

第五章 反馈神经网络



5 反馈神经网络

续

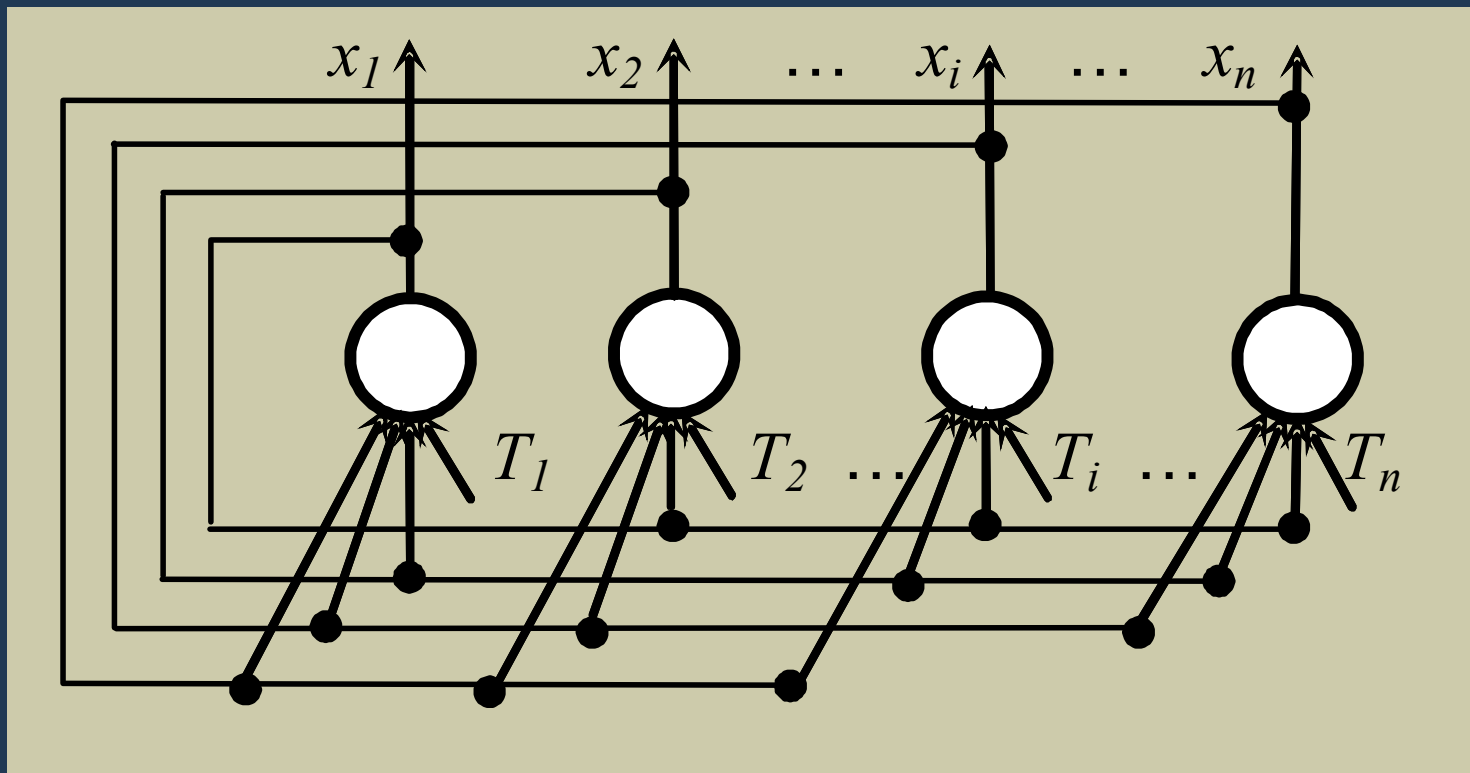
根据神经网络运行过程中的信息流向，可分为前馈式和反馈式两种基本类型。前馈网络的输出仅由当前输入和权重矩阵决定，而与网络先前的输出状态无关。

美国加州理工学院物理学家 J.J.Hopfield 教授于 1982 年提出一种单层反馈神经网络，后来人们将这种反馈网络称作 Hopfield 网。

Hopfield 网络分为离散型和连续型两种网络模型，分别记作 DHNN (Discrete Hopfield Neural Network) 和 CHNN (Continues Hopfield Neural Network)，本章重点讨论前一种类型。

5.1 离散型 Hopfield 神经网络

5.1.1 网络的结构与工作方式



离散型反馈网络的拓扑结构



(1) 网络的状态

DHNN 网中的每个神经元都有相同的功能，其输出称为状态，用 x_j 表示。

所有神经元状态的集合就构成反馈网络的状态

$$X=[x_1, x_2, \dots, x_n]^T$$

反馈网络的输入就是网络的状态初始值，表示为

$$X(0)=[x_1(0), x_2(0), \dots, x_n(0)]^T$$

反馈网络在外界输入激发下，从初始状态进入动态演变过程，变化规律为

$$x_j = f(\text{net}_j) \quad j=1, 2, \dots, n$$



DHNN 网的转移函数常采用符号函数

$$x_j = \text{sgn}(net_j) = \begin{cases} 1 & net_j \geq 0 \\ -1 & net_j < 0 \end{cases} \quad j=1,2,\dots,n \quad (5.1)$$

式中净输入为

$$net_j = \sum_{i=1}^n (w_{ij}x_i - T_j) \quad j=1,2,\dots,n \quad (5.2)$$

对于 DHNN 网，一般有 $w_{ii}=0$ ， $w_{ij}=w_{ji}$ 。

反馈网络稳定时每个神经元的状态都不再改变，此时的稳定状态就是网络的输出，表示为

$$\lim_{t \rightarrow \infty} X(t)$$



(2) 网络的异步工作方式

网络运行时每次只有一个神经元进行状态的调整计算，其它神经元的状态均保持不变，即

$$x_j(t+1) = \begin{cases} \text{sgn}[net_j(t)] & j = i \\ x_j(t) & j \neq i \end{cases} \quad (5.3)$$

(3) 网络的同步工作方式

网络的同步工作方式是一种并行方式，所有神经元同时调整状态，即

$$x_j(t+1) = \text{sgn}[net_j(t)] \quad j=1,2,\dots,n \quad (5.4)$$

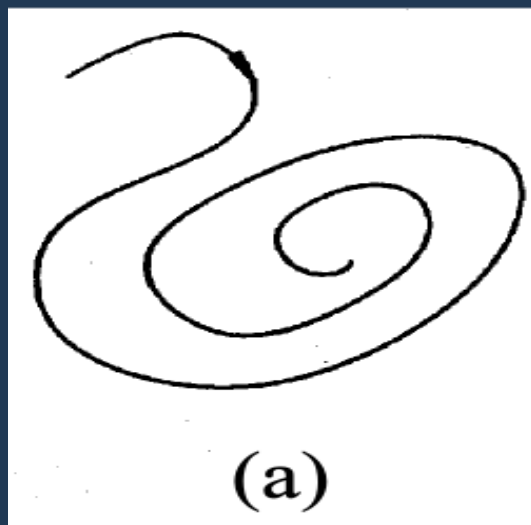


5.1.2 网络的稳定性与吸引子

5.1.2.1 网络的稳定性

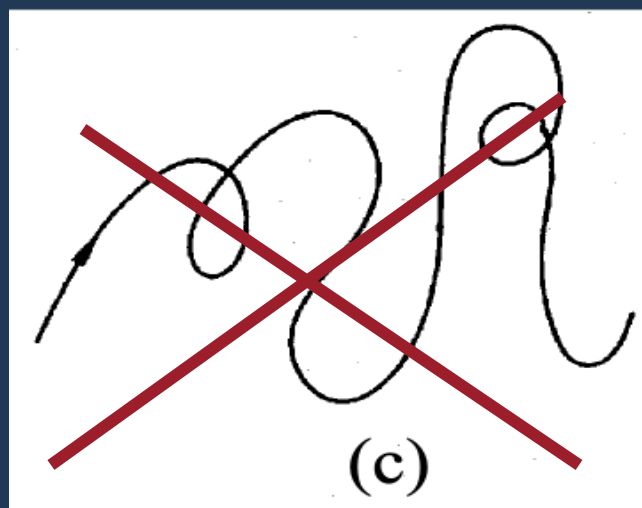
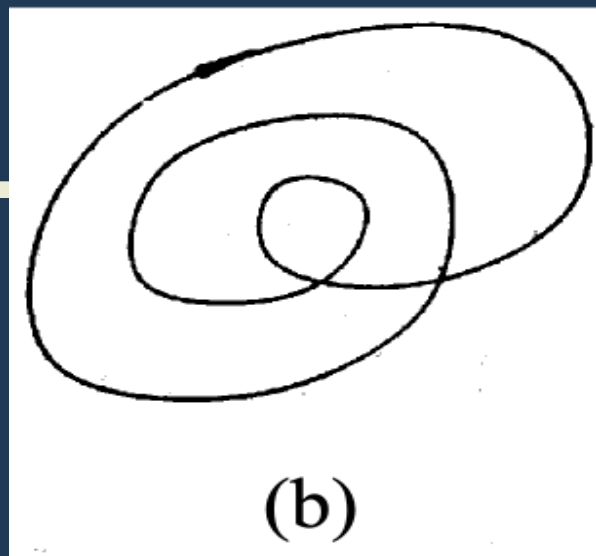
DHNN 网实质上是一个离散的非线性动力学系统。网络从初态 $X(0)$ 开始，若能经有限次递归后，其状态不再发生变化，即 $X(t+1) = X(t)$ ，则称该网络是稳定的。

如果网络是稳定的，它可以从不任一初态收敛到一个稳态：



若网络是不稳定的，由于DHNN网每个节点的状态只有1和-1两种情况，网络不可能出现无限发散的情况，而只可能出现限幅的自持振荡，这种网络称为有限环网络。

如果网络状态的轨迹在某个确定的范围内变迁，但既不重复也不停止，状态变化为无穷多个，轨迹也不发散到无穷远，这种现象称为混沌。



5.1.2.2 吸引子与能量函数

网络达到稳定时的状态 X ，称为网络的 **吸引子**。

如果把吸引子视为问题的解，从初态朝吸引子演变的过程便是求解计算的过程。

若把需记忆的样本信息存储于网络不同的吸引子，当输入含有部分记忆信息的样本时，网络的演变过程便是从部分信息寻找全部信息，即联想回忆的过程。

定义 5.1 若网络的状态 X 满足

$$X=f(WX-T)$$

则称 X 为网络的吸引子。



5.1.2.2 吸引子与能量函数

定理 5.1 对于 DHNN 网，若按异步方式调整网络状态，且连接权矩阵 W 为对称阵，则对于任意初态，网络都最终收敛到一个吸引子。

定理 5.1 证明：

定义网络的能量函数为：

$$E(t) = -\frac{1}{2} X^T(t) W X(t) + X^T(t) T \quad (5.5)$$

令网络的能量改变量为 ΔE ，状态改变量为 ΔX ，有

$$\Delta E(t) = E(t+1) - E(t) \quad (5.6)$$

$$\Delta X(t) = X(t+1) - X(t) \quad (5.7)$$



将式 (5.4)、(5.6) 代入 (5.5)，则网络能量可进一步展开为

$$\begin{aligned}\Delta E(t) &= E(t+1) - E(t) \\ &= -\frac{1}{2}[X(t) + \Delta X(t)]^T W [X(t) + \Delta X(t)] + [X(t) + \Delta X(t)]^T T - [-\frac{1}{2}X^T(t)WX(t) + X^T(t)T] \\ &= -\Delta X^T(t)WX(t) - \frac{1}{2}\Delta X^T(t)W\Delta X(t) + \Delta X^T(t)T \\ &= -\Delta X^T(t)[WX(t) - T] - \frac{1}{2}\Delta X^T(t)W\Delta X(t)\end{aligned}\tag{5.8}$$

将 $\Delta X(t) = [0, \dots, 0, \Delta x_j(t), 0, \dots, 0]^T$ 代入上式，并考虑到 W 为对称矩阵，有

$$\Delta E(t) = -\Delta x_j(t) \left[\sum_{i=1}^n (w_{ij} x_i - T_j) \right] - \frac{1}{2} \Delta x_j^2(t) w_{jj}$$



$$\Delta E(t) = -\Delta x_j(t) \text{net}_j(t) \quad (5.9)$$

上式中可能出现的情况：

情况 a : $x_j(t)=-1, x_j(t+1)=1$ ，由式 (5.7) 得 $\Delta x_j(t)=2$ ，由式 (5.1) 知， $\text{net}_j(t) \geq 0$ ，代入式 (5.9)，得 $\Delta E(t) \leq 0$ 。

情况 b : $x_j(t)=1, x_j(t+1)=-1$ ，所以 $\Delta x_j(t)=-2$ ，由式 (5.1) 知， $\text{net}_j(t) < 0$ ，代入式 (5.9)，得 $\Delta E(t) < 0$ 。

情况 c : $x_j(t)=x_j(t+1)$ ，所以 $\Delta x_j(t)=0$ ，代入式 (5.9)，从而有 $\Delta E(t)=0$ 。

由此可知在任何情况下均有 $\Delta E(t) \leq 0$ 。



由于网络中各节点的状态只能取 1 或 -1，能量函数 $\mathcal{E}(t)$ 作为网络状态的函数是有下界的，因此网络能量函数最终将收敛于一个常数，此时 $\Delta\mathcal{E}(t)=0$ 。综上所述，当网络工作方式和权矩阵均满足定理 5.1 的条件时，网络最终将收敛到一个吸引子。

综上所述，当网络工作方式和权矩阵均满足定理 5.1 的条件时，网络最终将收敛到一个吸引子。

定理 5.2 对于 DHNN 网，若按同步方式调整状态，且连接权矩阵 W 为非负定对称阵，则对于任意初态，网络都最终收敛到一个吸引子。



证明：由式 (5.8) 得

$$\Delta E(t) = E(t+1) - E(t)$$

$$= -\Delta \mathbf{X}^T(t) [W\mathbf{X}(t) - T] - \frac{1}{2} \Delta \mathbf{X}^T(t) W \Delta \mathbf{X}(t)$$

$$= -\sum_{j=1}^n \Delta x_j(t) \text{net}_j(t) - \frac{1}{2} \Delta \mathbf{X}^T(t) W \Delta \mathbf{X}(t)$$

$$= -\Delta \mathbf{X}^T(t) \text{net}(t) - \frac{1}{2} \Delta \mathbf{X}^T(t) W \Delta \mathbf{X}(t)$$

前已证明，对于任何神经元 j ，有 $-\Delta x_j(t) \text{net}_j(t) \leq 0$

因此上式第一项不大于 0，只要 W 为非负定阵，第二项也不大于 0，于是有 $\Delta E(t) \leq 0$ ，也就是说 $E(t)$ 最终将收敛到一个常数值，对应的稳定状态是网络的一个吸引子。



5.1.2.2 吸引子与能量函数

以上分析表明，在网络从初态向稳态演变的过程中，网络的能量始终向减小的方向演变，当能量最终稳定于一个常数时，该常数对应于**网络能量的极小状态**，称该极小状态为**网络的能量井**，**能量井对应于网络的吸引子**。



5.1.2.3 吸引子的性质

性质 1 : 若 X 是网络的一个吸引子, 且阈值 $T=0$, 任意一个节点的净输入都不为零, 则 $-X$ 也一定是该网络的吸引子。

证明 : $\because X$ 是吸引子, 即 $X=f(WX)$, 从而有

$$f[W(-X)]=f[-WX]=-f[WX]=-X$$

$\therefore -X$ 也是该网络的吸引子。



性质 2 : 若 X^a 是网络的一个吸引子, 则与 X^a 的海明距离 $dH(X^a, X^b)=1$ 的 X^b 一定不是吸引子。

证明: 不妨设 $x_1^a \neq x_1^b$, $x_j^a = x_j^b$, $j=2,3,\dots,n$ 。

$\because w_{11}=0$, 由吸引子定义, 有

$$x_1^a = f\left(\sum_{i=2}^n w_{1i} x_i^a - T_1\right) = f\left(\sum_{i=2}^n w_{1i} x_i^b - T_1\right)$$

由假设条件知, $x_1^a \neq x_1^b$, 故

$$x_1^b \neq f\left(\sum_{i=2}^n w_{1i} x_i^b - T_1\right)$$

$\therefore X^b$ 不是吸引子



5.1.2.4 吸引子的吸引域

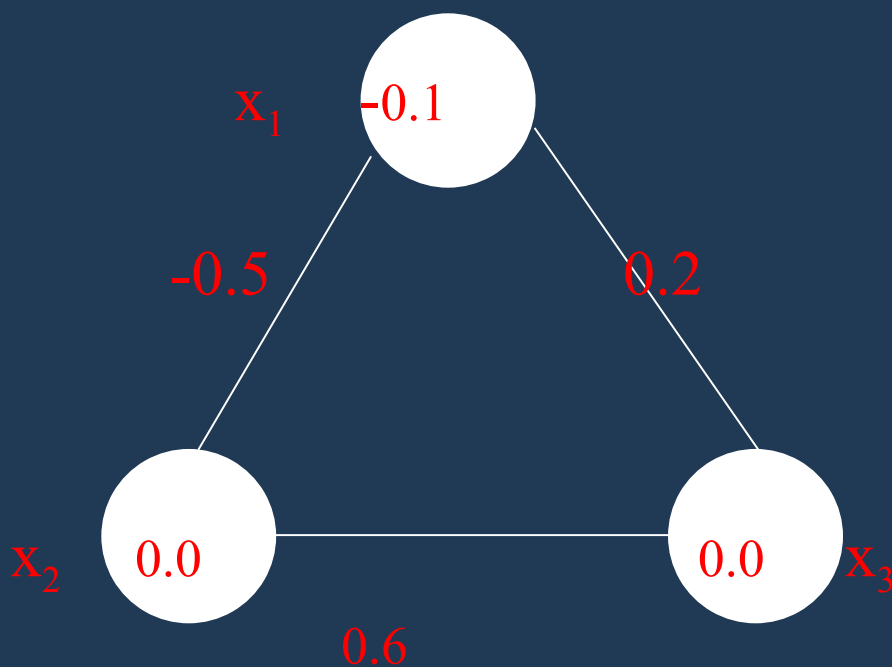
能使网络稳定在同一吸引子的所有初态的集合，称为该吸引子的**吸引域**。

定义 5.2 若 X^a 是吸引子，对于异步方式，若存在一个调整次序，使网络可以从状态 X 演变到 X^a ，则称 X 弱吸引到 X^a ；若对于任意调整次序，网络都可以从状态 X 演变到 X^a ，则称 X 强吸引到 X^a 。

定义 5.3 若对某些 X ，有 X 弱吸引到吸引子 X^a ，则称这些 X 的集合为 X^a 的**弱吸引域**；若对某些 X ，有 X 强吸引到吸引子 X^a ，则称这些 X 的集合为 X^a 的**强吸引域**。



例 5.1 设有 3 节点 DHNN 网，用无向图表示如下，权值与阈值均已标在图中，试计算网络演变过程的状态。



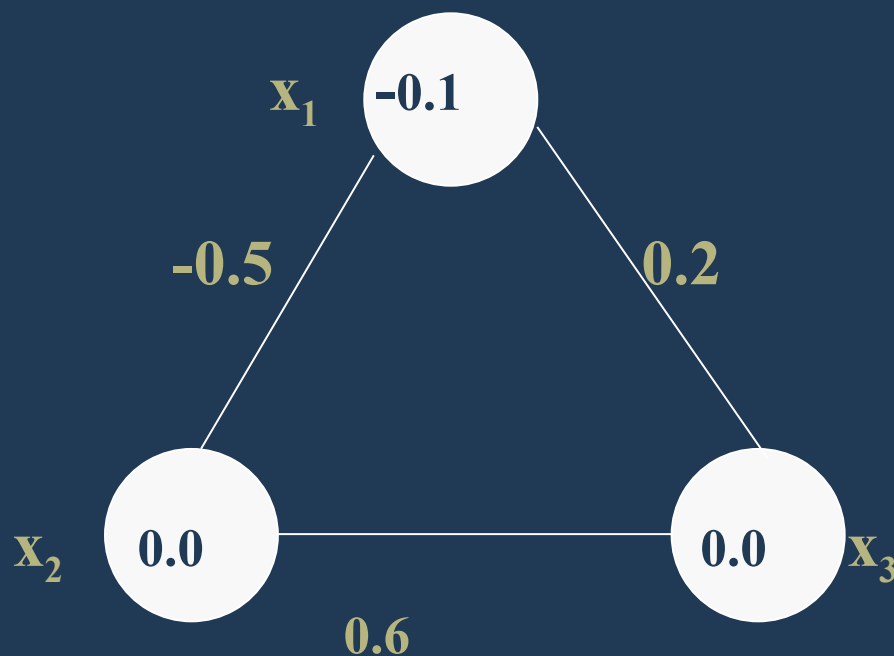
解：设各节点状态取值为 1 或 0 ， 3 节点 DHNN 网络应有 $2^3=8$ 种状态。不妨将 $\mathcal{X}=(x_1, x_2, x_3)^T=(0,0,0)^T$ 作为网络初态，按 $1 \rightarrow 2 \rightarrow 3$ 的次序更新状态。

第 1 步：更新 x_1 ，

$$x_1 = \text{sgn}[(-0.5) \times 0 + 0.2 \times 0 - (-0.1)]$$

$$= \text{sgn}(0.1) = 1$$

其它节点状态不变，网络状态由 $(0,0,0)^T$ 变成 $(1,0,0)^T$ 。如果先更新 x_2 或 x_3 ，网络状态将仍为 $(0,0,0)^T$ ，因此初态保持不变的概率为 $2/3$ ，而变为 $(1,0,0)^T$ 的概率为 $1/3$ 。



第2步：此时网络状态为 $(1,0,0)^T$ ，更新 x_2 后，得

$$x_2 = \text{sgn}[(-0.5) \times 1 + 0.6 \times 0 - 0] = \text{sgn}(-0.5) = 0$$

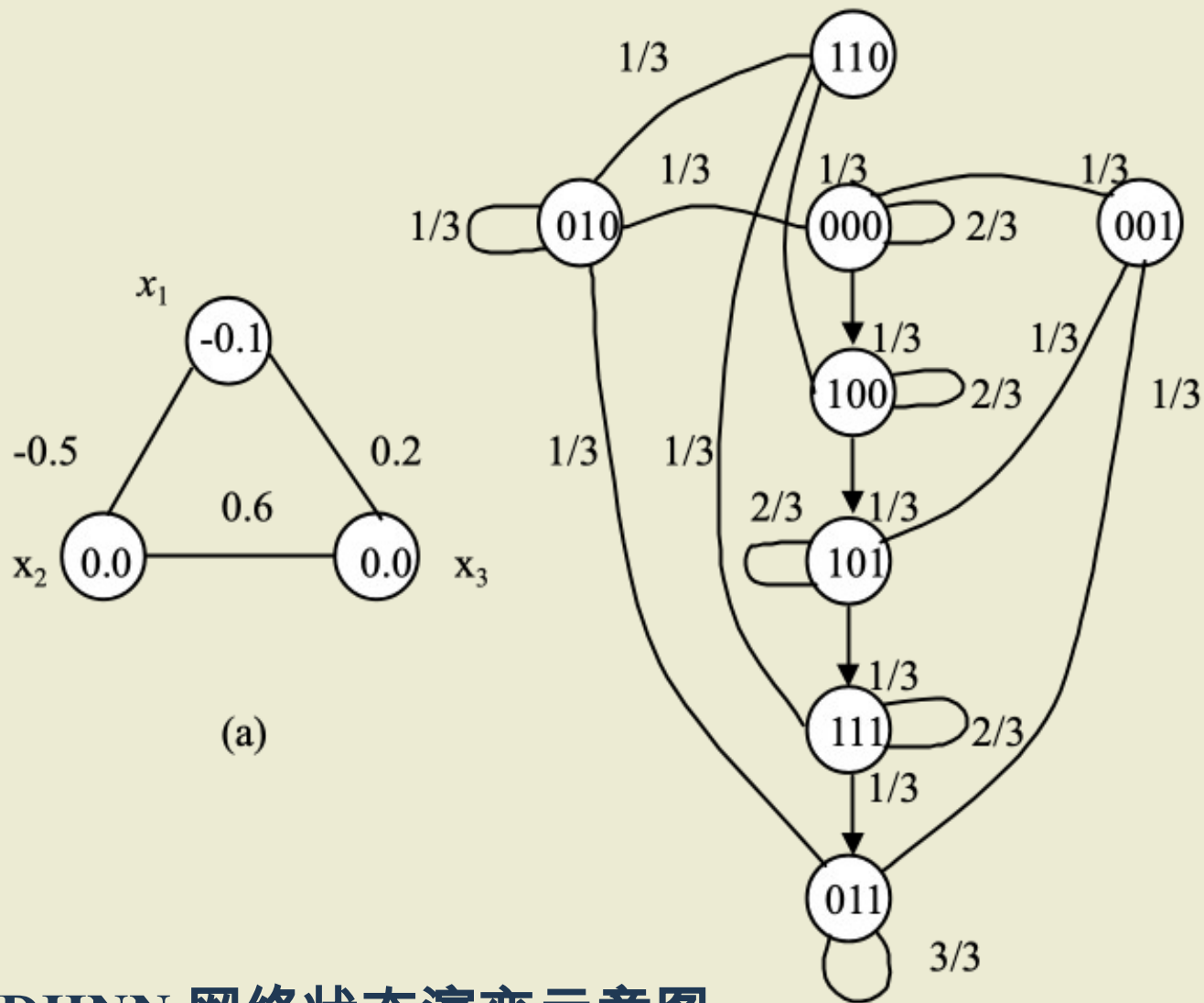
其它节点状态不变，网络状态仍为 $(1,0,0)^T$ 。如果本步先更新 x_1 或 x_3 ，网络相应状态将为 $(1,0,0)^T$ 和 $(1,0,1)^T$ ，因此本状态保持不变的概率为 $2/3$ ，而变为 $(1,0,1)^T$ 的概率为 $1/3$ 。

第3步：此时网络状态为 $(1,0,0)^T$ ，更新 x_3 得

$$x_3 = \text{sgn}[0.2 \times 1 + 0.6 \times 0 - 0] = \text{sgn}(0.2) = 1$$

同理可算出其它状态之间的演变历程和状态转移概率。





DHNN 网络状态演变示意图



5.1.3 网络的权值设计

为了使所设计的权值满足要求，权值矩阵应符合以下要求：

- (1) 为保证异步方式工作时网络收敛， W 应为对称阵；
- (2) 为保证同步方式工作时网络收敛， W 应为非负定对称阵；
- (3) 保证给定样本是网络的吸引子，并且要有一定的吸引域。

5.1.3.2 外积和法

设给定 P 个模式样本 X^p ， $p=1,2,\dots,P$ ， $x \in \{-1,1\}^n$ ，并设样本两两正交，且 $n > P$ ，则权值矩阵为记忆样本的外积和

$$W = \sum_{p=1}^P X^p (X^p)^T \quad (5.16)$$



若取 $w_{jj}=0$ ，上式应写为

$$W = \sum_{p=1}^P [X^p (X^p)^T - I] \quad (5.17)$$

式中 I 为单位矩阵。上式写成分量元素形式，有

$$w_{ij} = \begin{cases} \sum_{p=1}^P x_i^p x_j^p & i \neq j \\ 0 & i = j \end{cases} \quad (5.18)$$

下面检验所给样本能否称为吸引子。

因为 P 个样本 X^p ， $p=1,2,\dots,P$ ， $x \in \{-1,1\}^n$ 是两两正交的，有

$$(X^p)^T X^k = \begin{cases} 0 & p \neq k \\ n & p = k \end{cases}$$



$$\begin{aligned}
WX^k &= \sum_{p=1}^P [X^p (X^p)^T - I] X^k \\
&= \sum_{p=1}^P [X^p (X^p)^T X^k - X^k] \\
&= X^k (X^k)^T X^k - PX^k \\
&= nX^k - PX^k = (n - P)X^k
\end{aligned}$$

因为 $n > P$, 所以有

$$f(WX^p) = f[(n - P)X^p] = \text{sgn}[(n - P)X^p] = X^p$$

可见给定样本 X^p , $p=1,2,\dots,P$ 是吸引子。

