ПРИКЛАДНАЯ ДИСКРЕТНАЯ МАТЕМАТИКА

2019 Математические основы информатики и программирования № 45

УДК 621.396:621.372

ИНФОРМАЦИОННАЯ ЁМКОСТЬ СЕТИ ХОПФИЛДА С КВАНТОВАННЫМИ ВЕСАМИ

М. С. Тарков

Институт физики полупроводников им. А. В. Ржанова СО РАН, г. Новосибирск, Россия

Использование бинарных и многоуровневых мемристоров при аппаратной реализации нейронных сетей вызывает необходимость квантования их весовых коэффициентов. Исследуется влияние числа уровней квантования весов сети Хопфилда на её информационную ёмкость и устойчивость к искажениям входных данных. Показано, что при числе градаций весов порядка 20 ёмкость сети Хопфилда — Хебба с дискретными весами приближается к ёмкости её варианта с непрерывными весами. Для проекционной сети Хопфилда подобного результата удаётся достичь лишь при числе градаций порядка 100. Эксперименты показали, что: 1) бинарные мемристоры следует использовать в сетях Хопфилда — Хебба, редуцированных путём обнуления всех весов, модули которых строго меньше максимального для данной строки матрицы весов; 2) в проекционных сетях Хопфилда с дискретными весами следует использовать многоуровневые мемристоры с числом градаций (уровней) значительно больше двух, причём конкретное число уровней зависит от размерности хранимых эталонных векторов, их конкретного набора и допустимого уровня шума во входных данных.

Ключевые слова: *сети Хопфилда — Хебба, проекционные сети Хопфилда, информационная ёмкость, квантование весов, бинарные и многоуровневые мемристоры.*

DOI 10.17223/20710410/45/11

INFORMATIONAL CAPACITY OF THE HOPFIELD NETWORK WITH QUANTIZED WEIGHTS

M. S. Tarkov

Rzhanov Institute of Semiconductor Physics SB RAS, Novosibirsk, Russia

E-mail: tarkov@isp.nsc.ru

The use of binary and multilevel memristors in the hardware neural networks implementation necessitates their weight coefficients quantization. In this paper, we investigate the Hopfield network weights quantization influence on its information capacity and resistance to input data distortions. It is shown that, for a weight level number of the order of tens, the capacitance of Hopfield — Hebb network with the quantized weights approximates the capacitance of its version with continuous weights. For a Hopfield projection network, similar result can be achieved only for a weight levels number of the order of hundreds. Experiments show that: 1) binary memristors should be used in Hopfield — Hebb networks, reduced by zeroing all weights in a given row, in which absolute values are strictly less than the maximum weight in the row; 2) in the Hopfield projection networks with quantized weights, multilevel memristors with a weight levels number significantly more than two should be used, with a specific levels number depending on the stored reference vectors dimension, their particular set and the permissible input data noise level.

Keywords: *Hopfield — Hebb networks, projection Hopfield networks, information capacity, weight quantization, binary and multilevel memristors.*

Введение

В процессе функционирования сети Хопфилда [1, 2] можно выделить два режима — обучение и классификацию. Веса сети вычисляются в режиме её обучения на основе известных векторов. В режиме классификации при фиксированных значениях весов и заданных значениях входов нейрона инициируется переходный процесс, завершающийся в одном из локальных энергетических минимумов. При вводе обучающих векторов x^k, $k = 1, \ldots, p$, вычисляются веса w_{ij} в соответствии с обобщённым правилом Хебба

$$w_{ij} = \frac{1}{N} \sum_{k=1}^{p} x_i^k x_j^k. \tag{1}$$

Важным параметром ассоциативной памяти является её информационная ёмкость, которая определяется как максимальное число хранимых в памяти образцов. Показано [1], что при использовании правила Хебба ёмкость памяти оценивается как $p_{\max} = N/(2\ln N)$. Такую сеть будем называть далее сетью Хопфилда — Хебба.

Проекционный метод обучения сети Хопфилда [3, 4] порождает матрицу весов, итерационно зависящую от последовательности обучающих векторов x^k, $k = 1, \ldots, p$:

$$\begin{aligned} y^k &= (W^{k-1} - E)x^k, \\ W^k &= W^{k-1} + \frac{y^k \cdot y^{kT}}{y^{kT} \cdot y^k} \end{aligned} \tag{2}$$

при начальном условии $W^0 = 0$ (E — единичная матрица). В результате матрица весов сети получает значение $W = W^p$. Применение метода проекций увеличивает ёмкость сети Хопфилда до $N - 1$. Далее такую сеть будем называть проекционной сетью Хопфилда.

1. Бинарные и многоуровневые мемристоры

Аппаратная реализация нейронной сети требует много памяти для хранения матрицы весов слоя нейронов и является дорогостоящей. Решение этой проблемы упрощается при использовании в качестве ячейки памяти устройства, называемого мемристором (резистором с памятью). Мемристор был предсказан теоретически в 1971 г. Леоном Чуа [5]. Первую физическую реализацию мемристора продемонстрировала в 2008 г. лаборатория фирмы «Hewlett Packard» в виде тонкоплёночной структуры $\mathrm{TiO_2}$ [6]. Мемристор ведёт себя подобно синапсу: он «запоминает» полный электрический заряд, прошедший через него. Память, основанная на мемристорах, может достигать степени интеграции 100 Гбит/см^2, в несколько раз более высокой, чем на основе технологии флэш-памяти. Эти уникальные свойства делают мемристор многообещающим устройством для создания массово-параллельных нейроморфных систем.

Бинарными называют мемристоры, реализующие два значения проводимости, многоуровневыми — мемристоры, реализующие множество дискретных уровней проводимости (количество уровней может достигать десятков и сотен). Бинарные и многоуровневые мемристоры [7 – 9] основаны на механизме переключения филамента и распространены более широко, чем аналоговые мемристоры, в которых проводимость можно изменять непрерывно, но материалы для них встречаются значительно реже, к тому же они требуют более сложного процесса обработки. Многоуровневые мемристоры устойчивы к статистическим флуктуациям по сравнению с аналоговыми мемристорами. Использование бинарных и многоуровневых мемристоров для задания весовых коэффициентов сетей Хопфилда делает актуальной задачу исследования влияния числа уровней квантования весов на информационную ёмкость сети и её устойчивость к уровню шумов входных данных. При этом задачу квантования весов можно рассматривать как обобщение задачи их редукции [10].

2. Оценка ёмкости сети Хопфилда

Для оценивания ёмкости сети Хопфилда [11] порождается множество M случайных векторов с компонентами из множества $\{-1,+1\}$. На множестве M строится сеть Хопфилда согласно правилам (1) или (2). Каждый из векторов множества M подаётся на вход построенной сети Хопфилда. Если после одного срабатывания сети её выход совпадает с её входом для всех векторов множества M, то в M добавляется новый случайный вектор и процесс повторяется (изначально $|M| = 1$). В противном случае ёмкость сети полагаем равной $|M| - 1$.

Результаты апробации данного алгоритма приведены в табл. 1 и 2. Здесь C — оценка ёмкости, усреднённая по 100 испытаниям; $N/(2\ln N)$ — асимптотическая оценка ёмкости сети Хопфилда — Хебба; $N-1$ — оценка максимальной ёмкости проекционной сети Хопфилда.

Таблица 1

Метод Хебба

N	32	64	128	256	512
C	5,14	7,51	11,22	18,07	29,88
$N/(2\ln N)$	4,62	7,69	13,19	23,08	41,04

Таблица 2

Проекционный метод

N	32	64	128	256	512
C	30,83	62,39	125,97	253,52	508,52
$N-1$	31	63	127	255	511

Из табл. 2 следует, что результаты алгоритма [11] очень близки к оценкам [4] для проекционной сети Хопфилда (разница составляет менее 1 %).

3. Ёмкость редуцированной сети Хопфилда

Редукция весов сети Хопфилда осуществляется следующим образом [10]:

1. Веса сети нормируются по формуле

$$w_{ij} \leftarrow \frac{w_{ij}}{\max\limits_{j=1,\ldots,N} w_{ij}}.$$

2. Нормированные веса редуцируются по правилу

$$w_{ij}^{'} = \begin{cases} +1, & w_{ij} = 1, \\ -1, & w_{ij} = -1, \\ 0, & -1 < w_{ij} < 1. \end{cases} \tag{3}$$

С целью сохранения симметрии матрицы весов правило редукции (3) для сетей Хопфилда приобретает следующий вид: вес w_{ij} обнуляется одновременно с весом w_{ji} при одновременном выполнении условий $w_{ij} < w_{i,\max}$ и $w_{ji} < w_{j,\max}$, где $w_{i,\max}$ и $w_{j,\max}$ — максимальные веса в строках i и j соответственно.

В результате редукции получаем сеть с весами из множества $\{-1, 0, +1\}$, которые могут быть реализованы на основе использования бинарных мемристоров. Значения ± 1 задаются максимальной проводимостью, а значение 0 — минимальной проводимостью мемристора. В табл. 3 приведены значения ёмкости редуцированной сети Хопфилда — Хебба, усреднённые по 100 экспериментам. Их значения близки к $\log_2 N$.

Таблица 3

Ёмкость редуцированной сети Хопфилда — Хебба

N	32	64	128	256	512
C	4,62	5,43	6,52	7,85	8,9
$\log_2 N$	5	6	7	8	9

4. Ёмкость сети Хопфилда с квантованными весами

Алгоритм квантования весов сети Хопфилда состоит в следующем:

1. Найти максимум $W_{\max}$ модуля весов сети Хопфилда.
2. Найти величину кванта (скачка) веса $\Delta = W_{\max}/(L-1)$, где $L \geqslant 2$ — число уровней квантования.
3. Преобразовать все веса w_{ij}, $i, j = 1, \ldots, N$, по правилу: если $(k-1)\Delta < w_{ij} < k\Delta$, то w_{ij} получает значение $(k-1)\Delta \cdot \mathrm{sign}(w_{ij})$, $k = 1, \ldots, L$, где

$$\mathrm{sign}(a) = \begin{cases} 1, & a > 0, \\ -1, & a < 0. \end{cases}$$

На рис. 1 представлены графики зависимости ёмкости сети Хопфилда от числа уровней (градаций) $L \in \{2, 4, 8, 16, 32, 64, 128, 256\}$ квантования весовых коэффициентов, $N \in \{32, 64, 128\}$. Эти графики показывают, что при малых значениях L ёмкость сети с квантованными весами мала, но для сети Хопфилда — Хебба, начиная с $L = 16$, становится близкой к ёмкости сети с непрерывными весами (табл. 1). Для проекционной сети Хопфилда зависимость ёмкости от L выражена гораздо сильнее. Проекционная сеть с квантованными весами по ёмкости приближается к проекционной сети с непрерывными весами только при приближении L к $128 \div 256$.

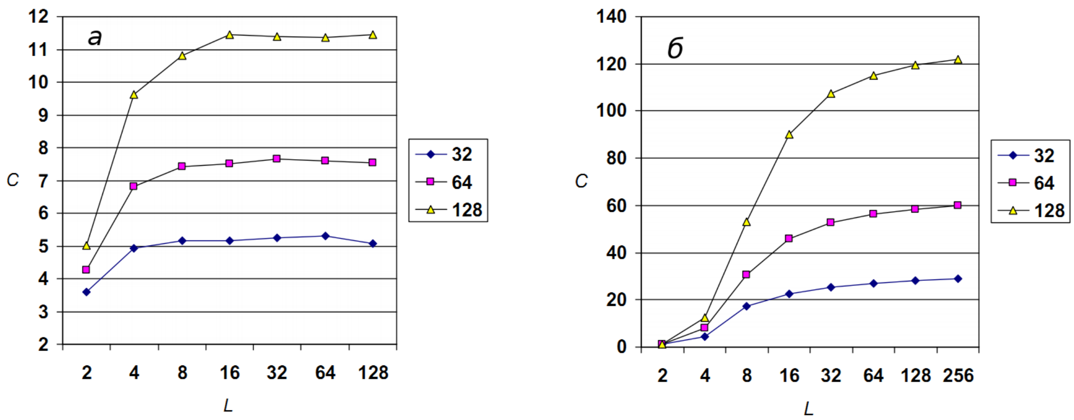

Рис. 1. Зависимость ёмкости сети Хопфилда — Хебба (*а*) и проекционной сети Хопфилда (*б*) от числа градаций весов

5. Выбор количества уровней квантования весов сети Хопфилда для фильтрации шумов

Выбор минимального количества уровней квантования весов для фильтрации шумов зависит от варианта сети Хопфилда (сеть Хопфилда — Хебба или проекционная сеть Хопфилда) и конкретного множества эталонных векторов. Для редуцированной сети Хопфилда — Хебба [10] количество уровней квантования равно двум независимо от размерности эталонных векторов. Веса таких сетей могут быть реализованы на бинарных мемристорах. При достаточно большой размерности векторов эти сети устойчивы к шумам при существенном сокращении числа связей (более 90 %) [10].

На рис. 2 и в табл. 4 приведены полученные экспериментально минимальные количества L уровней квантования весов проекционной сети Хопфилда в зависимости от доли искажённых пикселей ξ для набора из десяти эталонных векторов размерности $n \times n$, $n \in \{8, 16, 32, 64\}$, соответствующих изображениям цифр на рис. 3.

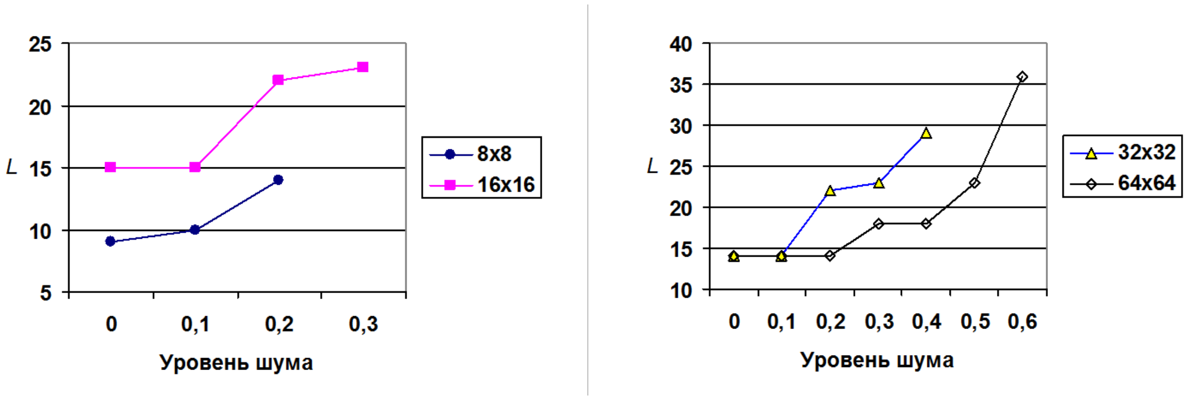

Рис. 2. Зависимость числа L уровней квантования от уровня шумов ξ

Таблица 4

Количество L уровней квантования весов проекционной сети Хопфилда для предельно допустимых долей ξ искажённых пикселей

$n \times n$	8×8	16×16	32×32	64×64
ξ	0,14	0,33	0,48	0,64
L	14	26	29	36

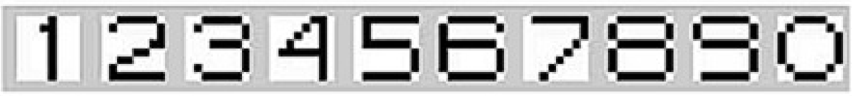

Рис. 3. Эталонные изображения

Результаты экспериментов показывают, что для обеспечения допустимых уровней шумов, соответствующих проекционной сети Хопфилда [10] с непрерывными весами, требуются десятки уровней весов.

Заключение

В настоящее время актуальной является задача построения нейронных сетей на основе использования многоуровневых (в том числе двухуровневых) мемристоров для аппаратной реализации синапсов. Поскольку число различимых уровней (градаций) проводимости мемристора ограничено, в работе исследуется влияние числа уровней квантования весов сети Хопфилда на её информационную ёмкость. Показано, что когда число градаций весов достигает 16, ёмкость квантованной сети Хопфилда — Хебба приближается к ёмкости её варианта с непрерывными весами. Для проекционной сети Хопфилда подобного результата удаётся достичь при числе градаций порядка сотен.

Эксперименты показали, что:

1. Бинарные мемристоры следует использовать в сетях Хопфилда — Хебба, редуцированных путём обнуления в строке матрицы всех весов, модули которых строго меньше максимального для данной строки.

2. В проекционных сетях Хопфилда с дискретными весами следует использовать многоуровневые мемристоры с числом уровней значительно больше двух, причём конкретное число уровней зависит от размерности хранимых эталонных векторов, их конкретного набора и допустимого уровня шума.

ЛИТЕРАТУРА

1. *Haykin S.* Neural Networks: A Comprehensive Foundation. Prentice Hall, Inc., 1999.
2. *Hopfield J.* Neural networks and physical systems with emergent collective computational abilities // Proc. NAS USA. 1982. V. 79. P. 2554–2558.
3. *Personnaz L., Guyon I., and Dreyfus G.* Collective computational properties of neural networks: New learning mechanisms // Phys. Rev. Ser. A. 1986. V. 34. No. 5. P. 4217–4228.
4. *Michel A. N. and Liu D.* Qualitative Analysis and Synthesis of Recurrent Neural Networks. N.Y.: Marcel Dekker Inc., 2002.
5. *Chua L.* Memristor — the missing circuit element // IEEE Trans. Circuit Theory. 1971. V. 18. P. 507–519.
6. *Strukov D. B., Snider G. S., Stewart D. R., and Williams R. S.* The missing memristor found // Nature. 2008. V. 453. P. 80–83.
7. *He W., Sun H., Zhou Y., et al.* Customized binary and multi-level HfO_{2-x}-based memristors tuned by oxidation conditions // Scientific Reports. 2017. V. 7. Article No. 10070.
8. *Yu S., Gao B., Fang Z., et al.* A low energy oxide-based electronic synaptic device for neuromorphic visual systems with tolerance to device variation // Adv. Mater. 2013. V. 25. P. 1774–1779.
9. *Тарков М. С.* Построение сети Хемминга на основе кроссбара с бинарными мемристорами // Прикладная дискретная математика. 2018. № 40. С. 105–113.
10. *Тарков М. С.* Редукция связей автоассоциативной памяти Хопфилда // Прикладная дискретная математика. 2017. № 37. С. 107–113.
11. *Folli V., Leonetti M., and Ruocco G.* On the maximum storage capacity of the Hopfield model // Frontiers in Comput. Neuroscience. 2017. V. 10. Article 144. www.frontiersin.org.

REFERENCES

1. *Haykin S.* Neural Networks: A Comprehensive Foundation. Prentice Hall, Inc., 1999.
2. *Hopfield J.* Neural networks and physical systems with emergent collective computational abilities. Proc. NAS, 1982, vol. 79, pp. 2554–2558.
3. *Personnaz L., Guyon I., and Dreyfus G.* Collective computational properties of neural networks: New learning mechanisms. Phys. Rev., ser. A, 1986, vol. 34, no. 5, pp. 4217–4228.

4. *Michel A. N. and Liu D.* Qualitative Analysis and Synthesis of Recurrent Neural Networks. N.Y., Marcel Dekker Inc., 2002.
5. *Chua L.* Memristor — the missing circuit element. IEEE Trans. Circuit Theory, 1971, vol. 18, pp. 507–519.
6. *Strukov D. B., Snider G. S., Stewart D. R., and Williams R. S.* The missing memristor found. Nature, 2008, vol. 453, pp. 80–83.
7. *He W., Sun H., Zhou Y., et al.* Customized binary and multi-level HfO_{2-x}-based memristors tuned by oxidation conditions. Scientific Reports, 2017, vol. 7, article number: 10070.
8. *Yu S., Gao B., Fang Z., et al.* A low energy oxide-based electronic synaptic device for neuromorphic visual systems with tolerance to device variation. Adv. Mater., 2013, vol. 25, pp. 1774–1779.
9. *Tarkov M. S.* Postroenie seti Khemminga na osnove krossbara s binarnymi memristorami [Construction of a Hamming network based on a crossbar with binary memristors]. Prikladnaya Diskretnaya Matematika, 2018, no. 40, pp. 105–113. (in Russian)
10. *Tarkov M. S.* Reduktsiya svyazey avtoassotsiativnoy pamyati Khopfilda [Reduction of synapses in the Hopfield autoassociative memory]. Prikladnaya Diskretnaya Matematika, 2017, no. 37, pp. 107–113. (in Russian)
11. *Folli V., Leonetti M., and Ruocco G.* On the maximum storage capacity of the Hopfield model. Frontiers in Comput. Neuroscience, 2017, vol. 10, article 144. `www.frontiersin.org`.