

УМЕНЬШЕНИЕ РАЗМЕРНОСТИ ЗАДАЧИ ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ

Как правило, практические задачи моделирования реальных процессов отличаются весьма значительным числом независимых переменных. Минимизация функции ошибки выхода нейронной сети, используемой для моделирования технологии металлургического производства, также относится к задачам большой размерности. Для моделирования одного передела производства используется двухслойная нейронная сеть [1]. Согласно теореме Хехт-Нильсена для аппроксимации зависимости между m входными и n выходными параметрами достаточно двухслойной нейронной сети с прямыми полными связями, $2m + 1$ нейронами первого и n нейронами второго слоя [2]. Тогда число весовых коэффициентов для нейронов первого слоя составит $(m + 1)(2m + 1)$, для второго слоя – $(2m + 2)n$.

Для наглядности рассмотрим задачу, где число входных параметров равно 20 ($m = 20$), число выходных параметров – 5 ($n = 5$). Получим, что число весов равно:

$$(m + 1)(2m + 1) + (2m + 2)n = (20 + 1)(2 \cdot 20 + 1) + (2 \cdot 20 + 2) \cdot 5 = 1071.$$

При таком большом числе переменных очень маловероятно случайно попасть в точку, близкую к оптимуму.

Логичным решением представляется перебор различных значений с некоторым шагом по всем переменным (наложение сетки). Для каждой комбинации значений нужно рассчитать ошибку выхода сети и выбрать лучшую комбинацию. При 4-х различных значениях каждой переменной число комбинаций переменных (число узлов сетки) будет таким:

$$4^{1071} \approx (2^2)^{1071} \approx 2^{2000} \approx (2^{10})^{200} \approx 1000^{200} \approx (10^3)^{200} \approx 10^{600}.$$

Как видно, не представляется возможным выполнить полный перебор.

Для уменьшения размерности задачи обучения нейронной сети предлагается выделять подсети, каждая из которых будет обучаться как отдельная нейронная сеть. При этом выходные параметры отдельной сети должны зависеть от части входных параметров первоначальной сети. Замена сети несколькими сетями меньшего размера позволяет уменьшить общее число неизвестных весовых коэффициентов и тем самым ускорить процесс обучения.

Проведем эксперимент, возьмем 3 малых сети [1] и одну общую сеть. Структура малых сетей следующая:

1-я сеть: 30 входов, 12 выходов;

2-я сеть: 25 входов, 6 выходов;

3-я сеть: 19 входов, 5 выходов.

Взято 200 строк данных для обучения, после удаления некомплектных данных (строк, в которых по какому-либо параметру отсутствуют данные) осталось 197 (рис. 1). Обучение сети на 40 итерациях заняло 37 минут.

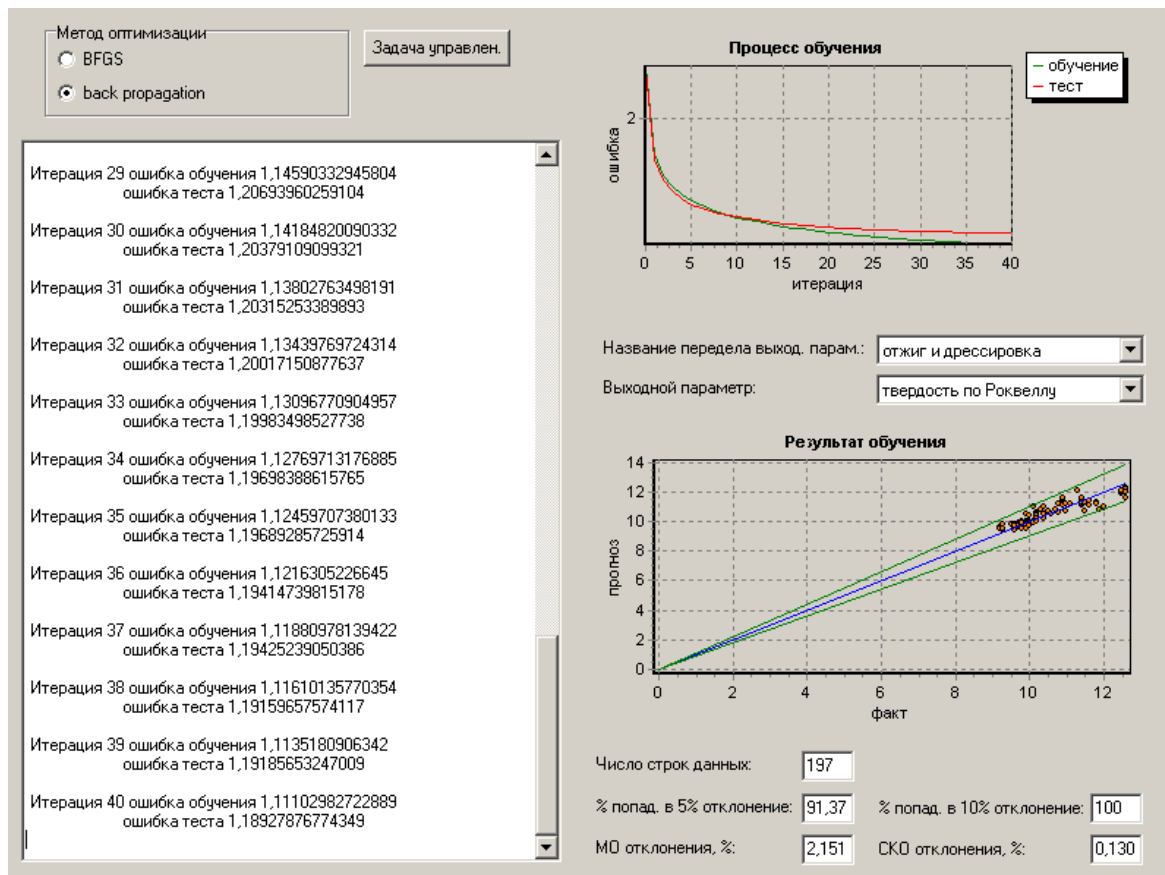


Рис. 1. Процесс обучения 3-х малых сетей

При обучении нескольких сетей на графике процесса обучения выводится сумма среднеквадратичных ошибок 3-х сетей. Каждая из сетей обучается по отдельности.

Параметр, который в варианте с 3 малыми сетями был и входом, и выходом, в большой сети определен как выходной параметр. В этой сети 50 входов и 23 выхода. Проведение тех же 40 итераций обучения данной сети (рис. 2), заняло уже 208 минут.

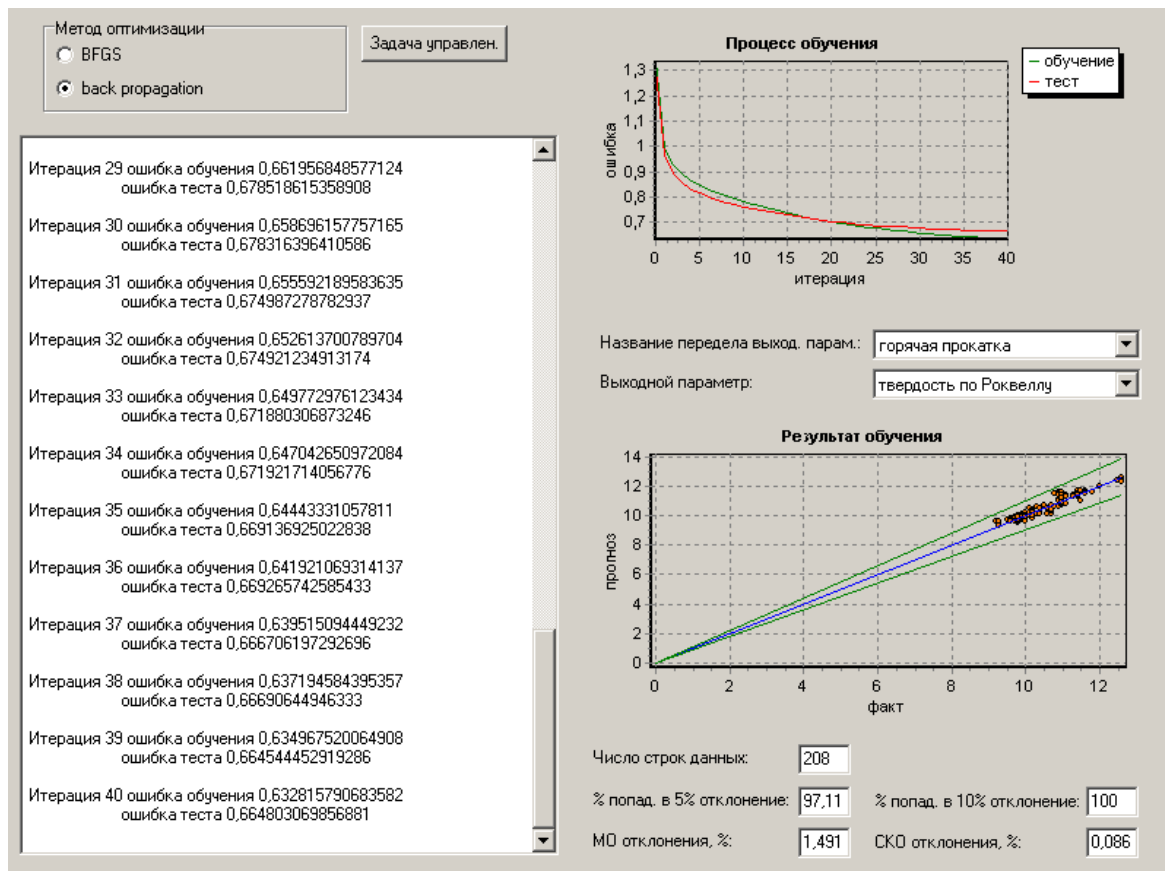


Рис. 2. Процесс обучения большой сети

Выигрыш по времени более чем в 5 раз (37 минут против 208). Для обучения большой сети требуется больше данных, иначе сеть “запомнит” данные и не научится обобщать информацию. Эффект запоминания происходит из-за избыточного количества весов для текущего объема обучающей выборки (или для невысокой сложности зависимости, представленной в данных). Сеть с большим количеством весов быстро обучится по данным, при этом аппроксимируемая нейронной сетью зависимость может оказаться гораздо сложнее.

Определим количество весовых коэффициентов в обоих случаях.

1-я сеть (30 входов, 12 выходов):

$$(m + 1)(2m + 1) + (2m + 2)n = (30 + 1)(2 \cdot 30 + 1) + (2 \cdot 30 + 2) \cdot 12 = 2635.$$

2-я сеть (25 входов, 6 выходов):

$$(m + 1)(2m + 1) + (2m + 2)n = (25 + 1)(2 \cdot 25 + 1) + (2 \cdot 25 + 2) \cdot 6 = 1638.$$

3-я сеть (19 входов, 5 выходов):

$$(m + 1)(2m + 1) + (2m + 2)n = (19 + 1)(2 \cdot 19 + 1) + (2 \cdot 19 + 2) \cdot 5 = 980.$$

Итого по всем 3 сетям:

$$2635 + 1638 + 980 = 5253.$$

Для большой сети (50 входов, 23 выхода):

$$(m + 1)(2m + 1) + (2m + 2)n = (50 + 1)(2 \cdot 50 + 1) + (2 \cdot 50 + 2) \cdot 23 = 7497.$$

Во втором случае весов оказалось больше почти в 1,5 раза.

Обучение сети выполнялось на основе метода обратного распространения ошибки. Этот метод принято считать классическим методом обучения нейронной сети,

он имеет невысокую сходимость, просто реализуется и не является требовательным к производительности используемой вычислительной системы. Изменение весов рассчитывается по формуле [3]:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}},$$

где $0 < \eta < 1$ – множитель, задающий скорость обучения, E – функция ошибки выхода сети. Веса изменяются, начиная с нейронов последнего слоя, далее меняются веса предпоследнего слоя и так далее до первого слоя.

Применение в эксперименте другого метода (BFGS) не позволило бы утверждать о преимуществе нескольких малых сетей. Метод BFGS использует в вычислениях приближение матрицы Гессе. Для сети с 50 входами и 23 выходами число весов будет более 7000. Тогда матрица будет содержать более 49 000 000 элементов. Из-за большой размерности задачи на современных персональных компьютерах не получилось выполнить расчеты. Выдавалось системное сообщение о недостаточном объеме свободной памяти. Попытка сохранить промежуточные результаты вычисления матрицы Гессе в переменные программы не привела к успеху, так как загрузка центрального процессора увеличивалась до 100%, размер файла подкачки приближался к предельному. В результате получить результаты обучения первой итерации не удалось.

Таким образом, обучение нейронной сети в виде нескольких сетей позволяет уменьшить число весовых коэффициентов и уменьшить время обучения. В случае большой вычислительной сложности метода оптимизации ошибки это решение может быть единственным вариантом обучения сети.

Библиографический список:

1. Качановский Ю.П. Основные принципы построения нейросетевой модели формирования свойств автолистовой стали [Текст] / Ю.П. Качановский, Е.А. Коротков // Вести высших учебных заведений Черноземья. – Липецк: ЛГТУ, 2009. №4. – с. 70-74.
2. Круглов В.В. Искусственные нейронные сети: теория и практика [Текст] / В.В Круглов, В.В Борисов. – М.: Горячая линия – Телеком, 2001. – 382 с.
3. Осовский С. Нейронные сети для обработки информации [Текст] / Пер. с польского И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 344 с.: ил.

Адрес для переписки:

evg.korotkov@mail.ru