情報処理学会論文誌　数理モデル化と応用　Vol. 3　No. 3　80–90 (Oct. 2010)

# ベイジアンネットワークを用いた株価予測について

左　　毅[†1]　北　栄輔[†1]

株価予測のために，過去データを用いた時系列分析に基づくモデルがしばしば用いられる．これらのモデルでは，予測したい株価を過去の株価などと残差の線形和で近似し，分布には正規分布が仮定される．しかし，実データに関するいくつかの知見によれば，株価収益率の頻度分布は必ずしも正規分布しないことが示されている．そうであれば，ホワイトノイズに基づくモデルでは精度良く予測できない可能性がある．そこで，本研究では，ベイジアンネットワークを用いる方法を示す．ところで，ベイジアンネットワークは離散的な値だけしか扱うことができないので，クラスタリング手法を用いて株価を離散値に変換する．解析例では，日経平均株価とトヨタ自動車株価をとる．解析結果より，収益率を適切に離散化すれば，相関係数解析や平均二乗誤差において，一般的な時系列分析モデルよりも精度良く予測できることが分かった．

# Stock Prices Prediction by Using Bayesian Network

Yi Zuo[†1] and Eisuke Kita[†1]

Time-series prediction algorithms are very often employed for predicting the stock price. In the algorithms, the stock price is assumed to be the weighted summation of the past stock price and the residual. The distribution is assumed to follow the normal distribution. However, recent results show that the stock price fluctuation dose not follow the normal distribution. Therefore, in this study, the Bayesian network is employed for the stock price prediction. The Bayesian network can deal with the discrete numbers alone. For the use of clustering algorithm, the stock price, which is expressed in the continuous number, has to be transformed to the discrete number. NIKKEI stock average (NIKKEI225) and TOYOTA Motor Corporation stock price are considered as numerical examples. The results show that the prediction accuracy of the present algorithm is better than that of the time-series prediction algorithms in the correlation coefficient analysis and the prediction error if the stock price is appropriately transformed to the discretized number.

## 1. 緒　　論

株価の精度良い予測は，投資家個人が投資行動を決定するためだけでなく，企業の経営方針や国家運営など様々な問題に関わっている．そこで，様々な株価の予測方法が提案されている．

株価予測には時系列分析に基づく方法が広く用いられている[1)]．この中には，AR モデル，MA モデル，ARMA モデル，ARCH モデルなどがある．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，今期株価を過去の撹乱項の移動加重和で近似する．ARMA モデルは AR と MA を組み合わせたモデルである．さらに，Auto Regressive Conditional Heteroskedasticity（ARCH）モデル[2)] 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ている．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収益率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，株価収益率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れば，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．したがって，そのような場合には予測精度が低下することが予想される．そこで，本研究では，ベイジアンネットワーク[3)] を用いて株価変化を推定することを提案する．

ベイジアンネットワークは，確率的なネットワークモデルである．確率変数をノードで表し，変数間の因果関係を非循環有向リンクで表す．さらに，変数間の定量的な依存関係を条件付き確率によって表す．本研究では，このベイジアンネットワークを用いて，過去の株価収益率から今期株価を予測する．このとき，株価収益率の分布を正規分布などによって仮定する必要がない．ところで，ベイジアンネットワークでは株価のような連続値をとる変数を扱うことができない．そこで，本研究では，クラスタリング手法を用いて株価を離散値に変換する．各ノードは，離散化された株価を確率変数として有し，各ノードの離散値間の因果関係をベイジアンネットワークによって推定する．

解析例として日経平均株価とトヨタ自動車の株価を用いる[4)]．日経平均株価は，その収益

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

率が正規分布に比較的近い場合の例として扱うのに対して，トヨタ自動車株価は収益率が正規分布から離れている例として用いる[5]．そして，提案手法による予測結果の精度を AR，MA，ARMA，ARCH と比較し，提案手法の有効性を検討する．また，ベイジアンネットワークによる予測精度は，株価を離散値に変換する方法にも依存する．そこで，頻度分布を均等に分割して離散化する方法のほかに，クラスタリング手法としてウォード法を用いて離散化する方法を示す．

本論文の構成は以下のようになっている．2 章では研究の背景として，時系列分析に基づく予測法と提案手法について説明する．3 章でベイジアンネットワークのアルゴリズムについて説明した後，4 章では，提案手法について説明する．5 章では解析結果を示し，6 章はまとめである．

## 2. 研 究 背 景

### 2.1 時系列分析法[1]

#### 2.1.1 AR モデル

AR モデルは自己回帰モデルである．今期の株価収益率を過去の株価収益率で表す．AR モデルを，以下では AR($p$) と表記する．$t$ 期の株価収益率を $r_t$ とすると，これは $p$ 期前までの収益率と，攪乱項 $u_t$ によって次のように表される．

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

ここで，$\alpha_i$（$i = 0, \cdots, p$）はモデルのパラメータである．また，$u_t$ は平均 0，分散 $\sigma^2$ のホワイトノイズに従う．

#### 2.1.2 MA モデル

MA モデルは移動平均モデルである．MA モデルを，以下では MA($q$) と表記する．このモデルでは，$t$ 期の株価収益率 $r_t$ を過去の撹乱項によって以下のように表す．

$$r_t = \beta_0 + \sum_{j=1}^{q} \beta_j u_{t-j} + u_t \tag{2}$$

ここで，$\beta_j$（$j = 0, \cdots, q$）はモデルのパラメータである．

#### 2.1.3 ARMA

ARMA モデルは，AR モデルと MA モデルを合わせたモデルであり，自己相関移動平均モデルとよばれる．ARMA モデルを，以下では ARMA($p, q$) と表記する．$t$ 期の株価収益率 $r_t$ は，株価収益率 $r_t$ と攪乱項 $u_t$ によって次のように表される．

$$r_t = \sum_{i=1}^{p} \alpha_i r_{t-i} + \sum_{j=1}^{q} \beta_j u_{t-j} + u_t \tag{3}$$

#### 2.1.4 ARCH

ARCH は Auto Regressive Conditional Heteroscedastic の略称で，Engle ら[2] により最初提案されたボラティリティ変動モデルの 1 つである．以下では，ARCH($p, q$) と表記する．

$t$ 期の株価収益率 $r_t$ は次のように表される．

$$r_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i} + u_t \tag{4}$$

ここで，攪乱項 $u_t$ は次式で与えられる．

$$u_t = \sigma_t z_t \tag{5}$$

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 モデルによれば，$\sigma_t^2$ は次式で近似される．

$$\sigma_t^2 = \beta_0 + \sum_{j=1}^{q} \beta_j u_{t-j}^2 \tag{6}$$

ARCH モデルから，様々な改良モデルが提案されている．

#### 2.1.5 モデル次数の決定

それぞれのモデルの次数 $p$，$q$ は，$p = 0, 1, \cdots, 10$ と $q = 0, 1, \cdots, 10$ について赤池情報量基準 AIC を求める．そして，AIC が最小となる $p$，$q$ を採用する．AIC は次式から評価する．

$$\mathrm{AIC} = \ln \hat{\sigma}^2 + \frac{2(p+q)}{T} \tag{7}$$

ここで，$\hat{\sigma}$ はモデルの残差 $\epsilon_1, \epsilon_2, \cdots, \epsilon_T$ の標本分散であり，$T$ は残差の総数である．

### 2.2 提案手法の概要

AR，MA，ARMA，ARCH モデルでは株価データの分布が正規分布に従うことを仮定している．しかし，実際の株価変動の解析から，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．例として，日経平均株価の株価収益率を**図 1** に示す．この頻

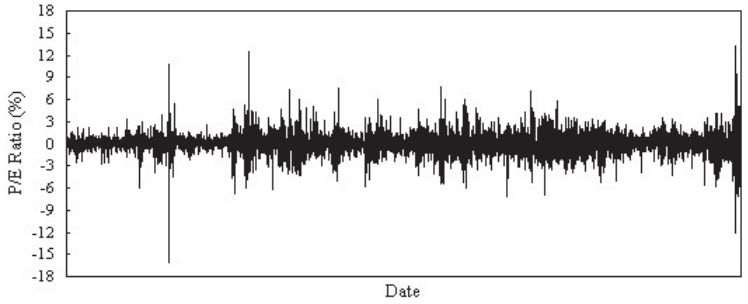


**図 1**　日経平均株価の株価収益率
Fig. 1　P/E ratio of NIKKEI stock average.

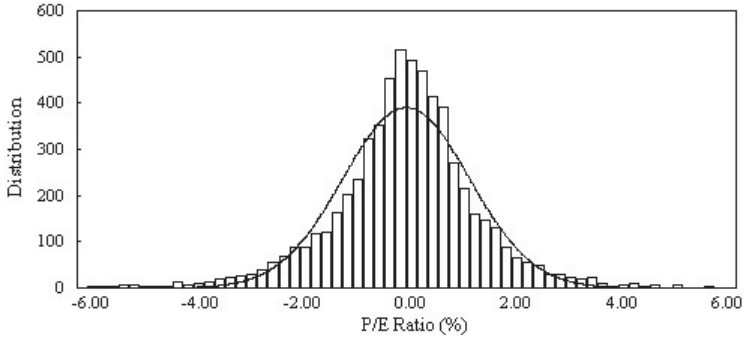


**図 2**　日経平均株価の株価収益率の頻度分布図
Fig. 2　Histgram of P/E ratio of NIKKEI stock average.

度分布に，それから求めた正規分布を重ねた図が**図 2** である．これより，分布は正規分布に近いけれども，少し異なっていることが分かる．後述するように，このずれは一般企業の株価ではもっと大きい場合がある．原ら[5] は TOPIX の頻度分布を解析し，ボラティリティを $\sigma$ とすると，実際の頻度分布ではイベントが $\pm\sigma$ 内にある頻度は正規分布よりも大きく，また，$\pm 3\sigma$ の外にある頻度は正規分布の 3 倍ほどであることを指摘している．このことを改善するために，頻度分布の確率密度関数の尖度や突度を修正する研究が示されている．これに対して，本研究ではベイジアンネットワークを用いて株価収益率の時系列分析を行う．

ベイズ推定によれば，過去に起きた事象の発生頻度から未来に起こる事象の発生頻度を求

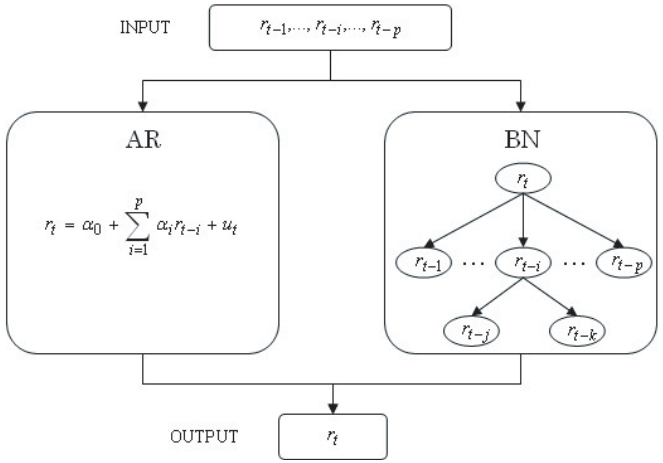


**図 3**　時系列分析法と提案手法の比較
Fig. 3　Comparison of time-series analysis and present models.

めることができる．そして，ベイズ推定に基礎をおくベイジアンネットワークを用いれば，確率変数をノードで，変数間の因果関係をリンクで表すことで因果関係を表す非循環有向グラフを得ることができる．本研究では，このベイジアンネットワークを用いて，過去の株価収益率から今期株価収益率を予測する．この方法では，株価収益率の分布を正規分布などによって仮定する必要がない．ただし，ベイジアンネットワークではノードにおかれる確率変数は連続値をとることができないので，クラスタリング手法などを用いて離散化する操作が必要である．

従来の時系列分析手法では，今期の株価収益率を過去の説明変数と攪乱項の線形結合で近似している．たとえば，AR モデルでは株価収益率を過去の株価収益率と攪乱項で表す．これに対して，提案手法では今期株価収益率と過去の株価収益率との間の非線形関係式をベイジアンネットワークによって定義して予測するという見方もできる（**図 3**）．

## 3. ベイジアンネットワーク

### 3.1 ベイジアンネットワークと条件付き確率表

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とし，依存関係をリンクとする非循環有向グラフで可視化する．さらに，変数間の定量的な依存関

係を変数間の条件付き確率によってモデル化する．

確率変数 $x_i$ が $x_j$ に依存していることを $x_j \to x_i$ と表現し，$x_j$ を親ノード，$x_i$ を子ノードとよぶ．$x_i$ の親ノードが複数ある場合，その親ノード集合を $Pa(x_i)$ と表現することにする．子ノードの親ノードに対する依存関係は条件付き確率 $P(x_i|Pa(x_i))$ で表される．

ここで，親ノード集合 $Pa(x_i)$ がとりうる状態が $M$ 個あり，その 1 つを記号 $Y^m$ とする．一方，子ノード $x_i$ のとりうる状態が $L$ 個あり，その 1 つを $X^l$ とする．このとき

$$\begin{matrix} P(X^1|Y^1), P(X^2|Y^1), \cdots, P(X^L|Y^1) \\ \vdots \qquad\qquad\qquad \vdots \\ P(X^1|Y^M), P(X^2|Y^M), \cdots, P(X^L|Y^M) \end{matrix}$$

を表にしたものを条件付き確率表（CPT）とよぶ．

### 3.2　グラフ構造の決定

本研究では，K2Metric[3),6)] をネットワークの評価値として採用し，K2 アルゴリズム[3)] を用いてネットワークのグラフ構造を決定する．

全ノード数を $N$，子ノード $x_i$ がとりうる状態の総数を $L$，親ノード集合 $Pa(x_i)$ がとりうる状態の総数を $M$ と表す．また，ノード $x_i$ について，その親ノード集合 $Pa(x_i)$ が状態 $Y^j$ を，ノード $x_i$ が状態 $X^k$ をとる場合の個数を $N_{ijk}$ とする．事前分布が一様分布であるとすると，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[3),6),7)] は次式で与えられる．

$$K2 = \prod_{i=1}^{N} \prod_{j=1}^{M} \frac{(L-1)!}{(N_{ij}+L-1)!} \prod_{k=1}^{L} N_{ijk}! \tag{8}$$

ここで，$N_{ij}$ と $N_{ijk}$ には次の関係がある．

$$N_{ij} = \sum_{k=1}^{L} N_{ijk} \tag{9}$$

K2 アルゴリズム[3)] はよくばり法を元に作られており，全木探索より少ない計算量でネットワークを構築できる．このアルゴリズムでは，あらかじめ変数間の全順序関係が分かっている必要がある．本研究で扱う株価では，時系列に基づく全順序関係がある．K2 アルゴリズムは，その全順序制約をうまく用いて，ある変数の親変数の組合せを探索する．

探索アルゴリズムを**図 4** に，アルゴリズムを以下に示す．

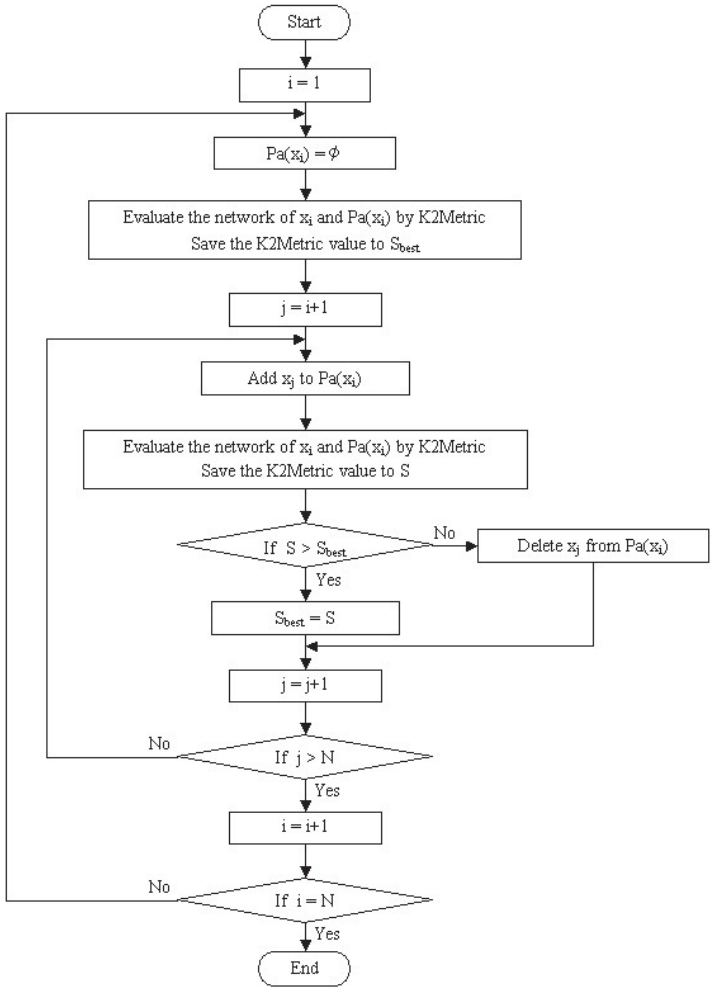


**図 4**　グラフ構造決定アルゴリズム
Fig. 4　Graph structure search algorithm.

（1）　$i = 1$

（2）　ノード $x_i$ に対する親ノード集合 $Pa(x_i)$ を空集合 $\phi$ として定義する．

(3)　$x_i$ と $Pa(x_i)$ から構成されるネットワークについて K2Metric を評価し，これを $S_{best}$ とする．

(4)　$j = i+1, \cdots, N$ について以下の操作を行う．

(a)　$x_j$ を $Pa(x_i)$ に加える．

(b)　$x_i$ と $Pa(x_i)$ から構成されるネットワークについて K2Metric を評価し，これを $S$ とする．

(c)　$S > S_{best}$ でないならば，$x_j$ を $Pa(x_i)$ から除外する．

(5)　$i = i+1$ とし，$i \leq N$ ならばステップ (2) へ戻る．

### 3.3 確率推論

ベイジアンネットワークの確率推論では，確率変数の確定値から，知りたい確率変数の事後確率を求め，これを用いて期待値などを計算する．

条件付き依存関係が成立しているネットワークにおいて，確率変数の確定値 $e$ に対する知りたい確率変数 $x_i$ の事後確率を $P(x_i|e)$ とすると，これは以下の手順で行われる[3]．

(1)　観測された変数の値 $e$ をノードにセットする．

(2)　知りたい確率変数 $x_i$ の条件付き確率 $P(x_i|e)$ を求める．

知りたい確率変数 $x_i$ の各値についての事後確率 $P(x_i|e)$ は，ありうるすべての状態で平均化する周辺化によって求める．そのために，条件付き確率表を用いる．

周辺化によれば，たとえば確率変数 $x_i$ が状態 $X^l$ をとる確率 $P(x_i = X^l|e)$ は次式で与えられる．

$$P(x_i = X^l|e) = \frac{\sum_{j=1, j\neq i}^{N} \sum_{x_j = X^1}^{X^L} P(x_1, \cdots, x_i = X^l, \cdots, x_N, e)}{\sum_{j=1}^{N} \sum_{x_j = X^1}^{X^L} P(x_1, \cdots, x_N, e)} \tag{10}$$

ここで，$\sum_{x_j = X^1}^{X^L}$ は確率変数 $x_j$ のとりうるすべての場合 $X^1, X^2, \cdots, X^L$ について総和をとることを意味する．

## 4. 提案手法

### 4.1 株価収益率の離散化

本研究では，株価指数の終値の日次データ[4] から求めた株価収益率[1] を扱う．定義式を以下に示す．

$$r_t = (\ln P_t - \ln P_{t-1}) \times 100 \tag{11}$$

ここで，$P_t$ は $t$ 期の株価終値を，$r_t$ は $t$ 期の株価収益率を示す．

ベイジアンネットワークの各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．しかし，株価収益率は連続値なので，それを何らかの方法で離散化する必要がある．各ノードは確率変数として離散化された株価収益率を持つので，3 章で述べた各ノードの状態総数は株価の離散値総数と等しい．そこで，ノードの状態総数にあたる離散化した株価収益率の離散値総数を $L$，離散値を $r^l$ とする．

離散値の集合は以下のように表される．

$$\{r^1, r^2, \cdots, r^L\} \tag{12}$$

離散化のために，本研究では等分割クラスタリングを用いた方法とウォード法を用いた方法を比較する．

クラスタリング法により離散化されたクラスタを $C_l$，その重心を $c_l$ とする．離散値 $r^l$ には，各クラスタの重心 $c_l$ をとることにする．つまり，

$$\{r^1, r^2, \cdots, r^L\} = \{c_1, c_2, \cdots, c_L\} \tag{13}$$

後述するように，本研究の解析例で得られたベイジアンネットワークは単純な構造をしている．特に，親ノードが 1 つの場合，親ノード集合のとりうる状態総数は，ノードのとりうる状態総数と等しくなる．つまり，このような場合は $M = L$ となる．

#### 4.1.1 等分割クラスタリングによる離散化

等分割クラスタリングでは，過去の株価収益率の頻度分布図において，それぞれのクラスタに含まれるイベント数がほぼ均等になるように複数のクラスタに等分割する．

#### 4.1.2 ウォード法による離散化

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．

$$D(C_i, C_j) = E(C_i \cup C_j) - E(C_i) - E(C_j) \tag{14}$$

ただし

$$E(C_i) = \sum_{z \in C_i} d(z, c_i)^2 \tag{15}$$

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 5**　株価収益率の全順序関係
Fig. 5　Total order relationship of P/E ratio.

### 4.2　株価収益率の推定

クラスタリングにより離散化された過去の株価収益率を用いて予測用のベイジアンネットワークを決定する．

本研究では K2 アルゴリズム[3] を用いているので，あらかじめ変数間の全順序関係が分かっている必要がある．株価では時系列に基づく全順序関係があるので，これに従ってネットワークを決定する（**図 5**）．

離散化された過去の株価収益率からネットワークを決定し，これを $B$ とする．これを用いて株価収益率 $r_t = r^l$ となる確率を推定した結果を $P(r^l|B)$ と定義する．$t$ 期の株価収益率として，$P(r^l|B)$ が最大となる $r^l$ を選択する．つまり，

$$r_t = \arg\max_{r^l}(P(r^l|B)) \tag{16}$$

### 4.3　提案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)　4.1 節のアルゴリズムにより，株価収益率を離散化する．
(2)　3 章のアルゴリズムにより，離散化された過去の株価収益率からベイジアンネットワーク $B$ を決定する．
(3)　$B$ を用いて式 (16) より株価収益率を予測する．

## 5. 数 値 実 験

### 5.1　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．式 (11) で変換した日次収益率を図 1 に，収益率の頻度分布を図 2 に示す．図 2 より日経平均株価の収益率分布が正規分布に比較的似ていることが分かる[5]．

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 6000 日間の日次株価収益率を使用する．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 日の株価を予測し，相関係数（CC）と平均二乗誤差（RMSE）を用いて，実際の株価に対する精度を比較する．実株価収益率を $r_t$，予測株価収益率を $r'_t$ とすると，相関係数（CC）と平均二乗誤差（RMSE）は以下のように定義される．

$$\mathrm{CC} = \frac{\sum_{t=1}^{n}(r_t - \overline{r})(r'_t - \overline{r'})}{\sqrt{\sum_{t=1}^{n}(r_t - \overline{r})^2} \cdot \sqrt{\sum_{t=1}^{n}(r'_t - \overline{r'})^2}} \tag{17}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(r_t - r'_t)^2} \tag{18}$$

ここで，$n$ はデータの組の総数を，$\bar{r}$，$\bar{r'}$ はそれぞれ $r_t$，$r'_t$ の平均値を示す．

**表 1**　日経平均株価収益率の分類区間（等分割クラスタリング）
Table 1　Classified section of P/E ratio of NIKKEI stock average (Uniform clustering).

| クラスタ | $C_l$ 区間 | データ数 | $c_l(r^l)$ |
|---|---|---|---|
| $C_1$ | [−16.14%, −1.12%) | 980 | −2.18% |
| $C_2$ | [−1.12%, −0.42%) | 980 | −0.73% |
| $C_3$ | [−0.42%, −0.00%] | 980 | −0.20% |
| $C_4$ | (+0.00%, 0.44%] | 1020 | 0.22% |
| $C_5$ | (0.44%, 1.07%] | 1020 | 0.72% |
| $C_6$ | (1.07%, 13.23%] | 1020 | 2.04% |

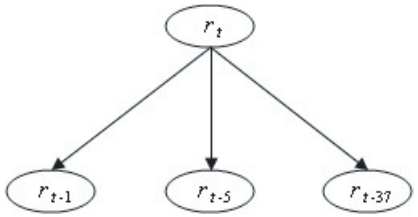


**図 6**　日経平均株価の株価収益率に対するベイジアンネットワーク（等分割クラスタリング）
Fig. 6　Bayesian network for P/E ratio of NIKKEI stock average (Uniform clustering).

#### 5.1.1　離散化法によるネットワークの比較

クラスタ数を 6 個として，4.1.1 項で説明した等分割クラスタリングで離散化してネットワークを構成する．

等分割クラスタリングによる離散値を**表 1** に示す．表中で，区間とはそのクラスタに含まれるデータ点の株価収益率の最大値と最小値を示す．データ数と平均値は，そのクラスタに含まれるデータ点総数とそれらの値の平均値を示す．データから決定されたベイジアンネッ

**表 2**　日経平均株価収益率の分類区間（ウォード法）
Table 2　Classified section of P/E ratio of NIKKEI stock average (Ward method).

| クラスタ | $C_l$ 区間 | データ数 | $c_l(r^l)$ |
|---|---|---|---|
| $C_1$ | $[-16.14\%, -3.03\%)$ | 150 | $-4.32\%$ |
| $C_2$ | $[-3.03\%, -0.85\%)$ | 1123 | $-1.59\%$ |
| $C_3$ | $[-0.85\%, -0.00\%]$ | 1667 | $-0.37\%$ |
| $C_4$ | $(+0.00\%, 1.23\%]$ | 2207 | $0.52\%$ |
| $C_5$ | $(1.23\%, 3.80\%]$ | 796 | $1.98\%$ |
| $C_6$ | $(3.80\%, 13.23\%]$ | 57 | $5.47\%$ |

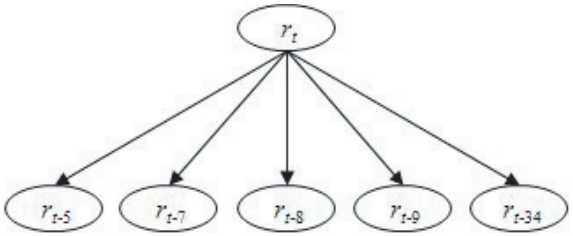


**図 7**　日経平均株価の株価収益率に対するベイジアンネットワーク（ウォード法）
Fig. 7　Bayesian network for P/E ratio of NIKKEI stock average (Ward method).

トワークを**図 6** に示す．この場合，株価 $r_t$ は 1 日前の株価 $r_{t-1}$，5 日前の株価 $r_{t-5}$，37 日前の株価 $r_{t-37}$ と関連づけられていることが分かる．

次に，クラスタ数を 6 個として，4.1.2 項で説明したウォード法で離散化してネットワークを構成する．

ウォード法で株価収益率を離散化した結果を**表 2** に示す．この離散化を用いて決定したネットワークを**図 7** に示す．この場合，株価 $r_t$ は 5 日前の株価 $r_{t-5}$，7 日前の株価 $r_{t-7}$，8 日前の株価 $r_{t-8}$，9 日前の株価 $r_{t-9}$，34 日前の株価 $r_{t-34}$ と関連づけられており，等分割クラスタリングの場合と異なっていることが分かる．

#### 5.1.2　離散化法の予測精度への影響

予測結果を実際の株価と比較した結果を**表 3** に示す．ここで，BN と BN2 は，それぞれ等分割クラスタリングによる離散化を用いた結果とウォード法によるクラスタリングを用いた結果を示す．また，AR，MA，ARMA，ARCH は，それぞれ AR モデル，MA モデル，ARMA モデル，ARCH モデルによる結果である．それぞれにおいて，1000 回シミュレーションを行い，その予測平均値を示す．比較に用いたモデルのパラメータは，2 章で述べた方法により決定した．

**表 3**　予測値と実測値の比較
Table 3　Comparison of predicted and actual stock prices.

| | 最大誤差 | 最小誤差 | CC | RMSE |
|---|---|---|---|---|
| BN | 5.5203 | 0.4172 | 0.7785 | 2.7521 |
| BN2 | 4.6893 | 0.1162 | 0.8028 | 2.3065 |
| AR(2) | 6.4808 | 0.0056 | 0.6928 | 2.7452 |
| MA(2) | 6.4808 | 0.0399 | 0.6942 | 2.7345 |
| ARMA(2,2) | 6.6313 | 0.2826 | 0.6840 | 2.7751 |
| ARCH(2,9) | 6.5119 | 0.1069 | 0.6974 | 2.7331 |

表 3 から，BN は AR，MA，ARMA，ARCH モデルより相関係数は高く，平均二乗誤差はほとんど同じであることが分かる．また，BN2 の予測結果の相関係数は BN や ARCH モデルよりも高くなっており，BN2 の平均二乗誤差は ARCH モデルなどよりも小さくなっていることが分かる．このことから，離散化法を変更することで，提案手法の精度が改善できることが分かる．

#### 5.1.3　離散値数の予測精度への影響

前項の結果より，ウォード法による離散化を用いた結果は，等分割クラスタリングによる離散化を用いた結果よりも精度が高くなることが分かった．そこで，本項では，ウォード法において離散値数を 2，4，6，8，10 として実験を行い，それらの精度を比較する．

シミュレーションに用いた離散値の一覧を**表 4** に，相関係数と最大誤差，最小誤差，平均二乗誤差の変化を**図 8** に示す．図 8 において，MaxErr，MinErr，RMSE，CC は，それぞれ最大誤差，最小誤差，平均二乗誤差，相関係数を示す．この図より，離散値数 $L = 6$ のときに相関係数は最も大きくなり，最大誤差と平均二乗誤差も最小となっていることが分かる．一方，最小誤差は離散値数にあまり依存しないことが分かる．

### 5.2　トヨタ自動車の株価への適用

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収益率の頻度分布を**図 9** に示す．この図より，分布が正規分布から外れていることが分かる．特に，図 9 から分かるように収益率 0% 付近にかなり多くのデータが分布している．

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 6000 日間の日次株価収益率を使用する．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 日の株価を予測し，相関係数（CC）と平均二乗誤差（RMSE）を用いて，実際の株価と比較する．

**表 4**　異なる離散値数による分類区間（ウォード法）
Table 4　Discrete number on different number of clusters (Ward method).

| $L$ | 2 | 4 | 6 | 8 | 10 |
|---|---|---|---|---|---|
| $c_1$ | −1.04% | −4.32% | −4.32% | −11.66% | −11.66% |
| $c_2$ | 0.99% | −0.86% | −1.59% | −4.07% | −4.07% |
| $c_3$ | – | 0.52% | −0.37% | −1.59% | −2.10% |
| $c_4$ | – | 2.21% | 0.52% | −0.38% | −1.16% |
| $c_5$ | – | – | 1.98% | 0.52% | −0.38% |
| $c_6$ | – | – | 5.47% | 1.57% | 0.52% |
| $c_7$ | – | – | – | 2.63% | 1.57% |
| $c_8$ | – | – | – | 5.47% | 2.63% |
| $c_9$ | – | – | – | – | 5.01% |
| $c_{10}$ | – | – | – | – | 11.48% |

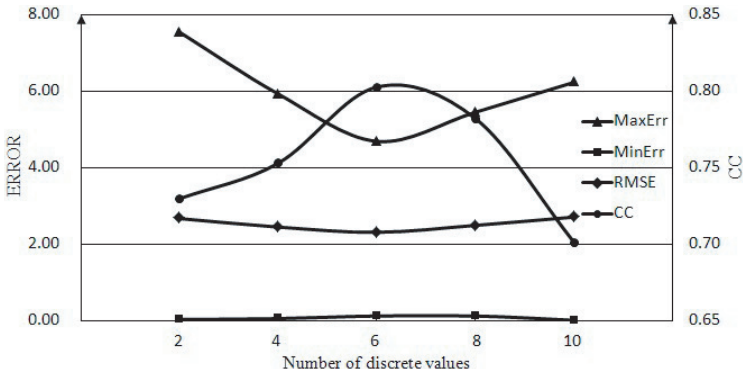


**図 8**　離散値数の精度への影響（ウォード法）
Fig. 8　Effect of number of discrete values to accuracy (Ward method).

### 5.2.1　離散化法によるネットワークの比較

先に述べたように，トヨタ自動車の株価収益率では，収益率 0% 付近にかなり多くのデータが分布している．そこで，クラスタ数を 7 個として，4.1.1 項で説明した等分割クラスタリングで離散化してネットワークを構成する．

等分割クラスタリングによる離散値を**表 5** に示す．表中で，区間とはそのクラスタに含まれるデータ点の株価収益率の最大値と最小値を示す．データ数と平均値は，そのクラスタに含まれるデータ点総数とそれらの値の平均値を示す．データから決定されたベイジアンネットワークを**図 10** に示す．この場合，株価 $r_t$ は 1 日前の株価 $r_{t-1}$ と関連づけられていることが分かる．

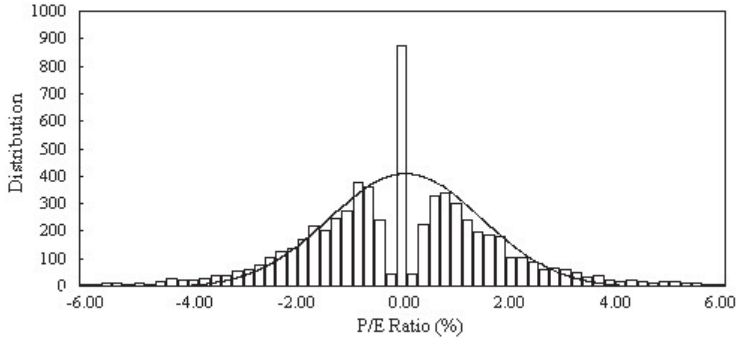


**図 9**　トヨタ自動車株価の株価収益率の頻度分布図
Fig. 9　Histgram of P/E ratio of Toyota Motor Corporation stock price.

**表 5**　トヨタ自動車株価の株価収益率の分類区間（等分割クラスタリング）
Table 5　Classified section of P/E ratio of Toyota Motor Corporation stock price (Uniform clustering).

| クラスタ | $C_l$ 区間 | データ数 | $c_l(r^l)$ |
|---|---|---|---|
| $C_1$ | [−21.13%, −1.603%) | 897 | −2.84% |
| $C_2$ | [−1.603%, −0.777%) | 897 | −1.16% |
| $C_3$ | [−0.777%, −0.00%) | 897 | −0.50% |
| $C_4$ | [+0.00%, −0.00%] | 720 | 0.00% |
| $C_5$ | (+0.00%, 0.80%] | 863 | 0.50% |
| $C_6$ | (0.80%, 1.63%] | 863 | 1.17% |
| $C_7$ | (1.63%, 16.25%] | 863 | 3.08% |

次に，クラスタ数を 7 個として，4.1.2 項で説明したウォード法で離散化してネットワークを構成する．

ウォード法で株価収益率を離散化した結果を**表 6** に示す．この離散化を用いて決定したネットワークを**図 11** に示す．この場合，株価 $r_t$ は 1 日前の株価 $r_{t-1}$，5 日前の株価 $r_{t-5}$，6 日前の株価 $r_{t-6}$，7 日前の株価 $r_{t-7}$，8 日前の株価 $r_{t-8}$，18 日前の株価 $r_{t-18}$ と関連づけられている．このように，クラスタリング手法を変更することで，得られるベイジアンネットワークはかなり異なることが分かる．


**図 10**　トヨタ自動車株価の株価収益率に対するベイジアンネットワーク（等分割クラスタリング）
Fig. 10　Bayesian network for P/E ratio of Toyota Motor Corporation stock price (Uniform clustering).

**表 6**　トヨタ自動車株価の株価収益率の分類区間（ウォード法）
Table 6　Classifying section of P/E ratio of Toyota Motor Corporation stock price (Ward method).

| クラスタ | $C_l$ 区間 | データ数 | $c_l(r^l)$ |
|---|---|---|---|
| $C_1$ | $[-21.13\%, -3.69\%)$ | 156 | $-5.31\%$ |
| $C_2$ | $[-3.69\%, -1.25\%)$ | 1062 | $-2.05\%$ |
| $C_3$ | $[-1.25\%, -0.00\%)$ | 1473 | $-0.69\%$ |
| $C_4$ | $[+0.00\%, -0.00\%]$ | 720 | 0.00% |
| $C_5$ | $(+0.00\%, 1.31\%]$ | 1456 | 0.72% |
| $C_6$ | $(1.31\%, 3.98\%]$ | 970 | 2.16% |
| $C_7$ | $(3.98\%, 16.25\%]$ | 163 | 5.85% |

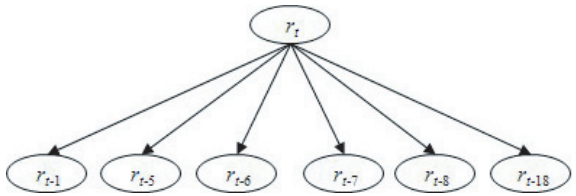


**図 11**　トヨタ自動車株価の株価収益率に対するベイジアンネットワーク（ウォード法）
Fig. 11　Bayesian network for P/E ratio of Toyota Motor Corporation stock price (Ward method).

#### 5.2.2　離散化法の予測精度への影響

予測結果を実際の株価と比較した結果を**表 7** に示す．ここで，BN と BN2 は，それぞれ等分割クラスタリングによる離散化を用いた結果とウォード法によるクラスタリングを用いた結果を示す．また，AR，MA，ARMA，ARCH は，それぞれ AR モデル，MA モデル，ARMA モデル，ARCH モデルによる結果である．それぞれにおいて，1000 回シミュレーションを行い，その予測平均値を示す．比較に用いたモデルのパラメータは，2 章で述べた方法により決定した．

**表 7**　予測値と実測値の比較
Table 7　Comparison of predicted and actual stock prices.

| | 最大誤差 | 最小誤差 | CC | RMSE |
|---|---|---|---|---|
| BN | 12.1722 | 0.1583 | 0.4849 | 5.0698 |
| BN2 | 7.1762 | 0.1808 | 0.6284 | 3.2669 |
| AR(9) | 10.3189 | 0.2119 | 0.4689 | 4.0866 |
| MA(6) | 10.2584 | 0.0242 | 0.4679 | 4.1016 |
| ARMA(9,6) | 10.6026 | 0.0539 | 0.4457 | 4.1434 |
| ARCH(9,9) | 10.2630 | 0.1632 | 0.4692 | 4.0669 |

表 7 から，BN，AR，MA，ARMA，ARCH モデルの予測結果では，相関係数が 0.5 より小さく，平均二乗誤差は 4 を超えており，日経平均株価での結果に比べるとかなり精度が低くなっていることが分かる．これに対して，BN2 では，相関係数は 0.6284 とほかの方法よりかなり大きい．また，平均二乗誤差は 3.2669 とほかの方法よりかなり小さい値を示している．BN2 の結果は日経平均株価での結果に比べると精度が低くなっているが，他の手法よりは良い結果を示している．

#### 5.2.3　離散値数の予測精度への影響

前項の結果より，ウォード法による離散化法による結果は，等分割クラスタリングを用いる離散化法による結果よりも精度が高くなることが分かった．そこで，本項では，ウォード法において離散値数（クラスタ数）を 3，5，7，9，11 として実験を行い，それらの精度を比較する．

シミュレーションに用いた離散値の一覧を**表 8** に，相関係数と最大誤差，最小誤差，平均二乗誤差の変化を**図 12** に示す．図 12 において，MaxErr，MinErr，RMSE，CC は，それぞれ最大誤差，最小誤差，平均二乗誤差，相関係数を示す．この図より，離散値数 $L = 7$ のときに相関係数（CC）は最も大きくなり，最大誤差と平均二乗誤差も最小となっていることが分かる．一方，最小誤差は離散値数にあまり依存しないことが分かる．

## 6. 結　　論

本研究では，ベイジアンネットワークを用いて株価収益率を予測する方法について述べた．提案手法では，株価収益率の分布をクラスタリング法により離散化し，その離散値を用いてベイジアンネットワークを決定する．決定したネットワークを用いて今期の株価収益率

**表 8**　異なる離散値数による分類区間（ウォード法）
Table 8　Discrete number on different number of clusters (Ward method).

| $L$ | 3 | 5 | 7 | 9 | 11 |
|---|---|---|---|---|---|
| $c_1$ | −1.51% | −5.31% | −5.31% | −8.20% | −8.20% |
| $c_2$ | 0.00% | −1.26% | −2.05% | −4.45% | −4.45% |
| $c_3$ | 1.59% | 0.00% | −0.69% | −2.05% | −2.68% |
| $c_4$ | - | 1.29% | 0.00% | −0.69% | −1.64% |
| $c_5$ | - | 5.85% | 0.72% | 0.00% | −0.69% |
| $c_6$ | - | - | 2.16% | 0.72% | 0.00% |
| $c_7$ | - | - | 5.85% | 2.16% | 0.72% |
| $c_8$ | - | - | - | 5.34% | 1.64% |
| $c_9$ | - | - | - | 10.80% | 2.74% |
| $c_{10}$ | - | - | - | - | 5.34% |
| $c_{11}$ | - | - | - | - | 10.80% |

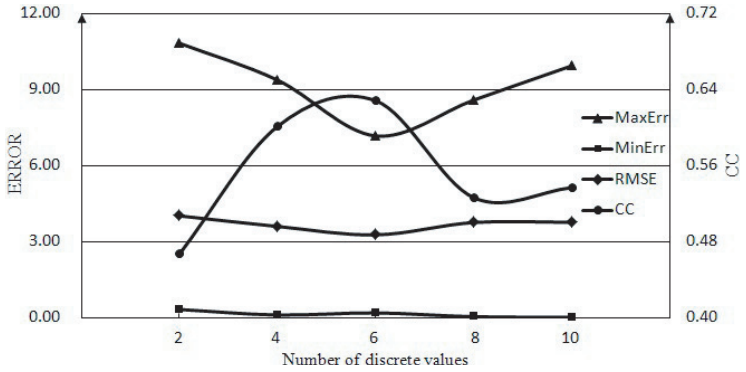


**図 12**　離散値数の精度への影響（ウォード法）
Fig. 12　Effect of number of discrete values to accuracy (Ward method).

を予測する．数値シミュレーションには日経平均株価とトヨタ自動車株価をとり，提案手法による予測結果を AR モデル，MA モデル，ARMA モデル，ARCH モデルによる結果と比較した．

最初の実験では，収益率が正規分布に近い日経平均株価を用いた．株価収益率の離散化法として等分割クラスタリングとウォード法を比較した．等分割クラスタリングを用いる方法では，AR モデル，MA モデル，ARMA モデル，ARCH モデルに比べて，提案手法は同等の精度を示し，相関係数ではより良い結果を示した．これに対して，ウォード法を用いる方法では等分割クラスタリングを用いる方法より誤差は 15% ほど小さくなり，相関係数も改善された．

次の実験では，収益率が正規分布と異なっているトヨタ自動車株価を用いた．AR モデル，MA モデル，ARMA モデル，ARCH モデルの予測精度は日経平均株価よりかなり低下した．提案手法の精度も同様に低下したが，AR モデル，MA モデル，ARMA モデル，ARCH モデルよりは高い結果を示した．以上のことから，ベイジアンネットワークを用いる予測法に一定の有効性があることを確認した．

しかし，提案手法には解決するべき問題もある．

第 1 は，株価収益率の離散化法である．株価収益率を離散化して近似する場合，離散値総数を多くすればより滑らかに分布を近似することができる．本研究では等分割クラスタリング法とウォード法によるクラスタリング法を比較し，ウォード法のほうが精度が良いことを示した．しかし，あまり細かく離散化するとベイジアンネットワークでの予測精度が低下する．そこで，精度を下げずに細かい離散化を可能とするような他のクラスタリング手法についても検討する必要がある．

第 2 に，ベイジアンネットワークで予測を行うためには，ある程度多くのデータが必要となる．そこで，より少ないデータ数で望みの精度を得られるようにアルゴリズムの改良が望まれる．

提案手法では，ベイジアンネットワークを学習するためにある程度多くのデータが必要となるのに対して，従来の時系列分析手法では，もっと少ないデータ数で予測することができる．そこで，組み合わせて使用するなど，今後両者の特長を生かした活用法についても検討を進めたい．

## 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化，三菱経済研究所 (1997).
2) Engle, R.F. and Ng, V.K.: Measuring and testing the impact of news on volatility, *Journal of Finance*, Vol.48, No.5, pp.1749–78 (1993).
3) 繁桝算男，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).
4) Yahoo!ファイナンス．http://quote.yahoo.co.jp/
5) 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).
6) Heckerman, D., Geiger, D. and Chickering, D.: Learning Bayesian networks: The combination of knowledge and statistical data, *Machine Learning*, Vol.20, pp.197–243 (1995).
7) Cooper, G.F. and E. Herskovits: A Bayesian method for the induction of proba-

bilistic networks from data, *Mach. Learn.*, Vol.9, No.4, pp.309–347 (1992).

(平成 22 年 4 月 16 日受付)
(平成 22 年 6 月 11 日再受付)
(平成 22 年 6 月 28 日採録)

**左　　毅**

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

**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，日本計算数理工学会等会員．