Распознавание числовой квазипериодической последовательности, включающей повторяющийся набор эталонных фрагментов

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

Рассматриваемая задача дополняет список изученных полиномиально разрешимых и NP-трудных задач комбинаторной оптимизации, возникающих в рамках нетрадиционного слабо изученного подхода к помехоустойчивому анализу и распознаванию числовых последовательностей с квазипериодической структурой [1]. Сущность этого подхода состоит в апостериорном (off-line) способе обработки последовательности при формализации содержательной задачи как задачи принятия решения.

Одна из возможных содержательных трактовок задачи состоит в следующем. Источник сообщений через канал связи с помехами передает информацию о некотором физическом объекте в виде повторяющегося упорядоченного эталонного набора импульсов различной формы, но одинаковой длительности. Имеется конечная совокупность отличающихся объектов. Каждому объекту этой совокупности соответствует единственный эталонный набор импульсов, а совокупности объектов — множество (словарь) эталонных наборов, размерности которых в общем случае различны. Словарь известен. На приёмную сторону через канал передачи поступает квазипериодическая последовательность импульсов, искаженная аддитивным шумом. Моменты времени появления импульсов в принятой (наблюдаемой) зашумленной последовательности и общее число переданных импульсов неизвестны. Требуется установить, какому объекту из совокупности соответствует принятая импульсная последовательность. Иными словами, требуется идентифицировать (распознать) принятую последовательность как последовательность, порожденную неизвестным эталонным набором из словаря.

Числовая квазипериодическая последовательность, в составе которой имеется повторяющийся упорядоченный набор из L ненулевых фрагментов размерности q, определяется следующей формулой общего члена:

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

где

$$
\begin{aligned}
&l(m,L) = (m-1) \bmod L + 1,\\
&u_{n-n_m}\bigl(l(m,L)\bigr) = 0, \text{ если } n - n_m \neq 0,\dots,q-1,\\
&\bigl(u_0(l(m,L)),\dots,u_{q-1}(l(m,L))\bigr) \in \mathbb{R}^q,\\
&0 < \bigl\|\bigl(u_0(l(m,L)),\dots,u_{q-1}(l(m,L))\bigr)\bigr\| < \infty
\end{aligned}
$$

при каждом $m \in \mathbb{M} = \{1,\dots,M\}$, причем $M \geqslant L$,

$$
(n_1,\dots,n_M) \in \Omega = \bigcup_{M=M_{\min}}^{M_{\max}} \Omega_M,
$$

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

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.

Заметим, что $l(m,L) \in \{1,\dots,L\}$ для любого $m \in \mathbb{M}$. Положим $U_j = \bigl(u_0(j),\dots,u_{q-1}(j)\bigr)$, и назовем U_j, $j = 1,\dots,L$, *эталонным вектором*, а $(U_1,\dots,U_L)$ — *эталонным набором*. Фрагмент $(x_{n_m},\dots,x_{n_m+q-1})$, $m \in \mathbb{M}$, последовательности x_n, $n = 0,\dots,N-1$, совпадающий с вектором U_j, $j = 1,\dots,L$, будем называть *эталонным фрагментом*.

Допустим, что $(U_1,\dots,U_L) \in \mathbb{W}$, где $\mathbb{W}$ — конечное подмножество (словарь, $|\mathbb{W}| = K$) множества всевозможных наборов, составленных из векторов размерности q:

$$
\mathbb{W} \subset \left\{ (U^{(1)},\dots,U^{(i)}) \;\middle|\; \begin{array}{l} U^{(k)} \in \mathbb{R}^q,\ 0 < \|U^{(k)}\| < \infty,\ k = 1,\dots,i, \\ 1 \leqslant i \leqslant M_{\max} \end{array} \right\}.
$$

Предполагается, что вектор $X = (x_0,\dots,x_{N-1})$, последовательность компонент которого включает повторяющиеся упорядоченые совокупности фрагментов, совпадающие с набором $(U_1,\dots,U_L)$ эталонных векторов, недоступен для непосредственой обработки из-за вектора помехи $E = (\varepsilon_0,\dots,\varepsilon_{N-1}) \in \Phi_{0,\sigma^2 I}$, где $\Phi_{0,\sigma^2 I}$ — нормальное распределение. Доступным для обработки считается вектор $Y = X + E$. При этом вектор X рассматривается как функция $X(n_1,\dots,n_M,U_1,\dots,U_L)$.

Задача распознавания состоит в следующем.

Дано: вектор Y и множество $\mathbb{W}$, включающее K эталонных наборов; если значения N^-, N^+, $T_{\min}$ и $T_{\max}$ неизвестны, то полагаем $N^- = 0$, $T_{\min} = q$, $T_{\max} = N^+ = N - q$.

Найти: набор $(U_1, \dots, U_L) \in \mathbb{W}$, доставляющий максимум функционалу правдоподобия $\mathcal{L}\big(X(n_1, \dots, n_M, U_1, \dots, U_L) \mid Y\big)$.

В работе показано, что максимально правдоподобное распознавание числовой квазипериодической последовательности (искаженной аддитивной гауссовской некоррелированной помехой), включающей повторяющийся набор эталонных фрагментов, в случае, когда суммарное число M фрагментов в последовательности неизвестно, сводится к задаче отыскания наборов $(n_1, \dots, n_M) \in \Omega$ и $(U_1, \dots, U_L) \in \mathbb{W}$ таких, что

$$\sum_{m \in \mathbb{M}} \Big\{ 2\big(Y_{n_m}, U_{l(m,L)}\big) - \big\|U_{l(m,L)}\big\|^2 \Big\} \to \max,$$

где $(\cdot,\cdot)$ — скалярное произведение векторов, а $Y_n = (y_n, \dots, y_{n+q-1})$, $n = 0, \dots, N - q + 1$. Обоснован точный эффективный алгоритм решения этой задачи, имеющий временную сложность

$$O\big(L_{\max} K (T_{\max} - T_{\min} + 1)(N - q + 1)\big) = O(KN^3),$$

где $L_{\max}$ — максимальная размерность эталонного набора в словаре $\mathbb{W}$. Этот алгоритм является ядром алгоритма распознавания, устойчивого к помехам.

Полиномиальная разрешимость близких по своей сути задач установлена в [2–4]. В [2–3] обоснованы точные алгоритмы, обеспечивающие решение задач помехоустойчивого обнаружения повторяющегося набора эталонных фрагментов размерности L в случаях, когда суммарное число M фрагментов в последовательности известно и неизвестно. Алгоритмы имеют временную сложность, соответственно,

$$O\big(M(T_{\max} - T_{\min} + 1)(N - q + 1)\big) = O(MN^2);$$
$$O\big(\min\{L, M_{\max}\}(T_{\max} - T_{\min} + 1)(N - q + 1)\big) = O(N^3).$$

В [4] обоснован алгоритм, гарантирующий максимально правдоподобное распознавание числовой последовательности, включающей повторяющийся набор эталонных фрагментов в случае, когда суммарное число фрагментов в последовательности известно. Этот алгоритм имеет временную сложность

$$O\big(MK(T_{\max} - T_{\min} + 1)(N - q + 1)\big) = O(MKN^2).$$

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

Литература

[1] *Кельманов А. В.* О некоторых полиномиально разрешимых и NP-трудных задачах анализа и распознавания последовательностей с квазипериодической структурой // ММРО-13 (в настоящем сборнике). — 2007. — С. **??**–**??**.

[2] *Кельманов А. В., Михайлова Л. В., Хамидуллин С. А.* Апостериорное обнаружение в квазипериодической последовательности повторяющегося набора эталонных фрагментов // ЖВМиМФ. (в печати).

[3] *Кельманов А. В., Михайлова Л. В., Хамидуллин С. А.* Оптимальное обнаружение в квазипериодической последовательности повторяющегося набора эталонных фрагментов // Докл. Всеросс. конф. «Дискретная оптимизация и исследование операций». — Владивосток-Новосибирск: Ин-т математики СО РАН, 2007. — `http://math.nsc.ru/conference/door07/`.

[4] *Кельманов А. В., Михайлова Л. В., Хамидуллин С. А.* Задача распознавания квазипериодической последовательности, включающей повторяющийся набор эталонных фрагментов // Докл. Всеросс. конф. «Дискретная оптимизация и исследование операций». — Владивосток-Новосибирск: Ин-т математики СО РАН, 2007. — `http://math.nsc.ru/conference/door07/`.