4E−7

単語音声サンプルからの音韻概念の獲得

児島 宏明　速水 悟　田中 和世　(電子技術総合研究所)

1　はじめに

単語音声に相当する時系列データを学習サンプルとし、単語の区別のみを教師として学習を行ない、音韻に相当する概念を獲得するタスクについて検討する。ここで音韻概念とは、ある音声がどのような音韻列として表記されるかを決定する規準の体系を指す。

ここで想定する学習の課題は、単語単位のサンプルデータとその単語が属するカテゴリとを与えて、単語の識別能力を獲得することである。すなわちこれは、通常の単語音声認識と同じタスクである。異なる点は、学習の結果として音韻に相当する概念が形成されるように学習方法を設計する点である。これは、例からの概念学習の一例となっている。

ニューラルネットと比較すると、単語を教師としてPDP的な学習を行ない、隠れ層に音韻に相当するものが現れるのと近いが、この場合は音韻の概念が明示的になりにくい点が異なる。人工知能的な学習においては仮説空間の探索的な手法を用いることが多いが、音声のような実数データを扱う際には、統計的な手法をベースとして記号処理的手法を融合するのが有効と考えられる。音韻を単位とする通常の音声認識の手法と比較すると、通常は音韻のカテゴリは既知として音韻の識別方法を学習するが、ここでは、音韻は未知として自動的に獲得することを目指している点が異なる。単語を単位としたのは、文法などの言語的な処理を捨象することによって問題を単純化するためである。

この問題設定を人間の幼児の言語音声習得過程とのアナロジーで考えてみる。例えば、「りんご」という単語を幼児が習得する過程を考えると、「りんご」という音声を、視覚をはじめとする他の感覚と対応付けながら、「りんご」の概念を体系化していると考えられる。ただし計算機実験では、視覚などの情報は直接単語カテゴリを教えることで代用している。音韻概念そのものは教えないという設定は、幼児の場合は通常音素そのものを外部から教えることはないことに対応する。

2　学習方式

学習用サンプルとして与える単語音声の時系列データの集合を S とし、その各要素を $\mathbf{s}_1, \mathbf{s}_2, \ldots$ とする。時刻 t におけるサンプル $\mathbf{s}_i$ の特徴ベクトルを $\mathbf{x}_i(t)$ とする。ここで s_i の時間長を T_i とすると、t は $1, 2, \ldots, T_i$ の整数値をとる。また、単語カテゴリの集合を Ω とし、その各要素を $\omega_1, \omega_2, \ldots$ とする。サンプル $\mathbf{s}_i$ が属する単語カテゴリを $L(\mathbf{s}_i)$ で表す。

学習のタスクは、サンプル $\mathbf{s}_i$ と教師情報 $L(\mathbf{s}_i)$ との組を与えて、L に相当する識別機構を推定することである。

以下に学習の手順を述べる。

基本的な方針は、異なる単語が同じ音韻表記となることがないという条件のもとで、クラスタ数の下限を推定することである。通常のクラスタリングと異なる点は、単語の区別に関して拘束があることである。

(1) $k = 2$ とする。

(2) すべてのサンプル中の全特徴ベクトルをクラスタリングにより k 個のクラスタに分け、生成されたクラスタを $c_1, c_2, \ldots, c_k$ とする。このときのクラスタリング関数を F とする。

(3) すべての $\mathbf{s}_i$ について $\{F(\mathbf{x}_i(1)), F(\mathbf{x}_i(2)), \ldots, F(\mathbf{x}_i(T_i))\}$ を求め、これを q_i とする。また、q_i 中で同じクラスタが連続して現れたものをひとつに縮約したものを p_i とする。

On Phonological Concept Acquisition from Spoken Word Samples.
Hiroaki KOJIMA, Satoru HAYAMIZU and Kazuyo TANAKA
Electrotechnical Laboratory

(4) $L(\mathrm{s}_i) \neq L(\mathrm{s}_j)$ でかつ $p_i = p_j$ となる (i,j) の組が存在すれば k を 1 増やして (2) へ。

(5) クラスタ $c_1, c_2, ..., c_k$ の任意の 2 クラスタの組のうち、中心間の距離が小さいものから順に合併をとり、これを新たなクラスタとしてすべて p_i を書き換えたときに、$L(\mathrm{s}_i) \neq L(\mathrm{s}_j)$ でかつ $p_i = p_j$ となる (i,j) の組が存在すれば合併を解消してすべての p_i を元に戻す。これをすべてのクラスタの組について実行する。未実行の組が無くなれば終了。

例えば、/a,i,u,e,o,Q/ の 6 音韻 (Q は無音) が 2 次元のパラメータで表され、図 1 のように分布していたとして、この 6 音韻を含む充分な単語サンプルが与えられたとする。$k = 7$ のときに図の点線のようにクラスタが分割されステップ (5) に到達したとすると、ステップ (5) で図中の c_6 と $c7$ の合併がとられ、最終的なクラスタ数は音韻数と同じ 6 となる。

このアルゴリズムには対象となるドメインに依存した知識が埋め込まれた形になっており、下のように、多くの部分が、音声に関する先験的な知見に基づいて決められている。

- ステップ (2) でクラスタに分割するのはのは、単語が音韻のようなより細かい単位から構成されることを仮定しているからである。
- クラスタリングが有効なのは概念の近さがパラメータ空間の距離に反映されることを仮定しているからである。
- ステップ (3) で縮約を行なうのは、ひとつの音韻は、ある時間範囲で連続した位置を占めることを仮定しているからである。
- k を充分大きくした後、最小限に減らす処理をしているのは、すなわち、生成されるクラスタ数 (音韻数) は必要最小限であることが望ましいということを前提としている。

3 検討

本手法により、単語カテゴリを教師情報として音韻に相当するカテゴリが生成される。しかし、生成されたものが通常の意味での音韻と同じであることは保証されない。

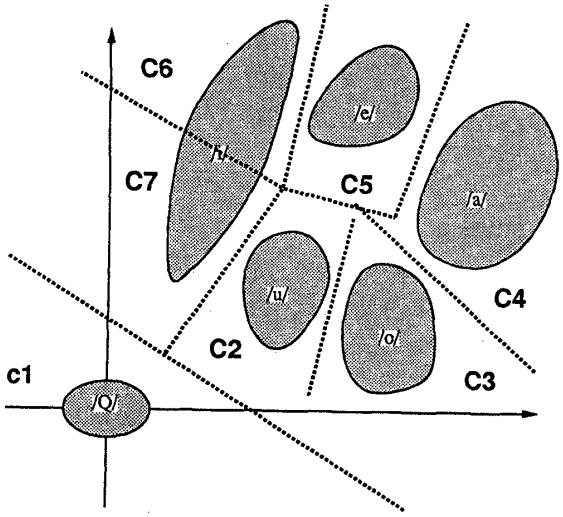

図 1: アルゴリズムの説明図

本手法の問題点としては、クラスタの境界は教師無しで求めたものであるため、必ずしも最適な判別面とはならないこと、最小音素対の仮説から生じるような音韻の概念と、ここで得られるクラスタとが対応しないこと、音韻の概念を階層構造を持った特徴 [1] として表現できないことなどが挙げられる。また、対象ドメインに依存する知識はアルゴリズムから分離し、汎用性のあるものにするのが望ましい。また、何が学習可能で何が先見的に与える必要がある知識かを再検討する必要がある。これらを含めていかに実現可能な手法を見い出すかが今後の課題である。

参考文献

[1] 児島、田中、太田: 「ボトムアップ型音声認識のための音素片の識別」音響講論, 2-P-12(1990,3)

[2] E. F. Jogensen (林 監訳) : 「音韻論総覧」 大修館 (1978)