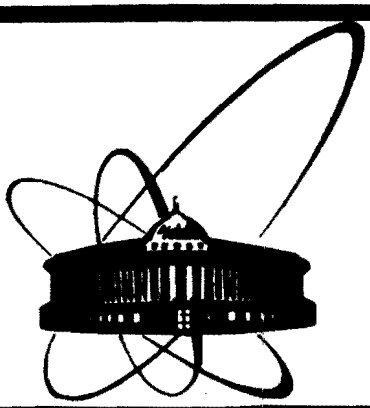

сообщения
объединенного
института
ядерных
исследований
дубна

Р17-87-145

В.А.Загребнов, Я.Л. ван Хеммен*

ЗАПОМИНАНИЕ ЭКСТЕНСИВНО БОЛЬШОГО ЧИСЛА ВЗВЕШЕННЫХ ОБРАЗОВ НАСЫЩЕННЫМИ НЕЙРОННЫМИ СЕТЯМИ

*Университет, Гейдельберг, ФРГ

1987

I. Введение

Формально нейронная сеть - это множество связанных между собой синапсами нейронов, находящихся в одном из двух возможных состояний, которое имеет такую динамику, что запоминаемые образы являются для неё устойчивыми аттракторами. Одной из главных проблем теории нейронных сетей является изучение возможных способов запоминания ими информации и описание механизма, который позволяет как находить (вспоминать), так и забывать хранящуюся в них информацию.

Обычно в качестве вполне реального приближения рассматривают сеть из N полностью связанных нейронов. Согласно работе /1/ нейрон может находиться только в одном из двух состояний (возбужденном и невозбужденном) и, следовательно, может быть ассоциирован с изинговским спином $S(i)$, $1 \le i \le N$, причем $S = +1$ соответствует возбуждению, а $S = -1$ соответствует покою. На этом языке образ есть некоторая конфигурация изинговских спинов.

Общепринято считать /2-7/, что все существенные особенности временного поведения сети описываются динамикой Монте - Карло, которая определяется гамильтонианом

$$H_N = -\frac{1}{2}\sum_{i \neq j}^{N} J_{ij}\, S(i) S(j) \quad . \qquad (I.1)$$

Это означает, что динамика сети сводится к "движению вниз с холмов ланшафта", соответствующего (свободной) энергии для H_N, а асимптотическая устойчивость предела определяется в рамках равновесной статистической механики, задаваемой (I.1). Ниже критической температуры T_c эргодичность в системе (I.1) нарушается и хранящиеся образы соответствуют притягивающим множествам (равновесные состояния или эргодические компоненты) в фазовом пространстве изинговского спинового стекла, которое мы рассмотрим ниже.

Следуя Хеббу /8/, будем считать, что память сосредоточена в синапсах, или точнее, - в распределении величин электрических потенциалов синапсов, которые определяют значения констант обменного взаимодействия в гамильтониане Изинга (I.1). При соответствующем выборе взаимодействий сеть работает как нечувствительная к отказам ассоциативная память, см., например, /2-6/. Дополнительные образы можно запомнить, если соответствующим образом модифицировать J_{ij}. Для облегчения моделирования обычно предполагают, что образы $\{\xi_{i\alpha} : 1 \le i \le N\}_\alpha$ для $1 \le \alpha \le q$ являются "случайными". Это означает,

что $\xi_{i\alpha} = \pm 1$ являются независимыми, одинаково распределенными случайными величинами, которые принимают значения ± 1 с равной вероятностью, а заданный образ является некоторой реализацией этой конечной последовательности.

Используемая для описания ассоциативной памяти модель Хопфилда /2/ соответствует следующему выбору обменного взаимодействия:

$$J_{ij} = \frac{1}{N} \sum_{\alpha=1}^{q} \xi_{i\alpha} \xi_{j\alpha} \quad . \tag{I.2}$$

В недавней работе /9/ Амит, Гутфренд и Сомполинский изучали эту сеть вблизи насыщения, т.е. когда $q = \alpha N$ и $\alpha > 0$. С помощью довольно изобретательного анализа в рамках теории среднего поля они показали, что при нулевой температуре ($T = 0$) система способна эффективно восстанавливать информацию, если $\alpha < \alpha_c \approx 0{,}14$. Однако при $\alpha = \alpha_c$ количество восстанавливаемых образов (скачком) сокращается, и при $\alpha > \alpha_c$ эффективное восстановление информации невозможно. Следовательно, система ведет себя таким образом, как если бы при $\alpha = \alpha_c$ происходил фазовый переход первого рода, после которого вся информация более или менее забывается. С точки зрения физиологии такая модель не внушает доверия. Поэтому предлагались и другие схемы работы памяти (как в режиме запоминания, так и в режиме воспоминания), но до сих пор их статус остается не очень удовлетворительным. Например, Паризи /10/ предложил модель памяти, "которая забывает", однако память, работающая по этой схеме, забывает все, за исключением самых последних образов.

В настоящей работе содержатся следующие результаты. Во-первых, предложен более простой и ясный путь, позволяющий получить основные результаты работы /9/, специальное внимание уделено анализу математического статуса используемого формализма. Во-вторых, предложено обобщение модели Хопфилда, такое, что число хранящихся образов в системе размера N также может быть равно N (случай полного насыщения), причем каждый образ имеет вес ε_ν. Если взвешивание образов соответствует временному порядку их поступления в память и $\varepsilon_\nu \to 0$ при $\nu \to \infty$, что в этом случае естественно, то при соответствующем выборе весов система запоминает экстенсивно большое число образов, однако с ростом ν их четкость постепенно уменьшается.

В разделе 2 дано описание модели и проведен её анализ в рамках теории среднего поля. В следующем разделе метод симметричной реплики использован для вычисления свободной энергии. Соответствующая этой модели ёмкость памяти при $T = 0$ найдена в разделе 4. Обсуждению полученных результатов посвящен заключительный раздел.

2. Модель Хопфилда для взвешенных образов

Мы будем рассматривать гамильтониан (I.I), в котором константы обменного взаимодействия имеют вид

$$J_{ij} = \frac{1}{N} \sum_{n=1}^{N} \varepsilon_n \xi_{in} \xi_{jn} \quad , \tag{2.I}$$

каждый образ имеет вес ε_n. Эти веса произвольны, за исключением общего требования: $0 \le \varepsilon_n \le 1$. Специальный выбор весов, соответствующий модели Хопфилда, следующий: $\varepsilon_n = 1$ для $1 \le n \le \alpha N$ и $\varepsilon_n = 0$ для $n > \alpha N$.

Так же, как и в работе /9/, мы вначале используем метод реплики и определим устойчивость некоторого конечного числа образов, которые мы обозначим ν , интегрируя по оставшимся, которые мы обозначим μ . Для этой и других целей удобно ввести в гамильтониан внешнее поле, которое выделяет ν - образы:

$$H_{ext} = -\sum_{\nu} h_\nu \sum_{i=1}^{N} \xi_{i\nu} S(i) \quad . \tag{2.2}$$

В рамках метода реплики начинают обычно /11/ с вычисления функции

$$\phi_N(n) = \frac{1}{N} \ln \langle Z_N^n \rangle \tag{2.3}$$

для положительных целых n , затем переходят к термодинамическому пределу $N \to \infty$, который равен $\phi(n)$, и, наконец, строят продолжение функции $\phi(n)$ (обычно для симметричной реплики) в окрестность $n = 0$. Тогда производная $\phi'(0)$ даёт нам $-\beta f(\beta)$, где $f(\beta)$ – плотность свободной энергии для обратной температуры β , а $Z_N = \mathrm{Tr} \exp(-\beta H_N)$ - статистическая сумма по всем изинговским конфигурациям. Угловые скобки в (2.3) обозначают усреднение по всевозможным образам $\{\xi_{in}\}$. Поскольку мы вначале интегрируем по μ -образам, то вместо $\langle Z_N^n \rangle$ сосредоточимся на вычислении

$$\left\langle \exp\left\{ \frac{\beta}{2N} \sum_{\mu,\rho} \varepsilon_\mu \left(\sum_{i=1}^{N} \xi_{i\mu} S_\rho(i) \right)^2 - \frac{1}{2} \beta n \sum_{\mu} \varepsilon_\mu \right\} \right\rangle \quad . \tag{2.4}$$

Здесь $\rho : 1 \le \rho \le n$ нумерует n реплик. Далее, вплоть до формулы (3.5), мы будем всюду опускать член $\left(-\frac{1}{2}\beta n \sum_{\mu} \varepsilon_\mu\right)$. С помощью соотношения

$$\exp\left(\frac{1}{2}\lambda a^2\right) = \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2 + \sqrt{\lambda}\, a z\right) \tag{2.5}$$

мы линеаризуем квадрат в экспоненте (2.4):

$$\left\langle \int \prod_{\mu,\rho} \frac{dm_{\mu\rho}}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\sum_{\mu,\rho} m_{\mu\rho}^2 + \sum_{\mu,\rho} m_{\mu\rho}\left[\left(\frac{\beta\varepsilon_\mu}{N}\right)^{1/2} \sum_{i=1}^{N} \xi_{i\mu} S_\rho(i) \right] \right\} \right\rangle , \qquad (2.6)$$

и преобразуем среднее по $\{\xi_{i\mu}\}$ к виду :

$$\int \prod_{\mu,\rho} \frac{dm_{\mu\rho}}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\sum_{\mu,\rho} m_{\mu\rho}^2 + \sum_{i,\mu} \ln\left(ch\left[\left(\frac{\beta\varepsilon_\mu}{N}\right)^{1/2} \sum_{\rho=1}^{n} m_{\mu\rho} S_\rho(i) \right] \right) \right\} . \qquad (2.7)$$

Если предположить, что член в фигурных скобках в (2.7) "мал" и заменить $\ln ch\, x$ для малых x на $\frac{1}{2}x^2$, то получим

$$\int \prod_{\mu,\rho} \frac{dm_{\mu\rho}}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\sum_{\mu,\rho} m_{\mu\rho}^2 + \sum_{\substack{i,\mu \\ \rho,\sigma}} \frac{\beta\varepsilon_\mu}{2N} m_{\mu\rho} m_{\mu\sigma} S_\rho(i) S_\sigma(i) \right. , \qquad (2.8)$$

где $1 \le \rho, \sigma \le n$. Для каждого фиксированного μ интегрирование по $m_{\mu\rho}$ может быть выполнено точно, и мы получаем

$$\det\left(Q_\mu\right)^{-1/2} , \qquad (2.9)$$

где Q_μ - симметричная $n \times n$ матрица, элементы которой имеют вид

$$\left(Q_\mu\right)_{\rho\sigma} = \delta_{\rho\sigma} - \beta\varepsilon_\mu \frac{1}{N} \sum_{i=1}^{N} S_\rho(i) S_\sigma(i) , \qquad (2.10)$$

здесь $\delta_{\rho\sigma}$ - символ Кронеккера. Таким образом, выражение (2.8) принимает вид

$$\prod_{\mu} \det\left(Q_\mu\right)^{-1/2} = \exp\left\{ -\frac{1}{2}\sum_{\mu} Tr\left(\ln Q_\mu\right) \right\} . \qquad (2.11)$$

Сделаем теперь следующее отступление /I2/. Для того чтобы (2.9) имело смысл, матрица Q_μ должна быть положительно определенной, т.е. её диагональные элементы должны быть положительными. Однако $(Q_\mu)_{\rho\rho} = 1 - \beta\varepsilon_\mu$ отрицательно для достаточно больших β . Причину этого легко усмотреть, если вернуться к переходу от (2.7) к (2.8). В то время, как выражение (2.7) хорошо определено для всех β (поскольку $\ln ch\, x \sim |x|$ при $|x| \to \infty$), выражение (2.8) таковым не является. Положение выглядит безвыходным, так как мы действительно нуждаемся в квадратичной аппроксимации $\ln ch\, x$ - в противном случае невозможны аналитические вычисления. Ниже мы покажем, однако, что положительная определенность восстанавливается в пределе $n \to 0$.

Зависящее от спиновых переменных выражение (2.11) представляет собой шумы, которые вносят μ -образы. Вклад этих шумов необходимо

учесть в гамильтониане для ν - образов. С точностью до оговорок, сделанных выше, это можно сделать строго . Из соотношений (2.3) и (2.11) получаем

$$\phi_N(n) = \frac{1}{N} \ln \left\langle Tr_{S_\rho} \exp\left[N\left\{ \frac{1}{2}\beta \sum_{\nu,\rho} \varepsilon_\nu \left(\frac{1}{N} \sum_{i=1}^{N} \xi_{i\nu} S_\rho(i) \right)^2 + \right.\right.\right.$$
$$\left.\left.\left. + \beta \sum_{\nu,\rho} h_\nu \left(\frac{1}{N} \sum_{i=1}^{N} \xi_{i\nu} S_\rho(i) \right) - \frac{1}{2N} \sum_{\mu} Tr\left(\ln Q_\mu\right) \right\} \right] \right\rangle . \qquad (2.12)$$

Первый след является суммой по всем 2^{nN} изинговским спиновым конфигурациям n реплик. Определим два типа параметров порядка:

$$m_{\nu\rho} = \frac{1}{N} \sum_{i=1}^{N} \xi_{i\nu} S_\rho(i) , \qquad 1 \le \rho \le n ;$$
$$q_{\rho\sigma} = \frac{1}{N} \sum_{i=1}^{N} S_\rho(i) S_\sigma(i) , \qquad 1 \le \rho < \sigma \le n . \qquad (2.13)$$

Тогда выражение в фигурных скобках в (2.12) можно представить в виде

$$F(\underset{\sim}{m}, \underset{\sim}{q}) = \beta \sum_{\nu,\rho} \left(\frac{1}{2}\varepsilon_\nu m_{\nu\rho}^2 + h_\nu m_{\nu\rho} \right) - \frac{1}{2N} \sum_{\mu} Tr \ln Q_\mu(\underset{\sim}{q}) , \qquad (2.14)$$

откуда следует, что более удобными переменными являются не спины $S_\rho(i)$, $1 \le i \le N$ и $1 \le \rho \le n$, а новые переменные $m_{\nu\rho}$ и $q_{\rho\sigma}$. Чтобы сделать этот переход, нам необходимо знать соответствующий "якобиан" $\mathcal{D}_N(\underset{\sim}{m}, \underset{\sim}{q})$. Как показано в работах /I3-I5/, в пределе $N \to \infty$ такое преобразование переменных возможно, причем с вероятностью единица (по отношению к априорной вероятностной мере) имеем

$$\mathcal{D}_N(\underset{\sim}{m}, \underset{\sim}{q}) = \exp\left\{ -N\, c^*(\underset{\sim}{m}, \underset{\sim}{q}) \right\} , \qquad (2.15)$$

где

$$c^*(\underset{\sim}{m}, \underset{\sim}{q}) = \sup_{\underset{\sim}{x}, \underset{\sim}{y}} \left\{ \underset{\sim}{m} \cdot \underset{\sim}{x} + \underset{\sim}{q} \cdot \underset{\sim}{y} - c(\underset{\sim}{x}, \underset{\sim}{y}) \right\} \qquad (2.16)$$

является преобразованием Лежандра /I6/ (строго) выпуклой функции

$$c(\underset{\sim}{x}, \underset{\sim}{y}) = \left\langle \ln Tr \exp\left\{ \sum_{\nu,\rho} x_{\nu\rho} \xi_\nu S_\rho + \sum_{(\rho,\sigma)} y_{\rho\sigma} S_\rho S_\sigma \right\} \right\rangle . \qquad (2.17)$$

Второе суммирование в (2.17) осуществляется только по парам (ρ, σ). След относится к n изинговским спинам S_ρ , $1 \le \rho \le n$, а в усреднении, обозначенном угловыми скобками, каждая переменная ξ_ν встречается только один раз, причем их число конечно. Кроме того, уже в формулировке (2.15) неявно содержится утверждение о том /I5/, что при $N \to \infty$ правая часть (2.15) не зависит от случайной конфигурации

μ – образов. По этой причине угловые скобки в выражении (2.12) можно опустить[1].

Если обозначить пару переменных $(\underset{\sim}{m},\underset{\sim}{q})$ через $\underset{\sim}{\mu}$, то из (2.12)-(2.17) получаем выражение

$$\phi(n)=\lim_{N\to\infty}\frac{1}{N}\ln\int d\underset{\sim}{\mu}\,\exp\{N[F(\underset{\sim}{\mu})-c^*(\underset{\sim}{\mu})]\}\,, \qquad (2.18)$$

которое с помощью стандартных аргументов (Лапласа) можно представить в виде

$$\phi(n)=\sup_{\underset{\sim}{\mu}}\{F(\underset{\sim}{\mu})-c^*(\underset{\sim}{\mu})\}\,. \qquad (2.19)$$

Для достаточно малых β это выражение является точным, однако оно становится совершенно формальным, если $\beta\max_{\mu}\varepsilon_\mu>1$. Если мы, несмотря на это, продолжим выкладки, то получим (см. приложение), что

$$\phi(n)=\max_{\underset{\sim}{\mu}}\{F(\underset{\sim}{\mu})-\underset{\sim}{\mu}\cdot\nabla F(\underset{\sim}{\mu})+c\,(\nabla F(\underset{\sim}{\mu}))\}\,, \qquad (2.20)$$

где $\underset{\sim}{\mu}$ удовлетворяет уравнению самосогласования (уравнение для неподвижной точки) вида

$$\underset{\sim}{\mu}=\nabla c\,(\nabla F(\underset{\sim}{\mu}))\,. \qquad (2.21)$$

Соотношение (2.20) имеет то же смысл, что и (2.19).

Матричные элементы $(Q_\mu(\underset{\sim}{q}))_{\rho\sigma}$ для $\rho<\sigma$ имеют вид

$$(Q_\mu)_{\rho\sigma}=\delta_{\rho\sigma}(1-\beta\varepsilon_\mu)-\beta\varepsilon_\mu q_{\rho\sigma}=(Q_\mu)_{\sigma\rho}\,. \qquad (2.22)$$

Используя соотношение

$$\frac{\partial}{\partial Q_{\rho\sigma}}\operatorname{Tr}\ln Q=2\,(Q^{-1})_{\rho\sigma}\,, \qquad (2.23)$$

легко проверить, что

---

[1] Такая возможность не является следствием свойства самоусреднения плотности свободной энергии, как это утверждали авторы работы /9/. Заметим, что усреднение необходимо провести под знаком логарифма. Отметим также, что параметры порядка $\tau_{\rho\sigma}$, введенные этими авторами, не имеют физического смысла, без них можно обойтись.

$$\nabla F(\underset{\sim}{\mu})=\begin{pmatrix}\beta(\varepsilon_\nu m_{\nu\rho}+h_\nu)\\ \frac{1}{N}\sum_{\mu}\beta\varepsilon_\mu(Q_\mu^{-1})_{\rho\sigma}\end{pmatrix}\,,\quad \underset{\sim}{\mu}=(\underset{\sim}{m},\underset{\sim}{q})\,. \qquad (2.24)$$

Теперь мы можем продолжить функцию (2.20) на $n=0$, предполагая, что реплика симметрична.

3. Симметричная реплика

Это означает, что реплики равны: $m_{\nu\rho}=m_\nu$ и $q_{\rho\sigma}=q$ $(\rho\neq\sigma)$. Такое требование совместимо с уравнением (2.21) для неподвижной точки. Далее, представим матрицу $Q_\mu(q)$ в виде

$$Q_\mu(q)=(1-\beta+\beta\varepsilon_\mu q)\,\mathbb{1}-\beta\varepsilon_\mu q n\,\Big|\frac{1}{\sqrt{n}}\underset{\sim}{\mathbf{1}}\Big\rangle\Big\langle\frac{1}{\sqrt{n}}\underset{\sim}{\mathbf{1}}\Big|\equiv a(n)-n\,b(n)P\,, \qquad (3.1)$$

где $\mathbb{1}$ – единичная матрица, $\underset{\sim}{\mathbf{1}}$ – единичный вектор $(1,\dots,1)\in\mathbb{R}^n$, а $P=P^2$ – проекционный оператор. Поскольку обратную матрицу можно представить в виде $Q^{-1}=c-d\cdot P$, то её матричные элементы легко найти, и

$$(Q^{-1})_{\rho\sigma}=-b(n)\left[a(n)\left(n\,b(n)-a(n)\right)\right]^{-1}\,. \qquad (3.2)$$

Следовательно,

$$\frac{1}{N}\sum_{\mu}\beta\varepsilon_\mu\left(Q_\mu^{-1}\right)_{\rho\sigma}\equiv\beta^2 q\,\tau(n)\,, \qquad (3.3)$$

где $\tau(n)$ таково, что явно не зависит от β. Тогда из (2.24) получаем

$$\nabla F(\underset{\sim}{\mu})=\begin{pmatrix}\beta(\varepsilon_\nu m_\nu+h_\nu)\\ \beta^2 q\,\tau(n)\end{pmatrix}\,. \qquad (3.4)$$

Используя (3.1), можно непосредственно убедиться, что невырожденные собственные значения матрицы $Q_\mu(q)$ равны $1-\beta\varepsilon_\mu(1-q)+\beta\varepsilon_\mu q n$, а $(n-1)$ – кратно вырожденные – равны $1-\beta\varepsilon_\mu(1-q)$. В пределе $n\to 0$ остается только величина $1-\beta\varepsilon_\mu(1-q)$, которая должна быть положительна, см. (3.7) ниже. В следующем разделе мы убедимся, что этот предел действительно положителен, см. также /9/.

С помощью (2.17), (2.20) и (3.1)-(3.4), а кроме того, добавляя константу, которую мы опускали, начиная с формулы (2.4), получаем

$$\phi(n)=-\tfrac{1}{2}\beta n\left(\tfrac{1}{N}\textstyle\sum\varepsilon_\mu\right)-\tfrac{1}{2}\beta n\sum_\nu m_\nu^2-\tfrac{1}{2N}\sum_\mu\left[\ln(1-\beta\varepsilon_\mu(1-q)+\beta\varepsilon_\mu q n)+(n-1)\times\right.$$
$$\left.\times\ln(1-\beta\varepsilon_\mu(1-q))\right]-\tfrac{1}{2}n(n-1)\tau(n)(\beta q)^2+\Big\langle\ln\operatorname{Tr}_{S_\rho}\exp\Big\{\beta\sum_{\nu,\rho}(\varepsilon_\nu m_\nu+h_\nu)\xi_\nu S_\rho+\tfrac{1}{2}\beta^2\sum_{\rho\neq\sigma}q\,\tau(n)\,S_\rho S_\sigma\Big\}\Big\rangle\,. \qquad (3.5)$$

Если воспользоваться представлением (2.5) и провести линеаризацию, то след можно вычислить и представить последний член в (3.5) в виде

$$-\tfrac{1}{2}n\beta^2 q\tau(n)+n\ln 2+\left\langle \ln\int_{-\infty}^{+\infty}\frac{dz}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\mathrm{ch}^n\{\beta[(\varepsilon\underset{\sim}{m}+\underset{\sim}{h})\cdot\underset{\sim}{\xi}+z\sqrt{q\tau(n)}]\}\right\rangle, \tag{3.6}$$

где $\varepsilon=\mathrm{diag}(\varepsilon_\nu)$ - диагональная матрица. Если выбрать теперь в качестве продолжения $\phi(n)$ её "очевидное" продолжение как вещественной функции, то получим

$$-\beta f(\beta)=\lim_{n\to 0}n^{-1}\phi(n)=-\frac{1}{2N}\sum_\mu\varepsilon_\mu-\frac{1}{2}\beta\sum_\nu\varepsilon_\nu m_\nu^2-\frac{1}{2N}\sum_\mu\Big[\ln(1-$$
$$-\beta\varepsilon_\mu(1-q))-\beta\varepsilon_\mu q(1-\beta\varepsilon_\mu(1-q))^{-1}\Big]-\tfrac{1}{2}\beta^2q(1-q)\tau+ \tag{3.7}$$
$$+\left\langle\int\frac{dz}{\sqrt{2\pi}}e^{-z^2/2}\ln\left[2\mathrm{ch}\{\beta((\varepsilon\underset{\sim}{m}+\underset{\sim}{h})\cdot\underset{\sim}{\xi}+z\sqrt{q\tau})\}\right]\right\rangle,$$

здесь подразумевается, что N велико и

$$\tau=\lim_{n\to 0}\tau(n)=\frac{1}{N}\sum_\mu\varepsilon_\mu^2\left[1-\beta\varepsilon_\mu(1-q)\right]^{-2}\geqslant 0. \tag{3.8}$$

Далее, необходимо выбрать такое решение уравнений для неподвижной точки

$$\underset{\sim}{m}=\ll\underset{\sim}{\xi}\,\mathrm{th}\{\beta[\varepsilon\underset{\sim}{m}+\underset{\sim}{h}]\cdot\underset{\sim}{\xi}+z\sqrt{q\tau}\}\gg, \tag{3.9а}$$

$$q=\ll\mathrm{th}^2\{\beta[\varepsilon\underset{\sim}{m}+\underset{\sim}{h}]\cdot\underset{\sim}{\xi}+z\sqrt{q\tau}\}\gg, \tag{3.9б}$$

которое обеспечивает максимум правой части (3.7). Двойные угловые скобки в (3.9) обозначают среднее и по конечному числу переменных $\{\xi_\nu\}$, и по гауссовскому распределению для переменной z. Если $\varepsilon_\mu=1$ для $1\leqslant\mu\leqslant\alpha N$ и $\varepsilon_\mu=0$ для $\mu>\alpha N$, тогда (3.7)-(3.9) воспроизводят результаты работы /9/ (если всюду положить $\underset{\sim}{h}=0$). Из (3.8) и (3.9) следует, что τ необходимо интерпретировать как константу, перенормирующую параметр порядка q. Это следствие шума, который порождает хвост из "бесконечного числа" ($N\to\infty$) образов.

4. Ёмкость памяти

Для заданного набора весовых факторов ε_μ ёмкостью памяти называется (максимальное) число образов, которые не теряют своей стабильности полностью. В этом разделе мы исследуем ёмкость памяти при $T=0$ для сети с весовыми факторами $\varepsilon_\mu=\mu^{-\alpha}$, $\alpha>0$. При соответствующем выборе α имеется постепенная потеря чёткости образов по мере роста индекса μ. Легко проверить, что для $\alpha>0$ среднее $N^{-1}\sum_\mu\mu^{-\alpha}$ сходится (при $N\to\infty$) к нулю (сходимость $\mu^{-\alpha}$ по Чезаро). Поэтому член $N^{-1}\sum_\mu\mathrm{Tr}(\ln Q_\mu)$, соответствующий фоновому шуму, также стремится (при $N\to\infty$) к нулю. Это не означает, однако, что им можно полностью пренебречь, поскольку вес ε_μ образа с номером μ также стремится к нулю при $\mu\to\infty$. Имеется, однако, особенность, связанная с предельными переходами, которые мы хотим изучать.

Мы будем рассматривать два предела $N\to\infty$ и $\beta\to\infty$, которые необходимо совершать в определенном порядке. Вначале мы должны совершить переход к термодинамическому пределу $N\to\infty$, а затем к пределу нулевой температуры $\beta\to\infty$. Поскольку параметр q при этом будет стремиться к единице, мы начнем с уравнения (3.9а). В пределе $\beta\to\infty$ функция $\mathrm{th}\{\ldots\}$ стремится к $\mathrm{sgn}\{\ldots\}$, а интеграл по гауссовскому распределению совпадает, с точностью до члена порядка T, с интегралом ошибок:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}dz\,e^{-z^2/2}\,\mathrm{th}\{\ldots\}\approx\sqrt{\frac{2}{\pi}}\int_0^{\varepsilon\underset{\sim}{m}\cdot\underset{\sim}{\xi}/\sqrt{\tau}}dz\,e^{-z^2/2}= \tag{4.1}$$
$$=\mathrm{erf}\,(\varepsilon\underset{\sim}{m}\cdot\underset{\sim}{\xi}/\sqrt{\tau}),$$

где

$$\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x dy\,e^{-y^2}, \tag{4.2}$$

так что уравнение (3.9а) принимает вид

$$\underset{\sim}{m}=\left\langle\underset{\sim}{\xi}\,\mathrm{erf}\,(\varepsilon\underset{\sim}{m}\cdot\underset{\sim}{\xi}/\sqrt{2\tau})\right\rangle_{\underset{\sim}{\xi}}. \tag{4.3}$$

Рассмотрим поведение одного выделенного образа с номером ν. Тогда $\underset{\sim}{m}$ имеет только одну ненулевую компоненту m и

$$m=\langle\xi_\nu\,\mathrm{erf}\,(m\varepsilon_\nu\xi_\nu/\sqrt{2\tau})\rangle=\mathrm{erf}\,(m\varepsilon_\nu/\sqrt{2\tau}). \tag{4.4}$$

Как следует из (2.13), чем ближе m к единице, тем лучше восстановление образа. Поскольку веса ε_ν стремятся к нулю в смысле Чезаро, параметр τ сходится к нулю при $N\to\infty$, а величина m (для фиксированного ν) стремится к единице. Аналогично для конечной группы фиксированных индексов $\{\nu\}$ и конечной температуры получаем

$$\underset{\sim}{m}=\left\langle\underset{\sim}{\xi}\,\mathrm{th}\{\beta\varepsilon\underset{\sim}{m}\cdot\underset{\sim}{\xi}\}\right\rangle_{\underset{\sim}{\xi}}, \tag{4.5}$$

а в случае единственной ненулевой компоненты

$$m = th\,\beta\varepsilon_\nu m\,. \qquad (4.6)$$

Следовательно, чем больше ν, тем меньше критическая температура $T_c(\nu)$, ниже которой появляется ветвь ненулевых значений m, соответствующих образу с индексом ν (компонента в направлении образа ν).

Обратимся теперь к уравнению (3.9б). В термодинамическом пределе оно сводится к равенству:

$$q = \left\langle th^2\{\beta\varepsilon \underset{\sim}{m}\cdot\underset{\sim}{\xi}\}\right\rangle_{\underset{\sim}{\xi}}\,, \qquad (4.7)$$

тогда при $\beta\to\infty$ правая часть сходится к единице экспоненциально быстро, и $C(\beta)=\beta(1-q)$ стремится к нулю. Как и выше, $\varepsilon = diag(\varepsilon_\nu)$.

Что произойдет для очень больших, но конечных N? Какое количество образов система может хранить при $T = 0$ в этом случае? Фактор шума τ в этом случае отличен от нуля. Его можно оценить, используя (3.8) и рассуждения, приведенные выше:

$$\tau = N^{-1}\sum_\mu\left[\varepsilon_\mu^{-1} - C(\beta)\right]^{-2} \xrightarrow[\beta\to\infty]{} N^{-1}\sum_\mu \mu^{-2x}\,. \qquad (4.8)$$

Это выражение, в свою очередь, можно оценить следующим образом:

$$\tau = N^{-1}\int_1^N d\mu\,\mu^{-2x} = (1-2x)^{-1}\left[N^{-2x} - N^{-1}\right]\,, \qquad (4.9)$$

если $x \neq 1/2$, и, соответственно,

$$\tau = N^{-1}\int_1^N d\mu\,\mu^{-1} = N^{-1}\ln N\,, \qquad (4.10)$$

если $x = 1/2$. Мы будем различать три случая: (а) $0 < x < 1/2$; (б) $x = 1/2$; (в) $x > 1/2$. В первом случае из (4.9) получаем

$$\tau = (1-2x)^{-1}N^{-2x}\,. \qquad (4.11а)$$

Во втором случае из (4.10) следует

$$\tau = N^{-1}\ln N\,, \qquad (4.11б)$$

в то время как в последнем случае имеем

$$\tau = (2x-1)^{-1}N^{-1}\,. \qquad (4.11в)$$

Теперь эти оценки для τ позволяют нам вернуться к выражению (4.4).

Функция $\psi(x) = erf(\eta x)$ является выпуклой $(\eta > 0)$ и монотонно возрастающей по $x \geq 0$. Она начинается в нуле и стремится к единице. Её изменения напоминают поведение $th(\eta x)$, за исключением наклона в нуле, который равен $2\eta/\sqrt{\pi}$, см. (4.2). Следовательно, уравнение

$$m = erf(\eta m) \qquad (4.12)$$

имеет единственное решение $m = 0$, если $2\eta/\sqrt{\pi} < 1$, т.е. $\eta < \eta_c = \frac{\sqrt{\pi}}{2}$. Для $\eta > \eta_c$ существует, кроме того, и нетривиальное решение, которое непрерывно стремится к нулю, когда $\eta \to \eta_c$ сверху.

Из (4.4) и (4.11) следует, что существует такое критическое значение ν_c, что образы с $\nu > \nu_c$ не запоминаются. Имеем

$$\left.\frac{\varepsilon_\nu}{\sqrt{2\tau}}\right|_{\nu=\nu_c} = \frac{1}{2}\sqrt{\pi} \Longleftrightarrow \sqrt{\tau}\,(\nu_c)^x = \sqrt{\frac{2}{\pi}}\,. \qquad (4.13)$$

Тогда в случае (а) получаем

$$(1-2x)^{-1/2}\frac{(\nu_c)^x}{N} = \sqrt{\frac{2}{\pi}}\,, \qquad (4.14)$$

так что величина ν_c является экстенсивной:

$$\nu_c = \left[(1-2x)\frac{2}{\pi}\right]^{1/2x} N\,. \qquad (4.15а)$$

В случае (б) имеем

$$\nu_c = \frac{2}{\pi}N/\ln N\,, \qquad (4.15б)$$

в то время, как в случае (в), когда $2x > 1$, эта величина равна

$$\nu_c = \left[(2x-1)\frac{2}{\pi}\right]^{1/2x} N^{1/2x}\,. \qquad (4.15в)$$

Следовательно, в последних двух случаях ν_c не является экстенсивной.

Оптимальную величину коэффициента $\alpha \equiv \left[(1-2x)\frac{2}{\pi}\right]^{1/2x}$ в (4.15а) можно найти, варьируя x. Это дает $\alpha_{max} = 0{,}103$ для $x = 0{,}280$. Тем не менее для этой величины показателя x значение ν_{max} необходимо уменьшить до $0{,}01\,N$, чтобы величина ошибки в воспроизведении информации не превышала 0,5 %. Наоборот, можно

фиксировать величину ошибки, скажем 0,5%, и найти величину $\varkappa$, которая делает соответствующее значение ν максимальным, см. (4.4). Эта процедура дает $\nu_{max} = 0{,}0138\,N$ для $\varkappa = 0{,}386$, что не является существенным улучшением.

Смысл оптимизации, приведенной выше, - прост. Если $\varkappa = 0$, то память находится в состоянии полной неразберихи /9/, т.е. никаких образов не хранит. С другой стороны, если $\varkappa \geq 1/2$, $\nu \to \infty$, то образы легко теряются в фоновом шуме и хранение экстенсивного их числа также становится невозможным. Оптимальное значение $\varkappa$ лежит между этими крайними значениями.

5. Обсуждение

Для последовательности весов вида $\varepsilon_\nu = \nu^{-\varkappa}$ существует такое критическое значение ν_c, что воспроизведение образов с индексами $\nu > \nu_c$ становится невозможным: образы теряются в фоновом шуме. Для $\nu < \nu_c$ корреляция между хранящимися в памяти и воспроизводимыми образами существует, но стремится к нулю (непрерывно), когда $\nu \to \nu_c$ снизу. В этом случае образы постепенно теряют четкость вплоть до полного исчезновения в точке ν_c. Если $0 < \varkappa < 1/2$, тогда ν_c является экстенсивной величиной, пропорциональной объему системы.

Все эти результаты, и в частности (4.15), легко понять качественно. При нулевой температуре образ с индексом ν должен быть устойчивой неподвижной точкой для динамики, которая задается соотношением

$$S(i) := \operatorname{sgn}\Big(\sum_j{}' J_{ij}\, S(j)\Big) \,. \tag{5.1}$$

Константа J_{ij} определена в (2.1) и, следовательно,

$$J_{ij} \sim \sum_n \varepsilon_n \xi_{in} \xi_{jn} \,, \tag{5.2}$$

тогда (с учетом $N-1 \approx N$) имеем

$$\sum_j J_{ij}\, \xi_{j\nu} \sim \sum_{n,j} \varepsilon_n \xi_{jn} \xi_{j\nu} \xi_{in} = \varepsilon_\nu \xi_{i\nu} N + \sum_{\mu \neq \nu, j} \varepsilon_\mu \xi_{i\mu} \xi_{j\mu} \xi_{j\nu} \,. \tag{5.3}$$

Здесь сумма по $\mu\,(\neq\nu)$ и $j\,(\neq i)$ является суммой независимых, одинаково распределенных случайных величин, поэтому её значение (по порядку величины) равно квадратному корню из дисперсии, т.е. равно $\big(N \sum_\mu \varepsilon_\mu^2\big)^{1/2}$. Значение ν_c определяется тогда из условия того, что оба члена в правой части (5.3) имеют одинаковую величину:

$$(\varepsilon_\nu N)^2 \approx N \sum_\mu \varepsilon_\mu^2 \,. \tag{5.4}$$

Следовательно, для $\nu = \nu_c$ получаем

$$\varepsilon_\nu^2 \approx N^{-1} \sum_\mu \varepsilon_\mu^2 \,, \tag{5.5}$$

где, по определению, $\varepsilon_\nu^2 = 1/\nu^{2\varkappa}$. Это соотношение, с точностью до (существенного) множителя $2/\pi$, который возникает из интеграла ошибок (4.4), воспроизводит, с учетом (4.8)-(4.10), условие (4.15). Без учета этого множителя оптимальная величина константы α была бы равна максимуму функции $(1-2\varkappa)^{1/2\varkappa}$, т.е. $e^{-1} = 0{,}368$, который достигается в точке $\varkappa = +0$.

В заключение суммируем полученные результаты. Мы предлагаем простой и строгий вывод термодинамических свойств модели Хопфилда вблизи насыщения. Этот анализ подтверждает основные результаты работы /9/, а также позволяет распространить модель на случай взвешенных образов. При соответствующем выборе ε_ν имеет место постепенное уменьшение четкости образов с ростом ν, что можно интерпретировать как забывание образа с ростом промежутка времени с момента его поступления в память. Однако необходимы дополнительные основания для того, чтобы решить, является ли процесс забывания внутренним свойством памяти либо связан с течением времени или с обоими этими факторами.

Один из авторов (Я.Л. ван Хеммен) благодарен ОИЯИ за гостеприимство, способствовавшее написанию настоящей работы. Он также благодарен Немецкому исследовательскому обществу (Бонн) за финансовую поддержку.

Приложение

Здесь мы хотим показать, что, если $c(\underset{\sim}{t})$ является строго выпуклой функцией, а $c^*(\underset{\sim}{m})$ - её преобразование Лежандра и функция $F(\underset{\sim}{m})$ гладкая, то выражение для

$$\sup_{\underset{\sim}{m}} \{ F(\underset{\sim}{m}) - c^*(\underset{\sim}{m}) \} \tag{П.1}$$

можно представить в виде

$$\max_{\underset{\sim}{\mu}} \{ F(\underset{\sim}{\mu}) - \underset{\sim}{\mu} \cdot \nabla F(\underset{\sim}{\mu}) + c\,(\nabla F(\underset{\sim}{\mu})) \} \,, \tag{П.2}$$

где $\underset{\sim}{\mu}$ удовлетворяет уравнению для неподвижной точки:

$$\underset{\sim}{\mu} = \nabla c\,(\nabla F(\underset{\sim}{\mu})) \,. \tag{П.3}$$

Доказательство сводится к следующему. Заметим, что $\underset{\sim}{t} \to \nabla c(\underset{\sim}{t})$ - это отображение $\mathbb{R}^n$ на $\mathbb{R}^n$. Обратное этому отображению существует и равно ∇c^*, см. /16/. Поскольку sup в (П.1) реализуется

для тех $\underset{\sim}{m}$, которые удовлетворяют соотношению $\nabla F(\underset{\sim}{m}) = \nabla c^*(\underset{\sim}{m})$, то мы немедленно получаем уравнение для неподвижной точки:

$$\underset{\sim}{m} = \nabla c\left(\nabla F(\underset{\sim}{m})\right). \tag{П.4}$$

Его решения обозначим символом $\underset{\sim}{\mu}$.

Рассмотрим теперь $c^*(\underset{\sim}{\mu})$. По определению,

$$c^*(\underset{\sim}{\mu}) = \sup_{\underset{\sim}{t}} \left\{\underset{\sim}{\mu}\cdot\underset{\sim}{t} - c(\underset{\sim}{t})\right\}. \tag{П.5}$$

Чтобы получить sup, мы должны найти такое $\underset{\sim}{t}$, что

$$\nabla c(\underset{\sim}{t}) = \underset{\sim}{\mu}. \tag{П.6}$$

Однако $\underset{\sim}{\mu}$ удовлетворяет уравнению (П.4). Сравнивая (П.6) с (П.3), получаем, что $\underset{\sim}{t} = \nabla F(\underset{\sim}{\mu})$. Если подставить это соотношение в (П.5) и вернуться к (П.I), тогда немедленно получим (П.2).

Литература

I. McCulloch W.S. and Pitts W.A. Bull. Math. Biophys., 5 (I943), II5.
2. Hopfield J.J. Proc. Nat. Acad. Sci. USA, 79 (I982) 2554; ibid. 8I (I984) 3088.
3. Little W.A. Math. Biosci., I9 (I974) I0I; Little W.A. and Shaw G.L. Math. Biosci., 39 (I978) 28I.
4. Peretto P. Biol. Cybern., 50 (I984) 5I.
5. Toulouse G., Dehaene S. and Changeux J.-P. Proc. Nat. Acad. Sci. USA, 83 (I986) I695.
6. van Hemmen J.L. and Kühn R. Phys. Rev. Lett., 57 (I986) 9I3.
7. van Enter A.C.D. and van Hemmen J.L. Phys. Rev., A 29 (I984) 355.
8. Hebb D. The Organization of Behavior (Wiley, New York, I949).
9. Amit D.J., Gutfreund H. and Sompolinsky H. Phys. Rev. Lett., 55 (I985) I530 and Ann. Phys. (N.Y.), to appear.
I0. Parisi G. J. Phys. A: Math. Gen, I9 (I986) L 6I7.
II. van Hemmen J.L. and Palmer R.G. J. Phys. A: Math. Gen., I2 (I979) 563.
I2. Grensing D., Kühn R. and van Hemmen J.L. to be published.
I3. van Hemmen J.L. Phys. Rev. Lett., 49 (I982) 409.
I4. van Hemmen J.L., van Enter A.C.D. and Canisius J. Z. Phys., 50 (I983) 3II.

I5. van Hemmen J.L. In: Heibelberg Colloquium on Spin Glasses, edited by van Hemmen J.L. and Morgenstern I., Leoture Notes in Physios, I92 (Springer, Heidelberg, I983), pp. 203–233, in particular the Appendix.
I6. Roberts A.W. and Varberg D.E. Convex Functions (Academic Press, New York, I973).

Рукопись поступила в издательский отдел
9 марта I987 года.

НЕТ ЛИ ПРОБЕЛОВ В ВАШЕЙ БИБЛИОТЕКЕ?

Вы можете получить по почте перечисленные ниже книги, если они не были заказаны ранее.

Д9-82-664	Труды совещания по коллективным методам ускорения. Дубна, 1982.	3 р. 30 к.
Д3,4-82-704	Труды IV Международной школы по нейтронной физике. Дубна, 1982.	5 р. 00 к.
Д11-83-511	Труды совещания по системам и методам аналитических вычислений на ЭВМ и их применению в теоретической физике. Дубна, 1982.	2 р. 50 к.
Д7-83-644	Труды Международной школы-семинара по физике тяжелых ионов. Алушта, 1983.	6 р. 55 к.
Д2,13-83-689	Труды рабочего совещания по проблемам излучения и детектирования гравитационных волн. Дубна, 1983.	2 р. 00 к.
Д13-84-63	Труды XI Международного симпозиума по ядерной электронике. Братислава, Чехословакия, 1983.	4 р. 50 к.
Д2-84-366	Труды 7 Международного совещания по проблемам квантовой теории поля. Алушта, 1984.	4 р. 30 к.
Д1,2-84-599	Труды VII Международного семинара по проблемам физики высоких энергий. Дубна, 1984.	5 р. 50 к.
Д17-84-850	Труды III Международного симпозиума по избранным проблемам статистической механики. Дубна.1984. /2 тома/	7 р. 75 к.
Д10,11-84-818	Труды V Международного совещания по проблемам математического моделирования, программированию и математическим методам решения физических задач. Дубна, 1983	3 р. 50 к.
	Труды IX Всесоюзного совещания по ускорителям заряженных частиц. Дубна, 1984 /2 тома/	13 р.50 к.
Д4-85-851	Труды Международной школы по структуре ядра, Алушта, 1985.	3 р. 75 к.
Д11-85-791	Труды Международного совещания по аналитическим вычислениям на ЭВМ и их применению в теоретической физике. Дубна,1985.	4 р.
Д13-85-793	Труды XII Международного симпозиума по ядерной электронике. Дубна 1985.	4 р. 80 к.
Д3,4,17-86-747	Труды V Международной школы по нейтронной физике. Алушта,1986.	4 р. 50 к.

Заказы на упомянутые книги могут быть направлены по адресу:
101000 Москва, Главпочтамт, п/я 79
Издательский отдел Объединенного института ядерных исследований

Загребнов В.А., ван Хеммен Я.Л. Р17-87-145

Запоминание экстенсивно большого числа взвешенных образов насыщенными нейронными сетями

Эффективность модели Хопфилда для нейронных сетей анализируется в случае, когда число запоминаемых образов экстенсивно велико и им приписаны некоторые веса. Если размеры системы равны N, тогда она может запомнить αN образов с соответствующими весами. Эти веса можно связать с временным порядком запомнившихся образов либо, при соответствующем выборе, с процессом постепенного забывания экстенсивно большого числа образов. Особое внимание уделено математической структуре модели.

Работа выполнена в Лаборатории теоретической физики ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1987

Перевод Г.Г.Сандуковской

Zagrebnov V.A., van Hemmen J.L. P17-87-145

Strong Extensively Many Weighted Patterns in a Saturated Neural Network

The performance of the Hopfield model of a neural network with extensively many weighted patterns is analyzed. If the system size is N, then N patterns, each provided with a suitable weight, are stored. The weights may be associated with a temporal order and, if appropriately chosen, they allow a gradual fading out of the extensively many stored patterns. Particular emphasis is put on the underlying mathematical structure.

The investigation has been performed at the Laboratory of Theoretical Physics, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1987