

Generative Model-based Analysis of Economic Temporal Networks (生成モデルによる経済テンポラルネットワークの研究)

数理情報学専攻 48-166229 溝上 力

指導教員 大西 立顕 准教授

1 導入

本研究では、国際貿易ネットワークと、株価相関ネットワークの解析を行った。まず、国際貿易ネットワークに非負値行列因子分解を適用し、分解された要素の類似度を解析することで、持続的な構造の抽出した [2]。さらに、株の分足の取引記録より相関ネットワークを構築しその構造変化を解析した。本発表では、株価相関ネットワークの解析について述べる。

2 株価相関ネットワーク

株価相関ネットワークの研究は、月毎のネットワークや、日足の価格を用いた解析が多く行われている [1]。本研究では、東証の 2000 年から 2013 年にかけての分足の株取引データから日毎のネットワークを構築し解析する。銘柄 (頂点) 間の枝の有無は、相関が統計的に有意であるかどうかに基づいて決めた。

まず、時刻によって価格の決定法や傾向が異なるため、全ての時刻における取引記録を扱うのは不適切である。取引開始 30 分と取引終了時刻の取引記録は取り除いて解析を行った。

次に、価格は非定常かつ銘柄によって平均が大きく異なるため、対数価格差

$$\log p_{i,t} - \log p_{i,t-\Delta t}, \quad (1)$$

の時系列を用いた。ここで、 $p_{i,t}$ は銘柄 i の時刻 t における価格を指す。

相関の算出に当たって、各銘柄は市場全体の影響を受けているため、銘柄間の相関を見るためにはこれを除去する必要がある。そのため、ここでは TOPIX に対する偏相関を用いて相関を算出した。

また、市場の銘柄は解析期間中に登場あるいは消滅するものが存在する。加えて、取引数が少ない銘柄は数週間価格がつかない場合もあるため、解析する期間中常に登場する中で取引頻度上位 200 銘柄を解析した。

最後に、相関係数の算出にあたって何分足のデータを扱うべきか解析を行った。相関が最も強い分足を用いるべきであり、1 分足から 5 分足まで比較したところ、1 分足 ($\Delta t=1$) のデータから相関を算出すると統計的に

有意な相関の割合が最大になることを確認した。

3 相関ネットワーク解析手法

株価相関ネットワークの解析のため、本研究では Edge Monitoring [4] を用いる。テンポラルネットワークの変化を検出する手法として、次数やクラスター係数などのネットワーク中心性に基づいた手法と、生成モデルに基づいた手法 [3] が挙げられる。それぞれ問題点として、ネットワークの構造に関する情報の多くが失われてしまうこと、構造推定に時間がかかる上にネットワーク構造に何らかの構造を仮定してしまっている点が挙げられる。Edge Monitoring は、生成モデルを元にした手法でありながら高速であり、クラスタ構造だけでなくスケールフリーネットワークなどのネットワーク構造を広く表現できることが特徴である。株価相関ネットワークは構造が非常に不安定かつ非常に長い時系列であるため、ネットワーク構造への仮定が少ない Edge Monitoring が有効であることが期待される。

次に、Edge Monitoring のアルゴリズムについて述べる (図 1)。まず、ネットワーク時系列の窓 w_s を抜き取る。各頂点ペアに独立にマルコフ過程を仮定し、パラメータの推定を行うことでその窓のネットワーク時系列の生成確率 $P(w_s)$ を算出する。そして、算出された生成確率間の距離 $d(P(w_s), P(w_{s+1}))$ を求め、この距離の値の外れ値を変化点とする。マルコフ過程における状態は各頂点ペア間の枝の有無で定まる。窓内における状態の推移の回数を用いて生成確率を算出する。

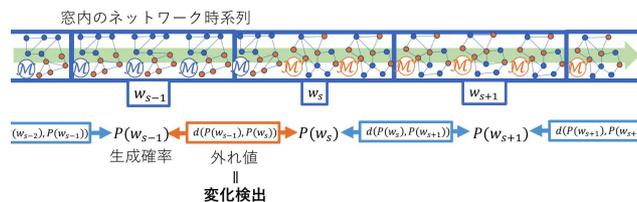


図 1. Edge Monitoring のプロセス

3.1 距離関数の提案

Edge Monitoring の距離関数としては従来 Kullback-Leibler ダイバージェンスとユークリッド距離が提案されている。これらの距離関数の問題点と

して、Kullback–Leibler ダイバージェンスは有界でないため z 値の算出ができない場合があること、ユークリッド距離は確率値をスケールでなく、差で評価するため、割合の変化を過小評価してしまうという問題点がある。本研究では、有界かつスケールで距離を評価する Jensen–Shannon ダイバージェンス

$$D_{JS}(p||q) = \frac{1}{2}D_{KL}(p||r) + \frac{1}{2}D_{KL}(q||r), \quad (2)$$

を提案する。ここで、 $r = \frac{1}{2}(p+q)$ であり、 $D_{KL}(p||q) = \sum_i p_i \log \frac{p_i}{q_i}$ である。

4 3 状態ネットワークモデル

ネットワークの頂点間の関係は枝の有無の 2 状態によって定まる。これは、相関ネットワークの例で考えると、正相関と相関という対極的な概念を有相関という一つの状態にまとめてしまっている。そこで本研究では、ネットワークにおいて 3 状態を考えることを提案する。そして、3 状態を考慮する形で Edge Monitoring を拡張する。Edge Monitoring を拡張するに当たって、マルコフ過程の状態遷移図を図 2 のような形で定める。

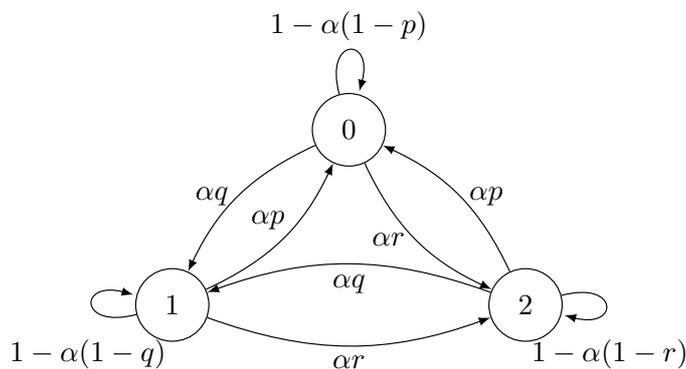


図 2. 3 状態に拡張した Edge Monitoring における状態遷移図

5 株価相関ネットワークへの適用結果

5.1 異なる距離関数の比較

まず、異なる距離関数による Edge Monitoring を適用した結果を図 3 に示す。提案した Jensen–Shannon ダイバージェンスによって 2008 年の金融危機や 2011 年の東北大震災といった実際の経済的な事件を検出していることが確認できる。

5.2 人工データにおける検証

人工的にネットワーク時系列を生成し、構造変化の時刻を検出できるか検証を行った。ここで、ネットワークの構造は絶えず変化しているものとし、その中である

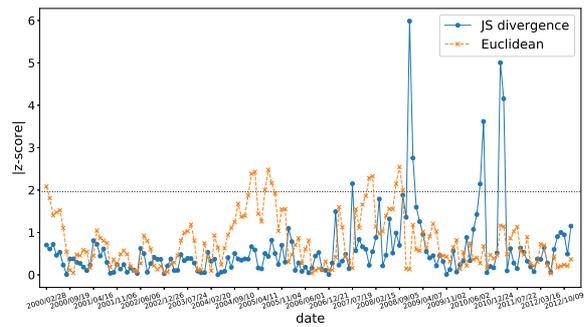


図 3. 異なる距離関数を用いた Edge Monitoring を株価相関ネットワークに適用した結果

時刻で頂点の 10% の枝の数が半分になる変化を検出できるか検証した。結果として、ユークリッド距離でなく Jensen–Shannon ダイバージェンスを用いることで構造変化を検出できることを確認した。

5.3 3 状態モデルにおける実験

3 状態モデルの有効性を確認する。3 状態を無相関、正相関、負相関と定めて Edge Monitoring と比較したところほぼ同一の結果を得た。このことから、負相関が株価相関ネットワークの構造変化を検出する上で重要でないと考えられる。次に、3 状態モデルを無相関、弱相関、強相関と定める。相関ネットワークでは、市場に強いショックが加わった時、強い相関として現れやすいため相関の強さでの分類は有効と考えられる。弱相関と強相関の閾値を相関係数の中央値である 0.18 として解析を行った。結果として、Edge Monitoring に比べて 2008 年の金融危機について 10 月の日経平均の暴落に直接対応して変化を検出することが示された。

参考文献

- [1] R. N. Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197, 1999.
- [2] C. Mizokami and T. Ohnishi. Revealing persistent structure of international trade by nonnegative matrix factorization. In *Complex Networks & Their Applications VI: Proceedings of Complex Networks 2017*, pages 1088–1099, 2018.
- [3] L. Peel and A. Clauset. Detecting Change Points in the Large-Scale Structure of Evolving Networks. In *29th AAAI conference on AI*, pages 2914–2920, 2015.
- [4] Y. Wang, A. Chakrabarti, D. Sivakoff, and S. Parthasarathy. Fast change point detection on dynamic social networks. In *International Joint Conference of Artificial Intelligence (IJCAI)*, 2017.