ГІБРИДНИЙ (КОМБІНОВАНИЙ) АЛГОРИТМ РОЗВ'ЯЗАННЯ ЗАДАЧІ РОЗПІЗНАВАННЯ МОВНИХ СИГНАЛІВ

Надія ТИМОФІЄВА
МННЦ ІТіС НАНУ та МОН України 03022, Київ, просп. Ак. Глушкова, 40,
тел.: (044) 502 53 32, E-mail: Nadija_Tymofijeva@uasoiro.org.ua

Abstract

A problem of speech recognition is presented as a problem of combinatorial optimization. It is shown that for it an objective function depends on several variables, which are the combinatorial configurations of different types: the n-element set partitions into subsets, the placements, the combinations without repetitions. For these definitions, the problem of speech recognition is divided into three subproblems. For a decision of each subproblem, an independent algorithm is constructed. A scheme of the structurization of a library of standard signals is described. The use of this structurization permits the dictionary look-up of the word with direction searching. The problem of speech recognition is conducted by hybrid algorithm, which is based on organization of iterative process which generates the sequence of solutions in accordance with built-in algorithms.

1. Вступ

Останні десятиліття одним із основних підходів в комбінаторній оптимізації є дослідження і розробка метаевристичних алгоритмів, що грунтуються на організації покрокових обчислень – ітераційного процесу, який породжує послідовність розв'язків у відповідності із вбудованими процедурами. Якщо вбудовані процедури – незалежні алгоритми, орієнтовані на розв'язання певних задач, то такі алгоритми називаються гібридними або комбінованими [1-3].

При розробці гібридних алгоритмів, задача що розглядається, розбивається на підзадачі, для розв'язання кожної з яких розробляється новий алгоритм або використовується відомий. При цьому, як правило, проводиться аналіз конкретної задачі з метою виявлення певних ознак, за якими необхідно проводити таке розбиття. Але в задачах комбінаторної оптимізації аргумент цільової функції досить часто залежить від кількох змінних, якими є різні типи комбінаторних конфігурацій. Тому за цією ознакою вони розділяються на підзадачі природно. Знання природи задач цього класу дозволяє використовувати для їх розв'язання відомі і нові алгоритми досить просто.

Якщо провести аналіз задачі розпізнавання мовних сигналів, то можна побачити, що вона розділяється на кілька підзадач, кожна з яких розв'язується розробленим спеціально для неї незалежним алгоритмом. Нижче розглянемо її як задачу комбінаторної оптимізації з метою виявлення тих ознак, за якими вона розділяється на підзадачі.

Спочатку розглянемо основні положення, що стосуються комбінаторної оптимізації.

2. Основні положення з комбінаторної оптимізації

Комбінаторна оптимізація – область математики, предметом якої є дослідження і розв'язання екстремальних задач на скінченній множині комбінаторного характеру [4].

Математичне моделювання задач комбінаторної оптимізації включає в себе змістовну постановку і побудову їх математичних моделей [5]. Для побудови конкретної математичної моделі необхідно визначити аргумент цільової функції (тип комбінаторної конфігурації), задати критерії, по яких вона оцінюється, і увести обмеження.

Задачі комбінаторної оптимізації, як правило, задаються на одній або кількох множинах, елементи яких мають будь-яку природу. Назвемо їх базовими. Для багатьох задач кожну з цих множин можна подати у вигляді графа, вершинами якого є її елементи, а кожному ребру поставлено у відповідність число $c_{ij} \in R$, яке називають вагою ребра; R – множина дійсних чисел. Для зручності в подальшому вважатимемо, що між елементами цих множин існують зв'язки, числове значення яких названо вагами. Для деяких задач між елементами заданих множин зв'язків не існує, але вагами виступають числа $c_{ij} \in R$, яким у відповідність поставлена деяка властивість цих елементів. Задачі комбінаторної оптимізації розділяються за способом задання вхідної інформації, яка описує властивості елементів базових множин і представлена множиною дійсних чисел або визначає зв'язки (ваги) між елементами однієї або кількох базових множин. Ці величини визначають значення цільової функції. В подальшому назвемо їх вхідними даними. Кількість таких множин і спосіб задання вхідних даних характеризують особливість будь-якої з цих задач.

В загальному вигляді задача комбінаторної оптимізації формулюється так. Задано скінченну множину A, між елементами яких існують зв'язки або ці елементи мають певні властивості. З елементів $a_j \in A$ утворюється скінченна множина комбінаторного характеру W – сукупність усіх комбінаторних об'єктів (комбінаторних конфігурацій) певного типу (перестановки, вибірки різних типів, робиття тощо). Необхідно знайти елемент w^* підмножини $W_0 \subseteq W \subset \overline{W}$, для якого функціонал $F(w^*)$ набуває екстремального значення при виконанні заданих умов чи обмежень, тобто

$$F(w^*) = \underset{w \in W_0 \subseteq W \subset \overline{W}}{\operatorname{ext}} F(w),$$

де $\overline{W}$ – множина, яка включає різноманітні типи комбінаторних конфігурацій. Функціонал $F(w)$ називають цільовою функцією. Для багатьох часткових випадків уводиться умова $F(w^*) \le b$, де b – задане дійсне додатне число. Якщо цільова функція залежить від кількох змінних, якими є різні типи комбінаторних об'єктів, то задача комбінаторної оптимізації полягає у знаходженні змінних $w^* \in W_0 \subseteq W \subset \overline{W}$, $w^{'**} \in W_0^{'} \subseteq \subseteq W^{'} \subset \overline{W}$, для яких

$$F(w^*, w^{'**}) = \underset{\substack{w \in W_0 \subseteq W \subset \overline{W} \\ w^{'} \in W_0^{'} \subseteq W^{'} \subset \overline{W}}}{\operatorname{ext}} F(w, w^{'}).$$

Побудуємо математичну модель задачі комбінаторної оптимізації, в якій між елементами заданих множин існують зв'язки, а вхідні дані задаються матрицями, одна з яких комбінаторна. Елементи комбінаторної матриці $Q(w^k)$ для багатьох задач набувають значення (0,1), а $g_{st}^k(w^k) = 1$, якщо a_s, a_t мають між собою зв'язки, і $g_{st}^k(w^k) = 0$ в іншому випадку, w^k – комбінаторна конфігурація (перестановка, розбиття n –елементної множини на підмножини тощо), яка є аргументом цільової функції, $k \in \{1,..., q\}$ – порядковий номер w^k у множині W, q – кількість елементів у W. Елементи матриці C задають кількість зв'язків між a_s, a_t заданої множини $A = \{a_1,..., a_n\}$. Послідовність елементів $Q(w^k)$ задамо функцією натурального аргументу $f(j)|_1^m$, а матриці C – функцією натурального аргументу $\varphi(j)|_1^m$. Комбінаторна функція $\beta(f(j), w^k)|_1^m$ – це $f(j)|_1^m$, яка змінюється в залежності від w^k, де значення m залежить від конкретної задачі. Цільова функція набуде вигляду:

$$F(w^k) = \sum_{j=1}^{m} \beta_j (f(j), w^k) \varphi(j).$$

Якщо аргумент цільової функції – перестановка або підмножина ізомофних комбінаторних конфігурацій, то комбінаторною матрицею може бути будь-яка із заданих.

2. Основні положення з розпізнавання мовних сигналів

Подана нижче математична модель задачі розпізнавання мовних сигналів описана у [6].

Мовний сигнал у [6] описується послідовністю $X_J = (x_1, ..., x_J)$, елемент x_j якої є значення сигналу у відліку j. Послідовність X_J одержують в результаті попередньої обробки сигналу, що поступає на мікрофон, з метою зменшення об'єму інформації. Довжина J різних реалізацій різна. Для розпізнавання реалізацій X_J створюється словник еталонних слів. Еталон слова словника описується послідовністю $E_h = (e_{h1},.., e_{hq_h})$, де h – номер слова у словнику, q_h – довжина сигналу еталона слова, $h \in \{1,..., K\}$, K – кількість слів-еталонів.

Задача розпізнавання мовних сигналів полягає у знаходженні для сигналу X_J найбільш правдоподібного еталону E_h з усіх можливих еталонних сигналів [6]. Так, як в цій задачі установлюється подібність сигналів, то в процесі обчислення інтегральної міри подібності знаходиться максимум функції

$$G_h(X_J) = \max_{v \in \tau_h(J)} G(X_J, vE_h), \qquad (1)$$

де $v \in \tau_h(J)$ – нелінійне розтягання початкового сигналу, $\tau_h(J)$ – множина можливих перетворень v початкового еталону E_h, $G(X_J, vE_h)$ – інтегральна міра подібності, $G_h(X_J)$ –значення інтегральної міри подібності [6].

Задачу (1) розв'язують шляхом порівняння еталону E_h із сигналом X_J одним із методів направленого перебору, наприклад методом динамічного програмування [6]. Для знаходження сигналу за номером h із усіх еталонних, якому відповідає вхідний сигнал, розв'язується шляхом перебору задача

$$h(X_J) = \arg\max_h G_k(X_J). \qquad (3)$$

Величина $h(X_J)$ повинна належати заданій області значень. Аргументом цільової функції в цих задачах, згідно з [6], є вхідний сигнал.

Як видно з математичної моделі (1)–(2), задача розпізнавання мовних сигналів досить природно розділяється на дві підзадачі: перебір еталонних сигналів і порівняння еталонного і вхідного сигналів. Тому для її розв'язання використовується гібридний алгоритм. Так, як тут має місце перебір варіантів, то вона відноситься до задач комбінаторної оптимізації.

У наступному параграфі визначимо тип комбінаторної конфігурації (аргумент цільової функції) і побудуємо математичну модель задачі розпізнавання як задачу комбінаторної оптимізації.

3. Представлення задачі розпізнавання мовних сигналів як задачі комбінаторної оптимізації

Розглянемо задачу порівняння еталонного і вхідного сигналів (задача (1)). Уведемо базову множину $A = \{a_1, ..., a_n\}$, де $a_j \in X_J$, $j = \overline{1, J}$, а $a_{J+i} \in E_h$, $i = \overline{1, q_h}$, $n = J + q_h$. Вхідні дані, якими є ваги між елементами $x_j \in X_J$ і $e_{h_i} \in E_h$ задамо несиметричною матрицею $C = \left\| c_{ij} \right\|_{q_h \, x \, J}$, номера стовпців якої збігаються з нумерацією елементів $x_j \in X_J$, $j = \overline{1, J}$, а номера рядків – з нумерацією елементів $e_{h_i} \in E_h$, $i = \overline{1, q_h}$. Аналогічне представлення вхідних даних подано розгорнутим графом слова у [6]. Як описано у [6], при поелементному розпізнаванні мовного сигналу для елемента $x_j \in X_J$ знаходиться йому подібний $e_{h_i} \in E_h$. Так, як з базової множини $A = \{a_1, ..., a_n\}$ вибираються почергово по два елементи у строгому порядку, то отримана комбінаторна конфігурація є розміщення без повторень. Позначимо її $\mu^k \in M$, M – їх всіляка множина. Для визначення елементів a_j, що вибираються з базової бібліотеки на k-му варіанті розв'язку задачі, уведемо комбінаторну (0,1)–матрицю $Q(\mu^k) = \left\| g_{ij}^k(\mu^k) \right\|_{q_h \, x \, J}$. Якщо $g_{ij}^k(\mu^k) = 1$, то з множини A вибрана пара (x_j, e_{h_i}), в іншому разі– значення $g_{ij}^k(\mu^k) = 0$.

Елементи матриці C подамо числовою функцією $\varphi(j)\,|_1^{n^*}$, а матриці $Q(\mu^k)$ – комбінаторною $\beta(\,f(j), \mu^k)|_1^{n^*}$, де $n^* = J \cdot q_h$.

Задача порівняння еталонного і вхідного мовних сигналів полягає у знаходженні такого розміщення без повторень $\mu^{k*} = (\mu_1^{k*}, ..., \mu_{q'}^{k*})$, для якого цільова функція

$$F(\mu^{k*}) = \max_{\mu^k \in M} \sum_{j=1}^{n^*} \varphi(j)\,\beta_j\,(f(j), \mu^k), \quad (3)$$

де $\sum_{j=1}^{n^*} \varphi(j)\,\beta_j\,(f(j), \mu^k)$ – інтегральна міра подібності, а $\varphi(j) = g_j'(x_j, e_{h_i})$ – елементарна міра подібності, яка визначає подібність між елементами еталонного і вхідного сигналів $q' = \min(J, q_h)$.

Отже, аргументом цільової функції задачі (1) є розміщення без повторень. Розглянемо задачу (2).

Позначимо $A = \{a_1, ..., a_n\}$ базову множину, де $a_1 \in X_J$, а $a_i \in E_{i-1}$, $i = \overline{2, n}$. В цій задачі як ваги між еталонним і вхідним сигналами виступають значення інтегральних мір подібності, одержаних згідно з виразом (3) і представлених матрицею C'. Номера стовпців цієї матриці збігаються з номерами еталонних сигналів, розміщених у бібліотеці. Рядок у ній один і відповідає номеру один вхідного сигналу. Так, як при порівнянні вхідного і еталонного сигналів з базової множини A вибираються два елементи, то утворений об'єкт є сполучення без повторень. Позначимо його $\mu'^k \in M'$, де M' – їх всіляка множина. Уведемо комбінаторну (0,1)–матрицю $Q(\mu'^k) = \left\| g_{1i}'^k(\mu'^k) \right\|_{1 \, x \, n}$. Якщо $g_{1i}'^k(\mu'^k) = 1$, то з множини A вибрана пара (a_1, a_j), в іншому разі – значення $g_{1i}'^k(\mu'^k) = 0$.

Елементи матриці C' подамо числовою функцією $\varphi'(j)\,|_1^{n-1}$, а матриці $Q(\mu'^k)$ – комбінаторною $\beta'(\,f'(j), \mu'^k)|_1^{n-1}$.

Задача пошуку еталонного сигналу, який відповідає вхідному, полягає у знаходженні такого сполучення без повторень $\mu'^{k*} = (a_1, a_i)$ із n елементів a_j по 2, для якого значення заданої цільової функції було б найбільшим, тобто

$$F(\mu'^{k*}) = \max_{\mu'^k \in M} \sum_{j=1}^{n-1} \varphi'(j)\,\beta_j'(f'(j), \mu'^k),$$

де $\varphi'(j) = \sum_{j=1}^{n^*} \varphi(j)\,\beta_j\,(f(j), \mu^k)$.

4. Гібридний алгоритм розпізнавання мовних сигналів

Виходячи з викладеного, задача розпізнавання мовних сигналів розділяється на дві підзадачі, аргументом цільової функції в одній є розміщення без повторень, а у другій – сполучення без повторень. Як видно з постановки задачі (2) пошук еталонного сигналу, подібного вхідному, потребує прямого перебору. Для розв'язання цієї задачі направленим перебором розглянемо підзадачу структуризації бібліотеки еталонних сигналів за певними ознаками. Упорядкуємо ці сигнали, що відповідають заданим словам, в алфавітному порядку за такою схемою [7].

1) З кожного бібліотечного сигналу виділимо сегмент постійної довжини q'', які є початком сигналів еталонних слів так, щоб вони відповідали частині фонеми першого звуку. Множину одержаних сегментів позначимо $A = \{a_1, ..., a_n\}$, а множину слів у словнику позначимо $B = \{b_1, ..., b_n\}$. Елементу $a_j \in A$ відповідає сегмент частини першої фонеми слова, якому відповідає елемент b_j словника.

2) розв'язавши задачу розбиття множини A на підмножини (кластеризацію), об'єднаємо однорідні сегменти в одну підмножину $\rho_s^k \subset \rho^k \subset \Theta$. Підмножина $\rho_s^{'k} \subset \rho^{'k}$ є підмножиною слів словника $B = \{b_1, ..., b_n\}$ з подібними початковими сегментами і ізоморфна підмножині $\rho_s^k \subset \rho^k$, $s \in \{1, ..., \eta^k\}$, η^k – кількість підмножин $\rho_s^k \subset \rho^k$, $\rho_s^{'k} \subset \rho^{'k}$. Як і в задачі розпізнавання, в цьому випадку значення функції $\varphi(j) = \sum_{j=1}^{q''} g_j''(a_{js}, a_{jt})$ – інтегральна міра подібності, а $g_j''(a_{js}, a_{jt})$ – елементарна міра подібності, яка встановлюється між сегментами $a_r, a_t \subset A$, $a_{jr} \in a_r$, $a_{jt} \in a_t$.

3) кожній одержаній підмножині $\rho_s^k \subset \rho^k$ поставимо у відповідність еталон сегмента a_j', який відповідає частині першої фонеми слова, що входить до $\rho_s^{'k} \subset \rho^{'k}$. Одержану множину сегментів позначимо $A' = \{a_1', ..., a_n'\}$. Аналогічно можна структуризувати бібліотеку еталонних сигналів по другій, третій фонемі.

Маючи еталони сегментів, упорядковані в алфавітному порядку, задача (1)–(2) розв'язується таким чином. При пошуку еталонного сигналу у бібліотеці вирізаємо сегмент вхідного сигналу X_J що відповідає частині першої фонеми довжиною q''. Задачу (1) розв'язуємо з використанням відомих методів, наприклад, методу динамічного програмування [6]. При цьому порівнюється сегмент вхідного сигналу довжиною q'' з еталонними сегментами структуризованої бібліотеки. Якщо значення функції (1) найбільше для підмножини $\rho_s^k \subset \rho^k$, то пошук вхідного слова проводиться у цій підмножині словника B по другій або третій фонемах.

Отже, задача розпізнавання мовних сигналів розділяється на три підзадачі, аргументом цільової функції в яких є комбінаторні конфігурації різних типів. Ця задача полягає у знаходженні таких комбінаторних конфігурацій $\rho^{i*} \in \Theta$, $\mu^{t*} \in M$, $\mu^{'k*} \in M'$, для яких задана цільова функція набуває екстремального значення, тобто

$$F(\rho^{i*}, \mu^{t*}, \mu^{'k*}) = \underset{\substack{\rho^i \in \Theta \\ \mu^t \in M \\ \mu^{'k} \in M'}}{\operatorname{ext}} F(\rho^i, \mu^t, \mu^{'k}),$$

де Θ – всіляка множина розбиттів n-елементної множини на підмножини. Структуризація бібліотеки еталонних сигналів проводиться один раз. Для розпізнавання вхідного сигналу організується ітераційний процес, на кожному кроці якого почергово розв'язуються задачі (1)–(2).

Саме ті задачі, цільова функція в яких залежить від кількох змінних (комбінаторних конфігурацій різних типів), вимагають для свого розв'язку розроблення гібридних алгоритмів, тобто для розв'язання кожної з підзадач розробляється свій алгоритм, які гармонійно доповнюють одне одного.

Література

1.Корбут А.А., Сигал И.Х., Финкельштейн Ю.Н. Гибридные алгоритмы в дискретной оптимизации // Изв. АН СССР. Техн. кибернетика.– 1986.– № 1. – С. 65–77.

2. Гуляницкий Л.Ф. Разработка гибридных методов дискретной оптимизации на основе G-алгоритмов // Компьютерная математика. – К.: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2005.– № 1.– С. 143–151

3. Гуляницкий Л.Ф., Тимофеева Н.К. О размещении разногабаритных элементов на печатных платах // УСИМ. – 1982. – N 3. – С. 50–53.

4. Пападимитриу Х., Стайглиц К. Комбинаторная оптимизация. Алгоритмы и сложность. Пер. с англ. – М.: Мир, 1985.– 510 с.

5. Глушков В.М. Введение в АСУ.– К.: Техніка, 1974. –320 с.

6. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов.– К.: Наукова думка, 1987. – 262 с.

7. Тимофієва Н.К. Комбінаторика в розпізнаванні мовних сигналів // Мат. десятої міжнар. наук. конф. імені академіка М.Кравчука (Київ 04).– К: Націон. техн. Ун-т України (КПІ). – 2004. – С. 528.