二进神经网络隐元数目最小上界研究

陆 阳 韩江洪 高 隽

(合肥工业大学 微型计算机应用研究所 合肥 230009)

摘 要 本文通过讨论布尔函数样本空间样本连通性和线性可分的关系,导出含一层隐元的二进神经网络隐元数目的最小上界为 2^{n-1},并构造出了必须用 2^{n-1} 个隐元实现的布尔函数.

关键词 二进神经网络,连通集,线性可分,最小上界

中图法分类号 TP183

1 引 言

在用前向网络研究布尔函数时,存在样本空间的线性可分问题,即只有用超平面可划分的样本子集才可以用一个隐元表示.许多研究证明,含有一层隐元的二进阈值神经网络能够完备表达任何布尔逻辑,并且在隐元数目对实现完备布尔逻辑的充分条件上取得了一些成果[1-5],但是在隐元数目对实现完备布尔逻辑的必要条件上,即必须采用多少隐元才可以完备表达布尔逻辑的问题上,目前还没有明确的结论.对于这个问题有一些建立在针对特殊事例基础上的实验估计结果[6-8],这些结果由于没有严格的证明过程,因此缺乏针对任何布尔逻辑都适用的理论基础.文献[3]中曾经阐明对于用含一层隐元的二进阈值神经网络实现的 n 元布尔函数,在隐元数目为 2^{n-1} 时,可以达到逻辑完备,但文中并未证明 2^{n-1} 就是隐元数目的最小上界,即必须用 2^{n-1} 个隐元才可以实现任何 n 元布尔函数.我们经过研究发现,2^{n-1} 确实是实现完备布尔逻辑的隐元数最小上界,本文将给出其严格的数学证明,并构造出一个只有用 2^{n-1} 个隐元才能实现的布尔函数.

收稿日期:2000-03-28

2 布尔空间样本连通性

F_2 表示二元域,f 是 $F_2^n \to F_2$ 的任意映射,则 f 称为 n 元布尔函数,F_2^n 构成了布尔函数样本空间.对所有 $X=(x_n,x_{n-1},\cdots,x_2,x_1)\in F_2^n$,以 $f(x_n,x_{n-1},\cdots,x_2,x_1)$ 表示 f 在 X 处的取值,简记为 $f(X)$,称 X 为 F_2^n 中的样本,$f(X)$ 为样本值,$f(X)\in\{0,1\}$.

在下面的讨论中,我们用 $C(A)$ 表示集合 A 中的元素数目,用 $d_H(X^p,X^q)$ 表示二个样本 X^p,X^q 之间的汉明距离.

定义 1 集合 $f^{-1}(1)=\{X\in F_2^n \mid f(X)=1\}$,$f^{-1}(0)=\{X\in F_2^n \mid f(X)=0\}$.

定义 2 $A\subseteq f^{-1}(1)$,若满足下列 2 个条件之一,则称集合 A 是 $f^{-1}(1)$ 中的一个连通集.

(1) $C(A)=1$;

(2) $C(A)\geqslant 2$,且 $\forall X^p,X^q\in A$, $X^p\neq X^q$,存在样本序列 $X^1,X^2,\cdots,X^k\in A$,满足:

$$\begin{cases} d_H(X^p,X^1)=1 & (1)\\ d_H(X^i,X^{i+1})=1 \quad i=1,\cdots,k-1 & (2)\\ d_H(X^k,X^q)=1 & (3)\end{cases}$$

定义 3　若 A 是连通集,且对 $\forall X^p \in A, X^q \notin A, d_H(X^p, X^q) = 1$,满足 $X^q \in f^{-1}(0)$,则称 A 是充分连通集.

定义 4　若样本 $X^g \in f^{-1}(1)$,且 $\forall X^i, d_H(X^i, X^g) = 1$,满足 $X^i \in f^{-1}(0)$,则称 X^g 为 $f^{-1}(1)$ 中的孤立样本.

显然,由定义 4 中孤立样本 X^g 构成的集合 $\{X^g\}$ 是充分连通集.以图 1 中的 5 变量布尔函数卡诺图为例,其中 10101 点是一个孤立样本, {01111, 11111, 11110, 10110} 是一个连通集, {00000, 00010} 和 {01111, 11111, 11110, 10110, 11010, 11011} 是二个充分连通集.

x_5x_4 \ $x_3x_2x_1$	000	001	011	010	110	111	101	100
00	1	0	0	1	0	0	0	0
01	0	0	0	0	0	1	0	0
11	0	0	1	1	1	1	0	0
10	0	0	0	0	1	0	1	0

图 1　5 变量布尔函数卡诺图

连通集的几何意义是指在 n 维布尔空间里由 2^n 个顶点构成的超立方体中,连通集中任何二个样本对应的顶点之间存在一条由数条边构成的路径,该路径只经过样本值为"1"的顶点.

连通集有如下性质:

性质 1　若 A, B 是连通集,且 $A \cap B \neq \phi$,则 $A \cup B$ 也是一个连通集.

性质 2　A 是充分连通集,样本 $X^p \in A$,若 $\exists X^q \in f^{-1}(1), d_H(X^p, X^q) = 1$,则 $X^q \in A$.

性质 3　任何一个非连通集都可以表示为几个充分连通集的并集.

3　连通性与线性可分的关系

只有一个输出的单隐层二进阈值神经网络可以表达为

$$y_j = U\left(\sum_{i=1}^{n} W_{ij}x_i - \theta_j\right)$$

$$U(x) = \begin{cases} 1 & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

$$O = U\left[\sum_{j=1}^{L} W'_j y_j - T\right]$$

其中, $X = (x_n, x_{n-1}, \cdots, x_2, x_1)$ 是输入, y_j 是第 j 个隐元输出, w_{ij} 是隐层权值, L 是隐元数目, O 是输出层输出, W'_j 是输出层权值.

在布尔函数的超平面分类中,连通性和分类特性存在重要关系,在下面的讨论中我们将隐元的分类超平面记为

$$R(X) = \sum_{i=1}^{n} W_i x_i - \theta = 0,$$

若 $\sum_{i=1}^{n} W_i x_i \geqslant \theta$,则 $X \in f^{-1}(1)$,否则 $X \in f^{-1}(0)$.

定理 1　对 $\forall X \in f^{-1}(1)$,存在一个超平面可以将 X 和 $f^{-1}(0)$ 分成二个区域,即 X 可以用一个隐元表示.

定理 2　A 是 $f^{-1}(1)$ 的非空子集,若存在一个超平面可以将 A 和 $f^{-1}(0)$ 分成二个区域,则 A 一定是连通集.

证明　反证法,假设 A 是非连通集.

(1) 若 $C(A) = 1$,根据定义 2, A 是连通集,矛盾.

(2) 若 $C(A) \geqslant 2$,根据性质 3, $A = \bigcup_{i=1}^{m} A_i$, $m \geqslant 2$, A_i 是一组充分连通集.不失一般性,在 A_i 中选取 A_1 和 A_2,一定能够找到一对样本 $X^p \in A_1, X^q \in A_2$,使 $d_H(X^i, X^j) \geqslant d_H(X^p, X^q) \geqslant 2$ 对 $\forall X^i \in A_1, X^j \in A_2$ 成立,即在所有由 A_1 中样本和 A_2 中样本构成的样本对中, X^p, X^q 是汉明距离最小的样本对,不妨设 $d_H(X^p, X^q) = l$.

$$X^p = (x_n^p, x_{n-1}^p, \cdots, x_2^p, x_1^p),$$
$$X^q = (x_n^q, x_{n-1}^q, \cdots, x_2^q, x_1^q),$$

X^p 和 X^q 存在 l 个不同码元,设为第 $Sl, \cdots, S2, S1$ 位码元不同,下面分二种情况讨论.

情况 1:　若 X^p 中的第 $Sl, \cdots, S2, S1$ 位码元不全部为"0",不失一般性,设 $S2$ 位码元为"1",则

$$X^p = (x_n, x_{n-1}, \cdots, x_{Sl}^p, \cdots, 1, \cdots, x_{S1}^p, \cdots, x_2, x_1),$$
$$X^q = (x_n, x_{n-1}, \cdots, x_{Sl}^q, \cdots, \underset{\substack{\uparrow \\ \text{第 } S2 \text{ 位}}}{0}, \cdots, x_{S1}^q, \cdots, x_2, x_1).$$

(a) 若 $W_{S2} \geqslant 0$,则因为 $X^q \in A$,所以

$$\begin{aligned} R(X^q) &= W_1x_1 + W_2x_2 + \cdots + W_{S1}x_{S1}^q + \cdots + \\ &\quad W_{S2}x_{S2}^q + \cdots + W_{Sl}x_{Sl}^q + \cdots + W_nx_n - \theta \\ &= W_1x_1 + W_2x_2 + \cdots + W_{S1}x_{S1}^q + \cdots + 0 + \\ &\quad \cdots + W_{Sl}x_{Sl}^q + \cdots + W_nx_n - \theta \\ &\geqslant 0. \end{aligned}$$

考虑样本

$$X^* = (x_n, x_{n-1}, \cdots, x_{Sl}^q, \cdots, 1, x_{S1}^q, \cdots, x_2, x_1),$$

$$R(X^*) = W_1x_1 + W_2x_2 + \cdots + W_{S1}x_{S1}^q + \cdots + W_{S2} + \cdots + W_{Sl}x_{Sl}^q + \cdots + W_nx_n - \theta,$$

$\because W_{S2} \geqslant 0 \qquad \therefore R(X^*) \geqslant R(X^q) \geqslant 0$

$\therefore X^* \in f^{-1}(1)$ 并且 $d_H(X^*, X^q) = 1$，根据性质 2，$X^* \in A_2$，而 $d_H(X^*, X^p) = l-1 < d_H(X^p, X^q)$，矛盾.

(b) 若 $W_{S2} < 0$，则因为 $X^p \in A$，所以

$$\begin{aligned} R(X^p) &= W_1x_1 + W_2x_2 + \cdots + W_{S1}x_{S1}^p + \cdots + W_{S2}x_{S2}^p + \cdots + W_{Sl}x_{Sl}^p + \cdots + W_nx_n - \theta \\ &= W_1x_1 + W_2x_2 + \cdots + W_{S1}x_{S1}^p + \cdots + W_{S2} + \cdots + W_{Sl}x_{Sl}^p + \cdots + W_nx_n - \theta \\ &\geqslant 0, \end{aligned}$$

考虑样本 $X^* = (x_n, x_{n-1}, \cdots, x_{Sl}^p, \cdots, 0, \cdots, x_{S1}^p, \cdots, x_2, x_1)$，

$$R(X^*) = W_1x_1 + W_2x_2 + \cdots + W_{S1}x_{S1}^q + \cdots + 0 + \cdots + W_{Sl}x_{Sl}^p + \cdots + W_nx_n - \theta,$$

$\because W_{S2} < 0 \qquad \therefore R(X^*) > R(X^p) \geqslant 0$

$\therefore X^* \in f^{-1}(1)$ 并且 $d_H(X^*, X^p) = 1$，根据性质 2，$X^* \in A_1$，而 $d_H(X^*, X^q) = l-1 < d_H(X^p, X^q)$，矛盾.

情况 2： 若 X^p 中第 $Sl, \cdots, S2, S1$ 位码元全部为"0"，则

$X^p = (x_n, x_{n-1}, \cdots, 0, \cdots, 0, \cdots, 0, \cdots, x_2, x_1)$，

$X^q = (x_n, x_{n-1}, \cdots, 1, \cdots, 1, \cdots, 1, \cdots, x_2, x_1)$；

当 $W_{S2} \geqslant 0$ 时，考虑样本 $X^* = (x_n, x_{n-1}, \cdots, 0, \cdots, 1, \cdots, 0, \cdots, x_2, x_1)$；

当 $W_{S2} < 0$ 时，考虑样本 $X^* = (x_n, x_{n-1}, \cdots, 1, \cdots, 0, \cdots, 1, \cdots, x_2, x_1)$.

同理可以得到和情况 1 类似的矛盾结果.

归纳情况 1 和情况 2，都是和假设矛盾的结果，所以原假设不成立. 证毕.

推论 1 $A_i (i = 1, \cdots, m)$ 是 m 个连通集，若 $\bigcup_{i=1}^m A_i$ 是非连通集，则 $\bigcup_{i=1}^m A_i$ 和 $f^{-1}(0)$ 不能被一个超平面所划分.

推论 2 若 A 是由孤立样本组成的集合 $\{X^{g1}, X^{g2}, \cdots, X^{gm}\}$，则对任意 $B \subseteq A$，$C(B) > 2$，B 和 $f^{-1}(0)$ 不能被一个超平面所划分.

证明 反证法. 假设 B 和 $f^{-1}(0)$ 可以被一超平面所划分，则根据定理 2，B 是连通集.

若 B 是连通集，则对于 $\forall X^p, X^q \in B$，$X^p \neq X^q$，在 B 中存在 $X^1, \cdots, X^k$，满足(1) ~ (3) 式，显然根据定义 4，X^p 和 X^q 都不是孤立样本，矛盾. 证毕.

定理 2 表明样本连通性是线性可分的必要条件，只有连通集才可能被一个超平面将其和 $f^{-1}(0)$ 划分在二个区域中. 因此，在 n 元布尔函数的样本空间中，如果能构造出一种样本结构，它具有最多的孤立样本或连通集，则由此样本结构形成的布尔函数将需要最多数目的隐元.

4 最多孤立样本问题

n 元布尔函数样本空间的样本有 2^n 个，对这 2^n 个样本如何排序会直接影响对样本空间不同类型样本分布情况的分析，样本的排序方法十分重要.

定义 4 $X = (x_n, x_{n-1}, \cdots, x_2, x_1)$ 是 n 元布尔函数样本空间的样本，$S = \phi$.

(1) 取 $X^0 = (0, 0, \cdots, 0)$ 为第一个样本，并设 $p = 0$.

(2) $S = S \cup \{X^p\}$，并将 X^p 从样本空间中去除，如果 X^p 是最后一个样本，则停止排序，否则继续(3).

(3) 从样本空间取一新样本 X^{p+1}，满足 $d_H(X^{p+1}, X^p) = 1$，并且对 $\forall X^q \notin S, X^q \neq X^{p+1}$，$d_H(X^q, X^p) = 1$，满足 $i < j$，其中 x_i 为 X^{p+1} 和 X^p 的相异位，x_j 为 X^q 和 X^p 的相异位.

(4) $p = p + 1$，返回(2).

称按上述(1) ~ (4) 步骤进入集合 S 的顺序为 n 元布尔函数样本空间样本的标准排序.

一个 3 元布尔函数的样本标准排序是：000，001，011，010，110，111，101，100. 布尔函数样本空间样本的标准排序过程实际上就是卡诺图中从左而右，从上而下的排列样本过程.

下面我们借助标准排序给出必须 2^{n-1} 个隐元才能实现的布尔函数的构造.

定理 3 若 n 元布尔函数样本空间的样本按标准排序排列，且对该排序中序号相邻的任意二个样本 X^i, X^j，满足 $f(X^i) \neq f(X^j)$，则实现由此样本和样本值构造的布尔函数需要的隐元数目等于 2^{n-1}.

证明 不失一般性，我们设样本标准排序和样本值的关系如下：

序号	样本排列
1	$f(0,0,\cdots,0,0) = 1$
2	$f(0,0,\cdots,0,1) = 0$
3	$f(0,0,\cdots,1,1) = 1$
4	$f(0,0,\cdots,1,0) = 0$
⋮	⋮
2^{n-1}	$f(1,0,\cdots,0,1) = 1$
2^n	$f(1,0,\cdots,0,0) = 0$

可见，所有序号为奇数的样本值都为“1”，序号为偶数的样本值都为“0”，并且对 $\forall X \in f^{-1}(1)$，如果 $d_H(X^T, X) = 1$，那么 X^T 必在序号为偶数的位置上，即 $f(X^T) = 0$. 所以根据定义4，$f^{-1}(1)$ 中所有样本都是孤立样本，孤立样本数目等于 2^{n-1}.

由于 $f^{-1}(1)$ 中的样本都是孤立样本，因此，根据推论2，对任何子集 $A \subseteq f^{-1}(1)$，$C(A) \geqslant 2$，A 和 $f^{-1}(0)$ 不能被一个超平面所划分，又根据定理1，$f^{-1}(1)$ 中的每个样本都能用一个隐元实现，总计需要 2^{n-1} 个隐元. 证毕.

图2以4元布尔函数为例(以卡诺图表示)，给出必需 2^{n-1} 个隐元才能实现的最多孤立样本构造. 在这个例子中，所需要的隐元数目为 8.

x_4x_3 \ x_2x_1	00	01	11	10
00	1	0	1	0
01	0	1	0	1
11	1	0	1	0
10	0	1	0	1

图 2 4 变量布尔函数卡诺图

5 结 论

我们研究了布尔函数样本空间连通性和线性可分的关系，得到了 $f^{-1}(1)$ 中任何子集和 $f^{-1}(0)$ 线性可分的必要条件. 将本文的定理3和文献3结合可以得出以下重要结论：

定理 4 具有一层隐元的二进阈值神经网络实现 n 元完备布尔函数的充要条件是隐元数目等于 2^{n-1}.

此定理说明 2^{n-1} 确实是含有一层隐元的二进神经网络完备实现 n 元布尔函数的隐元数最小上界.

参 考 文 献

[1] 马晓敏，杨义先，章照止. 二进神经网络学习算法研究. 计算机学报，1999，22(9)：931—935

[2] 刘喜成，韩承德. 感知器的布尔映射能力分析. 模式识别与人工智能，1997，10(12)：301—304

[3] 张军英，保 铮. 二进神经网络的最稳健设计. 电子学报，1997，25(10)：37—43

[4] 刘文举. 基于逻辑输入样本的神经网络算法. 电子学报，1993，21(5)：54—62

[5] Huang S C, Huang Y F. Bounds on the Number of Hidden Neurons in Multilayer Perceptrons. IEEE Trans on Neural Networks, 1991, 2(1): 47－55

[6] 张军英，保 铮. 关于二进前向网络隐节点数的实验研究. 1997年中国神经计算科学大会论文集，人民邮电出版社，1997

[7] Kim J H, Park S K. The Geometrical Learning of Binary Neural Networks. IEEE Trans on Neural Networks, 1995, 6(1): 237－247

[8] Gray D L, Michel A N. A Training Algorithm for Binary Feedforward Neural Networks. IEEE Trans on Neural Networks, 1992, 3(2): 176－194

RESEARCH ON THE MINIMAL UPPER BOUND OF THE NUMBER OF HIDDEN NODES IN BINARY NEURAL NETWORKS

Lu Yang, Han Jianghong, Gao Jun

(*Research Institute of Microcomputer*, *Hefei University of Technology*, *Hefei* 230009)

ABSTRACT

Through discussing the relationship between the connectivity and linear separability in Boolean space, we get the following conclusion: to realize any Boolean function using a binary neural network with only one hidden layer, 2^{n-1} hidden nodes are necessary. And we construct a special Boolean function which has to use 2^{n-1} hidden nodes.

Key Words Binary Neural Networks, Connected Set, Linear Separability, Minimal Upper Bound