

HMM 出力確率の共分散行列に対する制約条件の実験的考察

船田 哲男[†](正員) 岸尾 一史^{†*}

An Experimental Study on Conditions for Covariance Matrices of HMM Output Probability

Tetsuo FUNADA[†], Regular Member
and Kazufumi KISHIO^{†*}, Nonmember

[†] 金沢大学工学部, 金沢市

Faculty of Engineering, Kanazawa University, Kanazawa-shi,
920-8667 Japan

* 現在 (株) 松下通信金沢研究所

あらまし HMM 出力確率の共分散行列において、低次のパラメータ間にもみ相関を考慮した部分的な制約条件を加えたモデルを作成し認識実験を行った。その結果、学習データが制限されている条件のもとで、対角共分散や全共分散を仮定して学習したモデルより、学習外話者に対する認識率を上げる効果があることを実験的に確認した。

キーワード HMM 出力確率, 対角共分散行列, 全共分散行列, 部分全共分散行列

1. まえがき

HMM (Hidden Markov Model) で音声認識を行うと、その認識率は一般的に学習外話者に対する結果は学習内話者より低下することはよく知られている。そこで、学習外話者に対しても認識率を上げるため、モデルのパラメータ数を増やしてモデルの精度を上げることが必要となる。しかし、学習データ量が制限されている場合はパラメータ推定精度が低下するため限度がある。HMM 出力確率を対角行列に制限するとモデルのパラメータ数を抑えることができるのでしばしば用いられているが、この場合も話者数が増えると近似が不十分であり、複数個の分布関数の和で出力確率を仮定した混合分布が用いられる [1], [2]。もちろん、混合数が増えるとパラメータ数が増大するのでパラメータ推定精度が問題になってくる。あらかじめ、特徴ベクトルの相関を除去するような回転を施して対角行列で共分散行列を学習する方法なども提案されている [3]。

本研究では、学習データ量が制限されている条件のもとで学習外話者の認識率を上げることを目的として、HMM 出力確率の共分散行列の非対角要素に対する制約条件を種々仮定したモデルで認識実験を行った [4]。ここでの制約条件とは、特徴パラメータとして用いるケプストラム係数間で一部の異なる次数間にもみ相関

があるという仮定である。相関を仮定する次数を種々設定して単語音声の認識実験を行い、認識率との関係について実験的な考察を行った。

2. 出力確率分布

HMM の状態 s で観測される出力ベクトル \mathbf{o} に対する混合数 M の連続確率密度関数 $b_s(\mathbf{o})$ は

$$b_s(\mathbf{o}) = \sum_{m=1}^M w_{s,m} N(\mathbf{o}; \mu_{s,m}, \mathbf{R}_{s,m}) \quad (1)$$

で表される。ここで、 $w_{s,m}$ は状態 s での m 番目の混合重みである。また、 $N(\mathbf{o}; \mu_{s,m}, \mathbf{R}_{s,m})$ は平均 $\mu_{s,m}$ と共分散行列 $\mathbf{R}_{s,m}$ で表される多次元ガウス分布である。

本研究では、この共分散行列に特に制約を付けずに全要素を学習した HMM を全共分散 HMM、非対角要素をすべて 0 に設定して学習した場合を対角共分散 HMM と呼ぶことにする。また、ある次数の範囲内だけ部分的に相関を考慮し、その非対角要素の値も学習した場合を部分全共分散 HMM と呼ぶことにする。

例えば、 k 次までの低次同士のパラメータでのみ相関を考慮した場合は、

$$\mathbf{R}_{s,m} = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1k}^2 & & \mathbf{0} \\ \vdots & \ddots & \vdots & & \\ \sigma_{k1}^2 & \cdots & \sigma_{kk}^2 & & \\ & & & \ddots & \\ \mathbf{0} & & & & \sigma_{nn}^2 \end{bmatrix} \quad (2)$$

のように表現される。

3. 実験方法

本研究では、Left-to-Right の連続形 HMM を用いて単語音声の認識実験を行う。実験に用いた音声データは、サンプリング周波数 10 kHz、男性話者 10 名が都道府県名など 50 種類の単語を各 5 回発声した合計 2,500 個の単語音声である。音声の特徴パラメータには LPC ケプストラムを用い、LPC 分析次数と同じ次数のケプストラム係数まで用い、各次数ごとに独立に学習/認識を行った。HMM の学習は、認識対象話者 1 名を除く話者 9 名の第 1 発声から第 5 発声までの単語で行った。認識実験には、学習に用いなかった話者 (認識対象話者) 1 名の第 1 発声から第 5 発声までの単語を用いた。認識対象話者 (計 5 名) を順次入れ替え、この実験を 5 回繰り返した。HMM 出力確率の混合数 M は 1 とし、1 単語の状態数は 10 に固定した。

表1 全共分散 HMM におけるケプストラム次数と認識率 (%) の関係

Table 1 Dependency of recognition rates by full covariance HMM on cepstral analysis order.

次数 n	4	5	6						
認識率	87.2	89.1	90.4						
	10	11	12	13	14	15			
	91.7	91.8	91.0	89.9	90.5	90.6			

表2 対角共分散 HMM におけるケプストラム次数と認識率 (%) の関係

Table 2 Dependency of recognition rates by diagonal covariance HMM on cepstral analysis order.

次数 n	4	5	6						
認識率	81.7	84.7	86.3						
	10	11	12	13	14	15			
	92.1	92.8	92.9	92.0	92.4	92.6			

4. 実験結果と考察

4.1 全話者平均の結果と考察

全共分散 HMM で学習/認識した結果を表 1 に示す。表の各値は学習外話者 5 人の平均認識率である。分析次数 n ごとの認識率を調べたところ、次数が 11 のときに最も高い認識率となった。次数を 11 より高くしてパラメータ数を増やすと、パラメータ推定精度が下がるため認識率が低下すると考えられる。

次に、対角共分散 HMM で学習/認識した結果を表 2 に示す。ケプストラム次数ごとの認識率を調べたところ、次数が 12 のときに最も高い認識率となった。また、全共分散 HMM の場合と比較すると、ケプストラム次数が低次では認識率は低くなるが、高次まで利用すると対角共分散 HMM の結果がよくなった。このことから、共分散行列の非対角要素を 0 にすることで、HMM のパラメータ数が減少し、このことがパラメータ推定精度を上げ、高次での認識率の向上につながったと考えられる。

この 2 種の HMM を比較することにより、ケプストラム分析の次数が低次のときは全共分散、高次のときには対角共分散 HMM の場合に認識率が高くなるのがわかったので、低次のケプストラム同士の相関が重要であると推測される。そこで共分散行列の $1 \sim k$ 次までを全共分散、それより高次を対角共分散とする部分全共分散行列に設定して認識実験を行った。分析次数 n を 10~18、 k を 3~7 まで変えながら学習/認識した場合、 $n = 15$ 、 $k = 4$ のときに最も高い認識率が得られた。 $k = 4$ で固定し、 n を変えて実験を行ったときの認識結果を表 3 に示す。この結果は、全共分散

表3 部分全共分散 HMM ($k = 4$) におけるケプストラム次数と認識率 (%) の関係Table 3 Dependency of recognition rates by partially full covariance HMM ($k = 4$) on cepstral analysis order.

次数 n	10	11	12						
認識率	92.9	93.2	94.2						
	13	14	15	16	17	18			
	94.6	94.6	95.1	94.6	94.3	94.6			

表4 部分全共分散 HMM ($l = 3$) におけるケプストラム次数と認識率 (%) の関係Table 4 Dependency of recognition rates by partially full covariance HMM ($l = 3$) on cepstral analysis order.

次数 n	10	11	12						
認識率	91.8	91.8	93.0						
	13	14	15	16	17	18			
	92.5	92.9	92.6	92.2	90.7	90.3			

HMM、対角共分散 HMM の場合よりも高い認識率を示していることから、ケプストラム低次同士の相関が重要であり、出力共分散行列に部分的制約を設けたモデルが認識率向上に効果をもつことが確認できた。

次に、逆に高次のケプストラム同士の相関を仮定し、共分散行列の $n - l + 1$ 次から n 次までを全共分散、それよりも低次を対角共分散とする部分全共分散 HMM で学習/認識を行った。 n を 10~18、 l を 3~7 まで変えて実験を行った結果、 $n = 14$ 、 $l = 3$ のときに最も高い認識率となった。 $l = 3$ で固定し、 n を変えたときの結果を表 4 に示す。低次同士の相関を考慮した結果 (表 3) と比較すると、認識率は全体的に低くなっており、対角共分散 HMM の認識率とほぼ同等の結果となった。このことから高次同士の相関を考慮しても効果をもたらさないことが確認できる。

図 1 に全共分散 HMM、対角共分散 HMM、部分全共分散 HMM による認識結果を示す。部分全共分散では、 k と l をそれぞれ 3~7 まで変えたときの認識率を表している。それぞれの認識率は各 HMM で分析次数 n を変化させたときの最も高い値である。そのときの n の値を同図の各バーの上部に示す。

4.2 話者ごとの結果と考察

次に認識対象話者 (計 5 名) の各話者に対する結果について考察する。各 HMM に対して認識率が最低となった話者についての認識率を示すと表 5 のようになる。対角共分散 HMM では特定の話者に対する認識率が悪くなるのに対し、低いケプストラム次数間で相関を仮定した部分共分散 HMM では、認識率が最低とな

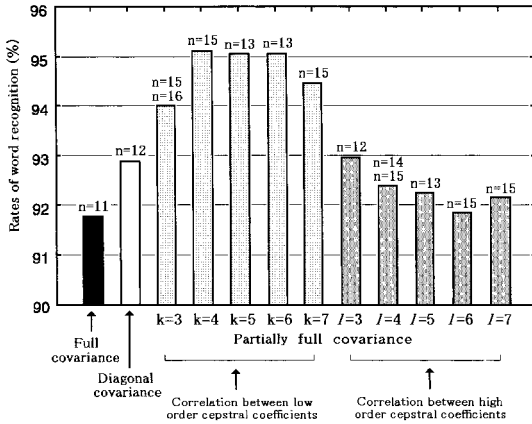


図1 HMM 出力確率共分散行列の制約条件に対する単語認識率の違い
 Fig.1 Word recognition rates for each covariance matrix of HMM output probability.

表5 各 HMM において認識率最低となった話者に対する単語認識率 (%)
 Table 5 Recognition rates of the lowest speaker for each HMM.

HMMの種類	全共分散	対角共分散	k = 3	k = 4	k = 5		
認識率	89.2	86.0	89.6	90.4	89.6		
	k = 6	k = 7	l = 3	l = 4	l = 5	l = 6	l = 7
	93.2	91.2	86.4	87.6	87.2	88.0	87.6

る話者に対する認識率が向上していることがわかる。これは $k = 6$ で最高の値を示す。話者によって認識率が変動することは当然であるが、最低の認識率を呈

する話者に対する認識率ができるだけ高くなることが望ましい。この点に関しても、分析次数の約半分の次数までの相関を考慮した部分共分散出力確率をもった HMM が効果的であると考えられる。

5. む す び

本論文では、HMM を用いた音声認識システムにおいて、学習外話者による音声の認識率向上を目的として、HMM 出力確率の共分散行列に着目し、これに制約条件を付加した場合の認識率を調べた。その結果、LPC ケプストラムを特徴として用いた場合、共分散行列の約半分の次数までを全共分散、それより高次を対角共分散とする部分全共分散 HMM を採用したモデルが効果的であることがわかった。今後、更に混合数を複数に設定した場合の出力確率について比較・検討する予定である。

文 献

- [1] 中川聖一, 確率モデルによる音声認識, 電子情報通信学会, コロナ社, 1988.
- [2] X.D. Huang, Y. Ariki, and M.A. Jack, "Hidden Markov models for speech recognition," Edinburgh University Press, 1990.
- [3] D. Rainton and S. Sagayama, "A new tied continuous mixture density HMM via orthogonalisation of the full covariance observation matrix," 日本音響学会講演論文集, 2-5-15, Oct. 1991.
- [4] 岸尾一史, 船田哲男, "単語音声認識における HMM 出力確率の共分散行列に対する制約条件の実験的考察," 日本音響学会講演論文集, 1-Q-11, March 2000.

(平成 12 年 8 月 8 日受付, 9 月 14 日再受付)