DOI https://doi.org/10.15407/csc.2022.01.015
УДК 004.93'1:519.157

М.І. ШЛЕЗІНГЕР, доктор фіз.-мат. наук, професор, гол. наук. співробітник, Міжнародний науково-навчальний центр інформаційних технологій і систем НАН і МОН України, 03187, м. Київ, просп. Академіка Глушкова, 40, Україна,
schles@irtc.org.ua

РОЗВ'ЯЗОК ОПТИМІЗАЦІЙНИХ ЗАДАЧ СТРУКТУРНОГО РОЗПІЗНАВАННЯ НА ОСНОВІ ЇХНЬОЇ РЕПАРАМЕТРИЗАЦІЇ

У статті наведено ключові ідеї нового наукового напряму, що сформувався внаслідок об'єднання досліджень проблем розпізнавання образів і проблем несуперечності обмежень. Ці наукові напрями традиційно зараховано до проблематики штучного інтелекту, але формалізують відмінні один від одного аспекти інтелектуальної активності. Створення спільної формальної схеми, що об'єднує обидва напрями, розширює та конкретизує поняття машинного мислення, на формалізацію якого ці напрями спрямовані, і викликає необхідність розроблення нових і вдосконалення відомих математичних методів оптимізації.

Ключові слова: *розпізнавання образів, несуперечність обмежень, динамічне програмування, супермодулярна максимізація, репараметризація задач оптимізації.*

Вступ

Двадцятип'ятиріччя роботи Центру, якому присв'ячено цей випуск, збігається з періодом формування нового магістрального напряму в науці про розпізнавання образів, яке стало результатом досліджень у багатьох десятках колективів усього світу. Дослідження в кожному з цих колективів мають своє забарвлення та своєрідність, і тому напрям, що сформувався, ще не отримав якоїсь усталеної назви й відомий в англомовній літературі за ключовими словами, такими як *Energy Minimization Methods, Graphical Models, Hidden Random Field, MinCut-MaxFlow* і, можливо, іншими. Внесок Центру у формування цього напряму полягає у створенні формальних моделей процесів мислення, що усвідомлено чи неусвідомлено виконує людина або інші живі істоти. Саме так було сформульовано мету Державної науково-технічної програми "Образний комп'ютер", у рамках якої було отримано основний внесок Центру у новий науковий напрям, названий теорією образного мислення у штучно створених механізмах, зокрема, в електронних пристроях [1, 2].

У статті наводяться ключові ідеї цього нового напряму, утвореного внаслідок об'єднання досліджень проблем розпізнавання образів [3] і проблем несуперечності обмежень [4] у рамках спільної формальної системи. Ці напрями мають багату передісторію незалежного один від одного розвитку і традиційно зараховуються до проблеми штучного інтелекту, але формалізують істотно відмінні аспекти інтелектуальної активності. Створення спільної формальної схеми, що об'єднує ці напрями, розширює та конкретизує поняття машинного мислення, на формалізацію якого ці напрями спрямовані, тож виникає необхідність розроблення нових і вдосконалення відомих математичних методів оптимізації.

Визначення основних понять і формулювання оптимізаційної задачі

Нехай T і K — скінченні множини, елементи яких називаються, відповідно, об'єктами і мітками. Довільна підмножина $\gamma \subset T$ називається віконцем за аналогією з ситуацією, коли T — це множена пікселів, на якій подано зображення. Позначимо $\bar{k}: T \to K$ функцію, що називається розміткою множини T мітками з множини K. Множину всіх можливих функцій такого формату позначимо K^T. Для заданої функції $\bar{k}: T \to K$ позначимо $k(t)$, $t \in T$ значення цієї функції для об'єкта t і $k(\gamma)$, $\gamma \subset T$ — її звуження на γ. Іншими словами, $k(t)$ — це мітка об'єкта $t \in T$, а $k(\gamma)$ — це фрагмент розмітки $\bar{k}$, що спостерігається у віконці $\gamma \subset T$ і сам є розміткою $k(\gamma): \gamma \to K$ віконця γ. Множину всіх можливих розміток віконця γ мітками з K позначимо K^{γ}.

Нехай $\Gamma \subset 2^T$ — підмножина віконець, що називається структурою, $\max_{\gamma \in \Gamma} |\gamma|$ — число, що називається порядком структури, і нехай для кожного віконця $\gamma \in \Gamma$ і кожного фрагмента $s \in K^{\gamma}$ задано число $g_{\gamma}(s) \in \{0,1\}$, що визначає припустимість чи неприпустимість фрагмента s у віконці γ. Для кожної розмітки $\bar{k} \in K^T$ сукупність цих чисел визначає число

$$g(\bar{k}) = \bigwedge_{\gamma \in \Gamma} g_{\gamma}(k(\gamma)) \in \{0,1\}, \quad (1)$$

яке розуміється як припустимість, $g(\bar{k}) = 1$, чи неприпустимість, $g(\bar{k}) = 0$, цієї розмітки.

Перевірка несуперечності обмежень полягає у розв'язку задач наступного формату.

Означення 1. Вхідними даними задачі є скінченні множини T і K, структура $\Gamma \subset 2^T$ і числа $g_{\gamma}(s) \in \{0,1\}$, визначені для всіх $\gamma \in \Gamma$ і усіх $s \in K^{\gamma}$; розв'язком задачі є число

$$G = \bigvee_{\bar{k} \in K^T} g(\bar{k}) = \bigvee_{\bar{k} \in K^T} \bigwedge_{\gamma \in \Gamma} g_{\gamma}(k(\gamma)) \in \{0,1\}, \quad (2)$$

яке відповідає на питання, чи існує така розмітка $\bar{k}^*: T \to K$, $\bar{k}^* \in K^T$, що $g_{\gamma}(k^*(\gamma)) = 1$ для всіх $\gamma \in \Gamma$.

Задачі наведеного формату називають $(\vee, \wedge)$-задачами згідно з логічними операціями у (2). Множина всіх можливих $(\vee, \wedge)$-задач утворює *NP*-повний клас проблем. Більш за це, *NP*-повний є навіть підклас $(\vee, \wedge)$-задач, визначених на структурах другого порядку. У підручнику [4] виконано вичерпний аналіз поліноміально розв'язних підкласів цього *NP*-повного класу.

У роботах [5, 6] наведено приклади змістовно зрозумілих множин зображень, розпізнавання яких зводиться до розв'язку $(\vee, \wedge)$-задач і які ілюструють абстрактні поняття науки про несуперечність обмежень. Водночас ці прості приклади показують, як треба розширити проблематику $(\vee, \wedge)$-задач, щоб вона стала адекватною реалістичним задачам розпізнавання, а не тільки штучним наочним прикладам. Вагомою мотивацією такого розширення став істотний прорив у практиці розпізнавання, а саме, в прикладних задачах розпізнавання усномовних сигналів [7], текстових рядків [8] і бінокулярного стереобачення [9]. Успішний розв'язок цих задач було досягнуто завдяки окремому випадку структури вхідних даних, коли вхідні дані утворюють послідовність, а розпізнавання зводиться до розв'язку певних оптимізаційних задач методами динамічного програмування. Ці результати доволі швидко і цілком заслужено стали загальновизнаними, і це стимулювало пошук такого узагальнення розроблених методів, яке стало би придатним для розпізнавання об'єктів зі складнішою структурою, зокрема, для розпізнавання двовимірних масивів даних, а не тільки одновимірних послідовностей.

Таким чином, багата передісторія незалежного один від одного розвитку наук про розпізнавання образів і несуперечності обмежень привела до формулювання наступної оптимізаційної задачі, яку зазвичай називають $(\max, +)$-задачею.

Як і в $(\vee, \wedge)$ - задачі, так і в $(\max, +)$-задачі, T і K — це скінченні множини, $\Gamma \subset 2^T$ — структура,

$\bar{k}: T \to K$ — розмітка, $k(t)$ — мітка об'єкта $t \in T$, $k(\gamma)$— фрагмент розмітки $\bar{k}$ у віконці $\gamma \in \Gamma$. На відміну від $(\vee, \wedge)$ - задачі, числа $g_\gamma(s)$ в (max,+)-задачі — це не бінарні, а реальні числа, які розуміються як вага фрагмента s у віконці γ. Сукупність чисел $g_\gamma(s)$, заданих для всіх $s \in K^\gamma$ і усіх $\gamma \in \Gamma$, визначає число

$$g(\bar{k}) = \sum_{\gamma \in \Gamma} g_\gamma(k(\gamma)), \tag{3}$$

яке розуміється як вага розмітки $\bar{k} \in K^T$. Проблема полягає у розв'язку задач наступного формату.

Означення 2. Вхідними даними (max,+) - задачі є скінченні множини T і K, структура $\Gamma \subset 2^T$ і числа $g_\gamma(s) \in \mathbb{R}$, визначені для всіх $\gamma \in \Gamma$ і усіх $s \in K^\gamma$; розв'язком (max,+) - задачі є число

$$G = \max_{\bar{k} \in K^T} g(\bar{k}) = \max_{\bar{k} \in K^T} \sum_{\gamma \in \Gamma} g_\gamma(k(\gamma)) \in \mathbb{R}. \tag{4}$$

Як видно, клас усіх (max,+) - задач не простіший за *NP*-повний клас проблем, бо містить у собі *NP*-повний клас усіх можливих$(\vee, \wedge)$ -задач. Більш за це, навіть підклас (max,+)-задач на структурах другого порядку не простіший за *NP*-повний класс проблем.

Осільки фундаментальна проблематичність $(\vee, \wedge)$- і (max,+)-задач зберігається й у цьому підкласі задач, далі розглядатимуться лише $(\vee, \wedge)$- і (max,+)-задачі на структурах другого порядку.

Нехай Γ — структура другого порядку, до того ж така, що $\{t\} \in \Gamma$ для будь-якого $t \in T$. Твердження $\{t\} \in \Gamma$ будемо записувати у більш простому вигляді $t \in T$, а твердження $\{t, t'\} \in \Gamma$ — у вигляді $tt' \in \Gamma$. Для кожного об'єкта $t \in T$ позначимо $N(t)$ множину $\{t' \in T \mid tt' \in \Gamma\}$. Пару $(k, t) \in K \times T$ назвемо вершиною з іменем k у об'єкті t, а пару $[(k,t),(k',t')]$ вершин, таких що $tt' \in \Gamma$, назвемо дужкою. Сукупність величин $g_\gamma(s)$, які визначають припустимість фрагмента s у віконці γ у $(\vee, \wedge)$-задачі або його вагу у (max,+)-задачі, подаватиметься у вигляді двох вагових функцій. Перша функція, позначена як q, визначена на множині $K \times T$ вершин, а її значення для вершини $(k,t) \in K \times T$ буде позначатися $q_t(k)$. Друга функція, позначена як g, визначена на множині $K \times K \times \Gamma$ дужок, а її значенння на дужці $[(k,t),(k't')] \in K \times K \times \Gamma$ позначатиметься $g_{tt'}(k,k')$. На основі цих позначень формулювання $(\vee, \wedge)$-задачі і (max,+)-задачі на структурах другого порядку набувають наступного вигляду.

Означення 3. Вхідними даними $(\vee, \wedge)$-задачі на структурі другого порядку є скінченні множини T і K, структура Γ другого порядку і функції $q: K \times T \to \{0,1\}$, $g: K \times K \times \Gamma \to \{0,1\}$; розв'язком задачі є число

$$G = \bigvee_{\bar{k} \in K^T} \left[\bigwedge_{tt' \in \Gamma} g_{tt'}(k(t), k(t')) \wedge \bigwedge_{t \in T} q_t(k(t)) \right] \in \\ \in \{0,1\}. \tag{5}$$

Означення 4. Вхідними даними (max,+)-задачі на структурі другого порядку є скінченні множини T і K, структура Γ другого порядку та вагові функції $q: K \times T \to \mathbb{R}$, $g: K \times K \times \Gamma \to \mathbb{R}$; розв'язком задачі є число

$$G = \max_{\bar{k} \in K^T} \left[\sum_{tt' \in \Gamma} g_{tt'}(k(t), k(t')) + \sum_{t \in T} q_t(k(t)) \right] \in \mathbb{R}. \tag{6}$$

Вхідні данні $(\vee, \wedge)$- и (max,+)-задач будемо записувати коротко у вигляді п'ятірки $\langle T, K, \Gamma, q, g \rangle$, не нагадуючи кожного разу, що T і K — скінченні множини тощо.

Поліноміально розв'язні підкласи (max,+)-задач

Відомо три підкласи (max,+)-задач, кожен з яких утворює класс *P*-задач. У перший підклас входять задачі на обмеженому класі структур, що називаються ациклічними. При виконанні цих обмежень (max,+)-задача розв'язується за поліноміальний час для довільних вагових функцій. Задачі цього класу називаються ациклічними.

Другий підклас є у певному сенсі протилежним до першого. До другого класу входять задачі з ваговими функціями обмеженого типу, відомими як супермодулярні функції. Якщо ваги дужок задовільняють ці обмеження, то (max,+)-задача розв'язується за поліноміальний час для довільної структури і довільних ваг вершин. Задачі цього класу називаються супермодулярними.

Наведені два підкласи задач відрізняються між собою, хоча й мають непорожній перетин.

Їхня відмінність полягає в суттєвій відмінності алгоритмів для їхнього розв'язку. Алгоритми розв'язку задач одного з підкласів у жодному сенсі не є модифікаціями або вдосконаленнями алгоритмів розв'язку задач іншого підкласу. У цьому плані набуває особливого значення третій підклас поліноміально розв'язних (max,+)-задач, який охоплює всі ациклічні та всі супермодулярні задачі. Розв'язок задач цього підкласу спирається на еквівалентні перетворення (max,+)-задач і зведення їх до специфічних задач особливого типу, що названі тривіальними. В англомовній літературі такі перетворення називаються репараметризацією.

Ациклічні (max,+)-задачі

Множина T об'єктів і структура Γ утворюють граф $\langle T,\Gamma\rangle$ з множиною T вершин і множиною Γ ребер.

Означення 5. Задача $\langle T,K,\Gamma,q,g\rangle$ називається ациклічною, якщо граф $\langle T,\Gamma\rangle$ є неорієнтованим деревом; структура Γ у цьому випадку теж називається ациклічною.

Методи розв'язку ациклічних (max,+)-задач відомі (див. наприклад, [3]) і грунтуються на послідовному вилученні об'єктів з множини T методами динамічного програмування. Ці методи доволі природно увійшли до практики розпізнавання образів, бо вони є досить простим узагальненням основних ідей робіт [7, 8], виконаних до формулювання (max,+)-задач у загальному вигляді та які стимулювали це узагальнення.

Нехай у певній прикладній ситуації відомо, що вона адекватно формалізується як ациклічна (max,+)-задача, але структура, що визначає цю ациклічність, невідома. Однак відомо, що вона є підмножиною іншої структури, яка є відомою. Задача, що виникає у цій ситуації, яку ми назвемо задачею з прихованою ациклічністю, визначається як розглянуто далі.

Означення 6. Задача $\langle T,K,\Gamma,q,g\rangle$ називається задачею з прихованою ациклічністю, якщо існує ациклічна задача $\langle T,K,\Gamma^*,q^*,g^*\rangle$ така, що $\Gamma^*\subset\Gamma$ і для довільної розмітки $\overline{k}\in K^T$ виконується рівність

$$\sum_{tt'\in\Gamma} g_{tt'}(k(t),k(t')) + \sum_{t\in T} q_t(k(t)) = \\ = \sum_{tt'\in\Gamma^*} g^*_{tt'}(k(t),k(t')) + \sum_{t\in T} q^*_t(k(t)). \tag{7}$$

Задачі з прихованою ациклічністю не є ациклічними, і тому до них безпосередньо незастосовні методи розв'язку ациклічних задач. Для застосування цих методів до конкретної задачі з прихованою ациклічністю, її треба було б спочатку перетворити на ациклічну, причому так, щоб виконувалася умова (7), а алгоритми таких перетворень не є апріорі очевидними. У розділі Еквівалентні перетворення описано методи розв'язку задач із прихованою ациклічністю, що не вимагають їхнього попереднього перетворення на ациклічні.

Супермодулярні (max,+)-задачі

У роботі [10] досліджено прикладну задачу сегментації зображення на сегменти двох типів і показано, що ця прикладна задача формалізується таким окремим випадком (max,+)-задачі, який зводиться до задачі поліноміальної складності про максимальний потік у мережі. Робота [10] є першою роботою, в якій було розв'язано практичну задачу через її зведення до (max,+)-задачі на структурі, що не є ациклічною. Невдовзі цей прорив було суттєво розширено у роботі [11], в якій сформульовано широкий клас (max,+)-задач, розв'язок яких зводиться до пошуку максимального потоку в мережі. Цей клас визначається на основі наступних понять.

Нехай N — множина позитивних цілих чисел, а $n:K\times T\to\mathrm{N}$ — функція, що є нумерацією множини $K\times T$ вершин і кожній вершині $(k,t)\in K\times T$ призначає число $n_t(k)$, що слугує номером мітки k в об'єкті t. Для кожного об'єкта t і кожної пари $k_1\neq k_2$ неоднакових міток функція n набуває неоднакові значення $n_t(k_1)\neq n_t(k_2)$ і в такий спосіб здійснює впорядкованність множини міток, свою для кожного об'єкта.

Означення 7. Задача $\langle T,K,\Gamma,q,g\rangle$ називається супермодулярною відносно заданої нумерації $n: K\times T\to \mathrm{N}$, якщо для довільної пари $tt'\in\Gamma$ об'єктів і довільних чотирьох міток k_1,k_2,k_3,k_4, таких що $n_t(k_1)<n_t(k_2)$ і $n_{t'}(k_{1'})<n_{t'}(k_{2'})$, виконується нерівність

$$g_{tt'}(k_1,k_{1'})+g_{tt'}(k_2,k_{2'})\geq g_{tt'}(k_1,k_{2'})+ \\ +g_{tt'}(k_2,k_{1'}). \tag{8}$$

Означення 8. Задача $\langle T,K,\Gamma,q,g\rangle$ називається супермодулярною, якщо існує така нумерація $n: K\times T\to \mathrm{N}$, стосовно якої ця задача є супермодулярною.

Означення 7 і 8 суттєво відрізняються одне від одного. Для довільної заданої задачі $z=\langle T,K,\Gamma,q,g\rangle$ і довільної заданої нумерації $n: K\times T\to \mathrm{N}$ зовсім не важко перевірити, чи відповідають задача z і нумерація n означенню 7. Означення 7 дає безпосередню вказівку на умови (8), яким має задовільняти пара (z,n) і які легко перевіряються. Якщо ж нумерацію не задано, то перевірка самої лише задачі без нумерації на відповідність означенню 8 є проблематичнішою. В означенні 8 нема вказівок на умови супермодулярності, які було би легко перевірити. Натомість формулюється допоміжна задача, яку треба розв'язати, щоб відповісти на питання про супермодулярність поданої задачі.

У роботі [11] показано алгоритм, який для заданих нумерації і (max,+)-задачі, супермодулярної для цієї нумерації, будує такий граф, що розв'язок заданої (max,+)-задачі зводиться до пошуку максимального потоку в цьому графі. В алгоритмі суттєвим чином використовується задана нумерація, і тому він є незастосовним, коли цю нумерацію не задано. У наступному розділі описано метод розв'зку супермодулярних задач, для реалізації якого не потрібно знати, яка саме впорядкованість вершин зумовлює цю супермодулярність, а достатньо знати, що така впорядкованість існує.

Еквівалентні перетворення (max,+)-задач (репараметризація)

Оберемо довільно і зафіксуємо для подальшого розгляду множину T об'єктів, множину K міток, структуру Γ другого порядку й ототожнимо множину всіх можливих (max,+)-задач $z=\langle T,K,\Gamma,q,g\rangle$ з лінійним простором, визначивши суму задач і помноження задачі на число $a\in\mathbb{R}$ як

$$\langle T,K,\Gamma,\ q_1,g_1\rangle+\langle T,K,\Gamma,\ q_2,g_2\rangle= \\ =\langle T,K,\Gamma,\ q_1+q_2,g_1+g_2\rangle, \\ a\cdot\langle T,K,\Gamma,\ q,g\rangle=\langle T,K,\Gamma,\ a\cdot q,a\cdot g\rangle.$$

Для кожної (max,+)-задачі $z=\langle T,K,\Gamma,q,g\rangle$ визначимо її числову характеристику

$$P(z)=\sum_{t\in T}\max_{k\in K}q_t(k)+\sum_{tt'\in\Gamma}\max_{k\in K}\max_{k'\in K}g_{tt'}(k,k'), \tag{9}$$

яку назвемо потенціалом задачі. Як видно з визначення (9), потенціал є опуклою функцією на множині задач. Позначимо

$$G(\bar{k})=\sum_{t\in T}q_t(k(t))+\sum_{tt'\in\Gamma}g_{tt'}(k(t),k(t'))$$

сумарну вагу розмітки $\bar{k}$ в задачі $z=\langle T,K,\Gamma,q,g\rangle$ і зауважимо, що для довільної розмітки $\bar{k}\in K^T$ виконується нерівність $G(\bar{k})\leq P(z)$.

Означення 9. Задача $z=\langle T,K,\Gamma,q,g\rangle$ називається тривіальною, якщо існує розмітка $\bar{k}^*$, для якої $G(\bar{k}^*)=P(z)$.

Розв'язок тривіальної задачі зводиться до обчислення її потенціалу за формулою (9).

Означення 10. Задачі $z^1=\langle T,K,\Gamma,\ q^1,g^1\rangle$ і $z^2=\langle T,K,\Gamma,\ q^2,g^2\rangle$ називаються еквівалентними, якщо для кожної розмітки виконується рівність

$$\sum_{t\in T}q_t^1(k(t))+\sum_{tt'\in\Gamma}g_{tt'}^1(k(t),k(t'))= \\ =\sum_{t\in T}q_t^2(k(t))+\sum_{tt'\in\Gamma}g_{tt'}^2(k(t),k(t')).$$

Позначимо $Z(z)$ множину задач, еквівалентних z. З означення 9 і нерівності $G(\bar{k})\leq P(z)$ безпосередньо випливає

Лема 1. Якщо для задачі z існує тривіальний еквівалент, то довільна задача $\arg\min\limits_{z'\in Z(z)}P(z')$ є тривіальною.

З леми 1 випливає, що розв'язок довільної (max,+)-задачі *z*, для якої існує тривіальний еквівалент, зводиться до обчислення числа $\min_{z' \in Z(z)} P(z')$. Або, детальніше, розв'язок такої (max,+)-задачі $\langle T, K, \Gamma, q, g \rangle$ зводиться до пошуку вагових функцій $q^*: K \times T \to \mathbb{R}$ і $g^*: K \times K \times \Gamma \to \mathbb{R}$, які мінімізують число

$$\sum_{t \in T} \max_{k \in K} q_t^*(k) + \sum_{tt' \in \Gamma} \max_{k \in K} \max_{k' \in K} g_{tt'}^*(k, k') \qquad (10)$$

за умови, що для всіх $\overline{k} \in K^T$ виконується рівність

$$\begin{aligned} &\sum_{t \in T} q_t(k(t)) + \sum_{tt' \in \Gamma} g_{tt'}(k(t), k(t')) = \\ &= \sum_{t \in T} q_t^*(k(t)) + \sum_{tt' \in \Gamma} g_{tt'}^*(k(t), k(t')). \end{aligned} \qquad (11)$$

Як видно, розв'язок (max,+)-задач, для яких існує тривіальний еквівалент, зводиться до пошуку мінімума опуклої функціїї на множині розв'язків системи лінійних рівнянь і, таким чином, до розв'язку деякої задачі опуклої оптимізації. Однак для практичного використання цього факту потрібна позитивна відповідь на два питання. По-перше, стандартні методи опуклої мінімізації безпосередньо непридатні для мінімізації (10) через експоненціальну кількість обмежень (11), і постає питання про наявність практично придатного алгоритму для розв'язку цієї задачі. По-друге, постає питання про те, наскільки обсяжною є множина задач, еквівалентних тривіальним, зокрема, у порівнянні з множинами ациклічних і супермодулярних задач.

Кількість обмежень в оптимізаційній задачі (10, 11) можна суттєво зменшити завдяки введенню в задачу допоміжних змінних $\varphi(t, t', k)$, визначених для кожної трійки $k \in K$, $t \in T$, $t' \in N(t)$. Назвемо ці змінні потенціалами та позначимо φ массив, у якому для кожної трійки $k \in K$, $t \in T$, $t' \in N(t)$ записано потенціал $\varphi(t, t', k)$. Позначимо також Φ множину всіх таких масивів. Задачу $\langle T, K, \Gamma, q, g \rangle$ назвемо нульовою, якщо для кожної розмітки $\overline{k} \in K^T$ виконується рівність

$$G(\overline{k}) = \sum_{t \in T} q_t(k(t)) + \sum_{tt' \in \Gamma} g_{tt'}(k(t), k(t')) = 0.$$

Лема 2. [12] Задача $\langle T, K, \Gamma, q, g \rangle$ є нульовою тоді і тільки тоді, коли існують потенціали $\varphi(t, t', k), k \in K, t \in T, t' \in N(t)$, що задовільняють систему лінійних рівнянь

$$q_t(k) = \sum_{t' \in N(t)} \varphi(t, t', k), \;\; t \in T, k \in K,$$

$$g_{tt'}(k, k') = \varphi(t, t', k) + \varphi(t', t, k'), tt' \in \Gamma, \;\; k, k' \in K.$$

Очевидно, що дві (max,+)-задачі є еквівалентними тоді і тільки тоді, коли їхня різниця є нульовою. Тому з леми 2 безпосередньо випливає, що мінімізація величини (10) за $|T \times K| + |\Gamma \times K \times K|$ змінними при $\left|K^{|T|}\right|$ обмеженнях (11) зводиться до мінімізації тієї ж самої величини

$$\sum_{t \in T} \max_{k \in K} q_t^*(k) + \sum_{tt' \in \Gamma} \max_{k \in K} \max_{k' \in K} g_{tt'}^*(k, k') \qquad (12)$$

при $|T \times K|$ обмеженнях

$$q_t^*(k) = q_t(k) - \sum_{t' \in N(t)} \varphi(t, t', k), t \in T, k \in K \qquad (13)$$

і $|\Gamma \times K \times K|$ обмеженнях

$$\begin{aligned} &g_{tt'}^*(k, k') = g_{tt'}(k, k') + \varphi(t, t', k) + \\ &+ \varphi(t', t, k'), \quad tt' \in \Gamma, \;\; k, k' \in K. \end{aligned} \qquad (14)$$

Якщо задані вагові функції *q* і *g* вхідної задачі та потенціали $\varphi(t, t', k)$, то обмеження (13) і (14) однозначно визначають вагові функції *q** і *g** шуканої еквівалентної задачі. Внаслідок цього, мінімізація числа (12) при обмеженнях (13) і (14) або, так само, мінімізація числа (10) при обмеженнях (11) зводиться до мінімізації числа

$$\begin{aligned} H(\varphi) = &\sum_{t \in T} \max_{k \in K} \left[q_t(k) - \sum_{t' \in N(t)} \varphi(t, t', k) \right] + \\ &+ \sum_{tt' \in \Gamma} \max_{k \in K} \max_{k' \in K} \left[g_{tt'}(k, k') + \varphi(t, t', k) + \varphi(t', t, k') \right] \end{aligned} \qquad (15)$$

за потенціалами $\varphi(t, t', k)$ без будь-яких обмежень на ці потенціали. Для довільних вагових функцій *q* і *g* вхідної (max,+)-задачі функція $H: \Phi \to R$, $\varphi \to H(\varphi)$, є опуклою функцією потенціалів, і таким чином, існує процедура, областю визначення якої є множина всіх можливих (max,+)-задач і яка для довільної вхідної задачі обчислює число $\min_{\varphi \in \Phi} H(\varphi)$. Це число є розв'язком не довільної (max,+)-задачі, а певного підкласу цих задач, який назвемо областю компетентності процедури на відміну від її області визначення. Область компетентності процедури обчислення числа $\min\limits_{\varphi\in\Phi} H(\varphi)$ формулює наступна теорема.

Теорема 1. Розв'язком (max,+)-задачі, для якої існує тривіальний еквівалент, є число $\min\limits_{\varphi\in\Phi} H(\varphi)$.

Теорема є очевидним наслідком леми 1 і доведеної еквівалентності оптимізаційних задач (10) і (15).

Таким чином, існують три підкласи поліноміально розв'язних (max,+)-задач. Перший підклас складається з ациклічних задач, які розв'язуються методами динамічного програмування. Другий підклас складається із задач, супермодулярних відносно заданого впорякування вершин. Задачі цього підкласу зводяться до пошуку максимального потоку в мережі або, так само, до пошуку мінімального розрізу графа. Третій підклас складають задачі, еквівалентні тривіальним, які зводяться до мінімізації опуклих функцій виду (15). Наступні дві леми визначають співвідношення між цими трьома підкласами.

Лема 3. Для довільної (max,+)-задачі з прихованою ациклічністю існує тривіальний еквівалент.

Доведення. У роботі [12] доведено, що довільна ациклічна (max,+)-задачі має тривіальний еквівалент. З умови (7) випливає, що для довільної (max,+)-задачі з прихованою ациклічністю існує еквівалентна ациклічна задача.

Лема 4. [12] Для довільної супермодулярної (max,+)-задачі існує тривіальний еквівалент.

Дві останні леми прозоро й лаконічно формулюють важливий результат, для якого, однак, незайвим є наступний коментар. Очевидно, что класс супермодулярных задач є значно ширшим, ніж клас задач, супермодулярних відносно заданої фіксованої нумерації вершин, а клас задач із прихованою ациклічністю є значно ширшим, ніж клас ациклічних задач. Щоб ту чи іншу супермодулярну (max,+)-задачу звести до пошуку максимального потоку в мережі, необхідно знати не тільки вхідні дані цієї задачі, а й нумерацію вершин, відносно якої ця задача є супермодулярною. Для еквівалентного перетворення задачі на тривіальну не треба знати цієї нумерації, а потрібно лише гарантувати її існування. Аналогічно, для розв'язку деякої ациклічної задачі методами динамічного програмування необхідно знати ту конкретну структуру, яка зумовлює цю ациклічність, і тому ці методи є незастосовними до задач із прихованою ациклічністю. Для еквівалентного перетворення задачі на тривіальну не треба знати цієї структури, а потрібна лише гарантія її існування.

Зауважимо при цьому, що множина задач, еквівалентних тривіальним, є ширшою за об'єднання множин супермодулярних і ациклічних задач. Як найпростіший приклад розглянемо задачу на множині T об'єктів, множині K міток та структурі Γ, що не є ациклічною. Нехай $\bar{k}^*: T \to K$ — деяка розмітка, а ваги вершин і дужок визначено як $q_t(k) = 0$ для всіх $k \in K$, $t \in T$, $g_{tt'}(k^*(t), k^*(t')) = 0$ для всіх $tt' \in \Gamma$, $g_{tt'}(k,k') = -1$ для всіх $tt' \in \Gamma$, $(k,k') \neq (k^*(t), k^*(t'))$.

Ця задача не є ациклічною і не є супермодулярною. Але потенціал цієї задачі дорівнює сумарній вазі розмітки $\bar{k}^*$, тобто

$$\sum_{t\in T}\max_{k\in K} q_t(k) + \sum_{tt'\in\Gamma}\max_{k\in K}\max_{k'\in K} g_{tt'}(k,k') = a$$
$$= \sum_{t\in T} q_t(k^*(t)) + \sum_{tt'\in\Gamma} g_{tt'}(k^*(t), k^*(t'))$$

і за означенням 9 ця задача є тривіальною. Із цього випливає, що ця задача і всі еквівалентні до неї задачі входять у область компетентності методу, заснованого на пошуку мінімального значення величини (15).

Таким чином, у трійці відомих поліноміально розв'язних підкласів (max,+)-задач підклас, що базується на еквівалентному перетворенні задач, посідає особливе місце, бо охоплює об'єднання двох інших підкласів, але не є тотожним з цим об'єднанням. Із цього факту аж ніяк не випливає рекомендація давати перевагу еквівалентній трансформації задач у всіх прикладних ситуаціях, а випливає лише достатньо чітко окреслена межа наших теперішніх знань про формальні властивості (max,+)-задач.

Прикінцеві зауваження

Нинішню ясність у розв'язку (max,+)-задач, які утворюють основу нового напряму в розпізнаванні образів, було досягнуто протягом останніх 25 років, і початковою точкою цього періоду є робота [10]. Чверть сторіччя, яка передувала цій початковій точці, була періодом усвідомлення необхідності розв'язку (max,+)-задач для прогресу науки та практики розпізнавання (див. наприклад [5, 6]). Як видно, зусилля в цьому напрямі були настільки тривалими, що вся проблематика (max,+)-задач отримала стійку репутацію неприступної фортеці, подолання якої не варто очікувати в найближчому майбутньому. Робота [10] стала першою шпаринкою в бастіоні цієї фортеці, яку майже відразу суттєво розширила робота [11]. Нарешті, ідея еквівалентної трансформації (max,+)-задач перетворила цю шпарину на широкий пролом, що відкрив вхід у фортецю для багатьох сотень досліджень, яким присвячується постійна конференція *Energy Minimization Methods in Computer Vision and Pattern Recognition* (див. наприклад, останній випуск праць цієї конференції [13] і, звичайно ж, попередні випуски). Таким чином, новий напрям у науці про розпізнавання образів можна вважати таким, що відбувся й відкрився для майбутніх досліджень.

ЛІТЕРАТУРА

1. *Гриценко В. И., Шлезингер М. И.* Формальные модели, задачи и алгоритмы образного мышления. *Автоматика-2011* (*AUTOMATICS-2011*) : Матеріали 18-ої міжнародної конференції з автоматичного управління, (Львів, 28–30 вересня, 2011). С. 110–113.
2. *Гриценко В. И., Шлезингер М. И.* Взаимосвязь проблем распознавания образов, машинного мышления и обучения. *Международній научно-технический журнал «Проблемы управления и информатики»*. 2020. 3. С. 108–136.
3. *Шлезингер М., Главач В.* Десять лекций по статистическому и структурному распознаванию. Киев : Наукова думка, 2004. 545 с.
4. *Rossi F., van Beek P., Walsh T.* Handbook of Constraint Programming. Elsevier, 2006. 975 p.
5. *Шлезингер М. И.* Синтаксический анализ двумерных зрительных сигналов в условиях помех. *Кибернетика*. 1976. 4. С. 113–129.
6. *Шлезингер М. И.* Математические средства обработки изображений. Киев : Наукова думка, 1989. 197 с.
7. *Винцюк Т. К.* Анализ, распознавание и интерпретация речевых сигналов. Киев : Наукова думка, 1987. 262 с.
8. *Ковалевский В. А.* Методы оптимальных решений в распознавании изображений. Москва : Наука, 1976. 328 с.
9. *Gimelfarb G. L.* Symmetrical approach to the problem of automating stereoscopic measurements in photogrammetry. *Cybernetics and System Analysis*. 1979. 15. P. 235–247.
10. *Ishikawa H., Geiger D.* Segmentation by grouping junctions. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1998. P. 125–131.
11. *Ковтун І. В.* Сегментація зображень на основі достатніх умов оптимальності в NP-повних класах задач структурної розмітки : дисертація на здобуття наукового ступеня кандидата технічних наук / Міжнародний науково-навчальний центр інформаційних технологій і систем. Київ, 2004.
12. *Шлезингер М. И., Гигиняк В. В.* Решение (max,+)-задач структурного распознавания при помощи их эквивалентных преобразований. *Управляющие системы и машины*. Киев, 2007. 1. С. 3–15. 2. С. 3–18.
13. Energy Minimization Methods in Computer Vision and Pattern Recognition. EMMCVPR 2017. Lecture Notes in Computer Science, 10746. Springer.

Надійшла 15.02.2022

REFERENCES

1. *Grytsenko V. I., Schlesinger M.I.*, 2011. "Formalnye modely, zadachy y alhorytmy obraznoho myshlenyya", Proceedings of the 18th International Conference on Automatic Control AUTOMATICS-2011 (Lviv, 28–30 Sept. 2011), pp. 110–113.
2. *Grytsenko V. I., Schlesinger M.I.,* 2020. "Vzaimosvyaz problem raspoznavaniya obrazov, mashinnogo myshleniya i obucheniya", International scientific and technical journal "Problemy upravleniya i informatiki", 3, pp. 108–136.

3. *Schlesinger M.I., Hlavach V.*, 2004. Desyat lektsyy po statystycheskomu y strukturnomu raspoznavanyyu, Naukova dumka, Kyiv, 545 p.
4. *Rossi F., van Beek P., Walsh T.*, 2006. Handbook of Constraint Programming, Elsevier, 975 p.
5. *Schlesinger M.I.*, 1976. "Sintaksicheskiy analiz dvumernykh zritelnykh signalov v usloviyakh pomekh", Kibernetika, 4, pp. 113–129.
6. *Schlesinger M.I.*, 1989. Matematicheskiye sredstva obrabotki izobrazheniy, Naukova dumka, Kyiv, 197 p.
7. *Vintsyuk T. K.*, 1987. Analiz, raspoznavaniye i interpretatsiya rechevykh signalov, Naukova dumka, Kyiv, 262 p.
8. *Kovalevskiy V. A.*, 1976. Metody optimalnykh resheniy v raspoznavanii izobrazheniy, Nauka, Moscow, 328 p.
9. *Gimelfarb G. L.*, 1979. "Symmetrical approach to the problem of automating stereoscopic measurements in photogrametry", Cybernetics and System Analysis, 15, pp. 235–247.
10. *Ishikawa H., Geiger D.,* 1998. "Segmentation by grouping junctions", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 125–131.
11. *Kovtun I. V.*, 2004. Sehmentatsiya zobrazhen na osnovi dostatnikh umov optymalnosti v NP-povnykh klasakh zadach strukturnoyi rozmitky, Ph. D. Thesis, International Research and Training Center for Information Technology and Systems, Kyiv.
12. *Schlesinger M.I., Giginyak V. V.*, 2007. "Resheniye (max,+)-zadach strukturnogo raspoznavaniya pri pomoshchi ikh ekvivalentnykh preobrazovaniy", Control systems and machines, Kyiv, 1, pp. 3–15, 2, pp. 3–18.
13. Energy Minimization Methods in Computer Vision and Pattern Recognition, EMMCVPR 2017, Lecture Notes in Computer Science, 10746, Springer.

Received 15.02.2022

M.I. Schlesinger, doctor of science, professor, chief researcher,
International Research and Training Center for Information Technologies
and Systems NAS and MES of Ukraine,
Glushkov ave., 40, Kyiv, 03187, Ukraine,
schles@irtc.org.ua

SOLUTION OF SOFT CONSTRAINS PROBLEMS VIA THEIR REPARAMETRIZATION

Introduction. The past quarter-century is characterized by the birth of a new scientific direction, formed as a result of combining research in pattern recognition problems and constraint satisfaction problems. These two scientific directions traditionally belong to the problem of artificial intelligence, but they formalize different aspects of intellectual activity. The formation of a single formal scheme that combines these two directions expands and concretizes the concept of machine thinking, on the formalization of which they are oriented, and necessitates the development of new and improvement of known mathematical optimization methods.

Objective. Comparison of three currently known polynomially solvable subclasses of the NP-hard class of optimization problems, which constitutes a mathematical formalization of a new scientific direction. Problems of the first subclass are solved by dynamic programming methods, problems of the second subclass are solved by supermodular maximization methods, and problems of the third subclass are solved by methods of equivalent transformation of optimization problems, also known as their reparametrization.

Result. The subclass of problems solved on the basis of their reparametrization includes subclasses solved using dynamic programming or supermodular maximization, and thus is the most extensive among the three currently known polynomially solvable subclasses.

Conclusion. Key moments in the process of formation of a new scientific direction are given.

***Keywords**: pattern recognition, constraint satisfaction, dynamic programming, supermodular maximization, reparametrization of optimization problems.*