

Аналитическое описание, сокращение избыточности данных и их обработка при сверхглубоком разложении

Дедус Ф. Ф. (мл.), Панкратов А. Н.

ffdedus@impb.ru

Пушкино, Институт математических проблем биологии РАН

Разработка специальной вычислительной технологии была вызвана необходимостью решения задач обработки реакций установки ядерно-магнитного резонанса (ЯМР), которые содержат несколько тысяч отсчетов в одном опыте. Особенностью сигналов ЯМР является их нестационарный характер, а также невозможность анализировать сигнал по частям.

Известно, что при этом применяют в основном цифровую обработку результатов ЯМР на основе быстрого дискретного преобразования Фурье. Очевидно, это вызывает существенные трудности, связанные, в первую очередь, с оценкой точности таких цифровых расчетов.

Задача ставилась следующим образом: выявить возможности обобщенного спектрально-аналитического метода (ОСАМ), связанные с обработкой подобных массивов данных большой размерности со значительным разбросом амплитуд в одном опыте.

Для этого понадобилось:

1. Исследовать возможности вычисления ортогональных функций выбранного базиса из числа классических с номерами до 5000 и выше с требуемой точностью.
2. Определить достижимую точность вычисления коэффициентов разложения с соответствующими номерами.
3. Определить возможные временные затраты на вычисления заданного количества коэффициентов разложения.

В результате выполненных исследований было установлено, что для решения указанных задач целесообразно использовать ортонормированные базисы Чебышева, Лагерра и Эрмита непрерывного аргумента. При этом точность аналитического описания подобных массивов находится в допустимых пределах.

Показано, что вполне осуществимы сверхглубокие разложения с вычислением до 10 000 членов ряда. Возможности современных ЭВМ позволяют вычислять такое количество коэффициентов разложения за несколько секунд. Достигнутые результаты по сверхглубокому и быстрому разложению сигналов могут реально увеличить возможности ОСАМ вообще и существенно расширить диапазон решаемых задач.

Существующие методы аналитического описания данных и методы сокращения избыточности развиваются по пути обработки малых объе-

мов данных для достижения быстроты сжатия с требуемой точностью. Предлагаемая же технология, наоборот, способна обеспечить требуемую точность при увеличении глубины разложения за минимальное время. На ряде тестовых задач можно определить оптимальную глубину разложения за минимальное время, а также длины обрабатываемых участков, которые соответствовали бы поставленным задачам наилучшим образом (точность описания, требуемый коэффициент сжатия и т. д.).

В случаях, когда необходимо описать сложные конфигурации и получить их адекватное математическое описание, целесообразно обеспечивать требуемую глубину разложения. Тогда непосредственно по полученным коэффициентам разложения можно вычислить (по имеющимся формулам) необходимые оценки и характеристики, в том числе статистические, геометрические и пр.

Весьма важным прикладным аспектом использования данного математического приема может явиться решение топографических задач (в том числе в составе специальных программных и программно-аппаратных комплексов специального назначения), связанных с описанием большого количества разнообразных топографических объектов (реки, дороги, и т. д.) с предельной глубиной целиком, без разбиения их на участки, а также обработка данных геодезических задач в пространстве коэффициентов разложения. Варьирование глубины разложения топографических объектов предоставляет возможность генерализации изображения, которая затруднена при кусочно-аппроксимативном характере описания. В частности, было показано, что фрактальность объектов находит естественное выражение при описании объектов со сверхглубоким разложением.

Применение рассматриваемой вычислительной технологии в задачах распознавания повторяющихся структур в геномных последовательностях живых организмов может сыграть решающую роль, особенно для обнаружения периодических данных и сверхдлинных (до 1000 символов и более) символьных последовательностей.

При разработке некоторых математических методов прогнозирования описание предыстории сигналов с соответствующей глубиной также будет способствовать получению достоверных оценок.

Для реализации эффективной обработки данных были разработаны параллельные алгоритмы вычисления коэффициентов разложения. При этом наиболее эффективные методы вычисления основаны на рекуррентных формулах.

Работа поддержана РФФИ, проекты № 06-01-08039 и № 06-07-89303.

Литература

- [1] *Панкратов А. Н., Бритенков А. К.* Обобщенный спектрально-аналитический метод: проблемы описания цифровых данных семейства ортогональных полиномов // Вестник Нижегородского университета им. Н. И. Лобачевского, серия Радиофизика. — 2004. — № 1(2).
- [2] *Дедус Ф. Ф., Махортых С. А., Устинин М. Н., Дедус А. Ф.* Обобщенный спектрально-аналитический метод обработки информационных массивов. — М.: Машиностроение, 1999. — 356 с.