

ニューラルネットワークによる次期の株価予測 — 株価予測におけるフィルタリングによる特徴量抽出 —

高穂 洋 荒井 隆行 大竹 敢 田中 衛

上智大学理工学部電気電子工学研究科 〒102-8554 東京都千代田区紀尾井町 7-1

E-mail: koguma@mamoru.ee.sophia.ac.jp,

あらまし 株価の時系列データを用いて次期の株価を予測することは困難である。つまり、一因に株価の時系列データに有益でない成分が混在していることがあげられる。先行研究において、ローパスフィルタを用いて株価の時系列データから長周期成分を除去し、ニューラルネットワークにより予測を行ったところ、比較的正確な予測結果が得られた。本論文ではさらに日数の異なる移動平均線の比を求め、その周波数特性を検討し、バンドパス処理により予測に必要と考えられる特徴量の抽出を行った。また、提案方式による実験において従来の手法と比較し、より優れた予測性能が得られた。

キーワード 時系列データ, ニューラルネットワーク, 予測, 特徴量

Prediction of the Next Stock Price using Neural Network — extraction the feature to predict next stock price by filtering —

Hiroshi TAKAHO Takayuki ĀRAI Tsuyoshi OTAKE and Mamoru TANAKA

Department of Electronics and Electrical Engineering, Sophia University 7-1 Kioi-cho, Chiyoda-ku, Tokyo,

102-8554, Japan

E-mail: koguma@mamoru.ee.sophia.ac.jp,

Abstract - Adding the dimension of time to databases produces time series databases (TSDB) and introduces new aspects and difficulties for data mining, knowledge discovery and prediction of sample points. In this paper, we introduce the method for the prediction of the next sample point with multi-layer neural network in TSDB. Predicting the next sample point in TSDB includes cleaning and filtering the time series data, identifying the most important predicting attributes, and extracting a set of association rules that can be used to predict the time series behavior in the future. Our method is based on signal processing techniques, and TSDB for the closing price of a stock are used as an example.

Keyword Time Series Database, Neural Network, Prediction, Feature

1. はじめに

時系列データとは、時間の経過とともに得られるデータのことである。本論文では株価の時系列データを利用し、ニューラルネットワーク[1]を用いて次期の株価（正確には天井度[2]）を予測する手法について提案する。また、特徴量の抽出技術についても言及する[3]。

次期の株価を予測することは、非常に難しいことである。一般には予測モデルには線形モデルなどが用いられるが、良い予測結果を得ることは困難である。我々は、株価を予測するということが非線形問題である[4]と考えており、そこで本論文では予測器として階層型ニューラルネットワークを用いている。

株価の時系列データには様々な成分が存在する[5]。図1の日経平均先物株価の時系列データに示されるように、短期間で見たとき日々目まぐるしく変化している成分である短周期成分や長期間で見たとき緩やかに変動する長周期成分が存在する。しかし、それらの成分

が混在した時系列データから次期の株価を予測することは困難であるため、それらの成分を分離することが要求される。一般的な手法として、 L 日移動平均フィルタ（ L 日移動平均線）などが良く使用される。株価の終値の時系列データ x を L 日移動平均フィルタを通過させることにより、ある n の終値 $x[n]$ から、 $x[n-L+1]$ までの移動平均値 $\hat{x}[n]$ を求める。それにより低周期成分を除去することができ、目的に応じた予測ができる[6]。

先行研究において、3日移動平均フィルタを用いた場合に最も良い予測結果が得られた。つまり、前処理を施した時系列データの周波数特性と予測対象となる天井度の周波数特性において、特定の周波数成分に起因したものと考えられる。従って、本論文では時系列データにバンドパス処理を行い、特定の周波数成分を抽出し、天井度の予測を行っている。

2. LPF によるフィルタリング処理

株価を予測するためには、短周期成分が必要となる。短周期成分は、元の時系列データを低域通過フィルタ (LPF) で処理した時系列データで割ることにより求められる。一般的には、株価予測で使用される低域通過フィルタは、移動平均フィルタである。移動平均フィルタを通過させた時系列データは、移動平均線と等価であり、次式で表される。

$$y_L[n] = \frac{1}{L} \sum_{k=0}^{L-1} x[n-k] \quad (1)$$

ここで、 $y_L[n]$ は出力信号、 $x[n]$ は入力信号、 L は移動平均の日数を示している。また、 L 日移動平均フィルタのインパルス応答は、次式で得られる [7-8]。

$$h[n] = \frac{1}{L} \sum_{k=0}^{L-1} \delta[n-k] \quad (2)$$

ここで時系列データを (1) 式で割ることによって、時系列データを正規化する。図 2 は、図 1 における株価の時系列データを離散フーリエ変換したのものであり、特定の周波数成分が多く存在していることがわかる。

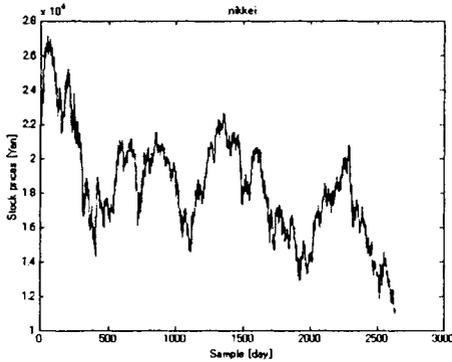


図 1: 日経平均先物の時系列データ

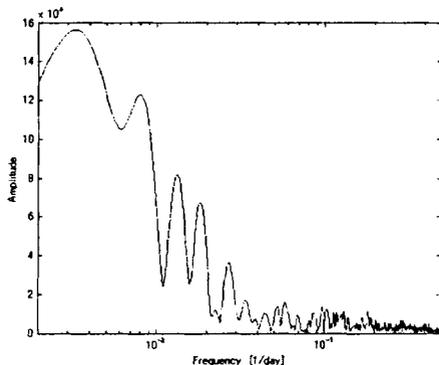


図 2: 日経平均先物の時系列データの周波数特性

3. エネルギー比による BP 処理

過去の研究では、3 日移動平均フィルタに関して最も良い予測結果が得られている。図 3 の予測対象となる天井度の周波数特性に関し、約 80 日周期成分にピークがある。それに対し、図 4, 5, 6, 7 の周波数特性を比較したとき、図 5 は約 80 日周期成分やそれよりも短周期成分において、均等な周波数特性である。そこで我々はそれらの成分が予測に有益であると考えバンドパスフィルタ (BPF) により抽出を行う。本論文では、エネルギー比を用いた BP 処理を提案する。BP 処理は日数の異なる移動平均線どうしを割ることにより行なっている。

提案手法では以下の手順で L_1, L_2 を決定する。まず、次式で示すように日数の異なる移動平均線の比 $y_p[n]$ を算出する。ここで、 y_{L_1} は L_1 日移動平均フィルタ、 y_{L_2} は L_2 日移動平均フィルタを示し、また $L_1 < L_2$ である。

$$y_p[n] = \frac{y_{L_1}[n]}{y_{L_2}[n]}, \quad (3)$$

次に、目的とする周波数成分と全エネルギーとの比を求める。

$$E[L_1, L_2] = \frac{Y_{L_1 L_2}[\omega_1]}{\sum_{\omega=0}^{\pi} Y_{L_1 L_2}[\omega]} \quad (4)$$

ここで $Y[\omega]$ は、 $y_p[n]$ の離散フーリエ変換である。式 (4) の $E[L_1, L_2]$ を使用し、 L_1, L_2 は次式により求められる。

$$L_1 \text{ and } L_2 = \arg \max E[L_1, L_2] \quad (5)$$

この処理の結果を表 1 に示す。

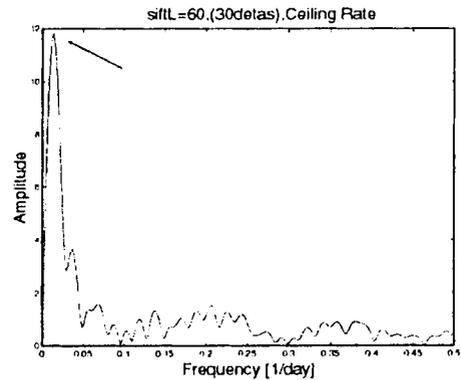


図 3: 天井度の時系列データにおける周波数特性

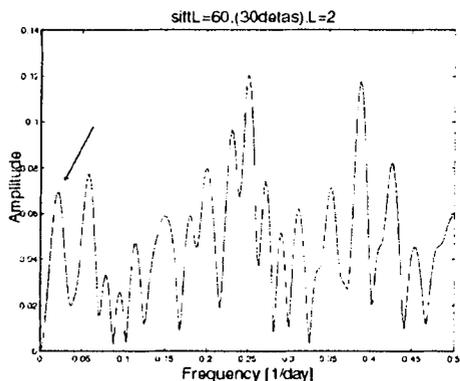


図 4: 2 日移動平均フィルタにより処理を施した時系列データの周波数特性

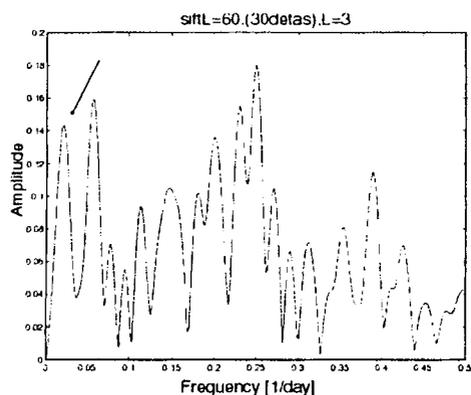


図 5: 3 日移動平均フィルタにより処理を施した時系列データの周波数特性

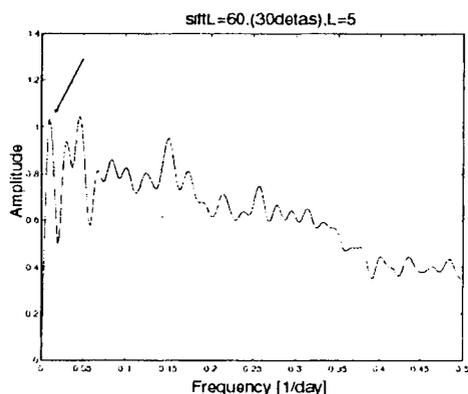


図 6: 5 日移動平均フィルタにより処理を施した時系列データの周波数特性

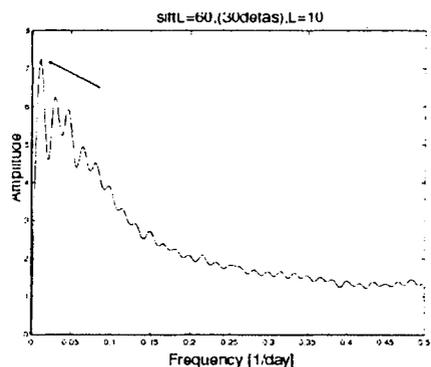


図 7: 10 日移動平均フィルタにより処理を施した時系列データの周波数特性

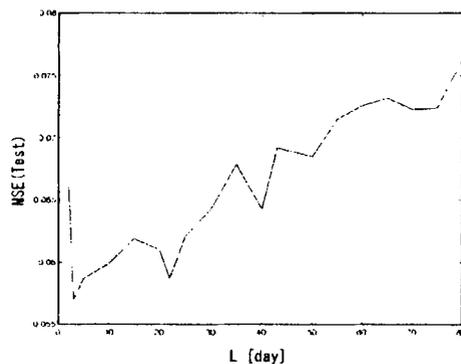


図 8: 移動平均フィルタの日数と予測時における平均二乗誤差との特性

表 1: BP 処理により選択されたの L_1, L_2

Frequency [1/day]	0.017	0.024	0.04	0.051	0.057
L_1	9	10	14	33	72
L_2	11	13	18	42	80

4. 予測実験

この章では 2, 3 章で前処理を行った時系列データに関して予測実験を行う。予測モデルとして階層型ニューラルネットワークを使用している。3 章で示した BP 処理により抽出された特徴量の有用性を確認するため、以下に示す 3 つの入力パターンでシミュレーションを行う。

1. 入力列 60 個 (59 日前までの計 60 個) に LPF のみを適用した時系列データ。
2. 入力列 55 個 (54 日前までの計 55 個) に LPF を適用した時系列データ及び入力列 5 個に BPF を適用した時系列データ。
3. 入力列 55 個 (54 日前までの計 55 個) に LPF を適用した時系列データ及び入力列 5 個に白色雑音を適用した時系列データ。

表 2 は、それぞれの入力パターンについて、20 回のシミュレーションを行った結果である。ここで予測性能は、実際のデータと予測結果との平均 2 乗誤差 (MSE) で示される。表 2 より、LPF を適用した時系列データ及び BPF を適用した時系列データを用いた時、平均 2 乗誤差が最小であった。

表 2: シミュレーション結果

input pattern	MSE(mean square error)
1.LPF only	0.0532
2.both LPF and BPF	0.0523
3. both LPF and WN	0.0571

5. 結論

予測実験より、LPF のみで処理を行った時系列データと比較して、BPF 処理で得た時系列データを加えて予測を行った場合、良い結果が得られた。従って、BPF で分離した特徴量は株価を予測するのに有益であるということが確認された。提案手法では移動平均フィルタを用いて成分の抽出を行ったが、より効果的に成分を抽出できるような BPF を作成することにより、予測精度の向上が期待される。

本研究において、ニューラルネットワークを用い予測モデルを作成したが、次期の株価を予測するためには最適であるとは断言できない。また、ニューラルネットワークを用いた場合、入力である特徴量と目的対象とのルール抽出が困難である。従って、非線形処理が可能であり、同時にルールの抽出が可能なモデルの作成を検討中である [9]。

文 献

- [1] 月本洋, "実践データマイニング", オーム社.
- [2] 滝沢孝安, 後藤康徳, 林知之 "絶対のパソコン投資術", "同友館
- [3] H.Takaho, T.Arai, T.Otake, and M.Tanaka, "Prediction of the Next Stock Price using Neural Network," in Proc. of NOLTA'02, Xi'an, China, 2002.10.
- [4] M. Last, Y. Kelin, and A. Kandel, and . Igcta, "Knowledge discovery in time series databases," Transactions on System, Man and Cybernetics, IEEE, vol.31 Issue.1
- [5] 廣松毅, 浪花貞夫, "経済時系列分析と実際," 多賀出版
- [6] C.W.Omlin and C.L.Giles, "Extraction of Rules from numerical data," in Proc. Of Knowledge-based Systems, vol.10, No.7, pp.413-419
- [7] L.Rabiner and C.Rader, "Digital Signal Processing", IEEE press, 1972
- [8] 吉野邦生, 荒井隆行, "デジタル信号処理と超関数", 海文堂
- [9] H. Tong, "Nonlinear Time Series: A Dynamical