

# 個別銘柄の連動類似度グラフおよびグラフ研磨手法を用いた株価予測

Finding a group of stocks to be focused based on the similarity graph approach  
- Effectiveness of graph polishing in detecting a herd -

羽室 行信 \*1      岡田 克彦 \*2  
Yukinobu HAMURO      Katsuhiko OKADA

\*1\*2 関西学院大学 経営戦略研究科  
Institute of Business and Accounting, Kwansei Gakuin University

Based on the assumption that any given stocks with similar characteristics tend to move in tandem in the future, we observe graph similarities of multiple stocks to the stock that suddenly rises; and identify a group of stocks that are expected to chase it in the following days. Our identification method proved to be accurate, allowing us to construct a portfolio that generates abnormal alpha using the standard asset pricing model. Also, we find larger alpha and higher Sharpe ratio when we use the graph polishing method.

## 1. はじめに

個別銘柄の日々の変動には様々な理由が考えられる。その銘柄に関する良いファンダメンタルズ情報が入ったり、その銘柄が属する業界の発展性を予感させる政治的決定がなされたのかもしれない。もしくはファンダメンタルズとは関係のないテクニカル指標に関する動きに多くのトレーダーが反応したのかもしれない。しかし、それらの理由を事前に見出すことは非常に困難である。コンピュータトレードも含めた多くのトレーダーは、様々な要因を勘案して最終的な売買の意思決定をするが、多くの場合、銘柄に関する過去の値動きを事前に確認するであろう。とするならば、地震や政治的な事件などの突発的なイベントを除けば、過去の値動きの中に将来の値動きに影響を与える情報が含まれているはずである。そういった仮定の基でこれまでも様々なテクニカル分析が実施されてきたものの、明確に予測性を備えた手法は開発されていない。

本稿では、これまでには計算量の観点から算出が難しかった銘柄間の関係性変化を求めることで、そこから将来の値動きの予測を試みる。ブラック・マンデーやリーマンショックの様な市場の大崩落のときでさえも、全ての銘柄が一斉に同一時刻に下落するわけではない。ある銘柄の値崩れが連鎖的に市場全体に広がっていく。このような現象は、大崩落のときだけでなく、日常的にも観測できることである。そこで、本稿では、大きな価格変動のあった銘柄と類似した値動きを示していれば、その銘柄も近い将来に大きな価格変動を引き起こすという仮定のもとと推薦システムで用いられる協調フィルタリングと同様の手法を用いることで、価格変動の予測モデルを構築する。協調フィルタリングの優れた点は、過去の実績データが十分なあれば（例えば、通販では顧客の購買行動の実績）、主体間（通販における顧客）の類似度計算だけで推薦（通販における商品）が可能となるという点にある。本稿では、主体を銘柄として、株価が突然上昇した銘柄と類似した銘柄は、近い将来、その株価も上昇するであろうとの仮定のもとと推薦システムを構築する。本稿の特徴は、その類似性を計算する方法にある。

## 2. 手法

### 2.1 銘柄間の類似性判定

日  $t$  における銘柄集合  $N_t$  について、変数  $v$  において類似した銘柄間に枝を張ったグラフ  $G_t^v = (N_t, E_t^v)$  を考える。枝集合  $E_t^v$  は式 (1) で示されるように、任意の二つの銘柄  $a, b \in N_t$  について、日  $t$  から過去  $d$  日間の株価推移の類似度が、与えられた閾値  $\rho$  以上であるような銘柄間に枝を張ることで構成する。

$$E_t^v = \{(a, b) | \text{sim}(\mathbf{x}_{t,d}^{v,a}, \mathbf{x}_{t,d}^{v,b}) \geq \rho; a, b \in N_t\} \quad (1)$$

ここで  $\mathbf{x}_{t,d}^{v,a}$  は、変数  $v$  についての銘柄  $a$  の日  $t-d+1$  から日  $t$  までの数値ベクトルで、 $v$  は表 1 に示された 8 変数のいずれか一つに対応する。また類似度関数  $\text{sim}(\cdot, \cdot)$  としては多様な定義を与えることが可能であるが、本稿ではピアソンの積率相関係数を用い、変数に応じて  $\rho = 0.8$  もしくは  $\rho = 0.9$  を設定した。また、 $d = 10$  に固定して実験をおこなった。ただし、定義上一日前の情報を参照している変数 ( $kerb, barb$ ) については、 $d = 9$  としている。

表 1: 類似度グラフの構築で用いる変数一覧

変数名 ( $v$ )	定義	$\rho$	$d$
$kerb_t$	$o_t/c_{t-1}$	0.9	9
$berb_t$	$c_t/o_t$	0.9	9
$range_t$	$h_t - l_t$	0.8	10
$volume_t$	$v_t$	0.9	10
$kMb_t$	$kerb_t - berb_t$	0.9	10
$shadowU_t$	$(h_t - \max(o_t, c_t))/range_t$	0.8	10
$shadowL_t$	$(\min(o_t, c_t) - l_t)/range_t$	0.8	10
$shadowv_t$	$shadowU_t \cdot shadowD_t$	0.8	10

$o_t$ :始値,  $h_t$ :高値,  $l_t$ :安値,  $c_t$ :終値,  $v_t$ :出来高,  $p_t$ :topix 終値

### 2.2 突然の上昇銘柄の算出

次に株価が突然上昇した銘柄リストの獲得方法について述べる。ある銘柄  $a$  の日  $t$  より  $d$  日前までの終値の変動ベクトルを  $\mathbf{c}_{t,d}^a$  で表す。また、第  $d$  成分のみ 1 で他の成分が 0 であるような  $d$  次元ベクトルを  $\mathbf{u}_{t,d} = (0, 0, \dots, 1)$  と表すと、2 つのベクトル  $\mathbf{c}_{t,d}^a$  と  $\mathbf{u}_{t,d}$  の相関係数が、閾値  $\sigma$  より大きければ、

銘柄  $a$  は日  $t$  に突然上昇したと定義する。そして、日  $t$  に株価が突然上昇した銘柄集合を

$$U_t = \{a | \text{sim}(\mathbf{c}_{t,d}^a, \mathbf{u}_{t,d}) \geq \sigma; a \in N_t\} \quad (2)$$

で表す。以下の実験では  $\sigma = 0.9$  としている。

突然上昇についての上述の定義の有効性を示すために、図 1 に、実データ終値ベクトル  $\mathbf{c}_{t,10}^a$  をランダムに 1 万件サンプリングし、突然上昇ベクトル  $\mathbf{u}_{t,10}$  との相関係数を求め、-0.8 から 0.8 までの 9 水準で平均をとりプロットしている。相関が高くなるに従って、突然上昇の傾向を表していることが分かる。

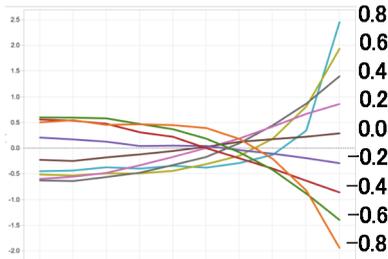


図 1: 突然上昇ベクトルとの相関係数と平均変動の関係。横軸は日  $t$  から日  $t+9$  に対応し、右列の数字は、各折れ線の相関係数である。

### 2.3 銘柄推薦

以上、日  $t$  に突然株価が上昇した銘柄群  $U_t$  と、その日の変数  $v$  の類似度グラフ  $G_t^v$  が与えられた時、日  $t+1$  に購入を推薦する銘柄群  $Q_t^v$  は式 (3) の通り選択される。

$$Q_t^v = \{b | (a, b) \in E_t^v, b \notin U_t\}_{a \in U_t} \quad (3)$$

これは、突然株価が上昇した銘柄  $a$  と変数  $v$  において過去の値動きが似た銘柄  $b$  の買いを推奨するというものである。

ただし、株価動向にはある一定のパターンが存在することが報告されている。[Jiang2015] では、米国において、日  $t-1$  で上昇した株式の中でニュースが存在する株式については、さらに日  $t$  も上昇する傾向 (momentum 傾向) が確認されている。米国株についてはこうした momentum 効果が短期間でも長期間でも報告されているが、一方で日本株には momentum 効果は存在しない。とりわけ短期の株価動向を見た場合、昨日上昇した株式は下落し、下落した株式が上昇するという reversal 傾向が色濃くみられるのである [Tsukioka2017]。このように、momentum や reversal 効果が基調として強い場合、突然上昇する銘柄に類似であるが故に上昇しているのか、単なる momentum / reversal の発現なのかの識別が困難である。そこで、式 (3) の推薦対象銘柄  $b$  に対して、式 (4) に示される通り、突然上昇ベクトルとの相関係数が -0.3 以上 0.3 以下の銘柄のみを対象とする制約を追加する。こうすることで、momentum 効果や reversal 効果に左右されない本手法独自の有効性を示すことが可能となる。以下の実験では、 $\lambda = 0.3$  と設定している。

$$|\text{sim}(\mathbf{c}_{t,d}^b, \mathbf{u}_{t,d})| \leq \lambda \quad (4)$$

### 2.4 類似度グラフの研磨

以上に示した手法の要は、変数  $v$  による類似度グラフ  $G_t^v$  が、銘柄が上昇した理由を示す類似性構造を持っているかどうかにある。前述の通り、類似度グラフは、2 つの銘柄間のみ

の関係性で構築されたものであり、そこにはノイズが入り込むこともありうる。そこで、複数の銘柄の関係性を考慮して、ノイズをクリーニングすることを考える。その方法としてグラフ研磨 (graph polishing) の手法を利用する [Uno14]。グラフ研磨とは、与えられたグラフに対して、部分的な密度に応じて辺を追加したら削除することで、より構造を明確化した新しいグラフを構築する操作のことである。具体的には、グラフ  $G = (V, E)$  について、 $V$  上の任意の節点ペア  $u, v$  について、その類似度  $s(u, v)$  がユーザの指定した閾値以上であれば、 $(u, v) \notin E$  であっても  $u, v$  に辺を追加し、また閾値に満たなければ、 $(u, v) \in E$  であっても辺を接続しないというルールに従って、新たなグラフを再構成する。研磨の類似度としては jaccard 係数、 $s(u, v) = |N(u) \cap N(v)| / |N(u) \cup N(v)|$  を用いている。ここで  $N(u)$  節点  $u$  の隣接節点集合である。また、類似度の閾値は  $\delta = 0.2, 0.3, 0.4$  として実験した。以上のように、提案手法の要である類似度グラフをクリーニングすることで、より精度の高い推薦が可能である。

### 2.5 評価

提案手法の有効性を評価するために、トレーディングシミュレーションを実施する。日  $t$  に推薦された銘柄集合  $U_t$  を日  $t+1$  の終値で購入し、 $h$  日間保有した後 (日  $t+h+1$ ) に売却し、日々の収益率を積み上げていく。ただし、購入数量は 1 日あたり 1 単位固定としている。つまり、日  $t$  の銘柄  $a$  の収益率を  $r_t^a$  で表すと、日  $t+1$  で購入した銘柄群の 1 日後 ( $t+2$ ) の総収益率は  $(1/|U_t|) \sum_{a \in U_t} r_{t+2}^a$  で計算される。なお、実験では  $h = 1, 2, 3, 4, 5$  と設定した。

以上のように、推薦された銘柄によりカレンダーポートフォリオを構成し、取引シミュレーションを実施した結果に対して、Lyon らの推奨する Fama-French Three Factor Model [Lyon1999] を用いて評価する。Three Factor Model では、以下の推定式が用いられる。

$$R_t^p - R_t^f = \alpha + \beta(R_t^m - R_t^f) + s(S_t) + h(H_t) + \epsilon_t \quad (5)$$

ここで  $R_t^p$  は日  $t$  のカレンダータイム・ポートフォリオの単純平均収益率、 $R_t^f$  は 10 年新発国債の利回り、 $R_t^m$  は東京証券取引所第 1 部、第 2 部の金融を含む全上場企業の時価総額加重平均リターン、 $S_t$  は小型株と大型株の時価総額加重平均収益率の差、 $H_t$  は成長株とバリュー株の時価総額加重平均リターンの差である。これらの重回帰結果として得られる  $\alpha$  の  $p$  値を評価値として用いる。この  $p$  値が低いということは、提案手法で構成したポートフォリオが、小型株、割安株への偏りをコントロールした後も超過リターンがあることを意味する。

Three Factor Model 以外にも、平均収益率やその標準偏差、Sharpe 比、max drawdown についても示す。Sharpe 比は、年あたりの収益率平均を収益率標準偏差で除した値として定義される。Sharpe 比とは、日々の収益率 (取引のない日の収益率は 0 とする) の標準偏差に対する比によって定義され (一般的には、日ではなく年率に換算して計算する)、リスクで基準化した平均的な業績を評価できる。また max drawdown とは累積収益率の最も大きい落ち幅のことで、実際の投資におけるリスクを評価する指標として用いられる。

### 3. 実験

実験では、トムソン・ロイターの Datastream より取得した米国の株価データを用いる。米国を対象とした理由は、提案手

法では、できるだけ活発に取引される銘柄で、できるだけ多く必要とする。そこで、全上場銘柄の時価総額上位 20%を対象にしたところ、米国で平均 3000 銘柄ほどで、日本や他の国では数百を下回る数字であったため今回の実験の対象とはしなかった。対象期間は 2000 年～2016 年 11 月の 17 年間弱、4234 営業日のデータである。

### 3.1 有意なモデル

表 2 は、全 8 変数 (表 1) と保有日数  $h = 1, 2, 3, 4, 5$  の 40 通りのシミュレーションを実施した結果、Three Factor Model の  $\alpha$  の p 値が 5%以下となった 9 つのモデルを p 値の小さい順に示している。 $v=\text{range}$  の保有日数違いのモデルが上位にきているが、これは、高値と低値の幅、すなわち 1 日の取引の中での価格変動幅の傾向が似ていれば、その銘柄も突然上昇する可能性が高いことを示唆している。その他の変数も含めて、全ての結果に対して意味付けをすることは困難ではあるが、多くのトレーダーが共通して利用している株価四本値と出来高の過去 10 日間の変動類似性に、将来の購買行動に結びつく情報が含まれていることが示唆される。

また、表 2 の最下行には、これら 9 つのモデルを統合した結果が示されている。延べ取引銘柄数や収益率は、9 つのモデルの数の合計になっている。もし 9 つのモデルが同じ情報を提供しているとすると、Sharpe 比も 9 つのモデルの平均に近くなり、max draw down はその合計になるはずであるが、Sharpe 比はいずれにモデルよりも高く、draw down はその合計値 (537.3) の半分程度である。これは、それぞれのモデルが上昇銘柄が示す類似性の異なる局面を捉えていると示唆されよう。

図 2 は、これら 9 つのモデルのシミュレーション結果 (左) と統合モデルの結果 (右) を示しており、9 つのモデルを統合することで、お互いの変動を打ち消しあって、より安定したカーブを描いていることが視認できる。なお、統合モデルで最も draw down が大きいのは、リーマンショックの時期であり、この大崩落を long 戦略 (値上がり期待できる株を買い、後に売り抜ける戦略) で、この程度の draw down で済んでいるのは逆にモデルの有効性を示していると考えてよいであろう。

### 3.2 グラフ研磨の有効性

次に、前節で求めた 9 つのモデルで用いたそれぞれの類似度グラフをグラフ研磨することで、統合モデルがどのように変化するかを表 3 に示している。研磨のパラメータである類似度の下限値  $\delta$  によって結果は異なるが、 $\delta = 0.4$  において、取引銘柄数は半分程度に押さえながらも、Sharpe 比が最大で draw down も最小になっており、グラフ研磨の有効性が認められる。しかしながらその差は大きくなく、統計的な検証が必要となるであろう。

また、グラフ研磨の効果が現れやすい変数とそうでない変数があるとも考えられる。そこで、前節で求めた 9 つのモデルに限定するのではなく、各モデル (変数  $\times$  保有日数) の中で最も p 値を低める  $\delta$  を求め、そこから前節同様に p 値 5% の条件でモデルを選択した。表 4 がその一覧である。表 2 と見比べると、2 つのモデルが追加されており、また最適な研磨パラメータ  $\delta$  はモデルにより異なることが分かる。また、これら 8 つの統合モデルの Sharpe 比は、表 3 に示されたいずれの統合モデルよりも高く (1.14)、グラフ研磨の有効性を伺うことができる。ただし、draw down についてはいずれの統合モデルよりも大きくなっており、グラフ研磨が全てにおいて優れた結果を生み出すわけではないことを確認した。

## 4. おわりに

過去の値動きに関する類似度グラフを用いた独自の推薦システムを提案し、銘柄の突然上昇を予測するモデルを構築した。米国の 2000 年から 2016 年の 17 年間のデータで検証し、その有効性について示した。また類似度グラフのノイズ除去としてのグラフ研磨手法を適用し、その有効性についても検証した。今後は、値動き以外の要因 (例えば産業属性やニュースへの exposure など) も考慮したモデルに拡張するとともに、全世界の上場銘柄を対象とした、より大規模なデータでの検証を行うことで一般性を高めたい。

## 謝辞

本研究は、JST CREST、科研費基盤研究 (B) 25285127、及び、統計数理研究所平成 27 年度公募型共同研究「重点研究 28-共研-4101」の研究助成を受けている。

## 参考文献

- [Hamuro2016] 羽室行信, 岡田克彦, Cheung Stephane, 銘柄類似度グラフの時系列構造変化に基づく株価予測, 2016 年度人工知能学会 (第 30 回), 北九州国際会議場, 2016 年 6 月.
- [Jiang2015] Jiang, H., S.Z. Li and H. Wang, 2015 “News Momentum”, working paper.
- [Lyon1999] Lyon, J., D., Brad M. Barber, Chih-Ling Tsai, “Improved Methods for Tests of Long-Run Abnormal Stock Returns”, *The Journal of Finance*, 54, 1, pp. 165–201, 1999.
- [Tsukioka2017] 月岡靖智, 山崎高広, 倉井龍太郎, 岡田克彦, ニュースモメンタム、それともニュース無しのリバーサル?, 2017 年度人工知能学会 (第 31 回), 名古屋市ウイックあいち, 2017 年 6 月.
- [Uno14] 宇野毅明, 中原孝信, 前川浩基, 羽室行信「データ研磨によるクリーク列挙クラスタリング」情報処理学会アルゴリズム研究会報告書, 2014-AL-146(2), pp. 1-8, 2014.

表 2: 有意なモデル一覧。Frama-French の 3 factor モデルによる  $\alpha$  の p 値が 5%以下のモデル (変数 × 保有日数) を選択することで得られた 9 つのモデル。最下行の「統合モデル」は、9 つのモデルを全て用いてシミュレーションした結果を示している。

変数	保有日数	延べ取引 銘柄数	$\alpha$ (%)	$\alpha$ の p 値 (%)	収益率 (年率%)	SD (年率%)	shape 比	max draw down		
								下落率 (%)	開始日	終了日
range	2	123,094	0.059	0.09	19.1	24.0	0.79	-60.3	2000/07/07	2000/10/17
range	4	36,292	0.056	1.05	18.9	27.3	0.69	-46.1	2000/06/09	2000/09/21
kerb	4	11,470	0.070	1.23	21.2	31.1	0.68	-69.2	2000/04/07	2000/05/17
kMb	2	14,603	0.065	2.29	19.5	30.8	0.63	-45.3	2006/01/26	2007/01/30
shadowL	1	89,673	0.037	2.79	13.3	23.2	0.57	-51.4	2008/04/21	2009/01/15
volume	5	112,152	0.027	4.21	9.8	18.9	0.52	-69.9	2007/10/12	2009/03/09
shadow	4	22,431	0.041	4.42	14.8	26.4	0.56	-60.0	2008/11/03	2009/03/03
shadowL	3	9,984	0.048	4.60	15.8	28.4	0.55	-68.1	2000/03/23	2000/07/21
volume	1	182,722	0.031	4.78	11.4	20.4	0.56	-67.0	2000/03/23	2001/10/09
統合モデル		602,421	0.037	2.79	144.0	146.7	0.98	-279.7	2008/06/05	2008/10/10

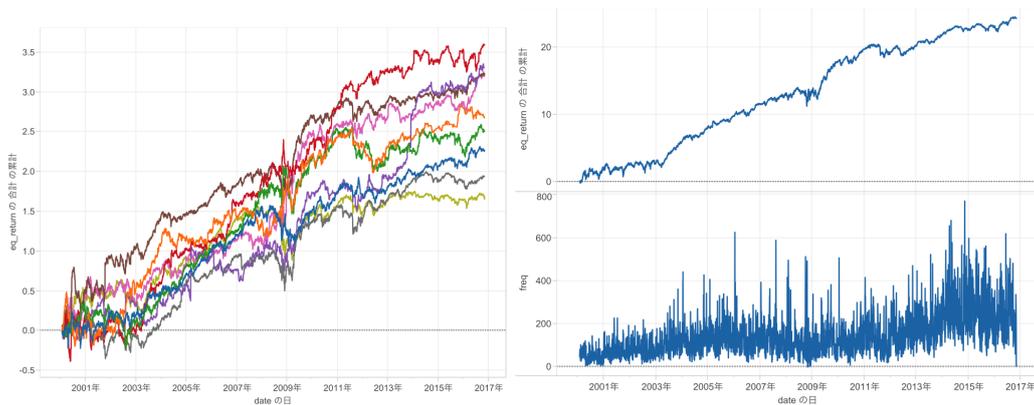


図 2: 有意なモデルによる取引シミュレーションの結果得られた累積収益率の推移。左図は、表 2 に示された 9 つのモデルを個別に実行した結果を示し、右図は、統合モデルの結果を示している。右図の下は、銘柄のポジション数 (保有銘柄数) を示している。9 つのモデルを統合することで、個々の変動が相殺され、より安定したカーブになっていることが視認できる。

表 3: 類似度グラフのグラフ研磨の効果。表 2 に示された有効な 9 つのモデルについて、類似度グラフを polishTH で示されたパラメータでグラフ研磨した場合の結果。一行目は、比較のため表 2 に示された統合モデルの結果を示している。

polishTH $\delta$	延べ取引 銘柄数	$\alpha$ (%)	$\alpha$ の p 値 (%)	収益率 (年率%)	SD (年率%)	shape 比	max draw down		
							下落率 (%)	開始日	終了日
-	602,421	0.037	2.79	144.0	146.7	0.98	-279.7	2008/06/05	2008/10/10
0.2	581,182	0.029	15.22	124.5	144.0	0.86	-264.0	2008/06/05	2009/03/09
0.3	435,285	0.092	0.00	133.2	133.1	1.00	-281.0	2008/06/05	2008/10/10
0.4	346,922	0.054	1.24	120.2	115.4	1.04	-263.0	2008/06/05	2008/10/10

表 4: 最適なグラフ研磨パラメータで選択した場合の有意なモデル一覧。各モデル (変数 × 保有日数) について、最も  $\alpha$  の p 値が低くなるようなグラフ研磨パラメータ  $\delta$  を選択し、全モデルから  $\alpha$  の p 値が 5%以下のモデルを選択した結果を示している。

変数	保有日数	最適な研磨 パラメータ ( $\delta$ )	延べ取引 銘柄数	$\alpha$ (%)	$\alpha$ の p 値 (%)	収益率 (年率%)	SD (年率%)	shape 比	max draw down		
									下落率 (%)	開始日	終了日
shadowL	1	0.3	22,627	0.092	0.00	263.8	269.6	0.97	-52.7	2008/07/08	2008/10/06
volume	1	0.4	148,379	0.051	0.47	160.7	219.2	0.73	-57.3	2007/01/24	2008/11/20
shadow	4	0.3	11,229	0.063	1.05	190.5	286.0	0.66	-50.3	2008/11/05	2009/04/07
shadow	5	0.4	17,202	0.053	1.27	181.3	262.2	0.69	-59.3	2008/07/22	2009/03/09
range	2	0.2	115,323	0.045	2.39	155.9	256.5	0.60	-59.6	2008/06/19	2008/10/10
volume	2	0.2	140,790	0.035	3.22	124.9	212.1	0.58	-89.5	2007/10/12	2008/10/10
kMb	2	0.2	14,771	0.064	4.21	188.8	335.2	0.56	-61.5	2006/01/26	2008/12/19
kerb	4	0.3	9,362	0.052	4.91	159.6	288.3	0.55	-72.5	2000/04/07	2001/09/20
統合モデル			479,683	0.035	3.22	142.5	124.6	1.14	-300.8	2008/06/05	2008/10/10