

固有ベクトルのずれの影響を軽減する固有値修正法

岩村 雅一[†] 大町真一郎[†] 阿曾 弘具[†]

A Modification of Eigenvalues to Compensate Estimation Errors of Eigenvectors

Masakazu IWAMURA[†], Shinichiro OMACHI[†], and Hirotomo ASO[†]

あらまし パターン認識で用いられる統計的識別関数は、パターンの確率分布が正しく与えられたときに最適になる。パターンの確率分布は一般に未知であるため、正規分布を仮定して、その平均ベクトルと共分散行列は学習サンプルからの推定値を用いる。しかし、推定に用いる学習サンプルが不足すると推定値が誤差を含み、認識性能が低下することが知られている。平均ベクトルや共分散行列の固有値の推定誤差についてはよく調査され、認識性能の低下を回避する方法が提案されている。しかし、共分散行列の固有ベクトルの推定誤差についてはこれまでほとんど考慮されてこなかった。本論文では、学習サンプルが十分用意できない場合においても高精度な認識を行うことを目的とし、共分散行列の固有ベクトルの推定誤差を考慮することで、2次識別関数の主要部分であるマハラノビス距離を正しく推定する手法を提案する。提案手法を文字認識に適用した結果、パラメータの推定に用いる学習サンプルが少ない場合でもマハラノビス距離の推定値を正しく推定することができ、認識性能の改善が見られた。

キーワード パターン認識, 識別関数, 固有ベクトル, 推定誤差, マハラノビス距離

1. ま え が き

ベイズの定理から導かれる統計的識別関数は、パターンの確率分布が正しく与えられたときに、誤認識によって生じる損失の期待値を最小に留める最適性を持つ。しかし、実際にはパターンの確率分布が未知であることが多く、パターンの確率分布として特定の分布を仮定し、分布のパラメータを学習用のサンプルから推定することがよく行われる。このとき、用いる学習サンプルが特徴量の次元数に比べて不足すると、パラメータの推定誤差のために認識性能が低下することが知られている。

2次識別関数やマハラノビス距離は、パターンの確率分布を正規分布と仮定して導かれる。正規分布は平均ベクトルと共分散行列をパラメータとして用いる。これらのパラメータを学習サンプルから推定したとき、推定値である標本平均ベクトルと標本共分散行列は誤差を含む。

竹下らは、2次識別関数の値を実質的に決定するマハラノビス距離が平均ベクトルと共分散行列のそれぞ

れの推定誤差に対してどのような影響を受けるか理論的に調査した [1]。その結果、学習サンプル数が 10 を超えるときは、平均ベクトルの推定誤差の影響は共分散行列の推定誤差に比べてほとんど無視できることを導いた。従って、2次識別関数やマハラノビス距離では共分散行列の推定誤差が認識精度を低下させているといえる。

共分散行列の推定誤差については、共分散行列が固有値展開により、その固有値と固有ベクトルに分解できることから、固有値の推定誤差と固有ベクトルの推定誤差に分けて考えられてきた。このうち固有値の推定誤差については、特に小さな固有値の推定誤差が認識性能に悪影響を及ぼすことが知られ、信頼性の低い推定値を用いない手法が提案されている。このような手法には、木村らの修正2次識別関数 (MQDF) [2]、加藤らの改良型マハラノビス距離 (MMD) [3]、孫らの簡素化マハラノビス距離 (SMD) [4]、Friedman の Regularized Discriminant Analysis [5] などがある。

それに対して、信頼性の低い固有値を補正して用いる手法も提案されている。分布のパラメータを補正する手法は、信頼性の低いパラメータを用いない手法に比べて難しい反面、正しく補正が行われれば認識性能がより良くなることが期待できる。また、補正法を

[†] 東北大学大学院工学研究科, 仙台市
Graduate School of Engineering, Tohoku University, Sendai-shi, 980-8579 Japan

導く理論自体がパターン認識に与える貢献も大きい。James と Stein は、標本共分散行列を改良する推定量 (James-Stein 推定量) を導き [6], Takeshita らは実験によって James-Stein 推定量の有効性を確認した [7]. 酒井らは標本共分散行列の固有値の平均的な偏りから標本固有値を補正する手法 (RQDF) [8] を提案し、良好な認識結果を得ている。

一方、固有ベクトルの推定誤差が認識性能に及ぼす影響についてはこれまであまり考慮されてこなかった。酒井らは、標本固有ベクトルの揺らぎのために、パターン認識で推定すべき共分散行列は真の共分散行列ではなく、真の共分散行列をぼかした行列 (ぼけ共分散行列) であると主張し、固有値を二回補正する手法 (WRQDF) を提案している [9]. しかし、何故ぼけ共分散行列を用いるべきかについての理論的根拠は示されていない。また、推定の際に標本固有値でも積分しているので、ぼけ共分散行列が標本固有値の値を反映していない。

本論文では共分散行列の固有ベクトルの推定誤差を考慮することで、学習サンプルが十分用意できない場合においてもマハラノビス距離を正しく求め、高精度な認識を行うことを目的とする。2. では、共分散行列の固有ベクトルの推定誤差によってマハラノビス距離に推定誤差が生じ、この誤差は固有値の誤差と解釈することができることを示す。3. では、固有値を修正することにより、辞書の作成に用いるサンプルが少ないときでも、マハラノビス距離を正しく推定する手法を提案する。4. では、提案手法を文字認識に適用し、マハラノビス距離が正しく推定されること、認識性能が改善されることを示す。

2. 固有ベクトルの推定誤差

2.1 固有ベクトルの推定誤差を表す行列

固有ベクトルの推定誤差を扱うために、固有ベクトルの推定誤差を表す行列を定義する。固有ベクトルは正規直交基底であるから、固有ベクトルの推定誤差は固有ベクトルが作る座標系の回転 (ずれ) と考えることができる。固有ベクトルの推定誤差は真の固有ベクトルが作る座標系と、サンプルから推定された固有ベクトルが作る座標系のずれである。従って、真の固有ベクトルが未知であれば、固有ベクトルの推定誤差を定義できないが、真の固有ベクトルが未知でも固有ベクトルの推定誤差を扱うために、固有ベクトルのずれを表す行列 $\Psi(N_1, N_2)$ を次のように定義する。

$$\Psi(N_1, N_2) = \hat{\Phi}_{N_1}^T \hat{\Phi}_{N_2} \quad (1)$$

$$= \begin{pmatrix} \hat{\phi}_1^{N_1} \cdot \hat{\phi}_1^{N_2} & \dots & \hat{\phi}_1^{N_1} \cdot \hat{\phi}_d^{N_2} \\ \vdots & \ddots & \vdots \\ \hat{\phi}_d^{N_1} \cdot \hat{\phi}_1^{N_2} & \dots & \hat{\phi}_d^{N_1} \cdot \hat{\phi}_d^{N_2} \end{pmatrix} \quad (2)$$

ここで d は特徴量の次元数、 $\hat{\phi}_k^N$ は N 個の学習サンプルから推定された固有ベクトル、 $\hat{\Phi}_N$ は

$$\hat{\Phi}_N = \begin{pmatrix} \hat{\phi}_1^N & \hat{\phi}_2^N & \dots & \hat{\phi}_d^N \end{pmatrix} \quad (3)$$

を表す。 $\Psi(N_1, N_2)$ の対角成分は誤差の大きさを表し、 $|\Psi(N_1, N_2)_{ii}|$ が 1 に近いほど誤差が小さい。また、非対角成分 Ψ_{ij} は固有ベクトル $\hat{\phi}_i^{N_1}$ の $\hat{\phi}_j^{N_2}$ 方向へのずれ、または固有ベクトル $\hat{\phi}_j^{N_2}$ の $\hat{\phi}_i^{N_1}$ 方向へのずれを表し、0 に近いほど誤差が小さい。もし誤差が全く無ければ、 $\Psi(N_1, N_2)$ は単位行列になる。

無限個のサンプルを用いて推定したときに真の値が得られるとすれば、 $\Psi(\infty, N)$ は本来の意味での固有ベクトルの推定誤差となる。なお、本論文での固有値、固有ベクトルは固有値が大きい順に並んでいるものとする。

2.2 固有ベクトルの誤差によるマハラノビス距離の誤差

固有ベクトルが誤差を含むときにマハラノビス距離に生じる誤差を、固有ベクトルのずれを表す行列 Ψ との関係に注目して考察する。

λ_k, ϕ_k を真の共分散行列の第 k 固有値、固有ベクトル、 μ を真の平均ベクトル、 x を未知入力ベクトル、 d を特徴量の次元数とすれば、マハラノビス距離は次式で表される。

$$d(x) = \sum_{k=1}^d \frac{1}{\lambda_k} \{\phi_k \cdot (x - \mu)\}^2 \quad (4)$$

マハラノビス距離は“各固有ベクトル軸 ϕ_k 方向の距離を $\frac{1}{\lambda_k}$ の重みを付けて計算する”と解釈できる。

いま、固有ベクトル ϕ_k が誤差 $\Delta\phi_k$ を含むときを考える。(4) 式の ϕ_k を $\phi_k + \Delta\phi_k$ で置き換えると固有ベクトルが誤差を含むときのマハラノビス距離 $\tilde{d}(x)$ が得られる。

$$\tilde{d}(x) = \sum_{k=1}^d \frac{1}{\lambda_k} \left[\{\phi_k \cdot (x - \mu)\}^2 + \{\Delta\phi_k \cdot (x - \mu)\}^2 + 2 \{\phi_k \cdot (x - \mu)\} \{\Delta\phi_k \cdot (x - \mu)\} \right] \quad (5)$$

さらに $\Delta\phi_k$ を固有ベクトルの線形和であると考えて

$$\Delta\phi_k = \sum_{i=1}^d \eta_{ki} \phi_i \quad (6)$$

とおくとき, (5) 式を変形すれば

$$\begin{aligned} \tilde{d}(\mathbf{x}) = & \sum_{k=1}^d \left\{ \frac{(1+\eta_{kk})^2}{\lambda_k} + \sum_{\substack{l=1 \\ l \neq k}}^d \frac{\eta_{lk}^2}{\lambda_l} \right\} \{\phi_k \cdot (\mathbf{x} - \boldsymbol{\mu})\}^2 \\ & + \sum_{i=1}^d \sum_{\substack{j=1 \\ j \neq i}}^d \left\{ \sum_{l=1}^d \frac{\eta_{li}\eta_{lj}}{\lambda_l} + \frac{2\eta_{ij}}{\lambda_i} \right\} \\ & \cdot \{\phi_i \cdot (\mathbf{x} - \boldsymbol{\mu})\} \{\phi_j \cdot (\mathbf{x} - \boldsymbol{\mu})\} \quad (7) \end{aligned}$$

が得られる. 第 2 項は, \mathbf{x} が d 次元正規分布 $N(\boldsymbol{\mu}, \Phi\Lambda\Phi^T)$ に従うとすれば, $\phi_i \cdot (\mathbf{x} - \boldsymbol{\mu})$ と $\phi_j \cdot (\mathbf{x} - \boldsymbol{\mu})$ は独立に 1 次元正規分布 $N(0, \lambda_i)$ と $N(0, \lambda_j)$ に従い, $\{\phi_i \cdot (\mathbf{x} - \boldsymbol{\mu})\} \{\phi_j \cdot (\mathbf{x} - \boldsymbol{\mu})\}$ の期待値は 0 になる. 従って, 第 2 項の期待値は 0 になる. ただし,

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d) \quad (8)$$

$$\Phi = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_d \end{pmatrix} \quad (9)$$

である. 第 1 項の $\frac{(1+\eta_{kk})^2}{\lambda_k} + \sum_{\substack{l=1 \\ l \neq k}}^d \frac{\eta_{lk}^2}{\lambda_l}$ は本来 $\frac{1}{\lambda_k}$ になるべきところであり, 固有ベクトルの誤差は固有値の見掛け上の大きさに影響を与えているといえる.

固有ベクトルのずれを表す行列 Ψ は η_{ij} を用いると

$$\Psi(\infty, N)_{ij} = \begin{cases} 1 + \eta_{ii}, & i = j \\ \eta_{ji}, & i \neq j \end{cases} \quad (10)$$

となるので, (7) 式の第 1 項は $\Psi(\infty, N)_{ij}$ を用いて

$$\sum_{k=1}^d \left\{ \sum_{l=1}^d \frac{\{\Psi(\infty, N)_{kl}\}^2}{\lambda_l} \right\} \{\phi_k \cdot (\mathbf{x} - \boldsymbol{\mu})\}^2 \quad (11)$$

と書くこともできる.

3. 固有ベクトルの推定誤差の影響を軽減する固有値修正法

酒井らは標本共分散行列の固有値の平均的な偏りから標本固有値の偏りを補正する手法を提案した [8].

同様に固有ベクトルの推定誤差の平均的な偏りから固有ベクトルの推定誤差を補正することは可能であろうか?

固有ベクトルの自由度は次元数 d を用いて $\frac{d(d-1)}{2}$ で表される. 固有値の自由度は d であるので, 高次元では固有ベクトルの自由度が固有値の自由度に比べて非常に大きくなる. 固有ベクトルの推定誤差は確率的にのみ得られるが, 自由度が高ければ各実現値における確率が小さくなるため, 一般に高次元特徴量を用いるパターン認識で固有ベクトルを補正することは現実的でない.

本論文では固有ベクトル自体を補正するのではなく, 固有ベクトルの推定誤差を打ち消すように固有値を修正する手法を提案する. 固有ベクトルが誤差を含んだときのマハラノビス距離は (7) 式で与えられる. 期待値が 0 である第 2 項を無視し, 第 1 項のみに注目すれば, (7) 式は固有値のみが変化したマハラノビス距離とみなすことができる. 従って, 固有値を修正することによって, 固有ベクトルの推定誤差によって生じる認識性能への悪影響を減少することができる.

3.1 提案手法

実際のパターン認識において学習サンプルが N 個あるとき, 以下の手順で固有ベクトルの推定誤差による認識性能への悪影響を軽減する.

(1) 辞書の作成に使用可能な N 個のサンプルから標本平均ベクトルと標本共分散行列を求め, 標本固有値と標本固有ベクトルを得る.

(2) 標本固有値を固有値の推定誤差を補正する固有値補正法 (例えば酒井らの手法 [8], James-Stein 推定量 [6] など) で補正し, 真の固有値を推定する. 補正された第 k 固有値を $\check{\lambda}_k$ とする.

(3) 次式を用いて真の固有値を修正する.

$$\tilde{\lambda}_k = \sum_{j=1}^d \left\{ \check{\Psi}_{jk} \right\}^2 \check{\lambda}_j \quad (12)$$

$\tilde{\lambda}_k$ が本手法で修正された第 k 固有値で, $\check{\lambda}_k$ の代わりに認識に用いる. $\left\{ \check{\Psi}_{jk} \right\}^2$ は (2) 式で定義した行列の成分を二乗して条件付き期待値をとったもので, モンテカルロ法により求める. 詳細は 3.3 で述べる.

3.2 提案手法の導出

(12) 式は以下のように導出する. まず, 真の共分散行列を Σ , N 個の学習サンプルから求めた標本共分散行列を $\hat{\Sigma}$ とおく. 共分散行列は固有値と固有ベク

トルに分解できるので,

$$\Sigma = \Phi \Lambda \Phi^T \quad (13)$$

$$\hat{\Sigma} = \hat{\Phi} \hat{\Lambda} \hat{\Phi}^T \quad (14)$$

となる. ここで $\hat{\Lambda}$, $\hat{\Phi}$ は

$$\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_d) \quad (15)$$

$$\hat{\Phi} = \begin{pmatrix} \hat{\phi}_1 & \hat{\phi}_2 & \dots & \hat{\phi}_d \end{pmatrix} \quad (16)$$

である. λ_k , ϕ_k は真の共分散行列の第 k 固有値, 固有ベクトル, $\hat{\lambda}_k$, $\hat{\phi}_k$ は標本共分散行列の第 k 固有値, 固有ベクトルである.

真の固有値は固有値の推定誤差を補正する固有値補正法 (酒井らの手法 [8], James-Stein 推定量 [6] など) を適用することで, 標本固有値から推定できるとする. このように推定した固有値を

$$\check{\Lambda} = \text{diag}(\check{\lambda}_1, \check{\lambda}_2, \dots, \check{\lambda}_d) \quad (17)$$

とおく.

もし, 真の固有ベクトル Φ が既知であれば真の共分散行列を

$$\Sigma \simeq \Phi \check{\Lambda} \Phi^T \quad (18)$$

のように推定できる. しかし, 真の固有ベクトルを推定することは困難であるので, 標本固有ベクトルを用いて真の共分散行列に近い行列を得ることを考える. すなわち,

$$\Sigma \simeq \hat{\Phi} \tilde{\Lambda} \hat{\Phi}^T \quad (19)$$

を満たすような対角行列

$$\tilde{\Lambda} = \text{diag}(\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_d) \quad (20)$$

を求める. $\tilde{\Lambda}$ は (18) 式, (19) 式から

$$\hat{\Phi} \tilde{\Lambda} \hat{\Phi}^T = \Phi \check{\Lambda} \Phi^T \quad (21)$$

とおくことで,

$$\tilde{\Lambda} = \hat{\Phi}^T \Phi \check{\Lambda} \Phi^T \hat{\Phi} \quad (22)$$

となる. しかし, 真の固有ベクトル Φ は未知であり, このままでは $\tilde{\Lambda}$ を求められない. そこで $\Phi^T \hat{\Phi}$ に注目すると, これは固有ベクトルの推定誤差を意味する. 固有ベクトルの推定誤差は真の固有ベクトルに依存し

ないので, 真の固有ベクトルが未知であっても観測可能である. $\Phi^T \hat{\Phi}$ を (1) 式の定義を用いて書き直すと,

$$\Phi^T \hat{\Phi} = \Psi(\infty, N) \quad (23)$$

となる. Φ を定数, $\hat{\Phi}$ を確率変数と考え, $\Psi(\infty, N)$ は確率変数である. 以後 $\Psi(\infty, N)$ を単に Ψ と書く. (23) 式を (22) 式に代入すると,

$$\tilde{\Lambda} = \Psi^T \check{\Lambda} \Psi \quad (24)$$

となる. Ψ は確率変数であるので, $\tilde{\Lambda}$ も確率変数である. 共分散行列は確率変数ではなく, 定数として求めたいので, 標本固有値が $\hat{\Lambda}$ であるときの固有ベクトルの推定誤差で両辺の条件付き期待値をとる.

$$E[\tilde{\Lambda}] = E[\Psi^T \check{\Lambda} \Psi] \quad (25)$$

(25) 式の第 k 対角要素を取り出すと,

$$E[\check{\lambda}_k] = E\left[\sum_{j=1}^d \{\Psi_{jk}\}^2 \check{\lambda}_j\right] \quad (26)$$

$$= \sum_{j=1}^d E[\{\Psi_{jk}\}^2] \check{\lambda}_j \quad (27)$$

となる. さらに

$$\tilde{\lambda}_k = E[\check{\lambda}_k] \quad (28)$$

$$\{\tilde{\Psi}_{jk}\}^2 = E[\{\Psi_{jk}\}^2] \quad (29)$$

とおき, (27) 式に代入すると (12) 式が得られる.

3.3 モンテカルル口法による提案手法の実現

(25) 式の右辺は次式で与えられる.

$$\int_{\Psi} \Psi^T \check{\Lambda} \Psi p(\Psi | \hat{\Lambda}, \Lambda) d\Psi \quad (30)$$

$p(\Psi | \hat{\Lambda}, \Lambda)$ は固有ベクトルの推定誤差の確率密度関数である. 高次元の場合, この値を厳密に求めるのは困難である. 本論文では (30) 式の値を条件付きモンテカルル口法 [10] を用いて求める.

$\Phi = I$ のときを考えても一般性を失わないことから

$$\hat{\Phi} = \Psi \quad (31)$$

$$\hat{\Sigma} = \Psi \check{\Lambda} \Psi^T \quad (32)$$

とおくと, (30) 式は次のように変形できる.

$$\begin{aligned}
 & \int_{\dot{\Psi}} \dot{\Psi}^T \check{\Lambda} \dot{\Psi} p(\dot{\Psi} | \hat{\Lambda}, \Lambda) d\dot{\Psi} \\
 &= \int_{\dot{\Psi}} \dot{\Psi}^T \check{\Lambda} \dot{\Psi} \frac{p(\dot{\Psi} | \hat{\Lambda}, \Lambda)}{G(\dot{\Psi})} d\dot{\Psi} \int_{\hat{\Lambda}} g(\dot{\Psi}, \hat{\Lambda}) d\hat{\Lambda} \\
 &= \int_{\dot{\Psi} \times \hat{\Lambda}} \dot{\Psi}^T \check{\Lambda} \dot{\Psi} \frac{p(\dot{\Psi} | \hat{\Lambda}, \Lambda) g(\dot{\Psi}, \hat{\Lambda})}{G(\dot{\Psi}) p(\dot{\Sigma} | \Lambda)} \\
 & \quad \cdot p(\dot{\Sigma} | \Lambda) d\dot{\Psi} d\hat{\Lambda} \\
 &= \int_{\dot{\Psi} \times \hat{\Lambda}} \dot{\Psi}^T \check{\Lambda} \dot{\Psi} \frac{p(\dot{\Psi}, \hat{\Lambda} | \Lambda)}{p(\hat{\Lambda} | \Lambda)} \\
 & \quad \cdot \frac{g(\dot{\Psi}, \hat{\Lambda})}{G(\dot{\Psi}) p(\dot{\Sigma} | \Lambda)} p(\dot{\Sigma} | \Lambda) d\dot{\Psi} d\hat{\Lambda} \\
 &= \int_{\dot{\Sigma}} \dot{\Psi}^T \check{\Lambda} \dot{\Psi} \frac{p(\dot{\Psi} \hat{\Lambda} \dot{\Psi}^T | \Lambda)}{p(\hat{\Lambda} | \Lambda) J(\dot{\Psi}, \hat{\Lambda})} \\
 & \quad \cdot \frac{J(\dot{\Psi}, \hat{\Lambda})}{p(\dot{\Psi} \hat{\Lambda} \dot{\Psi}^T | \Lambda)} \frac{g(\dot{\Psi}, \hat{\Lambda})}{G(\dot{\Psi})} p(\dot{\Sigma} | \Lambda) d\dot{\Sigma} \\
 &= \int_{\dot{\Sigma}} \dot{\Psi}^T \check{\Lambda} \dot{\Psi} w(\dot{\Sigma}) p(\dot{\Sigma} | \Lambda) d\dot{\Sigma} \tag{33}
 \end{aligned}$$

ただし,

$$G(\dot{\Psi}) = \int_{\hat{\Lambda}} g(\dot{\Psi}, \hat{\Lambda}) d\hat{\Lambda} \tag{34}$$

$$J(\dot{\Psi}, \hat{\Lambda}) = \frac{d\dot{\Psi} d\hat{\Lambda}}{d\dot{\Sigma}} \tag{35}$$

$$\begin{aligned}
 w(\dot{\Sigma}) &= \frac{p(\dot{\Psi} \hat{\Lambda} \dot{\Psi}^T | \Lambda)}{J(\dot{\Psi}, \hat{\Lambda})} \frac{J(\dot{\Psi}, \hat{\Lambda})}{p(\dot{\Psi} \hat{\Lambda} \dot{\Psi}^T | \Lambda)} \\
 & \quad \cdot \frac{g(\dot{\Psi}, \hat{\Lambda})}{p(\hat{\Lambda} | \Lambda) G(\dot{\Psi})} \tag{36}
 \end{aligned}$$

である。つまり、確率密度 $p(\dot{\Psi} | \hat{\Lambda}, \Lambda)$ を持つ固有ベクトルについて $\dot{\Psi} \check{\Lambda} \dot{\Psi}^T$ の期待値を取ることは、確率密度 $p(\dot{\Sigma} | \Lambda)$ を持つ共分散行列について $\dot{\Psi} \check{\Lambda} \dot{\Psi}^T w(\dot{\Sigma})$ の期待値を取ることに等しい。

$w(\dot{\Sigma})$ は Wishart 分布から導かれ、

$$\begin{aligned}
 w(\dot{\Sigma}) &= \frac{|\hat{\Lambda}|^{\frac{1}{2}(n-p-2)} \prod_{i < j} (\hat{\lambda}_i - \hat{\lambda}_j)}{|\dot{\Lambda}|^{\frac{1}{2}(n-p-2)} \prod_{i < j} (\dot{\lambda}_i - \dot{\lambda}_j)} \\
 & \quad \cdot \frac{\exp\left(-\frac{n-1}{2} \text{tr} \Lambda^{-1} \dot{\Psi} \hat{\Lambda} \dot{\Psi}^T\right)}{\exp\left(-\frac{n-1}{2} \text{tr} \Lambda^{-1} \dot{\Psi} \dot{\Lambda} \dot{\Psi}^T\right)} \\
 & \quad \cdot \frac{g(\dot{\Psi}, \hat{\Lambda})}{p(\hat{\Lambda} | \Lambda) G(\dot{\Psi})} \tag{37}
 \end{aligned}$$

となる (付録 1.)。 $g(\dot{\Psi}, \hat{\Lambda})$ は任意の関数で、選び方を工夫すれば期待値の分散を小さくすることができる

Algorithm 1 $\check{\lambda}_k$ の推定

- 1: 乱数を用いて正規分布 $N(0, \check{\Lambda})$ に従う Nt 個の人工サンプル X_1, \dots, X_{Nt} を作成する。 $\check{\Lambda}$ は真の固有値、もしくはその推定値である。
- 2: for $i = 1$ to t do
- 3: N 個の人工サンプル $X_{N(i-1)+1}, \dots, X_{Ni}$ から標本共分散行列 $\dot{\Sigma}_i$ を推定し、標本固有値 $\hat{\Lambda}_i$ と標本固有ベクトル $\hat{\Phi}_i$ を得る。ここで固有ベクトルの推定誤差 Ψ_i は $\Psi_i = \hat{\Phi}_i$ である。
- 4: $w(\Psi_i \hat{\Lambda}_i \Psi_i^T)$ を求める。
- 5: end for
- 6: 次式を用いて $\{\check{\Psi}_{jk}\}^2$ を推定する。

$$\{\check{\Psi}_{jk}\}^2 = \frac{\sum_{i=1}^t \{\Psi_{i,jk}\}^2 w(\Psi_i \hat{\Lambda}_i \Psi_i^T)}{\sum_{i=1}^t w(\Psi_i \hat{\Lambda}_i \Psi_i^T)} \tag{39}$$

- 7: (12) 式で $\check{\lambda}_k$ を得る。

が、本論文では $g(\dot{\Psi}, \hat{\Lambda})$ を定数とした。従って $G(\dot{\Psi})$ も定数となる。 $p(\hat{\Lambda} | \Lambda)$ は解析的に得られるが、特に高次元の場合、数値計算を行うのは困難である。しかし、固有ベクトルの推定誤差は正規性を持つので、

$$\int_{\dot{\Sigma}} w(\dot{\Sigma}) p(\dot{\Sigma} | \Lambda) d\dot{\Sigma} = 1 \tag{38}$$

を満たすように定めればよい (付録 2.)。

(33) 式に基づいて (25) 式を求める手順を Algorithm1 に示す。 N は学習に使用できるサンプル数、 t は固有ベクトルの推定誤差を求める回数である。

4. 提案手法の性能評価

4.1 認識実験

提案手法の有効性を確認するために文字認識実験を行う。実験は固有値だけが異なる 3 種類の辞書を作成し、それぞれの辞書を用いたときの認識率を比較する。

3 種類の辞書は、固有値として標本固有値を用いる “Control”，真の固有値を用いる “True eigenvalue”，提案手法で修正した固有値を用いる “Proposed method” である。全ての辞書で共通の標本平均ベクトル、標本固有ベクトルを用いる。

辞書を推定するのに用いるサンプルは、実際の文字画像と正規分布に従う人工サンプルの二種類を用いる。文字画像の特徴量は NIST Special Database 19 の数字サンプルを 64×64 の大きさに非線形正規化 [11] した後、196 次元の方向線素特徴量 [12] を抽出したものをを用いる。人工サンプルは平均ベクトル μ_0 と共分散

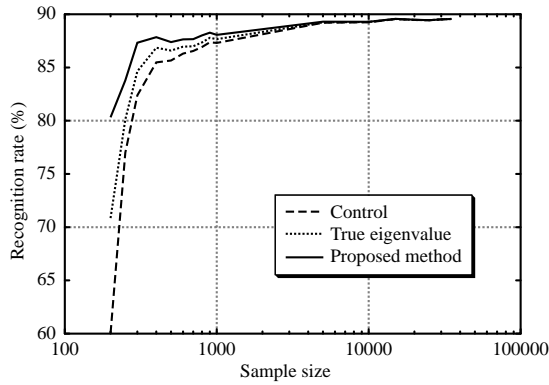


図 1 認識実験の結果 (文字画像)
Fig. 1 Result of recognition experiment.

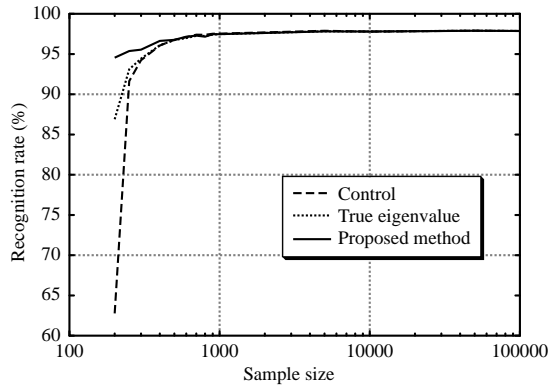


図 2 認識実験の結果 (人工サンプル)
Fig. 2 Result of recognition experiment.

行列 Σ_0 を定め、正規分布 $N(\mu_0, \Sigma_0)$ に従うように乱数を用いて作成する。 μ_0 と Σ_0 はクラス毎に文字画像の特徴量 36,000 文字分から推定した値を用いる。

辞書の作成に用いる学習サンプル数を文字画像の特徴量では 200 個から 35,000 個まで、人工サンプルでは 200 個から 100,000 個まで変化させる。テストサンプルは 1,000 個用いる。実際の文字画像では真の固有値が得られないため、酒井らの手法 [8] で補正した固有値を用いる。識別関数はマハラノビス距離を使用し、Algorithm1 のパラメータ t を 10,000 とした。

実際の文字画像を用いた認識実験の結果を図 1、人工サンプルを用いたときの結果を図 2 に示す。実際の文字画像、人工サンプル共に、“Proposed method” の認識率は全ての学習サンプル数で “True eigenvalue” より高く、提案手法の有効性が確認できた。“True eigenvalue” と “Proposed method” の認識率の差を

学習サンプル数との関係を考慮して見てみると、学習サンプル数が多いときには認識率にほとんど差はないが、少なくなるほど認識率の差は大きくなる。これは学習サンプル数が多いときには、補正されるべき推定誤差がもともと小さいためであると考えられる。また、真の固有値を用いた “True eigenvalue” (文字画像の特徴量では真の値に近付けた固有値) は、標本固有値をそのまま用いた “Control” よりも高い認識率が得られた。

4.2 マハラノビス距離の誤差に関する考察

誤差を含む固有ベクトルと提案手法で修正した固有値を用いるとき、行列表現のマハラノビス距離は (25) 式から次式で与えられる。

$$\begin{aligned} \tilde{d}(x) &= (x - \mu)^T \hat{\Phi} E [\hat{\Lambda}]^{-1} \hat{\Phi}^T (x - \mu) \\ &= (x - \mu)^T \hat{\Phi} E [\Psi^T \check{\Lambda} \Psi]^{-1} \hat{\Phi}^T (x - \mu) \\ &\sim (x - \mu)^T \Phi \check{\Lambda}^{-1} \Phi^T (x - \mu) \quad (40) \end{aligned}$$

従って、提案手法を用いることでマハラノビス距離を正しく求められることが期待される。提案手法を用いる場合、真の固有値を用いるよりもマハラノビス距離が正しく推定できることを実験により確認する。

実験のために 3 種類の辞書を用意する。“True eigenvalue” は真の平均ベクトル、真の固有値、標本固有ベクトルを用い、“Proposed method” は真の平均ベクトル、真の固有値を提案手法で修正した固有値、標本固有ベクトルを用い、“True” は真の平均ベクトル、真の固有値、真の固有ベクトルを用いる。

辞書と同じクラスに属するサンプルを用いて、“True” のマハラノビス距離に対する “True eigenvalue” のマハラノビス距離の比を求め、平均値を取る。“Proposed method” についても同様に求め、これらと比較する。

辞書の推定に用いる学習サンプル数を 200 個から 100,000 個まで変化させ、テストサンプルは 1,000 個ずつ用いる。特徴量は 4.1 で使用した人工サンプルの特徴量を用い、Algorithm1 のパラメータ t を 10,000 とした。

“True” のマハラノビス距離に対する “True eigenvalue”、“Proposed method” のマハラノビス距離の比の平均値を図 3 に示す。“True eigenvalue” では辞書の推定に用いる学習サンプル数が少ない程、マハラノビス距離の推定値が大きく偏っているが、“Proposed method” では学習サンプル数に依らずほぼ正しく推

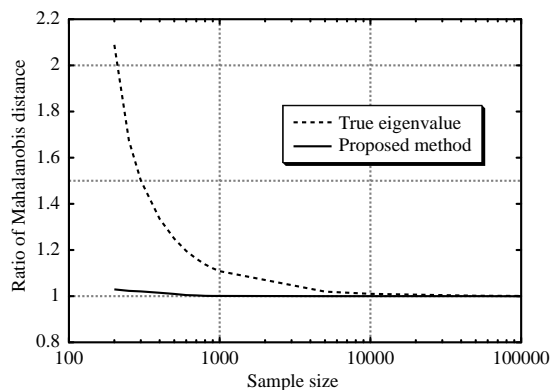


図3 マハラノビス距離の比
Fig. 3 Ratio of Mahalanobis distance.

定できている。

5. むすび

本論文では学習サンプルが十分用意できない場合においても高精度な認識を行うことを目的とし、まず、共分散行列の固有ベクトルの推定誤差によってマハラノビス距離に推定誤差が生じ、この誤差は固有値の誤差と解釈することができることを示した。そして、固有値を修正することにより、辞書の作成に用いるサンプルが少ないときでも、マハラノビス距離を正しく推定する手法を提案した。提案手法を文字認識に適用し、マハラノビス距離が正しく推定されること、認識性能が改善されることを確認した。マハラノビス距離の推定値を改善する効果はパラメータの推定に用いる学習サンプルが少ないほど高く、本手法は学習サンプルを十分用意できない場合において特に有効である。

謝辞 東北大学大学院工学研究科の佐藤俊治博士、氏家秀紀君、狩野芳正君を始め、阿曾研究室の皆様にご感謝致します。

文 献

- [1] 竹下鉄夫, 木村文隆, 三宅康二, “マハラノビス距離の推定値の確率分布に関する一性質,” 信学論 (D-II), vol.J76-D-II, no.1, pp.59-64, Jan., 1993.
- [2] F. Kimura, K. Takashina, S. Tsuruoka and Y. Miyake, “Modified quadratic discriminant functions and the application to chinese character recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.9, no.1, pp.149-153, Jan., 1987.
- [3] 加藤寧, 安倍正人, 根元義章, “改良型マハラノビス距離を用いた高精度な手書き文字認識,” 信学論 (D-II), vol.J79-D-II, no.1, pp.45-52, Jan., 1996.
- [4] F. Sun, S. Omachi and H. Aso, “Precise selection

of candidates for handwritten character recognition using feature regions,” IEICE Trans. Inf. & Syst., vol.E79-D, no.5, pp.510-515, May, 1996.

- [5] J. H. Friedman, “Regularized discriminant analysis,” Journal of American Statistical Association, vol.84, no.405, pp.165-175, 1989.
- [6] W. James and C. Stein, Estimation with quadratic loss, Proc. 4th Berkley Symp. on Math. Statist. and Prob., pp.361-379, 1961.
- [7] T. Takeshita and J. Toriwaki, “Experimental study of performance of pattern classifiers and the size of design samples,” Pattern Recognition Letters, vol.16, no.3, pp.307-312, Mar., 1995.
- [8] 酒井充, 米田政明, 長谷博行, 丸山博, 直江美知子, “固有値の偏り補正に基づく2次識別関数,” 信学論 (D-II), vol.J82-D-II, no.4, pp.631-640, Apr., 1999.
- [9] 酒井充, 米田政明, 長谷博行, 丸山博, “固有値を二重に補正した共分散行列を用いた2次識別関数,” 信学技報 PRMU99-30, pp.1-8, June, 1999.
- [10] 津田孝夫, モンテカルロ法とシミュレーション—電子計算機の確率論的応用—, 培風館, 東京, 三訂版, 1995.
- [11] 山田博三, 斉藤泰一, 山本和彦, “線密度イコライゼーション—相関法のための非線形正規化法,” 信学論 (D), vol.J67-D, no.11, pp.1379-1383, Nov., 1984.
- [12] 孫寧, 田原透, 阿曾弘具, 木村正行, “方向線素特徴量を用いた高精度文字認識,” 信学論 (D-II), vol.J74-D-II, no.3, pp.330-339, Mar., 1991.
- [13] R. J. Muirhead, Aspects of Multivariate Statistical Theory, John Wiley & Sons, Inc., New York, 1982.
- [14] 竹村彰通, 多変量推測統計の基礎, 共立出版株式会社, 東京, 1991.

付 録

1. (37) 式の導出

1.1 Wishart 分布

Wishart 分布は、 χ^2 -分布を多変量へ一般化したものである。正規分布 $N(\mathbf{0}, \Sigma)$ に従う各々独立な n 個の p 次元ベクトル $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ の平方和積和行列 \mathbf{W} を

$$\mathbf{W} = \sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t^T \quad (\text{A}\cdot\text{1})$$

とおくと、Wishart 分布は Σ が与えられたときの \mathbf{W} の分布を表し、 $W_p(n, \Sigma)$ と書く。その確率密度関数は次式で与えられる。

$$p(\mathbf{W}|\Sigma) = \frac{v(p, n) |\mathbf{W}|^{\frac{1}{2}(n-p-1)}}{|\Sigma|^{\frac{1}{2}n}} \exp\left(-\frac{1}{2} \text{tr} \Sigma^{-1} \mathbf{W}\right) \quad (\text{A}\cdot\text{2})$$

ここで $v(p, n)$ は

$$v(p, n) = \frac{1}{2^{\frac{1}{2}np} \pi^{\frac{1}{4}p(p-1)} \prod_{i=1}^p \Gamma[\frac{1}{2}(n+1-i)]} \quad (\text{A}\cdot 3)$$

である .

1.2 標本共分散行列の確率密度関数
標本共分散行列 $\hat{\Sigma}$, 標本平均ベクトル $\hat{\mu}$ を

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{t=1}^n (\mathbf{X}_t - \hat{\mu})(\mathbf{X}_t - \hat{\mu})^T \quad (\text{A}\cdot 4)$$

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t \quad (\text{A}\cdot 5)$$

とおくと , $\hat{\Sigma}$ の分布は $W_p(n-1, \frac{1}{n-1}\Sigma)$ で与えられ , $\hat{\Sigma}$ の確率密度関数は

$$p(\hat{\Sigma}|\Sigma) = (n-1)^{\frac{1}{2}(n-1)p} v(p, n-1) \frac{|\hat{\Sigma}|^{\frac{1}{2}(n-p-2)} \exp\left(-\frac{n-1}{2} \text{tr} \Sigma^{-1} \hat{\Sigma}\right)}{|\Sigma|^{\frac{1}{2}(n-1)}} \quad (\text{A}\cdot 6)$$

となる [13] .

1.3 (37) 式の導出

共分散行列 Σ , 標本共分散行列 $\hat{\Sigma}$ は固有値展開により ,

$$\Sigma = \Phi \Lambda \Phi^T \quad (\text{A}\cdot 7)$$

$$\hat{\Sigma} = \hat{\Phi} \hat{\Lambda} \hat{\Phi}^T \quad (\text{A}\cdot 8)$$

となる . $\Phi = I$ のときを考えても一般性を失わないので , (A.6) 式は次式のように書き直すことができる .

$$p(\hat{\Phi} \hat{\Lambda} \hat{\Phi}^T | \Lambda) = (n-1)^{\frac{1}{2}(n-1)p} v(p, n-1) \frac{|\hat{\Lambda}|^{\frac{1}{2}(n-p-2)} \exp\left(-\frac{n-1}{2} \text{tr} \Lambda^{-1} \hat{\Phi} \hat{\Lambda} \hat{\Phi}^T\right)}{|\Lambda|^{\frac{1}{2}(n-1)}} \quad (\text{A}\cdot 9)$$

(A.9) 式を $\hat{\Phi}$ と $\hat{\Lambda}$ の関数と考えると ,

$$p(\hat{\Phi} \hat{\Lambda} \hat{\Phi}^T | \Lambda) = p(\hat{\Phi}, \hat{\Lambda} | \Lambda) J(\hat{\Phi}, \hat{\Lambda}) \quad (\text{A}\cdot 10)$$

と変形できる . ヤコビアン $J(\hat{\Phi}, \hat{\Lambda})$ は

$$J(\hat{\Phi}, \hat{\Lambda}) = \frac{1}{\prod_{i < j} (\hat{\lambda}_i - \hat{\lambda}_j)} \quad (\text{A}\cdot 11)$$

である [14] .

(36) 式に (A.9) , (A.11) 式を代入して変形すれば (37) 式が得られる .

2. (38) 式の導出

(25) 式の右辺の積分表現は (33) 式で与えられる . 従って , (29) 式の積分表現は

$$\{\tilde{\Psi}_{jk}\}^2 = \int_{\hat{\Sigma}} \{\dot{\Psi}_{jk}\}^2 w(\hat{\Sigma}) p(\hat{\Sigma} | \Lambda) d\hat{\Sigma} \quad (\text{A}\cdot 12)$$

となる . 固有ベクトルの推定誤差の正規性から

$$\sum_k \{\tilde{\Psi}_{jk}\}^2 = \sum_k \{\dot{\Psi}_{jk}\}^2 = 1 \quad (\text{A}\cdot 13)$$

がいえ ,

$$\begin{aligned} \sum_k \{\tilde{\Psi}_{jk}\}^2 &= \sum_k \int_{\hat{\Sigma}} \{\dot{\Psi}_{jk}\}^2 w(\hat{\Sigma}) p(\hat{\Sigma} | \Lambda) d\hat{\Sigma} \\ &= \int_{\hat{\Sigma}} \left[\sum_k \{\dot{\Psi}_{jk}\}^2 \right] w(\hat{\Sigma}) p(\hat{\Sigma} | \Lambda) d\hat{\Sigma} \\ &= \int_{\hat{\Sigma}} w(\hat{\Sigma}) p(\hat{\Sigma} | \Lambda) d\hat{\Sigma} \end{aligned} \quad (\text{A}\cdot 14)$$

より , (38) 式が導かれた .

(平成 x 年 xx 月 xx 日受付)

岩村 雅一 (学生員)

平 10 東北大・工・通信卒 . 平 12 同大学院博士課程前期了 . 現在 , 博士課程後期在学中 . パターン認識に関する研究に従事 .

大町真一郎 (正員)

昭 63 東北大・工・情報卒 . 平 5 同大学院博士課程了 . 同年同大情報処理教育センター助手 . 平 8 同大工学部助手 . 平 11 同大学院工学研究科助教授 . 現在に至る . 博士 (工学) . その間 , 平 12 ~ 13 米国ブラウン大学客員助教授 . パターン認識 , コンピュータビジョン , 並列処理 , 文字認識システムの開発などの研究に従事 . IEEE , 情報処理学会 , 人工知能学会 , Pattern Recognition Society 各会員 .

阿曾 弘具 (正員)