

大域的クラスタ妥当性指標に基づく距離学習

福井 健一^{†1} 沼尾 正行^{†1}

本論文では、クラスタ間の近傍関係も考慮した大域的クラスタ妥当性指標に基づく距離学習法を提案する。提案法は、大域的クラスタ妥当性指標を適合度とした遺伝的アルゴリズム (GA) により、マハラノビス距離に基づく変換行列を学習する。評価実験では、妥当性指標の最適な平滑化半径の決定について示し、続いて距離学習により妥当性指標の値が改善することを確認すると共に、自己組織化マップ (SOM) による可視化を通じて提案法の有効性について考察する。

Metric Learning based on Global Cluster Validity Index

KEN-ICHI FUKUI^{†1} and MASAYUKI NUMAO^{†1}

This paper proposes a metric learning method based on the global cluster validity index that considers inter-cluster neighborhood relation. The proposed method learns a transformation matrix of Mahalanobis distance by genetic algorithm, where the global cluster validity index as the fitness function. The experiments reveals determination of the appropriate smoothing radius of the validity index, then confirms that the proposed distance metric learning improves the validity index. Also, we discuss the effectiveness of the proposed method via visualization by the self-organizing maps.

1. はじめに

データマイニングや機械学習において、クラスタリングや判別学習を行う際、データ点間の距離定義もしくは特徴の重み付けは、結果に大きな影響を与える。しかし、適切な距離定義や重み付けは難しく、一般には、ヒューリスティックに与える。そこで、観測データから

適切な距離学習の研究がこれまで多く行われている。

まず、教師なしの距離学習法には、ISOMAP¹⁵⁾、LLE (Local Linear Embedding)¹⁴⁾、Laplacian Eigenmaps³⁾ などがあり、データの非線形な分布構造を写し取る多様体学習とも呼ばれている。一方、教師ありの距離学習の多くは、データ対に対して must-link と cannot-link の2種類の制約を課し、なるべく制約を満たすような距離の変換をデータより求める。大域距離学習^{2),18)} は、特徴空間全体で共通の距離変換行列を持つ。一方、局所距離学習^{17),20)} は、距離変換行列が特徴空間上の位置やデータ対に依存する関数の形になっている。局所距離は、柔軟性がある一方で、学習に必要なパラメータが膨大であることや、過学習する傾向があることが問題である²⁰⁾。本研究は、教師ありの大域距離学習に属する。

本研究では、著者らが提案している大域的クラスタ妥当性指標⁷⁾ に基づいた距離学習を行う方法を提案する。提案法は、大域的クラスタ妥当性指標を適合度とした遺伝的アルゴリズム (GA) により、マハラノビス距離に基づく変換行列を学習する。先行の距離学習法と比較して本手法の利点は、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である。例えば、純度、F 値、エントロピーなど設計者が応用や目的に応じて選択することができる。本実験では、実データを用いた提案法の初期検証実験を行った。距離学習により妥当性指標の値が向上することを確認すると共に、自己組織化マップ (Self-Organizing Map: SOM)⁹⁾ の可視化を通じて提案法の有効性について考察する。

2. 大域的クラスタ妥当性指標

本節では、大域的クラスタ妥当性指標⁷⁾ について説明する。

2.1 通常の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は難しい問題である^{5),10),19)}。これまで提案されているクラスタリングの妥当性指標は、内部基準に基づく評価尺度と外部基準に基づく評価尺度の大きくふたつに分類される^{10),11),13)}。

内部基準は、データ間の距離のみに基づいて、得られたクラスタの compactness と separability⁵⁾ を評価する学習の観点からの評価である。Dunn-index⁶⁾、DB-index⁴⁾、CDBw⁸⁾ などがある。内部基準を用いたクラスタ妥当性指標のサーベイは文献^{5),12)} にある。一方、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評価指標としては、純度、エントロピー、F 値や相互情報量などがある^{13),16),19)}。

^{†1} 大阪大学産業科学研究所

The Institute of Scientific and Industrial Research (ISIR), Osaka University

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。一般に、全データに対するクラスタ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タラベルが得られる場合もある。そのような場合、直接対象となるデータを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。

2.2 準備

定義 1. (クラスタリング) v 次元のデータ $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,v})^t$ からなるデータセット $S = \{\mathbf{x}_i\}_{i=1}^N$ 、およびデータ点間の距離 $D(\mathbf{x}_i, \mathbf{x}_j)$ が与えられた下で、クラスタリングは S をクラスタの集合 $C = \{C_i\}_{i=1}^K$ への分割を与えるものとする。ただし、 $C_i \cap C_j = \emptyset (i \neq j)$ とする。また、 \mathbf{x}_i のクラスタ割り当てを $c(i) \in C$ とする。

定義 2. (クラス) クラスの集合を $T = \{T_i\}_{i=1}^L$ とし、 \mathbf{x}_i のクラス割り当てを $t(i) \in T$ とする。クラスはクラスタリングとは独立に S に与えられているものとする。

定義 3. (近傍関係) $d(C_i, C_j) : C \times C \mapsto \mathbb{R}$ をユークリッド距離やトポロジに基づく距離で与えられるクラスタ間距離とする。クラスタ間の近傍関係は、クラスタ間距離により与えられるものとする。

例) クラスタ間距離 ユークリッド距離に基づくクラスタ間距離は、最短距離法や最長距離法などの凝集型階層的クラスタリングで用いられる距離がある。一方、トポロジに基づく距離は、クラスタの近傍グラフ上の最短経路のホップ数や重みで与えられる。ここで近傍グラフは、ユークリッド距離に基づくクラスタ間距離に閾値を設けることや、 k -近傍法などで求められる。

本研究では、自身のクラスタ内のみならず近傍クラスタに属するデータのクラス T の分布により、クラスタ全体の大域的評価を行う。特に、クラス数よりもクラスタ数の方が多い ($K \gg L$) 場合のマイクロクラスタに対して、有効に働く指標である。外部基準に基づくクラスタ妥当性指標は、集合に基づく指標と、データ対に基づく指標の大きくふたつに分類される。それぞれに対する拡張方法を次節より述べる。

2.3 集合に基づくクラスタ妥当性指標の拡張

まず、集合に基づくクラスタ妥当性指標^{13),16)}の拡張について説明する。各指標の特性については詳しくは文献¹⁾で述べられている。

集合に基づく指標は、各クラスタに属する各クラスのデータ数の分布によって評価している。そこで、基本的な統計量に対して、クラスタ間距離による重みを導入することで、近傍のクラスタも含めた指標に拡張する。つまり、近傍のクラスタの重みを大きく、遠いクラスタの重みを小さくし、統計量の全体の総和を取る (図 1)。

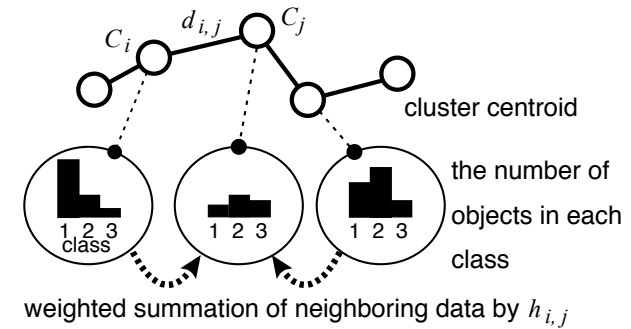


図 1 集合に基づくクラスタ妥当性指標の拡張の模式図。基本的な統計量はクラスタ間距離 $d_{i,j}$ によって重み付けられる。この図は、トポロジに基づくクラスタ間距離の例である。

Fig. 1 Extension of a set-based cluster validity measure. The basic statistics are weighted by the neighborhood relation based on inter-cluster distance $d_{i,j}$. This example shows topology-based distance.

具体的には、 N_i は第 i クラスタのデータ点数、 N_t は第 t クラスのデータ点数、 $N_{t,i}$ は第 i クラスタに含まれる第 t クラスのデータ点数を表すとする。ここで、 $N_{t,i}$ 、 N_i 、 N を重み関数 $h_{i,j}$ により次式のように重み付けする。

$$N'_{t,i} = \sum_{C_j \in C} h_{i,j} N_{t,j}, \quad (1)$$

$$N'_i = \sum_{l \in T} N'_{l,i} = \sum_{l \in T} \sum_{C_j \in C} h_{i,j} N_{l,j}, \quad (2)$$

$$N' = \sum_{C_i \in C} N'_i = \sum_{C_i \in C} \sum_{l \in T} \sum_{C_j \in C} h_{i,j} N_{l,j}. \quad (3)$$

ここで、 $h_{i,j}$ は非負の単調減少関数を用いる。ガウス関数で与える場合は、 $h_{i,j} = \exp(-d_{i,j}/\sigma^2)$ となる。 $d_{i,j}$ はクラスタ間距離であり、 $\sigma (> 0)$ は平滑化半径である。

例として、純度とエントロピーの拡張は、式 (1), (2), (3) を用いて次式で与えられる。

weighted Cluster Purity (wCP)

$$\text{wCP}(C) = \frac{1}{N'} \sum_{C_i \in C} \max_{l \in T} N'_{l,i}. \quad (4)$$

weighted Entropy (wEP)

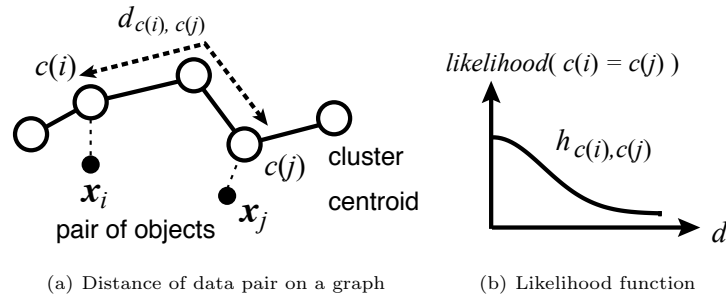


図 2 データ対に基づくクラスタ妥当性指標の拡張の模式図。データ対が同じクラスに属する尤もらしさを表す尤度関数を導入する。
Fig. 2 Extension of a pairwise-based cluster validity measure. A likelihood function is introduced to represent a degree that a data pair belongs to the same cluster.

$$\text{wEP}(\mathbf{C}) = \frac{1}{K} \sum_{C_i \in \mathbf{C}} \text{Entropy}(C_i), \quad (5)$$

$$\text{Entropy}(C_i) = -\frac{1}{\log N'} \sum_{l \in T} \frac{N'_{l,i}}{N'_i} \log \frac{N'_{l,i}}{N'_i}, \quad (6)$$

2.4 データ対に基づくクラスタ妥当性指標の拡張

次にデータ対に基づく指標^{1),19)}について述べる。まず、表1に示すペアワイズ分割表を拡張する。ここで、 a, b, c, d はデータのペアが同じクラス（クラスタ）に属している（属していない）数である。拡張の考え方としては、同じクラスタに属すデータ対の数の代わりに、同じクラスタである尤度 $\text{likelihood}(c(i) = c(j))$ を考える。これは、図2に例示するように、クラスタ間距離によりデータ対が同じクラスタである尤度を与える。尤度関数は集合に基づく指標と同じ重み関数を用いることができる。すなわち、 $\text{likelihood}(c(i) = c(j)) = h_{c(i), c(j)}$ とする。そして、この尤度の総和により次式のようにペアワイズ分割表を拡張する。

表 1 ペアワイズ分割表
Table 1 Cross table for pairwise of data

	$t(i) = t(j)$	$t(i) \neq t(j)$
$c(i) = c(j)$	a	b
$c(i) \neq c(j)$	c	d

$$a' = \sum_{\{i,j|t(i)=t(j)\}} h_{c(i), c(j)}, \quad (7)$$

$$b' = \sum_{\{i,j|t(i) \neq t(j)\}} h_{c(i), c(j)}, \quad (8)$$

$$c' = \sum_{\{i,j|t(i)=t(j)\}} (1 - h_{c(i), c(j)}) = a + c - a', \quad (9)$$

$$d' = \sum_{\{i,j|t(i) \neq t(j)\}} (1 - h_{c(i), c(j)}) = b + d - b'. \quad (10)$$

これら拡張した a', b', c', d' を用いて、例えばペアワイズ精度やペアワイズ F 値は以下のように定義される。

weighted Pairwise Accuracy (wPA)

$$\text{wPA}(\mathbf{C}) = \frac{a' + d'}{a' + b' + c' + d'}. \quad (11)$$

weighted Pairwise F-measure (wPF)

$$\text{wPF}(\mathbf{C}) = \frac{2 \cdot P \cdot R}{P + R}, \quad (12)$$

ここで、 $P = a' / (a' + b')$ は適合率、 $R = a' / (a' + c')$ は再現率を表している。

2.5 重み関数

集合に基づく指標とデータ対に基づく指標の拡張に用いた重み関数 $h_{i,j}$ には、任意の非負の単調減少関数を用いることができる。例えば、徐々に減少させる場合にはガウス関数、一定の距離で打ち切りたい場合は矩形の関数を用いる。ここで、 $h_{i,j} = \delta_{i,j}$ (δ はクロネッカーのデルタ) のとき、元々の指標となるため自然な拡張であるといえる。

次に、図3に重み関数の効果を示す。平滑化半径 σ が 0 に近づくに従って、クラスタ個別に評価することになり、反対に大きくなるに従って重みが一定に近づき、全データをひとつのクラスタとして評価することになる。

2.6 最適な平滑化半径の決定

大域的クラスタ妥当性指標は、通常の妥当性指標にクラスタ間の近傍関係の重み付けがなされている。そこで、大域的クラスタ妥当性指標に含まれる近傍成分を次のように定義する。
定義 4. (近傍成分) 評価値 $Eval$ と、ランダムに近傍関係が与えられた場合の評価値 $Eval_{rnd}$ の差分を、 $Eval$ に含まれる近傍成分とする。

ここで、 $Eval$ は任意の大域的クラスタ妥当性指標の評価値を表している。

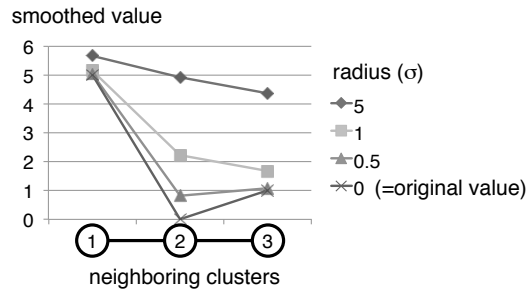


図3 平滑化半径の影響。半径が大きくなるに従って、近傍上で値が平滑化されていく。

Fig. 3 Example of the effect of smoothing radius. Values over the neighborhood relation of the clusters become smoother as the radius increases.

近傍成分 (NC) は次式により求められる。

$$NC(\sigma) = |Eval - \lim_{n \rightarrow \infty} Eval_{rnd(n)}| = |Eval(d_{i,j}) - Eval(\bar{d}_{i,j})|, \quad (13)$$

ここで、 $Eval_{rnd(n)}$ は n 回クラスタ間距離 $d_{i,j}$ をシャッフルしたときの評価値とする。 $n \rightarrow \infty$ で、平均クラスタ間距離 $\bar{d}_{i,j}$ を用いたときの評価値に収束する。この近傍成分が最大になるときの平滑化半径を最適な半径とする。すなわち、 $\sigma^* = \arg \max_{\sigma} NC(\sigma)$ である。この半径を用いて、あるデータセットに対する評価値は $Eval^* = Eval(\sigma^*)$ で求められる。

3. 大域的クラスタ妥当性指標に基づく距離学習

3.1 距離学習

本研究では、次式で与えられるマハラノビス距離に基づく距離学習を行う。

$$D_{i,j} = (\mathbf{x}_i - \mathbf{x}_j)^t \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j) \quad (14)$$

通常のマハラノビス距離では、 $v \times v$ 行列 \mathbf{M} は分散共分散行列の逆行列で定数として与えられる。本研究では \mathbf{M} の成分を変数として扱い、大域的クラスタ指数が最大になるような \mathbf{M} を求める。すなわち、

$$\mathbf{M}^* = \arg \max_{\mathbf{M}} Eval(D_{i,j}) \quad (15)$$

である。ただし、距離の定義を満たすために、 \mathbf{M} は半正定値実対称行列である。

3.2 GA による探索

式 (15) は、一般には多峰性の関数となることと、値が求まらない点がある関数となるため勾配法により求めるのは困難である。そこで、本研究では勾配を用いない方法として、多

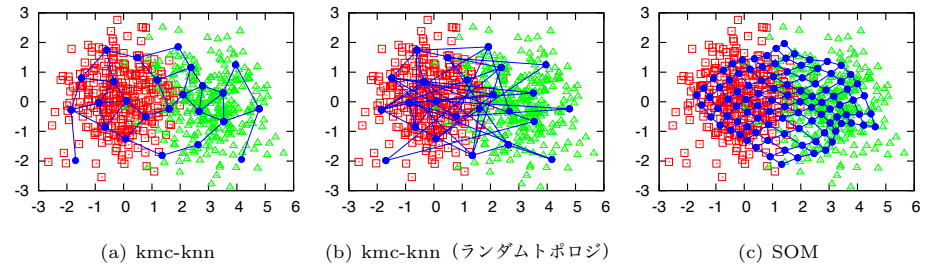


図4 人工データ、クラスタのプロトタイプとその近傍関係の図。□と△はそれぞれ異なるガウス分布から生成されたクラスのデータ点を表している。●がクラスタのプロトタイプ、それらを結ぶエッジが近傍関係を表している。
Fig. 4 Cluster prototypes (●) with topology-based neighbor relation on two dimensional synthetic data. The data points (□, △) were generated from two Gaussian distributions.

点並列探索法として広く用いられている遺伝的アルゴリズム (GA) に着目した。行列 \mathbf{M} を個体とし、遺伝子は行列 \mathbf{M} の各成分 (ただし、対称な成分は除く) とした。適合度関数は以下のように定義した。

$$F(\mathbf{M}) = \begin{cases} Eval(D_{i,j}) & \mathbf{M} \text{ が半正定値のとき} \\ 0 & \mathbf{M} \text{ が半正定値でないとき} \end{cases} \quad (16)$$

欠落点では適合度を 0 にしている。なお、本稿の GA の実装は、GALib^{*1} の SimpleGA を用いた。

4. 評価実験

4.1 評価実験 1: 大域的クラスタ妥当性指標の特性

データの特性を人為的に操作できるようにするため、2 クラス 2 次元のガウス分布から生成したデータを用意した。クラスタリングは k -means 法 (プロトタイプ数 $k1$ 個) を用い、近傍関係はクラスタプロトタイプ間の相互 k -近傍 (近傍数 $k2$ 個) により得た。人工データおよび、クラスタプロトタイプとその近傍関係を図 4 に示す。

平滑化半径を変化させた時の評価値、およびランダムトポロジに対する評価値と近傍成分のグラフを図 5 に示す。半径 $\sigma \rightarrow 0$ 、もしくは $\sigma \rightarrow \infty$ のとき $NC \rightarrow 0$ となる。また、近傍成分 NC が最大のとき、常に評価値 $Eval$ も最大になるとは限らないため、 NC を探索

*1 <http://lancet.mit.edu/ga/>

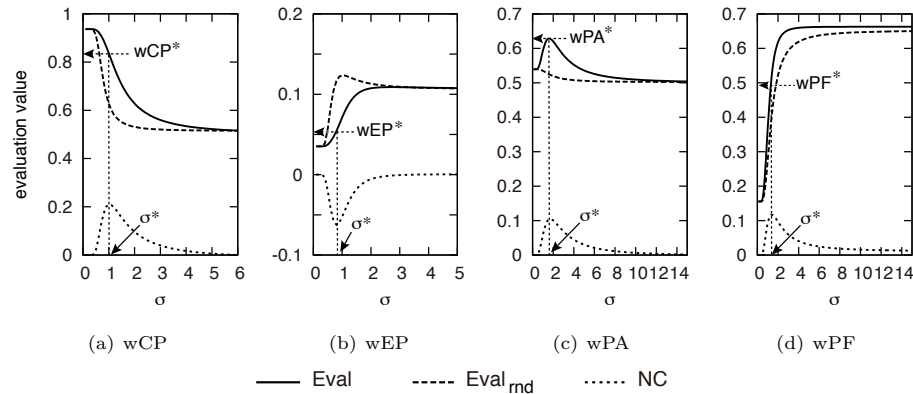


図5 人工データに対して kmc-knn ($k_1 = 25, k_2 = 4$) を用いたときの平滑化半径の影響. $Eval$ は大域的クラスタ妥当性指標の評価値, $Eval_{rnd}$ はランダムトポロジに対する評価値, NC は近傍成分を表している.
Fig.5 The effect of smoothing radius for synthetic data with kmc-knn ($k_1 = 25, k_2 = 4$); total evaluation value ($Eval$), $Eval$ with random topology ($Eval_{rnd}$), neighbor component (NC).

する必要があることを示している. NC が最大となる半径 σ^* のときの評価値を用いて距離学習を行う. その他, データ分布と評価値との関係については, 先行研究⁷⁾ に掲載されている.

4.2 評価実験 2: 大域的クラスタ妥当性指標に基づく距離学習

UCI Machine Learning Repository^{*1} に公開されている Iris データ (アヤメ科植物の分類データ) を用いた. データ数は 150, 次元数 4, 3 クラスのデータである.

クラスタプロトタイプと近傍関係の獲得には, 学習結果を視覚的に確認できる SOM を用いた. 標準的なバッチ学習 SOM を用い, ニューロンのトポロジはノード数 10×10 の正方形格子, 近傍半径は学習と共に減少させた. SOM の学習結果の可視化には, SOM_PAK^{*2} に含まれている U-matrix 表現を用いた. U-matrix 表現は隣接ノード間の距離を濃淡により表現することで, SOM 上のクラスタ構造を可視化する方法である. また, 距離学習に用いた GA は, 次世代の個体選択にはルーレット選択, 遺伝子の交叉には一様交叉を用いた.

まず図 6 に世代毎の最大適合度 (wPA の場合) の変化を示す. 個体数, 突然変異確率, 交

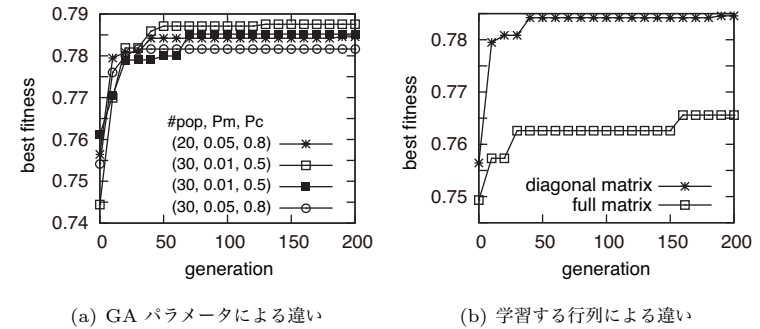


図6 各世代の最大適合度の変化. (a) #pop. は個体数, Pm は突然変異の確率, Pc は交叉確率を表している. (b) M が対角行列の場合と全成分を学習する場合とで比較している.
Fig.6 Change of the best fitness in each generation. (a) #pop. denotes population size, Pm denotes mutation probability, and Pc denotes crossover probability. (b) comparison between when M is diagonal or learn all components.

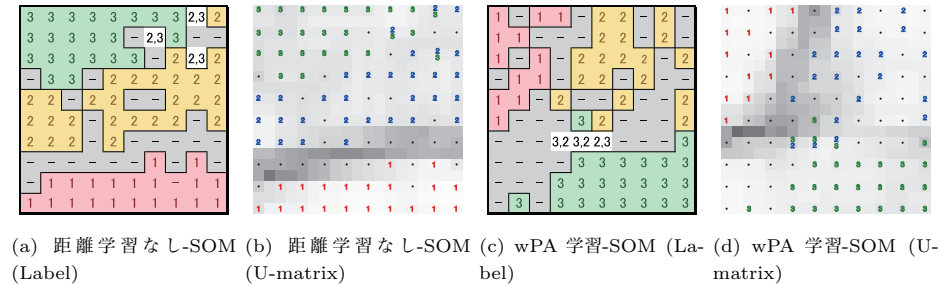


図7 提案距離学習あり/なしでの SOM による可視化結果の比較 (Iris データ使用). (a),(c) 代表クラスによるラベル表示, (b),(d) U-matrix 表示.
Fig.7 Comparison of the visualization results of SOM with/without proposed distance metric learning for Iris data. (a),(c) represented by dominant class, (b),(d) U-matrix representation.

叉確率を変えて実験を行った (図 6(a)). 最大適合度は, いずれも初期世代の 0.745-0.76 から, 70 世代あたりで 0.78-0.785 まで改善している. また, 行列の全成分を学習するよりも対角成分のみ学習 (他の成分は 0) した方が高い学習効果が得られた (図 6(b)). 対角成分数 v に対して全成分での学習は $(v + 1)/2$ 倍の探索空間になるため, 局所解に陥りやすいのが原因と考えられる. 図は wPA の場合であるが, 他の指標についても同様の傾向が確認された.

*1 <http://archive.ics.uci.edu/ml/>
*2 <http://www.cis.hut.fi/research/som-research/nncr-programs.shtml>

次に図 7 に, SOM 学習の可視化結果を示す. 最も違いが現れていた wPA の場合を示す. 代表クラスによるラベル表示では, 距離学習なしの SOM(図 7(a)) よりも wPA を基準に距離学習した SOM(図 7(c)) の方が, 各クラスが密集しクラス間は離れていることから, クラス間をより良く分離できているといえる. また, U-matrix 表現では, 元々, Iris データのクラス 1 は他のクラスと線形分離可能なので, 通常の SOM(図 7(b)) でもクラス 1 との境界はよく現れている. しかし, wPA を基準に距離学習した SOM(図 7(d)) では, クラス 2 と 3 の間にも若干の境界 (濃淡の濃い部分) が見られる. このことから, 距離学習の効果が確認できる.

5. おわりに

本研究では, 大域的クラスタ妥当性指標に基づく距離学習法を提案した. 本手法は, 大域的クラスタ妥当性指数を適合度関数として遺伝的アルゴリズムにより, 最適な変換行列を探索する. 評価実験では, 実データに対してある程度の妥当性指標の向上が確認された. 今後, GA の探索パラメータの調整, 探索性能の向上, 大規模データへの適用, 分類器の精度での比較, 既存の距離学習法との比較などの課題がある.

謝辞 本研究は科学研究費補助金若手研究 (B)(21700165) の支援を受けて行われた.

参考文献

- 1) Amigó, E., Gonzalo, J., Artiles, J. and Verdejo, F.: A comparison of extrinsic clustering evaluation metrics based on formal constraints, *Information Retrieval*, Vol.699, No.12, pp.461–486 (2009).
- 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) Belkin, M. and Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering, *Advances in Neural Processing Systems (NIPS)*, pp.585–591 (2002).
- 4) Davies, D.L. and Bouldin, D.W.: A Cluster Separation Measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol.1, No.4, pp.224–227 (1979).
- 5) Deborah, L.J., Baskaran, R. and Kannan, A.: A Survey on Internal Validity Measure for Cluster Validation, *International Journal of Computer Science & Engineering Survey (IJCSSES)*, Vol.1, No.2, pp.85–102 (2010).
- 6) Dunn, J.C.: Well Separated Clusters and Optimal Fuzzy Partitions, *Journal of*

- Cybernetics*, Vol.4, pp.95–104 (1974).
- 7) Fukui, K. and Numao, M.: Neighborhood-based Smoothing of Externall Cluster Validity Measures, *Proc. the 16th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2012)* (2012). (to appear).
- 8) Halkidi, M. and Vazirgiannis, M.: Clustering validity assessment using multi representatives, *Proc. 2nd Hellenic Conference on Artificial Intelligence*, pp.237–248 (2002).
- 9) Kohonen, T.: *Self-Organizing Maps*, Springer-Verlag (1995).
- 10) Kovács, F., Legány, C. and Babos, A.: Cluster Validity Measurement Techniques, *Engineering*, Vol.2006, pp.388–393 (2006).
- 11) 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. the 17th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-2011)*, pp.868–876 (2011).
- 12) Liu, Y., Li, Z., Xiong, H., Gao, X. and Wu, J.: Understanding of Internal Clustering Validation Measures, *Proc. IEEE International Conference on Data Mining (ICDM-10)*, pp.911–916 (2010).
- 13) Rendón, E., Abundez, I., Arizmendi, A. and Quiroz, E.M.: Internal versus External Cluster Validation Indexes, *International Journal of Computers and Communications*, Vol.5, No.1, pp.27–34 (2011).
- 14) Roweis, S. and Saul, L.: Nonlinear Dimensionality Reduction by Locally Linear Embedding, *Science*, Vol.290, pp.2323–2326 (2000).
- 15) 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).
- 16) Veenhuis, C. and Koppen, M.: *Data Swarm Clustering*, chapter 10, pp.221–241, Springer-Verlag (2006).
- 17) 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).
- 18) 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).
- 19) Xu, R. and Wunsch, D.: *CLUSTER VALIDITY*, Computational Intelligence, chapter 10, pp.263–278, IEEE Press (2008).
- 20) 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).