

УДК 004.9

DOI: 10.20310/1810-0198-2016-21-1-142-145

ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ ОБУЧЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ МЕТОДОМ ПОЛНОГО СКАНИРОВАНИЯ И АНАЛИТИЧЕСКАЯ МОДЕЛЬ ПАРАЛЛЕЛЬНОЙ ВЕРСИИ ЭТИХ ПРОЦЕССОВ

© О.В. Крючин

Описаны параллельные методы обучения искусственных нейронных сетей при помощи полного сканирования. Приводится аналитическая модель информационных процессов для этих методов.

Ключевые слова: аналитическая модель; информационные процессы; параллельные методы; искусственные нейронные сети.

Как известно, при подборе весовых коэффициентов $\overline{W}$ искусственной нейронной сети (ИНС) методом полного сканирования происходит перебор всех возможных вариантов их значений. При этом значение i -го весового коэффициента на I -й итерации вычисляется по формуле:

$$w_i^{(I)} = l_{0i} + s_i \left(\left[\frac{I}{\hat{s}(i)} \right] \bmod \left[\frac{l_{1i} - l_{0i}}{s_i} + 1 \right] \right), \quad i = \overline{0, l_w - 1}, \quad (1)$$

где l_{0i} , l_{1i} – нижний и верхний пределы, а s_i – шаг сканирования i -го весового коэффициента, l_w – размерность вектора весовых коэффициентов,

$$\hat{s}(i) = \begin{cases} 1, & i = l_w - 1; \\ \prod_{j=i+1}^{l_w-1} \left(\left[\frac{l_{1j} - l_{0j}}{s_j} \right] + 1 \right), & i < l_w - 1. \end{cases} \quad (2)$$

Общее количество итераций вычисляется по формуле:

$$I_F = \prod_{i=0}^{l_w-1} \left(\left[\frac{l_{1i} - l_{0i}}{s_i} \right] + 1 \right). \quad (3)$$

Использование подобного метода обучения приводит к необходимости задействовать значительное число элементов информационных ресурсов (ИР-элементов) – вычислительных узлов кластерной системы или процессоров суперкомпьютера. Применение большого числа ИР-элементов, в свою очередь, требует адаптировать метод обучения. Так, при использовании n ИР-элементов каждый из них, за исключением нуле-

вого (ведущего), перебирает J_F вариантов, а нулевой – $\hat{J}_F$. Значения J_F и $\hat{J}_F$ вычисляются по формулам:

$$J_F = \begin{cases} \left[\frac{I_F}{n} \right], & I_F \bmod n = 0; \\ \left[\frac{I_F}{n-1} \right], & I_F \bmod n \neq 0, \end{cases} \quad (4)$$

$$\hat{J}_F = \begin{cases} J_F, & I_F \bmod n = 0; \\ I_F - J_F(n-1), & I_F \bmod n \neq 0. \end{cases} \quad (5)$$

На рис. 1–2 представлены блок-схемы алгоритма для управляющего и неуправляющего ИР-элемента, соответственно, порядок выполнения информационного процесса для ведущего ИР-элемента следующий [1–2]:

- 1) формирование $n-1$ структур для неуправляющих ИР-элементов;
- 2) рассылка ИНС-структур на все используемые ИР-элементы;
- 3) перебор всех возможных вариантов из имеющихся на данном ИР-элементе и выбор оптимального набора весовых коэффициентов и связанной с ним погрешности;
- 4) получение с ИР-элементов подобранных весовых коэффициентов и соответствующих им значений невязки;
- 5) выбор наименьшего значения невязки и установка в ИНС соответствующих ему весовых коэффициентов.

Для неуправляющего ИР-элемента порядок выполнения информационного процесса другой:

- получение структуры;
- перебор всех возможных вариантов и выбор оптимального набора весовых коэффициентов и связанной с ним погрешности;
- отправка на управляющий ИР-элемент весовых коэффициентов и погрешности [3].

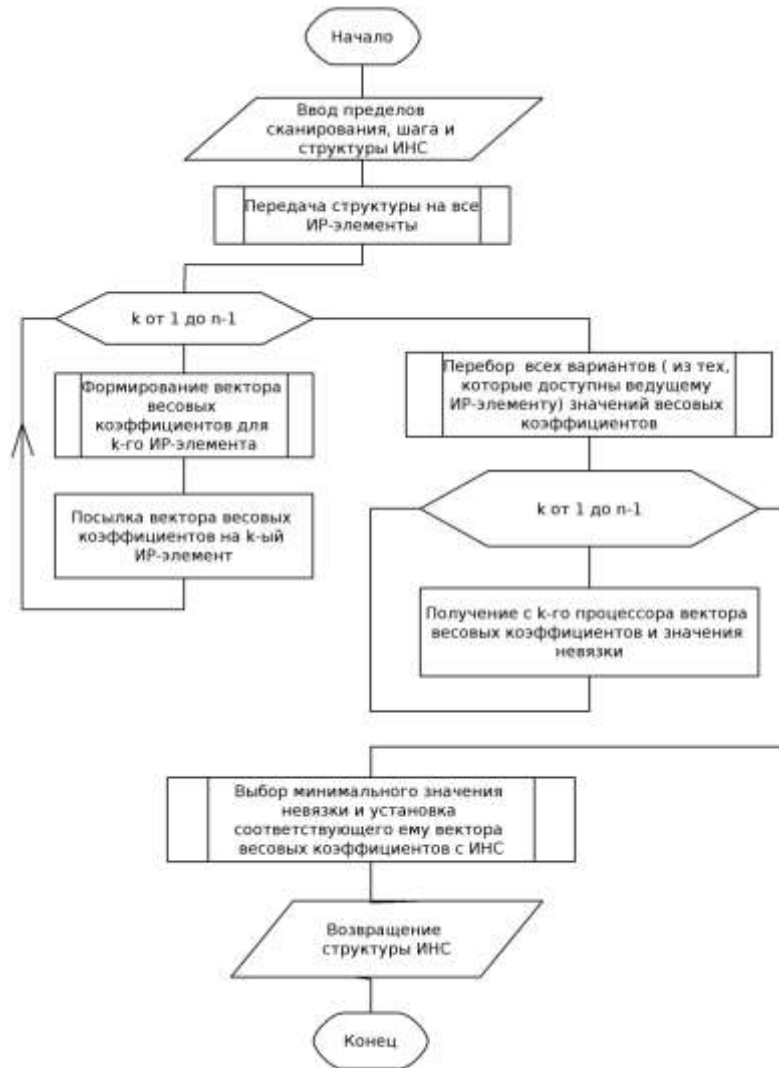


Рис. 1. Блок-схема информационного процесса обучения, запущенного на ведущем (управляющем) ИР-элементе при использовании метода полного сканирования



Рис. 2. Блок-схема информационного процесса обучения, запущенного на неведущем (неуправляющем) ИР-элементе при использовании метода полного сканирования

Для оценки эффективности этих процессов разработана аналитическая модель.

Количество мультипликативных операций, которые производит последовательный алгоритм полного сканирования, можно записать следующим образом:

$$z_{wF} = z_{\varepsilon} I_F + 2\sigma I_F, \quad (6)$$

где z_{ε} – количество операций, необходимых для последовательного вычисления невязки

$$\varepsilon = \frac{1}{N} \sum_{i=0}^{N-1} (\bar{d}_i - \bar{y}_i)^2 = \frac{1}{N} \sum_{i=0}^{N-1} \sum_{j=0}^{P-1} (d_{i,j} - y_{i,j})^2, \quad (7)$$

где $\bar{d}_i$, $\bar{y}_i$ – i -е выходные вектора моделируемого объекта и ИНС; N – количество строк в обучающей выборке; P – количество выходов объекта (размерность векторов $\bar{d}_i$ и $\bar{y}_i$); I_F – число итераций алгоритма (информационный процесс I_F раз выполняет

вычисление невязки, а для организации цикла ему требуется $2I_F$ операций – по 2 на каждой итерации, вперед и назад) [2].

Для приведения аддитивных операций к мультипликативным введен σ -коэффициент, который прямо пропорционален времени, затрачиваемому на одну мультипликативную операцию, и обратно пропорционален времени, затрачиваемому на одну аддитивную (мультипликативная операция занимает в σ раз больше времени, чем аддитивная и, следовательно, одну мультипликативную операцию можно заменить на σ аддитивную и наоборот) [4].

Для вычисления количества мультипликативных операций, производимых информационными процессом полного сканирования, необходимо рассмотреть этапы, из которых он состоит:

- 1) инициализация;
- 2) передача данных с ведущего ИР-элемента (элемента информационного ресурса, в качестве которого может выступать узел кластерной системы или компьютер вычислительной сети) на прочие;
- 3) перебор значений весовых коэффициентов, принадлежащих данному ИР-элементу;
- 4) передача данных со всех ИР-элементов на ведущий;
- 5) выбор ведущим ИР-элементом оптимальной конфигурации.

На первом этапе на ведущем ИР-элементе происходит $2l_w$ аддитивных операций (l_w – число весов сети – для организации цикла и l_w для присвоения), на втором – $l_w(n-1)$ мультипликативных и $2l_w(n-1)$ аддитивных. На неведущих на втором этапе происходит l_w мультипликативных и $2l_w$ аддитивных операций, но начаться они могут лишь после того, как ведущий ИР-элемент отправит данные, поэтому на неведущем ИР-элементе происходит $C_{Fk}^{(2)} = l_w(k + 2k\sigma + 2\sigma + 1) + \gamma(l_w, v)$ мультипликативных операций.

Количество операций на третьем этапе аналогично количеству операций, выполняемых последовательной версией алгоритма (за исключением того, что число итераций $\hat{J}_F$ для ведущего ИР-элемента и J_F для прочих).

На четвертом этапе ведущему ИР-элементу необходимо получить лучшие весовые коэффициенты и соответствующее им значение невязки с каждого ИР-элемента, поэтому он делает $(n-1)(l_w+1)$ мультипли-

кативных и $2(n-1)(l_w+1)$ аддитивных операций, а неведущий – l_w+1 мультипликативных и $2l_w+2$ аддитивных. Кроме того, ведущий ИР-элемент должен ожидать отправки.

На пятом этапе ведущий ИР-элемент делает $2n$ аддитивных операций. Если привести аддитивные операции к мультипликативным, то можно получить выражения, приведенные в табл. 1.

До начала получения ведущим ИР-элементом весовых коэффициентов и значений невязки он осуществляет 3 этапа (обозначим общее количество операций $\hat{C}_{F0} = 2\sigma l_w + l_w(n-1)(1+2\sigma) + z_\varepsilon \hat{J}_F + 2\sigma \hat{J}_F$), а прочие – 4 (обозначим общее количество операций $\hat{C}_{Fk} = l_w(k + 2k\sigma + \sigma + 1) + \gamma(l_w, v) + z_\varepsilon J_F + 2\sigma J_F + (l_w + 1)(1 + 2\sigma)$). Для выполнения первых 2 этапов ведущий ИР-элемент производит $C_{\varepsilon 0}^{(1)} + C_{\varepsilon 0}^{(2)}$, а k -й – $C_{\varepsilon k}^{(1)} + C_{\varepsilon k}^{(2)}$ операций.

Кроме того, необходимо учитывать время передачи невязки и весовых коэффициентов. Поскольку получение завершается после того, как значение невязки будет послано самым медленным процессором, то для информационного процесса, использующего параллельное вычисление невязки, требуется Z_{wF} операций, где Z_{wF} вычисляется по формуле:

$$Z_{wF} = \max_{k=1..n-1} (\max(\hat{C}_{F0} + k(l_w + 1)(2\sigma + 1), \hat{C}_{Fk} + \gamma(l_w + 1, v))) + (l_w + 1)(1 + 2\sigma) + 2\sigma n. \quad (8)$$

Исходя из вышесказанного, эффективность параллельного информационного процесса обучения можно выразить формулой:

$$\alpha_{wF}(Z) = \frac{Z_{wF}}{nZ_{wF}} = \frac{z_\varepsilon I_F + 2\sigma I_F}{nZ_{wF}}. \quad (9)$$

Таким образом, аналитическая модель выглядит следующим образом:

$$\alpha_{wF}^*(Z) = \frac{I_F N(\zeta_y + 1 + \sigma \hat{\zeta}_y + 2\sigma P) + I_F + 2\sigma I_F}{n \max_{k=1..n-1} \left(\hat{C}_{F0} + k(l_w + 1)(2\sigma + 1), \hat{C}_{Fk} + \gamma(l_w + 1, v) \right) + (l_w + 1)(1 + 2\sigma) + 2\sigma n}. \quad (10)$$

Таблица 1

Число мультипликативных операций на различных этапах информационного процесса

Этап	Ведущий ИР-элемент	Неведущий (k -й) ИР-элемент
1	$2\sigma l_w$	
2	$l_w(n-1)(1+2\sigma)$	$l_w(k + 2k\sigma + 2\sigma + 1) + \gamma(l_w, v)$
3	$Z_\varepsilon \hat{J}_F + 2\sigma \hat{J}_F$	$Z_\varepsilon J_F + 2\sigma J_F$
4	$n(l_w + 1)(1 + 2\sigma) + \gamma(l_w + 1)$	$(l_w + 1)(1 + 2\sigma)$
5	$2\sigma n$	

СПИСОК ЛИТЕРАТУРЫ

1. Крючин О.В., Арзамасцев А.А., Королев А.Н., Горбачев С.И., Семенов Н.О. Универсальный симулятор, базирующийся на технологии искусственных нейронных сетей, способный работать на параллельных машинах // Вестник Тамбовского университета. Серия Естественные и технические науки. Тамбов, 2008. Т. 13. Вып. 5. С. 372-375.
2. Крючин О.В. Параллельный алгоритм полного сканирования обучения искусственных нейронных сетей // В мире научных открытий. Красноярск, 2010. № 6.3 (12). С. 72-79.
3. Крючин О.В. Параллельные алгоритмы обучения искусственных нейронных сетей // Материалы 15 Международной конференции по нейрокибернетике. Т. 2. Симпозиум «Интерфейс "Мозг-Компьютер"», 3 Симпозиум по нейроинформатике и нейрокомпьютерам. Ростов н/Д, 2009. С. 93-97.
4. Крючин О.В., Арзамасцев А.А. Сравнение эффективности последовательных и параллельных алгоритмов обучения искусственных нейронных сетей на кластерных вычислительных системах // Вестник Тамбовского университета. Серия Естественные и технические науки. Тамбов, 2010. Т. 15. Вып. 6. С. 372-375.

Поступила в редакцию 30 ноября 2015 г.

Крючин Олег Владимирович, Тамбовский государственный университет им. Г.Р. Державина, г. Тамбов, Российская Федерация, кандидат технических наук, ведущий программист института математики, естественных и информационных технологий, e-mail: kryuchov@gmail.com

UDC 004.9

DOI: 10.20310/1810-0198-2016-21-1-142-145

INFORMATION PROCESSES OF ARTIFICIAL NEURAL NETWORKS TRAINING BY METHOD OF FULL SCANNING AND ANALYTICAL MODEL OF PARALLEL VERSION OF THESE PROCESSES

© O.V. Kryuchin

Parallel methods of artificial neural networks training with the help of full scanning are described. The analytical model of information processes of these methods is presented.

Key words: analytical model; information processes; parallel methods; artificial neural networks.

REFERENCES

1. Kryuchin O.V., Arzamastsev A.A., Korolev A.N., Gorbachev S.I., Semenov N.O. Universalnyy simulyator, baziruyushchiysya na tekhnologii iskusstvennykh neyronnykh setey, sposobnyy rabotat na parallelnykh mashinakh. *Vestnik Tambovskogo universiteta. Seriya Estestvennye i tekhnicheskie nauki – Tambov University Reports. Series: Natural and Technical Sciences*. Tambov, 2008, vol. 13, no. 5, pp. 372-375.
2. Kryuchin O.V. Parallelnyy algoritm polnogo skanirovaniya obucheniya iskusstvennykh neyronnykh setey. *V mire nauchnykh otkrytiy*. Krasnoyarsk, 2010, no. 6.3 (12), pp. 72-79.
3. Kryuchin O.V. Parallelnye algoritmy obucheniya iskusstvennykh neyronnykh setey. *Materialy 15 Mezhdunarodnoy konferentsii po neyrokibernetike. Vol. 2. Simpozium «Interfeys "Mozg-Kompyuter"», 3 Simpozium po neyroinformatike i neyrokompyuteram*. Rostov-on-Don, 2009, pp. 93-97.
4. Kryuchin O.V., Arzamastsev A.A. Sravnenie effektivnosti posledovatelnykh i parallelnykh algoritmov obucheniya iskusstvennykh neyronnykh setey na klasternykh vychislitelnykh sistemakh. *Vestnik Tambovskogo universiteta. Seriya Estestvennye i tekhnicheskie nauki – Tambov University Reports. Series: Natural and Technical Sciences*. Tambov, 2010, vol. 15, no. 6, pp. 372-375.

Received 30 November 2015

Kryuchin Oleg Vladimirovich, Tambov State University named after G.R. Derzhavin, Tambov, Russian Federation, Candidate of Technics, Leading Programmer of Mathematics, Natural Science and Information Technologies Institute, e-mail: kryuchov@gmail.com