

GAによる株取引戦略獲得における オーバーフィッティングに関する検討

A Study on Overfitting in Acquiring 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

We propose a new method to evaluate individuals for acquiring stock trading strategy using Genetic Algorithms. In our previous works, we presented an automatic trading method in financial markets. However, it had a tendency of overfitting in genetic searches. Our new approach, namely *neighborhood evaluation*, involves evaluation for neighboring points of genetic individuals as well as themselves. We examine the performance of our method in stock trading of twenty companies in the first section of Tokyo Stock Exchange for recent eleven years, and show the effectiveness of the neighborhood evaluation.

1. はじめに

近年の金融市場においては、アルゴリズムトレードと呼ばれる自動取引手法が注目されている [Brabazon 08]。特に機械学習の方法論を用いた自動取引については、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する。

GA を用いた自動取引手法に関するこれまでの研究において、訓練時とテスト時における損益に大きな差が生じやすいことが指摘されている [松井 09, Matsui 09]。本研究の目的は、この損益差を軽減する方法を提案することである。

訓練とテストにおける損益の差は訓練時のオーバーフィッティングによるものと考えられる。すなわち、訓練サンプル中に含まれる一種の特異点において大きな利益が得られている場合に、GA が過剰に適合してしまう現象が生じている可能性が高い。そこで本研究では、探索点周辺における解空間の形状を考慮した新しい評価手法として近傍評価法を導入する。提案手法では、解空間における単独の探索点だけで GA 個体を評価するのではなく、周辺の探索点を考慮した評価を行なう。これによって特異点による影響を減らすことが狙いであり、テスト時の損益改善が期待できる。本報告では、この近傍評価法を用いた自動取引シミュレーションを実施し、その効果を検証する。

2. GAによる株取引戦略の獲得

2.1 テクニカル分析への GA の適用

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が、本研究ではテクニカル分析を扱う。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等が代表的である。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」のように求められる。

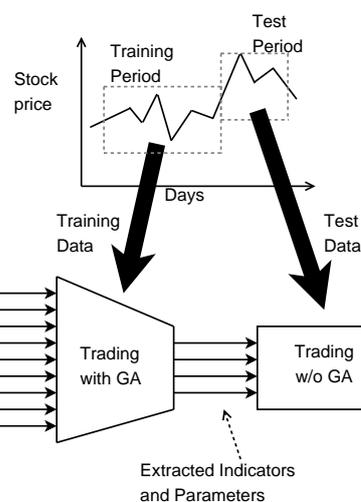


図 1: 全体構成

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な問題である。さらに、各指標の算出のためにパラメータを決定する必要があり、これも容易ではない。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。

本研究における株取引戦略獲得の流れを図 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。

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
田村町徳定字中河原 1, matsui@cs.ce.nihon-u.ac.jp

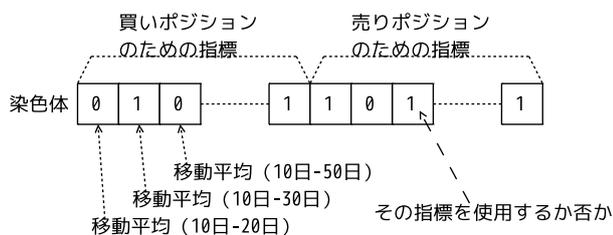


図 2: 遺伝子座表現の概念

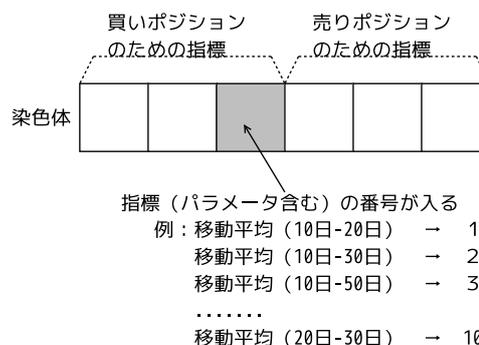


図 3: 対立遺伝子表現の概念

2.2 遺伝子コーディング法

2.2.1 直接コーディングと間接コーディング

GA を取引戦略獲得に用いる手法については、これまでにいくつかの研究が行なわれている。まず、テクニカル指標算出のためのパラメータを GA で探索する方法が提案されている [de la Fuente 06, 平林 08]。これらはパラメータをバイナリ型 GA で直接的に染色体上にコード化して探索する手法であり、ここでは直接コーディングと呼ぶ。この手法は従来広く用いられてきたが、探索範囲を単純に等間隔に区分するために効率が悪くなる傾向がある。

これに対して、テクニカル指標算出において多用されるパラメータだけに限定して効率的な探索を行なうための手法として間接コーディングが提案されている [松井 09, Matsui 09]。一般的なテクニカル指標には、典型的に使用されるパラメータ値が存在することが多い。そこで、そのような代表的な値だけに探索対象を限定することによって効率化を図る手法である。例えば、代表的なテクニカル指標である移動平均値には、算出期間のパラメータが存在する。このパラメータの探索範囲を {5, 10, 15, 20, 25, 30, 50, 75, 100, 200} のように代表値だけに限定することで、単純な均等割ではない効率的な探索を実現できる。本研究では、この間接コーディングを使用する。

2.2.2 遺伝子座表現と対立遺伝子表現

テクニカル指標を GA の染色体上に表現する手法としては、遺伝子座表現と対立遺伝子表現がある。前者は、染色体上の各遺伝子座 (ビット位置) にそれぞれのテクニカル指標を対応づける手法である (図 2)。従来はこれと類似した手法が用いられることが多かった (例えば [Dempster 01])。

これに対して、対立遺伝子表現は染色体上に配置する対立遺伝子にテクニカル指標を対応づける手法である (図 3)。著者らは 2 つの表現法を比較し、対立遺伝子表現が良好な結果を得ることを確認した [松井 09, Matsui 09]。そのため、本研究では対立遺伝子表現を使用する。

3. 従来法の問題点

間接コーディングと対立遺伝子表現を組み合わせた GA によって、ある程度機能する株取引戦略を獲得することができる。しかし、いくつかの問題点が残されている。特に大きな問題は、訓練時 (GA による探索時) に得られていた利益に対して、テスト時 (獲得した戦略による自動取引時) に得られる利益が少ないことである。これは、訓練時におけるオーバーフィッティングの可能性を示している。すなわち、訓練時の株価時系列データに過剰に適合した取引戦略が良好な適合度を獲得し、最終的な戦略として採用されてしまうために、テスト時における時系列データにうまく適合しない状態となっている可能性が高い。

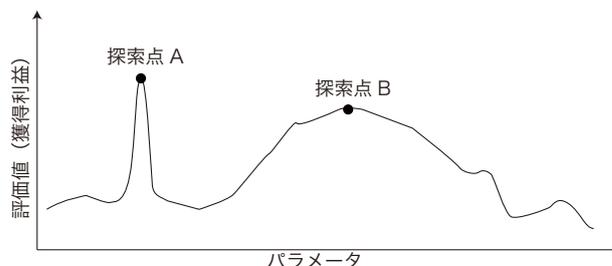


図 4: 特異点の概念

オーバーフィッティングの原因として考えられるのは、図 4 に示すような特異点の影響である。GA によってパラメータの最適値を探索する際に、ある特定のパラメータ値で、その周辺と比べて大幅に高い評価値 (利益) が得られる状態を特異点と呼ぶことにする。この図では、探索点 A が特異点である。通常の GA 探索は評価値が最大となる点を探索することが目的であるため (最大化問題の場合)、探索点 A を発見することができれば探索は成功である。しかし、本研究で扱っている株取引戦略獲得問題においては、GA 探索 (訓練) と実際の取引 (テスト) において探索空間の形状が異なったものになる。そのため、GA 探索における最適点がテストにおける最適点に一致するとは限らない。テスト時には GA は実行されないため、最適点を事前に知ることは困難である。訓練時の探索空間における最適点に適合した取引戦略を GA が発見したとしても、テスト時の探索空間に適合するとは限らない。これが、訓練とテストにおける利益の差が生じる原因となる。

4. 提案手法

前節で示した従来法の問題点を解決するために、本研究では新しい GA 個体評価手法を提案する。図 4 において、探索点 B は探索点 A に比べて評価値は低い。そのため、通常の GA 探索においては最適解とはならない。しかし、探索点 A の周辺が「急峻」な形状であるのに対して、探索点 B の周辺は「なだらか」である。「急峻」な形状に比べて「なだらか」な形状である方が、テスト時における利益の減少が少なくなることが期待できる。なぜならば、「急峻」な形状による高い評価値は、訓練データ中に存在する何らかの特殊な理由によって得られた可能性が高く、テストデータにおいても通用する可能性は低いからである。一方、「なだらか」な形状であれば、このような特殊な理由が存在する可能性は低いと考えられる。

以上の予測から、着目している探索点だけでなく、周辺の探索点の評価値も考慮し、これらを統合した新しい評価値を算出して用いる手法を提案する。本研究では、この新しい評価法を近傍評価法と呼ぶ。

著者らのこれまでの研究では、探索中の着目点 i に関する評価値 f_i は、

$$f_i = P_i^{long} + P_i^{short} \quad (1)$$

によって求めていた。ここで P_i^{long} と P_i^{short} は、それぞれ買い/売りポジションによる損益の総和である。

これに対して近傍評価法での評価値 F_i は、

$$F_i = F_i^{long} + F_i^{short} \quad (2)$$

によって求める。 F_i^{long} と F_i^{short} は、それぞれ買い/売りポジションによる近傍評価の値であり、次式によって求められる。

$$F_i^{long} = \sum_{j \in \text{IndexSet}^{long}} \frac{g_{j,i}^{long}}{\beta_{j,i}} \quad (3)$$

ここで、 IndexSet^{long} は、着目している GA 個体の染色体に含まれている買いポジション用の取引指標の集合であり、 j はその集合の1つの要素、すなわち1つの取引指標である。また、 $g_{j,i}^{long}$ と $\beta_{j,i}$ は以下によって求められる。

$$g_{j,i}^{long} = \sum_{k \in N_{j,i}} \alpha_{k,i} P_k^{long} \quad (4)$$

$$\beta_{j,i} = \sum_{k \in N_{j,i}} \alpha_{k,i} \quad (5)$$

ここで、 $N_{j,i}$ は取引指標 j に関する着目点 i の近傍点の集合であり、着目点 i 自身も含まれる。 $\alpha_{k,i}$ は、着目点 i と近傍点 k の位置関係によって定められる係数であり、近傍評価に与える影響の強さを決める。 β は、 α の効果を正規化するために用いる。また、 P_k^{long} は近傍点 k における買いポジションによる損益の総和である。なお、売りポジションについても同様に計算する。

5. 実験

5.1 設定

1999年初めから2009年末までの11年間の株価データ(日足)を用いて検証実験を行なった。この11年間を通して日経平均に採用され続け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、2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1年間でテストした。これを1セットとし、各セットを1年ずつずらしながら合計9セット実行した(図5)。

初期資金を500万円、取引単位を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。各指標の詳細な算出方法は著者らの以前の報告と同様であるため省略する[松井 09]。本実験では日足データを用いており、

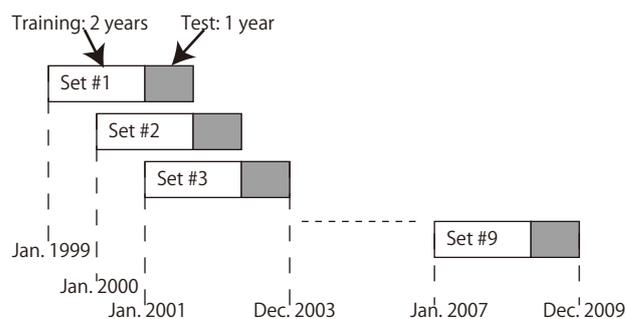


図 5: 実験セットのイメージ

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

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)を用い、一様交叉とランダム置き換え型突然変異を適用した。

個体の適応度計算には近傍評価法を用いた。表1に示したテクニカル指標のうち、チャンネルブレイクアウト以外の3種類についてはパラメータが2種類あるために、2次元の近傍となる。今回の実験では図6のように4近傍とした。この図は、算出期間50日と係数2.5の組み合わせを着目点とした場合の例である。一方、チャンネルブレイクアウトのパラメータは1種類であるため、1次元の近傍である。図7に算出期間25日の例を示した。なお、いずれの場合においても、近傍パラメータ α の値は着目点を1.0、近傍点を0.5とした。

5.2 結果

実験結果を表2に示した。乱数による偏りを考慮して同一条件で10回の試行を行ない、総利益が最大と最小の試行を除外した8試行を平均化した。金額は10,000円単位である(以下同様)。表中の総利益は全9セットを通しての利益の合計であり、最大ドロウダウンは「連敗」によって受けた損失の最悪値

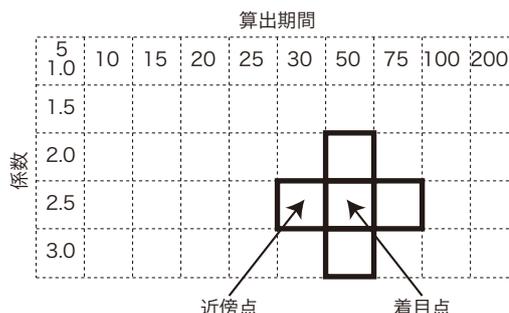


図 6: 2次元近傍の例: ボリンジャーバンド



図 7: 1次元近傍の例：チャンネルブレイクアウト

表 2: 実験結果

手法	提案法	従来法
取引回数	664.9	717.5
総利益	287.2	145.7
利益率	57.4%	29.1%
最大ドロウダウン	67.5	82.2
平均計算時間	35分20秒	1分55秒
訓練とテストの利益比	22.5%	9.7%

の平均である。また、表中の平均計算時間は全9セットの合計計算時間の平均値である。なお、実験に用いたCPUはIntel Xeon 2.66GHzである。表最下段の利益比は、訓練時に得られた利益に対するテスト時に得られた利益の比である。本実験では訓練期間を2年間、テスト期間を1年間としたため、前者を1年間の結果に換算し、全9セットで平均化した。

表2から、提案法は従来法に比べて大幅に総利益を増加させること、および最大ドロウダウンを低下させることが分かった。しかし、計算時間は大幅に増大した。また、本研究の主目的であるオーバーフィッティングに関しては、提案法は従来法に比べて大きく改善した。

図8に、全9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。しかし、1600日程度経過した時点から両手法の差が生じ始めている。

5.3 考察

今回の実験により、提案法は従来法に比べて大きな利益を得ることができることを確認した。また、テストと訓練における総利益の比も従来法に比べて改善することができた。近傍評価

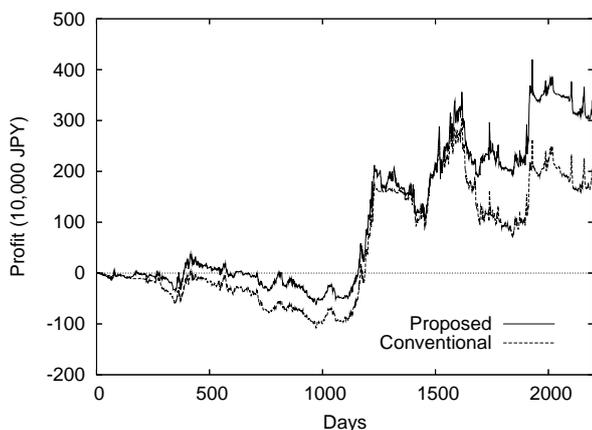


図 8: 総利益の推移

法により探索空間中の特異点の影響を減らすことができたのが、これらの結果につながったものと考えられる。しかし、表2に示したように、従来より改善されたものの依然としてオーバーフィッティングの影響は残っていると見える。

また、計算時間に関しては、提案法は従来法に比べて大幅に増大している。これは提案法の欠点であり、改善が求められる。例えば、既に評価を終えた探索点について評価値を記録しておくことによって、重複した評価を減らすような手法が考えられる。

6. むすび

GAを用いて株取引戦略を獲得するための手法について検討し、訓練時におけるオーバーフィッティングの影響を減らすための新しいGA 個体評価法として近傍評価法を提案した。今回の実験においては、提案法は従来法に比べて良好な結果を得ることができた。しかし、オーバーフィッティングを完全に除去する段階までは至らなかった。

今後の課題は、改善されたものの依然として残るオーバーフィッティングの影響を減らす方法について検討することと、計算時間の削減である。また、今回用いた形状以外の近傍領域についても検討する必要がある。

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] 松井藤五郎: "ファイナンスにおける機械学習," JSSST チュートリアル, 2008.
- [松井 09] 松井和宏, 佐藤晴夫: "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," 第23回人工知能学会全国大会, 3G3-OS12-4, 2009.
- [Matsui 09] K. Matsui and H. Sato: "A Comparative Study of Genotype Coding in Genetic Approaches for Acquiring Stock Trading Strategy," The 9th Asia-Pacific Complex Systems Conference, 2009.
- [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.
- [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.