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

1X08C082-1　竹山 湧祐
指導教員　後藤 正幸

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．特に因果関係の分析モデルと確率予測モデル，それら両方の特徴を持つベイジアンネットワーク (BN)[1] の研究が盛んである．

BN はモデル構築の際に，ノードを逐次的に追加し評価する方法と指定した空間内を全探索する方法の 2 つのアプローチ法が存在する．本研究では，全探索する方法を対象とする．

全探索法においては一般に，数あるモデルの中から情報量基準を用いて 1 つのモデルに決定する方法がある一方で，Ammar ら [2] 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ため，Ammar らの混合方法は最適であるとはいえない．一方，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．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り，因果関係の解釈が困難になる．BN は変数間の因果関係を有向グラフで表現しており，その解釈容易性が一つの特徴であるため，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，BN の良さを失うことにつながる．

そこで本研究では，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．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

BN は，確率変数間の定性的な依存関係を非循環有向グラフにより表現し，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表現した確率モデルである．

2.1　問題設定

学習データ総数 N，確率変数総数 n の観測値からなるデータ $\mathcal{D} = \{\boldsymbol{x}^1, \cdots, \boldsymbol{x}^N\}$，$\boldsymbol{x}^i = (x_1^i, \cdots, x_n^i)$ から，BN モデルを統計的に学習し，モデル上の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 $Pa(X_p)$ と書くこととすると，全確率変数 $\mathcal{X} = \{X_1, \cdots, X_n\}$ の同時確率分布は (1) 式で表現される．

$$\boldsymbol{P}(\mathcal{X}) = \prod_{i=1}^{n} P(X_i | Pa(X_i)) \tag{1}$$

2.2　モデルの統計的学習

統計的学習とは，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．条件付き確率の学習方法として，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た．1 つのモデル $T_l (l = 1, \cdots, m)$ 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．

$$\hat{\theta}_{ijk}^{T_l} = \frac{N_{ijk}}{\sum_k N_{ijk}} \tag{2}$$

モデルの構造は，全てのモデル $\mathcal{T} = \{T_1, \cdots, T_m\}$ 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めることである．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．

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

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 [2] 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．

第 1 段階では，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 (4) 式で表され，E_{ijk} は帰無仮説の期待頻度を表している．

$$G^2 = 2 \sum_{j,k} N_{ijk} \log \frac{N_{ijk}}{E_{ijk}} \tag{4}$$

第 2 段階では，各モデル T_l の $Pa(\mathcal{X})$ 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の出現回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．

$$\boldsymbol{P}_T(\mathcal{X}) = \prod_{i=1}^{n} P\left(X_i \middle| \sum_{l=1}^{M} Pa_{T_l}(X_i) \right) \tag{5}$$

4　提案手法

4.1　モデルの重要度を考慮した混合モデル

Ammar らのモデルの混合方法では，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ため，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 [3] 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，一般に混合モデルは複雑なモデルになってしまい，確率構造の解釈が容易ではない．

そこで，モデルクラスの最も複雑なモデルを 1 つ決定し，そのリンクを削除して得られる全ての BN モデルでモデルクラス $\mathcal{T}$ を構成することにより，混合モデルを 1 つの BN モデルで表現できる．

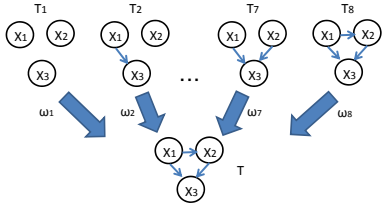

図 1．1 つのモデルで表現した BN 混合モデル

その上で，モデルの事後確率の漸近式で与えられる 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 を尤度，d を自由度，とすると (6) 式で計算することができる．

$$BIC(T_l) = 2\log L_{T_l} - d_{T_l}\log N$$
$$= 2\sum_{i,j,k} N_{ijk}\log \hat{\theta}_{ijk}^{T_l} - d_{T_l}\log N \qquad (6)$$

(6) 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 ω_l を (7) 式で与える．

$$\omega_l = \frac{e^{BIC(T_l)}}{\sum_{l=1}^{m} e^{BIC(T_l)}} \qquad (7)$$

(1) 式で得られるモデルの同時確率分布に，(7) 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 (8) 式で与えられる．

$$\boldsymbol{P}_T(\mathcal{X}) = \sum_{l=1}^{m} \omega_l \boldsymbol{P}_{T_l}(\mathcal{X}) \qquad (8)$$

4.2 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 変数の因果の強さを計量化した尺度で (9) 式で表される．

$$I(X_p; X_q) = \sum_{x_p, x_q} p(x_p, x_q)\log\frac{p(x_p, x_q)}{p(x_p)p(x_q)} \qquad (9)$$

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．

5.2 実験条件

本実験では，Yahoo ファイナンス [4] 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．実験に用いたノードは以下の 10 ノードである．

1. 日経が前週と比較して株価が上昇・下降の 2 値
2. 日経が 4 週前と比較して株価が上昇・下降の 2 値
3. 短期と長期の移動平均線のどちらが大きいかの 2 値
4. 短期と長期の移動平均線の差が 500 以上・以下の 2 値
5. ダウが前週と比較して株価が上昇・下降の 2 値
6. FTSE が前週と比較して株価が上昇・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．

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・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．

モデルの学習期間は 5 年，予測期間は 1 年とする．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．そのため，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した．

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 年までの 3 セットを予測する．

5.3 評価方法

本実験は以下の指標により評価する．

$$\text{正解率} = \frac{\text{正解した回数}}{\text{GC(DC) と予測した回数}} \qquad (10)$$

6 結果と考察

6.1 実験結果

表 1 に 2007〜2009 年までの正解率を示す．ここでは，3 年間に発生した全ての GC・DC を予測することができた．

表 1．実験結果

	BIC	Ammar	提案手法
GC	85.00%	82.76%	89.29%
DC	60.53%	58.97%	67.65%

次に，混合モデルのリンクの依存度合の一部を図 2 に示す．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，数値はノード間の相互情報量を表している．

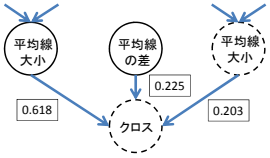

図 2．混合モデルの抜粋

6.2 考察

今回の実験では，Ammar らの手法は BIC で最適なモデルを 1 つ選択するよりも予測精度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，精度の低いモデルの影響を受けて精度が下がったと考えられる．

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 で選択されたモデルよりも優れた予測精度が得られることが示された．

また，相互情報量を計算して確率変数間の依存度合を可視化することで，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，確率構造の解釈し易いベイズ最適な BN 混合モデルを提案し，実験によりその有効性を示した．

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．

参考文献

[1] 繁枡算男, 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.

[2] S. Ammar and P. Leray,“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.

[3] 渡部洋,“ベイズ統計学入門,”福村出版,2009.

[4] Yahoo ファイナンス,“http://finance.yahoo.co.jp/”.