# 最小増加超距離木問題に対する$k$制限部分木交換近傍に基づく局所探索アルゴリズム

01013123　静岡大学　　安藤和敏　ANDO Kazutoshi
05001378　静岡大学　*水越雅紀　MIZUKOSHI Masaki

## 1.　はじめに

系統樹とは生物の進化の歴史を表す木であり, 系統樹を推定することは分子系統学における重要な課題である．系統樹の推定に用いられる主な方法は距離法である．$X = \{1, \cdots, n\}$ を考察の対象である生物種の集合とする．距離法では, まず生物種間の距離を表す行列 $M : X \times X \to \mathbb{R}_+$ を求める．次に葉の集合が $X$ であるような枝重み付き木 $(T, l)$ の中で $\|D_{(T,l)} - M\|_p$ を最小化するものを求める．ここで, 各 $x, y \in X$ に対して, $D_{(T,l)}[x, y]$ は $(T, l)$ における $x, y$ 間の距離であり, $\|\cdot\|_p$ は $L_p$-ノルムを表す．ここで，$p < +\infty$ である．本研究では, 求める $(T, l)$ として**超距離木**を考える．超距離木とは根から葉までの距離がすべて等しいような枝重み付き根付き木である．一般性を失わず $T$ は 2 分木であると仮定する．

DNA 配列から求めた生物種間の距離 $M[x, y]$ は真の距離の下限であることが多い．したがって，我々は $M \le D_{(T,l)}$ という制約の下で $\|D_{(T,l)} - M\|_p$ を最小化する超距離木 $(T, l)$ を求める問題を考える．ここで, $M \le D_{(T,l)}$ は任意の $x, y \in X$ に対して $M[x, y] \le D_{(T,l)}[x, y]$ が成り立つことを意味する．この問題は **$L_p$-最小増加超距離木問題**と呼ばれ, NP 困難であることが知られている [1]．$L_p$-最小増加超距離木問題に対するアルゴリズムとして分枝限定法 [3] や, 近似保証を持つ解を求める近似アルゴリズム等がある．

石川他 [2] は, NNI (最小近傍交換) 操作 や SS (部分木交換) 操作と呼ばれる 2 分木の変形操作に基づく $L_p$-最小増加超距離木問題に対する局所探索アルゴリズムを導入し, NNI 操作及び SS 操作に基づくアルゴリズムの計算時間はそれぞれ $\mathrm{O}(tn^3)$ 及び $\mathrm{O}(tn^4)$ であることを示した．ここで, $t$ はアルゴリズム中の木の変形操作の数である．本研究では, SS 操作にある種の制限を設けた 2 分木の変形操作である $k$ 制限部分木交換操作（$k$SS 操作) を導入し，$L_p$-最小増加超距離木問題に対する $k$SS 操作に基づく局所探索アルゴリズムを考察する．ここで, $1 \le k \le n$ であり, 1SS 操作は NNI 操作であり, $n$SS 操作は SS 操作である．我々は, $p = 1$ のときのこの局所探索アルゴリズムの時間計算量は $\mathrm{O}(n^2 + t \min\{2^k, n\} k^2 n)$ であることを示し，さらに，このアルゴリズムの近似性能と実際的計算量を数値実験によって検証する．

## 2.　局所探索アルゴリズム

行列 $M : X \times X \to \mathbb{R}_+$ は, 任意の $x, y \in X$ に対して $M[x, y] = M[y, x]$ かつ $M[x, x] = 0$ を満たすとき, **$X$ 上の相違行列**と呼ばれる．根付き 2 分木 $T = (V, E)$ に対して $L(T)$ を $T$ の葉集合とする．以降では根付き 2 分木 $T$ は, $L(T) = X$ を満たすものとする．また，本研究では根付き 2 分木を全ての枝が根から葉に向かって向き付けられた有向木とみなす．

$T$ を根 $r$ を持つ根付き 2 分木とする．$T$ の点への重み付け $h : V \to \mathbb{R}_+$ が, 任意の $v \in L(T)$ に対して $h(v) = 0$ かつ, 任意の $v \in V - r$ に対して $h(v) \le h(p(v))$ を満たすとき, $(T, h)$ を**単調点重み付き根付き木**という．ここで，任意の $v \in V - r$ に対して，$p(v)$ は $v$ の親である．任意の単調点重み付き根付き木 $(T, h)$ に対して，$D_{(T,h)}[x, y] = 2h(\mathrm{lca}(x, y))$ と定義する．ここで, $\mathrm{lca}(x, y)$ は $T$ 中の $x$ と $y$ の最小共通祖先である．最小増加超距離木問題は, 以下の問題と等価である．

$$\begin{array}{|ll} \min & \|D_{(T,h)} - M\|_p \\ \text{s.t.} & (T, h) \text{ は単調点重み付き根付き木}, \\ & M \le D_{(T,h)}. \end{array} \qquad (1)$$

したがって, 本研究では単調点重み付き根付き木を**超距離木**と呼び, 問題 (1) の最適解 $(T, h)$ を求める．

任意の $X$ 上の相違行列 $M$ と根付き 2 分木 $T$ が与えられたとき, $(T, h)$ が $M$ に対する最小増加超距離木であるような点重み関数 $h \colon V \to \mathbb{R}_+$ を見出す問題は **MUTT 問題**と呼ばれる．

任意の $v \in V$ に対して $v$ を根とする $T$ の部分木を $T_v$ と表す．また，$T$ の任意の内部点 $v$ に対して $v$ の 2 つの子を $v_+, v_-$ と表す．

**定理 2.1 (Wu et al. [3])** アルゴリズム 1 は $X$ 上の相違行列 $M$ と根付き 2 分木 $T = (V, E)$ が与えられたとき, $\mathrm{O}(n^2)$ 時間で MUTT 問題の最適解 $h$ を出力する．

$M$ と $T$ に対する MUTT 問題の最適解を $\mathrm{MUTT}(M, T)$ によって表す．アルゴリズム 1 を

**アルゴリズム 1:** MUTT アルゴリズム.

**for** $v \in X$ **do** $h(v) \leftarrow 0$;
**for** $T$ の各内部点 $v$ を後行順で **do**
  $h(v) \leftarrow \max\{h(v_+), h(v_-), \max\{M[x,y]/2 \mid x \in L(T_{v_+}), y \in L(T_{v_-})\}\}$;

**アルゴリズム 2:** 局所探索アルゴリズム.

$T \leftarrow$ 任意の根付き 2 分木, $h \leftarrow \mathrm{MUTT}(M,T)$;
**do**
  $\mathcal{N}(T)$ 中で $\|D_{(T',\mathrm{MUTT}(M,T'))} - M\|_p$ を最小にする根付き 2 分木を $T'$ とする;
  **if** $\|D_{(T,h)} - M\|_p > \|D_{(T',\mathrm{MUTT}(M,T'))} - M\|_p$ **then**
    $(T,h) \leftarrow (T', \mathrm{MUTT}(M,T'))$;
**while** $T$ は局所最適でない;

サブルーチンとして最小増加超距離木問題に対する局所探索アルゴリズム (アルゴリズム 2) を得る [2].

我々は, アルゴリズム 2 中の $\mathcal{N}(T)$ として以下で定義する $T$ の $k$ 制限部分木交換近傍を考える. $e = (v_1, w_1)$ と $f = (v_2, w_2)$ を $v_2 \neq v_1$ かつ $w_1$ と $w_2$ が子孫関係になく, $v_1$ から $v_2$ の道の長さが $k$ 以下であるような 2 つの枝とする. $T$ から枝 $e, f$ を削除した後 $(v_1, w_2), (v_2, w_1)$ を挿入する操作は, $e$ と $f$ による $k$ **制限部分木交換操作** ($k$SS 操作) と呼ばれる. $T$ から 1 回の $k$SS 操作によって得られる 2 分木の集合を $T$ の $k$ **制限部分木交換近傍** ($k$SS 近傍) と呼ぶ. 定理 2.1 より, アルゴリズム 2 における各反復の計算時間は $\mathrm{O}(\min\{2^k, n\}n^3)$ である.

## 3. $p = 1$ の場合のアルゴリズムの高速化

$M$ を $X$ 上の任意の相違行列, $T = (V, E)$ を根付き 2 分木, $h = \mathrm{MUTT}(M,T)$ とする. また, $T'$ を $T$ に対する $e = (v_1, w_1)$ と $f = (v_2, w_2)$ による $k$SS 操作で得られる 2 分木, $h' = \mathrm{MUTT}(M, T')$ とする.

**補題 3.1**

$$h'(v) = h(v) \qquad (v \in V \backslash V(P(v_1, v_2)) + \mathrm{lca}(v_1, v_2)).$$

ここで, $P(v_1, v_2)$ は $v_1$ から $v_2$ への $T$ 中の道である.

**補題 3.2**

$$\begin{aligned} &\frac{1}{2}(\|D_{(T',h')} - M\|_1 - \|D_{(T,h)} - M\|_1) \\ &= \sum_{v \in V(P(v_1,v_2))} (|L(T'_{v'_+})| \cdot |L(T'_{v'_-})| \cdot h'(v) \\ &\quad - |L(T_{v_+})| \cdot |L(T_{v_-})| \cdot h(v)). \end{aligned}$$

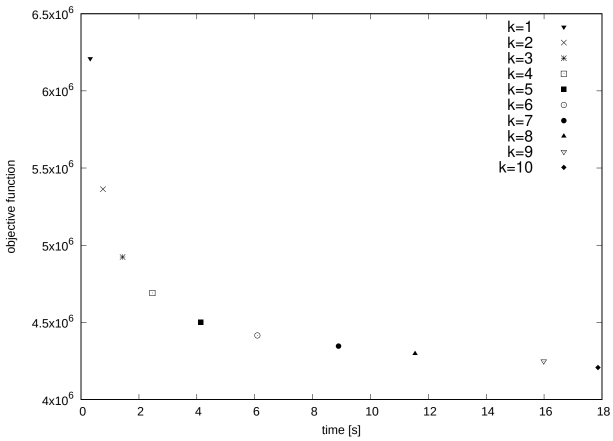


図 1: 実行時間と目的関数値のプロット.

ここで, $v'_+, v'_-$ は $v$ の $T'$ における 2 つの子である.

**定理 3.3** $X$ 上の相違行列 $M$ を入力とし, $p = 1$, $\mathcal{N}$ を $k$SS 近傍とするとき, アルゴリズム 2 の計算時間は $\mathrm{O}(n^2 + t\min\{2^k, n\}k^2 n)$ である.

## 4. 数値実験

$L_1$-最小増加超距離木問題に対する $k$ 制限部分木交換操作に基づく局所探索アルゴリズムの性能を数値実験によって検証した. アルゴリズムの入力として与える相違行列（$n = 500$）はランダムに生成した超距離木の距離行列をランダムに摂動させたものである. $k = 1, \ldots, 10$ のそれぞれに対して 100 個の相違行列を入力として, アルゴリズムの実行時間と出力の目的関数値の平均を図 1 に示す. $k = 3, 4$ 程度までは $k$ の増加したがって計算時間をあまり大きくすることなく目的関数値は大きく減少するが, $k$ が 5 を超えるあたりから計算時間の増加に対する目的関数値の減少の度合いは著しく小さくなる.


## 謝辞

本研究は JSPS 科研費 18K11180 の助成を受けたものである.


## 参考文献


[1] M. Farach et al.: A robust model for finding optimal evolutionary trees. *Algorithmica* **13** (1995) 155–179.

[2] 石川累, 安藤和敏: 最小増加超距離木問題に対する局所探索アルゴリズム. 数理解析研究所講究録 **2027** (2017) 15-29.

[3] B. Y. Wu et al.: Approximation and exact algorithms for constructing minimum ultrametric trees from distance matrices. *Journal of Combinatorial Optimization* **3** (1999) 199–211.