УДК 519.681.5

А.В. Анисимов, М.С. Яджак

ПАРАЛЛЕЛЬНЫЕ АЛГОРИТМЫ ВЫПОЛНЕНИЯ МАССОВЫХ ВЫЧИСЛЕНИЙ В НЕЙРОННЫХ СЕТЯХ

Сделан обзор результатов, касающихся параллельно организации массовых вычислений в нейронных сетях. Рассматриваются дву- и трехмерные многослойные сети прямого распространения с проективными и латеральными связями. Построены параллельные и параллельно-конвейерные алгоритмы, ориентированы на выполнение соответственно на вычислительных системах, реализующих SIMD-режим и квазисистолические вычислительные структуры.

Введение

Нейронные сети (НС) – удобный инструмент для решения многих практически важных задач [1–5]. Среди разнообразия используемых на практике двумерных сетей весьма распространёнными являются многоуровневые клеточные сети [6] для обработки изображений. Структурно их можно представить как трёхмерные многослойные. В отдельных случаях реализация НС осуществляется в последовательном режиме. Это имеет смысл, когда сеть состоит из небольшого числа нейронов и решение задачи можно получить за приемлемое время. Поэтому разработан ряд программных имитаторов НС для ПЭВМ [7–9]. Однако, если количество нейронов сети составляет десятки тысяч и больше или, если задачу необходимо решать в режиме реального времени, тогда возникает проблема оптимизации нейросетевых вычислений. Иногда от решения данной проблемы удавалось «уходить», используя различные аппаратные средства ускорения и реализации нейровычислений [10–13]. Но, ввиду специфики решаемых с их помощью задач и высокой стоимости, данные средства пока не стали широко распространёнными для выполнения массовых вычислений. Поэтому, например, отсутствует информация о том, в какой из прикладных отраслей каждый из разработанных нейрочипов [13, 14] может быть особенно полезным. Следовательно, одним из перспективных путей решения проблемы оптимизации по времени нейросетевых вычислений является использование возможностей распараллеливания обработки информации [15]. Данное направление имеет значительные резервы по усовершенствованию существующих и разработке новых методов и средств выполнения массовых вычислений в зависимости от типа используемой сети. Некоторые общие подходы к параллельной реализации НС рассмотрены в [16–18], где предложены способы распараллеливания вычислений как на внутринейронном (используя потоковые схемы или однородные вычислительные среды), так и на междунейронном уровнях. Однако, ориентация на конкретные типы сетей требует и более конкретных рекомендаций, вплоть до готовых алгоритмов, учитывающих параллелизм обоих уровней. Осуществить это позволяет установленная в [19] взаимосвязь между вычислениями, выполняемыми в НС и при решении задачи цифровой фильтрации (ЗЦФ). В работах [20–22] для решения ЗЦФ рассматриваются и анализируются различные параллельные методы организации вычислений (синхронный, асинхронный, гиперплоскостей Лемпорта, параллельно-конвейерный). В работе [23] подчёркивается конвейерный характер информационных процессов, протекающих в многослойных сетях. На основании изложенного естественным выглядит развитие параллельных методов решения ЗЦФ для эффективной организации массовых вычислений в НС.

Нами рассматриваются дву- и трёхмерные многослойные сети двух типов: 1) прямого распространения (полностью и частично связанные); 2) с проективными и латеральными связями. В сетях первого типа синаптические связи существуют только между нейронами соседних слоёв, а в НС второго типа, кроме некоторого количества синаптических связей между нейронами разных слоёв, существует заданное число связей между нейронами одного и того же слоя, т.е. латеральных.

Решаемая проблема состоит в развитии параллельных методов организации вычислений в одно- и двухмерной ЗЦФ для эффективного выполнения массовых вычислений в дву- и трёхмерных многослойных НС прямого распространения и с проективными и латеральными связями, а также в разработке соответствующих параллельных алгоритмов, ориентированных на реализацию на современных и перспективных высокопроизводительных вычислительных системах.

1. Организация параллельных вычислений в НС прямого распространения

В работах [20–22] рассматривается ЗЦФ, которая в общем случае состоит в выполнении некоторого числа пересчётов сглаживания массива значений переменных через плавающее окно заданного размера. Один из анализируемых в данных работах параллельных методов её решения – синхронный состоит: для вычислений на *v*-ом шаге, необходимые значения переменных берутся исключительно из (*v*-1)-го шага. Предлагаемые для реализации данного метода алгоритмы используют параллельный цикл типа SIM [15]. Если, учитывая результаты [19], в данных алгоритмах количество пересчётов сглаживания, число пересчитанных на *j*-ом шаге значений переменных и размер плавающего окна заменить соответственно на количество слоёв сети *S*, число нейронов N_j в *j*-ом слое, количество синаптических связей нейрона, а также для каждого нейрона произвольного слоя ввести активационную функцию, то получим параллельные алгоритмы выполнения массовых вычислений в многослойных НС прямого распространения. Для двумерной сети такой алгоритм имеет вид [24]:

$$
\begin{aligned}
&FOR \quad j = 1,\ S \quad DO\\
&FOR \quad i = 1,\ N_j \quad DO \quad SIM\\
&p_i = 0\ ;\\
&FOR \quad k = 1,\ N_{j-1} \quad DO\\
&p_i = p_i + w_{ki}^{j} * x_k^{j-1};\\
&x_i^j = \varphi_i^j(p_i).
\end{aligned} \tag{1}
$$

В приведённом алгоритме w_{ki}^j – вес, соответствующий связи между i-м нейроном j-го слоя и k-м нейроном $(j-1)$-го слоя; φ_i^j – активационная функция i-го нейрона j-го слоя; x_l^0 $\left(l = \overline{1,\ N_0}\right)$, x_m^S $\left(m = \overline{1,\ N_S}\right)$ – сигналы, поступающие соответственно на вход и выход сети; x_i^j – значение активационной функции, вычисленное i-м нейроном j-го слоя. В конструкции (1) считается, что количество синаптических связей нейронов j-го слоя равно числу нейронов $(j-1)$-го слоя, т.е. НС является полностью связанной. В случае линейных нейронов последний оператор в конструкции (1) примет вид: $x_i^j = p_i$. Для гомогенной НС имеет место тождество

$$
\begin{aligned}
&\varphi_1^1 \equiv \varphi_2^1 \equiv \ldots \equiv \varphi_{N_1}^1 \equiv \ldots \equiv \varphi_1^S \equiv\\
&\equiv \varphi_2^S \equiv \ldots \equiv \varphi_{N_S}^S \equiv \Phi,
\end{aligned}
$$

поэтому этот оператор в данном случае трансформируется в $x_i^j = \Phi(p_i)$.

На практике применяются и двумерные частично связанные НС прямого распространения. Такая архитектура позволяет заложить в сеть априорные сведения об желаемом законе обработки сигналов. Чтобы это учесть, необходимо в (1) заменить четвёртый оператор на

$$FOR \quad k \in G_i^{j-1} \quad DO,$$

где G_i^{j-1} – некоторое упорядоченное множество и $G_i^{j-1} \subset \left\{k_1 :\ k_1 = \overline{1,\ N_{j-1}}\right\}$.

Согласно приведённых алгоритмов распараллеливание вычислений в сетях осуществляется в основном на междунейронном уровне. Дальнейшее ускорение вычислительного процесса возможно вследствие распараллеливания вычисления значения активационной функции в нейронах. Известно, что процедуру взвешенного суммирования можно распараллелить в режиме, близком к полному двоичному дереву, который во многих ситуациях является оптимальным по времени. Однако, отходя от концепции неограниченного параллелизма, в отдельных случаях весьма эффективным может оказаться подход, состоящий (применительно к полностью связанным сетям) в удалении третьего и замене пятого оператора в (1) фрагментом:

$$
\begin{aligned}
& BEGIN \\
& \quad x_i^j = IF\,(k=1)\quad THEN \quad w_{1\,i}^j * x_1^{j-1} \\
& \quad ELSE \quad x_i^j + p_i\ , \\
& \quad p_i = w_{k+1,\,i}^j * x_{k+1}^{j-1} \\
& \qquad\qquad END\ .
\end{aligned}
$$

Здесь операторы $x_i^j = \dots$ и $p_i = \dots$, разделённые запятой, выполняются синхронно.

Для применения такого подхода к частично связанным сетям необходимо в (1) после удаления третьего оператора заменить пятый оператор на фрагмент:

$$
\begin{aligned}
& BEGIN \\
& \mathrm{x}_\mathrm{i}^\mathrm{j} = \mathrm{IF}(\mathrm{k} = \mathrm{d}_1) \quad \mathrm{THEN} \quad \mathrm{w}_{\mathrm{d}_1\ \mathrm{i}}^\mathrm{j} * \mathrm{x}_{\mathrm{d}_1}^{\mathrm{j-1}} \\
& \mathrm{ELSE} \quad \mathrm{x}_\mathrm{i}^\mathrm{j} + \mathrm{p}_\mathrm{i}\, , \\
& p_i = w_{d_{10},\,i}^j * x_{d_{10}}^{j-1} \\
& END\, .
\end{aligned}
$$

Здесь d_1 – первый элемент G_i^{j-1}, а d_{10} – элемент данного множества, следующий за элементом, определяемым значением переменной k.

В многослойных НС прямого распространения с большим количеством нейронов доля процедуры реализации активационной функции простого вида (линейная, пороговая) в общем времени выполнения нейросетевых вычислений является незначительной. Если используются более сложные функции активации нейронов (сигмоидная, гиперболический тангенс, арктангенс, синус), то: 1) согласно [25] их можно представить в виде степенного ряда и использовать в дальнейшем лишь четыре его члена; время реализации такого выражения можно уменьшить путём его распараллеливания; 2) для быстрой реализации таких функций подключать специальные аппаратные средства.

Параллельный алгоритм организации массовых вычислений в трёхмерных полностью связанных многослойных НС прямого распространения выглядит таким образом:

$$
\begin{aligned}
& FOR \quad t = 1,\, S \quad DO\, , \\
& FOR \quad (i,\, j) \in \left\{ (i_1,\, j_1):\ i_1 = \overline{1,\, N_t^1}\ ; \quad j_1 = \overline{1,\, M_t^1} \right\} \\
& \ DO \quad SIM \\
& FOR \quad r = 1,\, N_{t-1}^1 \quad DO\, , \\
& \qquad\qquad FOR \quad s = 1,\, M_{t-1}^1 \quad DO\, ,
\end{aligned}
$$

$$
\begin{aligned}
& BEGIN \\
& x_{i\,j}^t = IF \;\; (r = 1 \wedge s = 1) \\
& THEN \quad w_{11,\,i\,j}^t * x_{11}^{t-1} \\
& \ ELSE \quad x_{i\,j}^t + p_{i\,j}\, . \\
& \\
& p_{i\,j} = IF\ (s = M_{t-1}^1)\ THEN\ \ w_{r+1\,1,\,i\,j}^t * x_{r+1\,1}^{t-1} \\
& ELSE\ \ w_{r\ s+1,\,i\,j}^t * x_{r\ s+1}^{t-1} \\
& \qquad\qquad END; \\
& \qquad\qquad\quad x_{i\,j}^t = \psi_{i\,j}^t(x_{i\,j}^t).
\end{aligned}
\tag{2}
$$

В приведённой конструкции количество нейронов t-го слоя $N_t = N_t^1 \cdot M_t^1$; $w_{rs,\,i\,j}^t$ – вес, отвечающий связи между $(r,\, s)$-м нейроном $(t-1)$-го слоя и $(i,\, j)$-м нейроном t-гослоя;

$x_{l\,m}^0 \left(l = \overline{1,\, N_0^1};\ m = \overline{1,\, M_0^1} \right)$, $x_{cu}^S \left(c = \overline{1,\, N_S^1};\ u = \overline{1,\, M_S^1} \right)$ – сигналы, подающиеся соответственно на вход и выход сети; $x_{i\,j}^t$ – значение активационной функции, вычисленное $(i,\, j)$-м нейроном t-го слоя; $\psi_{i\,j}^t$ – активационная функция $(i,\, j)$-го нейрона t-го слоя. В конструкции (2) операторы

$x^t_{i\,j}=...$ и $p_{i\,j}=...$, разделённые запятой, выполняются синхронно.

Аналогично, как и для двумерных сетей, строится параллельный алгоритм организации массовых вычислений в трёхмерных частично связанных НС прямого распространения, который детально описан в [26].

Приведённые алгоритмы ориентированы на выполнение на параллельных вычислительных системах типа *SIMD,* т.е. последние наиболее естественно могут реализовать рассмотренные типы НС прямого распространения. Однако, если имеется механизм имитации *SIMD*-режима, это возможно и на любых других ЭВМ, включая и обычные последовательные.

2. Вычисления в НС с проективными и латеральными связями

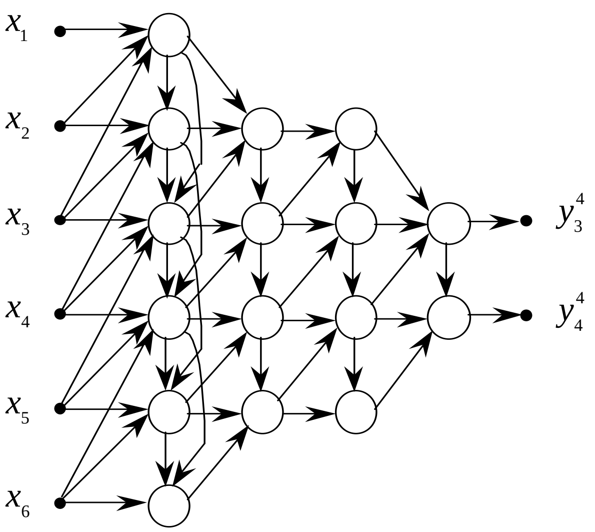
В работах [20–22] предлагаются и анализируются параллельно-конвейерные методы решения ЗЦФ. Согласно этих методов вычисления задаются в виде некоторого числа сдвинутых между собой ветвей. Каждая ветвь взаимодействует с определённым количеством своих соседей. Параллельно-конвейерные алгоритмы (ПКА), построенные на основе данных методов, используют параллельный цикл типа *SYNCH* [15]. Если в этих алгоритмах количество пересчитываемых значений переменных, число пересчётов сглаживания, размер плавающего окна заменить соответственно на $N^0=\max\limits_{1\le j\le S}\{N_j\}$; некоторый вспомогательный масив, определяющий структуру сети; количество синаптических связей, а также ввести для каждого нейрона активационную функцию и учесть количество тактов её реализации в операторах задержки *DELAY*, то получим ПКА организации массовых вычислений в многослойных НС с проективными и латеральными связями. Для двумерной сети такой алгоритм имеет вид [24]:

$$
\begin{array}{l}
FOR\quad i=1,\,N^0\quad DO\quad SYNCH;\\
DELAY\quad ((2+D)\cdot(i-1));\\
FOR\quad j=1,\,S_i\quad DO\quad;\\
BEGIN;\\
FOR\quad k=1,\,M_j\quad DO;\\
BEGIN;\\
x_i=IF\ (k=1)\quad THEN\quad v^j_{i\,i}*x_i ELSE\,x_i+p_i\\
p_i=v^j_{i+h(k,\ m_j),\,i}*x_{i+h(k,\ m_j)}\\
END;\\
x_i=f_j(x_i)\\
y^j_i=x_i\\
DELAY\ (2+D\cdot m_{j+1});\\
END.
\end{array}
\qquad (3)
$$

В конструкции (3) величина D – время реализации функции активации нейронов; $S_i\left(i=\overline{1,\,N^0}\right)$ – вспомогательный массив, к $S=\max\limits_{1\le i\le N^0}\{S_i\}$; M_j – количество синаптических связей нейронов j-го слоя, к $M_j=2m_j+1$, где $m_j\in\mathrm{N}$; операторы $x_i=...$ и $p_i=...$, разделённые запятой, выполняются синхронно; $v^j_{n\,i}$ – вес, соответствующий связи между n-м нейроном $(j-1)$-го слоя и i-м нейроном j-го слоя, если $n\ge i$, а также связи между n-м и i-м нейронами j-го слоя, если $n<i$; $x_l\ \left(l=\overline{1,\ N^0}\right)$, $y^S_i\ (\forall i:S_i=S)$ – соответственно входной и выходной сигналы НС; y^j_i – значение активационной функции i-го нейрона j-го слоя; функция $h(k,\,m_j)$ – условный оператор $IF\ \left(k/2=[k/2]\ \right),\quad THEN,\ k/2-m_j-1,\ ELSE\ \ (k+1)/2$; f_j – активационная функция нейронов j-го слоя. Запись $[a]$ обозначает целую часть числа a.

Рассмотрим пример архитектуры НС, вычисления в которой могут быть выполнены с помощью (3).

Для $N^0=6$; $S_1=S_6=1,\ S_2=S_5=3,\ S_3=S_4=4$; $M_1=5,\ M_2=M_3=M_4=3$ она показана на рисунке.



Рисунок

Согласно алгоритму (3) распараллеливание обработки информации в сети осуществляется как на междунейронном, так и на внутринейронном уровнях.

В работе [27] решена проблема организации параллельных вычислений в двумерной многослойной НС с проективными и латеральными связями в случае, когда время реализации активационной функции каждого из нейронов слоя, вообще говоря, является разным. Тогда алгоритм (3) немного изменится, прежде всего операторы задержки.

Первый выглядит так:

$$DELAY\ ((2+D^1)\cdot(i-1)),$$

второй – $DELAY\ (IF\ (D^{j+1} > D^j)\ \ THEN$
$2+D^j\cdot m_{j+1}+(D^{j+1}-D^j)\cdot(i-I_{j+1})$
$ELSE\ \ 2+D^{j+1}\cdot m_{j+1}+$
$+(D^j-D^{j+1})\cdot(P^j-i))$.

Здесь
$\forall j\,(j=\overline{1,S}):\ D^j=\max\{D_i^j:\ i\in\{i':S_{i'}\ge j\}\}$,
где D_i^j – время реализации активационной функции i-го нейрона j-го слоя; I_j, P^j – соответственно минимальный и максимальный номера нейронов j-го слоя. Отметим, что данный оператор задержки приведен при условии $m_{j+1}-P^j+P^{j+1}\le 0$. Кроме этого, десятый оператор в (3) будет выглядеть таким образом: $x_i=g_i^j(x_i)$, где g_i^j – активационная функция i-го нейрона j-го слоя.

Для организации массовых вычислений в трёхмерной НС с проективными и латеральными связями предлагается следующий ПКА [26]:

$FOR\ \ (i,j)\in\{(i_1,j_1):\ i_1=\overline{1,N_1};\ j_1=\overline{1,N_2}\}$
$DO\ \ SYNCH$
$DELAY((4n+2+(n+1)\cdot D)\cdot(i-1)+$
$+(2+D)\cdot(j-1))$
$FOR\ \ k=1,S_{i\,j}\ \ DO$
$BEGIN$
$FOR\ \ l=1,V_k\ \ DO$
$BEGIN$
x_{ij}=IF(l=1)
THEN$v^k_{ij,ij}*x_{ij}$ (4)
ELSE$x_{ij}+p_{ij}$,
$pij=v^k_{i+h(l,m_k,n)\,j+p(l,n),ij*}x_{i+h(l,m_r,n)\,j+p(l,n)}$
END;
$x_{i\,j}=f_k(x_{i\,j})$
$y_{i\,j}^k=x_{i\,j}$
$DELAY\ (D\cdot m_{k+1}\cdot(n+1)+D\cdot n+2)$;
END.

Здесь операторы $x_{i\,j}=\ldots$ и $p_{i\,j}=\ldots$, разделённые запятой, выполняются в синхронном режиме. Кроме этого, $N^0=N_1\cdot N_2$, где $N_1=\max\{N_t^1:\ t=\overline{1,S}\}$, $N_2=\max\{M_t^1:\ t=\overline{1,S}\}$;
$S_{i\,j}\ (i=\overline{1,N_1};\ j=\overline{1,N_2})$ – вспомогательный массив, определяющий структуру сети, к $S=\max\{S_{i\,j}:\ i=\overline{1,N_1};\ j=\overline{1,N_2}\}$;
$V_k=(2m_k+1)\cdot(2n+1)$, где $m_k,\ n\in\mathrm{N}$;
$v_{i\,j,\,r\,s}^k$ – вес, соответствующий связи между (i,j)-м и (r,s)-м нейронами k-го слоя, если $(i,j)\in\{(i_1,j_1):\ i_1\le r;\ j_1<s\}\cup\{(i_2,j_2):\ i_2<r;\ j_2\ge s\}$, или же связи между (i,j)-м нейроном $(k-1)$-го и

(r,s)-м нейроном k-го слоя при условии, что $(i,j)\in\{(i_1,j_1): i_1\ge r;\ j_1\ge s\}\cup\{(i_2,j_2): i_2>r;\ j_2<s\}$; $x_{i\,j}\ (i=\overline{1,N_1};\ j=\overline{1,N_2}),\ y_{i\,j}^{S}\ (\forall i,j: S_{i\,j}=S)$ – соответственно входной и выходной сигналы НС; $y_{i\,j}^{k}$ – значение активационной функции (i,j)-го нейрона k-го слоя. Функции $h(l,m_k,n)$ и $p(l,n)$ – соответственно такие условные операторы:

$IF\,(l/2=[l/2])\ THEN\,[(l-1)/(4n+2)]-m_k$
$ELSE\ \ [(l+2n+2)/(4n+2)]$;

$IF\ (l/2=[l/2])\ \ THEN$
$l/2-n-1-(2n+1)\cdot[(l-1)/(4n+2)]$

$ELSE\ \ l/2+0.5-(2n+1)\cdot[(l+2n+2)/(4n+2)]$.

В алгоритме (4) считается, что активационные функции нейронов различных слоёв имеют одинаковое время реализации. В противном случае этот алгоритм можно использовать, приняв $D=\max\{D_j:\ j=\overline{1,S}\}$, где D_j – количество тактов, необходимых для выполнения f_j.

Приведённые ПКА (3), (4) работают при условии, что $\forall j\ (j=\overline{1,S-1})$: $N_{j+1}\le N_j$. Они ориентированы на реализацию на специализированных вычислительных системах – квазисистолических структурах. В работе [28] такие структуры приведены для реализации массовых вычислений в двумерных многослойных НС с проективными и латеральными связями. Главное их отличие от чисто систолических состоит в том, что допускается одновременная передача данных из одной «инстанции» сразу в несколько «точек приёма». При этом учитываются различные типы и способы задания активационных функций нейронов.

3. Анализ параллельных вычислений

Будем считать, что время реализации операций сложения и умножения одинаково и равно одному такту. Выполнение вычислений в полностью связанных многослойных дву- и трёхмерной НС прямого распространения в последовательном режиме требует соответственно

$$\sum_{j=1}^{S}\left(2N_{j-1}\cdot N_j+\sum_{i=1}^{N_j}T_i^{\,j}\right)\quad \text{и}$$

$$\sum_{t=1}^{S}\left(2N_{t-1}^{1}\cdot M_{t-1}^{1}\cdot N_t^{1}\cdot M_t^{1}+\sum_{i=1}^{N_t^1}\sum_{j=1}^{M_t^1}U_{i\,j}^{\,t}\right)\qquad (5)$$

тактов, где $T_i^{\,j}$, $U_{i\,j}^{\,t}$ – времена реализации активационных функций i-го нейрона j-го слоя в двумерной сети и (i,j)-го нейрона t-го слоя в трёхмерной НС. Используя алгоритмы (1), (2), оценки (5) можно уменьшить соответственно до

$$\sum_{j=1}^{S}\left(2N_{j-1}+t^{\,j}\right)\quad \text{и}\quad \sum_{t=1}^{S}\left(N_{t-1}^{1}\cdot M_{t-1}^{1}+D_{0}^{\,t}\right),\ (6)$$

где $t^{\,j}=\max\{T_i^{\,j}:\ i=\overline{1,N_j}\}$,
$D_0^t=\max\{U_{i\,j}^{t}:\ i=\overline{1,N_t^1};\ j=\overline{1,M_t^1}\}$.

Если предположить, что каждый слой сети содержит N_e нейронов, а время реализации активационной функции нейронов j-го слоя равно $t_e^{\,j}$, то на основании (5), (6) получим, что ускорение алгоритма (1) равно N_e, а ускорение (2) равно $2N_e$ (эта оценка получена при условии, что $\forall j\ (j=\overline{1,S})$: $t_e^{\,j}<<N_e$). Приведённые оценки ускорения будут иметь место и в случае, когда количество синаптических связей K_j нейронов j-го слоя сети меньше N_{j-1}, т.е. когда НС уже не полностью, а частично связанная. При этом будем считать, что $\forall j\ (j=\overline{1,S})$: $K_j>>t_e^{\,j}$.

Если для полностью связанной НС прямого распространения предположить, что $\forall j\ (j=\overline{1,S})$: $N_j=N_0/2^{\,j}$, причём $N_0\ge 2^S$, а время реализации активационной функции нейронов равно t_0, причём $t_0<<N_0$, то в этом случае ускорения алгоритмов (1), (2) будут соответственно равны $N_0/3$ и $2N_0/3$ (здесь $N_0=N_0^1\cdot M_0^1$).

Если в (1), (2) для распараллеливания процедуры взвешенного суммирования использовать режим, близкий к полному двоичному дереву, то оценки (6) уменьшатся соответственно до

$$S+\sum_{j=1}^{S}\left(\log N_{j-1}+t^{j}\right) \quad \text{и}$$

$$S+\sum_{t=1}^{S}\left(\log(N_{t-1}^{1}\cdot M_{t-1}^{1})+D_{0}^{t}\right).$$

Алгоритмы (3), (4) для своего выполнения требуют соответственно

$$(2+D)\cdot(A+S-2)+M_{S}+D+$$
$$+\sum_{j=1}^{S-1}(M_{j}+D\cdot m_{j+1})$$

и

$$((4\mathrm{n}+2)+(\mathrm{n}+1)\cdot\mathrm{D})\cdot(\mathrm{I}-1)+$$
$$(2+\mathrm{D})\cdot(\mathrm{J}-1)+\mathrm{V_S}+\mathrm{D}+ \tag{7}$$
$$+\sum_{k=1}^{S-1}\left(V_{k}+D\cdot(m_{k+1}+1)\cdot(n+1)+2\right)$$

тактов. Здесь $A=\max\{i:\ S_{i}=S\}$, а также $\exists\ j$ такое, что $I=\max\{i':\ S_{i'\,j}=S\}$ и $\exists\ i$ такое, что $J=\max\{j':\ S_{i\ j'}=S\}$.

Последовательное выполнение вычислений в дву- и трёхмерной многослойных НС с проективными и латеральными связями требует соответственно

$$\sum_{i=1}^{N^{0}}\ \sum_{j=1}^{S_{i}}(2M_{j}+D-1) \quad \text{и}$$

$$\sum_{i=1}^{N_{1}}\sum_{j=1}^{N_{2}}\sum_{k=1}^{S_{i\,j}}(2V_{k}+D-1) \tag{8}$$

такта.

Для двумерной сети, в которой $N^{0}=1000,\ S=300$ и $S_{i}(i=\overline{1,\ 300})=i$, $S_{i}(i=\overline{301,\ 700})=S$, $S_{i}(i=\overline{701,1000})=N^{0}-i+1$ в предположении, что $\forall j\ (j=\overline{1,S})$: $M_{j}=M$ при $D=1$, $M=3$ на основании первых оценок в (7), (8) получим, что ускорение алгоритма (3) равно 300.64, а при $D=7,\ M=11$ оно несколько уменьшается и равно 258.68. С увеличением числа синаптических связей, например, при $M=21$ ускорение немного увеличится и будет равно 278.63. Рассмотрим сеть, в которой количество нейронов больше, чем в предыдущих примерах:

$N^{0}=10000,\ S=2000,\ S_{i}(i=\overline{1,\ 2000})=i$, $S_{i}(i=\overline{2001,\ 8000})=S$, $S_{i}(i=\overline{8001,10000})=$ $=N^{0}-i+1$. При $D=1,\ M=3$ получим, что ускорение алгоритма (3) в данном случае равно 2526.83. Следовательно, с увеличением количества нейронов сети ускорение ПКА, реализующего в ней массовые вычисления, тоже увеличивается.

В работе [26] приведены численные оценки ускорения алгоритма (4) для примеров трёхмерных многослойных НС с проективными и латеральными связями. Эти оценки подтверждают эффективность предложенного ПКА выполнения массовых вычислений.

Заключение. В работе развиты параллельные методы решения ЗЦФ для организации вычислений в дву- и трёхмерных многослойных НС прямого распространения и с проективными и латеральными связями. В результате предложены параллельные алгоритмы, ориентированы на реализацию на вычислительных системах, реализующих или имитирующих *SIMD*-режим, и квазисистолических вычислительных структурах.

Целесообразность использования предложенных алгоритмов подтверждена оценками их ускорения.

1. *Дубровин В.И., Морщавка С.В., Пиза Д.М. и др.* Распознавание растений по результатам дистанционного зондирования на основе многослойных нейронных сетей // Математичні машини і системи. – 2000. – № 2/3. – С. 113–119.
2. *Куссуль Э.М., Касаткина Л.М., Байдык Т.Н. и др.* Нейросетевые технологии распознавания рукописных текстов // Управляющие системы и машины. – 2001. – № 2. – С. 64–83.
3. *Годич О., Щербина Ю.* Застосування штучних нейронних мереж до розв'язування нелінійних задач про найменші квадрати // Вісн. Львів. ун-ту. Сер. прикл. математика та інформатика. – 2003. – Вип. 6. – С. 182–190.

4. *Chen K.-z., Leung Y., Leung K.S., Gao X.-b.* A Neural Network for Solving Nonlinear Programming Problems // Neural Computing and Applications. – 2002. – 11. – P. 103–111.
5. *Valensa M., Ludermir T.* Constructive neural networks for function approximation // Proc. of Internat. conf. on inductive modelling ICIM. – Lviv; 20–25 May 2002. – 4. – P. 22–27.
6. *Aizenberg N.N., Aizenberg I.N.* CNN–like networks based on multi–valued and universal binary neurons: learning and application to image processing // Proc. of Third IEEE Inter. Workshop on Cellular Neural Networks and Their Applications, Rome, 1994, IEEE94 TH0693–2. – P.153–158.
7. *Горбань А.Н., Россиев Д.А., Бутакова Е.В. и др.* Медицинские и физиологические применения нейроимитатора «MultiNeuron» // Нейроинформатика и её приложения: Материалы III Всерос. семинара, 6–8 октября 1995. – Красноярск: КГТУ, 1995. – Ч. 1. – С. 101–113.
8. *Резник А.М., Калина Е.А., Сычёв А.С. и др.* Многофункциональный нейрокомпьютер NEUROLAND // Математичні машини і системи. – 2003. – № 1. – С. 36–45.
9. *Гриценко В.И., Мисуно И.С., Рачковский Д.А. и др.* Концепция и архитектура программного нейрокомпьютера SNC // Управляющие системы и машины. – 2004. – № 3. – С. 3–14.
10. *Перспективные* системы обработки информации / Под ред. Я.А. Дуброва. – Львов, 1990. – 59 с. – (Препр. / АН УССР. Ин-т прикладных проблем механики и математики;90–6).
11. *Егоров В.М.* Трёхмерные нейроподобные оптические вычислительные структуры // Автометрия. – 1993. – № 3. – С. 38–43.
12. *Борисов Ю., Кашкаров В., Сорокин С.* Нейросетевые методы обработки информации и средства их программно-аппаратной поддержки // Открытые системы. – 1997.– № 4. – С. 38–40.
13. *Логовский А.* Технология ПЛИС и её применение для создания нейрочипов // Открытые системы. – 2000. – №10. http://www.osp.ru/os/2000/10/019.htm).
14. *Шахнов В., Власов А., Кузнецов А.* Элементная база параллельных вычислений // Открытые системы. – 2001. – № 5/6. http://www.osp.ru/os/2001/05-06/017.htm).
15. *Вальковский В.А.* Распараллеливание алгоритмов и программ. Структурный подход. – М.: Радио и связь, 1989. – 176 с.
16. *Вальковський В.О., Курбацький О.М., Полєтаєв С.М.* Тотальне розпаралелювання обчислень у нейронних мережах // Волинський математичний вісник. – 1997. – Вип. 4. – С. 22–24.
17. *Вальковський В.О., Фарід Т.М.* Про багатоваріантну обробку сигналів у нейроні // Вісник Держ. ун-ту «Львів. політехніка». Комп'ютерна інженерія та інформ. технології. – 1998. – № 349. – С. 39–45.
18. *Hrytsyk V.V., Aizenberg N.N., Bun R.A. and others.* The neural and neural-like networks: synthesis, realization, application and future // Information technologies and systems. – 1998. – N 1/2. – P. 15–55.
19. *Вальковский В.А.* К проблеме взаимосвязи между нейронными сетями и задачей цифровой фильтрации // Нейронные сети и модели: Тр. междунар. науч.-техн. конф. «Нейронные, реляторные и непрерывнологические сети и модели», Ульяновск, 19-21 мая 1998. – Т. 1. – С. 39–41.
20. *Val'kovskii V.A.* An optimal algorithm for solving the problem of digital filtering // Pattern Recognition and Image Analysis. – 1994. – 4, N 3. – P. 241–247.
21. *Вальковский В.А., Яджак М.С.* Оптимальный алгоритм решения двумерной задачи цифровой фильтрации // Проблемы управления и информатики. – 1999. – № 6. – С. 92–102.
22. *Яджак М.С.* Об оптимальном в одном классе алгоритме решения трёхмерной задачи цифровой фильтрации // Там же. – 2000. – № 6. – С. 66–81.
23. *Захаров И.С., Лопин В.Н.* Конвейерные информационные процессы в многослойных нейронных сетях // Радіоелектроніка, інформатика, управління. – 1999. – № 1. – С. 53–56.
24. *Яджак М.* Застосування алгоритмів розв'язування задач цифрової фільтрації для реалізації обчислень у нейронних мережах // Вісн. Львів. ун-ту. Сер. прикл. математика та інформатика. – 2003. – Вип. 6. – С. 227–232.
25. *Бодянский Е.В., Кулишова Н.Е., Руденко О.Г.* Обобщённый алгоритм обучения формального нейрона // Кибернетика и системный анализ. – 2002. – № 5. – С. 176–182.

26. *Вальковський В.О., Яджак М.С.* Про організацію паралельних обчислень у нейронних мережах // Відбір і обробка інформації. – 2003. – Вип. 19(95). – С. 138–144.
27. *Яджак М.С.* Паралельна організація обчислень у нейронних мережах одного типу // Вісн. Львів. ун-ту. Сер. прикл. математика та інформатика. – 2004. – Вип. 9. – С. 204–210.
28. *Яджак М.С.* Моделювання нейронних мереж з проективними та латеральними зв'язками на квазісистолічних структурах // Відбір і обробка інформації. – 2005. – Вип. 23(99). – С. 122–127.

Получено 10.01.2008

Об авторах:

Анисимов Анатолий Васильевич,
доктор физико-математических наук, профессор, декан факультета кибернетики,

Яджак Михаил Степанович,
кандидат физико-математических наук, старший научный сотрудник.

Место работы авторов:

Киевский национальный университет имени Тараса Шевченко
03187, Проспект Академика Глушкова 2, корп. 6, к. 29.
Тел.: 8(044)521 3554;
E-mail ava@unicyb.kiev.ua

Институт прикладных проблем механики и математики НАН Украины им. Я.С. Пстригача
79053, Львов, ул. Троллейбусная, буд. 14, кв. 125.
Тел.: 8(032)239 9969; 239 9914.
e-mail: dept@iapmm.lviv.ua
yadzhak@zadarma.com