

Секция «Биоинженерия и биоинформатика»

Гибридная MPI/OpenMP реализация алгоритма Нидлмана-Вунша

Кульдин Сергей Павлович

Студент

*Московский государственный университет имени М.В. Ломоносова, Факультет
вычислительной математики и кибернетики, Москва, Россия*

E-mail: develoooper@gmail.com

Алгоритм Нидельмана-Вунша (АНВ) – один из наиболее эффективных алгоритмов глобального выравнивания двух последовательностей. Используется в биоинформатике для выравнивания протеиновых и нуклеотидных последовательностей. Алгоритм основан на принципах *динамического программирования*. Временная и пространственная сложность алгоритма $O(n*m)$, где n и m – длины исходных последовательностей.

На данный момент не существует параллельных реализаций алгоритма Нидельмана-Вунша, позволяющих эффективно выравнивать последовательности с длинами сопоставимыми с длинами последовательностей кодирующими, например, хромосомы человека. Большинство имеющиеся реализации построено на модели общей памяти [1], что является существенным ограничением т.к. алгоритм при вычислении использует матрицу размера $n*m$ и при упомянутых длинах требует выделения нескольких терабайт памяти. В то же время, единственная найденная реализация на модели распределенной памяти [2], в отличие от предлагаемой гибридной, требует гораздо больших затрат на пересылки между вычислительными узлами и, как следствие, существенно проигрывает в производительности. Кроме того, в данной работе был проделан ряд оптимизаций по работе с памятью, позволяющих выравнивать еще более длинные последовательности за счет хранения части данных на жестком диске и использования возможностей параллельной файловой системы.

Кроме того, в первоначальном виде специфика АНВ не позволяла ему быть эффективно распараллеленным с использованием гибридной модели MPI/OpenMP, поэтому значительная часть работы заключалась в преобразовании самого алгоритма.

Приводимая в работе реализация была протестирована на таких суперкомпьютерах как BlueGene/P, "Ломоносов" и "Скиф-МГУ" и показала существенное увеличение производительности по сравнению с существовавшими реализациями, а также высокие показатели масштабируемости.

Литература

1. Martins W.S., Del Cuvillo J.B., Useche F.J. A Multithreaded Parallel Implementation Of A Dynamic Programming Algorithm For Sequence Comparison // In Pacific Symposium on Biocomputing, 2001.
2. Naveed T., Siddiqui I. S., Ahmed S. Parallel Needleman-Wunsch Algorithm for Grid // Proceedings of the PAK-US International Symposium on High Capacity Optical Networks and Enabling Technologies (HONET 2005), Islamabad, Pakistan, Dec 19 - 21, 2005.