

## Биологический алгоритм распознавания сильно скоррелированных образов

Академик А.Л.Микаэлян, Б.В.Крыжановский

Работа выполнена при поддержке РФФИ (проекты №02-01-00457, №01-07-90134) и программы "Интеллектуальные компьютерные системы" (проект 2.45).

**Аннотация.** Проведен анализ распознающей способности нейросети, способной хранить и обрабатывать информацию, закодированную в виде частотно-фазовой модуляции. Информативные сигналы в рассматриваемой сети передаются по межнейронным связям в виде квазимонохроматических импульсов на  $n$  разных частотах. За основу такой сети принят "параметрический" нейрон – обладающий кубической нелинейностью элемент, способный к преобразованию и генерации частот в процессах параметрического 4-волнового смешения. Показано, что с ростом числа несущих частот помехозащищенность рассматриваемой ассоциативной памяти резко возрастает. Одновременно резко возрастает и объем нейросетевой памяти, которая в  $n^2$  раз больше аналогичной величины в стандартной сети Хопфилда. Число образов, которые способна сохранять такая нейросеть, может во много раз превышать число нейронов.

Стандартные нейронные сети не приспособлены для распознавания сильно скоррелированных образов и так называемых *biased patterns*. Помимо этого они обладают малым объемом памяти. Так например, сеть Хопфилда [1] может хранить всего лишь  $M \approx N/2 \ln N$  рандомизированных  $N$ -мерных образов. При наличии корреляции между образами объем памяти ( $M$ ) резко уменьшается. Имеющиеся отдельные алгоритмы для распознавания похожих образов, например метод проекционной матрицы [2], достаточно сложны, и не позволяют ввести простое обучающее правило, обладающее биологическим принципом локальности [3].

В то же время, человек достаточно легко выделяет образ среди множества похожих даже при наличии больших искажений. Такую способность можно объяснить, положив в основу модели распознавания принципы функционирования описанного Розенблатом [4] биологического фото-перцептрона: а). Воздействия попадают на ретину (рис.1), которая в одних моделях работает по принципу "все или ничего" (т.е. выдает одиночные импульсы при надпороговых воздействиях), а в других моделях – по принципу частотной или амплитудной модуляции. б). Импульсы передаются в область ассоциации, представляющую собой набор связанных между собой блоков. Если алгебраическая

сумма поступающих на блок подавляющих и возбуждающих сигналов больше некоторого порога, то блок работает по принципу "все или ничего" или выдает частоту, соответствующую количеству и временному порядку принятых сигналов.

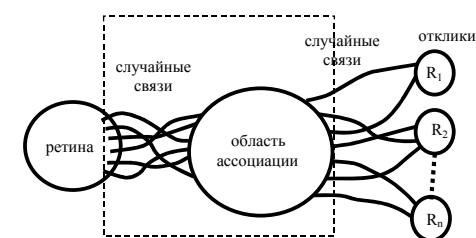


Рис. 1. Биологический фото-перцептрон. Пунктиром выделена часть, моделируемая векторной нейросетью.

Из возможных вариантов формального описания биологической модели мы выберем один, наиболее оптимальный с нашей точки зрения. Во-первых, мы постулируем бинарность сигналов, формируемых в ретине, и случайный характер их передачи в ассоциативную область. Во-вторых, мы примем, что в блоках ассоциативной области формируются частотно-модулированные сигналы, которыми эти блоки и обмениваются. В принятых допущениях распознавание образа можно, условно, разбить на два этапа. На первом, набор бинарных сигналов по случайным связям попадает на блоки ассоциативной области, где преобразуется в набор частотно-модулированных сигналов, т.е. набор векторов. На втором этапе, происходит распознавание образа ассоциативной памятью: блоки ассоциативной памяти обмениваются частотно-модулированными (векторными) сигналами до тех пор, пока система не придет в стабильное состояние, соответствующее распознанному образу.

Как будет видно далее, преобразование набора бинарных сигналов в набор векторных сигналов – это достаточное условие подавления негативного влияния корреляции на распознавание образов. Формальное описание этого процесса приведено в следующем пункте. Для описания работы ассоциативной области мы используем параметрическую модель нейронной сети, способной обрабатывать информацию, закодированную в виде частотно-фазовой модуляции [5]. За основу такой сети принят "параметрический" нейрон [6] – обладающий кубической нелинейностью элемент, способный к преобразованию и генерации  $Q$  частот в процессах параметрического четырехволнового смешения. Параметрическим нейроном мы будем моделировать работу целого блока ассоциативной памяти. Такой подход обоснован практически установленным фактом, что базовыми функциональными элементами, отвечающими за высокоуровневую деятельность

кору головного мозга, являются так называемые корковые колонки (блоки): сильно связанные группы нейронов, обладающие коллективными свойствами и способные к смешению частот и обработке частотно-модулированных сигналов (см. [7-10]). В [11] показано, что набору из  $Q$  частот можно поставить в соответствие набор  $Q$  ортогональных векторов и описание параметрической нейросети, оперирующей частотно-модулированными сигналами, свести к описанию системы взаимодействующих спинов. Поэтому, дальнейшее описание мы проведем на языке векторной (спиновой) модели, более привычном для нейронных сетей. Формализм предлагаемой модели описан в п.3.

Мы покажем, что параметрическая нейросетевая модель, соответствующая описанной выше биологической модели, обладает огромным объемом памяти и способностью распознавать образы даже при исключительно больших искажениях и наличии корреляции. Суть предлагаемого здесь формального описания состоит в следующем. Пусть имеется семейство  $N$ -мерных бинарных векторов  $\{Y_m\}$ , ( $m=1,2,\dots,M$ ), искаженные образы которых предстоит распознавать. Необходимая для этого ассоциативная память организуется следующим образом: каждому образу  $Y_m$  из пространства  $R^N$  ставится в однозначное соответствие образ  $X_m$  в некоем пространстве  $\mathcal{H}$  большей размерности; на семействе  $\{X_m\}$  строится ассоциативная память в виде описываемой ниже векторной нейросети. Процесс распознавания производится в следующем порядке: распознаваемый бинарный вектор  $Y \in R^N$  отображается в образ  $X \in \mathcal{H}$  и отображение предьявляется для распознавания векторной нейросети; при необходимости производится обратное отображение распознанного образа из  $\mathcal{H}$  в изначальное  $N$ -мерное пространство. Таким образом, задача распознавания большого числа бинарных коррелированных векторов сводится к задаче распознавания их отображений.

Алгоритм отображения, позволяющий использовать векторную модель для распознавания сильно скоррелированных бинарных векторов, состоит в следующем. Пусть имеется некий  $N$ -мерный бинарный вектор  $Y = (y_1, y_2, \dots, y_N)$ . Мысленно разделим его на  $n$  фрагментов, содержащих по  $k+1$  элементов каждый. Отдельный фрагмент можно рассматривать как целое число  $\pm q$ , записанное в двоичном коде: первый элемент фрагмента определяет знак (0 - знак "минус", 1 - "плюс"), а остальные  $k$  элементов - величину  $q$  (параметр  $k$  будем называть параметром отображения). Теперь фрагменту поставим в соответствие вектор  $\mathbf{x} = \pm \mathbf{e}_q$ , где  $\mathbf{e}_q$  - это  $q$ -й орт некоторого  $Q$ -мерного пространства ( $Q=2^k$ ). Тем самым, всему образу  $Y \in R^N$  в целом ставится в однозначное соответствие набор  $Q$ -мерных векторов, т.е. образ  $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ . Например, вектор  $Y=(01001001)$  можно разбить на два фрагмента по четыре элемента (0100) и (1001). Первому фрагменту (это "-4" в двоичном коде) ставим в соответствие вектор  $\mathbf{x}_1 = -\mathbf{e}_4$  в пространстве размерностью  $Q=8$ , а второму (это "+1" в двоичном коде) -

вектор  $\mathbf{x}_2 = +\mathbf{e}_1$ . Соответствующее отображение примет вид  $Y \rightarrow X = (\mathbf{x}_1, \mathbf{x}_2)$ . Существенно, что описываемое отображение взаимно однозначно, т.е. распознав отображение  $X$  можно однозначно восстановить его бинарный прообраз  $Y$ . Еще более существенно то, что процедура отображения практически сводит на нет имеющиеся корреляции. Например, рассмотрим два бинарных фрагмента (0000000001) и (0000000011), скоррелированных на 90%. Отличие фрагментов в одном только элементе приводит полному исчезновению корреляции между их отображениями в пространстве  $\mathcal{H}$ , какowymi являются различные орты  $\mathbf{e}_1$  и  $\mathbf{e}_2$  соответственно.

Рассмотрим полносвязную нейронную сеть из  $n$  нейронов, описываемых единичными векторами  $\mathbf{x}_i = x_i \mathbf{e}_q^{(i)}$ , где  $x_i = \pm 1$ ,  $\mathbf{e}_q^{(i)}$  - орт  $Q$ -мерного пространства,  $i = 1, 2, \dots, n$ . Состояние сети как целого определяется набором таких векторов  $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ .

Гамильтониан сети зададим в виде [11], аналогичном модели Хопфилда

$$H = -\frac{1}{2} \sum_{i,j=1}^n \mathbf{x}_i^+ \hat{T}_{ij} \mathbf{x}_j, \quad \hat{T}_{ij} = (1 - \delta_{ij}) \sum_{m=1}^M \mathbf{x}_{mi} \mathbf{x}_{mj}^+ \quad (1)$$

где  $\mathbf{x}_i$  - вектор-столбец,  $\mathbf{x}_i^+$  - вектор-строка, а величина межсвязи  $\hat{T}_{ij}$  между  $i$ -м и  $j$ -м нейронами -  $q \times q$  матрица, построенная по аналогии с обучающим правилом Хэбба [3] на эталонных образах  $X_m = (\mathbf{x}_{m1}, \mathbf{x}_{m2}, \dots, \mathbf{x}_{mn})$ ,  $m = 1, 2, \dots, M$ . Сеть (1) удобно интерпретировать как систему взаимодействующих  $Q$ -мерных спинов и использовать соответствующую терминологию. С учетом (1) входной сигнал на  $i$ -й нейрон, т.е. локальное поле действующее на  $i$ -й спин со стороны сети, запишется в виде:

$$\mathbf{h}_i = \sum_{j=1}^n \hat{T}_{ij} \mathbf{x}_j = \sum_{q=1}^Q A_q^{(i)} \mathbf{e}_q, \quad A_q^{(i)} = \sum_{j \neq i} \sum_{m=1}^M (\mathbf{e}_q^+ \mathbf{x}_{mi}) (\mathbf{x}_{mj}^+ \mathbf{x}_j) \quad (2)$$

Динамика физической системы определяется естественным образом:  $i$ -й спин под воздействием магнитного поля  $\mathbf{h}_i$  принимает положение, наиболее близкое к направлению этого поля, т.е. состояние  $i$ -го нейрона в момент времени  $t+1$  описывается выражением:

$$\mathbf{x}_i(t+1) = s \mathbf{e}_{\max}, \quad s = \text{sign}[A_{\max}^{(i)}(t)] \quad (3)$$

где индексом  $\max$  обозначена максимальная по модулю амплитуда  $A_q^{(i)} = A_q^{(i)}(t)$  в разложении (2). Динамика системы в целом состоит в последовательном изменении

состояний нейронов по правилу (3) и соответствует понижению энергии системы в процессе ее функционирования, т.е. алгоритм (3) сходится.

Определим, насколько эффективно такая нейросеть распознает искаженные образы. Пусть на вход системы подан искаженный  $m$ -й образ, т.е. начальные состояния нейронов сети заданы в виде  $\mathbf{x}_i = \hat{a}_i \hat{b}_i \mathbf{x}_{mi}$ , где  $\hat{a}_i$  - оператор мультипликативного шума, который с вероятностью  $a$  изменяет знак амплитуды  $x_{mi}$  вектора  $\mathbf{x}_{mi} = x_{mi} \mathbf{e}_{mi}$  и с вероятностью  $1-a$  оставляет его неизменным, оператор  $\hat{b}_i$  - с вероятностью  $b$  заменяет орт  $\mathbf{e}_{mi} \in \{\mathbf{e}_q\}$  на любой иной из набора  $\{\mathbf{e}_q\}$  и с вероятностью  $1-b$  оставляет его неизменным. Сеть правильно распознает эталонный образ  $X_m$ , если выход  $i$ -го нейрона, определяемый выражением (3), будет  $\mathbf{x}_i = \mathbf{x}_{mi}$ . В противном случае произойдет ошибка распознавания, т.е. сеть вместо  $X_m$  распознает иной образ. Для вероятности  $P$  этой ошибки, используя метод Чебышева-Чернова [13], детально описанный для данного рода задач в работах [5,6], получим:

$$P \leq n \exp \left[ -\frac{nQ^2}{2M} (1-2a)^2 (1-b)^2 \right] \quad (4)$$

Полученное неравенство устанавливает верхнюю границу для средней вероятности ошибки в рассматриваемой нами нейронной сети с параметрами  $(n; M; Q; a; b)$ . С ростом  $n$  эта граница сходится к нулю всякий раз, когда величина  $M$  как функция  $n$  растет медленнее, чем

$$\bar{M} = nQ^2 \frac{(1-2a)^2 (1-b)^2}{2 \ln n} \quad (5)$$

Это дает основание рассматривать величину (5) как асимптотически достижимую мощность ассоциативной памяти анализируемой нейронной сети. Сравнение (5) с аналогичными выражениями для параметрической оптической модели [5] и модели Поттса [12] показывает, что предложенная модель имеет в два раза больший объем памяти и, при прочих равных параметрах, может распознавать образы, искаженные на 20-30% сильнее.

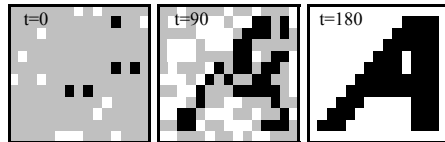


Рис.2 Распознавание буквы "А", у которой искажены 90% пикселей (выделены серым цветом).

Из (5) видно, что с ростом  $Q$  помехозащищенность рассматриваемой ассоциативной памяти резко возрастает. Одновременно резко возрастает и объем нейросетевой памяти, в  $Q^2$  раз больший чем в сети Хопфилда. Рис.2 демонстрирует большой объем памяти и высокую помехоустойчивость на примере сети из 180 нейронов с  $Q=32$ , в памяти которой записано 360 образов (32-цветных изображений), один из них - стилизованная буква "А". Сеть надежно распознает образ "А", у которого искажено 90% компонент за один цикл. При меньших искажениях ( $b \leq 70\%$ ) эта же сеть распознает до 1800 образов.

Обратимся теперь к проблеме распознавания бинарных образов. Задав некоторое значение параметра деления  $k$  и применив описанное выше отображение к набору бинарных векторов  $\{Y_m\} \in R^N$ ,  $m \in \overline{1, M}$ , получим соответствующий набор образов  $\{X_m\} \in \mathfrak{X}$ , на основе которых построим векторную ассоциативную память с параметрами: число нейронов векторной сети -  $n = N/(k+1)$ , число состояний векторного нейрона -  $Q = 2^k$ . Анализ проведем на примере "тенденциозных" образов (biased patterns), компоненты которых  $y_{mi}$  - случайные величины, принимающие значения 1 и 0 с вероятностями  $(1+\alpha)/2$  и  $(1-\alpha)/2$  соответственно, ( $-1 \leq \alpha \leq 1$ ). Пусть нам предстоит распознать искаженный  $m$ -й образ  $\tilde{Y}_m = (s_1 y_{m1}, s_2 y_{m2}, \dots, s_N y_{mN})$ , где случайная величина  $s_i$  с вероятностью  $p$  изменяет значение бинарной переменной  $y_{mi}$  и с вероятностью  $1-p$  оставляет ее неизменной. Отображением этого вектора в пространстве  $\mathfrak{X}$  является искаженный  $m$ -й образ  $\tilde{X}_m$ , который и предъявляется для распознавания векторной нейросети. Выражая мультипликативные шумы  $a$  и  $b$ , покрывающие отображение, как функции параметра  $p$  и подставляя соответствующие выражения в (4) для вероятности ошибки распознавания искаженного отображения  $\tilde{X}_m$  получим:

$$P = n \exp \left[ -\frac{\nu}{2\mu} (1-\alpha^3 \mu)^2 \right] \quad (6)$$

где

$$\nu = n(1-2p)^2 (1-p)^k, \quad \mu = MA^k / (1-p)^k, \quad A = (1+\alpha^2)[1+\alpha^2(1-2p)]/4.$$

При  $k=0$  выражение (6) описывает функционирование модели Хопфилда. Анализ (6) для данного случая показывает, что даже в отсутствие корреляций ( $\alpha = 0$ ) объем памяти не превышает относительно малого значения  $M_0 \approx N/2 \ln N$ . А наличие даже небольшой корреляции ( $\alpha > N^{-1/3}$ ) уменьшает число распознаваемых образов до величины порядка  $\alpha^{-3}$ , т.е. сеть практически перестает выполнять функции ассоциативной памяти.

С ростом параметра отображения  $k$  картина резко меняется. Сеть начинает функционировать как векторная модель, т.е. резко повышается объем памяти и снижается влияние корреляции. В частности, при небольших корреляциях, когда  $\alpha^3 < v$ , для объема памяти из (6) получаем оценочное выражение:

$$\bar{M} = M_0[(1-p)^2 / A]^k$$

При большей корреляции, когда  $\alpha^3 > v$ , объем памяти несколько ниже:

$$\bar{M} = \alpha^{-3}[(1-p) / A]^k$$

Однако и в том, и в другом случаях с ростом  $k$  имеет место экспоненциальный рост числа распознаваемых образов (рис.3) и рост надежности распознавания. На рис.4 показано, как с ростом параметра отображения спадает до нуля вероятность ошибки распознавания (кривые построены для корреляций  $\alpha=0.1, 0.2, 0.5, 0.6$  при  $M/N=2$  и искажениях  $p=20\%$ ). Как видим, при достижении некоторого критического значения параметра отображения  $k$  вероятность ошибки резко спадает, т.е. негативное влияние корреляции резко уменьшается.

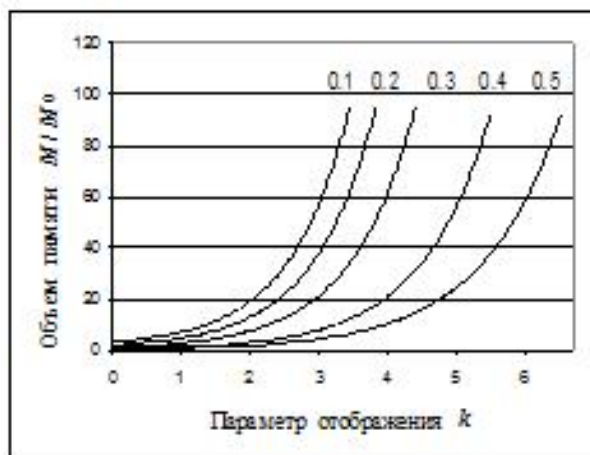


Рис.3 Рост объема памяти с ростом параметра отображения  $k$  ( $p=0.1 \div 0.5$ ).

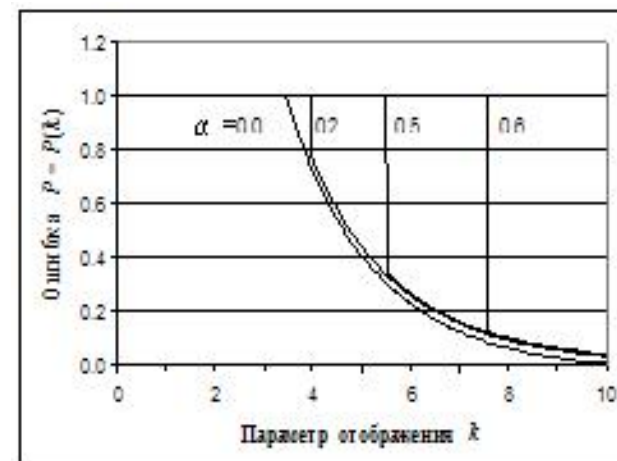


Рис.4 Уменьшение ошибки распознавания с ростом параметра отображения  $k$

Как видим, соответствующая биологическому прототипу параметрическая модель демонстрирует большой объем памяти и способность распознавать похожие образы. Основное допущение при моделировании состояло в том, что бинарные сигналы от фоторецепторов преобразуются в частотно-модулированные сигналы, которыми оперирует ассоциативная память. В проведенном выше анализе мы никак не использовали случайность связей между ретиной и ассоциативной областью, хотя она может играть большую роль в декорреляции распознаваемых образов. Действительно, в большинстве случаев образы заполняют сплошь целые фрагменты рецептивного поля и топологическое отображение в векторное пространство при небольшом значении параметра  $k$  не приводит к декорреляции. Однако, случайность передачи сигналов от рецепторов в ассоциативную область сводит на нет такую корреляцию. На алгоритмическом языке сказанное означает, что нумерацию компонент бинарных векторов полезно производить случайным образом, чтобы избежать фрагментарной корреляции.

Очевидно, что процесс распознавания образов мозгом значительно сложнее рассмотренной выше модели. Однако, если эта модель хоть как-то соответствует реальности, то можно утверждать, что размер биологической ассоциативной памяти и ее распознающая способность на порядки выше оценок, предлагаемых бинарными моделями, не учитывающими частотно-модулированный характер кодировки информации.

Действительно, нейронная колонка коры головного мозга (в нашей модели - это  $Q$ -мерный нейрон) содержит около 100 нейронов, соединенных возбуждающими и тормозящими связями, и может генерировать сигналы на различных частотах, число которых можно оценить как  $Q \sim 20 \div 40$ . Как следует из (5), при таком количестве частот ассоциативная память из таких колонок имеет огромный объем. Даже при весьма умеренном числе частот  $Q \sim 10$  объем ассоциативной памяти почти на два порядка превышает значения, характерные для сетей Хопфилда (см. рис.3). Проведем некоторые оценки. Характерный линейный размер нейронной колонки порядка 400 мкм. При скорости распространения сигналов по межсвязям  $\sim 0.1$  м/с возбуждение за время  $\sim 1.5$  мс (длительность нервных импульсов) охватывает пространство с линейными размерами  $\sim 1$  мм, на котором размещается порядка  $n \sim 30 \div 50$  колонок, вовлекая их в процесс одновременного возбуждения и анализа информации. Это означает, что участок коры головного мозга площадью  $\sim 1$  мм<sup>2</sup> способен запомнить  $M \sim 10^3 \div 10^4$  бинарных 150-мерных образов и в течение нескольких миллисекунд распознавать один из них.

## Литература

1. Hopfield J.J. // Proc.Nat.Acad.Sci.USA. 1982. V.79. P.2554-2558.
2. Personnaz L., Guyon H., Dreyfus G. // Phys.Rev.A. 1987. V.34. P.4217-4227.
3. Hebb D.O. The Organization of Behavior. N.Y.: Wiley, 1949.
4. Rozenblatt F. // Psychological Review. 1958. V.65. P.368-408.
5. Крыжановский Б.В., Микаэлян А.Л. // ДАН. 2002. Т. 383, №3, с.318-321.
6. Kryzhanovsky B.V., Mikaelian A.L. et al. // Opt.Mem.&Neural Nets. 2001. V.10. P.211-218.
7. Annios P.A., Beek B., Csermely T.J. and Harth E.M. // J.Theor.Biol.. 1970. V. 26. P.121-148.
8. Usher M., Schuster H.G. and Neibur E. // Neural Computation. 1993. V.5, P.370-386.
9. Farhat N. // SPIE'2000, San-Diego, 2000. P. 158-170,.
10. Hoppensteadt F.C., Izhikevich E.M. // IEEE Trans.Neural Nets. 2000. V.11. P.734-738.
11. Крыжановский Б.В., Литинский Л.Б. // Искусственный интеллект. 2002. Т.4. С.710-718.
12. Kanter I. // Phys.Rev.A. 1988. V.37(7). P. 2739-2742.
13. Chernov N. // Ann. Math. Stat. 1952. V.23. P.493-507.