

М. А. КОМАРОВ¹, Г. В. ОСИПОВ¹, М. С. БУРЦЕВ^{2,3}

¹ННГУ им. Н.И. Лобачевского, Нижний Новгород; ²ИПМ им. М.В. Келдыша РАН, Москва; ³НИИ нормальной физиологии им. П.К. Анохина РАМН, Москва
makomar@mail.ru

АЛГОРИТМ КЛАССИФИКАЦИИ НА ОСНОВЕ СПЕЦИАЛИЗАЦИИ И КОНКУРЕНЦИИ НЕЙРОНОВ*

Рассматривается задача разработки эффективного биологически релевантного нейросетевого алгоритма классификации образов. Нейронная сеть представляет собой ансамбль возбудимых элементов, каждый элемент моделируется системой уравнений Бонхоффера - Ван дер Поля, которая феноменологически описывает спайковую активность нейрона. Обучение происходит за счет специализации - добавления к сети новых нейронов, распознающих специфические признаки запоминаемого образа, и подавляющих по принципу конкуренции активность, препятствующую корректной классификации. Исследованы характеристики модели, произведено сравнение с классической сетью Хопфилда.

Введение

Синхронизованная популяционная активность нейронов рассматривается современной нейронаукой, как основа памяти [1-7]. Считается, что группы нейронов, распределенные в пространстве и активирующиеся одновременно или с фиксированными задержками, образуют «след памяти». В рамках теории функциональных систем [8] нейрональные группы являются проявлением активности функциональных систем на уровне межклеточного взаимодействия. Появление новых нейрональных групп в процессе обучения и созревания памяти соответствует обогащению имеющихся систем и формированию новых. Новые нейрональные группы, возникшие в процессе обучения, предназначены для получения животным адаптивного результата в изменившихся условиях, и поэтому должны хранить интеграцию значимых признаков проблемной ситуации и необходимых действий. Таким образом, в процессе обучения популяция нейронов, объединяющихся в новую группу, должна «запомнить» паттерн

* Данная работа выполнена при поддержке РФФИ, проекты № 06-02-16596, 08-02-97049, 08-02-92004, № 07-01-00180.

входной афферентации, соответствующий проблемной ситуации, и затем при его появлении в будущем стабильно генерировать выработанный ответ.

Описанный выше процесс образования и функционирования нейрональных групп выдвигает следующие требования к биологически релевантной нейросетевой модели:

- для эффективного запоминания признаков ситуации необходимо формировать нейрональные группы при однократном столкновении с проблемой;
- для стабильности получения результата функциональной системой, группы, появляющиеся при последующем обучении, не должны нарушать структуру уже имеющихся систем.

Сегодня, большинство методов обучения как в биологических, так и прикладных моделях нейронных сетей, основаны на правиле Хебба или постепенном накоплении некоторой статистики, и плохо удовлетворяют приведенным требованиям. Так, например, при использовании в модели пластичности зависимой от времени спайка (Spike-timing-dependent plasticity), нейрональные группы сначала нарастают при повторяющемся предъявлении входного паттерна, а потом неожиданно коллапсируют [7].

В данной работе предлагается оригинальный нейросетевой алгоритм, обеспечивающий запоминание паттернов входных возбуждений при однократном их предъявлении, и последующую классификацию входной афферентации по сформированным категориям. Алгоритм основан на гипотезе о специализации нейронов при обучении [9]. В нашей модели нейрональные группы образуются за счет добавления нейронов из «молчащего» пула, при этом связи между нейронами уже сформированных систем не изменяются, но их активность может регулироваться в результате торможения со стороны добавляемых элементов за счет принципа конкуренции [10-14].

Описание модели нейронной сети

В качестве модели отдельного нейрона мы использовали феноменологическую модель Бонхоффера-Ван дер Поля (первые два уравнения), воспроизводящую периодическую генерацию потенциалов действия, при воздействии постоянного внешнего возбуждения [11]:

$$\tau_1 \frac{dx_i(t)}{dt} = x_i(t) - \frac{x_i^3(t)}{3} - y_i(t) - z_i^1(t)(x_i(t) - v^{inh}) - z_i^2(t)(x_i(t) - v^{ex}) + S_i,$$

$$\frac{dy_i(t)}{dt} = x_i(t) - by_i(t) + a,$$

$$\tau_2 \frac{dz_i^1(t)}{dt} = \sum_j (g_{ij}^{inh} F(x_j)) - z_i^1, \quad \tau_3 \frac{dz_i^2(t)}{dt} = \sum_j (g_{ij}^{ex} F(x_j)) - z_i^2.$$

Здесь x_i переменная, описывающая динамику мембранного потенциала, y_i совокупность действия ионных токов, z_i^1 - синаптическую тормозную связь между нейронами, z_i^2 - синаптическую возбуждающую связь между нейронами. S_i параметры в модели, отвечающие внешним стимулам, прикладываемым к нейронам, v^{inh} и v^{ex} - потенциалы реверсии тормозных и возбуждающих синапсов. g_{ij}^{inh} , g_{ij}^{ex} матрицы коэффициентов, определяющих силу и топологию тормозных и возбуждающих связей в нейронной сети. $F(x_i)$ функция Хэвисайда. Значения параметров, неизменных во всех численных экспериментах были следующими: $a = 0.7$, $b = 0.8$, $\tau_1 = 0.08$, $\tau_2 = 3.1$, $\tau_3 = 1.15$, $v^{inh} = -1.5$, $v^{ex} = +1.5$.

Каждый акт обучения в нашей модели состоит в формировании представления для новых признаков входного паттерна и «подавления» активности уже имеющихся нейронов. «Подавление» происходит за счет принципа конкуренции достаточно хорошо известного в теории колебаний и нейродинамике [10-14]. Для иллюстрации рассматриваемого принципа рассмотрим сеть, состоящую из трех элементов: один автоколебательный элемент воздействует на два возбудимых элемента. Возбудимые элементы объединены взаимными тормозными синаптическими связями (Рис. 1).

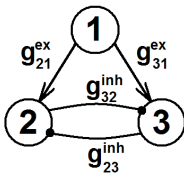


Рис.1 Сеть, состоящая из трех нейронов. 1 – автоколебательный элемент. 2, 3 - возбудимые элементы. Линии со стрелками на конце символизируют возбуждающие связи. Линии с кругами на конце обозначают тормозные связи.

Коэффициенты возбуждающих связей с нейрона 1 на нейроны 2 и 3 примем равными $g_{21}^{ex} = 0.5$, $g_{31}^{ex} = 0.49$. Следует заметить, что воздействие со стороны первого нейрона достаточно сильно для возбуждения

потенциала действия в обоих нейронах. Исследуем динамику ансамбля в зависимости от параметра $G^{inh} = g_{23}^{inh} = g_{32}^{inh}$. Бифуркационная диаграмма на Рис.2 показывает области существования различных режимов в системе. Для ее построения система для сети интегрировалась для каждого значения параметра тормозной связи при различных начальных условиях. По оси ординат откладывалась величина $m = 0.5(\max(x_2) + \max(x_3))$, где $\max(x_i)$ - максимум значения переменной x_i , за время реализации динамики системы (максимум детектировался в установившемся режиме).

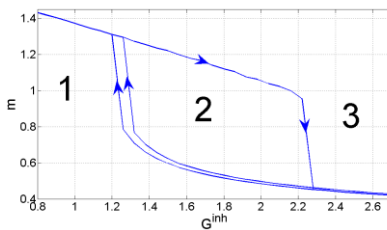


Рис.2. Диаграмма возможных режимов в системе, изображенной на Рис.1 в зависимости от параметра тормозной связи G^{inh} .

Рассмотрим подробно области, изображенные на бифуркационной диаграмме (Рис.2):

1) Область 1 – существует единственный устойчивый предельный цикл (обозначим его l_1), когда потенциалы действия синхронно генерируются элементами 2 и 3 (m максимально).

2) Область 2 – область мультистабильности. Сосуществуют три предельных цикла: l_1 , а также два предельных цикла l_2 и l_3 , когда один из элементов в результате сильного торможения подавил активность конкурента.

3) Область 3 – область бистабильности. Сосуществуют два предельных цикла - l_2 и l_3 .

Область 3 является областью конкуренции, согласно численным расчетам в зависимости от начальных условий в данной области возможны лишь два режима, когда второй элемент подавляет активность третьего, либо наоборот, третий осциллирует, второй подавлен.

Следует заметить, что при $g_{21}^{ex} > g_{31}^{ex} > g^*$, $G^{inh} = 2.5$ и эквивалентных начальных условиях для второго и третьего элемента

($x_2 = x_3, y_2 = y_3, z_2^{1,2} = z_3^{1,2}$), в конкуренции побеждает второй элемент. Диаграмма на Рис. 3 показывает разбиение пространства параметров на области, соответствующие различным режимам, реализующимся при эквивалентных начальных условиях второго и третьего элемента (нейроны идентичные, в начальный момент времени находятся в состоянии равновесия).

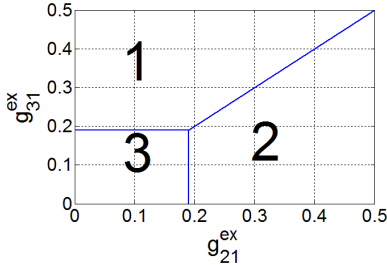


Рис.3. Плоскость параметров (g_{21}^{ex}, g_{31}^{ex}), $G^{inh} = 2.5$. Возможные режимы, реализующиеся при эквивалентных начальных условиях второго и третьего нейронов.

Области 1 и 2 на Рис.3 также соответствуют конкуренции (область 1 - второй нейрон подавил активность третьего, область 2 - третий нейрон подавил активность второго, область 3 соответствует одновременной активности обоих нейронов). Легко видеть, что в случае достаточно сильной тормозной связи между нейронами 2 и 3 ($G^{inh} = 2.5$) в конкуренции побеждает тот нейрон, который получает наибольшее возбуждение от нейрона 1.

Рассмотрим формирование нейрональных групп из пула (слоя) нейронов L_2 , получающих возбуждающий вход от некоторой популяции (слоя) нейронов L_1 . Будем считать, что в процессе обучения формируется нейрональная группа способная в будущем распознать (классифицировать) паттерн активации (эталонный объект или образ) L_1 присутствовавший в момент обучения. Поставим каждому классу (эталонному объекту) в соответствие некоторый нейрон слоя L_2 . Примем для определенности, что задача классификации решена, если при активации некоторого паттерна в L_1 , присутствует активность нейронов, соответствующих только одному классу.

Будем обозначать N_1 и N_2 число нейронов в первом и втором слое соответственно; g_{ij}^{21} - матрицу коэффициентов возбуждающих связей от нейронов первого слоя на нейроны второго слоя, $i = 1, \dots, N_2$, $j = 1, \dots, N_1$; g_{ij}^{22} - матрицу коэффициентов тормозных связей, между нейронами второго слоя, $i, j = 1, \dots, N_2$. Рассмотрим задачу классификации произвольного паттерна к двум классам, соответствующим двум эталонам. Каждому эталонному паттерну поставим в соответствие множество активных нейронов C_k . Таким образом, имеется два эталонных образа, которым соответствует два множества C_1 и C_2 . В общем случае, $C_1 \cap C_2 \neq \emptyset$. Потребуем, чтобы образам C_1 и C_2 соответствовала активность нейронов n_1 и n_2 из слоя L_2 . Для этого необходимо, чтобы нейроны $c_i \in C_1$ образовывали возбуждающие связи на нейрон n_1 , т.е. $g_{n_1 c_i}^{21} = 0.5|C_1|^{-1}$, $c_i \in C_1$. То же самое справедливо и для C_2 и n_2 : $g_{n_2 c_i}^{21} = 0.5|C_2|^{-1}$, $c_i \in C_2$. В силу того, что множества C_1 и C_2 могут иметь пересечение $C_3 = C_1 \cap C_2$, то при предъявлении, например, только множества C_1 , возбуждение может получить также и нейрон n_2 . В связи с этим, для корректной классификации необходимо организовать конкуренцию между множествами нейронов $C_1 \setminus C_3$ и $C_2 \setminus C_3$, т.е. между непересекающимися областями. Для этого в слое L_2 выделяются два нейрона n_3 и n_4 , которые получают возбуждение от областей $C_1 \setminus C_3$ и $C_2 \setminus C_3$ соответственно ($g_{n_3 c_i}^{21} = 0.5|C_1 \setminus C_3|^{-1}$, $c_i \in C_1 \setminus C_3$; $g_{n_4 c_i}^{21} = 0.5|C_2 \setminus C_3|^{-1}$, $c_i \in C_2 \setminus C_3$). Для обеспечения конкуренции, необходимо образовать взаимные ингибиторные связи между нейронами n_3 и n_4 : $g_{n_3 n_4}^{22} = g_{n_4 n_3}^{22} = 2.5$. Как обсуждалось выше, в конкуренции победит тот нейрон, который получит больше возбуждения от нейронов

слоя L_1 , т.е. исход конкуренции будет определяться соотношением активных нейронов из множеств $C_1 \setminus C_3$ и $C_2 \setminus C_3$. Наконец, нейрон, победивший в конкуренции, должен оказывать тормозное воздействие на нейроны, соответствующие другим эталонным паттернам (классам). Т.е. в нашем случае $g_{n_3 n_2}^{22} = g_{n_4 n_1}^{22} = 2.5$. В результате получается сеть, изображенная на Рис. 4. Нейрон n_1 соответствуют эталонному образу (классу) C_1 , n_2 - классу C_2 . Нейроны n_3 и n_4 ответственны за конкуренцию между образами.

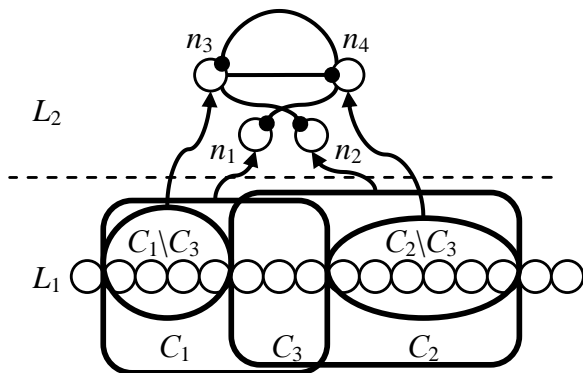


Рис.4. Архитектура рассматриваемой сети, описание в тексте.

Таким образом, для классификации произвольного паттерна активации, например, к классу C_1 , во-первых, необходимо, чтобы было достаточное количество активных нейронов из множества C_1 , в этом случае возможно возбуждение нейрона n_1 . Во-вторых, необходимо, чтобы нейрон n_3 подавил в процессе конкуренции нейрон n_4 . Это произойдет, если произвольный объект имеет большее пересечения с образом C_1 , чем с C_2 . При наличии более двух классов, необходимо, чтобы условия, описанные выше, были выполнены для каждой пары пересекающихся множеств.

Результаты моделирования

В первом эксперименте система обучалась трем образам, которые представляют собой рукописные цифры 1, 4 и 7 (Рис.5а). После организации связей (настройки параметров для динамической системы), сети предъявлялись паттерны, являющиеся искаженными версиями эталонных объектов. (Рис. 5б). Все паттерны были успешно классифицированы.

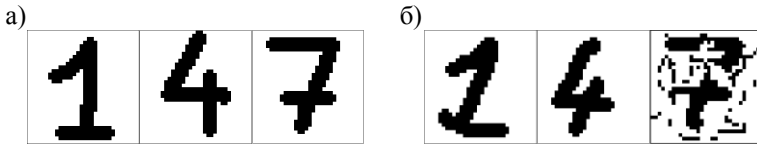


Рис. 5. а) Эталонные объекты. б) Объекты, подлежащие классификации.

В втором эксперименте исследовалась эффективность работы сети в сравнении с ассоциативной сетью Хопфилда. Мерой эффективности выступает величина E , равная вероятности правильного срабатывания сети при предъявлении зашумленного изображения. Величина E строилась в зависимости от двух параметров, P_1 - вероятность молчания нейрона эталонного паттерна, P_2 - вероятность активации нейрона, не принадлежащего эталону. Диаграммы на Рис.6 отображают зависимость величины E от значений P_1 и P_2 .

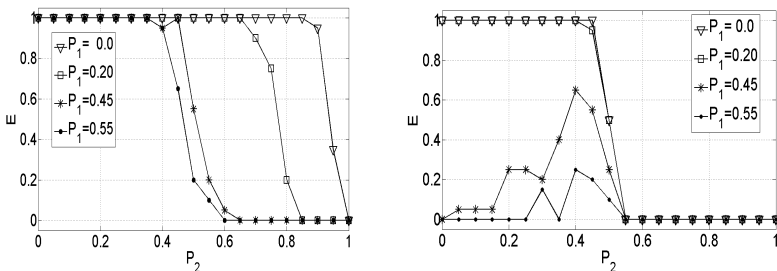


Рис.6. Вероятность правильного отклика сети в ответ на зашумленный образ. а) Предложенная модель б) Ассоциативная сеть Хопфилда.

Обе модели обучались трём, случайно сгенерированным паттернам активации. Каждый паттерн состоял из 213-ти нейронов, для каждой пары (P_1, P_2) усреднение проводилось по 20-ти обучающим активациям паттернов. Можно видеть, что обученная сеть достаточно устойчива к шумам и по эффективности распознавания зашумленных изображений намного превосходит классический вариант ассоциативной сети Хопфилда. Суммарная вероятность корректного отклика сети приблизительно в два раза выше для предложенного алгоритма, чем для сети Хопфилда. Это значит, что предложенный алгоритм обеспечивает более воспроизводимый результат классификации, генерируя «выученный» ответ, как для «слабых» (высокая P_1), так и «гиперактивных» (высокая P_2) паттернов. Тогда как для сети Хопфилда необходим некоторый «средний» уровень возбуждения (что демонстрирует пик $P_2 = 0.4$ при $P_1 \approx 0.5$, т.е. значит в суммарная активация близка к активации исходного образа).

Небольшая модификация архитектуры сети позволяет получить автоассоциативную память, при этом будет происходить одновременно и классификация и восстановление эталонного паттерна в слое L_1 . Вернемся к задаче классификации произвольного объекта к двум классам, соответствующим двум эталонным объектам C_1 и C_2 . Для возможности восстановления искаженного образа, прежде всего, необходимо добавление возбуждающих связей в слое L_1 между нейронами, входящими в одно множество. Т.е. $g_{c_j c_i}^{11} = 0.5|C_1|^{-1}, c_{i,j} \in C_1$, где g_{ij}^{11} - матрица коэффициентов, определяющих силу и топологию возбуждающих связей внутри слоя L_1 . Также и для второго множества: $g_{c_j c_i}^{11} = 0.5|C_2|^{-1}, c_{i,j} \in C_2$. Однако теперь нейроны $c_i \in C_3$ способны возбуждают как нейроны множества C_1 , так и множества C_2 . Таким образом, при предъявлении множества C_1 , в силу того, что $C_1 \cap C_2 \neq \emptyset$, возбуждаться будут также и нейроны множества C_2 . В связи с этим, разумно организовать обратные тормозные связи от нейронов слоя L_2 на нейроны слоя L_1 . Нейрон n_3 из

слоя L_2 , получающий возбуждение от нейронов $c_i \in C_1 \setminus C_3$ слоя L_1 , должен теперь образовать обратное торможение нейронов множества $C_2 \setminus C_3$, за тем, чтобы предотвратить нежелательное распространение активности в слое L_1 . Таким образом, $g_{c_i n_3}^{12} = 2.5, c_i \in C_2 \setminus C_3$. Аналогично $g_{c_i n_4}^{12} = 2.5, c_i \in C_1 \setminus C_3$. Здесь g_{ij}^{12} - матрица коэффициентов, определяющих силу и топологию тормозных связей, направленных от нейронов слоя L_2 на нейроны слоя L_1 . Такая модель также была обучена рукописным цифрам (Рис.5а) и тестировалась на искаженных образах (Рис.5б). Классификация прошла успешно. На Рис. 7 представлены мгновенные снимки распределения переменной x для элементов в слое L_1 в различные моменты времени. Классификация проводилась для искаженного изображения рукописной цифры 7 (Рис.5б). В начальные моменты времени, из-за шумовой составляющей изображения, возбуждается множество элементов, относящихся не только к эталонному изображению рукописной цифры 7 (Рис.5а). Однако, по прошествии некоторого времени, динамика сети представляет собой синхронные (со сдвигом фазы порядка 1 мс) колебания элементов, кодирующих эталонное изображение (сравните с Рис.5а).

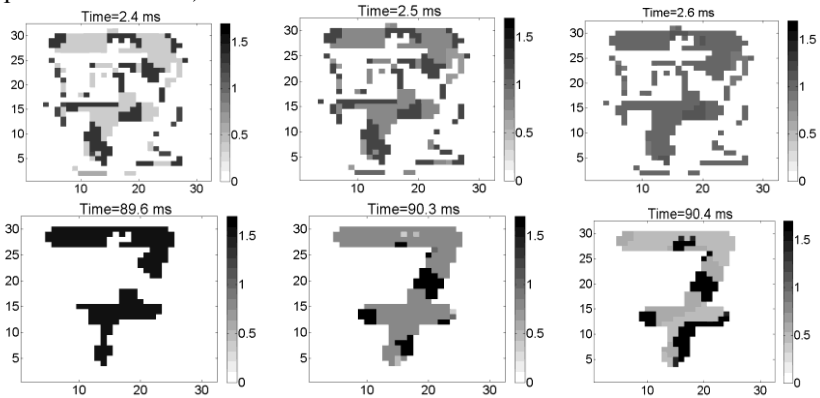


Рис.7. Мгновенные снимки распределения переменной x для элементов в слое L_1 в различные моменты времени.

Заключение

В работе предложена новая модель организации нейронной сети, позволяющей формировать популяции синхронизованно работающих элементов – нейрональных групп, способных категоризовать паттерны входного возбуждения по выученным эталонам. Важными отличиями модели от предложенных ранее являются:

- 1) обучение при однократном предъявлении запоминаемого паттерна,
- 2) неизменность уже имеющейся памяти при новом обучении.

Эти свойства не только делают модель более похожей на биологический аналог, но и дают существенные преимущества для ее прикладного использования. Применение предложенного нейросетевого алгоритма при разработке анимата или робота должно позволить добиться высокой скорости обучения, критичной для автономных систем реального времени, а также большей адаптивности за счет сохранения всей истории обучения.

Еще одним достоинством рассматриваемого алгоритма является постепенность усложнения динамики сети в процессе обучения. Для широкого диапазона вариации входного возбуждения динамика нейросети сходится к имеющемуся в памяти эталону, минимизируя появление «химерных» образов (см. Рис.6. и обсуждение в тексте). Пошаговое добавление только «нужных» нейронов в сеть, позволяет держать на оптимальном уровне сложность поведения сети, и уменьшить вычислительные ресурсы необходимые для ее работы.

Список литературы

1. Edelman G.M. Neural Darwinism: The Theory of Neuronal Group Selection. New York: Basic Books. 1987.
2. Abeles M. Corticonics: Neural Circuits of the Cerebral Cortex. Cambridge University Press. 1991.
3. Bienenstock E. A model of neocortex. Network: Comput. Neural Syst. 1995. Vol. 6. P. 179–224
4. Diesmann M., Gewaltig M.-O., Aertsen A. Stable propagation of synchronous spiking in cortical neural networks. Nature. 1999. Vol. 402. P. 529 - 533.
5. Ikegaya Y., Aaron G., Cossart R., Aronov D., Lampl I., Fester D., Yuste R. Synfire Chains and Cortical Songs: Temporal Modules of Cortical Activity. Science. 2004. Vol. 304. P. 559–564.
6. Izhikevich, E.M., Gally, J.A., and Edelman, G.M. Spike-timing Dynamics of Neuronal Groups. Cerebral Cortex. 2004. Vol. 14. P. 933–944.
7. Izhikevich E.M. Polychronization: Computation with Spikes. Neural Computation. 2005. Vol. 18. P. 245-282.

8. Анохин П.К. Принципиальные вопросы общей теории функциональных систем // Принципы системной организации функций. М., Наук. 1973. С. 5-61. [<http://www.keldysh.ru/pages/BioCyber/RT/Functional.pdf>]
9. Швырков В.Б. Системная детерминация активности нейронов в поведении // Успехи физиол. Наук. 1983. Т. 14. № 1. С. 45-66.
10. Рабинович М.И., Трубецков Д.И. Введение в теорию колебаний и волн // НИЦ «Регулярная и хаотическая динамика», 2000.
11. Afraimovich V.S., Rabinovich M.I., Varona P. Heteroclinic contours in neural ensembles. *Int. J. Bifurcation and Chaos*. 2004. Vol. 14. № 4. P. 1195-1208.
12. Rabinovich M.I., Varona P., Selverston A.I., Abarbanel H.D.I. Dynamical principles in neuroscience. *Review of modern physics*. 2006. Vol. 78. P. 1214-1260.
13. Komarov M.A., Osipov G.V., Suykens J.A.K., Rabinovich M.I. Numerical studies of slow rhythms emergence in neural microcircuits: Bifurcations and stability. *Chaos*. 2009. Vol. 1, P. 015107-015107-8.
14. Komarov M.A., Osipov G.V., Suykens J.A.K. Sequentially activated groups in neural networks. 2009. *EPL*. Vol. 86. P. 60006.