

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

武田 祐太[†]松井 藤五郎[‡]犬塚 信博^{*}[†]名古屋工業大学 工学部 情報工学科[‡]中部大学 生命健康科学部 臨床工学科^{*}名古屋工業大学大学院 工学研究科 情報工学専攻

1 はじめに

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や債券などの金融資産、あるいは不動産などに投資するよう指図し、運用成果を投資家に分配する金融商品のことである。（図1参照）

ファンドは自らが公示した投資方針に従って複数の金融商品へ投資を行う。そのため、投資家は投資先とするファンドを選択する際、運用方針の決定やリスク管理のためにファンドの性格を把握しておく必要がある。しかし、現状では類似したファンドを把握することは難しい。

そこで我々は、複数のデータ間の関係をリンクとして捉え、ノードとリンクによって関係をネットワークとして表現したネットワーク構造に基づく分析に注目した。この構造に着目した研究として、友人関係におけるコミュニティ抽出などの試みが行われている。

本稿では、投資信託間のネットワーク構造によってその特徴分析を試みる。

2 投資信託情報とその構造

ファンドについての基本的な情報を得るため、Yahoo!ファイナンスの提供する MORNING STER の投資信託情報の中から、投資割合上位 10 銘柄の公開されている国内株式型投資信託 691 個を抽出した。このデータには、各ファンドの投資割合上位 10 銘柄の銘柄名、銘柄業種、投資割合、ファンドの純資産や分配金、取引手数料などの情報が含まれている。

ファンドを丸、銘柄を四角で表現し、ファンドが投資先としている銘柄についてはファンドと銘柄の間にリンクを張ったネットワークを図2に示す。ファンド間や銘柄間にはリンクが存在しない二部グラフ構造となっている。

3 提案手法

図2より、投資先銘柄の似通ったいくつかのファンド群が存在していると考えられる。そこで我々は、まず次の2つのタイプ合計4つの手法を用いてファンドを分析した。

3.1 属性ベクトルを用いた手法

第1のタイプは、ネットワーク構造を用いず、銘柄 769 について各ファンドが投資割合上位 10 銘柄として投資しているか否かを 0/1 で表現した長さ 769 のベクトルと、それらの間のユークリッド距離に基づく方法である。我々はこのデータを対象に、(1) k -means 法および (2) 階層型クラスタリングであるウォード法 [1] によるクラスタリングを行った。

Clustering of Mutual Fund Based on the Network Structure of Funds and Brands

[†]Yuta TAKEDA, [‡]Tohgoroh MATSUI, ^{*}Nobuhiro INUZUKA

[†]*Nagoya Institute of Technology [‡]Chubu University

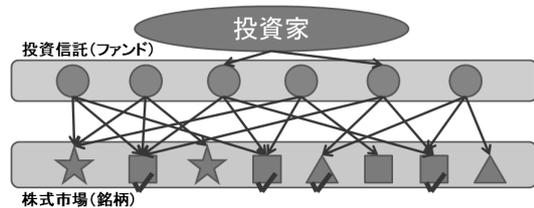


図1: ファンドの仕組み

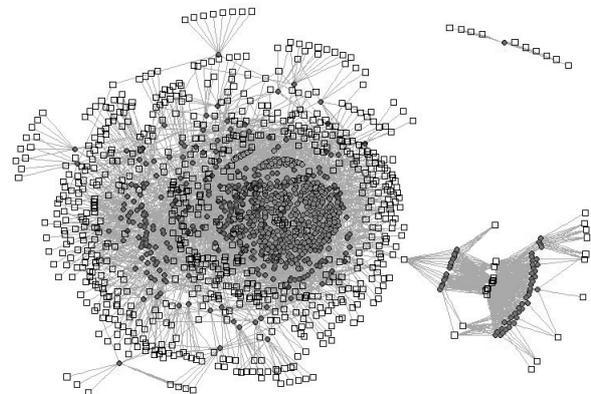


図2: ファンドと銘柄のネットワーク

3.2 ネットワーク構造を用いた手法

第2のタイプは、ネットワーク構造を利用したものである。ファンドをノードとし、5つ以上の銘柄に共通して投資しているファンド間にリンクをひいたネットワークについて、(3)GNアルゴリズム [2] を適用した。また、次に述べる (4)HITS アルゴリズム [3] と k -means 法を組み合わせる手法を用いた。ここでは HITS アルゴリズムによって多くのファンドが投資しているような人気の高い銘柄への投資を行っておらず、他のファンドと類似性の低いポートフォリオの構成となっているファンドを「独自性の高いファンド」として選別した。

HITS アルゴリズムでは、まず初めに各ファンドに hub、各銘柄に authority という値を与える。次に、hub の初期値を 1 とし、各 authority にはリンクの繋がっているファンドの hub の合計を格納する。そして、各 hub にリンクの繋がっている銘柄の authority の合計を格納する。この操作によって得られた hub 値を有するファンド数についてのヒストグラムを図3に示す。このとき、250 を閾値にして、それ以下のファンドを独自性の高いファンドとした。(4)の方法では、それ以外のファンドに k -means 法を適用した。

4 評価

評価のため、分散投資を目的として各手法に基づいて k 個のファンド ($k = 5, 10, 20, 30, 50$) から成るポートフォリオを

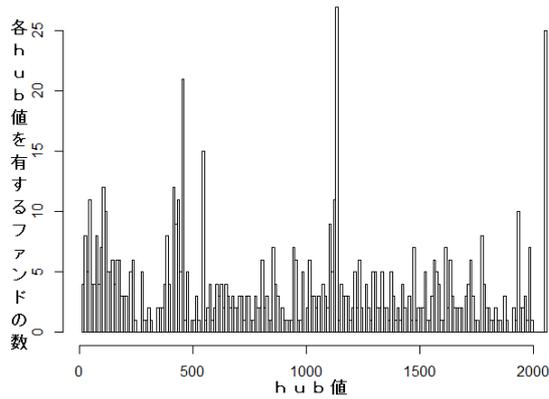


図 3: hub 値のヒストグラム

作成する。このため (1)~(4) の方法でファンドを k グループに分け、そこからランダムに 1 つずつファンドを選択し、これらによってポートフォリオを構成する。各手法が類似したファンドをまとめる効果があるならば、このポートフォリオによって幅広い銘柄、業種への分散投資が可能となる。

(3) の手法ではモジュラリティ最大となった分割 (実験では 209) からランダムに k 個のクラスタを選択した。

(4) の手法では、独自性の高いファンドとそうでないファンドの比率はおよそ 1 : 4 であったため、最初に独自性の高いファンド群から $\frac{1}{5}k$ 個のファンドをランダムで選び出した。次に、独自性の高くないファンド群を k -means 法によって $\frac{4}{5}k$ 個に分割した。

構成したポートフォリオが、銘柄と業種の全体数に対してどれほどの割合で投資先としてカバーできているかどうかを比較した結果を図 4、図 5 に示す。ここで全体数とは、銘柄で 769、業種で 34 である。比較のため、(5) 全体からランダムに k 個のファンドを選択した場合 (ランダム) と (6) 独自性の高いファンド群のみからランダムに k 個のファンドを選択した場合 (HITS+ランダム) についても計算した。

また、構成したポートフォリオがどれほど均等に投資できているかどうかを検証するため、ポートフォリオの構成ファンド全体が各銘柄・業種に対して投資している数を重みとし、投資先となり得る最大数である $k \times 10$ (業種の場合は全業種数である 34) 次元で構成されたベクトルについて分散を計算した。この値を比較した結果を図 6、図 7 に示す。

5 おわりに

本稿では、評価尺度として、銘柄と業種の投資先としてのカバー率およびその偏りを用いた。その結果、HITS アルゴリズムと k -means 法を組み合わせた手法により、他の手法よりも優れた結果が得られた。これは、初めから全体に対して k -means 法を適用してしまうと、クラスタ内の他のノードとほとんど類似していない独自性の高いファンドまで無理矢理クラスタに含まれてしまうからであると考えられる。そこで、あらかじめ独自性の高いファンドを除いてからクラスタリングをすることで、より類似性の高いもの同士のみで構成されたクラスタが得られた。独自性の高いファンドを先に選別しておいた後に、独自性の高いファンド群とクラスタのそれぞれからファンドを選択してポートフォリオを構成する

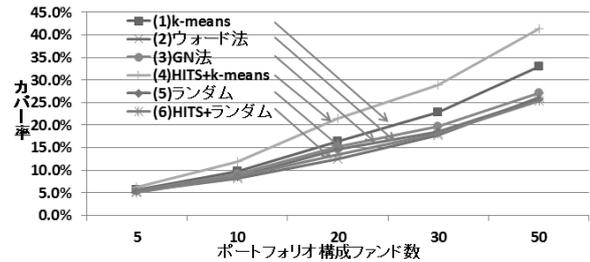


図 4: 銘柄カバー率の比較 (50 回平均)

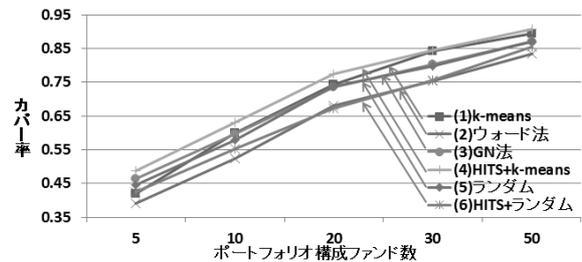


図 5: 業種カバー率の比較 (50 回平均)

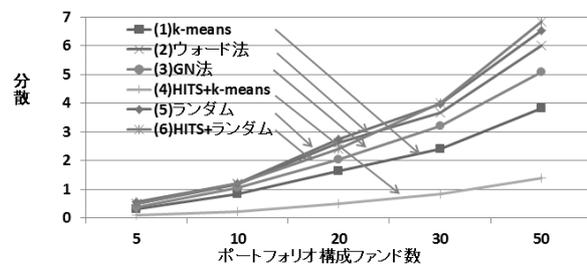


図 6: 銘柄の偏り (分散) 比較 (50 回平均)

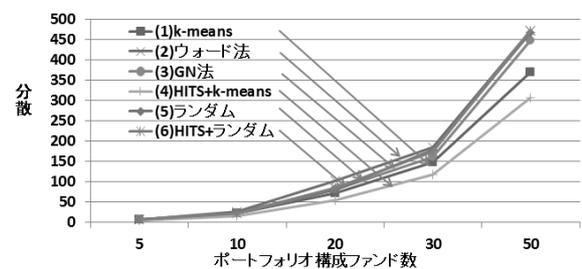


図 7: 業種の偏り (分散) 比較 (50 回平均)

ことで、分散した投資先を得られたと考えられる。

この実験では、投資したいファンドの数 k を用いて分割を行ったが、これが必ずしも最適な分割数であるとは限らない。その最適な分割数の検討が今後の課題として挙げられる。また、各 hub 値を有するファンドのヒストグラムの目視によって独自性の高いファンドとなる閾値を決定したが、この閾値の決定方法についての更なる検討も必要である。

参考文献

- [1] 神尾 敏弘, データマイニング分野のクラスタリング手法 (1), 人工知能学会誌, pp.59-65, (2003)
- [2] M.E.J.Newman and M.Girvan, "Finding and evaluating community structure in networks". Phys. Rev. E (2004)
- [3] Jon Kleinberg, "Authoritative sources in a hyperlinked environment", ACM-SIAM Symposium on Discrete Algorithms, pp.668-677 (1998)