

Обзор результатов работ ЛИИИТ ИВМиМГ СО РАН в 2023 г. по теме FWNM-2022-0005

Задачи 2023 г.

Одно из самых актуальных задач современной вычислительной и прикладной математики – создание и обоснование методов машинного обучения с приложениями к разработке методов анализа больших научных данных, проектированию и разработке платформ для сбора и хранения коллекций научных данных, разработке методов обеспечения информационной безопасности, разработке технологий защиты систем искусственного интеллекта от злонамеренных или случайных воздействий.

Отдельно выделена актуальная задача современной фундаментальной и прикладной науки – разработка фундаментальных основ цифровых двойников научных установок класса мегасайенс. Это программно-аппаратные комплексы, которые в режиме реального времени берут информацию с датчиков. На основе их обработки нейросети дают прогнозы и делают выводы для принятия решений. Наш институт вместе с коллегами из ЦКП «Сибирский кольцевой источник фотонов», ФИЦ «Институт катализа им. Г. К. Борескова СО РАН» и Конструкторско-технологического института научного приборостроения СО РАН занимается созданием цифрового двойника СКИФа.

Важнейшее фундаментальное научное направление – разработка многоуровневой распределенной информационно-вычислительной системы для хранения, обработки и анализа больших генетических данных. В этом направлении можно выделить следующие задачи: разработка методов математического моделирования сложных (больших) биологических систем, в том числе методов анализа больших графов, существенно нелинейных систем, процессов с запаздыванием и др.; разработка методов объяснимого (прозрачного) искусственного интеллекта в применении к моделированию и анализу больших генетических данных; разработка математических моделей с управлением применительно к живым системам.

Также важным направлением является создание методов научной аналитики и аналитической наукометрии.

Результаты

Начата разработка фундаментальных основ цифровых двойников научных установок класса мегасайенс. Подготовлен обзор публикаций на тему цифровых двойников крупномасштабной научной инфраструктуры. Были проанализированы цели и задачи таких цифровых двойников, а также проанализирована эффективность их использования. На основе обзора определены цели и задачи создания «Цифрового

двойника Сибирского кольцевого источника фотонов» (СКИФ). Также был проведен анализ необходимых вычислительных ресурсов и объема хранения данных.

С целью создания цифрового двойника позвоночника для моделирования его поведения после различных вмешательств была сформулирована задача автоматизации маркировки тел позвонков на рентгенограммах. Были описаны разработанные подходы к решению проблемы маркировки тел позвонков на рентгенограммах, а также описаны алгоритмы получения изображений позвоночника и изображений отдельных позвонков. Для обнаружения изображений каждого отдельного позвонка использовалась модель YOLOv5, а для решения задачи маркировки изображения каждого отдельного позвонка был создан ансамбль регрессионных моделей. После серии экспериментов ошибки разметки были оценены в 4,6% при использовании соотношения координат, полученных с помощью обученных моделей, и правильно размеченных данных. Выдвигались также теории о возможном усовершенствовании созданного программного обеспечения за счет разделения позвонков на два подкласса с целью снижения ошибки разметки до желаемых 0,8%.

Начата разработка многоуровневой распределенной информационно-вычислительной системы для хранения, обработки и анализа больших генетических данных, в том числе, в частности разработан программный комплекс для интеграции данных и оценки влияния точечных мутаций в районе TATA-box генов человека на изменение экспрессии этих генов. Синтетическая целевая оптимизация промоторов растений становится частью прогресса в постгеномном сельском хозяйстве наряду с гибридизацией культивируемых растений с дикими сородичами, а также селекцией с использованием маркеров. Поэтому здесь впервые были собраны все экспериментальные данные о влиянии мутаций проксимальных промоторов растений на экспрессию генов, которые смогли найти в PubMed. Некоторые из этих наборов данных ставят под сомнение как существование, так и уникальность искомого решения, которое могло бы однозначно оценить влияние мутации проксимального промотора на экспрессию генов при выращивании растений в различных условиях окружающей среды в ходе их развития. Это означает, что рассматриваемая обратная задача некорректна. Кроме того, были обнаружены экспериментальные данные о взаимозаменяемости *in vitro* растительных и человеческих TATA-связывающих белков, позволяющие применить регуляризацию Тихонова, что делает эту проблему корректной. В рамках этих фреймворков был создан веб-сервис Plant_SNP_TATA_Z-тестер и затем определены границы его применимости, используя те данные, которые ставят под сомнение как существование, так и уникальность искомого решения. Было подтверждено, что предсказанные с помощью

Plant_SNP_TATA_Z-тестера эффекты (мутаций проксимального промотора на экспрессию генов) статистически значимо коррелируют со всеми изученными экспериментальными данными. Наконец, было проиллюстрировано применение тестера Plant_SNP_TATA_Z для выявления сельскохозяйственно ценных мутаций в промоторах растений.

Проведено исследование аномального роста русскоязычных публикаций в материалах конференций по данным базы Скопус. Многие исследователи отмечают феномен быстрого роста количества публикаций российских исследователей в материалах конференций, индексируемых в международных базах данных. Это явление, как правило, связано с общим издательским давлением, вызванным особенностями национальной научной политики, и сопровождается, в том числе, системными нарушениями научной этики. Был представлен детальный анализ потока публикаций с российским участием типа документов конференций в базе данных Scopus и связанных с ней ресурсах. Его результаты ставят под сомнение некоторые, казалось бы, очевидные представления о низком качестве «домашних» конференций, о значимости влияния зарубежных «мусорных» конференций и других подозрительных практик, связанных с развитием этого явления.

Разработана и внедрена в эксплуатацию система мониторинга событий информационной безопасности в распределенных информационных системах. Основная задача системы – сбор и анализ событий информационной безопасности в институтах с целью выявления инцидентов информационной безопасности для их дальнейшего устранения, а также формирования наборов данных по выявленным событиям, инцидентам и исходным дампам трафика для научных исследований в области мониторинга информационной безопасности.

На базе ИВМиМГ формируется центральный узел сбора событий информационной безопасности (рисунок 6) от других научных организаций СО РАН. За счет централизации данных, собираемых с научных организаций, создается объект исследований, что позволяет сформировать научный центр исследований в области компьютерных атак и кибербезопасности.

Потенциальным технологическим заказчиком при коммерциализации результатов исследований планирует выступить лицензиат ФСТЭК и ФСБ в области технической и криптографической защиты информации - компания ООО «Системы информационной безопасности».

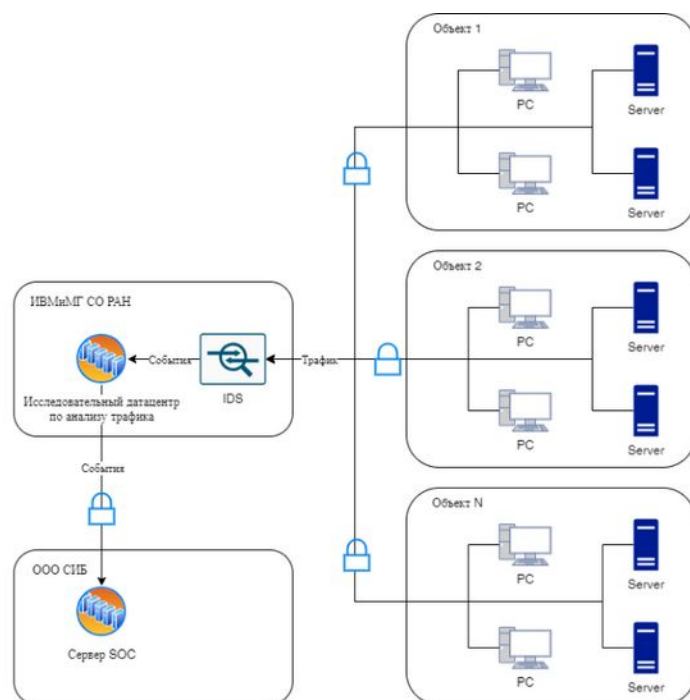


Рисунок 6 – Схема подключения ИВМиМГ СО РАН к системе мониторинга событий информационной безопасности

Список публикаций по теме за 2023 год

1 А. В. Иванов, И. А. Огнев, И. В. Никрошкин, Ю. А. Попова. Методика организации процесса мониторинга распределенных информационных систем // Безопасность цифровых технологий. Выход статьи ожидается в 4 номере в декабре 2023 года

2 С.В. Филонов, Н.Л. Подколотный, О.А. Подколотная, Н.Н. Твердохлеб, П.М. Пономаренко, Д.А. Рассказов, А.Г. Богомолов, М.П. Пономаренко Human_SNP_TATAdb – база данных о SNP, статистически достоверно изменяющих средство TATA-связывающего белка к промоторам генов человека: полногеномный анализ и варианты использования // Вавиловский журнал генетики и селекции. 2023;27(7):728-736. DOI 10.18699/VJGB-23-85.