第 20 卷 第 3 期
1994 年 5 月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 20, No. 3
May, 1994

短文

神经网络的知识获取与行为解释

钱大群
(上海交通大学自动控制系　上海　200030)

孙振飞
(上海工业大学计算机系　上海　200072)

摘　　要

神经网络的缺点之一在于它无法明显地表达其行为的含义. 针对神经网络的不同结构，通过给出节点取值不同的约束条件，描述了如何从神经网络提取知识，并对神经网络的行为进行解释. 举例说明了从神经网络提取产生式规划的策略.

关键词: 神经网络，知识获取，行为解释，产生式规则.

1、引言

与人工智能技术相比，由于神经网络有一些明显的缺陷，即神经网络无法如人工智能系统那样，以明显的方式表达它所包含的知识，以及向使用者解释怎样导出它的运算结果，从而不能帮助使用者提高认识能力和证实运算结果的正确性. 所以，研究神经网络的知识获取与行为解释方式，对推广神经网络的应用具有一定的意义.

目前，已有文章提出了一些可从神经网络中生成产生式规则的策略[1—3]. 但这些策略只适用于取值为—1，0，1 的前馈网络，且这些网络的节点大都代表一个命题. 本文拟从不同的角度描述神经网络的知识获取与行为解释的问题.

2、问题的一般描述

首先，设所讨论的神经网络是前馈网络，网络由 l 层的节点构成. N_i 是位于第 i 层的节点集合，$N_i = \{n_{ij} | j = 1, \cdots, m_i\}$，$i = 1, \cdots, l$，$n_{ij}$ 是 i 层上的第 j 个节点. W 是权集合，$W = \{W_{ijk}\}$，W_{ijk} 是从节点 n_{ki} 到节点 n_{k+1i} 的弧上的权. 在 t 时刻，节点 n_{ij} 的状态值可由 $s_{ij}(t)$ 表示，N_i 的状态向量由 $\boldsymbol{s}_i(t) = [s_{i1}(t), \cdots, s_{im_i}(t)]$ 表示. 在这前馈网络中，第一层(输入层)与第 l 层(输出层)上的节点代表一个命题或一个变量，$\boldsymbol{s}_1$ 代表第一层的命题的可信度或变量的已知值 (事实)，$\boldsymbol{s}_l$ 代表神经网络所导出的命题的可信度或变量的值，W_{ijk} 可视为节点 n_{ki} 蕴涵节点 n_{k+1j} 的可信度或加权与节点的权. 设

本文于 1992 年 4 月 24 日收到.

$s_i(t)=x_i$，$s_{ij}(t)=x_{ij}$，$i=1,\cdots,l$，$j=1,\cdots,m_i$，则可建立如下的产生式规则：

$$s_1(t)=x_1 \Rightarrow s_2(t)=x_2 \Rightarrow \cdots \Rightarrow s_l(t)=x_l.$$

就 $s_i(t)=x_i \Rightarrow s_{i+1}(t)=x_{i+1}$ 而言，可有下面更详细的产生式规则：

$$(s_{i1}(t)=x_{i1})(w_{j1i})\wedge(s_{i2}(t)=x_{i2})(w_{j2i})\wedge\cdots\wedge(s_{im_i}(t)=x_{im_i})(w_{jm_ii})$$
$$\Rightarrow (s_{i+1j}(t)=x_{i+1j}).$$

显然，采用这些产生式规则，可说明神经网络如何从当前的输入 $s_1(t)$ 导出神经网络的输出 $s_l(t)$。因此，神经网络的知识获取与行为解释的问题可变为如何从神经网络的行为中提取上述产生式规则的问题。求解产生式规则的基本思想是怎样选择节点，由这些节点导出的产生式规则和神经网络的输入，可以维持神经网络的当前的输出值。

根据节点的选取方式，上述问题可分为七类，即

1）当 $s_l(t)=x_l$，从 N_1 中寻找最少元素的子集 N_1'，其值 $s_1'(t)=x_1'$ 与 x_1 中 N_1' 的值相同，则有

$$s_1'(t)=x_1' \Rightarrow s_2'(t)=y_2 \Rightarrow \cdots \Rightarrow s_{l-1}'(t)=y_{l-1} \Rightarrow s_l(t)=x_l.$$

在此，$N_i'\subseteq N_i$，其值 $s_i'(t)=y_i$ 可与 x_i 中 N_i' 的状态值 x_i' 不同，$i=2,3,\cdots,l-1$。

2）当 $s_l(t)=x_l$，从 N_1 中寻找最少元素的子集 N_1'，其值 $s_1'(t)=x_1'$ 与 x_1 中 N_1' 的值相同，即使 N_1-N_1' 的状态值发生变化，仍有

$$s_1'(t)=x_1' \Rightarrow s_2'(t)=y_2 \Rightarrow \cdots \Rightarrow s_{l-1}'(t)=y_{l-1} \Rightarrow s_l(t)=x_l.$$

可见，1)与 2)的区别在于被删去的节点集合 (N_1-N_1') 的状态值是否可变化。

3）当 $s_l(t)=x_l$，从 N_i 中寻找最少元素的子集 N_i'，其值 $s_i'(t)=y_i$ 可与 x_i' 不同，$i=2,\cdots,l-1$，则有

$$s_1(t)=x_1 \Rightarrow s_2'(t)=y_2 \Rightarrow \cdots \Rightarrow s_{l-1}'(t)=y_{l-1} \Rightarrow s_l(t)=x_l.$$

即从 $s_1(t)=x_1$ 出发，寻找可推出 $s_l(t)=x_l$ 的最少数目的产生式规则（$N_i'(i=2,\cdots,l-1)$ 中的元素个数）。

4）当 $s_l(t)=x_l$，从 N_i 中寻找最少元素的子集 $N_i'(i=1,\cdots,l-1)$，有

$$s_1'(t)=x_1' \Rightarrow s_2'(t)=y_2 \Rightarrow \cdots \Rightarrow s_{l-1}'(t)=y_{l-1} \Rightarrow s_l(t)=x_l,$$

这是一个全局优化的问题。

5）当 $s_l(t)=x_l$，选择节点，使每一产生式规则具有最少的前提，并保证由 $s_1'(t)=x_1'$ 可导出 $s_l(t)=x_l$。

6）当 $s_l(t)=x_l$，从 N_1 中寻找最少元素的子集 N_1'，其值 $s_1'(t)=x_1'$，有

$$s_1'(t)=x_1' \Rightarrow s_2(t)=x_2 \Rightarrow \cdots \Rightarrow s_{l-1}(t)=x_{l-1} \Rightarrow s_l(t)=x_l.$$

这里 $s_i(t)$ 的值保持 $x_i(i=2,\cdots,l-1)$。

7）求出 $s_i(t)$ 的 m 个区域 $X_i^{(k)}$，$k=1,\cdots,m$，$i=1\cdots,l$，有

$$s_1(t)\subseteq X_1^{(k)} \Rightarrow s_2(t)\subseteq X_2^{(k)} \Rightarrow \cdots \Rightarrow s_l(t)\subseteq X_l^{(k)}\ k=1,\cdots,m,$$

这是一个求解联立方程的问题。求得的产生式规则可描述神经网络的一般行为。

在上述 1)—4）类中，由 x_i' 代替 $y_i(i=2,\cdots,l-1)$，以及将 5）与 1),2）相结合，即可产生新的问题描述。神经网络的不同之处也使上述问题具有如下特点。

1）当神经网络的传递函数是线性或 max-min 函数时，则它较适宜于表达可信度在神经网络上的计算与传播；当它是离散函数时，上面所描述的问题就相当于一个约束满足

问题;当传递函数是连续函数时,可考虑将此连续函数离散化,然后删选节点,构造产生式规则,但这将导致一定的误差;当传递函数是随机函数时, 则生成具有如下形式的产生式规则:

$$\boldsymbol{s}_k(t) = \boldsymbol{x}_k \Rightarrow \boldsymbol{s}_{k+1}(t) = \boldsymbol{x}_{k+1} \text{ with probability } p.$$

2) 在前馈网络中,节点通常代表一个命题或变量. 而回归网络常用网络的状态来表达信息.设回归网络的初始状态为 $s(t_0)$, 稳定状态为 $s(t_n)$, 可认为有如下的逻辑关系:

$$s(t_0) \Rightarrow s(t_n).$$

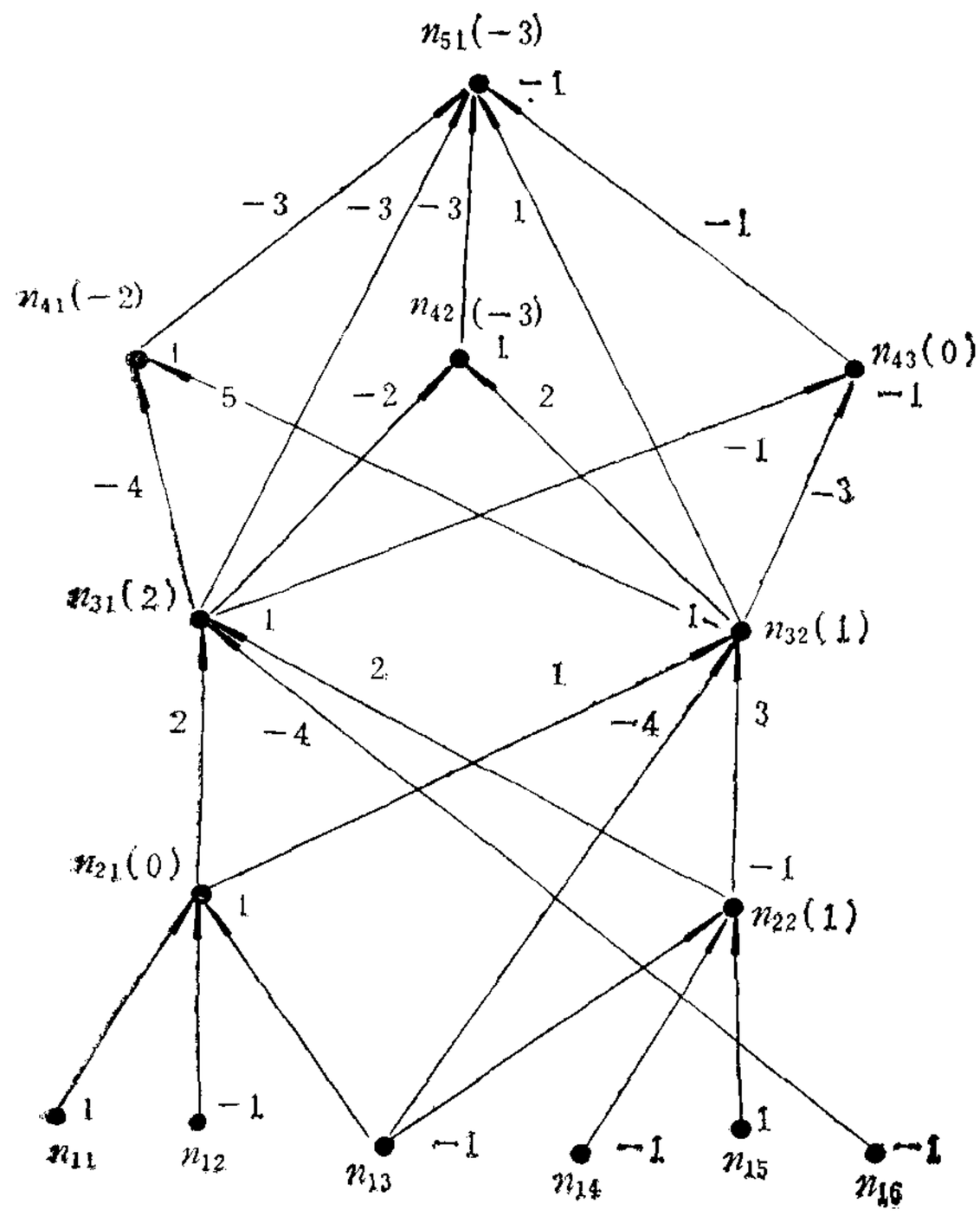
根据产生式规则生成的不同特点,生成过程可分为五类. 1) 前向搜索过程: 从输入节点出发向输出节点搜索,导出所有产生式规则; 2) 反向搜索过程: 从输出节点出发向输入节点反向搜索,导出针对该输出节点的产生式规则; 3) 启发式搜索过程: 引入启发式剪枝技术,缩小搜索空间; 4) 串行搜索过程: 以串行搜索方式生成产生式规则; 5) 并行搜索过程: 以并行搜索方式生成产生式规则. 详见文献[4].

3、例子

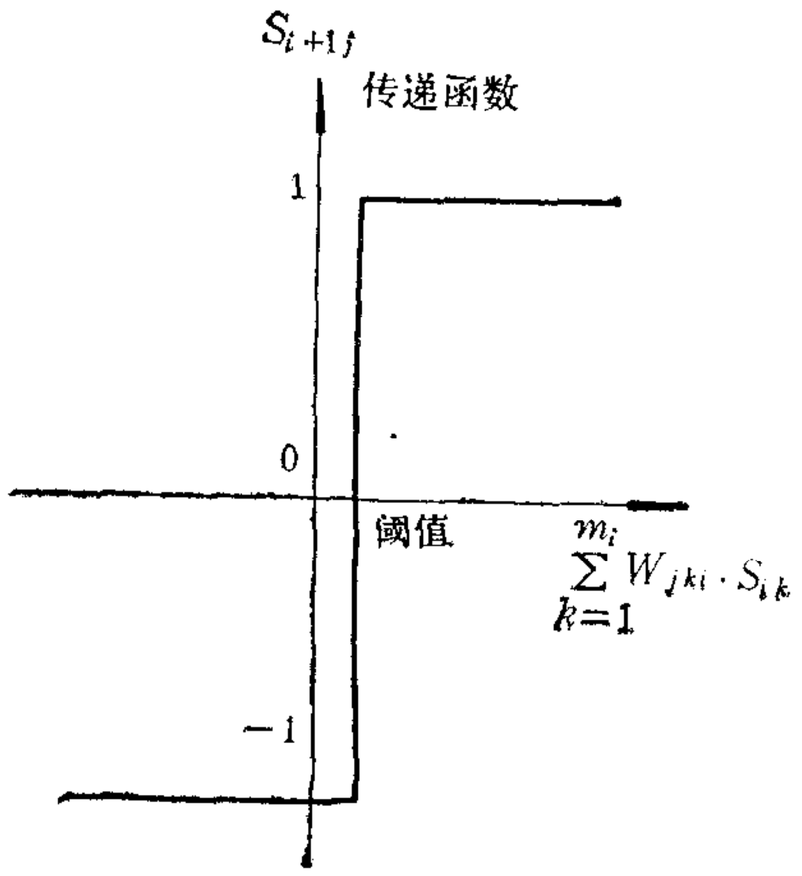
一个寻找最少元素集合 N_1' 的简单方法, 是基于集合运算的反向搜索策略. 设 $\boldsymbol{L}(i,j) = \{L^{(k)}(i,j)\}$ 是所有可维持节点 n_{ij} 的状态值 $s_{ij}(t) = x_{ij}$ 的 N_1 的子集类, 当要求维持 $s_{ij}(t) = x_{ij}$ 时, $(i,j) = (i_l, j_l)$, $l = 1, \cdots, m$, 可选择满足下述条件的集合N_1':

$$N_1' = \bigcup_{(i,j)} L^{(k_{ij})}(i,j), \quad \left|\bigcup_{(i,j)} L^{(k_{ij})}(i,j)\right| = \min_k \left|\bigcup_{(i,j)} L^{(k)}(i,j)\right|,$$

$L^{(k_{ij})}(i,j) \in L(i,j)$, $(i,j) = (i_l, j_l)$ $l = 1, \cdots, m$. 其中$|L|$表示集合 L 中元素的



(a)



(b)

图 1

个数。

在图 1 所示的神经网络中，节点 $n_{11},n_{12},n_{13},n_{14},n_{15},n_{16},n_{21},n_{22},n_{31},n_{32}$ 和 n_{51} 分别代表疾病症状或治疗措施，在每一节点旁括号内的数字是该节点传递函数的阈值，节点旁的值是它当前的状态值，弧旁的值是权。设允许被删除的节点状态值发生变化，用上述策略求出可维持 n_{51} 的状态值的集合 N_1'。

首先，n_{41},n_{42} 与 n_{31} 可维护 n_{51} 的值。如果选择 n_{31}，则须采用 n_{11}，n_{12}，n_{13} 与 n_{16} 来维护 n_{51} 的值，因而需要四个节点；如果选择 n_{31} 与 n_{41}（或 n_{42}），则 n_{31} 需要 n_{11},n_{12} 与 n_{16} 的支撑，n_{41}（或 n_{42}）需要 n_{11},n_{12} 与 n_{13} 的支撑，因而共需要四个节点；如果选择 n_{31},n_{41} 与 n_{42}，则也需要同样的四个节点的支撑。但是，在这三种选择中，第一种选择生成最少数目的产生式规则，即 $n_{11}(1)(2)\wedge n_{12}(-1)(-2)\wedge n_{13}(-1)(3)\wedge n_{16}(-1)(-4)\Rightarrow n_{31}(1)$，$n_{31}(1)(-3)\Rightarrow n_{51}(1)$。

参　考　文　献

[1] Bochereau L, Bourgine P. Rule Extraction and Validity Domain on a Multilayer Neural Network. Proc. IEEE and INNS Int. Joint Conf. on Neural Networks, San Diego, 1990: 97—100.

[2] Gallant S I. Connectionist Expert Systems. *Communication of ACM*, 1988, **31**: 152—169.

[3] Saitoh K, Nakano R. Medical Diagnostic Expert System Based on PDP Model. *Proc. IEEE Conf. on Neural Networks*, San Diego, 1988, **1**: 255—262.

[4] Qian D Q, Scaruffi P, Russi D. Generation of Production Rules from Neural Networks by Forward Search Procedures. Proc. Int. Young Computer Scientists Conf., Beijing, 1991: 413—418.

KNOWLEDGE ACQUISITION AND BEHAVIORAL EXPLANATION ON NEURAL NETWORKS

QIAN DAQUN

(*Dept. of Automatic Control, Shanghai Jiao Tong University, Shanghai* 200030)

SUN ZHENFEI

(*Dept. of Computer Engineering, Shanghai University of Technology, Shanghai* 200072)

ABSTRACT

One of the drawbacks of neural nets is that they can not explicitly express the meaning of their behaviors. This paper describes some problems on knowledge acquisition and behavioral explanation for neural nets with different structures, the constraints conditions are the numbers of nodes, and suggest a strategy of generating production rules for neural nets as an example.

Key words: neural net; knowledge acquisition; behavioral explanation; production rule.