

識別パターンを利用した遺伝的アルゴリズムの 部分解保護に向けて

東京工業大学 亀谷由隆

研究の背景

- 遺伝的アルゴリズム
 - 積み木仮説：部分解が生成され，またそれらの部分解が適切に組み合わされる
 - しかし生成された部分解が早い段階で交叉により破壊
- リンケージ学習 (linkage learning)
 - リンケージ適応手法，摂動に基づく手法，モデル構築手法
- GAP (GA with Patterns)
 - Gero & Kazakov (1995, 2001) による遺伝子工学アプローチの発展形
 - 高適合な染色体から部分列を重要なパターンとして抽出し，パターンの出現箇所を交叉から保護する
 - ビット表現だけではなく，順列表現も統一的に扱える
 - 各交叉における親染色体に依存した細粒度（遺伝子単位）の部分解保護
 - 適合度関数の評価を追加的に行う必要がない

パターンと部分解

- 染色体の翻訳

- 順列表現：そのまま

- 【例】染色体 $\langle a, c, b \rangle \rightarrow$ 系列データ $\langle a, c, b \rangle$

- ビット表現：座位-遺伝子のペアからなる集合データに翻訳

- 【例】染色体 $\langle 1, 0, 1, 1, 0 \rangle$

- \rightarrow 集合データ $\{(1, 1), (2, 0), (3, 1), (4, 1), (5, 0)\}$

- 部分系列（ギャップを含む）

- $\langle a, c, b \rangle$ の部分系列は

- $\langle a \rangle, \langle b \rangle, \langle c \rangle, \langle a, b \rangle, \langle a, c \rangle, \langle c, b \rangle, \langle a, c, b \rangle$

- 集合データ $\{(1, 1), (2, 0), (3, 1), (4, 1), (5, 0)\}$ の部分集合は例えば $\{(2, 0), (5, 0)\}$

正規表現風には書けば `".*a.*b.*"`

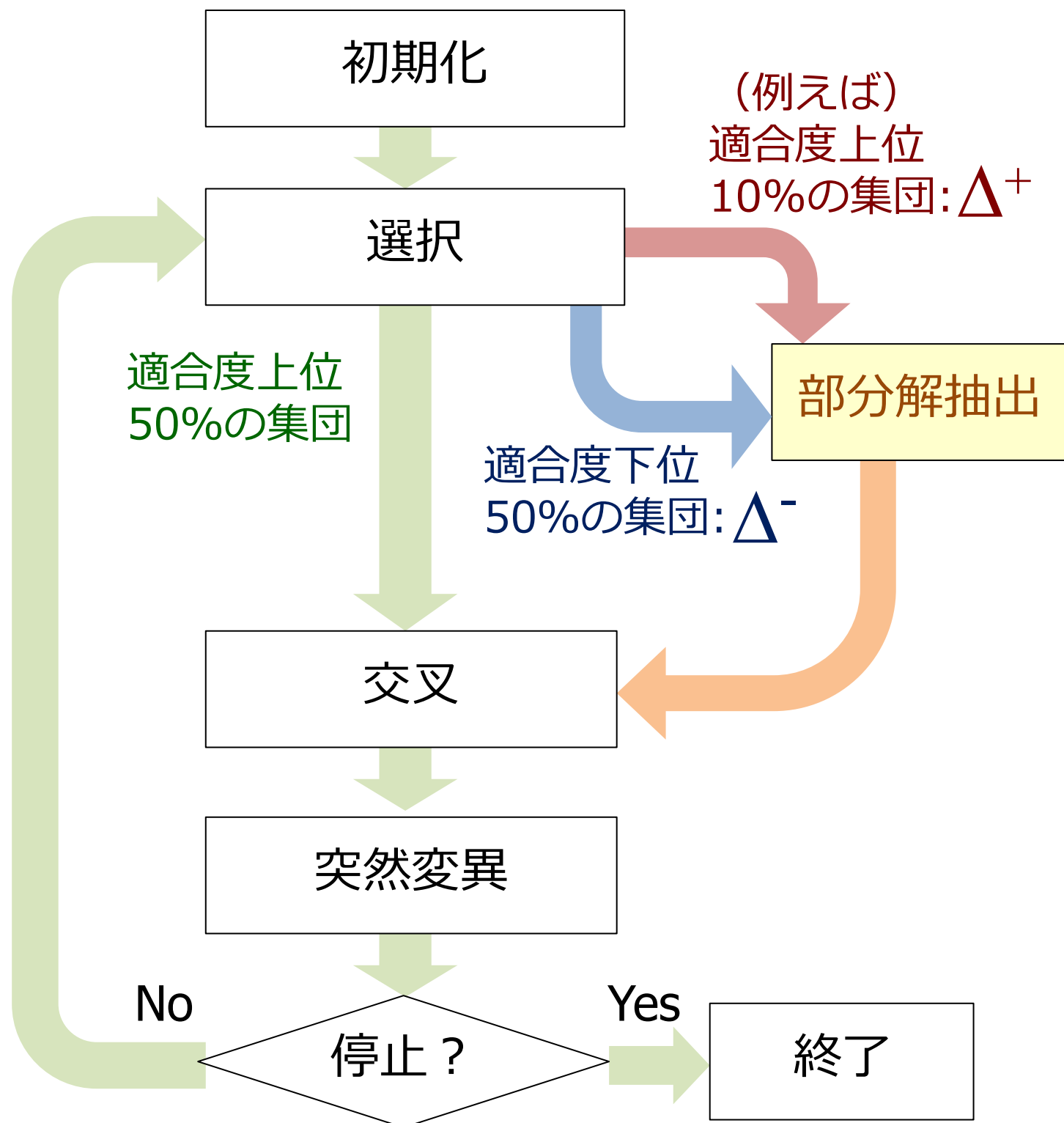
スキーマ表現 `#0##0` に対応

- パターン発見アルゴリズム：

- 順列表現：BIDE [Wang et al. 04]

- ビット表現：LCM [Uno et al. 04]

識別パターンに基づくGAP



- 識別パターン： Δ^+ と Δ^- の違いを際立たせるパターン

- 指標：F 値

$$F(s) = \frac{2 \cdot \text{Recall}(s)\text{Precision}(s)}{\text{Recall}(s) + \text{Precision}(s)}$$

$$\text{Recall}(s) = p(s | \Delta^+) = \frac{\sigma(s, \Delta^+)}{|\Delta^+|}$$

$$\text{Precision}(s) = p(\Delta^+ | s) = \frac{\sigma(s, \Delta^+)}{\sigma(s, \Delta^+) + \sigma(s, \Delta^-)}$$

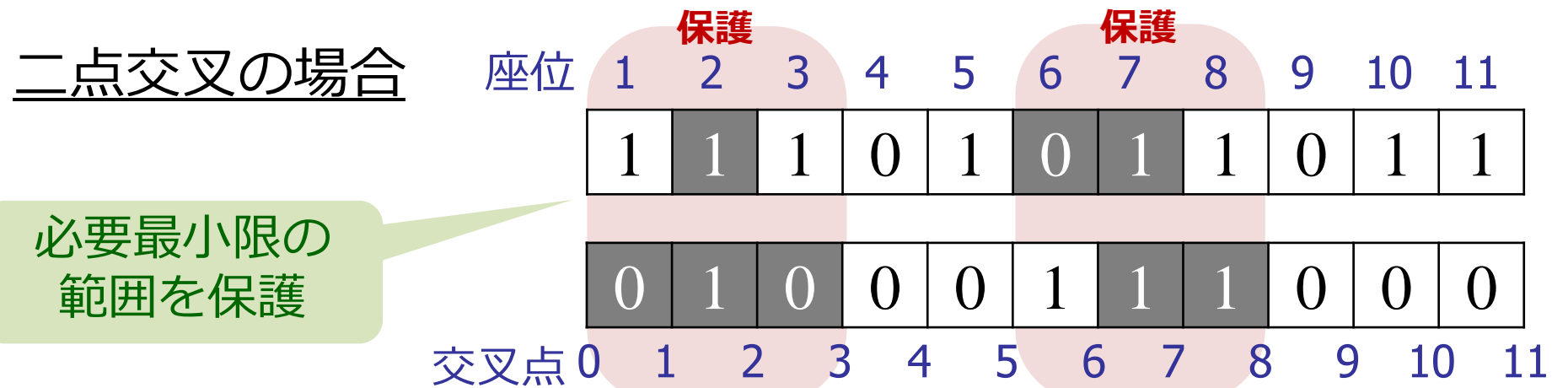
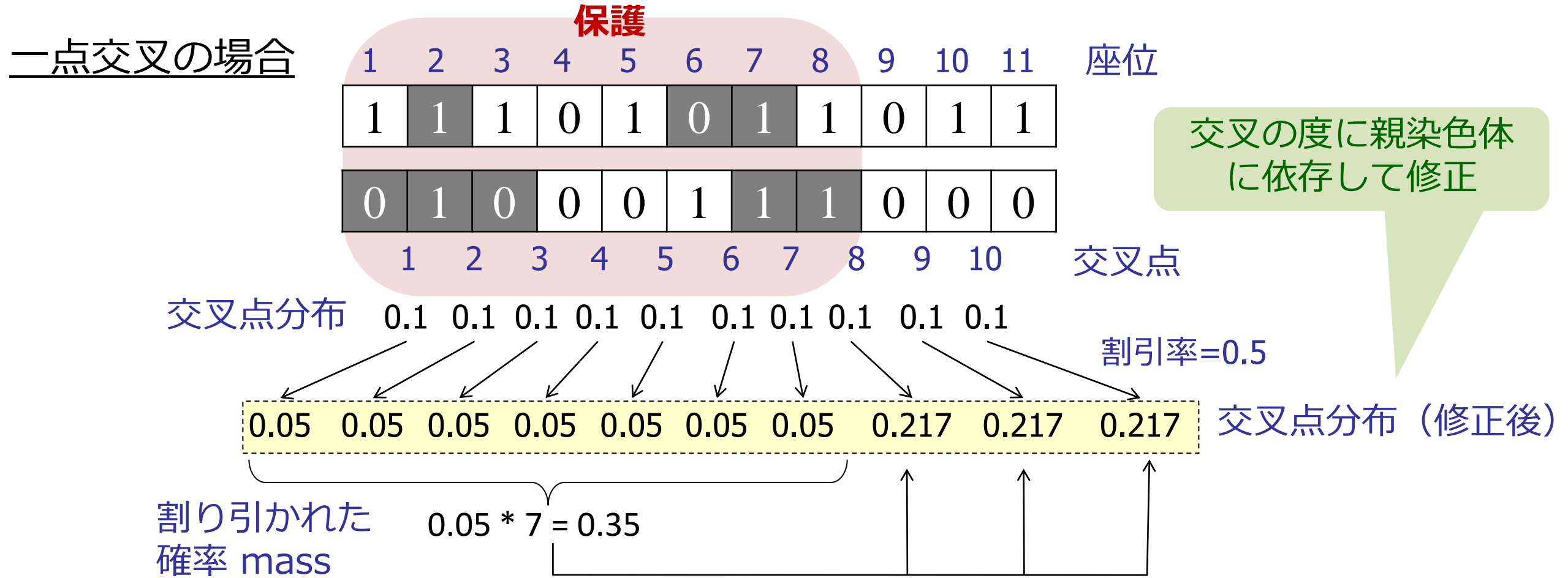
- F 値はパターンの包含関係に関し、反単調性を満たさない
→ 分岐限定法を利用

- 「弱い」識別パターンは除去

「 s' は s より弱い」 \Leftrightarrow
 $s \subset s'$ にも関わらず $F(s) \geq F(s')$

部分解保護

- 交叉点分布の修正に基づき部分解を保護



Royal Road 問題

- 停止までに要した世代数（試行100回の平均）を比較
- 比較相手に BOA [Pelikan et al. 2000] も含める

“tight” な符号化

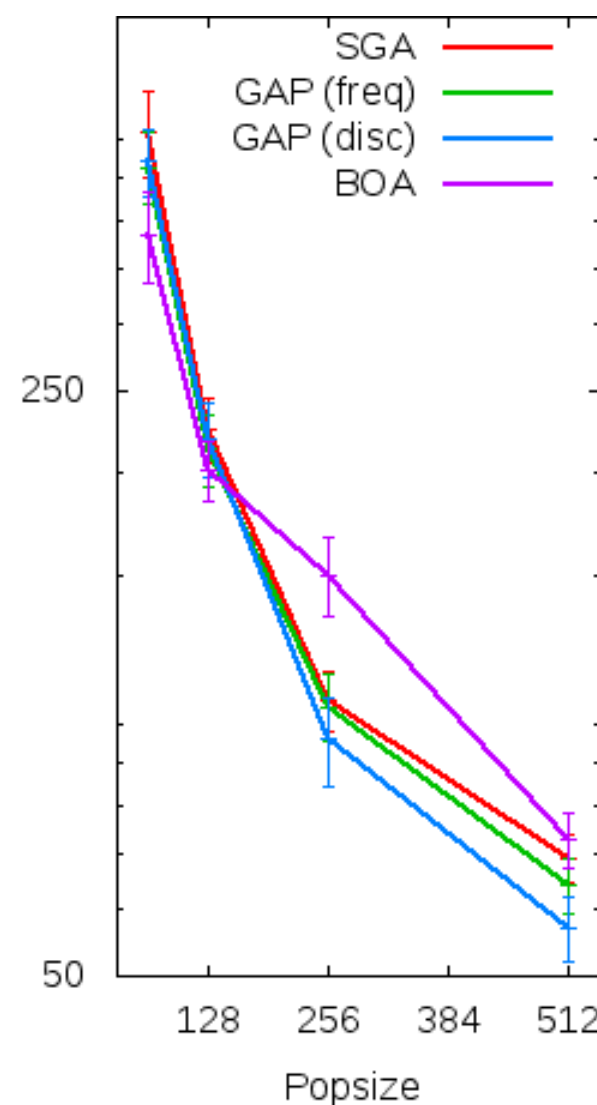
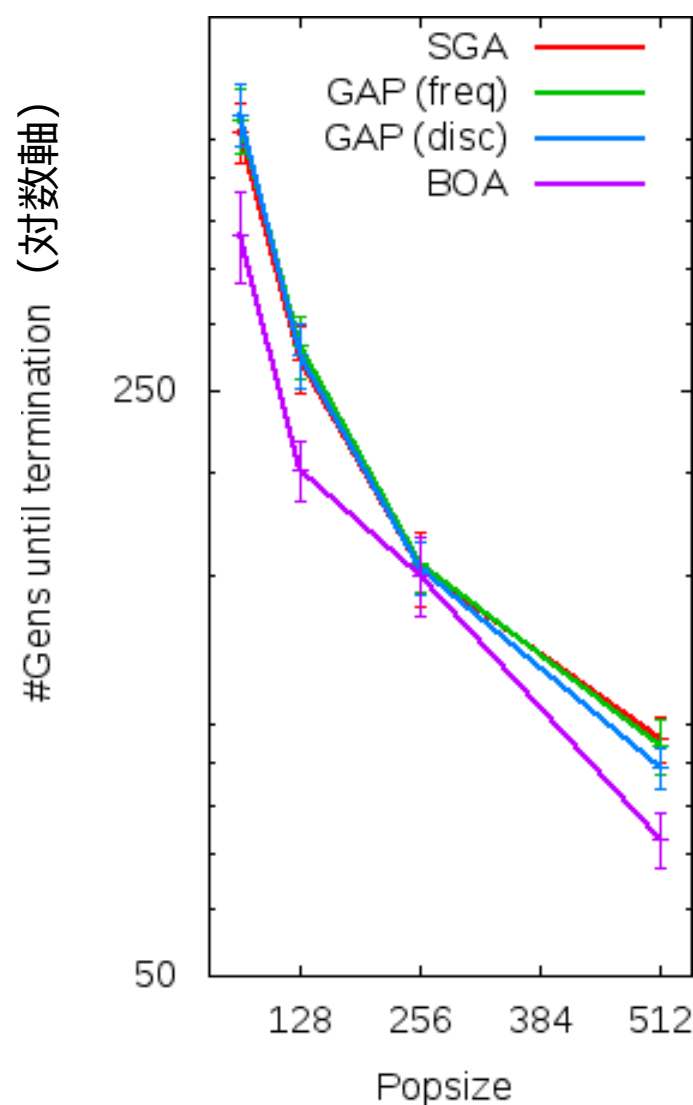
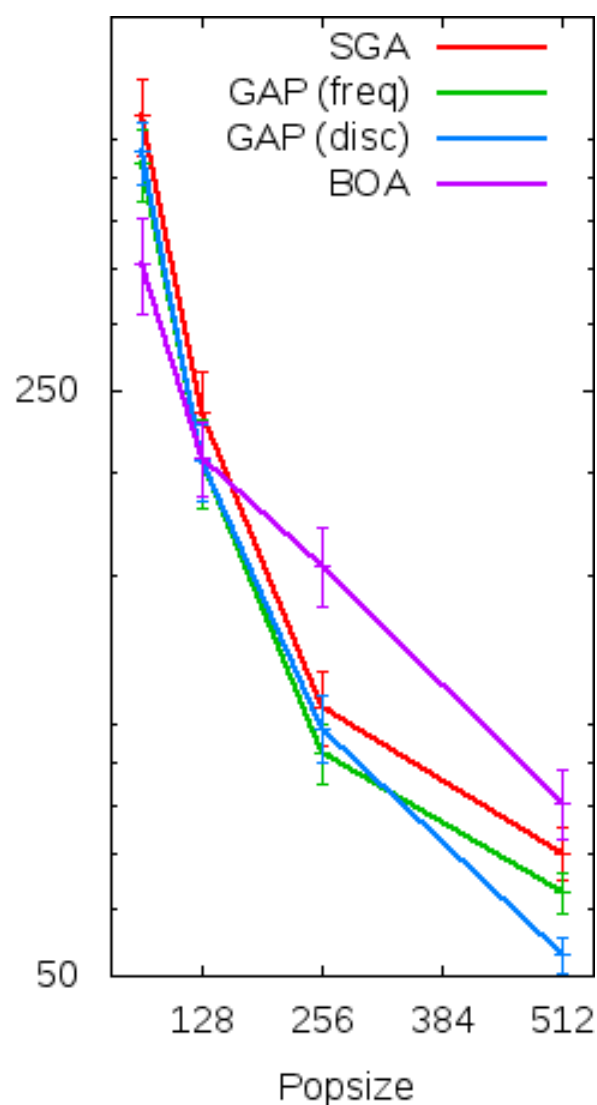
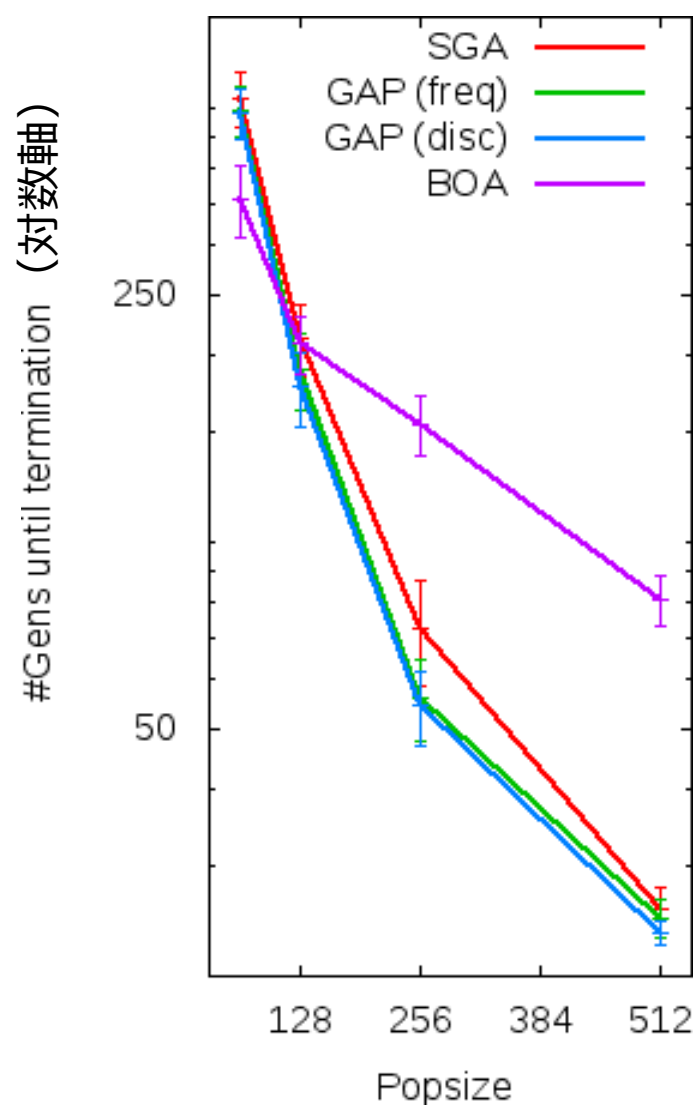
“loose” な符号化

二点交叉

一様交叉

二点交叉

一様交叉



巡回セールスマン問題 (Grötschels48)

最良個体の適合度
(試行100回の平均, 最適解 5046)

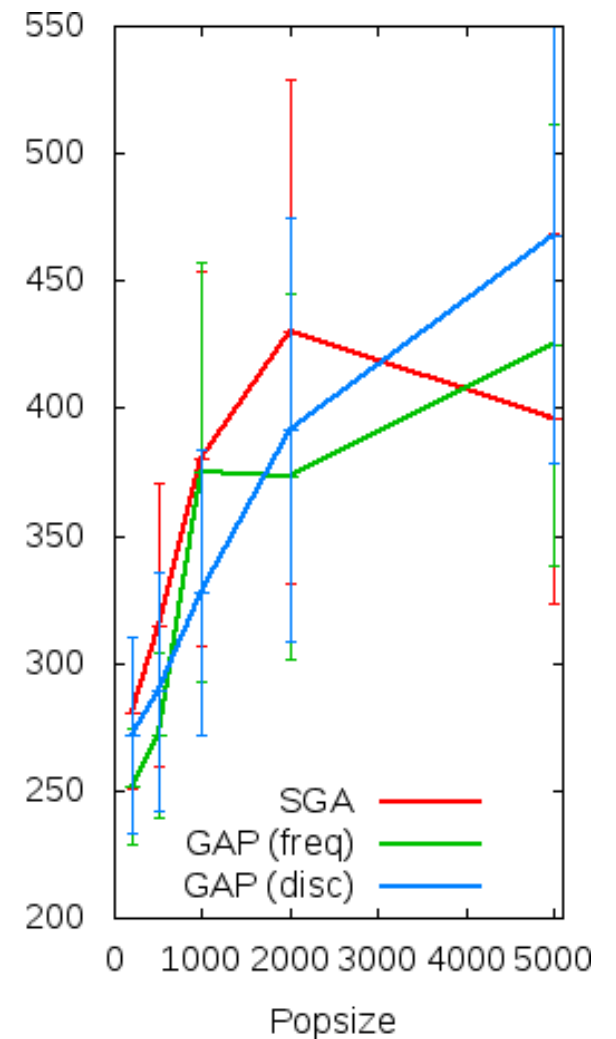
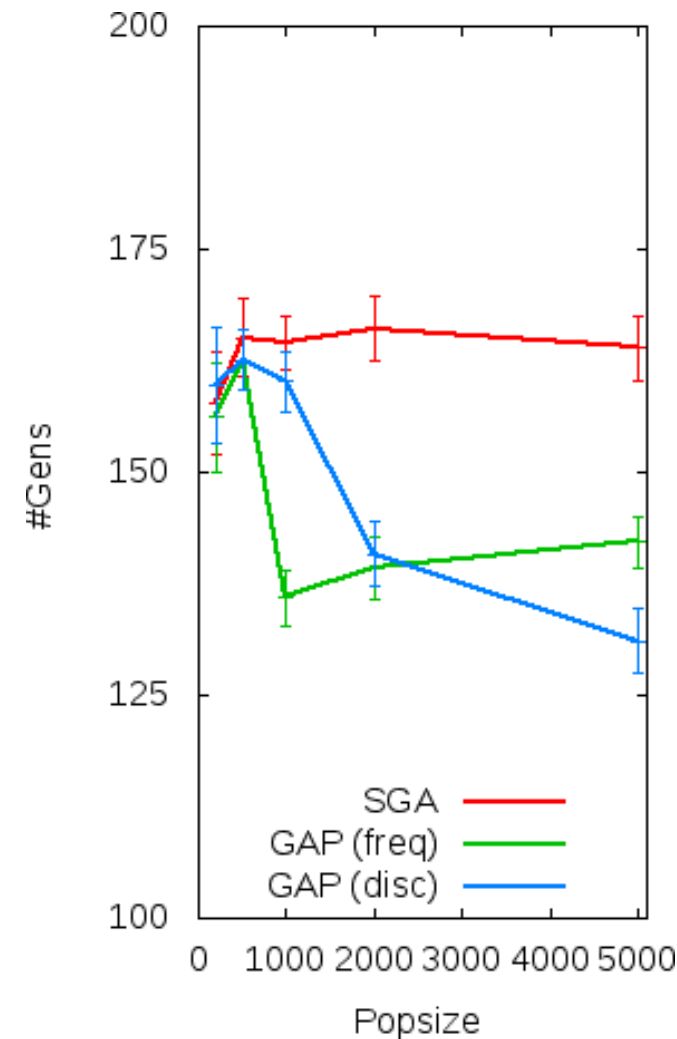
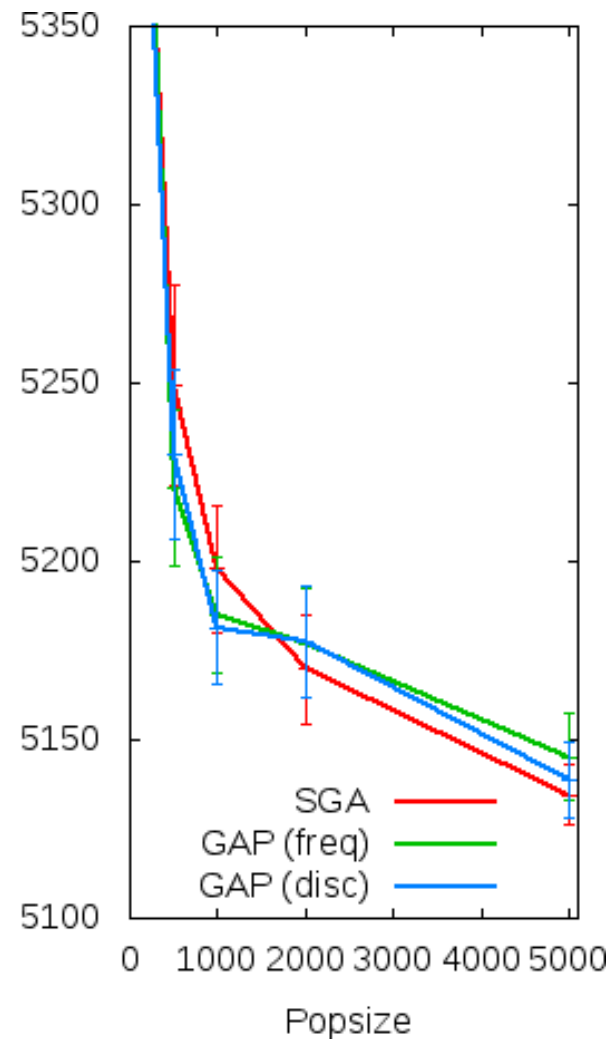
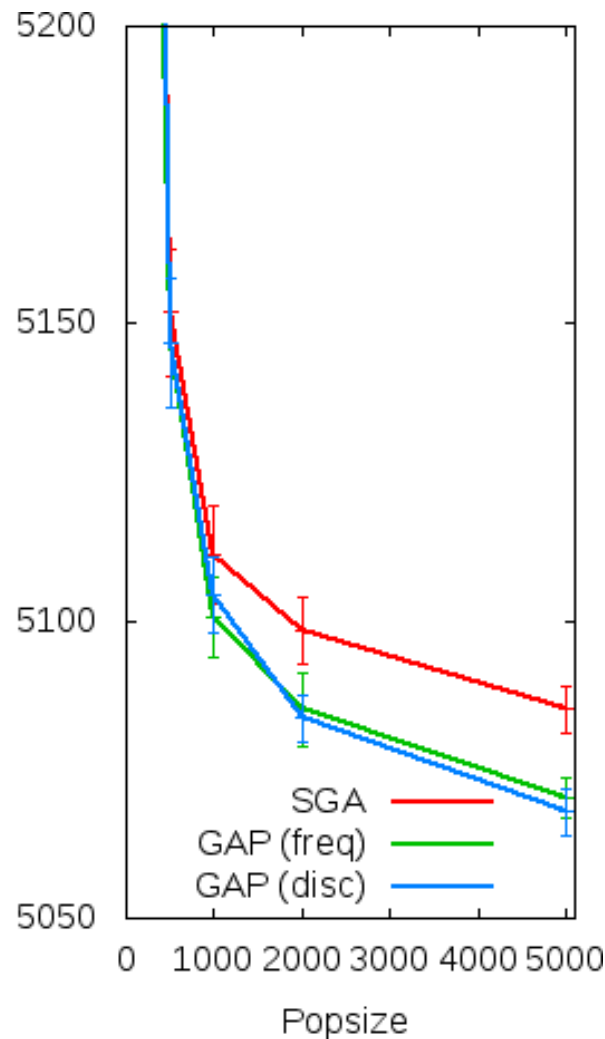
停止までに要した世代数
(試行100回の平均)

エッジ交叉

位置に基づく交叉

エッジ交叉

位置に基づく交叉



※誤差棒は 95% 信頼区間

巡回セールスマン問題 (Grötschels120)

最良個体の適合度 (最適適解 6942)

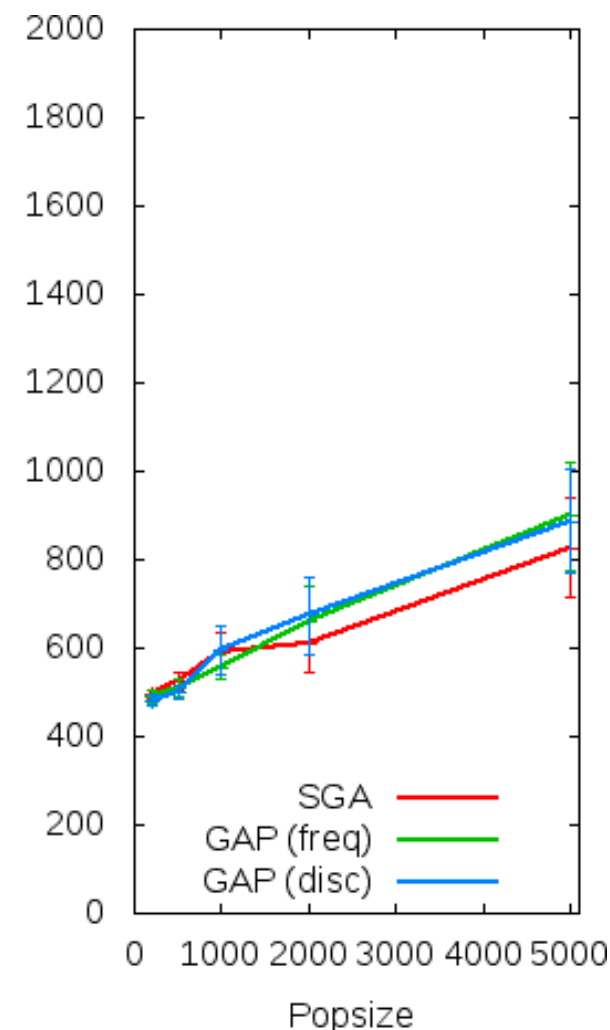
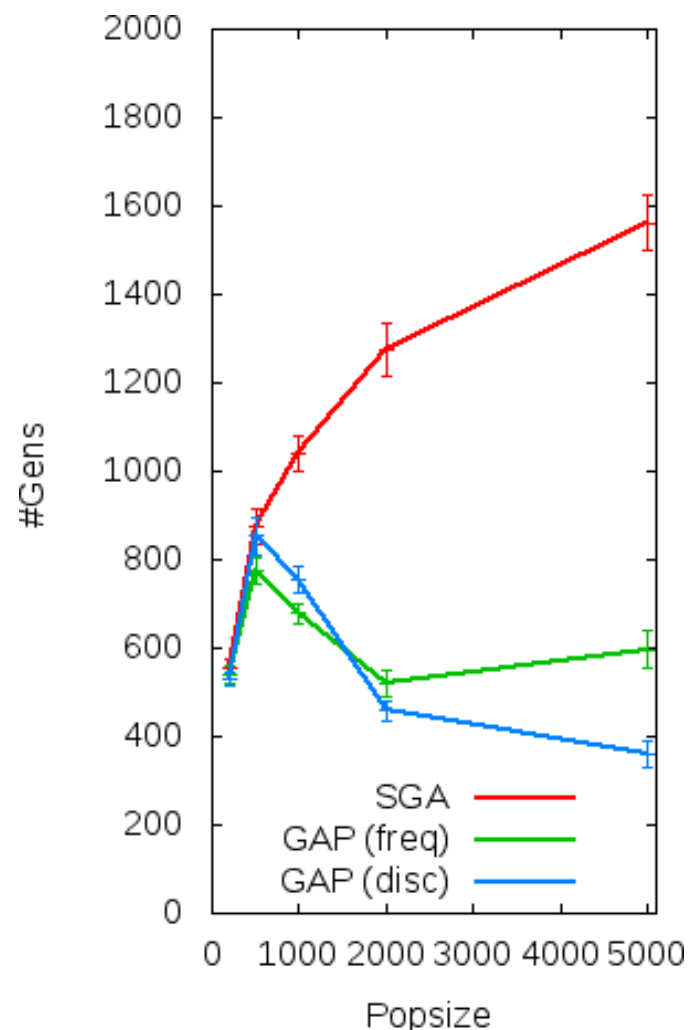
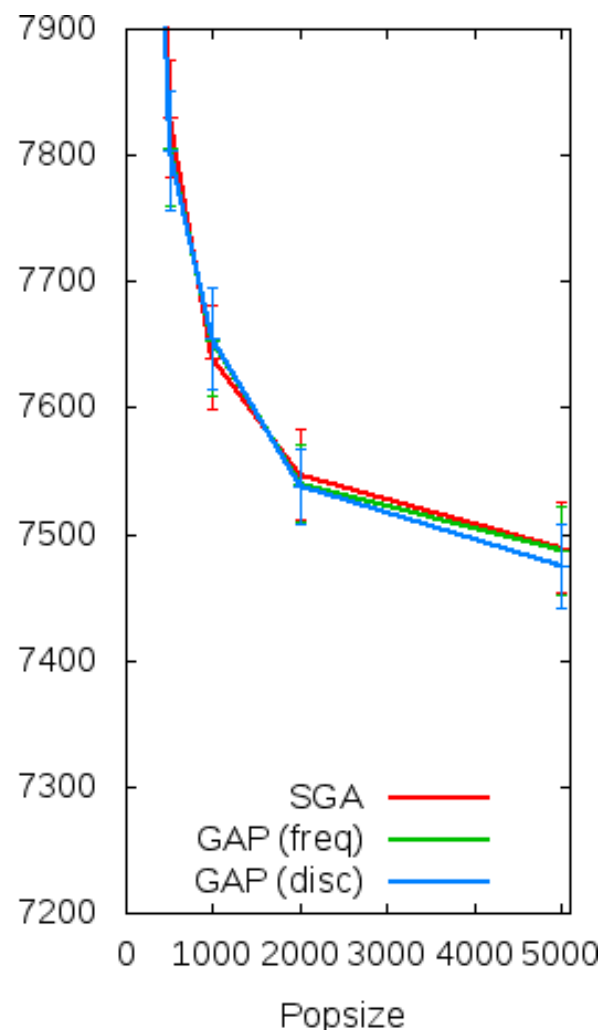
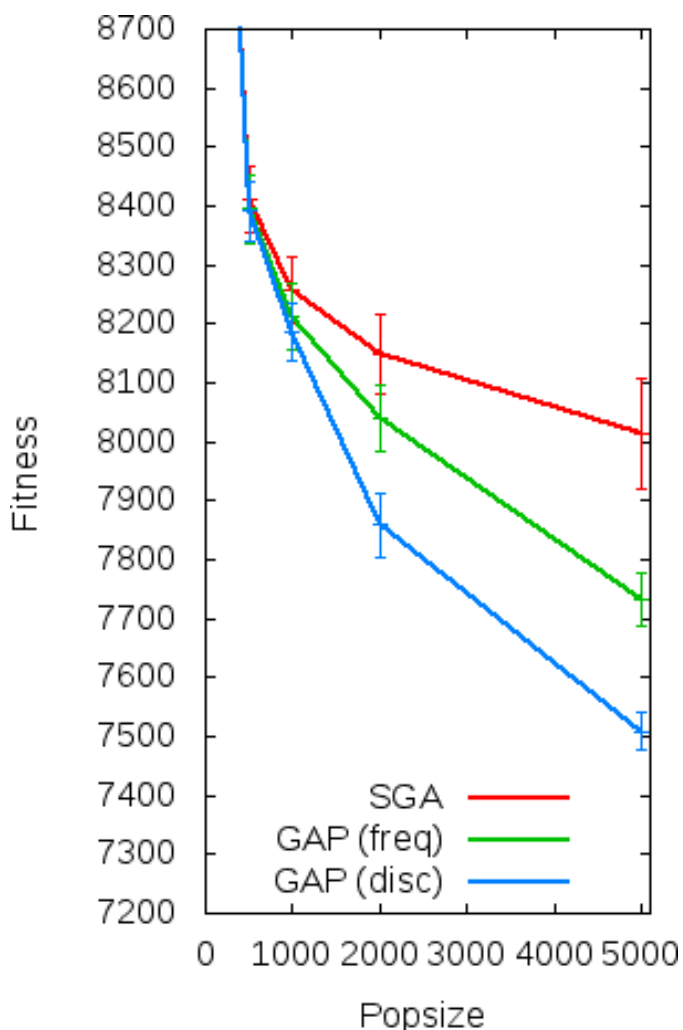
停止までに要した世代数

エッジ交叉

位置に基づく交叉

エッジ交叉

位置に基づく交叉



今後の課題

- だまし問題を用いた比較実験
- 世代に跨る優良個体・部分解ライブラリの導入
- 得られたパターンの観測に基づく進化過程の解析 [Le et al. 08]