

遺傳的アルゴリズムの基礎

經濟情報論講義

遺伝的アルゴリズムとは

- G.A. = Genetic Algorithm
 - 生物の進化にヒントを得た、最適化のためのアルゴリズムの一つである。
 - 選択淘汰 (Selection)
 - 交差 (Cross Over)
 - 突然変異 (Mutation)
- といった概念を用いて、システムをモデル化し、効率的な探索アルゴリズムを与える。

GAで用いられる概念

- 選択
 - 環境により適合した「種」は次世代で個体を増やし、環境に適合しない「種」は個体を減らす
- 交叉
 - 一定確率で二つの「種」の遺伝子配列が組み合わされて新しい種となること
- 突然変異
 - 遺伝子配列の中の特定のビットが一定確率で逆転して、別の種となること
- 世代交代
 - 上記の選択・交叉・突然変異によって、次第により環境に適した種が多数を占めるようになること

GAの原理的な手順

- 与えられた問題のパラメタを「遺伝子」として解釈
 - 初期集団の生成
 - パラメタの組ごとに複数の「種」を生成
 - 「種」は一定数の個体からなる
 - 適応度の評価
 - 選択・交叉・突然変異
 - 世代の進化
- 上記の手順を繰り返す。

GAの簡単な具体例

1. 遺伝子配列の生成
2. 各種の適合度の評価
3. ルーレット法
(次世代の個体数の決定)
4. 交叉と突然変異
5. 世代交代
6. 収束するまで、1～5を繰り返す

1. 遺伝子配列の生成

- $3x - 6 = 0$ の解をGAで求める
- いま、3bitの遺伝子をもった仮想生物を考える。(各種ごとに25個体とする)。
- 初期の遺伝子はランダムに与えられる。
 - 種A: 110 (=6)
 - 種B: 011 (=3)
 - 種C: 000 (=0)
 - 種D: 111 (=7)

2-1. 各種の適合度の評価

- A~Dの各種は、 x に代入されたとき、
 $f(x) = 3x - 6$ が0に近いほど「環境に適合している」と評価される。
逆に、0から遠いほど「誤差が大きい」
- 種A: $f(6) = 12$ 二乗して144
- 種B: $f(3) = 3$ 二乗して9
- 種C: $f(0) = -6$ 二乗して36
- 種D: $f(7) = 15$ 二乗して225

2-2. 各種の適合度の評価

- 種Aから種Dの誤差の最大二乗和は225なので、いま、適合度 $(F) = 225 - (\text{種の誤差二乗})$

とおくと、

$$\text{種A: } F = 81$$

$$\text{種B: } F = 216$$

$$\text{種C: } F = 189$$

$$\text{種D: } F = 0$$

- 適合度の合計は486

3. ルーレット法

- 適合度の高い種ほど次世代の個体数を増やし、低い種ほど次世代で個体数を増やす。
- ルーレット法では、適合度の合計を100%として個体数を適合度に応じて配分する。
- 種A: 16.7%
- 種B: 44.4%
- 種C: 38.9%
- 種D: 0%

4. 交叉と突然変異

- さらに、次世代の種には交叉と突然変異が一定の確率で起こることとすると、次世代は次のようになる。
- 種B(011)と種C(000)の交叉
 - 種E:001 (=1)
- 種Cの突然変異
 - 種F:100 (=4)

5. 世代交代

- ルーレット法で計算した、新たな個体数に交叉と突然変異によって発生した新たな種の個体数を勘案すると、
- 種A: 110 ($=6$) \Rightarrow 16個体
- 種B: 011 ($=3$) \Rightarrow 44個体
- 種C: 000 ($=0$) \Rightarrow 38個体
- 種E: 001 ($=1$) \Rightarrow 1個体
- 種F: 100 ($=4$) \Rightarrow 1個体

6-1. 繰り返し

- 第二世代の誤差二乗
 - 種A:144、種B: 9、種C: 36
 - 種E: 9、種F: 36
- 第二世代の適合度(最大誤差は225のまま)
 - 種A:81、種B:216、種C:189
 - 種E:216、種F:189
 - 適合度合計=891

6-2. 繰り返し

- ルーレット法による個体数の決定
 - 種A:9.1%、種B:24.2%
 - 種C:21.2%、種E:24.2%
 - 種F:21.2%
- 交叉と突然変異
 - 種B(011)と種C(000)との交叉
⇒種G:010
- 新たに環境への適合度を計算して繰り返す

GAはどのような問題に有効か

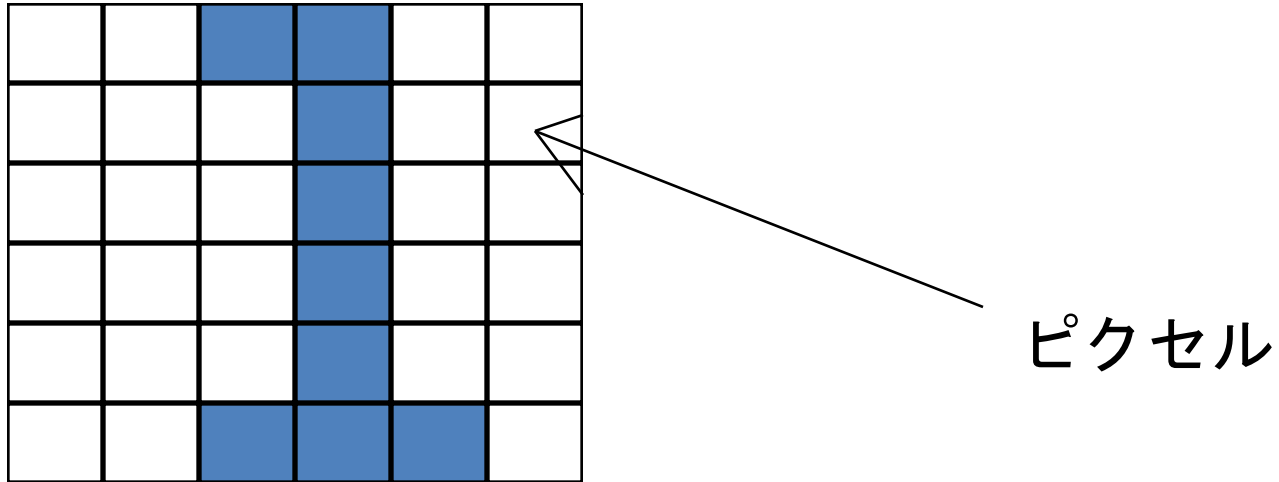
- ここまでの、一次方程式程度であれば、より簡単に解を求めることができるし、数学的な公式も存在する。
- GAが役に立つのは、
 - 数学的・理論的に解を得ることができない場合
 - 解空間があまりに広くて、通常の方法では探索しきれない場合。

GAの応用例

1. 画像復元手法としてのGA
2. GAによる画像復元の原理
3. 復元例

4. GAによる株取引エージェント
5. GAによる社会科学系統計分析

1-1. 画像復元手法としてのGA



各ピクセルについて、白か黒の2値の時、
1 bit/Pixel

明暗の階調が256ある時、
8bit/Pixel

1-2. 画像復元手法としてのGA

- 一般に、撮影された画像は、レンズの収差などによって劣化している。
 - 本来の画像を F 、劣化関数を H 、ノイズを n としたとき、
 - $G = H \cdot F + n$
 - 得られた画像 G から本来の画像 F を得るには、 H の逆関数を施せばよい。
- ⇒ H は知られていることも多いが、 H の逆関数を導くことは、一般に困難である。

2. GAによる画像復元の原理

- そこで、GAを用いて、画像復元を行う。
 1. F に相当する画像を遺伝子配列としてとらえ、さまざまな F の候補となる種 $F'(i)$ を生成する。
 2. H は分かっているので、 $F'(i)$ に劣化過程 H を施し、その結果として $G'(i)$ を得る。
 3. 各種から得られた G' を G と比較し、その差が小さいほど「適合度が高い」と考える。
 4. GAの手順に従って、世代交代を繰り返し行くと、より F に近い F' を得ることができる。

3-1. 復元例 (ピンぼけ)



3-2. 復元例 (光学系収差)



4. GAによる株取引エージェント

- 過去の株価時系列を学習して、売買を行うGAエージェントがいろいろと開発されている。
- 実験段階のものもあるし、現実取引企業が利用している場合もある。
- ただし、手法が浸透するにつれて、利益は次第に減少する。

GAの長所と短所

- 解空間が広大である場合に、GAで得られた解は、「十分に実用的な」近似解ではあるが、それが「真に近似的な」解であるという保証がない。
- すなわち、数学的な解とは異なる解を「最適解」として求めてしまう可能性がある。

ニューラルネットとの連携

- 一方、ニューラルネットは、十分な学習を行った場合に、数学的な解と一致することが知られている。
- 近年では、GAに対する理論的な考察が進むと同時に、GAを用いてニューラルネットを設計することで、効率よく正確な解を得ようとするなどの手法がとられる。

まとめ

- GA: 与えられた問題を、ある遺伝子配列を持った種の進化として捉えて、解を求める。
- 選択・交叉・突然変異・世代交代を繰り返す。
- 交叉と突然変異の確率は、経験的に定めるが、一般に交叉は1~5%、突然変異は0.1%程度とすることが多い。
- 真の解への収束の保証はない。