

銘柄類似度グラフの時系列構造変化に基づく株価予測

Prediction of Stock Market Index based on Structural Change on Similarity Graphs of Individual Stock Price Movement

羽室 行信^{*1} 岡田 克彦^{*2} Stephane Cheung^{*3}
Yukinobu HAMURO Katsuhiko OKADA Stephane CHEUNG

^{*1*2*3}関西学院大学 経営戦略研究科

Institute of Business and Accounting, Kwansei Gakuin University

This work presents the effectiveness using graph features in predicting the stock market movement. Specifically, we used return correlations among listed firms and expressed the market co-movement with edge density. We found the rising time series edge density corresponds to the excess market movement in the downside. We also demonstrate profitable trading strategies to long TOPIX when the edge density shows it is in the oversold condition.

1. はじめに

株式市場は日々変動するものであるが、稀に「セリング・クライマックス」という状態に陥る。これは、投資家心理が過度に悲観に傾いて、不確実性が高い銘柄だけではなく、ファンダメンタルズのしっかりした優良銘柄にも売り注文が浴びせられ、相場が陰の極みとなる状態である。このような局面においては、将来の先行きに対する大きな不安が広がっており、それを裏打ちするようなニュースがマーケットを席卷していることが常である。しかし、どれほど悲観的な相場環境にあったとしても、ファンダメンタルズ価値を大きく下回る水準で購入すれば、その後大きなリターンが見込める。問題は、「セリング・クライマックス」を客観的に示す指標がなく、投資家は事後的に認識するのみだという点である。

本稿では、こうした「セリング・クライマックス」を定量的に表現する手法を提案する。パニックに陥った投資家は、本来売つつもりがなかった銘柄まで売却してしまうため、その時点で銘柄間の相関が高まる。ある閾値を超えて銘柄間の相関が高まる時がファンダメンタルズ価値より下方乖離するタイミングだと認識し、投資機会と判断した。この手法をもちいて相場の大底で株式を購入する投資戦略をシミュレーションした結果、あらゆるケースでプラスのシャープレシオを達成した。

2. 手法

日 t における銘柄集合 N_t を構成する全銘柄間の関係性をグラフ $G_t = (N_t, E_t)$ で表現することを考える。枝集合は式 (1) で示される通り、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 ρ 以上であれば枝を張る。

$$E_t = \{(a, b) | \text{sim}(\mathbf{c}_a(t, d), \mathbf{c}_b(t, d)) \geq \rho; a, b \in N_t\} \quad (1)$$

ここで $\mathbf{c}_a(t, d)$ は、銘柄 a の日 $t-d+1$ から日 t までの終値ベクトルを表している。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る。全節点ペアに対する相関係数の計算には $|N_t|^2$ の計算コストが必要となるが、本研究では SketchSort 法を用いることで計算効率を高めている。

連絡先: 羽室行信, 関西学院大学, 西宮市上ヶ原一番町 1-155, TEL:0798-54-6572, hamuro@kwansei.ac.jp

田部井らにより開発された SketchSort 法 [Tabei10] は、与えられたベクトルデータ集合の中の 2 点間の距離が一定の閾値内のデータペアを効率的に全列挙する手法である。この手法は、閾値内のデータペアを見逃す確率を一定の値以下に抑える近似的アルゴリズムを用いることで高速化を達成している。計算コストは、節点サイズと出力される枝数に線形である。

以上の方法を全データ期間に適用することで、市場内で生じている全銘柄の関係性の変化をグラフの構造変化として表現することが可能となり、TOPIX のような市場インデックスでは捉えることのできなかった市場の動きを検知できることが期待できる。グラフの構造変化を捉える方法としては、日 t に生成されるグラフ G_t から特徴量 f_t を抽出し、それを時系列データ (f_1, f_2, \dots, f_n) として構成することで実現する。グラフ特徴量としては、近年様々な特徴量が提案されており、枝数、枝密度、平均次数、クラスタ係数、直径、中心性など様々な特徴量を利用できる。グラフ全体の構造 (市場全体の動き) を捉えることもできれば、節点の特徴 (単一銘柄の動き) を捉えることも可能である。

しかしながら特徴量によっては、どのような構造変化を捉えているのか、その意味解釈が困難になることもあり、本研究では、グラフの枝密度に注目した。枝密度 (density) は、式 (2) で定義されように、グラフ G_t が完全グラフであった場合の枝数に対する実際の枝数の割合として定義される。

$$\frac{|E_t|}{(|N_t|(|N_t| - 1)/2)} \quad (2)$$

枝密度が高くなるということは、多くの銘柄が同じような株価変動を示していることを意味する。ただし、類似度として相関係数を用いているので、株価の上昇局面で類似しているのか下降局面で類似しているのか、もしくは乱高下している局面で類似しているのかは区別できない。そこで、本稿では、式 (1) に示された類似度の条件に加え、式 (3,4) に示される制約を加えることで、局面を考慮した関係性グラフを構成する。

$$\text{sim}(\mathbf{c}_a(t, d), \mathbf{c}(d)) \geq \tau \quad (3)$$

$$\text{sim}(\mathbf{c}_b(t, d), \mathbf{c}(d)) \geq \tau \quad (4)$$

ここで $\mathbf{c}(d)$ はユーザにより指定されるサイズ d の定数ベクトルである。例えば、

$$\mathbf{c}(d) = (1, 2, \dots, d) \quad (5)$$

とすることで株価の上昇局面を表現でき、

$$c(d) = (d, d - 1, \dots, 1) \quad (6)$$

とすることで下降局面を表現できる。

3. 実験結果

対象銘柄集合 N_t として、日本の主な証券取引所に上場する全銘柄を対象とし、1987年から2014年までの28年間(一年の営業日数 250日 \times 28年 = 約7000日)のデータを用いた。類似度の条件としては、株価変動ベクトルのサイズ(日数) $d = 10$ 、相関係数の下限値 $\rho = 0.9$ とした。また局面 $c(d)$ については、式(5)と式(6)で示された上昇局面と下降局面の両方についてそれぞれ関係性グラフを構成した。またその関係性の強さは $\tau = 0.7$ とした。

図3に、下降局面における枝密度とTOPIXの推移を示す。また枝密度が0.2を超える日とその要因と推察される出来事を図中及び表1に示している。枝密度が0.2ということは、約4000銘柄の全ペア約8百万ペアのうち160万ペアが下降局面で似た株価変動を示していることを意味する。震災のような自然災害、リーマン・ショック、バブル崩壊のような経済界における大きな出来事など、その要因が比較的明らかな出来事もあれば、明確な要因がなく複数の要因が「複合」して生じていると考えられるケースも見られる。

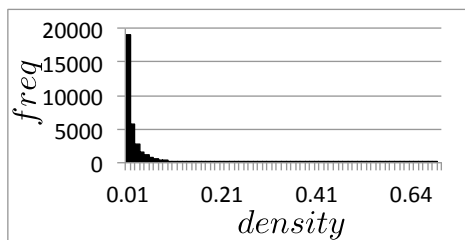


図1: 枝密度のヒストグラム

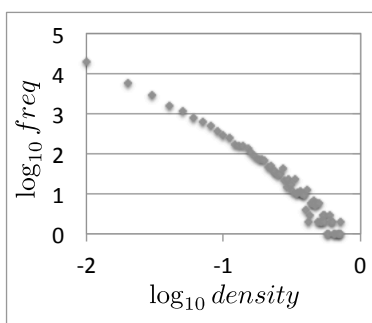


図2: log スケールで示された枝密度とその頻度の関係

枝密度とその頻度の分布は図1,2に示されるようにベキ分布に従っており、0.2を超えるような大きなピークは稀にしか出現しない。金融の世界では、株式リターンは正規分布を仮定して理論構築がされており、一般的なリスク管理手法のVaR(Value at Risk)なども、正規分布の前提が崩れれば機能しない。金融危機時に多くの金融機関がリスク管理に失敗した背景に、マーケットの動向がベキ分布であることが原因だとする見方も存在した。高い枝密度の出現パターンがベキ分布すること

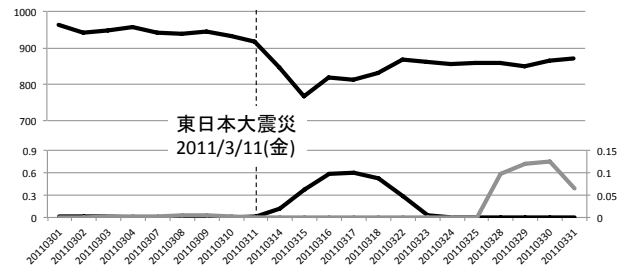


図4: 東日本大震災前後の動き。上の実線はTOPIXの変動を、下の実線は下降局面における枝密度を、そしてグレーの線は上昇局面における枝密度の変動を示している。また破線は震災が起こった日を示している。

表1: 枝密度が0.2を超えた日の出来事

出来事	日付	ピーク日付	枝密度
東日本大震災	2011/03/11	2011/03/17	0.592
バブル崩壊	1991/03 - 1993/05	1990/10/01	0.479
リーマン・ショック	2008/09/15	2008/10/10	0.472
複合		2014/02/05	0.321
ライブドア事件	2006/01/16	2006/02/20	0.270
複合		2014/10/17	0.261
日銀ゼロ金利政策解除	2006/07/14	2006/07/19	0.252
ギリシャ危機(格下げ)	2011/07/25	2011/08/09	0.235
衆議院解散自民党分裂	1993/06/18	1993/06/22	0.228
ブラックマンデー	1987/10/17	1987/10/26	0.226
サブプライム問題	2007/08 -	2008/01/17	0.215
複合		1993/11/30	0.213
複合		2004/05/17	0.202

は、こうした見方と整合的である。本稿の結果は、今後の金融リスク管理手法の開発について有益な示唆を与えるものだと考えられる。

次に、個々のケースについてより詳細に見てみる。図4は東日本大震災前後の枝密度とTOPIXの変動を示しており、図5はリーマン・ショックにおける変動を示している。震災のような誰の目からも明らかな出来事については市場も直ぐに反応するが、リーマン・ブラザーズ破綻のようにその影響が即座に判断のつかない出来事については市場は直ぐには反応しない。

図3,4,5から得られる仮説は、枝密度のピークが市場インデックス(TOPIX)の大底を検知しているのではないかというものである。

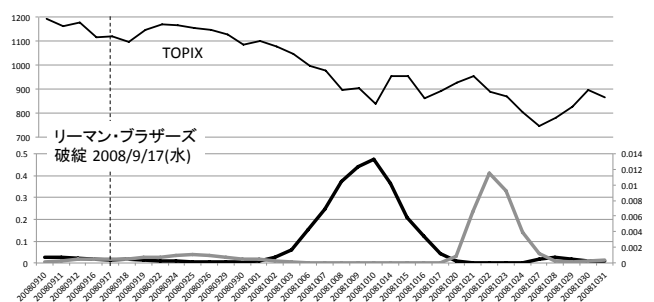


図5: リーマン・ショック前後の動き

4. TOPIX 大底検知

枝密度というグラフ特徴量の変化を見ることで市場インデックスの大底(下降から上昇への転換点)を検知できるとすれば、

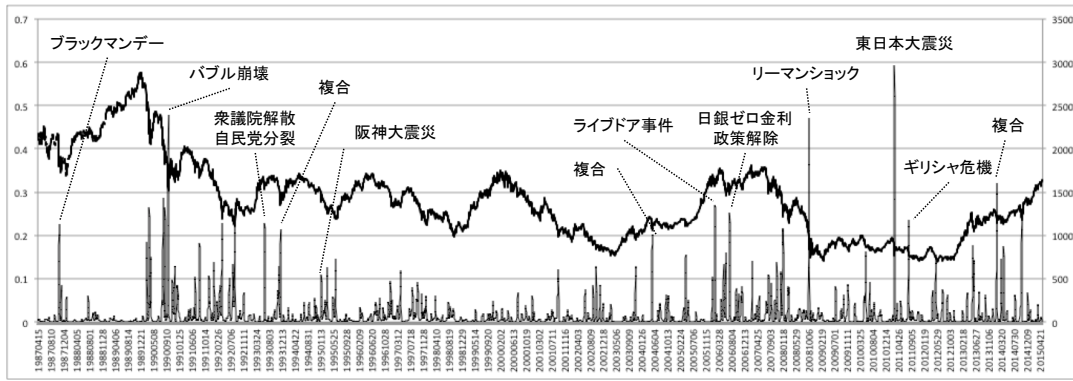


図 3: 1987 年から 2014 年までの枝密度と TOPIX のトレンド

表 2: 枝密度のピーク前後の収益率平均と標準偏差

ピーク 前後日数	0.05 ~ 0.1	0.1 ~ 0.15	0.15 ~ 0.2	0.2 ~ 0.3	0.3 ~
平均					
-20	-0.056 [#]	-0.077 [#]	-0.095 [#]	-0.113 [#]	-0.189 [#]
-10	-0.053 [#]	-0.073 [#]	-0.088 [#]	-0.091 [#]	-0.166 [#]
-5	-0.030 [#]	-0.039 [#]	-0.049 [#]	-0.050 [#]	-0.124 [#]
-1	-0.003 [*]	-0.003	-0.004	-0.002	-0.001
+1	0.002	-0.002	0.000	0.005	0.018
+5	0.001	0.001	0.003	0.017 [*]	0.036 [*]
+10	0.005 [*]	0.001	0.001	0.024 [*]	0.013
+20	0.006	0.022 ^{**}	0.030 ^{**}	0.038 [*]	0.068 [*]
#:P 値 <0.001, **:P 値 <0.01, *:P 値 <0.05, ;:P 値 <0.1					
標準偏差					
-20	0.050	0.048	0.048	0.065	0.073
-10	0.028	0.034	0.030	0.038	0.073
-5	0.030	0.034	0.035	0.033	0.073
-1	0.020	0.026	0.030	0.027	0.078
+1	0.020	0.024	0.025	0.035	0.065
+5	0.042	0.038	0.041	0.049	0.067
+10	0.051	0.047	0.053	0.091	0.117
+20	0.072	0.068	0.069	0.114	0.092
件数	308	110	42	44	13

その意義は大きいであろう。そこで本節では、枝密度による大底検知の可能性について検討する。

表 2 は、枝密度のピーク前後における TOPIX の収益率平均および標準偏差を示している。枝密度が上昇から下降に転じた日を起点として、その前後 1,5,10,20 日の収益率を、ピーク時の枝密度の水準別に示したものである。ピーク前の収益率は、日数と枝密度の水準に関わらず有意な収益率の下落が認められる。一方でピーク後の収益率は有意な結果は少ないが、概ねプラスの収益率が認められ、枝密度が市場インデックスの大底を示していると言えよう。またピーク時の前と後に関わらず、日数が経過するにしたがって分散は大きくなる傾向にあるが、これは対象のイベント以外の要因が関わってくるためと推察される。

次に、以上の結果を踏まえて、実際の投資戦略を構成することで、枝密度による大底検知の実践的効果について考察する。ここでは 2 つの投資戦略を想定する (図 6)。戦略 A は枝密度のピークによってピンポイントで TOPIX の大底を推定することで、最安値の時のみ購入しようとする作戦である。具体的には、枝密度がある閾値 σ を超えた後にピークを超えた日 (ピークを超えたかどうかはピークの翌日にしか判断できない) に TOPIX を購入し、再び閾値を下回った時に売却する。一方で戦略 B は、ドル・コスト平均法と呼ばれるもので、枝密度のピークが TOPIX の大底を数日の誤差で外すこともあるとの前提で、そのリスクを回避するために、ピーク前後の一定期間に定額で TOPIX を購入し続ける作戦である。具体的

表 3: 投資戦略 A,B の成績比較。mean はカレンダータイムポートフォリオにおける収益率日平均、SD はその標準偏差を示している。Sharpe は mean/SD で計算される。

σ	戦略 A (ピンポイント)			戦略 B (ドル・コスト法)		
	mean	SD	Sharpe	mean	SD	Sharpe
0.05	0.0027	0.020	0.136	0.0031	0.024	0.129
0.1	0.0018	0.024	0.075	0.0038	0.030	0.129
0.15	0.0040	0.027	0.146	0.0069	0.036	0.192
0.2	0.0014	0.028	0.051	0.0083	0.041	0.202
0.3	0.0083	0.025	0.325	0.0223	0.053	0.417

には、枝密度がある閾値 σ を超えた後、再び閾値を下回るまで毎日 TOPIX を買い増していき、閾値を下回った時点で全て売却する。

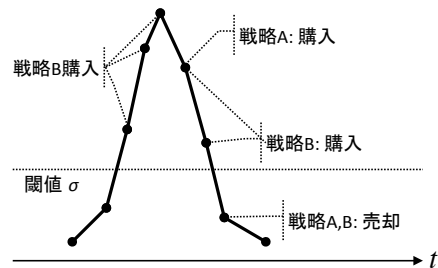


図 6: 大底検知に基づいた 2 つの戦略

戦略 A,B に従って 28 年間運用した結果を表 3 に示す。戦略 A,B とともに収益率の日平均はプラスを示しており、大底検知の戦略が機能していることが分かる。また戦略 B は A に比べてより高い収益率を達成できている一方で、バラツキも大きくなっている。これは、「セリング・クライマックス」が来る前から買い下がる行為をすることで、本当の大底が来た時には既に買いポジションをもってしまっていることに原因があると考えられる。相場の格言に「落ちてくるナイフを掴もうとするな (Don't try to catch a falling knife)」というものがあるが、まさに戦略 B はまさに相場の禁じ手を実践していることになり、実りも大きい (高い収益率) がリスク (バラツキ) も高いということであろう。

5. おわりに

日本の全上場銘柄について、過去の株価変動の類似度に基づいた類似度グラフを構成し、その時系列構造変化と TOPIX インデックスの関係について考察してきた。枝密度という基本的なグラフ特徴量が株価の大幅下落時の転換点の検知に有効であることを示した。

本稿で示した、銘柄の関係性の構造変化の解析は、ファイナンス分野においては初めての試みであり、その応用可能性は非常に高いと考えられる。とりわけ、本稿で扱ったグラフ特徴量としての枝密度の分布がべき乗分布になることは、金融危機時に議論されたマーケットを平均と分散でリスク管理することの危険性、すなわちマーケットがべき分布しているという議論と整合的なものであり、今後の研究の方向性に有意義な示唆を与えてくれている。また、枝密度以外の多様な特徴量を用いることで、本稿でも紹介したリーマン・ブラザーズ倒産から本格的に株価が暴落するまでの期間において、その予兆を検知できるかもしれない。

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「一般研究 1-27-共研-1031」の研究助成を受けている。

参考文献

[Tabei10] Yasuo Tabei, Takeaki Uno, Masashi Sugiyama, Koji Tsuda: Single Versus Multiple Sorting in All Pairs Similarity Search, The 2nd Asian Conference on Machine Learning (ACML2010), Tokyo, Japan, 2010.