

ベイジアンネットワークを用いた 株価予測法の精度改善

左 毅^{†1} 原田 昌 朗^{†1} 北 栄 輔^{†1}

これまでに、著者らのグループではベイジアンネットワークを用いた株価リターンの予測法を提案し、日経平均株価とトヨタ自動車株価を例として時系列分析法と比較した。その結果、提案手法は時系列分析法よりも高い精度を示した。本研究では、従来法の精度を向上させる手法について述べる。従来法では、株価リターンをノードとしてベイジアンネットワークを構築した。これに対して、改善手法では、株価リターンノードに加えて株価リターンの予測精度ノードを用いてベイジアンネットワークを構築し、株価リターンを予測する。予測精度において比較した結果、提案手法は、時系列分析法に比べて30%、ベイジアンネットワークを用いた従来法に比べて6%から12%程度精度を向上させることができることが分かった。

Improvement of Stock Price Forecast Using Bayesian Network

YI ZUO,^{†1} MASA AKI HARADA^{†1} and EISUKE KITA^{†1}

Authors have presented the forecast scheme of the stock return by using Bayesian network. The forecast scheme was compared with the time-series forecast schemes, which are popular traditional ones, in the prediction of NIKKEI stock average return and Toyota motor corporation stock price return. The present scheme showed more accurate forecast results than the time-series forecast schemes. In this study, we will present the improved scheme of the forecast scheme of the stock return using Bayesian network. In the previous scheme, Bayesian network for stock return forecast was constructed from the return nodes alone. The improved scheme adopts as the nodes, in addition to the return node, the return forecast error node. Numerical results show that the forecast accuracy of the present scheme is improved by 30% against the time-series forecast schemes and by 6% to 12% against the previous scheme using Bayesian network.

1. 緒 論

株価を予測するために様々な方法が提案されている。株価予測の代表的な手法は時系列分析である¹⁾。Auto Regressive (AR) モデルは自己回帰モデルであり、今期株価を過去の株価の線形結合で近似する。Moving Average (MA) モデルは移動平均モデルであり、今期株価を過去の攪乱項の移動加重和で近似する。AutoRegressive Moving Average (ARMA) モデルは AR と MA を組み合わせたモデルである。さらに、AutoRegressive Conditional Heteroskedasticity (ARCH) モデルは、ボラティリティをリスクの指標として用いて、その変化をとらえられる時系列モデルである。ARCH モデルは 1980 年代初期に提案されて以来、株や金利などによる価格変化の分析・予測に大きな威力を発揮し、その後多数の改良型が提案されている。

これらのモデルでは、予測しようとする株価リターンを説明変数の線形結合で近似し、重み係数を相関解析などにより決定する。このとき、株価リターンの頻度分布は正規乱数に従うと仮定されている。しかし、近年のいくつかの研究により、株価変化は正規乱数に完全には従わないことが指摘されているので、正規乱数を用いたモデルでは、十分な予測精度を得られない可能性がある。そこで、著者らは、以前の研究においてベイジアンネットワーク²⁾を用いた株価リターン推定法を提案した³⁾。この方法では、株価リターンの分布をクラスタリング手法によって離散値に変換する。つづいて、過去の株価リターンが特定の値をとるときに、次の時点の株価リターンが特定の値をとる確率を推測するためのベイジアンネットワークを構築して、将来の株価リターン予測に用いる。この方法を、日経平均株価とトヨタ自動車の株価リターン予測に適用し、時系列分析法よりも精度良く予測できることを示した。

本研究では、以前に提案した手法の予測精度改善方法について検討する。提案手法では、従来の手法で過去の株価リターンを予測した後、その予測値の誤差を評価する。誤差分布をクラスタリング法によって離散化して誤差分布の離散値を求め、誤差ノードを定義する。誤差ノードを株価リターンノードとともに用いてベイジアンネットワークを構築し、これを株価リターン予測に用いる。また、従来の方法では、求められたネットワークは Naive ベイズ構造であった。そこで、本研究では、計算コストを下げるために Naive ベイズ構造を仮

^{†1} 名古屋大学大学院情報科学研究科
Graduate School of Information Science, Nagoya University

定してネットワーク構造を探索する．解析例として日経平均株価とトヨタ自動車株価⁴⁾をとり，提案手法による予測結果を時系列分析法や従来³⁾の方法と比較する．

本論文の構成は以下のようになっている．2章では時系列分析法について簡単に紹介した後，ベイジアンネットワークを用いる予測法である従来法と本研究の方法を比較する．3章では Naive ベイズ構造に基づくネットワークの探索アルゴリズムについて説明する．4章と5章では，それぞれ以前に提案した手法と本研究で提案する手法について説明する．6章では解析結果を示し，7章はまとめである．

2. 研究の背景

2.1 時系列分析法¹⁾

2.1.1 AR モデル

AR モデル AR(p) では， t 期の株価リターン r_t を p 期前までのリターン r_{t-i} と，擾乱項 u_t によって次のように表す．

$$r_t = \alpha_0 + \sum_{i=1}^p \alpha_i r_{t-i} + u_t \quad (1)$$

ここで， α_i はモデルのパラメータである． u_t は平均 0，分散 σ^2 のホワイトノイズに従う．

本研究では，文献 1) に従って，予測誤差を最小とするように AR モデルのパラメータ次数 p を決定する．文献 1) では，予測誤差が独立同一の正規分布に従うと仮定すれば，予測誤差の平均 2 乗誤差は予測誤差の標準偏差 σ と同値なので，予測誤差を最小にする基準として式 (2) で定義される AIC を用い，これを最小とするようにモデルパラメータを決定している．

$$\text{AIC} = \ln \sigma^2 + \frac{2p}{T} \quad (2)$$

ここで， σ^2 は AR モデルの残差標本の分散であり， T は株価リターン標本の総数である． σ^2 は次式で求められる．

$$\sigma^2 = \frac{1}{T-p} \sum_{t=p+1}^T (r_t - \bar{r}_t)^2 \quad (3)$$

ここで r_t は式 (1) より求められる t 期のリターンの期待値を， \bar{r}_t は実測値を示す．

解析例では $p = 1, 2, \dots, 10$ について解析を行い，式 (2) が最小値となる p を採用する．

2.1.2 MA モデル

MA モデル MA(q) では， t 期の株価リターン r_t を過去の攪乱項 u_{t-j} によって以下のよう表す．

$$r_t = \beta_0 + \sum_{j=1}^q \beta_j u_{t-j} + u_t \quad (4)$$

ここで， β_j はモデルのパラメータである．

本研究では，文献 1) に従って，予測誤差を最小とするように MA モデルのパラメータ次数 q を決定する．文献 1) では，予測誤差が独立同一の正規分布に従うと仮定すれば，予測誤差の平均 2 乗誤差は予測誤差の標準偏差 σ と同値なので，予測誤差を最小にする基準として式 (5) で定義される AIC を用い，これを最小とするようにモデルパラメータを決定している．

$$\text{AIC} = \ln \sigma^2 + \frac{2q}{T} \quad (5)$$

ここで， σ^2 は次式で求められる．

$$\sigma^2 = \frac{1}{T-q} \sum_{t=q+1}^T (r_t - \bar{r}_t)^2 \quad (6)$$

解析例では，MA モデルのパラメータ次数 q が $q = 1, 2, \dots, 10$ について解析を行い，式 (5) が最小値をとるときの q を採用する．

2.1.3 ARMA モデル

ARMA モデル ARMA(p, q) では， t 期の株価リターン r_t を株価リターン r_{t-i} と攪乱項 u_{t-j} によって次のように表す．

$$r_t = \sum_{i=1}^p \alpha_i r_{t-i} + \sum_{j=1}^q \beta_j u_{t-j} + u_t \quad (7)$$

本研究では，文献 1) に従って，予測誤差を最小とするように ARMA モデルのパラメータ次数 p, q を決定する．文献 1) では，予測誤差が独立同一の正規分布に従うと仮定すれば，予測誤差の平均 2 乗誤差は予測誤差の標準偏差 σ と同値なので，予測誤差を最小にする基準として式 (8) で定義される AIC を用い，これを最小とするようにモデルパラメータを決定している．

$$\text{AIC} = \ln \bar{\sigma}^2 + \frac{2(p+q)}{T} \quad (8)$$

ここで、 $\bar{\sigma}^2$ は次式で求められる。

$$\bar{\sigma}^2 = \frac{1}{T-(p+q)} \sum_{t=(p+q)+1}^T (r_t - \bar{r}_t)^2 \quad (9)$$

解析例では、ARMA モデルのパラメータ次数 p, q が $p = 1, 2, \dots, 10$ と $q = 1, 2, \dots, 10$ について解析を行い、式 (8) が最小値をとるときの p, q を採用する。

2.1.4 ARCH モデル

ARCH モデル ARCH(p, q) では、 t 期の株価リターン r_t を次式で表す⁵⁾。

$$r_t = \alpha_0 + \sum_{i=1}^p \alpha_i r_{t-i} + u_t \quad (10)$$

また、擾乱項 u_t は次式で与えられる。

$$u_t = \sigma_t z_t \quad (11)$$

ここで、 $\sigma_t > 0$ であり、 z_t は平均 0、分散 1 の正規乱数である。

Engle の ARCH モデル⁵⁾ によれば、 σ_t^2 は次式で近似される。

$$\sigma_t^2 = \beta_0 + \sum_{j=1}^q \beta_j u_{t-j}^2 \quad (12)$$

本研究では、文献 1) に従って、予測誤差を最小とるように ARCH モデルのパラメータ次数 p, q を決定する。予測誤差の平均 2 乗誤差は予測誤差の標準偏差 $\bar{\sigma}$ と同値なので、予測誤差を最小にする基準として次式で定義される AIC を用いる。

$$\text{AIC} = \ln \bar{\sigma}^2 + \frac{2(p+q)}{T} \quad (13)$$

ここで、 $\bar{\sigma}^2$ は ARCH モデルの残差標本 $\sigma_{p+q+1}^2 - \bar{u}_{p+q+1}^2, \sigma_{p+q+2}^2 - \bar{u}_{p+q+2}^2, \dots$ の分散であり、 T は株価リターン標本の総数である。 $\bar{\sigma}^2$ は次式で求められる。

$$\bar{\sigma}^2 = \frac{1}{T-(p+q)} \sum_{t=(p+q)+1}^T (\sigma_t^2 - \bar{u}_t^2) \quad (14)$$

ここで、 σ_t^2 は t 期の株価リターン r_t の条件付き分散であり、 \bar{u}_t^2 は t 期株価リターンの予測

誤差の 2 乗である。

解析例では、ARCH モデルのパラメータ次数 p, q が $p = 1, 2, \dots, 10$ と $q = 1, 2, \dots, 10$ について解析を行い、式 (13) が最小値をとるときの p, q を採用する。

2.2 ベイジアンネットワークを用いた予測手法

時系列分析法では、株価リターンや誤差の分布が正規分布に従うと仮定し、今期の株価リターンを過去の株価リターンと擾乱項の線形結合で近似している。しかし、株価の頻度分布は完全には正規分布に従わないことが指摘されている⁶⁾。

ベイジアンネットワークでは、確率変数間の定性的な依存関係を確率変数をノード、依存関係をリンクとする非循環有向グラフで可視化し、変数間の定量的な依存関係を条件付き確率によってモデル化する。先行研究では、ベイジアンネットワークを用いて、過去の株価リターンどうしの因果関係をモデル化し、これを予測に用いている。このとき分布には正規乱数などによる仮定を用いないので、時系列分析法の問題点を改善できる可能性がある。

本研究で提案する手法では、従来手法の予測精度向上を目的として、株価リターンだけでなくその予測誤差も用いてベイジアンネットワークを構築する。従来手法のベイジアンネットワークを用いて、すでに測定されている過去の株価リターンを予測すれば、その予測値と実際値から誤差を評価できる。提案手法では、この誤差値のノードを定義し、株価リターンのノードとともに用いて、株価リターン予測用のベイジアンネットワークを構築する。

提案手法の着想は、時系列分析法における AR モデルと ARCH モデルの関係にある。図 1 に示すように、AR モデルでは予測株価リターン r_t を過去リターン r_{t-i} と擾乱項 u_t で近似する。ARCH モデルでは AR モデルにおける擾乱項 u_t のボラティリティを過去の擾乱項 u_{t-i} で近似する。つまり、ARCH モデルでは予測株価リターン r_t を過去リターン r_{t-i} と過去の擾乱項 u_{t-i} で近似している。一方、ベイジアンネットワークを用いた従来手法 (BN(Previous) と表記) では、予測株価リターン r_t を過去の株価リターン r_{t-i} のみの非線形関係式として定義するためにベイジアンネットワークを用いるのに対して、本研究の提案手法 (BN(Present) と表記) では予測株価リターン r_t を過去の株価リターン r_{t-i} と過去の予測誤差 u_{t-i} の両者の非線形関係式として定義するためにベイジアンネットワークを用いている。

しかし、ARCH と BN(Present) の予測精度向上に対するアプローチは少し異なる。ARCH モデルでは、AR モデルにおける擾乱項 u_t のボラティリティを過去の擾乱項 u_{t-i} で近似することにより、ボラティリティを時間に対して可変として精度を向上させている。これに対して、BN(Present) では、BN(Previous) で既知の株価リターンを予測して予測誤差を計測

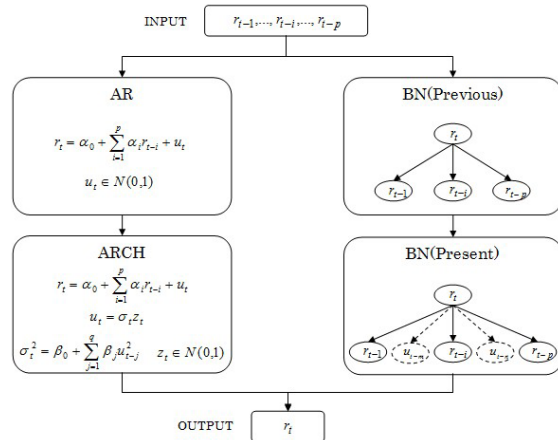


図 1 時系列分析法と提案手法の比較

Fig. 1 Comparison of time-series analysis and present schemes.

し、その誤差をもとに BN(Previous) で用いたベイジアンネットワークモデル自体を補正することで精度を改善しようとしている。結果として、ARCH モデルでは予測株価リターン r_t を過去リターン r_{t-i} と擾乱項 u_t の線形結合で近似するのに対して、BN(Present) では予測株価リターン r_t を過去リターン r_{t-i} と擾乱項 u_t の確率的依存関係に基づく非線形関係によってモデル化しているといえる。

3. ネットワーク探索アルゴリズム

3.1 Naive ベイズ構造と条件付き確率

確率変数 x_0 が x_i に依存していることを $x_i \rightarrow x_0$ と表現し、 x_i を親ノード、 x_0 を子ノードと呼ぶ。 x_0 の親ノードが複数ある場合、その親ノード集合を $Pa(x_0)$ と表現することにする。親ノード集合に属するノードを x_i 、その総数を M とすると $Pa(x_0)$ は次式のように定義される。

$$Pa(x_0) = \{x_1, x_2, \dots, x_M\} \quad (15)$$

子ノードの親ノード集合に対する依存関係は条件付き確率 $P(x_0|Pa(x_0))$ で表される。これは次式で与えられる。

$$P(x_0|Pa(x_0)) = \frac{P(x_0)P(Pa(x_0)|x_0)}{P(Pa(x_0))} \quad (16)$$

また、 $P(Pa(x_0)|x_0)$ は次式で与えられる。

$$P(Pa(x_0)|x_0) = \prod_{i=1}^M P(x_i|x_0) \quad (17)$$

3.2 K2Metric

本研究では、K2Metric^(2),7) をネットワークの評価値として採用する。全ノード数を N 、子ノード x_0 がとりうる状態の総数を L 、親ノード集合 $Pa(x_0)$ がとりうる状態の総数を M と表す。また、親ノード集合 $Pa(x_0)$ が状態 Y^j をノード x_0 が状態 X^k をとる場合の個数を N_{jk} とする。事前分布が一様分布であるとする、与えられたノード集合から構築されたネットワークについて、K2Metric^(2),7),8) は次式で与えられる。

$$K2 = \prod_{j=1}^M \frac{(L-1)!}{(N_j + L - 1)!} \prod_{k=1}^L N_{jk}! \quad (18)$$

ここで、 N_j と N_{jk} には次の関係がある。

$$N_j = \sum_{k=1}^L N_{jk} \quad (19)$$

3.3 グラフ構造の探索

ネットワークのグラフ構造を決定するために、K2 アルゴリズム⁽²⁾ を Naive ベイズ構造に対して修正したアルゴリズムを用いる。子ノードを x_0 とし、親ノードとなりうるノード x_i の総数を N とするときグラフ構造を探索するアルゴリズムは以下ようになる(図2)。

- (1) x_0 に対する親ノード集合 $Pa(x_0)$ を空集合 ϕ として定義する。
- (2) x_0 と $Pa(x_0)$ からなるネットワークの K2Metric を S_{best} とする。
- (3) $i = 1$ とする。
- (4) x_i を $Pa(x_0)$ に加える。
- (5) x_0 と $Pa(x_0)$ からなるネットワークの K2Metric を S とする。
- (6) $S \leq S_{best}$ ならば、 x_i を $Pa(x_0)$ から除外する。
- (7) $S > S_{best}$ ならば、 $S_{best} = S$ とする。
- (8) $i = i + 1$ とし、 $i \leq N$ ならば、ステップ (4) へ戻る。

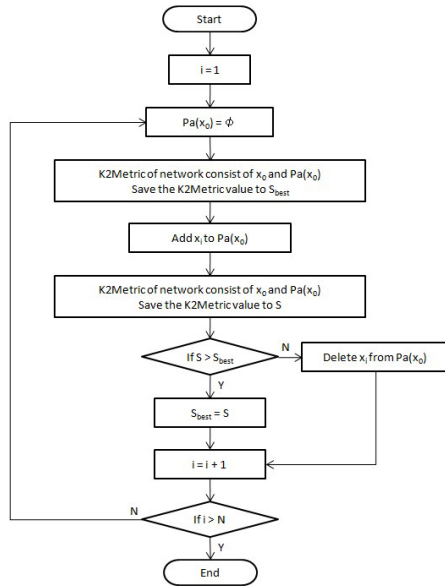


図2 グラフ構造探索アルゴリズム
Fig.2 Graph structure search algorithm.

(9) x_0 と $Pa(x_0)$ からなるネットワークを B とする.

4. 従来手法

4.1 株価リターンの離散化

本研究では、次式で定義される株価リターン¹⁾を扱う.

$$r_t = (\ln P_t - \ln P_{t-1}) \times 100 \quad (20)$$

ここで、 P_t は t 期の株価終値を、 r_t は t 期の株価リターンを示す. これは、株価指数の終値の日次データ⁴⁾から求める.

株価リターンは連続値だが、ベイジアンネットワークでノードが持つ確率変数では連続値を扱うことができない. そこで、何らかの方法によって株価リターン分布を離散化して離散値を得る. 以前の研究³⁾では、均等クラスタリング法とワード法によるクラスタリングを比較し、ワード法のほうが良い結果を示した(ワード法については、5.2節で説明する).

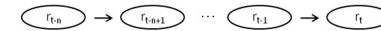


図3 株価リターンの全順序関係
Fig.3 Total order of stock return.

株価リターン離散値を r^l 、離散値総数を L とすると、離散値の集合は以下のように表される.

$$\{r^1, r^2, \dots, r^L\} \quad (21)$$

離散値総数 L は次のようにして定める異なる $L \geq 2$ について AIC を計算し、AIC が最小値を示す L を採用する. AIC は次式で定義される.

$$AIC = \ln \sigma^2 + \frac{2L}{T} \quad (22)$$

ここで、 σ^2 は残差標本の分散であり、 T は残差の総数である. σ^2 は次式で求められる.

$$\sigma^2 = \frac{1}{T - N} \sum_{t=N+1}^T (r_t - \bar{r}_t)^2 \quad (23)$$

ここで r_t は予測値を、 \bar{r}_t は実測値を示す. また、 N はネットワーク決定のために残差を求める標本数であり、 $N = 10$ とする.

株価リターン離散値を求める方法を箇条書きにすると以下ようになる.

- (1) 営業日ごとの株価リターンを式 (20) より求める.
- (2) 横軸を株価リターン、縦軸を頻度(その株価リターンをとった日数)として頻度分布図を描く.
- (3) 頻度分布図を異なる離散値総数(クラスタ数)でクラスタリングする.
- (4) 式 (22) で示される AIC が最小となる離散値総数 L とそのときの離散値を採用する.

4.2 ネットワークの構成と株価リターンの推定

クラスタリングにより離散化された過去の株価リターンをベイジアンネットワークに適用し、予測用のネットワークを決定する. ネットワークを決定するために、本研究では K2 アルゴリズム²⁾を用いる. そこで、株価リターンの時系列に基づく全順序関係に従って、ネットワークを決定する(図3).

離散化された過去の株価リターンから決定されたネットワークを B 、これを用いて株価リターン $r_t = r^l$ となる確率を推定した結果を $P(r^l|B)$ とする. t 期の株価リターンとして、 $P(r^l|B)$ が最大となる r^l を選択する. つまり、

$$r_t = \arg \max_{r^l} (P(r^l|B)) \quad (24)$$

4.3 従来手法のアルゴリズム

アルゴリズムを整理し直すと以下のようになる。

- (1) 株価リターンを離散化する。
- (2) 離散化された過去の株価リターンをノードとしてベイジアンネットワーク B を決定する。
- (3) B を用いて式 (24) より株価リターンを予測する。

5. 提案手法

5.1 提案手法のアルゴリズム

最初に提案手法のアルゴリズムについて示し、従来手法との違いについて述べる。

- (1) 4章で述べた従来手法により、すでに測定された過去の株価リターンを予測する。
- (2) 過去の株価リターンについて、予測値と実測値から誤差を求める。
- (3) 誤差分布をワード法によって離散化し、誤差の離散値を求める。
- (4) 過去の株価リターンと、それに対する評価から求めた誤差の離散値をノードとしてベイジアンネットワークを構築し、これを B' とする。
- (5) B' を用いて最尤推定して求めた株価リターンを予測値として採用する。

提案手法は、従来法に誤差を評価して離散化する過程を追加した手法となっている。そこで、以下では、誤差を評価して離散化する過程とネットワーク構築に関わる過程について詳しく述べることにする。

5.2 ウォード法

離散化のために、本研究ではウォード法を用いる。クラスタを C_i 、 C_i の重心を c_i 、データを z とする。ウォード法では、各対象からその対象を含むクラスタの重心までの距離の二乗の総和を最小化する。ウォード法の評価式は次式で表される。

$$D(C_i, C_j) = E(C_i \cup C_j) - E(C_i) - E(C_j) \quad (25)$$

ただし

$$E(C_i) = \sum_{z \in C_i} d(z, c_i)^2 \quad (26)$$

ここで、 $d(z, c_i)$ は z と c_i のユークリッド距離を示す。

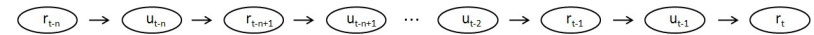


図 4 株価リターンと予測誤差の全順序関係

Fig. 4 Total order of stock return and prediction error.

5.3 予測誤差の離散化

すでに測定された過去の t 期において、式 (24) より求めた株価リターンの予測値を r_t 、実測値を \bar{r}_t とする。このとき、予測誤差 u_t は次式で与えられる。

$$u_t = r_t - \bar{r}_t \quad (27)$$

予測誤差は連続値なので、ベイジアンネットワークで適用できるように 5.2 節で示したワード法によって離散化する。

予測誤差離散値を u^l 、離散値総数を L とすると、予測誤差離散値の集合は以下のように表される。

$$\{u^1, u^2, \dots, u^L\} \quad (28)$$

離散値総数 L は次のようにして定める。異なる L について AIC を計算し、AIC が最小値を示す L を採用する。AIC は式 (22) と (23) で定義されたものを用いる。

予測誤差の離散値を求める方法を箇条書きにすると以下のようになる。

- (1) 営業日ごとの予測誤差を式 (27) より求める。
- (2) 横軸を予測誤差、縦軸を頻度（その予測誤差をとった日数）として頻度分布図を描く。
- (3) 頻度分布図を異なる離散値総数（クラスタ数）でクラスタリングする。
- (4) 式 (22) で示される AIC が最小となる離散値総数 L とそのときの離散値を採用する。

5.4 ネットワークの構成と株価リターンの推定

過去の株価リターン離散値とその予測誤差の離散値を用いて、図 4 に示す全順序関係に従って、3 章のアルゴリズムによりネットワークを決定する。

決定されたネットワークを B' 、これを用いて株価リターン $r_t = r^l$ となる確率を推定した結果を $P(r^l|B')$ とする。 $P(r^l|B')$ が最大となる r^l を t 期の株価リターンとする。つまり、

$$r_t = \arg \max_{r^l} (P(r^l|B')) \quad (29)$$

6. 数値実験

6.1 日経平均株価への適用

最初の例題として日経平均株価をとる。1985 年 2 月 22 日から 2008 年 12 月 30 日までの

表 1 日経平均株価の株価リターンにおける離散値数と AIC の関係
Table 1 Discrete number versus AIC on Nikkei stock average return.

Discrete number L	AIC
2	2.0442
3	1.7830
4	1.6066
5	1.5193
6	1.4597
7	1.8168
8	1.7478
9	1.8126
10	1.7729

日次株価リターンを用いてベイジアンネットワークを決定し、それを用いて 2008 年 12 月 1 日から 2008 年 12 月 30 日の株価リターンを予測する。そして、本研究で提案する手法を時系列分析手法や従来の研究で提案した方法³⁾と比較する。

すでに述べたように、時系列分析手法のモデルパラメータ p, q は、これらを 1 から 10 までとってシミュレーションを行い、AIC が最も小さな値を示す場合のモデルパラメータを用いる。

ベイジアンネットワークを用いた手法のうち従来法と提案手法についても、1 期前から 10 期前までの株価リターンや予測誤差のノードを用いてベイジアンネットワークを決定する。

6.1.1 ネットワークの決定

まず、株価リターンだけをノードとしてネットワークを決定する。このとき、株価リターンの離散値数によって結果が異なるので、離散値総数 $L = 2, 3, \dots, 10$ についてベイジアンネットワークを決定し、AIC (式 (22)) が最も小さな値を示すものを用いる。異なる離散値総数に対して決定されたネットワークの AIC を表 1 に示す。これより、離散値総数 $L = 6$ のとき AIC が最も小さいことが分かる。このときの具体的な離散値やクラスタのパラメータを表 2 に示す。ここで、 $c_l(r^l)$ は離散値を示し、 $(C_l)_{\min}, (C_l)_{\max}$ はその離散値が代表するクラスタに属するサンプルの最小値と最大値を示す。4 章のアルゴリズムにより、株価リターンのみで決定されたネットワークを図 5 に示す³⁾。この場合、株価リターン r_t は 5 日前の株価リターン r_{t-5} 、7 日前の株価リターン r_{t-7} 、9 日前の株価リターン r_{t-9} 、10 日前の株価リターン r_{t-10} と関連づけられていることが分かる。

次に、株価リターンとその誤差の両方をノードとしてネットワークを決定する。株価リターンの離散値総数を $L = 6$ として固定したうえで、誤差の離散値総数を $L = 2, 3, \dots, 10$

表 2 日経平均株価リターンの離散値

Table 2 Discrete number set of Nikkei stock average return.

Cluster	$(C_l)_{\min}, (C_l)_{\max}$	$c_l(r^l)$
C_1	$[-16.138\%, -3.000\%]$	-4.30%
C_2	$[-3.000\%, -0.730\%]$	-1.48%
C_3	$[-0.730\%, -0.065\%]$	-0.37%
C_4	$(-0.065\%, 0.947\%]$	0.39%
C_5	$(0.947\%, 3.800\%]$	1.73%
C_6	$(3.800\%, 13.235\%]$	5.45%

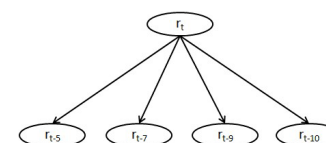


図 5 株価リターンのみで決定されたベイジアンネットワーク (日経平均株価)

Fig. 5 Bayesian network determined with Nikkei stock average return only.

表 3 日経平均株価予測誤差における離散値数と AIC の関係

Table 3 Discrete number versus AIC on forecast error of Nikkei stock average return.

Discrete number L	AIC
2	1.6659
3	1.4587
4	1.5764
5	1.5132
6	1.3203
7	1.4596
8	1.4599
9	1.4602
10	1.4606

としてベイジアンネットワークを決定し、AIC (式 (22)) を計算した結果を表 3 に示す。これより、離散値総数 $L = 6$ のとき AIC が最も小さいことが分かる。このときの具体的な離散値やクラスタのパラメータを表 4 に示す。次に、5 章のアルゴリズムにより、株価リターンと予測誤差で決定されたネットワークを図 6 に示す。この場合、株価リターン r_t は 5 日前の株価リターン r_{t-5} 、7 日前の株価リターン r_{t-7} 、9 日前の株価リターン r_{t-9} 、10 日前の株価リターン r_{t-10} 、2 日前の誤差 u_{t-2} 、7 日前の誤差 u_{t-7} と関連づけられていること

表 4 日経平均株価リターン予測誤差の離散値

Table 4 Discrete number set of forecast error of Nikkei stock average return.

Cluster	$(C_l)_{\min}, (C_l)_{\max}$	$c_l(u^l)$
C_1	$[-16.67\%, -2.87\%]$	-4.13%
C_2	$[-2.87\%, -0.92\%]$	-1.59%
C_3	$[-0.92\%, -0.00\%]$	-0.47%
C_4	$[0.00\%, 1.40\%]$	0.55%
C_5	$[1.40\%, 3.71\%]$	2.26%
C_6	$[3.71\%, 16.74\%]$	5.33%

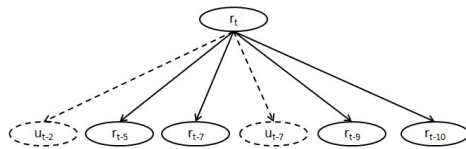


図 6 株価リターンと誤差で決定されたベイジアンネットワーク (日経平均株価)

Fig. 6 Bayesian network determined with Nikkei stock average return and error.

表 5 予測値と実測値の比較 (日経平均株価)

Table 5 Comparison of predicted and actual stock prices (Nikkei stock average).

	MaxE	MinE	CC	RMSE
AR(2)	6.4665	0.0214	0.6929	2.7477
MA(2)	6.5148	0.0174	0.6912	2.7538
ARMA(2,2)	6.6085	0.2443	0.6855	2.7696
ARCH(2,9)	6.4525	0.0976	0.6932	2.7440
BN(Previous)	4.2383	0.2442	0.8371	2.0727
BN(Present)	4.2383	0.1192	0.8629	1.9331

が分かる。予測誤差ノードを取り入れると、異なるネットワークが導かれることが分かる。

6.1.2 予測精度の比較 (1)

求めたネットワークで予測を行った結果を、時系列分析法や従来法と比較した結果を表 5 と図 7 に示す。ここで、AR, MA, ARMA, ARCH は、それぞれの時系列分析法による結果を、BN(Previous) は図 5 のネットワークによる予測結果を、BN(Present) は図 6 のネットワークによる予測結果を示す。それぞれの数値は相関係数 (CC), 平均二乗誤差 (RMSE), 最大誤差 (MaxE), 最小誤差 (MinE) を示す。

表 5 と図 7 から提案手法は最も大きな相関係数を示しており、平均二乗誤差と最大誤差

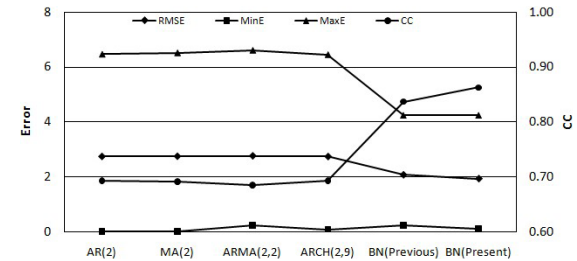


図 7 予測値と実測値の比較 (日経平均株価)

Fig. 7 Comparison of predicted and actual stock prices (Nikkei stock average).

表 6 予測期間が異なる場合の予測値と実測値の比較 (日経平均株価)

Table 6 Comparison of predicted and actual stock prices at different prediction period (Nikkei stock average).

	MaxE	MinE	CC	RMSE
AR(2)	5.7648	0.0604	0.9278	2.4923
MA(2)	5.7746	0.0758	0.9276	2.4953
ARMA(2,2)	5.9389	0.0076	0.9250	2.5150
ARCH(2,9)	5.8342	0.0178	0.9268	2.5116
BN(Previous)	6.2948	0.0057	0.9099	2.9905
BN(Present)	4.9016	0.0105	0.9375	2.3467

も最小となっていることが分かる。最小誤差はどの方法でも同程度であるが、平均二乗誤差では提案手法は時系列分析法よりも 30%程度、BN(Previous) よりも 6%程度改善されていることが分かる。最大誤差については、BN(Present) と BN(Previous) の精度は同程度となっている。これは BN(Present) と BN(Previous) のいずれにおいても、予測精度の改善とは予測確率の改善であって、予測値として示される株価リターンは表 2 における $c_l(r^l)$ の 6 種類でしかないことによると考えられる。

6.1.3 予測精度の比較 (2)

前項の実験結果では、予測期間は学習期間に含まれていた。そこで、学習期間に含まれない期間の株やりターンを予測することを考える。上記で示した手法を用いて、2009 年 1 月 1 日から 3 月 31 日における取引日での株価リターン予測を行う。

提案手法を時系列分析法や従来法と比較した結果を表 6 と図 8 に示す。これらの中で提案手法は最も大きな相関係数を示している。最小誤差はどの方法でも大きな違いはないけ

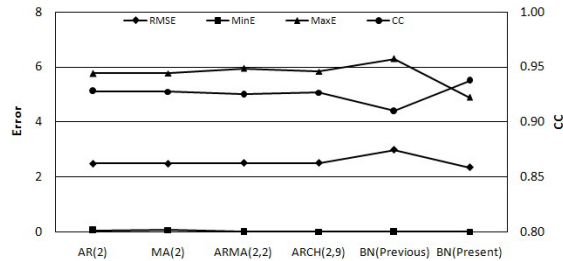


図 8 予測期間が異なる場合の予測値と実測値の比較 (日経平均株価)

Fig. 8 Comparison of predicted and actual stock prices at different prediction period (Nikkei stock average).

表 7 従来法と提案法の計算時間の比較

Table 7 Comparison of calculation cost of previous and present models.

	Network Search	Forecast
BN(Previous)	529 sec	0.691 sec
BN(Present)	908 sec	0.887 sec

れども、平均二乗誤差と最大誤差では提案手法が最も小さい値を示しており、提案手法の平均二乗誤差は、時系列分析法よりも 6%程度、最大誤差は 15%程度改善されていることが分かる。

6.1.4 計算時間

誤差ノードを用いる提案法は、誤差ノードを用いない従来法に比べて計算時間が増えることが予想される。そこで、日経平均株価リターン予測について、提案法と従来法の計算時間を比較する。計算時間は、ベイジアンネットワークの学習にかかる時間と、決定されたベイジアンネットワークを用いて株価リターンを予測するためにかかる時間に分けて示す。

比較結果を表 7 に示す。これより、従来法ではネットワーク決定に 529 秒かかり、株価リターン予測に 0.691 秒かかることが分かる。一方、提案法では、それぞれ 908 秒と 0.887 秒である。このように、株価リターン予測時間はネットワーク決定時間よりかなり短いことが分かる。ネットワーク決定時間をみると、提案法は従来法のおよそ 1.7 倍となっている。計算時間は増大しているけれども、最初にネットワーク決定のために 1 回だけ計算すればよい。予測にかかる時間について検討すると提案法は従来法の 1.3 倍程度であり、ネットワーク決定にかかる時間に比べれば計算時間の増加は比較的小さいといえる。

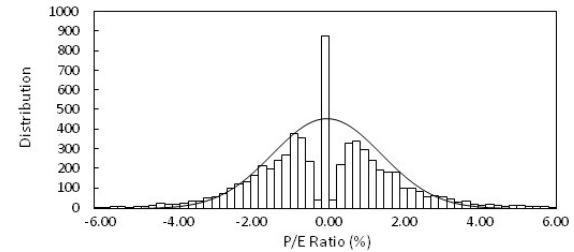


図 9 トヨタ自動車の株価リターンの頻度分布図

Fig. 9 Return histogram of Toyota Motor Corp. stock price.

6.2 トヨタ自動車株価への適用

次の例題として、トヨタ自動車の株価データを取り上げる。トヨタ自動車株価の日次リターンの頻度分布を図 9 に示す。この図より、分布が正規分布から外れており、特に、リターン 0%付近にかなり多くのデータが分布していることが分かる。

ベイジアンネットワークを決定するために 1985 年 2 月 22 日から 2008 年 12 月 30 日までの日次株価リターンを使用する。このベイジアンネットワークを用いて 2008 年 12 月 1 日から 2008 年 12 月 30 日の株価リターンを予測する。そして、本研究で提案する手法を時系列分析法や従来の研究で提案した方法³⁾と比較する。日経平均の場合と同様に、時系列分析法のモデルパラメータ p, q は、これらを 1 から 10 までとってシミュレーションを行い、AIC (式 (22)) が最も小さな値を示す場合のモデルパラメータを用いる。

また、ベイジアンネットワークを用いた手法のうち従来法と提案手法についても、1 期前から 10 期前までの株価リターンや予測誤差のノードを用いてベイジアンネットワークを決定する。

6.2.1 ネットワークの決定

まず、株価リターンだけをノードとしてネットワークを決定する。このとき、株価リターンの離散値数によって結果が異なるので、離散値総数 $L = 2, 3, \dots, 10$ についてベイジアンネットワークを決定し、AIC (式 (22)) が最も小さな値を示すものを用いる。異なる離散値総数に対して決定されたネットワークの AIC を表 8 に示す。これにより、離散値総数 $L = 5$ のとき AIC が最も小さいことが分かる。このときの具体的な離散値やクラスタのパラメータを表 9 に示す。4 章のアルゴリズムにより、株価リターンのみで決定されたネットワークを図 10 に示す³⁾。この場合、株価リターン r_t は 1 日前の株価リターン r_{t-1} 、4 日前の株価リターン r_{t-4} 、6 日前の株価リターン r_{t-6} 、8 日前の株価リターン r_{t-8} 、9 日前

表 8 トヨタ自動車株価リターンにおける離散値数と AIC の関係

Table 8 Discrete number versus AIC on Toyota Motor Corp. stock return.

Discrete number L	AIC
2	2.8452
3	2.5892
4	2.3913
5	2.3243
6	2.5309
7	2.7240
8	2.4935
9	2.5941
10	2.5994

表 9 トヨタ自動車株価リターンの離散値

Table 9 Discrete number set of Toyota Motor Corp. stock return.

Cluster	$(C_l)_{\min}, (C_l)_{\max}$	$c_l(r^l)$
C_1	$[-21.146\%, -3.900\%]$	-5.55%
C_2	$[-3.900\%, -1.285\%]$	-2.10%
C_3	$[-1.285\%, 0.000\%]$	-0.47%
C_4	$[0.000\%, 2.530\%]$	1.08%
C_5	$[2.530\%, 16.264\%]$	4.12%

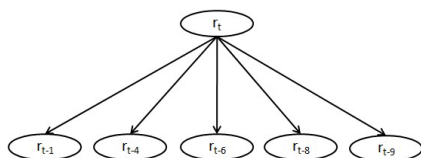


図 10 トヨタ自動車株価の株価リターンのみで決定されたベイジアンネットワーク

Fig.10 Bayesian network determined with Toyota Motor Corp. stock return only.

の株価リターン r_{t-9} と関連づけられていることが分かる。

次に、株価リターンとその誤差の両方をノードとしてネットワークを決定する。株価リターンの離散値総数を $L = 5$ として固定したうえで、誤差の離散値総数を $L = 2, 3, \dots, 10$ としてベイジアンネットワークを決定し、AIC (式 (22)) を計算した結果を表 10 に示す。これより、離散値総数 $L = 3$ のとき AIC が最も小さいことが分かる。このときの具体的な離散値やクラスターのパラメータを表 11 に示す。次に、5 章のアルゴリズムにより、株価リターンと予測誤差で決定されたネットワークを図 11 に示す。この場合、株価リターン r_t

表 10 トヨタ自動車株価リターン予測誤差における離散値数と AIC の関係

Table 10 Discrete number versus AIC on forecast error of Toyota Motor Corp. stock return.

Discrete number L	AIC
2	2.3225
3	2.0729
4	2.1618
5	2.4284
6	2.1506
7	2.3820
8	2.3823
9	2.6758
10	2.7555

表 11 トヨタ自動車株価リターン予測誤差の離散値

Table 11 Discrete number set of forecast error of Toyota Motor Corp. stock price return.

Cluster	$(C_l)_{\min}, (C_l)_{\max}$	$c_l(u^l)$
C_1	$[-25.262\%, -2.770\%]$	-4.36%
C_2	$[-2.770\%, 0.139\%]$	-0.98%
C_3	$[0.139\%, 19.980\%]$	1.66%

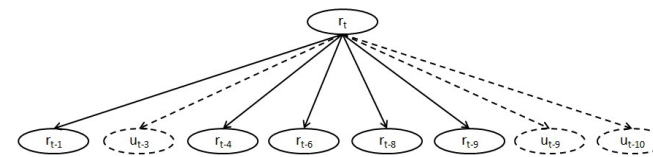


図 11 トヨタ自動車株価の株価リターンと予測誤差で決定されたベイジアンネットワーク

Fig.11 Bayesian network determined with Toyota Motor Corp. stock return and error.

は 1 日前の株価リターン r_{t-1} , 4 日前の株価リターン r_{t-4} , 6 日前の株価リターン r_{t-6} , 8 日前の株価リターン r_{t-8} , 9 日前の株価リターン r_{t-9} , 3 日前の誤差 u_{t-3} , 9 日前の誤差 u_{t-9} , 10 日前の誤差 u_{t-10} と関連づけられていることが分かる。予測誤差ノードを取り入れると、異なるネットワークが導かれることが分かる。

6.2.2 予測精度の比較 (1)

提案手法による予測結果を時系列分析法や従来法と比較した結果を表 12 と図 12 に示す。ここで、AR, MA, ARMA, ARCH は、それぞれの時系列分析法による結果を、BN(Previous) は図 10 のネットワークによる予測結果を、BN(Present) は図 11 のネット

表 12 予測値と実測値の比較 (トヨタ自動車株価)

Table 12 Comparison of predicted and actual stock prices (Toyota Motor Corp.).

	MaxE	MinE	CC	RMSE
AR(9)	10.4027	0.0937	0.4708	4.0789
MA(6)	10.3414	0.2012	0.4699	4.0853
ARMA(9,6)	10.5354	0.0711	0.4454	4.1409
ARCH(9,9)	10.2655	0.0302	0.4706	4.0709
BN(Previous)	6.8142	0.0491	0.6469	3.1942
BN(Present)	5.2172	0.1327	0.7056	2.8178

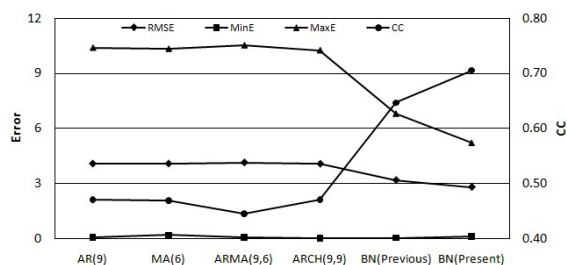


図 12 予測値と実測値の比較 (トヨタ自動車株価)

Fig. 12 Comparison of predicted and actual stock prices (Toyota Motor Corp.).

ワークによる予測結果を示す。それぞれの数値は相関係数 (CC), 平均二乗誤差 (RMSE), 最大誤差 (MaxE), 最小誤差 (MinE) を示す。

図 12 から, 提案手法は最も大きな相関係数を示しており, 平均二乗誤差と最大誤差も最小となっていることが分かる。また, 最小誤差はどの方法でも同程度である。平均二乗誤差をみると, 提案手法は時系列分析法よりも 30%程度向上している。従来手法 BN(Previous) と比べると, 平均誤差は 12%程度, 相関係数は 9%程度改善されている。

6.2.3 予測精度の比較 (2)

前項の実験結果では, 予測期間は学習期間に含まれていた。そこで, 学習期間に含まれない期間の株やリターンを予測することを考える。上記で示した手法を用いて, 2009年1月1日から3月31日の取引日における株価リターン予測を行う。

提案手法を時系列分析法や従来法と比較した結果を表 13 と図 13 に示す。これらの中で提案手法は最も大きな相関係数を示している。また, 最小誤差はどの方法でも 0 に近い値を示し, 平均二乗誤差と最大誤差は提案手法が最も小さい値を示している。提案手法の平均二

表 13 予測期間が異なる場合の予測値と実測値の比較 (トヨタ自動車)

Table 13 Comparison of predicted and actual stock prices at different prediction period (Toyota Motor Corp.).

	MaxE	MinE	CC	RMSE
AR(9)	7.5091	0.0615	0.7448	2.6657
MA(6)	7.6259	0.0319	0.7417	2.6859
ARMA(9,6)	7.1204	0.0401	0.7427	2.6739
ARCH(9,9)	8.0839	0.0597	0.7527	2.6992
BN(Previous)	7.8579	0.0415	0.7122	3.1494
BN(Present)	5.7047	0.0818	0.7984	2.6313

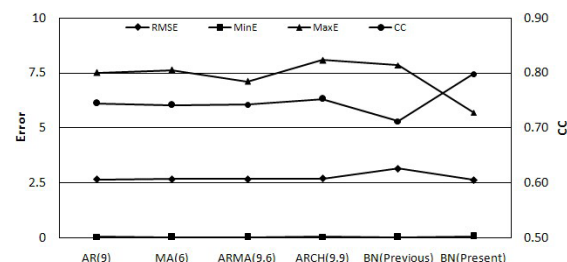


図 13 予測期間が異なる場合の予測値と実測値の比較 (トヨタ自動車)

Fig. 13 Comparison of predicted and actual stock prices at different prediction period (Toyota Motor Corp.).

乗誤差は時系列分析法よりも 2%程度, 最大誤差は 20%程度改善されていることが分かる。

7. 結 論

本研究では, 著者らがこれまでに提案しているベイジアンネットワークを用いた株価リターンの予測法の精度改善手法について述べた。

従来法では, 過去の株価リターンが特定の値をとるときに, 次の時点の株価リターンが特定の値をとる確率を推測するためのベイジアンネットワークを構築し, これを将来株価の予測に用いる。これに対して, 提案した方法では, 最初に従来法で過去の既知の株価リターンを予測し, 予測精度を評価する。この予測精度を株価リターンとともに用いてノードを定義してベイジアンネットワークを構築する。

例題として日経平均とトヨタ自動車の株価リターン予測を扱った。構成したネットワークを比較すると, いずれの場合においても, 提案手法のネットワークには予測誤差ノードが含

まれており、過去の予測精度が将来の予測に影響していることが分かった。

日経平均の株価リターンに適用した結果、平均誤差では、提案手法は時系列分析法よりも30%程度、ベイジアンネットワークを用いた従来手法よりも6%程度改善された。また、トヨタ自動車の株価リターンに適用した場合、平均誤差をみると、提案手法は時系列分析法よりも30%程度向上している。ベイジアンネットワークを用いた従来手法と比べても平均誤差は12%程度、相関係数は9%程度改善された。これらのことより、提案手法に一定の有効性があることが示された。

計算時間について従来法と比較したところ、株価リターン予測にかかる計算時間に比べてネットワーク決定にかかる計算時間の比重が大きかったことが分かったので、今後はこの時間を縮小する方法について検討を進めていく予定である。

参 考 文 献

- 1) 渡部敏明：日本の株式市場におけるボラティリティの変動—ARCH型モデルによる分析—，三菱経済研究所 (1997).
- 2) 繁樹算男，本村陽一，植野真臣：ベイジアンネットワーク概説，培風館 (2006).
- 3) 左 毅，北 栄輔：ベイジアンネットワークを用いた株価予測について，情報処理学会論文誌 数理モデル化と応用，Vol.3, No.3, pp.80-90 (2010).
- 4) Yahoo!ファイナンス，入手先(<http://quote.yahoo.co.jp/>).
- 5) Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation, *Econometrica*, Vol.50, pp.987-1008 (1982).
- 6) 原 章，長尾智晴：自動グループ構成手法 ADG を用いた人工株式市場の構築，情報処理学会論文誌，Vol.41, No.4, pp.1063-1072 (2000).
- 7) Heckerman, D., Geiger, D. and Chickering, D.: Learning Bayesian networks: The combination of knowledge and statistical data, *Machine Learning*, Vol.20, No.3, pp.197-243 (1995).
- 8) Cooper, G.F. and Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data, *Machine Learning*, Vol.9, No.4, pp.309-347 (1992).

(平成 23 年 4 月 18 日受付)

(平成 23 年 6 月 6 日再受付)

(平成 23 年 7 月 8 日再受付 (2))

(平成 23 年 8 月 22 日採録)



左 毅 (学生会員)

1981 年生。名古屋大学大学院情報科学研究科博士課程在学中。ベイジアンネットワーク、金融モデルの研究に従事。



原田 昌朗

1988 年生。名古屋大学大学院情報科学研究科修士課程在学中。ベイジアンネットワークの経済社会問題への応用研究に従事。



北 栄輔 (正会員)

1964 年生。1991 年名古屋大学大学院工学研究科機械工学専攻博士課程修了。2009 年より名古屋大学大学院教授。人工市場、交通シミュレーション、ベイジアンネットワーク、数値解析理論等の研究に従事。日本機械学会、日本計算工学会、計算数理工学会等会員。