ОБЪЕДИНЕННЫЙ
ИНСТИТУТ
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ

Дубна

P11-97-65

М.А.Назаренко, А.И.Чурин

АЛГОРИТМ ПОЛУЧЕНИЯ МИНИМАЛЬНОЙ И КРАТЧАЙШЕЙ ДИЗЪЮНКТИВНОЙ НОРМАЛЬНОЙ ФОРМЫ БУЛЕВА НЕЙРОНА

Направлено на I Открытую научную конференцию УНЦ ОИЯИ,
24 — 26 февраля 1997 г., Дубна

1997

1 Введение

В современных экспериментах ядерной физики существует целый ряд задач, решенных с использованием метода формальных нейронных сетей [1, 2, 3, 4]. Рассмотрим задачу реализации булевых нейронных сетей с жесткими, ненастраиваемыми связями. То есть требуется реализовать уже обученную нейронную сеть, предназначенную для использования, исключающего дополнительное или переобучение.

Элементом нейронной сети является формальный нейрон. Булев нейрон может быть реализован с использованием различных физических принципов преобразования информации [5]: аналоговая, оптическая, оптоэлектронная обработка, на основе ПЗС, магнитных материалов и др. В мировой практике является общепринятым [4, 5, 6, 7, 8] реализовывать булев нейрон на основе операций формального нейробазиса, которые будут перечислены ниже.

Мы предлагаем реализовывать булев нейрон как булеву функцию, используя все преимущества применения функциональных логических элементов. При этом достигается максимальная скорость принятия решения при абсолютно точном получении выходного сигнала, что является актуальным в современных задачах экспериментальной физики. Для решения поставленной задачи необходимо представить логическую функцию булева нейрона в соответствующей форме. Предлагается алгоритм получения минимальной и кратчайшей дизъюнктивной нормальной формы указанной функции.

2 Минимизация функции булева нейрона

Формальным булевым нейроном называется следующий исполнитель (см. [9]). Обозначим множество входных данных, состоящее из нулей и единиц, вектором $\vec{x}$. Количество компонент этого

вектора определяет число входов нейрона. Каждому входу соответствует величина w_i, называемая весом, где i обозначает номер входа нейрона. Сигнал внутреннего возбуждения N нейрона определяется по следующей формуле:

$$N = \sum_i x_i\, w_i.$$

Выходной сигнал y образуется применением логистической функции F к сигналу внутреннего возбуждения

$$y = F(N).$$

В случае булева нейрона логистическая функция имеет следующий вид

$$F(N) = \theta(N - C),$$

где C — величина порога возбуждения нейрона,

$$\theta(z) = \begin{cases} 0, & \text{при} \quad z \le 0, \\ 1, & \text{при} \quad z > 0. \end{cases} \tag{1}$$

Нейронная сеть формируется из некоторого числа формальных нейронов, выходы одних нейронов подаются на входы других, входы некоторых нейронов объявляются входами всей сети, выходным сигналом сети считается выходной сигнал какого-то множества нейронов сети. Более подробное и строгое определение формальной нейронной сети дано в работе [10].

Тем самым формальный булев нейрон имеет на входе булев вектор данных, на выходе дает также булеву величину. При этом по определению внутренние преобразования производятся, вообще говоря, с произвольными числами N и C. На основе этого формального определения нейрона строятся и методы его аппаратной реализации.

Булев нейрон реализует класс логических функций многих переменных, называемых линейно разделимыми функциями [9]. Минимизация представления этой булевой функции будет произведена в смысле получения минимальной и кратчайшей дизъюнктивной нормальной формы (ДНФ). Ниже следует, что эти представления совпадают [11, стр. 173].

При реализациях в виде схем любая ДНФ, по сравнению с другими формами представления, дает наименьшее число ступеней, равное 2, что является важным для обеспечения надежности и быстродействия схем [11, стр.99].

2.1 Линейно разделимые и монотонно возрастающие булевы функции

Введем следующие обозначения. Символом & будем обозначать функцию "логическое И" (конъюнкция), а символами + или $\vee$ — "логическое ИЛИ" (дизъюнкция). Знак конъюнкции будем опускать, если из контекста очевидно присутствие этой операции. Функцию "ИНВЕРСИЯ" будем обозначать чертой сверху. Символом x^σ будем обозначать результат следующей операции

$$x^\sigma := x\sigma \vee \bar{x}\bar{\sigma},$$

где σ – параметр, равный 0 либо 1, а x — логическая переменная. Очевидно выполнение следующего соотношения:

$$x^\sigma = \begin{cases} \bar{x}, & \text{при} \quad \sigma = 0, \\ x, & \text{при} \quad \sigma = 1. \end{cases} \tag{2}$$

Введем следующие определения, необходимые для дальнейших рассуждений.

Определение 1. *Логическим вектором $\vec{x}$ размерности n называется упорядоченный набор $(x_1, x_2, \ldots, x_n)$ булевых переменных x_j, $j = 1, 2, \ldots, n$. Будем говорить, что вектор $\vec{x}$ превосходит вектор $\vec{y}$, если для компонент этих векторов выполнены следующие неравенства*

$$x_j \geq y_j, \quad j = 1, 2, \ldots, n.$$

Определение 2. *Линейно разделимой называют булеву функцию многих переменных, представимую в следующем виде*

$$f(x_1, \ldots, x_n) = \theta\left(\sum_{i=1}^{n} w_i\, x_i - C\right),$$

где функция θ определяется соотношением (1).

Определение 3. *Функция $f(x_1, \ldots, x_n)$ называется монотонно возрастающей, если для любых таких наборов $\vec{\alpha}$ и $\vec{\beta}$, что $\vec{\alpha} \leq \vec{\beta}$, выполнено неравенство*

$$f(\vec{\alpha}) \leq f(\vec{\beta}).$$

Рассмотрим взаимные связи между классами линейно разделимых и монотонно возрастающих функций.

Теорема 1. *Пусть $f(x_1, \ldots, x_n)$ — произвольная линейно разделимая функция с внутренними параметрами $\{w_i\}_{i=1}^n$ и C. Тогда найдется такой набор $\vec{\sigma} = \{\sigma_1, \ldots, \sigma_n\}$, что функция*

$$f^{\vec{\sigma}}(x_1, \ldots, x_n) \equiv f(x_1^{\sigma_1}, \ldots, x_n^{\sigma_n}) \tag{3}$$

будет монотонно возрастающей.

Доказательство. Положим

$$\sigma_i = \begin{cases} 1, & \text{при} \quad w_i \geq 0, \\ 0, & \text{при} \quad w_i < 0. \end{cases} \tag{4}$$

Пусть для логического вектора $\vec{\alpha}$ имеет место равенство $f^{\vec{\sigma}}(\vec{\alpha}) = 1$. Покажем, что для любого $\vec{\beta}$, такого что $\vec{\beta} \geq \vec{\alpha}$ всегда выполнено соотношение $f^{\vec{\sigma}}(\vec{\beta}) = 1$. Тем самым требуется доказать следующее неравенство

$$\left(\sum_{i=1}^{n} w_i\, \alpha_i^{\sigma_i} - C\right) \leq \left(\sum_{i=1}^{n} w_i\, \beta_i^{\sigma_i} - C\right). \tag{5}$$

В силу нашего выбора $\vec{\beta} \geq \vec{\alpha}$, что, по определению, дает выполнение системы неравенств

$$\beta_i - \alpha_i \geq 0, \quad \bar{\beta}_i - \bar{\alpha}_i \leq 0, \qquad \forall i.$$

Тогда, учитывая соотношение (4), получаем следующие утверждения:

$$\begin{aligned}&\text{если } w_i \geq 0 \;\Rightarrow\; \sigma_i = 1 \;\Rightarrow\; w_i(\beta_i^{\sigma_i} - \alpha_i^{\sigma_i}) = w_i(\beta_i - \alpha_i) \geq 0,\\ &\text{если } w_i < 0 \;\Rightarrow\; \sigma_i = 0 \;\Rightarrow\; w_i(\beta_i^{\sigma_i} - \alpha_i^{\sigma_i}) = w_i(\bar{\beta}_i - \bar{\alpha}_i) \geq 0.\end{aligned}$$

Следовательно, сумма

$$\sum_{i=1}^{n} w_i \left(\beta_i^{\sigma_i} - \alpha_i^{\sigma_i}\right)$$

состоит только из неотрицательных слагаемых. Таким образом выполнено соотношение (5), и имеет место неравенство

$$\theta\left(\sum_{i=1}^{n} w_i\,\alpha_i^{\sigma_i} - C\right) \leq \theta\left(\sum_{i=1}^{n} w_i\,\beta_i^{\sigma_i} - C\right),$$

то есть $f^{\vec{\sigma}}(\vec{\beta}) = 1$.

Возьмем произвольные векторы $\vec{\alpha}$ и $\vec{\beta}$ такие, что $\vec{\alpha} \leq \vec{\beta}$. Если $f^{\vec{\sigma}}(\vec{\alpha}) = 0$ и $f^{\vec{\sigma}}(\vec{\beta}) = 0$, тогда справедливо неравенство

$$f^{\vec{\sigma}}(\vec{\alpha}) \leq f^{\vec{\sigma}}(\vec{\beta}). \tag{6}$$

Если $f^{\vec{\sigma}}(\vec{\alpha}) = 0$ и $f^{\vec{\sigma}}(\vec{\beta}) = 1$, то имеет место неравенство (6). При $f^{\vec{\sigma}}(\vec{\alpha}) = 1$, как было доказано выше, $f^{\vec{\sigma}}(\vec{\beta}) = 1$, поэтому неравенство (6) также справедливо. Следовательно, функция (3) для заданных соотношением (3) параметров $\{\sigma_1, \ldots, \sigma_n\}$ является монотонно возрастающей булевой функцией. Теорема доказана.

Следствие 1.1. *Любая линейно-разделимая функция может быть преобразована в монотонно возрастающую путем инвертирования соответствующих входных переменных, а именно тех, коэффициенты w_i при которых меньше 0. Или по-другому: ДНФ или КНФ функции булева нейрона будут отличаться от ДНФ или соответственно КНФ какой-то монотонно возрастающей функции только инверсией соответствующих переменных.*

Следствие 1.2. *Так как минимальная и кратчайшая ДНФ для монотонно возрастающей функции совпадают, то эти формы будут совпадать и для функции булева нейрона.*

Если монотонно возрастающая функция может быть получена из любой линейно разделимой функции, то является ли полученная монотонная функция линейно разделимой? Если да, чему будут равны параметры этой функции. На этот вопрос дает ответ. следующая теорема.

Теорема 2. *Если линейно-разделимая функция*

$$f(\vec{x}) = \theta\left(\sum_{i=1}^{n} w_i\,x_i - C\right),$$

имеет монотонно возрастающий образ $f^{\vec{\sigma}}(x_1, \ldots, x_n) = f(x_1^{\sigma_1}, \ldots, x_n^{\sigma_n})$. где

$$\begin{aligned}\sigma_i = 1, \quad &\textit{при} \quad w_i \geq 0,\\ \sigma_i = 0, \quad &\textit{при} \quad w_i < 0.\end{aligned}$$

то $f^{\vec{\sigma}}(x_1, \ldots, x_n)$ также является линейно разделимой функцией:

$$f^{\vec{\sigma}}(\vec{x}) = \theta\left(\sum_{i=1}^{n} w_i^{\star} x_i - C^{\star}\right),$$

где

$$w_i^{\star} = |w_i|, \quad C^{\star} = \sum_{\forall i: w_i < 0} |w_i| + C.$$

Доказательство. Докажем, что выражение

$$f(x_1^{\sigma_1}, \ldots, x_n^{\sigma_n}) = \theta\left(\sum_{i=1}^{n} w_i^{\star} x_i - C^{\star}\right) \tag{7}$$

является тождеством.

$$\theta\left(\sum_{i=1}^{n} w_i\,x_i^{\sigma_i} - C\right) = \theta\left(\sum_{i=1}^{n} |w_i|\,x_i - (C + \sum_{\forall i: w_i < 0} |w_i|)\right)$$

Тогда

$$\sum_{i=1}^{n} w_i\,x_i^{\sigma_i} = \sum_{i=1}^{n} |w_i|\,x_i - \sum_{\forall i: w_i < 0} |w_i|$$

$$\sum_{\forall i: w_i<0} w_i\bar{x}_i + \sum_{\forall i: w_i\geq 0} w_i x_i = \sum_{\forall i: w_i\geq 0} w_i x_i - \sum_{\forall i: w_i<0} w_i x_i - \sum_{\forall i: w_i<0} |w_i|$$

Так как $x_i + \bar{x}_i \equiv 1$, то

$$\sum_{\forall i: w_i<0} w_i = -\sum_{\forall i: w_i<0} |w_i|$$

Следовательно равенство 7 является тождеством. Теорема доказана.

Если из линейно разделимой функции можно получить монотонно возрастающую, то можно записать ее просто через базовые функции функциональных элементов. Для этого необходим алгоритм нахождения совершенной дизъюнктивной нормальной формы.

2.2 Алгоритм нахождения простых импликант по параметрам монотонно возрастающей линейно разделимой функции.

Определение 4. *Пусть $\sigma_{i_1}, \sigma_{i_2} ... \sigma_{i_r}$ - фиксированный набор чисел из 0 и 1 такой, что $1 \leq i_1 < i_2 < ... < i_r \leq n$. Множество всех вершин $(\alpha_1, \alpha_2, ..., \alpha_n)$ куба E^n таких, что*

$$\alpha_{i_1} = \sigma_{i_1}, \alpha_{i_2} = \sigma_{i_2}, ... \alpha_{i_r} = \sigma_{i_r},$$

называется $(n-r)-$мерной гранью куба N_K.

Исходя из приведенного определения конъюнкция ранга r, соответствующая данной грани, будет

$$K(x_1, ..., x_n) = x_{i_1}^{\sigma_1} \& ... \& x_{i_r}^{\sigma_r},$$

то есть $K(x_1, ..., x_n)$ принимает значение 1 только на множестве N_K.

Определение 5. *Грань N_K называется максимальной (относительно N_f), если не существует грани $N_{K'}$ такой, что*

1. $N_K \subseteq N_{K'} \subseteq N_f$;

2. Размерность грани $N_{K'}$ больше размерности грани N_K.

Определение 6. *Конъюнкция K, соответствующая максимальной грани N_K множества N_f, называется простой импликантой.*

Определение 7. *Д.Н.Ф. являющаяся дизъюнкцией всех простых импликант функции f, называется сокращенной ДНФ.*

Любая монотонная функция [11] может быть представлена в сокращенной дизъюнктивной нормальной форме, не содержащей отрицаний переменных и является ее единственной минимальной и кратчайшей ДНФ. То есть простые импликанты выражаются через конъюнкции некоторых переменных без инверсии.

Определение 8. *Будем называть набор $\{\alpha_i\}_{i=1}^n$ граничным вектором для монотонно возрастающей функции $f(x_1, ..., x_n)$, если $f(\vec{\alpha}) = 1$ и для $\forall$ набора $\vec{\beta}$ такого, что $\vec{\alpha} \geq \vec{\beta}$: $f(\vec{\beta}) = 0$.*

Лемма 1. *Для монотонно возрастающей функции граничные векторы однозначно задают максимальные грани.*

Доказательство

Действительно. Если набор $\vec{\alpha}-$ граничный, то для любого $\vec{\beta}$ такого, что $\vec{\beta} \geq \vec{\alpha}$: $f(\vec{\beta}) = 1$. Множество N_K : $\{\vec{\beta}_i\}_{i=1}^{2^r}$ таких, что $\vec{\beta} \geq \vec{\alpha}$ является $(N-r)-$мерной гранью. И этот набор является максимальным, так как не существует $N_{K'} \supset \vec{\gamma}$ такой, что $\vec{\gamma} \not\subset N_K$ и $f(\vec{\gamma}) = 1$. Лемма доказана.

Тогда простые импликанты монотонно возрастающей функции будут взаимно однозначно соответствовать граничным векторам, то есть $K = x_{i_1} \& x_{i_2} \& .. \& x_{i_r}$, где при $j = i_1 .. i_r$: $\alpha_j = 1$. Другими словами, любая монотонно возрастающая функция может быть однозначно задана граничными векторами.

Таким образом задача о нахождении всех конъюнкций совершенной ДНФ монотонно возрастающей линейно разделимой функции сводится к нахождению всех граничных векторов.

Пусть задана монотонно возрастающая и линейно разделимая функция, причем

$$f(\vec{x}) = \theta\left(\sum_{i=1}^{n} w_i\, x_i - C\right),$$

где из Теоремы 1 для всех $i : w_i \geq 0$. Необходимо найти все граничные векторы $\{\vec{\alpha}_j\}$.

Для этого предлагается следующий алгоритм:

Упорядочить индексы переменных в порядке возрастания весов w_i, то есть

$$w_{i_1} \geq w_{i_2} \geq \ldots \geq w_{i_n} (i_\nu \neq i_\mu \text{ } при \text{ } \nu \neq \mu)$$

Будем записывать граничные векторы в виде нулей и единиц в установленной последовательности. То есть первая позиция будет соответствовать переменной x_{i_1}, вторая x_{i_2} и так далее. Будем проверять в определенной последовательности наборы: являются ли они граничными сравнивая соответствующую им сумму

$$\sum_{j=1}^{n} w_j x_j$$

с C.

Последовательно перебираем векторы, фиксируя одну часть переменных, взвешенная сумма которых меньше C, и последовательно перебирая другую часть переменных. Когда общая взвешенная сумма превышает C – записывается граничный вектор.

Пример 1. Пусть для монотонно возрастающей линейно разделимой функции при $n = 6$ получены коэфициенты $w_1 = 6$, $w_2 = 4$, $w_3 = 3$, $w_4 = 3$, $w_5 = 2$, $w_6 = 1$, а $C = 7.5$. Тогда таблица граничных векторов будет выглядеть следующим образом:

$w_1 = 6$	$w_2 = 4$	$w_3 = 3$	$w_4 = 3$	$w_5 = 2$	$w_6 = 1$
1	1	0	0	0	0
1	0	1	0	0	0
1	0	0	1	0	0
1	0	0	0	1	0
0	1	1	1	0	0
0	1	1	0	1	0
0	1	1	0	0	1
0	1	0	1	1	0
0	1	0	1	0	1
0	0	1	1	1	0

Такой алгоритм позволяет перебрать все граничные векторы и записать монотонно возрастающую линейно разделимую функцию в виде ДНФ. В данном примере ДНФ запишется как $f = x_1x_2 + x_1x_3 + x_1x_4 + x_1x_5 + x_2x_3x_4 + x_2x_3x_5 + x_2x_3x_6 + x_2x_4x_5 + x_2x_4x_6 + x_3x_4x_5$.

3 Реализация булева нейрона через функциональные элементы

Пусть нейронная сеть обучена и все связи и параметры нейронов жестко заданы. Это означает, что коэффициенты линейно разделимых функций нейронов известны. Тогда для реализации функции булева нейрона через функциональные элементы необходимо получить ее представление в виде ДНФ или КНФ минимальной сложности. Следующий алгоритм позволяет получить минимальную и кратчайшую форму.

3.1 Алгоритм получения минимальной и кратчайшей ДНФ функции булева нейрона

Предлагается следующий алгоритм получения совершенной ДНФ:

– Преобразование линейно разделимой функции нейрона в монотонно возрастающую. При этом коэффициенты новой функции будут:

$$w_i^* = |w_i|, \quad C^* = \sum_{\forall i: w_i < 0} |w_i| + C.$$

– Номера проинвертированных переменных необходимо запомнить. ДНФ монотонной функции будет отличаться от ДНФ булева нейрона инверсиями перед этими переменными.

– По указанному алгоритму по параметрам монотонно возрастающей функции получаем перечень граничных векторов, по которым однозначно записываем функцию в ДНФ.

– Искомая запись функции в ДНФ будет ДНФ монотонно возрастающей функции с инверсией некоторых переменных.

Указанный алгоритм позволяет получить запись функции булева нейрона в ДНФ.

3.2 Оценка индекса простоты полученной ДНФ логической функции булева нейрона

Для практического применения указанного способа реализации булева нейрона необходимо знать максимальную сложность ДНФ. Индекс простоты для полученной ДНФ функции нейрона будет в точности равен индексу простоты для монотонных функций. Для всех булевых функций из замкнутых классов оценка получена Лупановым [12]. Для монотонных функций оценки были проведены в работах [13, 14].

$$L(A_1) \sim \rho \cdot \sqrt{\frac{2}{\pi}} \cdot \frac{2^n}{n^{3/2}},$$

где n – число переменных, а ρ – коэффициент, зависящий от выбранной полной конечной системы функциональных элементов.

4 Выводы

В работе представлен алгоритм получения минимальной и кратчайшей дизъюнктивной нормальной формы функции булева нейрона. В соответствии с этим алгоритмом реализация формального нейрона проводится без использования формальных нейронных преобразований. При этом нейронная сеть должна быть предварительно полностью обучена, найдены все ее внутренние параметры. Логическая функция, представленная в кратчайшей и минимальной ДНФ булева нейрона, реализуется через функциональные элементы.

Использование ДНФ при проектировании схемы дает наименьшее число ступеней, равное двум. Это позволяет достичь максимальной скорости принятия решения нейронной сетью, в том числе и за счет абсолютной параллельности вычислений.

Фиксация булевого вида,возможно, сужает класс решаемых задач и увеличивает сложность системы в целом. Невозможность переобучения делает систему менее гибкой.

Литература

[1] FENT J., FRÖCHTEICHT, GAEDE F. ET AL. A Newral Network Second Level Trigger for the H1-Experiment at HERA // August 1995.

[2] GOGIBERIDZE G. L., MEKHDIEV R. R. γ-π^0 Discrimination With a Shower Maximum Detector Using Newral Networks For the Solenoidal Tracker At RHIC // JINR Rapid Communications. 1996. **2**(76), P. 17–24.

[3] ASTVATSATUROV A., BUDAGOV J., SHIGAEV V. ET AL. Identification Of b-Jet With a Low p_T Muon Using ATLAS Tile Calorimeter Simulation Data And Artificial Neural Networks Technique // Communication of the JINR. E10 96 288. Dubna. 1996.

[4] Кисель И. В., Нескоромный В. Н., Ососков Г. А. Применение нейронных сетей в экспериментальной физике // Физика ЭЧАЯ, 1993. **24**(6), С. 1551–1595.

[5] Гамкрелидзе С. А., Федченко О. И. Методика определения оптимальных параметров микропроцессорных СБИС и СБИС ЗУ для нейрокомпьютеров // Нейрокомпьютер, 1992. **2**, С. 43–51.

[6] Кирсанов Э. Ю. Оценка производительности нейрокомпьютеров // Нейрокомпьютер, 1992. **2**, С. 37–42.

[7] Галушкин А. И., Кирсанов Д. В. Заказной цифровой нейрочип // Нейрокомпьютер, 1992. **2**, С. 67–72.

[8] CNAPS/VME Board Reference Manual 1.0 // *Adaptive solutions Inc.*, Beaverton OR, 1993.

[9] Уоссермен Ф. Теория и практика нейронных сетей. М.: Мир, 1994.

[10] Зинов В. Г., Назаренко М. А. Обучение формальных нейронных сетей // Препринт ОИЯИ Р11–97–35, Дубна, 1997.

[11] Яблонский С. В. *Введение в дискретную математику*. 2-е изд. М.: Наука. 1986. (Часть V, Глава I, стр 326)

[12] Лупанов О.Б.. Об одном подходе к синтезу управляющих систем – принципе локального кодирования. // Проблемы кибернетики. **Выпуск 14**, 1965, стр. 31–110

[13] Угольников А.Б.. О реализации функций из замкнутых классов схемами из функциональных элементов // Математические вопросы кибернетики, 1988. **Выпуск 1**(9), стр. 89–113

[14] Угольников А.Б.. О реализации монотонных функций схемами из функциональных элементов. // Проблемы кибернетики **Выпуск 14**, 1976, стр. 167–185.

Рукопись поступила в издательский отдел
28 февраля 1997 года.