

模块化神经网络的 Bayes 子网集结新算法研究

王 攀^{1,2}, 李幼凤³, 冯 珊²

(1. 武汉理工大学 自动化学院, 湖北 武汉 430070; 2. 华中科技大学 控制系, 湖北 武汉 430074; 3. 浙江大学 信息学院, 浙江 杭州 310027)

摘 要:针对模块化神经网络的重要命题——子网动态集成问题,提出一种基于改进的 Bayes 学习的子网集结新方法. 首先从处理复杂问题能力、计算开销、训练误差限等级的合理性、逼近正确率的构造等方面分析了已有方法的不足. 既而提出相应策略,其核心在于采用了简洁、相关性小的子网生成方法;同时以误差作为依据提出新的逼近正确率指标以确定子网的动态集结权值. 仿真实验对两种改进方法的测试误差进行了比较研究,结果表明了改进方法的有效性.

关键词:模块化神经网络; Bayes 学习; 子网集成

中图分类号: TP183 **文献标识码:** A **文章编号:** 1673-4785(2006)02-0079-05

Novel integrated algorithm of modular neural network's sub-nets based on Bayesian learning

WANG Pan¹, LI You-feng³, FENG Shan²

(1. School of Automation, Wuhan University of Technology, Wuhan 430070, China; 2. Department of Control, Huazhong University of Science and Technology, Wuhan 430070, China; 3. School of Information, Zhejiang University, Hangzhou 320027, China)

Abstract: Aiming at the important issue of modular neural network (MNN) —the dynamic integration of the sub-nets, a novel integrated algorithm based on the improved Bayesian learning is presented. Firstly, the drawbacks of the old algorithm are analyzed from four aspects—the processing ability for complex problems, computing cost, the rationality of trained error limit and the approximated accuracy. Then the corresponding strategy is presented whose key points are adopting a concise and lowly correlative sub-nets generating method, and then a new approximated accuracy index based on the error measure is presented to determine sub-nets' dynamic integration weights. The effectiveness of the above algorithms was demonstrated by simulation through the comparative research to two improved algorithms' test accuracy.

Key words: modular neural network; Bayesian learning; sub-nets' integration

学习机的集成是当前机器学习研究的一个主要方向,并且被广泛应用于现实之中^[1]. 而模块化神经网络则是学习机集成的典型范例^[2]. 模块化神经网络研究的一个核心问题是如何有效地集结子网输出,狭义地说,是如何确定各子网的输出投票权重.

到目前为止,提出的方法有很多,如文献[3]中的等权值法(即平均法)、RLS法、线性最优连接法(MSEOLCs)、模糊积分法(FI)、Stacked Generalization法、主元回归法(PCR)、部分最小二乘法(PLS)、并行递推预报误差法(PRPE). 此外,基于 Bayes 学习的子网集成法(这里简称为 Bayes 学习法)也是获取子网的输出投票权重的一种有效方法^[4-5]. 蔡骏博士提出的一种 Bayes 学习思路新

收稿日期:2006-01-04.

基金项目:国家自然科学基金资助项目(60174039).

颖,有重要的应用价值和潜在的理论价值.但是,在该算法的研究过程中也发现一些必须完善和改进之处.本文就此展开研究.

1 改进的子网集结 Bayes 学习法

从广义上讲,模块化神经网络就是由多个相互相对独立、相互联系、相互协同的神经网络组成的系统,它们有明确的任务分工、相对的独立性和各具特色的信息处理模式和风格.该网络系统的基本单元是单个神经网络——模块.据此,目前文献中出现的多种提法如神经网络集成(neural networks ensemble)、多神经网络(multiple neural networks)、混合神经网络(hybrid neural networks)、分布式神经网络(distributed neural networks)以及委员会机(committee machine)等都统一于上述较宽泛的定义.以神经网络集成为例:它是指多个神经网络学习同一任务再行集结的一种结构.它显然属于(广义)模块化神经网络:它的各模块独立处理同一问题,再利用相应的集成机制将各具特色的处理结果合理集成.模块化神经网络的学习机制可称为集成学习(ensemble learning),或协作学习(cooperation learning).

Bayes 学习法是把确定模块化神经网络的子网输出投票权重当作一种统计决策过程,递归运用 Bayes 决策分析,使得对子网的性能评估(即决策)越来越精确.其基本思想是:在各子网对训练样本集进行连接权值学习的同时,通过对测试样本集中信息的学习来得到各子网的输出投票权重.算法基本框架为^[4]:

1) 选定 K 个不同结构的子网,并设定各个子网的先验投票权重为 $w_j^0, j=1, 2, \dots, K$;

2) 由大到小设置 S 个等级的均方误差限 $e_i, i=1, 2, \dots, S$. 可以依次把 100, 10, 5, 作为分级均方误差限. S 的个数一般取 3~5;

3) 对每个级别 $i(i=1, 2, \dots, S)$ 进行如下递推操作:

在训练集上以 e_i 为均方误差限训练各个子网直至收敛;在测试集上观察各子网的水平逼近正确率 $j(\cdot), j=1, 2, \dots, K$; 根据 w_j^{i-1} 和 $j(\cdot)$ 计算各子网的后验投票权重如下:

$$w_j^i = \frac{w_j^{i-1} j(\cdot)}{\sum_{k=1}^K w_k^{i-1} j_k(\cdot)}, j=1, 2, \dots, K. \quad (1)$$

式中: e_i 是各子网的最终目标均方误差限,而水平逼近正确率 $j(\cdot)$ 是按如下方式进行定义的:

设测试集中有 L 个样本,如果一个子网能以小于 ϵ 的平方误差实现 $P(P-L)$ 个测试样本的逼近,则该子网在这一测试集上的水平逼近正确率为

$$j(\cdot) = P/L. \quad (2)$$

显然该算法比等权值法要合理得多,它不是对所有子网都一视同仁,而是充分考虑到了各个子网的不同.尽管各子网都在同一样本集上学习,但因为不同结构、不同初始权值,学习的结果——泛化性能也因“网”而异.这里,泛化性能是由逼近正确率来表现的.由式(1)知,逼近能力越好的子网所得到的投票权重越大.另外,该算法也考虑到各子网的学习有快有慢这一情况,因此它把权值学习过程分解成若干个连续的阶段,在每一阶段中间再嵌入 Bayes 学习.所谓 Bayes 学习,就是通过不断做实验,将信息不充分或不真实的先验知识改变为信息逐步完善准确的后验知识,从而提高人们对问题认识的把握度,进而提高决策精度.对于神经网络的参数学习和网络设计,常规的误差纠正学习方法都可看作是 Bayes 学习方法的一种近似^[5].

然而,该算法还存在较多缺陷,分析说明如下:

1) 处理复杂问题能力有限.因为所有子网都得先独立学习整个问题,如果这时每个子网都学习困难——要么学习时间很长,要么学习能力很差(直接结果是泛化性能差),那么再协作就毫无意义了.如果为使每个子网都能学习得较好而复杂化网络结构,那么模块化的意义也不存在了.

2) 在整个样本集上训练每个子网若干次,从耗费用时间上讲是不合算的.如果样本集很庞大,子网较多,结构较为复杂,那么训练时间将相当可观.另外还可能造成子网相关性的提高而不利于有效的最终决策.

3) 训练误差限等级的设置若为 100, 10, 5, ..., 那么在以 100、10 作为训练误差限时,要求计算测试误差达到 ϵ 的逼近正确率,似乎有些勉为其难.很可能出现逼近正确率全零的情况,这时式(1)就无意义,因为分母为零.

4) 水平逼近正确率显得过于绝对.对于 1.01 和 100 的测试误差,尽管都大于 ϵ ,但在相对意义上讲,1.01 较 100 正确,决不能等量齐观,同时,对于在逼近误差限附近的一些样本,若存在干扰噪声,则很容易使其中某些样本列为“逼近正确”,而其余的样本不在此之列,这有悖常理.

针对这些不足,作者对上述算法进行了如下改进:

1) 可采用与许多研究者类似的方法:将样本集顺序分成 K (子网个数) 份(这样做有一个前提,即样本集是随机产生的),然后在每份子样本集上训练一个子网,最后按投票方式组合各个子网输出. 这样先分后合,一方面可以降低子网的学习难度,缩短必要的训练时间;另一方面子网间协作的意义也社会化了. 如果说算法在未改进前,子网之间与其说是协作,不如说是竞争:类似于人类社会中的招投标(每个候选者都能独立完成该任务,只是候选者的规模、处理能力各不相同):每个子网相当于一个候选者;子网间的连接权值对应于某个评审委员会对各个候选者的打分;子网输出的加权综合其实是对所有独立候选者处理能力的一种兼收并蓄,扬长补短. 在改进之后,各个子网面对的不再是同一任务,而只是任务的某一部分,因此协作成为必要,这时子网的综合相应于任务的还原.

2) 将测试误差限也进行等级划分. 划分方法为:取当前训练误差限的下一级误差限为这一步的测试误差限. 比方说,若训练误差限等级设置为 100, 10, 5, ..., 而当前的训练误差限为 100, 那么在测试时,就以 10 为误差限,并计算这一水平(10) 下的逼近正确率. 依此类推,只是在到达最后一级训练误差限时,测试误差限保持上一级不变,也就是与当前训练误差限一致. 这样做,显然可以减少大误差限等级时式(1) 分母为零的几率,而且这种渐进学习更符合人类的理性化行为.

3) 对于上述的第 4) 点缺陷,引入一种模糊机制,模糊化“水平逼近正确”的概念,即让超出一定范围的平方测试误差都算作“水平逼近正确”. 用数学语言描述,即:设某一个测试样本的平方测试误差为 se , 计算该误差对模糊集合“水平逼近正确”的隶属度,设为 mb , 若 $mb > 0.5$, 则认为这一误差可以算作“水平逼近正确”. 这里存在一个模糊隶属度选择的问题. 因为当 $se < P$, 那么可认为该误差对模糊集合“水平逼近正确”的隶属度为 1; 当 $se > P$ (P 为某一常数), 那么就认为该误差对模糊集合“水平逼近正确”的隶属度为 0; 否则隶属度随 se 增大而减小, 在 $(0, 1)$ 开区间上取值. 因此文中选用 sigmoid 隶属度函数:

$$\text{sig}(x; a, c) = \frac{1}{1 + \exp[-a(x - c)]} \quad (3)$$

式中: a 为控制交叉点 $x = c$ 处的斜度. 这里, 取 c 为 5.5 倍的当前测试误差限.

4) 对“水平逼近正确率”重新定义以避免式(1)

分母为零. 譬如:

$$j(\quad) = \frac{1/\text{sse}_j(\quad)}{\sum_{k=1}^K 1/\text{sse}_k(\quad)} \quad (4)$$

式中: $\text{sse}_j(\quad)$ 表示第 j 个子网在训练集上以 \quad 为误差限的平方误差和. 显然, $\text{sse}_j(\quad)$ 越小 $j(\quad)$ 越大; 反之亦然. 与式(2) 定义比较, 这种定义更为精细一些. 另外, 由于该表达式进行了归一化, 因此是相对逼近正确率.

因此,改进后的 Bayes 模块化学习算法可描述为:

输入参数:训练样本集 TrainSet ,
测试样本集 TestSet ,
子网训练参数,包括学习率、最大迭代次数和训练误差限 .

输出参数:组合模块化神经网络模型,包括所有子网的内部权值向量以及子网的输出投票权重 w_j ($j = 1, 2, \dots, K$).

算法:
PROCEDURE
BEGIN

1) 构造 K 个不同结构的子网,并设定各个子网的先验投票权重为 $w_j^0, j = 1, 2, \dots, K$; 将训练样本集 TrainSet 顺序分成 K 个子样本集,记为 $T_j (j = 1, 2, \dots, K)$.

2) 由大到小设置 S 个等级的训练均方误差限 $e_i, i = 1, 2, \dots, S$ (S 一般取 3 ~ 5); 从而测试误差限可定,设为 $g_i, i = 1, 2, \dots, S$.

3) FOR $i = 1 : S$
在训练集 $T_j (j = 1, 2, \dots, K)$ 上以 e_i 为均方误差限训练第 j 个子网直至收敛; 在测试集上计算各子网的精确(或模糊) g_i 水平逼近正确率 $j(g_i), j = 1, 2, \dots, K$; 根据 w_j^{i-1} 和 $j(g_i)$ 计算各子网的后验投票权重如下:

$$w_j^i = w_j^{i-1} j(g_i), j = 1, 2, \dots, K.$$
$$\text{IF } \sum_{j=1}^K w_j^{i-1} j(g_i) = 0, \text{ THEN } w_j^i = w_j^{i-1}$$
$$\text{ELSE}$$

$$w_j^i = \frac{w_j^{i-1} j(g_i)}{\sum_{k=1}^K w_k^{i-1} k(g_i)}$$

END
FOR $j = 1 : K$
IF $w_j^i < 0.1/K$, THEN $w_j^i = 0$
END

归一化投票权重

END

END

这里作一说明:如果逼近正确率全零,即条件一成立,那么各个子网的投票权重保持上一次值不变;否则,使用 Bayes 定理对其修正.如果某一子网的投票权重过小,也就是条件二成立(具体来讲,就是小于 $0.1/K$),那么截断为零.在所有子网的投票权重重置完毕后,再次对其归一化.

2 仿真研究

仿真实验在 MATLAB6.1 环境下(处理器为 PIV1.4GHz)完成,并且使用其神经网络工具箱.为克服随机取值可能带来的偶然性,研究中对各实例皆作 5 次实验.限于篇幅,这里只报告 1 例.

对于改进算法中的 g_i 水平逼近正确率 g_i ,可分别按式(2)或(4)来计算.这样就得到 2 种算法,为方便起见,分别称作算法一和算法二.

例 一个多输入非线性动态系统的辨识^[6]

问题描述:

$$\begin{cases} x_1(k+1) = u_1(k) \times \sin(u_2(k) + u_3(k)) \times e^{-x_1^2(k)}, \\ x_2(k+1) = \frac{5}{4} \cdot \frac{u_1(k) \times u_2(k) \times e^{-u_3^2(k)}}{1 + x_2^2(l)}, \\ y(k) = x_1(k) + x_2(k). \end{cases} \quad (5)$$

该方程组表示一个 3 输入 1 输出的动态系统,式中: $x_1(k)$, $x_2(k)$ 为系统的状态变量.令 $z = [u_1(k), u_2(k), u_3(k), x_1(k), x_2(k)]$,则系统的辨识转化为 5 输入 1 输出的函数逼近问题: $y = f(z)$.

实验步骤:首先,构造训练样本集和测试样本集.置 x_1, x_2 的初始值为零, u_1, u_2, u_3 在 $[-1, +1]$ 上取均匀分布的随机数 800 个.取前 400 个用以训练,后 400 个用以测试.然后,设计网络模型.这里,选用 5 个前向 BP 网络,每个网络的结构如表 1.其中,每个隐层神经元的激活函数均为双曲正切函数,输出层神经元的激活函数均为线性函数.最后,训练并测试神经网络.训练参数设置:训练误差限为 $1e-8$,训练最大代数 3 000 代,学习率为 0.02.

表 1 子网结构

Table 1 Subnet's architecture

1	2	3	4	5
5-10-1	5-15-1	5-20-1	5-10-5	15-10-10-1

实验结果:对于算法一和算法二,5 次实验平均测试误差分别为 $3.5381e-5$, $1.0749e-6$.下面给出第 5 次实验的数据结果(表 2)和图形结果(图 1,其中浅色曲线对应算法一).

结果分析:由表 2 知,算法二的测试误差较算法一的要小,从图 1 也可很直观的看到.另外,由算法二得到的投票权重很特殊,只有一个网络的值为 1,其余的全为 0,有些“胜者为王”的意味.其他 4 次实验也是这种情况.从平均测试误差看,算法二的误差是算法一的 $1/30$ 还弱.同文献[6]中用小波网络方法得到的均方差结果($7.88e-4$)相比,显然这 2 种算法的结果要好得多.

表 2 第 5 次实验数据结果

Table 2 The data results of the 5-th experiment

子网	逼近正确率		投票权重	
	算法一	算法二	算法一	算法二
1	0.7475	0.3256	0.1798	0
2	0.7825	0.0434	0.3324	0
3	0.7625	0.6303	0.4878	1
4	0.365	0.0005	0	0
5	0.1475	0.0001	0	0
测试误差	算法一	$1.2671e-7$	算法二	$6.7944e-8$

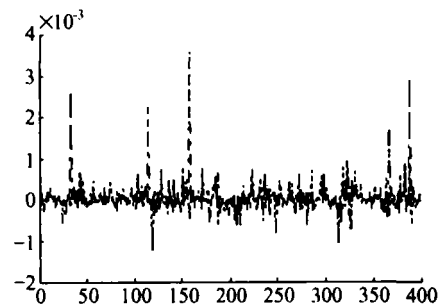


图 1 输出误差曲线(第 5 次实验)

Fig.1 The error curve of the 5-th experiment

3 结 论

本节针对文献[4]中的 Bayes 学习法的若干不足提出几点改进.实验结果表明,不论在速度上还是在精度上都令人满意.对于某些复杂问题,速度优势更是明显.然而,研究过程中也出现了一些需要进一步解决的问题,列举如下:

1) 训练误差限的等级划分是直接从文献[4]中获得的,对其依据尚没有深究.

2) 在子网学习时,测试误差限凭经验设定,在理

论上须作定量分析.

3)对投票权值的递归学习未作理论上的推理论证,其收敛性还有待研究.

同时,虽然从实验结果中反映所提出的算法有较好的鲁棒性,但例子还不太充分,今后尚需通过更广泛的问题加以验证.除此之外,将本算法与其他算法进行融合也颇有研究价值,譬如说,先用这种算法得到各个子网的先验投票权重,而后根据新的输入,对权重进行在线调整.下一步的工作要针对以上这些问题进行深入研究.

参考文献:

- [1] VALENTINI G MASULLI F. Ensembles of learning machines[A]. In: M. Marinaro and R. Tagliaferri (Eds.): WIRN VIETRI 2002[C]. LNCS 2486, 2002, 3 - 20.
- [2]王 攀. 软计算方法在若干决策与控制问题中的应用研究[D]. 武汉: 华中科技大学, 2003.
- WANG Pan. Applicational reesearches with soft computing for some decision and control issues[D]. Wuhan: Huazhong University of Science and Technology, 2003.
- [3]熊智华, 王 雄, 徐用懋. 基于分布式神经网络递推预报误差算法的非线性系统建模[J]. 信息与控制, 2000, 29 (5): 414 - 420.
- XIONG Zhihua, WANG Xiong, XU Yongmao. Distributed neural networks based nonlinear system model using parallel recursive prediction error algorithm[J]. Information and Control, 2000, 29 (5): 414 - 420.
- [4]蔡 骏. 面向决策支持的计算智能技术若干应用研究[D]. 武汉: 华中理工大学, 1996.
- CAI Jun. A study on the computational intelligence applied for decision support [D]. Wuhan: Huazhong University of Science and Technology, 1996.
- [5]AHMAD Z, ZHANG J. Bayesian selective combination

of multiple neural networks for improving long-range predictions in nonlinear process modeling [J]. Neural Comput & Applic, 2005, 14: 78 - 87.

- [6]李向武, 韦 岗. 基于小波网络的动态系统辨识方法及应用[J]. 控制理论与应用, 1998, 15 (4): 494 - 500.

LI Xiangwu, WEI gang. Dynamical system identification and its applications using wavelet network [J]. Chinese Control Theory and Applications, 1998, 15 (4): 494 - 500.

作者简介:



王 攀,男,1971 年生,毕业于华中理工大学,获工学博士学位.现为武汉理工大学自动化学院副教授,控制与决策研究所所长.主要研究方向为智能优化与控制、决策分析,生物医学智能化系统.先后近 10 次获得省部级科技进步、教研成果一、二、三等奖,发表论文 50 余篇.



李幼凤,女,1978 年生,浙江大学信息学院博士研究生,主要研究方向为计算智能、预测控制.



冯 珊,女,1933 年生,教授,博士生导师.清华大学本科,纽约州立大学石溪分校硕士.主要研究方向为智能决策支持系统,复杂系统建模与仿真.先后多次获得国家科技进步三等奖,省部级科技进步一、二等奖,发表论文 100 余篇.