

## 企業の株価データを用いた業種ごとの同類度分析

楊 小龍† 沖本 まどか† 上野 雄史† 齊藤 和巳†

†静岡県立大学 経営情報学部

## 1 はじめに

企業の株価は、互いに関係し影響を与えつつ変動する。よって、株価変動の類似性を相関係数などで求めて分析する研究が展開されている [1, 2]。その端緒となる代表研究として、Mantegna [1] は、ニューヨーク証券取引所の 1989 年 7 月から 1995 年 10 月の期間での S & P500 及びダウ工業株 30 種平均に用いられる企業の株価 (Standard & Poor's 500 Stock Index) を用いた実験により、同一業種の企業  $u$  と企業  $v$  のペアの株価変動の同類性とともに、類似度には階層性が見られることを示した。

しかしながら、Mantegna の実験結果は、最小全域木とデンドログラムの可視化に基づく定性分析である。言い換えれば、多様な市場比較評価などに向けては、妥当な定量評価指標が必要となる。そこで我々は、Newman が提案したネットワーク上でのノード属性の同類係数 (assortative mixing coefficients) の概念 [3] に基づき、同業種の株価変動における同類性の定量評価が可能であることを、日本の証券取引所を始めとし、イギリスやフランスなど複数国の株価データで実証的に示してきた [4, 5]。本稿では、これらの方法を発展させ、相対的な同類度の高さで業種をランキングする評価指標を提案する。実験では、17 または 33 業種分類の日本株価データを用い、提案法の有効性を検証する。

## 2 提案手法

企業間ネットワークを  $G = (V, E)$  とし、企業を  $v \in V$ 、企業  $u$  から  $v$  へのリンクを  $(u, v) \in E \subset V \times V$  とする。なお本稿では、Mantegna 法で構築される最小全域木を  $G$  として用いるが、本提案法は、任意のネットワーク  $G$  に適用可能である。業種数を  $N$  とし、各業種を  $\{1, \dots, N\}$  の整数と同一視する。企業  $v$  の業種を  $f(v) \in \{1, \dots, N\}$  とすれば、業種  $i$  の企業から業種  $j$  の企業へのリンク数は  $m(i, j) = |\{(u, v) \in E; f(u) = i, f(v) = j\}|$  となり、このリンク数  $m(i, j)$  を  $i, j$  成分とする  $N \times N$  混合行列 (Mixing Matrix) を構成できる。ここで、 $|\cdot|$  は集合の要素数を表す。いま、業種  $i$  からのリンク数を

$a(i) = \sum_{j=1}^N M(i, j)$  とし、業種  $j$  へのリンク数を  $b(j) = \sum_{i=1}^N M(i, j)$  とすれば、Newman の同類係数 (assortative mixing coefficients) [3] より同類度は次式で定義される。

$$r = \frac{\sum_{i=1}^N m(i, i) - \sum_{i=1}^N a(i)b(i)/|E|}{|E| - \sum_{i=1}^N a(i)b(i)/|E|}. \quad (1)$$

ここで、全企業が同業種へリンクすれば  $\sum_{i=1}^N M(i, i) = |E|$  となり、同類度は  $r = 1$  となることは容易に確認できる。

式 (1) は、ネットワーク  $G$  全体に対し求まる指標であるが、一般に、同類性が相対的に高い業種もあれば、逆に低い業種も存在し得る。以下では、業種  $i$  毎に同類度を定義するために、機械学習分野などで標準的に適用される OAA (One-Against-All) の考え方を採用し、業種  $i$  の同類度  $r(i)$  を次式で定義する。

$$r(i) = \frac{m(i, i) - a(i)b(i)/|E|}{(a(i) + b(i))/2 - a(i)b(i)/|E|}. \quad (2)$$

式 (2) は、業種  $i$  とこれを除く業種群を一つの業種と見なし、式 (1) を適用することで容易に導ける。ここで、業種  $i$  の企業が同業種  $i$  のみリンクすれば  $a(i) = b(i) = m(i, i)$  となり、業種  $i$  の同類度は  $r(i) = 1$  となることは容易に確認できる。また、本稿で対象とする Mantegna 法で構築される最小全域木のように、リンクの方向性を考えない無向ネットワークでは、 $a(i) = b(i)$  となり、式 (2) は以下のように単純化される。

$$r(i) = \frac{m(i, i) - a(i)^2/|E|}{a(i) - a(i)^2/|E|}. \quad (3)$$

以下は、式 (3) に基づく結果を報告する。

## 3 評価実験

本実験では、2000 年 1 月 4 日から 2009 年 12 月 30 日までの期間で、東証 1 部上場の株価データを用いた。ただし、この 10 年間のすべての営業日で株価の終値が得られた企業に限定したので、分析対象の企業数は 866 となった。なお、この期間の営業日の総数は 2,457 となった。既に述べているように、Mantegna 法 [1] に最小全域木を構築しネットワーク  $G$  とした。

まず、17 業種分類と 33 業種分類のそれぞれを用いて、式 (1) により求めた同類度は、それぞれ 0.487 と 0.451 となった。これらより、業種階層のレベルを変化させても、同類度は定量的に高いことを確認できる。

Assortativity Analyses of Industrial Sectors based on Stock Market Data  
†Xiaolong YANG †Madoka OKIMOTO †Takefumi UENO  
†Kazumi SAITO  
†School of Management and Information, University of Shizuoka

表 1: 17 業種分類での業種同類度

ランク	業種名	同類度	企業数
1	電力・ガス	.922	26
2	運輸・物流	.757	73
3	鉄鋼・非鉄	.720	71
4	不動産	.684	29
5	銀行	.669	121
⋮	⋮	⋮	⋮
13	エネルギー資源	.398	5
14	商社・卸売	.358	91
15	機械	.349	142
16	小売	.298	44
17	金融（除く銀行）	.119	155

表 2: 33 業種分類での業種同類度

ランク	業種名	同類度	企業数
1	電力・ガス	.922	26
2	海運	.908	11
3	鉄鋼	.743	40
4	不動産	.684	29
5	銀行	.669	121
⋮	⋮	⋮	⋮
28	鉱業	-.001	2
29	空運	-.001	2
30	水産・農林	-.003	4
31	証券、商品先物取引	-.003	139
32	精密機器	-.011	17

表 1 には、17 業種分類での業種同類度のランキング結果を示す。表では、同類度の高さでのランク、そのランクの業種名と同類度、及び、その業種に属す企業数を示している。実験結果より、同類度が高い業種として、電力・ガス業から銀行業までなどが抽出され、同じ業種の企業は類似した株価変動となることを示している。具体的には、地域毎に分割して事業を展開する電力・ガス業で特に高い同類度となることは直感的にも妥当な結果と考えられる。一方、同類度が低い業種として、エネルギー資源業から金融（除く銀行）業までなどが抽出され、これら業種では、同じ業種の企業でも類似した株価変動とならないことを示している。

表 2 には、33 業種分類での業種の同類度ランキング結果を示す。なお、本実験で利用したデータには、33 業種分類での保険業には属す企業が存在しなかったため、ランクは 32 までとなっている。実験結果より、同類度が高い業種については、電力・ガス業や銀行業などがランクされ、17 業種分類と類似した結果になった。詳細には、運輸・物流業と鉄鋼・非鉄業では、海運業と鉄鋼業のそれぞれで同類度が特に高いことが示された。一方、同類度が低い業種については、業種タイプが大幅に変わる傾向とともに、同類度も殆ど 0 まで小さくなったが、その中では、金融（除く銀行）業では、証券/商品先物取引業同類度が特に低いことが示された。また、鉱業業、空運業、及び、水産・農林業では、属す企業数が特に少ないが、証券、商品先物取引においては、ある程度の企業数での結果となっている。これらの結果より、17 または 33 業種分類の日本株価データを用い、直感的にも妥当な業種ランキングを得られたことより、提案法の有効性を検証できたと考える。

#### 4 おわりに

Newman が提案したネットワーク上でのノード属性の同類係数 (assortative mixing) の概念 [3] を発展させ、同類度の高さで業種をランキングする手法を提案した。実験では、17 または 33 業種分類の日本株価データを用い、直感的にも妥当な業種ランキングを得られたことより、提案法の有効性を検証した。今後は、日本だけでなく別の市場の株価データを用い、最小全域木とは異なるネットワーク構築法を採用するなどし、多様なデータにおいて提案法の有効性を検証する予定である。

謝辞 本研究は科学研究費補助金基盤研究 (C)(No. 25330635) の補助を受けた。

#### 参考文献

- [1] Mantegna, R.N. (1999), Hierarchical structure in financial markets, Euro. Phys. J. B Vol.11, pp.193-197.
- [2] Bonanno, G., Lillo, F., and Mantegna, R.N.(2001) High-frequency cross-correlation in a set of stocks, Quantit. Finan., Vol.1, pp.96-104.
- [3] M.E.J.Newman (2003), The structure and function of complex networks, SIAM Review 45, pp.167-256.
- [4] ブイ・スアン・フォン, 榊原 源基, 上野 雄史, 斉藤 和巳 (2012), 株価変動のネットワーク分析, NetECO 9.
- [5] ブイ・スアン・フォン, 榊原 源基, 上野 雄史, 斉藤 和巳 (2013), Mantegna 法に基づく各国株価データの業種同類度分析, 第 76 回情処全大.