**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**
**им. М.В. Ломоносова**

Факультет вычислительной математики и кибернетики

На правах рукописи

Коробкова Светлана Викторовна

# РАЗРАБОТКА, ИССЛЕДОВАНИЕ И РЕАЛИЗАЦИЯ АЛГОРИТМОВ ПРОГРАММНОГО КОМПЛЕКСА «НЕЙРОМАТЕМАТИКА»

Специальность 05.13.11 - математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва – 2008

Работа выполнена на кафедре автоматизации систем вычислительных комплексов факультета вычислительной математики и кибернетики Московского государственного университета имени М.В. Ломоносова

Научные руководители:

доктор технических наук
профессор Галушкин Александр Иванович

кандидат физико-математических наук
доцент Попова Нина Николаевна

Официальные оппоненты:

Член корреспондент РАН
доктор физико-математических наук
профессор Четверушкин Борис Николаевич

доктор физико-математических наук
профессор Тархов Дмитрий Альбертович

Ведущая организация:

Институт прикладной математики им. М.В. Келдыша
Российской академии наук

Защита состоится 25 апреля 2008 г. в $11^{00}$ на заседании диссертационного совета Д 501.001.44 в Московском государственном университете имени М.В. Ломоносова по адресу: 119991, Москва, Ленинские горы ГСП-2, МГУ, факультет вычислительной математики и кибернетики, ауд. 685.

С диссертацией можно ознакомиться в библиотеке факультета ВМиК МГУ. С текстом автореферата можно ознакомиться на официальном сайте ВМиК МГУ им. М.В. Ломоносова http://www.cmc.msu.ru в разделе «Наука» – «Работа диссертационных советов» – «Д 501.001.44»

Автореферат разослан _____ 2008 г.

Ученый секретарь
диссертационного совета
профессор _____ Трифонов Н.П.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

## 1. Актуальность исследования

Диссертационная работа посвящена актуальной теме исследования и разработки методов построения программных комплексов, ориентированных на нейросетевое решение практических задач с учетом аппаратных возможностей современных вычислительных систем, обеспечивающих параллельную обработку данных. Важными свойствами таких программных комплексов является обеспечение высокой эффективности работы готовых приложений при решении практических задач, удобный пользовательский интерфейс и средства контроля параллельного выполнения задач.

В настоящее время существует ряд реализаций конкретных нейросетевых пакетов. Исследуемый в диссертации подход направлен на разработку программной среды, позволяющей создавать решения разных практических задач с обеспечением возможности добавления новых алгоритмов нейросетевой обработки, снижения сложности программирования новых практических задач, повышения скорости работы готовых приложений.

Основное внимание при рассмотрении нейросетевого подхода к решению задач, как правило, уделяется различным алгоритмам нейросетевой обработки. Теоретические аспекты нейросетевой обработки данных активно изучаются как в России, так и за рубежом. В данной работе рассмотрены вопросы реализации методов нейросетевой обработки данных и методов построения программных средств на их основе.

Основными свойствами такого нейросетевого программного комплекса являются возможность реализации решения конкретных практических задач большой размерности с использованием возможности параллельного счета на кластерных системах и графических платах, возможность расширения программного комплекса путем добавления модулей решения практических задач, расширения библиотеки нейросетевых алгоритмов, поддерживаемых программным комплексом, а также поддержки новых средств аппаратного ускорения вычислений.

Предлагаемые в диссертационной работе методы построения нейросетевых программных систем реализованы на примере программного комплекса «Нейроматематика».

Особенностью решения практических задач большой размерности с помощью нейронных сетей является сложность формирования обучающей выборки достаточного размера и повышенные требования к точности и скорости работы алгоритмов решения задач.

Для реализации предъявленных требований необходима разработка нейросетевых алгоритмов и подходов, позволяющих оптимизировать разработку нейросетевых решений практических задач и улучшить время и точность их решения.

В диссертационной работе рассматриваются вопросы, связанные с формированием обучающей выборки и заданием начальных условий при обучении нейронных сетей и эффективной параллельной реализации нейросетевых алгоритмов.

При обучении нейронной сети возникает проблема формировании обучающей выборки, когда сбор данных является трудоемким или дорогостоящим процессом. Методы предлагаемые в данной работе направлены на уменьшение затрат на формирование выборки и сбор данных.

В настоящее время, как правило, задание начальных условий при обучении нейронных сетей осуществляются на базе случайного выбора. Задание специальных начальных условий позволяет отказаться от случайного выбора. Выбор начального состояния сети с учетом специфики задачи позволяет получить как ускорение процесса обучения, так и улучшение точности получаемого результата. В диссертационной работе предложены методы задания начальных условий нейронной сети для довольно широкого класса решаемых задач.

Большие объемы данных, которые требуется обрабатывать во многих практических задачах, и естественный внутренний параллелизм нейронных сетей обуславливают необходимость параллельной реализации нейросетевых алгоритмов решения практических задач большой размерности.

Предлагаемые и реализованные в рамках программного комплекса «Нейроматематика» методы направлены на повышение эффективности его работы.

## 2. Цели и задачи исследования

Целью исследования является разработка методов построения программных комплексов направленных на нейросетевую обработку данных и решение практических задач большой размерности.

## 3. Объект исследования

Объектом исследования являются программные комплексы, ориентированные на нейросетевую обработку данных.

## 4. Предмет исследования

Предметом исследования являются методы построения нейросетевых программных комплексов и их реализация в рамках программного комплекса «Нейроматематика». Для обеспечения эффективного решения практических задач с использованием нейросетевых комплексов, построенных по разработанной методике, требуется разработка методов повышения точности и скорости работы нейросетевых алгоритмов. Существенными факторами, влияющими на точность и быстродействие нейросетевых методов решения задач, помимо правильно выбранной архитектуры нейронной сети и алгоритмов обучения, являются:

- Формирование обучающей выборки для задачи, в частности, методы работы с малой выборкой;
- Задание начальных условий для обучения нейронной сети, позволяющих сократить время обучения нейронной сети;
- Эффективная реализация нейросетевых алгоритмов на параллельных вычислительных системах и графических платах.

Предметом исследования диссертационной работы также является реализация решения конкретных практических задач в рамках предлагаемого подхода.

## 5. Методологическая и теоретическая основа исследования

При выполнении исследования были использованы теоретические основы нейронных сетей, методы построения сложных программных систем, теория разностных схем. Исследование, проводимое в диссертационной работе, основано на методике разработки нейросетевых алгоритмов решения задач, предложенной проф. А.И. Галушкиным[1].

## 6. Информационная база исследования

В работе использованы материалы трудов отечественных и зарубежных авторов в области теории нейронных сетей и аппаратной поддержки нейросетевых вычислений, документация и описания существующих нейросетевых программных пакетов, а также работы, посвященные исследованиям в прикладных областях (аэродинамика, горно-обогатительная промышленность, обработка изображений). Также были использованы материалы, опубликованные на международных конференциях по нейронным сетям и прикладным областям.

## 7. Научная новизна исследования

- Предложен новый подход к построению программных комплексов, ориентированных на нейросетевую обработку данных. В рамках предложенного подхода разработан программный комплекс «Нейроматематика».
- Разработаны новые методы и алгоритмы, направленные на повышение эффективности нейросетевой обработки данных:
  - Предложены новые методы, позволяющие настраивать нейронную сеть в условиях ограниченности обучающей выборки.

---

[1] *Решение задач в нейросетевом логическом базисе / Галушкин А.И. // Нейрокомпьютеры: разработка, применение. - 2006. - № 2. - С. 49-70. - Библиогр.: 7 назв.*

- ♦ Предложен новый подход к заданию специализированных начальных условий для обучения нейронной сети, который позволяет, в отличие от ранее использовавшихся методов, ускорить процесс обучения и повысить точность работы обученной сети.
- ♦ Предложены новые методы распараллеливания нейросетевых алгоритмов для графических плат и многопроцессорных вычислительных систем с общей и распределенной памятью.

- Разработаны алгоритмы и модули для решения практических задач определения гранулометрического состава окомкованного материала и задач аэродинамики.

**8. Практическая значимость работы**

Результаты, полученные в данной работе, могут быть использованы при построении нейросетевых программных комплексов, а также для решения практических задач в различных областях (Горнодобывающая промышленность, самолетостроение, строительство и планирование городов, экономика, энергетика). Возможность применения разработанных методов показана при решении практических задач.

**9. Апробация результатов исследования**

Основные положения диссертационной работы докладывались и обсуждались на следующих конференциях и семинарах:

- 5-я Международная научно-техническая Конференция и Выставка Цифровая обработка сигналов и ее применения, DSPA'2003, Москва, Россия, март 2003;
- XLVI Научная конференция МФТИ, 2003;
- 6-я Международная научно-техническая Конференция и Выставка Цифровая обработка сигналов и ее применения, DSPA'2004, Москва, Россия, март 2004;
- 11ая Международная конференция по нейронной обработке информации, ICONIP'2004, Калькутта, Индия, ноябрь 2004;
- An IEEE sponsored international conference on Artificial Intelligence Systems, Divnomorsk 2005;
- XLVIII Научная конференция МФТИ, Москва, 2005;
- Международный семинар по струйным, отрывным и нестационарным течениям. Санкт-Петербург, Россия, июнь 2006;
- International Joint conference on Neural Networks, IJCNN 2006, Vancouver, Canada, July 2006;
- 9-я Международная научно-техническая Конференция и Выставка Цифровая обработка сигналов и ее применение - DSPA-2007, март 2007;
- XV Международная Конференция по вычислительной механике и современным прикладным системам, ВМСППС 2007, Украина, Алушта, май 2007;
- International Joint conference on Neural Networks, IJCNN'07, Orlando, USA, August 2007;
- Российско-ирландская летняя школа по научным вычислениям, Москва-Ирландия, август 2007;
- Научная конференция Тихоновские чтения, Москва октябрь 2007.

**10. Перечень базисных положений, выносимых на защиту**

На защиту выносятся следующие положения:

- Методы построения программных средств для решения задач большой размерности на основе нейросетевого подхода и разработка программного комплекса «Нейроматематика».

- Методы работы с малой выборкой при обучении нейронных сетей.

- Методы задания начальных условий при обучении нейронных сетей.

- Реализация нейросетевых алгоритмов на графических платах и вычислительных системах с параллельной архитектурой.

- Оценка эффективности разработанных методов и нейросетевых алгоритмов при решении задач определения гранулометрического состава окомкованного материала и расчета аэродинамических параметров среды.

**11. Краткое описание структуры диссертационной работы**

Диссертационная работа содержит введение, три главы, заключение, приложение, библиографию работ по теме диссертации. Диссертация содержит 55 рисунков и 11 таблиц. Общий объем диссертации составляет 104 страницы.

## СОДЕРЖАНИЕ РАБОТЫ

Диссертационная работа состоит из введения, основной части и заключения. Во введении представлена тема исследования, обосновывается его актуальность, отмечаются основные цели и задачи исследования, приводится описания объекта и предмета исследования, а также описание методов исследования и содержание информационной базы проводимого исследования. Также в данном разделе указывается научная новизна исследования, практическая применимость его результатов и практические работы, на которых проведена апробация результатов исследования.

Содержание диссертации логически разделено на три части. Первая часть, представленная первой главой диссертации, посвящена методам построения программных комплексов, ориентированных на нейросетевую обработку данных и их реализации в программном комплексе «Нейроматематика».

Вторая часть – вторая глава посвящена новым алгоритмам нейросетевой обработки информации, используемым в программном комплексе «Нейроматематика» и позволяющим повысить точность и сократить время решения задач.

Третья часть работы посвящена применению разработанных методов при решении практических задач и оценке их эффективности. Данная часть включает в себя 3 главу, в которой приведены постановки практических задач и алгоритмы их решения и представлены результаты решения задач и их сравнение с результатами, полученными классическими методами.

Первая глава посвящена разработке методов построения программных комплексов, ориентированных на нейросетевую обработку данных с учетом большой размерности. Предложена структура такого программного комплекса и ее реализация на примере пакета «Нейроматематика».

Отличительными чертами программного комплекса «Нейроматематика» являются возможность решения практических задач, описанных готовыми модулями программного комплекса «Нейроматематика», возможность использования заложенных в него нейросетевых алгоритмов для разработки решения других практических задач, что обеспечивается открытым интерфейсом программного комплекса. Программный комплекс «Нейроматематика» предоставляет возможность параллельной реализации расчета задач с различной аппаратной поддержкой. «Нейроматематика» может использоваться как специализированное средство нейросетевого решения реальных задач.

С точки зрения классификации программных комплексов, использующих нейросетевые методы обработки информации, данный программный комплекс можно отнести как к специализированным средствам решения практических задач, так и к классу универсальных пакетов, с возможностью аппаратной поддержки вычислений.

Согласно представленным ранее принципам построения нейросетевых программных комплексов нейросетевой программный комплекс, обеспечивающий возможность формирования готовых решения задач большой размерности можно представить двумя основными блоками:

- ядро программного комплекса
- модули решения практических задач (РПЗ).

Каждая составляющая часть программного комплекса, как ядро, так и модули решения практических задач, предлагается строить по технологии клиент-сервер (структура программного комплекса представлена на рис. 1).

Рис. 1 Предлагаемая структура нейросетевого программного комплекса

В первой главе диссертационной работы обоснована общая архитектура нейросетевых программных комплексов для решения практических задач большой размерности, а также структура и схема взаимодействия его составных частей, библиотек и структур данных с учетом разработанной методики.

При разработке программного комплекса «Нейроматематика» по предложенной методике был использован объектно-ориентированный подход, с использованием механизма виртуальных функций. Основным языком программирования был выбран язык C++, для реализации дополнительных возможностей распараллеливания на кластерные вычислительные системы использован интерфейс передачи сообщений MPI и механизм сокетов, для распараллеливания алгоритмов на графических платах – язык Cg и библиотека OpenGL.

Во второй главе представлены описания предлагаемых методов и алгоритмов нейросетевой обработки данных, позволяющих повысить точность и сократить время решения задач.

В главе исследована методика разработки нейросетевого алгоритма решения практических задач, предложенная А.И. Галушкиным. Следуя этой методике, определены этапы разработки, на которых следует искать пути повышения эффективности нейросетевых алгоритмов решения практических задач. Среди них:

1) формирование обучающей и тестирующей выборки (для решаемой задачи);
2) задание адекватных начальных условий;
3) параллельная реализация нейросетевых алгоритмов.

В этой главе разработаны методы формирования выборок, задания начальных условий для обучения нейронной сети. Также предлагаются перспективные методы алгоритмической и аппаратной параллельной реализации выбранных нейросетевых парадигм.

Проблема формирования корректных выборок (обучающей и тестовой) чрезвычайно важна для получения качественного результата работы сети после ее обучения. Получение экспериментальных данных большого размера бывает невозможно из-за сложности проведения экспериментов, их стоимости и временных затрат на их проведение. Кроме того, в некоторых

практических задачах возникает необходимость принятия решения на начальном этапе проведения эксперимента, когда выборка еще достаточно мала, но имеется возможность ее увеличения во времени. Ограниченность количества обучающих примеров приводит к своеобразным особенностям в разработке архитектуры, алгоритмов обучения и тестирования нейронных сетей.

Предложены следующие методы построения и обучения нейронных сетей в случае, когда доступна небольшая выборка известных примеров:

1) случайная генерация дополнительных примеров с математическим ожиданием в виде составляющих исходной малой выборки и различной дисперсией;
2) метод настройки коэффициентов нейронной сети, основанный на делении выборки на подвыборки;
3) комбинированный метод: усреднение результатов настройки по множеству вариантов методов а) и б).

Выбор начальных условий для настройки многослойной нейронной сети является специфическим для каждой конкретной задачи, решаемой нейронной сетью и ставит целью ускорение процесса обучения путем введения нейронной сети в область глобального экстремума функционала вторичной оптимизации.

Нейронную сеть можно рассматривать как нелинейный фильтр. В работе предлагается в качестве одного из способов задания начальных условий нейронной сети использовать выбор весовых коэффициентов нейронной сети, реализующих эквивалентный линейный фильтр.

Расширением способа задания начальных условий нейронной сети с помощью фильтров является задание начальных условий нейронной сети как решения упрощенной модельной задачи, соответствующей решаемой практической задаче. Например, можно рассматривать упрощенные по сравнению с реальной системой объекты, а также пренебречь шумами, которые обязательно присутствуют в реальной системе. Для модельной задачи генерируется обучающая выборка и на ней обучается нейронная сеть. Коэффициенты, полученные после обучения, считаются начальными условиями для нейронной сети при обучении на реальных данных.

Параллельная реализация работы нейронных сетей для расчета на классических кластерных архитектурах (с общей и разделяемой памятью) рассматривается и решается как с алгоритмической, так и с аппаратной сторон.

За основу параллельной реализации нейросетевых алгоритмов в данной работе выбраны клеточные нейронные сети[2].

Клеточную нейронную сеть можно разбить на группу меньших клеточных нейронных сетей, работающих независимо друг от друга на каждом расчетном шаге с последующих согласованием значений граничных нейронов.

При реализации клеточных нейронных сетей на кластерных архитектурах возникает проблема передачи данных между процессорными элементами. Такую передачу требуется выполнять после каждого шага расчета в процессе работы нейронной сети. Так, при размере нейронной сети $n \times n \times n$ и доступном количестве процессорных элементов $M$ каждый процессорный элемент должен производить вычисления подобласти из $n \times n \times \frac{n}{M}$ нейронов, а объем передаваемых пограничных данных составит $n \times n$ значений нейронов. При этом для каждого нейрона на каждом шаге расчета требуется выполнить не более $K$ операций с плавающей точкой. Общее количество операций с плавающей точкой на отдельном шаге для отдельного процессорного элемента составит $K \times n \times n \times \frac{n}{M}$ операций. Учитывая, что в каждом нейроне хранится $N$ значений типа float, то на каждом шаге требуется передавать $N \times n \times n$ значений с плавающей точкой. Следовательно, время расчета одного шага (для процессора выполняющего $Num_op \times 10^9$

[2] *Chua L. O., Yang L.* Cellular Neural Networks: Theory // IEEE Transactions on Circuits and Systems. – 1988, vol. 35. – No 10. – P. 1257 – 1272

операций в секунду) составит не менее $\frac{K \times n \times n \times \frac{n}{M}}{Num_op \times 10^9}$ секунд, а время передачи информации (для кластера, организованного на Gigabyte Ethernet, состоящего из $L$ узлов) составит $L \times \frac{N \times n \times n \times 32}{10^9}$ секунд. Сравним время счета одного шага для кластера из $M$ процессорных элементов и для варианта последовательной программы, получим соответственно $\frac{K \times n \times n \times \frac{n}{M}}{Num_op \times 10^9} + L \times \frac{N \times n \times n \times 32}{10^9}$ и $\frac{K \times n \times n \times n}{Num_op \times 10^9}$ секунд, т.е. при увеличении количества процессорных элементов в $M$ раз достигается ускорение работы в $\frac{K \times n \times M}{K \times n + 10^4 \times M}$ раз. Для исследуемой системы из 16 процессоров, при расчете области общей размерностью $1024 \times 1024 \times 1024$, при $K = 100$ было получено семикратное ускорение.

Для улучшения времени кластерной реализации клеточных нейронных сетей большой размерности, как и в варианте простого распараллеливания счета, разобьем исходную нейронную сеть на группу меньших клеточных нейронных сетей, считаемых независимо друг от друга на каждом расчетном шаге с последующих согласованием значений граничных нейронов., аналогично неоптимизированному случаю.

Для перехода к следующему шагу расчета необходимо согласование значений граничных нейронов.

Разработаны 3 варианта алгоритма передачи значений на границе области, рассчитываемой на одном процессорном элементе, позволяющих сократить объем передаваемой информации:

1) ***Метод флага***. Чтобы минимизировать количество передаваемой информации можно ввести дополнительную переменную, отвечающую за состояние границы. В случае, если нулевой процесс не изменял правую границу области, то ее пересылка первому процессу не требуется. Пусть значение переменной right_border_state = 1 означает, что значения нейронов на границе не были изменены, а значение 0, что хотя бы один из этих нейронов был изменен. Тогда перед передачей границы будем пересылать значение этой переменной соседнему процессу. Дальнейшая пересылка данных будет происходить, только если right_border_state = 0.
2) ***Метод экстраполятор.*** Еще одним способом минимизации пересылаемой информации является введение на границе блока прогнозирования. Данный блок по предыдущему (или по значениям в течение нескольких предыдущих итераций) значению нейронов границе выдает значение данных нейронов на следующей итерации. Тогда можно осуществлять пересылку граничных значений один раз для нескольких итераций счета для корректировки прогнозных значений. Данный блок может быть реализован с использованием алгоритмов прогнозирования разной степени точности – линейных, полиномиальных или нейросетевых с возможностью дообучения в процессе работы.
3) ***Метод двух экстраполяторов***. Данный способ является модификацией способа б). Для граничных значений нулевого процесса строится экстраполятор. Такой же экстраполятор строится и для первого процесса, для значений в ячейках, которые должны быть заменены на значения, полученные из нулевого процесса. Данный экстраполятор прогнозирует граничные значения для области, находящейся на нулевом процессорном элементе и посылает первому процессу только те значения, которые не совпали с полученными в результате прогноза. Аналогично экстраполятор работает в первом процессе, так первый процесс принимает данные от нулевого и записывает их в те ячейки, которые экстраполированы нулевым процессом (а соответственно и первым процессом) неверно. Для данного метода количество передаваемой информации зависит от точности экстраполятора: чем выше точность, тем меньше информации требуется передавать.

При реализации аппаратной поддержки нейросетевых алгоритмов на основе графических ускорителей нейронная сеть представляется в виде набора текстурных карт (карты входа, выхода, весовых значений, данных индексации) и расчет производится на графической плате. При этом происходит распараллеливание расчета на внутренние фрагментные процессоры графической платы, что позволяет избежать накладных расходов на пересылку данных.
В данной главе представлено исследование возможности применения технологии SLI (Scalable Link Interface), позволяющей эффективно организовать обмен данными между двумя видеокартами, для распараллеливания решения задачи на несколько графических плат. В работе показана перспективность такого пути параллельной реализации алгоритмов, а также рассмотрены проблемы, возникающие при его реализации.

В третьей главе приведены описания практических задач, на которых демонстрируется работа предлагаемых алгоритмов. Показано, что разработанные алгоритмы применимы к задачам различной физической природы, в частности, к задаче определения гранулометрического состава окомкованного материала и задачам аэродинамики. Исследовано применение пакета «Нейроматематика» и разработанных теоретических алгоритмов к решению данных задач. Приведены результаты тестирования разработанных методов на реальных данных и их сравнение с существующими методами. Показано, что предлагаемые методы позволяют решать данные задачи быстрее и с более высокой точностью.

Задача определения гранулометрического состава окомкованного материала состоит в том, чтобы по информации, взятой с поверхности заданного набора, определить гранулометрический состав (грансостав) – весовое содержание шарообразных тел в каждой из 6-ти категорий, определяемых размером.

Решение данной задачи показано при определении гранулометрического состава окомкованного материала в процессе обработки руды на горно-обогатительном комплексе.

В качестве входных данных нейронной сети были использованы значения высот, полученных на срезах с поверхности окомкованного материала, либо значения яркостей, полученного изображения на аналогичных срезах.

Эксперименты проводились с двухслойными нейронными сетями с прямыми последовательными связями. Сети обучались по правилу обратного распространения ошибки. В качестве функционала оптимизации использована среднеквадратичная ошибка выходов сети. Размер нейронной сети был определен в ходе проведения экспериментов.

На примере данной задачи были исследованы предложенные методы задания начальных условий и работы с малой выборкой при обучении нейронных сетей. Для задания начальных условий использован метод обучения сети на модельных данных, где материал представляет собой идеальные шары различных размеров. Для увеличения размера обучающей выборки использован предложенный в первой главе метод генерации дополнительных примеров.

При решении аэродинамических задач исследованы проблемы прогнозирования параметров среды в случае эжекции при обтекании крыла самолета на большой скорости и задача прогнозирования ветровой обстановки в застроенных кварталах. Для решения данных задач разработаны модули программного комплекса «Нейроматематика» «Эжекция» и «Ветер».

Исходными данными для решения задач аэродинамики являются:

– описание геометрии крыла (и, возможно, фюзеляжа) или геометрия зданий для задачи прогнозирования ветровой обстановки;

– описание физических и химических свойств набегающего потока и эжектируемых струй;

– описание граничных условий в областях эжекции и граничных условий вдува набегающего потока.

Результатом расчета является распределение газодинамических параметров в счетной области для заданного момента времени.

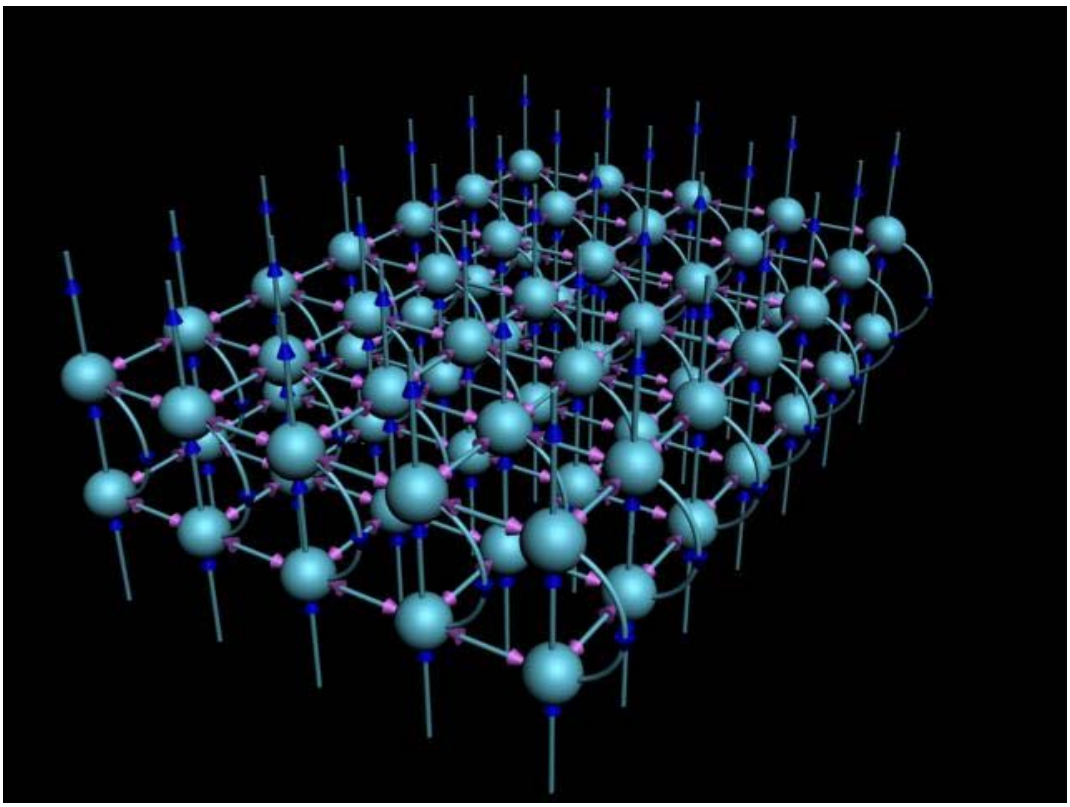

Рис. 3 Структура нейронной сети для решения двумерной задачи.

На рис. 3 представлен вариант нейронной сети для решения двумерной задачи. Для нее входом каждого нейрона является вектор $\overline{X} = \{X_{ij}\} = \{(U \quad V \quad E \quad P \quad \rho \quad \gamma)_{ij}\}$, где $U,V$ - скорости газа в данной ячейке по направлениям $x \quad u \quad y$ соответственно, $E$ - полная энергия газа в данной ячейке, $P$ - давление газа в данной ячейке, $\rho$ - плотность газа в данной ячейке, $\gamma$ - значение адиабаты газа в данной ячейке. На выходе получаются те же значения, рассчитанные через заданный промежуток времени - $\overline{Y} = \{Y_{ij}\} = \{(U \quad V \quad E \quad P \quad \rho \quad \gamma)_{ij}\}$. Для каждого нейрона первого или второго слоя производятся следующие преобразования – суммирование:

$$\sum = \begin{pmatrix} (w_{i+1}P_{i+1,j} + w_{i-1}P_{i-1,j}); & (w_{j+1}P_{i,j+1} + w_{j-1}P_{i,j-1}); \\ Speed_r \mid_x + Speed_l \mid_x; & Speed_u \mid_y + Speed_d \mid_y; \end{pmatrix},$$

Где $\begin{matrix} w_{i+1} = -w_{i-1}; \\ w_{j+1} = -w_{j-1}; \end{matrix}$ - веса, $Speed_{индекс}$ - вектор скоростей в ячейке соответствующей индексу (правой, левой, верхней или нижней ячейке). Функция активации нейрона первого слоя представлена следующей формулой:

$$f_A \begin{pmatrix} Speed_c \\ E_c \\ P_c \\ \rho_c \\ \gamma_c \\ \sum \end{pmatrix} = \begin{pmatrix} Speed_c - \frac{\alpha}{\rho_c}\sum_1 \\ E_c - \frac{\alpha}{\rho_c}(\sum_1 * \sum_2) \\ P_c \\ \rho_c \\ \gamma_c \end{pmatrix}.$$

Функция активации нейронов второго слоя выражается следующей формулой:

$$f_A\begin{pmatrix} Speed_c \\ E_c \\ P_c \\ \rho_c \\ \gamma_c \\ \sum \end{pmatrix} = \begin{pmatrix} Speed_c - \frac{\alpha}{\rho_c}\sum_1 \\ E_c - \frac{\alpha}{\rho_c}\sum_2 \\ \varepsilon\rho_c(\gamma_c - 1) \\ \rho_c \\ \gamma_c \end{pmatrix},$$

где $\varepsilon$ - удельная внутренняя энергия газа в данной ячейке.

Данная постановка легко переносится на трехмерный случай добавлением третьей компоненты скорости. Входом нейронной сети является вектор $\overline{X} = \{X_{ijk}\} = \{(U \quad V \quad W \quad E \quad P \quad \rho \quad \gamma)_{ijk}\}$, а выходом – $\overline{Y} = \{Y_{ijk}\} = \{(U \quad V \quad W \quad E \quad P \quad \rho \quad \gamma)_{ijk}\}$, где $W$ – скорость газа в направлении оси $z$.

При решении практической задачи определения гранулометрического состава окомкованного материала ошибка определения процентного содержания окатышей для крайних групп – менее 5 мм и более 18 мм в диаметре, была близка к нулю.

Для сравнения с существующими методами рассмотрен алгоритм оценки, основанный на определении размера окатыша по видимой части его площади и последующем подсчете процентного содержания окатышей каждого из классов. Предположим, что существует некий алгоритм, с помощью которого можно идеально точно определить линейный размер тела, если *X%* площади его поверхности видимы.

Существующие алгоритмы определения гранулометрического состава посредством выделения очертаний отдельных окатышей на изображении и подсчета количества окатышей каждого класса способны определить размер окатыша, начиная с 60% его видимой площади. Как можно видеть из Таблицы 1, точность такого алгоритма, определяющего грансостав по телам с видимыми 60% площади, сильно уступает нейросетевому алгоритму, результаты обработки которым приведены в Таблице 2.

*Таблица 1. Среднеквадратическая ошибка определения грансостава методом расчета линейных размеров в зависимости от процента видимой площади тела, требуемой для точного определения диаметра окатыша.*

| Категория<br><br>Требуемый процент видимой площади | +18мм | 18+16мм | 16+14мм | 14+10мм | 10+5мм | 5мм |
|---|---|---|---|---|---|---|
| 10% | 4.63 | 1.91 | 1.98 | 2.68 | 2.51 | 1.18 |
| 20% | 5.14 | 2.23 | 2.23 | 3.07 | 2.75 | 1.46 |
| 30% | 5.63 | 2.57 | 2.62 | 3.50 | 2.84 | 1.84 |
| 40% | 6.13 | 2.74 | 2.77 | 3.79 | 2.95 | 2.13 |
| 50% | 6.80 | 2.98 | 2.99 | 4.18 | 3.05 | 2.54 |
| 60% | 7.53 | 3.37 | 3.45 | 4.60 | 3.16 | 3.17 |
| 70% | 8.46 | 3.68 | 3.84 | 5.14 | 3.34 | 3.71 |
| 80% | 10.13 | 4.39 | 4.58 | 5.94 | 3.73 | 4.85 |
| 90% | 13.03 | 5.27 | 5.93 | 7.21 | 5.22 | 6.60 |
| 100% | 24.96 | 10.10 | 10.12 | 10.79 | 9.35 | 8.11 |

*Таблица 2. Среднеквадратическая ошибка аппроксимации гранулометрического состава окомкованного материала разработанным методом на тестовой выборке.*

| +18мм | 18+16мм | 16+14мм | 14+10мм | 10+5мм | 5мм |
|---|---|---|---|---|---|
| 2.36 | 3.16 | 3.19 | 2.55 | 1.98 | 1.28 |

При решении задач аэродинамики результаты расчета аэродинамических задач с помощью модулей «Ветер» и «Эжекция» полностью совпадают либо различаются не более чем на *0.01*% от результатов расчета тех же задач при помощи пакета Gas Dynamics Tool, который был принят за эталонный.

Время работы представленных модулей при расчете аэродинамических задач прогнозирования ветровой нагрузки на здание сиднейской оперы и исследования обтекания крыла самолета, которые были выбраны как тестовые, в несколько раз меньше времени решения этих задач другими методами. На рис. 4 представлены графики времени расчета задачи прогнозирования аэродинамических параметров среды для здания Сиднейской оперы.

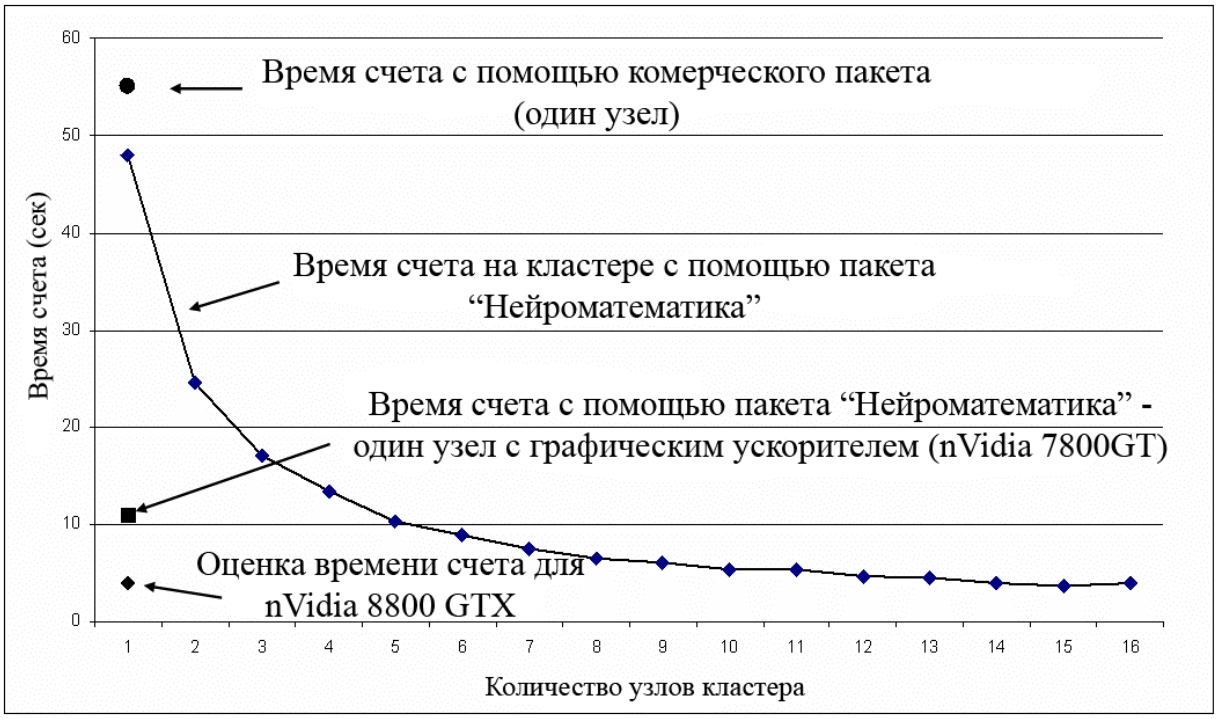


Рис. 4 Время расчета задачи прогнозирования ветровой обстановки в застроенных кварталах (здание сиднейской оперы)

В заключении приведены основные результаты работы.

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

- Предложен подход к построению программных систем нейросетевой обработки данных, позволяющий эффективно решать задачи большой размерности, в том числе, на вычислительных системах с параллельной архитектурой и графических платах.

- Разработаны новые методы и алгоритмы, позволяющие повысить точность расчета и сократить время работы нейронных сетей, в том числе:
    - Методы и алгоритмы формирования обучающей выборки в случае небольшого объема исходных данных для решения задач с помощью нейронных сетей.

    - Методы задания начальных условий нейронной сети, позволяющие повысить точность работы нейросетевого алгоритма и сократить время обучения нейронной сети. Метод основан на использовании коэффициентов соответствующего задаче оптимального фильтра в качестве начальных значений весов нейронной сети. Как развитие метода с использованием коэффициентов фильтра в качестве начальных условий предложен метод задания начальных условий для обучения нейронной сети, заключающийся в обучении нейронной сети на упрощенных модельных данных, результаты которого считаются начальными условиями нейронной сети.

    - Методы параллельной реализации нейросетевых алгоритмов для нейронных сетей прямого распространения и клеточных нейронных сетей в применении к кластерным вычислительным системам и графическим платам.

- На основе предложенного подхода, разработанных методов и алгоритмов реализован программный комплекс «Нейроматематика». Решение практических задач (определение гранулометрического состава окомкованного материала, исследования аэродинамических параметров среды и другие) с использованием разработанных средств подтвердило эффективность предложенных методов и разработанных программных средств.

## ПУБЛИКАЦИИ


1. Замараев Б.А., Коробкова С.В. Применение нейронных сетей для прогнозирования квартальной динамики ВВП. //Труды 5-ой Международной научно-технической Конференции и Выставки Цифровая обработка сигналов и ее применения, DSPA'2003 – М.: Издательское предприятие редакции журнала «Радиотехника», 2003 г., стр. 594-599
2. Коробкова С.В. Решение задачи экстраполяции временного ряда в нейросетевом логическом базисе. //Труды XLVI Научной конференции МФТИ – Издательство МФТИ, 2003 г., стр. 110-111
3. Galushkin A.I., Kazantsev P.A., Korobkova S.V. Neuromathematics: Development tendencies. //Applied and Computational Mathematics v.2, №1 – Elsevier, 2003 г., стр. 57-64
4. Кацин А.С., Куравский Л.С., Коробкова С.В. Нейросетевой алгоритм классификации в условиях малой выборки в задаче контроля усталостного разрушения тонкостенных конструкций. //Труды 6-ой Международной научно-технической Конференции и Выставки Цифровая обработка сигналов и ее применения, DSPA'2004 - М.: Издательское предприятие редакции журнала «Радиотехника», 2004 г., стр. 214-218
5. Галушкин А.И., Куравский Л.С., Кацин А.С., Коробкова С.В. Neural network classification algorithm for the small size training set situation in the task of thin-walled constructions fatigue destruction control. // Труды 11ой Международной конференции по нейронной обработке информации, ICONIP'2004 – Germany: Springer-Verlag, 2004 г., стр. 1279-1284
6. Коробкова С.В. Проблема малой выборки в решении задач распознавания образов. //Журнал Нейрокомпьютеры: разработка и применение. № 1-2 - Издательство «Радиотехника», 2005 г., стр. 34-39
7. Коробкова С.В., Томашевич Н.С. Система распознавания микробиологических объектов на изображении с помощью нейронных сетей. // Журнал Нейрокомпьютеры: разработка и применение. № 1-2 - Издательство «Радиотехника», 2005 г., стр. 56-61
8. Казанцев П.А., Лодягин А.М., Коробкова С.В. Нейросетевое распознавание гранулометрического состава шарообразных тел и его сравнение с линейными методами. //Материалы международной научно-технической конференции Интеллектуальные и многопроцессорные системы, том 2 - Таганрог: Издательство ТРТУ, 2005 г., стр. 251-257
9. Коробкова С.В. Анализ методов решения задач аппроксимации и экстраполяции функций. //Материалы международной научно-технической конференции Интеллектуальные и многопроцессорные системы, том 2 - Таганрог: Издательство ТРТУ, 2005 г., стр. 263-265
10. Казанцев П.А., Галушкин А.И., Коробкова С.В. Маршрут проектирования нейросетевого приложения. //Труды XLVIII Научной конференция МФТИ - – Издательство МФТИ, 2005 г., стр. 14-15
11. Казанцев П.А., Лодягин А.М., Коробкова С.В. Применение нейросетевого распознавания гранулометрического состава набора шарообразных тел на производстве. //Труды XLVIII Научной конференция МФТИ – Издательство МФТИ, 2005 г., стр. 32-33
12. Коробкова С.В. Проблемы эффективной аппроксимации многомерных функций с помощью нейронных сетей. //Известия ТРТУ. Тематический выпуск. Перспективные системы и задачи управления. № 3 (58) - Таганрог: Издательство ТРТУ, 2006 г., стр. 121-127
13. Коробкова С.В., Аляутдинов М.А., Воронков И.М., Галушкин А.И., Казанцев П.А., Лодягин А.М., Остапенко Г.П., Скрибцов П.В. Пакет программ для решения инженерных задач с плотными системами уравнений и сверхбольшим числом неизвестных на базе нейросетевых технологий. //Материалы VI Международной Конференции по неравновесным процессам в соплах и струях (NPNJ-2006) – Спб.: Вузовская книга, 2006 г., стр. 347-352
14. Galushkin A.I., Kazantsev P.A., Korobkova S.V., Lodyagin A.M., Panteleev S.V. Neural Network Recognition of Spherical Bodies Set Grain-Size Distribution Using Envelope of Surface.

//Труды международной совместной конференции по нейронным сетям IJCNN'2006 – IEEE Press, 2006 г., стр. 4849-4856
15. Коробкова С.В. Решение задач аэродинамики с помощью клеточных нейронных сетей. //Труды 9-ой Международной научно-технической Конференции и Выставки Цифровая обработка сигналов и ее применение - DSPA-2007 - М.: Издательское предприятие редакции журнала «Радиотехника», 2007 г., стр. 531-534
16. Аляутдинов М.А., Вронков И.М., Галушкин А.И., Казанцев П.А., Коробкова С.В., Лодягин А.М., Остапенко Г.П., Скрибцов П.В. Пакет «Нейроматематика» для решения сложных прикладных математических задач с использованием нейронных сетей. // Труды XV Международной конференции по вычислительной механике и современным прикладным системам (ВМСППС'2007) – М: Вузовская книга, типография «Знак почета» издательства МГУ, 2007 г., стр. 46-47
17. Alexander I. Galushkin, Pavel V. Skribtsov, Pavel A. Kazantsev, Svetlana V. Korobkova, Ilia M. Voronkov «Neuromathematics» - Open platform to accelerate complex applied math problems solutions using neural network algorithms. // Труды международной совместной конференции по нейронным сетям IJCNN’07 - IEEE, 2007 г., № 1314
18. Коробкова С.В., Аляутдинов М.А., Воронков И.М., Галушкин А.И., Казанцев П.А., Лодягин А.М., Остапенко Г.П., Скрибцов П.В. Решение сложных прикладных математических задач с использованием нейросетевых алгоритмов. Пакет «Нейроматематика». (часть 1) //Журнал «Современная электроника» - Издательство «СТА-ПРЕСС», №1, 2008 г. стр. 72-75
19. Коробкова С.В., Аляутдинов М.А., Воронков И.М., Галушкин А.И., Казанцев П.А., Лодягин А.М., Остапенко Г.П., Скрибцов П.В. Решение сложных прикладных математических задач с использованием нейросетевых алгоритмов. Пакет «Нейроматематика». (часть 2) //Журнал «Современная электроника» - Издательство «СТА-ПРЕСС» №2, 2008 г. стр. 70-73
20. Патент на изобретение № 2300753 Галуза Ю.П., Галушкин А.И., Коробкова С.В., Леонов А.С., Пантелеев С.В., Ряжских Л.Т. Система адаптивного нейросетевого определения гранулометрического состава частиц окомкованного и/или гранулированного материала.