ДИНАМИЧЕСКИЕ СЕТИ С ОБРАТНЫМИ СВЯЗЯМИ. НЕЙРОННЫЕ СЕТИ ХОПФИЛДА.*

М.Н. Рычагов,
доктор физико-математических наук, доцент

Московский институт электронной техники (МИЭТ)
Факультет электронных и компьютерных технологий
Зеленоград 103498 Москва
Тел: 532-99-63; Факс: 530-22-33
Электронная почта: mrychagov@miee.ru

1 Нейронные сети Хопфилда и оптимизация

Нейронная сеть Хопфилда (рис. 1) представляет собой слой адаптивных сумматоров с обратными связями, выходные сигналы которых, подвергаясь нелинейной обработке по заданному закону, поступают с некоторой временной задержкой на входы нейронов, в результате чего выходной сигнал нейронной сети формируется лишь после того, как сеть достигнет динамического равновесия. Поведение нейронной сети моделирует, таким образом, некоторый стохастический процесс, конечное состояние которого определяется входным вектором нейросети, являющимся по сути вектором внешних смещений.

Пусть состояние каждого i-го нейрона определяется его выходным сигналом F_i. В архитектуре нейросети, реализующей бинарные операции, функция F_i может принимать значения $F_i^0 = 0$ или $F_i^1 = 1$. Как видно из рис. 1, входной сигнал каждого нейрона H_i представляет собой суперпозицию двух сигналов: внешнего сигнала I_i и сигнала обратной связи, в виде суммы выходных сигналов других нейронов. Тогда

$$H_i = \sum_{j=1}^{J} \omega_{ij} F_j + I_i, \tag{1}$$

где ω_{ij} – вес синаптической связи, соединяющий j-й нейрон с i-м нейроном.

Каждый нейрон изменяет свое состояние в зависимости от заданного уровня активации S_i, так что

$$F_i = \begin{cases} \searrow F_i^0, & H_i < S_i \\ \nearrow F_i^1, & H_i > S_i. \end{cases} \tag{2}$$

*Материал подготовлен специально для участников ознакомительного семинара Консультационного Центра MATLAB компании SoftLine, 29 июня 2001 г.

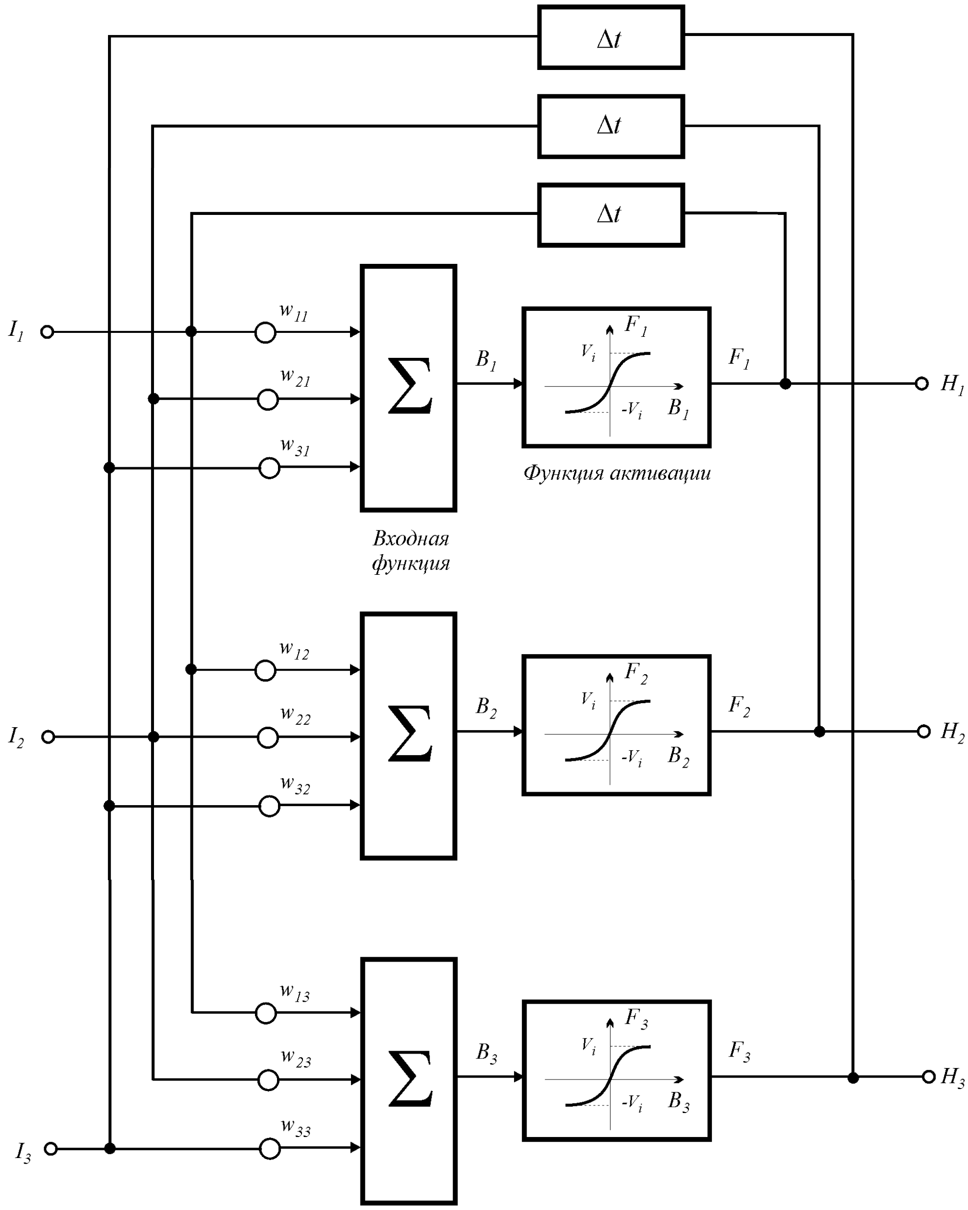

Рис. 1: Архитектура рекуррентной сети Хопфилда с тремя нейронами.

Если предположить, что весовые коэффициенты синаптических связей ω_{ij} являются фиксированными для всех i и j, то система уравнений (1)-(2) определяет стохастический процесс, который достигает устойчивых положений равновесия в зависимости от внешних значений I_i. Иными словами, данная нейросистема с рекуррентными связями функционирует как «ассоциативная» память, поскольку ее устойчивые состояния таковы, что если система инициализирована вблизи одного из устойчивых состояний, то последующий динамический релаксационный процесс приводит ее именно в это состояние: любая произвольная точка на фазовой диаграмме хопфилдовской нейросистемы ассоциируется с одним из таких состояний. Доказано [1], [2], что сходимость гарантирована, если ее матрица весовых коэффициентов **W** является симметричной и все диагональные элементы равны нулю. Доказательство сходимости может быть получено из анализа «энергетической» функции нейросистемы, а именно, функции Ляпунова [3], которая для рассматриваемой нейронной сети с обратными связями имеет вид

$$\mathcal{E} = -\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\omega_{ij}F_iF_j - \sum_{i=1}^{I} I_iF_i + \sum_{i=1}^{I} S_iF_i \tag{3}$$

и представляет собой квадратичный функционал состояния нейронной сети. Изменение функции $\mathcal{E}$ вследствие изменения состояния i-го нейрона на ΔF_i выражается в виде

$$\Delta\mathcal{E} = -\left[\sum_{j=1}^{J}\omega_{ij}F_i + I_i + S_i\right]\Delta F_i. \tag{4}$$

Из уравнения (2) следует, что величина ΔF_i принимает положительные значения только в том случае, когда $\sum\limits_{j=1}^{J}\omega_{ij}F_i+I_i+S_i > 0$, и наоборот, принимает отрицательные значения, если $\sum\limits_{j=1}^{J}\omega_{ij}F_i+I_i+S_i < 0$. Следовательно, произвольное изменение состояния нейрона в архитектуре нейросети Хопфилда приводит к уменьшению энергетической функции всей системы.

Если вместо бинарного представления сигналов выбрать «спиновую» модель, т.е. предположить, что функция F_i может принимать значения $F_i^0 = -1$ или $F_i^1 = 1$, то уровень активации S_i можно положить равным нулю. Тогда система (2) перепишется в виде

$$F_i = \begin{cases} \searrow -1, & H_i < 0 \\ \nearrow +1, & H_i > 0, \end{cases} \tag{5}$$

Для энергетической функции, отвечающей данной системе, получим

$$\mathcal{E} = -\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\omega_{ij}F_iF_j - \sum_{i=1}^{I} I_iF_i. \tag{6}$$

Поведение нейронной сети Хопфилда можно анализировать, используя графовую модель либо решая задачу о собственных векторах и собственных значениях рассматриваемой системы.

С точки зрения графовой модели, нейронная сеть представляет собой направленный граф, вершины которого образуют нейроны с приложенными к ним внешними смещениями, а ребра образованы синаптическими связями с весовыми коэффициентами w_{ij}. Приведем простое доказательство сходимости динамического процесса сети Хопфилда. Предположим, что k нейронов сети Хопфилда разделены на две группы: первая группа включает нейроны с $F_i^1 = 1$, вторая группа - с $F_i^1 = -1$. Нейроны одной группы соединены с нейронами другой группы ребрами графа. Выберем произвольный нейрон и будем анализировать «влияние», которому он подвержен со стороны нейронов своей группы и нейронов, принадлежащих группе с иным состоянием. Это влияние проявляется в виде суммы весовых коэффициентов всех синаптических связей данного нейрона. Если воздействие извне оказывается более значительным, чем влияние нейронов собственной группы, то нейрон изменяет в соответствии с (5) свое состояние и переходит таким образом в другую группу. В противном случае нейрон остается в собственной группе. Эта процедура многократно повторяется и именно она описывает динамику рассматриваемой нейросети. Сеть должна в итоге достигнуть стабильного состояния, так как сумма весовых коэффициентов синаптических связей (весов ребер), соединяющих нейроны одной группы с нейронами другой группы может только уменьшаться. Поскольку количество состояний нейронной сети является ограниченным, в конце концов должно быть достигнуто такое состояние, при котором воздействие нейронов собственной группы не будет превышать воздействие нейронов другой группы. Данная процедура сводится в теории графов к задаче нахождения минимального разреза графа [4], [5].

Рассмотрим теперь сеть Хопфилда, нейроны которой обладают непрерывной монотонной активацией, т. е. $F_i^0 \le F_i \le F_i^1$. Покажем, что такая нейронная сеть обладает свойствами процессора, производящего минимизацию квадратичной целевой функции вида

$$\mathcal{E}(F) = \frac{1}{2}\sum_{j=1}^{m}(\Psi_j^m - \Psi_j^c)^2, \tag{7}$$

где величины Ψ_j^m, $j = 1,\ldots,m$, соответствуют измерительным данным. Значения Ψ_j^c, $j = 1,\ldots,m$, представляют собой данные идеальных измерений, связанных с «нейронно-сетевым решением» с помощью преобразования

$$A\mathbf{F} = \Psi, \tag{8}$$

где A - матрица размерности $m \times n$. Согласно ранее введенным обозначениям, вектор $\mathbf{F}$ соответствует состоянию n нейронов; тогда входной вектор $\mathbf{I}$ и матрица синаптических коэффициентов W определяются соотношениями

$$\begin{aligned} \mathbf{I} &= A^T\Psi^m, \\ W &= -A^T A. \end{aligned} \tag{9}$$

Теперь перепишем целевую функцию для оптимизации в форме, аналогичной (3):

$$\mathcal{E}(\mathbf{F}) = -\frac{1}{2}\,\mathbf{F}^T\,W\,\mathbf{F} - \mathbf{I}^T\,\mathbf{F} + \frac{1}{2}\,\{\Psi^m\}^T\,\{\Psi^m\}. \tag{10}$$

Воспользуемся параметрическим представлением целевой функции $\mathcal{E}(\mathbf{F})$ относительно некоторого параметра η, характеризующего траекторию динамического процесса на фазовой плоскости. Тогда изменение целевой функции вдоль этой траектории можно представить в виде

$$\frac{d\mathcal{E}}{d\eta} = \sum_{i=1}^{n}\frac{\partial\mathcal{E}}{\partial F_i}\frac{dF_i}{d\eta}. \tag{11}$$

Из уравнения (7) получим

$$\frac{\partial \mathcal{E}}{\partial F_i} = -\left(\sum_{j=1}^{n} \omega_{ij} F_j + I_i\right). \tag{12}$$

Следовательно,

$$\frac{d\mathcal{E}}{d\eta} = -\sum_{i=1}^{n} \frac{dF_i}{d\eta}\left(\sum_{j=1}^{n} \omega_{ij} F_j + I_i\right). \tag{13}$$

Если потребовать, чтобы изменение состояний нейронов удовлетворяло соотношению

$$\frac{dF_i}{d\eta} = \lambda\left(\sum_{j=1}^{n} w_{ij} F_j + I_i\right), \tag{14}$$

при положительных λ значение производной (13) всегда будет отрицательным. Выбор $dF_i/d\eta$ в соответствии с (14) обязательно гарантирует уменьшение энергии системы. Предположив, как это было сделано ранее, что пороговое значение S_i (уровень активации) равно нулю, будем иметь эквивалентность уравнений (12) и (4).

Далее, следуя работе [6], рассмотрим дискретный случай. Изменение целевой функции после изменения состояния k-го нейрона примет вид

$$\begin{aligned} \mathcal{E} + \Delta\mathcal{E}_k &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(F_i + \Delta F_i \delta_{ik})(F_j + \Delta F_j \delta_{jk}) - \\ &- \sum_{i=1}^{n} I_i (F_i + \Delta F_i \delta_{ik}) + \frac{1}{2}\sum_{i=1}^{m} (\Psi_i^m)^2, \end{aligned} \tag{15}$$

так что

$$\begin{aligned} \Delta\mathcal{E}_k &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\,(F_j\,\Delta F_i\,\delta_{ik} + F_i\,\Delta F_j \delta_{jk} + \\ &+ \Delta F_i\,\Delta F_j\,\delta_{ik}\,\delta_{jk}) - \sum_{i=1}^{n} I_i\,\Delta F_i\,\delta_{ik}. \end{aligned} \tag{16}$$

Определим уравнения, описывающие динамику системы, в виде

$$F_i(t+1) = F_i(t) + \Delta F_i(t), \tag{17}$$

$$\Delta F_i(t) = \lambda_i \left(\sum_{j=1}^{n} W_{ij} F_j(t) + I_i\right). \tag{18}$$

Подставив (18) в (16), получим

$$\begin{aligned}\Delta\mathcal{E}_k &= -\left(\sum_{j=1}^{n} W_{kj}F_j + I_k\right)\Delta F_k - \frac{1}{2}W_{kk}\Delta F_k\Delta F_k \\ &= -(\Delta F_k)^2\left(\frac{1}{\lambda_k} + \frac{W_{kk}}{2}\right).\end{aligned} \tag{19}$$

Выбор параметра λ_kобеспечивает сходимость динамического процесса и его скорость для k-го элемента в $\mathbf{F}$. Целевая функция будет иметь отрицательные приращения

$$\Delta\mathcal{E} \leq 0, \quad \text{если} \quad \lambda_k \leq \frac{2}{|W_{kk}|} \quad \text{и} \quad W_{kk} < 0. \tag{20}$$

Монотонное убывание энергетической функции E является гарантированным, если λ_k принимает достаточно малые положительные значения.

Литература

[1] Cohen M.A., Grossberg S.O. Absolute stability of global pattern formation and parallel memory storage by compatitive neural networks // IEEE Transactions on Systems, Man and Cybernetics. - 1983. - V. 13. - P. 815 - 826.

[2] Amari Sh.-I. Mathematical foundations of neurocomputing // Proceedings of the IEEE. - 1990. - V. 78. - P. 1443 - 1463.

[3] Нейроинформатика / Горбань А.Н., Дунин-Барковский В.Л., Кардин А.Н. и др.; Отв. ред. Новиков Е.А.; РАН, Сиб. отд-е. Институт вычисл. моделирования. - Новосибирск: Наука, 1998. - 295 с.

[4] Bruck J. On the convergence properties of the Hopfield model // Proceedings of the IEEE. - 1990. - V. 78. - P. 1579 - 1585.

[5] Ревякин А.М. *Графы, матроиды и их инженерные приложения.* - М.: МИЭТ, 1991. - 178 с.

[6] Jeffrey W., Rosner R. Optimization algorithms: simulated annealing and neural network processing // The Astrophysical Journal. - 1986. - V. 310. - P. 473 - 481.