日本神経回路学会 第 9 回全国大会 講演論文集 pp.145-146, 札幌 (1999)

SMEM アルゴリズムによる NGnet の学習

NTT コミュニケーション科学基礎研究所　鈴木敏 (PY)　上田修功

Learning NGnets by SMEM algorithm

NTT Communication Science Laboratories　Satoshi Suzuki　Naonori Ueda

E-mail: satoshi@cslab.kecl.ntt.co.jp

Abstract

Recently the SMEM algorithm has been proposed to overcome the local optima problem associated with the EM algorithm for mixture models. In this report, we present experimental results of application of the SMEM algorithm to Normalized Gaussian networks (NGnet) and show the advantage over the conventional learning algorithm.

1　はじめに

筆者の一人は，混合モデルにおける EM アルゴリズムの局所最適性の問題を解決するための一手法として SMEM(Split and Merge EM) アルゴリズムを先に提案し，混合潜在変数モデルを含む混合分布推定問題に対しその有効性を示した [1]．本稿では，混合回帰モデルである正規化ガウス関数ネットワーク (NGnet)[2] に SMEM アルゴリズム適用し，SMEM アルゴリズムの回帰問題での有効性を示す．

2　NGnet と EM アルゴリズム

NGnet は，

$$y = \sum_{i=1}^{M} \frac{G_i(x|\mu_i,\Sigma_i)}{\sum_{j=1}^{M} G_j(x|\mu_j,\Sigma_j)} f_i(x;W_i)\ , \qquad (1)$$

$$G_i(x|\mu_i,\Sigma_i) \equiv (2\pi)^{-N/2}|\Sigma_i|^{-1/2} \times \exp[-\tfrac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)]$$

により与えられ，形式的には Jordan らの提唱した混合回帰モデル (通常，Mixture Expert network (MEnet) と呼ばれる) の gating network をガウス関数に置き換えたものと見なせる．但し，MEnet と NGnet の本質的な違いは，gating 関数の違いではなく，NGnet では入力変数を確率変数として取り扱っている点である．

通常，回帰問題では，入力は単なる変数と考えられる為，NGnet が MEnet よりも性能面において優れているとは言えないが，NGnet を用いると EM アルゴリズムの M ステップにおける最適解の計算が効率よく実行できる場合がある [2]．

モデル数 M，観測データ数 N，入力 x，出力 y，ネットワークパラメタ $\Theta = \{(\mu_i,\Sigma_i,W_i)\}_{i=1}^M$，モデル指標 i としたときの NGnet に対する EM アルゴリズムの Q 関数は，ステップ t でのパラメタ推定値を $\Theta^{(t)}$ とすると，

$$\begin{aligned} Q(\Theta|\Theta^{(t)}) &= \sum_{n=1}^{N}\sum_{i=1}^{M}[P(i|x,y,\Theta^{(t)}) \\ &\quad \times \log\{P(y|x,i,\Theta)P(x|i,\Theta)P(i|\Theta)\}] \\ &= \sum_{i=1}^{M} q_i(\Theta|\Theta^{(t)}) \end{aligned} \qquad (2)$$

と，$\{q_i\}_{i=1}^M$ の直和で表される為，SMEM アルゴリズムが適用できる．

ここで，$P(y|x,i,\Theta)$,$P(x|i,\Theta)$ はそれぞれモデル i により推定される出力分布および入力分布で，ガウス分布を仮定している．また，$P(i|\Theta)$ はモデル選択の事前確率で一様分布を仮定する．このとき，各モデルでの $f_i(x;W_i)$ による変換を線形変換とし，観測データ (x,y) を同時分布と見なせば，EM アルゴリズムの M ステップにおける最適解の計算が閉形式で求められ，効率の高い学習が可能となる．

3　SMEM アルゴリズム

SMEM アルゴリズムは，EM アルゴリズムの学習過程において，モデル要素の併合分割操作により，局所最適解からの脱出と，より良い解への誘導を図るアルゴリズムで，概要は以下の通りである．

1．通常の EM アルゴリズムを実行．(Θ^*, Q^* を収束後のパラメタ推定値および Q 関数値とする)
2．Θ^* に基づき，併合分割候補 $\{i,j,k\}$ の優先順位 c を併合分割基準により定める．
3．$c = 1,\cdots,c_{max}$ に対し，以下を実行．
   3-1. 併合分割を実行し，パラメタの再学習を行う．(Θ^{**}, Q^{**} を収束値とする)
   3-2. $Q^{**} > Q^*$ ならば，$Q^* \leftarrow Q^{**}, \Theta^* \leftarrow \Theta^{**}$ とし，2．へ．
4．Θ^* を最終結果として終了．

但し，2．の併合分割候補の計算は以下に従う．

<u>併合基準</u>：分布推定問題の併合基準で用いたモデル i への帰属確率 $P(i|x,\Theta)$ を，回帰問題では $P(i|x,y,\Theta)$ として置き換える．即ち，各データのモデル i への帰属確率 $P(i|x,y,\Theta)$ から成るベクトルを P_i として，

$$J_{merge}(i,j) = \left[\frac{P_i}{\|P_i\|}\right]^T \left[\frac{P_j}{\|P_j\|}\right] \qquad (3)$$

キーワード：SMEM アルゴリズム，回帰問題，NGnet

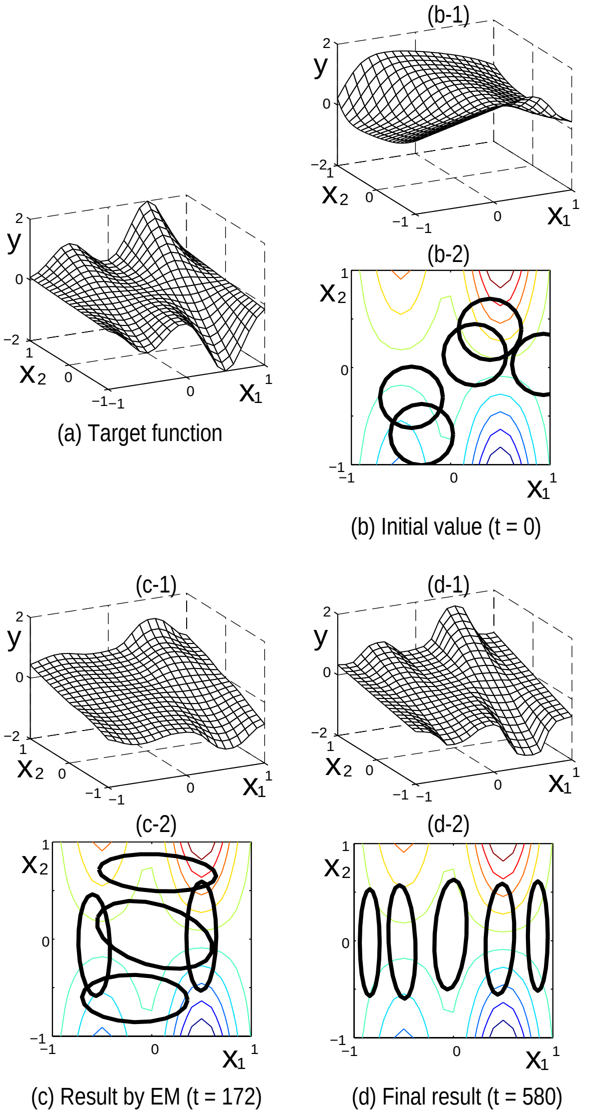

図 1: 学習過程の例

とする． J_{merge} の大きい組が高い優先順位を持つ．
分割基準：分布推定問題の分割基準で用いたモデル k による推定分布 $P(x|k,\Theta)$ および事後確率で重み付けした局所経験分布 $f_k(x)$ を，回帰問題では $P(x,y|k,\Theta)$ および $f_k(x,y)$ として置き換える．即ち，入力の局所経験分布 $f_k(x,y)$ と推定分布 $P(x,y|k,\Theta)$ との KL 情報量

$$J_{split}(k) = \int f_k(x,y) \log \frac{f_k(x,y)}{P(x,y|k,\Theta)} dx \qquad (4)$$

を用いる．但し，

$$f_k(x,y) = \frac{\sum_{n=1}^{N} \delta(x-x_n)\delta(y-y_n)P(k|x,y,\Theta)}{\sum_{n=1}^{N} P(k|x_n,y_n,\Theta)} \qquad (5)$$

である． J_{split} の大きいモデルが高い優先順位を持つ．

4 計算機実験

計算機実験では人工データを用いて，3 次元曲面の推定問題について検討した．以下，モデル数を 5 とした場合の実験結果を示す．

表 1: アルゴリズムの比較

	対数尤度		計算量	
	EM	SMEM	EM	SMEM
平均	-1739	-1426	52	662
標準偏差	182	124	35	477
最大値	-1372	-1226	156	2561
最小値	-2625	-1749	6	27

図 1 に学習過程の例を示す．(a) は目的関数で，2 次元入力から 1 次元出力を与える関数である．学習に用いるデータは入力空間から任意に選ばれた 1000 点を用い，出力範囲の 5% を標準偏差とするガウスノイズを出力に対して加えたものを用いた．(b) は学習の初期状態を表している．図 (b-1) は初期値として任意に与えられたパラメタにより推定された関数である．図 (b-2) では各モデルで推定された入力分布が太線で表されており，背後の細線は目的関数の等高線を示している．

(c) は通常の EM アルゴリズムによる学習結果である．この例では，EM ステップを 172 回繰り返すことにより学習が収束している．推定された関数は目的関数に近づいているが，十分とは言えない (c-1)．また，入力分布の推定も，変化の大きい領域を入力領域とするモデルがあり，最適な入力分布とはいえない (c-2)．

(d) は SMEM アルゴリズムによる学習結果で，受理されなかった EM ステップも含め，合計 580 回の EM ステップを要した．(c) に比べ，より良い解が得られていることが分かる (d-2)．モデル毎の入力分布の推定も，平面に近い部分が入力領域となるように学習され，妥当な推定結果である (d-2)．

通常の EM アルゴリズムと SMEM アルゴリズムとを比較した結果を表 1 に示す．これは，それぞれのアルゴリズムで 100 回の異なる初期値からの学習結果を用い，対数尤度および計算量 (総 EM ステップ数）の平均，標準偏差，最大値，最小値を比較した結果である．SMEM アルゴリズムによる尤度の大幅な改善が認められる．但し，計算量は，平均で約 10 倍を要した．

5 おわりに

SMEM アルゴリズムの NGnet への適用について検討した．従来の EM アルゴリズムに比べ，総ステップ数は増大するものの，推定精度は確実に向上しており，有効性が確認された．今後，高次元実データでの検証も行う予定である．

参考文献

[1] 上田，中野．(1999)．混合モデルのための併合分割操作付き EM アルゴリズム．信学論 D-II, vol.J82-S-II, no.5, pp.930-940.

[2] Xu,L. Jordan,M.I. and Hinton,G.E. An Aleternative Model for Mixtures of Experts. *NIPS 7*. The MIT Press, Cambridge, MA.