

РАЗРАБОТКА НЕЙРОКОМПЬЮТЕРА С БИНАРНОЙ МАТРИЦЕЙ ПАМЯТИ

*М.И.Дябин, Н.Г.Карпинский, Э.М.Куссуль, А.И.Половянюк,
В.Г.Редько, В.А.Сержеченко, О.В.Ургант*

Сообщается о разработке нейрокомпьютера, основанного на использовании автоассоциативной нейроподобной памяти с бинарными весами синапсов, изготовленного из стандартных отечественных интегральных схем. Излагается алгоритм распознавания образов, реализуемый в нейрокомпьютере. Описаны функциональная схема и принцип работы нейрокомпьютера. Приведены характеристики нейрокомпьютера, включая оценки времени обучения и распознавания.

Цель настоящей статьи — охарактеризовать состояние разработок нейрокомпьютера с бинарной матрицей памяти, проводимых на предприятии ИнТеСи при содействии Зеленоградского предприятия НПО "Вертекс". Разрабатываемый нейрокомпьютер предназначен для распознавания образов по алгоритму, предложенному Э.М.Куссулем с сотрудниками [1-3].

Распознавание образов предусматривает режим обучения, в котором входные образы предъявляются в виде набора признаков, и для каждого образа указывается класс, к которому этот образ принадлежит, и режим распознавания, в котором по предъявленному образу устанавливается его принадлежность к определенному классу (подробнее о методах распознавания образов см., например, [4]).

В алгоритмах [1-3] предусмотрена специальная процедура кодирования образов в форме бинарных векторов (компоненты векторов принимают значения 0 или 1), осуществляемая так, что близким образам соответствуют близкие (в смысле расстояния по Хеммингу) бинарные векторы. Классы также кодируются с помощью бинарных векторов. Соответствие между образами и классами запоминается в совокупности бинарных синаптических весов нейронной сети (см.ниже). Множество синаптических весов образуют бинарную матрицу памяти. Бинарное представление хранимой и обрабатываемой информации позволяет эффективно реализовать алгоритмы [1-3] аппаратно с помощью цифровых интегральных микросхем.

Описанная ниже схема нейрокомпьютера основана на использовании оперативного запоминающего устройства (ОЗУ) большой емкости, в котором запоминается принадлежность образов определенным классам в форме бинарной матрицы памяти, и поразрядного параллельного арифметико-логического устройства (АЛУ).

1. АЛГОРИТМ РАСПОЗНАВАНИЯ ОБРАЗОВ В НЕЙРОКОМПЬЮТЕРЕ С БИНАРНОЙ МАТРИЦЕЙ ПАМЯТИ

Алгоритм предусматривает три режима работы нейрокомпьютера:

- a) режим обучения;*
- б) режим распознавания;*
- в) режим распознавания с коррекцией.*

В режиме обучения предъявляются образы, закодированные в виде бинарных векторов \mathbf{X}_F , и указывается к какому классу эти образы принадлежат. Каждый класс кодируется своим бинарным вектором \mathbf{C}_k . Векторы \mathbf{X}_F и \mathbf{C}_k логически суммируются: $\mathbf{X} = \mathbf{X}_F \vee \mathbf{C}_k$. По векторам \mathbf{X} , соответствующим всем образам из обучающей выборки, формируется бинарная матрица памяти $\|w_{ij}\|$:

$$w'_{ij} = w_{ij} \vee (X_i \wedge X_j \wedge h_{ij}), \quad (1)$$

где w_{ij} и w'_{ij} — элементы матрицы до и после предъявления рассматриваемого образа, h_{ij} — случайные бинарные величины с заданной вероятностью единичных значений, определяющей интенсивность обучения. При этом в матрице $\|w_{ij}\|$, характеризующей бинарные веса связей между нейронами, формируется память о принадлежности входных образов определенным классам. Размерность всех векторов равна числу нейронов в сети n .

В режиме распознавания по предъявленному образу определяется класс, к которому он принадлежит. При этом нейронная сеть функционирует как автоассоциативная память, т.е. по векторам \mathbf{X}_F (характеризующим предъявленный образ) восстанавливаются векторы $\mathbf{X}' \approx \mathbf{X}_F \vee \mathbf{C}_k$.

Функционирование нейронной сети осуществляется путем подачи на сеть входного вектора, подсчета суммарного воздействия на каждый нейрон (с учетом весов w_{ij}) и сравнения этих воздействий с порогом, в результате чего получается выходной вектор \mathbf{X}' :

$$X'_i = \Theta(\text{net}_i - K), \quad \text{net}_i = \sum_{j=1}^n w_{ij} X_j, \quad (2)$$

где X_i и X'_i — компоненты входного и выходного векторов, K — порог нейронов, $\Theta(\cdot)$ — тета-функция. Далее подсчитывается число единичных компонент выходного вектора, т.е. определяется суммарная выходная активность сети, и если она не достигла заданного значения, то уменьшается порог нейронов, после чего вновь определяется выходной вектор.

Для того, чтобы восстановление векторов $\mathbf{X}_F \vee \mathbf{C}_k$ было по возможности максимально точным, осуществляется несколько итераций нейронной сети (полученный на некоторой итерации выходной вектор является

входным вектором для следующей итерации). На первой итерации на вход сети подается вектор \mathbf{X}_F .

По результирующему выходному вектору \mathbf{X}' определяется номер класса входного образа, как номер класса, для которого максимально скалярное произведение $(\mathbf{X}', \mathbf{C}_k)$:

$$k_N = \arg \max_k \{(\mathbf{X}', \mathbf{C}_k)\}. \quad (3)$$

Номер класса k_N является результатом распознавания.

В режиме распознавания с коррекцией происходит распознавание под контролем учителя, и если номер класса определен нейронной сетью неправильно ($k_N \neq k_T$, k_N и k_T — номера классов, определенные нейронной сетью и учителем), то происходит коррекция матрицы памяти $||\mathbf{w}_{ij}||$, содержащая разобучение и дообучение. При разобучении входной вектор \mathbf{X}_F объединяется с вектором неправильного класса $\mathbf{X} = \mathbf{X}_F \vee \mathbf{C}_{k_N}$ и "вычеркиваются" (с определенной вероятностью) соответствующие единицы в матрице памяти:

$$w'_{ij} = w_{ij} \sqrt{(X_i \wedge X_j \wedge h_{ij})}, \quad (4)$$

где h_{ij} — случайные бинарные величины, вероятность единичных значений которых определяет интенсивность разобучения, w_{ij} , w'_{ij} — элементы матрицы-памяти до и после ее модификации. При дообучении входной вектор объединяется с вектором правильного класса $\mathbf{X} = \mathbf{X}_F \vee \mathbf{C}_{k_T}$ и происходит такая же модификация матрицы памяти как и при обучении.

2. ФУНКЦИОНАЛЬНАЯ СХЕМА И ПРИНЦИП РАБОТЫ НЕЙРОКОМПЬЮТЕРА

Функциональная схема нейрокомпьютера представлена на рисунке.

Нейрокомпьютер содержит блок параллельной обработки информации (ПОИ) и устройство управления (УУ). Через УУ нейрокомпьютер связан с ЭВМ.

В блок ПОИ входят оперативное запоминающее устройство (ОЗУ), арифметико-логическое устройство (АЛУ), регистр сдвига (РС), буферные регистры БР1, БР2 и счетчик (С).

В реализованном варианте нейрокомпьютера блок ПОИ выполнен в виде блока, состоящего из 16 плат. На каждой из плат производятся операции над 64 разрядами 1024-х разрядных бинарных векторов. Увеличение размерности обрабатываемых векторов может быть обеспечено включением в нейрокомпьютер дополнительных плат (добавление одной платы увеличивает размерность вектора на 64 разряда).

Информационная емкость ОЗУ составляет 64 Мбита. ОЗУ организовано в виде матрицы 1024×65536 бита. Число разрядов в буферных

registрах, регистре сдвига и в АЛУ равно числу строк матрицы ОЗУ $n_y = 1024$ (рисунок).

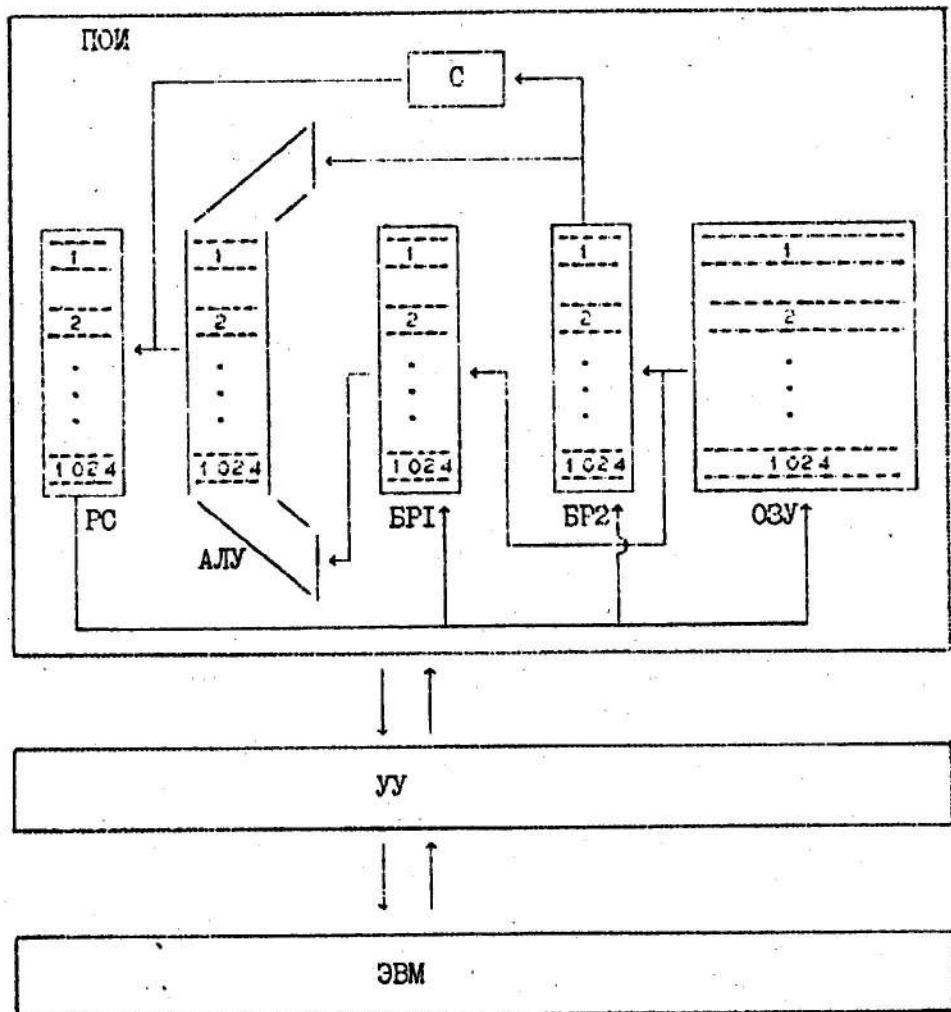


Рис. 1. Функциональная схема нейрокомпьютера.

Блок ПОИ осуществляет следующие операции над бинарными векторами размерности n_y : логические операции "И", "ИЛИ", "НЕ", "Исключенное ИЛИ" (выполняются в АЛУ), циклический сдвиг бинарного вектора (выполняется с помощью РС), пересылка вектора из заданного столбца ОЗУ в буферные регистры, пересылка вектора из регистра сдвига в какой-либо буферный регистр или в заданный столбец ОЗУ. Указанные операции выполняются одновременно для всех n_y разрядов.

Управление работой блока ПОИ осуществляется управляющей программой, написанной на специально разработанном языке "Нейроассем-

блер". Программа хранится в ЭВМ и частично в УУ. УУ формирует и подает команды программы в блок ПОИ. Таким образом, с помощью УУ в блоке ПОИ программным способом выполняются произвольные операции над бинарными векторами размерности n_y .

Кроме описанного программного способа работы для ускорения функционирования в режиме распознавания в нейрокомпьютере предусмотрены две операции, выполняемые аппаратным способом:

1) подсчет суммарного числа единиц бинарных величин, хранящихся в строках ОЗУ (осуществляется с помощью счетчика С, предназначен для определения значений net_i , см. формулу (2));

2) сравнение суммарного числа единиц с порогом.

В настоящем варианте нейрокомпьютера бинарная матрица памяти $||\mathbf{w}_{ij}||$ хранится в ОЗУ. Формирование матрицы $||\mathbf{w}_{ij}||$ в соответствии с (1), (4) осуществляется программным способом. Если размерность векторов \mathbf{X} превышает $n_y = 1024$, то обработка векторов \mathbf{X} и матриц $||\mathbf{w}_{ij}||$, $||\mathbf{h}_{ij}||$ осуществляется по частям, т.е. по блокам с размерностью n_y . При расчете векторов выходной активности \mathbf{X}' в соответствии с (2), при вычислении суммарной выходной активности нейронной сети и сравнении ее с заданным значением используются упомянутые аппаратно реализуемые операции.

Оценим времена обучения и распознавания в нейрокомпьютере. При оценке в соответствии с [1-3] будем полагать, что размерность векторов \mathbf{X} и количество единичных компонент в них составляет $n = 4096$ и $m \sim 100$. Время обучения нейрокомпьютера t_1 составляет:

$$t_1 \approx L \frac{4n^2}{N} N_0 f^{-1}, \quad (5)$$

где L — число образов в обучающей выборке, f — тактовая частота интегральных схем, используемых в нейрокомпьютере, N_0 — число тактов, требуемых для выбора вектора из ОЗУ и выполнения одной логической операции, N — число одновременно выполняемых операций. В (5) учтено, что при формировании каждого элемента матрицы $||\mathbf{w}_{ij}||$ осуществляются 4 логические операции. (Предполагается, что случайные бинарные величины $h_{ij} = h_i^{(1)} \wedge h_j^{(2)}$ формируются из пары случайных бинарных векторов $\mathbf{h}^{(1)}, \mathbf{h}^{(2)}$, которые были предварительно записаны в ОЗУ). Полагая в соответствии с оценками информационной емкости ассоциативной памяти [5, 6], что число образов в обучающей выборке составляет $L \sim (n/m)^2 \sim 10^3$, и учитывая, что в разработанном нейрокомпьютере $f = 8$ МГц, $N_0 = 12$, $N = 1024$, имеем $t_1 \sim 100$ с.

Время распознавания составляет:

$$t_r \sim N_{it}(mt_s + N_1 N_0 t_0) \quad (6)$$

N_{it} — число итераций, t_s — время поиска в ОЗУ столбца матрицы $\|w_{ij}\|$, соответствующего значению $X_j = 1$ и занесения единичных компонент w_{ij} в счетчик С (см. формулу (2), при $X_j = 0$ указанные операции не производятся), N_1 — число единичных компонент в выходных векторах \mathbf{X}' , N_0 — число циклов уменьшения порога, t_0 — время сравнения с порогом. В разработанном нейрокомпьютере $t_s = 8f^{-1}$, $t_0 = 4f^{-1}$. Оценим t_r сверху, полагая $N_{it} \sim 10$, N_1 , $N_0 < m$: $t_r < 40m^2f^{-1} \sim 50$ мс.

3. ТЕКУЩЕЕ СОСТОЯНИЕ РАЗРАБОТКИ

В настоящее время изготовлены опытные образцы нейрокомпьютера на основе матричных БИС типа 1515ХМ1 и динамической памяти КН565РУ5 в виде блока, состоящего из 3 плат УУ и 16 плат блока ПОИ, подключаемого в ЭВМ типа IBM PC/AT. Разработан специальный язык "Нейроассемблер", обеспечивающий управление нейрокомпьютером и реализацию нейросетевых алгоритмов. Создано несколько демонстрационных программ, иллюстрирующих процедуры обучения и распознавания на нейрокомпьютере. Ведется отладка системного и прикладного программного обеспечения для нейрокомпьютера, реализующего распознавание спектральных и визуальных образов.

Проведено тестирование работы нейрокомпьютера в режиме распознавания, выполненное следующим образом. Рассматривалась задача распознавания геометрических изображений цифр, инвариантного относительно поворотов, сдвигов и масштабирования. В качестве набора признаков были выбраны моменты Цернике, вычисленные с помощью методики, описанной в [7]. Кодирование, формирование векторов \mathbf{X}_F , \mathbf{C}_k , \mathbf{X} и обучение (т.е. расчет матрицы $\|w_{ij}\|$) проводились в ЭВМ. Число нейронов и активность нейронной сети составляли $n = 512$, $m = 168$. Матрица $\|w_{ij}\|$ записывалась в ОЗУ нейрокомпьютера, в котором затем осуществлялось распознавание в соответствии с описанным алгоритмом. Время распознавания определялось с помощью системного таймера PC/AT и при $N_{it} = 8$ составило 2,8 мс.

В настоящее время ведется разработка аппаратуры нейрокомпьютера на базе программируемых логических матричных БИС, что позволяет выполнить нейрокомпьютер в виде плат (слотов) встраиваемых в PC/AT и увеличить его быстродействие в 30–40 раз.

4. ВЫВОДЫ

Произведена разработка, изготовление и испытание нейрокомпьютера с бинарной матрицей памяти. Его новая аппаратная реализация предназначена для практического применения в различных областях распознавания образов. Авторы приглашают к сотрудничеству в практическом

применении нейрокомпьютера для научных и производственных задач.
Адрес: 103498, Москва, К-498, Зеленоград, корпус 445, кв.102. ИнТеСи.

ЛИТЕРАТУРА

1. Нейроподобные сети и нейрокомпьютеры // Сб.науч.тр. АН УССР. — Киев: Ин-т кибернетики им.В.М.Глушкова, 1991. 78 с.
2. Артыкуца С.Я., Байдык Т.Н., Куссуль Э.М., Рачковский Д.А. // Препринт. — Киев: Ин-т кибернетики им.В.М.Глушкова АН УССР, 1991. 20 с.
3. Куссуль Э.М., Байдык Т.Н. // Нейрокомпьютер. 1992. Т.1. N 1. С.16.
4. Искусственный интеллект. — В 3-х кн. Кн.2: Модели и методы: Справочник / Под ред.Д.А.Поспелова. — М.: Радио и связь, 1990. Гл.4. 304 с.
5. Willshaw D.J., Beneman O.P., Longuet-Higgins H.C. // Nature. 1969. V.220. N.5197. P.960.
6. Фролов А.А., Муравьев И.П. Информационные характеристики нейронных сетей. — М.: Наука, 1988. 160 с.
7. Perntoris S.J., Lisboa P.J.G. // IEEE trans. on neural networks. 1992. V.3. N 2. P.241.

Предприятие "Вертер",
г.Зеленоград

Поступила в редакцию
18 ноября 1993 г.
После переработки
10 марта 1994 г.

DESIGN OF A NEUROCOMPUTER WITH A BINARY MEMORY MATRIX

*M.I.Dyabin, N.G.Karpinskij, E.M.Kussul', A.I.Polovyanyuk,
V.G.Red'ko, V.A.Serechenko, O.V.Urgant*

A neurocomputer made of standard home integrated circuits has been developed based on an autoassociative neuro-leke memory with synapse binary weights. A pattern recognition algorithm realized in the neurocomputer has been presented. A neurocomputer functional diagram and operation principle have been discussed. Neurocomputer characteristics have been given including estimation of learning and recognition times.