Algorytmy i struktury danych 2020
Lista 4
Zadanie 8

Kwiecień 2020

1 Treść

Doktor Zseimel postanowił zaciągnąc się do pracy na platformę wiertniczą. Niestety został zakwalifikowany do kategorii osób o statusie "overqualified" i nie dostał wymarzonej pracy. Zaproponowano mu natomiast zostanie konsultantem strategicznym koncernu naftowego, który planuje budowę dużego rurociągu przebiegającego z zachodu na wschód przez pola naftowe, na których znajduje się n wież wiertniczych. Do kazdej wieży ma dochodzić odnoga głównego rurociągu po najkrótszej możliwej drodze (albo na północ, albo na południe) (zakładamy, że główny rurociąg będzie modelowany przez prostą, a odchodzące od niego odnogi będa prostymi podłączonymi do głównego rurociągu pod kątem prostym). Jakiego algorytmu powinien użyć doktor dla zadanych współrzędnych (x_i, y_i) $i = 1 \ldots n$ wież, aby wyznaczyć optymalne położenie głównego rurociągu (czyli takie, dla którego suma dlugosci odnóg jest minimalna)?
Wykaż, że można takie położenie wyznaczyć w czasie liniowym.

2 Rozwiązanie

Do rozwiązania potrzebne będą tylko wartości drugiej współrzędnej y_i. Odpowiedzią będzie także współrzędna y, na której powinien byc wybudowany rurociąg. Niech wyznaczone $y = y_s$, a suma odległości to s, tj. $\sum_{i=1}^{n} \|y_s - y_i\|$ Będę udowadniać, że y_s to mediana współrzędnych wejściowych, a dokładnie:

- Jeśli liczba wież n jest **parzysta**, szukanym y_s będzie dowolna wartość między górną a dolną medianą.
- Jeśli liczba wież n jest **nieparzysta**, szukanym y_s będzie y mediany.

Spróbujmy odtworzyć wyznaczanie y_s. Na początku zaczynamy od pierwszej wieży. W pierwszym kroku $y_s = y_1$, wtedy $s = 0$. Przy dwóch wieżach, najmniejszą sumą odległości będzie różnica między y_1 a y_2, niezależnie od wybranego $y_1 \leqslant y \leqslant y_2$ (lub $\geqslant$, jeśli $y_2 \geqslant y_1$). Musi być spełniony jedynie warunek $d_1 + d_2 = \|y_1 - y_2\|$, gdzie $d_1 = \|y_1 - y\|$ (odległość 1) i $d_1 = \|y_1 - y\|$ (odległość 2).
Dla wszystkich przypadków parzystych: załóżmy, że początkowo y_s wybieramy na wysokości górnej mediany. Wtedy suma odległości będzie sumą odległości $s_1 = n/2 - 1$ wież wyżej i $s_2 = n/2$ wież niżej, łącznie z dolną medianą. Jeżeli wybralibyśmy y rurociągu wyżej od górnej mediany o wartość k, wtedy dla wszystkich $n/2-1$ wież o wyższej współrzędnej y wartość zmalałaby o k, więc $ss_1 \geqslant s_1 - (n/2-1)*k$. "$\geqslant$", ponieważ, k może być na tyle duże, że odległości od wież o wyższym y niż mediana także mogą wzrosnąć. Równocześnie, dla wież poniżej rurociągu wartość s_2 wzrośnie o $k*(n/2+1)$, ponieważ trzeba doliczyć odległość do górnej mediany, więc $ss_2 = s_2 + (n/2+1)*k$. Widzimy więc, że $ss_1 + ss_2 = s_1 + s_2 + k > s_1 + s_2$. Natomiast zmieniając y_s rurociągu niżej od górnej mediany, ale nie niżej od dolnej, odległości rurociągu od $n/2$ wież wyżej wszystkie wzrosną o różnicę k, a odległości od $n/2$ wież niżej wszystkie zmaleją o różnicę k. $k*(n/2) - k*(n/2) = 0$, więc zmiana wysokości y_s między wartościami y dwóch median, nie powoduje wzrostu sum odległości. Analogiczny dowód przeprowadzamy dla dolnej mediany.
Jeżeli liczba wież jest nieparzysta, zakładam, że szukane y_s jest na wysokości mediany wszystkich y wież.

Wtedy dla $(n-1)/2$ wież o wyższym y suma odległości to s_1. Jeżeli przesunęlibyśmy rurociąg w górę o k, otrzymalibyśmy $s' = s - k(n-1)/2 + k(n+1)/2 = s + k > s$.
Do wyznaczania mediany w czasie liniowym można użyć algorytmu **Select (median of medians)**.