Санкт-Петербургский Государственный Университет

Математико-механический факультет

Кафедра системного программирования

Отзыв научного руководителя

на выпускную квалификационную работу студента 471 группы

Тарасенко Никиты Дмитриевича

В биоинформатике задача поиска генов или их подобий в геноме живых организме остаётся актуальной задачей, поскольку хотя в общих чертах эволюционное древо на уровне отдельных видов построено, исследование происхождения отдельных генов всё ещё не завершено. Алгоритмы, связанные со скрытыми марковскими моделями (СММ), которые используются для решения этой задачи инструменте HMMer – это вычислительно ёмкие алгоритмы, и их оптимизация для центральных процессоров, т.е. вычислителях с гомогенной архитектурой если и не достигла предела возможного, то близка к этому. При этом есть возможность и наработки по использованию гетерогенных вычислительных архитектур, в частности, связки центрального процессора и графической карты.

Перед Никитой была поставлена задача оптимизировать два связанных друг с другом алгоритма из алгоритмического конвейера HMMer – Forward и Backward-Forward. Оба они используют в своей работе динамическое программирование. Они в меньшей степени рассматриваются в литературе, нежели алгоритмы SSV/MSV и Витерби, так как их удельный вес в вычислениях в исходном HMMer сравнительно невелик. Однако после оптимизаций других алгоритмов оказывается, что они становятся новым узким местом программной системы.

Сперва Никита анализирует аналогичные решения и делает вывод, что прямых аналогов его разработка не имеет, а реализации в более общих фреймворках плохо подходят к специфическим СММ, использующимся в HMMer. Далее он аккуратно описывает эти алгоритмы и предлагает модель оптимизации, в которой различные подпоследовательности генома анализируются против СММ гена различными мультипроцессорами, а в рамках одного мультипроцессора строка матрицы динамического программирования обрабатывается одновременно несколькими потоками, что отражает хорошее понимание им как зависимостей по данным в этих алгоритмах, так и принципов программирования для графических карт.

В соответствии с этой моделью была написана реализация алгоритмов Forward и Backward-Forward с применением технологии NVIDIA CUDA. С данной реализацией был проведён ряд экспериментов, показавших ускорение как со стандартной версией HMMer, так и с оптимизированной под современные потоковые команды центрального процессора.

Вместе с тем данная работа не лишена заметных недостатков. В тексте присутствует небольшое количество орфографических и пунктуационных ошибок. Поскольку ресурсы каждой графической карты ограничены, особенно в части регистровой и разделяемой памяти мультипроцессоров, а также глобальной памяти, в данной работе необходим анализ максимального размера генома и СММ, которую можно обработать алгоритмами Forward и Backward-Forward при выбранной модели оптимизации, чего не было сделано. Не сделан также и анализ причин, по которым ускорение не столь существенно, если сравнивать реализации алгоритмов на графической карте и с помощью современных потоковых инструкций центрального процессора. Описание модели оптимизации неполное, хотя при личном общении и изучении кода понятно, что именно Никита сделал. Отсутствует в тексте и ссылка на публичный репозиторий кода, где выложены результаты работы.

Также необходимо отметить, что в течение учебного года Никита неоднократно и грубо нарушал сроки календарного плана, вследствие чего реальная работа над ВКР проходила в последние месяцы перед защитой, а её объём оказался заметно меньше, чем можно было бы сделать. Но вместе с тем Никита проявил себя как находчивый и целеустремлённый инженер с широким кругозором, что можно заметить по написанному им коду.

Работа не содержит внешних заимствований. Найденные заимствования относятся к устойчивым оборотам речи и литературы.

По совокупности вышеизложенного выполненная работа заслуживает оценки «хорошо».

Научный руководитель: старший преподаватель кафедры системного программирования СПбГУ

Сартасов С.Ю. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ «\_9 »\_\_\_июня\_\_\_\_\_\_\_\_\_ 2018г.