

潜在クラスを考慮したベイジアンネットワークによる株価予測手法の提案

1X11C043-5 小泉拓己
指導教員 後藤正幸

1 研究背景と目的

株価予測に関する研究は、単純な投資目的だけでなく、経営的な意思決定のためにも重要であり、関心を集めてきた。株価予測によく用いられる自己回帰モデルや移動平均モデルといった一般的な線形の時系列分析モデルに対して、株価変化に正規ノイズを仮定しない予測モデルとして、左らの手法 [1] がある。左らの手法では、変数間の関係を条件付き確率で表現するモデルであるベイジアンネットワークを用いて予測対象データから過去に遡って因果関係を構造学習することで、株価予測モデルを構築している。ベイジアンネットワークでは、ネットワーク構造を任意に設定することで数期先の株価を予測するなどの柔軟な構造変更が可能である。しかし左らの手法は、株価予測の際、どのような市場状況下であっても同じネットワークを用いた予測を行うため、市場状況を表現することができないという問題点があった。

この問題点の解決策として、単純な値動きを追うだけでは見えてこない隠れた市場状況を考慮することが挙げられる。そのためのアプローチとして、次の2点が考えられる。1つは、市場状況と相関の大きな変数として値動きの目安を示す指標であるテクニカル指標をベイジアンネットワークに取り入れることである。2つ目は、潜在クラスモデル (Latent Class Model; 以下 LCM) の導入である。ここでの潜在クラスとは任意の値動きのパターンの傾向であり、市場の状態を示す潜在的な離散変数である。この LCM の導入により、変数間の関係から抽出した市場状況を表現する潜在変数を用いて市場状況の差異を考慮することが可能になる。以上により本研究では、テクニカル指標・LCM を導入したベイジアンネットワークによる株価予測手法を提案する。実際の株価データを用いた実験を行い、提案手法の有効性を示す。

2 従来手法

本研究の従来手法である左らの手法は、ベイジアンネットワークを用いた株価予測手法であり、次の3ステージで成る。
 ステージ 1) 株価収益率の離散化
 ステージ 2) ベイジアンネットワーク BN の構造学習
 ステージ 3) 今期株価収益率の予測
 なお、左らの研究および本研究では、式 (1) で表される株価収益率を予測対象として用いる。

$$t \text{ 期の株価収益率 } r_t = (\ln P_t - \ln P_{t-1}) \times 100 \quad (1)$$

P_t : 第 t 期の株価の終値

2.1 株価収益率の離散化

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出来ない。そこで、連続値である株価収益率を離散値である H 個のクラスに量子化することで、離散化を行なう。クラスタリングには、等分割クラスタリング法、ウォード法の2種類の方法を用いる。

2.2 ベイジアンネットワーク BN の構造学習

式 (2) で示す $K2Metric$ をネットワークの評価値として採用し、 $K2$ アルゴリズム [1] を用いてネットワークのグラフ構造を決定する。ここでノード集合 = $\{A_1, A_2, \dots, A_i, \dots, A_j\}$ とし、 A_i の親ノード集合を \mathcal{B}_i とする。

- I : 全ノード数
- J_i : 親ノード集合 \mathcal{B}_i が取りうる離散値の組合せの総数
- K_i : ノード A_i が取りうる離散値の総数
- N_{ijk} : \mathcal{B}_i が離散値ベクトル \mathbf{b}_i^k を取り、 A_i が離散値 a_i^k を取る場合の数

$$K2Metric = \prod_{i=1}^I \prod_{j=1}^{J_i} \left\{ \frac{(K_i - 1)!}{(N_{ij} + K_i - 1)!} \prod_{k=1}^{K_i} N_{ijk}! \right\} \quad (2)$$

$$N_{ij} = \sum_{k=1}^{K_i} N_{ijk} \quad (3)$$

これを用いて、 $K2$ アルゴリズムで、株価の時系列に基づいた全順序関係を利用した、各変数の親ノードの組合せを探索していく。アルゴリズムを以下に示す。

- ステップ 1) $i = 1$
- ステップ 2) 親ノード集合 \mathcal{B}_i を空集合 \emptyset として定義する
- ステップ 3) ノード A_i と \mathcal{B}_i から構成されるネットワークについて $K2Metric$ を算出し、これを S_{best} とする
- ステップ 4) $j = i + 1, \dots, N$ について以下の操作を行う
 - (a) A_j を \mathcal{B}_i に追加する
 - (b) A_j と \mathcal{B}_i から構成されるネットワークについて $K2Metric$ を算出し、これを S とする
 - (c) $S > S_{best}$ でないならば、 A_j を \mathcal{B}_i から除外する
- ステップ 5) $i = i + 1$ とし、 $i \leq N$ ならばステップ 2 へ戻る

2.3 今期株価収益率の予測

2.2 節で構造を学習したベイジアンネットワーク BN に基づいて、各親ノードと子ノード間の条件付き確率を求める。 h 番目のクラスターの重心を r^h とし、 t 期の株価収益率 r_t^i が $r_t^i = r^h$ となる確率を $P(r^h | BN)$ と定義する。このとき、 r_t^i の予測値 \hat{r}_t^i は以下の式 (4) で求められる。

$$\hat{r}_t^i = \arg \max_{r^h} P(r^h | BN) \quad (4)$$

r^h : h 番目のクラスターの重心 ($1 \leq h \leq H$)

3 提案手法

3.1 概要と変更点

従来手法では、どのような市場状況であっても単一のネットワークによって予測するために、市場の状況変化に対応できないという問題点があった。そこで本研究では、多様な市場環境に対応した予測を行なうため、テクニカル指標と LCM を導入したベイジアンネットワークによる株価予測手法を提案する。提案手法の概要と従来手法との関係を図 1 に示す。

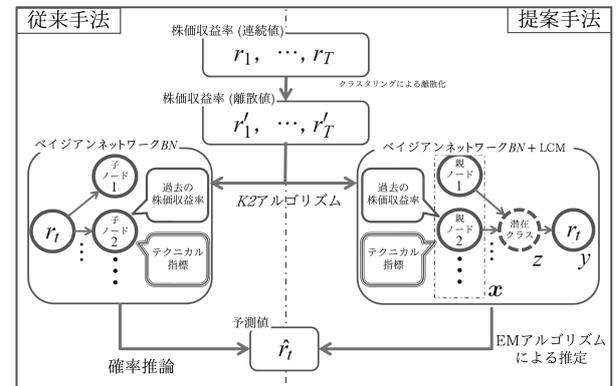


図 1. 提案手法概要

提案手法では、従来手法に次の修正を加える。まず、ステージ 2 にベイジアンネットワーク BN の構造学習方向の変更とテクニカル指標の導入を行う。具体的には、入力変数 \mathbf{u} から $K2$ アルゴリズムにより予測したいノード y の親ノード集合に組み込む説明変数 \mathbf{x} を決定し、ベイジアンネットワー

ク BN の構造を決める。この際に、3.2 節で述べるテクニカル指標を説明変数 \mathbf{x} に加える。次に、ステージ 3 に 3.3 節で述べるベイジアンネットワーク BN への LCM の導入を行う。

3.2 テクニカル指標の導入

株式市場の値動きは、自然発生的なノイズだけでなく、市場参加者の心理的側面が影響することもしばしばある。心理的側面に影響する因子の 1 つとしてテクニカル指標が挙げられる。本研究では、テクニカル指標の 1 つであり、新しいデータほど重みを大きくする、指数移動平均 EMA を用いる。 EMA は市場状況を表す指標として、株式市場参加者に一般的に用いられている。 t 期の EMA_t は式 (5) で表される。

$$EMA_t(M) = \frac{2}{M+1} \times \sum_{q=1}^{\infty} \frac{M-1}{M+1}^{q-1} P_{t-q} \quad (5)$$

M : タイムスパンパラメータ

テクニカル指標に以下の表 1 のような離散化ルールを定めることで、ネットワークに組み込むことが可能になる。

表 1. テクニカル指標の離散化ルール

条件	$EMA_t(M) < P_t$	$EMA_t(M) \geq P_t$
離散値	1	2

3.3 ベイジアンネットワーク BN への LCM の導入

株式市場では、観測されない市場状況によって値動きが決定されると仮定する。この仮定を反映するため、入力変数 \mathbf{u} から構造を決定したネットワーク BN 中の \mathbf{x} と y 間に潜在クラス z を導入する。変数 \mathbf{u} を入力したときのベイジアンネットワークの出力を $P_{BN}(\mathbf{x} | \mathbf{u})$ とすると、この確率モデルは式 (6) で表現できる。

$$P(\mathbf{x}, y | \mathbf{u}) = P_{BN}(\mathbf{x} | \mathbf{u}) \sum_z P(y | z) P(z | \mathbf{x}) \quad (6)$$

LCM のパラメータ $P(z | \mathbf{x}, y)$, $P(z)$, $P(\mathbf{x} | z)$, $P(y | z)$ は、EM アルゴリズムにより推定する。ここで、求めたい予測値 \hat{r}_t は式 (7) で求められる。

$$\hat{r}_t = \arg \max_y \sum_z P(y | z) \frac{P(\mathbf{x} | z) P(z)}{P(\mathbf{x})} P_{BN}(\mathbf{x} | \mathbf{u}) \quad (7)$$

4 実験と結果

提案手法の有効性を示すため、実際の株価データを用いた実験を行った。

4.1 実験条件

実験では、日経平均株価 5879 日分の株価収益率 (1991 年 1 月 4 日～2014 年 11 月 25 日) を用いて、現時点までの変数から次期の株価収益率を予測した。このうち、最新 220 日分 (22 日分 \times 10 セット) を評価用のテストデータとした。

株価収益率は 6 つのクラスに量子化することで離散化し (離散値数 6)、提案手法では潜在クラス数を 6 に設定した。左らの手法を従来手法とする。また、 EMA で日毎の変化をみる場合において、一般的に短期間の変動を見るために $M = 12$ 日、中長期間で $M = 26$ 日が使われることが多いことから、従来手法・提案手法に $M = 12$ 日、 $M = 26$ 日の EMA を表 1 の離散化ルールによって組み込んだ。

提案の効果を詳細に検討するため、テクニカル手法のみ導入したものを比較手法 1、LCM のみ導入したものを比較手法 2、どちらも導入したものを提案手法として実験を行った。

4.2 評価方法

予測性能の評価指標として一般的に用いられる相関係数 (CC) と平均二乗誤差 (RMSE) を用いる。 t 期の CC および RMSE は式 (8), (9) で表される。

$$CC = \frac{\sum_{t=1}^n (r_t - \bar{r})(r'_t - \bar{r}')}{\sqrt{\sum_{t=1}^n (r_t - \bar{r})^2} \sqrt{\sum_{t=1}^n (r'_t - \bar{r}')^2}} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (r_t - r'_t)^2} \quad (9)$$

各評価値はテストデータ 10 セットに対する結果の平均を用いる。また、比較手法 2 と提案手法では、初期値に依存して局所最適となることを考慮し、10 回の繰り返し実験を行い、テストデータ 10 セット \times 10 回 = 100 回の平均を評価値として与える。

4.3 結果と考察

実験結果を図 2 に示す。

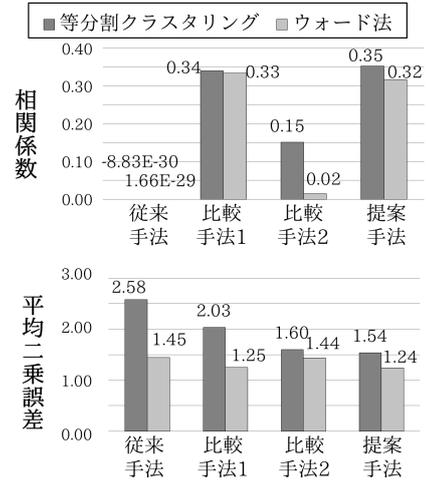


図 2. 実験結果の比較

各手法の比較によって、ベイジアンネットワークにテクニカル指標を組み込むことと、潜在クラスを組み込むことの両方により精度が向上することが検証された。このことは前者で市場状況と相関のある変数を取り入れ、後者で変数間の関係を少数の潜在変数へ縮約することによって、従来手法では考慮できなかった潜在的な市場状況を反映できたためと考察できる。一例として 2014 年 8 月 8 日～21 日の日経平均が 9 日続伸した場面では、従来手法 (ウォード法) で RMSE が 2.01 であったのに対し、提案手法 (ウォード法) で 1.19 であり、市場状況を良く反映し精度高く予測できている。

また比較手法 1 のウォード法に関しては、離散値 1 と 6 に所属するような極端な値動き自体が、市場状況に関する情報を潜在的に含んでいることが考えられる。テクニカル指標との組合せにより、提案手法の潜在クラスモデルと同等に潜在的な市場状況の反映がなされたと考えられる。

5 まとめと今後の課題

本研究では、多様な市場環境に対応した予測手法として、テクニカル指標と LCM を導入したベイジアンネットワークによる株価予測法を提案した。また、実データでの実験により、提案手法の有効性を検証した。

今後の課題として、他の銘柄への適用や適切な学習期間設定方法・テクニカル指標の組み込み基準の確立が挙げられる。

参考文献

- [1] 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, Vol.3, No.3, pp.80-90, 2010.
- [2] Thomas Hofmann: “Probabilistic Latent Semantic Indexing”, Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval (SIGIR-99), 1999.
- [3] Gregory F. Cooper, Edward Herskovits: “A Bayesian methods for the induction of probabilistic networks from data”, Machine Learning, Vol.9, pp.209-347, 1992.