

## 進化論的計算手法による金融データの予測

大野勝彦<sup>†</sup> 伊庭斉志<sup>‡</sup>

東京大学工学部電子情報工学科<sup>†</sup> 東京大学大学院新領域創成科学研究科基盤情報学専攻<sup>‡</sup>

### 1 はじめに

インターネット上での証券取引が盛んになり多くの個人投資家が生まれている。そのため株価などの金融データのシステム同定はより重要になってきている。システム同定にはさまざまな手法が使われている。例えば GP (遺伝的プログラミング) を用いたもの [1]、ニューラルネットワークを用いたもの [2]、PSO を用いたもの [3]、などがあげられる。この研究では GP とその拡張である STROGANOFF を金融データのシステム同定問題に応用する。

### 2 GP の拡張

GP の探索に多重回帰分析の統計的手法 (GMDH [4]) を追加することで、より高度な探索を目指す。このために、GMDH と GP を統合したシステムとして STROGANOFF が提案されている [5]。STROGANOFF では、適合度の計算には MDL 基準を用いる。MDL は木の記述長 (複雑さ) と近似の誤差のバランスを考慮するものであり、次のように定義される [6]。

$$MDL = 0.5N \log S_N^2 + 0.5kW \log N$$

MDL は最小値を最適値とする。各文字をそれぞれ説明すると、N はデータの数、 $S_N^2$  は最小二乗誤差、k は木のパラメータ (NODE に含まれる式の情報) の数で、W (WEIGHT) は木の大きさによるペナルティを調節する重み係数である。

### 3 株価の予測

#### 3.1 GP の実行

GP を用いて株価を算出する近似関数を探索し、売買を行う。

実験に用いるデータと条件は次のようにする。

- ・実験には日経 225 の一分毎のデータ (合計 33,177 分) を使う
- ・最初の 3,000 個を訓練データ、残りの 30,177 個をテストデータに使う。
- ・元手を 100 万円として、手数料は考えない。

株売買の方法は次のようにする。

1. 時刻  $t=3000$  から始める (テストデータ)。株保有フラグ  $ST=0$  にする。
2. 訓練データで得られた木構造を用いて予測値を生成する。
3. 一分後に上がると予想し、 $ST=0$  で、持ち金が株価より高ければ株を買い、 $ST=1$  とする。株を買えない場合は何もしない。
4. 下がると予想し、 $ST=1$  ならば株を売り、 $ST=0$  にし、持ち金に株価を加える。株を所有していない場合は何もしない。
5.  $t < 33177$  なら  $t=t+1$  として 2. に戻る。さもなければ利益を返して終了する。

終端記号は次の 5 通りの場合を考える。

A:  $y_1, \dots, y_{10}, R$

B:  $ave1, \dots, ave10, R$

C:  $m1, \dots, m10, R$

D:  $ave1, \dots, ave10, m1, \dots, m10, R$

E:  $v1, \dots, v10, r1, \dots, r10, R$

終端記号をそれぞれ説明する。ここで  $x(t)$  は時刻  $t$  における株価である。

$y_i$  は  $i$  分前の値 ( $=x(t-i)$ )、 $ave_i$ 、 $m_i$  はそれぞれ 10 分毎の平均値、分散値である。残りの  $v_i$ 、 $r_i$  はそれぞれ次の式で計算される。

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

非終端記号には  $|$ ,  $+$ ,  $-$ ,  $\times$ ,  $\div$  を使う。

上に述べた条件において、世代数 100、集団数 1,000、選択方式 TOURNAMENT (size=6) で GP をそれぞれの条件につき 10 回ずつ実行する。その実行結果を表 1 に示す。このデータの場合 100% 予測が的中すると利益は 80,103 円である。

表 1 実行結果

	平均二乗誤差 (訓練内)	的中率 (%)		利益 (円)	
		平均値	最良値	平均値	最良値
A	1.79e-06	55.02	62.78	12411	31256
B	1.22e-05	47.47	48.17	-4093.2	-2341.5
C	5.82e-04	50.42	51.00	127.03	305.13
D	1.28e-05	41.09	51.64	-19728	-3811
E	1.80e-06	61.38	62.56	28942	30897

Predicting Financial Data by means of Evolutionary Computation

<sup>†</sup> Katuhiko Ono · Department of Information and Communication Engineering, Faculty of Engineering, The University of Tokyo

<sup>‡</sup> Hitoshi Iba · Graduate School of Frontier Sciences, The University of Tokyo

### 3.2 STROGANOFF の実行

GP と同じ A~E の各条件において STROGANOFF 固有のパラメータの W の値を何種類か変えて実行する。ただし集団数は 100 とする。

STROGANOFF を W 毎に 10 回ずつ実行した結果を表 2 に示す。さらに条件 A で最良の結果を得た木構造による予測値と実際の日経 225 との比較を図 1 に示す。

表 2 STROGANOFF の実行結果

	W	SN <sup>2</sup> (訓練内)	的中率(%)		利益(円)	
			平均値	最良値	平均値	最良値
A	0.2	9.40e-06	62.3	62.4	30712	30762
	0.1	9.38e-06	62.3	62.4	30744	30762
	0.01	9.37e-06	62.2	62.3	30516	30823
B	0.2	1.25e-05	57.5	57.7	18636	19194
	0.1	1.25e-05	57.3	57.7	18594	19194
	0.01	1.24e-05	55.3	57.7	13266	19194
C	0.2	6.57e-04	50.0	50.3	1599	3156
	0.1	6.57e-04	50.0	50.3	1517	3156
	0.01	6.57e-04	50.0	58.2	841	4044
D	0.2	1.26e-05	57.6	57.7	18995	19194
	0.1	1.25e-05	57.2	57.7	18390	19194
	0.01	1.25e-05	54.9	57.7	13569	19194
E	0.2	7.25e-04	51.2	51.3	5785	6071
	0.1	7.25e-04	51.6	51.7	5381	5443
	0.01	7.24e-04	51.7	51.7	5443	5443

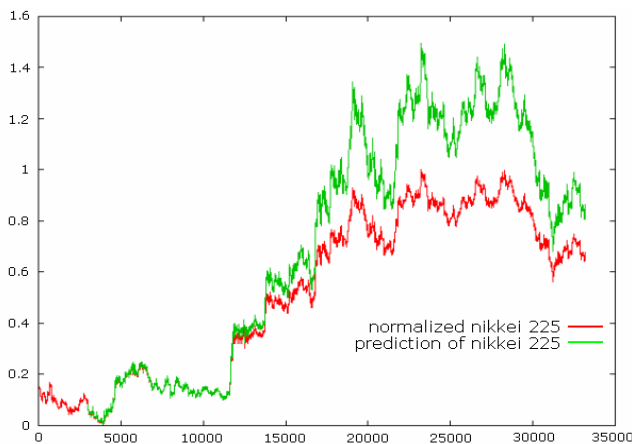


図 1 実際の値との比較

### 3.3 実行結果の考察・検討

STROGANOFF による予測はシンプルな GP に比べると最良値はほとんど同じだがどの試行でも非常に安定した結果が得られた。したがってシンプルな GP よりも一般的なデータに適合していると考えられる。

### 4 オプション価格の予測

次にオプション価格を求める式の探索をする。オプション価格を求める式としてはブラック・ショールズ式が有名であるがこの式はいくつかの仮定のために実際の値とは異なる。そこで GP を用いてより精度の高いオプション価格算出式を求めてみる。

オプション価格には実際の市場のデータを用いる。今回用いるのは、日経 225 の 2005 年のデータである。6 月を訓練に 7 月をテストデータに用いる。終端記号には株価, HV (ヒストリカルボラティリティ), 権利行使価格, 満期までの日数, 安全利率を用いる。非終端記号として {+, -, \*, %, sin, cos, exp, ln, sqrt, normal (正規分布関数)} を用いる。ターゲットとしてブラック・ショールズ式の値を用いる。

GP を 10 回実行した結果を表 3 に示す。HP はホームページの理論値、BS はブラック・ショールズ式により計算した値、また、使用したオプション価格の平均は 763.81 である。この表から GP の結果が訓練・テストとも最良の結果を得られたのがわかる。

表 3 理論値と GP の平均誤差

	理論値 (HP)	理論値 (BS)	GP 平均
訓練	26.9	155.3	24.3
テスト	54.1	154.3	34.1

### 4.1 考察・検討

GP によるオプション価格の予測にはまだ改善の余地が残されている。誤差の要因の 1 つには、実験に使ったデータに取引がない(満期まで遠すぎる)時期の価格のデータも含まれていることがあげられる。こういったデータを除外すると、さらに結果は向上すると考えられる。

### 5 おわりに

本論文では日経の平均株価とオプション価格を扱った。今後は他の金融データも扱う予定である。

### 参考文献

- [1] Iba, H. and Sasaki, T.: "Using Genetic Programming to Predict Financial Data" in Proc. 1999 Congress on Evolutionary Computation (CEC99), pp.244-250, 1999
- [2] Baba, N. and Kozaki, M.: "An intelligent forecasting system of stock price using neural network", Proceedings of IJCNN, 1, pp.371-377, 1992.
- [3] J. LEE, S. LEE, S. CHANG, N P MAHALIK, and B AHN: "A comparison of GA and PSO for Return Evaluation in Stock Market" IEICE TRANS. FUNDAMENTALS/COMMUN. /ELEC-TRON. /INF. & SYST. VOL. E85-A/B/C/D. No. 1 2002
- [4] Ivakhnenko, A. G. : "Polynomial Theory of Complex Systems" IEEE Tr. SMC, vol. SMC-1, no. 4, 1971
- [5] Iba, H., deGaris, H., Sato, T.: "Numerical Approach to Genetic Programming for System Identification", Evolutionary Computation vol. 3, no. 4, pp. 417-452, 1996
- [6] Tenorio, M.F. and Lee, W.: "Self-organizing Network for Optimum Supervised Learning" IEEE Tr. Neural Networks, vol. 1, 1990