

ФИЗИОЛОГИЧЕСКИЕ МЕХАНИЗМЫ ПОВЕДЕНИЯ ЖИВОТНЫХ:  
ВОСПРИЯТИЕ ВНЕШНИХ СТИМУЛОВ, ДВИГАТЕЛЬНАЯ  
АКТИВНОСТЬ, ОБУЧЕНИЕ И ПАМЯТЬ

УДК 612.821

НЕЙРОСЕТЕВАЯ МОДЕЛЬ РАСПОЗНАВАНИЯ ИЗВЕСТНОСТИ  
ЧИСЛОВЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

© 2020 г. Я. Б. Казанович\*

Институт математических проблем биологии РАН – филиал Института прикладной математики  
им. М.В. Келдыша РАН, г. Пушино, Россия

\*e-mail: yasha@impb.ru

Поступила в редакцию 15.07.2019 г.

После доработки 09.12.2019 г.

Принята к публикации 16.12.2019 г.

Одним из компонентов распознающей памяти является распознавание известности, когда испытуемый помнит сам факт того, что данный объект или событие встречались ему раньше, но может не помнить деталей объекта/события, а также контекста, в котором данный объект/событие имели место. Эксперименты Стэндинга [Standing, 1970; Standing, 1973] выявили парадоксально высокую емкость памяти при распознавании известности естественных изображений, слов и музыкальных мелодий. Имеющиеся нейросетевые модели распознавания известности продемонстрировали возможность распознавания известности с памятью порядка  $n^2$ , где  $n$  – число нейронов в модели. В данной работе предложена новая модель распознавания известности, ориентированная на распознавание известности временных последовательностей (в частности, последовательностей чисел), отличающаяся разреженным кодированием входных паттернов. С помощью компьютерных экспериментов показано, что удельная емкость памяти модели при определенных условиях безошибочного распознавания больше, чем в известных моделях хопфилдовского типа.

*Ключевые слова:* распознающая память, распознавание известности, нейросетевые модели, емкость памяти

DOI: 10.31857/S0044467720030053

Распознающая память [Ameen-Ali et al., 2015] – один из видов декларативной памяти – подразделяется на два типа [Yonelinas, 2002]: 1. вспоминание (recollection), при котором воспроизводится объект/событие вместе с контекстом; 2. распознавание известности (familiarity), при котором устанавливается факт знакомства с объектом/событием, но какая-либо контекстная информация отсутствует. При экспериментальном исследовании распознающей памяти время, затрачиваемое на вспоминание, обычно существенно больше, чем время, затрачиваемое на распознавание известности. Таким образом, можно думать, что сначала происходит распознавание самого факта, что данный объект знаком, а затем вспоминаются характеристики объекта и контекст, в котором объект ранее встречался, т.е. что эти два процесса диссо-

цированы, по крайней мере, на уровне поведения.

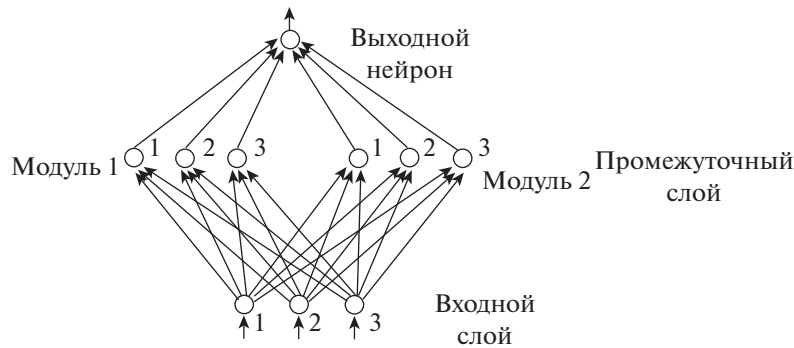
Некоторые аспекты распознавания известности являются парадоксальными с точки зрения того, что известно о функционировании различных видов памяти. Например, в то время как объем кратковременной памяти обычно оценивается как  $7 \pm 2$  элементарных объектов (чанков) [Miller, 1956], объем памяти при распознавании известности зрительных объектов практически неограничен [Standing, 1970; Standing, 1973]. Заметим, что память в экспериментах Стэндинга формировалась путем короткого (5 с) *однократного* предъявления большого числа стимулов (от 20 до 10000 стимулов), поэтому кажется удивительным, что ее емкость так радикально отличается в большую сторону от емкости кратковременной памяти.

В теоретической и экспериментальной литературе, посвященной распознающей памяти, нет единого мнения относительно того, являются ли два ее компонента различными стадиями одного процесса или это два разных когнитивных процесса. Если верно последнее, то возникают следующие вопросы: реализуются ли эти процессы различными структурами мозга и/или различными нейронными популяциями в структурах мозга; работают ли эти процессы с одними и теми же или с разными следами памяти; какие нейронные структуры задействованы в реализации указанных типов распознающей памяти и какие нейронные механизмы при этом используются.

Сторонники однопроцессорной теории (single-process theory – SPT) утверждают, что вспоминание является лишь более детальной версией распознавания известности. При этом распознающая память рассматривается как континуум, где сила следов памяти может варьировать от слабых до сильных [Squire et al., 2007; Medina, 2008; Rutishauser et al., 2008; Wixted, Squire, 2011; Slotnick, 2013]. Экспериментальные данные, свидетельствующие в пользу двухпроцессорной теории (dual-process theory – DPT), основаны на результатах фМРТ и на исследованиях пациентов с нарушениями мозга, приводящими к различным типам амнезии. Обзор теоретических аргументов в пользу DPT приведен в работе [Diana et al., 2006]. Согласно DPT, имеются две области, где реализуются процессы распознающей памяти, – это гиппокамп и медиальная височная кора. В одних работах приводятся аргументы в пользу того, что гиппокамп отвечает за вспоминание, а периринальная кора – за распознавание известности [Aggleton et al., 2005; Eichenbaum et al., 2007; Bowles et al., 2010; Montaldi, Mayers, 2010]. В других исследованиях получены экспериментальные данные, свидетельствующие, что гиппокамп участвует в обоих видах распознающей памяти [Wais et al., 2006; Merkow et al., 2015]. Кроме того, обнаружено, что повреждение энторинальной коры ведет к дефициту распознавания известности, но не вспоминания [Brandt et al., 2016]. Различие между структурами, ответственными за вспоминание и распознавание известности, распространяется и за пределы медиальной височной коры, в частности, на префронтальную кору [Scalici et al., 2017].

О.С. Виноградова на основе нейрофизиологических экспериментов по ориентационному рефлексу выдвинула теорию о ведущей роли поля СА3 гиппокампа в детекции новизны [Vinogradova, 2001]. Правда, следует отметить, что эксперименты Виноградовой относились к рабочей памяти грызунов, а не к распознающей (долговременной) памяти человека. Распознающая память в связи с обонятельными стимулами исследовалась в работах [Feinberg et al., 2012; Кедров, Анохин, 2019], ее функционирование продемонстрировало вовлеченность тех же структур медиальной височной коры, что и в случае стимулов других модальностей.

Хотя очевидно, что распознающая память является сложно организованной системой [Kafkas, Montaldi, 2018; Bastin et al., 2019], теория DPT позволяет рассчитывать на относительную независимость функционирования периринальной коры при распознавании известности. Опираясь на эту гипотезу, авторы ряда компьютерных моделей с той или иной степенью успешности воспроизвели как психофизические результаты Стэндинга, так и нейрофизиологические данные относительно активности в периринальной коре в процессе распознавания известности [Androulidakis et al., 2008; Bogacz, Brown, 2003]. Модели [Bogacz et al., 2001; Bogacz, Brown, 2002] реализуют распознавание известности на основе энергии в сети Хопфилда [Hopfield, 1982] (или ее аналога), используя тот факт, что модуль этой функции больше для известных паттернов, чем для новых. Ряд моделей были построены на основе двухслойных сетей с прямым распространением сигналов. Использовалось либо антихеббовское обучение [Brown, Xiang, 1998; Bogacz, Brown, 2003], либо обучение на основе так называемого алгоритма Info-max [Androulidakis et al., 2008], которое также приводит к ослаблению связей между слоями. В работе [Bogacz, Brown, 2002] было показано, что емкость распознающей памяти имеет порядок  $0.023n^2$ , где  $n$  – число элементов в нейронной сети. Заметим, что при стандартном хеббовском обучении в сетях Хопфилда емкость памяти составляет порядка  $0.14n$  [Amit, 1989]. Таким образом, моделирование показало, что на уровне формальной нейронной сети можно воспроизвести экспериментальные данные по радикальному увеличению емкости памяти при запоминании без контекста (с последующим распознаванием известности) по сравнению с запомина-



**Рис. 1.** Архитектура модели для случая  $n = 3, m = 2$ .  
**Fig. 1.** Architecture of connections for the case  $n = 3, m = 2$ .

нием с контекстом (с последующим воспоминанием).

В данной работе будет описана модель распознавания известности, которая отличается от разработанных ранее в следующих отношениях.

1. Модель реализует запоминание не стационарных бинарных паттернов, а последовательностей целых чисел. Таких моделей в настоящее время нет.

2. Модель использует разреженное кодирование входных паттернов, и такое кодирование принципиально для получения емкости памяти для распознавания известности порядка  $n^2$ .

3. При определенных условиях безошибочного распознавания модель имеет емкость памяти большую, чем модели хопфилдовского типа.

Как и в других моделях распознавания известности предполагается, что эта функция реализуется в периринальной коре. Согласно данным работы [Kafkas, Montaldi, 2018] в периринальной коре имеются две системы, одна из которых ответственна за распознавание известности, а другая за распознавание новизны. Поскольку в нашей модели активность в ответ на известные последовательности увеличивается, естественно ассоциировать эту модель с системой распознавания известности периринальной коры. Однако это не является существенным ограничением, так как распознавание новизны может быть инкорпорировано в модель путем дополнительного введения обратного торможения от выходного слоя к входному.

## МОДЕЛЬ

Модель предназначена для запоминания и распознавания известности последовательностей, состоящих из  $m$  целых чисел в диапазоне от 1 до  $n$ . Другие интерпретации использования модели обсуждаются в Заключении.

### *Архитектура нейронной сети*

Сеть состоит из входного слоя, содержащего  $n$  нейронов, промежуточного слоя, содержащего  $m$  однослойных модулей с  $n$  нейронами в каждом модуле, и выходного слоя, состоящего из одного нейрона, предназначенного для принятия решения относительно известности или неизвестности предъявленной последовательности (рис. 1).

Входная последовательность идентифицируется как новая, если в ответ на ее предъявление выходной нейрон генерирует сигнал 0; последовательность идентифицируется как известная, если в ответ на ее предъявление выходной нейрон генерирует сигнал 1.

Все связи в сети имеют тип снизу-вверх. Входной слой имеет связи все-на-всех со всеми нейронами промежуточного слоя. Все веса связей предполагаются неизменными и равными 1.

Все нейроны промежуточного слоя подают сигнал на выходной нейрон. Нейроны входного слоя получают бинарные сигналы и генерируют на выходе целочисленные сигналы, суммируя сигналы, приходящие на их входы. Нейроны промежуточного и выходного слоев генерируют бинарные сигналы 0 или 1.

Запоминание последовательностей происходит путем модификации связей между входным и промежуточным слоями. В отличие от стандартной процедуры запоминания,

основанной на модификации весов синаптических связей, в модели использовано запоминание путем модификации параметров, ограничивающих постсинаптический потенциал. В начальном состоянии эти параметры равны 0, т.е. нейроны второго слоя в ответ на предъявление любой входной последовательности генерируют на выходе нулевой сигнал, что означает, что любая последовательность, подаваемая на “необученную” сеть, идентифицируется как новая. В процессе запоминания параметры, ограничивающие постсинаптический потенциал, принимают целочисленные значения, которые монотонно растут. Связи модифицируются только в том случае, если входная последовательность идентифицирована как новая (не предъявлявшаяся ранее). Правило модификации связей таково, что однажды предъявленная последовательность будет в дальнейшем распознаваться как известная. В результате процесса запоминания все большее количество связей будет модифицировано, в силу чего могут возникать ошибки, когда новая последовательность распознается как известная (переполнение памяти).

#### Работа модели

Функционирование модели разбито на три этапа. На первом этапе сеть отвечает на входную последовательность генерацией определенного типа активности нейронов. В этот период в дискретные моменты времени  $t = 1, 2, \dots, m$  на нейроны входного слоя подаются внешние сигналы. В каждый момент времени сигнал приходит лишь на один из нейронов. Входные последовательности задаются в виде последовательности номеров нейронов  $s_1, s_2, \dots, s_m$ , где  $s_t = \{1, 2, \dots, n\}$  ( $t = 1, \dots, m$ ). Эта последовательность означает, что на нейрон входного слоя с номером  $s_t$  приходит сигнал 1 в момент времени  $t$ . Активность, возникающая в нейронах входного слоя, с помощью связей передается в промежуточный и затем в выходной слой.

На втором этапе происходит распознавание известности. Последовательность распознается как известная, если выходной нейрон генерирует сигнал 1. В случае, когда выходной нейрон генерирует сигнал 0, последовательность распознается как новая.

На третьем этапе, если входная последовательность идентифицируется как новая, происходит модификация связей между входным

и промежуточным слоями. При предъявлении каждой последовательности отсчет времени начинается заново с 0.

#### Правила функционирования нейронов

Нейроны в модели являются формальными. Это суммирующие или интегративно-пороговые устройства, которые не претендуют на воспроизведение динамики мембранного потенциала. Нейроны входного слоя являются беспороговыми сумматорами. Они суммируют все приходящие на них сигналы за период времени от 1 до  $t$ . Перед подачей входной последовательности, т.е. в момент времени 0, все потенциалы входных нейронов обнуляются,  $V_i^1(0) = 0$  ( $i = 1, \dots, n$ ). Далее эти потенциалы модифицируются в соответствии со следующей формулой:

$$V_i^1(t) = \begin{cases} V_i^1(t-1) + 1, & i = s_t, \\ V_i^1(t-1), & i \neq s_t, \end{cases} \quad (1)$$

$$t = 1, \dots, m, \quad i = 1, \dots, n.$$

Например, если входная последовательность равна (1, 2, 1), то в моменты времени 0, 1, 2, 3 выходные сигналы первых двух нейронов входного слоя будут соответственно (0, 0), (1, 0), (1, 1) и (2, 1).

Сумма всех сигналов, генерируемых нейронами входного слоя в дискретный момент времени  $t$ , равна  $t$ . Максимальное значение этой суммы достигается в последний момент предъявления элементов последовательности и равно  $m$ .

Нейроны промежуточного и выходного слоев – бинарные с порогом, равным  $m$ , т.е. их выходные сигналы равны 0, если суммарный входной сигнал, приходящий на них от входных нейронов, меньше  $m$ , и равны 1 в противном случае.

При подаче на нейрон промежуточного слоя входных сигналов выходной сигнал формируется в соответствии со следующей формулой:

$$V_{jk}^2(t) = \begin{cases} \theta \left( \sum_{i=1}^n \min(V_i^1(t), \lambda_{ijk}) - m \right), & j = s_k, \quad k \leq t, \\ 0, & j \neq s_k \quad \text{или} \quad k > t, \end{cases} \quad (2)$$

$$k = 1, \dots, m, \quad j = 1, \dots, n,$$

где  $\theta(x)$  – функция Хевисайда ( $\theta(x) = 0$  при  $x < 0$ ,  $\theta(x) = 1$  при  $x \geq 0$ );  $V_{jk}^2$  – выходной сигнал, генерируемый  $j$ -м нейроном  $k$ -го модуля

промежуточного слоя в момент  $t$ ;  $V_i^1(t)$  — сигнал, поступающий на нейроны промежуточного слоя от  $i$ -го нейрона входного слоя в момент времени  $t$ ;  $\lambda_{ijk}$  — параметр, ограничивающий рост постсинаптического потенциала, формируемого в результате синаптической передачи в  $j$ -м нейроне модуля  $k$  промежуточного слоя под действием сигнала от  $i$ -го нейрона входного слоя.

Заметим, что ненулевой сигнал в модуле  $k$  может генерировать только нейрон  $s_k$ . Выходной сигнал нейрона  $s_k$  модуля  $k$  до момента  $m$  равен 0. В момент  $m$ , когда предъявление последовательности заканчивается, он равен 0 или 1. Значение 1 имеет место в том случае, если последовательность известная или ошибочно распознается данным модулем как известная. Согласно приведенной формуле, сигналы, поступающие на нейрон промежуточного слоя, проходят через два порога: один порог ограничивает постсинаптический потенциал, а второй порог задает преобразование суммарного сигнала в выходной сигнал.

Из приведенной формулы следует, что для того чтобы нейрон  $j$  модуля  $k$  в промежуточном слое сработал (сгенерировал единичный выходной сигнал), необходимо, чтобы для всех  $i = 1, \dots, n$  имело место  $V_i^1(m) \leq \lambda_{ijk}$ . Другими словами, параметр  $\lambda_{ijk}$  характеризует “пропускную способность” синапса по передаче сигнала от пресинаптического нейрона к постсинаптическому. Если этот порог ниже, чем подаваемый на синапс сигнал, постсинаптический нейрон получит недостаточный входной суммарный сигнал для своего возбуждения.

Нейрон выходного слоя преобразует входные сигналы в выходные по следующей формуле

$$V_{\text{out}}^3(t) = \theta \left( \sum_{k=1}^m \sum_{j=1}^n V_{jk}^2(t) - m \right) = \theta \left( \sum_{k=1}^m V_{s_k k}^2 - m \right), \quad (3)$$

где  $V_{jk}^2(t)$  — входной сигнал, приходящий в момент времени  $t$  на выходной нейрон от  $j$ -го нейрона  $k$ -го модуля промежуточного слоя. Таким образом, для того чтобы в ответ на входную последовательность  $s_1, s_2, \dots, s_m$  выходной нейрон сработал (т.е. сгенерировал выходной сигнал, равный 1), нужно, чтобы в момент времени  $m$  сработали нейроны  $s_1$  в модуле 1,  $s_2$  в модуле 2, ...,  $s_m$  в модуле  $m$ . До момента  $m$  выходной нейрон генерирует ну-

левой сигнал, поскольку до этого момента сумма входных сигналов на выходной нейрон меньше  $m$ .

#### Модификация связей

До начала запоминания последовательностей все параметры, ограничивающие постсинаптические потенциалы нейронов промежуточного слоя, устанавливаются равными 0. Если последовательность распознается как новая, то связи между входным слоем и промежуточным слоем модифицируются в момент окончания предъявления всей последовательности. Модификации подвергаются только связи от входного слоя к нейронам всех модулей промежуточного слоя. Новая величина  $\lambda_{ijk}^{\text{new}}$  вычисляется по формуле

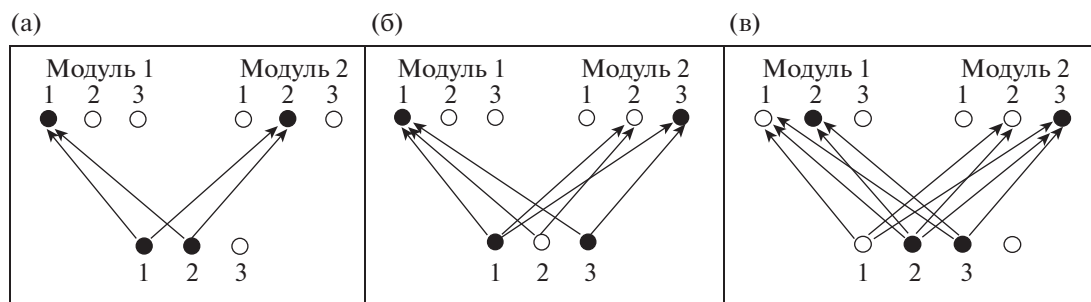
$$\lambda_{ijk}^{\text{new}} = \begin{cases} \max(V_i^1(m), \lambda_{ijk}^{\text{old}}), & j = s_k, \\ \lambda_{ijk}^{\text{old}}, & j \neq s_k, \end{cases} \quad (4)$$

$$i = 1, \dots, n, \quad j = 1, \dots, n, \quad k = 1, \dots, m,$$

где  $V_i^1(m)$  — сигнал, поступающий на нейроны промежуточного слоя от  $i$ -го нейрона входного слоя после предъявления всей входной последовательности;  $\lambda_{ijk}^{\text{old}}$  — текущее значение параметра, ограничивающего постсинаптический потенциал.

Указанная модификация связей гарантирует, что если какая-то входная последовательность предъявлена повторно, то при окончании второго предъявления в каждом модуле промежуточного слоя будет один активный нейрон и, следовательно, эта последовательность будет идентифицирована как известная. Более того, поскольку связи между входным и промежуточными слоями могут только расти, последовательность, которая однажды была идентифицирована как известная, в дальнейшем при любом предъявлении также будет распознана как известная.

По мере предъявления большого набора последовательностей (и соответственно модификации все большего количества связей) память модели может “переполниться”. В результате новые последовательности могут быть ошибочно распознаны как известные. Это происходит тогда, когда параметры  $\lambda_{ijk}$  достаточно велики, чтобы для новой последовательности выполнялись условия: для всех  $i = 1, \dots, n$  имеет место  $V_i^1(m) \leq \lambda_{ijk}$ . Чем меньше остается нулевых (не модифицированных) связей, тем больше вероятность та-



**Рис. 2.** Адаптации связей в нейронной сети с параметрами  $n = 3$ ,  $m = 2$  при последовательном предъявлении на вход трех последовательностей (а) (1, 2), (б) (1, 3), (в) (2, 3). Показаны только ненулевые связи между входным и промежуточным слоями. Нейроны, активные после предъявления всех элементов каждой последовательности, показаны черным цветом.

**Fig. 2.** Adaptation of connection strengths in a network with the parameters  $n = 3$ ,  $m = 2$ . Three sequences are sequentially presented at the input: (a) (1, 2), (б) (1, 3), (в) (2, 3). Filled black circles show the neurons that are active after all members of a particular sequence are presented.

кого события. Считая, что модификация связей имеет случайный и независимый характер, вероятность  $P_L$  того, что в процессе запоминания  $M$  последовательностей длины  $m$  будет модифицировано  $L$  связей, может быть вычислена, исходя из комбинаторных соображений:

$$P_L = \frac{\binom{N}{L} \binom{K-1}{K-L}}{\binom{N+K-1}{K}}, \quad (5)$$

$$1 \leq L \leq \min(K, N), \quad N = n^2, \quad K = mM.$$

## РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

### Пример 1

На рис. 2 приведены результаты адаптации связей в нейронной сети с параметрами  $n = 3$ ,  $m = 2$  при последовательном предъявлении на входе трех последовательностей (1, 2), (1, 3), (2, 3).

**Пример 2** (ошибка распознавания известности)

В табл. 1 приведены результаты модификации связей между входным и промежуточным слоями при предъявлении набора из 9 последовательностей. Последняя девятая последовательность ошибочно идентифицирована как известная.

### Оценка емкости памяти при распознавании известности

Для определения емкости памяти при распознавании известности на вход сети подавались различные случайные последовательности с независимыми компонентами, равномерно распределенными в диапазоне от 1 до  $n$ . Предъявление последовательностей продолжалось до тех пор, пока не возникала ошибка распознавания известности, т.е. новая последовательность идентифицировалась как известная. Емкость сети определялась как число последовательностей, распознавание известности которых проходило безошибочно. Средняя емкость  $C_{\text{авг}}$  определялась как среднее значение емкости, полученное при 100-кратном повторении вычисления емкости для разных наборов последовательностей. Данные о средней емкости  $C_{\text{авг}}$  при разреженном кодировании ( $m \ll n$ ), полученные в результате симуляции для параметра модели  $m = 3, 4, 5$ , характеризующего разреженность входных паттернов, приведены на рис. 3.

Приближение квадратичными полиномами значений  $C_{\text{авг}}$ , полученных в вычислительных экспериментах для значений параметра  $m = 3, 4, 5$ , дало следующие результаты:

$$y_1 = 0.0601x^2 + 6.5816x - 32.6615;$$

$$y_2 = 0.2x^2 + 4.8318x - 1.8075;$$

$$y_3 = 0.2612x^2 + 4.9592x - 5.145.$$

**Таблица 1.** Формирование связей между входным и промежуточным слоями при подаче на вход набора из 9 последовательностей,  $n = 3, m = 3$

**Table 1.** Modification of connections between the input and intermediate layers for the input set of 9 sequences,  $n = 3, m = 3$

| Номер | Входные последовательности | Связи между входным слоем и модулем 1                      | Связи между входным слоем и модулем 2 | Связи между входным слоем и модулем 3 |
|-------|----------------------------|--|---------------------------------------|---------------------------------------|
| 1     | 2, 3, 2                    | 0 0 0<br>0 2 0<br>0 1 0                                    | 0 0 0<br>0 0 2<br>0 0 1               | 0 0 0<br>0 2 0<br>0 1 0               |
| 2     | 2, 1, 3                    | 0 1 0<br>0 2 0<br>0 1 0                                    | 1 0 0<br>1 0 2<br>1 0 1               | 0 0 1<br>0 2 1<br>0 1 1               |
| 3     | 1, 3, 3                    | 1 1 0<br>0 2 0<br>2 1 0                                    | 1 0 1<br>1 0 2<br>1 0 2               | 0 0 1<br>0 2 1<br>0 1 2               |
| 4     | 3, 3, 2                    | 1 1 0<br>0 2 1<br>2 1 2                                    | 1 0 1<br>1 0 2<br>1 0 2               | 0 0 1<br>0 2 1<br>0 2 2               |
| 5     | 1, 3, 1                    | 2 1 0<br>0 2 1<br>2 1 2                                    | 1 0 2<br>1 0 2<br>1 0 2               | 2 0 1<br>0 2 1<br>1 2 2               |
| 6     | 3, 2, 2                    | 2 1 0<br>0 2 2<br>2 1 2                                    | 1 0 2<br>1 2 2<br>1 1 2               | 2 0 1<br>0 2 1<br>1 2 2               |
| 7     | 2, 1, 1                    | 2 2 0<br>0 2 2<br>2 1 2                                    | 2 0 2<br>1 2 2<br>1 1 2               | 2 0 1<br>1 2 1<br>1 2 2               |
| 8     | 1, 1, 1                    | 3 2 0<br>0 2 2<br>2 1 2                                    | 3 0 2<br>1 2 2<br>1 1 2               | 3 0 1<br>1 2 1<br>1 2 2               |
| 9     | 2, 3, 1                    | Последовательность ошибочно идентифицирована как известная |                                       |                                       |

В приведенных выражениях  $x$  является непрерывным аналогом переменной  $n$ , а  $u$  заменяет переменную  $C_{avr}$  (рис. 3).

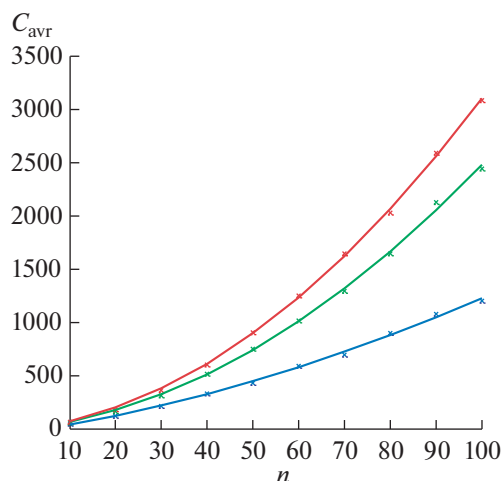
Удельная емкость в расчете на один модифицируемый синапс по порядку равна  $0.06 : 3 = 0.02$  для  $m = 3$ ;  $0.2 : 4 = 0.05$  для  $m = 4$ ;  $0.26 : 5 = 0.0512$  для  $m = 5$ . В последних двух случаях удельная емкость более чем в два раза превосходит удельную емкость 0.023, полученную для модели распознавания известности хопфилдовского типа [Bogacz, 2001; Bogacz, Brown, 2002].

### ЗАКЛЮЧЕНИЕ

Рассмотренная модель распознавания известности последовательностей целых чисел построена из простейших интегративно-пороговых элементов и обучается с помощью

простой процедуры снизу-вверх. Основное отличие модели от существующих моделей распознавания известности состоит в использовании нескольких модулей во втором слое и принципиальной роли разреженного кодирования входной информации. С помощью компьютерных вычислений показано, что модель имеет достаточно большую емкость памяти. Для сети из 100 входных нейронов при запоминании последовательностей чисел из пяти членов емкость памяти (среднее число безошибочно запоминаемых последовательностей) превосходит 3000 последовательностей при общем числе нейронов в сети 601. Существенно, что распознавание известности происходит в момент окончания предъявления последовательности с помощью параллельной процедуры, не требуя





**Рис. 3.** Графики зависимости средней емкости от размера входного слоя (аппроксимация квадратичными полиномами). Нижняя (синяя) кривая  $m = 3$ ; средняя (зеленая) кривая  $m = 4$ ; верхняя (красная) кривая  $m = 5$ .

**Fig. 3.** Dependence of the average capacity on the size of the input layer (approximation by quadratic polynomials). Lower curve (blue)  $m = 3$ ; middle curve (green)  $m = 4$ ; upper curve (red)  $m = 5$ .

просмотра списка запомненных последовательностей.

Предложенная модель распознавания известности была сформулирована для случая, когда входная последовательность представляет собой последовательность целых чисел в диапазоне от 1 до  $n$ . На самом деле, возможные условия для использования модели намного шире, чем это может показаться, исходя из сделанных ограничений.

Элементы последовательности можно интерпретировать как дискретизованные (в диапазоне 1, ...,  $n$ ) длительности временных интервалов (между событиями). В такой терминологии речь идет о запоминании (и распознавании известности) различных ритмов. Исходную информацию для задания временных интервалов могут предоставлять клетки времени [Eichenbaum, 2014].

Входная последовательность может быть образована, например, из номеров лампочек, расположенных в виде цепочки, которые загораются в определенной последовательности.

Входная последовательность может быть сформирована из звуков разной частоты. Предполагается, что имеется  $n$  возможных тонов, а последовательность звуков образует определенную “мелодию” [Nardo, 2009].

Модель может быть естественным образом преобразована для запоминания бинарных паттернов. В этом случае во входном слое одновременно активируются все нейроны с номерами, соответствующими положению 1 во входном паттерне. Отличие в функционировании такой модели будет состоять в том, что на выходе нейронов входного слоя могут быть только сигналы 0 или 1, поскольку при стационарном паттерне одни и те же номера входных нейронов не могут повторяться.

Следует отметить, что в то время как в модели хопфилдовского типа [Vogacz, Brown, 2002] емкость памяти при распознавании известности не зависит от степени разреженности входных векторов, в рассматриваемой модели она возрастает при увеличении длины входной последовательности. Однако при этом одновременно растет число модулей в сети, так что в случае неразреженной последовательности число связей достигает порядка  $n^3$ . Это существенно снижает удельную емкость памяти по сравнению с моделями хопфилдовского типа, в которых число связей имеет порядок  $n^2$ .

Если понизить порог срабатывания нейронов в промежуточном слое (в текущем варианте модели он равен  $m$ ), то можно получить распознавание известности всех подпоследовательностей меньшей длины, чем последовательности, которые использовались при обучении. Понижение порога приведет к тому, что известными окажутся также последовательности, которые не в точности совпадают с обучающими последовательностями. Таким образом, можно реализовать не точное, а ассоциативное запоминание и вспоминание.

Рассмотренная версия модели строилась из соображений минимизации числа возможных ошибок распознавания, поэтому появление ошибок распознавания имеет следующий характер. До тех пор, пока емкость памяти в значительной мере не исчерпывается, ошибок очень мало. Но после того, как число модифицированных связей между первым и вторым слоями превосходит определенный порог, число ошибок начинает быстро возрастать. Это не соответствует данным [Standing, 1973], согласно которым ошибки появляются уже при сравнительно небольшом числе запоминаемых паттернов, и их число медленно растет при увеличении числа паттернов до вероятности примерно 0,15 при числе запоминаемых паттернов порядка  $10^4$ . Этот недо-



статок модели является следствием ее “жесткости”. Из-за этого в модели возможны ошибки только одного рода — распознавание новых последовательностей как известных.

Для того чтобы сделать модель биологически более адекватной, необходимо внести стохастический компонент в процесс формирования связей между первым и вторым слоями. При обучении не все связи должны успешно формироваться и, кроме того, должны появляться случайные “ложные” связи, не обусловленные обучением. По всей видимости, ошибки распознавания в экспериментах Стэндинга являются следствием именно этих процессов, а не предельного насыщения памяти запомненными паттернами. Число нейронов в периринальной коре вполне достаточно для того, чтобы такое насыщение не возникло даже при запоминании  $10^4$  объектов.

Введение стохастики в процесс запоминания приведет к появлению ошибок как первого, так и второго рода, что позволит построить ROC-кривые для модели и сравнить их с экспериментальными ROC-кривыми. Это задача для дальнейших исследований.

Экспериментальные данные для распознавания известности нередко получаются путем предъявления испытуемому не одного примера, а пары примеров, в которой один из объектов уже предъявлялся ранее, а другой новый. Испытуемый должен определить, какой из объектов известный. Эта экспериментальная парадигма может быть реализована в результате небольшой модификации имеющейся модели. Пусть в ответ на первый пример на выходной элемент приходит сигнал  $I_1$ , а в ответ на второй пример — сигнал  $I_2$ . Если  $I_1 > I_2$ , тогда принимается решение, что первый пример является известным, если  $I_1 < I_2$ , тогда принимается решение, что второй пример является известным, если же  $I_1 = I_2$ , тогда решение принимается вероятностным способом.

Для аналитического исследования информационных характеристик нейронных сетей, в том числе с прямым распространением сигналов, разработана эффективная математическая техника [Дунин-Барковский, 1978; Фролов, Муравьев, 1987; Фролов, Муравьев, 1988; Frolov, Murav'ev, 1993]. Тем не менее мы воздержались от ее использования, так как рассматриваем изложенную модель как предварительный этап к построению биологически более адекватной модели распозна-

вания последовательностей. По этой же причине мы ограничились сравнением эффективности построенной модели лишь с известной моделью хопфилдовского типа, поскольку в работе [Bogacz, Brown, 2002] была приведена числовая характеристика удельной емкости памяти такого рода моделей при распознавании известности. Центральной задачей, на наш взгляд, является понимание, почему люди хорошо запоминают не произвольные паттерны (последовательности), а специфические паттерны, “имеющие смысл”. Например, случайные текстуры или перевернутые вверх ногами изображения будут плохо запоминаться. Музыка легче запоминается, когда она “мелодична”, а не является случайной последовательностью нот. Имеющиеся в настоящее время модели распознавания известности не способны воспроизводить эту особенность вспоминания у людей, поскольку в модель необходимо вносить трудно формализуемое семантическое осмысление запоминаемой информации. Без такой модели детальный анализ формальных информационных характеристик будет не более чем упражнениями, демонстрирующими возможности математического аппарата, но мало полезные для понимания функционирования мозга при распознавании известности. Решение этой задачи должно быть предметом будущих исследований.

Автор благодарен Т.С. Шмелинг за сообщение формулы (5).

## СПИСОК ЛИТЕРАТУРЫ

- Дунин-Барковский В.Л. Информационные процессы в нейронных структурах. М.: Наука, 1978.
- Кедров А.В., Анохин К.В. Однократное обучение условно-рефлекторному замиранию на запах у мышей: новая поведенческая модель для исследования клеточных механизмов формирования и извлечения обонятельной памяти. Журн. высш. нервн. деят. им. И.П. Павлова. 2019. 69 (4): 522–526.
- Фролов А.А., Муравьев И.П. Нейронные модели ассоциативной памяти. М.: Наука, 1987.
- Фролов А.А., Муравьев И.П. Информационные характеристики нейронных сетей. М.: Наука, 1988.
- Aggleton J.P., Vann S.D., Denby C., Dix S., Mayes A.R., Roberts N., Yonelinas A.P. Sparing of the familiarity component of recognition memory in a patient with hippocampal pathology. *Neuropsychologia*. 2005. 43 (12): 1810–1823.

- Ameen-Ali K.E., Easton A., Eacott M.J.* Moving beyond standard procedures to assess spontaneous recognition memory. *Neurosci. Biobehav. Rev.* 2015. 53: 37–51.
- Amit D.J.* Modeling brain function. Cambridge: Cambridge University Press, 1989.
- Androulidakis Z., Lulham A., Bogacz R., Brown M.W.* Computational models can replicate the capacity of human recognition memory. *Network: Comput. Neural Syst.* 2008. 19: 161–182.
- Bastin C., Besson G., Simon J., Delhaye E., Geurten M., Willems S., Salmon E.* An integrative memory model of recollection and familiarity to understand memory deficits. *Behav. Brain Sci.* 2019. 42: 1–66.
- Bogacz R., Brown M.W.* The restricted influence of the sparseness of coding on the capacity of the familiarity discrimination networks. *Network: Comput. Neural Syst.* 2002. 13: 457–485.
- Bogacz R., Brown M.W.* Comparison of computational models of familiarity discrimination in the perirhinal cortex. *Hippocampus.* 2003. 13: 494–524.
- Bogacz R., Brown M.W., Giraud-Carrier C.* Model of familiarity discrimination in the perirhinal cortex. *J. Comput. Neurosci.* 2001. 10: 5–23.
- Bowles B., Crupi C., Pigott S., Parrent A., Wiebe S., Janzen L., Köhler S.* Double dissociation of selective recollection and familiarity impairments following two different surgical treatments for temporal-lobe epilepsy. *Neuropsychologia.* 2010. 48: 2640–2647.
- Brandt K.R., Eysenck M.W., Nielsen M.K., von Oertzen T.J.* Selective lesion to the entorhinal cortex leads to an impairment in familiarity but not recollection. *Brain and Cognition.* 2016. 104: 82–92.
- Brown M.W., Xiang J.Z.* Recognition memory: Neuronal substrates of the judgement of prior occurrence. *Progr. Neurobiol.* 1998. 55: 149–189.
- Diana R.A., Reder L.M., Arndt J., Park H.* Models of recognition: A review of arguments in favor of a dual-process account. *Psychon. Bull. Rev.* 2006. 13: 1–21.
- Eichenbaum H.* Time cells in the hippocampus: a new dimension for mapping memories. *Nat. Rev. Neurosci.* 2014. 15: 732–744.
- Eichenbaum H., Yonelinas A.R., Ranganath C.* The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci.* 2007. 30: 123–152.
- Feinberg L.M., Allen T.A., Ly D., Fortin N.J.* Recognition memory for social and non-social odors: Differential effects of neurotoxic lesions to the hippocampus and perirhinal cortex. *Neurobiology of Learning and Memory.* 2012. 97: 7–16.
- Frolov A.A., Murav'ev I.P.* Information characteristics of neural networks capable of associative learning based on Hebbian plasticity. *Network: Comput. Neural Syst.* 1993. 4: 495–536.
- Hopfield J.J.* Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U. S. A.* 1982. 79: 2554–2558.
- Kafkas A., Montaldi D.* How do memory systems detect and respond to novelty? *Neurosci. Lett.* 2018. 680: 60–68.
- Medina J.J.* The biology of recognition memory. *Psychiatric Times.* 2008. 13: 13–16.
- Merkow M.W., Burke J.F., Kahanac M.J.* The human hippocampus contributes to both the recollection and familiarity components of recognition memory. *Proc. Natl. Acad. Sci. U. S. A.* 2015. 112: 14378–14383.
- Miller G.A.* The magical number seven, plus or minus two. Some limits on our capacity for processing information. *Psychol. Rev.* 1956. 63: 81–97.
- Montaldi D., Mayers A.R.* The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus.* 2010. 20: 1291–1314.
- Nardo D., Brunetti R., Cupellini E., Belardinelli M.O.* The influence of melodic and rhythmic redundancies on recognition memory for unknown musical themes. *Musicæ Scientiæ.* 2009. 13: 337–355.
- Rutishauser U., Schuman E.M., Mamelak A.N.* Activity of human hippocampal and amygdala neurons during retrieval of declarative memories. *Proc. Natl. Acad. Sci. U. S. A.* 2008. 105: 329–334.
- Scalici F., Caltagirone C., Carlesimo G.A.* The contribution of different prefrontal cortex regions to recollection and familiarity: a review of fMRI data. *Neurosci. Biobehav. Rev.* 2017. 83: 240–251.
- Slotnick S.D.* The nature of recollection in behavior and the brain. *NeuroReport.* 2013. 24: 663–670.
- Squire L.R., Zola-Morgan J., Clark R.E.* Recognition memory and the medial temporal lobe: a new perspective. *Nature Rev. Neurosci.* 2007. 8: 872–883.
- Standing L.* Perception and memory for pictures: Single trial learning of 2500 visual stimuli. *Psychol. Sci.* 1970. 19: 73–74.
- Standing L.* Learning 10 000 pictures. *Quat. J. Exp. Psychol.* 1973. 25: 207–222.
- Vinogradova O.S.* Hippocampus as comparator: Role of two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus.* 2001. 11: 578–598.
- Wais P.E., Wixted J.T., Hopkins R.O., Squire L.R.* The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron.* 2006. 49: 459–466.
- Wixted J.T., Squire L.R.* The medial temporal lobe and the attributes of memory. *Trends Cogn. Sci.* 2011. 15: 210–217.
- Yonelinas A.P.* The nature of recollection and familiarity: A review of 30 years of research. *J. Memory and Language.* 2002. 46: 441–517.

## NEURAL NETWORK MODEL OF FAMILIARITY RECOGNITION FOR NUMERICAL SEQUENCES

**Y. B. Kazanovich<sup>#</sup>**

*Institute of Mathematical Problems of Biology, the Branch of M.V. Keldysh  
Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Russia*

<sup>#</sup>*e-mail: yasha@impb.ru*

Familiarity recognition is a component of recognition memory that becomes apparent when a subject remembers the fact that an object or event were met by him earlier, but he may not remember object/event details and also the context when this object/event took place. The experiments of Standing (Standing, 1970; Standing, 1973) have shown extraordinary capacity of familiarity recognition memory when it has been tested by natural images, words or musical melodies. Existing neural network models of familiarity recognition demonstrated that the corresponding memory capacity can be of the order  $n^2$ , where  $n$  is the number of neurons in the model. Here a new model of familiarity recognition is suggested that is specialized on temporal sequences (in particular, sequences of integers) represented by sparse input patterns. By computer simulations it is shown that under special conditions of error-free recognition the model capacity is higher than in traditional Hopfield-type models.

*Keywords:* recognition memory, familiarity recognition, neural network models, memory capacity