

## О журнале «Нейроинформатика»

Уважаемый коллега!

Перед Вами первый номер электронного рецензируемого журнала «Нейроинформатика», цель которого – оперативная публикация оригинальных и обзорных статей по фундаментальным и прикладным исследованиям нейронных сетей.

Слово «Нейроинформатика» сравнительно недавно вошло в научный обиход и характеризует обширную область исследований, посвященную изучению информационных процессов в искусственных и естественных нейронных сетях. Специалистам хорошо известна Всероссийская научно-техническая конференция «Нейроинформатика», которая проводится ежегодно в конце января в МИФИ под эгидой Российской Ассоциации нейроинформатики. Число и научный уровень докладов, поступающих на конференцию, год от года растет. И сейчас формат конференции уже явно не вмещает все работы российских нейросетевиков. Поэтому необходим и данный журнал.

Тематика журнала включает в себя следующие направления:

- Нейробиология
- Нейронные сети и когнитивные науки
- Нейросетевые парадигмы и архитектуры: представление данных, обучение и оптимизация
- Нейронные сети и самоорганизация систем
- Нейросетевые системы обработки данных, распознавания образов и управления
- Приложения нейроинформатики в медицине, технике, экономике, в естественных и гуманитарных науках
- Модели адаптивного поведения
- Модели эволюции нейронных сетей

В первый номер вошли пять статей.

В статье Я.Б. Казановича и Р.М. Борисюка «Нейросетевая модель слежения за несколькими объектами» предлагается и исследуется модель, характеризующая психологические эксперименты, в которых человек, не теряя внимания, одновременно следит за несколькими независимо перемещающимися зрительными объектами. В модели используется многослойная осцилляторная сеть, в которой каждый слой предназначен для слежения за одним целевым объектом. Представленная в данной работе модель внимания основана на трех основных идеях. Во-первых, в качестве ключевых механизмов формирования фокуса внимания используются колебания и их синхронизация. Во-вторых, используется гипотеза о существовании специальной нейронной системы, центрального управляющего элемента, координирующего фокусировку внимания. В терминах модели зрительное внимание характеризуется как следствие синхронизации активности

центрального элемента и связанного ансамбля нейронов в первичных зонах зрительной коры. В-третьих, в модели используется резонанс. Его назначение – формализовать гипотезу об увеличении уровня нейронной активности в областях мозга, включенных в фокус внимания, и подавить до низкого уровня активность, возникающую в ответ на стимулы вне фокуса внимания. В работе проведен цикл компьютерных экспериментов и продемонстрирована адекватность модели психологическим данным.

Статья Ю.Р. Цоя и В.Г. Спицына «Эволюционный подход к настройке и обучению искусственных нейронных сетей» содержит аналитический обзор современных работ по применению эволюционных алгоритмов для настройки структуры и синаптических весов нейронных сетей.

Статья Г.А. Ососкова «Метод эластичных нейронных сетей и его робастная трактовка» содержит как краткий обзор по нейросетевым методам обработки данных в физике высоких энергий, так и изложение оригинальных методов выделения траекторий элементарных частиц с помощью эластичных нейронных сетей. Эластичная нейронная сеть использует уравнение трека частицы, зависящего от вариаций параметров нейросети таким образом, чтобы, изгибаясь при их изменении, кривая, описываемая этим уравнением, прошла как можно ближе к «своим» точкам, измеренным на треке, «отталкиваясь» от «чужих». В статье также описан формализм робастного подхода и на базе исследования совместной оценки параметров положения и масштаба получен экономичный алгоритм робастного оценивания.

В статье А.В. Чижова и А.А. Турбина «От моделей единичных нейронов к моделям популяций нейронов» составляется классификация современных моделей разного уровня описания: от детальных моделей единичных нейронов до моделей нейронных популяций, характеризующих среднюю возбудимость большого массива нейронов. Классифицируется иерархия моделей. Приводятся уравнения типичных одномерных популяционных моделей, т.е. континуальных моделей, различающих состояние нейронов между спайками по одному параметру. Излагаются две оригинальные модели нейронных популяций, рассматривающие распределение нейронов по предполагаемому времени до ожидаемого спайка или по времени, прошедшему от момента предыдущего спайка.

В статье Б.В. Крыжановского и Б.М. Магомедова «О вероятности обнаружения локальных минимумов энергии в обобщенной модели Хопфилда» исследуется обобщение модели Хопфилда, в котором каждый эталонный вектор записывается со своим весом. Это позволяет описывать нейронную сеть с невырожденным спектром минимумов энергии. В работе получены аналитические зависимости, устанавливающие связь между глубиной локального минимума и шириной его области притяжения. На основании этого вероятность нахождения локального минимума при случайной инициализации нейронной сети удалось представить как функцию глубины этого минимума. Развита аналитическая теория подтверждена компьютерными расчетами.

#### Предисловие

Представленные в номере статьи характеризуют типичные и вместе с тем разные аспекты нейросетевых исследований.

Приглашая специалистов к активному участию в работе журнала, хотелось бы отметить, что пытливого исследователя, да и просто любознательного человека, всегда будут волновать вопросы: как работает мозг, каковы процессы обработки информации в нейронных структурах, как активность нервной системы связана с мышлением, интеллектом человека? И очевидно, что у исследований в области нейроинформатики большие перспективы.

Главный редактор журнала, д.ф.-м.н. В.Г. Редько

# Нейросетевая модель слежения за несколькими объектами\*

Казанович Я.Б.<sup>1</sup>, Борисюк Р.М.<sup>1,2</sup>

<sup>1</sup>Институт математических проблем биологии РАН, Пущино, Московская область, yakov\_k@impb.psn.ru

<sup>2</sup>Центр теоретической и вычислительной нейробиологии, Университет г. Плимут, Плимут, Великобритания, rborisyuk@plymouth.ac.uk

**Аннотация.** Психологические эксперименты показывают, что человек способен одновременно следить за пятью независимо перемещающимися зрительными объектами. В данной работе предлагается нейросетевая модель этого феномена. Модель работает с набором идентичных зрительных объектов, движущихся в заданном поле. На начальном этапе функционирования при неподвижных объектах модель выбирает в фокус внимания подмножество объектов, которые исходно заданы как целевые. Остальные объекты считаются дистракторами, отвлекающими внимание от слежения за целевыми объектами. В дальнейшем при движении объектов модель должна по возможности сохранять начальное разделение объектов на целевые и дистракторы. Идея реализации модели состоит в использовании многослойной осцилляторной сети, в которой каждый слой предназначен для слежения за одним целевым объектом. Будет показано, что одновременное слежение за несколькими объектами может быть реализовано при подходящем выборе синхронизирующих/десинхронизирующих взаимодействий между слоями. Результаты имитационного моделирования показывают, что в случае пересекающихся объектов вероятность ошибки слежения (принятие дистрактора за целевой объект или целевого объекта за дистрактор) возрастает при увеличении числа дистракторов. Такая зависимость имеет место и в психологических экспериментах. Кроме того будет описано поведение модели в случае работы с пересекающимися объектами.

## 1. Введение

Селективное зрительное внимание играет важную роль в жизнедеятельности животных и человека, позволяя выделять из большого объема зрительной информации ту часть, которая в данный момент наиболее значима и которая должна быть подвергнута более детальной обработке. Психологи выделяют несколько типов селективного внимания, которые соответствуют различным стратегиям формирования фокуса внимания. Пространственно-ориентированное внимание характеризуется в пространственных терминах, таких как "проектор" или "увеличительное стекло", которые могут быть направлены на оп-

---

\* Работа поддержана Российским фондом фундаментальных исследований (проект 03-04-48482) и Президентской программой для ведущих научных школ (проект НШ 1872.2003.4).

ределенную часть зрительного поля [20, 48]. Объектно-ориентированное внимание привязано к определенным объектам независимо от того, в какой части пространства эти объекты находятся [19, 52].

Одной из экспериментальных парадигм в изучении объектно-ориентированного внимания является слежение за несколькими объектами (СНО). В традиционных экспериментах по СНО [49-52] испытуемому показывается экран дисплея, содержащий  $m$  простых по форме идентичных объектов (обычно используется до 10-12 объектов, таких как точки, круги, крестики и т.д.). В начале эксперимента некоторое подмножество этих объектов, содержащее от 1 до  $m/2$  объектов (предполагается, что число  $m$  четное), коротко подсвечивается вспышками света. Подсвеченные таким образом объекты являются целевыми в том смысле, что именно за их перемещениями испытуемый должен следить на последующей стадии эксперимента. Остальные объекты считаются дистракторами. Дистракторы предназначены для повышения трудности слежения, поскольку они отвлекают внимание от целевых объектов.

После разделения объектов на целевые и дистракторы начинается следующий этап эксперимента, при котором объекты движутся независимо друг от друга в случайно выбранных и постоянно меняющихся направлениях. На этом этапе яркость всех объектов одинакова. Задача испытуемого состоит в том, чтобы следить за движениями целевых объектов при неподвижных зрачках (взгляд фиксируется на центр экрана). Время от времени один из объектов подсвечивается вспышкой света. В этот момент испытуемый должен нажать определенную клавишу в зависимости от того, идентифицирует ли он подсвеченный объект как целевой или дистрактор. Качество слежения характеризуется числом сделанных ошибок. В другой модификации эксперимента через фиксированное время (несколько секунд) перемещение объектов останавливается, и испытуемый должен с помощью компьютерной мышки указать все целевые объекты.

Проведенные в работе [50] эксперименты показали, что тренированные испытуемые способны следить за движением 2-5 целевых объектов с достаточно высокой надежностью. Хотя вероятность ошибки при идентификации растет с увеличением числа целей, даже в случае 5 целевых объектов вероятность правильной идентификации достигает 85%.

При интерпретации экспериментальных результатов было показано, что в данном случае имеет место параллельная обработка информации о целевых объектах, что соответствует гипотезе разделенного внимания. Последовательное циклическое переключение внимания по всем целевым объектам здесь вряд ли возможно, поскольку с учетом быстроты движения объектов перенос "проектора" внимания с одного объекта на другой должен был бы происходить с неправдоподобной с точки зрения нейрофизиологии скоростью.

Последующие работы по СНО [9, 40, 46, 53, 54, 65, 67] подтвердили, что первичная зрительная система способна "следить за движением до пяти зрительных объектов, причем осуществляется это без кодирования их визуальных характеристик" [49]. В работе [65] СНО исследовалось в случае, когда объекты могли пересекаться в пространстве при движении, так что их изображения накладывались одно на другое. Эксперименты проводились как при наличии "подсказки", характеризующей положение объектов по глубине, так и без такой

подсказки. Результаты экспериментов показывают, что "хотя задача слежения становится более трудной... тем не менее, она выполнима даже в случае перемещения объектов по двумерной плоскости". При наличии "подсказки" надежность СНО приближается к ситуации с непересекающимися объектами.

В настоящее время феномен внимания является популярным объектом нейросетевого моделирования. Разрабатываемые модели можно подразделить на две категории. Коннекционистские модели [25, 28, 29, 42, 47, 63], основанные на так называемой процедуре "победитель получает все" (winner-take-all), функционируют путем подходящей модификации весов связей в иерархической нейронной сети. Модели такого типа мало подходят для работы с движущимися объектами, поскольку для каждого нового положения объектов веса связей приходится пересчитывать заново.

Другую разновидность моделей внимания представляют осцилляторные нейронные сети [10, 11, 13, 32, 33, 66] (см. также обзорную статью [2]). Такие сети более подходят для моделирования объектно-ориентированного внимания, так как они оперируют в фазово-частотном пространстве. Благодаря этому достигается инвариантность фокусировки внимания по отношению к расположению объектов в физическом пространстве зрительного поля.

Модель СНО, представленная в данной работе, основана на разработанной нами ранее модели внимания с центральным осциллятором (МВЦО) [10, 11, 33-35]. Каждый элемент МВЦО представляет собой осциллятор, динамика которого описывается тремя переменными: фазой колебаний, амплитудой колебаний и собственной частотой осциллятора. Взаимодействие между осцилляторами реализуется в терминах фазовой синхронизации, резонанса и адаптации собственной частоты осциллятора.

МВЦО представляет собой однослойную сеть локально связанных осцилляторов, так называемых *периферических осцилляторов* (ПО), работа которых координируется *центральным осциллятором* (ЦО) [3, 17]. ЦО связан со всеми ПО прямыми и обратными связями. ПО представляют колонки кортикальных нейронов в первичных зонах зрительной коры (поля 17-18), реагирующие на определенные локальные признаки на изображении. Для простоты в качестве таких признаков используется контраст между интенсивностью заданного пикселя и интенсивностью фона. ЦО играет роль центрального управляющего элемента системы внимания (central executive) [5, 14]. В МВЦО изолированные объекты представляются в сети синхронными ансамблями ПО, а фокус внимания формируется теми ПО, которые работают синхронно с ЦО.

В психологической литературе центральный элемент представляет собой систему, осуществляющую управление рабочей памятью с помощью внимания [5-7, 56]. Вопрос о локализации этой системы в мозге до сих пор является предметом дискуссии. До недавнего времени считалось, что функцию центрального элемента выполняет определенная локальная область в префронтальной коре [17, 41]. Однако новые исследования дают основание предположить, что центральный элемент, по всей видимости, реализуется распределенной нейронной сетью, включающей латеральную префронтальную, орбитофронтальную и медиальную префронтальную зоны коры [4, 8]. Данные, полученные с помощью функциональной магнитной томографии и позитронно-эмиссионной томогра-

фии, показывают, что "различные управляющие функции не только связаны с определенными зонами фронтальной коры, но и зависят от задних (в основном теменных) областей" [12]. Помимо областей новой коры, также и гиппокамп может играть важную роль в реализации функций центрального элемента, поскольку: гиппокамп а) занимает финальное положение в пирамиде конвергентных зон коры [16], б) координирует обработку информации практически во всех областях неокортекса [27], в) координирует работу системы внимания [18, 26, 64].

С технической точки зрения наличие в системе центрального элемента, имеющего связи со всеми остальными элементами сети, позволяет сократить число связей в системе до величины порядка  $n$ , где  $n$  – число элементов в системе, в то время как для полностью связанной сети потребовалось бы число связей порядка  $n^2$ . Это свойство делает системы с центральным элементом биологически реализуемыми даже в случае больших значений  $n$ .

В работах [11, 34] было показано, что с использованием принципов синхронизации и резонанса МВЦО может решать задачи последовательного выбора изолированных объектов в фокус внимания. В работе [1] МВЦО была использована для слежения за одним целевым объектом, движущимся среди объектов-дистракторов.

Идея одновременного слежения за  $m$  целевыми объектами состоит в том, чтобы объединить в единую сеть  $m$  копий МВЦО, каждая из которых должна следить за движениями одной цели. Реализация этой идеи упирается в решение следующих проблем. Во-первых, нужно предотвратить возникновение ситуации, когда один и тот же целевой объект одновременно попадает в фокус внимания нескольких МВЦО. Во-вторых, модель должна сохранять работоспособность в случае, когда объекты пересекаются в ходе своих перемещений. Поскольку объекты идентичны и движутся в случайных и постоянно изменяющихся направлениях, не существует безошибочной стратегии идентификации целевого объекта после того, как существенная его часть оказалась закрыта дистрактором. Разумной стратегией в данном случае является включение в фокус внимания любого из двух только что разошедшихся объектов, один из которых был целевым, а другой дистрактором. В случае существенного пересечения двух целевых объектов идентификация каждого из них также оказывается невозможной. Единственное, что может сделать в этом случае система внимания, это продолжать следить за обоими такими объектами после того, как они разойдутся. При этом возможно, что две МВЦО обменяются объектами, которые были у них в фокусе внимания. При следовании такой стратегии общее число объектов в фокусе внимания всегда остается равным  $m$ .

В разд. 2 будет описано, как согласовать работу нескольких МВЦО при слежении за набором из двух и более целевых объектов. В разд. 3, будут представлены результаты имитационных вычислений. Раздел 4 содержит обсуждение результатов.

## 2. Архитектура и принципы функционирования модели

Поскольку МВЦО является основным функциональным модулем модели СНО, опишем сначала архитектуру и принципы функционирования МВЦО. Архитектура МВЦО показана на рис. 1.

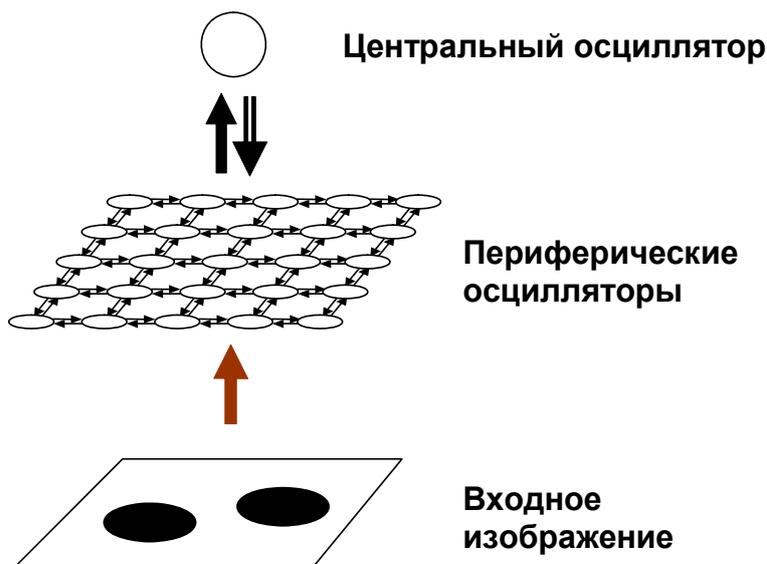


Рис. 1. Архитектура МВЦО. Коричневая стрелка показывает присвоение собственных частот периферическим осцилляторам. Черные стрелки показывают синхронизирующие связи, используемые (а) для связывания осцилляторов, кодирующих один объект, в синхронный ансамбль, (б) для синхронизации ансамбля ПО с ЦО. Полая стрелка показывает десинхронизирующие связи, используемые для предотвращения одновременной синхронизации ЦО с несколькими ансамблями ПО.

Входной информацией для сети является многоградационное монохроматическое изображение на плоской решетке пикселей. Решетка имеет те же размеры, что и сеть из периферических осцилляторов, так что каждый ПО получает внешний сигнал от пикселя, координаты которого идентичны координатам ПО. Для кодирования изображения в сети осцилляторов используется традиционный подход, согласно которому внешний сигнал задает собственную частоту осциллятора [38, 45]. Предполагается, что внешний сигнал формируется в латеральном генеральном ядре и зависит от контраста между интенсивностями пикселя и фона. В дальнейшем будут рассматриваться темные объекты на белом фоне, так что интенсивность фона всегда будет выше, чем интенсивность пикселей объектов. Величина собственной частоты осциллятора задается формулой

$$\omega_i = \lambda(B - I_i), \quad (0 \leq I_i \leq B),$$

где  $\omega_i$  – собственная частота  $i$ -го ПО,  $I_i$  – интенсивность  $i$ -го пикселя,  $B$  – интенсивность фона,  $\lambda$  – шкалирующий параметр. Таким образом, большей величине контраста соответствует большее значение собственной частоты осциллятора.

ПО, соответствующие пикселям объектов, называются *активными*. Динамика этих осцилляторов задается уравнениями (1.1)–(1.4) из Приложения 1. ПО, соответствующие пикселям фона называются *молчащими*. Молчащие осцилляторы не принимают участия в динамике системы.

Каждый ПО имеет синхронизирующие локальные связи со своими ближайшими соседями. Эти связи используются для объединения в синхронный ансамбль периферических осцилляторов, представляющих изолированный объект. Такой способ представления объектов в нейронной сети соответствует синхронизационной гипотезе о связывании признаков объектов в единое целое [57].

ЦО имеет десинхронизирующие прямые и синхронизирующие обратные связи на каждый ПО. Синхронизирующие связи используются для синхронизации ЦО с ансамблем ПО. Десинхронизирующие связи используются для распределения ансамблей ПО в фазово-частотном пространстве так, чтобы разные ансамбли ПО работали несинхронно. Это препятствует одновременной синхронизации с ЦО нескольких ансамблей ПО.

Синхронизирующие и десинхронизирующие взаимодействия между ПО и ЦО порождают конкуренцию между различными ансамблями ПО за синхронизацию с ЦО. Только один ансамбль ПО может "победить" в этой борьбе, поэтому в каждый момент времени только один объект может быть включен в фокус внимания (исключение составляют лишь короткие периоды переходной динамики).

Синхронная работа ЦО и ансамбля ПО имеет следствием резонанс: амплитуды колебаний осцилляторов в ансамбле быстро возрастают до высокого уровня. Амплитуды остальных осцилляторов (не работающих синхронно с ЦО) подавляются до низкого уровня. Высокая амплитуда колебаний ПО является критерием того, что данный ПО включен в фокус внимания. Амплитуда колебаний ЦО является в случае МВЦО постоянной.

Поскольку ЦО может войти в синхронизацию только с теми ПО, чьи собственные частоты находятся в некотором диапазоне относительно собственной частоты ЦО, для настройки ЦО на частоту определенного ансамбля ПО необходимо, чтобы собственная частота ЦО совпадала с собственной частотой данного ансамбля ПО. В связи с этим в МВЦО собственная частота ЦО сделана адаптирующейся к текущей частоте ЦО. Благодаря этому ЦО как бы "ищет" в фазово-частотном пространстве тот ансамбль ПО, который является наиболее подходящим кандидатом на синхронизацию с ЦО.

Примеры численных экспериментов с МВЦО по выбору объектов в фокус внимания и по слежению за движением объекта можно найти в работах [1, 11, 34].

Перейдем теперь к описанию модели СНО. Архитектура соответствующей осцилляторной сети приведена на рис. 2.

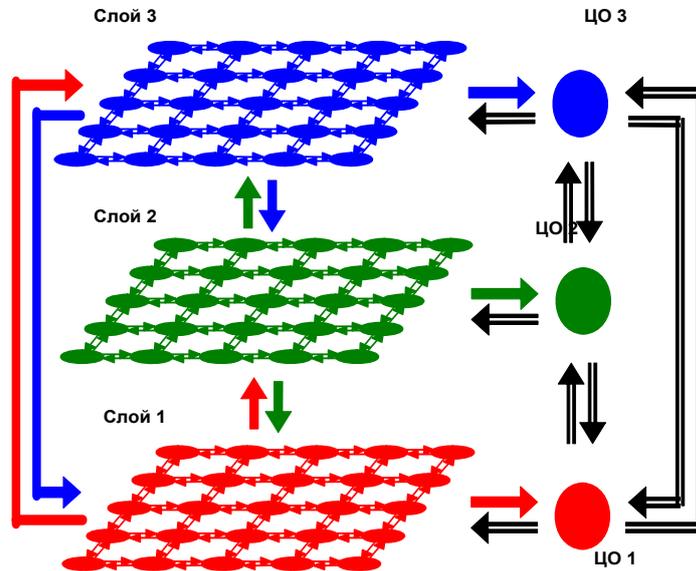


Рис. 2. Архитектура модели СНО. Слои сети раскрашены в разные цвета. Каждый слой предназначен для слежения за одним целевым объектом. Сплошные стрелки показывают синхронизирующие связи, полые стрелки показывают десинхронизирующие связи.

Показанная на этом рисунке сеть содержит 3 слоя (модуля) МВЦО, что соответствует случаю слежения за тремя целевыми объектами (в общем случае число слоев равно числу целевых объектов). Эти модули рассматриваются в модели как подсистемы системы внимания. Каждая подсистема предназначена для слежения за своим целевым объектом. В дальнейшем будет удобно использовать названия цветов (красный, синий, зеленый) для указания на конкретную подсистему внимания. Уравнения динамики модели приведены в Приложении 2 для общего случая  $m$  целевых объектов.

Периферические осцилляторы, имеющие одинаковые координаты на плоскости, но принадлежащие разным слоям, образуют колонку. ПО в колонке связаны сильными синхронизирующими связями типа "все на всех". Все ПО в колонке получают одинаковый внешний сигнал от соответствующего пикселя в зрительном поле. Как и в случае МВЦО, внешний сигнал определяется контрастом между интенсивностями в пикселе и на фоне. Этот сигнал задает собственные частоты ПО в колонке. В силу сформулированных условий собственные частоты в колонке идентичны, что способствует быстрой синхронизации осцилляторов в колонке.

Синхронизирующие локальные связи между ПО внутри слоя ограничены ближайшими соседями. Эти связи сделаны достаточно сильными, чтобы быстро синхронизировать все ПО, соответствующие изолированному объекту. Таким образом, наличие объекта в зрительном поле приводит к формированию синхронного ансамбля ПО, в который входят все колонки, активированные сигналами от данного объекта.

ЦО, принадлежащие различным слоям, соединены десинхронизирующими связями. Это сделано для того, чтобы предотвратить одновременную синхронизацию нескольких ЦО с одним и тем же ансамблем ПО при распределении целевых объектов между подсистемами внимания. Вследствие такой десинхронизации изолированные целевые объекты представляются в сети асинхронной активностью различных ансамблей ПО.

Как и в случае одного целевого объекта, если ансамбль ПО в  $k$ -ом слое работает синхронно с расположенным в этом слое ЦО, то амплитуды колебаний этих ПО быстро возрастают. Если амплитуда колебаний превышает заданный порог  $R$ , то считается, что осциллятор находится в резонансном состоянии. Если все ПО в  $k$ -ом слое, соответствующие пикселям некоторого объекта, находятся в резонансном состоянии, это интерпретируется как факт включения этого объекта в фокус внимания  $k$ -ой подсистемы внимания.

Опишем, что происходит с уже сформированным фокусом внимания при движении объектов. Если объекты движутся достаточно медленно и не пересекаются друг с другом, то фокус внимания обладает высокой устойчивостью. Дело в том, что резонансные ПО оказывают намного более сильное влияние на ЦО в своем слое, чем остальные ПО. Это препятствует переключению внимания на какой-либо ансамбль нерезонансных ПО. Ситуация изменяется, если скорость движения объектов становится достаточно большой по сравнению со скоростью протекания процессов синхронизации и возникновения резонанса. В этом случае возникают спонтанные переключения фокуса внимания с одного объекта на другой. Это приводит к ошибкам в идентификации целевых объектов и дистракторов.

Рассмотрим, что произойдет с фокусом внимания, если при движении имело место наложение двух объектов. Если оба объекта были вне фокуса внимания, то при их наложении фокус внимания не изменится. Временно образовавшийся дистрактор сложной формы так же не будет иметь возможность существенно влиять на фокус внимания, как и отдельные объекты-дистракторы.

Предположим, что произошло наложение двух объектов  $A$  и  $B$ , первый из которых входил в фокус внимания, а второй нет. В этом случае внимание распространится на весь составной объект, поскольку все ПО, соответствующие этому объекту будут включены в общий синхронный ансамбль ПО. Этот ансамбль будет работать синхронно с тем же ЦО, с которым был синхронизован ансамбль ПО, соответствующий объекту  $A$  до его пересечения с  $B$ .

После того как объекты  $A$  и  $B$  снова разойдутся и станут изолированными, соответствующие им ансамбли ПО возобновляют конкуренцию за то, чтобы быть включенными в фокус внимания. Поскольку связи от ЦО к ПО десинхронизирующие, только один из двух объектов имеет возможность выиграть в борьбе за синхронизацию с ЦО. Какой именно, зависит от случая – у системы нет информации, чтобы идентифицировать тот объект, который был в фокусе внимания до пересечения объектов. Если пересечение объектов  $A$  и  $B$  было достаточно значительным, то любой из них может быть выбран в фокус внимания с вероятностью 0.5, что, естественно, с той же вероятностью может повлечь возникновение ошибок при идентификации данных объектов.

Пусть произошло наложение двух объектов таких, что оба они были в фокусе внимания. В этом случае оба объекта останутся в фокусе внимания, несмотря на десинхронизирующие связи между центральными осцилляторами. Это достигается за счет того, что десинхронизация слаба по сравнению с синхронизирующим воздействием, получаемым центральными осцилляторами от общего ансамбля синхронно работающих ПО.

После того как объекты разошлись, десинхронизирующие связи между ЦО снова начинают играть решающую роль в динамике активности этих осцилляторов. Вследствие этого каждый объект "стремится" остаться в фокусе внимания ровно одного из модулей МВЦО. Конечно, вполне возможна такая ситуация, что объекты  $A$  и  $B$ , которые до пересечения друг с другом были в фокусе внимания, скажем, "красного" и "зеленого" модулей, поменяют следящие за ними подсистемы: после того как объекты  $A$  и  $B$  разойдутся, объект  $A$  окажется в фокусе внимания "зеленого" модуля, а объект  $B$  – "красного". Произойдет ли такая смена, зависит от того, насколько сильно пересеклись объекты  $A$  и  $B$  и сколько долго длилось это пересечение.

На самом деле, взаимодействие синхронизирующих и десинхронизирующих влияний на ЦО, реализованное в модели СНО, несколько более сложно, чем это было только что описано. Как показали вычислительные эксперименты, константная сила десинхронизирующего взаимодействия между центральными осцилляторами не может гарантировать нужной динамики ЦО. Если величина этой десинхронизирующей связи выбрана слишком большой, то ЦО может потерять синхронизацию с ансамблем ПО в тот момент, когда произошло наложение двух объектов, которые были в фокусе внимания. В результате только один ЦО сохранит синхронизацию с образовавшимся сложным объектом. Если же величина десинхронизирующей связи между ЦО выбрана слишком слабой, то после разделения сложного объекта оба ЦО могут оказаться синхронизованными с одним из объектов, а второй объект окажется вне фокуса внимания. Оказалось, что не удается подобрать константную величину связи между ЦО, которая бы позволила достаточно надежно избегать тех или иных из указанных ошибок.

С этой трудностью удалось справиться, предположив, что десинхронизирующее взаимодействие между ЦО усиливается после формирования сложного объекта из двух объектов, входивших в фокус внимания. В качестве механизма реализации этой гипотезы был использован резонанс, но в данном случае он был применен к центральному осцилляторам. Таким образом, в отличие от отдельной МВЦО, в модели СНО амплитуды колебаний ЦО не постоянны.

Резонансное возрастание амплитуды ЦО происходит, когда два объекта, находившихся в фокусе внимания, пересекаются друг с другом (см. уравнение (2.5) в Приложении 2). Благодаря последнему члену в уравнении (2.1) резонанс ЦО приводит к возрастанию десинхронизации между ЦО, в фокусе внимания которых находились указанные объекты. Поэтому в момент, когда объекты расходятся, десинхронизация между ЦО оказывается достаточно большой, чтобы предотвратить ситуацию, когда оба ЦО "следят" за одним объектом, в то время как второй объект остается вне фокуса внимания.

Для воспроизведения экспериментальных данных модель СНО помимо слежения за объектами должна быть способна выбрать в фокус внимания за-

данный набор целевых объектов. В экспериментальных условиях целевые объекты указываются испытуемому с помощью коротких вспышек света. Это делает целевые объекты заметными среди всего множества предъявленных объектов. В модели СНО для выбора объектов представление о заметности объектов формализуется в терминах силы действия периферических осцилляторов на центральные осцилляторы. Предполагается, что на этапе выбора целевых объектов те ПО, которые соответствуют пикселям, принадлежащим подсвечивавшимся объектам, оказывают на ЦО более сильное синхронизирующее воздействие, чем остальные ПО. В результате ансамбли ПО, соответствующие подсвечиваемым объектам, получают преимущество в конкуренции за синхронизацию с ЦО.

Идея о том, что фокусировка внимания связана с формированием в мозгу некоторой карты заметности была предложена в работе [36] и в дальнейшем активно использовалась в нейросетевых моделях зрительного поиска [28-30, 47]. Карта заметности представляет собой двумерную таблицу, которая кодирует заметность объектов среди их окружения в зрительной сцене и определяет приоритет включения объектов в фокус внимания. Для простоты в модели СНО карта заметности не вычисляется какой-либо нейронной схемой, а вводится формально как набор параметров  $s_i$ , соответствующих пикселям изображения и влияющих на силу воздействия ПО на ЦО, как это показано в уравнении (2.1) Приложения 2. Для различения подсвеченных и неподсвеченных объектов параметры  $s_i$  могут принимать одно из двух значений: большее значение  $S_{flashed}$  соответствует пикселям подсвеченных объектов, меньшее значение  $S_{non-flashed}$  соответствует пикселям неподсвеченных объектов. Для пикселей фона  $s_i = 0$ . Компьютерные вычисления показывают, что для того, чтобы с высокой степенью надежности обеспечить выбор в фокус внимания подсвеченных объектов, необходимо, чтобы величина  $S_{flashed}$  была в несколько раз больше, чем  $S_{non-flashed}$ . В описываемых далее компьютерных экспериментах были использованы значения  $S_{flashed} = 5$  и  $S_{non-flashed} = 0,2$ . После окончания подсветки целевых объектов и начала этапа движения все объекты имеют одинаковую заметность. Это отражается в модели заданием идентичных значений для всех  $s_i$ . В компьютерных экспериментах это значение полагалось равным 1.

### 3. Имитационные вычисления и сравнение с экспериментальными данными

Мы рассмотрим два случая функционирования модели СНО, соответствующие движениям объектов без пересечений и с пересечениями. В первом случае целью имитационного моделирования будет сравнить результаты функционирования модели с недавними экспериментальными данными по СНО, опубликованными в работе [46]. Во втором случае будет продемонстрировано, что в случае пересечения объектов фокус внимания в модели изменяется согласно процедурам, описанным в разделе 2. Тем самым, хотя число ошибок

слежения увеличивается, система слежения остается работоспособной в тех пределах, которые естественным образом вытекают из правил случайного движения объектов.

Авторы работы [46] экспериментировали с 12 объектами. Чтобы несколько ускорить вычисления мы ограничились 10 объектами, как это имело место в экспериментах [50]. Это не вносит существенных изменений в полученные результаты. В имитационных вычислениях в качестве объектов использовались черные квадраты размером  $7 \times 7$  пикселей, расположенные на белом фоне в поле размером  $30 \times 60$  пикселей. Рассматривались примеры с  $k$  целевыми объектами ( $2 \leq k \leq 5$ ). В соответствии с этим использовались сети с  $k$  слоями (модулями МВЦО).

Параметры модели были подобраны таким образом, чтобы она более или менее адекватно отражала временные характеристики, типичные для психологических экспериментов по СНО. В качестве единицы времени был выбран интервал 100 мсек. Собственные частоты всех ПО были заданы равными  $\omega_i = 5$  (5 колебаний за 100 мсек), что соответствует частоте 50 Гц. Таким образом, рабочей полосой частот модели является диапазон гамма-ритма. Амплитуды всех ПО в начальный момент устанавливались на значение 2, а в процессе работы системы варьировали в пределах (1, 11). Порог для резонансной амплитуды  $R = 8.8$ . Если амплитуда колебаний ПО превосходила  $R$ , считалось, что соответствующий пиксель включен в фокус внимания. Детальная информация о параметрах модели приведена в Приложении 2.

Время однократной работы модели (имеется в виду внутреннее время модели, а не длительность реальных вычислений на компьютере) будем называть **рабочим циклом**. Рабочий цикл составляет 7,2 сек и подразделяется на три фазы. Первая фаза (экспозиция) занимает 0,7 сек и используется для выделения целевых объектов. В это время объекты неподвижны. Единственное отличие целевых объектов от дистракторов задается с помощью карты заметности, которая формируется так, как это было описано в разделе 2. Ожидаемый результат работы модели при экспозиции состоит в том, что каждый целевой объект включается в фокус внимания одной из подсистем внимания. Ошибка может возникнуть в том случае, когда две подсистемы внимания фокусируются на одном целевом объекте или когда в фокус внимания одной из подсистем внимания включается объект-дистрактор. Компьютерные вычисления показали, что при выбранных параметрах модели такие ошибки весьма редки. Их вероятность не превосходит 0,005, поэтому практически они могут не приниматься во внимание.

Вторая фаза функционирования модели (слежение) занимает 6 сек. В течение этого времени объекты перемещаются в случайных направлениях. Скорость перемещения составляет 1 пиксель за 50 мсек, то есть каждые 50 мсек все квадраты подвергаются сдвигу на 1 пиксель в одном из четырех направлений: вверх, вниз, влево, вправо. Для каждого квадрата направление движения выбирается случайно и независимо с вероятностью 0,25. Для предотвращения столкновения квадратов вводится ограничение, согласно которому объекты должны отстоять друг от друга по меньшей мере на один пиксель. Сначала решается вопрос о том, будет ли объект двигаться в горизонтальном или вертикальном

направлении. Этот выбор делается с вероятностью 0,5. Для горизонтального движения затем решается вопрос, будет ли оно сделано влево или вправо. Этот выбор также делается с вероятностью 0,5. Аналогично, если принято решение о вертикальном движении, то его направление вверх или вниз выбирается с вероятностью 0,5. Если возникает опасность столкновения с другим объектом, направление движения меняется на противоположное. Если и в этом случае есть опасность столкновения, то в данный момент объект остается на месте. Аналогичные правила действуют для того, чтобы предотвратить пересечение объектом границы поля.

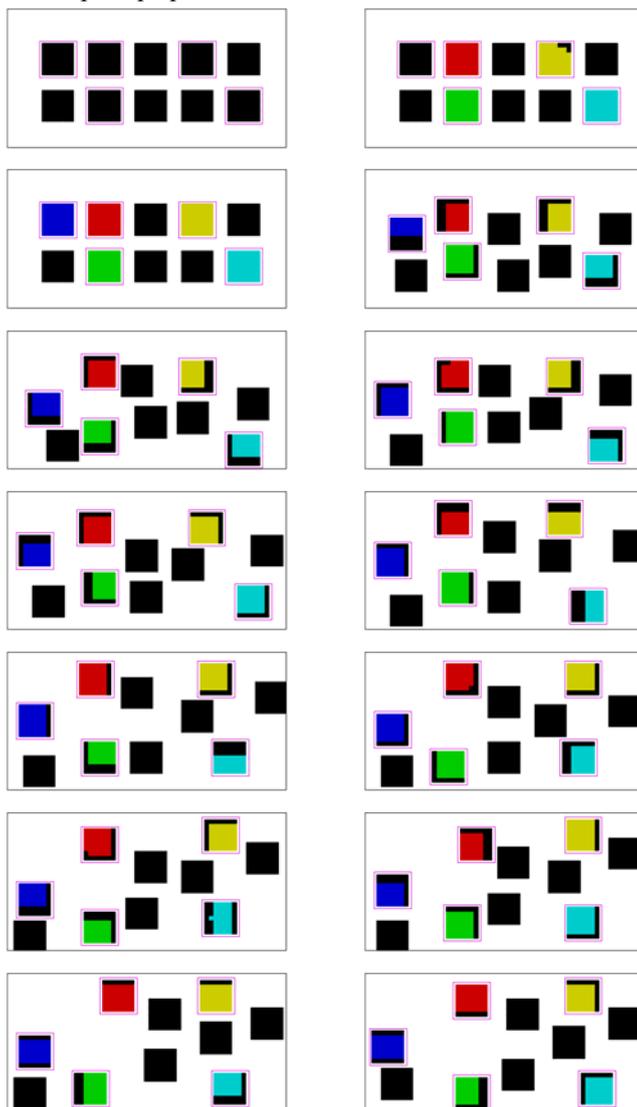
Третья фаза (тестирование) длится 0,5 сек. Во время этой фазы движение объектов прекращается. Эта фаза предназначена для того, чтобы дать системе время на принятие решения в ситуации, когда несколько объектов оказываются одновременно в фокусе внимания одной из подсистем внимания. Такая ситуация может время от времени возникать при движении объектов. Для стационарных объектов период времени 0.5 сек достаточен для того, чтобы подсистема внимания приняла решение, какой из этих объектов должен быть оставлен в фокусе внимания. Остальные объекты автоматически из фокуса внимания исключаются. Это достигается за счет уже упоминавшегося десинхронизирующего воздействия ЦО на ПО.

В финальный момент фазы тестирования происходит регистрация числа ошибок идентификации объектов как целевых и дистракторов. В соответствии с принятыми в данной модели принципами функционирования в этот момент в фокус внимания оказываются включенными ровно  $k$  квадратов, где  $k$  – число слоев в сети. Ошибки идентификации могут быть двух типов: целевой объект идентифицируется как дистрактор или дистрактор идентифицируется как целевой объект. В соответствии со стратегией, реализованной в модели, число ошибок всегда четно: если имеет место ошибка, когда целевой объект не попадает в фокус внимания, это неизбежно влечет включение в фокус внимания какого-то дистрактора.

На рис.3 показаны кадры из фильма, иллюстрирующего динамику формирования и переключения фокуса внимания при слежении за пятью целевыми объектами при наличии пяти дистракторов. Кадры упорядочены слева направо и сверху вниз. Временной интервал между кадрами 0,3 сек. Целевые объекты показаны с обведенной вокруг них сиреневой рамочкой. Объекты, находящиеся в текущий момент в фокусе внимания, покрашены в различные цвета. Цвет пикселя в каждом кадре зависит от состояния ПО в колонке, соответствующей этому пикселю. Если ПО в "красном" слое находится в резонансном состоянии, тогда пиксель окрашен в красный цвет. Аналогичный принцип используется для остальных цветов, кроме черного и белого. В черный цвет покрашены пиксели, находящиеся вне фокуса внимания. В белый цвет покрашены пиксели фона.

Левая верхняя рамка соответствует начальному моменту, когда в фокусе внимания еще нет объектов. Фокус внимания, включающий пять целевых объектов формируется на первом кадре второй строки. Второй кадр второй строки соответствует началу фазы слежения. Первоначально модель удерживает в фокусе внимания все изначально заданные целевые объекты, но в конце фазы

слежения происходит ошибочное переключение "зеленой" подсистемы внимания с целевого объекта на один из объектов дистракторов. Позднее эта подсистема внимания захватывает два объекта-дистрактора в фокус внимания, но на этапе тестирования (последний кадр в последней строке) в фокусе внимания этой подсистемы остается один дистрактор. В результате число ошибок идентификации в этом примере равно 2.



Нейросетевая модель слежения за несколькими объектами

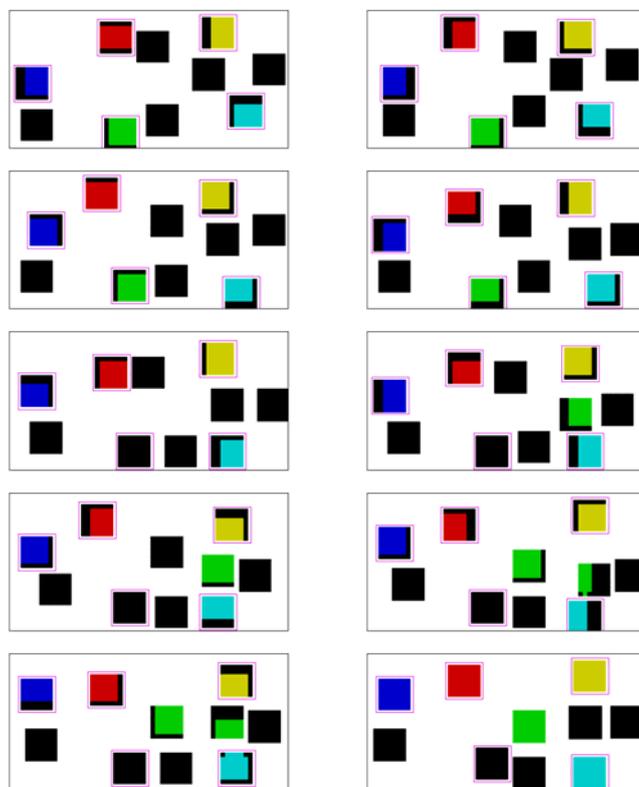


Рис. 3. Слежение за несколькими объектами при запрете на наложение объектов. Изображение содержит 5 целевых объектов и 5 дистракторов. Пиксели объектов, включенные в фокус внимания, окрашены в различные цвета, соответствующие различным подсистемам внимания. Пиксель имеет черный цвет, если ни один из осцилляторов в колонке не находится в резонансном состоянии. Пиксель имеет белый цвет, если он принадлежит фону.

Для оценки качества функционирования модели в терминах числа ошибок идентификации для каждого значения числа целевых объектов  $k = 2, 3, 4, 5$  было проделано по 50 рабочих циклов модели. Результаты вычислений показаны в таблице 1.

Таблица 1. Результаты идентификации целевых объектов и дистракторов моделью СНО

	Число целевых объектов			
	2	3	4	5
Общее число ошибок	2	38	86	134
Среднее число ошибок на рабочий цикл	0.04	0.76	1.72	2.68
Стандартное отклонение	0.3	1.1	1.5	1.6
Вероятность ошибки идентификации пробного объекта	0.006	0.09	0.179	0.268

Для проверки статистической значимости отличия средних значений числа ошибок в четырех группах, соответствующих четырем значениям  $k$ , был использован тест анализа дисперсии (ANOVA). Проверяемая гипотеза состояла в том, что все группы (каждая из которых содержит по 50 чисел) были получены из одного вероятностного распределения (или из разных распределений, но имеющих одно и то же среднее). Полученный результат:  $F = 43,7$  и  $p < 0,0001$ . Это означает, что гипотеза должна быть отвергнута.

Дальнейший анализ с помощью парного Т-теста дал следующие результаты:  $T_{23} = 4.6$ ,  $T_{34} = 3.7$ ,  $T_{45} = 3.1$ . Из этого следует, что гипотезы о том, что среднее в группе  $k$  равно среднему в группе  $(k-1)$  для  $k = 3, 4, 5$  должны быть отвергнуты.

В экспериментах [46] результат СНО для человека оценивался с помощью пробного объекта, который подсвечивался в финальный момент каждого опыта и должен был быть идентифицирован испытуемым как целевой или дистрактор. Чтобы исключить возможность угадывания на основе учета соотношения между числом целевых объектов и дистракторов, пробный объект выбирался из набора целевых объектов или из набора дистракторов с вероятностью 0,5. В наших экспериментах пробные объекты не использовались, однако наши данные относительно числа ошибок в каждом рабочем цикле дают возможность вычислить вероятность ошибки при использовании пробного объекта.

Пусть  $s$  – число объектов ( $s = 10$ ),  $k$  – число целевых объектов ( $k = 2, 3, 4, 5$ ),  $e$  – число целевых объектов, которые были ошибочно идентифицированы как дистракторы в одном рабочем цикле (следовательно, то же число дистракторов было ошибочно идентифицировано как целевые объекты). Тогда вероятность ошибки идентификации при использовании пробного объекта будет

$$P = 0.5 \left( \frac{e}{k} + \frac{e}{s-k} \right) = \frac{0.5se}{k(s-k)}.$$

Используя эту формулу, мы вычислили значения  $P$  для каждого рабочего цикла и затем усреднили эти значения по всем рабочим циклам. Результаты представлены в последней строке таблицы 1 и на рис. 4. Для сравнения на рис. 4 показаны также данные о качестве СНО для людей, приведенные в [46].

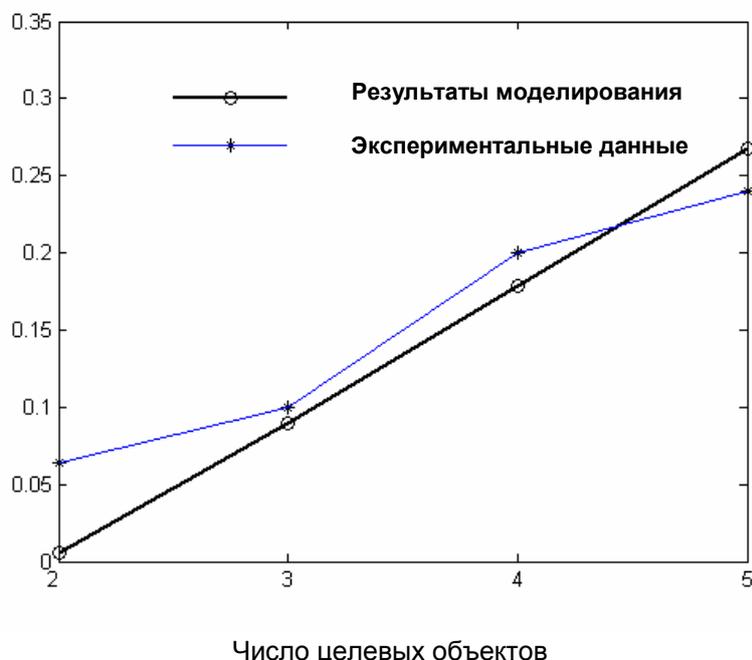


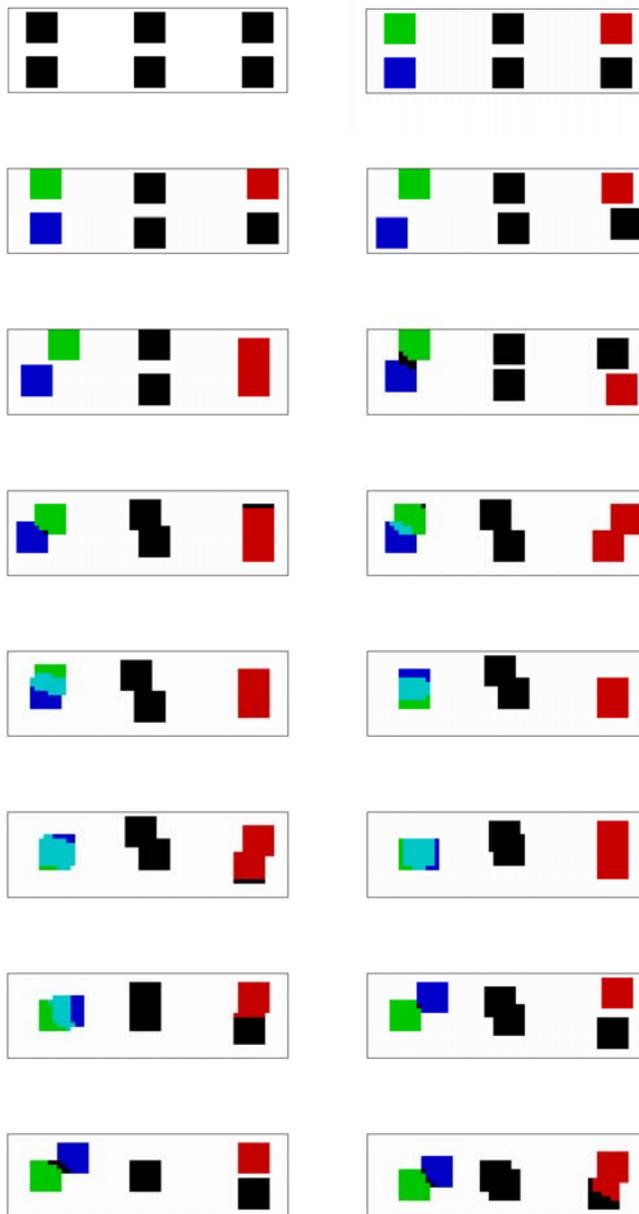
Рис. 4. Вероятность ошибки при тестировании пробного объекта в зависимости от числа целевых объектов. Экспериментальные данные взяты из работы [46].

Оба графика на рис. 4 имеют сходное поведение: вероятность ошибки возрастает с увеличением числа целевых объектов. Основное отличие между графиками имеет место для случая двух целевых объектов. В этом случае модель дает существенно меньшую вероятность ошибки, чем эксперимент.

Проиллюстрируем теперь функционирование модели в случае, когда допускается пересечение объектов при их движении. В качестве примера рассмотрим слежение за 6 объектами (3 целевых объекта и 3 дистрактора). Как и прежде, объектами являются черные квадраты размером 7 x 7 пикселей, расположенные на белом фоне. Размер поля 19 x 62 пикселей.

Рабочий цикл модели разделен на две фазы. Фаза экспозиции (1,2 сек) используется для указания целевых объектов и выбора их в фокус внимания. В этой фазе все квадраты представляют собой неподвижные изолированные объекты. Выбор целевых объектов осуществляется так же, как это было описано выше.

В фазе слежения, которая длится 9,6 сек, объекты движутся аналогично тому, как это было описано выше, но без ограничений на пересечение объектов. На рис. 5 показаны кадры из фильма, иллюстрирующего динамику формирования и переключения фокуса внимания.



Нейросетевая модель слежения за несколькими объектами

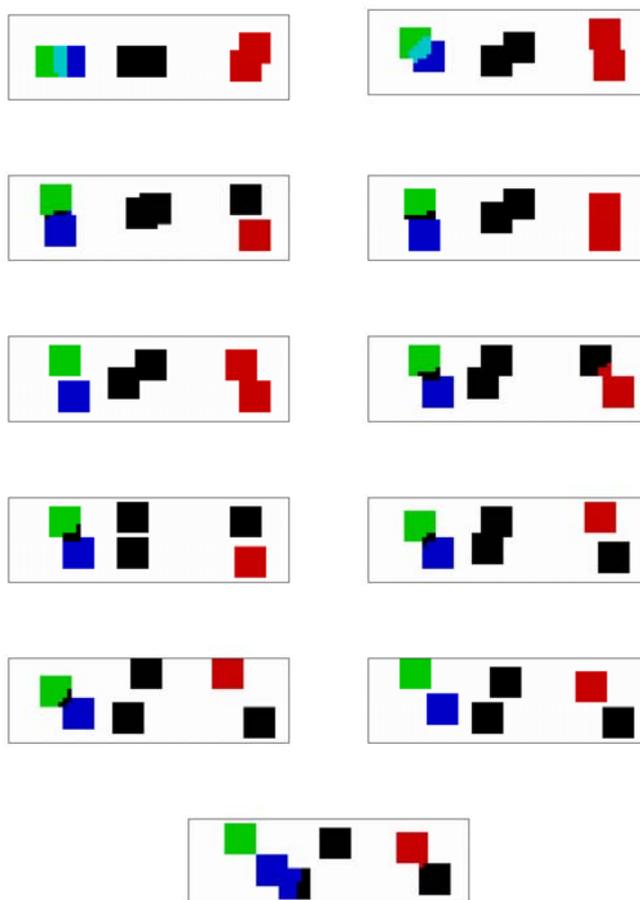


Рис. 5. Слежение за несколькими объектами при разрешении наложения объектов. Изображение содержит 3 целевых объекта и 3 дистрактора. Каждый пиксель имеет один из цветов: красный, синий, зеленый или смесь этих цветов. Пиксель имеет красный/зеленый/синий цвет, если представляющий этот пиксель ПО в "красном"/"зеленом"/"синем" слое находится в резонансном состоянии. Пиксель имеет голубой цвет, если оба осциллятора в "синем" и "зеленом" слоях находятся в резонансном состоянии. Пиксель имеет черный цвет, если ни один из осцилляторов в колонке не находится в резонансном состоянии. Пиксель имеет белый цвет, если он принадлежит фону.

Кадры упорядочены слева направо и сверху вниз. Временной интервал между кадрами 0,4 сек. Кадр в левом верхнем углу показывает начальное положение квадратов на изображении. На следующем кадре показан момент, когда формируется фокус внимания, содержащий 3 целевых объекта. Первый кадр во второй строке кадров соответствует началу фазы движения объектов. В даль-

нейшем это движение приводит к образованию различных комбинаций пересекающихся объектов.

Как и на рис. 3 цвета на рис. 5 это не цвета, в которые окрашены объекты (все объекты на входном изображении черные). Цвета используются для того, чтобы показать, какие объекты находятся в фокусе внимания той или иной подсистемы внимания. Цвет пикселя в каждом кадре зависит от состояния ПО в колонке, соответствующей этому пикселю. Если ПО в "красном" слое находится в резонансном состоянии, тогда пиксель окрашен в красный цвет. Аналогично для обозначения резонансных осцилляторов используются зеленый и синий цвета. Если в колонке несколько осцилляторов находятся в резонансном состоянии, то цвет пикселя формируется из смеси основных цветов. Например, пиксель имеет голубой цвет, если он одновременно включен в фокус внимания "зеленой" и "синей" подсистем внимания. Черный цвет используется для пикселей объектов вне фокуса внимания, белый цвет используется для пикселей фона. (Интенсивности зеленого, красного и синего подобраны так, чтобы их комбинация в одном пикселе давала серый цвет. На самом деле, на рис. 4 таких пикселей нет).

Рассмотрим, что происходит с фокусом внимания в процессе перемещения объектов.

Пара квадратов, расположенных в центре изображения, всегда находится вне фокуса внимания независимо от того, являются ли эти квадраты изолированными объектами или временно формируют сложный объект. ПО, представляющие пиксели этих объектов, всегда работают с низкой амплитудой.

Пара квадратов в правой части изображения показывает процесс переключения внимания в случае, когда сложный объект образуется в результате пересечения двух объектов, один из которых находится в фокусе внимания, а другой нет. Если время существования сложного объекта невелико, то после расхождения квадратов внимание остается на том же объекте, на котором оно было сфокусировано до наложения объектов. Если область пересечения объектов достаточно велика и время существования сложного объекта достаточно продолжительное, то это приводит к тому, что внимание распространяется на весь сложный объект. После того как квадраты снова разойдутся, любой из них может остаться в фокусе внимания, тогда как другой квадрат будет исключен из фокуса внимания.

Наконец, рассмотрим пару квадратов в левой части изображения. Эти квадраты оба были указаны как целевые объекты и вследствие этого они были выбраны в фокус внимания двух подсистем внимания "зеленой" и "синей". До тех пор пока эти квадраты не пересекаются, их движение устойчиво отслеживается этими подсистемами. Ситуация меняется после того, как возникает сложный объект с достаточно большой областью пересечения этих квадратов. Эта область на рисунке окрашена в голубой цвет, поскольку ее пиксели одновременно находятся в фокусе внимания обеих подсистем, "зеленой" и "синей". В процессе перемещения объектов размер этой голубой области меняется, она может увеличиваться или уменьшаться, исчезать и снова появляться. Существенно лишь то, что после того как квадраты разойдутся, движение каждого из них будет вновь отслеживаться одной из подсистем внимания. Обмен целевых

объектов между подсистемами внимания может произойти тогда, когда область пересечения квадратов в сложном объекте достаточно велика и не позволяет надежно идентифицировать каждый из квадратов после того, как они разойдутся.

Количественное исследование качества функционирования модели СНО в случае пересекающихся объектов является предметом дальнейших исследований. Предварительный опыт имитационных вычислений показал, что вероятность ошибок существенно зависит от скорости перемещения объектов. Для невысоких скоростей (именно такие скорости использовались в приведенном выше примере), вероятность ошибок (за исключением тех, которые неизбежны при пересечении целевых объектов и дистракторов) достаточно мала.

#### 4. Обсуждение

Представленная в данной работе модель внимания основана на трех основных идеях: Во-первых, в качестве ключевых механизмов формирования фокуса внимания нами используются колебания и их синхронизация. Экспериментальное подтверждение того, что осцилляторная активность и фазовая синхронизация имеют прямое отношение к феномену селективного зрительного внимания, было получено в работах по регистрации электроэнцефалограмм [26], магнитных энцефалограмм [24, 58], и локальных полевых потенциалов [21, 22, 23, 39, 44, 60]. В частности было показано, что модуляция синхронизации нейронной активности определяет исход при решении поведенческой задачи, требующей привлечения внимания [24, 61].

Во-вторых, в модели используется гипотеза о существовании специальной нейронной системы, центрального управляющего элемента координирующего фокусировку внимания. Теоретические и экспериментальные соображения в пользу этой гипотезы приведены в [5-7, 14, 56]. В терминах модели зрительное внимание обычно характеризуется как следствие синхронизации активности центрального элемента и связанного ансамбля нейронов в первичных зонах зрительной коры. СНО дает пример более сложной ситуации, когда внимание распределено между несколькими изолированными объектами. Мы предполагаем, что в этом случае центральный элемент как бы разбивается на несколько подсистем, работающих асинхронно. Синхронные же колебания используются в качестве "метки", позволяющей различать объекты, одновременно включенные в фокус внимания.

В третьих, в модели используется резонанс. Его назначение – формализовать гипотезу об увеличении уровня нейронной активности в областях мозга, включенных в фокус внимания, и подавить до низкого уровня активность, возникающую в ответ на стимулы вне фокуса внимания.

Связь между резонансом в модели и модуляцией кортикальной активности при внимании следует прояснить более подробно. Нейрофизиологические [43, 51] и томографические исследования [31, 55, 59] показали, что связанная с вниманием модуляция может проявляться уже на уровне первичных зон зрительной коры, однако проявление эффектов внимания увеличивается при движении вверх по иерархии кортикальных структур [62]. Более того, в работе [15]

функциональная магнитная томография была применена для исследования уровня нейронной активности, возникающей в экспериментах с фокусировкой внимания на движущихся объектах. Авторы не обнаружили какого-либо увеличения активности в первичной зрительной коре, но в теменных и фронтальных зонах коры активность по меньшей мере удваивалась. Как привести эти данные в соответствие с динамикой модели СНО?

В модели были использованы непосредственные связи от ПО к ЦО. Это, конечно, является сильным упрощением реальной ситуации. На самом деле, путь от первичных зон зрительной коры до областей, занятых центральным управляющим элементом проходит через многие промежуточные кортикальные структуры. К моменту, когда поток информации достигает верхних уровней коры, разница в активности между тем, что находится или не находится в фокусе внимания, становится вполне очевидной. И все же эта разница не так велика, как та, что имеет место в модели для резонансных и нерезонансных колебаний. Поэтому не следует воспринимать амплитуду колебаний в модели как реальное отражение наблюдаемой в эксперименте кортикальной активности. Амплитуду колебаний в модели следует скорее рассматривать как формальную характеристику, положительно коррелирующую с активностью в коре и определяющую силу взаимодействия между кортикальными осцилляторами и центральным управляющим элементом.

Теоретическая концепция, используемая для объяснения СНО, основана на идее индексации объектов, подлежащих слежению, на уровне предвнимания [49, 50]. Предполагается, что индексация происходит независимо и параллельно в нескольких участках зрительного поля. В отличие от этой теории в нашей модели индексация есть результат одновременной работы предвнимания и внимания. Предвнимание ответственно за то, чтобы присвоить каждому объекту метку в виде определенной колебательной активности. Вся информация об объекте кодируется в форме синхронных (синфазных) колебаний. Хотя такая метка в определенных пределах меняется во времени, она позволяет достаточно надежно отличать даже идентичные по своим признакам объекты. Различение объектов и выделение из них определенного подмножества, предназначенного для каких-то специальных процедур (в данном случае слежения) осуществляется системой внимания. В случае, когда в фокусе внимания должно находиться несколько объектов, система внимания функционирует как несколько взаимодействующих подсистем. Номер подсистемы, в фокусе внимания которой находится некоторый объект, может рассматриваться как индекс этого объекта.

Рассматриваемая модель СНО имеет жесткую архитектуру связей и априорно заданные и не изменяющиеся функции взаимодействия и параметры. Реальная биологическая система, реализующая СНО, несомненно, должна быть намного более гибкой. В частности, модель СНО с фиксированным числом модулей МВЦО не может адаптироваться к слежению за произвольно заданным числом целевых объектов. Конечно, можно предположить, что число активных ЦО может задаваться с помощью внутренней психологической установки и всегда соответствует числу целевых объектов. Нам, однако, представляется, что более правдоподобным решением является гибкая, зависящая от решаемой задачи, архитектура взаимодействий при формировании системы центральных

осцилляторов. Если необходимо следить за одним объектом, все центральные осцилляторы объединяются в один ансамбль с помощью синхронных взаимодействий. Если требуется следить за двумя объектами, то формируется две группы центральных осцилляторов, с синхронизирующими взаимодействиями между осцилляторами в группе и десинхронизирующими взаимодействиями между группами и т.д.

Зависимость вероятности ошибок в СНО от числа целевых объектов является причиной, по которой исследователи-экспериментаторы рассматривают СНО как пример параллельной обработки информации с ограниченными ресурсами [50]. Модель СНО показывает, чем может быть обусловлена эта ограниченность. Увеличение вероятности ошибок в модели связано с ограниченной емкостью фазового пространства, в котором одновременно вынуждено оперировать несколько ЦО. Увеличение числа ЦО приводит к тому, что им все труднее становится избегать синхронизации друг с другом, что приводит к непредсказуемым переключениям внимания на объекты-дистракторы. Пока число целевых объектов не превышает 5, вероятность таких переключений мала, если объекты стационарны, но эта вероятность существенно возрастает, когда объекты начинают двигаться с достаточно большой скоростью. Из-за перемещений объектов процессы синхронизации и резонанса не могут продвинуться достаточно далеко, что приводит к потере синхронизации ЦО с выбранным ранее в фокус внимания ансамблем ПО.

Демонстрируемое моделью в случае двух целевых объектов существенно лучшее качество слежения по сравнению с людьми, показывает, что в этом случае, вероятно, есть какие-то дополнительные факторы, которые делают данную задачу для людей достаточно сложной.

Хотя параметры модели были выбраны таким образом, чтобы временные соотношения приближались к тем, которые используются в экспериментах СНО, к этому факту не стоит относиться слишком серьезно. Модель слишком груба, чтобы в данном аспекте быть точной. Весьма небольшое число пикселей в объектах, ограничение взаимодействия ПО ближайшими соседями и многие другие особенности модели обусловлены необходимостью уложиться в разумное время при компьютерных вычислениях. На самом деле, модель довольно гибка в отношении воспроизведения различных временных показателей. Другой выбор параметров или даже просто изменение единицы времени позволяет увеличить или уменьшить длительность рабочего цикла. В связи с этим, при сравнении качества функционирования модели по сравнению с людьми мы использовали экспериментальные данные, усредненные по периодам времени 5, 9 и 13 сек. Данные для 5 сек в экспериментах с людьми давали меньшие вероятности ошибок, чем те, что были получены в наших вычислениях, но поведение этих вероятностей при увеличении числа целевых объектов остается тем же самым.

Использованная в модели стратегия принятия решений также является сильно упрощенной. Модель всегда старается следить за фиксированным и заданным заранее числом объектов. Эксперименты показывают, что стратегия людей более разумна [46, 50]. Если испытуемый чувствует, что правильная идентификация объекта в процессе слежения становится сомнительной, он

склонен прекратить слежение этого объекта и сосредоточить внимание на слежении меньшего числа объектов. Это может служить объяснением тому, что качество слежения для человека падает не резко, а постепенно, даже когда число целевых объектов превосходит 5. Этот же факт может быть причиной того, что качество слежения для человека при переходе от четырех целевых объектов к пяти падает меньше, чем в модели СНО.

Известно, что в экспериментах СНО качество слежения может быть улучшено, если испытуемый группирует целевые объекты в виртуальный многоугольник, а затем следит за деформациями этого многоугольника. [67]. Внешне слежение в этом случае выглядит весьма сходно, однако при его реализации могут быть задействованы более сложные когнитивные механизмы, которые, по меньшей мере, должны включать наличие у испытуемого представления о геометрических фигурах. Границы виртуального многоугольника являются воображаемыми линиями, которые формируются в результате внутреннего усилия. После этого многоугольник может рассматриваться как связный объект, фокусировка внимания на котором не требует разделенного внимания и может быть выполнена в рамках МВЦО. Трудность слежения в этом случае обусловлена, по-видимому, другими причинами, чем те, что рассмотрены в данной работе.

При разработке модели СНО мы старались исключить из операций, совершаемых моделью, традиционные процедуры обработки изображений, такие как анализ формы объектов, тестирование связности, распознавание и т.д. Благодаря этому модель может с равным успехом работать с неидентичными объектами или объектами, форма которых меняется во времени. Последнее важно, например, если объекты движутся в трехмерном пространстве, когда проекция их на сетчатку постоянно меняется. Другой причиной нашего подхода к моделированию СНО является тот факт, что перечисленные виды обработки зрительной информации требуют большего времени и требуют участия более высоких кортикальных структур. Нашей задачей было исследовать, может ли СНО быть объяснено в терминах простой нейронной сети, в которой единственная роль, выполняемая высокой структурой, состоит в том, чтобы направлять внимание на нужные объекты.

## Приложение 1. Математическое описание МВЦО

Осцилляторы, из которых строится сеть МВЦО, представляют собой обобщенные фазовые осцилляторы. Состояние такого осциллятора описывается тремя явно заданными переменными: фазой колебаний, амплитудой колебаний и собственной частотой осциллятора (осциллятор генерирует колебания с собственной частотой при отключении взаимодействия с другими осцилляторами сети).

Динамика МВЦО описывается следующими уравнениями

$$\frac{d\theta_0}{dt} = 2\pi\omega_0 + \frac{w_0}{n} \sum_{i=1}^n s_i a_i g(\theta_i - \theta_0), \quad (1.1)$$

Нейросетевая модель слежения за несколькими объектами

$$\frac{d\theta_i}{dt} = 2\pi\omega_i - a_0 w_1 h(\theta_0 - \theta_i) + w_2 \sum_{j \in N_i} a_j p(\theta_j - \theta_i) + \rho, \quad (1.2)$$

$$\frac{da_i}{dt} = \beta(-a_i + \gamma f(\theta_0 - \theta_i)), \quad (1.3)$$

$$\frac{d\omega_0}{dt} = -\alpha(2\pi\omega_0 - \frac{d\theta_0}{dt}). \quad (1.4)$$

В этих уравнениях  $\theta_0$  – фаза ЦО,  $\theta_i$  ( $i = 1, \dots, n$ ) – фазы ПО,  $\frac{d\theta_0}{dt}$  и  $\frac{d\theta_i}{dt}$  – текущие частоты осцилляторов,  $\omega_0$  – собственная частота ЦО,  $\omega_i$  – собственные частоты ПО,  $a_0$  – амплитуда колебаний ЦО (константа),  $a_i$  – амплитуды колебаний ПО,  $w_0, w_1, w_2$ , – положительные константы, задающие силу взаимодействия между осцилляторами,  $s_i$  – параметр, задающий различие между активными и молчащими ПО,  $s_i = 1$ , если ПО активен, в противном случае  $s_i = 0$ ,  $N_i$  – набор активных ПО в ближайшей окрестности осциллятора  $i$ ,  $\rho$  – гауссовский шум со средним 0 и стандартным отклонением  $\sigma$ , функции  $g, h, p$  задают взаимодействие между осцилляторами,  $f$  – функция, задающая амплитуду колебаний периферических осцилляторов и их переход в резонансное состояние,  $\alpha, \beta, \gamma$  – параметры (положительные константы). Значения  $\omega_i$  задаются внешними сигналами на ПО;  $\theta_0, \theta_i, \omega_0, a_i$  – внутренние переменные, задающие состояние сети.

Функции  $g, h, p$   $2\pi$  – периодические, нечетные, одномодальные в положительном полуинтервале периодичности, функция  $f$   $2\pi$  – периодическая, четная, положительная, одномодальная в интервале периодичности.

Приведем формулы для перечисленных функций, использовавшиеся при имитационных расчетах. На интервале  $(0, \pi)$  эти функции имеют вид

$$g(x) = \begin{cases} 10x, & 0 \leq x < 0.1, \\ -4x + 1.4, & 0.1 \leq x < 0.2, \\ -0.1x + 0.62, & 0.2 \leq x \leq \pi, \end{cases}$$

$$h(x) = 4.8x \exp(-4.8x + 1),$$

$$p(x) = \sin x,$$

$$f(x) = S((\cos x)^+),$$

где

$$y^+ = \begin{cases} y, & y \geq 0, \\ 0, & y < 0, \end{cases}$$

$$S(z) = \frac{\exp((z - 0.9)/0.02)}{1 + \exp((z - 0.9)/0.02)} + 0.1.$$

На интервал  $(-\pi, 0)$  функции  $g, h$  распространяются как нечетные, а функция  $f$  как четная.

Уравнения (1.1)–(1.2) – традиционные уравнения фазовой синхронизации, соответствующие архитектуре связей, приведенной на рис. 1. Эти уравнения описывают процессы синхронизации/десинхронизации в сети. Уравнение (1.1) задает динамику ЦО. Уравнение (1.2) задает динамику периферических осцилляторов. Шум  $\rho$  в (1.2) используется как дополнительный источник десинхронизации между синхронными ансамблями ПО. Шум помогает рандомизировать расположение различных ансамблей ПО в фазово-частотном пространстве, благодаря чему ЦО способен эти ансамбли различить.

Уравнение (1.3) описывает динамику амплитуд ПО. Интервал вариации функции  $f - (0, 1; 1, 1)$ . Поэтому амплитуда колебаний ПО возрастает до максимального значения  $a_{\max} = 1,1\gamma$ , если ПО работает синхронно (синфазно) с ЦО, и падает до низких значений  $a_{\min} = 0,1\gamma$ , если фаза ПО существенно отличается от фазы ЦО. Считается, что осциллятор находится в резонансном состоянии, если амплитуда колебаний превосходит порог  $R = 0,8a_{\max}$ . Параметр  $\beta$  задает скорость изменения амплитуды.

Уравнение (1.4) описывает адаптацию собственной частоты ЦО. В соответствии с этим уравнением значение  $2\pi\omega_0$  стремится к текущей частоте ЦО. Благодаря этой адаптации ЦО способен "искать" в фазово-частотном пространстве ансамбль ПО, наиболее подходящий для синхронизации

## Приложение 2. Математическое описание модели СНО

Уравнения для динамики модели СНО являются модификацией уравнений (1.1)–(1.4) применительно к архитектуре связей, представленной на рис. 2,

$$\frac{d\theta_0^k}{dt} = 2\pi\omega_0^k + \frac{w_0}{n_{res}} \sum_{i=1}^n s_i a_i^k g(\theta_j^k - \theta_0^k) - w_3 \sum_{l=1}^m a_0^l h(\theta_0^l - \theta_0^k), \quad (2.1)$$

$$\frac{d\theta_i^k}{dt} = 2\pi\omega_i^k - a_0^k w_1 h(\theta_0^k - \theta_i^k) + w_2 \sum_{j \in N_i} a_j^k p(\theta_j^k - \theta_i^k) + \frac{w_4}{m} \sum_{l=1}^m a_l^l p(\theta_i^l - \theta_i^k) + \rho, \quad (2.2)$$

$$\frac{da_i^k}{dt} = \beta(-a_i^k + \gamma f(\theta_0^k - \theta_i^k) + \zeta), \quad (2.3)$$

$$\frac{d\omega_0^k}{dt} = -\alpha(2\pi\omega_0^k - \frac{d\theta_0^k}{dt}). \quad (2.4)$$

В этих уравнениях верхний и нижний индексы нумеруют, соответственно, слои и осцилляторы в слое;  $k = 1, \dots, m$ , где  $m$  – число слоев. Нормирующий

параметр  $n_{res}$  равен числу резонансных осцилляторов, но не меньше, чем 49 (число пикселей в одном объекте). Последний член в (2.1) описывает взаимодействие между центральными осцилляторами (знак минус перед этим членом указывает на то, что взаимодействие десинхронизирующее). Член, предшествующий шуму  $\rho$  в (2.2), описывает взаимодействие ПО внутри колонки. Параметры  $w_3, w_4$  – положительные константы.

Параметры  $s_i$  в уравнении (2.1) формируют карту заметности,  $s_i > 0$  для пикселей объектов,  $s_i = 0$  для пикселей фона. В фазе экспозиции значения  $s_i = 5$  для пикселей целевых объектов,  $s_i = 0,2$  для пикселей дистракторов. В фазе слежения и тестирования все  $s_i = 1$ .

Уравнения (2.3)-(2.4) являются обобщением уравнений (1.3)-(1.4) для случая многослойной сети.

Амплитуды ЦО меняются в соответствии с уравнением, сходным с уравнением (2.3),

$$\frac{da_0^k}{dt} = \beta(-a_0^k + \gamma_1 r(\sum_{l=1, l \neq k}^m f(\theta_0^l - \theta_0^k))) + \zeta, \quad (2.5)$$

где  $\zeta$  - положительная константа,

$$r(x) = \begin{cases} x, & x \leq 1, \\ 1, & x > 1. \end{cases}$$

Функция  $r$  необходима для стандартизации интервала вариации амплитуд ЦО, который в данном случае равен  $(0, 1\gamma_1 + \zeta, \gamma_1 + \zeta)$ .

Параметры уравнений (2.1)-(2.5) и начальные значения переменных приведены в таблицах 2 и 3.

Таблица 2. Параметры, использовавшиеся при имитационном моделировании.

Параметры	Значения	Параметры	Значения
$w_0$	5	$w_2$	40
$w_3$	-12 (при экспозиции) -7 (при слежении и тестировании)	$w_4$	40
$\alpha$	2	$\beta$	1
$\gamma$	10	$\gamma_1$	1,7
$\zeta$	1	$\rho$	3

Таблица 3. Начальные значения переменных.

Переменные	Начальные значения	Переменные	Начальные значения
$\theta_0^k, \theta_i^k$	Равномерно распределены в $(0, \pi)$	$\omega_0^k$	6
$a_0^k$	10	$a_i^k$	2

## Литература

1. Казанович Я.Б., Борисюк Р.М. (2005) Осцилляторная нейросетевая модель слежения за движущимся объектом. Нейроинформатика-2005. VII Всероссийская научно-техническая конференция, т. 1. Москва, МИФИ. с. 25-35.
2. Казанович Я.Б., Шматченко В.В. (2004). Осцилляторные нейросетевые модели сегментации изображений и зрительного внимания. Нейроинформатика-2004. VI Всероссийская научно-техническая конференция. Лекции по нейроинформатике, т. 1, под ред. Ю.В. Тюменцева. Москва: МИФИ, с. 72-125.
3. Крюков В.И. (2004). Модель внимания и памяти, основанная на принципе доминанты и компараторной функции гиппокампа. Журнал высшей нервной деятельности, 54(1), 10-29.
4. Andres P. (2003). Frontal cortex as the central executive: time to revise our view. Cortex, 39, 871-895.
5. Baddeley A. (1996). Exploring the central executive. Quarterly Journal of Experimental Psychology, 49A, 5-28.
6. Baddeley A. (2002). Fractionating the central executive. In D. Stuss and R. T. Knight (Eds.), Principles of Frontal Lobe Function (pp. 246–260). New York: Oxford University Press.
7. Baddeley A. (2003). Working memory and language: an overview. Journal of Communication Disorders, 36, 189-208.

8. Barbas H. (2000). Connections underlying the synthesis of cognition, memory, and emotion in primate prefrontal cortices. *Brain Research Bulletin*, 52, 319–330.
9. Blaser E., Pylyshyn Z.W., Holcombe A.O. (2000). Tracking an object through feature space. *Nature*, 408, 196-199.
10. Borisyuk R., Kazanovich Y. (2003). Oscillatory neural network model of attention-focus formation and control. *BioSystems*, 71, 29-36.
11. Borisyuk R., Kazanovich Y. (2004). Oscillatory model of attention-guided object selection and novelty detection. *Neural Networks*, 17, 899-915.
12. Collett F., Van der Linden M. (2002). Brain imaging of the central executive component of the working memory. *Neuroscience and Behavioral Review*, 26, 105-125.
13. Corchs S., Deco G. (2001). A neurodynamical model for selective visual attention using oscillators. *Neural Networks*, 14, 981-990.
14. Cowan N. (1988). Evolving conceptions of memory storage, selective attention and their mutual constraints within the human information processing system. *Psychological Bulletin*, 104, 163-191.
15. Culham J., Brandt S.A., Cavanagh P., Kanwisher N.G., Dale A.M., Tootell R. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, 80, 2657-2670.
16. Damasio A. (1989). The brain binds entities and events by multiregional activation from convergent zones. *Neural Computation*, 1, 123-132.
17. D'Esposito M., Detre J.A., Alsop D.C., Shin R.R., Atlas S., Grossman M. (1991). The neural basis of central executive system of working memory. *Nature*, 378, 279-281.
18. Duncan J. (2001). An adaptive coding model of neural functions in prefrontal cortex. *Nature Reviews Neuroscience*, 2, 820-829.
19. Egeth H., Yantis S. (1997). Visual attention: control, representation, and time course. *Annual Review of Psychology*, 48, 269-297.
20. Eriksen C.W., St. James J.D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Perception and Psychophysics*, 40, 225-240.
21. Fell J., Fernandez G., Klaver P., Elger C.E., Fries P. (2003). Is synchronized neuronal gamma activity relevant for selective attention? *Brain Research Reviews*, 42, 265-272.
22. Fries P., Reynolds J.H., Rorie A.E., Desimone R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291, 1560-1563.
23. Fries P., Schroeder J-H., Roelfsema P.R., Singer W., Engel A.K. (2002). Oscillatory neural synchronization in primary visual cortex as a correlate of stimulus selection. *Journal of Neuroscience*, 22, 3739-3754.
24. Gross J., Scmitz F., Schnitzler I., Kessler K., Shapiro K., Hommel B., Schnitzler A. (2004). Modulation of long-range neuronal synchrony reflects temporal limitations of visual attention in humans. *Proc. Natl. Acad. Sci. (USA)*, 101, 13050-13055.
25. Grossberg S., Raizada R. (2000). Contrast sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40, 1413-1432.
26. Herrmann C.S., Knight R.T. (2000). Mechanisms of human attention: event related potentials and oscillations. *Neuroscience and Biobehavioral Reviews*, 25, 465-476.
27. Holscher C. (2003). Time space and hippocampal functions. *Review of Neuroscience*, 14, 253-284.
28. Itti L., Koch C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489-1506.
29. Itti L., Koch C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2, 194-203.
30. Itti L., Koch C., Niebur E. (1998). A model of saliency-based visual attention for rapid screen analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 1254-1259.

31. Kanwisher N., Wojciulik E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews Neuroscience*, 1, 91-100.
32. Katayama K., Yano M., Horiguchi T. (2004). Neural network model of selective visual attention using Hodgkin-Huxley equation. *Biological Cybernetics*, 91, 315-325.
33. Kazanovich Y.B., Borisjuk R.M. (1999). Dynamics of neural networks with a central element. *Neural Networks*, 12, 441-454.
34. Kazanovich Y., Borisjuk R. (2002). Object selection by an oscillatory neural network. *BioSystems*, 67, 103-111.
35. Kazanovich Y.B., Borisjuk R.M. (2003). Synchronization in oscillator systems with phase shifts. *Progress in Theoretical Physics*, 110, 1047-1058.
36. Koch C., Ullman S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4, 219-227.
37. Kryukov V.I. (1991). An attention model based on the principle of dominance. In A.V. Holden and V.I. Kryukov (Eds.), *Neurocomputers and Attention I. Neurobiology, Synchronization and Chaos* (pp. 319-352). Manchester: Manchester University Press.
38. Kuramoto Y. (1991). Collective synchronization of pulse coupled oscillators and excitable units. *Physica D*, 50, 15-30.
39. Liang H., Bressler S.L., Ding M., Desimone R., Fries P. (2003). Temporal dynamics of attention-modulated neuronal synchronization in macaque V4. *Neurocomputing*, 52-54, 481-487.
40. Liu G., Austen E.L., Booth K.S., Fisher B.D., Argue R., Rempel M.I., Enns J.T. (2005). Multiple-object tracking is based on scene, not retinal coordinates. *Journal of Experimental Psychology*, 31, 235-247.
41. Loose R., Kaufmann C., Auer D.P., Lange K.W. (2003). Human prefrontal and sensory cortical activity during divided attention tasks. *Human Brain Mapping*, 18, 249-59.
42. Moser M.C., Siltan M. (1998). Computational model of spatial attention. In H. Pashler (Ed.), *Attention* (pp. 341-393). London: UCL Press.
43. Motter B.C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, 70, 909-919.
44. Niebur E., Hsiao S.S., Johnson K.O. (2002). Synchrony: a neuronal mechanism for attentional selection? *Current Opinion in Neurobiology*, 12, 190-194.
45. Niebur E., Kammen D.E., Koch C. (1991). Phase-locking in 1-D and 2-D networks of oscillating neurons. In W. Singer and H. Schuster (Eds.) *Nonlinear Dynamics and Neuronal Networks*, (pp. 173-204). Vieweg Verlag, Berlin.
46. Oksama L., Hyönä J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition*, 11, 631-671.
47. Olshausen B.A., Anderson C.H., Van Essen D.C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13, 4700-4719.
48. Posner M.I., Snyder C.R.R., Davidson D.J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109, 160-174.
49. Pylyshyn Z.W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80, 127-158.
50. Pylyshyn Z. W., Storm R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, 3, 179-197.
51. Roelfsema P.R., Lamme V., Spekreijse H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, 395, 376-381.
52. Scholl B.J. (2001). Objects and attention: the state of the art. *Cognition*, 80, 1-46.

Нейросетевая модель слежения за несколькими объектами

53. Scholl B.J., Tremoulet P.D. (2000). Perception causality and animacy. *Trends in Cognitive Science*, 4, 299-309.
54. Sears C.R., Pylyshyn Z.W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology*, 54, 1-14.
55. Seifert A.E., Somers D.C., Dale A.M., Tootel R. (2003). Functional MRI studies of human visual motion perception: texture, luminance, attention and aftereffects. *Cerebral Cortex*, 13, 340-349.
56. Shallice T. (2002). Fractionation of the supervisory system. In D. T. Stuss and R. T. Knight (Eds.), *Principles of Frontal Lobe Function* (pp. 261-277). New York: Oxford University Press.
57. Singer W. (1999). Neuronal synchrony: A versatile code for the definition of relations. *Neuron*, 24, 49-65.
58. Sokolov A., Lutzenberger W., Pavlova M., Pressl H., Braun C., Birbauner N. (1999). Gamma-band MEG activity to coherent motion depends on task-driven attention. *Neuroreport*, 10, 1997-2000.
59. Somers, D.C., Dale A.M., Seifert A.E., Tootel R. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc. Natl. Acad. Sci. (USA)*, 96, 1663-1668.
60. Steinmetz P.N., Roy A., Fitzgerald P., Hsiao S.S., Johnson K.O., Niebur E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, 404, 187-190.
61. Tallon-Baudry C. (2004). Attention and awareness in synchrony. *Trends in Cognitive Sciences*, 8, 523-525.
62. Treue S. (2003). Visual attention: the where, what, how and why of salience. *Current Opinion in Neurobiology*, 13, 428-432.
63. Tsotsos J.K., Culhane S.M., Wai W.Y.K., Lai Y., Davis N., Nufl F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, 78, 507-545.
64. Vinogradova O.S. (2001). Hippocampus as comparator: role of the two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus*, 11, 578-598.
65. Visvanathan L., Mingolla E. (2002). Dynamics of attention in depth: evidence from multi-element tracking. *Perception*, 31, 1415-1437.
66. Wang D.L. (1999). Object selection based on oscillatory correlation. *Neural Networks*, 12, 579-592.
67. Yantis S. (1992). Multielement visual tracking: attention and perceptual organization. *Cognitive Psychology*, 24, 295-340.

Статья поступила 10 ноября 2005 г.

# Эволюционный подход к настройке и обучению искусственных нейронных сетей

Цой Ю.Р.<sup>1)</sup>, Спицын В.Г.<sup>2)</sup>

Томский политехнический университет, Томск, <sup>1)</sup> qai@mail.ru,  
<sup>2)</sup> spitsyn@ce.cctpu.edu.ru

**Аннотация.** В статье рассматривается применение эволюционных алгоритмов (ЭА) для настройки структуры и поиска весов связей искусственных нейронных сетей (ИНС) – нейроэволюционный (НЭ) подход. Дается общая характеристика НЭ подхода, формулируются его преимущества и недостатки. Рассматриваются проблемы разработки НЭ алгоритмов и возможные пути их решения. Обосновывается использование эволюционного поиска весов связей ИНС для решения ряда задач, использующих качественную оценку функционирования ИНС.

## Введение

Выбор топологии и настройка весов связей искусственной нейронной сети (ИНС) [1] являются одними из важнейших этапов при использовании нейросетевых технологий для решения практических задач. От этих этапов напрямую зависит качество (адекватность) полученной нейросетевой модели, системы управления и т.д. В данной статье в общем виде рассматривается нейроэволюционный (НЭ) подход, как один из возможных вариантов решения проблем настройки структуры и обучения ИНС.

Статья организована следующим образом. В первом разделе дан краткий обзор эволюционных алгоритмов. Во втором разделе в общем виде описывается нейроэволюционный подход. Вопросам эволюционного поиска топологии ИНС, настройки весов связей и одновременному решению обеих задач посвящены соответственно разделы 3, 4 и 5. Проблема генерации новых решений-кандидатов в результате использования оператора скрещивания обсуждается в разделе 6. В разделе 7 приводится общая характеристика задач, использующих качественную оценку ИНС, рассматриваются случаи, когда введение такой оценки позволяет упростить решение прикладных задач, а также расширить область применения методов нейроинформатики.

## 1. Краткий обзор эволюционных алгоритмов

Прежде чем перейти к описанию нейроэволюционного (НЭ) подхода кратко рассмотрим эволюционные вычисления (ЭВ), методы которых лежат в его основе. Эволюционные принципы наследственности, изменчивости и естественного отбора используются для решения как задач моделирования (например, модели молекулярной эволюции [2] и макроэволюционные модели в социологии и истории [3]), так и прикладных задач оптимизации. Идея использования механизмов эволюции для машинного обучения прослеживается и в хрестоматийном труде Алана Тьюринга «Могут ли машины мыслить?». Множество ал-

горитмов и методов, использующие для поиска решения эволюционные принципы, объединяют под общим названием *эволюционные вычисления (ЭВ)* или *эволюционные алгоритмы (ЭА)* [4] (соответственно, *evolutionary computation* и *evolutionary algorithms*). Выделяют следующие основные виды ЭА:

- генетический алгоритм (ГА) [5, 6];
- эволюционное программирование (ЭП) [7, 8];
- эволюционные стратегии (ЭС) [9, 10];
- генетическое программирование (ГП) [11, 12].

Здесь мы более подробно рассмотрим генетический алгоритм, как один из самых распространенных ЭА.

Идея ГА предложена Джоном Холландом в 60-х годах, а результаты первых исследований обобщены в его монографии «Адаптация в природных и искусственных системах» [5], а также в диссертации его аспиранта Кеннета Де Йонга [13]. Общая схема канонического ГА представлена на рис. 1. Необходимо отметить, что в настоящее время существует множество модификаций канонического ГА, предложенного в [5], и не все из них могут быть описаны схемой на рис.1.

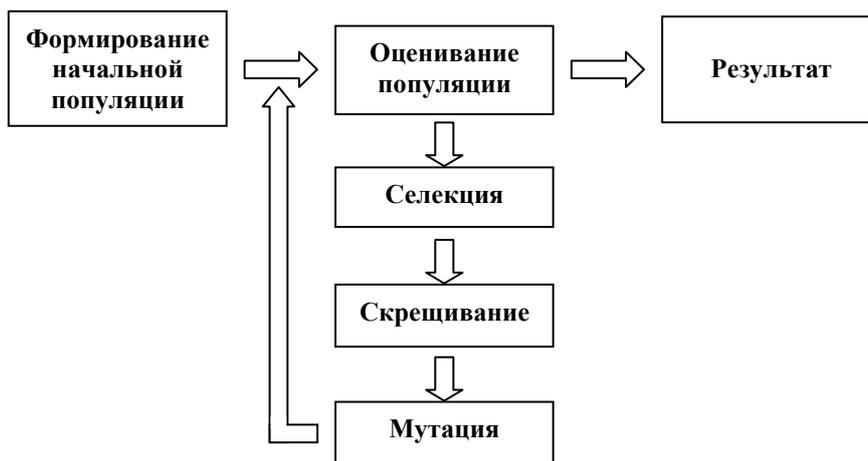


Рис. 1. Общая схема ГА. Комментарии к схеме в тексте.

ГА часто используются для решения оптимизационных задач, но, в отличие от большинства других алгоритмов (например, градиентные алгоритмы), на каждом этапе работы алгоритма рассматривается не одно потенциальное решение, а множество решений – *популяция*. Каждая *особь* популяции кодирует информацию о некотором решении в *хромосоме (генотипе)*, а декодированное из хромосомы решение иногда называют *фенотипом*.

В начале работы ГА популяция формируется, как правило, случайным образом (блок «**Формирование начальной популяции**» на рис. 1). Чтобы оценить качество закодированных решений (их *приспособленность*), используют *функцию приспособленности* (блок «**Оценивание популяции**»). По результатам оценивания особей наиболее приспособленные из них выбираются (блок «**Се-**

лекция») для скрещивания. В результате *скрещивания* выбранных особей посредством применения генетического оператора *кроссовера* создаются *потомки*, генетическая информация которых формируется в результате обмена хромосомной информацией между родительскими особями (блок «**Скрещивание**»). Созданные потомки формируют новую популяцию, причем часть потомков *мутирует* (с использованием генетического оператора *мутации*), что выражается в случайном изменении их генотипов (блок «**Мутация**»). Этап, включающий последовательность операций, обозначенных блоками «**Оценивание популяции**» – «**Селекция**» – «**Скрещивание**» – «**Мутация**», называется *поколением*. Эволюция популяции состоит из последовательности таких поколений.

Окончание эволюции может определяться следующими факторами:

- нахождение решения в результате эволюционного поиска;
- ограничение на число поколений, количество вычислений целевой функции, общее время работы ГА;
- вырождение популяции, когда степень разнородности хромосом в популяции становится меньше установленного значения<sup>1</sup>.

Как правило, в хромосоме кодируются параметры решения. Для этого используют целочисленное, либо вещественное кодирование. Определение приспособленности *i*-й особи в популяции производится в соответствии с оценкой решения, определяемого набором параметров, закодированных в хромосоме *i*-й особи, и зависит от рассматриваемой задачи. Важным при использовании ГА для решения прикладных задач является предположение, что в результате повторяющегося отбора наиболее приспособленных особей, их скрещивания и мутации будет производиться отсев неудовлетворительных решений и постепенное повышение качества (приспособленности) существующих «хороших» решений.

Упомянутые выше другие виды ЭА несколько отличаются от ГА. Так, основным отличием ЭС от ГА является значительно меньшая роль оператора кроссовера, а в ЭП кроссовер не используется совсем. Также ЭС, и ЭП изначально разрабатывались не только для целочисленного, но и для вещественного кодирования. ГП, в отличие от ГА, ЭС и ЭП, которые часто используются для решения оптимизационных задач, предназначено для решения задач автоматизированного программирования и машинного обучения. Более подробное описание перечисленных видов ЭА, а также их особенностей приведено, например, в [15] и [16].

## 2. Нейроэволюционный подход

Первые работы, посвященные применению ЭА для обучения и настройки ИНС, появились около 20 лет назад. Исследования в этой области, как правило, рассматривают решение следующих задач:

- поиск значений весов связей ИНС при фиксированной структуре;

---

<sup>1</sup> Степень разнородности хромосом в популяции может определяться различными способами в зависимости от выбранного варианта кодирования информации. Примеры числовых характеристик разнообразия хромосом в популяции [14]: суммарное расстояние по Хеммингу для всех пар особей в популяции; использование аналога момента инерции; суммарная энтропия по всем разрядам в хромосомах; использование концепции лингвистической сложности (linguistic complexity) и др.

Эволюционный подход к настройке и обучению искусственных нейронных сетей

- настройка структуры ИНС без поиска весов связей;
- настройка параметров алгоритма обучения;
- настройка параметров активационных функций нейронов;
- фильтрация обучающих данных;
- различные комбинации вышеперечисленных задач.

Предлагались различные варианты названий для данного направления исследований. В настоящей статье будет использоваться термин «нейроэволюция» (НЭ) [17], а соответствующий подход будем называть нейроэволюционным. Отметим, что предлагаемый вариант не является устоявшимся и окончательным.

Комбинация ИНС и ЭА дает возможность совместить гибкость настройки ИНС и адаптивность ЭА, что позволяет реализовать во многом унифицированный подход к решению широкого спектра задач классификации, аппроксимации и моделирования (см. раздел 7). Тем не менее, у такого синтеза есть и недостатки. К числу важнейших из них необходимо отнести:

1. Отсутствие, в общем случае, гарантии сходимости НЭ алгоритма. Данный недостаток обусловлен, в первую очередь, универсальностью схемы эволюционного поиска. При этом в условиях конечного времени нельзя гарантировать ни получение решения, ни качество этого решения. Пользователь задает условия поиска в виде целевой функции и параметров эволюционного поиска, но стохастичность работы ЭА часто вносит свои коррективы. Частичным решением этой проблемы является использование гибридного подхода. Например, типичным является разделение этапов глобального и локального поиска, где ЭА используется для поиска приближенного, грубого решения, которое затем «доводится до ума» с использованием более «точных» методов, например, градиентных алгоритмов оптимизации [18-23].

2. Вероятностный характер ЭА ведет также и к потере контроля над процессом поиска решения. Результат отдельных операций, осуществляемых в процессе выполнения ЭА, часто не несет смысловой нагрузки, поэтому «отследить логику» работы ЭА, как правило, не представляется возможным. Данный недостаток существенен при использовании ЭА и НЭ алгоритмов для синтеза правил логического вывода, где помимо самих правил имеет значение и цепочка «рассуждений», приведшая к их появлению.

Однако «логика» поиска не всегда является важным условием, а по результатам нескольких запусков часто можно оценить качество решений, получаемых с помощью НЭ алгоритма. Полученные приближенные оценки позволяют сделать заключение о применимости НЭ алгоритма для решения поставленной задачи и, если необходимо, внести поправки в процесс эволюционного поиска.

### **3. Эволюционная настройка весов связей ИНС**

Одной из типичных задач, решаемых в рамках НЭ подхода, является задача поиска весов связей ИНС при ее фиксированной структуре [3, 18-34]. При этом, как правило, рассматривается задача минимизации целевой функции (обычно функция ошибки выхода ИНС), а в качестве оптимизируемых параметров используются веса связей, значения которых подбираются с помощью ЭА. Одним из обоснований эволюционного обучения ИНС является «застывание» гради-

ентных алгоритмов в локальных экстремумах в процессе обучения ИНС. Кроме этого, как будет показано далее (см. раздел 7), эволюционный подход к поиску весов связей ИНС позволяет существенно упростить процесс решения ряда задач.

**3.1. Плюсы и минусы эволюционной настройки весов связей ИНС.** Существенными преимуществами эволюционной настройки весов связей ИНС являются:

1. Независимость от структуры ИНС и характеристик функций активации нейронов.
2. Отсутствие необходимости в обучающем множестве данных.

Первое преимущество дает возможность использовать единый подход к обучению ИНС с различной структурой.

Второе преимущество позволяет осуществлять обучение ИНС без информации об эталонных значениях выходных сигналов, а на основе оценки функционирования ИНС «в целом». Для градиентных алгоритмов оптимизации обучающее множество необходимо, чтобы подстраивать веса связей ИНС на основе расхождения значений реального и требуемого выходного сигнала ИНС. Подробнее этот вопрос обсуждается в разделе 7, здесь мы только отметим, что в случае эволюционной настройки весов связей ИНС появляется возможность использовать приближенные оценки функционирования ИНС, при этом оцениваться может не каждый выходной сигнал, а последовательность сигналов. Таким образом, упрощается поиск решения ряда задач.

К недостаткам использования ЭА для обучения ИНС следует отнести следующие:

1. Трудность «тонкой» настройки весов связей на поздних этапах эволюционного поиска.
2. Большие, по сравнению с градиентными алгоритмами, требования к объему оперативной памяти из-за использования популяции ИНС.
3. Проблема конкурирующих решений (см. подраздел 3.3).

Первая проблема во многом обусловлена использованием целочисленного кодирования весов связей ИНС, что может привести к значительным «скачкам» в пространстве поиска в результате применения операторов скрещивания и мутации. Данное свойство полезно на начальном этапе работы НЭ алгоритма, когда необходимо быстро исследовать пространство поиска, но затрудняет работу алгоритма на более поздних этапах.

Проблема конкурирующих решений заключается в том, что один и тот же набор весов связей может быть представлен в генетическом виде несколькими способами (см. подраздел 3.3).

**3.2. Кодирование информации о весах связей ИНС.** Как было сказано в начале текущего раздела, при эволюционном обучении ИНС, как правило, рассматривается задача минимизации целевой функции. Решение задачи часто заключается в отыскании такого набора значений весов связей ИНС, при которых функционирование ИНС оптимально с точки зрения выбранного критерия оценки.

Веса связей ИНС могут быть представлены с использованием как целочисленного, так и вещественного способов кодирования. Таким образом, хромосома каждой особи представляет вектор целочисленных, либо вещественных па-

раметров, соответствующих весам связей ИНС. Простые примеры кодирования весов связей показаны на рис. 2. В зависимости от выбранного варианта кодирования используются различные операторы скрещивания и мутации. Операторы для целочисленного кодирования достаточно подробно представлены в [6], для вещественного – в [35, 36]. Необходимо отметить, что использование нестандартных операторов, ориентированных на решение специфичных задач, способствует увеличению скорости работы и повышению качества результатов НЭ алгоритмов [37].

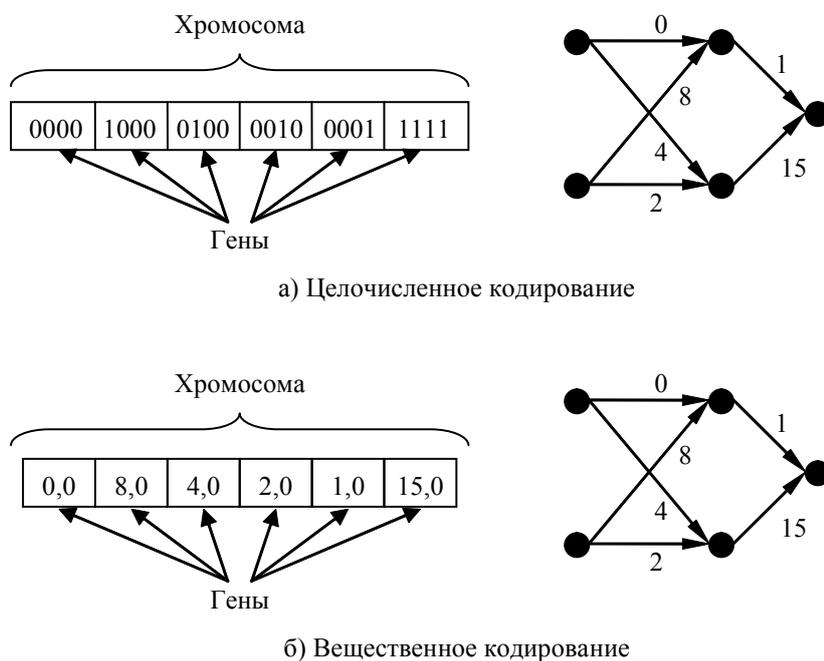


Рис. 2. Простые примеры целочисленного (а) и вещественного (б) кодирования весов связей для двух одинаковых ИНС. Здесь и на последующих рисунках кругами обозначены нейроны, стрелками показано направление связей, а числа возле стрелок соответствуют весам связей, если не указано иное.

**3.3. Проблема конкурирующих решений.** Одной из проблем, связанной с кодированием весов связей ИНС, является *проблема перестановки (permutation problem)* [28, 29], известная также как *проблема конкурирующих решений (competing conventions problem)* [38, 39]. Ее суть в том, что генетическое представление допускает существование нескольких вариантов хромосом, кодирующих одну и ту же ИНС (рис. 3).

Предложено несколько вариантов решения проблемы перестановки, среди которых необходимо отметить следующие:

1. Упорядочивание списка весов связей, кодируемых ИНС (выравнивание хромосом), по значениям весов [25], либо с использованием специальных меток [40]. Считается, что такие меры позволят уменьшить вероятность появления особей, представляющих одинаковые ИНС. Также возможно раннее удаление особей-дубликатов на этапе формирования популяции следующего поколения.

2. Разработка адаптивных операторов скрещивания [29, 41], уменьшение размера популяции, а также повышение вероятности мутации.

3. Уменьшение роли оператора кроссовера, либо полный отказ от его использования [42, 43]. При этом предполагается, что: (а) вероятность возникновения одинаковых ИНС в результате мутаций различных особей начальной популяции незначительна; (б) в случае, если такие особи все-таки появились, мутации приведут к меньшим «разрушениям», чем скрещивание.

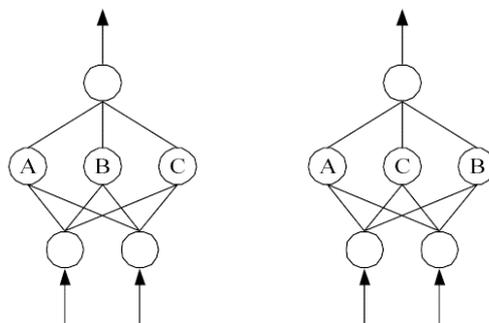


Рис. 3. Проблема перестановки. Две одинаковые ИНС могут быть закодированы различным образом «перестановкой» скрытых нейронов. Считается, что «скрещивание» таких сетей практически бесполезно (см. раздел 6) [40].

Отметим, что вопрос об актуальности проблемы конкурирующих решений остается открытым. Из-за того, что, в общем случае, вероятность появления в одной популяции двух одинаковых ИНС мала, проблема конкурирующих решений может игнорироваться либо вовсе не рассматриваться [17, 25, 44].

#### 4. Эволюционная настройка структуры ИНС

Еще одной типичной задачей НЭ подхода является задача эволюционной настройки структуры ИНС. В хромосоме кодируется топология ИНС, а обучение осуществляется, например, с использованием градиентных алгоритмов [45-47]. Каждая особь, представляющая сеть той или иной структуры, оценивается в зависимости от результатов обучения: чем лучше результат, тем более приспособлена особь. Поскольку выбор топологии ИНС является, как правило, сложной задачей, решаемой методом проб и ошибок, то эволюционный поиск нейросетевой структуры способен облегчить и в определенной степени автоматизировать процесс решения задачи настройки и обучения ИНС.

**4.1. Плюсы и минусы эволюционной настройки структуры ИНС.** Перечислим преимущества и недостатки эволюционной настройки структуры ИНС.

Преимущества:

1. Возможность автоматического поиска топологии ИНС и получения более точной нейросетевой модели за счет рассмотрения «нестандартных», нерегулярных топологий.

2. Независимость от характеристик функций активации нейронов.

Для упрощения задачи и повышения качества результатов в процессе поиска топологии ИНС возможно использование дополнительных регулирующих ограничений, помогающих избежать чрезмерного «разрастания» сети, которое выражается в быстром увеличении количества скрытых нейронов и связей между ними.

Недостатки:

1. Сложность оценки структуры ИНС без информации о значениях весов связей.

2. Сложность организации поиска топологии ИНС.

Первый недостаток представляет основную проблему эволюционной настройки структуры ИНС. Он, в основном, обусловлен чувствительностью результатов обучения к начальным условиям и значениям параметров алгоритма обучения. Таким образом, хромосоме, представляющей ИНС с некоторой структурой, могут соответствовать совершенно разные нейросетевые модели с точки зрения их качества и свойств (в таких случаях говорят, что одному генотипу соответствует множество фенотипов). Для уменьшения влияния случайных факторов на оценку структуры ИНС проводится несколько независимых операций обучения [45, 47], усредненный результат которых и используется в качестве оценки топологии ИНС. Также, несмотря на то, что существуют эвристические правила, согласно которым ИНС с более простой структурой обладают лучшими способностями к обобщению, этот вопрос все еще нельзя считать до конца решенным, т.к. многое зависит от результатов обучения [48, 49].

Сложность объективной оценки топологии ИНС порождает также сложность оценки отдельных структурных частей сети и возможных структурных модификаций. Данная проблема характерна также и для алгоритмов одновременной настройки и обучения ИНС (см. раздел 5), поэтому многие дальнейшие замечания относятся в равной степени и к эволюционному поиску структуры ИНС, и к одновременной настройке структуры и обучению ИНС. Для изменения структуры ИНС часто применяются следующие операции:

- добавление/удаление нейронов;
- добавление/удаление связей.

Встречаются также модификации этих операций, включающие, например, «расщепление» существующего нейрона [43], «перенаправление» связи [17] и др.

Итак, выбор типа структурного изменения ИНС в общем случае неоднозначен из-за того, что оценка необходимости модификации топологии ИНС и варианта этой модификации не всегда возможна. Для решения проблемы выбора операции структурной модификации ИНС известны следующие варианты:

- определение эффективности каждой операции преобразования топологии ИНС на основании оценки качества получаемых НС решений [50, 51];

- использование дополнительной информации об ИНС для ее модификации (например, вычисление «значимости» связей [43], эвристическая оценка структуры ИНС [42, 52]).

**4.2. Кодирование информации о структуре ИНС.** Будем рассматривать два достаточно общих класса кодирования информации о топологии ИНС [37, 53]:

1. *Прямое кодирование (direct encoding)* – информация о структуре ИНС представлена в хромосоме в явном виде матрицей смежности [54, 55], списком связей [17, 40, 52, 56] и т.д. (рис. 4).

2. *Косвенное кодирование (indirect encoding)* – для описания структуры ИНС используются специальные грамматики [46, 57-59] (рис. 5 и 6).

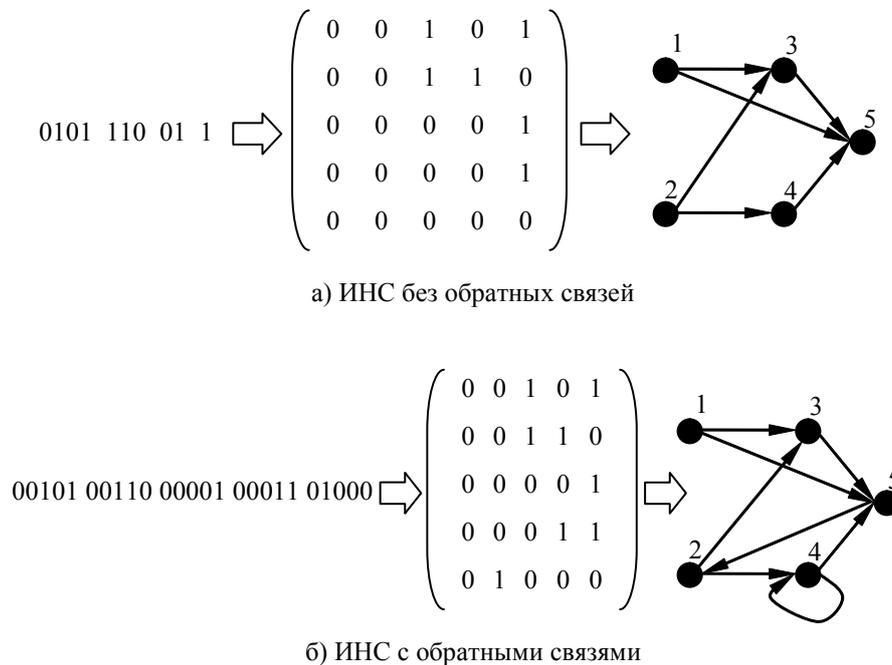


Рис. 4. Пример прямого кодирования структуры ИНС с использованием матрицы смежности для ИНС без обратных связей (а) и с обратными связями (б). Для каждого случая показаны (слева направо): хромосома, матрица смежности, структура ИНС с расставленными индексами нейронов.

Пример прямого кодирования структуры ИНС с использованием бинарной матрицы смежности  $C = \{c_{ij}, i, j = 1, \dots, N\}$ , где  $N$  – количество нейронов в ИНС, и соответствующие бинарные хромосомы представлены на рис. 4. Каждый элемент  $c_{ij}$  матрицы  $C$  принимает значения 1 или 0 и кодирует наличие связи (в случае  $c_{ij} = 1$ ) от  $i$ -го нейрона к  $j$ -му. Если топология ИНС не имеет обратных связей (рис. 4а), то в хромосоме достаточно записывать только части строк матрицы, находящиеся справа от главной диагонали.



1. Кодирование параметров структуры ИНС (количество слоев, число нейронов в каждом слое).

2. Введение специализированной грамматики, описывающей правила построения структуры ИНС. В данном варианте возможны два случая: (а) последовательность применения правил задана, но сами правила кодируются в хромосоме и отличаются у различных особей [46] (рис. 5); (б) в хромосоме кодируется последовательность применения правил, общих для всех особей, при этом набор правил и последовательность их применения для разных особей могут различаться [57, 59] (рис. 6).

Рассмотрим более подробно косвенное кодирование топологии ИНС. Иллюстративный пример эволюции правил построения структуры ИНС представлен на рис. 5 [37, 46]. Матрица смежности определяется в результате 3-шаговой итерационной процедуры, начиная с некоторого исходного состояния  $S$  (рис. 5в).

На первом шаге происходит декодирование символа  $S$  в матрицу четырех других символов  $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ , каждый из которых затем также декодируется в соответствующую ему матрицу согласно правилам (рис. 5б), закодированным в хромосоме особи (рис. 5а) и т.д. Результатом является матрица смежности, определяющая структуру ИНС (рис. 5в). В хромосоме кодируются только правила, определяющие 1-й и 2-й шаги построения структуры ИНС, а 16 различных правил, применяемых на 3-м шаге, задающие все возможные бинарные матрицы  $2 \times 2$  и обозначенные последовательно первыми 16-ю символами латинского алфавита (от «a» до «p»), определяются до начала работы алгоритма и не изменяются. Таким образом, какова бы ни была матрица смежности, сохраняется возможность закодировать ее содержимое в матрице  $4 \times 4$ , содержащей символы от «a» до «p», которая затем может быть закодирована четырьмя символами  $\{A, B, C, D\}$  в матрице  $2 \times 2$ . Тем самым обеспечивается возможность представить в хромосоме любую структуру ИНС. Ограничениями описываемого способа эволюции правил является фиксированное количество шагов построения матрицы смежности, а также невозможность использования рекурсивных декодирующих правил (например,  $A \rightarrow \begin{pmatrix} a & A \\ A & b \end{pmatrix}$ ).

Пример для эволюции последовательности применения правил для построения ИНС с бинарными весами связей<sup>1</sup>, алгоритм «Клеточного кодирования» (Cellular encoding) Фредерика Груо [57], показан на рис. 6 [60]. Сами правила определяются до начала работы алгоритма, а в хромосоме кодируется дерево, соответствующее последовательности выполнения этих правил. Количество шагов декодирования и их последовательность зависят от структуры дерева правил. Способы генетического кодирования древообразных структур описаны, например, в [61, 62].

<sup>1</sup> Хотя, фактически, алгоритм Ф. Груо помимо структуры ИНС настраивает также и веса связей, он может быть адаптирован к решению задачи поиска топологии ИНС путем исключения правил, относящихся к изменению весов связей.

Эволюционный подход к настройке и обучению искусственных нейронных сетей

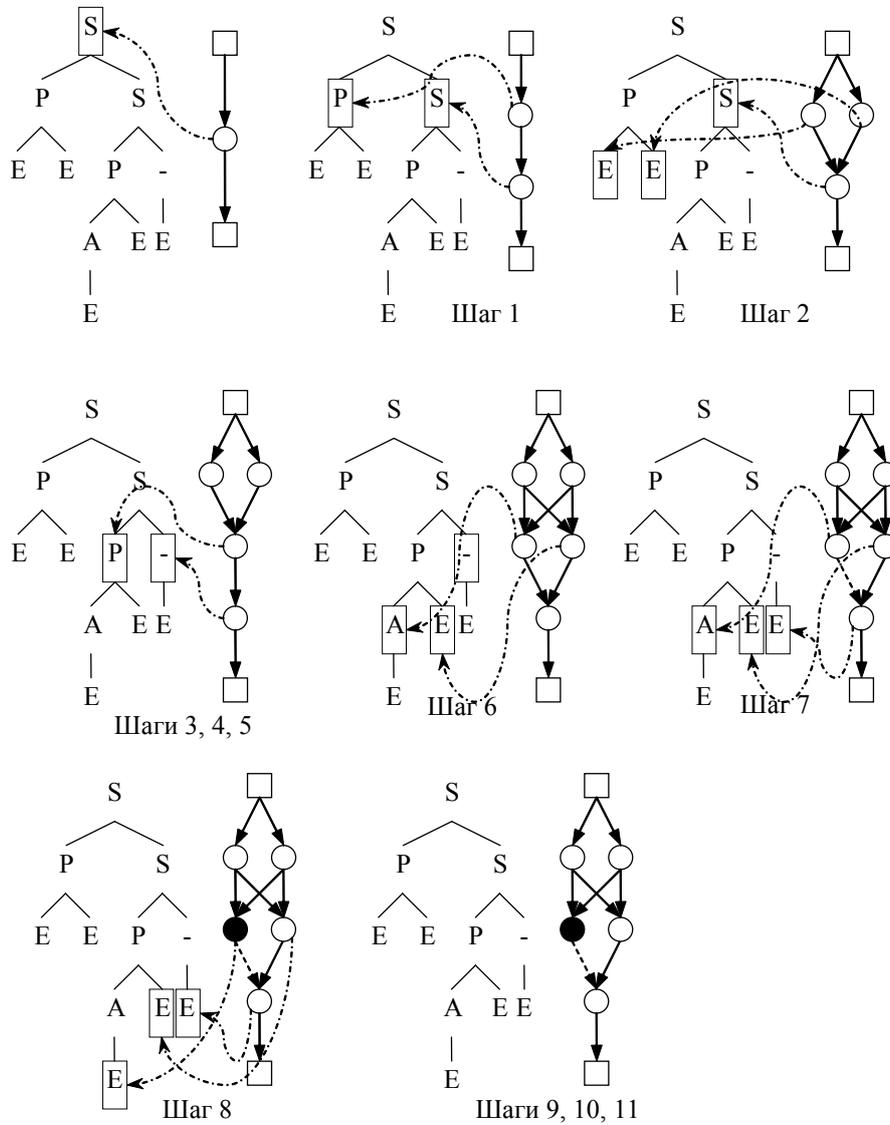


Рис. 6. Пример построения структуры ИНС с использованием специализированной грамматики и эволюцией последовательностей применения правил [60]. Описание операций, используемых при декодировании хромосомы, приведено в тексте.

Для примера, изображенного на рис. 6, используются следующие правила [60]:

- «S» – последовательное деление нейрона.
- «P» – параллельное деление нейрона.
- «A» – увеличение порога активационной функции нейрона.

«Е» – символ-терминатор, обозначающий окончание развития нейрона.

«-» – установка веса связи равным «-1».

На каждом шаге декодирования выполняется одно правило, записанное в узле дерева. Так, для примера на рис. 6, на первом шаге, в соответствии с правилом, обозначенным символом «S», нейрон делится последовательно, а на шаге 2 – параллельно, согласно правилу, обозначенному символом «P». После выполнения шагов 3, 4 и 5 два нейрона заканчивают развитие, а один делится последовательно и т.д.

Результатом является ИНС с 3-мя скрытыми слоями, в которой изменены значение порога функции активации одного скрытого нейрона (показано черным цветом) и величина одной связи (показано пунктиром). Кодирование в хромосоме дерева правил позволяет использовать рекурсивные операции для построения структуры ИНС [57]. Оригинальным и самостоятельным развитием описанного способа эволюции последовательности правил роста ИНС можно считать НЭ алгоритм, в котором учитываются пространственное расположение нейронов и длины связей между ними [63].

Заметим, что существуют неэволюционные подходы к настройке топологии сети либо во время ее обучения (алгоритм каскадной корреляции Фальмана [64] и его модификации [65] и др.), либо по завершении процесса обучения (алгоритмы Optimal Brain Damage (OBD) [66], Optimal Brain Surgeon (OBS) [67] и др.). Несмотря на то, что в [68] показана эффективность использования эволюционного подхода для настройки топологии ИНС по сравнению с алгоритмами OBD и OBS, однако детального исследования на эту тему, насколько известно авторам данной статьи, не проводилось.

## 5. Одновременная эволюционная настройка весов связей и структуры ИНС

Использование эволюционного подхода позволяет одновременно настраивать веса связей и структуру ИНС [17, 40, 42, 43, 50, 52, 54-59, 69-78]. При этом в хромосоме кодируется информация о весах и связях ИНС. Возможно использование как бинарного, так и вещественного кодирования для записи весов связей, а структура сети может быть представлена с использованием и прямого, и косвенного способов кодирования.

В силу того, что в случае одновременной настройки весов связей и структуры ИНС фактически комбинируются две различные задачи, пространство поиска многократно увеличивается и часто включает подпространства различной размерности (для ИНС с разным количеством связей). Для упрощения задачи поиска в таком сложном пространстве часто в явном или неявном виде вводятся ограничения на топологию ИНС и/или ее изменения:

- ограничение количества скрытых нейронов [17, 42, 50];
- ограничение количества связей [17, 59];
- ограничение изменения топологии ИНС (рассматриваются только растущие ИНС [40], либо только многослойные ИНС [73] и др.);

**5.1. Плюсы и минусы одновременной настройки весов связей и структуры ИНС.** Одновременное решение двух отдельных задач: настройки весов связей и структуры ИНС – позволяет в некоторой степени скомпенсировать недос-

татки, присущие каждой из них в отдельности и объединить их преимущества. С другой стороны, «платой» за это является огромное пространство поиска, а также объединение ряда недостатков, вызванных использованием эволюционного подхода. Суммируя, перечислим преимущества и недостатки.

Преимущества:

1. Независимость от структуры ИНС и характеристик функций активации нейронов.
2. Отсутствие необходимости в обучающей выборке.
3. Возможность автоматического поиска топологии ИНС и получения более точной нейросетевой модели.

Одним из основных преимуществ одновременной эволюционной настройки весов связей и структуры ИНС является возможность автоматизированного поиска ИНС, располагая только критерием оценки ИНС для осуществления эволюционного поиска. При этом, как и для эволюционного обучения ИНС (см. раздел 3), наличие обучающей выборки не является обязательным, а НЭ алгоритм может применяться для поиска ИНС с любыми активационными функциями нейронов.

По сравнению с отдельным эволюционным поиском структуры ИНС и настройкой весов связей, одновременное решение этих задач позволяет избежать некоторых недостатков. Так появление в популяции особей, которым соответствуют ИНС с различными топологиями, уменьшает значимость проблемы конкурирующих решений (см. подраздел 3.3), а наличие информации о весах связей позволяет обойти проблему субъективной оценки структуры ИНС, в силу того, что оценивается не структура нейросети, а вся ИНС «целиком».

Недостатки:

1. Сложность «тонкой» настройки весов связей на поздних этапах эволюционного поиска.
2. Большие, по сравнению с градиентными алгоритмами, требования к объему оперативной памяти из-за использования популяции ИНС.
3. Сложность организации поиска топологии ИНС.

**5.2. Кодирование информации о структуре и весах ИНС.** Поскольку веса связей кодируются вместе со структурой ИНС, то при записи информации о связи удобно записывать информацию о ее положении в сети и о величине веса вместе. Этим соображениям соответствует запись структуры ИНС в виде списка связей [17, 40, 52, 56, 74, 75]. При этом связи могут быть дополнительно сгруппированы по нейронам, для которых они являются входными [17, 56], что позволяет несколько сократить объем памяти, требуемый для записи информации об ИНС.

Для записи информации о структуре ИНС возможно также использование *кооперативного коэволюционного* подхода (*cooperative coevolution*) [79], когда решение записывается не в хромосоме каждой особи, а «разбивается на части» по хромосомам нескольких особей. Рассмотрим пример на основе алгоритма SANE Дэвида Мориарти [17]. В хромосоме особи записывается информация о входных и выходных связях одного нейрона. Во время оценки популяции формируется  $k$  ИНС, включающих нейроны, соответствующих различным особям (рис. 7), и производится их оценка. Оценка каждой особи вычисляется как средняя оценка ИНС, в состав которых входил нейрон, соответствующий этой осо-

би. Козволюционный подход также применяется и для эволюционной настройки весов связей ИНС [26, 27].

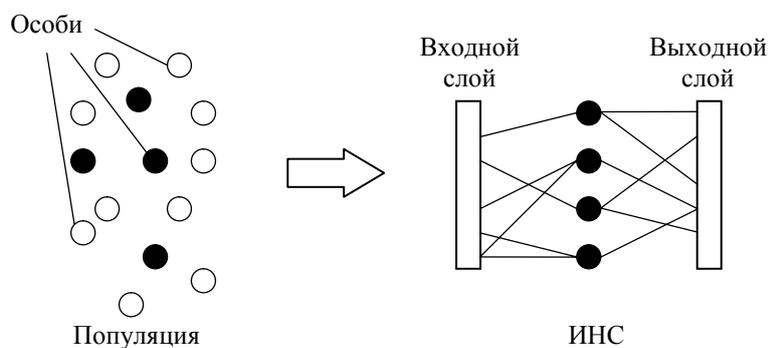


Рис. 7. Пример использования коэволюционного подхода к кодированию информации об ИНС для оценки особей. Черными кругами в популяции обозначены особи, соответствующие нейронам, включенным в структуру оцениваемой ИНС. Для определения приспособленности особей в популяции производится оценивание  $k$  различных ИНС. Рисунок представлен по [26].

Преимуществом использования коэволюционного подхода является уменьшение объема оперативной памяти необходимого для хранения информации о популяции. Также появляется возможность дополнительного распараллеливания эволюционного поиска за счет разбиения исходной задачи на отдельные подзадачи настройки связей скрытых нейронов. Недостатком является довольно сильное ограничение на структуру ИНС. Например, в [17] рассматриваются только сети прямого распространения с одним скрытым слоем, при этом количество скрытых нейронов, а также суммарное число входных и выходных связей для каждого скрытого нейрона фиксированы и задаются пользователем до начала работы алгоритма.

Информация о структуре ИНС и ее связях может также кодироваться отдельно. В работе [59] топология ИНС в хромосоме задается в виде дерева, а значения весов связей и величины порога функций активации нейронов записываются отдельно (рис. 8). Рассмотрим более подробно участок хромосомы, соответствующий верхней строке на рис. 8а. Рассматриваемый участок можно разбить на две части по выполняемым функциям:

1. Часть, кодирующая структуру ИНС («буквенная» часть участка хромосомы).
2. Часть, определяющая соответствие связей и нейронов с весами и порогами активационных функций соответственно («цифровая» часть участка хромосомы).

DTQaababaa ( bbaabba0571745 ) 736284668 ( 2 ) 867

Веса: { -1,64; -1,834; -0,295; 1,205; -0,807; -0,856; 1,702; -1,026; -0,417; -1,061 }

Параметры: { -1,14; 1,177; -1,179; -0,74; 0,393; 1,135; -0,625; 1,643; -0,029; -1,639 }

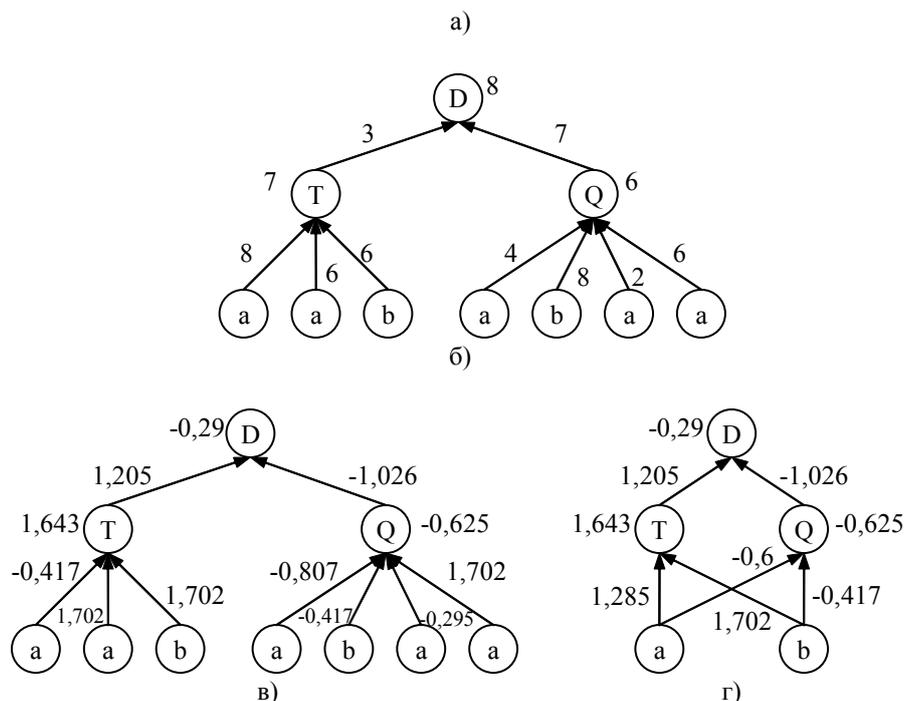


Рис. 8. Пример раздельного кодирования информации о структуре ИНС и значениях весов связей и параметрах функций активации [59]. Показаны: а) пример хромосомы; б) структура ИНС в виде дерева с индексами для весов связей и порогов функций активации нейронов; в) структура ИНС с «расставленными» весами связей и величинами параметров функций активации; г) структура ИНС в «стандартном» виде (связи от одинаковых входных нейронов объединены). Участки хромосомы, заключенные в круглые скобки, не учитываются при построении ИНС. Подробное описание правил декодирования хромосомы приведено в тексте.

В первой части рассматриваемого участка заглавные символы определяют количество входных связей нейронов сети («D», «T» и «Q» обозначают соответственно две, три и четыре входные связи), а строчные – входные нейроны ИНС («a» – первый входной нейрон, «b» – второй). Информация о связях «читается» в линейном порядке. Так, для части хромосомы «DTQaababaa» на рис. 8а, нейрон, соответствующий корню дерева (ИНС), обозначен символом «D» и поэтому имеет 2 входные связи от нейронов, обозначенных символами «T» и «Q». При этом «T-нейрон» имеет 3 входные связи от входных нейронов «a», «a» и «b»

(предполагается, что входные нейроны могут быть продублированы произвольное количество раз), а «Q-нейрон» имеет 4 входные связи от входных нейронов «a», «b», «a» и «a». Заметим, что оставшийся «отрезок» хромосомы «bbaabba» не используется. Таким образом, получаем структуру ИНС представленную в виде дерева на рис. 8б.

Во второй части участка хромосомы перечислены индексы для величин весов связей и порогов функций активации из отдельных одномерных массивов. Поскольку последние 7 символов из первой части пропускаются, то первые 7 символов из второй части «0571745» также не читаются. Следующие далее 9 индексов (по количеству связей) «расставляются» по соответствующим связям, начиная с корневого нейрона, при перечислении нейронов и связей справа налево, начиная от корня (рис. 8б). Для рассматриваемого примера индексы 7 и 3 определяют веса для входных связей к «корневому» нейрону («D-нейрон»), индексы 6, 2, 8 и 4 – веса связей для «Q-нейрона», а индексы 6, 6, 8 – веса связей для «T-нейрона».

Для определения величин порогов активационных функций следующий индекс в хромосоме (в данном случае «2», рис. 8а) пропускается, и записанные за ним индексы определяют величины порогов для нейронов в соответствии со значениями элементов отдельного массива пороговых величин. Порядок присвоения индексов – обратный. Результат расстановки индексов показан на рис. 8б, а соответствующая древовидная структура ИНС с определенными значениями весов связей – на рис. 8в.

Объединяя связи от одинаковых входных нейронов, получим «итоговую» ИНС на рис. 8г.

Для рассмотренного варианта раздельного кодирования структуры ИНС и весов связей характерна избыточность информации, записанной в хромосоме, что увеличивает требования к объему оперативной памяти, необходимого для хранения информации об одной ИНС, но, в то же время, оставляет «простор» для возможных дальнейших изменений структуры ИНС. Использование избыточности присутствует также в алгоритмах П. Ангелина [42] и К. Стенли [40, 73, 74].

Для одновременной настройки и обучения ИНС может использоваться косвенное кодирование информации [69]. В [69] рассматриваются бинарные ИНС, причем в хромосоме записывается информация только о росте, развитии ИНС, а величины весов связей по умолчанию считаются равными «1». В процессе декодирования хромосомы (см. рис. 6) значение веса связи может измениться и стать равным «-1».

Таким образом, существуют различные подходы к одновременному кодированию структуры и весов связей ИНС. Отметим, что чем сложнее запись информации об ИНС в хромосоме, тем, как правило, сложнее правила декодирования и больше вычислительные затраты на его реализацию. В [59, 70] показано, что подобные «сложные» способы кодирования могут использоваться для решения широкого круга задач, не только «нейросетевых», однако очень сложно сказать что-либо определенное об эффективности такого способа представления информации об ИНС.

## 6. Скрещивать или не скрещивать?

Оператор скрещивания (кроссовер, от англ. crossover) является основным оператором в генетическом алгоритме (ГА) и генетическом программировании (ГП). Одним из главных аргументов в пользу использования этого оператора является возможность рекомбинации параметров решений, представленных различными особями. Другими словами, в терминах *гипотезы о строительных блоках*<sup>1</sup> (*building blocks hypothesis*) [80], посредством кроссовера осуществляется поиск новых строительных блоков и смешивание уже существующих<sup>2</sup>. Однако в эволюционных стратегиях (ЭС) кроссовер используется крайне редко, и совсем не используется в эволюционном программировании (ЭП).

Применительно к концепции эволюционирующих ИНС существует серьезный аргумент против использования оператора кроссовера [37, 42]. Его смысл может быть сформулирован следующим образом. Каждый нейрон в ИНС несет определенную функциональную нагрузку, например, реализуя разделяющую гиперплоскость (для нейронов с сигмоидной, линейной и ступенчатой функциями активации), либо выделяя центр некоторого кластера в пространстве входных параметров (для нейронов с радиально-базисными функциями активации и карт Кохонена). Таким образом, реализуемое ИНС отображение  $X \rightarrow Y$ , где  $X$  – пространство входных сигналов,  $Y$  – пространство выходных сигналов ИНС, существенно зависит от характеристик нейронов, их функциональной нагрузки и векторов весов их входных связей. Использование кроссовера для рекомбинации параметров двух ИНС может стать причиной существенного искажения преобразований, соответствующих ИНС родительских особей, и, тем самым, привести к потере существовавших решений. Другими словами, «скрещивание» ИНС без учета их свойств даст скорее неудовлетворительный результат, чем будет способствовать поиску решения<sup>3</sup>.

Если рассматривать проблему перестановки (см. подраздел 3.3, рис. 3), то скрещивание различных хромосом, кодирующих одинаковые ИНС, может привести к тому, что в структуре ИНС особей-потомков одни скрытые нейроны с одинаковой функциональной нагрузкой будут продублированы, а другие скрытые нейроны будут отсутствовать.

---

<sup>1</sup> Гипотеза о строительных блоках предполагает существование различных частей хромосом разной длины – строительных блоков (СБ) и их комбинаций. Согласно этой гипотезе, наличие СБ и комбинаций СБ в некоторой хромосоме явно и практически всегда однозначно (положительно или отрицательно) сказывается на приспособленности соответствующей ей особи. Тогда эволюционный поиск сводится к *выявлению (identification)* «хороших» строительных блоков и к поиску их комбинаций, *смешиванию (mixing)*, способствующих повышению приспособленности особей.

<sup>2</sup> Существует и «биологический» аргумент в пользу кроссовера [81]. Его суть заключается в том, что скрещивание, отсутствующее у простейших, «изобретено» эволюцией для развития сложных организмов насекомых, рыб, земноводных, птиц и млекопитающих как ответ на усложнение их генетического кода, морфологии и динамичные условия жизни. Исходя из этих соображений, делается вывод о необходимости использования оператора кроссовера.

<sup>3</sup> Аналогичная проблема, когда использование кроссовера часто приводило к появлению «нежизнеспособных» решений существовала и при применении ГА для решения задач на графах (задача коммивояжера, поиск минимальных покрытий и др.), пока не были разработаны специализированные операторы, учитывающие специфику задачи.

Одним из естественных вариантов решения проблемы «скрещивания» ИНС является разработка адаптивных операторов скрещивания, которые учитывают различия в структуре скрещиваемых ИНС и создают «жизнеспособных» потомков, либо использование такого способа кодирования информации, который позволяет уменьшить ошибки, возникающие при использовании «стандартных» операторов скрещивания и мутации. Отметим, что разработка проблемно-ориентированных генетических операторов является одним из важнейших составляемых в разработке эффективных ЭА [82].

## 7. Нейроэволюционные алгоритмы и задачи с качественной оценкой ИНС

Обучение ИНС, с «классической» точки зрения, подразумевает, что существует обучающая выборка, использование которой дает возможность подстраивать веса связей ИНС с помощью градиентных алгоритмов, среди которых широко известны алгоритм обратного распространения ошибки, сопряженных градиентов, Левенберга-Марквардта и др. Однако, в ряде случаев формирование обучающей выборки сопряжено со следующими трудностями:

1. Необходимость переформулировки исходной задачи, в результате чего могут появиться дополнительные подзадачи, мало связанные с исходной проблемой, что делает решение громоздким. Например, один из вариантов решения задачи нейроуправления заключается в нейросетевой аппроксимации объекта управления с последующей оптимизацией состояния управляемого объекта посредством поиска соответствующих значений управляющих сигналов.

2. Сложность определения значений компонент вектора  $Y$  из обучающего множества. Данная проблема возникает при решении некорректных задач, а также, если необходимо оценить последовательность выходных сигналов ИНС. К таким задачам относятся задачи, связанные с адаптивным поведением, игровыми стратегиями, обработкой изображений и др.

Одним из возможных решений перечисленных проблем является обучение ИНС с использованием приближенной, интегральной оценки, отражающей качественные, внешние особенности ее функционирования. Будем называть оценку ИНС *качественной неполной* оценкой или, коротко, *качественной* оценкой<sup>1</sup>, если эта оценка позволяет с требуемой точностью оценить качество функционирования ИНС, но при этом ее значения недостаточно для подстройки весов связей ИНС с использованием градиентных алгоритмов. Примерами качественной оценки являются: время поддержания стабильного состояния объекта управления для задачи нейроуправления [17, 26, 40, 52, 83]; процент выигранных игр, либо оценка позиции для задачи поиска игровых стратегий [28, 72, 74, 77]; качество изображения для задачи обработки изображений [84] и т.д. Такие оценки более естественны и «интуитивны», однако, как следствие, они не могут быть использованы градиентными алгоритмами обучения для подстройки весов

---

<sup>1</sup> Использование термина «качественная» оценка, очевидно, является не самым лучшим вариантом, т.к. означает существование не только «некачественной», но также и «количественной» оценок. При этом возникает закономерный вопрос о формальном разделении этих оценок, и оказывается, что провести четкие границы совсем непросто.

связей. Поэтому «подходящий» обучающий алгоритм должен изменять веса связей ИНС, не имея информации о величине ошибки для каждого выхода этой НС.

Отметим, что в случае «стандартного» обучения ИНС, для ее оценки и обучения, фактически, используются различные величины: среднеквадратичная ошибка выхода ИНС отражает качество ее функционирования, в то время как разность реального и требуемого выходных сигналов ИНС используется для подстройки весов связей. Т.е. качественная оценка присутствует и здесь, но она не используется напрямую для обучения ИНС, для осуществления которого необходимо репрезентативное с точки зрения решаемой задачи множество наборов входных и соответствующих им выходных данных.

При использовании ЭА для обучения ИНС достаточно одной качественной оценки. Здесь важна *адекватность* оценки, т.е. если согласно используемой оценке качество первой ИНС лучше качества второй ИНС, то из этого должно следовать, что первая ИНС лучше второй. Кроме адекватности оценки также важно ее *существование* для любого решения, которое может быть закодировано в хромосоме. При этом оценка может вычисляться независимо от расхождения выходного сигнала ИНС с эталонным значением. Таким образом, отсутствие обучающей выборки не является серьезным препятствием для эволюционного обучения ИНС, если существует альтернативный способ оценки ее функционирования. Это позволяет расширить круг практических приложений нейроинформатики, а также упростить процесс решения многих задач, связанных с адаптивным нейроуправлением, поведением и моделированием.

Рассмотрим коротко задачи, для решения которых использование качественной оценки ИНС помогает упростить процесс поиска решения и повысить качество результатов.

**7.1. Адаптивное управление и поведение.** При решении задач, связанных с управлением сложным динамическим объектом, эталонный управляющий сигнал в общем случае неизвестен. Сложность синтеза исполняющего устройства (ИУ) часто связана со сложностью сопоставления оценки ИУ и необходимых изменений структуры и параметров ИУ. Здесь одним из возможных решений является использование *обучения с подкреплением (reinforcement learning)* [85], однако для его применения необходимо оценивать каждое действие ИУ, что не всегда доступно. Использование ЭА для обучения ИНС позволяет использовать оценку нейросетевого ИУ по серии управляющих сигналов. Некоторые общие вопросы применения эволюционирующих ИНС для решения задач теории регулирования рассматриваются в [34].

В случае, когда ОУ является искусственным интеллектуальным агентом (виртуальным либо реальным, например, роботом), а внешняя по отношению к агенту среда является аналогом среды обитания, то задача управления таким агентом может рассматриваться как задача адаптивного поведения. При этом использование ИНС в качестве ИУ агента позволяет добиться большой гибкости действий агента вместе с его обучаемостью [2, 3, 75]. Во многих исследованиях поведение агентов определяется его НС [3, 22, 26, 31, 75, 86], при этом структура сети часто фиксирована. Возможность изменения топологии ИНС в ходе эволюции позволяет выделить и изучить интересные аспекты поведенческой адаптации и повысить, в целом, уровень «интеллектуальности» поведения

агентов. Таким образом, НЭ подход представляет интерес для таких направлений исследований как «Искусственная жизнь» (*Artificial Life, A-Life*) [2, 87] и «Адаптивное поведение» (*Adaptive Behavior*) [2, 88].

Отдельным направлением в исследовании адаптивного поведения является синтез оптимальных игровых стратегий. Часто рассматриваются настольные игры (например, крестики-нолики [72] и го [28, 74, 77]). При этом возможно использование «автообучения», когда настройка весов ИНС осуществляется путем соревнования одной популяции ИНС с другой [77] (т.н. *конкурентная коэволюция (competitive coevolution)*).

**7.1.1. Проект NERO К. Стэнли.** Ярким примером применения НЭ подхода к решению задач адаптивного поведения является проект NERO (Neuro-Evolution of Robotic Operatives, [http://dmc.ic2.org/nero\\_public](http://dmc.ic2.org/nero_public)) [75]. Ключевой технологией проекта является модифицированный алгоритм NEAT для одно-временной настройки структуры и весов связей ИНС, разработанный Кеннетом Стэнли (Kenneth Stanley). Результатом проекта является интерактивная компьютерная игра, в которой пользователь сначала тренирует на специальном тренировочном полигоне команду агентов, а затем его команда «соревнуется» с агентами компьютера или другого пользователя на своеобразном «поле боя».

Поведение агентов игрока определяется соответствующими им ИНС, которые настраиваются в процессе обучения на полигоне. Действия агентов зависят от создающейся на тренировочном полигоне ситуации и определяются приоритетами поведения (расстояние до противника, подвижность агентов, активность стрельбы и др.), заданными пользователем. Более приспособленные с точки зрения выбранных приоритетов агенты используются для генерации агентов-потомков, которые замещают худших существующих агентов.

Для выбора действия (направление движения и стрельба) агентам доступна следующая информация:

- расположение агентов-противников, находящихся в поле зрения агента;
- находится ли агент-противник на линии огня;
- расположение препятствий (стены, различные объекты);
- направление стрельбы ближайшего агента-противника.

Использование для реализации проекта NERO НЭ алгоритма позволяет в реальном времени осуществлять поиск разнообразных стратегий поведения агентов. В процессе обучения агенты учатся решать такие задачи, как поиск маршрута движения в присутствии препятствий, преследование агентов-противников, использование стен для прикрытия от огня и избежания окружения и др. Отметим, что во время обучения агентов задачи ставятся в общем виде (дойти до определенной точки, атаковать цель и т.д.) и возникновение сложных стратегий поведения (например, укрытие за стенами) зависит от ситуации, которую создаст на тренировочном полигоне игрок.

**7.2. Многоагентные системы.** Поскольку во многих НЭ алгоритмах каждая особь представляет отдельную ИНС, то популяция может рассматриваться как эволюционирующая многоагентная система, в которой стратегия поведения каждого агента определяется соответствующей ему нейросетью [89]. Один из «классических» подходов к реализации стратегии поведения агента заключается в использовании набора правил формальной логики, состав которого формируется в результате действий агента. Среди недостатков такого подхода выделим

сложность синтеза правил, определяющих как приобретение агентом знаний, так и обмен знаниями между агентами, а также разрастание размера базы знаний агента в процессе накопления им «жизненного опыта». Применение ИНС, реализующей стратегию поведения агента, позволяет хранить знания в неявном виде в компактной форме, при этом сохраняется возможность дообучения и переобучения агента. Отметим, что взаимодействие НЭ-агентов в многоагентной системе возможно путем не только формального «общения» или противодействия, но и посредством применения операций селекции (своеобразное «соревнование» агентов) и скрещивания (обмен генетической информацией). При скрещивании агентов возможен обмен информацией, либо имевшейся у них в момент рождения (дарвиновская эволюция), либо накопленной ими в процессе жизни (эволюция ламарковского типа). В роли качественной оценки агента может использоваться показатель успешности его действий, определяемый в соответствии с поставленной задачей.

### Заключение

Аппроксимирующие возможности ИНС очень высоки, однако волнующий вопрос «Как научить ИНС делать все?» поднимавшийся, например, в [90], остается открытым. С одной стороны, концепция ИНС является универсальной по отношению к решению широкого спектра задач, а с другой, существуют проблемы (например, задачи нейроуправления), сама формулировка которых вынуждает искать «обходные» пути. В то же время, известен эволюционный подход к решению оптимизационных задач самой разной природы и претендующий на универсальность применения, наравне со случайным поиском и последовательным перебором, поскольку может рассматриваться как формализованный метод проб и ошибок.

Совместное использование ЭА и ИНС позволяет решать задачи настройки и обучения ИНС как по отдельности, так и одновременно. Одним из достоинств такого синтезированного подхода является во многом унифицированный подход к решению разнообразных задач классификации, аппроксимации, управления и моделирования. Использование качественной оценки функционирования ИНС (подробнее см. раздел 7) позволяет применять НЭ алгоритмы для решения задач исследования адаптивного поведения интеллектуальных агентов, поиска игровых стратегий, обработки сигналов. Несмотря на то, что количество проблем и открытых вопросов, касающихся разработки и применения НЭ алгоритмов (способы кодирования, генетические операторы, методы анализа и др.) велико, часто для успешного решения задачи с использованием НЭ алгоритма достаточно адекватного понимания проблемы и НЭ подхода, свидетельством чего является большое число интересных и успешных работ в данном направлении.

Данная статья не претендует на полноту излагаемого материала, так, например, «за бортом» остались такие задачи как синтез правил формальной логики, создание компьютерной музыки, обработка сигналов и изображений и др. Однако авторы надеются, что содержащейся в статье информации достаточно для формирования общего представления об использовании эволюционных прин-

ципов для настройки и обучения ИНС при решении как «классических» задач классификации и аппроксимации, так и для задач с качественной оценкой ИНС.

Авторы выражают благодарность рецензентам за ценные замечания, способствовавшие улучшению статьи.

## Литература

1. Осовский С. Нейронные сети для обработки информации. М.: Финансы и статистика, 2002.
2. Редько В.Г. Эволюционная кибернетика. М.: Наука, 2001.
3. Бурцев М.С. Исследование новых типов самоорганизации и возникновения поведенческих стратегий. Диссертация на соискание ученой степени кандидата физико-математических наук. Институт прикладной математики им. В.М. Келдыша РАН, Москва, 2005. См. также: <http://mbur.narod.ru/>
4. Beyer H.-G., Schwefel H.-P., Wegener I. How to analyse evolutionary algorithms. Technical Report No.CI-139/02. University of Dortmund, Germany, 2002. See also: <http://ls11-www.cs.uni-dortmund.de/people/schwefel/WelcomeE.html>
5. Holland J.H. Adaptation in Natural and Artificial Systems. The University of Michigan Press, 1975.
6. Емельянов В.В., Курейчик В.В., Курейчик В.М. Теория и практика эволюционного моделирования. М.: ФИЗМАТЛИТ, 2003.
7. Фогель Л., Оуэнс А., Уолш М. Искусственный интеллект и эволюционное моделирование. М.: Мир, 1969.
8. Родзин С.П. Гибридные интеллектуальные системы на основе алгоритмов эволюционного программирования // Новости искусственного интеллекта, 2000, № 3, с. 159-170.
9. Rechenberg I. Evolutionsstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution. Werkstatt Bionik und Evolutionstechnik, Stuttgart: Frommann-Holzboog, 1973.
10. Schwefel H.-P. Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie // Interdisciplinary Systems Research, 1977. Vol. 26, pp. 5-8.
11. Koza J. Genetic programming: a paradigm for genetically breeding computer population of computer programs to solve problems. Cambridge, MA: MIT Press, 1992.
12. Курейчик В.М., Родзин С.И. Эволюционные алгоритмы: генетическое программирование. Обзор // Известия РАН. ТИСУ, 2002, № 1, с. 127-137.
13. De Jong K. An analysis of the behavior of a class of genetic adaptive systems. Unpublished PhD thesis. University of Michigan, Ann Arbor, 1975. (University Microfilms No. 76-9381). See also: <http://www.cs.gmu.edu/~eclab/>
14. Mattiussi C., Waibel M., Floreano D. Measures of Diversity for Populations and Distances Between Individuals with Highly Reorganizable Genomes // Evolutionary Computation, 2004, no. 12(4), pp. 495-515. See also: <http://lis.epfl.ch/member.php?SCIPER=111729>
15. Whitley D. An Overview of Evolutionary Algorithms // Journal of Information and Software Technology, 2001, no. 43, pp. 817-831. See also: <http://www.cs.colostate.edu/~genitor>
16. Evolutionary Computation 1: Basic Algorithms and Operators / Back T., Fogel D.B., Michalewicz Z. (Eds.). Bristol and Philadelphia: Institute of Physics Publishing, 2000.
17. Moriarty D., Miikkulainen R. Efficient Reinforcement Learning through Symbiotic Evolution // Machine Learning, 1996, no. 22, pp. 11-33. See also: <http://www.nn.cs.utexas.edu/>

18. Montana D., Davis L. Training feedforward neural networks using genetic algorithms // Proceedings of the 11th International Joint Conference on Artificial Intelligence. San Mateo, CA: Morgan Kaufmann, 1989. P. 762–767.
19. Belew R. K., McInerney J., Schraudolph N. N. Evolving networks: Using the genetic algorithm with connectionist learning // Artificial Life II, 1992, pp. 511-547.
20. Heistermann J. Different learning algorithms for neural networks – a comparative study / In Y. Davidor, H.P. Schwefel, R. Manner (Eds.) Parallel Problems Solving from Nature, Workshop Proceedings. Springer-Verlag, 1994. P. 368-396.
21. Комарцова Л.Г. Исследование алгоритмов обучения многослойного персептрона // Нейрокомпьютеры: Разработка и применение. М.: Радиотехника, 2002, № 12.
22. Мосалов О.П., Редько В.Г. Модель эволюционной ассимиляции приобретенных навыков в нейросетевых системах управления адаптивных агентов // Научная сессия МИФИ-2005. VII Всероссийская научно-практическая конференция "Нейроинформатика-2005": Сборник научных трудов. В 2-х частях. Часть 1. М.: МИФИ, 2005. С.210-217. См. также: <http://www.iont.ru/projects/rfbr/90197/>
23. Valsalam V., Bednar J., Miikkulainen R. Constructing Good Learners using Evolved Pattern Generators // In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO). Washington, DC, 2005. See also: <http://www.nn.cs.utexas.edu/>
24. Thierens D., Suykens J., Vanderwalle J., De Moor B. Genetic Weight Optimization of Feedforward Neural Network Controller // Proceedings of International Conference on Artificial Neural Networks and Genetic Algorithms. Springer-Verlag, 1993. P. 658-663.
25. Thierens D. Non-Redundant Genetic Coding for Neural Networks. Technical report No.UU-CS-1998-46. Utrecht University, Netherlands, 1998. See also: <http://www.cs.uu.nl/research/techreps/aut/dirk.html>
26. Gomez F., Miikkulainen R. Incremental Evolution of Complex General Behavior // Adaptive Behavior, 1997, no.5, pp. 317-342. See also: <http://www.nn.cs.utexas.edu/>
27. Gomez F., Miikkulainen R. Active Guidance for a Finless Rocket using Neuroevolution // Proceedings of Genetic and Evolutionary Computation Conference 2003 (GECCO 2003). Washington, USA, 2003. See also: <http://www.nn.cs.utexas.edu/>
28. Perez-Bergquist A. S. Applying ESP and Region Specialists to Neuro-Evolution for Go. Technical Report CSTR01-24. The University of Texas at Austin, 2001. See also: <http://www.nn.cs.utexas.edu/>
29. Saravanan N., Fogel D. B. Evolving neural control systems // IEEE Expert, 1995, pp. 23-27.
30. Korning P.G. Training of Neural Networks by means of Genetic Algorithms Working on very long Chromosomes // International Journal of Neural Systems, 1995, Vol. 6, No. 3, pp. 299-316. См. также: <http://citeseer.ist.psu.edu/>
31. Lund H.H., Mayoh B.H. Specialization in Populations of Artificial Neural Networks // Proceedings of the 5-th Scandinavian Conference on Artificial Intelligence (SCAI'95). Amsterdam: IOS Press, 1995. См. также: <http://citeseer.ist.psu.edu/>
32. Wieland A. Evolving neural network controllers for unstable systems // Proceedings of the International Joint Conference on Neural Networks. Piscataway, New Jersey, 1991. pp. 667–673.
33. Yamamichi T., Saito T., Taguchi K., Torikai H. Synthesis of Binary Cellular Automata based on Binary Neural Networks // Proceedings of International Joint Conference on Neural Networks. Montreal, Canada, 2005. P. 1361-1365.
34. Вороновский Г.К., Махотило К.В., Петрашев С.Н., Сергеев С.А. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности. Харьков: Основа, 1997. См. также: <http://neuroschool.narod.ru/books/gannvirt.html>
35. Herrera F., Lozano M., Verdegay J.L. Tackling Real-Coded Genetic Algorithms: Operators and tools for the Behaviour Analysis // Artificial Intelligence Review, 1998, no. 12, pp. 265-319. See also: <http://decsai.ugr.es/~herrera/>

36. Deb K., Anand A., Joshi D. A Computationally Efficient Evolutionary Algorithm for Real-Parameter Optimization. KanGAL Report No.2002003. Indian Institute of Technology, Kanpur, 2002. See also: <http://www.iitk.ac.in/kangal/>
37. Yao X. Evolving artificial neural networks // Proceedings of the IEEE, 1999, no. 9(87), pp. 1423-1447. See also: <http://www.cs.bham.ac.uk/~xin/>
38. Whitley D., Starkweather T., Bogart C. Genetic Algorithms and Neural Networks: Optimizing Connections and Connectivity // Parallel Computing, 1990, Vol. 14, pp. 341-361. See also: <http://www.cs.colostate.edu/~genitor/>
39. Shaffer J., Whitley D., Eshelmann L. Combination of Genetic Algorithms and Neural Networks: A Survey of the State of the Art // In Proceedings of the International Workshop on Combination of Genetic Algorithms and Neural Networks (COGANN-92). Los Alamos, CA: IEEE Computer Society Press, 1992. pp. 1-37. See also: <http://www.cs.colostate.edu/~genitor/>
40. Stanley K.O., Miikkulainen R. Evolving Neural Networks through Augmenting Topologies // Evolutionary Computation, 2002, no. 2(10), pp. 99-127. See also: <http://www.nn.cs.utexas.edu/>
41. Radcliffe N. Genetic set recombination and its application to neural network topology optimization // Neural Computing and Applications, 1993, Vol.1, pp. 67-90.
42. Angeline P.J., Saunders G.M., Pollack J.B. An Evolutionary Algorithm that Constructs Recurrent Neural Networks // IEEE Transactions on Neural Networks, 1994, no. 5(1), pp. 54-65. See also: <http://citeseer.ist.psu.edu/>
43. Yao X., Liu Y. Making Use of Population Information in Evolutionary Artificial Neural Networks // IEEE Transactions on systems, man, and cybernetics - part B: Cybernetics, 1998, no. 28(3), pp. 417-425. See also: <http://www.cs.bham.ac.uk/~xin/>
44. Hancock P. Recombination Operators for the Design of the Neural Nets by Genetic Algorithm // Parallel Problem Solving from Nature (PPSN-II). North-Holland, 1992. pp. 441-450.
45. Fiszlelew A., Britos P., Perichisky G., Garcia-Martinez R. Automatic Generation of Neural Networks based on Genetic Algorithms // Revista Eletronica de Sistemas de Informacao, 2003, Vol. 2, no. 1. См. также: <http://citeseer.ist.psu.edu/>
46. Kitano H. Designing neural network using genetic algorithm with graph generation system // Complex Systems, 1990, no. 4, pp. 461-476.
47. Хомич А.В., Жуков Л.А. Эволюционный метод оптимизации структуры нейронной сети с учителем // Научная сессия МИФИ-2005. VII Всероссийская научно-практическая конференция "Нейроинформатика-2005": Сборник научных трудов. В 2-х частях. Часть 1. М: МИФИ, 2005. С. 11-18.
48. Caruana R.A., de Sa V.R. Benefiting from the variables that variable selection discards // Journal of Machine Learning Research, 2003, Vol. 3, pp. 1245-1264.
49. Царегородцев В.Г. Редукция размеров нейросети не приводит к повышению обобщающих способностей // Материалы XII Всероссийского семинара «Нейроинформатика и ее приложения», Красноярск, 2004. С. 163-165. См. также: <http://www.neuropro.ru/>
50. Igel C., Kreutz M. Operator adaptation in evolutionary computation and its application to structure optimization of neural networks // Neurocomputing, 2003, no. 55(1-2), pp. 347-361. See also: <http://www.neuroinformatik.ruhr-uni-bochum.de/PEOPLE/igel/publications.html>
51. Хомич А.В., Степанян И.В., Карпищук А.В. Диагностика хронического пылевого бронхита по данным акустической спирометрии с применением блочных нейронных сетей // Информационные процессы, 2005, Том 5, № 5, с. 405-413.
52. Цой Ю.Р., Спицын В.Г. Применение генетического алгоритма для решения задачи адаптивного нейроуправления // Научная сессия МИФИ-2005. VII Всероссийская научно-практическая конференция "Нейроинформатика-2005": Сборник научных

- трудов. В 2-х частях. Часть 1. М.:МИФИ, 2005. С. 35-43. См. также: <http://qai.narod.ru/Publications/>
53. Balakrishnan K., Honavar V. Evolutionary Design of Neural Architectures – A Preliminary Taxonomy and Guide to Literature. Technical Report CS TR #95-01. Iowa State University, 1995. See also: <http://www.cs.iastate.edu/~gannadm/homepage.html>
  54. Law D., Miikkulainen R. Grounding Robotic Control with Genetic Neural Networks. Technical Report No. AI94-223. The University of Texas at Austin, 1994. See also: <http://www.nn.cs.utexas.edu/>
  55. Derakhshani R. GETnet: A General Framework for Evolutionary Temporal Neural Networks // Proceedings of International Joint Conference on Neural Networks. Montreal, Canada, 2005. pp. 3150-3155.
  56. Fullmer B., Miikkulainen R. Using Marker-Based Genetic Encoding of Neural Networks to Evolve Finite State Behavior // In Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life (ECAL-91). Paris, 1991. pp.255-262. See also: <http://www.nn.cs.utexas.edu/>
  57. Gruau F. Genetic synthesis of Boolean neural networks with a cell rewriting developmental process // In Proceedings of the International Workshop on Combination of Genetic Algorithms and Neural Networks (COGANN-92). Los Alamos, CA: IEEE Computer Society Press, 1992. pp. 55-74. See also: <http://www.cs.colostate.edu/~genitor/>
  58. Шукович Г. Применение генетических алгоритмов и систем генерирующих графов для создания модулярных нейросетей // Программирование, 2002, № 1, с. 13-20.
  59. Ferreira C. Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence. Angra do Heroismo, Portugal, 2002. See also: <http://gene-expression-programming.com/GepBook/Introduction.htm>
  60. Whitley D. Genetic Algorithms and Neural Networks / Winter, Periaux, Galan and Cuesta (Eds.) Genetic Algorithms in Engineering and Computer Science. John Wiley, 1995. pp. 203-216. See also: <http://www.cs.colostate.edu/~genitor/>
  61. Palmer C.C., Kershbaum A. Computer science: representing trees in genetic algorithms // Back T., Fogel D., Michalewicz Z. (Eds.) Handbook of evolutionary computation. Bristol, New York: Institute of Physics Publishing and Oxford University Press, 1997. pp. G1.3:1 – G1.3:8.
  62. Скобцов В.Ю., Скобцов Ю.А. Модификации генетического программирования // Труды Международных научно-технических конференций «Интеллектуальные системы (IEEE AIS'04)» и «Интеллектуальные САПР» (CAD-2004). Научное издание в 3-х томах. Т.1. М.: Физматлит, 2004. С. 78-86.
  63. Kodjabachian J., Meyer J.A. Evolution and development of modular control architectures for 1-D locomotion in six-legged animats // Connection Science, 1998, Vol. 18, pp. 211-237. See also: <http://animatlab.lip6.fr/index.en.html>
  64. Fahlman S.E., Lebiere C. The cascade-correlation learning architecture / Touretzky D. S. (Ed.) Advances in Neural Information Processing Systems 2. San Mateo, CA: Morgan Kaufman Publishers. P. 524-532.
  65. Littmann E., Ritter H. Cascade network architectures // Proceedings of International Joint Conference on Neural Networks, 1992, Vol. 2. pp. 398-404.
  66. LeCun Y., Denker J.S., Solla S.A. Optimal brain damage / In Touretzky D. S. (Ed.) Advances in Neural Information Processing Systems 2. San Mateo, CA: Morgan Kaufman Publishers, 1990. pp. 598-605. See also: <http://yann.lecun.com/>
  67. Hassibi B., Stork D.G. Second order derivatives for network pruning: Optimal brain surgeon / In Hanson S.J., Cowan J.D., Giles C.L. (Eds.) Advances in Neural Information Processing Systems 5. San Mateo, CA: Morgan Kaufman Publishers, 1993. P. 164-171.
  68. Ragg T., Braun H., Landsberg H. A Comparative Study of Neural Network Optimization Techniques. // In Proceedings of the ICANNGA 97. Springer-Verlag, 1997.

69. Yao X. Evolutionary artificial neural networks // *International Journal of Neural Systems*, 1993, vol. 4, no. 3, pp. 203–222. See also: <http://www.cs.bham.ac.uk/~xin/>
70. Gruau F. Neural network synthesis using cellular encoding and the genetic algorithm. Unpublished PhD thesis. Lyon: l'Universite Claude Bernard, 1994. See also: <http://citeseer.ist.psu.edu/>
71. Chen C.-C., Miikkulainen R. Creating Melodies with Evolving Recurrent Neural Networks // *Proceedings of the 2001 International Joint Conference on Neural Networks (IJCNN-01)*. Piscataway, NJ: IEEE, 2001. P. 2241-2246. See also: <http://www.nn.cs.utexas.edu/>
72. James D., Tucker P. A Comparative Analysis of Simplification and Complexification in the Evolution of Neural Network Topologies // *Proceedings of Genetic and Evolutionary Computation Conference (GECCO-2004)*. New York, NY: Springer-Verlag, 2004. See also: <http://citeseer.ist.psu.edu/>
73. Паклин Н. Обучаем нейронную сеть генетическим алгоритмом. 2003. См. также: [http://paklin.newmail.ru/mater/gene\\_net.html](http://paklin.newmail.ru/mater/gene_net.html)
74. Stanley K.O., Miikkulainen R. Evolving a roving eye for go // *In Proceedings of Genetic and Evolutionary Computation Conference (GECCO-2004)*. New York, NY: Springer-Verlag, 2004. See also: <http://www.nn.cs.utexas.edu/>
75. Stanley K.O., Bryant B.D., Miikkulainen R. The NERO real-time video game. Technical report AI-TR-04-312. The University of Texas at Austin, 2004. See also: <http://www.nn.cs.utexas.edu/>
76. Pan Z., Kang L., Nie S. Evolving Both the Topology and Weights of Neural Networks // *Parallel Algorithms and Applications*, 1996, no. 9, pp. 299-307. See also: <http://citeseer.ist.psu.edu/>
77. Lubberts A., Miikkulainen R. Co-Evolving a Go-Playing Neural Network // *Proceedings of Coevolution: Turning Adaptive Algorithms upon Themselves, Birds-of-a-Feather Workshop, Genetic and Evolutionary Computation Conference (GECCO-2001)*. San Francisco, CA: Kaufmann, 2001. pp. 14-19. See also: <http://www.nn.cs.utexas.edu/>
78. Polani D., Miikkulainen R. Eugenic Neuro-Evolution for Reinforcement Learning // *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2000)*. San Francisco, CA: Kaufmann, 2000. pp. 1041-1046. See also: <http://www.nn.cs.utexas.edu/>
79. Potter M. A., De Jong K. A. Cooperative coevolution: An architecture for evolving co-adapted subcomponents // *Evolutionary Computation*, 2000, no. 8, pp. 1–29. See also: <http://www.cs.gmu.edu/~eclab/>
80. Goldberg D.E. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.
81. Holland J.H. Building Blocks, Cohort Genetic Algorithms, and Hyperplane-Defined Functions // *Evolutionary Computation*, 2000, no. 4(8), pp. 373-391.
82. Barnett L. Evolutionary Search on Fitness Landscapes with Neutral Networks. Unpublished PhD thesis. The University of Sussex, England, 2003. See also: <http://www.cogs.susx.ac.uk/users/lionelb/>
83. Igel C. Neuroevolution for Reinforcement Learning Using Evolution Strategies // *Proceedings of Congress on Evolutionary Computation 2003 (CEC 2003)*, Vol. 4. IEEE Press, 2003. pp. 2588-2595. See also: <http://www.neuroinformatik.ruhr-uni-bochum.de/PEOPLE/igel/publications.html>
84. Цой Ю.Р., Спицын В.Г., Чернявский А.В. Нейроэволюционное улучшение качества изображений // *Научная сессия МИФИ-2006. VIII Всероссийская научно-практическая конференция "Нейроинформатика-2006": Сборник научных трудов*. В печати.
85. Sutton R.S., Barto A.G. *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press, 1998. See also: <http://www-anw.cs.umass.edu/~rich/book/the-book.html>

Эволюционный подход к настройке и обучению искусственных нейронных сетей

86. Ackley D., Littman M. Interactions between learning and evolution // Artificial Life II. Reading, MA: Addison-Wesley, 1992. pp. 487-509.
87. Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems / Langton C.G. (Ed.) Redwood City, CA: Addison-Wesley. 1992.
88. Meyer J.-A., Guillot A. Simulation of Adaptive Behavior in Animats: Review and Prospects / In Meyer J.-A., Wilson S. W. (Eds.) Proceedings of the First International Conference on Simulation of Adaptive Behavior: from animals to animats. Cambridge, MA: MIT Press, 1991. pp. 2-14.  
See also: <http://www.poleia.lip6.fr/ANIMATLAB/#Publications>.
89. Тарасов В.Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика. М.: Эдиториал УРСС, 2002. 352 с.
90. Нейроинформатика / А.Н. Горбань, В.Л. Дунин-Барковский, А.Н. Кирдин и др. – Новосибирск: Наука. Сибирское предприятие РАН, 1998.  
См. также: <http://www.neuroschool.narod.ru/books/neurinf.html>

Статья поступила 20 декабря 2005 г.

После доработки 11 января 2006 г.

# Метод эластичных нейронных сетей и его робастная трактовка

Ососков Г.А.

Объединенный институт ядерных исследований, Дубна, Московской обл., soskov@jinr.ru

*Аннотация.* После краткого обзора методов обработки данных современных экспериментов физики высоких энергий описаны методы эластичных нейронных сетей, которые затем трактуются как частный случай робастного (устойчивого к засорениям выборки) подхода к оцениванию зависимостей. Описан формализм робастного подхода с выводом оптимальных весов для случаев равномерного засорения выборки. На базе исследования совместной оценки параметров положения и масштаба получен экономичный алгоритм робастного оценивания.

## 1. Введение

Распознавание кривых линий на сложном фоне шумовых точек и близких соседних кривых является классической проблемой распознавания образов во многих важных научно-технических приложениях. Особенно сложными такие проблемы оказываются в ядерной физике высоких энергий, где по данным измерений необходимо не только распознать траекторию отдельной заряженной частицы среди тысяч траекторий других частиц, но и дать оценку ее параметров, в том числе с максимально возможной точностью вычислить импульс частицы, определяемый кривизной ее траектории в магнитном поле. Хотя эти проблемы распознавания и последующей подгонки траекторий (треков) заряженных частиц имеют почти полувековую историю, методы их решения существенно менялись по мере эволюции экспериментальных установок от пузырьковых, где несколько треков регистрировались на стереофотографиях, до современных экспериментов с тяжелыми ионами, в которых рождаются тысячи треков, фиксируемых быстрыми электронными детекторами прямо в памяти компьютеров в виде массивов измеренных координат (см. рис.1). К тому же кругу проблем относится и обработка данных с детекторов черенковского излучения типа RICH, в которых черенковские фотоны, образующиеся при пролете частицы со скоростью выше скорости света в данной среде, отражаются от сферического зеркала на фотоматрицу, где они регистрируются в виде характерных колец (см. рис. 2).

Главные требования к обработке в современных экспериментах: максимальная скорость вычислений при предельно достижимой их точности и высокая эффективность методов оценки физических параметров, интересующих экспериментаторов. Реализация этих требований при наличии чрезвычайно сложной текстуры и зашумленности экспериментальных данных неизбежно натолкнулась на ограниченность традиционно применяемых классических комбинаторных методов, кластерного анализа и подгонки по методу наименьших квадратов, которые в этих условиях уже не обеспечивали либо точности, либо

### Метод эластичных нейронных сетей и его робастная трактовка

скорости вычислений, либо высокой эффективности оценок параметров, либо всего этого вместе [1].

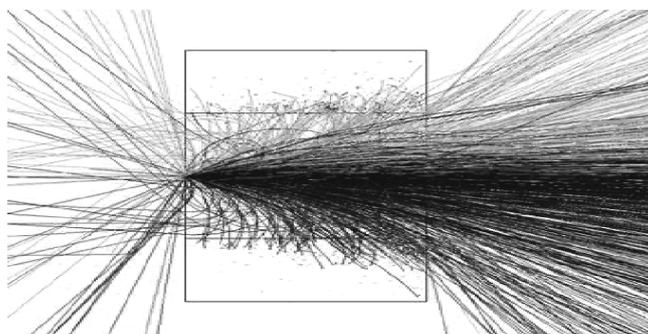


Рис.1. Смоделированное событие множественного рождения при центральном столкновении атомов золота при энергии 25 AGeV/c. Эксперимент CBM, Германия.

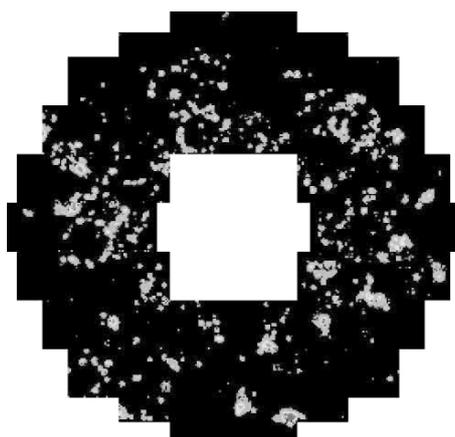


Рис.2. Снимок с фотоматрицы RICH детектора с изображениями колец черенковского излучения, зарегистрированных на шумовом фоне. Эксперимент CERES/NA-45, CERN.

Возьмем в качестве примера проблему распознавания треков. Если раньше при множественности детектируемых событий в несколько треков на событие основной методикой распознавания было «прослеживание в дороге» (Road Guidance – RG, см. статью Р.Стренда [2]), т.е. кластеризация соседних точек по угловой близости отрезков, их соединяющих, то для множественности в несколько сот треков на событие потребовались новые, более совершенные методы распознавания. Вполне естественным было обращение к аппарату искусственных нейронных сетей (ИНС). Однако помимо многих нейросетевых (и поэтому быстрых) реализаций того же RG-метода с помощью многослойных перцептронов [3], принципиально новым явился предложенный Денби и Петерсоном (ДП) метод, основанный на использовании полносвязных нейронных сетей

типа Хопфилда [4]. В ДП методе трек рассматривается как последовательность сегментов, соединяющих соседние измеренные точки. Идея метода состоит во введении бинарных нейронов  $S_{ij}$ , равных единице, если трек соединяет точки  $i$  и  $j$ , и нулю в противном случае. Принадлежность точек треку, как к некой гладкой кривой, обеспечивалась убывающей функциональной зависимостью синаптических весов  $w_{ijkl}$  для пары нейронов  $S_{ij}$  и  $S_{kl}$  от угла между сегментами трека, соответствующими этим нейронам, и от расстояния между этими сегментами. Энергетическая функция ИНС как билинейный функционал от  $S_{ij}$  и  $S_{kl}$  строилась так, чтобы набор гладких треков соответствовал ее минимуму. Несмотря на очевидную чувствительность такого метода к наличию шумовых измерений, ДП метод и аналогичный подход с так называемыми роторными ИНС [5], благодаря применению различных обобщений и системы пороговых обрезающих нашли весьма широкие применения в экспериментальной физике [6,7].

Тем не менее, как было продемонстрировано в исследовании [8], в современных экспериментах с множественностью в несколько тысяч треков на событие нейронные сети, как непосредственный инструмент распознавания треков, перестают удовлетворять экспериментаторов. Это может быть объяснено прежде всего тем, что при использовании методов типа ДП не применяется информация о параметрической форме трека, которая обычно известна заранее. ИНС используется только для решения комбинаторной задачи поиска точек и присоединения их к «своему» треку, в то время как, например, учет его локальной кривизны, известной из его уравнения, мог бы позволить сделать этот поиск более надежным. Недаром столь удачным в приложениях оказался опыт учета параметрической формы трека в синаптических весах роторной ИНС, позволивший выполнить обработку реальных экспериментальных данных [9] и провести анализ ионограмм [10].

Идея объединения этапов прослеживания и фитирования треков была предложена в работе [11] в виде **эластичной нейронной сети**, использующей гибкий шаблон, т.е. уравнение трека, зависящего от вариаций параметров таким образом, чтобы, изгибаясь при их изменении, кривая, описываемая этим уравнением (ее еще называют: *деформируемый шаблон*, *эластичная рука* и т.д.), прошла как можно ближе к «своим» точкам, измеренным на треке, соответствующем этому шаблону, «отталкиваясь» от «чужих». В качестве средства для реализации этой идеи были взяты бинарные нейроны, равные единице, если точка принадлежала шаблону и нулю в противном случае. Минимизация энергетической функции сети выполнялась при ограничении, учитывающем наличие шумовых точек, не принадлежащих никакому шаблону,

В данной статье после изложения метода эластичных нейронных сетей и близкого к нему подхода [8], также реализующего идею о совмещении этапов поиска и фитирования кривых, дано краткое введение в теорию робастных методов подгонки и показано, что оба метода «эластичного» трекинга могут рассматриваться как частные случаи робастного подхода.

## 2. Эластичные нейронные сети

Как отмечалось выше, идея работы [11] по объединению этапов распознавания и фитирования кривых привела к подходу учета известного уравнения трека путем создания эластичного шаблона и использования нейронов для сортировки точек. Такой подход немедленно порождает естественные вопросы о неизвестном числе шаблонов, инициализации их параметров и об организации подгонки сразу всех кривых (треков).

Проблема нахождения шаблонов (templates) и грубых начальных значений их параметров требует специального рассмотрения. Мы ограничимся здесь ссылкой на работы [11-13], в которых использовались эластичные шаблоны, и отметим, что всюду для их поиска применяется тот или иной вариант преобразования Радона-Хафа [14,15]. При практической реализации это преобразование сводится к гистограммированию в пространстве параметров с последующим поиском максимального значения. В силу приближенного характера процедуры нахождения шаблонов их число могло превышать число реальных треков в событии. Появляющиеся из-за этого лишние, искусственные треки должны быть удалены впоследствии в ходе специальной процедуры отбраковки.

Для поиска сразу всех треков, порожденных взаимодействием в однородном магнитном поле, в работе [11] предложено минимизировать следующий функционал:

$$E(\{S_{ia}\}; \vec{\pi}) = \sum_{i=1}^N \sum_{a=1}^M S_{ia} D_{ia}(\vec{x}, \vec{\pi}) + \lambda \sum_{i=1}^N \left( \sum_{a=1}^M S_{ia} - 1 \right)^2 \quad (1)$$

Здесь  $\pi$  – вектор параметров винтовой линии (геликоида), описывающей траекторию движения  $a$ -й частицы в пространстве;  $S_{ia}$  – бинарный нейрон, определяющий принадлежность  $i$ -й точки к  $a$ -му шаблону трека по правилу:  $S_{ia} = 1$ , если  $i$ -я точка принадлежит  $a$ -му треку,  $S_{ia} = 0$ , в противном случае,  $D_{ia}(\mathbf{x}, \pi)$  – квадрат расстояния от точки  $\mathbf{x}$  до  $a$ -го трека.

Минимизация  $E(S_{ia}; \pi)$  ведется при условии, что каждая точка может принадлежать только одному треку или ни одному ( $\sum_{ia} S_{ia} = 0$ ). В последнем случае функционал штрафует на величину  $\lambda$ , что определяет критическое расстояние  $\sqrt{\lambda}$ , до которого точки энергетически выгодно включать в трек, а после – считать такую точку шумовой ( $S_{ia} = 0, \forall i$ ).

Дальнейшие вычисления по поиску глобального минимума функционала (1) ведутся в соответствии с обычной схемой сетей Хопфилда: сеть термализуется (используется метод отжига [17], т.е. вводится модельная температура  $T$ , характеризующая случайный шум в процессе поиска минимума функционала (1), и  $T$  постепенно уменьшается) и применяется теория среднего поля. Для последовательности уменьшающихся температур  $T_k > T_{k-1} > \dots > T_0$  метод наискорейшего спуска дает пошаговый итерационный алгоритм поиска минимума [11]:

$$\Delta\pi_a^{(k)} = -\eta_a^{(k)} \sum_i V_{ia} \frac{\partial D_{ia}}{\partial \pi_a^{(k)}}, \quad (2)$$

где

$$V_{ia} = \frac{e^{-\beta D_{ia}}}{e^{-\beta \lambda} + \sum_{b=1}^M e^{-\beta D_{ib}}} \quad (3)$$

– факторы Поттса,  $\beta = 1/T$ .

Метод эластичных нейросетей был успешно применен в наших работах [12,13] для обработки данных, параметризуемых уравнением окружности: колец черенковского излучения и треков в однородном магнитном поле.

В первой работе [12] методом эластичных нейросетей (ЭН) осуществлялся поиск черенковских колец с одновременной оценкой их параметров по данным RICH детектора CERES/NA-45 (см. пример на рис.3). Глобальный поиск велся безо всяких априорных сведений о центрах и радиусах колец. Эта начальная информация была получена с помощью сведения трехмерного преобразования Хафа к двумерному и одномерному. Вначале выполнялся перебор допустимых триплетов измеренных точек (т.е. таких троек точек, через которые можно провести окружность с радиусом и центром, удовлетворяющим заданным неравенствам). Идея метода основана на том, что все точки, принадлежащие некоей окружности, должны отображаться в одну точку в пространстве параметров, так что, суммируя, мы должны получить пик в этом месте. Примеры работы алгоритма в этих двух режимах представлены на рис.3. Отсев ложных колец, получившихся из-за того, что преобразование Хафа нашло несколько лишних шаблонов (см. пример в левом нижнем углу рис.3), мог быть легко выполнен в последующем анализе с использованием критерия  $\chi^2$  и учетом статистики по числу и угловому распределению фотонов в кольце.

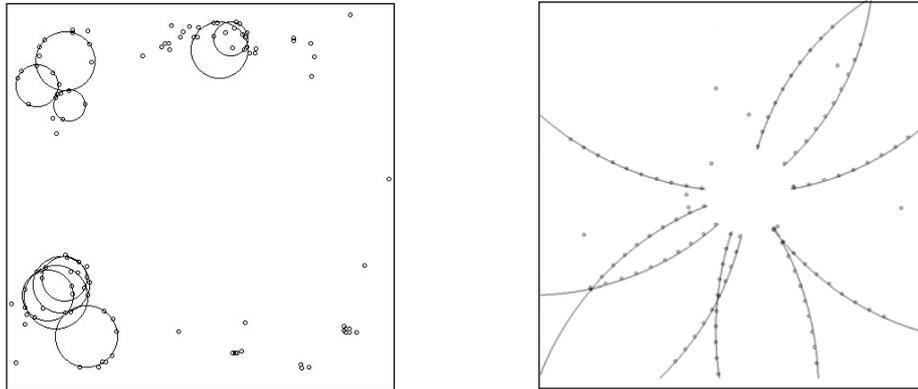


Рис.3. Примеры результатов работы метода эластичных нейросетей из [12]: слева – распознавание черенковских колец; справа – распознавание треков в магнитном поле.

Следует, однако, сделать важное замечание, касающееся издержек глобальности метода ЭН: при попытках увеличить число колец или множественности, т.е. числа треков, в районе поиска выше 12-13, метод переставал работать (не говоря уж о быстром росте машинного времени). Частично это происходило из-за ошибок в преобразовании Хафа, но главным образом из-за проблем с минимизацией функционала.

Поэтому во второй работе [13], касавшейся применения метода ЭН в задаче распознавания и определения параметров треков в системе дрейфовых трубок, мы отказались от глобального прослеживания сразу всех треков события и искали треки по данным преобразования Хафа по очереди. При прохождении частицы сквозь дрейфовую трубку регистрируется координата ее центра и расстояние от него до траектории прошедшей частицы (см. рис. 4). Главная неприятность здесь состоит в том, что знание этого расстояния не позволяет определить, слева или справа от центра прошла частица, возникает проблема *лево-право неопределенности*.

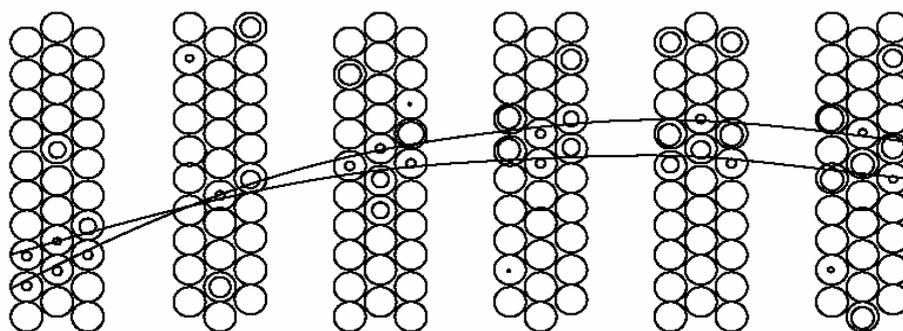


Рис.4. Типичное событие, детектированное дрейфовой камерой в магнитном поле.

Для учета этой двойственности в работе [13] вместо бинарных нейронов  $S_{ia}$ , используемых для определения принадлежности точки треку, был введен двумерный вектор-нейрон  $S_i = (s_i^+, s_i^-)$  с допустимыми значениями (1,0), (0,1), (0,0), что увеличило размерность минимизируемой функции и привело к появлению двух факторов Поттса.

Основой схемы минимизации явилась так называемая процедура имитационного отжига, для описания которой мы обратимся к работе [8] Джиласси и Харландера (ДХ). Для решения задачи одновременного поиска и подгонки треков ДХ предложили свой метод эластичного трекинга (ЕТ), исходя из другой идеи, не связанной с ИНС (хотя в [16] они показали, как можно трактовать ЕТ-метод как один из вариантов Хопфилдовой ИНС).

ДХ также использовали идею гибкого шаблона, но описывали его, исходя из физического рассмотрения взаимодействия положительно заряженного шаблона и отрицательно заряженных пространственных точек, измеренных на треке. Чем лучше гибкий шаблон пройдет по точкам, тем меньше будет энергия их взаимодействия. Пусть заряд для шаблона трека распределен с плотностью

$\rho_T(r)$ , а заряд множества измеренных точек имеет плотность  $\rho(r')$ . Вычисляя энергию взаимодействия  $E$  между двумя этими зарядами, получаем

$$E = - \int dr' dr \rho_T(r) V(r - r') \rho(r') \longrightarrow \min, \quad (4)$$

где  $V$  – потенциал, зависящий от расстояния измеренных точек до шаблона. ДХ выбрали потенциал Лоренца

$$V(x, t) = \frac{w^2(t)}{x^2 + w^2(t)} \quad (5)$$

с шириной, зависящей от температуры:  $w(t) = a + (b - a)\exp(-t/T)$ , где  $t$  – температура,  $T$  – температурная константа,  $a$  – максимальное расстояние, на котором точки еще приписываются к данному шаблону,  $b \approx \sigma_{res}$  – точность пространственного разрешения детектора. Очевидно, что  $b \ll a$ . Учитывая дискретность измерений, получаем вместо интеграла в (4) сумму

$$E(\pi, t) = - \frac{1}{N} \sum_i^N \frac{w^2(t)}{(\vec{x}_i - \vec{r}(\pi, \vec{x}))^2 + w^2(t)} \quad (6)$$

Здесь  $N$  – число точек на треке,  $\mathbf{x}_i$  и  $\mathbf{r}_i$  –  $i$ -я точка, измеренная в пространстве, и ее расстояние до шаблона, а  $\boldsymbol{\pi}$  как и раньше, вектор параметров геликоиды, описывающей траекторию движения частицы в однородном магнитном поле.

Функционал в формуле (6) зависит от точек только одного трека, хотя как и в предыдущем подходе, можно осуществлять одновременную подгонку сразу всех треков. Тем не менее сами ДХ этого не рекомендуют. Наличие зависимости потенциала от температуры позволяет применять технику симулированного отжига (simulated annealing) [17]: вначале, когда параметры известны только в грубом приближении, берется потенциал настолько широким (высокая температура  $t$ ), чтобы наверняка захватывать все измеренные точки. При этом минимизируемая функция  $E(\boldsymbol{\pi}, t)$  выполаживается так, что остается только один минимум (см. рис. 5). Хотя он может быть расположен и несколько в стороне от искомого глобального минимума, но может служить начальным приближением на следующем этапе поиска при понижении температурного параметра. При постепенном понижении  $t$  до нуля этот процесс сойдется к минимуму исходной функции.

ЭТ метод в такой постановке был успешно применен в работе [18] для распознавания треков по данным проекционной камеры установки STAR в условиях, предельных по загрузке (до 5 тыс. треков на событие), а также в ряде других экспериментальных исследований [19,20].

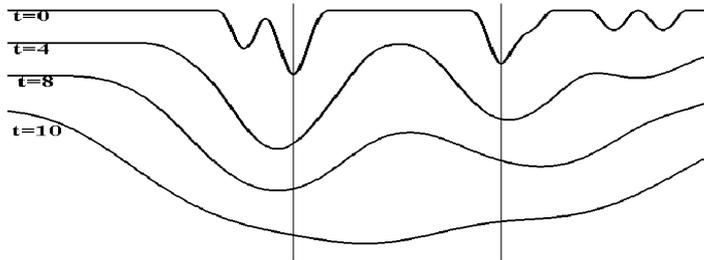


Рис.5. Вид  $E(\pi, t)$  при разных температурах.

### 3. Робастные методы оценки параметров и их применения

Наиболее распространенным методом оценки параметров функциональных зависимостей по данным измерений является метод наименьших квадратов (МНК). Возьмем в качестве простейшего примера линейную модель трека вне магнитного поля в координатной проекции  $(x, z)$  :  $y_i = y_0 + t_y z_i + \varepsilon_i$ ;  $i=1, 2, \dots, n$ , где  $y_i$  –  $i$ -е измерение,  $z_i$  – координата  $i$ -й плоскости детектора,  $\varepsilon_i$  – ошибка измерения, имеющая по предположению нормальное распределение с нулевым средним и среднеквадратичным значением  $\sigma_i$ . Обозначим вектор неизвестных параметров как  $\mathbf{p}^T = (y_0, t_y) = (p_1, p_2)$ . Для их определения следует найти минимум суммы квадратов невязок  $\varepsilon_i$ , которые в общем случае неравноточных измерений должны быть нормированы на их стандартные отклонения  $\sigma_i$ :

$$S(\bar{\mathbf{p}}) = \sum_i w_i \varepsilon_i^2 \implies \min_{\bar{\mathbf{p}}}, \quad (7)$$

где  $w_i = 1/\sigma_i^2$  – веса измерений. Чтобы найти минимум  $S(\mathbf{p})$  приравняем нулю производные этого функционала по параметрам. Это даст систему нормальных МНК-уравнений:

$$\frac{\partial S}{\partial p_j} = 2 \sum_i w_i \varepsilon_i \frac{\partial \varepsilon_i}{\partial p_j} = 0, \quad j = 1, 2 \quad (8)$$

решения которой дают искомые оценки параметров. МНК-оценки обладают целым рядом полезных свойств, гарантированных тем, что при непременном условии нормальности распределения невязок  $\varepsilon_i$  МНК оказывается частным случаем общего метода оценивания, называемого методом максимального правдоподобия (ММП), который действует при произвольных законах распределения оцениваемых параметров. Нарушение нормальности величин  $\varepsilon_i$  в трековых измерениях обычно является следствием засорения выборки посторонними измерениями как шумовыми, так и от соседних треков и неизбежно ведет к значительным искажениям оцениваемых параметров.

Квадратичность функционала (7) ведет к тому, что точки, далеко отстоящие от подгоняемой кривой, могут дать неоправданно большой вклад в функционал и привести к значительной потере точности оценок параметров. Чтобы избежать этого, следует учитывать измерения только из непосредственной окрестности подгоняемой функции, придавая остальным меньшие значения или вообще пренебрегая ими. Такую идею можно реализовать, придавая каждому измерению специальный вес, значение которого убывает с ростом невязки  $\epsilon_i$ , т.е. расстояния до подгоняемой кривой. Этот подход, называемый **робастным**<sup>1</sup>, был дан П. Хьюбером [21], предложившим, однако, иной метод его реализации. Предложение Хьюбера сводится к некоторому обобщению метода максимального правдоподобия. Подчеркивая эту связь с ММП, Хьюбер назвал свой подход М-оцениванием. С математической точки зрения предлагалось перейти от суммы квадратов в (7) к сумме некоторых симметричных **функций вклада**  $\rho(\epsilon)$ , которые также зависят от отклонения  $\epsilon$  точки от прямой, но растут медленнее, чем квадратичная парабола. Теперь минимизируемый функционал будет выглядеть так:

$$L(p) = \sum_i \rho(\epsilon_i). \quad (9)$$

Забываясь о выпуклости этого функционала, чтобы обеспечить единственность решения задачи его минимизации, П.Хьюбер предложил неограниченную функцию вклада  $\rho(\epsilon)$ , составленную из параболы в малой окрестности нуля, продолженной далее отрезками прямых. Однако это предложение оказалось мало устойчивым к сильным засорениям, так как неограниченная  $\rho(\epsilon)$  по-прежнему придавала неоправданно большой вес точкам, далеко отстоящим от подгоняемой линии. Более популярным оказался весовой подход Дж. Тьюки [22], упрощавший одновременно и вычислительную проблему минимизации функционала (9). Если продифференцировать (9) по параметрам, получим систему нелинейных уравнений

$$\frac{\partial L(p)}{\partial p} = \sum_{i=1}^n \frac{\partial \rho(\epsilon_i)}{\partial \epsilon_i} \frac{\partial \epsilon_i}{\partial p} = 0,$$

которая в обозначениях

$$w(\epsilon) = \frac{1}{\epsilon} \frac{\partial \rho(\epsilon)}{\partial \epsilon} \quad (10)$$

превращается в систему

$$\frac{\partial L(p)}{\partial p} = \sum_{i=1}^n w(\epsilon_i) \frac{\partial \epsilon_i}{\partial p} \epsilon_i = 0 \quad (11)$$

<sup>1</sup> robust (англ.) – крепкий, здоровый, в статистике – не чувствительный к шумам.

казалось бы полностью совпадающую с нормальными уравнениями МНК (8), но с тем важным отличием, что числовые весовые коэффициенты заменены на **весовые функции**  $w(\varepsilon)$ , которые приходится перевычислять на каждой итерации получившейся итеративной процедуры, названной процедурой Флетчера-Гранта-Хеблена (ФГХ) [23]. На каждой итерации ФГХ-процедуры выполняется взвешенный МНК, но с функциональными весами. Если нет каких-либо априорных соображений по выбору начальных значений весов, то можно инициировать ФГХ процедуру с помощью обычного МНК, взяв в качестве весов единицы  $w(\varepsilon) \equiv 1$ .

Важным аспектом ФГХ-процедуры является выбор весовой функции. Из формулы (10) следует, что в обычном МНК весовая функция является единичной константой, в то время как в весовой функции, соответствующей функции вклада Хьюбера, эта константа за пределами малого центрального участка спадает по гиперболе. Отсутствие теоретического обоснования по выбору  $w(\varepsilon)$  вызвало массу эвристических предложений (см. обзор и анализ в [24]), среди которых одним из наиболее эффективных с точки зрения шумоподавления оказалась весовая функция Тьюки [22], называемая **бивесовой** за применение биквадрата невязок:

$$w(\varepsilon) = \begin{cases} \left[ \frac{1 - \left( \frac{\varepsilon}{c_T \cdot \sigma} \right)^2}{1 - \left( \frac{\varepsilon}{c_T \cdot \sigma} \right)^2} \right]^2, & \text{если } |\varepsilon| < c_T \cdot \sigma \\ 0 & \text{в противном случае} \end{cases} \quad (12)$$

Выбор константы  $c_T$  в этой формуле определяет ширину коридора вокруг подгоняемой линии, за пределами которого все точки игнорируются. Использование  $c_T$  как температуры в схеме симулированного отжига позволяет справиться с проблемой локальных минимумов при минимизации функционала (9). Явное выражение для оптимальной весовой функции удалось вывести для случая равномерного засорения [25]. Для описания засоренного распределения воспользуемся моделью "больших ошибок" (*gross-error model*) Тьюки [26]:

$f_c(\varepsilon) = (1 - g) \varphi(\varepsilon) + g h(\varepsilon)$ , где  $\varphi(\varepsilon)$  – гауссово распределение с нулевым средним и среднеквадратичным значением  $\sigma_i$ ,  $g$  – параметр засорения, а  $h(\varepsilon) = 1/\Delta$  – плотность равномерного распределения в отрезке  $(-\Delta/2, \Delta/2)$ , где  $\sigma \ll \Delta$ . Составим логарифмическую функцию правдоподобия для такого распределения, приняв для простоты однопараметрическую зависимость  $L(p)$ :

$$L(p) = \ln \prod_i f(\varepsilon_i) = \sum_i \ln \left[ \frac{1-g}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right) + \frac{g}{\Delta} \right].$$

Приравнивая к нулю ее производную по параметру  $p$ , мы получим уравнение правдоподобия, которое примет вид (11), если обозначить

$$w_{opt}(\epsilon) = \frac{1 + c}{1 + c \cdot \exp\left(\frac{\epsilon^2}{2\sigma^2}\right)}, \quad c = \frac{g}{1 - g} \frac{\sqrt{2\pi}\sigma}{\Delta} \quad (13)$$

Как видно, единственный параметр  $c$  определяется засоренностью данных не во всем диапазоне измерений  $\Delta$ , а только в узкой полосе шириной  $\sigma\sqrt{2\pi}$  вокруг подгоняемой кривой, где и сосредоточено большинство полезных измерений. Часто константу  $c$  можно оценить заранее, исходя из устройства детектора.

Из соображений ускорения вычислений была предложена полиномиальная аппроксимация функции (13) многочленом 4-го порядка, которая оказалась ничем иным, как вышеупомянутым бивесом Тьюки (12).

Уже отмечалась вычислительная проблема минимизации функционала (9), который в силу своей нелинейности может иметь несколько локальных минимумов. Их число зависит не только от оцениваемых параметров  $p$ , но и от разброса точек вокруг подгоняемой кривой, т.е. от оценки среднеквадратичного среднего  $\sigma$ . Исследование совместной оценки  $p$  и  $\sigma$  было проведено в нашей работе [25] на основе рассмотрения геометрических свойств  $L(p, \sigma)$  как функции этих двух параметров. Результаты можно продемонстрировать на простой однопараметрической модели, когда множество локальных условных минимумов  $L(p, \sigma)$  для всех фиксированных  $\sigma > 0$  образует некий набор гладких кривых в полуплоскости  $\{(p, \sigma): \sigma > 0\}$  (см. рис. 6). Обозначим их  $\gamma_1, \gamma_2, \dots, \gamma_m$ . Тогда найдется достаточно большое  $\sigma_1 > 0$ , для которого в полуплоскости  $\{(p, \sigma): \sigma > 0\}$  существует только одна из этих кривых, обозначенная как  $\gamma_1$ . Она неограниченна и при  $\sigma \rightarrow \infty$  асимптотически сходится к среднему всех измеренных значений.

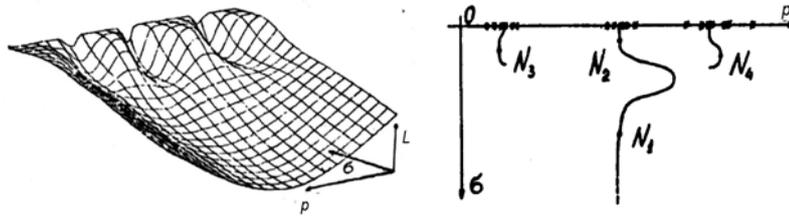


Рис.6. Слева: пример поверхности  $L(p, \sigma)$  с характерными оврагами; справа: кривые, соответствующие дну каждого из этих оврагов.

Пример на рис.6 показывает, что возможные случайные сгущения измерений приводят к появлению характерных «оврагов», иногда весьма «кривых», число которых возрастает с уменьшением  $\sigma$ . Детальное теоретическое изучение совместных оценок  $p$  и  $\sigma$  в работе [26] и проведенные расчеты на модельных данных показали, что всегда на кривой  $\gamma_1$  существует по крайней мере одна совместная оценка, являющаяся решением системы уравнений

$$\begin{cases} \partial L / \partial p = 0; \\ \partial^2 L / \partial p^2 + \sigma^{-1} \partial L / \partial \sigma = 0. \end{cases} \quad (14)$$

с условиями достижения минимума:  $\partial^2 L / \partial p^2 > 0, \partial / \partial \sigma (\partial^2 L / \partial p^2 + \sigma^{-1} \partial L / \partial \sigma) > 0$ .

Если в задаче выполнено условие  $\sigma \ll \Delta$ , то для достаточно большой  $\sigma_0$ , равной, например, размаху выборки  $x_n - x_1$ , мы окажемся как раз на кривой  $\gamma_1$ , так что, начиная с асимптотического значения  $p_0 = (\sum_i x_i)/n$ , мы можем постепенно двигаться вдоль этой кривой до достижения  $\sigma = \sigma_{min}$ , которая обычно оценивается заранее как среднеквадратичная ошибка ионизированного трека. В реальных вычислениях вместо перевычисления  $\sigma^2$  на каждой итерации как  $\sigma^2 = (\sum_i w_i \varepsilon_i^2) / (\sum_i w_i)$  мы, начиная с  $\sigma_0$ , вычисляем  $\sigma^{(k)} = (1-\delta) \sigma^{(k-1)}$  с малым  $\delta=0.05$ .

#### 4. Приложения и выводы

За последние годы в Объединенном Институте Ядерных Исследований был проведен целый ряд работ по анализу экспериментальных данных на базе применения робастного подхода. Робастные методы подгонки, несмотря на свою кажущуюся простоту, показали высокую эффективность в таких приложениях, как определение вершины взаимодействия с помощью вершинного детектора установки CERES, состоящего всего из двух координатных плоскостей [27], распознавание колец черенковского излучения и проведение идентификации частиц [28,29], поиск треков в экспериментах с высокой множественностью событий [18,30], решение задач калибровки и алайнмента трековых детекторов [31,32]. Отметим развитие теории весовых функций как на случаи неравномерного засорения выборки [30], так и на случаи, когда кроме координат, в детекторе фиксируется и амплитуда сигнала [29].

Если теперь сравнить процедуры минимизации функционалов (1), (6) и (9) с вычислительной точки зрения, то становится очевидным, что минимизация функционала (1) эластичных нейросетей с помощью итеративного метода (2) сводится, по сути, к взвешенному методу наименьших квадратов с весами (3). Потенциал Лоренца в методе ДХ также может рассматриваться как робастная функция вклада, а вся схема симулированного отжига практически полностью совпадает с вышеописанным методом совместного робастного оценивания параметров положения и масштаба.

Эластичные методы хотя и весьма эффективны в приложениях, но опираются на различные физические аналогии и эвристические обоснования, в то время как робастный подход основан на ясных статистических выводах и по существу всегда может быть сведен к тому или иному эластичному методу как к частному случаю.

#### Литература

1. Г.А.Ососков, А.Полянский, И.В.Пузынин, Современные методы обработки экспериментальных данных в физике высоких энергий // ЭЧАЯ, т.33, в.3, 2002, 676-745.
2. Р.Стренд, Распознавание оптических образов при экспериментах на трековых камерах с частицами высоких энергий // в сб. «Распознавание образов при помощи цифровых вычислительных машин», под ред. Л.Хармона, пер с англ., М., МИР, 1974,15-37.
3. New computing techniques in physics research III // in: Proc. AINENP, Ed. K.Becks. Perret-Gallix, World-Sci, Singapore, 1994.

4. И.В.Кисель, В.Н.Нескоромный, Г.А.Ососков, Применение нейронных сетей в экспериментальной физике // ЭЧАЯ, т.24,(1993), вып. 6,1551-1595.
5. L.Gislen, C.Peterson, B.Soderberg, Rotor neurons – basic formalism and dynamics // Neural Computation, 4, 1992,737.
6. G.Ososkov, Robust tracking by cellular automata and neural network with non-local weights // in “Applications and Science of Artificial Neural Networks”, S. K. Rogers, D. W. Ruck, Editors, Proc. SPIE 2492, (1995) 1180-1191.
7. Г.А.Ососков и др., Использование нейронных сетей для улучшения интерпретации эксперимента EXCHARM // Матем. Моделир., т.11, в.10, 1999, 116-126.
8. M.Gyulassy and M.Harlander, Elastic tracking and neural networks algorithms for complex pattern recognition // Comp.Phys.Comm. 66, 1991, 31-46.
9. S.Baginyan et al, Tracking by a Modified Rotor Model of Neural Network // Comp. Phys. Commun. (1994) v.79, 165-178.
10. Galkin et al, Feedback neural networks for ARTIST ionogram processing // Radio Science v.31, № 5 (1996) 1119-1128.
11. M. Ohlsson, C. Peterson, A. Yuille, Track finding with deformable templates - the elastic arms approach // Comput. Phys. Commun., 71, (1992), 77.
12. L. Muresan, R. Muresan, G. Ososkov, Yu. Panebratsev, Deformable Templates for Circle Recognition // JINR Rapid Communications, I[81]-97, Dubna, 1997, 27-44.
13. S.Baginyan, G.Ososkov, Finding tracks detected by a drift tube system // Comp.Phys.Comm, v. 108, No 1 (1998) 20-28.
14. G.Ososkov, V.Palichik, E.Tikhonenko, Robust Technique with Sub-Optimal Weight Function for Track Fitting in CMS Muon Strip Chamber // Abst. of Europhysics Conf. on Computat. Phys., Vol. 22F, EPS, Granada, Spain (1998) 323-324.
15. Hough P. V. C. A Method and Means for Recognizing Complex Patterns //, US Patent: 3,069,654, 1962.
16. Toft P. The Radon Transform. Theory and Implementation // Ph.D. Thesis, Department of Mathematical Modelling, Section for Digital Signal Processing, Technical University of Denmark, 1996. (см. <http://www.ei.dtu.dk/staff/ptoft/ptoft.html>)
17. S.Kirkpatrick et al, Optimization by simulated annealing // Science 220 (1983), 671.
18. B.Lasiuk, D.Lyons, G.Ososkov, T.Ullrich, Development of an Elastic Tracking Package // Proc. of CHEP'98, Chicago (1998).
19. R.Blankenbecler, A unified treatment of track reconstruction and particle identification // Comput. Phys. Commun., 81, 1994, 335-342.
20. M.Lindstrom, Track reconstruction in the STLAS detector using elastic arms // Nucl. Instr. Meth. A357, 1995, 129-149.
21. P.Huber, Robust statistics // Wiley, N-Y (1981).
22. F.Mosteller, W.Tukey, Data analysis and regression: a second course in statistics, //Addison - Wesley, N-Y (1977).
23. R.Fletcher et al // Comp.J. v.72, No 3, (1971) p 276.
24. А.Астапов и др., Численный анализ робастных регрессионных методов // Сообщ. ОИЯИ Р9-85-492, Дубна, 1985.
25. G.Ososkov, Robust regression for the heavy contaminated sample // Proc. 2-nd International Tampere Conference in Statistics, (Tampere, Finland) (1987), 615-626.
26. J.Tukey, Introduction to Today's Data Analysis, Critical Evaluation of Chemical and Physical. Structural Information // Nat. Acad. Sci., Washington, 1974, 3-14.
27. H.Agakishiev et al, New Robust Fitting Algorithm For Vertex Reconstruction in the CERES Experiment // Nucl. Instr. and Methods, A394 (1997), 225-231.
28. G.Agakishiev et al, Cherenkov Ring Fitting Techniques for the CERES RICH Detectors // Nucl. Instr. and Methods, A371 (1996) 243-247.
29. N.Chernov, E.Kolganova, G.Ososkov, Robust Methods for the RICH Ring Recognition and Particle Identification // Nuclear Inst. Meth. A433 (1999) 274-278.

Метод эластичных нейронных сетей и его робастная трактовка

30. Golutvin, Yu.Kiryushin, S.Movchan, G.Ososkov, V.Palichik, E.Tikhonenko, Robust estimates of track parameters and spatial resolution CMS muon chambers // *Comp.Phys.Comm*, v.126/1-2, (2000) 72-76.
31. A.Bel'kov, A.Moshkin, G.Ososkov, Autocalibration of PC OTR chambers based on the robust fitting approach // *HERA-B Note 02-067*, DESY, Hamburg, 2002, 37 pp.
32. O.Barannikova et al, Specifications of SVT global alignment package, *STAR Note 0364*, BNL, 1998.

Статья поступила 12 декабря 2005 г.

# От моделей единичных нейронов к моделям популяций нейронов\*

Чижов А.В.<sup>1)</sup>, Турбин А.А.<sup>2)</sup>

<sup>1)</sup> Физико-технический институт им. А.Ф.Иоффе РАН, Санкт-Петербург,  
[anton.chizhov@mail.ioffe.ru](mailto:anton.chizhov@mail.ioffe.ru)

<sup>2)</sup> Санкт-Петербургский государственный политехнический университет, С.-Петербург,  
[Turbin\\_A\\_A@mail.ru](mailto:Turbin_A_A@mail.ru)

**Аннотация.** Составляется классификация существующих моделей нейронных популяций и моделей единичных нейронов, используемых в популяционных моделях. Приводятся уравнения типичных одномерных популяционных моделей, т.е. континуальных моделей, различающих состояние нейронов между спайками по одному параметру, в том числе двух моделей авторов статьи.

## 1. Введение

Согласно субъективному представлению авторов в любой конкретный процесс обработки информации мозгом вовлечены системы нейронов разного уровня. По-видимому, важными являются активность нейронной сети, состоящей из ограниченного числа нейронов, наиболее специфичных для данного процесса, и фоновая активность популяций нейронов, захватывающая большие области мозга. Поэтому для моделирования информационных процессов мозга необходимо сочетание моделей разного уровня подробности. В настоящее время известны модели единичных нейронов, модели нейронных популяций и модели электроэнцефалографических сигналов. Однако трудно проследить связь моделей разных уровней друг с другом. Значительная доля работ в области вычислительной нейрофизиологии посвящена построению моделей популяций нейронов. Целью этих работ является математическое описание активности большого числа подобных нейронов в ответ на получаемый общий стимул. Переход от описания отдельных клеток к описанию популяции, т.е. переход от микроскопического описания к макроскопическому, требует введения ряда допущений, не единственен и допускает несколько разных подходов. По этим причинам до сих пор не существует общепринятой популяционной модели, и нет общепринятого вывода уравнений таких моделей из уравнений единичных нейронов. В настоящей статье приводится классификация существующих популяционных моделей, приводятся уравнения наиболее типичных моделей и собственных моделей авторов статьи.

Рассмотрим путь построения моделей макроскопического уровня из моделей единичных нейронов. Общий формальный путь вывода уравнений активности популяции нейронов известен [1]. Он носит название *Density Approach*. Под популяцией будем понимать набор бесконечного количества подобных нейронов, различающихся либо своими параметрами, например, порогом возбуждения, либо флуктуациями входных воздействий, например, синап-

---

\* Работа выполнена при финансовой поддержке РФФИ (проект № 04-01-00048).

тических токов. В общем таком случае можно записать уравнение единичного нейрона в виде

$$\frac{d\vec{v}}{dt} = \vec{F}(\vec{v}) + \vec{S},$$

где  $\vec{v}$  – вектор состояния нейрона (для классических уравнений Ходжкина-Хаксли это – вектор с компонентами  $(V, m, h, n)$ , т.е. мембранным потенциалом и безразмерными параметрами активации и инактивации натриевых и калиевых каналов);  $\vec{F}$  – одинаковая для всех нейронов часть правой части уравнений,  $\vec{S}$  – оставшаяся часть, отражающая различия собственных параметров нейрона либо различия входных параметров.

Популяция нейронов может быть представлена распределением нейронной плотности в пространстве параметров состояния, т.е. в фазовом пространстве, составленном из компонентов вектора  $\vec{v}$ . Уравнение для нейронной плотности  $\rho(t, \vec{v})$  можно записать, пользуясь формализмом перехода от уравнения Ланжевена для единичной «частицы»-нейрона к уравнению Фоккера-Планка-Колмогорова для вероятности нахождения этой частицы в фазовом пространстве состояний этой частицы [2]. Таким образом, из уравнения единичного нейрона получаем уравнение для нейронной плотности типа

$$\frac{\partial \rho}{\partial t} = - \frac{\partial}{\partial \vec{v}} \cdot (\vec{F}(\vec{v}) \rho) + J,$$

где член  $J$  представляет собой сумму диссипативных членов, отражающих влияние  $\vec{S}$ , которые могут быть явно выписаны с помощью формул из [2].

Однако получившееся уравнение многомерно, что делает такой общий подход неэффективным в случае применения к сложным моделям нейрона. Редукция, в смысле понижения размерности этих уравнений, составляет основную задачу популяционного моделирования. В основе редукции популяционных уравнений лежит редукция базовой модели единичного нейрона.

## 2. Классификация моделей единичных нейронов

Кратко рассмотрим этапы редукции уравнений единичного нейрона (более подробное описание см. в [3,4]). В наиболее реалистичных моделях единичных нейронов воссоздаётся геометрия дендритного и аксонного деревьев и решаются распределённые уравнения типа Ходжкина-Хаксли, уравнения в частных производных [3,4,5] (рис.1). В более упрощённых моделях отдельные неветвящиеся волокна деревьев заменяются элементами (compartments), для каждого из которых решаются согласованные одноточечные уравнения типа Ходжкина-Хаксли. Согласно теории Ролла [5] отдельные ветви дендритного и аксонного деревьев могут быть заменены эквивалентными цилиндрами в допущении о пассивности процесса электротонического распространения и при соблюдении закона ветвления «3/2»:  $R^{3/2} = r_1^{3/2} + r_2^{3/2}$ , где  $R$  – радиус ветвящейся ветви,  $r_1, r_2$  – радиусы дочерних ветвей.

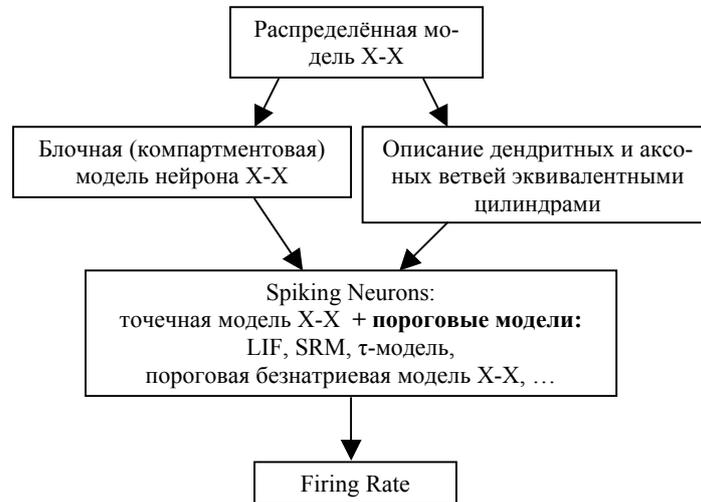


Рис.1. Иерархия моделей нейронов, используемых в популяционных моделях. Стрелки указывают направление упрощения моделей. Сокращения: «модель Х-Х» – модель Ходжкина-Хаксли, LIF – Leaked Integrate-And-Fire, SRM – Spike-Response-Model. Пояснения – в тексте.

Класс *Spiking Neurons* включает в себя модельные нейроны, генерирующие импульсы-спайки [3]. В этот класс входят одноточечные модели типа Ходжкина-Хаксли, линейные и нелинейные модели пороговых интеграторов (*Leaked Integrate-And-Fire (LIF)*). Уравнения линейного порогового интегратора записываются следующим образом:

$$\tau_m \frac{dV}{dt} = -V + RI(t), \quad V_{reset} < V < T, \quad (1)$$

где  $V$  – мембранный потенциал,  $I$  – внешний или синаптический ток,  $\tau_m$  – временная постоянная мембраны,  $R$  – входное сопротивление или полное сопротивление мембраны нейрона. Потенциал обретает значение  $V_{reset}$  при достижении порога  $T$ .

Относящаяся к этому же классу *Spike-Response-Model (SRM)* [3] расщепляет мембранный потенциал на его составляющие, отражающие форму спайка, после-спайковую релаксацию и постсинаптический потенциал. Постсинаптический потенциал вычисляется посредством интеграла-свёртки, ядро которого аппроксимирует свойства нейрона Ходжкина-Хаксли. А именно,

$$V(t) = \eta(t - \hat{t}) + \int_0^{t-\hat{t}} k(t - \hat{t}, t') I(t') dt', \quad (2)$$

где  $V$  – мембранный потенциал,  $\eta$  – спайковый и релаксационный потенциал, аппроксимирующий форму типичного спайка,  $k$  – ядро постсинаптического потенциала, вычисляющееся по полной модели нейрона,  $I$  – внешний или синаптический ток,  $\hat{t}$  – момент времени предыдущего спайка. Момент спайка определяется по условию достижения потенциалом порога  $T$ .

К классу *Spiking Neurons* принадлежит также предложенная авторами  $\tau$ -модель [6,7], характеризующая нейрон единственным параметром – предполагаемым временем до следующего спайка (см. п. 3.2.3). Модель порогового интегратора, *SRM*-модель и  $\tau$ -модель являются пороговыми моделями. Наименее редуцированная пороговая модель типа Ходжкина-Хаксли предложена в [8]; эта модель заменяет действие натриевых каналов динамическим пороговым критерием на потенциал.

Завершает последовательность упрощений модель типа *Firing Rate*. Она сводится к алгебраической зависимости частоты спайков от силы стимулирующего тока, рассчитанной для режима стационарной генерации  $f = f(I)$ .

### 3. Популяционные модели

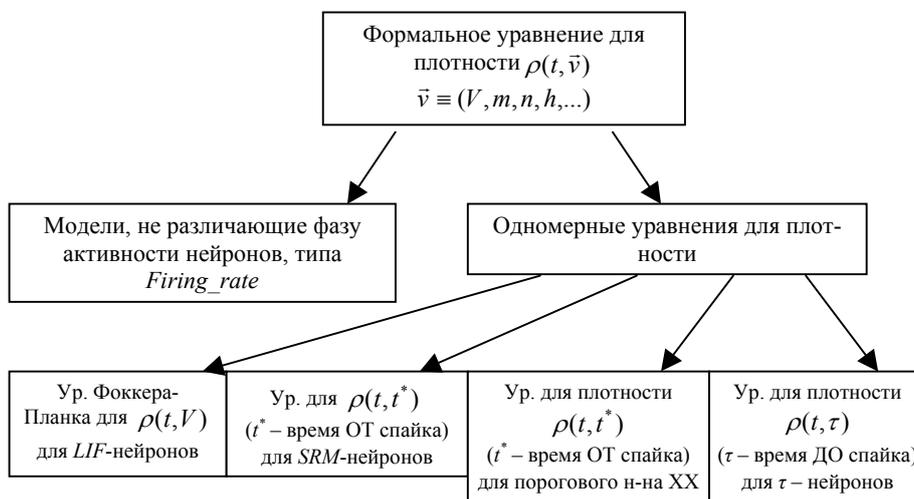


Рис.2. Иерархия популяционных моделей. Указанные модели охарактеризованы в соответствующих пунктах.

**3.1. Модель популяции Firing Rate.** Крайний случай редукции уравнений – популяционная модель типа *Firing Rate* (рис.2). Эта модель совпадает с аналогичной моделью для единичного «среднего» нейрона и сводится к алгебраической зависимости частоты спайков нейронов популяции (активности) от силы стимулирующего тока,

$$f = f(I). \quad (3)$$

Такая зависимость может быть рассчитана с помощью модели одного нейрона как зависимость частоты генерации спайков одним нейроном в стационарном случае. Однако, эта модель хорошо аппроксимирует активность популяции только в том случае, когда нейроны де-синхронизированы. В противном случае, например, при внезапно изменившемся, ступенчатом стимуле, первые спайки

нейронов синхронны и активность популяции может иметь ярко выраженные максимумы и минимумы, в то время как модель Firing Rate даёт постоянный уровень активности (рис.3). К этому классу моделей типа Firing Rate относятся модели типа Вилсона-Коуэна и модели ЭЭГ [9,10,11].

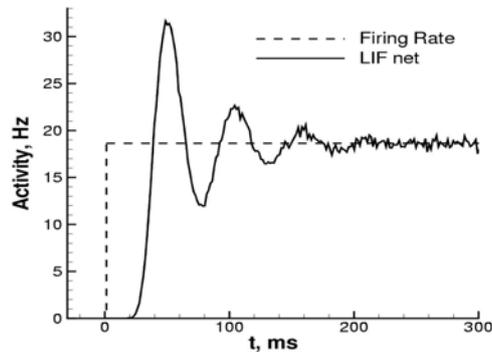


Рис.3. Ответ популяции нейронов-пороговых интеграторов и модели *Firing Rate* на внезапно возникающий в момент  $t=0$  стимул. Активность нейронов (Activity) представляет собой суммарное количество импульсов-спайков всех нейронов в единицу времени, нормированное на число нейронов.

**3.2. Одномерные модели распределения плотности.** Наиболее эффективными можно признать одномерные модели, в которых нейроны популяции различаются только по одному параметру состояния. При этом состояние нейронов может характеризоваться только этим одним параметром либо несколькими переменными, параметризованными этим единственным параметром. Ключевым для таких моделей является выбор этого единственного параметра.

**3.2.1. Уравнение Фоккера-Планка для  $\rho(t, V)$ .** Наиболее известный вариант выбора параметра состояния – потенциал  $V$  [1,12,13]. Существенным недостатком этого выбора является то, что потенциал не может однозначно характеризовать состояние нейрона, если потенциал немонотонен на межспайковом интервале. По этой причине такой подход работает только с упрощёнными моделями единичных нейронов, например с моделью линейного порогового интегратора. Пусть входной сигнал для нейрона, описываемого уравнением (1), состоит из тока, одинакового для всех нейронов популяции,  $I(t)$  и гауссова шума  $\eta(t)$ , т.е.

$$\tau_m \frac{dV}{dt} = -V + RI(t) + \eta(t), \quad \langle \eta(t) \rangle = 0, \quad \langle \eta(t) \eta(t') \rangle = \tau_m \sigma^2 \delta(t - t'), \quad (4)$$

где  $\delta(\cdot)$  – дельта-функция.

Это уравнение может описывать синаптически связанные или несвязанные нейроны популяции. В случае связанных нейронов ток  $I(t)$  должен представлять собой сумму внешнего тока, одинакового для всех нейронов, и среднего по популяции синаптического тока; флуктуации синаптического тока должны удовлетворять условиям гауссова шума  $\eta(t)$ . Тогда уравнение эволюции плотности распределения нейронов по значениям мембранного потенциала  $\rho(t, V)$  будет представлять собой уравнение Фоккера-Планка-Колмогорова, соответствующее

уравнению Ланжевена для одной «частицы», т.е. уравнению (4). Соответствующее уравнение Фоккера-Планка-Колмогорова имеет вид:

$$\tau_m \frac{\partial \rho}{\partial t} = \frac{\partial}{\partial V} [(V - RI) \rho] + \frac{\sigma^2}{2} \frac{\partial^2 \rho}{\partial V^2} + f \cdot \delta(V - V_{reset}) \quad (5)$$

$$f(t) = \frac{\sigma^2}{2} \frac{\partial \rho}{\partial V} \Big|_{V=T},$$

где  $f$  – активность популяции (firing rate), равная потоку нейронов через порог. Граничные условия:  $\rho(t, -\infty) = 0$ ,  $\rho(t, T) = 0$ .

**3.2.2. Уравнение для плотности  $\rho(t, t^*)$  для SRM-нейронов.** Другой разновидностью подхода *Density Approach* является подход *Refractory Density Approach* [14,3], рассматривающий распределение нейронов по времени, прошедшему от момента предыдущего спайка  $t^*$  (рис.4). Уравнение для нейронной плотности  $\rho(t, t^*)$  имеет вид:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial t^*} = -\rho H \quad (6)$$

$$\rho(t, 0) \equiv f(t) = \int_0^\infty \rho(t, t^*) dt^*,$$

где  $H$  – функция вероятности генерации спайка одним нейроном, так называемая *hazard function*. Обычно, это функция имеет в качестве своего аргумента мембранный потенциал, который сравнивается с пороговым значением, т.е.  $H=H(V(t), T)$ . Вид функции  $H$  зависит от предположения о разбросе параметров нейронов внутри популяции или от модели шумовой части входного сигнала. Простейший вид функции  $H$  – сигмоидная функция. Наиболее известна реализация *Refractory Density* подхода с SRM-нейроном [3], когда потенциал находится согласно уравнению (2), т.е.

$$V(t, t^*) = \eta(t^*) + \int_0^{t-t^*} k(t^*, t') I(t') dt' \quad (7)$$

Недостатки этой реализации подхода заключены в недостатках модели SRM, а также в неточности функции  $H$ . SRM-модель не может в полной мере аппроксимировать свойства реального нейрона. Так, синаптическое воздействие выражено только в действии тока в пренебрежении шунтирующим эффектом, а также модель даёт ошибку в чувствительности нейрона при подаче стимула в разную фазу его активности между спайками. К тому же, SRM-модель приводит к сложному интегральному виду уравнений.

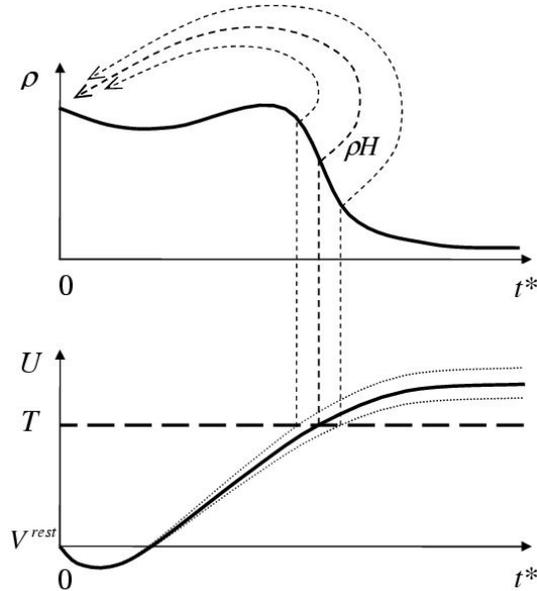


Рис.4. Схематическое представление эволюции нейронного континуума в фазовом пространстве переменной  $t^*$  – времени от предыдущего спайка. Распределению нейронной плотности  $\rho$  соответствует распределение среднего потенциала  $U$ . Отдельный нейрон «движется» вдоль оси  $t^*$  до пересечения потенциалом  $U$  порога  $T$ , когда он возвращается в состояние спайка  $t^*=0$ .

**3.2.3. Уравнение для плотности  $\rho(t, \tau)$ .** Среди скалярных переменных, наиболее полно описывающих состояние нейрона и имеющих физический смысл, согласно работам [7,15] можно признать время  $\tau$ , оставшееся до ожидаемого спайка. Соответственно, модель нейрона записывается в виде уравнения для  $\tau$ :

$$\frac{d\tau}{dt} = -1 + \frac{dg_E}{dt} L_E(\tau, g_E) + \frac{dg_I}{dt} L_I(\tau, g_I), \quad (8)$$

$$\tau \in [0, \tau_{MAX}(g_E, g_I)]; \quad \text{if } \tau = 0 \text{ then } \tau = \tau_{MAX}(g_E, g_I),$$

где функции  $L_E(g_E, g_I)$  и  $L_I(g_E, g_I)$  аппроксимируют чувствительность нейрона типа Ходжкина-Хаксли к изменениям возбуждающей и тормозной синаптических проводимостей  $g_E$  и  $g_I$ . Как показано в [6],  $\tau$ -модель точнее, чем  $SRM$ -модель воспроизводит чувствительность нейрона типа Ходжкина-Хаксли к возмущениям синаптических проводимостей, приходящих в разную фазу активности нейрона между спайками.

Данная модель нейрона приводит к модели популяции, т.е. к уравнению Фоккера-Планка для плотности  $\rho(t, \tau)$  распределения нейронов по параметру  $\tau$ :

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial \tau} \left[ \left( 1 - \frac{dg_E}{dt} L_E - \frac{dg_I}{dt} L_I \right) \rho \right] + \frac{\partial^2 (Q\rho)}{\partial \tau^2}, \quad (9)$$

где  $Q(\tau, g_E, g_I)$  – коэффициент диффузии, зависящий от параметров синаптического шума.

Данная модель популяции хорошо согласуется с расчётами прямым моделированием, т.е. с результатами моделирования большого числа единичных нейронов типа Ходжкина-Хаксли (рис.5). Недостатки этой модели заключаются в том, что  $\tau$ -модель работоспособна только в режиме генерации спайков, а также в том, что такие основные переменные как мембранный потенциал и проводимости ионных каналов не задействованы явно в уравнениях модели.

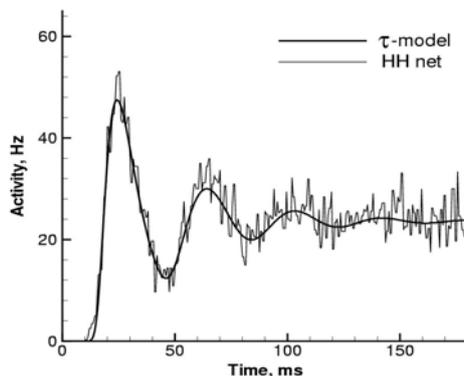


Рис.5. Сравнение  $\tau$ -модели ансамбля с прямым расчётом большого числа нейронов типа Ходжкина-Хаксли.

**3.2.4. Уравнение для плотности  $\rho(t, t^*)$  для нейронов типа Ходжкина-Хаксли.** Уравнение для нейронной плотности останется одномерным, если предположить, что переменные состояния всех нейронов могут быть параметризованы одной переменной (рис.4). В работах [8,16,17] предложено редуцировать уравнения типа Ходжкина-Хаксли до пороговой модели и параметризовать плотность, мембранный потенциал и безразмерные переменные активации ( $x$ ) и инактивации ( $y$ ) ионных каналов переменной  $t^*$  - временем, прошедшим от момента предыдущего спайка, т.е.  $\rho = \rho(t, t^*)$ ,  $V = V(t, t^*)$ ,  $x = x(t, t^*)$ ,  $y = y(t, t^*)$ . Тогда популяционная модель будет состоять из уравнения для плотности (6) и уравнений для потенциала и канальных переменных, записанных в терминах частных производных по времени  $t$  и параметру  $t^*$ . Основные уравнения этой модели для популяции пирамидных клеток области CA1 гиппокампа приведены ниже.

Для базовой односточечной модели нейрона типа Ходжкина-Хаксли используются аппроксимации кинетики натриевых и калиевых ионных токов  $I_{Na}$ ,  $I_{DR}$ ,  $I_A$ ,  $I_M$ ,  $I_H$  и значения проводимостей из работы [18], суммарный вклад кальциевых токов выражается в токе послеспайковой гиперполяризации  $I_{AHP}$ , аппроксимация для которого приведена в [19]. Пороговая безнатриевая модель типа Ходжкина-Хаксли получается обнулением натриевого тока. При этом для определения момента генерации спайка потенциал  $V$  сравнивается с порогом  $T$ . Приближённая зависимость порога от параметров состояния нейрона перед спайком выражена в однопараметрической зависимости  $T = T(dV/dt)$ , которая вычисляется сравнением полной и пороговой моделей и приведена на рис.6.

Полученная таким образом пороговая модель очень точно воспроизводит моменты спайков и потенциал на межспайковых интервалах, что видно на рис.7.

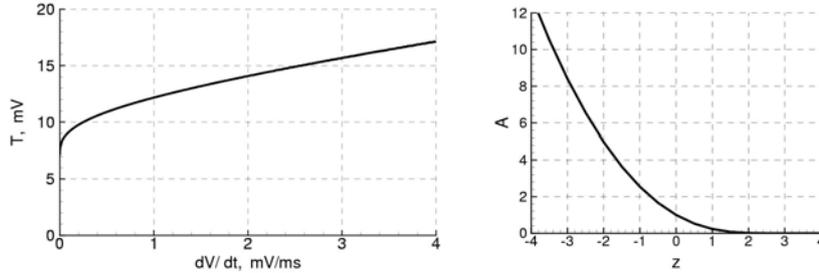


Рис. 6. Аппроксимирующие зависимости для динамического порога  $T$  и для вспомогательной функции  $A$ , аппроксимирующей функцию вероятности генерации спайка  $H$  в уравнении (10).

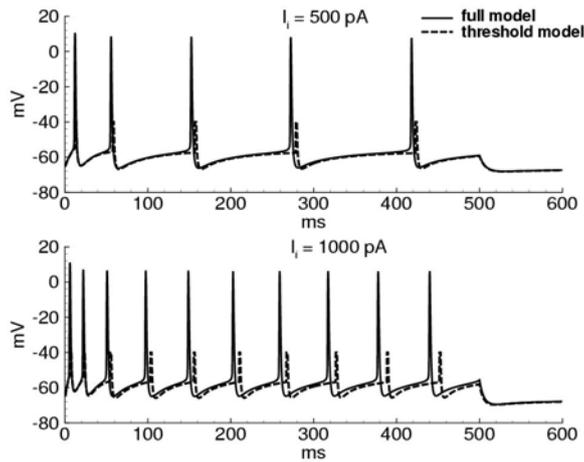


Рис.7. Сравнение полной и пороговой безнатриевой модели типа Ходжкина-Хаксли.

Уравнения в частных производных для нейронной плотности  $\rho$ , мембранного потенциала  $V$  и безразмерных переменных активации и инактивации ионных каналов  $x$  и  $y$  имеют вид:

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial t^*} &= -\rho H, \\ C \left( \frac{\partial V}{\partial t} + \frac{\partial V}{\partial t^*} \right) &= -I_{DR} - I_A - I_M - I_H - I_L - I_{AHP} - I_i, \\ \frac{\partial x}{\partial t} + \frac{\partial x}{\partial t^*} &= \frac{x_\infty(V) - x}{\tau_x(V)}, \\ \frac{\partial y}{\partial t} + \frac{\partial y}{\partial t^*} &= \frac{y_\infty(V) - y}{\tau_y(V)}, \end{aligned} \quad (10)$$

где  $C$  – мембранная ёмкость,  $I_L$  – ток утечки,  $I_i$  – внешний или синаптический ток; аппроксимация для функции  $H$  имеет вид:

$$H(V) = \frac{1}{\tau_m} \left( A(z) - \tau_m \frac{dz}{dt} \tilde{F}(z) \right), \quad z = \frac{T-V}{\sigma},$$

$$\tau_m = C / (g_{DR}(t, t^*) + g_A(t, t^*) + g_M(t, t^*) + g_H(t, t^*) + g_L + g_{AHP}(t, t^*)),$$

$A(z)$  определяется кривой на рис.6;

$$\tilde{F}(z) = \sqrt{\frac{2}{\pi}} \frac{\exp(-z^2)}{1 + \operatorname{erf}(z)},$$

где  $\sigma$  – дисперсия потенциала, соответствующая параметру шума при прямом моделировании единичных нейронов. Обоснование аппроксимации для  $H$  выходит за рамки настоящей статьи.

Граничные условия:

$$\rho(t, 0) = \int_{+0}^{\infty} \rho F dt^*,$$

$$V(t, 0) = V^{reset} = -40 mV,$$

$$x(t, 0) = x^{reset}, \quad y(t, 0) = y^{reset} \quad \text{для } I_{DR}, I_A, I_H; \quad I_{DR}: x^{reset} = 0.262, \quad y^{reset} = 0.473;$$

$$x(t, 0) = x(t, t^{*T}) + \Delta x^{reset} \quad \text{для } I_M; \quad I_A: x^{reset} = 0.743, \quad y^{reset} = 0.691;$$

$$x(t, 0) = x(t, t^{*T}) + \Delta x^{reset} \quad \text{для } I_{AHP}; \quad I_H: y^{reset} = 0.002;$$

$$I_M: \Delta x^{reset} = 0.18 (1 - x);$$

$$I_{AHP}: \Delta x^{reset} = 0.018(1 - x).$$

Значения численных параметров  $x^{reset}$ ,  $y^{reset}$ ,  $\Delta x^{reset}$  измерены по полной модели единичного нейрона на спайке при  $V = -40 mV$ .

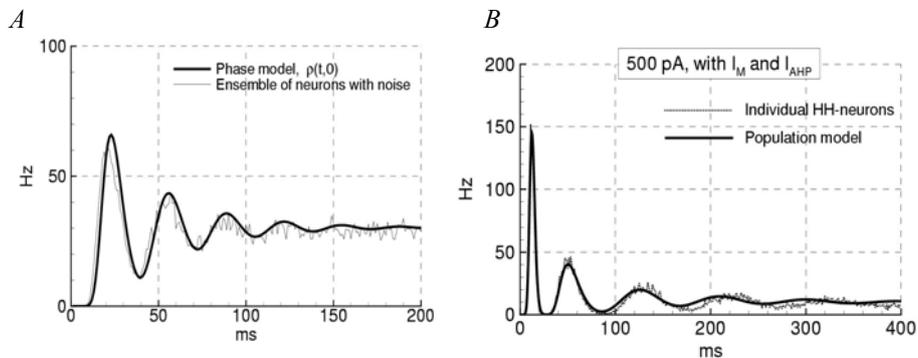


Рис.8. Согласование популяционной модели с расчётом прямым моделированием единичных нейронов при  $\sigma = 2 mV$ . A: неадаптирующиеся нейроны, ток стимуляции  $I_i = 300 pA$ ; B: адаптирующиеся нейроны,  $I_i = 500 pA$ .

Приведённая модель сравнивалась с расчётом прямым моделированием большого числа (4000 штук) единичных нейронов. При прямом моделировании разброс параметров нейронов и входных параметров выражался в дополнительном белом гауссовом шуме-токе с нулевым средним значением и дисперсией  $\sigma$ , т.е.  $\langle \eta(t) \rangle = 0$ ,  $\langle \eta(t) \eta(t') \rangle = \tau_m \sigma^2 \delta(t - t')$ . Популяционная модель очень точно

воспроизводит активность популяции неадаптирующихся клеток, т.е. в отсутствие токов  $I_M$ ,  $I_{AHP}$  (рис.8А). В случае учёта адаптации популяционная модель становится грубее, но такое согласование с «точным» решением можно признать вполне удовлетворительным (рис.8В).

#### 4. Заключение

Для описания любой фоновой активности нейросети, а также чисто популяционной ритмической или волновой активности актуально использование популяционной модели ансамбля нейронов. Такая модель должна выбираться из соображений точности описания единичных нейронов моделируемой нейросети. В настоящей статье приведён обзор существующих популяционных моделей и анализ их применимости в рамках выбранной модели единичного нейрона. Кратко можно сделать вывод, что если допустима упрощённая модель нейрона типа порогового интегратора, то можно использовать популяционную модель для плотности распределения нейронов по значениям мембранного потенциала. Если требуется рассматривать более подробную модель единичного нейрона, то в качестве упрощённой модели популяции всегда активных нейронов может быть порекомендована  $\tau$  – модель ансамбля из работы [7], рассматривающая распределение нейронов по предполагаемому времени до ожидаемого спайка. В качестве наиболее подробной модели популяции реалистичных нейронов может быть рекомендована модель из работы [8], рассматривающая распределение по времени, прошедшему от момента предыдущего спайка, пороговых нейронов типа Ходжкина-Хаксли.

#### Литература:

1. Omurtag A., Knight B.W., Sirovich L. On the Simulation of Large Populations of Neurons // Journal of Computational Neuroscience, 2000, v.8, pp 51–63.
2. Tuckwell H.C. Stochastic Processes in the Neurosciences. SIAM, Philadelphia, 1989.
3. Gerstner W., Kistler W.M. Spiking Neuron Models: Single Neurons, Populations, Plasticity. Cambridge: University Press, 2002. См. также <http://diwww.epfl.ch/~gerstner/SPNM>.
4. Покровский А.Н. Процессы управления в нервных клетках: Учеб. пособие. Л.: Изд-во ЛГУ, 1987.
5. Rall, W. Cable theory for dendritic neurons. // In Koch, C. and Segev, I., editors, Methods in Neuronal Modeling, pages 9-62, Cambridge. MIT Press, 1989.
6. Турбин А.А., Чижов А.В. Сравнение одномерных моделей реалистичных нейронов. // Устойчивость и процессы управления. Труды межд. конф. памяти В.И.Зубова. – СПб.: СПбГУ 2005г. с. 1207-1211.
7. Турбин А.А., Чижов А.В. Модель нейронного ансамбля. // Нейроинформатика-2003. Сб. научных трудов. М.: МИФИ. 2003, 133-140 с.
8. Чижов А.В. Модель популяций нейронов как элемент крупномасштабной нейросети // Нейрокомпьютеры: разработка, применение, 2004, No.2-3, стр.60-68.
9. Бахарев Б.В., Жадин М.Н., Агладзе Н.Н. Ритмические процессы в биоэлектрической активации коры головного мозга при реакции активации: качественный нелинейный анализ с учетом рефрактерности // Биофизика 2001, т.46(4), стр.715-723

От моделей единичных нейронов к моделям популяций нейронов

10. Rowe D.L., Robinson P.A., Rennie C.J. Estimation of neurophysiological parameters from the waking EEG using a biophysical model of brain dynamics // *J. Theor. Biology*, 2004, v.231(3), pp.413-433.
11. Dudkin A.O., Sbitnev V.I. Coupled map lattice simulation of epileptogenesis in hippocampal slices // *Biol. Cybern.* 1998, v.78(6), pp.479-486.
12. Nycamp D.Q., Tranchina D. A Population Density Approach That Facilitates Large-Scale Modeling of Neural Networks: Analysis and an Application to Orientation Tuning // *Journal of Computational Neuroscience*, 2000, v.8, pp.19-50.
13. Brunel N. and Hakim V. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates // *Neural Comput.*, 1999, v.11, pp.1621-1671.
14. Eggert J. and van Hemmen J. L. Modeling neuronal assemblies: theory and implementation // *Neural Computation*, 2001, v.13, pp.1923-1974.
15. Турбин А.А., Чижов А.В. Сравнение моделей популяционной нейронной активности. // *Нейроинформатика-2005. Сб. научных трудов. М.: МИФИ, 2005г. Ч.1, 122-126 с.*
16. Чижов А.В., Грэм Л. Популяционная модель нервной ткани для трактовки экспериментальных гамма и тета ритмов в гиппокампе // *Проблемы нейрокибернетики. Материалы 14-й междунар. конф. по нейрокибернетике. Ростов-на-Дону, 2005, с.198-203.*
17. Chizhov A.V., Graham L.J., Turbin A.A. Simulation of neural population dynamics with a refractory density approach and a conductance-based threshold neuron model // *Neurocomputing*, (accepted for publication), 2006.
18. Borg-Graham L. Interpretations of data and mechanisms for hippocampal pyramidal cell models // In P.S.Ulinski, E.G.Jones, and A.Peters, eds, *Cerebral Cortex*, 1999, v.13, pp. 19-138.
19. Kopell N., Ermentrout G.B., Whittington M.A., Traub R.D. Gamma rhythms and beta rhythms have different synchronization properties // *Neurobiology*, 2000, v.97(4), pp.1867-72.

Статья поступила 21 декабря 2005 г.  
После доработки 4 апреля 2006 г.

# О вероятности обнаружения локальных минимумов в обобщенной модели Хопфилда\*

Б.В.Крыжановский<sup>1)</sup>, Б.М.Магомедов<sup>2)</sup>

<sup>1)</sup> Институт оптико-нейронных технологий РАН, Москва, [kryzhanov@mail.ru](mailto:kryzhanov@mail.ru)

<sup>2)</sup> Институт оптико-нейронных технологий РАН, Москва, [bashir.magomedov@gmail.com](mailto:bashir.magomedov@gmail.com)

**Аннотация.** На основе анализа обобщенной модели Хопфилда получены выражения, устанавливающие связь между глубиной локального минимума и шириной области притяжения. На основании этого вероятность нахождения локального минимума при случайной инициализации нейронной сети удалось представить как функцию глубины этого минимума. В практических оптимизационных приложениях наличие таких выражений позволит по ряду уже найденных минимумов оценить вероятность нахождения более глубокого минимума и принять решение на остановку программы поиска или ее продолжение. Развитая аналитическая теория находится в хорошем согласии с результатами компьютерного эксперимента.

## 1. Введение

Обычно нейронная система ассоциативной памяти рассматривается как система, решающая задачи распознавания или восстановления образов. Однако, ее можно рассматривать и как систему, решающую оптимизационную задачу – нейронная сеть в процессе релаксации находит конфигурацию, соответствующую минимуму энергии [1]. Это свойство нейронной сети можно использовать для решения различных *NP*-полных задач. Стандартный подход сводится к нахождению такой архитектуры и параметров нейронной сети, при которых целевая (целевая) функция совпадает с понятием нейросетевой энергии. Успешное применение нейронной сети к задаче коммивояжера [2] инициировало широкие исследования нейросетевых подходов к решению задач теории графов [3], нейросетевой оптимизации обработки изображений [4] и большому ряду других приложений. Эта область теории нейронных сетей активно развивается по настоящее время.

Исследования в указанной области объединяет одно – сердцевиной подавляющего большинства нейросетевых оптимизационных алгоритмов является модель Хопфилда [1], а сам процесс оптимизации так или иначе сводится к нахождению в  $N$ -мерном конфигурационном пространстве глобального минимума некоего квадратичного функционала (энергии), построенного на заданной  $N \times N$ - матрице. Стандартный нейросетевой подход к решению указанной задачи сводится к процедуре случайного поиска оптимального решения: на первом этапе этой процедуры нейронная сеть случайным образом инициализируется, на втором этапе – нейронная сеть релаксирует в одно из возможных ста-

---

\* Работа выполнена при поддержке РФФИ (проект No 05-09-90049-в).

бильных состояний, т.е. оптимизирует величину энергии. Поскольку искомым результатом неизвестен и поиск проводится "вслепую", то нейросеть инициализируется раз за разом, с тем, чтобы найти как можно более глубокий минимум энергии. Несмотря на относительно успешное применение [2] этого подхода, всегда остается открытым вопрос – сколько таких случайных попыток следует сделать и когда проведенный поиск можно считать удачным (исходя из соотношения "затраченное время/полученный результат"). Сказанное справедливо как при спиновой [1-5], так и доменной [6,7] динамиках релаксации.

В настоящей работе, в продолжение проведенного в [10] анализа, получены выражения, устанавливающие связь между глубиной локального минимума энергии и шириной области притяжения. На основании этого, вероятность нахождения локального минимума при случайной инициализации нейронной сети удалось представить как функцию глубины этого минимума. В практических приложениях наличие таких выражений позволит по ряду уже найденных минимумов оценить вероятность нахождения более глубокого минимума и принять решение на остановку программы поиска или ее продолжение. Выражения получены на основе анализа обобщенной модели Хопфилда – нейронной сети с хэббовской [8] (корреляционной) матрицей межсвязей. Для такого типа матриц получено хорошее совпадение аналитических расчетов с результатами компьютерных экспериментов. Введенное в разделе 2 обобщение модели Хопфилда допускает возможность эвристического обобщения полученных результатов и на случай матриц иного типа.

## 2. Обобщенная модель Хопфилда

Рассмотрим модель Хопфилда. В теории нейронных сетей ее принято описывать как одномерную систему из  $N$  спинов-нейронов, которые могут ориентироваться вдоль, либо против заданной оси. Состояние такой нейронной сети характеризуется конфигурационным вектором  $\mathbf{S} = (s_1, s_2, \dots, s_N)$ , где  $s_i = \pm 1$ ,  $i \in (1, N)$ . Здесь мы будем рассматривать обобщенную модель, архитектура которой задается матрицей межсвязей

$$T_{ij} = \sum_{m=1}^M r_m s_{mi} s_{mj} \quad (1)$$

с нулевой диагональю ( $T_{ii} = 0$ ), организованной по правилу Хэбба на  $M$  образах –  $N$ -мерных бинарных эталонных векторах  $\mathbf{S}_m = (s_{m1}, s_{m2}, \dots, s_{mN})$ . Обобщение состоит в том, что каждый эталон  $\mathbf{S}_m$  добавляется в матрицу  $T_{ij}$  со своим статвесом  $r_m$ . Для упрощения выражений принята нормировка статвесов  $\sum r_m^2 = 1$ , не нарушающая общности рассмотрения. Это незначительное видоизменение модели оказывается весьма существенным, поскольку, в отличие от стандартной модели, позволяет описывать нейронную сеть с невырожденным спектром минимумов.

Энергия нейросети описывается выражением

$$E = - \sum_{i,j=1}^N s_i T_{ij} s_j, \quad (2)$$

а ее динамика заключается в следующем. Задается начальное состояние сети  $\mathbf{S}$  (начальные направления спинов устанавливаются в соответствии со знаками компонент вектора  $\mathbf{S}$ ). Затем вычисляется локальное поле  $h_i = -\partial E / \partial s_i$ , воздействующее на произвольно выбранный  $i$ -й спин со стороны всех остальных спинов системы в момент времени  $t$  и определяется энергия спина в этом поле  $\varepsilon_i = -s_i h_i$ . Если направление спина совпадает с направлением локального поля ( $\varepsilon_i < 0$ ), то его положение энергетически устойчиво и в последующий момент времени состояние спина остается неизменным. В противном случае ( $\varepsilon_i > 0$ ) положение спина неустойчиво и он разворачивается вдоль направления этого поля, переходя в состояние  $s_i(t+1) = -s_i(t)$  с энергией  $\varepsilon_i(t+1) < 0$ . Такая процедура последовательно применяется ко всем спинам нейронной сети. При каждом перевороте спина энергия сети понижается. Это означает, что сеть за конечное число шагов перейдет в стабильное состояние, соответствующее локальному минимуму энергии.

### 3. Размер области притяжения

Исследуем, при каких условиях эталонный образ  $\mathbf{S}_m$  будет являться неподвижной точкой [1], в которой энергия сети  $E$  достигает своего (локального) минимума  $E_m$ . Оценки получим в асимптотическом пределе  $N \rightarrow \infty$ . Определим область притяжения образа  $\mathbf{S}_m$  как совокупность точек  $N$ -мерного конфигурационного пространства, из которых нейронная сеть релаксирует в конфигурацию  $\mathbf{S}_m$ , и попробуем оценить размер этой области. Пусть начальное состояние сети  $\mathbf{S}$  находится в некоторой окрестности образа  $\mathbf{S}_m$ . Тогда вероятность того, что сеть сойдется к точке  $\mathbf{S}_m$ , опишется выражением:

$$Pr = 2^{-N} (1 + \operatorname{erf} z)^N. \quad (3)$$

Здесь  $\operatorname{erf} z$  – функция ошибок переменной  $z$ :

$$z = r_m \frac{\sqrt{N}}{\sqrt{2(1-r_m^2)}} \left( 1 - \frac{2n}{N} \right), \quad (4)$$

где  $n$  – хеммингово расстояние между  $\mathbf{S}_m$  и  $\mathbf{S}$ . Выражение (3) можно получить методами теоретико-вероятностного подхода (см. Приложение), повторив хорошо известные для случая  $r_m = 1/\sqrt{M}$ ,  $m = 1, 2, \dots, M$  выкладки [5].

Из (3) следует, что область притяжения определяется как совокупность близких к  $\mathbf{S}_m$  точек конфигурационного пространства, для которых справедливо соотношение  $n \leq n_m$ :

$$n_m = \frac{N}{2} \left( 1 - \frac{r_0}{r_m} \sqrt{1-r_m^2} \right), \quad (5)$$

где

$$r_0 = \sqrt{2 \ln N / N} . \quad (6)$$

Действительно, при  $n \leq n_m$  вероятность схождения в точку  $S_m$  с ростом  $N$  асимптотически стремится к единице; в противном случае ( $n > n_m$ ) имеем  $Pr \rightarrow 0$ . Это означает, что величину  $n_m$  можно рассматривать как радиус области притяжения локального минимума  $E_m$ .

Из (5) следует, что при  $r_m \rightarrow r_0$  радиус области притяжения стремится к нулю (рис.1). Это означает, что эталонные образы, прописанные в матрицу межсвязей (1) с статвесом, меньшим  $r_0$ , попросту не образуют локальных минимумов. Локальные минимумы имеются только в точках  $S_m$ , образы которых прописаны в матрице межсвязей с относительно большими статвесами  $r_m > r_0$ . Более того, если статвес одного из образов, например, образа  $S_1$ , сделать значительно больше чем у остальных ( $r_1 \rightarrow 1$ ), то область его притяжения ( $n_1 \sim N/2$ ) охватит половину всего  $N$ -мерного пространства (остальное пространство занимает область притяжения его негатива  $S_1^* = -S_1$ ). При этом сеть будет иметь всего лишь два локальных минимума в точках  $S_1$  и  $S_1^*$ , остальные локальные минимумы исчезнут. Этот предельный случай не представляет практического интереса для оптимизационных задач и ниже мы его рассматривать не будем.

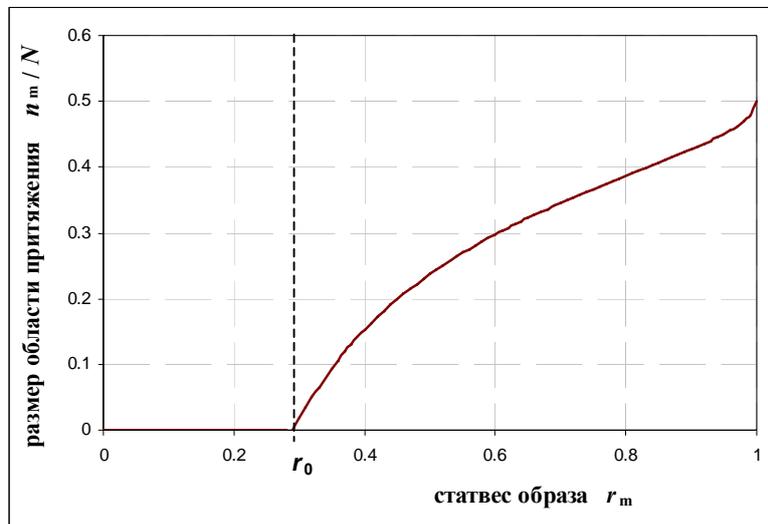


Рис.1. Характерный вид зависимости ширины области притяжения от статвеса образа. Локальный минимум существует только для образов, статвес которых в матрице межсвязей больше  $r_0$ . При  $r_m \rightarrow r_0$  размер области притяжения стремится к нулю, т.е. образы со статвесом  $r_m < r_0$  не создают локальных минимумов.

#### 4. Объем памяти нейронной сети в обобщенной модели

Из проведенного выше анализа следует, что не все из прописанных в матрице межсвязей образов образуют локальные минимумы и могут распознаваться нейронной сетью. В связи с этим следует дать четкое определение понятия объема нейросетевой памяти  $\bar{M}$ . Определим объем памяти как число тех образов, которые образуют локальные минимумы и, следовательно, могут быть восстановлены нейронной сетью.

Очевидно, что объем памяти зависит от того, как сформирована матрица межсвязей и, в частности, от распределения статвесов. Рассмотрим, например, случай, когда статвеса образов случайны и равномерно распределены на отрезке  $r_m^2 \in [0, 2/M]$  со средним  $\overline{r_m^2} = 1/M$ . В этом случае, вероятность того, что статвес образа удовлетворяет условию  $r_m > r_0$ , равна величине  $1 - Mr_0^2/2$ . Соответственно, объем памяти модели определится выражением:

$$\bar{M} = M(1 - Mr_0^2/2). \quad (7a)$$

Зависимость объема памяти от числа прописанных в матрице межсвязей образов  $M$ , представлена на рис.2. Как видим, максимальный объем памяти  $\bar{M} = N/4 \ln N$  достигается при числе образов  $M = N/2 \ln N$ . При  $M \geq N/\ln N$  объем памяти сети обращается в ноль – нет ни одного образа, образующего локальный минимум.

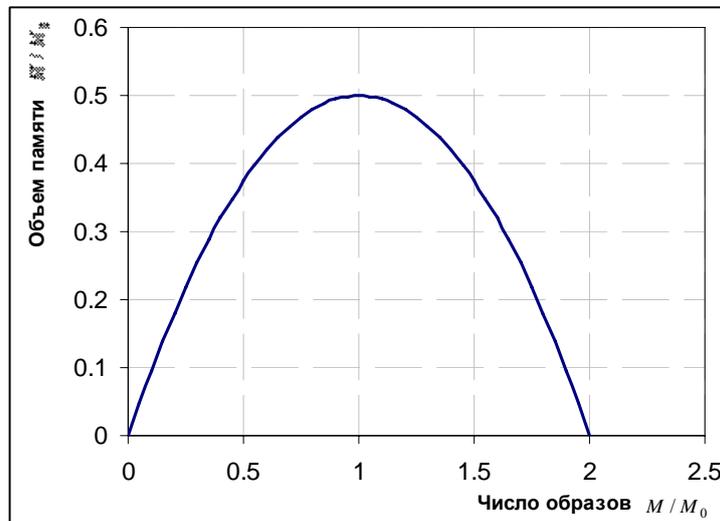


Рис.2. Вид зависимости объема памяти от числа образов  $M$ , прописанных в матрице межсвязей. Все величины нормированы на величину объема памяти  $M_0 = N/2 \ln N$  стандартной модели Хопфилда. Квадраты статвесов  $r_m^2$  эталонных образов равномерно распределены на интервале  $[0, 2/M]$ .

О вероятности обнаружения локальных минимумов в обобщенной модели Хопфилда

В случае, когда статвеса образов случайны и равномерно распределены на отрезке  $r_m \in [0, \sqrt{3/M}]$  со средним  $\overline{r_m^2} = 1/M$ , зависимость объема памяти от числа образов  $M$  следующим выражением:

$$\overline{M} = M(1 - r_0 \sqrt{M/3}). \quad (76)$$

Максимальный объем памяти  $\overline{M} = 2N/9 \ln N$  в этом случае достигается при числе образов  $M = 2N/3 \ln N$ , а при  $M \geq 3N/2 \ln N$  объем памяти сети обращается в ноль.

В другом предельном случае, когда все статвеса равны друг другу ( $r_m = 1/\sqrt{M}$ ), условие существования локальных минимумов  $r_m > r_0$  преобразуется в известное [5] ограничение  $M < N/2 \ln N$  на емкость памяти стандартной модели Хопфилда. В этом случае понятие "емкость памяти" совпадает с числом образов  $M$ , поскольку либо распознаются все прописанные в матрице межсвязей образы, либо не распознается ни один из них. Соответственно, максимальный объем памяти определяется выражением  $\overline{M} = N/2 \ln N$ .

В заключение этого раздела отметим, что удаление из матрицы межсвязей образов, прописанных в матрице межсвязей с статвесами, меньшими критического значения ( $r_m < r_0$ ), не влияет на размер объема памяти, но улучшает ее распознающие характеристики – размер областей притяжения оставшихся образов увеличится и, следовательно, сеть сможет распознавать более сильно искаженные образы.

## 5. Энергия локального минимума

Из анализа выражения (2) следует, что энергию локального минимума  $E_m$  с точностью до незначительной флуктуации порядка

$$\sigma_m = N\sqrt{1 - r_m^2} \quad (8)$$

можно представить в виде

$$E_m = -r_m N^2. \quad (9)$$

Тогда, с учетом соотношения (5) из (9) нетрудно получить выражение

$$E_m = E_0 \frac{1}{\sqrt{(1 - 2n_m/N)^2 + E_0^2/N^4}}, \quad (10)$$

где

$$E_0 = -N\sqrt{2N \ln N}, \quad (11)$$

устанавливающее связь между глубиной локального минимума и шириной его области притяжения. Как видим, чем шире область притяжения, тем глубже минимум, и наоборот – чем глубже минимум, тем шире область его притяжения (см. рис.3).

Введенная здесь величина  $E_0$  характеризует сразу два параметра нейронной сети. Во-первых, ею определяется полуширина распределения (10). Во-вторых, из (10) вытекает неравенство  $E_m \leq E_0$ , т.е.  $E_0$  является верхней границей спектра локальных минимумов. Эти результаты хорошо согласуются с ре-

зультатами компьютерных экспериментов, в ходе которых проверялось, существует ли в точке  $S_m$  локальный минимум или его нет.

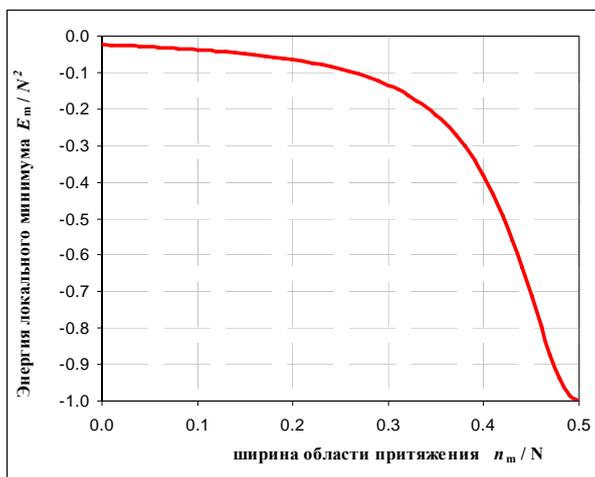


Рис.3. Зависимость энергии локального минимума от ширины области притяжения.

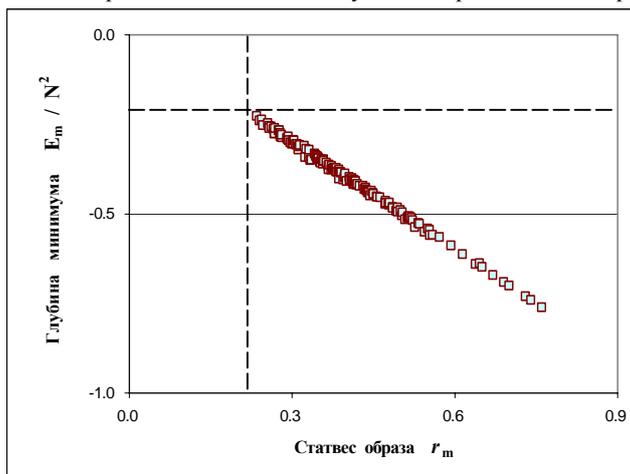


Рис. 4. Зависимость энергии  $E_m$  локального минимума от статвеса образа  $r_m$ .  $N=500$ ,  $M=25$ .

Результаты одного из экспериментов ( $N=500$ ,  $M=25$ ) приведены на рис.4. Как видим, хорошо прослеживается линейная зависимость энергии локального минимума от величины статвеса образа. Причем, экспериментальные точки, соответствующие локальным минимумам, расположены только в правом нижнем квадранте, где  $r_m > r_0$  и  $E_m < E_0$ . Кроме того, на рис.4 видно, что в соответствии с (8) флуктуации минимумов энергии с ростом статвеса уменьшаются.

## 6. Вероятность обнаружения минимума

Определим теперь вероятность  $W$  нахождения локального минимума  $E_m$  при случайном поиске. Искомая вероятность по определению совпадает с вероятностью того, что, задавая начальную конфигурацию, мы попадаем в область притяжения образа  $S_m$ . Следовательно, величина  $W = W(n_m)$  есть число точек в сфере радиуса  $n_m$ , приведенное к общему числу точек в  $N$ -мерном пространстве:

$$W = 2^{-N} \sum_{n=1}^{n_m} C_N^n. \quad (12)$$

Выражениями (8) и (12) в неявном виде задается связь между глубиной локального минимума и вероятностью его нахождения. Применяя к биномиальным коэффициентам асимптотическое разложение Стирлинга и заменяя в (12) суммирование интегрированием, можно представить искомую связь как

$$W = W_0 e^{-Nh}, \quad (13)$$

где  $h$  с точностью до постоянной совпадает с функцией Шеннона:

$$h = \frac{n_m}{N} \ln \frac{n_m}{N} + \left(1 - \frac{n_m}{N}\right) \ln \left(1 - \frac{n_m}{N}\right) + \ln 2. \quad (14)$$

Здесь  $W_0$  – не существенная для дальнейшего анализа медленная функция от  $E_m$ , которую можно получить асимптотической оценкой (12) при условии  $n_m \gg 1$ , а зависимость  $W = W(n_m)$  всецело определяется быстрой экспонентой.

Как следует из (12), вероятность обнаружения локального минимума малой глубины ( $E_m \sim E_0$ ) убывающе мала как  $W \sim 2^{-N}$ . Заметно отличной от нуля вероятность  $W$  становится только в случае достаточно глубоких минимумов  $|E_m| \gg |E_0|$ , размеры областей притяжения которых сопоставимы с величиной  $N/2$ . Выражение (13) с учетом (10) в этом случае преобразуется к зависимости  $W = W(E_m)$ , задаваемой выражением

$$W = W_0 \exp \left[ -NE_0^2 \left( \frac{1}{E_m^2} - \frac{1}{E_\infty^2} \right) \right], \quad (15)$$

где

$$E_\infty^2 = \sum E_m^2. \quad (16)$$

Здесь мы ввели новую константу  $E_\infty$ , хотя в принятой нами нормировке из (9) вытекает  $E_\infty = N^2$  и введение нового обозначения не имеет особого смысла, однако для других типов нормировок вполне возможна иная зависимость  $E_\infty$  от  $N$ .

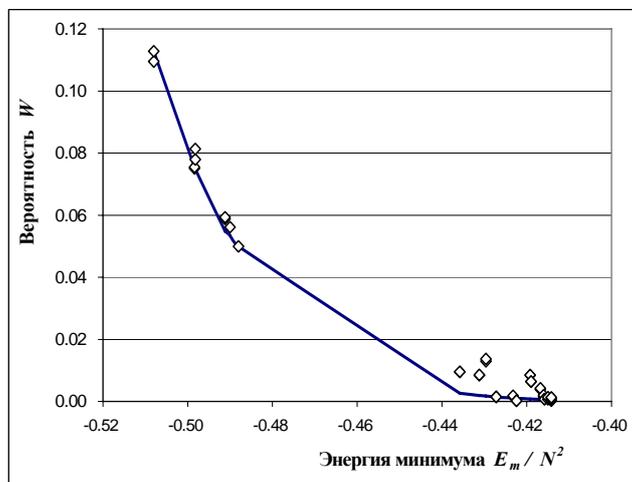


Рис. 5. Зависимость вероятности  $W$  обнаружения локального минимума от его глубины  $E_m$ .

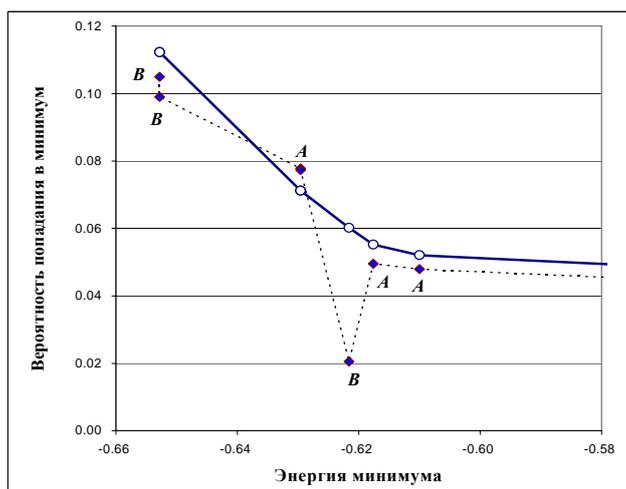


Рис.6. Сравнение предсказанных вероятностей (сплошная линия) и экспериментально найденных значений (точки, связанные пунктиром).

Как видим, вероятность обнаружения минимума растет с ростом его глубины. Эта зависимость подтверждается результатами компьютерных экспериментов, проведенных для матриц (1) с малым загрузочным параметром  $M/N \leq 0.05$ , при котором проведенный выше анализ справедлив. Сплошная кривая на рис.5 построена по формуле (13), точки – эксперимент. Как видим, хорошее согласие достигается прежде всего для наиболее глубоких локальных минимумов, которые соответствуют записанным в матрице межсвязей образам  $S_m$  (на рисунке – это область энергий  $E_m < -0.48N^2$ ). Экспериментально обна-

руженные минимумы малой глубины (точки в области  $E_m > -0.44N^2$ ) – это так называемые "химеры" [1]. В стандартной модели Хопфилда ( $r_m \equiv 1/\sqrt{M}$ ) они появляются при относительно больших загрузках  $M/N > 0.05$ . В рассматриваемом здесь более общем случае они могут появиться и раньше. Причины их появления хорошо исследованы методами статистической физики в работе [9], где показано, что химеры образуются как следствие интерференции минимумов  $S_m$ . При малой загрузке химеры отделены от минимумов  $S_m$  энергетической щелью, отчетливо видной на рис.5. С ростом параметра загрузки  $M/N$  соответствующие химерам минимумы опускаются и, при загрузке  $M/N \sim 0.1$ , сравниваются по глубине с локальными минимумами  $S_m$ . Более того, при  $M/N \geq 0.14$  появляются глубоко расположенные так называемые ложные (spurious) минимумы, образованные из больших фрагментов векторов  $S_m$ , а большинство локальных минимумов  $S_m$  исчезает. Эта экспериментально наблюдавшаяся картина имеет простое объяснение в рамках проведенного выше анализа: с ростом  $M$  уменьшается число образов, удовлетворяющих условию  $r_m > r_0$  существования локального минимума.

Несмотря на такую сложную картину энергетической поверхности, описываемая выражениями (13)-(15) зависимость "глубже минимум – больше область притяжения – больше вероятность попадания в этот минимум" как правило сохраняется при любых параметрах загрузки. Строгое обоснование этому феномену мы надеемся дать в последующих публикациях. Здесь же отметим, что зависимость (13) выполняется тем лучше, чем глубже описываемый этим выражением минимум. А поскольку для задачи оптимизации важно описать поведение именно наиболее глубоких минимумов, то это позволяет сформулировать эвристический подход для отыскания глобального минимума функционала (2) с произвольной наперед заданной матрицей (не обязательно хэббовской). Состоит он в том, чтобы использовать выражение (13) с неизвестными параметрами  $W_0$ ,  $E_0$  и  $E_\infty$ . Для этого запускается процедура случайного поиска и находится какое-то число минимумов. По полученным данным определяются характерные для данной матрицы значения  $E_0$  и  $E_\infty$  и величина подгоночного параметра  $W_0$ . Подстановка этих величин в (13) позволит оценить вероятность нахождения неизвестного более глубокого минимума  $E_m$  (если он существует) и принять решение на остановку программы поиска (если оценка пессимистична) или ее продолжение.

Такой подход опробован на хэббовских матрицах при достаточно больших загрузочных параметрах ( $M/N = 0.2 \div 2$ ), когда нейросетевая система переходит в состояние спинового стекла [9] и применение полученных выше результатов становится не совсем корректным. Результат одного из экспериментов приведен на рис.6. В ходе эксперимента по обнаруженным минимумам (точки А) были рассчитаны параметры  $W_0$ ,  $E_0$  и  $E_\infty$  и построена зависимость  $W = W(E_m)$  (сплошная кривая). Затем, в результате многократно повторяемой процедуры случайного поиска ( $\sim 10^5$  стартов нейронной сети) были установле-

ны другие минимумы (точки  $V$ ) и оценочные значения вероятностей попадания в них. Как видим, несмотря на разброс, предсказанные значения по порядку величины хорошо совпадают с реальными значениями вероятностей.

## 7. Обсуждение результатов

Проведенный анализ показал, что свойства обобщенной модели описываются двумя критическими параметрами  $r_0$  и  $E_0$ . Первый из них определяет минимальное значение статвеса, при котором образ образует локальный минимум и может быть восстановлен нейронной сетью. Второй – это верхняя граница локальных минимумов энергетической поверхности. Существенно, что величины обоих параметров никак не зависят от числа образов  $M$ , на которых построена матрица межсвязей. Существенно и то, что удаление из матрицы межсвязей образов, прописанных в матрице межсвязей с статвесами, меньшими критического значения ( $r_m < r_0$ ), хотя и не влияет на размер объема памяти, однако улучшает ее распознающие характеристики – размер областей притяжения оставшихся образов увеличится и, следовательно, сеть может распознавать более сильно искаженные образы. Вместе с этим, обратим внимание на тот факт, что при постоянном параметре загрузки  $M/N = const$ , с ростом размерности сети радиус области притяжения локального минимума уменьшается. Как следует из (5), такое уменьшение имеет место даже в том случае, когда загрузка сети с ростом  $N$  логарифмически уменьшается как  $M/N = const/2 \ln N$ . Это означает, что с ростом  $N$  уменьшается уровень допустимых искажений, при которых сеть способна восстанавливать образы (сеть меньших размеров способна распознавать более сильно искаженные образы).

Во избежание недоразумений следует указать на то, что связь между радиусом области притяжения  $n_m$  и объемом этой области  $W = W(n_m)$  существенно нелинейная (12). Это означает, что даже при очень большом значении радиуса (например  $n_m \sim 0.49N$ ) область притяжения может занимать очень малую часть  $N$ -мерного конфигурационного пространства, а доля пространства вне области притяжения (т.е. все точки, отстоящие от минимума на расстоянии от  $n_m$  до  $0.5N$ ) с ростом  $N$  нарастает экспоненциально и при достаточно больших значениях  $N$  стремится к  $1/2$ .

Действительно, рассмотрим простейший пример – случай стандартной модели Хопфилда ( $r_m^2 = M^{-1}$ ). На рис. 7 показана зависимость радиуса области притяжения от размерности сети при разных загрузках. Как видно из рис.7а, в случае постоянной загрузки ( $M/N = const$ ) радиус притяжения с ростом  $N$  быстро стремится к нулю. Это соответствует выражению (5), из которого следует, что радиус области притяжения отличен от нуля только при небольших размерах сети, когда  $2k \ln N \leq 1$ . Очевидно, что доля пространства, занимаемая всеми  $M$  локальными минимумами, в этом случае с ростом  $N$  стремится к нулю. В приведенном на рис.7б случае логарифмически уменьшающейся загрузки ( $M/N = const/2 \ln N$ ) радиус области притяжения с ростом  $N$  достаточно быстро уменьшается до некоторого (отличного от нуля) асимптотического значе-

О вероятности обнаружения локальных минимумов в обобщенной модели Хопфилда

ния  $n_m \sim 0.5N(1 - \sqrt{k})$ . Тем не менее, доля пространства, занимаемая всеми  $M$  локальными минимумами (величина, равная  $M \cdot W$ ), и в этом случае с ростом  $N$  стремится к нулю. Это означает, что с ростом  $N$  практически все пространство отходит к ложным минимумам и вероятность нахождения глубоких минимумов при случайном поиске также стремится к нулю.

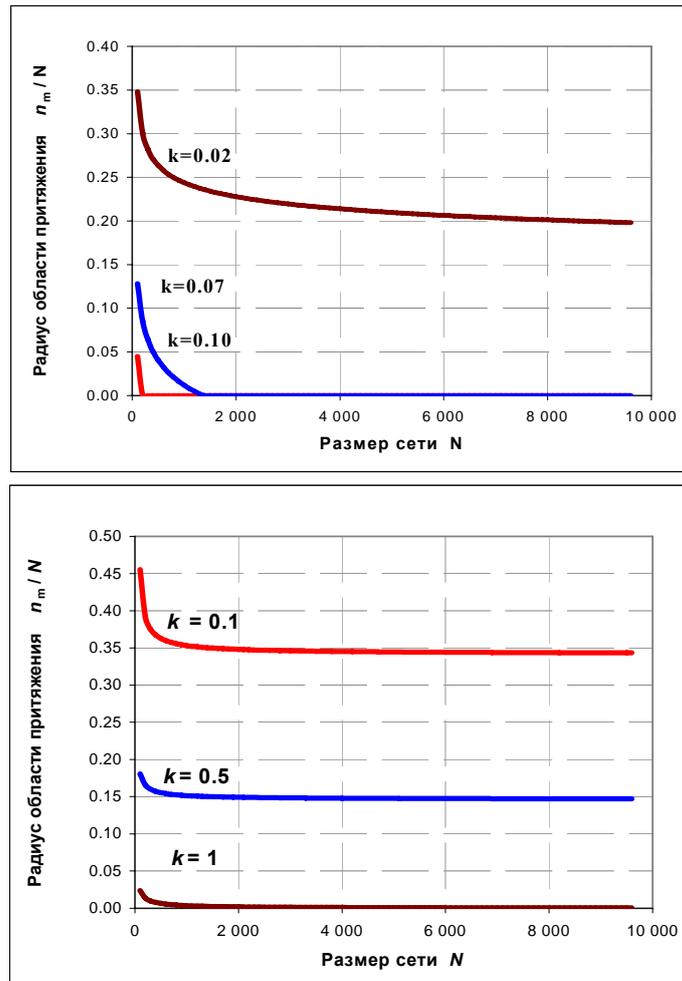


Рис.7. Размер радиуса области притяжения локального минимума в стандартной модели Хопфилда, как функция размерности пространства  $N$  для различных значений загрузки сети: а) постоянная загрузка  $M = kN$  при  $k = 0.02, 0.07, 0.1$ ; б) логарифмически уменьшающаяся загрузка  $M = kN / 2 \ln N$  при  $k = 0.1, 0.5, 1$ .

На рис.8 для наглядности сделанного утверждения приводится гипотетический пример, когда имеется  $M$  минимумов, причем радиус области притяже-

ния каждого из них очень велик и равен  $n_m = kN$ , где  $k = 0.47, 0.48, 0.49$ . Как видим, даже в этом случае, несмотря на нереально большой радиус областей притяжения, доля пространства, занимаемая этими минимумами, с ростом  $N$  стремится к нулю.

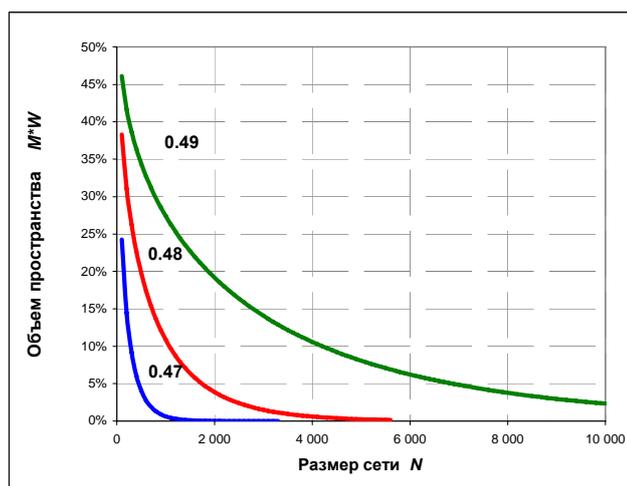


Рис.8. На рисунке показано уменьшение доли пространства, занимаемого  $M$  локальными минимумами в случае, когда радиус притяжения каждого из минимумов равен  $kN$ , при  $k=0.47, 0.48, 0.49$ .

В заключение сделаем несколько практических замечаний, относящихся к сравнению аналитической теории с компьютерным экспериментом. Во-первых, напомним, что картина энергетического ландшафта симметрична относительно изменения знака вектора состояния нейронной сети и, следовательно, каждому локальному минимуму в точке  $S$  соответствует точно такой же минимум в точке  $S^* = -S$ . Поэтому, экспериментально измеренную вероятность нахождения минимума с энергией  $E$  следует сравнивать с величиной  $W = W(E)$ , если только в ходе эксперимента фиксировалась координата минимума, в противном случае – с величиной  $2W(E)$ . Во-вторых, еще раз подчеркнем, что зависимость "глубже минимум – больше область притяжения – больше вероятность попадания в этот минимум" строго обоснована только в случае корреляционных (хэббовских) матриц. В более общем случае это не зависимость, а скорее экспериментально установленная тенденция, применимая к наиболее глубоким минимумам. Обоснование этой тенденции заключается в том, что любую наперед заданную матрицу можно представить в виде хэббовской матрицы (1), построенной на произвольном числе образов (например  $M \rightarrow \infty$ ) с произвольными статвесами. И если спектр минимумов функционала, построенного на такой матрице, имеет ярко выраженные минимумы большой глубины, то применительно к ним справедливы сделанные ранее выводы. Трудность наступает в случае, когда таких ярко выраженных минимумов нет и спектр минимумов ква-

О вероятности обнаружения локальных минимумов в обобщенной модели Хопфилда

зинепрерывен. В этом случае в оптимизационном эксперименте измеряется не вероятность  $W(E)$ , а частота попадания в минимум с заданной энергией – произведение вероятности на спектральную плотность, характеристики которой неизвестны. Соответственно, затрудняется применение развитой выше теории.

## 8. Приложение

Используем так называемое одношаговое приближение (one-step approximation) [5]. Нейронная сеть из состояния  $\mathbf{S}$  сойдется в точку  $\mathbf{S}_m$ , если для любого  $i \in [1, N]$  направление локального поля  $h_i$  совпадает с направлением  $i$ -го спина в эталоне, т.е.  $h_i s_i^{(m)} \geq 0$ . Вероятность этого события описывается выражением:

$$Pr = \prod_{i=1}^N Pr(h_i s_i^{(m)} \geq 0)$$

В асимптотическом пределе  $N \rightarrow \infty$  величину  $h_i s_i^{(m)}$  можно рассматривать как случайную величину со средним  $\xi = r_m N(1 - 2n/N)$  и дисперсией  $\sigma_\xi^2 = N(1 - r_m^2)$ . Соответственно, вероятность события  $h_i s_i^{(m)} \geq 0$  равна  $Pr(h_i s_i^{(m)} \geq 0) = \frac{1}{2}(1 + \operatorname{erf} \gamma)$ , где  $\gamma = \xi / \sigma_\xi$ . Возводя эту величину в  $N$  степень, получим выражение (3).

## Литература

1. J.J.Hopfield. Neural networks and physical systems with emergent collective computational abilities. Proc. Nat. Acad. Sci. USA. 79, 2554-2558 (1982).
2. J.J.Hopfield, D.W.Tank. Neural computation of decisions in optimization problems. Biological Cybernetics 52, 141-152 (1985); J.J.Hopfield, D.W.Tank. Computing with neural circuits: A model. Science 233, 625-633 (1986).
3. Y.Fu, P.W.Anderson. Application of statistical mechanics to NP-complete problems in combinatorial optimization. Journal of Physics A. 19, 1605-1620 (1986).
4. T.Poggio, F.Girosi. Regularization algorithms for learning that are equivalent to multi-layer networks. Science 247, 978-982 (1990).
5. R.McEllice, E.Posner, E.R.Rodemich. Capacity of the Hopfield associative memory. IEEE Trans. on Inf. Theory. 33, 461-482 (1987).
6. Б.В.Крыжановский, Б.М. Магомедов, А.Л.Микаэлян. Доменная модель нейронной сети. Доклады АН 401, 462-466 (2005).
7. В.Kryzhanovsky, В.Magomedov. Application of domain neural network to optimization tasks. ICANN 2005. Poland. LNCS 3697, Part II, pp.397-403.
8. D.O.Hebb The Organization of Behavior. New York: Wiley, 1949.
9. D.J.Amit, H.Gutfreund and H.Sompolinsky Statistical mechanics of neural networks near saturation //Annals of Physics 173, 30-67 (1987).
10. Б.В.Крыжановский, Б.М.Магомедов, А.Л.Микаэлян. Взаимосвязь глубины локального минимума и вероятности его обнаружения в обобщенной модели Хопфилда. Доклады АН 405, №3, с.1-5 (2005).

Статья поступила 16 декабря 2005 г.

После доработки 20 апреля 2006 г.

# NEUROINFORMATICS

Referred electronic journal

**Vol. 1, No 1**

English abstracts of articles

## **Neural Network Model of Multiple Object Tracking**

Yakov B. Kazanovich<sup>1</sup>, Roman M. Borisyuk<sup>1,2</sup>

<sup>1</sup>Institute of Mathematical Problems of Biology, Russian Academy of Science, Pushchino, Moscow region, yakov\_k@impb.psn.ru

<sup>2</sup>Centre for Theoretical and Computational Neuroscience, University of Plymouth, Plymouth, UK rborisyuk@plymouth.ac.uk

**Abstract.** Psychological experiments show that humans are able to simultaneously track up to five randomly moving visual objects. The paper presents a neural network model of this phenomenon. The model operates with a set of identical visual objects located in the visual field. At the stage of exposition a subset of objects marked as targets is selected into the focus of attention. Other objects are used as attention distractors. At this stage objects are motionless. Later on at the stage of tracking the model aims to preserve initial separation between targets and distractors while objects are moving. It will be shown that multiple object tracking can be implemented by a proper interplay of synchronizing and desynchronizing interactions in a multilayer oscillatory neural network where each layer is responsible for tracking a single target. Computer simulations show that in the case of non-overlapping objects the performance of the model decreases as the number of targets increases. The same dependence has been observed in the experiments with humans. Also the functioning of the model in the case of temporally overlapping objects is presented.

## Evolutionary approach to design and training of artificial neural networks

Tsoy Y.R.<sup>1)</sup>, Spitsyn V.G.<sup>2)</sup>

Tomsk Polytechnic University, Tomsk, <sup>1)</sup> qai@mail.ru,  
<sup>2)</sup> spitsyn@ce.cctpu.edu.ru

**Abstract.** This paper concerns application of evolutionary algorithms to design and training of artificial neural networks (ANNs). We give general characteristics of the neuroevolutionary approach and state its advantages and disadvantages. Problems of neuroevolutionary algorithms development as well as possible ways of their solutions are described. Also the use of evolutionary training of ANNs for some “inexact evaluation” tasks solution is substantiated.

## Elastic neural networks as a robust approach

G.A. Ososkov

Laboratory of Information Technologies, Joint Institute for Nuclear Research, 141980, Dubna, Russia, ososkov@jinr.ru

**Abstract.** After a brief survey of data handling methods for contemporary experiments of high energy physics, methods of elastic neural networks are expounded. Then they are considered as a particular case of a robust (stable to sample contaminations) approach to dependence estimations. A formalism of the robust approach is described with deriving optimal weights for the sample uniform contamination. An effective algorithm of the robust estimation is obtained on the basis of the joint estimate of position and scale parameters.

## From single-neuron models to models of neural populations

Chizhov A.V.<sup>1)</sup>, Turbin A.A.<sup>2)</sup>

<sup>1)</sup> Ioffe Physico-Technical Institute, Russian Academy of Science, [anton.chizhov@mail.ioffe.ru](mailto:anton.chizhov@mail.ioffe.ru)

<sup>2)</sup> St.-Petersburg State Polytechnic University, [Turbin\\_A\\_A@mail.ru](mailto:Turbin_A_A@mail.ru)

**Abstract.** We classify known neural population models and those single-neuron models used in the population models. The equations of typical one-dimensional population models are given, which describe neural continuums and distinguish states of the neurons between spikes by the means of a single parameter. The equations of such two models proposed by the authors of the present paper are given as well.

Аннотации

## On probability of finding local minima in the generalized Hopfield mode

Boris V. Kryzhanovsky<sup>1)</sup>, Bashir M. Magomedov<sup>2)</sup>

<sup>1)</sup> Institute of Optical Neural Technologies, RAS, Moscow [kryzhanov@mail.ru](mailto:kryzhanov@mail.ru)

<sup>2)</sup> Institute of Optical Neural Technologies, RAS, Moscow [bashir.magomedov@gmail.com](mailto:bashir.magomedov@gmail.com)

**Abstract.** In this paper, we obtain expressions relating the depth of a local minimum of energy to the width of the domain of attraction. Using these expressions, we were able to represent the probability of finding a local minimum under a random initialization of the neural network as a function of the depth of this minimum. In practical applications, these expressions will make it possible to estimate the probability of determining a deeper minimum from a series of already found minima and make a decision on whether the run of a search program must be terminated or continued. The expressions are obtained by analyzing the generalized Hopfield model, namely, a neural network with Hebb correlation matrix. For matrices of this type, the analytical theory excellently agrees with the computer experiment.

## Оглавление

<b>Редько В.Г.</b> Предисловие: О журнале «Нейроинформатика» .....	1
<b>Казанович Я.Б., Борисюк Р.М.</b> Нейросетевая модель слежения за несколькими объектами .....	4
<b>Цой Ю.Р., Спицын В.Г.</b> Эволюционный подход к настройке и обучению искусственных нейронных сетей .....	34
<b>Осоков Г.А.</b> Метод эластичных нейронных сетей и его робастная трактовка .....	62
<b>Чижов А.В., Турбин А.А.</b> От моделей единичных нейронов к моделям популяций нейронов .....	76
<b>Крыжановский Б.В., Магомедов Б.М.</b> О вероятности обнаружения локальных минимумов в обобщенной модели Хопфилда .....	88
<b>English abstracts of the articles</b> .....	102