A-004

巨大ナレッジグラフ向けの 1 メディアン問題に対する近似解法

Approximate Algorithms for the 1-Median Problem on Large-Scale Knowledge Graphs

植田　佳佑 †　　呉　偉 ‡
Keisuke Ueta†　　Wei WU‡

1　はじめに

p メディアン問題（p-median problem, pMP）とは，n 頂点を持つグラフ上で，p 個の施設を適切に配置することで，各顧客点から最も近い施設点までの重み付き総距離を最小化する組合せ最適化問題である．pMP は，代表的な施設配置問題として知られており，$p \geq 2$ のとき $\mathcal{NP}$ 困難であることが Kariv と Hakimi [1] によって示されている．一方で，$p = 1$ の場合の 1 メディアン問題（1-median problem, 1MP）は多項式時間で解くことが可能であるため，既存研究は少ない．しかし，入力グラフがナレッジグラフのような頂点数が膨大な場合は，多項式時間アルゴリズムであっても計算時間が非常に長くなってしまう．例えば，応用例として，ナレッジグラフ上に複数のユーザ履歴に基づく興味のある頂点（顧客点）が与えられたとき，ナレッジグラフ上で次にユーザに推薦すべき頂点（施設点）を探索する問題は，1MP の応用例とみなすことができる．このようなリアルタイム性が求められる推薦システムでは，短時間での応答が求められるが，既存の手法では計算コストが課題となる．

本研究では，大規模グラフかつ顧客点が比較的近接している 1MP に対して，探索範囲を適切に限定することにより，計算時間を短縮するアルゴリズムの設計を目標にする．提案手法の近似率を理論的に解析するとともに，既存の単純な厳密解法との比較実験を通して有効性を検証する．

2　問題説明

頂点集合 $V = \{1, 2, \ldots, n\}$ と辺集合 E からなる連結無向グラフ $G = (V, E)$ が与えられる．各辺 $\{i, j\} \in E$ には非負のコスト（距離）c_{ij}（$= c_{ji}$）が設定されており，m 個の頂点からなる顧客点集合 $M \subseteq V$ と，それぞれの顧客点 $i \in M$ には頂点重み w_i が与えられている．本研究で扱う 1MP は，各顧客点からの重み付き最短距離の総和が最小となる頂点（施設点）を求める問題である．本研究では，顧客点の数 m は施設候補点の数 n よりもはるかに少なく（$m \ll n$），m 個の顧客点はグラフ上で比較的に近接している状況を想定する．

記述の便宜上，頂点 i から j までのグラフ上の最短距離を $c_{ij}^{(\mathrm{sp})}$ と定義する．頂点 i $(\in V)$ が施設点として選ばれたときの評価値を $z(i) = \sum_{j \in M} w_j c_{ji}^{(\mathrm{sp})}$ とし，1MP は以下のように表せる：

$$\min_{i \in V} z(i) = \min_{i \in V} \sum_{j \in M} w_j c_{ji}^{(\mathrm{sp})}. \tag{1}$$

各 $j \in M$ を始点とするダイクストラ法を実行することで，問題 (1) に必要な $c_{ji}^{(\mathrm{sp})}$ をすべて計算できるため，1MP は $\mathcal{O}(m|E| + mn \log n)$ 時間で厳密に解くことができる．

3　提案手法

提案する 3 種類のアルゴリズムの基本的なアイデアを述べる．各顧客点 j を始点としたダイクストラ法を行う．ただし，始点 j と異なるすべての顧客点への最短距離が求まった時点で始点 j のダイクストラ法を打ち切る．また，このような共通の計算を行った後，それぞれの提案手法の違いを以下に説明する．

- TDA-SA（truncated Dijkstra algorithm with selective aggregation）：すべての顧客点から最短距離が計算された頂点のみを候補点として重みつき総距離を計算し，評価値が最小となる頂点を出力する．
- TDA-NNA（truncated Dijkstra algorithm with nearest-neighbor approximation）：少なくとも 1 つの顧客点からの最短距離が決定されている頂点を，評価する候補点とする．候補点 i の重みつき総距離を計算するときに，ある顧客点 j からの最短距離が確定されていなかった場合は，i と最も近い顧客点 j' を経由したパスの距離 $c_{jj'}^{(\mathrm{sp})} + c_{j'i}^{(\mathrm{sp})}$ を用いて $c_{ji}^{(\mathrm{sp})}$ を近似的に評価する．
- TDA-SPA（truncated Dijkstra algorithm with shortest-path approximation）：評価対象の候補点は TDA-NNA と同じである．しかし，候補点 i の重みつき総距離の計算時に，ある顧客点 j からの最短距離が確定されていなかった場合は，他の確定

†静岡大学 Shizuoka University
‡静岡大学 Shizuoka University

済みの顧客点を経由するパスの重みつき距離の中で，最も短いもの（$\min_{j':c_{j'i}^{(\mathrm{sp})}\text{が計算済み}} c_{jj'}^{(\mathrm{sp})}+c_{j'i}^{(\mathrm{sp})}$）を用いて $c_{ji}^{(\mathrm{sp})}$ を近似的に評価する．

4 近似精度

提案アルゴリズムに対して，以下の理論成果を得られた．

定理 1. $m \leq 3$ のとき，3 つの提案手法によって得られる解は最適解である．

定理 2. TDA-SA の近似率は 2 である．

補題 1. 定理 2 で示された TDA-SA の近似率がタイトである．

定理 3. TDA-NNA と TDA-SPA の近似率は 1.618 であり，近似率の下界は 1.2 である．

全ての $w_i = 1$ となる単位重みの特殊ケースに関して，以下の成果を得られた．

定理 4. 単位重みの 1MP に対して，$m \geq 3$ の場合 TDA-SA の近似率は $\left(2-\frac{4}{m+1}\right)$ である．

アルゴリズムの計算方法により，定理 4 で示された近似率は TDA-NNA, TDA-SPA に対しても有効である．

補題 2. 単位重みの 1MP に対して，$m \geq 3$ の場合，定理 4 で示された TDA-SA の近似率がタイトである．

5 計算実験

5.1 計算環境と問題例の生成

2 節の最後に述べた，各顧客点からダイクストラ法行うことで構築される厳密解法及び，3 つの提案手法の実装に C++言語を用いた．計算実験には，Xeon E-2286G CPU (4.0 GHz) と 64 GB のメモリを搭載した PC を用いた．

計算実験では以下の 2 種類のグラフを用いる：

- **RDU**（random graph with distance-restricted source selection and uniform vertex weights）: $|E| = 4n$ のランダムグラフである．顧客点はランダムに 1 点選択した後，ダイクストラ法で近傍の $\max\{2m, \lfloor \log_2 n \rfloor\}$ 個の中から m 点をランダム選択し，重みを 1 として設定する．
- **GDU**（grid graph with distance-restricted source selection and uniform vertex weights）: グリッドグラフであり，顧客点の選択および重みの設定方法は RDU と同様である．

用いる問題例のサイズは，$n \in \{10^6, 10^7\}$，$m \in \{2, 8, 32\}$，グラフタイプ $t \in \{\mathrm{RDU}, \mathrm{GDU}\}$ であり，各 (n, m, t) の組合せごとに 10 個，計 120 個の問題例を生成した．

5.2 計算結果

解の質に関しては，すべての提案手法がすべての問題例に対して最適解を得ることができた．厳密手法および 3 種類の提案手法における 120 問の平均計算時間を表 1, 2 に示す．

表 1: RDU 問題例に対する平均計算時間 (ミリ秒)

n	m	厳密解法	TDA-SA	TDA-NNA	TDA-SPA
10^6	2	1361	20	11	22
	8	5462	50	25	55
	32	21855	435	263	586
10^7	2	19542	8	4	9
	8	78120	1222	670	1320
	32	312908	1889	1036	2346

表 2: GDU 問題例に対する平均計算時間 (ミリ秒)

n	m	厳密解法	TDA-SA	TDA-NNA	TDA-SPA
10^6	2	342	0	0	0
	8	1395	0	0	0
	32	5675	0	0	1
10^7	2	4353	0	0	0
	8	18156	0	0	0
	32	70731	0	0	1

TDA-SA, TDA-NNA, TDA-SPA は，非常に優れた計算効率性を示した．特に，10^7 頂点を持つグリットグラフに対しては，1 ミリ秒以下で最適解を得ることができた．

6 おわりに

本研究では，1MP に対する 2 近似，および 1.618 近似アルゴリズムを提案した．また，計算実験により顧客点が少なく互いに近接している場合において，巨大グラフ上での有効性を示した．今後の展望として，TDA-NNA, TDA-SPA に対して，よりタイトな近似率を求めたい．また，提案手法の近似率の知見を活かし，新たな厳密解法を提案したい．

参考文献

[1] Kariv, O. and Hakimi, S. L.: An algorithmic pproach to network location problems. II: The p-medians, *SIAM Journal on Applied Mathematics*, Vol. 37 (1979), 539–560.