

非線形施設配置問題における貪欲法の遅延評価による高速化——複数施設が確率的に貢献する場合

武藤 伸明^{†1} 斉藤 和巳^{†1} 池田 哲夫^{†1}
永田 大^{†1} 伏見 卓恭^{†1}

施設配置問題は、施設を開設するコストと利用者が施設を利用するコストの和を最小化する問題である。本論文では、利用コストに複数施設が確率的に貢献するタイプの施設配置問題を考える。その解法として、コストの総和を目的関数とした貪欲法、あるいは解品質の向上のため単位開設コストあたりのコストの総和の差分を評価関数とする貪欲法が考えられる。本論文では、これらの目的関数および評価関数の数理構造を分析し、これらの貪欲法に遅延評価と呼ばれる手法が導入可能なことを示す。また計算機実験によって、遅延評価の導入による高速化の効果を明らかにする。

Speed-up of Greedy Method by Lazy Evaluation on Nonlinear Facility Location Problem — Case of Probabilistic Contribution by Multiple Facilities

NOBUAKI MUTOH,^{†1} KAZUMI SAITO,^{†1} TETSUO IKEDA,^{†1}
DAI NAGATA^{†1} and TAKAYASU FUSHIMI^{†1}

We address the problem of locating a number of facilities requiring opening costs, so as to maximize the expected benefits of users in the setting that multiple facilities can contribute to users with some probabilities. In this paper, after showing that this location problem has a submodular property, we present two solution methods based on a greedy algorithm, and propose their speed-up method by incorporating a technique called lazy evaluation on the submodular problem. In our experiments using three real data sets, we show that the proposed method can stably produce desirable results.

1. はじめに

工場や倉庫、公共施設、無線基地局などの施設の適切な配置は、多様な経済活動などを支える重要な問題である。特に、容量制約なし施設配置問題の研究には長い歴史があり、最近でも、精度保証付きの近似アルゴリズムが幾種も提案されるなど高度に発展が続いている³⁾。そこでの主要な解法は、整数計画問題として定式化される施設配置問題を線形計画問題に緩和し、双対性などの線形計画問題に特有な構造を土台にして開発されている。ゆえに、これら最新解法の適用には、線形計画問題への妥当な緩和が必要となるものの、現実には、このような緩和が実質的に困難な非線形性を有する施設配置問題も多数存在する。

本論文では、線形計画問題への妥当な緩和が困難な非線形施設配置問題として、複数施設が各利用者へ確率的に独立貢献する状況において開設コストを考慮し、利用者の期待利得の総和を最大化する問題に焦点を当てる。このような枠組みとなる問題の具体例としては、新商品などのクチコミ宣伝による影響最大化において情報源とすべき人々たちを選定する問題¹⁾ や、汚染などを検出するセンサを施設と見なして複数設置する問題⁵⁾ などである。特に、情報源の人たちと一般ユーザ（利用者）との人間関係の、あるいは汚染源とセンサの遠近により、情報伝達や、汚染検出の成功確率は変化すると自然に想定できる。よって、それら期待値の最大化問題としての定式化を考える。また、情報源として機能させたりセンサを設置したりするにはコストが必要であり、これを一般に施設の開設コストと考える。

この非線形施設配置問題を離散最適化の観点でとらえれば NP-完全クラスに属し、大規模になれば妥当な計算時間で厳密解を求めるのが一般に困難となる。ただし、本問題はサブモジュラ⁶⁾ と呼ばれる構造を持つと想定できる。サブモジュラ関数最大化問題は、有限ピボット集合に基づくクラスタリング問題¹⁰⁾ もその例であり、変数選択問題⁴⁾ など潜在的に幅広い応用が存在する。また、その望ましい性質として、いわゆる貪欲法で効率良く求まる近似解により、ある程度妥当な精度で、最悪ケースの解品質を理論的に保証することができる⁷⁾。さらに、遅延評価と呼ばれる手法の導入で貪欲法をさらに効率化することができる⁵⁾、非常に大規模かつ複雑な問題でも妥当な計算時間で精度保証付き近似解を得ることが可能になる。本論文の構成は以下のとおりである。まず、複数施設が各利用者へ確率的に独立貢献する状況で開設コストを考慮した非線形施設配置問題を提案し定式化するとともに、

^{†1} 静岡県立大学
University of Shizuoka

標準的な容量制約なし施設配置問題³⁾ や既存研究^{1),5)} における定式化との関係について述べる。次に、この非線形施設配置問題がサブモジュラ性を持つことを示す。次いで、施設追加による目的関数の増減分と、単位開設コストあたりの増減分を評価関数とする、2種の貪欲法を解法として示すとともに、その計算量について述べる。また、サブモジュラ性に基づき、2種の貪欲法に遅延評価が導入可能であることを示すとともに、それぞれの高速化アルゴリズムを提案し、その計算量について考察する。次に、貪欲解の近似精度について議論する。最後に、3種の実ネットワーク構造上での非線形施設配置問題において、施設開設コストのレンジを変化させ、各種手法の解品質と計算効率を実証評価するとともに、これらの手法の性能に関する性質を考察する。

2. 問題設定

容量制約なし施設配置問題の標準的な表記法や定式化³⁾ に基づき、複数施設が各利用者へ確率的に独立な貢献をする状況において開設コストを考慮し、利用者の期待利得の総和を最大化する非線形施設配置問題を提案し定式化する。いま、施設候補集合と利用者集合を \mathcal{F} と \mathcal{D} とし、施設と利用者を $i \in \mathcal{F}$ と $j \in \mathcal{D}$ で表記する。また、施設 i が利用者 j に貢献できる確率を $p_{i,j}$ とし、貢献に成功したときの利用者 j の利得を d_j とする。つまり、開設施設集合 $X \subset \mathcal{F}$ に対して、利用者の期待利得の総和 $C_S(X)$ を次式で定義する。

$$C_S(X) = \sum_{j \in \mathcal{D}} d_j \left(1 - \prod_{i \in X} (1 - p_{i,j})\right). \quad (1)$$

一方、施設 i の開設コストを f_i とし、開設施設集合 X に対する総コスト $C_F(X)$ を次式で定義する。

$$C_F(X) = \sum_{i \in X} f_i. \quad (2)$$

よって、本論文で提案する非線形施設配置問題とは、次式で定義する目的関数 $C(X)$ を最大化する開設施設集合 $X \subset \mathcal{F}$ を求める問題として定式化される。

$$C(X) = C_S(X) - C_F(X). \quad (3)$$

なお、標準的な施設配置問題³⁾ では、利用者の利得ではなくコストを導入にすることで最小化問題として定式化するが、本論文では、後述するサブモジュラ最大化問題との対応を考慮し、式 (3) で定義した目的関数の最大化を考える。

以下では、まず、標準的な容量制約なし施設配置問題³⁾ との関係について詳しく述べる。式 (3) で定義した目的関数は次式のように変形できる。

$$C(X) = \sum_{j \in \mathcal{D}} d_j - \sum_{j \in \mathcal{D}} d_j \prod_{i \in X} (1 - p_{i,j}) - \sum_{i \in X} f_i. \quad (4)$$

ここで、式 (4) 右辺の第 1 項は開設施設集合 X に依存しないため、符号を反転させれば、式 (3) の最大化問題は、次式で定義する目的関数 $B(X)$ の最小化問題と等価である。

$$B(X) = \sum_{j \in \mathcal{D}} d_j \prod_{i \in X} (1 - p_{i,j}) + \sum_{i \in X} f_i. \quad (5)$$

いま、各施設 i に対して、開設施設集合 X に含まれるかを示す指示変数 x_i を導入する。すなわち、 $i \in X$ ならば $x_i = 1$ 、さもなければ $x_i = 0$ で定義する。一方、指示変数 x_i を用いれば、式 (5) の目的関数 $B(X)$ は次のように書きかえられる。

$$\begin{aligned} B(X) &= \sum_{j \in \mathcal{D}} d_j \prod_{i \in \mathcal{F}} (1 - p_{i,j})^{x_i} + \sum_{i \in \mathcal{F}} x_i f_i \\ &= \sum_{j \in \mathcal{D}} d_j \exp\left(\sum_{i \in \mathcal{F}} x_i \log(1 - p_{i,j})\right) + \sum_{i \in \mathcal{F}} x_i f_i. \end{aligned} \quad (6)$$

ここで、対数関数 $\log(1 - p_{i,j})$ の値は定数であり、指数関数 $\exp(\cdot)$ の引数は線形式となる。よって、たとえば、指数関数 $\exp(\cdot)$ を Taylor 展開して 1 次近似すれば線形近似式を得ることもできる。しかしながら、指示変数は $x_i \in \{0, 1\}$ より、ある程度その差分が大きい状況を扱う必要があり、このような近似では妥当な精度で解を得ることは一般に困難と考えられる。

次に、提案した非線形施設配置問題の、情報源選定問題とセンサ配置問題での解釈について述べる。情報源選定問題では、選定した施設集合 X のすべてから同一情報が独立に発信され、利用者 j が施設 i からの情報受信に成功する確率が貢献確率 $p_{i,j}$ と見なされ、情報受信に成功すれば利用者 j は利得 d_j があると考えられる。なお、Kempe ら¹⁾ はネットワーク上での情報拡散モデルに基づく情報源選定モデルを扱っているため、情報受信の成功確率は互いに独立ではなく、シミュレーションによる確率推定が必要となり、改良法²⁾ は提案されているものの、大規模なネットワークを対象とすればきわめて膨大な計算量が必要となる。また、Kempe ら¹⁾ の定式化では、開設コストは考慮されていない。センサ配置問題では、汚染などの発生源が利用者 j と見なされ、その発生確率が利用者の利得 d_j で、施設 i に対応するそれぞれのセンサにより、各発生源に対する検出確率が貢献確率 $p_{i,j}$ と考える。なお、Leskovec ら⁵⁾ も発生確率を考慮したセンサ配置問題を対象としているが、Leskovec ら⁵⁾ の定式化では検出確率は考慮されず、目的関数が線形式で規定される問題を扱っている。

本論文の問題設定の応用例としては、特定ユーザのブログなどに広告を配置し、できるだけ多くの一般ユーザに対して新商品の情報発信をする問題などが想定できる。明らかに、特定ユーザ (i) に広告配置を依頼するにはコスト (f_i) が必要となり、一般ユーザ (j) が特定ユーザのブログなどを閲覧するかは確率的 ($p_{i,j}$) となるとともに、新商品の情報発信の観点においては、一般ユーザがどれか1つの広告を目にすることが重要となる。また、一般ユーザから得られる利得 (d_j) もユーザに依存すると考えられる。一方、このような応用においては、非常に大規模な問題となっても妥当な処理時間の範囲で、ある程度の近似精度で解を求めることが重要となってくる。すなわち、本論文では、インターネットなど新たなメディア上での非線形施設配置問題が想定応用であり、厳密解を求めることよりも、ある程度の精度でスケラビリティを重視する方法論について探求する。

3. サブモジュラ性

集合を定義域とする実数値関数 g が与えられ、集合の任意の要素 i と包含関係 $T \subset U$ を満たす任意の集合ペアに対して、以下のサブモジュラ不等式が成り立つとき、関数 g はサブモジュラ関数と呼ばれる。

$$g(T \cup \{i\}) - g(T) \geq g(U \cup \{i\}) - g(U). \quad (7)$$

以下では、式 (3) で定義した非線形施設配置問題の目的関数 $C(X)$ がサブモジュラ関数となることを示す。

いま、包含関係を満たす2つの開設施設集合を $T \subset U \subset \mathcal{F}$ とし、任意の施設を $i \in \mathcal{F}$ とする。このとき、任意の $i \in \mathcal{F}, j \in \mathcal{D}$ に対して $0 \leq p_{i,j} \leq 1$ であるから、式 (1) で定義した利用者の期待利得の総和関数 $C_S(X)$ に対して、以下の不等式が成り立つ。

$$\begin{aligned} C_S(T \cup \{i\}) - C_S(T) &= \sum_{j \in \mathcal{D}} p_{i,j} d_j \prod_{i' \in T} (1 - p_{i',j}) \\ &\geq \sum_{j \in \mathcal{D}} p_{i,j} d_j \prod_{i' \in U} (1 - p_{i',j}) = C_S(U \cup \{i\}) - C_S(U). \end{aligned} \quad (8)$$

すなわち、関数 $C_S(X)$ を X の関数と見なせば、サブモジュラ不等式が成り立ち、関数 $C_S(X)$ はサブモジュラ関数となる。また、式 (3) で定義した目的関数 $C(X)$ に対しても、式 (8) の結果を用いて、以下の不等式の成立を確認できる。

$$\begin{aligned} C(T \cup \{i\}) - C(T) &= C_S(T \cup \{i\}) - C_S(T) - f_i \\ &\geq C_S(U \cup \{i\}) - C_S(U) - f_i = C(U \cup \{i\}) - C(U). \end{aligned} \quad (9)$$

よって、本論文の非線形施設配置問題の目的関数 $C(X)$ についても、サブモジュラ不等

式が成り立ち、サブモジュラ関数となる。

4. 貪欲法に基づく解法

貪欲法とは、すでに選定した開設施設集合を固定し、ある評価関数値を最大にする施設を求め、目的関数が増加するならば開設施設集合に追加することで、結果の開設施設集合を求める方法である。すでに選定した開設施設集合を X とし、新たに追加を試みる施設を i とするとき、本論文では、以下で述べる2種の評価関数を考える。また、 $C_S(\emptyset) = 0$ と定義し、 $C(\emptyset) = 0$ とする。第1の評価関数は、施設追加による目的関数の単純な増減分に基づき次式で定義する。

$$H_{SG}(i; X) = C(X \cup \{i\}) - C(X) = \sum_{j \in \mathcal{D}} p_{i,j} \mu(j; X) - f_i. \quad (10)$$

ここで $\mu(j; X)$ は、次式で定義され、新たに選択する施設 i に依存せず、利用者 j とすでに選定した開設施設集合 X で決まる値である。

$$\mu(j; X) = d_j \prod_{i' \in X} (1 - p_{i',j}). \quad (11)$$

第2の評価関数は、施設追加による単位開設コストあたりの目的関数の増減分に基づき次式で定義する。

$$H_{CG}(i; X) = \frac{C(X \cup \{i\}) - C(X)}{f_i} = \frac{1}{f_i} \sum_{j \in \mathcal{D}} p_{i,j} \mu(j; X) - 1. \quad (12)$$

以降では、式 (10) で定義した評価関数に基づく貪欲を SG (Simple Greedy) 法と、式 (10) に基づく貪欲を CG (Cost-based Greedy) 法と呼ぶ。なお、CG 法の評価関数は、次のように変形できる。

$$H_{CG}(i; X) = \frac{C_S(X \cup \{i\}) - C_S(X)}{f_i} - 1. \quad (13)$$

つまり、CG 法については、単位開設コストで利用者の期待利得の総和を最も改善する施設を選択する方法と見なすこともできる。

施設数を制御する変数を k とし、 $H(i; X)$ を $H_{SG}(i; X)$ または $H_{CG}(i; X)$ とするとき、SG 法と CG 法に共通する貪欲アルゴリズムの詳細は以下となる。

G1. $k \leftarrow 1, X_0 \leftarrow \emptyset$ とし、各利用者 $j \in \mathcal{D}$ に対し $\mu(j; \emptyset) \leftarrow d_j$ と初期化する；

4 非線形施設配置問題における貪欲法の遅延評価による高速化——複数施設が確率的に貢献する場合

- G2. $\hat{i}_k = \arg \max_{i \in \mathcal{F} \setminus X_{k-1}} \{H(i; X_{k-1})\}$ を求め、 $X_k \leftarrow X_{k-1} \cup \{\hat{i}_k\}$ とする；
 G3. $C(X_{k-1}) \geq C(X_k)$ ならば $K = k - 1$ とし、 $\hat{X}_K = \{\hat{i}_1, \dots, \hat{i}_K\}$ を出力し終了する；
 G4. 各利用者 $j \in \mathcal{D}$ に対し $\mu(j; X_k)$ を求め、 $k \leftarrow k + 1$ としステップ G2. へ戻る．

ここで、 $\mathcal{F} \setminus X_{k-1}$ は $\{i : i \in \mathcal{F} \wedge i \notin X_{k-1}\}$ で定義される集合差を表す．また、ステップ G2. における最良追加要素のタイプブレークを一意に設定すれば、貪欲法で求まる結果の施設集合 \hat{X}_K は常に同一となる．

利用者数を $M = |\mathcal{D}|$ 、施設候補数を $N = |\mathcal{F}|$ 、そして結果の開設施設集合数を $K = |\hat{X}_K|$ とする．また、すべてのデータはメモリ上に格納されているとし、乗算回数に着目して貪欲法の主たる計算量を分析する．ステップ G2. では、集合 $\mathcal{F} \setminus X_{k-1}$ の各施設について、各利用者の期待利得を求めて最良施設 \hat{i}_k を求めるが、式 (10) または (12) を用いれば、その主たる計算量は $(N - k) \times M$ となる．一方、ステップ G4. で、各利用者に対し $\mu(j; X_k)$ については、 $\mu(j; X_k) = (1 - p_{i_k, j})\mu(j; X_{k-1})$ で求まるので、その計算量は M となる．よって、ステップ G2. から G4. の反復を $(K + 1)$ 回行うので、一般に $1 \ll K \ll N$ より、貪欲法の主たる計算量は $K \times M \times N$ となる．

5. 遅延評価による高速化

式 (3) で定義した目的関数の最大化問題はサブモジュラ性を満たすことより、式 (10) または (12) で定義した評価関数において逆単調性が成り立つことを示す．ここで、集合を定義域とする実数値関数 h に対して、包含関係 $T \subset U$ を満たす任意の集合ペアに対して $h(T) \geq h(U)$ が成り立つとき、関数 h は逆単調性を持つという．まず、SG 法の評価関数 $H_{SG}(i; X)$ については、施設 i を固定し、施設集合 X に対する関数と見なせば、式 (9) の関係より、逆単調性の成立が確認できる．また、CG 法の評価関数 $H_{CG}(i; X)$ についても、施設 i は固定されるので、評価関数 $H_{SG}(i; X)$ の定数倍 ($1/f_i$ 倍) するだけなので、同様に逆単調性の成立が確認できる．

逆単調性に基づく遅延評価 (lazy evaluation) の導入により、貪欲法の探索効率を向上させる方法について述べる．いま、 $X_0 = \emptyset$ とし、貪欲法の各反復 k で求まる解 (開設施設集合) リストを X_1, X_2, \dots とすれば、逆単調性より、評価関数値 $H(i; X_k)$ は $0 \leq k$ で非増加である．したがって、 $s < k$ となる任意の非負整数 s に対し、関数値 $H(i; Z_s)$ は $H(i; Z_k)$ の上限値となる．この関係式を利用して、探索効率を向上させるのが遅延評価付き貪欲法である．本論文では、遅延評価付き SG 法と CG 法を、それぞれ SGLE (SG with

Lazy Evaluation) 法と CGLE (CG with Lazy Evaluation) 法と呼ぶ．施設 $i \in \mathcal{F}$ に対する評価上限値を $\zeta(i)$ と表記し、評価上限値で施設を降順にソートして構成したリストの上位 n 番目の施設を r_n とするとき、SGLE 法と CGLE 法に共通する遅延評価付き貪欲アルゴリズムの詳細は以下となる．

- GL1. $k \leftarrow 1$, $X_0 \leftarrow \emptyset$ とし、各利用者 $j \in \mathcal{D}$ に対し $\mu(j; \emptyset) \leftarrow d_j$ 、各施設 $i \in \mathcal{F}$ に対し $\zeta(i) \leftarrow \infty$ に初期化し、施設リストを任意に構成し各 r_n を定める；
 GL2. $G \leftarrow 0$ とし、 $n = 1, \dots, N$ の順で GL2-1 から GL2-2 の処理を繰り返す．
 GL2-1. $\zeta(r_n) \leq G$ なら、 $\zeta(\hat{i}_k) \leftarrow 0$, $X_k \leftarrow X_{k-1} \cup \{\hat{i}_k\}$ とし、ステップ GL3 に進む；
 GL2-2. $\zeta(r_n) \leftarrow H(r_n; X_{k-1})$ とし、 $G < \zeta(r_n)$ なら、 $\hat{i}_k \leftarrow r_n$, $G \leftarrow \zeta(r_n)$ とする；
 GL3. $C(X_{k-1}) \geq C(X_k)$ ならば $K = k - 1$ とし、 $\hat{X}_K = \{\hat{i}_1, \dots, \hat{i}_K\}$ を出力し終了する；
 GL4. 各利用者 $j \in \mathcal{D}$ に対し $\mu(j; X_k)$ を求め、評価上限値 $\zeta(i)$ で施設を降順にソートしたものを r_1, \dots, r_N とし、 $k \leftarrow k + 1$ としステップ GL2. へ戻る．

ここで G は、処理過程の時点で最良評価値を格納する変数である．遅延評価付き貪欲法では、ステップ GL2 の第 k 反復目の最良施設選定において、評価上限値 $\zeta(i)$ を用いて、各施設 i を降順にソートしたリストを利用する．このリストの先頭から順に探索する過程で、ある施設 $i \in \mathcal{F}$ での実際の改善値と比較して、探索リスト上で未探索の先頭要素での上限が小さくなれば、その時点で探索を終了させ最良施設 \hat{i}_k が求まる．明らかに、遅延評価付き貪欲法 SGLE 法と CGLE 法の出力結果は SG 法と CG 法の結果とそれぞれ一致する．

遅延評価付き貪欲法の主たる計算量について述べる．ステップ GL2-2. では、施設 r_n に対して評価関数値 $H(r_n; X_{k-1})$ を求めるのに、式 (10) または (12) を用いれば、その計算量は M となる．よって、ステップ GL3 の終了条件を満たすまでの反復で探索した施設数の平均割合を α ($0 < \alpha < 1$) とすれば、ステップ GL2 の総計算量は $\alpha \times M \times N$ となる．一方、ステップ GL4. では、貪欲法のステップ G4. と同じ計算量 M に加えて、評価上限値による $(N - k)$ 個の施設をソートするので $(N - k) \log(N - k)$ の計算量が必要となるが、このソーティング処理には既知の評価上限値リストを用いるので、期待値計算では必要な利用者数との積は不要となる．よって、ステップ G2. から G4. の反復を $K + 1$ 回行うので、一般に $1 \ll K \ll N$ かつ $\log N \ll M$ より、遅延評価付き貪欲法の主たる計算量は $\alpha \times K \times M \times N$ となる．すなわち、遅延評価付き貪欲法の計算効率は α に依存するが、この量を解析的に求めることは一般に困難なため、本論文では実験により評価する．

6. 貪欲解の近似精度

貪欲法で得られる施設集合 (貪欲解) \hat{X}_K の最適解 X^* に対する近似精度を論じるため、Nemhauser ら⁷⁾ の結果を土台に、本論文の問題設定に特化して導かれる関係式について述べる。まず、以下の導出で用いる基本不等式として、貪欲解 \hat{X}_K と任意の施設 i に対して、アルゴリズムの終了条件 (G3) より、 $C(\hat{X}_K \cup \{i\}) - C(\hat{X}_K) \leq 0$ が成立すること、および、任意の施設集合 X と施設 i のペアに対して、 $C(X \cup \{i\}) - C(X) \geq -f_i$ が成立することを注意しておく。いま、貪欲解 \hat{X}_K と最適解 X^* の集合差を $X^* \setminus \hat{X}_K = \{a_1, \dots, a_Q\}$ 、および、 $\hat{X}_K \setminus X^* = \{b_1, \dots, b_R\}$ とする。ここで、 $R \leq |\hat{X}_K| = K$ である。このとき、目的関数 $C(X)$ のサブモジュラ性より、以下に示す不等式が成り立つ。

$$\begin{aligned} C(\hat{X}_K \cup X^*) - C(\hat{X}_K) &= \sum_{t=1}^Q (C(\hat{X}_K \cup \{a_1, \dots, a_t\}) - C(\hat{X}_K \cup \{a_1, \dots, a_{t-1}\})) \\ &\leq \sum_{t=1}^Q (C(\hat{X}_K \cup \{a_t\}) - C(\hat{X}_K)) \leq 0. \end{aligned} \quad (14)$$

$$\begin{aligned} C(X^* \cup \hat{X}_K) - C(X^*) &= \sum_{t=1}^R (C(X^* \cup \{b_1, \dots, b_t\}) - C(X^* \cup \{b_1, \dots, b_{t-1}\})) \\ &\geq \sum_{t=1}^R (C((X^* \cup \hat{X}_K \setminus \{b_t\}) \cup \{b_t\}) - C(X^* \cup \hat{X}_K \setminus \{b_t\})) \\ &\geq -\sum_{t=1}^R f_{b_t} \geq -\sum_{i \in \hat{X}_K} f_i = -C_F(\hat{X}_K). \end{aligned} \quad (15)$$

よって、式 (14) から式 (15) を引くことにより、貪欲解による最適解の近似精度に関する上限として次式を得ることができる。

$$C(X^*) \leq C(\hat{X}_K) + C_F(\hat{X}_K) = C_S(\hat{X}_K). \quad (16)$$

ここで、上記関係式 (16) は、貪欲アルゴリズムの終了条件の性質のみ利用しているため、評価関数が H_{SG} か H_{CG} に依存することなく、SG 法でも CG 法でも成り立つ。したがって、最適解の性能上限として、SG 法と CG 法で求まる $C_S(\hat{X}_K)$ の最小値を採用できる。

7. 評価実験

本章では、実験で用いる評価データと実験設定について述べた後、実験結果について報告する。また、実験結果を詳しく分析し、各種手法の特性について分析する。

7.1 評価データと実験設定

ネットワークのノード群を利用者や施設と見なした非線形施設配置問題で各種手法の特性を評価する。実験では、3 種の実ネットワークを用いた。詳細には、Watts らがスモールワールドに関する論文で用いた電力ネットワーク¹¹⁾ (ノード数 4,941, リンク数 6,594)、ブログのエントリとトラックバックから構成される有向グラフに対し、トラックバックは相互承認が必要なことから無向化して構築したブログネットワーク⁹⁾ (ノード数 12,047, リンク数 39,960)、および、ウィキペディア内の人名一覧に登場する人物を対象に、記事中に 6 回以上共起した 2 人の人物をリンクすることから得られる無向グラフの最大連結成分を抽出した人名ネットワーク²⁾ (ノード数 9,481, リンク数 122,522) を用いた。これらのネットワークでは、リンクの意味付けとともに、ある程度の幅でリンク密度 (ノード数とリンク数の比) も異なるので、それぞれに固有の性質を持ったネットワークと考えられる。ただし、これらは複雑ネットワークに特有の性質⁸⁾ を持つ点では共通である。

実験では、利用者集合 \mathcal{D} と施設候補集合 \mathcal{F} をネットワークの全ノードとして設定した。すなわち、 $\mathcal{D} = \mathcal{F}$ である。施設 i による利用者 j への貢献確率については、ノード i とノード j 間のネットワーク上での最短パス長を $d(i, j)$ を求め、施設と利用者の任意のペアに対して $p_{i,j} = 1/(1 + d(i, j))$ で定義した。明らかに、 $d(i, i) = 0$ に設定すれば、任意のネットワークにおいて $0 \leq p_{i,j} \leq 1$ を満たす。今回の実験では、全利用者の利得はすべて等しく $d_j = 1$ に設定した。各施設の開設コストについては、その最大コスト f_{max} を正の実数として与え、 $1 \leq f_i \leq f_{max}$ を満たす実数を一様ランダムに割り当てて設定した。なお、 $f_{max} = 1$ のときは、全施設の開設コストは $f_i = 1$ となる。また実験では、 $f_{max} = (1.2)^q$ と設定し、 q は非負整数値として、 $f_{max} = (1.2)^q < 10000$ を満たす範囲のすべての最大開設コストを用いて実験した。また、以下で示す結果は、それぞれの最大開設コスト f_{max} に対して、すべての施設の開設コストを一様ランダムに割り当てなおすことを 100 回繰り返し平均を求めた値である。

7.2 実験結果

図 1 では、CG 法と SG 法の解品質を比較するため、最大開設コスト f_{max} の設定に対し、電力ネットワーク、ブログネットワーク、および、人名ネットワークにおける結果の目

6 非線形施設配置問題における貪欲法の遅延評価による高速化——複数施設が確率的に貢献する場合

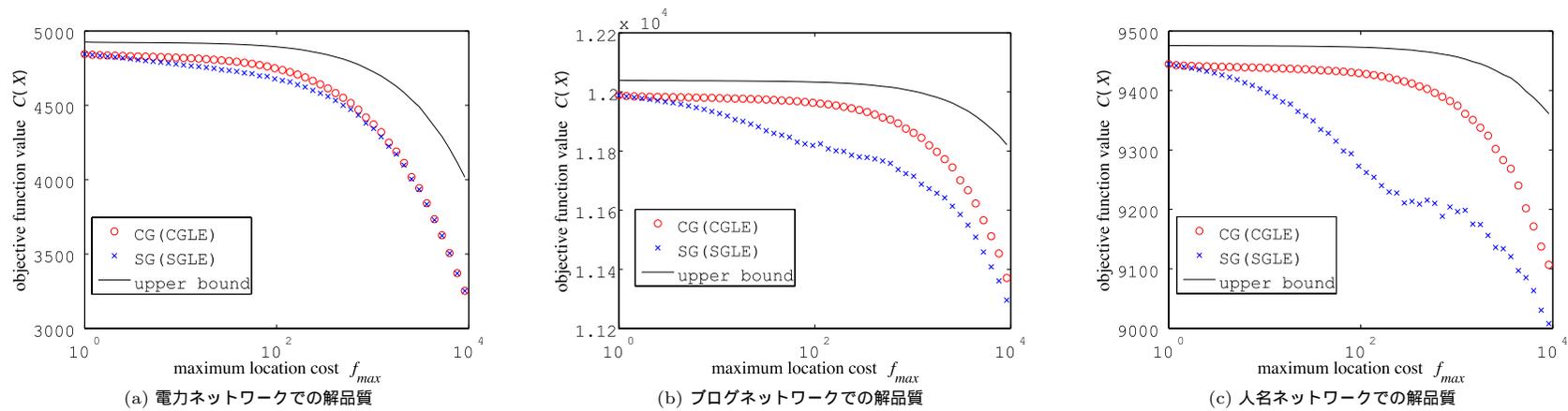


図 1 解品質の比較

Fig.1 Comparison of solution quality.

的関数の値を比較している．ここで，CGLE 法と SGLE 法の結果は，CG 法と SG 法とそれぞれ完全に一致するので省略している．図より，どのネットワークへの適用でも，CG 法の結果は SG 法より優れていることが分かる．詳細には， $f_{max} = 1$ のときは全施設の開設コストは等しく，CG 法と SG 法は等価になり，両者の結果は一致している．これに対し，ある程度の範囲まで最大開設コスト f_{max} が大きくなると，CG 法と SG 法の性能差が増大し，さらに f_{max} が大きくなると，両者の性能差は減少する傾向が見取れる．すなわち，ある程度の範囲で開設コストが異なる状況において，SG 法と比較して，CG 法の有効性は顕著になることが示唆される．

図 1 には，6 章で導出した，貪欲解から求まる最適解の性能上限 (upper bound) も示している．ここで，上限値として SG 法と CG 法で求まる $C_S(\hat{X}_K)$ の最小値を採用しているが，後述するように，本実験範囲の任意の最大開設コスト f_{max} に対し，CG 法でも SG 法でもほとんど等しい上限値が与えられた．図より，どのネットワークへの適用でも，特に CG 法で求まる貪欲解の性能は，この上限値に対して妥当な精度が得られていることが分かる．詳細には，上限値に対する CG 法での性能比は， $f_{max} < 100$ の範囲で 97% から 99% 程度以上となり，全体でも，電力ネットワークでは 80% 程度以上，プログと人名ネットワークでは 96% 程度以上となっている．

図 2 では，CGLE 法，CG 法，SGLE 法，および，SG 法の処理効率を比較するため，最

大開設コスト f_{max} の設定に対し，電力ネットワーク，プログネットワーク，および，人名ネットワークにおける結果を得るまでに要した処理時間 (sec.) を比較している．ここで，各手法のプログラムは C++ 言語で実装し，実験には Intel Xeon X5450 3.0 GHz CPU のサーバを用いた．図より，どのネットワークへの適用でも，処理効率の面で CGLE 法は CG 法よりも，SGLE 法は SG 法よりも顕著に優れていることが分かる．すなわち，逆単調性に基づく遅延評価導入の有効性が実証されたと考える．また図において，CG 法と SG 法，および，CGLE 法と SGLE 法のそれぞれのペアを比較すれば，解品質の比較結果と類似して， $f_{max} = 1$ のときには，それぞれのペアの結果は一致し，ある程度の範囲まで最大開設コスト f_{max} が大きくなると，それぞれのペアの性能差が増大し，さらに f_{max} が大きくなると，それぞれのペアの性能差は減少する傾向が見取れる．ただし，CG 法と SG 法では，ある程度 CG 法の方が計算時間を要するのに対し，CGLE 法と SGLE 法では，逆に，CGLE 法の処理効率の方が優る結果となっている．つまり，CG 法の評価関数の方が SG 法のよりも，特に，ある程度の範囲で開設コストが異なる状況において，遅延評価の導入効果が大きいことが示唆される．各手法がこのような特性を持つ理由については，以下で述べる分析を通して考察する．

7.3 実験結果の分析

図 3 では，実験結果を分析するため，最大開設コスト f_{max} の設定に対し，電力ネット

7 非線形施設配置問題における貪欲法の遅延評価による高速化——複数施設が確率的に貢献する場合

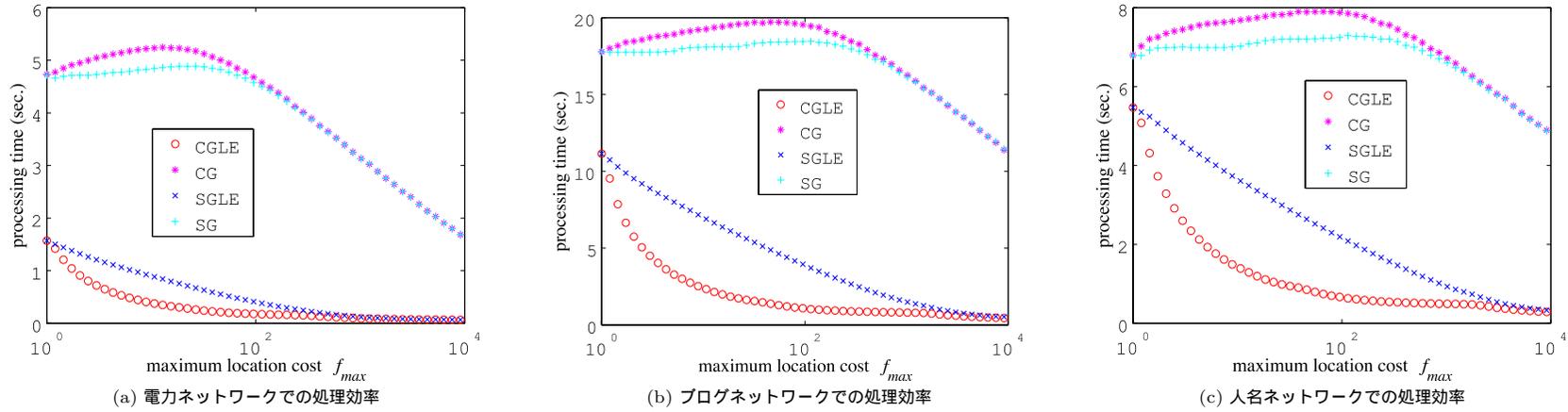


図2 処理効率の比較
Fig. 2 Comparison of processing time.

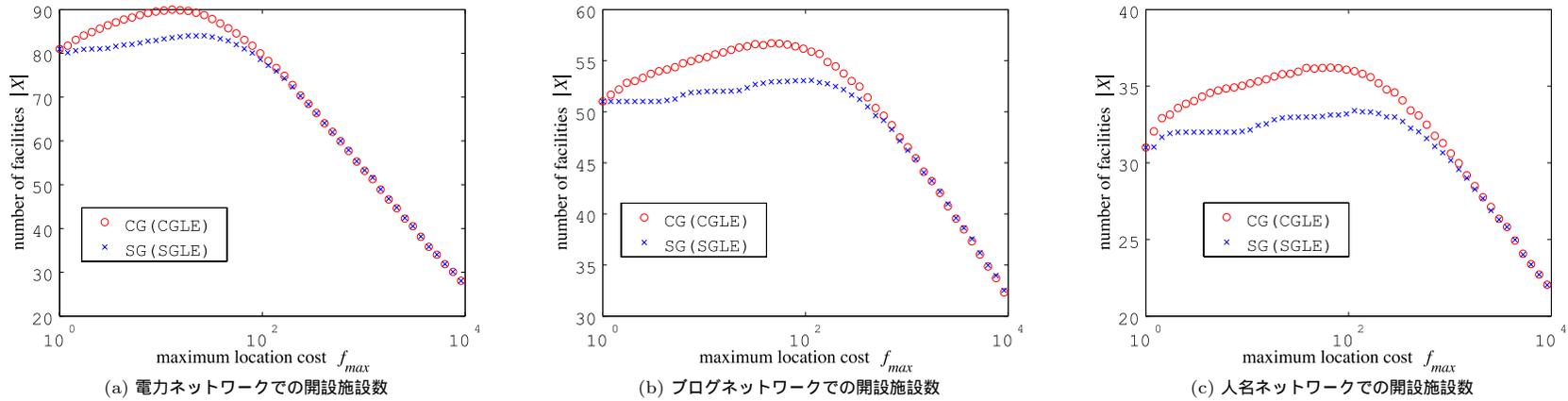


図3 開設施設数の比較
Fig. 3 Comparison of the number of facilities.

ワーク、プログネットワーク、および、人名ネットワークにおけるCG法とSG法の結果の開設施設数を比較している。ここでも、CGLE法とSGLE法の結果は、CG法とSG法とそれぞれ完全に一致するので省略している。図より、どのネットワークへの適用でも、 $f_{max} = 1$

のときには、CG法とSG法の結果は一致し、ある程度の範囲まで最大開設コスト f_{max} が大きくなると、CG法とSG法の両方で開設施設数が増える傾向があり、特に、CG法において顕著に増加する傾向がみられる。一方、 f_{max} がさらに大きくなれば、両者の結果はほ

8 非線形施設配置問題における貪欲法の遅延評価による高速化——複数施設が確率的に貢献する場合

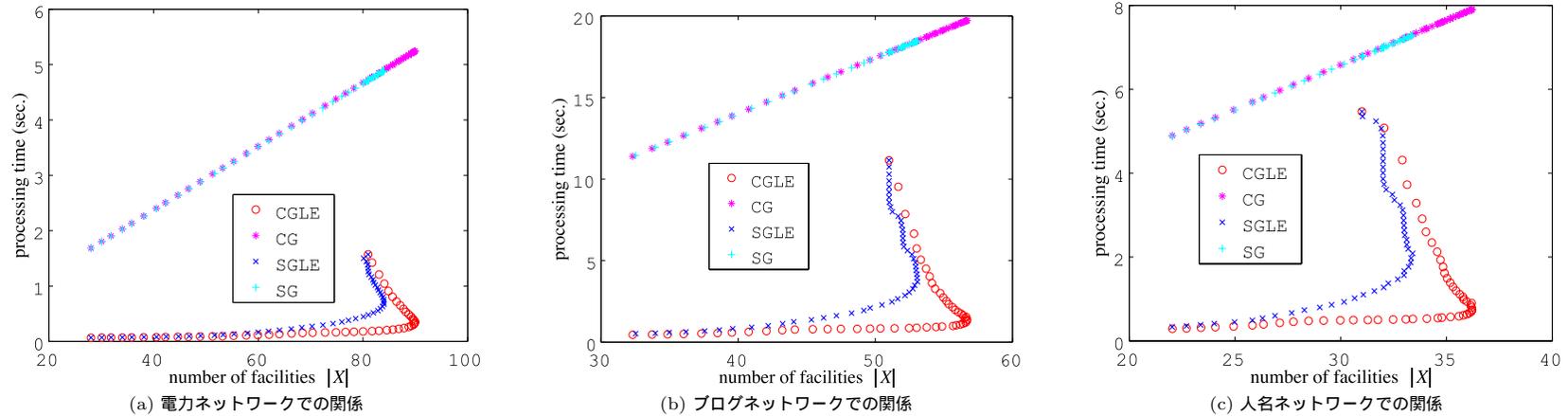


図4 開設施設数と処理間の関係

Fig. 4 Relationship between the number of facilities and processing time.

とんど等しくなる．ここで，図1で示した目的関数の値に関しては，最大開設コスト f_{max} が大きくなるにつれ，目的関数の $-C_F(X)$ の項がより大きく影響するので，ほぼ単調に減少している．これに対して，開設施設数に関しては，最大開設コストが増大により単調に減少するとは限らず，ある程度の範囲で開設コストが異なる状況では，むしろ多数の施設を開設することで，特にCG法において目的関数値の減少を緩和させていることが示唆される．

図4には，さらに詳しく実験結果を分析するため，図3で示した開設施設数と，図2に示した各手法の処理時間に対する散布図を示す．図より，どのネットワークにおいても，CG法とSG法に関しては，開設施設にほとんど完全に比例した処理時間を要していることが分かる．これに対し，CGLE法とSGLE法に関しては，結果の開設施設数が同じでも，最大開設コストの設定により，処理時間が顕著に異なる場合が見取れる．特に，ある程度の範囲まで最大開設コスト f_{max} が大きくなると，施設数は増えるものの，処理時間は顕著に減少する．つまり，この範囲では，CGLE法とSGLE法の反復回数は増大しているものの，遅延評価の効果が大きく貢献すると考えられる．また，図2に示したように，CG法とSG法では，ある程度CG法の方が計算時間を要するのに対し，CGLE法とSGLE法では，逆に，CGLE法の処理効率の方が優る結果となっている．この理由については，SGLE法の評価関数 $H_{SG}(i; X)$ とCGLE法の評価関数 $H_{CG}(i; X) = H_{SG}(i; X)/f_i - 1$ を比較する

と，最大開設コスト f_{max} がある程度の範囲のとき， H_{SG} が小さい施設の数より H_{CG} が小さい施設の数が多いので，遅延評価が有効に機能すると考えられる． f_{max} が1に近い範囲では， H_{CG} と H_{SG} の違いが少ないので両者の計算時間はほぼ変わらない．また f_{max} が大きくなると， H_{SG} が小さい施設も多くなり，両者の計算時間は大差なくなる．

図5では，さらに詳しく解品質を分析するため，最大開設コスト f_{max} の設定に対し，電力ネットワーク，プログネットワーク，および，人名ネットワークにおける結果において，CG法とSG法の目的関数値の差分 $\Delta C = C(X_{CG}) - C(X_{SG})$ ，利用者の期待利得総和の差分 $\Delta C_S = C_S(X_{CG}) - C_S(X_{SG})$ ，および，開設施設集合の総開設コストの差分 $\Delta C_F = C_F(X_{CG}) - C_F(X_{SG})$ を比較している．ここで， X_{CG} と X_{SG} は，CG法とSG法で求めた開設施設集合を表す．図より，どのネットワークにおいても，期待利得総和の差分 ΔC_S に関しては顕著な差はみられず，CG法とSG法での目的関数値の差分 ΔC は，ほとんど総開設コストの差分 ΔC_F に起因していることが見て取れる．すなわち，SG法と比較してCG法では，開設コストが比較的低い施設を適切に選択できていることが示唆される．また，期待利得総和の差分 ΔC_S が小さいことより，最適解における性能の上限値推定についてはCG法でもSG法でもほとんど差はないことが示唆される．

上述した実験結果や分析結果をまとめると，ある程度の範囲で開設コストが異なる状況に

9 非線形施設配置問題における貪欲法の遅延評価による高速化——複数施設が確率的に貢献する場合

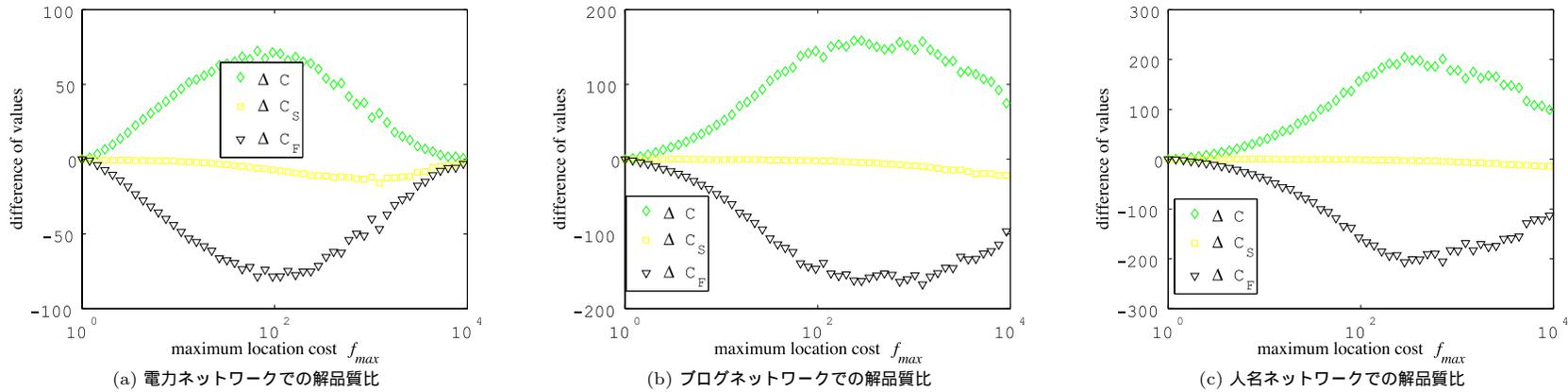


図5 開設施設数と ΔC , ΔC_S , ΔC_F の関係
 Fig. 5 Relationship between the number of facilities and ΔC , ΔC_S , ΔC_F .

において、SG法と比較して、CG法の有効性は顕著になることが示唆され、さらに、このような状況において、他の手法と比較して、CG法に遅延評価を導入したCGLE法の計算効率が高くなっている。現実的な施設配置問題における施設の開設コストを考えれば、これら値は施設ごとに異なるのが一般的であり、極度に開設コストが高い施設も多くは存在しないことが想定できる。したがって、現実的な状況での多様な施設配置問題において、CGLE法を用いれば、開設コストが比較的低い施設を適切に効率良く選択できると考えられ、このような問題を解くための基本法としてCGLE法は重要な役割を果たすことが期待できる。

8. おわりに

本論文では、複数施設が各利用者へ確率的に独立貢献する状況で開設コストを考慮した非線形施設配置問題を提案し定式化した。また、この非線形施設配置問題がサブモジュラ性を持つことを示し、施設追加による目的関数の増減分と、単位開設コストあたりの増減分を評価関数とする、2種の貪欲法SG法とCG法について述べるとともに、遅延評価導入によるそれぞれの高速化法SGLE法とCGLE法を提案した。3種の実ネットワーク構造上での非線形施設配置問題での評価実験では、施設開設コストのレンジを変化させ、各種手法の解品質と計算効率を実証評価し結果を考察した実験結果より、ある程度の範囲で施設開設コストが異なる時、他の方法と比較して、CGLE法は解品質と計算効率の両面で優れている

ことが示唆された。今後は、さらに多様な非線形施設配置問題について探求するとともに、それら問題にSGLE法やCGLE法を適用し、その有効性などの検証を進める。

謝辞 本研究は、科学研究費補助金基盤研究(C)(No.22500133)の補助を受けた。

参考文献

- 1) Kempe, D., Kleinberg, J. and Tardos, E.: Maximizing the spread of influence through a social network, *Proc. 9th International Conference on Knowledge Discovery and Data Mining*, pp.137–146 (2003).
- 2) Kimura, M., Saito, K., Nakano, R. and Motoda, H.: Extracting Influential Nodes on a Social Network for Information Diffusion, *Data Mining and Knowledge Discovery*, Vol.20, No.1, pp.70–97 (2010).
- 3) ベルンハルト コルテ, イェンス フィーゲン: 組合せ最適化第2版—理論とアルゴリズム, シュプリンガー・ジャパン (2009).
- 4) Krause, A. and Guestrin, C.: Near-optimal Nonmyopic Value of Information in Graphical Models, *Proc. 21st Conference in Uncertainty in Artificial Intelligence*, pp.324–331 (2005).
- 5) Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J. and Glance, N.: Cost-effective outbreak detection in networks, *Proc. 13th International Conference on Knowledge Discovery and Data Mining*, pp.420–429 (2007).
- 6) 室田一雄: 離散凸解析の考えかた—最適化における離散と連続の数理, 共立出版 (2007).

10 非線形施設配置問題における貪欲法の遅延評価による高速化——複数施設が確率的に貢献する場合

- 7) Nemhauser, G., Wolsey, L. and Fisher, M.: An analysis of the approximations for maximizing submodular set functions, *Math. Program.*, Vol.14, pp.265–294 (1978).
- 8) Newman, M.E.J.: The structure and function of complex networks, *SIAM Review*, Vol.45, No.2, pp.165–256 (2003).
- 9) 齊藤和巳：ウェブサイエンス入門—インターネットの構造を解き明かす，NTT 出版 (2007).
- 10) 齊藤和巳，武藤伸明，池田哲夫，入月卓也，永田 大，伊藤かの子：遅延評価導入による局所改善クラスタリング法の高速化，情報処理学会論文誌 数理モデル化と応用，Vol.3, No.1, pp.62–72 (2010).
- 11) Watts, D.J. and Strogatz, S.H.: Collective dynamics of ‘small-world’ networks, *Nature*, Vol.393, pp.440–442 (1998).

(平成 22 年 4 月 21 日受付)

(平成 22 年 6 月 11 日再受付)

(平成 22 年 7 月 26 日採録)



武藤 伸明

昭和 42 年生．平成 2 年横浜国立大学工学部電子情報工学科卒業．平成 4 年横浜国立大学大学院工学研究科博士課程前期修了．平成 7 年横浜国立大学大学院工学研究科博士課程後期修了．同年静岡県立大学経営情報学部助手．平成 19 年静岡県立大学経営情報学部准教授．専門は離散幾何学，グラフ論等．博士（工学）（横浜国立大学）．電子情報通信学会員．



齊藤 和巳（正会員）

昭和 38 年生．昭和 60 年慶應義塾大学理工学部数理科学科卒業．同年 NTT 入社．平成 3 年より 1 年間オタワ大学客員研究員．平成 19 年静岡県立大学経営情報学部教授．機械学習，複雑ネットワーク等の研究に従事．博士（工学）（東京大学）．平成 8 年情報処理学会論文賞等受賞．電子情報通信学会，人工知能学会，日本神経回路学会，日本応用数理学会各会員．



池田 哲夫（正会員）

昭和 31 年生．昭和 54 年東京大学理学部情報科学科卒業．昭和 56 年東京大学大学院理学系研究科情報科学専攻修士課程修了．同年日本電信電話公社（現 NTT）電気通信研究所入所．平成 14 年岩手県立大学ソフトウェア情報学部教授．平成 18 年静岡県立大学経営情報学部教授．専門は，データベース工学，情報検索，社会情報システム等．博士（工学）（東京大学）．ACM，IEEE CS，言語処理学会各会員．



永田 大

昭和 61 年生．平成 21 年静岡県立大学経営情報学部経営情報学科卒業．平成 21 年より静岡県立大学大学院経営情報学研究科経営情報学専攻在学中．複雑ネットワーク，データマイニング，離散幾何学に興味を持つ．



伏見 卓恭

昭和 63 年生．平成 21 年静岡県立大学経営情報学部経営情報学科中退．平成 21 年静岡県立大学大学院経営情報学研究科経営情報学専攻飛び級入学．現在，在学中．複雑ネットワーク，データマイニング，離散幾何学に興味を持つ．