

前馈过程神经网络的网络结构与泛化能力^{*})

葛利^{1,2} 陈广胜³

(哈尔滨商业大学计算机与信息工程学院 哈尔滨 150056)¹

(哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001)² (东北林业大学材料科学与工程学院 哈尔滨 150040)³

摘要 基于提高过程神经网络泛化能力的角度,对前馈过程神经网络网络结构对泛化能力的影响进行研究,得出以下结论:其过程神经元隐层(时变隐层)起主要作用,一般神经元隐层(非时变隐层)并非是必须的,对于相同特征的样本,过程神经元对样本特征的抽取能力远远高于传统神经元。给出了一个基于提高泛化能力的前馈过程神经网络网络结构构造算法,并应用一个实例验证了其有效性。

关键词 过程神经网络,网络结构,泛化能力,网络结构构造算法

Network Structure and Generalization Ability of Feedforward Process Neural Networks

GE Li^{1,2} CHEN Guang-sheng³

(College of Computer and Information Engineering, Harbin University of Commerce, Harbin 150056, China)¹

(College of Computer science and Technology, Harbin Engineering University, Harbin 150001, China)²

(College of Material Science and Engineering, Northeast Forestry University, Harbin 150040, China)³

Abstract The influence of structure of feedforward process neural networks on its generalization ability was studied, based on raising generalization ability of feedforward process neural networks. The results were obtained: the process neuron hidden layer operates important function, the general neuron hidden layer isn't absolutely necessary, the process neuron's extractive ability on stylebook character outclasses tradition neuron. Based on improving generalization ability of feedforward process neural network, an algorithm for constructing network structure of feedforward process neural networks was proposed. The effectiveness of the algorithm is proved by an example.

Keywords Process neural networks, Structure of network, Generalization ability, Algorithm constructing network structure

过程神经网络是人工神经网络在时间域上的一种扩展,在文献[1,2]被首次提出,它在实际中对于求解与过程有关的众多问题有着一定的应用价值。由于这种模型提出时间较短,尚有许多问题有待于研究和完善,特别是学习速度和泛化能力问题,它们是评价一种神经网络性能的两个重要指标,而网络泛化能力是衡量网络实际应用能力的关键因素。本文对于前馈过程神经网络的网络拓扑结构进行了分析,对网络中过程神经元隐层数、隐层节点个数,一般神经元隐层数及隐层节点个数对网络泛化能力的影响进行了研究,在此基础上,给出一种基于提高网络泛化能力的前馈过程神经网络网络结构构造的算法。

1 前馈过程神经网络网络结构分析

前馈过程神经网络是由若干个过程神经元按一定的拓扑结构组成的前向网络。一类前馈过程神经网络拓扑结构如图1所示,其输入为时变函数,输出为常量,拓扑结构为 $n-m-l-1$ 。输入层有 n 个节点,用于完成 n 个时变函数向网络的输入;第一隐层为过程神经元隐层(时变隐层),由 m 个节点单元组成,第二隐层为一般神经元隐层(非时变隐层),有 l 个节点单元;第四层为输出层,输出为 y 。

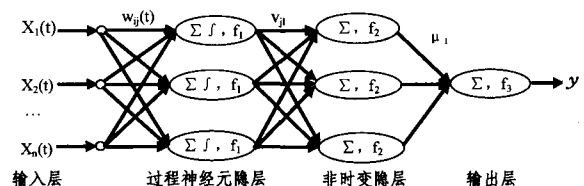


图1 过程神经网络

前馈过程神经网络既具有传统前馈神经网络的特点,又兼有过程神经网络本身的特色。研究表明,对于传统前馈神经网络而言,神经网络的隐层作为输入模式的内部表示,其作用是完成对一类输入模式中所含的区别于其他输入模式的特征提取^[3];而对于过程神经网络而言,与传统神经网络输入是与时间无关的常量相比,过程神经网络输入是一个时变函数(或称过程),函数中所包含的系统信息特征显然优于传统神经网络中的几何点式输入,而对于过程(输入函数)中系统特征的提取过程由过程神经元隐层完成,过程神经元隐层由若干过程神经元组成,过程神经元结构如图2所示,其特征提取的过程由(1)式完成。

过程神经元输入与输出之间的关系为:

$$y = f(w(t) \oplus x(t)) \otimes K(\cdot) - \theta \quad (1)$$

^{*})国家自然科学基金资助项目(30671645),黑龙江省科技计划资助项目(GC05B605,GC05A118),黑龙江省自然科学基金资助项目(F200506)。葛利 副教授,博士研究生,主要研究方向为神经网络、进化计算、数据库;陈广胜 副研究员,博士,主要研究方向为神经网络木材材质预测、小波无损检测。

这里“ \oplus ”表示某种空间聚合运算，“ \otimes ”表示某种时间（过程）聚合运算，其中，空间聚合运算“ \oplus ”可以选择一般的加权求和形式，还可以选择其他合理方式构成，激励函数 f 也可以选取任何有界函数的形式^[1]。

与传统神经网络中一般神经元相比，过程神经元既包含了一般神经元隐层具有的空间聚合（多输入）运算，又考虑了时间的累计效应，增加了时间聚合运算“ \otimes ”，过程神经元隐层不仅具有一般神经元对空间特征的提取功能，又具有一般神经元所不具备的、对应于系统时变函数输入的对时间累积特征进行提取的作用，对样本过程模式中时间特征的提取依靠时间聚合运算“ \otimes ”和激励函数完成，而对空间累积效应的提取由空间聚合运算“ \oplus ”和激励函数来完成，二者的共同作用完成了对过程神经元整体特征的提取，因此，过程神经元隐层既具有一般神经元隐层对样本离散常量特征的提取功能，又具有过程神经元所特有的对样本过程模式特征提取的功能。

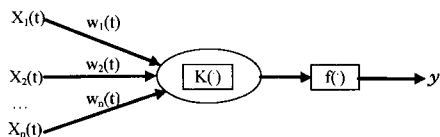


图2 过程神经元示意图

其次，过程神经网络的输入是一个过程即依赖于时间的函数的特点。与传统神经网络输入为与时间无关的常量相比，过程神经网络输入所采用的时变函数所包含的整体信息特征量显然优于传统神经网络的离散式的数据输入，传统神经网络理论上的几何点式的瞬时输入，只是对其的一种近似模拟，在实际应用中，是不可能实现的；因此，从输入样本的形式上考虑，过程神经网络可处理的样本形式更为高级，单个样本所包含的信息量更大，当然，对时变样本的处理过程和作用机理也更为复杂。

2 网络结构与泛化能力

2.1 过程神经元隐层数与传统神经元隐层数

对于传统神经网络隐层数的选取问题，一直以来还没有完善的理论支持^[3]。对于隐层的层数，许多学者作了理论上的研究，Lippmann 和 Cyberko 指出：有两个隐层，就能解决任何形式的分类问题^[3]，hornik^[4]等人的研究进一步指出：有一个隐层的神经网络，只要隐节点足够多，就可以以任意精度逼近一个非线性函数。既然一个一般（非时变）神经元隐层可以以任意精度逼近一个非线性函数，并且过程神经元隐层（时变隐层）具有一般神经元隐层的对离散数据特征的抽取功能，那么，在过程神经网络结构中，一般神经元隐层是否是必需的呢？为此，笔者选取 $y = at \sin(bt+c) + dt \cos(et+f)$ 产生输入时变函数，其中 $a, b, c, d, e, f \in (-\infty, +\infty)$ ，其值由随

表1 测试结果

网络序号	隐层数	隐层节点数	训练误差	测试误差
1	2	6,6	0.0999	0.0675
2	2	4,4	0.0999	0.0386
3	2	3,3	0.1000	0.0350
4	1	6	0.1000	0.0454
5	1	4	0.0999	0.0314
6	1	3	0.0999	0.0293
7	1	2	0.0999	0.0323

机函数产生，构成 16 个样本，选取 13 个样本作为过程神经网络训练集，3 个样本作为测试集，测试集独立于训练集，采用上述相同的训练集和测试集，训练、测试几种不同结构的网络，学习精度 $\epsilon \leq 0.1$ ，结果见表 1，其中，第一隐层均为过程神经元隐层，第二隐层均为一般神经元隐层。

从实验结果可以看到，在满足相同的学习精度下，具有两个神经元隐层（第一隐层为过程神经元隐层，第二隐层为一般神经元隐层）的过程神经网络泛化能力均低于只有单过程神经元隐层网络的泛化能力，并且，在训练误差相同的情况下，随着隐节点数和隐层数的增加，测试误差不断增大；由此可见，第 1—第 4 种网络结构均存在不同程度的“过拟合”现象，其中第 1 种网络“过拟合”情况最为严重，网络泛化能力最低。在隐层数相同的情况下，网络的泛化能力随着隐节点数的减少而不断升高，在单过程神经元隐层节点数降至 3 时泛化性能最高。两个隐层的情况下，当隐节点数降至 4,4 时，网络的泛化性能有很大的提高，特别是，将序号为 1 和 4,2 和 5,3 和 6 的网络结构对比可知，在第一层节点数相同的前提下，第二层节点数的等数量增加，测试误差并没有明显变化，再将序号为 3,4 的网络结构比较，在两层总节点数相同的情况下，第一隐层节点数的不同，导致网络的测试误差差异较大，且第 3 号网络的测试误差与第一隐层节点数相同的第 6 号网络接近。为进一步验证以上实验结果，仍采用上述方法随机产生多组实验数据，在其他实验条件不变的情况下，实验结果仍存在上述特征，其原因在于过程神经元不仅具有其特有的对样本过程模式特征提取的功能，而且还具有一般神经元对样本离散常量特征的提取功能，并且其所具有的对样本离散常量特征提取功能并没有因此减弱，由此可见，在过程神经网络结构中，第一隐层——过程神经元隐层起主要作用，依照传统神经网络 hornik 有关一个隐层的结论可知，过程神经网络结构中的一般神经元隐层并非是必需的。

再取一组实测数据进行预测实验，共组成 18 个样本，以其中 15 个样本作为训练集，另外 3 个样本作为测试集，网络学习精度 $\epsilon \leq 0.03$ ，训练过程神经网络，再采用同样的训练、测试样本及学习精度，应用传统神经网络进行训练，输入节点个数与离散数据的输入个数相同，输出节点个数相同，以保证两种网络使用完全相同的训练和测试样本，两种神经网络网络结构预测值误差结果对比如表 2 所示。

表2 两种网络结构预测结果对比

网络类型	网络结构	测试误差 (sse)
传统神经网络	1 个一般神经元隐层 20 个隐节点	0.000976
过程神经网络	1 个过程神经元隐层 2 个隐节点	0.000730

从表 2 可以看到，在上述相同条件下，过程神经网络的测试误差低于传统神经网络（测试误差依据平方和误差 sse 衡量），网络的泛化性能和预测结果均优于传统神经网络的结果，并且在相同的学习精度及样本个数下收敛，过程神经网络所需的过程神经元隐层节点个数远远少于传统神经元的一般神经元隐层节点个数。过程神经网络的测试误差在一个过程神经元隐层具有 2 个隐节点的网络结构时达到最小，在增大过程神经元隐层节点数，测试误差出现下降，这说明具有两个隐节点的网络已可以对 15 个过程样本进行很好的拟合，且不存在“过拟合”现象，即在过程神经网络中，隐节点个数在远小于样本个数的情况下，仍可以收敛到指定精度，这说明对于相

（下转第 150 页）

[7] 胡峰,代劲,王国胤. 一种决策表增量属性约简算法[J]. 控制与决策, 2007, 32(3): 268-277

[8] Liu H, Yu L. Feature Selection for Data Mining [EB/OL]. http://www.public.asu.edu/~huanliu/feature_selection.html,

[9] Dorigo M L M. Ant colony system: a cooperative learning approach to the traveling salesman problem[J]. IEEE Transactions on Evolutionary Computing, 1997, 1(1): 53-56

[10] 吴春明, 陈治, 姜明. 蚁群算法中系统初始化及系统参数的研究[J]. 电子学报, 2006, 34(8): 1530-1533

(上接第 138 页)

同特征的样本,过程神经元对样本的特征抽取能力大大高于传统神经元,这一结果也从侧面验证了上述结论。

2.2 隐节点数

对于过程神经网络来说,过程神经元隐层节点数的选取无疑是十分重要的。过程神经元隐节点数选取过小,将导致对样本过程模式特征提取能力太弱,训练无法收敛;节点数选取过多,会导致两个问题:一是网络的计算量增大,由于过程神经网络自身特点,已产生较大的计算量,再增加网络计算量,导致的可能结果是,对整个过程神经网络性能的整体影响;二是由于抽取能力太强而产生的“过拟合”现象,而对于过程神经网络而言,“过拟合”现象的出现,会导致较传统神经网络而言泛化误差的迅速降低。其原因在于:过程神经网络输入时变函数的产生,是由多个几何点式的常量经函数拟合得到,而拟合的过程不可避免地存在误差,因此,输入时变函数,在包含优于传统神经网络常量输入的大量有用信息特征的同时,无疑也增加了噪声,而噪声的产生多少与具体的函数拟合方法和精度有关,但从噪声的量上看,过程神经网络输入函数所包含的噪声量无疑应大于传统神经网络常量输入的情况,因此,当过程神经网络一旦产生“过拟合”现象时,会出现随着训练误差的逐渐变小,测试误差迅速增大的情况,因为过程神经元隐层所具有的对过程模式和空间累积效应的双重特征提取功能,导致网络对输入函数中所包含的大量噪声的快速提取,出现了网络泛化性能的快速下降。

表 3 测试误差对比

网络序号	隐层数	隐层节点数	最小误差要求	测试误差
1#	2	4,4	0.10	0.0386
2#	2	6,6	0.10	0.0675
3#	2	6,6	0.05	0.3067

从表 3(各实验条件同表 1)中看到,当网络出现“过拟合”现象时,随着隐层节点数增加和最小误差要求的提高,网络泛化能力迅速下降。

3 前馈过程神经网络网络结构构造算法

基于上述讨论结论,给出前馈过程神经网络网络结构构造算法:

- 1)选择单过程神经元隐层;
- 2)选择足够少的过程神经元隐节点数;
- 3)训练网络;
- 4)训练误差是否满足要求,如果满足转 6),否则转 5);
- 5)增加隐节点数或增加一个一般神经元隐层,转 3);
- 6)挑选测试集样本;
- 7)测试误差是否满足要求,如果满足转 9),否则转 8);
- 8)增加样本训练集的数量,转 3);
- 9)结束。

其中步骤 5)可选择性地增加过程神经元隐节点数或增加一个一般神经元隐层。增加一个过程神经元隐节点出现“过拟合”现象时,可采用增加若干个一般神经元隐节点的方法避免“过拟合”现象。算法中,要求训练集与测试集的数据应能够很好地表示输入域的整体情况,测试集与训练集应相

互独立。

4 应用实例

在林业木材科学研究中,木材生长密度是进行木材合理利用决策的重要依据。选取某一生长条件下的特定树种作为样本,对木材生长密度进行等距采样,根据木材林地条件对数据进行合理的早晚材划分,得到木材生长密度的时间序列,通过动态调整网络过程神经元的互连权函数和一般神经元的权值,拟合时间序列数据内在的映射关系,实现对木材生长密度的预测。

选取具有代表性的、共有 30 个年轮的某地区木材生长密度数据,共组成 18 个样本。过程神经网络结构构造过程如下:

- 1)初始化时,网络层数为 1,即只有 1 个过程神经元隐层,过程神经元隐层节点数为 2,学习精度 $\epsilon < 0.03$,测试误差限用平方和误差 sse 表示,取值为 0.005。以其中 14 个样本作为训练集,另外 3 个样本作为测试集,训练 132 次收敛,利用 4 测试样本预测后的平方和误差 sse 为 $0.0077 > 0.005$,测试误差不满足要求。
- 2)合并测试集中的一个样本到样本集中,仍然使用上述网络结构训练,经 173 代进化后收敛,对 3 个测试样本进行生长密度预测结果见表 4,平均绝对误差为 0.0134,平均相对误差为 1.51%,平方和误差 sse 为 $0.00073 < 0.005$,测试误差满足要求,网络构造和训练结束。

表 4 测试样本预测结果

测试集样本序号	1	2	3
实测密度值	0.8800	0.8460	0.8960
预测密度值	0.8646	0.8488	0.8740
绝对误差	0.0154	0.0028	0.0220
相对误差(%)	1.7500	0.3310	2.4550

结束语 对过程神经网络的结构与泛化能力进行了分析,得出了如下结论:其过程神经元隐层起主要作用,一般神经元隐层并非是必需的,对于相同特征的样本,过程神经元对样本的特征抽取能力大大高于传统神经元;给出了一个基于提高泛化能力的前馈过程神经网络网络结构构造算法,并应用一个实例验证了其有效性。同时与传统神经网络类似,样本的数量、质量也不可避免地对过程神经网络的泛化能力产生影响,除此以外,过程神经网络中时变输入函数的拟合精度,基于正交基展开的过程神经网络中正交基的选取,对过程神经网络的泛化能力都有一定的影响,而对于影响过程神经网络泛化能力的上述因素的研究,还有待于进一步深入探讨。

参 考 文 献

[1] 何新贵,梁久祯. 过程神经网络的若干理论问题[J]. 中国工程科学, 2000, 12(2): 40-44

[2] 何新贵,梁久祯,许少华. 过程神经元网络的训练及其应用[J]. 中国工程科学, 2001, 4(3): 31-35

[3] 焦李成. 神经网络系统理论[M]. 西安:西安电子科技大学出版社, 1990

[4] Hornik, Stinchcombe K M, White H. Multi-Layered Feed-Forward Neural Networks are Universal Approximations [J]. Neural networks, 1990, (2): 359-366