

УДК 517.9

MSC 47J20, 49J40, 65K15, 90C25

Training of Artificial Neural Network

I.Sh. Didmanidze, G.A. Kakhiani, D.Z. Didmanidze

Batumi Shota Rustaveli state university, Batumi, Georgia

E-mail: ibraimd@mail.ru, g.kakhiani@bsu.edu.ge

The methodology of neural networks is even more often applied in tasks of management and decision-making, including in the sphere of trade and finance. The basis of neural networks is made by nonlinear adaptive systems which proved the efficiency at the solution of problems of forecasting.

Key words: Artificial neural network, nonlinear adaptive systems, synchronization of neural networks, neural network.

Обучение искусственной нейронной сети

И.Ш. Дидманидзе, Г.А. Кахиани, Д.З. Дидманидзе

Батумский государственный университет Щота Руставели, Батуми, Грузия

E-mail: ibraimd@mail.ru, g.kakhiani@bsu.edu.ge

Методология нейронных сетей всё чаще применяется в задачах управления и принятия решений, в том числе в сфере торговли и финансов. Основу нейронных сетей составляют нелинейные адаптивные системы, которые доказали свою эффективность при решении задач прогнозирования.

Ключевые слова: искусственная нейронная сеть, нелинейные адаптивные системы, синхронизация нейронных сетей, нейронная сеть.

1. Введение

Эффективность обучения многослойных нейронных сетей связана с количеством слоев, количеством нейронов в скрытых слоях и начальным уровнем инициализации синоптических коэффициентами между нейронами.

Чем больше количество нейронов в скрытых слоях, тем больше точность работы сети, однако слишком большое количество нейронов в скрытых слоях может привести к так называемый, переобучению нейронной сети. Делается следующие выводы:

1. Нейронная сеть с одним скрытым слоем в состоянии отобразить любой входной вектор в любой выходной
2. Количество нейронов в скрытом слое должен быть меньше набора объема данных, предназначенных для обучения нейронной сети
3. Мощность нейронной сети может быть увеличена как за счёт увеличения количества скрытых слоев, так и за счёт количества самих нейронов в скрытых слоях
4. Начальная инициализация синоптических связей внутри нейронной сети должна производиться в достаточно узком диапазоне значений.

2. Основная часть

Поведение искусственной нейронной сети зависит не только от коэффициентов синаптических связей, но также и от вида функции активации (возбуждения) нейрона. Известны три основные функции активации: единичный скачок или жесткая пороговая функция, линейный порог или гистерезис, сигмоидальная функция или сигмоид. При использовании линейного порога выходное значение нейрона устанавливается в одно из двух положений в зависимости от входящих данных. При использовании линейного порога выходное значение нейрона пропорциональна входному значению. В случае с сигмоидальной функцией изменения выходного сигнала происходит пропорционально входному, однако эта зависимость нелинейна. Нейронные сети, предназначенные для решения конкретной задачи свои свойства, получают в результате обучения. Обучение подразумевает или определение коэффициентов синаптических связей или определение особых правил, согласно которым и будут изменяться синаптические связи в зависимости от реакции тех или иных нейронов в сети. Обучение сети производится с целью выработать желаемое (адекватное) множество реакций нейронной сети на множество входных данных сети. Каждое такое входящее или исходящее множество рассматривается как вектор. Обучение происходит путём подачи на вход каждого множества и выработки соответствующих синаптических связей согласно заранее выработанным правилам. Нейронные сети могут быть обучены одновременно, в этом случае имеет место синхронизация нейронных сетей. Различают два алгоритма обучения нейронной сети: обучение с учителем и обучение без учителя. При обучении с учителем подразумевается, что для каждого входящего вектора имеется заранее определенный выходящий вектор (целевой) значений и их вместе называют обучающим примером. Обычно процесс обучения происходит на определенном наборе обучающих примеров. Сети подаётся входной вектор, вычисляется исходящий вектор и происходит его сравнение с целевым вектором, разница (ошибка) посредством обратной связи возвращается сети и синаптические связи меняются с целью минимизации возвращенной ошибки. Примеры подаются последовательно один за другим пока суммарная ошибка сети по всем примерам не будет меньше заранее определенного порогового значения. Примерами алгоритмов обучения с учителем являются: алгоритм обучения персептрона (модель Розенблатта) и алгоритм обратного распространения ошибки.

Рассмотрим алгоритм обратного распространения ошибки. Согласно этому алгоритму на каждой итерации процесса обучения сети скорость обучения каждого слоя должна быть подобрана таким образом чтобы минимизировать среднеквадратическую ошибку сети:

$$\min E(y_j(t+1)),$$

где $j = \overline{1, m}$, m – количество нейронов в последнем слое. Выходное значение j -го нейрона зависит от функций активации и в общем случае определяется следующим образом:

$$w_{ij}(t+1) = w_{ij}(t) - a(t) \frac{\partial E}{\partial w_{ij}}, \quad (1)$$

$$T_j(t+1) = T_j(t) - a(t) \frac{\partial E}{\partial T_j(t)}. \quad (2)$$

Среднеквадратическую ошибку сети можно выразить следующим образом:

$$E = \frac{1}{2} \sum_j (y_j - t_j)^2. \text{ Тогда для отыскания } a(t) \text{ необходимо решить следующее}$$

$$\text{уравнение: } \frac{\partial E}{\partial a(t)} = \frac{\partial E}{\partial y_j(t+1)} \frac{\partial y_j(t+1)}{\partial a(t)}. \text{ В ряде работ для поиска адаптивной скоро-}$$

сти обучения предлагается использование метода линейного спуска. Однако этот подход связан с большим объемом вычислений. Поэтому предлагаем воспользоваться приближительным методом нахождения $a(t)$, который основывается на разложении в ряд Тейлора функции активации нейрона

Пусть выходное значение j -го нейрона последнего слоя вычисляется следующим образом:

$$y_j(t) = F(S_j(t)), \quad S_j(t) = \sum_i y_i(t) w_{ij}(t) - T_j(t), \quad (3)$$

где $y_i(t)$ – значение i -го нейрона скрытого слоя. Для поиска значения j -го нейрона в $t+1$ момент времени в (3) подставим (1) и (2):

$$S_j(t+1) = \sum_i y_i(t) \left(w_{ij} - a \frac{\partial E}{\partial w_{ij}} \right) - T_j + a \frac{\partial E}{\partial T_j} = \sum_i y_i w_{ij} - T_j + a \left(\sum_j y_j \frac{\partial E}{\partial w_{ij}} - \frac{\partial E}{\partial T_j} \right). \quad (4)$$

Введем следующие обозначения:

$$a_j = \sum_j y_j \frac{\partial E}{\partial w_{ij}} - \frac{\partial E}{\partial T_j}. \quad (5)$$

Тогда (4) можно переписать следующим образом:

$$S_j(t+1) = S_j(t) - a a_j. \quad (6)$$

Для $t+1$ момента времени выходное значение j -го нейрона будет равно:

$$y_j(t+1) = F(S_j(t+1)).$$

Раскроем последнее выражение в ряд Тейлора с точностью первых двух членов:

$$y_j(t+1) = F(0) + F'(0) F(S_j(t+1)), \quad (7)$$

где $F'(0) = \frac{\partial F}{\partial S_j}$, когда $S_j = 0$. С учётом (7) в (6) получим:

$$y_j(t) = F(0) + F'(0) S_j(t) - a F'(0) a_j. \quad (8)$$

Так как $y_j(t) = F(0) + F'(0) S_j(t)$, тогда (8) можно представить следующим образом:

$$y_j(t+1) = y_j(t) F(0) - a F'(0) a_j. \quad (9)$$

Для адаптивной скорости обучения следует обеспечить выполнение следующего условия:

$$E = \frac{1}{2} \sum_j (y_j(t+1) - t_j)^2 \rightarrow \min.$$

Тогда $\frac{\partial E}{\partial a} = \sum_j (y_j(t) - t_j - aF'(0)a_j)(-F'(0)a_j) = 0$. Из последнего выражения отыщем $a(t)$, получим:

$$a(t) = \frac{\sum_j (y_j(t) - t_j) a_j}{F'(0) \sum_j a_j^2}. \quad (10)$$

Так как $\frac{\partial^2 E}{\partial a^2} > 0$, тогда для $a(t)$ – достигается минимум среднеквадратической ошибки. Отыщем выражение для a_j . Для этого введем определение:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial S_j} \frac{\partial S_j}{\partial w_{ij}} = (y_j - t_j) F'(S_j) y_j, \quad (11)$$

$$\frac{\partial E}{\partial T_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial S_j} \frac{\partial S_j}{\partial T_j} = -(y_j - t_j) F'(S_j). \quad (12)$$

Если (11) и (12) подставить в (5) получим:

$$a_j = \left(1 + \sum_i y_i^2\right) (y_j - t_j) F'(S_j). \quad (13)$$

Исходя из принципа независимости слоёв, предположим, что:

$$\gamma_j = y_j - t_j. \quad (14)$$

Если учесть (14) и (13) в (10) получим выражение для расчета примерной скорости адаптивного обучения нейронной сети:

$$a(t) = \frac{\sum_j \gamma_j^2 F'(S_j)}{F'(0) \left(1 + \sum_i y_i^2\right) \sum_j \gamma_j^2 \left(F'(S_j)\right)^2}, \quad (15)$$

где γ_j – ошибка j -го нейронного элемента сети.

3. Заключение

Следует отметить, что для каждого слоя сети $a(t)$ рассчитывается отдельно. Как показали численные эксперименты в случае применения адаптивной скорости обучения сети $a(t)$ может принимать большие значения. Это может привести к десинхронизации процесса обучения сети, когда весовые коэффициенты нейронной сети резко меняются в одном направлении. В результате чего изменения среднеквадратической ошибки по времени получит колебательный характер. Поэтому рекомендуется ограничивать значение $a(t)$ по его абсолютному значению.

Полученные выше выражения для адаптивной скорости обучения позволяют ощутимо повысить скорость обучения нейронной сети и обойти проблему подбора скорости обучения произвольным образом. Это является ощутимым преимуществом по сравнению со стандартным алгоритмом обратного распространения ошибки. Хотя при удачном подборе скорости обучения сети время работы данного алгоритма не превышает время работы стандартного алгоритма градиентного спуска.

ЛИТЕРАТУРА

1. Didmanidze Ibraim, Kakhiani Gregori, Shatashvili Tamara, Dumbadze Zurab. The process of learning in artificial neural networks. XXXV international conference “Problems of decision making under uncertainties” (PDMU-2020). Abstracts. May 11-15, 2020, Baku-Sheki, Republic of Azerbaijan, p. 36.
2. Дидманідзе І.Ш., Кахіані Г.А., Дидманідзе Д.З. Ефективність стратегії управління капиталом в торгівлі на біржі. матеріали II Міжнародної научно-практичної конференції «Актуальні проблеми розвитку ресторанного, гостиничного і туристичного бізнесу в умовах глобальної інтеграції: досягнення і перспективи». 19-20 вересня 2019 року. Харків–Варна. ХДУХТ 2019. С.194–197.
3. Didmanidze I., Kakhiani G., Didmanidze D., Dumbadze Z. Neural network’s modeling problem. XXXIII International conference “Problems of decision making under uncertainties” (PDMU-2019). Abstracts. January 24 – February 1, 2019, Hurgada, Egypt. P. 35.
4. Didmanidze Ibraim, Kakhiani Grigol, Kakhidze Revaz. Dumbadze Zurab. Securities portfolio management through artificial neural networks. Матеріали 6-ої міжнародної конференції «Математичне моделювання, оптимізація інформаційні технології» (ММОТИ-2018). 2018 г. Кишинь, 12-16 листопада 2018 г. С. 73–75.
5. Didmanidze I., Shatashvili A., Kakhiani G., Fomina T., Dumbadze Z. Selection of adaptive artificial neural networks learning speed. XXXI International conference “Problems of decision making under uncertainties” (PDMU-2018). Abstracts. July 3-8, 2018, Lankaran-Baku, Republic of Azerbaijan. P. 53.
6. Didmanidze I., Kakhiani G., Didmanidze D., Dumbadze Z. Task of modeling the single line neural network’s. XXXII international conference “Problems of decision making under uncertainties” (PDMU-2018). Abstracts. August 27-31, 2018, Prague, Czech Republic. P. 38.
7. Головкин В.А. Нейронные сети: обучение, организация и применение. М.: ИПРЖР, 2001. 256 с.

Надійшла 18.05.2021.

Навчання штучної нейронної мережі
І.Ш. Дідманідзе, Г.А. Кахіані, Д.З. Дідманідзе
Батумський державний університет Шота Руставелі
E-mail: ibraimd@mail.ru, g.kakhiani@bsu.edu.ge

Методологія нейронних мереж все частіше застосовується в задачах управління та прийняття рішень, в тому числі в сфері торгівлі і фінансів. Основу нейронних мереж складають нелінійні адаптивні системи, які довели свою ефективність при вирішенні задач прогнозування.

Ключові слова: штучна нейронна мережа, нелінійні адаптивні системи, синхронізація нейронних мереж, нейронна мережа.