

可変長パターンのための KMCE 学習法を用いた音声認識の評価について

小糸泰吉[†] 片桐滋[†] 大崎美穂[†]

[†]同志社大学

1 はじめに

これまでの研究では、固定次元パターン分類のための KMCE (Kernel Minimum Classification Error) 学習法を発展させた、KMCE の学習法において、カーネル写像と状態遷移モデルを用いるハイブリッド型可変長パターン分類器のための最小分類誤り学習法がある。過去にいくつか実験を行い、小規模なモデルサイズにおいて動作することを確認した。本研究では、一つの音声データセットを用いて、音声認識による評価実験を行った。

2 ハイブリッド型可変長パターン分類器

2.1 状態遷移モデル

状態遷移モデルで構成される分類器を用いて、入力可変長パターン標本 $X = \{x_1, \dots, x_t, \dots, x_T\}$ を J 個の(単語)クラス $\{C_j\}_{j=1}^J$ のいずれか 1 つに分類する問題を考える。この問題は以下の分類規則に従う。

$$C(X) = C_k \quad \text{iff} \quad k = \arg \max_j C_k(X; \Lambda) \quad (1)$$

動的計画法 (DTW: Dynamic Time Warping) によって計算された KMCE 学習法における識別関数 $g_{\varphi_{j,t} \theta_{j,t}}(x_t)$ の最大累積値を識別関数として以下のように定義する。

$$G_j(X; \Lambda) = \frac{1}{T} \sum_{t=1}^T g_{\varphi_{j,t} \theta_{j,t}}(x_t) \quad (2)$$

$$= \frac{1}{T} \sum_{t=1}^T \sum_{m=1}^M \tau_{m, \varphi_{j,t} \theta_{j,t}} K(x_t, p_m) \quad (3)$$

ここで DTW により最適な経路 $\{(\varphi_{j,1} \theta_{j,1}); (\varphi_{j,2} \theta_{j,2}); \dots (\varphi_{j,T} \theta_{j,T})\}$ は求められる。また、 $\{p_m\}_{m=1}^M$ は全ての音素モデルのプロトタイプを集約した集合、 $\{\tau_{m, \varphi_{j,t} \theta_{j,t}}\}_{m=1}^M$ は学習によって調整されるカーネルの重みである。 $K(\cdot, \cdot)$ はカーネルである。

2.2 ハイブリッド型分類器設計のための KMCE 学習法

分類は式(1)に従って行うが、学習はフレームごとに音声特徴ベクトルを対象に行う。 X が属するクラスを C_y 、最も誤りやすいクラスを C_{y^*} とすると、時刻 t における $(\varphi_{y,t} \theta_{y,t})$ を (y, t) と略記して、KMCE 学習法の誤分類尺度はカーネルによって写像された空間 H 上での幾何マージン r_H の正負を反転したものとして以下のように定義する。

$$D_{y,t}(x_t; \Lambda) = - \frac{\tau_{y,t}^T k(x_t) - \tau_{y^*,t}^T k(x_t)}{\sqrt{(\tau_{y,t} - \tau_{y^*,t})^T K (\tau_{y,t} - \tau_{y^*,t})}} \quad (4)$$

ここで $k(x)$ は M 次元のベクトルであり、経験的カーネルマップと呼ばれる。また、 K はプロトタイプカーネルによって構成される $M \times M$ の半正定値性を満たす行列であり、グラム行列と呼ばれる。

$$K(x) = [K(x, p_1) \quad K(x, p_2) \quad \dots \quad K(x, p_M)] \quad (5)$$

$$K = \begin{bmatrix} K(p_1, p_1) & K(p_1, p_2) & \dots & K(p_1, p_M) \\ K(p_2, p_1) & K(p_2, p_2) & \dots & K(p_2, p_M) \\ \vdots & \vdots & \ddots & \vdots \\ K(p_M, p_1) & K(p_M, p_2) & \dots & K(p_M, p_M) \end{bmatrix} \quad (6)$$

今回の KMCE 学習法は勾配法などの手法を用いて Λ は経験的平均損失を最小にする。この Λ を最小にすることを目的とする。

$$L(\Lambda) = \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^T \ell(D_{y,t_n}(x_t; \Lambda)) \quad (7)$$

ここで $\ell(D_{y,t_n}(x_t; \Lambda))$ は

$$\ell(D_{y,t_n}(x_t; \Lambda)) = - \frac{1}{1 + \exp(-\alpha D_{y,t_n}(x_t; \Lambda))} \quad (\alpha > 0) \quad (8)$$

で表せられる平滑分類誤り損失であり、 α は損失関数の平滑を決めるハイパーパラメータである。

Evaluation of speech recognition using a KMCE learning method for variable length patterns
[†]Taichichi Koito, [†]Sigeru Katagiri, [†]Miho Ohsaki, [†]Doshisha University

3 評価実験

3.1 データセットおよび評価手法

音声データとして ETL-WD-I データセットを用いた, 提案手法の動作確認のために動作確認のため単語音声の認識実験を行った. このデータセットは男女各 10 名, 計 20 名の話者による 492 語の読み上げ単語音声で構成され各クラス 20 標本, 492 クラスとなる. また, 特徴量は 12 次元のメル周波数ケプストラム係数 (Mel-Frequency Cepstral Coefficient: MFCC) と 1 次元のパワー, さらに, それらの一次微分値を加えた計 26 次元のベクトルを用いる. 評価方法として, 大きく分けて 2 つの実験を行った. 以下これらを実験 I と実験 II とする. 実験 I は, 学習回数を 50 回とし, 学習用標本, 試験用標本共に男女各 5 名とし, 合わせて 10 名の 492 単語とした Hold-Out 法を用いた. 実験 II では, 実験 I の結果をもとに, 学習回数を 100 回に増やし推移を調べた. また, 学習用標本に対する評価を Closed Test と呼び, 試験用標本に対する評価を Open Test と呼ぶこととする. 本実験に用いた各音素の状態遷移モデルは, 音素の状態数が 3 状態の場合で検証した.

3.2 結果と考察

提案手法に基づく, 行った実験 I の中で認識率が高かった Closed Test, Open Test の認識率を表 1 に示す. また, 実験 I におけるそれら二つの経験的平均損失の推移を図 1 に示す. さらに, 実験 II における経験的平均損失の推移を図 2 に示す.

表 1 学習用標本と試験用標本に対する分類率 (%)

	Closed Test	Open Test
実験 I における認識率	81.0%	79.5%
実験 II における認識率	91.05%	89.91%

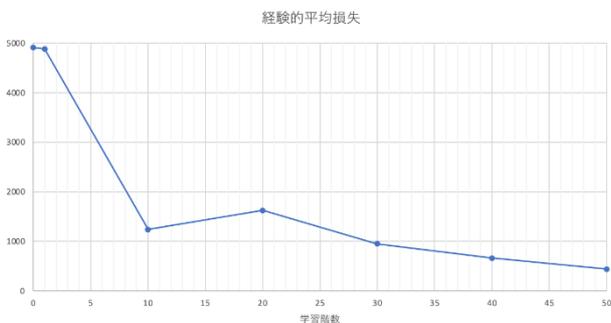


図 1 実験 I の認識率が高かった経験的平均損失の推移

表 1 から, 学習用標本だけでなく, 試験用標本の値について, 十分に学習でき高い認識率となっていることがわかる. 学習回数を 50 回にして

実験を行った図 1 の経験的平均損失の推移のグラフから学習回数が 20 回までの間に大幅に下がっている事がわかる. カーネル幅などのパラメータによる最適な値を探索する際に今回のデータセットでは学習回数を抑えることで時間短縮する事ができると考えられる. また, 経験的平均損失は 50 回以降も下がり続けると考えられる. 学習回数を 100 回に増やし行った実験 II でも同様に学習回数が 10 回で大幅に経験的平均損失の変化が見られた. その後, 少しずつ下がり続け学習回数が 100 回以降も下がり続ける可能性があると考えられる.

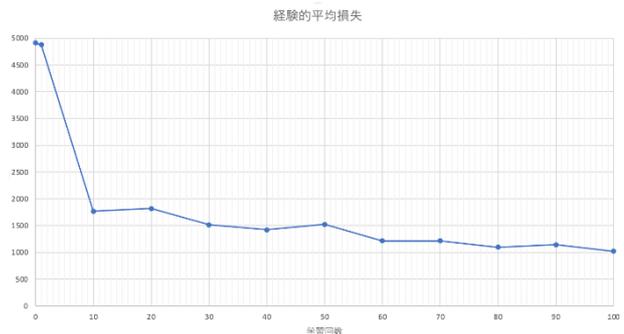


図 2 実験 II の経験的平均損失の推移

4 おわりに

本研究では, KMCE 学習法に基づくカーネルと状態遷移モデルを用いるハイブリッド型可変長パターン分類器の実験的評価を行った. 実験結果として提案手法によって学習が行われることがわかった. 今後実験をするにおいて, 損失の減少が見込めるため, さらに学習回数を増やして実験を行うことが考えられる.

謝辞: 本研究は科研費 18H03266 の支援を受けて行った.

参考文献

- 1) 渡辺秀行, 片桐滋, 足立守, 大崎美穂. カーネルに基づく高次元空間における大幾何マージン最小分類誤り学習法. 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol. 110, No. 330, pp. 55-60, Dec 2010.
- 2) 大越俊. 最小分類誤り学習法に基づくカーネルと状態遷移モデルを用いるハイブリッド型可変長パターン分類器. 同志社大学大学卒業論文, 2019.