

Отзыв на дипломную работу
«Скрытые Марковские модели переменного порядка для
анализа данных ChIP-seq»

студентки 444-й группы Атамановой Анны Михайловны

В выпускной квалификационной работе студенткой А.М. Атамановой решалась задача анализа данных ChIP-seq.

Появление эксперимента ChIP-seq позволило биологам изучать взаимодействие молекулы ДНК с белками в клетках живого организма. Это важно, потому что белки, связанные с молекулой ДНК, могут влиять на экспрессию находящихся по соседству генов. Сама экспрессия генов обуславливает то, какую функцию ген выполняет или не выполняет в клетке. Результат исследования взаимодействия ДНК и белков позволит объяснить качественные изменения в жизни клетки, то есть, например, называть причины выключения тех или иных генов.

К сожалению, как и большинство биологических экспериментов, ChIP-seq подвержен влиянию шумов из-за неполной фиксации белков на ДНК, неточности работы антител и ошибок секвенирования. Как следствие этого, задача анализа данных ChIP-seq сводится к разметке наблюдений на две категории: сигнал и шум.

Большинство существующих методов для анализа результатов ChIP-seq основаны на скрытых Марковских моделях первого порядка. Выбор порядка модели продиктован простотой реализации, а также экспоненциальной зависимостью количества параметров модели от порядка. Минусом моделей первого порядка является то, что они не учитывают сложные зависимости в биологических данных.

Основной идеей предложенного А.М. Атамановой подхода является использование Марковской цепи переменного порядка для моделирования категорий последовательных наблюдений. Модель автоматически подбирает нужный порядок и обучает только необходимое число параметров.

А.М. Атаманова реализовала инструмент для анализа данных ChIP-seq с использованием предложенного ею подхода. В качестве языка разработки студенткой был выбран язык Python, что позволило в короткие сроки реализовать прототип модели и сосредоточиться на экспериментах. Для оптимизации производительности некоторые модули инструмента переписаны на языке Cython, позволяющем транслировать надмножество языка Python в код на языке C.

В процессе работы студентка А.М. Атаманова проявила себя как знающий и трудолюбивый специалист, способный решать сложные тех-

нические задачи. Студентка самостоятельно разобралась с предметной областью по имеющимся источникам и освоила инструментарий для исследований с использованием языка Python.

В результате сравнения студенткой было показано, что предложенная модель является более правдоподобной для данных ChIP-seq (имея в виду функцию правдоподобия), чем широко используемая скрытая Марковская модель первого порядка. Таким образом, А.М. Атаманова выполнила поставленную перед ней задачу. Также полученный результат может найти применение и при анализе других биологических экспериментов.

К недочётам работы можно отнести отсутствие публикации полученных результатов и то, что текст написан недостаточно научным языком.

Считаю, что, несмотря на указанные недостатки, дипломная работа заслуживает оценки «отлично», а студентка А.М. Атаманова — присвоения квалификации бакалавра.

Руководитель дипломной работы
д.ф.-м.н., профессор Терехов А. Н.