修士論文 要旨

多次元割当問題の近似解法

数理情報学専攻 56207 黒木 裕介
指導教員 杉原 厚吉 教授
松井 知己 教授

概要

多次元割当問題は，割当問題の多次元への自然な拡張であり，データ結合問題のモデルになっている．本研究では，k 次元割当問題の入力は，頂点が有理 d 次元空間に埋め込まれた完全 k 部グラフであるとし，目的は，頂点集合の k クリークへの分割で，分割を構成するクリークの重みの総和が最小になるものを見つけることである．ただし，各 k クリークの重みは，クリークに含まれるすべての辺の重みの総和で定める．さらに各辺の重みは両端点のユークリッド距離の二乗で定める．

本研究では，上記の k 次元割当問題は，多くの仮定にもかかわらず，$k \geqslant 3$ において NP 困難であることを示した．さらに，二次錐計画緩和と多項式時間乱択丸め手続きを用いた $(5/2 - 3/k)$ 近似解法を提案する．

1 多次元割当問題

多次元割当問題を定義する．サイズ n の k 個の頂点集合 $V_1, V_2, \ldots, V_k$ が与えられているとし，族 $\mathcal{F} = \{V_1, V_2, \ldots, V_k\}$ と定義する．完全 k 部グラフ $G = (V_1, V_2, \ldots, V_k; E)$ を，頂点集合 $V_1, V_2, \ldots, V_k$ と辺集合 $E = \bigcup_{\{U,V\} \in \binom{\mathcal{F}}{2}} \{\{u, v\} \mid u \in U,\ v \in V\}$ で定める．G の頂点集合全体 $V_1 \cup V_2 \cup \cdots \cup V_k$ を $\widehat{V}$ で表す．グラフ G のすべての k クリークの重みが与えられたとき，G の頂点集合 $\widehat{V}$ の (互いに交わりのない) n 個の k クリークへの分割で，選ばれた n 個の k クリークの重みの総和を最小にするものを見つける問題を k 次元割当問題と呼ぶ．

以下の定義と仮定を導入する．辺重みベクトル $\boldsymbol{w} \in \mathbb{R}^E$ が与えられたとき，クリーク Q の重みを，Q で誘導される辺の重みの総和で定める．さらに，G の頂点は d 次元空間に埋め込まれており，辺の重みは両端点間のユークリッド距離の平方で定まっていると仮定する．以下では，問題の入力は k 個の，サイズ n の有理 d 次元ベクトルの集合 $V_1, V_2, \ldots, V_k$ (つまり，$V_1, V_2, \ldots, V_k \subseteq \mathbb{Q}^d$ かつ $|V_1| = |V_2| = \cdots = |V_k| = n$) であると仮定する．この仮定の下で，本研究で扱う問題は，各クラスタがすべての $V_i \in \mathcal{F}$ とちょうど 1 頂点ずつで交わるという条件の中で，二乗誤差の総和 (同じクラスタに含まれるすべての 2 頂点間の二乗距離の総和) を最小化するような n クラスタを求める問題である．

二乗誤差の総和を最小化する n クラスタリング問題は広く研究されている．また，データ結合問題の背景として考えられる確率論的モデルから自然に定義され，さらに，最適解に関する不変性を持つことから，問題設定には十分な意義がある．

本研究では，上記の問題が NP 困難であることを示し，さらに二次錐計画緩和を用いた乱択 $(5/2 - 3/k)$ 近似解法を提案する．すなわち，任意の問題例に対し，提案手法により求まる目的関数値の期待値が，最適値の $(5/2 - 3/k)$ 倍で抑えられる．この結果は，Bandelt–Crama–Spieksma [1] により得られていた近似率 $(4 - 6/k)$ を改善している．

2 定式化と緩和問題

任意の頂点部分集合 $V \subseteq \widehat{V}$ に対して，V と $\widehat{V} \setminus V$ を結ぶ E の辺の集合を $\delta(V)$ と書く．互いに交わりのない任意の頂点部分集合の組 $U, V \subseteq \widehat{V}$ に対して，辺部分集合 $\delta(U) \cap \delta(V)$ を，$E(U, V)$ や $E(V, U)$ と書く．単一集合 $\{v\}$ は，誤解の恐れがないときには，簡単のため v と書く．G の辺の列 (e_1, e_2, e_3) を，$\{e_1, e_2, e_3\}$ で誘導されるグラフが G において長さ 3 の閉路をなすとき，G の三角形と呼ぶ．任意のベクトル $\boldsymbol{x} \in \mathbb{R}^E$ の辺 $\{u, v\} \in E$ に対する要素 $x(\{u, v\})$ を，簡単のため $x(u, v)$ や $x(v, u)$ と記す．

0-1 変数のベクトル $\boldsymbol{x} \in \{0, 1\}^E$ を導入する．任意の辺重みベクトル $\boldsymbol{w} \in \mathbb{R}^E$ に対し，多次元割当問題は以下のように定式化できる:

$$
\begin{aligned}
\text{ILP: min. } & \textstyle\sum_{e \in E} w(e) x(e) \\
\text{s. t. } & \textstyle\sum_{u \in U} x(u, v) = 1 \quad (\forall U \in \mathcal{F},\ \forall v \in \widehat{V} \setminus U), \\
& x(e_1) \geqslant x(e_2) x(e_3) \\
& \qquad\qquad (\text{for } G \text{ の各三角形 } (e_1, e_2, e_3)), \\
& x(e) \in \{0, 1\} \qquad (\forall e \in E).
\end{aligned}
$$

任意の頂点 $v \in \widehat{V}$ について，($\mathbb{Q}^d$ 中の) 位置を $\boldsymbol{v} \in \mathbb{Q}^d$ で表す．緩和問題 QCLPR は以下のように書ける:

QCLPR:

$$
\begin{array}{lll}
\text{min.} & z & \\
\text{s. t.} & z \geqslant \sum_{\{u,v\}\in E} \|\boldsymbol{v}-\boldsymbol{u}\|^2 x(u,v), & \\
& z \geqslant \sum_{\{u,v\}\in \delta(U)} \|\boldsymbol{v}-\boldsymbol{u}\|^2 x(u,v) & \\
& \quad + \displaystyle\sum_{u\in U} \sum_{\{V,V'\}\in\binom{\mathcal{F}\setminus\{U\}}{2}} \left\| \sum_{v\in V} x(u,v)\boldsymbol{v} - \sum_{v'\in V'} x(u,v')\boldsymbol{v}' \right\|^2 & \\
& & (\forall U\in\mathcal{F}), \\
& \sum_{u\in U} x(u,v) = 1 & (\forall U\in\mathcal{F},\ \forall v\in\widehat{V}\setminus U), \\
& x(e)\geqslant 0 & (\forall e\in E).
\end{array}
$$

緩和問題 QCLPR は，凸二次制約の下で線形目的関数を最小化する問題であり，二次錐計画問題に等価変形できる．二次錐計画問題は，内点法などにより，多項式時間で (誤差 ε を許して) 最適解を求めることができる．

例 1. $k=3, n=2, d=2$ の問題例を示す．集合 V_1, V_2, V_3 を示す記号をそれぞれ◯, ●, ■とし，V_1 を構成する頂点を①, ②と記し，ほかも同様とする．半径 1 の円周の 6 等分点上に，順番に①, ❶, 1, ②, ❷, 2 を配置した頂点集合を入力とする．

最適解は図 1(a) に示した分割で，最適値は 10 である．緩和問題 QCLPR の最適値は 8.000 である (解は図 1(b) を参照). □

(a) 最適解．線が描かれている (いない) 辺は変数の値が 1 (0).

(b) QCLPR の最適解．太線，細線の辺は変数の値がそれぞれ 0.8333, 0.1667.

図 1. 小さな問題例 $(k=3, n=2, d=2)$.

3 乱択近似アルゴリズム

緩和問題 QCLPR の実行可能解を $(z,\boldsymbol{x})$ とする．任意の頂点部分集合の組 $\{U,V\}\in\binom{\mathcal{F}}{2}$ に対して，$\boldsymbol{x}$ のうち $E(U,V)$ に対応する成分からなる部分ベクトル $\boldsymbol{x}|_{E(U,V)}$ は，完全 2 部グラフ $(U,V;E(U,V))$ で定まる割当多面体

$$
\left\{ \widetilde{\boldsymbol{x}}\in\mathbb{R}_{+}^{E(U,V)} \;\middle|\; \begin{array}{l} \sum_{u\in U}\widetilde{x}(u,v)=1\ (\forall v\in V), \\ \sum_{v\in V}\widetilde{x}(u,v)=1\ (\forall u\in U) \end{array} \right\}
$$

に含まれる．割当多面体の整数性から，部分ベクトル $\boldsymbol{x}|_{E(U,V)}$ は，2 部グラフ $(U,V;E(U,V))$ 中の完全マッチングの特性ベクトルの凸結合によって表現できる．そこで，次の乱択丸め手続きを得る:

手続き 2.

入力． 頂点集合 U と QCLPR の実行可能解 $(z,\boldsymbol{x})$ の部分ベクトル $\boldsymbol{x}|_{\delta(U)}$.

出力． 0-1 ベクトル $\boldsymbol{X}\in\{0,1\}^{\delta(U)}$.

各頂点部分集合 $V\in\mathcal{F}\setminus\{U\}$ に対して以下を行う．

ステップ 1． 部分ベクトル $\boldsymbol{x}|_{E(U,V)}$ を 2 部グラフ $(U,V;E(U,V))$ 中の完全マッチングの特性ベクトルの凸結合によって表現する; 完全マッチング M に対応する凸結合係数を $\lambda(M)$ と記す;

ステップ 2． $(U,V;E(U,V))$ 上の完全マッチングを一つ，完全マッチング M を選ぶ確率を $\lambda(M)$ として，選ぶ;

ステップ 3． 部分ベクトル $\boldsymbol{X}|_{E(U,V)}$ を，選ばれた完全マッチングの特性ベクトルとする． □

乱択近似アルゴリズムを以下のように設計する．

アルゴリズム 3.

入力． 集合 $V_1,\dots,V_k\subseteq\mathbb{Q}^d$ $(|V_1|=\cdots=|V_k|=n)$.

出力． ILP の実行可能解 $\boldsymbol{X}$.

ステップ 1． QCLPR を解いて最適解 $(z^*,\boldsymbol{x}^*)$ を得る;

ステップ 2． ランダムに頂点集合 $U\in\mathcal{F}$ を選ぶ;

ステップ 3． 手続き 2 を部分ベクトル $\boldsymbol{x}^*|_{\delta(U)}$ に適用し，$\delta(U)$ の成分に対応する 0-1 ベクトル $\boldsymbol{X}_U$ を求める;

ステップ 4． 0-1 ベクトル $\boldsymbol{X}\in\{0,1\}^E$ を以下で定める:

$$
X(e)=\begin{cases} X_U(e) & (\forall e\in\delta(U)), \\ \displaystyle\sum_{u\in U} X_U(u,v)X_U(u,v') & (\text{その他}). \end{cases} \quad \square
$$

次の定理が主結果である．

定理 4. アルゴリズム 3 によって，目的関数値の期待値が $(5/2-3/k)\,z^{**}$ 以下の ILP の実行可能解が求まる．ただし，z^{**} は集合 $V_1,\dots,V_k\subseteq\mathbb{Q}^d$ $(|V_1|=\cdots=|V_k|=n)$ で定まる多次元割当問題の最適値であるとする． □

4 困難性

本論文で扱う多次元割当問題は，$k\geqslant 3$, $d\geqslant 2$ のとき NP 困難であることを，planar 3DM と呼ばれる NP 完全問題 [2] から帰着することで示した．

参考文献

[1] Bandelt, H.-J., Crama, Y., Spieksma, F. C. R.: Approximation algorithms for multi-dimensional assignment problems with decomposable costs. *Discrete Applied Mathematics*, **49** (1994), 25–50.

[2] Dyer, M. E., Frieze, A. M.: Planar 3DM is NP-complete. *Journal of Algorithms*, **7** (1986), 174–184.