

УДК 33

DOI: 10.34670/AR.2021.30.36.017

Вопросы эффективности обучения нейронных сетей при анализе временных рядов

Губарева Елена Алексеевна

Кандидат физико-математических наук, доцент,
кафедра математики и информатики,
Государственный университет управления,
109542, Российская Федерация, Москва, просп. Рязанский, 99;
e-mail: gubel@inbox.ru

Хашин Сергей Иванович

Кандидат физико-математических наук, доцент,
кафедра информационных технологий и прикладной математики,
Ивановский государственный университет,
153025, Российская Федерация, Иваново, ул. Ермака, 39;
e-mail: khash2@mail.ru

Аннотация

Ряд теорем математики (Такенса, Горбаня, Колмогорова и др.) говорят о возможности сведения задачи прогноза временных рядов к задаче аппроксимации непрерывной функции нескольких переменных. Для построения такой модели можно использовать искусственные нейронные сети. Искусственные нейронные сети (ИНС) предлагается в основном использовать для решения таких задач экономики и управления, где формализация этапов решения или достаточно сложная, или вообще является невозможной. Обсуждается вопрос улучшения эффективности обучения искусственных нейронных сетей, используемых для построения математической модели с целью прогнозирования поведения динамической системы. Правильный выбор характеристик ИНС позволяет качественно решить поставленную задачу прогнозирования временного ряда. Основной этап – это подбор и обучение нейронной сети, который требует оценки всех возможных архитектур и всестороннего анализа существующих алгоритмов обучения нейронных сетей, чтобы выбрать оптимальный. В статье рассматриваются вопросы обучения ИНС по блокам и предложена тактика выбора размера блока и скорости обучения ИНС.

Для цитирования в научных исследованиях

Губарева Е.А., Хашин С.И. Вопросы эффективности обучения нейронных сетей при анализе временных рядов // Экономика: вчера, сегодня, завтра. 2021. Том 11. № 9А. С. 121-127. DOI: 10.34670/AR.2021.30.36.017

Ключевые слова

Временной ряд, нейронные сети, обучение ИНС, обучение по блокам, скорость обучения.

Введение

Современное развитие математики и информационных технологий дает широкий спектр возможностей для решения тех задач в области экономики и управления, где формализация этапов решения либо достаточно сложная, либо вообще является невозможной. Сегодня в самых различных областях жизнедеятельности [Гафаров, Галимянов, 2018; Ширяев, 2009; Николенко, Кадури, Архангельская, 2018] активно используются искусственные нейронные сети (ИНС), в том числе на их основе разрабатываются новые подходы к исследованию динамических задач в области экономики и управления [Ширяев, 2009; Gubareva, Shemyakova, 2021]. Хорошо обученная на ограниченном числе данных ИНС может показывать вполне достоверные результаты для данных, которые не использовались в процессе обучения.

Основная часть

Как известно, анализ временного ряда, порожденного некоторой динамической системой, проводится с целью предсказания значений его последующих состояний X_t на основе имеющихся данных о его предшествующих состояниях: $X_{t-1}, X_{t-2} \dots$. Классический подход к решению этой задачи состоит в использовании статистических методов, значительную часть которых представляют методы регрессионного анализа (теории наименьших квадратов) или их модификации и аналоги [Андерсон, 1976]. Вопрос же возможности использования ИНС для прогнозирования временного ряда рассматривался уже со второй половины XX века.

В частности, было доказано (теорема Такенса) [Takens, 1981], что для любого временного ряда существует «глубина погружения» k . То есть существует такая функция $\varphi(\cdot)$ нескольких переменных, которая явно не зависит от t , что на ее основании обеспечивается однозначность предсказания:

$$X_{t+1} = \varphi(X_t, \dots, X_{t-k+1}) \quad (1)$$

Согласно этой теореме, задачу прогнозирования временного ряда можно свести к задаче наилучшей аппроксимации нелинейной функции $\varphi(\cdot)$ от нескольких переменных. Для аппроксимации используются наборы данных, которые получают в результате проведения тех или иных измерений (анализа истории временного ряда).

Нелинейная модель, которую используют для решения задачи прогнозирования временного ряда $X_t, t = 0, 1, \dots$ имеет вид:

$$\tilde{X}_{t+1} = f(X_t, \dots, X_{t-k+1}) \quad (2)$$

Здесь функция $f(\cdot)$ – некоторая нелинейная функция нескольких переменных, $\tilde{X}_{t+1}$ – прогнозируемое значение ряда, $X_t, \dots, X_{t-k+1}$ – наблюдаемые значения, а k – порядок модели (глубина погружения).

ИНС, которая будет использоваться для прогнозирования временных рядов, должна

представлять собой некоторую функцию f от двух векторных аргументов: $f(X, W)$, где $X \in \mathbb{R}^n$ – входные параметры, и $W \in \mathbb{R}^m$ – внутренние параметры (веса).

В процесс обучения нейронной сети, проводимого с наличием известного для каждого входного вектора X_i целевого вектора Y_i , который ожидается получить на выходе, вектор внутренних параметров W подбирается таким образом, чтобы для всех векторов-строк X_i обучающей матрицы X выполнялось условие:

$$y_i \approx f(X_i, W) \quad (6)$$

Мерой погрешности приближения чаще всего служит функция:

$$S(W) = \sum_i (f(X_i, W) - y_i)^2 \quad (7)$$

которая должна принимать наименьшее возможное значение.

Теорема Горбаня [Горбань и др., 1998] показывает, что если в качестве функции активации нейронов выбрана нелинейная дважды непрерывно дифференцируемая функция, то любая непрерывная на замкнутом ограниченном множестве функция может быть равномерно приближена функциями, которые вычисляет ИНС.

Для активации нейронов часто используется логистическая функция (сигмоид):

$$g(z) = \frac{1}{1 + e^{-\alpha z}} \quad (3)$$

которая слабые сигналы усиливает сильнее, чем большие, и имеет простое выражение для своей производной:

$$g'(z) = \alpha \cdot g(z) \cdot (1 - g(z)) \quad (4)$$

Векторы X_i обучающей матрицы X последовательно подаются на вход ИНС, которая вычисляет соответствующий выходной вектор $f(X_i, W)$, сравнивает его с целевым вектором Y_i и вычисляет ошибку. Если ошибка не удовлетворяет выбранному критерию погрешности, то значения внутренних параметров подстраиваются за счет применения алгоритма минимизации функции ошибки $S(W)$. Так как внутренние параметры ИНС настраиваются таким образом, чтобы ошибка по всему обучающему массиву X соответствовала требуемому уровню, то один и тот же входной вектор X_i может с некоторой периодичностью подаваться на вход столько раз, сколько потребуется для работы алгоритма минимизации.

Для минимизации функции ошибки применяются различные методы. Например, в известном пакете нейросетевых вычислений TensorFlow [Пакет нейросетевых вычислений TensorFlow, [www](http://www.tensorflow.org)] стандартно используются десять методов (`tf.train.GradientDescentOptimizer`,

tf.train.AdadeltaOptimizer, tf.train.FtrlOptimizer и др.), которые все являются разновидностями или модификациями базового алгоритма градиентного спуска (GradientDescentOptimizer). Один шаг метода градиентного спуска (GradientDescentOptimizer) с данной скоростью обучения (LR)

означает переход от вектора внутренних параметров W_0 к W_1 :

$$W_0 \rightarrow W_1 = W_0 - LR \frac{\nabla S}{|\nabla S|} \quad (5)$$

где ∇S – градиент функции $S(W)$.

На этапе обучения возникает достаточное количество проблем различного рода: обучение длится очень долго или останавливается, не достигнув результата; в результате обучения был достигнут локальный, а не глобальный минимум функции ошибки и др. Чтобы избежать подобных ситуаций, необходимо уделять особое внимание выбору начального значения

внутренних параметров (весов) W_0 , применять случайное упорядочивание обучающей последовательности векторов, обратить внимание на выбор скорости обучения.

Использование искусственных нейронных сетей не вытесняет традиционные методы, а позволяет расширить круг решаемых задач. При этом качественное решение поставленной задачи в области анализа временных рядов с использованием ИНС возможно только при условии правильного выбора важнейших свойств нейронной сети: ее архитектура, размер блоков, скорость обучения и др. Этот вопрос до сих пор остается в основном эмпирическим, что является сдерживающим фактором использования ИНС для решения практических задач.

В нашей работе [Губарева, 2020] была предложена схема поэтапного построения ИНС, которая определяет основные характеристики ИНС, правильный выбор которых позволит качественно решить поставленную задачу прогнозирования временного ряда. Обращает на себя внимание тот факт, что основные части процедуры – это подбор и обучение нейронной сети: подготовка собранных данных, а именно их случайное упорядочивание, разбиение при необходимости на блоки и др., то есть проведение процедур, которые могли бы улучшить (и ускорить) процесс обучения; выбор архитектуры нейронной сети (количество входных нейронов, обусловленное, в частности, «глубиной погружения», многослойность, количество нейронов в слое, прямое распространение или двунаправленное распространение, связность и др.).

На этих этапах требуется оценка всех возможных архитектур нейронных сетей и способов их обучения. В случае обучения нейронной сети по блокам правильный выбор размера блока и величины скорости обучения позволит улучшить качество обучения.

В ряде работ, посвященных обучению нейронных сетей [Гафаров, Галимянов, 2018; Nasonov, Krylov, Petrova, Rychagov, 2016], предлагается устанавливать небольшой размер блока (обычно несколько десятков). Такая тактика на начальных эпохах обучения дает сеть гораздо лучшую, чем при использовании больших блоков на том же количестве начальных эпох.

В некоторых работах [Smith et al.] в результате сравнения эффектов от уменьшения скорости обучения и от увеличения размера блока был сделан вывод, что при выполнении определенных условий лучше сохранить скорость обучения, но размер блока при этом лучше увеличить.

Для некоторого класса нейронных сетей точность не ухудшается, а вычисления заметно ускоряются, если увеличить размер блока с нескольких десятков до нескольких тысяч и даже

десятков тысяч [Yang You, Gitman, Ginsburg, 2017].

Целью наших исследований было показать, что размер блока лучше не фиксировать, а динамически увеличивать, начиная с нескольких десятков и вплоть до размера блока, равного всей имеющейся обучающей выборке.

Результаты, изложенные в работе, были получены при проведении численных экспериментов обучения ИНС, которую планировалось использовать для прогнозирования исследуемых временных рядов.

Для проведения эксперимента с интернет-источника были взяты данные по ценам на нефть сортов Brent и WTI от EIA U.S. (Energy Information Administration) начиная с 20.05.1987 и по 24.08.2020 включительно [GitHub: Oil-prices, www]. Поскольку ежедневные данные слишком сильно колеблются (отражают зависимость от многих случайных факторов) и приведены не за все дни подряд (нет данных за выходные и праздничные дни), то для обучения ИНС были использованы цены, усредненные по неделям (в результате получили данные за 1737 недель).

Прогноз цены на нефть на следующую неделю строится на основе данных о стоимости цены за несколько предыдущих недель. Выбирая глубину погружения временного ряда в диапазоне от 5 до 15 недель, получали обучающую матрицу из 1730 строк (примерно) и 6-16 столбцов соответственно.

Для аппроксимации использовалась трехслойная нейронная сеть с 5-15 входными значениями (зависело от выбранной глубины погружения) и с функцией активации

$$f(x) = \frac{1}{1 + e^{-x}}$$

в первых двух слоях и тривиальной функцией $f(x) = x$ в выходном слое.

Заключение

На основе проведенных экспериментов можно предложить следующую тактику. В начале обучения нейронной сети, при случайном выборе начального приближения, можно рекомендовать выбрать блок минимального размера. Но при этом следует учесть, что вычисления в каждом блоке сопровождаются определенными накладными расходами и слишком маленький размер блока тоже будет неэффективен. Оптимальный начальный размер блока должен быть от нескольких десятков до сотен, а начальная скорость обучения должна быть как можно больше. Если предполагать, что все данные, на которых происходит обучение, имеют порядок единицы, то LR лучше всего взять порядка 0.1^m , где m – размерность пространства, количество обучаемых параметров.

В процессе обучения размер блока рекомендуется увеличивать, а скорость обучения при этом уменьшать. Скорость обучения следует снизить, оптимально – уменьшить в два раза, если в течение нескольких (3-4) шагов обучения целевая функция не уменьшилась.

Для контроля за размером блока следует на каждом шаге вычислять косинус угла между текущим направлением смещения и предыдущим. Если за последние несколько (4-8) шагов обучения среднее значение косинуса угла меньше 0,5, то размер блока следует увеличить, оптимально – вдвое. Разумеется, каждый шаг после этого будет занимать больше времени, но медленное продвижение в верном направлении предпочтительней, чем быстрое и хаотичное в неверном или просто случайном.

Стоит обратить внимание на то, что обучающие выборки могут оказаться столь велики, что не поместятся в оперативную память. В этом случае предельный размер блока должен определяться лишь доступным объемом памяти.

Библиография

1. Андерсон Т. Статистический анализ временных рядов. М.: Мир, 1976. 756 с.
2. Гафаров Ф.М., Галимянов А.Ф. Искусственные нейронные сети и приложения. Казань: Изд-во Казанского ун-та, 2018. 121 с.
3. Горбань А.Н. и др. Нейроинформатика. Новосибирск: Наука, 1998. 296 с.
4. Губарева Е.А. Нейронные сети в анализе временных рядов // Инновации и инвестиции. 2020. № 10. С. 150-153.
5. Николенко С., Кадурин А., Архангельская Е. Глубокое обучение. Погружение в мир нейронных сетей. СПб, Питер, 2018. 480 с.
6. Пакет нейросетевых вычислений TensorFlow. URL: <https://www.tensorflow.org>.
7. Ширяев В.И. Финансовые рынки. Нейронные сети. Хаос и нелинейная динамика. М.: Либроком, 2009. 232 с.
8. GitHub: Oil-prices. URL: <https://github.com/datasets/oil-prices>.
9. Gubareva E.A., Shemyakova E. Neural networks in the digital economy // In Popkova E.G., Ostrovskaya V.N., Bogoviz A.V. (eds.) Socio-economic Systems: Paradigms for the Future. Studies in Systems, Decision and Control, 314. Cham: Springer, 2021. P. 827-835.
10. Nasonov A.V., Krylov A.S., Petrova X., Rychagov M.N. Edge-directional interpolation algorithm using structure tensor // Electronic Imaging, Image Processing: Algorithms and Systems XIV – Society for Imaging Science and Technology. 2016. P. 1-4.
11. Smith S.L. et al. Don't Decay the Learning Rate, Increase the Batch Size arXiv:1711.00489.
12. Takens T. Detecting strange attractors in turbulence // Lec. Notes in Math. 1981.
13. Yang You, Gitman I., Ginsburg B. Scaling SGD batch size to 32k for imagenet training. 2017. arXiv:1708.03888.

Questions of efficiency of learning neural networks in analysis of time series

Elena A. Gubareva

PhD in Physical and Mathematical Sciences, Associate Professor,
Department of mathematics and informatics,
State University of Management,
109542, 99 Ryazanskii ave., Moscow, Russian Federation;
e-mail: gubel@inbox.ru

Sergei I. Khashin

PhD in Physical and Mathematical Sciences, Associate Professor,
Department of information technology and applied mathematics,
Ivanovo State University,
153025, 39 Ermaka str., Ivanovo, Russian Federation;
e-mail: khash2@mail.ru

Abstract

A number of theorems of mathematics (Takens, Gorbanya, Kolmogorov, etc.) speak of the possibility of reducing the problem of forecasting time series to the problem of approximating a continuous function of several variables. Artificial neural networks can be used to build such a model. Artificial neural networks (ANNs) are proposed to be mainly used for solving such problems of economics and management, where the formalization of the solution stages is either quite complicated or even impossible. The issue of improving the efficiency of teaching artificial neural networks used to construct a mathematical model in order to predict the behavior of a dynamic system is discussed. The correct choice of ANN characteristics allows us to qualitatively solve the

set problem of forecasting the time series. The main stage is the selection and training of a neural network, which requires an assessment of all possible architectures and a comprehensive analysis of existing neural network training algorithms in order to choose the optimal one. The article discusses the issues of teaching ANN in blocks and suggests a tactic for choosing the block size and learning rate of the ANN.

For citation

Gubareva E.A., Khashin S.I. (2021) Voprosy effektivnosti obucheniya neironnykh setei pri analize vremennykh ryadov [Questions of efficiency of learning neural networks in analysis of time series]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 11 (9A), pp. 121-127. DOI: 10.34670/AR.2021.30.36.017

Keywords

Time series, neural networks, ANN training, block learning, learning rate.

References

1. Anderson T. (1976) *Statisticheskii analiz vremennykh ryadov* [Statistical analysis of time series]. Moscow: Mir Publ.
2. Gafarov F.M., Galimyanov A.F. (2018) *Iskusstvennye neironnye seti i prilozheniya* [Artificial neural networks and applications]. Kazan': Publishing house of Kazan University.
3. *GitHub: Oil-prices*. Available at: <https://github.com/datasets/oil-prices> [Accessed 12/09/2021].
4. Gorban' A.N. et al. (1998) *Neiroinformatika* [Neuroinformatics]. Novosibirsk: Nauka Publ.
5. Gubareva E.A. (2020) Neironnye seti v analize vremennykh ryadov [Neural networks in time series analysis]. *Innovatsii i investitsii* [Innovations and investments], 10, pp. 150-153.
6. Gubareva E.A., Shemyakova E. (2021) Neural networks in the digital economy. In Popkova E.G., Ostrovskaya V.N., Bogoviz A.V. (eds.) *Socio-economic Systems: Paradigms for the Future. Studies in Systems, Decision and Control*, 314. Cham: Springer, pp. 827-835.
7. Nasonov A.V., Krylov A.S., Petrova X., Rychagov M.N. (2016) Edge-directional interpolation algorithm using structure tensor. *Electronic Imaging, Image Processing: Algorithms and Systems XIV – Society for Imaging Science and Technology*, pp. 1-4.
8. Nikolenko S., Kadurin A., Arkhangel'skaya E. (2018) *Glubokoe obuchenie. Pogruzhenie v mir neironnykh setei* [Deep learning. Immersion in the world of neural networks]. Saint Petersburg: Piter Publ.
9. *Paket neirosetevykh vychislenii TensorFlow* [The package of neural network computing TensorFlow]. Available at: <https://www.tensorflow.org> [Accessed 18/09/2021].
10. Shiryaev V.I. (2009) *Finansovye rynki. Neironnye seti. Khaos i nelineinaya dinamika* [Financial market. Neural network. Chaos and nonlinear dynamics]. Moscow: Librokom Publ.
11. Smith S.L. et al. *Don't Decay the Learning Rate, Increase the Batch Size* *arXiv:1711.00489*.
12. Takens T. (1981) Detecting strang attractors in turbulence. *Lec. Notes in Math*.
13. Yang You, Gitman I., Ginsburg B. (2017) *Scaling SGD batch size to 32k for imagenet training*. *arXiv:1708.03888*.