

一种含两层专家网络的委员会机器模型^{*}

金 健¹ 黄国兴² 梁道雷¹ 鲍 钰²

(华东师范大学信息科学技术学院 上海 200062)¹ (华东师范大学软件学院 上海 200062)²

摘 要 针对含单层专家网络的委员会机器在处理复杂问题时拟合能力不足的情况,本文提出了一种基于两层专家网络的委员会机器(CM-2LE)模型,并推导了其中的网络权值学习规则。对人造数据的整体检验和对实际的气象数据的逐次预报检验,通过调节隐含层节点数目,实验误差结果出现了较明显的减小过程,表明通过增加委员会机器中专家网络的层数,可以提高委员会机器的拟合能力。

关键词 委员会机器,模糊 C 均值聚类,模块化神经网络,短期预报

A Committee Machine with Two-layer Expert Nets

JIN Jian¹ HUANG Guo-Xing² LIANG Dao-Lei¹ Bao Yu²

(School of Information Science and Technology, East China Normal University, Shanghai 200062)¹

(Software Engineering Institute, East China Normal University, Shanghai 200062)²

Abstract A CM with two-layer expert net (CM-2LE) is presented to overcome the lack of approximating ability of CM with single layer of expert net for complicated problems. The derived learning rules of the CM-2LE is proposed. Experiments is performed on both synthetic and real-life climatic data. Results show that the increasing number of hidden layer of an expert net in the CM can improve the approximating ability of the CMs.

Keywords Committee machine, Fuzzy c-means, Modular neural network, Short-term prediction

1 引言

神经网络自上世纪中叶被提出以来,经历了曲折的发展历程。如今的电脑 CPU 已经发展成为双内核处理器,甚至是多内核处理器。神经网络也发展出了多网络系统,并且今后很可能成为神经计算领域的一个重要组成部分^[1]。委员会机器(Committee Machine, CM)具有一种并行处理结构,通过组合单个专家模块的输出得到系统的总体输出^[2]。其中,专家模块可以是人工神经网络、经验公式,或者其它算法^[3]。委员会机器大体上可分为两大类,静态结构和动态结构。动态结构中,输入信号被直接用于专家模块的组合策略;而在静态结构中,专家模块输出的组合并不与输入信号直接相关^[4]。文[5]介绍了一种含单层专家网络的动态委员会机器,文中称之为模块化模糊神经网络(Modular Fuzzy Neural Network, MFNN),它被用于建立短期气候预测模型。委员会机器模型中,门网络是用于对专家网络的输出组合的。在 MFNN 中,专家网络采用的是单层线性神经网络,门网络中采用了模糊 C 均值(fuzzy c-means, FCM)聚类。FCM 将输入样本进行粗分类,单层神经网络用于在各自类别中对输入样本进行相对精确拟合。尽管在 MFNN 中各专家网络的组合可以部分抵消因为采用单层线性网络而带来的精度低的问题,但对于像降水量预测这样的复杂问题,该模型在预测能力上还有一定局限。

本文介绍了一种含两层专家网络的委员会机器(Committee Machine with Two-Layer Experts, CM-2LE),尝试通过增加专家层神经网络的层数来提高委员会机器的拟合能

力,并使用一组人造数据和一组实际降水数据进行检验,将 CM-2LE 和 MFNN(即,CM-1LE)进行比较。

2 理论和方法

2.1 委员会机器(CM)

人工神经网络(ANN)由两部分组成,信息处理体和模拟生物神经网络进行信息处理的模式识别工具^[6]。它利用输入-输出模式对,将近似函数映射为对输出值的潜在控制规则,而该输出值是在复杂的实际空间中与输入值相对应的^[7]。

委员会机器,即多个专家模块的集成(专家通常由人工神经网络组成),先要经历一个学习过程,而后委员会可对一个外界新输入,以单个委员预测值的组合的形式给出总体输出。委员会机器可在很多方面得到应用,原因主要有以下几个:首先,委员会可以得到单个委员难以取得的性能。当多个委员的预测值进行组合后,其误差可以得到一定程度的抵消,从而使最终的误差减小。即使取极端的情况,多个委员使用相同的数据集进行训练,各委员的预测值也会不尽相同,有利于将结果组合。其次是委员会机器的模块化。通常情况下,输入到输出的映射由多个预测机组成,可以使单个预测机把更多注意力放在其特定的子空间上。委员会机器的最后预测值常常由各单个委员的预测值经过加权组合而成。在一些实际应用中,单个委员还可以自动将预测任务划分成多个具有实际意义的模块。最后,委员会机器可以降低计算复杂度。先将数据划分成多个子集,或者划分成多个子空间,而后训练与其对应的预测机,这样做在计算上可能会比使用单个预测机,而对所有数据进行训练会有更高的效率^[8]。委员会机器总体结

^{*}上海市科委项目(065115023)和华东师范大学 211 重点(521B0108)资助。金 健 博士研究生,研究方向为知识发现与数据挖掘;黄国兴 博士生导师,研究方向为智能信息系统、知识发现与数据挖掘;梁道雷 博士研究生,研究方向为知识发现与数据挖掘;鲍 钰 博士研究生,研究方向为知识发现与数据挖掘。

构如图 1 所示。

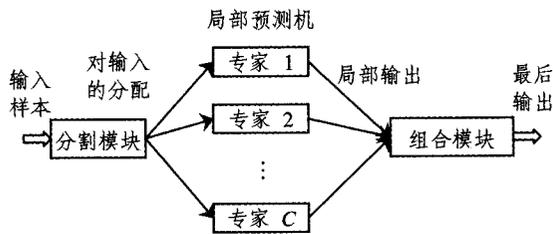


图 1 委员会机器结构

本文介绍的 CM-2LE 结构使用 FCM 聚类算法作为组合模块,经过聚类后每个输入样本不是完全属于某个类或某个专家,而是对每个类都有一个隶属度,这样更符合实际情况。

2.2 模糊 C 均值聚类 (FCM)

模糊 C 均值聚类由 Bezdek^[9]提出,通过使 C 均值目标函数最小化,来找到训练样本集的模糊划分。该目标函数为:

$$\min. J(P;U,V) = \sum_{c=1}^C \sum_{q=1}^Q \mu_{c,q}^m d_{c,q}^2 \quad (1)$$

$$\text{s. t. } \sum_{c=1}^C \mu_{c,q} = 1, \quad (2)$$

其中: $U = [\mu_{c,q}] \in [0, 1]$ —是输入样本集 P 的模糊隶属度矩阵; $V = [V_1, V_1, \dots, V_c]$, $v_c \in \mathcal{R}^R$ 是 c 个聚类中心; $d_{c,q}^2 = \|p_q - v_c\|^2$ 是样本 p_q 与聚类中心 v_c 之间的不相似度,此处用欧氏距离表示; $m \in (1, +\infty)$ 是模糊加权指数,用于确定结果中各类的模糊程度。

对 $J(P;U,V)$ 求最小化的过程可以归结为求以下两个量的迭代过程:

$$v_c = \frac{\sum_{q=1}^Q \mu_{c,q}^m p_q}{\sum_{q=1}^Q \mu_{c,q}^m}, \quad \forall c \quad (3)$$

$$\mu_{c,q} = \left(\sum_{i=1}^C \left(\frac{d_{i,q}}{d_{c,q}} \right)^{\frac{2}{m-1}} \right)^{-1}, \quad \forall c, q. \quad (4)$$

当相邻两次迭代后目标函数值的差别小于预设值 ϵ_U 时,训练结束。

在训练阶段,FCM 聚类将产生各类的聚类中心,我们可以根据这些聚类中心按式(4)求出未知样本的隶属度矩阵。

2.3 含两层专家网络的委员会机器 (CM-2LE)

CM-2LE 中的专家模块是一个两层神经网络,其中,第一层使用 log-sigmoid 传递函数,第二层使用纯线性传递函数(如图 2 所示)。图 2 中, z 是扩展输入向量,它把阈值看成是输入为 1 的特殊的权,因此集成了输入样本和 1。 X 是扩展权值矩阵,它集成了阈值。log-sigmoid 和纯线性传递函数分别定义为:

$$\text{logsig}(n) = (1 + e^{-n})^{-1} \quad (5)$$

$$\text{purelin}(n) = n \quad (6)$$

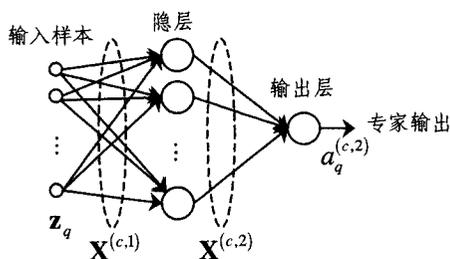


图 2 第 c 类的专家网络结构

整个委员会机器的最后输出可以概括为

$$y_q = \sum_{c=1}^C (\mu_{c,q} \cdot a_q^{(c,2)}) \\ = \sum_{c=1}^C (\mu_{c,q} \cdot X^{(c,2)} \cdot \text{logsig}(X^{(c,1)}) \cdot z_q) \quad (7)$$

CM-2LE 的训练规则采用梯度下降算法,通过求能量函数的最小化来实现。能量函数定义为:

$$E_q = e_q^2 = (y_q - t_q)^2 \quad (8)$$

网络中任一权值的学习规则可写成

$$x(k+1) = x(k) - \alpha g(k) \quad (9)$$

其中, α 是学习率,用于控制学习速度; g 是误差函数 E 对 $x(k)$ 的偏导,即 $\frac{\partial E}{\partial x}$ 。因此,关键问题是要寻找 g 。接下来,我们以第 c 类专家网络中第 l 层中第 $S^{(l)}$ 个神经元与 $(l-1)$ 层中第 $S^{(l-1)}$ 个神经元之间的梯度为例,即 $g_{s^{(c,l)}, s^{(c,l-1)}}^{(c,l)}$, 来作推导(如图 3 所示)。

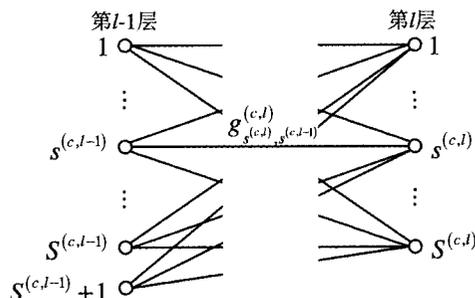


图 3 神经网络两层中神经元的连接关系

应用求偏导中的链规则可得

$$g_{s^{(c,l)}, s^{(c,l-1)}}^{(c,l)} = \frac{\partial E_q}{\partial x_{s^{(c,l)}, s^{(c,l-1)}}^{(c,l)}} = \frac{\partial E_q}{\partial n_{s^{(c,l)}}^{(c,l)}} \cdot \frac{\partial n_{s^{(c,l)}}^{(c,l)}}{\partial x_{s^{(c,l)}, s^{(c,l-1)}}^{(c,l)}} \quad (10)$$

$$\text{其中 } \begin{cases} \forall s^{(c,l-1)} \in \{1, 2, \dots, S^{(l-1)}, S^{(l-1)} + 1\} \\ \forall s^{(c,l)} \in \{1, 2, \dots, S^{(l)}\} \end{cases}$$

式(10)中的第二项可推导为

$$\frac{\partial n_{s^{(c,l)}}^{(c,l)}}{\partial x_{s^{(c,l)}, s^{(c,l-1)}}^{(c,l)}} = \frac{\partial (x_{s^{(c,l)}, s^{(c,l-1)}}^{(c,l)} \cdot z_{s^{(c,l)}}^{(c,l)})}{\partial x_{s^{(c,l)}, s^{(c,l-1)}}^{(c,l)}} \\ = \frac{\sum_{s=1}^{S^{(l-1)}+1} (x_{s^{(c,l)}, s}^{(c,l)} \cdot z_{s^{(c,l)}}^{(c,l)})}{\partial x_{s^{(c,l)}, s^{(c,l-1)}}^{(c,l)}} = z_{s^{(c,l)}}^{(c,l)} \quad (11)$$

其中, $n_{s^{(c,l)}}^{(c,l)}$ 是第 $s^{(c,l)}$ 个神经元的净输出(net output)。

式(10)中的第一项常被称为灵敏度因子。因此可定义

$$\delta_{s^{(c,l)}}^{(c,l)} = \frac{\partial E_q}{\partial n_{s^{(c,l)}}^{(c,l)}}, \quad (12)$$

利用式(11)和(12),式(10)可改写为

$$g_{s^{(c,l)}, s^{(c,l-1)}}^{(c,l)} = \delta_{s^{(c,l)}}^{(c,l)} \cdot z_{s^{(c,l)}}^{(c,l)} \quad (13)$$

对灵敏度因子的求取分两步走。

先考虑网络的输出层。这里我们只考虑 CM-2LE 中的情况,即专家网络为两层,且输出层只有一个神经元。

再次使链规则,有

$$\delta_{s^{(c,2)}}^{(c,2)} = \frac{\partial E_q}{\partial n^{(c,2)}} = \frac{\partial E_q}{\partial a^{(c,2)}} \cdot \frac{\partial a^{(c,2)}}{\partial n^{(c,2)}}$$

其中,等式右边第一项和第二项可分别推导为

$$\frac{\partial E_q}{\partial a^{(c,2)}} = \frac{\partial (y_q - t_q)^2}{\partial a^{(c,2)}} = 2e_q \cdot \frac{\partial \sum_{i=1}^C (\mu_{i,q} \cdot a^{(i,2)})}{\partial a^{(c,2)}} = 2\mu_{c,q} e_q \quad (15)$$

$$\frac{\partial a^{(c,2)}}{\partial n^{(c,2)}} = \frac{\partial \text{purelin}(n^{(c,2)})}{\partial n^{(c,2)}} = 1 \quad (16)$$

因此,式(14)可改写为

$$\delta^{(c,2)} = \frac{\partial E_q}{\partial n^{(c,2)}} = 2\mu_{c,q}e_q \quad (17)$$

对于第一层网络,有

$$\delta_s^{(c,1)} = \frac{\partial E_q}{\partial a_s^{(c,1)}} \cdot \frac{\partial a_s^{(c,1)}}{\partial n_s^{(c,1)}} \quad (18)$$

其中,对于 $\forall S^{(c,2)} \in \{1, 2, \dots, S^{(2)}\}$

$$n_s^{(c,2)} = x_s^{(c,2)} \cdot z^{(c,2)} = \sum_{s^{(c,1)}=1}^{S^{(1)+1}} (x_s^{(c,2)},_{s^{(c,1)}} z_s^{(c,2)}) \quad (19)$$

式(19)中的 $z_s^{(c,2)}$ 可能取 $a_s^{(c,1)}$, 也可能是 1。

则(18)式又写成

$$\begin{aligned} \frac{\partial E_q}{\partial a_s^{(c,1)}} &= \frac{\partial E_q}{\partial n_1^{(c,2)}} \cdot \frac{\partial n_1^{(c,2)}}{\partial a_s^{(c,1)}} + \dots + \frac{\partial E_q}{\partial n_{s^{(c,2)}}^{(c,2)}} \cdot \frac{\partial n_{s^{(c,2)}}^{(c,2)}}{\partial a_s^{(c,1)}} \\ &= \sum_{s^{(c,2)}=1}^{S^{(2)}} \left(\frac{\partial E_q}{\partial n_{s^{(c,2)}}^{(c,2)}} \cdot \frac{\partial n_{s^{(c,2)}}^{(c,2)}}{\partial a_s^{(c,1)}} \right) \\ &= \sum_{s^{(c,2)}=1}^{S^{(2)}} \left(\frac{\partial E_q}{\partial n_{s^{(c,2)}}^{(c,2)}} \cdot x_s^{(c,2)},_{s^{(c,1)}} \right) \end{aligned} \quad (20)$$

将式(14)代入式(20)可以得到

$$\frac{\partial E_q}{\partial a_s^{(c,1)}} = \sum_{s^{(c,2)}=1}^{S^{(2)}} (\delta_s^{(c,2)} \cdot \omega_{s^{(c,2)},s^{(c,1)}}) \quad (21)$$

式(18)等式右边第二项可推导为

$$\frac{\partial a_s^{(c,1)}}{\partial n_s^{(c,1)}} = \frac{\partial \text{logsig}(n_s^{(c,1)})}{\partial n_s^{(c,1)}} = \frac{e^{-n_s^{(c,1)}}}{(1+e^{-n_s^{(c,1)}})^2} \quad (22)$$

结合式(18)、(21)和(22)可得到

$$\delta_s^{(c,1)} = \left(\sum_{s^{(c,2)}=1}^{S^{(2)}} (\delta_s^{(c,2)} \cdot \omega_{s^{(c,2)},s^{(c,1)}}) \right) \cdot \frac{e^{-n_s^{(c,1)}}}{(1+e^{-n_s^{(c,1)}})^2} \quad (23)$$

特别地,在 CM-2LE 中,

$$\delta_s^{(c,l)} = \delta^{(c,2)} \cdot \omega_{1,s^{(c,1)}} \cdot \frac{e^{-n_s^{(c,1)}}}{(1+e^{-n_s^{(c,1)}})^2} \quad (24)$$

最后,将 CM-2LE 的权值学习规则概括为:

$$\omega_{s^{(c,l)},s^{(c,l-1)}}(k+1) = \omega_{s^{(c,l)},s^{(c,l-1)}}(k) - \alpha \delta_s^{(c,l)} z_s^{(c,l-1)} \quad (25)$$

其中

$$\delta_s^{(c,l)} = \begin{cases} 2\mu_{c,q}e_q & l=2 \\ \delta^{(c,2)} \cdot \omega_{1,s^{(c,1)}} \cdot \frac{e^{-n_s^{(c,1)}}}{(1+e^{-n_s^{(c,1)}})^2} & l=1 \end{cases} \quad (26)$$

整个系统的工作过程可以归结为两个阶段:训练阶段和测试阶段。

训练阶段:

1) 设置 FCM 聚类参数: C 指聚类数; m 指模糊加权指数; K_U 指最大迭代次数; ϵ_U 最小中止迭代误差(当相邻两次迭代后的目标函数之差达到此值,则中止迭代)。

在约束(2)下将隶属度矩阵 U 初始化为随机数。

2) 利用前文介绍的 FCM 聚类算法求出各类的聚类中心 V 和训练样本的隶属度 U 。

3) 设置专家网络参数: $S^{(1)}$ 指第 1 层的神经元数目; K_W 指最大训练次数; ϵ_W 指最小中止训练误差(当 Q 个训练样本的均方误差达到此值,则中止训练); α 指学习率,它控制误差收敛速度。

初始化所有专家网络中两层神经网络权值,CM-2LE 中

共有 C 个专家网络。

4) 利用式(25)对所有 Q 个样本都进行一次权值更新,我们称之为一次训练。重复多次训练,直到均方误差达到 ϵ_W 或最大训练次数达到 K_W 。

测试阶段:

1) 利用训练阶段得到的聚类中心和式(4)计算测试样本的隶属度矩阵。

2) 利用第 1 步得到的隶属度矩阵和训练阶段得到的神经网络权值,根据式(7)可求得最终预测值。

3 实验及分析

3.1 数据集

采用一组人造数据和一组连续 47 年的气象数据对 CM-2LE 进行实验检测。

第一组人造数据是根据下式进行采样的,

$$t = \left(1.0 + \sqrt{p_1} + \frac{1}{p_2} + p_3^{-1.5} \right)^2 \quad (27)$$

共采样 40 个样本,部分数据如表 1 所示。其中自变量取 1 到 5 间的整数,该组数据与文[10]中的完全相同。其中测试集含有几个与训练集中相同的样本,不过这不影响我们对 CM-2LE 和之前的 MFNN 的性能比较。

表 1 部分人造数据

No.	Training Set				Testing Set			
	p_1	p_2	p_3	t	p_1	p_2	p_3	t
1	1	3	1	11.110	1	1	5	9.545
2	1	5	2	6.521	1	3	4	6.043
3	1	1	3	10.190	1	5	3	5.724
4	1	3	4	6.043	1	1	2	11.250
...
20	5	5	4	12.680	5	3	2	15.390

气象数据取自广西全省 89 年气象观测站连续 47 年五月份日平均降水量资料,其中还包括 10 个经过挑选的相应气候参数作为预报因子,部分数据如表 2 所示。

表 2 连续 47 年广西全省五月份日平均降水数据

No.	p_1	p_2	p_3	...	p_8	p_9	p_{10}	t
1	176	76	180	...	173	96	15	2174
2	168	92	174	...	177	125	29	1536
3	169	76	176	...	174	114	8	3323
4	172	80	177	...	178	124	105	2619
...
47	175	76	179	...	175	140	27	2392

对两组数据采用不同的测试方式。对于人造数据,我们用 20 个样本作为训练集,另外 20 个作为测试集。而对于降水量数据,我们采用逐步预测方式对其测试。先将第 1~40 年的数据作为训练集,第 41 年的作为测试;接下来将第 1~41 年的数据作为训练集,第 42 年的作为测试;以此类推,一直到以第 47 年的数据作为测试。文[11]中也采用这种测试方式。

3.2 实验结果及分析

由于最佳学习率会随着隐层结点数或随着训练过程的变化而变化,我们在测试过程中通过手动方式人为地调节学习

率,尽量取较优值。

为了考察 CM-2LE 中隐含层在系统中表现出的效果,在与 MFNN 的比较实验时固定如 C , m , ϵ_U , K_U 和 K_W 等与 MFNN 公有参数的情况下,将第一层专家网络的神经元数从 1 开始逐步增加到 30。对人造数据实验产生的两组结果分别如表 3 和表 5 所示。

产生第一组结果所使用的参数设置为: $C = 3$, $m = 2$, $\epsilon_U = 1e-5$, $K_U = 100$, $K_W = 500$ 。经过调整学习率(最终确定为 $3e-2$),得到 CM-1LE 的训练样本的平均相对误差 (MRE_{tm}) 和测试样本平均相对误差 (MRE_{tst}) 分别为 8.15% 和 15.12%,明显不如表 3 所示的 CM-2LE 的结果。

表 3 CM-2LE 对人造数据实验结果 ($C=3$)

α	$S^{(1)}$	MRE_{tm}	MRE_{tst}
5.00E-03	1	14.49%	10.27%
5.00E-03	3	3.27%	5.58%
9.00E-03	5	1.27%	2.83%
9.00E-03	20	1.07%	3.61%
9.00E-03	30	1.12%	3.56%

在第二组结果中,聚类数 C 增加到 5,参数 m , ϵ_U , K_U 不变,其它参数及 CM-1LE 的测试结果如表 4 所示。可以看到,其测试结果也远不如 CM-2LE 的结果。

表 4 不同学习率和训练次数情况下 CM-1LE 的测试结果

α	K_W	MRE_{tm}	MRE_{tst}
7.00E-02	500	4.95%	14.20%
8.00E-02	500	4.55%	13.91%
9.00E-02	500	4.37%	13.98%
9.00E-02	5000	2.94%	17.35%

表 5 CM-2LE 对人造数据实验结果 ($C=5$)

α	$S^{(1)}$	MRE_{tm}	MRE_{tst}
9.00E-03	1	9.10%	8.07%
1.00E-02	3	9.01%	6.66%
1.00E-02	5	6.03%	4.34%
1.00E-02	20	1.22%	4.17%
4.00E-02	30	0.19%	3.48%

从上面几个表格可以看出,CM-2LE 的性能要比 CM-1LE 的好。表 3 和表 5 表明增加隐层神经元数可以显著提高委员会机器的拟合能力。而尽管调节了 CM-1LE 的学习率和训练次数,但其性能还是难以得到明显提高(见表 4)。

在对气象数据进行的实验中我们也手动改变了一些参数,参数改变遵循以下原则:

- 首先,将 m , K_U , ϵ_U , $S^{(1)}$ 和 K_W 分别设定为 2, 100, $1e-5$, 5 和 3000;
- 然后,将聚类数设定为一个较小值,比如 2;设定一个探测性的值 $1e-3$ 作为学习率值;接下来根据权值学习规则对 CM-2LE 进行训练。通过观察训练集的平均误差,将学习率调整为一个相对合适的值。如果误差曲线很陡,则减小学习率;否则增大学习率。如果通过调整得到的结果仍不能令人满意,或感觉到系统未被充分训练,或均方误差曲线不稳定不平滑,则增加聚类数目。

经过多次探测性的实验,我们得到的气象数据数值预报结果,如表 6 所示。其中, MAE_{tm} 和 MAE_{tst} 分别代表训练集和测试集的平均绝对误差。

表 6 气象数据 7 年的逐年预测结果

Year	C	α	MAE_{tm}	MRE_{tm}	MAE_{tst}	MRE_{tst}
41	2	9.00E-5	495.59	27.8%	59.22	2.8%
42	2	1.00E-05	494.18	27.8%	118.03	4.9%
43	3	1.00E-04	385.42	20.9%	22.48	1.2%
44	5	5.00E-04	382.39	21.5%	120.83	5.1%
45	5	7.00E-03	397.82	22.0%	312.01	11.9%
46	3	7.00E-02	442.52	25.8%	135.78	4.9%
47	3	7.00E-03	408.37	22.6%	9.41	0.4%

利用 CM-2LE 对 7 年气象数据的逐步预报性能曲线如图 4 所示。

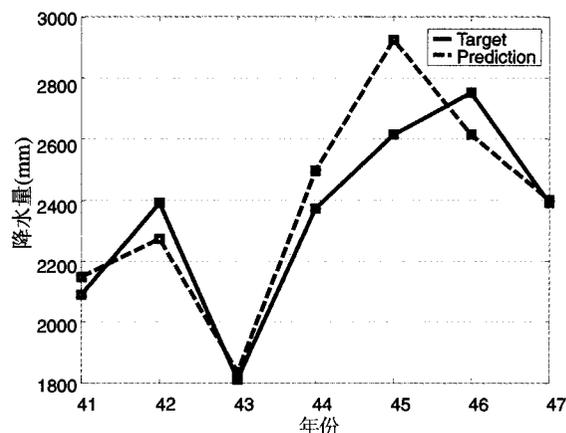


图 4 CM-2LE 对 7 年气象数据预报结果及相应的实况值

表 6 和图 4 表明 CM-2LE 的性能要优于改进的 CM-1LE,即 MFNNef^[11],文[11]中实验数据采用了与本文相同的 47 年气象数据。

毫无疑问,增加委员会机器中的隐层数和隐层神经元数可以提高委员会机器的性能,但这也不是无条件的。增加隐层数和隐层神经元数也可能引起过拟合现象,而且该现象在我们所做的实验中也出现了。我们通过将隐层神经元初始为一个较小值,并将学习率调整到一个满意的水平来克服。在某些情况下,增加聚类数也能从一定程度上提高 CM-2LE 模型的拟合能力。在实验过程中我们还发现,最佳学习率的值和聚类数、隐层神经元数之间存在一定关联,因为对聚类数和隐层神经元数的改变,常引起最佳学习率数量级的改变。在我们所做的实验中还发生另一个现象,即,在对人造数据进行实验时,测试误差比训练误差高;而对气象数据实验时,测试误差低于训练误差。后者也发生在文[5]和文[12]的实验中。

小结 针对含单层网络的委员会机器面对复杂问题而拟合能力不足的情况,我们构造了含两层专家网络的委员会机器,并命名为 CM-2LE。考察该模型的相关特性是非常重要的。为了提高整委员会的拟合能力,它可以为我们是否在确定是否扩大委员会成员的规模,或加强单个专家网络的拟合能力时提供依据。而且本文工作也是更复杂专家网络探索过程的一部分,在今后的专家网络的探索中,本文中采用的神经网络可以被其他更有针对性的算法(如遗传算法、SVM,等)所取代。只要其数据样本符合 $N-1$ 的输入-输出对格式,CM-2LE 模型也能应用于其他领域。如果数据是 $N-N$ 形式,只要相应增加输出层的神经元数即可,对网络的权值更新方法与 CM-2LE 相同。

(下转第 180 页)

同理可证 $\frac{\partial F(a,b)}{\partial b} = \ln(a+b-1) - \ln a > 0$ 。

所以, $F(a,b)$ 关于 a, b 严格递增, 即 $T(x, y, h)$ 是随 m 单调递减的。

由于 $T(x, y, h)$ 关于 m 是递减的, 如果能证明 m 关于 h 总是递减的, 则说明 $T(x, y, h)$ 关于 h 是递增的。

(2) 再证明 m 关于 h 总是递减的

$$\begin{aligned} \frac{\partial m}{\partial h} &= \frac{-4(4h(1-h)) - (3-4h)(4-8h)}{(4h(1-h))^2} \\ &= \frac{-16h+16h^2 - (12-16h-24h+32h^2)}{(4h(1-h))^2} \\ &= \frac{-(16h^2-24h+12)}{(4h(1-h))^2} \end{aligned}$$

要证明 $\frac{\partial m}{\partial h} < 0$, 就是要证明 $16h^2 - 24h + 12 > 0$, 即要证明 $4h^2 - 6h + 3 > 0$

由于判别式 $\Delta = (-6)^2 - 4 \times 4 \times 3 = -12 < 0$, 因此总有 $4h^2 - 6h + 3 > 0$, 即 m 关于 h 是递减的。

综上所述, T 关于 h 是递增的。(证毕)

定理 2 对 $h \in (0.5, 0.75)$, $T(x, y, h)$ 是随 h 单调递增的。

证明: 根据对 m 和 h 的定义, 要证明 $T(x, y, h)$ 关于 h 是递增的, 需要分别证明 $T(x, y, h)$ 关于 m 总是递减的, 并且 m 关于 h 也总是递减的。

(1) 先证明 $T(x, y, h)$ 关于 m 总是递减的。

由于 $T(x, y, h) = (\max(0, x^m + y^m - 1))^{\frac{1}{m}}$,

$$m = \frac{3-4h}{4h(1-h)}$$

考虑 $\frac{\partial T(x, y, h)}{\partial m}$, 若 $\frac{\partial T(x, y, h)}{\partial m} < 0$, 则 $T(x, y, h)$ 是随 m 单调递减的。

∵ 当 $0.5 < h < 0.75$ 时, 有 $m > 0$, $0 < x^m < 1$, $0 < y^m < 1$ 。

当 $x^m + y^m - 1 \leq 0$ 时, 有 $T(x, y, h) = 0$ 。

当 $x^m + y^m - 1 > 0$ 时, 有 $T(x, y, h) = (x^m + y^m - 1)^{\frac{1}{m}}$ 。

下面考虑 $x^m + y^m - 1 > 0$ 的情况。

对 $T(x, y, h) = (x^m + y^m - 1)^{\frac{1}{m}}$ 两边取对数, 有 $\ln T(x, y, h) = \frac{1}{m} \ln(x^m + y^m - 1)$ 。

两边关于 m 求导, 有

$$\begin{aligned} \frac{\partial T(x, y, h)}{\partial m} &= \left(-\frac{1}{m^2}\right) \ln(x^m + y^m - 1) + \\ &\quad \frac{1}{m} \times \frac{x^m \ln x + y^m \ln y}{x^m + y^m - 1} \end{aligned}$$

要证明 $\frac{\partial T(x, y, h)}{\partial m} < 0$, 即等价于证明

$$\frac{1}{m^2} \ln(x^m + y^m - 1) > \frac{1}{m} \times \frac{x^m \ln x + y^m \ln y}{x^m + y^m - 1}$$

其证明过程与定理 1 相似, 故省略。因此 $T(x, y, h)$ 是随 m 单调递减的。

由于 $T(x, y, h)$ 关于 m 是递减的, 如果能证明 m 关于 h 总是递减的, 则说明 $T(x, y, h)$ 关于 h 是递增的。

(2) 再证明 m 关于 h 总是递减的

其证明过程与定理 1 类似, 故略。

综上所述, T 关于 h 是递增的。(证毕)

结论 基于 Schweizer 范数的柔性概率逻辑算子, 可以通过 h 的变化, 使其值在 $[0, 1]$ 区间连续单调变化。实际上, 由定理 1 和定理 2 还可以推出, 当 $h \in (0.75, 1)$ 时有 $P(A|B) > P(A)$, 当 $h \in (0.5, 0.75)$ 时有 $P(A|B) < P(A)$ 。这一研究, 为实现逻辑框架内的概率逻辑不确定推理, 提供了一条有效途径。笔者下一步的研究重点是要建立基于 Frank 算子的柔性概率逻辑算子及其运算模型。

参 考 文 献

- 1 王万森, 何华灿. 基于泛逻辑学的逻辑关系柔性化研究[J]. 软件学报, 2005, 16(5): 754~760
- 2 何华灿, 等. 泛逻辑学原理[M]. 北京: 科学出版社, 2001
- 3 He Huacan, Wang Hua, Liu Yonghuai, et al. Princip of Universal Logics [M]. China: Science Press, 2006
- 4 王万森, 何华灿. 基于泛逻辑学的柔性命题逻辑研究[J]. 小型微型计算机系统, 2004, 25(12): 2116~2119
- 5 Roeper P, Leblanc H. Probability theory and probability logic [M]. university of toronto press incorporated, 1999
- 6 张文修, 梁怡. 不确定性推理原理[M]. 西安: 西安交通大学出版社, 1996
- 7 Klement E P, Mesiar R, Pap E. Triangular Norms. Dordrecht: Kluwer Academic Publishers, 2000
- 8 王万森, 何华灿. 基于泛逻辑学的概率命题逻辑研究与分析[J]. 计算机研究与发展, 2005, 42(7): 1204~1209
- 9 McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biology, 1943, 5: 115~133
- 10 Morshed J, Kaluarachchi J J. Application of artificial neural network and genetic algorithm in flow and transport simulations. Advances in Water Resources, 1998, 22: 145~158
- 11 Tresp V. Committee Machines. In: Handbook for Neural Network Signal Processing, Yu Hen Hu and J.-N. Hwang, eds. CRC Press, 2001
- 12 Bezdek J C. Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers Norwell, MA, USA, 1981
- 13 Shitong W. Fuzzy System, Fuzzy Neural Networks and the Design of Application Program Shanghai. Shanghai Scientific & Technical Press, 1998
- 14 Jin J, Wang J, Huang G, et al. Study on Influence of Effectiveness Factors on Modular Fuzzy Neural Network Based Prediction Model. In: Sixth World Congress on Intelligent Control and Automation, Conference Proceedings (WCICA 2006). Dalian, China, 2006
- 15 Dongjun Y, Shitong W. A New Fuzzy Neural Network Model and Its Modelling Application Research. Journal of East China Shipbuilding Institute, 1998, 12
- 16 Sharkey A J C. Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 1999
- 17 Haykin S. Neural Networks: A Comprehensive Foundation. Prentice Hall PTR Upper Saddle River, NJ, USA, 1998
- 18 Chen C H, Lin Z S. A committee machine with empirical formulas for permeability prediction. Computers & Geosciences, 2006, 32: 485~496
- 19 Bullinaria J A. Lecture 15: Committee Machines. 2004
- 20 Long J, Jian J, Cai Y. A Short-Term Climate Prediction Model Based on a Modular Fuzzy Neural Network. Advances in Atmospheric Sciences, 2005, 22: 428~435

参 考 文 献

(上接第 163 页)

另外, 本文实验过程中手工确定最佳学习率、聚类数、隐层神经元数和最大训练次数的工作非常烦琐, 且增加不确定性。因此, 确定一种自动确定上述参数的机制显得非常迫切。也有一些参数的确定已有现成方法, 在今后的工作中, 我们将会在模型中集成这些算法, 并作进一步探索。