

ДЕТЕКТИРОВАНИЕ СТРУКТУРНЫХ ИЗМЕНЕНИЙ ВРЕМЕННОГО РЯДА НА СУПЕРКОМПЬЮТЕРЕ

Никольский И.М., Фурманов К.К.¹

119991, Москва, ГСП-1, Ленинские горы, д. 1, стр. 52, факультет ВМК,

haifly@rambler.ru;

¹101000, Москва, ул. Мясницкая, д. 20, furmach@menja.net

В данной работе рассматривается задача поиска точек сдвига матожидания временного ряда большой длины. Предполагается, что длина ряда велика (от миллиона элементов), и его анализ будет производиться на суперкомпьютере. В связи с этим возникает необходимость разработки соответствующего параллельного алгоритма.

Будем считать, что данные имеют нормальное распределение с неизвестными параметрами. Предполагается только, что дисперсия постоянна, а матожидание может меняться. Анализ данных производится ретроспективно (т.е. длина временного ряда не меняется).

Подобная задача возникает во многих отраслях знания - генетике, климатологии, финансовой математике, в анализе поведения пользователей компьютерной сети.

В настоящее время существует несколько методов поиска точек сдвига среднего. Это прежде всего метод бинарной сегментации[1], а также алгоритмы, основанные на идее динамического программирования (метод PELT[2], метод оптимального разбиения и т.д.). Распараллеливание любого из перечисленных подходов представляет собой нетривиальную задачу.

В данной работе предложен легко параллелизуемый метод обнаружения сдвигов среднего. Основная его идея - разбиение на сегменты небольшой длины. Каждому вычислительному узлу выдается некоторый набор таких сегментов. Детектирование точки сдвига на отдельном сегменте производится с помощью статистики T_{\max} (предложена в [3]; в данной работе эта статистика исследована с помощью метода Монте-Карло). Предусмотрены дополнительные проверки, которые 1) уменьшают количество ложных обнаружений, 2) улучшают детектируемость вблизи границ сегментов.

Вычислительные эксперименты, проведенные на суперкомпьютере IBM Regatta (входит в суперкомпьютерный комплекс МГУ), показали хорошую масштабируемость данного алгоритма.

Литература.

1. Scott, A. J. and Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, 30(3):507-512.
2. Killick, R., Fearnhead, P., and Eckley, I. a. (2012). Optimal Detection of Change-points With a Linear Computational Cost. *Journal of the American Statistical Association*, 107(500):1590-1598
3. Sen, A. and Srivastava, M. S. (1975). On tests for detecting change in mean. *The Annals of Statistics*, 3(1):98-108.