

6B-2

連続音声認識に適した高速/省メモリな HMM 計算回路構成法

山本 正俊[†] 中村 一博[‡] 高木 一義[‡] 高木 直史[‡]

名古屋大学 工学部[†]
名古屋大学大学院 情報科学研究科[‡]

1 はじめに

現在、音声認識には時間的な変動に対する柔軟性から、音素や音韻、単語などのモデルに連続分布型 HMM (隠れマルコフモデル, Hidden Markov Model) が広く用いられている。連続分布型 HMM を用いる手法では認識性能が高い反面、出力確率計算における計算量が大きく、組み込み携帯機器での実時間応答の実現には高速化を行うことが重要である。このため、専用ハードウェアを用いて高速化を図る研究がなされてきた [1, 2, 3]。本稿では、使用頻度の高いデータである特徴ベクトルを回路内のメモリに保持することで、従来の回路構成よりも高速に出力確率の計算を実行する HMM 計算回路の構成法を提案する。

2 連続分布型 HMM における出力確率計算

連続分布型 HMM に基づく音声認識の処理時間における大部分を占める出力確率計算を式 (1) に示す。この計算は、時刻 t における P 次元の特徴ベクトル $\mathbf{o}_t = (o_{t,1}, o_{t,2}, \dots, o_{t,P})$ から、ガウス分布確率密度に基づく出力確率 $\log b_i(\mathbf{o}_t)$, ($i = 1, \dots, N$) を求める計算である。

$$\log b_i(\mathbf{o}_t) = \omega_i + \sum_{p=1}^P \rho_{i,p}(o_{t,p} + \mu_{i,p})^2 \quad (1)$$

ここで、 N は HMM の状態数、 ω_i , $\rho_{i,p}$, $\mu_{i,p}$ は HMM の学習データである。学習データは事前の学習により得られた、HMM 毎に固有の値である。現在の連続音声認識では、状態数 N は 3 から 10、次元数 P は 20 から 40 程度の値がよく用いられる。

3 連続音声認識に適した HMM 計算回路の構成法

本稿で提案する HMM 計算回路の回路構成を図 1 に示す。この回路は、特徴ベクトルを格納するためのレジスタアレイ RegO と、学習データを格納するためのレジスタアレイ Reg μ , Reg ρ , 出力確率計算の中間結果を格納するためのレジスタアレイ RA1, M 並列の演算ユニットから成る。この演算ユニットは、2つの加算器と2つの乗算器により構成される、式 (1) の計算を行うための回路である。提案する回路構成では、演算ユニットの並列度 M に対して、RegO のサイズを $2M \cdot P$, RA1 のサイズを $2M$ とする。この構成により、特徴ベクトル \mathbf{o}_t について同じデータを読み込む必要がなくなり、効率的に出力確率の計算を行うことが出来る。

A High-Speed Memory-Efficient HMM Circuit Structure for Continuous Speech Recognition

[†]Masatoshi YAMAMOTO, School of Engineering, Nagoya University

[‡]Kazuhiro NAKAMURA, Kazuyoshi TAKAGI, and Naofumi TAKAGI, Graduate School of Information Science, Nagoya University

4 提案構成法による HMM 計算回路の動作

図 2 に図 1 の回路のタイミングチャートを示す。初め

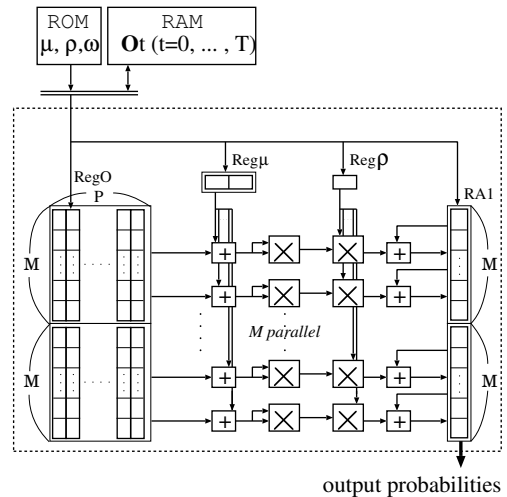


図 1 高速・省メモリ型 HMM 計算回路

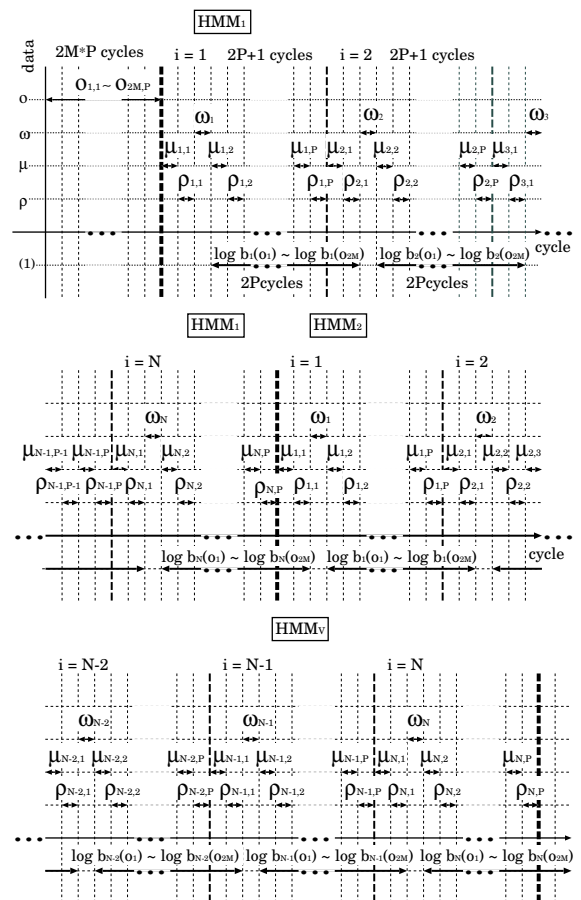


図 2 提案回路のタイミングチャート

表 1 提案構成法と [1] の回路構成法との比較

	conventional[1]	proposed
register size [bit]		
o	0	$2xMP$
μ	xNP	$2x$
ρ	xNP	x
ω	xN	0
RA1	yN	$2yM$
total	$x(2NP + N) + yN$	$x(2MP + 3) + y2M$
processing time [cycle]		
o	VTP	$\lceil \frac{T}{2M} \rceil 2MP$
μ	VNP	$\lceil \frac{T}{2M} \rceil VNP$
ρ	VNP	$\lceil \frac{T}{2M} \rceil VNP$
ω	VN	$\lceil \frac{T}{2M} \rceil VN$
other	0	2
total	$V(2NP + N + TP)$	$\lceil \frac{T}{2M} \rceil (2MP + VN(2P + 1)) + 2$

にレジスタアレイ RegO に $2M$ 個の P 次元特徴ベクトルを RAM から $2M \cdot P$ クロックサイクルを費やして格納する．この特徴ベクトルは全ての HMM の全ての状態において共通であるので，以後読み直す必要はない．続いて ROM から学習データ $\mu_{1,1}$ と $\rho_{1,1}$, ω_1 をそれぞれ Reg μ , Reg ρ , RA1 に各 1 サイクルで格納する，これにより式 (1) の計算を開始するためのデータがそろふ．以後， $\mu_{1,p}$ と $\rho_{1,p}$ を ROM から交互に読み込み，同時に式 (1) の $\log b_1(o_1), \dots, \log b_1(o_{2M})$ の計算を $2M$ 並列で行い，中間結果を RA1 に格納する．状態 $i = 2, \dots, N$ についても同様に計算を行う．各状態 i に対する式 (1) の計算は，次の状態 $i + 1$ の学習データ $\mu_{i+1,1}$ と $\rho_{i+1,1}$ を格納したときに同時に終了する．

5 従来の回路構成法との比較

これまでに提案された連続分布型 HMM 計算回路の中で最も高速な，主に孤立単語音声認識に用いられる，HMM 計算回路 [1] との比較を表 1 に示す．表 1 において，従来の回路構成 [1] を conventional，提案する回路構成を proposed としている．また， x は外部の RAM，ROM からの入力ビット幅， y は回路の出力ビット幅を表し， T は各 HMM への入力総フレーム数， V は HMM の総数を表す．register size は必要となるレジスタアレイのサイズについての比較であり， $o, \mu, \rho, RA1, \omega$ はそれぞれ，レジスタアレイ RegO, Reg μ , Reg ρ , RA1, および従来の回路構成において必要となるであろう， ω を格納するためのレジスタアレイのビット数であり，total はそれら全ての合計値である．また，processing time は HMM 計算に要するクロックサイクル数についての比較であり， o, μ, ρ, ω は，それぞれのデータを読み出すのに費やすクロックサイクル数であり，other は，データの読み出しが全て終わった後，式 (1) の計算が完了するまでにかかるクロックサイクル数である．

図 3, 4 に HMM の総数を 21000，状態数を 5，次元

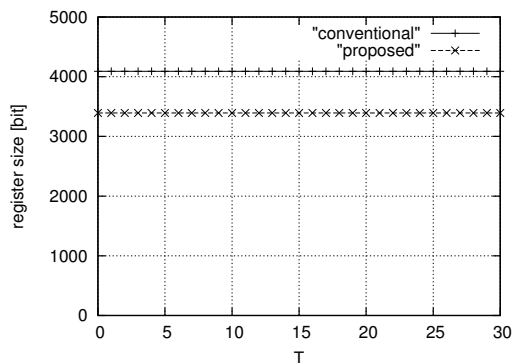


図 3 T を変化させた場合のメモリ量

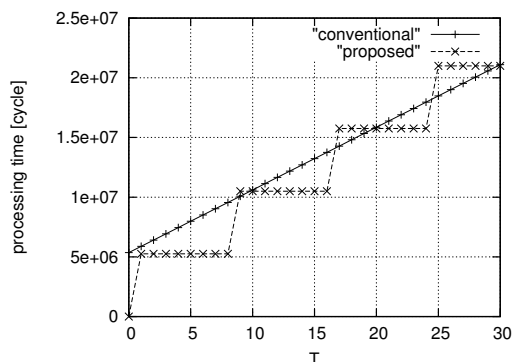


図 4 T を変化させた場合のサイクル数

数を 25，提案構成法の並列数を 4， x を 16， y を 18 とし，総フレーム数を 0 ~ 30 で変化させた場合のレジスタサイズとサイクル数のグラフを示す．

これらのグラフより，レジスタサイズは従来の構成法 [1] より小さく，処理速度も多くの場合で高速であることが分かる．また，このとき必要となる演算器の数は [1] が加算器と乗算器がともに 10 個ずつであるのに対し，提案する回路では 8 個ずつである．従来の構成法 [1] では総フレーム数を大きくすると，より高速に処理することが出来るが，連続音声認識の性質を考えると総フレーム数は小さいほうがよいので，提案する構成法が省メモリであり，高速に処理を行うことが出来るといえる．

6 まとめ

本稿では使用頻度の高いデータである特徴ベクトルを回路内のメモリに保持することで，連続音声認識のような総フレーム数の小さな処理に対して，従来の回路構成法に比べて必要となるメモリの量とサイクル数が少ない HMM 計算回路の構成法を提案した．今後は HDL を用いて提案構成法による HMM 計算回路の設計を行い，評価を行う予定である．

参考文献

- [1] 吉沢 真吾, 宮永 喜一, 吉田 則信, “一括並列処理による HMM 高速化手法及びその VLSI 設計,” 電子情報通信学会論文誌 A, Vol. J85-A, No. 12, pp. 1440-1450, Dec. 2002.
- [2] Stephen J. Melnikoff, Steven F. Quigley, and Martin J. Russell, “Speech Recognition on an FPGA Using Discrete and Continuous Hidden Markov Models,” Field-Programmable Logic and Applications, 202-211, Sep. 2002.
- [3] Binu Mathew, Al Davis, and Zhen Frag, “A Low-Power Accelerator for the SPHINX 3 Speech Recognition System,” Proc. of Int'l Conf. on Compilers, Architecture and Synthesis for Embedded Systems, pp. 210-219, 2003.