

機械学習を用いたポートフォリオの最適化

Dynamic portfolio reconfiguration by machine learning techniques

海野一則^{*1} 山田隆志^{*1} 寺野隆雄^{*1}
Kazunori UMINO Takashi YAMADA Takao TERANO

^{*1} 東京工業大学

This study proposes a new investment method to adapt to the characteristics and changes of asset markets. For this purpose, two kinds of filters, momentum filter and AntiCor filter both of which are observed in the markets, are used. The momentum filter, on the one hand, is to select assets with high momentum. The AntiCor filter, on the other hand, is to forecast the process of mean-reverting for an asset by using the mean-reverting characteristics of its similar assets. The application of the former proposed method to S&P 500 Index and that of the latter to Nasdaq-100 between 2000 and 2012 show that both the methods have the possibility to earn excess profits by setting the pertinent market environments and selecting the possible asset selections.

1. はじめに

株式トレードおよびポートフォリオ構成における機械学習の応用は数多く見られる。しかしながら、機械学習で容易に株価予測や有利なポートフォリオを構築できるわけではない。

株式市場においては、Fama [Fama 70]による効率的市場仮説があり、どのような有益な情報やテクニカル分析を利用しても市場平均をアウトパフォームできないと言われている。

1つは、マーケットにおいて常にモーメンタム(勢い)の高い株式が存在し、その後もモーメンタムが持続することである。例えば、Jegadeesh [Jegadeesh 93]では、高いモーメンタムの株式群が、その後3ヶ月から 12 ヶ月においてプラスの超過リターンが得られると述べている。

2つ目は、会社の事業分野が同じで、企業規模が近い場合を類似した会社と呼ぶ。そのような会社の株式間においては、相関の高い株価変動を示すと考えられる。Bonanno [Bonanno 02]では、日次データをもとに、相関性の近いものからツリーを構成する手法で、その変動の類似性を可視化し、さらに事業分野の近さが強く影響していることを示した。

我々は、上記の2つの現象を的確に検出し、ポートフォリオの再構成に利用することで、株式市場から超過収益を得られると考えた。そして、それらの指標の組み合わせやパラメータ設定を遺伝的アルゴリズムなどにより経験則等に頼らない形でのモデル構築を試みた。

また、類似した株式は株価変動において相関の高い変動を示すこと。変動の相関と株価の平均回帰性を利用した予測アルゴリズムとして AntiCor フィルター[Borodin 04]を採用した。

本論文では、2つのフィルターによる逐次的なポートフォリオの再構成モデルを提案する。

2. 関連研究

機械学習を利用した、株価予測・株式トレード・ポートフォリオ構築が行われている。ヘッジファンドなどでは、機械学習の技術者を雇用してアルゴリズム開発を積極的に行なっている [Patterson 10]。

価格データおよびテクニカル分析指標から、将来的な価格の予測や売買の意思決定を行うことが一般的である。さらに、何らかのランダムではない性質を抽出して、そこに特化した学

習モデルを構築するなどのより緻密なアプローチが必要となっている。

Chang [Chang 09]では、個別の株式において、変動を折れ線で近似し、同時に様々なテクニカル分析指標を与えることで、そのパターンをニューラルネットワークに学習させるものである。また、有効な指標の抽出には遺伝的アルゴリズムを用い、様々な指標の組み合わせから有効な指標をトレード収益により選択している。また、結果としては、上昇・持ち合い・下降トレンドの全てにおいて、同じ期間の Buy&Hold をアウトパフォームしている。

Yan [Yan 08]においては、Genetic Programming (GP)および Support Vector Machine (SVM)を用いて、多様なテクニカル分析指標から GP により様々な指標やパラメータの合成を行う。さらに、有効な組み合わせを複数用意し、その結果を投票により決める Voting システムにより優位性を示している。

ただし、上記の手法の問題点は、テスト期間が数年間程度と短く、リーマン・ショックのような大幅な下落に適応できるのか否かが不明である。さらに、個別株の予測のケースでは、多様な銘柄において優位であることは非常に難しい。

3. 提案手法の概要

関連する研究の欠点を補い、より長期の運用における優位性を主眼に置いた、実現可能な手法を構築した。

我々は、最も流動性が高く効率的と言われている米国市場を対象とすることとした。よって、特に 2000 年以降は上昇トレンドが崩れ、大きな下落と反転が発生している。よって、最も難しいと思われる 2000 年以降において検証を行うことが必要不可欠と考えた。よって、2000 年から 2012 年までを検証の区間とした。

また、ポートフォリオによる複数銘柄への分散化が行われていること。さらに、現実的な取引コストが考慮されていることが重要である。

本研究において、モーメンタム・フィルターでは、与えられた株式リストに対して抽出されたモデルを適用し、毎日ポートフォリオの再構成を行なっている。

さらに、一ヶ月毎にモデルの再抽出を行なっている。これは、マーケット環境の特性が変化しているという想定のもと、一定期間毎に有効な指標とパラメータを再度抽出する必要があると考えたからである。本手法では1ヶ月間隔と仮に決定しているが、今後は適正な期間を算出する手法が必要になると考えている。

モデルの例は、以下のようなものである。学習期間において、最もマーケット環境に適したテクニカル指標や、パラメータ(例え

ば、移動平均線の期間)を抽出したものである。次章でその詳細について述べる。

AntiCor フィルターでは、毎日ポートフォリオの再構成を行っている。

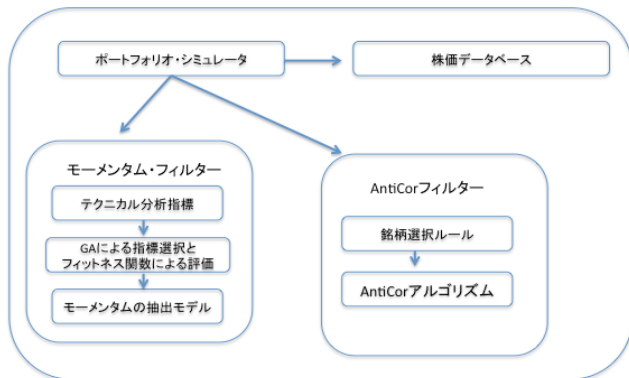


図 1 提案手法の概念図

表 1 モーメンタム・フィルターのパラメータ設定の実例

モーメンタム・フィルターのモデル例 (N 銘柄でポートフォリオを均等に構成する場合)	
MA1=5 日移動平均	
MA 2=10 日移動平均	
MA2 > MA1 {現在値}	
RSI (20) > 53 {平均値}	
RSI (10) > 60 {現在値}	
選択: RSI の高い上位 N 銘柄を選択	
ポジション: +100, -100 (売買サインを一度に実行)	

4. モーメンタム・フィルターの作成

株式市場に発生しているモーメンタム(勢い)は、株式市場における注目されている株式に発生しているものである。つまり、株式市場で注目されている株式は買い手と売り手のインバランスが発生し、買いが多くなれば株価が上昇する。言い換えれば、株式の長期に亘るインバランスをテクニカル分析指標から予測可能なモデル化を行うことである。個別株のモーメンタムがある程度の期間継続するものであれば、市場から超過収益 α を得られる可能性がある。

一般的なテクニカル分析に加え、それらを組み合わせた新たな指標を作成し、テクニカル指標およびパラメータの設定の最適な組み合わせを GA により選択する。

4.1 テクニカル分析

テクニカル分析として、適切なパラメータ設定による移動平均(MA)・MACD・RSI 等のテクニカル指標を用いた。

また Volatility はリターンの標準偏差である。

$$MA(t,1), MA(t,3), MA(t,5), MA(t,10), MA(t,20), MA(t,40), MA(t,80), MA(t,160)$$

$$Volatility(t,10), Volatility(t,20), Volatility(t,30)$$

$$RSI(t,10), RSI(t,20), RSI(t,40), RSI(t,80)$$

$$MACD(t,4,8,2), MACD(t,9,16,3), MACD(t,18,32,6)$$

さらに、上記指標から新たな指標を合成する

$$\langle 1 \rangle CMA(t,i,j) = MA(t,i) - MA(t,j) \quad (i > j)$$

$$\langle 2 \rangle CRSI(t,i,j) = RSI(t,i) - RSI(t,j) \quad (i > j)$$

$$\langle 3 \rangle SCMA(n,t,i,j) = CMA(t-n+1,i,j) +$$

$$CMA(t-n+2,i,j) + \dots + CMA(t,i,j)$$

$$\langle 4 \rangle SCRSI(n,t,i,j) = CRSI(t-n+1,i,j) +$$

$$CRSI(t-n+2,i,j) + \dots + CRSI(t,i,j)$$

$$\langle 5 \rangle SCMACD(n,t,i,j) = MACD(t-n+1,i,j) +$$

$$MACD(t-n+2,i,j) + \dots + MACD(t,i,j)$$

$$\langle 6 \rangle TSCMA(x,n,t,i,j) = \sum_{k=1..x} 1_{\{SCMA(n,t,i,j) > 0\}}$$

$$\langle 7 \rangle TSCRSI(x,n,t,i,j) = \sum_{k=1..x} 1_{\{SCRSI(n,t,i,j) > 0\}}$$

$$\langle 8 \rangle TSCMACD(x,n,t,i,j) = \sum_{k=1..x} 1_{\{SCMACD(n,t,i,j) > 0\}}$$

4.2 ポートフォリオ最適化

モーメンタム・フィルターにより、ポートフォリオのリターンを最大化するための最適化モデルを作成する。

テクニカル指標及びその拡張指標を GA により組み合わせを行い、最も適合度が高いモデルを抽出する。

このモデルを一定期間(現在は1ヶ月)毎に再抽出を行う。

4.3 GA による選択アルゴリズム

GA により表2に示した選択項目の組み合わせにより、モーメンタム・フィルターのモデルを構築する。また、テクニカル分析指標は複数の指標を利用可とした。

表 2 GA による選択項目

テクニカル分析指標	TAIndex={MA,CMA,SCMA, TSCMA, Volatility, RSI, MACD ...} *複数の指標を利用可
銘柄選択ルール	SetRule={現在値, 平均値, 中央値, 合計値, 順位, 百分位数}
ポジション増減ルール	PosAddPct={10,20,50,100} PosSubPct={-10,-20,-50,-100}
ポートフォリオ配分ルール	NStocks={3,4,5,...,20} Dvide={均等, 重みに応じて}
学習・適用期間	LearnPeriod={10,20,...,240} TestPeriod={10,20,...,240}

5. AntiCor フィルターの作成

AntiCor アルゴリズムとは、Follow-the-Loser ポートフォリオ戦略と呼ばれ Anti Correlation Strategy の略である。

下落した株式はどこかのタイミングで上昇し、大きく上昇した株式はどこかで下落を始める。これを平均回帰性と呼ぶ。つまり、株式価格は平均的な水準に対して、乖離と収束を繰り返しているとも考えられる。このような現象を検出するフィルターである。

類似した銘柄間では、価格変動においてある程度の相関が常に働いており、一時的に乖離しても最終的には縮小するという確率が高くなる。さらに、過去のデータを精査すれば、より乖離と収束の発生を予測しやすいと考えられる。

また、Follow-the-Loser とは、上昇している銘柄群に対して下落している銘柄群が追いつこうとする性質を利用している。つまり、上昇している銘柄群は平均回帰性により下落し、下落している銘柄群には逆のことが起こる。その両方の性質から、下落した銘柄群を買えばいずれ均衡する値までは上昇するだろうという予測ができる。この手法の欠点は、逐次的な売買を行いポートフォリオのリバランスを行うため、取引コストが大きいことである。

5.1 AntiCor アルゴリズムの概要

このアルゴリズムは、どのように $t+1$ 期におけるポートフォリオを構成するかというものであり、以下にその基本アルゴリズムを示す [Li 12].

x: price relative vector

w: window size

$X_t \in \mathbf{R}_+^m, t=1, \dots, n$

vectors $X_{t1}^{t2} = \{X_{t1}, \dots, X_{t2}\}$

$y_1 = \text{Log}(X_{t-2w+1}^{t-w})$

$y_2 = \text{Log}(X_{t-w+1}^t)$

It then calculates the cross-correlation matrix between

y_1 and y_2 ,

$$M_{\text{cov}}(i,j) = \frac{1}{w-1} (y_{1,i} - \bar{y}_1)^T (y_{2,j} - \bar{y}_2)$$

$$M_{\text{cor}}(i,j) = \begin{cases} \frac{M_{\text{cov}}(i,j)}{\sigma 1(i) \cdot \sigma 2(j)} & \sigma 1(i), \sigma 2(j) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

if asset i increase more than asset j ans their sequences in the window are positively correrated. Anticor claims a transfer from asset i to j with the amount equals the cross correlation value ($M_{\text{cor}}(i,j)$) minus their negative auto correlation value ($\min\{0, M_{\text{cor}}(i,i)\}$ and $\min\{0, M_{\text{cor}}(j,j)\}$). (p. 19)

6. 実データでの検証

検証期間は、2000年～2012年であり、銘柄の選択は、最も流動性の高い米国のSP500指数およびNasdaq100指数に含まれる600の銘柄から行った。また、データはOpen, High, Low, Close, Volumeからなる日次データである。

本検証では、最初にモーメンタムおよびAntiCorのフィルターにおいて、同じ銘柄群での検証を行い、さらにより多くの銘柄でモーメンタム・フィルターの検証を行った。

これは、モーメンタムの強い銘柄を選択するためには、なるべく多くの銘柄から選択するほうが有利だからである。SP500指数には、多様な業種や規模の異なる会社があるため、市場から抽出すべきモーメンタムの分散化が図れると考えた。

AntiCorフィルターは計算量が大いことを考慮して、ナスダック100銘柄の中より12の銘柄群に絞込んだ。

選択手法は、全期間における株価リターンに応じて分け、候補が多い場合はランダムサンプリングし12銘柄を選別した。本論文では、株価リターンを2000年の最初の取引日の終値の株価を1として、その株価の変化を倍率で示した。ただし、株式分割などは補正し、その間の配当は考慮していない。

ナスダック市場の銘柄のみを選択範囲としたのは、はハイテク産業に関係しているものが殆どであり、近いセクターに属しているものが多く、それにより株価の変動の相関が高いと考えたからである。

ポートフォリオ・シミュレーションにおいては、全期間における資産倍率に応じて異なる株式群に分けた。2000年から2012年までの米国株式市場はITバブル崩壊・その後の大幅なハイテク株の下落・リーマンショック・リーマンショックからの回復という過去に類を見ない激動の期間であった。このような環境において有効な手法であれば、今後の市場の変化が起こったとしても十分に有効であると考えられる。また、また、対象となるナスダック100指数およびSP500指数の銘柄は、2011年度に選ばれている銘柄を基にしており、その銘柄のデータを過去に遡ったデータを利用している。よって、上記の指数では毎年成長性の高い銘柄の入れ替えが行われており、本利用データの平均指数は同時期のSP500指数およびナスダック100指数と異なり同期間に4~5倍程度的大幅な上昇をしている。

よって、ここでは3種類のマーケット環境を意識した銘柄群に分けた。これは、強い上昇バイアスを排除するためでもある。

ナスダック100銘柄より2000年～2012年の期間において、

- ・ 株価リターンが0.0～0.8からランダムに12銘柄を選択
- ・ 株価リターンが0.0～2.0からランダムに12銘柄を選択
- ・ 株価リターンが3.0以上からランダムに12銘柄を選択

これらのケース毎の結果を示す。また、AntiCorフィルターおよびモーメンタム・フィルターでは、同じ12銘柄を株式データとして与え、自動的に4銘柄に分散投資を行った。

取引コストについては、取引金額の0.1%として計算する。また、米国株式の高い流動性の銘柄を選択するため、マーケットインパクト(売買の際に起こる価格変動)は考慮しない。また、米国のRiskFreeRate(無リスク利回り)を考慮すべきであるが、ここでは最終的に得られた幾何平均による年間の平均利回りと比較することで優位性を検証する。

6.1 シミュレーション(株価リターンが0.0～0.8)

図1の上のグラフは、AntiCorフィルターのシミュレーションで、青い線は取引コストを考慮しない場合。赤が取引コストを考慮した場合。そして、緑は市場平均を表したものである。手数料を考慮しても十分に超過収益を得られることがわかる。下のグラフはモーメンタム・フィルターによる結果で、市場平均をアンダーパフォーマンスしている。よって、株価リターンが低いマーケット環境においてはAntiCorフィルターが有効であることがわかる。

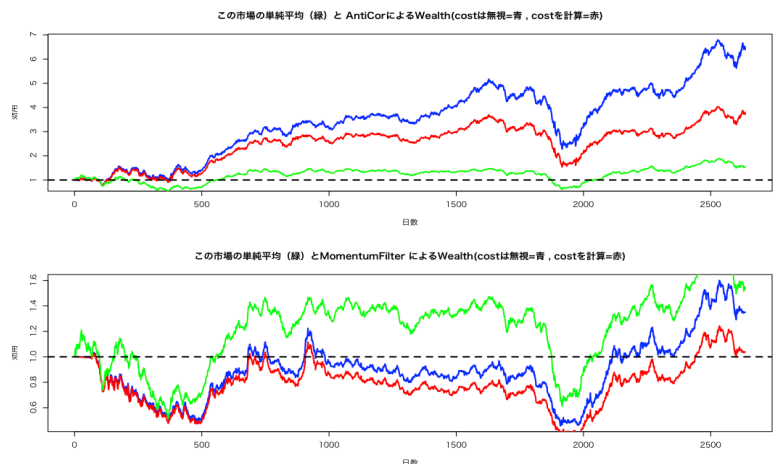


図1 低い株価リターンのみからポートフォリオを構成したケース

6.2 シミュレーション(株価リターン0.0～2.0)

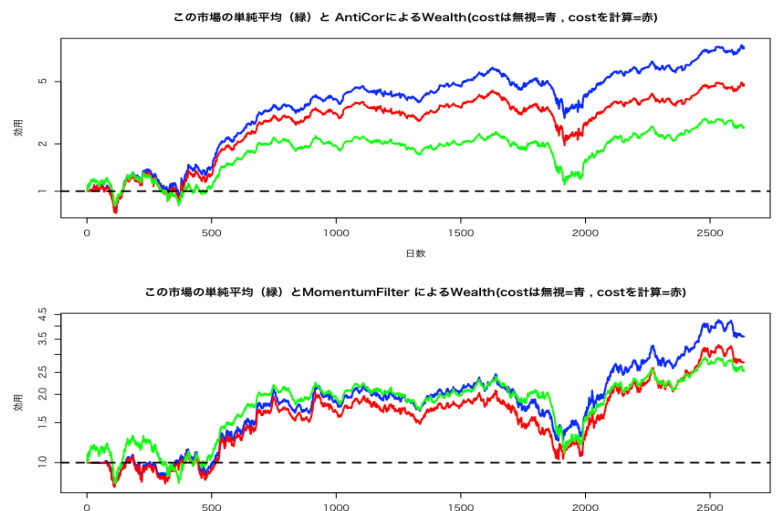


図2 緩やかな上昇トレンドを持つ市場を想定

図2では、緩やかな上昇トレンドを持つ市場においては、AntiCor フィルターでは市場をアウトパフォーマンスしており、モーメント・フィルターでは取引コストを考慮すると市場平均程度のリターンとなる。

6.3 シミュレーション(株価リターン 3.0~)

図3は、大きな上昇トレンドのケースであり、このケースでは両手法とも市場平均を上回っている。全て3倍以上上昇した銘柄のみから選択し、市場平均が 20 倍程度の株価リターンの状況でも、フィルターの効果は有効である。

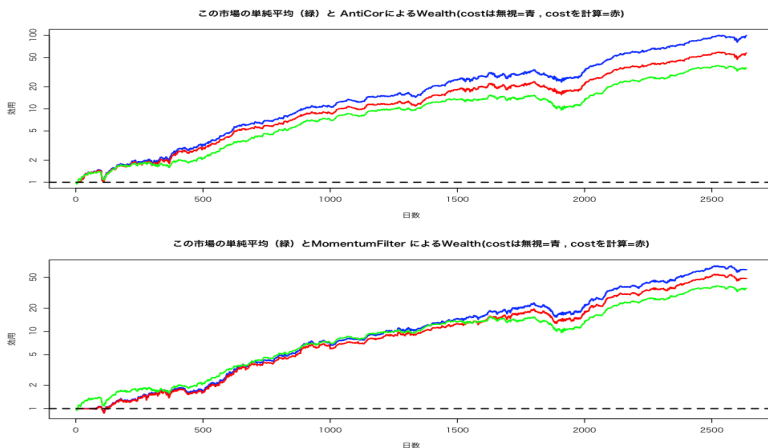


図 3 強い上昇トレンドを持つ市場を想定

手数料を考慮しても、AntiCor アルゴリズムは非常に有効である。ただし、取引コストを 0.1%としてあるため、取引コストがそれを超えた場合、収益は大きく減少するという欠点がある。

6.4 モーメント・フィルターによるシミュレーション

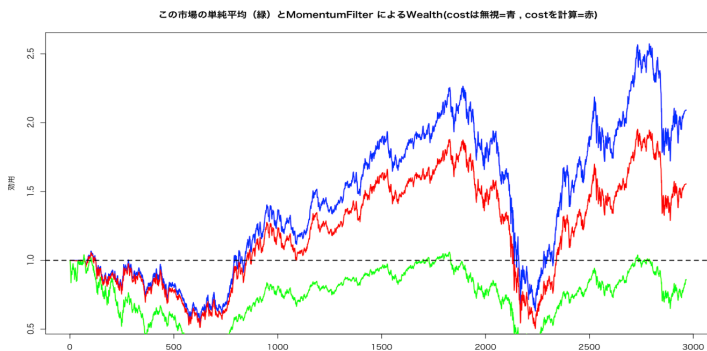


図 4 株価リターン 0.5 以下の 50 銘柄より 4 銘柄を選択

図4は、SP500指数から選択銘柄候補を 50 にしたものである。モーメントがより高頻度で発生することによりリターンが増加している。

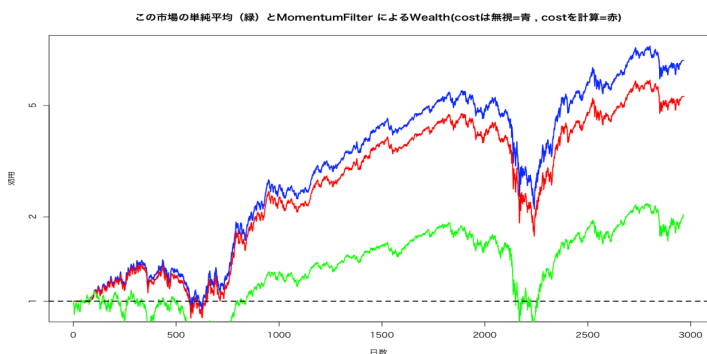


図 5 株価リターンが 2.0 以下の 237 銘柄から 4 銘柄を選択

図5では、より多くの株式(SP500 指数の 237 の銘柄)から選択することで、全ての株式の株価リターンが 2.0 以下であっても、図5に示したように、モーメントを検知することで、市場平均を大きくアウトパフォーマンスすることが可能である。

7. まとめ

本研究では、モーメント・フィルターおよび AntiCor フィルターの二つの手法を実装し、それぞれの特性を検証した。有効性が確認できたケースにおいては、効率的市場仮説において言及されていない恒久的な性質および株式間の関係性が存在していることを間接的に示していると考えている。

結果をまとめると、AntiCor フィルターはナスダック 100 銘柄から選択する場合、取引コストを 0.1%以内であれば、強いダウントレンドの市場でなければ超過収益があげられる可能性が高い。また、モーメント・フィルターは選択候補の銘柄数が 12 程度と少ない場合は有効性がないが、50 銘柄以上になれば有効性が高くなり、市場がゆるやかな下降トレンドであっても超過収益を得ることが可能である。

また、モーメントが分散している SP500 指数の方が相関の高いナスダック指数の銘柄より有利といえる。

今後の研究課題としては、より高精度に検出可能とするフィルターの改良と日本市場での検証を行いたいと考えている。

参考文献

- [Fama 70] E.F.Fama :Efficient Capital Markets:A Review of Theory and Emprical Work, Journal of Finance, Volume 25, Issue 2,Papers and Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association, N.Y.December, 28-30,pp383-417 (1970)
- [Jegadeesh 93] N.Jegadeesh, S.Titman :Returns to Buying Winners and Selling Losers:Implications for Stock Market Efficiency, The Journal of Finance, VOL. XLVIII, NO. 1 MARCH (1993)
- [Bonanno 02] G.Bonanno, G.Caldarelli, F.Lillo, R.N.Mantegna :Topology of correlation based minimal spanning trees in real and model markets: 1993-2000, Phys. Rev. E 68, 046130 (2003)
- [Borodin 04] A.Borodin, R.EL-Yaniv, V.Gogan :Can We Learn to Beat the Best Stock, Journal of Artificial Intelligence Research 21,pp579-594 (2004)
- [Chang 09] P.Chang, C.Y. Fan, C.H. Liu: Integrating a Poewise Linear Representation Method and a Neural Network Model for Stock Trading Points Prediction, IEEE Transactions on Systems,Man,and Cybernetics-Part C:Applications and reviews,VOL.39,No. 1,January(2009)
- [Yan 08] W.Yan, M.Sewell, C.D.Clack:Learning to Optimize Profits Beats Predicting Returns Comparing Techniques for Fincial Porfolio Optimisation, In GECCO'08: Proc. 10th annual conference on Genetic and evolutionary computation, pages1681-1688. ACM Press, (2008)
- [Patterson 10] S.Patterson : Letting the Machines Decide New Wave of Investment Firms Look to 'Artificial Intelligence' in Trade Decisions, The Wall Street Journal Tuesday, July 13, (2010)
- [Li 12] B.Li, S.C.H.Hoi :On-Line Portfolio Selection:A Survey , arXiv:1212.2129v1 (2012)