

非线性 CC 神经元模型的原理及其学习算法^{*}

The Principle & Learning Algorithm of Nonlinear CC Neural Model

黄金才 陈文伟 陈元

(国防科技大学管理科学与工程系 长沙 410073)

Abstract On the base of the principle of FP covering algorithm, the paper puts forward a new nonlinear neural model — CC model and the construct algorithm of neural networks based on CC model. The paper has proved in theory that the computing workload of neural networks based on CC model just is $O(p^3)$, where p is the number of the examples while their fault-tolerance ability is better and the number of their middle nodes is far fewer than those of neural networks based FP algorithm. The networks have provided with a clear geometrical and physical presentation of examples, the ability of incremental learning and noise data processing. The model and networks presented in the paper own greater meaningfulness in theory and application performance.

Keywords Neural Networks, Nonlinear, CC Model, FP Algorithm

1 M-P 神经元模型的工作原理和几何意义

1943年, McCulloch 和 Pitts^[1]根据神经元传递规律,第一次提出了神经元的数学模型。M-P 神经元模型一直沿用至今,它对神经网络的发展起到了奠基性的作用。每个神经元的状态由 M-P 方程决定: $S = f(\sum_{i=1}^n W_i X_i - \theta)$, θ 为阈值, f 为激励函数,一般取符号函数。令:它代表了 n 维空间中,以 X 为坐标变量,以 W 为坐标系数, θ 为常数项的一个超平面。当

$$f(X) = \sum_{i=1}^n W_i X_i - \theta = 0$$

样本点 X 落入超平面的正半区,即 $I(X) > 0$ 时,有 $f(I) = 1$; 当样本点 X 落入超平面的负半区,即 $I(X) < 0$ 时,有 $f(I) = 0$ 。从分类的角度看,一个神经元按输入将样本划分成为两类(0 和 1)。现在广泛使用的 BP 模型采用 Sigmoid 函数作为激励函数,但是它没有改变神经元分类的本质。神经网络实际上就是多个神经元组织起来的一种网状结构。

图 1 说明了神经网络完成对样本的分割。神经网络的学习算法就是确定各个超平面的位置(W 和 θ),使得每个分割区域只对应一种类型的样本,从而完成分类。

M-P 神经元模型通过线性超平面完成对样本空

间的划分。在此我们把基于 M-P 神经元模型的神经网络称之为线性神经网络。线性神经网络在学习和应用时存在明显的不足,以 BP 神经网络模型为例,它具有如下不足:

- (1) 收敛不稳定,容易陷入局部极小。
- (2) 非单调行为。增加隐层节点可以提高训练样本的性能,但是测试样本的性能会下降。隐层节点少,样本不收敛,隐层节点多,则系统过于泛化,测试样本性能下降。
- (3) 不符合生物原理。每个神经元根据传播的方向不同执行的是两种不同的工作模式。
- (4) 网络结构难以确定。隐层节点的个数的确定没有理论上或实践上的指导。

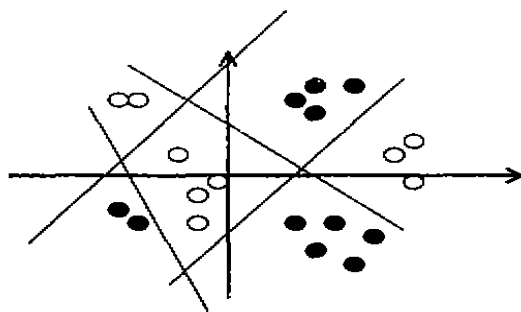


图 1

^{*} 本文研究得到国家自然科学基金项目的资助。黄金才 博士研究生,主要研究神经网络、数据开采等。陈文伟 博士生导师,主要研究人工智能、数据开采、数据仓库、决策支持系统等。

因此,神经网络研究的前沿课题就是通过解析方法来获取网络的权值或者构造其它更加有效的新的神经元模型。

本文主要从分类(学习样本的输出是离散的)角度来探讨新型神经元模型的构造和神经网络的解析求解。

2 FP 算法

FP 算法是以张钹院士为代表的研究小组于 1995 年提出的一种新的多层前馈神经网络学习算法。其网络结构如图 2 所示。其中 n 为输入向量维数, m 为输出向量维数, p 为样本个数。给定训练样本集 $K = \{r^0 = (x^0, y^0), r^1 = (x^1, y^1), \dots, r^{p-1} = (x^{p-1}, y^{p-1})\}$, 各分量取值 $\{1, -1\}$ 。其第一层网络权值 W 和阈值 θ 由下式给出:

$$w^i = (x^i)^T \quad i = 1, 2, \dots, p-1$$

$$\theta = \begin{cases} n-d_i-1 & d_i \text{ 为偶数} \\ n-d_i & d_i \text{ 为奇数} \end{cases}$$

$$W = (w^i), \theta = (\theta)$$

其中 $d_i = \min_{i \neq j} d(x^i, x^j) \quad i = 1, \dots, p-1$

$d(x, y)$ 表示向量之间的海明距离。

第二层网络权值 U 和阈值 ξ 由下式给出:

$$u_j = \begin{cases} 0 & \text{当 } y^0 = y_j^i \\ y_j^i & \text{其它} \end{cases} \quad i = 1, \dots, m; j = 1, \dots, p-1$$

$$\xi = -(y^0 + u_1^1 + \dots + u_{p-1}^1)$$

$$U = (u^i), \xi = (\xi)$$

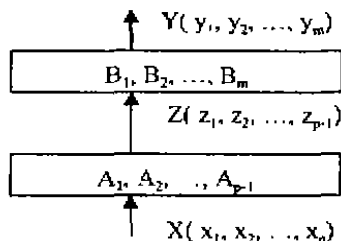


图 2

从网络权值和阈值的定义可以看出,FP 算法以向量的内积为相似性来联想与输入特征最相似样本的输出。可以比较容易地证明按照所定义的权值和阈值,上图的网络可以把 x^i 映射到 y^i , 即每个样本都是吸引中心,且每个样本的吸引域为 $D(x^i) = \{x | (x, x^i) \geq \theta^i / 2\}$ 。

文[2]已经证明,作为通用的联想记忆器,FP 给出的网络是最优的。但是从应用的角度来看,FP 算法并没有太大的价值。其隐层节点数目为 $p-1$, 在样本数量很多时,网络显得很庞大,而且网络对样本的依赖性很强。从知识发现的角度来看,在所发现的知识存储量小于原数据的存储量的情况下,所得到的知识才是有价

值的。显然,FP 算法无法满足这一要求,按照 FP 算法所设计的网络权值和阈值的存储量与原样本的存储量几乎是相当的。另外用内积来衡量样本之间相似度也是值得探讨的问题,至少它不易于理解。

不过,FP 算法对于探讨神经网络的解析求解方法以及神经网络的工作原理具有积极的指导意义。本文所提出 CC 神经元模型是受 FP 算法的启发而提出的。

3 CC 模型

同 FP 算法,我们认为每个神经元(不一定是样本)都有一个吸引域(作用域),位于吸引域内的输入数据响应同一种输出(类别),这也与 M-P 神经元模型对输入空间的划分思想是一致的,每个神经元的吸引域就是一个空间分割块,位于分割块中的输入特征对应同一种输出类型。基于这种思想,我们提出了一种更加容易理解和操作的 CC(Circle Covering)模型。

每个 CC 神经元模型的输出状态由下式决定:

$$y = \text{sgn}(\sum (x - a_i)^2 - c^2)$$

$A(a_1, a_2, \dots, a_n)$ 为神经元的特征向量,不同的神经元由不同的特征向量进行描述, c 为神经元的吸引域,即若新的输入样本特征与神经元特征向量的离合度位于吸引域内(小于 c),则输出该神经元所标识的类别,特征 u 和 v 离合度由下式定义:

$$d(u, v) = \sqrt{\sum (u_i - v_i)^2}$$

可以看到,特征 u 和 v 的离合度越小,二者越相似。为了便于表述,以 $c(A)$ 表示神经元 A 的吸引域的大小, $d(x, A)$ 表示样本 x 与神经元 A 的特征离合度, $y(A)$ 表示神经元 A 所覆盖样本的输出类别, sgn 符号函数定义如下:

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

CC 模型具有很好理解的物理意义和几何意义。它用超圆(非线性)而不是超平面(线性)来分割输入空间,超圆可以和数据的聚类结合起来,根据聚类的结果建立相应的网络结构。由于数据聚类的思想与 CC 神经元模型的超圆覆盖极为相似,因此可以比较容易地根据数据聚类结果来构造相应的神经网络,而不需要进行如 BP 模型那样的权值调整。

4 基于 CC 模型的神经网络学习算法

基于 CC 模型的神经网络把网络结构的确定与网络的学习结合起来,从而在保证系统的每个样本都能够被识别的情况下,达到整个神经网络的容错范围最大,网络结构最优(或次优)。

4.1 网络结构

网络结构如图 3 所示,其中第一层为 CC 神经元层,由 M 个 CC 神经元完成对整个样本空间的划分,即每个样本都应该位于 M 个 CC 神经元中的其中一个神经元的吸引域内。每个 CC 神经元对应一个输出类别,位于同一神经元吸引域的样本的输出类别是相同的。后面介绍的学习算法的目标就是找到满足上述要求 CC 神经元的最少个数。M-P 神经元层根据 CC 神经元层对输入特征的评价,输出该输入特征对应的输出类别。K 为给定样本的总的类别数目。

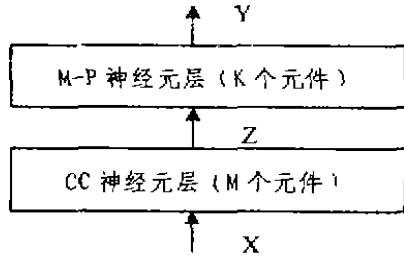


图 3

4.2 网络构造算法

网络构造算法由两个部分组成:(1)CC 神经元层的构造;(2)M-P 神经元层的构造。

4.2.1 CC 神经元的构造 与 ART 神经网络的学习思想相近。每个输入分量到 CC 神经元层都是全连接,其权值为 1.0。样本经当前 CC 神经元层中的每个 CC 神经元进行评价,选出与神经元特征最相似(d 最小)的同类神经元(其对应的输出类别与样本类别一致),将该样本加入到此神经元,调整其特征取值及吸引域。若调整后的此神经元吸引域包含异类样本,则以该样本为中心,建立新的 CC 神经元。其具体操作步骤如下:

样本集: $R = \{(x^1, y^1), \dots, (x^p, y^p)\}$

1) CC 神经元个数 $M=0$, 样本编号 $p=1$;

2) 若 $p > P$, 则转 9;

3) 对于样本 $p (1 \leq p \leq P)$, 若 $M=0$, 则转 8, 否则计算其输入与各个同类 CC 神经元特征的相似度 $d(x^p, A^m), m=1, \dots, M$;

4) 若无同类的神经元, 则转 8;

5) 设 $d(x^p, A^{n_0})$ 最小, 建立临时神经元:

$$A^j = (A^{n_0} * \text{Count}(A^{n_0}) + x^p) / (\text{Count}(A^{n_0}) + 1)$$

$$j = 1, \dots, n$$

$$c(A^j) = \max\{d(A^j, x') \mid x' = x^p \text{ 或 } x' \text{ 为被 } A^{n_0} \text{ 覆盖的样本}\}$$

其中, $\text{Count}(A^{n_0})$ 为 A^{n_0} 覆盖的样本个数, n 为样本维数。

6) 判断 A^j 是否覆盖了其他非本类的样本点或已经被

其他神经元覆盖的样本点。

7) 若无, 则 $A^m = A^j, p = p + 1$, 转 2; 否则转 8;

8) 生成新神经元

$$M = M + 1$$

$$A^M = x^p$$

$$c(A^M) = 0$$

$p = p + 1$, 转 2;

9) 扩展各神经元的吸引域:

$$c(A^m) = c(A^m) + \min\{d(A^m, A^n) - c(A^m) - c(A^n)\} / 2 \mid m < n, 1 \leq n \leq M \\ m = 1, 2, \dots, M$$

扩展各神经元吸引域的步骤要同时进行。

其中步骤 9 是为了在不影响每个样本都被识别的情况下, 尽量扩展每个 CC 神经元的吸引域。

4.2.2 M-P 神经元层的构造 CC 神经元层的 M 个元件到 M-P 神经元层的 K 个元件是全连接。其权值由下式定义:

$$W_{mk} = -(y(A^m))_k, \quad m = 1, \dots, M, \quad k = 1, \dots, K$$

$$\theta_k = -\sum_m W_{mk}$$

即权值 W_{mk} 取神经元 A^m 覆盖样本输出的第 k 个分量。

定理 1 基于上述方法构造的由 CC 神经元层和 M-P 神经元层构成的神经网络能够完成从样本输入到样本输出的映射。

证明: 对于样本 $r = \{x, y\}$, 由上面的 CC 神经元层的构造算法可以得知: 该样本 r 一定被输出为 y' 一个 CC 神经元所覆盖, 设第 m 个 CC 神经元。则样本 r 经过 CC 神经元层评价后的输出向量为:

$$(1, 1, \dots, 1, 0, 1, 1, \dots, 1)$$

那么对于输出层的 M-P 神经元 k :

$$y_k = \text{sgn}(\text{net}_k - \theta_k)$$

$$\text{net}_k = (1, 1, \dots, 1, 0, 1, 1, \dots, 1) \cdot (-y(A^1)_k, -y(A^2)_k, \dots, -y(A^M)_k)^T = -\sum_{m=1}^M y(A^m)_k$$

$$y_k = \text{sgn}(-\sum_{m=1}^M y(A^m)_k - \sum_m y(A^m)_k) = \text{sgn}(y(A^m)_k) = \text{sgn}(y'_k)$$

即 $y = \text{sgn}(y')$, 证毕。

对于取值连续的样本, 神经元的激励函数可以取(设样本分量取值范围为 $[-1, 1]$):

$$f(x) = \frac{2}{1 + e^{-x}} - 1, 0$$

5 CC 神经元模型及基于 CC 模型神经网络的分析

我们现在从下面几个方面来分析 CC 神经元模型

及基于 CC 模型的神经网络的性能和应用。

(1)相似度的定义。FP 算法采用两个向量之间的内积来定义相似度。虽然这种定义在某种情况下是适合的,但是它并不是一种易于理解和接受的相似度衡量标准。例如向量 0 和任意向量之间的内积都是 0,而 CC 神经元模型对样本之间相似度的定义要灵活得多,本文所推荐使用的相似度度量方法就很容易理解和被人接受,而且它的物理意义比较明确。

(2)学习复杂度。基于 M-P 模型的 BP 算法采用梯度下降法来确定网络的权值,其具有很好的适用能力,但是研究表明,BP 算法的复杂度是学习样本的指数关系。FP 算法的神经网络采用构造方法进行学习,其计算复杂度仅为 $O(p^2)$, p 为样本数量。从前面介绍的基于 CC 神经元的神经网络构造算法可以看出,每个样本都要与网络的 CC 神经元层中每个神经元进行比较(计算离合度),而 CC 神经元层中的神经元个数小于 p ,所以其学习复杂度要小于 $O(p^2)$,因此此算法对于大规模数据的学习是非常适合的。

(3)持续学习能力。基于 CC 模型的神经网络具有持续学习能力,在增加新的学习样本时,以前的样本不需要进行重新学习。当前的大多数神经网络权值和结构在增加新的学习样本时都需要重新进行学习,即不具有持续学习能力。持续学习能力是知识发现和数据挖掘领域对算法的一个重要要求。算法持续学习能力的研究一直是机器学习研究的一个重要方向。本文提出的基于 CC 神经元模型的神经网络对神经网络持续学习能力的研究是一个重要的参考和借鉴。

(4)易于理解的网络结构,对于与其他系统的有机集成,BP 网络的“黑箱”结构是一个难以逾越的障碍。FP 算法构造的神经网络虽然赋予每个神经元以明确的意义,但是这种神经网络抛弃了传统神经网络对样本类分割的本质属性。基于 CC 神经元的神经网络具有基于理解的物理结构和几何意义。它可以与数据开采的聚类算法结合建立基于聚类的分类器。而且网络构造完成之后,可以非常容易地从中提出用于分类的产生式规则。

(5)容错性。容错性是神经网络的一个基本特征,即对于有一定误差的输入,神经网络能够输出与之最近样本的输出。在基于 CC 模型的网络中,最特殊的情况就是每个 CC 神经元只覆盖一个样本。此时基于 CC 神经元模型的神经网络和 FP 构造的算法具有相同或相似系统结构,二者的吸引域是相同的。但是在一般情况下,CC 神经元覆盖的是具有相近特征的一类样本,CC 神经元的吸引域不但比被覆盖的每个样本的吸引域要大、容错性好,而且更加有意义。

(6)隐层神经元的个数,FP 算法确定的神经网络

隐层神经元个数为 $p-1$ 个。基于 CC 神经元模型的神经网络中隐层神经元个数最大为 p ,不过在一般情况下,其个数要远远小于 p 。下表是我们在 $[0,1] \times [0,1]$ 空间随机产生的两类数据(0 类和 1 类)样本数目与隐层神经元之间的关系:

样本数据规模	500	200	100	50	10
隐层神经元数目	141	55	32	14	6

由于实际数据具有明确的意义,同类数据会更加集中,因此在实际应用中对于同样规模的样本数据,隐层神经元个数要比上表中的还要小得多。

(7)噪声数据的处理。噪声数据一般情况下会影响神经网络的学习效率和应用效果。传统的神经网络无法确定哪些数据是噪声,而基于 CC 神经元模型的神经网络具有噪声数据处理的能力。在 CC 神经元构造完成以后,可以删除那些覆盖较少样本或具有较小吸引域的神经元,这些神经元对样本分类不具有代表性,其覆盖的样本可以被认为是噪声数据。

6 实例和应用

下面用 Takagi(1991 年)给出的一个仿真例子^[5](表 1)来比较一下基于 CC 模型的神经网络和 BP 神经网络的性能。其测试结果如表 2。

从学习结果和测试结果来看,基于 CC 模型的神经网络对分类学习是适用的,但它在学习时间和网络模型的可理解上都优于 BP 神经网络,学习结束后用于表示网络的参数要少于 BP 网络,即它能够以较少的数据量蕴涵同样的信息量。对于基于 CC 模型的神经网络的其他特性,限于篇幅,在此就不一一测试了。

本文在 FP 算法思想的基础上提出了一种 CC 神经元模型。它是采用非线性方式来完成对样本空间的划分。可以看到,这种基于 CC 模型的神经网络能够完成对样本的最优或次优划分。由于它学习复杂度低,仅为 $O(p^2)$,网络结构易于理解,使用较少的网络隐层神经元,具有持续学习能力和较强的容错能力。基于 M-P 模型的许多神经网络,如 BP 模型,从理论上讲是可以得到系统样本空间最优划分,但是要得到这种最优或次优划分是非常困难的,因为从解析的角度讲,获得最优的样本划分问题是一个 NPC 问题;从梯度学习方法讲,学习速度和容易陷入极小状态又是一个难以逾越的难题。用构造方法进行神经网络建模是近年来研究的热点,本文提出的 CC 神经元模型对于如何寻找有效的构造神经网络的方法具有积极的意义。基于 CC 神经元模型的神经网络对于大规模数据的聚类分析、分类、预测等应用是非常适合的。

(下转第 25 页)

因而 Ta 和 Tm 相差不大,而测试三中采用的微机计算性能差别较大,因而数据自动迁移大大缩短了并行程序运行的时间。

测试	使用微机	Ta(秒)	Tm(秒)
测试一	1,2	315	282
测试二	1,2,3,4	165	162
测试三	2,5,6,7	1015	421

4.2 计算节点负载变化对程序性能的影响

测试三使用微机 1,2,后台运行 CFD 计算程序,前台运行其他用户程序,测试得到:对于平均分配网格数据的并行程序,运行时间为 413 秒;对于数据自动迁移的并行程序,运行时间为 349 秒,由于采用的是在 X 方向划分,两个节点程序在 X 方向上的数据宽度随时间变化如图 5 所示。



图 6 数据宽度随时间的变化

结束语 采用 ADM 模型也会带来一定的额外开销。一般情况下由节点计算能力差异而导致的数据迁移会较大地提高并行程序的运行效率,并且额外开销较小;如果由于负载变化(特别是负载变化频繁时)导致较为频繁的数据迁移,额外开销较大,但是如果负载变化不很频繁,由于迭代过程中计算时间远远大于数据迁移所用的时间,其额外开销还是可以忽略的。

另外,ADM 模型不但适用于 CFD,对于数值天气预报等网格类计算程序也同样适用。

参考文献

- 1 William B. et al. Parallel Programmung with Polaris. IEEE Computer, 1996, 29(12): 78~82
- 2 Hall Mary W, et al Maximizing Multiprocessor Performance with the SUIF Compiler. IEEE Computer, 1996, 29(12): 84~89
- 3 冯百明. 基于分区的自动并行化程序重构技术研究.[博士学位论文]. 西安:西北工业大学, 1998
- 4 朱传琪, 臧斌宇, 陈彤. 程序自动并行化系统. 软件学报, 1996, 7(3): 180~186
- 5 金国华, 陈福接. KD-PARPRO: 一个基于知识的并行化工具-总体设计与功能描述. 软件学报, 1993, 4(6): 1~6
- 6 况正谦. 程序自动并行化的场循环相关分析与优化技术.[博士学位论文]. 西安:西北工业大学, 1998

(上接第 36 页)

表 1 仿真数据

学习样本				测试样本					
No.	类	x1	x2	x3	No.	类	x1	x2	x3
1	0	1	3	1	21	0	1	1	5
2	0	1	5	2	22	0	1	3	4
3	0	1	1	3	23	0	1	5	3
4	0	1	3	4	24	0	1	1	2
5	0	1	5	5	25	0	1	3	1
6	1	5	1	4	26	1	5	5	2
7	1	5	3	3	27	1	5	1	3
8	1	5	5	2	28	1	5	3	4
9	1	5	1	1	29	1	5	5	5
10	1	5	3	2	30	1	5	1	4
11	0	1	5	3	31	0	1	3	3
12	0	1	1	4	32	0	1	5	2
13	0	1	3	5	33	0	1	1	1
14	0	1	5	4	34	0	1	3	2
15	0	1	1	3	35	0	1	5	3
16	1	5	3	2	36	1	5	1	4
17	1	5	5	1	37	1	5	3	5
18	1	5	1	2	38	1	5	5	4
19	1	5	3	3	39	1	5	1	3
20	1	5	5	4	40	1	5	3	2

表 2 测试结果

类型	基于 CC 模型的神经网络	BP 神经网络
环境	Intel PII366/44RAM	
参数设置	3 个输入节点	3 个输入节点, 2 个隐节点, 1 个输出节点, 最大允许误差 0.2
学习时间	0.02 秒	0.96 秒
学习结果	学习的结果是 CC 神经元层有两个节点, 其参数为: 节点 1: $u = (1.0, 3.2, 3.4)$ $r = 1.99$ $Class=0$ 节点 2: $u = (5.0, 3.0, 2.4)$ $r = 2.14$ $Class=1$	得到 8 个连接参数和 3 个阈值参数
测试结果	正确率 100%	正确率 100% (输出大于 0.9 是 1, 小于 0.1 是 0)

参考文献

- 1 McCulloch G A, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematics Biophysics, 1943, 5: 115~133
- 2 张玲, 等. 多层前馈神经网络的学习和综合算法. 软件学报, 1995, 6(7): 440~448
- 3 Tomas H. Modular Learning in Neural Networks. John Wiley & Sons, Inc., 1992
- 4 张玲, 等. M-P 神经元的几何意义及其应用. 软件学报, 1998, 5: 334~338
- 5 Takagi H, Hayashi I. NN-driven fuzzy reasoning. Int. Journal of Approximate Reasoning, 1991, 5: 191~213