

АДАПТИВНЫЕ ПОИСКОВЫЕ АЛГОРИТМЫ ОБУЧЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

С.Ю. Белецкая, Д.Н. Лицман, А.Д. Поваляев, А.В. Гаганов

Рассматриваются задачи обучения искусственных нейронных сетей. Для их решения предлагаются адаптивные алгоритмы поисковой оптимизации, построенные на основе рандомизированного подхода

Ключевые слова: обучение нейронной сети, весовые коэффициенты, рандомизация, случайный вектор, адаптивные поисковые алгоритмы

В настоящее время при решении задач проектирования и управления в различных предметных областях широкое применение находят нейросетевые технологии. При этом эффективность нейросетевого моделирования и адекватность построенных моделей во многом зависят от используемых алгоритмов обучения сетей. В этой связи в современных исследованиях большое внимание уделяется разработке и совершенствованию алгоритмов обучения искусственных нейронных сетей [1].

Задача обучения нейронной сети состоит в определении таких значений весовых коэффициентов, которые обеспечивали бы минимальное отклонение выходных сигналов сети от ожидаемых эталонных значений. Обучение с учителем осуществляется на основе обучающей выборки:

$$\langle X^k, D^k \rangle, k = \overline{1, p},$$

где p - количество обучающих пар, X^k - вектор значений входных сигналов сети; D^k - вектор требуемых значений выходных сигналов [1,2].

Пусть Y^k - вектор выходных сигналов нейронной сети, полученный для входного вектора X^k . Тогда задача обучения сводится к задаче минимизации целевой функции, зависящей в конечном итоге от вектора весовых коэффициентов W :

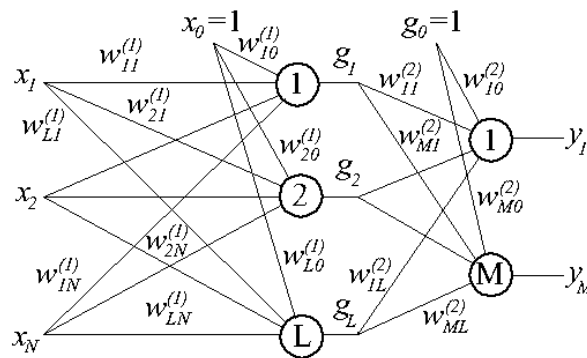
$$F(W) = \frac{1}{2} \sum_{k=1}^p \sum_{i=1}^M (y_i - d_i)^2 \rightarrow \min, \quad (1)$$

где M - количество выходных сигналов.

Если рассматривается единичная обучающая выборка $\langle X, D \rangle$, критерий (1) принимает вид:

$$F(W) = \frac{1}{2} \sum_{i=1}^M (y_i - d_i)^2 \rightarrow \min.$$

Формирование целевой функции для задачи обучения проиллюстрируем на примере многослойного персептрона. На рисунке представлена структурная схема двухслойного персептрона. Здесь x_1, x_2, \dots, x_N - входные сигналы нейронной сети, g_1, g_2, \dots, g_L - выходные сигналы первого слоя, y_1, \dots, y_M - выходные сигналы сети, $w_{ij}^{(m)}$ - значения весов. При этом верхние индексы в скобках (m) означают номер слоя нейрона.



Двухслойный персептрон

Выходные сигналы нейронных слоев определяются следующим образом:

$$g_l = f\left(\sum_{j=0}^N w_{lj}^{(1)} \cdot x_j\right), l = \overline{1, L},$$

$$y_i = f\left(\sum_{l=0}^L w_{il}^{(2)} \cdot g_l\right) = f\left(\sum_{l=0}^L w_{il}^{(2)} \cdot f\left(\sum_{j=0}^N w_{lj}^{(1)} \cdot x_j\right)\right), i = \overline{1, M}. \quad (2)$$

Здесь f - функция активации, одинаковая для всех нейронов сети.

С учётом (2) критерий оптимальности (3) примет вид

$$F(W) = \frac{1}{2} \sum_{k=1}^p \sum_{i=1}^M \left(f\left(\sum_{l=0}^L w_{il}^{(2)} \cdot f\left(\sum_{j=0}^N w_{lj}^{(1)} \cdot x_j\right)\right) - d_i \right)^2 \rightarrow \min.$$

Аналогично можно определить критерии оптимальности в задаче обучения при произвольном числе слоёв. При этом формируемые критерии ха-

Белецкая Светлана Юрьевна – ВГТУ, д-р техн. наук, профессор, тел. (473) 243-77-04

Лицман Дмитрий Николаевич – ВГТУ, аспирант, тел. (473) 243-77-04

Поваляев Анатолий Дмитриевич – ВГТУ, канд. физ.-мат. наук, профессор, тел. (473) 246-12-07

Гаганов Александр Владимирович – ВГТУ, директор НОЦ, e-mail: gaganov@hotmail.ru

рактируются сложной структурой, значительной нелинейностью, высокой размерностью, что обуславливает повышенные требования к используемым алгоритмам обучения.

В статье рассматриваются адаптивные поисковые алгоритмы обучения нейронных сетей. Они строятся на основе вероятностной переформулировки (рандомизации) решаемой задачи (1) и перехода к осредненному критерию оптимальности [3]:

$$G(W) = M[F(W)] \rightarrow \min_{\{W\}}, \quad (3)$$

где M – математическое ожидание. При этом множество возможных решений (значений весовых коэффициентов) W расширяется до множества случайных векторов $\{W\}$.

Рандомизированные итерационные процедуры перестройки весовых коэффициентов нейронной сети имеют следующий вид:

$$W^{N+1} = W^N + \alpha_N H^N, \quad (4)$$

где N – номер итерации; H^N – направление движения; α_N – значение шага. При этом вектор H^N также является случайным.

При дальнейшей алгоритмизации поисковая процедура (4) переформулируется с использованием различных вероятностных характеристик случайных векторов. Так, переходя в формуле (4) к математическим ожиданиям, получим

$$M[W^{N+1}] = M[W^N] + \alpha_N M[H^N]. \quad (5)$$

В работах [2,3] рассмотрены математические основы построения адаптивных поисковых алгоритмов. При этом на каждой N -й итерации осуществляется замена исходной целевой функции некоторой потенциальной функцией $v(W)$ с последующим использованием для её оптимизации процедур градиентного типа. Градиент потенциальной функции $v(W)$ статистически оценивается в процессе поиска по реализациям случайных векторов и адаптивно перестраивается на основе текущей информации. В конечном итоге формируется обобщенная рандомизированная процедура поиска оптимальных вариантов, которая затем может быть переформулирована в различных вероятностных характеристиках. Так, в работе [3] сформирована итерационная поисковая схема с использованием математического ожидания:

$$m^{N+1} = m^N + \alpha_N \cdot M_{W^N} M_{U^N} \left[\frac{\Psi(f(U^N) - c_N)}{\omega_n |W^N - U^N|^n} (W^N - U^N) \right], \quad (6)$$

где $m^N = M[W^N]$; ω_n – площадь поверхности единичной сферы в пространстве R^n ; U^N – вспомогательный случайный вектор; $c_N = \text{const}$ – средний уровень целевой функции, который используется для анализа текущей информации;

$\Psi(\cdot)$ – монотонная неубывающая функция, удовлетворяющая условиям: $\Psi(t) \cdot t > 0 \quad \forall t \neq 0$; $\Psi(0) = 0$.

Процедура (6) носит обобщенный характер и служит основой для построения комплекса адаптивных алгоритмов поискового типа, использующих только значения критерия оптимальности. При этом вариант алгоритмической схемы зависит от способов получения реализаций случайных векторов W и U , выбора функции $\Psi(t)$, вариантов определения уровня c_N , а также от выбранных стратегий статистической оценки математического ожидания. Так, если использовать одну реализацию вектора U^N , а в качестве реализации вектора W^N выбрать его математическое ожидание $m^N = M[W^N]$, можно перейти к следующей схеме [4,5]:

$$m^{N+1} = m^N + \alpha_N \left[\frac{\Psi(f(u^N) - c_N)}{\omega_n |m^N - u^N|^n} (m^N - u^N) \right], \quad (7)$$

где через u^N обозначена реализация случайного вектора U^N . Процедура (7) является адаптивным расширением метода переменного многогранника. При этом в качестве вершин многогранника рассматриваются реализации случайных векторов, а в качестве центра тяжести многогранника – перестраиваемое математическое ожидание m^N .

Таким образом, для алгоритмизации рассматриваемых задач настройки весовых коэффициентов нейронной сети может быть использована следующая обобщенная вычислительная схема:

1. Инициализация начальных значений весовых коэффициентов $W^0 = (w_1^0, \dots, w_n^0)$, где n – общее количество весовых коэффициентов по всем слоям нейронной сети. При этом $N=0$, где N – номер итерации.

2. Получение k -реализаций случайного вектора U^N : $u^{N,i}$, $i = \overline{1, k}$. Каждая реализация $u^{N,i}$ представляет собой n -мерный вектор $u^{N,i} = (u_1^{N,i}, \dots, u_n^{N,i})^T$. При этом в качестве первой реализации выбирается начальный вектор весовых коэффициентов сети, задаваемый в начале работы алгоритма: $u^{N,1} = W^0$. Если количество реализаций k равно $n+1$, то реализации $u^{N,i}$ могут рассматриваться как вершины переменного многогранника в пространстве R^n .

3. Вычисление значений целевой функции $F(u^{N,i})$, $i = \overline{1, k}$. После получения всех значений $F(u^{N,i})$ реализации $u^{N,i}$ упорядочиваются по

возрастанию значений целевой функции:
 $F(u^{N,1}) \leq \dots \leq F(u^{N,k})$.

4. Определение среднего уровня целевой функции $c_N = \frac{1}{k} \sum_{i=1}^k F(u^{N,i})$.

5. Разделение реализаций на группы перспективности в зависимости от значения уровня c_N :

$\Omega_p : f(u^{N,i}) \leq c_N$ - перспективная группареализаций;

$\Omega_v : f(u^{N,i}) > c_N$ - неперспективная группа.

6. Вычисление математического ожидания m^N по r перспективным реализациям:

$$m^N = \sum_{j=1}^r p_j^N u^{N,j}, \quad \sum_{j=1}^r p_j^N = 1.$$

При этом вероятности p_j^N определяются следующим образом:

$$p_i^N = \left| F(u^{N,i}) - c_N \right| / \sum_{j=1}^r \left| F(u^{N,j}) - c_N \right|, \quad i = \overline{1, r}.$$

7. Движение в случайных векторах и определение значения m^{N+1} . При этом используются обобщенная рандомизированная процедура (6) или ее различные варианты (например, схема (7)). Если движение осуществляется на основе схемы (7), то данная итерационная процедура применяется к реализации $u^{N,k}$ с наибольшим значением целевой функции.

8. Адаптивное уточнение шага α_N по схеме:

$$\alpha_N = \begin{cases} \alpha_N \gamma_1, & \text{если } F(m^{N+1}) \leq c_N \\ \alpha_N \gamma_2, & \text{если } F(m^{N+1}) > c_N \end{cases},$$

где $\gamma_1 > 1$; $0 < \gamma_2 < 1$. Параметры γ_1 и γ_2 задаются пользователем в начале работы алгоритма.

9. Корректировка списка реализаций следующим образом:

$$u^{N+1,k} = m^{N+1}, \quad F(u^{N+1,k}) = F(m^{N+1}).$$

10. Проверка сходимости в соответствии с условием

$$\sigma_N^2 = \frac{1}{k} \sum_{i=1}^k (F(u^{N,i}) - c_N)^2 < \varepsilon,$$

где ε – точность, задаваемая на входе алгоритма.

11. Если сходимость не достигнута, осуществляется переход к шагу 3 (при этом $N = N + 1$).

Воронежский государственный технический университет

ADAPTIVE SEARCH ALGORITHMS OF ARTIFICIAL NEURAL NETWORKS TRAINING

S.Yu. Beletskaja, D.N. Litsman, A.D. Povalyaev, A.V. Gaganov

The problems artificial neural networks training are considered. For solving of this problems the adaptive search engine optimization algorithms, based on randomized approach, are suggested

Key words: neural networks training, weights, randomization, random vector, adaptive search algorithms

После окончания работы алгоритма результирующие значения весовых коэффициентов нейронной сети W_{opt} определяются следующим образом:

$$W_{opt} = u^{N,1},$$

где $u^{N,1}$ – реализация с наименьшим значением целевой функции.

В работе [3] показано, что рассматриваемые адаптивные алгоритмы поисковой оптимизации имеют нелокальный характер. Доказываются теоремы о выходе итерационного процесса из зон локальных экстремумов и о сходимости в заданную область.

Необходимо заметить, что на основе различных интерпретаций обобщенной рандомизированной процедуры (6) могут быть построены и другие варианты адаптивных поисковых алгоритмов. Используемые при построении данных алгоритмов принципы рандомизации позволяют вынести поиск оптимальных решений на уровень множества случайных векторов, что обеспечивает возможность выявления статистических закономерностей в свойствах оптимизируемой функции и позволяет повысить эффективность оптимизационного процесса при обучении нейронных сетей.

Литература

1. Норенков И.П. Основы автоматизированного проектирования [Текст] / И.П. Норенков. - М.: МГТУ им. Баумана, 2002. 448 с.
2. Рутковская, Д. Нейронные сети, генетические алгоритмы и нечеткие системы [Текст] / Д. Рутковская, М. Пилиньский, Л. Рутковский. - М.: Горячая линия Телеком, 2013. - 384 с.
3. Каплинский, А.И. Моделирование и алгоритмизация слабоформализованных задач выбора наилучших вариантов систем [Текст] / А.И. Каплинский, И.Б. Русман, В.М. Умывакин. - Воронеж: Изд-во ВГУ, 1990. - 234 с.
4. Львович, Я.Е. Адаптивные методы оптимального проектирования [Текст] / Я.Е. Львович, С.Ю. Белецкая // Информационные технологии. - 2010. - № 7. - С. 1-32.
5. Батищев, Д.И. Оптимизация в САПР [Текст] / Д.И. Батищев, Я.Е. Львович, В.Н. Фролов. - Воронеж: Изд-во ВГУ, 1997. - 416 с.
6. Проблемы разработки интеллектуальных систем многоальтернативного моделирования [Текст] / С.Л. Подвальный, Т.М. Леденева, А.Д. Поваляев, Е.С. Подвальный // Вестник Воронежского государственного технического университета. - 2013. - Т. 9. - № 3-1. - С. 19-23.
7. Основы автоматизации проектирования, тестирования и управления жизненным циклом изделия [Текст] / В.Ф. Барабанов, А.Д. Поваляев, С.Л. Подвальный, С.В. Тюрин. - Воронеж. 2011.