第6回システム最適化に関するシンポジウム講演論文集　1999年12月

(17) 確率的最適化手法の情報理論による考察

A BASIC CONSIDERATION ON COMBINATORIAL OPTIMIZATION PROBLEM USING INFORMATION THEORY

須藤 敦史*・星谷　勝**
Atsushi SUTOH and Masaru HOSHIYA

* 博士(工学)(株)地崎工業 土木技術部 主任研究員(〒105-8488 東京都港区西新橋2-23-1)
** Ph.D.　武蔵工業大学教授 工学部土木工学科　(〒158-8557 東京都世田谷区玉堤1-28-1)

This paper deals with an interpretation on stochastic combinatorial optimization algorithm using Markov process and Information theory. Combinatorial optimization problems are essential for seeking a specific set among other alternatives. These problems cannot be solved by standard optimization techniques such as the Newton method. And, these problems are almost always exposed to the danger of falling in local minima. Under these circumstances, techniques which integrate biological evolution process (Genetic Algorithm), physical process (Simulated Annealing), and stochastic processing (Random Sampling) have been developed. These techniques are usually cast into a probabilistic and information framework via such paradigms as combinatorial optimization, stochastic based algorithm.

Key Words: *combinatorial optimization problems, stochastic processing, information entropy*

1. はじめに

組み合わせ最適化問題は,多数の解候補から目的関数の最大あるいは最小化かつ制約条件に適合するような特定な組み合わせを探し出す決定論的な問題であり,ニュートン法などの勾配法のように目的関数の接線勾配を利用できないため最適解の探索は難しく,加えて常に局所解への滞留問題も生じるため,最適解の探索には勾配法とは基本的に異なった解法が必要となる.

このような組み合わせ最適化問題の解法として Genetic Algorithm : GA[1], Simulated Annealing : SA[2], Random Sampling[3]を基本とした Modified Importance Sampling[4,5]などの組み合わせ最適化問題の解探索アルゴリズムに物理・生物科学的概念や確率・統計的な技法を導入した発見的探索法と称される近似解法が用いられている.

しかし,その有用性にも関わらず探索アルゴリズムの基本的な考え方が不明瞭であったり,構成アルゴリズムや制御パラメータの設定が経験的になることも指摘されている.したがって,このような発見的探索法を有効に活用して,汎用的な解法として発展させるためには,組み合わせ最適化問題の特徴を整理するとともに探索方法の基本的な考え方を明確にする必要が生じる.

そこで本研究では,組み合わせ最適化問題の特徴と探索上の問題点を整理し,探索アルゴリズムに確率・統計的な技法を導入する意義やその効果を確率過程(離散マルコフ決定過程)および情報理論により考察している.

2. 組み合わせ最適化問題,探索方法の整理

探索アルゴリズムを評価するときに探索(解)空間がどのようなものか明らかにすることが重要である.そこで,ここでは組み合わせ最適化問題の特徴を整理して発見的探索法における問題の考え方を考察する.

2.1 組み合わせ最適化問題の特徴

組み合わせ最適化問題は実数の目的関数 $f(x)$ を用いて式(1)のように定義される.

$$\min_{\{x\}}\{f(x) \mid x \in X\} \tag{1}$$

x : 離散状態変数ﾍﾞｸﾄﾙ, X : 状態空間における有限集合

ここで組み合わせ最適化問題は,解候補すべてを探索すれば最適解が求められる決定論的な問題である.しかし実際の問題では,一つ一つの解候補を評価することは計算量が増加するため,現実的には不可能となり,同時に局所解への滞留問題も生じる.加えて,勾配法のように目的関数からの直接情報が得られないため,何らかの方法により探索情報を得なければならず,また得られた情報を活用した探索アルゴリズムを構成しなければならない.

2.2 発見的探索法の基本的な考え方

基本的に発見的探索法は,最良解が得られれば良く,加

えて最適性は収束性で判断する，という前提で以下のような探索アルゴリズムを構成している．

(1) 評価尺度（目的関数）を設定し，解候補をより良い解候補で置き換えて改良していく解法である．

(2) 複数の解候補から直接探索情報を入手し，同時に解候補の集合特性（平均値・分散値）に着目することで計算量の軽減を図っている．

(3) ダーウインの進化論（GA），物理現象(SA)，離散マルコフ決定過程(MIS 法)などの利用により，基本的に勾配法とは異なった探索アルゴリズムを構成している．

以上のように，組み合わせ最適化問題（決定論的問題）における解候補を集合的に扱うこと（確率化）で，少ない計算量でより良い解候補を得ようとするのが発見的探索手法の特徴である．

２．３組み合わせ最適化問題の確率化

組み合わせ最適化問題の確率化は，解候補集合：$x_1, x_2, \cdots, x_n$（n 個）を発生させる確率分布：$p(x_1), p(x_2), \cdots, p(x_n)$ の期待値と分散値を探索情報を基に設定してゆく問題となり，解候補の期待値は式(2)となる．この時，確率分布の分散を極力小さく，期待値が最適解になるように設定すれば，出現する解候補が最適解となる確率は高くなる．

$$E[x] = \sum_{i=1}^{n} x_i p(x_i) \qquad (2)$$

$p(x_i)$：候補 x_i の出現確率，n：候補総数，

$$0 \le p_i \le 1, \sum_{i=1}^{n} p(x_i) = 1$$

したがって，効率的な探索法とはいかに早く解候補を発生させる確率分布の期待値を最適解に，同時に小さい分散値を設定するアルゴリズムを構成するかが決め手となり，GA ではダーウインの進化論，SA では物理現象，また MIS 法では離散マルコフ決定過程によりを構成している．

３．情報理論による発見的探索法の考察

Random Samplingを基本としたMIS 法や突然変異を用いないSimple GAは確率過程（離散マルコフ決定過程）によって定式化できるが，これを解説した文献[6]は少ない．

そこで単純なアルゴリズムを有するMIS 法を例に上げて，探索アルゴリズムを確率的な考え方で整理し，同時に情報理論により考察する．

３．１確率過程（離散マルコフ決定過程[7]）による考察

GA やMIS 法では，探索初回において解の事前情報がない場合に解候補集合は式(3)に示すように，N 個の確率変数ごとに領域 D_N^1（M：取り得る候補数）から，それぞれランダムに n 個の発生させる．

（第１回目-解候補 n 個）

$$x_{(1)}^n = \{x_1^n \in D_1^1, x_2^n \in D_2^1, \cdots, x_N^n \in D_N^1\} \qquad (3)$$

$x_{(1)}^n$：初回の解候補集合，x_i^n：i 番目の設計変数の実現値

次に N 個の確率変数ごとに解候補（n 個）から平均値以上の目的関数値を有する候補を選抜して，それらが分布する範囲を領域 $D_1^1, \cdots, D_N^1$ から選定し，次回の探索領域 $D_1^2, \cdots, D_N^2$ とする．

一方 SimpleGA では適応度(制約条件を満足する)の高い個体を選択し，その遺伝子(確率変数)を交配させて，次回の解候補集団を作成する．したがって，いずれの手法とも探索 t 回目の解候補集合は $t-1$ 回目の解候補集合に依存する離散マルコフ過程(連鎖)を示しており，これを条件付き確率で表すと式(4)となる．

$$p(x_{(t)}^n \mid x_{(1)}^n, x_{(2)}^n, \cdots, x_{(t-1)}^n) = p(x_{(t)}^n \mid x_{(t-1)}^n) \qquad (4)$$

式(4)を各確率変数ごとに推移確率行列 P による確率分布 π_t から π_{t+1} への状態推移で示すと式(5)となる．

$$\pi_{t+1} = \pi_t \cdot P \qquad (5)$$

π_t：状態確率分布($1 \times M$)，M：取りうる候補の数，

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{bmatrix}$$:推移確率行列

ここで探索初回では，領域 $D_1^1, \cdots, D_N^1$ の候補（M 個）の中からランダムに選択(推移確率行列の各要素 p_{ij} は確率 $1/M$)される．ここで推移確率行列 P において最適解となる要素の確率が１，それ以外の要素が０になれば，常に最適解のみが発生する唯一の分布状態となる．しかし，どの要素が最適解であるかは不明なため，何らかの方法で確率１になる要素の位置を求めなければならない．

そこで，要素位置を求める唯一の情報は，推移した候補（探索初回で残った候補）の個数 n_{ij} であるため，これを用いて確率１になる要素位置の推定を考える．

いま，解候補の確率分布を多項分布として，推移

（残った）解候補の個数n_{ij}で表すと式(6)となり，推移確率p_{ij}を未知数とすると式(6)の対数尤度関数$\log L$は式(7)となる[8].

$$F(p_{11},\cdots,p_{MM})=\frac{n_i!}{\prod_{i=1}^{M}\prod_{j=1}^{M}n_{ij}!}\prod_{i=1}^{M}\prod_{j=1}^{M}p_{ij}^{n_{ij}} \tag{6}$$

$$\because\ n_i=\sum_{j=1}^{M}n_{ij}$$

$$\log L=\log(n_i!\Big/\prod_{i=1}^{M}\prod_{j=1}^{M}n_{ij}!)+\sum_{i=1}^{M}\sum_{j=1}^{M}n_{ij}\log p_{ij} \tag{7}$$

$$\log L'=\sum_{i=1}^{M}\sum_{j=1}^{M}n_{ij}\log p_{ij} \tag{8}$$

$$p_{ij}\geq 0,\ i,j=1,2,\cdots M \tag{9. a}$$

$$\sum_{j=1}^{M}p_{ij}=1,\ i=1,2,\cdots M \tag{9. b}$$

ここで式(7)の右辺第1項は定数となるため，式(9)の制約条件の下で式(8)を最大にする推移確率p_{ij}を求める問題となる．よって，ラグランジェの未定係数法により推移確率行列の要素p_{ij}は，式(10)のように残った解候補の個数n_{ij}で表される．

$$p_{ij}=n_{ij}/n_i \tag{10}$$

したがって，推移確率行列の各要素p_{ij}は推移した解候補数をその総数で除した値となり，同時に分布も得られる．

3．2情報理論[9]による探索アルゴリズムの考察

観測データは未知特性に関するあいまい性を減少させるものであり，未知特性に関してどれだけの情報をもたらすかを定量的に定義できる．そこで情報理論では定量的な評価指標として情報エントロピーを導入している．ここで情報エントロピーとは「不確定性の度合い」といった極めて抽象的なものを定量的に評価する指標であり，事象の情報エントロピーはその確率分布の形態（分散の状態や度合い）によって一意的に定まる．

ここで離散型の確率分布を有する事象A，Bを考える．事象Aの情報エントロピーは，式(12)のように定義される．（Bについても同様）

$$\mathbf{A}=\begin{pmatrix}A_1 & A_2 & \cdots & A_n\\ p_1 & p_2 & \cdots & p_n\end{pmatrix}\quad 0\leq p_i\leq 1\quad \sum_{i=1}^{n}p_i=1 \tag{11. a}$$

$$\mathbf{B}=\begin{pmatrix}B_1 & B_2 & \cdots & B_m\\ q_1 & q_2 & \cdots & q_m\end{pmatrix}\quad 0\leq q_j\leq 1\quad \sum_{j=1}^{m}q_j=1 \tag{11. b}$$

$$H(\mathbf{A})\equiv H(p_1,\cdots,p_n)=-\sum_{i=1}^{n}p_i\log p_i \tag{12}$$

ただし$\log p$の底はeとし，$0\log 0=0$とする．

ここで情報エントロピー$H(\mathbf{A})\equiv H(p_1,\cdots,p_n)$の基本的性質を示す．

$$\min H(p_1,\cdots,p_n)=0 \tag{13. a}$$

$$H(p_1,\cdots,p_n)\geq 0 \tag{13. b}$$

$$\max H(p_1,\cdots,p_n)=H(1/n,1/n,\cdots,1/n)=\log n \tag{13. c}$$

式(13. a)は確率要素p_iの一つが 1.0 でそれ以外の要素p_iがすべて0（最適解が選択された時：確定的現象）の場合に成立し，式(13. b)では情報エントロピーは正値であることを示している． また式(13. c)はすべての事象が等しい確率で生起する場合に成立し，この時の情報エントロピーは最大となる．

ここで式(13. c)から分かるように，組み合わせ最適化問題における解候補の一様分布からの選抜は情報エントロピーが最大になるようにサンプルされている．これは解に関する事前情報が全くない場合における偏りのない探索を意味しており，推移確率行列の要素を設定する合理的な手法である．（無差別原理）

いま，情報エントロピーをによる簡単な探索事例としてサイコロを振る確率モデルを考える．

$$X_1=\begin{pmatrix}a_1, & a_2, & a_3, & a_4, & a_5, & a_6\\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6}\end{pmatrix} \tag{14}$$

上段の記号a_iはiの目が出る事象X_1，下段は個々の事象a_iの生起確率であり，このときの情報エントロピーは式(15)となる．つまりサイコロの一つの目が出現する不確定性の評価値を$-\log p(a_i)=-\log 1/6$とすれば，事象全体の不確定さの平均量が情報エントロピーである．

$$H(X_1)=\sum_{x}^{6}-\frac{1}{6}\log\frac{1}{6}=1.792 \quad (15)$$

次に１と６の目が候補から削除されたとすると事象の確率(16)となり，その時の情報エントロピーは式(17)となる．

$$X_2=\begin{pmatrix} a_1, & a_2, & a_3, & a_4, & a_5, & a_6 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \end{pmatrix} \quad (16)$$

$$H(X_2)=\sum_{x}^{4}-\frac{1}{4}\log\frac{1}{4}=1.386 \quad (17)$$

同様に２と５の目が候補から削除されたすると，情報エントロピーは式(18)となる．

$$H(X_3)=\sum_{x}^{2}-\frac{1}{2}\log\frac{1}{2}=0.693 \quad (18)$$

ここで，候補数が少なくなるにつれて情報エントロピーは小さくなっていることより，解候補の不確定性の減少を示している．したがって，情報エントロピーが最小となるような解候補は，ある特定な組み合わせ（最適解）であり，これは前節における推移確率行列において最適解となる要素確率を１，それ以外を０にする唯一の分布を形成することと同様の操作を行っていることとなる．

一方，組み合わせ最適化問題の探索過程における解候補は一つ前の解候補に依存するマルコフ過程（連鎖）を示し，その事象は条件付き確率で表される．いま，確率モデルX，Yとその結合確率モデル$Z=X\cdot Y$が定義される事象の条件付き確率の情報量を考える．

$$X=\begin{pmatrix} a_1 & a_2\cdots & a_M \\ p_1 & p_2\cdots & p_M \end{pmatrix} \quad (19.\mathrm{a})$$

$$Y=\begin{pmatrix} b_1 & b_2\cdots & b_M \\ q_1 & q_2\cdots & q_M \end{pmatrix} \quad (19.\mathrm{b})$$

$$Z=X\cdot Y=\begin{pmatrix} a_i\cap b_j & i=1,2,\cdots M \\ r_{ij} & j=1,2,\cdots,M \end{pmatrix} \quad (19.\mathrm{c})$$

r_{ij}：事象a_i，b_jが同時に生起する確率

ここで事象a_iが得られたときの事象b_jに対する条件付き確率の情報エントロピー（相互情報量）は次式となる．

$$H(Y\mid X)=\sum_{j=1}^{M}\sum_{i=1}^{M}-P(a_i)P(b_j\mid a_i)\log P(b_j\mid a_i) \quad (20)$$

$$P(b_j\mid a_j)=\frac{P(a_i\cap b_j)}{P(a_i)} \quad (21)$$

また，式(20)は式(21)より式(22)となる．

$$H(Y\mid X)=\sum_{j=1}^{M}\sum_{i=1}^{M}-P(a_i\cap b_j)\log P(b_j\mid a_i) \quad (22)$$

この時，条件付き確率の情報エントロピーにおいて事象a_iとb_jが同時に生起した確率$P(a_i\cap b_j)$を実際に推移した解候補の数n_{ij}で置き換えると式(8)で示した対数尤度関数$\log L'$と同じになる．また次回の探索確率は情報エントロピー最大とする条件付き確率$P(b_j\mid a_i)$として式(22)から求められ，推移確率p_{ij}を求める問題と同様に要素値は残った候補数をその総数で除した値となる．

５．まとめ

本研究では，組み合わせ最適化問題の特徴を整理し，同時に発見的探索法の離散マルコフ決定過程および情報理論による考察を行い，以下の結論が得られた．

(1) 組み合わせ最適化問題の確率化は，解候補が生起する確率分布を候補集合の情報より，分散が小さく期待値を最適解近傍に設定する問題に置き換えている．

(2) 離散マルコフ決定過程では，推移確率行列の唯一の分布が最適解を生起する状態となり，推移した候補の数からその要素位置は推定される．

(3) 情報理論では，解候補は情報エントロピーが最大になるように生起させ，情報エントロピーが最小となる特定な組み合わせが最適解となる．

参考文献

1) Goldberg, D. E.: *Genetic Algorithm*, Addison-Wesly, 1983.

2) Aarts, E. and Korst, E.: *Simulated Annealing and Boltzmann Machine*, Jhon Wiley, 1989.

3) Rubinstein, R. Y.: *Simulation and Monte Carlo Method*, John Wiley, 1981.

4) 須藤敦史,星谷勝,宮沢和樹:遺伝的要素を考慮したｲﾝﾎﾟｰﾀﾝｽｻﾝﾌﾟﾘﾝｸﾞによる離散型変数を有するｼｽﾃﾑの最適化,土木学会論文集,第519号, I-32, pp.223-232, 1995.

5) Hoshiya, M. and Sutoh, A.: Optimization Analysis by Importance Sampling, GA procedure and other MCS-Based Algorithms, *Jour. of Prob. Eng. Mch*, Vol. 12, No. 14, pp. 221-223, 1997.

6) 山村雅幸,織田悦子,小林重信:ﾏﾙｺﾌ過程によるSimple GAの解析,日本機械学会,第2回FANｼﾝﾎﾟｼﾞｳﾑ論文集, pp. 383-388, 1992.

7) 森村英典,高橋幸雄:ﾏﾙｺﾌ解析,日科技連, 1985.

8)西尾勝:自然科学の統計学,東京大学出版会, 1996.

9)有本卓:確率･情報ｴﾝﾄﾛﾋﾟｰ,森北出版, 1994.