

ОТЗЫВ

официального оппонента на диссертацию

Ерещенко Алексея Владимировича

«Применение графовых нейронных сетей для анализа молекулярных структур»,

представленную на соискание учёной степени кандидата технических наук по специальности 1.2.1 «Искусственный интеллект и машинное обучение»

Актуальность темы. Медицинская химия является перспективным направлением для разработки и применения методов машинного обучения, способных успешно работать со специальными видами данных. В частности, разработка лекарственных средств обладает большим потенциалом для внедрения алгоритмов на основании машинного обучения с целью удешевления и ускорения процесса создания новых лекарств.

Растущее количество и качество доступных структурных данных органических молекул открывает возможности для разработки более точных алгоритмов с использованием искусственного интеллекта. В этой связи представляют интерес графовые нейронные сети (GNN), которые представляют данные в виде набора вершин и ребер. Данная архитектура обладает большой гибкостью возможного для применения математического аппарата, может учитывать информацию о пространственном положении объекта, кроме того, графовое представление является весьма близким к молекулярным структурам, которые часто описывают именно графом, состоящим из атомов молекулы, где ребрами обозначаются формируемые связи. Помимо этого, GNN способны работать с объектами в трехмерном пространстве, что является важным для многих задач медицинской химии.

Рассматриваемое в данной работе практическое применение разработанных алгоритмов на основе GNN является актуальным, и направлено на решение таких задач разработки лекарственных средств как поиск и классификация сайтов связывания белковых молекул, разметка пространства на поверхности белка, предсказание аффинности малых молекул к заданному молекулярному окружению, а также оценка физико-химических свойств малых молекул.

Содержание работы. Диссертационная работа состоит из введения, трех глав и заключения.

Во введении обоснованы актуальность темы исследования, научная новизна, достоверность полученных результатов, их практическая и теоретическая значимость, представлена цель работы.

В первой главе представлен разработанный алгоритм на основании GNN для поиска сайтов связывания. Дана постановка задачи, описаны используемые методы, разработанная нейросетевая архитектура, формирование обучающих данных. Продемонстрирована интеграция пространственной информации трехмерной структуры в признаковое описание, используемое нейросетью. Показано преимущество реализованного алгоритма по сравнению с другими современными методами, используемыми для решения подобного рода задач.

Во второй главе реализован алгоритм разметки пространства на поверхности белка с использованием GNN, архитектура которой является модификацией модели, представленной в первой главе. Дается постановка задачи, описывается созданный с использованием обученных нейросетей алгоритм, демонстрируется его точность по сравнению с существующими аналогами. Проводится эксперимент, в рамках которого ответы моделей нейросетей разметки пространства используются в качестве дескрипторов для обучения других нейросетевых моделей, а именно для GNN модели классификации сайтов связывания и GNN модели предсказания аффинности малой молекулы к данному молекулярному окружению. Производится валидация разработанных моделей на независимой выборке данных, их сравнение с другими существующими алгоритмами.

В третьей главе проводится исследование применения архитектуры GNN для разработки моделей предсказания свойств малых молекул. Представлен алгоритм предсказания наличия ингибирующей активности малой молекулы против 75 заданных мишеней. Дана постановка задачи, описан подход к сбору данных и обучению разработанной GNN модели, показано сравнение модели с иными моделями машинного обучения. Также представлен алгоритм предсказания растворимости малой молекулы, использующий GNN.

Научная новизна.

В данной диссертационной работе предложен новый алгоритм поиска объемных сайтов связывания на поверхности трехмерной структуры белка с применением разработанной архитектуры GNN и методов кластеризации.

Также предложена модификация архитектуры GNN поиска сайтов связывания для решения задачи разметки пространства на поверхности трехмерной белковой структуры в соответствии с экспертно разработанными биохимическими типами. Впервые продемонстрировано, что обученные GNN разметки пространства могут быть применены для создания признакового описания, которое может быть использовано для обучения других нейросетевых моделей.

Кроме того, предложен новый подход по предсказанию наличия ингибирующей активности против заданной мишени с применением разработанной GNN, обученной на ансамблях трехмерных представлений малых молекул.

Теоретическая значимость.

Теоретическая значимость диссертационной работы заключается в исследовании применения методов машинного обучения на основе GNN для работы с трехмерными белковыми структурами, в рамках решения различных задач медицинской химии. Изучена возможность применения предсказаний разработанных GNN разметки пространства на поверхности белковых структур в качестве биохимического описания пространства, которое могло бы быть использовано как признаковое представление другими нейросетевыми моделями.

Практическая значимость.

Практическая значимость диссертационной работы подтверждается заявленным внедрением разработанных алгоритмов поиска сайтов связывания, разметки пространства сайта связывания, классификации сайтов связывания, а также предсказания аффинности малых молекул во ФГУП «ВНИИА». Кроме того, в рамках данной работы реализованным алгоритмом была размечена крупная библиотека соединений, что также имеет практическую значимость для специалистов предметной области.

Достоверность и обоснованность результатов.

Достоверность и обоснованность результатов, полученных в ходе данной диссертационной работы, подтверждается их согласованностью с известными исследованиями в предметной области, сравнением с известными аналогами, экспертной валидацией.

Результаты диссертационной работы были изложены на четырех конференциях и были опубликованы в четырех публикациях, из которых три индексируются в базах Web of Science и Scopus (Q1), одна входит в список ВАК. По результатам работы также был зарегистрирован патент и было получено авторское право на код.

Замечания по диссертации.

1. В главе 2 был представлен график обучения разработанной GNN модели аннотации пространства по одному из рассматриваемых классов, с целью демонстрации особенности динамики обучения. Было бы целесообразно показать графики обучения для всех рассматриваемых классов.
2. В главе 3, в рамках решения задачи предсказания наличия ингибирующей активности, представлены результаты сравнения разработанной архитектуры 3D GNN с другими архитектурами машинного обучения,

использующими иное признаковое описание, и обученными на тех же данных. Однако, не было проведено сравнение с уже существующими и обученными моделями, решающими подобные задачи. Такой эксперимент позволил бы более полно оценить качество разработанной модели и алгоритма.

3. В главе 3, применение GNN, не использующей данные о трехмерной структуре молекулы, было рассмотрено в рамках решения задачи предсказания растворимости. Целесообразно было бы обучить модель на подобном представлении данных также и в рамках решения задачи предсказания наличия ингибирующей активности, чтобы более ясно показать преимущества использования трехмерных структурных данных.

Указанные замечания не влияют на общую положительную оценку диссертации.

Общее заключение.

Диссертационная работа Ерещенко Алексея Владимировича на тему «Применение графовых нейронных сетей для анализа молекулярных структур» является законченной научно-квалификационной работой, имеющей практическую и теоретическую значимость и посвященной актуальному научному направлению. Автореферат полно отражает содержание диссертационной работы. Работа соответствует целям и задачам исследования, а также паспорту научной специальности.

Диссертационная работа «Применение графовых нейронных сетей для анализа молекулярных структур» Ерещенко Алексея Владимировича соответствует требованиям Положения о присуждении ученых степеней, утвержденного постановлением Правительства РФ от 24 сентября 2013 года № 842, а Ерещенко Алексей Владимирович заслуживает присуждения учёной степени кандидата технических наук по специальности 1.2.1 «Искусственный интеллект и машинное обучение».

Официальный оппонент
Куравский Лев Семёнович,

доктор технических наук, профессор, Почётный работник науки и техники РФ,
Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный психолого-педагогический университет» (ФГБОУ ВО МГППУ), декан факультета информационных технологий.

16.02.2026

Докторская диссертация защищена по специальности 05.13.18 Математическое моделирование, численные методы и комплексы программ (2003).

Тел.: 7 916 400-57-81

E-mail: l.s.kuravsky@gmail.com

