

## A Bayes Optimal Predicting Method of Bayesian Network with Visualization of Causal Relationship

Yusuke TAKEYAMA <sup>†1</sup>, Takashi ISHIDA <sup>†2</sup> and Masayuki GOTO <sup>†1</sup>

### Abstract

A Bayesian network is one of the useful models for pattern recognition problems and it has the features of both stochastic prediction and causal models. A Bayesian network expresses the causal relationship between variables with directed graphs. Usually, the structure of a Bayesian network is statistically estimated using a set of training data and the model selection has been applied in conventional methods when Bayesian network structures were estimated. However, it is not necessary to choose one model for the purpose of prediction. From the viewpoint of Bayesian statistics, it is well known that prediction using the mixture model on model class is Bayes optimal. In general, the mixture model that is given by a weighted sum of all models with the posterior probability on the model class is the Bayes optimal prediction. In this paper, we propose a new Bayes optimal prediction on a Bayesian network model class using the mixture model. A mixture model sometimes becomes a complex expression due to the weighted sum of all models on a model class, and it results in loss of the usefulness as a causal model. Since the easiness of interpretation is one of the merits of a Bayesian network, using a mixed model only for improvement in predictive accuracy may lead to losing the merit of a Bayesian network. Therefore, we propose a new method that is configured with the mixture model utilizing the characteristics of the Bayesian network by organizing model classes properly. Furthermore, we propose a method to quantitatively assess the strength of the causal relationship between the nodes on the mixed Bayesian network model. In addition, the effectiveness of the proposed method is clarified via a numerical experiment on an application to a prediction problem of buying and selling of shares in a stock market.

Key words: Bayesian network, mixture model, BIC, mutual information, Bayesian model averaging

---

<sup>†1</sup> Department of Industrial and Management Systems Engineering, School of Creative Science and Engineering, Waseda University

<sup>†2</sup> Media Network Center, Waseda University

Received: March 20, 2012

Accepted: March 21, 2013

# 因果関係の可視化を考慮した ベイジアンネットワークのベイズ最適な予測法

竹山 淳祐<sup>†1</sup>, 石田 崇<sup>†2</sup>, 後藤 正幸<sup>†1</sup>,

パターン認識問題において、因果モデルと確率予測モデル、双方の特徴を持つ手法であるベイジアンネットワークが注目されている。本研究では、ベイジアンネットワークモデルのクラスを対象としたベイズ最適な予測法を新たに提案する。一般に、モデルクラス内の全モデルの事後確率による混合モデルが予測に対してベイズ最適だが、一般的なモデルクラスでは混合モデルは複雑なモデルになるため、解釈容易性を持つ因果モデルとしての有用性を失ってしまう。そこで、モデルクラスをうまく構成することにより、ベイジアンネットワークの特徴を活かした混合モデルを構成し、さらにノード間の因果関係の強さを定量的に評価する方法を提案する。また、株の売買指標の予測問題に適用し提案手法の有効性を示す。

キーワード：ベイジアンネットワーク、混合モデル、BIC、相互情報量、Bayesian Model Averaging

## 1 研究背景と目的

近年、情報技術の発展に伴い、データマイニングやパターン認識の技術が注目を集めている。特に、因果モデルと確率予測モデル、それら両方の特徴を持つ手法であるベイジアンネットワーク（以下、BN）の研究が盛んである [1], [2]。BN モデルの特徴として、(1) 確率変数間の因果関係をグラフィカルなネットワークで表現することができる、(2) ある観測値データが与えられたもとで、その他の変数の条件付き確率を推論する現実的な計算アルゴリズムが与えられている、(3) 確率モデルとしての整合性を持った推論が可能である、などがある。BN モデルの長所は、主にその確率推論の因果関係の可視性と確率推論としての完全性にあり、条件付き確率によってモデル化されたネットワーク上で柔軟な推論を行うことができる。そのような長所を生かし、BN モデルは最近でも様々な応用分野への適用が研究されている [3]。例えば、上山ら [4] は BN モデルを株価動向分析に適用し、足立ら [5] は、BN モデルを用いた複合機故障診断について検討している。マーケティング分野における BN モデルの活用例も多く見られるが、例えば、早矢仕ら [6] は、ネットワーク構造変化検出と広告効果測定に BN モデルを応用している。その他、本村 [7] による BN モデルを利用したユーザ適応システムや椿ら [8] による BN モデルを用いた学生タイプ別教育効果分析など、様々な応用分野へと広がりを見せており、また、申ら [9] は、BN モデルを動的なシステムに拡張した Dynamic BN モデル

を構築し、交通渋滞量の予測確率分布を求めると共に交通信号制御へ応用する方法を提案している。以上のように、BN モデルの応用分野は現在も広がっており、経営工学分野においても、市場分析や人間工学、社会現象のモデル化などの多方面で、統計的な分析モデルとしての価値が高まっていると言える。

一方、BN モデルは学習データから統計的にモデル構造とパラメータを推定する必要がある。特に、BN モデルの有用性が認識された当初から、そのモデル構造をどのように与えるかについて様々な研究がなされてきたが、多くの場合、モデル構築方法の 1 つとして、情報量基準によって最適なモデルを選択する方法が用いられている [10], [11]。これらのモデル選択基準としては、AIC や BIC、MDL などの様々な基準 [12], [13] が考えられるが、それぞれのモデル構造を比較し、最良なモデルを選択するというプロセスが必要になる。しかし、BN モデルを未知の事象を予測する問題に適用する場合、数あるモデルの中から 1 つのモデルを選択することが最適である保証はない。特に、演绎推論だけでなく、統計的な学習に対してもベイズ理論 [14] の枠組みで予測に最適な方法を構築した場合、仮定されるモデルクラス上の全てのモデルで混合をとる方法がベイズ最適であることが知られている [12], [15]。複数のモデルの混合を取って予測モデルとする方法は、Bayesian Model Averaging (BMA) [16]-[18] として知られており、様々な研究が行われている。しかし、一部の確率モデルのクラスを除き、混合モデルは一般に元の確率モデルのクラスに属しておらず、予測モデルとしては複雑で、モデル構造の解釈が難しくなってしまうという問題がある。

一方、BN モデルを対象とした研究においては、数

<sup>†1</sup> 早稲田大学創造理工学部経営システム工学科

<sup>†2</sup> 早稲田大学メディアネットワークセンター

受付：2012 年 3 月 20 日、再受付（3 回）

受理：2013 年 3 月 21 日

あるモデルの中から情報量基準などを用いて 1 つのモデルに決定する方法がある一方で, Ammar ら [19] は, モデルクラスの中で, ある一定の割合以上のモデルが保有するリンクを残して構成した BN モデルを混合モデルと提案している. しかし, モデルの中には予測精度の高いモデルや低いモデルが混在しているが, Ammar らの混合方法は, それらを単純に平均化する方法と見なすことができ, 最適であるとはいえない. 一方で, 先に述べたように, モデルクラス内の全モデルの事後確率による混合モデルが, 予測に対するベイズ最適であることが知られているが [12], 一般的のモデルクラスに対して混合モデルを計算すると, 必要以上に複雑なモデルになり, 因果関係の解釈も困難になってしまふ. BN は変数間の因果関係が有向グラフで表現されるため, モデルの構造の解釈が容易であるという特徴を持っているが, 予測精度の向上のために複雑な混合モデルを用いることは, BN の良さを失うことにつながる. 木構造のモデルや回帰モデルでは, 混合モデルが元のモデルクラスのモデルで表現できることが示されているが [20]-[22], BN モデルについても, 同様のモデルクラスを構成し, 混合 BN モデルを 1 つの BN モデルの構造で表現することができれば, 予測精度の向上と共に, BN モデルが持つモデルの解釈容易性を保持することが可能と考えられる.

そこで本研究では, ベイズ最適な予測精度を持つつ, 混合モデルの確率構造の解釈容易性を伴った BN の混合予測モデルの構成法を提案する. 具体的には, ベイズ最適な混合 BN モデルが一つの BN モデルの構造で与えられるようなモデルクラスの設定法を提案すると共に, その混合 BN モデルの計算法について示す. これにより, 混合 BN モデルが, やはり一つの BN モデルの構造で与えられ, ノード間の因果関係を理解し易いという BN モデルの持つメリットを保持することが可能である. さらに, 得られる混合 BN モデルはモデルクラスに含まれる BN モデルのうち, 最も構造が複雑なモデルの構造となってしまうため, 確率構造の解釈容易性を目的として, 新たに相互情報量を用いて, ノード間の因果関係の強さを定量的に評価する方法を提案する. また, 評価実験として株取引の売買指標であるゴールデンクロス (GC), デッドクロス (DC) の予測問題 [23]-[26] に提案手法を適用し, その有効性を検証する.

## 2 ベイジアンネットワーク

ベイジアンネットワークとは, 複数の確率変数間の定性的な依存関係をグラフ構造によって表し, 個々の確率変数間の定量的な関係を条件付き確率で表した確

率モデルである. 対象とする現象を構成するそれぞれの事象をノードで表し, ノード間の依存関係は原因から結果となる変数の向きをもつ有向リンクで表す. ベイジアンネットワークで用いるグラフは非循環有向グラフと呼ばれ, どのようなリンクの通り方をしても一度通ったノードに戻ることがないグラフを指す.

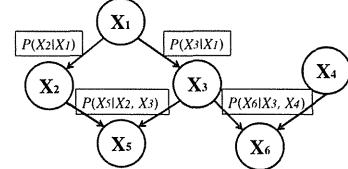


図 1 ベイジアンネットワーク例

各ノードは, ある事象が生起したかどうかを表す確率変数となっている.

### 2.1 問題設定

学習データの総数を  $N$ , 確率変数の総数が  $n$  の観測値からなるデータ  $\mathcal{D} = \{x^1, \dots, x^N\}$ ,  $x^i = (x_1^i, \dots, x_n^i)$  から, BN モデルを統計的に学習し, モデル上的一部の確率変数の観測値を入力したもとで状態が未知の確率変数に対して事後分布を出力する確率推論を行うことで予測値を得る.

2 つの確率変数  $X_p, X_q$  の依存性を BN では向きを持つリンクによって  $X_q \rightarrow X_p$  と表し,  $X_q$  を親ノード,  $X_p$  を子ノードと呼ぶ. 親ノードが複数ある場合, 子ノード  $X_p$  の親ノードの集合を  $\mathcal{P}a(X_p)$  と書くこととすると, 全変数を合わせた  $n$  次元確率変数  $\mathbf{X} = (X_1, \dots, X_n)$  の同時確率分布は (1) 式で表現される.

$$P(\mathbf{X}) = \prod_{i=1}^n P(X_i | \mathcal{P}a(X_i)) \quad (1)$$

### 2.2 モデルの統計的学习

統計的学習 [27] とは, 学習データからモデルの構造と条件付き確率を推定することである. BN モデルでは, 取り扱う確率変数によってノード数が決定されるが, これらのノード間の因果関係の有無によって BN モデルの構造が決定される. この BN モデルの構造が一つの確率モデルを表わしており, どのモデルが適しているかを学習データから統計的に推定する必要がある. 本研究では, これらの候補となるモデル集合を  $\mathcal{T} = \{T_1, T_2, \dots, T_m\}$  とする.

条件付き確率の学習方法としては, 最尤推定法とベイズ推定法があるが, 本研究では最尤推定法を用いる<sup>1</sup>.

<sup>1</sup>ベイズ推定への拡張は容易である. 二項確率の事前分布にベータ分布を仮定すれば, ラプラス推定量で置きかえることによってベイズ推定が構成できる.

1 つのモデル  $T_l (l = 1, \dots, m)$  が与えられたとき,  $N_{ijk}$  を学習データのうち  $X_i$  の親ノード集合が  $j$  番目のパターンであるもとで,  $X_i$  が  $k$  番目の値をとるデータ数と定義すると, 条件付き確率の最尤推定量  $\hat{\theta}_{ijk}^{T_l}$  は (2) 式で得られる.

$$\hat{\theta}_{ijk}^{T_l} = \frac{N_{ijk}}{\sum_k N_{ijk}} \quad (2)$$

モデルの構造は, 全てのモデル  $\mathcal{T} = \{T_1, \dots, T_m\}$  の中から情報量基準を最大にするモデルを選択する.

### 2.3 確率推論

確率推論 [28] とは, 確率変数の値が観測されたもとで, 構築したモデル構造を利用して, 目的変数の事後確率を求ることである. これは,  $X_i$  の親ノードの観測情報を  $e^+$ , 子ノードの観測情報を  $e^-$  とすると (3) 式で表される.

$$P(X_i|e^+, e^-) = \frac{P(X_i|e^+)P(e^-|X_i)}{P(e^-|e^+)} \quad (3)$$

### 3 従来研究

BN におけるモデルの混合に関する研究として, Ammar ら [19] の研究がある. Ammar らは, 真のモデルに近づけることを目的にモデルの混合を行っている. Ammar らのモデル構築方法は大きく分けて 2 段階に分かれている.

一般に, ノード数  $n$  が与えられたもとでの BN のモデル数  $m = f(n)$  は (4) 式で計算することができる.

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i),$$

$$f(0) = 1 \quad (4)$$

(4) 式より, ノード数が 5 個 ( $n=5$ ) の時点でモデル数は 29,281 通りとなり, ノード数の増加に伴いモデル数は指数的に増加することがわかる. そのため, あらゆる可能な BN モデルを考慮して, 混合モデルを構成することは計算量的に現実的ではない. Ammar ら [19] は, モデルクラスを探索的に与え, その上で混合を取る方法を与えている.

第 1 段階 [29] では, ノード間が独立か否かを判定するために,  $G^2$  検定を用いて各ノードの親ノードになりうるノードを限定し, 探索するモデル数を削減している.  $G^2$  検定の統計量は (5) 式で表され,  $E_{ijk}$  は帰無仮説の期待頻度を表している.

$$G^2 = 2 \sum_{j,k} N_{ijk} \log \frac{N_{ijk}}{E_{ijk}} \quad (5)$$

$G^2$  検定で探索空間を削減した上で, 考え得る全てのモデルでモデルクラスを構成する. このようにして構成したモデルクラスを  $\mathcal{T} = \{T_1, T_2, \dots, T_m\}$  とおく.

第 2 段階 [30] では, 得られたモデルクラス  $\mathcal{T}$  上の各モデル  $T_l$  の  $\mathcal{P}a^{T_l}(X)$  を用いて, 各リンクを無向リンクに変換した後, 各確率変数間の無向リンクを保有するモデル数をリンクの重みとしたモデルを混合モデルとして提案している. さらに, 混合する  $M$  個のモデルのうち 20 % 以下のモデルでしか存在しないリンクを削除し, 残ったリンクを評価値の高くなるように再度有向リンクに変換することで, BN の混合モデルが得られる. 混合モデルの因果関係の条件付き確率  $P_{\mathcal{T}}(X_i | \cup_{l=1}^m \mathcal{P}a^{T_l}(X_i))$  を, 各モデルの推定量の平均によって得ることにより, この混合モデルの確率構造を次式のように与えることができる.

$$P_{\mathcal{T}}(X) = \prod_{i=1}^n P_{\mathcal{T}}\left(X_i \mid \cup_{l=1}^m \mathcal{P}a^{T_l}(X_i)\right) \quad (6)$$

### 4 提案手法

#### 4.1 提案への展開

一般に, モデルクラスの中には, 予測精度の高いモデルや低いモデルが混在しているため, Ammar らの提案した混合モデルのように各モデルの重要度を考慮せずにモデルの混合をした場合, 予測精度の低いモデルの影響を受け, 逆に予測精度を下げてしまう恐れがある. ベイズ統計の理論 [12] によると, 考えうる全てのモデルを事後確率の重み付けで混合することがベイズ最適な予測となることが知られている. しかし, 一般のモデルクラスでは, 混合モデルは複雑なモデルになってしまい, 因果関係の解釈が容易ではない. 一方, ベイジアンネットワークは変数間の因果関係を有向グラフで表現しており, その解釈容易性が 1 つの特徴にもなっている. したがって, 単に予測精度の向上のために混合モデルを用いることは, ベイジアンネットワークの良さを失うことにつながる.

そこで, ベイジアンネットワークのモデルクラスをうまく構成することにより, ベイジアンネットワークの混合モデルが 1 つのモデルで表現できるような構成法を提案する. その上で, BIC を用いて漸近近似的に事後確率を求め, 混合モデルを計算し, これが 1 つの BN モデルで表現されることを示す. 予測に最適である混合モデルが, 1 つのモデルで表現できれば, 因果関係の解釈は容易である. さらに, モデルの解釈容易性を向上させるため, 相互情報量を用いて, ノード間の因果関係の強さを定量的に評価する方法を提案する.

以上により, 本研究では, ベイズ最適な予測精度を持ちつつ, 混合モデルの因果関係の解釈容易性を伴つ

たベイジアンネットワークの混合予測モデルの構成法を提案する。

#### 4.2 モデルの重要度を考慮した混合モデル

本研究では、何らかの方法に基づいてノードの先行順序関係が与えられるものとする。そのための方法としては、下記のような方法が考えられる。

- (1) 分析者が対象問題の技術的知見に基づいて因果関係を決定する方法
- (2) 全ノード間でリンクを仮定し、そのリンクの方向の組み合わせについてモデル選択により決定する方法
- (3)  $G^2$  検定によりノード間の結合の有無を検定した後、全てのリンクの方向の組み合わせについてモデル選択により決定する方法

本提案手法は、このようなノード間の先行順序関係のみが与えられたもとで、最も複雑なモデルの部分集合となるモデル全てで混合を取る方法である。提案手法の手順をまず以下に示し、その後に各 STEP の詳細を述べる。

**STEP1)** 与えられているノードの間の先行順序関係から、各ノードの親ノード候補集合を決定する。

**STEP2)** 各ノードに対し、全ての親ノード候補とリンクを結んだ、最も複雑なモデルを構成する。このモデルのリンク総数を  $k$  本とする。

**STEP3)** STEP2で得られたモデルに対し、 $k$  本のリンクのうちの一部を削除して得られる  $2^k$  通りのモデルからなるモデルクラスを構成する。

**STEP4)** モデルクラスに含まれる全てのモデルに対して BIC ((7) 式) を計算し、それに基づいて各モデルの事後確率の近似値を計算する (式 (8))。

□

STEP1～STEP3 のようにモデルクラスを構成することで、モデルを混合しても 1 つの BN で表現することができる。ここでは、表 1 のような各ノードの親ノード候補集合が得られた場合を例に挙げる。

表 1 親ノード候補例

ノード	親ノード候補
$X_1$	-
$X_2$	$X_1$
$X_3$	$X_1, X_2$

このとき、提案する混合モデルを図 2 に示す。この場合、モデルクラス中で  $T_8$  が最も複雑なモデルとなる。

STEP4において、モデルクラスに含まれる各モデルに対して、モデルの事後確率の漸近式で与えられる

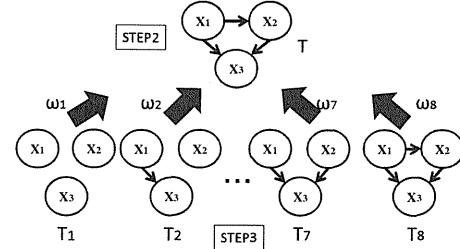


図 2 1 つのモデルで表現した BN 混合モデルの例

Bayesian Information Criterion(BIC) を用いて、モデルの重みを決定する。BN における各モデル  $T_l$  の BIC は、 $L_{T_l}$  を尤度、 $d_{T_l}$  をパラメータ数(自由度)とすると (7) 式で計算することができる。

$$\begin{aligned} BIC(T_l) &= -2\log L_{T_l} + d_{T_l} \log N \\ &= -2 \sum_{i,j,k} N_{ijk} \log \hat{\theta}_{ijk}^{T_l} + d_{T_l} \log N \end{aligned} \quad (7)$$

(7) 式で得られた BIC をもとに、モデル  $T_l$  の事後確率の近似値  $\omega_l$  は (8) 式で与えられる。

$$\omega_l = \frac{e^{-BIC(T_l)}}{\sum_{l=1}^m e^{-BIC(T_l)}} \quad (8)$$

(1) 式で得られるモデルの同時確率分布に、(8) 式で得られる各モデルの重みとの積を取り、全てのモデルの和をとることで混合モデルが得られる。混合モデルの同時確率分布は (9) 式で与えられる。

$$P_{\mathcal{T}}(\mathbf{X}) = \sum_{l=1}^m \omega_l P_{T_l}(\mathbf{X}) \quad (9)$$

ここで、 $P_{T_l}(\mathbf{X})$  は BN モデル  $T_l$  の構造もとで、(1) 式によって計算される同時確率を表わす。

提案したモデルクラス構成法では、得られたモデルクラス  $\mathcal{T}$  内の任意のモデル  $T_l, T_{l'} \in \mathcal{T}$  に対して、 $\mathcal{P}a^{T_l}(X_i) \subseteq \mathcal{P}a^{T_{l'}}(X_i)$ 、または  $\mathcal{P}a^{T_l}(X_i) \supseteq \mathcal{P}a^{T_{l'}}(X_i)$  が成り立っている。従って、 $\mathcal{P}a^{T_l}(X_i)$  に含まれない親ノードからの条件付き確率を 0 と見なせば、

$$\overline{\mathcal{P}a}^{\mathcal{T}}(X_i) = \cup_{l=1}^m \mathcal{P}a^{T_l}(X_i) \quad (10)$$

を定義することによって、

$$P_{\mathcal{T}}(X_i | \overline{\mathcal{P}a}^{\mathcal{T}}(X_i)) = \sum_{l=1}^m \omega_l P_{T_l}(X_i | \mathcal{P}a^{T_l}(X_i)) \quad (11)$$

を計算することができる。さらに、この条件付き確率を用いれば、(9) 式は、

$$P_{\mathcal{T}}(\mathbf{X}) = \prod_{i=1}^n P_{\mathcal{T}}(X_i | \overline{\mathcal{P}a}^{\mathcal{T}}(X_i)) \quad (12)$$

と書き換えることができる。これは、この条件付き確率を持つ1つのBNモデルによって混合モデルが表現できることを意味する。

#### 4.3 依存度合の可視化

提案したBNの混合モデルは、1つの混合モデルで表すことができるが、最も複雑なBNモデルの形で表現され、可能性のある全てのノード間でリンクが結ばれてしまうため、ノード間の因果関係を視覚的に解釈することができない。そこで、混合モデルの解釈容易性を向上させるため、相互情報量を用いて、ノード間の因果関係の強さを定量的に評価する方法を提案する。相互情報量とは、2変数の因果関係の強さを計量化した尺度で(13)式で表される。

$$I(X_p; X_q) = \sum_{x_p, x_q} P(x_p, x_q) \log \frac{P(x_p, x_q)}{P(x_p)P(x_q)} \quad (13)$$

相互情報量を用いることで、任意で定める閾値以上のリンクに注目し、混合モデルの因果関係の構造を把握することが容易になる。

### 5 評価実験

#### 5.1 予測対象

本研究では、日経平均株価のゴールデンクロス(GC)及びデッドクロス(DC)と呼ばれる株の売買指標[23]-[26]を予測対象とする。一般に、短期移動平均線(直前6週の株価平均)が長期移動平均線(直前13週の株価平均)を上回った点をGC、短期移動平均線が長期移動平均線を下回った点をDCと呼ぶ。GC・DCは株式市場の動向の転換点を表しており、それぞれ買い・売りの指標として用いられている。また、GC・DCは発生した時点で、既に市場の変動は緩やかになり始めているため、GC・DCの発生を前もって予測することができればより多くの利益を得ることができる。

#### 5.2 実験条件

本実験では、Yahoo ファイナンス[31]よりデータを収集した。データは、アメリカのダウ工業平均株価(ダウ)、イギリスのFinancial Times Stock Exchange 100(FTSE)、日本の日経平均株価(日経)の3指標の週データを用いた。この3指標は、取引が世界の株式市場の上位を占めているだけでなく、各地域の代表的な都市のためとりあげられることが多い。また、これらのデータは時差があるため「日経 >FTSE> ダウ」という先行順序関係をもとに、各ノードの親ノード候補集合を決定する。

実験に用いたノードは以下の10ノードである。

- (1) 前週と比較して日経株価が上昇・下降の2値
- (2) 4週前と比較して日経株価が上昇・下降の2値
- (3) 長短期の移動平均線のどちらが大きいかの2値
- (4) 長短期の移動平均線の差500以上・以下の2値
- (5) 前週と比較してダウ株価が上昇・下降の2値
- (6) 前週と比較してFTSE株価が上昇・下降の2値

上記6つの現在の値を表すノードに加えて、1~3の翌週の値を表すノードを用いる。以上の9つのノードを用いて、GC発生・DC発生・発生しないの3値を持つノードを予測する。モデルの学習期間は5年、予測期間は1年とし、これを期間を1年ずつずらして3セットの実験を行う。

予測の際は、現在の値を表すノードを観測情報として(3)式を用いて確率推論を行い、“GC発生”、“DC発生”、“発生しない”の3つの事象の事後確率を求める。その際、本研究で扱うGC・DC発生の予測問題においては、多くの平常時はこのどちらも発生しないため、必然的に“GC・DCは発生しない”の事後確率が高くなってしまう傾向がある。従って、GC発生、またはDC発生の確率が、ある閾値を超えた時に発生すると判定する<sup>2</sup>。本稿では、GC発生、またはDC発生の確率が0.4を上回った場合を発生すると判定することとした。0.4という値は予備実験によって0.1単位で確率値を変化させてもっとも良い結果を与えたものである。また、この閾値の影響については、6.2節で考察する。

また、GC・DCが発生した時点で市場の変動は緩やかになり始めているため、GC・DCの発生を前もって予測すればより多くの利益を得ることができる。本研究では、馬場ら[26]によるニューラルネットワークを用いたGC・DCの予測で用いられた正解基準と同様、GC・DCが発生すると予測してから7週以内に発生した場合を正解とする。

実験は以下の3手法で比較を行う。

- BIC基準でモデルを1つ選択する方法 (BIC)
- Ammarらの混合方法 (Ammar)
- BICを用いた提案手法 (提案手法)

各手法で2007年から2009年の3年分のデータに対して予測を行い、以下の指標により評価する。

$$\text{正解率} = \frac{\text{正解した回数}}{\text{GC(DC)と予測した回数}} \quad (14)$$

<sup>2</sup>ベイズ統計の分野では、事象Aの事前確率が非常に高いとき、別の事象Bが起こった可能性が高いことを示す観測結果が得られても、やはり事象Aの事後確率が高くなってしまうという事実がよく知られている[32]。この場合、事後確率の最大値によって事象を予測していると結果は事前確率が高い事象だけになってしまうため、問題によっては、閾値などによって検出するといった方法が必要である。

## 6 結果と考察

### 6.1 予測精度に関する結果

図3に2007年を予測年とする際に用いた提案法による混合モデルを示す。実線は現在のノード、点線は翌週もしくは目的変数のノードを表している。

また、比較手法で2007年の予測に用いられたモデルとして図4にBICにより選択されたモデル、図5にAmmarらの手法による混合モデルを示す。

2007~2009年を予測年とした場合のGCの正解率

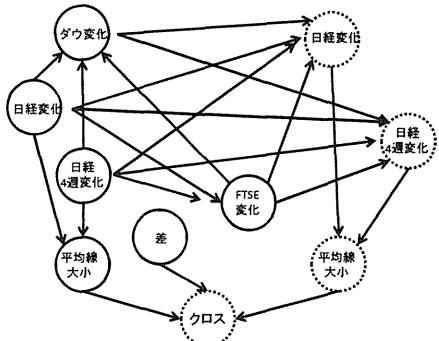


図3 提案法による混合モデル (2007年予測)

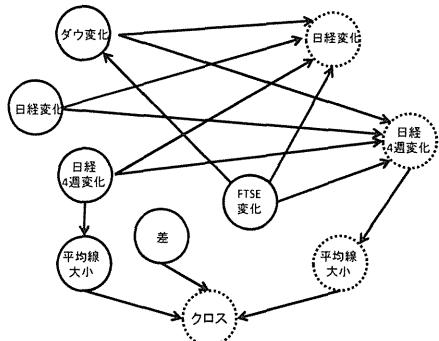


図4 BICにより選択されたモデル (2007年予測)

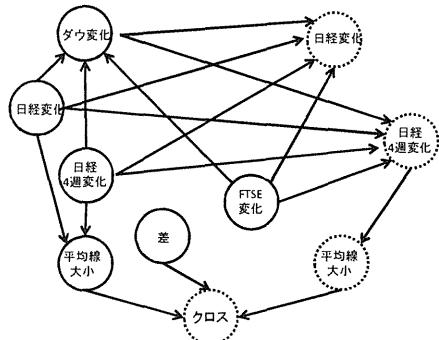


図5 Ammarらによる混合モデル (2007年予測)

を表2に、DCの正解率を表3に示す。なお、3年間に発生したGC・DCの全てを予測することができた。

表2 GCの予測実験結果

予測年	BIC	Ammar	提案手法
2007	80.0%	80.0%	87.5%
2008	40.0%	40.0%	55.6%
2009	61.1%	57.9%	65.0%
計	60.53%	58.97%	67.57%

表3 DCの予測実験結果

予測年	BIC	Ammar	提案手法
2007	83.3%	90.9%	87.5%
2008	100.0%	100.0%	100.0%
2009	80.0%	73.3%	83.3%
計	85.00%	82.76%	87.10%

この結果、BICによる漸近事後確率に基づいてモデルを混合することにより、BICでモデル選択を行う方法、Ammarらの重要度を考慮しない方法と比較して優れた予測精度を得られたことがわかる。

### 6.2 閾値の影響について

本実験では、BNモデルを用いてDC・GCの発生確率が0.4を上回った場合にDC・GCの発生を予測するものとしたが、この閾値は変化させることができる。一般に、閾値を大きくとると、DC・GCの発生を予測した際の予測の確信度は大きいため、精度は高まるが、DC・GCの発生の予測がされ難くなる。逆に閾値を低くとると、頻繁にDC・GCの発生の予測がされるようになり、精度は低下すると考えられる。この閾値の設定は、BNモデルによる予測に共通する問題であるため、従来法と提案法の双方に影響を与える。この影響について考察するため、この閾値を0.3、0.5とした場合の3年間合計の予測精度を表4、表5に示す。

表4 GCの予測実験結果

閾値	BIC	Ammar	提案手法
0.3	58.97%	57.50%	65.79%
0.4	60.53%	58.97%	67.57%
0.5	84.21%	80.00%	100.0%

表5 DCの予測実験結果

閾値	BIC	Ammar	提案手法
0.3	58.70%	60.87%	62.75%
0.4	85.00%	82.76%	87.10%
0.5	(予測なし)	(予測なし)	(予測なし)

これらの表より、閾値を0.5とするとGCの予測精度は全体的に向上するが、DCの予測では3年間で一度もDC発生予測をすることがないという結果となつ

た。また、閾値を 0.3 とした場合では、全ての手法において 0.4 の場合よりも予測精度が低下している。これは、予測確率が低い場合でも発生予測を出してしまうことによる精度劣化と考えられる。以上のことから、0.4 という閾値は、従来法と提案法の両者にとって適当な値であることが伺える。

なお、本研究では予備実験として行った正解率の評価によって閾値を0.4と設定したが、実問題に適用する際にあらかじめ閾値をどのように設定しておくかについては検討が必要である。本研究で採用した閾値は、テストデータに対する性能が最もよいという意味で採用した値であるが、実際の応用場面においては、その都度、適切な値を設定しなければならない。そのための方法としては、学習データとして得られているデータを分割し、クロスバリデーションによって適切な値を探索する方法がある。現実上の応用場面においては、十分な量の学習データが得られれば、このような方法は有用と考えられる。

その一方で、十分な量の過去のデータが得られない場合も含めて、正解率の評価に依存しない方法で事前に閾値を設定した上で予測を行う場面も想定されるが、実務上では以下のような様々な状況を考慮した上で閾値を設定して予測を行う必要があると考えられる。

- (1) GC・DC の予測問題では、株式を保有状況によって GC と DC のどちらを予測したいのかという投資家の観点を考慮する。
  - (2) 投資のリスクを取るかリスクを避けるかは投資家の態度や考え方方に帰する部分であり、この投資家の戦略を考慮する。
  - (3) 本研究の閾値は予測確率という意味付けがなされている量に対する閾値であるため、実務者の経験等によってこの“予測確率”の高低の基準をどのレベルに取るかを考慮する。

また、本研究のモデルでは、学習データに用いるデータが得られた期間と実際に予測を行う期間で統計的構造が大きく変わらないことを前提としているため、

- (4) 社会情勢や経済情勢によって不安定で非定常な株価推移になっていると判断される場合には、本モデルによる予測に加えて実務家・専門家の判断を加えて経験的な知見を加味した判断を行う、ことが考えられる。

### 6.3 混合モデルの結合強度の可視化

次に、構成された BN モデルの可視化について検討する。本稿の冒頭でも述べたように、BN の長所として因果関係の可視性が挙げられるが、混合モデルではネットワークが複雑になってしまいその解釈が困難に

なってしまう場合がある。そこで、相互情報量を用いてノード間の因果関係の強さを定量的に評価することによって、確率構造の解釈容易性を向上させることを考える。

まず、提案法の混合 BN モデルにおいてノード間の相互情報量を付与したものを図 6 に示す。このグラフから、相互情報量の値が 0.6 以上となる因果関係の強いもの、約 0.2 の値をとるもの、0.1 以下の値をとるものと、大きく 3 つのグループに分けられる。

そこで、さらにこの関係性をより強調して可視化した混合モデルを図 7 に示す。今回の実験では、上記の結果から相互情報量の閾値として 0.1 と 0.6 を用いてリンクの太さによって因果関係の強さを表すこととした。これにより、混合モデルを用いることで予測精度の向上を図りつつ、BN の長所である解釈容易性も向上できる。

## 6.4 考察

表2, 表3より、今回の実験では Ammar らの手法は BIC で最適なモデルを 1 つ選択するよりも正解率が低い結果となった。理由として、全てのモデルを同じ重要度で混合しているため、正解率の低いモデルの影響を強く受けたのが原因と考えられる。

一方、今回提案した混合方法は事後確率でモデルを混合しており、BIC でモデルを 1 つ選択した場合よりも優れた正解率が得られたことから有効性が示された。また、GC 及び DC の正解率に差があることについては、2008 年に GC 発生の直前で回避した期間があり、その期間に対する予測が全て不正解となってしまったことが原因と考えられる。2008 年を除いた GC の正解率は約 85% となるため、妥当な結果が得られたと考えられる。

さらに図7は、図3を相互情報量を用いて確率変数間の依存度合を可視化したものである。提案手法の混合モデル(図3)やAmmarらの混合モデル(図5)では、モデル選択で単一のモデルが選択される手法(図

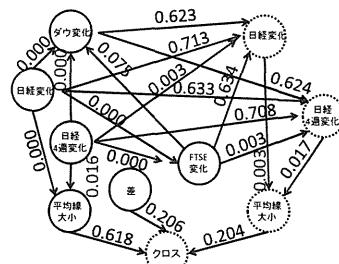


図6 相互情報量を付与した提案法の混合モデル

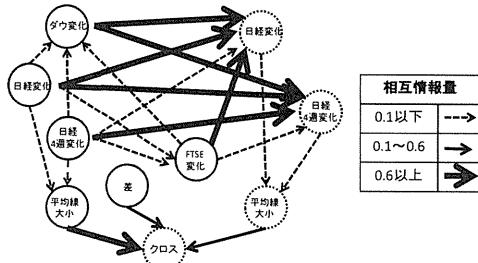


図7 ノード間の関連性を可視化した混合モデル

4) に比べてモデルのリンク構造が複雑になってしまう傾向がある。これは、提案の混合モデルによって予測精度が向上する一方で、BN がもつ因果関係を視覚的にとらえることができるという利点が失われてしまうことになる。しかし、図7 のように相互情報量を用いてグラフの可視化を行うことで、ベイズ最適な予測精度を保つつつ、ユーザが任意で定める閾値以上のリンクを注目し、混合モデルの因果関係の構造を把握することが容易になったといえる。

## 7 まとめと今後の課題

本研究では、予測を目的としたベイジアンネットワーク(BN)の統計的学習において、ベイズ最適な混合BNモデルが一つのBNモデルの構造で与えられるようなモデルクラスの設定法を提案すると共に、その混合BNモデルの計算法について示した。さらに、得られる混合BNモデルはモデルクラスに含まれるBNモデルのうち、最も構造が複雑なモデルの構造となってしまうため、確率構造の解釈容易性を目的として、新たに相互情報量を用いて、ノード間の因果関係の強さを量的に評価する方法を提案した。提案法により、BNモデルを用いてベイズ最適な予測精度の高い予測が構成できると共に、因果関係の解釈容易性という本来備えているモデルの長所を有した方法を与えることができたと考えられる。株価の予測問題への適用により、実験的にその有効性を示すことができた。

今後の課題として、今回用いたBICによる事後確率の漸近近似値ではなく厳密解での混合方法の導出や、他のデータを用いた評価実験による検証も必要である。また、DC・GCの発生確率の閾値についても今回は予備実験における正解率の評価によって値を設定したが、事前に何らかの客観的な手法によって閾値を設定する体系的な方法論についても検討する必要がある。

## 謝 辞

著者の一人竹山は、本研究を行う際に、早稲田大学創造理工学部経営システム工学科後藤研究室の鈴木健史氏に多くの時間を割いていただきました。また、本研究の内容について活発に議論していただいた細谷剛氏、三川健太氏、雲居玄道氏を始め、早稲田大学創造理工学部経営システム工学科後藤研究室の皆様に深く感謝いたします。

本研究の成果の一部は、科学研究費(23510192)の助成を受けたものである。

## 参 考 文 献

- [1] 本村陽一, 佐藤泰介: “ベイジアンネットワーク…不確定性のモデリング技術…”, 人工知能学会誌, Vol. 15, No. 4, pp. 1-8 (2000)
- [2] 繁栢算男, 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] 本村陽一, 岩崎弘利: 「ベイジアンネットワーク技術…ユーザ・顧客のモデル化と不確実性推論」, 東京電機大学出版局 (2006)
- [4] 上山 薫, 左 納, 上島康孝, 北 栄輔: “ベイジアンネットワークの株価動向分析への応用”, 人工知能学会研究会資料 SIG-FIN-004-07, pp. 40-45 (2007)
- [5] 足立康二, 山田紀一, 上床弘毅, 安川 薫: “ベイジアンネットワークによる複合機故障診断技術”, 富士ゼロックス テクニカルレポート, No. 19, pp. 78-87 (2010)
- [6] 早矢仕裕, 山西健司: “ネットワーク構造変化検出と広告効果測定への応用(機械学習とその応用)”, 電子情報通信学会技術研究報告 IBISML, 情報論的学習理論と機械学習, Vol. 111, No. 87, pp. 59-66 (2011)
- [7] 本村陽一: “ベイジアンネットワークを利用したユーザ適応システム”, オペレーションズ・リサーチ, 経営の科学, Vol. 56, No. 9, pp. 518-523 (2011)
- [8] 植美智子, 岩崎 晃: “ベイジアンネットワークを用いた学生タイプ別教育効果分析における測定精度・予測精度の検証”, 日本教育情報学会学会誌, Vol. 26, No. 4, pp. 25-36 (2011)
- [9] 申芝仙, 崔成優, 李大弘, 李義頤: “交通渋滞の確率モデルと交通信号制御”, 電気学会論文誌 C, 電子・情報・システム部門誌, Vol. 131, No. 2, pp. 303-310 (2011)
- [10] Lam, W. and Bacchus, F.: “Learning Bayesian Belief Networks: An Approach Based on the MDL Principle”, *Comput. Intell.*, Vol. 10, No. 4, pp. 269-293 (1994)
- [11] Suzuki, J.: “Learning Bayesian Belief Networks based on the Minimum Description Length Principle: Basic Properties”, *IEICE Trans. Fundam.*, Vol. E82-A, pp. 2237-2245 (1999)

- [12] 松嶋敏泰: “統計モデル選択の概要”, オペレーションズ・リサーチ, Vol. 41, No. 7, pp. 369-374 (1996)
- [13] 赤池弘次: “AIC と MDL と BIC”, オペレーションズ・リサーチ, Vol. 41, No. 7, pp. 375-378 (1996)
- [14] Bernardo, J. M. and Smith, A. F.: Bayesian Theory, John Wiley & Sons (1994)
- [15] Matsushima, T., Inazumi, H. and Hirasawa, S.: “A Class of Distortionless Codes Designed by Bayes Decision Theory”, *IEEE Trans. Inf. Theory*, Vol. 37, No. 5, pp. 1288-1293 (1991)
- [16] Claeskens, G. and Hjort, N. L.: Model Selection and Model Averaging, Cambridge University Press (2008)
- [17] Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T.: “Bayesian Model Averaging: A Tutorial”, *Stat. Sci.*, Vol. 14, No. 4, pp. 382-417 (1999)
- [18] Ando, T.: “Bayesian Model Averaging and Predictive Information Criterion for Model Selection”, *J. Jpn. Stat. Soc.*, Vol. 38, No. 2, pp. 243-257 (2008)
- [19] Ammar, S. and Lerav, P.: “Mixture of Markov Trees for Bayesian Network Structure Learning with Small Datasets in High Dimensional Space”, *Lecture Notes in Computer Science 2011*, Vol. 6717, pp. 229-238 (2011)
- [20] Willems, F. M. J., Shtarkov, Y. M. and Tjalkens, T. J.: “The Context-TreeWeighting Method: Basic Properties”, *IEEE Trans. on Inf. Theory*, Vol. 41, No. 3, pp. 653-664 (1995)
- [21] 鈴木友彦, 後藤正幸, 俵信彦: “線形回帰モデルのベイズ最適な予測法に関する研究”, 日本経営工学会論文誌, Vol. 51, No. 1, pp. 59-69 (2000)
- [22] 須子統太, 野村亮, 松嶋敏泰, 平澤茂一: “決定木モデルにおける予測アルゴリズムについて”, 電子情報通信学会技術研究報告 COMP, コンピュテーション, Vol. 103, pp. 93-98 (2003)
- [23] 小巻秦之: “景気の転換点予測モデルの有効性”, 財務省財務総合政策研究所「フィナンシャル・レビュー」, pp. 42-69 (2001)
- [24] 原田信行: “景気指標としての株価”, 日本経済学会, pp. 1-25 (2002)
- [25] 磯貝明文: “ニューラルネットワークによる株価予測: 実務上の使いやすさの観点から”, システム/制御/情報: システム制御情報学会誌, Vol. 44, No. 9, pp. 492-498 (2000)
- [26] 馬場則夫, 西田将人, 甲斐良隆: “ニューラルネットを活用した従来型テクニカル分析手法改善の試み”, 電気学会論文誌 C, Vol. 126, No. 11, pp. 1324-1331 (2006)
- [27] 磯崎隆司: “ペイジアンネットワークにおける因果発見”, 人工知能学会誌, Vol. 25, No. 6, pp. 811-818 (2010)
- [28] 本村陽一, 阪井尚樹: “確率伝播法を用いたペイジアンネットの EM アルゴリズム”, 電子情報通信学会技術研究報告 NC, ニューロコンピューティング, Vol. 105, No. 657, pp. 85-90 (2006)
- [29] Tsamardinos, I., Aliferis, C. and Statnikov, A.: “Time and Sample Efficient Discovery of Markov Blankets and Direct Causal Relations”, *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 673-678 (2003)
- [30] Ammar, S., Leray, P., Defourny, B. and Wehenkel, L.: “Probability Density Estimation by Perturbing and Combining Tree Structures Markov Networks”, *Proceedings of the 10th ECSQARU '09*, pp. 156-167 (2009)
- [31] Yahoo ファイナンス: “<http://finance.yahoo.co.jp/>”
- [32] 渡部洋: 「ペイズ統計学入門」, 福村出版 (1999)