

# МОСКОВСКИЙ ИНСТИТУТ ЭЛЕКТРОННОЙ ТЕХНИКИ

Институт системной и программной инженерии  
и информационных технологий (Институт СПИНТех)

## Лабораторный практикум по курсу "Нейронные сети"

Лабораторная работа 5

Изучение нейронной сети Хопфилда.

Информационная емкость ассоциативной памяти.

Это занятие компьютерного практикума посвящено изучению нейронной сети Хопфилда и определению информационной емкости архитектуры, на основе которой осуществляется ее реализация.

Как обычно, прежде чем приступить собственно к программированию, настоятельно рекомендуется ознакомиться с материалом лекций, а также с дополнительными материалами, имеющими отношение к ассоциативной памяти и сети Хопфилда.

Файлы, включенные в задание:

*digits.m* – набор данных, представляющих собой шаблоны цифр от 0 до 9;

*visualize\_number.m* – программа отображения произвольного числа из набора данных, формируемого программой *digits.m*;

*show\_digits.m* – программа отображения всего массива образцов, формируемого программой *digits.m*;

*noisy\_digit.m* – программа, в которой моделируется добавление шума к исходным данным (к каждой отдельной цифре);

*train\_hop\_net.m* – программа, реализующая обучение сети Хопфилда;

*test\_hop\_net\_5.m* – программа тестирования реализованной ассоциативной памяти и извлечения сохраненных образцов (для 5-ти итераций);

*change\_digits\_example.m* – программа изменения исходных данных, представляемых на вход сети Хопфилда.

### 1 Ассоциативная память

Наиболее распространенной группой нейронных сетей, используемых для решения инженерных задач, являются так называемые *нейронные сети прямого распространения* (в частности, *многослойный перцептрон*). Эти нейронные сети, составлявшие предмет наиболее интенсивных исследований до середины 90-х годов, не имеют обратных связей, т.е. связей, идущих от выходов сетей к их входам, что гарантирует безусловную устойчивость нейронной сети. В сетях этой группы распространение сигнала осуществляется, таким образом, от входа к выходу. Эта группа нейронных сетей подробно рассматривается в настоящем учебном курсе «Нейронные сети».

Нейронные сети прямого распространения не могут войти в режим, при котором их выходные значения беспрерывно блуждают от состояния к состоянию. Такое поведение, однако, широко распространено в нейронных сетях с обратными связями.

Модели нейронных сетей данного класса были предложены и проанализированы М. Коэном и С. Гроссбергом (англ. Cohen, Grossberg), а также Д. Хопфилдом (англ. Hopfield). Коэну и Гроссбергу удалось сформулировать и доказать теорему, определившую класс нейронных сетей с обратными связями, выходы которых достигают в динамическом процессе устойчивого состояния. Д. Хопфилд установил, что полносвязная нейронная сеть с симметричной матрицей связей (т.н. нейронная сеть Хопфилда) может представлять собой систему, минимизирующую энергию.

В общем случае сеть с обратной связью моделирует *ассоциативную память* (англ. *Associative memory*). В отличие от обычной машинной памяти (памяти произвольного доступа, или RAM), в которой пользователь задает адрес памяти и оперативное запоминающее устройство (ОЗУ) возвращает слово данных, хранящееся по этому адресу, ассоциативная память оперирует с входным словом данных, поиск которого производится во всей памяти, чтобы выяснить, хранится ли оно где-нибудь в массиве данных. Если слово найдено, возвращается список одного или более адресов хранения, где слово было найдено. Таким образом, ассоциативная память в терминах программирования соответствует *ассоциативному массиву*.

Ассоциативная память представляет собой распределенную память, которая обучается на основе ассоциаций, подобно мозгу живых существ. Это свойство считается основной характерной чертой человеческого мозга со времен Аристотеля, вследствие чего все модели познания в качестве одной из основных операций используют в той или иной форме ассоциативную память.

Существуют два типа ассоциативной памяти: *автоассоциативная* (англ. *Autoassociative memory*) и *гетероассоциативная* (англ. *Heteroassociative memory*).

Решение интеллектуальной задачи с помощью автоассоциативной памяти сводится к запоминанию передаваемых ей образцов (векторов). Затем в эту сеть последовательно подставляются неполные описания или зашумленные представления хранимых в ней исходных образов, и ставится задача *распознавания* конкретного образа.

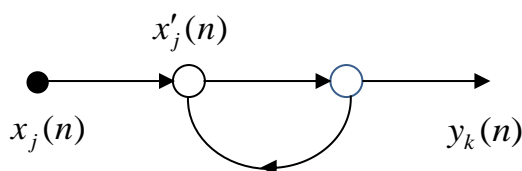
Гетероассоциативная память отличается от автоассоциативной тем, что произвольному набору входных данных ставится в соответствие другой произвольный набор выходных сигналов.

В теории интеллектуальных систем распознавание образов формально определяется как процесс, в котором получаемый образ/многомерный сигнал должен быть отнесен к одному из заранее predetermined классов (категорий). Нейронную сеть Хопфилда в этом смысле можно охарактеризовать как автоассоциативную память, которая может завершить или исправить образ, но не может ассоциировать полученный образ с другим образом. В инженерной практике такая сеть может быть использована также как фильтр и для решения некоторых задач оптимизации.

## 2 Дискретная сеть Хопфилда

### 2.1 Поведение динамической системы с обратной связью

Понятие обратной связи характерно для динамических систем, в которых выходной сигнал возвращается на вход системы и оказывает влияние на сигналы входных элементов, составляющих систему. Граф прохождения сигнала в системе с обратной связью показан на рис. 1.



$D$  $F$ 

Рис. 1. Модель передачи сигнала в системе с одной обратной связью

Предполагается, что эта система линейна и содержит прямую и обратную связи, которые характеризуются операторами  $D$  (англ. Direct) и  $F$  (англ. Feed-back). Если  $n$  – дискретная переменная, характеризующая состояние системы (её параметров, сигналов) во времени, а  $x_j(n)$  – входной сигнал,  $x'_j(n)$  – внутренний сигнал,  $y_j(n)$  – выходной сигнал, то имеет место закономерность:

$$y_k(n) = D[x'_j(n)], \quad (5.1)$$

$$x'_j(n) = x_j(n) + F[y_k(n)], \quad (5.2)$$

где квадратные скобки обозначают действие операторов  $D$  и  $F$ .

Исключая переменную  $x'_j(n)$  в (5.1) и (5.2), будем иметь

$$y_k(n) = \frac{D}{1 - DF} [x_j(n)] \quad . \quad (5.3)$$

В [1] показано, что если оператор  $D$  – это умножение сигнала на фиксированный вес  $w$ , оператор  $F$  является оператором единичной задержки, задерживающим выходной сигнал по отношению к входному на один шаг дискретизации, а  $x_j(n-l)$  – входной сигнал, задержанный на  $l$  единиц дискретизации, то выходной сигнал  $y_k(n)$  можно представить в виде:

$$y_k(n) = \sum_{l=0}^n w^{l+1} x_j(n-l) \quad . \quad (5.4)$$

Динамика системы с обратной связью представлена на рис. 2.

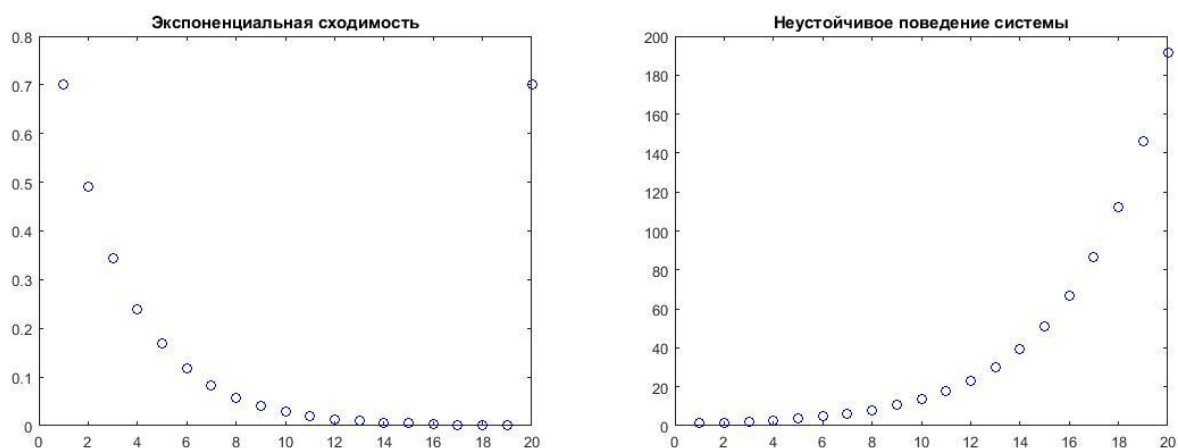


Рис. 2. Реакция системы с обратной связью  
для различных значений  $w$ : (а)  $|w| < 1$ ; (б)  $|w| \geq 1$

Здесь можно выделить два случая:

1. Если  $|w| < 1$ , то выходной сигнал  $y_k(n)$  демонстрирует экспоненциальную сходимость.

Этот случай соответствует устойчивому поведению системы;

2. Если  $|w| \geq 1$ , то выходной сигнал  $y_k(n)$  является расходящейся во времени функцией, что характеризует неустойчивое поведение системы. При этом, если  $|w|=1$ , то расходимость является линейной; если же  $|w| > 1$  - то экспоненциальной.

Анализ динамики поведения нейронных сетей с обратной связью осложняется тем обстоятельством, что нейроны представляют собой элементы с нелинейной функцией активации.

## 2.1 Архитектура нейронной сети Хопфилда

Дискретная сеть Хопфилда, архитектура которой представлена на рис. 3, обладает следующими характеристиками:

- архитектура сети образована одним слоем нейронов (входные элементы, представляющие собой входной образец, не учитываются);
- каждый элемент сети имеет связь со всеми остальными элементами сети, но связь элемента самого с собой отсутствует;
- за один шаг обновляется только один нейрон, в отличие, например, от сети с обратным распространением ошибки, где все элементы слоя могут изменяться одновременно;
- состояние нейронов сети обновляются в случайном порядке, но в среднем каждый элемент должен обновляться равномерно. Например, в случае сети из 100 элементов после 1000 обновлений каждый элемент должен обновиться приблизительно 10 раз;
- Выходные значения нейрона ограничены величинами 0 или 1 (или -1, +1).

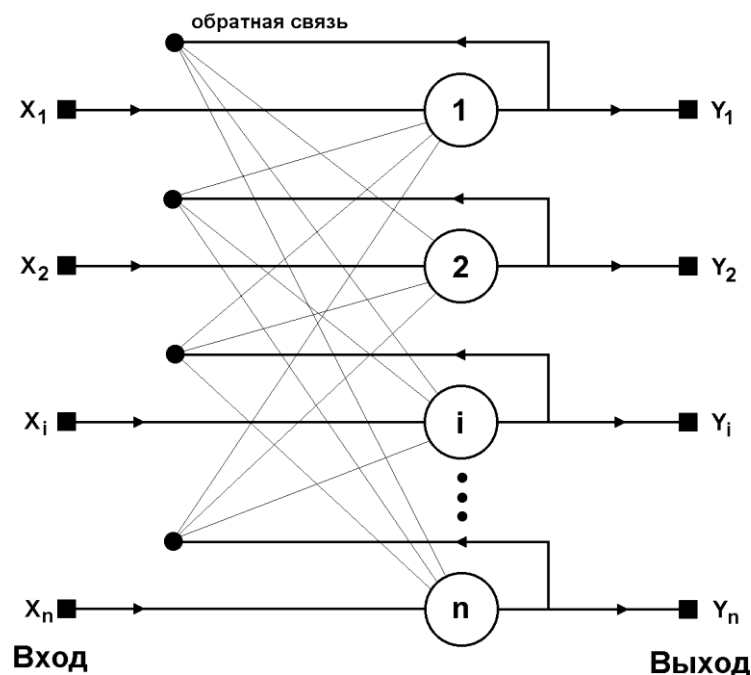


Рис. 3. Архитектура нейронной сети Хопфилда

Представленная на рис. 3 архитектура является рекуррентной в том смысле, что для каждого входного «паттерна» выход сети используется повторно в качестве входа до тех пор, пока не будет достигнуто устойчивое состояние.

Значение активности нейрона в сети Хопфилда получается на основе использования некоторого правила активации. При этом состояние сети в произвольный момент времени задается вектором состояний всех её элементов.

В качестве входных данных сети Хопфилда можно использовать +1 для обозначения активного состояния нейрона, и -1 - для состояния «выключено».

Комбинированный ввод элемента вычисляется по формуле:

$$net_j = \sum_{i=1}^n s_i w_{ij}, \quad (5.5)$$

где  $s_i$  обозначает состояние нейрона с номером  $i$ . Когда состояние нейрона изменяется, его состояние рассчитывается в соответствии с правилом

$$s_j = \begin{cases} +1, & \text{если } net_j > 0, \\ -1, & \text{если } net_j < 0. \end{cases} \quad (5.6)$$

Эта зависимость описывается «сигнум» функцией (от лат. *signum* – знак). Если комбинированный ввод оказывается равным нулю, то элемент остается в состоянии, в котором он пребывал перед обновлением.

Сеть работает следующим образом: входной вектор задает начальные состояния всех элементов (нейронов) сети. Случайным образом выбирается нейрон, состояние которого будет изменено. Выбранный элемент получает взвешенные (умноженные на значение весов) сигналы от всех остальных элементов и изменяет свое состояние. Случайным образом выбирается другой элемент, и процесс повторяется. Устойчивое состояние достигается в том случае, если ни один из элементов сети Хопфилда, будучи выбранным для обновления, не меняет своего состояния.

## 2.2 Фаза сохранения прототипов рекуррентной сетью

Сеть Хопфилда ведет себя, таким образом, как память, а процедура сохранения отдельного вектора в «фундаментальную память» представляет собой вычисление прямого произведения вектора с ним самим. Матрица, задающая весовые коэффициенты для сети Хопфилда, имеет вид:

$$\mathbf{W} = \mathbf{x}^T \mathbf{x}, \quad (5.7)$$

при этом все ее диагональные элементы должны равняться нулю.

## 2.4 Энергетические характеристики процесса извлечения образа

Хопфилдом было доказано, что нейронная сеть, представленная на рис. 3, должна сходиться к устойчивому набору значений системы. Изменение состояния нейронов сети осуществляется по пороговому значению, равному нулю, что приводит к следующей функции энергии архитектуры с обратными связями:

$$E = -\frac{1}{2} \sum_j \sum_i s_j s_i w_{ij}. \quad (5.8)$$

Если нейрон  $j$  изменяет свое состояние на  $\Delta s_j$ , то изменение энергии будет равно

$$\Delta E = -\Delta s_j \sum_{ii} s_i w_{ij}. \quad (5.9)$$

Изменение энергии следует рассматривать как функцию  $\Delta s_j$  и  $\sum_{ii} s_i w_{ij}$ . Результат Хопфилда заключается в том, что  $\Delta s_j$  и  $\sum_{ii} s_i w_{ij}$  имеют одинаковые знаки, а поэтому в итерационном процессе энергия системы уменьшается.

## 2.3 Информационная емкость сети Хопфилда

В литературе можно найти оценки, касающиеся числа образцов фундаментальной памяти  $M$ , которые можно сохранить с помощью нейронной сети Хопфилда, если требовать правильного распознавания большинства образцов [2]:

$$M_{\max} = \frac{N}{2 \ln N}, \quad (5.10)$$

где  $N$  – количество нейронов сети Хопфилда, равное размерности входного вектора (двоичного слова), а  $\ln$  – натуральный логарифм.

## 3 Задания

### 3.1 Задание 1

Основываясь на материале лекций, а также [1], продемонстрировать вывод соотношения (5.4) из (5.3).

### 3.2 Задание 2

Написать программу, с помощью которой вычисляются значения выходного сигнала  $y_k(n)$  в (5.4). Проанализировать поведение системы с одной обратной связью (рис. 1) для случая  $|w| < 1$  и  $|w| \geq 1$ . Как ведет себя система, если  $|w| = 1$ ?

### 3.3 Задание 3

Найдите набор весовых коэффициентов сети Хопфилда  $\mathbf{W}$ , соответствующий сохранению единственного образца  $\mathbf{x}_{10} = 5_{10}$ , представленного в биполярной двоичной системе счисления (+1 – активный нейрон, -1 – нейрон, находящийся в выключенном состоянии), т.е.  $\mathbf{x}_2^{\{-1,+1\}} = [1; -1; 1]$ . Ответ должен соответствовать архитектуре нейронной сети с количеством нейронов, допускающим запоминание максимального десятичного числа  $15_{10}$ , т.е.  $\mathbf{x}_2^{\{-1,+1\}} = [1; 1; 1; 1]$ . Нижние символы соответствуют представлению числа в двоичной биполярной и десятичной системах счисления, а верхние подчеркивают биполярность двоичной системы.

Указание: Используйте функцию `dec2bin(d)`, возвращающую строку двоичных символов (0 и 1), эквивалентную десятичному числу  $d$ .

### 3.4 Задание 4

Напишите программу-функцию, вычисляющую значения элементов матрицы весовых коэффициентов нейронной сети Хопфилда  $\mathbf{W}$ , предназначенной для одновременного запоминания нескольких образцов. Входные аргументы функции:  $N$  – количество нейронов сети Хопфилда,  $\mathbf{x}$  – вектор, содержащий образцы для запоминания, размерность которого не превышает  $N \times M_{\max}$ . Если количество элементов, представляемых для запоминания, превышает  $M_{\max}$  (см. формулу (5.4)), то функция должна выдавать сообщение об ошибке.

Указание: Если набор входных образцов имеет вид  $\mathbf{x}_{10} = \{2, 5\}$ , или, что то же самое  $\mathbf{x}_2^{\{-1,+1\}} = [-1 \ 1 \ -1; 1 \ -1 \ 1]$ ; то соответствующей искомой матрицей является матрица

$\mathbf{W} =$

0      -2      2

```

-2      0      -2
 2      -2      0.

```

### 3.5 Обработка идеальных входных данных

Изучим способность рекуррентной нейронной сети запоминать и извлекать визуальные образы. Такими образами могут, в частности, выступать шаблоны цифр от 0 до 9, которые формируются программой `digits.m`. Вывод отдельной цифры на экран (например, цифры «4») производится командой `visualize_number(four)` или нескольких цифр одновременно (например, девяти) командой `show_digits({zero one two three four five six seven eight})` или командой `show_digits(digits)`, которая отображает все образцы одновременно.

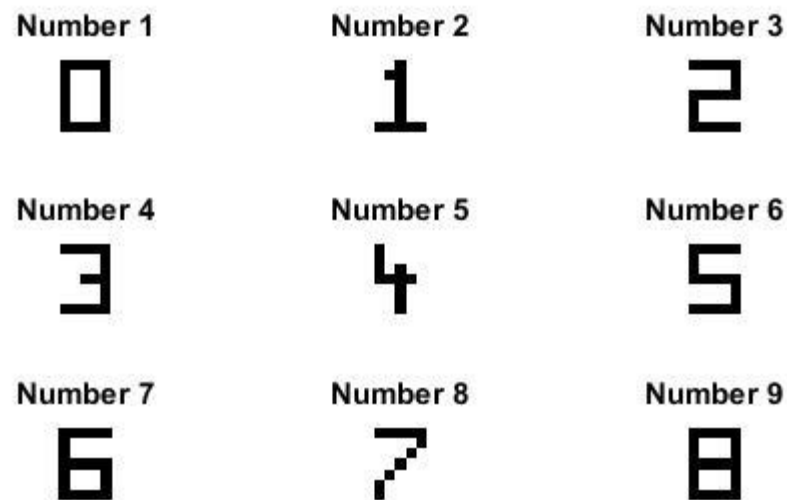


Рис. 4. Отображение девяти элементов фундаментальной памяти

Настройка весовых коэффициентов сети Хопфилда производится с помощью команды `w = train_hop_net({three seven})`, где `w` – весовые коэффициенты нейронной сети (см. также задание 4). После завершения стадии запоминания (в данном случае образов двух цифр – 3 и 7), можно проверить способность рекуррентной архитектуры извлекать элементы из фундаментальной памяти:

```

x = three;
y = test_hop_net(w, x, 1);
visualize_number(y),

```

где `x` - входной образ (может представлять собой и зашумленную версию исходного вектора), а `1` задаёт одну итерацию. Тестирование нейронной сети Хопфилда в течение 5 итераций и отображение результата `y = test_hop_net_5(w, x)`.

**Задание:** Требуется подготовить свой собственный набор шаблонов цифр (например, как в почтовой индексации) и продемонстрировать извлечение подготовленных образов.

**Указание:** Можно воспользоваться программой `change_digit_example.m`, имеющейся в каталоге.

### 3.6 Обработка зашумленных входных данных

Продемонстрировать извлечение элементов фундаментальной памяти при подаче на вход сети Хопфилда зашумленного входного вектора.

Указание: Можно воспользоваться программой `noisy_digit(zero, 0.3)`, в которой второй параметр определяет уровень шума, добавляемого к исходному вектору.

## 4 Литература

- [1] Хайкин С. Нейронные сети. Полный курс. – Изд-во «Вильямс», Москва, 2006. – 1104 с.
- [2] Каллан Р. Основные концепции нейронных сетей / Пер. с англ. — М.: Изд-во «Вильямс», 2001. — 287 с.
- [3] Осовский С. Нейронные сети для обработки информации/ Пер. с польск. – М.: Финансы и статистика, 2004, – 344с.