

(19)



**Евразийское
патентное
ведомство**

(11) **048157**

(13) **B1**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

(45) Дата публикации и выдачи патента
2024.10.30

(21) Номер заявки
202490930

(22) Дата подачи заявки
2024.05.08

(51) Int. Cl. **G06N 3/06** (2006.01)
G06F 7/72 (2006.01)
G06F 17/17 (2006.01)
G06F 7/38 (2006.01)

(54) **УСТРОЙСТВО ДЛЯ КУСОЧНО-ЛИНЕЙНОЙ АППРОКСИМАЦИИ СИГМОИДЫ**

(43) **2024.10.24**

(96) **2024000088 (RU) 2024.05.08**

(71)(73) Заявитель и патентовладелец:
**ФЕДЕРАЛЬНОЕ
ГОСУДАРСТВЕННОЕ
АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО
ОБРАЗОВАНИЯ "СЕВЕРО-
КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ
УНИВЕРСИТЕТ" (RU)**

(56) CN-A-104484703
RU-C2-2724784
CN-A-110659015
RU-C1-2731332
RU-C1-2747371

(72) Изобретатель:
**Бабенко Михаил Григорьевич,
Кучуков Виктор Андреевич, Валуева
Мария Васильевна, Валуев Георгий
Вячеславович (RU)**

(74) Представитель:
Алиханов А.А. (RU)

(57) Изобретение относится к вычислительной технике и предназначено для кусочно-линейной аппроксимации функции активации сигмоиды искусственных нейронных сетей (ИНС), работающих в системе остаточных классов. Техническим результатом заявляемого изобретения является расширение функциональных возможностей, а именно возможность работать с числами, представленными в системе остаточных классов, а также повышение быстродействия и сокращение оборудования за счет сокращения узлов аппроксимации и оптимального выбора аргументов аппроксимации. Сущность изобретения основана на аппроксимации сигмоиды многочленами наилучшего равномерного приближения Чебышева первой степени с использованием функции Хевисайда, представленными в системе остаточных классов.

048157
B1

048157
B1

Изобретение относится к вычислительной технике и предназначено для кусочно-линейной аппроксимации функции активации сигмоиды искусственных нейронных сетей (ИНС), работающих в системе остаточных классов.

Одним из подходов к увеличению быстродействия является разработка аппаратных ускорителей. Аппаратные реализации ИНС являются сложными и, как правило, процесс обучения производится до аппаратного проектирования. Одним из подходов к улучшению производительности ИНС является использование альтернативных непозиционных систем счисления, например системы остаточных классов, которая позволяет оптимизировать операции сложения и умножения чисел. Однако реализация нелинейного слоя в системе остаточных классов, к которому относятся функции активации, и, в частности, сигмоида, являются вычислительно сложными. Решению данной задачи посвящено данное изобретение.

Известно устройство для аппроксимации функций (патент RU 2010324, опубл. 30.03.1994). Сущность изобретения состоит в создании простого и помехоустойчивого устройства для аппроксимации функций времяимпульсного аргумента с использованием итерационного метода реализации оператора усреднения путем функционального обобщения процесса преобразования кода в частоту как для цепи обратной связи местного контура, так и для цепи прямой связи главного контура при противофазном использовании четырех опорных частот и исключением благодаря этому двухтактного разделения процессов и коммутирования входного и выходного кодов в цепях прямой и обратной связи главного контура. Устройство для аппроксимации функций содержит два двоичных умножителя, блок сравнения, два реверсивных счетчика, два элемента И, два элемента 2И - ИЛИ, элемент НЕ, соединенные между собой функционально.

Недостатком данного устройства являются ограниченные функциональные возможности, а именно невозможность работы в системе остаточных классов и неоптимальность выбора точек аппроксимации.

Известна нелинейная адаптивная система автоматического управления (патент RU 2267147, опубл. 27.12.2005), которая относится к адаптивным системам автоматического управления объектами с неизвестным математическим описанием и произвольными возмущающими воздействиями при наличии в задании системы управления нелинейных ограничений в форме равенств и неравенств на управляемые переменные, управляющие воздействия и траектории перехода объекта управления в требуемое состояние. В системе блок формирования задания, вычислитель выходных сигналов системы измерительных устройств и генератор управляющих воздействий оснащены нелинейными преобразователями с сигмоидальными статическими характеристиками, обеспечивающими выполнение системой управления ограничений в форме неравенств. Вычислитель системы измерительных устройств образован параллельным соединением пропорционального, интегрирующего и дифференцирующего блоков, а генератор управляющих воздействий и наблюдатель переменных состояния выполнены в виде многомерных самонастраивающихся ПИД-регуляторов, реализующих алгоритмы модифицированных фильтров Калмана.

Недостатком данной системы являются ограниченные функциональные возможности, а именно невозможность работы в системе остаточных классов и неоптимальность выбора точек аппроксимации.

Известен способ управления реакцией нейропроцессора на входные сигналы (патент RU 2724784, опубл. 25.06.2020), в котором с учетом нелинейных свойств поверхности разделения классов объектов в N-мерном континуальном пространстве нейропроцессор (НП) выполняет линейную аппроксимацию функций в подобластях области Ω , а параметры нейропроцессора настраиваются путем обучения на основе экспериментальных или теоретических данных о поведении нейропроцессора в заданных точках N-мерного континуального пространства. При обучении нейропроцессора на обучающий вход поочередно подаются эталонные сигналы d_1, d_2, \dots, d_s , а на сигнальные входы НП подаются соответствующие им стимулы x_1, x_2, \dots, x_N . Эталонные сигналы выбирают так, чтобы в каждой подобласти $\Omega^* \in \Omega$ находилась хотя бы одна обучающая выборка. При повторении эпох выполняется минимизация погрешностей функции в вершинах N-мерного куба, ограничивающего каждую подобласть Ω^* .

Недостатком данного способа являются ограниченные функциональные возможности, а именно невозможность работы в системе остаточных классов.

Известен способ формирования порога решающего устройства на основе нейрорегулятора (патент RU 2731332, опубл. 01.09.2020). В способе в режиме обучения при регистрации изменения информативных параметров на выходе параметризатора, выделяющего из набора входных данных совокупности вторичных параметров, являющихся определяющими для решаемой задачи, а также из уставки в виде требуемых реакций бинарного вида формируют обучающую выборку и осуществляют обучение искусственной нейронной сети (ИНС) в виде нейрорегулятора, в котором формируется порог, на основе которого в решающем устройстве вычисляется реальная выходная реакция Z, а в вычитателе вычисляется сигнал ошибки управления, корректируются весовые коэффициенты нейронной сети с использованием алгоритма обратного распространения ошибки, при обучении ИНС используется якобиан решающего устройства, представляющий собой производную от функции Хевисайда, для расчета якобиана в качестве аппроксимации функции Хевисайда используется сигмоидальная функция; в режиме регулирования обученный нейрорегулятор формирует порог, а параметризатор формирует

вторичные параметры, которые сравниваются с полученным порогом, после чего в бинарном виде формируется реальная выходная реакция Z решающего устройства.

Недостатком данного способа являются ограниченные функциональные возможности, а именно невозможность работы в системе остаточных классов.

Известен цифровой аппроксиматор на основе непрерывных кусочно-линейных функций без предварительного табулирования коэффициентов аппроксимации (патент RU 83636, опубл. 10.06.2009), который содержит блок памяти для хранения значений координат Y и блок памяти для хранения значений координат X , эти блоки присоединены соответственно к первому и второму буферным регистрам, первый буферный регистр подключается к перемножителю с инвертором и сумматору, второй буферный регистр подсоединяется к вычитателю, вычитатель и сумматор подключены к делителю, делитель подключен выходом к перемножителю, к которому своим выходом подключен второй сумматор, к второму сумматору своими выходами подключены блок памяти для хранения значений координаты X и блок памяти для хранения значений аргумента функции, перемножитель подключен выходом к накопительному сумматору.

Недостатком данного способа являются ограниченные функциональные возможности, а именно невозможность работы в системе остаточных классов.

Техническим результатом заявляемого изобретения является расширение функциональных возможностей, а именно, возможность работать с числами, представленными в системе остаточных классов, а также повышение быстродействия и сокращение оборудования за счет сокращения узлов аппроксимации и оптимальном выборе аргументов аппроксимации.

Данный технический результат достигается тем, что в устройство для кусочно-линейной аппроксимации сигмоиды, содержащее 3 блока памяти, умножитель, два сумматора, накапливающий сумматор, при этом в качестве первого блока памяти используют блок хранения аргумента \bar{Z} , содержащего матрицу $k \times n$, где k - количество аргументов аппроксимации, n - количество модулей p_i системы остаточных классов (СОК), в качестве второго блока памяти используют блок хранения коэффициента u , содержащего матрицу $k \times n$, в качестве третьего блока памяти используют блок хранения коэффициента v , содержащего матрицу $k \times n$, в качестве накапливающего сумматора используется многовходовый блок суммирования по модулям СОК, дополнительно введено k блоков аппроксимации при этом i -е выходы блока хранения аргумента \bar{Z} , блока хранения коэффициента u и блока хранения коэффициента v подсоединены соответственно к первому, второму и третьему входу i -го блока аппроксимации, на оставшиеся n входов остатка блоков аппроксимации подается значение X , представленное остатками от деления X на модули СОК p_i , т.е. (x_1, x_2, \dots, x_n) , выходы блоков аппроксимации соединены с соответствующими входами многовходового блока суммирования по модулям СОК, выход которого является выходом устройства, при этом каждый i -й блок аппроксимации содержит два сумматора и умножитель по каждому модулю p_j , в качестве которых взяты сумматоры по модулю и умножитель по модулю, при этом n входов остатка соединены с первыми входами соответствующих n умножителей и первых сумматоров по модулю p_j , на вторые входы первых сумматоров по модулю p_j поступают с первого входа блока аппроксимации значения $|\bar{z}_i|_{p_j}$, выходы первых сумматоров по модулю p_j соединены с входами блока вычисления функции Хевисайда, выход которой соединен с управляющим входом многовходового мультиплексора, на вторые входы умножителей по модулю p_j поступают с второго входа блока аппроксимации значения $|u_i|_{p_j}$, выходы умножителей по модулю p_j соединены с первыми входами соответствующих n вторых сумматоров по модулю p_j , вторые входы которых соединены с третьим входом блока аппроксимации, на который поступают значения $|v_i|_{p_j}$, выходы вторых сумматоров по модулю p_j соединены с первым информационным входом многовходового мультиплексора, на вторые входы которого подаются нули, а выходы являются выходом блока аппроксимации.

Сущность изобретения основана на следующем математическом аппарате. В системе остаточных классов любое число $X < P$ однозначно представляется набором остатков x_i от деления числа X на взаимно простые модули СОК p_i , где $x_i \equiv X \pmod{p_i}$, $P = \prod_{i=1}^n p_i$ рабочий диапазон СОК, $i = \overline{1, n}$. При этом операции сложения и умножения выполняются параллельно независимо по каждому модулю p_i .

Широко используемой функций активации является сигмоида, которая задается уравнением

$$s(x) = \frac{1}{1 + e^{-x}}. \quad (1)$$

Однако в системе остаточных классов операции деления и возведения экспоненты в отрицательную степень является вычислительно сложными и необходима аппроксимация функции (1) кусочно-линейными функциями. Рассмотрим аппроксимацию сигмоиды многочленами $Q_n(x)$ наилучшего равномерного приближения Чебышева первой степени, для которых всех точках аппроксимации погрешность не превышает ε . Таким образом, исходная область определения функции разбивается $k+1$ точкой z_i между которыми строятся прямые $A_i \cdot x + B_i \cdot y + C_i = 0$, коэффициенты которых равны

$$A_i = s(z_{i+1}) - s(z_i),$$

$$B_i = z_{i+1} - z_i,$$

$$C_i = z_i \cdot (s(z_i) - s(z_{i+1})) + s(z_i) \cdot (z_{i+1} - z_i).$$

Функция Хевисайда $H(x)$ имеет вид

$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

и по сути является инвертированной функцией знака, реализация которой в системе остаточных классов возможна, например, с использованием устройств по патентам RU 2747371, EA 038389 и др.

Тогда аппроксимация функции сигмоиды $s(x)$ возможна по формуле

$$y = \sum_{i=0}^{k-1} H(x - z_i) \cdot (u_i \cdot x + v_i), \quad (2)$$

где коэффициенты вычисляются следующим образом

$$u_0 = -\frac{A_0}{B_0},$$

$$v_0 = -\frac{C_0}{B_0},$$

$$u_i = \frac{A_{i-1}}{B_{i-1}} - \frac{A_i}{B_i},$$

$$v_i = \frac{C_{i-1}}{B_{i-1}} - \frac{C_i}{B_i}.$$

Реализуем функцию (2) в СОК, но поскольку дробные значения в СОК реализуются сложно, умножим все значения на коэффициент масштабирования. Тогда в СОК с модулями $\{p_1, p_2, \dots, p_n\}$ функция (2) примет вид

$$|g_i|_{p_j} = |x_j - |z_i|_{p_j}|_{p_j},$$

$$|t_i|_{p_j} = H(g_i) \cdot (|u_i|_{p_j} x_j + |v_i|_{p_j}), \quad (3)$$

$$y_j = \left| \sum_{i=0}^{k-1} t_i \right|_{p_j},$$

где $i = \overline{0, k-1}, j = \overline{1, n}$.

Устройство поясняется фигурами 1 и 2. На фиг. 1 представлена структурная схема устройства, которая содержит блок хранения аргумента \bar{Z} 1, блок хранения коэффициента u 2, блок хранения коэффициента v 3, k блоков аппроксимации 4.1-4.к, многоходовый блок суммирования по модулям СОК 5. На фиг. 2 представлена структурная схема блока аппроксимации, которая содержит n первых сумматоров по модулю p_i 6.1-6.н, n вторых сумматоров по модулю p_j 9.1-9.н, n умножителей по модулю p_j 7.1-7.н, блок вычисления функции Хевисайда 8, многоходовый мультиплексор 10.

Поясним связи между блоками и работу устройства на примере.

Рассмотрим пример вычисления аппроксимации сигмоиды с областью определения $[-10; 10]$ и допустимой погрешностью $\varepsilon = 10^{-2}$ для $x = -4$. Значение сигмоиды в этой точке равно 0.0179. Для получения многочленов наилучшего равномерного приближения Чебышева первой степени значения аргументов не являются равноудаленными, таким образом, для данной погрешности достаточно $k+1=10$ точек z_i аппроксимации $[-10; -3.96; -2.46; -1.47; -0.53; 1.03; 1.95; 3.11; 5.42; 10]$.

Поскольку сигмоида имеет высокую кривизну около 0, основные точки аппроксимации сгруппированы там. Для реализации в СОК масштабируем данные значения на 10^2 . Получим $[-10000; -3964; -2464; -1472; -525; 1026; 1946; 3110; 5418; 10000]$.

Возьмем систему остаточных классов $\{127, 255, 511, 512\}$. Для введения знака в СОК диапазон P может быть разделен, чтобы число X удовлетворяло выражению

$$-\frac{P-1}{2} \leq X \leq \frac{P-1}{2}, \text{ если } P \text{ нечетное,}$$

$$-\frac{P}{2} \leq X \leq \frac{P}{2} - 1, \text{ если } P \text{ четное.}$$

Также, чтобы избежать вычитания в первом выражении (3) будем брать $|\bar{z}_i|_{p_j} = |p_j - z_i|_{p_j}$.

Тогда в блок хранения аргумента \bar{Z} 1 будет храниться $k=9$ чисел, представленных в СОК $\{127, 255, 511, 512\}$, т.е. будет содержаться матрица

$$\bar{z} = \begin{pmatrix} 94 & 55 & 291 & 272 \\ 27 & 139 & 387 & 380 \\ 51 & 169 & 420 & 416 \\ 75 & 197 & 450 & 448 \\ 17 & 15 & 14 & 13 \\ 117 & 249 & 507 & 510 \\ 86 & 94 & 98 & 102 \\ 65 & 205 & 467 & 474 \\ 43 & 192 & 203 & 214 \end{pmatrix}.$$

В блоке хранения коэффициента u 2 будут находиться представленные в СОК коэффициенты u прямых $ix+v$, т.е.

$$u = \begin{pmatrix} 3 & 3 & 3 & 3 \\ 36 & 36 & 36 & 36 \\ 69 & 69 & 69 & 69 \\ 86 & 86 & 86 & 86 \\ 39 & 39 & 39 & 39 \\ 43 & 171 & 427 & 428 \\ 47 & 175 & 431 & 432 \\ 73 & 201 & 457 & 458 \\ 112 & 240 & 496 & 497 \end{pmatrix}$$

В блоке хранения коэффициента v 3 будут находиться представленные в СОК коэффициенты v прямых $ix+v$, т.е.

$$v = \begin{pmatrix} 85 & 219 & 159 & 99 \\ 52 & 116 & 86 & 313 \\ 10 & 231 & 408 & 75 \\ 85 & 95 & 357 & 109 \\ 20 & 193 & 408 & 368 \\ 29 & 72 & 414 & 246 \\ 98 & 121 & 326 & 21 \\ 107 & 210 & 136 & 319 \\ 88 & 10 & 355 & 190 \end{pmatrix}$$

Значение $X=-4$, масштабированное и представленное в СОК, будет иметь вид (64, 80, 88, 96) данные значения подаются на 4 входа остатка каждого из 9 блоков аппроксимации.

На первый вход i -го блока аппроксимации $4.i$ поступает i -я строка матрицы g с блока хранения аргумента \bar{z} 1, на второй вход i -го блока аппроксимации поступает i -я строка матрицы u с блока хранения коэффициента u 2, на третий вход i -го блока аппроксимации поступает i -я строка матрицы v с блока хранения коэффициента v 3.

Таким образом, на входы первого блока аппроксимации 4.1 поступают $x=(64, 80, 88, 96)$, $z_1=(94, 55, 291, 272)$, $u_1=(3, 3, 3, 3)$ и $v_1=(85, 219, 159, 99)$. Рассмотрим работу первого блока аппроксимации 4.1 , в остальных блоках вычисления происходят аналогичным образом. В первом сумматоре по модулю p_1 6.1 происходит вычисление $|64+94|_{127}=31$, в первом сумматоре по модулю p_2 6.2 происходит вычисление $|80+55|_{255}=135$ в первом сумматоре по модулю p_3 6.3 происходит вычисление $|88+291|_{511}=379$ в первом сумматоре по модулю p_4 6.4 происходит вычисление $|96+272|_{512}=368$. Таким образом, на вход блока вычисления функции Хевисайда 8 поступают значения $g_1=(31, 135, 379, 368)$, что соответствует числу 6000, которое меньше $\frac{p}{2} - 1$, значит число положительное и на выходе будет значение 1.

В умножителе по модулю p_1 7.1 происходит вычисление $|64 \cdot 3|_{127}=65$, в умножителе по модулю p_2 7.2 происходит вычисление $|80 \cdot 3|_{255}=240$, в умножителе по модулю p_3 7.3 происходит вычисление $|88 \cdot 3|_{511}=264$, в умножителе по модулю p_4 7.4 происходит вычисление $|96 \cdot 3|_{512}=288$. Во втором сумматоре по модулю p_1 9.1 происходит вычисление $|65+85|_{127}=23$, во втором сумматоре по модулю p_2 9.2 происходит вычисление $|240+219|_{255}=204$ во втором сумматоре по модулю p_3 9.3 происходит вычисление $|264+159|_{511}=423$ во втором сумматоре по модулю p_4 9.4 происходит вычисление $|288+99|_{512}=387$ И поскольку значение функции Хевисайда, поступающее на управляющий вход многовходового мультиплексора 10 равно 1, то на выход поступает значение с первого входа, а именно (23, 204, 423, 387) с выходов вторых сумматоров по модулю p_j 9.j.

В остальных блоках аппроксимации $4.i$ вычисления происходят аналогично. Таким образом, на входы соответствующих блоков вычисления функции Хевисайда 8 i -го блока аппроксимации $4.i$ поступает i -я строка матрицы

$$g = \begin{pmatrix} 31 & 135 & 379 & 368 \\ 91 & 219 & 475 & 476 \\ 115 & 249 & 508 & 0 \\ 12 & 22 & 27 & 32 \\ 81 & 95 & 102 & 109 \\ 54 & 74 & 84 & 94 \\ 23 & 174 & 186 & 198 \\ 2 & 30 & 44 & 58 \\ 107 & 17 & 291 & 310 \end{pmatrix}.$$

Соответствующие значения функции Хевисайда равны (1, 0, 0, 0, 0, 0, 0, 0, 0). Таким образом, только с выхода первого блока аппроксимации 4.1 будет подаваться ненулевое значение (23, 204, 423, 387), на выходах остальных блоков аппроксимации будет значение (0, 0, 0, 0).

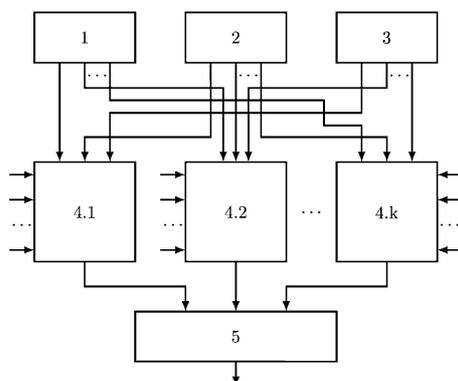
Значения с выходов блоков аппроксимации поступают на входы многовходового блока суммирования по модулям СОК 5, в котором параллельно по каждому модулю происходит вычисление $|23+0+0+0+0+0+0+0+0|_{127}=23$ и т.д. На выходе многовходового блока суммирования по модулям СОК 5 будет значение (23, 204, 423, 387), которое является аппроксимированным значением сигмоиды, представленным в СОК. При переводе данного значения в дробное, получим 0.018819 что соответствует с заданной погрешностью исходному значению.

Поскольку все вычисления выполняются над целочисленными значениями малой размерности, увеличивается скорость вычисления, а неравномерное задание узлов аппроксимации позволяет сократить используемое оборудование.

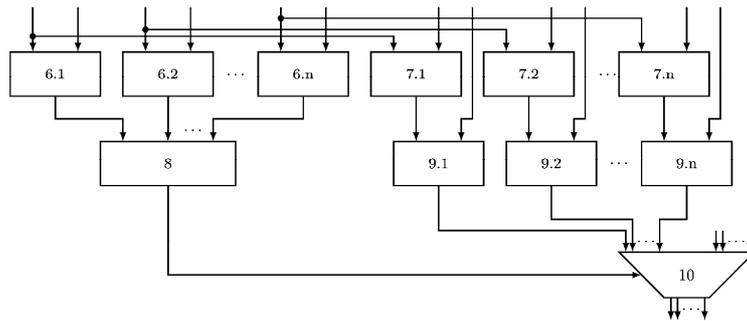
Реализация всего устройства возможна с использованием программируемых логических интегральных схем (ПЛИС) и может использоваться как отдельное устройство, так и вычислительный блок аппаратной реализации нейронной сети.

ФОРМУЛА ИЗОБРЕТЕНИЯ

Устройство для кусочно-линейной аппроксимации сигмоиды, содержащее три блока памяти, умножитель, два сумматора, накапливающий сумматор, отличающееся тем, что в качестве первого блока памяти используют блок хранения аргумента \bar{z} , содержащего матрицу $k \times n$, где k - количество аргументов аппроксимации, n - количество модулей p_i системы остаточных классов (СОК), в качестве второго блока памяти используют блок хранения коэффициента u , содержащего матрицу $k \times n$, в качестве третьего блока памяти используют блок хранения коэффициента v , содержащего матрицу $k \times n$, в качестве накапливающего сумматора используется многовходовый блок суммирования по модулям СОК, дополнительно введено k блоков аппроксимации, при этом i -е выходы блока хранения аргумента \bar{z} , блока хранения коэффициента u и блока хранения коэффициента v соединены соответственно с первым, вторым и третьим входами i -го блока аппроксимации, на оставшиеся n входов остатка блоков аппроксимации подается значение X , представленное остатками от деления X на модули СОК p_i , т.е. (x_1, x_2, \dots, x_n) , выходы блоков аппроксимации соединены с соответствующими входами многовходового блока суммирования по модулям СОК, выход которого является выходом устройства, при этом каждый i -й блок аппроксимации содержит два сумматора и умножитель по каждому модулю p_j , в качестве которых взяты сумматоры по модулю и умножитель по модулю, при этом n входов остатка соединены с первыми входами соответствующих n умножителей и первых сумматоров по модулю p_j , на вторые входы первых сумматоров по модулю p_j поступают с первого входа блока аппроксимации значения $|\bar{z}_{ij}|_{p_j}$, выходы первых сумматоров по модулю p_j соединены с входами блока вычисления функции Хевисайда, выход которой соединен с управляющим входом многовходового мультиплексора, на вторые входы умножителей по модулю p_j поступают с второго входа блока аппроксимации значения $|u_{ij}|_{p_j}$, выходы умножителей по модулю p_j соединены с первыми входами соответствующих n вторых сумматоров по модулю p_j , вторые входы которых соединены с третьим входом блока аппроксимации, на который поступают значения $|v_{ij}|_{p_j}$, выходы вторых сумматоров по модулю p_j соединены с первым информационным входом многовходового мультиплексора, на вторые входы которого подаются нули, а выходы являются выходом блока аппроксимации.



Фиг. 1. Структурная схема устройства



Фиг. 2. Структурная схема блока аппроксимации

