

ПЕРСПЕКТИВЫ СОЗДАНИЯ ОПТОЭЛЕКТРОННЫХ НЕЙРОКОМПЬЮТЕРОВ В КОНТЕКСТЕ РАЗВИТИЯ СОВРЕМЕННЫХ НЕЙРОТЕХНОЛОГИЙ

Национальные и международные программы исследований в области нейροкомпьютеров сегодня охватывают разнообразные проблемы их создания и практического применения. Учебные курсы по нейронным сетям уже включены в учебные программы по информатике во многих университетах мира [1,2]. Наблюдается оживление этого сектора и в промышленности, где число фирм, специализирующихся в данной области, исчисляется сотнями. На рынке появляются новые промышленные изделия – нейрочипы, нейροкомпьютеры и нейроакселераторы для ПЭВМ, а также программные продукты, реализующие нейротехнологии на базе современных ЭВМ.

В последнее время активно ведутся работы по построению моделей обработки информации в нервной системе [2,3]. Большинство моделей основывается на схеме формального нейрона У. МакКаллока и У. Питтса, согласно которой нейрон представляет собой пороговый элемент, на входах которого имеются возбуждающие и тормозящие синапсы. В модели нейрона такого типа определяется взвешенная сумма входных сигналов (с учетом весов синапсов), а при превышении этой суммы порога нейрона вырабатывается выходной сигнал. Согласно существующим представлениям [3,4], нейронная сеть рассматривается как мультистабильная система, устойчивые состояния которой (аттракторы) соответствуют запомненным состояниям. Если внешнее воздействие устанавливает сеть в состояние, близкое к одному из аттракторов, то, будучи предоставлена сама себе, сеть за одну или несколько итераций конвергенции перейдет в состояние аттрактора. Таким образом, если исходное состояние отражает неполный или искаженный набор данных, ранее занесенных в память, то после конвергенции весь этот набор данных будет восстановлен. На этом базируется свойство ассоциативной памяти, позволяющее применять нейροкомпьютеры для распознавания образов, дешифрации закодированных данных, прогнозирования сложных процессов и т. п.

В моделях уже построены нейронные сети, выполняющие различные алгоритмы обработки информации [5]: ассоциативная память, разбиение множества образов на кластеры, состоящих из подобных друг другу образов, топологически корректное отображение одного пространства переменных в другое, распознавание зрительных образов, инвариантное относительно деформаций и сдвигов в пространстве решения задач комбинаторной оптимизации. Подавляющее число работ относится к исследованию алгоритмов нейросетей с прагматическими целями. Предполагается, что практические задачи будут решаться нейροкомпьютерами – искусственными нейроподобными сетями, созданными на основе микроэлектронных вычислительных систем. Спектр задач для разрабатываемых нейροкомпьютеров достаточно широк: распознавание зрительных звуковых образов, создание экспертных систем и их аналогов, управление роботами, создание нейропротезов для людей, потерявших слух или зрение. Достоинства нейροкомпьютеров – параллельная обработка информации и способность к обучению.

Несмотря на чрезвычайную активность исследований по нейронным сетям и нейροкомпьютерам, многое в этих исследованиях требует уточнения и развития. В особенности, при рассмотрении алгоритмов работы нейросетей сильно выражается некоторое их упрощение и рассмотрение отдельных функций работы нервной системы. Следует признать, что сегодня мы далеки от полного понимания структуры и функций относительно простых биологических систем и поэтому можно говорить о воспроизведении в нейροкомпьютерах лишь отдельных свойств нервной системы, в частности, способности к ассоциативному запоминанию информации и параллельному выполнению операций над всеми элементами обрабатываемых данных. Использование только этих свойств нервной системы позволяет эффективно осуществлять распознавание образов и принимать решения в условиях дефицита информации, т.е. решать те задачи, для которых отсутствуют надежные формальные методы [2].

Создаваемые сегодня нейροкомпьютеры базируются на относительно простых принципах, состоящих в систематическом применении линейных решающих правил для классификации состояний фрагментов системы.

Одним из основных достоинств нейровычислителя является то, что его основу составляют относительно простые, чаще всего - однотипные, элементы (ячейки), имитирующие работу нейронов мозга - "нейроны". Каждый нейрон характеризуется своим текущим состоянием по аналогии с нервными клетками головного мозга, которые могут быть возбуждены или заторможены. Он обладает группой синапсов - однонаправленных входных связей, соединенных с выходами других нейронов, а также имеет аксон - выходную связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов. Общий вид нейрона приведен на рисунке 1.

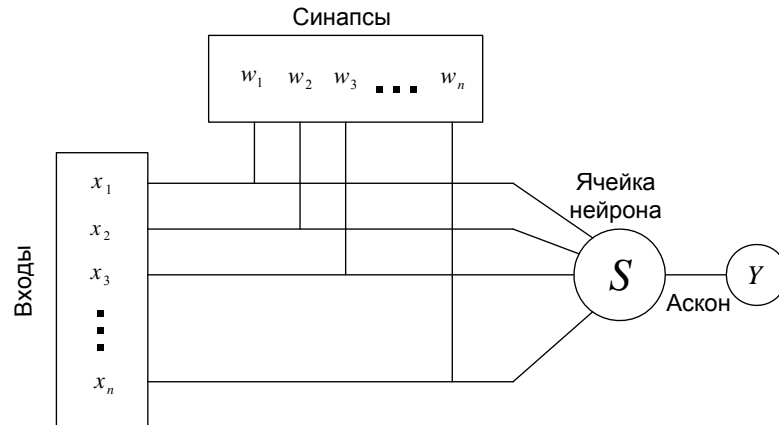


Рис. 1. Общий вид нейрона.

Каждый синапс характеризуется величиной синаптической связи или ее весом W_i , который по физическому смыслу эквивалентен электрической проводимости [6]. Текущее состояние нейрона определяется, как взвешенная сумма его входов:

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

Выходом нейрона является функция его состояния:

$$y = f(x) , \quad (2)$$

которая называется активационной и может иметь различный вид [5,6]. Одной из наиболее распространенных является нелинейная функция с насыщением, так называемая логистическая функция (функция S-образного вида):

$$f(x) = \frac{1}{1 + e^{-\alpha x}} . \quad (3)$$

При уменьшении α данная функция становится более пологой, при $\alpha = 0$ вырождается в пределе в горизонтальную линию на уровне 0.5, при увеличении α - приближается по внешнему виду к функции единичного скачка с порогом p в точке $x = 0$. Из выражения (3) очевидно, что выходное значение нейрона лежит в диапазоне $[0,1]$. Одно из ценных свойств логистической функции - простое выражение для ее производной, в частности

$$f'(x) = \alpha f(x)(1 - f(x)) . \quad (4)$$

Параллелизм обработки достигается путем объединения большого числа нейронов в слои и соединения определенным образом различных нейронов между собой. Очевидно, что все весовые коэффициенты синапсов одного слоя нейронов можно свести в матрицу W , в которой каждый элемент w_{ij} задает величину i -ой синаптической связи j -ого нейрона. Таким образом, процесс, происходящий

в нейронной сети (НС), может быть записан в матричной форме: $Y = F(XW)$, где X и Y - соответственно входной и выходной сигнальные векторы, $F(XW)$ - активационная функция, применяемая поэлементно к компонентам вектора XW . Теоретически число слоев и число нейронов в каждом слое может быть произвольным.

Для того чтобы нейронная сеть работала – ее надо обучить [3]. От качества обучения зависит способность сети решать поставленные перед ней проблемы. На этапе обучения, кроме параметра качества подбора весовых коэффициентов, важную роль играет время обучения. Как правило, эти два параметра связаны обратной зависимостью и их приходится выбирать на основе компромисса. Обучение НС может вестись с учителем или без него. В первом случае сети предъявляются значения как входных, так и желательных выходных сигналов, и она по некоторому внутреннему алгоритму подстраивает веса своих синаптических связей. Во втором случае выходы НС формируются самостоятельно, а веса изменяются по алгоритму, учитывающему только входные и производные от них сигналы.

Рассматривая классификацию НС можно выделить следующие системы: бинарные (цифровые) и аналоговые НС, предварительно обученные (неадаптивные) и самообучающиеся (адаптивные) нейронные сети, что крайне важно при их аппаратной реализации. Бинарные оперируют с двоичными сигналами, и выход каждого нейрона может принимать только два значения: логический ноль ("заторможенное" состояние) и логическая единица ("возбужденное" состояние). В аналоговых сетях выходные значения нейронов могут принимать непрерывные значения, что могло бы иметь место после замены активационной функции нейронов персептрона на логистическую функцию.

Сети также можно классифицировать по топологии (числу слоев и связей между ними). При этом нелинейность активационной функции имеет конкретный смысл: так как, если бы она не обладала данным свойством или не входила в алгоритм работы каждого нейрона, результат функционирования любой m -слойной НС с весовыми матрицами $W_i, i = \overline{1, m}$ для каждого слоя i сводился бы к

перемножению входного вектора сигналов X на матрицу $W = \sum_{i=1}^m W_i$, то есть фактически такая m -слойная НС эквивалентна однослойной НС с весовой матрицей единственного слоя.

Выигрыш при использовании нейросетевого подхода, по сравнению с традиционным, заключается в том, что при его использовании возможно проводить вычисления параллельно, а это в свою очередь дает возможность реализовать системы управления более высокого порядка при приемлемых показателях сходимости и, следовательно, добиться более высокого качества управления. Нейросетевой подход к реализации многомерных пространственных систем управления во многом снимает проблемы стоявшие перед разработчиками по необходимости выполнения векторно-матричных операций высокой размерности в реальном времени.

В общем случае проектирование нейросистем сложный и трудоемкий процесс [3], в котором выбор конкретного алгоритма - это только один из нескольких шагов процесса проектирования. Он, как правило, включает: исследование предметной области, структурно-функциональное проектирование, топологическое проектирование и т.д. Вопросы, стоящие перед разработчиками нейросетевой элементной базы и нейровычислителей, во многом сложны и требуют дополнительных исследований: как целесообразней реализовать нейрочип - со встроенными нелинейными преобразованиями (пусть и фиксированного набора, но реализованных аппаратно) или позволить программисту - разработчику нейронной сети, самому формировать активационную функцию программно (размещая соответствующий код в ПЗУ).

Следует отметить, что эффективное применение нейрокомпьютеров характерно для случаев, требующих резкого сокращения времени обработки при решении пространственных задач повышенной размерности, которые существуют практически в любой сфере деятельности.

Однако, несмотря на положительные результаты исследований, следует отметить, что ожидаемый рост приложений нейротехнологий едва ли приведет в обозримом будущем к замене ЭВМ традиционного типа нейрокомпьютерами. Приобретая в результате новейших усовершенствований способность быстро усваивать неформализованные данные, нейрокомпьютеры должны будут пройти еще длительный путь эволюции, прежде чем смогут столь же успешно оперировать формализованной информацией, компактно выражающей накопленный опыт и знания. Поэтому в обозримом будущем при построении интеллектуальных ЭВМ наиболее вероятно применение гибридных систем, сочетающих элементы традиционной и нейросетевой архитектур. Для этих систем предпочтительна программная реализация нейросетевых компонент с использованием векторных сопроцессоров в качестве нейроакселераторов при необходимости повысить производительность нейровычислений. Аппаратные

нейрокомпьютеры на базе нейрочипов или (в более отдаленном будущем) оптоэлектронных и квантовых процессоров по-видимому продолжают укреплять свои позиции в качестве высокопроизводительных параллельных процессоров для адаптивных систем реального времени [7].

Продолжая анализ вышеуказанных проблем необходимо указать на большие возможности, связанные с использованием оптоэлектронных технологий в плане дальнейшего развития нейроархитектур. В данном случае речь идет об эффективности использования оптоэлектронных процессоров в системе развития нейротехнологий, что позволяет с большим оптимизмом смотреть в будущее нейрокомпьютеров и нейротехнологий.

Действительно, если носителем информации в нейрокомпьютере является свет, то можно резко, на два-три и более порядков повысить производительность вычислений за счет следующих факторов [4,7,8]:

- Свет позволяет физически организовать обрабатываемую информацию в виде двумерных пространственных кадров большой размерности (128x128, 256x256, 512x512, 1024x1024 пикселей), в отличие от электричества, которое позволяет физически организовать обрабатываемую информацию в виде одномерных временных последовательностей импульсов напряжения.

- Лучи света пересекаются в трехмерном пространстве без взаимодействия друг с другом, поэтому в трехмерном пространстве можно полностью параллельно умножать двумерный входной сигнал на четырехмерную матрицу весовых коэффициентов и получать двумерную матрицу извещенных сумм.

- Оптическая элементная база (LCLV) позволяет полностью параллельно выполнять над двумерными кадрами арифметические операции сложения, вычитания и нелинейного преобразования, поэтому возможно полностью оптическое параллельное вычисление двумерного выходного сигнала или двумерного обучающего сигнала ошибки.

- Оптическая элементная база (фоторефрактивные кристаллы, матрицы нелинейных эталонов и т.п.) позволяет работать не только с бинарными (0:1) двумерными сигналами, но и с аналоговыми двумерными сигналами, аппаратно можно реализовать до восьми градаций яркости на одном пикселе, а объединяя в один макропиксел группу из нескольких простых пикселей — и того больше.

- Такая сильная сторона электроники, как высокая точность представления данных, очень важная для цифровых вычислителей, при аппаратной реализации НС имеет гораздо меньшее значение, чем такая сильная сторона оптики как возможность организовать в ограниченном объеме огромное число независимых параллельно действующих каналов связи.

- Оптический сигнал можно модулировать не только по амплитуде, как электронный, но и по фазе, и по поляризации, и даже по цвету — по длине волны, например, записывать трехмерные голограммы лучом с длиной волны λ_1 , а выполнять неразрушающее считывание лучом с λ_2 . Это расширяет возможности разработчиков нейрокомпьютеров.

- Пассивные элементы оптики, такие как двумерные и трехмерные голограммы и диффракционные решетки, с помощью которых могут выполняться трудоемкие операции умножения двумерного оптического входного сигнала на четырехмерную матрицу весовых коэффициентов, вообще не требуют энергии. Несущий информацию двумерный оптический сигнал просто перераспределяется в соответствии с шагом сформированной в фоторефрактивном кристалле трехмерной или на стекле двумерной диффракционной решетки. В интегральной электронике что-то подобное невозможно. При перемножении массивов данных в СБИС все элементы активны, непременно должно работать множество сумматоров, умножителей, элементов памяти, каждый из которых потребляет и рассеивает энергию. Это приводит к тому, что для обработки одинакового (большого) объема информации оптический нейрокомпьютер, скорее всего, потратит гораздо меньше энергии, чем электронный аналог.

- Элементная база для оптических нейрокомпьютеров в основном создана, свойства этого оборудования уже долго исследовались, эксплуатационные качества оптимизированы. Для создания оптических нейрокомпьютеров необходимо только, как это уже давно делают за рубежом, собрать готовые образцы элементной базы в одном месте и построить из них такую архитектурную конструкцию, которая бы реализовывала правильные, т.е. адекватные сильным сторонам оптической элементной базы нейросетевые алгоритмы решения социально важных задач.

Поэтому все больше возрастает интерес к использованию оптических нейрокомпьютеров в качестве специализированных вычислительных структур с высокой производительностью во многих направлениях прикладного характера.

Таким образом, на основании проведенного исследования состояния и перспектив развития нейротехнологий, посредством теоретического анализа раскрываются преимущества оптики для аппаратной реализации пороговой логики и нейрокомпьютеров в перспективе создания нового поколения мощных оптических нейрокомпьютеров, способных воспринимать и параллельно

обрабатывать информацию в реальном масштабе времени, с использованием новых методов обучения и алгоритмов работы.

ЛИТЕРАТУРА

1. Морозов А.А., Клименко В.П., Резник А.М. Нейрокомпьютеры и нейротехнологии: накануне нового старта // УсиМ – 1997 – №1/3.
2. А.И.Галушкин Некоторые исторические аспекты развития элементной базы вычислительных систем с массовым параллелизмом (80- и 90- годы) // Нейрокомпьютер, №1. 2000. - С.68-82
3. Проблемы построения и обучения нейронных сетей / под ред. А.И.Галушкина и В.А.Шахнова. - М. Изд-во Машиностроение. Библиотечка журнала Информационные технологии №1. 1999. - 105 с.
4. Кожем'яко В.П., Хамді Рамі Р.М., Гордієнко Д.В., Баштовий Ю.В. Синтез нейронних систем на квантрон-автоматах з використанням енергетичного нуля // Вісник ВПІ – 1999 – №6.
5. А.И. Власов. Аппаратная реализация нейровычислительных управляющих систем // Приборы и системы управления - 1999, №2, С.61-65.
6. Э.Ю. Кирсанов Цифровые нейрокомпьютеры: Архитектура и схемотехника / Под ред. А.И.Галушкина. - Казань: Казанский Гос. У-т. 1995. 131 с.
7. Кожем'яко В.П. Погляд на природу штучного інтелекту // Вісник ВПІ - 1997 р. - №1.
8. Степанов М.В. Оптические нейрокомпьютеры: современное состояние и перспективы // Зарубежная радиоэлектроника – 1997 – №2.