



УДК 004.8.032.26, 681.513.8

*Е.Н. Бендерская, К.В. Никитин*

## **РЕКУРРЕНТНАЯ НЕЙРОННАЯ СЕТЬ КАК ДИНАМИЧЕСКАЯ СИСТЕМА И ПОДХОДЫ К ЕЕ ОБУЧЕНИЮ**

*E.N. Benderskaya, K.V. Nikitin*

### **RECURRENT NEURAL NETWORK AS DYNAMICAL SYSTEM AND APPROACHES TO ITS TRAINING**

Представлены результаты аналитического исследования рекуррентных нейронных сетей (РНС) и их обобщающая классификация, выполненная с позиций динамических систем и с учетом нового класса РНС – резервуарных РНС. Систематизация знаний в данной предметной области позволила выделить основные динамические режимы работы РНС, а также определить наиболее перспективные направления в развитии методов обучения РНС с учетом выявленных достоинств и недостатков существующих подходов.

**РЕКУРРЕНТНАЯ НЕЙРОННАЯ СЕТЬ. ДИНАМИЧЕСКАЯ СИСТЕМА. АЛГОРИТМЫ ОБУЧЕНИЯ. РЕЗЕРВУАРНЫЕ ВЫЧИСЛЕНИЯ. НЕУСТОЙЧИВАЯ ДИНАМИКА.**

This paper presents the results of analytical study of recurrent neural networks (RNN) and their summary classification, done from the position of dynamical systems with regard to a new type of RNN – reservoir RNN. Knowledge systematization in the subject field allowed to distinguish main dynamical regimes of RNN and to describe the most perspective trends in the development of training algorithms in terms of the detected advantages and drawbacks of existing approaches.

**RECURRENT NEURAL NETWORK. DYNAMICAL SYSTEM. TRAINING ALGORITHMS. RESEVOIR COMPUTING. UNSTABLE DYNAMICS.**

Рекуррентные нейронные сети (РНС) – нейронные сети (НС) с обратными связями, благодаря которым они имеют собственную динамику, что отличает их от статических НС, например, нейронных сетей прямого распространения. Состояние РНС в текущий момент времени определяется не только входными сигналами, но и предыдущим состоянием. За счет этого РНС обладают бесконечной памятью, что отличает их от НС с задержками, которые запоминают информацию о входах только за ограниченный промежуток времени.

Ввиду того что РНС является динамической системой, целесообразно провести анализ возможности привлечения результатов исследования динамических систем в смежных областях науки, таких, как физика, теория нелинейных динамических систем, теория хаоса и других, для синте-

за РНС и решения вопросов их обучения [6, 9, 12, 18]. Необходимо отметить, что при этом термины «нейрон» и «НС» могут вообще не использоваться – вместо этого используются такие понятия, как связанная решетка, структура, среда, конечный автомат с памятью, клеточный автомат. Результатом такого анализа станет выявление перспективных направлений развития РНС в части их синтеза и обучения, что актуально ввиду больших потенциальных возможностей РНС, отмеченных во многих научных работах.

#### **Классификация РНС**

Систематизация существующих на сегодняшний день различных РНС позволила предложить обобщенную классификацию НС, выделив особенности, присущие РНС.

1. Структура:
  - 1) топология:
    - без обратных связей (прямого распространения, НС радиальных базисных функций);
    - с обратными связями (рекуррентные):
      - полносвязные (каждый с каждым);
      - с локальными связями (в небольшой окрестности);
      - с произвольными, случайными связями;
    - специфическая топология (самоорганизующиеся НС, сети теории адаптивного резонанса);
  - 2) дискретизация в пространстве:
    - дискретное пространство (решетка);
    - непрерывное пространство (среда);
  - 3) гомогенность связей:
    - гомогенные связи (одинаковые по типу и силе);
    - гетерогенные связи (различные по типу или силе);
  - 4) детерминированность:
    - полностью определенные связи;
    - случайные связи.
2. Тип элементов:
  - 1) модель нейрона (реалистичность):
    - искусственные нейроны (искусственные НС);
    - импульсные нейроны (импульсные НС);
  - 2) динамика нейрона:
    - статические нейроны (статические НС);
    - динамические нейроны (динамические НС):
      - нейроны с затухающей памятью;
      - осцилляторы (в т. ч. хаотические);
  - 3) функция активации нейрона:
    - аддитивная;
    - мультипликативная (шунтирующая);
    - расстояния (с нормализацией и без нормализации);
  - 4) передаточная функция нейрона:
    - пороговая (симметричная и несимметричная);
      - линейная с насыщением/без насыщения;
      - сигмоидальная (симметричная и несимметричная);
    - соревновательная (с жестким и мягким максимумами);
    - радиально-базисная;
    - В-сплайн.
3. Тип связей:
  - 1) постоянство:
    - фиксированные связи;
    - переменные связи (сила связи изменяется в течение работы НС);
  - 2) характер связей (для импульсных НС):
    - электрические – сильная связь (токовая);
    - химические – импульсно-связанные (pulse-coupled).
4. Динамика НС:
  - 1) наличие динамики:
    - динамика отсутствует (статические НС);
    - динамика есть (динамические НС)
  - 2) объем памяти:
    - конечная память (НС с задержками, КИХ-фильтры);
    - бесконечная память (РНС, БИХ-фильтры);
  - 3) установившийся режим:
    - аттрактор типа фиксированная точка;
    - предельный цикл (осцилляции, периодические колебания);
    - сложный аттрактор (тор, хаотический аттрактор);
  - 4) устойчивость динамики:
    - устойчивая;
    - на границе устойчивости (автоколебания);
    - неустойчивая;
  - 5) дискретизация по времени:
    - дискретное время (описывается разностными уравнениями);
    - непрерывное время (описывается дифференциальными уравнениями);
  - 6) синхронность динамики (для дискретного времени):
    - асинхронная (состояние изменяется сразу только у одного элемента);
    - синхронная (состояние изменяется сразу у всех элементов);
    - распределенная (состояние изменяется сразу у группы элементов);
5. Алгоритм обучения:
  - 1) с учителем:

- глобальное – на основе оптимизации: обратное распространение; Больцмана – случайное; генетические алгоритмы;
  - локальное – на основе правила Хебба;
- 2) без учителя – самоорганизация (правило Кохонена, синаптическая временная пластичность);
  - 3) с подкреплением (с критиком);
  - 4) без обучения.
6. Решаемая задача:
- 1) извлечение (генерация) признаков;
  - 2) классификация;
  - 3) ассоциативная память;
  - 4) аппроксимация;
  - 5) предсказание;
  - 6) оптимизация;
  - 7) управление, идентификация, диагностика;
  - 8) кластеризация, сегментация.

**Структура РНС.** Структура РНС является ее основной характеристикой, от которой напрямую зависят все ее свойства. Выделяют полносвязные РНС и РНС с локальными связями. В полносвязных РНС каждый нейрон связан с каждым. Это приводит к подчинению всех нейронов сети какому-то одному глобальному режиму. В РНС с локальными связями каждый нейрон связан лишь с нейронами в некоторой окрестности вокруг него. Такие связи формируются детерминированным или случайным образом. Локальная связность дает нейронам больше независимости, в разных областях РНС могут преобладать разные динамические режимы.

Существуют примеры РНС с изменяющейся структурой: в ходе работы нейронной сети у нее могут появляться (исчезать) нейроны и связи.

При небольшом количестве элементов структура РНС представляет собой *решетку* в одно-, двух- и максимум трехмерном *дискретном* пространстве. Состояние РНС характеризуется совокупностью значений выходов ее нейронов. Для математического описания таких РНС используются системы ДУ. Если количество элементов велико, то от дискретного пространства состояний переходят к *непрерывному*, и структура РНС представляет собой непрерывную *среду* с

распределенными параметрами. Динамика таких бесконечномерных в пространстве систем описывается специальными волновыми уравнениями – уравнениями в частных производных или интегрально-дифференциальными уравнениями.

Для анализа структуры РНС может применяться подход, основанный на теории графов. Он заключается в вычислении различных показателей, таких, как матрица достижимости, пути, циклы, маршруты, кластерный индекс и др. С помощью этих показателей можно приблизительно оценить взаимодействие элементов в РНС.

РНС является динамической системой, работающей в дискретном или непрерывном времени. В случае дискретного пространства значений выходов РНС изменение ее состояний может быть описано системой разностных или дифференциальных уравнений:

$$X[t+1] = g1(X[t], U[t])$$

$$dX/dt = g2(X, U),$$

где  $X = (x_1, x_2, \dots, x_n)$  – состояние системы;  $n$  – размерность системы;  $x_i$  – значение  $i$ -го выхода РНС;  $U$  – входное воздействие;  $g1$  и  $g2$  – функция и отображение соответственно, описывающие зависимости состояния системы от предшествующих состояний и входного воздействия (определяют эволюцию системы во времени).

В зависимости от вида  $g1$  и  $g2$  РНС будет принадлежать к разному классу динамических систем, классификация которых частично совпадает с классификацией динамики для РНС, представленной выше. Одним из важнейших классификационных признаков является линейность или нелинейность  $g1$  и  $g2$ , от чего зависит принадлежность РНС к классу линейных или нелинейных динамических систем.

Анализ системы ДУ может проводиться в зависимости от цели и особенностей динамической системы (линейная/нелинейная, непрерывность/дискретность по времени):

- 1) использование физических принципов работы динамической системы (анализ с позиций превращения энергии);
- 2) применение вариационных принципов

пов для системы ДУ;

3) применение формальных групп и теории инвариантов;

4) геометрический подход;

5) анализ существования и устойчивости периодических движений.

Нейродинамика — область знаний, в которой НС рассматриваются как динамические системы, но основной упор делается на проблему устойчивости [8].

### Анализ динамики РНС

**РНС с дискретным пространством состояний.** *Устойчивость.* Одним из главных свойств РНС как динамической системы является устойчивость. РНС может быть устойчива (не устойчива) в малом и в большом. Устойчивость в большом гарантирует устойчивость во всем пространстве состояний, а устойчивость в малом — лишь в определенных точках. Теоремы Ляпунова дают аналитические критерии устойчивости нелинейных систем:

для устойчивости в большом необходимо подобрать специальную функцию, удовлетворяющую условиям второй теоремы Ляпунова;

для устойчивости в малом необходимо произвести линеаризацию системы в точке равновесия и проверить в ней устойчивость.

Как правило, устойчивость в большом удается выявить лишь для небольшого класса РНС (например, НС Хопфилда) [7]. А устойчивость в малом свидетельствует лишь о поведении системы в точках равновесия. Поэтому разработаны специальные численные характеристики, позволяющие судить об устойчивости системы по ее фазовой траектории.

*Основные типы динамики.* В зависимости от устойчивости различают три основных типа динамики РНС:

- устойчивая динамика, при которой за конечное время РНС сходится к устойчивому состоянию равновесия (в данном режиме при изменении входного сигнала возможны переходы системы из одного состояния в другое);

- колебательная динамика: состояние РНС описывает замкнутую циклическую

траекторию — предельный цикл; с одной стороны можно считать, что предельный цикл кодирует некоторую информацию, а с другой — что происходят колебания, при этом информация передается в фазовых соотношениях;

- неустойчивая динамика: траектории РНС при сколь угодно близких начальных состояниях расходятся со временем; если рост состояния системы никак не ограничен, то оно уходит в бесконечность, в противном случае устанавливается хаотическая динамика, при которой состояние системы спонтанно изменяется во времени внутри определенной области фазового пространства [2, 7].

Необходимо отметить, что существует еще хаотическая динамика, при которой изменения состояния РНС носят непредсказуемый характер, и со стороны случайного наблюдателя процесс является стохастическим. С точки зрения физики процесса, система «уходит в себя» и слабо реагирует на входные сигналы. В хаотическом режиме количество информации, содержащееся в состоянии РНС, большое, но извлечь эту информацию сложно. Во многих лабораториях исследуется поведение динамических систем в хаотических режимах и предлагаются варианты интерпретации этой динамики [10, 16]. Данный тип динамики является неустойчивым в малом, но устойчивым в большом и в некоторых случаях динамику можно рассматривать как колебательную.

Различные динамические режимы РНС могут рассматриваться как отдельные этапы решения конкретной задачи. Например, хаотический режим соответствует начальному поиску решений — выбору между различными гипотезами, колебательный режим отражает процессы переключения внимания с одного решения на другое, а устойчивый режим отвечает найденному решению [9].

Динамика РНС зависит как от ее структуры, так и от свойств составляющих ее элементов. При этом возможны случаи, когда РНС из устойчивых элементов при определенной структуре неустойчива и наоборот — когда РНС из неустойчивых, хаотических элементов является устойчи-

вой. Однако при исследовании устойчивых режимов НС, как правило, формируется из устойчивых элементов, при исследовании колебательных режимов – из осцилляторов-колебательных элементов и, наконец, при исследовании хаотических режимов – из хаотических осцилляторов [2, 7].

Теория бифуркаций, катастроф и другие разделы теории нелинейной динамики [5] позволяют аналитически исследовать динамику довольно простых систем (2-3 порядка). Это уровень нейрона (синапса). На уровне РНС эти теории не работают. Поэтому в общем случае для РНС произвольной структуры единственный возможный способ анализа динамики – численное моделирование и вычисление различных показателей динамики. Для РНС со сложными динамическими режимами можно вычислить специальные показатели, характеризующие эти режимы. Эти показатели определяются с использованием теории вероятности и математической статистики, теории случайных процессов, теории информации, теории детерминированного хаоса, теории фракталов и теории синхронизации.

*Переходные и установившиеся режимы.* Динамику РНС можно представить в виде суммы двух составляющих – установившейся и переходной. Переходная составляющая со временем затухает, после чего остается только постоянная составляющая. Для распознавания образов используются переходные и установившиеся составляющие (режимы) [1, 2, 6, 7]:

в переходном режиме после подачи входного образа его динамика накладывается на динамику РНС – по «снимку» результирующей динамики через некоторое время можно распознать входной образ; в данном случае используется свойство кратковременной памяти – учета предыстории изменения входного сигнала (стимула);

в установившемся режиме после подачи входного образа через какое-то время система выходит на аттрактор (фиксированное состояние, предельный цикл или хаотический аттрактор) – по этому аттрактору можно распознать входной образ.

*РНС с устойчивыми режимами.* РНС

данного класса являются устойчивыми в большом. После изменения их состояния или после подачи входного сигнала через некоторое время они сходятся к устойчивым точкам равновесия – фиксированным аттракторам.

Самым известным представителем таких РНС является НС Хопфилда (НСХ). Это полносвязная РНС с симметричной матрицей связей. Цель обучения НСХ – задать весовые коэффициенты таким образом, чтобы аттракторы совпадали с распознаваемыми образами.

На макроуровне НСХ описывается специальной функцией энергии (функцией Ляпунова). В ходе функционирования НСХ ее энергия уменьшается до тех пор, пока НСХ не попадет в состояние, соответствующее аттрактору. НСХ получила два основных применения – ассоциативная память и оптимизация.

НСХ породила целое направление в исследовании РНС – аттракторные НС [15]. Были предложены разные модификации, улучшающие ее основные недостатки (небольшой объем памяти, наличие ложных аттракторов – химер), а также скорость работы. В простейшем случае изменялся алгоритм обучения и, как следствие, – связность между нейронами. В некоторых подходах состояние каждого нейрона описывалось не одним числом, а вектором. Также использовались более сложные нелинейные модели нейронов и связей (терминальные аттракторы, синергетический компьютер Хакена). Во всех случаях принцип работы НС оставался тем же, что и у НСХ.

В целом, несмотря на большое количество моделей РНС с устойчивыми режимами, широкого распространения они не получили, а стали скорее удобными моделями для исследования (НСХ). В конечном итоге критике была подвергнута идея точечных аттракторов и предсказуемая динамика [17]:

в человеческом мозге основные вычисления осуществляются через хаотическую динамику, а сходимости к чему-то постоянному означает состояние покоя;

РНС с точечными аттракторами не могут обрабатывать динамические образы.

*РНС с колебательными режимами или осцилляторные НС.* Осцилляторными НС (ОНС) называются НС, состояние которых изменяется во времени по периодическому или близкому к периодическому закону. Иногда их еще называют НС с аттракторами типа «предельный цикл». С точки зрения биологии в основе ОНС лежит гипотеза о взаимосвязи феномена восприятия и колебательной активности нейронов.

Все модели ОНС различаются по следующим главным параметрам:

- тип элементов-осцилляторов;
- тип связей;
- структура.

В качестве осцилляторов могут использоваться детальные модели биологических нейронов (Ходжкина–Хаксли), упрощенные модели биологических нейронов (Хиндмарш–Роуз, Фитцху–Нагумо, Ижикевича), модели «интегрирование и возбуждение», модели популяций нейронов (Вилсона–Кована), а также модели физических осцилляторов, не имеющие никакого отношения к биологическому нейрону (генератор Ван-Дер-Поля, фазовый осциллятор) [2, 5, 17]. Главное требование к осциллятору заключается в том, чтобы он формировал на выходе колебательную активность. Одним из ключевых свойств элементов ОНС является их частота. В ОНС все элементы могут иметь одну частоту, либо быть разбиты на группы с разными частотами.

Тип связей зависит от модели осциллятора. Если это модель биологического нейрона, то связи осуществляются через химические синапсы, если нет, то через электрические. В зависимости от модели осциллятора электрическая связь может быть линейной или нелинейной. Связи могут обладать свойствами саморегуляции (внутренняя динамика) и пластичности (самообучение).

Наконец, наиболее важной характеристикой, отличающей одну модель осцилляторной НС от другой, является ее структура. Структура включает в себя следующие составляющие:

топология (связность) — полносвязная, с локальными связями, со слабыми связя-

ми и др.;

баланс возбуждающих и подавляющих связей (осцилляторов), наличие глобального ингибитора;

распределение силы связей, а также задержка (гомогенные и гетерогенные).

Все исследования ОНС можно разделить на две группы:

математическое исследование явлений синхронизации в ОНС;

применение ОНС для построения моделей обонятельной, зрительной коры и двигательной системы, а также памяти и внимания.

В рамках первой группы рассматривается проблема формирования нейронными структурами сложных пространственно-временных неоднородных структур (спиральных и плоских волн, солитонов). В рамках второй группы рассматривается применение ОНС для моделирования различных функций головного мозга. Исследованию ОНС посвящено большое количество научных работ. Часть из них затрагивает такой вопрос, как возникновение более сложных динамических режимов — хаотических.

*РНС с хаотическими режимами.* Хаос составляет базовую форму коллективной нейронной активности для всех процессов и функций восприятия. Он играет роль контролируемого источника шума с целью обеспечения непрерывного доступа к предыдущим запомненным образам и запоминанию новых. Хаос позволяет системе быть всегда активной, избавляя ее от необходимости пробуждаться или входить в устойчивое состояние каждый раз, когда изменяется входное воздействие.

Большинство исследователей сходятся во мнении, что наилучшим с точки зрения хранения и обработки информации является режим упорядоченного хаоса [5]. С одной стороны, в этом режиме есть все преимущества хаоса, с другой стороны, этим режимом можно управлять.

Множество состояний, через которое проходит траектория хаотической системы, называется *хаотическим аттрактором*. В зависимости от количества, силы связей и меры хаотичности элементов хаотическая

система может находиться в различных динамических режимах [16]. Наибольший интерес представляет упорядоченный режим работы решеток с формированием кластеров — хаотическая синхронизация. Этот режим находится между двумя крайними режимами — полным порядком и полным хаосом. Он позволяет использовать полезные свойства хаотических систем для решения задач кластеризации [1, 9].

**РНС с непрерывным пространством состояний.** При очень большом количестве нейронов РНС рассматривается не как дискретная в пространстве система, а как непрерывная среда. В этом случае для описания используются волновые уравнения [11]. Они определяют динамику нейронного поля  $u(r, t)$ , формируемого средой. Поле  $u(r, t)$  характеризует активность РНС в момент времени  $t$  в точке с координатами  $r$ . Для задания связности используется специальная функция связи  $w(r)$ , определяющая зависимость силы связи от расстояния, например функция типа «мексиканская шляпа».

Динамику РНС-сред исследуют с помощью специального метода — анализа неустойчивости Тьюринга. Волновое уравнение среды имеет в качестве одного из состояний равновесия состояние покоя. Далее уравнение линеаризуется в окрестности этого состояния и вычисляются показатели, характеризующие тип динамики. При определенном подборе параметров (весовой функции, функции активации) динамика может быть неустойчивой. Такая неустойчивость носит название *неустойчивость Тьюринга* в честь ученого, впервые ее обнаружившего. Эта неустойчивость приводит к образованию стационарных неоднородных структур, названных *диссипативными структурами* или *паттернами Тьюринга* [10]. *Диссипативный* означает, что структура появляется за счет диссипативных процессов рассеивания энергии в открытой системе.

Вид и форма паттернов Тьюринга зависят в первую очередь от числа измерений среды. Для одномерной среды характерно появление глобальных периодических устойчивых структур, получивших назва-

ние «бугорковые» (bump) или «непрерывные» (continuous) аттракторы [14].

При определенных условиях в средах могут появляться более сложные динамические режимы — распространяющиеся волны. Они могут перемежаться с состояниями покоя, а могут существовать постоянно в форме солитонов. В двумерных средах были обнаружены спиральные волны. Есть гипотеза о том, что в реальных биологических системах из-за гетерогенности и нарушения симметрии волны не могут распространяться бесконечно и быстро затухают.

В целом процессы, наблюдаемые в средах, также могут быть обнаружены и в дискретных решетках, но для этого необходима соответствующая размерность для дискретного варианта.

Исследование полезных для решения задач обработки информации свойства таких РНС пока находится в начальной стадии. Поэтому в дальнейшем анализе проблем обучения РНС основное внимание сосредоточено на РНС с дискретным пространством состояний.

## Обучение РНС

### Классические подходы к обучению РНС.

Одно из важнейших свойств НС — возможность обучения. Под обучением, согласно [8] понимается процесс, в котором параметры НС настраиваются за счет моделирования среды, в которую сеть помещена. Тип обучения зависит от способа подстройки этих параметров [8, 11]. Для обучения в НС из внешней среды должны поступать сигналы, изменяющие свободные параметры НС, что влечет за собой другую ответную реакцию НС на входные сигналы.

Обучение РНС в общем случае классифицируется как NP-полная задача даже для НС с одним скрытым слоем [12]. Для ряда случаев (определенная структура РНС и простые обучающие примеры) обучение имеет полиномиальную сложность.

К РНС применимо утверждение: «Чем сложнее устроена обучаемая машина, тем сложнее алгоритм обучения». Поскольку РНС сложнее обычных НС прямого рас-

пространения, то и алгоритмы их обучения сложнее алгоритмов обучения последних.

Для обучения РНС можно выделить несколько различных по своей сути групп подходов:

- использование известных соотношений между параметрами РНС и ее динамикой; пример – НСХ, НС Хакена и другие аттракторные РНС, в которых путем задания весовых коэффициентов можно закодировать желаемые аттракторы на базе правила Хэбба [8];

- использование алгоритмов обучения с учителем на базе методов оптимизации по типу алгоритма обратного распространения ошибки; пример – алгоритмы обратного распространения во времени (Back Propagation Through Time, ВРТТ), рекуррентное обучение в реальном времени (Real-Time Recurrent Learning, RTRL), рекуррентное обратное распространение (Recurrent Back Propagation, RBP) [8], алгоритмы, использующие Калмановскую фильтрацию; из-за того, что в этих алгоритмах функционал оптимизации формируется как сумма ошибок на некотором временном интервале, и при этом параметры на каждом шаге настраиваются одни и те же, существуют проблемы со сходимостью методов, временем работы и вычислительными затратами;

- использование алгоритма обучения с учителем при рассмотрении РНС как частично рекуррентной – сигналы с обратных связей рассматриваются как отдельные входные сигналы – контекстные нейроны, в результате задача обучения упрощается и сводится к обычным алгоритмам обучения НС прямого распространения; к данному типу обучения относятся НС Элмана, НС Джордана и другие НС;

- использование алгоритмов обучения без учителя (правило обучения Кохонена, синаптическая пластичность импульсных НС);

- отсутствие обучения в классическом смысле изменения параметров системы (неявное обучение) – настройка весовых коэффициентов случайными значениями; при этом функции обучения возложены

на специальное устройство – считыватель, который занимается классификацией динамики РНС; данный принцип лег в основу новой парадигмы НС – резервуарные вычисления [18].

Последний вариант обучения (без явного обучения через изменение весовых коэффициентов самой РНС) представляет собой новую парадигму обучения, в которой способ связи с окружающей средой представлен в изменении фазовых портретов поведения РНС и формировании определенной динамики в ответ на определенное воздействие (можно сопоставить с кратковременной памятью в НС прямого распространения при отклике на входное воздействие) [8]. Новая парадигма обучения связана с новым подходом в вычислениях – резервуарными вычислениями.

**Резервуарные вычисления.** Сложность и неэффективность существующих алгоритмов обучения РНС вынуждает искать новые подходы, стратегии для использования их вычислительных возможностей. Одним из таких подходов стало новое направление – резервуарные вычисления (РВ, Reservoir Computing, RC) [17, 18, 20].

Основная идея РВ заключается в использовании РНС как резервуара с богатой динамикой и мощными вычислительными возможностями. При этом резервуар формируется случайным образом, что исключает необходимость проводить его обучение. При подаче на вход резервуара непрерывного сигнала он попадает в определенный динамический режим или состояние. Это состояние зависит от входного сигнала. Резервуар формируется таким образом, чтобы для похожих входных сигналов это состояние было похожим, а для разных – разным. Выход резервуара соединен со специальными устройствами – считывателями, которые по состоянию резервуара решают поставленную задачу – производят классификацию, предсказание, кластеризацию [6]. В качестве считывателей могут использоваться статические обучающие машины с простыми алгоритмами обучения. Таким образом, резервуар интегрирует в своем состоянии динамику входного образа.



А считыватели по состоянию резервуара распознают входной образ.

Причины появления направления РВ связаны с тем, что статические модели НС в силу своих возможностей не могли решать задачу распознавания динамических образов, как и динамические НС с задержками, поэтому начали использоваться более сложные рекуррентные НС. Но проблема обучения РНС заставила искать новые подходы, и в результате появились резервуарные вычисления, сочетающие в себе богатую динамику РНС и мощные возможности по обучению статических НС. Было предложено несколько вариантов реализации резервуарных вычислений:

- сеть с эхо-состояниями (Echo State Network, ESN) [20], резервуар является обычной аналоговой РНС, функционирующей в дискретном времени;

- обратное распространение с декорреляцией (Backpropagation-Decorrelation, BPDC) [19]; модель похожа на ESN, используется алгоритм обучения BPDC, являющийся упрощением алгоритма RTRL; если в RTRL обучается вся РНС, то BPDC построен таким образом, чтобы обучались только считывающие модули;

- машина неустойчивых состояний (Liquid State Machine, LSM)[17], резервуар является импульсной РНС;

- эволюно (Evolino – EVolution of systems with Linear Output) [21] – резервуар составлен из специальных нейронов, моделирующих длительную кратковременную память, выходной слой – линейный;

- временная РНС (temporal RNN) – биологические резервуары [13] и представляющие собой настоящие кортикальные НС;

- прочие резервуары, представляющие собой произвольные динамические системы, как чисто аналитические, так и реальные оптические, биологические, физические, квантовые и другие системы.

Для формирования резервуаров существует три базовых способа [18]:

- использование общих рекомендаций для создания «хороших» резервуаров вне зависимости от решаемой задачи – выбор

топологии, связности, силы связей, задержек;

- адаптация резервуара – обучение без учителя с использованием примеров входных данных:

- на глобальном уровне – резервуар формируется так, чтобы при заданных входных данных он находился в нужном динамическом режиме и обладал необходимыми свойствами (например, разделения);

- на локальном уровне – самоорганизация резервуара с использованием правил синаптической пластичности при подаче входных данных;

- обучение резервуара с учителем с использованием примеров входных и соответствующих им выходных данных; в этом случае для конкретной задачи генерируется множество резервуаров с разными параметрами, оценивается качество распознавания для каждого резервуара и из них выбирается наилучший.

В базовом варианте резервуарные вычисления обладают универсальными вычислительными возможностями в смысле аппроксимации произвольных нелинейных динамических систем с затухающей памятью. Если добавить *обратные связи* со считывателя на резервуар, то появляется возможность аппроксимации систем с постоянной памятью, т. е. машин Тьюринга.

Считыватели являются простыми статическими обучающими машинами: взвешенная линейная регрессия, адалина (с обучением в реальном времени по методу наименьших квадратов), персептрон,  $k$  ближайших соседей, машины опорных векторов, статические НС.

Для представления о классах динамических систем на базе РВ и степени сложности систем на биологически подобных нейронах приведем пример: для описания МНС из 125 нейронов с наличием шумовой составляющей и динамическими синапсами необходимо 250 связанных нелинейных стохастических ДУ для нейронов и около 400 разностных уравнений третьего порядка для синапсов.

Использование существующих показателей динамических систем (например, эн-

тропии Колмогорова, фрактальной размерности, показателей синхронизации и др.) для определения режимов функционирования РНС даже для простых случаев хаотических нейронных сетей [1, 9] оказывается неэффективным [3, 4] в силу специфики использования динамических режимов для решения задач обработки информации. Это необходимость выявления особенностей именно совместной относительной динамики элементов в неоднородных нелинейных динамических системах большой размерности. Поэтому для анализа работы сложных РНС требуется использование методов ки-

бернетической физики и методов из новых разделов нелинейной динамики [4].

В данной статье проведен аналитический обзор существующих подходов к исследованию РНС и показаны предпосылки к появлению нового направления в области нейроинформатики – резервуарных вычислений. Проведенная систематизация разновидностей РНС и методов их обучения позволили предложить обобщенную классификацию РНС и показать место РНС с неустойчивой динамикой среди множества других классов РНС.

#### СПИСОК ЛИТЕРАТУРЫ

1. **Бендерская, Е.Н.** Осцилляторные нейронные сети с хаотической динамикой в задачах кластерного анализа [Текст] / Е.Н. Бендерская, С.В. Жукова // Нейрокомпьютеры: разработка, применение. – 2011. – № 7. – С. 74–86.
2. **Бендерская, Е.Н.** Моделирование нейронной активности мозга и бионспирированные вычисления [Текст] / Е.Н. Бендерская, К.В. Никитин // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. –СПб.: Изд-во Политехн. ун-та, 2011. –№ 6-2 (138). – С. 34–40.
3. **Бендерская, Е.Н.** Возможности применения некоторых характеристик синхронизации для выявления самоорганизующихся кластеров в осцилляторной нейронной сети с хаотической динамикой [Текст] / Е.Н. Бендерская // Нейрокомпьютеры: разработка, применение: научно-технический журнал. – М.: Научный центр нейрокомпьютеров, 2012. – № 11. – С. 69–73 .
4. **Магницкий, Н.А.** Новые методы хаотической динамики [Текст] / Н.А. Магницкий, С.В. Сидоров. – М.: Едиториал УРСС, 2004. – 320 с.
5. **Малинецкий, Г.Г.** Нелинейная динамика и хаос. Основные понятия: Учеб. пособие [Текст] / Г.Г. Малинецкий, А.Б. Потапов. – М.: Ком-Книга, 2006. – 240 с.
6. **Никитин, К.В.** Новый подход к применению рекуррентных нейронных сетей для решения задач распознавания образов [Текст] / К.В. Никитин, Е.Н. Бендерская // Научно-технические ведомости СПбГПУ. –СПб.: Изд-во Политехн. ун-та, 2007. – № 4-2 (52). – С. 85–92.
7. **Тюкин, И.Ю.** Адаптация в нелинейных динамических системах [Текст] / И.Ю. Тюкин, В.А. Терехов. –М.: Изд-во ЛКИ, 2008. – 384 с.
8. **Хайкин, С.** Нейронные сети: полн. курс [Текст] / С. Хайкин; Пер. с англ. – 2-е изд., испр. – М. –СПб. –Киев: ИД «Вильямс», 2006. – 1103 с.
9. **Benderskaya, E.N.** Nonlinear Trends in Modern Artificial Intelligence: A New Perspective [Text] / E.N. Benderskaya // Beyond AI: Interdisciplinary Aspects of Artificial Intelligence. Topics in Intelligent Engineering and Informatics. – Springer, 2013. – Vol. 4. – P. 113–124.
10. **Benderskaya, E.N.** Multidisciplinary Trends in Modern Artificial Intelligence: Turing's Way [Text] / E.N. Benderskaya, S.V. Zhukova // AIECM – Turing 2013, Book Chapters: Artificial Intelligence, Evolutionary Computation and Metaheuristics. – Springer, 2013. – P. 320–343.
11. **Coombes, S.** Waves, bumps, and patterns in neural field theories [Text] / S. Coombes // Biological Cybernetics. – 2005. – Vol. 93. – № 2. – P. 91–108.
12. **Dasgupta, B.** On the Complexity of Training Neural Networks with Continuous Activation Functions [Text] / B. Dasgupta, H. Siegelmann, E.D. Sontag // IEEE Transactions on Neural Networks. – 1995. – Vol. 6. – № 6. – P. 1490–1504.
13. **Dominey, P.F.** Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning [Text] / P.F. Dominey // Biological Cybernetics. – 1995. – Vol. 73. – № 3. – P. 265–274.
14. **Dunin-Barkowski, W.L.** Hebb-Hopfield neural networks based on one-dimensional sets of neuron states [Text] / W.L. Dunin-Barkowski, N.B. Osovets // Neural Processing Letters. – 1995. – Vol. 2. – № 5. – P. 28–31.
15. **Feng, J.** Fixed-point attractor analysis for a class of neurodynamics [Text] / J. Feng, D. Brown // Neural Computation. – 1998. – Vol. 10. – P. 189–213.

16. **Kaneko, K.** Life: an introduction to complex systems biology [Text] / K. Kaneko. – Berlin: Springer-Verlag, 2006. – 369 p.

17. **Maass, W.** Real-time computing without stable states: a new framework for neural computations based on perturbations [Text] / W. Maass, T. Natschläger, H. Markram // Neural Computation. – 2002. – Vol. 11. – P. 2531–2560.

18. **Schrauwen, B.** An overview of reservoir computing theory, applications and implementations [Text] / B. Schrauwen, D. Verstraeten, J.V. Campenhout // Proc. of the 15th European Symp. on Artificial Neural Networks. – 2007. – P. 471–482.

19. **Steil, J.J.** Backpropagation-decorrelation: recurrent learning with  $O(N)$  complexity [Text] / J.J. Steil // Proc. IJCNN. – 2004. – Vol. 1. – P. 843–848.

20. **Jaeger, H.** The «echo state» approach to analyzing and training recurrent neural networks: GMD report 148 [Text] / H. Jaeger. – German National Research Center for Information Technology. – St. Augustin, 2001. – 43 p.

21. **Schmidhuber, J.** Training Recurrent Networks by Evolino [Text] / J. Schmidhuber, D. Wierstra, M. Gagliolo, F. Gomez // Neural computation. – 2007. – Vol. 19. – № 3. – P. 757–779.

## REFERENCES

1. **Benderskaia E.N., Zhukova S.V.** Ostsilliatornye neironnye seti s khaoticheskoi dinamikoii v zadachakh klaster'nogo analiza / Neirokomp'iutery: razrabotka, primeneniye; Radiotekhnika. – 2011. – № 7. – S. 74–86. (rus)

2. **Benderskaia E.N., Nikitin K.V.** Modelirovaniye neironnoi aktivnosti mozga i bionspirovannyye vychisleniia / Nauchno-tekhnicheskie vedomosti SPbGPU. Informatika. Telecommunicacii. Upravleniye. – St.-Petersburg: Izd-vo Politehn. un-ta, 2011. – № 6-2(138). – S. 34–40. (rus)

3. **Benderskaia E.N.** Vozmozhnosti primeneniia nekotorykh kharakteristik sinkhronizatsii dlia vyavleniia samoorganizuiushchikhsia klasterov v ostsilliatornoii neironnoi seti s khaoticheskoi dinamikoii / Neirokomp'iutery: razrabotka, primeneniye: nauchno-tekhnicheskii zhurnal. – Moscow: Nauchnyi tsentr neirokomp'iuterov, 2012. – № 11. – S. 69–73. (rus)

4. **Magnitskii N.A., Sidorov S.V.** Novyye metody khaoticheskoi dinamiki. – Moscow: Editorial URSS, 2004. – 320 s. (rus)

5. **Malinetskii G.G., Potapov A.B.** Nelineinaia dinamika i khaos. Osnovnyye poniatii. – Moscow: KomKniga, 2006. – 240 s. (rus)

6. **Nikitin K.V., Benderskaia E.N.** Novyyi podkhod k primeneniuiu rekurrentnykh neironnykh seteii dlia resheniia zadach raspoznavaniia obrazov / Nauchno-tekhnicheskie vedomosti SPbGPU. – St.-Petersburg: Izd-vo Politehn. un-ta, 2007. – № 4-2 (52). – S. 85–92. (rus)

7. **Tiukin I.Iu., Terekhov V.A.** Adaptatsiia v nelineinykh dinamicheskikh sistemakh. – Moscow: Izd-vo LKI, 2008. – 384 s. (rus)

8. Khaikin S. Neironnye seti: poln. Kurs; per. s angl; 2-e izd., ispr. – Moscow–St.-Petersburg–Kiev: ID “Vil'iams”, 2006. – 1103 s. (rus)

9. **Benderskaya E.N.** Nonlinear Trends in Modern Artificial Intelligence: A New Perspective / Beyond AI: Interdisciplinary Aspects of Artificial Intelligence.

Topics in Intelligent Engineering and Informatics. – Springer, 2013. – Vol. 4. – P. 113–124.

10. **Benderskaya E.N., Zhukova S.V.** Multidisciplinary Trends in Modern Artificial Intelligence: Turing's Way / AIECM – Turing 2013, Book Chapters: Artificial Intelligence, Evolutionary Computation and Metaheuristics. – Springer, 2013. – P. 320–343.

11. **Coomes S.** Waves, bumps, and patterns in neural field theories / Biological Cybernetics. – 2005. – Vol. 93. – № 2. – P. 91–108.

12. **Dasgupta B., Siegelmann H., Sontag E.D.** On the Complexity of Training Neural Networks with Continuous Activation Functions / IEEE Transactions on Neural Networks. – 1995. – Vol. 6. – № 6. – P. 1490–1504.

13. **Dominey P.F.** Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning / Biological Cybernetics. – 1995. – Vol. 73. – № 3. – P. 265–274.

14. **Dunin-Barkowski W.L., Osovets N.B.** Hebb-Hopfield neural networks based on one-dimensional sets of neuron states / Neural Processing Letters. – 1995. – Vol. 2. – № 5. – P. 28–31.

15. **Feng J., Brown D.** Fixed-point attractor analysis for a class of neurodynamics / Neural Computation. – 1998. – Vol. 10. – P. 189–213.

16. **Kaneko K.** Life: an introduction to complex systems biology. – Berlin: Springer-Verlag, 2006. – 369 p.

17. **Maass W., Natschläger T., Markram H.** Real-time computing without stable states: a new framework for neural computations based on perturbations / Neural Computation. – 2002. – Vol. 11. – P. 2531–2560.

18. **Schrauwen B., Verstraeten D., Campenhout J.V.** An overview of reservoir computing theory, applications and implementations / Proc. of the 15th European Symp. on Artificial Neural Networks. – 2007. – P. 471–482.

19. **Steil J.J.** Backpropagation-decorrelation: recurrent learning with  $O(N)$  complexity / Proc. IJCNN. – 2004. – Vol. 1. – P. 843–848.
20. **Jaeger H.** The «echo state» approach to analyzing and training recurrent neural networks: GMD report 148. – St. Augustin: German National Research Center for Information Technology, 2001. – 43 p.
21. **Schmidhuber J., Wierstra D., Gagliolo M., Gomez F.** Training Recurrent Networks by Evolino / Neural computation. – 2007. – Vol. 19. – № 3. – P. 757–779.
- 

**БЕНДЕРСКАЯ Елена Николаевна** – доцент кафедры компьютерных систем и программных технологий Санкт-Петербургского государственного политехнического университета, кандидат технических наук.

195251, Россия, Санкт-Петербург, Политехническая ул., д. 21.  
Тел. +7(812)297-42-18, e-mail: helen.bend@gmail.com

**BENDERSKAYA, Elena N.** *St. Petersburg State Polytechnical University.*

195251, Politekhnikeskaya Str. 21, St.-Petersburg, Russia.  
E-mail: helen.bend@gmail.com

**НИКИТИН Кирилл Вячеславович** – аспирант кафедры компьютерных систем и программных технологий Санкт-Петербургского государственного политехнического университета.

195251, Россия, Санкт-Петербург, Политехническая ул., д. 21.  
Тел. +7(812)297-42-18, e-mail: execiter@mail.ru

**NIKITIN, Kirill V.** *St. Petersburg State Polytechnical University.*

195251, Politekhnikeskaya Str. 21, St.-Petersburg, Russia.  
E-mail: execiter@mail.ru