

適応フィルタによるノイズ除去と階層型HMMを用いた 経済時系列の認識

高木, 昇
九州大学大学院経済学府

<https://doi.org/10.15017/3000321>

出版情報 : 経済論究. 115, pp.37-47, 2003-03-15. 九州大学大学院経済学会
バージョン :
権利関係 :

適応フィルタによるノイズ除去と階層型HMMを用いた経済時系列の認識

高 木 昇

1 まえがき

時系列データを認識する統計的な手法として隠れマルコフモデル (Hidden Markov Model, 以下HMM) が用いられ、音声をはじめとする1次元データの解析に用いられている。その応用範囲についても、たんぱく質の配列、生体の信号処理など、実際のデータ処理へと拡大されている [1] [2]。これにともなって、本来のHMMが前提としている信号にガウスノイズが重畳している条件が満足されないケース (有色ノイズの重畳) が発生し、いくつかの拡張がなされている [3] [4]。

その中の有力な方法である逆フィルタ法では、有色ノイズと逆の特性をもつフィルタを通すことによりノイズ除去を行い、ノイズ除去された時系列にHMMを適用する。しかし、この方法では、ノイズが除去されるとともに、元の信号も変換を受けるため、新たに状態を再定義し、metastateを導入する必要がある [3]。更に、このmetastateの数は、元のモデルの状態数のべき乗に比例して増大するので、状態数が4以上になると実際の計算が困難となる。

著者らは、以前、ノイズ除去の方法を工夫することにより、基本的なHMMの手法が適用できる方法を提案した [5]-[7]。本報告ではこの手法を基本として用いる。基本的な手法は、以下の通りである [5]-[7]。有色ノイズを含んだ信号を予測フィルタに通してイノベーションを生成する。このイノベーションを用いて、信号を仮想的なノイズと考えて、もとの信号に含まれるノイズを特定する。こののち、ノイズを入力信号から差し引く処理をする。

適応フィルタを用いて時間域でノイズを除去する方法は、フーリエ変換などを用いて周波数域でノイズ除去を行う方法に比べて、処理にともなう新たなノイズ発生が少ないこと、元の信号に遅延を発生することもないこと、ノイズスペクトルの情報を必要としないなどの利点を有する [6]。

本報告の手法は、基本的に通常のHMMに帰着できるので、階層型のHMMを構成しさまざまな経済時系列の認識に応用する基礎的な考察を行う。

以下、2.では、適応フィルタを用いたノイズ除去による信号の抽出アルゴリズムについて述べる。3.では既知の発生メカニズムをもつケースについて、本論文の手法により実施された有色ノイズ除去およびHMM適用の例を示し、4.では階層型HMMによる経済時系列認識の基礎について述べる。

2 適応フィルタによるノイズ除去

2.1 適応フィルタの構成

以下では、ノイズは有色であり、ノイズのない信号に加算的に加えられていると仮定する。時刻 n について

$$x(n) = s(n) + \varepsilon(n) \tag{1}$$

ここで、 $s(n)$ はノイズのない信号、 $\varepsilon(n)$ はノイズ、 $x(n)$ は観測信号である。観測信号からノイズを除去するには、信号成分とノイズ成分を分離することが必要である。この場合、経験的に、信号の周波数変化は小さく、ノイズの周波数変化は大きいという性質を用いる。

適応フィルタを用いたノイズ除去の方法について、以下のように概要をまとめることができる。最初に、信号を予測誤差フィルタ (FIRフィルタ) に通してイノベーションを求める。すなわち信号をフィルタを通すことにより白色化する。この段階で、信号も変換を受けるが、これ以降の処理ではフィルタを通過した信号は利用しないので問題はない。

次に、このイノベーションを入力として、ノイズのスペクトルを推定する。通常の処理では信号のスペクトルを推定するが、ここでは、ノイズを仮想的な信号と考え、本来の信号を仮想的なノイズと考える。このスペクトルを推定する方法として、適応フィルタを用いる。通常、オンラインの信号処理ではFIRフィルタを用いて、実時間でこのスペクトルを推定する必要があるが、ここで取り上げる問題では実時間処理は必要ないので、FIRフィルタの設計問題をオフラインで計算することにする。これらの関係を図1に示す。

2.2 予測誤差フィルタの設計

予測誤差フィルタの設計には、通常の最小2乗予測フィルタの設計方法を用いる。ただし、入力に対して必ずしもオンライン処理は必要ではないので、入力を繰り返し用いて計算の精度をたかめる方法をとっている。このフィルタの出力をイノベーションとして用いて、後の段階でのノイズスペクトル

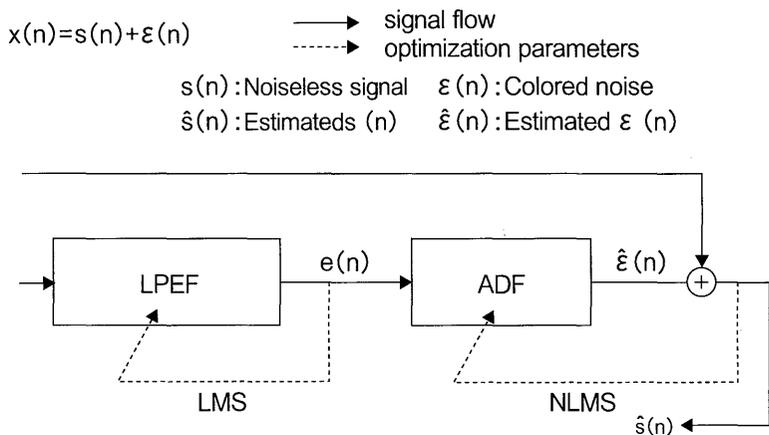


図1 ノイズ除去のシステム構成概要

ルの予測に用いる。

最初の、予測誤差フィルタを設計する方法に関しては、基本的には、次に示すようなFIRフィルタ係数の推定問題となる。

$$y(n) = \sum_{i=1}^M h_i(n)x(n-i) \quad (2)$$

$$e(n)^2 = [x(n) - y(n)]^2 \quad (3)$$

ここで、 $e(n)$ は予測誤差であり、これを最小にするように係数 $h_i(n), i=1, 2, \dots, M$ を決定する(時刻 n は、データ入力ごとに h_i を更新することを意味する)。係数の更新アルゴリズムは、予測誤差を最小化する方向に設定する。一般的には、次のように行われる [6]。

$$h_i(n+1) = h_i(n) + \mu_i x(n-i)e(n) \quad (4)$$

ただし、通常のノイズ除去フィルタにおいては、オンライン処理をすることが必要なため、係数をできるだけ早く、目的の値に近付けるように工夫がなされている。しかし、ここで考察しているケースでは、必ずしもオンラインで係数の計算をする必要がないので、線形適応フィルタの係数の計算は十分な回数実施することを仮定する。

具体的には、FIRフィルタの係数タップを調整し、誤差の系列の性質を観測しながら収束を判断する。すなわち、出力信号である $e(t)$ のスペクトルを計算し、これが平坦であるほど予測フィルタが良好であると判断する。具体的には、 $e(t)$ に対してFFTを適用し、これらの値に対する平均値からのずれ(分散)を求める。分散が小さくなるように適合度を高めていく。以下のシミュレーションでは係数タップの数を20にしておく。

2.3 ノイズスペクトルの推定とノイズ除去

イノベーションを入力として、ノイズスペクトルを推定する適応フィルタを設計する。この構成概要を図3に示す。この適応フィルタにおいては、FIRフィルタが用いられ検出されるであろうノイズな

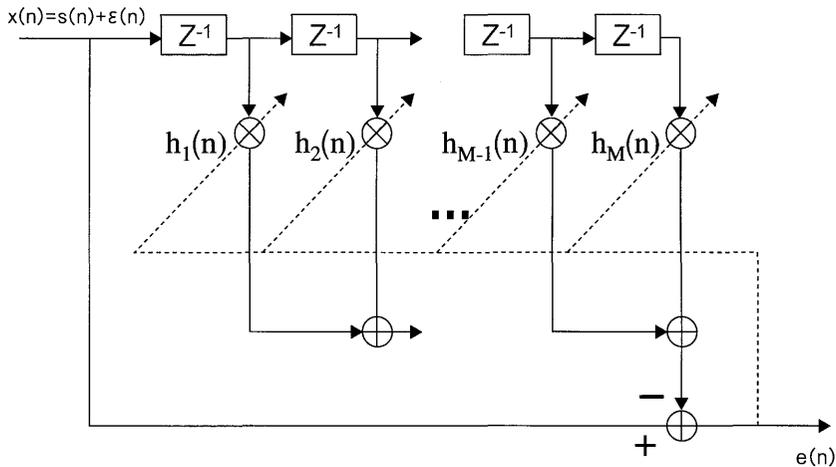


図2 予測誤差フィルタの構成

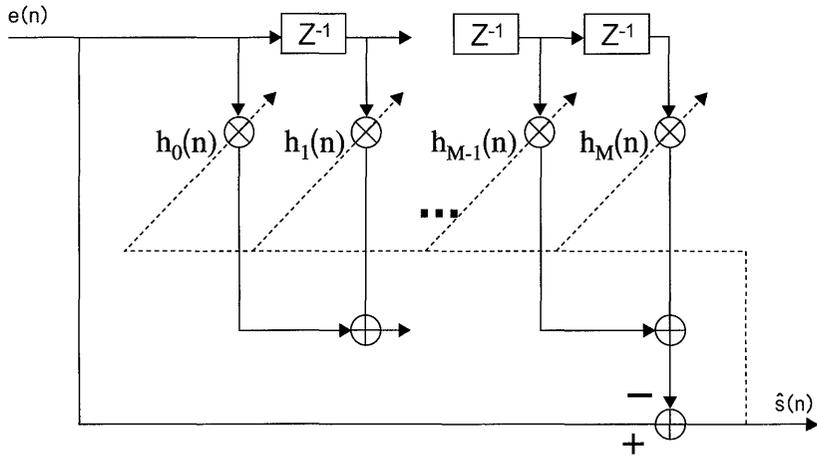


図3 イノベーションからの有色ノイズスペクトル推定

しの信号の予測値 $\hat{s}(n)$ を用いての調整が行われる。

しかし、信号とノイズの違いはあるが、基本的には線形予測フィルタを通過させることにより、ノイズを除去する手順と同じである。ただし、この場合、入力信号はイノベーションであり、一方、出力は有色ノイズである。

文献 [5] - [7] では、次のような方法を提案している。

$$g(n+1) = g(n) + \mu \frac{e(n)}{|e(n)|} \hat{s}(n) \tag{5}$$

ここで、 $e(n)$ は誤差として得られるイノベーションのベクトルである。

しかし、本論文ではスペクトルの性質を用いることにより係数を最適化している。すなわちノイズのスペクトルは広い周波数範囲にわたり滑らかとなっていると考えられ、これらを適合度とする評価を実施する。

すでに述べた信号からのイノベーション検出と同じアルゴリズムとなる。前に述べた原信号の白色化に対する線形フィルタの適用の場合と同様に、係数をオンラインで逐次近似する必要はないので、十分な逐次計算回数による最適化アルゴリズムを適用する。

この場合、係数最適化の評価としては、出力となる有色ノイズへの近似度合となる。ただし、有色スペクトルの形状は分かっていないので、次のような仮定をおいて近似する。すなわち、ノイズのない信号のスペクトルは、直線状ではないが平坦な形状をしており、これにスムーズな曲線を当てはめて、当てはめが最適であるほど適合度がたかいとする。このような信号を加えることにより、ノイズが除去された信号が結果として得られるようにしたい。そのためには、適応フィルタを通過した信号において、変動の大きな周波数成分が除去されていることが必要である。

3 階層型HMMの利用

3.1 経済時系列の特性

以下では、本報告で述べる適応フィルタによる有色ノイズ除去と通常の階層的HMMを利用した時系列認識について述べるが、最初に、主要な対象とする経済時系列の性質について整理しておく。本報告では、経済時系列として、株価や為替レートなどの金融時系列、自動車販売や電力消費などの需要データ、インターネットトラフィックなどのサービス需要、企業の財務データの推移などを対象としている。

音声データなど、通常のHMMの処理対象とする分野と比較した場合の、これらのデータの特徴は、時系列そのものの認識を目的としていることにある。音声の場合、短時間スペクトルなどをベクトル表現して用いるが、経済時系列の場合には財務指標の時系列などを除いて、単独に時系列を用いる場合が多い。

次にあげられるのが、離散化するレベルがそれほど多くないことであり、特に、定性的あるいは相対的な識別が重視されることにある。音声データの識別の場合には、短時間スペクトルなどの特徴ベクトルの要素の種類は膨大となるが、経済時系列の場合には、要素の数は多くはない。例えば、企業の経営状況を示す財務指標（決算データを基礎として数値の間の比率などを求めたもの）の種類は数百におよぶが、これを主成分分析すると3つ程度に集約できることが知られている。また、マクロ経済指標に関しても、それぞれが密接に推移することが多く時系列として見た場合の系列相関が極めて強いケースが多い。

次の特徴として、ノイズの解釈が困難であることがあげられる。音声を含めて、生成に物理モデルの存在するシステムでは、ノイズの性質や加わり方についてモデルやノイズの識別が可能であることが多いが、経済現象の場合には物理モデルが存在しないため、ノイズの意味が明確ではない。経済時系列予測の場合においては、モデルとして自己回帰をとることが多いことを考慮すると、ノイズ除去の結果として、平滑化された時系列が得られることをモデルとすることが妥当であると思われる。

3.2 2階層のHMM

これまでの議論で、適応フィルタを用いて有色ノイズの影響を抑制することができるので、この後段の処理として、通常のHMMを適用して状態遷移などを推定する。用いるHMMのシステム構成は、通常の方法と同じである。HMMは以下のパラメータを用いて記述される。

$S=[s_0, s_1, \dots, s_M]$: 状態の集合

$O=[o_1, o_2, \dots, o_N]$: 観測された記号列

$A=[a_{ij}]$: 状態 $s_i \rightarrow s_j$ の状態遷移の確率

$B=b_j(o_n)=Pr(o_n|s_n=j)$: 時刻 n における記号 o_n の出力確率、状態は j

$\pi=[\pi_i=Pr(s_1=i)]$: 初期状態の確率

パラメータの集合である $\lambda=(A,B,\pi)$ は、すでにカテゴリへの所属が判明している学習データを用い

てなされる[2][8]。なお、HMMのアルゴリズムについては多数の解説書があるので、詳細は省略する。

時間 n において状態 i に存在する確率を $a(n, i)$ (前向き確率)、時間 n に状態 j を出発して時刻 $N + 1$ に終了状態 M に存在する確率 $\beta(n, j)$ (後向き確率) を逐次近似により計算しておく。

このようにノイズを含む時系列からのノイズ除去を行うとともにこれを入力するHMMを構成する。

次に、この第1階層のHMMによりカテゴリ化されたシンボルを、次の第2階層のHMMへの入力として用いる。

以下では、2段目のHMMにおけるシンボルを3などに限定し、これらのシンボルが生成される確率的なシステムを仮定する。例えば、シンボルBBBを基本的に生成するシステムは、第1段目のHMMの認識誤差を吸収するためのいくつかの派生的なシンボル列、例えば、BBA, BBCも受け付けるシステムにする必要がある。

4 応用例

4.1 シミュレーション・システムと有色ノイズ

以下では、最初に、生成方法が既知である時系列に有色が加わった信号から、適応フィルタを用いてノイズを除去し、この時系列の識別に通常のHMMを用いるシミュレーションを示す。

時系列の生成には、HMMによる時系列認識の基本的なモデルである、複数の状態と複数の出力信号レベルをもつ信号発生システムを仮定する。一般的には、HMMが認識対象とする時系列の生成モデルでは、状態数が多い場合もとりあげられている。しかし、本論文で取り扱うような、有色ノイズで乱された時系列をHMMにより認識する場合には、もとの発生システムの出力レベル数は、それほど多くとることはできない。文献 [5] では、レベルが2の場合に限定して議論している。

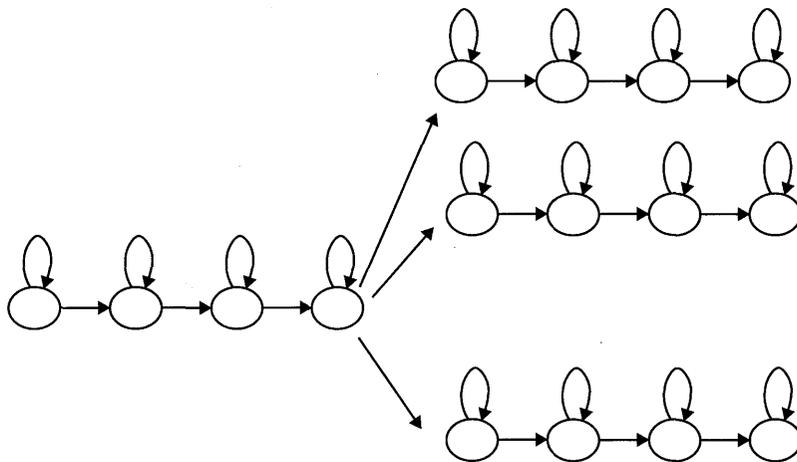


図4 2階層HMMの構成概念

以下では、信号のレベルを5までに拡大した場合に、有色レベルを除去する方法の性質を調べる。信号発生モデルとして、比較的認識システムの構成が容易である、図5に示すような、left to rightの状態遷移をもつHMMの基本モデルにより生成された時系列に、有色ノイズを発生するデジタルフィルタからの信号を加えた時系列を考察する。この時系列を入力として、ここで示す方法でノイズを除去し、もとの信号との2乗誤差を求めて精度を確認する。

従って、時系列的には、1とゼロとが繰り返し出現する系列となり、これに、図6に示すノイズ生成回路からのノイズが重畳されたものを、観測信号としている。

図6には、有色ノイズを発生するデジタルフィルタを示す。ノイズのレベルが大きくなるに従って、パルス状の時系列はその振幅に変動が見られるようになる。図6～8には、もとの信号、ノイズを加算した除去前とノイズ除去後の信号の例を示している。これらの結果について、次に示すようなノイズ除去効果の評価式をもとに計算をしたものが、図9である。

$$SN \text{ 改善比率} = \text{出力の } SN \text{ 比} / \text{入力 } SN \text{ の比} \tag{6}$$

図では、横軸に入力のSN比を、縦軸に出力のSN比を示す。これらの結果より分かるように、本報告で用いたノイズ除去の方法は、新しいノイズの発生や、位相のずれなどの付随する問題を発生しない

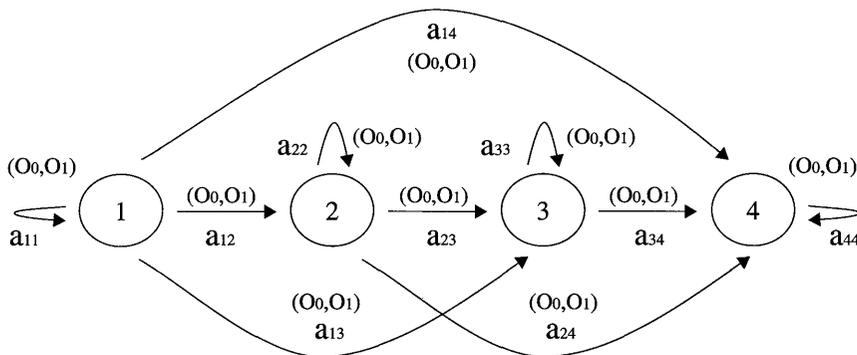


図5 left to rightの遷移状態をもつ4状態HMM

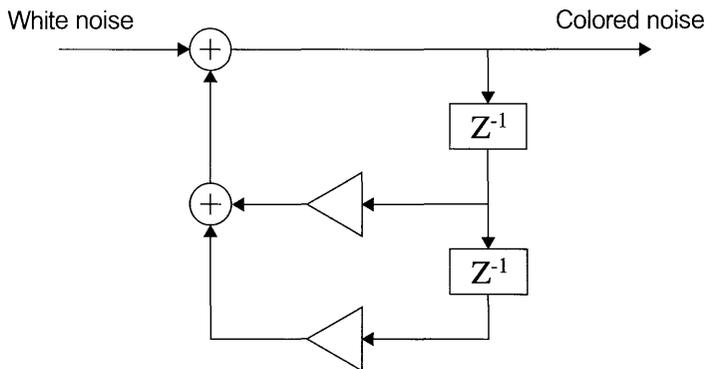


図6 有色ノイズの発生回路

方法であることが確認できる。

4.2 時系列識別の応用例

次に、本論文の手法を用いて、8つの時系列生成のカテゴリを仮定した場合の時系列識別への応用を示す。表2に示すような8つの出力信号の生成パターンを仮定しておく。これらの発生システムから生成された信号に、図5に示す有色ノイズを重畳させ識別すべき信号を作成する。

ただし、以前に議論した2レベルの信号の場合と異なり、経済時系列の場合には明確な離散的なレベルを区別することがそれほど意味をなさないこともあるので、信号レベルの可能な生成パターンをすべて単独のカテゴリとするのではなく、これらの中で類似するパターンを1つのカテゴリとして統

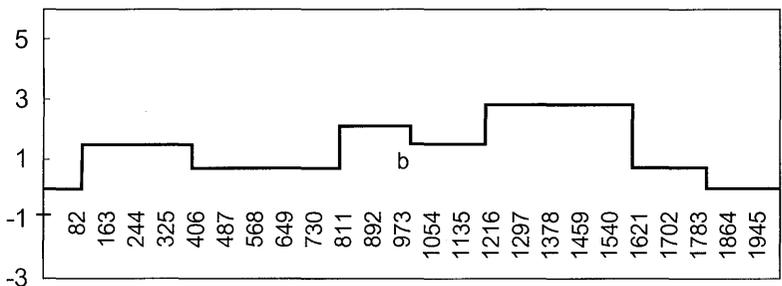


図7 もとの信号

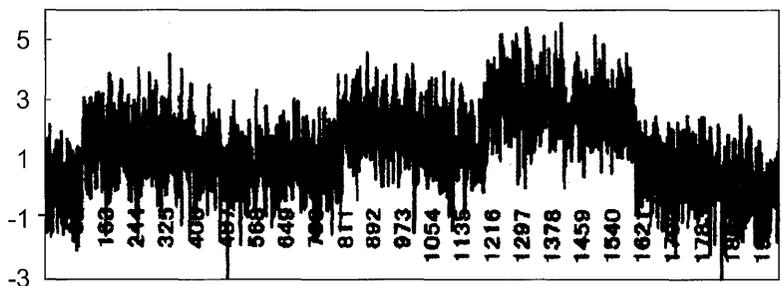


図8 ノイズを重畳した信号

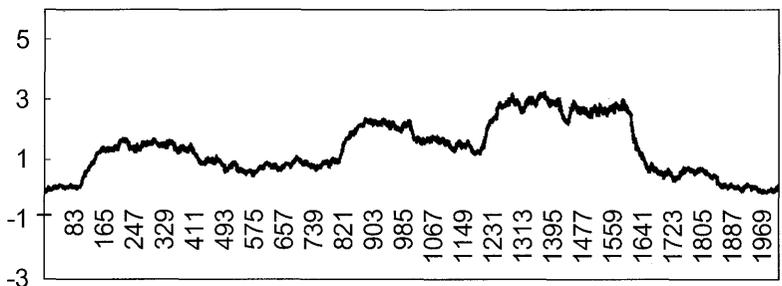


図9 ノイズ除去後の信号

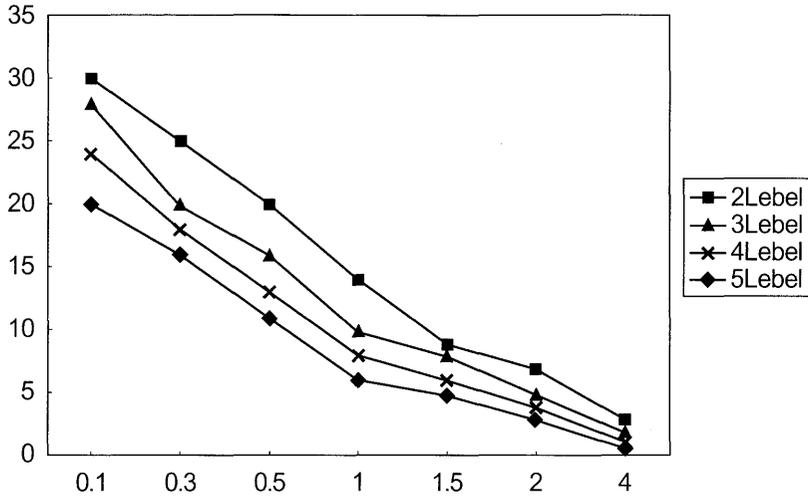


図10 ノイズ抑制の効果

合するケースを考察する。表1にはこれを示す。

同様に第2階層についても可能な生成パターンをすべてとるのではなくいくつかのカテゴリに区分化している。これを表2に示す。

表3には、ノイズレベルと識別の率を示している。なお、信号が000, 111の場合には、そのままではSNRは定義できないので、別の信号レベルとの相対値をとり定義としている。この結果から分かるように、原信号に付加されるノイズのレベルとして、SN比が15以上であるならば認識は確実に実施されている。

表1 8つの時系列カテゴリと信号列の例

no	pattern	no	pattern
A	000 111 222 333 444	B	001 112 223 334
C	010 121 232 343	D	011 122 233 344
E	100 211 322 433	F	101 212 323 434
G	110 221 332 443		

表2 8つの信号系列の時系列カテゴリと信号列の例

no	pattern	no	pattern
a	AAA BBB CCC..	b	AAB BBC CCD..
c	ABA BCB CDC..	d	ABB BCC CDD..
e	BAA CBB DCC..	f	BAB CBC DCD..
g	BBA CCB DDC..		

表3 4つの信号レベル系列のカテゴリ識別率 (%)

no	SNR=5	SNR=10	SNR=15	SNR=20
A	24	65	100	100
B	10	55	100	100
C	15	42	100	100
D	19	50	100	100
E	20	65	100	100
F	15	55	100	100
G	10	32	100	100

4.3 第2階層における認識

次に、第1階層で識別された時系列をシンボルとして表し、これを入力の記号列とした場合に、第2階層のHMMにおいて行われるカテゴリ識別の結果について述べる。

表4 8つの時系列カテゴリの識別率 (%)

no	SNR=5	SNR=10	SNR=15	SNR=20
a	24	65	100	100
b	10	25	70	79
c	15	22	88	80
d	19	30	80	78
e	20	35	70	80
f	18	45	80	70
g	19	32	75	79

5 むすび

本報告では、以前に著者らが示した、信号に有色ノイズが重畳しているケースにおいて、ノイズ除去の方法を工夫することにより、基本的なHMMの手法を適用する方法を基礎として階層型HMMを構成することを提案した。ノイズを含んだ信号を予測フィルタに通してイノベーションを生成し、このイノベーションを用いて、信号を仮想的なノイズと考慮して、もとの信号に含まれるノイズを特定する。こののち、ノイズを入力信号から差し引く処理をする。これにより、HMMにおける状態を拡張するためのmetastateの導入などは必要ない。応用例として、複数の信号レベルをもつ経済時系列の認識の基本的モデルを考察した。

今後、実際の時系列認識への拡張を検討する予定である。

参考文献

- [1] S. E. Levinson, L. R. Rabiner and M. M. Sondhi, "An introduction to the application of probabilistic functions of Markov processes to automatic speech recognition", Bell Systet. Tech. J, vol.62, pp.1035-1074, 1983.

- [2] L. Rabiner, “A tutorial on hidden markov models and selected applications in sepeech processing”, Proc. IEEE, vol.77, pp.257-285, 1989.
- [3] L. Venkataramanan, J. L. Walsh, R. Kuc and F. J. Sigworth, “Identification of hidden markov models for ion channel currents-Part I: Colored background noise”, IEEE Trans., Signal Processing, vol.46, no.3, pp. 1901-1915, 1998.
- [4] D. R. Fredkin and J. A. Rice: “Fast evaluation of the likelihood of an HMM: Ion channel currents with filtering and colored noise”, IEEE Trans., Signal Processing, vol.49, no.3, pp.625-623, 2002.
- [5] 高木昇, 時永祥三, 適応的フィルタリングによるノイズ除去法と隠れマルコフモデルを用いた時系列認識, 信学技報DSP2002-100, pp.13-18, 2002.
- [6] 川村新, 藤井健作, 伊藤良生, 副井裕, “線形予測分析に基づく騒音抑圧法”, 電子情報通信学会論文誌, vol.J85-A, no.4, pp.415-423, 2002.
- [7] L. E. Baum, Y. Petirc, G. Soules and N. Weiss, “A maximization technique occuring in the statistical analysis of probabilistic functions of markov chains”, Ann. Math. Stat, vol.40, pp.164-171, 1970.
- [8] D. R. Fredkin and J. A. Rice, “Fast evaluation of the likelihood of an HMM: Channel currents with filtering and colored noise”, IEEE Transactions on Signal Porcessing, vol.49, no.3, pp.625-633, March 2002.