

Крючин О.В., Воробьева А.А.  
Тамбовский Государственный Университет им. Г.Р. Державина  
Россия, г. Тамбов  
kryuchov@gmail.com

### Построение аналитической модели для оценки эффективности использования параллельных или распределенных вычислений для подбора числа скрытых нейронов искусственной нейронной сети

Как известно, полный процесс обучения искусственных нейронных сетей (ИНС) состоит из трех вложенных уровней — подбора структуры сети, подбора активационных функций и подбора весовых коэффициентов. Именно из-за такой большой вложенности этот процесс требует значительных временных затрат, которые являются чрезмерными даже при использовании современных компьютеров. Как показано в работе [1] решением этой проблемы является использование параллельных или распределенных вычислений. При использовании распараллеливания на уровне подбора структуры сети временные затраты на межпроцессорную передачу данных являются незначительными по сравнению с временными затратами на подбор активационных функций конкретной структуры, поэтому в данном случае нет большой разницы между кластерной или распределенной системой (то есть локальной сетью), поскольку скорость передачи данных между элементами информационных ресурсов (ИР-элементов), то есть узлов не важна.

Алгоритм подбора структуры сети подробно рассмотрен в работе [2]. Для того чтобы построить аналитическую модель именно для подбора числа скрытых нейронов необходимо обозначить число элементов начальной структуры нейронной сети как  $S^{(0)}$ . После начала обучения в сеть добавляется один дополнительный скрытый нейрон и число элементов становится равным  $S^{(1)}$ . Затем добавляется еще один скрытый нейрон и количество элементов становится  $S^{(2)}$ . Аналогично после добавления следующих нейронов. Поскольку типы сети могут быть различны, и каждый имеет свои особенности, то конкретные значения  $S$  определяются выбранным типом.

Как следует из работ [2-3] каждая итерация подбора числа скрытых нейронов требует следующих действий:

1. добавить новый скрытый нейрон (возможно, добавить слой, если таковы требования конкретного типа сети);
2. произвести подбор активационных функций нейронов (обучить нейроны); при этом производится также подбор весовых коэффициентов (для каждой комбинации активационных функций);
3. проверить качество полученной структуры (критерием качества, как правило, выступает значение невязки).

Если обозначить количество вариантов, которые необходимо перебрать для обучения структуры сети (то есть, для подбора числа скрытых нейронов) как  $I_S$ , то число необходимых мультипликативных операций, которые требуются для этого можно выразить формулой:

$$z_S = \sum_{i=0}^{I_S-1} (z_{\mu}(S^{(i)}) + z_N(S^{(i)}, \mu_i) + \lambda_{Si}). \quad (1)$$

здесь  $z_N(S^{(i)}, \mu_i)$  — число мультипликативных операций, требуемых для добавления нейрона  $\mu_i$  в  $i$ -ую структуру (то есть, операций, которые необходимы для перестроения сети, изменения весов и тп — это зависит, во-первых, от типа ИНС, поскольку, например, добавление нового скрытого нейрона в сеть Вольтерри приводит к геометрическому увеличению структуры, а во-вторых — от программной реализации ИНС),  $z_{\mu}(S^{(i)})$  — число мультипликативных операций, которые требуются для подбора активационных функций нейронов (включая весовые коэффициенты) в  $i$ -ой структуре (подробно рассмотрено в работах [3-4]),  $\lambda_{Si}$  — число операций, необходимых для проверки критерия останова обучения (то есть, как правило, для вычисления значения невязки) при  $i$ -ой структуре.

Если используются кластерная или распределенная системы, то на каждый ИР-элемент можно отправить конкретную структуры с тем, чтобы он самостоятельно подбирал активационные функции нейронов (а также весовые коэффициенты). Таким образом,  $k$ -ый ИР-элемент для выполнения заданной ему работы (то есть для обучения нейронов и весов) выполняет  $z_\mu(S^{(k-1)})$  мультипликативных операций. Кроме того, необходимо учитывать, что для передачи структуры сети также выполняется ряд операций, число которых можно рассчитать как  $S^{(k-1)} + 2\sigma S^{(k-1)} + \gamma(S^{(k-1)}, v)$  и часто ему необходимо ожидать завершения работы других ИР-элементов, что можно также привести к число мультипликативных операций равных  $kS^{(k-1)} + 2k\sigma S^{(k-1)}$ , то общее число операций, выполняемых на ИР-элементе на первом (для них) этапе обучения (то есть когда ИР-элемент первый раз получил структуру ИНС и произвел для нее подбор активационных функций и весовых коэффициентов), вычисляется по формуле

$$C_{Sk}^{(0)} = (k + 1)S^{(k-1)} + 2(k + 1)\sigma S^{(k-1)} + z_\mu(S^{(k-1)}) + \gamma(S^{(k-1)}, v) + \sum_{i=0}^{k-1} Y(S^{(i)}, \mu_i), \quad (2)$$

где  $k$  — номер ИР-элемента. Таким образом, ведущий ИР-элемент при первом проходе осуществляет операций

$$C_{S0}^{(0)} = \max_{k=1..n-1} \left( C_{Sk}^{(0)} + \gamma(S^{(k-1)}, v) + S^{(k-1)}(1 + 2\sigma) + \lambda_{Sk-1} \right), \quad (3)$$

здесь  $C_{Sk}^{(0)}$  — число операций, которые выполняет для обучения структуры  $k$ -ый ИР-элемент,  $\gamma(S^{(k-1)}, v)$  — число операций, необходимых для возвращения обученной структуры,  $S^{(k-1)}(1 + 2\sigma)$  — число операций, необходимых для получения структуры ведущим ИР-элементом,  $\lambda_{Sk-1}$  — число операций, необходимых для проверки сети на адекватность.

Как правило, количество итераций подбора числа скрытых слоев  $I_S$  значительно превосходит количество ИР-элементов  $n$ , а поэтому  $k$ -ый ИР-элемент последовательно выполняет обучения структур с номерами  $(k - 1)$ ,  $(n + k - 2)$ ,  $(2n - 3 + k)$ ,  $\dots$ ,  $(In - I - 1 + k)$ ,  $\dots$ ,  $(I_{Sk})$ . Номер последней структуры  $I_{Sk}$ , обучаемой  $k$ -ым ИР-элементом, можно вычислить по формуле (4), а количество структур  $N_S(I_{Sk})$  — по формуле (5).

$$I_{Sk} = \begin{cases} I_S - 1 + k - (I_S \bmod (n - 1)), & I_S \bmod n - 1 \notin (1; k); \\ I_S - n + k - (I_S \bmod (n - 1)), & I_S \bmod n - 1 \in (1; k), \end{cases} \quad (4)$$

$$N_S(I_{Sk}) = \begin{cases} \left\lfloor \frac{I_S}{n-1} \right\rfloor + 1, & I_S \bmod n - 1 \notin (1; k); \\ \left\lfloor \frac{I_S}{n-1} \right\rfloor, & I_S \bmod n - 1 \in (1; k). \end{cases} \quad (5)$$

Исходя из вышесказанного, число мультипликативных операций, которые осуществляет  $k$ -ый ИР-элемент (включая пустые, то есть ожидание, выраженное через число операций), может быть выражено формулой

$$C_{Sk}^\wedge = C_{Sk}^{(0)} + \gamma(S^{(k-1)}, v) + S^{(k-1)} + 2\sigma S^{(k-1)} + \lambda_{Sk-1} + \sum_{i=1}^{N_S(I_{Sk})-1} (e_S(i)), \quad (6)$$

$$e_S(i) = 4S^{(j)} + 8\sigma S^{(j)} + z_\mu(S^{(j)}) + 2\gamma(S^{(j)}, v) + Y(S^{(j)}, \mu_j) + \lambda_{Sj}, j = i(n - 1) + k - 1. \quad (7)$$

Следовательно, общее число операций  $Z_S$  вычисляется по формуле

$$Z_S = \max_{k=1..n-1} (C_{Sk}^\wedge) + \psi_S, \quad (8)$$

где  $\psi_S$  — задержки, возникающие из-за асинхронности процесса, а коэффициент эффективности — по формуле

$$\alpha_S(Z) = \frac{Z_S}{nZ_S}. \quad (9)$$

Таким образом, в рамках данной работы построена аналитическая модель, которая показывает эффективность параллельного или распределенного подбора структуры ИНС.

#### **Литература**

1. Крючин О.В, Вязовова Е.В., Арзамасцев А.А. Реализация параллельного алгоритма подбора структуры искусственной нейронной // Вестник ТГУ, т.18, вып.4, 2013 — С. 1394-1395.
2. Крючин О.В. Построение модели информационных процессов выбора структур искусственной нейронной сети // Вестник ТГУ, т.19, вып.6, 2014 — С. 1836-1840.
3. Арзамасцев А.А., Крючин О.В. Реализация параллельного алгоритма подбора активационных функций искусственной нейронной сети // ЭФТЖ, т.6, 2011— С. 52-61.
4. Арзамасцев А.А., Крючин О.В., Слетков Д.В., Зенкова Н.А. Количество мультипликативных операций при подборе весовых коэффициентов искусственной нейронной сети // Вестник ТГУ, т.18, вып.1, 2013 — С. 176-182.