オペレーションズ・リサーチ第二
近似アルゴリズム

担当: 前原貴憲 (理研AIP)

2019年11月7日

1　背景

これまで多項式時間で解ける組合せ最適化問題を見てきた．しかし，一般には多くの組合せ最適化問題は多項式時間では厳密に解くことができない (もしくはできないと考えられている)．そのような問題に対する対応は概ね３つある．

- 汎用性を諦める：入力を特殊なケースに制限する．たとえば一般グラフの代わりに二部グラフのみを扱うなど．
- 効率性：指数時間だが指数の底が小さいもの，たとえば $O(1.1^n)$ 時間などを求める．
- 最適性を諦める：多項式時間でそれなりによい解を求める．

３つ目のアプローチは「それなりによい」の部分に理論保証がつくかつかないかで２つに別れる．特に理論保証はないけれど良さそうな解を出力するものを**ヒューリスティクス (heuristics)**といい，理論保証のあるものを**近似アルゴリズム**という．

この授業では近似アルゴリズムについて議論する．

2　頂点被覆問題

$G = (V, E)$ を無向グラフ，$w\colon V \to \mathbb{R}_{\geq 0}$ を頂点重みとする．頂点部分集合 C が **頂点被覆 (vertex cover)** であるとは，であって，すべての枝が C のいずれかに接続することを言う．重み和最小の頂点被覆を求める問題を **重み付き最小頂点被覆問題 (minimum weighted vertex cover problem)** という．

2.1 LP緩和

IP を書くと次のようになる.

$$\begin{array}{ll} \text{minimize} & \sum_{u \in V} w_u x_u \\ \text{subject to} & x_u + x_v \geq 1, \\ & x_u \in \{0, 1\} \end{array} \tag{2.1}$$

残念ながらこの LP は整数性をもたない. $w = 1$ で三角形が反例となる.

2.2 近似アルゴリズム

(多項式時間で求まる) LP 解を使い，離散的な解を求めるアプローチを**線型計画に基づく近似 (LP-based approximation)**という.

LP 緩和問題をいて最適解 x^* を得る (これは多項式時間でできる). ここから離散的な解 x を次のように作る：

$x_u^* \geq 1/2$ であれば $x_u = 1$ とし，そうでなければ $x_u = 0$ とする.

定理 1. こうして作った解 x は頂点被覆であって最適解の 2 倍以下の目的関数値をもつ (このことを「近似率 2 の解である」という).

証明. まず，x が頂点被覆であることを確認する. 任意の枝 (u, v) について $x_u^* + x_v^* \geq 1$ なので，x_u^* もしくは x_v^* は 1/2 以上である. したがって x_u もしくは x_v は 1 となるので $x_u + x_v \geq 1$ が成立するので，x は頂点被覆である.

x の作り方から $x_u \leq 2x_u^*$ が成立する. 従って

$$\sum_{u \in V} w_u x_u \leq 2 \sum_{u \in V} w_u x_u^* = 2\text{LP} \leq 2\text{IP} \tag{2.2}$$

よって x の目的関数値は最適解の 2 倍以下である. □

一般に，LP 緩和問題を解いて得られた連続解から離散的な解を構成する手続きを「**丸め (rounding)**という.

2.3 集合被覆問題

頂点被覆問題の拡張として集合被覆問題を考える. 有限集合 E と，E の部分集合たち $\mathcal{F} = \{F_1, \ldots, F_n\}$ が与えられる. できるだけ少ない $S \subseteq \{1, \ldots, n\}$ を選ぶことで E の全部の要素をカバーしたい.

この問題は頂点被覆問題の一般化である. 実際，グラフがあって F_i を頂点 i に隣接する枝集合とした場合が頂点被覆問題である.

2.4 LP 表現

$V = \{1, \ldots, n\}$ とする．すると問題は

$$\begin{array}{ll} \text{maximize} & \sum_{u \in V} x_i \\ \text{subject to} & \sum_{u \in V; F_u \ni e} x_u \geq 1, \quad e \in E \\ & x_u \in \{0, 1\} \end{array} \tag{2.3}$$

となる．

3 近似アルゴリズム

この問題を解いた解 x^* から解を構築したい．ひとつの集合が任意個の要素をカバーする可能性があるので，頂点被覆のように簡単に閾値を切ることが難しい．

x_i は 0 のとき F_i を使わない，$x_i = 1$ のとき F_i を使う，ということを考えると，一般に「x_i は 1 になる確率」のような気がする．そこで以下の手続きを考える．

- 各 F_i を確率 x_i で選択する．

定理 2. このアルゴリズムの出力の期待値は最適値以下．

証明. $E[|S|] = \sum_{i \in V} x_i = \mathrm{LP} \leq \mathrm{IP}$ なので従う． □

しかし，この手続きの出力は実行可能解になるとは限らない (カバーされない要素があるかもしれない)．

定理 3. 任意の要素 $e \in E$ がこのアルゴリズムの出力でカバーされない確率は $1/e$ 以下．ただし $e = 2.71...$ は自然対数の底．

証明. e がカバーされるのは，$F_u \ni u$ なる F_u が少なくとも 1 つ選ばれることなので，その否定の確率は

$$\prod_{F_u \ni e} (1 - x_u) \leq e^{-\sum_{F_u \ni e} x_u} \leq e^{-1} \tag{3.1}$$

となる．ただし不等式 $1 - x \leq e^{-x}$ を使った． □

したがって，この手続きを T 回繰り返すと，要素 e がカバーされない確率を e^{-T} 以下にできる． E のうちどの要素もカバーされない確率はこれらの合計 me^{-T} 以下なので，$T = 1 + \log m$ にとると出力が集合被覆になる確率を $1/e$ 以下にできる．そのかわり，目的関数値は $T = 1 + \log m$ 倍悪化する．

よって，この手続きを出力が集合被覆になるまで繰り返すと，期待的に e 回で近似率 $1/(1 + \log m)$ の解が得られる．

4　演習問題

問題 1. 集合被覆問題に対する貪欲法は以下で与えられる：以下の手続きを全ての要素を被覆するまで繰り返す「まだ被覆されていない要素たちを最も多く被覆する集合 F_u を選ぶ」．この手続きが $1/\log m$ 近似を達成することを示せ．
ヒント：以下の手順に従って示せる．
1. 最適値を α^* とする．このとき貪欲法によって選ばれる集合 F_u は毎ステップ $1/\alpha^*$ 以上の割合の要素を被覆することを示せ．
2. 1 より，各反復で残り要素数が $1-1/\alpha^*$ 倍以下になる．このことから $\log(m)\alpha^*$ 回繰り返すと全ての要素が被覆されることを示せ．