グラフ分割問題の解構造とAR(1)モデル

01107771 小樽商科大学 加地太一 KAJI Taichi

1 はじめに

本研究の大きな背景はメタヒューリスティックスアルゴリズムの優位性を明確にするため、理論的にその求まる解の特性を考察することにある。その研究の一つとして、Eikelder等[2]は巡回セールスマン問題(TSP)を対象としてLocal Searchによって得られる解の値、および要求される反復数の理論的な期待値を検討している。そこでは解のインスタンスt_0がコストc_0をもつときその近傍のすべてが与えられたコストcより大である確率（ステップ確率）を必要としている。これを求めるのにEikelder等はTSPの限定した近傍構造に対して問題独自に導出している。著者はグラフ分割問題に対してあらたにその解の特性を分析したい。そこで、上記のステップ確率を導出するため、解構造がAR(1)モデルにしたがっているという仮定のもとそのステップ確率を導出する方法を考察した[5]。そこで問題はグラフ分割問題の解構造が真にAR(1)モデルに従っているかということに帰着される。本論分ではグラフ分割問題の解構造を分析し、AR(1)モデルとして仮定できるであろうかをシミュレーションを通して検証したい。

2 グラフ分割問題と近傍構造

グラフ分割問題はNP困難な組合せ最適化問題の一種であり、VLSI設計、配置問題、配送問題、プログラム分割問題など多くの場面で利用されている。このモノグラフで扱うグラフ分割では、固定された大きさ$M_1, M_2, \ldots, M_k$の成分からなる頂点集合のk分割を対象としている。

ここで、n個のノード集合Vとエッジ集合Eからなるグラフを$D(V,E)$とし、大きさ$M_1, M_2, \ldots, M_k$の成分をもつVのk分割$t = \{V_1, V_2, \cdots, V_k\}$をグラフ分割問題の解$t$、$V_i$をブロックと呼ぶ。各ブロックのノード数は固定された値（ブロックサイズ）M_iをもつ。また、すべての可能な解からなる集合を解空間Tと呼ぶ。解tを形成する各ブロックは分割の成分であるから、当然網羅的かつ排他的条件$V_1 \cup V_2 \cup \cdots \cup V_k = V$かつ$V_i \cap V_j \; for \; i \neq j$を満たしている。

エッジ$(i,j) \in E$のコストをe_{ij}とするとき、解tの評価値$f(t)$を

$$f(t) = \sum_{1 \leq l < m \leq k} \sum_{i \in V_l, j \in V_m} e_{ij}$$

と定義する。

グラフ分割問題は次式のように、解空間Tの上で評価値$f(t)$を最小にする解t_0を見出す問題として定式化される。

$$f(t_0) = \min_{t \in T} f(t) \tag{1}$$

また、今回の報告では分割数を$k = 2$に限定した。さらに、Local Searchに使用する近傍系構造$N(x)$は二つのブロックの間の頂点を交換する方法を用いており、それは以下のように定義される。

$$\begin{aligned} N(x) = \{x' = (V_1, V_2, \cdots, V_s', \cdots, V_w', \cdots, V_k) \\ : V_s' = (V_s \setminus \{i\}) \cup \{j\}, V_w' = (V_w \setminus \{j\}) \cup \{i\} \\ , i \in V_s, j \in V_w, s \neq w\} \end{aligned} \tag{2}$$

3 LandscapeとAR(1)モデル

本論文ではグラフ分割問題の解構造であるlandscapeがAR(1)モデルに従うかをシミュレーションを通して解析するものである。

すべての可能な分割tの集合をTとしよう。そして、与えられた分割からある基本操作で別の分割を導出することを移動と呼び、その移動の集合によって、任意の分割に対して近傍集合が定義できる。各分割tに対しての評価値（コスト）を$f(t)$で表すこととする。その写像$f : T \mapsto R$を分割問題の評価値と呼びその評価値の形状の族をlandscapeと呼ぶこととする。landscapeという言葉は理論生物学に端を発し、組合せ的対象から実数値への写像に関する問題に広がりつつある。最近、Weiberger[4]は不偏的ランダムウォークの概念がlandscape構造を調査するための適切な手法であることを示した。ここで、"landscapeは統計的にisotropicである"という重要な付加的条件を課することが必要である。

ランダムに選んだ点(分割)を出発点としてランダムに選ばれた近傍点に移動し、この点から再びランダムに選ばれた近傍点に移動することなどを繰り返すことによって得られた評価値(fitness)の系列を考える。この評価値の系列の統計が、選ばれ

た出発点にかかわらず同じであるとき、考えている landscape は統計的に等方的なランドスケープ (statistically isotropic landscape) であるという。このとき、分割問題の探索過程のステップ t_i における評価値 $f(t_i)$ がつくる時系列は、あるランダムウォークの結果であると考えることができ、定常確率過程 (stationary process)[1] を形成する。また、定常確率過程の一つの部分タイプとして AR(1) プロセス (autoregressive process of order one) [4] と呼ばれる定常過程がある。それは次の方程式をみたす過程である。

$$X_t = z_L X_{t-1} + N_t \tag{3}$$

ただし、N_t は無相関確率変数の定常系列である。Weinberger [4] は、"AR(1) プロセスが、N-K 問題や 組合せ問題を含めて、landscape の広いクラスの上でランダムウォーク の統計をうまく捕らえるであらうと考えることには十分納得のいく理由がある"と述べている。また、Gaussian かつ Markov である定常確率過程は、その自己相関関数は指数的な減衰を持つことが分かっており、この性質を調べることにより分割問題が AR(1) であるかを判明することが可能となろう。分割問題の landscape が AR(1) 方程式に従うことにより、解の特性を考察するために必要となるすべての統計量が求まるものである。

4 シミュレーションによる AR(1) モデルの分析

本論分では分割問題の解空間における Landscape 上のランダムウオークが評価値について、AR(1) モデルに従う確率過程になることをシミュレーションによって調べる。解空間上の偏りのないランダムウォークによる landscape から得られたある解の時系列が得られ、その時系列上の解のコストからある自己相関関数が求まる。landscape から得られた自己相関関数が以下の式を満たすとき、AR(1)landscape と呼ぶ [4]。

$$\rho(s) = \rho(1)^s = exp(-s/\lambda) \tag{4}$$

ただし、sは 2 点間の距離を表し、$\lambda = -1/\log\rho(1)$ は相関長を表す。グラフ分割問題の landscape が AR(1) であるか判定するためには、landscape 上をランダムウォークすることによって、そのウォーク t_i にそったコスト $f(t_i)$ をサンプリングし、以下の式より時系列の自己相関関数の値を導出して検証することとなる。

$$\rho(s) = \frac{1}{\sqrt{var(f(t_i)) \cdot var(f(t_{i+s}))}} \times \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \{f(t_i) - E[f(t_i)]\} \times \{f(t_{i+s}) - E[f(t_{i+s})]\} \tag{5}$$

この (5) 式の値が、(4) 式の値に近似されるならば、すなわち自己相関関数が指数的減衰性を示すならば、グラフ分割問題の landscape は AR(1)landscape として呼ぶことが可能であろう [3]。

また、AR(1)landscape であることにより、確率過程 $E[f(t_k)]$ は次の再帰方程式で支配される。

$$E[f(t_k)] = \mu + \rho(1)\{E[f(t_{k-1})] - \mu\} + \Delta \tag{6}$$

ただし、定常過程の定義 [1] から $E[f(t_k)] = \mu$ for all k が成立する。また、Δ は平均 zero、分散 d^2 をもつ白色雑音である。この AR(1) 方程式 (6) 式から今後必要となるすべての統計量が導出可能となる。これらの推定される理論値についても実験値との整合性を調査していきたい。

5 おわりに

グラフ分割問題の landscape が AR(1) であれば、Local Search で求まる局所解のコスト、および必要なステップ数の確率密度などの理論的推定値を求めることが可能となる。それによりこのアルゴリズムの求めうる値の能力、また、必要な計算時間の予測が可能となる。そこで、本論文ではグラフ分割問題の landscape が AR(1) であるかを検証するために、その解構造をシミュレーションを通して分析し解析を試みる。そのために、解のランダムウォークによる時系列上の自己相関関数の特徴が重要なポイントとなる。解の時系列から得られたその自己相関関数の特徴、その他推定される理論値についての特性は発表当日報告する。

参考文献

[1] Karlin,S. and Taylor,H.M., A First Course in Stochastic Processes, Second Edition, Academic Press,1975.

[2] Osman,I.O. and Kelly,J.P.(ed.), Meta-Heuristics: Theory & Applications, Kluwer Academic Publishers,pp.605-618,1996.

[3] Stadler,P.F. and Schnabl, W., "The Landscape of the Traveling Salesman Problem", Phisics Letters A, Vol.161,Num.4,pp.337-344,1992.

[4] Weinberger,E., "Correlated and Uncorrelated Fitness Landscapes and How to Tell the Difference", Biological Cybernetics,Vol.63,pp.325-336,1990.

[5] 加地, Local Search の確率的解析による性能評価, 商学討究（小樽商科大学）, 第 51 巻, 第 4 号,pp.193-211,2001.