О некоторых полиномиально разрешимых и NP-трудных задачах анализа и распознавания последовательностей с квазипериодической структурой

Кельманов А. В.
kelm@math.nsc.ru
Новосибирск, Институт математики им. С. Л. Соболева СО РАН

Рассматривается нетрадиционный подход к помехоустойчивому компьютерному анализу и распознаванию числовых последовательностей, сущность которого состоит в апостериорном (off-line) способе обработки последовательности в сочетании с формализацией содержательной задачи как задачи принятия решения (проверки гипотез). Изложены результаты по исследованию сложности, решению и систематизации дискретных экстремальных задач, к которым сводится реализация этого подхода в случае, когда последовательности включают квазипериодически чередующиеся информационные фрагменты (подпоследовательности), имеющие одну и ту же размерность (число членов).

Задачи, входящие в анализируемую совокупность, возникают в приложениях, связанных с обработкой массивов зашумленных структурированных данных — результатов измерения характеристик изучаемых объектов различной природы. Эти задачи типичны, в частности, для электронной разведки, дистанционного зондирования, телекоммуникации, геофизики, обработки речевых сигналов, биометрики, медицинской и технической диагностики, радиолокации, гидроакустики, криминалистики, поиска по мультимедийным базам данных и др.

В основе трех хорошо изученных традиционных подходов лежат последовательный (on-line) и апостериорный способы обработки последовательности в сочетании с формализацией содержательной задачи как задачи оценивания (оптимальной фильтрации), а также последовательный способ обработки в комбинации с формализацией задачи как задачи проверки гипотез. Эти подходы имеют глубокую историю и связаны с фундаментальными работами Колмогорова, Котельникова, Пугачева, Ширяева, Харкевича, Андерсона, Вальда, Винера, Калмана, Пэйджа, Хинкли и множества других отечественных и зарубежных исследователей. При реализации традиционных подходов проблемы комбинаторной оптимизации как правило не возникают. В противоположность этому, реализация рассматриваемого в работе подхода сопряжена с решением специфических задач комбинаторной оптимизации с целью выбора наилучшего из множества допустимых решений, мощность которого растет экспоненциально при увеличении длины последовательности.

Числовая последовательность, включающая квазипериодически чередующиеся ненулевые информационные фрагменты размерности q, определяется следующей формулой общего члена:

$$x_n = \sum_{m \in \mathbb{M}} u_{n-n_m}(m), \quad n = 0, \ldots, N-1,$$

где $u_{n-n_m}(m) = 0$, если $n - n_m \neq 0, \ldots, q-1$; $\bigl(u_0(m), \ldots, u_{q-1}(m)\bigr) \in \mathbb{R}^q$, $0 < \bigl\|(u_0(m), \ldots, u_{q-1}(m))\bigr\| < \infty$ при каждом $m \in \mathbb{M} = \{1, \ldots, M\}$, а $(n_1, \ldots, n_M) \in \Omega$, причем

$$\Omega = \bigcup_{M=M_{\min}}^{M_{\max}} \Omega_M,$$

где

$$\Omega_M = \left\{ (n_1, \ldots, n_M) \left| \begin{array}{l} 0 \leqslant n_1 \leqslant N^+ \leqslant N-q, \;\; 0 \leqslant N^- \leqslant n_M \leqslant N-q \\ 0 < q \leqslant T_{\min} \leqslant n_m - n_{m-1} \leqslant T_{\max} \leqslant N-q, \\ m = 2, \ldots, M \end{array} \right. \right\},$$

а $M_{\min}$ и $M_{\max}$ находятся из решения системы неравенств, входящих в определение множества Ω_M, в которой N^-, N^+, $T_{\min}$ и $T_{\max}$ — целые числа. Термин «квазипериодическая последовательность» (т. е. последовательность, квазипериодически изменяющая свои свойства) обусловлен спецификой ограничений, входящих в определение множества Ω_M.

Положим $U_m = \bigl(u_0(m), \ldots, u_{q-1}(m)\bigr)$ и назовем U_m, $m \in \mathbb{M}$, *информационным вектором*, последовательность его компонент — *информационной последовательностью*. Фрагмент $(x_{n_m}, \ldots, x_{n_m+q-1})$, $m \in \mathbb{M}$, последовательности x_n, $n = 0, \ldots, N-1$, совпадающий с вектором U_m, будем называть *информационным фрагментом*.

Предполагается, что вектор $X = (x_0, \ldots, x_{N-1})$, последовательность компонент которого содержит чередующиеся информационные фрагменты, недоступен для непосредственой обработки из-за вектора помехи $E = (\varepsilon_0, \ldots, \varepsilon_{N-1}) \in \Phi_{0,\sigma^2 I}$, где $\Phi_{0,\sigma^2 I}$ — нормальное распределение. Доступным для обработки считается вектор $Y = X + E$. При этом вектор X рассматривается как функция $X(n_1, \ldots, n_M, U_1, \ldots, U_M)$, совокупность аргументов которой уточняется при формулировке различных вариантов задач анализа и распознавания.

Все рассмотренные в докладе дискретные экстремальные задачи выявлены (возникают) в результате формализации содержательных задач обработки данных как задач принятия решения (о среднем $X(\cdot)$ случайного гауссовского вектора $Y \in \Phi_{X(\cdot),\sigma^2 I}$), доставляющего максимум функционалу правдоподобия. К идентичным формулировкам экстремальных задач приводит минимизация функционала $\|Y - X(\cdot)\|^2$ суммы квадратов уклонений.

Совокупность выявленных экстремальных задач включает следующие классы, объединяющие содержательно похожие задачи:

1) обнаружение в числовой последовательности повторяющегося фрагмента;
2) распознавние последовательности, включающей повторяющийся фрагмент;
3) обнаружение и идентификация фрагментов;
4) распознавние последовательности, включающей фрагменты из алфавита;
5) обнаружение фрагментов в последовательности и разбиение этой последовательности на серии идентичных фрагментов;
6) распознавние последовательности, включающей серии идентичных фрагментов;
7) обнаружение в числовой последовательности повторяющегося набора фрагментов;
8) распознавание последовательности, включающей повторяющийся набор фрагментов;
9) обнаружение и идентификация наборов фрагментов;
10) кластеризация последовательностей.

Для части экстремальных задач из анализируемой совокупности установлена полиномиальная разрешимость либо NP-трудность, обоснованы точные и приближенные алгоритмы их решения (см. [1-3] и цитированные там работы). Однако статус вычислительной сложности многих задач из этой совокупности пока не выяснен. Установление статуса комбинаторной сложности этой части задач представляется важным делом ближайшей перспективы, поскольку они являются специальными случаями задач обработки последовательностей с более сложной структурой.

Работа поддержана РФФИ, проекты № 06-01-00058 и № 07-07-00022.

Литература

[1] *Кельманов А. В.* Апостериорный подход к решению типовых задач анализа и распознавания числовых квазипериодических последовательностей: обзор результатов // ММРО-12 — Москва: МаксПресс, 2005. — С. 125–128.

[2] *Кельманов А. В.* Проблемы оптимизации в типовых задачах помехоустойчивой апостериорной обработки числовых последовательностей с квазипериодической структурой // Докл. 3-й Всеросс. конф. «Проблемы оптимизации и экономические приложения». — Омск: ОмГТУ, 2006. — С. 37–41.

[3] *Кельманов А. В.* Полиномиално разрешимые и NP-трудные варианты задачи оптимального обнаружения в числовой последовательности повторяющегося фрагмента // Докл. Всеросс. конф. «Дискретная оптимизация и исследование операций». — Владивосток-Новосибирск: ИМ СО РАН, 2007. — `http://math.nsc.ru/conference/door07/`.