

要 旨

時系列を用いて予測や計画を行う場合には、その内容を十分に表現する適切なモデルを決定することが重要である。本稿では、対象となる時系列を拡張標本自己相関を求めることによってパターン化し、学習を伴うパターン認識技法と情報量基準を用いることにより、その適切な自己回帰移動平均モデルを同定することを試みる。

1. 緒 言

時系列データは多くの分野で観測され、時間に関する順序づけや観測値間での相関性を有する特徴を利用して、主として予測や計画などに使用されている。このため、対象となる時系列を表現する適切なモデルを決定することは、効果的な予測などを行う上で重要となる。

広く使用されている時系列モデルは、 p 次の自己回帰(AR)と q 次の移動平均(MA)の2つの線形フィルタからなる自己回帰移動平均(ARMA)モデル¹⁾である。時系列モデリングとは、この p と q の次数を決定することであり、解法として多くの統計的なアプローチが提案されているが、操作において複雑性などを有する。

本稿では、種々の時系列をパターン化し、それぞれに適するARMAモデルを同定することを試みる。そのための手法として、線形識別関数に基づく学習²⁾と赤池の情報量基準³⁾を導入し、未知パターンに対する適応可能性の増大と、モデルの決定過程の複雑さの軽減を図る。

2. 時系列モデルとパターン認識

対象となる時系列モデルARMA(p, q)過程 Z_t は、以下の関係によって生成される。

$$\phi(B)Z_t = C + \theta(B)a_t \quad (1)$$

ここで、 $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ 、 $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ 、 B は $BZ_t = Z_{t-1}$ 、 $Ba_t = a_{t-1}$ のようなバックシフトオペレータ、 C は任意定数、 a_t は平均0、分散 σ_a^2 の白色雑音過程を表現し、 ϕ と θ は推定されるべきパラメータである。

パターン認識に必要なことは、異なるパターンを識別できるような特徴を得ることであり、それによって正しい分類が行われる。この意味で、パターン認識アプローチは、特徴抽出と分類の2つのステージからなる。これを時系列モデリングに適用する場合、パターンマッチング過程は以下の3つのステージを構成する²⁾。

- (1) ARMAモデルに対するプロトタイプパターンの学習。
- (2) ある時系列に特有の特徴の抽出。
- (3) その時系列パターンに最もよく照合するプロトタイプパターンを示しているARMAモデルへの分類。

図1には、このパターン認識アプローチの構成を示す。

3. 特徴抽出

一般に、特徴はある目的関数の最適化に基づいて選択される。本稿では、時系列についてある特定の値をもた

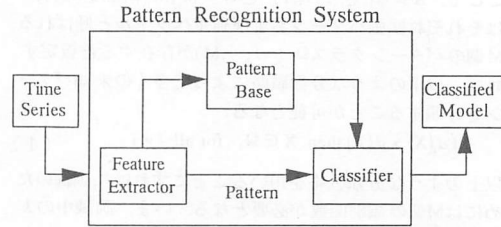


Fig.1 Formulation by Pattern Recognition

らす可能性がある拡張標本自己相関関数(ESACF)に基づく反復的な最適化法を用い、切断点¹⁾を利用して特徴を同時に選ぶ³⁾。直接的に自己相関関数を求めることが困難のため、以下の標本自己相関関数を使用する。

$$Y_k = \frac{c_k}{c_0}, \quad c_k = \frac{1}{n} \sum_{t=1}^{n-k} Z_t Z_{t+k} \quad (2)$$

ここで、 n は利用可能な観測値の数を示す。

実際、ESACFアプローチは、2次元テーブル上に p と q の種々の値に対応している拡張標本自己相関の集合を配置することによって特徴づけが可能である。いま、ESACF表の行と列が、それぞれAR過程とMA過程の次数を示すものとすれば、2つの標準偏差内のESACFは三角形を形成する。したがって、この三角形の頂点の行と列の座標は、それぞれ次数 p のARと次数 q のMAの推定値に対応する。

本稿では、あるパターンクラスはARMAモデルの1つを表現し、あるパターンは、特定の時系列のESACF値から変換したバイナリ数で構成される。このバイナリ数は、ESACF値が2つの標準偏差内にあれば0を割り付け、それ以外には1を割り当てる。ESACFアプローチによる三角形パターンの一例を表1に示す。この例は、理想的なARMA(1,2)モデルのプロトタイプパターンを表している。

Table 1. Pattern by ESACF approach

AR	MA							
	0	1	2	3	4	5	6	7
0	1	1	1	1	1	1	1	1
1	1	1	0	0	0	0	0	0
2	1	1	1	0	0	0	0	0
3	1	1	1	1	0	0	0	0
4	1	1	1	1	1	0	0	0

4. 線形識別関数に基づく学習

パターン認識問題を解決するためには、識別に基づくいくつかのルールを確立することが必要である。本稿では、対象クラスが線形分離可能であると仮定し、あるパターンを他のパターンから識別する方式として、以下のような線形識別関数を導入する。

$$d(\mathbf{X}) = w_1x_1 + w_2x_2 + \dots + w_nx_n + w_{n+1} = \mathbf{W}^T\mathbf{X} \quad (3)$$

ここで、 $\mathbf{X} = (x_1, x_2, \dots, x_n, 1)^T$ と $\mathbf{W} = (w_1, w_2, \dots, w_n, w_{n+1})^T$ はそれぞれ拡張パターンおよび荷重ベクトルと呼ばれる。M個のパターンクラス Ω_i ($i=1, \dots, M$) が存在すると仮定すれば、以下のような分類戦略によって多くの未知パターンを分類することが可能となる。

$$\text{if } d_i(\mathbf{X}) > d_j(\mathbf{X}) \text{ then } \mathbf{X} \in \Omega_i \text{ for all } j \neq i \quad (4)$$

以上のような分類戦略を用いることにすれば、訓練のためにはM個の識別関数が必要となる。いま、訓練中のk回目の反復ステップでクラス Ω_i に属するパターン $\mathbf{X}(k)$ がシステムに与えられると仮定すれば、M個の識別関数

$$d_i[\mathbf{X}(k)] = \mathbf{W}_i^T\mathbf{X}(k), \quad j=1, 2, \dots, M$$

が以下のように評価される。

$$\begin{aligned} \text{if } d_i[\mathbf{X}(k)] > d_j[\mathbf{X}(k)] \text{ then } \mathbf{W}_i(k+1) &= \mathbf{W}_i(k) \\ & \quad j=1, 2, \dots, M, \quad j \neq i \\ \text{if } d_i[\mathbf{X}(k)] \leq d_j[\mathbf{X}(k)] \text{ then } \mathbf{W}_i(k+1) &= \mathbf{W}_i(k) + c\mathbf{X}(k), \\ \mathbf{W}_i(k+1) &= \mathbf{W}_i(k) - c\mathbf{X}(k), \quad \mathbf{W}_i(k+1) = \mathbf{W}_i(k) \\ & \quad j=1, 2, \dots, M, \quad j \neq i, \quad j \neq l \end{aligned}$$

ここで、 c はある正の訂正増分である。

5. 訓練サンプルの選択と情報量基準 AIC

ここでは、各訓練サンプルのパターンベクトルは、表1と同じ形式をもつ特徴テーブルに対応するように選んだ。さらに、ARMAモデルの次数の範囲に制限を設け、それぞれに分類番号を対応づけた。

また、情報量基準 AIC は、統計的モデルの適切さの基準として真の分布との間の Kullback の情報量をとり、その推定値として導入されたもので

$$\text{AIC} = -2(\text{最大対数尤度}) + 2(\text{パラメータ数})$$

と定義される。モデル決定においては、種々のモデルのパラメータを最尤法で推定した場合、その中で AIC を最小にするものが最良のモデルとして選ばれる。真のモデルと近似モデルとの推定誤差が正規分布にしたがうものとする、標準偏差を σ_n 、データ数を N とすれば、パラメータ m の場合の AIC は以下のように表される。

$$\text{AIC}(p) = N \log \sigma_n^2 + 2p \quad (5)$$

6. 同定システムの動作手順

プロトタイプパターンの形成やマッチングの労力を軽減するために、以下のような方式を用いる。

1) 拡張標本自己相関から抽出した既知パターンを持つ多くの訓練サンプルが集められ、荷重ベクトルの更新

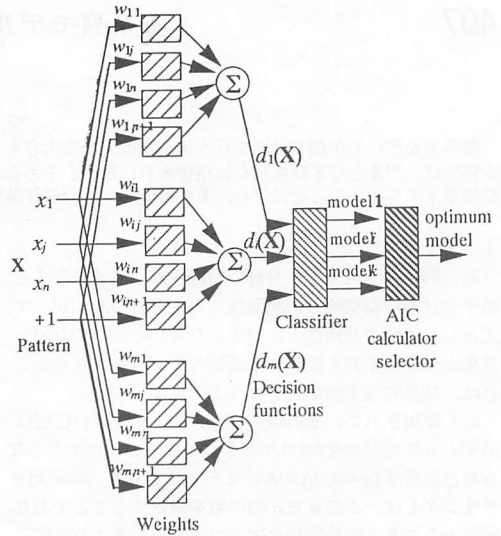


Fig.2 Basic frame of identification system

アルゴリズムが種々の識別関数を得るためにこれらのサンプルに適用される。この処理は、すべてのパターンが正確に分類されるまで続けられ、その結果として適切な識別関数が得られ、格納される。

- 2) 未知入力パターンに対し、そのパターンベクトルがまずESACFを用いて決定される。さらに、そのパターンベクトルの決定値が、すでに格納されているそれぞれの識別関数を用いて計算される。その結果として、大きな決定値を示す識別関数に関連する、いくつかのARMAモデルが選ばれる。
 - 3) これらのARMAモデルにAICを適用し、最小の値を持つモデルを最良のモデルとして選定する。
- 同定システムの基本的な枠組みを図2に示す。

7. 結言

拡張標本自己相関関数を用いて時系列をパターン化し、その特徴にパターン認識技法と情報量基準を適用することによって、適切なARMAモデルを同定するための一手法を示した。

[参考文献]

- 1) W.Vandaele : 時系列入門, 多賀出版
- 2) K.C.Lee and S.J.Park : Decision Support in Time Series Modeling by Pattern Recognition, Decision Support Systems, vol.4, pp.199-207, 1988
- 3) R.S.Tsay and G.C.Tiao : Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models, Journal of American Statistical Association, vol.79, no.385, pp.84-96, 1984