

模块化神经网络子网的动态集成方法学研究

王攀¹, 李幼凤², 范衡^{1,3}, 冯帅¹

(1. 武汉理工大学自动化学院, 湖北武汉 430070; 2. 浙江大学信息学院, 浙江杭州 310027;
3. 丹麦科技大学管理工程系, 2800 Kgs 林拜, 丹麦)

摘要: 提出一个模块化神经网络的广义定义, 它包含了几乎所有多神经网络(系统)。简要分析了模块化神经网络子网集成的相关概念和问题。针对一类模块化神经网络, 提出了 5 种基于“分而治之”原理和自适应组合的新型动态集成方法。它们之间的主要区别在于: 距离测度(绝对距离测度和相对距离测度); 个体数目(有些全部参与集成, 有些则是部分参与); 集成策略和规则(数据驱动和数据/知识驱动)。仿真实验验证了这些方法的有效性。同时, 还提出了一种基于“一专多能”思想的子网训练方法。

关键词: 模块化神经网络; “分而治之”原理; 动态集成; 距离测度; 数据与/或知识驱动策略
中图分类号: N 945.12 **文献标志码:** A

Methodological research for dynamic integration of modular neural network's sub-networks

WANG Pan¹, LI You-feng², FAN Zhun^{1,3}, FENG Shuai¹

(1. School of Automation, Wuhan Univ. of Technology, Wuhan 430070, China;

2. School of Information, Zhejiang Univ., Hangzhou 320027, China;

3. Dept. of Management Engineering, Technical Univ. of Denmark, 2800 Kgs Lyngby, Denmark)

Abstract: A generalized definition of modular neural networks (MNN) that almost include all multi-neural networks (systems) is presented. Some relative concepts and problems of MNN' dynamic integration are briefly analyzed. Five new dynamic integration algorithms are presented for a kind of modular neural networks that are based on the principle of “divide and conquer” and adaptive combination. The main differences among each sub-method are: distance measure (absolute distance measure and relative distance measure) and computing method; selected and integrated number of the candidates (some integrate all networks and the others select only in part); integrated strategy and rule (some are based on data driven or knowledge driven, the others are based on data and knowledge driven). Empirical studies show these algorithms' effectiveness and potential. Meanwhile, a technical approach based on the thought of “experienced-in-one-aspect-and-feasible-in-others” is presented for MNN sub-networks' training.

Keywords: modular neural network; principle of “divide and conquer”; dynamic integration; distance measure; data and/or knowledge driven strategy

0 引言

模块化神经网络是当前软计算集成领域的研究热点^[1-5]。模块化方法(模块化神经网络则是该类方法的范例)的提出既是人类解决面临的日益复杂化的问题之必然要求, 也是人类处理问题时采用“分而治之”、“集思广益”策略的具体体现。当面对大规模复杂问题时, 单一神经网络

往往捉襟见肘, 其固有的缺陷显得尤为突出, 例如, 结构设计困难, 收敛速度慢, 泛化性较差, 难解释性等等。在硬件设计方面, 单一神经网络也可能因结构复杂而较难实现。同时, 社会生活中的经验法则也在向我们表明后面的事实: 群体的分工协作在效率和质量上往往优越于个人的单干独行; 多个具有不同背景、经历、个性和特色的人处理相同问题然后再集思广益地予以综合往往比任何一个个体处理该

收稿日期: 2007-05-17; 修回日期: 2007-10-11。

基金项目: 国家自然科学基金(60174039)资助课题

作者简介: 王攀(1971-)男, 副教授, 博士, 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。E-mail: jfpwang@

tom.com

问题可靠和有效。为此,人们将“模块化(modularity)”的概念引入到神经网络的设计中来,希望通过这种在更高程度对人脑的模拟来提升神经网络的性能。

作者认为: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)。

模块化神经网络方法由于思路自然、方法先进、可操作性强,已在诸多领域获得了广泛应用。如在工程上,模块化神经网络在岩土类型识别、大坝工程结构安全检测、混凝土性能预估等多个方面都大有用武之地;在军事上,有应用模块化神经网络成功获取复杂的非线性导弹制导系统的最优制导律之范例;在商业上, Mitchell 等的研究揭示了模块化神经网络在商业决策领域应用的竞争力^[6];在生命系统信息处理上,该方法也有许多成功应用的范例: Ehrlich 等报道了基于蛋白质序列信息,采用模块化神经网络预测水合作用位的工作^[7]; Pan 和 Sensen 则论证了用混合专家网络集成 3 个预测机 GraiEXP、Genscan、MZEF 可提高基因预测的性能^[8]……

1 子网动态集成的相关概念

模块化神经网络的子网动态集成问题是指对于不同的输入(广义地说是不同的环境和状态),参与集成的网络数量、集成的权重都发生与时俱进的动态调整,从而灵活机动地应对丰富的问题世界。这种模式,因其反映了人的许多启发式处理问题的特征,辅以计算机的强大计算能力,因而极具生命力。目前,已有这方面的应用研究,如文献^[9]。

首先,简要探讨一下频繁需要动态集成子网的原因:对不同问题或同一问题的不同阶段或同一阶段但有不同输入信息或问题完全相同但系统所处的环境不同等等,各子网对于整体(全局)决策的地位和作用都会有所不同,有时甚至相去甚远。这一点,在社会实践、科学研究中累见不鲜。客观地反映这一事实,是有效运用模块化神经网络的一个关键,也是有效避免形而上学方法论的基本要求。

本文重点研究图 1 描述的一类模块化神经网络的子网动态集结问题。其中,任务分解器执行的是输入/输出空间分解,决策机构完成各子网的加权集结。

图 1 反映了很广的一类模块化神经网络:其中, $X \in R^n, Y \in R^m$ 分别为输入和输出, $y_i (i=1, 2, \dots, K)$ 为子网输出;任务分解器的作用是将整体任务分解—如输入或输入/输出空间分解、映射分解、不同类任务的按类分解等。而这

种分解在时序上并非一定要先于各子网络工作,也可边工作边分解;决策机构将根据具体情况对各个子网的输出汇总、分析、处理,如有选择地动态集结各子网输出,在神经网络集成这一特殊的情形下以加权和的形式得到最终结果、判断面向问题的最可靠子网等。对于神经网络集成: $w_i (i=1, 2, \dots, K)$ 是子网决策(或输出)的权重,表征子网 i 在网络综合中的重要性。

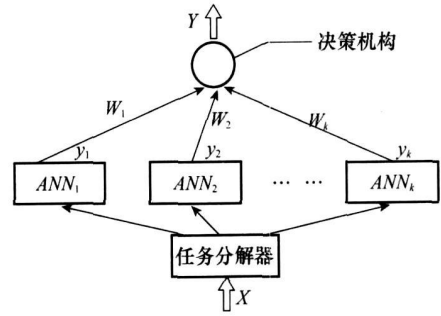


图 1 系统结构图

相应于上述结构,可得到如下形式化描述

$$MANN = \langle ANN_s, TDM, DM, X, Y \rangle \quad (1)$$

式中: $ANN_s = \{ ANN_1, \dots, ANN_k \}$ 为子神经元网络集合; TDM 为任务分解器; DM 为决策机构; X 为系统输入; Y 为系统输出; 特别地,对于神经网络集成的情形: $Y = f(W, Y_{ANN_1}, \dots, Y_{ANN_k})$ 其中: f 为输出函数。

2 一类动态集成算法

本节中模块化神经网络动态集成的基本研究思路是:首先,将样本集按某种聚类算法(如采用模糊 C 均值算法)进行分类;然后针对每类构造并学习一个子网;最后,按选择规则以加权方式将获选子网进行综合,即

$$Y = \sum_{i=1}^K w_i \cdot y_i \quad (2)$$

式中, $y_i, w_i (i=1, 2, \dots, K)$ 的定义如上。

下面给出 5 种选择、集结子网的新动态集成算法(算法 1 主要基于文献^[10]的工作,因此在本文中较多地作为与后面算法(有较大的改进)对比之用)。

2.1 算法 1

先用模糊 C 均值法求出各类的聚类中心 $C_i (i=1, 2, \dots, K)$, 然后求出某个新的输入 X 对各子网的隶属度 η_i , 令 $w_i = \eta_i$, 从而按式(2)得到相应输出。

η 是以这样的方式确定的: $J = \sum_{i=1}^K \eta_i^2 d_i$ 为性能指标, 其中 $\sum_{i=1}^K \eta_i = 1, \eta \in [0, 1]; d_i = \|X - C_i\|^2$, 并利用拉格朗日乘子法解出 η , 即

$$\begin{cases} \eta_i = 1, \eta_{i \neq i} = 0, & \text{当 } d_i = 0 \\ \eta_i = \left(\frac{1}{d_i} \right) / \left(\sum_{j=1}^K \frac{1}{d_j} \right), & \text{其它} \end{cases} \quad (3)$$

该算法是在文献^[10]中的模糊隶属度法基础上稍作改动得来。原算法中

$$d_i = \sum_{j=1}^{N_i} \|X - X_j\|^2 \quad (4)$$

式中, X_j 是第 i 类样本集中的样本, N_i 是第 i 类样本集的样本个数。作者认为: 样本子集中并非所有点对确定新样本点都有同样作用。显然, 远离样本中心的点作用不仅不大, 按文献[10]的方法可能还会起到反作用, 如新输入远离某类中心但该类点集稀疏时; 另一方面, 若样本个数庞大, 那么原有方法计算量将很大。因此取最能体现子类的代表点集(如本文的聚类中心)来代表该类样本集, 将新样本与之距离作为与该类的距离更为合理。

由式(3)不难看出, 本算法中的连接权值实际上是归一化了的输入到各类聚类中心距离平方的倒数。输入到某一类聚类中心的距离越小, 说明此输入对该类的隶属度越大, 从而该类所对应的子网在综合中所占的比重应该越大。反之亦然。该算法流程图如图 2 所示。

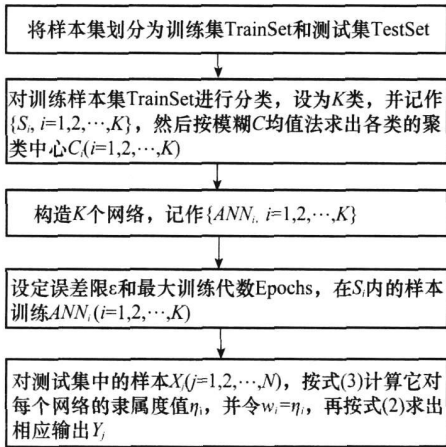


图 2 算法 1 流程框图

许多时候, 人类处理问题时所采用的启发式智能往往能起到一举中的, 切中要害的神来功效。如“抓重点”的办法, 既有深刻的辩证唯物主义的哲学基础, 又有丰富的成功实践运用。因此, 若将人的启发式与计算智能成功结合, 往往会起到几近完美的功效。

2.2 算法 2

本算法较算法 1 的改进在于整个网络的最终输出不是所有子网的加权和, 而是利用所谓“主成分”(这里的主成分指对系统输出贡献率较大的子网)的思想仅选取其中隶属度较大的几个网络来综合, 这显然能合理使用资源并避免在特定问题上性能较差的网络影响效果。

该算法的基本步骤为:

步骤 1 计算 X 对各子网的隶属度 $\eta_i (i=1, 2, \dots, K)$;

步骤 2 对 $\{\eta_i\}$ 由大到小排序, 设排序结果为 $\{\eta'_i\}$, 排序后相应的下标集为 I ;

步骤 3 计算 $w_j = \frac{\eta'_j}{\sum_{i=1}^K \eta'_i} (j=1, 2, \dots, K)$ 。

找出 $k = \max \{k | w_k \geq const, k \in I\}$, 其中 $const$ 为一常数, 在 $[0, 1]$ 间取值。令 $I' = I(1 : k)$ 。

步骤 4 按 $w_i = n'_i / \sum_{j \in I'} n'_j$ 综合 I' 所对应的几个子网, 得到系统输出。

该算法的连接权值仍然是以输入对各类聚类中心距离平方倒数来度量的, 但在综合时只选取了最近邻的几个网络, 这样做是考虑到有时输入样本相对各类的隶属度差异较大, 甚至差别殊异。这时将那些隶属度太小的子网综合进来没有太大意义, 相反, 将其作为“噪声”剔除出去, 能够减少它们对主流网络的干扰, 使隶属度较大的网络能更好的决策, 同时也使计算时间大为缩短, 提高了计算效率。

2.3 算法 3

有时候, 各类别在几何尺寸上差别甚远, 有些类别疆域广阔, 有些类别却是弹丸之地。此时, 若仍然采用绝对距离度量样本属于类别的程度显然会出现误判, 原本属于大尺寸类别的样本会因为它与小尺寸类别的较小距离而更大程度地归于该类, 虽然此时样本与小尺寸类别几无关联。上述现象在问题世界中较普遍。因此, 为应对这一局面, 对算法 2 可在度量距离上作进一步改进: 采用的距离测度 d_i 不再是绝对距离, 而是一种充分考虑类别尺寸差异的相对距离测度。如可定义

$$d_i = \|X - G\| / da_i, i = 1, 2, \dots, K \quad (5)$$

式中, $da_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \|X_j - G\|$, 表示第 i 类的平均距离。

当然, 相对距离测度还有其它的定义方法。限于篇幅, 这里不拟展开。

下面提出进一步充分挖掘人的启发式智能的两种方法。

2.4 算法 4

在算法 2 的基础上加入启发式规则, 使得权值确定更为合理。具体地, 在使用“主成分”时加入如下规则:

规则 1 如果 η'_i 远远大于 $1/K$, 则只选用与之对应的子网络, 连接权值 $w_i = 1 (i=1)$ 。

规则 2 如果 η'_i 接近于 $1/K$, 则选用所有子网络, 连接权值 $w_i = \eta'_i (i=1, 2, \dots, K)$ 。

规则 3 其他情况, 则使用算法二来挑选子网络, 并确定相应的连接权值。

其中: $\eta'_i = \max \{\eta_i | i=1, 2, \dots, K\}$ 。

该算法考虑了最大隶属度与连接权值的均值 $1/K$ 之间的关系, 并充分考虑到两个较极端的情况, 显然从情理上讲更具合理性。本算法的一个关键是确定合适的隶属度函数, 换句话说, 即如何量化模糊语言“远远大于”和“接近于”。

2.5 算法 5

类似于算法 4, 不同的是将算法 3 中的绝对距离测度改为相对距离测度来确定子网间连接权重的。

3 仿真研究

应用本文方法对不同系统作了仿真研究, 效果良好。限于篇幅, 只对 Mackey-Glass 混沌时间序列预测作报告。

仿真实验在 Matlab6.1 环境下(处理器为 PIV 1.4GHz)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, 研究中对实例作了五次实验。

例 Mackey-Glass 混沌时间序列预测^[11]

Mackey-Glass 时间序列是由如下的延迟性微分方程产生的

$$x(t) = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \quad (6)$$

当 $\tau > 17$ 时, 该方程呈混沌特性。 τ 的值越大, 混沌程度越高。 这里, 取 $\tau=30$ 。 学习目的是建立一个一步超前预测器, 即需确定输出层神经元的激活函数均为线性函数。

$$x(t) = f(z(t)) \quad (7)$$

式中, $z(t)=[x(t-1), \dots, x(t-n)]^T$ 为时间序列过去的预测。 这里, 取 $n=10$ 。

离散化(令采样周期 $T_s=1$)后的方程为

$$x(k+1) = \frac{0.2x(k-\tau)}{1+x^{10}(k-\tau)} - 0.9x(k) \quad (8)$$

令 τ 时刻前, x 在 $[-0.2, +0.2]$ 上取均匀分布的随机值。 取该混沌序列的前 2 000 个点作为样本集, 并将前面的 1 000 个点作为训练集, 后 1 000 个点作为测试集。 如此产生五组不同数据(令初始化不同), 以供五次实验用。 实验中选用 10 个前向 BP 网络, 即 $K=10$, 结构如表 1。 表中: 第 1、3 列为网络标号; 第 2、4 为神经网络各层节点个数。 每个隐层神经元的激活函数如下

$$f(x) = \frac{1-e^{-2x}}{1+e^{-2x}} \quad (9)$$

表 1 子网结构

1	10-20-1	6	10-20-18-1
2	10-30-1	7	10-20-20-1
3	10-31-1	8	10-15-20-1
4	10-30-1	9	10-21-22-1
5	10-18-25-1	10	10-20-25-1

表 2 Mackey-Glass 混沌时间序列预测结果(其中时间以秒作为单位)

	实验 1	实验 2	实验 3	实验 4	实验 5
算法 1	5.320 5e-4	1.687 2e-3	2.469 7e-4	4.113 5e-4	9.365 8e-4
	平均测试时间: 109.73		平均均方差: 7.628 3e-4		
算法 2	1.522 6e-6 (0.2)	2.919 0e-4 (0.2)	1.351 8e-6 (0.2)	1.630 8e-6 (0.2)	8.384 8e-7 (0.2)
	平均测试时间: 10.98		平均均方差: 5.944 9e-5		
算法 3	1.947 5e-6 (0.2)	2.840 9e-4 (0.2)	1.104 9e-6 (0.2)	1.978 4e-6 (0.2)	7.545 6e-7 (0.2)
	平均测试时间: 13.92		平均均方差: 5.797 5e-5		
算法 4	1.522 6e-6 (0.4, 2.0, 1.05)	2.919 0e-4 (0.2, 2.5, 1.5)	1.351 7e-6 (0.6, 2.5, 1.5)	1.517 0e-6 (0.6, 2.5, 1.5)	8.374 8e-7 (0.4, 2.0, 1.5)
	平均测试时间: 11.45		平均均方差: 5.942 6e-5		
算法 5	1.947 5e-6 (0.4, 2.0, 1.05)	2.840 5e-4 (0.2, 2.5, 1.5)	1.087 2e-6 (0.4, 1.5, 1.02)	1.978 2e-6 (0.2, 2.5, 1.5)	7.485 5e-7 (0.4, 2.0, 1.5)
	平均测试时间: 11.46		平均均方差: 5.796 2e-5		

4 “一专多能”动态集成方法

本节中, 作者提出一种训练子网时基于“一专多能”思想的动态集成方法。 其特点在于训练子网时, 并非一个训练集对应一个子网的方式, 而是一个训练集(称为主训练集)加一个或几个最临近训练集(称为辅助训练集)对应一

最后, 训练和测试神经网络。 先将每组中训练集样本分成 K 类, 然后让每类对应一个子网, 最后按类训练相应子网。 训练时, 训练函数采用“trainlm”, 即采用 Levenberg-Marquardt 算法来调整网络的权值。 训练完毕后, 将网络权值加以保存, 用以测试。

训练参数设置: 训练误差限为 $1e-8$, 训练最大代数为 2 000 代, 学习率为 0.02。 测试时, 各种参数均凭经验设置。

如表 2 所示, 其中括号中的数据是相应的参数取值。 对应于算法 2 和算法 3, 只有一个参数值, 它表示算法中的选择阈值。 对应于算法 4 和算法 5, 有 3 个参数值, 它们分别表示算法规则 3 中的选择阈值, 规则 1 中的条件阈值和规则 2 中的条件阈值(准确来讲, 是条件阈值的 K 倍)。 从测试时间上看, 算法 1 所花时间(109.73 s)最长, 算法 2 所花时间(10.98 s)最短, 前者几近为后者的 10 倍。 其他 3 种算法所花时间则比算法 2 稍长一些。 究其原因, 若选择的网络少, 那么所花时间也少。 由表 2 知, 算法 2 在 5 组数据中的选择阈值均为 0.2, 该值很容易满足, 几乎对每个样本而言, 只需选用最邻近的 1 个网络即可。 从测试平均误差看, 算法 1(7.628 3e-4)明显较差, 而该算法基本上与文献[10]中描述的算法相似, 是一种静态方法。 后 4 种算法都属于动态集成方法, 实验结果显示: 合理的动态集成具有优势。 同时与文献[8]中的图形结果比较而言, 本文得到的误差曲线显然要好一些。

个子网, 其中以主训练集为核心, 同时兼顾辅助训练集。 表现于指标函数, 即为

$$E = \sum_{x \in S_j} e^2(x) + 1/J \sum_{x \in S_{j,aux}} e^2(x) \quad (10)$$

式中, J 为一根据具体问题适当选取的正整数。 这样处理, 是现实社会中按“一专多能”思想培养人才的模式在模块化子网训练中的生动体现。

