HCPC 勉強会 2017/03/28 ー座標圧縮ー

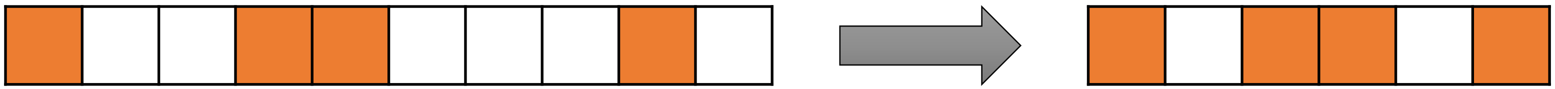
担当：D1 鈴木 浩史

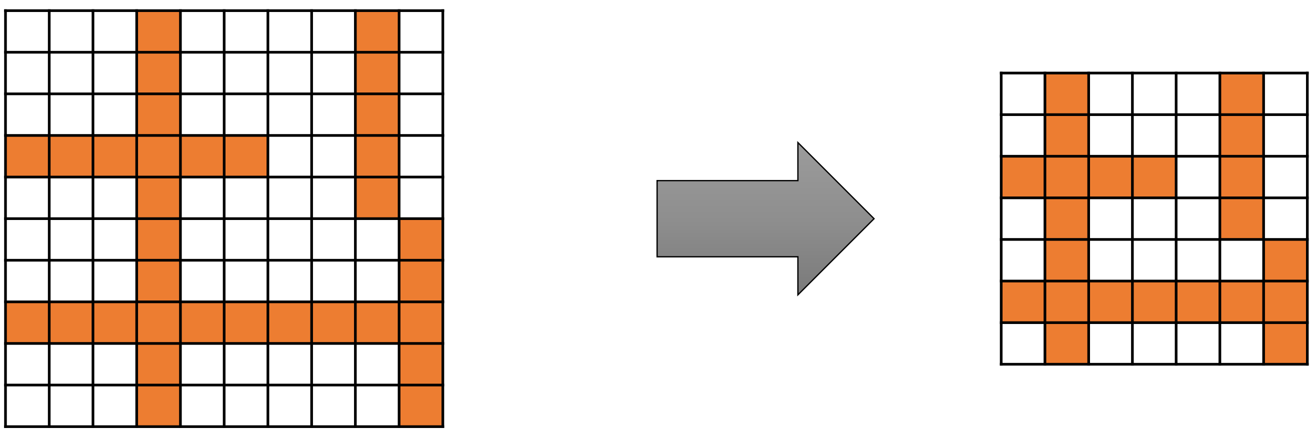
座標圧縮とは

- 広大な領域を対象とする問題はコンピュータの手におえない
  - 例①：10^{18}の直線領域はメモリに乗らない
  - 例②：$10^9 \times 10^9$の格子領域も同様
    - 単位領域あたり1bitで持っても~10^5TB

- それ，不要な領域，ありませんか？
  - はいプロ 世界一メモリ節約が上手 メモリ界のtourist メモリ不足に終焉を告げる者 実質**座圧** メモリ節約のために生まれてきた男

座圧の感覚





座圧入門（1次元）

ABC036 C問題「座圧」

- 長さnの数列$a = (a_1 \dots a_n)$が与えられるので，以下を満たす数列$b = (b_1 \dots b_n)$を求めよ．
  - $b_i \geq 0$
  - $a_i < a_j \rightarrow b_i < b_j$
  - $a_i = a_j \rightarrow b_i = b_j$
  - 各b_iは考えうる中で最小
- 制約
  - $1 \leq n \leq 10^5$
  - $0 \leq a_i \leq 10^9$

$a = (3\ 3\ 1\ 6\ 1)$

$b = (1\ 1\ 0\ 2\ 0)$

※一意に定まる

考察

- bはaの各要素の大小関係を保存する
- ソートしてみるとわかりやすい

$a = (3\ 3\ 1\ 6\ 1)$をソートした列は以下のようになる.

$$a_{sort} = (1\ 1\ 3\ 3\ 6)$$

1を0に置き換えると, a_{sort}の各要素の大小関係は

$$a'_{sort} = (0\ 0\ 1\ 1\ 2)$$

でも表現可能であり, 元の順序に戻すと

$$b = (1\ 1\ 0\ 2\ 0)$$

要するに座圧とは

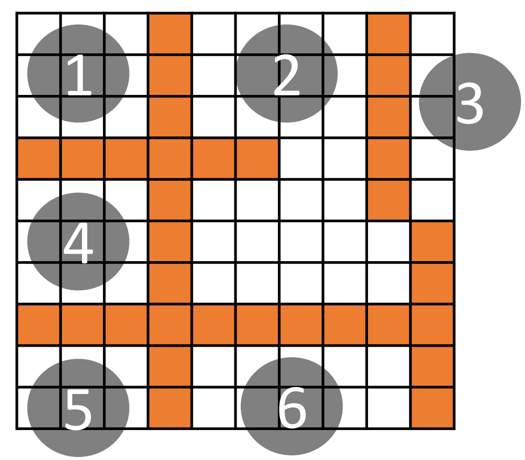
- 各値の大小関係（や位置関係など）を崩さないように,
新しい数値を割当てること
- 座圧後の座標幅は$O(n)$
  ※問題によって定数倍が異なる

- 座圧するにはソートと再割り当てが基本で$O(n\log n)$
- サンプルコードはtsukasa_diaryのgithubに

座圧入門（2次元）

蟻本「領域の個数」

- $w \times h$の格子領域にn本の縦 or 横の直線が引かれている．直線によって区切られている領域の個数を求めよ．



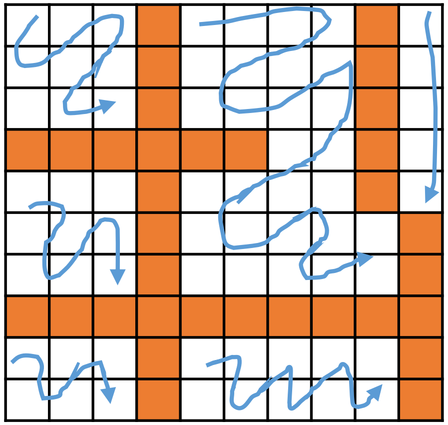
- 制約
  - $1 \leq w, h \leq 10^6$
  - $1 \leq n \leq 500$
  - 直線は始点と終点で表現

領域の数え方

- まだ訪問していない格子があれば探索を開始
  - 直線を超えないようにBFSなど

⇒探索を開始した回数が領域の個数

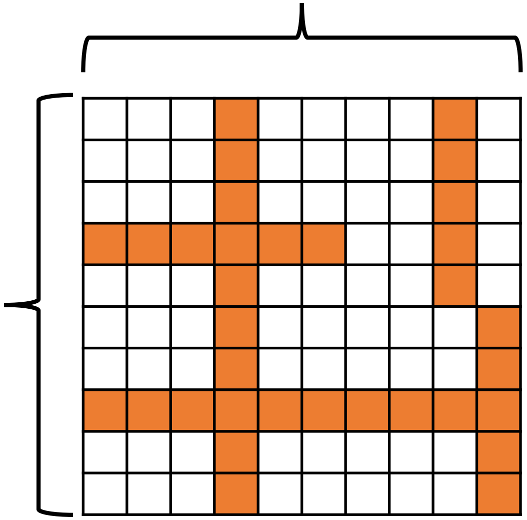
- 右の例では6回探索している
- 領域も6個



問題点

- 座標幅がでかい

最大で$10^6 \Rightarrow 10^6 \times 10^6$の領域はMLE

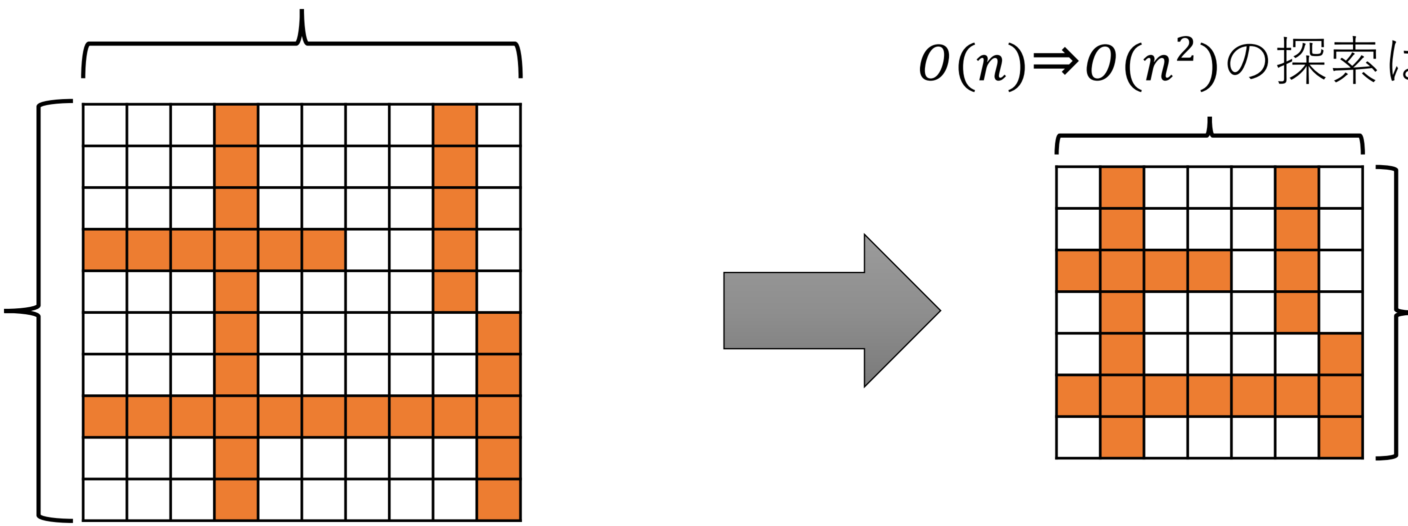


座圧する

- 各端点の座標を使って座圧⇒優勝

最大で10^6 ⇒$10^6 \times 10^6$の領域はMLE

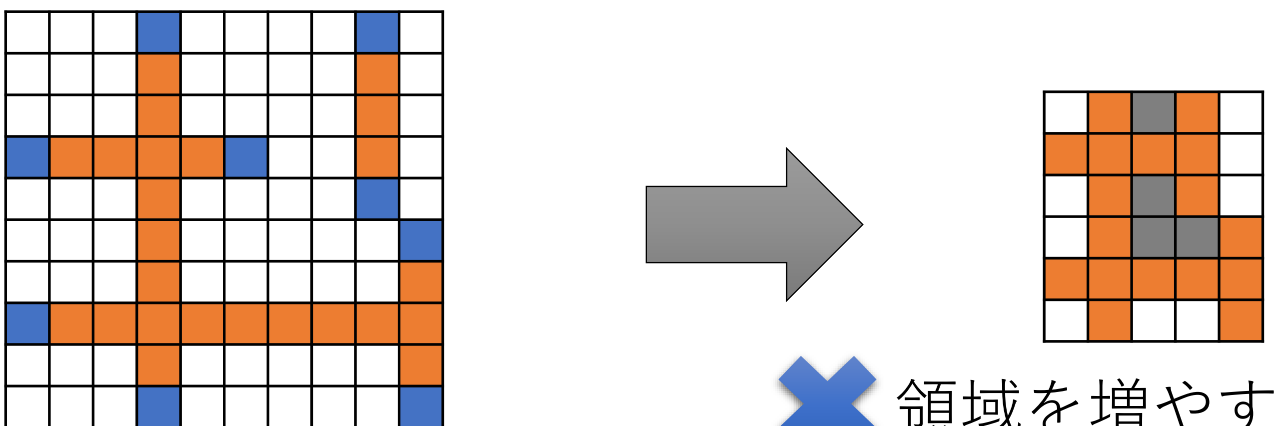
$O(n)$⇒$O(n^2)$の探索は余裕



座圧の際の注意点

- 各軸について，端点の座標，座標の最小値**1**，および最大値w, hを入れて，前問と同様に座圧⇒間違い

x軸について$(1\ 4\ 6\ 9\ 10) \rightarrow (1\ 2\ 3\ 4\ 5)$

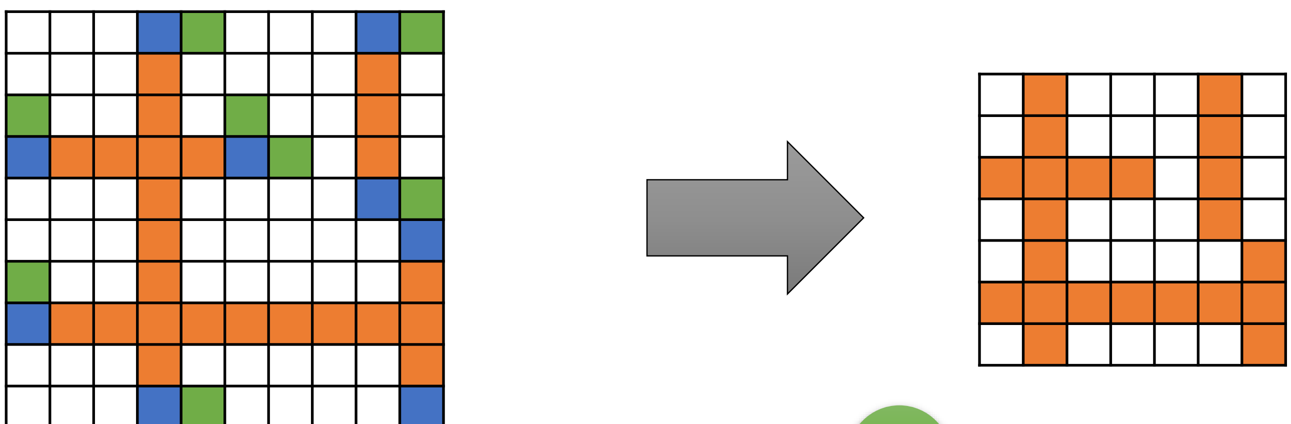


✖ 領域を増やすプロ

当問題の正しい座圧

- さらに端点の右と上に余白座標を追加して座圧し，余白座標以外の相対位置だけ取り出す

x軸について$(1\ 4\ 6\ 9\ 10) \to (1\ 4\ 5\ 6\ 7\ 9\ 10) \to (1\ 2\ 4\ 6\ 7)$



○ 座圧のプロ

サンプルコード

- tsukasa_diaryのgithub（verifyできてません）

座圧の応用

AOJ 2426「Treasure Hunt」

- n個の点と，m個の矩形領域が与えられる．
- 各矩形領域に含まれる点の数を求めよ．

- 制約
  - $1 \le n \le 5000$
  - $1 \le m \le 5 \times 10^5$
  - 矩形領域は軸に平行で左上と右下の座標が入力
  - 座標値は-10^9~10^9ですべて整数

考察

- 各点について各領域に含まれるかどうか検査

⇒$O(nm)$はTLE

- 点のある座標に1，そうでない座標に0を割り当て，2次元累積和を前計算

⇒各矩形領域に対して$O(1)$

⇒累積和の領域が大きい…

⇒おや…？

0	0	1	1	0	0
0	0	0	0	0	0
0	0	1	0	0	0
1	0	0	1	0	0
0	0	0	0	0	0
0	0	1	0	0	1

0	0	1	2	2	2
0	0	1	2	2	2
0	0	2	3	3	3
1	1	3	5	5	5
1	1	3	5	5	5
1	1	4	6	6	7

$(0,0)$と(x, y)による矩形領域内の点数の2次元累積和
⇒任意の矩形領域の点数を$O(1)$で計算可能
※次ページで補足

2次元累積和（補足）

- x軸方向の累積和を取った後，y軸方向の累積和

0	1	1	0
0	1	0	0
1	0	1	0
0	1	0	1

x軸方向

0	1	2	2
0	1	1	1
1	1	2	2
0	1	1	2

y軸方向

0	1	2	2
0	2	3	3
1	3	5	5
1	4	6	7

- 左上が(x_1, y_1)，右下が(x_2, y_2)の矩形領域の和は
  - $sum[y_2][x_2] - sum[y_1-1][x_2] - sum[y_2][x_1-1] + sum[y_1-1][x_1-1]$

0	1	2	2
0	2	3	3
1	3	5	5
1	4	6	7

$$5 - 2 - 1 + 0 = 2$$

座圧した領域で累積和

0	0	1	1	0	0
0	0	0	0	0	0
0	0	1	0	0	0
1	0	0	1	0	0
0	0	0	0	0	0
0	0	1	0	0	1

座圧

0	1	1	0
0	1	0	0
1	0	1	0
0	1	0	1

累積和

0	1	2	2
0	2	3	3
1	3	5	5
1	4	6	7

※座圧してから行列を作る

各矩形領域の点数を知る

- 入力の矩形領域を座圧した領域に写像
  - 左上の点で，各軸についてlower_boundとなる点を探す
  - 右下の点で，各軸についてupper_boundとなる点を探す
  - その元座標の座圧後の座標に写像

0	0	1	1	0	0
0	0	0	0	0	0
0	0	1	0	0	0
1	0	0	1	0	0
0	0	0	0	0	0
0	0	1	0	0	1

写像

0	1	1	0
0	1	0	0
1	0	1	0
0	1	0	1

0	1	2	2
0	2	3	3
1	3	5	5
1	4	6	7

サンプルコード

- tsukasa_diaryのgithub $O(n^2 + (n + m) \log n)$

おしまい

しっかり練習して座圧のプロになろう！