7M-04

確率的主成分分析と t-SNE 法を用いた 高次元 Artifical Bee Colony アルゴリズム

坂本　昇汰 †　　山口　智 ‡
千葉工業大学 †　　千葉工業大学 ‡

1　はじめに

Artificial Bee Colony (ABC) アルゴリズム [1] は少ないパラメータ設定で高い探索性能が持っている．その反面，解の更新の仕方として解候補の設計変数の一つを更新するため，設計変数間に依存性があるものは解の更新が難しくなる．そこで主成分分析 (PCA) を用いることで個体群の相関に合わせた座標変換を行い，複数の設計変数の更新を行う．しかし，大規模な問題になると次元の呪いや探索が不十分な設計変数によって，理想的な座標変換を行うことができず，探索が遅くなる．そこで本研究では，PCA に確率モデルを取り入れた確率的主成分分析 (PPCA)[3] を用いて次元圧縮を行う．これにより探索する次元を制限することで次元の呪い等の影響を減らし効率的な探索ができると考える．

2　提案手法

高次元の ABC アルゴリズムに探索する次元を PPCA によって定める．また，圧縮した次元に対しては，探索を行わないと最適解を導くのに世代数がかかる．そこで圧縮した次元の重み付けとして GA[4] や t-SNE を用いて複数の次元の更新も試みる．

2.1　次元圧縮

次元圧縮の方法は以下の通りである．個体群を一つの行列として扱う．この行列の共分散行列を C して対数尤度関数 (1) 式を用いる．

$$l(W,\sigma^2) = -\frac{ND}{2}\log(2\pi) - \frac{N}{2}\log|C| - \frac{1}{2}X^T C^{-1} X \quad (1)$$

$x_i(i=1,...,N)$ は個体，$X=[x_1,...,x_n]$ 個体群とした．
(1) 式を W で偏微分し，極値を求めることにより W の最尤推定量は (2) 式になる．

$$\widehat{W} = P(\Lambda + \sigma^2 I)^{\frac{1}{2}} \quad (2)$$

$(\lambda_i, a_i)(i=1,2,...,N)$ を分散共分散の固有値と固有ベクトルの組とし，$\Lambda = \mathrm{diag}(\lambda_1,\lambda_2,...,\lambda_N)$，$P=(a_1,a_2,...,a_N)$ とおいた．PCA では，軸の数を決定するために寄与率や累積寄与率から決定するが PPCA では，データに確立構造を入れてパラメータを最尤法で求めているので，(3) 式の AIC でより客観的に次元数を決めることができる．このとき，AIC が最小になるように新しい次元 D_{new} を決める．

$$AIC = -2\widehat{W} + 2(D - D_{\mathrm{new}}) \quad (3)$$

D が現在の次元数で D_{new} は新しい次元数である．

2.2　GA による重み付け

GA を用いた方法として蜂の探索式をすべての次元でまとめて行い，世代数を減らすものがあるが計算量が増える問題がある．そこで次元圧縮した個体群を一つの次元として GA を用いた手法を取り入れる．

$$v_{[i,j]} = x_{[i,j]} + r * (x_{[i,j]} - x_{[k,j]}) \quad \text{if} \quad r_{ij} < \alpha \quad (4)$$

v は新しい個体．x は現在の個体．i,k は 1 から N の整数乱数，j は i ではない乱数，r は 0 から 1 の一様乱数

2.3　t-SNE 重み付け

GA を行う方法では，圧縮した次元数が多くなると一世代あたりのかかる時間が大きくなる．そこで，t-SNE を用いて一定世代数ごとに点の再生成を行う．

まず，圧縮した点に対して t-SNE を行う．その後，各点とのユークリッド距離を用いる．探索のたびに一定の確率に t-SNE で生成した点を動かす．その後，点との距離を戻すことで圧縮した次元を行うことで動くと考える．また，t-SNE を行うため，それに誤差がある．そこで k3n error[5] を用いて生成誤差を加える．

$$\mathrm{ts}v_{[i,j]} = \mathrm{ts}x_{[i,j]} + r * (\mathrm{ts}x_{[i,j]} - \mathrm{ts}x_{[k,j]}) \quad \text{if} \quad r_{ij} < \alpha (5)$$

$\mathrm{ts}v, \mathrm{ts}x$ は圧縮した点を t − SNE した点である．t-SNE で探索させた次元の評価は次に PPCA を行う前に行う．

3　結果

本実験では，既存手法として，通常の ABC アルゴリズム，PCA を用いた ABC アルゴリズム，PPCA を行い GA による重み付けを行ったものを Proposed1，PPCA を行い t-SNE による重み付けを行ったものは Proposed2 として行った．ベンチマーク関数として，単峰性として Sphere 関数，変数間依存性を持った関数として Ridge 関数の 2 つ関数を 1000 次元で行った．また，パラメータ設定としては，個体数は 100，scouts bee の limit を次元数の 10 倍，PCA，PPCA を行う頻度は 100 探索ごと，PPCA の圧縮した次元の重み付けによる探索頻度は $\frac{\text{現在の次元数}}{\text{次元数の 5 倍}}\%$ とした．(4) 式，(5) 式の α は 0.7 とした．探索を終了条件としては関数値が 10^{-6} 以下に到達するか，世代数が 10 万を超える

A High Dimensional Artificial Bee Colony Algorithm Using Probabilistic principal component analysis with t-SNE
† Shota Sakamoto, Chiba Institute of Technology
‡ Satoshi Yamaguchi, Chiba Institute of Technology

と探索打ち切りとした．Sphere 関数の 1000 次元の関数値の推移を図 1 に，Ridge 関数の 1000 次元の関数値の推移を図 2 にまとめた．

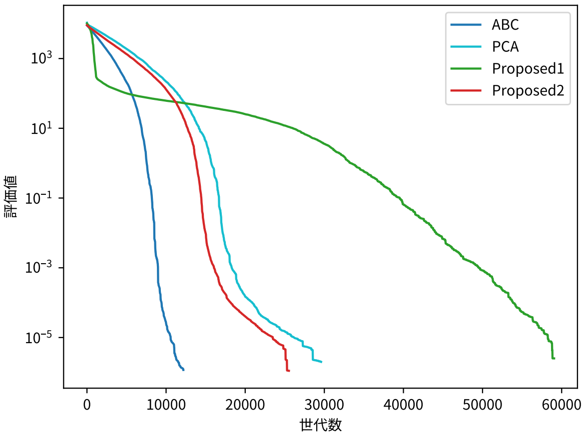

図 1　Sphere 関数 1000 次元の関数値の推移

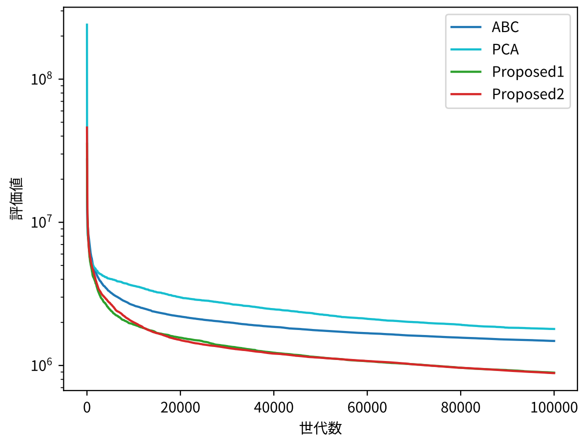

図 2　Ridge 関数 1000 次元の関数値の推移

4 考察

図 1 の Sphere 関数の推移は通常の ABC アルゴリズムが一番早く探索終了しており，次に t-SNE を用いた Proposed2，PCA を用いた手法，GA の考えを入れた Proposed1 の順となっている．しかし，10000 世代以内では Proposed1 の方が ABC アルゴリズムよりも関数値が小さくなっている．この理由としてははじめの方は圧縮する次元数は少ないため GA を用いた探索によって更新がしやすくなったためと考える．圧縮する次元数が多くなると GA を用いた方法では更新が留まり最終的に探索終了は一番遅くなる結果となった．Proposed2 は通常の ABC アルゴリズムを横に伸ばしたような形となり，PCA を用いた ABC と同じような形となった．このような形となった理由として，Sphere 関数と PCA との相性が悪いためとなる．PCA を用いた方法より t-SNE で重み付けしたものの更新や PCA による探索を行わなかった次元によってわずかに早くなったと考えられる．

図 2 の Ridge 関数の関数値の推移においてはどの手法も 10 万世代数到達してから探索を終了してしまった．10 万世代の関数値を見ると Proposed1，Proposed2 がほぼ同じ値で 4 つの手法の中では一番探索が進んでいる．次に ABC アルゴリズム，PCA を用いた手法の順となっている．Proposed1，Proposed2 が ABC アルゴリズム，PCA を用いた手法よりも良くなっている理由としては変数間依存性のある問題では，次元を制限することで特定の次元の更新が多くなり結果として探索性能の向上したと考える．PCA を用いると変数間依存性の問題を解くのに向いているが遅くなっている理由としては，探索ができていない次元や次元の呪いによって理想の相関を得ることができず遅くなっていると考える．探索後半では ABC アルゴリズムよりも下がり幅が大きいため，10 万世代数を上限ではなく最適解が得れるまで行えば通常の ABC アルゴリズムより早くなると考える．提案手法は次元を制限したため，探索ができていない次元や次元の呪いの影響が少なくなったため理想の探索軸を得ることができたと考えられる．

5 おわりに

ABC アルゴリズムの探索において探索する次元を制限し，次元の呪いの影響を少なくして効率的な探索ができると考え行った．変数間依存性のある Ridge 関数では，ABC アルゴリズム，PCA よりも効率的な探索が行うことができた．そのため，高次元の変数間依存性がある問題では，PPCA を行い，制限することによって次元の呪いの影響を小さくすることができたと言える．変数間依存性がない Sphere 関数では，遅くなる結果となった．しかし，探索序盤では探索の進みが良かったことから変数間依存性がない問題でも一部の次元に対して別の方法で探索を行うことで早く探索ができる可能性を感じた．また，多峰性や変数間依存性が弱い関数に関してはあまり良い結果を得ることができなかった．

参考文献

[1] D. Karaboga: “An idea based on honeybee swarm for numerical optimization”, Technical Report TR06, Erciyes University, Engineering Faculty:Computer Engineering Department (2005)
[2] 森　大輔，山口　智:”主成分分析を取り入れた Artificial Bee Colony アルゴリズム”: 電気学会論文誌　C　Vol.135 No.4,pp423-425(2014)
[3] 小野原拓:”確率的主成分分析と次元圧縮”: 大学院研究年報 理工学研究科編, 2012
[4] 前田 陽一:”人工蜂コロニーアルゴリズムのためのハイブリッド探索手法”: 知能と情報（日本知能情報ファジィ学会誌）Vol.30,No.2,pp556-pp563(2018)
[5] Kaneko, Hiromasa. ”k-nearest neighbor normalized error for visualization and reconstruction – A new measure for data visualization performance.” Chemometrics and Intelligent Laboratory Systems 176 (2018): 22-33.