Vol.2013-MPS-95 No.1
2013/9/26

# 大域的クラスタ妥当性指標に基づく距離学習における適応度景観の可視化

女鹿野 大志[1,a)]　福井 健一[2,b)]　小野 智司[1,c)]　沼尾 正行[2]　中山 茂[1]

**概要**：本研究は，クラスタ間の近傍関係を考慮した大域的クラスタ妥当性指標に基づく距離学習において，問題の特性を明らかにする試みである．様々な進化アルゴリズムを距離学習に適用した結果，自己適応型差分進化（jDE）が比較的品質の良い解を発見できたものの，遺伝的アルゴリズム（GA）や共分散行列適応進化戦略（CMA-ES）ではランダム探索と比較して良好な解を得ることができなかった．この原因を探り，より効率的なアルゴリズムを模索するために，本研究では距離学習問題の適応度景観を観察することで問題の性質を調べる．実験により距離学習問題は，局所的な激しい起伏を含み，かつ大域的な傾向も極めて弱いものの，最良解近辺では凸状の傾向が見られることを確認した．

**キーワード**：距離学習，クラスタリング，適応度景観，差分進化

# Fitness Landscape Visualization in Distance Metric Learning based on Global Cluster Validity Measure

Taishi Megano[1,a)]　Ken-ichi Fukui[2,b)]　Satoshi Ono[1,c)]　Masayuki Numao[2]　Shigeru Nakayama[1]

***Abstract:*** This study is an attempt to clarify the problem property of metric learning based on a clustering index with neighbor relation that simultaneously evaluates inter- and intra-clusters. Although self-adaptive differential evolution (jDE) could find a good solution, it was difficult for other evolutionary algorithms to find sufficient solutions. A simple fitness landscape analysis by changing a variable showed that the landscape of this problem involves convex-concave local trends much greater than a slight global trend, which makes the algorithms difficult to search. This paper analyzes fitness landscapes by changing all variables at a time to clarify the reason why jDE could find the sufficient solutions. Experiments have shown that, near the best solutions obtained by jDE, averaging sampled solutions for each distance reveals apparent trends in a convex way, which allows jDE finding the solution.

***Keywords:*** Metric Learning, Clustering, Fitness Landscape, Differential Evolution

## 1. はじめに

データマイニングや機械学習において，距離定義はクラスタリングや判別学習の結果に多大な影響を与える．対象の特殊性が明らかな場合を除き，一般には対象に適した距離尺度を定義することは難しい．そこで近年では，データから適切な距離尺度を学習する距離学習に関する研究が行われている．一般に距離学習は教師なし学習と（半）教師あり学習に分けられる．本研究では後者を扱う．

（半）教師あり学習は，クラスラベルやデータ対の制約条件に基づいて，距離測度の変換関数を求める方法である．大域的距離学習 [1–5] はデータ空間全体に対して，制約をなるべく満たすような単一の変換関数を持つ．一方，局所

[1] 鹿児島大学大学院理工学研究科情報生体システム工学専攻
鹿児島市郡元 1 丁目 21-40
[2] 大阪大学産業科学研究所
大阪府美穂ヶ丘 8-1
[a)] sc108052@ibe.kagoshima-u.ac.jp
[b)] fukui@ai.sanken.osaka-u.ac.jp
[c)] ono@ibe.kagoshima-u.ac.jp

的距離学習 [6,7] は，各データ点の近傍でのみ制約を満たすように，変換関数がデータ空間上における位置の関数の形になっている．

福井らは，教師ありの大域的距離学習を大域的クラスタ妥当性指標に基づいて距離学習を行う方式を提案している [8]．この方法は，任意のクラスタ妥当性指標に基づいた距離学習が行えること，データ対ではなくクラスラベルを使用して，直接クラスタ妥当性指標を改善することが利点として挙げられる．多くの教師あり学習は *must-link* と *cannot-link* と呼ばれる 2 種類のデータ対の制約に基づいている．この制約は，データ数 $n$ の増加に従って，$O(n^2)$ で増加する一方，クラスラベルを用いると，クラスラベルの数はデータ数と同じ $n$ となる．任意のクラスタ妥当性指標については，純度，F 値，エントロピーなど，設計者が目的に応じて選択することができる．

本研究では，大域的クラスタ妥当性指標を適応度として，実数表現 GA（Real-Coded Genetic Algorithm: RCGA）[9]，粒子群最適化（Particle Swarm Optimization: PSO）[10]，共分散行列適応進化戦略（Covariance Matrix Adaptation Evolution Strategy: CMA-ES）[11]，および，適応型差分進化（Self-Adaptive Differential Evolution: jDE）[12] を用いて距離学習を行った [13–15]．実験の結果，jDE は RCGA，PSO，CMA-ES と比較して，よりよい探索性能を示した．

本稿では，上記の距離学習問題において，jDE が他の進化アルゴリズムに比べ良好な解を得られた原因を探り，より効率的なアルゴリズムを模索するために，適応度景観を観察することで問題の性質を調べる．適応度景観に関する研究では，組合せ問題を対象とした手法が提案されている [16,17]．その一方で，本研究で扱う距離学習問題のような，連続値の適応度景観を観察する方法は確立されておらず，適応度景観を分析することは難しいとされている．

## 2. 研究分野の概要

### 2.1 距離学習

本研究の距離学習は，他の大域的距離学習と同様に，マハラノビス距離に基づく距離変換行列を学習する．データセット $D = \{x_i = (x_{i,1} \cdots x_{i,v})^t \in R\}_{i=1}^N$ が与えられた時，マハラノビス距離は以下に定義される．

$$d_{i,j}^2 = (x_i - x_j)^t M (x_i - x_j) \tag{1}$$

ここで，$d_{i,j}$ はクラスタ間距離であり，$M = (m_{k,l})$ は $v \times v$ 行列である．元々のマハラノビス距離では，$M$ は入力データの分散共分散行列の逆行列で与えられ，等分散等共分散化する効果を持っている．一方，マハラノビス距離に基づく距離学習では，$M$ の成分を設計変数として学習する．ただし，$M$ は距離の公理を満たすために半正定値行列である必要がある．

### 2.2 クラスタ妥当性指標

クラスタリングはデータに内在する類似集合を抽出することを目的としているため，クラスタリング精度を定量的に評価することが難しい問題である [18–20]．これまでに提案されているクラスタリングの妥当性指標は，内部基準に基づく距離尺度と外部基準に基づく距離尺度に大別される [19,21,22]．本研究では，外部基準を用いたクラスタ妥当性指標を扱う．外部基準は各データ点のカテゴリやクラスがクラスタによって正確に捉えられているかどうかを評価する，ユーザ視点からの評価である．外部基準を用いたクラスタ妥当性指標としては，純度，エントロピー，F 値や相互情報量などがある．

一般に，全データに対するクラスラベルを得るのはコストがかかるが，一部分のみや，同じ領域の模擬データについてはクラスラベルが得られる場合がある．そのような場合，対象となるデータを直接評価することはできないが，間接的な評価として用いることができる．

## 3. 進化アルゴリズムによる距離学習

### 3.1 概要

本研究では，進化アルゴリズムを距離学習に適用した [8]．大域的クラスタ妥当性指標を適応度として，進化アルゴリズムによりマハラノビス距離に基づく変換行列を学習する．

### 3.2 設計変数

進化アルゴリズムで距離学習を解く場合，個体は距離測度変換行列 $M$ とし，$M$ の上三角成分を設計変数とする．例えば，問題が 2 次元の場合，

$$M = \begin{bmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{bmatrix} \tag{2}$$

に対して，対応する個体ベクトルは $p = (m_{1,1}, m_{1,2}, m_{2,2})$ である．

本問題は，$M$ が半正定値行列であるため以下の制約を含む．

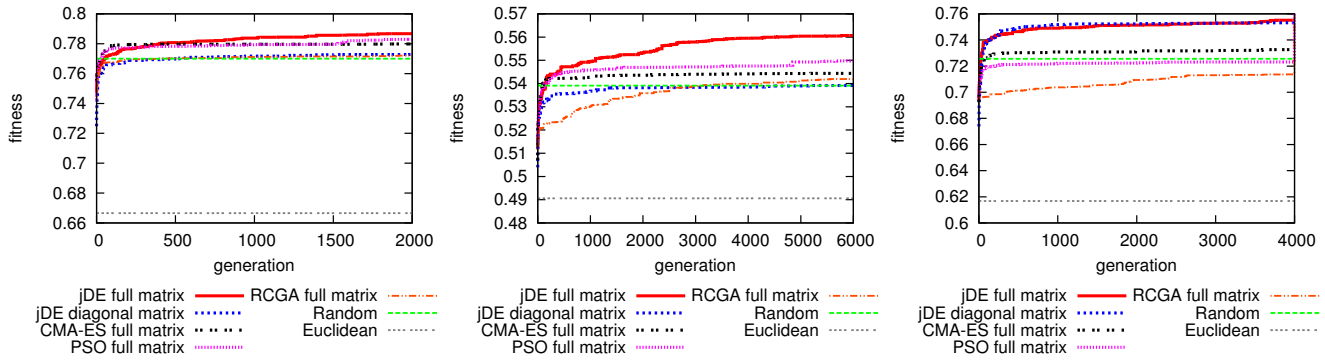
$$\begin{aligned} |m_{i,i}| &> \sum_{j(i \neq j)} |m_{i,j}| \\ m_{i,i} &\in [0,1], m_{i,j} \in [-1,1] \, (i \neq j) \end{aligned} \tag{3}$$

$M$ が半正定値行列であるためには，$M$ が優対角行列かつ対角成分が正である必要がある．優対角の条件を満たさない場合には，$m_{i,j}^{repair} = m_{i,j} / \sum_j |m_{i,j}| \, (i \neq j)$ により，修正する．

### 3.3 目的関数

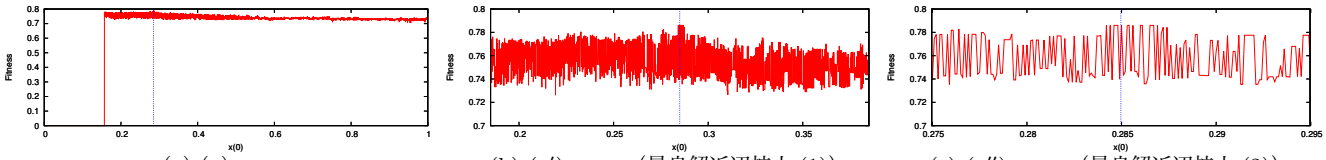
平滑化クラスタ妥当性指標 $Eval$ を最大化するようにマハラノビス距離 $M$ を最適化する．

$$M^* = \arg\max_M Eval\left(Clustering\left(d_{i,j}^2\right)\right) \tag{4}$$




(a) Iris　(b) Glass　(c) Wine

**図 1**　最良個体の適応度の推移



(a) (a) $m_{1,1}$　(b) (a′) $m_{1,1}$　(最良解近辺拡大 (1))　(c) (a″) $m_{1,1}$　(最良解近辺拡大 (2))

**図 2**　適応度景観の例（Iris）

個体の評価は，その個体が表す $M$ による距離尺度（式 (1)）においてクラスタ構造を学習して行う．ここで，クラスタ構造とはクラスタとその近傍関係を示す．本研究では，ベクトル量子化と位相保存が同時に得られる Self-Organizing Map(SOM) [23] を用いる．

クラスタ構造の学習後，得られたクラスタ構造に対して平滑化クラスタ妥当性指標により，クラスラベルを用いてクラスタ内部と近傍関係の質を評価し，変換行列候補の適応度とする．

### 3.4　適応度の推移

著者らは，UCI Machine Learning Repository[*1] に公開されているデータセットを用いて実験を行っている [8,13–15]．Iris，Glass，Wine の 3 種類のデータセットについて，各アルゴリズムの最良解の推移を図 1 に示す．クラスタ妥当性指標は，クラスタの近傍関係による平滑化を導入した F 値（weighted Class F-measure: wCF）を用いた．次元の少ない Iris においては jDE，PSO の性能が優れているものの，RCGA に関してはランダム探索で得られた結果と比較して同程度の探索性能しか見られない．次元の多い Wine においては，jDE のみがランダム探索より性能が優れており，他のアルゴリズムはランダム探索と同程度，もしくは悪い結果が得られた．これより，jDE はすべての問題でランダム探索および他のアルゴリズムよりも適応度の高い解が発見できていることがわかる．

## 4.　適応度景観の可視化

### 4.1　一部の次元による適応度景観

著者らは，問題の性質を調べるため，jDE により発見した最良解の設計変数のうち，1 つまたは 2 つを変更した場合の適応度景観を示している [14]．設計変数の 1 つを変更した場合の適応度景観を図 2 に示す．破線は最良解の値である．

図 2 より，探索空間で適応度の激しい変動が生じており，大域的な変動よりも局所的な変動が激しいことがわかる．よって，PSO や RCGA，CMA-ES ではランダム探索と比較して良好な解を得ることができなかったということがわかったものの，jDE にて比較的品質の良い解が発見できた原因を特定できなかった．

### 4.2　全次元による適応度景観

探索によって得られた最良解を用いて，設計変数をすべて変更することで適応度景観を観察する．個体 $x$ を $(x_1, x_2, x_3, \ldots, x_{n-1}, x_n)$，最適解 $x^{(best)}$ を $\left(x_1^{(best)}, x_2^{(best)}, x_3^{(best)}, \ldots, x_{n-1}^{(best)}, x_n^{(best)}\right)$ とする（$n$ は個体 $n$ の次元数）．最良解からのユークリッド距離 $\delta\left(x, x^{(best)}\right)$ を基準としてサンプリングを行う．本稿では，直交座標系にてサンプリングする方法，極座標系にてサンプリングする方法の 2 つの手法を提案する．サンプリングした個体の，最良解からの距離と適応度を用いて適応度景観を観察する．

[*1] http://archive.ics.uci.edu/ml

表 1　データセットの基本情報

| データセット | 次元数（Diagonal, Full） |
|---|---|
| Iris | 4, 10 |
| Glass | 9, 45 |
| Wine | 13, 91 |

**手法 1: 直交座標系にてサンプリングする方法**

データの各次元において，最良解の値を変化させながらサンプリングを行う．各次元の値は式（5）により求める．

$$x_i = \delta' \left(2 \cdot rand - 1\right) + x_i^{(best)} \tag{5}$$

$rand$ は [0,1] の一様乱数，$\delta'$ はサンプリングの制御パラメータである．全次元のユークリッド距離 $\delta\left(x, x^{(best)}\right)$ は，式 (6) を用いて求める．

$$\delta = \frac{\sqrt{\sum_{i=1}^{n}\left(x_i^{(best)} - x_i\right)^2}}{\sqrt{n}} \tag{6}$$

**手法 2: 極座標系にてサンプリングする方法**

極座標系にて，動径 $r$ を固定してサンプリングを行う．この方法により，最良解から同距離のサンプル数を一定にすることができる．

$$x_i = \begin{cases} rcos\theta_1 + x_1^{(best)} & (i = 1) \\ rsin\theta_1 cos\theta_2 + x_2^{(best)} & (i = 2) \\ rsin\theta_1 \cdots sin\theta_{i-2} cos\theta_{i-1} + x_i^{(best)} & \\ \qquad (i = 3, 4, \ldots, m-1) & (7) \\ rsin\theta_1 \cdots sin\theta_{i-2} sin\theta_{i-1} + x_i^{(best)} & (i = m) \end{cases}$$

$$\theta_i \in \begin{cases} [0, \pi] & (i = 1, 2, \ldots, m-2) \\ [0, 2\pi] & (i = m-1) \end{cases} \tag{8}$$

$m$ はサンプリング対象の次元数で，$\theta_i$ の値は一様乱数で与えられる．式 (7) を用いて，極座標系にて原点から等距離にある個体をサンプリングする．その後直交座標系に変換し，最適解 $x^{(best)}$ の値に応じて平行移動することで，最適解から等距離にある個体を生成する．手法 2 では，$\delta\left(x, x^{(best)}\right) = r/\sqrt{n}$ となる．

### 4.3 実験

#### 4.3.1 実験設定

jDE により発見した最良解を用いて適応度景観を観察する．本稿で提案する 2 種類の手法を用いてサンプリングを行う．データセットの基本情報を表 1 に示す．手法 1 は，変域を $\delta' \in [0, 1]$ とし，$\delta'$ の値を 0.001 ずつ変化させながら，各 $\delta'$ につき 100 個体のサンプリングを行う．手法 2 に関しては，変域を $r \in [0, 1]$ とし，$r$ の値を 0.001 ずつ変化させながら，各 $r$ の値につき 100 個体のサンプリングを行う．すべての手法において，式（3）の制約を満たさない個体が生成された場合は，個体を切り捨てて再生成を行う．

#### 4.3.2 適応度景観の観察

それぞれのデータセットで作成した適応度景観を図 3 から図 6 に示す．距離 $\delta\left(x, x^{(best)}\right) = 0$ の地点に最良解があり，最良解からの適応度の広がりを示している．エラーバーは各距離 100 個体生成した平均値における標準偏差である．図 3，図 4 は距離測度変換行列 $M$ の対角成分のみを最適化した結果（Diagonal）の適応度景観で，それぞれ手法 1，手法 2 でサンプリングした結果である．図 3 と図 4 を比較すると，サンプリング方法によって適応度景観が変化していることがわかる．これは，手法 1 にてサンプリングを行うと，サンプル数が正規分布のようになり，サンプル数の偏りが発生しているためである．

Iris，Glass を見ると，大域的に緩やかな凸状の傾向が見られる．Wine では，図 3(c) に大域的な傾向が見られるが，図 4(c) は最良解付近に平坦な形状が広がっており，進化アルゴリズムで探索することが困難であることがわかる．

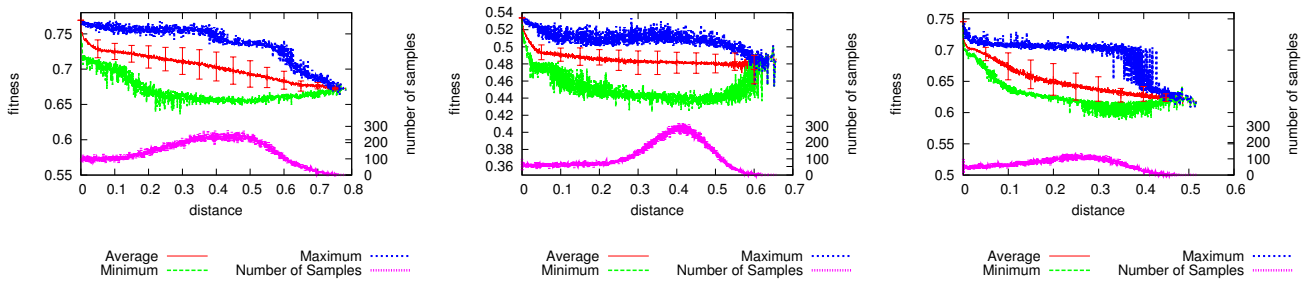
図 5 から図 6 は $M$ の非対角成分も合わせて最適化をした結果（Full）の適応度景観で，それぞれ手法 1，2 によるサンプリング結果である．Glass（図 6(b)）の最良解付近では，平均値で見ると傾向が確認でき，最大値で見ると平坦な形状が広がっている．これは，ある設計変数の値が変化すると，適応度が大幅に変化するためだと考えられる．手法 1 の Wine は制約を満たす個体がサンプリングできなかったため，適応度景観を調べることができなかった．これは，Wine の次元数が多いことで，制約の 1 つである優対角行列を満たす個体の生成が困難であるためだと考えられる．

図 3 から図 6 において，データセットの次元数が多くなると，サンプリング可能な最良解からの距離が近くなっている．これは，次元数の増加に伴い，最良解からの距離が遠くなると，制約条件を満たす個体の生成が困難になるためである．また，Iris（図 4(a)）や，Wine（図 3(c)），Glass（図 6(b)）の適応度の変動が激しくなるのは，超球の表面積が次元の指数に応じて増加するため，単位面積あたりのサンプル数が相対的に少なくなり，適応度の高い個体が発見できなかったものと考えられる．

また，サンプリングした個体の偏りについて調べるため，手法 2 の Full で生成された全個体について解析を行った．Iris の個体は，対角成分，非対角成分ともに各次元での分散が見られた．一方，Glass の個体は対角成分は変化しておらず，非対角成分の一部のみが変化している．これは，式（3）の制約により，探索空間内にサンプリングが安易な箇所，困難な箇所が存在し，サンプリングが安易な箇所に偏ったものと考えられる．

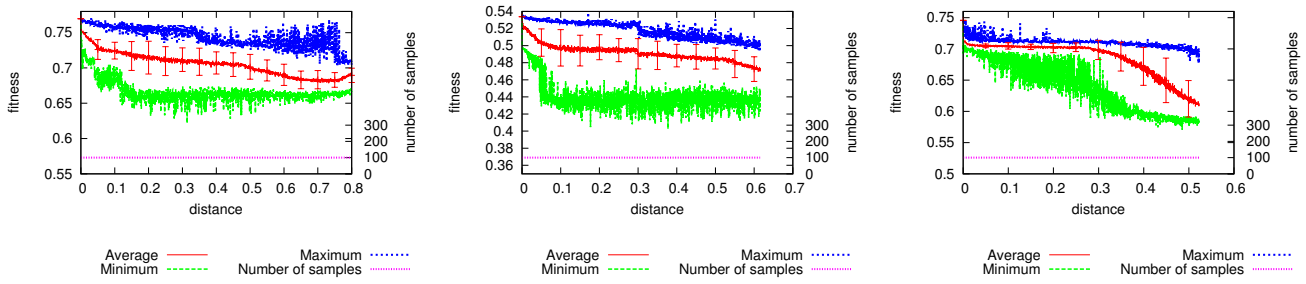
### 4.4 考察

jDE にて発見した最良解を用いて，2 つの手法にてサンプリングを行い，適応度景観を観察した．各手法において



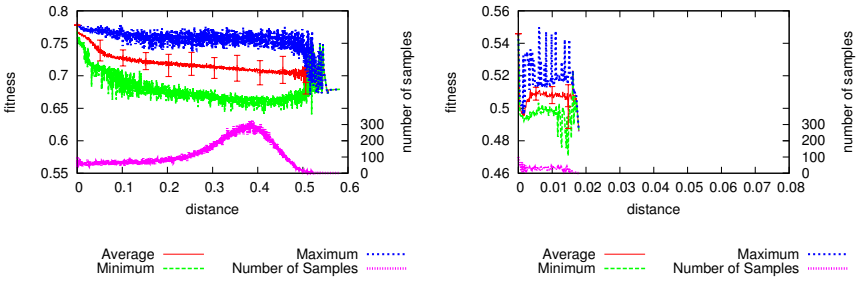

(a) Iris (b) Glass (c) Wine

**図 3** 適応度景観: 手法 1（直交座標系, Diagonal）



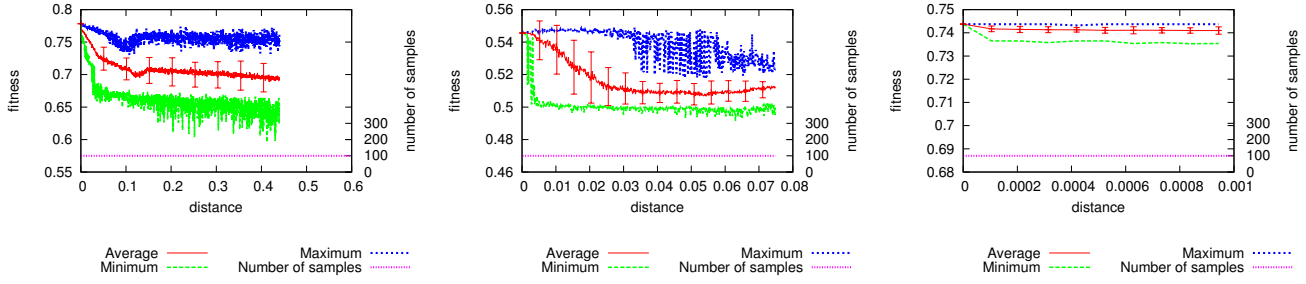

(a) Iris (b) Glass (c) Wine

**図 4** 適応度景観: 手法 2（極座標系, Diagonal）




(a) Iris (b) Glass

**図 5** 適応度景観: 手法 1（直交座標系, Full）




(a) Iris (b) Glass (c) Wine

**図 6** 適応度景観: 手法 2（極座標系, Full）

適応度景観が観察できたものの，手法によって同じデータセットでも適応度景観が変化しており，サンプリングに偏りがあることがわかった．手法 1 は，サンプル数に偏りがあり，サンプル数の少ない部分で適応度景観の信頼性が低くなる．手法 2 は，サンプル数が一定で，サンプル数の偏りが無くなったため，適応度景観の信頼性が高くなった．しかし，サンプリングした個体について分析すると，データセットの次元が多くなると個体の分散が小さくなっており，サンプリングの偏りが見られた．偏ったサンプリングから得られた適応度景観もまた偏りがあると考えられる．

今回のサンプリング方法で観察した適応度景観は，Iris において，いずれの手法でも大域的な傾向が見られる．Glass，Wine では，探索空間に大域的な傾向は見られず，局所的な変動が激しいことがわかった．また，すべてのデータセットにおいて最良解付近に凸状の傾向が見られた．

## 5. おわりに

本研究では，連続値において適応度景観を観察する手法を 2 種類提案し，クラスタ間の近傍関係を考慮した大域的クラスタ妥当性指標に基づく距離学習において適応度景観を観察した．実験により，距離学習問題は局所的な激しい起伏を含み，かつ大域的な傾向も極めて弱いものの，最良解付近では凸状の傾向が見られることを確認した．今後の課題として，偏りが無いサンプリングができるよう，個体のサンプリング方法を改良する必要がある．

### 参考文献

[1] Xing, E. P., Ng, A. Y., Jordan, M. I. and Russell, S. J.: Distance Metric Learning with Application to Clustering with Side-Information, *Advances in Neural Information Processing Systems (NIPS)*, pp. 505–512 (2002).

[2] Bar-Hillel, A., Hertz, T., Shental, N. and Weinshall, D.: Learning Distance Functions Using Equivalence Relations, *Proc. the 20th International Conference on Machine Learning (ICML-03)*, pp. 11–18 (2003).

[3] Goldberger, J., Roweis, S., Hinton, G. and Salakhutdinov, R.: Neighbourhood Components Analysis, *Advances in Neural Information Processing Systems*, pp. 513–520 (2004).

[4] Zha, Z.-J., Mei, T., Wang, M., Wang, Z. and Hua, X.-S.: Robust Distance Metric Learning with Auxiliary Knowledge, *Proc. International Joint Conference on Artificial Intelligence (IJCAI-09)*, pp. 1327–1332 (2009).

[5] Bian, W. and Tao, D.: Learning a Distance Metric by Empirical Loss Minimization, *Proc. International Joint Conference on Artificial Intelligence (IJCAI-11)*, pp. 1186–1191 (2011).

[6] Yang, L., Jin, R., Sukthankar, R. and Liu, Y.: An Efficient Algorithm for Local Distance Metric Learning, *Proc. the National Conference on American Association for Artificial Intelligence (AAAI)*, pp. 543–548 (2006).

[7] Weinberger, K. Q., Blitzer, J. and Saul, L. K.: Distance Metric Learning for Large Margin Nearest Neighbor Classification, *Journal of Machine Learning Research (JMLR)*, Vol. 10, pp. 207–244 (2009).

[8] 福井健一，沼尾正行：大域的クラスタ妥当性指標に基づく距離学習，情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告， Vol. 2012-MPS-87, No. 32, pp. 1–6 (2012).

[9] Eshelman, L. J. and Schaffer, J. D.: Real-coded genetic algorithms and interval-schemata., *Foundation of Genetic Algorithms 2*, pp. 187–202 (1993).

[10] Kennedy, J. and Eberhart, R.: Particle Swarm Optimization, *Proc. IEEE International Conference on Neural Networks (ICNN'95)*, pp. 1942–1948 (1995).

[11] Hansen, N. and Ostermeier, A.: Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation, *Proceedings of The 1996 IEEE International Conference on Evolutionary Computation*, pp. 312–317 (1996).

[12] Brest, J., Greiner, S., Boskovic, B., Mernik, M. and Zumer, V.: Self-Adapting Control Parameters in Differential Evolution: A Comparative Study on Numerical Benchmark Problems, *IEEE Transactions on Evolutionary Computation*, Vol. 10, No. 6, pp. 646–657 (2006).

[13] 福井健一，小野智司，沼尾正行：大域的クラスタ妥当性指標に基づく差分進化による距離学習，人工知能学会データ指向構成マイニングとシミュレーション研究会 (SIG-DOCMAS)， No. B201 (2012).

[14] 小野智司，福井健一，堤田沙由里，澤井陽輔， 中山茂，沼尾正行：大域的クラスタ指標に基づく距離学習への適応型差分進化法の適用，第 4 回進化計算学会研究会，pp. 52–57 (2013).

[15] Fukui, K., Ono, S., Megano, T. and Numao, M.: Evolutionary Distance Metric Learning Approach to Semi-Supervised Clustering with Neighbor Relations, *Proc. of 25th IEEE International Conference on Tools with Artificial Intelligence (ICTAI-13)* (2013(to appear)).

[16] Merz, P. and Freisleben, B.: Fitness landscape analysis and memetic algorithms for the quadratic assignment problem, *Evolutionary Computation, IEEE Transactions on*, Vol. 4, No. 4, pp. 337–352 (2000).

[17] Malan, K. M. and Engelbrecht, A. P.: A survey of techniques for characterising fitness landscapes and some possible ways forward, *Information Sciences*, Vol. 241, pp. 148 – 163 (2013).

[18] Deborah, L. J., Baskaran, R. and Kannan, A.: A Survey on Internal Validity Measure for Cluster Validation, *International Journal of Computer Science & Engineering Survey (IJCSES)*, Vol. 1, No. 2, pp. 85–102 (2010).

[19] Kovács, F., Legány, C. and Babos, A.: Cluster Validity Measurement Techniques, *Engineering*, Vol. 2006, pp. 388–393 (2006).

[20] Tenenbaum, J. B., de Silva, V. and Langford, J. C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science*, Vol. 290, pp. 2319–2323 (2000).

[21] Kremer, H., Kranen, P., Jansen, T., Seidl, T., Bifet, A., Holmes, G. and Pfahringer, B.: An Effective Evaluation Measure for Clustering on Evolving Data Streams, *Proc. Conf. Knowledge Discovery and Data Mining*, pp. 868–876 (2011).

[22] Rendon, E., Abundez, I., Arizmendi, A. and Quiroz, E.: Internal versus External Cluster Validation Indexes, *International Journal of Computers and Communications*, Vol. 5, No. 1, pp. 27–34 (2011).

[23] Kohonen, T.: *Self-Organizing Maps*, Springer-Verlag (1995).