РОССИЙСКАЯ АКАДЕМИЯ НАУК

СИБИРСКОЕ ОТДЕЛЕНИЕ

А В Т О М Е Т Р И Я

№ 1 1998

УДК 519.68

С. Г. Пудов

(Новосибирск)

ВЛИЯНИЕ АВТОСВЯЗИ НА УСТОЙЧИВОСТЬ КЛЕТОЧНО-НЕЙРОННОЙ АССОЦИАТИВНОЙ ПАМЯТИ

Рассматривается клеточно-нейронная ассоциативная память (КНАП), которая представляет собой нейронную сеть Хопфилда со структурой связей клеточного автомата. Получены формулы для расчета весов автосвязи, которые обеспечивают максимально сильную устойчивость сети к искажениям заданного прототипа. Также предложены две стратегии выбора окончательной величины веса автосвязи из величин, рассчитанных для всех прототипов по отдельности. Полученные результаты справедливы для полносвязной ассоциативной памяти, обученной с помощью любого метода.

Введение. Клеточно-нейронная ассоциативная память (КНАП) представляет собой нейронную ассоциативную память Хопфилда [1] со структурой связей клеточного автомата. Большинство из существующих методов обучения, разработанных для сетей Хопфилда, не применимы для КНАП, поскольку они основываются на решении матричных уравнений и производят полносвязную матрицу весов связей [2, 3]. Методы, в которых принимается во внимание структура связей нейросети, можно разбить на две неравные группы по критерию определения автосвязи. В первую группу входит метод обучения Мишеля [4, 5], основывающийся на сложных матричных преобразованиях, причем получаемая матрица весов связей зависит от произвольно выбираемого параметра. Этот метод определяет ненулевые веса автосвязи, однако из-за произвола в выборе параметра ничего нельзя сказать об оптимальности полученных весов с точки зрения устойчивости к искажениям.

Во вторую группу входят методы, такие как Хебба [1], персептрона, и алгоритм, предложенный автором в [6], которые не определяют автосвязь (т. е. полагают этот вес равным нулю). Метод Хебба является наиболее простым и при этом имеет ряд существенных недостатков, таких как малая способность к хранению информации (порядка 0,15 от числа связей нейронов) и отсутствие гарантии индивидуальной устойчивости хранимых образов. Метод обучения персептрона [7] адаптирован для обучения КНАП [6], при этом гарантируется индивидуальная устойчивость хранимых прототипов и, более того, их количество может превышать число соседей нейрона. Однако нейросеть, обученная по этому методу, обладает низкой устойчивостью к искажениям хранимых образов. В связи с этим предложена модификация этого метода [6], которая ориентирована на повышение устойчивости к искажениям при сохранении достоинств базового метода: простоты, большого числа хранимых образов, а также гарантии их индивидуальной устойчивости. В разработанном методе, как и в случае обучения персептрона, автосвязь не используется. Однако было бы заманчиво ее использовать, поскольку можно предположить, что она окажет существенное влияние на характеристики сети: ускорит сходимость, уменьшит число колебаний и улучшит устойчивость (т. е. окажет стабилизирующее влияние подобно обратной связи). Кроме того, связь «сам с собой» реализуется проще, чем связь с каким-либо соседом. Поэтому возникает идея: рассчитывать веса автосвязи после завершения алгоритма обучения. Это особенно актуально именно для КНАП, поскольку число связей здесь ограничено.

В данной работе исследуется влияние автосвязи на одну из основных характеристик КНАП — устойчивость к искажениям. Целью работы является получение алгоритмов (или формул) для расчета весов автосвязи, максимально улучшающих устойчивость клеточно-нейронной ассоциативной памяти к искажениям хранимых образов. Следует отметить, что результаты, полученные в этой работе, могут быть применены и к полносвязной сети Хопфилда, обученной по одному из методов. Здесь так же, как и в случае обучения КНАП, существующие алгоритмы не определяют веса автосвязи оптимальным образом. В качестве примера можно привести метод проекторов [2]. В [8] было показано, что для полносвязной сети, обученной по этому методу, уменьшение влияния автосвязи улучшает устойчивость к искажениям, уменьшает время сходимости, а также расширяет способности сети по хранению информации. Таким образом, правильно изменив только веса автосвязей, можно добиться серьезных улучшений качеств ассоциативной памяти.

Статья организована следующим образом: в разд. 1 дается формальное представление КНАП, в разд. 2 и 3 получаются веса автосвязей для случаев 1-искажений и k-искажений соответственно.

1. **Формальное представление КНАП.** Следуя [9], определим КНАП тройкой: $\mathbf{N} = \langle \mathbf{C}, \mathbf{W}, \Phi \rangle$, где $\mathbf{C}$ — прямоугольный массив, состоящий из m строк и n столбцов клеток (или нейронов) с состояниями $c_{ij} \in \{-1, 1\}$; $\mathbf{W} = \{W_{ij}\}$ — множество *весовых векторов* вида $W_{ij} = (w_1, ..., w_q)$, где w_k является вещественным числом, определяющим вес связи между нейроном с координатами (i, j) и его k-м соседом; Φ — правило функционирования КНАП.

Множество пар координат формирует *пространство имен* M, на котором определены *именующие функции* $\varphi(i, j)$: $M \to M$. Множество всех клеточных массивов с одним и тем же пространством имен M обозначается как $\mathbf{C}(M)$. Для каждой клетки (i, j) множество других клеток, которые связаны с ней, составляют ее соседство, определяемое *шаблоном связей*:

$$T(i, j) = \{\varphi_1(i, j), ..., \varphi_q(i, j)\}, \tag{1}$$

где $\varphi_k(i, j) \neq \varphi_l(i, j)$ для каждого $(i, j) \in M$ и для всех $k, l = 1, ..., q,\ k \neq l$.

Для массива $C \in \mathbf{C}(M)$ состояния всех соседей нейрона (i, j) представляются в виде вектора $C_{ij} = (c_{\varphi_1(i,j)}, ..., c_{\varphi_q(i,j)})$, или для краткости $C_{ij} = (c_1, ..., c_q)$. В дальнейшем также будет использоваться вектор

$$D_{ij} = c_{ij} C_{ij} = (d_1, ..., d_q), \tag{2}$$

называемый *приведенным соседством* нейрона (i, j). Оба вектора C_{ij} и W_{ij} состоят из q элементов каждый, нумерация которых порождена нумерацией именующих функций в (1), и их скалярное произведение определяется как $\langle C_{ij}, W_{ij} \rangle = \sum_l c_l w_l$.

Правило Φ функционирования КНАП описывается следующей итеративной процедурой.

Процедура. Пусть $C(t) \in C(M)$ обозначает массив после итерации t, C_{ij} и W_{ij} — состояния соседей и весовой вектор нейрона (i, j) соответственно. Тогда

1) все клетки в $C(t)$ вычисляют следующую функцию:

$$f(C_{ij}, W_{ij}) = \begin{cases} 1, & \text{если } \langle C_{ij}, W_{ij} \rangle \geq 0, \\ -1 & \text{иначе,} \end{cases} \tag{3}$$

и результат вычислений становится состоянием соответствующей клетки в $C(t + 1)$;

2) если $C(t + 1) = C(t)$, то $C(t) = \Phi(C(0))$ — результат вычислений, который соответствует стабильному состоянию КНАП.

Множество весовых векторов **W** формируется путем обучения КНАП. Понятие Розенблатта [7] о линейной различимости образов, которые должны храниться в КНАП (называемые в дальнейшем *прототипами*), и метод обучения персептрона используются как основа для разработки методов обучения КНАП. Поэтому дадим их краткое описание.

Прототипы $P^0, ..., P^{L-1}$ *линейно различимы*, если существует множество весовых векторов $\mathbf{W} = \{W_{ij}\}$ такое, что

$$\langle D_{ij}^K, W_{ij} \rangle > 0 \ \forall \ (i, j) \in M \tag{4}$$

для любого $K = 0, ..., L-1$, где D_{ij}^K — приведенное соседство нейрона (i, j) в прототипе P^K.

Метод обучения персептрона описывается следующим образом:

1) из множества прототипов $\{P^0, ..., P^{L-1}\}$ формируется обучающая последовательность в виде $\{P^{(i)}: P^{(i)} = P^K$ при $i = k(\text{mod} L)\}$;

2) весовые векторы, которые не включают в себя вес автосвязи, вычисляются по следующей итеративной процедуре:

$$\begin{cases} W_{ij}^{(0)} & \text{произвольно,} \\ t \geq 0: & \\ W_{ij}^{(t+1)} = W_{ij}^{(t)}, & \text{если } \langle D_{ij}^{(t)}, W_{ij}^{(t)} \rangle > 0, \\ W_{ij}^{(t+1)} = W_{ij}^{(t)} + D_{ij}^{(t)} & \text{иначе.} \end{cases}$$

Вычисления прекращаются, когда для всех (i, j) выполняется условие (4). Это значит, что все прототипы P^K *индивидуально устойчивы*, т. е. в такой сети $\Phi(P^K) = P^K$ для всех P^K.

Действуя в соответствии с вышеописанной процедурой (Ф), КНАП производит хранение и восстановление прототипов. Образ, который должен быть распознан, подается на вход путем установки КНАП в соответствующее начальное состояние $C(0)$. Затем применяется итеративная процедура, в результате которой сеть либо приходит в устойчивое состояние, являющееся ближайшим к $C(0)$, либо возникают колебания в форме $C(t) = C^1, C(t+1) = C^2, C(t+2) = C^1, \ldots$.

Основываясь на методе обучения персептрона, в [6] был предложен метод обучения КНАП, который, как и его предшественник, определял весовые векторы без компонент автосвязи. Моделирование обученной по предложенному методу КНАП показало, что введение веса автосвязи $w_0 > 0$ существенно улучшало способности КНАП. В частности, устойчивость к искажениям возрастала в 7—10 раз в нейросети, обученной на прототипах, нарисованных толщиной в одну клетку и представляющих собой символы алфавита, цифры и т. д.

Для того чтобы исследовать это явление и определить оптимальное значение веса автосвязи, необходимы следующие обозначения, относящиеся к КНАП. Пусть P^K — прототип, хранимый в КНАП, и C_{ij}^K — состояние соседства нейрона с именем (i, j) в этом прототипе. Тогда $C_{ij}^K(k)$ называется *k-искажением соседства* C_{ij}^K, если $C_{ij}^K(k)$ отличается от C_{ij}^K в состояниях k нейронов. Образ $P^K(k)$ называется *k-искажением* прототипа P^K, если для любого (i, j) $C_{ij}^K(s)$ является s-искажением соседства для $s \leq k$. Прототип P^K называется *k-аттрактором*, если для любого s-искажения $P^K(s)$, $s \leq k$, $\Phi(P^K(s)) = P^K$. Если результат получается за одну итерацию, то P^K — *сильный k-аттрактор*. КНАП, хранящая $P^0, ..., P^{L-1}$, называется *устойчивой* (*сильно устойчивой*) *к k-искажениям*, если любой P^K является k-аттрактором (сильным k-аттрактором). Аналогично, если нейрон для любого k-искажения своего соседства распознает истинное состояние за один шаг распознающей процедуры, то он называется *сильно устойчивым к k-искажениям.*

2. **Случай 1-искаженных образов.** Пусть КНАП обучена хранить множество прототипов $\mathbf{P} = \{P^0, ..., P^{L-1}\}$. Для дальнейших рассуждений предположим, что вес автосвязи равен 0 и любой $P^K \in \mathbf{P}$ индивидуально устойчив (иными словами, для любого (i, j) и для всех P^K величина $\langle D_{ij}^K, W_{ij}\rangle > 0$). Такие весовые векторы могут быть получены, например, с помощью метода обучения персептрона. Если же нейросеть обучена так, что веса автосвязей не равны нулю (например, по методу [2]), то они выбрасываются из дальнейшего рассмотрения, но при этом все равно остается требование индивидуальной устойчивости в такой сети. Заметим, что в дальнейшем C_{ij}^K (и D_{ij}^K) будет обозначать только состояние соседства нейрона (i, j) без его собственного состояния. Поэтому, если в нейросеть ввести положительный вес w_0, то для прототипа P^K функция (3) примет следующий вид:

$$f(C_{ij}^K, W_{ij}) = \begin{cases} 1, & \text{если } \langle C_{ij}^K, W_{ij}\rangle + c_{ij}^K w_0 \geq 0, \\ -1 & \text{иначе,} \end{cases} \tag{5}$$

где c_{ij}^K — состояние нейрона (i, j) в этом образе.

Сначала рассмотрим случай 1-искажений. Для определения влияния w_0 на нейрон (i, j) будут использованы следующие предположения.

1. Пусть P^K — хранимый образ, $P^K(1)$ — его 1-искажение, и состояние нейрона (i, j) искажено. Тогда этот нейрон сможет корректно распознать свое состояние, т. е. прийти к состоянию c_{ij} в P^K после одного шага итеративной процедуры Φ с функцией (5), если

$$\langle W_{ij}, D_{ij}^K\rangle - w_0 \geq 0,$$

иначе это искажение может зафиксироваться (напомним, что для приведенного соседства D_{ij}^K в соответствии с (2) нормализованное состояние нейрона $d_{ij} = c_{ij}^2$, т. е. $d_{ij} = 1$ для любого исходного состояния c_{ij}). По этой причине влияние w_0 не должно превосходить влияния всех остальных весов связей для любого прототипа P^K, т. е.

$$w_0 \leq \min_K \langle D_{ij}^K, W_{ij}\rangle. \tag{6}$$

2. Пусть в $P^K(1)$ состояние нейрона (i, j) не искажено, но искажен один из его соседей. Если

$$\langle D_{ij}^K(1), W_{ij}\rangle + w_0 \geq 0 \tag{7}$$

(где $D_{ij}^K(1)$ — это 1-искажение D_{ij}^K), тогда нейрон (i, j) не изменит своего состояния после применения одного шага итеративной процедуры Φ. В этом случае чем больше w_0, тем больше сильная устойчивость нейрона. Таким образом, из (6) и (7) получаем

$$w_0 = \min_K \langle D_{ij}^K, W_{ij}\rangle. \tag{8}$$

3. **Случай k-искаженных образов.** Пусть образы $P^0, ..., P^{L-1}$ хранятся в КНАП, C_{ij}^K, как и раньше, обозначает состояние соседства нейрона (i, j) без состояния самого нейрона. Введение автосвязи делает необходимым рассмотрение *состояния расширенного соседства* нейрона как $E_{ij}^K = c_{ij}^K \cup C_{ij}^K$ (c_{ij}^K — состояние нейрона (i, j)). Аналогично (2) определяется вектор $G_{ij}^K = c_{ij}^K E_{ij}^K$ сос-

тояния приведенного расширенного соседства (в дальнейшем термин «расширенного» будет опускаться из-за разных буквенных обозначений соседства). Вектор W_{ij}^K определяется как $w_0^K \cup W_{ij}$, где w_0^K — вес автосвязи нейрона (i, j) для прототипа P^K (объяснение выбора различных весов автосвязей для разных прототипов будет дано позже). Отметим, что $\langle G_{ij}^K, W_{ij}^K \rangle = \langle D_{ij}^K, W_{ij} \rangle + w_0$, так как нулевая компонента G_{ij}^K равна $g_0^K = (c_{ij}^K)^2 = +1$ (это следует из вышеприведенных определений).

Пусть $D_{ij}^K(k)$ будет k-искажением приведенного соседства, тогда

$$\langle D_{ij}^K(k), W_{ij} \rangle = \langle D_{ij}^K, W_{ij} \rangle - 2(d_{s_1}^K w_{s_1} + \ldots + d_{s_k}^K w_{s_k}),$$

где $\{s_j : j = 1, \ldots, k\}$ — множество имен искаженных нейронов в $D_{ij}^K(k)$. Для k-искажения нормализованного соседства $G_{ij}^K(k)$ подобное скалярное произведение есть

$$\langle G_{ij}^K(k), W_{ij}^K \rangle = \langle D_{ij}^K, W_{ij} \rangle + w_0 - 2(d_{r_1}^K w_{r_1} + \ldots + d_{r_k}^K w_{r_k}), \tag{9}$$

где

$$\delta(k) = 2(d_{r_1}^K w_{r_1} + \ldots + d_{r_k}^K w_{r_k}) \tag{10}$$

представляет собой *влияние искажений*. Необходимо отметить, что в множестве $\{r_j : j = 1, \ldots, k\}$ имен искаженных нейронов (в $G_{ij}^K(k)$) может быть и имя самого нейрона (i, j). Если это так, то влияние w_0 отрицательно ($w_0 - 2w_0 < < 0$).

Из (9) и (5) следует, что сильная устойчивость нейрона (i, j) к любым k-искажениям $D_{ij}^K(k)$ ($G_{ij}^K(k)$) выражается неравенством

$$\langle D_{ij}^K(k), W_{ij} \rangle \geq 0 \quad (\langle G_{ij}^K(k), W_{ij}^K \rangle \geq 0). \tag{11}$$

Итак, теперь проблема ставится следующим образом: определить значение w_0 так, чтобы (11) (неравенство для $G_{ij}^K(k)$) выполнялось для наибольшего k. Для этого рассмотрим влияние искажений (10). Величина $\delta(k)$ ограничивается сверху величиной $k \max |w_i|$, и, используя эту оценку, получаем [6]

$$w_0 = \min\left(\min_K \langle D_{ij}^K, W_{ij} \rangle, \max_i |w_i|\right). \tag{12}$$

В [6] показано, что $w_0 = \max |w_i|$ максимально увеличивает величину

$$k_{ij} = \min_K \frac{\langle D_{ij}^K, W_{ij} \rangle + w_0}{2 \max |w_i|},$$

которая оценивает максимально допустимые искажения любого из хранимых прототипов. Эту грубую оценку можно уточнить, принимая во внимание несколько наибольших компонент весового вектора (а не одну, как это сделано выше), и (12) использовать для приближенных вычислений w_0. Для того чтобы более точно вычислить w_0, используем следующие предпосылки.

1. Влияние искажений (10) оценивается более точной величиной

$$\delta(k) \leq 2\max(|w_{i_1}| + \ldots + |w_{i_k}|). \tag{13}$$

Заметим, что для некоторого P^K в (13) может быть достигнуто равенство, т. е. оценка в точности совпадет с (10).

2. Веса автосвязи w_0^K, $K = 0, \ldots, L - 1$, вычисляются отдельно для каждого прототипа P^K, и окончательная величина w_0 определяется из этих значений в соответствии с заданными условиями.

Предположим теперь, что w_0^K уже определено и все веса связей нейрона (i, j) переименованы так, что

$$|w_1| \geq |w_2| \geq \ldots \geq |w_{q+1}|,$$

и пусть вес автосвязи получил номер t_0, т. е. $w_0^K = w_{t_0}$. Тогда, используя (9), можно записать, что нейрон (i, j) сильно устойчив к k-искажениям в своем расширенном соседстве, если

$$f_{ij}(w_{t_0}) = \langle D_{ij}^K, W_{ij} \rangle + w_{t_0} - 2(|w_1| + |w_2| + \ldots + |w_k|) \geq 0. \tag{14}$$

Максимальное k (оценка числа допустимых искажений) может быть найдено как

$$k_{\max} = \max\{k: f_{ij}(w_{t_0}) \geq 0\}. \tag{15}$$

Для определения w_{t_0}, которое наибольшим образом увеличивает (15), рассмотрим два случая.

1. Если $t_0 \geq k$ (т. е. $w_{t_0} \leq |w_k|$) в (14), то с увеличением w_{t_0} возрастает $f_{ij}(w_{t_0})$.

2. Если $t_0 \leq k$ в (14), то с увеличением w_{t_0} убывает $f_{ij}(w_{t_0})$ (соответственно $k_{\max}$ может уменьшиться), так как

$$f_{ij}(w_{t_0}) = \langle D_{ij}^K, W_{ij} \rangle + w_{t_0} - 2(|w_1| + \ldots + |w_{t_0}| + \ldots + |w_k|) =$$

$$= \langle D_{ij}^K, W_{ij} \rangle - w_{t_0} - 2(|w_1| + \ldots + |w_{t_0 - 1}| + |w_{t_0 + 1}| + \ldots + |w_k|).$$

Отсюда видно, что наибольшее увеличение $k_{\max}$ получается для $w_{t_0} = |w_{k_{\max}}|$. Как вычислять такие веса автосвязи $w_0^K = w_{t_0}$ на практике? Алгоритм может быть следующим (для нейрона (i, j)).

1. Берем прототип P^K и находим $k_1 = \max\{k: f_{ij}(0) \geq 0\}$ — максимальное число такое, что нейрон (i, j) становится сильно устойчивым к k_1-искажениям на прототипе P^K без автосвязи.

2. Если $k_1 \geq 1$, то вычисляем значение $y_1 = 2(|w_1| + \ldots + |w_{k_1+1}|) - \langle D_{ij}^K, W_{ij} \rangle$ и рассматриваем следующие случаи:

а) если $y_1 \geq |w_{k_1}|$, то $w_0^K = |w_{k_1}|$, так как $k_{\max} = k_1$;

б) если $y_1 \leq |w_{k_1+1}|$, то $w_0^K = |w_{k_1+1}|$, так как $k_{\max} = k_1 + 1$;

в) иначе $w_0^K = y_1$.

Таким образом, получаем точное значение w_0^K, которое максимально увеличивает сильную устойчивость нейрона (i, j) на выбранном прототипе P^K (в случае, когда $k_1 \geq 1$). Возникает следующий вопрос: как определить окончательную величину для w_0, одинаковую для всех прототипов? Допустим, что для каждого прототипа было вычислено значение w_0^K, т. е. данный нейрон без автосвязи сильно устойчив, по крайней мере, к 1-искажениям. Тогда можно использовать следующие подходы:

1) найти прототип P^K, для которого на первом шаге вышеописанного алгоритма получено минимальное допустимое число k-искажений, и выбрать $w_0 = w_0^K$ (в этом случае, если минимальное k равно 0, то w_0 вычисляется по формуле (8));

2) взять в качестве w_0 среднее значение от w_0^K для всех прототипов, т. е. $w_0 = (w_0^0 + \ldots + w_0^{L-1})/L$.

Первый подход гарантирует то, что максимальное число допустимых искажений (для всех прототипов) после расчета веса автосвязи станет наибольшим из возможных.

Отметим, что если для нейрона (i, j) на некотором прототипе P^K величина $k_1 = 0$, т. е. нет сильной устойчивости даже к 1-искажениям, то вес автосвязи вычисляется по формуле (8). Таким образом, объединены результаты двух последних разделов и четко определены области применения предложенных формул.

Заключение. Показано, как влияет величина веса автосвязи на основную характеристику КНАП, называемую сильной устойчивостью к k-искажениям хранимых прототипов. Получено выражение для определения величин (w_0^K), которые обеспечивают максимальную сильную устойчивость для каждого конкретного прототипа. Также предложены две стратегии выбора окончательной величины автосвязи, которая должна быть одной и той же для распознавания всех хранимых образов. Полученные результаты справедливы не только для КНАП, но и для полносвязной ассоциативной памяти Хопфилда, обученной с помощью любого метода (необходимо только требование индивидуальной устойчивости прототипов без автосвязи).

СПИСОК ЛИТЕРАТУРЫ

1. **Hopfield J. J., Tank D. W.** Computing with neural circuits: a model // Science. 1986. **233**. P. 625.
2. **Personnaz L., Guyon I., Dreyfus G.** Collective computational properties of neural networks: New learning mechanisms // Phys. Rev. A. 1986. **34**. P. 4217.
3. **Cottrell M.** Stability and attractivity in associative memory networks // Biological Cybernetics. 1988. **58**. P. 129.
4. **Li J., Michel A. N., Porod W.** Analysis and synthesis of a class of neural networks: linear systems operating on a closed hypercube // IEEE Trans. Circuits and Systems. 1989. **36**, N 11. P. 1405.
5. **Liu D., Michel A. N.** Sparsely interconnected neural networks for associative memories with applications to cellular neural networks // IEEE Trans. Circuits and Systems — II: Analog and Digital Signal Process. 1994. **41**, N 4. P. 295.
6. **Пудов С. Г.** Обучение клеточно-нейронной ассоциативной памяти // Автометрия. 1997. № 2. С. 107.
7. **Rosenblatt F.** Principles of Neurodynamics. Washington: Spartan, 1959.
8. **Gorodnichy D. O.** Desaturating coefficient for projection learning rule // Lecture Notes in Computer Science. 1996. **1112**. P. 469.
9. **Bandman O. L.** Cellular-neural computations, formal model and possible applications // Lecture Notes in Computer Science. 1995. **964**. P. 21.

Поступила в редакцию 12 мая 1997 г.