

論文 / 著書情報
Article / Book Information

論題(和文)	木構造確率分布を用いた音声認識
Title(English)	
著者(和文)	渡辺隆夫, 篠田浩一, 高木啓三郎, 山田栄子, 服部浩明
Authors(English)	Koichi Shinoda
出典(和文)	日本音響学会平成5年度秋季研究発表会講演論文集, Vol. 1-8-7, No. , pp.
Citation(English)	, Vol. 1-8-7, No. , pp.
発行日 / Pub. date	1993,

○渡辺隆夫 篠田浩一 高木啓三郎 山田栄子 服部浩明 磯健一 (NEC・情報メディア研究所)

1. まえがき

近年、HMMによる音声認識において、混合連続確率分布の利用が盛んである。連続確率分布HMMでは、ガウス分布数は、認識単位の種類×認識単位のモデルの状態数×状態での混合数である。tied mixture型の連続確率分布を使用した場合にはガウス分布数を通常の連続確率分布より少なくできる。いずれの場合でも、認識時にはガウス分布の個数分の確率計算が必要である。認識処理の高速化にとって、この確率計算部分の高速化は分析処理や単語照合(探索)処理の高速化と並んで重要な課題である。本文では、連続確率分布HMMによる音声認識の高速化の一方法として、木構造表現した確率分布を用いた高速認識法を提案する。

2. 方法

2.1 基本的な考え方

入力パターンを $X = x_1, \dots, x_t, \dots, x_T$ とする。出力確率における混合成分であるガウス分布の全状態、全認識単位についての集合を

$$Y = \{N_1[\cdot], \dots, N_k[\cdot], \dots, N_K[\cdot]\},$$

とする(各 $N_k[\cdot]$ はガウス分布)。認識時には、各時刻 t において、確率値集合

$$B_t = \{N_k[x_t] \mid k=1, \dots, K\} \quad (1)$$

を求めることが必要である。(1)式の確率値集合を効率よく求めることを考える。

基本的な考え方は、認識の結果、ラベル k が時刻 t に対応付けられる可能性のある k に対しては確率値の計算をなるべく厳密に行い、そうでない k に対しては計算を手抜きするというものである。これをもう少し具体的に表現すると次のようになる。

確率値の計算を、確率値が大きくなる k に対しては正確に行い、小さくなる k に対してはおおまかに行う。

2.2 木構造確率分布による高速化

上記の原理を、ガウス分布集合 Y を木構造表

現し、(1)式の計算を木探索によって実行することにより実現する。

ガウス分布集合 Y の要素を木の葉ノードに対応させる。中間ノードは要素を集めたクラスタを表しており、ガウス分布(クラスタ代表分布)を付随させる。このクラスタ代表分布はクラスタ内の要素に対応するガウス分布の集合を近似的に単一ガウス分布で表現したものである。

認識時には、木の根ノードから各ノードにおいて確率分布の計算を行い、確率値の高い子ノードを選択する処理を繰り返す。選択されなかったノードの下位にある要素(すなわち選択されなかったクラスタに属するガウス分布)に対しては確率値の計算を行わず、かわりに上位のノードで計算した確率値によって近似する。以上の処理により、(1)式の確率値集合の近似値を効率よく計算することができる。

2.3 木構造確率分布の設計

木構造設計法としてトップダウンクラスタリングを採用する。すなわち、ガウス分布集合 Y を初期集合として、確率分布集合を指定された個数のクラスタに分割する処理を繰り返す。クラスタリング法として K -means法を採用する。以下に手順を示す。

- ① 初期クラスタ代表分布を与える
- ② 処理③、④をクラスタ分割が収束するまで繰り返す
- ③ 各要素がどのクラスタに属するかを判定する。ここでは、クラスタ代表分布 $N_m[x]$ と要素のガウス分布 $N_k[x]$ との距離として、Kullback Divergenceを用いて判定を行った。

$$D[k, m] = \int (N_k[x] - N_m[x]) (\log(N_k[x]) - \log(N_m[x])) dx \\ = \sum_i \left\{ \frac{\sigma_{k(i)}^2 + \Delta_{km(i)}^2}{\sigma_{k(i)}^2} + \frac{\sigma_{m(i)}^2 + \Delta_{km(i)}^2}{\sigma_{m(i)}^2} \right\}$$

$$\Delta_{km(i)}^2 = (\mu_{k(i)} - \mu_{m(i)})^2$$

$N_k[x]$, $N_m[x]$ は無相関ガウス分布とし、 $\mu_{k(i)}$, $\mu_{m(i)}$ は平均ベクトルの第 i 要素、 $\sigma_{k(i)}^2$, $\sigma_{m(i)}^2$ は

*Speech Recognition Using Tree-Structured Probability Density Function.

By T. Watanabe, K. Shinoda, H. Takagi, E. Yamada, H. Hattori, K. Iso (NEC Corporation)

