

УДК 681.513.7; 681.5.013

**Д.А. Дзюба**

**Институт проблем математических машин  
и систем НАН Украины**

# МОДИФИКАЦИЯ МЕТОДА КОНТРОЛИРУЕМОГО ВОЗМУЩЕНИЯ УПРАВЛЕНИЯ

**Ключевые слова:** нейроруправление, рекуррентный перцептрон, динамические объекты

## Введение

Задача управления сложными динамическими объектами интересна в первую очередь тем, что во многих практически важных случаях крайне проблематично выяснить, какое управление будет оптимальным в том или ином смысле, а порой даже неизвестно, какое управление вообще обеспечит сколько-нибудь приемлемое поведение объекта управления.

Очевидно, что нестабильная природа управляемого объекта требует адаптивной системы управления, которая может приспособиться к особенностям данного объекта в текущем режиме, а при смене свойств объекта, или изменения режима управления – достаточно быстро изменить свое поведение.

Нейронные сети обладают достаточно сильными адаптивными свойствами, что делает естественным их применение в задачах такого типа. Выбор метода обучения нейронной сети является, крайне важной проблемой – ведь при изменении свойств объекта мы заранее не знаем, каким должно быть новое управление, метод обучения должен позволять тем или иным путем его определить, и обучить ему нейронную сеть.

Для решения этой задачи необходимо проанализировать поведение объекта, выяснить, как он будет реагировать на разные сигналы управления в данном состоянии, и, вообще говоря, было бы желательно построить якобиан положений объекта по управляющим сигналам.

В схеме нейроруправления, предложенной независимо К.С. Нарендрой и К. Пасарати [1],

*Предложено развитие метода контролируемого возмущения управления, позволяющее полностью автоматизировать процесс обучения нейроконтроллера управлению динамическим объектом без предварительного построения его модели. Проведены эксперименты, подтверждающие эффективность метода.*

*Запропонований розвиток метода контрольованого збурення керування, який дозволяє повністю автоматизувати процес навчання нейроконтроллера керуванню динамічними об'єктами без попередньої побудови його моделі. Проведені експерименти, що підтверджують ефективність метода.*

*Further development of control perturbation method is proposed. It allows to fully automate a neurocontroller training process and makes possible to control dynamic objects without their prior model development. Experiments that confirm efficiency of the method are described.*

Б. Вербосом [2] и Джорданом и Румельхартом [3], и получившей впоследствии значительную популярность, для получения ошибки нейроконтроллера вместо непосредственного вычисления якобиана объекта управления используется механизм обратного распространения ошибки через прямую нейросетевую модель объекта управления, предварительно обученного точно воспроизводить динамику его поведения. Также в работе [1], было указано, что для эмуляции поведения динамического объекта наилучшей моделью нейронных сетей являются рекуррентные многослойные перцептроны, позже Х. Сигельман и др. доказали это строго [4]. Однако при динамическом изменении характеристик объекта, должна меняться и его модель, а это в свою очередь сложно обеспечить, т.к. для успешного обучения нейроэмулятора нужно накопить некоторую статистику поведений объекта в различных ситуациях. При попытке следовать этому подходу, модель объекта будет «отставать» от его реальных характеристик, а, следовательно, управление не будет в полной мере соответствовать динамике объекта.

В рамках развития подхода с использованием нейроэмулятора были предложены методы нейросетевого прогнозирующего управления [5] и системы адаптивных критиков [6]. В схеме нейросетевого прогнозирующего управления также производится обучение нейроэмулятора прямой динамике объекта, который затем используется для предсказания поведения объекта на несколько шагов вперед. Работа

систем адаптивных критиков основана на применении принципа Беллмана для выбора контроллером управляющего действия, оптимального в некотором заданном смысле. В них выделяется отдельный нейросетевой модуль на основе многослойного перцептрона, называемый «критик», который обучается по наблюдению за реальным поведением системы возвращать оценку прогнозируемой ошибки управления на некоторое количество шагов вперед, в зависимости от поданного на вход управляющего воздействия.

Впрочем, в этих подходах возникает аналогичная проблема, при быстром изменении характеристик объекта его модель устаревает, и такие подходы не дают возможности оптимизировать управление.

Одним из возможных путей решения является определение характеристик объекта без построения промежуточной модели, а за счет применения «пробных» управляющих сигналов непосредственно к объекту, и анализа его реакций на эти сигналы.

#### Метод контролируемого возмущения управления

Применение нейронной сети для управления некоторым объектом требует ее обучения на примерах успешного управления таким объектом. В то же время, 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.

Одним из способов анализа поведения объекта без предварительного построения его модели является использование пробных шагов управления и анализ полученных состояний объекта. В частности можно применить к объекту последовательно два сигнала управления, отличающихся на небольшую величину. Если при этом сделать существенное, но справедливое для многих практических случаев предположение, что управляющие сигналы определяют некоторые силы, прикладываемые к объекту, то можно сделать некоторые выводы, которые в конечном итоге дадут достаточно информации для обучения нейроконтроллера.

Для начала, найдем производную от положения объекта по прикладываемому управлению в окрестности текущего состояния объекта и окрестности приложенного управления

(объект в общем случае нелинеен, хотя в рамках данного метода необходимо, чтобы этой нелинейностью можно было пренебречь при рассмотрении разности между двумя пробными сигналами – отсюда можно получить ограничения на величину разности пробных сигналов для каждого конкретного объекта управления). Если бы мы могли применить один пробный сигнал, вернуть объект в исходное состояние, и применить второй пробный сигнал, то задача была бы тривиальной. К сожалению, это невозможно. Тем не менее, мы можем оценить, каким было бы состояние объекта, если бы в начальном состоянии мы применили второй пробный сигнал вместо первого. Для получения этой оценки воспользуемся следующей процедурой. Пусть  $X_0$  – положение объекта в некоторый момент времени. Применим к нему последовательно два сигнала управления (предположим, что положение и управление одномерно – обобщение на многомерный случай выполняется путем последовательного применения алгоритма для одномерного случая к различным каналам управления, более детально оно было рассмотрено в [7]):  $U - h$  и  $U + h$ . Положение после применения первого управления обозначим как  $X_1$ , после второго – как  $X_2$ . Тогда можно оценить положение  $X_1'$ , в котором оказался бы объект, если бы к нему было применено второе управление вместо первого, как:

$$X_1' = X_0 + (X_2 - X_1) + (X_0' - X_1')dt \quad (1)$$

$$\frac{\partial X}{\partial U} = \frac{X_1' - X_1}{2|h|} \quad (2)$$

где  $X_0$  – скорость до применения управления,  $X_1$  – скорость после применения первого управляющего сигнала.

Соответственно, производная от положения объекта по управлению записывается как

Обучение нейроконтроллера с использованием этой производной выглядит следующим образом:

- Нейроконтроллер получает на вход положение и скорость объекта, генерирует некоторое управление  $U$  на выходе;
- Применяются последовательно два управления, измененные на небольшую величину  $\pm h$  относительно выданного контроллером;
- С помощью производной (2) производится оценка, в какую сторону нужно сместить управление, чтобы лучше достичь целевого со-

стояния;

- Нейронная сеть контроллера обучается генерировать ответ  $U + T$  при текущих входах, где  $T$  – небольшая величина, смещающая управление в сторону лучшего достижения це-

левого состояния.

Схематически процесс обучения нейроконтроллера при использовании этого метода изображен на рис. 1.



Рис. 1. Общая схема метода контролируемого возмущения управления.

### Модификация метода контролируемого возмущения управления

В описанном методе пункт 3 неявно содержит достаточно нетривиальный вопрос, как все-таки определить, какое из управлений лучше в каком-либо смысле, например, в какой мере то или иное управление соответствует достижению целевого состояния. В обычном методе контролируемого возмущения управления этот вопрос решался сравнением  $X_1$  и  $X_1'$ , положение, более близкое к целевому, считалось лучшим, и соответственно выбира-

лось направление коррекции. Однако не всегда подобный выбор является корректным (например, если положение уже близко к целевому, а скорость далека от целевой, то нужно приводить в соответствие скорость, а не положение), и только аккуратный подбор коэффициентов скорости и режимов обучения позволял получить действительно качественный контроллер в итоге. В рамках модификации метода, этот критерий формализован в следующем общем виде:

$$Q(U) = \alpha(X_1 - X_0)(X_t - X_1) + \beta(X_1' - X_0')(X_t' - X_1')dt \quad (3)$$

где  $Q(U)$  – количественный показатель степени достижения цели при управлении  $U$  в данный момент,  $X_0$  – положение объекта до применения управления,  $X_1$  – положение объекта после применения управления,  $X_t$  – целевое положение объекта,  $X_0'$  – скорость до применения управления,  $X_1'$  – скорость после применения управления,  $X_t'$  – целевая скорость,  $\alpha, \beta$  – некоторые члены, которые будут определены ниже.

Такой вид выбран из соображений достижения одновременно целевого положения и



скорости, при этом  $\alpha$  и  $\beta$  не должны быть фиксированными коэффициентами, они должны меняться в зависимости от того, более важно в данном состоянии быстрее приближаться к цели, или приводить скорость к целевому значению. Для этого можно сравнить время, которое уйдет на приведение скорости к целевому значению при максимальном допустимом управляющем сигнале, с расстоянием, которое пройдет объект за это время при данной скорости и таком же управлении. Если полученное расстояние больше разности текущего и целевого положений, то преимущественный вклад должен вносить первый член, если меньше – то второй. Точно вычислить эти члены для нелинейного объекта с заранее неизвестным поведе-

нием невозможно, однако эксперименты показали, что даже если в этом моменте воспользоваться линейным приближением, процесс обу-

$$\gamma = (\dot{X}_t - \dot{X}_1)^2 / 2\ddot{X}_{max} |X_t - X_1| - |\dot{X}_t - \dot{X}_1| / \ddot{X}_{max} \quad (4)$$

где  $\ddot{X}_{max}$  – ускорение, которое получает объект при максимальном допустимом сигнале управления (в линейном приближении всюду одинаковое), остальные обозначения совпадают с обозначениями для формулы (3). Тогда

$$\begin{aligned} \alpha &= e^\gamma / (e^\gamma + 1) \\ \beta &= e^{-\gamma} / (e^{-\gamma} + 1) \end{aligned} \quad (5)$$

что завершает определение формулы (3). Соответственно, возвращаясь к методу обучения контроллера, если , то нейронной сети в качестве целевого значения для обучения в данной точке поступает  $U + h_T$ , где  $h_T$  – некоторая малая величина, значение которой определяется эмпирически, а если , то целевым значением будет  $U - h_T$ .

Упрощенная версия предложенного метода уже исследовалась ранее на задаче управления моделью перевернутого маятника [7, 8], и продемонстрировала там значительное превосходство над PID-контроллерами. Внесенные в метод коррективы ускорили процесс обучения, приблизив его к монотонности.

В данной работе исследовалось применение этого метода для обучения нейронной сети управлению стендовой моделью вертолета из

чения приводит нас к достаточно хорошему управлению. В случае линейного приближения, это соотношение может быть выписано как

виртуальной лаборатории Automatic Control Telelab [9]. Ввиду технических ограничений, исследования проводились на математической модели этого вертолета, разработанной компанией Numusoft. Модель представляет собой закрепленный на стенде вертолет, оснащенный двумя винтами. В проведенных экспериментах производилось управление углом азимута.

### Экспериментальные исследования

На рис. 2 показано сравнение двух идентичных по структуре нейроконтроллеров, один из которых обучен по упрощенному методу, другой – по изложенному в данной статье. Следует отметить, что контроллер, обученный по упрощенному методу, был получен в результате длительной процедуры подбора оптимальных параметров обучения, в то время как обученный по модифицированному методу был получен полностью автоматически, без какой-либо «ручной» оптимизации. При этом качество управления при использовании модифицированного метода хоть и не критически, но ощутимо превосходит качество управления, достигнутое при упрощенном методе: полностью отсутствуют эффекты переправления, а целевое положение достигается несколько быстрее.

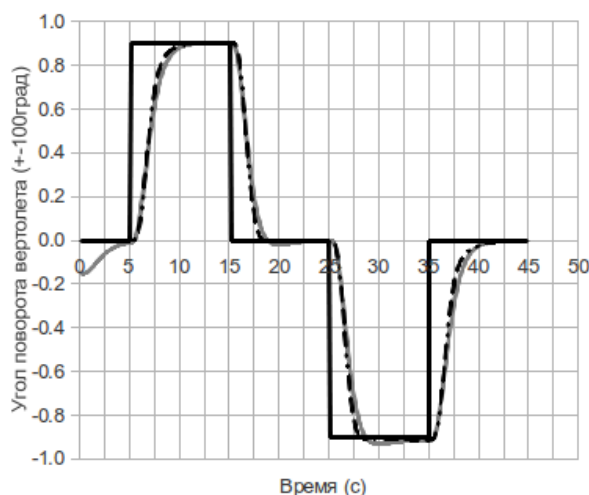


Рис. 2. Сравнение нейроконтроллеров, обученных по обычному и модифицированному методам.

Штрих-пунктирная линия – модифицированный метод обучения,  
сплошная серая – упрощенный,  
сплошная черная – уставка.

Более интересные результаты получаются при сравнении обычного и модифицированного метода при добавлении внешних возмущений, «толкающих» модель в случайные стороны со случайной амплитудой, так, что в отсутствие

управления модель отклонилась бы от целевого положения более чем на 10 градусов за одну секунду. Графически результаты такого эксперимента отображены на рис. 3.

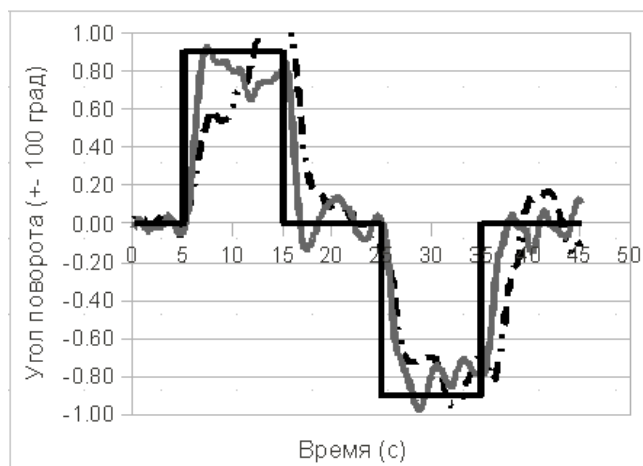


Рис. 3. Сравнение нейроконтроллеров, обученных по модифицированному (серая линия) и обычному (пунктирная линия) методу при непрерывных существенных внешних возмущениях.

Видно, что контроллер, полученный при использовании модифицированного метода, удерживает модель в окрестности целевого положения существенно лучше. Нейроконтроллер, обученный по модифицированному правилу не только более стабильно удерживает модель в окрестности целевого положения, но также существенно быстрее достигает такого положения. Тем не менее, графики приведены

для одной конкретной реализации шума, и их одних недостаточно для того, чтобы сделать окончательные выводы. Была проведена серия из 100 экспериментов такого типа, с различными реализациями шума, и вычислено среднее от суммарного среднеквадратического отклонения реального положения модели от целевого положения. Результаты приведены в табл.

Таблица

Суммарное среднеквадратическое отклонение реального положения модели от целевого для разных методов обучения и разных условий

	Без возмущений	С возмущениями
Обычный	26.15	56.44
Модифицированный	24.78	29.68

Отсюда видно, что хотя для случая управления моделью без внешних возмущений выигрыш нейроконтроллера, полученного по модифицированному методу, достаточно несущественный — но при переходе к более сложной ситуации, когда модель непрерывно подвергается внешним воздействиям, преимущество контроллера с модифицированным методом обучения становится весьма значимым.

Впрочем, следует принимать во внимание тот факт, что в немодифицированном методе параметры обучения оптимизировались вруч-

ную для достижения оптимального управления, и оптимизация проводилась для ситуации управления моделью без внешних возмущений. Возможно, если бы ручная оптимизация проводилась для случая управления моделью в условиях существенных внешних воздействий, преимущество нейроконтроллера, обученного по модифицированному методу, оказалось бы не столь значимым.

Тем не менее, можно утверждать, что модифицированный метод обучения дает на выходе такие нейроконтроллеры, которые лучше при-

способлены к различным внешним условиям, чем нейроконтроллеры, полученные при ручной оптимизации параметров в рамках немодифицированного метода.

#### Выводы

Предложенная модификация метода контролируемого возмущения управления позволяет перевести процесс обучения в полностью автоматический режим, без необходимости ручного подбора параметров, сохранив и даже несколько повысив качество управления, которое обеспечит полученный в результате контроллер. В такой модификации обучение происходит чрезвычайно быстро, например контроллер, результаты работы которого отображены на рис. 2, был получен всего лишь за 11 проходов целевой траектории, т.е. в случае обучения на реальном объекте это заняло бы менее 500 секунд. При этом не требуется никакая априорная информация об объекте управления, а качество управления, достигаемое таким контроллером, находится на уровне лучших результатов международных соревнований по управлению данным объектом (где допускается любое предварительное исследование, как самого объекта, так и его модели).

#### Список литературы:

1. K.S.Narendra, K.Parthasarathy Identification and control of dynamical systems using neural networks. In: IEEE Transactions on Neural Networks, 1,

1990, p. 4-27.

2. P.Werbos, Back propagation through time: what it does and how to do it. In: Proc. IEEE, Vol. 78, No. 10, October 1990

3. M.I.Jordan and D.E.Rumelhart. Forward models: Supervised learning with a distal teacher. In: Cognitive Science, Vol. 16, pp.313- 355, 1990.

4. H.T.Siegelmann, B.G.Horne, and C.L.Giles. Computational capabilities of recurrent NARX neural networks. In: IEEE Trans. Systems, MAN, and Cybernetics - Part. B: Cybernetics, 1997, 27(2): 208-215.

5. M.T.Hagan, H.B.Demuth. Neural Networks for Control. In: Proceedings of the 1999 American Control Conference, San Diego, CA, 1999, pp. 1642-1656.

6. D.Prokhorov and D.Wunsch. Adaptive critic designs. In: IEEE Transactions on Neural Networks, 8(5), 1997, p. 997-1007.

7. Д.А.Дзюба, А.Н.Чернодуб. Применение метода контролируемого возмущения для модификации нейроконтроллеров в реальном времени. Математические Машины и Системы, 2011. № 1. с. 20 - 28.

8. Д.А.Дзюба, А.Н.Чернодуб. Обучение рекуррентной нейронной сети методом контролируемого возмущения для управления динамическими объектами. Proceedings of the 2010 Knowledge - Dialogue - Solution Conference, Sep. 06 - 10, Kyiv (Ukraine)

9. M.Casini, D.Prattichizzo, A.Vicino. The Automatic Control Telelab. A Web-based technology for Distance Learning // IEEE Control Systems Magazine. 2004. N 24 (3). - P. 36 - 44

#### Сведения об авторе



**Дзюба Дмитрий Александрович** - м.н.с. отдела нейротехнологий в Институте проблем математических машин и систем НАН Украины. Основные области научных интересов - рекуррентные и аттракторные нейронные сети, проблемы нейроуправления, компьютерное зрение

E-mail: [ddziuba@immsp.kiev.ua](mailto:ddziuba@immsp.kiev.ua)

Статья поступила в редакцию 12.04.2011 г.