

Проектирование и разработка ПО

Архитектура ПО реализована в следующей логике:

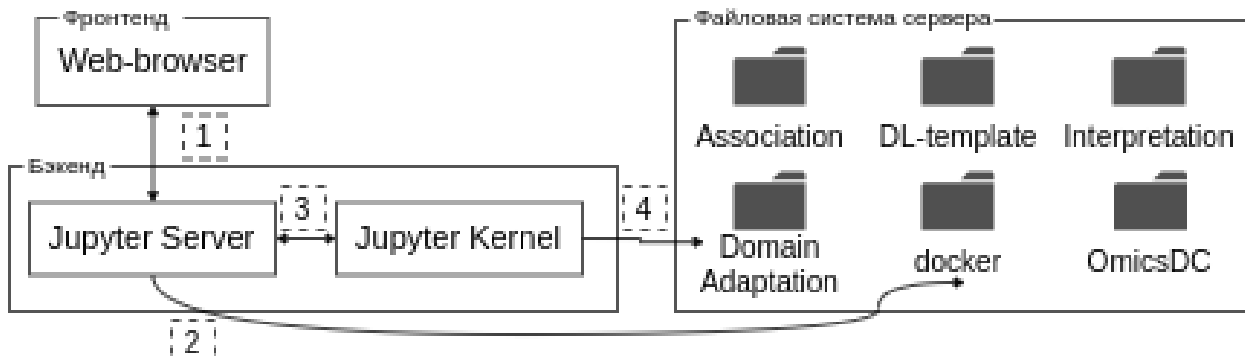


Рис.1. Взаимодействие фронтенда, бэкенда и файловой системы в рамках ПО

Направления запросов и получения ответов между компонентами ПО:

- 1) Получение бэкендом (Jupyter Server) команды пользователя, такой как запуск загрузки данных, выполнения блока кода, и т.д.
- 2) В зависимости от команды, Jupyter Server может напрямую взаимодействовать с файловой системой сервера (№2), например, для загрузки данных или чтения содержимого файлов, либо передать команду на выполнение Jupyter Kernel (№3).
- 3) Jupyter Kernel выполняет переданную команду (исполняет Python, R или bash код, меняет текущую директорию, и т.д.) и возвращает результат в Jupyter Server. Тот, в свою очередь, предоставляет его пользователю в Web-browser.

Требуемые директории в составе файловой системы:

- Директория «docker». Содержит файл Dockerfile, позволяющий контейнеризовать программное обеспечение.
- Директория «DL_template». Хранит файлы и директории, обеспечивающие функционирование основного модуля:
 - artifacts – директория, содержащая результаты работы модуля – checkpoints моделей и полногеномные предсказания моделей;

- data – директория, в которую пользователь должен загрузить исходные данные для построения моделей;
 - assets – директория, содержащая два поясняющих изображения для шаблона Template.ipynb;
 - models – директория, содержащая примеры различных типов нейронных сетей;
 - Template.ipynb – редактируемый файл-шаблон, представляющий пример работы с модулем и осуществляющий запуск работы модуля;
 - README.mb – файл с документацией модуля для пользователя.
- Директория «OmicsDC», включающая поддиректории OmicsDC и storage и файлы Examp1.ipynb и README.mb:
 - OmicsDC – директория, содержащая имплементацию функционала модуля загрузки данных;
 - storage – директория для сохранения результатов по умолчанию;
 - README.mb – файл с документацией модуля для пользователя;
 - Example.ipynb – редактируемый файл-шаблон, представляющий пример работы с модулем и осуществляющий запуск работы модуля.
- Директория «Association», включающая поддиректории Association и storage:
 - Association – директория, содержащая имплементацию функционала модуля исследования признаков на ассоциацию;
 - storage – директория для сохранения результатов по умолчанию.
- Директория «DomainAdaptation», включающая поддиректории DomainAdaptation и storage и файлы:
 - DomainAdaptation – директория, содержащая имплементацию функционала модуля трансферного обучения;

- storage – директория для сохранения результатов по умолчанию;
- conda_requirements.yml – файл с необходимыми библиотеками и их версиями для установки с помощью conda;
- pip.requirements – файл с необходимыми библиотеками и их версиями для установки с помощью pip.
- Директория «Interpretation». Содержит необходимый файл для функционирования модуля «Интерпретация обученных нейросетевых моделей»:
 - lrp.py: файл, содержащий имплементацию функционала модуля интерпретации обученных моделей на языке программирования Python.

ПО реализовано как библиотека программных и аналитических средств, направленных на решение задачи предсказания ГФЭ. При этом ПО предоставляет пользователю возможность работать как с общим веб-интерфейсом фреймворка, так и получать доступ к исходному коду отдельных программных модулей. Данное свойство обеспечено на уровне архитектуры ПО, что включает разработку веб-интерфейса на основе JupyterLab и стандартизацию исходного кода программных модулей.

ПО предоставляет возможность независимого запуска отдельных программных модулей, т.е. возможность запуска одних модулей не зависит от наличия других модулей. Данное требование обеспечивает возможность установки ПО в различной комплектации, что позволяет, с одной стороны, кастомизировать состав программных модулей для конкретного заказчика, и, с другой стороны, позволяет разграничить права различных сторон на программные модули при разработке фреймворка. Вместе с независимостью применения модулей поддерживается доступ к любому из них через пользовательский интерфейс, как единую точку входа. Дополнительно, ПО предусматривает возможность отчуждения программных модулей в виде их

исходного кода, таким образом, предоставляя возможность применения ПО без интеграционной площадки как таковой в целях эксплуатации конкретных программных модулей.

ПО предоставляет функционал рекомендаций при принятии решений, выраженный в получении пользователем результата применения фреймворка (лучшего предсказания расположения ГФЭ), а также функционала по интерпретации и объяснению полученных результатов работы нейросетевых моделей на данных пользователя.

Фронтенд ПО основан на веб-интерфейсе JupyterLab с использованием сторонних Jupyter плагинов. Взаимодействие пользователя с фронтендом обеспечено для браузера класса Chrome (версии 108 и выше).

Фронтенд, аналогично классическому JupyterLab, предоставляет пользователям следующие возможности:

- Загрузку пользовательских данных на сервер.
- Загрузку публичных омиксных данных, а также основных сборок генома с удаленных серверов.
- Выбор, конфигурацию и обучения моделей глубокого обучения на пользовательских данных.

Тестирование

ПО было протестировано на экспериментальных данных по расположению вторичных структур ДНК, в частности, участков Z-ДНК (Shin, Nam et al. 2016, Kouzine, Wojtowicz et al. 2017, Zhang, Yin et al. 2022 *). На основе экспериментальной разметки Z-ДНК и омиксных данных (Beknazarov, Jin et al. 2020 *) с помощью представляемого ПО были предсказаны участки генома, образующие Z-структуры для полных геномов человека и мыши. Эти предсказания легли в основу целого ряда исследований, результаты которых опубликованы в высокорейтинговых научных журналах (Zhang, Yin et al. 2022,

Beknazarov, Konovalov et al. 2023, Latyshev, Pavlov et al. 2023, Umerenkov, Herbert et al. 2023 *). Демонстрационных экземпляров ПО сопровождается модельными примерами, иллюстрирующими предсказания участков Z-ДНК.

Приобретение и поставка ПО

Разработка была ПО осуществлена на базе Центра искусственного интеллекта НИУ ВШЭ, который регулирует приобретение полного пакета ПО или его отдельных модулей. Приобретение ПО планируется на лицензионной основе. Поставка ПО осуществляется предоставлением дистрибутива, описанного выше.

Документирование

Документирование ПО осуществлено его разработчиками. Прилагаемая к дистрибутиву ПО документация описывает назначение, структуру, установку, эксплуатацию и техническую поддержку.

Обучение и квалификация персонала

Для эксплуатации ПО требуется наличие администратора, в функции которого входит модерация новых программных модулей, которые могут быть добавлены в ПО, а также мониторинг инфраструктуры, необходимой для функционирования ПО. Дополнительных специальных условий эксплуатации, кроме соблюдения минимальных требований к вычислительным ресурсам технических средств, на которых ПО развернуто, не требуется. Пользователь ПО должен быть уверенным пользователем ПК.

Особого обслуживания ПО не требуется.

Разработка ПО и его техническая поддержка осуществляется сотрудниками Центра искусственного интеллекта НИУ ВШЭ и Международной лаборатории биоинформатики Факультета компьютерных наук НИУ ВШЭ, обладающими необходимой для этого квалификацией. Вся инфраструктура разработки, а также разработчики ПО, осуществляющие

техническую его поддержку, размещаются по адресу: АУК "Покровский бульвар", Москва, Покровский б-р, д. 11.

Поддержка версий и доработка

Доработка ПО в случае добавления новых методов в предоставляемые модули, таких как, например, обучение нейронных сетей других типов или моделей других архитектур, предполагается соответствующим редактированием шаблона основного модуля и входит в функции администратора со стороны пользователя.

Устранение сбойных ситуаций

Устранение сбойных ситуаций при развертывании ПО на серверном оборудовании, соответствующем перечисленным в документации требованиям, осуществляется силами технической поддержки. При развертывании ПО на серверном оборудовании должно быть предусмотрено бесперебойное питание технических средств. В случае использования различных серверов (таких как сервер приложений, сервер для хранения объектов на файловой системе) должны быть проведены мероприятия по обеспечению и поддержке сетевой связности между серверами.

Время восстановления после отказа, вызванного сбоем электроснабжения, сетевого оборудования, программных средств, нефатальным сбоем сервера хранилища данных не превышает трех суток.

Время восстановления после отказа, вызванного неисправностью технических средств, не превышает времени, необходимого на устранение неисправностей или замену технических средств.

На этапах разработки программы разработчиком предусмотрена невозможность возникновения отказов в работе программы из-за некорректных действий оператора.

*

Beknazarov, N., S. Jin and M. Poptsova (2020). "Deep learning approach for predicting functional Z-DNA regions using omics data." Sci Rep **10**(1): 19134.

Beknazarov, N., D. Konovalov, A. Herbert and M. Poptsova (2023). "Z-Flipons conserved between human and mouse are associated with increased transcription initiation rates." bioRxiv: 2023.2010.2031.564984.

Kouzine, F., D. Wojtowicz, L. Baranello, A. Yamane, S. Nelson, W. Resch, K. R. Kieffer-Kwon, C. J. Benham, R. Casellas, T. M. Przytycka and D. Levens (2017). "Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome." Cell Syst **4**(3): 344-356 e347.

Latyshev, P., F. Pavlov, A. Herbert and M. Poptsova (2023). "Unsupervised domain adaptation methods for cross-species transfer of regulatory code signals." Frontiers in Big Data **6**: 1140663.

Shin, S.-I., S. Ham, J. Park, S. H. Seo, C. H. Lim, H. Jeon, J. Huh and T.-Y. Roh (2016). "Z-DNA-forming sites identified by ChIP-Seq are associated with actively transcribed regions in the human genome." DNA Research **23**(5): 477-486.

Umerenkov, D., A. Herbert, D. Konovalov, A. Danilova, N. Beknazarov, V. Kokh, A. Fedorov and M. Poptsova (2023). "Z-flipon variants reveal the many roles of Z-DNA and Z-RNA in health and disease." Life Sci Alliance **6**(7).

Zhang, T., C. Yin, A. Fedorov, L. Qiao, H. Bao, N. Beknazarov, S. Wang, A. Gautam, R. M. Williams and J. C. Crawford (2022). "ADAR1 masks the cancer immunotherapeutic promise of ZBP1-driven necroptosis." Nature: 1-9.