

УДК 004.93:007.52

А.В. Волик, С.Н. Герасин, Е.В. Слипченко

ИССЛЕДОВАНИЕ ДИНАМИКИ СРЕДНЕГО УРОВНЯ И ДИСПЕРСИИ АКТИВНОСТИ НЕЙРОННОЙ СЕТИ С ЛОКАЛЬНЫМИ СВЯЗЯМИ

1. Введение

Синтез математических методов и подходов, а именно: математического моделирования, теории информации, формальных нейронных сетей и нейробиологии породил новое научное направление — вычислительные нейронауки, общей идеей которого является тезис «изучение через моделирование» [1].

При этом в имитационных экспериментах на моделях, аккумулирующих тот или иной набор известных параметров и свойств реальных нейронов, воспроизводятся некоторые особенности их активности, анализируются возможные механизмы и формулируются предположения, доступные экспериментальной проверке. Одним из направлений исследований в данной области является изучение организации нейронных структур как многомерных рекуррентных сетей и определение влияния свойств отдельных клеток на динамику активности и свойства всей сети. В работе [2] описан критический режим функционирования нейронной сети, состоящей из вероятностных возбуждающих элементов с локальными связями, получаемый с помощью имитационного моделирования на ЭВМ. Суть этого режима состоит в том, что для некоторого диапазона изменения параметров взаимодействия элементов сети, начальная конфигурация «горящих» элементов сохраняется длительное время, на несколько порядков превышающее время релаксации отдельного элемента. В работах [3, 4] обсуждаются вопросы применимости полученного режима для моделирования гиппокампа и объяснения кратковременной памяти.

Цель работы — нахождение и исследование динамических характеристик стохастической нейронной сети, по которым можно оценить поведение сети в целом.

Задачи исследования состоят в выводе и качественном анализе уравнений для средней активности и дисперсии активности нейронной сети и выявления возможных критических режимов.

2. Уравнения динамики макрохарактеристик сети с локальными связями

На уровне макроподхода к моделированию поведения нейронной сети, критический режим характеризуется наличием метастабильного состояния среднего уровня активности (из одних и тех же начальных условий различные траектории эволюционируют либо к верхнему уровню, либо к нижне-

му уровню стационарной активности, но средний, по множеству траекторий, уровень активности остается постоянным) и неограниченным ростом дисперсии [2, 4]. Для моделирования критического поведения нейронной сети с локальными связями рассмотрим две макрохарактеристики: $E(t)$ — средний уровень активности сети и $z(t)$ — дисперсия среднего уровня активности, которые удовлетворяют системе уравнений:

$$\begin{cases} \frac{dE}{dt} = -E + S(aE) + \frac{S''(aE)}{2} a^2 z, \\ \frac{dz}{dt} = -2z(1 - aS'(aE)). \end{cases} \quad (1)$$

В работе [3] рассматривалась нейронная сеть с локальными связями. Для этой сети из обратного уравнения Колмогорова была получена следующая система

$$\begin{cases} \frac{dq_i(t)}{dt} = -q_i(t) + M\{c_i(y)\}, \\ \frac{dR_{ij}^0(t)}{dt} = -2R_{ij}^0 + M\{y_j c_i(y)\} + M\{y_i c_j(y)\}, \end{cases} \quad (2)$$

где $y(t) = (y_1, y_2, \dots, y_N)$ — вектор, описывающий поведение элементов сети; $q_i(t) = M\{y_i(t)\}$ — среднее значение переменной, характеризующей поведение i -го элемента; $R_{ij}^0(t) = M\{y_i(t)y_j(t)\}$ — среднее от произведения характеристик i -го и j -го элементов.

Пусть T_{ij} — ковариация характеристик i -го и j -го элементов

$$T_{ij} = M\{(y_i - q_i)(y_j - q_j)\} = R_{ij} - q_i q_j.$$

Отсюда

$$R_{ij} = T_{ij} + q_i q_j.$$

Дифференцируя по t и учитывая (1), получим уравнение для T_{ij}

$$\begin{aligned} \frac{dT_{ij}}{dt} = & -2T_{ij} - q_i M\{c_j(y)\} - q_j M\{c_i(y)\} + \\ & + M\{y_i c_j(y) + y_j c_i(y)\}. \end{aligned} \quad (3)$$

Предположим, что $c_i(y)$ имеет следующий вид:

$$c_i(y) = S_E\left(\sum_j a_{ij} y_j\right),$$

где S_E — сигмоидная функция.

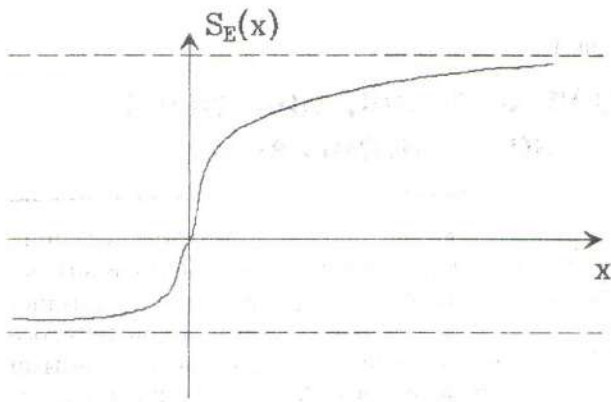


Рис. 1. График активационной функции $S_E(x)$

Обозначим

$$\xi_i = \sum_j a_{ij} y_j, \quad V_i = M(\xi_i),$$

тогда $h = \xi_i - V_i$ — отклонение случайной величины ξ_i от среднего значения.

Разложим функцию C_i в ряд Тейлора по h :

$$c_i(h) = S(V_i + h) = S(V_i) + S'(V_i)h + \frac{S''(V_i)}{2}h^2 + \dots$$

Здесь будем предполагать, что

$$\|h\| = \sqrt{\int (\xi_i - V_i)^2 dP_i}$$

мало. Найдем математическое ожидание $c_i(y)$:

$$M\{c_i(y)\} = S(V_i) + \frac{S''(V_i)}{2}M\{(\xi_i - V_i)^2\} + \dots$$

Вычислим $M\{(\xi_i - V_i)^2\}$:

$$\begin{aligned} M\{(\xi_i - V_i)^2\} &= M\left\{\left[\sum_j a_{ij}(y_j - q_j)\right]^2\right\} \\ &= M\left\{\sum_j \sum_k a_{ij} a_{ik} (y_j - q_j)(y_k - q_k)\right\} \\ &= \sum_j \sum_k a_{ij} a_{ik} T_{jk}. \end{aligned}$$

Тогда

$$M\{c_i(y)\} = S\left(\sum_j a_{ij} q_j\right) + \frac{S''(V_i)}{2} \sum_j \sum_k a_{ij} a_{ik} T_{jk} + \dots$$

Подставляя полученное равенство в первое уравнение (1) и отбрасывая члены выше второго порядка, получим уравнение для $q_i(t)$:

$$\frac{dq_i}{dt} = -q_i + S\left(\sum_l a_{il} q_l\right) + \frac{S''\left(\sum_l a_{il} q_l\right)}{2} \sum_k \sum_l a_{ik} a_{il} T_{lk}.$$

Чтобы получить уравнение для T_{ij} , вычислим $M\{y_j c_i\} - q_j M\{c_i\}$.

$$\begin{aligned} M\{y_j c_i\} - q_j M\{c_i\} &= M\{(y_j - q_j) c_i(y)\} = \\ &= M\{(y_j - q_j)(S(V_i) + S'(V_i)h + \dots)\} = \\ &= S'(V_i) M\left\{(y_j - q_j) \left(\sum_k a_{ik} (y_k - q_k)\right)\right\} = \\ &= S'(V_i) \sum_k a_{ik} T_{jk} + \dots \end{aligned}$$

Подставляя полученное выражение в уравнение (3) и отбрасывая члены выше первого порядка, получим систему обыкновенных дифференциальных уравнений для описания динамики активности и дисперсии активности каждого элемента сети, а также взаимные ковариации активностей элементов

$$\begin{cases} \frac{dq_i}{dt} = q_i + S\left(\sum_l a_{il} q_l\right) + \frac{S''\left(\sum_l a_{il} q_l\right)}{2} \sum_{k,l} a_{il} a_{ik} T_{lk}, \\ \frac{dT_{ij}}{dt} - 2T_{ij} + S'\left(\sum_l a_{il} q_l\right) \sum_k a_{ik} T_{jk} + S'\left(\sum_l a_{jl} q_l\right) \sum_k a_{jk} T_{ik}. \end{cases} \quad (4)$$

Чтобы получить уравнения для макрохарактеристик сети, предположим, что средняя активность E и дисперсия активности Z всей сети подчиняются тем же уравнениям, что и характеристики некоторого типичного среднего нейрона, поэтому в системе (1) положим $N=1$ и обозначим

$$E(t) = q_1; \quad Z(t) = T_{11}; \quad a = a_{11},$$

тогда

$$\begin{cases} \frac{dE}{dt} = -E + S(aE) + \frac{S''(aE)}{2} a^2 Z, \\ \frac{dZ}{dt} = -2Z(1 - aS'(aE)). \end{cases} \quad (5)$$

Аналогичный подход к получению уравнений для макрохарактеристик сети применялся в работе [5]. Для описания активности совокупности возбуждающих и тормозных нейронов там использовалась система дифференциальных уравнений типа системы Ходжкина-Хаксли, полагая, что она описывает динамику активности некоторого типичного нейрона, представляющего средние характеристики совокупности. При этом основное внимание было уделено переходам сети в периодический режим и роли пейсмекерных нейронов в этом процессе.

3. Качественное исследование системы уравнений для макрохарактеристик. Сравнение с имитационной моделью

Проведем качественное исследование системы (5) и сравним с результатами имитационного моделирования [6]. Предположим, что функция $S(x)$ имеет вид

$$S(x) = \frac{1}{1 + \exp[-b(x - \theta)]} - \frac{1}{1 - \exp[b\theta]}$$

и обладает следующими свойствами:

а) $S(0) = 0$;

б) $S(\infty) = \frac{\exp[b\theta]}{1 + \exp[b\theta]}$;

в) $S'(x) = \frac{b \exp[-bx + b\theta]}{(1 + \exp[-bx + b\theta])^2}$;

г) $S''(x) = \frac{b^2 \exp[-bx + b\theta](\exp[-bx + b\theta] - 1)}{(1 + \exp[-bx + b\theta])^3}$;

д) $\max_x S'(ax) = \frac{b}{4}$ при $x = \frac{\theta}{a}$.

Функция $S(x)$ характеризует отклик нейронной сети на поступившее воздействие $x(t)$ и зависит от двух параметров b и θ , причем параметр θ пропорционален величине эффективного порога сети, а параметр b характеризует степень распределенности порогов у элементов сети.

Заметим, что все бифуркации, возможные в системе (5), зависят от параметра a , характеризующего величину связи между элементами. Изменение параметров b и θ не влияет на качественный характер происходящих перестроек. Поэтому в качестве изменяемого параметра мы будем рассматривать параметр a , считая, что b и θ фиксированы.

Найдем стационарные состояния системы (5), для этого рассмотрим систему нелинейных уравнений:

$$\begin{cases} -E + S(aE) + \frac{S''(aE)}{2} a^2 Z = 0, \\ 2Z(1 - aS'(aE)) = 0. \end{cases}$$

Эта система эквивалентна двум системам уравнений, дающим две серии решений:

$$\begin{cases} Z = 0, \\ S(aE). \end{cases} \quad (6)$$

$$\begin{cases} aS'(aE) = 1, \\ Z = 2 \frac{E - S(aE)}{a^2 S''(aE)}. \end{cases} \quad (7)$$

При анализе систем удобна геометрическая интерпретация. Решениями второго уравнения системы (6) являются точки пересечения графиков $f(E) = S(aE)$ и $f(E) = E$. Решениями первого уравнения системы (7) являются такие значения E , при которых касательные к графику функции $f(E) = S(aE)$ параллельны прямой $f(E) = E$.

Для определения устойчивости стационарных состояний рассмотрим матрицу линеаризации системы (5):

$$A = \begin{pmatrix} -1 + aS'(aE) + \frac{S''(aE)}{2} a^2 Z & \frac{S''(aE)}{2} a^2 \\ 0 & 2(-1 + aS'(aE)) \end{pmatrix}$$

Для стационарных решений, определяемых системой (2), т. е. решений вида $(E_0, 0)$, матрица принимает вид:

$$A = \begin{pmatrix} -1 + aS'(aE_0) & \frac{S''(aE_0)}{2} a^2 \\ 0 & 2(-1 + aS'(aE_0)) \end{pmatrix}$$

Отсюда находим собственные значения

$$\lambda_1 = -1 + S'(aE_0), \quad \lambda_2 = 2\lambda_1.$$

Для стационарных решений, определяемых системой (6), матрица A принимает вид:

$$A = \begin{pmatrix} S'''(aE)a(E - S(aE)) & \frac{S''(aE)}{2} a^2 \\ 4(E - S(aE)) & 0 \end{pmatrix}$$

Определитель этой матрицы Δ равен произведению собственных значений и позволяет судить об устойчивости стационарных состояний.

$$\Delta = -2(E - S(aE))a^2 S''(aE).$$

Диапазон изменения параметра a разбивается на четыре участка (b и θ фиксированы), в соответствии с этим рассмотрим четыре случая, для каждого из которых дадим описание фазового портрета.

Случай 1. $0 < a < \alpha_1$. В этом случае существует единственное стационарное состояние (устойчивое) $E=0, Z=0$, к которому эволюционируют все траектории системы. При $a = \alpha_1$ происходит бифуркация и из бесконечности рождаются два стационарных состояния: E_Λ, Z_Λ и E_Π, Z_Π , причем $E_\Lambda \xrightarrow{\alpha \uparrow \alpha_1} \infty$, а $Z_\Pi \xrightarrow{\alpha \uparrow \alpha_1} -\infty$. (В момент бифуркации

касательная к графику $f(E) = S(aE)$ становится параллельной прямой $f = E$.

Случай 2. $\alpha_1 < \alpha < \alpha_2$. В этом случае в полуплоскости $Z > 0$ существует неустойчивое (седловое) стационарное решение. При $\alpha = \alpha_2$ происходит сложная бифуркация, в результате которой рождаются три стационарных состояния: на оси $Z=0$ устойчивое и неустойчивое состояния и в полуплоскости $Z > 0$ седловой стационар. (В момент бифуркации происходит касание графиков функций $f(E) = S(\alpha E)$ и $f(E) = E$).

Случай 3. $\alpha_2 < \alpha < \alpha_3$. В этом случае в полуплоскости $Z \geq 0$ существуют пять стационарных состояний: два устойчивых и одно неустойчивое на оси $Z=0$ и два седловых при $Z > 0$. При $\alpha = \alpha_3$ происходит бифуркация слияния седла (E_A, Z_A) с устойчивым узлом $(E=0, Z=0)$ и неустойчивым узлом на оси E ; при увеличении a сложная точка опять распадается на три стационарных решения. (В момент бифуркации происходит касание графиков функций $f(E) = S(\alpha E)$ и $f(E) = E$ в точке $E=0$).

Случай 4. $a > a_3$. В этом случае опять существуют 5 стационарных решений и фазовый портрет аналогичен случаю 3, но теперь одно устойчивое и одно седловое стационарные состояния переместились в квадрант $E < 0, z \geq 0$.

Перейдем теперь к сравнению результатов качественного исследования и имитационного моделирования [6]. Покажем, как можно объяснить возникновение критического режима, наблюдаемого на имитационной модели.

Предположим, что наша система всегда начинает эволюционировать из одних и тех же начальных условий $(E(0), Z(0))$. Проследим изменение финальных состояний траектории, выходящей из одной точки, при увеличении параметра связи a . При малых значениях параметра a траектория будет стремиться к состоянию фоновой активности $(E(0), Z(0))$. При увеличении параметра a неустойчивый стационар и оба седла будут сдвигаться влево и при $\alpha \approx \alpha_{кр}^1$ сепаратриса седла (E_d, Z_d) пройдет вблизи зафиксированной нами начальной точки $(E(0), Z(0))$. При этой величине связи система будет длительное время эволюционировать к своему стационарному состоянию, поскольку траектория системы пройдет вблизи седла и надолго задержится в его окрестности. При $\alpha_{кр}^1 < \alpha < \alpha_{кр}^2$ траектория, идущая из начальной точки, вообще не попадает в стационар $(E=0, Z=0)$, т. к. попадает в область между сепаратрисами седел и переменная Z будет неограниченно возрастать, а E будет принимать некоторое промежуточное значение. При $\alpha > \alpha_{кр}^2$ траектория, идущая из начальной точки, будет стремиться к стационарному состоянию $(E_0, 0)$, соответствующему ВУ активности.

На рис. 2 и 3 приведены графики зависимостей $E(t)$ и $Z(t)$ при различных значениях параметра α .

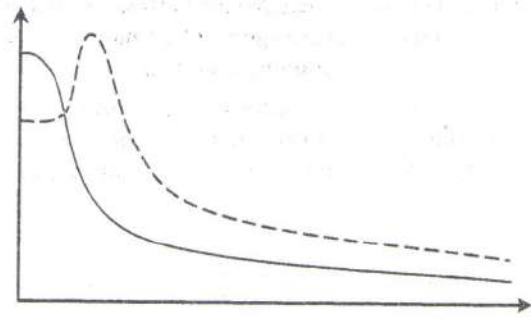


Рис. 2. График зависимости дисперсии и среднего уровня активности при $\alpha < \alpha_{кр}^1$ (----- $E(t)$, — $Z(t)$)

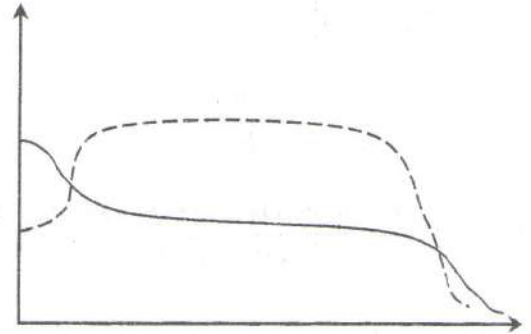


Рис. 3. График зависимости дисперсии и среднего уровня активности при $\alpha \approx \alpha_{кр}^1$ (----- $E(t)$, — $Z(t)$)

Описанная ситуация, которая объясняет возникновение критического режима, диапазон изменения параметра $\alpha_{кр}^1 < \alpha < \alpha_{кр}^2$, является критическим для данного начального уровня активности сети $(E(0), Z(0))$ и фиксированных b и θ . Переход в критический режим связан с увеличением времени приближения к стационарному состоянию. В имитационной модели это соответствует длительному сохранению конфигурации начально активных элементов, что в свою очередь позволяет моделировать кратковременную память. Начальное «пятно» активных элементов как бы «раздумывает» — погаснуть или разгореться. И в зависимости от результата этих длительных «раздумий» оно либо затухает до фоновой активности, либо разгорается до ВУ активности. При этом средний, по множеству реализаций, уровень активности остается промежуточным, а дисперсия растет. Важно подчеркнуть, что критический режим, как в случае имитационного моделирования, так и в системе (7), сохраняется для целого интервала изменения параметра взаимодействия $\alpha \in (\alpha_{кр}^1, \alpha_{кр}^2)$.

В заключение приведем асимптотическую формулу больших времен релаксации при приближении к критическому режиму. Предположим, что значение параметра α близко к $\alpha_{кр}^1$ ($\alpha < \alpha_{кр}^1$). Тогда траектория, идущая от точки $(E(0), Z(0))$, будет

проходить вблизи сепаратрисы седла, и, следовательно, время τ , которое она будет проводить в окрестности седла, будет увеличиваться по мере приближения α к $\alpha_{кр}^1$. Известно, что асимптотическая формула для τ имеет вид:

$$\tau(\delta) \approx \frac{1}{\mu} \ln \frac{1}{\delta},$$

где δ — расстояние от начальной точки до сепаратрисы (малый параметр); μ — положительное собственное значение седла (μ может обращаться в ноль при $\alpha = \alpha_2$ и $\alpha = \alpha_3$).

Выводы

Дано математическое описание возникновения критического режима в сети из возбуждающих элементов с локальными связями. Для этого получена система двух уравнений, описывающая динамику среднего и дисперсии уровня активности сети. Показано, что критический режим существует в целой области изменения параметров и характеризуется «метастабильностью» среднего уровня активности и большой дисперсией. Критический режим — основа для моделирования кратковременной памяти.

Проведено полное двухпараметрическое качественное исследование системы уравнений, описывающих динамику активности сети из возбужда-

ющих и тормозных элементов со связями типа «все на всех». Получено семь режимов функционирования сети. На плоскости параметров указаны границы, разделяющие различные режимы. Получено хорошее соответствие модельных исследований с экспериментальными данными.

Данные исследования могут использоваться как при анализе динамики реальных и искусственных нейронных сетей, так и для исследования некоторых механизмов памяти.

Список литературы: 1. Интеллектуальные процессы и их моделирование. — М.: Мир, 1987. 2. Волик А. В., Герасин С. Н. / Модели динамики нейронной сети с коллективными эффектами // Бионика интеллекта. — 2005. — № 2 (62). — С. 79–82. 3. Коваленко Е. И., Крюков В. И., Борисюк Р. М., Борисюк Г. Н., Кириллов А. Б. Нейронная память: имитационная модель поля СА₃ гиппокампа. — В сб.: Взаимодействующие марковские процессы и их применение к математическому моделированию биологических систем. — Пущино, 1982. С. 77–107. 4. Крюков В. И. Марковские процессы взаимодействия и нейронная активность. — В сб.: Взаимодействующие марковские процессы в биологии. Пущино, 1977. — С. 127–145. 5. Кащенко С. А., Майоров В. В. Об одном дифференциально-разностном уравнении, моделирующим импульсную активность нейрона // Математическое моделирование. — 1993. — Т. 5. — № 12. — С. 13–25. 6. Майоров В. В., Мышкин И. Ю. Математическое моделирование нейронов сети на основе уравнений с запаздыванием // Математическое моделирование. — 1990. — Т. 2. — № 11. — С. 64–76.

Поступила в редакцию 13.09.2005