K-019

手話単語を構成するためのサブユニットHMMの自動生成

Self-organizing Word Level Subunits for Japanese Sign Language Using HMM

中村 光希† 酒向 慎司† 北村 正†
Koki Nakamura Shinji Sako Tadashi Kitamura

1. はじめに

手話は日常会話レベルで数千という語彙を要するが，これらの表現方法は全く独立したものでなく，単語間で共通要素が存在し，それらの組み合わせによって多様性があると考えられる．よって手話認識システムを構成する上で個々の単語を個別にモデル化するよりも，共通要素とみなせる基本的な単位 (サブユニット) をモデル化する方が効率的である．しかし手話表現の自由度の高さから，発声様式を分類した音声言語における音素のような，手話表現を分類する枠組みやそれに基づいた辞書に相当するものは充分に整備されていない．

我々は，手話映像からサブユニットを自動的に定めることを目指して，単語単位で学習された隠れマルコフモデル (HMM) の状態パラメータをクラスタリングすることで単語動作における共通要素を分類する手法を提案した [1]．しかし単語認識性能の改善は見られたが，サブユニットは過剰に分類される傾向があり共通要素の分類という点では問題があった．そこで本稿では，手話単語を構成する上で最適に分類されたサブユニットを得るためのクラスタリング手法について述べる．

2. 手話単語構成のためのサブユニットの決定

2.1 クラスタリングによるサブユニットの自動生成

特定の単語について学習された HMM (単語 HMM) は，一連の単語動作を複数の状態系列で近似したものである．単語内の基本的な動作が各状態に相当することから，多数の単語間で類似した状態を共通要素として考えると，このような分類は状態パラメータ空間におけるクラスタリングの問題として扱うことができる (図 1)．本手法では，まず単語 HMM の各状態を全て個別のクラスタとし，最近傍の 2 つのクラスタから順次統合していく．統合を終了した時点で得られたクラスタをサブユニットと定めることで，手話単語のデータから自律的にサブユニットを生成可能である．

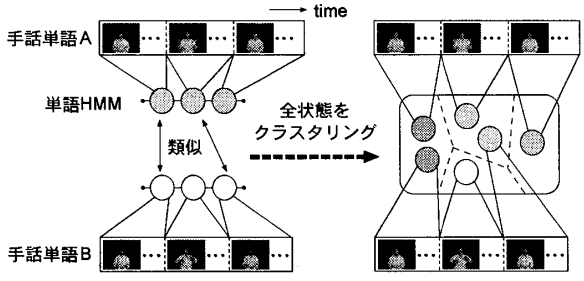

図 1: クラスタリングによるサブユニットの形成

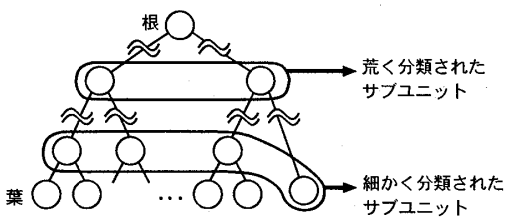

図 2: 木の深さによるサブユニットの違い

2.2 木構造を用いたサブユニットの最適な分類方法

ここで，適切なサブユニットを定めるためには，どの段階で統合を終了するかが問題となる．この方法では，クラスタを 2 つずつ統合していくので最終的に 2 分木が形成され，木の各ノードがある段階で統合を終了したときに得られる 1 つのサブユニットに対応する．そのため，サブユニットは葉に近いほど細かく，逆に根に近いほど荒く分類されたものとなる (図 2)．

過剰に分類されたサブユニットは，本来区別の不要な動作まで個別化したものであり，これは木において深すぎる位置に存在する．しかし，動作を分類するという意味では，上位のサブユニットでも対応可能であり，過剰に個別化された複数のサブユニットは上位のサブユニットで代用可能であると考えられる．

よって，サブユニットの最適な分類を行うには，2 分木上で最適な深さのサブユニットを求めれば良い．そのため，ここではある動作を表現する上で実際に用いられた複数の異なるサブユニットに対し，これら全てを代用できる汎用的な上位のサブユニットを求める．

2.2.1 木における最適な深さのサブユニットの決定

最適な深さのサブユニットを求めるため，まず，同じ単語のデータ N 個に対して 2 分木上の全ノードを用いて Viterbi パスを求めることで，その単語を表す N 個のノード系列を得る．これらは同じ単語を表す系列であるため，データ間で同じ動作に対応する区間が存在し (図 3 左)，その区間には似たノードが並ぶと考えられる．また，似たノードは木において近い位置関係に存在するため，これらのノード全てを包含する部分木の根 (上端ノード) を求めることで汎用的なノードが得られる (図 3 右)．

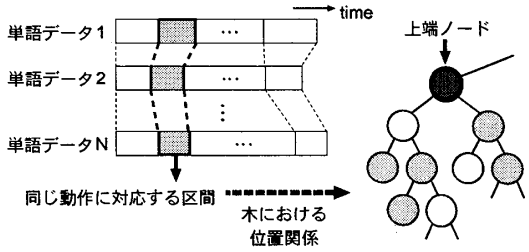

図 3: 同じ動作に対応するノードとそれらの上端ノード

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

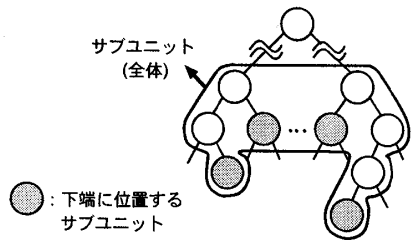

図 4: 提案法により得られる木構造をもつサブユニット

これをクラスタリングにおける統合終了のノードと定め，複数の単語の全区間について同様の操作を行うことで最適に分類されたサブユニットを得る．

ここで，サブユニットは図 4 のように木構造として表される．このことから，サブユニットの種類としては下端ノード数だけ存在するが，動作によっては不必要に分類されたサブユニットを用いなくとも，その動作を表すために十分な深さをもつより上方のサブユニットで代用可能であると考えられる．

なお，クラスタリングした段階では各ノードは単語 HMM の状態の集合でしかないので，Viterbi パスを求めるためにあらかじめ各ノードを HMM で学習しておく．これは単語 HMM の学習データから学習可能である．

2.2.2 異なるデータ間における同じ動作区間の決定

手話は同じ単語であっても，話者による動作速度の違いや手話に関係ない不要動作の挿入などが起こるため，異なるデータ間で同じ動作に対応する区間を定めることは難しい．ここでは簡単のため次のような方法を考える．

- 単語 HMM の学習データに対し Viterbi パスを求める．
- この際，そのデータを用いて学習されたノードのみが並ぶという制約を課す．

これにより，図 5 のように，データのある区間には木構造上の根から葉までのどれかのノードが並び，並ぶ個数もデータ間で共通となる．よって異なるデータでもそれらが同じ単語のデータならば，左から n 番目に同じ動作を表すノードが並ぶことになる．

3. 評価実験

3.1 実験条件

提案法により生成されたサブユニットを用いて単語認識実験を行った．ここでは，一定の個数に達するまで統合を行うクラスタリング方法 (従来法) を比較対象とした．また，サブユニット生成用の単語と認識対象の単語は同一とし，単語 HMM は left-to-right 型で状態数は 20，各画像フレームから抽出した主成分得点 50 次元を特徴ベクトル系列とした．その他の実験条件を表 1 に示す．

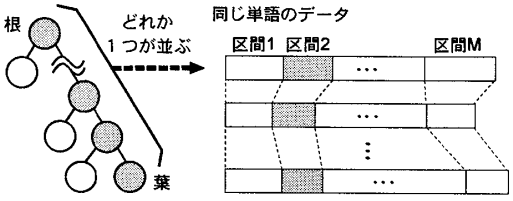

図 5: データのある区間に並び得るノード

表 1: 実験条件

手話画像のデータベース	RWC マルチモーダルデータベース [2]
画像サイズ	横 320× 縦 240 pixel
使用単語数	100
学習用データ	単語当り 6 データ ×100 単語=600 データ
評価用データ	単語当り 2 データ ×100 単語=200 データ

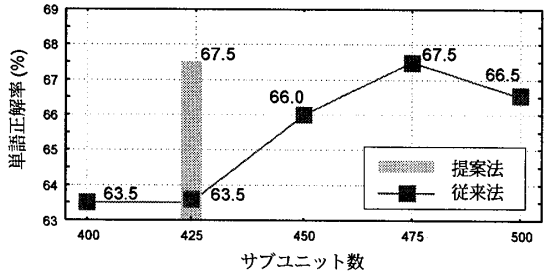

図 6: 従来法と提案法による単語認識結果の比較

生成されたサブユニットは木構造を持ち，述べ数は 861 個，下端に位置するものは 424 個であった．よってサブユニットの種類としては 424 である．一方，従来法では，サブユニットの種類はクラスタリング終了時のクラスタ数に等しいので，クラスタ数が 400~500 付近になるまでクラスタリングを行いサブユニットを生成した．

3.2 実験結果

実験結果を図 6 に示す．サブユニット数が 424 のとき，提案法が高い認識率を示していることがわかる．また，同等の認識率を得るために，提案法ではサブユニット数を削減できていることがわかる．よって，提案法で得られたサブユニットは従来法に比べて単語動作を効率的に表現できているといえる．

4. むすび

単語 HMM の全状態を階層的クラスタリング手法により分類することで．単語動作を構成するためのサブユニットを自動生成する方法を提案した．この際，クラスタリング時に形成される木構造と単語データを用いることでクラスタリングの終了条件を決定した．単語認識実験を行い従来法との比較を行ったところ，より少ないサブユニット数で同等の認識率を得られることが確認できた．

また，本来同じ単語の動作でも話者よって動作遷移が異なるが，今回はその単語のデータから学習されたサブユニットのみを用いて Viterbi パスを求めることで，強制的に動作遷移を同期させた．今後は，非同期な動作遷移のサブユニット系列中から同じ動作区間の対応付けを行うことで，その単語により適したサブユニットが求まると考えられる．

参考文献

[1] 中村 光希 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講演論文集，D-12-124，p.233，Mar. 2009.

[2] 速水 悟 他，“身振りと発話のマルチモーダルデータベース”，信学技報，PRAM97-95，pp.1–8，Sep. 1997.