局所最適解のクラスタリングによる多様な解の生成法と解釈性

05001029　国立情報学研究所　*橋本進　HASHIMOTO Susumu

1.　はじめに

現実の問題をそのまま最適化問題にモデル化することは一般に難しく，なんらかの制約や選好などが緩和されることが多い．例えば，自宅から会社まで車で通勤する最も良い経路を求める問題を考える．地図データを用いて最短距離の経路を求める問題（最短路問題）にモデル化することは難しくないが,「コンビニが経路上にあってほしい」「この道路は右折待ちで渋滞しやすい」などの細かい情報が無視され得るため，モデル上の最適解が現実の問題において受容できる解であるとは限らない．実務ではこのギャップを想定して，1 つの最適解ではなく多数の近似解を出力することがしばしば要求される．

また，現実の問題にとって良い解が得られたとして,「なぜ」その解が良いかを知ることは，実務者にとって非常に重要である．上記の問題を考えると，例えば「得られた良い経路の中で，自宅からA地点までの経路が特に重要である」ことが分かれば，A地点より先の経路上で通行止めが起こった場合でも，A地点以降で経路を変更すればある程度早く会社に着くだろうという推測を立てられる．しかし，これまでの応用研究の多くは短時間でモデル上の良い解を得ることを目的としており，解の解釈性に関する研究はほとんど無い．

本研究は，上記 2 つの問題を解決する，局所最適解のクラスタリングを用いた組合せ最適化問題に対するアルゴリズムのフレームワークを提案し，最短路問題と巡回セールスマン問題に対する数値実験で有効性を示す．数値実験の結果と巡回セールスマン問題に関する説明は紙面の都合上省略する．

2.　最短路問題と局所探索法

連結な枝重み付きグラフ $G = (V, E, w)$ と始点 $s \in V$，終点 $t \in V$ が与えられた時に，重みが最小となる $s-t$ パスを求める問題を最短路問題という．以下では，枝 $(v_1, v_2) \in E$ の重みを $w(v_1, v_2)$ と表記し，任意のパス $p = v_{i_1} v_{i_2} \dots v_{i_k}$ に対して $w(p) = \sum_{j=1}^{k-1} w(v_{i_j}, v_{i_{j+1}})$ と定義する．本研究では，任意の頂点 $v_1, v_2 \in V$, $v_1 \neq v_2$ に対して $w(v_1, v_1) = 0$, $w(v_1, v_2) > 0$ を仮定する．この問題はダイクストラ法 [1] などの解法で高速に解けることが知られているが，著者の知る限りでは局所探索法は提案されていない．そのため，本研究では最短路問題に対する局所探索法を新たに提案する．この局所探索法では，まず G 上での全点対間最短距離を求め，それを用いて枝重み付き完全グラフ $G' = (V, V \times V, w')$ を作る．そして G' 上の $s-t$ パス p の近傍 $N(p)$ を次のように定義する．

完全グラフ上の最短路問題に対する解の近傍

G' 上の最短路問題の任意の解 $p = s v_{i_1} \dots v_{i_k} t$ に対して，次の (i) から (iii) のいずれか 1 回の操作で得ることができるパスの集合を p の近傍 $N(p)$ と定義する．ただし，$V_p = \{v_{i_1}, \dots, v_{i_k}\}$, $V_p^c = V \setminus (V_p \cup \{s, t\})$ とする．
(i) $v_1 \in V_p$ を選び，p から削除する
(ii) $v_1 \in V_p$ と $v_2 \in V_p^c$ を選び，p 上の v_1 を v_2 に変更する
(iii) $v_2 \in V_p^c$ を選び，p の先頭・末尾以外の場所に挿入する

この近傍を用いて，次の局所探索法を提案する．

Algorithm 1 最短路問題に対する局所探索法

Require: $G = (V, E, w)$, $s, t \in V$
Ensure: $s-t$ パス p
全点対間最短距離 $d : V \times V \to \mathbb{R}_{\geq 0}$ と各点から終点 t への最短距離 $d_t : V \to \mathbb{R}_{\geq 0}$ を求める
Algorithm 2 で $s-t$ パス p_{init} を生成する
$p \leftarrow p_{init}$
Algorithm 3 で w' を計算する
while $\max_{p' \in N(p)} \{w'(p) - w'(p')\} > 0$ **do**
　$w'(p) - w'(p') > 0$ を満たす $p' \in N(p)$ をランダムに選ぶ
　$p \leftarrow p'$
end while
return p

Algorithm 2 最短路問題に対する初期解生成法

Require: $G=(V,E,w)$, $s,t\in V$, $d_t:V\to\mathbb{R}_{\geq 0}$
Ensure: $s-t$ パス p

$p \leftarrow s$
$u \leftarrow s$
while $u \neq t$ **do**
  $d_t(u) > d_t(u')$ を満たす u の G 上の隣接点 u' をランダムに選ぶ
  $p \leftarrow pu'$ (u' を p の末尾に追加する)
  $u \leftarrow u'$
end while
return p

Algorithm 3 枝重み付き完全グラフの構成法

Require: $G=(V,E,w), d:V\times V\to\mathbb{R}_{\geq 0}, \lambda>1$
Ensure: $w':V\times V\to\mathbb{R}_{\geq 0}$

for $v_1,v_2\in V$, $v_1\neq v_2$ **do**
  if $(v_1,v_2)\in E$ **then**
    $w'(v_1,v_2)\leftarrow w(v_1,v_2)$
  else
    $w'(v_1,v_2)\leftarrow \lambda d(v_1,v_2)$
  end if
end for
return w'

Algorithm 1 は G' 上の局所探索法になっているが，必ず G 上の $s-t$ パスを出力することが示せる．

3. 提案するフレームワークと計算例

本節は，組合せ最適化問題に対するアルゴリズムのフレームワークを以下に提案する．

局所最適解のクラスタリング

Step1. ランダム多点局所探索法により k 個の局所最適解 $X=\{x_1,\dots,x_k\}$ を求める．
Step2. 解の距離（乖離度合い）$d_X:X\times X\to\mathbb{R}_{\geq 0}$ を全対に対して計算する．
Step3. d_X を用いて解をクラスタリングする．

本手法は，局所探索法と解の距離を定義できる様々な組合せ最適化問題に応用できる．本研究の数値実験では，解の距離に Jaccard 距離，クラスタリング手法にデータ研磨によるマイクロクラスタリング [2] を用いている．このクラスタリング手法は無向グラフの密な構造を列挙する安定な手法であり，本研究では解の距離 d_X に閾値を設けて生成される解の隣接グラフに対して適用している．閾値は，近さ 0 の枝とそれを除いた上位 δ の割合の枝を隣接グラフに張るように設定している．

最後に，最短路問題の計算例を用いて本手法の結果が解釈でき得ることを説明する．

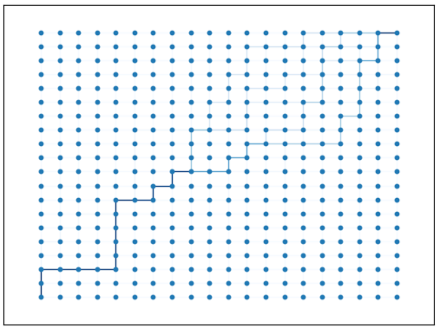

図 1: クラスタの計算例 ($n=555$)

$[10,20]$ の一様ランダムな枝重みを与えた 20×20 のグリッドグラフに対して，始点と終点をそれぞれ最も左下の点，右上の点，$k=1000, \lambda=1.1, \delta=0.3$ として提案手法を適用し，得られたクラスタのうち最も大きいものを図 1 に示す．各枝の濃さはクラスタ内でその枝を含む解の割合を示しており，濃いほど割合が高い．図 1 より，左下は経路が固定されているが，それ以降は様々な経路が得られており，このクラスタの重要な部分は左下の経路であると解釈できる．この解釈性により，現実の要求に合致しているクラスタを選び，重要な部分を固定して焼きなまし法などの精度の高い手法を適用するなど，様々な応用が考えられる．

参考文献

[1] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, Vol.1, 1959, pp.269–271.

[2] T. Uno, H. Maegawa, T. Nakahara, Y. Hamuro, R. Yoshinaka and M. Tatsuta, "Micro-clustering by data polishing," *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 1012–1018