

人工神经网络在天气预报中的应用研究

张承福

(北京大学物理系, 100871)

提 要

介绍了将人工神经网络用于灾害性天气(暴雨)预报研究的情况。分析了天气预报问题的数学提法及困难所在,神经网络用于天气预报的原理,暴雨预报的特点及我们对网络模型的改进。结果表明,神经网络确可通过学习从原始数据中提取足够的分类信息,达到较好的预报准确率,值得进一步研究。

关键词: 神经网络 灾害性天气预报 改进

1 问题的提出

自然界事物的变化与发展都有其内在规律与因果关系,这是可预报的基础,完全随机发生的事件是不可能预报的。预报问题分为两大类,一类是所谓正问题:已知其规律及当前状态,要求其演化与发展。对于正问题,已有系统的、较成熟的数学、物理及计算方法,是相对较容易的任务。但是实际工作中更多遇到的是另一类问题,即反问题:已知事物的行为、表现或实验数据,要求其内在规律,以便作预报与推断。反问题求解一般比正问题困难得多,至今并无系统的、理想的方法。天气预报中很多问题实际上是困难的反问题。

求解反问题的一种常用的方法,是建立模型(动力学建模),假设其规律具有某种形式,式中包含若干待定参数,由拟合实验数据定出。对于本质上是线性的问题,上述方法是很有用的。但是大多数反问题是非线性的,非线性形态变化多端,若无足够的信息与知识,建立模型是极困难的或实际上不可能的。

求解反问题的另一种方法,是用人工智能技术建立专家系统等智能系统。与建模方式不同,它不追求建立全局的、定量的数学模型,而是由专家的经验、知识总结出一套知识库和规则库。每一条规则都是局域的,可以是

定量的、定性的或半定量的。好的专家系统很有实用价值,其性能好坏完全取决于规则库的质量,即取决于专家知识的深浅及总结的优劣。困难在于,高水平的专家并不多;而且,经验知识往往“能意会、难言传”,要总结成计算机能识别的规则,实非易事。人较善于总结低阶规律(因子与推论具有低阶相关性),例如发烧可能有炎症,闷热可能要下雨等等。民间气象谚语中总结的多半是这一类信息。但是复杂事物往往是高阶的,即其规则隐含在多个因子的某种组合中,而可能的组合数是随因子数增加指数地增加的,要从此庞大的高维空间总结出规则来,不是常人容易做到的。

能否直接从原始数据中总结规律呢?实际上,对于低维问题常用的数据拟合(曲线拟合、曲面拟合)及内插、外推技术,就属于此类。再复杂的问题,只要能在二维或三维空间中直观地显示其行为,人就可能总结出其近似规律来。但是,对于高维问题,缺乏直观显示的方法,人也不善于建立整体的高维形象,上述方法难以推广。

总之,求解反问题的困难主要源于非线性、高维与高阶性。此外,对天气预报等大量实际问题,还有数据的不精确性(存在误差、

* 本工作受非线性研究攀登项目的资助。参加本工作的还有陈国新、那基坤、王耀生等;汪之义提供了部分资料。

涨落,其它偶然因素等)、不完备性(因子不完备,采样数据不充分等)等因素,致使以完备性、存在唯一性、精确性等为基础的传统数学方法难以完全适用。而人工神经网络的出现,正是在一定程度上弥补了上述种种方法的不足,提供了求解反问题的一种有力工具。

人工神经网络研究已有 30 余年历史,近 10 年来形成第二次研究热潮,已引起各行各业的广泛关注。其根本原因,即发展的原动力,就在于它开拓了新的思路,为求解某些困难问题提供了新的手段和方法。人工神经网络是借鉴大脑和神经系统贮存、处理信息的某些特性而抽象出来的一种数学模型。其基本特征,也即它与传统方法的重要差别在于:

- (1)大规模并行处理。若硬件实施,速度极快;且便于信息的综合。
- (2)良好的容错性与鲁棒性。这也正是传统方法中的“瓶颈”问题。
- (3)有学习功能,网络的大量参数均由学习(即训练)获得,而不是由人设定。这一点十分重要,这是从原始数据中“提取”信息,逼近规律,而不是由人赋予规律。正如杂技演员的高难动作难以用“编程”控制,但可以通过训练而获得。应该说,没有学习功能的系统很难说是“智能”的,它只是人的智能的快速体现而已。若系统能从学习中“获得”某种技能,且有“举一反三”的能力(即具有推广性),则可认为它具有某种程度的智能。限于篇幅,关于神经网络的一般介绍请参阅有关文献,例如文献[1-5]。

近几年来,我们与北京气象学院、广西气象台等合作,开展了人工神经网络在灾害性天气预报(暴雨预报)中的应用研究。改进了网格模型和算法,摸索出了一套有效的训练方案,取得了较好的结果,表明神经网络方法可以在天气预报中发挥积极作用。

2 基本原理

模式识别、分类、预报、复杂系统控制等等实际问题,在数学上可归结为映射关系 $f: \vec{x} \rightarrow \vec{y}$, 其中, $\vec{x}$ 是待识别、待分类的模式,预报因子,系统的状态等; $\vec{y}$ 则是识别、分类、预报的结果,或应加的控制量等等。一般 f 是非线性的、未知的,只知道若干代表点之间的关

系,即给定样本集 $\vec{x}^\alpha, \vec{y}^\alpha, \alpha=1, 2, \dots, M$, 要求出近似的映射关系来。这是一个典型的反问题。若输出是一维的,且是二值的,则成为二分类问题。感知器(Perceptron)^[6]是可完成二分类问题的最简单的神经网络(图 1)。

$$y = \sigma(\sum_i w_i x_i - \theta)$$

$$\sigma(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (1)$$

其性能由联接权值 $\vec{w}$ 及阈值 θ 决定,它们并非人为设定,而是由学习得到。学习方法不唯一,原则是调整 $\vec{w}$ 与 θ , 使网络输出与目标的差别不断减少。感知器的分类决策类似于委员会的加权投票决策。每个输入神经元代表一个委员,按其“表现”(即该分量的一阶信息量)调整权重。学习过程是达到“集思广义”的优化过程。此模型图象清楚,学习简便,但功能亦有限,只能完成一阶分类问题(线性可分问题),故不宜一般使用。

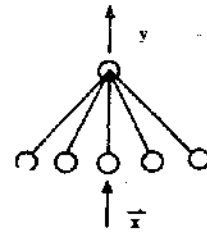


图1 简单二分类网络示意图

若在输入层与输出层之间加一层或多层神经元(称隐层,隐单元),理论上可证明,此种多层前传网络原则上可实现任何非线性映射。三层前传网络结构如图 2 所示,输入记为 $\vec{x}$, 隐层输出记为 $\vec{\xi}$, 输出为 $\vec{o}$, 层间的联接权及阈值分别为 $\vec{w}, \vec{\theta}$ 及 $\vec{w}^{(1)}, \theta^{(1)}$ 。网络传输关

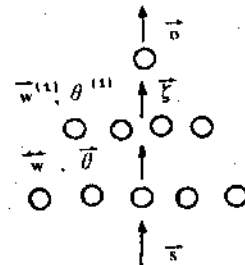


图2 多层前传网络示意图

系的数学表达式为

$$\begin{aligned} o &= \sigma(\vec{w}^{(1)} \cdot \vec{z} - \theta^{(1)}) \\ \vec{z} &= \sigma(\vec{w} \cdot \vec{s} - \theta) \end{aligned} \quad (2)$$

其中 $\sigma(x)$ 为神经元激活函数, 可取多种形式, 通常取为 s 型函数:

$$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \begin{cases} 1 & x \gg 1 \\ 0 & x = 0 \\ -1 & x \ll -1 \end{cases} \quad (3)$$

多层网络功能是强的, 关键是学习, 即如何调整联接权和阈值。通常采用误差反传播 (BP, 即 Back Propagation) 算法^[3], 它本质上就是一种梯度下降优化算法。其原理是, 首先引入误差函数, 亦称能量函数 (这是一种物理类比), 可有多种形式, 最常用的形式是

$$E = \frac{1}{2} \sum_{\alpha=1}^M (\alpha^{\alpha} - o^{\alpha})^2 = E(\vec{w}, \vec{\theta}, \vec{w}^{(1)}, \vec{\theta}^{(1)}) \quad (4)$$

其中, α^{α} 是 α 样本的目标输出, o^{α} 是输入为 s^{α} , 按式 (2) 得到的输出, 它是联接权及阈值的函数, 因而 E 也是。初始时权值任取, 按梯度下降法原则调整之, 以使 E 单调下降, 若 E 小于某一给定范围, 则认为学习完成。若以 w_x 代表任一权值或阈值分量, 则其修改公式为

$$\Delta w_x = -\eta \frac{\partial E}{\partial w_x} \quad (5)$$

其中 η 为小正量, 控制学习速率。 $\frac{\partial E}{\partial w_x}$ 的计算及许多细节、技巧在此不能一一细说, 仅指出一点, 所谓“误差反传播”, 从数学上看, 就是计算复合函数微商的逐级内传过程。

大量计算表明, 用 BP 算法确实可以逼近许多复杂的非线性映射问题, 为原来感到十分棘手的反问题求解提供了一种有力的工具, 在识别、预报、控制等许多领域已有应用。但在实际工作中也发现不少问题, 主要有: 学习似乎很好, 但推广性 (即对未训练的检测集的检验正确率) 差; 长时间逗留于局域极小, 学习难以进展, 学习效率低, 稍大的实际问题往往要很多小时甚至几天的计算, 等等。这些

问题不解决, 研究工作难以顺利开展。神经网络提供了一种方法与工具, 但不是万能的。针对不同的具体问题, 必需对工具作相应的调整、改进, 才能有好的效果。因此, 我们强调, 应用本身就是一项研究, 而不是简单的套用。

3 灾害性天气预报的特点及我们的改进考虑

我们在多年专家系统研究的基础上, 用 10 年的实际数据, 对四川、广西的暴雨这一灾害性天气预报进行研究。四川省暴雨集中在每年 5 月中旬至 9 月中旬, 预报未来 12—48 小时内有无日降水量大于 50mm, 5 站以上成片暴雨。广西省暴雨集中在每年 5—6 月, 预报未来 24 小时有无日降水量 ≥ 100 mm, 10 站以上成片特大暴雨。

3.1 暴雨预报问题的特点

此问题具有如下特点:

(1) 与暴雨有关的因子很多, 按专家系统总结出的因子, 四川有 29 个, 广西有 39 个, 即这是一个高维问题。而且, 统计分析表明, 因子的低阶信息量极弱, 即属于高阶问题。因此, 必须用多层网络, 我们一般用三层网络。

(2) 分析表明, 两类样本 (有无大暴雨) 的因子在高维空间中聚团性差, 错综交缠, 结构复杂。加上气象变化多端, 存在偶然因素、涨落、不确定性, 此问题远比单纯的数学模型复杂。

(3) 数据量大, 四川有二千多组数据, 广西有近千组数据。由于因子空间维数高, 训练样本集必需足够大。因此, 必需对 BP 算法作大的改进, 否则在微机上进行研究是很困难的。

(4) 大暴雨是小概率事件, 四川省的气候概率约为 20%—25%, 广西省报的是特大暴雨, 概率更小, 约为 10%—13%。预报的成功界限指数定义为:

$$CSI = \frac{c_f}{c_f + w_f} \times 100\% \quad (6)$$

其中, c_f 为正确报出的暴雨数, w_f 是错报与

漏报数之和。对于小概率事件,提高CSI是很严的要求。若以 R_r 代表预报暴雨的正确率, N_r 代表预报无暴雨的正确率, A 代表暴雨概率,则式(6)可表示为

$$CSI = \frac{R_r}{1 + (1/A - 1)N_r} \times 100\% \quad (7)$$

由附表可见,要提高CSI, R_r 和 N_r 必需有很高的正确率才行; A 越小,CSI 越低,任务越困难。而且,概率越小,涨落的影响也越大。

附表

CSI	$R_r = N_r = 0.8$		$R_r = N_r = 0.9$	
	$A = 0.1$	28.6%	47%	
	$A = 0.2$	44.4%	64%	

3.2 对网络结构及算法的改进

针对以上特点,我们对网络结构,训练算法等进行多方改进,使研究工作能在微机上顺利开展。主要有:

(1)学习的根本目的是逼近规律,有推广性,而不是单纯追求学习正确率。这方面也曾走过弯路:训练样本分类很好,但检测效果反而很差。为此,我们采取两项措施:a. 在训练时,随时观察检测效果。训练集的 CSI_{train} 与检测集的 CSI_{test} 随训练的变化如图3所示,结果发现训练前期两者同步上升,表明确实在

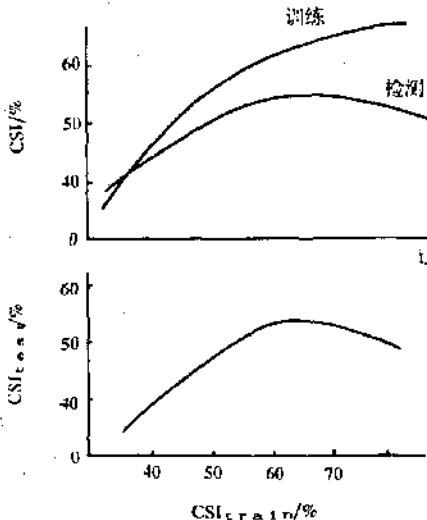


图3 训练集与检测集成功界限指数 CSI_{train} 与 CSI_{test} 间的关系

逐步逼近规律;但到后期, CSI_{train} 的上升反而使 CSI_{test} 下降。究其原因,可能是过于迁就某些不典型、不精确的样本,反会偏离总体规律。因此,训练应适时而止(而不是“误差越小越好”),能达到的 CSI_{test} 也是有一定上限的。

b. 隐层神经元数目不宜过多,过多的可调参数虽然有利于样本集的逼近,但也增加了任意性,往往离规律更远。我们一般取6—10个隐单元。其实,上述两点在简单的曲线拟合中就有表现:用低阶多项式虽不能很好逼近每一样本点,但能大致逼近规律;用高阶多项式虽能精确通过每一样本点,但振荡剧烈,往往与规律相差甚远,对于不精确的数据更明显。两者道理是相通的。

(2)不论取何种误差函数形式,梯度下降算法都是用一个指标来指导训练的。对于单纯的数学问题这也许很好;但对于复杂的实际问题,效率不会高,甚至会行不通。比如,球类教练员靠一个比分指标去指导比赛,军事指挥员靠一个指标去指挥战役,管理部门靠一个指标指导发展等,不论这个指标如何合理,会有好结果吗?神经网络的训练方法提供了一种工具,但如何用好这工具,根据不同的实际问题,不同的训练阶段,不同的具体表现,调整和改变方法,是必不可少的。即目前的网络只是一种工具,并不能代替人的指挥作用。为做到这一点,我们采取两项措施:a. 掌握全面信息。在训练中对每个样本(包括检测集)的输出状态用图形直观地、动态地显示出来,以便发现问题,决定调整策略,观察调整效果。b. 灵活的调整手段。学习效率,权重分配曲线(我们用灵活的权重分配曲线代替通常的误差函数)都可及时的方便的键盘调整。进一步,隐层神经元的函数形式,数目也可及时调整。总之,充分的信息与灵活的手段,使得学习效率有成量级的提高,使研究能在微机上顺利开展。

(3)多网络投票决策:如附表所示,即使 N_r 与 R_r 达到80%—90%,CSI仍不太高。图

3 也表明,一个网络的 CSI_{test} 有一定的上限,难以突破。进一步提高 CSI 的一个方案是:独立训练一批网络,从中挑选一批较好的,再进行投票决策(类似由一批专家组成委员会,投票决策)。这又是一个“集思广议”的过程。

其它一些考虑这里不宜一一细述。

4 结果与讨论

对四川、广西暴雨预报的研究中,我们用 6 年数据进行训练,4 年数据用作检测。 CSI_{train} 与 CSI_{test} 的同步上升表明,网络训练确在逼近规律,可从原始数据中提取分类信息。每次训练网络的结果稍有涨落,一般可达:

对四川数据 CSI_{test} : 50%—55%

对广西数据 CSI_{test} : 35%—42%

四川的 CSI 明显高于广西,主要原因是四川报的是 3 站以上暴雨,概率较高;广西是 10 站以上特大暴雨,概率低得多。

用多网络投票方式,确实可使 CSI_{test} 有进一步提高,四川可达 ~60%,广西可达 45%—50%。但与理论分析相比,提高幅度不够大,原因何在呢?

数据分析表明,数据集中存在一类异常类,其主要表现为:a. 虽然每个网络是从随机初条件独立训练的,但训练集及检测集中报错的样本有很大几率集中在一小群样本中。b. 若将这些样本除外,则训练要容易得

多,检测效果也有大幅提高。即这一小群“异常”的样本的行为是与大多数正常样本“顶牛”的。形成异常类的原因何在?是偶然因素?气象条件的突变性?因子不够可靠?还是另有需要因子没有包括进来?有待进一步分析,特别需要有经验的气象专家的帮助。如何区分异常类与正常类,这是提高预报正确率的重要环节。

目前的工作还是初步的,还有大量问题有待研究。通过前一阶段工作我们体会到,灾害性气象预报研究是十分重要的,但也是十分复杂困难的。神经网络提供了一种有力的工具,但不能单打一,必需综合各种有效手段,吸取专家的有益的经验,才能取得好的结果。

参考文献

- 1 Harold Szu(新华龄). 电脑人脑化:神经网络——第六代计算机. 北京大学出版社,1993.
- 2 张承福. 神经网络系统. 力学进展,1988,18(2):145.
- 3 Kohonen T., Self-Organization and Associative Memory. Springer-Verlag, Berlin, 1984.
- 4 Hopfield J. J., Proc. Natl. Acad. Sci. U. S. A. 81(1984), 2554—2558; ibid. 81(1984), 3088—3092; Biol. Cybern. 52(1985), 141—152.
- 5 Rumelhart D. E. et al. Parallel distributed processing: Explorations in the microstructure of cognition. MIT press, Cambridge MA (1986).
- 6 Rosenblatt F. Principles of neurodynamics. Spartan. New York (1962).

On the Application of Artificial Neural Network for Weather Forecast

Zhang Chengfu

(Department of Physics, Peking University 100871)

Abstract

The research on the application of artificial neural networks for calamitous (rainstorm) weather forecast is reported. The mathematical description of weather forecast, the difficult points of the problem, the principle of neural network for forecast, the characteristics of rainstorm forecast and our improvements of neural network models are described and analyzed. The results show that by learning from samples, the networks can extract enough classified informations from original data and reach good forecast index. It means further works in this direction are worth.

Key Words: neural network calamitous weather forecast improvement