

УДК 00.004.27 + 004.3

РЕАЛИЗАЦИЯ ИСКУССТВЕННОГО НЕЙРОНА НА FPGA**Гонтаренко Б.В., Беловолова М.А., Зинченко Ю.Е.**Донецкий национальный технический университет
кафедра компьютерной инженерииE-mail: tuofil@gmail.com belovolovama@gmail.com**Аннотация**

Гонтаренко Б.В., Беловолова М.А., Зинченко Е.Ю. Реализация искусственного нейрона на FPGA. В данной статье рассмотрены принципы реализации искусственного нейрона – основной составляющей части искусственных нейронных сетей. Так же определены основные преимущества реализации нейронных сетей на программируемых логических интегральных схемах типа FPGA.

Введение

Современные компьютеры могут выполнять сложные вычисления, оперируют сложными функциями управления и могут хранить огромные объемы информации. Но существуют классы проблем, с которыми человек может справиться гораздо быстрее и лучше, чем компьютер. Примеры просты - распознавание символов, интерпретация изображений или чтение текста.

В отличие от компьютеров, человеческий мозг может адаптироваться к новым ситуациям и улучшать свои знания с помощью обучения. Он может обрабатывать недостоверную или неполную информацию и достигать желаемого результата. Это возможно благодаря адаптации человеческого мозга.

Искусственные нейронные сети (ИНС) - математические модели, а также их программные или аппаратные реализации, построенные по принципу организации и функционирования биологических нейронных сетей - сетей нервных клеток живого организма [1]. Это понятие возникло при изучении процессов, протекающих в мозгу, и при попытке смоделировать эти процессы.

Вычислительные системы, основанные на искусственных нейронных сетях, имеют ряд качеств, которые отсутствуют в машинах с архитектурой фон Неймана (но присущи мозгу человека):

- обучаемость;
- способность к обобщению;
- ассоциативное, распределенное хранение данных;
- адаптация к изменению окружающей среды;
- потенциально высокая производительность;
- отказоустойчивость аппаратной реализации.

Искусственные нейронные сети представляют собой систему соединенных и взаимодействующих между собой простых процессоров - искусственных нейронов. Такие процессоры обычно довольно просты, особенно по сравнению с процессорами, которые используются в персональных компьютерах. Каждый процессор подобной сети имеет дело только с сигналами, которые он периодически получает, и сигналами, которые он периодически посылает другим процессорам. И, тем не менее, будучи соединенными в достаточно большую сеть с управляемым взаимодействием, такие процессоры вместе способны выполнять довольно сложные задачи.

В целом, сложность реализации любой ИНС зависит от сложности реализации отдельного нейрона. Данная статья посвящена именно этой проблеме.

Актуальность поставленной задачи подтверждается тем, что развитие концепций реализации искусственных нейронных сетей на базе программируемых логических схем (в большей степени из-за удобства работы с ними и возможности быстрого их программирования) только набирает обороты.

Преимущества аппаратной реализации ИНС на FPGA

Существуют два способа реализации искусственных нейронных сетей - аппаратная и программная. Программная реализация, уступая аппаратной по скорости работы и автономности, обладает рядом очевидных преимуществ, связанных с простотой использования и внедрения в информационно-управляющую систему.

Однако специализированные аппаратные средства предлагают заметные преимущества в определенных ситуациях. К основным достоинствам аппаратной реализации ИНС можно отнести скорость, надежность, безопасность и дополнительные режимы эксплуатации.

Что касается аппаратной реализации нейронных сетей на FPGA, то она выгодно отличается от реализации на специальных DSP-процессорах (поскольку они выпускаются серийно) и от реализации на ASIC-микросхемах (поскольку они не подлежат переконфигурированию). Реализация на FPGA наиболее точно передает параллельную архитектуру нейронов и предоставляет возможность гибкого реконфигурирования всей нейронной сети и её составляющих – искусственных нейронов [2]. Так же FPGA – это сравнительно доступные схемы небольшой стоимости, что позволяет быстро и недорого реализовать необходимую систему. Кроме того, конфигурацию основанных на FPGA нейронных сетей легко изменить.

Математическая модель искусственного нейрона

Как было сказано выше, искусственный нейрон – это узел любой искусственной нейронной сети, представляющий собой значительно упрощенную модель биологического нейрона. Для большего понимания процессов, происходящих внутри биологической нейронной сети, МакКаллох и Питтс еще в 1949 году разработали математическую модель нейрона, которая используется до сих пор.

Математически искусственный нейрон обычно представляют как некоторую нелинейную функцию от единственного аргумента – линейной комбинации всех входных сигналов. Эту функцию называют функцией активации, функцией срабатывания или передаточной функцией нейрона. Полученный результат отправляется на единственный выход нейрона. Математическая модель искусственного нейрона представлена на рис. 1.

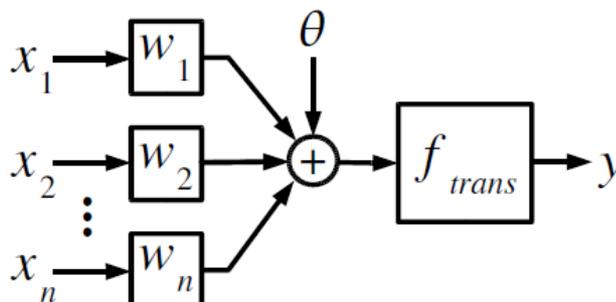


Рисунок 1 – Математическая модель искусственного нейрона

Функция, которую реализует эта модель, имеет вид: $y = \text{ftrans} \left(\sum_i x_i w_i + \theta \right)$. Все входы x_i умножаются на значения своих весов w_i , суммируются и дополняются некоторым значением θ . Затем сумма при помощи функции активации передается на выход нейрона.

Значение θ используется для инициализации нейрона. Под инициализацией имеется в виду смещение функции активации нейрона по горизонтальной оси, т.е. формирование порога чувствительности нейрона [3]. Кроме того, иногда к выходу нейрона специально добавляют некоторую случайную величину, которая называется смещением. Смещение можно рассматривать как сигнал на дополнительном, всегда нагруженном синапсе (синапсы – это связи, по которым выходные сигналы одних нейронов попадают на входы других).

Принципы и варианты реализации структурных составляющих

Структура искусственного нейрона, полностью соответствующая математической модели, представлена на рис. 2. Входы X_1 и X_2 умножаются, соответственно, на свои веса W_1 и W_2 , после чего складываются. К сумме добавляется переменная смещения B , результат суммирования проходит через блок функции активации, на выходе которого и формируется выходное значение нейрона.

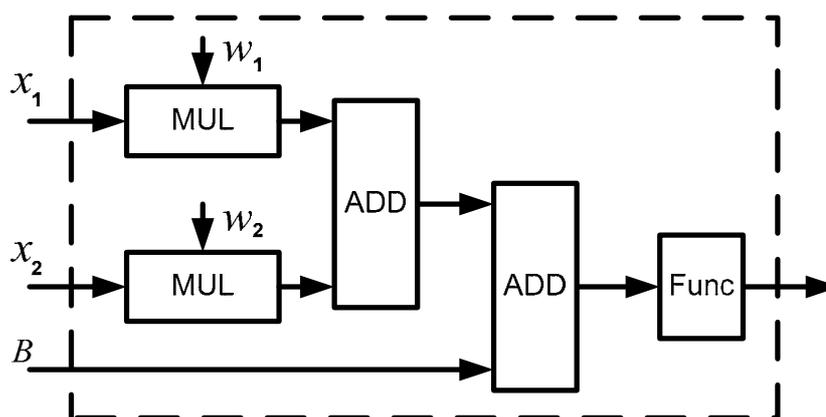


Рисунок 2 – Структура искусственного нейрона

Что касается реализации структуры входных/выходных портов на языке VHDL, стоит отметить, что есть два возможных варианта: использовать дополнительные входы реализуемой схемы для весовых коэффициентов, либо использовать для этой цели отдельную RAM. Выход схемы в обоих случаях будет один.

К основным компонентам искусственного нейрона относятся умножители, сумматор-аккумулятор и блок функции активации нейрона. Поскольку модели ИНС во многом зависят от массивных параллельных вычислений, чтобы обеспечить высокую скорость работы в режиме реального времени, нейронные сети должны быть реализованы с помощью применения параллельных аппаратных структур. Это, в свою очередь, становится причиной возникновения проблем при аппаратной реализации искусственных нейронных сетей, связанных как со сложностью реализации, так и с занимаемыми объемами кристалла.

Все это следует учитывать на этапе разработки основной структурной составляющей части ИНС – искусственного нейрона, что логично.

Умножители. В идеальном случае, количество умножителей должно быть равно количеству входов нейрона – только таким образом может быть достигнута максимальная

параллельность вычислений. С другой стороны, поскольку приходится иметь дело с дробными числами (как с фиксированной, так и с плавающей запятой), реализовать абсолютно параллельную ИНС не представится возможным, поскольку арифметические операции над дробными числами слишком ресурсоемки в плане занимаемого места на кристалле ПЛИС.

Есть два пути решения этой проблемы: упрощение и оптимизация умножителей и уменьшение количества умножителей, например, за счет использования всеми входами нейрона одного умножителя.

В разрабатываемом дипломном проекте, который посвящен реализации параметризованного искусственного нейрона в контексте создания нейронных сетей, будут рассмотрены оба варианта использования умножителей. Проблемы с использованием полностью параллельного нейрона не возникнет, поскольку размер нейронной сети не слишком большой, но с целью изучения времени работы нейрона будет рассмотрен и нейрон с одним умножителем.

Для обеспечения необходимой точности будут использованы умножители 16x16 бит. Поскольку проблем с переполнением возникнуть не может – все числа будут достаточно малы, выходное значение умножителя будет так же 16-битное.

Сумматор-аккумулятор. Сложность этого элемента так же определяется сложностью алгоритма обработки дробных чисел, но в плане потребления ресурсов кристалла здесь все проще – сумматор один, и с увеличением количества входов его размер не меняется. Используются 16-битные числа.

Функция активации. Функция активации определяет зависимость сигнала на выходе нейрона от взвешенной суммы сигналов на его входах. В большинстве случаев она является монотонно возрастающей и имеет область значений $(-1,1)$ или $(0,1)$. Бывает нескольких типов, причем в искусственных нейронных сетях чаще всего используется нелинейная функция активации - сигмоида. Функции активации представлены на рис. 3.

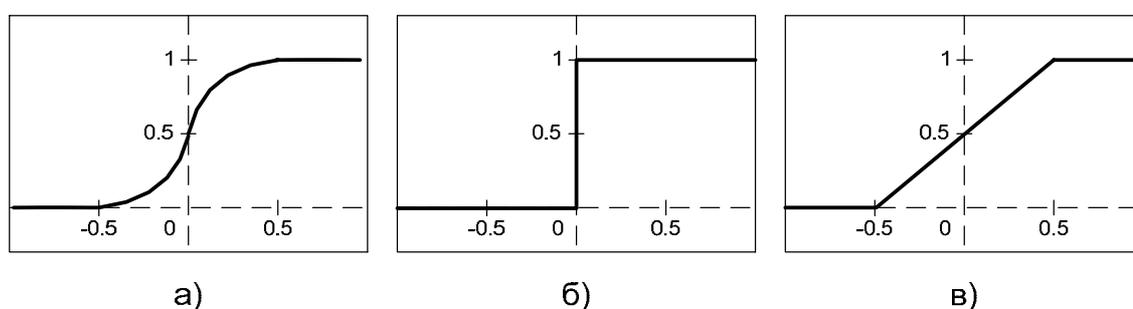


Рисунок 3 – Сигмоидальная (а), пороговая (б) и линейная (в) функции активации нейрона

В разрабатываемом дипломном проекте реализация функций активации искусственного нейрона занимает едва ли не важнейшее место, поскольку прямая реализация нелинейной сигмоидальной функции хоть и возможна, но очень проблематична [4].

Математически функция представляется следующим образом:

$$\sigma(x) = \frac{1}{(1 + \exp(-tx))}$$

Существует несколько способов упрощения аппаратной реализации сигмоидальной функции: использование поисковых таблиц (lookup table, по сути - память), различные виды кусочно-линейной аппроксимации, комбинационная аппроксимация.

В практической части дипломного проекта будут реализованы несколько вариантов из приведенных выше классов, будут приведены сравнительные характеристики по быстродействию и занимаемому на кристалле месту.

Выводы

При реализации искусственных нейронных сетей на FPGA следует понимать, какую роль играет возможность реконфигурации аппаратных средств и разрабатывать стратегии для эффективного использования аппаратных ресурсов программируемых интегральных схем. Немаловажными и проблемными этапами разработки нейронной сети являются: реализация нейрона в целом и привязанного к нему умножителя в частности и реализация функции активации нейрона и её аппроксимация.

Список литературы

1. Искусственная нейронная сеть. Материал из Википедии. Электронный ресурс. <http://ru.wikipedia.org/wiki>
2. Каллан Р. Основные концепции нейронных сетей — Вильямс, 2001. — 288 с.
3. Круглов В.В. Искусственные нейронные сети. Теория и практика — М.: Горячая линия - Телеком, 2001. — 382 с.
4. Бейд С.Л. Стохастические нейронные сети. IEEE Workshop on FPGAs for Custom Computing Machines. IEEE Computer Society Press, 2007. — 189 с.