ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Июль—август 2009. Том 16, № 4. С. 31–46

УДК 519.2+621.391

ОБ ОДНОЙ ЗАДАЧЕ ПОИСКА УПОРЯДОЧЕННЫХ НАБОРОВ ФРАГМЕНТОВ В ЧИСЛОВОЙ ПОСЛЕДОВАТЕЛЬНОСТИ *)

А. В. Кельманов, Л. В. Михайлова, С. А. Хамидуллин

Аннотация. Рассматривается дискретная экстремальная задача, к которой сводится один из вариантов проблемы помехоустойчивого off-line обнаружения в числовой последовательности повторяющегося упорядоченного набора фрагментов. Анализируется вариант проблемы, в котором фрагменты из искомых наборов в отсутствие помехи совпадают с элементами из заданного упорядоченного эталонного набора векторов. Обоснован новый точный полиномиальный алгоритм решения редуцированной задачи, гарантирующий оптимальность решения по критерию минимума суммы квадратов уклонений, а также по критерию максимума правдоподобия в случае, когда помеха аддитивна и является гауссовской последовательностью независимых одинаково распределённых случайных величин. Трудоёмкость предложенного алгоритма меньше, чем у известного аналога.

Ключевые слова: дискретная экстремальная задача, числовая последовательность, упорядоченный набор фрагментов, off-line алгоритм.

Введение

Объект исследования работы — проблемы оптимизации в задачах анализа данных и распознавания образов. Предмет исследования — дискретная экстремальная задача, к которой сводится один из вариантов проблемы off-line обнаружения в зашумлённой числовой последовательности упорядоченных наборов фрагментов в предположении, что скрытые в шуме фрагменты из искомых наборов совпадают с компонентами заданного упорядоченного эталонного набора векторов. Цель работы —

*) Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проекты 09–01–00032 и 07–07–00022) и АВЦП Рособразования (проект 2.1.1/3235).

обоснование нового, более эффективного алгоритма решения этой задачи. В работе дано развернутое доказательство результатов, которые в краткой форме были представлены в [1].

Одна из возможных содержательных проблем, которая приводит к рассматриваемой математической задаче, состоит в следующем. Источник сообщений передаёт информацию об активном состоянии физического объекта в виде эталонного набора импульсов, имеющих одну и ту же известную длительность, но различную известную форму. Каждому импульсу соответствует некоторое промежуточное активное состояние объекта. Порядок импульсов в наборе фиксирован и неизменен. Пассивному состоянию соответствует отсутствие каких-либо импульсов. На приёмную сторону через канал передачи поступает последовательность квазипериодически чередующихся импульсов, искажённая аддитивным шумом. Термин «квазипериодически» означает, что интервал между двумя последовательными импульсами не одинаков, а лишь ограничен сверху и снизу некоторыми константами. Моменты времени появления импульсов в принятой (наблюдаемой) зашумлённой последовательности неизвестны. Требуется обнаружить упорядоченные наборы импульсов в наблюдаемой последовательности, т. е. определить моменты времени, когда объект находился в активном состоянии. Ситуации, в которых возникает эта задача, характерны, в частности, для электронной разведки, геофизики, гидроакустики, телекоммуникации, распознавания речевых сигналов и других приложений (см. [2, 3] и цитированные там работы).

1. Постановка задачи

В [2, 3] приведена математическая модель сформулированной содержательной проблемы и показано, что в рамках этой модели решение задачи состоит в помехоустойчивом off-line обнаружении в числовой последовательности таких упорядоченных наборов квазипериодических фрагментов, что элементы этих наборов совпадают с компонентами из упорядоченного эталонного набора векторов. Там же установлено, что поиск решения как по критерию минимума суммы квадратов уклонений, так и по критерию максимального правдоподобия (в случае, когда помеха аддитивна и является гауссовской последовательностью независимых одинаково распределённых случайных величин) сводится к решению следующей дискретной экстремальной задачи.

Задача SRTFS (Searching for Recurring Tuples of Fragments in a Sequence) — поиск упорядоченных наборов фрагментов в числовой последовательности, «похожих» на эталонный набор векторов по критерию минимума суммы квадратов уклонений.

Дано: числовая последовательность $y_0, \ldots, y_{N-1}$, эталонный набор $(U_1, \ldots, U_L)$ векторов из $\mathbb{R}^q$ и целые числа N^-, N^+, $T_{\max}$ и $T_{\min}$ (в общем случае эти числа не являются частью входа задачи). Найти: целочисленный набор $(n_1, \ldots, n_M)$, доставляющий максимум целевой функции

$$G(n_1, \ldots, n_M) = \sum_{m=1}^{M} g_{l(m|L)}(n_m), \tag{1}$$

где

$$l(m|L) = (m-1) \bmod L + 1, \tag{2}$$

$$g_i(n) = 2\langle Y_n, U_i \rangle - \|U_i\|^2, \quad n \in \{0, \ldots, N-q\}, \quad i \in \{1, \ldots, L\}, \tag{3}$$

$Y_n = (y_n, \ldots, y_{n+q-1})$ — вектор-фрагмент последовательности, $\langle \cdot, \cdot \rangle$ — скалярное произведение, а $\| \cdot \|$ — норма вектора, при следующих ограничениях

$$0 \leqslant n_1 \leqslant N^+ \leqslant N-q, \quad 0 \leqslant N^- \leqslant n_M \leqslant N-q, \tag{4}$$

$$1 \leqslant q \leqslant T_{\min} \leqslant n_m - n_{m-1} \leqslant T_{\max} \leqslant N-q, \quad m = 2, \ldots, M. \tag{5}$$

Замечание. В общем случае этой задачи, когда значения N^-, N^+, $T_{\max}$ и $T_{\min}$ не являются частью входа, ограничения (4) и (5) имеют вид

$$0 \leqslant n_1 \leqslant N-q;\ 0 \leqslant n_M \leqslant N-q;\ 1 \leqslant q \leqslant n_m - n_{m-1} \leqslant N-q,\ m = 2, \ldots, M.$$

Поэтому во всех приведённых ниже формулах для общего случая следует положить $N^- = 0$, $T_{\min} = q$ и $T_{\max} = N^+ = N - q$.

Заметим, что из (4) и (5) следует двойное неравенство

$$M^- \leqslant M \leqslant M^+, \tag{6}$$

где [2]

$$\begin{aligned} M^- &= \begin{cases} \lfloor (N^- - N^+ + T_{\max} - 1)/T_{\max} \rfloor + 1, & N^- > N^+, \\ 1, & N^- \leqslant N^+, \end{cases} \\ M^+ &= \lfloor (N-q)/T_{\min} \rfloor + 1. \end{aligned} \tag{7}$$

Фрагмент в задаче SRTFS соответствует импульсу из содержательной задачи, сформулированной во введении, а набор векторов $(U_1, \ldots, U_L)$ —

эталонному набору импульсов. Переменная M обозначает число фрагментов в последовательности $y_0, \ldots, y_{N-1}$, а M^- и M^+ — границы снизу и сверху для M. Элементы набора $(n_1, \ldots, n_M)$ суть начальные номера искомых фрагментов. Величины N^- и N^+ задают краевые условия на начальные номера первого и последнего фрагментов, а $T_{\min}$ и $T_{\max}$ — минимальный и максимальный интервалы между двумя последовательными фрагментами. Ограничение $T_{\min} \leqslant n_m - n_{m-1} \leqslant T_{\max} \leqslant N - q$, $m = 2, \ldots, M$, называется условием *квазипериодичности* фрагментов, а неравенство $q \leqslant T_{\min}$ — условием, запрещающим перекрытие фрагментов. Формула (2) определяет очередность поиска фрагментов и повторяемость наборов фрагментов в последовательности. Функция (3) является редуцированной [2, 3] мерой «сходства» между вектором из эталонного набора и фрагментом последовательности.

Возможны два варианта задачи SRTFS: 1) число M является частью входа задачи (т. е. известно); 2) это число не является частью входа (неизвестно и является оптимизируемой величиной). В [2] установлено, что оба варианта задачи SRTFS разрешимы за полиномиальное время. В этой работе для первого варианта задачи обоснован алгоритм, имеющий временну́ю сложность $O(M(T_{\max} - T_{\min} + q)(N - q + 1))$ в случае, когда числа N^-, N^+, $T_{\min}$ и $T_{\max}$ являются частью входа задачи. В общем случае трудоёмкость этого алгоритма равна $O(MN^2)$. Этот же алгоритм, реализующий схему динамического программирования, можно использовать как базовый в полиномиальной процедуре отыскания оптимального решения варианта задачи SRTFS, в котором число M является оптимизируемой величиной [2]. Для этого достаточно найти решения задачи SRTFS для каждого допустимого значения $M \in \{M^-, M^- + 1, \ldots, M^+\}$ и выбрать из этих решений наилучшее. Как показано в [2], временна́я сложность такой процедуры равна $O((M^+)^2(T_{\max} - T_{\min} + q)(N - q + 1))$ в частном случае и $O((M^+)^2 N^2)$ в общем случае. Более эффективный (менее трудоёмкий) алгоритм решения варианта задачи SRTFS с оптимизируемой величиной M обоснован в работе [3]. Трудоёмкость этого алгоритма равна $O(M^+(T_{\max} - T_{\min} + q)(N - q + 1))$ в частном случае и $O(M^+ N^2)$ в общем случае. Ниже обоснован ещё менее трудоёмкий алгоритм. При $M^+ > L$ его временна́я сложность оказывается в M^+/L раз меньше, чем у алгоритма, предложенного в [3], что является значимым для ряда распространённых приложений, в которых $M^+ \gg L$, в частности, в упомянутых во введении приложениях, когда L сравнимо с $\lg(M^+)$.

2. Алгоритм решения задачи

Для каждого $M \in [M^-, M^+]$, используя ограничения (4)–(6), определим множество

$$\begin{aligned}\Omega_M &= \Omega_M(N, N^-, N^+, T_{\min}, T_{\max}, q)\\ &= \{(n_1, \dots, n_M) \mid \ 0 \leqslant n_1 \leqslant N^+ \leqslant N - q;\ 0 \leqslant N^- \leqslant n_M \leqslant N - q;\\ &\qquad 1 \leqslant q \leqslant T_{\min} \leqslant n_m - n_{m-1} \leqslant T_{\max} \leqslant N - q,\ m = 2, \dots, M\}\end{aligned} \quad (8)$$

допустимых наборов в задаче SRTFS и объединённое множество

$$\Omega = \Omega(N, N^-, N^+, T_{\min}, T_{\max}, q) = \bigcup_{M=M^-}^{M^+} \Omega_M \quad (9)$$

таких наборов всевозможных размерностей. Далее, следуя [2, 3], для каждого $M \in [M^-, M^+]$ определим область $\omega_m(M) = [n'_m(M), n''_m(M)]$, $m = 1, \dots, M$, где

$$\begin{aligned}n'_m(M) &= \max\{(m-1)T_{\min}, N^- - (M-m)T_{\max}\},\\ n''_m(M) &= \min\{N^+ + (m-1)T_{\max}, N - q - (M-m)T_{\min}\},\end{aligned}$$

допустимых значений m-й компоненты набора $(n_1, \dots, n_M)$. Кроме того, для каждого $M \in [M^-, M^+]$ такого, что $M \geqslant 2$, определим множество

$$\gamma^-_{m-1}(n) = \{j \mid {}^-n_{m-1}(n) \leqslant j \leqslant {}^+n_{m-1}(n)\},\ n \in \omega_m(M),\qquad m = 2, \dots, M, \quad (10)$$

где

$$\begin{aligned}{}^-n_{m-1}(n) &= \max\{(m-2)T_{\min},\ n - T_{\max})\},\\ {}^+n_{m-1}(n) &= \min\{N^+ + (m-2)T_{\max},\ n - T_{\min}\},\end{aligned}$$

допустимых значений $(m-1)$-й компоненты набора $(n_1, \dots, n_M)$ при условии, что m-я компонента фиксирована и равна n. Наконец, определим объединённые множества

$$\omega^+ = \bigcup_{M=M^-}^{M^+} \omega_1(M),\quad \omega^- = \bigcup_{M=M^-}^{M^+} \omega_M(M),\quad \omega^1 = \bigcup_{M=M^-}^{M^+} \bigcup_{m=1}^{M} \omega_m(M),$$

$$\omega^2 = \bigcup_{M=M^-}^{M^+} \bigcup_{m=2}^{M} \omega_m(M),\quad \gamma^-(n) = \bigcup_{m=2}^{M^+} \gamma^-_{m-1}(n),\quad n \in \omega^2. \quad (11)$$

Напомним [2, 3], что $\Omega_M \neq \varnothing$, $M \in [M^-, M^+]$, и $\Omega \neq \varnothing$ тогда и только тогда, когда

$$\lfloor (N - q)/T_{\min} \rfloor \geqslant \lfloor (N^- - N^+ + T_{\max} - 1)/T_{\max} \rfloor. \tag{12}$$

Условие (12) гарантирует совместность ограничений, входящих в определение (8).

Справедлива следующая

Теорема 1. *Пусть условие* (12), *связывающее входные данные задачи* SRTFS, *выполнено. Тогда максимум целевой функции* (1) *определяется по формуле*

$$G_{\max} = \max_{i \in \{1,\ldots,L\}} \max_{n \in \omega^-} G_i(n), \tag{13}$$

а значения функции $G_i(n)$, $i = 1, \ldots, L$, $n \in \omega^-$, *находятся по рекуррентным формулам*

$$G_i(n) = \begin{cases} g_1(n), & i = 1, \\ -\infty, & i = 2, \ldots, L, \end{cases} \qquad n \in \omega^+ \backslash \omega^2, \tag{14}$$

$$G_i(n) = \max_{j \in \gamma^-(n)} G_{l(i-1|L)}(j) + g_i(n), \quad i = 1, \ldots, L, \quad n \in \omega^2 \backslash \omega^+, \tag{15}$$

$$G_i(n) = \begin{cases} \max\{g_1(n), \max\limits_{j \in \gamma^-(n)} G_L(j) + g_1(n)\}, \; i = 1, \\ \max\limits_{j \in \gamma^-(n)} G_{i-1}(j) + g_i(n), \qquad i = 2, \ldots, L, \end{cases} \quad n \in \omega^+ \cap \omega^2, \tag{16}$$

где множества ω^-, ω^+, ω^2 *и* $\gamma^-(n)$ *определяются формулами* (11).

Доказательство. Так как выполнено условие (12), система ограничений совместна и $\Omega \neq \varnothing$. Используя определение (11), запишем (9) в эквивалентном виде:

$$\Omega = \begin{cases} \{(n_1) \mid n_1 \in \omega^-\}, \text{ если } M^+ = 1; \\ \{(n_1, \ldots, n_M) \mid \; n_M \in \omega^-; \, n_{m-1} \in \gamma^-(n_m), \, m = 2, \ldots, M; \\ \quad n_1 \in \omega^+; \; M \in [M^-, M^+]\}, \text{ если } M^+ > 1, M^- > 1; \\ \{(n_1) \mid n_1 \in \omega^-\} \cup \{(n_1, \ldots, n_M) \mid \; n_M \in \omega^-; \\ n_{m-1} \in \gamma^-(n_m), \; m = 2, \ldots, M; \, n_1 \in \omega^+; \, M \in [2, M^+]\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{если } M^+ > 1, M^- = 1. \end{cases} \tag{17}$$

Пусть значения аргументов N, N^-, N^+, $T_{\min}$, $T_{\max}$ и q, входящих в определение множества Ω, таковы, что $M^- = M^+ = 1$. Тогда из условий задачи следует, что $M = 1$. Следовательно, $\omega^+ = \omega^- = \omega^1$ и $\omega^2 = \varnothing$ согласно (11) и $\omega^- = \Omega$ в соответствии с (17). Кроме того, $l(m|L) = 1$

для любого фиксированного L, так как $m \in \{1, \dots, M\}$ согласно (1), а $M = 1$. Поэтому из (1) и (2) имеем $G = g_1(n_1)$, где $n_1 \in \omega^- = \omega^+$. Формулы (14) и (13) очевидны.

Пусть значения N, N^-, N^+, $T_{\min}$, $T_{\max}$ и q таковы, что $M^+ > 1$. Заметим, что по определению (11) множество ω^1 объединяет области допустимых значений всех компонент набора $(n_1, \dots, n_M)$ для всех допустимых значений M. Опираясь на (7), (9) и (11), для каждого $n \in \omega^1$ определим множество

$$\begin{aligned}\Psi(n) = \Psi(n|N, N^+, T_{\min}, T_{\max}, q) = \{(n_1, \dots, n_m) \mid n_m = n;\\ 0 \leqslant n_1 \leqslant N^+ \leqslant N - q; 1 \leqslant q \leqslant T_{\min} \leqslant n_k - n_{k-1} \leqslant T_{\max} \leqslant N - q;\\ k = 2, \dots, m;\ m \in [M^-(n), M^+(n)]\},\end{aligned} \quad (18)$$

где $M^+(n) = \lfloor n/T_{\min} \rfloor + 1$, а $M^-(n) = \lfloor (n - N^+ + T_{\max} - 1)/T_{\max} \rfloor + 1$, если $n > N^+$ и $M^-(n) = 1$, если $n \leqslant N^+$. Множество (18) включает наборы, у которых последняя компонента фиксирована и равна n, а размерность m наборов из этого множества лежит в интервале $[M^-(n), M^+(n)]$. При этом из определений (9) и (18) видно, что

$$\Psi(n) = \Omega(n + q, n, N^+, T_{\min}, T_{\max}, q|N^+, T_{\min}, T_{\max}, q).$$

Используя (17), запишем (18) в эквивалентном виде

$$\Psi(n) = \begin{cases} \{(n_1) \mid n_1 = n\}, \text{ если } M^+(n) = 1; \\ \{(n_1, \dots, n_m) \mid n_m = n;\ n_{k-1} \in \gamma^-(n_k),\ k = 2, \dots, m; \\ \quad n_1 \in \omega^+;\ m \in [M^-(n), M^+(n)]\}, \\ \qquad\qquad\qquad\qquad \text{если } M^+(n) > 1, M^-(n) > 1; \\ \{(n_1) \mid n_1 = n\} \cup \{(n_1, \dots, n_m) \mid n_m = n; \\ \quad n_{k-1} \in \gamma^-(n_k),\ k = 2, \dots, m;\ n_1 \in \omega^+;\ m \in [2, M^+(n)]\}, \\ \qquad\qquad\qquad\qquad \text{если } M^+(n) > 1, M^-(n) = 1. \end{cases}$$

Кроме того, для каждого $n \in \omega^1$ и $i = 1, \dots, L$ определим множество

$$\begin{aligned}\Psi_i(n) = \{(n_1, \dots, n_m) \mid (n_1, \dots, n_m) \in \Psi(n),\\ m \in [M^-(n), M^+(n)],\ l(m|L) = i\}.\end{aligned} \quad (19)$$

Это множество объединяет наборы $(n_1, \dots, n_m)$ из множества $\Psi(n)$ такие, что для последнего слагаемого $g_{l(m|L)}(n_m)$ в сумме $G(n_1, \dots, n_m)$ справедливо $n_m = n$ и $l(m|L) = i$. Наконец, для каждого $n \in \omega^1$ такого, что $M^+(n) > 1$, и каждого $i = 1, \ldots, L$ определим множество

$$\Psi_i^-(n) = \begin{cases} \{(n_1) \mid n_1 \in \gamma^-(n)\}, \text{ если } M^+(n) = 2, i = 2; \\ \varnothing, \text{ если } M^+(n) = 2, i \neq 2; \\ \{(n_1, \ldots, n_{m-1}) \mid l(m|L) = i;\ n_{m-1} \in \gamma^-(n); \\ \quad n_{k-1} \in \gamma^-(n_k),\ k = 2, \ldots, m-1;\ n_1 \in \omega^+; \\ \quad\quad m \in [M^-(n), M^+(n)]\}, \\ \qquad\qquad\qquad \text{если } M^+(n) > 2, M^-(n) > 2; \\ \{(n_1) \mid n_1 \in \gamma^-(n)\} \cup \{(n_1, \ldots, n_{m-1}) | \, l(m|L) = i; \\ \quad n_{m-1} \in \gamma^-(n);\ n_{k-1} \in \gamma^-(n_k),\ k = 2, \ldots, m-1; \\ \quad\quad n_1 \in \omega^+;\ m \in [3, M^+(n)]\}, \\ \qquad\qquad\qquad \text{если } M^+(n) > 2, M^-(n) = 2, i = 2; \\ \{(n_1, \ldots, n_{m-1}) \mid l(m|L) = i;\ n_{m-1} \in \gamma^-(n); \\ n_{k-1} \in \gamma^-(n_k), k = 2, \ldots, m-1;\ n_1 \in \omega^+; \\ m \in [3, M^+(n)]\}, \text{ если } M^+(n) > 2, M^-(n) = 2, i \neq 2. \end{cases} \tag{20}$$

Множество (20) включает по одному поднабору каждого из наборов (размерность которых не меньше 2) множества (19). Каждый поднабор включает все компоненты соответствующего набора за исключением последней компоненты.

Из определений (17)–(20) следуют равенства

$$\Psi(n) = \bigcup_{i=1}^{L} \Psi_i(n), \quad n \in \omega^1, \tag{21}$$

$$\Omega = \bigcup_{n \in \omega^-} \Psi(n) = \bigcup_{n \in \omega^-} \bigcup_{i=1}^{L} \Psi_i(n), \tag{22}$$

$$\Psi_i^-(n) = \begin{cases} \bigcup\limits_{j \in \gamma^-(n)} \Psi_L(j), & i = 1, \\ \bigcup\limits_{j \in \gamma^-(n)} \Psi_{i-1}(j), & i = 2, \ldots, L, \end{cases} \quad n \in \omega^2. \tag{23}$$

Целевую функцию (1) представим в виде

$$G(n_1, \ldots, n_M) = \sum_{j=1}^{J} \sum_{k=1}^{L} g_k(n_{(j-1)L+k}) + \sum_{k=1}^{i-1} g_k(n_{JL+k}) + g_i(n_M),$$

где $J = \lfloor M/L \rfloor$ — число полных повторов эталонного набора, а $i = l(m|L)$ — номер элемента из этого набора, соответствующий последнему фрагменту в числовой последовательности.

Пусть $n \in \omega^1$. Тогда согласно (11) найдутся $M \in [M^-, M^+]$ и $m \in [1, M]$ такие, что $n \in \omega_m(M)$. Поэтому существует такой набор $(n_1, \ldots, n_m, \ldots, n_M) \in \Omega_M$, в котором $n_m = n$, что его компоненты удовлетворяют ограничениям, входящим в определение (8). Для поднабора $(n_1, \ldots, n_m)$ этого набора из (8) имеем $0 \leqslant n_1 \leqslant N^+ \leqslant N - q$, $1 \leqslant q \leqslant T_{\min} \leqslant n_k - n_{k-1} \leqslant T_{\max} \leqslant N - q$, $k = 2, \ldots, m$, где $n_m = n$. Это значит, что в соответствии с (18) $(n_1, \ldots, n_{m-1}, n) \in \Psi(n)$, т. е. $\Psi(n) \neq \varnothing$, $n \in \omega^1$. Отсюда и из (19), (21) следует, что $\Psi_i(n) \neq \varnothing$, $n \in \omega^1$, тогда и только тогда, когда найдётся $m \in [M^-(n), M^+(n)]$ такое, что $l(m|L) = i$. В противном случае $\Psi_i(n) = \varnothing$.

Положим $\eta = (n_1, \ldots, n_m)$. Рассмотрим семейство задач

$$\{\langle i, n\rangle,\ i = 1, \ldots, L,\ n \in \omega^1\}$$

вычисления максимума функции $G_i(\eta | n_m = n,\ l(m|L) = i)$. Для максимума этой функции положим

$$G_i(n) = \begin{cases} \max\limits_{\eta \in \Psi_i(n)} G(\eta), & \Psi_i(n) \neq \varnothing, \\ -\infty, & \Psi_i(n) = \varnothing, \end{cases} \quad i = 1, \ldots, L,\ n \in \omega^1, \tag{24}$$

учитывая, что множество $\Psi_i(n)$ может быть пустым.

Покажем справедливость формул (14)–(16). Из (11) имеем $\omega^1 = \omega^+ \cup \omega^2$. Поэтому множество ω^1 можно разбить на три непересекающихся подмножества: $\omega^+ \setminus \omega^2$, $\omega^+ \cap \omega^2$ и $\omega^2 \setminus \omega^+$. Каждому из этих подмножеств соответствует свой вариант вычисления значений функции (24), представленный в формулах (14)–(16). Рассмотрим эти варианты.

Пусть $n \in \omega^+ \backslash \omega^2$. Тогда из (11) и (18) имеем $M^-(n) = M^+(n) = 1$. Поскольку $m \in \{M^-(n), \ldots, M^+(n)\} = \{1\}$, равенство $l(m|L) = i$ справедливо при единственном значении $i = 1$. Отсюда следует, что $\Psi_i(n) \neq \varnothing$ при $i = 1$ и $\Psi_i(n) = \varnothing$ для каждого $i = 2, \ldots, L$. В случае $i = 1$ из (19), (21) и (11) находим $\Psi_1(n) = \{(n_1) \mid n_1 = n\}$. Поэтому

$$G_1(n) = \max_{\eta \in \Psi_1(n)} G(\eta) = \max_{n_1 \in \{n\}} \sum_{k=1}^{1} g_{l(k|L)}(n_k) = g_1(n),$$

что устанавливает справедливость первого случая формулы (14). Справедливость второго случая этой формулы для $i = 2, \ldots, L$ следует непосредственно из (24).

Допустим, что $n \in \omega^2 \setminus \omega^+$. Тогда из (11) и (18) имеем $M^-(n) > 1$, $M^+(n) > 1$ и $\gamma^-(n) \neq \varnothing$. Как и в предыдущем варианте, в зависимости от значения $i = 1, \ldots, L$ возможны два случая: $\Psi_i(n) = \varnothing$ и $\Psi_i(n) \neq \varnothing$.

Если значение $i = 1, \ldots, L$ таково, что $\Psi_i(n) = \varnothing$, то согласно (20) $l(m|L) \neq i$ для каждого $m \in [M^-(n), M^+(n)]$. При этом для левой части формулы (15) из (24) имеем $G_i(n) = -\infty$. С другой стороны, так как в соответствии с (10), (11) и (18) для каждого $j \in \gamma^-(n)$ выполнены равенства $M^-(j) = M^-(n) - 1$ и $M^+(j) = M^+(n) - 1$, то очевидно, что $l(k|L) \neq l(i-1|L)$ для каждого $k \in [M^-(j), M^+(j)]$. Поэтому из (20) имеем $\Psi_{l(i-1|L)}(j) = \varnothing$ для каждого $j \in \gamma^-(n)$. Тогда из (24) следует, что $G_{l(i-1|L)}(j) = -\infty$, $j \in \gamma^-(n)$, и $\max\limits_{j \in \gamma^-(n)} G_{l(i-1|L)}(j) = -\infty$. Отсюда для правой части (15) находим $\max\limits_{j \in \gamma^-(n)} G_{l(i-1|L)}(j) + g_i(n) = -\infty$, так как $g_i(n)$ конечно. Таким образом, показано равенство правой и левой частей формулы (15), что устанавливает её справедливость.

Рассмотрим второй случай, когда $i = 1, \ldots, L$ и $\Psi_i(n) \neq \varnothing$. Поскольку $M^-(n) > 1$, размерность m каждого набора в множестве $\Psi_i(n)$ больше единицы. Далее, из определений (19) и (20) следует, что если $n \in \omega^1$ и $M^+(n) > 1$, то множество $\Psi_i^-(n)$ пусто или непусто одновременно с множеством $\Psi_i(n)$. Причём, если $\Psi_i(n) \neq \varnothing$ и $\Psi_i^-(n) \neq \varnothing$, то между наборами из множеств $\Psi_i(n)$ и $\Psi_i^-(n)$ существует взаимно однозначное соответствие: $(n_1, \ldots, n_{m-1}, n) \in \Psi_i(n)$ тогда и только тогда, когда $(n_1, \ldots, n_{m-1}) \in \Psi_i^-(n)$. Следовательно, в соответствии с (19) и (20) при каждом фиксированном n максимизация по наборам $\eta = (n_1, \ldots, n_{m-1}, n) \in \Psi_i(n)$ эквивалентна максимизации по наборам $\eta^- = (n_1, \ldots, n_{m-1}) \in \Psi_i^-(n)$. Поэтому, используя (23) и обозначив через $m(\eta)$ размерность набора η, получим цепочку равенств

$$
\begin{aligned}
G_i(n) &= \max_{\eta \in \Psi_i(n)} \sum_{k=1}^{m(\eta)} g_{l(k|L)}(n_k) \\
&= \max_{\eta \in \Psi_i(n)} \left\{ g_{l(m|L)}(n_m) + \sum_{k=1}^{m(\eta)-1} g_{l(k|L)}(n_k) \right\} \\
&= \max_{\eta \in \Psi_i(n)} \left\{ g_i(n) + \sum_{k=1}^{m(\eta)-1} g_{l(k|L)}(n_k) \right\} \\
&= g_i(n) + \max_{\eta^- \in \Psi_i^-(n)} \sum_{k=1}^{m(\eta^-)} g_{l(k|L)}(n_k) \\
&= g_i(n) + \max_{j \in \gamma^-(n)} \max_{\eta^- \in \Psi_{l(i-1|L)}(j)} \sum_{k=1}^{m(\eta^-)} g_{l(k|L)}(n_k)
\end{aligned}
$$

$$= g_i(n) + \max_{j \in \gamma^-(n)} G_{l(i-1|L)}(j),$$

которая устанавливает справедливость (15).

Рассмотрим последний вариант вычислений, когда $n \in \omega^+ \cap \omega^2$. Этот вариант является комбинацией уже рассмотренных. Заметим, что в этом варианте, как и в предыдущем, $M^+ > 1$, $\gamma^-(n) \neq \varnothing$, однако $M^- = 1$ в соответствии с (11) и (18). При этом возможны те же два случая: $\Psi_i(n) \neq \varnothing$ и $\Psi_i(n) = \varnothing$.

Рассмотрим первый случай. Если значение $i \in \{1, \ldots, L\}$ таково, что $\Psi_i(n) \neq \varnothing$, то множество $\Psi_i(n)$ можно разбить на два непересекающихся подмножества, по крайней мере одно из которых непусто:

$$\Psi_i(n) = \{\eta \in \Psi_i(n) \mid m(\eta) = 1\} \cup \{\eta \in \Psi_i(n) \mid m(\eta) > 1\}. \tag{25}$$

В соответствии с этим разбиением возможны 3 подслучая: первое подмножество пусто, а второе непусто; первое подмножество непусто, а второе пусто; оба подмножества непусты. Проанализируем каждый из них по отдельности.

Пусть $\{\eta \in \Psi_i(n) \mid m(\eta) = 1\} = \varnothing$. Это равенство справедливо для всех $i = 2, \ldots, L$, так как $l(m(\eta)|L) = l(1|L) = i$ лишь при $i = 1$ в соответствии с (19). Поэтому из (25) для этих значений i имеем $\Psi_i(n) = \{\eta \in \Psi_i(n) \mid m(\eta) > 1\}$ и справедлива та же цепочка равенств, что и при доказательстве формулы (15), т. е. формула (15) применима при вычислении значения $G_i(n)$, $i = 2, \ldots, L$. Для окончательной записи второго случая формулы (16) следует заметить, что $l(i-1|L) = i$ при $i = 2, \ldots, L$.

Два оставшихся подслучая возможны только при $i = 1$, что соответствует первому случаю формулы (16). Пусть оба подмножества в разложении (25) непусты. Ясно, что для наборов из этих подмножеств справедливы такие же утверждения и формулировки, как и при доказательстве формул (14) и (15). Поэтому из (24) и (25) получаем

$$\begin{aligned}
G_1(n) &= \max_{\eta \in \{\Psi_1(n)|m(\eta)=1\} \cup \{\Psi_1(n)|m(\eta)>1\}} G(\eta) \\
&= \max\{\max_{\eta \in \{\Psi_1(n)|m(\eta)=1\}} G(\eta), \max_{\eta \in \{\Psi_1(n)|m(\eta)>1\}} G(\eta)\} \\
&= \max\{g_1(n),\, g_1(n) + \max_{j \in \gamma^-(n)} G_L(j)\}.
\end{aligned}$$

Эти равенства устанавливают справедливость формулы (16) для второго из рассматриваемых подслучаев.

Пусть $\{\eta \in \Psi_i(n) \mid m(\eta) > 1\} = \varnothing$, тогда в соответствии с (25) $\Psi_i(n) = \{\eta \in \psi_i(n) \mid m(\eta) = 1\}$. Повторяя доказательство для первого варианта вычислений, устанавливаем равенства $G_1(n) = g_1(n)$ и

$$g_1(n) + \max_{j\in\gamma^-(n)} G_L(j) = -\infty.$$

Тем самым равенство $G_1(n) = \max\{g_1(n), g_1(n) + \max\limits_{j\in\gamma^-(n)} G_L(j)\}$ завершает рассмотрение первого случая.

Рассмотрим второй случай. Заметим, что $\Psi_i(n) = \varnothing$ при $i = 2, \dots, L$. Действительно, так как $\omega^+ \cap \omega^2 \subseteq \omega^+$, то для $i = 1$ имеем $\Psi_1(n) \neq \varnothing$ при каждом $n \in \omega^+ \cap \omega^2$ в соответствии с (11). Повторяя доказательство, приведённое для варианта $n \in \omega^2 \setminus \omega^+$, получим, что правая и левая части второго случая формулы (16) совпадают. Формулы (14)–(16) доказаны.

Наконец, из (1), (17), (22) и (24) имеем

$$\begin{aligned} G_{\max} &= \max_{\eta\in\Omega} G(\eta) = \max_{\eta\in \bigcup\limits_{n\in\omega^-} \bigcup\limits_{i=1}^{L} \Psi_i(n)} G(\eta) \\ &= \max_{n\in\omega^-} \max_{i\in\{1,\dots,L\}} \max_{\eta\in\Psi_i(n)} G(\eta) = \max_{n\in\omega^-} \max_{i\in\{1,\dots,L\}} G_i(n). \end{aligned}$$

Теорема 1 доказана.

Для отыскания оптимального набора $(\widehat{n}_1, \dots, \widehat{n}_{\widehat{M}})$ определим вспомогательные функции. Функцию $r(i,n)$, $i \in \{1, \dots, L\}$, $n \in \omega^1$, — индикатор остановки процесса — определим, опираясь на формулы (14)–(16), так что

$$r(i,n) = \begin{cases} -1, & n \in \omega^+ \setminus \omega^2, \quad i = 2, \dots, L, \\ 0, & n \in \omega^+ \setminus \omega^2, \quad i = 1, \\ 0, & n \in \omega^+ \cap \omega^2, \quad i = 1, \ \max\limits_{j\in\gamma^-(n)} G_L(j) < 0, \\ 1, & n \in \omega^+ \cap \omega^2, \quad i = 1, \ \max\limits_{j\in\gamma^-(n)} G_L(j) \geqslant 0, \\ 1, & n \in \omega^+ \cap \omega^2, \quad i = 2, \dots, L, \\ 1, & n \in \omega^2 \setminus \omega^+, \quad i = 1, \dots, L. \end{cases} \tag{26}$$

Далее, для значений $n \in \omega^2$ и $i \in \{1, \dots, L\}$ определим функцию

$$I(i,n) = \arg\max_{j\in\gamma^-(n)} G_{l(i-1|L)}(j), \tag{27}$$

которую назовём *условно-оптимальным управлением* на (i,n)-м шаге оптимизации. Кроме того, используя (26), определим функцию

$$\nu(m) = \begin{cases} \arg\max\limits_{n\in\omega^-} \{\max\limits_{i\in\{1,\dots,L\}} G_i(n)\}, & m = 1, \\ I(l(i^* - (m-1) + 1|L), \nu(m-1)), & m = 2, \dots, M^*, \end{cases} \tag{28}$$

где
$$i^* = \arg \max_{i \in \{1,\dots,L\}} G_i(\nu(1)), \tag{29}$$
$$M^* = \min\{m \mid m \in \mathbb{N},\, r(l(i^* - m + 1|L), \nu(m)) = 0\}. \tag{30}$$

Следствие 1. *Пусть выполнены условия теоремы* 1. *Тогда компоненты оптимального набора* $\widehat{\eta} = (\widehat{n_1}, \dots, \widehat{n}_{\widehat{M}})$ *определяются по правилу*
$$\widehat{n}_m = \nu(\widehat{M} - m + 1), \quad m = 1, \dots, \widehat{M}, \tag{31}$$
где $\widehat{M} = M^*$.

Доказательство. Из определения (26) и формулы (16) следует, что $r(i, n) = 0$ тогда и только тогда, когда одновременно выполнено
$$i = 1,\ n \in \omega^+,\ G_1(n) = g_1(n). \tag{32}$$
Далее, из (30) имеем $r(l(i^* - M^* + 1|L), \nu(M^*)) = 0$. Из этого равенства и (32) находим
$$\left.\begin{array}{l} l(i^* - M^* + 1|L) = 1,\ \nu(M^*) \in \omega^+, \\ G_{l(i^* - M^* + 1|L)}(\nu(M^*)) = g_1(\nu(M^*)). \end{array}\right\} \tag{33}$$
Кроме того, $\nu(1) \in \omega^-$, $\nu(m) \in \gamma^-(\nu(m-1))$, $m = 2, \dots, M^*$, в соответствии с (27) и (28). Отсюда в силу (17) следует, что $(\nu(M^*), \dots, \nu(1)) \in \Omega$ и $M^* \in [M^-, M^+]$.

Покажем, что набор $(\nu(M^*), \dots, \nu(1))$ оптимален. Из (1) находим
$$\begin{aligned} G(\nu(M^*), \dots, \nu(1)) &= \sum_{m=1}^{M^*} g_{l(m|L)}(\nu(M^* - m + 1)) \\ &= \sum_{m=1}^{M^*} g_{l(M^* - m + 1|L)}(\nu(m)). \end{aligned} \tag{34}$$
Остаётся показать, что правая часть (34) равна оптимальному значению целевой функции.

Из (28), (29) и (13) имеем
$$G_{\max} = \max_{n \in \omega^-} \max_{i \in \{1,\dots,L\}} G_i(n) = G_{i^*}(\nu(1)). \tag{35}$$

Опираясь на (14)–(16), (26) и (27), устанавливаем, что для $n \in \omega^2$ и $i = 1, \dots, L$ таких, что $r(i, n) = 1$, справедливо равенство
$$G_i(n) = G_{l(i-1|L)}(I(l(i-1|L), n)) + g_i(n). \tag{36}$$

Из формул (2), (26)–(30) находим, что $l(i^* - m + 1|L) \in \{1, \dots, L\}$, $m = 1, \dots, M^*$; $\nu(m) \in \omega^2$, $m = 1, \dots, M^* - 1$; $r(l(i^* - m + 1|L), \nu(m))$, $m = 1, \dots, M^* - 1$. Поэтому равенство (36) справедливо при подстановке $i = l(i^* - m + 1|L)$, $n = \nu(m)$, $m = 1, \dots, M^* - 1$. Сделав эту подстановку, из (26), (27) и (36) получаем рекуррентную формулу

$$\begin{aligned} &G_{l(i^*-m+1|L)}(\nu(m)) \\ &\quad = G_{l(i^*-m+1-1|L)}(I(l(i^* - m + 1 - 1|L), \nu(m))) + g_{l(i^*-m+1|L)}(\nu(m)) \\ &\qquad\qquad = G_{l(i^*-(m+1)+1|L)}(\nu(m+1)) + g_{l(i^*-m+1|L)}(\nu(m)). \quad (37) \end{aligned}$$

Наконец, используя (37) и очевидное равенство $l(M^* - m + 1|L) = l(i^* - m + 1|L)$, $m = 1, \dots, M^*$, из (35) находим

$$\begin{aligned} G_{\max} &= G_{i^*}(\nu(1)) = G_{l(i^*-1|L)}(\nu(2)) + g_{i^*}(\nu(1)) \\ &= G_{l(i^*-2|L)}(\nu(3)) + g_{l(i^*-1|L)}(\nu(2)) + g_{i^*}(\nu(1)) = \dots \\ &= G_{l(i^*-M^*+1|L)}(\nu(M^*)) + \sum_{m=1}^{M^*-1} g_{l(i^*-m+1|L)}(\nu(m)) \\ &= \sum_{m=1}^{M^*} g_{l(i^*-m+1|L)}(\nu(m)). \end{aligned}$$

Следствие 1 доказано.

Представим алгоритм решения задачи SRTFS в виде следующих последовательно выполняемых шагов.

Шаг 1 (анализ корректности входных данных). Проверка справедливости неравенства (12). Если оно не выполнено, то входные данные не удовлетворяют условию совместности системы ограничений задачи SRTFS и последующие шаги не производятся.

Шаг 2 (подготовка данных для оптимизации). Вычисление значений функции $g_i(n)$, $i = 1, \dots, L$, для каждого $n \in \{0, \dots, N - q\}$ по формулам (2) и (3).

Шаг 3 (прямой ход, пошаговая оптимизация). Вычисление последовательности значений $G_i(n)$ условных максимумов и значения $G_{\max}$ глобального максимума по формулам (13)–(16).

Шаг 4 (обратный ход). Вычисление вспомогательных функций $r(i, n)$, $I(i, n)$, $\nu(m)$, а также значений i^* и M^* по формулам (26)–(30). Определение компонент оптимального набора по формуле (31).

3. Временна́я сложность

Для оценивания временно́й сложности алгоритма воспользуемся его пошаговой записью (см. п. 2).

Следствие 2. *Временна́я сложность алгоритма решения задачи* SRTFS *равна* $O(L(T_{\max} - T_{\min} + q)(N - q + 1))$ *в частном случае, когда значения переменных* $T_{\min}$, $T_{\max}$, N^- *и* N^+ *являются частью входа задачи, и* $O(LN^2)$ *в общем случае.*

Доказательство. Из (10) и (11) заключаем, что мощности множеств ω^+, ω^-, ω^2, $\omega^+ \cap \omega^2$, $\omega^2 \setminus \omega^+$ и $\omega^+ \setminus \omega^2$ не превосходят $N - q + 1$. Кроме того, $|\gamma^-(n)| \leqslant T_{\max} - T_{\min} + 1$, $n \in \omega^2$.

На первом шаге алгоритма число операций есть константа, не зависящая от размера входа задачи. Второй шаг требует $O(Lq(N - q + 1))$ элементарных операций. Анализируя формулы (13)–(16), легко установить, что трудоёмкость третьего шага алгоритма равна

$$O\Big(L|\omega^+ \setminus \omega^2| + \sum_{n\in\omega^+\cap\omega^2} (|\gamma^-(n)| + 1) + (L-1) \sum_{n\in\omega^+\cap\omega^2} |\gamma^-(n)| \\ + L \sum_{n\in\omega^2\setminus\omega^+} |\gamma^-(n)| + L|\omega^-|\Big) = O(L(T_{\max} - T_{\min} + 1)(N - q + 1)).$$

На четвёртом шаге согласно (26)–(31) число элементарных операций оценивается величиной

$$O\Big(L|\omega^1| + L \sum_{n\in\omega^2} (|\gamma^-(n)| + 1) + L|\omega^-| + M^+\Big) \\ = O(L(T_{\max} - T_{\min} + 1)(N - q + 1)),$$

так как $M^+ \leqslant N - q + 1$. Суммируя затраты по времени на втором, третьем и четвёртом шагах, находим оценку временно́й сложности для частного случая задачи SRTFS. Из этой оценки, используя границы сверху и снизу для входящих в неё переменных, получаем оценку временно́й сложности для общего случая анализируемой задачи.

В [3] был обоснован алгоритм решения варианта задачи SRTFS, в котором число M является оптимизируемой величиной. Временна́я сложность этого алгоритма оценивается величиной $O(M^+(T_{\max}-T_{\min}+q)(N-q+1))$ в частном случае и $O(M^+N^2)$ — в общем случае. Если $L < M^+$, то легко видеть, что предлагаемый алгоритм в M^+/L раз эффективнее по сравнению с алгоритмом, предложенным в [3].

Заключение

Рассмотренная задача входит в большое семейство (насчитывающее несколько сотен элементов) актуальных оптимизационных задач [4], к которым сводятся типовые проблемы помехоустойчивого анализа структурированных данных и распознавания образов. Для решения этих задач особенно важны наиболее «быстрые» точные и приближённые полиномиальные алгоритмы. В связи с этим в настоящей работе для решения одной из таких задач обоснован новый более эффективный по сравнению с известным аналогом точный полиномиальный алгоритм, реализующий схему динамического программирования. Демонстрационная версия программы доступна через Интернет [4].

Открытым остаётся вопрос о разрешимости обобщения задачи SRTFS на случай, когда вместо поиска в числовой последовательности повторяющегося набора фрагментов, элементы которого упорядочены в соответствии с эталонным набором векторов, требуется найти повторы фрагментов с точностью до всевозможных перестановок элементов эталонного набора векторов. Алгоритм решения этой задачи представляет значительный интерес для ряда упомянутых во введении приложений. Обоснование алгоритма решения этой задачи представляется делом ближайшей перспективы.

ЛИТЕРАТУРА

1. **Кельманов А. В., Михайлова Л. В., Хамидуллин С. А.** Оптимальное обнаружение в квазипериодической последовательности повторяющегося набора эталонных фрагментов // Материалы Российской конф. «Дискретная оптимизация и исследование операций». Владивосток, 7–14 сентября 2007. — Новосибирск: Изд-во Ин-та математики СО РАН, 2007. — С. 178. http://math.nsc.ru/conference/door07/DOOR_abstracts.pdf.
2. **Кельманов А. В., Михайлова Л. В., Хамидуллин С. А.** Апостериорное обнаружение в квазипериодической последовательности повторяющегося набора эталонных фрагментов // Журн. вычисл. математики и мат. физики. 2008. Т. 48. № 12. С. 2247–2260.
3. **Кельманов А. В., Михайлова Л. В., Хамидуллин С. А.** Оптимальное обнаружение в квазипериодической последовательности повторяющегося набора эталонных фрагментов // Сиб. журн. вычисл. математики. 2008, Т. 11, № 3. С. 311–327.
4. http://math.nsc.ru/ serge/qpsl/

Кельманов Александр Васильевич,
e-mail: kelm@math.nsc.ru

Михайлова Людмила Викторовна,
e-mail: mikh@math.nsc.ru

Хамидуллин Сергей Асгадуллович,
e-mail: kham@math.nsc.ru

Статья поступила
10 февраля 2009 г.
Переработанный вариант —
5 марта 2009 г.