

ОТЗЫВ официального оппонента
на диссертацию на соискание ученой степени
кандидата физико-математических наук
Иванова Дмитрия Александровича
на тему «Нейроморфные методы оптимизации систем искусственного
интеллекта для задач обучения с подкреплением»
по специальности 2.3.5 Математическое и программное обеспечение
вычислительных систем, комплексов и компьютерных сетей

Актуальность работы

Актуальность диссертационной работы Д.А. Иванова определяется ее ориентацией на исследование возможности применения методов упрощения нейронных сетей в задачах обучения с подкреплением. Она обусловлена необходимостью развертывания интеллектуальных агентов на устройствах с необходимостью экономии энергии, а также с ограниченными вычислительными ресурсами и памятью, например, таких как роботы, дроны и встраиваемые системы. Квантование весов и активаций позволяет значительно уменьшить объем модели и ускорить вывод, что критически важно для принятия решений в реальном времени, в то время как прунинг и другие методы сокращения устраниют избыточные связи, снижая риск переобучения и улучшая обобщающую способность. Это особенно важно в обучении с подкреплением, где агенты взаимодействуют со средой в онлайн-режиме, а эффективное использование вычислительных ресурсов напрямую влияет на скорость обучения и возможность итеративного обновления политик и функций награды. Предлагаемый автором нейроморфный подход представляет собой интересный методологически обоснованный путь преодоления ограничений традиционной фон-неймановской архитектуры для задач обучения с подкреплением.

Основное содержание работы

Первая глава диссертации содержит критический анализ проблем производительности и энергоэффективности, присущих системам ИИ на фон-неймановских вычислителях. Автор фокусируется на особенностях систем ИИ для задач обучения с подкреплением, где малый размер пакета данных усугубляет проблему доступа к памяти, и показывает недостаточность существующих методов оптимизации, что создает предпосылки для разработки новых решений. Во второй главе осуществлен переход от констатации проблем к поиску их решений через анализ биологических аналогов. Ценность данной части работы заключается не только в систематизации известных принципов работы естественного биологического интеллекта, но и в оценке их

практической применимости для создания энергоэффективных систем ИИ. В третьей главе сформулированы методы решения обозначенной проблемы. Разработанные автором алгоритмы комбинированной оптимизации (структурная разреженность + квантование; структурная + временная разреженность) направлены на непосредственное сокращение объема вычислений и числа обращений к памяти, что является ответом на выявленные в первой главе узкие места в системах ИИ для задач обучения с подкреплением. Четвертая глава содержит экспериментальную верификацию предложенных в третьей главе алгоритмов, демонстрируют возможность многократного сокращения размеров нейронных сетей без ущерба для их производительности. Полученные результаты имеют большую ценность для разработки систем, функционирующих в условиях жестких ограничений по времени и энергии.

Научная новизна, обоснованность и достоверность научных положений, выводов и рекомендаций, сформулированных в диссертации

Наиболее важными результатами, составляющими научную новизну диссертации, являются следующие:

- Проведен обзор работ и предложена собственная систематизация известных принципов работы естественного интеллекта (мозга живых существ) в контексте их адаптации для применения в системах искусственного интеллекта (ИИ), в частности, в попытке преодоления ограничений фон-неймановской вычислительной архитектуры в задачах построения ИИ систем на основе обучения с подкреплением. Данное направление на настоящий момент является актуальным и бурно развивающимся с еще неустоявшейся терминологией и понятийным аппаратом, поэтому результаты в данной области являются полезными и вызывают интерес у научного сообщества, что подтверждается большим числом ссылок на публикацию автора диссертации по этой теме.
- Адаптация и применения методов упрощения структуры нейросетей за счет обрубания связей и квантизации применительно к задаче обучения с подкреплением, где нейросети используются либо в подходе «агент-критик» для моделирования собственно процесса взаимодействия агента с окружающей средой и критика, оценивающего награду или штраф за результат взаимодействия для дообучения агента, либо в сети глубокого обучения непосредственно для моделирования Q-функции.
- Рассмотрение и обоснование автором концепции временной разреженности нейронной сети для задачи обучения с подкреплением и предложенный в работе алгоритм ее реализации.

Достоверность и обоснованность полученных результатов обеспечивается проведением большого числа вычислительных экспериментов в нескольких тестовых окружениях.

Основные результаты и положения изложены в 3 работах, опубликованных в изданиях, рекомендованных для защиты в диссертационном совете МГУ им. М.В. Ломоносова по специальности 2.3.5 Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Замечания по диссертационной работе

- В предложенном подходе обучения с подкреплением на основе модели «агент-критик» делается акцент на то, что упрощается только нейросеть агента (актера), а нейросеть критика остается исходной. Это с одной стороны позволяет сохранить высокое качество, с другой стороны не исследуется вопрос, а как упрощение модели критика может повысить вычислительную и потенциально энерго- эффективность, за которые борется автор. Это особенно актуально, т.к. в задачах обучения с подкреплением дообучение модели обычно происходит непрерывно.
- В исследованиях эффективности предложенных методов упрощения нейросетевых моделей недостаточно внимания уделено вопросам выбора исходной архитектуры, которая, собственно, и упрощается. Возможно, в случае, если исходная архитектура не обладает большой структурной избыточностью, предложенные методы квантизации и прореживания все-таки существенно ухудшают качество?
- В качестве среды для экспериментов и оценки качества предлагаемых алгоритмов и методов автор берет «игровые» окружения Atari и MuJoCo. Неочевидно, насколько результаты, полученные в данной прикладной области, переносятся на другие важные области, упомянутые в работе, например, обучение с подкреплением в задачах оптимального управления. Возможно, имело бы смысл более строго описать область и исследованные ограничения применения предложенных методов.
- В работе присутствуют опечатки и в отдельных частях некоторая небрежность изложения. В частности, начало разделов с описанием особенностей адаптации методов квантирования и прореживания в задачах обучения с подкреплением на страницах 68 и 71 повторяется дословно, что корректно с логической точки зрения, но стилистически выглядит неаккуратно.

Вместе с тем, указанные замечания к диссертации не умаляют значимости диссертационного исследования. Диссертация отвечает требованиям,

установленным Московским государственным университетом имени М.В. Ломоносова к работам подобного рода. Содержание диссертации соответствует специальности 2.3.5 Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (по физико-математическим наукам), а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова, а также оформлена согласно требованиям Положения о совете по защите диссертаций на соискание ученой степени кандидата наук Московского государственного университета имени М.В. Ломоносова.

Таким образом, соискатель Иванов Дмитрий Александрович заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 2.3.5 Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Официальный оппонент:

кандидат физико-математических наук, доцент
доцент кафедры интеллектуальных информационных технологий факультета
вычислительной математики и кибернетики
МГУ имени М.В. Ломоносова
Петровский Михаил Игоревич _____ « 11 » 12. 2025 г.

Контактные данные:

Тел.: +7 495 939-17-89, email: michael@cs.msu.su

Специальность, по которой официальным оппонентом защищена диссертация:
05.13.11, Математическое и программное обеспечение вычислительных машин,
комплексов и компьютерных сетей

Адрес места работы:

119991 ГСП-1 Москва, Ленинские горы, д.1, стр. 52, 2-й учебный корпус,
факультет ВМК
кафедра интеллектуальных информационных технологий факультета
вычислительной математики и кибернетики МГУ имени М.В. Ломоносова
Тел.: +7 495 939-17-89, email: michael@cs.msu.su