D-002

株価データを用いた企業分布の可視化に関する研究

Visualization of Firm Distribution using Stock Price Data

芝田 遥† 荒川 正幹†
Haruka Shibata Masamoto Arakawa

1. はじめに

現在日本は、円高、超低金利、高失業率、少子高齢化などの影響により厳しい経済環境に置かれており、将来に備えた資産形成のためには各個人の積極的な投資が必要とされている。しかし、家計における資産の大半は預貯金、保険、年金であり、十分な投資が行われていないのが現状である。特に株式投資については、証券優遇税制や手数料自由化、ネット証券の充実などにより、一般個人でも容易に投資できる環境が整いつつあるにも関わらず、依然として低調である。そこで我々は、株式投資を活発にし、日本経済の活性化に寄与することを目標とした研究を進めている。

本研究の目的は、株価データを用いた企業分布の可視化である。市場における過去の株価に対して多変量解析手法を適用し、企業の分布を二次元平面へと写像する。本稿では、この写像のために generative topographic mapping（GTM）[1]を用いることを提案する。また、写像の質を評価するための指標として中点 RMSE[2]を採用し、主成分分析および self organizing map（SOM）[3]による写像との比較を行った。

2. 手法

2.1 データ

企業の分布を可視化するため、東証一部上場企業の株価データを用いた。2001 年 1 月から 2011 年 12 月までの期間に東証一部に上場していた企業について、Yahoo! ファイナンス[4]より日足株価を収集した。そして、月毎の騰落率を求め、対数変換を行い説明変数とした。企業数は 1,210、次元数は 131 である。なお、データ収集のためのプログラムは、C#言語および VB Script を用いて開発した。

2.2 SOM

SOM[3]は、多次元データを二次元空間へ非線型写像するためのニューラルネットワークである。データ間の隣接関係を保持しつつ、多次元データの分布を二次元の平面上に可視化することが可能である。学習アルゴリズムが容易であり、多くの分野に応用されている。

2.3 GTM

GTM は Bishop ら[1]によって提案された教師なし学習手法であり、主成分分析や SOM と同様に多次元データの可視化に利用可能である。

図 1 に GTM による写像の概要を示す。GTM では、潜在変数空間に一定間隔で配置した点を、基底関数と重み行列を利用し入力データ空間へ非線型写像する。そして、写像された点を中心としたガウス分布を用いて、入力データの確率密度分布を推定する。最適化は、対数尤度を目的関数とし EM アルゴリズムを利用することで実現される。最適な写像が求まった後、データ点に対する各ガウス分布の寄与を求め、潜在変数空間上での座標値の重み付き平均を計算することで、データ点が潜在変数空間へ写像される。

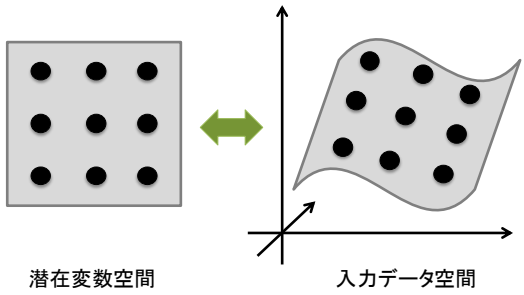

図 1　GTM の概要

2.4 中点 RMSE

高次元空間から低次元空間への写像を評価するひとつの方法は、次式で定義される RMSE を用いることである。

$$\mathrm{RMSE} = \left(\sum_{i}^{n} \| x_i - xc_i \|^2 / n \right)^{\frac{1}{2}} \qquad (1)$$

ここで、n はデータ点の数、x_i は i 番目のデータベクトル、xc_i は低次元空間へ写像したデータ点を再度高次元空間へと写像したときのデータベクトルである。写像の精度が高い場合、xc_i は x_i の近くに位置するため RMSE の値は小さい（図 2）。RMSE は写像の評価指標として有用ではあるが、RMSE が小さいからといって必ずしも良い写像であるとは限らない。例えば SOM においては、マップサイズを十分に大きくすることで RMSE を 0 にすることが可能である。

そこで本研究では、写像の質を評価するための指標として中点 RMSE の使用を提案する。中点 RMSE はデータ点間の全ての中点を対象として RMSE を求める指標である。写像が線型であれば、高次元空間における二点間の中点は、低次元空間においても対応する二点の中点となるため、RMSE とデータの分散で補正した中点 RMSE（修正済み中点 RMSE）の値は一致する。しかし写像が非線型の場合、必ずしもこの関係は成立せず、修正済み中点 RMSE の値は写像の非線形性に応じて変化する。

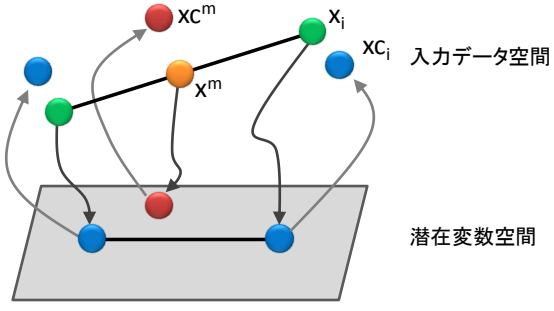

図 2　中点 RMSE の概要

† 宇部工業高等専門学校, Ube National College of Technology

3. 結果

3.1 写像手法の比較

東証一部上場企業の月次騰落率の対数を用いて、各写像手法による企業分布の可視化を行った。GTM および SOM については、いくつかのパラメータを設定することが可能である。本研究で用いたパラメータの値を以下に示す。

GTM： 基底関数の数 2, 3, 4, 5, 6, 7, 8, 9, 10
基底関数の標準偏差 0.1, 0.4, 0.7, 1.0, 1.3
SOM： 近傍半径の初期値 2, 4, 6, 8, 10
学習係数の初期値 0.005, 0.010, 0.015, 0.020

なお、GTM と SOM の両方において、繰り返し回数は 100 回、マップサイズは 30 とした。これらの条件を用いて各手法による二次元への写像を行い、RMSE および中点 RMSE を求めた結果を図 3 に示す。横軸が RMSE、縦軸が修正済み中点 RMSE であり、各点が写像を表している。

主成分分析においては RMSE=0.328 であった。線型写像であるため、修正済み RMSE の値も同様である。なお、第二成分までの累積寄与率は 12.7%であった。

これに対し、GTM および SOM においては、修正済み中点 RMSE の方が RMSE よりも大きく、全ての点が対角線の上側にプロットされる結果となった。両手法の学習においては、データ点を基準とした最適化が行われるため、これは妥当な結果である。RMSE は、写像の非線型性を増すことによって容易に減少させることが可能であるが、中点 RMSE は約 0.31 が下限であり、これを大きく下回る写像は得られなかった。また、GTM による写像は、SOM による写像と比較し、広い範囲に分布している。これは、GTM の方が多様な写像が可能であることを意味している。

最適な写像は目的によって異なるため厳密に定義することは困難であるが、RMSE と中点 RMSE の散布図を用いることで、適切な写像の選択が可能である。RMSE と中点 RMSE のバランスのとれた写像が、必要最低限の非線型性で高い精度を実現する写像である。本研究においては、RMSE=0.308、中点 RMSE=0.315 の写像を用いて企業分布の可視化を行った。

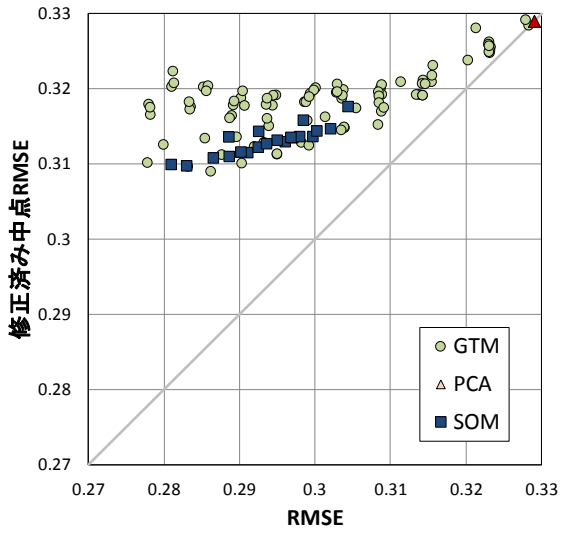

図 3　RMSE と中点 RMSE の散布図

3.2 GTM による可視化

GTM による企業分布の可視化結果を図 4 に示す。パラメータはマップサイズ 30、基底関数の数 4、基底関数の標準偏差 0.4、繰り返し回数 100 回である。主成分分析による可視化では、中心付近に多くの企業が集中して分布するのに対し、GTM による可視化ではマップ全体にほぼ均等な密度で分布していることが確認できる。

この写像の妥当性を検証するため、業種ごとの企業分布を確認した。業種分類としては、標準的な東証 17 業種を用いた。例として、銀行、小売業に分類される企業の分布を図 4 に示す。銀行は図の右下の領域に集中して分布しているのに対し、小売業は図全体に広く分布している。これは、銀行の株価が互いに類似した挙動を示すのに対し、小売業では大きく異なることを示している。小売業には非常に多様な企業が含まれていることを考慮すると、これは妥当な結果であるといえる。

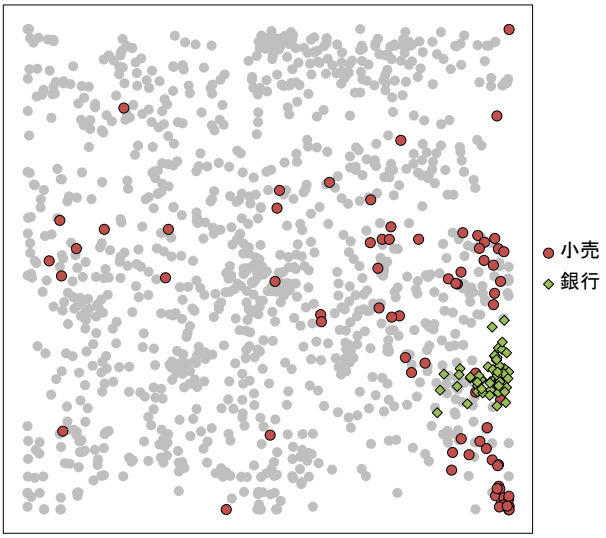

図 4　GTM による株価データの可視化

4. おわりに

株価データを用いて企業の分布を可視化するため、GTM、SOM、主成分分析による写像を行った。また、中点 RMSE を用いて各手法による写像の質を評価した。その結果、SOM、主成分分析に対する GTM の優位性が示された。

提案手法による企業分布の可視化は、株式投資における銘柄選択において非常に有用である。得られたマップを見ることで、各銘柄の特徴を容易に把握することができ、目的に合った効率的なポートフォリオの構築が可能となる。

参考文献

[1] C. M. Bishop, M. Svensén, C. K. I. Williams, “Developments of the Generative Topographic Mapping”, Neural Computation, Vol. 10, 215-234 (1998).
[2] 荒川 正幹, 宮尾 知幸, 船津 公人, “ドラッグライクネスモデルの構築とその可視化”, Journal of Computer Aided Chemistry, Vol. 9, 70-80 (2008).
[3] T. Kohonen, Self-Organizing Maps 3rd Ed., Springer, 2000.
[4] Yahoo! ファイナンス, http://finance.yahoo.co.jp/