

1 はじめに

1.1 背景

投資信託（ファンド）とは、販売会社を通じて複数の投資家から集めた資金を、資産運用の専門家が株式などに投資するよう指示し、運用成果をリターンとして投資家に分配する金融商品のことである。各ファンドは運用についての方針を持っており、それに従って複数の金融商品へ投資を行う。

ファンドはその運用方針によってパッシブとアクティブの二つに分けることができる。パッシブファンドは株式相場の指標として用いられるインデックス（日経平均株価や東証株価指数など）に連動する運用方針を持ち、アクティブファンドはインデックス以上の利益を得ることを目的とする運用方針を持つファンドである。ファンドの中には公表している運用方針と実際の運用内容が違うファンドが存在している。そのため、投資するファンドが公表している運用方針と実際の運用方針が一致しているかを投資家は知る必要がある。

過去の運用実績からファンドを分類する研究 (Alimi ら [1]) もあるが設立間もないファンドに適用できない。運用実績を必要としない手法として、武田ら [2] はファンドが投資している株式会社に着目し k -means 法でクラスタリングする手法を提案したがこの手法はアクティブファンドを分割できなかった。

1.2 目的

本論文は「実際の運用方針に基づいたファンドのグループ分け」を目的に、2 節でアクティブファンドもクラスタリングできる手法を提案し、3 節で提案手法によってファンドが運用方針に基づき分けられているかを運用実績であるトラッキングエラーを用いて評価する。

2 提案手法

運用方針が類似しているファンドはその投資先株式も類似していると考えられる。そこで本論文では「投資先株式が類似したファンド同士は実際の運用方針も類似している」と考え、運用方針が類似したファンドにグループ分けするために投資先類似度に着目したファンドのクラスタリング手法を提案する。本節では投資先株式会社の情報から投資先類似度の定義と、用いるクラスタリング手法としてスペクトラルクラスタリングの説明を行う。

2.1 投資先類似度

本論文ではファンド間の類似度を「ファンド間で共通する投資先株式銘柄数」とする。

ファンド v が投資している株式銘柄の集合を $I(v)$ 、そして v の投資割合が高い上位 10 の株式銘柄の集合

を $I_{10}(v)$ とする。この時ファンド v_i と v_j 間の類似度 $w(v_i, v_j)$ は以下のように定義する。

$$w(v_i, v_j) = |I_{10}(v_i) \cap I_{10}(v_j)|$$

2.2 スペクトラルクラスタリング

スペクトラルクラスタリング [3] ではグラフ分割の問題としてクラスタリングを行う。この手法ではデータをグラフとして与え、グラフの同じクラスターで構成されたサブグラフ内は密であり、異なるクラスターで構成されたサブグラフ間は疎であるよう分割することでクラスタリングを行う。スペクトラルクラスタリングのアルゴリズムを図 1 に示す。

入力: 類似度行列 W , 分割数 k

出力: 各対象が分割されたクラスター C_1, \dots, C_k

1. W から $D_{ii} = \sum_j W_{ij}$ となる対角行列 D を作る
2. W, D から $L = I - D^{-1/2} W D^{-1/2}$ となるラプラシアン行列 L を作る (I は単位行列)
3. L の固有値, 固有ベクトルを計算し固有値の値が小さい固有ベクトルを k 個列に並べた行列 U を作る
4. U をユークリッドノルムで割り正規化する
5. U に対して k -means 法を行いクラスター C を割り当てる

図 1: スペクトラルクラスタリングのアルゴリズム

本論文ではスペクトラルクラスタリングに与える類似度行列 W は以下のように定義される。

$$W = \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{pmatrix}$$

ここで $w_{ij} = w(v_i, v_j)$ は本節で説明したファンド v_i と v_j 間の類似度であり、 n は与えるファンドの総数である。

また与える分割数 k はインデックスごとに分割することを目的として、分割数を増やしていったときに日経平均株価 (日経 225) と東証株価指数 (TOPIX) の二つのインデックスが最初に分割される分割数とした。

3 実験と評価

本章では実験に用いるデータの説明を行った後、提案手法を用いて実際のファンドに対しクラスタリングを行い、結果をクラスターごとに色分けしたネットワーク図で示す。そして各クラスターをトラッキングエラーを用いて評価する。

なお実験には 2015 年 5 月に取得した Yahoo! ファイナンスで公開されている日本国内の株式会社のみで投資している国内株式型ファンドを使用した。

3.1 実験と評価

3.1.1 ネットワーク図 ファンドをクラスタリングし、各クラスタごとにノード(ファンド)を色分けしたネットワーク図を示す。エッジの重みは前節で定義した投資先類似度と同義である。

ネットワーク図は重みが大きいノード間は近くに、重みが小さいノード間は遠くに描画される。そのため投資先銘柄が類似したファンド群は近くに、そうでないファンドは遠くに描画がされる。

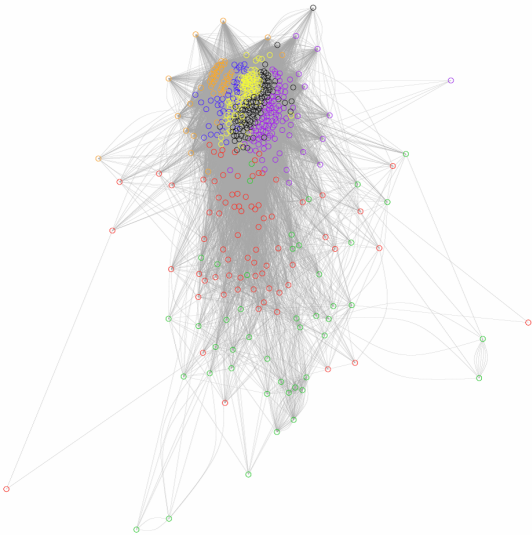


図 2: クラスタごとに色分けしたネットワーク図

スペクトラルクラスタリングでファンドを分割し対応するノードを色分けしたネットワークが図 2 である。その時の分割数は 7 であった。また、TOPIX ファンド群はすべて黄クラスタ、日経 225 ファンド群はすべて橙クラスタにまとめられた。

ネットワーク図におけるノードが密集した部分が似通った株式銘柄に投資しているパッシブファンド、その他がアクティブファンドに近いファンド群に相当する。図 2 からスペクトラルクラスタリングはパッシブファンドを青黄紫橙黒の 5 クラスタに、アクティブファンドを赤緑の 2 クラスタに分割しており、 k -means 法を用いた武田の手法でできなかったアクティブファンドの分割も行えている。

3.1.2 トラッキングエラー トラッキングエラーとは指定した期間において特定のインデックスとファンドのリターンの変動にどれだけ差異があるかを示す尺度である。算出したいファンド F の n ヶ月リターンを Fr_1, \dots, Fr_n 、インデックス I の n ヶ月リターンを Ir_1, \dots, Ir_n とすると、ファンド F のインデックス I とのトラッキングエラー $E(F, I)$ は以下のように計算される。

$$E(F, I) = \sqrt{\frac{1}{n} \sum_{i=1}^n (Fr_i - Ir_i)^2 \times 12}$$

図 3 の縦軸は各ファンドごとに「TOPIX、日経 225、JPX400、ジャスダック指数、東証マザーズ指数」の 5

つのインデックスとの 6 ヶ月間のトラッキングエラーを計算し、その中で最小のトラッキングエラーを示しており、折れ線グラフの色は図 2 でのノードの色に対応している。

図 3 からアクティブファンドが属する赤クラスタと緑クラスタがパッシブファンドが属するクラスタよりもトラッキングエラーの値が高かった。パッシブファンドの場合、インデックスに連動させることを目的にしているのでトラッキングエラーの値は小さくなる。対してアクティブファンドはインデックスを超過するリターンを目標としているのでパッシブファンドと比べインデックスとのトラッキングエラーの値は大きくなる。トラッキングエラーを示したグラフから投資先上位 10 銘柄の情報だけで運用実績から見てもアクティブであるファンド群とパッシブであるファンド群に分けることができたと言える。

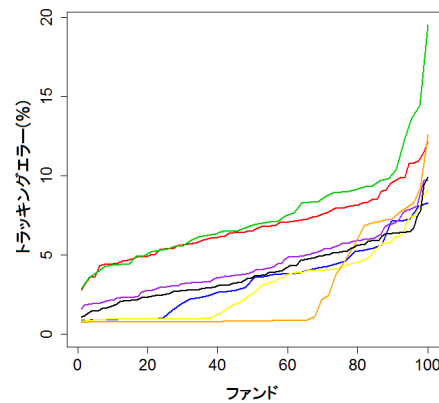


図 3: ファンドのトラッキングエラー

4 おわりに

ファンドの中には公表されている運用方針と実際の運用方針が異なるものが存在する。本論文では実際の運用方針に基づいたファンドの分類を行うために投資先類似度に基づきファンドをクラスタリングする手法を提案した。

提案手法を用いることで、運用間もないファンドであっても投資家は自分の投資方針とマッチした運用をしているファンドを見つけることができ、ファンドの運用会社は公表している運用方針と実際の運用方針が一致していることを投資家へ示すことができる。

参考文献

- [1] A. Alimi, M. Zandieh, and M. Amiri: "Multi-objective portfolio optimization of mutual funds under downside risk measure using fuzzy theory," International Journal of Industrial Engineering Computations, 3(5):859–872 (2012)
- [2] 武田祐太, 松井藤五郎, 犬塚信博, 「投資信託と銘柄のネットワーク構造に基づいた投資信託の分類」, 第 75 回情報処理学会全国大会, 5M-9 (2013)
- [3] U. von Luxburg, "A tutorial on spectral clustering," Statistics and Computing, 17(4):395–416 (2007)