它是一种既能保持样本在高维空间的某种“结构”,同时又能够保留原变量大部分信息的多元统计分析方法^[8].在实践中,收集到的数据往往是多指标的,而且各个指标之间不是相互独立的,其之间存在着一定程度的相关性.因子分析的目的是通过少数几个变量来描述这众多变量间的协方差关系.这少数几个变量是潜在的,不能通过观察得到,称为因子.根据变量间相关性的把高度相关的变量分组,使得同组内的变量之间相关性较高,不同组的变量相关性较低.这些组内高度相关的变量,受某一共同因素所影响,这个共同因素称之为公共因子.

因子分析可以通过确定 m 个“公共因子”正交变量,最优地显示 n 个变量 $X_i (X_i = x_1, x_2 \cdots x_n)$ 之间的差异($n > m$).经过标准化处理和数学变换,提取公共因子,再利用这些公共因子来描述和代替原始变量,可以反映和解释原始变量之间的复杂关系.具体过程如下:

计算 X_i 的相关系数矩阵 $R = r_i \times r_j$, 其中

$$r_{ij} = \frac{1}{n-1} \sum_{k=1}^n (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j)$$

$$i, j = 1, 2, \cdots, n, \bar{X} = \frac{1}{n} \sum_{k=1}^n X_{ki}$$

计算相关系数矩阵 R 的特征根,并根据前 q 个特征根的累积百分数大于或等于 85% 选取公共因子.计算特征根对应的单位特征向量,将其规格化为 a_{ij} , 可得到载荷矩阵 $A(a_{ij})$.

通过为因子载荷矩阵 A 右乘一个正交矩阵进行结构简化,化简后每个变量仅在一个公共因子上有较大的载荷,而其余公共因子上的载荷比较小,这使得载荷矩阵的每一列元素的平方值或尽可能大或尽可能小,即向 1 和 0 两级分化.然后利用下面公式计算各个因素的得分,并根据各个因素得分对数据进行降维处理.

$$X_i = \sum_{j=1}^m a_{ij} F_j + c_i \epsilon_i$$

$$(j = 1, 2 \cdots m; j = 1, 2 \cdots n; n > m)$$

式中, $F_j = F_1, F_2 \cdots F_m$ 为第 j 个因子, a_{ij} 为因子载荷,表示因子 F_j 与变量 X_i 之间的相关关系, c_i 为权重, ϵ_i 为特殊因子变量.

解出 F_j 的表示形式,将其作为原始数据的属性,通过计算从而重新表示原始数据.

通过以上过程,可以将大量原始数据的维度降低、结构得到简化,并在诸多的因子中为每一个变量找出包含多变量系统的信息和影响系统各种因素的最佳子集.

2.2 RBF 神经网络

RBF 神经网络是一种特殊的三层局部逼近神经网络,具有学习快、不会陷入局部最优的优点^[9]. RBF 神经网络是根据在人的大脑皮层中具有局部调节和交叠的感受域提出的,是一种新颖的、有效的前馈式神经网络,具有较高的运算速度和较强的非线性映射能力,可以以任意精度全局逼近一个非线性函数^[10].

如图 1 所示, RBF 神经网络由输入层、隐含层、输出层三层组成.其中,隐含层和输出层采用径向基函数作为激励函数,在这里我们用经常采用的一般高斯函数表达式作为径向基函数,公式如下:

$$R_i(x) = \exp\left(-\frac{1}{2\sigma^2} \|x - c_i\|^2\right)$$

$$i = 1, 2, \cdots, n$$

式中, $R_i(x)$ 为隐含层第 i 个单元的输出, $i = 1, 2, \cdots, n$, $\|\cdot\|$ 为欧氏范数, x 为 N 维输入矢量, C_i 为隐含层第 i 个单元高斯函数的中心, σ 为该隐含节点的宽度, n 为隐含层节点数.

RBF 神经网络的输入和输出之间是一种映射关系^[11],可表示为

$$F(x) = \sum_{j=1}^n w_j \exp(-\frac{1}{2\sigma^2} \|x - c_j\|^2)$$

$j = 1, 2, \cdots, n$

式中, w_j 为隐含层到输入层的权值, $j = 1, 2 \cdots n, n$ 为隐含层的节点的个数。

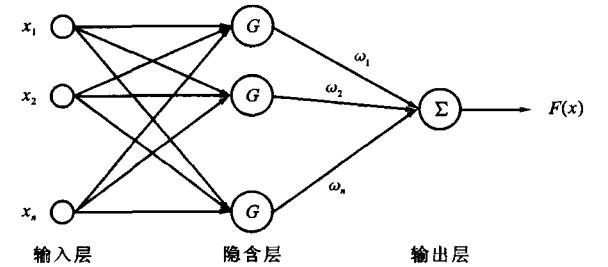


图1 RBF神经网络结构图

2.3 FA-RBF神经网络算法

针对原始数据维数较多而且不易处理的现象,可以将以上介绍的两种算法联合起来,构成一种新的算法来进行改进。

首先利用因子分析法确定原始数据的主要因子,并对待预测的数据进行降维处理,然后将这些降维后的数据作为神经网络的输入数据,并构建合适的神经网络结构,通过计算得到结果值。利用因子分析法改进的 RBF 神经网络降低了样本和输入数据的维度,使得神经网络的结构更为简洁,设计更为方便,并可以在一定程度上提高网络的训练时间和运算速度。FA-RBF 神经网络算法的流程图如图 2 所示:

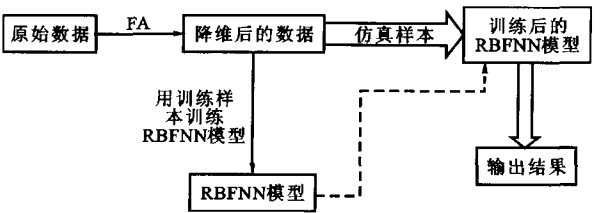


图2 FA-RBF神经网络算法流程图

FA-RBF神经网络算法步骤如下:

- Step1:原始数据标准化,计算相应的相关系数矩阵;
- Step2:计算相关系数矩阵的特征向量和特征值,及方差贡献率和累计方差贡献率,确定主要因子;
- Step3:转置载荷因子矩阵,计算将要被作为新的样本的得分;
- Step4:根据主要因子的个数来设计新的 RBF

神经网络,并确定隐含层的神经元的个数;
Step6:利用训练好的神经网络和降维后的数据对网络进行仿真。

3 算法仿真分析

将FA-RBF算法应用到私家车保有量的预测问题中,来检验算法的效果。

据国家统计局年鉴显示,自 1980 年代中国开始出现私人汽车到 2008 年底,我国私人汽车保有量约为 3 885 万辆,占全国汽车保有量的 60% 以上。私家车保有量的快速增长带来了能源紧缺、环境污染等相关的问题。如果对这种快速增长不从战略的高度加以科学引导,其带来的负面效应也将会成为社会发展的阻碍因素。所以,私家车保有量的预测研究已经成为国民经济、能源、环境等领域的一个重要问题。

影响私家车保有量的因素有多种,如:人均国内生产总值,全社会消费品零售总额,全社会固定资产投资总额,运营公交车辆数,公交营运总数,公交车营运总里程,道路总长,居民人均可支配收入,居民储蓄款余额,汽油(93 号) 年均价等。由于影响私家车保有量的因素较多,数据量较大,且各个因素之间存在一定程度的相关性,所以其较为适合用 FA-RBF 算法来进行求解。

这里选取 1981 — 1998 年的 18 组数据作为样本,用 $x_1 \sim x_{10}$ 表示 10 个初始数据的影响因素, Y 表示预测得到的私家车保有量的值。为了消除误差,首先利用以下公式对原始数据进行标准化操作得到 x' ,标准化后的数据仍可以保证 $X(x_1, \cdots, x_{10})$ 的特征不变^[12]。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

将标准化后的数据作为输入利用 RBF 神经网络进行预测。利用 Matlab 建立输入层有 10 个神经元和输出层有 1 出神经元的 RBF 神经网络。将 1981 — 1998 年的 18 组数据作为训练数据对神经网络进行训练,将 1999 — 2008 年的 10 组标准化后的数据作为预测数据,进行私家车保有量的预测,预测结果可见表 1。

表1 RBF、PCB-RBF 和 FA-RBF 算法的对比

神经网络算法	预测精确度	训练次数	误差平方和
RBF	77.48%	586	2.349 3
PCB-RBF	80.65%	372	1.890 1
FA-RBF	81.84%	289	1.769 4

利用 FA-BP 神经网络算法对预测样本进行预测并利用 SPSS 对标准化后的数据进行分析. 经过计算得到特征根和方差贡献率, 得到公共因子为居民储蓄款余额、全社会消费品零售总额、公交车营运总里程、道路总长, 其可以解释原始数据的 92.514% (表 2). 进而确定各个主要因素对每个原始指标的线性系数, 得到因子载荷矩阵. 转置因子载荷矩阵, 使其结构尽量简单, 并计算各个因子的得分, 降低数据的维数.

表 2 通过方差确定公共因子

因素名称	特征根	方差 贡献率/%	方差累积 贡献率/%
居民储蓄款 余额	2.813	40.436	40.467
全社会消费品 零售总额	2.176	31.238	71.675
公交车营运 总里程	1.257	14.674	84.393
道路总长	1.036	10.753	92.514

利用降维后的数据作为输入来构建新的 RBF 神经网络, 将 1981—1998 年的 18 组数据作为训练数据对神经网络进行训练, 对 1999—2008 年的私家车保有量进行预测. 类似的, 参考文献[3]介绍的 PCA-RBF 方法对数据进行降维处理并进行预测, 三种方法的预测结果可见表 1.

FA-RBF 神经网络算法和 PCA-RBF、RBF 神经网络算法的预测结果对比如表 1 所示, 改进后的算法的预测精度有所提高, 训练时间及误差平方和都明显降低. 产生这种结果的原因是通过因子分析的方法, 降低了原始数据的维度, 使得输入神经网络的数据量减少, 确定的神经网络的结构更为简洁, 收敛速度更快. 图 3 显示了两个算法预测数据对比曲线, 曲线说明了 FA-RBF 算法优于 PCA-RBF 和 RBF 算法且具有较强的应用性, 值得更多的研究和推广.

4 结束语

本文提出了一种将因子分析法和 RBF 神经网络结合起来的 FA-RBF 神经网络算法. 因子分析法可用来降低原始数据的维数, 确定主要因素, 减少冗余数据, 消除不同因素之间的相关性的影响, 神经网络则很擅长用来求解复杂非线性问题. FA-RBF 神经网络算法结合了这两种算法的优点, 可以应用于输入数据相关性较高、规模较大的复杂非线性预测问题. 实验仿真结果表明, FA-RBF 算法较之于

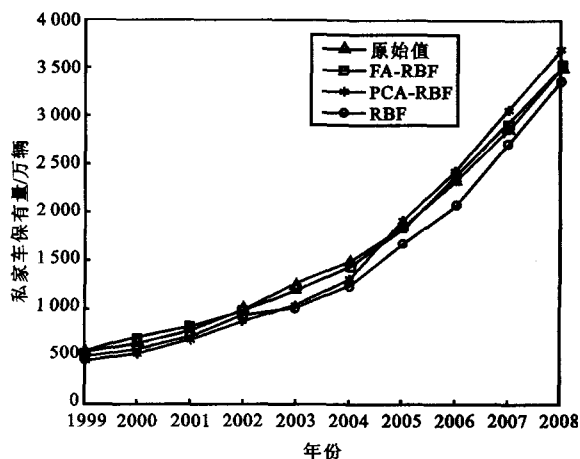


图 3 FA-RBF 和 RBF 预测结果的对比曲线

PCA-RBF 和 RBF 算法在自学习性、收敛速度和预测精度等方面均有了较大的提高.

文中验证了与神经网络相结合的多元统计分析, 可以提高用神经网络处理问题的效率^[5]这一观点. 根据这一思想, 倘若将其算法应用于神经网络的预处理过程中并与神经网络的设计相结合将会是神经网络研究的一个重要的方向之一.

参考文献:

- [1] Powell M I D. Radial basis function for multivariable interpolation[C]// Review IMA Conf on algorithms for the Approximation of Functions and Data. RMCS Shrivaniham, 1985.
- [2] Ding S F, Xu L. Research and progress of cluster algorithms based on granular computing[J]. JDCTA: International Journal of Digital Content Technology and its Applications, 2010, 4(5): 96—104.
- [3] 李艳, 吴介军, 寇晓东, 等. 基于 PCA-RBF 网络的高校保密项目风险预测[J]. 科技管理研究, 2011(5): 101—103.
- [4] 杨俊, 陈贤富. 基于 KPCA 和 RBF 网络的文本分类研究[J]. 微电子学与计算机, 2010, 27(3): 122—125.
- [5] 江帆, 高涛, 刘金安, 等. 基于独立分量分析和径向基神经网络的人脸识别方法[J]. 国外电子元器件, 2008(10): 64—66.
- [6] Ding S F, Jia Z Z, Su C Y, et al. An improved-based BP neural network algorithm based on factor analysis[J]. JCIT: Journal of Convergence Information Technology, 2010, 5(4): 103—108.
- [7] 吴倩, 宋永发. 基于因子分析和 RBF 神经网络的快速估算模型[J]. 工程管理学报, 2010, 24(1): 65—68.
- [8] 蔡建琼, 于慧芳, 朱志洪. SPSS 统计分析实例精选[M]. 北京: 清华大学出版社, 2006.

5 结束语

利用全局路径规划方法寻找群体自主运动的最短路径,通过对运动曲线控制顶点的改变实现了群体运动的可控性.通过引导分量矢量场、范围保持分量矢量场和个体间自主的碰撞避免实现了个体运动的自主性,动画师可以通过调节各个分量的权值对个体运动进行间接的交互控制,同时,由于引导分量的矢量场是在运动模拟过程中动态生成的,因此动画师还能通过动态地改变引导分量来实现对个体运动的实时交互控制.实验结果表明,该方法既较好地模拟了群体动画中个体的自主运动,又满足实时控制的要求,实现了自主性与可控性相结合的群体动画运动控制.进一步的工作是对团体操等特殊应用情况下加入更多的自主运动规则,以便在更多的情况下模拟个体的自主运动^[9].

参考文献:

- [1] Craig W Reynolds. Steering behaviors for autonomous characters[C]// Proceedings of Game Developers Conference, San Francisco, 1999: 763—782.
- [2] Takeshi Sakuma, Tomohiko Mukai, Shigeru Kuriyama. Psychological model for animating crowded pedestrians [J]. Virtual Humans and Social Agents. Comput. Animat. Virtual Worlds, 2005, 16(3/4): 343—351.
- [3] Marta B Villamil, Braun A, Soraia R Musse. A rules—based model used to describe group dynamics for games [C]// Computer Graphics and Image Processing, SIIG-GRAPI 2003. Brazil, Sao Carlas, 2003: 91—98.
- [4] Bayazit O B, Lien J, Amato N M. Better group behaviors in complex environments using global roadmaps[C]// Proceedings of the eighth international conference on Artificial Life, Sydney, 2002: 37—57.
- [5] Musse S R, Thalmann D. Hierarchical model for real time simulation of virtual human crowds [J]. IEEE Transactions on Visualization and Computer Graphics archive, 2001, 7(2): 152—164.
- [6] Ulicny B, Ciechomski P H, Thalmann D. Crowd brush: interactive authoring of real—time crowd scenes[C]// Proceedings of the ACM Siggraph/Eurographics Symposium on Computer Animation (SCA' 04). France, Grenoble, 2004: 243—252.
- [7] Sung M, Gleicher M, Chenney S. Scalable behaviors for crowd simulation[J]. Computer Graphics Forum, 2004, 23(3): 519—528.
- [8] Stylianou S, Fyrillas M, Chrysanthou Y. Scalable pedestrian simulation for visual cities[C]// Proceedings of the ACM Symposium on Virtual Reality Software and Technology. HongKong, 2004: 65—72.
- [9] Xiaogang Jin, Jiayi Xu, Charlie C L Wang, et al. Interactive control of large—crowd navigation in virtual environments using vector fields[J]. IEEE Comput Graph Appl, 2008, 28(6): 37—46.

作者简介:

姚巍 男, (1986—), 硕士. 研究方向为计算机图形学、模式识别.

刘惠义 男, (1961—), 博士, 教授. 研究方向为计算机图形学、虚拟现实、模式识别.

沈赞芳 女, (1985—), 硕士. 研究方向为计算机图形学、模式识别.

(上接第 108 页)

- [9] 史忠植, 王文杰. 人工智能[M]. 北京: 国防工业出版社, 2007: 387—417.
- [10] 马辉. 建设工程项目快速投资估算方法研究[D]. 天津: 天津大学, 2006.
- [11] 王洪斌, 杨香兰, 王洪瑞. 一种改进的 RBF 神经网络学习算法[J]. 系统工程与电子技术, 2002, 24(6): 38—41.
- [12] 曾希君, 于博. 基于改进的 BP 神经网络私家车保有量

的预测研究[J]. 计算机工程与设计, 2010, 31(3): 605—608.

作者简介:

杨华 男, (1990—). 研究方向为人工智能、传感器网络.

王珂 男, (1979—), 讲师. 研究方向为人工智能、传感器网络、地理信息系统.