УДК 004.032.26:519.87

Д.В. Маршаков, Д.В. Фатхи

МОДЕЛЬ АППАРАТНОЙ РЕАЛИЗАЦИИ ИСКУССТВЕННОГО НЕЙРОНА НА ОСНОВЕ ЦВЕТНЫХ ВРЕМЕННЫХ СЕТЕЙ ПЕТРИ

В работе рассматриваются основные принципы построения цветных временных сетей Петри для моделирования и анализа свойств искусственных нейронных сетей. Представлена модель искусственного нейрона на основе цветных временных сетей Петри и дано формальное описание функционирования модели.

Искусственные нейронные сети, аппаратная реализация, моделирование, сети Петри, сигмоидальная функция активации

D.V. Marshakov, D.V. Fatkhi

THE MODEL OF THE HARDWARE IMPLEMENTATION OF ARTIFICIAL NEURON BASED ON COLORED TIMED PETRI NETS

The paper deals with the basic principles of colored timed Petri nets as a tool for modeling and analyzing the properties of artificial neural networks. The model of artificial neuron based on colored timed Petri nets and a formal description of a functioning model are given.

Artificial neural networks, hardware implementation, simulation, Petri nets, sigmoid activation function

Введение

Современные искусственные нейронные сети (ИНС) в силу своих способностей к моделированию нелинейных процессов, обобщению, работе с зашумленными данными и т.д. находят применение во многих прикладных областях. При этом все более возрастает удельных вес аппаратных нейросетевых реализаций. Это связано с требованиями к обеспечению высоких скоростей решения сложных задач с использованием массового параллелизма [1, 2], с которыми пока не справляются традиционные компьютеры.

Одной из перспективных аппаратных реализаций ИНС является построение их на основе ПЛИС – матричных структур, обеспечивающих параллельную обработку информации и обладающих высокой степенью интеграции, масштабируемостью, развитыми средствами проектирования, реализации выполнения алгоритмов обучения [1-4].

Для правильного воплощения ИНС до её реализации необходимо предварительное моделирование структуры и поведения нейронов сети.

Мощным средством для моделирования параллельных и распределенных процессов являются сети Петри [5]. С помощью сетей Петри и их модификаций можно наглядно представить динамику функционирования систем и составляющих их элементов.

Применение цветных сетей Петри в контексте моделирования ИНС рассматривается в работах [6, 7]. В рассматриваемых моделях используются только пороговые функции активации (так называемые пороговые логические элементы), в которых обработка данных сводится к формированию на выходе бинарных дискретных сигналов. ИНС с пороговой функцией активации, они обладают рядом недостатков, среди которых наиболее существенным является насыщение нейронов от больших значений входных сигналов.

Более приемлемой в практике моделью ИНС является модель, состоящая из нейронов с непрерывной (как правило, сигмоидальной) функцией активации [1, 3, 4]. При этом пространство решений, в котором работает такая сеть, не ограничивается дискретными значениями и состоит из полного спектра возможных входов и их значений, благодаря чему сеть способна аппроксимировать практически любую функциональную зависимость. Кроме того, сигмоидальная функция активации обладает свойством усиливать слабые сигналы и предотвращает насыщение нейрона от больших сигналов [8].

Необходимо отметить, что в ИНС с пороговыми элементами обучение производится лишь на её модели, занося полученные весовые коэффициенты в память кристалла на этапе проектирования. Использование в аппаратных реализациях многослойных нейронных сетей нейронов с сигмоидальной функцией активации может позволить производить обучение нейросети на кристалле в реальном масштабе времени.

В [9] отмечается, что при проектировании ИНС особое внимание современными разработчиками уделяется реализации ИНС с нейронами, обладающими сигмоидальной функцией активации.

Для исключения указанных недостатков ИНС с пороговой функцией, а также для обеспечения обучения аппаратных реализаций ИНС в реальном масштабе времени, необходима разработка сетевой модели ИНС с реализацией сигмоидальной функции активации.

1. Математическая модель искусственного нейрона

На рис. 1 представлена модель нейрона, лежащая в основе искусственных нейронных сетей. В состав нейрона входят умножители (синапсы), сумматор и устройство, реализующее функцию активации.

Синапсы выполняют скалярное произведение n-мерного вектора входных данных $x = x_1, x_2, ..., x_n$ и n-мерного вектора весовых данных $w = w_1, w_2, ..., w_n$.

Сумматор выполняет сложение взвешенных входов:

$$\Sigma_j = \sum_{i=1}^{n} w_i x_i \quad (1)$$

где w_i (i=1,2,…,n) – вес синапса,

x_i (i=1,2,…,n) – входной сигнал,

n – число входов нейрона,

j – номер нейрона.

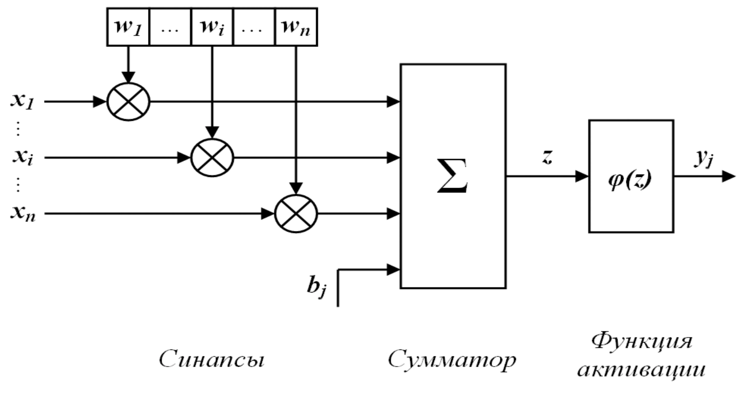

Рис. 1. Структура искусственного нейрона

Нейрон, показанный на рис. 1, дополнен скалярным смещением b_j. Смещение суммируется со взвешенными входами и приводит к сдвигу аргумента функции активации на величину b_j:

$$z = \Sigma_j + b_j, \tag{2}$$

где b_j – значение смещения.

Функция активации $\varphi(z)$ определяет выходной сигнал нейрона. Сигмоидальная функция активации имеет вид (3).

$$y_j = \varphi(z) = \frac{1}{1+e^{-z}} \tag{3}$$

Математическая модель нейрона описывается системой рассмотренных соотношений:

$$\begin{cases} z = \Sigma_j + b_j, \\ \varphi(z) = \dfrac{1}{1+e^{-z}}; \end{cases} \tag{4}$$

2. Аппаратное представление искусственного нейрона

Один из вариантов архитектурного решения искусственного нейрона на основе ПЛИС Xilinx Spartan IIE рассмотрен в работе [4] и представлен на рис. 2.

Структура нейрона состоит из различных вычислительных подблоков. Для определения значения функции активации используется таблица (LUT), хранящаяся в блоке памяти RAM. В структуре содержится один умножитель (MUL) и один сумматор (ADD). Входные сигналы (input_i) вводятся параллельно в ROM1. CONTROL UNIT, управляя стартовыми сигналами (start), конечными сигналами (reset) и счетчиком сигналов нейрона (num) координирует функции блоков, обеспечивает синхронизацию входных сигналов и соответствующих им весов. Первый входной сигнал и соответствующий ему вес сохраняется в ROM2, после чего блок MUL последовательно их перемножает и заносит в ROM3. Та же операция повторяется и для других входных сигналов, после чего сумматор ADD суммирует взвешенные входы (sum) и добавляет к ним значение смещения BIEST. Результат суммирования (sum_out) поступает в таблицу определения значений сигмоидальной функции LUT, после чего полученное значение сохраняется в блоке OUT, подающем сигнал окончания процесса.

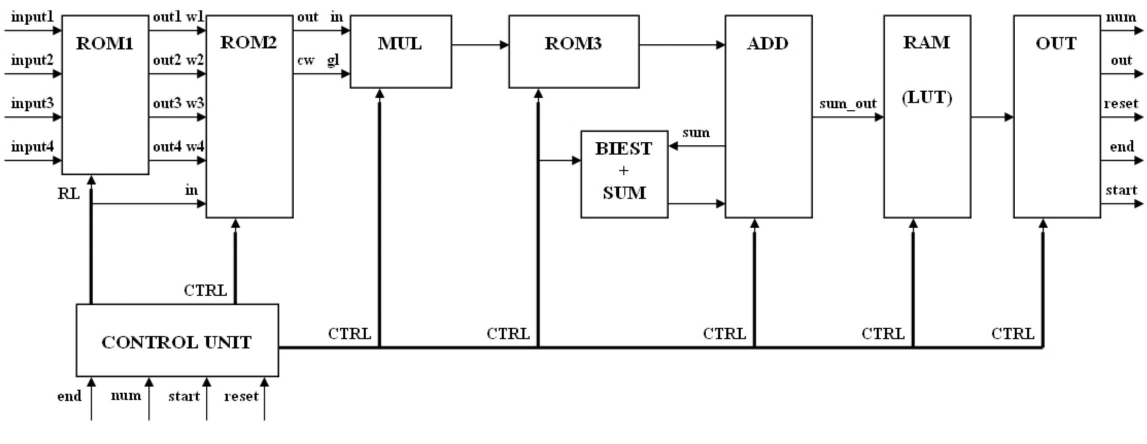

Рис. 2. Структурная схема аппаратного представления нейрона
на основе ПЛИС Xilinx Spartan IIE

3. Модель искусственного нейрона на основе цветной временной сети Петри

В качестве модели искусственного нейрона предлагается модель, модифицирующая модели, рассмотренные в [6, 7], на основе цветных временных сетей Петри в контексте нейронных сетей с сигмоидальной функцией активации.

Формальное определение модели. Модифицированная цветная временная сеть Петри – *CTPNNSIG*, представляющая искусственный нейрон с сигмоидальной функцией активации задается следующим набором:

$$STPNNSIG = (P, T, A, M, C, F, H, \mu, \tau, OC, CT, SIG), \quad (5)$$

где: $P = \{p_1, ..., p_n\}$ – конечное непустое множество позиций;

$T = \{t_1, ..., t_m\}$ – конечное непустое множество переходов;

A – конечное непустое множество дуг такое, что $A \subseteq (P \times T) \cup (T \times P)$;

M – функция маркировки, определенная на P так, что для $p \in P$ $M(p): C(p) \to Z$, где Z – множество целых чисел;

C – конечное множество цветов, состоящее из двух наборов цветов $C(p)$ и $C(t)$, такое, что $C = C(p) \cap C(t)$:

- $C(p_j) = \{\langle цвет_{jp1}, приоритет_{jp1}\rangle, \langle цвет_{jp2}, приоритет_{jp2}\rangle, \dots, \langle цвет_{jpc}, приоритет_{jpc}\rangle\}$ – набор присвоения маркерам каждого места $p \in P$ цветов и соответствующих им приоритетов, где p_c – число цветных маркеров в позиции $p_j \in P$;
- $C(t_j) = \{\langle цвет_{jt1}, приоритет_{jt1}\rangle, \langle цвет_{jt2}, приоритет_{jt2}\rangle, \dots, \langle цвет_{jtc}, приоритет_{jtc}\rangle\}$ – набор присвоения маркерам цветов и соответствующих им приоритетов каждому переходу $t \in T$ цветов, где t_c – количество цветов, связанных с переходом $t_j \in T$.

Цвета и их приоритеты связаны с каждой позицией $C(p) \in C$ и каждым переходом $C(t) \in C$ в сети. Для связи цвета с его приоритетом цвета определяются в порядке убывания приоритетов, а именно: красный (*к*), зеленый (*з*), синий (*с*), черный (*ч*), желтый (*ж*) и белый (*б*). Каждому приоритету этих цветов соответствует числовое обозначение в порядке убывания – 1, 2, 3, 4, 5 и 6, соответственно. Например, запись $\langle к, 1\rangle$ отражает красный цвет с максимальным приоритетом.

F – функция цветной кратности $F(a): A(C) \to Z$, описывающая тип и количество цветов, связанных с каждой дугой $a \in A$, где Z – множество целых чисел.

H – множество ингибиторных дуг. При этом $P_h = \{p_h \mid (p_h, t) \in H\}$ – множество ингибиторных позиций, $p_h \in P_h$ и $p_h \subseteq P_h$. Функция $H(p_h\langle цвет\rangle, t): H \to C(p_h) \times C(t)$ описывает каждую ингибиторную дугу $H(p_h\langle цвет\rangle, t)$ во множестве H.

μ – функция, описывающая цветные маркеры $C(t) \in C$ в позиции $p \in P$, $\mu(p): p \to \langle I, C(p)\rangle$. Атрибут I используется для описания процессов, моделируемых нейроном:

- $\mu(p): p \to \langle x_i, C(p)\rangle$ – передачи входных данных x_i (i=1,2,…,n) для j-го нейрона;
- $\mu(p): p \to \langle w_i, C(p)\rangle$ – передачи весовых коэффициентов w_i (i=1,2,…,n) для j-го нейрона. Маркеры x_i и w_i имеют одинаковый цвет;
- $\mu(p): p \to \langle b_j, C(p)\rangle$ – передачи значения смещения b_j для j-го нейрона;

- $\mu(p)\colon p \to \langle out_j, C(p)\rangle$ – передачи конечного результата с выхода *j*-го нейрона;
- $\mu(p)\colon p \to \langle Вход_i, C(p)\rangle$ – передачи нейрону информации о входных данных x_i, соответствующих им весовых данных w_i, и значения смещения b_j.

τ – функция времени срабатывания, определяющая время выполнения операций в переходах $t \in T$ (для каждого цвета).

OC – функция вычисления, определяющая параметры, необходимые для выполнения операций сложения и умножения переходов $t \in T$ в соответствии с имеющимся цветом $C(t)$:

- $OC_{j1}(t)\colon T \to \langle x_i \times w_i,\ \tau,\ C(t)\rangle$ – функция, определяющая параметры для операции взвешивания входных данных: $x_i \times w_i$;
- $OC_{j2}(t)\colon T \to \langle sum_j,\ \tau,\ C(t)\rangle$ – функция, определяющая параметры для операции сложения взвешенных данных *j*-го нейрона: $sum_j = \sum_{i=1}^{n} x_i \times w_i$;
- $OC_{j3}(t)\colon T \to \langle sum\_out_j,\ \tau,\ C(t)\rangle$ – функция, определяющая параметры для операции суммирования смещения и взвешенных данных *j*-го нейрона: $sum\_out_j = sum_j + b_j$.

CT – функция связи $CT(t)\colon T \to \langle CT_i,\ \tau,\ C(t)\rangle$. Определяет параметры, необходимые для срабатывания «связующих переходов» $t \in T$ в соответствии с имеющимся цветом. Эти переходы моделируют поведение нейронов при приёме-передаче данных между собой.

SIG – функция вычисления $SIG(t)\colon T \to \langle \varphi(sum\_out_j),\ \tau,\ C(t)\rangle$. Определяет необходимые параметры для выполнения переходов $t \in T$ сигмоидальной функции активации в соответствии с появлением цвета $C(t)$. Выходной результат выполнения операции вычисления сигмоидальной функции активации $\langle out_j, C(p)\rangle$ рассчитывается с помощью атрибута φ: $\varphi(sum\_out_j) = 1/(1 + \exp(-sum\_out_j))$.

В *CTPNNSIG* существует функциональная зависимость между цветовым приоритетом включения переходов и цветовым приоритетом маркеров, размещаемых во входных позициях этих переходов./

***CTPNNSIG*-модель искусственного нейрона.** *CTPNNSIG*-модель искусственного нейрона представлена на рис. 3. По аналогии с аппаратной моделью на моделируемый нейрон поступают четыре входных сигнала.

Для разрешения воздействия на моделируемый нейрон со стороны соседних нейронов используются ингибиторные дуги $H(p_4, t_5)$, $H(p_7, t_5)$. При наличии маркера в позициях p_4, p_7 ингибиторная дуга запрещает нейрону получение данных от других нейронов до завершения текущих вычислений моделируемым нейроном. Когда указанные позиции пусты, моделируемый нейрон готов к дальнейшей связи с соседними нейронами.

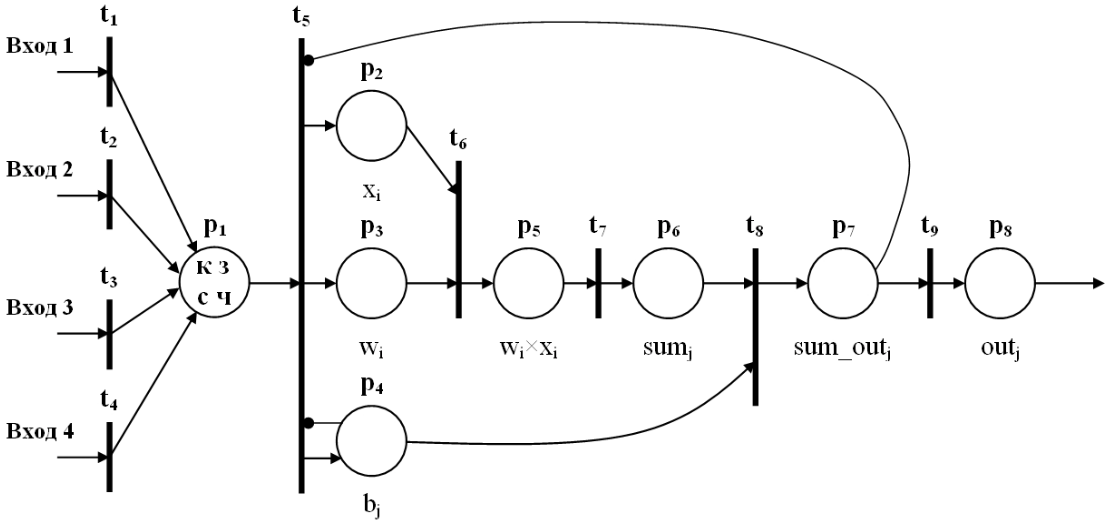

Рис. 3. *CTPNNSIG*-модель искусственного нейрона

Поясним работу этой модели. В позицию p_1 заносятся поступившие в нейрон входные данные $\langle Вход_1, (к,1)\rangle$, $\langle Вход_2, (з,2)\rangle$, $\langle Вход_3, (с,3)\rangle$, $\langle Вход_4, (ч,4)\rangle$, помеченные цветными маркерами. При отсутствии маркеров в позициях p_4 и p_7, срабатывает переход t_5, осуществляя передачу входных сигналов x_i в позицию p_2, соответствующих им весовых данных w_i в позицию p_3 и значения смещения b_j моделируемого нейрона в позицию p_4. При этом маркер в позиции p_4 запрещает переходу t_5 принимать новые данные. Переход t_5 выполняет операцию умножения $w_i \times x_i$ – он последовательно обрабатывает маркеры, согласно их цветовому приоритету, и заносит полученные результаты в позицию p_5. Переход t_7 выполняет операцию суммирования взвешенных входных сигналов, занося результат sum_j в позицию p_6 с присвоенным ему максимальным цветовым приоритетом. Переход t_8 производит суммирование маркеров позиций p_6 и p_4, добавляя к sum_j смещение.

Полученный результат sum_out_j заносится в позицию p_7. При этом ингибиторная дуга $H(p_7,t_5)$, запрещает переходу t_5 получение данных. Переход t_9 производит вычисление сигмоидальной функции активации $\varphi(sum\_out_j)$ и передает результат out_j в позицию p_8. При этом маркер покидает позицию p_7, снимая запрет для перехода t_5, и нейрон готов к дальнейшей работе.

Формальное описание работы модели. Формальное описание алгоритма работы модели и её динамическое поведение можно представить как множество всех состояний модели, которые могут быть достигнуты ею в определенный момент времени при срабатывании конечного числа переходов.

В исходном состоянии модели Сост$_1$ позиция p_1 содержит поступающие в моделируемый нейрон данные, помеченные цветными маркерами: $\langle Вход_1, (к,1)\rangle$, $\langle Вход_2, (з,2)\rangle$, $\langle Вход_3, (с,3)\rangle$, $\langle Вход_4, (ч,4)\rangle$. Каждый цвет присваивает маркерам приоритет для организации последовательности событий на переходах $t \in T$.

Состояния модели описываются атрибутами $МАР_i$, $ПЕР_i$ и $ИНГ_i$.

Атрибут MAP_i показывает распределение маркеров в различных позициях текущего состояния модели в соответствии с функцией M_i.

Атрибут $ПЕР_i$ описывает работу переходов в текущем состоянии и имеет три параметра: $T_{сраб}$ – показывает сработанные переходы, $T_{гот}$ – новые переходы готовые к срабатыванию и $T_{мин}$ – отображает минимальное время срабатывания переходов, описываемых атрибутами $T_{сраб}$ и $T_{гот}$. Целью параметра $T_{мин}$ является оценка общего времени работы переходов до текущего состояния.

Атрибут $ИНГ_i$ показывает распределение маркеров в ингибиторных местах и надписи на ингибиторных дугах, которые соответствуют текущему состоянию модели в соответствии с функцией H_i.

Формально состояния модели имеют следующее описание:

Сост₁:

$$MAP_1: \mu(p_1) = \langle Вход_1, (к,1)\rangle, \langle Вход_2, (з,2)\rangle, \langle Вход_3, (с,3)\rangle, \langle Вход_4, (ч,4)\rangle$$
$$ПЕР_1: T_{мин}: t_5 = \langle (к,1), (з,2), (с,3), (ч,4)\rangle$$

Сост₂:

$$MAP_2: \mu(p_2) = \langle x_1, (к,1)\rangle, \langle x_2, (з,2)\rangle, \langle x_3, (с,3)\rangle, \langle x_4, (ч,4)\rangle,$$
$$\mu(p_3) = \langle w_1, (к,1)\rangle, \langle w_2, (з,2)\rangle, \langle w_3, (с,3)\rangle, \langle w_4, (ч,4)\rangle,$$
$$\mu(p_4) = \langle b_j, (к,1)\rangle$$
$$ПЕР_2: T_{мин}: t_5 = \langle (w_i \times x_i), \tau_1, ((к,1), (з,2), (с,3), (ч,4))\rangle$$
$$ИНГ_2: H(p_4\langle к,1\rangle, t_5)$$

Сост₃:

$$MAP_3: \mu(p_2) = \langle x_2, (з,2)\rangle, \langle x_3, (с,3)\rangle, \langle x_4, (ч,4)\rangle,$$
$$\mu(p_3) = \langle w_2, (з,2)\rangle, \langle w_3, (с,3)\rangle, \langle w_4, (ч,4)\rangle, \ \mu(p_5) = \langle w_1 x_1, (к,1)\rangle,$$
$$\mu(p_4) = \langle b_j, (к,1)\rangle$$
$$ПЕР_3: T_{мин}: t_5 = \langle (w_i \times x_i), \tau_1, ((к,1), (з,2), (с,3), (ч,4))\rangle$$
$$ИНГ_3: H(p_4\langle к,1\rangle, t_5)$$

Сост₄:

$$MAP_4: \mu(p_2) = \langle x_3, (с,3)\rangle, \langle x_4, (ч,4)\rangle,$$
$$\mu(p_3) = \langle w_3, (с,3)\rangle, \langle w_4, (ч,4)\rangle, \ \mu(p_5) = \langle w_1 x_1, (к,1)\rangle, \langle w_2 x_2, (з,2)\rangle,$$
$$\mu(p_4) = \langle b_j, (к,1)\rangle$$
$$ПЕР_4: T_{мин}: t_5 = \langle (w_i \times x_i), \tau_1, ((к,1), (з,2), (с,3), (ч,4))\rangle$$
$$ИНГ_4: H(p_4\langle к,1\rangle, t_5)$$

Сост₅:

$$MAP_5: \mu(p_2) = \langle x_4, (ч,4)\rangle,$$
$$\mu(p_3) = \langle w_4, (ч,4)\rangle, \ \mu(p_5) = \langle w_1 x_1, (к,1)\rangle, \langle w_2 x_2, (з,2)\rangle, \langle w_3 x_3, (с,3)\rangle,$$
$$\mu(p_4) = \langle b_j, (к,1)\rangle$$
$$ПЕР_5: T_{мин}: t_5 = \langle (w_i \times x_i), \tau_1, ((к,1), (з,2), (с,3), (ч,4))\rangle$$
$$ИНГ_5: H(p_4\langle к,1\rangle, t_5)$$

Сост₆: $MAP_6 : \mu(p_5) = \langle w_1 x_1, (\kappa,1)\rangle, \langle w_2 x_2, (з,2)\rangle, \langle w_3 x_3, (c,3)\rangle, \langle w_4 x_4, (ч,4)\rangle$,

$\mu(p_4) = \langle b_j, (\kappa,1)\rangle$

$ПЕР_6 : T_{мин} : t_7 = \langle sum_j, \tau_2, ((\kappa,1),(з,2),(c,3),(ч,4))\rangle$

$ИНГ_6 : H(p_4\langle \kappa,1\rangle, t_5)$

Сост₇: $MAP_7 : \mu(p_6) = \langle sum_j, (\kappa,1)\rangle$,

$\mu(p_4) = \langle b_j, (\kappa,1)\rangle$

$ПЕР_7 : T_{мин} : t_8 = \langle (sum_j + b_j), \tau_3, (\kappa,1)\rangle$

$ИНГ_7 : H(p_4\langle \kappa,1\rangle, t_5)$

Сост₈: $MAP_8 : \mu(p_7) = \langle (sum\_out_j), (\kappa,1)\rangle$

$ПЕР_8 : T_{мин} : t_9 = \langle \varphi(sum\_out_j), \tau_4, (\kappa,1)\rangle$

$ИНГ_8 : H(p_7\langle \kappa,1\rangle, t_5)$

Сост₉: $MAP_9 : \mu(p_6) = \langle out_j, (\kappa,1)\rangle$

Выводы

Достоинство предложенной модели состоит в использовании сигмоидальной функции активации, позволяющей исключить насыщение нейрона от больших сигналов, а также обеспечить возможность производить обучение нейронной сети в реальном масштабе времени при аппаратной реализации.

Модель в сравнении с известными является более адекватной процессам, реализуемым искусственными нейронами. В качестве входных и выходных сигналов могут быть не только бинарные сигналы, но и любые другие. Это позволяет достаточно полно обеспечивать свойства искусственного нейрона, необходимые при разработке конфигурации желаемой ИНС, с последующей аппаратной реализацией на ПЛИС.

Формальное описание функционирования модели представляет собой довольно простой метод для разработки и проверки алгоритмов обучения и диагностики, которые могут применяться на модели. Результаты анализа таких свойств сети, как условия выполнения переходов и время срабатывания переходов, проверяют реальную динамику функционирования модели.

ЛИТЕРАТУРА

1. Himavathi S., Anitha D., Himavathi S. Feedforward Neural Network Implementation in FPGA Using Layer Multiplexing for Effective Resource Utilization. Muthuramalingam A. // IEEE Transactions on Neural Networks. 2007. Vol.18. No.3. P. 880-888.

2. Грибачев В.П. Элементная база аппаратных реализаций нейронных сетей // Компоненты и технологии. 2006. №8. С. 12-15.

3. Lee Y., Ko S.-B. FPGA Implementation of a Face Detector Using Neural Networks // IEEE CCECE/CCGEI. Ottawa. 2006. P. 1914-1917.

4. Sahin S., Becerikli Y., Yazici S. Neural Network Implementation in Hardware Using FPGAs // ICONIP 2006. Springer-Verlag Berlin Heidelberg, 2006. Part III. P. 1105-1112.

5. Питерсон Дж. Теория сетей Петри и моделирование систем. М.: Мир, 1984. 264 с.

6. Koriem S.M. CN-Nets for Modeling and Analyzing Neural Networks // Journal of King Saud University. Computer & Information Sciences. 2001. Vol.13. №1. P. 19-47.

7. Крюкова Д.Ю., Суконщиков А.А. Разработка системы моделирования сложных систем на базе нейронных сетей Петри // Актуальные проблемы управления и экономики: история и современность: мат. науч. конференции. Вологда: филиал СЗАГС, 2006. С. 144-148.

8. Хайкин С. Нейронные сети: полный курс, 2е издание. Пер. с англ. М.: Издательский дом «Вильямс», 2006. 1104 с.

9. Omondi A., Rajapakse J., Bajger M. FPGA Implementations of Neural Networks. Berlin: Springer, Germany, 2006. 360 p.

Маршаков Даниил Витальевич –
аспирант Донского государственного технического университета, ассистент кафедры «Вычислительные системы и информационная безопасность» института энергетики и машиностроения Донского государственного технического университета

Marshakov Daniil Vital'evich –
Postgraduate Student of Don State Technical University, Technical Assistant of the Department "Computing Systems and Information Security", Institute of Energy and Machine Building of Don State Technical University

Фатхи Денис Владимирович –
кандидат технических наук, старший научный сотрудник 1 научно-исследовательской лаборатории Ростовского военного института ракетных войск

Fatkhi Denis Vladimirovich –
Candidate of Technical Science, Senior Research Worker of 1st Research Laboratory, Rostov Military Institute of Rocket Forces.

Статья поступила в редакцию 20.03.2011, принята к опубликованию 20.08.2011