

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
МИНИСТЕРСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ ПО АТОМНОЙ ЭНЕРГИИ
МИНИСТЕРСТВО ПРОМЫШЛЕННОСТИ, НАУКИ И ТЕХНОЛОГИЙ
РОССИЙСКОЙ ФЕДЕРАЦИИ
РОССИЙСКАЯ АССОЦИАЦИЯ НЕЙРОИНФОРМАТИКИ
МОСКОВСКИЙ ИНЖЕНЕРНО-ФИЗИЧЕСКИЙ ИНСТИТУТ
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ)

НАУЧНАЯ СЕССИЯ МИФИ–2003

НЕЙРОИНФОРМАТИКА–2003

**V ВСЕРОССИЙСКАЯ
НАУЧНО-ТЕХНИЧЕСКАЯ
КОНФЕРЕНЦИЯ**

**ЛЕКЦИИ
ПО НЕЙРОИНФОРМАТИКЕ**

Часть 1

По материалам Школы-семинара
«Современные проблемы нейроинформатики»

Москва 2003

УДК 004.032.26 (06)

ББК 32.818я5

М82

НАУЧНАЯ СЕССИЯ МИФИ–2003. V ВСЕРОССИЙСКАЯ НАУЧНО-ТЕХНИЧЕСКАЯ КОНФЕРЕНЦИЯ «НЕЙРОИНФОРМАТИКА–2003»: ЛЕКЦИИ ПО НЕЙРОИНФОРМАТИКЕ. Часть 1. – М.: МИФИ, 2003. – 188 с.

В книге публикуются тексты лекций, прочитанных на Школе-семинаре «Современные проблемы нейроинформатики», проходившей 29–31 января 2003 года в МИФИ в рамках V Всероссийской конференции «Нейроинформатика–2003».

Материалы лекций связаны с рядом проблем, актуальных для современного этапа развития нейроинформатики, включая ее взаимодействие с другими научно-техническими областями.

Ответственный редактор

Ю. В. Тюменцев, кандидат технических наук

ISBN 5–7262–0471–9

© *Московский инженерно-физический институт
(государственный университет), 2003*

Содержание

А. А. Фролов, Д. Гусек, И. П. Муравьев. Информационная эффективность ассоциативной памяти типа Хопфилда с разреженным кодированием	28
Введение	29
Описание модели	34
Информация, извлекаемая из сети за счет коррекции искаженных эталонов	39
Аналитические подходы	43
Одношаговое приближение	44
Статистическая нейродинамика	47
Компьютерное моделирование	48
Плотное кодирование	48
Разреженное кодирование	56
Информационная эффективность	64
Заключение	68
Литература	69

А. А. ФРОЛОВ¹⁾, Д. ГУСЕК²⁾, И. П. МУРАВЬЕВ¹⁾

¹⁾ Институт высшей нервной деятельности и нейрофизиологии РАН,
Москва,

²⁾ Институт информатики Академии наук Чешской республики,
Прага

E-mail: aafrolov@mail.ru

ИНФОРМАЦИОННАЯ ЭФФЕКТИВНОСТЬ АССОЦИАТИВНОЙ ПАМЯТИ ТИПА ХОПФИЛДА С РАЗРЕЖЕННЫМ КОДИРОВАНИЕМ

Аннотация

Аттракторная нейронная сеть хопфилдового типа исследуется аналитически и путем компьютерного моделирования. Вычисляются информационная емкость, качество воспроизведения и размер областей притяжения. Выявляются асимптотические свойства нейронных сетей путем компьютерного моделирования для больших размеров сетей (до 15000 нейронов). Показано, что существующие аналитические подходы недостаточны для случая большой разреженности. Основная причина их недостатков заключается в игнорировании различия в поведении нейронов, активных и неактивных на предшествующих тактах процесса воспроизведения. Показано, что размер областей притяжения меняется немонотонно при возрастании разреженности: вначале он возрастает, а потом — убывает. Поэтому в пределе при $p \rightarrow 0$ разреженность ухудшает способность сети к коррекции эталонов, в то же время повышая информационную емкость и качество воспроизведения. Информационная эффективность сети, достигаемая за счет коррекции искаженных эталонов, используется как обобщенная характеристика способности сети выполнять функцию автоассоциативной памяти. Существует оптимальная разреженность, для которой информационная эффективность максимальна. Оптимальная разреженность оказывается близкой к уровню нервной активности мозга.

A. FROLOV¹⁾, **D. HUSEK**²⁾, **I. MURAVIEV**¹⁾

¹⁾Institute of Higher Nervous Activity and Neurophysiology
of the Russian Academy of Sciences, Moscow

²⁾Institute of Computer Science,
Academy of Sciences of the Czech Republic, Prague

E-mail: aafrolov@mail.ru

INFORMATIONAL EFFICIENCY OF SPARSELY ENCODED HOPFIELD-LIKE ASSOCIATIVE MEMORY

Abstract

A sparsely encoded Hopfield-like attractor neural network is investigated analytically and by computer simulation. Informational capacity, recall quality and attractor basin sizes are evaluated. Asymptotic properties of the neural network are revealed by computer simulation for a large network size (up to 15000 of neurons). This shows that all existing analytical approaches fail in the case of a large sparseness. The main reason of their invalidity is that they ignore the difference in behavior of previously active and nonactive neurons during the recall process. It is shown that the size of attraction basins changes nonmonotonically while sparseness increases: it increases initially and then decreases. Thus at the limit case for $p \rightarrow 0$ sparseness worsens the network ability to correct distorted prototypes while it improves both informational capacity and recall quality. The gain of information provided by the network due to correction of the distorted prototypes is used as cumulative index of the network ability to perform the functions of autoassociative memory. There exists an optimal sparseness for which the gain is maximal. Optimal sparseness obtained is corresponded to brain neural activity.

Введение

Исследуемая нейронная сеть хопфилдового типа является полностью связанной сетью, которая действует как автоассоциативная память для статистически независимых бинарных паттернов на основе корреляционного правила обучения Хебба. Кодирование называется *разреженным*, если количество активных нейронов n в записанных паттернах (эталонах) мало по сравнению с общим количеством нейронов в сети N . Типичные вопросы, возникающие при исследовании свойств сетей такого типа:

1. Как много эталонов может быть записано и воспроизведено в таких сетях? Это проблема *информационной емкости*.
2. Какова доля ошибок в финальных (выходных) паттернах по сравнению с воспроизводимыми эталонами? Это проблема *качества воспроизведения*.
3. Как сильно эталоны могут быть искажены при сохранении свойства воспроизводимости? Это проблема *размера областей притяжения устойчивых состояний* (аттракторов) в окрестности эталонов.

Важный теоретический вклад в оценку емкости сети и качества воспроизведения сетей хопфилдового типа с разреженным кодированием сделан методами статистической физики (метод реплик) (см. [2, 6, 10, 15, 24, 26] и др.). Было показано, что разреженность приводит к возрастанию емкости сети, причем не только если она определяется количеством запомненных эталонов (что является слабым результатом, потому что энтропия каждого эталона уменьшается при разрежении), но это верно даже если рассматривается *информационная емкость*, которая определяется общей *энтропией* записанных эталонов. Например, в предельном случае, когда $p = n/N \rightarrow 0$, эта энтропия была оценена в работе [2] как $(N^2/2) \log_2 e \simeq 0.72N^2$, в то время, как для оригинальной сети Хопфилда с плотным кодированием ($p = 0.5$) эта энтропия составляет только около $0.14N^2$ [5, 14].

Качество воспроизведения сетей хопфилдового типа обычно измеряется коэффициентом корреляции между эталонами и стабильными паттернами в их окрестностях. Эти паттерны являются аттракторами сетевой нейродинамики и рассматриваются как выходные (финальные) паттерны в процедуре декодирования (воспроизведения). Хорошо известно, что при относительно малом числе записанных эталонов (малой информационной нагрузке) качество воспроизведения сетей хопфилдового типа довольно высоко (корреляция между выходными и записанными паттернами близка к 1), но качество воспроизведения резко падает, когда информационная нагрузка близка к насыщению. Для оригинальной сети Хопфилда с плотным кодированием оценка корреляции между выходными паттернами и эталонами при максимальной нагрузке, сделанная методом реплик, дает 0.968 [5]. Ее оценка, сделанная тем же методом в работе [10] для сети с разреженным кодированием, показывает, что качество воспроизведения для сети с разреженным кодированием меняется немонотонно при возрастании разреженности. Вначале оно убывает, и коэффициент корреляции между финальными паттернами и эталонами при максимальной нагрузке достигает 0.82 для

$p \simeq 0.02$, а потом возрастает и стремится к 1, когда $p \rightarrow 0$. Таким образом, в пределе высокой разреженности качество воспроизведения и информационная емкость выше, чем в случае плотного кодирования. Однако, при увеличении разреженности они возрастают очень медленно. Например, качество воспроизведения при максимальной нагрузке достигает величины 0.9 только при p около 10^{-5} и величины 0.95 только при p около 10^{-44} .

Почти ничего не известно о влиянии разреженности на размер областей притяжения вокруг эталонов, что является основной характеристикой способности сети к коррекции эталонов после их частичного искажения. Анализ этого размера требует исследования динамики сети, что довольно трудно проделать точно. Трудности возникают из-за наличия корреляции между историей активности сети и матрицей связей. Исследование первых нескольких шагов по времени было проведено *Gardner et al.* [12] для сети с плотным кодированием и параллельной (синхронной) динамикой. Однако, сложность их расчетов экспоненциально возрастает в зависимости от числа рассматриваемых временных шагов, поэтому практически приходится ограничиваться малым числом шагов. Исследование этой ранней стадии нейродинамики часто оказывается недостаточным для заключения о размере областей притяжения, так как начальное возрастание корреляции между текущим паттерном активности сети и воспроизводимым эталоном может наблюдаться даже вне области притяжения (см. [4] и многие другие). С другой стороны, начальное убывание корреляции может наблюдаться внутри области притяжения (см. раздел «Компьютерное моделирование» настоящей статьи, с. 48–64). *Amari & Maginu* [4] предложили теорию статистической нейродинамики (Statistical Neurodynamics, SN) с целью исследования долговременного поведения активности сети с параллельной динамикой. Они вывели систему рекурсивных уравнений для двух макропараметров: 1) корреляции между текущим паттерном нейронной активности и воспроизводимым эталоном и 2) дисперсии шума. Их оценки информационной емкости и качества воспроизведения для сети Хопфилда с плотным кодированием оказались очень близки к оценкам, полученным методом реплик, а их оценки размера областей притяжения качественно совпадают с результатами компьютерного моделирования. *Okada* [22] обобщил SN-метод *Amari–Maginu* и вывел рекурсивные уравнения для более чем двух макропараметров. Он показал, что возрастание порядка системы рекурсивных уравнений делает оценки информационной емкости все более близкими к получаемым методом реплик. *Coolen & Sherington* [7] предложили более изощренный и точный теоретический подход к анализу параметров сети

в случае асинхронной (последовательной) динамики, свободной от предположения, сделанного в теориях *Amari-Maginu* и *Okada*, что распределение синаптических возбуждений нейронов является гауссовским. Недавно *Koyama et al.* [21] предложили разумное приближение к теории, построенной в работе [12], и вывели нейродинамические уравнения для полного процесса воспроизведения с параллельной динамикой. Как и указанная теория, этот метод дает достаточно точную оценку информационных и динамических свойств сети Хопфилда с неразрезанным кодированием.

Имеются лишь несколько статей, в которых исследуется размер областей притяжения в сетях хопфилдового типа с разреженным кодированием. *Horner et al.* [16] исследовали их для асинхронной, а *Okada* [22] — для синхронной динамики. Метод, предложенный в работе [16], основан на интерполяции между первыми шагами воспроизведения (когда может быть сделано разложение по степеням шага воспроизведения) и последними шагами воспроизведения (когда может использоваться стационарное решение). *Okada* использовал SN-приближение, близкое к предложенному *Amari-Maginu* для плотного кодирования. К сожалению, влияние разреженности на размер области притяжения систематически в этих статьях не анализировалось.

Обобщенной характеристикой способности сети выполнять функцию автоассоциативной памяти является информация I_g , извлекаемая из сети за счет коррекции искаженных записанных эталонов. Следуя *Дунину-Барковскому* [1], относительную информацию, извлекаемую из сети $E = I_g/N^2$, где N^2 — общее количество элементов, способных к модификации (для хопфилдовой сети — число синаптических связей), мы называем *информационной эффективностью*. Информация, извлеченная из сети, не может, очевидно, превысить информацию, загруженную в сеть $I_l = LNh(p)$, где L — количество записанных эталонов, и $h(p) = -p \log_2 p - (1-p) \log_2 (1-p)$ — функция Шеннона. Как относительную информационную нагрузку мы используем $\alpha = I_l/N^2 = Lh(p)/N$. Эта мера информационной нагрузки учитывает, что возрастание разреженности приводит к убыванию энтропии эталонов, в отличие от параметра α , определенного как L/N , который часто используется в качестве меры нагрузки сети, начиная с появления статьи Хопфилда [14]. В случае плотного кодирования обе эти меры совпадают, так как $h(0.5) = 1$. Для разреженного кодирования мы предпочитаем учитывать энтропию эталонов при расчете информационной нагрузки по двум причинам. Во-первых, в этом случае относительная информационная нагрузка измеряется в тех же единицах (бит/синапс), что и

информационная эффективность, и поэтому эти две величины легко сравнивать. Во-вторых, информационная емкость, измеряемая в битах на синапс, стремится к конечной величине $(\log_2 e)/2$ [10], когда разреженность возрастает, в то время как емкость, определенная как L/N , расходится как $1/|p \log_2 p|$.

Информационную эффективность можно сравнить с ее верхним пределом, задаваемым энтропией матрицы связей. Для полносвязной нейронной сети, обучаемой по правилу Хебба, верхний предел информационной эффективности был оценен как $E_{lim} = (\log_2 e)/4 \simeq 0.36$ [11, 23]. Для плотного кодирования максимальная информационная эффективность значительно меньше и составляет только 0.092 [16]. Однако, она возрастает при возрастании разреженности и, например, для $p = 0.1$ оценивается величиной 0.143 [16].

Основная цель настоящей лекции заключается в оценке информационной эффективности сети хопфилдового типа при разреженном кодировании. Она требует оценки всех трех упомянутых выше характеристик эффективности сети: информационной емкости, качества воспроизведения и размера областей притяжения. В статье [10] мы исследовали только информационную емкость и качество воспроизведения. В настоящей лекции мы исследуем также размер областей притяжения. Исследование производилось аналитически и путем компьютерного моделирования. В качестве аналитической техники мы использовали статистическую нейродинамику [4] и одношаговое (Single Step, SS) приближение [18]. Система рекурсивных уравнений, используемая здесь для оценок размеров областей притяжения, была выведена в работах [9] и [10]. К сожалению, мы не могли сравнить ее с системой уравнений, независимо выведенных *Shigemitsu & Okada* [25], так как их статья опубликована на японском языке. На основе результатов, полученных для плотного кодирования, может возникнуть сомнение в точности SS-приближения [22], которое игнорирует корреляции между историей активности сети и матрицей связей. Однако, как показано в работе [10], различие между результатами, полученными *методом реплик* (RM), SN и SS-приближениями убывает при возрастании разреженности. Более того, результаты компьютерного моделирования (см. раздел «Компьютерное моделирование», с. 48–64) показывают, что при разреженном кодировании результаты SS-приближения дают даже лучшее согласие с результатами компьютерного моделирования, чем методы RM и SN. Так как результаты, полученные с помощью SS-приближения, легко интерпретируются и справедливы по крайней мере для первых шагов нейродинамики,

мы сохранили в наших исследованиях эту «наивную» [22] технику, как один из используемых аналитических подходов.

Большинство результатов, представленных в этой лекции, получено путем компьютерного моделирования. Результаты показывают, что все существующие пока аналитические подходы дают только качественную оценку параметров сети хопфилдового типа с разреженным кодированием и параллельной динамикой. Основной недостаток существующих аналитических методов заключается в игнорировании различия в поведении нейронов, активных и неактивных в предшествующих состояниях динамики. Различие между результатами компьютерного моделирования и аналитическими становится очевидным только при достаточно больших размерах сети. Например, для сети Хопфилда с плотным кодированием количество нейронов в моделируемой сети должно превышать, по крайней мере, несколько тысяч (в моделировании, проведенном в работах [19] и [20] оно достигало 10^5). Для моделирования таких больших сетей Коринг [19] разработал специальный алгоритм, позволяющий обойтись без записи матрицы связей в компьютерную память. Мы применили алгоритм Коринга для сетей с разреженным кодированием и дополнительно модифицировали его, избежав необходимости записи в компьютерную память полного набора запомненных эталонов. Поэтому, для моделирования сети большого размера мы были ограничены только временем счета. В наших компьютерных экспериментах размер сети достигал 15000 нейронов.

В последующих разделах лекции мы опишем вначале модель и технику, используемую в компьютерном моделировании (с. 34–39), далее выведем основные уравнения для вычисления информационной эффективности (с. 39–43), затем рассмотрим аналитические подходы (с. 43–48). Результаты проведенного компьютерного моделирования сравниваются с аналитическими результатами (с. 48–64), после чего используются для оценки информационной эффективности (с. 64–68). Завершается лекция кратким обсуждением результатов (с. 68–69).

Описание модели

Как и в работе [10], мы использовали для обучения сети *корреляционное правило Хебба*, более эффективное, чем обычное правило Хебба [6, 26]. Правило Хебба называется корреляционным, если компоненты матрицы

связей определяются уравнением

$$J_{ij} = \frac{1}{Np(1-p)} \sum_{l=1,L} (X_i^l - p)(X_j^l - p), \quad i \neq j, \quad J_{ii} = 0, \quad (1)$$

где $X_i^l \in \{0, 1\}$ — компоненты эталонов (1 — для активного, 0 — для неактивного нейронов), $p = n/N$ — вероятность нейрона быть активным и L — количество записанных эталонов. В нашей модели эталоны предполагаются равномерно и независимо распределенными на множестве всех паттернов с $n = pN$ компонентами 1 и $(N - n)$ компонентами 0 (т. е. количество активных нейронов в каждом эталоне фиксировано). Мы показали в работе [10], что такой тип кодирования обеспечивает возрастание информационной емкости по сравнению с кодированием, когда количество активных нейронов в эталонах случайно и только их среднее количество равно n .

Мы ограничились анализом только случая параллельной динамики, когда паттерн нейронной активности \mathbf{X} в каждый момент времени t вычисляется как

$$X_i(t+1) = \Theta(\eta_i(t) - T(t)), \quad i = 1, \dots, N, \quad (2)$$

где

$$\eta_i(t) = \sum_j J_{ij} X_j(t) \quad (3)$$

представляет собой синаптическое возбуждение, Θ — пороговая функция Хевисайда и T — порог активации. Порог $T(t)$ выбирается в каждый момент времени так, чтобы количество активных нейронов было бы равно $n = pN$. Таким образом, в каждый такт времени активны только n нейронов, имеющих наибольшее синаптическое возбуждение. Поскольку количество активных нейронов в каждом эталоне фиксировано и тоже равно n , этот выбор порога активации способствует стабилизации активности сети в окрестности одного из эталонов. Следовательно, тип кодирования эталонов согласуется с используемой простой процедурой воспроизведения, которая позволяет обойтись без точного контроля порога активации. Это второе преимущество кодирования с фиксированным числом активных нейронов в эталонах. Для того, чтобы исключить конкуренцию между нейронами, имеющими одинаковые синаптические возбуждения, равные порогу активации, к возбуждениям добавлен шум χ_i , различный для всех

нейронов сети, величина которого столь мала, что не меняет распределения нейронов по величине синаптического возбуждения от активных нейронов сети, но достаточна, чтобы упорядочить их по величине суммарного возбуждения, если вклады от активных нейронов сети одинаковы. Для того, чтобы обеспечить стабилизацию активности сети, распределение шума χ_i по нейронам полагается фиксированным во все моменты времени. Тогда, как и в случае, когда порог активации фиксирован [13], в нейронной динамике существуют только два типа аттракторов – точки или циклы длины 2.

Для доказательства введем $\eta'_i(t) = \eta_i(t) + \chi_i$, где $\eta_i(t)$ – синаптические возбуждения от активных нейронов сети, а χ_i – синаптический шум. Тогда все значения $\eta'_i(t)$ различны, даже если некоторые значения $\eta_i(t)$ равны. Введем порог $T'(t)$ такой что n нейронов с $\eta'_i(t) > T'(t)$ активны и $N - n$ нейронов с $\eta'_i < T'$ – неактивны. В этом случае уравнение (2) заменяется на

$$X_i(t+1) = \Theta(u_i(t)), \quad (4)$$

где $u_i(t) = \eta'_i(t) - T'(t)$.

Введем функцию $F(t) = \mathbf{X}^T(t+1)\mathbf{J}\mathbf{X}(t) + [\mathbf{X}^T(t+1) + \mathbf{X}^T(t)]\chi$, где χ – вектор с компонентами χ_i . Поскольку число активных нейронов фиксировано на каждом временном шаге, $\mathbf{X}^T(t+2)\mathbf{e} = \mathbf{X}^T(t)\mathbf{e} = n$, где \mathbf{e} – вектор из N единиц. Следовательно,

$$\begin{aligned} \Delta &= F(t+1) - F(t) = [\mathbf{X}^T(t+2) - \mathbf{X}^T(t)] [\mathbf{J}\mathbf{X}(t+1) + \chi] = \\ &= [\mathbf{X}^T(t+2) - \mathbf{X}^T(t)] [\mathbf{J}\mathbf{X}(t+1) + \chi - T'(t+1)\mathbf{e}] = \sum_i \delta_i, \end{aligned} \quad (5)$$

где $\delta_i = [X_i(t+2) - X_i(t)]u_i(t+1)$. Величины δ_i неотрицательны, поскольку, если $u_i(t+1) > 0$, то, в соответствии с (4), $X_i(t+2) = 1$ и $X_i(t+2) - X_i(t) \geq 0$ независимо от $X_i(t)$. Если $u_i(t+1) < 0$, то согласно (4), $X_i(t+2) = 0$ и $X_i(t+2) - X_i(t) \leq 0$ независимо от $X_i(t)$. Следовательно, в обоих случаях $\delta_i \geq 0$ и, следовательно, $\Delta \geq 0$. Поскольку количество состояний сети конечно, через конечное число шагов Δ достигает нулевого значения, когда все $\delta_i = 0$. Так как в используемой процедуре восстановления $u_i \neq 0$ для всех i , это возможно только если $X_i(t+2) = X_i(t)$ для всех i . Таким образом, в процессе восстановления достигается неподвижная точка или цикл длины два. Выходным (финальным) паттерном \mathbf{X}^f процедуры воспроизведения полагается стабильное состояние сети или первый паттерн цикла длины 2.

Сходство между текущим паттерном нейронной активности $\mathbf{X}(t)$ и восстанавливаемым эталоном \mathbf{X}^l задается так называемым *пересечением* (overlap) этих паттернов:

$$m(\mathbf{X}^l, \mathbf{X}(t)) = \sum_{i=1, N} (X_i^l - p) X_i(t) / [Np(1-p)]. \quad (6)$$

Качество воспроизведения измеряется средним пересечением между восстанавливаемым эталоном и финальным паттерном $m_f = m(\mathbf{X}^l, \mathbf{X}^f)$. Так как средняя активность нейрона равна p , а ее дисперсия равна $p(1-p)$, то среднее пересечение фактически равно коэффициенту корреляции между паттерном нейронной активности и восстанавливаемым эталоном.

Для компьютерного моделирования сетей большого размера при ограничениях на компьютерную память мы используем процедуру, близкую к процедуре, предложенной в [19] для плотного кодирования. Эта процедура позволяет моделировать работу сети без записи матрицы связей. С использованием (1)–(3) можно легко выразить синаптическое возбуждение прямо через пересечения между текущими паттернами активности сети и записанными эталонами:

$$\begin{aligned} \eta_i(t) &= \frac{1}{Np(1-p)} \sum_{l=1, L} (X_i^l - p) \sum_{j \neq i} (X_j^l - p) X_j(t) = \\ &= \sum_{l=1, L} \left[(X_i^l - p) m(\mathbf{X}^l, \mathbf{X}(t)) - \frac{1}{Np(1-p)} (X_i^l - p)^2 X_i(t) \right] = \quad (7) \\ &= A_i - \frac{1-2p}{Np(1-p)} B_i - pC - \frac{LpX_i}{N(1-p)}, \end{aligned}$$

где

$$\begin{aligned} A_i &= \sum_{l=1, L} X_i^l m(\mathbf{X}^l, \mathbf{X}(t)), \\ B_i &= \sum_{l=1, L} X_i^l X_i(t), \\ C &= \sum_{l=1, L} m(\mathbf{X}^l, \mathbf{X}(t)). \end{aligned}$$

Для разреженного кодирования более эффективно представлять записанные эталоны как векторы длины n , компонентами которых являются

номера активных нейронов вместо векторов длины N с компонентами состояния (0 или 1) всех нейронов. Тогда более удобно представлять пересечения в виде

$$\begin{aligned} m(\mathbf{X}^l, \mathbf{X}(t)) &= \frac{1}{Np(1-p)} \left[\sum_{i=1, N} X_i^l X_i(t) - Np^2 \right] = \\ &= \frac{1}{Np(1-p)} \left[\sum_{j=1, n} X_{i(j)}(t) - Np^2 \right], \end{aligned}$$

где $i(j)$ — номер активного нейрона, задаваемый вектором, представляющим l -тый эталон. Поэтому требуется только n сложений для вычисления каждого пересечения и Ln сложений для вычисления всех пересечений. Аналогично, для вычисления всех N величин A_i и B_i требуется только $2Ln$ сложений и в целом требуется примерно $3Ln$ сложений для вычисления всех N величин синаптических возбуждений η_i . Заметим, что для обычной процедуры, основанной на прямом вычислении η_i согласно (3), необходимо nN сложений. Таким образом, для малых информационных нагрузок ($L < N/3$) модифицированная процедура имеет преимущество перед обычной даже по числу вычислений.

Описанная процедура позволяет обойтись без записи матрицы связей, но не набора эталонов. Однако компьютерная память, требуемая для записи матрицы связей и требуемая для записи набора эталонов, сравнимы. Фактически, для записи около $N^2/2$ независимых коэффициентов матрицы связей как действительных чисел требуется $2N^2$ байтов памяти (если положить, что для записи каждого числа требуется 4 байта). В то же время для записи Ln индексов активных нейронов в эталонах как переменных типа «слово» требуется $2Ln$ байтов (если положить, что для записи слова требуется 2 байта). Для разреженного кодирования мы имеем $L \simeq \alpha N / |p \log_2 p|$. Поэтому, память, требуемая для записи набора эталонов, по порядку только в $|\log_2 p|$ раз меньше, чем требуемая для записи матрицы связей. Чтобы обойтись без записи эталонов, мы использовали процедуру, в которой эталоны заново генерируются на каждом шаге воспроизведения, с помощью одной и той же последовательности псевдослучайных величин. Для генерации L эталонов требуется Ln псевдослучайных величин. Для этого требуется примерно то же компьютерное время, что и для вычисления синаптических возбуждений по уравнению (7). Таким образом, при моделировании работы сети можно обойтись не только без записи в компьютерную память

матрицы связей, но и без записи набора эталонов, оплачивая эту экономию памяти уменьшением скорости счета примерно в два раза.

Информация, извлекаемая из сети за счет коррекции искаженных эталонов

Для вычисления количества информации, извлекаемой из памяти за счет коррекции искаженных эталонов, мы применяем основные понятия теории информации. *Взаимная информация* случайных величин A и B вычисляется по формуле

$$I(A, B) = H(A) - H(A|B) = H(B) - H(B|A).$$

Взаимная условная информация случайных величин A и B при условии, что задана случайная величина C , вычисляется по формуле

$$I(A, B|C) = H(A|C) - H(A|B, C),$$

где

$$H(A) = - \sum_A \mathcal{P}\{A\} \log_2 \mathcal{P}\{A\}$$

есть энтропия случайной величины A и

$$H(A|B) = - \sum_{A,B} \mathcal{P}\{A, B\} \log_2 \mathcal{P}\{A|B\}$$

есть условная энтропия случайной величины A .

Информация, полученная из сети за счет коррекции входного паттерна, задается выражением

$$i_g = I(\mathbf{X}^l, \mathbf{X}^f | \mathbf{X}^{in}) = H(\mathbf{X}^l | \mathbf{X}^{in}) - H(\mathbf{X}^l | \mathbf{X}^{in}, \mathbf{X}^f), \quad (8)$$

где \mathbf{X}^{in} и \mathbf{X}^f — начальный и финальный паттерны в процессе воспроизведения эталона \mathbf{X}^l , соответственно. В (8) первый член дает величину информации, требуемой для нахождения эталона, когда известен лишь начальный паттерн (начальная неопределенность восстанавливаемого эталона), а второй — остаточную неопределенность, когда становится известным

также финальный паттерн. Их разность есть информация, извлеченная из сети в процессе воспроизведения.

Для нахождения $H(\mathbf{X}^{in}|\mathbf{X}^l)$ введем вероятности $q_\mu = \mathcal{P}\{X_i^{in} = 1|X_i^l = \mu\}$ того, что нейроны, принадлежащие ($\mu = 1$) и не принадлежащие ($\mu = 0$) восстанавливаемому эталону \mathbf{X}^l активны в начальном паттерне \mathbf{X}^{in} . Поскольку количество активных нейронов в начальном паттерне совпадает с количеством активных нейронов в эталоне, то $H(\mathbf{X}^l|\mathbf{X}^{in}) = H(\mathbf{X}^{in}|\mathbf{X}^l) = Nh_{in}$, где

$$h_{in} = ph(q_1) + (1-p)h(q_0). \quad (9)$$

Два члена в (9) дают количество информации (в пересчете на один нейрон), требуемой для нахождения нейронов, принадлежащих эталону, среди активных и неактивных нейронов начального паттерна, соответственно.

Поскольку количество активных нейронов в начальном паттерне полагается равным $n = pN$, то

$$p = q_1p + q_0(1-p) \quad (10)$$

и, следовательно, начальное пересечение определяется по формуле

$$m(\mathbf{X}^l, \mathbf{X}^{in}) = m_{in} = q_1 - q_0 = (q_1 - p)/(1-p). \quad (11)$$

Для нахождения $H(\mathbf{X}^l|\mathbf{X}^{in}, \mathbf{X}^f)$ введем вероятности

$$p_{\mu\nu} = \mathcal{P}\{X_i^f = 1|X_i^l = \mu, X_i^{in} = \nu\}$$

того, что данный нейрон, который *принадлежит* ($\mu = 1$) или *не принадлежит* ($\mu = 0$) воспроизводимому эталону и *активен* ($\nu = 1$) или *не активен* ($\nu = 0$) в начальном паттерне, активен в финальном паттерне. Тогда

$$P_{11} = \mathcal{P}\{X_i^{in} = 1, X_i^f = 1\} = p_{11}q_1p + p_{01}q_0(1-p)$$

есть вероятность того, что данный нейрон активен в начальном и финальном паттернах,

$$P_{10} = \mathcal{P}\{X_i^{in} = 1, X_i^f = 0\} = (1-p_{11})q_1p + (1-p_{01})q_0(1-p)$$

вероятность того, что он активен в начальном и не активен в финальном паттерне,

$$P_{01} = \mathcal{P}\{X_i^{in} = 0, X_i^f = 1\} = p_{10}(1-q_1)p + p_{00}(1-q_0)(1-p)$$

вероятность того, что он не активен в начальном и активен в финальном паттерне, и

$$P_{00} = \mathcal{P}\{X_i^{in} = 0, X_i^f = 0\} = (1 - p_{10})(1 - q_1)p + (1 - p_{00})(1 - q_0)(1 - p)$$

вероятность того, что он не активен в начальном и не активен в финальном паттерне.

Тогда $H(\mathbf{X}^l | \mathbf{X}^{in}, \mathbf{X}^f) = Nh_f$, где

$$h_f = P_{11}h(p_{11}q_1p/P_{11}) + P_{10}h((1 - p_{11})q_1p/P_{10}) + P_{01}h[p_{10}(1 - q_1)p/P_{01}] + P_{00}h[(1 - p_{10})(1 - q_1)p/P_{00}]. \quad (12)$$

Четыре члена в (12) дают количества информации, которые требуются, соответственно, для нахождения нейронов, принадлежащих записанному паттерну среди нейронов, *активных* и в начальном и в финальном паттернах, *активных* в начальном и *не активных* в финальном паттерне, *не активных* в начальном и *активных* в финальном паттерне и *не активных* в начальном и в финальном паттернах.

Так как количество активных нейронов в финальном паттерне полагается равным $n = pN$, то

$$p = p_{11}q_1p + p_{01}q_0(1 - p) + p_{10}(1 - q_1)p + p_{00}(1 - q_0)(1 - p). \quad (13)$$

Полезно также ввести вероятности

$$p_1 = p_{11}q_1 + p_{10}(1 - q_1), \quad p_0 = p_{01}q_0 + p_{00}(1 - q_0)$$

того, что нейрон, принадлежащий или не принадлежащий восстанавливаемому эталону, активен в финальном паттерне. Так как количество активных нейронов в финальном паттерне равно Np , то

$$p = p_1p + p_0(1 - p). \quad (14)$$

Финальное пересечение $m(\mathbf{X}^l, \mathbf{X}^f)$, которое равно

$$m_f = p_1 - p_0 = (p_1 - p)/(1 - p), \quad (15)$$

может быть легко выражено через вероятности $p_{\mu\nu}$.

Если эталон восстановлен без ошибок, т.е. $p_1 = p_{11} = p_{10} = 1$ и $p_0 = p_{01} = p_{00} = 0$, то $H(\mathbf{X}^l | \mathbf{X}^{in}, \mathbf{X}^f) = 0$ (дополнительной информации об эталоне не требуется, потому что он полностью определен финальным паттерном). Однако, даже в этом случае выигрыш информации не равен энтропии эталона, потому что начальный паттерн уже содержит некоторую информацию о нем. В соответствии с (10) и (9), эта информация

$$\begin{aligned} I(\mathbf{X}^l, \mathbf{X}^{in}) &= H(\mathbf{X}^l) - H(\mathbf{X}^l | \mathbf{X}^{in}) = H(\mathbf{X}^{in}) - H(\mathbf{X}^{in} | \mathbf{X}^l) = \\ &= N[h(p) - p h(q_1) - (1-p)h(p(1-q_1)/(1-p))]. \end{aligned}$$

Она мала, только когда высок уровень искажения эталонов. Она равна нулю, когда входной паттерн статистически независим от эталона (т.е. $q_1 = p$). Однако, в большинстве статей (например, [3,6,24]) эта информация игнорируется. Это приводит к переоценке количества информации, извлекаемой из сети. Это количество, очевидно, равно нулю, если начальный паттерн совпадает с восстанавливаемым эталоном (потому что эталон уже полностью определен), в то время как игнорирование информации, данной начальным паттерном, приводит к ненулевой оценке для количества информации, извлекаемой из сети. Информация, содержащаяся относительно эталона в начальном паттерне, учитывается в наших работах (например, [11]) и в работе [16]. Аналогично, в большинстве статей (например [16]) информация о восстанавливаемом эталоне, содержащаяся в начальном паттерне, игнорируется при оценке финальной неопределенности в восстанавливаемом эталоне. То есть, вместо условной энтропии $H(\mathbf{X}^l | \mathbf{X}^{in}, \mathbf{X}^f)$ в (8) используется $H(\mathbf{X}^l | \mathbf{X}^f)$. Для совпадения этих условных энтропий достаточно выполнения двух условий. Во-первых, начальный паттерн является статистически независимым от эталона (т.е. $q_1 = q_0 = p$) и, во-вторых, активность нейронов в финальном паттерне не зависит от их активности в начальном паттерне (т.е. $p_{11} = p_{10} = p_1$ и $p_{01} = p_{00} = p_0$). Тогда $H(\mathbf{X}^l | \mathbf{X}^{in}, \mathbf{X}^f) = H(\mathbf{X}^l | \mathbf{X}^f) = N h'_f$, где

$$h'_f = p h(p_1) + (1-p) h(p_0). \quad (16)$$

Два члена в (16) определяют среднюю информацию, необходимую для нахождения нейронов, активных в эталоне, среди, соответственно, активных и неактивных нейронов в финальном паттерне. Уравнение (16) совпадает с уравнением (9), если заменить p_1 на q_1 и p_0 на q_0 .

Однако, в общем случае игнорирование информации, имеющейся в начальном паттерне, ведет к недооценке количества информации, извлекаемой из сети. Например, если начальный паттерн совпадает с эталоном

(т.е. $q_1 = 1, q_0 = 0$) вычисление по (8) дает ожидаемое нулевое количество информации, в то время как при использовании $H(\mathbf{X}^l|\mathbf{X}^f)$ вместо $H(\mathbf{X}^l|\mathbf{X}^{in}, \mathbf{X}^f)$ нулевое количество получается, только если эталоны воспроизводятся точно (т.е. $p_1 = 1, p_0 = 0$). В остальных случаях оно оказывается парадоксально отрицательным.

Извлеченная из памяти информация о всех записанных эталонах $I_g = Li_g$ и, соответственно, информационная эффективность

$$E = I_g/N^2 = i_{in} - i_f, \quad (17)$$

где

$$i_{in} = \alpha h_{in}/h(p) \quad (18)$$

и

$$i_f = \alpha h_f/h(p). \quad (19)$$

Когда информационная нагрузка мала, эталоны воспроизводятся точно и области притяжения имеют большой размер (т.е. $h_f = 0$ и $h_{in} \simeq h(p)$). Тогда информационная эффективность близка к α . Однако при возрастании информационной загрузки h_{in} убывает (из-за убывания размера областей притяжений) и h_f возрастает (из-за уменьшения качества воспроизведения). Поэтому возрастание E за счет возрастания α преодолевается убыванием $h_{in} - h_f$ и информационная эффективность начинает падать. В связи с этим, для каждого значения разреженности существует оптимальная информационная нагрузка, обеспечивающая максимальную информационную эффективность. Ее оценка в зависимости от p является основной целью настоящей лекции.

Аналитические подходы

Размер областей притяжения рассчитывается здесь с использованием двух аналитических подходов: одношагового приближения (SS) и статистической нейродинамики (SN). Соответствующие нейродинамические уравнения для нейронных сетей с разреженным кодированием были получены в работах [9] и [10].

Одношаговое приближение

Принципиальной особенностью SS-приближения [18] является то, что на каждом временном шаге нейродинамики игнорируется статистическая зависимость между активностью сети и матрицей связей и учитывается только один макропараметр нейродинамики: пересечение текущего паттерна с воспроизводимым эталоном $m(t)$. Пренебрежение статистической зависимостью справедливо только для первого шага нейродинамики, когда начальная активность сети действительно устанавливается независимо от матрицы связей. Именно поэтому данный подход называется *одношаговым приближением* или приближением первого шага. Другими, более точными аналитическими методами, а также с помощью компьютерного моделирования (см. работы [4, 5] и многие другие) было показано, что SS-приближение очень неточно в случае плотного кодирования ($p = 0.5$). Однако, оно оказалось достаточно точным для случая разреженного кодирования [10]. Более того, в разделе «Компьютерное моделирование» будет показано, что в этом случае оно дает даже лучшее соответствие с экспериментом, чем SN-приближение.

При воспроизведении эталона \mathbf{X}^l синаптические возбуждения каждого нейрона на первом временном шаге выражаются, в соответствии с уравнениями (1) и (3), в виде

$$\eta_i = (X_i^l - p)m_{in} + N_i,$$

где m_{in} — пересечение между воспроизводимым эталоном и начальным паттерном \mathbf{X}^{in} и N_i — шум, обусловленный записью в матрицу связей других эталонов. Для используемой модели [10]

$$\mathcal{M}\{N_i\} = -\alpha p/h(p), \quad \sigma^2 = D\{N_i\} = \alpha r/h(p), \quad (20)$$

где

$$r = p(1 - p). \quad (21)$$

Для больших N и L распределение N_i близко к нормальному. Тогда

$$p_{01} = p_{00} = p_0 = \Phi(\theta_0), \quad p_{11} = p_{10} = p_1 = \Phi(\theta_1), \quad (22)$$

где

$$\theta_0 = \theta + m_{in}p/\sigma, \quad \theta_1 = \theta - m_{in}(1 - p)/\sigma, \quad (23)$$

$$\Phi(x) = 1/(2\pi)^{1/2} \int_x^{\infty} \exp(-u^2/2) du$$

и θ — масштабированный порог возбуждения. В нашей модели он выбирается так, что общее количество активных нейронов остается фиксированным и равным n , т. е. θ выбирается так, чтобы выполнялось уравнение (14). В результате первого шага воспроизведения пересечение меняется на

$$m(1) = p_1 - p_0. \quad (24)$$

В SS-приближении эти уравнения полагаются выполняющимися для всех временных шагов (естественно, m_{in} нужно заменить на $m(t)$ и $m(1)$ на $m(t+1)$).

Финальное стабильное состояние удовлетворяет условию

$$m(t+1) = m(t) = m_f. \quad (25)$$

Кривые, характеризующие поведение активности сети в зависимости от α , m_{in} и p представлены на рис. 1. Пусть начальное состояние активности сети для данного α характеризуется точкой (m_{in}, α) . Если эта точка лежит под кривой, пересечение между текущим и восстанавливаемым паттернами смещается в процессе воспроизведения вправо к финальному пересечению, задаваемому правой частью кривой. Поэтому данная часть кривой задает величину пересечения между эталоном и финальным стабильным состоянием в его окрестности, т. е. качество воспроизведения. Пересечение смещается влево для любой точки выше кривой. Поэтому левая часть кривой соответствует границе области притяжения. Эта часть отсутствует при $p = 0.5$. Критическое значение α_{cr} , которое дает информационную емкость сети для данного значения p , соответствует максимуму каждой кривой.

Рис. 1 показывает, что SS-приближение предсказывает монотонное убывание размера областей притяжения при возрастании разреженности т. е. при убывании p .

В предельном случае

$$p \rightarrow 0, \\ p_1 = \Phi(\theta_1) \simeq m(1), \quad p_0 = \Phi(\theta_0) \simeq p(1 - m(1)), \\ \theta_0 - \theta_1 \simeq \frac{m_{in}}{\sigma} = m_{in} \frac{\sqrt{\ln(1/p)}}{\sqrt{\alpha \ln 2}}.$$

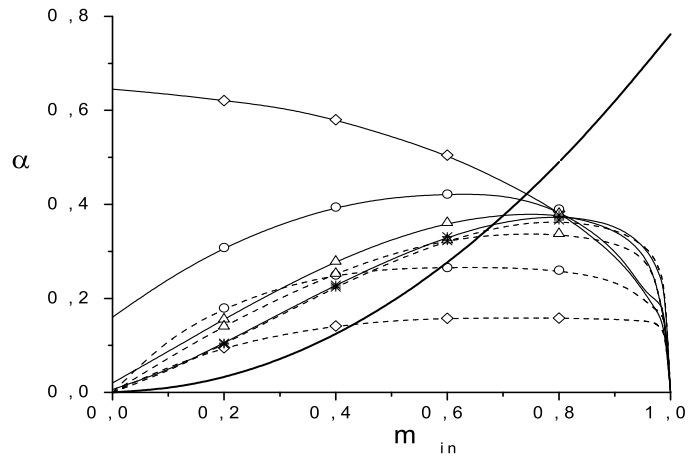


Рис. 1. Области притяжения, предсказываемые SS (тонкие сплошные линии) и SN (пунктирные линии). Толстая сплошная линия соответствует предельному случаю большой разреженности ($p \rightarrow 0$); \diamond — для $p = 0.5$, \circ — для $p = 0.1$, Δ — для $p = 0.02$, $*$ — для $p = 0.004$.

Тогда

$$\theta_0 \simeq \sqrt{2 \ln(1/p)}, \quad \theta_1 \simeq \sqrt{2 \ln(1/p)} (1 - m_{in} / \sqrt{2\alpha \ln 2}).$$

Так как $\ln(1/p) \gg 1$, то

$$\theta_1 \ll -1, \quad \text{т.е. } m(1) \simeq 1, \quad \text{если } \alpha < \frac{m_{in}^2}{2 \ln 2},$$

$$\theta_1 \gg 1, \quad \text{т.е. } m(1) \ll 1, \quad \text{если } \alpha > \frac{m_{in}^2}{2 \ln 2}.$$

Поэтому

$$\alpha = \frac{m_{in}^2}{2 \ln 2} \tag{26}$$

дает границу области притяжения в этом предельном случае. Формула (26) дает ту же зависимость α от m_{in} , что полученная в работе [3], однако с другим коэффициентом. В этой работе границы областей притяжения соответствуют требованию, чтобы вероятность ошибки (на первом шаге) хотя бы в одном нейроне и хотя бы в одном из записанных эталонов была мала, в то время как в наших расчетах мала ошибка в каждом отдельно взятом нейроне при воспроизведении каждого отдельно взятого эталона.

Статистическая нейродинамика

Метод статистической нейродинамики был разработан *Amari & Maginu* [4] для плотного кодирования и модифицирован в работах [9, 10, 25] для разреженного кодирования. Принципиальной особенностью этого подхода является учет статистической зависимости между активностью сети и матрицей связей. Это достигается включением дополнительной переменной $c(t)$ в систему рекурсивных нейродинамических уравнений. В SN-приближении остаются справедливыми все уравнения SS-приближения, кроме уравнения (21), которое задает дисперсию синаптических возбуждений $\mathcal{D}\{N_i(t)\}$. В SN-приближении оно принимает следующий вид

$$r(t) = p(1-p) + 2p(1-p)m(t)m(t-1)c(t) + c^2(t) \left[r(t-1) + \frac{\alpha p(1-p)}{h(p)}(1 - c^2(t-1)) \right], \quad (27)$$

где

$$c(t) = \frac{p \exp[-\theta_1^2(t-1)/2] + (1-p) \exp[-\theta_0^2(t-1)/2]}{\sqrt{2\pi\alpha r(t-1)/h(p)}}, \quad (28)$$

$$c(0) = 0.$$

Параметры конечного стабильного состояния находятся из условий

$$m(t+1) = m(t) = m(t-1) = m_f, \quad c(t+1) = c(t) = c^f.$$

Кривые, характеризующие размер областей притяжения в SN-приближении также показаны на рис. 1. Видно, что для $p = 0.5$ результаты SS- и SN-приближений сильно различаются, но разница убывает при возрастании разреженности. Размер областей притяжения в SN меньше, чем в

SS. Так как на первом шаге оба подхода дают одинаковые результаты, из этого следует немонотонное изменение пересечения в процессе воспроизведения, если начальная активность находится внутри области притяжения, предсказываемой SS, но вне области притяжения, предсказываемой SN. В этом случае пересечение вначале возрастает, но потом уменьшается до нуля. Такое немонотонное поведение $m(t)$ действительно наблюдалось при компьютерном моделировании (см. [4, 16, 22] и др. работы). Заметим, что отличие размеров областей притяжений, предсказываемых SS и SN, убывает, когда убывает p . Они практически неразличимы уже при $p = 0.004$.

Как показано на рис. 1, SN предсказывает немонотонную зависимость размера областей притяжения от p . Когда p убывает от 0.5 до 0.1, кривые на рис. 1 смещаются вверх и влево (что означает возрастание размера области притяжения). Когда p продолжает убывать, они смещаются вверх и вправо (что означает убывание размера области притяжения).

Компьютерное моделирование

Мы тестировали сети размером от 200 до 15000 нейронов с p от 0.02 до 0.5. Начальные пересечения были в интервале от 0.1 до 1. Общее число испытаний для каждой комбинации p , N , α и m_{in} было 2000 для $N < 3000$, 1000 для $3000 \leq N \leq 5000$ и 250 для $N \geq 5000$.

Плотное кодирование

Полученные распределения для пересечений m_f между финальными состояниями активности сети и восстанавливаемыми эталонами показаны на рис. 2. Общие свойства этих распределений хорошо известны (см. статьи [5, 16, 19] и многие другие). Каждое распределение состоит из двух хорошо разделенных мод. Первая мода близка к $m_f = 1$ («правильные» аттракторы) и соответствует восстановленным, а вторая мода — невосстановленным эталонам («ложные» аттракторы). Когда начальный паттерн находится достаточно далеко от границы внутри области притяжения, преобладает первая мода. Когда он далеко от границы вне области притяжения — вторая. Для каждого значения начального пересечения m_{in} существует узкая переходная область вблизи критического значения α , где обе моды имеют сходный размер. Для $m_{in} = 1$ переход происходит между $\alpha = 0.14$ и $\alpha = 0.15$, для $m_{in} = 0.5$ — между 0.12 и 0.13, для $m_{in} = 0.3$ — между

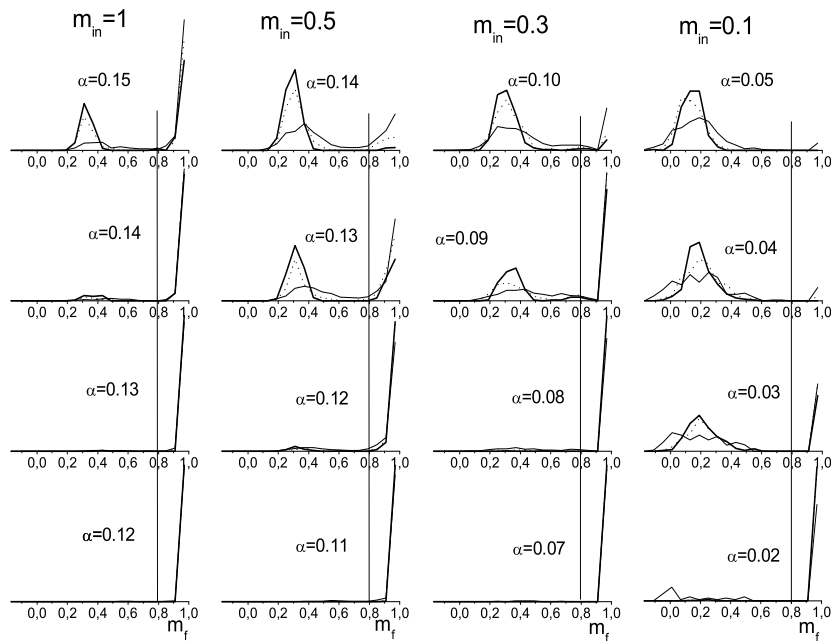


Рис. 2. Распределение финальных пересечений для $p = 0.5$ и различных значений N , α и m_{in}
 Распределение финальных пересечений для $p = 0.5$ и различных значений N , α и m_{in} . Вертикальные линии соответствуют выбранным порогам для разделения траекторий, сходящихся к «правильным» ($m_f \simeq 1$) и «ложным» ($m_f \ll 1$) аттракторам. Здесь: $N = 1000$ — тонкие сплошные линии, $N = 3000$ — пунктирные линии, $N = 5000$ — толстые сплошные линии.

0.08 and 0.09, и для $m_{in} = 0.1$ — между 0.02 и 0.03. Переходная область сужается при возрастании размера сети.

Для более точного вычисления $\alpha_{cr}(m_{in})$ мы вычисляем «площадь» первой моды, т. е. вероятность $P(\alpha, m_{in}, N)$ того, что траектория, стартовая из начального состояния, имеющего пересечение с эталоном m_{in} , заканчивается около соответствующего эталона. Эта вероятность зависит от выбора границы разделения мод m_b . Хотя этот выбор в какой-то сте-

пени произволен, это не критично из-за хорошей разделенности мод. Как в [5, 10, 16], мы выбираем $m_b = 0.8$. Эта граница показана на рис. 2 вертикальными линиями. Так как P изменяется от 0 до 1, удобно [5, 8] сделать преобразование

$$P = \frac{\exp F}{1 + \exp F}, \quad (29)$$

где F меняется от $-\infty$ до $+\infty$. Полученные оценки F для N в области от 200 до 5000 показаны на рис. 3 светлыми знаками. Тонкие сплошные линии — результат их аппроксимации по формуле

$$F = a_0 + a_1\alpha + a_2N(\alpha - \alpha_{cr}) + a_3 \ln N. \quad (30)$$

Для проверки справедливости этой аппроксимации мы дополнили наши данные на рис. 3а ($m_{in} = 1$) данными, представленными в работе [19] для больших размеров сети (N меняется до 32000) и аппроксимировали полный набор данных по (30). Данные, взятые из работы [19], показаны на рис. 3а темными знаками. Видно, что все данные достаточно единообразны и могут быть хорошо аппроксимированы выбранной формулой.

В работах [5, 8, 10, 16, 17] оценки F аппроксимировались аналогичной формулой, но без последнего (логарифмического) члена. Его влияние важно лишь тогда, когда N достаточно велико, а α очень близко к α_{cr} . Например, немонотонная зависимость P от N проявляется для $\alpha = 0.14$ только когда N достигает 10000.

Экспериментальная оценка α_{cr} соответствует толстой прямой линии, которая разделяет на рис. 3а два множества кривых. Верхние кривые вогнуты и стремятся к $+\infty$ ($P \rightarrow 1$), когда $N \rightarrow \infty$. Нижние — выпуклы и стремятся к $-\infty$ ($P \rightarrow 0$), когда $N \rightarrow \infty$. Когда α становится близким к α_{cr} , кривизна линий уменьшается. Поэтому, верхние кривые при $\alpha \simeq \alpha_{cr}$ имеют очень широкую нижнюю часть: F близка к константе для большой области значений N . Мы предсказываем (на основе используемой аппроксимации), что при возрастании N кривые в конечном счете поворачивают вверх и $P \rightarrow 1$. Но при приближении α к α_{cr} все большие значения N требуются для того, чтобы F начало возрастать.

Как *Amit et al.* [5], *Horner et al.* [16] и другие, мы рассматриваем значение α_{cr} , полученное в случае, когда начальное состояние совпадает с одним из эталонов ($m_{in} = 1$), как экспериментальную оценку информационной емкости. При совмещении наших данных и данных, полученных

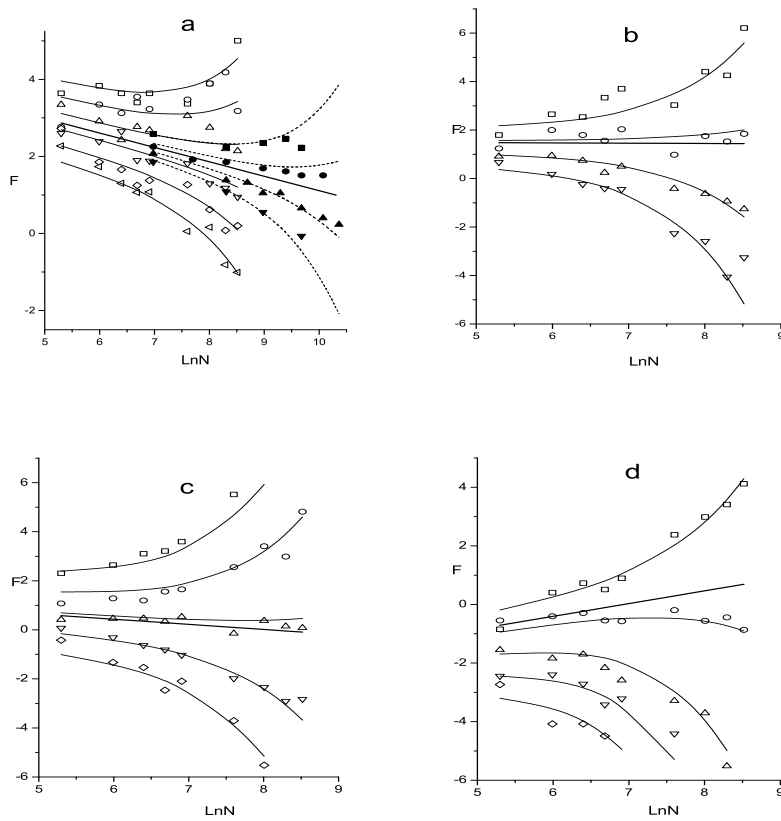


Рис. 3. Зависимость F от α и $\ln N$ для $p = 0.5$ и $m_{in} = 1$ (а), $m_{in} = 0.5$ (b), $m_{in} = 0.3$ (c) и $m_{in} = 0.1$ (d). Светлые точки — наши результаты, темные — данные [19]. На (а) для наших результатов α меняется от 0.13 (верхняя кривая) до 0.155 (нижняя кривая) с шагом 0.005, для данных [19] — от 0.14 до 0.146 с шагом 0.002, на (b) α меняется от 0.11 до 0.14 с шагом 0.01, на (c) α меняется от 0.07 до 0.11 с шагом 0.01, на (d) α меняется от 0.02 до 0.06 с шагом 0.01. Толстые прямые линии соответствуют $\alpha = \alpha_{cr} = 0.1425$ на (а), $\alpha = \alpha_{cr} = 0.121$ на (b), $\alpha = \alpha_{cr} = 0.091$ на (c) и $\alpha = \alpha_{cr} = 0.027$ на (d).

в работе [19], $\alpha_{cr} = 0.1429 \pm 0.0003$. Если используются только наши данные, то $\alpha_{cr} = 0.1425 \pm 0.002$. Если только данные работы [19], то $\alpha_{cr} = 0.1430 \pm 0.0002$. Таким образом наши данные хорошо согласуются с данными работы [19] и дают довольно точную оценку α_{cr} . Все приведенные оценки хорошо согласуются также со значением $\alpha_{cr} = 0.1455 \pm 0.001$, полученным в работе [16] для асинхронной (последовательной) динамики, и близки (хотя и достоверно отличны) от значений $\alpha_{cr} = 0.138$ и $\alpha_{cr} = 0.158$, предсказанных RM- и SN-приближениями в [5] и [4], соответственно.

В работе [25] SN-приближение обобщено на случай, когда корреляции между шумовыми членами учитываются на нескольких шагах нейродинамики. В этом обобщении приближение первого порядка соответствует уравнениям Амари-Магину [4]. Приближение второго порядка дает более точную оценку информационной емкости: $\alpha_{cr} = 0.142$. Однако далее, когда порядок приближения возрастает, точность приближения уменьшается: (0.140, 0.139), достигая значения 0.138, предсказанного RM.

Для случая, когда начальные паттерны отличались от эталонов (рис. 3b, 3c, 3d), аппроксимация данных по формуле (30) дает: $\alpha_{cr} = 0.121 \pm 0.003$ для $m_{in} = 0.5$, $\alpha_{cr} = 0.091 \pm 0.0015$ для $m_{in} = 0.3$ и $\alpha_{cr} = 0.027 \pm 0.001$ для $m_{in} = 0.1$. Эти значения приведены на рис. 4 вместе с кривыми, полученными, полученными SN-методом; эти данные взяты из работы Окада [22]. Точка, построенная при $m_{in} = 0.8$, показывает полученную оценку информационной емкости. Видно, что значения, полученные нами с помощью компьютерного моделирования (кроме значений для информационной емкости), лежат между кривыми, полученными по методу Окада по приближениям второго и четвертого порядка. Это свидетельствует в пользу точности как этого метода, так и экспериментальных оценок.

Наши экспериментальные данные хорошо согласуются также с результатами компьютерного моделирования сети с асинхронной динамикой, приведенными в [16]. Однако результаты компьютерного моделирования, приведенные в [22], дают заниженную величину области притяжения. Например, при $m_{in} = 0.1$ в этой работе приведена оценка $\alpha_{cr} \simeq 0.02$ вместо 0.027, полученной в наших экспериментах. Эта разность может быть объяснена двумя причинами. Первое, в работе [22] моделировалась сеть значительно меньшего размера ($N = 1000$). Второе, в этой работе граница области притяжения оценивалась как точка, в которой среднее финальное пересечение становилось равным 0.5. Однако эта оценка является довольно грубой, так как среднее конечное пересечение определяется двумя фактора-

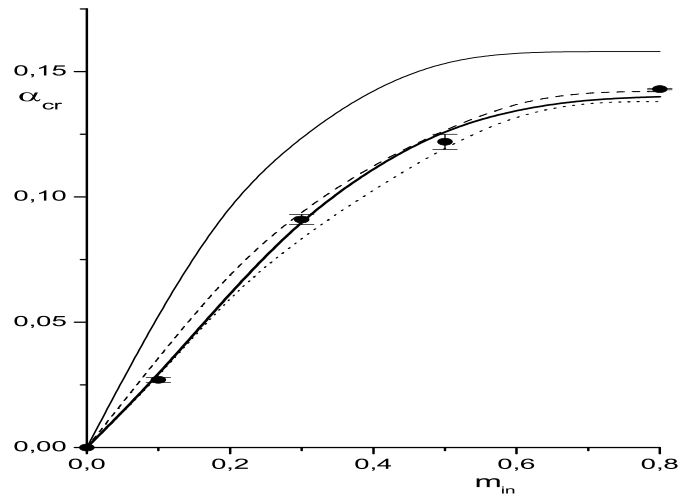


Рис. 4. Базы притяжения для $p = 0.5$

Получены SN-методом первого порядка (сплошная линия), второго порядка (штриховая линия) и четвертого порядка (пунктирная линия). Точками представлены результаты компьютерного моделирования. Результаты, полученные SN-методом второго и четвертого порядка, взяты из работы [22].

ми. Во-первых, оно зависит от перераспределения финальных пересечений между модой, близкой к единице (конечные состояния в окрестности эталонов), и модой, близкой к нулю (ложные аттракторы). Во-вторых, оно зависит от положения этих мод на оси финальных пересечений. Так как влияние этих факторов по-разному проявляется при увеличении N , нельзя экстраполировать данные, полученные для финальных пересечений при относительно малых значениях N , на случай $N \rightarrow \infty$.

На рис. 5 показаны средние значения параметров активности сети в зависимости от шага нейродинамического процесса при $\alpha = 0.09$, $m_{in} = 0.3$ и $N = 5000$. Значение m_{in} выбрано очень близко к границе области притяжения для данного α . Как показано на рис. 2 и рис. 3, при этих значениях α и m_{in} две моды конечных пересечений m_f почти равны по площади (т. е. $P \simeq 0.5$). Кривые получены усреднением отдельно траекторий, сходящихся

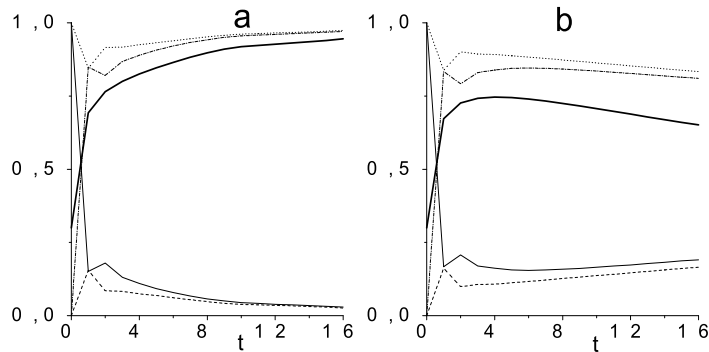


Рис. 5. Осредненные характеристики воспроизведения
 Получены по траекториям, сходящимся к правильным **(а)** и ложным **(б)** аттракторам для $p = 0.5$, $m_{in} = 0.3$, $\alpha = 0.09$ and $N = 5000$. Здесь p_{11} — пунктирные линии; p_{10} — штрих-пунктирные линии; p_{01} — тонкие сплошные линии; p_{00} — штриховые линии; m — толстые сплошные линии.

к правильным (рис. 5а) и ложным (рис. 5б) аттракторам (около 250 траекторий для каждого случая). На первом шаге нейродинамики усредненные параметры для траекторий, сходящихся к правильным и ложным аттракторам, совпадают между собой и с предсказанием SS-приближения. На этом шаге пересечение с воспроизводимым эталоном возрастает. Для траекторий, сходящихся к правильным аттракторам, оно продолжает возрастать и на следующих шагах. Однако для траекторий, сходящихся к ложным аттракторам, оно возрастает только на нескольких первых шагах, а затем медленно падает. Это хорошо знакомое поведение траекторий, сходящихся к правильным и ложным аттракторам, которые начинаются внутри области притяжения, предсказанной SS-приближением. Более интересно взглянуть на поведение вероятностей $p_{\mu\nu}$, определенных в разделе «Информация, извлекаемая из сети за счет коррекции искаженных эталонов» (см. с. 41). Для плотного кодирования поведение нейронов, активных и неактивных в воспроизводимых эталонах, симметрично, т. е. $p_{01} = 1 - p_{10}$, а $p_{00} = 1 - p_{11}$. На

первом шаге, в соответствии с предсказаниями SS- и SN-приближений, вероятности p_{11} и p_{10} равны. На следующем шаге они становятся различными, затем медленно сближаются для траекторий, сходящихся к правильным аттракторам, и остаются различными для траекторий, сходящихся к ложным аттракторам. Нейроны, которые были активны в начальном паттерне, имеют большую вероятность активироваться: $p_{11} > p_{10}$, а $p_{01} > p_{00}$. Таким образом, нейроны как бы «помнят» свою предысторию в течение многих шагов нейродинамики. Этот факт игнорируется SN-приближениями всех порядков. Однако для плотного кодирования различие между p_{11} и p_{10} не так велико. Поэтому это приближение оказывается достаточно точным для плотного кодирования эталонов. Для разреженного кодирования, как показано далее, различие в поведении нейронов, активных и неактивных в начальном паттерне, более значительно.

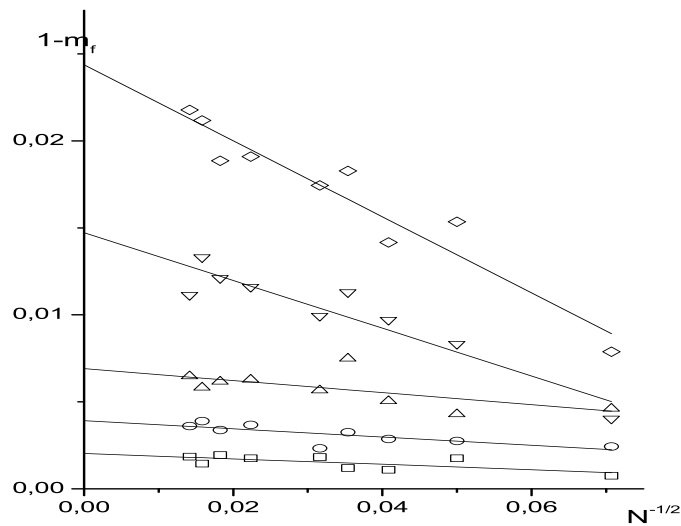


Рис. 6. Зависимость качества воспроизведения от α и N для $p = 0.5$, полученные компьютерным моделированием. Значения α меняются от 0.1 (нижняя линия) до 0.14 (верхняя линия) с шагом 0.1

Для оценки качества воспроизведения мы использовали средние значения финальных пересечений m_f для траекторий, сходящихся к правильным аттракторам из начальных состояний, совпадающих с эталонами (т. е. когда $m_{in} = 1$). Результаты показаны на рис. 6. Для каждого значения α значения средних финальных пересечений были аппроксимированы прямыми линиями в зависимости от $1/\sqrt{N}$, и их пересечение с осью ординат рассматривалось как асимптотическая оценка качества воспроизведения для данного значения α . Полученные значения m_f приведены на рис. 7. Они сравниваются с оценками качества воспроизведения, полученными с помощью RM- и SN-приближений, а также с данными, полученными *Корингом* [19] для гораздо больших размеров сети (N до 50000). Точки, полученные с помощью компьютерного моделирования, опять лежат между кривыми, полученными с помощью RM- и SN-приближений, а точки, полученные нами, практически совпадают с данными *Коринга*. Следует заметить, что на рис. 7 значения $(1 - m_f)$ умножены на 5, чтобы привести их в тот же диапазон значений, что и для разреженного кодирования.

Разреженное кодирование

Распределения финальных пересечений m_f , полученных при $p = 0.02$, показаны на рис. 8. Главной особенностью этих распределений является то, что они имеют две явно различающиеся моды только в случае, когда начальные состояния достаточно далеки от эталонов ($m_{in} = 0.1$ и $m_{in} = 0.3$). Когда $m_{in} = 0.5$ или $m_{in} = 1$ присутствует только одна мода, которая медленно смещается к меньшим значениям m_f при возрастании α . Та же особенность имеется при $p = 0.1$. Можно все-таки предположить, что распределение имеет две моды и при больших значениях m_{in} , которые слишком близки и имеют слишком большую дисперсию, чтобы их было легко разделить. При возрастании N моды становятся более узкими и их можно разделить, как это показано на рис. 8 для $p = 0.02$, $m_{in} = 0.5$, $\alpha = 0.34$ и $N = 15000$. Наличие дополнительной моды вблизи моды, близкой к $m_f = 1$, можно заметить даже для плотного кодирования (см. рис. 2, $m_{in} = 0.3$, $\alpha = 0.09$). Эту моду можно объяснить наличием метастабильных состояний в окрестности эталонов, которые исчезают при возрастании N [20].

Несмотря на плохое разделение мод, для оценки α_{cr} мы использовали ту же процедуру, что и при плотном кодировании. В качестве границы для разделения правильных и ложных аттракторов мы приняли $m_b = 0.75$ (вер-

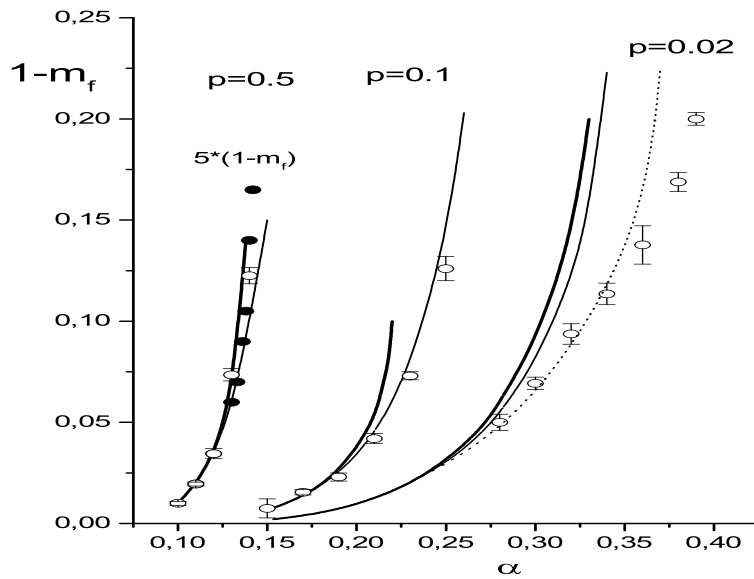


Рис. 7. Зависимость качества воспроизведения, полученная методом RM (толстые сплошные линии), SN (тонкие сплошные линии) и SS (пунктирные линии) и экстраполяцией результатов компьютерного моделирования к пределу $N \rightarrow \infty$ (светлые точки), от α и p . Темные точки — данные, взятые из работы [19].

тикальные линии на рис. 8). При $m_{in} = 0.1$ и $m_{in} = 0.3$ эта граница имеет тот же смысл, что и при плотном кодировании. Она явно разделяет хорошо выраженные моды для правильных и ложных аттракторов. При $m_{in} = 0.5$ и $m_{in} = 1$ вероятность сходимости нейродинамического процесса к правильным аттракторам сильно зависит от выбора m_b . Выбор $m_b = 0.75$ представляется разумным, потому что он обеспечивает изменение вероятности P , что некоторая траектория сойдется к аттрактору в окрестности эталона, от значения 1 при малых α до 0 при больших α . Оказалось, что выбор m_b в некотором разумном диапазоне от 0.7 до 0.8 не так сильно влияет на оценку α_{cr} , как на оценку величины P для каждой комбинации α и N .

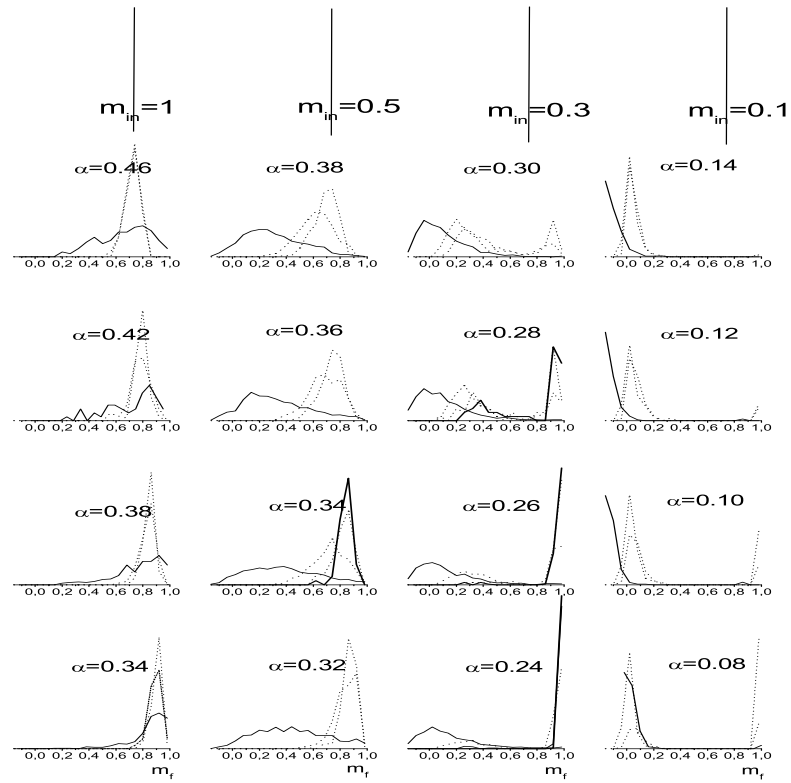


Рис. 8. Распределение финальных пересечений для $p = 0.02$ и различных значений N , α и m_{in} . Вертикальные линии соответствуют выбранным порогам ($m_b = 0.75$) для разделения траекторий, сходящихся к правильным и ложным аттракторам. Тонкие сплошные линии — $N = 850$ для $m_{in} = 0.1$, $N = 1100$ для $m_{in} = 0.3$ и $N = 1000$ для $m_{in} \geq 0.5$. Штриховые линии — $N = 5000$. Пунктирные линии — $N = 10000$. Толстые сплошные линии — $N = 15000$.

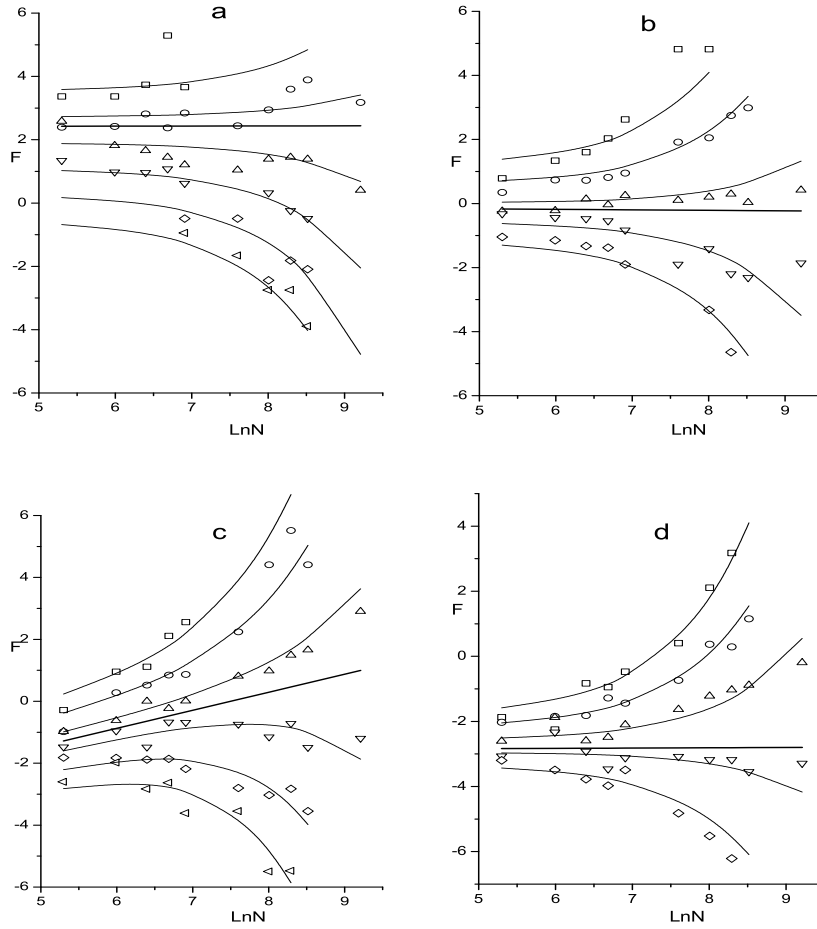


Рис. 9. Зависимость F от α и $\ln N$ для $p = 0.1$ и $m_{in} = 1$ (a), $m_{in} = 0.5$ (b), $m_{in} = 0.3$ (c) и $m_{in} = 0.1$ (d). Значения α изменяются с шагом 0.02 от 0.23 (верхняя кривая) до 0.33 (нижняя кривая) на (a), от 0.21 до 0.29 на (b), от 0.15 до 0.25 на (c) и от 0.05 до 0.13 на (d). Толстые прямые линии соответствуют $\alpha = \alpha_{cr} = 0.257$ на (a), $\alpha = \alpha_{cr} = 0.252$ на (b), $\alpha = \alpha_{cr} = 0.2$ на (c) и $\alpha = \alpha_{cr} = 0.104$ на (d).

Для получения оценки α_{cr} вероятности P трансформировались, как и для плотного кодирования, в значения F по формуле (29), а значения F аппроксимировались формулой (30). Результаты аппроксимации показаны на рис. 9 ($p = 0.1$) и рис. 10 ($p = 0.02$).

Как и для плотного кодирования, формула (30) обеспечивает хорошее приближение для зависимости F от α и N . Среднеквадратичная ошибка рассогласования экспериментальных значений F от рассчитанных по формуле (30) была примерно той же, что и для плотного кодирования (в диапазоне от 0.35 до 0.45). Информационная емкость (оцененная как α_{cr} при $m_{in} = 1$) составляет 0.257 ± 0.006 для $p = 0.1$ и 0.411 ± 0.007 для $p = 0.2$. Ее можно сравнить с предсказаниями RM-, SN- и SS-приближений, которые, соответственно, дают 0.22, 0.27 и 0.42 для $p = 0.1$; 0.31, 0.325 и 0.365 для $p = 0.02$. Таким образом, для $p = 0.1$, как и для плотного кодирования, компьютерное моделирование дает оценку информационной емкости между значениями, предсказанными RM- и SN-приближениями. В противоположность этому, для $p = 0.02$ компьютерное моделирование дает удивительно высокую оценку информационной емкости. Эта оценка даже больше предсказания SS-приближения. Для того, чтобы проверить, не является ли полученная завышенная оценка информационной емкости следствием выбора m_b , она была пересчитана при $m_b = 0.8$. Изменение m_b привело лишь к статистически незначимым изменениям информационной емкости: $\alpha_{cr} = 0.396 \pm 0.008$.

Гистограммы, представленные на рис. 8, также показывают, что полученная завышенная оценка информационной емкости при $p = 0.02$ не является столь нереалистичной. Распределение финальных пересечений m_f остается близким к единице вплоть до $N = 10000$ даже при $\alpha = 0.42$. То же самое демонстрирует рис. 10а: вероятность P продолжает возрастать при увеличении N даже при $\alpha = 0.42$. Мы можем предсказать ее падение при $\alpha = 0.42$ и $N > 10000$ только на основе используемой аппроксимации. Таким образом простейшее SS-приближение дает наиболее точную оценку информационной емкости в пределе большой разреженности кодирования. Однако это не столь уж и большое его преимущество, поскольку все три приближения SS, SN и RM дают все более близкие результаты при увеличении разреженности [10].

Результаты для областей притяжения представлены на рис. 11. При $p = 0.1$ (рис. 11а) данные компьютерного моделирования очень близки к полученным SN-методом. Тогда, как при $p = 0.02$ (рис. 11б) они очень близки к полученным SS-методом.

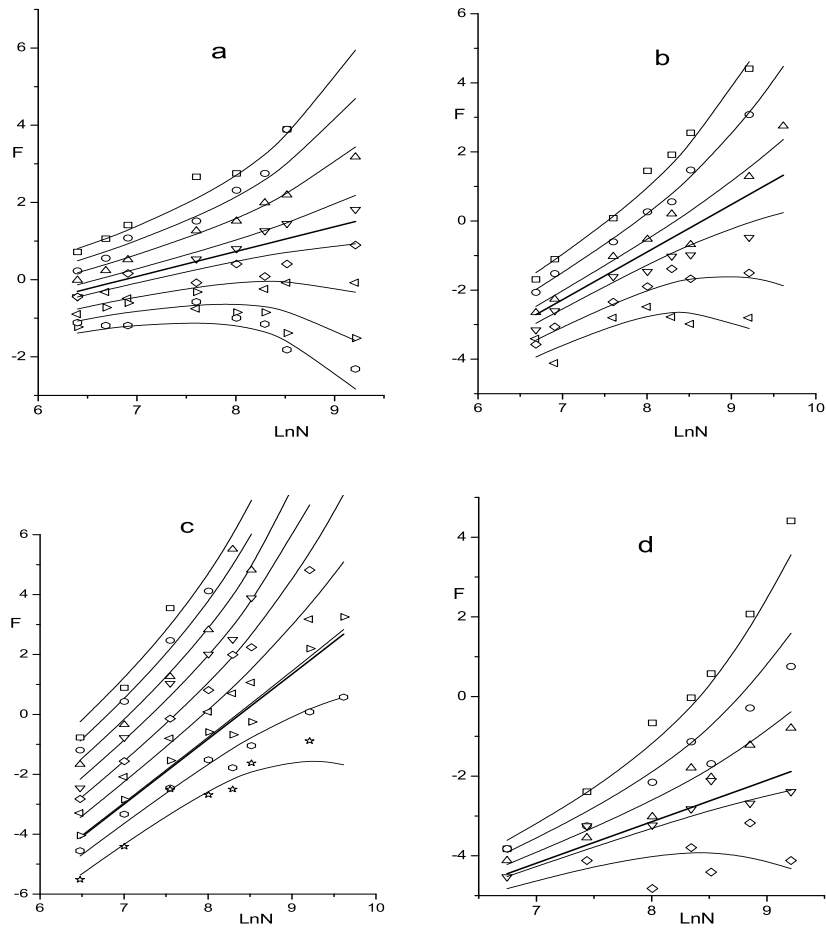


Рис. 10. Зависимость F от α и $\ln N$ для $p = 0.02$ и $m_{in} = 1$ (a), $m_{in} = 0.5$ (b), $m_{in} = 0.3$ (c) и $m_{in} = 0.1$ (d). Значения α изменяются с шагом 0.02 от 0.34 (верхняя кривая) до 0.48 (нижняя кривая) на (a), от 0.28 до 0.38 на (b), 0.14 до 0.3 на (c) и от 0.06 до 0.14 на (d). Толстые прямые линии соответствуют $\alpha = \alpha_{cr} = 0.41$ на (a), $\alpha = \alpha_{cr} = 0.33$ на (b), $\alpha = \alpha_{cr} = 0.261$ на (c) и $\alpha = \alpha_{cr} = 0.115$ на (d).

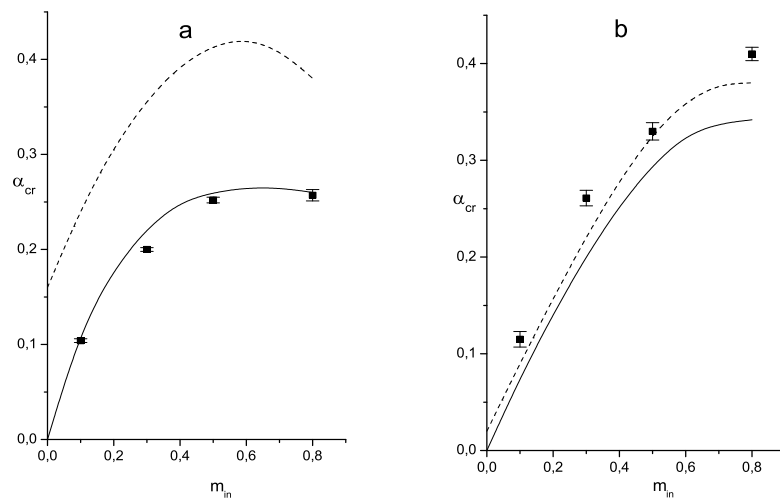


Рис. 11. Размеры баз притяжения для $p = 0.1$ **(a)** и $p = 0.02$ **(b)**. Получены SN-методом (сплошные линии) и SS-методом (штриховые линии). Темные точки представляют результаты компьютерного моделирования.

При $p = 0.02$ области притяжения оказались даже больше, чем предсказывает SS-метод. Это означает, что вблизи границы области притяжения траектории, сходящиеся к правильным аттракторам, вначале будут удаляться от эталонов и только затем приближаться к ним. Такое поведение траекторий противоположно тому, что наблюдается в случае плотного кодирования. Среднее пересечение $m(t)$ между текущим паттерном нейронной активности и воспроизводимым эталоном для $p = 0.02$, $\alpha = 0.24$, $m_{in} = 0.3$ и $N = 15000$ показано на рис. 12а. В полном согласии с предсказанием SS-приближения, на первом шаге среднее пересечение уменьшается, однако затем оно увеличивается, приближаясь к единице. Заметим, что для выбранных параметров сети все 300 тестируемых траекторий сошлись к аттракторам в окрестности соответствующих эталонов. Как и для информационной емкости, полученное для $m_{in} = 0.3$ значение $\alpha_{cr} = 0.26 \pm 0.008$

не выглядит переоценкой критического значения информационной нагрузки при данном начальном пересечении, так как даже при $\alpha = 0.28$ (см. рис. 8 и рис. 10с) мода для правильных аттракторов продолжает возрастать вплоть до $N = 15000$ и мы можем предсказать ее падение при дальнейшем увеличении N только на основе используемой аппроксимации зависимости F от α и N .

Для $p = 0.1$ поведение пересечения $m(t)$ то же самое, что и для $p = 0.5$: оно монотонно возрастает для траекторий, сходящихся к правильным аттракторам, и может вначале увеличиваться, но затем уменьшается для ложных аттракторов (т. е. размер областей притяжения меньше предсказанного SS-методом, см. рис. 11а). Колебания $p_{\mu\nu}$ также, как и разница между p_{11} и p_{10} , возрастают для разреженного кодирования. Таким образом влияние начального состояния нейрона на его активность в процессе восстановления эталонов увеличивается при увеличении разреженности кодирования. Заметим, что для $p = 0.02$ даже $m(t)$ имеет слабые колебания, несмотря на гладкое и монотонное изменение порога активации $T(t)$ (рис. 12b).

К сожалению, невозможно напрямую сравнить наши расчеты с результатами, полученными в работах [16] и [22] для $p = 0.1$. В первой работе границы области притяжения представлены только для таких параметров сети, которые обеспечивают извлечение из нее максимальной информации. Для разреженного кодирования этот максимум достигается с случае, когда начальный паттерн не содержит активных нейронов, не принадлежащих восстанавливаемому эталону (т. е. когда начальный паттерн является фрагментом эталона с числом активных нейронов, меньшим чем в эталоне). В отличие от этого мы моделировали случай, когда уровень активности в начальных паттернах был тот же, что в эталонах. В работе [22] рассматривался только случай постоянного порога активации T . В этом случае области притяжения вокруг эталонов значительно меньше, чем для случая переменного порога, рассмотренного нами.

Как и для плотного кодирования, качество воспроизведения оценивалось по среднему финальному пересечению для правильных траекторий (т. е. траекторий с $m_f > 0.75$), стартующих из неискаженных эталонов. Результаты приведены на рис. 13. Для каждого значения α данные компьютерного моделирования аппроксимировались прямой линией в зависимости от $1/\sqrt{N}$. Пересечение прямой с осью ординат рассматривалось как асимптотическая оценка качества воспроизведения для данного α . Полученные асимптотические значения m_f показаны на рис. 7. Они сравниваются на этом рисунке со значениями, полученными методами RM, SN и

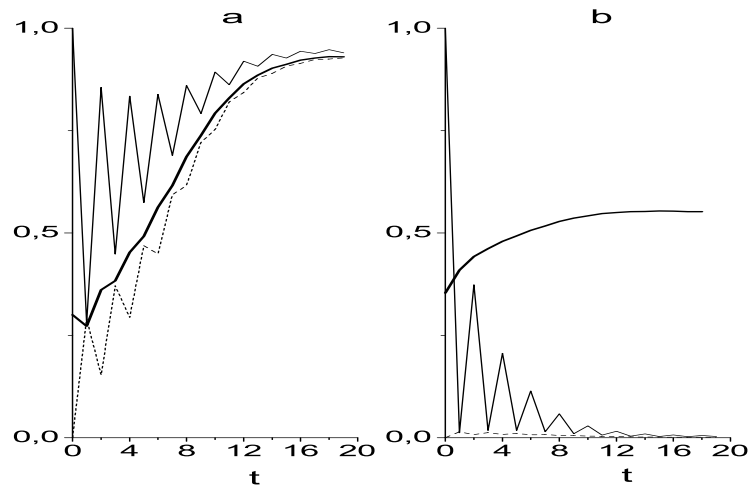


Рис. 12. Усредненные характеристики траекторий нейронной активности для $p = 0.02$, $m_{in} = 0.3$, $\alpha = 0.24$ и $N = 15000$

(a) Сплошная тонкая линия — p_{11} ; штриховая линия — p_{10} ; сплошная толстая линия — пересечение m между текущим паттерном нейронной активности и воспроизводимым эталоном. (b) Сплошная тонкая линия — p_{01} ; штриховая линия — p_{00} ; сплошная толстая линия — средний порог активации T .

SS. Для $p = 0.1$ экспериментальные данные очень близки к полученным SN-методом, тогда как для $p = 0.02$ они более близки к полученным SS-методом.

Информационная эффективность

Количество информации, извлекаемой из сети благодаря коррекции начальных паттернов, задается формулами (8), (9) и (12). Для каждого значения p оно оценивалось аналитически тем методом, который оказался наиболее точным в сравнении с данными компьютерного моделирования. Для $p = 0.5$ это был SN-метод второго порядка, предложенный в работе [22], для $p = 0.1$ — SN-метод первого порядка, предложенный в работе [10] и для

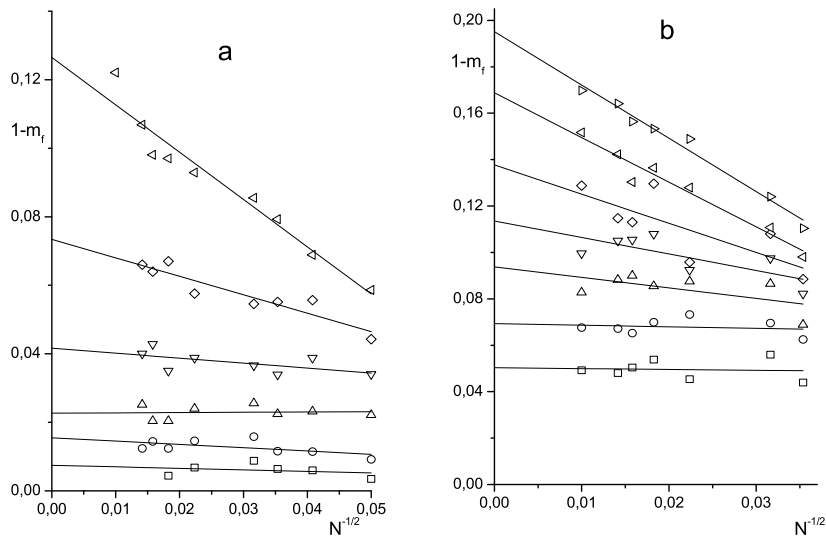


Рис. 13. Зависимость качества воспроизведения от α и N для $p = 0.1$ (а) и $p = 0.02$ (б), полученные компьютерным моделированием (а) Значения α изменяются от 0.15 (нижняя линия) до 0.25 (верхняя линия) с шагом 0.2. (б) Значения α изменяются от 0.28 (нижняя линия) до 0.40 (верхняя линия) с шагом 0.2.

$p = 0.02$ — SS-метод (который фактически является SN-методом нулевого порядка). Результаты этих оценок приведены на рис. 14 в зависимости от информационной нагрузки α . Для каждого значения α начальное пересечение с воспроизводимым эталоном бралось на границе базы притяжения. Задание начального пересечения полностью определяет значения вероятностей q_1 и q_0 (см. формулу (11)), а тем самым и начальную энтропию воспроизводимого эталона Nh_{in} , заданную формулой (9). Конечная энтропия воспроизводимого эталона определяется конечными значениями вероятностей $p_{\mu\nu}$. Во всех используемых методах влияние начального состояния нейрона на его конечное состояние игнорируется, т. е. p_{10} полагается равным p_{11} , а p_{00} полагается равным p_{01} . Тогда вероятности $p_{11} = p_{10} = p_1$ и

$p_{01} = p_{00} = p_0$ полностью определяются значением конечного пересечения m_f (см. уравнение (15)), и поэтому конечная энтропия воспроизводимого эталона Nh_f , заданная формулой (12), полностью определяется значением m_{in} (взятым для данного α на границе базы притяжения) и качеством воспроизведения, т. е. m_f .

Коэффициенты информационной эффективности E , рассчитанные по формуле (17) для различных значений p , показаны на рис. 14 толстыми линиями. Значения i_{in} , рассчитанные по формуле (18), которые задают начальную энтропию воспроизводимых эталонов, показаны тонкими линиями. Значения i_f , рассчитанные по формуле (19), задающие конечную энтропию эталонов, показаны штриховыми линиями. Значения $i'_f = \alpha h'_f / h(p)$, где h'_f рассчитано по формуле (16), показаны пунктирными линиями.

Когда информационная нагрузка α относительно мала, размеры баз притяжения велики ($h_{in} \simeq h(p)$), а качество воспроизведения близко к единице ($h_f \ll h(p)$). Тогда $E \simeq \alpha$, т. е. из сети извлекается почти столько же информации, сколько в нее заложено. Когда α возрастает, различие между α и E возрастает как за счет уменьшения баз притяжения, так и за счет уменьшения качества воспроизведения. Для $p = 0.5$ максимальная информационная эффективность составляет 0.092, что совпадает с результатом работы [16]. Для $p = 0.1$ она составляет 0.167, что превышает значение 0.143, полученное в работе [16]. Это различие не может объясняться тем фактом, что в этой работе при расчете конечной энтропии воспроизводимого эталона пренебрегается информацией, данной начальным паттерном (т. е. используется формула (16) вместо (12)). Такое пренебрежение наоборот приводит к завышению информационной эффективности. Основная причина расхождения нашей оценки информационной эффективности и оценки работы [16] объясняется тем, что в этой работе использовался метод, основанный на RM-методе, а этот метод дает заниженную оценку информационной емкости при $p = 0.1$ (см. раздел «Компьютерное моделирование», с. 48–64). При такой разреженности кодирования мы оценивали информационную эффективность SN-методом первого порядка, который, как показано на рис. 11а, дает намного более точную оценку всех характеристик сети. Различие между нашей оценкой и результатом работы [16] могло бы быть еще больше, если бы мы считали информационную эффективность для случая, когда начальные паттерны являются фрагментами эталонов, как это сделано в работе [16], а не для случая, когда начальные паттерны содержат активные нейроны, совпадающие и не совпадающие с активными нейронами эталонов, как это сделано в наших исследованиях.

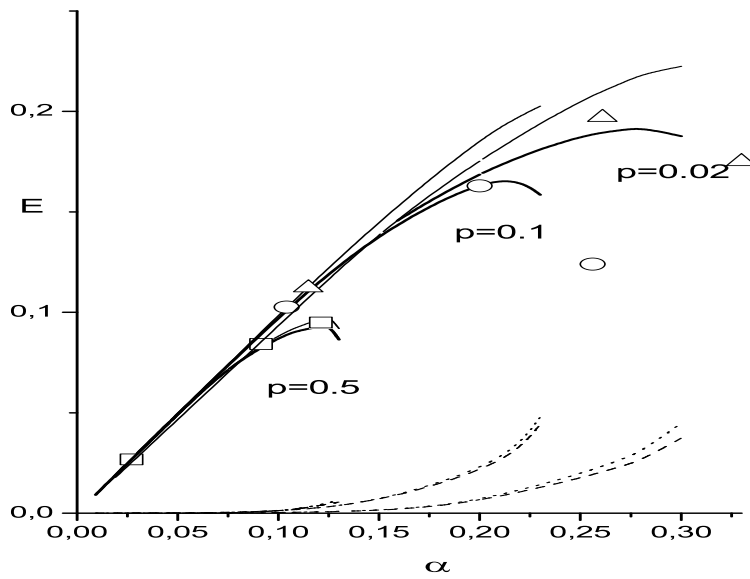


Рис. 14. Информационная эффективность E (сплошные толстые линии) в зависимости от α и p
 Рассчитаны по SN-методу второго порядка ($p = 0.5$), первого порядка ($p = 0.1$) и нулевого порядка ($p = 0.02$), и с помощью компьютерного моделирования (\diamond — для $p = 0.5$, \circ — для $p = 0.1$ и Δ — для $p = 0.02$). Тонкие сплошные линии — i_{in} ; штриховые линии — i_f ; пунктирные линии — i'_f .

Как хорошо известно [11, 16], в случае разреженного кодирования восстановление эталонов по их фрагментам позволяет извлечь из сети больше информации, чем их восстановление по начальным паттернам, содержащим активные и неактивные нейроны эталонов.

Для $p = 0.02$ информационная эффективность оценивалась SS-методом. Ее максимальное значение составило 0.192. Компьютерное моделирование дает даже слегка большее значение.

Для $p \rightarrow 0$ уравнения (9–11) и (18) дают $i_{in} = \alpha(1 - m_{in})$. В этом пределе все три метода SS, SN и RM предсказывают высокое качество воспроизведения (т.е. $i_f = 0$) [10]. Оба метода SS и SN (см. уравнение (26)) предсказывают, что размер области притяжения в этом случае задается

формулой $m_{in} = \sqrt{2\alpha \ln 2}$, а потому $E = \alpha(1 - \sqrt{2\alpha \ln 2})$. Максимальное значение информационной емкости составляет $E = 2/(27 \ln 2) \simeq 0.11$. Таким образом, существует оптимальная разреженность кодирования, при которой информационная эффективность максимальна. Эта эффективность по крайней мере не менее значения 0.2, достигаемого при $p = 0.02$.

Заключение

В настоящей работе мы исследовали аналитически и с помощью компьютерного моделирования влияние разреженности на размер областей притяжения в полносвязной автоассоциативной сети хопфилдового типа. Компьютерное моделирование показало, что существующие аналитические подходы недостаточно точны при больших разреженностях. Для $p = 0.02$ траектории, описывающие активность сети, демонстрируют парадоксальное поведение: пересечение с восстанавливаемым эталоном вначале убывает, но потом траектория возвращается и активность сети приближается к восстанавливаемому эталону. Поэтому, размер области притяжения оказывается даже больше, чем получаемый в SS-приближении. Такое поведение не предсказывается существующими аналитическими методами. Основной причиной их неспособности отразить это свойство является игнорирование различия в поведении нейронов, активных и неактивных на предыдущих шагах.

Компьютерное моделирование показало, что размер области притяжения меняется немонотонно по мере возрастания разреженности: вначале он возрастает, а затем убывает. Поэтому в пределе $p \rightarrow 0$ разреженность ухудшает способность к коррекции искаженных эталонов, хотя улучшает как информационную емкость, так и качество воспроизведения. Количество информации I_g , извлекаемой из сети за счет коррекции предъявленных искаженных вариантов записанных эталонов, используется как обобщенная характеристика способности сети к выполнению функции ассоциативной памяти. Существует оптимальная разреженность (p порядка 10^{-2}), для которой информационная эффективность $E = I_g/N^2$ максимальна. Эта эффективность составляет около 0.2. Интересно, что оптимальная разреженность соответствует нейронной активности мозга. Мы полагаем, что это совпадение не случайно и отражает важную особенность работы мозга.

Максимальная информационная эффективность оказывается примерно вдвое выше, чем для плотного кодирования ($p = 0.5$) и чем в предельном

случае ($p \rightarrow 0$). Однако, она остается значительно ниже максимального предела информационной эффективности $E = 1/(4 \ln 2) \simeq 0.36$, задаваемого энтропией матрицы связей [11]. Это означает, что в принципе процедура воспроизведения может быть улучшена путем повышения количества информации, извлекаемой при коррекции искаженных паттернов. С другой стороны, оценка информационной эффективности, сделанная в данной работе, относится только к части записанной информации, извлекаемой из сети в процессе воспроизведения. Другая часть информации может быть извлечена за счет распознавания новых (далеких от записанных эталонов) и знакомых (близких к одному из записанных эталонов) паттернов среди тестирующего множества входных паттернов информации [11]. Таким образом, перед тем, как пытаться улучшить процедуру коррекции, уместно оценить также количество информации, которая может быть извлечена из сети за счет распознавания входных паттернов как новых и знакомых.

Литература

1. Дунин-Барковский В. Л. Информационные процессы в нейронных структурах. – М.: Наука, 1978. – 166 с.
2. Фролов А. А., Мушинский А. М., Цодыкс М. В. Имитационная модель автоассоциативной памяти в виде нейронной сети с малым уровнем активности // *Биофизика*. – 36. – с. 339–343.
3. Amari S. Characteristics of sparsely encoded associative memory // *Neural Networks*. – 1989. – 2. – pp. 451–457.
4. Amari S., Maginu K. Statistical neurodynamics of associative memory // *Neural Networks*. – 1988. – 1. – pp. 63–73.
5. Amit D. J., Gutfreund H., Sompolinsky H. Statistical mechanics of neural networks near saturation // *Annals of Physics*. – 1987. – 173. – pp. 30–67.
6. Buhmann J., Divko R., Schulten K. Associative memory with high information content // *Physical Review A*. – 1989. – 39. – pp. 2689–2692.
7. Coolen A. C. C., Sherington D. Order-parameter flow in the fully connected Hopfield model near saturation // *Physical Review E*. – 1993. – 49. – pp. 1921–1934.
8. Forrest B. M. Content-addressability and learning in neural networks // *Journal of Physics A*. – 1988. – 21. – pp. 245–255.
9. Frolov A. A., Husek D., Muraviev I. P. Statistical neurodynamics of sparsely encoded Hopfield-like associative memory // *Neural Netw. World*. – 1996. – 6. – pp. 609–617.
10. Frolov A. A., Husek D., Muraviev I. P. Informational capacity and recall quality in sparsely encoded Hopfield-like neural network: Analytical approaches and computer simulation // *Neural Networks*. – 1997. – 10. – pp. 845–855.

11. Frolov A. A., Muraviev I. P. Informational characteristics of neural networks capable of associative learning based on Hebbian plasticity // *Network*. – 1993. – **4**. – pp. 495–536.
12. Gardner E., Derrida B., Mottishaw P. Zero temperature parallel dynamics for infinite range spin glasses and neural networks // *J. Physique*. – 1987. – **48**. – pp. 741–755.
13. Goles-Chacc E., Fogelman-Soulie F., Pellegrin D. Decreasing energy functions as a tool for studying threshold networks // *Discrete Mathematics*. – 1985. – **12**. – pp. 261–277.
14. Hopfield J. J. Neural network and physical systems with emergent collective computational abilities // *Proceedings of the National Academy of Science, USA*. – 1982. – **79**. – pp. 2544–2548.
15. Horner H. Neural networks with low levels of activity: Ising vs. McCulloch-Pitts neurons // *Z. Physik B*. – 1989. – **75**. – pp. 133–136.
16. Horner H., Bormann D., Frick M., Kinzelbach H., Schmidt A. Transients and basins of attraction in neural network models // *Z. Physik B*. – 1989. – **76**. – pp. 381–398.
17. Husek D., Frolov A. A. Evaluation of the informational capacity of Hopfield network by computer simulation // *Neural Network World*. – 1994. – **4**. – pp. 53–65.
18. Kinzel W. Learning and pattern recognition in spin glass models // *Z. Physik B*. – 1985. – **60**. – pp. 205–213.
19. Kohring G. A. A high-precision study of the Hopfield model in the phase of broken replica symmetry // *Journal of Statistical Physics*. – 1990. – **59**. – pp. 1077–1086.
20. Kohring G. A. Convergence time and finite size effects in neural networks // *Journal of Physics A: Math. Gen.* – 1990. – **23**. – pp. 2237–2241.
21. Koyama H., Fujie N., Seyama H. Results from the Gardner–Derrida–Mottishaw theory of associative memory *Neural Networks*. – 1999. – **12**. – pp. 247–257.
22. Okada M. Notions of associative memory and sparse coding // *Neural Networks*. – 1996. – **9**. – pp. 1429–1458.
23. Palm G., Sommer F. T. Information capacity in recurrent McCulloch–Pitts network with sparsely coded memory states // *Network*. – 1992. – **3**. – pp. 177–186.
24. Perez-Vicente C. J., Amit D. J. Optimized network for sparsely coded patterns // *Journal of Physics A: Math. Gen.* – 1989. – **22**. – pp. 559–569.
25. Shigemitsu S., Okada M. Statistical neurodynamics of the sparsely encoded associative memory (in Japanese) // *The Brain & Neural Networks*. – 1996. – **3**. – pp. 58–64.
26. Tsodyks M. V., Feigelman M. V. The enhanced storage capacity in neural network with low activity level // *Europhysical Letters*. – 1988. – **6**. – pp. 101–105.

Александр Алексеевич ФРОЛОВ, кандидат физико-математических, доктор биологических наук, профессор, заведующий лабораторией математической нейробиологии обучения ИВНД и НФ РАН, специалист в области биомеханики и нейросетевого моделирования функций нервной системы (двигательное управление, ассоциативная память, биофизические механизмы обучения и памяти), автор более 200 статей, 6 изобретений и 2 монографий (по нейросетевым моделям ассоциативной памяти).

Душан ГУСЕК, научный сотрудник Института информатики Академии наук Чешской республики. Область научных интересов: нейронные сети, представление знаний, представление информации и базы данных. Автор более 50 научных публикаций.

URL: <http://www.cs.cas.cz/~dusan>

Игорь Петрович МУРАВЬЕВ, научный сотрудник лаборатории математической нейробиологии обучения ИВНД и НФ РАН, специалист в области нейросетевого моделирования ассоциативной памяти, автор более 20 статей и 2 монографий (по нейросетевым моделям ассоциативной памяти).