

УТВЕРЖДАЮ

Проректор по науке и
международной деятельности
Юго-Западного
государственного

университета,

Д.Т.Н., доцент

Д.В. Титов

05

2023 г.

Отзыв

ведущей организации ФГБОУ ВО

«Юго-Западный государственный университет»

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 диссертационной работе решаются актуальные и важные задачи реализации распределенных нейронных сетей в вычислительных системах различного назначения и их элементах. Распределенные нейронные сети улучшают обработку информации за счет возможности использования больших объемов данных и повышения скорости обработки данных. Использование таких сетей является целесообразным в различных сферах, таких как финансы, медицина, военная техника, промышленность, транспорт, игровая индустрия, реклама, кино и телевидение. Таким образом, распределенные нейронные сети являются важным инструментом в обработке данных и помогают увеличить

производительность в некоторых отраслях промышленности.

Усложнение элементов вычислительных систем, увеличение числа возможных режимов эксплуатации, степени распределенности систем в условиях санкционного давления на экономику приводят к тому, что возрастает значимость экономии вычислительных ресурсов, оптимизации использования существующих вычислительных систем.

В связи с этими факторами, диссертационная работа Бахтина В.В. «Метод синтеза нейросетевых устройств для реализации режима Fog Computing», посвященная улучшению эксплуатационно-технических характеристик вычислительных систем при помощи оптимальной декомпозиции искусственной нейронной сети и её реализации в каскаде нейросетевых устройств, является актуальной для практического применения в различных сферах использования вычислительных систем.

Оценка структуры и содержания работы

Работа включает в себя 153 страницы, из них – 115 страниц основного текста и 5 приложений объёмом 38 страниц. Работа содержит 10 таблиц, 43 рисунка и 104 наименования в списке использованной литературы. Приложения включают в себя копии актов внедрения и листинг разработанной программы декомпозиции.

Во *введении* представлено описание текущего состояния дел в области использования вычислительных систем и их элементов, основанных на распределенных нейронных сетях. Описаны основные достоинства и недостатки существующих методов синтеза нейросетевых устройств и декомпозиции нейронных сетей, осуществлена постановка цели и задачи исследования.

В *первой главе* проведён анализ объекта и предмета исследования. Анализ структуры вычислительных систем позволил определить существующие недостатки в математическом описании распределенных нейронных сетей. Выделив некоторые общие особенности, автор использует их

для создания более точной математической модели.

Рассмотрев способы синтеза нейросетевых устройств и реализации распределенных нейронных сетей Бахтин В.В. сформулировал постановку научной задачи предлагаемого исследования: метод должен обеспечивать поиск оптимального по определенным параметрам решения для декомпозиции нейронной сети на каскад распределенных вычислителей при заданных ограничениях на параметры вычислительной системы.

Во второй главе представлены результаты создания математической модели искусственной нейронной сети, ориентированной на туманные вычисления. Анализ, предшествующий описанию существующих моделей, позволил сформулировать некоторые недостатки и учесть их в разработке собственной модели. Математическое описание выходных характеристик позволяет автоматизировать расчёты по разработанным моделям. Основной идеей математической модели распределенной нейронной сети является разделение нейронной сети на блоки, исходя из параметров каскада устройств, который будет реализовывать блочную сеть, будут меняться параметры полученных блоков и связей между ними, а соответственно и нагрузка на элементы вычислительной системы. Математическая модель позволяет получать оптимальную в рамках задачи загрузку вычислительных узлов при распределении блоков нейронной сети между различными устройствами при помощи балансировки размеров декомпозированных блоков.

В третьей главе автор приводит формулировку метода синтеза устройств реализации искусственных нейронных сетей, ориентированных на туманные вычисления. Так, метод синтеза позволяет выбрать оптимальный вариант декомпозиции нейронной сети по заранее определенным параметрам и продолжать работу даже в случае отказа или сбоя части устройств в каскаде.

Метод синтеза нейросетевых устройств, разработанный Бахтиным В.В. основывается на декомпозиции монолитной нейронной сети с помощью разработанной математической модели и последующей Парето-оптимизации вариантов декомпозиции, полученных различными способами и для различных

архитектур вычислительной системы. Ограничения, которые накладываются на условия функционирования метода состоят в том, что декомпозируется многослойная нейронная сеть, принадлежащая к одной из архитектур: сети прямого распространения, сверточные и рекуррентные. Для корректной работы метода синтеза устройств требуется, чтобы к моменту декомпозиции нейронные сети уже были обучены.

В *четвертой главе* представлены алгоритмы преобразования классической нейронной сети в нейронную сеть, адаптированную для туманных вычислений, запуска и работы распределенной нейронной сети и 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. Автор разработал ПО, реализующее декомпозицию монолитной нейронной сети различными способами, и позволяющее осуществить распределенный запуск блочной нейронной сети.

В *пятой главе* автор представляет практические результаты, полученные в ходе внедрения результатов работы в реальную вычислительную систему в модификацию программно-аппаратного комплекса зала заседаний ЗАО «Проминформ». Приведено описание системы, а также описание доработок в программно – аппаратный комплекс, которые произвел автор для внедрения своих результатов.

Проведенный сравнительный анализ результатов моделирования и работы реальной системы позволяет сделать вывод об адекватности

теоретических положений, высказанных автором при разработке модели, метода и алгоритмов работы вычислительной системы.

На практике автором было получено уменьшение затрат на создание прототипа системы биометрической идентификации на 27%, а также сокращение энергопотребления прототипа системы биометрической идентификации на 12,7%.

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 следующие пункты:

1. Разработана математическая модель распределенной нейронной сети, позволяющая получать требуемую загрузку вычислительных узлов при распределении блоков нейронной сети между различными устройствами при помощи балансировки размеров декомпозированных блоков.

2. Разработан метод синтеза нейросетевых устройств для реализации туманных вычислений, позволяющий выбрать оптимальный вариант декомпозиции обученной нейронной сети по заранее определенным параметрам и продолжать работу даже в случае отказа или сбоя части устройств в каскаде.

3. Разработан алгоритм декомпозиции нейронной сети, позволяющий реализовать диагностику и адаптацию каскада вычислительных устройств в случае отказа или сбоя, а также проводить многократную декомпозицию в глубину, например, если потребуется декомпозировать отдельный блок еще на несколько блоков.

4. Разработан алгоритм поиска оптимальной декомпозиции нейронной сети, позволяющий находить Парето-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.

Результаты работы обсуждались на различных научно-технических конференциях, в том числе с международным участием, а также опубликованы в журналах, входящих в перечень, рекомендуемых ВАК (5 статей); журналах, индексируемых в базах Scopus (3 статьи). На результаты диссертационной работы получены 2 свидетельства о государственной регистрации программ для ЭВМ, а также 2 акта внедрения в реальные вычислительные системы и учебный процесс.

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

Согласно актам внедрения, результаты исследования позволили 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, полученные автором, могут быть использованы в вычислительных системах при различных условиях эксплуатации и степени распределенности элементов системы.

Замечания

1. Автор считает величину вычислительной мощности узла постоянной, однако на практике данная величина может зависеть от различных параметров и меняться с течением времени.

2. Автором недостаточно полно описано, каким образом учитываются накладные расходы на передачу и получение данных между устройствами в рамках разработанной модели распределенной нейронной сети.

3. Эффективность разработанного метода синтеза нейросетевых устройств оценена только в двух ситуациях. Не очевидно, как поведет себя данный алгоритм в других ситуациях.

4. Следовало бы расширить сравнение результатов применения метода синтеза нейросетевых устройств для реализации режима туманных вычислений с существующими методами.

5. Из текста диссертации не ясен вопрос выбора конкретных моделей аппаратных средств.

При этом следует отметить, что указанные замечания не являются определяющими при рассмотрении представленной диссертации и не влияют на ее общую положительную оценку.

Заключение

Представленная диссертация Бахтина В.В. представляет собой завершённую научно-квалификационную работу, имеющую научное и практическое значение.

Диссертационная работа соответствует требованиям п. 9 «Положения о присуждении ученых степеней», утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842, предъявляемым к кандидатским диссертациям по специальности 2.3.2 – Вычислительные системы и их элементы, а её автор Бахтин Вадим Вячеславович заслуживает присуждения ученой степени кандидата технических наук.

Отзыв на автореферат и диссертацию обсужден и утвержден на расширенном заседании кафедры «Вычислительная техника» ФГБОУ ВО «ЮЗГУ» «31» мая 2023 г. протокол № «11».

Зав. каф. Вычислительная техника,

д.т.н., доцент

____ / И. Е. Чернецкая /

Доцент каф. Вычислительная техника,

д.т.н., доцент

____ / Э.И. Ватутин /

Доцент каф. Вычислительная техника,

к.т.н., доцент

____ / В.С. Панищев /

«31» мая 2023 г.

Федеральное государственное бюджетное образовательное учреждение высшего образования «Юго-Западный государственный университет»

Адрес: 305040, Курская область, г. Курск, ул. 50 лет Октября, 94,

Телефон: +7 (4712) 50-48-20,

Email: rector@swsu.ru