

一个自生成的神经树网络模式分类器*

李吉桂 林国扶 李冠英

(华南师范大学计算机科学系, 广州 510631)

摘要 本文研究一种神经树网络(NTN, Neural Tree Network)模型, 提出一种能自动生成其体系结构的有监督的竞争学习算法, 指出该模型能解决复杂的多类模式分类问题. 模拟数字例子——手写体数码识别的结果也表明该分类器是有效的.

关键词 神经网络, 模式分类, 决策树, 神经树网, 竞争学习, 聚类.

人工神经网络是一种以大规模并行处理和大量连接为基础的计算模型. 它在手写体识别、语音识别和图象识别方面取得了令人鼓舞的结果. 然而目前存在的一些问题却妨碍它进一步推广应用. 其中之一是如何将问题分解成适当的子系统并进行求解. 易知决策树^[1]是一种解决复杂模式分类的方法, 它能将复杂的分类问题转化为一系列简单的分类问题. 因而不少人采用一种类树状结构来组织神经网络, 以解决复杂的多类模式分类问题. 如岩田等人提出的大规模神经网络模型“Comb NET”^[2], 以及文献[3]提出的用于识别大量印刷体朝文的两层树状神经网络复合模型等. 不过, 这些系统只是按照树结构来组织神经网络, 而且其结构也是事先决定的, 缺乏灵活性.

本文研究一种在学习中能自动生成其结构的树状神经网络(称为神经树网络, 或简称为神经树网, NTN). 树的根节点、内部节点由一简单神经网络组成(本文提出的网络特别简单, 见图1). 我们提出了一种有监督的竞争学习算法. 模拟数据的结果说明它是有效的.

1 神经树网的结构

神经树网的结构(图1)由学习时生成的树结构组成, 其根节点、内部节点由简单的双层神经网络组成. 它只有一个输入层 L_A 和一个输出层 L_B , 没有隐含层. L_A 层有 n 个单元, L_B 层有 P 个单元; L_A 到 L_B 向前全连接, 设 W 为它们的连接矩阵. $A_k = (a_1^k, a_2^k, \dots, a_n^k)$ 为网络的第 k 个输入向量(样本模式), 对应地 L_A 层有 n 个单元; $B_k = (b_1^k, b_2^k, \dots, b_P^k)$ 为网络输入 A_k 时的输出(向量), 对应地 L_B 层有 P 个单元. n 是输入向量的维数, 由特征抽取确定, 在分类过程中不变. P 为 L_B 层的单元数. 对应于根节点的网络, P 表示希望将被识别模式分为 P 类(例如数字字符有 10 类), 对于其它内部节点, P 表示归并到该单元的类别数 P_i ($P_i \geq 2, i$ 为

* 本文 1994-05-27 收到, 1994-12-29 定稿

作者李吉桂, 1942年生, 教授, 主要研究领域为知识工程, 模式识别等. 林国扶, 1969年生, 技术员, 主要研究领域为知识工程. 李冠英, 1934年生, 教授, 主要研究领域为专家系统.

本文通讯联系人: 李吉桂, 广州 510631, 华南师范大学计算机科学系

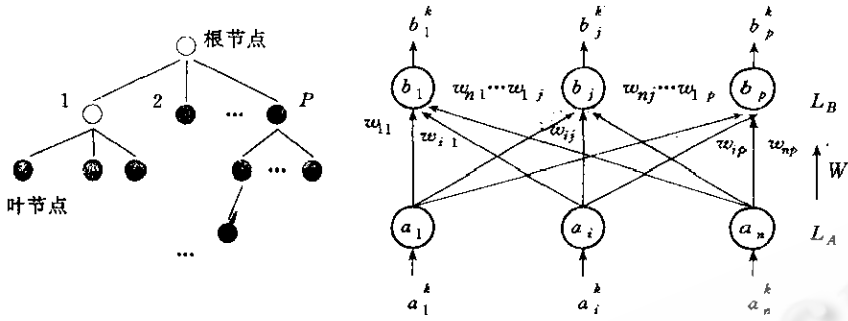


图1 神经树网的结构

内部节点标识,对于训练例, P_i 为已知).

神经树网开始学习时,只有一个根节点 T ,学习时,根节点的网络一般将样本空间分为 P 个子空间(区域),每个子空间赋予一个子节点.对于每个子节点例如 T_i ,计算其子空间所含样本的类别数 P_i ,于是对应于节点 T_i 的网络,其 L_A 有 n 个单元, L_B 有 P_i 个单元.它又将 T_i 对应的子空间进行划分,如此继续,直到每个节点仅包含一个类别的样本为止.仅含一个类别的节点称为叶节点,对应于一个类,常用该类名称进行标识.

训练完毕,其树结构、各节点对应的网络的权重和叶节点对应的标识暂时确定.若将其用于识别,当被识别样本落入对应的叶节点时,该叶节点的标识即为被识别样本的类别标识.

2 神经树网的学习算法

如上所述,神经树网从根节点开始,将样本空间逐层逐层地划分.每次划分都依赖于所在节点的神经网络.因此必须使这些网络有良好的分类性能.对于图1,常用的算法有感知机算法^[4]和LMS学习算法^[5].易知这些算法不是线性可分的.即使采用非线性单元的感知机算法,也只能部分解决线性不可分问题,且收敛速度也较慢^[6].另一种较流行的学习算法是采用自组织聚类技术的竞争学习算法.该算法对每一个输入向量,不断修改其聚类中心,每个聚类中心由一个神经元的连接权向量表示,这些权向量一开始赋予随机值.学习时采用一种“胜者为王”的竞争学习规则,当输入一个向量时,为了找到最接近该向量的聚类中心,各个神经元之间就进行竞争.被选中的神经元对应的聚类中心(该神经元的连接权向量)将被修改,使之更接近于输入向量.

设 W_i 为输出单元 i 和输入单元之间的连接权向量, $D(X, W_i)$ 为输入向量 X 和 W_i 的距离, Y_i 为单元 i 的输出, P 为输出层的单元数, t 为迭代次数, m 为输入模式向量的序号,则最简单的竞争学习规则 SCL^[7]如下所述:

$$W_i(m+1) = W_i(m) + a(t)(X - W_i(m)) \cdot Y_i \tag{1}$$

其中
$$Y_i = \begin{cases} 1 & \text{当 } D(X, W_i) \leq D(X, W_j), j=1, 2, \dots, P, i \neq j \\ 0 & \text{其它} \end{cases} \tag{2}$$

在上式中, $D(X, W_i) = (\sum_{k=1}^n (X_k - W_{ik})^2)^{\frac{1}{2}}$, $X = (x_1, x_2, \dots, x_n)$, $a(t)$ 为时刻 t 的学习率,是随 t 单调下降的,常取 $a(t) = a_0/t, 0 < a_0 < 1$.

SCL 算法存在一个“死单元”问题,即输出层的某些单元,因从未赢得竞争而变成“死单元”。为此,Kohonen 提出一种自组织映射学习方法(KSFM),该算法在学习前先定义得胜单元的邻域,在学习过程中,得胜单元及其邻域单元的权值都被修改^[8]。易知 KSFM 算法的计算量较大。文献[9]提出一种既能克服“死单元”问题,又不增加计算量的算法,称为“良心”竞争学习算法(FSLC)。该算法在计算输入模式 X 与胜者单元的距离时,考虑一个“良心”系数,即记录每个输出单元赢得竞争的次数。设 $F_i(m)$ 为第 i 个输出单元在网络输入第 m 个模式时获胜的次数。某单元得胜的次数越多,其对应的权被修改的次数越多,跟随着其获胜的机会就会减少,而其它单元获胜的机会就会增多。具体地,其计算距离的公式可改为:

$$DF(X, W_i) = D(X, W_i) \cdot F_i \quad (3)$$

FSCL 的学习规则与(1)式相同,但输出单元的值变为:

$$Y_i = \begin{cases} 1 & \text{当 } DF(X, W_i) \leq DF(X, W_j), j=1, 2, \dots, P, j \neq i \\ 0 & \text{其它} \end{cases} \quad (4)$$

虽然 FSCL 算法不增加计算量就解决了“死单元”问题,但文献[10]指出其与传统模式识别方法中的统计理论不兼容。为了避免这种情况,我们将学习过程分为两步:第 1 步,在竞争时,采用(3)式和(4)式,动态地修改其权重矩阵;第 2 步,在学习结束时,即权值的变化小于某一正数 $\epsilon > 0$ 时,按照(2)式判定其分类。为了缩短学习时间和提高分类精度,我们提出一个有监督的改进型 FSCL 算法,简称为 SFSCCL 算法。SFSCCL 法在监督学习时,每个输入模式向量 X 归属于哪一类是事先知道的。网络输入 X 后,输出层的单元互相竞争,并产生出胜者。例如设单元 g 为胜者。若单元 g 对应的是 X 的希望分类(事先已知),则将单元 g 对应的连接权向量向 X 靠拢,否则使之偏离 X ,其学习规则可表示为

$$W_g(m+1) = \begin{cases} W_g(m) + a(t)(X - W_g(m)), & g \text{ 为希望分类} \\ W_g(m) - a(t)(X - W_g(m)), & g \text{ 为别的分类} \end{cases} \quad (5)$$

综上所述 SFSCCL 法可描述为:

第 1 步:首先将 L_A 层到 L_B 层单元之间的所有连接权随机地赋予 $[0, 1]$ 区间的值,且将所有输出单元的获胜次数初始化为 1,即 $F_j = 1, j = 1, 2, \dots, p$ 。

第 2 步:对每个输入模式 X ,进行下述操作:

(1)求
$$DF(X, W_k) = \min_{i=1}^P DF(X, W_i)$$

其中 $DF(X, W)$ 的计算公式与(3)式相同。

(2)根据单元 g 是否对应于希望分类,修改 W_g 的权值:

$$\Delta W_g = \begin{cases} a(t)(X - W_g) \\ -a(t)(X - W_g) \end{cases} \quad a(t) = a_0/t, 0 < a_0 < 1$$

(3)修改单元 g 的获胜次数, $F_k = F_g + 1$ 。

第 3 步:对不同的时刻 t ,重复步骤 2,直到权值变化 $|\Delta W|$ 小于某一正数 $\epsilon > 0$ 。

由于学习率 $a(t)$ 随 t 的增大而逐渐趋于零,从而保证学习过程是收敛的,因而可以逐层逐层地生成神经树网。

3 神经树网的分类性能

如上所述,神经树网从根节点开始,将样本逐层逐层地进行划分。文献[11]指出多类分

类器中的线性机器形成的决策区域是凸的. 我们将证明本文提出的 SFSCCL 算法, 其神经树网的非叶节点也形成凸的决策区域.

设 R_i 为对应于内节点 t 的样本子空间(区域), P_i 为 R_i 中的样本类别数. 根据 SFSCCL 学习算法, 在学习结束时, 按(6)式将输入模式向量进行分类. 即 X 分为第 i 类, 当且仅当

$$Q_i(X) \leq Q_j(X), j \neq i, 1 \leq i, j \leq P_i \tag{6}$$

其中 $Q_i(X) = D(X, W_i)$. 设由(6)式定义的集合表示为:

$$R_i = \{X | D(X, W_i) = \underset{j=1}{P_i} \text{MIND}(X, W_j)\}$$

设 $X, Y \in R_i$, 易知 R_i 是凸的, 当且仅当

$$aX + (1-a)Y \in R_i, 0 < a < 1$$

事实上, 因 $X \in R_i$, 故有

$$D(X, W_i) \leq D(X, W_j) \quad j \neq i, 1 \leq j \leq P_i$$

即有

$$\begin{aligned} & \sum_{k=1}^n (x_k - W_{ik})^2 \leq \sum_{k=1}^n (x_k - W_{jk})^2 \\ \Rightarrow & \sum_{k=1}^n x_k^2 - 2 \sum_{k=1}^n x_k \cdot W_{ik} + \sum_{k=1}^n W_{ik}^2 \leq \sum_{k=1}^n x_k^2 - 2 \sum_{k=1}^n x_k \cdot W_{jk} + \sum_{k=1}^n W_{jk}^2 \\ \Rightarrow & -2 \sum_{k=1}^n x_k \cdot W_{ik} + \sum_{k=1}^n W_{ik}^2 \leq -2 \sum_{k=1}^n x_k \cdot W_{jk} + \sum_{k=1}^n W_{jk}^2 \end{aligned} \tag{7}$$

$$\text{同理有:} \quad -2 \sum_{k=1}^n y_k \cdot W_{ik} + \sum_{k=1}^n W_{ik}^2 \leq -2 \sum_{k=1}^n y_k \cdot W_{jk} + \sum_{k=1}^n W_{jk}^2 \tag{8}$$

(7) × a + (8) × (1-a) 整理得:

$$-2 \sum_{k=1}^N [ax_k + (1-a)y_k] W_{ik} + \sum_{k=1}^N W_{ik}^2 \leq -2 \sum_{k=1}^N [ax_k + (1-a)y_k] W_{jk} + \sum_{k=1}^N W_{jk}^2$$

上式两边加上 $\sum_{k=1}^N [ax_k + (1-a)y_k]^2$, 化简得:

$$\sum_{k=1}^N \{ [ax_k + (1-a)y_k] - W_{ik} \}^2 \leq \sum_{k=1}^N [ax_k + (1-a)y_k] - W_{jk} \}^2$$

也即 $D(aX + (1-a)Y, W_i) \leq D(aX + (1-a)Y, W_j) \quad j \neq i, 1 \leq i, j \leq P_i$

故有 $aX + (1-a)Y \in R_i$, 即证得 R_i 是凸的.

上述分析表明, 神经树网将样本空间逐层逐层地划分为凸区域, 直到每个区域只含一个类别的样本为止, 这与决策树类似. 不过决策树是建立在传统的贝叶斯方法上, 须先求出样本数据的统计密度函数. 神经树网却不用求出密度函数. 其次, 神经树网的结构是在学习中自动生成的. 因而网络结构可以较好地适合于分类问题.

4 模拟数据例子

我们设计的模拟数据例子是将神经树网用于手写体数码的识别. 实验结果表明, 其效果是令人满意的.

我们用 8 × 8 的方格手写“0”到“9”10 个数码, 获得 100 个手写数码点阵模式. 每个模式用 1 个 64 维的 0, 1 向量表示, 作为输入模式向量训练神经树网.

100 个样本共有 10 个类别,每一类各有 10 个样本.学习刚开始时,神经树网的根节点中的神经网络有 10 个输出单元,此神经网络迭代 426 次就收敛,学习结束后只有 8 个有效单元,这时的聚类结果为(见图 2):

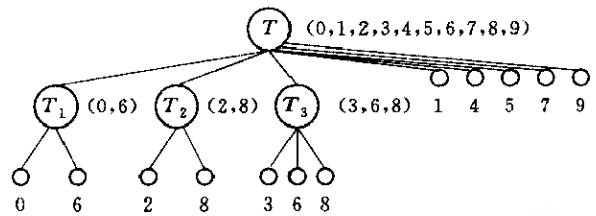


图2 一个用数码识别的神经树网络

$((0,6), (2,8), (3,6,8), 1, 4, 5, 7, 9)$

也即根节点 T 生成 8 个子节点,其中 5 个节点仅包含一个类别的样本,便成了叶节点. T_1 包含有两个类的样本,因而神经网络只有两个输出单元,算法迭代了 134 次网络收敛,聚类结果为 $(3,5)$,即生成两个叶节点.节点 T_2, T_3 经过再学习,各迭代 144、294 次,学习结束.聚类结果分别为 $(2,8)$ 、 $(3,6,8)$.至此整个神经树网的训练学习结束,生成一个三层的树状网络结构(如图 2 所示),有 4 个内节点、12 个叶节点,共利用 15 个有效的输出神经单元,总迭代次数才 998 次,这要比一般的神经网络快得多.

识别时,待识别的模式依次提供给神经树网,模式进入神经树网的根节点,经各节点中的神经网络把模式分到适合的子节点,依此类推,直到模式进入叶节点,该叶节点的标识便是模式的类别.首先用训练样本作为测试数据,来测试此神经树网的分类性能.结果表明,对于训练样本,神经树网的识别率达到 100%.

对于没有训练过的样本,我们的实验表明:对不同的人、不同风格的手写体,例如“2”和“ α ”等会出现较显著的误差,但稍经训练后,对同一人、同一风格的手写体,即使没有训练过的样本,也是相当准确的.我们解决的办法是设各内部节点的神经网络的输出单元数都为 P 个(被识别模式的类别数).易知,这时可能增加一些计算量,产生一些无效单元.但这些单元在识别未经训练过的模式时,恰好可以用于动态进行标识(或者称为动态学习),无须重新系统选择训练例子.

参考文献

- 1 吴佑寿等.汉字识别——原理·方法与实现.北京:高等教育出版社,1992.85—89.
- 2 Iwata Akira, Tohma Takashi, Matsuo Hiroshi *et al.* A large scale neural network 'ComNET' and its application to chinese character recognition. INNC July 9—13, PARIS, 1990. 83—86.
- 3 Bae Cho Song, H Kim Jin. Recognition of large—set print HANGUL by two—stage BP neural classifier. Patten Recognition, 1992, 25(11):1553—1360.
- 4 Rosenblatt F. The perception: a perceiving and recognizing automation. Cornell Aeronautical Laboratory Report 85—460—1, 1957.
- 5 Widrow B, Hoff W. Adaptive switching circuits. WESCON Convention Record, Part 4, 96—104, 1960.
- 6 John Hertz. Introduction to the theory of neural computation. 1990.
- 7 S Grossberg. Competitive learning: from interactive activation to adaptive resonance. Cognitive Science, 1987, 11: 23—63.
- 8 T Kohonen. Self organization and associative memory. Springer Verlag, 1984.
- 9 S C Ahalt, A K Krishnamurthy, P Chen *et al.* Competitive learning algorithms for vector quantization. Neural Network, 1990, 3(3):277—290.

- 10 Yin Hao, Lengelle Regis, Gaillard Paul. Inverse—Step competitive learning. IEEE, IJCNN, 1991. 839—844.
- 11 边肇祺等. 模式识别. 北京:清华大学出版社, 1992.
- 12 Kohonen T. Learning vector quantization for patten recognition. Helsinki Univ. of Tec. , Dept. of TPTR, 1986. TKK—F—AG01.

A SELF—GROWING NEURAL TREE NETWORK PATTERN CLASSIFIER

Li Jigui Lin Guofu Li Guanying

(Department of Computer Science, South China Normal University, Guangzhou 510631)

Abstract This paper is about a neural tree network (NTN) model. A supervised competitive learning algorithm is presented to grow this architecture, which can effectly solve the complex multicategory classification. A simulation result also presented on handwritten numeric recognition shows effectiveness of this classifier.

Key words Neural network, pattern classify, decision tree, neural tree network, competitive learning, clustering.