Informatyczny kącik olimpijski (31): Diamenty

Zadanie, któremu przyjrzymy się tym razem, nosi nazwę *Diamenty*.

Mając danych n diamentów, z których i-ty ma masę m_i i wartość w_i (wyrażone liczbami całkowitymi), należy z nich wybrać dokładnie k diamentów $i_1, i_2, \ldots, i_k$ o najlepszym współczynniku wartości, który wyraża się wzorem:

$$\frac{w_{i_1} + w_{i_2} + \ldots + w_{i_k}}{m_{i_1} + m_{i_2} + \ldots + m_{i_k}}.$$

Pierwszymi przychodzącymi do głowy rozwiązaniami są algorytmy zachłanne: brać k o największej wartości, albo o najlepszym stosunku $\frac{w_i}{m_i}$, albo jeszcze wedle jakiegoś innego kryterium. Pomysły te są niepoprawne, czego sprawdzenie pozostawiam Czytelnikowi. To, który diament opłaca się wziąć, zależy bowiem od uprzednio wybranych diamentów. Nie działa także (choć trudniej wskazać kontrprzykład – jak?) pomysł, aby po każdym kolejno wybranym j-tym diamencie aktualny współczynnik

$$\frac{w_{i_1} + w_{i_2} + \ldots + w_{i_j}}{m_{i_1} + m_{i_2} + \ldots + m_{i_j}}$$

był najlepszy z możliwych.

Aby uzyskać poprawne rozwiązanie, można natomiast użyć programowania dynamicznego. Wystarczy dla każdej liczby diamentów d i każdej ich masy m obliczać maksymalną łączną wartość odpowiadającą temu zestawowi parametrów, dodając po jednym diamencie do zbioru już przetworzonych. Jeśli t jest tablicą wyników przed dodaniem i-tego diamentu, to wyniki t' po dodaniu tegoż diamentu są następujące:

$$t'[d][m] = \max(t[d][m], t[d-1][m-m_i] + w_i),$$

przy czym należy pamiętać, że drugi człon bierzemy pod uwagę tylko, gdy $d-1$ i $m-m_i$ są nieujemne. Po przetworzeniu wszystkich diamentów wynikiem jest największa z liczb postaci $\frac{t[k][i]}{i}$. Niestety, to rozwiązanie istotnie zależy od łącznej masy M wszystkich diamentów i ma złożoność czasową $O(n \cdot k \cdot M)$, czyli $O(n^2 \cdot M)$.

W tym miejscu chciałbym zachęcić Czytelnika do zastanowienia się, przed przeczytaniem reszty artykułu, jak lepiej rozwiązać ten problem. Nie jest bowiem łatwo o dobry pomysł na poprawienie tego wyniku.

W rozwiązaniu zastosujemy kluczowe spostrzeżenie, że można do wyznaczenia wartości żądanego współczynnika wykorzystać wyszukiwanie binarne. Aby tak zrobić, musimy mieć sposób na sprawdzanie, czy wybierając k diamentów, możemy uzyskać współczynnik nie mniejszy od danego. Konkretnie, dla danego x, chcemy umieć stwierdzić, czy istnieją takie diamenty, że

$$\frac{w_{i_1} + w_{i_2} + \ldots + w_{i_k}}{m_{i_1} + m_{i_2} + \ldots + m_{i_k}} \geqslant x.$$

Na pierwszy rzut oka nie wydaje się to łatwiejsze niż wyjściowy problem, ale faktycznie jest. Możemy przecież teraz pomnożyć obie strony przez mianownik, aby pozbyć się niewygodnego dzielenia! Otrzymujemy kolejno równoważne warunki:

$$w_{i_1} + w_{i_2} + \ldots + w_{i_k} \geqslant (m_{i_1} + m_{i_2} + \ldots + m_{i_k}) \cdot x,$$
$$(w_{i_1} - m_{i_1}x) + (w_{i_2} - m_{i_2}x) + \ldots + (w_{i_k} - m_{i_k}x) \geqslant 0.$$

Sprowadziliśmy zatem nasz warunek do takiego, który jesteśmy w stanie łatwo sprawdzić. Wystarczy, że obliczymy dla każdego diamentu

$$u_i = w_i - m_i x$$

i zdecydujemy, czy wśród liczb u_i jest k takich, których suma jest nieujemna. Jak to sprawdzić? Wystarczy wziąć k największych liczb u_i i sprawdzić, czy ich suma jest nieujemna (gdyż jest to największa możliwa do uzyskania suma). Warto zaznaczyć, że te k największych liczb u_i niekoniecznie reprezentuje diamenty, które maksymalizują współczynnik wartości, a jedynie takie, które dowodzą, że da się osiągnąć pewien końcowy współczynnik.

W ten sposób możemy wyszukać ten współczynnik z pewną dokładnością, ale jak dowiedzieć się, które diamenty wchodzą w skład najlepszego zestawu? Wystarczy zastosować ten sam schemat dla x równego największemu znalezionemu współczynnikowi i wybrać te k diamentów, które odpowiadają k największym liczbom u_i. Owe diamenty dają wynik co najmniej x, a więc dokładnie tyle, ile chcemy.

Na koniec zastanówmy się jeszcze nad złożonością takiego rozwiązania. Niewątpliwie, w każdym kroku wyszukiwania binarnego należy posortować n liczb, a więc taki krok będzie miał złożoność $O(n \log n)$. Ile jednak takich kroków będzie? Poszukiwania rozpoczynamy od przedziału długości W (suma wszystkich wartości diamentów), natomiast kontynuujemy je tak długo, aż długość przedziału spadnie poniżej najmniejszej możliwej odległości między dwoma różnymi wynikami, czyli najmniejszej liczby, o którą mogą różnić się ułamki o mianownikach nie większych niż M, która to liczba jest rzędu $O(\frac{1}{M^2})$. Ostatecznie rozwiązanie ma złożoność czasową $O(\log(M^2 W) \cdot n \log n)$, czyli równoważnie $O(n \log n \cdot (\log M + \log W))$.

Tomasz KULCZYŃSKI