



УТВЕРЖДАЮ

И.о. директора ИСП РАН
академик РАН А.И. Аветисян

« 09 » 02 2026 г.

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Федерального государственного бюджетного учреждения науки «Институт системного программирования им. В.П. Иванникова» Российской академии наук (ИСП РАН)
на диссертационную работу 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 исследования

Вычислительные методы на основе искусственного интеллекта находят все большее применение для решения ресурсоемких исследовательских и практических задач различных научных областей. Медицинская химия обладает большим потенциалом для внедрения подобных решений с целью удешевления и ускорения таких ключевых направлений, как разработка лекарственных средств.

Несмотря на растущее количество разрабатываемых вычислительных алгоритмов, как с использованием машинного обучения, так и без него, различные задачи медицинской химии остаются трудными для решения и обладают потенциалом для создания более качественных методов. Расширяющееся количество доступных структурных данных органических соединений открывает возможности для разработки более точных методов с применением машинного обучения.

Графовые нейронные сети являются перспективной архитектурой для работы с органическими структурными данными, которые зачастую рассматривают в виде графов. Этот вид нейросетей представляет данные в виде набора вершин и ребер, подходит для обработки пространственной информации, и обладает большим разнообразием возможной реализации, предоставляя пространство для применения различного математического аппарата.

В данной диссертационной работе рассматриваются различные области применения алгоритмов, основанных на графовых нейронных сетях, а именно поиск сайтов связывания белковых молекул, оценка их биохимической активности, предсказание фармацевтических свойств малых молекул и первичный анализ их возможных мишеней.

Содержание работы

Диссертационная работа состоит из введения, трех глав и заключения.

Во введении обоснованы актуальность темы исследования, научная новизна и практическая значимость полученных результатов, представлена цель исследования, изложены выносимые на защиту положения, предоставлена информация по апробации работы и публикациям по теме диссертации, описан личный вклад автора.

В первой главе рассматривается применение графовых нейронных сетей для поиска сайтов связывания белков на основании их трехмерной структуры. Представлено введение в предметную область и ее значимость, проводится обзор существующих подходов к решению данной задачи. Дана постановка задачи, описаны подход к формированию входных данных и способ описания рассматриваемых структур в виде графов. Представлено описание разработанной архитектуры графовой нейронной сети, а также двух исследуемых модификаций – модель, использующая информацию о химическом составе рассматриваемого объекта, и модель, использующая только информацию о позициях атомов. Описан подход к обучению реализованных графовых нейронных сетей и формированию обучающей и валидационной выборок, представлены показатели точности обученных моделей на валидационной выборке. Представлен разработанный алгоритм поиска объемных сайтов связывания, использующий разработанные нейросети. Приведены результаты тестирования реализованного алгоритма поиска сайтов связывания на экспертно сформированной независимой выборке данных, описываются использованные оценочные метрики. Проводится сравнение алгоритма с рядом существующих решений на данной выборке, также приводятся результаты экспертного анализа отдельных примеров.

Во второй главе рассматривается применение графовой нейронной сети для разметки сайта связывания трехмерной белковой структуры в соответствии с экспертно разработанными функциональными группами, на основании белкового окружения данного пространства. Представлено введение в предметную область и ее значимость, проведен обзор возможных сфер применения предлагаемой разметки пространства. Дана постановка задачи, описание входных данных, представлена архитектура разработанной графовой нейронной сети, являющейся модификацией нейронных сетей, реализованной в главе 1. Представлен подход к обучению графовых нейронных сетей и формированию обучающей и валидационной выборок. Описан алгоритм разметки сайта связывания трехмерной белковой структуры, созданный с применением реализованных графовых нейронных сетей. Приведены результаты тестирования разработанных нейросетей, а также сравнение точности аннотации пространства с их применением с существующими алгоритмами. Продемонстрировано применение аннотации пространства разработанными графовыми нейронными сетями для формирования признакового пространства для обучения других нейронных сетей, на примере моделей классификации сайтов связывания и модели предсказания аффинности малой молекулы к заданному окружению. Представлены результаты сравнения этих моделей с существующими аналогами. Показаны эксперименты, направленные на выявление вклада признаков, сгенерированных аннотационными графовыми нейронными сетями, в производительность обученной модели предсказания аффинности.

Третья глава посвящена изучению использования графовых нейронных сетей для анализа органических молекулярных структур, без использования информации о макромолекулярном белковом окружении. Исследование приведено на примере решения двух практических задач – предсказания возможности наличия ингибирующей активности малой молекулы против заданной мишени, а также предсказания растворимости. Дано введение в предметную область и ее значимость, представлен обзор форматов хранения данных, которые могут быть использованы для обучения моделей машинного обучения, рассмотрены подходы к сбору признакового описания органических молекулярных структур. Приведены постановки задач, описание входных данных, применяемых методов машинного обучения, описываются архитектуры подготовленных в рамках исследования моделей, подготовка обучающих и валидационных данных, методика обучения. Продемонстрированы результаты тестирования моделей, их сравнение, представлен эксперимент по применению разработанной графовой нейронной сети предсказания наличия ингибирующей активности для аннотации крупной открытой библиотеки органических структур.

Основные результаты диссертации и их научная новизна

1. Разработан комплексный алгоритм с применением графовых нейронных сетей, кластеризации и алгоритмического формирования трехмерного пространства для поиска сайтов связывания белковых молекул, создающий по результату работы объемное представление зоны связывания на поверхности молекулы. Показана эффективность алгоритма по сравнению с существующими актуальными решениями. Новизна заключается в создании оригинального алгоритма и собственной нейросетевой архитектуры.

2. Предложен алгоритм с применением графовых нейронных сетей для классификации пространства поверхности трехмерной белковой структуры с использованием экспертно разработанных функциональных групп. Показана эффективность разработанного алгоритма по сравнению с существующими актуальными решениями. Научная новизна заключается в создании оригинального алгоритма и архитектуры нейронной сети, являющейся модификацией разработанной графовой нейронной сети в рамках решения задачи поиска сайтов связывания. Также показана применимость оценок графовых нейронных сетей классификации пространства для разработки и обучения других нейросетевых моделей, на примере решения задач предсказания свойств сайтов связывания, а также оценки аффинности малых молекул относительно данной белковой молекулы.

3. Реализованы алгоритмы предсказания определенных свойств малых молекул, а именно наличие ингибирующей активности против тех или иных белковых мишеней, а также растворимость, с применением графовых нейронных сетей. Продемонстрировано преимущество разработанной графовой нейронной сети по сравнению с известными алгоритмами машинного обучения в задаче предсказания наличия ингибирующей активности против заданной мишени. Научная новизна заключается в разработанной архитектуре нейронной сети, совмещающей представление

молекулы в виде ансамбля графов ее возможных конформаций в трехмерном пространстве с ее физико-химическими дескрипторами, и в оригинальности предложенного подхода.

Соответствие паспорту научной специальности

Область исследования и содержание диссертации соответствует паспорту специальности 1.2.1 «Искусственный интеллект и машинное обучение», в частности по пунктам:

4. Разработка методов, алгоритмов и создание систем искусственного интеллекта и машинного обучения для обработки и анализа текстов на естественном языке, для изображений, речи, биомедицины и других специальных видов данных.

5. Методы и технологии поиска, приобретения и использования знаний и закономерностей, в том числе — эмпирических, в системах искусственного интеллекта. Исследования в области совместного применения методов машинного обучения и классического математического моделирования. Методы и средства использования экспертных знаний.

7. Разработка специализированного математического, алгоритмического и программного обеспечения систем искусственного интеллекта и машинного обучения. Методы и средства взаимодействия систем искусственного интеллекта с другими системами и человеком-оператором.

Достоверность полученных результатов

Достоверность результатов работы подтверждена согласованностью полученных результатов известным фактам и исследованиям в рассматриваемой предметной области, экспериментальной проверкой разработанных алгоритмов на представительных наборах данных, сравнительным анализом разработанных моделей и алгоритмов с ведущими аналогами, а также успешным внедрением реализованных моделей в платформу для разработки лекарственных молекул ФГУП «ВНИИА».

Основные положения диссертации получили апробацию на четырех конференциях и были опубликованы в четырех публикациях, из которых три индексируются в Q1 Scopus/WOS, одна входит в список ВАК. Кроме того, по результатам проведенной работы было зарегистрировано программное средство, был получен патент.

Значимость полученных результатов

Результаты, полученные в диссертации Ерещенко А.В., имеют научную и практическую значимость.

Теоретическая значимость работы состоит в анализе возможности использования графового представления трехмерных молекулярных структур и построения на его основе графовых нейронных сетей для задач медицинской химии и создания лекарств. Помимо этого, рассмотрено использование аннотаций трехмерного пространства на поверхности

белковых молекул, полученных с помощью созданных в ходе исследования графовых нейронных сетей, в роли биохимических дескрипторов, способных заменить или расширить традиционные признаки при обучении других графовых нейросетевых моделей для решения релевантных задач.

Практическая значимость работы состоит в эффективном внедрении предложенных моделей и алгоритмов в промышленную практику. В частности, в создаваемой программной платформе ФГУП «ВНИИА» интегрированы разработанные алгоритмы поиска объемных сайтов связывания белков, их аннотации в соответствии с функциональными свойствами, их классификации, а также алгоритм прогнозирования аффинности малых молекул. Дополнительно, алгоритм оценки ингибирующей активности по отношению к целевым мишеням использовался для аннотации открытого набора из 80 000 соединений без ранее известных данных об активности.

Рекомендации по использованию результатов

Разработанные нейросетевые подходы обладают значительным потенциалом в области компьютерного дизайна лекарств, в частности для решения таких задач как поиск сайтов связывания в белках без известных ингибиторов, поиск аллостерических сайтов связывания, классификация сайтов связывания, предсказание аффинности химического соединения к заданному белковому окружению, оценка селективности лекарственных молекул.

Замечания

1. В главе 2 представлена архитектура графовой нейронной сети, являющейся модификацией архитектуры, представленной в главе 1. Было бы целесообразно сравнить ее эффективность с реализованной в главе 1 нейросетью в рамках рассматриваемой в главе 2 задачи.
2. В главе 2 было бы полезно добавить визуальную диаграмму разработанной нейросетевой архитектуры, как это было сделано в главе 1.
3. В главе 3 указано, что одна из моделей сравнения, модель градиентного бустинга с применением программной библиотеки CatBoost, была протестирована с использованием трех вариантов отпечатков Моргана, 512 бит, 1024 бит, 2048 бит. В рамках данной задачи автором был выбран 512 битный вектор, так как другие варианты не дали прироста в точности, однако проведенные сравнительные эксперименты не были представлены в работе.

Указанные замечания носят рекомендательный характер и не влияют на общую положительную оценку диссертационной работы.

Заключение

Диссертационная работа Ерещенко А.В. «Применение графовых нейронных сетей для анализа молекулярных структур» представляет собой законченную научно-

квалификационную работу. Представленная работа посвящена актуальному научному направлению, полученные результаты обладают практической и теоретической значимостью, а также обладают научной новизной.

Результаты диссертации были опубликованы в трех научных журналах, индексируемых в базах Scopus и Web of Science, одном научном издании из списка ВАК, были представлены на четырех тематических конференциях. Кроме того, реализованные алгоритмы были внедрены в разрабатываемой платформе для генерации лекарственных молекул во ФГУП "ВНИИА", был получен патент, зарегистрировано программное средство. Структура и содержание работы соответствует целям и задачам исследования. Автореферат верно и полно отражает содержание диссертационной работы.

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

Диссертационная работа, автореферат и отзыв на работу рассмотрены на заседании семинара подразделения ИСП РАН «30-го центра Цифровой экосистемы научного центра мирового уровня «Цифровой биодизайн и персонализированное здравоохранения» ИСП РАН» (протокол № 1 от 2 февраля 2026 г.).

Сведения о составителе отзыва

Заведующий 30-го центра Цифровой экосистемы научного центра мирового уровня «Цифровой биодизайн и персонализированное здравоохранения» ИСП РАН,

кандидат физико-математических наук  Карпулевич Евгений Андреевич

Сведения о ведущей организации

Федеральное государственное бюджетное учреждение науки Институт системного программирования им. В.П. Иванникова Российской академии наук (ИСП РАН)

Почтовый адрес: 109004, г. Москва, ул. Александра Солженицына, д. 25

Телефон: +7(495) 912-44-25

Сайт: <https://www.ispras.ru/>

E-mail: info-isp@ispras.ru