

## ПРИМЕНЕНИЕ МЕТОДА КОНТРОЛИРУЕМОГО ВОЗМУЩЕНИЯ ДЛЯ МОДИФИКАЦИИ НЕЙРОКОНТРОЛЛЕРОВ В РЕАЛЬНОМ ВРЕМЕНИ

---

**Анотація.** Запропоновано новий метод навчання нейронних мереж керуванню динамічними об'єктами, який базується на оцінці локальних значень якобіану положення об'єкта по керуванню за рахунок використання малих збурень управління, що прикладається до об'єкта. Метод не вимагає апріорної інформації про об'єкт керування і може використовуватись для адаптивного керування об'єктами різної природи. Експерименти з використанням запропонованого методу для навчання рекурентних нейромереж у системах керування різними об'єктами демонструють переваги такого нейрокерування над стандартним PID-керуванням.

**Ключові слова:** нейрокерування, нейроконтролери, рекурентні нейронні мережі, адаптивне керування.

**Аннотация.** Предложен новый метод обучения нейронных сетей управлению динамическими объектами, базирующийся на оценке локальных значений якобиана положения объекта по управлению за счет использования малых возмущений прикладываемого управления. Метод не требует наличия априорной информации об управляемом объекте и может использоваться для адаптивного управления объектами различной природы. Эксперименты по использованию предложенного метода для обучения рекуррентных нейросетей в системах управления различными объектами демонстрируют преимущества такого нейроуправления над стандартным PID-управлением.

**Ключевые слова:** нейроуправление, нейроконтроллеры, рекуррентные нейронные сети, адаптивное управление.

**Abstract.** New method for training neural networks to control dynamic objects is proposed. The method is based on estimation of local values of Jacobian of object's position on control signal by introducing a small perturbation in applied control. The method doesn't require any a priori information about controlled object and may be applied for control of objects of different nature. Experiments on application of the proposed method for recurrent neural networks training in control systems of different objects illustrates advantages of such neurocontrol comparing to standard PID-control.

**Key words:** neurocontrol, neurocontrollers, recurrent neural networks, adaptive control.

### 1. Введение

Существующие подходы к нейроуправлению можно разделить на два направления: 1) имитация уже имеющейся системы управления (человека или сложного контроллера); 2) синтез управления на основе анализа модели поведения объекта. Системы первого типа получают требуемое управление в явном виде, и задача сводится к выбору типа и метода обучения нейронной сети. В системах второго типа оценивают значение производной от положения объекта по управлению, т.е. якобиан объекта, и эту информацию используют для обучения нейронной сети (прямого нейроконтроллера).

Для получения значений якобиана используются различные методы. Если известна точная математическая модель объекта управления, якобиан вычисляется аналитически. Обычно точная модель объекта неизвестна, поэтому используются приближенные методы оценки его значений. Один из наиболее известных методов состоит в обучении нейроэмулятора копированию поведения объекта и нахождению якобиана объекта путем обратного распространения единичного сигнала через нейроэмулятор [1–3]. Преимуществом этого метода является простота получения якобиана, а недостатком – трудоемкость обучения нейроэмулятора, а также непригодность метода в случаях, когда

характеристики объекта меняются со временем (при этом необходимо регулярное переобучение нейроэмулятора, что часто недопустимо).

В последние годы получил распространение метод адаптивной критики [4–6], основанный на вычислении значения производной от оценки стоимости по управлению. Этот метод близок к предыдущему, но не тождественен ему.

Известен также метод инверсного нейроэмулятора, в котором нейронную сеть вначале обучают оценивать значение приложенного управления, используя предыдущее и следующее положения управляемого объекта. В рабочем режиме на вход нейронной сети подают текущее положение объекта и целевое положение, а с ее выхода получают управление, которое переводит объект из текущего положения в целевое [7, 8]. В этом методе оценку производной от положения объекта по управлению выполняет нейронная сеть, которая действует как инверсный нейроэмулятор.

Таким образом, в различных методах нейроуправления явно или неявно решается задача нахождения значений якобиана объекта и качество ее решения в значительной степени определяет эффективность получаемого нейроуправления. Мы предлагаем новую схему, в которой для вычисления якобиана используются результаты воздействия на объект двух близких сигналов управления. Это позволяет обучать нейронную сеть управлению объектом без применения дополнительных моделей и дает возможность адаптировать управление к меняющимся характеристикам динамического объекта. Для демонстрации эффективности предлагаемой схемы проведено экспериментальное сравнение ее работы с PID-управлением на примере задачи стабилизации положения перевернутого маятника на движущейся тележке.

## 2. Метод контролируемого возмущения управления

Применение нейронной сети для управления некоторым объектом требует ее обучения на примерах успешного управления таким объектом. Однако во многих практически важных случаях такие примеры отсутствуют, в частности, при адаптивном управлении объектами, когда обучение должно производиться в режиме реального времени. В таких случаях для обучения нейронной сети приходится использовать текущие данные, получаемые путем сравнения траектории наблюдаемого поведения объекта с заданной целевой траекторией. Предлагаемый нами метод контролируемого возмущения управления состоит в том, чтобы последовательно отклоняя значение управления в разные стороны, сравнивать полученные реакции объекта, и, в зависимости от того, какое из отклонений больше соответствует переходу в целевое состояние, генерировать сигнал ошибки для обучения нейронной сети. Метод может быть применен к объектам с многомерным управлением, однако следует учитывать, что за один цикл работы метода вычисляется производная положения только

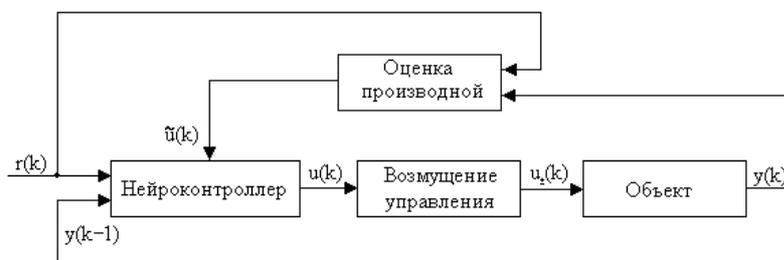


Рис. 1. Общая схема нейроуправления по методу контролируемого возмущения

по одной компоненте управления. Для вычисления производных по другим компонентам цикл необходимо повторять.

Схема нейроуправления по предлагаемому методу представлена на рис. 1.

Суть метода состоит в следующем: пусть  $\xi(k) = X_0$  – положение

управляемого объекта в  $k$ -й момент времени. Рассмотрим случай дискретного времени,

где каждый временной шаг намного меньше характерного времени изменения динамики объекта. Применим на следующем временном шаге управление  $U_1 = U(k) - h$ , где  $U(k)$  – текущее управление,  $h$  – вектор с одной ненулевой компонентой (по которой берется производная), величина которой достаточно мала. Обозначим полученное положение как  $X_1$ . На следующем временном шаге применим управление  $U_2 = U(k) + h$  и обозначим полученное положение как  $X_2$ .

Чтобы определить, какое из этих управлений больше соответствует достижению целевого положения, найдем оценку положения  $X_1'$ , в котором оказался бы объект, если бы в положении  $X_0$  к нему было приложено управление  $U_2$  вместо  $U_1$ :

$$X_1' = X_0 + (X_2 - X_1) + (\dot{X}_0 - \dot{X}_1)dt. \quad (1)$$

Здесь второе слагаемое представляет фактическое смещение системы при применении второго управления, а последний член учитывает изменение скорости объекта.

Производная от положения по одной компоненте управления может быть найдена как

$$\frac{\partial \xi(k)}{\partial u_i} \approx \frac{X_1' - X_1}{2|h|} = \frac{X_0 + X_2 - 2X_1 + (\dot{X}_0 - \dot{X}_1)dt}{2|h|}, \quad (2)$$

где индекс  $i$  соответствует номеру ненулевой компоненты вектора  $h$ , по которой берется частная производная в данный момент.

Для обучения нейронной сети следует выполнить следующие действия:

1. На вход нейронной сети подать значение текущего состояния объекта  $\xi(k)$  (координаты, скорости). Ответом сети является управление  $U(k)$ .

2. К управляемому объекту последовательно применять управление  $U(k) + h$  и  $U(k) - h$  с запоминанием состояний  $X_0$ ,  $X_1$  и  $X_2$ . Номер ненулевой компоненты  $h$  фиксируется в рамках одного цикла обучения. На каждом следующем цикле он может меняться, последовательно пробегая все значения.

3. Используя формулу (2), вычислить производную  $\frac{\partial \xi(k)}{\partial u_i}$ .

4. Выполнить один шаг обучения нейронной сети, входом которой служит текущее состояние объекта  $\xi(k)$ , а целевым – вектор  $\tilde{U}(k) = U(k) + \alpha \cdot \text{sign}\left(\frac{\partial \xi}{\partial u_i}\right) \cdot \text{sign}[r(k) - \xi(k)]$ , полученный из  $U(k)$  изменением  $i$ -й компоненты на малую величину  $\alpha$  в сторону наибольшего приближения положения объекта  $\xi(k)$  к целевому положению  $r(k)$ .

Величина  $h$  выбирается достаточно малой, чтобы нелинейность управляемого объекта слабо проявлялась на разнице между управлениями  $U(k) + h$  и  $U(k) - h$ , однако эта разница должна быть достаточной, чтобы эффект управления существенно превышал уровень флуктуаций, вызванных шумом.

Описанный в пп. 1–4 процесс представляет собой полный цикл обучения нейронной сети. Повторив его достаточное количество раз, можно научить сеть поддерживать заданное положение управляемого объекта.

### 3. Структура нейронной сети

В системах нейрорегулирования обычно используются рекуррентные нейронные сети, обучаемые по методу обратного распространения ошибки во времени. Однако, как было

показано в [17], метод рекуррентного обучения в реальном времени (RTRL, real-time recurrent learning) дает примерно такие же результаты при меньшей вычислительной сложности. Поэтому для реализации метода контролируемого возмущения мы использовали рекуррентный перцептрон, к которому применялся метод обучения RTRL. Архитектура нейронной сети изображена на рис. 2. Сеть включает один скрытый слой нейронов, слой рецепторов и один или два (соответственно размерности управления) эффекторных нейрона. Использовались нейроны с сигмоидной активационной функцией. На входы скрытого слоя нейронов поступали текущие и задержанные на несколько тактов значения положения и скорости объекта, уставки, а также задержанные значения выхода эффекторных нейронов. Для выбора приемлемого варианта конфигурации сети были выполнены предварительные эксперименты, в которых было установлено, что ввод линии задержек входного сигнала существенно улучшает качество управления, если задержки составляют не менее двух тактов, но дальнейшее увеличение числа задержек мало влияет на управление. Задержки выходного сигнала мало влияют на качество обучения, если присутствует хотя бы одна. Наилучшее соотношение между качеством управления и сложностью сети получено при задержке выхода на 3 такта. В экспериментах с изменением числа нейронов скрытого слоя от 3 до 20 не было выявлено четкого оптимума, поэтому в дальнейших экспериментах число скрытых нейронов было принято 8 для задач одномерного управления и 16 – для двумерного.

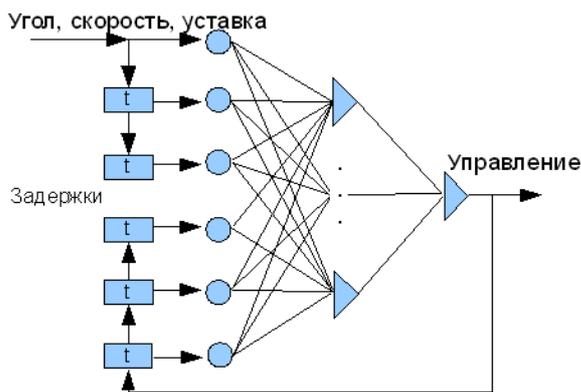


Рис. 2. Структура рекуррентной сети для одномерного случая

Итоговая структура нейронных сетей была следующей: на вход подавалось значение текущего угла, угловой скорости, уставки по углу. Эти данные поступали также в линию задержек на 2 такта, выход сети попадал на вход через линию задержек в три такта. Было использовано две сети: для одномерного и двумерного случаев. Для одномерного случая сеть имела структуру 12 входов, 8 скрытых нейронов, 1 нейрон на выходе. Для двумерного случая – 24 входа, 16 скрытых нейронов, 2 нейрона на выходе.

#### 4. Схема эксперимента

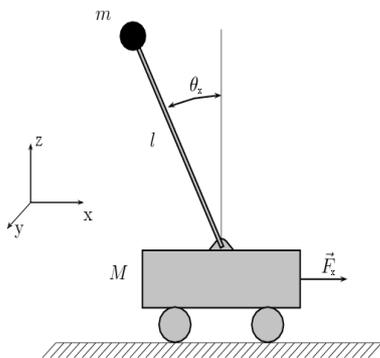


Рис. 3. Схема перевернутого маятника на тележке

Экспериментальное исследование метода контролируемого возмущения проводилось с использованием двумерного перевернутого маятника на подвижной платформе, который является классической моделью нелинейного динамического объекта [11]. Нелинейность четко проявляется при отклонении маятника на угол свыше  $12^\circ$  от вертикальной оси.

Схема перевернутого маятника показана на рис. 3. Тележка массы  $M$  может горизонтально перемещаться по двум осям ( $X$ ,  $Y$ ). На тележке закреплен маятник массы  $m$  длины  $l$ , образующий углы  $\theta_x, \theta_y$  с вертикальной осью. В случае, когда

$\theta_x = \theta_y = 0$ , система находится в состоянии равновесия. По условию задачи, на систему

действуют внешние силы, которые дестабилизируют ее, вынуждая маятник отклоняться от состояния равновесия. Если маятник отклоняется слишком далеко от положения равновесия, он падает. Тележка оснащена двигателем, который прикладывает к тележке силы  $F_x, F_y$ , которые перемещают ее вместе с основанием маятника в плоскости XY.

Задача контроллера состоит в поддержании маятника в вертикальном положении в условиях внешних сил (шума) как можно более долгое время, получая на вход значение текущих углов отклонения  $\theta_x, \theta_y$  и генерируя значение управляющих воздействий  $F_x, F_y$ .

Более качественной считается та система управления, которая может более длительное время удерживать маятник от падения при воздействии внешнего шума.

Были проведены две серии экспериментов. В первой серии один из углов маятника был зафиксирован ( $\theta_y = 0$ ) и управление проводилось только по оси X. Положение тележки не ограничивалось. Во второй серии ни один из углов не фиксировался, положение тележки также не ограничивалось. Задачей контроллера было поддержание как равновесия, так и заданного угла наклона маятника (последнее требовало неограниченного

движения тележки с ускорением).

Схема включения рекуррентной нейронной сети в систему управления показана на рис. 4. На вход сети поступает информация об углах по обеим осям, на выходе генерируются сигналы управления для каждой оси. При выполнении экспериментов с одномерным маятником вместо двух координат на вход поступала одна и управление генерировалось только для одной оси.

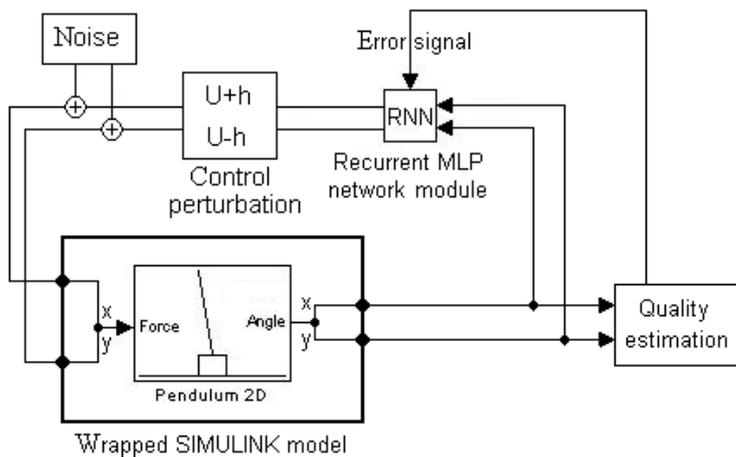


Рис. 4. Управление перевернутым маятником с помощью рекуррентной нейронной сети

## 5. Экспериментальное исследование качества управления

Для оценки качества управления использовалось время в секундах, которое проходит от установки маятника в положение равновесия до его отклонения от вертикали на угол свыше 55 градусов при заданном уровне шума. Такой критерий часто используется для оценки управления в ситуациях, когда маятник достаточно далек от положения равновесия, и нелинейные эффекты вносят ощутимый вклад [12, 13].

В целом эксперимент был построен следующим образом: модель маятника, импортированная из среды Simulink (составная часть пакета MATLAB), приводилась в положение равновесия, потом последовательно к ней применялся один шаг случайного управления с заданной максимальной амплитудой (шум), один шаг управления с помощью контроллера – рекуррентной сети, или PID-контроллера, в зависимости от режима эксперимента, потом снова шум, снова управление и так далее.

Шум представлял собой случайный импульс управления, равномерно распределенный в интервале  $[-n \cdot U_{\max}, n \cdot U_{\max}]$ , где  $n$  – уровень шума (в наших экспериментах принимал значения от 0 до 1),  $U_{\max}$  – управление, которое применяется к системе, когда на выходном нейроне нейроконтроллера будет единица.

Шум и управление последовательно применялись до тех пор, пока отклонение маятника не превышало заданный порог. Время, которое прошло от момента установки маятника в положение равновесия до момента, когда отклонение превышало порог (время удержания маятника), интерпретировалось как качество управления, маятник устанавливался в положение равновесия, и процесс повторялся снова.

Эксперименты показали, что время удержания при управлении рекуррентной сетью меняется немонотонно в процессе обучения. Поэтому использовался следующий порядок выполнения экспериментов:

1. Цикл обучения повторялся 3000 раз.
2. Обучение временно прекращалось, и нейронная сеть использовалась в режиме управления при фиксированном уровне шума ( $n = 0,65$ ) до тех пор, пока угол маятника не превысит граничный угол ( $55^\circ$ ). Эксперимент повторялся 200 раз, вычислялось среднее значение времени удержания маятника.

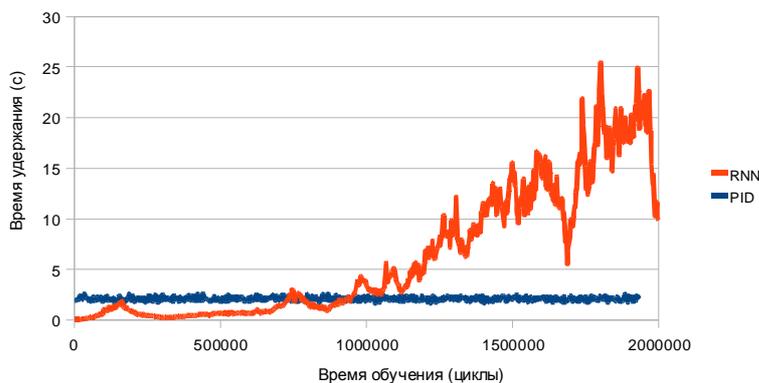


Рис. 5. Зависимость времени удержания двумерного маятника от времени обучения рекуррентной сети. Для сравнения приведено время удержания PID-контроллером

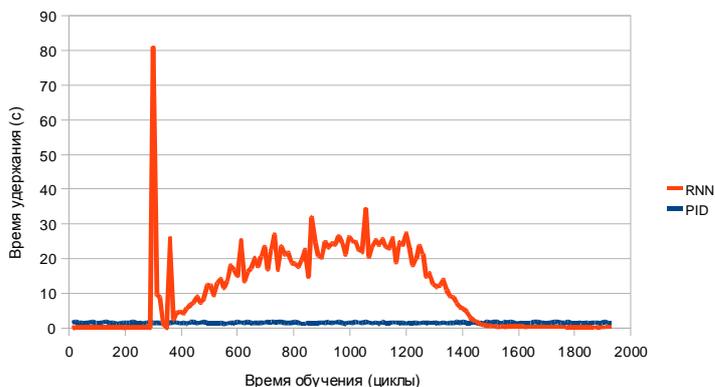


Рис. 6. Зависимость времени удержания одномерного маятника от времени обучения рекуррентной сети при увеличенной скорости обучения. Для сравнения приведено время удержания PID-контроллером

представлены результаты при значительно большем коэффициенте скорости обучения.

Тем не менее, максимальный результат при обучении в таком режиме также на порядок превосходит PID-управление.

Траектория отклонений маятника, которые он проходит при удержании его с помощью рекуррентной сети и PID-контроллера при фиксированном уровне шума, отображена на рис. 7. Видно, что отклонение маятника от целевого положения (нуля) при

3. Полученное значение отображалось в виде одной точки на графике, пп. 1 и 2 повторялись.

Результаты для управления двумерным маятником изображены на рис. 5, коэффициент скорости обучения составляет 0,0001. Из графика следует, что достигаемое рекуррентной сетью максимальное время удержания маятника на порядок превосходит время удержания с помощью PID-контроллера при том же уровне шума. Однако для достижения такого превосходства нейронная сеть требует длительного обучения. Выполненные нами эксперименты по ускорению обучения нейронной сети путем увеличения коэффициента скорости модификации ее весовых коэффициентов до 0,01 показали, что вместе с ускорением возрастает нестабильность обучения, ведущая к падению качества управления. Это видно на рис. 6, где

управлении с помощью PID-контроллера существенно больше, чем при управлении рекуррентной сетью.

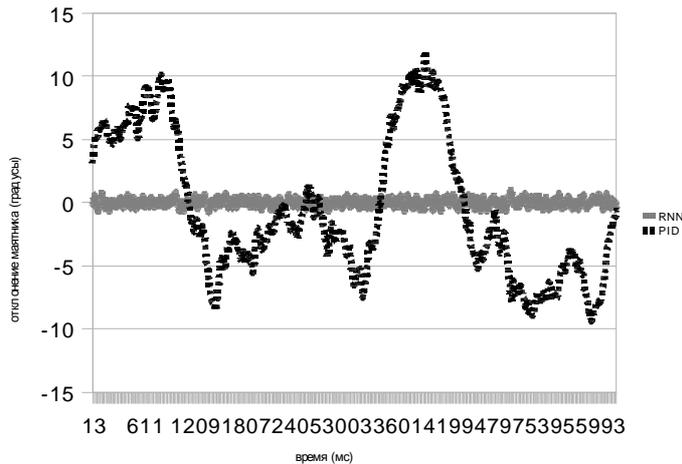


Рис. 7. Траектория отклонений одномерного маятника при большом ( $n = 0,35$ ) уровне шума. Сплошным показано управление рекуррентной сетью, пунктир – PID-контроллер

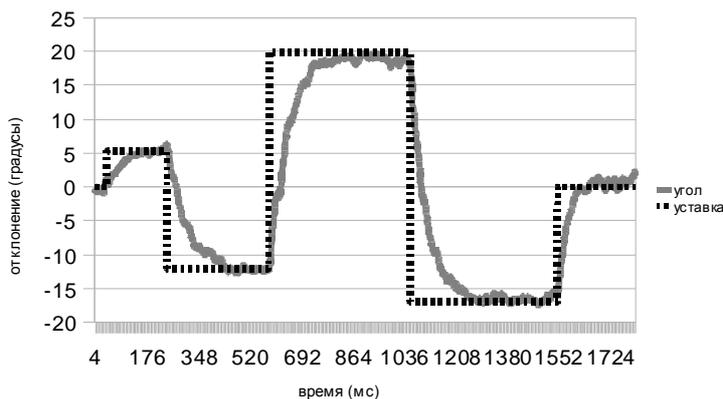


Рис. 8. Изменение угла отклонения маятника при изменении уставки при высоком ( $n = 0,3$ ) уровне шума, управление рекуррентной нейронной сетью

с лучшими результатами соревнований по управлению этой моделью, доступными на сайте лаборатории [14]. Использовалась такая же рекуррентная сеть, как и для управления перевернутым маятником. Отличие состояло в увеличении числа задержек входных значений (7 вместо 3) и времени задержек (в пределах от 1 до 50 тактов), а также в увеличении числа нейронов скрытого слоя (10 нейронов вместо 8). Сравнение результатов управления оптимизированным PID-контроллером и рекуррентной сетью, обученной по методу управляемого возмущения управления, приведено на рис. 9.

## 6. Выводы

Предложенный метод контролируемого возмущения управления позволяет получить оценку локального значения якобиана объекта по управлению, которая может быть использована для обучения рекуррентной нейронной сети управлению этим объектом. Полученный в результате нейроконтроллер обладает хорошими характеристиками, существенно превосходящими управление с помощью PID-контроллера на различных

Также рекуррентная сеть позволяет поддерживать заданный угол наклона маятника (что требует непрерывного ускорения тележки). График следования угла отклонения маятника в уставке показан на рис. 8. Использованный нами PID-контроллер в принципе не мог обеспечить достижение такой уставки.

Помимо этого, начата серия экспериментов на модели вертолета из виртуальной лаборатории Automatic Control Telelab [14]. На первом этапе исследования проводятся на математической модели этого вертолета, разработанной компанией Humusoft. Модель представляет собой закрепленный на стенде вертолет, оснащенный двумя винтами. В проведенных экспериментах производилось управление только углом азимута (т.к. только для этого режима доступны данные для сравнения с другими методами). На данный момент при управлении математической моделью вертолета нейронная сеть, обученная по методу контролируемого возмущения управления, дает результаты, сравнимые

задачах. Следует учитывать, что процесс обучения нейронной сети неустойчив, и для

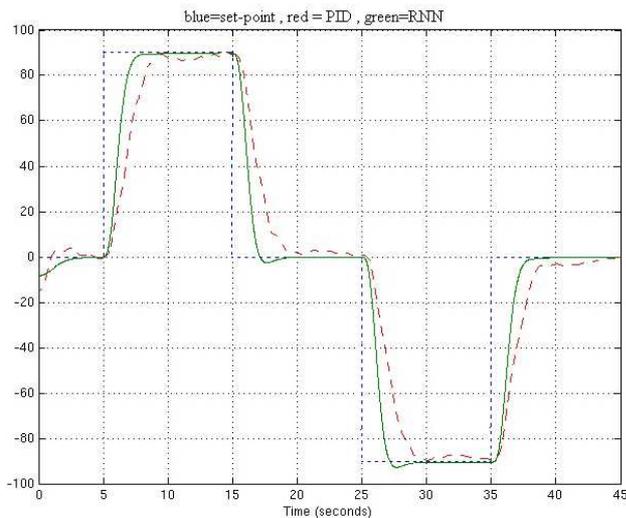


Рис. 9. Управление PID-контроллером физической моделью вертолета (пунктирная линия) и рекуррентной нейронной сетью математической моделью вертолета (сплошная линия). Коротким пунктиром нарисована уставка

достижения высокого качества управления необходимо постоянно тестировать сеть в процессе обучения.

Безусловным преимуществом метода является возможность быстро обучаться управлению объектами, о которых нет никакой априорной информации, без длительного процесса построения модели объекта. Для объектов, реакция которых на управление происходит с некоторой задержкой, применение этого метода потребует либо внесения соответствующей задержки в явном виде в схему вычисления якобиана, либо увеличения шага времени таким образом, чтобы воздействие на предыдущем шаге уже отражалось на положении объекта на следующем.

Начатые эксперименты по управлению моделью вертолета

подтверждают эффективность предложенного метода и доказывают возможность его применения без какого-либо дополнительного анализа объекта управления. Тем не менее, для окончательных выводов о качестве получаемых контроллеров необходимо завершить эту серию экспериментов, а также исследовать поведение контроллера при динамическом изменении характеристик объекта управления.

## СПИСОК ЛИТЕРАТУРЫ

1. Jordan M.I. Forward models: Supervised learning with a distal teacher / M.I. Jordan, D.E. Rumelhart // *Cognitive Science*. – 1990. – Vol. 16. – P. 313 – 355.
2. Narendra K.S. Identification and control of dynamical systems using neural networks / K.S. Narendra, K. Parthasarathy // *IEEE Transactions on Neural Networks*. – 1990. – Vol. 1. – P. 4 – 27.
3. Werbos P. Backpropagation through time: what it does and how to do it / P. Werbos // *Proc. IEEE*. – 1990. – Vol. 78, N 10. – P. 1550 – 1560.
4. Prokhorov D. Adaptive critic designs / D. Prokhorov, D. Wunsch // *IEEE Transactions on Neural Networks*. – 1997. – N 8(5). – P. 997 – 1007.
5. Lendaris G.G. A Retrospective on Adaptive Dynamic Programming for Control / G.G. Lendaris // *Proc. of IJCNN*, (Atlanta, Georgia, USA, June 14–19 2009). – Atlanta, Georgia, USA, 2009. – P. 1750 – 1757.
6. Barto A.G. Reinforcement learning and adaptive critic methods / A.G. Barto // *Handbook of Intelligent Control* / D.A. White, D.A. Sofge [eds.]. – New York: Van Nostrand Reinhold, 1992. – P. 469 – 491.
7. Омату С. Нейроуправление и его приложения / С. Омату, М. Халид, Р. Юсоф; пер. с англ. – М.: ИПРЖР, 2000 – 272 с.
8. Ronco E. Incremental Polynomial Controller Networks: Two Self-Organising Non-Linear Controllers / E. Ronco // *Ph.D. Dissertation Thesis*. – Glasgow, 1997. – 207 p.
9. Wei W. Recurrent neuro-controller for an inverted pendulum using evolution strategy / W. Wei, W. von Seelen // *International Journal of Systems Science*. – 2001. – Vol. 32, N 5. – P. 643 – 650.
10. Reznik A.M. Dynamic Associative Memory Based on Open Recurrent Neural Network / A.M. Reznik, D.A. Dziuba // *Proc. of IJCNN'09*, (Atlanta, Georgia, USA, June 14–19, 2009). – Atlanta, Georgia, USA, 2009. – P. 2657 – 2663.
11. Slotine J.-J. E. *Nonlinear Control Systems Design* / J.-J. E. Slotine, W. Li. – Englewood Cliffs: Prentice-Hall, 1991. – P. 193.

12. Стюарт Р. Искусственный интеллект: современный подход / Р. Стюарт, П. Норвиг; пер. с англ. – М.: Издательский дом "Вильямс", 2006. – [2-е изд.]. – 1408 с.
13. Michie D. BOXES: An experiment in adaptive control / D. Michie, R.A. Chambers // Machine Intelligence. – 1968. – N 2. – P. 125 – 133.
14. Casini M. The Automatic Control Telelab. A Web-based technology for Distance Learning / M. Casini, D. Prattichizzo, A. Vicino // IEEE Control Systems Magazine. – 2004. – N 24 (3). – P. 36 – 44.
15. Різник А.М. Динамічні рекурентні нейронні мережі / А.М. Різник // Математичні машини і системи. – 2009. – № 2. – С. 3 – 26.
16. Різник А.М. Динамічна автоасоціативна пам'ять, заснована на відкритій рекурентній нейронній мережі / А.М. Різник, Д.О. Дзюба // Математичні машини і системи. – 2010. – № 2. – С. 45 – 51.
17. Orlando De Jesus Backpropagation Algorithms for a Broad Class of Dynamic Networks / Orlando De Jesus, Martin T. Hagan // IEEE Transactions on Neural Networks. – 2007. – Vol. 18, N 1. – P. 14 – 27.

*Стаття надійшла до редакції 14.01.2011*