

УДК 519.72

РАСПОЗНАВАНИЕ КВАЗИПЕРИОДИЧЕСКОЙ ПОСЛЕДОВАТЕЛЬНОСТИ, ВКЛЮЧАЮЩЕЙ НЕИЗВЕСТНОЕ ЧИСЛО НЕЛИНЕЙНО-РАСТЯНУТЫХ ЭТАЛОННЫХ ПОДПОСЛЕДОВАТЕЛЬНОСТЕЙ¹⁾

© 2021 г. А. В. Кельманов^{1,2}, Л. В. Михайлова^{1,*}, П. С. Рузанкин^{1,2,**}, С. А. Хамидуллин¹¹ 630090 Новосибирск, пр-т акад. Коптюга, 4, Ин-т матем. им. С.Л. Соболева, Россия² 630090 Новосибирск, ул. Пирогова, 2, Новосибирский гос. ун-т, Россия

*e-mail: mikh@math.nsc.ru

**e-mail: ruzankin@math.nsc.ru

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Рассматривается неизученная экстремальная задача, которая индуцируется одной из задач помехоустойчивого распознавания квазипериодической последовательности, а именно, задачей распознавания последовательности Y длины N как последовательности, порожденной некоторой последовательностью U , принадлежащей заданному конечному множеству W (алфавиту) последовательностей. Каждая последовательность U из W порождает экспоненциальное по мощности множество $\mathcal{X}(U)$ последовательностей, объединяющее все последовательности длины N , которые в качестве подпоследовательностей включают переменное число допустимых квазипериодических (флуктуационных) повторов последовательности U . Каждый квазипериодический повтор порождается допустимыми преобразованиями последовательности U , а именно, сдвигами и растяжениями. Задача распознавания состоит в выборе последовательности U из W и аппроксимации последовательности Y элементом X из множества $\mathcal{X}(U)$ последовательностей. Критерием аппроксимации является минимум суммы квадратов расстояний между элементами последовательностей. Мы показываем, что рассматриваемая задача эквивалентна задаче суммирования элементов двух числовых последовательностей, в которой требуется минимизировать сумму неизвестного числа M слагаемых, каждое из которых является разностью невзвешенной автосвертки растянутой на переменную длину последовательности U (путем кратных повторов ее элементов) и взвешенной свертки этой растянутой последовательности с подпоследовательностью из Y . Мы доказываем, что рассматриваемая экстремальная задача и вместе с ней задача распознавания разрешимы за полиномиальное время. Примерами численного моделирования проиллюстрирована применимость алгоритма к решению модельных прикладных задач помехоустойчивой обработки ECG-подобных и PPG-подобных квазипериодических сигналов (electrocardiogram-like and photoplethysmogram-like signals). Библ. 9. Фиг. 5.

Ключевые слова: числовые последовательности, распознавание, квазипериодическая последовательность, полиномиальная разрешимость, разность взвешенных сверток.

DOI: 10.31857/S0044466921070097

ВВЕДЕНИЕ

В работе рассматривается неизученная экстремальная задача, которая индуцируется одной из задач помехоустойчивого распознавания квазипериодической последовательности. Цель работы — доказательство полиномиальной разрешимости данной задачи и построение алгоритма, гарантирующего получение оптимального решения. Поводом для проведения исследований послужило отсутствие каких-либо эффективных (полиномиальных) вычислительных алгоритмов с априорно гарантированными оценками точности для ее решения.

¹⁾ Работа выполнена при финансовой поддержке РФФИ, проекты 19-07-00397 и 19-01-00308, программы ФНИ РАН, проект 0314-2019-0015, а также программы Тор-5-100 Минобрнауки РФ.

Рассматриваемая задача актуальна для помехоустойчивого мониторинга природных объектов, типичное состояние которых во времени квазипериодически повторяется с флуктуациями. То есть расстояние между двумя последовательными повторами лежит в заданном интервале, а типичное состояние допускает некоторую вариативность от повтора к повтору. А именно, для прикладных задач, когда требуется помимо обнаружения этих типовых повторов идентифицировать (распознать) либо сам объект, либо состояние объекта среди множества допустимых.

Подобный характер повторяемости состояний типичен, в первую очередь, для биомедицинских задач. В частности, задач анализа и распознавания ECG и PPG сигналов. Поэтому для иллюстрации работы алгоритма далее приведены примеры модельных прикладных задач распознавания ECG и PPG-подобных квазипериодических сигналов.

1. ФОРМУЛИРОВКА ЗАДАЧИ, ЕЕ ИСТОКИ И ТРАКТОВКИ

Рассматриваемая экстремальная задача имеет следующую формулировку.

Задача 1. Дано: числовая последовательность $Y = (y_1, \dots, y_N)$, совокупность $W = \{U^{(1)}, \dots, U^{(K)} \mid U^{(k)} = (u_1^{(k)}, \dots, u_{q_k}^{(k)}), k = 1, \dots, K\}$, натуральные числа T_{\max} и ℓ . Найти: числовую последовательность $U = (u_1, \dots, u_{q(U)}) \in W$, набор $\mathcal{M} = \{n_1, \dots, n_m, \dots\}$ номеров последовательности Y , набор $\mathcal{P} = \{p_1, \dots, p_m, \dots\}$ натуральных чисел, набор $\mathcal{J} = \{J^{(1)}, \dots, J^{(m)}, \dots\}$ сжимающих отображений, в котором $J^{(m)} : \{1, \dots, p_m\} \rightarrow \{1, \dots, q(U)\}$, а также размерность M этих наборов, доставляющих минимум целевой функции

$$F(U, \mathcal{M}, \mathcal{P}, \mathcal{J}) = \sum_{m=1}^M \sum_{i=1}^{p_m} \{u_{J^{(m)}(i)}^2 - 2y_{n_m+i-1}u_{J^{(m)}(i)}\}, \tag{1.1}$$

при ограничениях

$$q(U) \leq p_m \leq \ell \leq T_{\max} \leq N, \quad m = 1, \dots, M, \tag{1.2}$$

$$p_{m-1} \leq n_m - n_{m-1} \leq T_{\max}, \quad m = 2, \dots, M, \tag{1.3}$$

$$p_M \leq N - n_M + 1, \tag{1.4}$$

на элементы искомых наборов \mathcal{M} , \mathcal{P} , и при ограничениях

$$\begin{aligned} J^{(m)}(1) &= 1, \quad J^{(m)}(p_m) = q(U), \\ 0 \leq J^{(m)}(i) - J^{(m)}(i-1) &\leq 1, \quad i = 2, \dots, p_m, \quad m = 1, \dots, M, \end{aligned} \tag{1.5}$$

на элементы искомых сжимающих отображений.

Приведем несколько трактовок задачи 1.

Из формулировки задачи 1 и вида целевой функции (1.1) следует, что задача 1 – задача об оптимальном (в смысле минимума (1.1)) суммировании элементов двух последовательностей, одна из которых Y задана, а другая U принадлежит заданному множеству последовательностей. Эквивалентная перезапись целевой функции (1.1) в виде

$$F(U, \mathcal{M}, \mathcal{P}, \mathcal{J}) = \sum_{m=1}^M \left\{ \sum_{i=1}^{p_m} u_{J^{(m)}(i)}^2 - 2 \sum_{i=1}^{p_m} y_{n_m+i-1} u_{J^{(m)}(i)} \right\}$$

позволяет трактовать задачу 1 как задачу минимизации суммы разностей взвешенных сверток. Действительно, при каждом $m = 1, \dots, M$ выражение $\sum_{i=1}^{p_m} u_{J^{(m)}(i)}^2$ в фигурных скобках – невзвешенная автосвертка последовательности $u_{J^{(m)}(i)}$, $i = 1, \dots, p_m$, полученной из элементов U растяжением путем дублирования элементов, а выражение $\sum_{i=1}^{p_m} y_{n_m+i-1} u_{J^{(m)}(i)}$ – свертка этой растянутой последовательности с подпоследовательностью из Y , имеющей ту же длину p_m (коэффициент 2 – вес этой свертки).

Другая возможная трактовка — задача совместного выбора последовательности $U \in W$ и аппроксимации последовательности Y последовательностью $X \in \mathcal{X}(U)$, по критерию минимума суммы квадратов расстояний между элементами Y и X , т.е. задача

$$\|Y - X\|^2 \rightarrow \min_{U, \mathcal{X}(U)}. \quad (1.6)$$

Здесь $\mathcal{X}(U)$, $U \in W$, — множество допустимых аппроксимирующих последовательностей, порожденных последовательностью U . Каждый элемент $X = (x_1, \dots, x_N) \in \mathcal{X}(U)$ однозначно определяется наборами \mathcal{M} , \mathcal{P} , \mathcal{J} , удовлетворяющими ограничениям (1.2)–(1.5), по правилу

$$x_n = \sum_{m=1}^M h_{n-n_m+1}^{(m)}, \quad n = 1, \dots, N, \quad (1.7)$$

где

$$h_i^{(m)} = \begin{cases} u_{J^{(m)}(i)}, & \text{если } i = 1, \dots, p_m, \\ 0, & \text{если } i < 1, \quad i > p_m, \end{cases} \quad m = 1, \dots, M. \quad (1.8)$$

Равенство (1.8) устанавливает связь между элементами последовательностей $h_i^{(m)}$ и U . Из этого равенства видно, что каждая из последовательностей $h_i^{(m)}$ — растянутая до длины p_m последовательность U , а кратность дублирования ее элементов определяется формулой

$$k_t^{(m)} = \{i | J^{(m)}(i) = t, i \in \{1, \dots, p_m\}\}, \quad t = 1, \dots, q(U), \quad (1.9)$$

причем

$$p_m = k_1^{(m)} + \dots + k_{q(U)}^{(m)}, \quad m = 1, \dots, M. \quad (1.10)$$

Формула (1.7) — сумма M растянутых последовательностей вида (1.8). Таким образом, последовательность X включает в себя M повторов растянутых последовательностей U . Значение индекса $n = n_m$, $n_m \in \mathcal{M}$, определяет начальный номер m -го повтора, а значение $p = p_m$, $p_m \in \mathcal{P}$, и отображение $J = J^{(m)}$, $J^{(m)} \in \mathcal{J}$, — длину этого повтора и кратности дублирования элементов из U в этом повторе.

Неформально можно сказать, что каждая последовательность $X \in \mathcal{X}(U)$ включает в себя неизвестное число допустимых квазипериодических повторов последовательности U . При этом каждый повтор определяется 1) сдвигом U на переменную величину, которая между соседними повторами не превышает $T_{\max} \leq N$; 2) допустимым растяжением последовательности U , путем дублирования ее компонент.

Из (1.7) и (1.8) следует, что для каждого $U \in W$ между элементами множества $\mathcal{X}(U)$ и наборами \mathcal{M} , \mathcal{P} , \mathcal{J} , удовлетворяющими ограничениям (1.2)–(1.5), существует взаимнооднозначное соответствие, т.е. $X = X(U, \mathcal{M}, \mathcal{P}, \mathcal{J})$, поэтому задачу (1.6) можно эквивалентно записать в виде

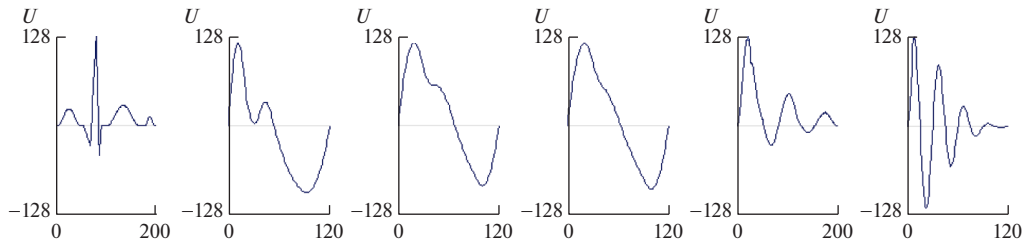
$$\|Y - X\|^2 = \|Y - X(U, \mathcal{M}, \mathcal{P}, \mathcal{J})\|^2 \rightarrow \min_{U, \mathcal{M}, \mathcal{P}, \mathcal{J}}. \quad (1.11)$$

Наконец, преобразовав $\|Y - X\|^2$ с учетом (1.7) и (1.8), имеем

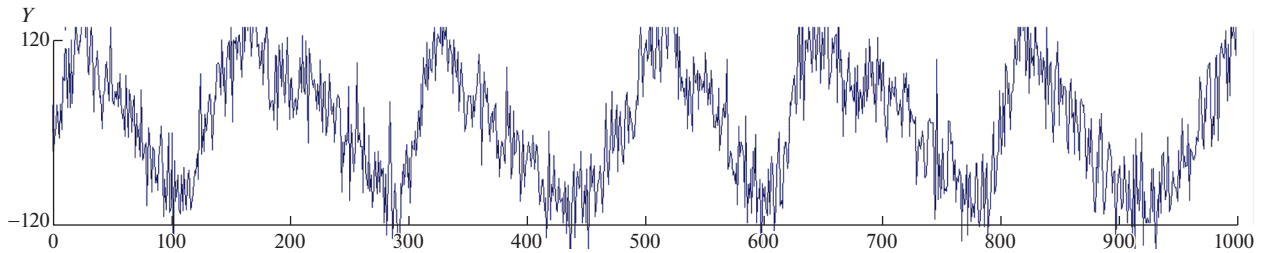
$$\sum_{n=1}^N (x_n - y_n)^2 = \sum_{n=1}^N y_n^2 + \sum_{m=1}^M \sum_{i=1}^{p_m} \{u_{J^{(m)}(i)}^2 - 2y_{n_m+i-1} u_{J^{(m)}(i)}\}.$$

Легко видеть, что первое слагаемое в правой части этого равенства — константа, не зависящая от переменных задачи 1, а второе совпадает с целевой функцией задачи 1. Поэтому задача (1.11), а вместе с ней и задача (1.6), эквивалентна оптимизационной задаче 1.

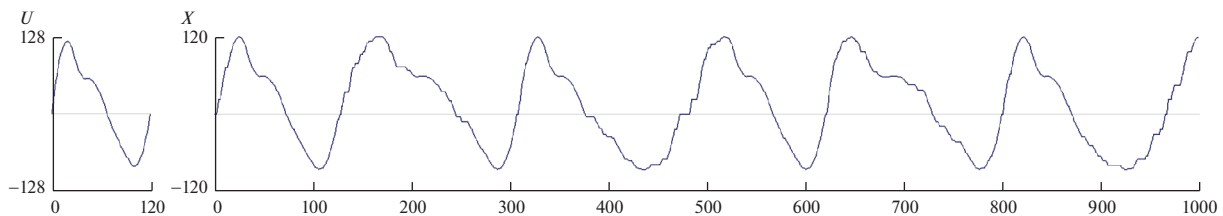
Чтобы проиллюстрировать данную интерпретацию, на фиг. 1 и 2 в виде графиков изображены входные данные задачи 1. Здесь и далее считаем, что номерам элементов последовательностей из алфавита и последовательности Y соответствуют дискретно-временные отсчеты непрерывных сигналов. На фиг. 1 приведен пример алфавита W , содержащего 6 последовательностей, на фиг. 2 — пример последовательности Y , порожденной одной из последовательностей алфавита.



Фиг. 1. Пример входных значений задачи 1. Алфавит W .



Фиг. 2. Пример входных значений задачи 1. Последовательность Y , подлежащая обработке.



Фиг. 3. Пример последовательности U из алфавита W и одной из последовательностей $X \in \mathcal{X}(U)$.

На фиг. 3а изображена последовательность U из алфавита W , на фиг. 3б – пример последовательности $X \in \mathcal{X}(U)$, построенной по правилам (1.7), (1.8), (1.9), (1.10) для некоторых допустимых значений $\{p_m\}$ и $\{k_i^{(m)}\}$. Ступенчатость сигнала на этой фигуре обусловлена кратными повторениями элементов из U .

2. БЛИЗКИЕ ПО ПОСТАНОВКЕ ЗАДАЧИ

Задача 1 является обобщением ранее исследованных задач распознавания. Частный случай задачи 1, в котором $q_1 = \dots = q_K = q$, $p_m = q$, $J^{(m)}(i) = i$, $m = 1, \dots, M$, исследовался в [1]. В [2] была рассмотрена модификация этого частного случая, когда M является частью входа задачи. В этом частном случае и его модификации, как и в задаче 1, требуется распознать квазипериодическую последовательность, но все повторы порождающей последовательности в ней идентичны. В [1] и [2] были построены алгоритмы, позволяющие получить оптимальное решение за время $\mathcal{O}(KT_{\max}N)$ и $\mathcal{O}(KMT_{\max}N)$. Другой частный случай, когда $K = 1$, был исследован в [3]. В этом частном случае алфавит состоит из единственной последовательности, поэтому задачу можно интерпретировать как задачу аппроксимации. В этой работе приведен точный алгоритм, позволяющий получить точное решение за время $\mathcal{O}(T_{\max}^3N)$.

Близкими в постановочном плане являются задачи распознавания квазипериодических последовательностей по усеченным данным [4] и [5]. В этих задачах, как и в задаче 1, предполагалось, что распознаваемая квазипериодическая последовательность включает в себя искаженные допустимым образом повторы некоторой последовательности, принадлежащей заданному мно-

жеству. В отличие от задачи 1, здесь предполагалось, что множество допустимых преобразований последовательности состоит из всевозможных ее подпоследовательностей, полученных удалением некоторого количества начальных и конечных отсчетов. Для этих задач в цитируемых работах получены точные полиномиальные алгоритмы для их решения.

Алгоритм решения задачи 1 является подходящим инструментом для решения прикладных задач распознавания и анализа сигналов, имеющих квазипериодическую структуру в виде флуктуирующих повторов участков сигнала. Распознавание и анализ таких сигналов актуальны для различных приложений, имеющих дело с обработкой импульсных сигналов, полученных от природных источников: биомедицинские, геофизические и т.п. Далее, в разделе численное моделирование, будут приведены примеры обработки медицинских сигналов, имеющих вид квазипериодических последовательностей (ECG, PPG).

3. РАЗМЕР МНОЖЕСТВА ДОПУСТИМЫХ РЕШЕНИЙ

Приведенная выше интерпретация позволяет заметить, что число допустимых решений задачи 1 совпадает с мощностью множества $\mathcal{X} = \bigcup_{U \in W} \mathcal{X}(U)$. Из комбинаторных соображений легко видеть, что при каждом $U \in W$ справедлива оценка

$$\begin{aligned} |\mathcal{X}(U)| &\leq (N - q(U) + 1) \sum_{M=1}^{M_{\max}(U)} q(U)^{(T_{\max} - q(U))M} (T_{\max} - q(U) + 1)^{2M-1} \leq \\ &\leq q(U)^{(T_{\max} - q(U))M_{\max}(U)} (N - q(U) + 1) (T_{\max} - q(U) + 1)^{2M_{\max}(U)-1} M_{\max}(U), \end{aligned}$$

где $M_{\max}(U) = \lfloor N/q(U) \rfloor$ – максимально возможное число повторов последовательности U в Y . Данная верхняя оценка позволяет оценить влияние, которое параметры задачи 1 оказывают на мощность множества допустимых решений.

С другой стороны, за исключением тривиального случая $T_{\max} = q(U)$ справедлива нижняя оценка

$$|\mathcal{X}(U)| \geq 2^{\lfloor \frac{N - q(U) + 1}{q(U) + 1} \rfloor}, \quad U \in W.$$

Отсюда следует, что

$$|\mathcal{X}| = \sum_{U \in W} |\mathcal{X}(U)| \geq K 2^{\lfloor \frac{N - q_{\max} + 1}{q_{\max} + 1} \rfloor},$$

где $q_{\max} = \max_{U \in W} q(U)$, а K – мощность алфавита последовательностей. Это означает, что если q_{\max} ограничено некоторой константой (что типично для приложений), мощность множества \mathcal{X} допустимых решений задачи 1 экспоненциально растет с ростом N . Очевидно, что перебрать напрямую элементы этого множества за приемлемое время вряд ли удастся, так как N – часть входа задачи. Несмотря на этот экспоненциальный рост, ниже будет приведен алгоритм, позволяющий получить оптимальное решение за полиномиальное время.

4. ОСНОВЫ АЛГОРИТМА

Для построения алгоритма решения задачи 1 нам потребуется следующая вспомогательная

Задача 2. Дано: числовые последовательности $Y = (y_1, \dots, y_N)$, $U = (u_1, \dots, u_{q(U)})$, и натуральные числа T_{\max}, ℓ . Найти: набор $\mathcal{M} = \{n_1, \dots, n_m, \dots\}$ номеров последовательности Y , набор $\mathcal{P} = \{p_1, \dots, p_m, \dots\}$ натуральных чисел, набор $\mathcal{J} = \{J^{(1)}, \dots, J^{(m)}, \dots\}$ сжимающих отображений, в котором $J^{(m)} : \{1, \dots, p_m\} \rightarrow \{1, \dots, q(U)\}$, а также размерность M этих наборов, которые минимизируют целевую функцию

$$G(\mathcal{M}, \mathcal{P}, \mathcal{J}) = F(\bullet | U) \tag{4.1}$$

при ограничениях (1.2), (1.3), (1.4) на элементы искомого набора \mathcal{M} , \mathcal{P} , и при ограничениях (1.5) на элементы искомого сжимающих отображений.

Для полноты изложения приведем алгоритм ее решения, базирующийся на результатах, полученных в [3]. Проанализировав ограничения (1.2), (1.3), (1.4), входящие в условия задачи 2, легко видеть, что справедливо

Утверждение 1. Пусть выполнены условия задачи 2, тогда для компонент наборов \mathcal{M} и \mathcal{P} справедливо:

1) $n_m \in \omega$, $m = 1, \dots, M$, где

$$\omega = \{1, \dots, N - q(U) + 1\};$$

2) $n_m \in \omega^+$, $m = 2, \dots, M$, где

$$\omega^+ = \{q(U) + 1, \dots, N - q(U) + 1\};$$

3) если $n_m = n$, $m = 1, \dots, M$, $n \in \omega$, то $p_m \in \delta(n)$, где

$$\delta(n) = \{q(U), \dots, \min\{\ell, N - n + 1\}\};$$

4) если $n_m = n$, $m = 2, \dots, M$, $n \in \omega^+$, то $n_{m-1} \in \gamma(n)$, где

$$\gamma(n) = \{k \mid \max\{n - T_{\max}, 1\} \leq k \leq n - q(U)\};$$

5) если $n_m = n$ и $n_{m-1} = j$, $m = 2, \dots, M$, $n \in \omega^+$, $j \in \gamma(n)$ то $p_{m-1} \in \theta(n, j)$, где

$$\theta(n, j) = \{q(U), \dots, \min\{\ell, n - j\}\}.$$

Опираясь на это утверждение, выпишем алгоритм решения задачи 2.

Алгоритм \mathcal{A}_1

Вход: Y , U , T_{\max} и ℓ .

Прямой ход алгоритма.

Шаг 1. Для каждого $n = 1, \dots, N - q(U) + 1$ выполним:

Для каждого $p = q(U), \dots, \min\{\ell, N - n + 1\}$ выполним:

(1) Вычислим

$$w_{s,t} = u_t^2 - 2y_{n+s-1}u_t, \quad s = 1, \dots, p, \quad t = 1, \dots, q(U).$$

(2) Найдем значение W^* по формулам

$$W^* = W_{p,q(U)},$$

$$W_{s,t} = \min\{W_{s-1,t}, W_{s-1,t-1}\} + w_{s,t}, \quad s = 1, \dots, p, \quad t = 1, \dots, q(U),$$

$$W_{s,t} = \begin{cases} 0, & s = 0, \quad t = 0, \\ +\infty, & s = 0, \quad t = 1, \dots, q(U), \\ +\infty, & s = 1, \dots, p, \quad t = 0, \end{cases}$$

и последовательность $J^*(1), \dots, J^*(p)$ по формулам

$$J^*(p) = q(U);$$

$$J^*(i-1) = \begin{cases} J^*(i), & \text{если } W_{i-1, J^*(i)} \leq W_{i-1, J^*(i)-1}, \\ J^*(i) - 1, & \text{если } W_{i-1, J^*(i)} > W_{i-1, J^*(i)-1}, \end{cases}$$

$$i = p, p-1, \dots, 2.$$

(3) Положим $W^*(n, p) = W^*$; $J^*(n, p) = \{J^*(i), i = 1, \dots, p\}$.

Шаг 2. Вычислим совокупность значений $G(n, p)$, $p = q(U), \dots, \ell$, $n = 1, \dots, N - p + 1$, по формулам

$$G(n, p) = \begin{cases} W^*(n, p), & n \in \omega, \quad p \in \delta(n), \\ \min\{0, \min_{j \in \gamma(n)} \min_{i \in \theta(n, j)} G(j, i)\} + W^*(n, p), & n \in \omega^+, \quad p \in \delta(n), \end{cases}$$

а так же значение G^* по формуле

$$G^* = \min_{n \in \omega} \min_{p \in \delta(n)} G(n, p).$$

Положим $F_A = G^*$.

Обратный ход алгоритма.

Шаг 3. Вычислим $\pi(n)$ и $I(n)$, $n = 1, \dots, N - q(U) + 1$, по формулам

$$I(n) = \begin{cases} 0, & \text{если } n \in \omega \setminus \omega^+, \\ 0, & \text{если } \min_{j \in \gamma(n)} \min_{i \in \theta(n,j)} G(j, i) \geq 0, \quad n \in \omega^+, \\ \arg \min_{j \in \gamma(n)} \left\{ \min_{i \in \theta(n,j)} G(j, i) \right\}, & \text{если } \min_{j \in \gamma(n)} \min_{i \in \theta(n,j)} G(j, i) < 0, \quad n \in \omega^+, \end{cases}$$

и

$$\pi(n) = \begin{cases} 0, & \text{если } I(n) = 0, \\ \arg \min_{i \in \theta(n, I(n))} G(I(n), i), & \text{если } I(n) > 0, \quad n \in \omega. \end{cases}$$

Шаг 4. Найдем компоненты вспомогательных наборов (π_1, \dots, π_M) и (v_1, \dots, v_M) и их размерность M по формулам

$$\begin{aligned} \pi_1 &= \arg \min_{p \in \{q(U), \dots, \ell\}} \left\{ \min_{n \in \{1, \dots, N-p+1\}} G(n, p) \right\}, \\ v_1 &= \arg \min_{n \in \{1, \dots, N-\pi_1+1\}} G(n, \pi_1), \\ \pi_m &= \pi(v_{m-1}), \quad v_m = I(v_{m-1}), \quad m = 2, \dots, M, \end{aligned}$$

где M – наименьшее значение m , при котором $\pi(v_m) = 0$.

Шаг 5. Положим: $M_{A_1} = M$, $\mathcal{M}_{A_1} = \{v_M, \dots, v_1\}$, $\mathcal{P}_{A_1} = \{\pi_M, \dots, \pi_1\}$, $\mathcal{J}_{A_1} = \{J^*(v_M, \pi_M), \dots, J^*(v_1, \pi_1)\}$.

Выход: M_{A_1} , \mathcal{M}_{A_1} , \mathcal{P}_{A_1} , \mathcal{J}_{A_1} и G_{A_1} .

Утверждение 2 (см. [3]). *Алгоритм \mathcal{A}_1 находит точное решение задачи 2 за время $\mathcal{O}(T_{\max}^3 N)$.*

5. АЛГОРИТМ

Опираясь на вспомогательную задачу 2 и утверждение 2, сформулируем алгоритм решения задачи 1.

Алгоритм \mathcal{A}

Вход: Y, W, T_{\max} и ℓ .

Шаг 1. Для каждого $U \in W$ выполним алгоритм \mathcal{A}_1 с входными данными Y, U, T_{\max} и ℓ . Положим $M(U) = M_{A_1}$, $\mathcal{M}(U) = \mathcal{M}_{A_1}$, $\mathcal{P}(U) = \mathcal{P}_{A_1}$, $\mathcal{J}(U) = \mathcal{J}_{A_1}$, $G_U = G_{A_1}$.

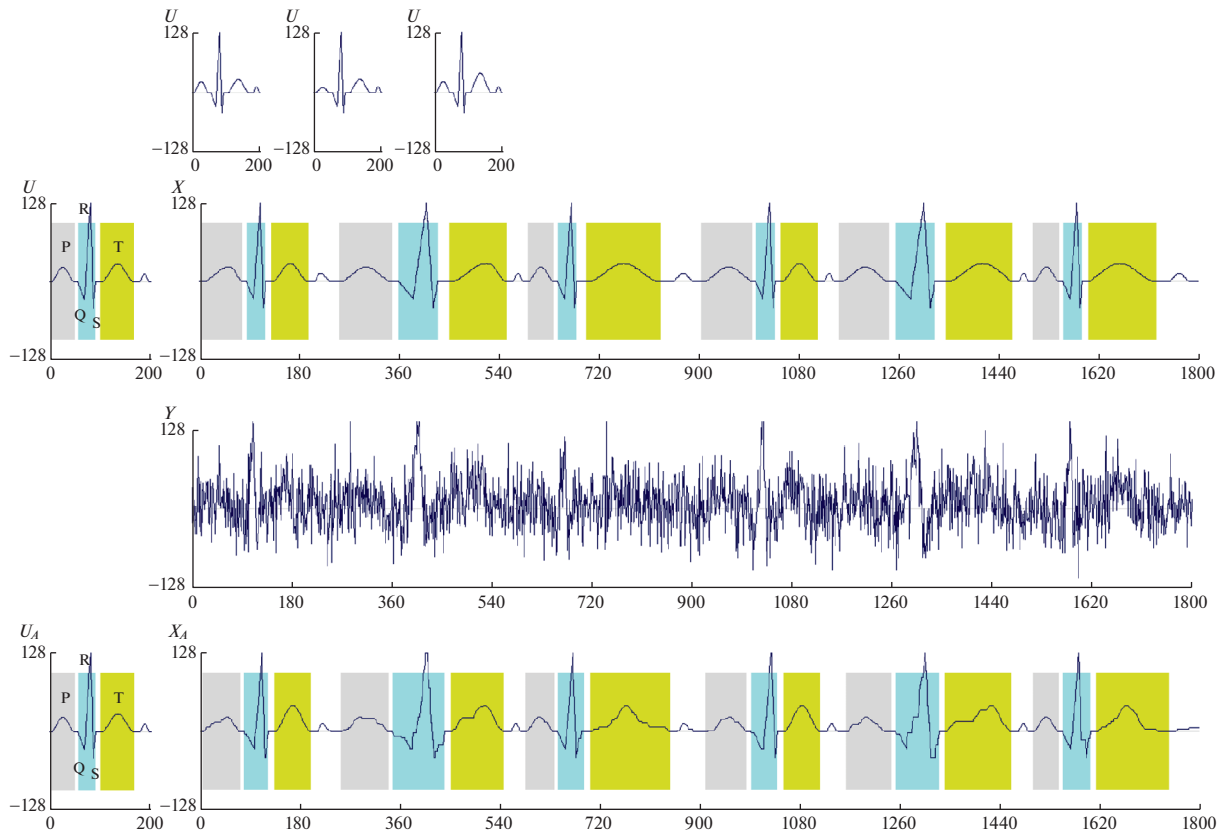
Шаг 2. Вычислим

$$\begin{aligned} F_A &= \min_{U \in W} G_U, \\ U_A &= \arg \min_{U \in W} G_U. \end{aligned} \tag{5.1}$$

Шаг 3. Положим $M_A = M(U_A)$, $\mathcal{M}_A = \mathcal{M}(U_A)$, $\mathcal{P}_A = \mathcal{P}(U_A)$, $\mathcal{J}_A = \mathcal{J}(U_A)$.

Выход: $U_A, M_A, \mathcal{M}_A, \mathcal{P}_A, \mathcal{J}_A$ и F_A .

Теорема 1. *Алгоритм \mathcal{A} находит точное решение задачи 1 за время $\mathcal{O}(KT_{\max}^3 N)$.*



Фиг. 4. Пример 1. Распознавание ECG-подобной последовательности.

Доказательство. Оптимальность решения, найденного алгоритмом \mathcal{A} , следует из Утверждения 2 и следующей цепочки равенств:

$$\begin{aligned}
 F^* &\stackrel{1}{=} \min_{U, \mathcal{M}, \mathcal{P}, \mathcal{J}} F(U, \mathcal{M}, \mathcal{P}, \mathcal{J}) \stackrel{2}{=} \min_U \min_{\mathcal{M}, \mathcal{P}, \mathcal{J}} F(U, \mathcal{M}, \mathcal{P}, \mathcal{J}) \stackrel{3}{=} \\
 &\stackrel{4}{=} \min_U \left\{ \min_{\mathcal{M}, \mathcal{P}, \mathcal{J}} F(U, \mathcal{M}, \mathcal{P}, \mathcal{J} | U) \right\} \stackrel{5}{=} \min_U \min_{\mathcal{M}, \mathcal{P}, \mathcal{J}} G(\mathcal{M}, \mathcal{P}, \mathcal{J}) \stackrel{6}{=} \min_U G_U = F_{\mathcal{A}}.
 \end{aligned}$$

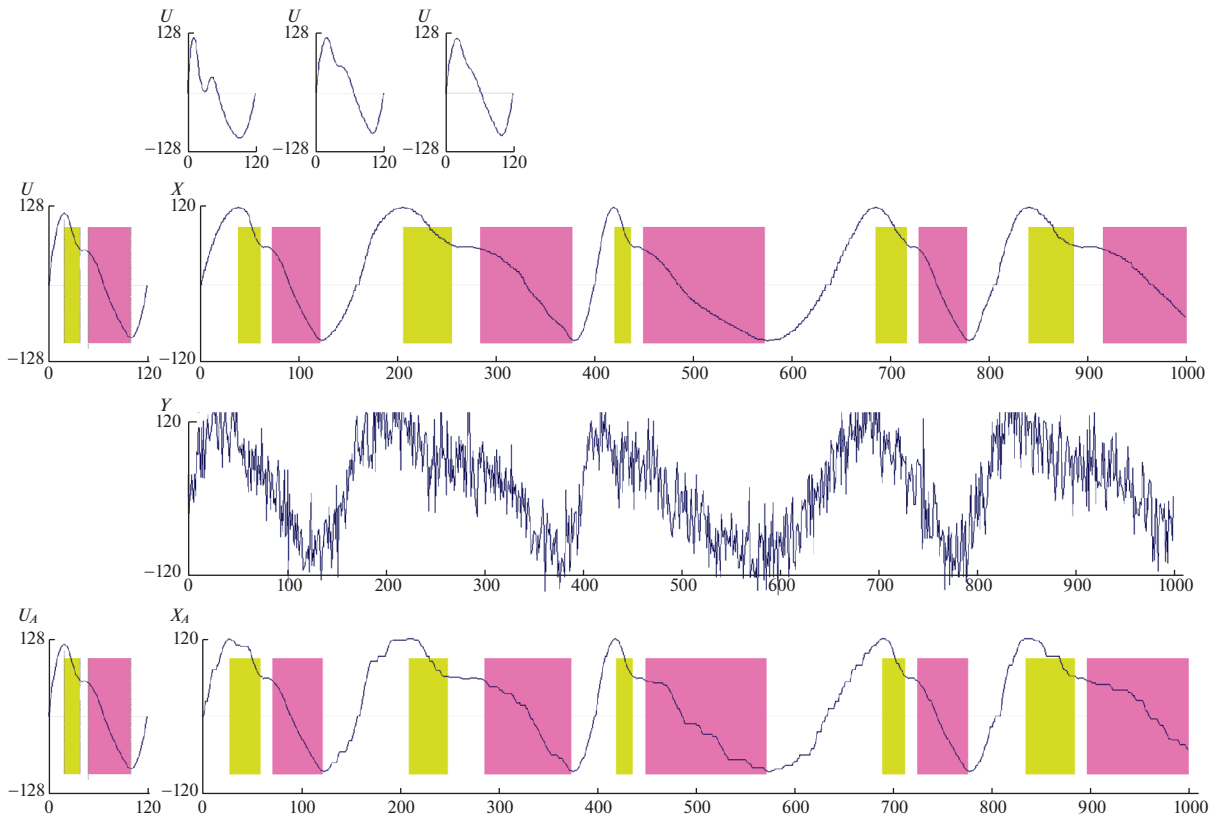
В этой цепочке равенство 1 – определение оптимального решения задачи 1. Равенство 2 следует из вида целевой функции (1.1). Равенство 3 очевидно. Равенство 4 следует из формулы (4.1), равенство 5 – определение оптимального решения задачи 2 при каждом фиксированном значении U . Наконец, равенство 6 следует из формулы (5.1) в пошаговой записи алгоритма \mathcal{A} .

В соответствии с утверждением 2, время работы алгоритма \mathcal{A}_1 оценивается величиной $\mathcal{O}(T_{\max}^3 N)$. На шаге 1 алгоритма \mathcal{A} требуется K раз выполнить алгоритм \mathcal{A}_1 , поэтому трудоемкость шага 1 алгоритма \mathcal{A} есть величина $\mathcal{O}(KT_{\max}^3 N)$. Шаг 2, очевидно, требует $\mathcal{O}(K)$ операций, а шаг 3 – константного числа операций. Суммируя временные затраты на всех шагах алгоритма, получаем итоговую трудоемкость алгоритма \mathcal{A} . Теорема доказана.

Замечание 1. Если T_{\max} является частью входа задачи, а мощность K алфавита – фиксированный параметр, то время работы алгоритма равно $\mathcal{O}(N^4)$. Таким образом, алгоритм \mathcal{A} является полиномиальным.

6. ПРИМЕРЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ

Для иллюстрации работоспособности алгоритма приведем два примера обработки последовательностей (временных рядов), которые можно интерпретировать как квазипериодические последовательности флуктуирующих ECG-подобных (фиг. 4) и PPG-подобных (фиг. 5) импульсов



Фиг. 5. Пример 2. Распознавание PPG-подобной последовательности.

при наличии аддитивной помехи. В действительности, с математической точки зрения не имеет значения, какие именно последовательности включены в состав алфавита W . Выбор именно ECG и PPG-подобных сигналов обусловлен желанием проиллюстрировать применимость алгоритма для биомедицинских приложений. Форма этих импульсов, а так же характерные участки и значимые точки описаны экспертами, см., например, [6]–[9]. На фиг. 4 (ECG-импульс) этими участками являются P-Q-R-S-T-U, на фиг. 5 (PPG-импульс) – систолическая точка, диастолическая точка, диастолическая точка. Раскраска фигур маркирует характерные участки или интервалы между значимыми точками.

Рассмотрим подробнее фиг. 4, иллюстрирующую помехоустойчивую обработку ECG-подобного сигнала. В верхнем ряду изображен алфавит. Ниже слева изображена последовательность U – одна из последовательностей из алфавита. Справа от нее изображена программно-генерированная последовательность – квазипериодическая последовательность, порожденная флуктуирующими повторами импульса U . Под модельной последовательностью изображена последовательность Y – результат сложения модельной последовательности и последовательности независимых одинаково распределенных гауссовских случайных величин с нулевым математическим ожиданием. Важно отметить, что только Y и W являются входными данными для алгоритма. Последовательность U и модельная последовательность X , изображенные во втором ряду фигуры, приведены для иллюстрации; эти данные недоступны.

В нижней части фигуры представлены результаты работы алгоритма \mathcal{A} : слева последовательности U_A , которая является оценкой для модельной, ненаблюдаемой последовательности. Справа изображены компоненты последовательности X_A , восстановленной по формулам (1.7) и (1.8) с использованием четырех наборов U_A , M_A , \mathcal{P}_A и \mathcal{F}_A , полученных в результате работы алгоритма \mathcal{A} . Данные на фиг. 4 получены при $K = 3$, $q(U) = 203$, $U \in W$, $T_{\max} = 370$, $N = 1800$, максимальная амплитуда сигнала – 128, уровень шума $\sigma = 35$.

Фигура 5 показывает результаты обработки PPG-подобного сигнала. Эта фигура имеет такую же структуру, как и фиг. 4. Данные на фиг. 5 получены при $K = 3$, $q(U) = 120$, $U \in W$, $T_{\max} = 240$, $N = 1000$, максимальная амплитуда сигнала – 128, уровень шума $\sigma = 35$.

Приведенные примеры показывают, что построенный алгоритм позволяет с вполне приемлемым качеством обрабатывать данные в виде квазипериодических последовательностей флуктуирующих импульсов. Во-первых, неизвестная последовательность U и определенная алгоритмом последовательность U_A совпадают в обоих примерах (распознавание произведено корректно). Во-вторых, визуальное сравнение графиков ненаблюдаемой последовательности X и восстановленной последовательности X_A демонстрирует лишь незначительные отклонения одного графика от другого и практически точное совпадение всех характерных точек.

ЗАКЛЮЧЕНИЕ

В работе показано, что одна из неисследованных задач дискретной оптимизации полиномиально разрешима. Полиномиальная разрешимость доказана конструктивно, т.е. построен алгоритм, гарантирующий получение оптимального решения задачи и получена полиномиальная оценка его трудоемкости.

Результаты численного моделирования продемонстрировали, что предложенный алгоритм может служить подходящим инструментом для решения задач помехоустойчивого распознавания и анализа квазипериодических импульсных последовательностей. Приведены примеры обработки ECG и PPG-подобных сигналов.

Остается неизученной модификация задачи 1, в которой число суммируемых сверток является частью входа задачи. Значительный математический интерес так же представляет дискретная экстремальная задача, когда алфавит последовательностей не задан, т.е. требуется по входной последовательности Y распознать последовательность U как элемент континуального множества числовых последовательностей, имеющих фиксированную конечную длину. Исследование этих задач представляется делом ближайшей перспективы.

СПИСОК ЛИТЕРАТУРЫ

1. *Kel'manov A.V., Khamidullin S.A., Okol'nishnikova L.V.* Recognition of a quasiperiodic sequence containing identical subsequences-fragments // *Pattern Recognition and Image Analysis*. 2004. V. 14. № 1. P. 72–83.
2. *Kel'manov A.V., Khamidullin S.A.* Recognizing a quasiperiodic sequence composed of a given number of identical subsequences // *Pattern Recognition and Image Analysis*. 2000. V. 10. № 1. P. 127–142.
3. *Kel'manov A.V., Khamidullin S.A., Mikhailova L.V., Ruzankin P.S.* Polynomial-time solvability of one optimization problem induced by processing and analyzing quasiperiodic ECG and PPG signals // *Communications in Computer and Information Science*. 2020. V. CCIS 1145. P. 88–101.
4. *Kel'manov A.V., Khamidullin S.A.* Algorithm of recognition of a quasiperiodic sequence composed of a given number of truncated subsequences // *Pattern Recognition and Image Analysis*. 2001. V. 11. № 1. P. 43–46.
5. *Кельманов А.В., Хамидуллин С.А.* Распознавание числовой последовательности по фрагментам квазипериодически повторяющейся эталонной последовательности // *Сиб. ж. индустр. матем.* 2004. Т. 7. № 2. С. 68–87.
6. *Rajni R., Kaur I.* Electrocardiogram signal analysis – an overview // *Int. J. Comput. Appl.* 2013. V. 84. № 7. P. 22–25.
7. *Al-Ani M.S.* ECG Waveform Classification Based on P-QRS-T Wave Recognition // *UHD Journal of Science and Technology*. 2018. V. 2. № 2.
8. *Shelley K., Shelley S.* Pulse Oximeter Waveform: Photoelectric Plethysmography. In: Carol Lake, Hines R., Blitt C. (eds). *Clinical Monitoring*. W.B. Saunders Company. 2001. P. 420–428.
9. *Elgendi M.* On the analysis of fingertip photoplethysmogram signals // *Current Cardiology Reviews*. 2012. V. 8. № 1. P. 14–25.