

K-057

HMMに基づく手の動きと形状特徴の ストリーム別クラスタリングを用いた手話認識手法

HMM-based Sign Language Recognition using Partial-Clustering of Hand Movement and Shape

有賀 光希[†]
Koki Ariga

酒向 慎司[†]
Shinji Sako

北村 正[†]
Tadashi Kitamura

1. はじめに

手話は主に手や指の様々な動き(手指動作)によって意味を成すが、これらの動作は単語ごとに独立したのではなく、単語間で共通する基本動作の逐次的な連鎖から様々な単語を表現していると言える。また手指動作は、おおまかに大域的な手の位置変化(動き)と局所的な掌の形状変化(形状)という性質の異なる2つの表現要素があり、それぞれの基本動作が時間的に遷移していくと考えられる。このような視点から手話認識を考えると、個々の単語を個別にモデル化するよりも、動き要素と形状要素それぞれの基本動作モデルを単語間で共有する方が効率的であるが、表現要素ごとに基本動作をどう分類するかが問題となる。

本研究では複数の単語に含まれる共通要素を、手話の動画像から学習される統計モデルに基づいて、手の動き特徴と形状特徴それぞれについて個別に自動分類することを目指す。そのために、動きと形状の2つのストリームについて単語単位で隠れマルコフモデル(Hidden Markov Model, HMM)を学習し、単語全体の状態パラメータをストリーム別にクラスタリングする。

2. 手話単語のモデル化

手話のような時系列パターンのモデリング手法として有効性が示されているHMMに基づき、手話の性質を考慮した単語のモデル化手法を述べる。

2.1 HMMの状態に基づく共通要素の自動分類

先述のように、単語を構成する基本動作モデルの導入は、単語間でモデルを共有できるという点で効率的である。しかし手話の構成を示した一般的な枠組みが定まっていないことから、我々は、単語単位で学習されたHMM(単語HMM)の各状態は単語動作中の短区間を近似したものであると考え、複数の単語HMMの状態パラメータ空間をクラスタリングすることで、単語動作中の類似区間を自動で分類する方法を提案した[1]。これによって異なる単語間で類似した状態の共有が図れるため、全体のパラメータ数を削減できる上、状態当りの学習量増加が見込める。

2.2 動きと形状の各構成要素に基づく単語モデル

手話は、手の動き特徴(手の位置変化や運動方向等)と形状特徴(掌の形状や向き、それらの変化等)という性質の異なる特徴があり、それぞれの構成要素の組合せによって意味を成していると言える。よって、手の動きは共通でも形状の違いで意味を区別するというように、手

話単語の共通要素の出現傾向はそれぞれ異なると考えられる。そこで、前節のように類似動作の分類を時間方向に行う場合、あらかじめ両特徴を分離し、個別にクラスタリングする方が効率的に共通要素をまとめられると考えられ、その方法を以下で述べる。

1. マルチストリームHMMによる動きと形状の分離

動きと形状という振る舞いの異なる特徴を、それぞれを異なるストリームとしてみなし、HMMの出力確率をストリームごとに学習する方法を用いる。これはマルチストリームHMMと呼ばれ、本研究では、各状態の出力確率は、それぞれ重み付けされた動き特徴ベクトルの出力確率と形状特徴ベクトルの出力確率の積となる。このようにストリームごとに重みを設定することで、動きと形状の優先度を考慮したモデル化が可能である。

2. 共通要素のストリーム別の分類

動きと形状のそれぞれの共通要素を独立に分類するために、単語ごとに学習されたマルチストリームHMMの全状態をストリーム別にクラスタリングする(図1)。これにより、両特徴を分離せずクラスタリングする場合に比べて共通要素を効率的にまとめることができるため、全体のパラメータ数のさらなる削減が見込める。また、各ストリームの共通要素の複雑さの違いも、ストリームごとにクラスタ数を変えることで考慮できる。

3. 評価実験

3.1 実験条件

本手法の評価として孤立単語認識実験を行う。状態共有を行わない場合を従来法1、ストリーム間で同一に状態共有する場合を従来法2、ストリーム別に状態共有す

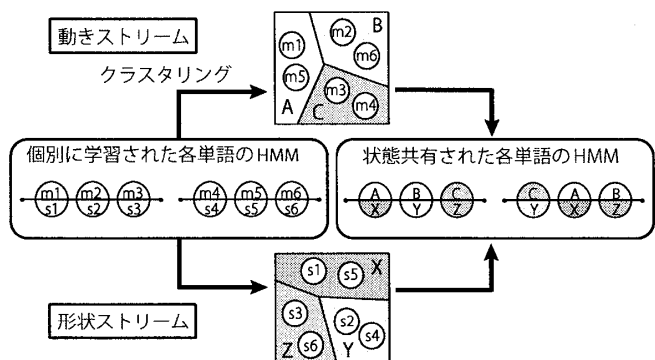


図1: ストリーム別クラスタリングによる状態共有

[†]名古屋工業大学, Nagoya Institute of Technology

表 1: 実験条件

	従来法 1	従来法 2	提案法
状態共有	なし	ストリーム非分離	ストリーム別
単語当りの状態数	5~35	25	
総状態数	500~3500	300~1050	
データベース	RWC データベース (98 年度版)[4]		
認識対象単語数	100 (3 セット)		
特徴量	動き: 両手の中心座標 4 次元, Δ , Δ^2 形状: 両手画像の主成分得点 12 次元, Δ , Δ^2		
実験データ	学習用: 3 名 \times 2 回 \times 100 単語 = 600 データ 認識用: 1 名 \times 2 回 \times 100 単語 = 200 データ leave-one-out 法 (4 セット)		
ストリーム重み	動作 0.7, 形状 0.3		

る場合を提案法として比較する。3 手法とも全単語は同一の状態数とした。よって従来法 1 では総状態数は単語数 \times 状態数となる。従来法 2 ではクラスタ数が総状態数である。提案法ではストリームごとにクラスタ数を変えられるが、従来法と同パラメータ数で比較することを考え、両ストリームのクラスタ数は同一とした(実験 1)。

また、動きと形状の特徴において共通要素の複雑さの違いを見るため、ストリームごとにクラスタ数を変え、全学習データに対するモデルの尤度から個々のストリームにおける分類傾向の違いを見る(実験 2)。

手の動き特徴として、画像から掌の中心座標を抽出し、座標の遷移を人物の体格差に関して正規化したものを用いる [2]。形状特徴は、掌を中心として切り出した画像の主成分得点を用い [3]、予備実験より両手 12 次元(片手 6 次元ずつ)とした。このとき累積寄与率は約 40%で、掌の概形と向きがおおよそ判別できる。ストリーム重みも予備実験より動作 0.7, 形状 0.3 とした。その他、実験条件を表 1 にまとめる。

3.2 実験結果

実験 1 の結果を図 2 に示す。従来法 1 に比べ、状態共有した場合はより少ないパラメータ数でありながら高い認識率を得ている。これは 1 つの状態を複数の単語データから学習できるようになったことで単語モデル全体のロバスト性が向上した効果と言える。提案法ではさらに認識率が向上し、特に総状態数が少ないときでも高い性能を維持されている。よって、動き特徴と形状特徴をそれぞれ個別に分類したことで、より効率的な状態共有ができたと考えられる。

また実験 2 の結果として、モデルに対する全学習データの尤度の総和の増加量を図 3 に示す。形状特徴は細かく分類しなければ高い尤度を示せないが、動き特徴は比較的荒い分類でも変動が少ないことから、特徴ごとに分類の精度が異なることが確認できる。よって、今後特徴ごとにクラスタ数を最適化することで、両特徴の共通要素を複雑さの違いに応じて定められると考えられる。

4. むすび

本研究では、単語間で共有可能な手話の共通要素を、手の動きと形状のそれぞれについて自動で分類する方法を提案した。これは、手指動作の特徴を動きと形状の 2

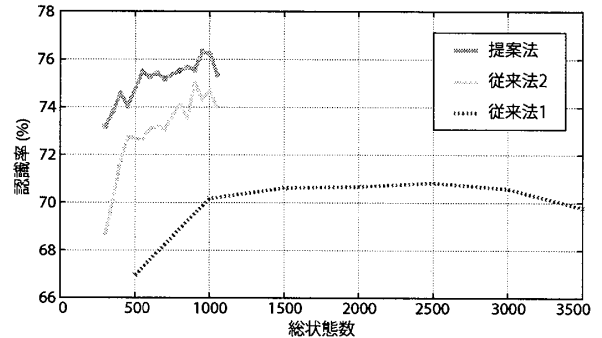


図 2: ストリーム別クラスタリングの効果の検討

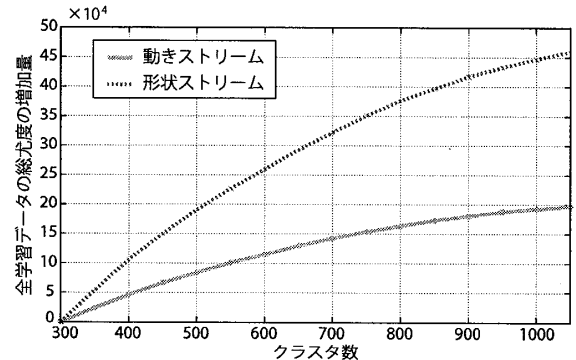


図 3: 各ストリームのクラスタ数の検討

つのストリームとして学習した HMM において、複数の HMM の全状態をストリーム別にクラスタリングすることで行った。単語認識実験では、ストリームを分離せずクラスタリングする場合よりも高い性能を示した。

今回は、動きと形状の完全な時間同期を仮定してマルチストリーム HMM により同一の状態遷移確率で扱ったが、位置変化と形状変化の非同源性も今後考慮すべきである。また手話の音韻論の研究では、手話の構成要素は位置、運動、手形状の 3 要素があり、左右の手で出現傾向も異なるという考えがあるため、このような知見をクラスタリングに利用することが今後の課題として挙げられる。

謝辞 本研究の一部は(財)立松財団平成 21 年度特別研究助成、ならびに文部科学省科学研究費補助金(課題番号:22500506)によって行われた。

参考文献

- [1] 中村 他, “手話単語認識のためのサブユニット HMM の自動生成”, 電子情報通信学会総合大会, D-12-124, p.233, 2009.
- [2] 有賀 他, “HMM に基づく人物や動作の多様性を考慮した手話認識手法”, 信学技報, WIT2010-11, pp.55-60, 2010.
- [3] 柳 他, “手の動作と形状を用いた HMM 手話認識”, 電子情報通信学会総合大会, D-12-119, p.285, 2004.
- [4] 矢部 他, “ジェスチャ認識システム評価用ジェスチャデータベースの開発”, 信学技報, IE2000-22, pp.45-50, 2000.