

УДК 004.032.26(045)

Е. И. Чумаченко, канд. техн. наук,
В. С. Горбатюк

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ

Национальный технический университет Украины «КПИ», e-mail: lobach21@mail.ru

Предложен алгоритм построения искусственных нейронных сетей для решения задачи прогнозирования, который позволяет автоматизировать процесс построения нейронной сети.

Ключевые слова: прогнозирование, искусственные нейронные сети, многослойный персептрон.

Математическая постановка задачи прогнозирования. Пусть заданы n дискретных отсчетов $\{y_1, y_2, \dots, y_n\}$ в последовательные моменты времени $t_1, t_2, \dots, t_n$. Тогда задача прогнозирования (рис. 1) заключается в предсказании значения y_{n+k} в некоторый будущий момент времени t_{n+k} , где k – длительность прогноза:

$$y_{n+k} = F(y_1, y_2, \dots, y_n),$$

где F – функциональный преобразователь.

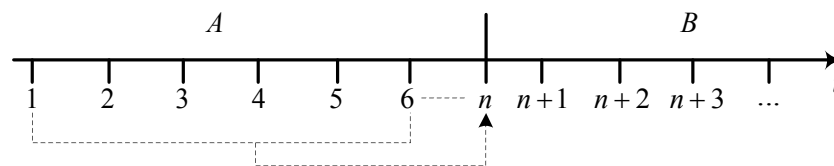


Рис. 1. Графическая иллюстрация постановки задачи прогнозирования: A – известные значения; B – прогнозируемый период

Обзор существующих методов прогнозирования. Перечислим основные методы прогнозирования.

Метод скользящего среднего [1]. Метод основывается на простой модели, которая предполагает, что текущее значение y_t , ($t=1 \dots n$) ряда $\{y_1, y_2, \dots, y_n\}$ является суммой среднего арифметического некоторого количества предыдущих значений и некоторой случайной составляющей: $y_t = C + \varepsilon_t$, $C = \frac{1}{p} \sum_{i=1}^p y_{t-i}$, где p – количество учитываемых предыдущих значений, так называемый «размер предыстории»; ε_t – случайная составляющая. Очень важным является знание таких характеристик ε_t , как закон распределения, математическое ожидание, дисперсия.

Метод взвешенного скользящего среднего [1]. Следующим шагом в модификации модели является предположение о том, что более поздние значения ряда более адекватно отражают ситуацию. Тогда каждому значению присваивается вес тем больший, чем более «свежее» значение добавляется:

$$y_t = C + \varepsilon_t,$$

$$C = \frac{\sum_{i=1}^p \alpha_i * y_{t-i}}{\sum_{i=1}^p \alpha_i},$$

где α_i – соответствующие веса значений.

Метод группового учета аргументов [2] – это набор алгоритмов прогнозирования (а точнее математического моделирования), который основывается на разбиении исходных данных на две выборки: обучающую и проверочную с использованием опорных функций некоторого вида, параметры которых находятся из обучающей выборки, а проверка того, насколько хорошо они моделируют заданный ряд, выполняется на проверочной выборке.

Искусственные нейронные сети (ИНС) [3]. Основным элементом нейронной сети – формальный (искусственный) нейрон. Он представляет собой математическую модель биологической нервной клетки.

Искусственные нейронные сети представляют собой систему соединённых и взаимодействующих между собой искусственных нейронов.

Искусственные нейронные сети не программируются в привычном смысле этого слова, они обучаются. В процессе обучения нейронная сеть способна выявлять сложные зависимости между входными и выходными данными, а также выполнять обобщение.

Способности нейронной сети к прогнозированию напрямую следуют из ее способности к обобщению и выделению скрытых зависимостей между входными и выходными данными. После обучения сеть способна предсказать будущее значение некой последовательности на основе нескольких предыдущих значений и/или каких-то существующих в настоящий момент факторов.

Главным преимуществом ИНС перед другими методами прогнозирования является то, что сети с одинаковым успехом могут прогнозировать процессы, регулярная составляющая которых имеет любой закон изменения значений, в то время как большинство остальных методов наиболее подходит для процессов, регулярная составляющая которых принадлежит к определенному классу (очевидно, что метод полиномиального сглаживания больше всего подходит для процессов с полиномиальной регулярной составляющей, метод сглаживания рядами Фурье – для процессов с периодической регулярной составляющей). Еще одним важным преимуществом нейронных сетей является возможность обучения.

Постановка задачи прогнозирования для нейронных сетей. Применительно к нейронным сетям входные данные будут те же: n дискретных отчетов $\{y_1, y_2, \dots, y_n\}$. Меняется лишь сама задача прогнозирования, поскольку общий вид функционального преобразователя F уже определен – будет использоваться нейронная сеть. Соответственно задача прогнозирования для ИНС состоит в построении сети – выборе топологии сети (сеть прямого распространения, рекуррентная сеть); определении структуры сети (количества скрытых слоёв, количества нейронов в слоях); определении параметров сети (вида активационной функции – сигмоида, пороговая функция, значений весовых коэффициентов и порогов); выборе алгоритма обучения сети (алгоритм обратного распространения ошибки с постоянным шагом обучения, с адаптивным шагом), а также в выборе критерия остановки обучения сети, при которых сеть предсказывала бы с максимальной точностью значение y_{n+k} .

Основные топологии нейронных сетей. Приведем основные существующие топологии нейронных сетей.

Многослойный персептрон (МП) [4]. Эта топология представляет собой сеть, состоящую из нескольких последовательно соединённых слоёв нейронов, причем каждый нейрон слоя i соединён со всеми нейронами слоя $i-1$, где i – номер слоя. Первый слой является входами сети, в процессе обучения или работы сети на него подаются исходные данные. За ним следует один либо несколько скрытых слоев, и последний слой – выходной слой, который представляет собой выходы сети.

Преимущества: наибольшая изученность, простота, устойчивость.

Недостатки: рассматривает обучающие примеры независимо друг от друга, тем самым упуская возможную связь между примерами.

Рекуррентные нейронные сети [4]. В отличие от персептрона в рекуррентных сетях связи между нейронами могут быть не только последовательными, т. е. от слоя с меньшим «порядковым номером» к слою с большим, но и обратными, в том числе связь может быть между нейроном и самим собой.

Наиболее популярным частным случаем рекуррентных сетей является сеть Эльмана [5]. Рекуррентная сеть Эльмана характеризуется частичной рекуррентностью в виде обратной связи между скрытым и входным слоем, реализуемой с помощью единичных элементов запаздывания.

Преимущества: рекуррентные ИНС учитывают возможную связь между примерами, как следствие лучше подходят для моделирования последовательностей; способность к обобщению большая, чем у МП.

Недостатки: неустойчивость, сложность, обучение в общем случае более медленное, чем у МП.

Радиально-базисные нейронные сети [6]. Искусственные нейронные сети, использующие в качестве активационных функций радиально-базисные (такие сети сокращённо называются RBF-сетями). Общий вид радиально-базисной функции:

$$f(x) = \varphi\left(\frac{x^2}{\sigma^2}\right), \text{ например, } f(x) = e^{-\frac{x^2}{\sigma^2}}, \text{ где } x - \text{ вектор входных сигналов нейрона; } \sigma - \text{ ширина окна функции; } \varphi(y) - \text{ убывающая функция (чаще всего равная нулю вне некоторого отрезка).}$$

Преимущества: большая скорость обучения, содержат всего один скрытый слой, выходной слой имеет линейную активационную функцию.

Недостатки: обладают плохими экстраполирующими свойствами, все связи между входами и нейронами скрытого слоя равны единице, т. е. все входные переменные имеют равную значимость при нахождении выходов сети.

Построение нейронной сети для задачи прогнозирования. Построение ИНС заключается в последовательном решении следующих задач:

- выбор топологии;
- структурный синтез ИНС: определение количества и типов (числовые, бинарные, нечеткие и др.) входов сети, количества скрытых слоев и количества нейронов в них, количества выходов сети (тип выходов совпадает с типом входов);
- параметрический синтез ИНС: выбор типа активационных функций для слоев (сигмоида, линейная, пороговая и т. д.), определение значений весовых коэффициентов;
- выбор алгоритма обучения и критерия остановки обучения.

Алгоритм решения задачи прогнозирования с помощью нейронных сетей:

1. Выбор топологии. В работах [7; 8] обосновывается выбор МП в качестве топологии сети для задачи прогнозирования. В работе [9] путем создания искусственных временных последовательностей было установлено, что если прогнозируемый процесс включает в себя МА (Moving Average – скользящее среднее) составляющую или мультипликативный шум, то рекуррентные нейронные сети намного лучше, чем МП справляются с задачей прогнозирования. Следует начать с выбора МП в качестве топологии сети, так как МП обучается быстрее и проще чем рекуррентные ИНС.

2. Структурный синтез ИНС:

– определение количества и типов входов сети. Тип входов сети зависит от конкретного прогнозируемого процесса и совпадает с типом значений процесса. Количество входов выбирают в зависимости от количества имеющихся точек в обучающей выборке и допустимой сложности сети. Очевидно, что чем больше входов имеет сеть, тем более сложные скрытые зависимости между входными данными она может находить, тем лучше

она может прогнозировать. Большее количество входов увеличивает сложность сети и уменьшает количество обучающих примеров. Поэтому следует начать с какого-то малого количества входов $p = 0,05 * n \dots 0,15 * n$, где n – количество точек в обучающей последовательности;

– определение количества скрытых слоев и количества нейронов в них. Не существует универсального правила выбора количества скрытых слоев и количества нейронов в них. Поэтому следует начать с количества скрытых слоев $t = 1$ и малого количества нейронов в единственном скрытом слое $s^{(1)} = 3 \dots 5$, где верхний индекс указывает на номер скрытого слоя;

– определение количества выходов сети. Количество выходов сети зависит от поставленной задачи. В случае задачи прогнозирования, когда требуется предсказать одно будущее значение y_{n+k} , выбирается один выход, который и будет прогнозировать это значение.

3. Параметрический синтез ИНС:

– выбор типа активационных функций слоев. В качестве активационной функции скрытых слоев обычно выбирают сигмоиду $f(u) = \frac{1}{1 + e^{-u}}$, поскольку она является наиболее распространенной активационной функцией. В качестве активационной функции выходного нейрона следует выбрать линейную функцию $f(u) = u$, так как он должен предсказывать нужное значение, которое в общем случае является действительным числом, а область значений сигмоиды лежит в границах $[0 \dots 1]$;

– определение значений весовых коэффициентов. Значения весовых коэффициентов определяется в ходе обучения сети.

4. Выбор алгоритма обучения и критерия останова обучения.

– выбор алгоритма обучения. Наиболее известным и изученным алгоритмом обучения нейронных сетей является алгоритм обратного распространения ошибки [10], а большинство остальных алгоритмов является его модификациями. Специально для оптимизации вычислений на вычислительных машинах были разработаны численно-оптимизированные методы обучения сети, например метод BFGS (Broyden–Fletcher–Goldfarb–Shanno method);

– выбор критерия останова обучения. В качестве критерия останова обычно принимают какое-то наперед заданное значение среднеквадратической ошибки (СКО) сети на проверочной либо на обучающей выборке. При обучении сети после каждой итерации следует вычислять значение СКО, и если оно меньше либо равно нужного значения, то следует прекратить обучение.

5. Обучение сети. Если после обучения сети её СКО является неудовлетворительной, то следует вернуться к пункту 2 и увеличить количество входов сети, после чего заново обучить сеть. Если же количество входов достигло 50 % от количества точек в обучающей последовательности и требуемая СКО не была достигнута, следует вернуться к количеству входов $p = 0,05 * n \dots 0,15 * n$ и начать увеличивать количество нейронов в скрытом слое, каждый раз выполняя обучение сети и находя значение СКО. Если количество нейронов в скрытом слое стало больше либо равно утроенному количеству входов $s^{(1)} \geq 3 * p$ и требуемая СКО не была достигнута, то следует вернуться к начальному количеству нейронов $s^{(1)} = 3 \dots 5$ и добавить еще один скрытый слой с количеством нейронов в нем $s^{(2)} = s^{(1)} = 3 \dots 5$, постепенно наращивая количество нейронов в слоях и количество входов, выполняя обучение сети после каждого изменения её структуры. Если в результате требуемая СКО так и не была достигнута, то либо данная топология сети не подходит для этой задачи прогнозирования и следует изменить топологию сети, либо для прогнозирования нужного

значения y_{n+k} ряда $\{y_1, y_2, \dots, y_n\}$ знания самих предыдущих значений недостаточно, и тогда следует искать дополнительные данные.

Обработка исходных данных. В работах [7 – 9] указана важность предварительной обработки исходных данных для получения лучшей модели прогнозирования на основе ИНС. Приведем наиболее используемые виды обработки исходных данных.

1. Поскольку большинство известных активационных функций имеют диапазон значений в пределах $[0...1]$ (либо $[-1...1]$), то нейронные сети хуже работают с большими входными значениями. Для того чтобы избежать этого, следует нормализовать входные данные:

$$x_t^n = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}},$$

где x_t, x_t^n – исходные и нормированные входные данные соответственно, $x_{\max}, x_{\min}$ – максимум и минимум входных данных соответственно.

2. В работе [9] показано, что нейронные сети хуже справляются с данными, где присутствуют резкие скачки значений (outliers), поэтому, если в результате обучения сеть будет иметь слишком большую ошибку на проверочной выборке, то стоит попробовать сгладить исходные данные (но поскольку в результате сглаживания существует риск утери важной информации, очень полезной может быть любая информация о характере шума, что может помочь в выборе алгоритма сглаживания и подборе его параметров). Больше про алгоритмы сглаживания исходных данных приведено в работе [9].

Пример использования предложенного алгоритма построения ИНС для задачи прогнозирования. Для проверки алгоритма построения ИНС на реальном примере был использован интернет-портал http://indiawaterportal.org/met_data/, который предоставляет базу данных по различным атмосферным показателям в Индии, откуда была взята информация по среднегодовой температуре района Индии с 1901 по 2002 год.

Параметры сети, выбранные последовательно, согласно приведенному алгоритму:

- тип сети – многослойный персептрон;
- тип входов – числовые;
- количество входов – 15;
- количество скрытых слоев – 1;
- количество нейронов в скрытом слое – 5;
- количество выходов – 1;
- активационная функция скрытого слоя – сигмоида, выходного слоя – линейная;
- алгоритм обучения сети – BFGS (Broyden–Fletcher–Goldfarb–Shanno method);
- критерий остановки обучения – ошибка сети на обучающей последовательности

должна быть около 0,01.

После обучения исходной сети требуемая СКО не была достигнута и увеличение количества входов не дало ощутимого результата. Поэтому согласно описанному алгоритму начали увеличивать количество нейронов в скрытом слое, и при количестве нейронов $s^{(1)} = 14$ была достигнута ошибка на обучающей выборке $E^{об} = 0,006513$. Среднеквадратическая ошибка сети на проверочной выборке составила $E^{пр} = 0,049353$. Примеры в проверочную выборку отбирались случайным образом с учетом количества входов нейронной сети p (т. е. значения $y_t, y_{t+1}, \dots, y_{t+p-1}$ вместе с ожидаемым значением y_{t+p} , где $t = 1 \dots n - p$; n – количество имеющихся точек, образуют один пример). В проверочную выборку было отобрано 30 % всех имеющихся примеров (соответственно в обучающую – 70 %).

Графический результат работы сети показан на рис. 2:

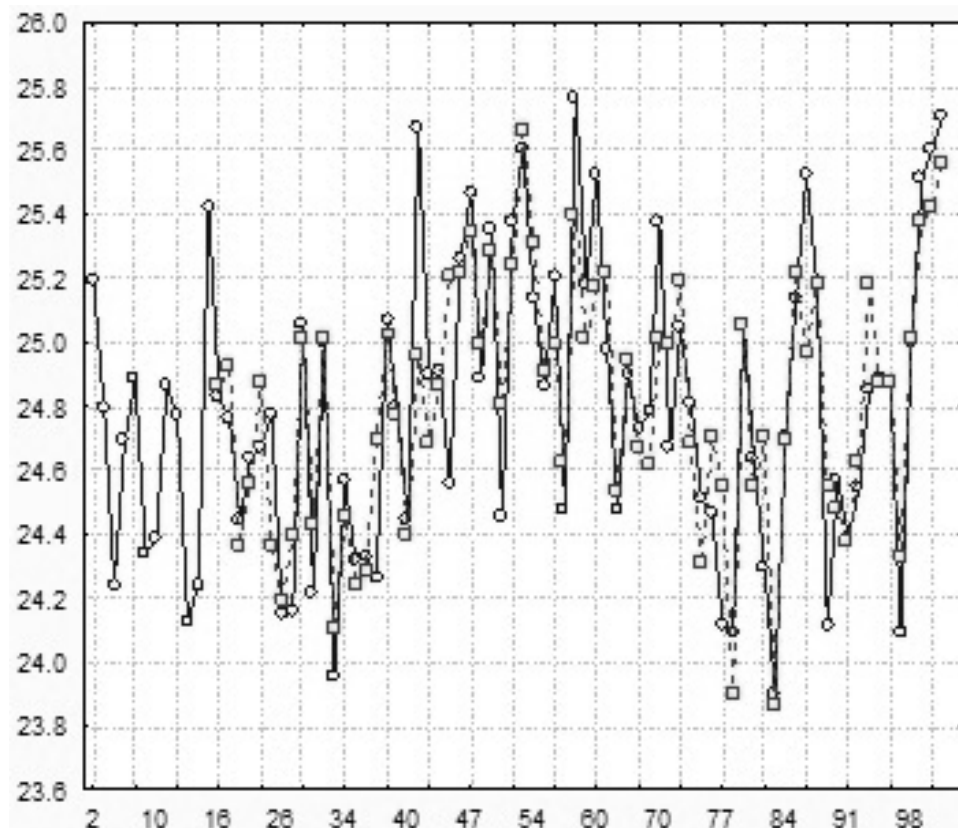


Рис. 2. Результат обучения ИНС, построенной по описанной методике: —○— — исходные данные; —□— — значения, предсказанные ИНС

Выводы. Как и ожидалось, сеть хуже справилась с большими скачками, поэтому один из возможных вариантов улучшения работы сети — это сглаживание этих скачков ([9]).

Характер изменения различных процессов почти всегда меняется с течением времени. Поэтому в отличие от задачи классификации, где большее количество разнообразных обучающих примеров обычно приводит к лучшей работе сети, в задаче прогнозирования более логичным может оказаться выбор в качестве обучающих точек лишь некоторого ограниченного количества последних значений $\{y_{n-k}, y_{n-k+1}, \dots, y_n\}$, даже при наличии большего количества точек, поскольку слишком «старые» значения могут лишь «навредить» и ухудшить прогнозирующие качества сети.

Список литературы

1. Алесинская Т. В. Методы скользящего среднего и экспоненциального сглаживания: учеб. пособие по решению задач по курсу «Экономико-математические методы и модели». / Т. В. Алесинская. — Таганрог: Изд-во ТРТУ, 2002. — 153 с.
2. Ивахненко А. Г. Метод группового учета аргументов — конкурент методу стохастичної апроксимації / А. Г. Ивахненко // Автоматика. — 1968. — № 3. — С. 58–72.
3. Мак-Каллок У. С. Логическое исчисление идей, относящихся к нервной активности / У. С. Мак-Каллок, В. Питтс // В сб.: «Автоматы»; под ред. К. Э. Шеннона и Дж. Маккарти. — М.: Изд-во иностр. лит., 1956. — С. 363–384.
4. Розенблатт Ф. Принципы нейродинамики: Перцептроны и теория механизмов мозга / Ф. Розенблатт // Principles of Neurodynamic: Perceptrons and the Theory of Brain Mechanisms. — М.: Мир, 1965. — 480 с.

5. *Elman J. L.* Finding structure in time / J. L. Elman // Cognitive Science. – 1990. – P. 179–211.
6. *Круглов В. В.* Сети с радиальными базисными элементами (RBF) / В. В. Круглов, В. В. Борисов // Нечеткая логика и искусственные нейронные сети. – М.: Изд-во Физматлит, 2001. – 221 с.
7. *Amir F. A.* A comparison between neural-network forecasting techniques / F. A. Amir, I. S. Samir – case study: river flow forecasting. // IEEE Transactions on neural networks. – Vol. 10, No. 2. – 1999. – P. 402–409.
8. *Mohsen H.* Artificial neural network approach for short term load forecasting for Illam region / H. Mohsen, S. Yazdan // World Academy of Science, Engineering and Technology 28 2007. – P. 280–284.
9. *Jerome T. C.* Recurrent neural networks and robust time series prediction / T. C. Jerome, R. M. Douglas, L. E. Atlas // IEEE transactions on neural networks. – Vol. 5, No. 2. – 1994. – P. 240–254.
10. *Галушкин А. И.* Синтез многослойных систем распознавания образов / А. И. Галушкин. – М.: Энергия, 1974, – 368 с.

О. І. Чумаченко, В. С. Горбатюк

Використання штучних нейронних мереж для задачі прогнозування

Запропоновано алгоритм побудови штучних нейронних мереж для розв'язання задачі прогнозування, який дозволяє автоматизувати процес побудови нейронної мережі.

H. I. Chumachenko, V. S. Horbatiuk

The use of artificial neural networks for the forecasting task

The algorithm of artificial neural networks construction was offered to solve the forecasting task, which enables the process of neural network construction automatization.