NP-ПЪЛНОТА

КОНТРОЛНО № 5 ПО ДИЗАЙН И АНАЛИЗ НА АЛГОРИТМИ

ЗА СПЕЦИАЛНОСТ "КОМПЮТЪРНИ НАУКИ", 1. ПОТОК

(СУ, ФМИ, 27 МАЙ 2021 Г.)

Задача. Разглеждаме алгоритмичната задача за разпознаване "Подматрица":

— Вход: две числови матрици A и B.

— Въпрос: Матрицата A съдържа ли подматрица, еднаква на B
с точност до разместване на редове и разместване на стълбове?

Подматрица на A наричаме всяка матрица, образувана от елементите на A в пресечните точки на няколко избрани реда и няколко избрани стълба на A. Не е задължително избраните редове и стълбове да са един до друг. Формално: подматрица на A наричаме всяка матрица с елементи a_{ij}, $i \in R$, $j \in S$, където R е произволно множество от индекси на редове на A, а пък S е произволно множество от индекси на стълбове на A.

В задачата "Подматрица" се пита съществуват ли такива множества R и S, че породената от тях подматрица на A да е еднаква на B.

Тук две матрици се смятат за еднакви, ако едната се получава от другата чрез разместване на редове и разместване на стълбове.

Да се докаже, че "Подматрица" е NP-пълна задача.

(5 точки)

РЕШЕНИЕ

Неформално описание на идеята на решението: То се състои от две стъпки. До нашата задача "Подматрица" се свежда задачата "Изоморфен подграф", обаче само частният случай, когато графите са двуделни, и то с допълнително изискване подграфът да е индуциран и да е казано кой дял от единия граф на кой дял от другия граф съответства. Тази задача на свой ред е NP-пълна, което се доказва с редукция от общия случай на "Изоморфен подграф": разцепваме всяко ребро на две ребра, добавяйки връх по средата на реброто.

Подробно описание на решението. Алгоритмичната задача “Подматрица” е формулирана по-горе. Да припомним точната формулировка на задачата “Изоморфен подграф”:

Вход: неориентирани графи $G(\widetilde{V}\ ;\ \widetilde{E}\ )$ и $H(\widetilde{U}\ ;\ \widetilde{F}\ )$; всяка наредена двойка има за свой първи елемент множеството от върховете на съответния граф, а за втори елемент — множеството от ребрата му.

Въпрос: Графът G съдържа ли подграф, изоморфен на H ?

Следният частен (и променен!) случай на задачата “Изоморфен подграф” да наречем например “Индуциран двуделен подграф”:

Вход: два неориентирани двуделни графа $K(V_1 ; V_2 ; E)$ и $L(U_1 ; U_2 ; F)$, където V_1 и V_2 са двата дяла (върховете) на графа K, а пък E е множеството от ребрата на K, като E удовлетворява изискването за двуделност: $E \subseteq V_1 \times V_2$; аналогично, U_1 и U_2 са двата дяла (върховете) на графа L, а F е множеството от ребрата на L, като F удовлетворява изискването за двуделност: $F \subseteq U_1 \times U_2$.

Въпрос: Графът K съдържа ли индуциран подграф, изоморфен на L , като се иска изоморфизмът да запазва номерацията на дяловете? По-формално, съществуват ли инекции $f_1 : U_1 \rightarrow V_1$ и $f_2 : U_2 \rightarrow V_2$, такива че за всички $u_1 \in U_1$ и $u_2 \in U_2$ е вярно, че $\{u_1 ;\ u_2\} \in F$, ако и само ако $\{ f_1(u_1) ;\ f_2(u_2)\} \in E$?

Ще построим две полиномиални редукции:
“Изоморфен подграф” $\propto$ “Индуциран двуделен подграф” $\propto$ “Подматрица”.

Първа редукция: “Изоморфен подграф” $\propto$ “Индуциран двуделен подграф”. Тази редукция строим по следния начин:

Изоморфен подграф $(G(\widetilde{V}\ ;\ \widetilde{E}\ ) ,\ H(\widetilde{U}\ ;\ \widetilde{F}\ ))$
$V_1 \leftarrow \widetilde{V}$
$V_2 \leftarrow \widetilde{E}$
$E \leftarrow \varnothing$
for each $e = \{v_1 ;\ v_2\} \in \widetilde{E}$ **do**
    $E \leftarrow E \cup \{(v_1 ;\ e) ;\ (v_2 ;\ e)\}$
$U_1 \leftarrow \widetilde{U}$
$U_2 \leftarrow \widetilde{F}$
$F \leftarrow \varnothing$
for each $f = \{u_1 ;\ u_2\} \in \widetilde{F}$ **do**
    $F \leftarrow F \cup \{(u_1 ;\ f\ ) ;\ (u_2 ;\ f\ )\}$
return Индуциран двуделен подграф $(K(V_1 ; V_2 ; E) ,\ L(U_1 ; U_2 ; F))$

Всеки от двата цикъла има линейна сложност: обхожда по веднъж ребрата на съответния граф. Затова редукцията има полиномиална (линейна) сложност.

Забележка: В нашия курс по алгоритми предполагаме, че ребрата се пазят не в един списък, общ за целия граф, а че за всеки връх има отделен списък на ребрата, излизащи от него. В псевдокода по-горе за удобство е използвано първото представяне. Това не променя идеята на алгоритъма, нито неговата времева сложност. Ако използваме по един списък от ребра за всеки връх, то ще бъде нужно и обхождане на върховете, но сложността остава линейна, защото дължината на входа по порядък е общият брой на върховете и ребрата.

Коректност на редукцията:

Функцията “Изоморфен подграф”, извикана с параметри $G(\widetilde{V}\ ;\ \widetilde{E}\ )$ и $H(\widetilde{U}\ ;\ \widetilde{F}\ )$, връща истина.

$\Updownarrow$ (От ред № 11 на алгоритъма.)

Функцията “Индуциран двуделен подграф”, извикана с параметри $K(V_1\,;V_2\,;E)$ и $L(U_1\,;U_2\,;F)$, връща истина.

$\Updownarrow$ (От определението на задачата “Индуциран двуделен подграф”.)

Съществуват две инекции $f_1 : U_1 \to V_1$ и $f_2 : U_2 \to V_2$ със следното свойство: за всички $u_1 \in U_1$ и $u_2 \in U_2$ е вярно, че $\{f_1(u_1)\ ;\ f_2(u_2)\} \in E$ тогава и само тогава, когато $\{u_1\ ;\ u_2\} \in F$.

$\Updownarrow$ (От редове № 1 – № 10 на алгоритъма.
В посока надолу полагаме $f = f_1$, $u_2 = \{w_1\,;w_2\}$; ту $u_1 = w_1$, ту $u_1 = w_2$.
В посока нагоре полагаме $f_1 = f$, $f_2(\{w_1\,;w_2\}) = \{f(w_1)\ ;\ f(w_2)\}$.)

Съществува инекция $f : \widetilde{U} \to \widetilde{V}$ със следното свойство: за всички $w_1 \in \widetilde{U}$ и $w_2 \in \widetilde{U}$ е вярно, че ако $\{w_1\ ;\ w_2\} \in \widetilde{F}$, то $\{f(w_1)\ ;\ f(w_2)\} \in \widetilde{E}$.

Това, че на предпоследната стъпка $\{f_1(u_1)\ ;\ f_2(u_2)\} \in E$ влече $\{u_1\,;u_2\} \in F$, следва от 2-регулярността на дяловете U_2 и V_2.

Доказахме следната еквивалентност:

Функцията “Изоморфен подграф”, извикана с параметри $G(\widetilde{V}\ ;\ \widetilde{E}\ )$ и $H(\widetilde{U}\ ;\ \widetilde{F}\ )$, връща истина тогава и само тогава, когато графът $G(\widetilde{V}\ ;\ \widetilde{E}\ )$ съдържа подграф, изоморфен на графа $H(\widetilde{U}\ ;\ \widetilde{F}\ )$.

Това твърдение повтаря формулировката на задачата “Изоморфен подграф”, следователно функцията “Изоморфен подграф” решава едноименната задача, тоест редукцията е коректна.

От теорията е известно, че задачата “Изоморфен подграф” е NP-трудна. Щом тя се свежда до “Индуциран двуделен подграф” за полиномиално време, то задачата “Индуциран двуделен подграф” също е NP-трудна.

Втора редукция: “Индуциран двуделен подграф” $\propto$ “Подматрица”. Идеята е да вземем матриците на съседствата на двата графа. Без ограничение предполагаме, че върховете на всеки дял са номерирани с първите няколко цели положителни числа; можем да отъждествим всеки връх с неговия номер.

```
Индуциран двуделен подграф (K(V1; V2; E), L(U1; U2; F))
    A [V1] [V2], B[U1] [U2]: числови матрици
    for each v1∈V1 do
        for each v2∈V2 do
            A [v1] [v2] ← 0
    for each (v1; v2) ∈ E do // v1∈ V1, v2∈ V2
        A [v1] [v2] ← 1
    for each u1∈U1 do
        for each u2∈U2 do
            B [u1] [u2] ← 0
    for each (u1; u2) ∈ F do // u1∈ U1, u2∈ U2
        B [u1] [u2] ← 1
    return Подматрица (A , B)
```

Тази редукция има времева сложност $|V_1|.|V_2|+|E|+|U_1|.|U_2|+|F|$, което е квадратична (а оттам и полиномиална) функция на дължината на входа $|V_1|+|V_2|+|E|+|U_1|+|U_2|+|F|$.

Коректност на редукцията: Че функцията “Индуциран двуделен подграф” решава едноименната задача, следва от това, че A и B са матрици на съседствата за двуделните графи K и L съответно. Разместването на редовете на B е аналог на инекцията f_1; разместването на стълбовете на B съответства на избора на f_2. Изборът на множествата R и S от редове и стълбове на матрицата A съответства на избора на върхове от двата дяла на графа K, тоест $R = f_1(U_1)$, $S = f_2(U_2)$. Това, че подграфът е индуциран, следва от равенството на съответните елементи на матрицата B и подматрицата на A. На редовете на B се съпоставят редове на A и на стълбовете на B — стълбове на A, защото за двуделните графи сме казали кой дял от единия граф на кой дял от другия граф е съпоставен.

С това доказахме, че задачата “Индуциран двуделен подграф” е NP-трудна и има полиномиална редукция “Индуциран двуделен подграф” $\propto$ “Подматрица”. Следователно задачата “Подматрица” също е NP-трудна.

Че задачата “Подматрица” принадлежи на класа NP, ще докажем, съставяйки алгоритъм, който проверява предложено решение за полиномиално време. Сертификатът (тоест предложеното решение) се състои от две инекции, които описват кой ред и стълб на A отговаря на съответно избран ред или стълб на B. Представяме инекциите с масиви: индексите са номера на редове / стълбове на B, а стойностите на елементите са редове / стълбове на A съответно.

```
Проверка на подматрица (A[1...m1][1...n1], B[1...m2][1...n2], f[1...m2], g[1...n2])
for m ← 1 to m2 do
   if f[m] ∉ ℤ or f[m] < 1 or f[m] > m1
      return false // недопустим номер на ред
for n ← 1 to n2 do
   if g[n] ∉ ℤ or g[n] < 1 or g[n] > n1
      return false // недопустим номер на стълб
for m ← 1 to m2 do
   for n ← 1 to m2 do
      if f[m] = f[n]
         return false // повтаряне на номер на ред
for m ← 1 to n2 do
   for n ← 1 to n2 do
      if g[m] = g[n]
         return false // повтаряне на номер на стълб
for m ← 1 to m2 do
   for n ← 1 to n2 do
      if A[f[m]][g[n]] ≠ B[m][n]
         return false // подматрицата на A е различна от B
return true
```

Най-лошият случай за времето на алгоритъма е, когато A съдържа подматрица, еднаква на B с точност до разместване на редове и разместване на стълбове. Времевата сложност на алгоритъма в най-лошия случай е равна по порядък на

$$m_2 + n_2 + (m_2)^2 + (n_2)^2 + m_2\, n_2 \;=\; \mathrm{O}((m_1\, n_1 + m_2\, n_2)^2),$$

където степенуваният сбор в дясната страна е порядъкът на дължината на входа (без дължината на сертификата). С други думи, времевата сложност е не повече от квадратична, следователно е полиномиална. Сертификатът е къс, защото

$$m_2 + n_2 \;=\; \mathrm{O}(m_1\, n_1 + m_2\, n_2),$$

тоест дължината му не надхвърля дължината на входа (достатъчно е дължината на сертификата да не надвишава някакъв полином от дължината на останалата част от входа).

С това доказахме, че алгоритмичната задача “Подматрица” е от класа NP. По-горе установихме, че тази задача е NP-трудна. Следователно тя е NP-пълна.

СХЕМА ЗА ТОЧКУВАНЕ

Цялото контролно носи 5 точки, разпределени по следния начин:
— за първата редукция (описание, коректност и бързина): 2 точки;
— за втората редукция (описание, коректност и бързина): 2 точки;
— за проверката на предложено решение (сертификат): 1 точка.