

Басканова Т.Ф., Ланкин Ю.П. Нейросетевые алгоритмы самостоятельной адаптации.// Научная сессия МИФИ-99. Всероссийская научно-техническая конференция "Нейроинформатика-99". Сборник научных трудов. В 3 частях. Ч.1.- М.: МИФИ, 1999.- С.17-24.

## НЕЙРОСЕТЕВЫЕ АЛГОРИТМЫ САМОСТОЯТЕЛЬНОЙ АДАПТАЦИИ

*Басканова Т.Ф., Ланкин Ю.П.*  
Красноярский государственный университет,  
Институт биофизики СО РАН

### Аннотация

В работе предложен обучающий алгоритм для нового поколения нейронных сетей, обладающего рядом нетрадиционных возможностей,- нейросетей с самостоятельной адаптацией.

### Представление нейросетей

Данная работа посвящена разработке алгоритмов в концепции построения нейронных сетей с самостоятельной адаптацией (самоадаптирующихся нейросетей), предложенных в работах [1,2], позволяющей значительно расширить возможности нейросетевых методов управления и обработки информации.

Пусть функционирование нейрона нейросети описывается уравнением

$$\alpha_i = \arctg \rho_i, \quad \rho_i = \sum_j \alpha_j x_{ij} + A_i, \quad (1)$$

где  $A_i$  - внешние входные сигналы;  $\alpha_j$  - входные сигналы от других нейронов;  $x_{ij}$  - веса межнейронных связей.

В работе [3] показано, что можно получить сколь угодно точное приближение любой непрерывной функции многих переменных, используя операции сложения и умножения на число, суперпозицию функций, линейные функции, а также одну произвольную непрерывную нелинейную функцию одного переменного. Для нейросетей это означает, что для получения указанного результата от функции

активации нейрона требуется только нелинейность. Адаптивные возможности предлагаемых алгоритмов с ускоренным стохастическим поиском позволяют в полной мере реализовать потенциальные возможности нейросетей.

Целевая функция с помощью которой оценивается успешность обучения (адаптации) может быть задана, например, следующим образом:

$$H = \frac{1}{2} \sum_i (\alpha_i - \alpha_i^*)^2, \quad (2)$$

где  $\alpha_i^*$  - требуемые значения  $\alpha_i$ . Однако существуют и другие возможности [1,2].

В работе [1] приведен метод случайного поиска с целью адаптации нейросети, построенный на флуктуациях весов межнейронных связей. Обладая всеми необходимыми качествами для создания самоадаптирующихся нейросетей, этот метод требует изменения  $N^2$  параметров на каждом пробном шаге, где  $N$  - число формальных нейронов полносвязной нейросети. Метод адаптации, предложенный в работе [2] дает возможность сократить число изменяемых параметров поиска с  $N^2$  до  $N$  за счет варьирования величинами  $\rho_i$  вместо  $x_{ij}$ , что обеспечивает ускорение обучающих алгоритмов. Дальнейшее ускорение осуществляется прогнозированием последующего направления поиска с использованием предыдущего опыта адаптации. Комбинация этих методов позволяет существенно ускорить процедуру случайного поиска, приближая ее быстрдействие к детерминированным алгоритмам и сохраняя ее преимущества перед другими методами при построении самоадаптирующихся нейросетей.

#### **Алгоритм самоадаптации нейросетей**

В процессе адаптации (обучения) нейросети происходит подстройка весовых коэффициентов  $x_i$  до тех пор, пока не будет получен набор  $x_{ij}$ , удовлетворяющих условию  $H \leq \mu$ , где  $\mu$  максимально допустимая величина ошибки.

В работе предложенного в публикации [2] алгоритма можно выделить три повторяющихся этапа:

1. Сохранение текущих значений  $\rho_i^0$  для каждого нейрона и  $H^0$  для сети в целом.

2. Подбор методом случайного поиска величин  $\rho_i$  таким образом, чтобы выполнялось условие  $H < H^0$ . При этом ошибка в уровне активности каждого нейрона сети может быть получена как

$$\Delta \rho_i = \rho_i - \rho_i^0.$$

3. Подстройка весовых коэффициентов осуществляется с помощью преобразования

$$x_{ij} = x_{ij} + \Delta \rho_i \alpha_j y,$$

аналогичного обобщенному правилу [4,5] Уидроу-Хоффа [6] (дельта-правилу), а коэффициент  $y$  пропорционален скорости изменения  $x_{ij}$ .

Изменение  $x_{ij}$  ведет к уменьшению  $\Delta \rho_i$ , приближающему величины  $\rho_i$  и, следовательно,  $\alpha_i = f(\rho_i)$  к желаемым. При этом  $H$  стремится к минимально возможным значениям. То, что наилучшее приближение для  $\rho_i$  существует следует, из теоремы, приведенной в работе [7]. Для упрощения доказательства теоремы заменим  $\rho_i$ - $A_i$  на  $U$ .

Для  $U \in R$  добиться наилучшего приближения можно линейной комбинацией  $\sum_j x_{ij} \alpha_j$  линейно независимых элементов

$\alpha_1, \dots, \alpha_n \in R$ . То есть можно найти такой элемент  $\sum_j x_{ij}^0 \alpha_j$ , что

$$\left\| U - \sum_{j=1}^n x_{ij}^* \alpha_j \right\| = \Delta = \inf_{x_{i1}, \dots, x_{in}} \left\| U - \sum_{j=1}^n x_{ij} \alpha_j \right\|.$$

*Теорема:* Элемент наилучшего приближения существует.

*Доказательство:* Исследуем соотношение (неравенство треугольника)

$$\left\| U - \sum_{j=1}^n x'_{ij} \alpha_j \right\| - \left\| U - \sum_{j=1}^n x''_{ij} \alpha_j \right\| \leq \left\| \sum_{j=1}^n (x'_{ij} - x''_{ij}) \alpha_j \right\| \leq \sum_{j=1}^n |x'_{ij} - x''_{ij}| \|\alpha_j\|.$$

Из него следует, что функция

$$F_U(x_{i1}, \dots, x_{in}) = \left\| U - \sum_{j=1}^n x_{ij}^* \alpha_j \right\|$$

является непрерывной функцией аргументов  $x_{ij}$  при  $\forall U \in R$ .

Пусть  $|\bar{x}|$ -евклидова норма вектора  $\bar{x} = x_{i1}, \dots, x_{in}$ . Функция

$F_0(x_{i_1}, \dots, x_{i_n}) = \|x_{i_1}\alpha_1 + \dots + x_{i_n}\alpha_n\|$  непрерывна на единичной сфере  $|\bar{x}| = 1$  и, следовательно, в некоторой точке  $\tilde{x}_{i_1}, \dots, \tilde{x}_{i_n}$  достигает своей нижней грани  $\tilde{F}$  по сфере, причем  $\tilde{F} \neq 0$ , так как равенство  $\tilde{F} = \|\tilde{x}_{i_1}\alpha_1 + \dots + \tilde{x}_{i_n}\alpha_n\| = 0$  противоречит линейной независимости элементов  $\alpha_1, \dots, \alpha_n$ . Для любого  $\bar{x} = x_{i_1}, \dots, x_{i_n} \neq (0, \dots, 0)$  справедлива оценка

$$\|\tilde{x}_{i_1}\alpha_1 + \dots + \tilde{x}_{i_n}\alpha_n\| = F_0(x_{i_1}, \dots, x_{i_n}) = |\bar{x}| F_0\left(\frac{x_{i_1}}{|\bar{x}|}, \dots, \frac{x_{i_n}}{|\bar{x}|}\right) \geq |\bar{x}| \tilde{F}.$$

Пусть  $\gamma > 2\|U\|/\tilde{F}$ . Функция  $F_U(x_{i_1}, \dots, x_{i_n})$  непрерывна в шаре  $|\bar{x}| \leq \gamma$ ; следовательно в некоторой точке шара  $(x_{i_1}^*, \dots, x_{i_n}^*)$  она достигает своей нижней грани  $F^0$  по шару. Имеем  $F^0 \leq F_U(0, \dots, 0) = \|U\|$ . Вне этого шара выполняется соотношение  $F_U(x_{i_1}, \dots, x_{i_n}) \geq \|x_{i_1}\alpha_1 + \dots + x_{i_n}\alpha_n\| - \|U\| > \tilde{F}2\|U\|/\tilde{F} - \|U\| = \|U\| > F^0$ .

Таким образом

$$F_U(x_{i_1}, \dots, x_{i_n}) \geq F^0 = F_U(x_{i_1}^0, \dots, x_{i_n}^0)$$

при всех возможных  $x_{i_1}, \dots, x_{i_n}$ , что и требовалось доказать.

Отметим, что элементов наилучшего приближения может быть множество.

#### Адаптивная процедура

Для выполнения процедуры случайного поиска может быть использовано множество алгоритмов.

Одним из простейших является алгоритм с возвратом на неудачном шаге [8]. Он сводится к построению последовательности  $\{\rho_i^k\}$  по правилу:

$$\rho_i^{k+1} = \rho_i^k + \gamma^k \xi, \quad k = 0, 1, \dots,$$

где  $\gamma^k$  - некоторая положительная величина;  $\bar{\xi} = (\xi_1, \dots, \xi_n)$  - какая-либо реализация  $n$ -мерной случайной величины  $\bar{\xi}$  с известным законом распределения. Например, координаты  $\xi_h$  случайного век-

тора  $\bar{\xi}$  могут представлять независимые случайные величины, равномерно распределенные на отрезке  $[-1,1]$ .

Пусть  $k$ -е приближение уже известно. Функция  $H(\alpha(\rho_i))$  в уравнении (2) достигает минимума на некотором множестве  $Q \in E^n$ ,  $\rho_i^k \in Q$  ( $k \geq 0$ ).

Алгоритм действует следующим образом. С помощью датчика случайных чисел получают некоторую реализацию случайного вектора  $\bar{\xi}$  и в пространстве  $E^n$  определяют точку  $r_i^k = \rho_i^k + \gamma \bar{\xi}$ ,  $\gamma = \text{const} > 0$ . Если  $r_i^k \in Q$  и  $H(\alpha(r_i^k)) < H(\alpha(\rho_i^k))$ , тогда  $\rho_i^{k+1} = r_i^k$ . Если  $r_i^k \in Q$  но  $H(\alpha(r_i^k)) \geq H(\alpha(\rho_i^k))$ , или  $r_i^k \notin Q$ , то сделанный шаг считается неудачным и полагается  $\rho_i^{k+1} = \rho_i^k$ .

В ситуации, когда требуемая точность достигнута -  $H \leq \mu$ , или  $\rho_i^k = \rho_i^{k+1} = \rho_i^{k+2} \dots = \rho_i^{k+N}$ , если  $N$  достаточно велико, точка  $\rho_i^k$  принимается в качестве искомой точки минимума.

### Ускорение адаптивной процедуры

На этапе функционирования описываемые нейросети совпадают по быстрдействию с нейросетями, обученными по детерминированным алгоритмам, но при обучении даже с учетом предложенных усовершенствований<sup>1</sup>, случайный поиск остается процессом достаточно медленным. Для его ускорения в работе [2] предложено использовать прогнозирование последующего направления поиска с использованием предыдущего опыта адаптации. Набор численных методов, пригодных для прогноза наилучшего изменения величин  $\rho_i$  (а для алгоритма, описанного в [1], величин  $x_{ij}$ ), достаточно широк.

Приведем один из подходов, позволяющих осуществить подобную процедуру. Он представляет собой простую самонастраиваемую модель, основанную на вычислении так называемой экспоненциальной средней [9].

**Поиск с вычислением экспоненциальной средней.** В процессе случайного поиска возникает последовательность значений  $\{\rho_i\}$ , которые удовлетворяют условию улучшения функции оценки. Это

---

<sup>1</sup> Сокращения изменяемых параметров поиска с  $N^2$  до  $N$ , что, например, в 1000-нейронной полносвязной сети позволяет сократить число параметров поиска с  $10^6$  до  $10^3$ .

временной ряд  $\rho_t$ , поведение которого прогнозируется ( $\rho_t$  - значение  $\rho_t$  в некоторый момент времени). Пусть временной ряд, генерируемый некоторой моделью, можно представить в виде можно представить в виде двух компонент

$$\rho_t = \delta_t + \varepsilon_t,$$

где величина  $\varepsilon_t$  генерируется случайным неавтокоррелированным процессом с нулевым математическим ожиданием и конечной дисперсией, а величина  $\delta_t$  может быть сгенерирована либо детерминированной функцией, либо случайным процессом, либо какой-нибудь их комбинацией.

Вычисление и анализ тенденции динамического ряда можно осуществить с помощью его экспоненциального сглаживания. В его основе лежит расчет экспоненциальных средних. Экспоненциальное сглаживание описывается рекуррентной формулой

$$S_t = \nu \rho_t + \beta S_{t-1},$$

где  $S_t$  - значение экспоненциальной средней в момент  $t$ ;  $\nu$  - параметр сглаживания,  $\nu = \text{const}$ ,  $0 < \nu < 1$ ;  $\beta = 1 - \nu$ . Или через значения временного ряда  $\rho_t$

$$S_t = \nu \sum_{l=0}^{N-1} \beta^l \rho_{t-l} + \beta^N S_0,$$

где  $N$  - количество членов ряда;  $S_0$  - некоторая величина, характеризующая начальные условия для первого применения формулы при  $t=1$ .

Так как  $\beta < 1$ , то при  $N \rightarrow \infty$   $\beta^N \rightarrow 0$ , а

$$\nu \sum_{l=0}^{N-1} \beta^l \rightarrow 1.$$

Тогда

$$S_t = \nu \sum_{l=0}^{\infty} \beta^l \rho_{t-l}.$$

Таким образом, величина  $S_t$  является взвешенной суммой всех членов ряда.

Пусть ряд генерируется моделью

$$\rho_t = a_1 + \varepsilon_t,$$

где  $a_1 = \text{const}$ ;  $\varepsilon_t$  - случайное неавтокоррелированное отклонение, или шум, со средним значением 0 и дисперсией  $\sigma^2$ .

Применим к нему процедуру экспоненциального сглаживания. Тогда

$$S_t = v \sum_{l=0}^{\infty} \beta^l \rho_{t-l} = v \sum_{l=0}^{\infty} \beta^l (a_1 + \varepsilon_{t-l}) = a_1 + v \sum_{l=0}^{\infty} \beta^l \varepsilon_{t-l}.$$

Найдем математическое ожидание

$$M(S_t) = M(\rho_t) = a_1$$

и дисперсию

$$D(S_t) = M[(S_t - a_1)^2] = M\left[\left(v \sum_{l=0}^{\infty} \beta^l \varepsilon_{t-l}\right)^2\right] = v^2 \sum_{l=0}^{\infty} \beta^{2l} \sigma^2 = \frac{v}{2-v} \sigma^2.$$

Так как  $0 < v < 1$ ,  $D(S_t) < D(\rho_t) = \sigma^2$ .

Таким образом, экспоненциальная средняя  $S_t$  имеет то же математическое ожидание, что и ряд  $\rho$ , но меньшую дисперсию. При высоком значении  $v$  дисперсия экспоненциальной средней незначительно отличается от дисперсии ряда  $\rho$ .

Чтобы использовать экспоненциальную среднюю для краткосрочного прогнозирования используем ряд, который генерируется моделью

$$\rho_t = a_{1,t} + \varepsilon_t,$$

где  $a_{1,t}$  - варьируемый во времени средний уровень ряда;  $\varepsilon_t$  - случайное неавтокоррелированное отклонение с нулевым математическим ожиданием и дисперсией  $\sigma^2$ .

Прогнозная модель имеет вид

$$\hat{\rho}_\tau(t) = \hat{a}_{1,t},$$

где  $\hat{\rho}_\tau(t)$  - прогноз, сделанный в момент  $t$  на  $\tau$  единиц времени (шагов) вперед;  $\hat{a}_{1,t}$  - оценка  $a_{1,t}$ . Экспоненциальная средняя  $S_t$  служит оценкой для единственного параметра модели  $a_{1,t}$

$$\hat{a}_{1,t} = S_t.$$

Все свойства экспоненциальной средней являются одновременно свойствами прогнозной модели.

Если  $S_{t-1}$  - прогноз на 1 шаг вперед, то величина  $(\rho_t - S_{t-1})$  является погрешностью этого прогноза, а новый прогноз  $S_t$  получается в ре-

погрешностью этого прогноза, а новый прогноз  $S_t$  получается в результате корректировки предыдущего прогноза с учетом ошибки.

### Литература

1. Ланкин Ю.П. Самоадаптирующиеся нейронные сети./ Препринт ТО №3 Института биофизики СО РАН, Теоротдел.- Красноярск, 1997.- 21 с.
2. Ланкин Ю.П. Адаптивные сети с самостоятельной адаптацией./ Препринт ТО №4 Института биофизики СО РАН, Теоротдел.- Красноярск, 1998.- 17 с.
3. Горбань А.Н. Обобщенная аппроксимационная теорема и вычислительные возможности нейронных сетей.// Сибирский журнал вычислительной математики.- Новосибирск: РАН. Сиб. отделение, 1998.- 1, №1.- С.11-24.
4. Rumelhart D.E., Hinton G.E. & Williams R.J. Learning representations by back-propogating errors.// Nature.- 1986.- 323.- P.533-536.
5. Барцев С.И., Охонин В.А. Адаптивные сети обработки информации./ Препринт №59Б Института биофизики СО АН СССР.- Красноярск, 1986.- 19 стр.
6. Widrow B., Hoff M.B. Adaptive Switching Circuits.// IRE WESCOW Conv. Record.- 1960.- Pt.- 4.- P.96-104.
7. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы.- М.: Наука. Гл. ред. физ. мат. лит, 1987.- 600 с.
8. Васильев Ф.П. Численные методы решения экстремальных задач.- М.: Наука. Гл. ред. физ. мат. лит, 1988.- 552 с.
9. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования.- М.: Статистика, 1979.- 252 с.