神经网络模型的透明化及输入变量约简

姚立忠[1] 李太福[2] 易 军[2] 苏盈盈[2] 胡文金[2] 肖大志[2]
(西安石油大学电子工程学院 西安 710065)[1] (重庆科技学院电气与信息工程学院 重庆 401331)[2]

摘 要 由于神经网络很容易实现从输入空间到输出空间的非线性映射,因此,神经网络应用者往往未考虑输入变量和输出变量之间的相关性,直接用神经网络来实现输入变量与输出变量之间的黑箱建模,致使模型中常存在冗余变量,并造成模型可靠性和鲁棒性差。提出一种透明化神经网络黑箱特性的方法,并用它剔除模型中的冗余变量。该方法首先利用神经网络释义图可视化网络;再利用连接权法计算神经网络输入变量的相对贡献率,判断其对输出变量的重要性;最后利用改进的随机化测验对连接权和输入变量贡献率进行显著性检验,修剪模型,并以综合贡献度和相对贡献率均不显著的输入变量的交集为依据,剔除冗余变量,实现NN模型透明化及变量选择。实验结果表明,该方法增加了模型的透明度,选择出了最佳输入变量,剔除了冗余输入变量,提高了模型的可靠性和鲁棒性。因此,该研究为神经网络模型的透明化及变量约简提供了一种新的方法。

关键词 神经网络模型,透明化,网络释义图,变量选择
中图法分类号 TP183 **文献标识码** A

Visualize Black-box of NN Model and its Application in Dimensionality Reduction

YAO Li-zhong[1] LI Tai-fu[2] YI Jun[2] SU Ying-ying[2] HU Wen-jin[2] XIAO Da-zhi[2]
(School of Electronic Engineering, Xi'an Shiyou University, Xi'an 710065, China)[1]
(School of Electrical and Information Engineering, Chongqing University of Science and Technology, Chongqing 401331, China)[2]

Abstract Since the neural network can easily fit the nonlinear mapping from the input space to the output space, the users of artificial neural network directly use it to get the black-box model with data pairs including input variables and output variables, often without taking dependencies between the input variables and output variables into account. So, there are often redundant variables in the model which would result in poor reliability and robustness. An approach to increase visual capability for black-box properties of neural network was proposed. Firstly, the network interpretation diagram is employed to make the network transparency. Then, connection weights method is used to compute the relative contribution of each input variable for estimating the importance to the output variable. Lastly, the significance tests of the connection weights and contribution ratios of input variables are implemented using the improved randomization tests for trimming the model, and the redundant variables can be eliminated by the intersection of the variables which are not significant for the overall contribution and the relative contribution rate to realize dimensionality reduction of neural network model. The experimental results indicate that the method can increase the transparency of model, select the best input variable set, eliminate redundant input variables, and improve the reliability and robustness of the model. Therefore, the study provides a new approach to visualize neural network model and eliminates redundant input variables.

Keywords Neural network model, Visualize, Network interpretation diagram, Variables reduction

1 引言

神经网络模型具有自适应学习、高容错自组织功能,适合解决从输入空间到输出空间之间的非线性映射问题[1]。它仅通过对训练样本的学习,就能任意逼近一个高度非线性的函数,被视为"万能函数逼近器"。其已在函数逼近[2]、模式识别[3]、故障诊断[4]、预测控制[5]等领域得到广泛应用。为此,神经网络应用者可能未考虑输入变量和输出变量之间的相关性,就直接用神经网络实现输入变量与输出变量之间的建模。

但是,神经网络建模方法具有"黑箱"特性。利用该方法建立的模型,不能确定输入变量对输出变量的贡献率及输入变量与输出变量之间的关系,模型中的参数不能直观地解释真实系统,难以观察模型输入变量的灵敏性[6]。另外,神经网络模型的可靠性和鲁棒性受模型输入变量选择的影响很大[7],不合理变量的引入难以保证神经网络模型的可靠性和鲁棒性。因此,研究弱化神经网络黑箱特性的方法及如何从大量的影响因素中选择出一组对期望输出影响最大的输入变量子集,是神经网络建模面临的重要问题。

为了弱化神经网络模型的黑箱特性,Ozesmi等[8]提出了神经网络释义图(Neural Interpretation Diagram, NID),实现

到稿日期:2012-02-14 返修日期:2012-04-25 本文受国家自然科学基金(61174015,51075418),重庆市自然科学基金(CSTC2010BB2285)资助。
姚立忠(1985—),男,硕士生,主要研究方向为人工智能、复杂系统建模与稳健优化,E-mail:yaolizhong225@163.com;李太福(1971—),男,博士,教授,主要研究方向为人工智能、复杂系统建模与稳健优化。

了神经网络模型的可视化，但该方法只对模型做了定性的解释，而没有量化变量的贡献度，另外当模型复杂度增加时，这种定性解释也几乎无法实现。Gevrey 等人[9]研究和比较了几种评价变量贡献度方法的优劣，包括“加森(Garson)方法”、“偏导数(Partial Derivatives)”、“输入扰动(Input Perturbation)”、“灵敏度分析(Sensitivity Analysis)”等。然而，Olden 等[10]认为 Gevrey[9]在未知数据分布和贡献大小的条件下评价各方法优劣是无效的，并通过已知数据证实了连接权法是衡量各变量贡献度的理想方法。为了解决模型复杂度增加时，神经网络释义图(NID)解释力较差的问题，Olden[11]利用随机化方法修剪了神经网络释义图，但该方法采用加森算法计算输入变量的相对贡献率，忽略了权重之间的状态。为此，本文改进了随机化准则，利用连接权法得到相对贡献率进行随机化测验。

然后，本文结合神经网络释义图、连接权法和改进的随机化测验 3 种方法，研究了神经网络模型的透明化及输入变量约简。该方法首先利用神经网路释义图可视化神经网络模型；再利用连接权法计算神经网络输入变量的相对贡献率，判断其对输出变量的重要性；最后利用改进的随机化测验对连接权和输入变量的贡献率进行显著性检验，修剪模型，并以综合贡献度和相对贡献率不显著变量的交集为依据，剔除冗余实现神经网络模型的变量约简。

2 神经网络黑箱建模

为了便于说明问题，本文通过构造样本数据和系统模型，模拟实际生产过程。假设某系统固有模型：$Y=5X_1{}^3-10X_2{}^2+7X_3{}^2+8X_4+X_3X_4$，自变量矩阵 $X=(X_1,X_2,X_3,X_4,X_5,X_6)$，样本容量为 1500 组，其中 X_i 是(0,1)之间的随机数向量，X_5，X_6 为冗余变量。

选用 3 层 BP 前馈神经网络对该系统进行建模，将样本分为训练集和检验集，分别包含 1000 组、500 组样本。经反复训练最终确定网络的拓扑结构为 6-6-1，如图 1 所示，其中 $H_i(i=1,2\cdots,6)$为隐含层神经元。

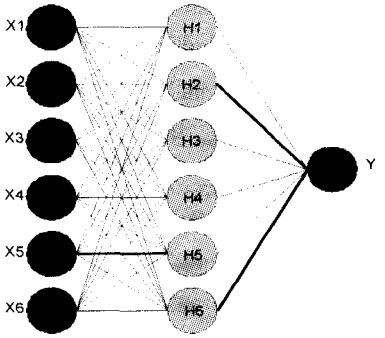

图 1 系统的神经网络模型

该模型训练集和检验集的绝对误差和相对误差曲线如图 2—图 5 所示。

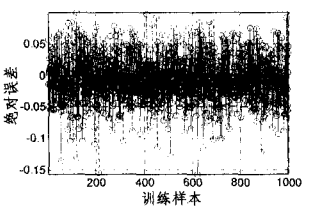

图 2 训练集绝对误差曲线

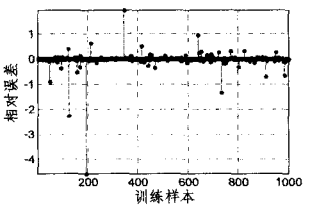

图 3 训练集相对误差曲线

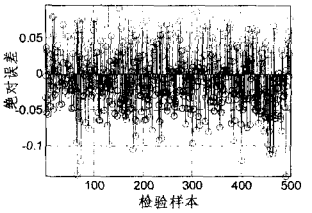

图 4 检验集绝对误差曲线

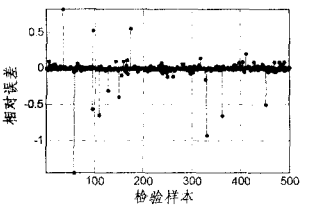

图 5 检验集相对误差曲线

将训练集和检验集的绝对误差 e 和相对误差 σ 按大小分别分为 3 类，绝对误差 1 类：$e\leqslant 0.02$；2 类：$0.02<e\leqslant 0.03$；3 类：$e>0.03$。相对误差 1 类：$\sigma\leqslant 2\%$；2 类：$2\%<\sigma\leqslant 5\%$；3 类：$\sigma>5\%$。绝对误差和相对误差落在每类中的样本个数占各自总样本数的百分比，如图 6、图 7 所示。

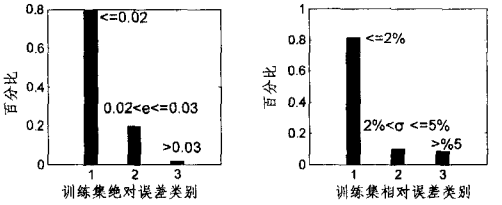

图 6 训练集不同类绝对误差和相对误差的百分比

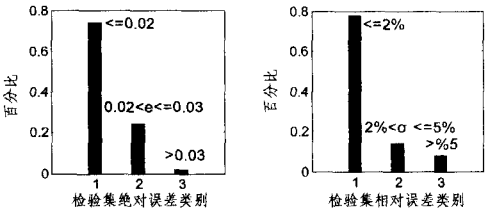

图 7 检验集不同类绝对误差和相对误差的百分比

以均方误差(MSE)的大小(见式(1))，来观察该模型的鲁棒性。给检验样本输入集分别加 2%、3%、4%和 5%的扰动，如 $x=x+x\cdot(-a+2a\cdot rand(1))$，其中 a 为扰动百分比，得到的均方误差如表 1 所列。

$$MSE=\frac{1}{n}\sum_{1}^{n}(y_i-\hat{y}_i)^2 \tag{1}$$

表 1 不同扰动下检验样本的 MSE

扰动	未加扰动	2%	3%	4%	5%
MSE	0.0018	0.0259	0.0491	0.0963	0.1404

通过以上分析可以得出：在系统模型包含冗余变量的同时，神经网络仍获得了较好的泛化能力和一定的可靠度及鲁棒性能。然而，神经网络模型的复杂度随输入变量的增加而快速增长[12]，并且神经网络建模属于黑箱建模，不能科学地解释输入变量对输出变量的重要程度。因此，包含冗余变量的神经网络，其建模精度、可靠性和鲁棒性不可能达到最佳。另外，对于软传感器模型而言，冗余变量会增加检测系统外围装置，严重浪费资源；对于企业而言，冗余变量不利于企业提高产品质量、降低生产成本和节能减排。因此，研究神经网络模型的透明化及变量约简具有重要的科学意义和学术价值。

3 神经网络模型的透明化及变量约简

3.1 神经网络释义图

Ozesmi[8]提出了神经网络释义图，该图提供了一种基于神经网络连接权的可视化解释方法。以图 8 为例进行说明，连接权的绝对值大小$|W_{ij}|(i=A,B;j=1,2,3)$用线的粗细表示：粗线表示较大的连接权，细线表示较小的连接权；线的

形状表示连接权的作用状态：实线代表着对输出起正作用的刺激状态，虚线代表着对输出起负作用的抑制状态。

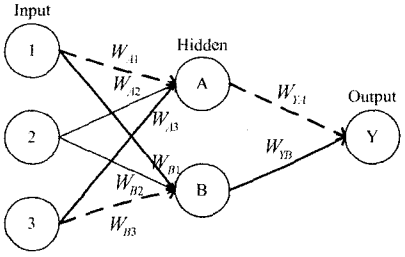

图 8　神经网络释义图

通过对连接权大小和状态的跟踪，研究者可以鉴别出单个变量或多个变量之间对输出变量的影响。对输出变量的正向刺激或负向抑制由两个因素决定：输入层到隐含层的连接权 W_{ij} 和隐含层到输出层的连接权 W_{Yi}。正向刺激的输入变量由正的 W_{ij} 和正的 W_{Yi}，或负的 W_{ij} 和负的 W_{Yi} 决定。负向抑制的输入变量由负的 W_{ij} 和正的 W_{Yi}，或正的 W_{ij} 和负的 W_{Yi} 决定。因此，W_{ij} 和 W_{Yi} 乘积的大小和状态决定了输入变量对输出变量影响的大小和状态。

虽然神经网络释义图实现了神经网络模型的可视化，其借助连接权的大小和状态可以定性地描述输入变量或输入变量之间与输出变量的关系，但该方法有两个明显的缺陷：(1)缺少输入变量对输出变量贡献率的定量分析；(2)随着输入变量维数和隐含层神经元数目的增多，模型结构的复杂度急剧增加，此时神经网络释义图对模型几乎没有了解释能力。这样，若再利用该方法来解释模型，就显得毫无价值。为此，下文依次利用连接权法和改进的随机化测验来分别解决以上两个问题。

3.2　连接权法

连接权法包含 4 个指标：输入层到隐含层和隐含层到输出层的连接权矩阵、输入层-隐含层-输出层连接权贡献度、综合连接权贡献度和相对贡献率。以图 8 为例进行介绍，具体算法如下。

(1)记录输入层到隐含层和隐含层到输出层的连接权矩阵，如表 2 所列。

表 2　连接权矩阵

	Hidden A	Hidden B
Input1	$W_{A1}=-0.7147$	$W_{B1}=0.6854$
Input2	$W_{A2}=0.1270$	$W_{B2}=0.3319$
Input3	$W_{A3}=0.8324$	$W_{B3}=-0.9157$
Output	$W_{YA}=-0.7577$	$W_{YB}=0.8650$

(2)计算输入层-隐含层-输出层连接权贡献度 C

输入层-隐含层-输出层连接权贡献度，表征每个变量通过隐含层神经元对输出的贡献大小。其值为输入层到隐含层的连接权和隐含层到输出层的连接权的乘积，其表达式为：

$$C_{ij}=W_{ij}\times W_{Yi},i=A,B;j=1,2,3 \tag{2}$$

例：$C_{A1}=W_{A1}\times W_{YA}=-0.7147\times(-0.7557)=0.5401$，表明输入变量 X_1 通过隐含层神经元 A 对输出 Y 的贡献度大小为 0.5401。输入层-隐含层-输出层贡献度如表 3 所列。

表 3　输入层-隐含层-输出层贡献度

	Hidden A	Hidden B
Input1	$C_{A1}=0.5401$	$C_{B1}=0.5929$
Input2	$C_{A2}=-0.0960$	$C_{B2}=0.2871$
Input3	$C_{A3}=-0.6290$	$C_{B3}=-0.7921$

(3)综合连接权贡献度 OI

OI 表征每个输入变量对输出变量的贡献大小。‘+’表示起正向刺激作用；‘−’表示起负向抑制作用。绝对值越大表示对输出的贡献度越大，其表达式为：

$$OI_j=\sum_{i=A}^{B}C_{ij},j=1,2,3 \tag{3}$$

例：$OI_{X1}=\sum_{i=A}^{B}C_{i1}=0.5401+0.5929=1.1330$，表明 X_1 对 Y 的综合贡献度为 1.1330。

(4)相对贡献率 RI

RI 表明每个输入变量整体对输出变量的重要程度，以百分比的形式给出。

$$RI_i=\frac{OI_i}{\sum_{i=1}^{3}|OI_i|}\times 100\% \tag{4}$$

若其大于 0，则表示该变量对输出变量起正作用；若其小于 0，则表示该变量对输出起负作用；若其等于 0，则表示该变量对输出变量没有影响。

计算的综合连接权贡献度 OI 和相对贡献率 RI，如表 4 所列。

表 4　综合贡献度(OI)和相对贡献率(RI)

	OI	RI
Input1	$OI_{X1}=1.1330$	$RI_{X1}=0.4127$
Input2	$OI_{X2}=0.1911$	$RI_{X2}=0.0696$
Input3	$OI_{X3}=-1.4211$	$RI_{X3}=-0.5177$

根据表 4 可以得出，X_1、X_2 对输出 Y 起正向刺激作用，相对贡献率分别为 41.27%和 6.96%；X_3 对 Y 起负向抑制作用，相对贡献率为−51.77%。因此，连接权法弥补了神经网络释义图的第一个缺陷，实现了输入变量对输出变量贡献率的定量分析。

3.3　改进的随机化测验

随机化测验实质上是一种剪枝技术，它根据统计理论对统计指标进行显著性计数检验，剔除不显著的指标，保留显著的指标，实现修剪的目的[11]。为了解决神经网络释义图的第二个问题，利用随机化测验修剪释义图，进一步增加其“可解释”能力，是一种有效的方法。

Olden[11] 利用“加森(Garson)算法”(见式(5))计算得到相对贡献率，进行随机化测验。由式(5)可知，加森算法利用绝对权重求解输入变量的相对贡献率，并没有考虑到连接权之间的方向，因此获得的相对贡献率和显著指标来解释输入变量的重要性时，会误导人们对模型的认知。

$$RI_i=\frac{\sum_{j=1}^{H}\frac{|W_{ij}||W_{jk}|}{\sum_{i=1}^{N}|W_{ij}||W_{jk}|}}{\sum_{i=1}^{N}\sum_{j=1}^{H}\frac{|W_{ij}||W_{jk}|}{\sum_{i=1}^{N}|W_{ij}||W_{jk}|}} \tag{5}$$

$$RI_i=\frac{\sum_{j=1}^{H}W_{ij}W_{jk}}{\sum_{i=1}^{N}|\sum_{j=1}^{H}W_{ij}W_{jk}|} \tag{6}$$

为此，本文改进了随机化准则，利用基于连接权法的输入变量相对贡献率(见式(6))进行随机化实验，并以相对贡献率和综合贡献度不显著变量的交集为依据进行变量选择。改进的随机化准则充分考虑了连接权之间的方向，具体步骤如下：

(1)根据标准化后的大量数据样本，构建一系列神经网络模型，每个神经网络模型都采用小随机数的初始权重和带有动量项及学习速率的训练方式。

(2)根据建立的一系列模型，选择出具有最佳预测性能的神经网络模型，并记录下该模型的初始权重和终止权重。根据终止权重和连接权法计算出标准的 C、OI、RI，操作如下：

1)计算输入层-隐含层-输出层连接权贡献度 C；

2)计算每个变量的综合连接权贡献度 OI；

3)计算每个变量的相对贡献率 RI。

(3)随机改变训练样本输出集的顺序。

(4)用改变顺序后的样本和(2)中记录的初始权重，重新训练神经网络模型，并记录模型的终止权重。

(5)大量重复(3)、(4)，并以 $COUNT$ 记录重复次数(如999次)，每次都根据(4)中记录的终止权重，计算(2)中的1)、2)、3)，得到随机化的 C、OI、RI。

(6)分别计算输入层-隐含层-输出层连接权贡献度 C、综合连接权贡献度 OI、相对贡献率 RI 的显著程度 P(即统计检验的概率)。

1)若其标准值大于0，则 $P=(N+1)/(COUNT+1)$，N 为随机化值大于等于标准值的个数；

2)若其标准值小于0，则 $P=(M+1)/(COUNT+1)$，M 为随机化值小于等于标准值的个数；

(7)若 $P\leqslant$显著水平 α(如0.05)，则保留该连接权，否则剔除。

(8)以综合贡献度和相对贡献率不显著变量的交集为依据剔除冗余变量。

4 仿真研究

4.1 神经网络释模型的可视化

图1的神经网络释义图，如图9所示，该图很显然存在3.1节所描述的问题。因此，下文利用连接权法定量计算该模型输入变量的贡献率，再利用改进的随机化测验来修剪神经网络释义图，进一步增加该神经网络模型的透明化程度。

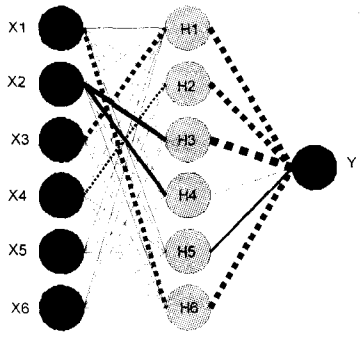

图9 系统的神经网络释义图(NID)

4.2 输入变量贡献率分析

应用连接权法得到图9神经网络模型的输入变量对 Y 的综合贡献度和相对贡献率，如表5所列。通过该表可以发现，冗余输入变量 X_5、X_6 对输出变量 Y 的综合贡献度和相对贡献率几乎为0。X_1，X_3，X_4 起着正向刺激作用，X_3 相对贡献率最高，为25.893%，其次为 X_4 和 X_1，分别为19.1601%、18.574%；X_2 起着负向抑制作用，相对贡献率为−36.237%；对 Y 相对贡献率最大的输入变量为 X_2，如图10所示。

表5 输入变量的综合贡献度和相对贡献率

	X_1	X_2	X_3
OI	0.489497	−0.95499	0.68238
RI	0.18574	−0.36237	0.25893
	X_4	X_5	X_6
OI	0.504943	0.001644	0.001924
RI	0.191601	0.000624	0.00073

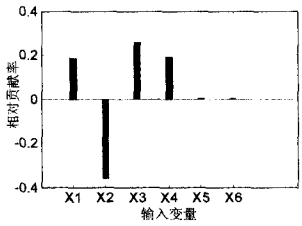

图10 输入变量的相对贡献率

并且根据输入变量对 Y 的综合贡献度，可增加神经网络释义图的透明度，如图11所示。黑色的输入神经元代表对 Y 起着正向刺激作用，综合贡献度为'+'；灰色的输入神经元代表对 Y 起着负向抑制作用，综合贡献度为'−'。

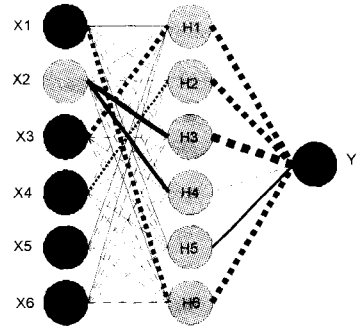

图11 基于连接权法的神经网络释义图

4.3 改进的随机化准则

对图11进行随机化测验，得到了输入-隐含层-输出连接权的随机化 P 值，如表6($\alpha=0.05$)所列，通过表6可知在共36条连接权中，只有4条连接权的统计重要性是显著的。输入变量的综合贡献度指标有4个是显著的，分别是 X_1，X_2，X_3，X_4；相对贡献度只有 X_2 是显著的。

表6 随机化的 P 值

	隐含层 H_1	隐含层 H_2	隐含层 H_3	隐含层 H_4	隐含层 H_5	隐含层 H_6	综合权重	相对贡献
	$P(C_{1i})$	$P(C_{2i})$	$P(C_{3i})$	$P(C_{4i})$	$P(C_{5i})$	$P(C_{6i})$	$P(OI_i)$	$P(RI_i)$
X_1	0.62275	0.24825	***0.0305***	0.38325	0.4235	0.54425	***0.00125***	0.1905
X_2	0.3905	0.468	0.44375	***0.01727***	0.48625	0.4755	***0.00025***	***0.04***
X_3	0.3265	***0.01325***	0.64925	0.65475	0.45875	0.63775	***0.00025***	0.10225
X_4	0.5305	0.46125	0.60125	0.70175	0.53475	0.61625	***0.00125***	0.202
X_5	0.8415	0.3555	0.69225	0.2365	0.5935	0.7545	0.50575	0.51125
X_6	***0.00425***	0.426	0.33675	0.50875	0.5185	0.60225	0.53525	0.5405

注：C_{ji} 代表隐含层神经元 $H_j(j=1,2,\cdots,6)$ 与输入神经元 $i(i=1,2,\cdots,6)$ 之间的输入隐含层输出连接权贡献度；OI_i 代表第 i 个输入神经元变量的的综合贡献度；RI_i 表示第 i 个输入神经元变量的相对贡献率；C_{ji}、OI_i、RI_i 的 P 值分别是基于3999次随机化实验计算得到的。加粗斜体的数字表示统计水平显著($\alpha=0.05$)。

根据表 6 的 P 值移除 NN 模型中不显著的连接权，得到了新的神经网络释义图，如图 12(α=0.05)所示。与图 11 相比，该神经网络释义图较简洁，透明化程度更高，更易于解释输入变量与输出变量之间的关系，如表 7 所列。

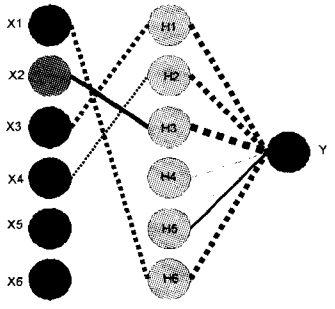

图 12　修剪后的神经网络释义图

表 7　X 与 Y 的关系

输入变量	通过隐含层	W_{ij}	W_{Yi}	对 Y 综合作用	神经元表现
X_1	H6	'－'	'－'	'＋'	黑色
X_2	H3	'＋'	'－'	'－'	灰色
X_3	H1	'－'	'－'	'＋'	黑色
X_4	H2	'－'	'－'	'＋'	黑色

根据 OI_i 和 RI_i 的显著程度指标，得到不显著输入变量的交集为(X_5，X_6)，剔除冗余信息后，得出了最佳的输入变量子集(X_1，X_2，X_3，X_4)。在本例中最佳输入变量子集恰好对应了 4 条显著的连接权。在其它模型中，显著的输入变量可能包含多条显著的连接权。

5　剔除冗余变量后的模型研究

采用本文第 1 节的研究方法对新变量集(X_1，X_2，X_3，X_4，Y)进行神经网络建模，最终的拓扑结构为 4-6-1。该模型训练集和检验集的绝对误差和相对误差曲线如图 13－图 16 所示。

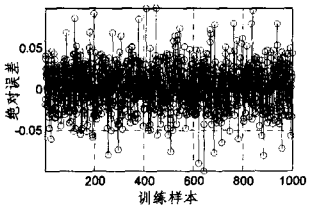

图 13　训练集绝对误差曲线

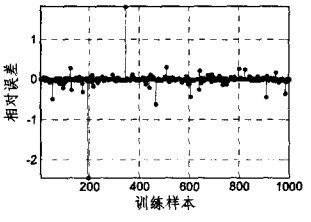

图 14　训练集相对误差曲线

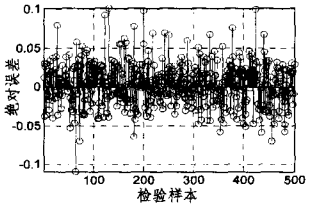

图 15　检验集绝对误差曲线

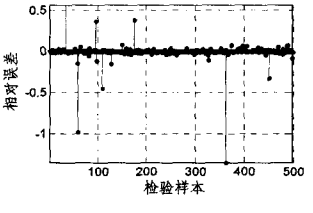

图 16　检验集相对误差曲线

绝对误差和相对误差落在每类中的样本个数所占各自总样本数的百分比，如图 17、图 18 所示。

冗余变量剔除前后，训练集和检验集不同类别的绝对误差和相对误差所占各自总样本数的统计百分比，如表 8、表 9 所列。

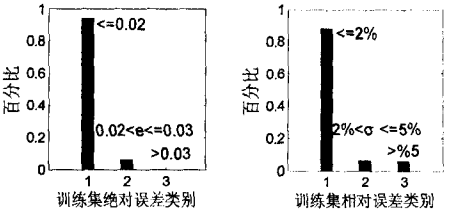

图 17　训练集不同类绝对误差和相对误差的百分比

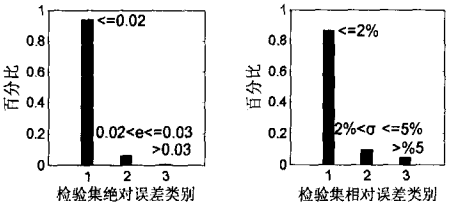

图 18　检验集不同类绝对误差和相对误差的百分比

表 8　训练样本绝对误差和相对误差的统计百分比

冗余变量	绝对误差百分比			相对误差百分比		
	1 类	2 类	3 类	1 类	2 类	3 类
X_1，X_2	$\leqslant$0.02	0.02$<e\leqslant$0.03	$>$0.03	$\leqslant$2%	2%$<\sigma\leqslant$5%	$>$5%
剔除前	0.792	0.194	0.014	0.816	0.099	0.085
剔除后	0.94	0.06	0	0.885	0.062	0.053

表 9　检验样本绝对误差和相对误差的统计百分比

冗余变量	绝对误差百分比			相对误差百分比		
	1 类	2 类	3 类	1 类	2 类	3 类
X_1，X_2	$\leqslant$0.02	0.02$<e\leqslant$0.03	$>$0.03	$\leqslant$2%	2%$<\sigma\leqslant$5%	$>$5%
剔除前	0.738	0.244	0.018	0.778	0.14	0.082
剔除后	0.938	0.058	0.004	0.864	0.094	0.042

剔除冗余后，检验样本在不同扰动下的 MSE 大小与变化如表 10 所列。

表 10　剔除冗余后检验样本的 MSE 与变化

扰动	未加扰动	2%	3%	4%	5%
MSE	0.00079	0.0214	0.0423	0.0842	0.1303
与剔除前比较	↓0.00101	↓0.0045	↓0.0068	↓0.0121	↓0.0101

通过以上分析，该系统模型的输入变量由剔除冗余前的 6 变为 4 个，有效地降低了模型的复杂度，提高了建模效率。另外，由表 8－表 10 可知，该系统模型的可靠性和鲁棒性也得到了明显的改善。

结束语　(1)利用神经网络释义图和连接权法，实现了神经网络模型的可视化和输入变量贡献率的定量计算。

(2)改进的随机化实验，通过修剪神经网路释义图，增加了 NN 模型的透明度；并以输入变量综合贡献度和相对贡献率不显著变量的交集为依据，可以有效地选出最佳变量子集，降低了模型复杂度，提高了模型的可靠性和鲁棒性。

(3)本文结合神经网络释义图、连接权法和改进的随机化测验 3 种方法，为 NN 模型的透明化及输入变量约简提供了一种新方法。

参 考 文 献

[1]　Wang Ai-jie, Liu Chun-shuang, et al. Modeling denitrifying sulfide removal process using artificial neural networks[J]. Journal of Hazardous Materials, 2009, 168(1): 1274-1279

(下转第 278 页)

图 3 和图 4 分别给出了 ORL 和 Yale 人脸数据库上，在每类不同训练样本数的情况下，LGME 与其它 4 种方法提取的人脸图像特征用于人脸识别时随特征维数变化的识别率曲线比较。表 1 给出了 ORL 和 Yale 人脸数据库上，在每类不同训练样本数的情况下，LGME 与其它 4 种方法提取的人脸图像特征用于人脸识别时的最高识别率及相应特征维数进行对比。由图 3、图 4 和表 1 可以看出，相比于 PCA、LDA、LPP、MFA 等特征提取方法，LGME 方法提取的人脸图像特征在用于人脸识别时，具有较高的识别率。这很容易理解，因为 PCA 主要保持了数据空间的全局结构信息；LDA 主要保持了数据空间的判别信息；LPP 主要保持了数据空间的流形结构信息；MFA 通过构建固有关系图和惩罚图来描述类内数据的紧密度和类间数据的分离度，使得提取的特征在保持数据空间的流形结构信息的同时，具有较强的判别力；而 LGME 同时使用类间数据的局部和整体间距信息，对类间数据分离度进行了充分描述，从而使得 LGME 方法提取的数据特征具有更强的判别力。

结束语 本文基于图嵌入框架，在构建惩罚图时采用全部的不同类样本数据对，并适当地强调间距较小的不同类样本数据对的作用，提出了一种局部和整体间距嵌入（LGME）方法，并将其用于提取人脸图像特征。与 MFA 不同，由于 LGME 同时使用类间数据的局部和整体间距信息，对类间数据分离度进行了充分的描述，使得 LGME 方法提取的数据特征具有更强的判别力。在 ORL 和 Yale 人脸库上的实验结果表明，与 PCA、LDA、LPP、MFA 等特征提取方法相比，本文提出的 LGME 方法提取的人脸图像特征在用于人脸识别时，具有较高的识别率，且更具鲁棒性。

参 考 文 献

[1] Turk M, Pentland A. Eigenfaces for recognition [J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71-86

[2] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19 (7): 711-720

[3] Tenenbaum J B, Silva V D, Langford J C. A global geometric framework for nonlinear dimensionality reduction [J]. Science, 2000, 290(12): 2319-2322

[4] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290(12): 2323-2326

[5] Shashua A, Levin A, Avidan S. Manifold pursuit: A new approach to appearance based recognition [C] // Proceedings of 16th International Conference on Pattern Recognition. Quebec City, Canada, 2002, 3: 590-594

[6] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation [J]. Neural Computation, 2003, 15(6): 1737- 1396

[7] He X F, Niyogi P. Loaclity preserving projections [C] // Proceedings of Neural Information Processing System. 2003: 153-160

[8] Gui J, Jia W, Zhu L, et al. Locality preserving discriminant projections for face and palmprint recognition [J]. Neurocomputing, 2010, 73(13-15): 2696-2707

[9] Zhang L M, Qiao L S, Chen S C. Graph-optimized locality preserving projections [J]. Pattern Recognition, 2010, 43(6): 1993-2002

[10] Chen H T, Chang H W, Liu T L. Local discriminant embedding and its variants [C] // Proceedings of 2005 International Conference on Computer Vision and Pattern Recognition. San Diego, USA, 2005(2): 846-853

[11] Yan S C, Xu D, Zhang B Y, et al. Graph embedding and extensions: A general framework for dimensionality reduction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51

[12] Zhang T, Tao D, Li X, et al. Patch alignment for dimensionality reduction [J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(9): 1299-1313

[13] Zhang S W, Lei Y K, Wu Y H. Semi-supervised locally discriminant projection for classification and recognition [J]. Knowledge-Based Systems, 2011, 24(2): 341-346

[14] Cheng M, Fang B, Tang Y Y, et al. Incremental embedding and learning in the local discriminant subspace with application to face recognition [J]. IEEE Transactions on Systems, Man, and Cybernetics, 2010, 40(5): 580-591

[15] Fang B, Cheng M, Tang Y Y, et al. Improving the discriminant ability of local margin based learning method by incorporating the global between-class separability criterion [J]. Neurocomputing, 2009, 73: 536-541

（上接第 251 页）

[2] 周孑民，朱再兴，刘艳军，等. 基于 Elman 神经网络的动力配煤发热量及着火温度的预测[J]. 中南大学学报：自然科学版，2011，42(12)：3871-3875

[3] El-Midany T T, El-Baz M A, Abd-Elwahed M S. A proposed framework for control chart pattern recognition in multivariate process using artificial neural networks[J]. Expert Systems with Applications, 2010, 37(1): 1035-1042

[4] 吴伟，李楠，郭茂耘. 粗糙集及 PSO 优化 BP 网络的故障诊断研究[J]. 计算机科学，2011，38(11)：200-203

[5] 满红，邵诚. 基于 Hammerstein-wiener 模型的连续搅拌反应釜神经网络预测控制[J]. 化工学报，2011，62(8)：2275-2280

[6] 胡包钢，王泳，杨双红，等. 如何增加人工神经网路模型的透明度[J]. 模式识别与人工智能，2007，20(1)：72-83

[7] Salari D, Daneshvar N, Aghazadeh F, et al. Application of artificial neural networks for modeling of the treatment of wastewater contaminated with methyl tert-butyl ether (MTBE) by UV/H_2O_2 process[J]. Journal of Hazardous Materials, 2005, B125(1): 205-210

[8] Ozesmi S L, Ozesmi U. An artificial neural network approach to spatial habitat modeling with interspecific interaction[J]. Ecological Modelling, 1999, 116(1): 15-31

[9] Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models[J]. Ecological Modeling, 2003, 160(1): 249-264

[10] Olden J D, Joy M K, Death R G. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data[J]. Ecological Modeling, 2004, 178(1): 389-397

[11] Olden J D, Jackson D A. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks[J]. Ecological Modeling, 2002, 154 (1): 135-150

[12] 张昭昭，乔俊飞，韩红桂. 一种基于神经网络复杂度的修剪算法[J]. 控制与决策，2010，25(6)：821-830