

УДК: 519.7, 612.8

О механизме обучения нейронной сети как адаптивной системы

Лоренц В.А.*¹, Гавриков В.Л.**², Хлебопрос Р.Г.***^{1,3}

¹ФГАОУ ВПО Сибирский Федеральный Университет,

²Красноярский Государственный Педагогический Университет им. В.П.Астафьева,

³Красноярский Научный Центр СО РАН

Аннотация. Анализируются динамики ошибок обучения нейронных сетей, полученные при одинаковых условиях решаемой задачи и параметрах сетей, но с разными начальными условиями. А также предпринята попытка предложить возможную теоретическую модель единого механизма обучения искусственных и естественных адаптивных систем.

Ключевые слова: нейронная сеть, обучение, ошибка обучения.

ВВЕДЕНИЕ

Проблема обучения в настоящее время представляет интерес для ряда научных дисциплин, таких как педагогика, психология, нейрофизиология, информатика, генетика и т.п. [1–3], и это неугасающее внимание обусловлено необходимостью глубже проникнуть в механизм обучения как жизненно и эволюционно важного процесса, поскольку это ведет к пониманию не только этого процесса, но и пониманию живого вообще.

Биофизический подход к этой проблеме основан на изучении искусственных адаптивных систем, таких как нейронные сети. Под искусственной адаптивной системой понимается модель, в которой в результате изменения характеристик внешних и внутренних свойств объекта происходит соответствующее изменение структуры и параметров функционирования с целью обеспечения стабильности [4]. И модельный подход при исследовании механизмов обучения представляется чуть ли не единственно возможным. Нейросети, воспроизводящие в некоторой степени способность живых организмов к самообучению, являются универсальной эвристической моделью живого [5]. Поскольку современная наука давно отошла от попыток воспроизвести с помощью нейросетей полноценную нервную деятельность, то они рассматриваются лишь как структуры, воспроизводящие реализацию конкретных функций живого [6]. Это обстоятельство открывает возможность как бы произвести редукцию проблемы: наше исследование здесь заключается в создании абстрактного процесса обучения абстрактной модели (нейронной сети) и выявлении общих свойств, понятий механизмов, одинаково характерных для процессов обучения искусственных и естественных адаптивных систем.

Сравнительный анализ закономерностей обучения человека и животных привлекал многих исследователей, однако нами в ряде работ было показано, что качественно эти процессы поддаются приемлемому описанию с помощью аппарата теории

* forest555@yandex.ru ,

** gavrikov@mail.kspu.ru

*** olikru@yandex.ru

катастроф [1,7]. Нами также было выявлено принципиальное сходство между процессом обучения нейронной сети, с одной стороны, и животных и человека, с другой [8]. Большинство динамик обучения нейронной сети при решении простой задачи со скрытой закономерностью характеризуются постепенным снижением ошибки обучения, что роднит этот процесс с «методом проб и ошибок». С другой стороны, при некоторых условиях наблюдается эффект резкого выхода на определенный уровень «понимания задачи» нейросетью, сопровождающийся улучшением обучаемости, что характерно для понимания посредством *инсайта* у высших животных. Главная неожиданность здесь заключается в том, что обнаруженный эффект не вписывался в первоначально встроенный алгоритм обучения нейронной сети и может быть интерпретирован как следствие влияния начальных условий и уникальной структуры нейронной сети. Благодаря открытию данного эффекта родилась гипотеза о существовании механизма обучения, одинаково характерного для любой адаптивной системы. Задачами настоящей статьи, таким образом, являются исследования нескольких феноменов обучения нейронных сетей:

А) Обучение одинаковых нейронных сетей одной и той же задаче при разных начальных условиях может характеризоваться разнообразием траекторий обучения. Имеют ли эти динамики форму абсолютно разную в каждом случае или подобную друг другу, либо имеет место несколько типов траекторий?

Б) Частотное распределение нейросетей по траекториям их обучения в каждый момент времени может быть равномерным, независимым, а может демонстрировать группировку по нескольким классам. Какой вариант имеет место в реальности?

И в качестве цели настоящего исследования мы преследуем возможность сформулировать некие теоретические механизмы реализации данных явлений.

МЕТОДИКА

В эксперименте была использована нейросетевая надстройка «Модели», реализующая оперативный синтез аналитических моделей, функционирующая в среде MS Excel и в математическом отношении осуществляющая нелинейную многомерную регрессию. В качестве интерполирующего используется один из вариантов многомерных представлений в виде интегралов Фурье с заменой интегралов конечными суммами. Используемая нейронная сеть по своей структуре может называться двухслойной, так как содержит «основной» слой нелинейных элементов типа $A \cdot \sin(\omega \cdot x + \varphi)$ и дополнительный слой линейных элементов.

При оптимизации используется метод быстрого вычисления многомерных градиентов, известный в рамках нейросетевой идеологии под именем *back propagation* (обратного распространения ошибки) или метод множителей Лагранжа, лежащий также в основе теории оптимального управления [9, 10]. Суть метода обратного распространения ошибки интуитивно проста. В нулевой момент времени нейронная сеть, состоящая из некоторого количества нейронов (условных элементарных компонентов сети, выполняющих роль преобразователей входных сигналов), имеет различные стартовые, заданные случайным образом, величины связей между нейронами (весовые коэффициенты), – в данном исследовании назовем их *начальные условия*, – которые в каждый момент времени (на каждой *итерации*) изменяются по следующему принципу. Во-первых, оценивается, как каждая пара нейронов повлияла на полученное выходное значение сети. А затем увеличиваются те значения весовых коэффициентов связей между нейронами, которые уменьшали ошибку обучения, и снижается величина связи между теми нейронами, которые не способствовали решению задачи. И таким образом достигается приближение предсказаний нейронной сети к желаемому результату с высокой точностью.

Среди задаваемых параметров, используемых в работе, следующие: *количество*

нейронов, количество итераций, размер обучающей и тестовой выборки. Обучающая выборка содержит последовательность входных данных нейронной сети, используемых для настройки внутрисетевых связей, т.е. непосредственно для процесса обучения. А тестовая выборка – это совокупность данных, имеющих ту же внутреннюю закономерность, что и задачи обучающей выборки, но не включенных в нее. Отслеживаемыми параметрами нейронной сети являлись *ошибка обучения*, представляющая собой среднеквадратичное отклонение предсказаний сети от эмпирических входных данных обучающей выборки, и *ошибка прогноза*, т.е. среднеквадратичное отклонение предсказаний сети для тестовых задач, не участвующих в обучении.

Нейронной сети в данном эксперименте предлагалось решить задачу распознавания функции $\sin(x)$. Значения этой функции подавались на единственный вход нейронной сети, и они же являлись эталоном, с которым сеть должна была сравнивать прогнозируемые значения выхода. Из практики известно, что нейронные сети легко решают подобные задачи, и, таким образом, выбор задания был обусловлен требованием использования наиболее простой задачи, содержащей скрытую закономерность, которую данная нейросеть способна решить, и ход решения которой легче контролировать. В течение эксперимента отслеживалась обучаемость нейронной сети и динамика ошибки обучения. Под обучаемостью нейросети здесь понимается такое количественное значение ошибок обучения и прогноза, которое позволяет воспроизвести кривую синуса с приемлемой точностью, что, по нашему мнению, соответствует решению поставленной задачи в рамках данного эксперимента. Результаты всех экспериментов проверялись на отсутствие эффекта переобучения.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Как было отмечено выше, в определенных условиях нейронные сети демонстрируют в процессе обучения эффект резкого выхода на новый уровень прогресса в решении задачи, не встроенный изначально в их алгоритм обучения. Предположительно, таким образом нейросеть избегает «застывания» в локальных минимумах, что обеспечивает лучшую обучаемость системы. Другое предположение касается структуры распределения траекторий обучения одинаковых нейронных сетей, имеющих лишь разные начальные условия. В каждый момент времени нейросети могут быть распределены случайным образом в виде равномерного облака, размазанного по оси величины ошибки обучения, а также могут подчиняться какой-либо закономерности: нормальному распределению, либо мы сможем наблюдать их группировку около небольшого дискретного числа уровней.

Для проверки последней гипотезы было проведено 200 экспериментов, в ходе которых отслеживалась динамика ошибок обучения нейронных сетей с одинаковыми параметрами (количество нейронов – 9, количество итераций – 50) и условиями задачи, предлагаемой для решения, при разных начальных условиях. Примеры абсолютных значений ошибок обучения восьми нейросетей (из 200) в динамике отражены на рис. 1, из которого можно заметить, что в зависимости от начальных условий наблюдается разнообразие форм динамик обучения нейронных сетей. Особенно вариабельны начальные стадии траекторий падения ошибки обучения.

Визуальный анализ показывает, что выделяется несколько типов динамик обучения нейронных сетей. Один из них характеризуется постепенным снижением величины ошибки обучения нейронной сети, что не противоречит алгоритму ее работы и является теоретически ожидаемым (сплошные линии на рис. 1). Второй тип содержит медленную стадию снижения ошибки на первых итерациях, которая затем переходит в более быструю – достаточно резкий спад за 1–2 итерации – и затем имеет последующее замедление, как и в первом случае (жирный пунктир на рис. 1). Эффект резкого выхода

на новый, лучший уровень «понимания» задачи нейросетью является характерным свойством большинства нейронных сетей данного класса при решении плохо определенных задач со скрытой закономерностью [8]. Третий же тип динамик не содержит предварительной стадии с низкой скоростью падения ошибки обучения, и отличается значительным ее уменьшением уже на первых итерациях с последующим плавным снижением (светлый пунктир на рис. 1).

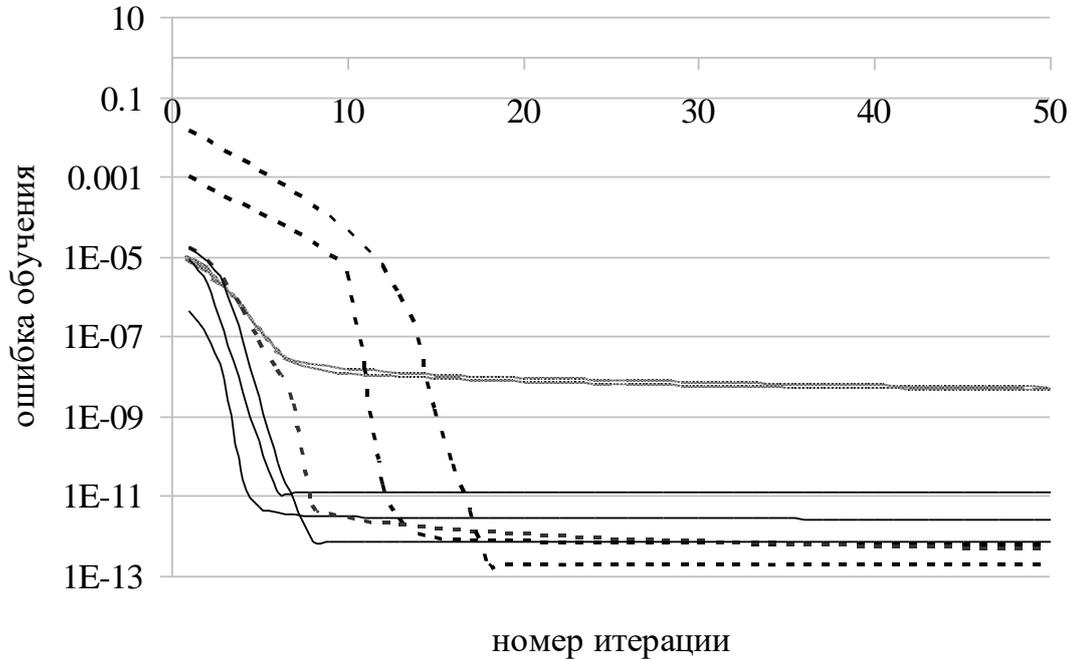


Рис. 1. Динамики абсолютных ошибок обучения 8 нейронных сетей с одинаковыми параметрами и разными начальными условиями (в логарифмической шкале).

Для более детального анализа далее была предпринята попытка построить частотные распределения ошибок обучения в фиксированные моменты времени (т.е. на уровне отдельных итераций). Полученные распределения по 23 логарифмическим классам показаны на рис. 2,а–г.

Рис. 2,а отображает распределение нейронных сетей по их ошибкам обучения на первой итерации в пределах некоторого небольшого числа логарифмических классов. Размер ошибок обучения сетей, входящих в класс, возрастает с номером класса, поэтому, как и следовало ожидать, на первой итерации нейронные сети сгруппированы по своим ошибкам среди классов с большим номером. Тем не менее, уже в этом распределении можно заметить группирование некоторых нейронных сетей вблизи небольшого числа выделенных классов. Это наблюдение может указывать на общую дальнейшую тенденцию к некоему «квантованию» нейронных сетей по уровням их обученности, ведь с момента запуска поиска нейросетью решения задачи и подачи случайного набора весовых коэффициентов прошла лишь одна рабочая итерация.

Картина, представленная на рис. 2,б и являющаяся срезом динамик ошибок обучения нейросетей в середине их траектории (в момент 9-ой итерации), демонстрирует разделение нейронных сетей среди нескольких классов с разными уровнями обученности, что является свидетельством снижения величины ошибки обучения по некоторым «избранным» траекториям в течение процесса обучения. По сравнению с картиной на рис. 2,а здесь более четко выделились отдельные группы нейронных сетей (отмечены на рис. 2,б стрелками).

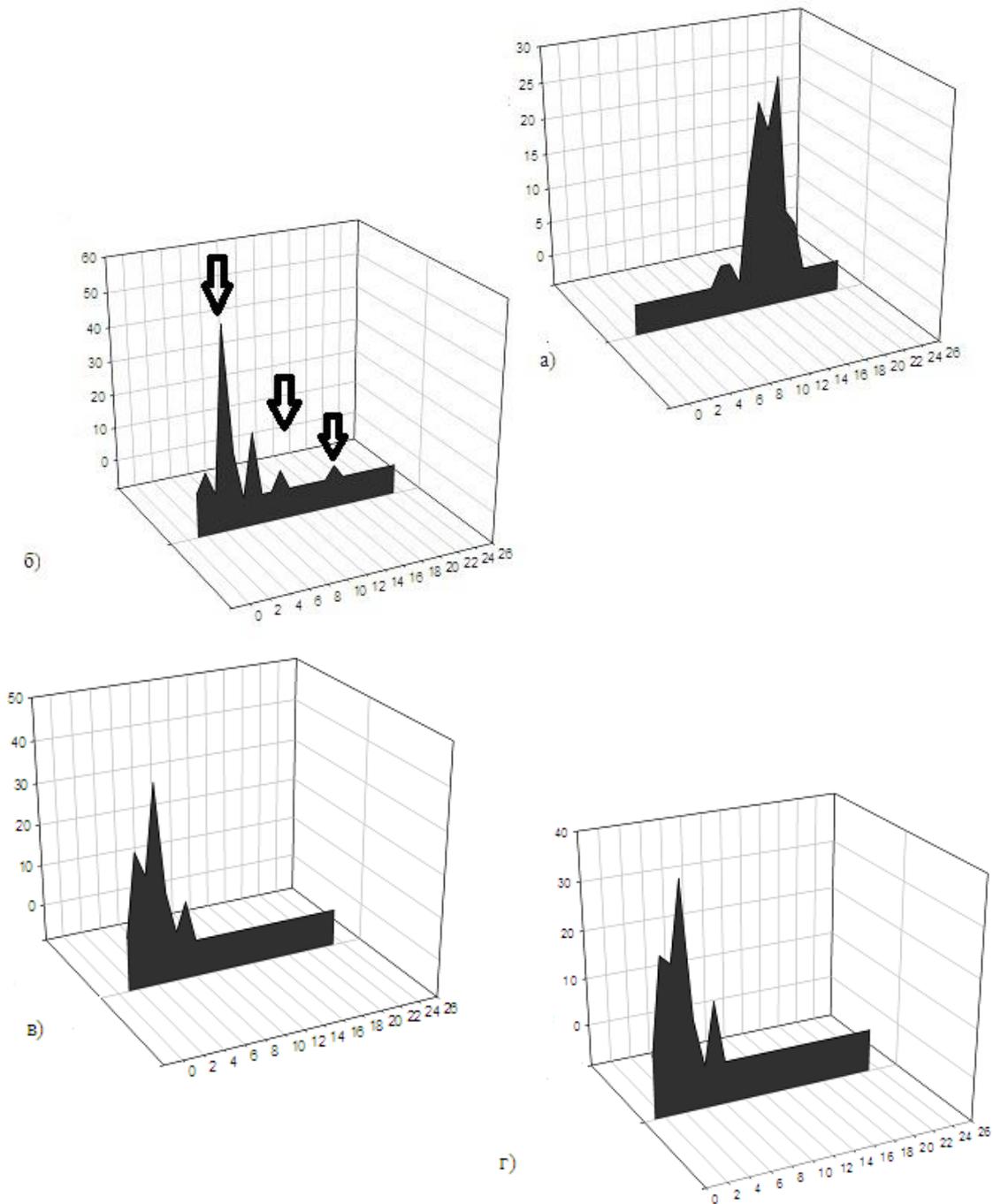


Рис. 2. Распределение абсолютных ошибок обучения нейронных сетей с одинаковыми параметрами по логарифмическим классам на 1-ой (а), 9-ой (б), 42-ой (в) и 51-ой (г) итерациях (OY – номер класса, OZ – число нейросетей, вошедших в класс).

Следует отметить, однако, что термин «траектория» в данном случае употребляется достаточно условно. Если принять во внимание хоть и небольшую, но все же разницу в значениях ошибок обучения одного класса, то более правильным будет употребление «коридор ошибок» или «избранный туннель ошибок». Тем не менее, если не останавливаться на нюансах терминологии, очевидным становится вывод о «квантованном» характере распределения ошибок обучения нейронной сети, а значит и об избирательности уровней «понимания» задачи, уровней обучаемости. Также заметим наличие одного явного максимума, которое может свидетельствовать о возможном существовании «наиболее вероятной из всех» траектории обучения нейронной сети при решении данной задачи.

Рис. 2,в и г, отражающие распределение ошибок обучения нейронных сетей на 42 и 51-ой итерациях, то есть в конце процесса обучения, демонстрируют сохранение тенденции к группированию ошибок обучения вблизи некоторых избранных значений. В данном случае большинство нейронных сетей сосредоточено в первом и втором классах, имеющих самые маленькие ошибки обучения, многообразие классов исчезает. Это распределение может свидетельствовать о том, что вблизи «технического нуля» величины ошибок обучения нейронных сетей может существовать и большее число выделенных траекторий, наблюдавшихся ранее, на 9-й итерации (рис. 2,б), которые сложно определить ввиду малости значений ошибок, которые отличаются по своим значениям от стартовых ошибок обучения на несколько порядков.

Односторонние переходы между отдельными уровнями обученности, вероятно, могут осуществляться спонтанно с кривых, обладающих низкой скоростью обучения, на кривые с более высокой скоростью обучения (но не наоборот) в направлении улучшения обучаемости. Природа этих переходов пока не ясна и требует дальнейших исследований, тем не менее, представляется возможным выдвинуть на теоретическом уровне предположения, касающиеся механизма формирования именно такой картины процесса обучения.

Поскольку динамики обучения, как живых систем, так и искусственных, начинаются с некоторого (ненулевого) уровня начальной ошибки, то эту ошибку условно можно назвать состоянием «незнания». Для нейронных сетей характерно иметь как достаточно высокое значение стартовой ошибки (1 и близкие к ней значения в нормированной шкале), так и низкие начальные значения ошибок (порядка 10^{-7}). Живые системы, как правило, в начальной фазе обучения новым задачам за пределами инстинктивного поведения всегда демонстрируют высокие значения ошибок. При этом скорость обучения естественных адаптивных систем на начальной стадии процесса обучения обычно невысока, в то время как искусственные нейронные системы могут иметь вначале как относительно низкую, так и относительно высокую скорость обучения. Отличительной особенностью живых систем является способность к забыванию, отвлечению внимания, следование по «неверному» пути решения, характеризующееся кратковременными увеличениями ошибки обучения. При этом нейронные сети запрограммированы таким образом, что во время процесса обучения ошибка может либо снижаться, либо не изменяться, но не может возрастать.

Таким образом, если воспользоваться качественным анализом поведения этих систем, то фазовый портрет их процессов обучения начинается с некоторой неустойчивой особой точки, предположим, с фокуса, двигаясь по которому на начальном этапе обучения, эти системы уменьшали величину ошибки обучения и медленно снижали скорость обучения. При этом естественные адаптивные системы в некоторых случаях могут демонстрировать прирост ошибки обучения, возвращаясь к состоянию «незнания», а алгоритм же большинства искусственных адаптивных систем позволяет совершать движение только в направлении снижения ошибки. Выше были расклассифицированы динамики обучения нейронных сетей, имеющих разные начальные условия, на три типа, два из которых являются крайними проявлениями возможных динамик: тип первый, характеризующийся монотонным снижением ошибки обучения (см. рис. 1, светлый пунктир), и второй тип (рис. 1, жирный пунктир), которому свойственно 3 фазы с разными скоростями снижения ошибки. Наличие нескольких фаз с разными скоростями обучения может свидетельствовать о переходе на траекторию обучения, описываемую уравнением с другими коэффициентами, то есть смене правой части уравнения типа:

$$X' = f(X, t),$$

где X – ошибка обучения нейронной сети. Таким образом, для динамик второго типа мы

получаем зависимость, обладающую кусочно-линейной формой или имеющую разрывы, фазовый портрет которой состоит из нескольких областей с различными фазовыми траекториями, а фазовые траектории не пересекаются нигде, кроме особых точек типа седла. Переход между разными уровнями обучаемости этих систем, вероятно, может осуществляться спонтанно в целях достижения лучшей обучаемости посредством движения по фазовой траектории, принадлежащей особой точке типа седло.

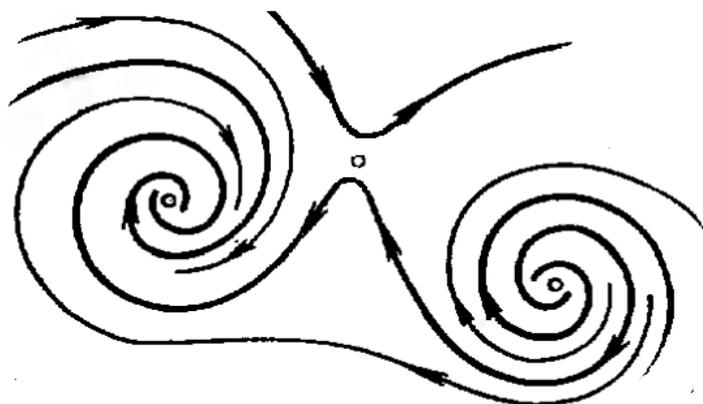


Рис. 3. Фазовый портрет, включающий три особые точки: два фокуса и седло между ними (динамика процесса обучения системы происходит справа налево).

Завершающая фаза обучения, как у искусственных, так и у естественных систем, в случае решения задачи характеризуется нулевым (или почти нулевым) значением ошибки обучения. При подходе к этой точке, которую можно назвать состоянием «понимания», скорость обучения может быть небольшой и даже нулевой. То есть качественно это устойчивая особая точка, завершающая процесс обучения. На рис. 3 показан возможный фазовый портрет описанного процесса, направление динамики которого показано стрелками.

Для визуальной демонстрации предложенной гипотезы построены фазовые портреты для нескольких нейронных сетей, демонстрирующих разные типы динамик (рис. 4–6). Портреты построены в осях величины ошибки обучения и нормированной скорости (т.е. скорости обучения, деленной на величину текущей ошибки).

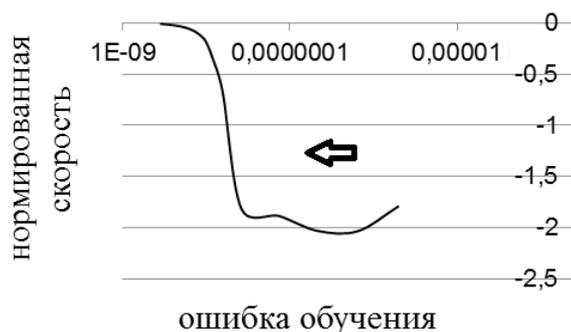


Рис. 4. Фазовый портрет динамики типа «кошка».

На рис. 4 показан фазовый портрет процесса обучения одной из нейронных сетей, динамику обучения которой можно отнести к первому типу (см. рис. 1, светлый пунктир). Примечательно то, что этот рисунок согласуется с теоретическим предположением о существовании в системе трех особых точек: двух фокусов и седла между ними обеспечивающего переход между траекториями с разными скоростями обучения (на рис. 3 - это самая нижняя кривая).

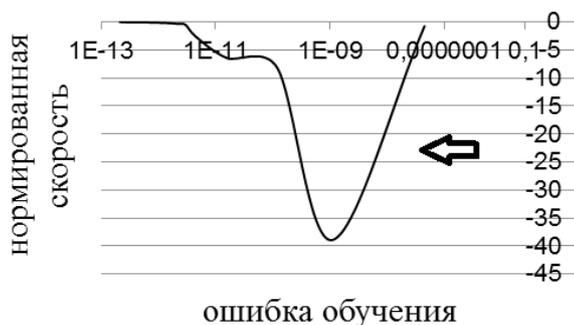


Рис. 5. Фазовый портрет динамики типа «обезьяна» с двумя изломами траектории обучения.

Рис. 5 демонстрирует фазовый портрет нейронной сети, которая обучалась решению задачи со скрытой закономерностью по траектории второго типа (см. рис. 1, жирный пунктир). Согласно теоретическим предположениям наблюдается начальное снижение ошибки обучения с относительно высокой скоростью, затем излом и переход фазовой траектории на дугу седла, и далее новый переход на конечный участок траектории с относительно низкой скоростью обучения, приводящий сеть к решению задачи.

Среди нейронных сетей, обучавшихся по второму типу, иногда встречались такие, у которых в динамике обнаруживалось более трех фаз с различными скоростями обучения. Среди типов, выделенных на рис. 1, этот случай не представлен, поскольку является сравнительно редким частным случаем динамики второго типа. На рис. 6 представлен фазовый портрет одной из таких сетей, у которой, по-видимому, фазовая траектория совершала переключение между тремя седлами. Таким образом, общая картина топологической местности, вдоль которой движется фазовая кривая нейронной сети, содержит $(2n - 1)$ особых точек, где n принадлежит множеству целых чисел.

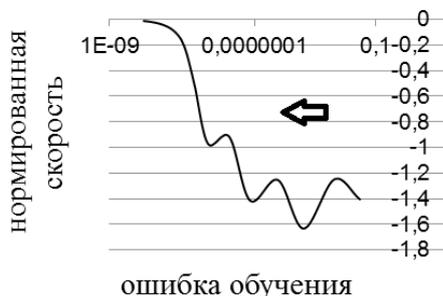


Рис. 6. Фазовый портрет динамики типа «обезьяна» с несколькими изломами траектории обучения.

Выдвинутые теоретические предположения могут распространяться и на процессы, связанные с обучением естественных адаптивных систем, поскольку обладают четко выраженной неоднородностью, которая может иметь смысл внутреннего переключателя, позволяющего адаптивной системе обучаться посредством мгновенного перехода между уровнями с разной степенью обученности.

ВЫВОДЫ

Проведенное экспериментальное исследование привело к формулировке гипотезы о «квантовании» уровней обучения искусственной адаптивной системы. Гипотеза такова: процесс обучения искусственных адаптивных систем происходит не непрерывным образом, а по некоторому конечному числу избранных траекторий или «туннелей» обучаемости с разными скоростями снижения ошибки. Существование таких траекторий позволяет системе быть более гибкой: избегать застревания в локальных минимумах и быстро (за несколько итераций) приходиться к верному решению задачи.

Вероятен переход между траекториями, и возможные причины наблюдаемого явления могут быть заложены в самом механизме процесса обучения.

Предложен фазовый портрет процесса обучения искусственной адаптивной системы на примере нейронной сети. Предположительно топологическая местность, по которой движется фазовая кривая нейронной сети в процессе обучения, содержит $(2n-1)$ особых точек, среди которых встречаются фокусы (неустойчивые, и один устойчивый, отвечающий за состояние решения задачи) и седла, где $n \in \mathbb{Z}$. Таким образом, фазовая траектория испытывает излом четное число раз, приводящий к смене правой части уравнения $X' = f(X, t)$, где X – ошибка обучения нейронной сети, а выявленные типы обучения, по сути, являются разными реализациями одного и того же процесса. Вероятно, этот механизм может быть общим для искусственных и естественных адаптивных систем, однако алгоритмически нейронные сети не могут демонстрировать увеличение ошибки обучения, в то время как живые системы склонны отвлекаться, рассредоточивать внимание, забывать, решать задачу неверным способом и т.п., т.е. могут увеличивать ошибку обучения в процессе.

Изучение данного явления может быть продуктивным при использовании методов, используемых в науках о нервной деятельности, чтобы определить те явления или системы, которые могут быть ответственны за процесс обучения у живых систем.

СПИСОК ЛИТЕРАТУРЫ

1. Гавриков В.Л., Хлебопрос Р.Г. Две динамические модели научения типа «кошка Торндайка». *Вестник Красноярского государственного педагогического университета им. В.П.Астафьева*. 2009. Т. 2. С. 47–55.
2. Lin C.S., Chang C.C., Chiu J.S., Lee Y.W., Lin J.A., Mok M.S., Chiu H.W., Li Y.C. Application of an Artificial Neural Network to Predict Postinduction Hypotension During General Anesthesia. *Medical Decision Making*. 2011. V. 31. № 2. P. 308–314.
3. Резникова Ж.И. Интеллект и язык животных и человека. *Основы когнитивной этологии: учебное пособие для вузов*. М.: ИКЦ «Академкнига», 2005. 518 с.
4. Фомин В.Н., Фрадков А.Л., Якубович В.А. *Адаптивное управление динамическими объектами*. М.: Наука, 1981.
5. Барцев С.И., Барцева О.Д. *Эвристические нейросетевые модели в биофизике: приложение к проблеме структурно-функционального соответствия: монография*. Красноярск: Изд-во Сибирского федерального университета, 2010. 115 с.
6. Барцев С.И., Охонин В.А. *Адаптивные сети обработки информации: препринт № 59Б*. Красноярск: ИФ СО АН СССР, 1986. 20 с.
7. Гавриков В.Л., Хлебопрос Р.Г. Континуальность типов научения: динамическое моделирование на основе теории катастроф. *Вестник Томского государственного университета*. 2010. Т. 331. С. 163–170.
8. Лоренц В.А., Гавриков В.Л., Хлебопрос Р.Г. Анализ обучения нейронной сети задачам, содержащим скрытую закономерность. *Вестник КрасГАУ*. 2012. Т. 5. С. 88–92.
9. Горбань А.Н. *Обучение нейронных сетей*. М.: ParaGraph, 1990. 160 с.
10. Носков М.В., Симонов К.В., Щемель А.Л. Нелинейная многопараметрическая регрессия данных наблюдений. Вопросы математического анализа. *ИЦП КГТУ*. 2003. Т. 7. С. 103–120.

Материал поступил в редакцию 22.10.2012, опубликован 07.02.2012.