非定常データからの動的パラメータ推定とモデル選択に関する研究

数理情報学専攻 48096217　**櫻井瑛一**
指導教員　**山西健司 教授**

1　はじめに

本研究では非定常なデータに対するモデル選択を考える．ここで非定常なデータとは時間区間ごとでは定常であるが，各時間区間での情報源は複雑度が異なる情報源から出力されていることを考える．近年，モデルの変化は起きるものと仮定し，変化するモデルの系列を推定するということが考えられるようになってきた．

本研究は変化するモデル系列を求めるという**動的モデル選択**（Dynamic Model Selection，以下 DMS と略す）の枠組み [5] を用いる．[5] では変化するモデル系列を予測的 MDL(Minimum Description Length) [2] により，最適なモデル系列を選択する動的モデル選択規準を考え，その規準を満たすモデル系列を求める手法を確立した．

しかし，動的モデル選択では二つの問題が存在する．一つは，モデルに付随するパラメータの変動があったとき，いかに追随し推定を行うかについて考えてこられなかったという問題である．他方は，モデルが変化する事は過去を全て説明できるモデルが入れ替わるときである (switching) としており，変化点を境に完全にモデルが切り替わる場合を扱えていないという問題である．

本研究では二つの問題を解消するアルゴリズムの設計と解析を与えた．前者に対してはパラメータが区間的に変化しても，これに追従できる動的パラメータ推定を提案し，従来の動的モデル選択と組み合わせて効果を生み出す手法を与えた．後者に対しては直近の変化点からパラメータを初期化して推定しなおし，モデル系列を選択する resetting 型のアルゴリズムを提案した．

2　動的モデル選択

本研究での基礎となる動的モデル選択の定義を行う．観測されたデータ列を $x^T=\{x_1,\dots,x_T\}$ とし，それの生成源として，確率モデルクラスを考える．$(\mathcal{C}_k=\{P_k(X|\theta)\},\mathcal{C}_{k-1}\subset\mathcal{C}_k,(1\le k\le K))$ を考える．ここでの，k はクラスの複雑さをあらわす指標であり，今後は k をモデルの次数と呼ぶ．そして，モデルクラス $\mathcal{C}_k$ に対する $t-1$ までの出力から学習した時刻 t での逐次的なパラメータ推定値を $\theta_k^{(t)}$ とする．そして，モデルは遷移し，その確率は次のように定まるとする．

$$P(k_1)=1/K,$$
$$P(k_{t+1}|k_t)=\begin{cases}1-\alpha & ,\text{if } k_{t+1}=k_t, k_t\neq 1\text{or}K,\\ 1-\alpha/2 & ,\text{if } k_{t+1}=k_t, k_t=1\text{or}K,\\ \alpha/2 & ,\text{if } |k_{t+1}-k_t|=1.\end{cases}$$

そして，α の推定を変化の頻度で行う．モデルの系列 $k^T=\{k_1,\dots,k_T\}$ の下での記述長とモデルの系列自体の記述長 $\ell(k^T)$ を合わせた総記述長を $\ell(x^T,k^T)$ としたときに，MDL の二段階符号化による定義にならい，動的モデル選択規準を次のように定義する．

定義 1. モデル系列 k^T の下で総記述長を

$$\ell(x^T,k^T)=\sum_{t=1}^{T}\Big(-\log P_k(x_t|\theta_k^{(t)}) \\ -\log P(k_t|k^{t-1}:\alpha(k^{t-1}))\Big)$$

とする．この量を最小とするモデル系列 k^T を選択することを**動的モデル選択**と定義し，その選択の規準となる総記述長 $\ell(x^T,k^T)$ を**動的モデル選択規準**と定義する．

パラメータが過去データ全てに依るため，モデル遷移は過去を含め今をより良く説明するモデルが現れた時点で発生する．それゆえ，**switching 型動的モデル選択**とよぶ．そして，最適なモデル系列を求めるアルゴリズムが動的計画法により構成でき，次が成立する．

定理 2. 定義 1. の規準を最小化するモデル系列は $O(KT^2)$ で求まり，その総記述長は次で抑えられる:

$$\ell(x^T,k^T)\le\min_m\Bigg\{\min_{\boldsymbol{t},\boldsymbol{k}}\sum_{j=0}^{m}\sum_{t=t_j}^{t_{j+1}-1}-\log P\left(x_t|\theta_{k(j)}^{(t)}\right) \\ +(T-1)H\left(\frac{m}{T-1}\right)+\frac{1}{2}\log(T-1) \\ +(m+1)\log 2+\log K\Bigg\}.$$

ただし，$\boldsymbol{t}=\{t_0,\dots,t_{m+1}\}$ はモデルの変化時刻を表し，$\boldsymbol{k}=\{k(1),\dots,k(m)\}$ で $k(j)$ は時間区間 $[t_j,t_{j+1}-1]$ の間でのモデルをあらわす．

3 本研究での問題設定

本研究では特に，出力 y_t を補助入力 x_t から予測関数 $f(x,\theta)$ を用いて予測を行う逐次予測問題を考える．その予測損失を凸な関数 $L(y,z)$ を用いてはかり，動的モデル選択を考える際には，その損失から確率を定め $P(L)$ とする．その例として，論文では二乗ロスとガウス分布を用いた AR モデルを考え数値実験を行った．

4 動的パラメータ推定

Swithcing 型動的モデル選択のパラメータ推定法 $\theta_k^{(t)}$ には真の構造変化に対しての追従性が必要になる．そこで，連続型集合型予測枠組みと [1] を組み合わせ，以下のようなアルゴリズムを構築した． この結果，次が

Algorithm 1 動的パラメータ推定

Initialization

$0<\eta\le\eta^*, 0<\alpha<1$, $\pi_1^s(\theta)=\pi_1(\theta)$, $\int\pi_1(\theta)\mathrm{d}\theta=1$.

for $t=1$ to T **do**

$\hat{y}_t=\int\pi_t^s(\theta)/\Pi_t^s f(x_t,\theta)\mathrm{d}\theta$, $\Pi_t^s=\int\pi_t^s(\theta)\mathrm{d}\theta$.

$\tilde{\pi}_t^m(\theta)=\pi_t^s(\theta)\mathrm{e}^{-\eta L(y_t,f(x_t,\theta))}$,$\tilde{\Pi}_t^m=\int\tilde{\pi}_t^m(\theta)\mathrm{d}\theta$.

$\pi_{t+1}^s(\theta)=(1-\alpha)\tilde{\pi}_t^m(\theta)+\alpha\pi_1^s(\theta)\tilde{\Pi}_t^m$.

end for

成立する．

定理 3. 累積予測損失の上界は次を満たす:

$$\sum_{t=1}^{T}L(y_t,\hat{y}_t)\le\min_{\boldsymbol{t},m}\left(-\frac{1}{\eta}\left(\sum_{i=0}^{k}\log Z[t_i,t_{i+1}-1]\right.\right.$$
$$\left.\left.+(T-1)\left(H\left(\frac{m}{T-1}\right)+D\left(\frac{m}{T-1}\Big|\Big|\alpha\right)\right)\right)\right),$$
$$Z[t_i,t_{i+1}-1]=\int\pi_1(\theta)\mathrm{e}^{-\eta\sum_{t'=t_i}^{t_{i+1}-1}L(y_{t'},f(x_{t'},\theta))}\mathrm{d}\theta.$$

ただし，$\boldsymbol{t}=\{t_0,\dots,t_m,t_{m+1}\},(1=t_0<t_1<\cdots<t_{m+1}=T+1)$ であり，$0\le m\le T-1$ である．

5 動的モデル選択の拡張

前章での動的なパラメータ推定では過去の影響を受けていたが，直近の変化点 t_c から推定しなおす $\theta_k^{(t_c,t)}$ 場合を考える．このような推定を行う動的モデル選択を **resetting 型動的モデル選択**と定義する．これにより，switching 型とは違い，局所に最良なモデルを選択できる．

ここで，パラメータ推定として最尤推定量に漸近する推定を行うとする．その場合，switching 型と動的なパラメータ推定を組み合わせた場合と比べ，前章で明らかになった追従性に必要なエントロピーのコストだけ上界が下がる．一方計算量は変化頻度と変化時点を記憶するため $O(KT^3)$ となる．そこで，データ圧縮における研究 [3,4] を援用し，変化点に依存する遷移確率推定を用いる動的モデル選択アルゴリズムを提案した．これにより計算すべき状態が減り，計算量は $O(KT^2)$ となる．その上界は，[3] を用いれば次のように書ける．

定理 4. [3] を応用した resetting 型 DMS は遷移確率を

$$\hat{\alpha}(k^t)=\frac{\pi(t-t_c+1)}{Z_\infty-Z_{t-t_c}},\pi(t)=\frac{1}{t^{1+\epsilon}},Z_n=\sum_{j=1}^{n}\pi(j)$$

で推定し，求めた k^T の下で上界は以下を満たす:

$$\ell(x^T,k^T)\le\min_m\left\{\min_{\boldsymbol{t},\boldsymbol{k}}\sum_{j=0}^{m}\sum_{t=t_j}^{t_{j+1}-1}-\log P\left(x_t|\theta_{k(j)}^{(t_j,t)}\right)\right.$$
$$+m\log\frac{T}{m}+\epsilon(m+1)\log\frac{T}{m+1}$$
$$\left.+(m+1)\log(1+\epsilon)-m\log\frac{\epsilon}{2}+\log K\right\}.$$

6 まとめ

Switching 型動的モデル選択で構造変化への追従性が要求されその手法を提案した．また，モデルの推移が限定されていたため，resetting 型を提案した．そして，switching 型と動的パラメータ推定を組み合わせた場合と resetting 型と逐次最尤推定を組み合わせた場合の情報論的・計算論的限界を導いた．その結果，両者を計算論的には同じにした場合，情報論的上界は前者 $<$ 後者となり，後者の有意性を理論的に保証した．

参考文献

[1] M. Herbster and M. K. Warmuth. Tracking the best expert. In *Proceedings of the 12th International Conference on Machine Learning*, pp. 286–294. Morgan Kaufmann, 1995.

[2] J. Rissanen. Universal coding, information, prediction, and estimation. *IEEE Transaction on Information Theory*, Vol. 30, pp. 629–636, 1984.

[3] G. I. Shamir and N. Merhav. Low complexity sequential lossless coding for piecewise stationary memoryless sources. *IEEE Transactions on Information Theory*, Vol. 45, pp. 1498–1519, 1999.

[4] F. M. J. Willems. Coding for a binary independent piecewiseidentically-distributed source. *IEEE Transactions on Information Theory*, Vol. 42, pp. 2210–2217, 1996.

[5] K. Yamanishi and Y. Maruyama. Dynamic model selection with its applications to novelty detection. *IEEE Transactions on Information Theory*, Vol. 53, pp. 2180–2189, 2007.