

УДК 519.95

УВЕЛИЧЕНИЕ СКОРОСТИ ОБУЧЕНИЯ ПРИ ИСПОЛЬЗОВАНИИ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ С ЖЕСТКИМ ЯДРОМ

© В.В. Максимушкин

Maximushkin V.V. Increase in the rate of teaching when using artificial neuron hard-nucleus nets. The article contains a detailed and demonstrative account of the issue.

При выборе метода решения любой задачи исследователь руководствуется двумя критериями: точностью, с которой данный метод может решить поставленную задачу, и временем, которое надо затратить, чтобы получить решение.

Очевидно, что быстрых и точных методов решения не существует. Всегда приходится искать компромисс между скоростью и точностью.

В случае искусственных нейронных сетей (ИНС) точность решения задачи определяется количеством слоев, количеством связей между нейронами и видами активационных функций каждого нейрона. Время, за которое может быть получено решение, определяется методом обучения.

Рассмотрим слоистую ИНС, в которой каждый нейрон связан с каждым нейроном предыдущего слоя и не имеет обратных связей. Возьмем также случай обучения с учителем.

Для рассматриваемого случая в настоящее время существуют достаточно много методов обучения. Перечислим основные.

1. Градиентные методы обучения.
2. Алгоритмы метода сопряженных градиентов.
3. Квазиньютоновские алгоритмы, использующие матрицу Гессе вторых частных производных.

Каждый метод имеет свои преимущества и недостатки и применяется в разных задачах [1].

Возникает вопрос, можно ли найти способ увеличить скорость обучения ИНС, не зависящий от метода обучения?

Предположим, необходимо построить ИНС-модель сложного объекта А, составной частью которого является объект А1, ИНС-модель которого уже создана. Определена структура модели, активационные функции нейронов и коэффициенты синаптических связей, при которых она показывает адекватные эксперименту результаты. Для разработки ИНС-модели объекта А представляется верным использовать уже имеющуюся модель объекта А1 в неизменном виде так, чтобы при дальнейших модификациях, проверке адекватности, выборе структуры и т. д. эта часть, представляющая собой «жесткое ядро», не подвергалась изменениям, а вновь сконструированная сеть обучалась бы лишь за счет выбора структуры, активационных функций нейронов и значений коэффициентов синаптических связей во вновь введенной части.

Главными достоинствами предлагаемой структуры ИНС является то, что обучение сети с «жестким ядром» будет происходить быстрее, чем в традиционном случае, так как при этом снижается размерность решаемой задачи минимизации (подстраивается лишь часть весовых коэффициентов, имеющихся во вновь введенных структурах) [2].

Время, затраченное на получение решения, нельзя оценивать непосредственно в единицах времени, так как процесс вычислений сильно зависит от производительности вычислительной техники, на которой проводится эксперимент. Более объективной будет оценка количества операций, которые необходимо совершить, чтобы обучить ИНС.

Попробуем получить количественное описание нашего предположения.

Рассмотрим произвольную искусственную нейронную сеть. Время обучения ИНС будет пропорционально следующему значению:

$$\tau \sim kN^m, \quad (1)$$

где k – количество проходов обучения, N – количество операций, требуемых для минимизации функции по одной переменной, m – количество переменных. В среднем N может принимать значения порядка 100, а m и k – порядка 10.

Предположим, что требуется создать ИНС для моделирования некоторого объекта А, который можно разбить на два более простых объекта А1 и А2.

Предположим так же, что полноценный объект А можно смоделировать при помощи ИНС с количеством связей между нейронами порядка 20. Два более простых объекта А1 и А2 можно было бы представить как две ИНС с количеством связей порядка 10 каждая.

Воспользуемся формулой (1) для оценки времени, требуемой для обучения классической нейронной сети и сети с жестким ядром. В первом случае время можно будет оценить так: $\tau \sim 10 \cdot 100^{20} = 10^{23}$.

В случае ИНС с жестким ядром затраченное время будет оцениваться как: $\tau \sim 10 \cdot 100^{10} + 10 \cdot 100^{10} = 20^{13}$.

Однако обучение ИНС с жестким ядром не сводится к простому соединению двух обученных сетей. Чтобы сеть давала более адекватные результаты, требуется провести обобщающее обучение.

Его можно осуществить двумя способами: 1) объединить два жестких ядра и провести классическое обучение; 2) объединить жесткие ядра и добавить новый «обобщающий» слой, за счет которого и провести окончательное обучение ИНС.

Оценим суммарное время в случае обобщения.

В первом случае в процессе обобщающего обучения предполагается, что N – количество итераций для минимизации одной переменной будет порядка 10, а не 100, как в начале. Это объясняется тем, что результирующая сеть и без обобщающего обучения должна давать приемлемые результаты, так что на окончательную подстройку не должно затрачиваться много времени.

Применив формулу (1), получаем:

$$\tau \sim 10 \cdot 100^{10} + 10 \cdot 100^{10} + 10 \cdot 10^{20} \approx 10^{21}.$$

В случае обобщающего слоя оценка времени будет иметь другой вид:

$$\tau \sim 10 \cdot 100^{10} + 10 \cdot 100^{10} + 10 \cdot 100^{10} = 30^{13}.$$

Очевидно, что самым выгодным является способ обучения ИНС с жестким ядром и с использованием обобщающего слоя. Но даже в случае обобщающего обучения по классической схеме мы получаем выигрыш в количестве операций.

Несложно записать наши выкладки в общем виде.

Пусть имеется ИНС, которую можно разделить на L жестких ядер. В результате общее время обучения можно будет оценить по формуле:

$$\tau \sim l \cdot k \cdot N^{m-l+1} + k \cdot N^l.$$

Зависимость общего времени обучения от количества жестких ядер демонстрируется на рис. 1.

Отсюда видно, что оптимальное число разбиений, при котором общее время обучения минимально, рассчитывается следующим образом: половина от количества переменных плюс единица. Если записать в символическом виде, то получаем:

$$l = \left[\frac{m}{2} \right] + 1.$$

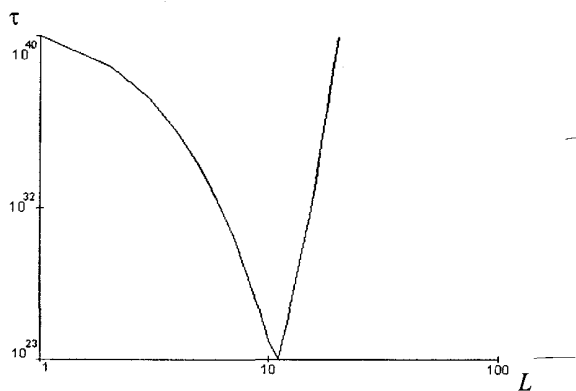


Рис. 1. Зависимость общего времени обучения τ от количества жестких ядер L .

Для проверки предложенной гипотезы было создано программное средство, реализующее обучение ИНС с жестким ядром. Для обучения использовался алгоритм градиентного спуска с выбором параметра скорости обучения, который использует эвристическую стратегию изменения этого параметра в процессе обучения.

Эта стратегия заключается в следующем. Вычисляются выход и погрешность инициализированной нейронной сети. Затем на каждом цикле обучения вычисляются новые значения настраиваемых параметров и новые значения выходов и погрешностей. Если отношение нового значения погрешности к прежнему превышает величину 1,04, то новые значения настраиваемых параметров во внимание не принимаются. При этом параметр скорости настройки уменьшается с коэффициентом 0,7. Если новая погрешность меньше прежней, то параметр скорости настройки увеличивается с коэффициентом 1,05 [1].

Был проведен следующий эксперимент.

Искусственной нейронной сети для аппроксимации было предложено уравнение Менделеева – Клапейрона для одного моля газа $pV = RT$. Исследовалась зависимость температуры T от давления p и объема V . Для приближения данных к экспериментальным, в функцию были добавлены случайные числа в промежутке [0; 1]. В результате получилась следующая формула:

$$T = \frac{pV}{R} + rnd(1).$$

В качестве активационной функции нейронов использовалась логистическая функция или сигмоид

$y = \frac{1}{1 + e^{-x}}$. Диапазон выходных значений данной функции находится в промежутке [0; 1], значит выходные данные надо пропорционально уменьшить так, чтобы они находились в этом диапазоне, иначе говоря, провести масштабирование.

В результате было получено следующее выражение:

$$T = \frac{\left(\frac{pV}{R} + rnd(1) \right)}{12}.$$

Вид функции представлен на рис. 2.

Сначала была создана следующая ИНС: 2 входа, 10 нейронов в скрытом слое, 1 выход. Время обучения составило 11 минут. Критерием оценки служила суммарная квадратическая ошибка:

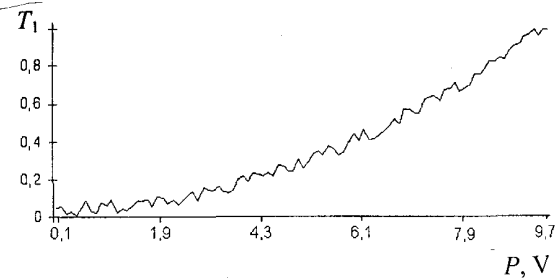


Рис. 2. График экспериментальной функции

$$E = \frac{1}{2} \sum_s \sum_j (d_j^s - y_j^s)^2,$$

где d – целевое значение, y – выход сети, S – количество элементов в обучающем множестве, j – количество нейронов в выходном слое.

Применяя формулу (1), получим оценку времени обучения этой ИНС: $\tau \sim 10 \cdot 100^{30} = 10^{33}$.

Суммарная квадратическая ошибка была равна $E_{\text{общ}} \approx 0,217121$. Для достижения такого значения ошибки потребовалось совершить 24701 цикл обучения. На каждом цикле подстраивались все весовые коэффициенты. Число связей равнялось 30, а значит, было произведено 741030 операций.

Затем было решено обучить две ИНС. Первая имела следующую структуру: 1 вход, 5 нейронов в скрытом слое, 1 выход. Эта сеть впоследствии стала жестким ядром основной ИНС.

Применяя формулу (1) для оценки скорости обучения ИНС с жестким ядром, получаем следующее выражение: $\tau_1 \sim 10 \cdot 100^{10} = 10^{13}$ – скорость обучения жесткого ядра, и $\tau_2 \sim 10 \cdot 100^{10} = 10^{13}$ – скорость обучения ИНС с жестким ядром. В результате сложения получаем: $\tau \sim \tau_1 + \tau_2 = 20^{13}$. Очевидно, что обучение ИНС с жестким ядром будет проходить быстрее.

Было проведено несколько экспериментов, и средняя суммарная квадратическая ошибка оказалась равна $E_{\text{ж.я.}} \approx 0,399308$. Для этого потребовалось в среднем

10944 цикла обучения. Число связей равнялось 10, а значит было произведено 109440 операций.

Затем эта сеть была включена в основную в качестве жесткого ядра. Структура сети с жестким ядром была выбрана такой же, как и структура первоначальной классической ИНС: 2 входа, 10 нейронов в скрытом слое, 1 выход.

Сеть с жестким ядром обучалась до достижения суммарной квадратической ошибки, полученной при обучении ИНС с классической структурой. В среднем количество циклов обучения сети с жестким ядром оказалось равным 6621. Число свободных связей в этой ИНС равнялось 15, а значит, было произведено 99315 операций.

Нетрудно получить суммарное количество операций, выполненных для достижения той же величины ошибки. Оно равно 208755, что в 3,5 раза меньше, чем количество операций, выполненных для обучения классической ИНС.

ЛИТЕРАТУРА

1. *Медведев В.С., Потемкин В.Г.* Нейронные сети. МАТЛАВ 6 / под общ. ред. к. т. н. В.Г. Потемкина. М.: ДИАЛОГ-МИФИ, 2002. 496 с. (Пакеты прикладных программ; Кн. 4).
2. *Максимушкин В.В.* Построение искусственных нейронных сетей с жестким ядром // Актуальные проблемы информатики и информационных технологий: матер. Междунар. (VIII Тамб. межвуз.) науч.-практ. конф. Тамбов: Изд-во ТГУ им. Г.Р. Державина, 2004. 122 с.

Поступила в редакцию 15 июня 2005 г.