

Распознавание большого количества образов при помощи нейронных сетей с использованием многопроцессорных систем

Ясинский Ф.Н., д-р физ.-мат. наук, Мочалов А.С., асп.

Рассматривается алгоритм распараллеливания обучения искусственных нейронных сетей с использованием многопроцессорных систем. Представлены результаты испытания алгоритма на кластере ИГЭУ «Энергет».

Ключевые слова: нейронные сети, распознавание образов, обучение нейронных сетей, многопроцессорные системы, параллельный алгоритм, алгоритм обратного распространения ошибки, выбор структуры нейронной сети.

Recognition of Large Number of Patterns with Using Neuronal Networks and Multiprocessor Systems

F.N. Yasinsky, Doctor of Physics and Mathematics, A.S. Mochalov, Post Graduate Student

This article describes the parallel training algorithm of artificial neural networks using multi-processor systems. Also in this article presents the results of testing the described algorithm on a cluster ISPU «ENERGY».

Keywords: neural networks, pattern recognition, learning of neural networks, multiprocessor systems, parallel algorithm, backpropagation algorithm, selection of neural network structure.

Как известно нейронные сети не программируются в привычном смысле этого слова, они обучаются. Возможность обучения – одно из главных преимуществ нейронных сетей перед традиционными алгоритмами. Технически обучение заключается в нахождении коэффициентов связей между нейронами.

В случае, если нейронная сеть должна обучаться распознаванию очень большого количества образов, время обучения нейронной сети будет велико.

Представим, что нам нужно создать систему распознавания очень большого количества китайских иероглифов. Иероглифы сами по себе имеют сложное начертание. Кроме того, по условию системе придется различать очень большое количество иероглифов. Соответственно, время обучения нашей программной системы будет велико.

Однако есть возможность ускорить обучение системы путем применения распараллеливания с использованием многопроцессорных систем [1]. Поступим следующим образом: распределим задачу по множеству нейронных сетей. Для этого каждую нейронную сеть обучим отличать ограниченный набор иероглифов. Условно говоря, каждая нейронная сеть станет «экспертом» по своей группе иероглифов. Попробуем применить выбранный способ распараллеливания для задачи распознавания иероглифических символов.

Введем ряд терминов. *Область компетенции* процесса – группа иероглифов, которые обязана различать нейронная сеть на данном процессе. *Область вне компетенции* – группа иероглифов, которые нейронная сеть не должна различать, а лишь определять, что они не относятся к ее *области компетенции*.

Условимся, что в параллельном алгоритме каждый процесс будет содержать одну нейронную сеть. Общая схема применения алгоритма

для 3-х иероглифов при использовании 3-х нейронных сетей представлена на рис. 1.

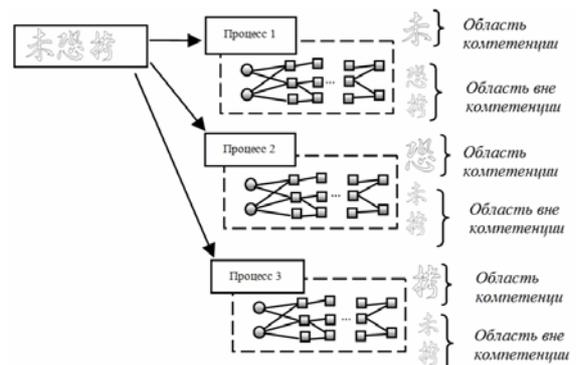


Рис. 1. Распределение образов по нейронным сетям в выбранном способе распараллеливания на примере задачи распознавания 3-х иероглифов

После обучения каждая нейронная сеть на каждом процессе [2] должна вернуть:

- число от 0 до 1 как степень уверенности в том, что входной символ принадлежит ее *области компетенции*;
- код распознанного символа из своей *области компетенции*.

Процессы, участвующие в распознавании, можно условно разделить на 2 вида (рис. 2) [2]:

- главный процесс;
- рабочий процесс.

Распределением данных между процессами занимается особый выделенный процесс, который условно можно назвать главным процессом. После распределения данных каждый процесс выполняет работу над своим блоком данных и затем отправляет результат главному процессу, который обобщает все принятые данные и выносит решение о результате распознавания. В качестве результата распознавания главный процесс должен выбрать результат той нейронной сети, которая наиболее уверена в своем ответе.

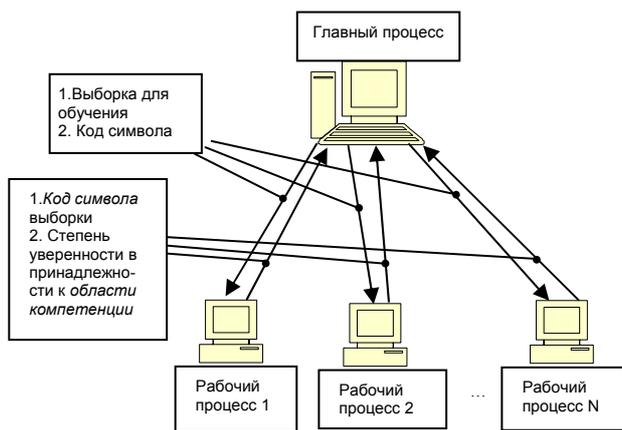


Рис. 2. Общая схема организации работы процессов в алгоритме

Роль главного процесса:

- подготавливает данные (обучающие выборки, символы для распознавания, коды символов);
- рассылает данные рабочим процессам;
- принимает пару *код символа – степень уверенности в принадлежности к области компетенции* от рабочих процессов;
- обобщает результаты и делает вывод о значении распознаваемого символа;
- выполняет действия рабочего процесса для своей части данных.

Роль рабочего процесса:

- получает обучающую выборку, символ для распознавания, код символа;
- инициализирует нейронную сеть и обучает её полученной выборке;
- выполняет распознавание;
- отправляет главному процессу пару *код символа – степень уверенности в принадлежности к области компетенции*.

Преимущества алгоритма:

- возможно ускорение обучения в случае правильной настройки алгоритма;
- затраты времени на обмен данными между нейронными сетями при обучении отсутствуют.

Недостатки алгоритма:

- хоть нейронная сеть и не должна различать свои символы из *области вне компетенции*, но определять принадлежность входного образа к ним она должна. Таким образом, в обучающую выборку текущей нейронной сети должны входить символы из *области компетенции* других нейронных сетей;
- возможен конфликт интересов нейронных сетей, т. е. ситуация, когда одна нейронная сеть ошибочно опознает поданный на распознавание символ как иероглиф из своей *области компетенции* и вернет большую степень уверенности в принадлежности символа к своей *области компетенции*, чем та сеть, которая действительно училась распознаванию поданного иероглифа, что может повлечь за собой ошибку распознавания.

лифа, что может повлечь за собой ошибку распознавания.

Для успешной работы параллельного алгоритма необходимо также выбрать правильную структуру нейронных сетей.

При выборе структуры нужно руководствоваться следующими правилами [3]:

1. Количество нейронных элементов в промежуточном слое должно быть меньше количества тренировочных образцов.
2. Мощность нейронной сети можно увеличивать как за счет количества нейронов в слое, так и за счет количества слоев. Если на нейронную сеть накладываются ограничения пункта 1 и она не может решить поставленную задачу, то необходимо увеличивать количество скрытых слоев нейронной сети.

Чем больше количество нейронных элементов в скрытом слое, тем больше точность. Однако при слишком большой размерности скрытого слоя может наступить явление, которое называется перетренировкой сети. Перетренированная нейронная сеть может хорошо аппроксимировать функцию только на тренировочных наборах. Таким образом, слишком большое количество нейронов в скрытом слое ухудшает обобщающие способности нейронных сетей. Поэтому количество нейронных элементов в скрытом слое должно быть меньше количества тренировочных образцов. С другой стороны, при слишком малой размерности скрытого слоя можно попасть в нежелательный локальный минимум или процесс обучения будет слишком длительным. Поэтому необходим разумный компромисс [4].

Для обеспечения требуемой точности и обобщающей способности можно использовать нейронную сеть с двумя скрытыми слоями, размерность которых меньше, чем при использовании трехслойной сети. Однако нейронные сети, которые имеют несколько скрытых слоев, обучаются значительно медленнее [4].

Как уже упоминалось выше, выходной результат каждого процесса (т. е. нейронной сети на каждом процессе) будет представлен рядом нулей и единиц, формирующих в результате двоичное представление десятичного числа, характеризующий класс образа (код символа).

Также каждая нейронная сеть должна обязательно возвращать помимо кода распознанного символа, степень уверенности в принадлежности, т. е. уверенность, что нейронная сеть опознала «свой» символ (рис. 3).

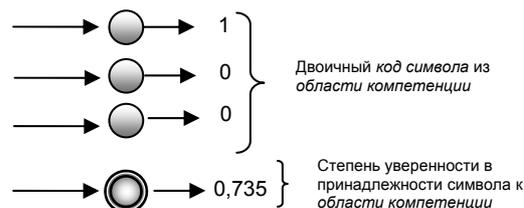


Рис. 3. Выходные данные нейронных сетей в рассматриваемом способе распараллеливания

Результаты обучения системы, распознающей последовательно 2, 3 и 4 иероглифа, с использованием указанного способа распараллеливания на 2, 3 и 4-х процессах соответственно по сравнению с однопроцессорным вариантом представлены на рис. 4–6, в табл. 1–3. Замеры проводились на кластере ИГЭУ «Энергет». На каждом процессе обучалась одна нейронная сеть. При замерах результатов работы параллельного алгоритма каждая нейронная сеть на каждом процессе специализировалась на распознавании одного из 2, 3 или 4-х символов-иероглифов, в зависимости от испытания.

Для обучения каждой нейронной сети использовался алгоритм обратного распространения ошибки. На каждом процессе применялась нейронная сеть следующей структуры: первый слой нейронной сети содержал 900 нейронов, второй – 20 нейронов, третий – 4 нейрона.

Испытание №1:

Количество символов: 2

Шаг: 0,01

Максимально допустимая среднеквадратичная ошибка: 0,04

Результат: табл. 1, рис. 4.

Таблица 1. Результаты испытания для 2-х символов

Процесс	Последовательный вариант		Параллельный вариант	
	Количество итераций	Время, с	Количество итераций	Время, с
0	23400	10,84109	24240	11,18837
1			24000	10,98687

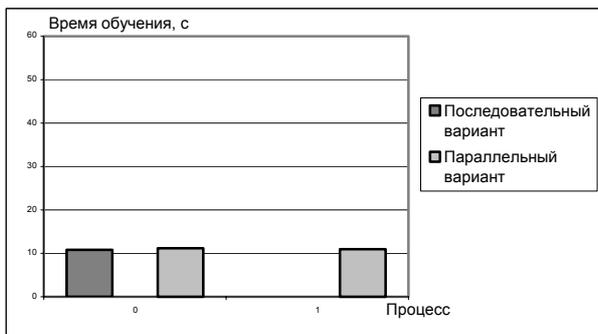


Рис. 4. Результаты испытания для 2-х символов

Вывод. Для двух символов время обучения в последовательном и параллельном вариантах практически не отличается.

Испытание №2:

Количество символов: 3

Шаг: 0,01

Максимально допустимая среднеквадратичная ошибка: 0,04

Результат: табл. 2, рис. 5.

Таблица 2. Результаты испытания для 3-х символов

Процесс	Последовательный вариант		Параллельный вариант	
	Количество итераций	Время, с	Количество итераций	Время, с
0	69660	32,03527	33480	15,37338
1			29160	13,38122
2			39600	18,08049

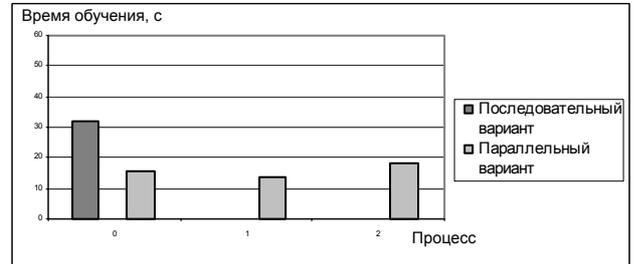


Рис. 5. Результаты испытания для 3-х символов

Вывод. Для трех символов при обучении у параллельного варианта появляется преимущество. Это происходит за счет того, что в последовательном варианте нейронной сети приходится учиться различать 3 образа, а при этом в параллельном варианте каждая нейронная сеть учится различать 2 образа – символ из своей области компетенции и символы вне области компетенции.

Испытание №3:

Количество символов: 4

Шаг: 0,01

Максимально допустимая среднеквадратичная ошибка: 0,04

Результат: табл. 3, рис. 6.

Таблица 3. Результаты испытания для 4-х символов

Процесс	Последовательный вариант		Параллельный вариант	
	Количество итераций	Время, с	Количество итераций	Время, с
0	118560	54,13879	32880	15,11914
1			35520	16,38798
2			36720	17,66580
3			44400	20,56264

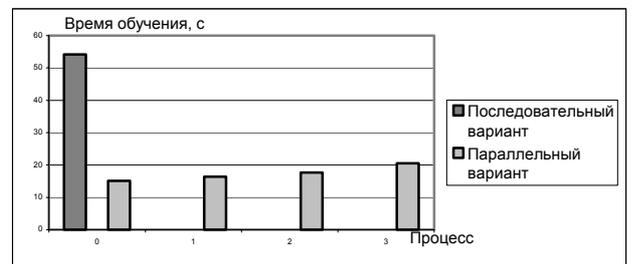


Рис. 6. Результаты испытания для 4-х символов

Вывод. Для трех процессов при обучении у параллельного варианта появляется значительное преимущество. Это происходит за счет того, что в последовательном варианте нейронной сети приходится учиться различать 4 образа, а при этом в параллельном варианте каждая

нейронная сеть учится различать 2 образа – символ из своей *области компетенции* и символы вне *области компетенции*.

Анализ полученных результатов позволяет говорить о том, что выбранный алгоритм распараллеливания дает хорошие результаты. Однако масштаб рассмотренной задачи невелик. Необходимо продолжать разработку и испытание алгоритма.

Список литературы

1. Царегородцев В.Г. Перспективы распараллеливания программ нейросетевого анализа и обработки данных:

мат-лы III Всеросс. конф. «Математика, информатика, управление – 2004». – Иркутск, 2004.

2. Мочалов А.С. Распознавание иероглифических текстов при помощи нейронных сетей на многопроцессорных системах: мат-лы науч.-техн. конф. «Энергия-2010». – Иваново, 2010

3. Головкин В.А. Нейронные сети: обучение, организация и применение. – М.: ИПРЖР, 2002

4. Саймон Хайнкин. Нейронные сети. Полный курс. – Вильямс, 2006.

5. Евдокимов П.В., Гаврилов А.И. Определение оптимальных параметров нейронной сети при построении тематических моделей технологических процессов // Вестник ИГЭУ. – 2007. – Вып. 4.

6. Анисимов А.А., Горячев М.Н. Идентификация электромеханических систем с использованием искусственной нейронной сети // Вестник ИГЭУ. – 2008. – Вып. 3.

Ясинский Федор Николаевич,
ГОУВПО «Ивановский государственный энергетический университет имени В.И. Ленина»,
доктор физико-математических наук, профессор кафедры высокопроизводительных вычислительных систем,
телефон (4932) 26-98-29.

Мочалов Александр Сергеевич,
ГОУВПО «Ивановский государственный энергетический университет имени В.И. Ленина»,
аспирант кафедры высокопроизводительных вычислительных систем,
телефон (4932) 26-98-29.