

Сравнение различных алгоритмов обучения

Алгоритм обучения	α_1	α_2	α_3	Кол-во тестов	Кол-во успешн. тестов	Среднее кол-во итераций	Среднее время, сек	E_S
SA	Адапт.	0,1	Адапт.	100	100	15,7	0,0847	1,3E-7
	0,5	0,5	0,5	100	98	16,5	0,0747	1,12E-6
BP	0,05	0,05	—	100	100	129	0,578	8,3E-6
HL	0,03	0,03	—	100	100	19,9	0,0763	3,7E-6

SA- алгоритм послонного обучения; **BP**-алгоритм обратного распространения ошибки; **HL**-алгоритм Хебба

информации. В качестве эталонных данных используются значения \bar{Y} , полученные на предыдущем шаге алгоритма.

Таблица 2

5. Пункт 4 продолжается до тех пор, пока суммарная среднеквадратичная ошибка прямого слоя не станет меньше заданной.

Сжатый массив

0,765	0,816	-0,285
0,128	0,0386	-0,556
0,482	0,395	-0,798
0,0498	-0,115	-0,894
0,439	0,464	-0,189
0,374	0,404	-0,11
0,347	0,327	-0,355
0,376	0,238	-0,99
0,156	0,0427	-0,702
0,177	0,112	-0,466

5. ЭКСПЕРИМЕНТЫ

Анализ возможностей метода послонного обучения проводился на различных числовых массивах и изображениях. Как показали эксперименты, эффективность его зависит от входного массива данных. Один из примеров входного массива чисел показан в таблице 1. В этом случае нейронная сеть состоит из 6 входных, 3 скрытых и 6 выходных нейронных элементов. В качестве функции активации использовалась функция гиперболического тангенса. В таблице 2 представлен сжатый массив данных, который получается после обучения нейронной сети.

Таблица 1

Входной массив

0,9	0,1	0,9	0,9	0,9	0,9
0,6	0,7	0,6	0,6	0,6	0,6
0,9	0,8	0,9	0,9	0,9	0,9
0,7	0,9	0,7	0,7	0,7	0,7
0,7	0,1	0,7	0,7	0,7	0,7
0,6	0,0	0,6	0,6	0,6	0,6
0,7	0,4	0,7	0,7	0,7	0,7
0,9	0,9	0,9	0,9	0,9	0,9
0,7	0,8	0,7	0,7	0,7	0,7
0,6	0,6	0,6	0,6	0,6	0,6

Пусть α_1 – шаг обучения для модификации весовых коэффициентов прямого слоя w_{ij} , α_2 – шаг обучения для модификации весовых коэффициентов обратного слоя w_{ji}' и α_3 – скорость модификации выходов скрытого слоя y_j . В таблице 3 представлена информация о временной сложности различных алгоритмов обучения. Как следует из таблицы, при использовании алгоритма послонного обучения получаются лучшие результаты по сравнению с другими алгоритмами.

ЗАКЛЮЧЕНИЕ

В работе предложен новый метод обучения рециркуляционных нейронных сетей. Данный алгоритм называется методом послонного обучения. Результаты экспериментов с предложенным методом показали, что он позволяет достичь более высокой скорости обучения по сравнению с другими алгоритмами.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- Jolliffe I.T. Principal Component Analysis. Springer-Verlag, 1986
- P. Baldi and K. Hornik. Learning in linear networks; A Survey // IEEE Trans. Neural Networks, Vol. 6, pp. 837-857, 1995
- C. Fyfe and R.Baddley. Non-linear data structure extraction using simple hebbian networks. Biological Cybernetics, 72(6); 533-541, 1995.
- E. Oja. Principal components, minor components and linear networks // Neural Networks, vol.5, pp.927-935, 1992
- V.Golovko, V.Gladyschuk. Unsupervised training algorithm for recirculation neural networks // Proceedings of ICNNAI'99, Brest, 1999, p.19-25
- H.Ando, S.Suzuki, T.Fujita. Unsupervised visual learning of three-dimensional objects using a modular network architecture // Neural Networks, vol. 12, pp. 1037-1051, 1999

УДК 681.324.01

Головко В.А., Савицкий Ю.В., Фоменкова Н.Е.

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ И МОДЕЛИРОВАНИЯ НЕЛИНЕЙНЫХ СИСТЕМ

ВВЕДЕНИЕ

Способность нейронных сетей к обобщению результатов создает предпосылки для построения на базе их различного рода прогнозирующих систем. Пусть дан временной ряд $x(t)$ на промежутке $t=1, m$. Тогда задача прогнозирования состоит

в нахождении продолжения временного ряда на неизвестном промежутке времени, т.е. необходимо определить значения $x(m+1)$, $x(m+2)$ и так далее.

Для прогнозирования временных рядов используется ме-

Фоменкова Наталья Евгеньевна. Аспирант каф. ЭВМиС БГТУ. БГТУ, Беларусь, г. Брест, ул. Московская 267.

год скользящего окна. Он характеризуется длиной окна T , которая равняется количеству элементов ряда одновременно подаваемых на нейронную сеть. Тогда нелинейную модель прогнозирования можно представить, как

$$X(k)=F[X(k-1), X(k-2), \dots, X(k-T)], \quad (1)$$

где F – нелинейная функция.

Как было показано в работах [1,2] многослойная нейронная сеть с двумя или более слоями способна аппроксимировать любую непрерывную функцию. Это является основой для использования нейронных сетей в прогнозировании. Прогнозирующие свойства нейронной сети можно использовать для идентификации поведения нелинейных динамических систем, например, для построения странных аттракторов [3].

Это базируется на теореме Такенса [4], который показал, что можно восстановить некоторые свойства аттрактора в фазовом пространстве используя временную последовательность только одной составляющей. Для построения аттракторов можно использовать метод псевдофазового пространства [5]. Он заключается в том, что можно восстановить топологию аттрактора исходя из наблюдения сигнала $X(t)$ и выбирая в качестве системы координат $X(t)$ и $X(t+\tau)$. Таким образом, если спрогнозировать при помощи нейронной сети сигнал $X(t)$, то на основе этих данных можно построить фазовую траекторию системы.

1. АРХИТЕКТУРА СЕТИ

В качестве архитектуры сети для прогнозирования и построения аттракторов будем использовать рекуррентную сеть Джордана [6]. Такая сеть состоит из одного выходного нейронного элемента с линейной функцией активации, скрытого нелинейного слоя и входного распределительного слоя (см. рисунок 1). При этом выходной нейронный элемент сети соединяется посредством контекстного нейрона со всеми

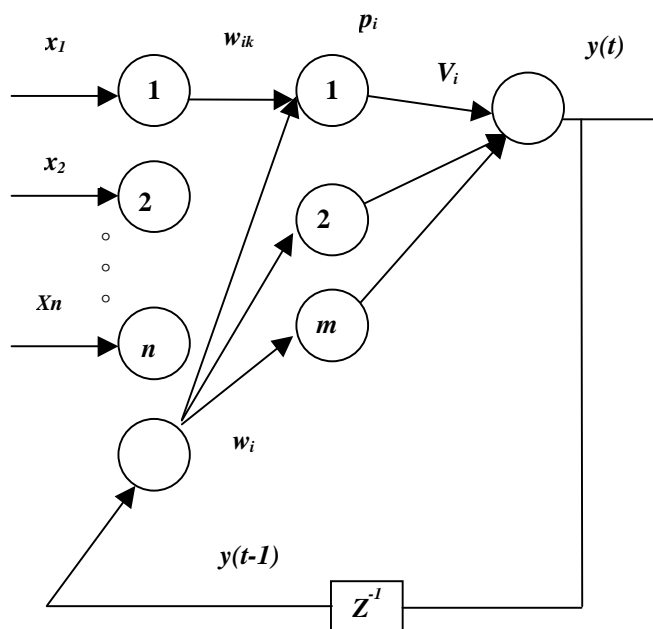


Рисунок 1. Архитектура рекуррентной сети

нейронами скрытого слоя через элемент задержки Z^{-1} . Это позволяет учитывать результаты преобразования сетью информации на предыдущем этапе для обработки входного вектора на следующем этапе функционирования сети.

Выходное значение сети определяется следующим образом:

$$y(t) = \sum v_i p_i(t) - T \quad (2)$$

где v_i – синаптическая связь от i – го нейрона скрытого слоя к выходному нейрону, T – пороговое значение нейронного элемента выходного слоя, $p_i(t)$ – выходное значение i -го нейрона скрытого слоя.

Выходное значение i – го нейрона скрытого слоя определяется как функция нелинейного преобразования от взвешенной суммы.

$$p_i(t) = F(S_i(t)), \quad (3)$$

где F – функция активации нейронных элементов скрытого слоя; $S_i(t)$ – взвешенная сумма i -го нейрона скрытого слоя.

Взвешенная сумма определяется как

$$S_i(t) = \sum_{k=1}^n x_k(t)w_{ki} + w_i y(t-1) - T_i, \quad (4)$$

где w_{ki} – синаптическая связь между k -м и i -м нейронами соответственно входного и скрытого слоев; T_i – пороговое значение i -го нейрона скрытого слоя;

w_i синаптическая связь между контекстным и i -м нейронном скрытого слоя.

2. ОБУЧЕНИЕ

Для обучения рекуррентной сети используется алгоритм обратного распространения ошибки, адаптированный к архитектуре сети. Он минимизирует суммарную среднеквадратичную ошибку сети:

$$E_s = \frac{1}{2} \sum_{k=1}^L (y^k - e^k)^2, \quad (5)$$

где y^k – выходное значение нейронной сети для k -го входного образа, e^k – эталонное значение нейронной сети для k -го входного образа, L – размерность обучающей выборки.

Рассмотрим правила модификации синаптических связей. Пусть среднеквадратичная ошибка для одного образа равняется:

$$E = \frac{1}{2} (y(t) - e)^2 \quad (6)$$

Тогда в соответствии с методом градиентного спуска для выходного слоя сети можно записать следующие выражения:

$$v_i(t+1) = v_i(t) - \alpha_1 \frac{\partial E}{\partial v_i(t)} = v_i(t) - \alpha_1 (y(t) - e) p_i(t) \quad (7)$$

$$T_i(t+1) = T_i(t) - \alpha_1 \frac{\partial E}{\partial T_i(t)} = T_i(t) - \alpha_1 (y(t) - e)$$

Соответственно для скрытого слоя

$$w_{ki}(t+1) = w_{ki}(t) - \alpha_2 (y(t) - e) v_i F'(S_i) x_k(t) \quad (8)$$

$$T_i(t+1) = T_i(t) - \alpha_2 (y(t) - e) v_i F'(S_i)$$

$$w_i(t+1) = w_i(t) - \alpha_2 (y(t) - e) v_i F'(S_i) y(t-1)$$

где $F'(S_i)$ – производная функции активации нейронных элементов скрытого слоя по взвешенной сумме, α_1 и α_2 шаг обучения соответственно для выходного и скрытого слоев сети.

Для обучения сети можно использовать адаптивный шаг обучения. В этом случае [7]

$$\alpha_1 = \frac{1}{1 + \sum_{i=1}^m p_i^2}$$

$$\alpha_2 = \frac{\sum_i v_i^2 F'(S_i)}{F'(0) \left(1 + \sum_k x_k^2\right) \sum_i v_i^2 (F'(S_i))^2}, \quad (9)$$

где $F'(0) = \frac{\partial F}{\partial S_i}$ при $S_i = 0$; $F'(S_i) = \frac{\partial F}{\partial S_i}$

В качестве функции активации может использоваться сигмоидная функция или гиперболический тангенс. При использовании сигмоидной функции

$$F'(S_i) = p_i(1 - p_i) \quad (10)$$

$$F'(0) = \frac{1}{4}$$

Если в качестве функции активации применяется гиперболический тангенс, то

$$F'(S_i) = (1 - p_i^2) \quad (11)$$

$$F'(0) = 1$$

При расчете адаптивного шага по приведенным выше выражениям рекомендуется ограничивать его по абсолютной величине. Это необходимо для обеспечения стабильности обучения. Как уже отмечалось, для обучения рекуррентной сети используется алгоритм обратного распространения ошибки. В этом случае состояние контекстного нейронного элемента в начальный момент времени принимается равным нулю.

3. ПРОГНОЗИРОВАНИЕ СТРАННЫХ АТТРАКТОРОВ

Рассмотрим применение рекуррентных сетей для построения странных аттракторов. В этом случае используется метод псевдофазовой плоскости. Пусть нам известен сигнал $X(t)$ на небольшом промежутке времени. Тогда при помощи рекуррентной сети мы можем пролонгировать данный сигнал на упреждающий промежуток времени. Отобразив полученные значения $X(t)$ на псевдофазовую плоскость можно идентифици-

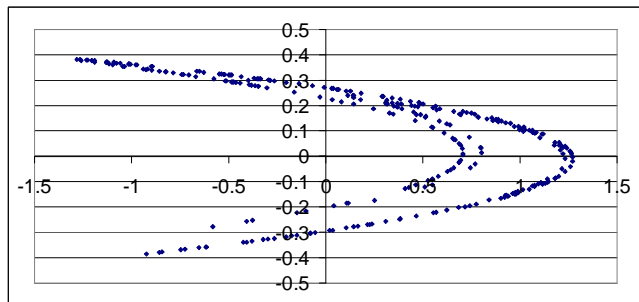


Рисунок 2. Оригинальный аттрактор

цировать поведение динамической системы. Рассмотрим построение странного аттрактора, который описывается отображением Энона:

$$X_{n+1} = 1 - \alpha X_n^2 + Y_n, \quad (12)$$

$$Y_{n+1} = \beta X_n,$$

где $\alpha = 1.4$; $\beta = 0.3$.

Отображение Энона имеет фрактальную размерность 1,26.

Общий вид аттрактора представлен на рисунке 2 рассчитанный для 300 точек по формулам приведенным выше.

Предположим, что нам известны значения координаты X для $n=15$. Возьмем рекуррентную нейронную сеть с 3 входными 2 скрытыми и одним выходным нейронными элементами. В качестве функции активации будем использовать гиперболический тангенс. Обучим такую сеть на известных данных и спрогнозируем значения ряда $X(k)$ на 300 шагов вперед. Для обучения и прогнозирования будем применять метод скользящего окна размерностью 3 и шагом 1. При этом будем использовать члены временного ряда умноженные на коэффициент 0,1.

В таблице 1 приведены результаты обучения сети при использовании адаптивного и постоянного шага. Как следует из таблицы применение адаптивного шага обучения более эффективно по сравнению с постоянным шагом.

Таблица 1

Шаг	Кол-во итераций	Суммарная среднеквадратичная ошибка	
Адаптивный	7652	0,00002	
Постоянный	$\alpha=0,1$	20875	0,00002
	$\alpha=0,2$	15234	0,00002
	$\alpha=0,3$	9310	0,00002

Графики эталонной и прогнозируемой функции приведены на рисунке 3 для $n=40$. При этом темная линия изображает эталонный ряд, а светлая линия – прогнозируемый ряд.

Как следует из рисунка 3, значение эталонного ряда отличаются от прогнозируемого за пределами обучающей выборки.

Такое отличие обуславливается хаотической природой сигнала X . В результате оказывается недостаточным значений сигнала X на конечном интервале времени для его предска-

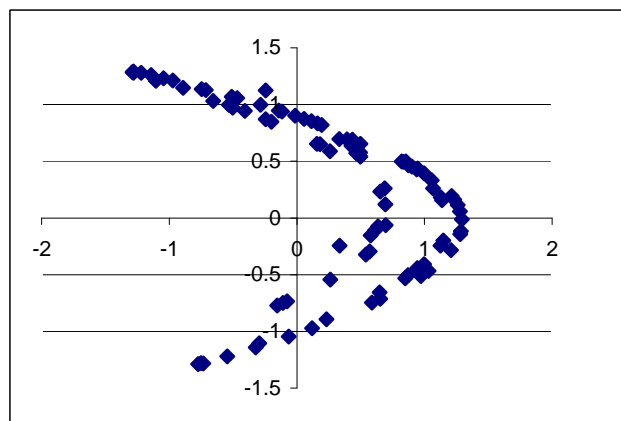


Рисунок 4. Прогнозируемый аттрактор

ния. Это одно из фундаментальных положений теории хаоса.

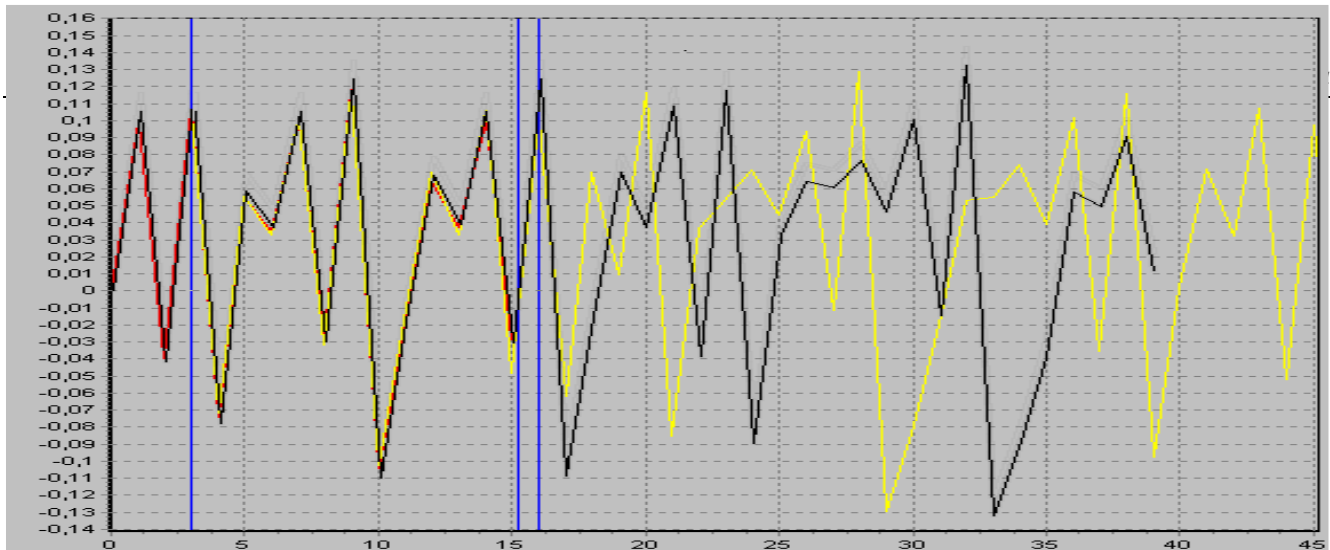


Рисунок 3 .Прогнозируемый и эталонный ряды

Рассмотрим построение аттрактора Энона. Для этого отобразим полученные при помощи рекуррентной сети 300 значений ряда X на псевдофазовую плоскость (рисунок 4).

Как следует из рисунка, форма прогнозируемого аттрактора соответствует оригинальному аттрактору. Таким образом, рекуррентная нейронная сеть позволяет моделировать поведение нелинейных динамических систем.

ЗАКЛЮЧЕНИЕ

В работе рассмотрена технология применения рекуррентных нейронных сетей для прогнозирования и построения странных аттракторов. Для моделирования аттракторов использовалось отображение Энона. В дальнейшем планируется моделирование других хаотических процессов при помощи нейронных сетей.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Hornik K., Stinchcombe M., White H. Multilaer feedforward networks are universal approximators// Neural Networks, 1989, vol.2,359-366
2. Cubenko G. Approximation by superpositions of a sigmoidal function// Mathematics of Control,Signals and Systems, 1989, vol.2, N.4,303-314
3. Kim J.,Kasabov N. HyFIS: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems// Neural Networks, 1999,vol.12, N.9, 1301-1319
4. Takens F. Detecting Strange Attractors in Fluid Turbulence. In dynamical Systems and Turbulence, 1981, Springer, Berlin
5. Шустер Г. Детерминированный хаос: Введение. – М.: Мир, 1988, 240 с.
6. Jordan M. Attractor dynamics and parallelism in a connectionist sequential machine//proceedings of the 8 Annual Conference of the Cognitive Science. – Hillsaale: Erlbaum – 1986, p.531-546
7. В.А. Головки. Нейроинтеллект: теория и применение. Книга 1: Организация и обучение нейронных сетей с прямыми и обратными связями. Брест., БПИ, 1999, 264 с.

УДК 681.324.01

Саченко А., Кочан В., Турченко В., Головки В., Савицкий Ю., Дунец А.

ПРОГНОЗИРОВАНИЕ ОШИБОК СЕНСОРНЫХ УСТРОЙСТВ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

ВВЕДЕНИЕ

В подавляющем большинстве современных систем обработки сигналов сенсоров нормируется погрешность измерения выходного сигнала сенсора, а не физической величины. Например, при измерении температуры с помощью сенсоров Honeywell Pt100 [1] и блока Hydra 2625A Fluke [2] соотношение погрешностей элементов измерительного канала больше пятидесяти. Большинство работ по обработке сигнала сенсора [3, 4, 5], рассматривают вопросы, не связанные с повышением его точности. Надежное повышение точности сенсора независимо от условий эксплуатации обеспечивается его периодической поверкой по образцовому или калибровочной с помощью специального калибратора на месте эксплуатации, но операции, реализующие эти методы, достаточно трудоемки [6]. Снижение трудоемкости достигается прогнозированием дрейфа сенсора в межповерочный (межкалибровочный) ин-

тервал [7]. В этом случае наиболее эффективным является применение методов искусственного интеллекта, в частности нейронных сетей [8, 9, 10]. Известно [11], что качество обучения нейронных сетей в сильной степени зависит от объема данных, используемых для обучения. Это обуславливает основное противоречие, возникающее при использовании нейронных сетей для коррекции дрейфа сенсоров. Высококачественное обучение нейронной сети позволяет резко снизить погрешность прогноза и тем самым настолько увеличить межповерочный интервал, что полученное в результате поверок количество данных окажется недостаточным для высококачественного обучения нейронной сети.

Для разрешения этого противоречия, то есть для искусственного увеличения количества точек обучения прогнозирующей нейронной сети, предложено использовать «исторические» данные (данные о дрейфе однотипных сенсоров в

Саченко А.А. Директор института компьютерных информационных технологий Тернопольской академии народного хозяйства.

Турченко В. Аспирант каф. ЭВМиС Бресткого государственного технического университета.

Дунец А.П. Аспирант каф. ЭВМиС Бресткого государственного технического университета.

БГТУ, Беларусь, г. Брест, ул. Московская 267.