

遺伝的アルゴリズムにおける complexity と収束速度との関係に関する検討*

6N-4

小山毅*1, 津本周作*2, 田中博*2†

*1 大阪大学医学部 *2 東京医科歯科大学難治疾患研究所情報医学研究部門医薬情報†

1 はじめに

遺伝的アルゴリズムにおける重要な問題点として、染色体の表現形式とそれに対する演算子のアルゴリズム形式の設定によって収束速度が大きく変化すること、および突然変異発生の確率等のパラメータの設定による収束速度の変化が挙げられる。これらは、適応される各領域に依存する現象であることが指摘されてきた。今回我々は、この2つの問題に対して、巡回セールスマン問題(TSP)を対象とした種々のアプローチについて、情報理論的な「複雑さ(complexity)」の概念を導入し、各手法についての比較・検討を行った。結果として、ある形式で与えられる複雑さを最小化する手法が最適な収束速度に対応することが示され、この測度の重要性が示唆された。

2 巡回セールスマン問題(TSP)とGA

TSPはGA出現以前からNP hardの問題として知られ[1]、さまざまな近似解法が提唱されてきた[3]。GAがTSPに対して応用されるようになってからも、GAの古典的な枠組みでは、TSPの近似解を得る手法は簡単には得られなかったために、さまざまなGAにおける手法が開発された[2]。TSPの問題解決の手法を考える上での問題点は次の様なものが挙げられる: 1) 問題空間が離散的なベクトル空間であること、2) 解空間上にすべての要素が含まれていなければならないこと、3) 解に順序が必要であり、順序によりfitnessが大きく変わってしまうことなどである。これに対し古典的GAにおいては、1) 染色体はbit列表現で表されること、2) それぞれのbit列に従属関係がないことが暗に仮定されている、3) cross-overあるいはmutationは、bitがそれぞれ独立である故に、いかなる場所においても適応可能である。と問題空間に一定の制限を加えていることが特徴である。しかも、genetic operator自体がこの問題空間の仮定に基づいて構築しているために、問題空間が、解決すべき問題に不適合である場合、問題空間を構成する染色体の表現(いわゆる宣言部)のみならず、それに対する演算子(いわゆる手続き部)までも変更しなければならない。これらはいわゆる山村ら[4]、小林ら[5]において、「コード化の評価規範」として完備性、健全性、非冗長性、及び「交叉の評価規範」として形質遺伝性の4つの規範を用いて論じられており、TSPにおける様々な表現形式(ordinal representation, path representation, PMX等)が、交叉の評価規範を部分的にしか満たさず、結局としてランダムサーチとperformanceに変化がないことが例示されている。

TSPでのコード化における大きな問題点は、上記の2論文に暗に示されているように、染色体そのもののビット列が独立でないことにある。つまり、染色体のある一つのビットの変化が他のビットの保有している情報そのものに影響を与え、解の満たす条件を完全に破壊してしまうことがその問題の本質である。したがって、このようなビットの変化による情報の変化の程度を示す測度を定義することによって、各手法の比較を行うことができることが予想される。

*Comparison of several methods of GA for TSP based on the concepts of complexity

†Tsuyoshi Koyama, Shusaku Tsumoto and Hiroshi Tanaka

*Osaka University, Medical School

†Medical Research Institute, Tokyo Medical and Dental University 1-5-45 Yushima, Bunkyo-ku, Tokyo 113, Japan

3 TSP の complexity

前節で論じたように、TSP における問題の複雑度は主にビット間の独立性によって依存していると考えられる。ここで、TSP の解空間をビット列ではなく、ベクトル空間 $s_k = \{v_1, v_2, \dots, v_n\}$ として表されると仮定し、それぞれの点が link している edge を $E(v_i) = \{e_1, e_2, \dots, e_j\}$ と表すことにする。ここで、ある染色体の population $\{s_i\}$ (一定) においてある解 s_k 上のある点 v_i を交叉によって他の値に変化させる場合に、TSP の解の条件を満たさなくなる確率を $p(v_i)$ と表し、この点における情報量として、 $\log_2(p(v_i) + 1)$ を定義すれば、ある解表現における各点の独立性を示すことができる。つまり、この情報量が 0 の時、その点はある変化に対して独立であることができる。この情報量の各点に関する総和 $\sum \log_2(p(v_i) + 1)$ をもって、解 s_i の複雑度の指標と考える。これによってある解表現の安定性を定量的に観測でき、TSP における初期の random に与えた解の安定性と収束速度とを比較することが可能である。

4 実例による検討

大阪 20 都市の位置を座標化した例を用い、これによって、初期の random に全解を発生させ、最適解 (経路長 212.45km) と complexity を求めておき、これに対して GA (ordinal representation, path representation で一点交叉、二点交叉) によって解を求め、収束世代数、集団内で得られた最適解の割合 (以下最適率) 各集団での平均の complexity を求めた。集団サイズ 50、交叉確率 0.20、突然変異の確率 0.01 においては、ordinal representation の一点交叉で、収束までに平均 1986 世代、得られた最適率は平均で 62% であり、complexity は 0.75 であったのに対し、二点交叉では、それぞれ平均 1550 世代、75%、0.56 であった。path representation の一点交叉で、収束までに平均 3986 世代、得られた最適率は平均で 31% であり、complexity は 0.86 であったのに対し、二点交叉 (PMX) では、それぞれ平均 1350 世代、82%、0.31 であった。この傾向は、集団サイズの大きさによって変化を認めなかった。また、初期の complexity の集団平均と収束世代数については、負の相関 (ordinal representation では、一点交叉で $r=0.62$, 二点交叉で $r=0.52$, path representation では、一点交叉で $r=0.74$, 二点交叉で $r=0.36$) が得られた。従って、より適切な交叉においては、complexity に対する収束世代数の変化は小さくなっていく傾向にあると言える。

5 おわりに

我々の定義した complexity によって、ある解の収束しやすさの指標を与えることができた。しかしながら、解を安定させることのできる表現は局所最適解に陥る危険が多いことも考えられる。今回検証した例が小規模であったため、この効果は無視できたが、実際には都市数が増えた場合は、この因子を含めた complexity の定義が必要であると考えられる。TSP 自体の複雑さには我々が想定した因子以外の関与も考えられるので、それらについて究明、より最適な complexity の定義を探求していく予定である。

参考文献

- [1] Galey, M. and Johnson, D. *Computers and Intractability*, W.H. Freeman, San Francisco, 1979.
- [2] Michalewicz, Z. *Genetic Algorithm + Data Structure = Evolutionary Program*, Springer Verlag, 1992.
- [3] Paterson, M.S. (ed) *Proc. of the 17th Colloquium on Automata, Languages, and Programming*, Springer-Verlag, Lecture Notes in Computer Science, Vol. 443, 1990.
- [4] 山村雅幸, 小野貴久, 小林重信. 形質の遺伝を重視した遺伝的アルゴリズムに基づく巡回セールスマン問題の解法, 人工知能学会誌 7, 1049-1059, 1992. Vol. 443, 1990.
- [5] 小林重信. 遺伝的アルゴリズムの現状と課題, 計測と制御 32, 1993.