

Отзыв

официального оппонента Виноградова Геннадия Павловича
на диссертационную работу Бахтина Вадима Вячеславовича
«Метод синтеза нейросетевых устройств для реализации режима fog computing»,
представленную на соискание ученой степени кандидата технических наук
по специальности 2.3.2 – Вычислительные системы и их элементы

Актуальность темы

В настоящее время возрастает значимость экономии вычислительных ресурсов, оптимизации использования существующих вычислительных систем. Расширение обработки данных вычислительной системой зачастую требуется в реальных эксплуатационных задачах на производстве и в организационной деятельности. При этом для увеличения возможностей обработки данных, как правило, необходимо приобретение и интеграция в систему дополнительных вычислительных узлов, что потребует выделения материальных ресурсов. В то же самое время, часто встречаются ситуации, когда вычислительные мощности уже имеющихся устройств задействованы не в полном объеме. Поэтому актуальной задачей сегодня является разработка и реализация методов синтеза нейросетевых устройств для реализации распределенных нейронных сетей, которые позволят сбалансировать нагрузку на уже имеющиеся устройства таким образом, чтобы она была оптимально распределена между узлами вычислительной системы, имеющими запас производительности, при этом не вызвав на них существенного ухудшения технико-экономических и эксплуатационных характеристик.

Диссертационная работа Бахтина В.В. «Метод синтеза нейросетевых устройств для реализации режима fog computing» посвящена улучшению эксплуатационно-технических показателей вычислительных систем и их элементов на основе декомпозиции искусственной нейронной сети и реализации полученных блоков в каскаде нейросетевых устройств.

Автор сосредоточил внимание на создании модели, метода, алгоритмов работы нейросетевых устройств в режиме туманных вычислений, позволяющих повысить качественные и эксплуатационные характеристики вычислительных систем и их элементов. Методы, разработанные для расчёта основных характеристик моделей, могут быть использованы для решения других задач, в которых необходимо реализовывать распределенные нейронные сети или находить оптимальные по заданным параметрам декомпозиции нейронных сетей.

Считаю, что диссертационная работа Бахтина В.В. является актуальной для практического применения в различных сферах использования вычислительных систем.

Оценка структуры и содержания работы

Диссертационная работа Бахтина В.В. состоит из введения, пяти глав,

заклучения, списка использованной литературы. Работа содержит 115 страниц основного текста, 10 таблиц, 43 рисунка и 5 приложений (38 страниц), где представлены копии актов внедрения результатов диссертационной работы, свидетельства о государственной регистрации программ для ЭВМ и листинг разработанной программы для декомпозиции нейронной сети. Список использованной литературы включает 104 наименования.

Во **введении** обоснована актуальность темы диссертационного исследования, показана степень научной разработанности проблемы, сформулированы цель и основные задачи, определены объект и предмет исследования, обоснованы методы, применяемые в представленном исследовании.

В первой главе **Анализ объекта и предмета исследования. Постановка научной задачи и частных задач исследования** проведён анализ существующих методов синтеза нейросетевых устройств, реализации распределенных нейронных сетей, сформулированы научные задачи исследования. Рассмотрены способы реализации распределенных нейронных сетей существующими методами на основе различных аппаратных архитектур и программных реализаций.

Бахтин В.В. рассмотрел классификацию методов синтеза нейросетевых устройств, условия их применения и основные подходы к реализации распределенных нейронных сетей. Сопоставительный анализ показал, что с учетом ограничений, которые выставляются в представленной работе, и необходимых для этого усовершенствований, рассматриваемая задача еще не была решена.

Проанализировано несколько методов синтеза нейросетевых устройств и реализации распределенных нейронных сетей, сформулировал постановку научной задачи исследования: метод должен обеспечивать поиск оптимального по стоимости, энергопотреблению, времени выполнения нейросетевых вычислений и времени выполнения основных вычислений, решения для декомпозиции нейронной сети на каскад распределенных вычислителей при заданных ограничениях на выбранную архитектуру вычислительной системы (число доступных вычислительных устройств), их вычислительные мощности и пропускной способности каналов связи.

Во второй главе **Разработка математической модели искусственной нейронной сети, ориентированной на туманные вычисления** проведен анализ существующих архитектур нейронных сетей и накладываемых на них ограничений по материалам отечественных и зарубежных исследователей. За основу для разработки математической модели принята многослойная нейронная сеть. Разработаны математические модели нейронных сетей с архитектурами: сети прямого распространения, сверточные и рекуррентные. Для корректной работы метода синтеза устройств требуется, чтобы к моменту декомпозиции нейронные сети уже были обучены.

Представлены входные данные модели: монолитная многослойная нейронная сеть X , результат работы которой – последовательность сигналов, и ожидаемый результат: последовательность нейронных сетей, где результат

вычисления блоков совпадает с результатом работы сети X . Дано определение декомпозиции монолитной нейронной сети. Получены общие рекуррентные формулы описания каждой из блочных нейронных сетей для различных входных параметров декомпозиции.

Сформулированы и доказаны теоремы о существовании блочной нейронной сети и об эквивалентности монолитной и блочной нейронных сетей. Адекватность математической модели подтверждается доказательствами сформулированных теорем, а также, с помощью проверки соответствия входных и выходных параметров блочной и монолитной нейронных сетей для всех возможных наборов входных параметров.

Разработанная модель позволяет учитывать требуемую загрузку вычислительных узлов при распределении блоков нейронной сети между различными устройствами. Математическая модель позволяет не только рассчитать параметры декомпозиции, но и на основе всех вариантов декомпозиции получить оптимальное разбиение для конкретного случая.

Математические выражения, определяющие параметры разработанной модели разработаны таким образом, чтобы обеспечить наиболее простую реализацию программных расчётов. Этим обеспечивается дополнительная эффективность расчётов по разработанной модели.

В третьей главе **Разработка метода синтеза устройств реализации искусственных нейронных сетей, ориентированных на туманные вычисления, и его отказоустойчивой модификации** сформулирован метод синтеза, представлены различные способы декомпозиции нейронной сети в зависимости от входных параметров, и рассмотрена модификация метода синтеза, позволяющая обеспечить отказоустойчивость.

Метод синтеза устройств реализации искусственных нейронных сетей, ориентированных на туманные вычисления, требует совмещения знания о параметрах нейронной сети и ее архитектуре с выбранными параметрами вычислительного каскада. Здесь метод формирует оптимальную по выбранным параметрам последовательность нейросетевых устройств для реализации режима fog computing размерности D .

Стоит отметить, что разработанный метод имеет также отказоустойчивую реализацию, основанную на активной отказоустойчивости, реализуемой методами диагностики и последующей адаптации. Реконфигурация, то есть перераспределение функций в рамках всего вычислительного каскада, заключается в том, что в процессе адаптации замещающий узел системы начнет выполнять дополнительно функции отказавшего узла.

Можно согласиться с выводами автора о том, что метод, разработанный в данной главе, имеет преимущества по сравнению с существующими при решении задачи декомпозиции искусственной нейронной сети и реализации полученных блоков в каскаде нейросетевых устройств.

Преимущества заключаются в возможности выбрать оптимальный вариант декомпозиции по заранее определенным параметрам, и организовать вычислительный каскад таким образом, чтобы продолжать работу даже в случае отказа части устройств в каскаде.

В четвертой главе **Разработка алгоритмов и программ декомпозиции монолитной нейронной сети и функционирования блочной нейронной сети** описана разработка и реализация алгоритма преобразования классической нейронной сети в нейронную сеть, адаптированную для туманных вычислений, алгоритма запуска и работы распределенной нейронной сети, а также проведена оценка сложности алгоритмов. Помимо этого, разработан алгоритм выбора оптимального варианта декомпозиции нейронной сети для реализации на распределенных вычислительных устройствах.

Для выполнения основных функций автором разработан новый алгоритм декомпозиции монолитной нейронной сети, который позволяет выбрать оптимальный способ декомпозиции в зависимости от входных параметров. Помимо этого, разработан модифицированный алгоритм запуска и работы распределенной нейронной сети, отличающийся возможностью диагностики работы соседних вычислительных узлов, а также реконфигурации одиночного вычислительного узла таким образом, чтобы он мог принять на себя вычисления от соседнего отказавшего узла. Следует отметить, что все представленные алгоритмы имеют полиномиальную сложность.

С целью тестирования корректности разработанных алгоритмов было разработано ПО, реализующее декомпозицию монолитной нейронной сети различными способами, и позволяющее осуществить распределенный запуск блочной НС. Разработанные программы были написаны на языке Java.

Разработанные алгоритмы действительно позволяют найти оптимальную декомпозицию монолитной нейронной сети сразу по нескольким важным для вычислительной системы параметрам, за счет чего стало возможным расширение обработки данных вычислительной системой без увеличения ее стоимости.

В пятой главе **Схемотехническое моделирование, прототипирование и оценка эффективности метода синтеза нейросетевых устройств на схемотехнических моделях и физических устройствах** дано описание схемотехнического моделирования и прототипирования каскадов нейросетевых устройств, а также внедрения основных результатов диссертационной работы в модификацию программно-аппаратного комплекса зала заседаний ЗАО «Проминформ».

Разработанные модель, метод и алгоритмы были адаптированы и внедрены в состав программно-аппаратного комплекса в виде прототипа системы биометрической идентификации.

Предложенные нововведения позволили уменьшить затраты на создание прототипа системы биометрической идентификации на 27%, а также сократить энергопотребление прототипа системы биометрической идентификации на 13%, что позволяет значительно экономить ресурсы при горизонтальном масштабировании числа рабочих мест. Важно отметить, что результаты, полученные в рамках моделирования, и практические результаты, полученные в реальной системе, не противоречат друг другу.

В **заключении** представлены основные результаты и выводы диссертации. Выводы полностью отражают основные научные достижения

автора.

Научная новизна диссертационной работы

К основным научным результатам, полученным в диссертации, следует отнести:

1. Разработана *математическая модель искусственной нейронной сети для синтеза нейросетевых устройств, ориентированных на туманные вычисления*. Она отличается от существующих тем, что с ее помощью возможно балансировать размеры декомпозированных блоков нейронной сети в зависимости от характеристик физических устройств, входящих в вычислительный каскад. Это *позволяет* учитывать требуемую загрузку вычислительных узлов при распределении блоков нейронной сети между различными устройствами.

2. Разработан *метод синтеза устройств реализации искусственных нейронных сетей, ориентированных на туманные вычисления*. Он отличается от существующих тем, что учитывает параметры: мощность устройств, оптимальный объем передаваемых между устройствами данных, возможность учета пропорциональности блоков нейронной сети по слоям или по нейронам, а также имеет возможность реализации диагностики и реконфигурации. Это *позволяет* выбрать оптимальный вариант декомпозиции нейронной сети по заранее определенным параметрам и продолжать работу даже в случае отказа или сбоя части устройств в каскаде.

3. Разработан *алгоритм декомпозиции монолитной нейронной сети на каскад блоков блочной нейронной сети, адаптированной для туманных вычислений*. Он отличается от существующих тем, что предлагает способ унификации хранения в памяти монолитной нейронной сети и блоков блочной нейронной сети. Это *позволяет* проводить многократную декомпозицию в глубину, например, если потребуется декомпозировать отдельный блок еще на несколько блоков.

4. Разработан *алгоритм выбора оптимального варианта декомпозиции нейронной сети для реализации на распределенных вычислительных устройствах*. Он отличается от существующих тем, что он реализует многокритериальную оптимизацию путем нахождения Парето-оптимальных вариантов. Это *позволяет* находить оптимальную декомпозицию монолитной нейронной сети сразу по нескольким важным для вычислительной системы параметрам.

Достоверность и обоснованность полученных результатов

Полученные в диссертационной работе результаты не противоречат теоретическим положениям, известным из научных публикаций отечественных и зарубежных исследователей, и подтверждаются результатами апробации и внедрения предложенных в диссертации модели, метода и алгоритмов в реальные вычислительные системы.

Методология и методы исследования базируются на математическом аппарате и методах дискретной математики, теории булевых функций и автоматов, теории вероятности и математической статистики, теории искусственных нейронных сетей, теории алгоритмизации.

Основные результаты изложены в 20 публикациях, из них 5 статей в ведущих рецензируемых научных изданиях; 3 публикации в трудах международных конференций, индексированных в МБЦ Scopus, 2 свидетельства о государственной регистрации программ для ЭВМ.

Результаты внедрения подтверждаются актами, представленными в приложении. Все это позволяет считать полученные результаты обоснованными и достоверными.

Автореферат в полной мере отражает основные положения, содержание и выводы диссертации. Опубликованные работы дают достаточно объемное представление о содержании диссертации.

Практическая значимость полученных автором результатов

Практическая значимость работы состоит в том, что предложенный инструментарий в виде модели, метода, алгоритмов реализован и внедрен в составе аппаратного и программного обеспечения элементов вычислительных систем, реализующих распределенные нейронные сети.

По основным показателям отмечено, что внедрение результатов работы позволило уменьшить затраты на создание прототипа системы биометрической идентификации на 27% (акт ЗАО «Проминформ»), а также сократить энергопотребление прототипа системы биометрической идентификации на 12,7% (акт ЗАО «Проминформ»). Результаты работы внедрены в учебный процесс кафедры «Автоматика и телемеханика» (акт ФГАОУ ВО «ПНИПУ»).

Замечания

1. В аналитическом обзоре представлено недостаточное количество работ авторов, исследующих проблемы синтеза нейросетевых устройств непосредственно для распределенных вычислений.

2. Для разработки математической модели автор использует различный математический аппарат, но не всегда обосновывает его выбор.

3. В качестве модели вычислительной системы рассматривается граф, но типы рассматриваемых взаимосвязей между элементами не определены. В связи этим, автор не учитывает ошибки в каналах передачи информации между элементами.

4. Автор принимает величину вычислительной мощности узла за константу. В реальной системе данная величина может зависеть от нескольких факторов, в т. ч. от технического состояния элемента в данный момент времени.

5. Произведено недостаточное сравнение результатов применения метода синтеза нейросетевых устройств с существующими аналогами.

Указанные замечания не снижают общего положительного впечатления от проделанной работы и полученных результатов.

Заключение

Диссертационная работа Бахтина Вадима Вячеславовича «Метод синтеза нейросетевых устройств для реализации режима fog computing» представляет собой завершённую научно-квалификационную работу, в которой содержится постановка и решение актуальной научно - технической проблемы - улучшения эксплуатационно-технических показателей вычислительных систем и их элементов на основе декомпозиции искусственной нейронной сети и реализации полученных блоков в каскаде нейросетевых устройств.

Диссертация выполнена на достаточно высоком научном и техническом уровне с использованием современных инструментальных средств и методик, а основные выводы достаточно обоснованы и подтверждены результатами внедрения.

Основные положения, выносимые на защиту, соответствуют пунктам паспорта научной специальности 2.3.2 «Вычислительные системы и их элементы»:

1. Математическая модель распределенной искусственной нейронной сети для синтеза нейросетевых устройств, ориентированных на туманные вычисления.
2. Метод синтеза устройств реализации искусственных нейронных сетей для работы в режиме туманных вычислений и его модификация, позволяющая обеспечить отказоустойчивость.
3. Алгоритм декомпозиции монолитной нейронной сети на каскад блоков нейронной сети, адаптированной для туманных вычислений.
4. Алгоритм выбора оптимального варианта декомпозиции нейронной сети для реализации на распределенных вычислительных устройствах.
5. Результаты апробации разработанных модели, метода и алгоритмов, реализованных в структуре аппаратного и программного обеспечения элементов вычислительных систем.

Основные положения соответствуют пункту 7 «Разработка научных методов и алгоритмов организации параллельной и распределенной обработки информации, многопроцессорных, многоядерных, многомашинных и специальных вычислительных систем» и пункту 4 «Теоретический анализ и экспериментальное исследование функционирования вычислительных систем и их элементов в нормальных и экстремальных условиях с целью улучшения их технико-экономических и эксплуатационных характеристик» паспорта научной специальности 2.3.2 «Вычислительные системы и их элементы».

Диссертационная работа соответствует требованиям п. 9 «Положения о присуждении ученых степеней» от 24 сентября 2013 г. № 842, предъявляемым к кандидатским диссертациям, а её автор Бахтин Вадим Вячеславович заслуживает присуждения ученой степени кандидата технических наук по

специальности 2.3.2 - Вычислительные системы и их элементы.

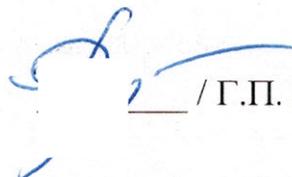
Я, Виноградов Геннадий Павлович, даю своё согласие на включение своих персональных данных в документы, связанные с работой диссертационного совета, и их дальнейшую обработку.

Официальный оппонент:

Доктор технических наук, доцент,
профессор кафедры Информатики и
прикладной математики

Тверского государственного
технического университета

«9» 06 2023 г.

 / Г.П. Виноградов /

Виноградов Геннадий Павлович

Докторская диссертация защищена по специальности 05.13.06 – Автоматизация и управление технологическими процессами и производствами (по отраслям).

ФГБОУ ВО «Тверской государственный технический университет»

170026, Тверская область, г. Тверь, наб. Аф. Никитина, д. 22,

телефон: +7 (4822) 78-91-90,

почта: wgp272ng@mail.ru


ВЕРЯЮ
член Совета
государственного
технического
университета