

УДК 519.688+612.76

**ДИНАМИЧЕСКИЕ ХАРАКТЕРИСТИКИ НЕЙРОСЕТЕВОЙ МОДЕЛИ  
ПРОСТРАНСТВЕННОЙ ПАМЯТИ**

В.А. Ляховецкий, А.С. Потапов

Рассмотрена модель запоминания последовательности движений (пространственных позиций), основанная на гетероассоциативной нейронной сети. Предложен критерий оценки правильности запомненного материала как числа итераций, необходимого для перехода в устойчивое состояние. Этот критерий хорошо согласуется с психофизиологическими данными и может использоваться для обучения нейросетевой модели без учителя. Предложена интерпретация разных типов ошибок с позиций нелинейной динамики.

**Ключевые слова:** гетероассоциативные нейронные сети, рабочая память, повторные ошибки.

**Введение**

Основной функцией памяти является извлечение информации по содержанию, в чем технологии искусственного интеллекта до сих пор существенно проигрывают человеку. В этой связи актуальным представляется сопоставление характеристик компьютерных моделей и памяти человека. В настоящей работе это сопоставление выполняется на примере темпоральной рабочей памяти, которая используется для запоминания и последующего воспроизведения последовательности движений или пространственных положений.

Существуют три класса теорий, рассматривающих кодирование упорядоченной во времени информации [1]. Позиционные теории предполагают, что каждый элемент последовательности хранится в стеке рабочей памяти независимо от других элементов, а воспроизведение элементов в нужном порядке обеспечивается обходом стека предопределенным образом [2]. Популярными в настоящее время теории «соревновательного сцепления» (competitive queuing) являются развитием позиционных теорий. Они также предполагают, что элементы последовательности хранятся в рабочей памяти независимо друг от друга. Порядок воспроизведения элементов определяется «активностью» каждого элемента, определяющейся на протяжении запоминания всей последовательности. Выходные сигналы элементов последовательности конкурируют друг с другом, и только самый «активный» из них может «выиграть» в определенный момент времени. После того как элемент воспроизведен, осуществляется торможение его выходного сигнала, препятствующее его повторному воспроизведению [3]. Теории «последовательного связывания» (chaining) предполагают, что последующий элемент последовательности может быть воспроизведен лишь на основе информации о предыдущем элементе [1]. В настоящее время модели последовательного связывания реализуются на базе искусственных нейронных сетей (ИНС) типа ассоциативной памяти (автоассоциативные сети Хопфилда, гетероассоциативные сети Коско) [4, 5]. Нейросетевые модели позволяют объяснить как электрофизиологические данные, так и когнитивные аспекты рабочей памяти [4].

В нейросетевых моделях элементы последовательности хранятся распределенно, определяя матрицу весов связей нейронов. При использовании простых алгоритмов расчета матрицы весов емкость сетей достаточно мала, и они «самостоятельно», без вмешательства экспериментатора в работу сети, допускают ошибки при воспроизведении запомненного [6]. Это свойство сетей позволяет без дополнительных допущений объяснить малую емкость рабочей памяти, в отличие от моделей, основанных на теориях соревновательного сцепления, в которых необходимы дополнительные допущения для объяснения ошибок запоминания.

Одной из интересных проблем разработки моделей памяти является введение критерия оценки верности запомненного материала без обращения к запомненным объектам внешнего мира. Значение этого критерия характеризует качество запоминания и может позволить спланировать стратегию обучения модели. Проблема оценки верности запомненного материала – одна из важных проблем также в когнитивной психологии. Некоторым образом человеку удается понять, какой материал запомнен им правильно, а какой – ошибочно. При этом отличить правильный ответ от неверного можно и неосознанно [7]. Часто правильность ответа влияет на время, требуемое на воспроизведение запомненного, – средние времена правильных и неверных ответов испытуемого различны. При воспроизведении последовательности движений руки среднее время правильно воспроизведенных движений ниже среднего времени ошибочно воспроизведенных движений [8].

Нейросетевая модель [5], основанная на теории последовательного связывания, хорошо подходит для описания распределений ошибок при немедленном воспроизведении запомненных последовательностей движений руки, в частности, воспроизводит эффект новизны, наблюдаемый в психофизиологическом эксперименте. В настоящей работе предпринята попытка, оставаясь в рамках внутреннего представления данных, используемого в модели [5], предложить объективный критерий, позволяющий отличить верные ответы от неверных, не обращая повторно к входной последовательности стимулов.

**Методика**

Предложенная нейросетевая модель используется для описания процессов запоминания и воспроизведения человеком последовательности движений руки по плоскости, поэтому целесообразно начать изложение методики с описания психофизического эксперимента.

С помощью сенсорного экрана регистрировались координаты положений указательного пальца руки (т.е. «рабочей точки» руки) испытуемых при немедленном воспроизведении запомненной последовательности. Для регистрации использовался емкостной сенсорный экран 3M Touch Systems 18.90", подключенный с помощью NovRAM кабеля через микроконтроллер к USB-порту персонального компьютера. Специальное программное обеспечение, разработанное авторами, позволяло регистрировать параметры касаний сенсорного экрана. Опыты проводились без обратной связи – до окончания эксперимента испытуемому не сообщалось о правильности или ошибочности его ответов. Локоть испытуемого не касался поверхности стола и не был фиксирован. В экспериментах принимали участие правши, определение доминирующей руки проводилось по специальному тесту [9].

Изучалось воспроизведение последовательностей движений в двух группах правшей (по 15 человек в каждой): в группе А испытуемые начинали работать правой рукой, а продолжали левой, в группе Б – наоборот. Анализировались ошибки испытуемых при многократном (3 раза) запоминании и воспроизведении последовательности движений правой или левой руки по 6 полям, расположенным на листе формата А4, разбитом на 6×4 квадратов.

Анализ данных проводили, исходя из гипотезы о двух способах кодирования информации в кратковременной моторной памяти, согласно которой при запоминании последовательности кодируется положение стимулов (позиционное кодирование конечных точек движений) и (или) изменение их положения (кодирование векторов перемещений руки) [5]. Для двух последовательных серий движений выделялись одинаковые повторные ошибки по положению (испытуемый дважды неправильно воспроизводил положение руки с расхождением не более чем на 2 см) и по направлению движения (испытуемый дважды неправильно воспроизводил направление движения руки с расхождением не более чем на 15°). Регистрировалось среднее время движений испытуемых при правильных и ошибочных ответах.

**Математическая модель.** Модель типа двунаправленной гетероассоциативной памяти содержит два слоя формальных нейронов, соединенных двунаправленными связями [5] (рисунок).

На входной слой модели подается вектор  $Z_i$ . Нейроны этого класса сетей работают с бинаризованными данными, поэтому элементы векторов  $Z$  переводятся в двоичную систему счисления [10]. Согласно вышеупомянутой гипотезе о двух возможных типах кодирования в моторной памяти, элементы векторов запоминаемой последовательности  $Z_i$  содержат информацию о рабочих точках руки  $S_i$ , которые можно определить в полярной или декартовой системах координат.

**Полярная система координат.** Элементами векторов являются направление  $\alpha$  и величина перемещения  $\rho$ . Направление  $\alpha$  хранится дискретно, с шагом 11,5°, максимальная величина перемещения составляет 6 квадратов (из одного угла листа в другой по диагонали), поэтому для хранения  $\alpha$  и  $\rho$  в бинаризованной форме требуется 5 и 3 элемента соответственно. Параметры  $\alpha$  и  $\rho$  хранятся в рамках двух подсетей, между которыми отсутствуют связи, что соответствует имеющимся в литературе сведениям о раздельном запоминании человеком направления и амплитуды движения. Для расчета симметричной матрицы весов  $W_p$ , используемой при запоминании последовательности величин перемещений, применялся алгоритм расчета, предложенный в [10]. Для расчета асимметричной матрицы весов  $W_a$ , используемой при запоминании последовательности направлений перемещений, применялся алгоритм расчета, предложенный в [6]. Схема кодирования смещений рабочей точки в полярной системе координат может использоваться при векторном способе внутреннего представления информации, т.е. при кодировании траектории движения руки.

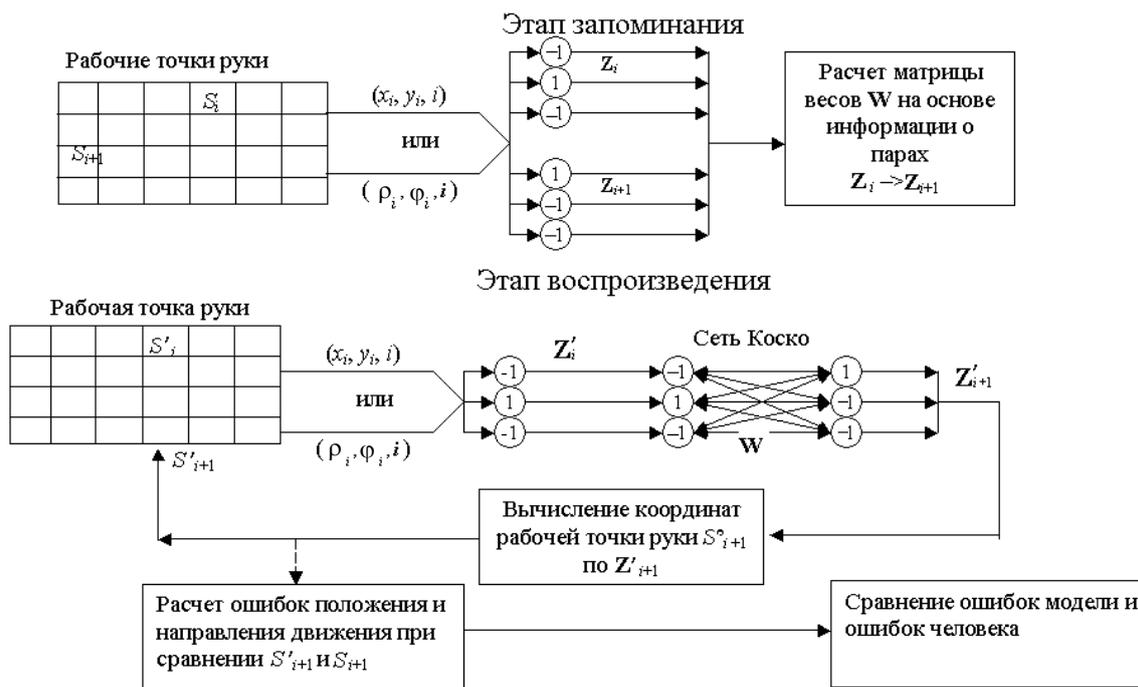


Рисунок. Схема модели

**Декартова система координат.** Модель предназначена для имитации запоминания целей, расположенных на листе, разбитом на  $6 \times 4$  квадрата, поэтому для хранения координат в бинаризованной форме требуется 3 и 2 элемента соответственно. Такая схема кодирования может использоваться при позиционном способе внутреннего представления информации, т.е. при кодировании положения руки.

Для обеих рассмотренных выше схем кодирования вектор  $\mathbf{Z}$  содержит номер цели, поэтому модель способна запоминать последовательности, содержащие повторяющиеся элементы.

Сеть характеризуется вычисляемой в явной форме матрицей весов  $\mathbf{W}$  и не нуждается в тренировке. Этап запоминания заключается в расчете матрицы  $\mathbf{W}$  по «цепочке» пар  $(\mathbf{Z}_{i-1}, \mathbf{Z}_i)$  векторов запоминаемой последовательности:  $\mathbf{W} = \sum_i \mathbf{Z}_{i-1}^t \mathbf{Z}_i$ . Модель предназначена для имитации запоминания шести движений руки, поэтому  $i = 1-6$ .

На этапе воспроизведения элементы  $z_{ik}$  вектора  $\mathbf{Z}_i$  вычисляются через элементы  $z_{(i-1)m}$  вектора  $\mathbf{Z}_{i-1}$  и элементы  $w_{mk}$  матрицы  $\mathbf{W}$ :

$$z_{ik}^0 = \text{sgn} \left( \sum_m z_{(i-1)m} w_{mk} \right),$$

где  $k, m = 1 \dots K$ ,  $K$  – число нейронов одного слоя сети, которое зависит от выбранной схемы кодирования,

$$\text{sgn}(a) = \begin{cases} 1, & \text{если } a \geq 0, \\ -1, & \text{иначе.} \end{cases}$$

Далее активность распространяется на первый слой сети,  $z_{(i-1)k} = \text{sgn} \left( \sum_m z_{im}^0 w_{mk} \right)$ , и затем вновь на второй слой,  $z_{ik}^1 = \text{sgn} \left( \sum_m x_m w_{mk} \right)$ . Двустороннее распространение активности продолжается до тех пор, пока не будет достигнуто устойчивое состояние,  $z_{im}^t = z_{im}^{t+1}$ .

При каждой схеме кодирования с моделями проводили 10000 опытов с нейронной сетью. Регистрировалось количество итераций, за которые сеть достигнет устойчивого состояния для верно и ошибочно запомненных векторов.

### Результаты и их обсуждение

Зависимость числа итераций, за которое нейронная сеть достигнет устойчивого состояния, от числа ошибочно запомненных битов представлена в табл. 1. Из таблицы видно, что для обеих схем кодирования сеть достоверно быстрее приходит в устойчивое состояние при правильных ответах (0 ошибочных бит). Это означает, что, хотя информация о числе ошибочных битов не доступна сети, для оценки пра-

вильности ответа модель может использовать количество итераций, необходимых для перехода сети в устойчивое состояние.

Ошибочных битов	0	1	2	3	4	5	6	7	8
<b>Схема 1</b>	1,3	1,6	1,6	1,7	1,8	1,7	1,7	1,7	1,8
<b>Схема 2</b>	1,4	1,6	1,7	1,8	1,8	1,7	–	–	–

Таблица 1. Число итераций сети для различного числа ошибочно запомненных битов

Можно предположить, что, так как разница в среднем числе итераций в зависимости от правильности ответа невелика (0,3 итерации для Схемы 1; 0,2 итерации для Схемы 2), то, основываясь на таком критерии, модель будет иногда принимать ошибочный ответ за верный. Это означает, что введенный в рассмотрение критерий правильности ответа будет «слабым». Однако такое свойство модели не будет недостатком – и человек, как показано в приведенном психофизическом опыте, не всегда способен отличить свои правильные ответы от ошибочных.

При запоминании движений руки испытуемые достаточно часто допускают повторные ошибки, т.е. при двукратном воспроизведении запомненной серии движений дважды повторяют свою ошибку одинаковым образом (дважды подряд касаются окрестности одной и той же неверной точки, дважды подряд выбирают одно и то же неверное направление). Среднее время правильных и ошибочных ответов испытуемых представлено в табл. 2. Среднее время движений, связанных с повторными ошибками, значимо (по критерию знаков) меньше среднего времени движений, связанных с иными ошибками, и не отличается от среднего времени правильно запомненных движений. Причем эта закономерность проявляется уже для первого ошибочного движения.

Группа	Рука	Положение			Направление		
		Правильные ответы	Ошибки		Правильные ответы	Ошибки	
			Повторные	Иные		Повторные	Иные
<b>А</b>	Правая	1649	1447	2139	1667	1624	2004
	Левая	1602	1671	2038	1789	1809	2029
<b>Б</b>	Левая	1583	1493	1939	1592	1723	1707
	Правая	1592	1412	1966	1518	1549	1828

Таблица 2. Среднее время движений при правильных и ошибочных ответах испытуемых, мс

По-видимому, испытуемые уверены в правильности таких ошибочных движений и не пытаются корректировать их при повторном заучивании последовательности движений руки, т.е. не могут отличить такие неверные ответы от правильных – число повторных ошибок не уменьшается от первой серии заучивания к третьей.

Нейросетевые модели могут допускать как случайные, так и повторные ошибки. Эти ошибки хорошо объяснимы с позиций нелинейной динамики ИНС: поведение подобных рекуррентных ИНС можно описать через траектории в фазовом пространстве, задаваемом активностями нейронов сети. Сходимость ассоциативной памяти к некоторому устойчивому состоянию означает наличие точечного аттрактора в фазовом пространстве. Задача запоминания некоторых образов такой сетью сводится к выбору таких весов связей, которые бы в качестве аттракторов определяли запомненные образы.

В нашем исследовании затрачиваемое время при совершении повторных ошибок, совершаемых человеком, близко ко времени, затрачиваемому при правильных ответах. Такое соотношение времен воспроизводится нейросетевой моделью, для которой повторные ошибки обусловлены «ложными» точечными аттракторами, не соответствующими запоминаемым образам и возникающими как побочный эффект при обучении.

При совершении случайных ошибок человек затрачивает больше времени на принятие решения. Эта особенность также воспроизводится нейросетевой моделью: при возникновении в фазовом пространстве ИНС неточечных аттракторов сеть (при попадании в такой аттрактор) будет колебаться между разными ответами, выбор между которыми (при превышении максимального числа итераций) даст случайную ошибку.

### Заключение

Критерий оценки правильности запомненного материала нейросетевой моделью как числа итераций, необходимого для перехода искусственных нейронных сетей в устойчивое состояние, предложенный в данной работе, хорошо согласуется с психофизиологическими данными и может использоваться для обучения нейросетевой модели без учителя.

Установлено наличие сходства между процессами воспроизведения запомненных образов человеком и искусственными нейронными сетями не только в случае случайных, но и в случае повторных ошибок. В дальнейшем планируется расширить интерпретацию указанных особенностей функционирования памяти с позиции нелинейной динамики не только для установления сходства при совершении ошибок, но и для воспроизведения полезных свойств памяти человека в компьютерных системах. К таким свойствам относится обобщение образов и их извлечение по отдельным фрагментам. В случае робототехнических систем подобные свойства могут быть полезны при реализации адаптивных программ управления движением, формируемых на основе примеров, а также для реализации топографической памяти.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации.

### Литература

1. Henson R. Serial order in short-term memory // *The Psychologist*. – 2001. – V. 14. – P. 71–74.
2. Conrad R. Order error in immediate recall of sequences // *J. of Verbal Learning and Verbal Behavior*. – 1965. – V. 6. – P. 49–54.
3. Grossberg S. Behavioral contrast in short-term memory: serial binary memory models or parallel continuous memory models? // *J. of Mathematical Psychology*. – 1978. – V. 17. – P. 199–219.
4. Durstewitz D., Seamans J.K., Sejnowski T. Neurocomputational models of working memory // *Nature Neuroscience Supplement*. – 2000. – V. 3. – P. 1184–1191.
5. Ляховецкий В.А., Боброва Е.В. Воспроизведение запомненной последовательности движений правой и левой руки: позиционное и векторное кодирование // *Журнал высшей нервной деятельности*. – 2009. – № 1. – С. 45–54.
6. Xu Z.-B., Leung Y., He X.-W. Asymmetric bidirectional associative memories // *IEEE Trans. Systems, Man and Cybernetics*. – 1994. – V. 24. – P. 1559–1564.
7. Аллахвердов В.М. Методологическое путешествие по океану бессознательного к таинственному острову сознания. – СПб: Речь, 2003. – 368 с.
8. Rabbit P. An elementary preliminary taxonomy for some errors in laboratory choice RT-tasks // *Acta Psychologica*. – 1970. – V. 33. – P. 56–76.
9. Oldfield R.C. The assessment and analysis of handedness: the Edinburg inventory // *Neuropsychologia*. – 1971. – V. 9. – P. 97–113.
10. Kosko B. Bidirectional associative memories // *IEEE Trans. Systems, Man and Cybernetics*. – 1988. – V. 18. – P. 49–60.

- Ляховецкий Всеволод Александрович* – Институт Физиологии им. И.П. Павлова РАН, кандидат технических наук, младший научный сотрудник, v\_la2002@mail.ru
- Потапов Алексей Сергеевич* – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, доктор технических наук, профессор, pas.aicv@gmail.com