ОПТИМИЗАЦИЯ ПРОЦЕССА ОБУЧЕНИЯ ЭКСТРАПОЛИРУЮЩЕЙ НЕЙРОННОЙ СЕТИ

Маньяков Н.В. (Брестский государственный технический университет), Махнист Л.П. (Брестский филиал Института современных знаний им. А.М. Широкова)

1. Введение. Во многих областях прикладного моделирования все чаще встречается задача экстраполяции временных последовательностей: прогнозирование курса валют, прогнозирование среднегодового стока вод и т.д. Решение данных задач актуально с точки зрения принятия решений на упреждающий период для разработки стратегий управления и предопределения состояний. Прочное место для решения этой задач заняли нейронные сети. В предлагаемой работе аргументировано построение прогнозирующей нейронной сети с точки зрения теории нелинейной динамики. Предложен также алгоритм адаптивного изменения параметра градиентного обучения для достижения более быстрой сходимости процесса обучения.

2. Прогнозирование динамики ряда в нейросетевом базисе. Метод экстраполяции временной последовательности основан на рассмотрении временного ряда как ряда измерений некоторой фазовой переменной динамической системы. В соответствии с теоремой Такенса все фазовое пространство может быть реконструировано из измерений только одной фазовой переменной рассматриваемого сигнала с сохранением основных топологических свойств:

Пусть дана динамическая система с решением $(x(t), y(t), \ldots, z(t))$, принадлежащем фазовому пространству размерности d. Используя только одну координату $x(t)$ можно при некоторых условиях построить фигуру в пространстве точек с задержками $(x(t), x(t+t), x(t+2t), \ldots, x(t+(D-1)t))$, которая будет диффеоморфна аттрактору в реальном фазовом пространстве. Размерность D определяется по формуле $D=2[d_F]+1$, где d_F – фрактальная размерность аттрактора и [.] – целая часть числа.

Для использования этой реконструкции необходимо определить временную задержку t и размерность пространства вложения *D*.

Временная задержка t является периодом между компонентами точек в реконструированном псевдофазовом пространстве. Исходя из этого, необходимо выбирать t таким образом, чтобы координаты радиус-вектора, формирующего псевдофазовое пространство вложения, были независимы насколько это возможно. Для определения временной задержки можно использовать автокорреляционную функцию $C(t)$, выбирая при этом в качестве t точку, в которой $C(t)$ первый раз достигает e^{-1} либо время, где абсолютная величина функции автокорреляции достигает первого минимума. Но наиболее оптимальный выбор временной задержки основан на использовании первого минимума функции взаимной информации.

При вычислении размерности псевдофазового пространства вложения можно либо выделить из погружения в пространство большой размерности методом главных компонент наиболее значимые, либо непосредственно вычислить фрактальную размерность приближая ее корреляционной. Однако, первый из них эффективен только для линейных систем, а второй характеризуется временной сложностью. Поэтому в большинстве исследований размерность пространства вложения определяют используя метод «ложных соседей» [1].

После нахождения параметров вложения и задержки можно построить вложение временного ряда в псевдофазовое пространство $\left(x(t), x(t+\tau), \ldots, x(t+(N-1)\tau)\right)$

размерности $N = 2[m]+1$ где m – фрактальная размерность временного ряда, а [.]– целая часть числа. При этом, зная (N-1) координату, можно однозначным образом определить оставшуюся, т.к. временной ряд лежит на поверхность в псевдофазовом пространстве задержек. Тем самым задача прогнозирования временного ряда сводится к задаче аппроксимации функции многих переменных функциями одного переменного. Данная 13 проблема Гилберта может быть решена с использованием двухслойных нейронные сети прямого распространения без обратных связей [2].

Возьмем нелинейный многослойный персептрон (MLP) с как минимум (D-1) нейроном в распределительном слое (где D является размерностью пространства вложения), единственным скрытым слоем нейронов с нелинейной функцией активации и одним нейроном в последнем слое. При обучении сети, на нейроны распределительного слоя подаются значения [$x(t)$, $x(t$+t), ..., $x(t$+(D-2)t)] а в качестве цели берется значение $x(t$+(D-1)t). Такая конструкция нейронной сети позволяет аппроксимировать динамику системы и прогнозируемые значения будут сходиться к аттрактору, подобному исходному.

3. Методика послойного обучения сети. Основным методом обучения многослойных гетерогенных нейронных сетей прямого распространения без обратных связей является метод обратного распространения, представимый для компактности в матричной форме [3]:

Модификация синаптических связей и порогов многослойной гетерогенной нейронной сети производится в соответствии с формулами:

$$w_{j_{n-1}j_n}^{(n)}(t+1) = w_{j_{n-1}j_n}^{(n)}(t) - \alpha^{(n)} \cdot \frac{1}{L} \cdot \sum_{k=1}^{L} C^{(n)} \cdot M_{j_n j_{n-1}}^{(n)} \cdot Y^{(n-1),k}$$

$$T_{j_n}^{(n)}(t+1) = T_{j_n}^{(n)}(t) - \alpha^{(n)} \cdot \frac{1}{L} \cdot \sum_{k=1}^{L} C^{(n)} \cdot M_{j_n(m_{n-1}+1)}^{(n)} \cdot Y^{(n-1),k}$$

$C^{(n)}$ вычисляется рекуррентно:

$$C^{(n)} = C^{(n+1)} \cdot W^{(n+1)} \cdot MF_n' , \quad C^{(N)} = \varepsilon_N^k \cdot MF_N'$$

$$\varepsilon_N^k = \left(\left(y_1^{(N),k} - t_1^k \right) \ \left(y_2^{(N),k} - t_2^k \right) \ \dots \ \left(y_{m_N}^{(N),k} - t_{m_N}^k \right) \right),$$ а

$MF_n' = diag\left(F_n'\left(S_1^{(n),k}\right) \ F_n'\left(S_2^{(n),k}\right) \ \ F_n'\left(S_{m_n}^{(n),k}\right) \right)$ – матрица размерности $m_n \times m_n$, а матрица $M_{j_n j_{n-1}}^{(n)}$ размерности $m_n \times (m_{n-1} + 1)$ состоит из числа 1 на позиции $j_n j_{n-1}$ и нулей в качестве остальных элементов матрицы.

Изменение синаптических связей и порогов сети производится начиная с последнего N-ого до первого слоя сети.

При таком обучении шаг $\alpha^{(n)}$ может оставаться как постоянным, так и адаптивным. В последнем случае сначала, после подачи всех элементов обучающего множества, изменяются синаптические связи последнего слоя. Затем, после того как на модифицированную сеть были поданы все элементы обучающего множества, изменяются синаптические связи предшествующего слоя. И так далее, до достижения первого слоя сети. Затем осуществляется переход к началу алгоритма. Процедура обучения останавливается после того, как ошибка сети не превышает заданную после двух последовательных итераций. Шаги обучения $\alpha^{(n)}$ выбираются для наилучшей минимизации ошибки сети при изменении синаптических связей соответствующих слоев.

Теорема. При послойном обучении многослойной гетерогенной нейронной сети прямого распространения без обратных связей адаптивный шаг для каждого слоя определяется в соответствии с формулами:

$$\alpha^{(n)} = \frac{L \cdot \sum_{j_{n-1}=1}^{m_{n-1}} \sum_{j_n=1}^{m_n} \left(\sum_{k=1}^{L} C^{(n)} \cdot M_{j_n j_{n-1}}^{(n)} \cdot Y^{(n-1),k} \right)}{\sum_{j_{n-1}=1}^{m_{n-1}} \sum_{i_n=1}^{m_n} \sum_{l_{n-1}=1}^{m_{n-1}} \sum_{l_n=1}^{m_n} \left(\sum_{k=1}^{L} \left(\left(K_{l_{n-1} l_n}^{(n),k} \right)^T \cdot U^{(n),k} \cdot \left(K_{j_{n-1} j_n}^{(n),k} \right) \right) \right)},$$

где $K_{j_{n-1} j_n}^{(n),k} = M_{j_n j_{n-1}}^{(n)} \cdot Y^{(n-1),k}$,

а $U^{(n),k} = \left(W^{(n+1)} \cdot MF_n' \right)^T \cdot U^{(n+1),k} \cdot \left(W^{(n+1)} \cdot MF_n' \right) + W^{(n+1)} \cdot MF_n''$ вычисляется рекуррентно с начальным условием $U^{(N),k} = \left(MF_N' \right)^2 + DE^{(N),k} \cdot MF_n'$

ЛИТЕРАТУРА:

1. M.B. Kennel, R. Brown and H.D.I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, Physical Review A 45, 1992. – PP. 3403-3411.

2. K.-I. Funahashi. On the approximate realization of continuous mappings by neural networks, Neural Networks, vol. 2, 1989. – PP. 183-192.

3. Маньяков Н.В., Махнист Л.П. Матричная алгоритмизация обучения многослойных нейронных сетей с использованием градиентных ме-

тодов // Вестник Брестского государственного технического университета. – Брест: БГТУ, 2002. - №5 (17): Физика, математика, химия. – С. 60-64.