多点探索法によるヒューリスティックな統計推論

5Ｋ－８

相澤 彰子
(学術情報センター)

1 はじめに

遺伝的アルゴリズムにおける問題の容易性の分析において，Walsh 解析や重回帰分析的な手法により問題に内在する非線形性を数値で表現する [1][2] などの方法が検討されている．しかし一般に，これらの手法が適用可能であるのは，問題のサイズが小さく全解探索が可能な場合に限られる．そこで本稿では，与えられた解集合に基づき有用な部分解を推測する方法について予備的な検討を行った結果を報告する．

2 分割の定義

L 個の 2 値変量 $(x^{(1)}\cdots x^{(L)})=\boldsymbol{x}$, $x^{(i)}\in\{0,1\}$ を実数値に対応させる関数 f が与えられているとする．このとき，分割 Π_i $(0\leq i\leq 2^L-1)$ を次のように定義する．まず，添字 i を長さ L の 2 進数で表現したときに，その中に含まれる '1' の総数を K_i（Π_i のレベルと呼ぶ）として，'1' のビット位置に対応する K_i 個の数の集合 $\{n_1,\cdots,n_{K_i}\}$ を考える．分割 Π_i とは，解の全空間 A $(\boldsymbol{x}\in A,\ |A|=N=2^L)$ を，変量 $x^{(n_1)},\cdots,x^{(n_{K_i})}$ の値を基準として $N_\Pi=2^{K_i}$ 個の部分空間に分けるものである．以下，分割 Π_i の添字 i の 2 進数表現の '0' を '*' で，'1' を '.' で表記する．

Π が生成する N_Π 個の部分空間の大きさは等しい．これを $N_{\overline{\Pi}}$ と表記すると，$N_\Pi\cdot N_{\overline{\Pi}}=N$ が成り立つ．$i=0$ のとき Π_0 は解の全空間 A をただ 1 つの要素として含み，$i=2^L$ のとき Π_{2^L} はそれぞれ 1 個の解から構成される $|A|$ 個の部分空間の集合を表す．

Π_i が生成する部分空間は，GA では競合するスキーマに対応する．たとえば $L=4$ のとき，レベル 2 の分割 '**..' $(=\Pi_3)$ は { **00, **01, **11, **10 } の 4 $(=2^2)$ 個の部分空間を構成する．各部分空間は 4 $(=2^{4-2})$ 個の解を含んでいる．**00 を例にとると，対応する解は $\{0000,0100,1100,1000\}$ の 4 個である．

3 ハッセ図による分割の表現

各分割を 1 つのノードに対応させ，添字 i の 2 進数表現に関して近接するノードをリンクで結ぶことにより，図 1 のようなハッセ図を構成することができる．このとき，すべてのノードは L 個の近傍を持ち，レベル k のノードは，$(L-k)$ 個の上位ノードと k 個の下位ノードを持つ．

Heuristic Statistical Inference for Population-based Search
Akiko N. AIZAWA (akiko@rd.nacsis.ac.jp)
National Center for Science Information Systems
3-29-1 Otsuka, Bunkyoku, Tokyo 113, Japan

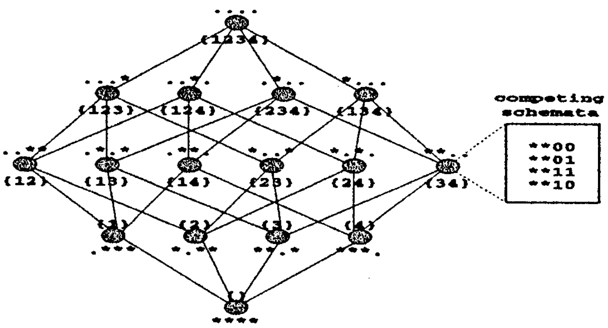

図 1: ハッセ図による分割の表現

4 分割による解空間の扱い

分割 Π_i が，与えられた m 個の解よりなる解集合 X を t 個の部分集合 $\{x_{11},\cdots,x_{m_1 1}\},\cdots\{x_{t1},\cdots,x_{tm_t}\}$ に分割するものとする．Π_i による分割の群内分散 $\sigma_W^2(\Pi_i)$ と群間分散 $\sigma_B^2(\Pi_i)$ は，$\overline{f(x)}=\frac{1}{m}\sum_{ij}f(x_{ij})$ として，次式で与えられる．

$$\sigma_W^2(\Pi_i)\ =\ \frac{\sum_{ij}f(x_{ij})^2}{m}-\frac{1}{m}\sum_i\frac{\left(\sum_j f(x_{ij})\right)^2}{m_i}\quad(1)$$

$$\sigma_B^2(\Pi_i)\ =\ \frac{1}{m}\sum_i\frac{\left(\sum_j f(x_{ij})\right)^2}{m_i}-\overline{f(x)}^2\quad(2)$$

σ_T^2 を $\frac{1}{m}\sum\left(f(x_{ij})-\overline{f(x)}\right)^2$ とすると，定義から，$\sigma_W^2+\sigma_B^2=\sigma_T^2$ はすべての Π_i について等しい．レベル L では，$\sigma_B^2(\Pi_{2^L})=\sigma_T^2$，レベル 0 では，$\sigma_B^2(\Pi_0)=0$，またすべての Π_i について，$0\leq\sigma_B^2(\Pi_i)\leq\sigma_T^2$ が成立する．さらに，リンクで結ばれた 2 個のノードについて $K_{i_1}>K_{i_2}$ のとき必ず，$\sigma_B^2(\Pi_{i_1})>\sigma_B^2(\Pi_{i_2})$ が成立する．

5 Walsh 係数との対応

Walsh 解析とは L 次元の 2 値信号を L 個の独立な基底に分解する方法である [1]．解の全集合が与えられた場合，式 (1) の $\sigma_B^2(\Pi_i)$ と Paley 順序による Walsh の係数との間には次の関係が存在する．まず，Π_i に対応する 2^L 個のノードに対して，添字が i に等しい Walsh 係数を配置する（図 2）．このとき $\sigma_B^2(\Pi_i)$ は，ノード i の下位のノード（ノード i を含みノード 0 を除く $2^{K_i}-1$ 個）に配置された Walsh 係数の和に等しい．ここでノード 0 が特別な扱いとなるのは，$w_0^2=\overline{f(x)}^2$ であり，群間分散ではこの平均値からのずれを計算しているためである．

$$\sigma_B^2(\Pi_{14}) = w_2^2 + w_4^2 + w_6^2 + w_8^2 + w_{10}^2 + w_{12}^2 + w_{14}^2$$

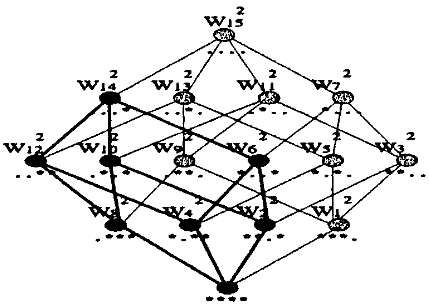

図 2: 群間分散と Walsh 係数の対応

6 分割の評価基準

Walsh 係数の絶対値は，評価値の計算に大きな影響を持つ「有用な」部分解を特徴づけているといえる．そこで本稿では，各ノードにおいて独立に計算できる分割の評価基準として，5. で述べた群間分散と Walsh 係数の関係に注目して以下を定義する．

$$C(\Pi_i) = \frac{\sigma_B^2(\Pi_i)}{K_i} \tag{3}$$

各ビットが独立で，かつ重みが等しい場合，レベル 0 以外のノードはすべて等しい $C(\Pi_i)$ の値を持つことになる（図6）.

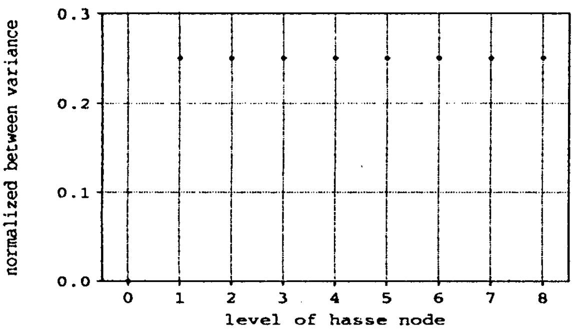

図 3: 各ビットが独立で重みが等しい場合

7 予備実験

文献 [3] で与えられる $NK-landscape$ 関数を用いて予備実験を行った．まず，ストリングの各ビット位置 i に対して，依存関係のある近傍ビット数 n_i を定める．$n_i = 1$ のとき，ビット i は独立である．次に，2^{n_i} 個の重みをランダムに生成して，長さ n_i のサブストリングに対応させる．関数の最終的な値は，L 個のサブストリングに対応する重みの総和とする．

図4, 図5, 図6 に，(1) 各ビットが独立の場合（n_i がすべて 1）．(2) 先頭 4 ビットの間に依存関係が存在する場合（$n_1 = 4, n_2 = n_3 = n_4 = 0, n_5 = \cdots = n_8 = 1$），(3) 全体的に依存関係が存在する場合（$n_i$ がすべて 4），の各々について $C(\Pi_i)$ を求めた結果を示す．

(1) では，レベルが低く重みが大きな分割ほど大きな値を示しビット単位での探索の有望性を示している．一方 (2) では，$C(\ldots.****)$ および近傍のノードの値が大きく，先頭 4 ビットの依存関係を示している．(3) では $C(\ldots\ldots..)$ の値が最も大きく，全解探索が必要であることを示唆している．

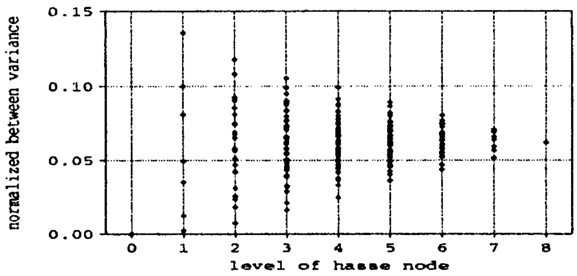

図 4: 各ビットが独立の場合

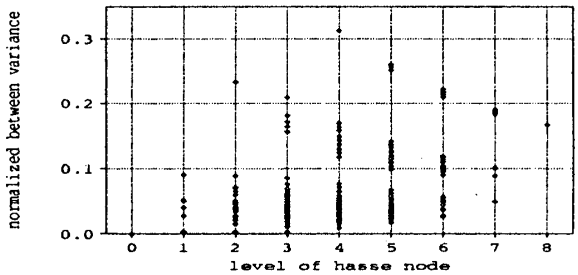

図 5: 先頭 4 ビットの間に依存関係が存在する場合

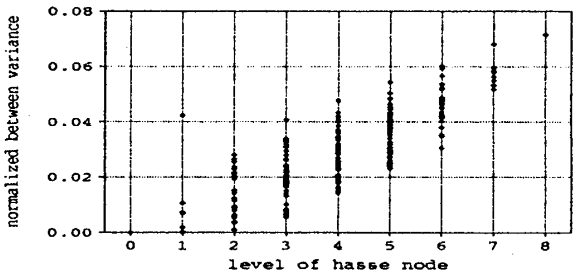

図 6: 全体的に依存関係が存在する場合

8 おわりに

本稿では，依存関係を含む部分解の抽出について，Walsh 係数との関係等を中心に予備的な検討を行った．本稿で定義した分割の評価基準は群間分散に基づいており，解のサンプル集合から推測することが可能である．ハッセ図による表現では，与えられる解集合によらず群間分散の大小関係が定まり，近接するノードの評価値に相関が存在する．これを利用してハッセ図の上を有効に探索する手法の検討が今後の課題である．

参考文献

[1] Goldberg,D.E.:Genetic Algorithms and Walsh Functions: Part I, A Gentle Introduction, *Complex Systems*, **3** (1989).

[2] Davidor, Y.:Epistasis Variance: A Viewpoint on GA-hardness, in *FOGA* (1991).

[3] Manderic,B et.al.: The Genetic Algorithm and the Structure of the Fitness Landscape, in *proc of 4'th ICGA* (1991).