社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2006－SLP－64 (33)
2006／12／22

頑健なパラメタ推定のためのクロスバリデーションEM法の提案

篠崎 隆宏† Mari Ostendorf††

† 京都大学学術情報メディアセンター (南館)
〒606-8501 京都市左京区吉田二本松町
†† Department of Electrical Engineering, University of Washington
Box 352500, Seattle, Washington, U.S.A. 98195-2500
E-mail: †staka@ar.media.kyoto-u.ac.jp, ††mo@ee.washington.edu

あらまし EMアルゴリズムの欠点である過学習の問題を補うため、従来の自己尤度に代えてクロスバリデーション尤度を用いる新しい最尤学習アルゴリズムの提案を行う。並列化EM学習と同様に学習セットを区画化し、各区画ごとに求めたモデルの十分統計量を用いることで、提案手法は従来のEM学習と同程度の計算量で実行可能である。人工的なデータを用いた分析実験により、提案法が従来のEMアルゴリズムと比較して過学習に対して頑健であることを示す。中国語放送音声を用いた大語彙連続音声認識実験により、提案手法がEM学習と比較してより多くのパラメタを有効に活用し、単語誤り率の削減に有効であることを示す。
キーワード EMアルゴリズム, 過学習, クロスバリデーション

Cross-validation EM Algorithm for Robust Parameter Estimation

Takahiro SHINOZAKI† and Mari OSTENDORF††

† Academic Center for Computing and Media Studies [South Building], Kyoto University
Yoshida Nihonmatsu-cho, Sakyo-ku, Kyoto, 606-8501 Japan
†† Department of Electrical Engineering, University of Washington
Box 352500, Seattle, Washington, U.S.A. 98195-2500
E-mail: †staka@ar.media.kyoto-u.ac.jp, ††mo@ee.washington.edu

Abstract A new maximum likelihood training algorithm is proposed that compensates for weaknesses of the EM algorithm by using cross-validation likelihood in the expectation step to avoid overtraining. By using a set of sufficient statistics associated with a partitioning of the training data, as in parallel EM, the algorithm has the same order of computational requirements as the original EM algorithm. Analyses using a GMM with artificial data show the proposed algorithm is more robust for overtraining than the conventional EM algorithm. Large vocabulary recognition experiments on Mandarin broadcast news data show that the method makes better use of more parameters and gives lower recognition error rates than EM training.
Key words EM training, overtraining, cross-validation

1. はじめに

モデル学習において一般に問題となるのが過学習である。最尤推定法によりモデルを推定しようとする場合、学習セットに対する尤度はモデルのパラメタ数に対して単調に増加するが、同時に学習セットに対する尤度と新しいデータに対する尤度の解離も大きくなる。このためパラメタ数の増大にしたがって、学習セットに現れない新しいデータに対するモデルの一般性が失われてしまう。この問題に対処するため、これまでに最小記述長基準 (MDL) やベイズ情報量基準 (BIC) といった尤度とパラメタ数に応じたペナルティのバランスをとる情報料基準や、クロスバリデーションやブートストラップなどのデータドリブンな方法が試みられてきた [1]。

隠れマルコフモデル (HMM) のようにモデルが隠れ変数を含む場合、最尤推定問題を直接解くことが難しくなるため、繰り返し学習法であるEM法が広く用いられている [2]。EM学習の各学習ステージでは、モデルに含まれる隠れ変数の共起確率分布を初期モデルと学習データが与えられた条件で推論し、その分布に基づいた学習セットに対する期待対数尤度を最大とするようにパラメタを更新する。EM法を用いる場合においても

モデルの精度と新しいデータに対する一般性のバランスをとるために、パラメタ数を調整することは重要である。加えて、混合ガウス分布を用いるようなモデルにおいては、有限のパラメタ数でも学習過程において尤度が無限大に発散する不安定性の問題が存在する。例えば、2つのガウス分布からなる2混合の混合ガウス分布の場合、片方のガウス分布が学習セット全体をカバーし、もう片方のガウス分布が学習データ中のある一点を非常に小さい分散でカバーするようにすると、いくらでも大きな尤度が得られてしまう。しかしながら、そのようなモデルは学習セットに特化し過ぎているため、新しいデータに対する一般性の観点からは性能の低いモデルとなる。パラメタ推定の際にとり得る分散に下限値を設定することでこの問題はある程度避けることが可能であるが、設定する下限値が大き過ぎればモデルの性能を制限することになり、小さ過ぎれば不安定性の問題を取り除くことができなくなるため、実験に基づいた調整が必要となる [3]。また情報料基準は、これまでに一定の条件の元ではHMMのモデルの選択に応用されて有効性が示されているものの [4], [5]、尤度が発散し得るような条件では原理的に意味をなさず不安定性の問題に対しては無力である。

不安定性の問題も広い意味で過学習の極端な場合と見ることが出来るが、これら過学習の問題はモデルパラメタの推定と尤度を用いたモデルの評価を同じデータを用いて行うことに起因しており、パラメタ推定と評価に用いるデータを分離することで解決することが出来る。クロスバリデーションは別途評価用のデータを用意することなく学習データのみを用いて効果的にデータの分離を行える手法であり、広くその有効性が認められている。しかしながら、計算量の問題もあり、これまでクロスバリデーションは基本的に学習アルゴリズムの外側で学習結果のモデルの比較に用いられる場合がほとんどであった。クロスバリデーションをモデルの学習アルゴリズム内部に組み込む試みとては、HMMの状態クラスタリングに用いる決定木をクロスバリデーション尤度に基づいて作成する研究が挙げられる [6]。決定木は尤度を基準に質問を順次選択することにより成長させるが、クロスバリデーション尤度を用いることで従来の自己尤度を用いた方法と比べてよりよいHMMの状態共有構造が得られ、また決定木のサイズを自動的に決定出来ることが示されている。

本研究では、クロスバリデーションをEMの学習アルゴリズム枠組内に効果的に組み込むことにより従来のEMアルゴリズムの過学習に対する脆弱性を改善する、新しい学習アルゴリズム、クロスバリデーションEM(CV-EM)の提案を行う。クロスバリデーション尤度を用いた決定木学習アルゴリズムが基本的に尤度をモデル選択に用いるのに対し、本論文で提案するアルゴリズムはクロスバリデーション尤度をモデル選択ではなく隠れ変数の期待値計算に使用する点が、最大の相異点である。提案手法は十分統計量を用いたEMアルゴリズムの並列化技術を応用することで、従来のEMアルゴリズムと同程度の計算量で実行可能である。

実験ではまず提案手法の基本的な振る舞いを分析するため、無作為に設定した混合ガウス分布からランダムに抽出したデータを用いて様々な条件でモデルの学習を行い、学習データとは独立に抽出したテストデータに対する尤度の評価を行った。分析では、CV-EMにより学習されたモデルは従来のEMによるモデルと比較して、常に同等かより大きな尤度を与えることが示された。一般にEMアルゴリズムでは学習ループの繰り返しとともに学習データに対する尤度は単調増加が保証されているが、過学習の問題が存在するためにテストデータに対しては最適な繰り返し数が存在する。CV-EMはEMと比較して過学習に対して頑健であり、さらに最適な繰り返し数を自動的に決めることが可能である利点を持つ。ついで、中国語放送音声をタスクとしてCV-EMを大語彙連続音声認識に応用し、認識率による評価を行った。認識実験の結果、CV-EMはモデルのサイズに対して安定した性能を与えるとともに、認識率の向上に有効であることが示された。

以下 **2.** で提案アルゴリズムについて説明し、**3.** で人工データによる混合ガウス分布を用いた分析、および **4.** で実際の中国語放送音声を用いた認識実験について述べる。**5.** において未決問題の提起および今後の課題について議論し、最後に **6.** でまとめを行う。

2. クロスバリデーションEM学習

EMアルゴリズムは不完全データに対する最尤推定問題を実質的に完全データに対する推定問題に変換する非常に一般的な枠組であるが、指数分布族に属する確率分布については十分統計量を用いた以下の2つのステップを繰り返すことにより、効率的に実行することが出来る。

- *E-step:* 観測値 O と初期モデルパラメタ $\theta^{(p)}$ が与えられた条件で求めた隠れ変数の共起確率分布をもとに、期待十分統計量 $t^{(p)} = E[t|O, \theta^{(p)}]$ を計算する。
- *M-step:* 得られた統計量を基に、全ての変数において観測値が与えられた場合と同様に最尤法により、モデルパラメタ $\theta^{(p+1)}$ を更新する。

(例として、混合ガウス分布の場合の期待十分統計量 $t^{(p)}$ は、現在のモデル $\theta^{(p)}$ を用いて計算した、要素ガウス分布毎の事後占有確率の和および事後確率で重みづけされた一次および二次の観測値の平均である。) このアルゴリズムの問題点は、E-stepとM-stepで同じ学習データを用いるため、ある段階で一部のサンプルが特定の要素ガウス分布に対して高い尤度を持つと次の段階ではさらに高い尤度が割り当てられるなど、過学習の悪循環に陥りやすいことである。提案手法であるCV-EMの中心となる概念は、E-stepにおける期待十分統計量の計算対象となるデータとM-stepにおけるモデルパラメタ推定において用いられるデータの分離である。期待十分統計量の計算において用いられるモデルの学習に使用されるデータと、期待値計算の対象となるデータが独立しているため、潜在的な過学習の問題を軽減することが出来る。

2.1 アルゴリズム

CV-EMの学習手順は、並列化EM学習のそれと類似している [7]。並列化EM学習 (図1) では、プロセスを並列化し学習ターンアラウンド時間を短縮するために、学習データは K 個

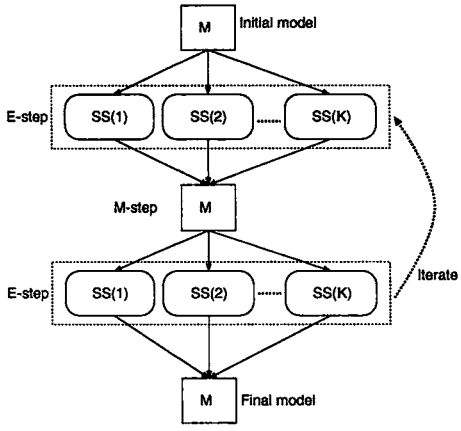

図 1 Parallel EM training. SS(i) denotes the sufficient statistics for the i-th data subset.

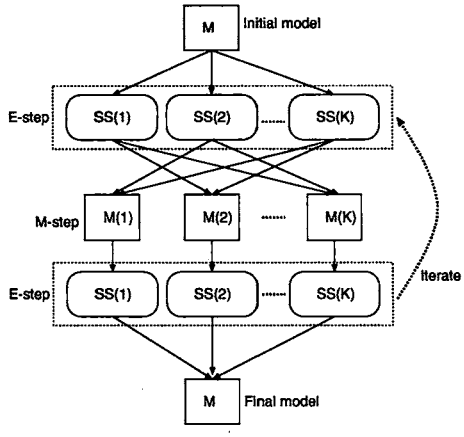

図 2 CV-EM training. M(i) denotes the i-th CV model estimated without using the i-th data subset.

の部分集合に区分化される。そして、E-step において十分統計量は各学習サブセット毎に独立に計算される。サブセット毎の統計量を一つに統合化した後、M-step においてモデルパラメタの更新が行われる。図 2 に示す CV-EM アルゴリズムにおいては、学習データの区分化は E-step と M-step で使用するデータを分離しつつ効率的にクロスバリデーションモデル集合を求めることを目的として行われる。具体的には、初期モデルを用いた第一回目の E-step は従来の並列化 EM とまったく同じであり、K 個の部分学習セットそれぞれに対して十分統計量が計算される。次に M-step において、学習部分集合に対して全ての十分統計量を統合して一つのモデルを推定する代わりに、統合の際どれか一つの部分集合を除外することで K 個のクロスバリデーションモデルを生成する。引き続く E-step ではそれぞれのモデルはそのパラメタ推定の際に除外した十分統計量に対応する学習サブセットに対して隠れ変数の共起確率を評価し、新しい期待十分統計量を求めるために使用される。従来の EM 法と同様に E-step と M-step は繰り返し行われ、最終的なモデルは全ての十分統計量を統合することにより出力される。

学習セットをモデルパラメタ推定用と隠れ変数の共起確率計算用に分割する CV-EM には、実質的な学習セットサイズの減少に伴うパラメタ推定結果のバイアスや分散増大の潜在的な問題が存在する。しかしながらこの問題の影響は、K 次クロスバリデーションにおいて K の値を大きくとることにより、十分小さくすることが可能である。これは CV-EM において各クロスバリデーションモデルが、学習セット全体の $(K-1)/K$ のデータを用いて推定されるためである。反対に CV-EM では学習データ分割の効果により、パラメタ推定において学習データに依存した楽観的で自信過剰なバイアスと分散を減らすことが出来る。これは E-step と M-step で用いられるデータが独立しているため、学習過程において確率分布が特定のサンプルに依存して高い尤度を得ることが出来なくなるためである。よって全体として、K 次クロスバリデーションを EM の繰り返しループ中に組み込む CV-EM を用いることにより過学習の問題が軽減され、モデル性能の向上が期待できる。また EM 学習のもう一つの問題点である解の局所性についても、大域最適解への収束の保証は依然ないものの、CV-EM では個々のサンプル点に依存したモデル探索尤度空間の凹凸が軽減すると考えられることから、有利に働くと予想される。

2.2 アルゴリズム動作の詳細

一般に K 次クロスバリデーションは K 個のモデルを必要とするため、もしこれらのモデルが個別に初めから作られるとするならば、単純にモデルを作成する場合と比較して K 倍の計算量が必要となることになる。しかしながら CV-EM では十分統計量を用いることで、クロスバリデーションモデルの作成を効率的に行っている。学習データのサイズが K と比較して十分大きい場合の CV-EM の計算量は、ほぼ E-step において K 個の学習サブセットに対し十分統計量を求めるために必要な計算量である。さらに従来の並列化 EM と CV-EM における E-step の違いは、各学習サブセットにおいて同じモデルを用いるか違うモデルを用いるかのみであり計算量としては同じであるため、全体として CV-EM の計算量は従来の EM の計算量と同程度となる。CV-EM のストレージコストは、各学習サブセットに対して計算された十分統計量を保持するためのコストが支配的であり、クロスバリデーションの次数 K に比例する。

CV-EM は学習過程において隠れ変数を含む複数のモデルを作成し、異なるモデルを用いて計算した十分統計量を統合する処理を繰り返す。この際に潜在的な問題として、個々のモデルが隠れ変数について異なる解釈を行い、結果として十分統計量の統合が意味をなさなくなる危険が考えられる。この問題の具体例としては、もし複数の混合ガウス分布が独立に学習された場合、各混合ガウス分布に属する個々の要素ガウス分布間には対応関係がなく、それらの統計量を意味のあるかたちで統合することは出来ない。CV-EM においてこの問題は、アルゴリズムを同一のモデルで初期化し、大きな K を用いることで避けることが出来る。これは CV-EM の各学習ステージにおいて任意のクロスバリデーションモデルのペアはそれぞれの学習セッ

トの $(K-2)/(K-1)$ を共有するため、K が大きければ全てのモデルはほぼ共通のデータから学習され (例として K が 21 の場合は 95%)、モデル間でパラメタの推定結果が大きく異なる可能性は低く、隠れ変数間の対応が保存される為である。

3. 人工データによる分析

3.1 実験条件

提案手法の基本的な振る舞いを調べるため、無作為に設定した 4 次元 8 混合の混合ガウス分布からランダムに抽出した学習用データと評価用データを使った実験を行った。学習セットのサイズやクロスバリデーションの次数を変えたさまざまな条件でモデルの学習を行い、テストセットに対する尤度を求めることで評価を行った。モデルの学習は、学習データ全体の平均と分散を基に初期化した 8 混合のガウス分布に対し、従来法の EM および提案法の CV-EM をそれぞれ適用することにより行った。テストセットのサンプル数は 1000 である。実験結果から偶然性を取り除くために、各実験条件において独立に抽出したデータを用いた実験を 100 回繰り返し、平均をとった値を最終的な評価値として用いた。

3.2 学習データ量および繰り返し回数とモデル性能

図 3 にさまざまな学習セットサイズに対して EM および CV-EM を適用して学習した混合ガウス分布の評価用データに対する尤度を、学習の繰り返し回数とともに示す。CV-EM を用いた学習における、クロスバリデーションの次数 K は 10 である。図において零番目の繰り返しは、初期モデルを用いた結果を示している。また、CV-EM 学習では各学習段階において、$K-1$ 個の十分統計量から作成される K 個のクロスバリデーションモデルとともに、K 個全ての十分統計量を統合した一般モデルを作成し、尤度の評価に用いた。EM 学習と CV-EM 学習において同じ初期モデルが用いられた場合、繰り返し学習の一回目の結果は、EM と CV-EM で E-step が同じため同一となる。両者の違いは 2 回目以降に現れる。

図に見られるように CV-EM 学習によるモデルは EM 学習によるモデルと比較して、常に同程度以上の尤度を与えることが分かる。CV-EM のもっとも顕著な利点は大きな学習繰り返し数に対する頑健性である。EM アルゴリズムでは各学習段階において学習セットに対する尤度が単調増加することが保証されているが、このことは評価データに対しては当てはまらない。特に学習データサイズが小さい場合、EM 学習はモデルパラメタを特定の学習データに特化させる傾向が強く、学習の繰り返しが進行するとともにモデルの一般性が失われ易い。結果として、評価用データに対する尤度は学習の繰り返しに対し初めのうち増加するが最適点が存在し、その後は過学習のために減少に転じる。図に見られるように、CV-EM においても過学習の影響は存在するものの、従来の EM 学習に比べてはるかに安定していることが分かる。学習の最適繰り返し数が未知の場合や、HMM のような複合モデルにおいて要素分布が異なる最適繰り返し数を持つような場合には、大きな学習回数を安全に指定することが出来ることから学習回数に対する頑健性はモデル性能の向上に貢献する。

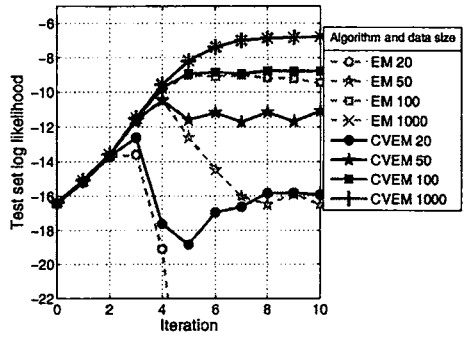

図 3 Test set likelihood of GMMs trained by EM and CV-EM with varying training set sizes.

過学習の影響は一般に、パラメタ数に比較して多くの学習サンプルを用いてモデルを推定する場合には少なくなる。実際図より、データ量の増大とともに EM と CV-EM のどちらを用いても正確なモデルの推定が行われるようになるため、両者の性能差が減少することが分かる。

3.3 学習時の尤度

図 4 に学習過程において E-step で得られた EM および CV-EM の学習セットに対する尤度を示す。E-step は並列化されているため、学習サブセットごとに得られた尤度を平均して示している。図において零番目の繰り返しは、初期モデルを用いた一回目の E-step で得られる尤度である。従来の EM アルゴリズムと異なり、CV-EM の場合には尤度は単調増加とならない。EM の場合には学習セットサイズが小さくなるにつれて尤度が増大するが、CV-EM の場合は減少する。言い替えると、EM の場合データ量の減少とともに E-step 尤度と評価セットに対する尤度の差が増大するのに対し、CV-EM の E-step 尤度は評価データに対する尤度と同じ傾向を示す。EM の E-step 尤度が学習セットサイズの減少とともに増大するのは、モデルの推定と尤度評価に同一のデータを用いているために尤度評価のバイアスが増大し、自信過剰となるためである。これに対して CV-EM の E-step 尤度は各学習サブセットに対し、そのサブセットのデータを用いずに推定されたモデルを用いたクロスバリデーション尤度であり、より信頼性の高い評価となっているためである。学習データ量の増大とともに EM と CV-EM の E-step 尤度は同じ値に向かって収束していくが、これはデータ量が大きければモデルパラメタおよび尤度が、学習手法にかかわらず適切に推定されるようになるためである。CV-EM の信頼性の高い E-step 尤度を用いることで、CV-EM においては最適な学習繰り返し数を自動決定することも可能である。

3.4 クロスバリデーション次数

図 5 にクロスバリデーション次数 K の、モデル性能に対する影響を示す。この実験では同じ学習セットに対し、異なる K の値を用いて GMM を学習し、評価セットに対する性能を求めた。CV-EM の学習繰り返し数は 10 である。クロスバリデーションの次数が小さい場合、実効的な学習セットサイズが減少

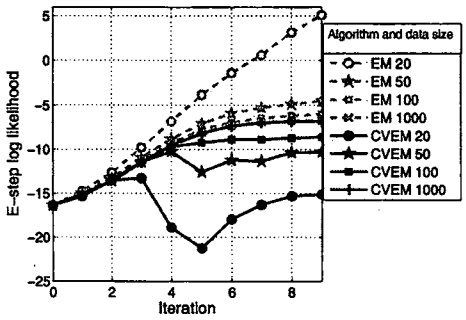

図 4 Training set likelihood obtained in the E-step.

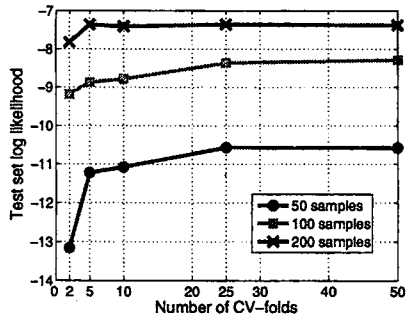

図 5 Number of cross-validation folds and the model performance.

しモデルの推定精度が悪化すること、およびクロスバリデーションモデル間で共有される学習データが減少し隠れ変数間の対応を保持するのが困難になることから、モデル性能は低くなる。しかしながら、図より次数 K を 20 程度以上とすれば、K の値にかかわらず安定した性能が得られることが分かる。

4. 大語彙連続音声認識

4.1 実験条件

中国語 Hub4 および TDT4 の放送ニュース音声を音響モデルの学習セットとして用いた大語彙連続音声認識実験を行った。学習セットのデータ量は合計で約 97 時間である。認識率の評価は中国語 RT-04 の開発セットおよび評価セットを用いて行った。開発セットおよび評価セットは放送音声で、データ量はそれぞれ約 30 分および一時間である。実験は中国語音声用に構成された Decipher 認識システムを用いて行った [8]。認識システムの構成にはいくつかのバリエーションが存在するが、本研究ではトライグラム言語モデルおよび最尤学習により学習した音響モデルを用いたシステムを使用している。発音辞書の語彙数は 49k である。音響モデルは単語内状態共有トライフォンの混合重みを全て展開した GENONE モデルである。母音はトーンタイプに応じてそれぞれ 4 通りの音素ラベルを割り当ててそれらを個別にモデル化しており、音素セット全体のサイズは 70 である。音響モデルの学習の際、Decipher システムの標準構成に従い、小さなバリアンスフロアリング閾値 (10^{-20}) を用いている。認識システムは声道長正規化および MLLR による話者適応を行っている。認識結果の評価には文字誤り率を用いている。ベースライン音響モデルは並列化 EM により学習している。並列化 EM の並列数および CV-EM のクロスバリデーションの次数 K はともに 21 である。CV-EM 学習において各学習サブセットが統計的に出来るだけ均一となるよう、学習サブセットは学習セット全体を無作為に並べ替えた後に区分化して定義した(注1)。モデルの学習は、EM 学習および CV-EM 学習とも同じ HMM を初期モデルとし、それぞれ 5 回の繰り返しを行うことにより行った。

表 1 CER for development set and evaluation set

Model		Dev04 set		Eval04 set	
# states	# mixes	EM	CV-EM	EM	CV-EM
2500	32	8.8	9.2	18.9	18.9
5000	32	9.3	9.1	18.9	18.8
1800	128	9.1	8.8	18.8	18.4
3000	128	9.4	8.9	18.7	18.6
6000	128	9.9	9.6	19.7	18.6

4.2 実験結果

表 1 に開発セットおよび評価セットに対する文字誤り率 (CER) を示す。探索ビーム幅などのシステムパラメタは、開発セットを用いてベースラインシステムに対して調整された値をそのまま提案手法である CV-EM により学習したモデルを用いたシステムにも使用している。表より、CV-EM が評価用データにおいて EM と比較して常に同じかより低い誤り率を与えていること、およびモデルのパラメタ数に対して EM より頑健であることが分かる。開発セットにより最適なモデルサイズを選択した場合、EM の代わりに CV-EM を用いることで評価用セットにおける誤り率が 18.9% から 18.4% へ減少し、相対値で 2.6% の誤り率を削減することが出来た。この結果は、NIST の Matched Pairs Sentence-Segment Word Error (MAPSSWE) test を用いた場合、$p = .021$ で統計的に有意である。

5. 考察と課題

本論文において示したクロスバリデーションを用いてモデルの過学習を軽減するアイデアは十分統計量を用いて並列化できる繰り返し学習アルゴリズムであれば、EM 以外にも識別学習などに広く応用することが出来る [9], [10]。識別学習が尤度に基づく最適化と誤り率による評価の不整合を取り除くことを目的としているのに対し、クロスバリデーションを用いた学習は過学習などパラメタ学習時の推定誤差を取り除くことを目的としているため、両者は相補的に働くと考えられる。

(注1)：CV-EM に用いる学習サブセットの定義において、クロスバリデーションが話者独立に行われるようにするため、同一の話者からの音声が全て同一のサブセットに配分されるような制約を設けた実験も行ったが、話者を考慮しない場合と比べて文字誤り率に有意な差はなかった。

CV-EM の学習プロセスは、クロスアダプテーション [11] に幾分類似していると言える。クロスアダプテーションが同じ音声データに対して複数の音響特徴量を抽出するなどして異なる観点から統計量を計算して用いるのに対し、CV-EM では同じ特徴量を用いながら異なる学習サブセットを設定することで複数の統計量を計算し用いることなどが両者の違いである。CV-EM のアイデアを話者適応の変換行列 [12] を求めるために適用し、クロスアダプテーションと性能の比較をすることも出来るであろう。

CV-EM の E-step において得られる尤度は混合ガウス分布を用いた実験で示したように、評価セットに対する尤度と似た動きをとり、学習の繰り返しに対して単調とはならない。これは尤度の評価が公正であるためであり、学習セットのみを用いて最適な繰り返し数を決定することが可能となるなど CV-EM の長所ではあるが、従来の EM の収束性に関する議論が少なくともそのままでは成り立たないことを示している。学習データ量が大きくかつクロスバリデーションの次数も十分に大きい場合には、過学習や不安定性の問題が小さくなるとともに各クロスバリデーションモデルが学習セットのほぼ全体を用いて学習されるために CV-EM の結果は EM の結果に漸近していくと考えられ、実際両者が漸近していくことが GMM を用いた分析で示された。学習データ量が少ない場合、本論文では直感的な説明とともに CV-EM が EM と比較してよりよい結果を与えることを実験で示したが、理論的な特徴づけは今後の課題である。CV-EM の収束性に関して関係しそうな文献としては、各学習ステップで尤度増加の保証がない場合についての繰り返し学習アルゴリズムの収束条件について述べている文献 [13] などが挙げられる。

近年研究が進められている変分ベイズ法 [14] は評価データを用いずにモデルの構造選択が行えるなど、クロスバリデーションを用いたモデル学習と似た効果が得られることが示されている。クロスバリデーションを用いたモデル学習の変分ベイズ法に対する利点としては、全てがデータ駆動であり、直感的あるいは経験的な事前分布の設定が不要であること、十分統計量によって表現できる限り尤度以外の目的関数を用いた学習アルゴリズムに対しても適応が可能であることなどが挙げられる。

今後の課題としては、CV-EM を話者適応などさまざまな用途に応用することや、過学習に対する頑健性をより高めることなどが挙げられる。頑健性の向上のためには、学習の繰り返しの際に E-step 尤度が減少する場合にはパラメタの更新を行わない、あるいは前段の推定結果やパラメタ数の少ないモデルとのインターポレーションを各学習段階においてクロスバリデーション尤度を用いて行うことなどが、考えられる。

6. ま と め

パラメタ推定を過学習に対して頑健に行うため、クロスバリデーションを EM アルゴリズム内に組み込んだ CV-EM の提案を行った。十分統計量を活用することで、提案手法の実行に必要な計算量は従来の EM とほぼ同程度である。アルゴリズムの挙動を調べるため人工的なデータを用いて混合ガウス分布の学習と評価を行い、CV-EM が評価セットに対し従来の EM と比較して常に同等以上の尤度を与え過学習に対して頑健であることを示した。また、CV-EM の E-step で得られる尤度は EM で得られる尤度と異り信頼性が高く最適な繰り返し学習回数の自動決定が可能であることを示した。さらに提案手法を混合ガウス分布を用いた大規模な HMM の学習に応用し、従来法と比較してモデルパラメタ数の増加に対して頑健であり、認識誤り率を削減出来ることを示した。

謝　辞

本研究の大部分は第一著者が Department of Electrical Engineering, University of Washington (Seattle, Washington, U.S.A) および International Computer Science Institute (Berkeley, California, USA) に滞在中に、行われた。本研究は契約番号 HR0011-06-C-0023 により DARPA の支援を得て行われた。配布における制限はない。本論文の見解は著者のものであり、資金を提供した機関の見解を反映するものではない。

文　献

[1] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning", Springer-Verlag, (2001).

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", J. of the Royal Statistical Society, Series B 39, No. 1, pp. 1–38 (1977).

[3] H. Melin, J. W. Koolwaaij, J. Lindberg, and F. Bimbot, "A comparative evaluation of variance flooring techniques in HMM-based speaker verification", Proc. ICSLP, Sydney, pp. 2379–2382 (1998).

[4] S. Chen and P. S. Gopalakrishnan, "Clustering via the Bayesian Information Criterion with applications in speech recognition", Proc. ICASSP, pp. 645–648 (1998).

[5] K. Shinoda and T. Watanabe, "Acoustic modeling based on the MDL criterion for speech recognition", Proc. EuroSpeech, pp. 99–102 (1997).

[6] T. Shinozaki, "HMM state clustering based on efficient cross-validation", Proc. ICASSP, Toulouse, Vol. 1, pp. 1157–1160 (2006).

[7] S. Young *et al.*, "The HTK Book", Cambridge University Engineering Department (2005).

[8] M. Y. Hwang, X. Lei, W. Wang, and T. Shinozaki, "Investigation on Mandarin broadcast news speech recognition", Proc. ICSLP, pp. 1233–1236 (2006).

[9] L. R. Bahl, P. F. Brown, P. V. de Souza, and R. L. Mercer, "Maximum mutual information estimation of hidden Markov model parameters for speech recognition", Proc. ICASSP, pp. 49–52 (1986).

[10] D. Povey and P.C. Woodland, "Minimum phone error and i-smoothing for improved discriminative training", Proc. ICASSP, Vol. 1, pp. 105–108 (2002).

[11] H. Soltau, B. Kingsbury, L. Mangu, D. Povery, G. Saon, and G. Zweig, "The IBM 2004 conversational telephony system for rich transcription", Proc. ICASSP, vol. 1, pp. 205–208 (2005).

[12] C. J. Leggetter and P. C. Woodland, "Flexible speaker adaptation using maximum likelihood linear regression", Proc. Eurospeech, pp. 1155–1158 (1995).

[13] A. Gunawardana and W. Byrne, "Convergence theorems for generalized alternating minimization procedures", Journal of Machine Learning Research, Vol. 6, pp. 2049–2073 (2005).

[14] S. Waterhouse, D. MacKay, T. Robinson, "Bayesian Methods for Mixture of Experts", Proc. NIPS8, pp.351-357, (1995).