

ОТЗЫВ

на выпускную квалификационную работу бакалавра
обучающегося группы 17.Б04-мм Попова В.В.
“Скрытые марковские модели на графах сборки”

В настоящее время наблюдается бурный рост биоинформатики – области знания, находящейся на стыке молекулярной биологии, информатики и математики. Как следствие этого возникает огромное количество принципиально новых задач, которые до этого никто не решал.

Определение кодирующих последовательностей является одной из основных задач современной биоинформатики. Конечно же, для ее решения уже существует целый ряд методов, которые, тем не менее, не лишены недостатков.

Один из главных недостатков существующих подходов заключается в том, что предсказание кодирующих последовательностей может выполняться лишь на уже полностью восстановленной первичной последовательности ДНК (т.е. методы ожидают на вход либо полностью собранный геном, либо сборку генома достаточно хорошего качества). Однако, во многих случаях полноценная сборка генома невозможна. Такая ситуация, например, возникает при анализа (мета-)геномов природных сообществ. Как следствие этого, некоторая часть генов может быть пропущена, что несомненно может повлиять на получаемые выводы.

Полная информация о сборке как правило доступна в виде так называемого графа сборки, содержащего больше информации о собираемых геномах, нежели в несвязных фрагментах. Попову В.В. была предложена задача изучения возможности применения скрытых марковских моделей (СММ) для поиска кодирующих последовательностей генов на графах сборки геномов.

В рамках работы было разработано обобщение алгоритма FragGeneScan на ациклические ориентированные графы. Конечно, подобного рода обобщение достаточно естественно, так как графы подобного вида допускают топологическую сортировку, а следовательно возможно и обобщение стандартных подходов и методов (например, алгоритма Витерби поиска максимально правдоподобной траектории СММ) на случай входных данных подобного рода.

Разработанный метод протестирован на наборе модельных данных разного вида (три генома разного GC состава), что позволяет надеяться на корректность его работы, а также на применимость к реальным данным.

К сожалению, работа написана достаточно тезисно и непоследовательно, некоторые важные аспекты (например, детали алгоритма FragGeneScan) в принципе не приводятся. Также отсутствуют результаты, полученные на реальных данных. Складывается впечатление, что автору банально не хватило времени на завершение экспериментов и обоснование корректности полученных результатов.

В свете всего вышесказанного, считаю, что работа может быть оценена на оценку “хорошо”.



к.ф.-м.н. доцент А.И. Коробейников