

ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ

УДК 007.001.362

Н.Г. Аксак, И.В. Новосельцев, А.С. Солдатов

АДАПТАЦИЯ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ НА ВЫСОКОПРОИЗВОДИТЕЛЬНЫЕ СИСТЕМЫ С ОБЩЕЙ ПАМЯТЬЮ

Введение

В настоящее время в машиностроении все больше функций управления и контроля, которые раньше выполнялись человеком, переключаются на автоматические системы управления. Необходимым этапом создания систем управления нелинейными динамическими объектами является построение адекватных математических моделей. Наиболее широкое распространение получили параметрические модели. Несмотря на огромное количество работ в этой области, многообразие видов нелинейности не позволяет создать единую теорию идентификации нелинейных систем. Поэтому область применения классических подходов к построению систем управления нелинейными динамическими объектами ограничена.

Альтернативой классическим методам являются системы управления, основанные на нейронных сетях [1]. Нейросетевой подход, сочетающий в себе способность компьютера к обработке чисел и способность мозга к обобщению и распознаванию, особенно эффективен в задачах распознавания и диагностики. Сдерживающим фактором широкого применения нейросетевых алгоритмов является то, что реализации нейронных сетей ресурсоемки, особенно когда речь идет о классификации изображений, а время обучения очень велико.

С переходом компьютерной индустрии на многоядерные архитектуры, появилась возможность увеличить производительность нейронных сетей за счет распараллеливания однотипных операций. Для эффективного использования вычислительных возможностей высокопроизводительных систем требуется разработка методик

распараллеливания, на основе которых может осуществляться реализация нейронных сетей.

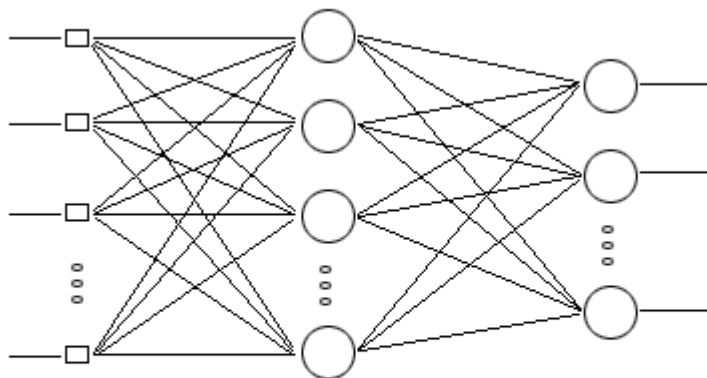
Постановка задачи

Для всех нейронных сетей (НС) присущ принцип параллельной обработки сигналов. Параллелизм выражается в том, что одновременно обрабатывается большое число цепочек нейронов, но программные модели нейронных сетей на традиционных компьютерах лишены свойства высокого параллелизма мозга.

Вследствие того, что в настоящий момент отсутствуют универсальные методики и рекомендации по адаптации нейросетевых моделей на системы с общей памятью, возникает необходимость разработки и исследования методов и подходов, позволяющих решить поставленные задачи с учетом особенностей конкретной предметной области.

Основной раздел

В данной работе предлагается адаптация моделей пересептронного типа (рис.1) на высокопроизводительные системы с общей памятью.



$$X = [x_1, x_2, \dots, x_i] \quad Y^{(1)} = [y_1^{(1)}, y_2^{(1)}, \dots, y_j^{(1)}] \quad Y^{(2)} = [y_1^{(2)}, y_2^{(2)}, \dots, y_k^{(2)}]$$

Входной слой Скрытый слой Выходной слой

Рисунок 1 - Архитектура многослойного персептрона

Функционирование нейрона определяется следующим соотношением

$$y_m^{(L)}(t) = \varphi \left(\sum_{n=1}^i w_{nm}^{(L)}(t) y_n^{(L-1)}(t) + b_m^{(L)} \right)$$

где $w_{nm}^{(L)}$ - синаптический вес связи n -го нейрона слоя $L-1$ с m -ым нейроном слоя L , $b_m^{(L)}$ порог m -ого нейрона, $\varphi(\bullet)$ - функция активации нейрона.

При обучении нейронных сетей наиболее ресурсоемкими являются операции умножения матрицы на вектор.

Реализация вычисления нейронов на однопроцессорных системах
На однопроцессорной системе вычисление нейронов каждого слоя осуществляется последовательно. Сначала производятся вычисления вектора скрытого слоя, что представляет собой последовательные вычисления $y_1^{(1)}$, $y_2^{(1)}$, $y_j^{(1)}$, затем аналогично вычисляются значения выходного слоя. Такие операции требуют больших временных затрат.

Таким образом, реализация нейронной сети представляет собой циклы вычисления нейронов. Для прямого прохода по НС достаточно двух циклов, в каждом из которых производятся вычисления нейронов соответствующих слоев.

На рисунке 2-а представлен граф алгоритма последовательного вычисления нейронов на однопроцессорной системе в процессе прямого прохода по НС.

Адаптация вычисления нейронов многослойного персептрона на многопроцессорных системах с общей памятью с помощью технологии OpenMP

Технология OpenMP является средством программирования компьютеров с общей памятью, базирующихся на традиционных языках программирования. За основу берется последовательная программа. В OpenMP любой процесс состоит из нескольких нитей управления, которые имеют общее адресное пространство, но разные потоки команд и отдельные стеки [2].

На рисунке 2-б представлен граф алгоритма реализации нейронной сети на четырехядерной системе с помощью технологии OpenMP.

Модель производительности параллельного алгоритма [3] реализации НС основана на двух параметрах:

1. Ускорение $S_p = T_1 / T_p$, где T_1, T_p - время выполнения параллельного алгоритма на одном и p ядрах соответственно;

2. Эффективность $\varepsilon = S_p/p$, определяющая реальную выгоду от использования многоядерных систем по сравнению с однопроцессорными.

Как видно из графа, на четырехядерной системе, происходит параллельное вычисление четырех нейронов, что уменьшает общее время прямого прохода при обучении. При количестве процессорных элементов равном количеству нейронов в каждом слое достигается максимальная производительность.

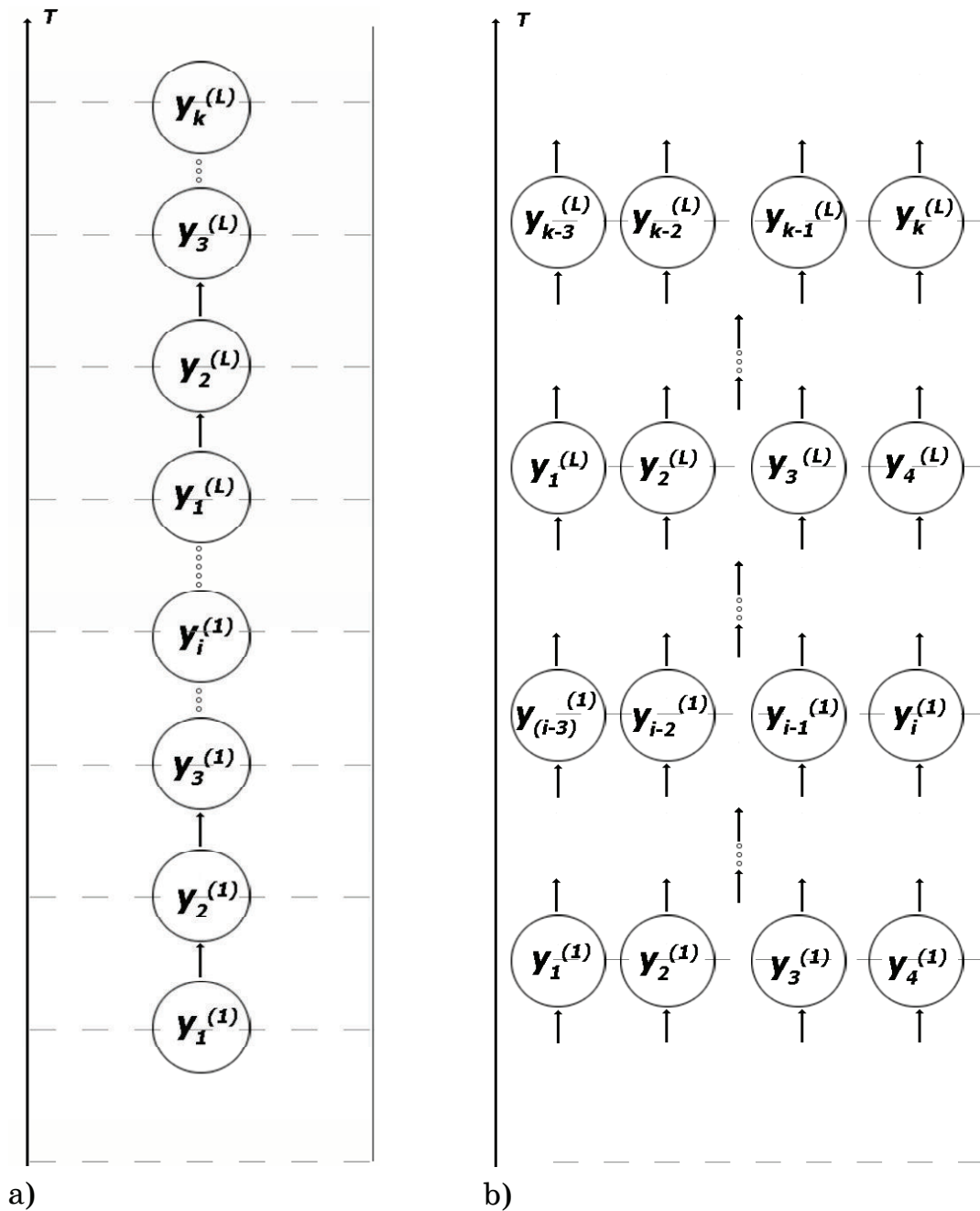


Рисунок 2 – Граф вычисления многослойного персептрона на однопроцессорной а) и на многоядерной системе б)

В общем случае для адаптации на мультиядерной системе предлагается использовать метод «Программирования задачами» - вычисление нейронов путем деления нейромодели на n -частей, где n - количество ядер. На рисунке 3 изображено отображение нейронов скрытого слоя на n -ядерную вычислительную систему

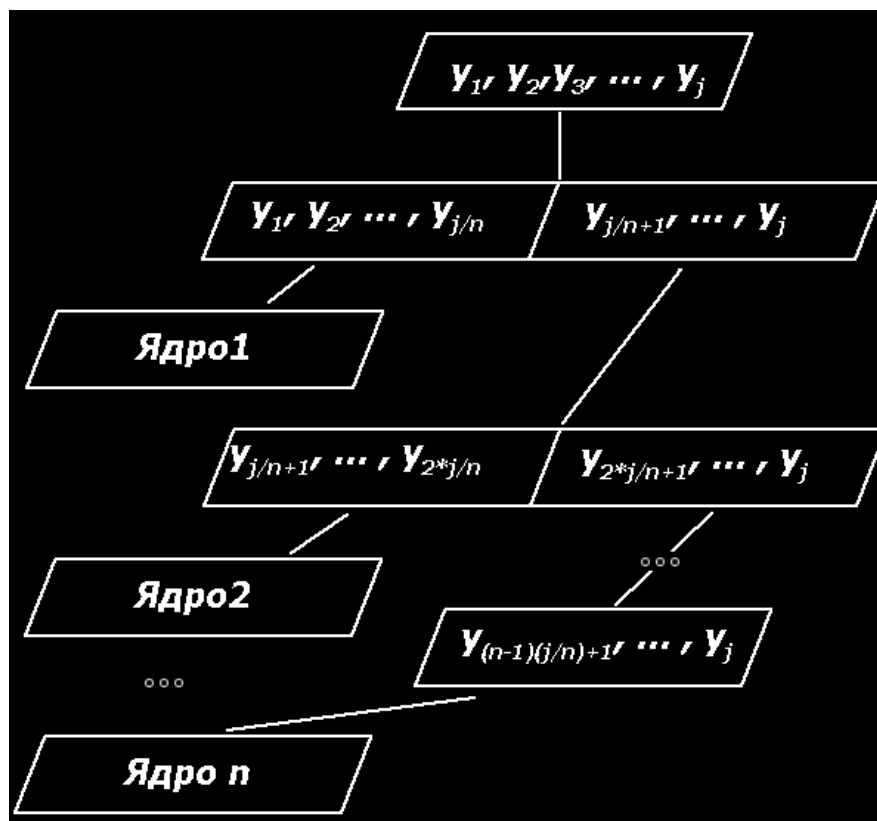


Рисунок 3 - Декомпозиция вычисления нейронов на многоядерной системе

Выводы

В работе предложен метод адаптации моделей персептронного типа на многоядерные вычислительные системы на основе стандарта OpenMP. Адаптация многослойного персептрона на двухядерном процессоре позволяет увеличить производительность в 1,87 раз и в 3,8 раз на четырехядерном процессоре.

Применение данных методик позволит значительно повысить производительность нейросетевых моделей в системах управления нелинейными динамическими объектами.

ЛИТЕРАТУРА

1. Комашинский В.И., Смирнов Д.А. Нейронные сети и их применение в системах управления и связи.- М.: Горячая линия-Телеком, 2003. – 94с.

2. Shameem Akhter, Jason Roberts. Multi-Core Programming. Increasing Performance through Software Multi-threading - ISBN 0-9764832-4-6, 2006.
3. Бухановский А.В., Иванов С.В. Проектирование прикладного математического обеспечения параллельной обработки данных: Учебные материалы Зимней школы- практикума «Технологии параллельного программирования 2006». – Санкт-Петербург – Нижний Новгород, 2006

Получено 11.03.2008 г.