УДК 681.513.7

Д.В. НОВИЦКИЙ

ВОССТАНОВЛЕНИЕ АССОЦИАТИВНОЙ ПАМЯТИ В СЛУЧАЕ УНИЧТОЖЕНИЯ ЧАСТИ НЕЙРОНОВ

Анотація. *Розглядається задача відновлення нейронної асоціативної пам'яті у випадку, коли частину її нейронів було цілком знищено. На моделі мережі типу Хопфілда показано, що для цілковитого відновлення функціонування пам'яті достатньо інформації про таку кількість початково завантажених образів, що дорівнює числу вилучених нейронів. У чисельному експерименті таке відновлення здійснюється шляхом донавчання, подібно до звичайного навчання такого типу мереж.*
Ключові слова: *асоціативна пам'ять, мережа Хопфілда, знищення нейронів, відновлення пам'яті.*

Аннотация. *Рассматривается задача восстановления ассоциативной памяти в случае, если часть ее нейронов была полностью уничтожена. На модели сети типа Хопфилда показано, что для полного восстановления функционирования памяти достаточно информации о ранее запомненных образах в количестве, равном числу удаленных нейронов. В численном эксперименте восстановление проводится путем дообучения, аналогично обычному обучению ассоциативных нейросетей.*
Ключевые слова: *асоциативная память, сеть Хопфилда, уничтожение нейронов, восстановление памяти.*

Abstract. *We consider re-learning ability of a Hopfield-type network after killing some neurons. Neurons were "killed" by means of nullification of corresponding rows and columns of the synaptic matrix. We show that one can restore recognition ability of this network using re-training with the vectors, which was memorized before. The number of needed vectors is equal to the number of deleted neurons. It does not depend on network's size and on volume of stored data.*
Keywords: *associative memory, Hopfield network, neurons killing, memory recovery.*

1. Введение

Известно, что сети нейронной ассоциативной памяти типа Хопфилда отличаются большой информационной избыточностью. Так, при применении классического правила обучения Хопфилда (Хэбба) сеть с n нейронами может корректно запомнить не более $0{,}14n$ образов [1]. При применении псевдоинверсного правила обучения это число возрастает до $0{,}25n$; количество весов такой сети равно n^2. Хотя такая избыточность и требует значительных вычислительных ресурсов при больших n, она обеспечивает устойчивость при различного рода искажениях. Так, например, можно удалить (приравняв соответствующие веса нулю) значительную часть связей такой сети без заметного ухудшения характеристик ассоциативной памяти [2].

Интересен вопрос, можно ли восстановить информацию, содержащуюся в сети, если часть нейронов была полностью «убита» и заменена «пустыми» – со всеми весами связей, равными нулю? Мы покажем теоретически, что для этого достаточно знать столько обучающих векторов (из числа содержавшихся в памяти до повреждения), сколько нейронов было удалено из сети. При этом восстанавливаются все ранее запомненные образы, а не только те, которые повторно вводились в сеть.

Численные эксперименты показывают, что для восстановления памяти может быть использовано дообучение системы по тем же правилам, что и при обычном псевдоинверсном алгоритме. Хотя полученная таким образом матрица весов не тождественна исходной, система показала способность к воспоминанию, близкую к первоначальной.

В данной работе мы используем модель ассоциативной памяти на основе сетей Хопфилда с псевдоинверсным правилом обучения [3]. В сетях этого типа запоминаются биполярные векторы: $v_k \in \{-1,\ 1\}^n$, $k = 1...m$. Пусть эти векторы образуют столбцы матрицы V размером $m \times n$. Синаптическая матрица C дается соотношением

$$C = VV^+, \tag{1}$$

где V^+ – матрица, псевдообратная к V по Муру-Пенроузу. Ее можно вычислить напрямую по формуле $V^+ = \left(V^T V\right)^{-1} V^T$ или по формулам Гревиля [4, 5].

Ассоциативный поиск осуществляется с помощью процедуры экзамена: входной вектор x_0 служит начальной точкой итераций вида

$$x_{t+1} = f\left(Cx_t\right), \tag{2}$$

где f – монотонная нечетная функция, такая что $\lim_{s \to \pm\infty} f\left(s\right) = \pm 1$. К векторному аргументу она применяется покомпонентно. Устойчивую неподвижную точку этого отображения будем называть аттрактором. Максимальное расстояние по Хэммингу между входным вектором x_0 и запомненным образом v_k такое, что процедура экзамена все еще сходится к v_k и называется аттракторным радиусом.

2. Теория восстановления памяти

Рассмотрим поведение сети Хопфилда, обученной с помощью псевдоинверсного алгоритма, в которой часть нейронов удалена путем обнуления значений веса связей на их входах и выходах. При таком удалении сеть теряет способность к конвергенции, вследствие чего происходит разрушение ассоциативной памяти (АП), ее содержимое становится недоступным. Покажем, что такую АП можно полностью восстановить путем повторного запоминания лишь части векторов из числа запомненных ранее. Пусть из сети удалено p нейронов. Тогда исходную (проекционную) матрицу сети можно представить в виде

$$C = \begin{pmatrix} X & Y \\ Y^T & Z \end{pmatrix}, \tag{3}$$

здесь X, Y, Z – матрицы размером $(n-p)\times(n-p)$, $(n-p)\times p$, $p \times p$ соответственно. Они связаны соотношениями

$$\begin{cases} X = X^2 + YY^T \\ Z = Z^2 + Y^T Y \\ Y = XY + Y^T Z \end{cases}. \tag{4}$$

Если известна только матрица X усеченной сети, соотношения (4) можно разрешить относительно Y, что позволит восстановить исходную проекционную матрицу C. Но это решение неоднозначно: Y можно домножить справа на произвольную ортогональную матрицу U размером $p \times p$. Однако неоднозначность может быть снята путем дообучения на некотором наборе векторов из исходного множества.

Количество вновь запоминаемых векторов, необходимое для восстановления памяти, равно числу удаленных нейронов p и не зависит от размеров сети и объема запомненных ею данных.

Теорема 1. Если известны матрица X, а также произведение исходной матрицы C на p векторов из исходного запоминаемого множества, матрица C восстанавливается однозначно.

Доказательство. Матрица Y находится из системы (2) с точностью до умножения на ортогональною матрицу U размером $m \times m$. При этом Z определяется однозначно. Пусть $V_{n,p}$ – матрица, образованная из p линейно независимых векторов, содержавшихся в исходной памяти. Разделим ее на блоки $V_{n-p,p}$ $V_{p,p}$, относящиеся к сохранившимся и удаленным нейронам соответственно. Затем найдем U, решая систему

$$UY_0^T V_{n-p,p} + ZV_{p,p} = V_{p,p}, \tag{5}$$

где Y_0 – какое-нибудь решение системы (4). U находится в виде

$$U = (I - Z)V_{pp}\left(Y_0^T V_{n-p,p}\right)^+ .\blacksquare$$

3. Экспериментальные результаты

Возможность восстановления связей утраченных нейронов подтверждают эксперименты по дообучению [6]. Использованный там алгоритм базируется на проекционном правиле обучения АП. Для дообучения предъявлялись p образов из числа запомненных ранее. Такой метод не дает точного решения систем (2)–(3), однако обеспечивает его приближение. О качестве приближения можно судить по спектрам получающихся матриц.

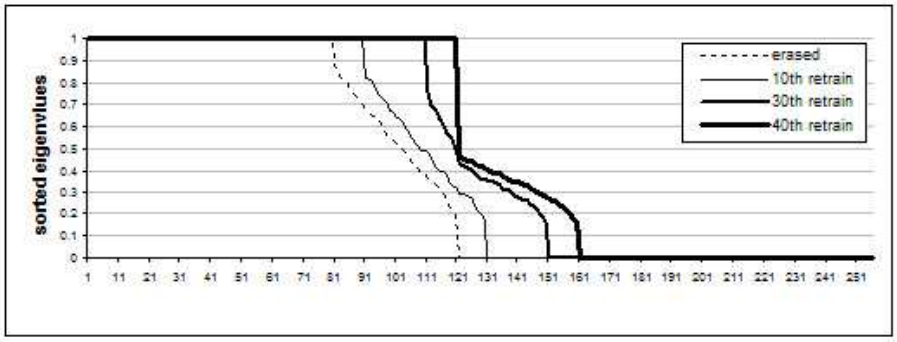

Рис. 1. Эволюция спектра синаптической матрицы при обучении. M=120

На рис. 1 приведены данные для матрицы связей 256×256, в которой после запоминания 120 векторов было обнулено по 40 строк и столбцов. Графики отражают спектры до обучения, после обучения на 10, 30 и 40 векторах, запомненных сетью ранее. Как видно, при удалении нейронов ранг матрицы сохраняется, но величины 40 из 120 собственных значений заметно уменьшились. При запоминании каждого следующего вектора ранг матрицы увеличивается на единицу, причем добавляется одна компонента с единичным собственным значением и происходит перераспределение и сокращение значений 40 ранее ослабленных компонент спектра. После запоминания 40 векторов они образуют “хвост” спектра матрицы и по значению не превосходят 0,47. Образование “хвоста” вызвано тем, что собственные векторы сети, образовавшейся после обнуления части связей, имеют составляющие, ортогональные ранее запомненным векторам. При обучении восстановление ранее запомненных собственных векторов сопровождается запоминанием ортогональных составляющих, собственные значения которых менее 0,5.

Приведенные на рис. 1 данные относятся к случаю плотного заполнения памяти $(m/n = 0,47)$ при удалении более 15% нейронов. На рис. 2 приведены аналогичные данные для такой же сети при малом заполнении памяти $(m = 40,\ m/n = 0,15)$ и удалении всего 10 нейронов (менее 4%). Приведены спектры в начале обучения, после обучения на 5, 9 и 10 векторах, запомненных сетью до разрушения связей, а также после дополнительного обучения на новых 20 векторах. Отметим, что появление 20 новых единичных собственных значений при запоминании новых векторов практически не повлияло на 10 дополнительных малых составляющих спектра, образовавшихся в результате удаления нейронов.

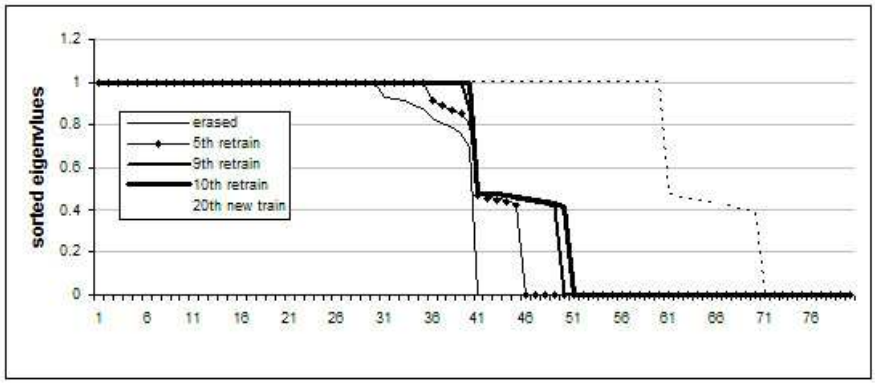

Рис. 2. Эволюция спектра при обучении на запомненных данных и последующем обучении на 20 новых векторах. M=40+20

При малом заполнении памяти $(m/n < 0,2)$ появление новых составляющих спектра практически не влияет на поведение сети. Однако при высоком заполнении памяти появление дополнительных компонент увеличивает число ложных аттракторов сети и сокращает объем свободной ассоциативной памяти. Применяя метод разнасыщения синаптической матрицы, удается значительно уменьшить их влияние. Более радикальным решением может быть возведение синаптической матрицы в высокую степень, после чего в ней сохранятся лишь собственные значения, близкие к единице.

На рис. 3 показана динамика аттракторных свойств АП по мере ее обучения. Сеть с «убитыми» нейронами полностью теряет способность к конвергенции (лишь для аттрактора №100 был случайно зафиксирован ненулевой радиус). По мере дообучения аттракторный радиус тех образов, которые были повторно предъявлены сети, сразу восстанавливается до первоначального значения, всех остальных – постепенно увеличивается. В момент, когда число векторов для дообучения сравнивается с числом уничтоженных нейронов, аттракторные свойства системы полностью восстанавливаются.

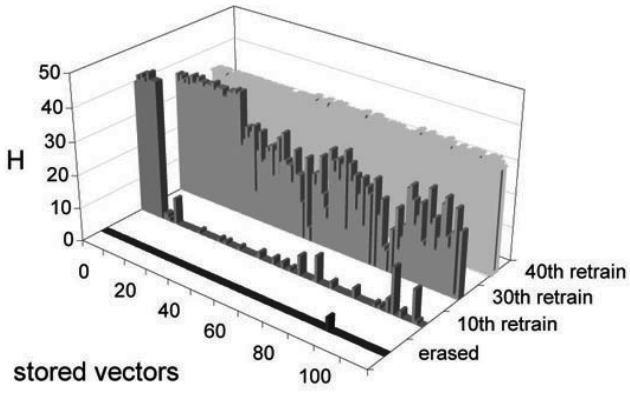

Рис. 3. Изменение радиуса аттракторов сети в процессе дообучения

4. Выводы

Явление восстановления АП с уничтоженными нейронами с помощью подмножества ранее содержавшихся в ней образов напоминает широко известные примеры излечения больных амнезией путем напоминания пациентам ярких событий из их прошлого.

СПИСОК ЛІТЕРАТУРЫ

1. Amit D. Modeling Brain function / D. Amit // The world of attractor networks. – Cambridge Univ. Press, 1989. – 528 p.
2. Сычев А.С. Селекция связей в нейронных сетях с псевдоинверсным алгоритмом обучения / А.С. Сычев // Математические машины и системы. – 1998. – № 2. – С. 25 – 30.
3. Personnaz L. Collective computational properties of neural networks: New learning mechanisms / L. Personnaz, I. Guyon, G. Dreyfus // Phys. Rev. A. – 1986. – Vol. 34, N 5. – P. 4217 – 4228.
4. Albert A. Regression and the Moore-Penrose pseudoinverse / A. Albert. – New-York-London: Academic Press, 1972. – 180 p.
5. Кириченко Н.Ф. Псевдообращение матриц в проблеме проектирования ассоциативной памяти / Н.Ф. Кириченко, А.М. Резник, С.П. Щетенюк // Кибернетика и системный анализ. – 2000. – № 3. – С. 18 – 27.
6. Associative Memories with "Killed" Neurons: the Methods of Recovery / A.M. Reznik, A.S. Sitchov, O.K. Dekhtyarenko [et al.] // Proc. of the International Joint Conference on Neural Networks. – Portland: Neural Networks, 2003. – Vol. 4. – P. 2579 – 2582.

Стаття надійшла в редакцію 19.11.2012