

5L-10

遺伝的プログラミングを用いた 株価指数予測*

佐々木 崇† 伊庭 斉志† 石塚 満†

東京大学工学部電子情報工学科‡

1 はじめに

本研究では遺伝的プログラミング（GP）を用いて実時系列データ予測を試みる。時系列予測問題へのGPの適用としては太陽の黒点の数の平均を予測する[Ang96]等の研究がなされている。しかしながら、実時系列データに対する適用例はまだ少ない。その理由としては実時系列データは様々な要因によって複雑に変動するので一般に予測が難しいということが考えられる。そこで本稿では、株価指数を用いて実時系列データ予測におけるGPの有用性を実証する。株価指数とは株式市場全体の株価動向を把握するための指数である。東京証券取引所の第一部上場銘柄のうち市場を代表する225銘柄を対象に平均をとって算出される日経225や第一部上場全銘柄の現在の時価総額を基準時価総額（1968年1月4日の時価総額）で割ったものを100倍して算出されるTOPIX等がその代表的なものである。本研究では実時系列データとして日経225を取り扱う。

2 実験方法

過去の日経225のデータをもとに未来の日経225の値を予測する近似関数を見つけ出すことを試みる。そこで実験方法としてはまず、1993年の上半期における日経225の1分毎のデータを0と1の間の数値をとるように正規化した後、訓練データ（前半の3000個）とテストデータとに分ける。近似関数は訓練データ内での適合度が良くなる方へと進化していく。そして訓練データ内で最も適合度の良い近似関数がテストデータで評価される。

2.1 近似関数の評価方法

訓練データにおける近似関数の評価方法は実際の値と近似関数から得られる値との平均2乗誤差を用いた。この値はそのまま適合度として扱われる。従って、適合度は0に近いほど良いと評価される。テストデータにおける近似関数の評価方法として株売買シミュレーションを作成した。この株売買シミュレーションは100万円を元手に近似関数から得られる予測値に従って株の売買をし、日経225の増減の予測的中率とどのくらいの利益が得られるかを表示するというものである。

2.2 実行条件

株価指数予測を行なうのにA～Gの6通りの条件のもとでGPを実行した。

- 条件A：終端記号としては $\{y_1, \dots, y_{10}, \text{R}\}$ を用いた。ここで y_i は i 分だけ前の日経225の値を表している。つまり、時刻 t 分における日経225の値を $x(t)$ とすると $y_i = x(t-i)$ である。また R はランダムに定数値を発生させる終端記号である。
- 条件B：終端記号としては $\{\text{ave}_1, \dots, \text{ave}_{10}, \text{R}\}$ を用いた。 ave_i は日経225の10分毎の平均値を表しており、次式で計算される。

$$\text{ave}_i = \frac{\sum_{k=1}^{10} x(t-10*(i-1)-k)}{10}$$

- 条件C：終端記号としては $\{m_1, \dots, m_{10}, \text{R}\}$ を用いた。ここで m_i は日経225の10分毎の分散を表しており、次式で計算される。

$$m_i = \frac{\sum_{k=1}^{10} (x(t-10*(i-1)-k) - \text{ave}_i)^2}{10}$$

- 条件D：終端記号として ave_i と m_i と R を用いた。
- 条件E：終端記号として y_i と ave_i と m_i と R を用いた。

* Predicting Stock Price Index with Genetic Programming

† Takashi Sasaki, Hitoshi Iba, Mitsuru Ishizuka

‡ University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

e-mail: sasaki@miv.t.u-tokyo.ac.jp

- 条件 F：終端記号として $\{d1, \dots, d10, \text{㊄}\}$ を用いた。ここで $d_i = x(t-i) - x(t-i-1)$ である。
- 条件 G：終端記号として $\{v1, \dots, v10, r1, \dots, r10, \text{㊄}\}$ を用いた。ここで v_i は

$$v_i = |x(t-i) - x(t-i-1)|$$

で計算される値であり、 r_i は

$$r_i = \frac{x(t-i) - x(t-i-1)}{x(t-i-1)}$$

で計算される値である。どの条件においても非終端記号としては四則演算である $\{+, -, *, \%\}$ を用いており、選択方式はトーナメント方式を用いている。A から E に対しては予測するのは現在の値である。F と G に対しては予測するのは現在の値と 1 分前の値の差である。集団数 1000、世代数 100 で GP を適用した。

3 実行結果と考察

各条件において GP を複数回実行した。その実行結果を表 1 に示す。実行結果を見ると条件 B,C,D,E では予測の平均的中率と最良的中率の両方とも 50 パーセント以下であり、利益も条件 C の場合を除いてマイナスの値となっている。条件 A,F,G では平均的中率は 50 パーセント以上、最良的中率は 60 パーセント以上であり、利益もプラスになっている。特に条件 G については平均的中率も 60 パーセント以上である。この結果から以下のようなことがわかる。

1. 日経 225 の増減の予測には過去の値の平均値や分散はあまり有効ではない。
2. 単純に過去の値や差を用いる予測は安定性に欠ける。

最も GP が効果的であった条件 G における正規化された日経 225(normalized nikkei225) と差の予測値(prediction of difference) の関係のグラフを図 1 に、予測から得られる利益(profit)と最適利益(optprofit)の関係のグラフを図 2 に示す。図 1 のグラフでは見易いように差の予測値を 2 倍している。このグラフからは日経 225 の増減に対応して差の予測値が正負の値をとっていることが読みとれる。従って、この予測に基づいて売買をすると利益があがることになる。また、図 2 のグラフの最適利益は予測が 100 パーセントの中した場合の利益である。

4 おわりに

本研究では日経 225 予測における GP の有用性を株売買シミュレーションを用いて実証した。現在、他

条件	的中率 (%)		利益 (円)	
	平均値	最良値	平均値	最良値
A	53.11	62.49	7548.56	30083.00
B	47.28	47.97	-3745.85	-2516.68
C	49.22	49.96	3256.91	3567.25
D	42.82	48.31	-16039.87	-5644.75
E	37.55	37.97	-27780.45	-27284.50
F	53.72	61.98	11373.54	30314.06
G	61.90	62.27	29734.02	30423.50

表 1: 実行結果

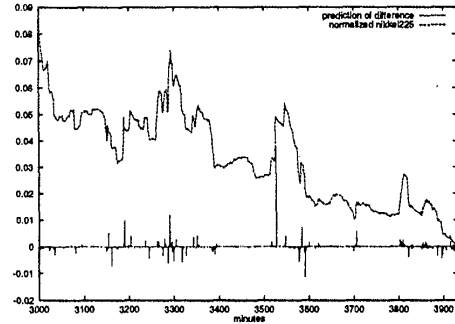


図 1: 正規化された日経 225 と差の予測

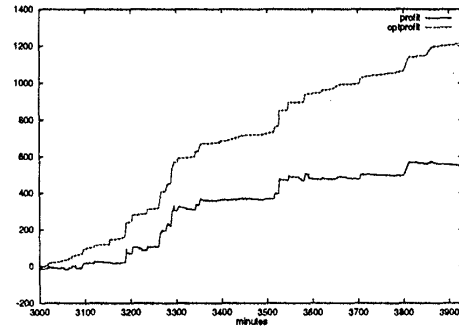


図 2: 予測から得られる利益と最適利益

の予測システムとの比較を行い、GP の有用性をさらにわかりやすい形で実証することを研究中である。また、終端記号として日経 225 に影響を与える様々な指数を取り入れ利益にどのような影響がでるか調べる研究も行なう予定である。

謝辞 日経 225 等の様々なデータの提供、および有益なアドバイスをくださった投資工学研究所の関係各位に深く感謝致します。

参考文献

[Ang96] P. Angeline. Two Self-Adaptive Crossover Operators for Genetic Programming. In P. Angeline and K. Kinnear, editors, *Advances in Genetic Programming 2*. MIT Press, 1996.

[伊庭 96] 伊庭齊志. 遺伝的プログラミング. 東京電機大学出版局, 1996.