

ДИНАМИЧЕСКАЯ СМЕНА МЕТОДОВ ОБУЧЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

КОЛЕСНИКОВ Д.О., КОНСТАНТИНОВА Л.И.

Обсуждается вопрос увеличения скорости обучения искусственных нейронных сетей за счёт динамической смены методов их обучения. Проводится вычислительный эксперимент, который подтверждает это. Предлагается общая схема такого обучения.

1. Цель исследований

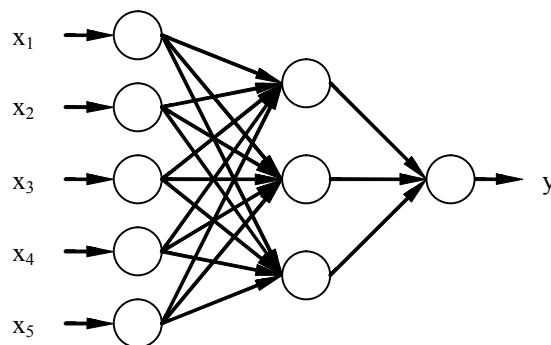
Проблема обучения нейронных сетей чрезвычайно мало изучена. В настоящее время существует значительное количество методов обучения искусственных нейронных сетей (ИНС) [1]. Как правило, метод обучения ИНС заключается в последовательной коррекции весов нейронов, из которых состоит ИНС. К таким методам, например, можно отнести дельта-правило, алгоритм обратного распространения ошибки, стохастические методы обучения и другие. **Целью исследований** является экспериментальное подтверждение возможности увеличения скорости обучения нейронной сети за счет динамической смены методов обучения и разработка общей схемы такого обучения. Сама идея динамической смены методов обучения ИНС совершенно естественно следует из наблюдения за скоростными характеристиками различных методов. Например, если создать нейронную сеть, то, как правило, можно подобрать два таких метода, что обучение ИНС займёт меньше времени, если вначале использовать один метод обучения, а затем, через некоторое время, другой, по сравнению с обучением ИНС только одним из этих методов.

2. Вычислительный эксперимент

В качестве реального примера приведём результаты следующего вычислительного эксперимента. В среде MathLab была создана нейронная сеть прямого распространения [2]. Структурно она состоит из трёх слоёв: 5 нейронов во входном слое, 3 - в скрытом и 1 - в выходном слое. В нейронах первых двух слоёв использовалась в качестве активационной логистическая функция (**logsig**), у нейронов выходного слоя активационная функция линейная (**purelin**).

Функциональное предназначение данной сети – преобразование двоичного кода в десятиричное число. На вход подаётся пятиразрядный двоичный код, на выходе нейронная сеть должна выработать десятиричное соответствие этому коду. Таким образом, данная сеть является конвертором, осуществляющим преобразование $x = (x_1, \dots, x_5) \rightarrow y$, где $x_1, \dots, x_5 \in \{0, 1\}$, $y \in \{0, 1, \dots, 31\}$. Обученная сеть пра-

вильно конвертирует из двоичной системы в десятиричную. Структура сети показана на рисунке.



Структура нейронной сети

Эксперимент заключался в следующем. Создавалась нейронная сеть указанной выше структуры. Уровень ошибки полагался равным 0,5. Далее вычислялось время обучения для трёх различных случаев:

- 1) сеть обучалась согласно методу случайных приращений (функция обучения **trainr**);
- 2) нейронная сеть обучалась по методу градиентного спуска с учетом моментов и с адаптивным обучением (функция обучения **traingdx**);
- 3) обучение проводилось путём комбинации первых двух методов во времени: пока уровень ошибки был более 10, сеть обучалась с использованием функции **trainr**, по достижении этого уровня сеть начинала обучаться с использованием функции **traingdx**.

Таким образом, осуществлялось обучение нейронной сети с использованием двух методов по отдельности и их комбинация.

После формирования программ производилась наработка статистики, т.е. проводилось обучение нейронной сети тремя описанными выше способами определённое количество раз (100 раз). Подсчитывалось среднее время обучения для каждого из трёх способов обучения. В итоге оказалось, что обучение с использованием комбинированного метода (вначале функция **trainr**, затем **traingdx**) занимает в среднем примерно в 1,5 раза меньше времени, чем обучение с использованием метода градиентного спуска с учетом моментов и с адаптивным обучением (функция **traingdx**), и намного меньше, чем обучение с использованием метода случайных приращений (функция **trainr**).

Таким образом, в некоторых случаях комбинация нескольких методов может давать преимущество по скорости обучения по сравнению с применением методов, которые входят в эту комбинацию по отдельности. Следует отметить, что данный факт, хотя и является известным, однако никаких систематических исследований в этой области обнаружено авторами не было. В первую очередь, это связано с тем, что сами методы, алгоритмы обучения не изучены как информационные объекты. Процесс обучения зависит от множества па-

раметров и, в первую очередь, от самой нейронной сети, для обучения которой он применяется.

Для того чтобы иметь возможность динамически оперировать сменой методов обучения в целях ускорения его процесса, сами методы должны быть подвергнуты тщательному изучению. Это изучение должно выявить некоторые числовые характеристики, меняющиеся во времени, по значениям которых управляющая система будет производить смену методов. Чрезвычайно интересным является вопрос о возможности применения для подобного управления метанейронной сети, которая по значениям параметров выбирала бы на каждом шаге (или же с некоторой периодичностью) метод обучения, наиболее эффективный по скорости на данном этапе.

3. Общая схема обучения с помощью динамической смены его методов

Опишем общую схему обучения нейронной сети с помощью динамической смены методов обучения. Осуществим формализацию его процесса. Пусть W — матрица весов, тогда метод обучения может быть записан как оператор $F(W)$, где $F: W \rightarrow W$ — отображение множества W всевозможных значений переменной W в себя. Процесс обучения заключается в построении последовательности приближений к искомому решению W^* . При этом искомое решение далеко не всегда определяется однозначно, т.е. существует некая область W_{ANN} , такая, что любое множество W из W_{ANN} является множеством весов, для которых критерий прекращения обучения CF достигает значения истины:

$$\forall W \in \Omega [W \in \Omega_{ANN} \Leftrightarrow CF(W) = True].$$

Очевидно, что для того, чтобы иметь возможность динамически менять методы обучения, необходимо определить некоторые их характеристики. Эти характеристики удобнее выражать в численном виде. Таким образом, мы вводим в рассмотрение параметр P , который является вектором числовых характеристик метода обучения. Тогда метод обучения можно записать так: $F(W, P)$. В процессе обучения матрица весов нейронной сети последовательно корректируется: $W_1 \rightarrow W_2 \dots \rightarrow W^*$.

Конечный элемент этой последовательности уже является оптимальным решением, т.е. по достижении этого элемента обучение заканчивается.

Наряду с последовательностью, составленной из матриц весов, параллельно существует последовательность значений числовых характеристик: $P_1 \rightarrow P_2 \rightarrow \dots \rightarrow P^*$.

Последний элемент P^* последовательности является по значению таким, что процесс обучения прекращается: $CF(W^*, P^*) = True$.

Рассмотрим теперь случай, когда для обучения нейронной сети имеется некоторая совокупность Φ методов обучения: $\Phi = \{F_j(W, P)\}, j = 1, \dots, n$.

Для динамической смены метода обучения должен существовать селектор S , который выбирает оптимальный метод обучения на каждом шаге или же через фиксированное число шагов: $S(P) = k$, где k — номер метода обучения $F_k(W, P)$.

В таком случае, процесс обучения с помощью динамической смены методов обучения можно записать следующим образом: $W_{z+1} = F_{S(P)}(W_z, P)$, где z — номер итерации (или время) обучения.

Основная сложность при реализации описанного метода заключается в определении вектора p числовых характеристик методов обучения и реализации селектора S . Этот вопрос может служить источником дальнейших исследований в области изучения методов обучения нейронных сетей.

4. Заключение

Научная новизна исследований состоит в том, что впервые проведён вычислительный эксперимент, подтверждающий возможность увеличения скорости обучения нейронной сети за счет динамической смены его методов, а также в построении общей схемы такого обучения.

Практическая ценность исследований заключается в том, что предложенная схема обучения, после разрешения указанных выше вопросов, может быть использована для увеличения скорости обучения нейронных сетей. Кроме этого, экспериментальное подтверждение увеличения скорости обучения при смене методов является основанием для дальнейших перспективных с практической точки зрения исследований в области разработки новых методов, которые увеличат скорость обучения ИНС.

Сравнение с существующими аналогами провести затруднительно, так как, хотя комбинирование методов обучения и используется эпизодически в практике нейронных сетей, в каждом конкретном случае комбинация методов определялась спецификой конкретной практической задачи, и общий подход никем не рассматривался. Таким образом, предложенная схема обучения, по сути, *не имеет аналогов* в теории нейронных сетей. Если же сравнивать существующие методы обучения и предложенную схему их комбинирования, то в некоторых случаях динамическая смена методов дает временной выигрыш, что подтвердил проведённый вычислительный эксперимент.

Литература: 1. Hagan M. T., Demuth H. B., Beale M. H. Neural Network Design. Boston, MA: PWS Publishing, 1996. 2. Demuth H.B., Beale M.H. Neural Network Toolbox User's Guide. Natick: MathWorks Inc, 1997. 700 p.

Поступила в редколлегию 16.07.2004

Рецензент: д-р техн. наук, проф. Руденко О.Г.

Колесников Дмитрий Олегович, канд. техн. наук, старший преподаватель кафедры ПО ЭВМ ХНУРЭ. Научные интересы: искусственный интеллект. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 70-21-446; e-mail: dok@kture.kharkov.ua.

Константинова Лариса Ивановна, канд. физ.-мат. наук, доцент кафедры ВМ ХНУРЭ. Научные интересы: алгебра. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 70-21-436.