

論文 / 著書情報  
Article / Book Information

論題(和文)	Aggregated cross-validation尤度を用いた混合ガウス分布最適化アルゴリズムの提案
Title(English)	
著者(和文)	篠崎 隆宏, 古井 貞熙, 河原 達也
Authors(English)	Takahiro Shinozaki, SADAOKI FURUI
出典(和文)	日本音響学会2008年春季講演論文集, Vol. , No. 2-10-1, pp. 67-68
Citation(English)	, Vol. , No. 2-10-1, pp. 67-68
発行日 / Pub. date	2008, 3

## Aggregated cross-validation 尤度を用いた 混合ガウス分布最適化アルゴリズムの提案\*

©篠崎 隆宏 (東工大), 古井貞熙 (東工大), 河原 達也 (京大)

### 1 はじめに

有限の学習データから学習されるモデルの精密さと汎化性のバランスをとることは、統計に基づくモデル推定における一般的な問題である。一般に、モデルのパラメタ数が増えればより複雑な確率分布を表現できるようになるが、他方でモデルの一般性が失われていき、新しいデータに対して期待した性能が得られない結果となる。本研究の目的は、統計的パターン分類に幅広い応用を持つ混合ガウス分布の新しいデータに対する性能を最大とするよう、モデル構造の最適化を行うことである。

これまでに、クロスバリデーション (CV) 尤度を用いて混合ガウス分布の要素数およびその配置を最適化する手法を提案し、混合ガウス分布 HMM を用いた大語彙音声認識への応用において情報量基準 MDL を用いた場合よりも高い性能が得られることを示した [1]。CV は学習データのみを用いながらバイアスの少ないモデル評価を可能とする有効で一般的な手法であるが、これまで混合分布の最適化には応用されてこなかった。これは主に、単純な方法では計算量が非現実的に大きくなってしまふことが理由だったと思われる。我々が提案したアルゴリズムでは  $K$  個 (数十程度) に分割した学習セットの各区画に対してあらかじめ求めた十分統計量を用いることで、効率的な CV 尤度の評価を可能としている。

CV の効果を最大とするためには、学習セットの  $K$  個の区画への振り分けをランダムに行い、また各区画を同じサイズとすることが望ましいと考えられ、前回の研究においてもその様な実装を行った。しかしながら、最適化の過程で混合要素対の併合を行うたびに混合数の 2 乗のオーダーの多くのモデルが比較されることから、固定された区画化に依存して偶然に高い CV 尤度を与えるモデルが含まれる可能性が少なからずあると考えられる。このため、CV 尤度にも依然として正のバイアスが存在すると考えられ、モデル性能のさらなる向上の上で課題であった。

本研究では、Bagging [2] に似たアイデアを CV と組み合わせることで汎化性をさらに向上させる aggregated cross-validation 法 (Ag-CV 法) の提案を行い、混合ガウス分布 HMM の最適化に応用する。大

語彙連続音声認識実験において提案法がモデルの汎化性の向上に有効であり、CV を用いた混合分布の最適化よりも高い認識性能が得られることを示す。

### 2 Aggregated cross-validation (Ag-CV) 法

$K$ -fold クロスバリデーション ( $K$ -fold CV) 法では、図 1 に示すように、 $K$  個に区画化した学習セットの 1 区画を評価用に取り除き、残りの  $K-1$  の区画を用いてモデルの学習を行う。そして得られたモデルを残しておいた 1 区画に適用し、評価を行う。これを  $K$  個全ての区画に対して行い、それらの結果を足し合わせたものを最終的な評価値とする。この操作により学習データ全体を有効に使いつつ、モデル学習と評価におけるデータの重複を避けることができ、バイアスの少ない評価値を得ることができる。得られた評価値を基に学習データ量に応じた、汎化性能とモデルの精密さとのバランスのとれたモデル構造を選択することが出来る。

他方 Bagging では与えられた学習セットから複数のサブセットをランダム抽出し、それぞれのサブセットを用いて複数のモデルを学習し、識別時にはそれらのモデルを並列実行し結果を統合することで、全体としての汎化性能を向上させる。

今回提案する aggregated cross-validation 法 (Ag-CV 法) では図 2 に示すように、各 CV モデルの学習に Bagging と同様なランダムサンプリングによる複数モデルの学習を導入する。具体的には  $K$  個中  $i$  番目の区画を除いた  $K-1$  個の区画の中から、 $K'$  個の区画をランダムに非復元抽出することを  $N$  回行い、 $N$  個のモデルを学習する。得られたモデルを  $i$  番目の区画に対して並列に適用し、それら  $N$  個の結果を平均したものを  $i$  番目の区画に対する評価値とする。この操作を全ての区画  $i = 1, 2, \dots, K$  に対して行い、各区画に対する評価値を足し合わせたものを最終的な評価値とする。区画のランダム抽出を繰り返す際、 $N$  個のモデルが全て異なるようにするとすると、 $N$  の最大値は  $K-1$  個中  $K'$  個の要素を取り出す組み合わせ数となる。また、従来の CV は  $K' = K-1, N = 1$  とした特殊な場合として、Ag-CV に含まれる。

\* Gaussian mixture optimization based on aggregated cross-validation likelihood. by Takahiro Shinozaki (Tokyo Institute of Technology), Sadaaki Furui (Tokyo Institute of Technology), and Tatsuya Kawahara (Kyoto University)

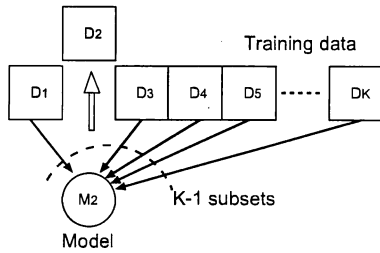


Fig. 1 Cross-validation

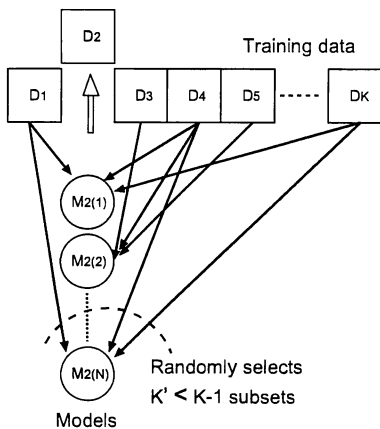


Fig. 2 Aggregated cross-validation

### 3 Ag-CV 法による混合分布最適化

Ag-CV 法を用いた混合ガウス分布の最適化は、Ag-CV 尤度を最大とするよう、与えられた混合ガウス分布の混合要素対を貪欲法により順次選択/併合することにより行う。具体的なプロセスは Ag-CV 尤度を用いる以外は前回提案した CV 尤度を用いた最適化 [1] と同様である。また、Ag-CV 尤度の計算も CV 尤度の場合と同様にして十分統計量を用いることで効率的に行うことが出来る。ただし、CV 尤度を用いた場合と比較して  $N$  倍のモデルの推定と評価が必要になるため、学習セットの分割数  $K$  が同じであれば、計算量は増加する。

Bagging と同様に、Ag-CV 法において  $i$  番目の区画に対して適用される  $N$  個のモデルは同一では意味がなく、適度に異なっている必要がある。このため、本研究では区画の選択数  $K'$  を予備実験より  $K' = \frac{1}{2}K$  となるように選んでいる。また、混合ガウス分布 HMM の最適化は、混合ガウス分布に対する最適化を各状態ごとに独立に適用することで実行される。

### 4 実験条件

音響モデルの学習セットとして、CSJ 学会講演から無作為に抽出した 30 時間の音声を用いた。1000 状態に状態共有された 1 混合トライフォンモデルを初

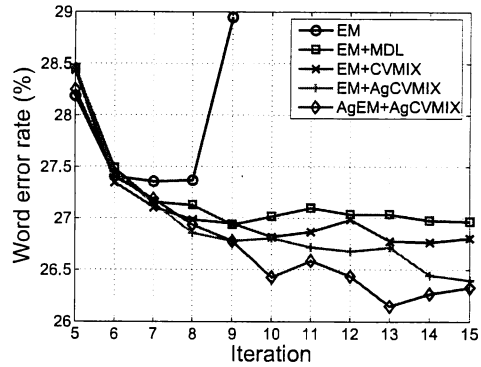


Fig. 3 Word error rates

期モデルとして、EM(5 回)+提案法を用いた混合要素の縮約最適化+混合要素の分割、の一連の操作を 1 学習ループとして、これを繰り返すことにより、混合ガウス分布 HMM の学習をおこなった。テストセットは CSJ 学会講演評価セット 10 講演である。詳細は文献 [1] と同様である。

### 5 実験結果

図 2 に EM と混合要素分割のみで混合分布の縮約最適化を行わない場合、MDL 法/ CV 法/ Ag-CV 法によりそれぞれ縮約最適化を行った場合、および EM 法の代わりに Ag-EM 法 [3] を Ag-CV 混合最適化と組み合わせた場合の単語誤り率を示す。CV による混合分布最適化は  $K = 30$ 、Ag-CV は  $K' = 6$ 、 $K' = 3$ 、 $N = 10$  として行った。Ag-EM 法は  $K = 12$ 、 $K' = 6$ 、 $N = 12$  とした。図の横軸は EM ループ 5 回と混合要素の操作を 1 つの学習ループとして数えたものである。EM のみのベースラインでは 1 ループごとにモデルの混合数が単純に 2 倍に増加するため、ループ数が増えると過学習が顕著となる。Ag-CV 法により混合分布の最適化を行うことで、混合数を自動的に制御しつつ、ベースラインの最小値と比較して相対的に 3.5% 誤り率が削減された。さらに Ag-EM と組み合わせることによって最大で 4.4% 誤り率が削減され、提案法の有効性が示された。

謝辞 本研究は科研費(19700167)の助成を受けたものである。

### 参考文献

- [1] 篠崎, 河原, 音講論(春), 41-42, 2007.
- [2] Breiman, Machine Learning, 24(2), 123-140, 1996.
- [3] 篠崎, Ostendorf, 河原, SP(2007-12), 223-228, 2007.