【土木計画学研究・講演集 No.22(2) 1999年10月】

修正情報量基準に基づくモデルの自動選択方法に関する研究*

A Study on Automatic Model Selection based on a Modified Information Criterion

紀伊雅敦**，土井健司***

Masanobu KII, Kenji DOI

1. はじめに

モデル選択においては適合度検定，予備検定，修正重相関係数，情報量基準などの様々な手法が開発されてきた．特に情報量基準はその理論的な背景が明快であり，計算が簡便であることから広く用いられている．

情報量基準は期待される誤差を最小とするモデルを選択する基準であることから，パラメータ推定においてその基準を最小とする目的関数を設定することが効率的である．例えば非分離型の確率モデル族ではモデルパラメータが零か非零かによって異なるモデルであるため，このような目的関数を設定できればパラメータ推定とモデル選択を同時に行うことが可能性である．しかし，従来定義されている情報量基準ではパラメータ数は離散値であることから，そのままではパラメータ推定の目的関数として用いることはできない．

ところで，情報量基準には異なる観点から導かれたいくつかの基準が存在する．代表的な3つの情報量基準についてまとめたものを表―1に示す．これらの第1項は全て-2×[最大対数尤度]であり，第2項はモデルパラメータ数と定数の積で与えられている．しかし実問題への適用において，予測誤差を最小とするための基準としてどのような定数を用いるべきかは明らかではない．

Watanabe[4]は，パラメータ数をパラメータに関して微分可能な関数とし，かつ従来の情報量基準を一般化した形式の修正情報量基準を提案している．但しその基準は未知の定数パラメータを含んでおり，その特定は感度分析により行われている．

本研究では，Cross Validation により検証されるモデルの平均対数尤度の平均値を期待平均対数尤度と見なすことで，未知の定数パラメータを推定する方法について新たに提案する．

表―1　従来情報量基準とその一般形

	式形	理論的根拠
AIC[1] Akaike's Information Criterion	$-2L(\hat{\theta})$ $+2m$	真の分布とモデルの分布との Kullback-Leibler 情報量の漸近的不偏推定量
BIC[2] Baysian Information Criterion	$-2L(\hat{\theta})$ $+m\log n$	モデルの漸近的な事後確率の対数
MDL[3] Minimum Description Length	$-2L(\hat{\theta})$ $+m(\log n+1)$	モデルの推定誤差の情報量とモデル自体を記述するために必要な情報量の和
一般形	$-2L(\hat{\theta})+A\cdot m$	

但し，$L(\hat{\theta})=\sum_{i=1\sim n}\log p(\hat{\theta};y_i\,|\,\mathbf{x}_i)$

$\hat{\theta}$；モデル $p(\theta;y|\mathbf{x})$ の最尤推定量

m；パラメータ数，n；サンプル数

以下，2章において Watanabe の提案する修正情報量基準について説明し，3章においてその定数パラメータの特定化を行う方法を示す．

3. 修正情報量基準[4]

Watanabe は非分離型であり非階層的な確率モデルの族からのモデル選択を自動的に行うための情報量基準として次式を提案している．

$$I_a(\theta)=2n\,L_{emp}(\theta)+A\,F_a(\theta) \tag{1}$$

但し，$$L_{emp}(\theta)=-\frac{1}{n}\sum_{i=1}^{n}\log p(\theta;y_i\,|\,\mathbf{x}_i) \tag{2}$$

$$F_a(\theta)=\sum_{i=1}^{m}f_a(\theta_i) \tag{3}$$

ここで $f_a(\theta)$ は次の2つの条件を満たす関数である．

(1) $f_a(\theta)\rightarrow f_0(\theta)$ when $a\rightarrow 0$

$$F_0(\theta)=\sum_{i=1}^{m}f_0(\theta_i)$$

$$f_0(\theta)=\begin{cases}0 & (\text{if } \theta=0)\\ 1 & (\text{if } \theta\neq 0)\end{cases}$$

* キーワード：計画手法論，モデル選択

** 学生員，工修，東京工業大学大学院　博士課程
（〒152 東京都目黒区大岡山2-12-1,
TEL 03-5734-2695, FAX 03-5734-3578）
*** 正員，工博，東京工業大学大学院情報環境学専攻

$$(2)\ |\theta_1| \le |\theta_2| \to 0 \le f_\alpha(\theta_1) \le f_\alpha(\theta_2) \le 1$$

この $f_\alpha(\theta)$の例として $f_\alpha(\theta)$=1-exp($-\theta^2/\alpha^2$)等があげられている. この $I_\alpha(\theta)$はθに関して2階微分可能であり, この情報量基準をパラメータ推定の目的関数として用いることが可能である. 但し, $I_\alpha(\theta)$は A=2, $\alpha\to 0$ の時, AIC であり, 予測誤差の理論上の期待値を最小化するものとなる. またA=log(n), $\alpha\to 0$ の時はBICであり, 期待される事後確率を最大化するものとなる. 次章において Cross Validation を用いた A の推定方法について提案する.

4. Cross Validation による情報量基準の選択

前述のように, モデル選択においては, 観点によって情報量基準は異なるパラメータ A を持つことが示されているが, どのような A を取るべきかは, 誤差の感度分析により試行錯誤的に決められていた. ここでは Cross Validation により推定されるモデルの尤度を期待平均対数尤度と見なして, 予測誤差を適切に表現するための, データに基づくA の推定方法について示す.

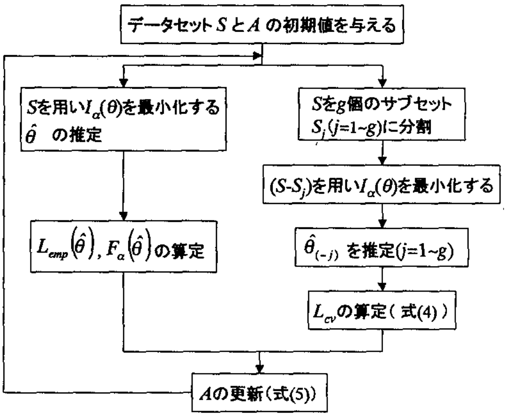
まず, 与えられたデータセット(S)を g 個のサブセットに分割し, そのj番目のサブセット(S_j)を除いたデータセット(S-S_j)を用い, $I_\alpha(\theta)$を最小とするパラメータの推定量を$\hat{\theta}(-j)$とする. このとき Cross Validation による平均対数尤度の期待値は

$$L_{CV} = -\frac{1}{g}\sum_{j=1}^{g}\frac{1}{n_j}\sum_{i\in S_j}\log p\left(\hat{\theta}(-j), y_i \mid \mathbf{x}_i\right) \qquad (4)$$

で与えられる. 但し, n_j は S_j のデータ数である. そこで I_{CV}=2$N\,L_{CV}$とおき, $I_{CV} \approx I_\alpha(\hat{\theta})$を仮定すると, 式(2)より

$$A = \frac{2n\left(L_{CV} - L_{emp}\left(\hat{\theta}\right)\right)}{F_\alpha\left(\hat{\theta}\right)} \qquad (5)$$

とおける. これは, L_{emp}と比較してL_{CV}が大きいとき, すなわち, データセット(S-S_j)に対するモデルの誤差と比較して(S_j)に対するモデルの誤差が大きい場合, パラメータ数で表される項の影響を大きくすることにより, モデルの汎化性能を向上させるための I_αを設定するものである. この A を用い再びパラメータを推定することで, 期待平均対数尤度のより高い$\boldsymbol{\theta}$を推定できることが期待される. 本手法を用いたA の推定のフローを図—1に示す.



図—1 本提案手法によるパラメータ推定の流れ

5. まとめ

本論では従来の情報量基準を一般化した形である修正情報量基準について示し, Cross Validation により対象データに適する定数パラメータの同定方法を示した. この方法は理論的に導かれたものではないが, 従来の手法と比較して, 予測誤差を小さくするようなモデルを効率よく選択することが可能であり, 実用的な方法であると考えられる.

本研究で提案した手法は, 非分離型の確率モデル族からのモデル選択全般に利用可能な方法である. 特にコネクショニストモデルのように考慮すべき変数が多数存在し, 比較するべきモデルの数が増大する場合に, 効率的にモデル選択を行うことが可能であり, また従来のパラメータ推定手法と比較してデータの持つ情報を生かし未知データに対する推定誤差を減少させるパラメータ推定を行いうることが期待される.

仮想データを用いたシミュレーション, 及び実データへの適用結果については講演時に発表する.

【参考文献】

1) H. Akaike; A new look at statistical model identification, IEEE Transactions on Automatic Control, vol. 19, pp.716-723, 1974

2) G. Schwarz; Estimating the dimension of a model, Annals of Statistics, vol. 6, No. 2, pp.461-464, 1978

3) J. Rissanen; Universal Coding, Information, Prediction, and Estimation, IEEE Transaction on Information Theory, vol. 30, pp.629-636, 1984

4) S. Watanabe; A Modified Information Criterion for Automatic Model and Parameter Selection in Neural Network Learning, IEICE Transaction on Information and systems, Vol. E78-D, No.4, 1995