

GAによる株取引戦略獲得における指標とパラメータのコーディング法に関する検討

Coding of Indicators and Parameters to Acquire Stock Trading Strategy Using GA

松井 和宏 佐藤 晴夫
Kazuhiro MATSUI Haruo SATO

日本大学工学部情報工学科
Department of Computer Science, College of Engineering, Nihon University

Automatic trading methods are important issues in recent financial markets. In this paper, we compare some genotype coding of technical indicators and their parameters to acquire stock trading strategy using Genetic Algorithms. There are two types of genotype coding for technical indicators: Allele-based coding and locus-based one. Also, we have two kinds of parameter coding: Direct coding and indirect one. We examine the performance of these coding methods in stock trading of twenty companies in the first section of Tokyo Stock Exchange for recent ten years. In preliminary results, the allele-based indirect coding is superior to the other ones.

1. はじめに

近年の金融市場においては、自動取引手法が注目されている [Brabazon 08]。特に、機械学習の技術を用いた自動取引手法については、さまざまな手法が提案されている [松井 08]。本研究では、その中でも遺伝的アルゴリズム (GA) を用いる手法に着目し、取引指標とそのパラメータの遺伝子コーディングについて検討する。

2. GAによる株取引戦略の獲得

2.1 テクニカル分析とGAの適用

本研究ではテクニカル指標に基づいた株取引戦略について扱う。一般に、株取引のためのアプローチはファンダメンタル分析とテクニカル分析に大別される。前者は企業の財務状況などの分析に基づいた取引であり、後者は過去の価格変動だけに基づいた取引である。

テクニカル分析では多数の指標が用いられる。これらは過去の価格変動から何らかの方法によって算出されるものであり、移動平均、ボリンジャーバンドといったものが代表的である。また、これらのテクニカル指標には算出のためのパラメータが付随する。例えば、代表的なテクニカル指標である移動平均には、算出するための期間というパラメータがある。これによって、「10日間の移動平均」、「50日間の移動平均」のように算出される。

テクニカル指標は多数知られているが、その中から取引のために実際に使用する指標を決定するのは困難である。さらに、各指標の算出のためのパラメータを決定するのも容易ではない。本研究では、これらの問題の解決にGAを適用する。すなわち、使用するテクニカル指標とそのパラメータの情報を遺伝子上にコード化しGA探索を適用することによって、取引に適した指標とパラメータの組み合わせを獲得することを目指す。本研究の目的は、GAの適用のために指標とパラメータを遺伝子上にコード化する方法について比較検討することである。

2.2 システム全体の流れ

本研究で用いたシステムの全体構成を図1に示す。まず、あらかじめ定められた期間の株価時系列データを用いて訓練を行
連絡先: 松井和宏, 日本大学工学部情報工学科, 福島県郡山市
田村町徳定字中原原 1, matsui@cs.ce.nihon-u.ac.jp

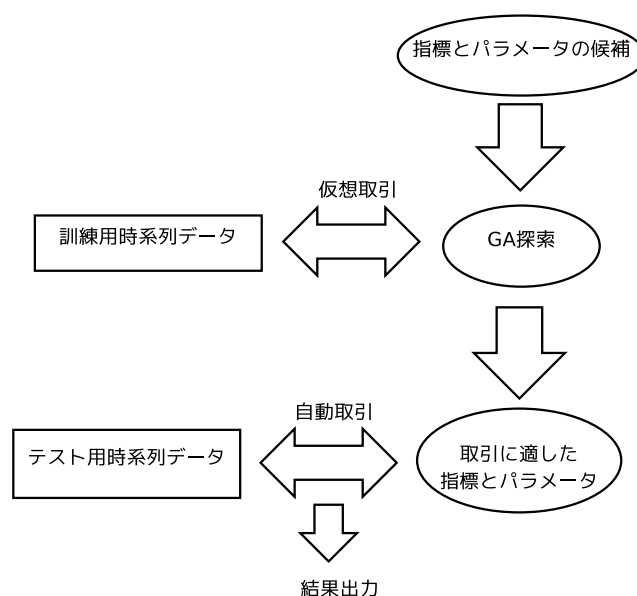


図 1: 全体構成

なう。これは株取引に有効なテクニカル指標とそのパラメータの組み合わせをGAによって探索するものである。次に、訓練によって得られた最良解を用いてテストを行なう。これは実際の自動取引のシミュレーションであり、テスト時にはGAは適用しない。このテストにおける性能を向上させることが重要であり、そのためには訓練時のオーバーフィッティングを回避することが求められる。

3. 遺伝子コーディング

3.1 先行研究

GAを自動取引戦略獲得に用いる方法については、これまでいくつかの研究が行なわれている。まず、テクニカル指標の組み合わせをGAで探索する手法が提案されている [Dempster 01]。これは後述する遺伝子座表現と類似した手法であるが、指標算出のためのパラメータは固定されており、GA探索の対象とは

なっていない。

また、テクニカル指標算出のためのパラメータを GA で探索する手法としても提案されている [de la Fuente 06],[平林 08]。これらは、使用するテクニカル指標を固定しておいてパラメータをバイナリ型 GA で探索する手法であり、後述する直接コーディングに対応する。

3.2 パラメータのコーディング方法

取引戦略獲得のためにテクニカル指標算出パラメータを GA で探索するためには、パラメータを染色体上にコード化する必要がある。従来の研究 (例えば [de la Fuente 06],[平林 08]) では、バイナリ型の GA による探索が主として用いられてきた。しかし、バイナリ型 GA によるパラメータ探索には、以下の問題点があると考えられる。

1. 一般的なバイナリコーディングでは、探索対象となる値の範囲を等間隔で区切ってバイナリコードに割り当てることが多い。しかし、テクニカル指標算出のためのパラメータは、必ずしも等間隔に区切ることが望ましいと言えない場合もある。例えば、移動平均算出期間を最大 100 日間の範囲で探索する場合を考える。4 日間、5 日間、6 日間の移動平均を比べると、算出期間は 1 日ずつしか変わらないが、短期間の売買では結果が異なってくることを考えられる。それに対して 99 日間と 100 日間の移動平均を比べると、算出期間は同じく 1 日の違いであるが、売買結果に与える影響は少ないと予想できる。このように、単純なバイナリコーディングがパラメータ探索に適しているかどうかは検討の余地が残る。
2. 探索対象となるパラメータが実数値の場合、近年研究が進んでいる実数値 GA [小林 09] を用いた方がバイナリコーディングよりも良好な結果が得られる可能性がある。

そこで、特に第一の問題点の観点から、新しいパラメータコーディング法について検討する。一般に、テクニカル指標算出のためのパラメータは、多用される値が存在することが多い。そこで、そのような主要な値だけに GA 探索の対象を限定することが考えられる。例えば、上に挙げた移動平均算出期間の場合、代表的な値として {5, 10, 20, 50, 100} のように対数的に区切った値に限定することができる。これによって、より効率的なパラメータ探索の実現が期待できる。さらに、値を限定することによって計算量が削減できるとともに、不必要に細かい探索を行わないことによるオーバーフィッティングの抑制も考えられる。

なお、以下では、従来から用いられてきたバイナリコーディングによって直接的に値をコード化する方法を直接コーディング、主要な値に限定してコード化する方法を間接コーディングと呼ぶ。

3.3 遺伝子座表現

次に、テクニカル指標の表現方法について 2 つの方法を検討する。第一の手法が遺伝子座表現である。これは、染色体上の各遺伝子座 (各ビット位置) にそれぞれのテクニカル指標を対応付けるものである。図 2 に、その概念を示す。各遺伝子座は、それぞれ「10 日移動平均」と 20 日移動平均の「ゴールクロス」のようなテクニカル指標に対応している。そのビットが「1」であれば、その指標を使用し、そうでなければ使用しない。なお、同図のように本研究ではテクニカル指標とそのパラメータを合わせてコード化しており、パラメータのコード化は間接コーディングとなっている。

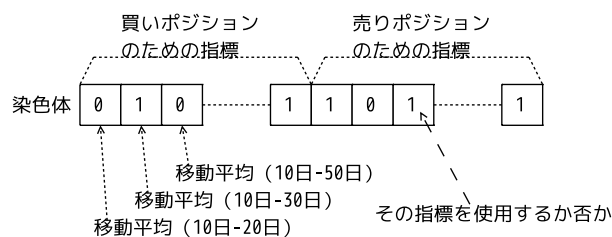


図 2: 遺伝子座表現の概念

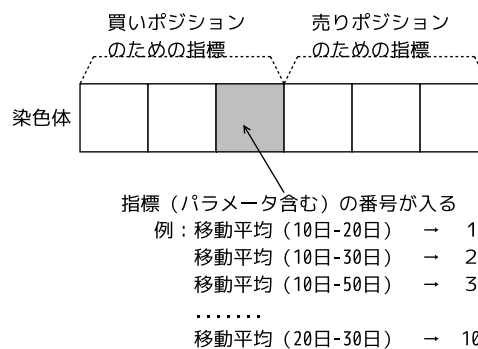


図 3: 対立遺伝子表現の概念

遺伝子座表現における染色体長は、使用するテクニカル指標の候補総数と一致する。そのため、指標の候補が多数の場合には、染色体長も長くなる。また、遺伝子座表現はバイナリコードで表現可能であるため、一般的な GA を容易に適用可能である。

3.4 対立遺伝子表現

第二のテクニカル指標表現法が対立遺伝子表現である。これは、染色体上に配置する対立遺伝子にそれぞれのテクニカル指標を対応付けるものである。対立遺伝子とは、各遺伝子座に置かれる値である。図 3 に、その概念を示す。同図のように、各指標を番号付けておき、その値を対立遺伝子として使用する。対立遺伝子表現では事前に染色体長を定める必要 (図では長さ 3) があるが、前述の遺伝子座表現に比べて、一般に大幅に短くすることができる。また、対立遺伝子の総数は使用するテクニカル指標の候補総数と一致する。そのため、指標の候補が多数の場合には、対立遺伝子数も多くなる。なお、遺伝子座表現と同様に、対立遺伝子表現においても指標とそのパラメータを合わせてコード化している。

4. 実験

4.1 設定

1999 年初めから 2008 年末までの 10 年間の株価データを用いて検証実験を行なった。この 10 年間を通して日経平均に採用されていた銘柄の中からランダムに 20 銘柄を抽出し、取引対象とした。

各銘柄について 4 年間の株価推移を訓練サンプルとして GA 探索を行ない、直後の半年間でテストした。これを 1 セットとし、各セットを半年ずつずらしながら合計 12 セット試行した (図 4)。

初期資金は 500 万円、取引単位は 1 単元 (最小取引単位) とし、手数料を 1 回の取引当たり 1000 円と設定した。表 1 に

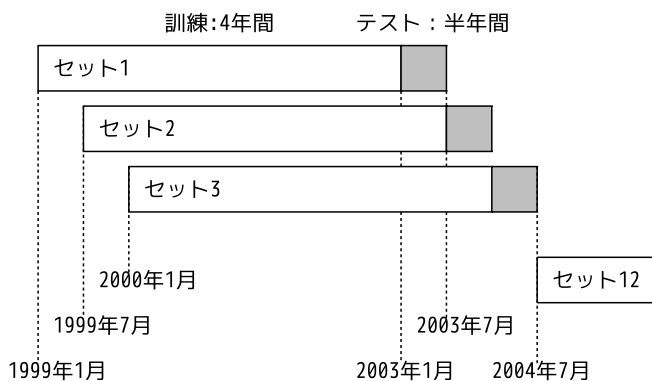


図 4: 実験セットのイメージ

表 1: 使用したテクニカル指標

指標	パラメータ
単純移動平均	算出期間 (短期、長期)
指数平滑移動平均	算出期間 (短期、長期)
ボリンジャーバンド	算出期間、係数
チャンネルブレイクアウト	算出期間

示したテクニカル指標を使用した。間接コーディングにおける算出期間は、{5, 10, 15, 20, 25, 30, 50, 75, 100, 200} の 10 通り、ボリンジャーバンドの係数は {1.0, 1.5, 2.0, 2.5, 3.0} の 5 通りとした。これらのパラメータの組み合わせを考慮に入れた指標数は 150 である。さらに、「買い」の指標と「売り」の指標を考慮して 2 倍となるため、最終的な指標の総数は 300 となる。また、直接コーディングにおいては、8 ビットのバイナリ表現とする。各指標の詳細な算出方法は付録に譲る。本実験では日足データを用いており、GA 個体から得られた指標の組み合わせに適合した場合には、翌日の始値でポジションを開く、または閉じるものとする。

GA で用いる個体数は 50、探索世代数は 5000 とした。世代交代モデルは MGG (最小世代ギャップモデル) [佐藤 96] を用い、交叉は一様交叉とした。突然変異としては、対立遺伝子表現ではランダム置き換え型、遺伝子座表現ではビット反転型を用いた。個体の適応度は取引対象期間で得られた利益とした。

本実験では、間接コーディングによる対立遺伝子表現、間接コーディングによる遺伝子座表現、直接コーディングによる遺伝子座表現の 3 種類を比較した。3 つめの直接コーディングによる遺伝子座表現は、先行研究における代表的な手法として、比較対象とした。

4.2 結果

全 12 セットのテスト期間を通して、表 2 の結果が得られた。金額は 1 万円単位である (以下同様)。表中の最大ドロウダウンは、「連敗」によって受けた損失の最悪値である。各手法ともに利益を上げているが、対立遺伝子表現を用いた場合の利益が最も高くなった。また、表中の平均計算時間は 12 セットの平均である (実験に用いた CPU は Intel Xeon 2.8GHz)。計算時間においては、直接コーディングと間接コーディングで大幅な違いが生じた。

図 5 に、全 12 セットのテスト期間を通しての総利益の推移を示した。いずれの手法においても、前半の期間において利益

表 2: 実験結果

	間接 対立遺伝子	間接 遺伝子座	直接 遺伝子座
取引回数	266	217	186
総利益	237.0	131.9	162.8
最大ドロウダウン	35.4	46.8	28.5
平均計算時間	1 分 09 秒	1 分 54 秒	16 分 12 秒

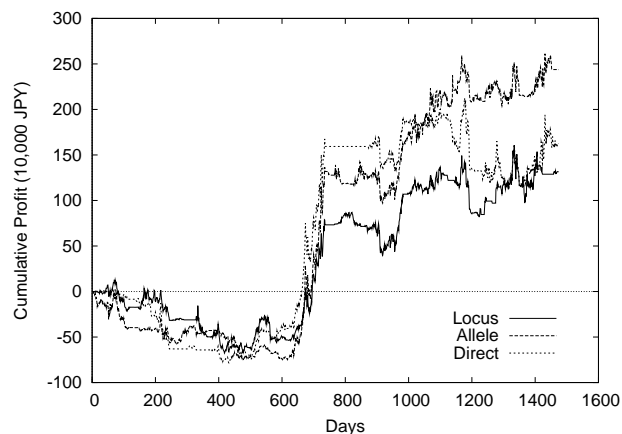


図 5: 総利益の推移

が負の値 (損失) となっている時期があるが、その後回復し、最終的には利益を得ることができた。

表 3, 4, 5 に、訓練とテストにおける性能比較をまとめた。本実験では訓練期間を 4 年間、テスト期間を半年間としたため、両者を 1 年間の結果に換算し、全 12 セットで平均化した。いずれの実験においても、訓練時の利益に比べてテスト時の利益が減少した。

4.3 考察

今回の実験の結果、間接コーディングによる対立遺伝子表現が最も大きい利益を得るとともに、計算時間も短くなるという結果となった。計算時間に関しては、間接コーディングではあらかじめ決められたパラメータの値しか用いないのに対して、直接コーディングでは用いられる値の数が多いために大きな差が生じたと考える。

また、直接コーディングを用いた場合は、取引回数が少なくなっている。これは、訓練によって得られた指標の組み合わせが訓練時の時系列データにオーバーフィットしてしまい、テストデータにおいて指標に適合する回数が減少しているものと考えられる。

図 5 に示した利益推移からは、セット毎の利益変動が大きい

表 3: 訓練とテストの比較: 対立遺伝子表現

	訓練	テスト	比
取引回数	52.5	44.3	84%
利益	104.7	39.5	38%

表 4: 訓練とテストの比較: 遺伝子座表現

	訓練	テスト	比
取引回数	51.0	36.2	71%
利益	98.2	22.0	22%

表 5: 訓練とテストの比較: 直接コーディング

	訓練	テスト	比
取引回数	40.2	31.0	77%
利益	106.5	27.1	25%

いことが分かる。特に損失が生じている時期が存在することは、改善の余地が残る。しかし、全 12 セットの最後の期間となる 2008 年後半は、世界的な金融危機によって市場全体で株価が大きく下落していた時期（グラフ中の 1400 日付近）であるが、そのような状況下でも資産を増やしていることは注目に値すると考える。

表 3, 4, 5 に示した訓練とテストの比較に関しては、いずれの手法においても訓練時の利益に比べてテスト時の利益が大幅に減少している。これはオーバーフィッティングの傾向があることを示唆しており、改善の余地が残る。

5. むすび

GA を用いて株取引手法を獲得するためのテクニカル指標とそのパラメータの表現方法について比較検討した。今回の実験においては、間接コーディングによる対立遺伝子表現が最も大きい利益を上げることができた。また、直接コーディングに比べて、間接コーディングは計算量を大幅に減らすことができることを確認した。

今後の課題は、訓練時におけるオーバーフィッティングを防ぐことと、セット間での利益の差を減らして安定した利益を上げられるようにすることである。また、今回用いたテクニカル指標は大別して 4 種類と比較的少なかったため、より多くの指標を追加することも課題である。

謝辞

本研究の一部は、科研費補助金（基盤研究（C）20500215）の補助を受けた。

参考文献

- [Brabazon 08] A. Brabazon and M. O'Neill: "An Introduction to Evolutionary Computation in Finance," IEEE Computational Intelligence Magazine, pp.42–55, Nov., 2008.
- [松井 08] 松井: "ファイナンスにおける機械学習," JSSST チュートリアル, 2008.
- [Dempster 01] M.A.H. Dempster and C.M. Jones: "A Real-time Adaptive Trading System Using Genetic Programming," Quantitative Finance, 1, pp. 397–413, 2001.

[de la Fuente 06] D. de la Fuente, et al.: "Genetic Algorithms to Optimise the Time to Make Stock Market Investment," Proc. GECCO, pp.1857–1858, 2006.

[平林 08] 平林, 伊庭: "遺伝的アルゴリズムによる外国為替取引手法の最適化," 人工知能学会全国大会, 3H1-2, 2008.

[小林 09] 小林: "実数値 GA のフロンティア," 人工知能学会誌, 24, 1, pp. 147–162, 2009.

[佐藤 96] 佐藤, 小野, 小林: "遺伝的アルゴリズムにおける世代交代モデルの提案と評価," 人工知能学会誌, 12, 5, pp.734–744, 1996.

付録

本研究で用いたテクニカル指標の算出方法を以下に記す。なお、日足データを用い、日付 t における終値を c_t 、高値を h_t 、安値を l_t とする。また、日付 $t-1$ は t の前日を意味する。

1. 単純移動平均

過去 n 日間の終値の単純平均であり、日付 t において式 (1) で算出される。本研究では、「ゴールデンクロスで買い、デッドクロスで売る」というルールを用いた。

$$SMA_n(t) = \frac{1}{n} \sum_{i=0}^{n-1} c_{t-i} \quad (1)$$

2. 指数平滑移動平均

最近の価格を重視し、古い価格の重みを減らすような重みを付けた移動平均であり、式 (2) で算出される。

$$EMA_n(t) = \begin{cases} SMA_n(t) & (t = 0) \\ EMA_n(t-1) \\ \quad + \alpha(p_t - EMA_n(t-1)) & (t \geq 1) \end{cases} \quad (2)$$

3. ボリンジャーバンド

価格変動幅の標準偏差に基づく指標であり、式 (3) で算出される。本研究では、株価がバンド上限を上回ったときに買い、バンド下限を下回ったときに売るというルールを適用した。

$$BB_n(t) = SMA_n(t) \pm \alpha\sigma \quad (3)$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-1} (p_t - SMA_n(t))^2} \quad (4)$$

4. チャンネルブレイクアウト [Dempster 01]

n 日間での価格変動幅を超えたかどうかで判断する指標であり、式 (5) と (6) で算出される。本研究では、株価が $U_n(t)$ を上回ったときに買い、 $L_n(t)$ を下回ったときに売るというルールを適用した。

$$U_n(t) = \max\{h_{t-i} | 1 \leq i \leq n\} \quad (5)$$

$$L_n(t) = \min\{l_{t-i} | 1 \leq i \leq n\} \quad (6)$$