電子情報通信学会論文誌 2008/8 Vol. J91–D No. 8

確率密度推定に基づく RDSP 法を用いた音素データの階層クラスタ分析

斧城　悠大†（学生員）　岩田　一貴†（正員）
末松　伸朗†（正員）　林　朗†（正員）

Hierarchical Cluster Analysis of Phoneme Data Using RDSP Methods Based on Probability Density Estimation

Yuta ONOSHIRO†, *Student Member*, Kazunori IWATA†, Nobuo SUEMATSU†, and Akira HAYASHI†, *Members*

† 広島市立大学大学院情報科学研究科，広島市
Graduate School of Information Sciences, Hiroshima City University, Hiroshima-shi, 731–3194 Japan

あらまし　標本の母集団が混合モデルで表されるときに有効な階層クラスタ分析の手法として RDSP 法がある．本論文では，ノンパラメトリックな確率密度推定を用いて RDSP 法を拡張する．また，音素データを使った実験において，提案手法の有効性を検証する．

キーワード　階層クラスタ分析，混合モデル，確率密度推定，音素データ

1. まえがき

階層クラスタ分析は，主に二つの目的で使われる．一つはラベル付けと呼ばれるもので，ラベルが未知の標本に対して，標本をいくつかのクラスタ（標本の集合）に分類するために使われる．もう一つの目的は，ラベルの階層構造の同定と呼ばれ，ラベルが既知の標本に対して，クラスタの背後にある階層構造を同定するために使われる．本論文では，ラベルの階層構造の同定のための階層クラスタ分析を取り扱う．RDSP（Redundancy-based DisSimilarity among Probability distributions）[1] はこの目的のために提案されたクラスタ間の非類似度であり，標本の母集団が混合モデルである場合に，ある種の理論的妥当性が示されている．しかし，RDSP は確率密度関数についての測度であるため，その値を計算するためには，クラスタを生成した部分母集団の確率密度関数を推定しなくてはならない．確率密度関数の推定については，[1] で述べられているようなパラメトリック推定による方法とノンパラメトリック推定による方法がある．データ解析においては，ノンパラメトリックな推定に基づく方法を用いた方が確率密度関数を仮定しないで済むので，確率密度関数を柔軟に表すことができるという点で実用的なことが多い．そこで，本論文ではノンパラメトリックな確率密度推定に基づく RDSP 法を提案する．そして，母集団が混合モデルに比較的近いと思われる音素データに対してクラスタの階層構造の推定を行うことで，既存の手法及びパラメトリック推定に基づく RDSP 法と比較した場合の提案手法の有効性を確認する．

2. 確率密度推定に基づく RDSP 法

2.1 混合モデルにおける階層クラスタ分析

標本の母集団が図 1 のような混合モデルに従う場合を考える．混合モデルでは，各離散時間ステップにおいて，ある事前確率 ω により一つの部分母集団が選ばれ，選ばれた部分母集団の確率密度関数に従って標本が生成される．各部分母集団に番号を付け，その番号をラベル番号と呼ぶ．d 次元ユークリッド空間 $\mathbb{R}^d$ 上の時間ステップ $i \in \mathbb{N}$ における確率変数を X_i，その標本を x_i，ラベル番号の全体集合を $\mathcal{L} \triangleq \{1, \ldots, M\}$ とする．標本空間 $\mathbb{R}^d$ 上の部分母集団 $m \in \mathcal{L}$ の確率密度関数を P_m，確率密度関数の集合を $\mathcal{P}(\mathcal{L}) \triangleq \{P_m | m \in \mathcal{L}\}$ とする．また，標本 $x \in \mathbb{R}^d$ が部分母集団 $m \in \mathcal{L}$ の確率分布に従うことを $x \sim P_m$ と表すと，部分母集団の事前確率 ω は $\omega(m) \triangleq \Pr(X_i \sim P_m)$ と表記できる．任意の $n \in \mathbb{N}$ に対して，母集団から出力された標本の集合を $x^n \triangleq (x_1, ..., x_n) \in \mathbb{R}^{dn}$ と表す．

混合モデルにおいて，同じ部分母集団から生成されたクラスタに対する階層クラスタ分析を考える．任意の $m \in \mathcal{L}$ に対して，クラスタ $\boldsymbol{x}^{(m)} \triangleq \{x_i \in \mathbb{R}^d \,|\, x_i \sim P_m, i = 1, \ldots, n\}$ をクラスタ m と呼ぶ．また，任意の $\underline{\mathcal{L}} \subseteq \mathcal{L}$ に対して，$\boldsymbol{x}^{(\underline{\mathcal{L}})} \triangleq \{x_i \in \mathbb{R}^d \,|\, x_i \sim P_m \in \mathcal{P}(\underline{\mathcal{L}}), i = 1, \ldots, n\}$ をクラスタ $\underline{\mathcal{L}}$ と呼ぶ．すなわち，クラスタ $\underline{\mathcal{L}}$ は $\underline{\mathcal{L}}$ に含まれるラベル番号をもつクラスタが統合されたものである．以後では，便宜上，上記の m や $\underline{\mathcal{L}}$ をクラスタラベルと呼ぶことがある．ラベルの階層構造を同定するためのクラスタ分析は，一般に表 1 のような手順で行われる．各手法で異なるのは手順 (1) の非類似度 d のみである．

2.2 RDSP 法

混合モデルに対する階層クラスタ分析において，複

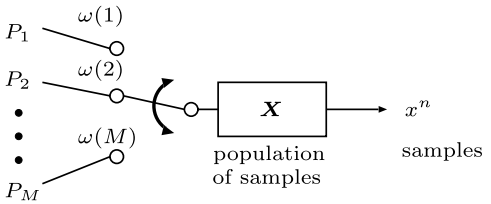


図 1　混合モデルに基づく母集団
Fig. 1　Population based on mixture models.

表 1　階層クラスタ分析の手順
Table 1　Procedures in hierarchical cluster analysis.

（1） クラスタ間の非類似度 d を設定し，ラベル番号の集合 $\mathcal{S}$ を $\mathcal{L}$ の各要素一つのみを含む集合の集合とする．
（2） 集合 $\mathcal{S}$ の要素数が 1 になるまで次のステップを繰り返す．
- 式 (1) を満たすクラスタの対 $\underline{\mathcal{L}}_1, \underline{\mathcal{L}}_2 \in \mathcal{S}$ ($\underline{\mathcal{L}}_1, \underline{\mathcal{L}}_2 \subset \mathcal{L}$) を一つのクラスタに統合し，統合したクラスタのラベルを $\underline{\mathcal{L}}_3 = \underline{\mathcal{L}}_1 \cup \underline{\mathcal{L}}_2$ と表す．

$$\left(\underline{\mathcal{L}}_1, \underline{\mathcal{L}}_2\right) = \underset{\underline{\mathcal{L}}', \underline{\mathcal{L}}'' \in \mathcal{S}:\, \underline{\mathcal{L}}' \neq \underline{\mathcal{L}}''}{\operatorname{argmin}} d(\underline{\mathcal{L}}', \underline{\mathcal{L}}''). \tag{1}$$

- $\mathcal{S} \leftarrow \mathcal{S} - \underline{\mathcal{L}}_1 - \underline{\mathcal{L}}_2 + \underline{\mathcal{L}}_3$

（3） クラスタの統合過程を示すデンドログラムを作成し，縦軸の適当な位置で切断することにより，クラスタを分類する．

数のクラスタ間の階層構造を適当に測るための非類似度として，RDSP が提案されている [1].

［定義 1］（RDSP [1]） 任意の部分集合 $\underline{\mathcal{L}} \subseteq \mathcal{L}$ に対して，複数の確率密度関数 $\mathcal{P}(\underline{\mathcal{L}})$ の間の RDSP は

$$\{RDS(\mathcal{P}(\underline{\mathcal{L}}))\}^2 \triangleq \sum_{m \in \underline{\mathcal{L}}} \lambda_{\underline{\mathcal{L}}}(m) E_{P_m}\left[\log \frac{P_m(x)}{Q_{\underline{\mathcal{L}}}(x)}\right], \tag{2}$$

と定義される．ただし，$E_{P_m}[\cdot]$ は P_m についての期待値を表し，$Q_{\underline{\mathcal{L}}}$ は次のように定義される．

$$Q_{\underline{\mathcal{L}}}(x) \triangleq \sum_{m \in \underline{\mathcal{L}}} \lambda_{\underline{\mathcal{L}}}(m) P_m(x), \tag{3}$$

$$\lambda_{\underline{\mathcal{L}}}(m) \triangleq \frac{\omega(m)}{\sum_{m \in \underline{\mathcal{L}}} \omega(m)}. \tag{4}$$

表 1 の手順 (1) における非類似度 d に RDSP を用いた階層クラスタ分析の手法を RDSP 法と呼ぶ．RDSP 法では，任意の $\underline{\mathcal{L}}'$, $\underline{\mathcal{L}}'' \subset \mathcal{L}$ に対して，RDSP を

$$d_R(\underline{\mathcal{L}}', \underline{\mathcal{L}}'') = \left\{RDS(\mathcal{P}(\underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''))\right\}^2, \tag{5}$$

$$\approx \frac{1}{n_{\underline{\mathcal{L}}'} + n_{\underline{\mathcal{L}}''}} \times \sum_{m \in \underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''} \left| \sum_{x \in \boldsymbol{x}^{(m)}} \log \frac{P_m(x)}{Q_{\underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''}(x)} \right|, \tag{6}$$

と計算する．ただし，任意の $\underline{\mathcal{L}} \subseteq \mathcal{L}$ に対して，$n_{\underline{\mathcal{L}}} \triangleq |\boldsymbol{x}^{(\underline{\mathcal{L}})}|$ とする．式 (6) で近似するのは，期待値に関する積分計算を避けるためである [1]．統合されたクラスタ間の RDSP を計算する際には，統合されている個々のクラスタの確率密度関数 P_m を使うことに注意されたい．例えば，$\underline{\mathcal{L}}' = \{1, 2\}$，$\underline{\mathcal{L}}'' = \{3, 4\}$ のとき，式 (5) の右辺は $RDS(P_1, P_2, P_3, P_4)^2$ と計算される．

2.3　ノンパラメトリックな確率密度推定

RDSP を計算する際には，各部分母集団の確率密度関数 P_m を推定する必要がある．ここでは，標本の集合 $x^n = (x_1, ..., x_n) \in \mathbb{R}^{dn}$ を使って P_m を

$$\hat{P}_m(x; \mathbf{H}_m) \triangleq \frac{1}{n_m} \sum_{i=1}^{n} I_{x_i \sim P_m} \times |\mathbf{H}_m|^{-\frac{1}{2}} K\left(\mathbf{H}_m^{-\frac{1}{2}}(x - x_i)\right), \tag{7}$$

により推定する．ただし，$n_m \triangleq |\boldsymbol{x}^{(m)}|$，$K$ は多変量カーネル関数，$\mathbf{H}_m$ はバンド幅行列と呼ばれるカーネル関数のパラメータで $d \times d$ 正定値実対称行列である．また，I_C は条件 C が真のとき 1，それ以外のときは 0 となる指示関数である．確率密度推定の精度は主にバンド幅行列の選択によって決まることが知られている [2]．本論文では，Zhang ら [2] が提案したマルコフ連鎖モンテカルロ法に基づくバンド幅行列の選択方法を使う．これは，目標とする確率密度関数 P_m と推定した $\hat{P}_m$ との Kullback-Leibler ダイバージェンスを最小とするバンド幅行列

$$\mathbf{H}_m^* \triangleq \underset{\mathbf{H}_m}{\operatorname{argmin}} E_{P_m}\left[\log \frac{P_m(x)}{\hat{P}_m(x; \mathbf{H}_m)}\right], \tag{8}$$

を求める方法である．具体的には，式 (7) のバンド幅行列 $\mathbf{H}_m$ を下三角行列 $\mathbf{B}_m^{-1}$ により $\mathbf{H}_m = \left(\mathbf{B}_m^{-1}\right)\left(\mathbf{B}_m^{-1}\right)^T$ と Cholesky 分解し，式 (7) を $\mathbf{B}_m^{-1}$ について書き直した一つ抜き推定量

$$\hat{P}_{m,i}(x_i; \mathbf{B}_m^{-1}) \triangleq \frac{1}{n_m - 1} \times \sum_{j=1}^{n} I_{j \neq i} I_{x_j \sim P_m} |\mathbf{B}_m| K\left(\mathbf{B}_m(x_i - x_j)\right), \tag{9}$$

を求め，その対数ゆう度

$$L_m(x^n; \mathbf{B}_m^{-1}) \triangleq \sum_{i=1}^{n} I_{x_i \sim P_m} \log \hat{P}_{m,i}(x_i; \mathbf{B}_m^{-1}), \tag{10}$$

を最大化する $\left(\mathbf{B}_m^{-1}\right)^*$ により $\mathbf{H}_m^*$ を求める．$\left(\mathbf{B}_m^{-1}\right)^*$ は次の事後分布 β から，Metropolis-Hastings（M-H）アルゴリズム [3] によってサンプリングされた $\mathbf{B}_m^{-1}$ の標本平均で近似される．

$$\beta(\mathbf{B}_m^{-1}|x^n) \triangleq \left(\prod_{k=1}^{d}\prod_{l=1}^{k}\alpha(b_{kl};\gamma)\right)\left(\prod_{i=1}^{n}\hat{P}_{m,i}(x_i;\mathbf{B}_m^{-1})\right). \tag{11}$$

ただし，α は事前分布を表し，$\alpha(b_{kl};\gamma) \triangleq 1/(1+\gamma b_{kl}^2)$ とする．ここで，b_{kl} は $\mathbf{B}_m^{-1}$ の (k,l) 要素，γ は事前分布 α のハイパパラメータである．

3. 実　　験

実データを用いた階層クラスタ分析において，確率密度推定に基づく RDSP 法の有効性を検証するために，音素データの階層クラスタ分析を行った．

3.1 音素データ

男性 437 人からサンプリングした 4509 個の音素データ [4] に対して階層クラスタ分析を行った．データの特徴には，音素データから生成した長さ 256 の対数ピリオドグラム [4] に対する主成分分析により得られた第 1 から第 10 主成分を用いた（つまり，$d = 10$）．音素データは母音と子音に大別され，一般に母音は 3 種類の音素（1: aa，2: ao，4: iy），子音は 2 種類の音素（3: dcl，5: sh）に分類される [5]．これは一般に母音間及び子音間の差が母音と子音の差よりも小さいためである [6]．更に，各種類の音素データは，発音した話者の方言（A: New England，B: Northern，C: North Midland，D: South Midland，E: Southern，F: New York City，G: Western，H: Army Brat）に従って 8 種類に細分類される．これは方言の違いによる差が音素の違いによる差よりも小さいためである．階層クラスタ分析の目標は，各方言によって細分類された標本の集合をクラスタとし（$M = 5 \times 8 = 40$），図 2 に示すようなクラスタ間の階層構造を同定することである．クラスタ分析に用いた手法は **3.2** で説明する最短距離法，Ward 法 [7]，RDSP 法（パラメトリック）[1]，RDSP 法（ノンパラメトリック）の 4 種類である．RDSP 法（パラメトリック）と RDSP 法（ノンパラメトリック）の間で異なる点は，確率密度関数 P_m の推定方法のみである．そのため，これらの手法の実験結果の違いは，P_m の推定結果にのみ依存することに注意されたい．

3.2 階層クラスタ分析の手法

階層クラスタ分析の各手法における非類似度は次のとおりである．

（a） 最短距離法

任意の $\underline{\mathcal{L}}'$，$\underline{\mathcal{L}}'' \subset \mathcal{L}$ に対して，非類似度は

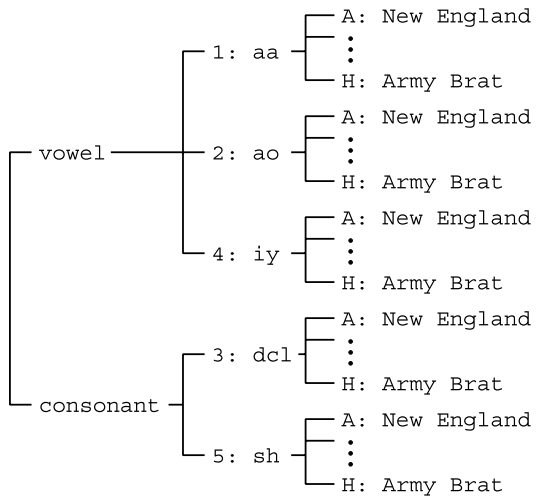


図 2　音素データの階層構造
Fig. 2　The hierarchical structure of phoneme data.

$$d_N(\underline{\mathcal{L}}', \underline{\mathcal{L}}'') = \min_{x' \in \boldsymbol{x}^{(\underline{\mathcal{L}}')},\ x'' \in \boldsymbol{x}^{(\underline{\mathcal{L}}'')}} \|x' - x''\|, \tag{12}$$

で与えられる．ただし，$\|\cdot\|$ はベクトルのノルムを表す．

（b） Ward 法

任意の $\underline{\mathcal{L}}'$，$\underline{\mathcal{L}}'' \subset \mathcal{L}$ に対して，非類似度は

$$d_W(\underline{\mathcal{L}}', \underline{\mathcal{L}}'') = \frac{n_{\underline{\mathcal{L}}'} n_{\underline{\mathcal{L}}''}}{n_{\underline{\mathcal{L}}'} + n_{\underline{\mathcal{L}}''}} \|\bar{x}^{(\underline{\mathcal{L}}')} - \bar{x}^{(\underline{\mathcal{L}}'')}\|^2, \tag{13}$$

で与えられる．ただし，任意の $\underline{\mathcal{L}} \subseteq \mathcal{L}$ に対して，$\bar{x}^{(\underline{\mathcal{L}})}$ は $\boldsymbol{x}^{(\underline{\mathcal{L}})}$ の平均ベクトルを表す．

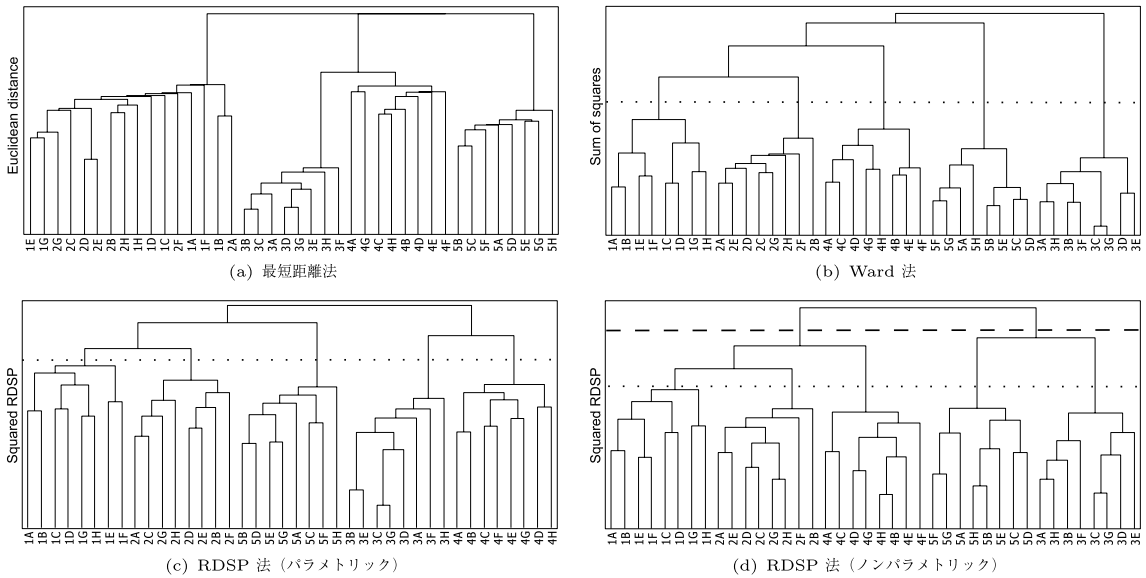
（c） RDSP 法（パラメトリック）

任意の $\underline{\mathcal{L}}'$，$\underline{\mathcal{L}}'' \subset \mathcal{L}$ に対して，非類似度は式 (6) の各部分母集団の確率密度関数 P_m をパラメトリックな正規分布 $\hat{N}_m$ に置き換えた

$$d_R(\underline{\mathcal{L}}', \underline{\mathcal{L}}'') \approx \frac{1}{n_{\underline{\mathcal{L}}'} + n_{\underline{\mathcal{L}}''}} \times \sum_{m \in \underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''} \left| \sum_{x \in \boldsymbol{x}^{(m)}} \log \frac{\hat{N}_m(x; \hat{\boldsymbol{\mu}}_m, \hat{\boldsymbol{\Sigma}}_m)}{\hat{Q}_{\underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''}(x; \hat{\boldsymbol{\mu}}_{\underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''}, \hat{\boldsymbol{\Sigma}}_{\underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''})} \right|, \tag{14}$$

で与えられる [1]．ここで，式 (14) の $\hat{Q}_{\underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''}$ は

$$\hat{Q}_{\underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''}(x; \hat{\boldsymbol{\mu}}_{\underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''}, \hat{\boldsymbol{\Sigma}}_{\underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''}) \triangleq \sum_{m \in \underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''} \hat{\lambda}_{\underline{\mathcal{L}}' \cup \underline{\mathcal{L}}''}(m) N_m(x; \hat{\boldsymbol{\mu}}_m, \hat{\boldsymbol{\Sigma}}_m), \tag{15}$$



(a) 最短距離法　(b) Ward 法

(c) RDSP 法（パラメトリック）　(d) RDSP 法（ノンパラメトリック）

図 3　各手法における音素データのデンドログラム

Fig. 3　Dendrograms for each method for phoneme data. (a) Nearest neighbor method, (b) Ward method, (c) RDSP method with parametric estimation, (d) RDSP method with nonparametric estimation.

を表し，$\hat{\lambda}_{\underline{\mathcal{L}}'\cup\underline{\mathcal{L}}''}(m) \triangleq n_m / \sum_{m\in\underline{\mathcal{L}}'\cup\underline{\mathcal{L}}''} n_m$ である．各 $m \in \{1A,\ldots,5H\}$ に対して，平均 $\hat{\boldsymbol{\mu}}_m$ 及び共分散行列 $\hat{\boldsymbol{\Sigma}}_m$ は，それぞれクラスタ m の標本平均及び共分散行列とした．

(d)　RDSP 法（ノンパラメトリック）

任意の $\underline{\mathcal{L}}'$, $\underline{\mathcal{L}}'' \subset \mathcal{L}$ に対して，非類似度は式 (6) の各部分母集団の確率密度関数 P_m をノンパラメトリックに確率密度推定した $\hat{P}_m$ に置き換えた

$$d_R(\underline{\mathcal{L}}',\underline{\mathcal{L}}'') \approx \frac{1}{n_{\underline{\mathcal{L}}'}+n_{\underline{\mathcal{L}}''}} \times \sum_{m\in\underline{\mathcal{L}}'\cup\underline{\mathcal{L}}''} \left| \sum_{x\in\boldsymbol{x}^{(m)}} \log \frac{\hat{P}_m(x;\mathbf{H}_m)}{\hat{Q}_{\underline{\mathcal{L}}'\cup\underline{\mathcal{L}}''}(x;\mathbf{H}_{\underline{\mathcal{L}}'\cup\underline{\mathcal{L}}''})} \right|, \tag{16}$$

で与えられる．ここで，式 (16) の $\hat{Q}_{\underline{\mathcal{L}}'\cup\underline{\mathcal{L}}''}$ は

$$\hat{Q}_{\underline{\mathcal{L}}'\cup\underline{\mathcal{L}}''}(x;\mathbf{H}_{\underline{\mathcal{L}}'\cup\underline{\mathcal{L}}''}) \triangleq \sum_{m\in\underline{\mathcal{L}}'\cup\underline{\mathcal{L}}''} \hat{\lambda}_{\underline{\mathcal{L}}'\cup\underline{\mathcal{L}}''}(m)\hat{P}_m(x;\mathbf{H}_m), \tag{17}$$

を表す．**2.3** で述べたように，各バンド幅行列は M-H アルゴリズムを使って計算した．ただし，多変量カーネル関数 K は十次元標準正規分布，M-H アルゴリズムにおける反復回数を 10000 回，burn-in 期間を 2500 回，$\mathbf{B}_m^{-1}$ の初期値を単位行列，$\gamma = 1$，proposal 分布を 55 次元正規分布とした(注1)．proposal 分布の共分散行列は，サンプリングの受理確率が 0.2〜0.3 になるように調節した．

3.3　実 験 結 果

階層クラスタ分析の結果を図 3 (a)〜(d) に示す．図 3 において，横軸の文字は各クラスタを表し，1 けた目は音素を示す数字，2 けた目は話者の方言を示すアルファベットを表す．点線及び破線は，それぞれ 5 種類の音素及び母音と子音を分類できる線である．OS が Windows Vista，CPU が Intel Xeon 2.66 GHz (2 プロセッサ)，メモリが 2 GByte の計算機において，統計解析ソフト R を用いて実験を行った結果，各手法の実行時間は，最短距離法が 100.91 s，Ward 法が 0.22 s，RDSP 法（パラメトリック）が 63.32 s，RDSP 法（ノンパラメトリック）がバンド幅行列の推定に 74799.39 s，クラスタ分析に 282.44 s の合計 75160.49 s であった．

はじめに，最短距離法との比較から，Ward 法及び RDSP 法（パラメトリック，ノンパラメトリック）の

(注1)：この場合，$\mathbf{B}_m^{-1}$ は 10×10 の実対称行列の下三角行列なので，その要素数は 55 である．よって proposal 分布は 55 次元となる．

表 2　各 m に対する $\sum_{x\in\boldsymbol{x}(m)} \log \hat{N}_m(x;\hat{\boldsymbol{\mu}}_m,\hat{\boldsymbol{\Sigma}}_m)$
Table 2　$\sum_{x\in\boldsymbol{x}(m)} \log \hat{N}_m(x;\hat{\boldsymbol{\mu}}_m,\hat{\boldsymbol{\Sigma}}_m)$ for each m.

	1	2	3	4	5
A	−1592	−2390	−1453	−2265	−1776
B	−3940	−5164	−3430	−6412	−4041
C	−3812	−6335	−3769	−6981	−4514
D	−2877	−6098	−3608	−6337	−3977
E	−3071	−4789	−3321	−5867	−3570
F	−1712	−1946	−1369	−2388	−1709
G	−4415	−5126	−3667	−5702	−4274
H	−1352	−1453	−1086	−2063	−1201

表 3　各 m に対する $\sum_{x\in\boldsymbol{x}(m)} \log \hat{P}_m(x;\mathbf{H}_m)$
Table 3　$\sum_{x\in\boldsymbol{x}(m)} \log \hat{P}_m(x;\mathbf{H}_m)$ for each m.

	1	2	3	4	5
A	−752	−1199	−561	−1117	−787
B	−1952	−2715	−1535	−3479	−1965
C	−1944	−3338	−1659	−3865	−2101
D	−1400	−3138	−1532	−3453	−1837
E	−1487	−2499	−1467	−3264	−1714
F	−858	−959	−569	−1247	−700
G	−2247	−2718	−1680	−3018	−2041
H	−658	−683	−456	−1039	−509

有効性を確認する．最短距離法は，図 3 (a) に示すように，鎖状効果を引き起こしてしまった結果，5 種類の音素の分類に失敗している．一方，図 3 (b)〜(d) に示すように，その他の手法は鎖状効果を引き起こすことなく音素の分類に成功している．よって，Ward 法及び RDSP 法は鎖状効果を引き起こしにくいことが分かる．

次に，Ward 法及び RDSP 法（パラメトリック）との比較から，RDSP 法（ノンパラメトリック）の有効性を確認する．母音と子音の分類については，図 3 (b)〜(d) に示すように，RDSP 法（ノンパラメトリック）のみが分類に成功し，その他の手法は失敗している．RDSP 法（ノンパラメトリック）が Ward 法に比べて良い理由は，Ward 法がクラスタ間の非類似度にクラスタの平均と平方和のみしか使わないのに対し，RDSP 法はクラスタ間の非類似度にクラスタの確率密度関数そのものを用いるからである（式 (13) と (16) を参照）．また，RDSP 法（ノンパラメトリック）が RDSP 法（パラメトリック）に比べて良い理由は，データ数が十分にあれば，クラスタの確率密度関数についての仮定がない分，ノンパラメトリック推定の方がパラメトリック推定よりも確率密度関数を柔軟に表すことができるので，クラスタの確率密度関数を精度良く推定できるからである．表 2，表 3 に，各 $m \in \{1A,\ldots,5H\}$ に対する $\hat{N}_m$ と $\hat{P}_m$ のクラスタ m についての対数ゆう度を示す．表中の数字 1〜5 は音素を表し，アルファベット A〜H は方言を表す．表から，実際に $\hat{P}_m$ の方が $\hat{N}_m$ より対数ゆう度という意味で精度良く確率密度関数 P_m を推定していることが確認できる．

4. むすび

ノンパラメトリックな確率密度推定に基づく RDSP 法を提案した．また，音素データの階層クラスタ分析において，既存の手法と比較した場合の提案手法の有効性を示した．

文　献

[1] K. Iwata and A. Hayashi, “A redundancy-based measure of dissimilarity among probability distributions for hierarchical clustering criteria,” IEEE Trans. Pattern Anal. Mach. Intell., vol.30, no.1, pp.76–88, 2008.

[2] X. Zhang, M.L. King, and R.J. Hyndman, “A bayesian approach to bandwidth selection for multivariate kernel density estimation,” Computational Statistics and Data Analysis, vol.50, pp.3009–3031, 2006.

[3] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, “Equations of state calculations by fast computing machines,” J. Chemical Physics, vol.21, pp.1087–1091, 1953.

[4] T. Hastie, R. Tibshirani, and J. Friednam, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer-Verlag, New York, 2001.

[5] T.J. Hazen, K. Saenko, C. La, and J.R. Glass, “A segment-based audio-visual speech recognizer: Data collection, development, and initial experiments,” Proc. 6th International Conf. on Multimodal Interfaces, pp.235–242, State College, PA, USA, Oct. 2004.

[6] P.B. de Mareüil, C. Corredor-Ardoy, and M. Adda-Decker, “Multi-lingual automatic phoneme clustering,” Proc. 14th International Congress of Phonetic Sciences, pp.1209–1212, San Francisco, USA, 1999.

[7] J.H. Ward, “Hierarchical grouping to optimize an objective function,” J. American Statistical Association, vol.58, no.301, pp.236–244, 1963.

（平成 19 年 12 月 21 日受付，20 年 3 月 22 日再受付）