

О вероятности обнаружения локальных минимумов в обобщенной модели Хопфилда*

Крыжановский Б.В.¹⁾, Магомедов Б.М.²⁾

¹⁾ Институт оптико-нейронных технологий РАН, Москва, kryzhanov@mail.ru

²⁾ Институт оптико-нейронных технологий РАН, Москва, bashir.magomedov@gmail.com

Аннотация. На основе анализа обобщенной модели Хопфилда получены выражения, устанавливающие связь между глубиной локального минимума и шириной области притяжения. На основании этого вероятность нахождения локального минимума при случайной инициализации нейронной сети удалось представить как функцию глубины этого минимума. В практических оптимизационных приложениях наличие таких выражений позволит по ряду уже найденных минимумов оценить вероятность нахождения более глубокого минимума и принять решение на остановку программы поиска или ее продолжение. Развита аналитическая теория находится в хорошем согласии с результатами компьютерного эксперимента.

1. Введение

Обычно нейронная система ассоциативной памяти рассматривается как система, решающая задачи распознавания или восстановления образов. Однако, ее можно рассматривать и как систему, решающую оптимизационную задачу – нейронная сеть в процессе релаксации находит конфигурацию, соответствующую минимуму энергии [1]. Это свойство нейронной сети можно использовать для решения различных NP -полных задач. Стандартный подход сводится к нахождению такой архитектуры и параметров нейронной сети, при которых целевая (целевая) функция совпадает с понятием нейросетевой энергии. Успешное применение нейронной сети к задаче коммивояжера [2] инициировало широкие исследования нейросетевых подходов к решению задач теории графов [3], нейросетевой оптимизации обработки изображений [4] и большому ряду других приложений. Эта область теории нейронных сетей активно развивается по настоящее время.

Исследования в указанной области объединяет одно – сердцевиной подавляющего большинства нейросетевых оптимизационных алгоритмов является модель Хопфилда [1], а сам процесс оптимизации так или иначе сводится к нахождению в N -мерном конфигурационном пространстве глобального минимума некоего квадратичного функционала (энергии), построенного на заданной $N \times N$ - матрице. Стандартный нейросетевой подход к решению указанной задачи сводится к процедуре случайного поиска оптимального решения: на первом этапе этой процедуры нейронная сеть случайным образом инициализируется, на втором этапе – нейронная сеть релаксирует в одно из возможных ста-

* Работа выполнена при поддержке РФФИ (проект No 05-09-90049-в).

бильных состояний, т.е. оптимизирует величину энергии. Поскольку искомым результатом неизвестен и поиск проводится "вслепую", то нейросеть инициализируется раз за разом, с тем, чтобы найти как можно более глубокий минимум энергии. Несмотря на относительно успешное применение [2] этого подхода, всегда остается открытым вопрос – сколько таких случайных попыток следует сделать и когда проведенный поиск можно считать удачным (исходя из соотношения "затраченное время/полученный результат"). Сказанное справедливо как при спиновой [1-5], так и доменной [6,7] динамиках релаксации.

В настоящей работе, в продолжение проведенного в [10] анализа, получены выражения, устанавливающие связь между глубиной локального минимума энергии и шириной области притяжения. На основании этого, вероятность нахождения локального минимума при случайной инициализации нейронной сети удалось представить как функцию глубины этого минимума. В практических приложениях наличие таких выражений позволит по ряду уже найденных минимумов оценить вероятность нахождения более глубокого минимума и принять решение на остановку программы поиска или ее продолжение. Выражения получены на основе анализа обобщенной модели Хопфилда – нейронной сети с хэббовской [8] (корреляционной) матрицей межсвязей. Для такого типа матриц получено хорошее совпадение аналитических расчетов с результатами компьютерных экспериментов. Введенное в разделе 2 обобщение модели Хопфилда допускает возможность эвристического обобщения полученных результатов и на случай матриц иного типа.

2. Обобщенная модель Хопфилда

Рассмотрим модель Хопфилда. В теории нейронных сетей ее принято описывать как одномерную систему из N спинов-нейронов, которые могут ориентироваться вдоль, либо против заданной оси. Состояние такой нейронной сети характеризуется конфигурационным вектором $\mathbf{S} = (s_1, s_2, \dots, s_N)$, где $s_i = \pm 1$, $i \in (1, N)$. Здесь мы будем рассматривать обобщенную модель, архитектура которой задается матрицей межсвязей

$$T_{ij} = \sum_{m=1}^M r_m s_{mi} s_{mj} \quad (1)$$

с нулевой диагональю ($T_{ii} = 0$), организованной по правилу Хэбба на M образах – N -мерных бинарных эталонных векторах $\mathbf{S}_m = (s_{m1}, s_{m2}, \dots, s_{mN})$. Обобщение состоит в том, что каждый эталон \mathbf{S}_m добавляется в матрицу T_{ij} со своим статвесом r_m . Для упрощения выражений принята нормировка статвесов $\sum r_m^2 = 1$, не нарушающая общности рассмотрения. Это незначительное видоизменение модели оказывается весьма существенным, поскольку, в отличие от стандартной модели, позволяет описывать нейронную сеть с невырожденным спектром минимумов.

Энергия нейросети описывается выражением

$$E = - \sum_{i,j=1}^N s_i T_{ij} s_j, \quad (2)$$

а ее динамика заключается в следующем. Задается начальное состояние сети \mathbf{S} (начальные направления спинов устанавливаются в соответствии со знаками компонент вектора \mathbf{S}). Затем вычисляется локальное поле $h_i = -\partial E / \partial s_i$, воздействующее на произвольно выбранный i -й спин со стороны всех остальных спинов системы в момент времени t и определяется энергия спина в этом поле $\varepsilon_i = -s_i h_i$. Если направление спина совпадает с направлением локального поля ($\varepsilon_i < 0$), то его положение энергетически устойчиво и в последующий момент времени состояние спина остается неизменным. В противном случае ($\varepsilon_i > 0$) положение спина неустойчиво и он разворачивается вдоль направления этого поля, переходя в состояние $s_i(t+1) = -s_i(t)$ с энергией $\varepsilon_i(t+1) < 0$. Такая процедура последовательно применяется ко всем спинам нейронной сети. При каждом перевороте спина энергия сети понижается. Это означает, что сеть за конечное число шагов перейдет в стабильное состояние, соответствующее локальному минимуму энергии.

3. Размер области притяжения

Исследуем, при каких условиях эталонный образ \mathbf{S}_m будет являться неподвижной точкой [1], в которой энергия сети E достигает своего (локального) минимума E_m . Оценки получим в асимптотическом пределе $N \rightarrow \infty$. Определим область притяжения образа \mathbf{S}_m как совокупность точек N -мерного конфигурационного пространства, из которых нейронная сеть релаксирует в конфигурацию \mathbf{S}_m , и попробуем оценить размер этой области. Пусть начальное состояние сети \mathbf{S} находится в некоторой окрестности образа \mathbf{S}_m . Тогда вероятность того, что сеть сойдется к точке \mathbf{S}_m , опишется выражением:

$$Pr = 2^{-N} (1 + \operatorname{erf} z)^N. \quad (3)$$

Здесь $\operatorname{erf} z$ – функция ошибок переменной z :

$$z = r_m \frac{\sqrt{N}}{\sqrt{2(1-r_m^2)}} \left(1 - \frac{2n}{N} \right), \quad (4)$$

где n – хеммингово расстояние между \mathbf{S}_m и \mathbf{S} . Выражение (3) можно получить методами теоретико-вероятностного подхода (см. Приложение), повторив хорошо известные для случая $r_m = 1/\sqrt{M}$, $m = 1, 2, \dots, M$ выкладки [5].

Из (3) следует, что область притяжения определяется как совокупность близких к \mathbf{S}_m точек конфигурационного пространства, для которых справедливо соотношение $n \leq n_m$:

$$n_m = \frac{N}{2} \left(1 - \frac{r_0}{r_m} \sqrt{1-r_m^2} \right), \quad (5)$$

где

$$r_0 = \sqrt{2 \ln N / N}. \quad (6)$$

Действительно, при $n \leq n_m$ вероятность схождения в точку S_m с ростом N асимптотически стремится к единице; в противном случае ($n > n_m$) имеем $Pr \rightarrow 0$. Это означает, что величину n_m можно рассматривать как радиус области притяжения локального минимума E_m .

Из (5) следует, что при $r_m \rightarrow r_0$ радиус области притяжения стремится к нулю (рис.1). Это означает, что эталонные образы, прописанные в матрицу межсвязей (1) с статвесом, меньшим r_0 , попросту не образуют локальных минимумов. Локальные минимумы имеются только в точках S_m , образы которых прописаны в матрице межсвязей с относительно большими статвесами $r_m > r_0$. Более того, если статвес одного из образов, например, образа S_1 , сделать значительно больше чем у остальных ($r_1 \rightarrow 1$), то область его притяжения ($n_1 \sim N/2$) охватит половину всего N -мерного пространства (остальное полу-пространство занимает область притяжения его негатива $S_1^* = -S_1$). При этом сеть будет иметь всего лишь два локальных минимума в точках S_1 и S_1^* , остальные локальные минимумы исчезнут. Этот предельный случай не представляет практического интереса для оптимизационных задач и ниже мы его рассматривать не будем.

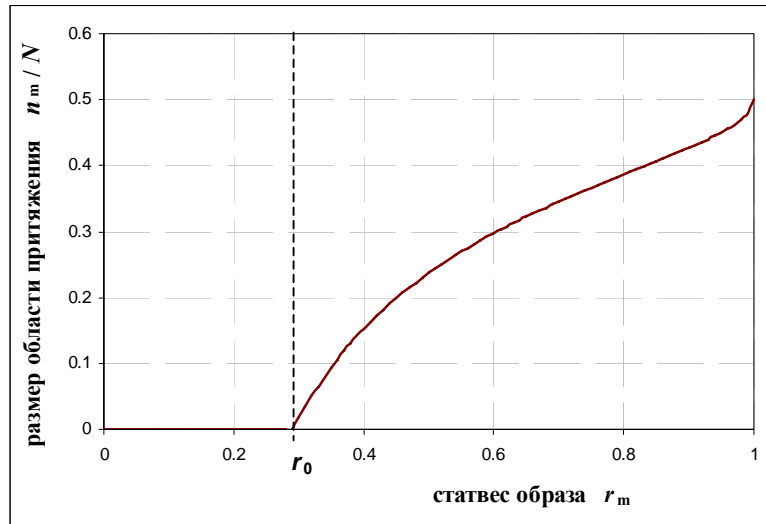


Рис.1. Характерный вид зависимости ширины области притяжения от статвеса образа. Локальный минимум существует только для образов, статвес которых в матрице межсвязей больше r_0 . При $r_m \rightarrow r_0$ размер области притяжения стремится к нулю, т.е. образы со статвесом $r_m < r_0$ не создают локальных минимумов.

4. Объем памяти нейронной сети в обобщенной модели

Из проведенного выше анализа следует, что не все из прописанных в матрице межсвязей образов образуют локальные минимумы и могут распознаваться нейронной сетью. В связи с этим следует дать четкое определение понятия объема нейросетевой памяти \bar{M} . Определим объем памяти как число тех образов, которые образуют локальные минимумы и, следовательно, могут быть восстановлены нейронной сетью.

Очевидно, что объем памяти зависит от того, как сформирована матрица межсвязей и, в частности, от распределения статвесов. Рассмотрим, например, случай, когда статвеса образов случайны и равномерно распределены на отрезке $r_m^2 \in [0, 2/M]$ со средним $\overline{r_m^2} = 1/M$. В этом случае, вероятность того, что статвес образа удовлетворяет условию $r_m > r_0$, равна величине $1 - Mr_0^2/2$. Соответственно, объем памяти модели определится выражением:

$$\bar{M} = M(1 - Mr_0^2/2). \quad (7a)$$

Зависимость объема памяти от числа прописанных в матрице межсвязей образов M , представлена на рис.2. Как видим, максимальный объем памяти $\bar{M} = N/4 \ln N$ достигается при числе образов $M = N/2 \ln N$. При $M \geq N/\ln N$ объем памяти сети обращается в ноль – нет ни одного образа, образующего локальный минимум.

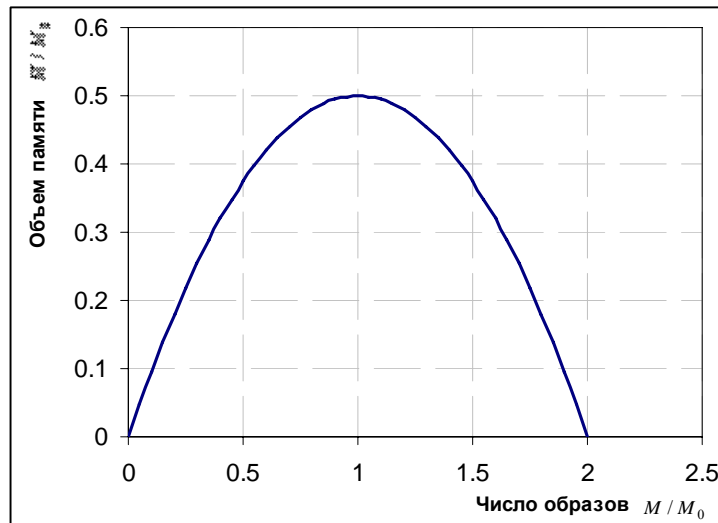


Рис.2. Вид зависимости объема памяти от числа образов M , прописанных в матрице межсвязей. Все величины нормированы на величину объема памяти $M_0 = N/2 \ln N$ стандартной модели Хопфилда. Квадраты статвесов r_m^2 эталонных образов равномерно распределены на интервале $[0, 2/M]$.

В случае, когда статвеса образов случайны и равномерно распределены на отрезке $r_m \in [0, \sqrt{3/M}]$ со средним $\overline{r_m^2} = 1/M$, зависимость объема памяти от числа образов M следующим выражением:

$$\overline{M} = M(1 - r_0 \sqrt{M/3}). \quad (76)$$

Максимальный объем памяти $\overline{M} = 2N/9 \ln N$ в этом случае достигается при числе образов $M = 2N/3 \ln N$, а при $M \geq 3N/2 \ln N$ объем памяти сети обращается в ноль.

В другом предельном случае, когда все статвеса равны друг другу ($r_m = 1/\sqrt{M}$), условие существования локальных минимумов $r_m > r_0$ преобразуется в известное [5] ограничение $M < N/2 \ln N$ на емкость памяти стандартной модели Хопфилда. В этом случае понятие "емкость памяти" совпадает с числом образов M , поскольку либо распознаются все прописанные в матрице межсвязей образы, либо не распознается ни один из них. Соответственно, максимальный объем памяти определяется выражением $\overline{M} = N/2 \ln N$.

В заключение этого раздела отметим, что удаление из матрицы межсвязей образов, прописанных в матрице межсвязей с статвесами, меньшими критического значения ($r_m < r_0$), не влияет на размер объема памяти, но улучшает ее распознающие характеристики – размер областей притяжения оставшихся образов увеличится и, следовательно, сеть сможет распознавать более сильно искаженные образы.

5. Энергия локального минимума

Из анализа выражения (2) следует, что энергию локального минимума E_m с точностью до незначительной флуктуации порядка

$$\sigma_m = N\sqrt{1 - r_m^2} \quad (8)$$

можно представить в виде

$$E_m = -r_m N^2. \quad (9)$$

Тогда, с учетом соотношения (5) из (9) нетрудно получить выражение

$$E_m = E_0 \frac{1}{\sqrt{(1 - 2n_m/N)^2 + E_0^2/N^4}}, \quad (10)$$

где

$$E_0 = -N\sqrt{2N \ln N}, \quad (11)$$

устанавливающее связь между глубиной локального минимума и шириной его области притяжения. Как видим, чем шире область притяжения, тем глубже минимум, и наоборот – чем глубже минимум, тем шире область его притяжения (см. рис.3).

Введенная здесь величина E_0 характеризует сразу два параметра нейронной сети. Во-первых, ею определяется полуширина распределения (10). Во-вторых, из (10) вытекает неравенство $E_m \leq E_0$, т.е. E_0 является верхней границей спектра локальных минимумов. Эти результаты хорошо согласуются с ре-

зультатами компьютерных экспериментов, в ходе которых проверялось, существует ли в точке S_m локальный минимум или его нет.

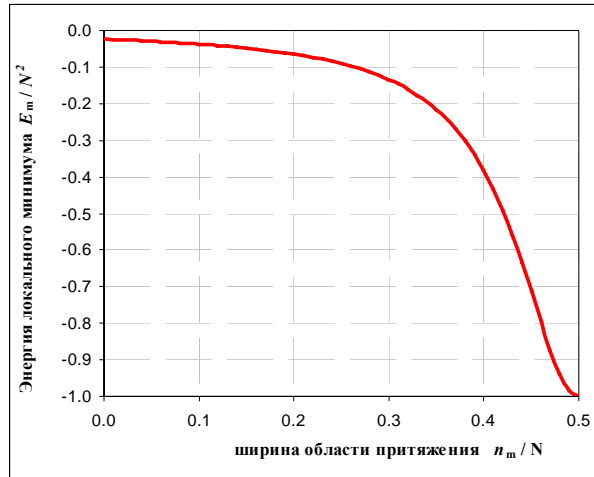


Рис.3. Зависимость энергии локального минимума от ширины области притяжения.

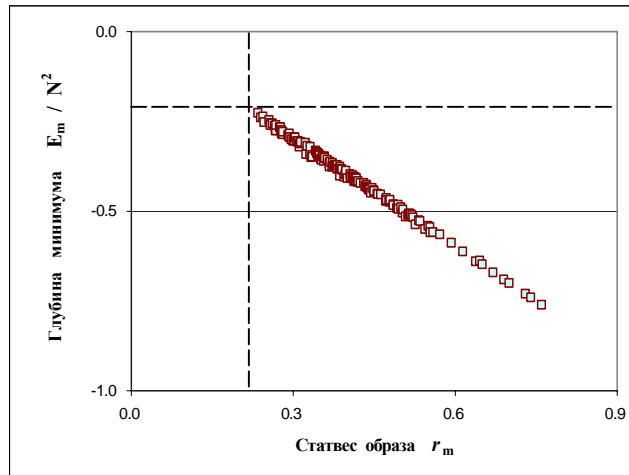


Рис. 4. Зависимость энергии E_m локального минимума от статвеса образа r_m . $N=500$, $M=25$.

Результаты одного из экспериментов ($N=500$, $M=25$) приведены на рис.4. Как видим, хорошо прослеживается линейная зависимость энергии локального минимума от величины статвеса образа. Причем, экспериментальные точки, соответствующие локальным минимумам, расположены только в правом нижнем квадранте, где $r_m > r_0$ и $E_m < E_0$. Кроме того, на рис.4 видно, что в соответствии с (8) флуктуации минимумов энергии с ростом статвеса уменьшаются.

6. Вероятность обнаружения минимума

Определим теперь вероятность W нахождения локального минимума E_m при случайном поиске. Искомая вероятность по определению совпадает с вероятностью того, что, задавая начальную конфигурацию, мы попадаем в область притяжения образа S_m . Следовательно, величина $W = W(n_m)$ есть число точек в сфере радиуса n_m , приведенное к общему числу точек в N -мерном пространстве:

$$W = 2^{-N} \sum_{n=1}^{n_m} C_N^n. \quad (12)$$

Выражениями (8) и (12) в неявном виде задается связь между глубиной локального минимума и вероятностью его нахождения. Применяя к биномиальным коэффициентам асимптотическое разложение Стирлинга и заменяя в (12) суммирование интегрированием, можно представить искомую связь как

$$W = W_0 e^{-Nh}, \quad (13)$$

где h с точностью до постоянной совпадает с функцией Шеннона:

$$h = \frac{n_m}{N} \ln \frac{n_m}{N} + \left(1 - \frac{n_m}{N}\right) \ln \left(1 - \frac{n_m}{N}\right) + \ln 2. \quad (14)$$

Здесь W_0 – не существенная для дальнейшего анализа медленная функция от E_m , которую можно получить асимптотической оценкой (12) при условии $n_m \gg 1$, а зависимость $W = W(n_m)$ всецело определяется быстрой экспонентой.

Как следует из (12), вероятность обнаружения локального минимума малой глубины ($E_m \sim E_0$) убывающе мала как $W \sim 2^{-N}$. Заметно отличной от нуля вероятность W становится только в случае достаточно глубоких минимумов $|E_m| \gg |E_0|$, размеры областей притяжения которых сопоставимы с величиной $N/2$. Выражение (13) с учетом (10) в этом случае преобразуется к зависимости $W = W(E_m)$, задаваемой выражением

$$W = W_0 \exp \left[-NE_0^2 \left(\frac{1}{E_m^2} - \frac{1}{E_\infty^2} \right) \right], \quad (15)$$

где

$$E_\infty^2 = \sum E_m^2. \quad (16)$$

Здесь мы ввели новую константу E_∞ , хотя в принятой нами нормировке из (9) вытекает $E_\infty = N^2$ и введение нового обозначения не имеет особого смысла, однако для других типов нормировок вполне возможна иная зависимость E_∞ от N .

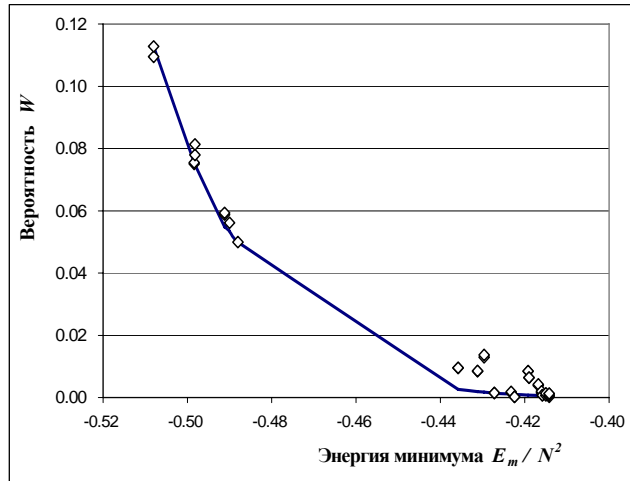


Рис. 5. Зависимость вероятности W обнаружения локального минимума от его глубины E_m .

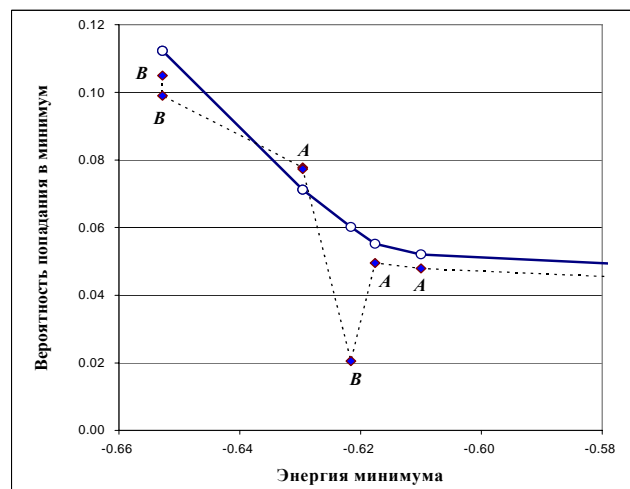


Рис.6. Сравнение предсказанных вероятностей (сплошная линия) и экспериментально найденных значений (точки, связанные пунктиром).

Как видим, вероятность обнаружения минимума растет с ростом его глубины. Эта зависимость подтверждается результатами компьютерных экспериментов, проведенных для матриц (1) с малым загрузочным параметром $M/N \leq 0.05$, при котором проведенный выше анализ справедлив. Сплошная кривая на рис.5 построена по формуле (13), точки – эксперимент. Как видим, хорошее согласие достигается прежде всего для наиболее глубоких локальных минимумов, которые соответствуют записанным в матрице межсвязей образам S_m (на рисунке – это область энергий $E_m < -0.48N^2$). Экспериментально обна-

руженные минимумы малой глубины (точки в области $E_m > -0.44N^2$) – это так называемые "химеры" [1]. В стандартной модели Хопфилда ($r_m \equiv 1/\sqrt{M}$) они появляются при относительно больших загрузках $M/N > 0.05$. В рассматриваемом здесь более общем случае они могут появиться и раньше. Причины их появления хорошо исследованы методами статистической физики в работе [9], где показано, что химеры образуются как следствие интерференции минимумов S_m . При малой загрузке химеры отделены от минимумов S_m энергетической щелью, отчетливо видной на рис.5. С ростом параметра загрузки M/N соответствующие химерам минимумы опускаются и, при загрузке $M/N \sim 0.1$, сравниваются по глубине с локальными минимумами S_m . Более того, при $M/N \geq 0.14$ появляются глубоко расположенные так называемые ложные (spurious) минимумы, образованные из больших фрагментов векторов S_m , а большинство локальных минимумов S_m исчезает. Эта экспериментально наблюдавшаяся картина имеет простое объяснение в рамках проведенного выше анализа: с ростом M уменьшается число образов, удовлетворяющих условию $r_m > r_0$ существования локального минимума.

Несмотря на такую сложную картину энергетической поверхности, описываемая выражениями (13)-(15) зависимость "глубже минимум – больше область притяжения – больше вероятность попадания в этот минимум" как правило сохраняется при любых параметрах загрузки. Строгое обоснование этому феномену мы надеемся дать в последующих публикациях. Здесь же отметим, что зависимость (13) выполняется тем лучше, чем глубже описываемый этим выражением минимум. А поскольку для задачи оптимизации важно описать поведение именно наиболее глубоких минимумов, то это позволяет сформулировать эвристический подход для отыскания глобального минимума функционала (2) с произвольной наперед заданной матрицей (не обязательно хэббовской). Состоит он в том, чтобы использовать выражение (13) с неизвестными параметрами W_0 , E_0 и E_∞ . Для этого запускается процедура случайного поиска и находится какое-то число минимумов. По полученным данным определяются характерные для данной матрицы значения E_0 и E_∞ и величина подгоночного параметра W_0 . Подстановка этих величин в (13) позволит оценить вероятность нахождения неизвестного более глубокого минимума E_m (если он существует) и принять решение на остановку программы поиска (если оценка пессимистична) или ее продолжение.

Такой подход опробован на хэббовских матрицах при достаточно больших загрузочных параметрах ($M/N = 0.2 \div 2$), когда нейросетевая система переходит в состояние спинового стекла [9] и применение полученных выше результатов становится не совсем корректным. Результат одного из экспериментов приведен на рис.6. В ходе эксперимента по обнаруженным минимумам (точки А) были рассчитаны параметры W_0 , E_0 и E_∞ и построена зависимость $W = W(E_m)$ (сплошная кривая). Затем, в результате многократно повторяемой процедуры случайного поиска ($\sim 10^5$ стартов нейронной сети) были установле-

ны другие минимумы (точки B) и оценочные значения вероятностей попадания в них. Как видим, несмотря на разброс, предсказанные значения по порядку величины хорошо совпадают с реальными значениями вероятностей.

7. Обсуждение результатов

Проведенный анализ показал, что свойства обобщенной модели описываются двумя критическими параметрами r_0 и E_0 . Первый из них определяет минимальное значение статвеса, при котором образ образует локальный минимум и может быть восстановлен нейронной сетью. Второй – это верхняя граница локальных минимумов энергетической поверхности. Существенно, что величины обоих параметров никак не зависят от числа образов M , на которых построена матрица межсвязей. Существенно и то, что удаление из матрицы межсвязей образов, прописанных в матрице межсвязей с статвесами, меньшими критического значения ($r_m < r_0$), хотя и не влияет на размер объема памяти, однако улучшает ее распознающие характеристики – размер областей притяжения оставшихся образов увеличится и, следовательно, сеть может распознавать более сильно искаженные образы. Вместе с этим, обратим внимание на тот факт, что при постоянном параметре загрузки $M/N = const$, с ростом размерности сети радиус области притяжения локального минимума уменьшается. Как следует из (5), такое уменьшение имеет место даже в том случае, когда загрузка сети с ростом N логарифмически уменьшается как $M/N = const/2 \ln N$. Это означает, что с ростом N уменьшается уровень допустимых искажений, при которых сеть способна восстанавливать образы (сеть меньших размеров способна распознавать более сильно искаженные образы).

Во избежание недоразумений следует указать на то, что связь между радиусом области притяжения n_m и объемом этой области $W = W(n_m)$ существенно нелинейная (12). Это означает, что даже при очень большом значении радиуса (например $n_m \sim 0.49N$) область притяжения может занимать очень малую часть N -мерного конфигурационного пространства, а доля пространства вне области притяжения (т.е. все точки, отстоящие от минимума на расстоянии от n_m до $0.5N$) с ростом N нарастает экспоненциально и при достаточно больших значениях N стремится к $1/2$.

Действительно, рассмотрим простейший пример – случай стандартной модели Хопфилда ($r_m^2 = M^{-1}$). На рис. 7 показана зависимость радиуса области притяжения от размерности сети при разных загрузках. Как видно из рис.7а, в случае постоянной загрузки ($M/N = const$) радиус притяжения с ростом N быстро стремится к нулю. Это соответствует выражению (5), из которого следует, что радиус области притяжения отличен от нуля только при небольших размерах сети, когда $2k \ln N \leq 1$. Очевидно, что доля пространства, занимаемая всеми M локальными минимумами, в этом случае с ростом N стремится к нулю. В приведенном на рис.7б случае логарифмически уменьшающейся загрузки ($M/N = const/2 \ln N$) радиус области притяжения с ростом N достаточно быстро уменьшается до некоторого (отличного от нуля) асимптотического значе-

ния $n_m \sim 0.5N(1 - \sqrt{k})$. Тем не менее, доля пространства, занимаемая всеми M локальными минимумами (величина, равная $M*W$), и в этом случае с ростом N стремится к нулю. Это означает, что с ростом N практически все пространство отходит к ложным минимумам и вероятность нахождения глубоких минимумов при случайном поиске также стремится к нулю.

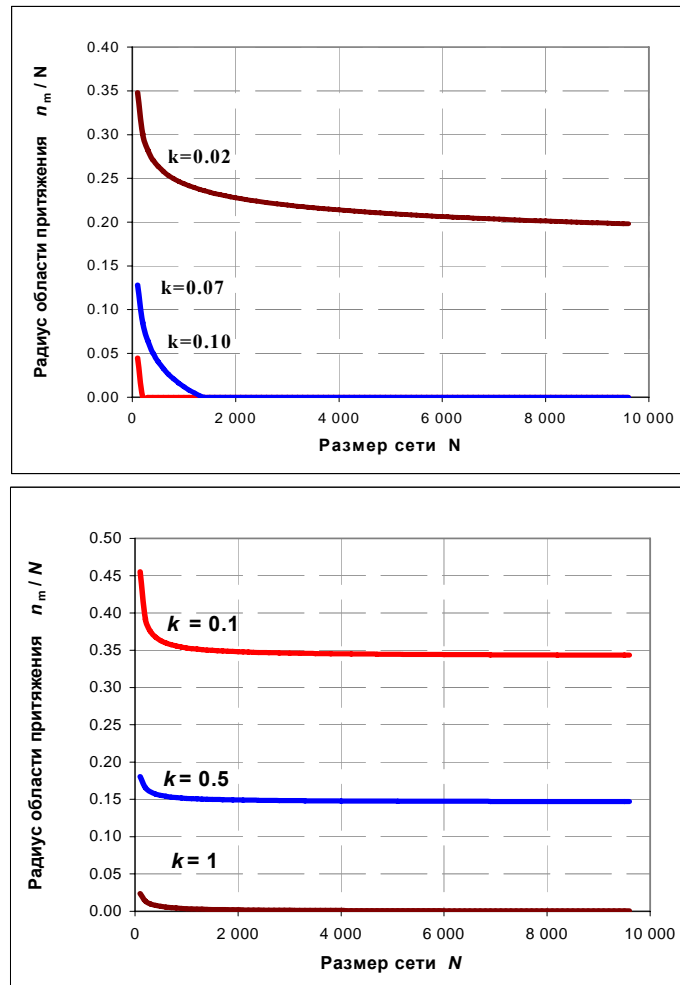


Рис.7. Размер радиуса области притяжения локального минимума в стандартной модели Хопфилда, как функция размерности пространства N для различных значений загрузки сети: а) постоянная загрузка $M = kN$ при $k = 0.02, 0.07, 0.1$; б) логарифмически уменьшающаяся загрузка $M = kN / 2 \ln N$ при $k = 0.1, 0.5, 1$.

На рис.8 для наглядности сделанного утверждения приводится гипотетический пример, когда имеется M минимумов, причем радиус области притяже-

ния каждого из них очень велик и равен $n_m = kN$, где $k = 0.47, 0.48, 0.49$. Как видим, даже в этом случае, несмотря на нереально большой радиус областей притяжения, доля пространства, занимаемая этими минимумами, с ростом N стремится к нулю.

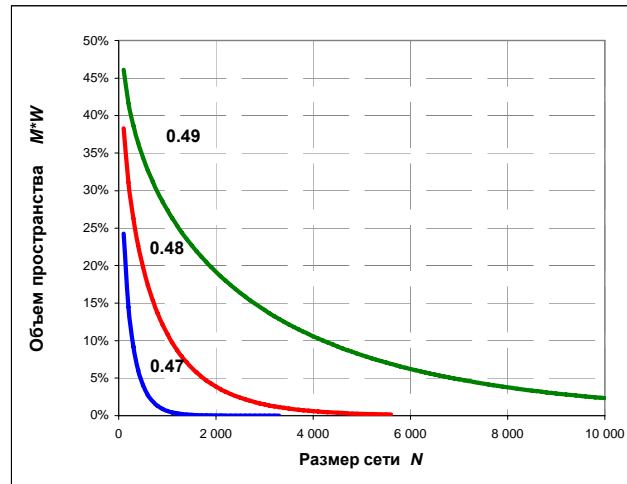


Рис.8. На рисунке показано уменьшение доли пространства, занимаемого M локальными минимумами в случае, когда радиус притяжения каждого из минимумов равен kN , при $k=0.47, 0.48, 0.49$.

В заключение сделаем несколько практических замечаний, относящихся к сравнению аналитической теории с компьютерным экспериментом. Во-первых, напомним, что картина энергетического ландшафта симметрична относительно изменения знака вектора состояния нейронной сети и, следовательно, каждому локальному минимуму в точке \mathbf{S} соответствует точно такой же минимум в точке $\mathbf{S}^* = -\mathbf{S}$. Поэтому, экспериментально измеренную вероятность нахождения минимума с энергией E следует сравнивать с величиной $W = W(E)$, если только в ходе эксперимента фиксировалась координата минимума, в противном случае – с величиной $2W(E)$. Во-вторых, еще раз подчеркнем, что зависимость "глубже минимум – больше область притяжения – больше вероятность попадания в этот минимум" строго обоснована только в случае корреляционных (хеббовских) матриц. В более общем случае это не зависимость, а скорее экспериментально установленная тенденция, применимая к наиболее глубоким минимумам. Обоснование этой тенденции заключается в том, что любую наперед заданную матрицу можно представить в виде хэббовской матрицы (1), построенной на произвольном числе образов (например $M \rightarrow \infty$) с произвольными статвесами. И если спектр минимумов функционала, построенного на такой матрице, имеет ярко выраженные минимумы большой глубины, то применительно к ним справедливы сделанные ранее выводы. Трудность наступает в случае, когда таких ярко выраженных минимумов нет и спектр минимумов ква-

зинепрерывен. В этом случае в оптимизационном эксперименте измеряется не вероятность $W(E)$, а частота попадания в минимум с заданной энергией – произведение вероятности на спектральную плотность, характеристики которой неизвестны. Соответственно, затрудняется применение развитой выше теории.

8. Приложение

Используем так называемое одношаговое приближение (one-step approximation) [5]. Нейронная сеть из состояния \mathbf{S} сойдется в точку \mathbf{S}_m , если для любого $i \in [1, N]$ направление локального поля h_i совпадает с направлением i -го спина в эталоне, т.е. $h_i s_i^{(m)} \geq 0$. Вероятность этого события описывается выражением:

$$Pr = \prod_{i=1}^N Pr(h_i s_i^{(m)} \geq 0)$$

В асимптотическом пределе $N \rightarrow \infty$ величину $h_i s_i^{(m)}$ можно рассматривать как случайную величину со средним $\xi = r_m N(1 - 2n/N)$ и дисперсией $\sigma_\xi^2 = N(1 - r_m^2)$. Соответственно, вероятность события $h_i s_i^{(m)} \geq 0$ равна $Pr(h_i s_i^{(m)} \geq 0) = \frac{1}{2}(1 + \operatorname{erf} \gamma)$, где $\gamma = \xi / \sigma_\xi$. Возводя эту величину в N степень, получим выражение (3).

Литература

1. J.J.Hopfield. Neural networks and physical systems with emergent collective computational abilities // Proc. Nat. Acad. Sci. USA. 79, 2554-2558 (1982).
2. J.J.Hopfield, D.W.Tank. Neural computation of decisions in optimization problems // Biological Cybernetics 52, 141-152 (1985); J.J.Hopfield, D.W.Tank. Computing with neural circuits: A model. Science 233, 625-633 (1986).
3. Y.Fu, P.W.Anderson. Application of statistical mechanics to NP-complete problems in combinatorial optimization // Journal of Physics A. 19, 1605-1620 (1986).
4. T.Poggio, F.Girosi. Regularization algorithms for learning that are equivalent to multi-layer networks // Science 247, 978-982 (1990).
5. R.McEllice, E.Posner, E.R.Rodemich. Capacity of the Hopfield associative memory // IEEE Trans. on Inf. Theory. 33, 461-482 (1987).
6. Б.В.Крыжановский, Б.М. Магомедов, А.Л.Микаэлян. Доменная модель нейронной сети // Доклады АН 401, 462-466 (2005).
7. В.Kryzhanovsky, В.Magomedov. Application of domain neural network to optimization tasks // ICANN 2005. Poland. LNCS 3697, Part II, pp.397-403.
8. D.O.Hebb The Organization of Behavior. – New York: Wiley, 1949.
9. D.J.Amit, H.Gutfreund and H.Sompolinsky Statistical mechanics of neural networks near saturation //Annals of Physics 173, 30-67 (1987).
10. Б.В.Крыжановский, Б.М.Магомедов, А.Л.Микаэлян. Взаимосвязь глубины локального минимума и вероятности его обнаружения в обобщенной модели Хопфилда // Доклады АН 405, №3, с.1-5 (2005).