

## 知識ベースに基づく連続音声認識システム

6N-3

## —帰納的学習による認識ルールの自動生成—

竹之内 正一郎† 辻野 克彦† 溝口 理一郎# 角所 収#

(† 大阪大学大学院 # 大阪大学 産業科学研究所)

## 1. まえがき

筆者等が開発してきた連続音声認識エキスパートシステム<sup>(1)</sup>では、音声認識のルールの入力是人間のエキスパートが持つ知識にのみ依存していた。このようなルールの入力方法は少数の話者に対してはエキスパートの負担も比較的少なく、それによって得られるルールの質も良好である。しかし、システムを不特定多数の話者に適応させようとすると、特殊な音声認識ルールの出現やパラメータの再調整等の問題に対処する必要が生じ、それによって人間のエキスパートが自分の持つ知識を抽出し表現することが困難になってくる。

一般に、自分の持つ知識を正確に記述しようとする困難を伴うが、例題を用いることによって表現することは比較的容易である。特に音声認識の場合、話者が何を言ったのかがはっきりしており例題も与え易い。したがって、例題から音声認識ルールを自動生成することが可能になればエキスパートからのルール抽出という最大のボトルネックを回避できると考えられる。これらのことを考慮して、我々は帰納的学習を用いた子音認識ルールの自動生成システムの開発を行っている。本稿では、本システムの概要について述べる。

## 2. ルール生成システムの概要

ルール生成システムの構成を図1に示す。

## 2.1 データ生成部

音声データは実際の音声をA/D変換し特徴パラメータの抽出を行ったものである。記号化モジュールはパラメータの分布を考慮してデータの記号化を行い、さらにセグメンテーション(各音韻の境界の認識)を施してセグメントデータとしルールインダクションモジュールへ送る。また、学習の結果得られた新しい属性の属性値をセグメントデータごとに算出する。

## 2.2 ルール学習部

①ルールインダクションモジュール：与えられた訓練集合に対して機能的学習を行い、決定木を生成する。

②評価モジュール：ルールインダクションモジュールによって得られた決定木を基に属性生成モジュールに必要な情報を生成する。

③学習知識：新しい属性を得るためのヒューリスティクスの集まりである。

④属性学習モジュール：簡単なプロダクションシステムになっており評価モジュールで生成された情報と学習知識をもとに新しい属性を生成する。

⑤手続き生成モジュール：属性学習モジュールからの指示を受けて属性値を求める手続きを生成し、これをデータ生成部へ送る。

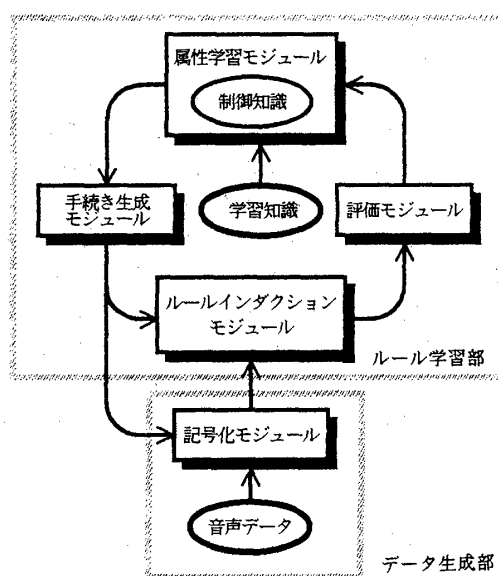


図1 ルール生成システムの構成

## 3. 帰納的学習によるルールの自動生成

## 3.1 ID3

本システムでは、決定木生成のアルゴリズムとしてID3<sup>(2)</sup>を採用している。システムに与えられるデータは多数の属性を持っており、決定木で実際に認識(分類)される過程で、ある属性が選択され、その属性値によって振り分けられていく。ID3では、認識に使用する属性の選択の基準として、その属性を使用することによって得られる情報量の多さを使っている。

## 3.2 生成される認識ルールの例

本システムに与えられるセグメントデータの属性としては、音声を表現するための特徴であるパワー、ホルマント周波数、有声無声区間、高域周波数成分の変化を与えている。具体的には図2に示される形式をしている。

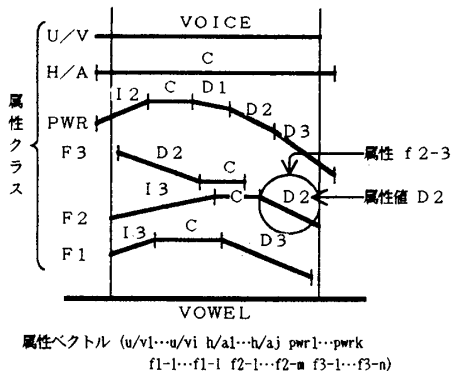


図2 セグメントデータの構造

図中に示されている pwr1, f1-1 といったものが属性で、これらは属性ベクトルを構成する。また、時間的に連続している属性群をひとまとめにして属性クラス (U/V, H/A 等) としている。属性の数は最も多くの属性を必要とするデータに合わせてあり 30 個前後である。また訓練集合として使用されるデータには属性ベクトルの他にそのデータの正解音韻が与えられる。子音認識では、正確な音韻の特定が困難であり、音韻をある程度グループ化している。我々のシステムでは音韻グループとして VOWEL, PTK, BDGZ, R, BDGZR, MN, SH, W, Y の 9 個を使用している。以上のような構造を持つデータを多数用意して訓練集合を構成し、これに ID3 を適用することによって図 3 (a) のような決定木が得られる。

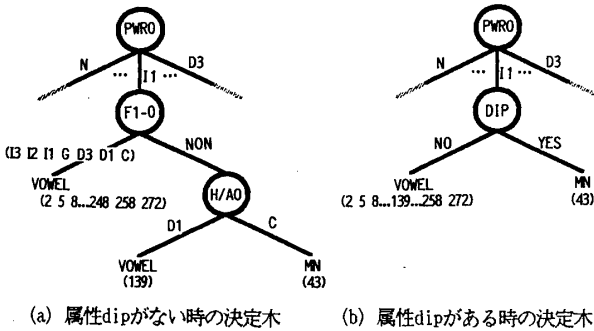


図3 生成された決定木の例

4. 属性の学習

4.1 属性の重要性

ID3 は基本的には誤りを含まないデータ群を対象にしたアルゴリズムであり、誤りを含むデータに対しては決定木を生成できないことがある。セグメントデータは多数の属性を持っているので、多少の誤差があったからといって決定木を生成できないことはほとんど無いが、その代わり決定木の深さが異常に深くなってしまふ。したがって、決定木のひとつひとつのリーフがカバーできるセグメントデータの領域が非常に制限されたものになる。実際の子音認識では、訓練集合中のデータ以外のセグメントデータに対してこの決定木を適用するので、これは大きな問題点である。

従って、認識率の高いルールを生成するためには誤差の少ない属性を使用する必要がある。しかし、多数の音声データを見て、その属性を正確に定義するのは人間にとっては非常に困難である。したがって、ルールの自動生成においてはこのような属性の定義をいかにして獲得するかが非常に重要となる。

4.2 属性学習のための知識

本システムは、最初は与えられた属性のみで決定木を生成する。そして、そのツリーの形、使用されたデータ、属性を総合的に評価しその結果をみて、利用できそうな属性を生成していく。そして、その新しい属性も含めて再度決定木を生成する。このようなサイクルを繰り返し行うことによって次第に優れた属性を学習していく。このサイクルにおいて新しい属性の候補を提案するのが学習知識である。現在、この知識として表 1 のようなものを用意している。

表 1 学習知識の例

①	ひとつの音韻グループを認識するのに使用された属性を組み合わせて新しい属性とする。
②	属性値の定義を一般化して新しい属性とする。
③	属性値の定義を特殊化して新しい属性とする。
④	ある特定のパターンが属性クラス内に現れるか否かを属性とする。

以上のような知識を使用すると、例えば DIP (子音に特徴的に現れるパワーのくぼみ) が存在するか否かという属性が得られる。この属性 DIP を加えた時の決定木を図 3 (b) に示す。図 3 (a) と (b) を比較してみると、(a) の方では、たまたま誤差のためにデータ 43 の属性 f1-0 の属性値がデータ 139 の値と同じであった為に決定木が深くなっている。一方、(b) では VOWEL と MN をより一般的に区別する属性 DIP があるために、決定木が深くなるのが避けられている。また、属性 DIP は h/a0 に比べて誤差を含みにくくなっており、したがって決定木自体の認識率も向上する。

5. まとめ

本稿では現在検討中の子音認識ルール自動生成システムについてその概要を述べた。現在このシステムを Symbolics 3600 システム上にインプリメント中である。今後、学習知識や属性学習モジュール内の制御知識を充実させてさらに効率よく属性を学習するように改良していきたいと考えている。

[参考文献]

[1] 辻野他: "知識ベースに基づく連続音声認識システム - 知識ベース構築の問題点とその解決法 -", 情報処理学会第33回全国大会, 6N-1 (1986).  
 [2] Quinlan, J. R.: "LEARNING EFFICIENT CLASSIFICATION PROCEDURES AND THEIR APPLICATION TO CHESS END GAMES", Machine Learning, Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (Eds.), Springer-Verlag, New York, pp. 463-482, 1984.