

時系列モデルと学習 —金融時系列を例として—

柴田 里程

慶應義塾大学理工学部
shibata@math.keio.ac.jp

上辻 茂男

慶應義塾大学理工学部
shigeo@math.keio.ac.jp

■学習アルゴリズムと時系列データ■

本稿では、時系列解析、特に近年注目を浴びている金融時系列の解析に学習アルゴリズムがどのような役割を果たすか、伝統的な時系列モデルと比較しながらその展望を述べることにする。

学習アルゴリズムと呼ばれるものの中には、ボルツマンマシン (Boltzman Machine)、ニューラルネットワーク (Neural Network)、自己組織化写像 (Self Organizing Map) などさまざまなものがあるが、その背後にはいずれも、いくつかのユニットを辺で結んだグラフ構造をしたモデルを想定している。辺に重みを設定し、その重みを変化させることによって学習を実現するアルゴリズムがつまり学習アルゴリズムであるが、特に学習とはいっていないものの、回帰樹 (回帰木, Regression Tree)、分類樹 (分類木, Classification Tree) やグラフィカルモデルの当てはめなどもデータから1つのグラフィカルなモデルを構築し、それを元にさまざまな判断を行おうとしている点では何ら違いはない。

これらのさまざまな方法が生まれる原因としてはいくつか考えられる。まず、よくいわれる区別が「教師あり」の学習か「教師なし」の学習かである。ニューラルネットワーク、回帰樹、分類樹は「教師あり」、ボルツマンマシン、自己組織化写像やグラフィカルモデルは「教師なし」である。さらに、目的が「予測」なのか「分類」なのか「モデルを得ることそのもの」なのかによっても違ってくる。ニューラルネットワークや回帰樹はもっぱら予測を目的とし、分類樹、自己組織化写像は分類をもっぱら目的としている。もちろん、予測を目的とする場合は教師ありのことが多く、分類が目的の場合は教師なしの場合が多い。一方、ボルツ

マンマシンやグラフィカルモデルはモデルそのものを得ることを目的としている。さらに、各ユニットが離散値をとるのか連続値をとるかによる違いも大きく、ランダムなメカニズムを内在しているかどうか大きな違いを生む。

これらの背景を前提に、時系列からの学習、あるいはグラフ構造を持ったモデルの時系列データへの当てはめを考えてみると、「時間」という一定方向の流れがある時系列データの場合には、ニューラルネットワークのような複数の入力と出力を持ったモデルを考え、遅れを持ったいくつかの値を入力とし、予測値を出力とするのが1つの自然な選択である。実際、このようなニューラルネットワークを金融時系列に適用してうまくいった、あるいはうまくいくという報告は数多く、たとえばFernández-Rodríguez¹⁾らは、マドリッド株式市場の1996年1月2日から1997年10月12日までの6931個の指数データから、投資戦略決定システムをニューラルネットワークで構築し売買したら、市場が弱気相場や平衡状態の場合には、途中で一切売買を行わずに保持したときより多くの利益が得られたと報告している。

一方、一連の時系列をそれぞれ1つのパターンと見なすならば視覚情報の特徴抽出処理における自己組織化の考え方を一般化した自己組織化写像 (Kohonen²⁾) によって分類を行うことも考えられる。実際、Cottrell³⁾らは翌日の平均電気消費量を予測するシステムを構築するために、ポーランドの1時間単位の電力消費量 (1986年1月1日から1994年12月31日までの3287日) データを元に、1日の電力消費量時系列を100種に分類した。ポーランドの祝日などや気象状況に関する情報は用いなかったものの、日曜日、月曜日、夏の平日、冬

の平日などその特徴別にうまく分類され、伝統的な時系列モデルに基づく予測より良い予測が得られると報告している。またSerrano-Cinca⁴⁾は会社の健全性を判定するシステムを構築するためにまず自己組織化写像でデータを分類し有意な変量を見出すことが有用であったと報告している。

このような報告のいずれも魅力的ではあるが、当然、伝統的な時系列モデルよりもなぜ良い結果が得られるのか、入念なデータ解析を行いモデルを構築すればさらに良い結果が得られるのではないかというような疑問がわく。一般にはモデルがグラフィカルな構造を持っていることが非線型な時系列モデルも含め従来の時系列モデルより有利であると認識されているがどうだろうか。当然、データの背後にある現象がこのようなグラフィカルな構造で動いていなければ、かえって不安定な動きを示すことになる。その意味では伝統的な時系列モデルよりも一段と慎重な扱いが必要なモデルであり、安易な適用は危険である。本稿の後半では、金利時系列を例にこのことを示す。

また、通常用いられる学習モデルは確率的な動きを表現するランダムな要素を内在していないため、時系列の非ランダムな部分をローカルに閾値近似しているに過ぎないが、果たしてそれで十分なのかという疑問もわく。この意味では、各ニューロンが入力によって確率的に活性化する確率的ニューラルネットワークの方が適していると思われるが、残念ながらまだあまり研究が進んでいない。最近、著者らは確率的ニューラルネットワークの尤度を最大となるように学習させる新たなバックプロパゲーション法を開発し、その収束性も証明できたが、詳細はここでは省略せざるを得ない。

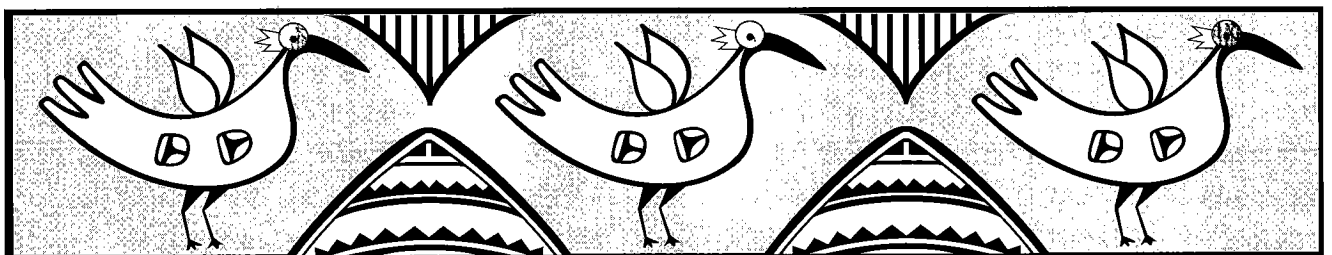
また、一部で用いられている遺伝的アルゴリズム (Genetic Algorithm, Varetto⁵⁾) も一種の学習アルゴリズムとして捉えられていることも多いが、データに対するモデルがこのアルゴリズムの背後にあるわけではなく、別途に想定したモデルを大量のデータに対して当てはめるときに利用できる計算アルゴリズムに過ぎない。

い。したがって、上記のようなグラフィカルな構造を持ったモデルに基づく学習アルゴリズムとは区別すべきであろう。

■金融時系列■

金融時系列は、為替、株価、債券価格、先物商品価格、オプション価格など金融に関係した時系列であるが、他の経済時系列と異なり、きわめて流動性の高い市場で定まる金融商品の価格である。したがって経済現象でありながら「見えざる手」によってごく短時間のうちに均衡が図られるといった側面が強く、熱力学的なモデルを応用することも可能である。実際、有名なBlack and Scholesによるオプション価格モデルは熱拡散方程式を念頭においた確率微分方程式を前提としている。しかし、経済現象である以上、人間の介入と市場を取り巻く社会の動きから無関係ではあり得ない。実際、金融時系列を解析してみれば、誤差を十分考慮したとしても、決して理論通りには動いていないことがすぐ分かる。そのため、金融時系列解析で一種の定番となっているGARCH (Generalized Autoregressive Conditional Heteroskedasticity) モデルでは、分散の変動を線形モデルでモデル化することによってブラウン運動あるいはランダムウォークを拡張し理論との乖離を吸収しようとしている。しかし、このモデルでは市場参加者の心理状態や比較的長期にわたる金融環境の変化まで追いかけるのはなかなか困難である。そこに学習アルゴリズムの登場する余地がある。

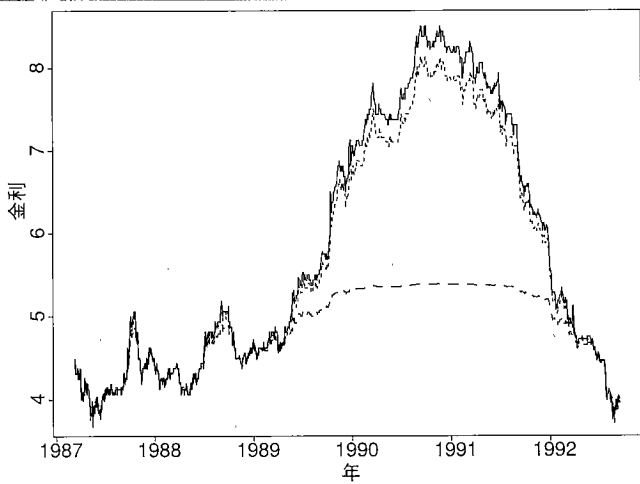
以下で説明の例として取り上げる時系列は3カ月満期の日本円金利 (1986年12月1日から1992年9月16日までの2115日) 時系列である。この時系列はShibata and Miura⁶⁾で対象とした時系列の一部であり、ちょうど1990年頃から始まったバブル経済の時期を含んでいるので、時系列解析の対象としても興味深い。彼らはノンランダムな長期変動と短期変動、それにランダムな変動の3つの時系列に分解することによって、公定歩合の変化のような緩やかな変動、それよりは急速な変化、さらにリスクを短期にカバーしようとする市場参加者



	H=1	2	3	4
I=1	0.00284	0.00281	0.00284	0.00284
2	0.00284	0.00284	0.00270	0.00270
3	0.00284	0.00284	0.00250	0.00235
4	0.00280	0.00222	0.00227	0.00226

表-1 学習後の予測値の平均2乗誤差

	H=1	2	3	4
I=1	-294.56	-289.76	-282.62	-276.60
2	-292.78	-284.58	-281.72	-273.79
3	-290.74	-280.76	-283.18	-279.40
4	-290.12	-288.61	-287.07	-275.62



の心理などを表すランダムな変動を抽出した。これはある意味では学習アルゴリズムの対極に位置する伝統的なアプローチである。

なお、以下では最初の100日間の時系列を学習あるいはパラメータ推定に用い、残りを検証に利用している。

■自己回帰モデルとニューラルネットワークモデル■

ここでは、話を分かりやすくするため、伝統的なモデルとしては自己回帰モデルAR(p)

$$Z_t = \sum_{j=1}^p a_j Z_{t-j} + \epsilon_t \quad (1)$$

を取り上げ、学習モデルとしては隠れ層が1層のニューラルネットワークを取り上げ比較する。

日本円金利 x_t から平均を除いた時系列 Z_t に、最初の100日間、つまり学習期間の実現値をもとに自己回帰モデルを当てはめると、 $p=1, 2, 3$ の中でAIC (Akaike's

Information Criterion) を最小とする p は1であり、

$$Z_t = 0.9093Z_{t-1} + \epsilon_t \quad (2)$$

というモデルが得られる。このとき ϵ_t の分散は0.00272である。

一方、 $x_t, x_{t-1}, \dots, x_{t-I+1}$ を入力とし x_{t+1} を出力と対照する値としてニューラルネットワーク

$$y_{t+j} = h_0 + \sum_{k=1}^H h_k f\left(\sum_{j=1}^I w_{jk} x_{t-I+j} + w_k\right) \quad (3)$$

に学習させた結果は表-1のようになる。ここで I は入力ユニットの数、 H は隠れユニットの数であり活性化関数 f としてはシグモイド関数を用いている。

もちろん、見かけ上、モデルが複雑になればなるほど誤差は減少するが、これはいわゆる与えられたデータへの当てはめ過ぎである可能性も高いので、自己回帰モデルのときの p の選択と同じように平均2乗誤差以外の基準で、当てはめ過ぎでもなく当てはまり不足でもないちょうどよいモデルを選び出す必要がある。

この場合、式(3)の y_{t+1} と x_{t+1} の差を ϵ_{t+1} とし、これが正規分布すると仮定すればAICは自己回帰モデルのときと同様に計算でき、

$$\begin{aligned} \text{AIC} &= (n-I) \log 2\pi\hat{\sigma}^2 + (n-I) \\ &+ 2\{(I+2) \times H + 1\}, \\ \hat{\sigma}^2 &= \frac{1}{n-I} \sum_{t=I+1}^n (x_t - y_t)^2 \end{aligned} \quad (4)$$

となる。ここで注意すべきことはモデルの初期値への依存度を統一したうえで尤度を計算する必要がある点である。これは、自己回帰モデルの場合も同様で選択の対象となる p の最大値 P をあらかじめ定め、 Z_1, Z_2, \dots, Z_P を条件とする条件付き密度に基づいてAICを計算している。これは、時系列という過去への依存性のある場合特有の問題で、このように条件とする値の個数を統一してからでないと同様な比較とならず、大きなモデルを選択してしまう危険にもつながる。このニューラルネットワーク場合には I の最大を $\bar{I}=4$ として統一している。次の表-2がAICの値を表にしたものであるが、実際 \bar{I} の代わりに I を用いると $I=4, H=2$ の場合の値と $I=4, H=3$ の場合の値が逆転する。もちろん、この場合の最小値は $I=1, H=1$ のときであり結果は左右されないが、左右されることも多いので注意が必要である。

得られた自己回帰モデルとニューラルネットワークモデルを用いて1987年3月11日以降の1日先の金利の予測を行った結果が図-1である。実線が実際の金利(年利%)、点線がAR(1)モデルによる予測、破線がニュー

ラルネットワークによる予測値である。

自己回帰モデルは前日の値だけから予測することになり、観測値の動きによく追従している。しかし、平均を除いた部分に関するモデル化であるので学習期間よりかなり高いレベルで金利が推移する個所では追い付ききれない部分も見受けられる。

一方、ニューラルネットワークは、この図ではあまりはつきりしないが、学習期間に近いところではある程度よく追従するものの、そこから離れると変化に追いついていない。

次の表-3は学習期間での平均2乗誤差とそれ以降の期間での平均2乗誤差をまとめたものである。平行移動とあるのは単純に x_t を x_{t+1} の予測値としたときである。

この結果から分かるように、学習期間で最も誤差の少ないのは、自己回帰、次にニューラルネットワーク、そして平行移動である。ニューラルネットワークはきわめて単純な平行移動よりは少しは良い予測を与えるものの、自己回帰ほど良くはならないことが分かる。

学習期間からはずれたところで平行移動が安定して予測できるのは当然といえば当然であるが、それでも自己回帰はニューラルネットワークよりはずっとましである。このように学習期間から離れるとニューラルネットワークが急速に追従性を失うのは皮肉にもその非線形性にある。この場合学習したモデルは

$$y_{t+1} = 5.407 - 7.487f(-1.022x_t + 2.632)$$

であるが、この式から y_{t+1} は5.407より大きな値はとり得ないことが分かる。そのため実際の金利が5.407を超えて増加すればするほど誤差は増え続けることになる。

実は学習期間でも、 I, H の値次第ではニューラルネットワークは不自然な予測値を与えることがある。その例が $I=3, H=3$ のときの予測値を示した図-2である。実線が実際の金利、点線がニューラルネットワークによる予測である。9番目の予測値と81番目の予測値がかなり低い値となっている。これは y_{t+1} の入力 x_t, x_{t-1}, x_{t-2} に関する勾配をまとめた表-4をみれば明らかである。つまり、入力特定の値の組をとると勾配が急に変化する。これが原因である。式(3)は係数が過去の値に依存する自己回帰モデル

$$y_{t+1} = \sum_{j=1}^I \alpha_j (x_{t-1}, \dots, x_{t-I}) x_{t-j} + \epsilon_t$$

で近似することができるので、表-4の勾配はちょうどこの表現での係数 α_j に当たる。

また、 $I=4, H=3$ の場合には図-3で点線で示された予測値から分かるようにもっと極端なことが起きる。

	学習期間	学習期間以降
ニューラル	0.00284	2.02
自己回帰	0.00272	0.0360
平行移動	0.00288	0.00160

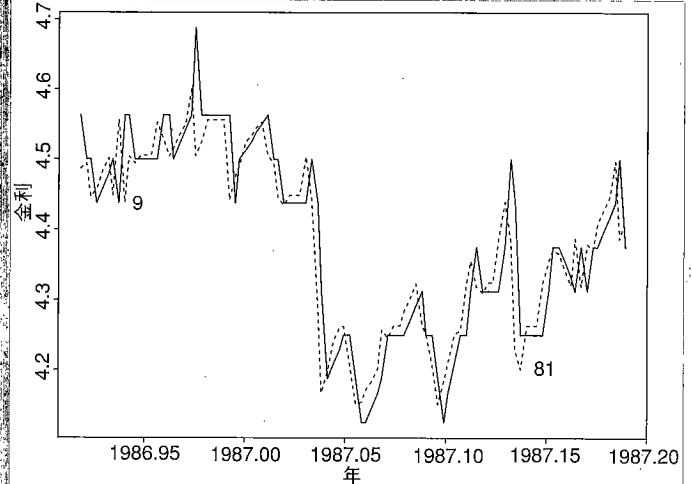


図2 $I=H=3$ の場合の予測値

番号	x_{t-2}	x_{t-1}	x_t
8	-0.0703	0.0463	0.801
9	2.25	-1.54	-0.104
10	-0.118	0.0789	0.922
		⋮	
80	-0.416	0.283	0.953
81	-0.438	0.299	0.876
82	-0.0159	0.00868	0.956

表4 9番目と81番目付近の予測変量の勾配

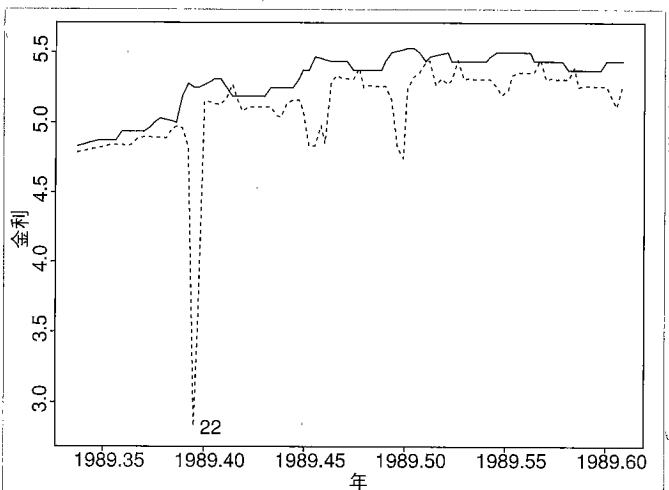
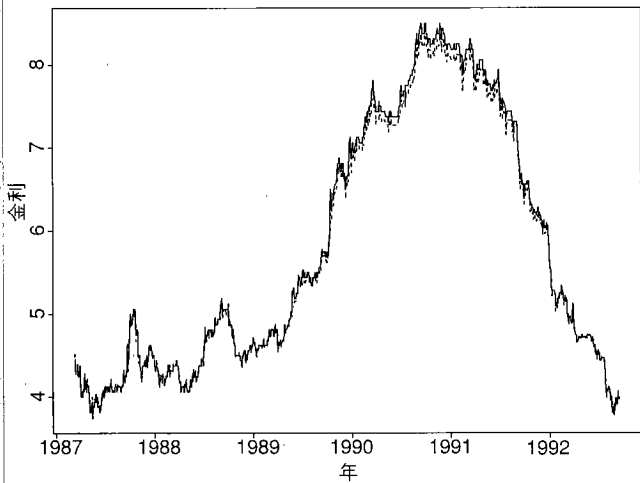


図3 $I=4, H=3$ の場合の予測値

番号	x_{t-3}	x_{t-2}	x_{t-1}	x_t
20	0.232	0.273	-0.398	-0.157
21	1.61	1.90	-2.55	-1.26
22	6.49	7.67	-10.2	-5.12
23	11.63	13.7	-18.2	-9.20
24	0.927	1.093	-1.491	-0.70



	学習期間	学習期間以降
ARMA(1,1)	0.00262	0.0430
ARMA(2,2)	0.00289	0.00655
ARMA(3,3)	0.00443	0.108

22番目の予測値が大幅に落ちているが、これも表-5の勾配をみるとこのあたりで勾配が急速に変化しているのが原因である。

ここで、ARモデルの代わりにより一般的なARMA (Auto-Regressive Moving-Average) モデルも試してみよう。

ARMA(p, q)モデル

$$Z_t = \sum_{i=1}^p \alpha_i Z_{t-i} + \sum_{j=0}^q \beta_j \epsilon_{t-j}$$

のうちARMA(1,1), ARMA(2,2), ARMA(3,3)をAICで比較するとARMA(2, 2)が最小値を与える。このモデルのもとでの予測値は図-4から分かるように学習期間以降でもほとんど見分けがつかないほど良好である。またこれら3つのモデルについて予測誤差は表-6のようになる。

ARMA(2, 2)の予測誤差は学習期間ではAR(1)モデルに比べて少し増加するものの、学習期間以降でもほと

んど同一のレベルに収まっている。ARMAモデルの利点は誤差に系列相関を導入できる点にあり、その効果が遺憾なく発揮されている。このような相関のある誤差を学習モデルで扱うとしたら、ランダム性の内在する確率的ニューラルネットワークのような学習モデルを考える必要があり、上記の結果はその必要性を示している。

■まとめ■

ニューラルネットワークに代表される学習モデルはグラフ構造を持ったきわめて非線型な現象もモデル化できる可能性を秘めているという点では魅力的であるが、時系列に対しては非線型性がその時系列を生成する現象とかなりの整合性を持たないかぎりその有効性を発揮できないばかりか、不自然な結果を得ることになる。

やはり、十分なデータ解析を行ったうえでの学習モデルの構築が必要であり、安易な利用は危険である。しかし、予備的な解析の道具としての活用には有効であると思われる。

参考文献

- 1) Fernández-Rodríguez, F., González-Martel, C. and Sosvilla-Rivero, S.: On the Profitability of Technical Trading Rules Based on Artificial Neural Network: Evidence from Madrid Stock Market, *Economic Letters*, Vol.69, pp.89-94 (2000).
- 2) Kohonen, T.: Self-organized Formation of Topologically Correct Feature Maps, *Biological Cybernetics*, Vol.43, pp.59-69 (1982).
- 3) Cottrell, M., Girard, B. and Rousset, P.: Long Term Forecasting by Combining Kohonen Algorithm and Standard Prediction, *SAMOS* (1997).
- 4) Serrano-Cinca, C.: Let Financial Data Speak for Themselves, In *Visual Explorations in Finance*, Eds. Deboeck, G. and Kohonen, T., Springer, London, pp.3-23 (1998).
- 5) Varetto, F.: Genetic Algorithms Applications in the Analysis of Insolvency Risk, *J. Bankign and Finance*, Vol.22, pp.1421-1439 (1998).
- 6) Shibata, R. and Miura, R.: Decomposition of Japanese Yen Interest Rate Data Through Local Regression, *Financial Engineering and the Japanese Markets*, Vol.4, pp.125-146 (1997).

(平成12年11月29日受付)

