O złożoności obliczeniowej

Dr Leszek PLASKOTA

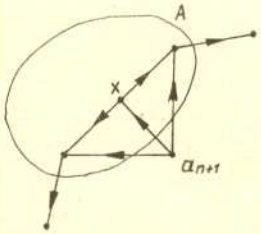
Wielu z nas nie zdaje sobie sprawy z przydatności tej czy innej metody rozwiązywania jakiegoś problemu, dopóki sami nie jesteśmy zmuszeni do rozwiązania konkretnego zadania. Nie jest źle, gdy znana nam metoda jest stosunkowo prosta i mało kosztowna. Gorzej jednak, jeśli takiej metody nie znamy. Wtedy zaczynamy się często zastanawiać, czy nasze zadanie w ogóle można rozwiązać tanim kosztem. Negatywna odpowiedź na to pytanie oznacza, że złożoność zadania jest „duża". Właśnie złożoność (złożoność obliczeniowa) zadań, rozumiana jako minimalny koszt potrzebny do znalezienia rozwiązania, będzie przedmiotem naszych rozważań.

Zauważmy przede wszystkim, że złożoność zadania jest własnością zadania. Dlatego nie należy jej mylić ze złożonością (kosztem) konkretnej metody. Zilustrujemy to na przykładzie rozwiązywania oznaczonego układu n równań z n niewiadomymi. W tym przypadku za miarę złożoności metody przyjmujemy liczbę wykonanych operacji arytmetycznych, takich jak dodawanie, odejmowanie, mnożenie i dzielenie. Znana, teoretyczna metoda używająca wyznaczników wymaga wykonania co najmniej $n!$ operacji arytmetycznych. Nawet dla niewielkich n liczba $n!$ jest tak duża, że koszt metody wyznacznikowej jest kolosalny. Dla przykładu, jeśli dostępny nam komputer wykonuje milion operacji arytmetycznych na sekundę, to rozwiązanie układu z $n = 20$ niewiadomymi trwałoby ponad sto tysięcy lat (!). Dlatego też w praktyce obliczeniowej stosuje się najczęściej metody wymagające około n^3 operacji arytmetycznych. Najbardziej znaną jest algorytm eliminacji Gaussa. Metody o złożoności proporcjonalnej do n^3 były (i są nadal!) tak powszechnie stosowane, iż niektórzy zaczęli wierzyć, że szybsze metody w ogóle nie istnieją. Dopiero stosunkowo niedawno, bo w 1969 roku, za sprawą Strassena i później innych matematyków okazało się, że można konstruować metody wyraźnie szybsze, przynajmniej dla dużych n. Jednak do dziś nie wiadomo, jaka jest dokładnie złożoność rozwiązywania układu równań liniowych. Najlepsze znane dolne ograniczenie złożoności wynosi n^2.

Zapewne każdy z nas grał kiedyś w „dwadzieścia pytań". Gra polega na takim zadawaniu pytań, aby jak najszybciej odgadnąć pomyślany przez kogoś przedmiot. W naturalny sposób można tę grę następująco sformalizować.

Niech f będzie liczbą całkowitą, o której na początku wiemy tylko, że należy do zbioru $F = \{1, 2, \ldots, n\}$. Zakładamy, iż możemy zadawać pytania w rodzaju: „czy liczba f należy do zbioru A?", gdzie A jest pewnym podzbiorem zbioru F. Zakładamy również, że na każde pytanie otrzymujemy prawdziwą odpowiedź („tak" lub „nie"). Należy odgadnąć liczbę f. Za miarę złożoności metody przyjmujemy maksymalną liczbę zadanych pytań, potrzebnych do odgadnięcia liczby f. Jaka jest złożoność tego zadania?

Rozwiązanie zadania M 551.
Zastosujmy indukcję. Sprawdzenie dla $n = 1$ jest oczywiste. Załóżmy teraz, że każdy układ n punktów $a_1, \ldots, a_n$ ma centrum. Rozpatrzmy punkty $a_1, \ldots, a_{n+1}$. Niech x będzie centrum dla $a_1, \ldots, a_n$. Rozpatrzmy A – zbiór punktów, do których można dojść z x w jednym kroku. Są możliwe dwa przypadki:
1. Istnieje strzałka, idąca z x do a_{n+1}; wtedy x jest centrum dla $a_1, \ldots, a_{n+1}$;
2. Każda strzałka łącząca a_{n+1} i $z \in A$ idzie od a_{n+1} do z – wtedy a_{n+1} jest centrum dla $a_1, \ldots, a_{n+1}$ (patrz rysunek).



Rozwiązanie zadania M 552.
Nie. Łatwo obliczyć r z następującej zależności

$$r(1+p)^{n-1} + r(1+p)^{n-2} + \ldots$$
$$\ldots + r = k(1+p)^n,$$

skąd

$$r = kp[1-(1+p)^{-n}].$$

Jak widać, przy $n \to \infty$ wielkość raty zmierza do $k \cdot p$.

Rozwiązanie zadania F 274.
Przy zadanej temperaturze atomy Na znajdują się w stanie $^2S_{1/2}$. Rzut wektora momentu magnetycznego atomu na kierunek pola wynosi

$$\mu_z = m_I g \mu_B,$$

gdzie $m_I = \pm 1/2$, $g = 2$ (czynnik Landego), μ_B – magneton Bohra. Siła rozszczepiająca wiązkę atomów lecących wzdłuż osi y, wynosząca

$$F_z = \pm \mu_z \frac{dB}{dx},$$

w ciągu czasu $t = l/v$ przyspieszy atomy o masie M do prędkości

$$v_z = \pm \mu_z \frac{dB}{dx} \frac{l}{vM}.$$

powodując przesunięcie na ekranie detektora o

$$x = \pm \mu_z \frac{dB}{dx} \frac{l(L+l/2)}{Mv^2}.$$

Energia kinetyczna atomów wynosi

$$Mv^2/2 = (3/2)kT.$$

Stąd poszukiwana odległość wynosi

$$\Delta = 2x = 2\mu_B \frac{dB}{dx} \frac{l(L+l/2)}{3kT} \approx 2\text{cm}.$$

Narzucającą się metodę znalezienia f opiszemy w następujący sposób. Załóżmy, że po zadaniu k pytań wiemy, że f należy do zbioru $A \subset F$. Jeśli A jest zbiorem jednoelementowym, to $A = \{f\}$. Jeśli nie, to zbiór A dzielimy na dwa rozłączne podzbiory A_1 oraz A_2 o liczebności różniącej się co najwyżej o jeden i zadajemy $(k+1)$-sze pytanie: „czy f należy do A_1?". Postępujemy tak dalej ze zbiorem A_1, jeśli otrzymaliśmy odpowiedź „tak", lub ze zbiorem A_2 w przeciwnym przypadku.

Łatwo zauważyć, że kładąc na początku $A = F$ znajdziemy w ten sposób liczbę f po zadaniu nie więcej niż $p(n) = \lceil \log_2 n \rceil$ pytań ($\lceil a \rceil$ oznacza najmniejszą liczbę całkowitą nie mniejszą od a), oraz że dla niektórych f będziemy musieli zadać dokładnie $p(n)$ pytań. A więc $p(n)$ jest złożonością opisanej metody. Czy istnieje lepsza? Nie. Aby się o tym przekonać, wystarczy zauważyć, że po zadaniu k pytań zbiór F dzieli się na co najwyżej 2^k rozłącznych podzbiorów elementów nierozróżnialnych ze względu na dotychczas zadawane pytania. Wobec tego dla $k < p(n)$ jeden z tych podzbiorów, powiedzmy B, jest co najmniej dwuelementowy. A więc w przypadku, gdy szukana liczba należy do B, nie wskażemy na nią za pomocą k lub mniej pytań. W ten sposób pokazaliśmy, że nasza metoda jest najtańsza, czyli optymalna, a jej złożoność jest złożonością zadania.

Rozpatrzone przez nas zadania mogły być rozwiązane dokładnie kosztem skończonym. W praktyce jednak często spotykamy zadania o złożoności nieskończonej lub tak dużej, że przekracza ona nasze możliwości obliczeniowe. W takim przypadku musimy się zadowolić jedynie rozwiązaniem przybliżonym. Na przykład, przy rozwiązywaniu układów równań liniowych z liczbą niewiadomych rzędu dziesięciu tysięcy eliminacja Gaussa może okazać się zbyt kosztowna. Z pomocą przychodzą wtedy tak zwane metody iteracyjne, szczególnie przydatne w przypadku układów rozrzedzonych, czyli takich, w których tylko część współczynników w każdym równaniu jest niezerowa. Stosując metody iteracyjne zyskujemy na czasie, ale musimy zadowolić się rozwiązaniem przybliżonym.

Dla zadań o dużej lub nieskończonej złożoności obliczania rozwiązania dokładnego wygodnie jest wprowadzić pojęcie ε-złożoności. Jest to minimalny koszt potrzebny do znalezienia rozwiązania z dokładnością ε. Oczywiście, wymaga to zdefiniowania w jakiś sposób błędu metody.

Dla ilustracji rozpatrzmy jeszcze raz grę w „dwadzieścia pytań", tym razem jednak ze zbiorem $F = (0, 1)$. Łatwo zauważyć, że teraz złożoność zadania jest nieskończona. A jaka jest ε-złożoność? Zanim odpowiemy na to pytanie, wyjaśnijmy najpierw, co będziemy rozumieli przez metodę, jej błąd oraz koszt. Ponieważ możemy pytać tylko o przynależność szukanej liczby do jakiegoś podzbioru, każdą metodę można rozbić na dwa etapy. W pierwszym zbieramy pewną informację o f przez zadawanie pytań. Uzyskaną informację o f, będącą ciągiem odpowiedzi „tak" lub „nie", oznaczymy krótko przez $N(f)$. W drugim etapie, na podstawie informacji $N(f)$ konstruujemy w jakiś sposób przybliżenie $U(f) = \Phi(N(f))$ dla szukanej liczby f. Każda metoda U jest więc złożeniem operatora informacji N z operatorem Φ, zwanym algorytmem (nazwa jest w pełni uzasadniona, gdyż Φ przetwarza uzyskane dane o zadaniu). Błąd metody U zdefiniujemy przez najgorsze jej zachowanie. A więc:

$$e(U) = \sup_{f \in F} |f - U(f)|.$$

Przez koszt metody U ($cost(U)$) rozumiemy maksymalną liczbę zadanych pytań, a więc obliczenie $y = N(f)$. Formalnie powinniśmy włączyć do kosztu całkowitego koszt obliczenia przybliżenia $\Phi(y)$. Okazuje się jednak, że jest on zaniedbywalnie mały w porównaniu z kosztem zadania jednego pytania. Zgodnie z definicją ε-złożoność zadania, $comp(\varepsilon)$, można zapisać w następujący sposób:

$$comp(\varepsilon) = \inf cost(U),$$

przy czym infimum wzięte jest po wszystkich metodach U o błędzie $e(U)$ nie przekraczającym ε.

Po tej porcji definicji możemy już wykazać, że ε-złożoność naszego zadania wynosi $q(\varepsilon) = \lceil \log_2(1/\varepsilon) \rceil - 1$. W tym celu zauważmy, że tak jak w przypadku dyskretnym, po zadaniu nie więcej niż k pytań zbiór F zostaje podzielony na co najwyżej 2^k rozłącznych podzbiorów nierozróżnialnych ze względu na zadawane pytania. Wobec tego jeden z tych podzbiorów, nazwijmy go B, ma średnicę nie mniejszą od $1/2^k$. Dlatego dla dowolnej liczby g mamy

$$\sup_{f \in B} |f - g| \geq 1/2^{k+1}.$$

A więc każda metoda korzystająca z nie więcej niż k pytań daje błąd co najmniej $1/2^{k+1}$. W celu uzyskania błędu nie większego od ε musimy więc zadać co najmniej $q(\varepsilon)$ pytań. Aby zakończyć dowód, wystarczy teraz wskazać metodę, która osiąga żądaną dokładność wyniku i korzysta z $q(\varepsilon)$ pytań. Metodę tę zapiszemy w następujący sposób (porównaj z przypadkiem dyskretnym):

początek połóż $a_0 = 0$, $b_0 = 1$, $c = 1/2$;
dla $k = 1, 2, ..., q(\varepsilon)$ **wykonuj**
początek zadaj pytanie:
„czy f jest mniejsza od c?";
jeśli uzyskałeś odpowiedź „tak", to
połóż $a_k = a_{k-1}$, $b_k = c$, w przeciwnym przypadku
połóż $a_k = c$, $b_k = b_{k-1}$;
połóż $c = (a_k + b_k)/2$
koniec;
połóż $U(f) = c$
koniec.

Przykład gry w „dwadzieścia pytań", chociaż stosunkowo prosty, jest jednak dość typowy. Występują w nim wszystkie podstawowe elementy modelu złożoności obliczeniowej, takie jak: informacja, algorytm, błąd i koszt metody czy w końcu ε-złożoność. Oczywiście, błąd metody można zdefiniować na różne sposoby. Dla przykładu, można wprowadzić błąd względny,

$$e_u(U) = \sup_{f \in F} |f - U(f)|/f.$$

Jednak dla $\varepsilon < 1$ złożoność naszego zadania jest wtedy nieskończona. Podobnie, przyjęty przez nas sposób oceny metody przez najgorsze jej zachowanie nie jest jedyny. Można na przykład zdefiniować błąd i koszt średni metody, co prowadzi do tak zwanego modelu przypadku średniego.

Na koniec, w kontekście naszego przykładu, zwrócimy jeszcze uwagę na pewien aspekt złożoności dotyczący sposobu uzyskiwania informacji o szukanym elemencie. Zauważmy, że kolejne pytania dowolnej metody oraz ogólna ich liczba mogły istotnie zależeć od otrzymywanych „po drodze" odpowiedzi. Dopuszczaliśmy więc tak zwaną informację adaptacyjną. Problem istnienia metod optymalnych korzystających z informacji nieadaptacyjnych ma w ogólności duże znaczenie praktyczne. Nieadaptacja umożliwia bowiem równoległe uzyskiwanie kolejnych porcji informacji. Właśnie obliczenia równoległe są ostatnio bardzo szybko rozwijającą się gałęzią informatyki. W grze w „dwadzieścia pytań" okazuje się, że informacja, z której korzystała wskazana przez nas metoda optymalna, jest równoważna nieadaptacyjnym pytaniom o kolejne bity rozwinięcia dwójkowego szukanej liczby. A więc: „czy na i-tym miejscu po przecinku w rozwinięciu dwójkowym liczby f stoi zero?", dla $i = 1, 2, \ldots, q(\varepsilon)$. Na przykład, dla $f = 3/5$ otrzymalibyśmy ciąg odpowiedzi „nie", „tak", „tak", „nie", itd. Podobną informację nieadaptacyjną można wskazać dla dyskretnego zbioru F. Zainteresowanym proponujemy jeszcze zastanowienie się nad różnicą między adaptacją i nieadaptacją w przypadku, gdy wolno nam zadawać jedynie pytania typu „czy f jest mniejsza od g", gdzie g jest dowolną liczbą z przedziału $(0, 1)$.

PODWÓJNY ROZPAD BETA

Od ponad 50. lat badanie rozpadów beta jąder atomowych dostarcza informacji o strukturze materii jądrowej. W tym rozpadzie jeden z neutronów jądra rozpada się na trzy cząstki: proton, elektron i antyneutrino. Jedną z bardzo istotnych, nieznanych własności neutrin jest ich masa spoczynkowa. Zwykle przyjmuje się, że neutrino jest cząstką o masie zero, ale w większości teorii tzw. wielkiej unifikacji przewiduje się, że neutrino powinno mieć masę. Masywne neutrina są również dyskutowane w kontekście tzw. czarnej materii we Wszechświecie. Dlatego pomiar masy neutrina mógłby w istotny sposób ograniczyć obecne spekulacje teoretyczne.

Badanie widma elektronów w rozpadach beta pozwala w zasadzie wyznaczyć masę neutrin m_ν (maksymalna energia elektronów = energia rozpadu $- m_\nu c^2$). Ale błędy doświadczalne i teoretyczne (wynikające z niepełnej znajomości efektów jądrowych) pozwalają jedynie podać górną granicę na masę neutrin elektronowych $m_\nu \leq 25$ eV (grupa z Zurichu podaje nawet $m_\nu < 18$ eV). Jedynie grupa doświadczalna z Moskwy podaje wynik niezerowy dla masy neutrin elektronowych $m_\nu = 26 \pm 6$ eV. Inną interesującą zagadką jest to, czy neutrino jest istotnie różne od swojej antycząstki (tzw. neutrino Diraca), czy też są to raczej dwa różne stany spinowe tej samej cząstki (tzw. neutrino Majorany). Już prawie 50 lat temu zauważono, że na oba pytania (o naturę i masę neutrin) można szukać odpowiedzi w badaniach tzw. podwójnego rozpadu beta. Zjawisko to zachodzi wówczas, gdy zwykły rozpad beta jest zabroniony energetycznie lub silnie stłumiony, natomiast rozpad podwójny jest dozwolony. Model standardowy oddziaływań z neutrinami dirakowskimi przewiduje w takim rozpadzie emisję dwóch antyneutrin, co spowoduje, że widmo energetyczne dwóch elektronów w stanie końcowym będzie szerokie. Jeśli neutrina są typu Majorany, to możliwy jest rozpad beta, gdzie neutrino wyemitowane w jednym rozpadzie może być pochłonięte w drugim. Wówczas widmo energetyczne elektronów będzie miało bardzo silne maksimum przy energii równej energii rozpadu. Taki pomiar nie jest łatwy, gdyż procesy takie są bardzo rzadkie. Czas życia jąder rozpadających się poprzez podwójny rozpad beta jest rzędu 10^{20} lat! (Dla porównania, czas życia Wszechświata ocenia się na 10^{10} lat.)

W latach 60. dokonano istotnego przełomu uzyskując wyniki pośrednio z pomiarów geochemicznych, w ktorych mierzono zawartość w starych skałach izotopów ksenonu i kryptonu pochodzących z podwójnego rozpadu ^{130}Te i ^{82}Se. W sierpniu 1987 r. podwójny rozpad beta został zaobserwowany po raz pierwszy bezpośrednio w laboratorium Uniwersytetu Kalifornijskiego w Irvine. Zaobserwowano wyraźnie ślady dwóch elektronów wylatujących z próbki selenu 82. Do tej pory nie zauważono przypadków, w których suma energii elektronów równa byłaby energii rozpadu – tzn. przypadków bez emisji neutrin. Obecny stan wyników doświadczeń i analizy teoretycznej można podsumować w następujący sposób: jeśli neutrina są cząstkami Majorany, to ich masa jest mniejsza od 1 eV.

J. K.