

遺伝的アルゴリズムに基づく組み合わせ最適化手法による 制約のある割り当て問題への応用

有 澤 正 樹*

概 要

本学の教養課程で開講されている大学基礎演習は新入生必修の科目であり、全38クラスからなる全8学科の学生の混成クラスとして運営されているため、全学生を各クラスに均等に割り当てる作業が組み合わせ最適化問題となっている。本論文では、この問題に遺伝的アルゴリズムを適用するため、複数の染色体からなる個体、個体内の染色体間による交叉、劣等染色体による交叉、遺伝子座に順序性のない染色体に考慮した円弧交叉などの特徴を有した、制約のある割り当て問題適応型遺伝的アルゴリズムを提案し、2018年度実績データに基づく解の導出によりその有効性を示している。

1 はじめに

近年、各大学では18歳人口の減少に対応すべく入試業務やカリキュラム編成業務など、研究・教育以外の業務が急増している。本学においても教養教育の充実を図るカリキュラム改定において、新入生必修の大学基礎演習を開講するなど、その動きは激しい。このような中、この大学基礎演習のクラス割り当て作業が組み合わせ最適化問題となっており、担当教員の大きな負担となっている。本論文では、この問題に対応すべく、制約のある割り当て問題適応型遺伝的アルゴリズムを提案し、これによりクラス割り当て案が迅速に求められることを示す。

2 制約のある割り当て問題 (CAP)

本学の教養課程で開講されている大学基礎演習は、全学部学生1年次必修の重要な科目として位置づけられている。そのため、担当教員は専任教員に限定し、さらに異学部異学科間の学生交流を狙うなど、より特色のある科目とするために全学科混成の指定クラスとなっている。その結果、全38クラスを同日同時限に一斉開講し、そこに新入生および再履修者合わせて681名(2018年度)の学生を全8学科混成で38クラスに均等に割り当てる必要が生じた。その他の条件を併せると、次のようにまとめられる。

1. 全クラス全学科混成とする
2. 全クラスの学生数を可能な範囲で均等にする
3. クラス内で同学科1年次を2名以上とする

*茨城キリスト教大学生生活科学部

4. 再履修者は当人の所属学科教員担当のクラスに配置し、同学科1年次2名+再履修者の3名とする
5. 3～4の条件を優先し、一部の学科の割り当てがないクラスを許す

ここではこれを制約のある割り当て問題 (CAP; Constrained Allocation Problem) と呼ぶ。

従来、この学生のクラス割り当て作業は、新入生数が確定する3月末日頃からの数日間に担当教員による試行錯誤により行われており、入学辞退や復学などによる急な増減への逐次対応も要する非常に負担の大きい業務となっている。

組み合わせ最適化問題 (COP; Combinatorial Optimization Problem) に関しては、OR (Operations Research) 分野を代表とする様々な手法が開発されているが、ここでは比較的容易に実装が可能で汎用性のある遺伝的アルゴリズムを採用する。

3 遺伝的アルゴリズム (GA)

遺伝的アルゴリズム (GA; Genetic Algorithms) は、生物の進化の仕組みをモデル化したアルゴリズムであり、確率的探索法の1つとしてCOPなどに広く応用されている。COPは解の候補となる組み合わせの中から最も評価値の高い組み合わせを求める問題であるが、GAにおいてはこの解の候補となる組み合わせを表現型と呼ぶ。また、染色体を構成する遺伝子の集まりを遺伝子型と呼び、各遺伝子の取り得る遺伝子候補を対立遺伝子と呼ぶ。そして、表現型から遺伝子型への写像をコーディングと呼ぶ。また、染色体により特徴づけられた個体の集まりを個体群と呼んでいる。

GAの一般的な手順は

1. 無作為による初期個体群の生成
2. 各個体の適合度評価
3. 2に基づく個体の存続・死滅・交配の選択、および交配による染色体の交叉 (子の生成) を伴う次世代個体群の生成
4. 低確率での突然変異
5. 2に戻り、目標とする適合度に到達するまで繰り返す

となる。最終的に最も適合度の高い個体の遺伝子型をデコードして得た表現型が準最適解となる。GAは全探索ではなく部分探索であるため、最適解を保証するものではない [1, 2]。

4 CAPにおける問題の最適化

GAにおいて、解の表現型をどのような形で染色体内に遺伝子型としてコーディングするかは重要な問題である。本CAPにおいて最も単純なコーディングは履修者総数にあたる681の遺伝子をもつ染色体を規定し、遺伝子座にクラス番号を置くことである。しかしこれでは 38^{681} 通りもの、もはや計算不能ともいえる探索空間を設定することになり、現実的ではない。このため、ここでは次のように問題を整理する。

学科視点でこの問題を捉え、2節に挙げた条件3～5に照らすとき、各学科定員の近傍となる総履修者数と総クラス数に鑑みると、1クラスあたりに割り当てられる同学科の学生数は3名、2名、もしくは0名のいずれかとなる。このように考えるとCAPは次に示す問題へと帰着する。

「学科数 $8 \times$ クラス数 38 の行列の要素に $3, 2, 0$ のいずれかの数値を設定し、学科ごとの合計値を各履修者総数に、クラスごとの合計値を可能な範囲で均等にする。」

5 CAP適応型GA

5.1 コーディング

4節に示したCAPの最適化により、対立遺伝子として $3, 2, 0$ の3値を取る $8 \times 38 = 304$ の遺伝子からなる染色体を有した個体を規定する。

このときの解候補の組み合わせ総数 P_{all} は

$$P_{all} = (3^{38})^8 \simeq 1.11 \times 10^{145} \quad (1)$$

で求められるが、これでもまだ広大であることに変わりない。

そこで304の遺伝子座に遺伝子の初期値を割り当てる初期個体生成にあたり、予め数理的処理により学科ごとに必要な対立遺伝子3値の数を求めておき、その必要数に応じて学科ごとに無作為に配置することを考える。すなわちこれは、学科ごとの履修者総数と割り当て条件を満たし、クラスごとの履修者数にばらつきのある個体の生成を意味している。2018年度データを基に算出した対立遺伝子の必要数を表1に示す。

このとき、全8学科を識別子 $i (i = 1, 2, \dots, 8)$ で表し、各学科の対立遺伝子 $3, 2, 0$ の割り当て数をそれぞれ k_i, l_i, m_i と置くと、総クラス数 38 の各学科のクラス割り当ての組み合わせ総数 p_i は、重複度 (k_i, l_i, m_i) の 38 -重複置換の総数として次の多項係数で与えられる。

表1 対立遺伝子 $3, 2, 0$ の必要数

学科	1年次数	再履修者数	合計	3名クラス数	2名クラス数	0名クラス数
1	95	0	95	19	19	0
2	67	0	67	1	32	5
3	85	4	89	13	25	0
4	90	1	91	15	23	0
5	88	1	89	13	25	0
6	75	3	78	4	33	1
7	99	0	99	23	15	0
8	72	1	73	1	35	2
合計	671	10	681			

$$p_i = \left(\frac{38}{k_i, l_i, m_i} \right) = \frac{38!}{k_i! \cdot l_i! \cdot m_i!} \quad (2)$$

よって、この個体による解候補の組み合わせ総数 P_{part} は p_i の総乗として次式で求められる。

$$P_{part} = \prod_{i=1}^8 p_i \quad (3)$$

式 (2) (3) により表 1 のデータを基に算出すると

$$P_{part} \simeq 2.61 \times 10^{68} \quad (4)$$

となり、これにより探索空間を P_{all} に比して約 10^{77} 削減している。

5.2 複数染色体型個体内GA

5.1節で述べたコーディングと初期個体の生成方法により、探索空間の大幅な削減を行うことができた。しかし、この初期個体の生成法に伴う制約により、各学科内での対立遺伝子ごとの総数を保存しなければならないため、いわゆる単純 GA は採用できない。そのためここでは複数の染色体からなる個体を採用する。

一般的な GA では解の遺伝子型である個体は 1 つの染色体で構成することが多く、交配 (交叉) は個体集団の中から選ばれた個体同士で行われる。これに対しここでは、1 つのクラスへの 8 学科の割り当てを 1 つの遺伝子型、すなわち遺伝子数 8 の染色体として捉え、38 クラス分の 38 染色体からなる個体を全体の遺伝子型とする。

一般的にこのような複数染色体型の個体を採用する GA においては、最適化する変数の個数に応じた染色体数を設定し、異個体間での交配時には同種の染色体同士だけで交叉させる並列処理が行われるが [3, 4]、先に述べた制約により CAP においてはこれも採用することはできない。

そこでここでは、個体集団を形成せず、制約を満たすために 1 個体内の染色体同士で交叉を行うことにより世代交代を行う方法を採用する。これにより、本 GA で行われる遺伝子交換は 1 個体内での交叉のみとなるため、交配という概念は失われる。

5.3 適合度と選択

一般的に個体の環境への適合度評価は、より適合度の高い個体が高く評価されるように設計された関数などを用いて行われ、適合度の高い個体ほど高確率で選択、交配 (交叉) され次世代個体集団を構成していく。ここで CAP に適用する複数染色体型個体内 GA においては、各染色体の遺伝子値の合計が各クラスの履修者数を表現していることから、この合計値により適合度評価を行うことが妥当であろう。ところが CAP におけるこの合計値の最適値は $\frac{681}{38} \simeq 17.92$ から分かる通り、割り切れる場合を除いては 17 または 18 のように幅がある。また、交叉についても特殊で、全染色体の同一遺伝子座の遺伝子値の合計が学科ごとの総履修者数を表しているため、交叉する 2 つの親染色体が死滅し、交叉によ

て生み出された2つの子染色体に完全な世代交代が行われることにより、学科ごとの総履修者数を保存する必要がある。そのため、適合度の低い染色体の交叉から高い染色体を生み出すことが必要とされるのである。以上のことから、本GAにおいては次のような方法を採用する。

適合度評価 各染色体の遺伝子値の合計。各クラスの履修者数を表しており、単調増加関数などではないため、その大小が適合度そのものを表すものではない。

交叉の選択 適合度評価値が最大と最小の任意の染色体。履修者数最大クラスと最少クラスに該当する。交叉により生み出された2つの子染色体と交叉に選ばれなかった36の染色体が次世代を構成する。

交叉の終了 適合度評価値の最大と最小の差が1以下になったとき。全クラスの履修者数が均等になったときを意味する。この値は均等な履修者数の理論値が整数のときは0、実数のときは1を取るが、互いに排他的であるため、1以下の条件で双方を満たす。

5.4 交叉

交叉対象として選択された2つの親染色体は、交叉によって次世代の子染色体2体に置き換わる。ここでは交叉方法として円弧交叉を提案する。円弧交叉とは、遺伝子座に順序性がない染色体に考慮するために染色体をリング状に捉え、そのリングの一部を交換する方法である。図1に円弧交叉の概念図を示す。ただし遺伝子座につけられた番号はラベルであり、その順序性に意味をもたない。図1のリングを遺伝子座1-8または3-4の境界で切り離し、従来の帯状染色体として捉えれば1点交叉であるが、それ以外の境界で切り離したものであれば2点交叉と同じである。

ところで、2点交叉を含む多点交叉は、その解説論文などを素直に解釈すると、“異なる n 点で $n+1$ 個に分離された染色体の一部を交換すること”と解されるだろう[1, 2, 5]。これを遺伝子数8の帯状染色体に対する2点交叉に適用すると、図2に示すように、帯状染色体の両端である1-8間は切り離す2点のどちらにも選ばれないことになり、リング状に捉えたときとの違いは明らかであろう。

円弧交叉は、帯状染色体における1-8間の選択を許す2点交叉と見なすこともできるが、乱数による2点の選択方法に関する実装の揺らぎにより差異が生じる。図2を例に取

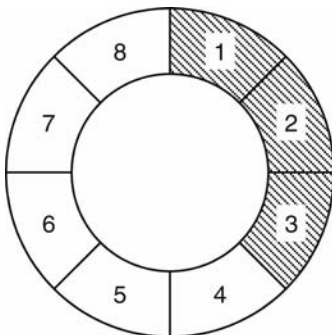


図1 円弧交叉



図2 2点交叉

れば、恐らく多くの実装では始点の4と終点の6を乱数により決定すると思われるが、“始点>終点”となる乱数を得た場合に、これを破棄し乱数を再生成するなどの処理を採れば、選択確率において円弧交叉との間に差異が生じるからである。

本論文での円弧交叉の実装は、無作為に1~8の始点と1~7の終点までの長さ（時計回り）を生成し、図1に示すリングの一部を交換している。なお、円弧交叉の性能評価については7節で述べる。

5.5 突然変異

5.1節で示したコーディングにより、学科ごとの総履修者数を表している38染色体の同一遺伝子座の遺伝子値の合計を保存しなければならないため、一般的な他の対立遺伝子への置き換えや転座などの方法は採用できない。そこでここでは、交叉対象となる染色体の選択に突然変異の要素を取り入れる。

5.3節で述べたように、CAP適応型GAでの交叉は相対的に適合度評価値の離れている2つの染色体をもって行われるが、十分に低い確率で1つの染色体を適合度評価値に関わらず無作為に選択することとした。これにより通常は交叉することなく保存されるはずの親染色体が死滅し、新しい可能性を秘めた子染色体へと変化を遂げる。しかし、詳しくは7節で述べるが、今回適用するCAPにおいては、突然変異を設定せずとも十分に最適値が求まることが分かっている。従って、実装においては100世代以降50世代ごとの一定間隔で突然変異を設定している。

6 CAPへの適用

6.1 初期個体の生成

今回の実装でパラメータとして入力が必要とするデータは、クラス数、学科数、学科ごとの1年次

表2 初期個体の遺伝子配列の例

C [†]	i=1	2	3	4	5	6	7	8	E [‡]
1	2	2	3	2	3	2	3	2	19
2	3	2	2	2	2	2	3	2	18
3	2	2	2	3	2	2	2	2	17
4	3	2	3	3	3	2	3	2	21
5	3	2	3	2	2	2	2	0	16
6	3	0	3	3	2	2	3	2	18
7	3	2	2	3	2	2	2	2	18
8	3	0	2	2	2	2	3	2	16
9	2	2	2	3	3	2	2	2	18
10	2	2	3	2	2	3	2	2	18
11	3	2	3	2	2	2	2	2	18
12	3	2	2	2	2	2	3	2	18
13	2	2	2	3	2	3	3	2	19
14	3	2	2	2	2	2	3	2	18
15	3	2	2	3	3	2	3	2	20
16	3	2	2	3	2	2	3	2	19
17	2	2	2	2	2	0	2	2	14
18	2	3	3	2	2	3	2	2	19
19	3	2	2	3	2	2	3	2	19
20	2	2	2	2	3	2	3	2	18
21	2	2	2	3	3	2	2	2	18
22	3	0	2	2	2	2	3	2	16
23	2	0	2	2	2	2	3	2	15
24	2	2	2	2	2	2	3	2	17
25	3	2	3	3	2	2	3	2	20
26	3	2	2	2	3	2	3	2	19
27	2	2	3	3	2	2	2	2	18
28	2	2	2	2	3	3	3	2	19
29	2	2	3	3	2	2	3	0	17
30	3	2	3	2	2	2	2	2	18
31	3	2	2	3	3	2	3	2	20
32	2	2	2	2	2	2	2	2	16
33	3	0	2	2	3	2	2	2	16
34	2	2	3	2	3	2	2	2	18
35	2	2	2	2	2	2	3	3	18
36	3	2	2	2	3	2	3	2	19
37	2	2	2	2	3	2	3	2	18
38	2	2	3	3	2	2	2	2	18
計	95	67	89	91	89	78	99	73	681

[†]染色体番号(クラス番号), [‡]適合度評価値(クラス人数)

数および再履修者数である。表 2 に無作為に生成された初期個体の遺伝子配列の一例を示す。5.1 節で述べたように最下行の合計値は各学科の総履修者数に一致し、最右列のクラス人数にばらつきが認められる。これを見ると最適条件を満たす初期個体の生成に期待がかかるかも知れないが、式 (4) から明らかなようにその可能性は極めて低い。実際に 100 万個 (1.00×10^6) の初期個体生成を行ったところ、そこに条件を満たすものはなかった。

6.2 選択と交叉

最適解が求まるまでに各世代で交叉した染色体とその適合度評価値の変化の様子を図 3 に示す。この例では初期個体を第 1 世代として 51 回の交叉後の第 52 世代ですべての条件を満たした個体が生み出されている。適合度評価値の変化を見ると、初期個体時には最大値 21、最小値 14 だったものが、第 52 世代では目標値である最大値 18、最小値 17 に変動を繰り返しながらも緩やかに収束していることが分かる。次に選択され交叉した染色体の動向を見ると、連続して同じ染色体が選択される期間がある一方で、あたかも無作為に選択されているかのように見える箇所も認められるなど様々である。また、5.5 節で述べたように、この例では収束までに 100 世代を要していないので、突然変異は起こっていない。

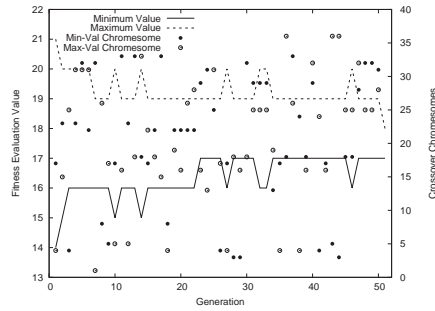


図 3 交叉染色体と適合度評価値の変化

表 3 最適化個体の遺伝子配列

C^{\dagger}	$i=1$	2	3	4	5	6	7	8	E^{\ddagger}
1	2	2	3	2	2	2	3	2	18
2	3	2	2	2	2	2	3	2	18
3	2	2	2	3	3	2	2	2	18
4	2	2	2	3	2	2	3	2	18
5	3	0	2	2	3	2	3	2	17
6	3	0	3	3	2	2	3	2	18
7	3	2	2	3	2	2	2	2	18
8	3	0	3	3	2	2	3	2	18
9	2	2	2	3	3	2	2	2	18
10	2	2	3	2	2	3	2	2	18
11	3	2	3	2	2	2	2	2	18
12	3	2	2	2	2	2	3	2	18
13	2	2	2	3	2	2	3	2	18
14	3	2	2	2	2	2	3	2	18
15	3	2	2	3	2	2	2	2	18
16	3	2	2	2	2	2	3	2	18
17	2	2	3	2	3	2	2	2	18
18	2	2	3	2	2	2	2	2	17
19	3	0	2	3	2	3	3	2	18
20	2	2	2	2	3	2	3	2	18
21	2	2	2	3	3	2	2	2	18
22	3	2	2	2	3	3	3	0	18
23	2	2	2	2	3	2	3	2	18
24	3	2	2	2	2	2	3	2	18
25	3	2	2	2	2	2	3	2	18
26	3	2	2	2	2	2	3	2	18
27	2	2	3	3	2	2	2	2	18
28	2	3	3	2	3	0	3	2	18
29	2	2	3	3	2	2	3	0	17
30	3	2	3	2	2	2	2	2	18
31	3	2	2	3	2	2	2	2	18
32	2	2	2	3	3	2	2	2	18
33	3	0	2	2	3	3	3	2	18
34	2	2	3	2	3	2	2	2	18
35	2	2	2	2	2	2	3	3	18
36	3	2	2	2	2	2	3	2	18
37	2	2	2	2	3	2	3	2	18
38	2	2	3	3	2	2	2	2	18
計	95	67	89	91	89	78	99	73	681

\dagger 染色体番号 (クラス番号), \ddagger 適合度評価値 (クラス人数)

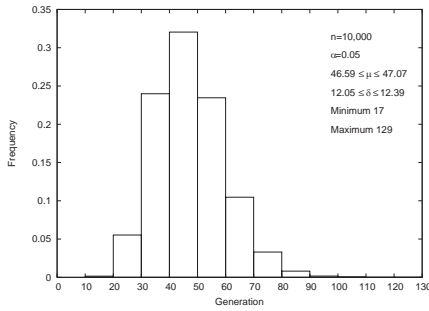


図4 最適化に要した世代数の分布

表4 各交叉方法による最適化に要した世代数の比較

基本統計量	1点交叉	2点交叉	円弧交叉
最小	16	18	17
最大	132	222	129
μ 下方信頼限界	54.48	54.10	46.59
μ 上方信頼限界	55.09	54.73	47.07
σ 下方信頼限界	15.57	16.03	12.05
σ 上方信頼限界	16.01	16.48	12.39

$n=10,000$, $\alpha=0.05$

6.3 クラス割り当ての決定 (デコード)

最適化された個体の遺伝子配列を表3に示す. 表中最下行の各学科の総履修者数は初期値のまま保存されており, 最右列のクラス人数が17または18に均等化されていることが分かる. また, 表中の太文字で表された“**3**”は, 再履修者1名と新入生2名, 合わせて3名を割り当てる箇所を示しており, 2節で示したCAPにおける条件4を満たしている一例である. 再履修者については学生所属学科の担当教員によるサポートの必要性があるため, 再履修者を割り当てるクラスはすべて別クラスとなるように選択している. この作業はヒューリスティックなものとなるが, 十分に自由度が高いため容易な作業である. これが終われば, まず再履修者に指定のクラスを割り当て, あとは学科別履修者名簿の先頭から表3に沿った数ずつクラス番号を割り当てればよい.

7 性能評価

最適化に要した世代数の分布を図4に示す. 図からも明らかなように, 非常に若い世代で最適化が完了していることが分かる. また, その他の交叉方法との比較を含め, 表4に基本統計量を示す. 表からも明らかなように, 円弧交叉においては母平均および母標準偏差ともに他に比べ小さく, その優位性が認められる. また, 母平均および母標準偏差の推定値から, 概ね100世代もあれば最適解に至ると判断できるため, 5.5節で述べたように, 突然変異の発生を100世代以降に設定している.

なお, 最適値への収束途中の若い世代に高頻度で突然変異を発生 (突然変異率0.01 ~ 0.1相当) させるシミュレーションにおいては, 収束に要する世代数が多少大きくなる傾向が認められたものの, 突然変異の頻度や有無に関わらず未収束の例は見られなかった.

8 おわりに

本論文ではCAPにGAを適用するために, 複数の染色体からなる個体, 個体内の染色体間による交叉, 劣等染色体による交叉, 遺伝子座に順序性がない染色体に考慮した円弧交叉などの特徴を有したCAP適応型GAを提案し, 実データによるシミュレーションによりその有効性を示した.

大学内には時間割編成に見られるような大規模な問題から, 本論文で取り上げた比較的小規模な学生の割り当て問題まで, 数多くのCOPが存在している. そしてその多くが独

自の複雑な制約条件をもち、かつ頻繁に制約条件が変更されるなど、標準化が難しく大規模なシステム化が困難なものがほとんどであろう。

今回、筆者の必要に迫られ、CAPへのGAの適用を検討するに至ったが、GAは比較的簡単なアルゴリズムであるためソフトウェアでの実装は比較的容易なものの、表現型から遺伝子型へのコーディングの良し悪しが結果を大きく左右し、自由度も高く調整が難しい一面も有している。そのため、問題に対する個別の設計が必要であり、決して万能な手段とはいえない。従って、本質的には問題の対象そのものの設計段階における複雑化の回避が重要であろう。

参考文献

- [1] 北野宏明, “遺伝的アルゴリズム,” 人工知能学会誌, Vol.7, No.1, pp.26-37, 1992.
- [2] 坂和正敏, 田中雅博, “日本ファジィ学会編ソフトコンピューティングシリーズ遺伝的アルゴリズム,” 朝倉書店, 1995.
- [3] 河村拓昌, 大森博司, “遺伝的アルゴリズムによる立体トラス構造物の形態創生,” 日本建築学会構造系論文集, No.538, pp.115-121, 2000.
- [4] 篠田秀男, 加藤友彦, “遺伝的アルゴリズムを用いた大学の時間割編成の最適化,” 福岡工業大学研究論集, Vol.42, No.1, pp.19-22, 2009.
- [5] A.J. Umbarkar, P.D. Sheth, “Crossover Operators in Genetic Algorithms: A Review,” *ICTACT Journal on Soft Computing*, Vol.6, Issue 1, pp.1083-1092, 2015.

Creating a Combinatorial Optimization Method Based on Genetic Algorithms to Solve Constrained Allocation Problems

Masaki Arisawa

As the University Basic Seminar offered among the general education courses is required for all first year students and mixes the students from eight departments into 38 separate classes, allocating students equally is a combinatorial optimization problem. In order to apply genetic algorithms to this problem, I created a genetic algorithm adapting to constrained allocation problems, featuring individuals consisting of multiple chromosomes, crossing over between chromosomes within an individual, crossing over by inferior chromosomes and crossing over considering chromosomes without locus sequencing. A solution based on the actual data of the 2018 academic year demonstrates the effectiveness of the genetic algorithm.