前向神经网络隐含层节点数的一种优化算法*)

夏克文　李昌彪　沈钧毅
(西安交通大学电子与信息工程学院 西安 710049)

摘　要　由于前向神经网络隐含层节点数的确定尚无理论依据，为此提出一种基于黄金分割原理的优化算法，首先确定网络隐含层节点数频繁出现的区间范围；将网络总误差作为试验结果，然后利用黄金分割法搜索其区间中的理想数值；兼顾高精度的需要，将隐含层节点数频繁出现的区间作拓展，可以求得逼近能力更强的节点数。算法分析和仿真例子表明，此优化算法是切实可行的，不仅能找到理想的隐含层节点数，而且能起到节省成本、提高搜索效率等功效。

关键词　前向神经网络，隐含层节点数，黄金分割，优化算法

An Optimization Algorithm on the Number of Hidden Layer Nodes in Feed-forward Neural Network

XIA Ke-Wen　LI Chang-Biao　SHEN Jun-Yi
(School of Electronic & information Engineering, Xi'an Jiaotong University, Xi'an 710049)

Abstract　There is not any theory able to design the number of hidden layer nodes in feed-forward neural network, so an optimization algorithm is presented based on the principle of golden section, first determining the interval which frequent appears the number of hidden layer nodes, using the total error of network as the result of experiment, then searching the perfect number of hidden layer nodes in the interval by the method of golden section. Considered the need of high precision, the more perfect number of hidden layer nodes enhancing the approximation capability of network can be searched in the extended range of interval. The algorithm analysis and the simulation experiment show the optimization algorithm not only can design the perfect number of hidden layer nodes, but also can save cost and improve search efficiency.

Keywords　Feed-forward neural network, Number of the Hidden layer nodes, Golden section, Optimization algorithm

1　引言

前向神经网络能够以任意精度逼近任意连续函数及平方可积函数[1,2]，而且可以精确实现任意有限训练样本集的拟合[3,4]。在实际应用中，如何估计隐含层神经元数目，一直是确定前向网络结构的难点和关键，而且目前尚无严格的理论依据。对此，人们提出了不少算法，其中"裁减算法"[5]是通过不断修剪并最终确定网络结构，但这显然会增加训练过程中的计算量，从而大大降低收敛速度。因此探讨实用的优化算法是非常必要的。

理论上已经证明3层前向神经网络就可足以逼近任意连续函数[6]，所以只需对3层前向网络给出一种较好的隐含层节点数的估算方法即可。尽管围绕此问题进行研究和发表的文章甚多，但是所得出的结论却千差万别[6]。一般地，随着隐含层节点数目的增多，网络的逼近能力得到提高，但网络的泛化能力会降低。为了兼顾网络的逼近能力和泛化能力，在隐含层节点数的设计中，尽量能够设法用规模较小的网络去逼近训练样本，这样使网络具有较好的泛化能力[5]。为此，作者在总结前人的经验上，提出一种实用的优化算法，主要工作和创新如下：

(1)根据实践经验，提出隐含层节点数"频繁"出现的一个区间范围的估算公式。

(2)提出基于黄金分割原理的搜索寻优方法，起到节省成本、提高搜索效率等功效。

(3)考虑到高精度的需求，提出根据黄金分割原理来拓展隐含层节点数所在区间的办法，再进行搜索寻优。

2　优化原理与算法

2.1　黄金分割法的基本原理

黄金分割法，即0.618法，是单峰值优化问题的一种求解方法。用这种方法能解决单一变量的极值问题或者以较少的试验次数、较低的成本迅速找到最优方案，因而它是一种较先进的常用的优选法，已在生产实践和科学试验中有着广泛的应用。

从数学的角度讲，黄金分割法就是把任意一条长为 L 的直线段分割成二部分，其中一部分的长为 $x=0.618\times L$，这种分割法叫黄金分割法。在生产实践和科学试验中，我们所说的黄金分割法是指在试验范围0.618处作一次试验，再在其对称点(即0.382处)作第二次试验，比较两点试验结果，去掉"劣"点部分，保留"优"点部分。然后在留下部分继续进行类似试验，再舍"劣"取"优"，逐步缩小试验范围，直至取得满意的结果。应用0.618法，每次至少可以去掉试验范围的0.382。因此，可以用较少的试验次数迅速找到最佳点。

黄金分割法的具体步骤如下：

(1)先在区间$[a,b]$上确定第一个试验点的位置 $x_1=0.618\times(b-a)+a$，并记录第一个试验点的试验结果。

(2)确定第二个试验点位置，即 x_1 的对称点处 $x_2=0.382\times(b-a)+a$，并记录试验结果。

(3)比较两次试验结果，舍"劣"取"优"。如果 x_1 试验结果 y_1 为优，就去掉$[a,x_2)$，留下$[x_2,b]$；如果 x_2 试验结果 y_2

*)基金项目：国家自然科学基金项目(编号：60173058)、中国石油天然气集团公司"九五"重点攻关项目(编号：2001-6-1)资助。夏克文　博士，博士后，高级工程师，研究生导师，主要从事计算智能与数据挖掘的研究。

为优，就去掉$(x_1,b]$，留下$[a,x_1]$；如果 x_1、x_2 试验结果一样，就去掉$[a,x_2)$和$(x_1,b]$，留下$[x_2,x_1]$。

(4)在留下部分重复上述试验，直到满意的结果为止。

2.2 隐含层节点数的优选法

设3层前向神经网络的输入层、隐含层和输出层的节点数分别为 n_i、n_h 和 n_o，在隐含层节点数的设计中，不少作者采用如下公式：

$$n_h=\sqrt{n_i+n_o}+m \tag{1}$$

其中 m 为常数，$m\in[1,10]$。

为求取合适的隐含层节点数，在大量试验和实际应用中，我们发现：比较理想的隐含层节点数，一般"频繁"出现在如下区间或范围：

$$a=\frac{n_i+n_o}{2}\leqslant n_h\leqslant(n_i+n_o)+10=b \tag{2}$$

再逐一计算全部可能构成的网络对同一样本集在一定迭代次数下的拟合精度或总误差来从中择优。

设神经网络的总误差为

$$E=\frac{1}{2}\sum_{j=1}^{n_o}(Q_j-O_j)^2 \tag{3}$$

这里 Q_j，O_j 分别为网络的期望输出和实际输出。

由上可知，若 a、b 数值不是很大，在区间$[a,b]$上逐一全部计算来选择最佳的隐含层节点数，其工作量不是很大。但是，当 a、b 数值较大时，以及当在$[a,b]$范围内万一找不到理想结果且需要扩展区间进行寻优时，显然，以上那种逐一全部计算寻优的方法就不可取了。

为此，我们提出一种基于黄金分割原理进行隐含层节点数的寻优算法。算法的主要思想是：首先在$[a,b]$内寻找理想的隐含层节点数，这样就充分保证了网络的逼近能力和泛化能力。为满足高精度逼近的要求，再按黄金分割原理拓展搜索区间，即得到区间$[b,c]$(其中 $b=0.618\times(c-a)+a$)，在区间$[b,c]$中搜索寻优，则得到逼近能力更强的隐含层节点数。在实际应用根据要求，从中选取其一即可。

这样设计的优点是：第一，充分保证能找到隐含层节点数"频繁"出现的区间中的理想值；第二，考虑到高精度的要求，将"频繁"出现的区间进行拓展可以求得逼近能力更强的隐含层节点数。

具体步骤如下：

(1)按照式(2)求得 a、b，在区间$[a,b]$内进行黄金分割优选，找到理想的隐含层节点数。其中试验结果为网络在一定迭代次数下对同一样本集训练后的总误差 E，误差小为"优"，误差大为"劣"。

最后得到隐含层节点数 n_h，且满足如下关系：

$$E(n_h)=\min\{E(a),E(b),E(x_1),E(x_2),\cdots\}$$

这里，$x_1,x_2,\cdots$为$[a,b]$中的黄金分割试验点位置。

(2)将区间$[a,b]$进行拓展，得到区间$[b,c]$，这里 c 是由黄金分割关系"$b=0.618(c-a)+a$)"得到。再按照步骤(1)搜索理想的隐含层节点数。

(3)比较所得两区间的隐含层节点数的逼近能力和泛化能力，根据实际需要从中选取其一。

3 仿真试验与应用分析

为了检验本文提出的隐含层节点数优化算法的应用效果，下面举例进行详细的仿真分析。

如表1所示，样本取自 MX 气田某一关键井储层的测井数据，即条件属性由自然伽码(c_1)、补偿中子(c_2)、声波时差(c_3)、深测向(c_4)等测井信息组成；ϕ 为期望输出，数据取自岩芯分析孔隙度(单位为%)。

表1 样本数据表

编号	c_1	c_2	c_3	c_4	ϕ(%)
1	57.79	10.23	61.93	31.33	5.83
2	63.40	7.18	57.93	62.19	1.05
3	68.79	10.39	63.76	38.76	3.88
4	68.22	10.35	62.18	35.67	4.68
5	67.22	11.68	61.05	37.60	4.41
6	58.07	13.05	66.36	26.98	6.37
7	62.05	9.65	64.65	26.14	5.99

设3层前向神经网络的输出层节点为1个，记为O，即输出的是网络计算的孔隙度数值。

输入层节点为4个，即 c_1，c_2，c_3，c_4，为保证网络学习训练稳定，各属性数据须作如下归一化处理：

$$x'=(x-x_{min})/(x_{max}-x_{min})$$

这里 $x\in[x_{min},x_{max}]$，x_{min}、x_{max} 分别为最小值和最大值。考虑到各属性的变化范围，其取值区间均为[0,100]，即 $x_{min}=0$，$x_{max}=100$。

在隐含层节点数的搜索寻优中，令"试验结果"为神经网络的总误差，其中激励函数取 S 型函数，学习训练采用既具有高斯-牛顿法的局部特性又具有梯度法的全局特性的著名的 Levenberg-Marquardt 算法[7]。设搜索步长为0.001，取控制参数 $\varepsilon=10^{-6}$，当迭代至第 $k+1$ 次，如有$|E(k+1)-E(k)|\leqslant\varepsilon$，或者迭代次数超过 10000 次时，则迭代停止，输出总误差。

下面按照2.2节中的优化算法来搜索理想的隐含层节点数。

(1)因为 $n_i=4$、$n_o=1$，则有求得 $a=2$，$b=15$。在区间[2,15]上确定第一个试验点的位置 $x_1=0.618\times(15-2)+2=10$，并记录隐含节点数为10时的试验结果，即网络总误差为0.000000987。

(2)确定第二个试验点位置，即 $x_2=0.382\times(15-2)+2=7$，并记录试验结果，即为0.000033940。

(3)比较两次试验结果，舍"劣"取"优"。因为 x_1 试验结果为"优"，则留下区间[7,15]。

(4)在留下的区间[7,15]里重复上述试验，最后得到 $n_h=10$ 为区间[2,15]中理想的值，有

$$E(10)=\min\{E(2),E(15),E(10),E(7),E(12),E(9),E(11)\}$$

(5)利用黄金分割法求得拓展区间为[15,23]，此区间右端点"23"是按等式"$15=0.618\times(c-2)+2$"计算出来的。

(6)仿照上述步骤(1)～(4)，即可求得区间[15,23]中的理想值为20，网络的总误差为 4.15×10^{-7}，有 $E(20)=\min\{E(15),E(23),E(20),E(18),E(21),E(19)\}$。

(7)比较隐含层节点数为10和20时的逼近能力和泛化能力。节点数为20时网络的误差小、逼近能力强，但网络的泛化能力没有节点数为10的强。在实际应用中根据需要，从中择优选一个即可。

为全面而方便地进行对比分析，我们将隐含层节点在整个区间[2,23]中的所有网络的学习训练结果列表出来，如表2所示。

表 2 神经网络仿真的迭代次数和总误差

隐节点数	迭代次数	总误差	隐节点数	迭代次数	总误差
2	10000	0.000949444	13	10000	0.001175002
3	10000	0.000055523	14	10000	0.000004807
4	10000	0.000030904	15	10000	0.000212522
5	8776	0.000105445	16	10000	0.000028057
6	10000	0.000012537	17	1473	0.000113128
7	733	0.000033940	18	10000	0.000222287
8	10000	0.000054020	19	10000	0.000020770
9	10000	0.000013394	20	1446	0.000000415
10	369	0.000000987	21	2784	0.000000999
11	10000	0.000018163	22	1585	0.000016172
12	4051	0.000022619	23	4565	0.000001000

由表 2 可知，隐含层节点数为 20 时的网络逼近精度最好，其次是节点数为 10，再次是 21、23。显然与本文提出的优化算法的仿真结果相吻合，这说明了本文的优化算法的应用效果是令人满意的，而且还具有试验次数少、精度高、简单、直观、有效、节省成本等优点。

结论 从黄金分割原理出发，提出的一种前向网络隐含层节点数的优化算法是切实可行的，不仅能搜索到理想的隐含层节点数，而且能起到节省成本、提高搜索效率等功效。算法中既充分考虑到能保证在隐节点数“频繁”出现的区间中搜索出理想的结果；又兼顾实际应用中高精度的需要，可将搜索区间做进一步拓展以求得逼近能力更强的隐含层节点数。

该优化算法简单、实用，具有很好的应用价值。

参 考 文 献

1 Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks, 1989, 2(5): 359～366

2 Blum E K, Li L K. Approximation theory and feed forward networks. Neural Networks, 1991, 4(4): 511～515

3 Huang S C, Huang Y F. Bounds on number of hidden neurons in multiplayer perceptrons. IEEE Trans on Neural Networks, 1991, 2(1): 47～55

4 Sartori M A, Antsaklis P J. A simple method to drive bounds on the size and to train multiplayer neural networks. IEEE Trans on Neural Networks, 1991, 2(3): 467～471

5 Reed R. Pruning algorithm——a survey. IEEE Trans on Neural Networks, 1993, 4(5): 740～747

6 靳蕃编著. 神经计算智能基础. 成都：西南交通大学出版社，2000. 135～136

7 赵弘，周瑞祥，林廷圻. 基于 Levenberg-Marquardt 算法的神经网络监督控制. 西安交通大学学报，2002，36(5)：523～527

(上接第 51 页)

在的问题：失效节点个数增加将会导致坐标空间内碎片的产生，一些节点掌管了大量的“区”。为了处理这个问题，Can 有一个节点重新定位的算法，用来合并那些可以合并的“区”。

4.3 结构化的文件查询算法优点

结构化的文件查询算法有很多优点：(1)效率高，目标的定位可以在确定的跳数内完成，节点间通信的平均延迟较小；(2)流量小，每次路由转发只选取一个节点，方向性强，不会像“Flooding”算法一样带来大量的网络流量；(3)扩展性好，节点路由表的大小和路由所需的跳数随着系统规模的扩大呈对数级增长，通信性能随着系统规模的扩大优美降级，具有良好的可扩展性；(4)可靠性好，节点之间不同路径数随系统规模的扩大呈对数级增长，如果路由路径中有节点失效，路由算法可以绕过失效节点，选择另一条路径，到达目的地，通信可靠性较高。

可以看出，结构化的 P2P 系统综合考虑了系统的可扩展性、通信的效率和路由可靠性，在三者之间作了有效的折衷，比前面介绍非结构化 P2P 系统更适合于构造大规模分布式存储系统。

总结与展望 上述一些算法是目前常见的 P2P 网络结构中所采用的文件查询算法。它们的优点和缺点反映出了设计者最初针对不同的网络结构特点所做的工作。根据现有文件查询算法的不足，我们提出了如下一些值得改进的方向。

非结构化的 P2P：非结构 P2P 网络的特征就是存放数据的位置和网络的拓扑结构完全无关。我们在这种网络上查询文件时只能是随机地查找，如果不考虑收敛性问题，非结构的 P2P 网络对文件共享和开发应用来讲是一个较好的选择。

因此改善收敛性问题是非结构化 P2P 网络发展的关键和研究方向。

例如，采用“并行随机漫步”[10]来替代“Flooding”洪泛算法，改善收敛性和文件检索的效率。

结构化的 P2P：在结构化的 P2P 网络系统中，当我们为了访问数据而创建标识符时，主要存在如下两个问题：

1. 数据的局域性被破坏。由于文件是根据 fileID 来确定存放位置，这使得文件可能被保存在距离访问请求比较远的地方，数据的局域性(locality)受到一定的影响，这也就意味着本来在一个小范围内经常访问的数据，可能会被随机存放在外部空间，失去了预先定位、高效查询的机会。

2. 对应用层的信息不能合理利用。实际上，大量的网络应用会自然地分层描述文件数据，这说明了文件之间隐式存在着远近关系。但是结构化的 P2P 网络根据文件标识符来分配存储节点的算法，明显无法利用这种现有的关系来合理分配资源。

同时结构化 P2P 网络缺乏一定的灵活性(resilience)，这是由于路由表被分散到了系统中的各个节点上，每个节点都要参与路由，当节点加入或者是离开网络时，有效地维护路由信息是比较困难的。

参 考 文 献

1 The gnutella homepage: http://gnutella.wego.com

2 Iarke I, Sandberg O, Wiley B, Hong T. Freenet: A distributed anonymous information storage and retrieval system. In: Proc. of ICSI Workshop on Design Issues in Anonymity and Unobservability. June 2000

3 Stoica I, Morris R, Karger D, Kaashoek MF, Balakrishnan H. Chord: A scalable peer-to-peer lookup service for internet applications. In: Proc. of SIGCOMM, Aug. 2001

4 Rowstron A, Druschel P. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In: Proc. of the 18th IFIP/ACM Intl. Conf. on Distributed Systems Platforms, Nov. 2001

5 Ratnasamy S, Francis P, Handley M, Karp R. A scalable content-addressable network. In: Proc. of SIGCOMM 2001, Aug. 2001

6 Gummadi K, Gribble S, et al. The Impact of DHT Routing Geometry on Resilience and Proximity. In: Proc. SIGCOMM'03, Aug. 2003

7 Bosselaers A, et al. SHA: a design for parallel architectures. In: Proc. Eurocrypt'97. April 1997

8 Zhao B Y, Kubiatowicz J, Joseph AD. Tapestry: An infrastructure for fault-tolerant wide-area location and routing. Technical Report UCB/CSD-01－1141, 94720, April 2001

9 Saroiu S, Gummadi PK, Gribble SD. Exploring the design space of distributed peer-to-peer systems: Comparing the web, triad and chord/cfs. In: Proc. of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02), March 2002

10 Lv C, Cao P, ECohen, Li K, Shenker S. Search and replication in unstructuredpeer-to-peer networks. Preprint, Optional 2001