第１２回：１２月２３日（水）

12　第１２回

今回は動的計画法の他の問題への応用について学ぶ.

12.1　文書整形問題に対する動的計画法

文書整形問題 (改行位置決定問題)

与えられた文章の一つの段落を,

- 右端をそろえて,
- 単語間 (あるいは文字間) のスペースをなるべく理想的な幅にし,
- 単語をハイフンで分割することによる改行をできるだけ少なくなるように

整形する問題を考える. この問題は組版に応用を持つ [1]. 文書整形の例を「講義資料と課題」2 ページに示す.

ここでは単語の分割は行わず, 単語間のスペースをどのように取ればよいかのみを検討する. そこで, 理想のスペース幅からの乖離をコストとし, コストの総和が最小となるようなレイアウト (改行位置) を決定する問題を考える [2].

与えられた文章には n 個の単語 $w_1, \ldots, w_n$ がこの順に現れるものとし, その幅をそれぞれ $\ell_1, \ldots, \ell_n$ とする. またテキストの幅を L とする. 単語 w_i から w_j $(1 \leq i \leq j \leq n)$ を一行にレイアウトするときのスペース幅 $s(i,j)$ を次式で定める.

$$s(i,j) = \frac{L - \sum_{k=i}^{j} \ell_k}{\max\{j-i, 1\}}.$$

理想のスペース幅 s を想定し, 単語 w_i から w_j を一行にレイアウトすることのコスト $c(i,j)$ を次のように定義する.

$$c(i,j) = \begin{cases} +\infty & s(i,j) \leq 0 \text{ のとき}, \\ (s - s(i,j))^2 & s(i,j) > 0 \text{ かつ } j < n \text{ のとき}, \\ \max\{s - s(i,j), 0\}^2 & s(i,j) > 0 \text{ かつ } j = n \text{ のとき}. \end{cases}$$

最後の $\max\{s - s(i,j), 0\}^2$ は，上で定めたスペース幅 $s(i,j)$ が理想のスペース幅 s を超えるならば，スペース幅は s とするのが理にかなっていることによる.

改行位置決定問題

入力: n 個の単語 $w_1, \ldots, w_n$ の幅 $\ell_1, \ldots, \ell_n$, 理想のスペース幅 s, テキストの幅 L.

出力: 総コスト最小のレイアウト.

部分問題の定義

整数 $i = 1, 2, \ldots, n$ に対して
$f^*(i) \triangleq \quad w_i, \ldots, w_n$ をレイアウトしたときの最小コスト
と定義する．我々が知りたいのは $f^*(1)$ とそれを実現する改行位置である．

再帰式の導出

最後の単語 w_n だけを持つ問題が最もサイズの小さな部分問題を与え，このときが境界条件となる．

境界条件:

$$f^*(n) = c(n, n).$$

再帰式:
$i = 1, 2, \ldots, n-1$ に対しては，最適なレイアウトの第一行目に配された単語列を $w_i, w_{i+1}, \ldots, w_j$ とすると，第一行目で生じるコストは $c(i, j)$ であり，残りの単語 $w_{j+1}, w_{j+2}, \ldots, w_n$ で生じる最小コストは $f^*(j+1)$ と表せる．ただし，$j = n$ のときは $w_i, w_{i+1}, \ldots, w_n$ を一行に配したコスト $c(i, n)$ だけが生じる．以上より以下を得る．

$$f^*(i) = \min\{\min_{i \leq j \leq n-1} \{c(i, j) + f^*(j+1)\}, c(i, n)\}.$$

アルゴリズムの設計

前処理として $c(i, j)$ $1 \leq i \leq j \leq n$ を計算しておく．$i = n$ の境界条件 $f^*(n) = c(n, n)$ の計算から始め，再帰式に従って $f^*(i)$ を $i = n-1, n-2, \ldots, 1$ の順に決定していく．$f^*(n)$ が決定できたらバックトラックにより最適な改行位置を求める．

計算量の上界の解析

一つの $f^*(i)$ の計算時間は $O(n)$ であるので，コスト $c(i, j)$ の計算が終わっていれば，$f^*(i), i = 1, 2, \ldots, n$ の計算は $O(n^2)$ 時間で済む．コスト $c(i, j)$ を個々に定義通り計算すると一つあたり $O(n)$ 時間，すべてのコストの計算時間が $O(n^3)$ となるが，$\sum_{k=i}^{j} \ell_k = \sum_{k=i}^{j-1} \ell_k + \ell_j$ を利用すると，

コスト $c(i,j)$ を $O(1)$ 時間できる．これにより，コスト $c(i,j)$ の計算も含め， 全体で $O(n^2)$ 時間で済む．計算領域は $O(n^2)$.

参考文献

[1] 黒木裕介, 組版におけるオペレーションズ・リサーチ, オペレーションズ・リサーチ, Vol. 60, No. 9, 2015, pp. 536–542.

[2] http://www.co.mi.i.nagoya-u.ac.jp/~yagiura/typeset/ (2020 年 9 月 1 日にアクセス確認)

【課題１２－１】改行位置決定問題に次の条件を付加した問題を動的計画法で解くための部分問題の定義，境界条件，再帰式の導出を行え．
「スペース幅が基準値 $s > 0$ から 30% 以上外れる行（違反行）の個数が K 以下」（条件を満たすレイアウトが存在しないこともある）.

12.2 巡回セールスマン問題に対する動的計画法

巡回セールスマン問題は最も有名な NP-完全問題であろう．グラフのすべての点を丁度 1 回通る 1 個の有向閉路をツアー (tour) と呼ぼう．

> 巡回セールスマン問題 (TSP)
> **入力:** 完全有向グラフ $G = (V = \{1, 2, \ldots, n\}, E)$, 各有向枝 $e = (i,j)$ の重み $d_{i,j}$.
> **出力:** G のツアー $C \subseteq E$ の中で重み和を最小にするもの.

部分問題の定義

点の部分集合 $S \subseteq \{2, 3, \ldots, n\} = V \setminus \{1\}$, 点 $i \in S$ に対し

$f^*(S,i) \triangleq$ 点 1 から出発し，S 内の点を丁度 1 回づつ通り点 i へ至る経路の中で距離和の最小値.

と定義する．出発点 s は常に点 $1 \in V$ に固定しておき，$f^*(V \setminus \{1\}, t)$, $t \in V \setminus \{1\}$ が決定できれば，もとの TSP のツアーの最小費用は

$$\min\{f^*(V \setminus \{1\}, t) + d_{t,1} \mid t = 2, 3, \ldots, n\}$$

として求められる．

再帰式の導出

$|S|=1$ のときを境界条件とする．
境界条件：$i=2,3,\ldots,n$ に対し

$$f^*(\{i\},i)=d_{1,i}.$$

$|S|\geq 2$ のときを考える．ここで，$f^*(S,i)$ の最短経路が点 t に入る直前の点を j とすると，点 1 から点 j への経路は部分問題 $f^*(S\setminus\{t\},j)$ の最適解となっている（最適性の原理）．もちろん，どの点が点 i の直前になるかは分からないので全てを試した中で重み和最小のものを選ぶ次の式が再帰式として得られる．
再帰式：点の部分集合 $S\subseteq\{2,3,\ldots,n\}$ ($|S|\geq 2$), 点 $i\in S$ に対し

$$f^*(S,i)=\min\{f^*(S\setminus\{i\},j)+d_{j,i}\mid j\in S\setminus\{i\}\}.$$

アルゴリズムの設計

再帰式に基づき，部分問題の最適値をすべて計算する．この後，$\min\{f^*(V\setminus\{1\},i)+d_{1,i}\mid i=2,3,\ldots,n\}$ を計算するとツアーの最小費用が得られる．バックトラックにより最適ツアーを求める．

計算量の上界の解析

部分問題の個数は $O(n2^n)$ であり，部分問題 1 題あたり $O(n)$ 時間で解ける．よって，計算時間は $O(n^2 2^n)$，計算領域は $O(n2^n)$ である．

第１２回の講義は以上である．