ОБУЧЕНИЕ РЕКУРРЕНТНОЙ НЕЙРОННОЙ СЕТИ МЕТОДОМ КОНТРОЛИРУЕМОГО ВОЗМУЩЕНИЯ ДЛЯ УПРАВЛЕНИЯ ДИНАМИЧЕСКИМИ ОБЪЕКТАМИ

Д. А. Дзюба, А. Н. Чернодуб

***Аннотация**: Предложен новый метод нейроуправления с применением нелинейного нейроконтроллера на основе рекуррентного прерцептрона, обучаемого в врежиме реального времени с использованием метода контролируемого возмущения управления динамическим объектом. Обучение нейроконтроллера осуществляется без использования нейроэмулятора, предварительно обученного прямой или инверсной динамике объекта управления. Приводятся данные численных экспериментов по сравнению качества нового метода нейроуправления и традиционного ПИД-управления на примере решения задачи стабилизации перевернутого маятника.*

***Ключевые слова**: нейроуправление, рекуррентные нейронные сети, нелинейные динамические объекты, инверсная динамика, стабилизация перевернутого маятника, обучение с подкреплением.*

***ACM Classification Keywords**: I.2.8 Problem Solving, Control Methods, and Search - Control theory, I.2.9 Robotics - Autonomous vehicles, I.2.6 Learning - Connectionism and neural nets.*

1. Вступление

Применение искусственных нейронных сетей для создания обучаемых систем управления динамическими объектами является перспективным направлением в теории управления, теории искусственного интеллекта, нейрофизиологии [Омату и др., 2000]. Нейронные сети обладают многими интригующими свойствами, которые делают их мощным инструментом для создания управляющих систем: способность к обучению на примерах, способность к гладкой интерполяции и экстраполяции данных на ранее не виденных нейросетью примерах, возможность синтеза нелинейных контроллеров, способность адаптации к изменяющимся свойствам объекта управления и внешней среды в режиме реального времени, большая по сравнению с классической фон-Неймановской архитектурой устойчивость к повреждениям своих элементов в силу изначально заложенного в нейросетевую архитектуру параллелизма. Среди ранних примеров систем нейроуправления можно назвать работу Б. Видроу [Widrow, 1986]. Он применил линейные нейронные сети с линией задержек на входе для синтеза системы управления, состоящей из двух модулей, один из которых был обучен инверсной динамике объекта и работал как прямой контроллер без обратной связи, а второй выполнял функцию восстановления исходного сигнала из зашумленного сигнала на входе объекта. Минусами этой схемы является ее неспособность работать с нестабильными объектами, необходимость наличия точной математической или имитационной модели объекта для обучения модуля фильтрации шума. В работе Д. Псалтиса и А. Сидериса [Psaltis & Sideris, 1988] в качестве нейросети использован статический многослойный персептрон с линией задержек, что позволяет синтезировать нелинейный контроллер, позволяющий обеспечивать повышенное качество управления при работе с нелинейными объектами управления, все остальные минусы схемы Б. Видроу остались неразрешенными. В этой же работе Д. Псалтис и др. предложили метод специализированного инверсного управления, в котором контроллер на основе многослойного персептрона с линией задержек обучался управлять объектом на основе ошибки между реальным выходом объекта и целевым выходом

(уставкой). Для применения этой схемы необходимо знание якобиана объекта управления, либо аппроксимация формирующих якобиан частных производных объекта управления по значениям его входов и выходов, на основе использования аппарата вычислительной математики. В другой схеме нейроуправления, предложенной независимо К.С. Нарендрой и К. Пасарати [Narendra & Pasarthy, 1990], Б. Вербосом [Werbos, 1990] и Джорданом и Румельхартом [Jordan & Rumelhart, 1990], 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 поведения объекта управления. Также в работе [Narendra & Pasarthy, 1990], было указано, что для эмуляции поведения динамического объекта наилучшей моделью нейронных сетей являются рекуррентные многослойные персептроны, позже Х. Сигельман и др. доказали это строго [Siegelmann, 1997].

В описанных методах нейроуправления с обучением нейроконтроллера по ошибке отклонения объекта управления от целевой траектории используется мгновенное значение ошибки на каждом такте работы системы. Нейроконтроллер на этапе функционирования выполняет действия, ориентированные на минимизацию именно текущей ошибки, что может дать худшее интегральное качество управления в долгосрочной перспективе. Для преодоления этого недостатка, были предложены методы нейросетевого прогнозирующего управления [Hagan & Demuth, 1999] и системы адаптивных критиков [Prokhorov & Wunsch, 1997]. В схеме нейросетевого прогнозирующего управления также производится обучение нейроэмулятора прямой динамике объекта, который затем используется для предсказания поведения объекта на несколько шагов вперед. При этом пробуются разные управляющие воздействия, и выбирается то, которое дает наименьшую интегральную ошибку. Перебор различных управляющих воздействий осуществляет специальный оптимизационный модуль в составе системы управления. Работа систем адаптивных критиков основана на применении принципа Беллмана для выбора контроллером управляющего действия, оптимального в смысле определения наилучшей стратегии. В них выделяется отдельный нейросетевой модуль на основе многослойного персептрона, 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. Обучение контроллера, входящего в систему управления адаптивного критика, выполняется по принципу минимизации отклика возвращаемого критиком. В ходе обучения системы, оба модуля обучаются попеременно: критик обучается выполнять более качественные оценки (итерация по значениям), а контроллер обучается осуществлять более правильные действия, ведущие к лучшим оценкам (итерация по стратегиям). В случае, когда критик прогнозирует ошибку только на один шаг вперед, схема управления на основе адаптивной критики фактически вырождается в схему с обратным распространением ошибки через прямой нейроэмулятор. Кроме этих схем, также были предложены схемы обучения нейронной сети для корректировки коэффициентов ПИД-контроллера [Ruano et. al., 1992], схемы по использованию нейроконтроллеров параллельно с существующими классическими ПИД-контроллерами [Kawato et. al., 1988], нейросистемы на основе многомодульных самоорганизующихся нейросетей [Wolpert & Kawato, 1998], [Ronco et. al, 1996] и др.

Наблюдаемая тенденция к усложнению нейроуправления отражает стремление возможно более точно представить инверсную модель нелинейного объекта управления, используемую при обучении нейроконтроллера. Модель на основе прямого нейроэмулятора с механизмом обратного распространения ошибки дает возможность учитывать нелинейность объекта и фокусировать управление в локальной области, допускающей линейную аппроксимацию его инверсной динамики. Однако создание эффективного нейроэмулятора для объекта со сложным нестационарным поведением часто

проблематично. В таких случаях информацию, необходимую для обучения нейроконтроллера получают с использованием более сложных методов типа прогнозирующего нейроуправления, адаптивной критики и т.п. Однако, всегда ли оправдано такое усложнение нейроуправления?

Мы предлагаем новую схему нейроуправления, в которой нейроконтроллер, основанный на рекуррентной нейронной сети, обучается управлению оперируя непосредственно объектом в режиме реального времени. Мы используем контролируемое малое возмущение управления, реакция на которое служит для определения направления коррекции работы контроллера. В нашей схеме отсутствует нейроэмулятор, а направление коррекции управления оценивается аналитически на каждом шаге обучения системы, что в итоге приводит к формированию инверсной модели объекта управления непосредственно в нейроконтроллере. Для оценки эффективности предлагаемой схемы проводится экспериментальное сравнение ее работы с ПИД-управлением на примере задачи стабилизации положения перевернутого маятника.

2. Метод контролируемого возмущения управления

Для решения задачи управления с помощью нейроконтроллера необходимо предоставить сети либо пример оптимального управления (которое в ряде случаев неизвестно), либо знак производной от качества управления по управляющему сигналу для коррекции текущего выхода сети (обучение с подкреплением). Именно решению второй задачи и посвящена данная работа.

Для обучения с подкреплением используется информация о том, в какую сторону следует менять управление, чтобы его улучшить. По сути, это означает, что строится некоторая инверсная модель объекта управления, которая давала бы правильный знак производной от состояния объекта по управлению, и который можно было бы трансформировать в обучающий сигнал нейросети. Часто для получения такой инверсной модели используют нейроэмулятор, однако, поскольку нас интересует только знак, а не точная величина производной, можно воспользоваться более простым методом, а именно — определить его анализируя непосредственно поведение объекта. Если мы сравниваем состояние объекта после применения двух близких управлений — то, при соблюдении некоторых условий, мы можем в некотором приближении определить производную от состояния объекта по примененному управлению.

Найти производную от состояния по управлению можно следующим образом: пусть X_0 — обобщенные координаты системы в некоторый момент времени, T — целевое положение. Применим на следующем временном шаге некоторое управление $U_1 - h$, где h – вектор с одной ненулевой компонентой, по которой берется производная, значение которой достаточно мало (критерий достаточности будет определен ниже), и обозначим полученные координаты как X_1, а на шаге после этого — $U_1 + h$, и обозначим полученные координаты как X_2. Теперь наша задача состоит в том, чтобы определить, какое из этих управлений в большей степени соответствует достижению целевого положения. Точный ответ мог бы дать эксперимент, в котором объект вернули из состояния 1 в состояние 0, применили второе управление, и сравнили полученные состояния с целевым. Однако, поскольку мы имеем дело с реальным объектом, а не его моделью, это невозможно. Тем не менее, эту задачу можно решить в некотором приближении. Нами было использовано первое приближение, в нем точка, в которой оказалась бы система, если бы в состоянии 0 к ней было приложено управление $U_1 + h$, определяется как:

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

здесь первый член — собственно исходное положение системы X_0, второй член — это фактическое смещение системы, полученное в результате применения второго управления, а третий член вводит поправку на то, что за счет применения первого управления, скорость системы изменилась.

Тогда производная от положения по некоторой компоненте управления может быть записана как

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

где индекс i соответствует номеру единственной ненулевой компоненты вектора h.

Теперь, когда у нас есть возможность сравнивать результаты различных управлений, мы можем применить следующий метод:

- На вход нейроконтроллера подается текущее состояние системы (координаты, скорости), и сеть генерирует некоторое управление U;
- К системе последовательно применяется управление $U + h$ и $U - h$, где номер ненулевой компоненты h последовательно пробегает все значения, с запоминанием полученных состояний;
- Для каждой пары полученных состояний системы вычисляется производная по соответствующей компоненте управления описанным выше способом;
- Выполняется один шаг обучения рекуррентной сети, где входом служит исходное состояние системы, а целевым значением — вектор U, смещенный согласно знаку производных в сторону приближения к целевому положению T.

Величина h выбирается достаточно малой, чтобы нелинейность системы слабо проявлялась на разнице между управлениями $U + h$ и $U - h$, однако не слишком малой, поскольку она связана со скоростью обучения нейронной сети.

Описанный процесс — это один шаг обучения сети, повторив его достаточное число раз, можно научить рекуррентную нейронную сеть поддерживать заданное положение системы.

3. Модель перевернутого маятника

В качестве полигона для испытания системы нейроуправления мы выбрали программную модель объекта «тележка с перевернутым маятником» в силу нескольких причин. Во-первых, этот объект является классическим объектом в теории управления [Michie & Chambers, 1968] и сейчас насчитывается более 2000 научных статей с описанием алгоритмов управления применительно к этому динамическому объекту [Стюарт и Норвиг, 2006], в том числе на основе как рекуррентных одномодульных [Wei et. al, 2001], так и нерекуррентных многомодульных [Ronco et. al, 1996] нейронных сетей. Во-вторых, это нелинейный динамический объект с сильно выраженной нелинейностью при отклонении маятника больше $12\circ$ от вертикальной оси [Ronco et. al, 1996], также он является сложным объектом для расчета управления на основе аппарата классической теории оптимального управления [Slotine & Li, 1991].

Схема перевернутого маятника показана на рис. 1. Тележка массы M может горизонтально перемещаться в одной степени свободы, вправо и влево. На тележке закреплен маятник массы m длины l, образующий угол θ с вертикальной осью. В случае, когда $\theta = 0$, 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 вниз. Для преодоления этого, тележка оснащена мотором, на который может подаваться сила F , которая двигает тележку вместе с маятником вправо или влево.

Динамика перевернутого маятника описывается системой уравнений:

$$(M+m)\ddot{x} - ml\ddot{\theta}\cos\theta + ml\dot{\theta}^2\sin\theta = F \quad (3)$$

$$ml(-g\sin\theta - \ddot{x}\cos\theta + l\ddot{\theta}) = 0 \quad (4)$$

Целью задачи контроллера является поддержание маятника в вертикальном положении в условиях внешних помех как можно более долгое время, получая на вход значение текущего угла отклонения θ и генерируя значение управляющего воздействия F .

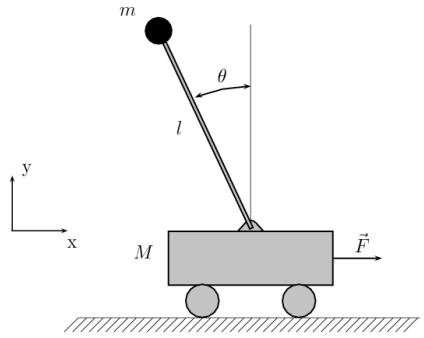

Рис. 1. Схема перевернутого маятника на тележке

Более качественной считается та система управления, которая может удерживать маятник на весу более долгое время. В ходе испытаний, для упрощения восприятия результатов экспериментов, проводился контроль только отклонения маятника, положение тележки в пространстве не фиксировалось. Так же было сделано, например, в работе [Ronco et. al, 1996].

4. Экспериментальное исследование качества управления полученной нейросети

В нашей работе в качестве оценки качества управления мы использовали время в секундах, которое проходит от установки маятника в положение равновесия до выхода угла маятника за определенную границу при некотором уровне шума (в этой работе приведены результаты для порогового угла, равного примерно 10 градусам). Этот критерий позволяет эффективно сравнивать управление в ситуациях, когда маятник достаточно далек от положения равновесия, и нелинейные эффекты вносят ощутимый вклад, и достаточно популярен [Michie & Chambers, 1968], [Стюарт & Норвиг, 2006], .

В целом эксперимент был построен следующим образом: модель маятника, импортированная из среды Simulink, приводилась в положение равновесия, потом к ней последовательно применялся один шаг случайного управления с заданной максимальной амплитудой (шум), и один шаг управления с помощью контроллера — рекуррентной сети, или PID, в зависимости от режима эксперимента. Если угол отклонения маятника не превысил заданный порог — шум и управление применялись повторно, и так до тех пор, пока угол отклонения маятника не достигал порогового значения. Время, которое прошло от момента установки маятника в положение равновесия до момента, когда отклонение превышало порог, записывалось как результат управления, маятник устанавливался в положение равновесия, и процесс повторялся снова. Схема управления показана на Рис. 2.

Структура нейронной сети была следующей: на вход подавалось значение текущего угла, угловой скорости, и уставка по углу, эти данные попадали в линию задержек на 2 такта, выход сети попадал на вход через линию задержек в три такта. Таким образом сеть получала 9 внешних входов и 3 задержанных выхода, итого на вход попадало 12 чисел. Было проведено несколько экспериментов с разным числом нейронов в скрытом слое (от 3 до 12 нейронов). Эксперименты показали, что изменения числа скрытых нейронов в таких пределах дают не очень большие изменения качества управления, и, поскольку задача оценки влияния этого параметра на эффективность управления не являлась предметом данного исследования, размер скрытого слоя был выбран равным 6 нейронам.

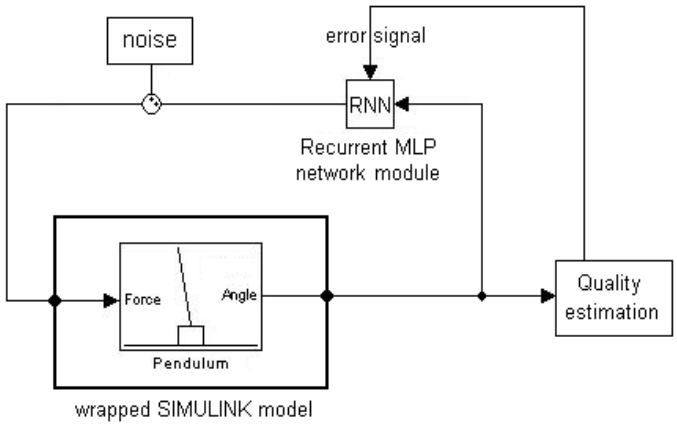

Рис. 2. Схема управления с помощью нейронной сети

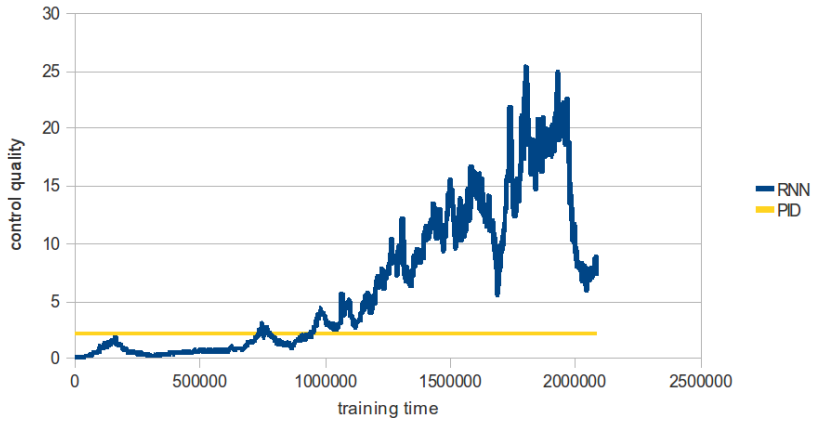

Рис. 3. Зависимость качества управления от времени обучения рекуррентной сети. Для сравнения приведено качество управления для PID-контроллера

Полученные данные изображены в виде графика на Рис. 3. Эксперименты показали, что качество управления рекуррентной сетью меняется немонотонно в процессе обучения, поэтому для оценки сеть использовалась в таком режиме:

- Процесс обучения, описанный в разделе 2, повторялся 3000 раз, с установкой маятника в положение равновесия в начале серии повторов;
- Сеть переключалась в режим управления, маятник устанавливался в положение равновесия, замерялось время до превышения порогового угла отклонения, замер повторялся 200 раз. Усредненная за это время характеристика качества управления использовалась как оценка сети на данном этапе обучения.

5. Выводы

Предложенный метод обучения рекуррентных сетей задачам управления может быть полезен в случаях, когда нет примеров оптимального управления некоторым объектом, но есть возможность оценить его

состояние с точки зрения достижения некоторого целевого положения. Отсутствие в схеме нейроэмулятора позволяет существенно сократить время обучения, однако накладывает некоторые ограничения на объект управления — а именно, предложенная функция оценки знака производной состояния по управлению должна быть корректной для него.

Использованные нами рекуррентные сети имеют большой потенциал в подобных задачах, т.к. с одной стороны, рекуррентные связи обеспечивают отслеживание динамического состояния объекта, а с другой — нелинейность, присущая нейросетям, позволяет более эффективно управлять нелинейными объектами в сравнении с аналитически простыми линейными методами.

Обучаемые с помощью предлагаемого метода нейроконтроллеры могут адаптироваться к изменяющимся свойствам объекта управления, что делает их более выигрышными по сравнению с классическими неадаптивными контроллерам в тех случаях, когда параметры объекта меняются в ходе его функционирования.

Эти общие соображения полностью подтверждаются результатами экспериментов. Однако вопрос об оптимальной структуре сети и режиме обучения пока остается открытым.

Список литературы

[Омату и др., 2000] С. Омату, М. Халид, Р. Юсоф. Нейроуправление и его приложения, пер. с англ. — М: ИПРЖР, 2000.

[Резник, 2009] А.М.Різник. Динамічні рекурентні нейронні мережі: Математичні Машини і Системи, 2009, №2, с.3-26.

[Резник и Дзюба, 2010] А.М. Різник, Д.О.Дзюба. Динамічна автоассоціативна пам'ять, заснована на відкритій рекурентній нейронній мережі: Математичні Машини і Системи, 2010, №2, р. 45-51.

[Стюарт & Норвиг, 2006] Р. Стюарт, П. Норвиг. Искусственный интеллект: современный подход, 2-е изд.: Пер. с англ. — М. : Издательский дом "Вильяме", 2006.

[Jordan & Rumelhart, 1990] M.I. Jordan and D.E. Rumelhart. Forwardmodels: Supervised learning with a distal teacher. In: Cognitive Science, Vol. 16, pp.313- 355, 1990.

[Hagan & Demuth, 1999] 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.

[Narendra & Pasarthy, 1990] K.S. Narendra, K. Parthasarathy K, Identification and control of dynamical systems using neural networks. In: IEEE Transactions on Neural Networks, 1, 1990, p. 4-27.

[Kawato et. al., 1988] M.Kawato, Y. Uno, M. Isobe and R.Suzuki. Hierarchical neural network model for voluntary movement with application to robotics. In: IEEE Control Systems Magazine. Vol. 8, pp.8- 16, 1988.

[Reznik & Dziuba, 2009] A.M. Reznik, D.A. Dziuba. Dynamic Associative Memory Based on Open Recurrent Neural Network. In: Proceeding of IJCNN'09, Atlanta, Georgia, USA, June 14-19, 2009.

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

[Psaltis & Sideris, 1988] Demetri Psaltis, Athanasios Sideris, and Alan A. Yamamura. A Multilayered Neural Network Controller. In: IEEE Control Systems Magazine (1988), v. 8, Issue 2, p.17-21.

[Ronco et. al, 1996] E. Ronco, J. Gawthrop, Y. Matter. Incremental Modular controllers network. In: Proceeding of the International Conference on Intelligent and Cognitive Systems (ICISC'96).

[Ruano et. al., 1992] A. E. B. Ruano, P. J. Fleming, and D. I. Jones. Connectionist approach to PID tuning. In: IEEE Proceedings, Part D, 129:279-285, 1992.

[Siegelmann, 1997] 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.

[Slotine & Li, 1991] J.-J. E. Slotine, W. Li. Nonlinear Control Systems Design. In: Prentice-Hall, 1991, p. 193.

[Sutton & Barto, 1998] R.S. Sutton, A.G. Barto. Reinforcement Learning: An introduction. A Bradfoord book, 1998.

[Michie & Chambers, 1968] D. Michie and Chambers. BOXES: An experiment in adaptive control. In: Dale E. and Michie D. (Eds.), Machine Intelligence 2, 1968, p. 125-133, Elsevier/North-Holland, Amsterdam, London, New York.

[Wei et. al, 2001] W. Wei, W. von Seelen. Recurrent neuro-controller for an inverted pendulum using evolution strategy. In: International Journal of Systems Science, 2001, volume 32, number 5, pages 643-650.

[Werbos, 1990] P. Werbos, Backpropagation through time: what it does and how to do it. In: Proc. IEEE, Vol. 78, No. 10, October 1990.

[Widrow, 1986] Bernard Widrow. Adaptive Inverse Control. In: IFAC Adaptive Systems in Control and Signal Processing, Lund, Sweden, 1986.

[Wolpert & Kawato, 1998] D.M. Wolpert , M. Kawato. Multiple Paired Forward and Inverse Models for Motor Control. In: Neural Networks, 1998, Vol. 11 , Issue 7-8, pp. 1317 – 1329.

Информация об авторах

Д.А. Дзюба — *аспирант Института Программных Систем НАНУ, работает в Институте Проблем Математических Машин и Систем НАНУ. Адрес: г. Киев, ул. академика Глушкова 42, ИПММС НАНУ, отдел Нейротехнологий; e-mail: ddziuba@immsp.kiev.ua*

Научные интересы: динамическая ассоциативная память, нейроуправление, рекуррентные нейронные сети, интеллектуальная обработка изображений.

А.Н. Чернодуб *– аспирант Института Программных Систем НАНУ, работает в Институте Проблем Математических Машин и Систем НАНУ. Адрес: г. Киев, ул. академика Глушкова 42, ИПММС НАНУ, отдел Нейротехнологий; e-mail: achernodub@immsp.kiev.ua*

Научные интересы: интеллектуальная обработка изображений, биометрическая идентификация, нейронные сети, нейроуправление.